ABDELKADER MOKKADEM

Critères de mélange pour des processus stationnaires : estimation sous des hypothèses de mélange : entropie des processus linéaires

Thèses d'Orsay, 1987

<http://www.numdam.org/item?id=BJHTUP11_1987__0213__P0_0>

ORSAY
n° d'ordre :

UNIVERSITE PARIS SUD

Centre d'Orsay

T H E S E

De Doctorat d'Etat Es Sciences Mathématiques

présentée pour obtenir le grade de

DOCTEUR ES-SCIENCES

par

Abdelkader MOKKADEM

Sujet : *CRITERES DE MELANGE POUR DES PROCESSUS STATIONNAIRES.*
ESTIMATION SOUS DES HYPOTHESES DE MELANGE.
ENTROPIE DES PROCESSUS LINEAIRES.

Soutenue le 25 septembre 1987 devant le Jury composé de :

DACUNHA-CASTELLE	*Didier*
BRETAGNOLLE	*Jean*
ELIE	*Laure*
GUYON	*Xavier*
LE BRETON	*Alain*
RAIS	*Mustapha*

A mon père....

ABSTRACT

There is three part in this thesis. In the first part we study the ergodic and mixing properties of some non linear or polynomial autoregressive random processes. We obtain sufficient conditions for geometric ergodicity and geometric absolute regularity of such processes. The results apply to the ARMA and bilinear processes. The technics used come from the Markov chain theory and the real algebraic and differential geometry. In the second part we study kernel estimators under strong mixing hypothesis ; we bound the p-mean risks and the uniform risk for the estimator of the density and some functionals ; we also propose estimators of the entropy and information of random variables and bound their risks. In the third part we study the entropy of linear processes ; we obtain an inequality between the entropy of a process and those of its linearly filtered ; an equality is obtained in some cases ; we close this part with applications particularly for the maximum entropy principle.

Key words : Autoregressive process. Markov chain. Geometric ergodicity. Strong mixing. Absolute regularity. Kernel estimation. Uniform risk. Entropy. Linear processes.

Je tiens tout d'abord à exprimer toute ma reconnaissance à Didier Dacunha-Castelle qui m'a accueilli,dirigé et encouragé; sa présence et son soutien m'ont permis de surmonter bien des moments difficiles .

Je remercie aussi Jean Bretagnolle et Alain Le Breton qui ont spontanément et très cordialement accepté de participer à mon jury de thèse .

Les suggestions et remarques de Laure Elie m'ont été très utiles;je la remercie pour les discussions fructueuses qu'elle m'a accordées .

Xavier Guyon qui a toujours un nouveau modèle de processus dans la poche a été un stimulant efficace dans la réalisation de ce travail;je lui suis redevable sur bien des points .

Avec Mustapha Rais qui m'a proposé un second sujet de thèse sur le theorème de Hilbert différentiable pour les groupes linéaires finis,j'ai découvert la théorie des invariants;les nombreuses conversations que nous avons eues ont nettement elargi ma vision des mathématiques;je tiens à l'en remercier .

Cette thèse a également bénéficié de l'amitié et de l'émulation que j'ai trouvées auprès de mes collègues du laboratoire de statistiques d'Orsay et du laboratoire de mathématiques de Poitiers; il est donc tout naturel que mes remerciements leur soient aussi adressés .

Je tiens encore à remercier mesdames Baillet,Parvan et Zielinsky d'Orsay ainsi que mesdames Ferré,Messy et Raynaud et monsieur Boisnard de Poitiers pour leur amabilité et leur aide dans la réalisation matérielle de ce travail .

Enfin, pour terminer, mes pensées vont à tous mes proches et en particulier à ma femme et mes enfants qui m'ont prodigué confiance et affection .

TABLE DES MATIERES

Page 4 - Introduction

Page 23 - Chapitre I : *Ergodicité et ergodicité géométrique des modèles autorégressifs non linéaires.*

Page 47 - Chapitre II : *Propriétés de mélange des processus* ARMA *vectoriels.*

Page 58 - Chapitre III : *Propriétés de mélange des processus autorégressifs polynomiaux.*

Page 133 - Chapitre IV : *Etude des risques des estimateurs à noyaux.*

Page 161 - Chapitre V : *Estimation de l'entropie et de l'information de variables aléatoires absolument continues.*

Page 175 - Chapitre VI : *Entropie des processus linéaires.*

I N T R O D U C T I O N

INTRODUCTION

A. MOKKADEM

Le travail présenté ici est composé de trois parties.

La première est consacrée à l'étude des propriétés d'ergodicité et de mélange de certaines classes de processus. Dans le chapitre I (voir aussi [1] et [2]) nous étudions les propriétés d'ergodicité et d'ergodicité géométrique de la chaîne de Markov $(X_n)_{n \in \mathbb{N}}$ définie par :

(1) $$X_n = \phi(X_{n-1}) + \varepsilon_{X_{n-1},n} ;$$

où $(\varepsilon_{x,n})_{n\in\mathbb{Z}}$ est une suite de champs aléatoires indépendants et de même loi de probabilité sur $\mathbb{R}^d$ et ϕ est une application mesurable de $\mathbb{R}^d$ dans $\mathbb{R}^d$. Nous faisons les deux hypothèses suivantes :

(C1) la loi de $\varepsilon_{x,n}$ est dominée par la mesure de Lebesgue sur $\mathbb{R}^d$;

(C2) pour tout borélien A de mesure de Lebesgue non nulle dans $\mathbb{R}^d$ et tout compact K dans $\mathbb{R}^d$, il existe un entier n_o tel que :

$$\operatorname*{Inf}_{x\in K} P^{n_o}(x,A) > o$$

($P(x,\cdot)$ désigne la loi de transition de la chaîne $(X_n)_{n\in\mathbb{N}}$).
En utilisant des résultats de R.L. Tweedie [13] et E. Nummelin, P. Tuominen [12], on établit diverses conditions d'ergodicité et d'ergodicité géométrique de la chaîne $(X_n)_{n\in\mathbb{N}}$. Nous obtenons en particulier le résultat suivant :

s'il existe $s > o$, $M > o$ et $\rho < 1$ tels que

$$(2) \qquad E|\phi(x)+\varepsilon_{x,n}|^s < \rho|x|^s \quad \text{pour } |x| > M \quad \text{et} \quad \sup_{|x|\le M} E|\phi(x)+\varepsilon_{x,n}| < \infty$$

alors $(X_n)_{n\in\mathbb{N}}$ est géométriquement ergodique.

La condition (2) est nettement affaiblie dans certains cas, en particulier quand X_n est unidimensionnel $(d = 1)$.
Il faut noter que sous les conditions (C1),(C2) et (2) la chaîne $(X_n)_{n\in\mathbb{N}}$ est apériodique et Harris-récurrente ; un article de Y. Davydov [14] donne alors ses propriétés de mélange.
Parmi les modèles auxquels s'appliquent les résultats du chapitre I on peut citer les modèles à seuils, certains modèles à coefficients aléatoires [2] et le modèle bilinéaire simple unidimensionnel [3]. Les automaticiens utilisent souvent des modèles du type (1) où $\varepsilon_{x,n}$ est une gaussienne centrée dont la variance dépend de x . Pour terminer, ajoutons que dans la note [3] nous donnons aussi une condition suffisante d'existence pour certains processus bilinéaires.

Notre intérêt dans cette première partie s'est ensuite porté sur des classes de processus stationnaires qui admettent une représentation markovienne ; une de nos contributions dans ce domaine a été d'établir des critères simples pour que de tels processus soient géométriquement absolument réguliers. Rappelons tout d'abord que le coefficient de régularité absolue (ou régularité complète) d'un processus stationnaire en loi $(X_n)_{n\in\mathbb{Z}}$ est défini pour tout entier positif k par :

$$(3) \qquad \beta(k) = E \sup\{|P(B|\mathcal{A}_o) - P(B)| \ , \ B \in \mathcal{A}^k\}$$

où $\mathcal{A}_o$ est la tribu engendrée par $\{X_o, X_{-1}, X_{-2}, \ldots\}$ et $\mathcal{A}^k$ celle engendrée par $\{X_k, X_{k+1}, X_{k+2}, \ldots\}$; notons que, d'après un résultat de V.A. Volkonskii et Y.A. Rozanov [15], $\beta(k)$ peut aussi s'écrire :

$$(4) \qquad \beta(k) = \sup \frac{1}{2} \sum_{i=1}^{I} \sum_{j=1}^{J} |P(A_i \cap B_j) - P(A_i)P(B_j)|$$

où le supremum est pris sur toutes les paires de partitions finies $\{A_1, \ldots, A_I\}$ et $\{B_1, \ldots, B_J\}$ de l'espace de probabilité Ω sur lequel est défini le processus $(X_n)_{n \in \mathbb{Z}}$; ces partitions étant astreintes à satisfaire :

$$A_i \in \mathcal{A}_o \quad \text{et} \quad B_j \in \mathcal{A}^k .$$

Le processus est dit absolument régulier (ou complètement régulier) si :

$$(5) \qquad \lim_{k \to \infty} \beta(k) = o ;$$

il est géométriquement absolument régulier s'il existe un nombre réel positif $\xi < 1$ tel que :

$$(6) \qquad \beta(k) = O(\xi^k) .$$

La notion de régularité absolue a été introduite par A.N. Kolmogorov. Elle peut être comparée à deux autres notions de mélange fréquemment utilisées : le mélange fort (M. Rosenblatt [16]) défini à partir du coefficient :

$$(7) \quad \alpha(k) = \sup\{|P(A \cap B) - P(A)P(B)| \ , \ A \in \mathcal{A}_o \ \text{ et } \ B \in \mathcal{A}^k\}$$

et le mélange uniforme (I.A. Ibragimov et Y.V. Linnik [17]) défini à partir de :

$$(8) \qquad \phi(k) = \sup\{|P(B|A) - P(B)| \ , \ A \in \mathcal{A}_o \ , \ B \in \mathcal{A}^k\} ;$$

on peut constater alors sans difficulté que :

(9) $$\alpha(k) \leq \beta(k) \leq \phi(k)$$

de sorte que les processus absolument réguliers sont en particulier fortement mélangeants.

Sous des hypothèses sur la vitesse de convergence vers zéro des coefficients de mélange fort ou des coefficients de régularité absolue, diverses propriétés asymptotiques sont obtenues comme on peut le voir par exemple dans [15],[17],[18],[19],[20] . Certains résultats obtenus sous l'hypothèse de régularité absolue ne sont plus valables sous la seule hypothèse de mélange fort (voir H.C.P. Berbee [21]).

