THÈSES

présentées à

L'UNIVERSITÉ DES SCIENCES ET TECHNIQUES
DE LILLE FLANDRES-ARTOIS

pour obtenir le grade de

DOCTEUR ÈS SCIENCES MATHÉMATIQUES

par

Michel DELECROIX

1ere Thèse : **"SUR L'ESTIMATION ET LA PRÉVISION NON-PARAMÉTRIQUE DES PROCESSUS ERGODIQUES"**

2eme Thèse : **"CYCLES ÉCONOMIQUES ET TACHES SOLAIRES"**

Soutenues le 2 juillet 1987 devant la Commission d'Examen :

Président : H. CAUSSINUS (Université de Toulouse)

Directeur de Recherche et
Rapporteur : D. BOSQ (Université de Paris VI)

Rapporteurs { P. DEHEUVELS (Université de Paris VI)
C. GOURIEROUX (Université de Lille I)

Examinateurs { L. DEVROYE (Mac Gill University, Montréal)
J.M. GRANDMONT (Ecole Polytechnique)
de Lille I)

THÈSES

présentées à

L'UNIVERSITÉ DES SCIENCES ET TECHNIQUES
DE LILLE FLANDRES-ARTOIS

pour obtenir le grade de

DOCTEUR ÈS SCIENCES MATHÉMATIQUES

par

Michel DELECROIX

1ere Thèse : **"SUR L'ESTIMATION ET LA PRÉVISION NON-PARAMÉTRIQUE DES PROCESSUS ERGODIQUES"**

2eme Thèse : **"CYCLES ÉCONOMIQUES ET TACHES SOLAIRES"**

Soutenues le 2 juillet 1987 devant la Commission d'Examen :

Président : H. CAUSSINUS (Université de Toulouse)

Directeur de Recherche et Rapporteur : D. BOSQ (Université de Paris VI)

Rapporteurs : P. DEHEUVELS (Université de Paris VI)
C. GOURIEROUX (Université de Lille I)

Examinateurs : L. DEVROYE (Mac Gill University, Montréal)
J.M. GRANDMONT (Ecole Polytechnique)
P. JACOB (Université de Lille I)

A mes parents,

A Danièle,

A Xavier, Philippe et Olivier .

Monsieur le Professeur Henri Caussinus nous a fait le grand honneur d'accepter la présidence du jury de cette thèse. Nous tenons à lui exprimer notre respectueuse gratitude.

Sans Monsieur le Professeur Denis Bosq, ce travail n'aurait sans doute pas été entrepris. Il nous a puissamment aidé à le mener à bien, par une direction de recherche patiente et efficace. Notre goût de la recherche en mathématique et notre intérêt pour la statistique non-paramétrique ont été éveillés par lui. Nous tenons à lui exprimer ici notre profonde reconnaissance.

Monsieur le Professeur Paul Deheuvels nous a fait l'honneur de bien vouloir juger notre travail. Qu'il trouve ici l'expression de notre profonde reconnaissance.

Monsieur le Professeur Christian Gouriéroux a toujours fait preuve à notre égard d'une très grande et très affable disponibilité. Les nombreuses discussions que nous avons eues avec lui nous ont permis d'enrichir la dernière partie de la thèse et d'améliorer, en de nombreux points, la présentation du travail, qu'il a bien voulu juger. Nous l'en remercions très vivement.

Nos remerciements vont également à Monsieur le Professeur Luc Devroye qui a bien voulu faire partie de la commission d'examen malgré ses nombreuses occupations.

Monsieur le Professeur Jean-Michel Grandmont a bien voulu nous proposer un deuxième sujet de thèse en économie mathématique, en discuter avec nous, et juger la synthèse que nous en avons faite. Il nous a révélé cette matière par le choix d'un thème particulièrement motivant. Qu'il veuille bien trouver ici l'expression de notre profonde reconnaissance.

Monsieur le Professeur Pierre Jacob a accepté de participer à notre jury. Ses conseils nous ont été précieux, notamment pour la fin de la première partie du travail. Nous l'en remercions vivement.

Tous les membres du laboratoire de Probabilités Statistique de Lille I ont su nous entourer de leur sympathie pendant l'élaboration de ce travail. Qu'ils en soient tous remerciés et particulièrement notre ami Michel Carbon, qui a su partager bureau, problèmes mathématiques, découragements et espoirs.

Arlette Lengaigne a accepté, avec un sourire inébranlable, les manuscrits les plus bruts pour en faire un travail impeccablement dactylographié, dans les délais réduits qu'impose l'époque difficile où chacun entend "boucler" avant les vacances. Il n'est que justice de l'en remercier vivement.

Nous remercions enfin tout le personnel technique de notre U.F.R. qui a assuré la mise en page et la présentation finale de cet ouvrage.

S O M M A I R E

N.B. - le • signifie la présence d'une page bleue intercalée entre les diverses parties de la thèse.

INTRODUCTION.

Le travail présenté comprend quatre parties distinctes qui concernent toute la statistique non-paramétrique des processus, chacune éclairant un aspect particulier de ce thème général. Dans cette introduction, nous nous efforcerons d'abord de dégager les idées nous ayant amené à démontrer les divers résultats obtenus, puis nous présenterons de façon systématique et plus complète chacune des quatre parties.

PRESENTATION GENERALE

Nous abordons à la fin du travail le problème bien connu de la prédiction de l'élément X_{T+h} d'une série temporelle à partir des valeurs $X_1, \ldots, X_T$ déjà observées à l'instant T. Suivant les travaux de Collomb (1), Sarda et Vieu (2), Härdle et Vieu (3), Doukhan et Portal (4) Bosq (5), par exemple, on se propose d'approcher X_{T+h} par une moyenne pondérée aléatoire des X_{t+h} , $r \leq t \leq T-h$. La pondération $\hat{\alpha}_{t,T}$ attachée à X_{T+h} se base sur la proximité de la séquence $(X_t, \ldots, X_{t-r+1})$ avec la dernière séquence observée de même longueur $(X_T, \ldots, X_{T-r+1})$. Le prédictogramme (cf. Collomb 1978 (6)) est le plus élémentaire des prédicteurs basés sur ce principe, son lissage amène à définir une classe de prédicteurs qui s'écrivent

$$\text{(I)} \qquad N_T((X_T, \ldots, X_{T-r+1}))/D_T((X_T, \ldots, X_{T-r+1})) \quad ,$$

où D_T et N_T sont des estimateurs respectifs de la densité g des r-uples $(X_t, \ldots, X_{t-r+1})$, et du produit $R.g$, où R est la régression de X_{T+h} sur $(X_T, \ldots, X_{T-r+1})$.

Pour obtenir la convergence du prédicteur et justifier la méthode utilisée, on doit donc démontrer des convergences d'estimateurs de paramètres fonctionnels, en particulier de densités de probabilité.

L'estimation des densités d'un processus stationnaire trouve ainsi un champ supplémentaire d'application, outre les intérêts propres qu'elle possède au niveau des identifications de modèles, et, plus généralement, dans la détermination des lois de dimensions finies du processus observé, donc de l'obtention de sa loi de probabilité.

Ce travail traite essentiellement de ces problèmes, liés, de "statistique non-paramétrique des processus". Les trois premières parties concernent plus spécifiquement l'estimation de paramètres fonctionnels ; la quatrième qui complète le travail avec l'estimation de la régression est axée sur sur les propriétés des prédicteurs non-paramétriques qu'on en déduit, et l'étude de leurs performances par rapport aux prédicteurs paramétriques classiques.

Dans les trois premières parties, nous nous sommes efforcés d'apporter plus particulièrement trois types d'innovations par rapport aux résultats pré-existants :

a) Proposer une approche unifiée de tous les problèmes d'estimation évoqués ci-dessus. Les estimateurs usuels de la densité, par la méthode du noyau ou celle des fonctions orthogonales, construits récursivement ou non, ceux de la fonction de répartition, ou des dérivées de la densité (sans souci d'exhaustivité) peuvent tous s'écrire sous la forme :

$$\sum_{t=1}^{T} K_{t,n}(X_t) \quad ,$$

où les $K_{t,n}$, $n \geq 1$, $1 \leq t \leq n$, forment une suite de séquences d'applications à valeurs dans un espace fonctionnel. On a donc cherché des conditions générales (sur les $K_{t,n}$) concrètement vérifiables, sous lesquelles ce type de sommes convergeaient. C'est l'objet de la première partie de cette thèse.

b) En proposer une approche fonctionnelle. Pour obtenir la convergence L^p de l'estimateur d'un élément g de cet espace, on recourt fréquemment à une évaluation de l'écart entre l'estimateur et g, en chaque point x, pour en déduire par intégration l'écart L^p-global cherché, (avec des exceptions notables comme la convergence L^2 de l'estimateur d'une densité par la méthode des fonctions orthogonales). On verra Bosq et Lecoutre (7) pour une synthèse, dans le cas de l'échantillon.

Nous préférons établir les convergences directement dans l'espace fonctionnel (supposé muni d'une structure d'espace de Banach) auquel appartient la fonction estimée, ce qui amène à traiter des problèmes techniques nouveaux pour le sujet étudié, résolus par l'utilisation systématique des opérateurs telle que la préfiguraient en particulier les travaux de Bosq (8)).

c) Proposer des résultats valables pour une classe de processus beaucoup plus large que celles qu'on a utilisées jusqu'alors, essentiellement celle des processus ergodiques, dont nous allons maintenant justifier l'introduction.

Les travaux récents sur le sujet concernent en effet l'estimation d'une densité ou d'une régression lorsque le processus observé (X_t), $t \in \mathbb{Z}$ est Φ-mélangeant, (Sarda et Vieu (2), Collomb (1), Doukhan (9)) ou L^2-mélangeant (Peligrad 1987 (10)). C'est-à-dire qu'en appelant $\mathcal{M}_{-\infty}^{t}$ et $\mathcal{M}_{t'}^{\infty}$, les tribus respectivement engendrées par les X_s, pour $s \leq t$ et $s \geq t'$, on a respectivement :

$$\text{(II)} \quad \lim_{n\to\infty} \Phi(n) = 0 \,, \quad \text{avec } \{ \sup_{t} \sup_{\substack{A\in\mathcal{M}_{-\infty}^{t} \\ B\in\mathcal{M}_{t+n}^{\infty}}} |P(A/B) - P(A)| \} = \Phi(n) \quad (\Phi\text{-mélangeance})$$

$$\text{(III)} \quad \lim_{n\to\infty} \beta(n) = 0 \ , \quad \text{avec} \quad \{\sup_t \sup_{\substack{X\in M_{-\infty}^t \\ Y\in M_{t+n}^{\infty}}} \rho(X,Y)\} = \beta(n) \quad (L^2 \text{ mélangeance})$$

($X \in M_{-\infty}^t$ signifiant selon l'abus usuel que X est une variable aléatoire réelle $M_{-\infty}^t$ mesurable, ρ représentant le coefficient de corrélation).

Ces hypothèses de mélange s'avèrent souvent trop fortes puisque (cf. Gastwirth et Rubin (1)) les processus auto-régressifs par exemple ne les vérifient pas forcément. On peut les affaiblir encore en imposant au processus une condition dérivée d'une mesure intrinsèque du degré d'indépendance asymptotique des variables le composant (cf. Bosq - Delecroix 1985 (12)), ou bien la "forte mélangeance" de Rosenblatt i.e.

$$\text{(IV)} \qquad \lim_{n\to\infty} \alpha(n) = 0 \quad \text{avec} \quad \{\sup_t \sup_{\substack{A\in M_{-\infty}^t \\ B\in M_{t+n}^{\infty}}} |P(A \cap B) - P(A)\ P(B)|\} = \alpha(n)$$

(voir à ce sujet Delecroix (13), ou Doukhan (9)). Mais cette solution ne résout pas encore totalement le problème, car certains processus linéaires ne sont pas même α-mélangeants (cf. Withers (4), Tuan D. Pham et Lanh T. Tran (15), Akonom (16)). Nous avons donc cherché des conditions naturelles plus faibles, si possible minimales, pour lesquelles les estimateurs usuels de paramètres fonctionnels, en particulier la densité, convergeaient.

Entre autres références, les travaux de M. Rosenblatt ((17), (18), (19)) éclairent ce problème de définition d'une indépendance asymptotique entre variables d'un processus. Ainsi, dans le cas d'un processus strictement stationnaire et Markovien la "forte mélangeance" est très proche de l'"uniforme ergodicité" (voir (18)) pour une définition) qui implique elle-même l'ergodicité (simple) du processus, condition beaucoup plus faible que les précédentes, et dont la définition classique est :

$$\text{(V)} \quad \forall A\ ,\ A \in \mathcal{N}_{-\infty}^{0}\ ,\quad \forall B\ ,\ B \in \mathcal{N}_{0}^{\infty}\ ,\quad \lim_{n\to\infty} \frac{1}{n} \sum_{1}^{n} P(A \cap \tau^{k} B) = P(A)\ P(B)\ .$$

(τ est l'opérateur de translation défini sur l'espace produit infini E des états du processus, $\mathcal{N}_{-\infty}^{0}$ et $\mathcal{N}_{0}^{\infty}$ les tribus engendrées par les applications coordonnées x_s sur E , respectivement pour $s \leqslant 0$ et $s \geqslant 0$).
L'ergodicité est évidemment impliquée par tous les types de mélangeance (cf. Rosenblatt (17), ou Ash et Gardner (20)), et paraît être une condition minimale d'obtention de lois des grands nombres, a fortiori de convergences d'estimateurs fonctionnels. C'est pourquoi nous avons cherché à obtenir nos théorèmes sous cette hypothèse, ou des hypothèses équivalentes.

QUELQUES CONSIDERATIONS TECHNIQUES

L'introduction de l'hypothèse d'ergodicité fait perdre les deux "piliers" techniques utilisés dans les démonstrations relatives aux processus mélangeants. Le premier consiste en des majorations de covariances des fonctions de variables d'indices différents X_t , et $X_{t'}$ (voir Doukhan et Portal (4) à ce sujet, ou Bosq (21) pour une approche hilbertienne).
Le second est l'utilisation d'inégalités de type Bernstein (Bosq (22), Delecroix (13), Collomb (1), Peligrad (10), et Carbon (23) pour la plus achevée). Aucun n'admet d'équivalent sous la seule hypothèse d'ergodicité, alors qu'une étude bibliographique montre la difficulté de transposer à des suites de séquences de variables de lois différentes, (les $K_{t,T}(X_t)$), les théorèmes ergodiques à valeurs dans les espaces de Banach (cf. Mourier (24) en référence classique, Beck (26), et Krengel (25), pour une synthèse).

Les articles de Warren et Beck (27), Gyorfi, Gyorfi et Vajda (28), font cependant ressortir un principe qui permet de résoudre naturellement le problème de covariance évoqué ci-dessus. Le premier traite de variables à valeurs dans un Banach et "orthogonales" (selon une définition particulière), le second de variables à valeurs dans un espace de Hilbert. Cette hypothèse permet de faire apparaître des variables orthogonales par une simple soustraction de leurs espérances conditionnelles.

C'est la méthode que nous avons retenue, car elle permet d'autre part d'introduire naturellement l'hypothèse souhaitée d'ergodicité. Supposant en effet les $K_{t,n}$ à valeurs dans un espace de Hilbert H, on décompose $\sum_{t=1}^{T} K_{t,T}(X_t)$ en

$$\text{(VI)} \qquad \sum_{t=1}^{T} \{K_{t,T}(X_t) - E^{\mathcal{F}_{t-1}}(K_{t,T}(X_t))\} + \sum_{t=1}^{T} E^{\mathcal{F}_{t-1}}(K_{t,T}(X_t))$$

L'orthogonalité des variables introduites dans la première somme permet alors de prouver aisément sa convergence à zéro, sous les hypothèses adéquates. Reste le problème de la convergence de $\sum_{t=1}^{T} E^{\mathcal{F}_{t-1}}(K_{t,T}(X_t))$. Nous montrons que celle-ci s'obtient sous une condition relative aux densités de transition $g_{X_t}^{\mathcal{F}_{t-1}}$ du processus (densité de X_t relativement à la tribu engendrée par X_{t-1}, X_{t-2}, ...) vérifiée par les processus ergodiques sous de faibles conditions additionnelles. Ceci résout bien notre problème.

Dans le cas où le paramètre estimé n'est pas élément d'un espace de Hilbert (par exemple la densité considérée comme élément d'un espace L^P, $P \neq 2$), on utilisera encore la décomposition(IV), des artifices de démonstration permettant d'utiliser une pseudo-orthogonalité des variables composant la première somme.

Finalement, l'ordre d'exposition des résultats obtenus suivra la logique mathématique. En une première partie, on étudie, à partir de la décomposition (IV), la convergence asymptotique des sommes $\sum_{t=1}^{T} K_{t,T}(X_t)$, les $K_{t,T}$ étant à valeurs dans un espace de Hilbert. On montre ensuite comme annoncé que les conditions nécessaires à cette convergence sont usuellement vérifiées par les processus ergodiques. On donne enfin deux exemples directs d'application des théorèmes obtenus (estimation de la loi et de la fonction de répartition). Dans la seconde partie on étudie la convergence dans L^1 et L^2 des estimateurs usuels de la densité vers la limite g des $g_{X_t}^{F_{t-1}}$, et dans la troisième partie les convergences ponctuelles uniformes de ces estimateurs sous les hypothèses d'ergodicité introduites au départ, en utilisant les idées développées dans la première partie. Enfin, la quatrième partie comprend un théorème de convergence de l'estimateur de la régression N_T/D_T défini en (I), suivi d'une étude de convergence du prédicteur correspondant, dans L^2 et uniforme. Elle se termine par une comparaison pratique du prédicteur non-paramétrique utilisé avec les prédicteurs paramétriques usuels.

Nous allons maintenant présenter plus précisément le détail des résultats obtenus, partie par partie, pour terminer cette introduction.

PREMIERE PARTIE

La première partie de la thèse concerne le devenir asymptotique d'estimateurs de la forme $\sum_{t=1}^{T} K_{t,T}(X_t)$, les X_t formant un processus à valeurs dans $\mathbb{R}^d$, et les $K_{t,T}$, $T \geq 1$, $1 \leq t \leq T$ une suite de séquences d'applications à valeurs dans un Hilbert séparable H , mesurables par rapport aux tribus boréliennes de $\mathbb{R}^d$ et H. Nous notons $< , >_H$ et $\| \, \|_H$ les produit scalaire et norme de H , et rappelons que les estimateurs usuels de la densité, de la fonction de répartition, etc ..., se mettent sous cette forme.

α) Dans la proposition I, nous étudions la convergence à 0 , presque sûre et en moyenne, de la quantité :

$$\text{(VII)} \qquad (A_1(T))^2 = \left\| \sum_{t=1}^{T} \{K_{t,T}(X_t) - E^{\mathcal{F}_{t-1}}(K_{t,T}(X_t))\} \right\|_H^2$$

L'orthogonalité des variables utilisées permet de conclure pour la convergence en moyenne quand la somme des normes des $K_{t,T}(X_t)$, $1 \leq t \leq T$ tend vers 0 ; et la convergence presque sûre résulte d'un lemme de Van Ryzin (29), de type martingale, qui permet de conclure sous la seule condition :

$$\text{(VIII)} \qquad \sum_{T=1}^{\infty} E\{|E^{\mathcal{F}_T}(A_1^2(T+1)) - A_1^2(T)|\} < \infty \quad .$$

Nous avons alors dégagé des conditions plus fortes impliquant les conditions précédentes,donc les convergences cherchées. Nous les notons H1), H2), H3) dans le texte. Elles ne sont pas nécessaires stricto sensu, mais ont l'avantage d'être assez facilement vérifiables pour les diverses familles $K_{t,T}$ utilisées ensuite, c'est pourquoi nous les avons privilégiées.

La proposition 1 permet d'affirmer que, sous H1), H2), H3), la limite de $\sum_{t=1}^{T} K_{t,T}(X_t)$ est en fait celle de $\sum_{1}^{T} E^{F_{t-1}}(K_{t,T}(X_t))$. On suppose alors que les densités de transition $g_{X_t}^{F_{t-1}}$ (F_{t-1} engendrée par X_{t-1}, X_{t-2}, ...) existent, relativement à une mesure μ σ-finie sur $\mathbb{R}^d$ telle que les $L^p(\mu)$, $1 \leq p < \infty$ soient séparables. Il faut donc chercher la limite de :

$$\text{(IX)} \qquad \sum_{t=1}^{T} \int_{\mathbb{R}^d} K_{t,T}(x)\, g_{X_t}^{F_{t-1}}(x)\, d\mu(x) = \sum_{t=1}^{T} [K_{t,T}]\, (g_{X_t}^{F_{t-1}})$$

les $[K_{t,T}]$ étant les opérateurs définis de $L^1(\mu)$ dans H par l'intégrale de Bochner utilisée ci-dessus, sous réserve d'existence.

Sans entrer dans les problèmes techniques concernant cette existence, on peut ici fournir l'intuition de la démonstration effectuée dans le théorème 1. En imposant des conditions supplémentaires (<u>H4 à H7</u>) aux applications $(K_{t,T})$, les sommes d'opérateurs correspondants $\{ \sum_{t=1}^{T} [K_{t,T}]\}$ convergent vers un opérateur C. Si de même on suppose (<u>conditions B1 à B4</u> imposées au processus) que la suite des $g_{X_t}^{F_{t-1}}$ converge vers une limite g , il est naturel d'obtenir la convergence de $\sum_{1}^{T} [K_{t,T}]\, (g_{X_t}^{F_{t-1}})$ vers Cg .
C'est le travail effectué dans le théorème 1, sous la formalisation au demeurant plus lourde (cf. les "<u>conditions B et I</u>" imposées aux processus et aux opérateurs), rendue nécessaire pour tenir compte des cas où la convergence des $g_{X_t}^{F_{t-1}}$ vers g s'effectue dans un espace fonctionnel B différent de $L^1(\mu)$.

Finalement, en supposant les $(K_{t,T})$ bien choisies (nous fournirons par la suite de nombreuses familles vérifiant les conditions H1 à H7 évoquées ci-dessus , dans les diverses applications) la convergence de

$\sum_{1}^{T} K_{t,T}(X_t)$ vers Cg est obtenue si le processus (X_t) vérifie au minimum une condition du type (conditions B2 et B4 du texte) :

$$\text{(X)} \qquad \lim_{T\to\infty} \left|\left|\frac{1}{T} \sum_{t=1}^{T} g_{X_t}^{F_{t-1}} - g\right|\right| = 0$$

la norme étant prise dans un espace fonctionnel $(L^1(\mu), L^2(\mu)$, $C^o_{(\mathbb{R}^d)} \ldots)$ auquel appartiennent les $g_{X_t}^{F_{t-1}}$ et g , la convergence s'effectuant en moyenne ou p.s. On montre alors, second point développé dans cette partie, que les processus ergodiques vérifient ce type de conditions.

Un cas particulier illustre bien le fait : si (X_t) est strictement stationnaire et markovien, h la densité des couples (X_t , X_{t+1}) par rapport à $\mu \otimes \mu$, g celle des variables X_t , obtenir (X), c'est montrer la convergence de la suite :

$$\text{(XI)} \qquad \frac{1}{T} \sum_{t=1}^{T} L(X_t) = \frac{1}{T} \sum_{t=1}^{T} \frac{h(X_t,.)}{g(X_t)}$$

(L est l'application qui à un élément (x) de $\mathbb{R}^d$ associe la fonction $(y) \longrightarrow \frac{h((x),(y))}{g((x))}$) . Lorsque $L(X_t)$ est Bochner-intégrable, il résulte du classique théorème ergodique à valeurs dans les Banach (Parthasarathy (30), Krengel (25)) que la convergence a alors lieu vers g ... Finalement, le <u>lemme 4</u> et le <u>corollaire 1</u> développent cette idée en montrant que, pour les processus stationnaires au moins, l'ergodicité implique B2) et B4) sous des conditions techniques additionnelles, automatiquement vérifiées si on veut obtenir la convergence (X) dans $L^1(\mu)$ par exemple.

En fait les exemples de processus vérifiant (X) pourront être enrichis par la suite. La stricte stationnarité n'apparaît pas nécessaire à l'obtention de B2) à B4), dans le cas d'un processus Markovien tel que la densité initiale ne soit pas la densité invariante, par exemple, de toute façon si les X_t sont indépendantes. C'est pourquoi nous avons laissé les hypothèses Bi) sous leur forme générale un peu abstraite, en les scindant en une condition sur les $(g_{X_t}^{F_{t-1}} - g_{X_t})$ (aspect "ergodique") et une condition sur les g_{X_t} (aspect "stationnaire"). Le corollaire 4 montre déjà que les résultats d'estimation démontrés, sous les conditions Bi), concernent au minimum la classe des processus stationnaires et ergodiques, c'est-à-dire une classe de processus beaucoup plus large que les processus stationnaires et mélangeants utilisés jusque là. Nous fournissons quelques exemples usuels (processus linéaires p. ex.) pour lesquels les résultats s'appliquent.

Nous terminons enfin la première partie par deux applications directes du théorème 1. On suppose que le processus observé vérifie (X), la norme étant prise dans $L^1(\mu)$. On démontre alors (théorèmes 2 et 3) :

i) La convergence de la fonction de répartition empirique :

$$\hat{F}_T((x)) = \frac{1}{T} \sum_{t=1}^{T} I_{\{\prod_1^d]-\infty,\, x_i]\}}(X_t) \qquad (*)$$

vers la fonction de répartition F de la loi de densité g, au sens où, pour tout compact C de $\mathbb{R}^d$ on a :

$$\lim_{T\to\infty} \int_C |\hat{F}_T - F|^2 \, d\mu = 0 \quad .$$

ii) La convergence de la loi empirique $\frac{1}{T} \sum_{t=1}^{T} \delta_{X_t}$ vers la loi P de densité g , au sens de la convergence faible p.s. : il faut pour cela assimiler les lois de probabilités à des éléments d'un espace de Hilbert à noyau reproduisant (cf. Guilbart (31)).

(*) $(x) = (x_1, \ldots, x_d)$.

On notera que la convergence de $\hat{F}_T$ pourrait se déduire de ii) ; i) et ii) montrent la facilité d'usage du théorème 1 dans des conditions originales, les démonstrations se trouvant ici notablement allégées par le fait que les opérateurs $[K_{t,T}]$ sont de la forme $\frac{1}{T}$ $[K]$.

DEUXIEME PARTIE

Dans la seconde partie du travail, supposant encore que les $g_{X_t}^{\mathcal{F}_{t-1}}$ appartiennent à $L^1(\mu)$ ou $L^2(\mu)$ et y convergent vers g au sens de Cesaro, on estime g par les méthodes classiques.

α) En un premier point, les $g_{X_t}^{\mathcal{F}_{t-1}}$ et g étant dans $L^2(\mu)$, on utilise la méthode usuelle de projection (ou encore des "fonctions orthogonales") (cf. Bosq (32) pour une introduction). L'estimateur retenu $\hat{f}_T^3$ s'écrit :

$$\text{(XIII)} \qquad \hat{f}_T^3 = \sum_{i=1}^{q(T)} \hat{a}_i \, e_i \quad , \quad \hat{a}_i = \frac{1}{T} \sum_{t=1}^{q(T)} e_i(X_t)$$

les e_i constituent une base orthonormale de $L^2(\mu)$, $q(T)$ une suite de réels croissant vers l'infini. Il a été largement étudié dans le cadre de l'échantillon (cf. Bosq et Lecoutre (7) pour une synthèse), et dans celui des processus stationnaires et mélangeants (Delecroix (13), Doukhan (9)).

On démontre alors sa convergence vers g en utilisant les résultats de la première partie : l'estimateur s'écrit $\sum_{t=1}^{T} K_{t,T}(X_t)$, les opérateurs $[K_{t,T}]$ associés n'étant autre que $(1/T \, . \, P_T)$ où P_T est la projection orthogonale sur l'espace engendré par $(e_1 , \ldots, e_{q(T)})$: c'est le théorème 4 qui généralise tous les résultats obtenus précédemment, au niveau des conditions imposées aux (X_t) .

Sous les mêmes hypothèses générales, on étudie ensuite un estimateur dérivé du précédent, et récursif dans le sens où il vérifie la relation :

(XIV) $$\hat{f}^4_{T+1} = (1 - \frac{1}{T+1})\ \hat{f}^4_T + \frac{1}{T+1} \sum_{j=1}^{q(T+1)} e_j(X_{T+1})\ e_j$$

Défini pour la première fois dans Delecroix (13), il amène un gain de temps dans les calculs et converge (cf. théorème 4) vers g, moyennant des conditions un peu plus restrictives que celles qu'exige la convergence de $\hat{f}^3_T$.

Le point suivant concerne l'estimation par projection des dérivées partielles $D^i(g)$ de g, $(D^i(g)(x) = \frac{\partial}{\partial x_i} g((x_1,\dots,x_d)))$, supposées exister et appartenir à $L^2(\mu)$: ce problème intervient naturellement dans l'identification non-paramétrique de la dérive d'une diffusion (Banon et Nguyen (33), Bhattacharaya (34), Brodeau (35)). Nous utilisons l'estimateur employé par exemple dans Greblicky - Pawlak (36) (et non celui de Walter (37) !), en supposant l'existence des $(D^i e_j)$, $1 \leq i \leq d$, $j \geq 1$ et leur appartenance à L^2, hypothèse vérifiée pour les bases classiques de $L^2(\mathbb{R}^d)$. Nous démontrons sa convergence au <u>théorème 6</u> pour peu que les $(D^i g_{X_t}^{\mathcal{F}_{t-1}})$ existent et convergent dans $L^2(\mu)$ vers les $D^i(g)$, tandis que les $g_{X_t}^{\mathcal{F}_{t-1}}$ y convergent vers g au sens de Cesaro, c'est-à-dire que l'espace B (cf. la "condition B") sera ici l'espace de Sobolev W^{12}.

β) On aborde ensuite l'estimation de g par la méthode, surabondamment étudiée dans le cas de l'échantillon (cf. Bosq et Lecoutre (7) pour une synthèse), dite du "noyau de Parzen-Rosenblatt". On utilise ainsi l'estimateur qui s'écrit :

(XV) $$\frac{1}{T} \{ \sum_{t=1}^{T} [h(T)]^{-d} . K [(h(T))^{-1} . (\cdot - (X_t))] \} .$$

(K densité de probabilité, $h(T)$ élément de $\mathbb{R}$), et ensuite sa version récursive $\hat{f}^7_T$ définie par :

(XVI) $$\hat{f}_T^7 = (1 - \gamma_T)\, \hat{f}_{T-1}^7 + \gamma_T \{[h(T)^{-d}]\, K[(h(T))^{-1} (\cdot - X_T)]\} \quad .$$

Le théorème 7 fournit les conditions exactes sous lesquelles ces deux estimateurs convergent vers g dans $L^2(\mu)$, en supposant que $(1/T) \sum_{t=1}^{T} g_{X_t}^{F_{t-1}}$ fasse de même. Sa démonstration utilise les résultats de la 1ère partie. Le lemme 8 permet, quant à lui, de déterminer concrètement les densités de probabilité K qui ont les propriétés requises dans le théorème. Il faut évidemment que la suite de réels $h(t)$ tende vers 0 , "pas trop vite", et que les γ_t vérifient des conditions de régularité exhibées pour les 2 cas particuliers les plus usuels.

Nous démontrons ensuite un théorème (théorème 8) de convergence des deux estimateurs de g , dans $L^1(\mu)$, quand $(1/T) (\sum_{t=1}^{T} g_{X_t}^{F_{t-1}})$ fait de même. (Rappelons que les processus ergodiques vérifient cette hypothèse sans aucune condition technique additionnelle sur les $g_{X_t}^{F_{t-1}}$, d'où son intérêt). La démonstration apportée n'utilise plus guère de la 1ère partie que la décomposition (VI), l'orthogonalité (tant recherchée) ayant disparu : on prouve la convergence à 0 de la 1ère partie de (VI) par une technique à notre connaissance originale, quitte à imposer au processus (X_t) des conditions supplémentaires assez générales. Le lemme 9 permet de dégager concrètement les processus les vérifiant. Rappelons pour mémoire la difficulté technique usuelle d'obtention de ces "théorèmes L^1" : dans le cadre de l'échantillon, on verra Devroye et Györfi (38).

On termine en démontrant que les dérivées partielles de l'estimateur classique de la méthode du noyau convergent vers celles de g (supposées exister) sous des conditions additionnelles mises en évidence aux théorèmes 10 et 11, (correspondant aux convergences respectives dans $L^2(\mu)$ et $L^1(\mu)$).

Cette estimation simultanée de g et des $D^i g$ permettra l'obtention de convergences ponctuelles au chapitre suivant, outre l'intérêt déjà évoqué pour l'identification non-paramétrique d'une diffusion. On notera de plus que la méthode se généraliserait aisément aux dérivées d'ordre supérieur, ce qui permet, lorsque les X_t sont réels, une estimation de la fenêtre $h(T)$ optimale (cf. Deheuvels (39)).

TROISIEME PARTIE

Dans la troisième partie de l'ouvrage, on aborde le problème de la convergence ponctuelle (uniforme ou non) d'estimateurs de la densité, fondamentale dans l'étude du prédicteur utilisé à la fin du travail, entre autres. Du point de vue technique, le travail se sépare nettement des résultats obtenus jusqu'ici. En considérant l'estimateur comme élément de L^∞ par exemple, les variables $\{K_{t,n}(X_t) - E^{\mathcal{F}_{t-1}}(K_{t,n}(X_t))\}$ introduites en (VI), n'offrent plus aucun caractère d'orthogonalité, et les résultats du chapitre I ne s'appliquent guère, au moins en leur globalité.

Nous proposons deux approches différentes du problème :

α) En un premier point, on étudie la convergence souhaitée sous une hypothèse vérifiée, comme jusqu'alors, par les processus ergodiques sous des conditions additionnelles de régularité (<u>lemme 4</u>). On suppose ainsi les $g_{X_t}^{\mathcal{F}_{t-1}}$ dans $C^0(\mathbb{R}^d)$ (espace séparable des fonctions continues sur $\mathbb{R}^d$ qui tendent vers 0 à l'infini) et y convergent vers une limite g au sens de Cesaro. On a donc :

$$\lim_{T\to\infty} \left(\sup_{x\in\mathbb{R}^d} \left| \frac{1}{T} \sum_{1}^{T} g_{X_t}^{\mathcal{F}_{t-1}}(x) - g(x) \right| \right) = 0 \quad \text{(en moyenne et p.s.).}$$

On considère l'estimateur classique de la méthode du noyau $\hat{f}_T^6$. La décomposition (VI) utilisée au départ permet de voir que sous l'hypothèse ergodique définie ci-dessus, pour les noyaux classiques,

$\sup_{x\in\mathbb{R}^d} |\hat{f}_T^6(x) - g(x)|$ tend vers 0 si $h(T)$ décroît vers 0, à condition qu'on ait aussi la convergence à 0 de :

$$\sup_{x\in\mathbb{R}^d} \left| \sum_{t=1}^{T} \left\{ \frac{1}{T.(h(T))^d} K\left(\frac{x-X_t}{h(T)}\right) - E^{F_{t-1}} \left[\frac{1}{T(h(T))^d} . K\left(\frac{x-X_t}{h(T)}\right) \right] \right\} \right|$$

C'est le point délicat. Le théorème 11 résout le problème. Pour suggérer simplement sa démonstration, très technique, on peut dire qu'elle utilise fondamentalement deux idées classiques :

i) Choisir un noyau lipschitzien (cf. Bosq et Bleuez (40) pour l'utilisation originelle de la méthode) qui ramène l'étude du supremum sur un compact à celle du supremum pris sur un nombre fini de points (x_i).

ii) En un point fixé x utiliser l'orthogonalité (cette fois rétablie) des variables étudiées, et obtenir une convergence presque complète (comme dans la démonstration usuelle de la loi des grands nombres), grâce à la décomposition de la somme des T variables en 2 parties, de 1 à $\Phi(T)$, et de $\Phi(T)$ à T, où $\Phi(T)$ est le plus grand carré entier inférieur ou égal à T. On obtient la convergence souhaitée, sous des conditions restrictives sur la suite $h(t)$, précisées par le corollaire 3 (qui fournit des exemples de suites utilisables).

β) Le théorème précédent (comme ceux de la seconde partie) ne permet pas d'obtenir les vitesses de convergence des estimateurs étudiés. D'abord, ces théorèmes basés sur une approche fonctionnelle (cf. lemme 2 chapitre I) ne donnent pas de relations explicites entre la norme de la différence des $(\sum_1^T E^{F_{t-1}}(K_{t,n}(X_t)))$ avec la limite cherchée, celle de $\{\frac{1}{T}\sum_1^T g_{X_t}^{F_{t-1}} - g\}$, et les normes des $[K_{t,T}]$ utilisées. Ensuite, l'évaluation même de $\left\|\frac{1}{T}\sum_1^T g_{X_t}^{F_{t-1}} - g\right\|$ reste problématique (cf. Krengel (25) "vitesse de convergence dans le théorème ergodique").

Aussi proposons-nous en une seconde partie une évaluation de l'écart ponctuel entre $\hat{f}_T^6$ et g dans le cas où le processus observé est stationnaire et mélangeant : nous reprenons un article (Delecroix (13)) qui permet cette évaluation sous l'hypothèse de forte mélangeance (cf. IV), la plus faible possible, pour tous les estimateurs de la densité étudiés jusqu'ici. On y trouvera de même les théorèmes de convergence ponctuelle d'estimateurs de la densité de transition. Cet article est au demeurant représentatif des techniques utilisées dans les cas d'observations mélangeantes, évoquées du début de cette introduction, et très différentes des approches jusqu'ici employées. Notons enfin que l'article concerne des estimateurs construits à partir d'un processus à temps continu, le passage au temps discret s'avérant immédiat.

QUATRIEME PARTIE

La quatrième partie du travail est consacrée à la prédiction non-paramétrique dans les séries temporelles stationnaires. Pour prévoir $\psi(X_{T+k})$, on construit un estimateur $\hat{R}_T$, non-paramétrique, de la régression R de $\psi(X_{t+k})$ sur $(X_t , X_{t-1} ,\dots, X_{t-r+1})$, pour un entier r au choix du statisticien, à partir des valeurs observées $X_1,\dots,X_T$. Le prédicteur proposé est alors $\hat{R}_T(X_T,\dots,X_{T-r+1})$: c'est la méthode utilisée par exemple par Collomb 1984 (1), Bosq 1987 (5), ou Härdle et Vieu 1987 (3). Nous la développons sous trois aspects :

α) On démontre la convergence uniforme presque sûre sur tout compact de $\hat{R}_T$ vers R, sous des hypothèses ergodiques du type de celles qu'on a utilisées jusqu'ici, $\hat{R}_T$ étant construit par la méthode du noyau. C'est l'objet du <u>théorème 12</u> qui s'appuie en particulier sur le résultat de convergence uniforme obtenu dans la partie III. On en déduit que la différence entre le prédicteur et $R(X_T ,\dots, X_{T-r+1})$ tend vers 0 , si $T \to \infty$, sous des conditions de régularité suffisantes. Ceci justifie l'utilisation du prédicteur choisi.

β) On étudie la convergence L^2 du prédicteur sous des hypothèses de mélangeance. C'est l'objet de l'article (12) publié en 1985 en collaboration avec D. Bosq, que nous reproduisons. Le prédicteur utilisé appartient ici à une classe contenant notamment ceux qui dérivent de l'emploi de la méthode des fonctions orthogonales.

γ) On procède à une étude concrète du prédicteur étudié de façon théorique en α). On dégage notamment un principe de construction (établir des moyennes pondérées aléatoires des X_{t+h}) qui semble généralisable par la suite, et une méthode non-paramétrique de détermination d'intervalles de prévision. Celle-ci permet une comparaison objective avec les performances d'autres prédicteurs classiques (méthode de Box-Jenkins) tels qu'ils sont définis en Monfort-Gouriéroux (41). Les comparaisons ont été effectuées sur de nombreuses séries, observées ou simulées. Les résultats obtenus démontrent une légère supériorité de la méthode non-paramétrique. C'est l'article (42) publié en collaboration avec M. Carbon, que nous reprenons ici.

BIBLIOGRAPHIE DE L'INTRODUCTION.

[16] AKONOM J. (1987) - Thèse d'Etat.
Université des Sciences et Techniques de Lille Flandres-Artois.

[20] ASH and GARDNER (1975) - Topics in stochastic processes.
Academic Press.

[33] BANON G. et HUNG NGUYEN T. (1981) - Recursive estimation in diffusion model.
S.I.A.M. J-Control and Optimization (19) n° 5, 676-685.

[34] BATTACHARYA P.K. - Estimation of a probability density function and its derivatives.
Sankhya Series, Ser. A, 29, 373-382.

[26] BECK A. - On the strong law of larges numbers
Ergodic Theory. Academic Press ed. F.B. Wright, New-York 21-53.

[27] BECK A. et WARREN P. (1968) - A strong law of large numbers for weakly orthogonal sequences of Banach-space valued random variables.
MRC Technical Summary - Report n° 848, Univ. of Wisconsin.

[40] BLEUEZ J. et BOSQ D. (1978) - Etude d'une classe d'estimateurs non-paramétriques de la densité.
Annales Inst. H. Poincaré 14, 4, 479-498.

[5] BOSQ D. (1987) - La statistique non-paramétrique des processus.
Rendiconti del seminario Matematico de Torino.

[7] BOSQ D. et LECOUTRE J.P. (1987) - Théorie de l'estimation fonctionnelle.
Economica Paris.

[8] BOSQ D. (1977) - Sur l'estimation sans biais d'un paramètre fonctionnel.
Pub. Interne de l'UER de Math., Lille, n° 104, USTL.

[12] BOSQ D. et DELECROIX M. (1985) - Non-parametric prediction of a
Hilbert space valued Random variable.
Stochastic Processes and their applications 19, 271-280.
North-Holland.

[21] BOSQ D. (1983) - Sur la prédiction non-paramétrique de variables
aléatoires et de mesures aléatoires.
Z. Wachen. Verw. Geb. 64, 541-553.

[22] BOSQ D. - Sur l'estimation de la densité d'un processus
stationnaire et mélangeant.
C.R.A.S. Série A, tome 277, 535-538.

[32] BOSQ D. (1980) - Estimation de la densité par projection sur
un sous-espace de dimension finie.
Portugaliae Math. 37, 1,2, 93-111.

[35] BRODEAU (1984) - Techniques d'estimations de paramètres pour des
équations différentielles stochastiques linéaires commandées.
Rapport de recherche n° 437, I.M.A.G.

[23] CARBON M. (1986) - Une inégalité de type Bernstein pour les processus
mélangeants à temps continu. Application à l'estimation
de la densité spectrale.
Pub. ISUP, fasc. 2,3 p. 15-33.

[41] CARBON M. et DELECROIX M. (1987) - Comparaison de méthodes paramétriques
et non-paramétriques de prévision.
Pub. IRMA, Lille, Vol. 9, n° 11.

[1] COLLOMB (1984) - Propriétés de convergence presque-complète du
prédicteur à noyau.
Z. Wahrs. verw. Geb. 66, 441-460.

[6] COLLOMB (1978) - Estimation non-paramétrique de la régression :
régressogramme et méthode du noyau.
Pub. Lab. Stat. Prob. Univ. Toulouse 0778159

[38] DEVROYE L. et GYÖRFI L. (1984) - Non-parametric density estimation : the L_1 view.
Wiley, New-York.

[39] DEHEUVELS P. (1977) - Estimation non-paramétrique de la densité par histogrammes généralisés.
Revue de Stat. Appl. 35, 5-42.

[13] DELECROIX M. (1980) - Sur l'estimation des densités d'un processus stationnaire à temps continu.
Pub. ISUP, XXV, 1,2, 17-39.

[4] DOUKHAN et PORTAL (1983) - Moments de variables aléatoires mélangeantes.
C.R.A.S., t. 297.

[9] DOUKHAN (1986) - Fonctions d'Hermite et statistique des processus mélangeants.
Collection "Approches non-paramétriques en analyse chronologique" 1985
Institut des Hautes Etudes de Belgique, p. 99-116.

[11] GASTWIRTH et RUBIN (1975) - The asymptotic distribution theory of the empiric C.D.F. for mixing stochastic processes.
Annals of Stat. 3, 4, 809-824.

[41] MONFORT-GOURIÉROUX (1983) - Cours de séries temporelles.
Economica.

[36] GREBLICKY and PAWLAK (1984) - Hermite series estimates of a probability density and its derivatives.
J. Multivariate analysis 15, 174-182.

[31] GUILBART C. (1978) - Etude des produits scalaires sur l'espace des mesures. Estimation par projections. Tests à noyaux.
Thèse Doctorat ès Sciences. U.S.T.L.

[28] GYORFI L. GYORFI Z. and VAJDA Z. (1977) - A strong law of large numbers and some applications.
Studia Scientiarum Mathematicarum Hungarica 12, 233-244.

[3] HÄRDLE et VIEU (1987) - Non-parametric prediction by the kernel method (à paraître).

[25] KRENGEL U. (1985) - Ergodic theorems
De Gruyter Studies in Mathematics 6. Walter De Gruyter Berlin - New-York.

[24] E. MOURIER (1953) - Elements aléatoires à valeurs dans un espace de Banach.
Annales de l'IHP B. 13, 161-244.

[30] PARTHASARATHY (1967) - Probability measures on metric spaces.
Academic Press. New-York.

[10] PELIGRAD M. (1987) - Properties of uniform consistensy of the kernel estimates of density and of regression functions under dependance assumptions.
A paraître.

[17] ROSENBLATT M. (1971) - Markov Processes.
Structure and asymptotic behavior.
Springer-Verlag.

[18] ROSENBLATT M. (1972) - Uniform ergodicity and strong mixing.
Z. Wahrs. verw. Gebiete, 24, 79-84

[19] ROSENBLATT M. (1985) - Stationary sequences and Random fields.
Birkhäuser Boston, Inc.

[2] SARDA et VIEU (1986) - Vitesse de convergence uniforme de l'estimateur à noyau de la régression pour des observations dépendantes.
C.R.A.S. t. 102, Série 1, n° 11.

[15] TUAN D. PHAM et LANH T. TRAN (1985) - Some mixing properties of time series models.
Stochastic processes and their applications.
297-303 North-Holland.

[29] VAN RYZIN (1969) - On strong consistensy of density estimate.
AMS, Vol. 40 n° 5, 1765-1772.

[37] WALTER G. (1977) - Properties of Hermite series estimation of a probability density.
Annals. Statis. 5, 6, 1258-1264.

[14] WITHERS C.S. (1981) - Conditions for linear processes to be strong mixing.
Z. Wahrs. verw. Gebiete, 57, 477-480.

CONVERGENCE D'ESTIMATEURS DE LA FORME $\sum_{t=1}^{T} K_{t,T}(X_t)$

(Chapitres I et II)

Comme annoncé dans l'introduction, on va ici étudier le devenir asymptotique d'une suite d'éléments aléatoires d'un espace de Hilbert de la forme :

$$(1) \qquad \sum_{t=1}^{n} K_{t,n}(X_t) \quad ,$$

le processus (X_t) , $t \in \mathbb{Z}$, observé dans $\mathbb{R}^d$, étant astreint à des conditions d'ergodicité portent sur les densités de transition.

Nous traiterons successivement trois points dans cette $1^{ère}$ partie de notre travail.

Dans la première partie du chapitre I, après avoir introduit précisément notations, définitions et les hypothèses Hi) et Bi) menant à l'obtention des convergences souhaitées (point A), nous démontrerons les théorèmes de convergence adéquats (point B).

Dans la seconde partie de ce chapitre, nous étudierons les types de processus vérifiant les hypothèses introduites dans les théorèmes de convergence, montrant que, grossièrement parlant, il s'agit des processus ergodiques, admettant des densités de transition suffisamment régulières.

Enfin, au chapitre II, nous retrouverons successivement, (points A et B), comme applications directes des théorèmes généraux démontrés auparavant, les convergences classiques de la fonction de répartition empirique et de la loi empirique.

C.I. - 1ère PARTIE

LES THEOREMES DE CONVERGENCE

(A) - Les conditions suffisantes de convergence : hypothèses Hi et Bi.

a) Les applications $K_{t,n}$

Dans ce qui suit les $K_{t,n}$, $n \geq 1$, $1 \leq t \leq n$ représentent une suite de séquences d'applications mesurables de $\mathbb{R}^d$ dans un espace de Hilbert H séparable, dont nous noterons $<,>_H$ et $|| \ ||_H$ produit scalaire et norme. Pour tout couple (t,n) , nous poserons :

(2) $$\alpha_{t,n} = \sup_{y \in \mathbb{R}^d} ||K_{t,n}(y)||_H \quad ,$$

(3) $$\gamma_{t,n} = \sup_{y \in \mathbb{R}^d} ||K_{t,n}(y) - K_{t,n+1}(y)||_H \quad .$$

Cela permet la formulation des trois hypothèses suivantes :

H1) $$\lim_{n \to \infty} \left(\sum_{t=1}^{n} \alpha_{t,n}^2 \right) = 0 \quad ,$$

qui assurera qu'en (2) le supremum soit fini (pour n assez grand)

H2) $$\sum_{n \geq 1} \alpha_{n,n}^2 < \infty$$

H3) $$\sum_{n=1}^{\infty} \left\{ \left(\sum_{t=1}^{n} \gamma_{t,n}^2 \right)^{1/2} \cdot \left[\left(\sum_{t=1}^{n} \alpha_{t,n}^2 \right)^{1/2} + \left(\sum_{t=1}^{n} \alpha_{t,n+1}^2 \right)^{1/2} \right] \right\} < \infty$$

hypothèse assurant, elle, la finitude des $\gamma_{t,n}$.

Les conditions précédentes ne sont pas, en toute généralité, réductibles les unes aux autres, mais si, à t fixé, $\alpha_{t,n}$ est décroissante en n , H3) se réduit à

H'3) $$\sum_{n=1}^{\infty} \{(\sum_{t=1}^{n} \alpha_{t,n}^2)\ (\sum_{t=1}^{n} \gamma_{t,n}^2)\}^{1/2} < \infty$$

b) Les densités $g_{X_t}^{F_{t-1}}$

α) On supposera que le processus des observations dans $\mathbb{R}^d$ est constitué de variables (X_t) , $t \in \mathbb{Z}$, dont les lois sont absolument continues par rapport à μ , mesure σ-finie sur $\mathbb{R}^d$, de densités g_{X_t}. On suppose de plus que, pour tout t la loi conditionnelle de X_t par rapport à F_{t-1} , tribu engendrée par X_{t-1} , X_{t-2} ,..., existe, et se caractérise par une densité conditionnelle, notée $g_{X_t}^{F_{t-1}}$, par rapport à μ.