Il apparait, par conséquent, important de pouvoir déterminer quand un processus est absolument régulier ; c'est A.N. Kolmogorov qui semble avoir posé le problème de trouver des critères effectifs pour qu'un processus soit absolument régulier. L'une des premières réponses apparaît dans le travail de V.A. Volkonskii et Y.A. Rozanov [15]. Quelques temps après I.A. Ibragimov et Y.A. Rozanov [22] étudient systématiquement les diverses notions de mélange pour les processus gaussiens.

Toujours pour les processus linéaires K.C. Chandra [23], V.V. Gorodetskii [24] et C.S. Withers [25] établissent une condition suffisante pour qu'un processus linéaire unidimensionnel soit fortement mélangeant; T.D.Pham,L.T.Tran dans[26] montrent que sous cette même condition le processus linéaire de dimension quelconque est en fait absolument régulier. Parallèlement à ces résultats se développait une série de travaux sur les propriétés de mélange des processus markoviens. Ainsi, dans [27], J.L. Doob montre qu'un processus markovien Doeblin récurrent est uniformément mélangeant. Dans un important ouvrage [28] M. Rosenblatt définit une nouvelle notion de mélange pour les processus markoviens (mélange en norme L^p) et donne des conditions nécessaires et suffisantes pour qu'un processus markovien soit fortement mélangeant ou ρ-mélangeant ; il donne également des conditions de mélange pour une marche aléatoire sur un groupe. La propriété de régularité absolue des processus markoviens est étudiée par Y. Davydov [14] qui fait ressortir le lien étroit qui existe entre cette propriété de mélange et la propriété de récurrence au sens de Harris. C'est ce lien qui nous a amenés à choisir la notion de régularité absolue dans notre travail.

Notre premier résultat dans ce domaine ([4] et [5]) concerne les processus ARMA vectoriels et fait l'objet du chapitre II.

On y considère un processus $(Y(t))_{t\in\mathbb{Z}}$ ARMA vectoriel d'équation

$$(10) \qquad \sum_{i=o}^{P} B(i)\, Y(t-i) = \sum_{k=o}^{Q} A(k)\, \varepsilon(t-k) ,$$

où $B(i)$ et $A(k)$ sont des matrices réelles de dimensions respectives $\ell \times \ell$ et $\ell \times r$; $B(o) = Id$; $\varepsilon(t)$ est une suite de variables aléatoires indépendantes, de même loi et à valeurs dans $\mathbb{R}$.

On suppose que :

$$(11) \qquad E|\varepsilon| < \infty \quad \text{et} \quad E\,\varepsilon = o$$

(C4) les zéros du polynôme $\det(\sum_{i=o}^{P} B(i)\, z^i)$ sont en valeur absolue strictement plus grands que 1 .

Sous ces conditions H. Akaïke [29] montre qu'il existe un processus markovien $(X(t))_{t\in\mathbb{Z}}$ tel que :

(12) $$X(t) = F\cdot X(t-1) + G\,\varepsilon(t)$$

et

(13) $$Y(t) = H\cdot X(t)$$

où F , G et H sont des matrices ; nous utilisons alors les équations (12) et (13) pour obtenir le résultat suivant :

Si la loi de probabilité de $\varepsilon(t)$ est équivalente à la mesure de Lebesgue sur $\mathbb{R}^r$ alors le processus $(Y(t))_{t\in\mathbb{Z}}$ est géométriquement absolument régulier.

Comme on le voit, notre condition sur la loi de $\varepsilon(t)$ est différente de celle de C.S. Withers [25] et T.D. Pham, L.T. Tran [26].

L'approche du chapitre II va être nettement approfondie dans le chapitre III où nous considérons un processus stationnaire en loi dans $\mathbb{R}^d$, $(X_n)_{n\in\mathbb{Z}}$ admettant une représentation markovienne polynômiale dans le sens suivant [6] :

il existe un processus stationnaire markovien $(Z_n)_{n\in\mathbb{Z}}$ à valeurs dans $\mathbb{R}^p$ et un entier positif k tels que :

(14) $$Z_n = \phi(Z_{n-1}, e_n)$$

et

(15) $$X_n = \psi(Z_{n-k}, Z_{n-k+1}, \dots, Z_{n+k})$$

où $(e_n)_{n\in\mathbb{Z}}$ est une suite de variables aléatoires i.i.d. à valeurs dans $\mathbb{R}^m$, ϕ est une application polynômiale de $\mathbb{R}^p \times \mathbb{R}^m$ dans $\mathbb{R}^p$ et ψ est une application mesurable de $\mathbb{R}^{p(2k+1)}$ dans $\mathbb{R}^d$.

Il est clair que les propriétés de mélange de $(X_n)_{n\in\mathbb{Z}}$ sont les mêmes que celles de $(Z_n)_{n\in\mathbb{Z}}$

Nous supposons que ϕ est de degré un relativement à Z_{n-1} ; l'équation (14) s'écrit alors sous la forme matricielle :

$$(16) \qquad Z_n = (A + \sum_\delta B_\delta \, e_n) Z_{n-1} + \sum_\delta C_\delta \, e_n^\delta + D \ ,$$

où la sommation sur δ est étendue à une partie finie de $\mathbb{N}^m$; après avoir déterminé presque sûrement l'espace des états du processus $(Z_n)_{n\in\mathbb{Z}}$, nous montrons qu'il est géométriquement absolument régulier sous les hypothèses suivantes :

(H_1') $\exists$ $s > o$ tel que

$$E|\sum_\delta C_\delta \, e_n^\delta + D|^s < \infty \quad \text{et} \quad E||A + \sum_\delta B_\delta \, e_n||^s < 1 \ ;$$

(H_2) les valeurs propres de A sont en valeur absolue strictement inférieures à un ;

(C5) La loi de probabilité de e_n est dominée par une mesure lebesguienne sur une variété algébrique lisse V dans $\mathbb{R}^m$ et sa densité f par rapport à cette mesure lebesguienne est telle que l'ensemble $\{f > o\}$ soit un un voinisage ouvert de zéro dans V .

On notera, comme celà apparaît dans nos démonstrations, que sous les hypothèses citées, le processus $(Z_n)_{n\in\mathbb{Z}}$ est en particulier Harris-récurrent .

Pour établir ces résultats, nous démontrons au préalable un théorème de continuité de l'image d'une mesure par une application polynômiale (théorème 3.2, chapitre III) ; ce théorème apparait comme une extension du théorème de continuité de la translation. Plus exactement nous montrons le résultat suivant.

Soit $\psi(z,x)$ une application de classe C^1 de $U \times V$ dans W où U et W sont des variétés algébriques contenues respectivement dans $\mathbb{R}^{n_1}$ et $\mathbb{R}^{n_2}$, V est une variété algébrique lisse contenue dans $\mathbb{R}^{n_3}$; on suppose que ψ est polynômiale en x.

Soit θ une mesure positive régulière sur V absolument continue par rapport à une mesure Lebesguienne sur V et θ_z son image dans W par l'application $\psi(z,\cdot)$.

Alors, si z_o est tel que l'image de V par $\psi(z_o,\cdot)$ contienne un ouvert non vide dans la partie régulière de W, on a :

$$\liminf_{z \to z_o} \theta_z(A) \geq \theta_{z_o}(A)$$

pour tout Borélien A de W ; de plus, si θ est une mesure finie, alors :

$$\lim_{z \to z_o} \theta_z(A) = \theta_{z_o}(A)$$

pour tout Borélien A de W.

Outre la théorie des processus markoviens et le théorème ergodique, les outils que nous utilisons proviennent de la géométrie algébrique et différentielle réelle.
Les résultats obtenus s'appliquent naturellement aux processus ARMA et aux processus bilinéaires. ∎

La deuxième partie de la thèse porte sur l'estimation par la méthode du noyau pour des observations non indépendantes.
Soit $(X_n)_{n\in\mathbb{Z}}$ un processus stationnaire en loi, fortement mélangeant et à valeurs dans $\mathbb{R}^d$; soit $f(x)$ la densité de probabilité de X_n.
Suivant M. Rosenblatt [30] , $f(x)$ peut être estimée par :

$$(17) \qquad f_N(x) = \frac{1}{N\, h_N^d} \sum_{j=1}^{N} K\left(\frac{x-X_j}{h_N}\right)$$

où K est un noyau positif et (h_N) est une suite de nombres réels positifs qui converge vers zéro.

Il existe une vaste littérature concernant cette estimation dans le cas i.i.d. (pour une bibliographie récente on peut voir par exemple L. Devroye et L. Györfi [31]).
Dans [8] et l'article [7] qui constitue le chapitre IV notre objectif est d'obtenir une approche unifiée de l'étude des risques :

$$(18) \qquad E\| x^k f_N(x)-x^k f(x)\|_q^p \quad , \quad p>o \quad , \quad o<q\le\infty \quad , \quad k\in\mathbb{N}^d \ ;$$

pour celà nous utilisons, quand q est supérieur ou égal à deux, la propriété de contraction de la transformée de Fourier sur $L^{q'}(\mathbb{R}^d)$ où q' est le conjugué de q . Quand q est inférieur à deux nous utilisons le facteur x^k qui apparaît dans (18). Nous obtenons alors sous certaines hypothèses sur f , K et le mélange du processus $(X_n)_{n\in\mathbb{Z}}$ les résultats suivants :

si $2 \leq q \leq \infty$, $q' = \frac{q}{q-1}$:

$$(19) \qquad E\left(\int_{\mathbb{R}^d} |x^k f_N(x) - x^k f(x)|^q \, dx\right)^{p/q} = 0\left(\frac{1}{N^{p/2} h_N^{dp/q'}} + h_N^{2p}\right)$$

si $o < q < 2$:

$$(20) \qquad E\left(\int_{\mathbb{R}^d} |x^k f_N(x) - x^k f(x)|^q \, dx\right)^{p/q} = 0\left(\frac{1}{N^{p/2} h_N^{dp/2}} + h_N^{2p}\right) ;$$

On a en particulier pour la norme $\sup(q = \infty)$

$$(21) \qquad E \sup_{\mathbb{R}^d} |x^k f_N(x) - x^k f(x)|^p = 0\left(\frac{1}{N^{p/2} h_N^{dp}} + h_N^{2p}\right)$$

En dehors d'un travail récent de P. Doukhan, J. Leon et F. Portal [32] sur le risque $E \sup_S |f_N(x) - f(x)|$ avec S compact il ne nous semble pas que le risque $E \sup_{\mathbb{R}^d} |f_N(x) - f(x)|$ ait été étudié sous des hypothèses de mélange.

Nos hypothèses de régularité sur f sont écrites dans le langage des distributions. Nous montrons qu'elles sont équivalentes à l'hypothèse introduite par L. Devroye et L. Györfi [31] dans le cas $k = o$ et $p = q = d = 1$:

$$(22) \qquad \sup_{a>o} \int |(f * \phi_a)''| < \infty$$

où ϕ est une densité positive à support compact et de classe C^2 ; on constate en particulier que l'hypothèse (22) de L. Devroye et L. Györfi signifie exactement que f est absolument continue et sa dérivée presque partout f' est une fonction à variation bornée.

Les résultats (19) et (20) sont utilisés à la fin du chapitre IV pour étudier les risques des estimations de la régression, de la densité conditionnelle et d'autres fonctionnelles.

Le chapitre V ([9]) est consacré à l'estimation de l'entropie de X_n par la méthode du noyau. Cette entropie est définie par :

$$(23) \qquad H(X_n) = -\int f(x) \text{ Log } f(x)\, dx$$

quand l'intégrale existe.

L'entropie joue un rôle important en théorie de l'information ainsi que dans certaines branches de la physique.

Diverses estimations ont été proposées dans le cas unidimensionnel ($d = 1$) et pour des observations i.i.d.

I.A. Ahmad et P.E. Lin [33] proposent l'estimateur :

$$(24) \qquad \hat{H}_N = -\frac{1}{N} \sum_{i=1}^{N} \text{Log } f_N(X_i)$$

où f_N est défini par (17) ; ils étudient la consistence en moyenne de $\hat{H}_N$. Leur démonstration est malheureusement incomplète.

B.L.S. Prakasa Rao [34] propose l'estimateur :

$$(25) \qquad \hat{H}_N = -\int_{k_N}^{k_N} f_N(x) \text{ Log } f_N(x)\, dx$$

où k_N est une suite croissante de nombres réels. Un autre estimateur intéressant est proposé par O. Vasicek [35] mais il n'est pas utilisable en multidimensionnel. Nous proposons les estimateurs :

(26) $$H_N = -\frac{1}{r_N} \operatorname{Log} \int_{\mathbb{R}^d} f_N^{r_N+1}(x)\,dx$$

où $r_N = h_N^\gamma$ et γ est un nombre réel positif et :

(27) $$\tilde{H} = - \int_{\mathbb{R}^d} f_N(x) \operatorname{Log} f_N(x)\,dx \;;$$

nous majorons ensuite les risques moyens d'ordre p de ces estimateurs. Le résultat étant valable en multidimensionnel, on l'utilise pour estimer l'information mutuelle de deux variables aléatoires X et Y ; cette information vérifie en effet la relation

(28) $$I(X,Y) = H(X) + H(Y) - H(X,Y) \quad \blacksquare$$

La troisième partie, constituée du chapitre VI, est indépendante des deux premières ; elle est consacrée à l'étude de l'entropie des processus linéaires [10] étude que nous avons commencée dans [11]. La définition de l'entropie d'une variable aléatoire donnée par la formule (23) plus haut est étendue dans le chapitre VI à toute variable aléatoire ayant un moment absolu d'ordre $s > o$ fini . Elle est notée $h(X)$ au lieu de $H(X)$ pour éviter la confusion avec l'entropie de processus que nous définissons maintenant.

Soit $X = (X_n)_{n\in\mathbb{Z}}$ un processus stationnaire en loi admettant un moment absolu d'ordre $s > o$ fini ; l'entropie de X est définie par :

$$(29) \qquad H(X) = \lim_n \frac{1}{n+1} h(X_o,\dots,X_n)$$

Notre objectif dans le chapitre VI est de trouver une relation entre l'entropie du processus X et celle d'un processus Y quand X est obtenu par un filtre linéaire appliqué à Y .

Plus précisément, nous considérons un processus stationnaire en loi réel $Y = (Y_n)_{n\in\mathbb{Z}}$ dans une classe de processus harmonisables qui contient les processus de carré intégrables et certains processus symétriques α-stables [36]. On suppose que le processus $X = (X_n)_{n\in\mathbb{Z}}$ est donné par :

$$(30) \qquad X_n = \int_{-\pi}^{\pi} e^{in\lambda}\, \phi(\lambda)\, dZ_Y$$

où dZ_Y désigne la mesure spectrale aléatoire de Y et ϕ est intégrable pour la mesure dZ_Y .

Nous montrons alors que :

$$(31) \qquad H(X) \geq \frac{1}{2\pi}\int_{-\pi}^{\pi} \mathrm{Log}|\phi(\lambda)|d\lambda + H(Y) \ ;$$

nous obtenons également des cas où l'égalité à lieu dans (31) ; parmi ces cas citons celui où Y est régulier ou symétrique α-stable. En particulier l'entropie d'un processus X régulier ou symétrique α-stable non singulier s'écrira exactement :

$$(32) \qquad H(X) = \frac{1}{2\alpha\pi}\int_{-\pi}^{\pi} \mathrm{Log}\ f(\lambda)\ d\lambda + H(W)$$

où f et W sont respectivement la densité spectrale et l'innovation de la partie régulière de X et $\alpha = 2$ si le processus X est de carré intégrable. Ce type de résultats remonte à C.E. Shannon [37] qui établit l'égalité dans (31) quand Y est Gaussien ; sa méthode consiste à calculer directement $h(X_o,\ldots,X_n)$ puis à passer à la limite dans (29). Plus récemment M. Kanter [38] donne l'inégalité (31) quand $Y = (Y_n)_{n\in\mathbb{Z}}$ est une suite de variables aléatoires indépendantes et de carré intégrable. Enfin L.A. Shepp, D. Slepian et A.D. Wyner [39] démontrent (31) quand $\phi(\lambda)$ est un polynôme trigonométrique.

La suite du chapitre contient diverses applications ; notamment une minoration de l'erreur de prédiction non linéaire et une discussion du principe du maximum d'entropie de J.P. Burg [40].

Enfin, nous terminons le chapître en donnant une extension des résultats au cas où Y n'est pas harmonisable. Plus exactement nous montrons que si $Y = (Y_n)_{n\in\mathbb{Z}}$ admet un moment absolu d'ordre un et si $X = (X_n)_{n\in\mathbb{Z}}$ est défini par la série convergente en moyenne ;

$$(33) \qquad X_n = \sum_{j=-\infty}^{+\infty} a_j \, Y_{n-j} \quad \text{avec} \quad \sum_{j=-\infty}^{+\infty} |a_j| < \infty$$

alors :

$$(34) \qquad H(X) \geq \frac{1}{2\pi} \int_{-\pi}^{\pi} \mathrm{Log}|\phi(\lambda)|\, d\lambda + H(Y)$$

où

$$(35) \qquad \phi(\lambda) = \sum_{j=-\infty}^{+\infty} a_j \, e^{ij\lambda}$$

l'égalité est obtenue si $\phi(\lambda)$ ne s'annule pas sur $[-\pi,\pi]$.

REFERENCES

[1] A. MOKKADEM - *Sur un modèle autorégressif non linéaire, ergodicité et ergodicité géométrique.*
J.T.S.A., vol.8,n°2,1987,195-204.

[2] A. MOKKADEM - *Le modèle non linéaire* AR(1) *général. Ergodicité et ergodicité géométrique.*
C.R.A.S., t.301, série I, 1985, 889-892.

[3] A. MOKKADEM - *Conditions suffisantes d'existence et d'ergodicité géométrique des modèles bilinéaires.*
C.R.A.S., t.301, série I, 1985, 375-377.

[4] A. MOKKADEM - *Mixing properties of* ARMA *processes.*
submitted to Stoch. Proc. Appl.

[5] A. MOKKADEM - *Sur le mélange d'un processus* ARMA *vectoriel.*
C.R.A.S., t.303, série I, 1986, 519-521.

[6] A. MOKKADEM - *Propriétés de mélange des processus autorégressifs polynômiaux.*
Prépublication. Soumis aux Annales de l'I.H.P..

[7] A. MOKKADEM - *Study of risks of kernel estimators.*
To appear

[8] A. MOKKADEM - *Etude des risques des estimateurs à noyaux.*
C.R.A.S., t.301, série I, 1985, 447-450.

[9] A. MOKKADEM - *Estimation of the entropy and information of absolutely continuous random variables.*
To appear in I.E.E.E. Trans. Inf. Theory.

[10] A. MOKKADEM - *Entropie des processus linéaires.*
Prépublication. Soumis à Prob. and Math. Stat.

[11] A. MOKKADEM - *Entropie de processus et erreur de prédiction.*
C.R.A.S., t.298, série I, 1984, 493-496.

[12] E. NUMMELIN, P. TUOMINEN - *Geometric ergodicity of Harris-recurrent Markov chains.*
Stoch. Proc. Appl.,3,1982,187-202.

[13] R.L. TWEEDIE - *Sufficient conditions for ergodicity and recurrence of Markov chains.*
Stoch. Proc. Appl.,3,1975,385-403.

[14] Y.A. DAVYDOV -*Mixing conditions for Markov chains.*
Th. Prob. Appl.,28,1973,313-328.

[15] V.A. VOLKONSKII, Y.A. ROZANOV - *Some limit theorems for random functions, part II.*
Th. Prob. Appl.,6,1961,186-198.

[16] M. ROSENBLATT - *A central limit theorem and a strong mixing condition.*
Proc. Nat. Acad.Sci.,U.S.A.,42,1956,43-47.

[17] I.A. IBRAGIMOV, Y.V. LINNIK - *Independent and stationary sequences of random variables.*
Walter-Noordhoof publishing, 1974.

[18] H.TAKATA - L_∞ *bounds for asymptotic normality of weakly dependent summands using Stein's methods.*
Ann. Prob.,9,1981,676-683.

[19] K.YOSHIHARA - *Limiting behavior of U-statistics for stationary absolutely regular processes.*
Z. Wahr. Verw. Gebiete,35,1976,237-252.

[20] R.C. BRADLEY - *Absolute regularity and functions of Markov chains.*
Stoch. Proc. Appl.,14,1983,67-77.

[21] H.C.P. BERBEE -*Random walks with stationary increments and renewal theory.*
Mathematical Center Tract n°112, AMSTERDAM, 1979.