Cette hypothèse posée, nous supposerons éventuellement que les $g_{X_t}^{F_{t-1}}$ appartiennent à un autre espace fonctionnel que $L^1(\mu)$, et dirons alors que

" Le processus (X_t) , $t \in \mathbb{Z}$, vérifie la condition (B) "

pour signifier que les g_{X_t} et $g_{X_t}^{F_{t-1}}$ existent, et sont éléments de $B \cap L^1(\mu)$. (Au minimum, le processus observé vérifie donc toujours "la condition $L^1(\mu)$" !). B sera muni d'une structure d'espace de Banach, dont nous notons $\| \ \|_B$ la norme.

β) Nous supposerons de plus que les densités $g_{X_t}^{F_{t-1}}$ et g_{X_t} convergent fortement dans B , introduisant ainsi 4 hypothèses éventuelles :

B1 $$\lim_{t\to\infty} \text{p.s. } \|g_{X_t}^{F_{t-1}} - g_{X_t}\|_B = 0 \quad \text{et} \quad \lim_{t\to\infty} \|g_{X_t} - g\|_B = 0$$

B2 $$\lim_{n\to\infty} \text{p.s. } \|(n^{-1}) \cdot [\sum_{t=1}^{n} (g_{X_t}^{F_{t-1}} - g_{X_t})]\|_B = 0$$

et $$\lim_{n\to\infty} \|(n^{-1}) \cdot (\sum_{t=1}^{n} g_{X_t}) - g\|_B = 0$$

B3 $$\lim_{t\to\infty} E\{||g_{X_t}^{\mathcal{F}_{t-1}} - g_{X_t}||_B\} = 0 \quad \text{et} \quad \lim_{t\to\infty} (||g_{X_t} - g||_B) = 0$$

B4 $$\lim_{n\to\infty} E\{||(n^{-1}) \cdot [\sum_{t=1}^{n} (g_{X_t}^{\mathcal{F}_{t-1}} - g_{X_t})]||_B\} = 0$$

et $$\lim_{n\to\infty} ||(n^{-1}) \cdot (\sum_{t=1}^{n} g_{X_t}) - g||_B = 0 \quad ,$$

g désignant un élément de B .

Les exemples de processus vérifiant ces conditions Bi) seront fournis dans le 3^e paragraphe ...

c) *Les opérateurs* $[K_{t,n}]$

α) L'existence des $g_{X_t}^{\mathcal{F}_{t-1}}$ permet d'écrire pour tous t et n :

(4) $$E^{\mathcal{F}_{t-1}} [K_{t,n}(X_t)] = \int_{\mathbb{R}^d} K_{t,n}(x) \cdot g_{X_t}^{\mathcal{F}_{t-1}}(x) \, d\mu(x) \quad ,$$

le second membre représentant l'intégrale de Bochner de la fonction $(g_{X_t}^{\mathcal{F}_{t-1}}(.)) \cdot K_{t,n}(.)$, à valeurs dans H .

L'espérance conditionnelle introduite existe en effet (cf. Neveu (1)), dès que $K_{t,n}(X_t)$ est Bochner-intégrable c'est-à-dire que

$$\int_{\mathbb{R}^d} ||K_{t,n}(x)||_H \cdot g_{X_t}(x) \, d\mu(x) < \infty$$

C'est donc vrai si $||K_{t,n}(.)||_H$ et g_{X_t} sont dans des espaces L^p et L^q conjugués, sur $\mathbb{R}^d$, donc toujours sous H1. Dans ces conditions (4) résulte de l'unicité de l'intégrale de Bochner, de celle de l'espérance conditionnelle, et du fait que l'on puisse les permuter avec toute forme linéaire continue sur H.

β) L'intégrale introduite en (4) n'étant pas forcément définie pour tout élément de B , lorsque le processus "vérifie la condition B" , alors qu'elle l'est sur $B \cap L^1(\mu)$, nous supposerons de surcroît l'existence d'une suite de séquences d'opérateurs $[K_{t,n}]$, de B dans H , "vérifiant la condition I" , c'est-à-dire tels que

$$(5) \qquad \forall g , \quad g \in B \cap L^1(\mu) , \quad [K_{t,n}](g) = \int_{\mathbb{R}^d} g(x) \, . \, K_{t,n}(x) \, d\mu(x) \, .$$

Nous imposerons enfin quatre hypothèses aux $[K_{t,n}]$:

H4) La suite des opérateurs $\{\sum_{t=1}^{n} [K_{t,n}]\}$ converge fortement vers un opérateur C de B dans H , i.e. :

$$\forall g, \; g \in B , \quad \lim_{n\to\infty} ||(\sum_{t=1}^{n} [K_{t,n}])(g) - C(g)||_H = 0 \, .$$

H5) t étant fixé $\lim_{n\to\infty} ||K_{t,n}|| = 0$

H6) $\sup_{n\geq 1} \{\sum_{t=1}^{n} ||K_{t,n}||\} < \infty$

H7) $\sup_{n\geq 1} \{\sum_{t=1}^{n-1} t \, . \, ||[K_{t,n}] - [K_{t,n+1}]|| + n \, ||K_{n,n}||\} < \infty$

(Pour H5), H6), H7) nous notons $|| \; ||$, sans indice, la norme usuelle d'opérateur de $K_{t,n}$:

$$||K_{t,n}|| = \sup_{g/||g||_B \leq 1} (||[K_{t,n}](g)||_H)$$

Sous ces hypothèses, nous pourrons démontrer les théorèmes de convergence de $\sum_{t=1}^{n} K_{t,n}(X_t)$ vers Cg , C étant l'opérateur introduit par H4), g la limite introduite dans les conditions Bi ...

(B) - Enoncés et démonstrations .

a) Principe de la démonstration

Pour obtenir la convergence souhaitée, on majorera $|| \sum_{t=1}^{n} K_{t,n}(X_t)-Cg||_H$ par la somme des deux quantités :

(6) $$A_1(n) = || \sum_{t=1}^{n} Y_{t,n} ||_H \quad , \quad \text{où} \quad Y_{t,n} = K_{t,n}(X_t) - E^{F_{t-1}} [K_{t,n}(X_t)]$$

(7) $$A_2(n) = || \sum_{t=1}^{n} \{E^{F_{t-1}} [K_{t,n}(X_t)]\} - Cg||_H$$

décomposition possible sous H1, nous l'avons vu. La convergence à 0 de $A_1(n)$ résultera de l'orthogonalité des $Y_{t,n}$, $1 \leq t \leq n$, celle de $A_2(n)$ de l'égalité

(8) $$A_2(n) = ||[\sum_{t=1}^{n} K_{t,n}] (g_{X_t}^{F_{t-1}}) - Cg||_H$$

justifiée ci-dessus, et des hypothèses (B_i) et H4) - H7) ...

b) Comportement asymptotique de $A_1(n)$

Proposition 1.- Si la suite de séquences d'applications $K_{t,n}$, $n \geq 1$, $1 \leq t \leq n$, vérifie H1) , on a :

$$\lim_{n\to\infty} E\{(A_1(n))^2\} = 0$$

Si elle vérifie de surcroît H2) et H3) $A_1^2(n)$ converge vers 0 presque sûrement.

Démonstration :

α) D'après (6) : $(A_1(n))^2 = \langle \sum_{t=1}^{n} Y_{t,n} \,,\, \sum_{t=1}^{n} Y_{t,n} \rangle_H$, mais comme les $Y_{t,n}$ sont orthogonales, par propriété standard de l'espérance conditionnelle, il vient

$$(9) \qquad E\left[A_1^2(n)\right] = \sum_{t=1}^{n} E(||Y_{t,n}||_H^2)$$

Maintenant, par définition des $Y_{t,n}$, on aura :

$$E(||Y_{t,n}||_H^2) = E(||K_{t,n}(X_t)||_H^2) - E(||E^{F_{t-1}}\left[K_{t,n}(X_t)\right]||_H^2)$$

$$\leqslant E(||K_{t,n}(X_t)||_H^2) \leqslant \sup_{y\in\mathbb{R}^d} ||K_{t,n}(y)||_H^2 \ ,$$

ce qui revient à écrire d'après (1) et (9)

$$(10) \qquad E\left[A_1^2(n)\right] \leqslant \sum_{t=1}^{n} \alpha_{t,n}^2$$

et H1) implique bien la conclusion souhaitée.

β) Pour démontrer la convergence p.s. à 0 de $A_1^2(n)$, nous allons utiliser un lemme de Van Ryzin (2).

Lemme 1.- Soit $(F_n)_{n\geq 1}$, une suite croissante de sous-tribus de F , (Z_n) et (Z'_n) , $n \geq 1$, deux suites de v.a.r. définies sur (Ω,F,P), adaptées aux sous-tribus F_n . Si on suppose alors Z_1 intégrable, les Z_n p.s. positives, que pour tout n on ait p.s. $E^{F_n}(Z_{n+1}) \leqslant Z_n + Z'_n$, et qu'enfin la série $\sum_{n=1}^{\infty} E(Z'_n)$ converge, alors la suite Z_n converge presque sûrement vers une limite finie.

Nous utiliserons ce lemme en prenant $Z_n = [A_1(n)]^2$, et $Z'_n = |Z_n - E^{F_n}(Z_{n+1})|$. Sous H1) Z_1 est intégrable, et la suite Z_n converge vers 0 en moyenne. Les autres conditions du lemme étant directement vérifiées, la démonstration de la convergence p.s. à 0 de $A_1^2(n)$ revient donc à prouver que, sous H1), H2) et H3) on a :

$$(11) \qquad \sum_{n=1}^{\infty} E(Z'_n) = \sum_{1}^{\infty} E\{|E^{F_n}(A_1^2(n+1)) - A_1^2(n)|\} < \infty$$

Par définition on a :

$$A_1^2(n+1) = \left< \sum_{t=1}^{n} Y_{t,n+1} + Y_{n+1,n+1} \, , \, \sum_{t=1}^{n} Y_{t,n+1} + Y_{n+1,n+1} \right>_H$$

Mais $E^{F_n}(<Y_{n+1,n+1} \, , \sum_{t=1}^{n} Y_{t,n+1}>)$ est nul, puisque $\sum_{t=1}^{n} Y_{t,n+1}$ est F_n mesurable et que $E^{F_n}(Y_{n+1,n+1}) = 0$. Il vient donc :

$$E^{F_n}(A_1^2(n+1)) = \left|\left| \sum_{t=1}^{n} Y_{t,n+1} \right|\right|_H^2 + E^{F_n}(||Y_{n+1,n+1}||_H^2)$$

et

$$(12) \quad \sum_{n=1}^{\infty} E(Z'_n) \leqslant \sum_{n=1}^{\infty} E(||Y_{n+1,n+1}||_H^2) + \sum_{n=1}^{\infty} E\left| \left|\left| \sum_{t=1}^{n} Y_{t,n+1} \right|\right|_H^2 - \left|\left| \sum_{t=1}^{n} Y_{t,n} \right|\right|_H^2 \right|$$

La première des deux séries majorantes convergera, d'après le calcul mené en α, si :

$$(13) \qquad \sum_{n=1}^{\infty} (\alpha_{n,n}^2) < \infty ,$$

et ce n'est autre que H2). Reste donc à obtenir la convergence de la 2ème série majorante de (12), sous H1), H2), H3)

γ) Pour tout n, on majore successivement

$$E\left\{\left|\,\|\sum_{t=1}^{n} Y_{t,n+1}\|_H^2 - \|\sum_{t=1}^{n} Y_{t,n}\|_H^2\right|\right\}$$

par

$$E\left[\{\|\sum_{t=1}^{n} (Y_{t,n}-Y_{t,n+1})\|_H\} \cdot \{\|\sum_{t=1}^{n} Y_{t,n}\|_H + \|\sum_{t=1}^{n} Y_{t,n+1}\|_H\}\right]$$

puis

$$\{E(\|\sum_{t=1}^{n} (Y_{t,n}-Y_{t,n+1})\|_H^2)\}^{1/2} \cdot \{E[\|\sum_{t=1}^{n} Y_{t,n}\|_H + \|\sum_{t=1}^{n} Y_{t,n+1}\|_H]^2\}^{1/2}$$

puis

$$\{E(\|\sum_{t=1}^{n} (Y_{t,n}-Y_{t,n+1})\|_H^2)\}^{1/2} \cdot \{[E(\|\sum_{t=1}^{n} Y_{t,n}\|_H^2)]^{1/2} + [E(\|\sum_{t=1}^{n} Y_{t,n+1}\|_H^2)]^{1/2}\}$$

d'après les inégalités de Schwarz et Minkowski. En appliquant de nouveau le calcul mené au point α), cela donne, par définition des $\alpha_{t,n}$ et $\gamma_{t,n}$ ((1) et (2)) :

$$(14)\; E\{\left|\,\|\sum_{t=1}^{n} Y_{t,n+1}\|_H^2-\|\sum_{t=1}^{n} Y_{t,n}\|_H^2\right|\}\leqslant(\sum_{t=1}^{n}\gamma_{t,n}^2)^{1/2}\{(\sum_{t=1}^{n}\alpha_{t,n}^2)^{1/2}+(\sum_{t=1}^{n}\alpha_{t,n+1}^2)^{1/2}\}$$

Par conséquent, la condition H3), implique bien que la 2ème série majorante sur (12) soit convergente, et H1), H2), H3) permettent d'obtenir (11), donc la convergence p.s. à 0 de $A_1^2(n)$ Q.E.D.

c) *Théorème de convergence*

Pour obtenir la convergence à 0 de $A_2(n)$, nous utiliserons d'abord le lemme suivant, dû à Fritz (3) et qui s'appuie sur le théorème de Banach-Steinhaus.

Lemme 2.- Soit $[K_{t,n}]$, $1 \leq t \leq n$, $n \geq 1$, une suite de séquences d'opérateurs définis sur un Banach B, à valeurs dans un Banach H. Si les $[K_{t,n}]$ vérifient les conditions H4), H5), H6) ; on aura, pour toute suite d'éléments (y_t) de B convergeant vers y :

$$\lim_{n\to\infty} \left(\sum_{t=1}^{n} [K_{t,n}] (y_t) \right) = Cy .$$

Si les $[K_{t,n}]$ vérifient de plus H7), la conclusion reste valable lorsque la suite (y_t) converge vers y au sens de Cesaro.

Nous aurons également besoin du lemme suivant, qui est une version dans $\mathbb{R}$, légèrement élargie, du précédent

Lemme 3.- Soit $(\delta_{t,n})$, $n \geq 1$, $1 \leq t \leq n$, une suite de séquences de réels positifs et β_t une suite de réels positifs convergeant vers 0. On suppose que

$$\lim_{n\to\infty} \delta_{t,n} = 0 \quad (t \text{ fixé}) \quad \text{et} \quad \sup_n \left(\sum_{t=1}^{n} \delta_{t,n} \right) < \infty .$$

Alors la suite $\left(\sum_{t=1}^{n} (\delta_{t,n} \cdot \beta_t) \right)$ converge vers 0 si $n \to \infty$.

Démonstration :

Faute de l'hypothèse : $\left(\sum_{1}^{n} \delta_{t,n}\right) \longrightarrow K$, quand $n \to \infty$, qui ramènerait exactement au lemme 2), on peut raisonner ainsi : pour tout majorant L des $\left(\sum_{t=1}^{n} \delta_{t,n} \right)$ et tout ε , $\varepsilon > 0$, il existe N_ε tel que

$$(t \geq N_\varepsilon) \Longrightarrow \beta_t \leq \varepsilon/L$$

Alors, pour $n \geq N_\varepsilon$ on peut écrire :

$$\sum_{t=1}^{n} (\delta_{t,n} \cdot \beta_t) = \sum_{1}^{N_\varepsilon} (\delta_{t,n} \cdot \beta_t) + \sum_{N_\varepsilon+1}^{n} (\delta_{t,n} \cdot \beta_t) \leq \sum_{1}^{N_\varepsilon} (\delta_{t,n} \cdot \beta_t) + \varepsilon .$$

Comme $\operatorname{Lim}_n \left(\sum_{1}^{N_\varepsilon} \delta_{t,n} \cdot \beta_t \right) = 0$ d'après l'hypothèse, la démonstration est achevée.

Nous pouvons alors énoncer.

Théorème 1.- Supposons que le processus (X_t), $t \in \mathbb{Z}$, vérifie la condition (B), et qu'il existe des opérateurs $[K_{t,n}]$ de B dans H, vérifiant la condition (I).

a) Si H1), H4), H5), H6) sont vérifiées, l'hypothèse B3) implique :

$$\lim_{n\to\infty} E\{\| \sum_{t=1}^{n} K_{t,n}(X_t) - Cg\|_H\} = 0$$

En supposant de surcroît H7) vérifiée, la conclusion demeure sous l'hypothèse affaiblie B4).

b) Si H1), H2), H3), H4), H5), H6) sont vérifiées, l'hypothèse B1) implique :

$$\lim_{n\to\infty} \text{p.s.} \{\| \sum_{t=1}^{n} K_{t,n}(X_t) - Cg\|_H\} = 0$$

En supposant ici encore H7) vérifiée de surcroît, la conclusion demeure valable sous l'hypothèse affaiblie B2).

Démonstration :

α) Comme nous l'avons noté dans le premier paragraphe, les hypothèses faites sur le processus (X_t) et les opérateurs $[K_{t,n}]$, jointes à l'hypothèse H1) ,justifient la majoration de $\|\sum_{t=1}^{n} K_{t,n}(X_t) - Cg\|_H$ par $A_1(n) + A_2(n)$, quantités définies en (6) et (8). La proposition 1 permettant d'affirmer que $A_1(n)$ tend vers 0 en moyenne sous H1) et p.s. sous H1), H2), H3), reste bien à vérifier que les conditions H4), H5), H6), H7) entraînent la convergence à 0 de $A_2(n)$, sous les conditions Bi), soit en moyenne, soit p.s.

β) Pour la convergence en moyenne on écrit :

$$E(A_2(n)) = E\{\|\sum_{t=1}^{n}([K_{t,n}] \cdot g_{X_t}^{\mathcal{F}_{t-1}}) - Cg\|_H\}$$

$$\leq E\{\sum_{t=1}^{n} \|K_{t,n}\|.\|g_{X_t}^{\mathcal{F}_{t-1}}-g_{X_t}\|_B + \|\sum_{t=1}^{n}([K_{t,n}]\ g_{X_t}) - Cg\|_H\}$$

$$\leq \sum_{t=1}^{n} \{\|K_{t,n}\| \cdot E\{\|g_{X_t}^{\mathcal{F}_{t-1}}-g_{X_t}\|_B\}\} + \|\sum_{t=1}^{n}([K_{t,n}]\ g_{X_t}) - Cg\|_H$$

Sous B3) on a : $\lim_t \|g_{X_t}-g\|_B = 0$, donc , d'après le lemme 2), si les $[K_{t,n}]$ vérifient H4), H5), H6) nous obtiendrons :

$$\lim_{n\to\infty} \|\sum_{t=1}^{n}([K_{t,n}]\ g_{X_t}) - Cg\|_H = 0 .$$

De même, sous B3) $\lim_{t\to\infty} E(\|g_{X_t}^{\mathcal{F}_{t-1}}-g_{X_t}\|_B) = 0$, et les $\|K_{t,n}\|$ vérifient les conditions du lemme 3) sous H5) et H6), on en conclut donc :

$$\lim_{n\to\infty} \{\sum_{t=1}^{n} \|K_{t,n}\| \cdot E[\|g_{X_t}^{\mathcal{F}_{t-1}}-g_{X_t}\|_B]\} = 0$$

Ceci démontre la première assertion du théorème.

Lorsque l'on veut affaiblir B3) en B4), on utilise la décomposition :

$$(15)\qquad A_2(n) = \Big|\Big| \sum_{t=1}^{n-1} t \cdot ([K_{t,n}] - [K_{t+1,n}]) \Big(\frac{1}{t} \cdot \sum_{k=1}^{t} (g_{X_k}^{F_{k-1}} - g_{X_k})\Big)$$

$$+ n\,[K_{n,n}] \Big(\frac{1}{n} \sum_{t=1}^{n} (g_{X_k}^{F_{k-1}} - g_{X_k})\Big) + \Big(\sum_{t=1}^{n} [K_{t,n}]\, g_{X_t} - Cg\Big)\Big|\Big|$$

D'où :

$$E(A_2(n)) \leqslant \sum_{t=1}^{n-1} t \cdot ||[K_{t,n}]-[K_{t+1,n}]|| \cdot \{E[||\frac{1}{t} \cdot \sum_{k=1}^{t} (g_{X_k}^{F_{k-1}} - g_{X_k})||_B]\}$$

$$+ n \cdot ||K_{n,n}|| \cdot \{E[||\frac{1}{n} \cdot \sum_{k=1}^{n} (g_{X_k}^{F_{k-1}} - g_{X_k})||_B]\}$$

$$+ ||\sum_{t=1}^{n} [K_{t,n}] \cdot g_{X_t} - Cg||_H$$

Comme B4) implique que $\lim_n ||\frac{1}{n} \sum_{t=1}^{n} g_{X_t} - g||_B = 0$, lorsque H7 est vérifiée, le lemme 2 fournit la convergence à 0 du 3^e majorant. Pour les deux premiers, on peut utiliser le lemme 3, en posant simplement :

$$\begin{cases} \delta_{t,n} = t \cdot ||[K_{t,n}] - [K_{t,n+1}]|| , & 1 \leqslant t \leqslant n-1 \\ \delta_{n,n} = n \cdot ||K_{n,n}|| \\ \beta_t = E(||\frac{1}{t} \sum_{k=1}^{t} (g_{X_k}^{F_{k-1}} - g_{X_k})||_B) \end{cases}$$

puisque H7) implique bien que $\sup_n \left(\sum_{t=1}^{n} \delta_{t,n}\right) < \infty$, et H5) que $\lim_{n\to\infty} \delta_{t,n} = 0$, pour tout t fixé, alors que B4) donne la convergence à 0 de β_t ...

Le a) est aussi entièrement démontré.

γ) Nous devons ensuite démontrer que les conditions H4) à H6), et B1) entraînent :

$$\operatorname*{Lim}_{n\to\infty} \text{p.s.} \ (A_2(n)) = 0 \quad .$$

Or B1) implique que $\operatorname*{Lim}_{t\to\infty} ||g_{X_t} - g||_B = 0$ et $\operatorname*{Lim}_{t\to\infty} \text{p.s.} \ (||g_{X_t}^{F_{t-1}} - g_{X_t}||_B) = 0$. D'après le lemme (2), on aura donc :

$$\operatorname*{Lim}_{t\to\infty} || \sum_{t=1}^{n} [K_{t,n}] \ (g_{X_t}) - Cg||_H = 0 \quad ,$$

et de même, sur un ensemble de probabilité 1 :

$$\operatorname*{Lim}_{t\to\infty} ||(\sum_{t=1}^{n} [K_{t,n}]) \ (g_{X_t}^{F_{t-1}} - g_{X_t})||_H = 0$$

ce qui permet de conclure directement.

Enfin, si on affaiblit B1) en B2) et que l'on rajoute l'hypothèse H7), il suffit d'utiliser la seconde partie du lemme (2) pour obtenir le même résultat, et achever la démonstration.

Q.E.D.

Dans les chapitres qui suivent, nous exhiberons des applications $K_{t,n}$ et les opérateurs $[K_{t,n}]$ correspondants, vérifiant les conditions H1 à H7. Nous allons maintenant montrer que les résultats ainsi démontrés s'appliquent à de nombreux processus (X_t) , $t \in \mathbb{Z}$, à valeurs dans $\mathbb{R}^d$, vérifiant les conditions B i) : c'est l'objet de la deuxième partie du chapitre.

C.I. - 2ème PARTIE

ETUDE DES PROCESSUS VERIFIANT LES CONDITIONS Bi .

a) Variables indépendantes

Si l'on suppose que, pour tout t , $g_{X_t}^{F_{t-1}} = g_{X_t}$, le théorème s'applique lorsque $\|g_{X_t} - g\|_B \to 0$ quand $t \to \infty$, ou que $\lim_{n\to\infty} \|(n^{-1}) (\sum_{t=1}^{n} g_{X_t}) - g\|_B = 0$. Ce cas de variables indépendantes non équidistribuées n'a jamais été abordé jusqu'ici, à notre connaissance, et relève donc des résultats obtenus.

A fortiori, ces résultats englobent aussi le cas de v.a. indépendantes de même densité g , les (Bi) étant toutes automatiquement vérifiées ! Evidemment, on peut alors obtenir le théorème 1) sous des hypothèses allégées : H1), H2), H3), H4) suffisent en ce cas.

b) La stricte stationnarité

<u>Lemme 4</u>.- Supposons le processus (X_t) , $t \in \mathbb{Z}$, défini sur (Ω,A,P), strictement stationnaire et "vérifiant la condition B" pour un Banach séparable B , et qu'il existe une application g qui à y de $\mathbb{R}^d$ et une suite $(x) = (x_1 , x_2 \ldots)$ de $(\mathbb{R}^d)^\infty$ associe $g_{X_t}(y/_{X_{t-1}=x_1, X_{t-2}=x_2}, \ldots)$, pour tout t , $t \in \mathbb{Z}$.

Alors, si pour tout élément u du dual de B , l'application $(x) \to \langle u,g(.,(x))\rangle$ (*) est $\mathcal{B}_{(\mathbb{R}^d)}-\mathcal{B}_{\mathbb{R}}$ mesurable, les variables Y_t qui à ω de Ω associent la densité conditionnelle

---

(*) Nous adoptons la notation classique $\langle u,f\rangle = u[f]$, $f \in B$, u étant dans son dual.

$g_{X_t}(\cdot / X_{t-1}(\omega), X_{t-2}(\omega), \ldots)$, forment une suite d'applications mesurables à valeurs dans B , strictement stationnaire.

Si $E(||Y_1||_B)$ est finie , il existe donc une fonction f définie sur $(B)^\infty$ et invariante par translation telle que, en moyenne et p.s., on ait :

$$||(n^{-1}) \cdot \sum_{t=1}^{n} g_{X_t}^{F_{t-1}} - f[(Y_t), t \in \mathbb{Z}]||_B \longrightarrow 0$$

Si le processus est ergodique (*), a fortiori s'il est "mélangeant" en un sens quelconque, la limite $f((Y_t), t \in \mathbb{Z})$ est égale à l'intégrale de Bochner $E(Y_1)$.

Démonstration :

α) Les applications $(x) \to \langle u, g(., (x))\rangle$ étant mesurables, pour tout élément u du dual de B , l'application

$$(x) \to g(., (x))$$

est mesurable pour les tribus boréliennes de $(\mathbb{R}^d)^\infty$ et B , puisque, B étant séparable, les notions de mesurabilités "faible" et "forte" coïncident (cf. Krengel (4) p. 60 par exemple). On en déduit la mesurabilité de l'application définie sur l'ensemble S des suites (x_i) , $i \in \mathbb{Z}$, d'éléments de $\mathbb{R}^d$, à valeurs dans B , par :

$$L^*((x_i , i \in \mathbb{Z})) = g(., (x)^*) , (x)^* = (x_{-1} , x_{-2} , \ldots) ,$$

(*) C'est-à-dire que la tribu des invariants est dégénérée.

comme enfin la "translation arrière" $T \quad ((x_i, \ i \in \mathbb{Z}) \xrightarrow{T} (x_{i-1}, \ i \in \mathbb{Z}))$ est mesurable sur $(\mathcal{S}, \mathcal{B}_{\mathcal{S}})$, la mesurabilité des Y_t provient de l'égalité :

$$Y_t(\omega) = L^*\{T^t[(X_i(\omega) \quad , \quad i \in \mathbb{Z})]\} \quad .$$

La stationnarité stricte de la suite Y_t découle alors de celle des (X_t) , $t \in \mathbb{Z}$, puisque pour tout k , élément de $\mathbb{N}^*$, tout choix d'indices $(t_1,\ldots,t_k)$ de $\mathbb{Z}^k$, tout choix de boréliens $A_1,\ldots,A_k$, on peut écrire, en posant :

$$A = \bigcap_{j=1}^{k} \{(T^{t_j})^{-1} \ [(L^*)^{-1} \ (A_j)]\}$$

$$P[\bigcap_{j=1}^{k} \{Y_{t_j+h} \in A_j\}] = P[\{\omega \ / \ T^h((X_i(\omega) \ , \ i \in \mathbb{Z})) \in A\}]$$

$$= \mu[(T^h)^{-1} \ (A)]$$

où μ est la mesure induite sur $\mathcal{S}$ par le processus (X_t) , $t \in \mathbb{Z}$. L'invariance de (X_t) , $t \in \mathbb{Z}$, entraîne que $\mu((T^h)^{-1}) = \mu$, donc l'égalité cherchée :

$$P[\bigcap_{j=1}^{k} \{Y_{t_j+h} \in A_j\}] = \mu(A) = P[\bigcap_{j=1}^{k} \{Y_{t_j} \ \in A_j\}] \ .$$

β) Le processus (Y_t) , $t \in \mathbb{Z}$, induit à son tour une mesure ν invariante par translation sur $(B)^\infty$. Si nous appelons $\mathcal{F}$ la tribu des boréliens $\mathcal{I}$ de $(B)^\infty$ tels que :

$$((\alpha_i) \ , \ i \in \mathbb{Z}) \in \mathcal{I} \Longleftrightarrow ((\alpha_{i+1}) \ , \ i \in \mathbb{Z}) \ \in \mathcal{I}$$

et (y_i) , $i \in \mathbb{Z}$, le processus canonique formé des applications coordonnées de $(B)^\infty$, on peut déduire des théorèmes 9.4 et 9.5 de Parthasarathy (5), que si y_1 est ν-intégrable on a :

$$||(n^{-1}) \cdot (\sum_{t=1}^{n} y_t) - E^{\mathcal{F}}(y_1)||_B \longrightarrow 0 \quad , \quad \text{si} \quad n \to \infty \quad ,$$

ν p.s. et en moyenne.

Si le processus est ergodique, $\mathcal{F}$ est dégénérée et $E^{\mathcal{F}}(y_1)$ n'est autre que l'espérance de y_1. Dans le cas général $E^{\mathcal{F}}(y_1)$ représente une application f sur B^{∞}, invariante par translation :

$$f([(y_i) \ , \ i \in \mathbb{Z}]) = f([(y_{i+1}) \ , \ i \in \mathbb{Z}])$$

γ) Les conclusions du lemme sont une retranscription du point précédent, en remarquant simplement que l'ergodicité des (X_t) , $t \in \mathbb{Z}$, entraîne celle des (Y_t) , $t \in \mathbb{Z}$. En effet, si B est un élément de $\mathcal{F}$ et $\bar{B}$ l'ensemble des suites (x) de $\mathcal{S}$ telles que la suite $(L^*(T^t((x)))$, $t \in \mathbb{Z})$, soit dans B , $\bar{B}$ sera lui-même un borélien de $\mathcal{S}$ invariant par translation, car

$$\begin{aligned}
(x) \in \bar{B} &\Longleftrightarrow (L^*(T^{t+1}((x)))\ , \ t \in \mathbb{Z}) \ , \ \text{est élément de } B \ , \ (\text{car } B \in \mathcal{F})\\
&\Longleftrightarrow (L^*[T^t(T(x))] \ , \ t \in \mathbb{Z}) \quad \text{est dans } B\\
&\Longleftrightarrow T((x)) \in \bar{B} \ .
\end{aligned}$$

La suite X_t , $t \in \mathbb{Z}$, étant ergodique , $P[\{\omega / \ (X_t(\omega) \ , t \in \mathbb{Z}) \in I\}]$ vaut 0 ou 1 si I est un borélien de $\mathcal{S}$ invariant par translation. Donc :

$$\begin{aligned}
\nu(B) &= P[\{\omega \ / \ (Y_t(\omega), \ t \in \mathbb{Z}) \in B\}]\\
&= P[\{\omega \ / \ L^*(T^t((X_i(\omega), \ i \in \mathbb{Z}))\ , \ t \in \mathbb{Z}) \in B\}]\\
&= P[\{\omega \ / \ (X_t(\omega), \ t \in \mathbb{Z}) \in \bar{B}\}] = 0 \text{ ou } 1 \ , \text{ si } B \in \mathcal{F} \ .
\end{aligned}$$

On sait enfin (cf. Ash et Gardner (6)) que toutes les conditions de "mélangeance" usuelles impliquent l'ergodicité ....

Corollaire 1.- Si le processus (X_t), $t \in \mathbb{Z}$, strictement stationnaire et vérifiant la condition B est tel que l'application g (cf. lemme 4) soit mesurable par rapport au couple (y, (x)) , et bornée sur $\mathbb{R}^d \times (\mathbb{R}^d)^\infty$, il existera h , $h \in B$ tel que :

$$||(n^{-1}) \cdot (\sum_{t=1}^{n} g_{X_t}^{F_{t-1}}) - h||_B \longrightarrow 0$$

p.s. et en moyenne, lorsque B est un espace $L^p(\mu)$, $1 \leq p < \infty$, ou l'espace $C_o(\mathbb{R}^d)$ des fonctions continues sur $\mathbb{R}^d$, s'annulant à l'infini.

Si (X_t) , $t \in \mathbb{Z}$, est ergodique, h vaut g_{X_t} .

<u>Démonstration</u> :

α) Si $B = L^p(\mu)$, $1 \leq p < \infty$, il est séparable (cf. Adams (7)) et son dual s'identifie à l'espace $L^q(\mu)$ conjugué. Pour toute forme linéaire continue u sur B , on a donc :

$$<u, g(.,(x))> = \int_{\mathbb{R}^d} A(y) \cdot g(y,(x))\, d\mu(y) ,$$

où $A \in L^q(\mu)$. Si g vérifie les conditions du corollaire, la mesurabilité des applications $(x) \to <u, g(.,(x))>$ est donc immédiate, et de même l'intégrabilité de $||Y_t||_{L^p(\mu)}$ exigée dans le lemme 4) puisque :

$$\forall\omega \quad , \quad ||Y_t(\omega)||^p_{L^p(\mu)} = \int_{\mathbb{R}^d} \{g[y,(X_{t-1}(\omega),X_{t-2}(\omega),\ldots)]\}^p\, d\mu(y)$$

$$\leq M^{p-1}$$

β) Pour cela, on note que si $A \in L^q(\mu)$, la fonction qui à (y,ω) associe $A(y) \,.\, g[y \,,\, (X_{t-i}(\omega),\ i \geq 1)]$, est bornée par $M \,.\, |A(y)|$ sur $\mathbb{R}^d \times \Omega$, donc $\mu \otimes P$ intégrable. Cela permet d'écrire

$$E\{\int_{\mathbb{R}^d} A(y).g(y,(X_{t-i}(\omega),\ i\geq 1))d\mu(y))\} = \int_{\mathbb{R}^d} A(y)\{\int_{\Omega} g(y,(X_{t-i}(\omega),i\geq 1)))dP(\omega)d\mu(y)\}$$

On en déduit successivement :

(16) $\quad * \quad \int_{\Omega} g(.,(X_{t-i}(\omega),\ i \geq 1))\ dP(\omega) = g_{X_t}(.)$

En prenant $A = I_C$, on a en effet, pour tout C de $\mathcal{B}_{\mathbb{R}^d}$

$$\int_C \{\int_{\Omega} g(y,(X_{t-i}(\omega),\ i \geq 1))\ dP(\omega)\}\ d\mu(y) = E\{P_{X_t}^{\{(X_{t-i}),i\geq 1\}}(C)\}$$

$$= P_{X_t}(C) = \int_C g_{X_t}(y)\ d\mu(y)$$

ce qui suffit à l'égalité annoncée.

(17) $\quad * \quad \langle u,\ g_{X_t}\rangle = E(\langle u,\ Y_t\rangle)$ pour toute forme linéaire u sur $L^p(\mu)$

Ceci prouve que g_{X_t} est l'intégrale de Bochner de Y_t par rapport à P.

Q.E.D.

γ) L'espace $C_o(\mathbb{R}^d)$ est séparable (cf. Rudin (8)) et d'après le théorème de représentation de Riesz (même référence, par exemple), pour toute forme linéaire u du dual de B , on a :

(18) $$\langle u,\ g(.,(x))\rangle = \int_{\mathbb{R}^d} g(y,(x))\ d\nu(y)$$

pour une mesure ν sur $\mathcal{B}_{\mathbb{R}^d}$. Ici encore, les hypothèses du corollaire impliquent donc toutes celles du lemme 4) par un raisonnement similaire à celui que nous venons de mener.

où $M = \sup_{y,(x)} (g(y,(x)))$, car g est une densité en y ...

On peut donc appliquer le lemme 4 , et ne reste à démontrer que l'égalité de h avec la densité des X_t , lorsque (X_t) , $t \in \mathbb{Z}$, est ergodique.

L'égalité :

$$E\left[\int_{\mathbb{R}^d} g(y,(X_{t-i}(\omega),i \geq 1))\, d\nu(y)\right] = \int_{\mathbb{R}^d} \left\{\int_{\Omega} g(y,(X_{t-i}(\omega),i \geq 1))\, dP(\omega)\right\} d\nu(y)$$

justifiée par la positivité de g , pour toute mesure ν sur $\mathcal{B}_{\mathbb{R}^d}$, permet encore d'écrire successivement (16), en prenant $\nu = I_C \cdot \mu$. puis (17), grâce à (18). On a donc ici encore : $E(Y_t) = g_{X_t}$ lorsque le processus est ergodique, ce qui achève la démonstration.

c) Quelques cas particuliers.

Le corollaire précédent montre que les conditions B2) et B4) sont vérifiées pour les processus stationnaires assez réguliers et vérifiant la condition B pour un espace $L^p(\mu)$, $1 \leq p < \infty$, ou $C_o(\mathbb{R}^d)$. De surcroît les conditions de régularité imposées à g sont usuelles comme le montrent les deux exemples qui suivent

α) Si (X_t) , $t \in \mathbb{Z}$, est markovien d'ordre k , tel que les lois des n-uples $(X_{t_1},\ldots,X_{t_n})$ soient dominées par $\mu^{\otimes n}$, pour tout n , on aura alors

$$g(.,(x)) = g(.,(x_{-1},\ldots,x_{-k})) = \frac{f_{n+1}(x_{-n},\ldots,x_{-1},.)}{f_n(x_{-n},\ldots,x_{-1})} I_{\{f_n \neq 0\}}$$

$\mu^{\otimes n}$ p.s. , où f_n désigne la densité de $(X_1,\ldots,X_n)$. $g(.,(x))$ est alors mesurable par rapport au couple $(y,(x))$, et le corollaire s'applique pourvu que g soit uniformément bornée sur $\mathbb{R}^d \times (\mathbb{R}^d)^k$.

β) Si (X_t) , $t \in \mathbb{Z}$ est un processus strictement stationnaire, admettant une représentation de la forme

(19) $$X_t = \sum_{t=1}^{\infty} a_i X_{t-i} + \varepsilon_t ,$$

où les ε_t sont des variables à densité formant un bruit blanc, et les a_i une suite de matrices $d \times d$, les $g_{X_t}^{F_{t-1}}$ existent, $g(y,(x))$ représentant la densité au point y d'une variable égale à $\varepsilon_t + \sum_{i=1}^{\infty} a_i x_{-i}$. Si les (ε_t) sont gaussiens, par exemple, ceci assure effectivement les conditions du corollaire 1). Sous des conditions de régularité suffisantes, les processus ARMA vérifient (19) et sont ergodiques (cf. Pham et Tran (9) et Withers (10)).

D'une façon plus générale si X_t , strictement stationnaire admet une représentation de la forme

(20) $$X_t = \psi(X_{t-1},X_{t-2},\dots) + \varepsilon_t$$

il suffira que ψ soit continue, de $(\mathbb{R}^d)^{\infty}$ dans $\mathbb{R}^d$, et la densité des variables ε_t constituant le bruit blanc assez régulière, comme dans le cas gaussien, pour arriver aux mêmes conclusions et appliquer le lemme 4).

C.II. - DEUX APPLICATIONS CLASSIQUES.

(A) - Estimation de la fonction de répartition .

a) Définition de l'estimateur

α) Considérons ici encore un processus (X_t) , $t \in \mathbb{Z}$, à valeurs dans $\mathbb{R}^d$, tel que les densités $g_{X_t}^{F_{t-1}}$ et g_{X_t} existent, relativement à une mesure σ-finie sur $\mathbb{R}^d$, et vérifient :

$$(21) \qquad \int_{\mathbb{R}^d} |(n^{-1}) \,.\, \sum_{t=1}^{n} (g_{X_t}^{F_{t-1}} - g_{X_t})| \, d\mu \longrightarrow 0 \quad , \quad \text{si} \quad n \to \infty \quad ,$$

p.s. ou en moyenne, tandis que

$$(22) \qquad \int_{\mathbb{R}^d} |(n^{-1}) \,.\, (\sum_{t=1}^{n} g_{X_t}) - g| \, d\mu \longrightarrow 0 \quad , \quad \text{si} \quad n \to \infty$$

pour un élément g de $L^1(\mu)$.

g sera alors une densité de probabilité, puisque la 2ème condition imposée implique que

$$\lim_{n\to\infty} \left| \int_{\mathbb{R}^d} (\frac{1}{n} \sum_{1}^{n} g_{X_t}) \, d\mu - \int_{\mathbb{R}^d} g \, d\mu \right| = 0$$

et donc : $\int_{\mathbb{R}^d} g \, d\mu = 1$, la positivité de g résultent de la convergence μ-p.s. d'une sous-suite des densités $(n^{-1} \,.\, \sum_{1}^{n} g_{X_t})$ vers leur limite L^1.

Les conditions imposées ci-dessus ne sont autres que "la condition $L^1(\mu)$" et les conditions B2) ou B4) du chapitre I, et nous savons pouvoir identifier g à la densité commune des X_t si le processus est strictement stationnaire, et ergodique, les $g_{X_t}^{\mathcal{F}_{t-1}}$ vérifiant les conditions imposées au corollaire 1.

On peut donc, en tous cas, définir la fonction de répartition F d'une variable de densité g :

$$F((t_1,\dots,t_d)) = \int_{\prod_{i=1}^{d}]-\infty,t_i]} g \cdot d\mu$$

Le but de ce paragraphe est d'étudier le comportement asymptotique d'un estimateur $\hat{f}_n^1$ de F, dont on montrera qu'il converge, sous les seules hypothèses évoquées ci-dessus, dans les espaces $L^2(\nu)$ associés aux mesures ν bornées sur $(\mathbb{R}^d, \mathcal{B}_{\mathbb{R}^d})$.

β) L'estimateur $\hat{f}_n^1$ sera défini par

$$\hat{f}_n^1 = (n^{-1}) \sum_{t=1}^{n} T(X_t) \tag{23}$$

T représentant l'application qui a tout $x = (x_1,\dots,x_d)$ de $\mathbb{R}^d$ associe la fonction indicatrice de l'ensemble :

$$\prod_{j=1}^{d} [x_j,\infty[$$

T est une application mesurable de $(\mathbb{R}^d, \mathcal{B}_{\mathbb{R}^d})$ dans $L^2(\nu)$ qui est un espace de Hilbert séparable. (Nous savons (cf. démonstration du lemme 4) Chap. I) que la mesurabilité cherchée est obtenue si pour toute forme linéaire u sur $L^2(\nu)$ l'application $x \to \langle u, T(x)\rangle$ est mesurable, mais ceci revient à écrire que pour tout g de $L^2(\nu)$

$$x \longrightarrow \int g(y)\,.\,I_{(\prod_1^d [x_j,\infty[)}(y)\; d\nu(y) = \int g(y) \times I_{(\mathbb{R}^+)^d}(y-x)\; d\nu(y)$$

est $\mathcal{B}_{\mathbb{R}}-\mathcal{B}_{\mathbb{R}}$ mesurable, ce qui découle immédiatement du théorème de Fubini.

Finalement $\hat{f}_n^1$ est donc bien de la forme $\sum_{t=1}^{n} K_{t,n}^1(X_t)$ pour une suite de séquences d'applications mesurables $K_{t,n}^1$, définies par $K_{t,n}^1 = (n^{-1})\,.\,T$, $n \geq 1$, $1 \leq t \leq n$. Le théorème démontré au chapitre I va alors nous permettre d'obtenir la convergence souhaitée.

b) Théorème de convergence

Théorème 2.- Soit (X_t) , $t \in \mathbb{Z}$ un processus à valeurs dans $\mathbb{R}^d$, tel que les densités g_{X_t} et $g_{X_t}^{\mathcal{F}_{t-1}}$ existent et vérifient (22). Soit F la fonction de répartition d'une variable ayant la densité g limite de (22).

Alors pour toute mesure ν finie sur $\mathbb{R}^d$, nous aurons

$$\lim_{n\to\infty} \int_{\mathbb{R}^d} (\hat{f}_n^1(x) - F(x))^2\, d\nu(x) = 0 \quad , \quad \text{p.s.} \quad \text{(resp. en moyenne)}$$

si

$$\int_{\mathbb{R}^d} |(n^{-1})\ \{\sum_{t=1}^{n} (g_{X_t}^{\mathcal{F}_{t-1}} - g_{X_t})\}|\, d\mu \quad \text{tend vers } 0 \text{ p.s. (resp. en moyenne)}$$

Sous les mêmes hypothèses, nous aurons donc

$\lim_{n\to\infty} \int_C |\hat{f}_n^1(x) - F(x)|\, dx = 0$, p.s. ou en moyenne, pour tout compact C de $\mathbb{R}^d$.

Démonstration :

α) Pour tout x de $\mathbb{R}^d$, on a :

$$||K^1_{t,n}(x)||^2_{L^2(\nu)} = (n^{-2}) \,.\, ||I_{(\prod_{j=1}^{d} [x_j,\infty[)}||^2_{L^2(\nu)} = O(n^{-2})$$

et $$||K^1_{t,n}(x)-K^1_{t,n+1}(x)||^2_{L^2(\nu)} = (n^{-1}-(n+1)^{-1})^2||.I_{\prod_{j=1}^{d} [x_j,\infty[}||^2_{L^2(\nu)} = O(n^{-4})$$

On en déduit immédiatement que les $K^1_{t,n}$ vérifient H1), H2), H3).

β) Appelons $[H]$ l'opérateur défini ponctuellement, au point $(y) = (y_1,\ldots,y_d)$ par

$$(24) \qquad ([H](g))\,(y) = \int_{A_y} g(x)\, d\mu(x) \ , \quad \text{où } A_y = \prod_{i=1}^{d}]-\infty \, , \, y_i]$$

$[H](g)$ est alors bien défini si $g \in L^1(\mu)$, et appartient à $L^2(\nu)$ (c'est une fonction bornée par $||g||_{L^1(\mu)}$ et ν est finie). Nous allons alors poser, pour $n \geq 1$ et $1 \leq t \leq n$:

$$[K^1_{t,n}] = (n^{-1}) \,.\, [H]$$

Il vient :

* Les $[K^1_{t,n}]$ vérifient la "condition I"

Ceci revient à vérifier que pour tout g , $g \in L^1(\mu)$, $[H]\ (g)$ est égal à l'intégrale de Bochner $\int_{\mathbb{R}^d} g(x) \,.\, T(x)\, d\mu(x)$, de par la définition des $(K^1_{t,n})$.

Or, pour tout a , $a \in L^2(\nu)$, et g , $g \in L^1(\mu)$, l'application ℓ définie sur $\mathbb{R}^d \times \mathbb{R}^d$ par

$$\ell(x,y) = a(y)\, g(x) \,.\, I_{A_y}(x)$$

est $\mu \otimes \nu$ intégrable : à x fixé , on a en effet :

$$\int_{\mathbb{R}^d} |I_{A_y}(x).a(y)| \, d\nu(y) \leq ||a||_{L^1(\nu)} < \infty \quad (a \in L^2(\nu), \quad \nu \quad \text{finie})$$

ce qui implique l'intégrabilité de $|\ell|$.

Alors l'égalité des 2 intégrales itérées de ℓ donne

$$<a, [H](g)>_{L^2(\nu)} = \int_{\mathbb{R}^d} <a \, , \, g(x) \, . \, T(x)>_{L^2(\nu)} \, d\mu(x)$$

ce qui revient à écrire que pour toute forme linéaire u sur $L^2(\nu)$, on a :

$$<u, [H](g)> = \int_{\mathbb{R}^d} <u, \, g(.) \, . \, T(.)> d\mu$$

Les propriétés standard de l'intégrale de Bochner, et son unicité permettent alors de conclure.

* Les $[K^1_{t,n}]$ vérifient les conditions H4), H5), H6), H7)

H4) est vérifiée "par nature" puisque, pour tout n, $\sum_{t=1}^{n} [K_{t,n}]$ est égal à $[H]$

Pour les autres propriétés, il suffit de constater que

$$||[H](g)||^2_{L^2(\nu)} = \int_{\mathbb{R}^d} \{\int_{A_y} g(x) \quad d\mu(x)\}^2 \, d\nu(y)$$

$$\leq \{ \sup_{y \in \mathbb{R}^d} \int_{A_y} |g(x)| \, d\mu(x)\}^2 \, . \, \nu(\mathbb{R}^d)$$

$$\leq \nu(\mathbb{R}^d) \, . \, ||g||^2_{L^1(\mu)}$$

et que donc $||H|| \leq [\nu(\mathbb{R}^d)]^{1/2}$, et $||K^1_{t,n}|| = O(n^{-1})$ pour tous t et n .

H5) et H6) en résultent directement, H7) de même en notant que pour tous t , $t \in \{1,\dots,n\}$, $[K_{t,n}] = [K_{t+1,n}]$.

γ) Sous les hypothèses apportées dans le théorème, le processus vérifie les conditions $L^1(\mu)$ et B2, ou B4, comme nous l'avions noté en introduction. Maintenant les $[K^1_{t,n}]$ vérifient la condition I , et $\sum_{t=1}^{n} [K^1_{t,n}]$ converge vers H. Toutes les autres conditions Hi) étant vérifiées, le théorème 1 s'applique.

Comme la limite g en (22) est une densité de probabilité et que $[H]$ associe à une densité la fonction de répartition correspondante, on obtient exactement la première conclusion du théorème. En choisissant pour ν la mesure de Lebesgue sur un compact C , on en déduit, a fortiori,

$$\lim_{n\to\infty} \int_C |\hat{f}^1_n(x) - F(x)| \, dx = 0 \qquad \text{Q.E.D.}$$

B - Estimation de la loi .

a) *Introduction*

α) Considérons ici encore un processus (X_t) , $t \in \mathbb{Z}$, vérifiant (21) et (22). Le but de ce paragraphe est de démontrer qu'alors

$$\frac{1}{n} \sum_{t=1}^{n} \delta_{X_t} \Longrightarrow P$$

où P est la loi de densité g (la limite apparue en (22)), au sens de la convergence faible sur l'espace des probabilités $\mathcal{P}$, définies sur $\mathbb{R}^d$.

A cette fin, suivant la théorie développée par C. Guilbart (11) et récemment par C. Suquet (12), nous allons identifier $\mathcal{P}$ à un sous-espace d'un espace de Hilbert séparable, de manière à utiliser les résultats du chapitre I.