[22] I.A. IBRAGIMOV, Y. ROZANOV -*Processus aléatoires gaussiens.*
MIR, Moscou, 1974.

[23] K.C. CHANDRA - *Strong mixing properties of linear stochastic processes.*
J.A.P.,11,1974,401-408.

[24] V.V. GORODETSKI - *On the strong mixing property for linear sequences.*
Th. Prob. Appl.,22,1977,411-413.

[25] C.S. WITHERS - *Conditions for linear processes to be strong mixing.*
Z. Wahr. Verw. Gebiete,1981,477-480.

[26] T.D. PHAM, L.T. TRAN -*Some mixing properties of time series models.*
Stoch. Proc.Appl.19,1985,297-303.

[27] J.L. DOOB - *Stochastic processes.*
Wiley, New-York,1953.

[28] M. ROSENBLATT -*Markov processes. Structure and asymptotic behaviour.*
Springer, Berlin, 1971.

[29] H. AKAIKE - *Markovian representation of stochastic processes.*
Ann. Inst. Stat. Math.,26,1974,363-387.

[30] M. ROSENBLATT - *Remarks on some non parametric estimates of a density function.*
Ann. Math. Stat.,27,1956,832-835.

[31] L. DEVROYE, L. GYÖRFI - *Non parametric density estimation. The* L_1 *view.*
Wiley, New-York, 1985.

[32] P. DOUKHAN, J. LEON, F. PORTAL - *Principe d'invariance faible pour la mesure empirique d'une suite de variables aléatoires mélangeantes.*
Prépublication d'Orsay,1985.

[33] I.A. AHMAD, P.E. LIN - *A nonparametric estimation of the entropy for absolutely continuous distributions.*
I.E.E.E. Trans. Inf. Theory, vol.I,t.22,1976,372-375.

[34] B.L.S. PRAKASA RAO - *Non parametric functional estimation.*
Academic Press, 1983.

[35] O. VASICEK - *A test for normality based on sample entropy.*
J.R.S.S., serie B, 38,1976,54-59.

[36] S. CAMBANIS - *Complex symetric stable variables and processes. In : Contribution to Statistics : Essays in Honor of Norman L. Johnson.*
ed. Sen, P.K., New-York, North-Holland,1982,63-79.

[37] C.E. SHANNON - *The mathematical theory of communications.*
Bell System Technical Journal, 1948.

[38] M. KANTER - *Lower bound for non linear prediction error in moving average processes.*
Ann. Prob.,7,1979,128-138.

[39] L.A. SHEPP, D. SLEPIAN, A.D. WYNER - *On prediction of moving average Processes.*
Bell System Technical Journal,59,1980,367-415.

[40] J.P. BURG - *A new analysis technique for time series data. In modern spectrum analysis.*
ed. D.G. Childer, Wiley, New-York, 1978.

C H A P I T R E I

SUR UN MODELE AUTOREGRESSIF NON LINEAIRE. ERGODICITE ET ERGODICITE GEOMETRIQUE.

Abdelkader MOKKADEM.

On établit des conditions suffisantes d'ergodicité et d'ergodicité géométrique pour la chaîne de Markov (X_n) définie par :

$$X_n = \varphi(X_{n-1}) + \varepsilon_{X_{n-1},n}.$$

On donne également des exemples d'application.

We obtain sufficient conditions of ergodicity and geometric ergodicity for the Markov chain (X_n) defined by :

$$X_n = \varphi(X_{n-1}) + \varepsilon_{X_{n-1},n}.$$

We give also some applications.

I. **INTRODUCTION**.-

Soit une suite de champs aléatoires sur $\mathbb{R}^d$ à valeurs dans $\mathbb{R}^d$, indépendants : $\varepsilon_{x,n}$, $x \in \mathbb{R}^d$, $n \in \mathbb{Z}$.

Soit φ une fonction de $\mathbb{R}^d$ dans $\mathbb{R}^d$ mesurable.

L'objet de ce travail est l'étude du comportement de la chaîne de Markov (X_n), $n \in \mathbb{N}$, définie par :

(1) $$X_n = \varphi(X_{n-1}) + \varepsilon_{X_{n-1},n}.$$

Un modèle voisin a été étudié par P. Doukhan et M. Ghindès (1980 a,b) qui considèrent :

(2) $$X_n = \varphi(X_{n-1}) + \varepsilon_n.$$

Ils établissent la proposition suivante pour le modèle (2) :

On suppose les ε_n indépendants de même loi F équivalente à la mesure de Lebesgue. Alors si φ est une fonction mesurable bornée, la chaîne (X_n) est Doeblin-récurrente.

Ils en déduisent alors que la chaîne est géométriquement ergodique et appliquent ce résultat à l'étude statistique du modèle (2).

La propriété d'ergodicité géométrique est plus faible que la récurrence Doeblin, mais elle est suffisante pour établir certains résultats statistiques. En fait, la différence entre les deux propriétés est que : dans le cas Doeblin-récurrent le processus (X_n) est uniformément mélangeant, alors que dans le cas d'ergodicité géométrique il est fortement mélangeant (Rosenblatt (1971)).

On se propose dans ce travail d'établir des conditions suffisantes d'ergodicité géométrique de la chaîne (X_n) définie par le modèle (1).

On notera que (1) apparait comme le modèle AR(1) le plus général ; en particulier le modèle du type $X_n = \psi(X_{n-1}, \eta_n)$ peut s'écrire sous la forme (1) avec par exemple $\varphi = 0$ et $\varepsilon_{x,n} = \psi(x, \eta_n)$.

On supposera dans tout ce qui suit que $\varepsilon_{x,n}$ admet une densité $f_x(v)$, par rapport à la mesure de Lebesgue μ sur $\mathbb{R}^d$, indépendante de n. La chaîne (X_n) admet alors pour densité de transition :

$$p(x,y) = f_x(y-\varphi(x)).$$

Sa loi de transition sera notée $P(x,.)$.

On utilisera dans la suite les hypothèses suivantes :

(U) : Pour n fixé les variables $|\varepsilon_{x,n}|$ sont uniformément intégrables.

(C1) : Pour tout Borélien A de μ-mesure non nulle et tout compact K dans $\mathbb{R}^d$, il existe un entier n_o strictement positif tel que :

$$\operatorname*{Inf}_{x\in K} P^{n_o}(x,A) > 0.$$

La condition (C1) implique :

(C2) : (X_n) est apériodique.

La condition (C2) est aussi vérifiée si par exemple :

(C3) : $\exists n_1 \in \mathbb{N}^*$, tel que $P^{n_1}(x,.) \sim \mu$ pour tout x.

De telles conditions sont satisfaites dans le modèle (2) dès que la densité de ε_n est strictement positive et φ est bornée sur tout compact.

Sous l'hypothèse (C1), la mesure $\sum 2^{-n}P^n(x,.)$ est équivalente à μ pour tout x ; et donc la chaîne (X_n) est fortement irréductible (voir Tweedie, 1976).

Soit π une mesure sous-invariante de (X_n) ; dans le cas où (X_n) est ergodique, on prendra pour π, la probabilité invariante de (X_n).

Dans les parties II et III on établit des conditions suffisantes d'ergodicité et d'ergodicité géométrique du modèle (1) dans le cas vectoriel. Dans la partie IV, on se restreint au cas scalaire (d=1) ; les conditions suffisantes qu'on obtient sont alors plus faibles que celles obtenues dans le cas vectoriel. Enfin dans la partie V on donne un exemple d'application.

II. ERGODICITE DE (X_n) DANS LE CAS VECTORIEL.

Le lemme suivant va nous permettre d'utiliser des compacts pour les critères d'ergodicité et de récurrence établis par Tweedie (1975).

LEMME 1. Soit K un compact de $\mathbb{R}^d$, alors sous la condition (C1) :

$$\mu(K) > 0 \Longrightarrow 0 < \pi(K) < \infty.$$

Preuve : La chaîne (X_n) étant fortement irréductible il est clair alors que $\pi >> \mu$ et donc $\pi(K)>0$. Puisque μ et π sont σ-finies, on peut trouver A tel que $\mu(A)>0$ et $\pi(A)<\infty$ (en effet, on prend A tel que $\mu(A)>0$: on a $A=\cup A_i$ avec $\pi(A_i) < \infty$ pour tout i et donc il existe i_o tel que $\mu(A_{i_o})>0$ et $\pi(A_{i_o})<\infty$) ; sous la condition (C1), il existe n tel que $b = \inf_{x\in K} P^n(x,A)>0$; on a :

$$\pi(A) \geqslant \int_{\mathbb{R}^d} \pi(dx)P^n(x,A) \geqslant \pi(K).b$$

et donc $\pi(K) < \infty$.

On peut maintenant énoncer des conditions suffisantes d'ergodicité de (X_n).

PROPOSITION 1. Si (C1) est vérifiée et s'il existe $M>0$, $\eta>0$ et $s>0$ tels que :

(3) $E|\varphi(x)+\varepsilon_{x,n}|^s \leqslant |x|^s-\eta$ pour $|x|>M$ et $\sup E|\varphi(x)+\varepsilon_{x,n}|^s<\infty$.

Alors (X_n) est une chaîne ergodique.

Preuve : On applique le critère d'ergodicité de Tweedie (1975).

On prend $g(x) = a|x|^s$ avec $a>\eta^{-1}$; il est facile alors de voir que pour $|x|$ assez grand

$$\int p(x,y)g(y)dy - g(x) + 1 < 0$$

Remarque 1 : s'il existe $s\leqslant 1$, $A>0$, $M>0$, tels que $\sup_x E|\varepsilon_{x,n}|^s<A<\infty$, $|\varphi(x)|^s \leqslant |x|^s-A$ et $\sup_{|x|\leqslant M} |\varphi(x)|<\infty$. Alors (3) est vérifiée. Cela correspond à un résultat connu (Doukhan-Ghindès, 1980 a) dans le cas du modèle (2) avec $s=1$ et $\varepsilon_{x,n} = \varepsilon_n$.

La condition (3) porte sur le moment de $|\varepsilon_{x,n}|^s$; cette contrainte peut-être levée dans le cas où $E\varepsilon_{x,n} = 0$ pour tout x ;

on obtient alors :

PROPOSITION 2. Si les hypothèses (U) et (C1) sont satisfaites et $E\varepsilon_{x,n} = 0$ pour tout x, et si φ possède pour $|x|$ assez grand la propriété :

(4) $\forall\ K>0,\ \exists\ K'$ tel que : $|\varphi(x)| > K' \Longrightarrow |\varphi_i(x)| > K$

pour tout composante φ_i de φ ;

alors s'il existe $M>0$ et $\eta>0$ tels que :

$\sum_{i=1}^{d} |\varphi_i(x)| < \sum_{i=1}^{d} |x_i| - \eta$ pour $|x|>M$ et $\sup_{|x|\leq M} |\varphi(x)| < \infty$,

la chaîne (X_n) est ergodique.