β) Selon la terminologie usuelle, nous appelons "noyau reproduisant" sur $\mathbb{R}^d$ une fonction N bi-mesurable (pour $\mathcal{B}_{\mathbb{R}^d}$) , à valeurs réelles, symétrique et semi-définie positive. Une telle fonction détermine alors un unique (à une équivalence près) espace H_N , de Hilbert, qui l'admette comme "noyau reproduisant" c'est-à-dire que

* Toutes les applications $K(.,y)$, $y \in \mathbb{R}^d$, sont dans H_N
* Le produit scalaire $\langle , \rangle_{H_N}$ vérifie :

$$\forall f \ , \ f \in H_N \ , \quad \langle f, K(.,y)\rangle_{H_N} = f(y)$$

En pratique, H_N sera la fermeture de l'ensemble H_o composé des combinaisons linéaires finies d'éléments $K(.,y)$, muni du produit scalaire :

$$(K(.,x), K(.,z)) = K(x,z) \quad .$$

γ) Les auteurs sus-cités ont alors démontré que l'on pouvait munir $\mathbb{R}^d$ d'un noyau N continu et borné tel que, en définissant sur l'ensemble M des mesures à signes bornées sur $(\mathbb{R}^d, \mathcal{B}_{\mathbb{R}^d})$, un produit scalaire par :

$$(25) \qquad \langle \mu,\nu\rangle_N = \int_{\mathbb{R}^d\times\mathbb{R}^d} N \, d\mu \otimes \nu$$

on en fasse un espace préhilbertien séparé, que nous noterons M_N .

De plus, l'application ψ

$$(26) \qquad \begin{cases} M_N \longrightarrow H_N \\ \mu \longrightarrow \int K(.,y) \, d\mu(y) \end{cases}$$

est une injection canonique, $\mathcal{J}(M_{\mathcal{N}})$ étant un sous-espace vectoriel dense de $H_{\mathcal{N}}$. Enfin, si $f \in H_{\mathcal{N}}$ et $\mu \in \mathcal{M}$, on a :

(27) $$<f, \mathcal{J}(\mu)>_{\mathcal{N}} = \int f \, d\mu \, .$$

ω) Comme on peut choisir $\mathcal{N}$ tel que $H_{\mathcal{N}}$ soit séparable et que la topologie trace de $H_{\mathcal{N}}$ sur $\mathcal{J}(M_{\mathcal{N}})$ coïncide (à une isométrie près) avec la topologie faible, l'application Δ

$$\Delta \left\{ \begin{array}{l} \mathbb{R}^d \longrightarrow \mathcal{P} \\ x \longrightarrow \delta_x \end{array} \right.$$

peut être considére comme à valeurs dans l'espace de Hilbert séparable $H_{\mathcal{N}}$ en assimilant δ_x à $\mathcal{J}(\delta_x) = K(.,x)$ ...

De surcroît cette application est mesurable pour les tribus boréliennes : elle est continue, puisque si $x_n \to x$, on a $\delta_{x_n} \overset{\mathcal{W}}{\Longrightarrow} \delta_x$, donc $\Delta(x_n) \xrightarrow[H_{\mathcal{N}}]{} \Delta(x)$ (on peut voir aussi que pour toute f de $H_{\mathcal{N}}$, $x \longrightarrow <f, \mathcal{J}(\delta_x)>_{H_{\mathcal{N}}} = f(x)$, est mesurable ....) .

Finalement, en posant

(28) $$\hat{f}_n^3 = \frac{1}{n} \sum_{t=1}^{n} \delta_{X_t} = \sum_{t=1}^{n} K_{t,n}^3(X_t)$$

avec $K_{t,n}^3(x) = (n^{-1}) \cdot \Delta(x)$, on est bien ramené à étudier une suite d'éléments aléatoires à valeurs dans un Hilbert séparable, comme au chapitre I ... On en déduit :

b) Théorème de convergence

Théorème 3. Considérons un processus (X_t) tel que les densités $g_{X_t}^{F_{t-1}}$ et g_{X_t} existent et vérifient

$$\lim_{n\to\infty} \int_{\mathbb{R}^d} |(n^{-1}) \cdot (\sum_{t=1}^{n} g_{X_t}) - g| \, d\mu = 0$$

$$\text{Lim p.s.}_{n\to\infty} \int_{\mathbb{R}^d} (n^{-1}) \, |\sum_{t=1}^{n} (g_{X_t}^{F_{t-1}} - g_{X_t})| \, d\mu = 0$$

Alors si P est la probabilité de densité g, on a :

$$\text{Lim p.s.}_{n\to\infty} (n^{-1} \cdot (\sum_{t=1}^{n} \delta_{X_t})) = P \quad , \text{ au sens de la convergence faible.}$$

Démonstration :

α) Vérifions les conditions H1), H2), H3) pour les $K_{t,n}^3$.

Comme on a :

$$\forall x \, , \quad x \in \mathbb{R}^d \quad , \qquad ||\Delta(x)||_H = \langle \psi(\delta_x), \psi(\delta_x) \rangle = N(x,x) \; ,$$

et que N est bornée, on en déduit directement que $\alpha_{t,n} = O(n^{-1})$, $\gamma_{t,n} = O(n^{-2})$, donc H1), H2) et H3).

β) Considérons l'opérateur $[T]$ qui à tout élément h de $L^1(\mu)$ associe l'élément de H égal à $\int N(.,x)\, h(x)\, d\mu(x)$ (i.e. égal à $\psi(\rho)$ où ρ est la mesure à signe de densité h). Alors en posant $[K_{t,n}^1] = (n^{-1})\,[T]$, on définit une suite d'opérateurs "vérifiant la condition I" au sens du chapitre I.

Ceci revient à prouver que l'on a :

$$\int_{\mathbb{R}^d} \Delta(x)\ g(x)\ d\mu(x) = [T](g)$$

pour tout $g \in L^1(\mu)$. Or au point y , $y \in \mathbb{R}^d$, $\int_{\mathbb{R}^d} \Delta(x)\ g(x)\ d\mu(x)$ prend comme valeur :

$$\int_{\mathbb{R}^d} \{\int_{\mathbb{R}^d} \Delta(x)\ g(x)\ d\mu(x)\}\ d(\delta_y)$$

soit
$$< \int_{\mathbb{R}^d} \Delta(x)\ g(x)\ d\mu(x),\ \psi(\delta_y)>_{H_N} \qquad \text{(cf. (27))}$$

ou
$$\int_{\mathbb{R}^d} <\psi(\delta_y)\ ,\ \Delta(x)\ g(x)>_{H_N} d\mu(x)$$

par propriété standard de l'intégrale de Bochner. D'où enfin :

$$\begin{aligned}(\int_{\mathbb{R}^d} \Delta(x)\ g(x)\ d\mu(x))\ (y) &= \int_{\mathbb{R}^d} \{<\psi(\delta_y),\ \psi(\delta_x)>_{H_N}\}\ g(x)\ d\mu(x)\\ &= \int_{\mathbb{R}^d} <\delta_y,\delta_x>_N\ g(x)\ d\mu(x)\\ &= \int_{\mathbb{R}^d} N(y,x)\ g(x)\ d\mu(x)\\ &= ([T](g))\ (y)\end{aligned}$$

pour tout y de $\mathbb{R}^d$ ... C.Q.F.D.

γ) Puisque $[K^3_{t,n}] = (n^{-1})\ [T]$, la condition H4) est automatiquement vérifiée, et l'opérateur limite est $[T]$, qui à la limite g des g_{X_t} associera donc la loi de probabilité $[T]g$ dont g est densité.

Pour obtenir H5), H6), H7), reste à déterminer $||[T]||$. Or si $g \in L^1(\mu)$, on a

$$||[T]g||_{H_N} = ||\rho||_{M_N} \qquad \text{si} \qquad [T](g) = \emptyset(\rho)$$

Mais nous avons vu que ρ n'est autre que la mesure de densité g par rapport à μ. Il vient :

$$||[T]g||_{H_N} = \{\int_{\mathbb{R}^d \times \mathbb{R}^d} N(x,y)\ d(\rho \otimes \rho)\}^{1/2}$$

$$= \{\int_{\mathbb{R}^d \times \mathbb{R}^d} N(x,y)\ g(x)\ g(y)\ d\mu(x)\ d\mu(y)\}^{1/2}$$

$$\leqslant \{\sup_{(x,y)} |N(x,y)|\}^{1/2} \cdot \{\int_{\mathbb{R}^d} |g(x)|\ d\mu(x)\} .$$

Par conséquent : $||[T]g||_{H_N} = O(||g||_{L^1(\mu)})$, donc $||[T]|| < M$.

On en déduit que $||[K^3_{t,n}]|| = O(n^{-1})$, et H5) et H6) sont automatiquement vérifiées. Il en est de même pour H7) car $[K^3_{t,n}] = [K^3_{t,n+1}]$, $t = 1,\dots,n$.

ω) On conclut alors du théorème 1 que sous les conditions du théorème 3 on a

$$||\sum_{t=1}^{n} K^3_{t,n}(X_t) - [T]g||_{H_N} \longrightarrow 0 \ \text{ p.s. quand } \ n \to \infty$$

(on a évidemment aussi $E(||\sum_{t=1}^{3} K^3_{t,n}(X_t) - [T]g||_{H_N}) \longrightarrow 0$, si $n \to \infty$, en remplaçant la convergence p.s. par la convergence en moyenne, dans les hypothèses du théorème).

Mais ceci signifie que p.s. la suite de probabilités $\frac{1}{n} \sum_{t=1}^{n} \delta_{X_t}$ converge vers P , probabilité de densité g , dans M_N . Comme la topologie de M_N coïncide avec la topologie faible, on en déduit le théorème.

ESTIMATION D'UNE DENSITE ET DE SES DERIVEES :
Les points de vue L^1 et L^2.

(Chapitres III à V)

ESTIMATION PAR PROJECTION

(Chapitre III)

Dans le cas où le processus observé "vérifie la condition $L^2(\mu)$ " , μ étant une mesure σ-finie régulière sur $(\mathbb{R}^d, \mathcal{B}_{\mathbb{R}^d})$, les conditions (B_i) expriment la convergence en moyenne quadratique des densités g_{X_t} , ou $(n^{-1}) (\sum_1^n g_{X_t})$, vers un élément g de $L^2(\mu)$. $L^2(\mu)$ étant un Hilbert séparable, g peut s'approcher par la suite de ses projetés g_n (projection orthogonale) sur des sous-espaces fermés M_n de $L^2(\mu)$, de dimensions $q(n)$ croissant vers l'infini avec n : une suite d'estimations $\hat{f}_n^3$ des g_n peut ainsi fournir un estimateur convergent de g, alors que l'estimation de g_n se ramène en pratique à celle d'un nombre fini de paramètres, ses coordonnées sur une des bases de M_n (cf. Bosq (13) ou Delecroix (14) pour une discussion générale). Comme la densité g est usuellement la densité commune des X_t (cf. chap. I), $\hat{f}_n^3$ sera un "estimateur, par projection, d'une densité de probabilité" .

α) Considérons une base orthonormale (e_i) , $i \in \mathbb{N}^*$, de $L^2(\mu)$, dont nous supposons, une fois pour toutes, qu'elle vérifie :

(29) $$\exists\, M , \quad \sup_{i\in\mathbb{N}^*} \ \sup_{x\in\mathbb{R}^d} |e_i(x)| \leq M \ .$$

(Cette hypothèse, nullement indispensable, s'avérera simplificatrice. Elle est peu contraignante, puisque vérifiée par les bases usuelles de $L^2(\mathbb{R}^d)$)
M_n sera alors l'espace engendré par $e_1,\ldots,e_{q(n)}$, $q(n)$ représentant une suite d'entiers tels que $\lim\limits_{n\to\infty} q(n) = \infty$. Cela donne :

(30) $$g = \sum_{i=1}^{\infty} a_i\, e_i \ , \quad g_n = \sum_{i=1}^{q(n)} a_i\, e_i \ , \quad a_i = \int_{\mathbb{R}^d} g \cdot e_i \cdot d\mu$$

A l'observation $X_t = x$, on va associer un élément de $L^2(\mu)$ de la forme $\sum_{i=1}^{q(n)} \hat{a}_i e_i$, qui puisse être considéré comme projection sur M_n de la densité h d'une loi "proche" de δ_x . Or d'après (30), a_i , le coefficient de g sur l'élément e_i représente l'espérance d'une variable $e_i(X)$, X étant de densité g. On prendra donc $\hat{a}_i = e_i(x)$, en assimilant la loi de X à δ_x , ce qui amène à la définition suivant de l'estimateur cherché :

$$(31) \qquad \begin{cases} \hat{f}_n^3 = \sum_{t=1}^{n} K_{t,n}^3(X_t) \\ K_{t,n}^3(x) = (n^{-1}) \, . \, \{\sum_{i=1}^{q(n)} e_i(x)\, e_i\} \end{cases} .$$

Les $K_{t,n}^3$ *sont bien à valeurs dans un espace de Hilbert séparable.* De plus elles sont mesurables pour les tribus boréliennes (il suffit, comme d'habitude, de vérifier que pour tout f de $L^2(\mu)$, $x \longrightarrow \langle f, K_{t,n}^3(x)\rangle_{L^2(\mu)}$ est mesurable, ce qui est direct). Donc $\hat{f}_n^3$ est bien de la forme évoquée au chapitre I, dont nous pouvons utiliser les résultats.

β) La détermination de $\hat{f}_n^3$ s'avérant parfois lourde au niveau calcul, on peut en proposer une version modifiée rendant récursif l'estimateur obtenu. *Posons en effet*

$$(32) \qquad \hat{f}_n^4 = \sum_{t=1}^{n} K_{t,n}^4(X_t) \quad \text{avec} \quad K_{t,n}^4(x) = (n^{-1}) \, . \sum_{j=1}^{q(t)} e_j(x)\, e_j$$

On obtient :

$$(33) \qquad \hat{f}_n^4 = \sum_{t=1}^{q(n)} \hat{b}_{t,n}\, e_t \quad , \quad \text{avec} \quad \hat{b}_{j,n} = (n^{-1}) \, . \sum_{t=\alpha(j)}^{n} e_j(X_t)$$

où $\alpha(j) = \inf \{i \ / \ q(i) \geq j\}$ alors que $\hat{f}_n^3$ s'écrivait (cf. 31) :

$$\hat{f}_n^3 = \sum_{j=1}^{q} \hat{a}_{j,n} \, e_j \qquad \text{avec} \qquad \hat{a}_{j,n} = (n^{-1}) \; . \sum_{t=1}^{n} e_j(X_t) \; .$$

Si, pour un n, $q(n+1) = q(n)+1$, la détermination de $\hat{b}_{q(n+1),n+1}$ n'exigera qu'un calcul de la valeur $e_{q(n)+1}(X_{n+1})$, puisque $\alpha(q(n+1)) = n+1$ en ce cas, alors que le calcul de $\hat{a}_{q(n+1),n+1}$ demande $(n+1)$ calculs du même type, la détermination des autres coefficients $\hat{a}_{j,n+1}$, ou $\hat{b}_{j,n+1}$, pour $j = 1,\ldots,q(n)$, à partir des $\hat{a}_{j,n}$ et $\hat{b}_{j,n}$, exigeant globalement les $q(n)$ calculs des $e_j(X_{n+1})$ dans chaque méthode. L'emploi de $\hat{f}_n^4$ permet donc de réaliser un gain de temps dans les calculs, qui se trouvent facilités, de surcroît, par la formule récursive :

$$(n+1) \; \hat{f}_{n+1}^4(x) = n \; \hat{f}_n^4(x) + \sum_{j=1}^{q(n+1)} e_j(X_{n+1}) \; . \; e_j(x) \quad .$$

$\hat{f}_n^4$ se basant sur le même principe que $\hat{f}_n^3$ (approcher f par $\sum_{i=1}^{q(n)} a_i \, e_i$, puis estimer les a_i (cf. (33) ...) convergera donc sous le même type de conditions, mais comme nous allons le voir, avec une contrainte plus stricte sur $q(n)$, qui devra croître très lentement vers l'infini : dans la mesure où l'estimateur $\hat{b}_{j,n}$ de a_j n'utilise que les $(n-\alpha(j))$ dernières variables de l'échantillon, la croissance trop rapide de $\alpha(j)$ (donc celle de $q(n)$) vers l'infini empêcherait évidemment $\hat{f}_n^4$ de converger.

γ) La 1ère partie du chapitre consistera en l'obtention concrète de la convergence de $\hat{f}_n^3$ et $\hat{f}_n^4$ vers g dans $L^2(\mu)$. En une seconde partie, nous étudierons le problème de la convergence d'estimateurs dérivés de $\hat{f}_n^3$, et définis précisément alors, vers les dérivées partielles de g.

C.III - 1ère PARTIE

CONVERGENCE DE L'ESTIMATEUR DE LA DENSITE LIMITE

Théorème 4. - Supposons que le processus (X_t) , $t \in \mathbb{Z}$, à valeurs dans $\mathbb{R}^d$, vérifie la "condition $L^2(\mu)$" , pour une mesure μ σ-finie, régulière, sur $\mathbb{R}^d$ et qu'il existe g dans $L^2(\mu)$ telle que :

$$\lim_{n\to\infty} \left(\left\|(n^{-1}) \cdot \left(\sum_{t=1}^{n} g_{X_t}\right) - g\right\|_{L^2(\mu)}\right) = 0 \ .$$

Soit (e_i) , $i \in \mathbb{N}$ une base orthonormale uniformément bornée de $L^2(\mu)$, et $q(n)$ une suite croissante d'entiers positifs vérifiant $\lim_{n\to\infty} [q(n)^{-1}] = \lim_{n\to\infty} [\frac{q(n)}{n}] = 0$.

a) Supposons que $\lim_{n\to\infty} E\left(\left\|(n^{-1}) \cdot \sum_{t=1}^{n} \left(g_{X_t}^{F_{t-1}} - g_{X_t}\right)\right\|_{L^2(\mu)}\right) = 0$,

alors

α) $\lim_{n\to\infty} E\left(\|\hat{f}_n^3 - g\|_{L^2(\mu)}\right) = 0$,

β) $\lim_{n\to\infty} E\left(\|\hat{f}_n^4 - g\|_{L^2(\mu)}\right) = 0$, pourvu que $q(n)$ vérifie :

$$\sup_n \left\{\left[\sum_{t=1}^{q(n)} t \sqrt{q(t+1)-q(t)}\right]/n\right\} < \infty \ . \tag{34}$$

b) Supposons que $\lim_{n\to\infty} \text{p.s.} \left(\left\|(n^{-1}) \cdot \sum_{t=1}^{n} \left(g_{X_t}^{F_{t-1}} - g_{X_t}\right)\right\|_{L^2(\mu)}\right) = 0$,

alors :

α) Si $\sum_{n\geq 1} (q(n)/n^2) < \infty$ et que

$\sum_{n\geq 1} \{n^{-1} . (q(n))^{1/2} (q(n+1)-q(n))^{1/2}\} < \infty$, on a

$$\lim_{n\to\infty} \text{p.s.} \ (\|\hat{f}_n^3 - g\|_{L^2(\mu)}) = 0 \ .$$

En particulier, si $q(n)$, croissante, vérifie $1 - \frac{1}{n} \leq \frac{q(n)}{q(n+1)}$ pour n assez grand, et $\sum_{n\geq 1} \frac{q(n)}{n^{3/2}} < \infty$, toutes les conditions requises sont obtenues.

β) Si $\sum_{n\geq 1} (q(n)/n^2) < \infty$, $q(n)$ vérifiant (34), on a

$$\lim_{n\to\infty} \text{p.s.} \ (\|\hat{f}_n^4 - g\|_{L^2(\mu)}) = 0 \ .$$

c) Si $q(n)$ est la partie entière d'une fonction t positive continue, strictement croissante sur $\mathbb{R}^+$, les conditions de b) α) sont vérifiées lorsque :

$$\sum_{k\geq 1} \left[\sqrt{k} \ / \ t^{-1}(k)\right] < \infty$$

(d'où les choix possibles de $t(x) = x^\alpha$, $0 < \alpha < 2/3$, ou $t(x) = \text{Log } x$), la condition du a) β) étant elle vérifiée si

$$\sup_k \{(\sum_{j\leq k} t^{-1}(j)) \ . \ \left[t^{-1}(k)\right]^{-1}\} < \infty \ ,$$

(d'où le choix possible de $t(x) = \text{Log } x$).

<u>Démonstration</u> :

α) Pour tout x de $\mathbb{R}^d$, on a :

$$\|K_{t,n}^3\|^2_{L^2(\mu)} = (n^{-2}) \ . \ \|\sum_{i=1}^{q(n)} e_i(x) \ e_i\|^2_{L^2(\mu)} = (n^{-2}) \ . \ (\sum_{i=1}^{q(n)} e_i^2(x))$$

et

$$||K^3_{t,n}-K^3_{t,n+1}||^2_{L^2(\mu)} = ||(n^{-1}) \cdot (\sum_{i=1}^{q(n)} e_i(x)\, e_i) - (n+1)^{-1} (\sum_{i=1}^{q(n+1)} e_i(x) e_i)||^2_{L^2(\mu)}$$

$$= (n(n+1))^{-2} \cdot (\sum_{i=1}^{q(n)} e_i^2(x)) + (n+1)^{-2} \cdot \sum_{i=q(n)+1}^{q(n+1)} e_i^2(x)$$

Ce qui donne, puisque les e_i sont uniformément bornées :

$$(\alpha^3_{t,n}) = \sup_{x\in \mathbb{R}^d} ||K^3_{t,n}(x)||_{L^2(\mu)} = \mathcal{O}(\sqrt{q(n)}/n) \quad .$$

et $$(\gamma^3_{t,n}) = \sup_{x\in \mathbb{R}^d} ||K^3_{t,n}(x)-K^3_{t,n+1}(x)||_{L^2(\mu)}$$

$$= \mathcal{O}\{[q(n)] \cdot (n^{-4}) + [q(n+1)-q(n)]\ (n+1)^{-2}\}^{1/2}$$

Ceci montre, pour $K^3_{t,n}$ que H1) est vérifiée si $\lim\limits_{n} (n^{-1}) \cdot [q(n)] = 0$, H2) et H3) en ajoutant les conditions de b α).

β) Le même calcul mené pour les $K^4_{t,n}$ donnera cette fois :

$$(\alpha^4_{t,n})^2 = \mathcal{O}((n^{-2}) \cdot q(t)) \ ,$$

et $$(\gamma^4_{t,n})^2 = \mathcal{O}((n^{-4}) \cdot q(t)) \ ,$$

d'où découlent directement les équivalences , valables pour les $K^4_{t,n}$

* H1) $\Longleftrightarrow$ $(\sum_{t=1}^{n} q(t))/n^2 \longrightarrow 0$, si $n \to \infty$
* H2) $\Longleftrightarrow$ $(\sum_{n\geq 1} (q(n)/n^2)) < \infty$
* H3) $\Longleftrightarrow$ $\sum_{n\geq 1} \{(\sum_{t=1}^{n} q(t))/n^3\} < \infty$

Comme $q(n)$ est croissante et que $\lim\limits_{n\to\infty} (q(n)/n) = 0$, la condition équivalente à H2) assure l'ensemble δ) Soit P_n l'opérateur de projection orthogonal sur M_n, de $L^2(\mu)$ dans $L^2(\mu)$. Posons $[K^3_{t,n}] = (n^{-1})\, P_n$, $n \geq 1$, $1 \leq t \leq n$. Alors :

* Les $[K^3_{t,n}]$ vérifient la "condition I", ce qui revient à montrer que si $g \in L^1(\mu) \cap L^2(\mu)$, on a :

$$P_n(g) = \int_{\mathbb{R}^d} \Big(\sum_{i=1}^{q(n)} e_i(x)\, e_i\Big) \cdot g(x)\, d\mu(x)$$

Ce sera vrai si les deux éléments ont même produit scalaire avec tous les e_i. Or pour tout $j_o \in \mathbb{N}^*$, on a (propriété standard de l'intégrale de Bochner) :

$$\Big\langle e_{j_o} , \int_{\mathbb{R}^d} \Big(\sum_{i=1}^{q(n)} e_i(x)\, e_i\Big)\, g(x)\, d\mu(x)\Big\rangle_{L^2(\mu)}$$

$$= \int \Big\langle e_{j_o} , \sum_{i=1}^{q(n)} e_i(x)\, e_i\Big\rangle_{L^2(\mu)} g(x)\, d\mu(x)$$

et cette dernière quantité qui vaut $\int_{\mathbb{R}^d} e_{j_o}(x)\, g(x)\, d\mu(x)$ si $j_o \leq q(n)$, et 0 sinon, n'est autre que $\langle P_n(g) , e_{j_o}\rangle$

Q.E.D.

* $\sum_{t=1}^{n} [K^3_{t,n}]$ converge fortement vers $Id_{L^2(\mu)}$, car $q(n) \uparrow \infty$ cette propriété classique des projections découle immédiatement de l'égalité :

$$\text{si} \quad g \in L^2(\mu) \quad \|P_n(g)-g\|^2_{L^2(\mu)} = \Big\| \sum_{i=q(n)+1}^{\infty} \Big(\int_{\mathbb{R}^d} g \cdot e_i\, d\mu\Big) \cdot e_i \Big\|^2_{L^2(\mu)}$$

$$= \sum_{i \geq q(n)+1} \Big(\int_{\mathbb{R}^d} g\, e_i\, d\mu\Big)^2$$

alors que $\int g^2 \, d\mu = \sum_{i=1}^{\infty} \left(\int g \, e_i \, d\mu\right)^2 < \infty$.

* D'autre part, si $g \in L^2(\mu)$, on a :

$$||P_n(g)||^2_{L^2(\mu)} = ||\sum_{i=1}^{q(n)} \left(\int_{\mathbb{R}^d} g \, e_i \, d\mu\right) . e_i||^2_{L^2(\mu)} = \sum_{i=1}^{q(n)} \left(\int_{\mathbb{R}^d} g \, e_i \, d\mu\right)^2$$

$$\leq ||g||^2_{L^2(\mu)}$$

dès lors $||[K^3_{t,n}]|| = n^{-1} \, ||P_n|| = \mathcal{O}(1/n)$ pour tous t et n .

γ) Par le même raisonnement, nous obtenons, en posant $[K^4_{t,n}] = (n^{-1}) . P_t$:

$$[K^4_{t,n}](g) = \int_{\mathbb{R}^d} K^4_{t,n}(x) . g(x) \, d\mu(x)$$

pour tout élément g de $L^1(\mu) \cap L^2(\mu)$, c'est-à-dire que les $[K^4_{t,n}]$ vérifiant la "condition I".

La convergence des $\left(\sum_{t=1}^{n} [K^4_{t,n}]\right)$ vers $Id_{L^2(\mu)}$ découle de même immédiatement de celle de P_n vers cette même limite, puisque $[K^4_{t,n}] = (n^{-1}) . P_t)$. Comme enfin $||K^4_{t,n}|| = ||(n^{-1}) \, P_t|| = O(n^{-1})$, H4), H5), H6) sont donc vérifiées pour les $[K^4_{t,n}]$.

Pour H7), on remarque que, si $g \in L^2(\mu)$

$$||[K_{t,n}](g) - [K_{t+1,n}](g)||^2_{L^2(\mu)} = ||(n^{-1}) . \{P_t(g) - P_{t+1}(g)\}||^2_{L^2(\mu)}$$

$$= (n^{-2}) . \{\sum_{j=q(t)+1}^{q(t+1)} a_j^2\} \quad , \quad a_j = \int_{\mathbb{R}^d} g \, e_j \, d\mu \ldots$$

Comme chaque a_j^2 se majore par $||g||^2_{L^2(\mu)}$, il vient :

$$||[K_{t,n}](g) - [K_{t+1,n}](g)||^2_{L^2(\mu)} \leq (n^{-2})\,[q(t+1)-q(t)]\,.\,||g||^2_{L^2(\mu)} \quad ,$$

d'où

$$||[K_{t,n}] - [K_{t+1,n}]|| \leq (n^{-1})\,.\,\{q(t+1) - q(t)\}^{1/2}$$

Donc H7) est vérifiée si $\sup_n \{(n^{-1})\ [\sum_{t=1}^{n} t\,.\,\sqrt{q(t+1) - q(t)}]\} < \infty$.

λ) Du point δ) résulte que les $[K^3_{t,n}]$ vérifient H4) à H7), du point α) que H1) est vérifiée sous les conditions générales du théorème 4, H2 et H3 sous les conditions du b) α), pour les $K^3_{t,n}$, le théorème 1 permet donc de conclure en ce qui concerne $\hat{f}^3_n$.

Si l'on suppose maintenant que $1 - \frac{q(n)}{q(n+1)} < 1/n$, pour n assez grand, la suite $q(n)/n^2$ sera décroissante à partir d'un certain rang. Alors la condition $\sum_n (q(n)/n^2) < \infty$ impliquera $q(n)/n \longrightarrow 0$, si $n \to \infty$, et de plus $q(n+1)-q(n)$ se majorera par $\frac{q(n+1)}{n}$. Finalement la condition $\sum_{n\geq 1} (n^{-3/2})\,.\,(q(n)) < \infty$ assurera les 3 conditions du b).

μ) Pour $\hat{f}^4_n$, nous avons vu au point γ) que sous la condition (34) H4), H5), H6), H7) sont vérifiées et au point β) que $\lim_n (q(n)/n) = 0$, H1) est vérifiée, et si $\sum_n (q(n)/n^2) < \infty$, H2) et H3) le sont de surcroît. Ceci donne les points a) β) et b) β) du théorème, grâce au théorème 1

ω) Si $q(n) = [t(n)]$, $q(n+1) - q(n)$ ne vaudra 1 que si l'on a $t(n) < k \leq t(n+1)$ pour un certain entier k , et 0 sinon, en supposant que $t(n)$ croisse moins vite que n. La 2$^{\text{ème}}$ condition du b) α) s'écrira :

$$\sum_{k=1}^{\infty} \sqrt{k}/[(t^{-1}(k)] < \infty$$

Ainsi pour $t(x) = \mathrm{Log}(x)$, obtient-on instantanément les conditions du b) α), pour $t(x) = x^{\alpha}$, la $3^{\text{ème}}$ se ramène à :

$$\sum_{k\geq 1} (k^{1/2} \,.\, / \, k^{1/\alpha}) < \infty$$

ce qui impose $\alpha < 2/3$, les autres conditions sur $q(n)$ étant alors vérifiées.

Pour vérifier la condition (34) on note que :

$$\sum_{t=1}^{n} t \sqrt{q(t+1)-q(t)} = \sum_{i\in A_n} (i) \; , \qquad \text{où}$$

$$A_n = \{ i \;/\; i \leq n \;,\; \exists k \in \mathbb{N} \;,\; t(i) < k \leq t(i+1) \} \;\ldots$$

et $$\sup_n \{ (n^{-1}) \,.\, \sum_{t=1}^{n} t \sqrt{q(t+1)-q(t)} \} = \sup_k \{ (t^{-1}(k))^{-1} \,.\, (\sum_{j\leq k} t^{-1}(j)) \} \;\ldots$$

Si nous prenons $t(k) = \mathrm{Log}(k)$, la condition nécessaire à (34) est donc vérifiée

Le théorème est démontré.

C.III - 2ème PARTIE

ESTIMATION DES DERIVEES PARTIELLES

a) *Introduction*

α) Dans un certain nombre de problèmes, dont nous ne retiendrons ici, comme exemple, que celui de l'estimation non paramétrique de la dérive d'une diffusion (cf. Banon (15) ou Bhattacharya (16), entre autres), il est naturel de chercher à identifier, à partir de variables observées X_t , non seulement leur "densité limite" g (au sens des conditions B_i) mais aussi, sous réserve d'existence, les dérivées partielles de celles-ci, que nous noterons alors $D^i g$, $1 \leq i \leq d$ $(D^i g(x_1,\ldots,x_d) = \frac{\partial}{\partial x_i} g(x_1,\ldots,x_d))$.

Supposons que g et ses dérivées partielles soient éléments de $L^2(\lambda)$, (λ étant la mesure de Lebesgue sur $\mathbb{R}^d$) , noté aussi $L^2(\mathbb{R}^d)$, ou L^2 , par la suite, (étant entendu que le choix de λ n'est pas absolument nécessaire aux résultats qui suivent, mais en simplifie l'obtention). En choisissant une base (e_j) , $j \geq 1$, orthonormale, de $L^2(\mathbb{R}^d)$, nous pouvons écrire :

$$D^i g = \sum_{i=1}^{\infty} C_j^i \, e_j \qquad , \qquad C_j^i = \int_{\mathbb{R}^d} (D^i g) \, e_j \, . \, d\lambda \, .$$

La méthode d'estimation ici développée, et utilisée notamment par Bosq (17) et Greblicki-Pawlak (18) , s'appuie sur l'égalité suivante :

$$(34) \qquad C_j^i = - \int_{\mathbb{R}^d} (D^i \, e_j) \, . \, (g) \, d\lambda$$

vérifiée sous des conditions de régularité à détailler, dont l'existence des $(D^i \, e_j)$ que nous supposons désormais vérifiée, ainsi que leur appartenance à $L^2(\mu)$ et leur continuité.

Ceci nous ramène au problème évoqué précédemment : pour estimer $D^i g$, on considérera l'estimateur qui à une observation x associe une fonction $\sum_{j=1}^{q(n)} \hat{C}_j e_j$, projection sur M_n d'une densité "proche de δ_x" en posant

$$(35) \quad \begin{cases} \hat{f}_n^5 = \sum_{t=1}^{n} K_{t,n}^5(X_t) \\ \text{avec} \\ K_{t,n}^5(x) = -(n^{-1}) \cdot \sum_{j=1}^{q(n)} \{(D^i(e_j))(x)\} e_j \end{cases}$$

$K_{t,n}^5$ est donc une application mesurable bien déterminée à valeurs dans L^2 (même justification que pour la mesurabilité de $\hat{f}_n^3$) , nous allons démontrer sa convergence vers $D^i(g)$, dans L^2, dans les paragraphes qui suivent. Pour mémoire rappelons qu'il existe une autre méthode usuelle d'estimation des $(D^i g)$, développée dans le cas de l'échantillon par Walter (19), notamment.

β) Il est naturel de supposer, a priori que les g_{X_t} et $g_{X_t}^{\mathcal{F}_{t-1}}$ sont éléments de C^{12} , ensemble des fonctions continûment différentiables sur $\mathbb{R}^d$, et qui sont comme leurs dérivées, de carré intégrable. Comme C^{12} n'est pas complet, nous considérerons en fait que les g_{X_t} et $g_{X_t}^{\mathcal{F}_{t-1}}$ appartiennent à sa fermeture topologique, c'est-à-dire l'espace de Sobolev W^{12} formé des fonctions f de carré intégrable, admettant des dérivées partielles $\tilde{D}^i f$, $i = 1,\ldots,d$, au sens des distributions, de carrés intégrables également.

Comme $\tilde{D}^i g = D^i g$, si $g \in C^{12}$ (cf. Reinhardt (20)), l'hypothèse d'appartenance des g_{X_t} et $g_{X_t}^{\mathcal{F}_{t-1}}$ à C^{12} implique donc de toutes façons que le processus "vérifie la condition W^{12} ". De surcroît W^{12} est un espace de Hilbert séparable pour la norme (Adams p. 47 (7))

$$||g||^2_{W^{12}} = \int_{\mathbb{R}^d} g^2 \, d\lambda + \sum_{i=1}^{d} \int_{\mathbb{R}^d} (\tilde{D}^i g)^2 \, d\lambda \; .$$

γ) Nous avons supposé les e_j dans C^{12} pour définir $\hat{f}^5_n$. Cette hypothèse n'est, en fait, par très contraignante : les bases usuelles (f_i) de $\mathbb{R}$ (fonctions d'Hermite, etc ...) sont constituées de fonctions continûment différentiables de dérivées de carré intégrable (on peut au moins choisir de telles versions), et dans ces conditions, les bases-produit (e_i) sur $\mathbb{R}^d$:

$$e_i(x_1,\ldots,x_d) = \prod_{j=1}^{d} f_{i_j}(x_j)$$

sont bien éléments de C^{12}.

ω) Pour justifier l'idée de la méthode utilisée, l'égalité (34), il nous suffira finalement de démontrer le lemme suivant :

Lemme 5. Si f et g sont dans W^{12}, on a

$$\int_{\mathbb{R}^d} (\tilde{D}^i f) \, g \, d\lambda = - \int_{\mathbb{R}^d} (\tilde{D}^i g) \, f \, d\lambda \; .$$

Démonstration :

α) Nous utilisons deux résultats d'Adams.

* C_o^∞ ensemble des fonctions continûment différentiables, à support compact sur $\mathbb{R}^d$, est dense dans W^{12} (p. 56).

* Par définition de la dérivée au sens des distributions, si $g \in W^{12}$, pour tout élément Φ de C_o^∞, on a :

$$\int_{\mathbb{R}^d} (\tilde{D}^i g) \cdot \Phi \, d\lambda = - \int_{\mathbb{R}^d} (D^i \Phi) \, g \, d\lambda$$

β) Soit alors $g \in W^{12}$, il existe une suite d'éléments g_n de C_o^∞ tels que l'on ait :

$$(36) \qquad \int_{\mathbb{R}^d} (g-g_n)^2 \, d\lambda + \sum_{i=1}^{d} \int_{\mathbb{R}^d} (\tilde{D}^i g - D^i g_n)^2 \, d\lambda \xrightarrow[n \to \infty]{} 0 \ .$$

De plus on a pour tout n, si $f \in W^{12}$:

$$\int_{\mathbb{R}^d} (\tilde{D}^i f) \, g_n \, d\lambda = - \int_{\mathbb{R}^d} (D^i g_n) \, f \, d\lambda$$

en application directe de α).

γ) Finalement on aura, pour tout n

$$\int_{\mathbb{R}^d} \left[(\tilde{D}^i f) \, g + (\tilde{D}^i g) \, f\right] d\lambda = - \int_{\mathbb{R}^d} (\tilde{D}^i f) \, (g_n - g) \, d\lambda$$

$$+ \int_{\mathbb{R}^d} \left[(\tilde{D}^i f) \, g_n + (D^i g_n) \, f\right] d\lambda + \int_{\mathbb{R}^d} \left[- D^i g_n + \tilde{D}^i g\right] f \, d\lambda \ .$$

Le terme du milieu s'annulant, on en tire, d'après l'inégalité de Schwartz :

$$0 \leq \overline{\operatorname{Lim}_n} \left| \int_{\mathbb{R}^d} \left[(\tilde{D}^i f) \, g + (\tilde{D}^i g) f\right] d\lambda \right| \leq ||\tilde{D}^i f||_{L^2} \cdot \operatorname{Lim}_n (||g - g_n||_{L^2})$$

$$+ ||f||_{L^2} \cdot \operatorname{Lim}_n (||\tilde{D}^i g - D^i g_n||_{L^2})$$

ce qui achève la démonstration : le terme de droite est nul d'après (36).

b) Théorème de convergence

Théorème 6.- Soit (e_i) , $i \geq 1$, une base orthonormale de $L^2(\mathbb{R}^d)$ constituée d'éléments de C^{12} , uniformément bornée, $q(n)$ une suite d'entiers croissant vers l'infini, $M(n)$ défini par

$$M(n) = \sum_{j=1}^{q(n)} \{ \sup_{x \in \mathbb{R}^d} (D^i(e_j)(x))^2 \} .$$

Supposons que le processus (X_t) , $t \in \mathbb{Z}$ vérifie la condition W^{12} , et qu'il existe un élément g de C^{12} tel que $\lim\limits_{n\to\infty} (\|(n^{-1}) . (\sum_{t=1}^{n} g_{X_t}) - g\|_{W^{12}}) = 0$.

a) Si $\lim\limits_{n\to\infty} E(\|(n^{-1}) . \sum_{t=1}^{n} (g_{X_t}^{F_{t-1}} - g_{X_t})\|_{W^{12}}) = 0$, et que $\lim\limits_{n} [M(n)/n] = 0$, on a :

$$\lim_{n\to\infty} E(\|\hat{f}_n^5 - D^i g\|_{L^2}) = 0 .$$

b) Si $\lim\limits_{n\to\infty}$ p.s. $(\|(n^{-1}) . \sum_{t=1}^{n} (g_{X_t}^{F_{t-1}} - g_{X_t})\|_{W^{12}}) = 0$, on aura : $\lim\limits_{n\to\infty}$ p.s. $(\|\hat{f}_n^5 - D^i g\|_{L^2}) = 0$, pourvu que $\lim\limits_{n} (M(n)/n) = 0$, $\sum_{n} [(M(n)/n^2] < \infty$, et $\sum_{n\geq 1} (n^{-1}) \sqrt{M(n)\,[M(n+1) - M(n)]} < \infty$.

c) Si nous supposons que $M(n) = O[q(n)^\beta]$ et que $q(n) = [t(n)]$, t étant positive, continue strictement croissante sur $\mathbb{R}^+$, les conditions de a) et b) sont vérifiées si

$$\sum_{1}^{\infty} \{[t^{-1}(k)]^{\beta/2} / t^{-1}(k)\} < \infty ,$$

d'où le choix possible de $q(n)$ en $[n^\alpha]$, ou $q(n) = [\mathrm{Log}\, x]$, si (e_i) est une base formée des produits de d fonctions d'Hermite, par exemple.

Démonstration :

α) Un calcul analogue à celui qui a permis la détermination de $\alpha^3_{t,n}$ et $\gamma^3_{t,n}$, basé sur l'orthonormalité des (e_i) donne directement, de par la définition de $M(n)$

$$(\alpha^5_{t,n}) = \sup_{x\in\mathbb{R}^d} ||K^5_{t,n}(x)||_{L^2} = O((n^{-1}) \cdot \sqrt{M(n)})$$

$$(\gamma^5_{t,n}) = \sup_{x\in\mathbb{R}^d} ||K^5_{t,n}(x) - K^5_{t+1,n}(x)|| = O(\frac{\sqrt{M(n)}}{n^2} + \frac{\sqrt{M(n+1)-M(n)}}{(n+1)})$$

ce qui montre que H1) est vérifiée si $\lim\limits_{n\to\infty} [M(n)/n] = 0$, H2, si $\sum\limits_n \frac{M(n)}{n^2} < \infty$, H3) si de surcroît $\sum\limits_1^\infty (n^{-1}) (\sqrt{[M(n)][M(n+1)-M(n)]}) < \infty$

β) Posons alors $P'_n = P_n \circ \tilde{D}^i$, composition de l'opérateur "dérivation partielle (faible) d'ordre i", défini de W^{12} dans L^2, et de la projection, de L^2 dans L^2, sur le sous-espace engendré par $e_1,\dots,e_{q(n)}$. On a alors

* Les $[K^5_{t,n}]$, égaux à $(n^{-1}) P'_n$, vérifient la condition I. Comme pour les $[K^3_{t,n}]$, ceci reviendra finalement à montrer que :

$$\forall j_o,\ j_o \in \mathbb{N},\quad \forall g,\ g \in L^1 \cap W^{12},\ \text{on a :}$$

$$\langle e_{j_o}, P^n(\tilde{D}^i(g))\rangle_{L^2} = - \int_{\mathbb{R}^d} \langle e_{j_o} (\sum_{j=1}^{q(n)} [(D^i e_j)(x)] \cdot e_j\rangle_{L^2} g(x)\, dx$$

et on vérifie directement que les 2 quantités valent

$$- \int_{\mathbb{R}^d} (g(x)) \cdot (D^i e_{j_o}(x))\, dx \quad \text{si } j_o \leq q(n), \quad 0 \text{ sinon}.$$

* $\sum\limits_{t=1}^n [K^5_{t,n}]$ qui vaut P'_n converge fortement vers $\tilde{D}^i$ puisque P_n converge vers Id_{L^2}, et que $q(n) \uparrow \infty$

* $||[K^4_{t,n}]|| = (n^{-1})\ ||P_n \cdot (\tilde{D}^i)|| = 0(n^{-1})$.

En effet P_n est contractante, nous l'avons vu, et $\tilde{D}^i$ de même, puisqu'on a toujours $||\tilde{D}^i(g)||_{L^2} \leq ||g||_{W^{12}}$, par définition de cette dernière norme.

Ainsi les $[K^5_{t,n}]$ "vérifient la condition I", et les conditions H4) à H7), d'après ce qui précède (pour H7), il suffit de voir que $[K^5_{t,n}] = [K^5_{t+1,n}]$, $t = 1,\dots,n-1$) .

γ) Nous avons vérifié toutes les hypothèses nécessaire à l'obtention du théorème 1. Il s'applique donc ici encore, ce qui démontre le théorème 5.

Supposons maintenant que $e_i(x_1,\dots,x_d) = \prod_{j=1}^{d} f_{i_j}(x_j)$, les f_{i_j} étant choisies dans la base des fonctions d'Hermite sur $\mathbb{R}$. Les f_{i_j} étant uniformément bornées, on déduit de la formule :

$$f'_k = \sqrt{\frac{k}{2}}\, f_{k-1} - \sqrt{\frac{k+1}{2}}\, f_{k+1}$$

(cf. Szego (21)) que

$$\sup_{(x)} [D^k(e_i(x)))] = 0(\sqrt{i_k})$$

En choisissant une numérotation des e_i telle que

$$(e_i = \prod_{j=1}^{d} f_{i_j}) \Longrightarrow i \geq \{(\sup_{j=1,\dots,d} i_j)\}^{1/2}$$

On aura donc : $\sup_{(x)} [D^k(e_i(x))] = 0\ (i^{1/4})$, et cette majoration se révèle évidemment <u>très</u> large.

Avec le même principe, les fonctions trigonométriques, dans le cas où le processus prend ses valeurs dans $[-\pi,\pi]^d$, fourniraient une base uniformément bornée telle que $\sup_{(x)} [D^k(e_i(x)] = 0(\sqrt{i})$ ... D'où finalement la majoration "naturelle" de $M(n)$ par $O[(q(n))^{\beta}]$ pour un $\beta > 0$. ($= 5/4$, pour les produits de fonctions d'Hermite avec la minoration utilisée ci-dessus sur les indices, mais on peut évidemment obtenir mieux !).

Dès lors, si $q(n) = [t(n)]$, $M(n+1) - M(n)$ n'étant non nulle que lorsque $q(n)$ "saute" un entier, comme nous l'avons déjà vu, la convergence de la série utilisée en H3) se ramène à la condition du théorème

$$\sum_{1}^{\infty} \{[t^{-1}(k)]^{\beta/2} / t^{-1}(k)\} < \infty$$

soit simplement : $\sum_{1}^{\infty}\{n^{\beta/2\alpha - 1/\alpha}\} < \infty$ si $t(x) = x^{\alpha}$, la condition étant vérifiée dès que $\beta < 2$, donc pour les fonctions d'Hermite, et a fortiori si $t(k) = \text{Log } k$ ...

Pour de tels choix de $q(n)$ H1) et H2) étant vérifiées, la démonstration est achevée.

ESTIMATION PAR LA METHODE DU NOYAU

(Chapitres IV et V)

Les résultats qui suivent concernent l'étude de l'estimateur d'une densité limite g, (au sens des conditions Bi), défini par $\hat{f}_n^6 = \sum_{t=1}^{n} K_{t,n}^6(X_t)$, où $K_{t,n}^6(x)$ est la fonction qui à y de $\mathbb{R}^d$ associe

(36) $$(n^{-1}) \cdot \left[h(n)^{-d}\right] \cdot K((h(n))^{-1}(y-x))$$

$h(n)$ représentant une suite de réels décroissant vers 0, K une densité de probabilité sur $\mathbb{R}^d$ (par rapport à une mesure μ σ-finie régulière). Nous supposerons de surcroît toujours K bornée, donc élément de $L^2(\mu)$.

Nous dirons alors que K est un "noyau", reprenant la terminologie classique de la méthode introduite par Parzen et Rosenblatt ((22),(23)). Du point de vue fonctionnel, cet "estimateur de la méthode du noyau ", $\hat{f}_n^6$, s'avère d'ailleurs très naturel : $(\sum_1^n K_{t,n}^6)$ associe à une observation x de $\mathbb{R}^d$ la densité d'une loi "pratiquement concentrée" en x, très proche de δ_x (qui au sens des distributions est limite des distributions régulières définies par les fonctions $\left[h(n^{-1}]\right)^{+d} \cdot K((h(n))^{-1}(\cdot-x))$, si $h(n) \downarrow 0 \ldots$).

Comme dans le cas de l'estimation par projection, on peut améliorer $\hat{f}_n^6$, de façon à diminuer le temps de calcul inhérent à sa détermination : à l'introduction d'une nouvelle observation X_{n+1}, l'obtention de $\hat{f}_{n+1}^6(y)$ exige le calcul de $(n+1)$ valeurs $K\left[(y-X_t)/h(n+1)\right]$, alors qu'en posant :

(37) $$\hat{f}_n^7 = \sum_{t=1}^{n} K_{t,n}^7(X_t) \text{ , avec}$$

$$K_{t,n}^7(x) = \beta_{t,n} \cdot \left[h(t)\right]^{-d} \cdot K\left[(h(t))^{-1} \cdot (\cdot-x)\right]$$

on réduira à 1 le nombre de ces opérations.

Dans le cas où l'on prend $\beta_{t,n} = (n^{-1})$, $1 \leq t \leq n$, on a exactement la relation de récurrence :

$$n\, \hat{f}_n^7(y) = (n-1)\, \hat{f}_{n-1}^7(y) + \left[h(n)\right]^{-d} K\left[(h(n))^{-1}(y-X_n)\right]$$

L'autre solution usuelle est de choisir $\beta_{t,n} = (h(t))^d \cdot [\sum_{t=1}^{n} (h(t))^d]^{-1}$,
et nous adapterons toujours à ces deux exemples les résultats relatifs à $\hat{f}_n^7$, les $\beta_{t,n}$ étant, pour des raisons qui apparaîtront clairement, soumis aux conditions minimales.

(38)
$$\forall n \geq 1 \ , \ \forall t \ , \ 1 \leq t \leq n \ , \ \beta_{t,n} \geq 0 \ ,$$
$$\lim_{n\to\infty} \left(\sum_{t=1}^{n} \beta_{t,n} = 1 \right) \ , \ \lim_{n\to\infty} \beta_{t,n} = 0 \quad (t \text{ fixé})$$

(On pourra consulter Deheuvels (24) pour obtenir une formulation un peu plus générale d'un estimateur récursif).

Techniquement notre travail sera donné en 4 points.

La première partie du C.IV, technique, consistera à obtenir quelques lemmes indispensables aux points suivants, la seconde partie de ce chapitre concernant alors la convergence vers g de $\hat{f}_n^6$ et $\hat{f}_n^7$ dans $L^2(\mu)$, lorsque le processus (X_t) , $t \in \mathbb{Z}$, "vérifie la condition $L^2(\mu)$" .

Nous attaquerons ensuite le problème de la convergence de ces estimateurs sous les seules hypothèses d'ergodicité des (X_t) et d'existence de $g_{X_t}^{\mathcal{F}_{t-1}}$: dans ce cas (cf. corollaire 1) le processus vérifie automatiquement la "condition $L^1(\mu)$" et nous démontrerons la convergence de $\hat{f}_n^6$ et $\hat{f}_n^7$ vers g dans $L^1(\mu)$ également. Evidemment la démonstration apportée, originale à notre connaissance, se démarque des techniques jusque là utilisées : les estimateurs ne sont plus à valeurs Hilbert ! Ceci fait l'objet de la 3ème partie du C. IV.

Enfin, le 4e et dernier point abordé consiste en une brève démonstration des convergences des dérivées de $\hat{f}_n^6$ et $\hat{f}_n^7$ vers celles de g , dans $L^2(\mu)$ puis $L^1(\mu)$: c'est l'objet du C.V. .

C.IV - 1ère PARTIE

QUELQUES LEMMES TECHNIQUES

Lemme 6.- Soit $(\Omega, \mathcal{A}, \mu)$ un espace mesuré, f une application intégrable de Ω dans $\bar{\mathbb{R}}$, f_n une suite d'applications intégrables de Ω dans $\bar{\mathbb{R}}$. Si $f_n \to f$, presque partout et que $\int_\Omega |f_n| \, d\mu \longrightarrow \int_\Omega |f| \, d\mu$, lorsque $n \to \infty$, on a :

$$\lim_{n\to\infty} \int_\Omega |f_n - f| \, d\mu = 0 \quad .$$

Démonstration : Classique, elle se trouve par exemple dans Parthasarathy (5) p. 206.

Lemme 7.- Soit μ une mesure σ-finie régulière sur $(\mathbb{R}^d, \mathcal{B}_{\mathbb{R}^d})$, β un réel positif, g une variable aléatoire réelle définie sur $(\mathbb{R}^d, \mathcal{B}_{\mathbb{R}^d})$. On définit alors $[H_\beta](g)$ ponctuellement en posant :

$$\{[H_\beta](g)\}(y) = \int_{\mathbb{R}^d} \frac{1}{\beta^d} \cdot K\left(\frac{y-x}{\beta}\right) g(x) \, d\mu(x)$$

K étant un noyau au sens défini dans l'introduction. Alors :

a) $[H_\beta]$ est un opérateur linéaire contractant bien défini sur $L^1(\mu)$ à valeurs dans $L^1(\mu)$, tel que $[H_\beta]$ converge fortement vers $Id_{L^1(\mu)}$ lorsque $\beta \downarrow 0$, et que $[H_\beta](y)$ soit égal à l'intégrale de Bochner $\int_{\mathbb{R}^d} \frac{1}{\beta^d} K\left(\frac{\cdot - x}{\beta}\right) g(x) \, d\mu(x)$.

b) Les conclusions sont identiques si l'on considère $[H_\beta]$ défini sur $L^2(\mu)$, à valeurs dans $L^2(\mu)$.

Démonstration :

α) Soit $g \in L^1(\mu)$. Pour tout $y, [H_\beta](g)(y)$ est alors défini, puisque K est borné. Pour obtenir les autres propriétés recherchées, considérons une fonction A réelle bornée sur $\mathbb{R}^d$, et $\mathcal{B}_{\mathbb{R}^d} - \mathcal{B}_{\mathbb{R}}$ mesurable, et la fonction ℓ définie par

$$\ell_A(x,y) = A(y) \cdot \beta^{-d} \cdot K[(y-x)/\beta] \cdot g(x)$$

g, K et A étant mesurables, ℓ le sera par rapport au couple (x,y), et de surcroît on a :

$$\iint |\ell_A(x,y)|\, d\mu(x)\, d\mu(y) \leq \|A\|_{L^\infty} \cdot \int_{\mathbb{R}^d} |g(x)|\, d\mu(x) \,.$$
$$\left\{\int_{\mathbb{R}^d} \beta^{-d} \cdot K[(y-x)/\beta]\, d\mu(y)\right\}$$

$$\leq \|A\|_{L^\infty} \cdot \|g\|_{L^1(\mu)} \qquad \text{(K est une densité).}$$

On en conclut alors successivement :

* $[H_\beta](g)$ est un élément de $L^1(\mu)$: la mesurabilité provient de celle de ℓ_1 et du théorème de Fubini, l'intégrabilité de ce que :

$$\int_{\mathbb{R}^d} |[H_\beta](g)|\, d\mu \leq \iint |\ell_1|\, d\mu^{\otimes 2}$$

(en appelant ℓ_1 la fonction ℓ_A où $A \equiv 1$) .

* $\|[H_\beta]\| \leq 1$: en prenant encore $A \equiv 1$ ou a bien, d'après ce qui précède

$$\|[H_\beta]\|_{L^1(\mu)} \leq \|g\|_{L^1(\mu)}$$

* $[H_\beta](g) = \int_{\mathbb{R}^d} \beta^{-d} \cdot K[(\cdot - x)/\beta]\, g(x)\, d\mu(x)$,

au sens de l'intégrale de Bochner : le dual de $L^1(\mu)$ étant $L^\infty(\mu)$, on déduit de l'intégrabilité de ℓ_A que, pour toute forme linéaire u sur $L^1(\mu)$, grâce au théorème de Fubini :

$$<u\ ,\ [H_\beta](g)> = \int_{\mathbb{R}^d} <u\ ,\ \beta^{-d}\ .\ K[(\cdot - x)]\ .\ g(x)>\, d\mu$$

Ceci démontre, comme d'habitude, la propriété recherchée.

β) Reste à démontrer que si $\beta \downarrow 0$ $[H_\beta]$ converge fortement vers $Id_{L^1(\mu)}$. Or, si $g \in L^1(\mu)$, on peut écrire, moyennant un simple changement de variable

$$\begin{aligned} \|[H_\beta](g)-g\|_{L^1(\mu)} &= \int_{\mathbb{R}^d} \left|\int_{\mathbb{R}^d} K(u)\ g(y-\beta u)\ d\mu(u) - g(y)\right| d\mu(y) \\ &\leq \iint_{\mathbb{R}^d\times\mathbb{R}^d} |h_\beta(u,y) - h(u,y)|\ d\mu^{\otimes 2}(u,y) \end{aligned} \tag{39}$$

où $h(u,y) = K(u)\ g(y)$ et $h_\beta(u,y) = K(u)\ g(y-\beta u)$, puisque $\int K(u)\ d\mu(u) = 1$. Alors

* $\forall\beta$, $\beta > 0$, on a :

$$\iint_{\mathbb{R}^d\times\mathbb{R}^d} |h_\beta(u,y)|\,d\mu^{\otimes 2}(u,y) = \iint_{\mathbb{R}^d\times\mathbb{R}^d} |h(u,y)|\ d\mu^{\otimes 2}(u,y)$$

* Si g est continue, pour tout (x,y) : $\lim\limits_{\beta\downarrow 0} h_\beta(x,y) = h(x,y)$.

Cette propriété, jointe à la précédente, permet de conclure, grâce au lemme 6 que $\lim\limits_{\beta\downarrow 0} \|[H_\beta](g) - g\|_{L^1(\mu)} = 0$, d'après (39).

* L'ensemble des fonctions continues étant dense dans $L^1(\mu)$, les $\|H_\beta\|$ étant toutes bornées par 1 (cf. α)) le théorème de Banach - Steinhaus permet de conclure à la convergence forte, sur $L^1(\mu)$ de H_β, vers $Id_{L^1(\mu)}$, lorsque $\beta \downarrow 0$.

Le a) est totalement vérifiée.

γ) Si $g \in L^2(\mu)$, l'existence de $\{[H_\beta](g)\}(y)$ se justifie par l'appartenance de K et g à ce même espace. De plus

* Pour tout y on a : $[H_\beta](g)\ (y) = \int_{\mathbb{R}^d} K(u)\ g(y-\beta u)\ d\mu(u)$.

Alors l'appartenance de $[H_\beta](g)$ à $L^2(\mu)$ résulte de l'intégrabilité par rapport à $\mu^{\otimes 3}$ de la fonction h définie par :

$$h(u,u',y) = K(u)\ K(u)\ g(y-\beta u)\ .\ g(y-\beta u')\ .$$

Celle-ci découle de ce que :

$$\iiint |h|\ d\mu^{\otimes 3} \leqslant \iint K(u)\ K(u')\ .\ \{\int |g(y-\beta u).g(y-\beta u')|\ d\mu(y)\}\ d\mu(u)\ d\mu(u')$$

$$\leqslant ||g||^2_{L^2(\mu)}\ .\ \iint K(u)\ K(u')\ d\mu(u)\ d\mu(u') \leqslant ||g||^2_{L^2(\mu)}.||K||^2_{L^1(\mu)}$$

$$< \infty$$

De :

$$\int \{[H_\beta](g)\}^2\ d\mu = \int_{\mathbb{R}^d} \{\int_{\mathbb{R}^d} K(u)\ g(y-\ u)\ d\mu(u)\} \times$$

$$\times \{\int_{\mathbb{R}^d} K(u')\ g(y-\ u')\ d\mu(u')\}\ d\mu(y)$$

on déduit alors la propriété cherchée.

* Les inégalités précédentes permettent de plus d'écrire :

$$|||H_\beta|(g)||^2_{L^2(\mu)} \leqslant \iiint |h|\ d\mu^{\otimes 3} \leqslant ||g||^2_{L^2(\mu)}\ .\ (||K||^2_{L^1(\mu)})$$

Ce qui donne : $||H_\beta|| \leqslant 1$.

* En considérant enfin A élément de $L^2(\mu)$, on obtient ici

$$\iint_{\mathbb{R}^d\times\mathbb{R}^d} |\ell_A(x,y)|\ d\mu(x)d\mu(y) \leqslant \int_{\mathbb{R}^d} |g(x)|d\mu(x).$$

$$\{\int_{\mathbb{R}^d} |A(y).\beta^{-d}.K[(y-x)/\beta]|d\mu(y)\}$$

$$\leqslant ||g||_{L^1(\mu)} . ||A||_{L^2(\mu)} . \beta^{-d/2}.||K||_{L^2(\mu)} < \infty,$$

ce qui justifie l'intégrabilité de ℓ_A et a fortiori, l'égalité :

$$<u, [H_\beta](g)> = \int_{\mathbb{R}^d} <u, \beta^{-d} K[(\cdot - x)/\beta]\, g(x)> d\mu(x)$$

pour toute forme linéaire du dual de $L^2(\mu)$, qui est réflexif.

On en conclut, comme de coutume, que

$$[H_\beta](g) = \int_{\mathbb{R}^d} \beta^{-d} . K[(\cdot-x)/\beta]g(x)\ d\mu(x) \quad .$$

On a donc toutes les propriétés cherchées pour $[H_\beta]$, en tant qu'opérateur de $L^2(\mu)$ dans $L^2(\mu)$. Reste à démontrer la convergence forte

ω) Soit alors $g \in L^2(\mu)$. On peut écrire, puisque $\int_{\mathbb{R}^d} K(u)\ d\mu(u) = 1$,

$$||[H_\beta](g)-g||^2_{L^2(\mu)} = \int_{\mathbb{R}^d}\{\int_{\mathbb{R}^d} K(u)\ [g(y-\beta u)-g(y)]\ d\mu(u)\}^2\ d\mu(y)$$

$$= \iint_{\mathbb{R}^d\times\mathbb{R}^d} K(u)\ K(u')\ \{\int_{\mathbb{R}^d}[g(y-u\beta)-g(y)]\,[g(y-u'\beta)-g(y)]$$

$$d\mu(y)\}\ d\mu(u)\ d\mu(u')$$

(égalité justifiée par les calculs ayant montré l'intégrabilité de h , au point précédent).

Alors * Si g est continue et bornée, on a , pour tous u et u' :

$$\lim_{\beta \downarrow 0} \int_{\mathbb{R}^d} [g(y-u\beta) - g(y)] [g(y-u'\beta) - g(y)] \, d\mu(y) = 0 \quad .$$

En effet, l'inégalité de Schwartz et le fait que g soit bornée ramènent cette conclusion à :

$$\lim_{\beta \downarrow 0} \int_{\mathbb{R}^d} |g(y-u\beta) - g(y)| \, d\mu(y) = 0 \quad .$$

Or ceci découle directement du lemme 6 parce que g est continue et que $\int_{\mathbb{R}^d} |g(y-u\beta)| \, d\mu(y) = \int_{\mathbb{R}^d} |g(y) \, d\mu(y)|$ pour tout β.

* On en déduit que si g est continue et bornée, par simple application du théorème de Lebesgue, d'après (40) :

$$\lim_{\beta \downarrow 0} \| [H_\beta](g) - g \|^2_{L^2(\mu)} = 0$$

puisque les fonctions utilisées en (40) se majorent toutes par $2K(u) \, K'(u) \, . \, \|g\|^2_{L^2(\mu)}$ intégrable en (u,u') ...

* Comme l'ensemble des fonctions continues à support compact est dense dans $L^2(\mu)$, le théorème de Banach - Steinhaus permet d'affirmer que $\forall g$, $g \in L^2(\mu)$:

$$\| [H_\beta](g) - g \|_{L^2(\mu)} \longrightarrow 0 \quad , \quad \text{si} \quad \beta \downarrow 0$$

puisque pour tout β , $\|H_\beta\| \leq 1$ .... Q.E.D.

Lemme 8.- Soit K une densité de probabilité sur $\mathbb{R}^d$, par rapport à une mesure μ σ-finie régulière, on suppose K bornée, et continûment différentiable

a) Si les $(D^i K)$ sont bornées sur $\mathbb{R}^d$ et que $||u||_{\mathbb{R}^d} \cdot K(u)$ est μ intégrable, on a :

(41) $$\int_{\mathbb{R}^d} K(u)\, K(cu)\, d\mu(u) = ||K||^2_{L^2(\mu)} + 0|c-1| \quad , \quad c \in \mathbb{R} .$$

b) Si les $D^i(K)$ vérifient, pour tout $j \in \{1,\dots,d\}$

$$\int_{\mathbb{R}^d} |uj|\, \left|\frac{\partial K}{\partial u_j} (u_1,\dots,u_d)\right| du_1 \dots du_d < \infty \ , \quad \text{alors on a :}$$

(42) $$\int \left|\frac{K(cu)}{K(u)} - 1\right| K(u)\, d\mu(u) \leq \left|\left(\frac{1}{c}\right)^d - 1\right| \ , \quad c \in [1,\infty[.$$

c) Si $\mu = \mu_o^{\otimes d}$ et que K est le produit de d densités bornées sur $\mathbb{R}$: $K(x_1,\dots,x_d) = \prod_{i=1}^{d} K_o(x_i)$, il suffira donc que K_o soit dérivable et tel que $\int_{\mathbb{R}^d} |x|\, K_o(x)\, dx < \infty$, $\int_{\mathbb{R}^d} |x|\, K_o'(x)\, dx < \infty$, pour vérifier (41) et (42) ... Pour $\mu_o = \lambda_{\mathbb{R}}$, le noyau gaussien est un exemple typique.

Démonstration :

a) Le résultat découle directement de l'égalité :

$$\int_{\mathbb{R}^d} K(u) K(cu) = d\mu(u) = ||K||^2_{L^2(\mu)} = \int_{\mathbb{R}^d} K(u) \ \left\{\sum_{j=1}^{d} (c-1)(u_j) .\ D^j K[u + \theta_j (c-1) u_j]\right\} d\mu(u)$$

obtenue d'après la formule de Taylor .

b) Posons $A(c) = \int_{\mathbb{R}^d} |K(cu) - K(u)| \, d\mu(u)$ Alors :

$$A(c) \leq \sum_{j=1}^{d} \int_{\mathbb{R}^d} |K(u_1,\dots,u_{j-1},u_j,cu_{j+1},\dots,cu_d) - K(u_1,\dots,u_{j-1},cu_j,cu_{j+1},\dots,cu_d)| \, du_1\dots du_d$$

$$\leq \sum_{j=1}^{d} \int_{\mathbb{R}^{d+1}} \left|\frac{\partial K}{\partial u_j}(u_1,\dots,u_{j-1},x,cu_{j+1},\dots,cu_d)\right| I_{A_c^{u_j}}(x) \, du_1\dots du_d \, dx$$

avec $A_c^{u_j} = [u_j, cu_j]$ si $u_j > 0$, $[cu_j, u_j]$ sinon. Donc :

$$A(c) \leq \sum_{j=1}^{d} \int_{\mathbb{R}^d} du_1\dots du_{j-1}.dx.du_{j+1}\dots du_d \left\{\left|\frac{\partial K}{\partial u_j}(u_1,\dots,u_{j-1},x,cu_{j+1},\dots,cu_d)\right| . |x| . \left|\frac{1}{c} - 1\right|\right\}$$

d'où :

$$A(c) \leq \sum_{j=1}^{d} \left|1 - \frac{1}{c}\right| \int_{\mathbb{R}^d} |x| \left|\frac{\partial K}{\partial u_j}(u_1,\dots,u_{j-1},x,cu_{j+1},\dots,cu_d)\right| du_1\dots du_{j-1} \, dx \, du_{j+1}\dots du_d$$

$$\leq \left|1 - \frac{1}{c}\right| \left\{\sum_{j=1}^{d} \frac{1}{c^{d-j}}\right\} \left\{\sup_{1 \leq j \leq d} \int_{\mathbb{R}^d} |u_j| \left|\frac{\partial K}{\partial u_j}(u_1,\dots,u_d)\right| du_1\dots du_d\right\}$$

$$= O\left[-\frac{1}{c^d} + 1\right]$$ sous l'hypothèse du lemme. Q.E.D.

c) Le résultat annoncé résulte directement de l'égalité

$$\{D^k(K)\}(u_1,\dots,u_d) = K_o(u_1)\dots K_o(u_{k-1}) \, K_o'(u_k) \, K_o(u_{k+1}) \dots K_o(u_d)$$

Lemme 9.- Soit (X_t) , $t \in \mathbb{Z}$, un processus strictement stationnaire à valeurs dans $\mathbb{R}^d$, tel que l'espérance des $||X_t||_{\mathbb{R}^d}$ existe.

Soit $\alpha(n)$ une suite de réels croissant vers l'infini et $(\beta_{t,n})$, $n \geq 1$, $1 \leq t \leq n$, une suite de séquences de réels strictement positifs tels que $\lim\limits_n (\sum\limits_1^n \beta_{t,n}) = 1$ et $\lim\limits_n \beta_{t,n} = 0$, pour chaque t

a) Posons $N_n = \sum\limits_{t=1}^{n} \beta_{t,n} \cdot I_{\{||X_t||>\alpha(n)\}}$, alors N_n tend vers 0 en moyenne. Sous les conditions :

(43) $\sum\limits_{n\geq 1} \beta_{n,n} [\alpha(n)]^{-1} < \infty$ et $\sum\limits_{n\geq 1} \{(\alpha(n))^{-1} \cdot (\sum\limits_{t=1}^{n} |\beta_{t,n}-\beta_{t,n+1}|)\} < \infty$,

N_n tend vers 0 presque sûrement. C'est donc le cas si $\beta_{t,n} = (n^{-1})$, $1 \leq t \leq n$, pourvu que $\sum\limits_1^{\infty} (n^{-1}) (\alpha(n))^{-1} < \infty$.

b) Dans ce dernier cas, si la suite $(\alpha(n))$, $n \geq 1$, vérifie : $\sum\limits_1^{\infty} (\frac{1}{\alpha(n)}) < \infty$, la suite M_n définie par

$$(M_n) = (n^{-1}) (\alpha(n))^{-1} \cdot \sum_{t=1}^{n} ||X_t|| \cdot I_{\{||X_t||>\frac{(\alpha(n))^2}{2}\}}$$

converge aussi vers 0 , en moyenne et presque sûrement.

Démonstration :

α) On a $E(N_n) = \sum\limits_{t=1}^{n} \beta_{t,n} \cdot P(||X_t|| > \alpha(n))$

$= O(\sum\limits_{t=1}^{n} \beta_{t,n} \cdot (\alpha(n))^{-1})$ et le lemme 3 p. 10 permet de conclure.

D'après le lemme 1, si F_n est la tribu engendrée par $\{X_i \ , \ i \leq n\}$, pour obtenir la convergence presque sûre, il suffit de montrer que :

$$\sum_1^\infty E \ |E^{F_n}(N_{n+1}) - N_n| < \infty$$

soit a fortiori la convergence des deux séries :

$$\left\{\begin{array}{l} \sum_1^\infty E \ |\beta_{n+1,n+1} \ . \ E^{F_n}(I_{\{||X_{n+1}||>\alpha(n+1)\}})| \\ \qquad\qquad \text{et} \\ \sum_1^\infty E \ | \sum_{t=1}^n (\beta_{t,n} \ . \ I_{\{||X_t||>\alpha(n)\}} - \beta_{t,n+1} \ . \ I_{\{||X_t||>\alpha(n+1)\}})| \end{array}\right.$$

La convergence de la première est obtenue si $(\sum_1^\infty \beta_{n,n} \ . \ [\alpha(n)]^{-1} < \infty)$.

Pour la seconde, on obtient comme condition suffisante, les convergences des séries

$$\left\{\begin{array}{l} \sum_1^\infty (\sum_1^n |\beta_{t,n} - \beta_{t,n+1}| E(|I_{\{||X_t||>\alpha(n+1)\}}) = O\{\sum_1^\infty (\sum_1^n |\beta_{t,n} - \beta_{t,n+1}| (\alpha(n+1)^{-1}\} \\ \qquad\qquad \text{et} \\ \sum_1^\infty (\sum_1^n \beta_{t,n} \ . \ P(\alpha(n) \leq ||X_t|| < \alpha(n+1))) = O\{\sum_{n=1}^\infty P[(\alpha(n) \leq ||X_t|| \leq \alpha(n+1)]\} < \infty \end{array}\right.$$

Ceci donne bien les conditions du lemme (a)). Elles sont vérifiées quand $\beta_{t,n} = (n^{-1})$, $1 \leq t \leq n$, si $\sum_{n=1}^\infty (n^{-1}) \ (\alpha(n))^{-1} < \infty$.

β) Nous utiliserons la même technique pour $M(n)$. On note d'abord que $E(M(n)) = O(\alpha(n)^{-1})$, car les (X_t) ont même loi et que $E(||X_t||)$ existe. Le lemme 1 ramène l'obtention de la convergence p.s. à la convergence des 2 séries, comme ci-dessus :

$$\sum_{n=1}^\infty E\{ (n+1)^{-1} \ .(\alpha(n+1))^{-1} \ . \ ||X_{n+1}|| \ . \ I_{\{||X_{n+1}||> \frac{[\alpha(n+1)]^2}{2}\}} \}$$

et

$$\sum_{n=1}^{\infty} E\{|(n+1)^{-1}(\alpha(n+1))^{-1} \cdot \sum_{t=1}^{n} ||X_t|| \cdot I_{\{||X_t||>\frac{[\alpha(n+1)]^2}{2}\}} - n^{-1}.(\alpha(n))^{-1}$$

$$\sum_{t=1}^{n} ||X_t|| \cdot I_{\{||X_t||>\frac{(\alpha(n))^2}{2}\}}|\}$$

Pour la première, on est ramené à obtenir la condition $\sum_1^{\infty} (n+1)^{-1}(\alpha(n+1))^{-1} < \infty$.
Pour la seconde, on doit obtenir la convergence de

$$\sum_{n=1}^{\infty} \left[\frac{1}{n\alpha(n)} - \frac{1}{(n+1)\alpha(n+1)}\right] \cdot \{\sum_{t=1}^{n} E[||X_t|| \cdot I_{\{||X_t||>\frac{1}{2}[\alpha(n+1)]^2\}}]\}$$

et de

$$\sum_{n=1}^{\infty} \frac{1}{n\alpha(n)} \cdot E\{\sum_{t=1}^{n} ||X_t||. I_{\{(\alpha(n))^2\leqslant 2||X_t||\leqslant(\alpha(n+1))^2\}}\} .$$

Dans les 2 cas, la condition $\sum_1^{\infty} (\alpha(n))^{-1} < \infty$ suffit à la convergence en majorant brutalement les espérances considérées par la constante $E(||X_t||)$.

Q.E.D.

C.IV - 2ème PARTIE
CONVERGENCE L^2

Théorème 7.- Supposons que le processus observé (X_t) , $t \in \mathbb{Z}$, vérifie la "condition $L^2(\mu)$ " , que g existe et soit telle que $\lim_{n\to\infty} (||(n^{-1}) . (\sum_{t=1}^{n} g_{X_t}) - g||_{L^2(\mu)}) = 0$, que la suite $h(n)$ vérifie : $\lim_{n\to\infty} (h(n)) = \lim_{n\to\infty} [n^{-1} . [h(n)]^{-d}] = 0$, et que le noyau K vérifie (41) et (42).

a) Si $\lim_n E(||(n^{-1}) . \sum_{t=1}^{n} (g_{X_t}^{F_{t-1}} - g_{X_t})||_{L^2(\mu)}) = 0$, on a aussi $\lim_n E(||\hat{f}_n^6 - g||_{L^2(\mu)}) = 0$, et de même $\lim_n E(||\hat{f}_n^7 - g||_{L^2(\mu)}) = 0$, pourvu que les $(\beta_{t,n})$, outre (38), vérifient :

(44) $\quad \lim_n (\sum_{t=1}^{n} \beta_{t,n}^2 [h(t)]^{-d}) = 0$, et

(45) $\quad \sup_{n\geq 1} \{\sum_{t=1}^{n-1} t\ |\beta_{t,n}-\beta_{t,n+1}|\} < \infty$, $\sup_{n\geq 1} \{\sum_{t=1}^{n-1} t\ \beta_{t,n}\ |(\frac{h(t+1)}{h(t)})^d - 1|\} < \infty$,

$\sup_n (n\ \beta_{n,n}) < \infty$.

b) Si $\lim_n$ p.s. $(||(n^{-1}) . \sum_{t=1}^{n} (g_{X_t}^{F_{t-1}} - g_{X_t})||_{L^2(\mu)}) = 0$, on aura aussi

α) $\lim_{n\to\infty}$ p.s. $(||\hat{f}_n^6 - g||_{L^2(\mu)}) = 0$, si les $h(n)$ vérifient $(\frac{h(n)}{h(n+1)})^{d/2} < 1 + \frac{1}{n}$, pour n assez grand, et $\sum_{n=1}^{\infty} n^{-3/2} . (h(n))^{-d} < \infty$.

β) $\lim_{n\to\infty}$ p.s. $(||\hat{f}_n^7 - g||_{L^2(\mu)}) = 0$, si les $\beta_{t,n}$ vérifient, outre (38) et (44) :

(46) $\quad \sum_n \beta_{n,n}^2 . (h(n))^{-d} < \infty$ et

$$\sum_n \{[\sum_{t=1}^{n} (\beta_{t,n}^2+\beta_{t,n+1}^2)(h(t))^{-d}]^{1/2}\ [\sum_{t=1}^{n} (\beta_{t,n}-\beta_{t,n+1})^2(h(t))^{-d}]^{1/2}\}$$

c) En particulier, si $\beta_{t,n} = (n^{-1})$, $1 \leq t \leq n$, ou si $\beta_{t,n} = (h(t))^d \,/\, [\sum_{t=1}^{n} (h(t))^d] < \infty$, les conditions imposées ci-dessus sont vérifiées si $\sum_n (h(n))^{-d} \,.\, (n^{-2}) < \infty$, pourvu que $(\frac{h(n)}{h(n+1)})^d$ soit inférieur à $1 + 1/n$, pour n assez grand

Démonstration :

α) Nous avons : $\hat{f}_n^6 = \sum_{t=1}^{n} K_{t,n}^6(X_t)$,

$K_{t,n}^6(x) = (n^{-1})\ (h(n))^{-d} \,.\, K[(. -x)/h(n)]$ et dès lors $\forall x$, $||K_{t,n}^6(x)||^2_{L^2(\mu)} = O\{(n^{-2})\ (h(n))^{-d}\}$ (moyennant un simple changement de variable dans le calcul de l'intégrale) , pour $t = 1,\ldots,n$.

Il vient : $\alpha_{t,n}^6 = O\ \{ n^{-1} \,.\, (h(n))^{-d/2}\}$.

De même : $||K_{t,n}^6(x) - K_{t,n+1}^6(x)||^2_{L^2(\mu)}$ vaut, par un calcul direct :

$$(n^{-2}) \,.\, (h(n)^{-d}) \,.\, ||K||^2_{L^2(\mu)} + (n+1)^{-2} \,.\, [h(n+1)^{-d}]\ ||K||^2_{L^2(\mu)}$$

$$- 2(n^{-1})\ [(n+1)^{-1}] \,.\, [(h(n+1)^{-d}] \int_{\mathbb{R}^d} K(u)\ K\{\frac{h(n)}{h(n+1)}\, u\}\ d\mu(u) ,$$

et se majore donc, en posant $\alpha(n) = \frac{h(n)}{h(n+1)} - 1$, puisque K vérifie (41), par :

$$||K||^2_{L^2(\mu)}\ \{(n^{-2})\ (h(n))^{-d} + (n+1)^{-2} \,.\, [h(n+1)]^{-d} - 2n^{-1}.(n+1)^{-1}[h(n+1)]^{-d}\}$$

(47)

$$+ O\ \{\alpha(n) \,.\, n^{-1} \,.\, (n+1)^{-1}\ [(h(n+1))^{-d}]\}$$

En décomposant la première accolade, on obtient :

$$\{(n^{-1}).(h(n))^{-d/2}-(n+1)^{-1}.(h(n+1))^{-d/2}\}^2 + 2n^{-1}.(n+1)^{-1}.[h(n+1)]^{-d/2}$$
$$\{(h(n))^{-d/2}-(h(n+1))^{-d/2}\}$$

majoré enfin par :

$$2[n^{-1}-(n+1)^{-1}]^2\ h(n)^{-d} + 2(n+1)^{-2}\ (h(n)^{-d/2} - h(n+1)^{-d/2})^2$$
$$+ 2\ n^{-1}.(n+1)^{-1}.(h(n+1))^{-d/2}[h(n)^{-d/2}-h(n+1)^{-d/2}]$$

soit encore :

$$\{2\ n^{-2}(n+1)^{-2}.(h(n))^{-d} + 2(n+1)^{-2}.(h(n))^{-d}.[(\frac{h(n)}{h(n+1)})^{d/2} - 1]^2$$
$$+ 2\ n^{-1}(n+1)^{-1}[h(n)\ h(n+1)]^{-d/2}\ [(\frac{h(n)}{h(n+1)})^{d/2} - 1]\}$$

Finalement, à partir de cette dernière majoration, de (47), et puisque $\alpha(n) \leq (\frac{h(n)}{h(n+1)})^{d/2} - 1$, il vient :

$$(\gamma^6_{t,n})^2 = 0\ \{n^{-4}\ .\ (h(n))^{-d}\} + 0\{n^{-2}\ .\ (h(n))^{-d}\ [(\frac{h(n)}{h(n+1)})^{d/2} - 1]\}$$

Par conséquent, pour le $K^6_{t,n}$, on obtiendra H1) si $\lim_n (n^{-1})\ (h(n))^{-d} = 0$, H2) si $\sum_n n^{-2}\ (h(n))^{-d} < \infty$, H3) si de surcroît $\sum_n (n^{-1})\ (h(n))^{-d} \sqrt{(\frac{h(n)}{h(n+1)})^{d/2} - 1} < \infty$

β) Puisque $K^7_{t,n} = \beta_{t,n}\{[h(t)]^{-d}\ .\ K[(. - x)/h(t)]\}$, on a immédiatement :

$$(\alpha^7_{t,n})^2 = \beta^2_{t,n}\ .\ (h(t))^{-d} \quad \text{et} \quad (\gamma^7_{t,n})^2 = (\beta_{t,n}-\beta_{t,n+1})^2\ (h(t))^{-d}\ .$$

Ceci montre que pour les $K^7_{t,n}$, H1) est vérifiée si (44) l'est, H2) et H3) l'étant sous les 2 conditions qui composent (46).

γ) Du lemme 7) résulte qu'en définissant $[K^6_{t,n}]$ comme l'opérateur $(n^{-1})\ [H_{h(n)}]$, $1 \leq t \leq n$, les $[K^6_{t,n}]$ vérifiant la "condition I". De plus, puisque $\operatorname{Lim}_n \downarrow h(n) = 0$ $\{\sum_{t=1}^{n} [K^6_{t,n}]\}$ converge vers $Id_{L^2(\mu)}$.

On aura de même, à partir du lemme : $||K^6_{t,n}|| = O(n^{-1})$, $1 \leq t \leq n$, et puisque ici $[K_{t,n}] = [K_{t+1,n}]$, $1 \leq t \leq n-1$, les conditions H4) à H7) sont vérifiées pour les $[K^6_{t,n}]$.

On déduit alors immédiatement du point α) ci-dessus et du théorème 1) la partie du théorème concernant $\hat{f}^6_n$, pourvu que l'on prouve que : $\sum_{n\geq 1} n^{-3/2}\ (h(n))^{-d}$ et $(\frac{h(n)}{h(n+1)})^{d/2} < 1 + \frac{1}{n}$ pour n assez grand réalisent les conditions imposées à $h(n)$ pour obtenir H2) et H3), ce qui est direct.

On peut d'ailleurs noter que la condition : $\operatorname{Lim}_n [(n^{-1})(h(n))^{-d}] = 0$, prise comme hypothèse générale, découle aussi des 2 conditions de b) α), la majoration de $(\frac{h(n)}{h(n+1)})^{d/2}$ donnant la décroissance de $(n^{-2})(h(n))^{-d}$.

δ) On déduit ensuite du lemme 7) qu'en posant $[K^7_{t,n}] = \beta_{t,n}[H_{h(t)}]$, les $[K^7_{t,n}]$ vérifient "la condition I". Comme de plus on a si $g \in L^2(\mu)$:

$$||\sum_{t=1}^{n} [K^7_{t,n}](g)-g||_{L^2(\mu)} \leq \sum_{t=1}^{n} \beta_{t,n}\ ||[H_{h(t)}](g)-g||_{L^2(\mu)} +$$

$$+ [(\sum_{t=1}^{n} \beta_{t,n}) - 1]\ \ ||g||_{L^2(\mu)} ,$$

$h_n(t)$ décroissant vers 0 , $\sum_{1}^{n} [K^7_{t,n}]$ converge vers $Id_{L^2(\mu)}$, pourvu que $\operatorname{Lim}_{n\to\infty} (\sum_{t=1}^{n} \beta_{t,n}) = 1$ c'est-à-dire sous (38).

On vérifie de même H5) et H6), sous (38), puisque $||[K^7_{t,n}]|| = O(\beta_{t,n})$.

Reste à vérifier H7), or on a :

$$||[K^7_{t,n}]-[K^7_{t+1,n}]||_{L^2(\mu)} \leq |\beta_{t,n}-\beta_{t+1,n}| \; ||[H_{h(t)}]||_{L^2(\mu)} +$$

$$+ \beta_{t+1,n} \; ||([H_{h(t)}]-[H_{h(t+1)}])||_{L^2(\mu)}$$

Maintenant, pour $g \in L^2(\mu)$:

$$||\{[H_{h(t)}]-[H_{h(t+1)}]\}(g)||_{L^2(\mu)} \leq ||(h(t))^{-d} \{\int K(\frac{\cdot - y}{h(t)}) \, g(y) \, d\mu(y) -$$

$$- \int K(\frac{\cdot - y}{h(t+1)}) \, g(y) \, d\mu(y)\}||_{L^2(\mu)} + ||[(h(t))^{-d} - (h(t+1)^{-d}] \,.$$

$$. \int K(\frac{\cdot - y}{h(t+1)}) \, g(y) \, d\mu(y)||_{L^2(\mu)}$$

$$\leq || \int [K(u) - K(\frac{h(t)}{h(t+1)} u)] \, g(. + u \, h(t)) \, d\mu(u)||_{L^2(\mu)}$$

$$+ |(\frac{h(t+1)}{h(t)})^d - 1| \, . \, ||[H_{h(t+1)}](g)||_{L^2(\mu)}$$

et le premier des deux majorants s'écrit :

$$\{\iiint [K(u) - K(\frac{h(t)}{h(t+1)} u)] \, [K(u') - K(\frac{h(t)}{h(t+1)} u')] \, g(x+uh(t))$$

$$(g(x+u'h(t)) \, d\mu^{\otimes 3}(u,u',x)\}^{1/2}$$

et se trouve à son tour majoré par :

$$\{\int |K(u) - K(\frac{h(t)}{h(t+1)} u)| \, d\mu(u)\} \, . \, ||g||_{L^2(\mu)} \, .$$

En regroupant les inégalités précédentes, vient :

$$||[K^7_{t,n}]-[K^7_{t+1,n}]|| = O\{|\beta_{t,n}-\beta_{t,n+1}| + \beta_{t+1,n} \, |(\frac{h(t+1)}{h(t)})^d - 1|$$

$$+ \beta_{t+1,n} \, . \, \{\int |K(u) - K\,(\frac{h(t)}{h(t+1)} u) \;\; d\mu(u)|\}$$

soit, puisque le noyau vérifie (42)

$$= O\{|\beta_{t,n}-\beta_{t+1,n}| + \beta_{t+1,n} \, |(\frac{h(t+1)}{h(t)})^d - 1|\} .$$

Dès lors H7 est vérifiée pour les $[K^7_{t,n}]$, si (45) l'est .

De ces résultats, du point β, du théorème 1, découlent les conclusions du théorème concernant $\hat{f}^7_n$.

ω) Pour achever, il nous faut vérifier que les conditions "minimales" sur $\beta_{t,n}$ exprimées en c) entraînent toutes celles qui précèdent, dans les 2 cas particuliers que nous étudions systématiquement.

* Si $(\beta_{t,n}) = n^{-1}$, $1 \leq t \leq n$, (38), (44) et (45) sont immédiates, au vu des hypothèses générales sur $h(n)$ et $(\frac{h(n)}{h(n+1)})^d$.
La 2ème condition de (46) s'écrit :

$$\sum_{n\geq 1} n^{-3} \, (\sum_{t=1}^{n} (h(t))^{-d}) < \infty$$

elle est donc vérifiée si $\sum_{n\geq 1} (n^{-2}) \, (h(n))^{-d} < \infty$, condition qui entraîne les deux manquantes, pour peu que $(n^{-2}) \, (h(n))^{-d}$ soit une suite décroissante c'est-à-dire que $(\frac{h(n)}{h(n+1)})^d < 1 + \frac{1}{n}$, pour n grand, par exemple ...

* Si $(\beta_{t,n}) = (h(t))^d \, / \, (\sum_{t=1}^{n} [h(t)]^d)$, un calcul de routine montre que si $(\frac{h(t+1)}{h(t)})^d - 1 < \frac{1}{t}$ pour t assez grand, il suffira d'avoir, comme ci-dessus, pour obtenir toutes les contraintes souhaitées :

$$\lim_{n\to\infty} [n^{-1} \, (h(n))]^{-1} = 0 \quad \text{et} \quad \sum_{n} (n^{-2}) \, (h(n))^{-d} < \infty ,$$

la 2ème condition impliquant alors la première. Q.E.D.

C.IV - 3ème PARTIE

CONVERGENCE L^1

Théorème 8.- Supposons que le processus observé (X_t), $t \in \mathbb{Z}$, admette des densités $g_{X_t}^{F_{t-1}}$ et g_{X_t}, par rapport à μ, mesure de Lebesgue sur $\mathbb{R}^d$, qu'il existe g telle que $\operatorname{Lim}_n (\|(n^{-1}) . (\sum_{t=1}^n g_{X_t}) - g\|_{L^1(\mu)}) = 0$, que $h(n)$ vérifie $\operatorname{Lim}_{n\to\infty} (h(n)) = \operatorname{Lim}_{n\to\infty} \{(n^{-1}) . [h(n)^{-3d}]\} = 0$, que le noyau K vérifie (41) et (42). Alors

a) Si $\operatorname{Lim}_{n\to\infty} E(\|(n^{-1}) . \sum_{t=1}^n (g_{X_t}^{F_{t-1}} - g_{X_t})\|_{L^1(\mu)}) = 0$, on a aussi $\operatorname{Lim}_{n\to\infty} E(\|\hat{f}_n^6 - g\|_{L^1(\mu)}) = 0$, et de même $\operatorname{Lim}_{n\to\infty} E(\|\hat{f}_n^7 - g\|_{L^1(\mu)}) = 0$, pourvu que les $\beta_{t,n}$ vérifient, outre (38) et (45) :

$$(48) \qquad \operatorname{Lim}_{n\to\infty} \{(h(n))^{-2d}\} . \{\sum_{t=1}^n \beta_{t,n}^2 . (h(t))^{-d}\} = 0$$

et que $\operatorname{Lim}_{n\to\infty} (\sum_1^n \beta_{t,n} . P(\|X_t\| > \frac{h(n)}{2}^{-2}) = 0$.

b) Si $\operatorname{Lim}_{n\to\infty}$ p.s. $(\|(n^{-1}) \sum_{t=1}^n (g_{X_t}^{F_{t-1}} - g_{X_t})\|_{L^1(\mu)}) = 0$, on aura

α) $\operatorname{Lim}_{n\to\infty}$ p.s. $(\|\hat{f}_n^6 - g\|_{L^1(\mu)}) = 0$, pourvu que $h(n)$ vérifie :

$$(49) \qquad \sum_{n\geq 1} (n^{-3/2}) . (h(n))^{-3d} < \infty, \quad \text{avec} \quad (\frac{h(n)}{h(n+1)})^d \leq 1 + \frac{1}{n}$$

pour n grand et que $\operatorname{Lim}_{n\to\infty}$ p.s. $[(n^{-1}) . \sum_{t=1}^n I_{\{\|X_t\| > \frac{(h(n))^{-2}}{2}\}}] = 0$, condition vérifiée sous les conditions du lemme 9

β) $\operatorname{Lim}_{n\to\infty}$ p.s. $(\|\hat{f}_n^7 - g\|_{L^1(\mu)}) = 0$ si les $\beta_{t,n}$ vérifient, outre (38), (45), (47) :

(50) $$\sum_{1}^{\infty} \beta_{nn}^{2}(h(n))^{-3d} < \infty$$

et

(51) $$\sum_{n\geq 1} [h(n)]^{-3d} \cdot \sqrt{(\sum_{t=1}^{n} \beta_{t,n}^{2}) [\sum_{t=1}^{n} (\beta_{t,n}-\beta_{t,n+1})^{2}]} \quad \text{et}$$

$$\sum_{n\geq 1} (h(n))^{-3d} \cdot [(\frac{h(n)}{h(n+1)})^{d} - 1] \sqrt{(\sum_{t=1}^{n} \beta_{t,n}^{2}) (\sum_{t=1}^{n} \beta_{t,n+1}^{2})}$$

et que $\underset{n\to\infty}{\text{Lim}}$ p.s. $(\sum_{t=1}^{n} \beta_{t,n} \cdot I_{\{||X_t||>\frac{(h(n))^{-2}}{2}\}})$

c) Si le processus (X_t) , $t \in \mathbb{Z}$ est strictement stationnaire et que les $||X_t||$ ont une espérance toutes les conditions nécessaires aux convergences de $\hat{f}_n^6$ sont réunies si $h(n)$ vérifie outre (49) :

(52) $$\sum_{n\geq 1} (n^{-1}) [h(n)]^{2} < \infty$$

($h(n) = n^{-\alpha}$, $0 < \alpha < 1/6d$ convient ...).

Dans les mêmes conditions, si $\beta_{t,n}$ vérifie (38), (45), (48), (42), (50), (51), on obtiendra les convergences de $\hat{f}_n^7$ vers g. Pourvu que $(\frac{h(n)}{h(n+1)})^{d}$ soit majoré par $1 + \frac{1}{n}$ pour n assez grand, ce sera donc le cas

* si $\beta_{t,n} = (n^{-1})$, $1 \leq t \leq n$, $h(n)$ vérifiant (52) et $\sum_{n\geq 1} (n^{-2}) (h(n))^{-3d} < \infty$

* si $\beta_{t,n} = (h(t))^{d} / (\sum_{t=1}^{n} h(t))^{d}$, $h(n)$ vérifiant (52) et $\sum_{n\geq 1} (n^{-2}) (h(n))^{-5d} < \infty$ (d'où le choix possible de $h(n)$ en $n^{-\alpha}$).

Démonstration :

α) Posons $K^{8}_{t,n} = [h(n)]^{-d} \, . \, K^{6}_{t,n}$. On déduit de la démonstration du théorème précédent que $(\alpha^{8}_{t,n})^{2} = \sup_{x\in\mathbb{R}^d} ||K^{8}_{t,n}(x)||_{L^2(\mu)} = (n^{-2}).(h(n))^{-3d}$

De même :

$$(\gamma^{8}_{t,n})^{2} = \sup_{x\in\mathbb{R}^d} ||(h(n))^{-d} \, . \, K^{6}_{t,n}(x) - (h(n+1))^{-d} \, K^{6}_{t,n+1}(x)||^{2}_{L^2(\mu)}$$

$$\leqslant 2(h(n))^{-2} \, (\gamma^{6}_{t,n})^{2} + 2 \, ((h(n))^{-d} - (h(n+1)^{-d})^{2} \, (\alpha^{6}_{t,n+1})^{2}$$

$$= O\{(h(n))^{-2d} \, . \, [(n^{-2}) \, (h(n))^{-d} \, . \, [(\frac{h(n)}{h(n+1)})^{d/2} - 1] + (n^{-4}) \, (h(n))^{-d}]\}$$

$$+ O\{(h(n))^{-2d} \, [(\frac{h(n)}{h(n+1)})^{d} - 1]^{2} \, . \, (n+1)^{-2} \, h(n+1)^{-d}\} \, .$$

On déduit alors de la proposition 1 que

* si $\lim_{n\to\infty} (n^{-1}) \, [h(n)^{-3d}] = 0$, H2) étant vérifiée pour les $K^{8}_{t,n}$ on aura :

$$\lim_{n\to\infty} [h(n)]^{-2d} \, . \, E(||\sum_{1}^{n} Z_{t,n}||^{2}_{L^2(\mu)}) = 0 \, ,$$

où $Z_{t,n} = K^{6}_{t,n}(X_t) - E^{F_{t-1}}(K^{6}_{t,n}(X_t))$

* si de surcroît $\sum_{n} (n^{-2}) \, (h(n))^{-3d} < \infty$

et $\sum_{n\geq 1} (n^{-1}) \, . \, (h(n)^{-3d}) \sqrt{(\frac{h(n)}{h(n+1)})^{d} - 1} < \infty$, H2) et H3) le sont et

$$\text{Lim p.s.}_{n\to\infty} \{[h(n)]^{-2} \, . \, (||\sum_{1}^{n} Z_{t,n}||^{2}_{L^2(\mu)})\} = 0$$

β) Appelons C_n l'hypercube $[- (h(n))^{-2} \, , + h(n)^{-2}]^{d}$ et I_n le sous-ensemble (aléatoire) des indices i , $i \in \{1,\dots,n\}$ tel que

$$i \in I_n \iff (||X_i|| < \frac{h(n)^{-2}}{2})$$

J_n représentant $\{1,\dots,n\}-I_n$, on peut alors écrire, si $Z_n = \sum_{t=1}^{n} Z_{t,n}$ et $E_n = ||Z_n||_{L^1(\mu)}$ (définis ci-dessus) :

$E_n \leq E_n^1 + E_n^2 + E_n^3$ avec $E_n^1 = \int_{C_n} |Z_n(x)|\, d\mu(x)$ $E_n^2 = \int_{\bar{C}_n} |\sum_{t\in I_n} Z_{t,n}|\, d\mu$,

$$E_n^3 = \int_{\bar{C}_n} |\sum_{t\in J_n} Z_{t,n}|\, d\mu \;.$$

Il vient :

$$* \quad E_n^1 = O\{|\mu(C_n)|^{1/2} \,.\, ||Z_n||_{L^2(\mu)}\} = O\{h(n)^{-d} \,.\, ||\sum_1^n Z_{t,n}||_{L^2(\mu)}\}$$

qui tend vers 0 si $n \to \infty$, d'après α) sous les conditions du théorème.

$* \quad E_n^2 \leq \sum_{t\in I_n} ||I_{\bar{C}_n} \,.\, \{Z_{t,n}\}||_{L^1}$, ce qui donne

$$E_n^2 = O\{n \,.\, \sup_{t\in I_n} |||I_{\bar{C}_n} \,.\, K_{t,n}^6(X_t)||_{L^1}|\}$$

Mais si $t \in I_n$, $x \in \bar{C}_n$, on aura :

$$||h(n))^{-1}.(x-X_i)||_{\mathbb{R}^d} \geq (h(n))^{-1}.(||x||-||X_i||) \geq (h(n))^{-1}$$
$$\{\sqrt{d}\,(h(n))^{-2} - \frac{1}{2}(h(n))^{-2}\}$$

Dès lors, si $t \in I_n$:

$$||I_{\bar{C}_n} \,.\, K_{t,n}^6(X_t)||_{L^1} = \frac{1}{n}\int_{\bar{C}_n} \frac{1}{(h(n))^d}\, K(\frac{x - X_t}{h(n)})\, dx \leq \frac{1}{n}\int_{L_n} K(u)\, du \;,$$

où $L_n = \{u \,|\, ||u|| > \frac{1}{2} h(n)^{-3}\}$. Il vient : $E_n^2 = O(\int_{L_n} K(u)\, du)$, qui tend vers 0 , si $n \to \infty$, puisque K est intégrable.

$$* \quad E_n^3 \leq (\mathrm{Card}\, J_n) \,.\, \sup_{t\in J_n} \{||I_{\bar{C}_n}\, (Z_{t,n}(X_t))||_{L^1(\mu)}\} \,.$$

Mais comme $||I_{\overline{C}_n}(Z_{t,n}(X_t)||_{L^1(\mu)}$ est $O(\frac{1}{n})$, il vient : $E_n^3 \leq \frac{\text{Card } J_n}{n}$.

Finalement :

* Si $\lim\limits_{n\to\infty} [(n^{-1})\ (h(n))^{-3d}] = 0$, on a $\lim\limits_{n} E(E_n^1) = 0$. Mais $E(\frac{\text{Card } J_n}{n})$ tend également vers 0 , puisque valant $\int_{L_n} (\frac{1}{n} \sum_1^n g_{X_t})\, d\mu$ qui se majore par $||\frac{1}{n} (\sum_1^n g_{X_t}) - g||_{L^1(\mu)} + \int_{L_n} g\, d\mu$.

Donc on aura : $\lim\limits_{n\to\infty} E(E_n) = 0$.