Il faut noter que la propriété (4) est trivialement vérifiée dans le cas scalaire.

Preuve de la proposition 2 : On prend $g(x) = A \sum_{i=1}^{d} |x_i|$ où les x_i sont les composantes de x et A est tel que $A\eta>2$.

Ecrivons l'intégrale : $\int_{\mathbb{R}^d} p(x,y)g(y)dy$ dans chacun des 2^d trièdres de $\mathbb{R}^d$; on obtient :

$$\int_{\mathbb{R}^d} p(x,y)g(y)dy = \sum_{\delta} A \int_{I_{\delta_2}} \int_{I_{\delta_2}} \cdots \int_{I_{\delta_d}} f_x(y-\varphi(x))(\delta_1 y_1 + \ldots + \delta_d y_d)dy_1 \ldots dy_d$$

où la somme est étendue à tous les δ de la forme :

$$\delta = (\delta_1, \ldots, \delta_1) \quad \text{avec} \quad \delta_i = +\ 1\ ;$$

$I_{\delta_i} =]-\infty, 0]$ si $\delta_i = -1$ et $I_{\delta_i} = [0, +\infty[$ si $\delta_i = +1$.

Faisons le changement de variable $u = y - \varphi(x)$; on obtient :

$$(5) \quad \int_{\mathbb{R}^d} p(x,y)g(y)dy \leq \sum_{\delta} A \left| \int_{J_{\delta_1}} \int_{J_{\delta_2}} \cdots \int_{J_{\delta_d}} (\delta_1 u_1 + \ldots + \delta_d u_d) f_x(u) du_1 \ldots du_d \right|$$
$$+ A \sum_{i=1}^{d} |\varphi_i(x)|$$

avec $J_{\delta_i} =]-\infty, -\varphi_i(x)]$ si $\delta_i = -1$ et

$J_{\delta_i} = [-\varphi_i(x), +\infty[$ si $\delta_i = +1$.

En vertu des hypothèses de la proposition, on peut choisir K tel que pour $|\varphi(x)|>K$, les intégrales du terme de droite de (5) soient assez petites, de sorte que :

$$\int_{\mathbb{R}^d} p(x,y)g(y)dy \leqslant \frac{1}{2} + A\sum_{i=1}^{d} |\varphi_i(x)|.$$

Pour $|\varphi(x)| \leqslant K$, l'intégrale $\int_{\mathbb{R}^d} p(x,y)g(y)$ est majorée par :

$$A2^{d-1}(\sup_x E|\varepsilon_{x,n}| + K)$$

mais sous l'hypothèse (U), $\sup_x E|\varepsilon_{x,n}| < \infty$; il est maintenant facile de voir que $|x|$ assez grand :

$$\int_{\mathbb{R}^d} p(x,y)g(y)dy - g(x) + 1 < 0.$$

III. ERGODICITE GEOMETRIQUE.

Rappelons que la chaîne ergodique (X_n) est dite géométriquement ergodique, s'il existe $\rho<1$ tel que :

$$\|P^n(x,.) - \pi\| = O(\rho^n) \quad \text{pour } \pi\text{-presque tout } x.$$

$\|.\|$ désignant la norme en variation.

Nummelin-Tuominen (1982) montrent que si de plus (X_n) est Harris-récurrente et apériodique alors :

$$\int \pi(dx) \; \|P^n(x,.)-\pi\| = O(\rho^n).$$

On peut montrer alors que le processus (X_n) est α-mélangeant avec un coefficient de α-mélange convergeant vers zéro géométriquement (Rosenblatt, 1971). Une telle propriété est importante dans l'étude statistique du processus (voir Doukhan-Ghindès, 1980 b et Mokkadem, 1985 b).

On va établir des critères d'ergodicité géométrique. Pour cela on a besoin du :

LEMME 2. Sous les conditions de la proposition 1 ou de la proposition 2, la (X_n) est Harris-récurrente et les mesures π et μ sont équivalentes.

C'est un résultat de Tweedie (1976) puisque la chaîne (X_n) est fortement irréductible et ergodique.

Comme conséquence, tout compact K tel que $\mu(K)>0$ est un petit ensemble au sens défini par Nummelin-Tuominen (1982).

On peut maintenant énoncer :

PROPOSITION 3. Si (C1) est vérifiée et s'il existe $s>0$, $M>0$ et $\rho<1$ tels que :

(6) $E|\varphi(x)+\varepsilon_{x,n}|^s < \rho|x|^s$ pour $|x|>M$ et $\sup\limits_{|x|\leq M} E|\varphi(x)+\varepsilon_{x,n}|^s < \infty$.

Alors (X_n) est géométriquement ergodique et π admet un moment d'ordre s.

De plus : $\exists\ \xi,\ \xi < 1$ tel que :

$$\int \pi(dx) \; r_n(x) = O(\xi^n)$$

où $r_n(x) = \sup\limits_{|h|<|x|^s} \int h(y)[P^n(x,dy)-\pi(dy)]$.

Preuve : On prend $g(x) = 1+|x|^s$ et α tel que $\rho<\alpha<1$; il est clair que $\int p(x,y)g(y)dy-\alpha g(x) < 0$ pour $|x|$ assez grand. Les résultats de Nummelin-Tuominen (1982) et Tweedie (1983) permettent alors de conclure.

Remarque 2 : si $\sup\limits_x E|\varepsilon_{x,n}|^s < \infty$ alors (6) est équivalente à :

(6') il existe $M'>0$ et $\rho'<1$ tels que :

$$|\varphi(x)| < \rho'|x| \quad \text{pour} \quad |x| > M' \quad \text{et} \quad \sup_{|x|\leq M'} |\varphi(x)| < \infty.$$

En effet : pour $s\leq 1$ c'est immédiat ; pour $s>1$, on utilise les deux inégalités suivantes :

$$(E|\varphi(x)+\varepsilon_{x,n}|^s)^{1/s} \leq |\varphi(x)| + (E|\varepsilon_{x,n}|^s)^{1/s}$$

et

$$|\varphi(x)| \leq (E|\varphi(x)+\varepsilon_{x,n}|^s)^{1/s} + (E|\varepsilon_{x,n}|^s)^{1/s}.$$

Ainsi si $\sup\limits_x E|\varepsilon_x|^s<\infty$, la condition (6') est suffisante pour l'existence d'un moment d'ordre s de π.

La proposition 3 peut être améliorée sous l'hypothèse suivante : il existe $a>0$ tel que $\sup\limits_x E(\exp a|\varepsilon_{x,n}|) < c < \infty$; on obtient alors :

PROPOSITION 4. Si (C1) est vérifiée et s'il existe $M > 0$ et $A > \frac{1}{a}\log C$ tels que :

(7) $|\varphi(x)|<|x|-A$ pour $|x|>M$ et $\sup\limits_{|x|\leq M} |\varphi(x)| < \infty$.

Alors la chaîne (X_n) est géométriquement ergodique et π admet une transformée de Laplace sur $[-a \quad a]$.

De plus $\exists \xi$, $\xi<1$, tel que :

$$\int \pi(dx)\, r_n(x) = O(\xi^n)$$

où $r_n(x) = \sup\limits_{|h|\leq e^{a|x|}} \int h(y)(P^n(x,dy) - \pi(dy))$.

Preuve : Il suffit de prendre $g(x) = e^{a|x|}$ et α tel que $\exp(\log C-aA) < \alpha < 1$. On procède ensuite comme pour la proposition 3.

IV. **CAS UNIDIMENSIONNEL.**

Dans le cas où $X_n \in \mathbb{R}$, les conditions établies plus haut peuvent être affaiblies.

On suppose que le modèle (1) est centré, c'est-à-dire $E\varepsilon_{x,n}=0$ pour tout x et tout n. On obtient alors :

PROPOSITION 5. Sous les conditions (U) et (C1), s'il existe $A>0$, $M>0$, $\alpha<0$, $\gamma<0$, $\alpha\gamma\leqslant 1$ tels que $\sup_{|x|\leqslant M} |\varphi(x)|<\infty$ et :

$$(8)\qquad \begin{aligned} \alpha x+A &\leqslant \varphi(x) \leqslant x-A \quad \text{pour} \quad x>M \\ x+A &\leqslant \varphi(x) \leqslant \gamma x-A \quad \text{pour} \quad x<-M \end{aligned}$$

Alors la chaîne (X_n) est ergodique.

Preuve :

Soit $a>0$ et $b = -\dfrac{a}{\alpha}$; on voit que :

$$(9)\qquad a\gamma + b \geqslant 0.$$

Choisissons a assez grand, de sorte que :

$$(10)\qquad \begin{aligned} 1 - aA &< -1 \\ 1 - bA &< -1 \end{aligned}$$

et soit $g(x) = \begin{cases} ax & \text{si } x \geqslant 0 \\ -bx & \text{si } x < 0 \end{cases}$. On a alors :

$$\int p(x,y)g(y)dy=a\int_{-\varphi(x)}^{+\infty} uf_x(u)du-b\int_{-\infty}^{-\varphi(x)} uf_x(u)du+a\varphi(x)\int_{-\varphi(x)}^{+\infty} f_x(u)du$$
$$-b\varphi(x)\int_{-\infty}^{-\varphi(x)} f_x(u)du.$$

Sous les hypothèses (U) et $E\varepsilon_{x,n} = 0$, on peut trouver K tel que pour $|B|>K$:

$$(a+b)\ \left|\int_{-\infty}^{B} uf_x(u)du\right| < \frac{1}{2} \qquad \text{et} \qquad (a+b)\left|\int_{B}^{+\infty} uf_x(u)du\right| < \frac{1}{2}\ .$$

Pour $|\varphi(x)| \leqslant K$ on a :

$$\int p(x,y)g(y)dy-g(x)+1 \leqslant (a+b)E|\varepsilon_{x,n}|+(a+b)K-g(x)+1$$

ce qui est négatif pour $|x|$ assez grand car l'hypothèse (U) implique que $\sup_x E|\varepsilon_{x,n}| < \infty$.