* Si de plus $\sum\limits_n (n^{-2})\ (h(n))^{-3d} < \infty$,

et $\sum\limits_{n\geq 1} (n^{-1})\ (h(n)^{-3d}) \sqrt{(\frac{h(n)}{h(n+1)})^d - 1} < \infty$, $E_n^1 \longrightarrow 0$ p.s. En supposant que $\frac{\text{Card } J_n}{n} \xrightarrow[\text{p.s.}]{} 0$, on aura : $E_n \xrightarrow[\text{p.s.}]{} 0$.

γ) D'après le lemme 7) les opérateurs $[K_{t,n}^6] = (n^{-1}) . [H_{h(n)}]$ considérés de $L^1(\mu)$ dans $L^1(\mu)$, vérifient "la condition I" et les conditions H4) à H7), puisque $\lim\limits_{n\to\infty} [H_{h(n)}] = Id_{L^1(\mu)}$ et $||H_\beta|| \leq 1$ pour tout $\beta > 0$. On en déduit que, sous les conditions du théorème, d'après le théorème 1, on a $(||(n^{-1})\ (\sum_1^n E^{F_{t-1}}(K_{t,n}^6(X_t)) - g||_{L^1(\mu)})$ qui tend vers zéro, en moyenne, ou p.s. selon la convergence à 0 , p.s. ou en moyenne de $||(n^{-1}) \sum_1^n (g_{X_t}^{F_{t-1}} - g_{X_t})||_{L^1(\mu)}$.

Ceci permet de conclure, à l'aide de β) pour la partie du théorème concernant $\hat{f}_n^6$, moyennant une légère généralisation des conditions sur $h(n)$.

δ) Pour utiliser la même technique qu'en α) pour les $K_{t,n}^6$, posons $K_{t,n}^9 = (h(n))^{-d} . K_{t,n}^7$, $1 \leq t \leq n$, $n \geq 1$. On obtient, à partir des calculs de $\alpha_{t,n}^7$ et $\gamma_{t,n}^7$:

$$* \quad (\alpha^9_{t,n})^2 = (h(n))^{-2d} . \{ \sup_{x\in\mathbb{R}^d} ||K^7_{t,n}(x)||^2_{L^2(\mu)} \}$$

$$= [h(n)^{-2d} . \beta^2_{t,n} . (h(t))^{-d}]$$

$$* \quad (\gamma^9_{t,n})^2 = \bigcirc\{[h(t)]^{-d} . ((h(n))^{-d} . \beta_{t,n}-(h(n+1)^{-d} . \beta_{t,n+1})^2\}$$

$$= \bigcirc\{(h(t))^{-d} . (h(n))^{-2d} . (\beta_{t,n}-\beta_{t,n+1})^2\}$$

$$+ \bigcirc\{(h(t))^{-d} . h(n)^{-2d} . (\beta_{t,n})^2 . [(\frac{h(n)}{h(n+1)})^d - 1]^2\}$$

En posant ici $U_{t,n} = K^7_{t,n}(X_t) - E^{\mathcal{F}_{t-1}}(K^7_{t,n}(X_t))$ et $U_n = \sum_{t=1}^{n} U_{t,n}$ on aura donc ici, d'après la proposition 1

$$* \lim_{n\to\infty} (h(n))^{-2d} . E(||U_n||^2_{L^2(\mu)}) = 0$$

$$\text{si} \quad \lim_{n\to\infty} \{(h(n))^{-2d} . \sum_{t=1}^{n} (h(t))^{-d} . \beta^2_{t,n})\} = 0$$

$$* \lim_{n\to\infty} \text{p.s.} \ (h(n))^{-2d} . ||U_n||^2_{L^2(\mu)}) = 0$$

$$\text{si} \quad \sum_n h(n)^{-3d} . \beta^2_{n,n} < \infty ,$$

$$\sum_{n\geq 1} (h(n))^{-3d} \sqrt{(\sum_1^n \beta^2_{t,n}) \ [\sum_{t=1}^{n} (\beta_{t,n}-\beta_{t,n+1})^2} < \infty ,$$

$$\text{et} \quad \sum_{n\geq 1} h(n)^{-3d} \ [(\frac{h(n)}{h(n+1)})^d - 1)] \sqrt{(\sum_1^n \beta^2_{t,n}) \ (\sum_1^n \beta^2_{t,n+1})} < \infty$$

(par un simple calcul de routine pour vérifier H1), H2), H3))

ε) On majore ensuite $||U_n||_{L^1(\mu)}$ par la somme de U^1_n, U^2_n, U^3_n, définis à partir des U^t_n comme l'étaient E^1_n, E^2_n, E^3_n (cf. point β). On obtiendra

* $U_n^1 = \int_{C_n} |Z_n| \, d\mu = \bigcirc(h(n)^{-d} \, . \, ||U_n||_{L^2})$ qui tend vers zéro en moyenne, ou p.s., sous les conditions dégagées ci-dessus

* $U_n^2 = \int_{\bar{C}_n} |\sum_{t\in I_n} U_{t,n}| \, d\lambda = \bigcirc\{\sum_{t\in I_n} \int_{\bar{C}_n} \beta_{t,n} \left(\frac{1}{h(t)}\right)^d . \, K\left(\frac{x - X_t}{h(t)}\right) d\lambda$.

Ici aussi si $x \in \bar{C}_n$ et $t \in I_n$ on a :

$$||(h(t))^{-1} \, (x - X_t)||_{\mathbb{R}^d} \geq (h(t)^{-1}) \left[\frac{1}{2} \, (h(n))^{-2}\right] \geq \frac{1}{2} \, (h(t))^{-3}$$

donc :

$$U_n^2 = \bigcirc\{\sum_{t=1}^{n} \beta_{t,n} \, . \int_{L_t} K(u) \, du\} \quad . \quad (L_t = \{u \, / \, ||u|| > \frac{(h(t))^{-3}}{2}\})$$

* $U_n^3 = \int_{\bar{C}_n} |\sum_{t\in J_n} U_{t,n}| \quad d\mu = \bigcirc(\sum_{t\in J_n} \beta_{t,n})$

ou encore $$U_n^3 = \bigcirc\{\sum_{t=1}^{n} \beta_{t,n} \, I_{\{||X_t|| > \frac{h(n)^{-2}}{2}\}}\} \quad .$$

ζ) D'après le lemme 6, les opérateurs $[K_{t,n}^7] = \beta_{t,n} \, . \, [H_{h(t)}]$, de $L^1(\mu)$ dans $L^1(\mu)$, vérifient la "condition I" , sont bornées par $\beta_{t,n}$ en norme, et tels que $(\sum_1^n [K_{t,n}^7]) \longrightarrow Id_{L^1(\mu)}$, car $(\sum_1^n \beta_{t,n}) \longrightarrow 1$. H4), H5), H6) sont donc vérifiées .

Pour H7) on utilisera la décomposition déjà exhibée au théorème 7

$$||[K_{t,n}^7]-[K_{t,n+1}^7]||_{L^1(\mu)} \leq |\beta_{t,n}-\beta_{t,n+1}| \, [||H_{h(t)}]||_{L^1(\mu)}$$

$$+ \beta_{t+1,n} \, ||[H_{h(t)}]-[H_{h(t+1)}]||_{L^1(\mu)}$$

Maintenant

$$||([H_{h(t)}]-[H_{h(t+1)}])(g)||_{L^1(\mu)}$$

$$\leq ||\int [K(u) - K(\frac{h(t)}{h(t+1)})u] g(.+uh(t))\ d\mu(u)||_{L^1(\mu)}$$

$$+ |(\frac{h(t+1)}{h(t)})^d - 1| \ .\ |||H_{[h(t+1)]}|(g)||_{L^1(\mu)}$$

$$\leq \{\int |K(u) - K(\frac{h(t)}{h(t+1)}\ u)|\ d\mu(u)\} \ .\ ||g||_{L^1(\mu)}$$

$$+ |(\frac{h(t+1)}{h(t)})^d - 1| \ .\ ||g||_{L^1(\mu)}$$

pour tout g de $L^1(\mu)$. Finalement, K vérifiant (42), on a

$$||[K^7_{t,n}]-[K^7_{t,n+1}]||_{L^1(\mu)} \leq |\beta_{t,n}-\beta_{t,n+1}| + ((\frac{h(t+1)}{h(t)})^d - 1)\ \beta_{t,n+1}$$

et H7) est vérifiée si (45) l'est ...

η) On conclut de ce qui précède que (45) étant vérifiée, la convergence en moyenne ou p.s. de $||\hat{f}^7_n - g||_{L^1(\mu)}$ se ramène à celles de U^1_n , U^2_n , U^3_n ...

U^2_n tend sûrement vers 0 car $\lim_{t\to\infty} \int_{L_t} K(u)\ du = 0$, et on peut alors appliquer le lemme 3. Sous (48) U^1_n tend vers 0 en moyenne, en ajoutant (50) et (51), U^1_n tend vers 0 presque sûrement ... Reste le problème de U^3_n , qui est résolu sous les conditions du théorème puisque $U^3_n = \sum_1^n \beta_{t,n}\ I_{\{||X_t||> \frac{(h(n))^{-2}}{2}\}}$.

ω) Dans le cas où les (X_t) , $t \in \mathbb{Z}$, sont strictement stationnaires et possèdent une espérance, le lemme (9) montre que la suite $(\sum_1^n I_{\{||X_t|| > (h(n))^{-2}\}})$ converge vers 0 en moyenne et presque sûrement pourvu que $\sum_{n\geq 1} (n^{-1}) (h(n))^2 < \infty$, (condition 52). Alors, ne reste bien à vérifier que (49) pour avoir les convergences de $\hat{f}_n^6$ vers g , en moyenne ou presque sûrement.

Pour $\hat{f}_n^7$, il nous faudra vérifier la suite de conditions (38), (43), (45), (48), (50), (51) pour conclure de même. Mais

* Lorsque $\beta_{t,n} = (n^{-1})$, on se ramène directement aux deux conditions du théorème.

* Lorsque $\beta_{t,n} = (h(t))^d \,/\, (\sum_1^n h(t))^d$, un calcul un peu moins direct montre qu'on vérifie (38) si $\lim_{n\to\infty} (\sum_1^n (h(t))^d) = \infty$, (43) si $\sum_1^\infty (n^{-1}) (h(n))^2 < \infty$, (45) toujours, (48), si $\lim_n (n^{-1}) (h(n)^{-3d}) = 0$, (50) si $\sum_n (n^{-2}) (h(n)^{-3d}) < \infty$, (51) si $\sum_n (n^{-2}) h(n)^{-5d} < \infty$ ... finalement on peut conclure comme dans le théorème. Q.E.D.

C.V.

ESTIMATION DES DERIVEES

Nous supposerons dans ce chapitre (comme au chapitre III, § 3), que la limite g des densités $(n^{-1} . \sum_{t=1}^{n} g_{X_t})$ admet des dérivées partielles continues. Pour estimer celles-ci, on utilisera les dérivées partielles de l'estimateur $\hat{f}_n^6$ dont nous avons vu au chapitre IV qu'il convergeait vers g dans L^1 et L^2 , sous les hypothèses suffisantes : l'existence des $D^i(\hat{f}_n^6)$ dépendant uniquement de la différentiabilité du noyau K , choisi à discrétion de l'utilisateur, la construction de l'estimateur cherché ne pose pas de problème particulier.

Comme dans le chapitre précédent, nous démontrerons un théorème de convergence L^2 de $(D^i \hat{f}_n^6)$ vers $(D^i g)$, lorsque les $g_{X_t}^{\mathcal{F}_{t-1}}$, g_{X_t} g , sont dans C^{12} et que les convergences de $(n^{-1} [\sum_{t=1}^{n} (g_{X_t}^{\mathcal{F}_{t-1}} - g_{X_t})])$ et $[n^{-1} (\sum_{t=1}^{n} g_{X_t}) - g]$ ont lieu dans W^{12} . Nous en déduirons un théorème de convergence L^1 de l'estimateur utilisé sous l'hypothèse d'existence des $g_{X_t}^{\mathcal{F}_{t-1}}$ et g_{X_t} et de leurs dérivées partielles, l'appartenance de celles-ci à $L^1(\mathbb{R}^d)$, et les convergences correspondantes dans W^{11} .

Pour fixer les notations, nous poserons comme d'habitude :

(53) $$\begin{cases} \hat{f}_n^{10} = \sum_{t=1}^{n} K_{t,n}^{10}(X_t) \\ \\ K_{t,n}^{10}(x) = (n^{-1}) \left[(h(n))^{-d-1}\right] \{D^i K \left[(\cdot - x) / h(n)\right]\} \end{cases} ,$$

pour $1 \leq t \leq n$, $n \geq 1$, en supposant l'appartenance à $C^1(\mathbb{R}^d)$ du noyau K (au sens du chapitre IV) borné.

THEOREME DE CONVERGENCE L^2.

Théorème 9.- Supposons que le processus (X_t) , $t \in \mathbb{Z}$, vérifie la condition $L^2(\mu)$, que de surcroît les $g_{X_t}^{F_{t-1}}$ et g_{X_t} (densités par rapport à la mesure μ de Lebesgue sur $\mathbb{R}^d$) soient comme le noyau K éléments de C^{12} et qu'il existe g dans C^{12} tel que $\lim_{n\to\infty} (||(n^{-1}) \sum_1^n g_{X_t} - g||_{W^{12}}) = 0$. Alors

a) Si $\lim_{n\to\infty} E(||(n^{-1}) . \sum_1^n (g_{X_t}^{F_{t-1}} - g_{X_t})||_{W^{12}}) = 0$, $h(n)$ vérifiant

$\lim_{n\to\infty} \downarrow h(n) = \lim_{n\to\infty} ((n^{-1}) . h(n)^{-d-2}) = 0$, on aura :

$$\lim_{n\to\infty} E(||\hat{f}_n^{10} - D^i g||_{L^2}) = 0$$

b) Supposons maintenant que

$\lim_{n\to\infty}$ p.s. $(||(n^{-1}) \sum_{t=1}^{n} (g_{X_t}^{F_{t-1}} - g_{X_t})||_{W^{12}}) = 0$, que les $h(n)$, outre les conditions du a) vérifient :

$$\sum_{n\geq 1} (n^{-2})\ (h(n)^{-d-2}) < \infty \quad \text{et} \quad \sum_{n\geq 1} (n^{-1})\ (h(n)^{-d-2})\ \sqrt{(\frac{h(n)}{h(n+1)})^{d/2} - 1} < \infty ,$$

et que l'on ait :

$$(54) \qquad \int_{\mathbb{R}^d} [D^i K(u)][D^i K(cu)]\ du = ||D^i K||^2_{L^2} + O(|c-1|) \quad ,$$

alors on obtient : $\underset{n\to\infty}{\text{Lim}}$ p.s. $(||\hat{f}^{10}_n - D^i g||_{L^2}) = 0$.

c) Si $K(x_1,\ldots,x_n) = \prod_1^n K_o(x_i)$, où K_o est la densité de $N(0,1)$, que $(\frac{h(n)}{h(n+1)})^{d-2} < 1 + \frac{1}{n}$ pour n assez grand et $\sum_{n\geq 1} n^{-3/2}\ (h(n))^{-d-2} < \infty$, on obtient les convergences souhaitées.

<u>Démonstration</u> :

α) Le calcul usuel donne :

$$(\alpha^{10}_{t,n})^2 = \sup_{x\in\mathbb{R}^d} (||K^{10}_{t,n}(x)||^2_{L^2}) = (n^{-2})\ (h(n)^{-d-2})\ \ ||D^i K||^2_{L^2}$$

$$(\gamma^{10}_{t,n})^2 = \sup_{x\in\mathbb{R}^d} (\ ||(n^{-1})\ .\ (h(n))^{-d-1}\ .\ (D^i K[(\cdot-x)/h(n)])$$

$$- (n+1)^{-1}\ (h(n+1))^{-d-1}\ \ (D^i K)\ \ [(\cdot-x)/h(n+1)]||^2_{L^2})$$

$$= (n^{-2}\ .\ (h(n))^{-d-2} + (n+1)^{-2}\ .\ (h(n+1))^{-d-2})\ ||D^i\ K||^2_{L^2}$$

$$- 2\ (n^{-1})(n+1)^{-1}\ .\ [h(n)]^{-1}\ .\ [h(n+1)]^{-d-1}\ .$$

$$\int_{\mathbb{R}^d} [(D^i K)(u)]\ \ [(D^i K)\ \ (\frac{h(n)}{h(n+1)}\ u)]\ \ du$$

Donc, sous les conditions du théorème :

$$(\gamma_{t,n}^{10})^2 = ||D_K^i||_{L^2}^2 \{n^{-2} . (h(n))^{-d-2} + (n+1)^{-2} [h(n+1)]^{-d-2} - 2 (n^{-1})$$

$$[(n+1)^{-1}] (h(n))^{-1} . (h(n+1))^{-d-1}\}$$

$$+ \bigcirc\{(\frac{h(n)}{h(n+1)} - 1) n^{-1} . (n+1)^{-1} . (h(n))^{-1} . (h(n+1))^{-d-1}\} .$$

et, par un calcul similaire à celui de la démonstration du théorème 7, on obtient :

$$\sum_1^n (\gamma_{t,n}^{10})^2 = \bigcirc\{n^{-3} . (h(n))^{-d-2}\} + \bigcirc\{n^{-1} . (h(n))^{-d-2} . [(\frac{h(n)}{h(n+1)})^{d/2} - 1]\}$$

H1) est donc vérifiée sous la condition a) du théorème, H2) et H3) sous les conditions du b).

β) Définissons $[K_{t,n}^{10}]$, $1 \leq t \leq n$, $n \geq 1$ comme celui qui à une fonction g de W_{12} associe la fonction définie ponctuellement par :

$$\{[K_{t,n}^{10}](g)\}(y) = (n^{-1}) . \int_{\mathbb{R}^d} [h(n)^{-d-1}] [(D^i K) ((y-x)/h(n))] g(x) dx$$

Alors * Cette définition a un sens , car g et $(D^i K)$ sont dans L^2.

* L'opérateur ainsi défini n'est autre que $(n^{-1}) [H_{h(n)}].(\tilde{D}^i g)$

$H_{h(n)}$ étant l'opérateur défini au lemme 7.

Ceci résulte directement du lemme 5. En posant $L = (h(n))^{-d-2} . K (y - \cdot) / h(n)$, g et L étant dans W^{12} , on peut écrire :

$$[K_{t,n}^{10}](g)(y) = (n^{-1}) (- \int_{\mathbb{R}^d} (\tilde{D}^i(L)) g d\lambda) = (n^{-1}) \int_{\mathbb{R}^d} L (\tilde{D}^i g) d\lambda$$

$$= (n^{-1}) . (\int_{\mathbb{R}^d} (h(n))^{-d} . K [(y-x)/h(n)] . [(\tilde{D}^i g(x)] dx$$

d'où la conclusion. Comme $\tilde{D}^i$ est un opérateur contractant de W^{12} dans L^2, et $[H_{h(n)}]$ un opérateur contractant de L^2 dans L^2, $[K^{10}_{t,n}]$ est un opérateur de W^{12} dans L^2, de norme majorée par (n^{-1}).

De surcroît, on sait que $\lim\limits_{\beta \downarrow 0} [H_\beta] = Id_{L^2(\mu)}$, ce qui implique la convergence forte de $\sum\limits_1^n [K^{10}_{t,n}]$ vers $(\tilde{D}^i)$, pourvu que $h(n) \downarrow 0$.

* Si $g \in L^1 \cap W^{12}$, $[K^{10}_{t,n}](g)$ coïncide avec l'intégrale de Bochner : $\int_{\mathbb{R}^d} K^{10}_{t,n}(x) \,.\, g(x)\, dx$, ce qui revient à vérifier, que pour toute forme linéaire u sur L^2

$$<u \,,\, [K^{10}_{t,n}](g)>_{L^2} = \int_{\mathbb{R}^d} <u \,,\, [K^{10}_{t,n}](g)>_{L^2} d\lambda \,.$$

Il faut donc montrer que pour tout élément A de L^2 :

$$\int A(y) \,.\, \{\int (h(n))^{-d} (D^iK) \,[(y-x)/h(n)]\, g(x)\, dx\}\, dy$$

$$= \int_{\mathbb{R}^d} \{\int A(y).(h(n))^{-d} (D^iK) \,[(y-x)/h(n)]\, dy\}\, g(x)\, dx$$

égalité obtenue sans problème.

* En résumé, les $[K^{10}_{t,n}]$ vérifient H4), H5), H6) et H7) ($[K^{10}_{t,n}] = [K^{10}_{t+1,n}]$) et "la condition I". Nous pouvons donc appliquer le théorème 1.

γ) Le théorème a) b) est application directe des 2 points précédents. Maintenant si $K(x_1,\dots,x_n) = \prod\limits_1^d K_o(x_i)$, K_o étant L^2 et continûment différentiable, la propriété imposée à K en b) s'écrit :

$$(||K_o||_{L^2})^{d-1} \,.\, \int_{\mathbb{R}} K_o'(u)\, \{K_o'(cu) - K_o'(u)\}\, du = O(c-1)$$

ce qui est évident si K_o'' existe et se trouve borné sur $\mathbb{R}$ ... D'où l'exemple cité en c).

De même si $\left(\frac{h(n)}{h(n+1)}\right)^{d+2} < 1 + \frac{1}{n}$, la suite $(n^{-1})\ (h(n))^{-d-2}$ étant décroissante, les conditions de b) impliquant celles de a) sont vérifiées sous celles de c) ! Q.E.D.

THEOREME DE CONVERGENCE L^1.

Théorème 10.- Supposons les densités $g_{X_t}^{F_{t-1}}$, g_{X_t} (par rapport à la mesure de Lebesgue sur $\mathbb{R}^d$) éléments de C^{11}, et de même, le noyau K dans C^{11}, à un support compact. Supposons qu'il existe g dans C^{11} tel que :

$$\lim_{n\to\infty} \|(n^{-1}) \cdot \sum_{t=1}^{n} (g_{X_t}) - g\|_{W^{11}} = 0 ,$$

et posons enfin $M_n = (n^{-1}) \sum_{t=1}^{n} (h(n).) \cdot \|X_t\| \cdot I_{\{\|X_t\| > \frac{1}{2}(h(n))^{-2}\}}$

Alors :

a) Si $\lim_{n\to\infty} E(\|(n^{-1}) \cdot \sum_{t=1}^{n} (g_{X_t}^{F_{t-1}} - g_{X_t})\|_{L^1}) = 0$,

on aura : $\lim_{n\to\infty} E(\|\hat{f}_n^{10} - g\|_{L^1}) = 0$, si $\lim_{n\to\infty} (n^{-1})\ (h(n))^{-3d-2} = 0$

et $\lim_n E(M_n) = 0$.

b) K vérifiant de plus (54), h(n) les conditions

$\sum_{n\geq 1} (n^{-2})\ \left[(h(n))^{-3d-2}\right] < \infty$, et $\sum_{n\geq 1} (n^{-1})(h(n))^{-3d-2} \cdot \left(\left(\frac{h(n)}{h(n+1)}\right)^{d/2} - 1\right)^{1/2} < \infty$

on aura : $\lim_{n\to\infty} \text{p.s. } (\|\hat{f}_n^{10} - g\|_{L^1}) = 0$, pourvu que

$$\lim_{n\to\infty} \text{p.s. } \|(n^{-1}) \sum_1^n (g_{X_t}^{F_{t-1}} - g_{X_t})\|_{L^1} = 0 \quad \text{et} \quad \lim_{n\to\infty} \text{p.s. } M_n = 0 .$$

c) Si X_t est strictement stationnaire et que $E(\|X_t\|)$ existe, toutes les conditions précédentes seront vérifiées si $\sum_1^\infty (n^{-1})\ h(n) < \infty$, $(\frac{h(n)}{h(n+1)})^{3d+2} < 1 + \frac{1}{n}$ pour n grand et $\sum_{n\geq 1} n^{-3/2} . (h(n))^{-3d-2} < \infty$.

<u>Démonstration</u> :

Elle s'inspire évidemment de celle du théorème 8 ; et nous allons simplement en rappeler les grandes lignes.

α) On pose $K_{t,n}^{11} = (h(n))^{-d} . K_{t,n}^{10}$, pour obtenir :

$$(\alpha_{t,n}^{10})^2 = (h(n))^{-2} . (\alpha_{t,n}^{9})^2 = O(n^{-2} . (h(n))^{-3d-2})$$

$$(\gamma_{t,n}^{10})^2 = O\{h(n)^{-2d} . (\gamma_{t,n}^{9})^2 + (h(n))^{-2d} . [(\frac{h(n)}{h(n+1)})^d - 1]^2 . (\alpha_{t,n}^{9})^2$$

K vérifiant (54).

Ceci implique que, en posant

$$A_n^{10} = \{\sum_1^n (K_{t,n}^{10}(X_t) - E^{F_{t-1}} |K_{t,n}^{10}(X_t)|\}_{L^2} = \sum_1^n Y_{t,n}$$

$$\begin{cases} \lim_{n\to\infty} E(h(n)^{-2d} . \|A_n^{10}\|_{L^2}^2) = 0 & \text{sous les conditions du a)} \\ \lim_{n\to\infty} \text{p.s. } (h(n)^{-2d} . \|A_n^{10}\|_{L^2}^2) = 0 & \text{sous les conditions du b)} \end{cases}$$

β) On en déduit alors la convergence de $\|A_n^{10}\|_{L^1}$ en la majorant par

$$\|I_{C_n} . A_n^{10}\|_{L^1} + \|I_{\bar{C}_n} . (\sum_{i\in I_n} Y_{t,n})\|_{L^1} + \|I_{\bar{C}_n} . (\sum_{i\in J_n} Y_{t,n})\|_{L^1}$$

avec $C_n = \left[-h(n)^{-2}, h(n)^{-2}\right]^{\otimes d}$ et $I_n = \{i \ / \ ||X_i|| \leqslant \frac{h(n)^{-2}}{2}\}$, puisque

* $||I_{C_n} A_n||_{L^1}$ se majore par $O\{h(n)^{-d} \cdot ||A_n||_{L^2}\}$ et on utilise α)

* $||I_{\bar{C}_n} \cdot (\sum_{t \in I_n} Y_{t,n})||_{L^1}$ se majore par
$(h(n))^{-1} \cdot \int_{L_n} \left[(D^i K)(u)\right] du$, où $L_n = \{u \ / \ ||u|| > \frac{h(n)^{-3}}{2}\}$,
donc par $h(n)^2 \int_{L_n} ||u|| \left[(D^i K)(u)\right] du$ qui vaut 0 pour n assez grand.

* $||I_{\bar{C}_n} \cdot (\sum_{t \in J_n} Y_{t,n})||_{L^1}$ se majore par

$$O\{\sum_{t \in J_n} n^{-1} \cdot \int_{\bar{C}_n} (h(n))^{-d-1} \{|D^i(K)| \ (\frac{y - X_t}{h(n)})\} \ dy\}$$

soit $O\{\sum_{t \in J_n} n^{-1} \cdot h(n) \cdot \int_{\mathbb{R}^d} ||X_t + uh(n)|| . |D^i(K)(u) \ | \ du \}$

soit enfin $O\{h(n)^2 \int_{\mathbb{R}^d} ||u|| \ |D^i K(u)| \ du\} + O\{ n^{-1} \cdot \sum_{t \in J_n} h(n) \cdot ||X_t||\}$

ce qui ramène la convergence à 0 de $||I_{\bar{C}_n} \cdot (\sum_{t \in J_n} (Y_{t,n}))||_{L^1}$ à celle de

$$(n^{-1}) \cdot \sum_{t=1}^{n} h(n) \cdot ||X_t|| \cdot I_{\{||X_t|| > \frac{h(n)^{-2}}{2}\}} \quad .$$

γ) Reste à démontrer que $||(n^{-1} \sum_{t=1}^{n} E^{F_{t-1}} (K_{t,n}^{10}(X_t)) - g||_{L^1} \longrightarrow 0$, si $n \to \infty$.

* K étant à support compact, on peut écrire que $\forall g$, $g \in W^{11}$:

$$\int_{\mathbb{R}^d} (D^i K)\ g\ d\lambda = \int (\tilde{D}^i g)\ K\ d\lambda \quad .$$

C_o^∞ étant dense dans W^{11} (cf. Adams (7) p. 56), on peut en effet reprendre la démonstration du lemme 5 , en utilisant une suite g_n d'éléments C_o^∞ convergeant vers g pour lesquels $\lim\limits_{n\to\infty} \int_{\mathbb{R}^d} |g - g_n|\ d\lambda = 0$, $\lim\limits_{n\to\infty} \int_{\mathbb{R}^d} |\tilde{D}^i g - \tilde{D}^i g_n|\ d\lambda = 0$, $\forall\ 0$, $= 1,\ldots,d$; et tels que :

$$0 \leq \overline{\lim_n} \left|\int_{\mathbb{R}^d} [(D^i K)\ g + (\tilde{D}^i g)\ K]\ d\lambda\right| \leq \lim_n \int_{\mathbb{R}^d} |D^i K|\ |g - g_n|\ d\lambda +$$

$$+ \lim_n \int_{\mathbb{R}^d} |K|\ |\tilde{D}^i - \tilde{D}^i g_n|\ d\lambda$$

d'où la conclusion , K étant à support compact et dans C^{11} , donc dans L^∞.

* Si nous définissons alors ponctuellement, pour $g \in W^{11}$, et $y \in \mathbb{R}^d$, l'opérateur $[K_{t,n}^{10}]$ comme dans la démonstration du théorème 9 , il est clair que la définition a un sens $(D^i K$ bornée, $g \in L^1)$, et qu'il coïncide avec $(n^{-1})\ |H_{h(n)}|\ |\tilde{D}^i(g)|$ pour tout g de W^{11} , d'après ce qui précède. On déduit ainsi la convergence de $(\sum\limits_1^n [K_{t,n}^{10}])$ vers $\tilde{D}^i$, et que $||[K_{t,n}^{10}]|| \leq (n^{-1})$ du lemme 7 , $[K_{t,n}^{10}]$ étant considéré comme opérateur de W^{11} dans L^1 .

* Reste enfin à voir que les $[K_{t,n}^{10}]$ vérifient la "condition I ", ce qui se ramène à vérifier l'égalité :

$$\int A(y)\ [\{[K_{t,n}^{10}](g)\}\ (y)]\ dy = \int \{\int A(y)\ [(K_{t,n}^{10}(x))(y)]\ dy\}\ g(x)\ dx$$

pour toute fonction A de L^∞, dual de L^1. Cette égalité résulte directement de l'intégrabilité de la fonction : $A(y)\ .\ (\tilde{D}^i K)\ [(y-x)\ /\ (h(n))]\ .\ g(x)$.

* Toutes les hypothèses requises sur les opérateurs $[K_{t,n}^{10}]$ étant vérifiées, on en conclut que

$$||(n^{-1}) \cdot \sum_{t=1}^{n} E^{\mathcal{F}_{t-1}}(K^{10}_{t,n}(X_t)) - g||_{L^1} = ||\sum_{t=1}^{n} [K^{10}_{t,n}]\, g_{X_t}^{\mathcal{F}_{t-1}} - g||_{L^1}$$

converge vers 0 , en moyenne ou p.s. lorsque $||(n^{-1}) \sum_{1}^{n} (g_{X_t}^{\mathcal{F}_{t-1}} - g_{X_t})||_{L^1}$ fait de même.

ω) Les points a) et b) du théorème découlent de ce qui précèdent. Pour le dernier point il suffit de voir que, d'après le lemme 9, si $\sum(n^{-1})\, h(n) < \infty$, M_n tend vers 0 en moyenne et p.s., lorsque les (X_t) sont intégrables et forment un processus strictement stationnaire, la routine habituelle ramenant les autres conditions sur $h(n)$ à celles de c).

Q.E.D.

THEOREMES DE CONVERGENCE PONCTUELLE

(Chapitre VI)

La convergence des estimateurs étudiés aux chapitres précédents a été obtenue dans l'espace $L^2(\mu)$ (éventuellement $L^1(\mu)$, moyennant un artifice de démonstration), ce qui n'a guère pour surprendre puisque les théorèmes du chapitre I concernent des éléments aléatoires à valeurs dans un Hilbert séparable. On aurait pu également considérer les estimateurs $\hat{f}_n^3$ et $\hat{f}_n^6$ (méthode de projection ou méthode du noyau) à valeurs dans W^{12}, qui répond aux mêmes critères, pour obtenir des théorèmes de convergences simultanées des estimateurs et de leurs dérivées vers les limites souhaitées g et $(D^i g)$, pour les "bons" (cf. lemme 4) processus ergodiques.

Le type de théorèmes obtenus ne résout évidemment pas le problème de la convergence au sens de L^∞ des estimateurs étudiés, et nous étudierons ici cette question. Les convergences ponctuelles de $\hat{f}_n^3$ et $\hat{f}_n^6$ vers la densité commune des X_t ont été déjà obtenues pour des processus strictement stationnaires et mélangeants (Bosq (25), Delecroix (14), Peligrad (26), Földes (27) Rosenblatt (28)) entre autres, mais jamais, à notre connaissance, pour des processus simplement ergodiques.

Nous démontrerons d'abord des théorèmes concernant des processus à valeurs réelles, tels que les dérivées des densités de transition existent, théorèmes déduits des résultats des chapitres précédents. Puis nous résoudrons le problème pour des processus à valeurs dans $\mathbb{R}^d$ tels que les densités $g_{X_t}^{\mathcal{F}_{t-1}}$ et g_{X_t}, $t \geq 1$, appartiennent à l'ensemble $C_o(\mathbb{R}^d)$ des fonctions continues sur $\mathbb{R}^d$ et tendant vers 0 à l'infini (c'est la fermeture dans L^∞ de $C(\mathbb{R}^d)$, elle constitue un espace de Banach séparable (cf. Rudin (8))).

En une deuxième partie, nous évoquerons le problème de la vitesse de convergence des estimateurs étudiés : montrant qu'une évaluation s'avère pratiquement impossible sous la seule hypothèse d'ergodicité, nous renverrons à notre article publié dans la revue de l'I.S.U.P. (1980, XXV, 1-2) pour une détermination dans le cas de processus mélangeants, cas qui contient en particulier celui de nombreux processus ARMA.

CONVERGENCES UNIFORMES SOUS DES HYPOTHESES D'ERGODICITE

(C.VI.I)

(A) - LE PROCESSUS EST A VALEURS REELLES

Les résultats qui suivent se basent sur le

Lemme 10.- Soit (f_n) , $n \geq 1$, et f des fonctions réelles de variables réelles, éléments de l'espace C^{11} (resp. C^{12}) , tels que $\lim_{n\to\infty} (\|f_n-f\|_{W^{11}}) = 0$ (resp. $\lim_{n\to\infty} (\|f_n-f\|_{W^{12}}) = 0$) .

Alors $$\lim_n \sup_{x\in\mathbb{R}} (|f_n(x) - f(x)|) = 0$$

Démonstration :

Soit $\delta > 0$, pour tout x réel on a l'identité :

$$f_n(x) - f(x) = \frac{1}{2\delta} \int_{x-\delta}^{x+\delta} [(f(t)-f(x)) - (f_n(t)-f_n(x))]\, dt$$
$$+ \frac{1}{2\delta} \int_{x-\delta}^{x+\delta} [f_n(t)-f(t)]\, dt$$

or $$(f(t)-f(x))- (f_n(t)-f_n(x)) = \int_x^t (f'(y)-f_n'(y))\, dy \quad ,$$ d'où :

* Si les f_n et f sont dans C^{11} :

$$\sup_{x\in\mathbb{R}} (|f_n(x)-f(x)|) \leq \|f_n'-f'\|_{L^1} + \frac{1}{2\delta} \|f_n-f\|_{L^1} \quad ,$$

* Si les f_n et f sont dans C^{12} :

$$\sup_{x\in\mathbb{R}} (|f_n(x)-f(x)|) \leq \|f_n'-f'\|_{L^2} \{\frac{1}{2\delta} \int_{x-\delta}^{x+\delta} \sqrt{|x-t|}\, dt\} + \|f_n-f\|_{L^2} \{\frac{\sqrt{2\delta}}{2\delta}\}$$

La conclusion est alors immédiate.

On en déduit le corollaire suivant :

Corollaire 2.- Soit (X_t) un processus strictement stationnaire à valeurs réelles

a) Si les $g_{X_t}^{F_{t-1}}$ sont, comme g, dans C^{12} et que

$$\lim_{n\to\infty} ||(n^{-1}) \sum_1^n g_{X_t}^{F_{t-1}} - g||_{W^{12}} = 0$$

en moyenne, et p.s., en choisissant le noyau K dans C^{12} tel que (41), (42), (54) soient vérifiées, et $(h(n))_{n\geq 1}$ telle que $\sum_1^\infty (n^{-3/2})\ h(n)^{-3} < \infty$, on aura la convergence à 0 de $\sup_{x\in\mathbb{R}} |\hat{f}_n^6(x)-g(x)|$, en moyenne, et p.s., si $n \to \infty$.

b) Si K est dans C^{11}, à support compact et vérifiant (41), (42), (54), et les h(n) tels que $h(n) = n^{-\alpha}$, $0 < \alpha < \frac{1}{10}$, on aura la même conclusion si les $g_{X_t}^{F_{t-1}}$ sont dans C^{11}, comme g, et tels que $\lim_{n\to\infty} ||(n^{-1}) \sum_1^n g_{X_t}^{F_{t-1}} - g||_{W^{11}} = 0$ en moyenne et p.s., et les (X_t) intégrables.

Ce corollaire découle directement des théorèmes 7 à 10 : ils montrent que sous les conditions apportées en a) et b), on a $||\hat{f}_n^6-g||_{W^{12}}$ et $||\hat{f}_n^6-g||_{W^{11}}$ qui tendent vers 0 en moyenne et p.s. , quand $n \to \infty$.

(B) - CAS GENERAL

Nous aurons besoin du lemme suivant :

Lemme 11.- Considérons l'opérateur $[H_\beta]$ défini ponctuellement au lemme 7, à partir d'une densité K bornée. Alors $[H_\beta]$ est un opérateur bien défini de $C_o(\mathbb{R}^d)$ dans $C_o(\mathbb{R}^d)$, de norme inférieure à 1. De plus $[H_\beta]$ converge fortement vers l'identité de $C_o(\mathbb{R}^d)$, si β tend vers 0 , et pour toute fonction g de $L^1(\mu) \cap C_o(\mathbb{R}^d)$, $[H_\beta](g)$ est égale à l'intégrale de Bochner : $\int \beta^{-d} . K((\cdot-x)/\beta)\ g(x)\ d\mu(x)$.

Démonstration :

α) Toute fonction g de $C_o(\mathbb{R}^d)$ étant bornée, on peut pour tout y calculer $[H_\beta](g)\ (y)$ car $K \in L^1$. De plus la fonction obtenue $[H_\beta](g)$ est dans $C_o(\mathbb{R}^d)$. En effet

* $[H_\beta](g)$ est continue : si x_k est une suite d'éléments de $\mathbb{R}^d$ tendant vers x , on aura, pour tout ε :

$$|([H_\beta](g))(x_k) - ([H_\beta](g))(x)| = \left|\int K(u)\ [g(x_k-u\beta)-g(x-u\beta)]\ du\right|$$

$$\leq \varepsilon \int |K(u)|\ du$$

dès que k est assez grand car g est uniformément continue (cf. Taylor R.L. et Tien-Chien-Hu (34)).

* $\lim_{||x||\to\infty} ([H_\beta](g))(x) = 0$, puisque si x_k est une suite d'éléments tels que $\lim_{k\to\infty} ||x_k|| = \infty$, on aura :

$$\lim_{k\to\infty} [([H_\beta](g))(x_k)] = \lim_k \int K(u) \,.\, g(x_k - \beta u)\, du \quad .$$

En posant $f_k(u) = g(x_k - \beta u) \,.\, K(u)$, on voit que les $|f_k|$ sont bornées par $C.|K|$, car g est bornée, et que $\lim_k f_k = 0$. Une application directe du théorème de Lebesgue permet de conclure.

β) L'inégalité

$$\sup_x |([H_\beta](g))(x)| = \sup_x \left|\int K(u) \,.\, g(x-u\beta)\, du\right|$$

$$\leq ||g||_{L^\infty} \{\int |K(u)|\, du$$

montre que $||[H_\beta]|| \leq 1$.

γ) Si $\beta \downarrow 0$, $[H_\beta]$ converge vers $Id_{C_o(\mathbb{R}^d)}$. En effet, pour tout δ

$$||[H_\beta]g-g||_{C_o(\mathbb{R}^d)} = \sup_{x\in\mathbb{R}^d} \left|\int \frac{1}{\beta^d} K(\tfrac{u}{\beta}) \,[g(x-u) - g(x)]\, du\right|$$

$$\leq \sup_{x\in\mathbb{R}^d} [\{ \sup_{||u||<\delta} |g(x-u)-g(x)|\} \,.\, \int_{||u||<\delta} \frac{1}{\beta^d} K(\tfrac{u}{\beta})\, du]$$

$$+ 2\, ||g||_{C_o(\mathbb{R}^d)} \,.\, \left|\int_{||u||>\delta} \frac{1}{\beta^d} K(\tfrac{u}{\beta})\, du\right|$$

L'uniforme continuité de g montre que $\forall\varepsilon$, $\exists\, \delta$:

$$||[H_\beta]\; g-g||_{C_o(\mathbb{R}^d)} \leq \varepsilon \,.\, + 2\, ||g||_{C_o(\mathbb{R}^d)} \,.\, \int_{||y||>\delta/\beta} K(y)\, dy \quad .$$

D'où la conclusion.

ω) Si $g \in L^1(\mu) \cap C_o(\mathbb{R}^d)$, l'intégrale $\int \beta^{-d} K[(\cdot-x)/\beta]\, g(x)\, d\mu(x)$ permutant avec les formes linéaires $f \to f(y)$ pour tout y de $\mathbb{R}^d$, coïncide avec $[H_\beta](g)$.

Q.E.D.

Nous pouvons alors démontrer le :

Théorème 11.- Considérons un processus (X_t) à valeurs dans $\mathbb{R}^d$, tel que $\sup_t \{E(||X_t||^2)\} < \infty$, et que les densités $g_{X_t}^{\mathcal{F}_{t-1}}$ et g_{X_t} soient dans $C_o(\mathbb{R}^d)$. On considère l'estimateur $\hat{f}_n^6$, pour un noyau K borné, et lipschitzien, et une suite $h(n)$ décroissante vers 0, $n\,h(n)$ étant croissante, et vérifiant

$$* \qquad 1 - \left(\frac{h(n)}{h(n-2\sqrt{n})}\right)^d = O\left(\frac{1}{\sqrt{n}}\right) ,$$

$$* \ \exists \varepsilon > 0 , \quad n^{sd-0,5}\,(h(n^2))^{-2d^2-3d} = O(n^{-\varepsilon}) , \quad \text{pour un } s > 0$$

Alors si $\lim_{n\to\infty} ||\frac{1}{n}\sum_1^n g_{X_t} - g||_{C_o(\mathbb{R}^d)} = 0$ et que de même

$||\frac{1}{n}\sum_1^n (g_{X_t}^{\mathcal{F}_{t-1}} - g_{X_t})||_{C_o(\mathbb{R}^d)}$ tend vers 0, p.s. (resp. en moyenne),

$\sup_{x\in D_n} |\hat{f}_n^6(x) - g(x)|$ tend aussi vers 0 p.s. (resp. en moyenne), D_n représentant le pavé $[-n^s , n^s]^d$

Démonstration :

α) En posant

$$Y_{t,n}(x) = (n^{-1})\,(h(n))^{-d} \,.\, \{K((x-X_t)/h(n)) - E^{\mathcal{F}_{t-1}}[K(x-X_t)/h(n)]\}$$

nous obtenons la décomposition usuelle :

$$\sup_{x\in\mathbb{R}^d} |\hat{f}_n^6(x)-g(x)| \leq \sup_{x\in\mathbb{R}^d} |\sum_{t=1}^n Y_{t,n}(x)| + ||\sum_{t=1}^n [L_{t,n}]\,(g_{X_t}^{\mathcal{F}_{t-1}}) - g||_{C_o(\mathbb{R}^d)}$$

en posant $L_{t,n} = n^{-1} . [H_{h(n)}]$, l'utilisation de ces opérateurs étant justifiée au lemme précédent, lemme montrant également que les $[L_{t,n}]$ vérifient les conditions usuelles (H4 à H7). On en déduit que le $2^{\text{ème}}$ majorant, ci-dessus, tend vers 0 en moyenne et p.s. sous les hypothèses générales sur le processus (X_t) introduite dans le théorème.

β) Posons alors $A_n = \sup_{x \in D_n} |\sum_1^n Y_{t,n}(x)|$, où D_n est le pavé $[-\xi_n , \xi_n]^d$, ξ_n étant définie par : $\xi_n = (E[\frac{n^s}{(h(n))^{d+2}}]) . (h(n))^{d+2}$

(E s'entendant ici comme "partie entière"). On peut décomposer D_n en $m(n) = (2E[\frac{n^s}{(h(n))^{d+2}}])^d$ pavés disjoints dont l'union recouvre D_n , et tel que chacun s'écrive $\prod_{i=1}^d a_i$, a_i étant un segment de longueur $[(h(n))^{d+2}]$. Appelons $t_1^n , \ldots , t_{m(n)}^n$ les centres de ces pavés. Tout x de D_n est dans un pavé élémentaire et un seul, et se trouve à une distance inférieure à $\sqrt{d}\,(h(n))^{d+2}$ de son centre, $t_i^n(x)$. Dès lors :

$$|\sum_1^n Y_{t,n}(x)| \leq |\sum_1^n Y_{t,n}(t_i^n(x))| + \sup_{\substack{x,y \in D_n \\ ||x-y|| \leq \sqrt{d}(h(n))^{d+2}}} |\sum_1^n [Y_{t,n}(x) - Y_{t,n}(y)]|$$

Mais K est lipschitzienne, donc :

$$|K(\frac{y-X_t}{h(n)}) - K(\frac{x-X_t}{h(n)})| \leq A \, ||\frac{y-x}{h(n)}||$$

et de même

$$|E^{\mathcal{F}_{t-1}}(K(\frac{y-X_t}{h(n)})) - E^{\mathcal{F}_{t-1}}(K(\frac{x-X_t}{h(n)})| \leq A' \, ||\frac{y-x}{h(n)}||$$

Dès lors :

$$\sup_{\substack{(x,y)\in D_n \\ ||x-y||\leqslant\sqrt{d}(h(n))^{d+2}}} \left|\sum_1^n (Y_{t,n}(x) - Y_{t,n}(y))\right| = \bigcirc(h(n))$$

et $$A_n = \sup_{1\leqslant i\leqslant m(n)} \left|\sum_1^n Y_{t,n}(t_i^n)\right| + \bigcirc(h(n))$$

D'après α) et ce qui précède il nous reste donc finalement à prouver que :

$$\sup_{1\leqslant i\leqslant m(n)} \left|\sum_1^n Y_{t,n}(t_i^n)\right| \longrightarrow 0$$

γ) Pour $x \in \mathbb{R}^d$, calculons $E((\sum_1^n Y_{t,n}(x))^2)$ soit

$$E\{\frac{1}{n^2}\left[\sum_{t=1}^n \frac{1}{h(n)^d} \cdot \{K(\frac{x-X_t}{h(n)}) - E^{F_{t-1}}\left[K(\frac{x-X_t}{h(n)})\right]\}\right]^2\}$$

L'orthogonalité des variables considérées donne :

$$E([\sum_1^n Y_{t,n}(x)]^2) \leqslant \frac{1}{n^2(h(n))^d} \sum_{t=1}^n \int K^2(u) \cdot g_{X_t}(x + u(h(n)) du$$

(55)
$$= \bigcirc(n^{-1}\ h(n)^{-d}) \quad , \qquad K \text{ étant bornée .}$$

* On en déduit en premier lieu que $E(\sup_{1\leqslant i\leqslant m(n)} |\sum_{t=1}^n Y_{t,n}(t_i^n)|)$ tend vers 0 sous les conditions du théorème : cette quantité se majore en effet par

$$\bigcirc\left[m(n) \cdot (n^{-1} \cdot (h(n))^{-d})^{1/2}\right] ,$$

soit $$\bigcirc\left[n^{sd-1/2} \cdot h(n)^{-d^2-(5/2)d}\right].$$

* On en déduit ensuite que $P[\{\sup_{1\le i\le m(n)} |\sum_{1}^{n} Y_{t,n}(t_i^n)|\} > \varepsilon\}$ se majore par :

(56) $$\bigcirc\{m(n)\ .\ n^{-1}\ .\ (h(n))^{-d}\} = \bigcirc\{n^{sd-1}\ .\ h(n)^{-d^2-3d}\}\ ,$$

ce qui permet de passer à la suite

ζ) On peut majorer $\sup_{1\le i\le m(n)} |\sum_{t=1}^{n} Y_{t,n}(t_i^n)|$ par la somme des 3 termes :

$$T_1^n = \sup_{1\le i\le m(n)} |\sum_{t=1}^{\Phi(n)} (Y_{t,n}(t_i^n) - Y_{t,\Phi(n)}(t_i^n))|$$

$$T_2^n = \sup_{1\le i\le m(n)} |\sum_{t=1}^{\Phi(n)} Y_{t,\Phi(n)}(t_i^n)|$$

$$T_3^n = \sup_{1\le i\le m(n)} |\sum_{t=\Phi(n)+1}^{n} Y_{t,n}(t_i^n)|$$

$\Phi(n)$ representant le plus grand carré entier inférieur ou égal à n.

Alors on aura

i) $$P(T_3^n > \varepsilon) \le m(n)\{\sup_{1\le i\le m(n)} P(|\sum_{t=\Phi(n)+1}^{n} Y_{t,n}(t_i^n)| > \varepsilon)\}$$

$$=\bigcirc\{m(n)\ .\ (n-\Phi(n))\ .\ n^{-2}\ .\ h(n)^{-d}\}$$

en reprenant le raisonnement de γ)

$$=\bigcirc(n^{sd-3/2}\ .\ h(n)^{-d^2-3d})$$

car $n - \Phi(n) < 2\sqrt{n}$. Donc T_3^n tend vers 0 p.s. si $n \to \infty$, sous les conditions du théorème.

ii) On peut majorer T_2^n par $T_4^n + T_5^n$, où

$$T_4^n = \sup_{x\in D_{\Phi(n)}} \left| \sum_{t=1}^{\Phi(n)} Y_{t,\Phi(n)}(x) \right| \quad \text{et} \quad T_5^n = \sup_{i\in I_n} \left| \sum_{t=1}^{\Phi(n)} Y_{t,\Phi(n)}(t_i^n) \right| ,$$

I_n représentant l'ensemble des indices i tels que $t_i^n \notin D_{\Phi(n)}$. Alors

* T_4^n tend p.s. vers 0 : cette suite prend la même valeur quand n varie entre 2 carrés parfaits , et il ne reste donc à vérifier que la convergence p.s. à 0 de la suite $A_{(n^2)}$, c'est-à-dire celle de

$$\sup_{1\leqslant i\leqslant m(n^2)} \left| \sum_{t=1}^{n^2} Y_{t,n^2}(t_i^{n^2}) \right|$$

ce qui découle directement de la majoration (56), qui montre que $\forall\varepsilon$, $\varepsilon > 0$

$$\sum_{n\geqslant 1} P\left[\left\{ \sup_{1\leqslant i\leqslant m(n^2)} \left| \sum_{1}^{n^2} Y_{t,n^2}(t_i^{n^2}) \right| > \varepsilon\right\}\right] = O\left(\sum_{n\geqslant 1} n^{2sd-2}\, h(n^2)^{-d^2-3d}\right)$$

convergente sous les hypothèses du théorème.