Pour $|\varphi(x)| > K$ on a :

$$\int p(x,y)g(y)dy-g(x)+1 \leqslant \frac{1}{2}+a\varphi(x)\int_{-\varphi(x)}^{+\infty} f_x(u)du-b\varphi(x)\int_{-\infty}^{-\varphi(x)} f_x(u)du-g(x)+1$$

on peut alors vérifier que cette expression est strictement négative pour $|x|>M$; faisons le pour $x > 0$; la vérification est du même type pour $x<0$. On a deux cas :

1. si $\varphi(x) > 0$:

$$\begin{aligned}\int p(x,y)g(y)dy-g(x) + 1 &\leq \frac{1}{2} + a\varphi(x) - ax + 1\\ &\leq \frac{1}{2} + a(x-A) - ax + 1\\ &\leq \frac{1}{2} + 1 - aA\end{aligned}$$

ce qui en vertu de (10) est strictement négatif.

2. si $\varphi(x) < 0$:

$$\begin{aligned}\int p(x,y)g(y)dy-g(x) + 1 &\leq \frac{1}{2} - b\varphi(x) - ax + 1\\ &\leq \frac{1}{2} - b(\alpha x+A) - ax + 1\\ &\leq \frac{1}{2} - bA + 1\end{aligned}$$

ce qui est encore strictement négatif en vertu de (10).

Comme conséquence, sous les conditions de la proposition 5, la chaîne (X_n) est Harris récurrente et μ et π sont des mesures équivalentes. On peut alors énoncer :

PROPOSITION 6. Sous les conditions (U) et (C1), s'il existe $M>0$, $\rho<1$, $\alpha<0$, $\gamma<0$, $\alpha\gamma<\rho$ tels que $\sup_{|x|\leq M} |\varphi(x)|<\infty$ et :

$$(11)\qquad \begin{aligned}\alpha x < \varphi(x) < \rho x \quad &\text{pour} \quad x > M\\ \rho x < \varphi(x) < \gamma x \quad &\text{pour} \quad x < -M\end{aligned}$$

Alors la chaîne (X_n) est géométriquement ergodique et π admet un moment d'ordre 1.

De plus : $\exists\xi$, $\xi < 1$ tel que :

$$\int \pi(dx)\cdot r_n(x) = O(\xi^n)$$

où $r_n(x) = \sup_{|h|<|x|} . \int h(y)(P^n(x,dy)-\pi(dy))$.

<u>Preuve</u> : On a $\alpha\gamma < \rho$; soit $\eta>0$,

tel que : $\rho < \eta^2 < \eta < 1$. Comme $\alpha\gamma < \eta^2$, on a : $-\frac{\gamma}{\eta} < -\frac{\eta}{\alpha}$; on prend alors deux nombres a et b, réels positifs, tels que : $-\frac{\gamma}{\eta} < \frac{b}{a} < -\frac{\eta}{\alpha}$. On a donc :

(12) $\qquad a\eta + b\alpha > 0$ et $b\eta + a\gamma > 0$

On prend maintenant $g(x) = \begin{cases} 1+ax & \text{si } s\geq 0 \\ 1-bx & \text{si } x<0 \end{cases}$. On a alors :

$$\int p(x,y)g(y)dy-\eta g(x) \leq (1-\eta)+(a+b)E|\varepsilon_x|+a\varphi(x)\int_{-\varphi(x)}^{+\infty} f_x(u)du$$
$$-b\varphi(x)\int_{-\infty}^{-\varphi(x)} f_x(u)du-\eta g(x).$$

En procédant maintenant comme dans la proposition 5 et en utilisant les inégalités (12), on voit facilement que $\int p(x,y)g(y)dy-\eta g(x)<0$ pour $|x|$ assez grand ; ce qui permet de conclure.

V. **EXEMPLE D'APPLICATION**.

Modèle bilinéaire simple.

Considérons le modèle :

(13) $$Z_t = a\, Z_{t-1} + b\, Z_{t-1}\, e_t + c\, e_t + d\, e_t^2 + f$$

où a, b, c, d, c, f sont des nombres réels et les e_t sont des variables aléatoires indépendantes de même loi, de densité $g(x)$ strictement positive. On suppose le modèle centré, c'est-à-dire :

$$E\, Z_t = 0, \quad E\, e_t = 0, \quad E(d\, e_t^2 + f) = 0.$$

Les propriétés des estimateurs des paramètres du modèle (13) semblent être difficiles à étudier (Tuan Dinh Pham-Lanh Tat Tran, 1980).

Ces auteurs établissent en particulier qu'une condition suffisante d'existence d'une solution de (13) stationnaire et de carré intégrable est :

(14) $$a^2 + b^2\, E\, e_t^2 < 1 \quad \text{et } E\, e_t^4 < \infty .$$

(cf. Mokkadem 1985 b, pour un autre type de condition d'existence).

Mais sous cette condition, Z_t est géométriquement ergodique ; en effet les conditions (C1) et (C3) avec $n_o=2$ sont faciles à vérifier pour le modèle (13) ; la proposition 3 permet de conclure. En particulier (Z_t) est géométriquement α-mélangeant. Il peut être préférable alors de modéliser autrement le processus (Z_t). En effet, si on calcule les corrélations de (Z_t) on voit que ceux sont celles d'un processus autorégressif linéaire d'ordre 1. Par conséquent Z_t s'écrit :

(15) $$Z_t = \rho\, Z_{t-1} + \varepsilon_t$$

où les ε_t sont orthogonaux. Le processus ε_t n'est pas un bruit blanc, mais il est géométriquement α-mélangeant et on peut voir que les estimations classiques dans (15) ont les propriétés usuelles. Par exemple, posons $E\, Z_t^2 = \sigma^2$, on peut estimer σ^2 et ρ par

$$\sigma_N^2 = \frac{1}{N}\sum_{j=0}^{N} Z_j^2 \quad \text{et} \quad \rho_N = \frac{C_N}{\sigma_N^2} \quad \text{où} \quad C_N = \frac{1}{N}\sum_{j=0}^{N-1} Z_j\, Z_{j+1}.$$

On a alors :

PROPOSITION 7.

1 - $\lim \rho_N = \rho$ p.s.

2 - Si $E|Z_t|^{2p+\delta} < \infty$, $p>0$, $\delta>0$, alors $E|\rho_N-\rho|^p = O(N^{-p/2})$

3 - Si $E|Z_t|^{4+\delta} < \infty$, $\delta>0$, alors $\sqrt{N}\,(\rho_N-\rho) \longrightarrow N(0,w)$ en loi

où $w = \lim \sigma^{-2} N E(C_N-\rho\sigma_N^2)^2$.

<u>Preuve</u> :

1 - est une conséquence du théorème ergodique.

2 - on écrit : $\rho_N-\rho = \dfrac{C_N-C}{\sigma^2} - \rho_N \dfrac{(\sigma_N^2-\sigma^2)}{\sigma^2}$

où $C = E(Z_t\, Z_{t+1})$.

On remarquera que $|\rho_N|\leqslant 1$ puis on applique un résultat de Yokoyama (1980).

3 - on utilise la même décomposition de $\rho_N-\rho$ puis on applique un théorème de limite centrale pour des variables aléatoires α-mélangeantes (Hall-Heyde, 1980).

La remarque qu'on vient de faire pour le modèle (13) est encore vraie pour tout modèle du même type que (1) et pour lequel les corrélations sont celles d'un processus autorégressif.

- APPENDICE -

On donne ici une condition suffisante pour que (C1) soit satisfaite.

(C′) : 1 - φ est bornée sur tout compact.

2 - Pour toute suite x_n de limite x_o, il existe une sous-suite x'_n telle que $f_{x'_n}(v)$ converge presque partout vers une densité $f_{\bar{x}}(v)$ strictement positive.

Une telle condition est satisfaite dans le modèle (2) dès que la densité de ε_n est strictement positive et φ est bornée sur tout compact.

La condition (C′) implique en particulier que $f_x(v)$ est une densité strictement positive pour tout x et par conséquent (C3) est vérifiée.

Soit maintenant K un compact ; prenons un pavé fermé I tel que :

$$\varphi(K) \subset I. \tag{16}$$

Soit A un Borélien de μ-mesure non-nulle.

Notons :

$$P_{x,z}(A) = \int_A f_x(y-z)dy. \tag{17}$$

Il est clair que :

$$P(x,A) = P_{x,\varphi(x)}(A) \tag{18}$$

et donc d'après (16) :

$$\operatorname*{Inf}_{K} P(x,A) \geqslant \operatorname*{Inf}_{K\times I} P_{x,z}(A). \tag{19}$$

Notons :

$$\alpha = \operatorname*{Inf}_{K\times I} P_{x,z}(A). \tag{20}$$

Il existe une suite (x_n, z_n) tel que :

$$\lim P_{x_n,z_n}(A) = \alpha. \tag{21}$$

$K\times I$ étant compact, on extrait une sous-suite (x'_n, z'_n) ayant une limite (x,z) ; on en extrait encore une sous-suite (x_n, z_n) de sorte que f_{x_n} converge presque partout vers une densité $f_{\bar{x}}$ strictement positive.

Notons :

(22) $$P_{\bar{x},z}(A) = \int_A f_{\bar{x}}(y-z)\,dy.$$

Il est clair que :

(23) $$P_{\bar{x},z}(A) > 0.$$

On a :

(24) $$|P_{x_n,z_n}(A) - P_{\bar{x},z}(A)| \leqslant |\int_A f_{x_n}(y-z_n)\,dy - \int_A f_{\bar{x}}(y-z_n)\,dy|$$
$$+ |\int_A f_{\bar{x}}(y-z_n)\,dy - \int_A f_{\bar{x}}(y-z)\,dy|.$$

Les deux termes de droite de (24) tendent vers zéro grâce au théorème de Scheffé et à la continuité de la translation.

On a donc d'après (21) :

(25) $$\alpha = P_{\bar{x},z}(A)$$

et par (23) :

(26) $$\alpha > 0.$$

Ce qui achève la démonstration.

REFERENCES

DOUKHAN P., GHINDES M. (1980 a), CRAS-Série I, T.290, 921-923.

(1980 b), CRAS-Série I, T.291, 81-83.

HALL P., HEYDE C.C. (1980), Martingale limit theory and its application. Academic Press.

MOKKADEM A. (1985 a), CRAS-Série I, T. 301, 375-377.

(1985 b), CRAS-Série I, T. 301, 447-450.