* Pour T_5^n , on remarque que, puisqu'il y a $2^d(E[\frac{n^s}{h(n)^{d+2}}])^d$ t_i^n dans D^n , et $2^d(E[\frac{\Phi(n)^s}{h(n)^{d+2}}])^d$ dans $D_{\Phi(n)}$, $I(n)$ comprendra au plus $E\{2^d[n^{sd} - (\Phi(n))^{sd}]\, h(n)^{-d^2-2d}\}^{+1}$ éléments, et dès lors : $P[T_5^n > \varepsilon]$ se majorera par $O\{n^{sd}\, [1 - (\frac{\Phi(n)}{n})^{sd}] \,.\, h(n)^{-d^2-2d} \,.\, (\Phi(n))^{-1} \,.\, (h(\Phi(n))^{-d}\}$ d'après (55).

Comme $\Phi(n) > n - 2\sqrt{n}$, on a : $(1 - (\frac{\Phi(n)}{n})^{sd}) = O(\frac{1}{\sqrt{n}})$ et $n/\Phi(n) < \frac{n}{n-2\sqrt{n}}$ reste bornée, pour n assez grand, $(n > 4)$ d'où

$$P[T_4^n > \varepsilon] = O(n^{sd-3/2}\, h(n)^{-d^2-3d})$$

ce qui assure la convergence p.s. à 0 de T_4^n sous les conditions du théorème.

iii) Pour T_1^n , nous aurons besoin de majorer :

$$E\{(\sum_{1}^{\Phi(n)} (Y_{t,n}(x) - Y_{t,\Phi(n)}(x)))^2\} \quad ,$$

qui, de par l'orthogonalité des $Y_{t,n}$ et les propriétés standard de l'espérance conditionnelle se majore par

$$\Phi(n) \int \left[\frac{1}{n(h(n))^d} K(\frac{x-u}{h(n)}) - \frac{1}{\Phi(n)[h(\Phi(n))]^d} K(\frac{x-u}{h(\Phi(n))})\right]^2 g_{X_t}(u)\, du$$

soit encore :

$$\mathcal{O}\{\Phi(n) (\frac{1}{n(h(n))^d})^2 . \int \left[K(\frac{x-u}{h(n)}) - K(\frac{x-u}{h(\Phi(n))})\right]^2 g_{X_t}(u)\, du$$

$$+ \quad \mathcal{O}\{\Phi(n) \left[\frac{1}{n[h(n)]^d} - \frac{1}{\Phi(n)[h(\Phi(n))]^d}\right]^2 \int K^2 (\frac{x-u}{h(\Phi(n))}) g_{X_t}(u)\, du\}$$

* Le premier majorant peut s'écrire

(57) $\mathcal{O}\{\Phi(n) . n^{-2} . h(n)^{-2d} . (\frac{1}{h(n)} - \frac{1}{h(\Phi(n))})^2 . \int ||x-u||^2 g_{X_t}(u)\, du\}$

soit, les X_t vérifiant $\sup_{t\in\mathbb{Z}} \{E(||X_t||^2)\} < \infty$, si $x \in D_n$ $(||x|| < \sqrt{d} . n^s)$

$$\mathcal{O}\{\frac{\Phi(n)}{n} . n^{-1+2s} . h(n)^{-2d-2} (1 - \frac{h(n)}{h(\Phi(n))})^2\} = \mathcal{O}\{n^{-2+2s} . h(n)^{-2d-2}\} ,$$

puisque nous avons supposé $1 - \frac{h(n)}{h(\Phi(n))} < 1 - \left[\frac{h(n)}{h(n-2\sqrt{n})}\right]^d = \mathcal{O}(\frac{1}{\sqrt{n}})$

* Pour le second, on obtient par l'intégration habituelle :

$$\mathcal{O}\{(\Phi(n))^{-1} (h(\Phi(n)))^{-d} . [1 - \frac{\Phi(n) [h(\Phi(n))]^d}{n (h(n))^d}]^2\}$$

et $n(h(n))$ étant supposée croissante, on a :

$$1 - \frac{\Phi(n)\ (h(\Phi(n)))^d}{n\ (h(n))^d} \leq 1 - \frac{(n-2\sqrt{n})\ (h(n-2\sqrt{n}))^d}{n\ (h(n))^d}$$

$$\leq \frac{h(n-2\sqrt{n}))^d}{(h(n))^d} \left(\frac{h(n)^d}{h(n-2\sqrt{n})^d} - 1\right) + \frac{2}{\sqrt{n}} \frac{(h(n-2\sqrt{n}))^d}{h(n)^d}$$

Mais $\left[\dfrac{h(n-2\sqrt{n})}{h(n)}\right]^d$ est borné, ce qui donne finalement comme majoration du 2ème terme :

$$\bigcirc\{(\Phi(n))^{-1}\ [h(\Phi(n))]^{-d}\ \frac{1}{n}\} = \bigcirc\left(\frac{1}{n(n-2\sqrt{n})}\ \frac{1}{h(n)^d}\right) = \bigcirc(n^{-2}\ h(n)^{-d})$$

Finalement, nous aurons donc :

$$P(T_1^n > \varepsilon) = \bigcirc(m(n)\ .\ n^{-2+2s}\ h(n)^{-2d-2})$$

$$= \bigcirc(n^{sd-2+2s}\ .\ h(n)^{-d^2-4d-2})$$

ce qui montre la convergence presque sûre à 0 de T_1^n sous les conditions du théorème.

λ) Des points précédents découle que $\sup\limits_{x\in D_n} |\hat{f}_n^6(x) - g(x)| \longrightarrow 0$ p.s. et en moyenne si $n \to \infty$.

Q.E.D.

Le théorème précédent peut être amélioré si l'on impose des conditions supplémentaires au processus (X_t) , et au noyau (K). Cela donne :

Corollaire 3.- Supposons que les (X_t) forment un processus strictement stationnaire vérifiant les conditions du théorème 1.1., et tel que de plus existe un nombre p , $p > 2/s$, tel que $E(||X_t||^p) < \infty$. Supposons de plus que K est une densité bornée nulle en dehors d'un compact. Alors :

$$\lim_{n\to\infty} \text{(p.s.)} \left(\sup_{x\in\mathbb{R}^d} |\hat{f}_n^6(x) - g(x)| \right) = 0 \quad .$$

Si $d = 1$, le choix usuel $h(n) = (\text{Log } n)^{-1}$ assure la convergence si les X_t ont un moment d'ordre 5.

Démonstration :

α) Posons $\bar{X}_t^n = X_t \cdot I_{\{||X_t|| < \sqrt{d}(n^s-(h(n))^2)-Ah(n)\}}$,

la constante A étant déterminée par l'implication

$$(||\frac{x - y}{h(n)}|| > A) \Longrightarrow (K(\frac{x - y}{h(n)}) = 0) \quad .$$

Posons

$$\bar{Y}_{n,t}(.) = (n^{-1}) \cdot [h(n)]^{-d} \cdot \{K(\frac{\cdot - \bar{X}_t^n}{h(n)}) - E^{\mathcal{F}_{t-1}} [K(\frac{\cdot - \bar{X}_t^n}{h(n)})]\} ,$$

et $R_n = \sum_{t=1}^{n} \bar{Y}_{n,t}$. Alors si x vérifie $||x|| > \sqrt{d}\ (n^s - (h(n))^2)$, on a pour tout t , $1 \leqslant t \leqslant n$:

$$||x - \bar{X}_t^n|| \geqslant ||x|| - ||\bar{X}_t^n|| \geqslant Ah(n)$$

donc $R_n(x)$ vaudra 0. C'est donc vrai si $x \notin D_n$ puisque si $x \in D_n$, on a $||x|| \leqslant \sqrt{d}\ (n^s - (h(n))^2)$ (cf. la démonstration du théorème 11).

β) Si nous posons $T'_n = \sup_{x\in\mathbb{R}^d} |R_n(x)|$, la convergence p.s. à 0 de T'_n dépend donc de celle de $\sup_{x\in D_n} |R_n(x)|$. Mais celle-ci découle immédiatement de la démonstration du théorème 11 : il suffit de la reprendre pour voir que (55) reste vraie en remplaçant les g_{X_t} par les $g_{\bar{X}_t^n}$ et de même pour (56) car $\int ||x-u||^2\ g_{\bar{X}_t^n}(u)\ du$ se majore ici encore par $(||x||^2 + \sup_t E(||\bar{X}_t^n||^2))$ soit a fortiori par $(||x||^2 + \sup_t E(||X_t||^2))$ ...

Finalement, sous les conditions du théorème : $\lim_{n\to\infty}$ p.s. $(T'_n) = 0$.

γ) Si nous posons $T_n = \sup_{x\in\mathbb{R}^d} \left| \sum_{t=1}^{n} Y_{t,n}(x) \right|$, il suffit finalement de montrer que $T_n - T'_n \longrightarrow 0$, p.s., si $n \to \infty$, pour obtenir la convergence souhaitée, c'est-à-dire qu'il suffit de prouver :

$$\sum_{n\geq 1} P(|T_n - T'_n| > \varepsilon) < \infty , \qquad \forall \varepsilon > 0 .$$

Mais $\{T_n = T'_n\} \supset \bigcap_{t=1}^{n} \{X_t = \bar{X}_t^n\}$ puisque T_n se définit à partir des X_t, T'_n à partir des $\bar{X}_t^n$. Il vient :

$$\sum_n P(\{|T_n - T'_n| > \varepsilon\}) \leqslant \sum_n P\left[\bigcup_{t=1}^{n} \{||X_t|| > \sqrt{d}\,(n^s - h^2(n)) - Ah(n)\}\right]$$

$$= \mathcal{O}\left\{\sum_n \, . \, n \, . \, \frac{E(||X_t||^p)}{\left[\sqrt{d}(n^s - h^2(n)) - Ah(n)\right]^p}\right\}$$

$$= \mathcal{O}\left\{\sum_n n^{-sp+1}\right\}$$

et on conclut puisque $-sp+1 < -1$ sous l'hypothèse

ω) Si $d = 1$, $h(n) = (\mathrm{Log}\, n)^{-1}$, $n^{sd-1/2}(h(n^2))^{-2d^2-3d}$ est $\mathcal{O}(n^{-\varepsilon})$ pourvu que $sd - \frac{1}{2} < 0$, soit $s < \frac{1}{2}$, $d > 2/s$ s'obtient alors pour $p = 5$ ...

Q.E.D.

VITESSES DE CONVERGENCE SOUS DES HYPOTHESES DE MELANGE

(C.VI.II)

Les techniques utilisées depuis le départ de ce travail ne permettent pas l'obtention de vitesses de convergence pour 2 raisons complémentaires :

a) "Il n'y a pas de vitesse de convergence dans le théorème ergodique" Krengel dixit (4) (p. 14), donc il n'y a pas de vitesse de convergence (type loi du log - itéré etc...) de la quantité $||(n^{-1}) \sum_{1}^{n} g_{X_t}^{F_{t-1}} - g||$, qui peut tendre vers 0 à une vitesse arbitraire (même référence). De plus, on ne peut même pas déduire cette vitesse des coefficients de mélangeance plus forts utilisés usuellement.

b) Les démonstrations des lemmes de base utilisés (lemmes 1 et 2), ne reposant pas sur des majorations de l'écart entre la suite étudiée et la limite, on ne peut là encore obtenir l'évaluation d'une vitesse dans les convergences qu'ils induisent.

Le problème se retrouve dans ce chapitre quoiqu'ici on ait pu exhiber une majoration exploitable de $P\left[\sup_{t \in \mathbb{R}^d} |\sum_{1}^{n} Y_{t,n}(x)| > \varepsilon\right]$.
Aussi peut-on proposer une autre approche du problème basée sur des hypothèses différentes, la mélangeance du processus étant essentiellement requise.

Plutôt qu'une ré-écriture de résultats acquis pour un processus réel à temps continu, nous proposons donc l'article que nous avons publié sur le sujet dans la revue de l'I.S.U.P. sous sa forme originale. On y trouvera, pour une classe d'estimateurs f_n comprenant ceux de la densité et du noyau (y compris les versions récursives) des majorants de l'erreur quadratique ponctuelle intégrée : $E\left[(f_n(x) - g(x))^2\right]$ et de la probabilité,

$P(|f_n(x) - g(x)| > \varepsilon)$, suffisants à entraîner les convergences, et dépendant des coefficients de mélangeance ... Rappelons que l'ordre de ceux-ci a été de surcroît déterminé pour quelques processus classiques (voir par exemple Wither (10) pour les processus linéaires).

L'article annoncé a été reproduit aux pages suivantes.

Pub. Inst. Stat. Univ. Paris
1980, XXV, fasc. 1-2, 17-39

SUR L'ESTIMATION DES DENSITES D'UN PROCESSUS STATIONNAIRE A TEMPS CONTINU

par

M. Delecroix

Université des Sciences et Techniques de Lille

U.E.R. de Mathématiques pures et appliquées

59655 - Villeneuve d'Ascq Cedex

Summary :

Density estimation for a continuous parameter and stationary process. Considering such a process X_t, we obtain pointwise consistency (a.s. and L^2) when estimating the density of X_t, the one of (X_t, X_{t+h}), and the conditional density of X_{t+h}/X_t, under additional mixing conditions. The class of estimates includes all the classical examples used in density estimation.

Introduction.

Depuis vingt ans environ, de nombreux travaux ont été consacrés à l'estimation non paramétrique d'une densité de probabilité, à partir d'une suite d'observations indépendantes. Sans vouloir prétendre à l'exhaustivité, on citera ici les articles récents de Bleuez et Bosq (1), Bertrand-Retali (2), P. Deheuvels (3), J. Geffroy (4), en renvoyant d'autre part à la synthèse bibliographique établie sur le sujet par Fryer (5).

Certains des résultats obtenus ont été généralisés dans le cas où les variables observées ne sont plus indépendantes par Bosq (6), Földes (7), Rosenblatt (8), Roussas (9), puis par Banon (10) et Nguyen (11).

Dans tous ces travaux, les variables observées forment un processus stationnaire au sens strict. Cette hypothèse est conservée dans ce qui suit, puisqu'elle assure l'existence d'une répétition d'observations sur de mêmes densités à estimer. La convergence des estimateurs utilisés est d'autre part assurée soit par une hypothèse de "ψ-mélangeance" sur le processus observé ((6) et (7)), soit par une hypothèse markovienne assortie de conditions supplémertaires, comme la condition de Doeblin ((9) par exemple). En fait, les deux types d'hypothèse impliquent la non corrélation asymptotique du processus ("L^2-mixing").

Dans ce travail, on n'imposera de conditions qu'au coefficient introduit par Rosenblatt (12), qui mesure la "mélangeance" du processus observé, mais se trouve être le plus faible de tous ceux qu'on utilise usuellement. En ce sens, on généralise tous les travaux sus-cités, les résultats obtenus concernant une très large classe de processus.

De même, la classe d'estimateurs utilisée regroupe les estimateurs définis à la fois par la méthode du "noyau", et celle des "fonctions orthogonales", et même pour certains résultats, les estimateurs splines.

Enfin, on supposera que le processus observé est à temps continu, problème moins étudié jusqu'ici. Les résultats obtenus se transcrivent aisément dans le cas où l'on dispose d'une <u>suite</u> de variables aléatoires observées.

Le travail qui suit consiste en l'obtention de conditions suffisantes de convergence ponctuelle en moyenne quadratique et presque sure pour les estimateurs des densités marginales et de transition du processus observé.

I. *Notations* - *Définitions des estimateurs.*

Comme annoncé ci-dessus, on considère une fois pour toutes un processus X_t, $t \in [0,\infty[$, chaque X_t étant une variable aléatoire réelle définie sur un espace probabilisé $(\Omega, \mathcal{A}, P)$, tel que :

1) (stationnarité) $\forall h \geq 0$, $\forall t_1,\dots,t_k$, k réels positifs distincts, la loi du k-uple $(X_{t_1},\dots,X_{t_k})$ est la même que celle de $(X_{t_1+h},\dots,X_{t_k+h})$

2) (mesurabilité) L'application $(t,\omega) \to X_t(\omega)$ est mesurable sur $(\mathbb{R}^+, \mathcal{B}_{\mathbb{R}^+}) \otimes (\Omega, \mathcal{A})$, à valeurs dans $(\mathbb{R}, \mathcal{B}_{\mathbb{R}})$

3) Il existe une mesure σ-finie, μ, sur $(\mathbb{R}, \mathcal{B}_{\mathbb{R}})$ telle que pour tout n de $\mathbb{N}^*$, la probabilité image de P par le vecteur $(X_{t_1}, \ldots, X_{t_n})$ soit dominée par $\mu^{\otimes n}$.

On appelle f et g_h $(h > 0)$ deux versions strictement positives de $\frac{dP_{X_t}}{d\mu}$ et $\frac{dP_{(X_t, X_{t+h})}}{d(\mu \otimes \mu)}$, respectivement, et g_h^a une version de la densité conditionnelle de X_{t+h} conditionnée par $X_t = a$, définie par :
$g_h^a(x) = \frac{g_h(a,x)}{f(a)}$.

Le problème est d'estimer ces trois densités à partir d'un échantillon de (X_t) observé sur $[0,T]$...

En appelant M_0^T et M_{T+h}^{∞} les tribus respectivement engendrées par $\{X_t,\ 0 \leq t \leq T\}$ et $\{X_t,\ t \geq T+h\}$ $(h \geq 0)$, on pose :

$$(1) \qquad \alpha(h) = \sup |P(A \cap B) - P(A)\,.\,P(B)| \,,$$

le sup étant pris sur tous les évènements A de M_0^T et B de M_{T+h}^{∞}.

La classique "strong mixing condition" de Rosenblatt (12) s'écrit :

$$\lim_{h \to \infty} \alpha(h) = 0 \,.$$

Nous aurons à imposer des conditions à la fonction α pour obtenir la convergence des estimateurs de f, g_h, g_h^a.

Introduisons trois familles $K^i_{(s,T)}$ $i = 0,1,2$, indicées dans $A = \{(s,T) | 0 \leq s \leq T\}$, de fonctions définies sur $\mathbb{R}^2$, à valeurs réelles, bi-mesurables, et définissons les estimateurs $\hat{f}_T$, $(\hat{g}_h)_T$, $(\hat{g}_h^a)_T$ de f, g_h, g_h^a, aux points t et (t,t') $(t \in \mathbb{R},\ t' \in \mathbb{R})$, par :

(2) : $$(\hat{f_T})(t) = (T^{-1}) \cdot \int_0^T K^o_{(s,T)}(X_s,t)ds$$

(3) : $$(\hat{g_h})_T(t,t') = (T^{-1}) \cdot \int_0^{T-h} K^1_{(s,T)}(X_s,t)K^2_{(s,T)}(X_{s+h},t')ds$$

(4) : $$(\hat{g^a_h})_T(t) = \left[(\hat{g_h})_T(a,t)\right] \cdot \left[(\hat{f_T})(a)\right]^{-1} \cdot I_{\{(\hat{f_T})(a) \neq 0\}}$$

Nous allons montrer que les estimateurs usuels de la densité s'écrivent sous la forme précédente, en fournissant des familles K^i qui vérifient les conditions fondamentales suivantes :

H_1) $(\hat{f_T})(t)$ et $(\hat{g_h})_T(t,t')$ sont des estimateurs asymptotiquement sans biais de $f(t)$ et $g_h(t,t')$,

H_2) Il existe six fonctions réelles à valeurs positives A_i, $i = 0,1,2$ et h_i, $i = 0,1,2$, telles que $\forall x, x \in \mathbb{R}$, $\forall t, t \in \mathbb{R}$, $\forall (s,T), (s,T) \in A$

$$|K^i_{(s,T)}(x,t)| \leq A_i(t) \cdot h_i(s,T),$$

les h_i étant des fonctions croissantes en s à T fixé, et en T, à s fixé.

H_2) est une condition technique qui servira dans toutes nos démonstrations, H_1) est la condition clé du problème : on consultera (13) pour constater qu'elle s'interprète très précisément en langage d'opérateurs intégraux, et le problème de déterminer toutes les familles K^i vérifiant H_1) (pour un certain type de processus X_t donné a priori), et fournissant donc de "bons" estimateurs de la forme (2) (3) (4), reste ouvert.

Nous allons simplement reprendre les deux méthodes usuelles d'estimation de la densité, citer les familles K^i correspondantes et les classes de processus pour lesquels elles vérifient H_1) et H_2).

II. Familles K^i usuelles.

1) Méthode du noyau.

Nous reprendrons, en les adaptant au temps continu, les quatre estimateurs cités en (3), qui semblent les plus généraux qu'on ait actuellement définis par cette méthode ((5),(7),(8) sont séquentiels, (7) a été utilisé en (11)).

On se donne une fonction K mesurable en x et y, positive, bornée, telle que $\int_{-\infty}^{+\infty} K(x,y)dy = 1$, une fonction δ définie sur $\mathbb{R}^+$, à valeurs dans $\mathbb{R}^{+*}$, décroissante, et telle que $\lim_{t\to\infty} \delta(t) = 0$, enfin une application H de $\mathbb{R}^{+*}$ dans $\mathbb{R}^+$, décroissante. On suppose de plus que :

$$\lim_{T\to\infty} \int_0^T \delta(t)dt = +\infty \quad \text{et} \quad \lim_{T\to\infty} \int_0^T \delta(t).H[\delta(t)].dt = +\infty$$

(avec les hypothèses d'intégrabilité correspondantes). On pose alors :

(5) $$K^0_{(s,T)}(x,t) = T.(\int_0^T \delta(s)ds)^{-1} \,.\, K[t, \frac{x-t}{\delta(s)}]$$

et $$K^1_{(s,T)}(x,t) = K^2_{(s,T)}(x,t) = \sqrt{T} \,.\, (\int_0^T \delta(s)ds)^{-\frac{1}{2}} \,.\, K[t, \frac{x-t}{\sqrt{\delta(s)}}]$$

(6) $$K^0_{(s,T)}(x,t) = K^1_{(s,T)}(x,t) = K^2_{(s,T)}(x,t) = [\delta(T)]^{-1}.\, K[t, \frac{x-t}{\delta(T)}]$$

(7) $$K^0_{(s,T)}(x,t) = T.(\int_0^T \delta(t).H[\delta(t)]dt)^{-1}.\, H[\delta(s)] \,.\, K[t, \frac{x-t}{\delta(s)}]$$

et $$K^1_{(s,T)}(x,t) = K^2_{(s,T)}(x,t) = \sqrt{T} \,.\, (\int_0^T \delta(t).H[\delta(t)]dt)^{-\frac{1}{2}} \,.\, (H[\delta(s)])^{\frac{1}{2}} \,. \\ .\, K[t, \frac{x-t}{\sqrt{\delta(s)}}]$$

(8) $$K^0_{(s,T)}(s,t) = K^1_{(s,T)}(x,t) = K^2_{(s,T)}(x,t) = [\delta(s)]^{-1} \,.\, K[t, \frac{x-t}{\delta(s)}]$$

Un calcul de routine montre alors que H_1) est vérifiée lorsque $\hat{f}_T$ et $(\hat{g}_h)_T$ sont définies à l'aide des familles K^i citées en (4) (5) (6) (7) pour une très large classe de processus X_t, ceux pour lesquels f et g_h sont bornées par exemple.

De même H_2) découle directement du fait que K est bornée. Ainsi en (4) $A_o(t) = \sup_{(x,y)\in\mathbb{R}^2} [K(x,y)]$ et $h_o(s,T) = T.(\int_0^T \delta(s)ds)^{-1}$ ...

2) Méthodes des fonctions orthogonales.

On appelle $(e_i)_{i\in\mathbb{N}}$ une base orthonormale de $L^2(\mu)$ (supposé séparable afin de simplifier). Les $(e_i\, e_j)_{(i,j)\in\mathbb{N}\times\mathbb{N}}$ en forment une de $L^2(\mu \otimes \mu)$.
On se donne alors trois fonctions r_i, $i = 0,1,2$, définies sur $\mathbb{R}^+$, à valeurs dans $\mathbb{R}^+$, croissantes vers l'infini, et on pose :

$$(9) \qquad K^i_{(s,T)}(x,t) = \sum_{j=0}^{[r_i(s)]} e_j(x)\, e_j(t) \qquad i = 0,1,2$$

ou

$$(10) \qquad K^i_{(s,T)}(x,t) = \sum_{j=0}^{[r_i(T)]} e_j(x)\, e_j(t) \qquad i = 0,1,2$$

[ ] représente la partie entière. (9) fournit un estimateur séquentiel.

Supposons alors que f et g_h soient de carrés intégrables. Ils s'écrivent dans $L^2(\mu)$ et $L^2(\mu \otimes \mu)$: $\sum_{i=0}^{\infty} a_i\, e_i$, et $\sum_{i,j=0}^{\infty} a_{ij}\, e_i\, e_j$. Sous des conditions usuelles qui portent sur les a_i et a_{ij}, notamment, on obtient la convergence de $\sum_{i=0}^{n} a_i\, e_i(t)$ vers $f(t)$ et celle de $\sum_{i,j=0}^{n,n'}$ vers $g_h(t,t')$ ($n\to\infty$, $n'\to\infty$), (pour presque tous les t et (t,t') par exemple). Pour de tels t et (t,t'), H_1) est alors vérifié (calcul direct).

Les fonctions majorantes utilisées en H_2) peuvent de même se calculer cas par cas, pour toutes les bases orthonormales courantes de $L^2(\mathbb{R})$ (voir (1)), et de façon triviale si les e_i sont bornées uniformément.

Remarques sur les méthodes splines.

Choisissons une famille de subdivisions $(x_i^s)_{0\leq i\leq N(s)}$ de $[0,1]$, telles que, $\forall s$, $s \in \mathbb{R}^+$:

- $0 = x_0^s < \dots < x_{N(s)}^s = 1$
- $\max\limits_{1\leq j\leq N(s)} (x_j^s - x_{j-1}^s) \; / \min\limits_{1\leq j\leq N(s)} (x_j^s - x_{j-1}^s) \leq C_0, \qquad C_0 \in \mathbb{R}^{+*}$
- $\lim\limits_{s\to\infty} \min\limits_{1\leq j\leq N(s)} (x_j^s - x_{j-1}^s) = 0$

Soit alors $x \in [0,1]$, appelons $S^s_{\psi_x}$ la spline cubique interpolant l'application ψ_x, où

$$\psi_x \begin{cases} [0,1] \to \mathbb{R} \\ \quad t \to 1_{[0,t]}(x) \end{cases}$$

sur les noeuds $(x_i^s)_{1\leq i\leq N(s)}$, et vérifiant $(S^s_{\psi_x})''(0) = (S^s_{\psi_x})''(1) = 0$.

En posant $K^o_{(s,T)}(x,t) = \frac{d}{dt}[S^s_{\psi_x}(t)]$, on obtient une famille de fonctions qui vérifient H_1) et H_2) en ce qui concerne $\hat{f}_T$, pour certains processus X_t.

En effet, les résultats établis par Berlinet en (14) montrent que si la fonction de répartition F commune des X_t est élément de $H^2([0,1]$, $(\hat{f}_T)(t)$ est un estimateur asymptotiquement sans biais de $f(t)$, et qu'on a, d'autre part $\sup\limits_{\substack{x\in[0,1]\\ t\in[0,1]}} |K^o_{(s,T)}(x,t)| \leq D \,.\, \left(\min\limits_{1\leq j\leq N(s)} (x_j^s - x_{j-1}^s)\right)^{-1}$.

De plus l'estimateur ainsi construit est séquentiel. (Tout ceci suppose bien sûr que f soit à support dans $[0,1]$ mais peut être transcrit pour un intervalle $[a,b]$ quelconque).

III. *Convergences en moyenne quadratique*.

Pour établir celles-ci, nous utiliserons les 2 résultats suivants

Proposition 1.- (On reprend les notations de I - (1)). Deux variables aléatoires réelles f et g mesurables respectivement par rapport à M_0^T et M_{T+h}^∞ vérifient : $|\mathrm{cov}(f,g)| \leq 4.\alpha(h).\|f\|_\infty.\|g\|_\infty$ (cov représente la covariance).

Démonstration : voir (15)

Proposition 2.- Soit I_t, $t \in [0,T]$, un processus réel défini sur $(\Omega, \mathcal{Q}, P)$ tel que chaque I_t soit élément de $L^2(\Omega, \mathcal{Q}, P)$, alors :

$$\mathrm{Var}\left[\int_0^T I_t dt\right] = \int_0^T \int_0^T \mathrm{Cov}(I_t, I_{t'}) dt dt'$$

Démonstration : classique

Nous pouvons alors énoncer :

Théorème 1.- Sous les hypothèses H_1) et H_2, la condition

(11) $$\lim_{T\to\infty}\{(\int_0^T \alpha(u)du) \cdot (\int_0^T h_o^2(s,T) \cdot ds) \cdot T^{-2}\} = 0$$

est suffisante pour que $\hat{f}_T(t)$ converge en moyenne quadratique vers $f(t)$, lorsque $T \to \infty$, et de même la condition

$$(12) \qquad \lim_{T\to\infty}\{(\int_0^T \alpha(u)du) \cdot (\int_0^T h_1^2(u,T) \cdot h_2^2(u,T)du) \cdot T^{-2}\} = 0$$

est suffisante pour que $(\hat{g_h})_T(t,t')$ converge en moyenne quadratique vers $g_h(t,t')$.

<u>Démonstration</u> :

a) Commençons par la convergence de f_T. H_1) étant vérifiée, il suffit de prouver que $\mathrm{Var}\left[\hat{f}_T(t)\right] \to 0$ lorsque $T \to \infty$. Une application directe des propositions (1) et (2) donne :

$$\mathrm{Var}\left[(\hat{f}_T)(t)\right] \leq 4.T^{-2}\int_0^T\int_0^T \alpha(|s-s'|)\,\|K^o_{(s,T)}(X_s,t)\|_\infty\,\|K^o_{(s',T)}(X_{s'},t)\|_\infty\,ds\,ds'$$

soit, grâce à H_2) :

$$\mathrm{Var}\left[(\hat{f}_t)(t)\right] \leq 4.T^{-2}.A_o^2(t).\int_0^T\int_0^T \alpha(|s-s'|)h_o(s,T)h_o(s',T)ds\,ds'$$

L'intégrale double s'écrit successivement :

$$2\int_0^T h_o(s,T)\{\int_0^s [h_o(s',T)].\alpha(s-s')ds'\}ds$$

par symétrie, puis

$$2\int_0^T h_o(s,T)\{\int_0^s h_o(s-u,T)\alpha(u)du\}ds$$

et

$$2\int_0^T \alpha(u)\{\int_u^T h_o(s-u,T)h_o(s,T)ds\}du.$$

Comme h_o est positive et croissante (H_2), l'intégrale se majore finalement par

$$2(\int_0^T \alpha(u)du)(\int_0^T h_o^2(s,T)ds)$$

et ceci achève la démonstration, en donnant :

(13) $$\mathrm{Var}[(\hat{f}_T)(t)] \leq 8.T^{-2}.A_o^2(t).(\int_0^T \alpha(u)du)(\int_0^T h_o^2(s,T)ds)$$

b) Pour $(\hat{g}_h)_T$, la technique est identique. Désignons par V la variance, posons $I_s = K^1_{(s,T)}(X_s,t).K^2_{(s,T)}(X_{s+h},t')$, $f(s) = h_1(s,T)h_2(s+h,T)$, et enfin $\beta = 4[A_1(t)A_2(t')]^2$.

Les propositions 2 et 3, le lemme de Scharwz, donnent une majoration de $T^2.V[(\hat{g}_h)_T(t,t')]$ par :

$$\int_0^h ds\{\int_0^{s+h}[V(I_s).V(I_{s'})]^{\frac{1}{2}}\}ds' + \int_{s+h}^{T-h}\alpha(s'-s-h)\|I_s\|_\infty\|I_{s'}\|_\infty\, ds'\}.$$

$$+\int_h^{T-2h} ds\{\int_0^{s-h}\alpha(s-s'-h)\|I_s\|_\infty\ \|I_{s'}\|_\infty\, ds' + \int_{s-h}^{s+h}[V(I_s)V(I_{s'})]^{\frac{1}{2}}\, ds' +$$

$$+\int_{s+h}^{T-h}\alpha(s'-s-h)\|I_s\|_\infty\ \|I_{s'}\|_\infty\, ds'\}$$

$$+\int_{T-2h}^{T-h} ds\{\int_0^{s-h}\alpha(s-s'-h)\|I_s\|_\infty\|I_{s'}\|_\infty ds' + \int_{s-h}^{T-h}|V(I_s)V(I_{s'})|^{\frac{1}{2}}ds'\}.$$

Grâce à H_2) cette majoration devient :

$$\beta \ . \int_0^h ds\{\int_0^{s+h} f(s)f(s')ds' + \int_{s+h}^{T-h} \alpha(s'-s-h)f(s)f(s')ds'\} +$$

$$+ \beta \int_h^{T-2h} ds\{\int_0^{s-h} \alpha(s-s'-h)f(s)f(s')ds' + \int_{s-h}^{s+h} f(s)f(s')ds' +$$

$$+ \int_{s+h}^{T-h} \alpha(s'-s-h)f(s)f(s')ds'\} +$$

$$+ \beta \int_{T-2h}^{T-h} ds\{\int_0^{s-h} \alpha(s-s'-h)f(s)f(s')ds' + \int_{s-h}^{T-h} f(s)f(s')ds'\}$$

Maintenant la croissance des h_i donne directement, de par la définition de $f(s)$

$$\int_0^h ds.f(s).\{\int_0^{s+h} f(s')ds'\} \leq (\int_h^{2h} h_1(u,T)h_2(u,T)du)(\int_{T-2h}^{T} h_1(u,T)h_2(u,T)du)$$

$$\int_h^{T-2h} f(s)ds\{\int_{s-h}^{s+h} f(s')ds'\} \leq (\int_{2h}^{T-h} h_1(u,T)h_2(u,T)du)(\int_{T-2h}^{T} h_1(u,T)h_2(u,T)du)$$

$$\int_{T-2h}^{T-h} dsf(s)\{\int_{s-h}^{T-h} f(s')ds'\} \leq (\int_{T-h}^{T} h_1(u,T)h_2(u,T)du)(\int_{T-2h}^{T} h_1(u,T)h_2(u,T)du)$$

A fortiori la somme des 3 intégrales précédentes se majore par $\int_0^T h_1^2(u,T)h_2^2(u,T)du$.

Reste à majorer la somme des 4 intégrales doubles qui par symétrie se réduit à :

$$2\int_h^{T-h} f(s)ds \ . \int_0^{s-h} \alpha(s-s'-h)f(s')ds'$$

Un calcul identique à celui qu'on a mené pour $\hat{f}_T$ ramène cette intégrale à

$$2\int_0^{T-2h} \alpha(u)du \ . \int_{u+h}^{T-h} f(s-h-u)f(s)ds$$

qui se majore à son tour par (croissance des h_i)

$$2\left(\int_0^T \alpha(u)du\right) \cdot \left(\int_0^T h_1^2(s,T)h_2^2(s,T)ds\right).$$

En regroupant ce qui précède, il vient :

$$(14) \qquad T^2 \cdot V\left[(\hat{g}_h)_T(t,t')\right] \leq \beta\left[1+2\int_0^T \alpha(u)du\right]\left[\int_0^T h_1^2(s,T)h_2^2(s,T)ds\right].$$

Cela achève la démonstration, puisque sous H_1) $(\hat{g}_h)_T$ est asymptotiquement sans biais...

Pour obtenir la convergence en moyenne quadratique de $(\hat{g}_h^a)_T$ nous avons malheureusement dû augmenter les conditions imposées aux familles K_i choisies, en supposant ici que $K^o_{(s,T)}$ est une famille de fonctions strictement positives et que $K^1 = K^o$. Dans la pratique, on obtient comme exemple immédiat d'utilisation les familles définies en (6) et (7). En fait, la condition imposée a pour but d'obtenir :

Proposition 3.- Sous H_2), si $K^o_{(s,T)}$ est une famille de fonctions positives, et si $K^1 = K^o$, il vient :

$$(15) \qquad \left|(\hat{g}_h^a)_T(t)\right| = O\left[h_2(T,T)\right]$$

Démonstration :

Il suffit de remarquer que dans tous les cas :

$$|(\hat{g}^a_h)_T(t)| = \left|\int_0^{T-h} K^1_{(s,T)}(X_s,a)K^2_{(s,T)}(X_{s+h},t)ds\right| . \left\{\left|\int_0^T K^o_{(s,T)}(X_{s},t)ds\right|\right\}^{-1}$$

$$\leq \{\sup_{0\leq s\leq T} \sup_{x\in R} |K^2_{(s,T)}(x,t)|\}.\left\{\int_0^{T-h} |K^1_{(s,T)}(X_s,a)|ds\right\} \left\{\left|\int_0^T K^o_{(s,T)}(X_s,t)ds\right|\right\}^{-1}$$

$= 0\ [h^2(T,T)]$, sous les hypothèses de la proposition.

Remarques.-

1) La démonstration qui précède montre que l'on peut affaiblir la condition $K^1 = K^o$, dès lors que l'on puisse obtenir une majoration comme (15). Il est clair que $0 \leq K^1_{(s,T)} \leq K^o_{(s,T)}$ $\forall(s,T) \in A$, suffit, par exemple, mais en pratique, il sera plus avantageux d'essayer de majorer directement par une constante ou une fonction faiblement croissante le rapport

$$\left\{\int_0^{T-h} |K^1_{(s,T)}(X_s,a)|ds\right\} . \left\{\left|\int_0^T K^o_{(s,T)}(X_s,a)ds\right|\right\}^{-1}$$

ou globalement $|(\hat{g}^a_h)_T(t)|$, pour chaque choix de familles K^o, K^1, K^2. Dès lors que cette majoration soit possible, le théorème qui suit sera valide en utilisant un résultat du type (15).

2) Sur les exemples qui relèvent de la proposition 3, et de façon générale de la méthode du noyau, le fait d'obtenir $K^o_{(s,T)} > 0$ ne pose aucun problème. Cette hypothèse impliquant : $I_{\{(\hat{f}_T)(a)\neq 0\}} = 1$, $\forall T \in \mathbb{R}^+$, simplifie les calculs qui suivent. On peut se contenter (cf. remarque 1) de conditions particulières sur $I_{\{(\hat{f}_T)(a)\neq 0\}}$, et ici encore il est plus intéressant de revoir le problème pour chaque choix de K^o : nous nous sommes contentés d'une hypothèse simple, afin d'alléger la démonstration. [La question a déjà été soulevée en (6) : "définition d'un estimateur strict", pour les fonctions orthogonales].

Théorème 2.- Sous les hypothèses H_1) et H_2), si $K^o_{(s,T)}$ est une fonction strictement positive pour tout (s,T), si on choisit dans (3) $K^1 = K^o$, et si

(15) $$\lim_{T\to\infty} \{T-2 \;.\; (\int_0^T \alpha(u)du) \;.\; (\int_0^T h_o^2(s,T)ds) \;.\; (h_2^2(T,T))\} = 0$$

alors $(\hat{g_h^a})_T(t)$ converge vers $(g_h^a)(t)$ en moyenne quadratique, lorsque $T \to \infty$.

Démonstration :

Il s'agit de calculer $A = E\{[(\hat{g_h^a})_T(t)-(g_h^a(t)]^2\}$.

Pour T assez grand les quantités $\bar{f}_T(a) = E[(\hat{f}_T)(a)]$ et $(\bar{g}_h)(a,t) = E[(\hat{g}_h)_T(a,t)]$ sont strictement positives puisque de limites $f(a)$ et $g_h(a,t)$ qui sont strictement positives. En utilisant la définition (4) on peut alors écrire l'identité :

$$A = E\{[(\hat{g}_h)_T(a,t).(\bar{f}_T(a))^{-1}-(\bar{f}_T(a))^{-1}.[(\hat{g_h^a})_T(t)].[\hat{f}_T(a)-\bar{f}_T(a)]-(g_h^a)(t)]^2\}$$

soit $A = B_1+B_2+B_3+B_4+B_5$, en développant, avec

$$B_1 = [(\bar{g}_h)(a,t).(\bar{f}_T(a))^{-1}-g_h^a(t)]^2$$

$$B_2 = [\bar{f}_T(a)]^{-2}.\mathrm{Var}[(\hat{g}_h)_T(a,t)]$$

$$B_3 = [\bar{f}_T(a)]^{-2}.E\{[(\hat{g_h^a})_T(t)]^2.[\hat{f}_T(a)-\bar{f}_T(a)]^2\}$$

$$B_4 = -2.[\bar{f}_T(a)]^{-2}E\{[(\hat{g_h^a})_T(t)].[(\hat{g}_h)_T(a,t)].[\hat{f}_T(a)-\bar{f}_T(a)]\}$$

$$B_5 = 2[(g_h^a)(t)].[\bar{f}_T(a)]^{-1}.E\{[(\hat{g_h^a})_T(t)].[\hat{f}_T(a)-\bar{f}_T(a)]\}$$

La proposition 3 permet d'écrire :

$$|B_3| = O\{[h_2(T,T)]^2.\mathrm{Var}(\hat{f}_T(a))\}$$

$$|B_5| = O\{[h_2,(T,T)].[\mathrm{Var}(\hat{f}_T(a))]^{\frac{1}{2}}\}$$

$$|B_4| = O\{[h_2(T,T)].(E\{[(\hat{g}_h)_T(a,t)]^2\}\{\mathrm{Var}(\hat{f}_T(a))\})^{\frac{1}{2}}$$

Supposons alors que $(\hat{g}_h)_T(a,t)$ et $\hat{f}_T(a)$ convergent en moyenne quadratique vers $(g_h)(a,t)$ et $f(a)$ (conditions du th. 1) et que, de même, $[h_2(T,T)]^2.\mathrm{Var}(\hat{f}_T(a))$ tende vers 0 lorsque $T \to \infty$. Alors :

$$|B_2+B_3+B_4+B_5| = O\{(h_2(T,T)) \, . \, [\mathrm{Var}(\hat{f}_T(a))]^{\frac{1}{2}}\} = o(1)$$

Or les formules (13) et (14) montrent que la condition (15) est suffisante à toutes ces convergences, puisque (H_2) h_2 est croissante en s et T. Dès lors, sous cette condition on a exactement :

$$(16) \qquad A = \{[\bar{g}_h(a,t)].[\bar{f}_T(a)]^{-1}-(g_h^a)(t)\}^2+O\{h_2(T,T).T^{-1}.(\int_0^T \alpha(u)du)^{\frac{1}{2}}(\int_0^T h_o^2(s,T)ds)^{\frac{1}{2}}\}.$$

Le premier des deux termes dépend du biais de $(\hat{g}_h)_T$ et $(\hat{f}_T)$ et tend vers 0 (H_1). Ceci achève la démonstration.

IV. *Convergence presque sûre.*

Nous aurons ici à travailler sur des processus qui présentent une condition de mélangeance précise comme le montre le :

Théorème 3.- Supposons H1 et H2 vérifiées, et de plus : $\alpha(T) = O[e^{-T}]$, pour T assez grand. Alors, sous la condition :

$$(17) : \qquad \forall\varepsilon > 0, \quad \sum_{n=1}^{\infty} \exp\{-\varepsilon.\sqrt{n}.[h_o(n,n)]^{-1}\} < +\infty$$

$(\hat{f}_T)(t)$ converge p.s. vers $f(t)$ lorsque $T \to \infty$. De même sous la condition :

$$(18) \qquad \forall\varepsilon > 0, \quad \sum_{n=1}^{\infty} \exp\{-\varepsilon.\sqrt{n}.[h_1(n,n).h_2(n,n)]^{-1}\} < +\infty,$$

$(\hat{g_h})_T(t,t')$ converge p.s. vers $g_h(t,t')$ lorsque $T \to \infty$.

Corollaire.- Sous H1 et H2, si $\alpha(T) = O[e^{-T}]$ pour T assez grand et si (17) et (18) sont vérifiées, $(\hat{g_h^a})_T(t)$ converge p.s. vers $(g_h^a)(t)$ lorsque $T \to \infty$.

Démonstration : (inspirée de (16))

a) $(\hat{f_T})(a)$ étant un estimateur asymptotiquement sans biais de $f(a)$, sous H1, il suffit de vérifier que $[(\hat{f_T}(a)-\bar{f_T}(a)]$ tend p.s. vers 0 lorsque $T \to \infty$. On se donne à cette fin un réel ε, $\varepsilon > 0$, et on cherche à majorer :

$$C = P[(\hat{f_T})(a)-(\bar{f_T})(a) > \varepsilon] = P[\int_0^T X_s ds > \varepsilon], \text{ où on a posé :}$$

$$X_s = K^o_{(s,T)}(X_s,a) - E[K^o_{(s,T)}(X_s,a)].$$

La majoration classique : $P[I>\varepsilon] \leq e^{-t\varepsilon}\, E[e^{tI}]$, $t > 0$, permet d'écrire :

$$(19) \qquad C \leq e^{-t\varepsilon T} . D, \text{ et on va majorer } D = E[\exp.(t\int_0^T X_s ds)].$$

b) On introduit un réel $k = k(T)$, à notre disposition, et ℓ l'entier tel que $2\ell k \leq T < 2(\ell+1)k$. On pose alors

$$I_j = [2(j-1)k,(2j-1)k], \qquad I'_j = [(2j-1)k,\ 2jk] \qquad j = 1,\dots,\ell+1$$

$$I_j = \int_{I_j} X_s ds \qquad Z_j = \int_{I'_j} X_s ds \qquad j = 1,\dots,\ell$$

$I_{\ell+1} = 0$, $\int_{2\ell k}^T X_s ds$, $\int_{I_{\ell+1}} X_s ds$ selon que $2\ell k = T$, $2k\ell < T \leq (2\ell+1)k$, ou $(2\ell+1)k > T$.

$Z_{\ell+1} = 0$, ou $\int_{(2\ell+1)k}^{T} X_s ds$, selon que $(2\ell+1)k \leq T$ ou $(2\ell+1)k > T$.

De cette sorte,

$$\int_0^T X_s ds = \sum_{i=1}^{\ell+1} I_i + \sum_{i=1}^{\ell+1} Z_i$$

c) La définition des X_s jointe à H2 donne :

$$|I_i| \leq 2A_o(a) \int_{I_i} h_o(s,T)ds \quad \text{et} \quad |Z_i| \leq 2A_o(a) \int_{I_i'} h_o(s,T)ds, \qquad i = 1,\dots,\ell$$

et h_o étant positive croissante, tous les $|I_i|$ et $|Z_i|$, $i = 1,\dots,\ell$, se majorent par : $2A_o(a)\int_{(2\ell+1)k}^{2(\ell+1)k} h_o(s,T)ds$, soit, a fortiori, par :
$2A_o(a)\ \sqrt{k}\left[\int_{(2\ell+1)k}^{2(\ell+1)k} h_o^2(s,T)ds\right]^{\frac{1}{2}}$.

En appelant $(t^{-1}/2)$ cette dernière quantité, on peut écrire :

$$\forall i = 1,\dots,\ell \qquad 0 \leq 2t\|I_i\|_\infty \leq 1 \quad \text{et} \quad 0 \leq 2t\|Z_i\|_\infty \leq 1$$

L'inégalité classique : $(0 \leq 2t\|I\|_\infty \leq 1) \Rightarrow \{E[\exp(2tI)]\} \leq \exp\{3t^2 E[I^2]\}$

donne finalement :

$$\text{(20)} \qquad \text{si} \quad t = [4\ A_o(a)]^{-1} \cdot \{k.\int_{(2\ell+1)k}^{2(\ell+1)k} h_o^2(s,T)ds\}^{-\frac{1}{2}},$$

$$\prod_{i=1}^{\ell} E[e^{2tI_i}] \leq \exp\{3t^2 . \sum_{i=1}^{\ell} E(I_i^2)\}$$

et

$$\prod_{i=1}^{\ell} E[e^{2tZ_i}] \leq \exp\{3t^2 \sum_{i=1}^{\ell} E(Z_i^2)\}$$

d) $E[I_i^2] = E([\int_{I_i} X_s ds]^2) = \mathrm{Var}\{\int_{I_i} K^o_{(s,T)}(X_s,a)ds\}.$

Un calcul identique à celui du théorème 1 (formule (13)), donne :

$$E(I_i^2) \leq 8\ A_o^2(a)\ .\ (\int_{I_i} \alpha(u)du)\ .\ (\int_{I_i} h_o^2(s,T)ds)$$

et de même $\quad E(Z_i^2) \leq 8\ A_o^2(a)\ .\ (\int_{I'_i} \alpha(u)du)\ .\ (\int_{I'_i} h_o^2(s,T)ds)$, pour $i = 1,\dots,\ell$.
En substituant dans (20), on obtient une majoration commune de $\prod_{i=1}^{\ell} E[e^{2tI_i}]$ et $\prod_{i=1}^{\ell} E[e^{2tZ_i}]$ par :

$$(21) \qquad \exp\{(3/2)\ .\ (\int_0^T \alpha(u)du)\ .\ k^{-1}\}\ ,$$

puisque $\sum_{i=1}^{\ell} E(I_i^2) \leq 8\ A_o^2(a)\ \{\sum_{i=1}^{\ell}\ (\int_{I_i} \alpha(u)du)\ .\ (\int_{I_\ell} h_o^2(s,T)ds)\}$

$$\leq 8[A_o(a)]^2\ (\int_0^T \alpha(u)du)\ .\ \int_{I_\ell} h_o^2(s,T)ds$$

et $\int_{I_\ell} h_o^2(s,T) \leq \int_{(2\ell+1)k}^{2(\ell+1)k} h_o^2\ (s,T)ds$ ... (et de même pour les Z_i).

e) On majore ensuite, par récurrence :

$$E = E_\ell = |E\{\exp.[2t.(\sum_{i=1}^{\ell} I_i)]\} - \prod_{i=1}^{\ell} E(\exp[2tI_i])|$$

Il suffit de remarquer que $E_\ell = |A+B|$ avec :

$$A = E[\exp\{2t.\sum_{i=1}^{\ell-1} I_i\}\ .\ \exp.\{2t.I_\ell] - E[\exp.\{2t\sum_{i=1}^{\ell-1} I_i\}]\ .\ E[\exp\ 2tI_\ell]$$

$$B = E[\exp.\{2t.\sum_{i=1}^{\ell-1} I_i\}]\ .\ E[\exp\ 2tI_\ell] - E[\exp\{2tI_\ell\}]\ .\ \prod_{i=1}^{\ell-1} E[\exp\{2tI_i\}]$$

d'où, en majorant A grâce à la proposition (1) :

$$|E_\ell| \leq 4\alpha(k)\ \|\exp.\{2tI_\ell\}\|_\infty .\ \|\exp\{2t\sum_{i=1}^{\ell-1} I_i\}\|_\infty + \|E[\exp 2tI_\ell]\|_\infty .\ |E_{\ell-1}|$$

On utilise la majoration de $|I_\ell|$ vue en b) pour écrire :

$$\|\exp\{2tI_i\}\|_\infty \leq \exp\{2t|I_i|\} \leq \exp\{4t.A_o(a).\int_{I_i} h_o(s,T)ds\}, \quad i = 1,\dots,\ell$$

et en remarquant ici encore que $\int_{I_i} h_o(s,T)ds \leq \int_{I_\ell} h_o(s,T)ds$, il vient :

$$|E_\ell| \leq 4\ \alpha(k)\ .\ \exp\{4t\ A_o(a)\ .\ \ell\ .\int_{I_\ell} h_o(s,T)ds\} + |E_{\ell-1}|\ .\ E\{\exp.2t\ I_\ell\}$$

on itère le procédé avec :

$$|E_{\ell-1}| \leq 4\ \alpha(k)\ \exp\{4t\ A_o(a).\sum_{i=1}^{\ell-1}\int_{I_j} h_o(s,T)ds\} + E\{\exp[2tI_{\ell-1}]\}\ .\ |E_{\ell-2}|$$

d'où

$$|E_\ell| \leq 2.[4\alpha(k)]\ .\ \exp\{4t\ A_o(a).\ell\int_{I_\ell} h_o(s,T)ds\} + |E_{\ell-2}|\ .\ E\{\exp(2tI_{\ell-1})\}.$$
$$.\ E\{\exp(2tI_\ell)\}$$

etc... jusque

$$|E| \leq 4(\ell-1).\alpha(k).\exp\{4t\ A_o(a).\ell.\int_{I_\ell} h_o(s,T)ds\}\ .$$

(On obtient bien sûr le même résultat pour les Z_i avec I'_ℓ au lieu de I_ℓ).
Enfin en reprenant la définition de t en c) on voit que la majoration peut s'écrire :

(22) $$|E| \leq 4(\ell-1)\ \alpha(k)\ \exp(\ell)$$

f) En combinant (22) et (21), on obtient finalement une majoration commune de $|E\{\exp 2t(\sum_i^\ell Z_i)\}|$ et $|E\{\exp 2t(\sum_i^\ell I_i)|$ par

(23) $$4(\ell-1)\ \alpha(k)\ .\ \exp(\ell) + \exp.\{3/2(\int_0^T \alpha(u)du)k^{-1}\}.$$

Or

$$D = E\left[\exp\{t.\int_0^T X_s ds\}\right] = E\{\exp.\left[t.\int_{I_{\ell+1}\cup I'_{\ell+1}} X_s ds + t.(\sum_i^\ell I_i) + t(\sum_i^\ell Z_i)\right]\}$$

$$\leq \exp t\left|\int_{I_{\ell+1}\cup I'_{\ell+1}} X_s ds\right| . \{E\left[\exp\{2t \sum_i^\ell I_i\}\right] . E\left[\exp 2t.\sum_i^\ell Z_i\}\right]\}^{\frac{1}{2}}$$

(inégalité de Schwarz) donc

$$D \leq e^2\left[4(\ell-1)\alpha(k)\exp(\ell) + \exp\{3/2(\int_0^T \alpha(u)du)\ .\ k^{-1}\}\right],$$

grâce à (23), la définition de t et la majoration usuelle de $|X_s|$ ...

Comme

$$\exp\{-\varepsilon Tt\} = \exp\{-\varepsilon' T\left[k\int_{(2\ell+1)k}^{2(\ell+1)k} h_o^2(s,T)\right]^{-\frac{1}{2}}\} \quad \text{où} \quad \varepsilon' = \varepsilon(4\ A_o(a))^{-1}$$

$$\leq \exp\{-\varepsilon'(T/k)\ .\ \left[h_o(T,T)\right]^{-1}\} \qquad \text{(croissance de } h_o)$$

(19) donne :

(24) $$C \leq e^2\ .\ \exp(-\varepsilon'(T/k)\left[h_o(T,T)^{-1}\right])\ .\ \left[4(\ell-1)\exp(\ell)\alpha(k)\ +\right.$$

$$\left.+ \exp\{3/2.(\int_0^T \alpha(u)du)\ .\ k^{-1}\}\right].$$

g) Prenons $k = k(T) = \sqrt{T}$, supposons $\alpha(T) = O[\exp(-T)]$, alors, d'après (24)

$$C = P[\hat{f}_T(a)-\bar{f}_T(a) > \varepsilon] = O\{\exp[-\varepsilon' \,.\, \sqrt{T} \,.\, [h_o(T,T)]^{-1}]\}$$

Comme $P[\hat{f}_T(a)-\bar{f}_T(a) < -\varepsilon] = P[\int_0^T (-X_s)ds > \varepsilon]$ relève de la même majoration que C on a enfin :

$$(25) \qquad \forall\varepsilon > 0 \qquad P[|\hat{f}_T(a)-\bar{f}_T(a)| > \varepsilon] = O\{\exp[-\varepsilon' \sqrt{T}[h_o(T,T)]^{-1}]\}.$$

Reste à montrer que ceci démontre le théorème grâce à la condition (18) or si $\alpha_T = \hat{f}_T(a) - \bar{f}_T(a)$, en notant simplement que :

$$|\alpha_T| \leq |\alpha_T - \alpha_{[T]}| + \alpha_{[T]}$$

et $|\alpha_T-\alpha_{[T]}| = |\frac{1}{T}\int_{[T]}^T X_s ds| = O(\frac{h_o(T,T)}{T}) = o(1)$ sous les hypothèses du théorème 3 (majorations usuelles), il est clair que $\alpha_{[T]} \xrightarrow[T\to\infty]{p.s} 0$ est condition suffisante de convergence p.s de α_T vers 0 lorsque $T \to \infty$.

Or la convergence de $\alpha_{[T]}$ est assurée par (18), grâce à (25) (lemme de Borel-Cantelli pour une suite de variables).

i) Il est clair qu'à de minimes variantes près, la même démonstration s'applique à $(\hat{g}_h)_T$, et le corollaire est trivial.

Remarque.- Les conditions choisies au g) ci-dessus peuvent être modulées, par exemple en prenant $k = T^{1/4}$, et en transformant $\alpha(u) = O[e^{-u}]$ en $\alpha[u^{1/4}] = O[\exp(-u^{3/4})]$, et de même (18) en $\sum_{n=1}^{\infty} \exp\{-\varepsilon.T^{3/4}.(h_o(T,T))^{-1}\}$. De façon générale on perdra en conditions sur α ce qu'on gagnera sur h_o. Nous avons adopté la solution "moyenne".

Bibliographie.-

(1) J. BLEUEZ et D. BOSQ : *Etude d'une classe d'estimateurs non paramétriques de la densité.* Annales de l'Institut Henri Poincaré - Vol. XIV, n°4, 1978, p. 479-484.

(2) M. BERTRAND-RETALI : *Convergence uniforme d'un estimateur de la densité par la méthode du noyau.* Revue Roumaine de Mathématiques pures et appliquées, Tome XXIII, n°3, 1978.

(3) P. DEHEUVELS : *Conditions nécessaires et suffisantes de convergence ponctuelle presque sûre et uniforme presque sûre des estimateurs de la densité.* C.R. de l'Acad. des Sciences de Paris - Série A, t.278, (29/4/74) p.1217-1220.

(4) J. GEFFROY : *Sur l'estimateur de la densité dans un espace métrique.* C.R. de l'Acad. des Sciences de Paris - Série A, t.278, 1974, p.1449.

(5) M.J. FRYER : *A review of some non-parametric methods of density estimation.* J. Inst. Maths. Applic. 20, p. 335-354, (1977).

(6) D. BOSQ : *Sur l'estimation de la densité d'un processus stationnaire et mélangeant.* C.R. de l'Acad. des Sciences de Paris - Série A, t.277, p.535-

(7) A. FÖLDES : *Density estimation for dependent sample.* Studia Scientiarum Mathematicamum Hungarica 9, (1974) 443-452.

(8) M. ROSENBLATT : *Density estimates and Markov sequence.* In Non Parametric techniques in statistical Inference (M. Puri, ed.) 199-210.

(9) G.G. ROUSSAS : *Non parametric estimation of the transition distribution function of a Markov process.* A.M.S. 1969, Vol. 40, n°4, p. 1886-1900.

(10) G. BANON : *Estimation non paramétrique de densités de probabilité pour les processus de Markov.* Thèse, Université Paul Sabatier - Toulouse (1977).

(11) T.H. NGUYEN et G. BANON : *Strong consistent recursive estimation of the coefficient of the Ito equation.* To appear in Annals of Stat.

(12) M. ROSENBLATT : *A central limit theorem and a mixing condition.* Proc. of nat. Acad. of Sciences USA. 42, p. 43-47 (1956).

(13) D. BOSQ : *Sur l'estimation sans biais d'un paramètre fonctionnel.* Publication Internes de l'UER de Math. Pures et Appliquées n°104 (mars 1977) Université de Lille I.

(14) A. BERLINET : *Sur les méthodes splines en estimation de la densité*
C.R. de l'Acad. des Sciences de Paris - t. 288 série A - p. 847-850 (7/5/1979).

(15) V.A. VOLKONSKII et YU.A. ROZANOV : *Some limit theorems for random functions I*
Theory of probability and its applications 1959 - Vol. IV, (p. 178-197).

(16) D. BOSQ : *Inégalité de Bernstein pour les processus stationnaires et mélangeants*
C.R. de l'Acad. des Sciences de Paris, t. 281, p. 1095-1098 (22/12/75).

Reçu en Septembre 1979

REGRESSION ET PREDICTION DANS
LES PROCESSUS ERGODIQUES.

(Chapitre VII)

Le chapitre qui suit est consacré à l'étude de quelques problèmes relatifs à la régression et la prédiction dans les processus ergodiques, les résultats obtenus généralisant, quant aux hypothèses faites sur le processus observé, les travaux déjà réalisés en ce domaine, dont une excellente illustration est fournie par Collomb 1984 (29) par exemple, qui fournit aussi une bibliographie pratiquement exhaustive sur le sujet.

α) Considérons un processus (X_t) , $t \in \mathbb{Z}$, à valeurs dans $\mathbb{R}^d$, et une application ψ de $\mathbb{R}^d$ dans un espace de Hilbert séparable H. Supposant ψ bornée, on peut alors définir pour tout t , r et k étant deux éléments fixés de $\mathbb{N}^*$, l'espérance conditionnelle $E[\psi(X_{t+k})/(X_t,\ldots,X_{t-r+1})]$. On supposera l'existence d'une version commune R de cette espérance, indépendante de t : si $(x) = (x_1,\ldots,x_r)$ est un élément de $(\mathbb{R}^d)^r$ et $X_t^{(r)}$ le r-uple aléatoire $(X_t ,\ldots, X_{t-r+1})$, on a donc :

$$(58) \qquad \forall t , \quad t \in \mathbb{Z} , \qquad R((x)) = E[\psi(X_{t+k}) \;/\; X_t^{(r)} = (x)]$$

On supposera enfin que R , bornée comme ψ , est uniformément continue sur $\mathbb{R}^d$, à valeurs dans H .

Nous étudions alors un estimateur de R , fonction de l'échantillon observé $X_1,\ldots,X_n$, qui se base sur l'idée naturelle suivante : approximer $E[\psi(X_{t+k}) \;/\; X_t^{(r)} = (x)]$ par une moyenne pondérée des $n - r - k+1$ valeurs $\psi(X_{t+k})$ observables pour l'échantillon, la pondération utilisée, aléatoire, reposant sur la proximité de $X_t^{(r)}$ et de (x) , $t = r ,\ldots,n-k$. Une telle pondération est obtenue de façon élémentaire en posant :

$$R_n((x)) = \sum_{t=r}^{n-k} \left[\chi_{t,n} \Big/ \left(\sum_{t=r}^{n-k} \chi_{t,n}\right)\right] \cdot \psi(X_{t+k})$$

avec $$\chi_{t,n} = \prod_{j=0}^{r-1} I_{[-h(n),h(n)]^d}(X_{t-j} - x_{j+1})$$

c'est une version du classique "prédictogramme". Nous utiliserons l'estimateur un peu plus général que l'on obtient en substituant à $\frac{1}{2}\, I_{[-1,1]}$ une densité de probabilité sur $(\mathbb{R}^d)^r$ qui soit bornée. On obtient finalement $R_n(\cdot) = N_n(\cdot) \,/\, D_n(\cdot)$, avec

$$(59)\quad \begin{cases} N_n(\cdot) = \displaystyle\sum_{t=r}^{n-k} (n-r-k+1)^{-1} \cdot (h(n))^{-dr} \cdot K\left[(\cdot - X_t^{(r)}) / h(n)\right] \cdot \psi(X_{t+k}) \\ D_n(\cdot) = \displaystyle\sum_{t=r}^{n-k} (n-r-k+1)^{-1} \cdot (h(n))^{-dr} \quad K\left[(\cdot - X_t^{(r)}) / h(n)\right] \end{cases}$$

étant entendu que le quotient est nul lorsque le dénominateur l'est : c'est l'estimateur qui a été étudié notamment par Collomb (1984) dans l'article sus-cité (g est alors à valeurs réelles) ainsi que par Doukhan (30), Peligrad (26), Härdle et Vieu (31).

Dans la première partie du chapitre nous démontrerons la convergence uniforme presque sûre, vers 0, sur tout compact, de $(R_n - R)$, lorsque $n \to \infty$, en supposant l'existence des densités de transition $g_{X_t^{(r)}}^{F_{t-r}}$, appartenant à $C_o\left[(\mathbb{R}^d)^p\right]$ et telles que

$$(60)\qquad \operatorname*{Lim\ p.s.}_{n\to\infty} \left\| (n^{-1}) \cdot \left(\sum_{t=1}^{n} g_{X_t^{(r)}}^{F_{t-r}}\right) - g \right\|_{C_o\left[(\mathbb{R}^d)^r\right]} = 0$$

pour un élément g de $C_o\left[(\mathbb{R}^d)^r\right]$. Cette condition est obtenue par une adaptation directe du lemme 4 pour les processus stationnaires ergodiques à densités suffisamment régulières : c'est une simple extension de l'hypothèse :

$$\exists h\ ,\ h \in C_o[(\mathbb{R}^d)]\ ,\ \underset{n\to\infty}{\text{Lim}}\ \text{p.s.}\ ||(n^{-1})\quad (\sum_{t=1}^{n} g_{X_t}^{F_{t-1}}) - h||_{C_o[(\mathbb{R}^d)]} = 0\ ,$$

que nous avons déjà utilisée. A titre comparatif, Collomb impose la "φ-mélangeance" au processus, condition beaucoup plus forte que l'ergodicité, mais n'exige pas l'existence des densités conditionnelles, les techniques de démonstration étant différentes des nôtres.

β) La façon d'envisager l'estimateur de la régression évoquée en α) débouche naturellement sur le problème de la prévision de la variable $\psi(X_{n+k})$ à partir de $X_1,\ldots,X_n$, un prédicteur possible P_n se déduisant directement de R_n en posant $P_n = R_n(X_n^{(r)})$.

On obtient la convergence P.S. à 0 de $||P_n - (X_n^{(r)})||_H$ en corollaire du théorème de convergence uniforme de $R_n(\cdot)$, à la fin de la 1ère partie.

Nous proposons ensuite en 2ème partie du chapitre une autre approche du problème (sous forme d'un article publié en 1985 en commun avec D. Bosq, et que nous reproduisons ici) : en considérant un processus stationnaire à temps continu, et Markovien, ψ une applications à valeurs dans H , comme ci-dessus, et R la régression de $\psi(X_{t+k})$ sur X_t , on démontre que

$$E(||\hat{R}_T(X_T) - R(X_T)||_H^2)$$

tend p.s. vers 0 , lorsque $T \to \infty$, pour une classe de prédicteurs $\hat{R}_T$ qui comprend aussi bien ceux qui dérivent de la méthode du noyau (comme ci-dessus) que de la méthode de projection évoquée au chap. 3. Le problème traité est relativement similaire à celui de la convergence de P_n : le caractère Markovien n'est guère indispensable, comme noté dans l'article, et le passage du temps continu au temps discret immédiat. La différence fondamentale résulte de ce que la loi marginale des X_t est, dans l'article, supposée connue et le processus mélangeant ce qui amène à un type de démonstration tout à fait différent.

γ) Les problèmes posés ci-dessus amènent enfin à celui de l'optimalité des prédicteurs P_n de $\psi(X_{n+k})$ proposés, au sens de l'obtention de la condition : $\lim_n \| E^{\mathcal{F}_n}[\psi(X_{n+k})] - P_n(X_1,\ldots,X_n) \| = 0$, p.s. ou en moyenne, puisque P_n dépend effectivement des n variables observées, et que au mieux, l'observateur peut espérer approcher $\psi(X_{n+k})$ par $E^{\mathcal{F}_n}(\psi(X_{n+k}))$ au vu de l'échantillon. Dans le cas d'un r processus Markovien, ceci découlera de ce qui précède (avec $r = 1$ pour l'article). Dans le cas général, lorsque $E^{\mathcal{F}_t}(\psi(X_{t+h}))$ est une fonction $\emptyset[(X_{t-i}), i \geq 0]$ le problème reste ouvert, l'estimation de $R(X_t^{(r)})$ ne le résolvant pas, lorsque r est fixé.

Nous ne connaissons actuellement sur le sujet que les résultats de D. Bosq (32) (basés sur une autre idée), mais il paraît possible de le résoudre en utilisant une suite d'estimateursde la régression $E(\psi(X_{t+h} / X_t^{(r_n)})$ où $r_n \uparrow \infty$, ce qui doit faire l'objet de travaux futurs.

Nous avons cependant déjà abordé le problème sous le biais pratique du choix du "bon r " dans une mise en oeuvre informatique de la méthode proposée : ceci fait, entre autres, l'objet de la $3^{\text{ème}}$ partie du chapitre, nous y proposons une méthode de définition d'un intervalle de confiance pour le prédicteur P_n , identique aux intervalles que proposent les méthodes paramétriques classiques (Box-Jenkins ou régression linéaire), (voir Monfort-Gouriéroux (33) pour une introduction), et qui permet une comparaison des méthodes paramétriques et non-paramétrique sur diverses séries classiques ou simulées. Pour chaque série, nous étudions la précision de la prédiction non-paramétrique selon la longueur de la régression (le r) utilisé. Simulations à l'appui, nous discutons encore de diverses améliorations possibles de la méthode (désaisonnalisation, différenciation, récursivité, pondération tenant compte des corrélations ...), et proposons une comparaison avec la méthode de Box-Jenkins.

CONVERGENCE PRESQUE SURE DU PREDICTEUR SOUS DES HYPOTHESES ERGODIQUES

(C.VII.I)

Théorème 12.- Supposons que les densités $g_{X_t^{(r)}}^{F_{t-r}}$ du processus observé (X_t) , $t \in \mathbb{Z}$, vérifient (60) , et que R, définie en (58), soit, comme ψ, bornée et uniformément continue. Si le noyau K borné utilisé en (59) est lipschitzien, l'estimateur $R_n(\cdot)$ défini en (59) vérifiera :

$$\lim_{n\to\infty} \text{ p.s. } \{ \sup_{(x)\in G} \|R_n(x) - R(x)\|_G\} = 0$$

pour tout compact G de $(\mathbb{R}^d)^r$ sur lequel g ne s'annule pas, pourvu que la suite $h(n)$ décroisse vers 0 , de telle sorte que $n\,h(n)$ soit croissante et que $h(n)$ vérifie :

$$* \quad 1 - \left(\frac{h(n)}{h(n-2\sqrt{n})}\right)^{dr} = 0\left(\frac{1}{\sqrt{n}}\right)$$

$$* \quad \exists \varepsilon\ ,\ \varepsilon > 0\ ,\quad n^{sdr-0,5}\ (h(n^2))^{-2d^2r^2-3dr} = 0(n^{-\varepsilon})$$

pour un nombre s , $s > 0$.

Démonstration :

α) On commence par démontrer que

$\sup_{(x)\in D_n} |D_n(x) - g(x)|$ $(R_n = N_n / D_n$, cf. (59)) tend vers 0 p.s. , si $n \to \infty$, où D_n est le pavé $[- n^s , n^s]^{d'}$, avec $d' = dr$. On décompose $D_n(x) - g(x)$ en :

$$\sum_{t=r}^{n-k} Y_{t,n}(x) + L_n(x) \quad ,$$

avec

$$\{(n-r-k+1)^{-1} \cdot \sum_{t=r}^{n-k} \int_{\mathbb{R}^{d'}} [h(n)]^{-d'} K\left(\frac{(x)-(y)}{h(n)}\right) \cdot g_{X_t^{(r)}}^{F_{t-r}}(y)\, dy - g(x)\} = L_n(x)$$

et

$$Y_{t,n}(x) = (n-r-k+1)^{-1}\ h(n)^{-d'}\ K\left(\frac{(x)-X_t^{(r)}}{h(n)}\right) - E^{\mathcal{F}_{t-r}}\ (n-r-k+1)^{-1}\ .$$

$$(h(n))^{-d'}\ K\left(\frac{(x)-X_t^{(r)}}{h(n)}\right)\}$$

* La démonstration permettant de montrer que $\sup_{x\in D_n} |\sum_1^n Y_{t,n}(x)|^2$

converge vers 0 p.s. se calque sur celle du théorème 11, la seule variante venant de ce que 2 variables $Y_{t,n}$ et $Y_{t',n}$ ne sont ici orthogonales que si $|t - t'| > r$. Mais on peut écrire

$$E((\sum_1^n Y_{t,n}(x))^2) = \sum_1^n E(Y_{t,n}^2(x)) + \sum_{t=1}^n \sum_{t'/|t'-t|\leq r} E(Y_{t,n}(x))\ E(Y_{t'n}(x))$$

le second terme est $O[nr \cdot (n^{-2}\quad h(n)^{-d'})]$, par application de l'inégalité de Schwartz, donc de l'ordre du premier, r étant fixé ... La majoration utilisée en (55) demeure donc valable avec $d' = dr$ au lieu de d , et le reste de la démonstration aussi ...

* Le lemme 11 permet d'affirmer que $\sup_{x\in\mathbb{R}^{d'}} [L_n(x)]$ tend vers 0 si $n \to \infty$, sous la condition (60) .

β) On montre ensuite $\sup_{x\in D_n} ||N_n(x) - R(x)\ g(x)||_H$ tend également vers 0 p.s. , en décomposant ici $N_n(x) - R(x)\ g(x)$ en $\sum_{t=1}^n Y^*_{t,n}(x) + L^*_n(x)$, avec

$$L^*_n(x) = (n-r-k+1)^{-1}\ h(n)^{-d'}\ (\sum_{t=1}^n E^{\mathcal{F}_{t-r}}\ \{\psi(X_{t+k})\ .\ K\left(\frac{x-X_t^{(r)}}{h(n)}\right)\} - R(x)\ g(x)$$

et

$$Y^*_{t,n}(x) = (n-r-k+1)^{-1}\ h(n)^{-d'}\ \{K(\frac{x-X_t^{(r)}}{h(n)}) \cdot \psi(X_{t+k})$$

$$- E^{F_{t-r}}\ [K(\frac{x-X_t^{(r)}}{h(n)}) \cdot \psi(X_{t+k})]\}$$

i) On démontre d'abord que $\sup_{(x)\in\mathbb{R}^d} ||L^*_n(x)||_H$ tend vers 0 , si $n \to \infty$, presque sûrement, sous la condition (60). En effet, on peut écrire :

$$E^{F_{t-r}}\ [K(\frac{x-X_t^{(r)}}{h(n)})\ \psi(X_{t+k})] = E^{F_{t-r}}[K(\frac{x-X_t^{(r)}}{h(n)}) \cdot R(X_t^{(r)})]$$

$$= \int_{\mathbb{R}^{d'}} K(\frac{x-y}{h(n)}) \cdot R(y)\ g^{F_{t-r}}_{X_t^{(r)}}(y)\ dy$$

en appliquant les propriétés standard de l'espérance conditionnelle et de l'intégrale de Bochner (permutation avec les formes linéaires, pour la 2ème égalité).

Par conséquent, il s'agit de prouver la converge p.s. à 0 du supremum, pour $(x) \in (\mathbb{R}^d)^r$, de

$$||\int_{\mathbb{R}^{d'}} [\frac{1}{h(n)^{d'}} \cdot K(\frac{x-y}{h(n)})]\ [\frac{1}{n-r-k+1} \sum_{t=r}^{n-k} g^{F_{t-r}}_{X_t^{(r)}}(y)]\ R(y)\ dy - R(x)g(x)||_H$$

quantité qui se majore par la somme des suprema, pour $(x) \in \mathbb{R}^{dr}$, des quantités :

$$||\int_{\mathbb{R}^{d'}} [\frac{1}{(h(n))^{d'}}\ K(\frac{x-y}{h(n)})]\ [\frac{1}{n-r-k+1} (\sum_{t=r}^{n-k} g^{F_{t-r}}_{X_t^{(r)}}(y))]\ [R(y)-R(x)]\ dy||_H$$

et

$$\left\|\int_{\mathbb{R}^{d'}}\left\{\left[\frac{1}{(h(n))^{d'}}\, K\left(\frac{x-y}{h(n)}\right)\right]\left[\frac{1}{n-r-k+1}\left(\sum_{t=r}^{n-k} g^{\mathcal{F}_{t-r}}_{X_t^{(r)}}(y)\right]\, dy - g(x)\right\} R(x)\right\|_H$$

Le second majorant, tend vers 0 p.s. sous la condition (60), d'après le lemme 11, puisque R est bornée. D'autre part, R étant uniformément continue, pour tout ε, $\varepsilon > 0$, existe un α tel que $(\|y-x\| \leqslant \alpha) \Longrightarrow (\|R(y)-R(x)\|_H < \varepsilon)$, et donc le premier majorant, δ_n est lui-même inférieur à la somme des suprema, pour $(x) \in (\mathbb{R}^d)^r$ de

$$2\Theta \cdot \int_{\{y\,|\,\|y-x\|>\alpha\}}\left[(h(n))^{-d'}\, K\left(\frac{x-y}{h(n)}\right)\right] \cdot \left[\frac{1}{n-k-r+1}\left(\sum_{t=r}^{n-k} g^{\mathcal{F}_{t-r}}_{X_t^{(r)}}(y)\right]\, dy$$

et

$$\varepsilon \left\{\int_{\{y\,|\,\|y-x\|\leqslant\alpha\}}\left[h(n)^{-d'} \cdot K\left(\frac{x-y}{h(n)}\right)\right] \cdot \left[\frac{1}{n-k-r+1}\left(\sum_{r}^{n-k} g^{\mathcal{F}_{t-r}}_{X_t^{(r)}}(y)\, dy\right\}\right.$$

où Θ vaut : $\{\sup_{(x)\in\mathbb{R}^{d'}} \|R(x)\|_H\}$.

La suite des $\dfrac{1}{n-r-k+1} \displaystyle\sum_{r}^{n-k} g^{\mathcal{F}_{t-r}}_{X_t^{(r)}}(\cdot)$ convergeant dans $C_o(\mathbb{R}^{d'})$, on aura

$$\sup_n \left(\sup_y \frac{1}{n-k-r+1} \sum_{r}^{n-k} g^{t-r}_{X_t^{(r)}}(y)\right) < \infty\,, \quad \text{et donc, pour tout } \varepsilon > 0$$

$$0 \leqslant \overline{\mathrm{Lim}}\ \text{p.s.}\ \delta_n \leqslant c \cdot \varepsilon + O\left(\mathrm{Lim}_n \int_{\{u\,|\,\|u\|>\alpha/h(n)\}} K(u)\, du\right)$$

ce qui donne la converge à 0 souhaitée, c ne dépendant pas de n.

ii) Pour montrer que $\sup_{x\in D_n} \left(\left\|\sum_1^n Y^*_{t,n}(x)\right\|_H^2\right)$ converge p.s. vers 0, on utilise encore la démonstration du théorème 11 qui s'adapte directement en tenant compte, ici aussi, de ce que

* Majorer $E(||Y^*_{t,n}(x)||^2_H)$ revient à majorer $E(||\frac{1}{n-r-k+1}\, h(n)^{-d'} \cdot \psi(X_{t+h}) \cdot K(\frac{x-X_t^{(r)}}{h(n)})||^2_H)$, et comme g est bornée, on obtient $O(\frac{1}{n^2 h(n)^{d'}})$.

* 2 variables $Y^*_{t,n}(x)$ et $Y^*_{t,n}(x)$ ne sont orthogonales que si $|t-t'| > r+k$. La remarque effectuée ci-dessus (en α) s'applique alors de même, et on aura bien $E(||\sum_1^n Y^*_{t,n}(x)||^2) = O(\frac{1}{nh(n)^{d'}})$.

ω) Pour conclure, on remarque que $||R_n(x)||_H$ est bornée par la même constante que les $||\psi(x)||_H$, puisque

$$R_n(x) = \frac{\sum_r^{n-k} \psi(X_{t+k})\, K\left[(x-X_t^{(r)}) \,/\, h(n)\right]}{\sum_r^{n-k} K\left[(x-X_t^{(r)}) \,/\, h(n)\right]}$$

et que K est une fonction positive. Alors, de l'identité

$$R_n(x) - R(x) = (\frac{N_n(x)}{g(x)} - R(x)) - R_n(x) \times \frac{D_n(x) - g(x)}{g(x)})$$

on tire la majoration :

$$\sup_{(x)\epsilon E} ||R_n(x) - (x)||_H \leq \sup_{(x)\epsilon E} \{\frac{1}{|g(x)|}\, ||N_n(x)-R(x)g(x)||_H\}$$

$$+ \sup_{(x)\epsilon E} \{||R_n(x)||_H \cdot |\frac{D_n(x)-g(x)}{g(x)}|\}$$

$$= O\{|\inf_{x\epsilon E} g(x)|^{-1}\} \{ \sup_{(x)\epsilon D_n} ||N_n(x)-R(x)g(x)||_H$$

$$+ \sup_{(x)\epsilon D_n} ||D_n(x)-g(x)||\}$$

pourvu que E , borné soit inclus dans D_n , ce qui a donc lieu asymptotiquement. Si E est compact et que g ne s'annule pas sur E , les points α) et β) permettent de conclure.

Q.E.D.

On peut déduire directement du théorème 12 des corollaires relatifs à la fonction de répartition conditionnelle, et à la loi conditionnelle de X_{t+k} , sachant que $X_t^{(r)}$ vaut (x) , en reprenant comme applications g celles que l'on a utilisées dans le chapitre II, à savoir respectivement

* $(x) \xrightarrow{g} I_{A(x)}$, $A(x) = \prod_{j=1}^{d} [x_i,\infty[$, $I_{A(x)}$ étant considéré comme élément d'un espace $L^2(\lambda_K)$, K compact de $\mathbb{R}^d$.

* $(x) \xrightarrow{g} \delta_{(x)}$, $\delta_{(x)}$ étant assimilé à un élément d'un espace auto-reproduisant,

puisque ces deux applications sont naturellement bornées (cf. chap. II).

De même, on obtient le résultat suivant, relatif à la prévision :

Corollaire 4.- Supposons que le processus (X_t) , $t \in \mathbb{Z}$, vérifie les conditions du théorème (12), les X_t étant à valeurs dans un compact G de $\mathbb{R}^d$.

Si la régression R commune des variables X_{t+k} , sur (r)-uples $X_t^{(r)}$ est continue, et que les $h(n)$ vérifient les conditions du théorème (12), on aura :

$$\lim_{n\to\infty} \text{p.s.} \left(\| R_n(X_n^{(r)}) - R(X_n^{(r)}) \|_{\mathbb{R}^d} \right) = 0$$

<u>Démonstration</u> : Il suffit d'appliquer le théorème précédent, avec l'identité pour application ψ : ψ est alors bornée et R uniformément continue car les $X_n^{(r)}$ sont à valeurs dans un compact. Donc le supremum sur G , ou encore sur $\mathbb{R}^d$, des $||R_n(x) - R(x)||_{\mathbb{R}^d}$ tend p.s. vers 0 , le corollaire en résulte.

Le résultat précédent concernant la convergence p.s. de $R_n(X_n^{(r)})$, nous proposons maintenant, comme annoncé, un article relatif à la convergence en moyenne du prédicteur, dans des hypothèses nettement différentes de celles que l'on a utilisées jusqu'ici.

CONVERGENCE DU PREDICTEUR SOUS
DES HYPOTHESES DE MELANGE

(C.VII.II)

Stochastic Processes and their Applications 19 (1985) 271-280
North-Holland

NONPARAMETRIC PREDICTION OF A HILBERT SPACE VALUED RANDOM VARIABLE

D. BOSQ and M. DELECROIX

U.E.R. de Mathématiques Pures et Appliquées (Bât. M2), Université des Sciences et Techniques de Lille I, France

Received 1 June 1983
Revised 2 April 1984

Let ξ_t, $t \in \mathbf{R}$, be a Markovian, measurable, strictly stationary process taking values in a measurable space $(E, \mathscr{B})$, and g a mapping from E into a separable Hilbert space H. A statistical nonparametric predictor of $g(\xi_{T+h})$ is studied in the paper. That predictor, based on the observations of the process between the times O and T generalizes the 'predictogram'; its asymptotic consistency is proved and some applications are given.

Markov processes * nonparametric prediction * regression estimators

1. Introduction

1.1. Defining the problem

Let (ξ_t), $t \in \mathbb{R}$, be a Markovian, measurable, strictly stationary process, defined on a probability space $(\Omega, \mathscr{A}, P)$, and observed between the times 0 and T. Let $(E, \mathscr{B})$ be its value-space, and g a mapping from $(E, \mathscr{B})$ into a separable Hilbert space H. We wish to predict the value of $g(\xi_{T+h})(h > 0)$ from a statistical point of view, from the observations, assuming only knowledge of the common distribution μ of the ξ_t. We shall use a predictor generalizing the classical 'predictogram' (cf. [4]) and based on the stationarity of the (ξ_t, ξ_{t+h})'s distributions. Indeed, that investigation seems to be of some interest even when the process-law is known exactly, for the construction of the probabilistic predictor, theoretically determined in that case, is sometimes excessively complicated.

1.2. A natural estimator of regression

In the sequel, we shall denote by R the conditional expectation of $g(\xi_h)$ relative to ξ_0, so that, for every point x of E, we have:

$$R(x) = E[g(\xi_h) \,|\, \xi_0 = x] = E[g(\xi_{t+h}) \,|\, \xi_t = x],$$

for every t of $\mathbf{R}^+$, because of the stationarity of (ξ_t, ξ_{t+h})'s distributions. Considering then a family (K_T), $T \in \mathbb{R}^+$, of Borelian, real valued, and bounded functions defined on E^2, we shall define an estimate $\hat{R}_T(x)$ of $R(x)$ by

$$\hat{R}_T(x) = (T-h)^{-1} \int_0^{T-h} K_T(\xi_t, x) g(\xi_{t+h})\, dt \quad (T > h). \tag{1}$$

For each t and ω, $(\omega \in \Omega)$, $K_T(\xi_t(\omega), x) g(\xi_{t+h}(\omega))$ denotes the product of the real $K_T(\xi_t(\omega), x)$ by the element $g(\xi_{t+h}(\omega))$ of H, so that $K_T(\xi_t, x) g(\xi_{t+h})$ represents an H-valued random variable that will be Bochner integrable if we assume the Bochner-integrability of $g(\xi_{t+h})$ (K_T is bounded). Thus, $\hat{R}_T(x)$ is well defined in (1).

To see then that $\hat{R}_T(x)$ is a somewhat natural estimate of $R(x)$, it is enough to choose as an example the simplest family of functions K_T we can use, corresponding to the 'regressogram'. In that case, on E^2, we define $K_T(x, y)$ as:

$$\sum_{i=1}^{h(T)} (\mu(A_{i,T}))^{-1} I_{A_{i,T}}(x) I_{A_{i,T}}(y).$$

($h(T)$ is some sequence of integers such that $\lim_{T\to\infty} h(T) = \infty$, $(A_{1,T}, \ldots, A_{h(T),T})$ some partition of E, for every T). Assuming that $\sup_{1 \le i \le h(T)} \mu(A_{i,T}) \to 0$ as $T \to \infty$, the same meaning of (1) is quite clear: $\hat{R}_T(x)$ selects the only ξ_t sufficiently near to x, and then takes the mean of the corresponding $g(\xi_{t+h})$.

Moreover, K_T families usually used in density-estimation techniques (cf. [5] for example, for a list), more or less based on the same idea as the regressogram can also be used. We shall give more details later.

1.3. The predictor of $g(\xi_{t+h})$

Considering the Markovian character of the process, we define the predictor of $g(\xi_{T+h})$ as $\hat{R}_T(\xi_T)$. In the sequel, we shall prove that $\hat{R}_T(\xi_T)$ is asymptotically consistent for $g(\xi_{T+h})$, for the H distance and give some applications of the results we obtain. See [1] for a study of the same problem in the case of a discrete time process and [2] for applications.

2. Technical preliminaries

2.1. Prediction error

This will obviously be defined as

$$E_T = E(\|\hat{R}_T(\xi_T) - g(\xi_{T+h})\|^2),$$

where $\|\cdot\|$ denotes the H-norm. Now, we can write

$$E_T = E(\|\hat{R}_T(\xi_T) - R(\xi_T)\|^2) + E(\|R(\xi_T) - g(\xi_{T+h})\|^2),$$

and, as the second term of that sum, the probabilistic prediction error, is fixed, we shall say that $\hat{R}_T$ is consistent if the limit of the statistical prediction error is 0, that is

$$\Delta'_T \xrightarrow{T\to\infty} 0, \qquad \Delta'_T = E(\|\hat{R}_T(\xi_T) - R(\xi_T)\|^2). \tag{2}$$

Indeed, if the Markovian hypothesis were omitted in the following proofs, condition (2) would still be obtained, under slight modifications, but it would not be enough to get an asymptotic efficiency of the predictor.

2.2. Choosing the K_T

For every point x in E, we have by stationarity that

$$\begin{aligned} E(\hat{R}_T(x)) &= E[K_T(\xi_0, x)g(\xi_h)] = E[K_T(\xi_0, x)R(\xi_0)] \\ &= \int_E K_T(y, x)R(y)\,\mathrm{d}\mu(y) = [K_T R](x). \end{aligned} \tag{3}$$

Obviously the product used here in $K_T(\xi_0, x)g(\xi_h)$ for example, keeps the same meaning as in (1). Conversely, $[K_T R]$ represents the transform of R by the integral operator associated to K_T, as defined in the third formula of (3).

It clearly appears then, from (3), that the minimal condition we need to obtain (2) is the convergence to the $L^2(\mu)$ identity of the operators associated with the (K_T) on the set $\mathcal{R}_E$ of all possible regressions of $g(\xi_h)$ on ξ_0. Indeed, to obtain (2), we shall need more, and choose a family of functions (K_T) satisfying

$$\lim_{T\to\infty} E(\|[K_T \cdot R](\xi_T) - R(\xi_T)\|^2) = 0. \tag{C}$$

Thus, our predictor will finally be consistent if

$$\lim_{T\to\infty} \Delta_T = 0, \qquad \Delta_T = E(\|\hat{R}_T(\xi_T) - [K_T R](\xi_T)\|^2). \tag{4}$$

2.3. Remarks on (C)

We shall only note that:

(a) Condition (C) is true for many functions used in classical density estimation (cf. [5]) and large classes of regressions.

First, if we call (h_j), $j \in \mathbb{N}$, an orthonormal basis of (H), and $\mathcal{R}$ the set of all the mappings: $x \to \langle R(x), h_j\rangle$ ($\langle\cdot\,,\cdot\rangle$ denotes H scalar product), with $j \in \mathbb{N}$, $R \in \mathcal{R}_E$, the strong convergence of the operators associated with functions K_T, to $L^2(\mu)$-identity, on $\mathcal{R}$, implies (C) for every regression R of $\mathcal{R}_E$.

Then, we can see that strong convergence of these operators to the $L^2(\mu)$ identity, on very large classes of real-valued functions is quite easily obtained when using a regressogram as defined in Section 1.2 or the other classes of functions usual in density estimation. That is:

(α) The Parzen–Rosenblatt kernels, for which

$$K_T(x, y) = [h(T)]^{-1} K[(x-y)(h(T))^{-1}]$$

($\lim_{T\to\infty} h(T) = 0$, K is a probability density).

(β) The orthogonal functions kernels, for which

$$K_T(x, y) = \sum_{i=1}^{h(T)} \lambda_{i,T} e_i(x) e_i(y).$$

($\lim_{T\to\infty} h(T) = \infty$, $(e_i)_{i\in\mathbb{N}}$ is an orthonormal $L^2(\mu)$-basis, simplest weights $\lambda_{i,T}$ can be chosen identically equal to 1).

(b) Condition (C) shows the necessity of knowing the common distribution of the ξ_t when choosing the K_T. But, in practice, if μ is not known, it is not usually difficult to estimate; see [5] for that problem.

2.4. *Decomposition of K_T*

As K_T is the kernel of a Hilbert–Schmidt operator on $L^2(\mu)$, there exists in $L^2(\mu)$ an orthonormal sequence f_{jT} such that:

$$K_T = \sum_{j\in J_T} \alpha_{jT} f_{jT} \otimes f_{jT} \tag{5}$$

(J_T is finite or equal to $\mathbb{N}$). We assign moreover to K_T the additional hypothesis

$$\sum_{j\in J_T} |\alpha_{jT}| M_{jT}^2 < \infty, \qquad M_{jT} = \sup_{x\in E} |f_{jT}(x)|. \tag{6}$$

Clearly condition (6) is more stringent than the usual $L^2(\mu)$ convergence of the series introduced in (5) and even more than the uniform convergence used in the famous Mercer theorem. But it is automatically satisfied in the case of a regressogram, and more generally for most of the orthogonal functions kernels introduced in the last paragraph (formula β)). It is then sufficient to choose orthogonal polynomials (e_i) which are uniformly bounded, for the condition $\sum_j |\alpha_{jT}| < \infty$ is always true!

On the other hand, condition (6) is quite difficult to verify for Parzen–Rosenblatt kernels, introduced in formula (α), for which our method of decomposition seems to be less well adapted. Even in that case, however, solutions can be found in particular cases. Assuming, for example, μ to be the uniform distribution on $[-\pi, \pi]$, from the expansion of any even bounded density K in Fourier series, we can derive a decomposition of $K(x-y)$ in a series of products of the type '$\cos px \cos py$', or '$\sin px \sin py$' and then, directly computing the α_{jT}, we can obtain (6) for many Parzen–Rosenblatt kernels. See [3] for an improvement of condition (6).

2.5. Decomposition of Δ_T

We can now write

$$\hat{R}_T(\xi_T) = (T-h)^{-1} \int_0^{T-h} \left\{ \sum_{j \in J_T} [\alpha_{jT} f_{jT}(\xi_t) f_{jT}(\xi_T) g(\xi_{t+h})] \right\} \mathrm{d}t,$$

$$[K_T R](\xi_T) = \int_E \left\{ \sum_{j \in J_T} [\alpha_{jT} f_{jT}(x) f_{jT}(\xi_T)] \right\} R(x) \, \mathrm{d}\mu(x),$$

and thus

$$\hat{R}_T(\xi_T) - [K_T R](\xi_T) = \sum_{j \in J_T} \alpha_{jT} f_{jT}(\xi_T)(\hat{r}_{jT} - r_{jT})$$

with

$$\hat{r}_{jT} = (T-h)^{-1} \int_0^{T-h} f_{jT}(\xi_t) g(\xi_{t+h}) \, \mathrm{d}t,$$

$$r_{jT} = \int_E f_{jT}(x) R(x) \, \mathrm{d}\mu(x) = E[\hat{r}_{jT}]. \tag{7}$$

So, we have

$$\Delta_T = \sum_{j', j \in J_T} \alpha_{jT} \alpha_{j'T} I_{jj'T}$$

where

$$I_{jj'T} = (T-h)^{-2} E\left\{ \left\langle \int_0^{T-h} [f_{jT}(\xi_T) f_{jT}(\xi_t) g(\xi_{t+h}) - r_{jT}] \, \mathrm{d}t, \right.\right.$$
$$\left.\left. \int_0^{T-h} [f_{j'T}(\xi_T) f_{j'T}(\xi_t) g(\xi_{t+h}) - r_{j'T}] \, \mathrm{d}t \right\rangle \right\} \tag{8}$$

and finally, assuming that $\int_E \|g(y)\|^2 \, \mathrm{d}\mu(y) < +\infty$,

$$I_{jj'T} = (T-h)^{-2} \int_0^{T-h} \int_0^{T-h} I_{jj'T}^{tt'} \, \mathrm{d}t \, \mathrm{d}t'$$

with

$$I_{jj'T}^{tt'} = E[f_{jT}(\xi_T) f_{j'T}(\xi_T) \langle f_{jT}(\xi_t) g(\xi_{t+h}) - r_{jT}, f_{j'T}(\xi_{t'}) g(\xi_{t'+h}) - r_{j'T} \rangle]. \tag{9}$$

From (8) and (9) we shall now prove our main results.

3. Main results

3.1. Preliminary lemma

We shall use a lemma of Gastwirth and Rubin (cf. [6, p. 817]) adapted to our study. For that, let us consider W and Z, two random variables taking values in spaces F and G respectively, such that a conditional distribution P_Z^W of Z given W exists. For any point y of F, we put

$$\mu(y) = P_Z^{W=y} - P_Z,$$

and denote by $\Delta(y)$ the total variation of $\mu(y)$.

Let φ and ψ be two random variables taking values in a Hilbert space H, such that $E(\|\varphi\|^q)<\infty$ and $E(\|\psi\|^r)<\infty$ for two numbers q and r belonging to $[1,\infty]$. Then we have

Lemma 1

$$E|\langle\varphi(W)-E[\varphi(W)], \psi(Z)-E[\psi(Z)]\rangle| \\ \leq 2^{1/q}E(\|\varphi\|^q)^{1/q}E(\|\psi\|^r)^{1/r}\bar{D}^{q,r}(W,Z)$$

where

$$\bar{D}^{q,r}(W,Z) = \left[\int_F (\Delta(y))^{s/r+1}\,\mathrm{d}P_W\right]^{1/s} \quad (s^{-1}+q^{-1}+r^{-1}=1).$$

Indeed, in the original version of Gastwirth and Rubin, Lemma 1 is proved for real valued variables φ and ψ. Extension to the case of H-valued variables is straightforward, and based on two classical results for such variables:

$$|E(\langle\varphi,\psi\rangle)| \leq E(\|\varphi\|\,\|\psi\|) \quad \text{and} \quad E(\langle\varphi,h\rangle) = \langle E(\varphi),h\rangle, \quad h\in H.$$

We shall apply the lemma, assuming here the existence of a number q belonging to $[2,\infty]$ such that $\int \|g(y)\|^q\,\mathrm{d}\mu(y) = E[\|g(\xi_t)\|^q]<\infty$. Then, if we choose W as ξ_{t_1+h}, and Z as $\{\xi_{t_2+l}, l\geq 0\}$ (taking values in $E^{\mathbb{R}^+}$), with $t_1+h<t_2$, we can write, considering the Markovian character of the process

$$|I_{jj'T}^{tt'}| \leq 2^{1/q}D[t_2-(t_1+h)]E(\|\varphi_0\|^q)^{1/q}E(\|\Psi_0\|^q)^{1/q}, \tag{10}$$

defining, in (10), $D[t_2-(t_1+h)]$ as the quantity $\bar{D}^{q,q}(W,Z)$ introduced in Lemma 1, and φ_0 and ψ_0 respectively as

$$\varphi_0 = f_{jT}(\xi_{t_1})g(\xi_{t_1+h}) - r_{jT}$$

and

$$\psi_0 = f_{jT}(\xi_T)f_{j'T}(\xi_T)[f_{j'T}(\xi_{t_2})g(\xi_{t_2+h}) - r_{j'T}].$$

Note that, q being fixed, the process stationarity ensures that the quantity D depends only on the difference $(t_2-(t_1+h))$, and the hypothesis made on g ensures that $E(\|\varphi_0\|^q)$ and $E(\|\psi_0\|^q)$ are finite.

Now we can prove

3.2. Proposition 2

Proposition 2. *Under the following hypotheses,*

(a) *the process (ξ_t) is strictly stationary, measurable, and Markovian,*

(b) *g is such that $E(\|g(\xi_0)\|^q)<\infty$ for a number q belonging to $[2,\infty]$, and the numbers $D(t)$ defined in* (10) *satisfy $\int_0^\infty D(t)\,\mathrm{d}t<\infty$,*

(c) *the functions K_T satisfy C and Condition 6,*

(d) $\lim_{T\to\infty}(T^{-1})\{\sum_{j\in J_T}|\alpha_{jT}|M_{jT}^2\}^2=0$, *the predictor $\hat{R}_T(\xi_T)$ is consistent.*

Proof. (a) Let us consider t and t' such that $t+h<t'$. Then from (10) we have, by the definition of the M_{jT},

$$\begin{aligned}|I_{jj'T}^{tt'}| &\le 2^{1/q}M_{jT}M_{j'T}D[t'-(t+h)]\times E[\|f_{jT}(\xi_0)g(\xi_h)-r_{jT}\|^q]^{1/q}\\ &\quad\times E[\|f_{j'T}(\xi_0)g(\xi_h)-r_{j'T}\|^q]^{1/q}\\ &\le 4 2^{1/q}M_{jT}^2M_{j'T}^2D[t'-(t+h)]\times E[\|g(\xi_h)\|^q]^{2/q}.\end{aligned}$$

For t and t' such that $t'+h<t$, we obtain the same bound with $D[t-(t'+h)]$ and if $|t-t'|<h$, Schwarz inequality gives $|I_{jj'T}^{tt'}|\le 4M_{jT}^2M_{j'T}^2E(\|g(\xi_h)\|^2)$. Then, a fortiori, $|I_{jj'T}^{tt'}|\le K$ with $K=4\cdot 2^{1/q}M_{jT}^2M_{j'T}^2E(\|g(\xi_0)\|^q)^{2/q}$.