NUMMELIN E., TUOMINEN P. (1982), Geometric ergodicity of Harris recurrent Markov chains. Stoch. Proc. Appl. 12, 187-202.

ROSENBLATT M. (1971), Markov Processes. Structure and asymptotic behaviour. Springer Verlag.

TUAN D. PHAM, LANH T. TRAN (1980), CRAS-Série I, T. 290, 335-338.

TWEEDIE R.L. (1975), Sufficient conditions for ergodicity and recurrence of Markov chains on a general state space. Stoch. Proc. Appl. 3, 385-403.

(1976), Criteria for classifying general Markov chains. Adv. Appl. Prob. 8, 737-771.

(1983), The existence of moments for stationary Markov chains, J. Appl. Prob. 20, 191-196.

YOKOYAMA R. (1980), Moments bounds for stationary mixing sequences. Z. Wahr. verw. Gebiete 52, 45-57.

C. R. Acad. Sc. Paris, t. 301, Série I, n° 19, 1985 889

PROBABILITÉS. — *Le modèle non linéaire AR(1) général. Ergodicité et ergodicité géométrique.* Note de **Abdelkader Mokkadem**, présentée par Robert Fortet.

On établit des conditions suffisantes d'ergodicité et d'ergodicité géométrique du modèle non linéaire AR(1) général.

PROBABILITY THEORY. — The general non linear AR(1) model. Ergodicity and geometric ergodicity.

We obtain sufficient conditions for ergodicity and geometric ergodicity of the general non linear AR(1) model.

I. LE MODÈLE. — Soit une suite de champs aléatoires sur $\mathbb{R}^d$ à valeurs dans $\mathbb{R}^d$, indépendants: $\varepsilon_{x,n}$, $x \in \mathbb{R}^d$, $n \in \mathbb{Z}$. On suppose que $\varepsilon_{x,n}$ admet une densité $f_x(v)$ par rapport à la mesure de Lebesgue μ dans $\mathbb{R}^d$, indépendante de n.

On étudie dans le présent travail le comportement du processus (X_n) défini par:

$$X_n = \varphi(X_{n-1}) + \varepsilon_{X_{n-1},n}, \tag{1}$$

où φ est une fonction mesurable de $\mathbb{R}^d$ dans $\mathbb{R}^d$. Le modèle (1) est le modèle AR(1) le plus général; par exemple le modèle $X_n = \psi(X_{n-1}, \eta_n)$, certains modèles bilinéaires (Tuan, D. Pham et Lanh, T. Tran [10]) et certains modèles à coefficients aléatoires (D. F. Nicholls et B. G. Qinn [5]) s'écrivent sous la forme (1).

Le processus (X_n) défini par (1) est une chaîne de Markov homogène de densité de transition:

$$p(x, y) = f_x(y - \varphi(x)).$$

On notera $P(x, .)$ sa loi de transition.

Les hypothèses suivantes seront utilisées dans la suite:

(U) Pour n fixé les variables $|\varepsilon_{x,n}|$ sont uniformément intégrables.

(I) Pour $a > 0$, fixé, $E(\exp a|\varepsilon_{x,n}|) < C$ pour tout x.

(B) Pour tout borélien A de μ-mesure non nulle et tout compact K dans $\mathbb{R}^d$, il existe un entier n_0 strictement positif tel que:

$$\inf_{x \in K} P^{n_0}(x, A) > 0.$$

(C) La chaîne (X_n) est apériodique.

La condition (C) est vérifiée si par exemple:

(C') $\exists n_1 \in \mathbb{N}^*$ tel que $P^{n_1}(x, .) \sim \mu$ pour tout x.

Les conditions (B) et (C') peuvent être remplacées par des conditions plus fortes mais plus pratiques:

PROPOSITION 1. — *Les conditions* (B) *et* (C') *sont vérifiées si:*

(*a*) *φ est bornée sur tout compact.*

(*b*) *Pour toute suite x_n de limite x_0, il existe une sous-suite x'_n telle que $f_{x'_n}(v)$ converge presque partout vers une densité $\bar{f}_{x_0}(v)$ strictement positive.*

Sous l'hypothèse (B) la chaîne (X_n) est μ-irréductible; elle admet alors une mesure excessive (*cf.* [11]) qu'on notera π; dans le cas où (X_n) est ergodique π sera sa probabilité invariante.

Pour les définitions des différentes notions qu'on utilise on pourra se reporter à [6], [7], [8], [11] et [12].

Les résultats énoncés dans la présente Note sont démontrés dans [3].

0249-6291/85/03010889 $ 2.00

II. ÉTUDE DU CAS VECTORIEL. — La condition (B) permet d'utiliser un résultat de [12] pour obtenir le :

THÉORÈME 1. — *Si* (B) *est vérifiée et s'il existe* $M>0$, $\eta>0$ *et* $s>0$ *tels que :*

$$E|\varphi(x)+\varepsilon_{x,n}|^s \leqq |x|^s-\eta \tag{2}$$

pour

$$|x|>M \quad \text{et} \quad \sup_{|x|\leqq M} E|\varphi(x)+\varepsilon_{x,n}|^s<\infty,$$

alors (X_n) *est récurrente positive.*

Remarque 1. — Si pour $s\leqq 1$ et $|x|>M$ on a :

$$|\varphi(x)|^s \leqq |x|^s - A \quad \text{avec} \quad A>\sup_x E|\varepsilon_{x,n}|^s \quad \text{et} \quad \sup_{|x|\leqq M} |\varphi(x)|<\infty,$$

alors (2) est vérifiée. Cela correspond au résultat de Doukhan-Ghindes [2] dans le cas $\varepsilon_{x,n}=\varepsilon_n$ et $s=1$.

La condition du théorème 1 porte sur le moment de $|\varepsilon_{x,n}|^s$; on peut lever cette contrainte dans le cas où $E\varepsilon_{x,n}=0$. On a alors le :

THÉORÈME 2. — *Si* $E\varepsilon_{x,n}=0$ *et* (B) *et* (U) *sont vérifiées et si* φ *possède pour* $|x|$ *assez grand la propriété :*

$$\forall K>0,\ \exists K' \text{ tel que } |\varphi(x)|>K' \Rightarrow |\varphi_i(x)|>K \tag{3}$$

pour toute composante φ_i *de* φ, *alors s'il existe* $M>0$ *et* $\eta>0$ *tels que :*

$$\sum|\varphi_i(x)|<\sum|x_i|-\eta \quad \text{pour} \quad |x|>M \quad \text{et} \quad \sup_{|x|\leqq M}|\varphi(x)|<\infty. \tag{4}$$

la chaîne (X_n) *est récurrente positive.*

Pour la démonstration on applique les résultats de [12] en se plaçant dans chacun des 2^d trièdres de $\mathbb{R}^d$.

Comme conséquence on a :

PROPOSITION 2. *Sous les conditions du théorème* 1 *ou du théorème* 2, *la chaîne* X_n *est* π-*récurrente.*

Ce résultat permet l'étude de l'ergodicité géométrique de (X_n). On se réfère à [6] et [13] pour les résultats de base sur l'ergodicité géométrique des chaînes de Markov Harris-récurrentes et apériodiques. Dans [6] il est montré en particulier que si (X_n) est une chaîne de Markov Harris-récurrente, apériodique, géométriquement ergodique et de probabilité invariante π alors :

$$\int \pi(dx)\,\|P^n(x,\,.)-\pi\|=O(\rho^n),$$

où $\rho<1$ et $\|.\|$ désigne la norme en variation.

Une telle propriété entraîne que le processus (X_n) est fortement mélangeant avec un coefficient de mélange convergeant vers zéro géométriquement [9]. Ce type de propriété est utile pour l'étude statistique de (X_n) ([1], [4]).

Le résultat qui suit donne un critère d'ergodicité géométrique de (1).

THÉORÈME 3. — *Si* (B) *et* (C) *sont vérifiées et s'il existe* $M>0$, $s>0$ *et* $\rho<1$ *tels que :*

$$E|\varphi(x)+\varepsilon_{x,n}|^s<\rho|x|^s \quad \text{pour} \quad |x|>M \tag{5}$$

et

$$\sup_{|x|\leqq M} E|\varphi(x)+\varepsilon_{x,n}|^s<\infty,$$

alors (X_n) *est géométriquement ergodique et* π *admet un moment d'ordre s.*

Remarque 2. — Si $\sup_x E|\varepsilon_{x,n}|^s < \infty$, alors (5) est équivalente à: $|\varphi(x)| < \rho|x|$ pour $|x| > M$ et $\sup_{|x| \leq M} |\varphi(x)| < \infty$.

Remarque 3. — Dans le cas $\varepsilon_{x,n} = \varepsilon^n$, la condition d'ergodicité géométrique connue jusqu'ici est celle de [2]: φ bornée.

Sous l'hypothèse (I) le théorème 3 peut être amélioré; on obtient:

THÉORÈME 4. — *Si* (B), (C) *et* (I) *sont vérifiées et s'il existe* $M > 0$ *et* $A > (1/a) \log C$ *tels que*:

$$|\varphi(x)| < |x| - A \quad \textit{pour} \quad |x| > M \quad \textit{et} \quad \sup_{|x| \leq M} |\varphi(x)| < \infty, \tag{6}$$

alors (X_n) *est géométriquement ergodique et* π *admet une transformée de Laplace sur* $[-a, a]$.

III. CAS SCALAIRE. — Dans ce cas la relation d'ordre sur $\mathbb{R}$ permet d'affaiblir les conditions établies dans le cas vectoriel.

On suppose (U) vérifiée et $E\varepsilon_{x,n} = 0$. On obtient alors:

THÉORÈME 5. — *Sous la condition* (B), *s'il existe deux applications affines* $l(x)$ *et* $m(x)$ *décroissantes et deux constantes* $A > 0$ *et* $M > 0$ *tels que*:

$$\sup_{|x| \leq M} |\varphi(x)| < \infty$$

et pour $|x| > M$:

$$\left\{\begin{array}{ll} l(m(x)) \leqq x & \textit{pour} \quad x > 0, \\ m(l(x)) \geqq x & \textit{pour} \quad x < 0, \\ l(0) < 0, & m(0) > 0, \\ m(x) \leqq \varphi(x) \leqq x - A & \textit{pour} \quad x > 0, \\ x + A \leqq \varphi(x) \leqq l(x) & \textit{pour} \quad x < 0, \end{array}\right. \tag{7}$$

alors la chaîne (X_n) *est récurrente positive.*

Pour l'ergodicité géométrique on a:

THÉORÈME 6. — *Sous les conditions* (B) *et* (C), *s'il existe deux applications linéaires* $l(x)$ *et* $m(x)$ *décroissantes et deux constantes* $M > 0$ *et* $\rho < 1$ *tels que*:

$$\sup_{|x| \leq M} |\varphi(x)| < \infty$$

et pour $|x| > M$:

$$\left\{\begin{array}{ll} l(m(x)) < \rho x & \textit{pour} \quad x > 0, \\ m(l(x)) > \rho x & \textit{pour} \quad x < 0, \\ m(x) < \varphi(x) < \rho x & \textit{pour} \quad x > 0, \\ \rho x < \varphi(x) < l(x) & \textit{pour} \quad x < 0, \end{array}\right. \tag{8}$$

alors la chaîne (X_n) *est géométriquement ergodique et* π *admet un moment d'ordre* 1.