(b) Thus, the quantity $|\int_0^{t-n}\int_0^{t-n}I_{jj'T}^{tt'}\,\mathrm{d}t\,\mathrm{d}t'|$ is clearly less than

$$\begin{aligned}&K\left\{\int_0^h \mathrm{d}t\left[\int_0^{t+h}\mathrm{d}t'+\int_{t+h}^{T-h}D[t'-(t+h)]\,\mathrm{d}t'\right]\right\}\\ &+K\left\{\int_h^{T-2h}\mathrm{d}t\left[\int_0^{t-h}D[t-(t'+h)]\,\mathrm{d}t'+\int_{t-h}^{t+h}\mathrm{d}t'\right.\right.\\ &\left.\left.+\int_{t+h}^{T-h}D[t'-(t+h)]\,\mathrm{d}t'\right]\right\}\\ &+K\left\{\int_{T-2h}^{T}\mathrm{d}t\left[\int_0^{t-h}D[t-(t'+h)]\,\mathrm{d}t'+\int_{t-h}^{T-h}\mathrm{d}t'\right]\right\}\end{aligned}$$

and we obtain

$$\left|\int_0^{T-h}\int_0^{T-h}I_{jj'T}^{tt'}\,\mathrm{d}t\,\mathrm{d}t'\right|\le K\left\{2h(T-h)+2(T-h)\int_0^{T-h}D(t)\,\mathrm{d}t\right\}.$$

(c) Finally, from (9), we have

$$|I_{jj'T}|\le(T-h)^{-1}\left\{h+\int_0^{T-h}D(t)\,\mathrm{d}t\right\}8\cdot 2^{1/q}M_{jT}^2M_{j'T}^2E[\|g(\xi_0)\|^q]^{2/q} \quad (11)$$

and

$$\Delta_T = 0\left\{\left(\int_0^{T-h} D(t)\,\mathrm{d}t\right)(T-h)^{-1}\left(\sum_{(j,j')\in J_T^2} |\alpha_{jT}||\alpha_{j'T}|M_{jT}^2 M_{j'T}^2\right)\right\}.$$

Asymptotic consistency of $\hat{R}_T(\xi_T)$ under the hypothesis of Proposition (2) then follows.

3.3. Some remarks

(a) Condition $\int_0^\infty D(t)\,\mathrm{d}t < \infty$ is satisfied in practice for typical autoregressive, 'φ-mixing', and 'strong-mixing' processes (in that case $q = +\infty$, cf. [5]).

(b) It is quite easy to improve Proposition 2 in the case of the regressogram. Defining $K_T(x, y)$ as in Section 1.2, we obtain functions f_{jT} of formula (6), as the functions $(\mu(A_{i,T}))^{-1/2} I_{A_{i,T}}$, $i = 1, \ldots, h(T)$, so that (6) is automatically satisfied. Moreover we can write

Proposition 3 (*the regressogram case*). *Defining K_T as*

$$K_T(x, y) = \sum_{i=1}^{h(T)} [\mu(A_{i,T})^{-1}] I_{A_i}(x) I_{A_i}(y), \tag{12}$$

where $h(T)$ is some sequence of integers growing to $+\infty$, and for each T, $(A_{i,T})$, $i = 1, \ldots, h(T)$, a finite partition of E, chosen such that (C) is satisfied and $(\sum_1^{h(T)} [\mu(A_{i,T})^{-1}])^2 = o(T)$, the predictor $\hat{R}_T(\xi_T)$ is consistent under the hypothesis (a) *and* (b) *of Proposition* 2.

(c) In the same way, in the case of orthogonal function kernels, as defined in Section 2.3, formula β), Proposition 2 can be improved to

Proposition 4 (*the orthogonal functions case*). *Let $(e_i)_{i\in\mathbb{N}}$, be an orthonormal $L^2(\mu)$-basis such that* $\sup_i \sup_x |e_i(x)|$ *is finite, and define K_T by*

$$K_T(x, y) = \sum_{i=1}^{q(T)} e_i(x) e_i(y), \tag{13}$$

where $q(T)$ is some growing sequence of integers verifying $q(T) = o(T^{1/2})$. Assuming again that (C) *is satisfied as well as hypotheses* (a) *and* (b) *of Proposition* 2, *the predictor $\hat{R}_T(\xi_T)$ is consistent.*

(d) In the cases studied in Propositions 3 and 4, we can compute more exactly the order of magnitude of Δ'_T, refining particularly the calculus of $E(\|[K_T R](\xi_T) - R(\xi_T)\|)$, from K_T's definition. One can see [3] for some results of that kind.

4. Applications

(1) Let us suppose the variables ξ_t are real-valued and let g be the mapping from $\mathbf{R}$ into $L^2(\nu): x \to I_{[x,\infty[}$ (ν denotes a bounded measure on $(\mathbf{R}, \mathcal{B}_{\mathbf{R}})$) $\hat{R}_T(\xi_T)$ is then

formally written as the function

$$\hat{R}_T(\xi_T): u \to (T-h)^{-1} \int_0^{T-h} K_T(\xi_t, \xi_T) I_{[\xi_{t+h}, \infty[}(u)\, dt$$

and R, defined as $E[g(\xi_h) | \xi_0]$, satisfies

$$R(u) = E[g(\xi_h) | \xi_0 = u] = E\{[g(\xi_h)](u) | \xi_0\} = [I_{\{\xi_h < u\}} | \xi_0],$$

so that our predictor will be an estimator of that conditional distribution function.

Noting that $\|g\|$ is a bounded function ($\|g(x)\| = \{\nu([x, \infty[)\}^{1/2}$), we can define the numbers $D(t)$ introduced in (10) with $q = \infty$, and thus, they are coefficients of 'strong mixing' (cf. [5]). Finally, if we suppose that $\int_0^\infty D(t)\, dt < \infty$, and choose K_T as in (13), assuming that (e_i), $i \in \mathbb{N}$, is a usual $L^2(\mathbb{R})$-basis (Hermite polynomials for example), we obtain as in Proposition 4 that

$$\lim_{T \to \infty} \|\hat{R}_T(\xi_T) - R(\xi_T)\|_{L^2(\nu)} = 0$$

if $\lim_{T\to\infty} q^2(T)/T = 0$.

(2) Let $(E, \mathscr{B})$ be a Polish space equipped with its Borel σ-field, $\mathscr{M}(\mathscr{B})$ the set of bounded signed measures defined on $(E, \mathscr{B})$ and $\mathscr{P}(\mathscr{B})$ the set of probabilities on $(E, \mathscr{B})$. Then there exists a bounded, $\mathscr{B} \otimes \mathscr{B}$ measurable, reproducing kernel, $\mathscr{N}$, such that

$$\langle \lambda_1, \lambda_2 \rangle = \int_{E \times E} \mathscr{N}(x, y)\, d\lambda_1(x)\, d\lambda_2(y) \quad (\lambda_1, \lambda_2 \in \mathscr{M}(\mathscr{B}))$$

is a scalar product on $\mathscr{M}(\mathscr{B})$ which induces on $\mathscr{P}(\mathscr{B})$ a topology compatible with the weak topology and identify $\mathscr{M}(\mathscr{B})$ as a subspace of H_N, the reproducing kernel Hilbert space generated by $\mathscr{N}$. Finally, one can choose $\mathscr{N}$ so that H_N should be separable (cf. [7]).

If we define g by $g(x) = \delta_x$ $(x \in E)$, where δ_x denotes the Dirac measure on x, we can see that, for $\lambda \in \mathscr{M}(\mathscr{B})$,

$$E(\langle \delta_{\xi_h}, \lambda \rangle | \xi_0) = E\left[\int_E \mathscr{N}(x, \xi_h)\, d\lambda(x) | \xi_0\right] = \int_{E \times E} \mathscr{N}(x, y)\, d\lambda(x)\, d\nu^{\xi_0}(y)$$
$$= \langle \nu^{\xi_0}, y \rangle$$

where ν^{ξ_0} denotes the conditional distribution of ξ_h given ξ_0, and here, $\nu^{\xi_0} = R$. $\hat{R}_T(\xi_T)$ will be an estimator of that conditional distribution, defined as the discrete random measure:

$$\hat{R}_T(\xi_T) = (T-h)^{-1} \int_0^{T-h} K_T(\xi_t, \xi_T) \delta_{(\xi_{t+h})}\, dt.$$

Moreover $\|g\|$ is a.s. bounded ($\|g(x)\| = \sqrt{\mathscr{N}(x, x)}$) so that we can again define the numbers $D(t)$ with $q = +\infty$, and under the other hypothesis of Proposition 2, $\hat{R}_T(\xi_T) - R(\xi_T)$ will converge to zero for a norm compatible with the weak topology.

Acknowledgment

The authors are grateful to the two referees for their useful suggestions on the exposition of this paper.

References

[1] D. Bosq, Sur la prédiction non paramétrique de variables aléatoires et de mesures aléatoires. Z. Wahrsch. Verw. Geb. 64 (1983) 541-553.
[2] D. Bosq, Nonparametric prediction in stationary processes, in: Lecture Notes in Statistics n° 16, Specifying Statistic Models (Springer-Verlag) 69-84.
[3] D. Bosq, Sur la prédiction non paramétrique d'un processus stationnaire, Pub. Int. n° 150 - U.E.R. de Math. de l'Université de Lille I - Villeneuve d'Ascq (1979).
[4] G. Collomb, From non parametric regression to non parametric prediction: Survey of the mean square error and original results on the predictogram in Lecture Notes in Statistics n° 16, Specifying Statistic Models (Springer-Verlag) 182-204.
[5] M. Delecroix, Sur l'estimation des densités d'un processus stationnaire à temps continu, Pub. Inst. Stat. Univ. Paris 1980 - XXV - fasc. 1-2, 17-39.
[6] J.L. Gastwirth and H. Rubin, The asymptotic distribution theory of the empiric CDF for mixing stochastic processes, Ann. Statist. 3 (1975) 809-824.
[7] C. Guilbart, Etude des produits scalaires sur l'espace des mesures. Annales de l'I.H.P. (1979), Section B, Vol. XV n° 4, 333-354.

ETUDE PAR COMPARAISON DE L'EFFICACITE
DU PREDICTEUR

(C.VII.III)

PUB. IRMA - Lille, 1987
Vol. 9, n° II

COMPARAISON DE METHODES PARAMETRIQUE ET NON PARAMETRIQUE DE PREVISION (*)

par

Michel CARBON et Michel DELECROIX
Laboratoire de Probabilités et Statistique
U.F.R. de Mathématiques Pures et Appliquées
Université des Sciences et Techniques de Lille Flandres Artois
59655 - Villeneuve d'Ascq - Cedex (France)

Le travail présent consistera en une comparaison de 2 méthodes classiques de prévision dans les séries temporelles à temps discret.

a) La méthode dite de BOX-JENKINS consistant à approcher le processus par un modèle ARIMA adapté.

b) Une méthode non paramétrique consistant à prévoir X_{T+h} par une moyenne pondérée de valeurs X_{t+h} , $(t = T-h-1, \dots, k)$, la pondération tenant compte de la proximité de X_t , X_{t-1} , ..., X_{t-p} , avec $(X_T, X_{T-1}, \dots, X_{T-p})$.

Cette comparaison s'effectuera à partir d'exemples divers de séries temporelles traitées par les deux méthodes.

Mots clés : Prédiction
Méthodes non paramétriques
Méthode Box et Jenkins

Classification AMS : 62M20 - 60G25 - 62J02 - 60G10

(*) Preprint

COMPARAISON EMPIRIQUE DE PREDICTEURS

Considérons la trajectoire observée $(x_1,\dots,x_T)$ d'un processus aléatoire réel $(X_t)_{t\in\mathbb{Z}}$. L'étude qui suit se propose d'évaluer empiriquement les performances d'un prédicteur non-paramétrique de la variable X_{T+k} noté L_T, construit à partir des valeurs observées $x_1,\dots,x_T$, par comparaison avec le prédicteur usuel utilisé dans la méthode de Box-Jenkins.

(A) - Le prédicteur L_T : aperçu théorique

a) Principe de construction

Le prédicteur retenu généralise la classique méthode du lissage exponentiel. Dans celle-ci, on prévoit X_{T+k} par une moyenne pondérée déterministe des variables X_T , X_{T-1} ,..., X_1 :

$$\sum_{j=0}^{T-1} (1-\beta)\,.\,\beta^j\ X_{T-j}$$

la pondération étant donc forte pour les dernières valeurs observées et faible pour les premières ... Le prédicteur étudié ici est aussi une moyenne pondérée :

$$L_T = \sum_{j=r}^{T-k} \hat{\alpha}_{t,T}\,.\,X_{t+k}\quad, \tag{1}$$

des X_{t+k}, mais la pondération aléatoire affectée à X_{t+k} est fonction de la proximité de la dernière série de r valeurs observées $X_T^{(r)} = (X_T,\dots,X_{T-r+1})$ avec le r-uple $X_t^{(r)} = (X_t,\dots,X_{t-r+1})$, r étant un entier à disposition du statisticien.

Dans le classique "prédictogramme" $\hat{\alpha}_{t,T}$ vaut simplement :

$$\hat{\alpha}_{t,T} = \{\prod_{j=1}^{r} I_{[-h(T),h(T)]}(X_{t-j+1}-X_{T-j+1})\} \times$$

$$\left[\sum_{j=r}^{T-k} \{\prod_{j=1}^{r} I_{[-h(T),h(T)]}(X_{t-j+1}-X_{T-j+1})\}\right]^{-1}$$

pour un réel $h(T)$ à disposition du praticien, c'est-à-dire que l'on affecte à chaque X_{t+k} le poids $^1/_{k(T)}$, ou 0, selon qu'il fasse ou non partie des $k(T)$ X_{t+k} pour lesquels chaque composante de $X_t^{(r)}$ est à distance inférieure à $h(T)$ de la composante correspondante de $X_T^{(r)}$ ... On choisira "astucieusement" $h(T)$ pour que $k(T)$ ne vaille ni 0, ni $T-r-h$ ...

Nous considérerons un prédicteur qui généralise le précédent en choisissant comme poids $\hat{\alpha}_{t,T}$

$$(2) \qquad \hat{\alpha}_{t,T} = K\left[\frac{X_T^{(r)} - X_t^{(r)}}{h(T)}\right] \cdot \Big/ \{\sum_{t=r}^{T-k} K\left[\frac{X_t^{(r)} - X_T^{(r)}}{h(T)}\right]\}$$

où K est une densité de probabilité sur $\mathbb{R}^r$: dans le cas du prédictogramme on avait $K(x_1,\dots,x_r) = \prod_{j=1}^{r} \{\frac{1}{2} I_{[-1,1]}(x_j)\}$, et nous avons retenu pour l'étude pratique qui suit $K(x_1,\dots,x_r) = \prod_{j=1}^{r} (\frac{1}{\sqrt{2\pi}} e^{-x_j^2/2})$.

b) Survol des résultats de convergence

Les travaux effectués sur le sujet se basent sur l'idée suivante : à partir de (1) et (2) on peut écrire, en multipliant par les constantes "ad hoc"

$$L_T = \hat{R}_T(X_T^{(r)}) = N_T(X_T^{(r)}) \;/\; D_T(X_T^{(r)}) \quad ,$$

où

$$N_T((x)) = (T-r-k)^{-1} \cdot \sum_{t=r}^{T-k} X_{t+k} \quad \{h(T)^{-r} \cdot K\left[((x)-X_T^{(r)})/h(T)\right]$$

$$D_T((x)) = (T-r-k)^{-1} \cdot \sum_{t=r}^{T-k} (h(T))^{-r} \cdot K\left[((x)-X_t^{(r)}) \,/\, h(T)\right] \quad .$$

Or, dans le cas où le processus est strictement stationnaire et les (n)-uples $(X_t , X_{t-1} ,\ldots, X_{t-r+1})$ admettent une densité commune g , $D_T(x)$ n'est autre que le classique estimateur de $g(x)$, par la méthode du noyau, $N_T(x)$ en étant un de $R(x)\, g(x)$, où R est la régression de X_{t+k} sur $(X_t^{(r)})$:

$$R((x)) = E\left[X_{t+k} \,/\, (X_t^{(r)} = (x))\right] \qquad (x) \in \mathbb{R}^r$$

et dès lors, sous de bonnes conditions on obtiendra la convergence vers $R(x)$ de $\dfrac{N_T(x)}{D_T(x)}$ et par conséquent aussi

(3) $$\lim_{T\to\infty} \text{p.s.} \quad |\hat{R}_T(X_T^{(r)}) - R(X_T^{(r)})| = 0$$

Si le processus (X_t) , $t \in \mathbb{Z}$ est formé de v.a. strictement stationnaires φ-mélangeantes, et prenant leurs valeurs dans un compact de $\mathbb{R}$, $h(T)$ étant alors "astucieusement choisi" en fonction de T , (3) est ainsi vérifiée (cf. COLLOMB "Propriétés de convergence presque complète du prédicteur à noyau (Z. Wahrsch. 66 441-460) DELECROIX (1987) obtenant le même résultat si le processus est simplement ergodique, au lieu d'être φ-mélangeant ...

* Les résultats précédents montrent, concrètement, que si T est grand, et le processus $(X_t)_{t\in\mathbb{Z}}$ r-Markovien, et suffisamment régulier (ergodique, mélangeant etc ...), le prédicteur L_T sera très proche de

$\mathcal{R}(X_T^{(r)})$ c'est-à-dire de la meilleure information que le statisticien puisse espérer tirer des observations quant à la prédiction de X_{T+k} : $\mathcal{R}(X_T^{(r)}) = E^{\mathcal{F}_T}(X_{T+k})$, $\mathcal{F}_T$ étant la tribu engendrée par X_T , X_{T-1} ,...

En l'absence de l'hypothèse markovienne, il est clair que pour des processus suffisamment réguliers, si (r) est grand, l'approximation de $E^{\mathcal{F}_T}(X_{T+k})$ par $\mathcal{R}(X_T^{(r)})$ sera encore tout à fait convenable (l'étude théorique d'une suite de prédicteurs basés sur des (r_T)-uples $(X_T , \ldots, X_{T-r_T+1})$ de longueurs croissant lentement vers l'infini, reste à écrire mais leur convergence est plausible !).

Enfin, les affirmations précédentes n'ont de sens que si l'on choisit évidemment h(T) (fonction de T) comme prévu dans les théorèmes sus-cités ... Nous avons adopté le choix qui réalise la convergence la plus rapide de $D_T(x)$ vers g(x) , c'est-à-dire $h(T) = S_T / T^{1/r+1}$, S_T^2 représentant la variance empirique des observations (cf. Bosq et Lecoutre : "Estimation fonctionnelle", à paraître), et les simulations ont montré le caractère concrètement optimal de ce choix, très souvent indispensable à l'obtention de résultats sur l'ordinateur. On sait que, par contre, le choix du "noyau" K n'a guère d'influence sur les performances de D_T et N_T , donc $\mathcal{R}_T$ (même référence).

c) Construction d'un intervalle de prévision

Pour chaque valeur de t , nous pouvons écrire :

$$X_{t+k} = \mathcal{R}(X_t^{(r)}) + \varepsilon_t$$

ε_t étant, par définition de l'espérance conditionnelle $\mathcal{R}(X_t^{(r)})$, non corrélé avec celle-ci, et centré . Si le processus (X_t) , $t \in \mathbb{Z}$, est strictement stationnaire les ε_t sont de même loi. Si t_α^r est tel que $P(|\varepsilon_t| < t_\alpha^r) = \alpha$, l'intervalle de prévision choisi, au niveau α, sera alors, en théorie :

$$\left]R(X_T^{(r)}) - t_\alpha^r \; , \; R(X_T^{(r)}) + t_\alpha^r\right[$$

Pour estimer $t_\alpha^{(r)}$ on calcule pour chaque $t = r,\ldots,T$ l'écart :

$$\hat{\varepsilon}_t = |X_{t+k} - R_t^*(X_t^r)|$$

où R_t^* se détermine selon le même principe que $\hat{R}_T$, mais en tenant compte de toutes les suites $(X_{t'},\ldots, X_{t'-r+1})$ possibles extraites de l'échantillon pour $t' \neq t$, $t' \in \{r,\ldots,T\}$:

$$R_t^*(X_t^r) = \sum_{t'\neq t} X_{t'+k} \cdot K\left[(X_t^{(r)}-X_{t'}^{(r)})/h(T)\right] \left\{\sum_{t'\neq t} K\left[(X_t^{(r)}-X_{t'}^{(r)})/h(T)\right]\right\}^{-1}$$

Les $\hat{\varepsilon}_t$ sont donc T-r estimations de la loi commune des $|\varepsilon_t|$, en admettant que $R_t^*(X_t^r)$ soit pratiquement égal à $R(X_t^{(r)})$, et on estime t_α^r par le quantile empirique du niveau α des $\hat{\varepsilon}_t$, c'est-à-dire un nombre $\hat{t}_\alpha^r$ tel que α % des $\hat{\varepsilon}_t$ soient inférieurs à ce nombre, procédé qui permet aussi d'éliminer les valeurs aberrantes éventuelles ; cette méthode empirique n'admet pas encore de justification théorique complète mais offre l'avantage de n'exiger aucune hypothèse sur la loi des ε_t .

Finalement on retiendra comme intervalle concret :

$$\left]L_T - \hat{t}_\alpha^r \; , \; L_T + \hat{t}_\alpha^r\right[\; .$$

Ⓑ - Résultats observés, comparaison avec la méthode de Box-Jenkins

a) Principe de la comparaison

Pour juger des performances de la méthode proposée, nous l'avons mise en oeuvre sur séries simulées ou extraites d'ouvrages classiques traitant de la prévision par la méthode de Box-Jenkins. Pour chaque série, on a prévu les k dernières valeurs observées à partir des $(N-k)$ premières :

* Par la méthode de Box et Jenkins, au besoin pour plusieurs choix de paramètres dans les modélisations ARMA ou SARIMA possibles pour la série, après identification. On a repris parmi eux les choix proposés comme optimaux par les ouvrages de référence, lorsque c'était possible, ou bien ceux qui paraissaient s'imposer (Identifier un AR(1) simulé par un AR(1) théorique !)

* A l'aide du prédicteur L_T défini auparavant, pour diverses valeurs de r, y compris ici encore celles qui semblent naturelles ($r=1$ pour un AR(1)).

On calcule ensuite deux indicateurs d'écart pour chacune des théories étudiées (on note $\hat{X}_{N-k+j}$ la valeur prévue et X_{N-k+j} la vraie valeur) :

* une "erreur observée" (ERMO) moyenne arithmétique des divers quotients :

$$\frac{|\hat{X}_{N-k+j} - X_{N-k+j}|}{\hat{X}_{N-k+j}} .$$

* une "erreur de prévision" (ERMP) : moyenne arithmétique des demi-longueurs d'intervalles de prévisions divisées par la valeur prévue.

Pour la prévision $\hat{X}_{N-k+j}$ de X_{N-k+j}, on définit cet écart comme quotient de la demi-largeur de l'intervalle de prévision, par la valeur prévue ; on prend la moyenne des divers chiffres observés pour $j = 1,\dots,k$. Evidemment pour la méthode de Box et Jenkins, le calcul, intégré au programme, suppose la normalité des variables ε_t de variance σ^2 , et la demi-largeur retenue est alors fonction de 1,96 $\hat{\sigma}_T$; $\hat{\sigma}_T$ étant l'estimation de σ, effectués à partir des "résidus" (les $\hat{\varepsilon}_t$ définis ci-dessus) mais il s'agit bien de la même méthode d'évaluation théorique de ERMP qu'avec la méthode non-paramétrique. On pourra donc comparer rationnellement les résultats ...

Enfin, on a, au besoin, appliqué à la série (X_t) , $t \in \mathbb{Z}$, préalablement à la prédiction, les transformations nécessaires à la rendre stationnaire (différenciation, ou passage au Log ...) en la reconstituant ensuite : ce procédé usuel dans la méthode de Box - Jenkins n'avait jamais été utilisé pour la prédiction non-paramétrique.

b) *Commentaires des résultats*

Ils ont été regroupés dans les feuilles annexes.

Il est toujours difficile de tirer des conclusions définitives d'observations fractionnaires, mais il semble néanmoins possible, à partir des tableaux joints, de souligner les 4 points suivants.

1) Quand le processus est "bien stationnaire", au besoin après différentiation, il est également "bien prévu" par les deux méthodes, la même observation jouant à l'inverse : aucune méthode n'est "miraculeuse" par rapport à l'autre.

On notera d'ailleurs que la meilleure précision observée pour certaines séries tient essentiellement au type de valeurs prévues, grandes, qui diminuent l'erreur relative de prévision : dans les séries à forte fluctuation, prévoir des valeurs éloignées de la moyenne, si elle est proche de 0, amène ainsi à une (théorique !) grande précision alors que l'inverse se produit pour les valeurs proches de m ...

2) En comparant les prévisions optimales de chaque méthode, pour chaque série, il apparaît un léger avantage en faveur de la méthode non-paramétrique qui l'emporte 14 fois sur 17 pour l'ERMP, et parfois nettement, et 14 sur 17 pour l'ERMO. On notera d'ailleurs que les performances ERMO-ERMP ne concordent pas forcément, et que les séries qui devraient typiquement relever des modèles ARMA (AR(1) simulé etc ...) se prévoit finalement mieux par la méthode non-paramétrique : le prédicteur L_T paraît donc assez robuste et adaptable à beaucoup de situations.

3) Pour la méthode "non-paramétrique" les grandes valeurs de r fournissent en général les meilleurs résultats, ce qui paraît surprenant dans le cas des séries 1 ou 2-Markoviennes en théorie (AR1 - AR2). Nous ne savons pas interpréter du point de vue théorique ce phénomène qui peut résulter d'une meilleure convergence de L_T lorsque r est grand : il faudrait, à titre de vérification, évaluer $E((L_T - (X_t^{(r)}))^2)$.

On peut encore noter que si r est grand, L_T n'est plus guère que la "pondération" de 2 ou 3 (voire 1 !) valeurs X_{t+k} , les $\hat{\alpha}_{t,T}$ étant pratiquement tous négligeables en ce cas devant les plus grands d'entre-eux ... (ce qui pourrait amener à une nouvelle définition d'un estimateur en systématisant cette idée de ne sélectionner que les X_{t+h} "significatifs" en un sens à préciser).

4) Si la série présente une saisonnalité régulière, la méthode non-paramétrique de prédiction fonctionne très bien sans désaisonnalisation préalable, ce qui est logique intuitivement.

C - En guise de conclusion

Le travail présenté mérite encore d'être affiné, ne serait-ce qu'en multipliant les séries observées, de types les plus variés possibles. Nous pensons d'autre part pouvoir pour chaque série améliorer la prévision obtenue par les deux méthodes en utilisant des transformations préalables plus précises sur les séries observées (voire en pratiquant les techniques d'"intervention", comme le traitement des points aberrants etc ...).

Dans deux domaines particuliers, le prédicteur non-paramétrique peut aussi certainement s'améliorer, quoique les premiers essais en ce domaine n'aient guère été concluants.

* En introduisant une récursivité dans la prévision (prévoir X_{T+2} par $X_1 ,\dots, X_T$ et $\hat{X}_{T+1}$, puis déduire de $X_1 ,\dots, X_T$, $\hat{X}_{T+1}$, $\hat{X}_{T+2}$ la prévision de X_{T+3} , etc ...), méthode qui semble bien adaptée au cas markovien.

* En introduisant des pondérations tenant compte des corrélations calculées au préalable entre les variables, pour éliminer en quelque sorte, dans les $X_T^{(r)}$ dernières, celles qui n'influent guère sur X_{T+k} ...

Ces points de vue devraient être développés ultérieurement.

Nous soulignerons enfin l'aspect très simple de la méthode non-paramétrique, dont la mise en oeuvre est quasi-instantanée, pour des programmes de calcul très courts, et sa robustesse puisqu'elle résiste à la méthode a priori la mieux adaptée dans le cas de processus ARMA.

AR1

Série : X(t) = 0,9 X(t-1) + 1000,0 + Eps(t) où Eps(t) suit N(0,5)

n=95 5 prévisions (série simulée)

	Paramètres	ERMO en %	ERMP en %
B.J.	p = 1 ; q = 0	0,089	0,136
	p = 2 ; q = 0	0,089	0,136
	p = 0 ; q = 2	0,135	0,14
	p = 1 ; q = 1	0,089	0,136
N.P.	r = 1	0,098	
	r = 5	0,077	
	r = 10	0,085	
	r = 20	0,06	

AR1 limite

Série : X(t) = 0,99 X(t-1) + 1000,0 + Epst où Esp(t) suit N(0,5)

n=95 5 prévisions (série simulée)

	Paramètres	ERMO en %	ERMP en %
B.J.	d=1 ; p=0 ; q=1	0,0145	0,022
	d=1 ; p=0 ; q=2	0,013	0,022
	d=1 ; p=0 ; q=3	0,012	0,026
	d=1 ; p=0 ; q=4	0,012	0,024
N.P.	r = 5		0,04
	r = 10		0,03
	r = 20		0,02

MA6

$$X(t) = e(t) - 2{,}848\ e(t-1) + 2{,}6885\ e(t-2) - 1{,}64645\ e(t-3) + 2{,}972\ e(t-4) - 2{,}1492\ e(t-5) + 0{,}67716\ e(t-6) \text{ où } e(t) \sim N(0{,}5)$$

n=95 5 prévisions (série simulée)

	Paramètres	ERMO en %	ERMP en %
B.J.	p=0 ; q=6	3,03	5,328
	p=0 ; q=7	2,77	5,304
	p=1 ; q=0	3,04	5,3
	p=0 ; q=3	2,95	5,2
	p=0 ; q=2	3,06	5,2
	p=1 ; q=2	3,01	5,3
N.P.	r = 2	2,77	5,25
	r = 5	3,14	5,62
	r = 10	4,44	6,18
	r = 20	4,45	7,13
	r = 25	3,85	5,65

AR2

$$X(t) = 1000{,}0 + 0{,}7\ X(t-1) + 0{,}2\ X(t-2) + Eps(t) \text{ où } Eps(t) \text{ suit } N(0{,}5)$$

n=95 5 prévisions (série simulée)

	Paramètres	ERMO en %	ERMP en %
B.J.	p=2 ; q=0	0,0124	0,138
	p=1 ; q=0	0,0257	0,154
	p=3 ; q=0	0,0127	0,144
	p=0 ; q=5	0,068	0,158
	p=0 ; d=1 ; q=3	0,0256	0,148
N.P.	r = 2	0,014	0,143
	r = 5	0,02	0,15
	r = 10	0,024	0,146
	r = 20	0,02	0,079
	r = 30	0,015	0,074

ARMA(1,1)

$X(t) = 0,8\ X(t-1) + e(t) + 0,2\ e(t-1) + 1000,0$ où $E(t)$ suit $N(0,5)$

n=95 5 prévisions (série simulée)

	Paramètres	ERMO en %	ERMP en %
B.J.	p=1 ; q=0	0,149	0,282
	p=2 ; q=0	0,177	0,294
	p=1 ; q=1	0,1226	0,29
	p=1 ; q=2	0,17	0,296
	p=2 ; q=1	0,177	0,296
N.P.	r = 5	0,149	0,326
	r = 10	0,098	0,313
	r = 20	0,099	0,316
	r = 30	0,074	0,186

AR1 à Marge Exponentielle

$X(t) = 0,8\ X(t-1) + 1000,0 + eps(t)$ où $esp(t)$ suit $\exp(\frac{1}{300})$

n=95 5 prévisions (série simulée)

	Paramètres	ERMO en %	ERMP en %
B.J.	p=0 ; q=4	3,8	12,21
	p=0 ; q=2	3,75	11,7
	p=0 ; q=1	4,23	11,5
	p=1 ; q=0	1,6	11
	p=2 ; q=0	1,6	11,3
	p=3 ; q=0	1,8	11,1
	p=7 ; q=1	2,47	10,8
N.P.	r = 5	1,77	13,35
	r = 10	4,9	10,27
	r = 20	2,96	7,67
	r = 30	2,55	6,55

Série Perturbée

$$X_t = \varepsilon_t + \begin{cases} 0 & \text{avec la proba } 1/2 \\ \frac{1}{2} X_{t-1} & \text{avec la proba } 1/2 \end{cases} \quad \text{où} \quad \varepsilon_t = \begin{cases} \varepsilon_t & \text{avec proba } 2/3 \\ \frac{1}{4}\varepsilon_t & \text{avec proba } 1/3 \end{cases}$$

n=95 5 prévisions (série simulée)

	Paramètres	ERMO en %	ERMP en %
B.J.	p=1 ; d=0 ; q=0	153,8	218
	p=10; d=0 ; q=0	137,8	198,8
	p=7 ; d=0 ; q=0	146,5	213,5
	p=11; d=0 ; q=0	137,9	203,2
	p=3 ; d=0 ; q=0	152,5	214
N.P.	r = 5	80,13	
	r = 9	53,34	
	r = 10	51,93	
	r = 20	86,56	
	r = 30	64,5	

Série Perturbée n° 2

idem que la précédente remontée de 100

n=95 5 prévisions (série simulée)

	Paramètres	ERMO en %	ERMP en %
B.J.	p=1 ; q=0	1,625	3,568
	p=0 ; q=1	1,616	3,568
	p=7 ; q=0	1,824	3,608
	p=2 ; q=0	1,564	3,576
	p=0 ; q=3	1,492	3,578
N.P.	r = 5	1,99	4,02
	r = 10	1,71	4,12
	r = 20	1,69	6,41
	r = 30	1,85	4,51
	r = 40	2,02	3,58

Marges bénéficiaires (SARIMA)

n = 75 5 prévisions Pankratz (Wiley)

	Paramètres	ERMO en %	ERMP en %
B.J.	p=1 ; d=0 ; q= 0 ; P=2 ; D=1 ; Q=4 ; saison = 4	4,85	24,1
N.P.	r = 4	8,4	17,98
	r = 8	6,92	18,13
	r = 12	4,98	18,50
	r = 24	1,74	9,15

Consommation de cigares (SARIMA)

n = 90 ; 6 prévisions

Pankratz (Wiley)

	Paramètres	ERMO en %	ERMP en %
B.J.	p=1 ; d=1 ; q=0 ; P=1 ; D=2 ; Q=0 ; saison = 12	13,07	42,7
	p=2 ; q=0 ; P=1 ; D=1 ; Q=0 ; saison = 12	8,758	23,1
N.P.	r = 4	12,26	33,62
	r = 12	5,70	24,95
	r = 24	7,83	24,70

Change in business inventories

n = 50 10 prévisions Pankratz (Wiley)

	Paramètres	ERMO en %	ERMP en %
B.J.	p=1 ; q=0	36,98	156,9
	p=2 ; q=0	36,62	156,5
	p=3 ; q=0	39,12	172,3
	p=7 ; q=0	41,49	176,8
N.P.	r = 10	28,78	83,68
	r = 20	27,06	86,49

Série Perturbée

$$x(t) = 3000,0 \sin\left(\frac{\pi}{15} t\right) + \varepsilon(t) \quad \text{où} \quad \varepsilon(t) \quad \text{suit} \quad \exp\left(\frac{1}{300}\right)$$

n=195 5 prévisions (série simulée)

	Paramètres	ERMO en %	ERMP en %
B.J.	p=2 ; d=1 ; q=0 ; P=2 ; D=1 ; Q=0 ; saison = 30	21,88	47,1
	p=2 ; d=1 ; q=1 ; P=2 ; D=1 ; Q=1 ; saison = 30	25,08	40,7
N.P.	r = 30	12,07	43,065
	r = 60	11,6	39,56
	r = 15	11,89	40,20

Vitesse de rotation de la terre sur son axe polaire

n=29 1 prévision (série d'essai MULTICS)

	Paramètres	ERMO en %	ERMP en %
B.J.	d=1 ; p=1 ; q=0	16,83	53.64
	d=1 ; p=2 ; q=0	17,86	53,96
N.P.	r = 1	24,16	57,78
	r = 5	36,71	51,84
	r = 10	33,03	41,92

Charbon

n=80 10 prévisions Pankratz (Wiley)

	Paramètres	ERMO en %	ERMP en %
B.J.	p=1 ; q=0	3,83	23,6
	p=2 ; q=0	3,42	24,2
	p=1 ; q=1	3,52	23,9
N.P.	r = 2	3,14	20,73
	r = 5	4,04	20,58
	r = 10	3,5	23,30
	r = 25	6,41	13,18
	r = 30	5,45	14,24

Série : données d'un processus chimique

n=65   5 prévisions   Box-Jenkins (Wiley)

	Paramètres	ERMO en %	ERMP en %
B.J.	p=2 ; q=0	26,31	43,14
	p=0 ; q=5	28,20	42,78
	p=0 ; q=4	27,04	43,83
	p=0 ; q=3	26,15	43,52
	p=0 ; q=2	25,7	43,02
	p=0 ; q=1	26,26	43,19
	p=1 ; q=1	26,75	42, 9
N.P.	r = 5	23,10	54,55
	r = 10	29,88	57,49
	r = 20	35,01	52,95
	r = 30	22,89	58,94

Série : processus chimique de concentrations

n=187 10 prévisions Box-Jenkins (Wiley)

	Paramètres	ERMO en %	ERMP en %
B.J.	p=1 ; q=2	2,38	3,96
	p=1 ; q=1	2,48	4,01
	p=0 ; d=1 ; q=1	1,85	4,17
N.P.	r=5 ; d=0	2,71	4,12
	r=10 ; d=0	3,57	4,51
	r=20 ; d=0	2,36	4,72
	r=30 ; d=0	2,29	4,51
	r=40 ; d=0	2,75	5,31
	r=50 ; d=0	2,04	4,51
	r=75 ; d=0	2,79	3,366
	r=80 ; d=0	2,9	3,68
	r=1 ; d=1	2,11	4,225
	r=5 ; d=1	3,07	4,43
	r=10 ; d=1	1,91	4,43
	r=20 ; d=1	2,17	4,77
	r=30 ; d=1	2,81	4,93
	r=50 ; d=1	1,58	5,53
	r=75 ; d=1	2,66	5,75
	r=80 ; d=1	2,28	5,72

Etude série : X(t) = 0,9 . X(t-1) + esp(t) où eps suit U [-49,.49]

n=95 5 prévisions Kendall-Stuart (Griffin)

	Paramètres	ERMO en %	ERMP en %
B.J.	p=0 ; q=6	39,02	708,16
	p=0 ; q=5	33,38	549,29
	p=0 ; q=4	36,64	343,16
	p=0 ; q=3	43,79	332,61
	p=0 ; q=2	30,55	365,32
	p=0 ; q=0	41,84	286,14
N.P.	r=1	137,16	460
	r=5	456,04	1321
	r=10	63,49	253
	r=12	61,79	180
	r=20	131,09	395
	r=30	121,57	333

Série X(t) = 0,2 .X(t-1 + X(t-2) - 0,3 × (t-3) + eps(t) où eps suit U[-49,49]

n=95 5 prévisions Kendall-Stuart (Griffin)

	Paramètres	ERMO en %	ERMP en %
B.J.	p=5 ; q=0	61,16	244,14
	p=0 ; q=4	62,72	383,04
	p=0 ; q=3	44,61	461,88
	p=1 ; q=2	95,93	206,67
	p=1 ; q=0	52,87	197,17
	p=2 ; q=0	53,18	221,98
	p=3 ; q=0	70,05	204,24
	p=4 ; q=0	98,16	182,56
N.P.	r=1	115,69	323
	r=2	137,31	368
	r=10	529,2	630
	r=20	44,45	66,3
	r=30	45,99	59,74

CONCLUSION

Les résultats exposés au long du travail laissent évidemment en suspens de nombreux problèmes parmi lesquels nous pouvons citer, sans souci d'exhaustivité :

a) Une caractérisation plus précise des classes de processus vérifiant les conditions Bi) (donc pour lesquels les théorèmes obtenus s'appliquent) et une interprétation plus concrète de la limite g utilisée dans ces conditions.

b) L'étude des vitesses de convergence des estimateurs étudiés. Ce problème, comme le précédent, a déjà été évoqué dans l'introduction.

c) L'obtention des lois limites des estimateurs étudiés. Dans le cas d'un processus mélangeant nous avons déjà démontré quelques résultats (cf. Delecroix (I), (II), (III)) sur le sujet, et la seule hypothèse d'ergodicité ne semble pas a priori suffisante pour obtenir un théorème central-limite (cf. Rosenblatt (IV), chapitre (VI), pour le cas Markovien), mais le sujet reste ouvert. En particulier se pose le problème d'obtenir des approximations fortes des processus que constituent les estimateurs (cf. Stute (V) par exemple dans le cas de l'échantillon), jamais abordé, à notre connaissance, dans le cadre de nos hypothèses.

d) La convergence de l'estimateur de la régression d'un processus non borné (ou de la prédiction de ce processus). L'amélioration à apporter semble ici plus technique, réalisable à partir des méthodes utilisées par Härdle et Vieu (VI) ou Mack et Silvermann (VII) par exemple.

Arrêtant là l'énumération, nous dirons simplement espérer résoudre tout ou partie des problèmes évoqués ... ou laisser aux élèves que nous espérons diriger le soin d'y apporter des solutions !

Cela dit, il nous semble qu'au plan global, les parties 1-2-3 de la thèse concernant spécifiquement l'estimation de paramètres fonctionnels, et en particulier de la densité, réalisent un relatif achèvement sur le sujet : les conditions imposées au processus X_t , par exemple, ne semblent guère pouvoir être affaiblies. La $4^{\text{ème}}$ partie, par contre, nous paraît ouvrir un champ d'investigation beaucoup plus large pour le futur. Dans le domaine de la prédiction, qui intéresse en premier chef tous les utilisateurs de la statistique, l'introduction de méthodes non-paramétriques, simples, robustes et mathématiquement établies paraît susceptible d'amener un progrès réel, comme le montre la dernière partie du travail, consacrée aux comparaisons avec la méthode de Box-Jenkins.

Là aussi, le débat reste ouvert. Sans entrer trop avant dans le détail de suppositions que le quotidien de recherches futures risque de toutes façons de dévaluer, il semble bien qu'on puisse améliorer la méthode de prédiction apportée en y introduisant une certaine "dose" de modélisation paramétrique. On peut ainsi mélanger les méthodes en des phases successives permettant de travailler, à la fin, sur un processus stationnaire transformé du processus initial, avec un nombre T optimal de variables utilisées pour le prédiction, dans la régression de X_{t+h} sur $(X_t ,\dots, X_{t-T})$.

Cette statistique "semi-paramétrique" des processus, à notre connaissance embryonnaire devrait faire l'objet de travaux ultérieurs ... Nous espérons seulement les concrétiser par des résultats prolongeant de manière visible les théorèmes présentés dans cette thèse ! ...

BIBLIOGRAPHIE DE LA CONCLUSION.

I DELECROIX M. - Sur l'estimation des densités marginales et de transition d'un processus stationnaire et mélangeant. Thèse de 3ème Cycle.

II DELECROIX M. (1976) - Applications d'une inégalité de Stein à des lois limites d'estimateurs de la densité. (Pub. Inter, UER Math. Lille I, n° 93).

III DELECROIX M. (1976) - Lois limites pour des estimateurs de densité d'un processus stationnaire et L^2-mélangeant (Proceedings du Congrès de la Bernoulli Society, Grenoble).

IV ROSENBLATT M. (1971) - Markov Processes. Structure and asymptotic behavior. Springer-Verlag.

V STUTE W. (1982 b) - A law of the logarithm for kernel density estimators. Ann. Prob. 10, 414-426.

VI HÄRDLE et VIEU (1987) - Non-parametric prediction by the kernel method (à paraître).

VII MACK et SILVERMANN (1984) - Weak and strong uniform consistency of kernel regression estimators. Z. Wahrs. verw. Gebiete, 61, p. 405-415.

BIBLIOGRAPHIE GENERALE

[7] ADAMS Robert A. (1975) - Sobolev spaces.
Academic Press.

[6] ASH and GARDNER (1975) - Topics in stochastic processes.
Academic Press.

[15] BANON G. et HUNG NGUYEN T. (1981) - Recursive estimation in diffusion model.
S.I.A.M. I-control and optimization (19) n° 5, 676-685.

[16] BATTACHARAYA P.K. (1967) - Estimation of a probability density function and its derivatives.
Sankhya Series, Ser. A, 29, 373-382.

[13] BOSQ D. (1980) - Estimation de la densité par projection sur un sous-espace de dimension finie.
Portugaliae Math. 37, 1,2, 93-111.

[17] BOSQ D. (1970) - Contribution à la théorie de l'estimation fonctionnelle.
Pub. Inst. Stat. Univ. Paris, Vol. XIX 2, 1-96.

[25] BOSQ D. (1973) - Sur l'estimation de la densité d'un processus stationnaire et mélangeant.
C.R.A.S., Série A, tome 277, 535-538.

[32] BOSQ D. (1987) - Estimation d'un filtre linéaire d'ordre infini.
Pub. LSTA, Paris VI.

[29] COLLOMB (1984) - Propriétés de convergence presque complète du prédicteur à noyau.
Z. Wahrs. verw. Geb. 66, 441-460.

[24] DEHEUVELS P. (1974) - Conditions nécessaires et suffisantes de convergence ponctuelle presque sûre et uniforme presque sûre des estimateurs de la densité.
C.R.A.S., A, T. 278, p. 1217-1220.

[14] DELECROIX M. - Sur l'estimation des densités marginales et de transition d'un processus stationnaire et mélangeant.
Thèse de 3ème Cycle.

[30] DOUKHAN (1986) - Fonctions d'Hermite et statistique des processus mélangeants.
Collection "Approches non-paramétriques en analyse chronologique"
Institut des Hautes Etudes de Belgique p. 99-116.

[27] FOLDES A. (1974) - Density estimation for dependent sample.
Studia Scientarum Mathematicarum Hungarica 9, 443-452.

[3] FRITZ J. (1974) - Learning from an ergodic training sequence in limit theorems of probability theory.
Ed. Revesz North-Holland.

[18] GREBLICKY and PAWLAK (1984) - Hermite series estimates of a probability density and its derivatives.
J. Multavariate analysis, 15, 174-182.

[11] GUILBART C. (1978) - Etude des produits scalaires sur l'espace des mesures. Estimation par projection. Tests à noyaux.
Thèse Doctorat ès Sciences, Université des Sciences et Techniques de Lille.

[31] HÄRDLE et VIEU (1987) - Non-parametric prediction by the kernel method.
A paraître.

[4] KRENGEL U. (1985) - Ergodic theorems.
De Gruyter Studies in Mathematics, 6, Walter De Gruyter Berlin - New-York.

[33] MONFORT A. - GOURIÉROUX C. (1983) - Cours de séries temporelles.
Economica.

[1] NEVEU J. (1972) - Martingales à temps discret.
Masson, Paris.

[5] PARTHASARATHY (1967) - Probability measures on metric spaces.
Academic Press, New-York.

[22] PARZEN (1962) - On estimation of probability density function and mode.
A.M.S. 33, 1065-1076.

[26] PELIGRAD M. (1987) - Properties of uniform consistensy of the kernel estimates of density and of regression functions under dependance assumptions.
A paraître.

[20] REINHARD H. (1986) - Cours de mathématiques du signal.
Dunod Paris.

[23] ROSENBLATT M. (1956) - Remarques on some non-parametric estimates of a density function.
A.M.S. 27, 832-837.

[28] ROSENBLATT M. - Density estimates and Markov sequences in "Non-parametric techniques in statistical inference"
M. Puri, Ed, 199-219.

[8] RUDIN W. (1980) - Analyse réelle et complète.
Masson.

[12] SUQUET C. (1986) - Espaces autoreproduisants et mesures aléatoires.
Thèse de $3^{ème}$ cycle, soutenue à l'Université des Sciences et Techniques de Lille.

[21] SZEGÖ G. (1975) - Orthogonal polynomials.
American Mathematical Society Providence
Rhode Island.

[34] R. L. TAYLOR and TIEN-CHUNG HU (1987) - Kernel density estimator properties in $C_o(\mathbb{R})$.
(à paraître).

[9] TUAN D. PHAM et LANH. T. TRAN (1985) - Some mixing properties of time series models.
Stochastic processes and their applications.
297-303, North-Holland.

[2] VAN RYZIN (1969) - On strong consistensy of density estimate.
A.M.S. Vol. 40 n° 5, 1765-1772.

[19] WALTER G. (1977) - Properties of Hermite series estimation of a probability density.
Annals Statis. 5, 6, 1258-1264.

[10] WITHERS C.S. (1981) - Conditions for linear processes to be strong mixing.
Z. Wahrs. verw. Gebiete, 477-480.

R E S U M E

Le travail comprend quatre parties :

Dans la première partie, on étudie le comportement asymptotique de la somme de n éléments aléatoires $K_{t,n}(X_t)$, $1 \leq t \leq n$, forme sous laquelle s'écrivent les estimateurs fonctionnels classiques. Les (X_t) , $t \in Z$, forment un processus à valeurs dans $\mathbb{R}^d$ et les $K_{t,n}$ $(n \geq 1 , 1 \leq t \leq n)$ sont une suite de séquences à valeurs dans un Hilbert séparable. On démontre la convergence de la somme vers C.g , où C est la limite forte des sommes de n opérateurs intégraux associés aux $K_{t.n}$ et g la limite des densités de transition des variables X_t du processus par rapport à leur passé (i.e. la tribu engendrée par X_{t-1} , X_{t-2} ...). On montre que les processus simplement ergodiques vérifient usuellement les conditions imposées. On applique les théorèmes démontrés à l'estimation de la fonction de répartition et de la loi.

La deuxième partie est consacrée à l'étude des convergences L^1 et L^2 , presque sûre et en moyenne, d'estimateurs de la densité, récursifs ou non, par les méthodes du noyau et des fonctions orthogonales, dans les processus ergodiques.

On démontre dans la troisième partie la convergence uniforme des estimateurs de la densité par la méthode du noyau, sous l'hypothèse d'ergodicité, puis, sous des conditions plus fortes (faible mélangeance), on étudie leurs vitesses de convergence ponctuelle.

Dans la quatrième partie on étudie les estimateurs non-paramétriques de la régression dans les processus ergodiques, puis faiblement mélangeants. On en déduit un prédicteur non-paramétrique pour ces processus. On compare les performances de ce prédicteur avec celles des prédicteurs paramétriques usuels (Box et Jenkins) : la comparaison se base sur les intervalles de prévision fournis par chaque méthode, les calculs ayant été effectués sur de nombreuses séries simulées ou observées.

M O T S C L E S

- PROCESSUS ERGODIQUES
- ESTIMATION FONCTIONNELLE
- REGRESSION NON-PARAMETRIQUE
- ESTIMATION DE LA DENSITE
- PREDICTION NON-PARAMETRIQUE
- COMPARAISON DE PREDICTEURS
- STATISTIQUE ASYMPTOTIQUE
- STATISTIQUE DES PROCESSUS