IV. EXEMPLES DE MODÈLES DE TYPE (1). — 1. *Considérons le modèle* [10]:

$$X_t = a X_{t-1} + b X_{t-1} e_t + c e_t + d e_t^2 + f, \tag{9}$$

avec a, b, c, d, e, f des nombres réels et e_t des variables aléatoires réelles indépendantes de même loi et de densité strictement positive.

On suppose le modèle centré : $EX_t = 0$, $E e_t = 0$, $E(de_t^2 + f) = 0$.

Dans [10], Tuan Dinh Pham et Lanh Tat Tran établissent qu'une condition suffisante d'existence d'une solution de (9) stationnaire et de carré intégrable est :

(10) $$a^2 + b^2 \, E e_t^2 < 1.$$

En fait, on peut facilement vérifier les conditions (B) et (C') avec $n_0 = n_1 = 2$ pour le modèle (9); ce qui en appliquant le théorème 3 montre que, sous la condition (10), (X_t) est géométriquement ergodique et sa probabilité invariante admet un moment d'ordre 2.

2. *Soit le modèle* [5] :

(11) $$X_t = \theta X_{t-1} + \gamma_t X_{t-1} + \eta_t,$$

où $\theta \in \mathbb{R}$, (γ_t) et (η_t) sont des suites de variables aléatoires indépendantes de même loi et indépendantes entre elles. On suppose que η_t admet une densité strictement positive et un moment d'ordre s. Ici également les conditions (B) et (C') sont vérifiées avec $n_0 = n_1 = 1$ et par conséquent sous la condition : $E|\gamma_t + \theta|^s < 1$ la chaîne (X_n) définie par (11) est géométriquement ergodique et sa probabilité invariante admet un moment d'ordre s.

Remise le 1er juillet 1985, acceptée après révision le 21 octobre 1985.

RÉFÉRENCES BIBLIOGRAPHIQUES

[1] G. COLLOMB et P. DOUKHAN, *Comptes rendus*, 296, série I, 1983, p. 859-862.

[2] P. DOUKHAN et M. GHINDES, *Comptes rendus*, 290, série A, 1980, p. 921-923.

[3] A. MOKKADEM, Prépublication, 1985.

[4] A. MOKKADEM, *Comptes rendus*, 301, série I, 1985, p. 447-450.

[5] D. F. NICHOLLS et B. G. QINN, Random coefficient autoregressive model: An introduction, *Lectures Notes in Statistics*, 11, 1982.

[6] E. NUMMELIN et P. TUOMINEN, Geometric ergodicity of Harris recurrent Markov chains, *Stoch Proc. Appl.*, 12, 1982, p. 187-202.

[7] S. OREY, *Limit theorems for Markov chain transition probabilities*, Van Nostrand, London, 1971.

[8] D. REVUZ, *Markov chains*, North Holland, Amsterdam, 1984.

[9] M. ROSENBLATT, *Markov processes. Structure and asymptotic behaviour*, Springer Verlag, 1971.

[10] D. TUAN PHAM et LANH T. TRAN, *Comptes rendus*, 290, série A, 1980, p. 335-338.

[11] R. L. TWEEDIE, R-theory for Markov chains on general state space I, *Ann. Prob.*, 2, 1974, p. 840-864.

[12] R. L. TWEEDIE, Sufficient conditions for ergodicity and recurrence of Markov chains on a general state space, *Stoch. Proc. Appl.*, 3, 1975, p. 385-403.

[13] R. L. TWEEDIE, The existence of moments for stationary Markov chains, *J. Appl. Prob.*, 20, 1983, p. 191-196.

C.N.R.S. U.A. n° 743, Statistique appliquée, Bât. 425, 91405 *Orsay Cedex*.

C. R. Acad. Sc. Paris, t. 301, Série I, n° 7, 1985 375

PROBABILITÉS. — *Conditions suffisantes d'existence et d'ergodicité géométrique des modèles bilinéaires.* Note de **Abdelkader Mokkadem**, présentée par Robert Fortet.

Considérant des modèles bilinéaires donnés par des équations :

$$X_t=\sum_{j=1}^{p} a_j X_{t-j}+e_t+\sum_{j=1}^{q} b_j e_{t-j}+\sum_{j=1}^{Q}\sum_{j=0}^{P} b_{kj} X_{t-j-k} e_{t-k},$$

on établit des conditions suffisantes d'existence, d'ergodicité et d'ergodicité géométrique de (X_t).

PROBABILITY THEORY. — Sufficient conditions for the existence and geometric ergodicity of bilinear models.

Starting from the equations:

$$X_t=\sum_{j=1}^{p} a_j X_{t-j}+e_t+\sum_{j=1}^{q} b_j e_{t-j}+\sum_{k=1}^{Q}\sum_{j=0}^{P} b_{kj} X_{t-j-k} e_{t-k},$$

we obtain sufficient conditions for the existence, ergodicity and geometric ergodicity of (X_t) *for this kind of bilinear models.*

I. INTRODUCTION. — On considère le modèle bilinéaire :

$$X_t=\sum_{j=1}^{p} a_j X_{t-j}+e_t+\sum_{j=1}^{q} b_j e_{t-j}+\sum_{k=1}^{Q}\sum_{j=0}^{P} b_{kj} X_{t-j-k} e_{t-k}, \tag{1}$$

où e_t est une suite de variables aléatoires réelles indépendantes et de même loi.

Tuan Pham Dinh [3] a montré qu'on pouvait écrire ce modèle sous la forme :

$$\begin{cases} Z_t=AZ_{t-1}+BZ_{t-1}e_t+Ce_t+De_t^2+E, \\ X_t=H'Z_{t-1}+e_t, \end{cases} \tag{2}$$

où Z_t est un processus vectoriel de dimension $n=\mathrm{Max}\,(p,\ q+P,\ Q+p)$, les matrices A, B et les vecteurs C, D, E étant fonctions des paramètres du modèle (1).

Quand e_t est indépendant de $\{Z_{t-1}, Z_{t-2}, \ldots\}$, (2) est une représentation markovienne de (1). Le problème de l'existence d'une solution stationnaire Z_t est abordé dans [1] où une condition suffisante, portant sur le moment d'ordre deux de e_t, est établie.

Dans la présente Note, on montre que : *sous la condition* $E\log|A+Be_t|<0$, *le modèle* (2) *admet une solution stationnaire; de plus si* $E|A+Be_t|^s<1$ *et* $s\leqq 1$, *la solution est géométriquement ergodique.*

II. EXISTENCE.

THÉORÈME 1. — *Si* $E|\log|A+Be_t||<\infty$, *alors une condition suffisante pour l'existence d'une solution* Z_t *strictement stationnaire de* (2) *est* :

$$E\log|A+Be_t|<0. \tag{3}$$

On a de plus :

$$Z_t=Ce_t+De_t^2+E+\sum_{m\geq 1}\left(\prod_{j=1}^{m}(A+Be_{t-j})\right)(Ce_{t-m}+De_{t-m}^2+E). \tag{4}$$

Preuve. — On peut voir que, sous les conditions du théorème, toute solution de (2) doit vérifier (4). On voit également que si le terme de droite de (4) existe, il vérifie (2). Il suffit par conséquent de montrer que la série de terme général :

$$M_m=\prod_{j=1}^{m}(A+Be_{t-j})(Ce_{t-m}+De_{t-m}^2+E) \quad \text{converge presque sûrement.}$$

0249-6291/85/03010375 $ 2.00 © Académie des Sciences

Mais :

$$\log|M_m| \leqq \sum_{j=1}^{m} \log|A+B e_{t-j}| + |C e_{t-m} + D e_{t-m}^2 + E|.$$

Par la loi des grands nombres on obtient :

$$\limsup_m |M_m|^{1/m} \leqq \exp(E \log|A+B e_t|) < 1 \quad \text{presque sûrement.}$$

Ce qui donne la convergence voulue.

Remarques. — 1° L'écriture (4) montre que e_t est indépendant de $\{Z_{t-1}, Z_{t-2}, \ldots\}$ et donc Z_t est markovien.

2° Si $E|e_t|^r < \infty$, $r > 0$, on a par l'inégalité de Jensen :

$$\|A+B e_t\|_r < 1 \quad \Rightarrow \quad E \log|A+B e_t| < 0.$$

III. Ergodicité. — On considère le modèle (2) avec A, B, C, D, E scalaires et Z_t la solution définie par (4).

On suppose que :

(A) e_t admet une densité $f(x) > 0$ presque partout.

(B) $E|e_t|^s < \infty$ où $0 < s \leqq 1$.

Sous la condition (A), (Z_t) est une chaîne de Markov μ-irréductible, apériodique, μ étant la mesure de Lebesgue, et sa loi de transition $P(x, .)$ est fortement continue.

Théorème 2. — *Sous les conditions* (A) *et* (B), *s'il existe* $s' \leqq s$ *tel que* $E|A+B e_t|^{s'} < 1$, *alors la chaîne* (Z_t) *est ergodique et récurrente.*

Preuve. — On a $E|A+B e_t|^{s''} < 1$ pour tout $s'' \leqq s'$, on peut donc prendre $s' \leqq s/2$ de sorte que $E(e_t^2)^{s'}$ existe. On applique maintenant les théorèmes 4.2 et 4.3 de Tweedie [5] avec $g(x) = |x|^{s'}$; on obtient :

$$(5) \qquad \int P(x, dy)\, g(y) - g(x) + \leqq E|A+B e_t|^{s'} |x|^{s'} + E|C e_t + D e_t^2 + F|^{s'} - |x|^{s'} + 1,$$

ce qui, pour $|x|$ assez grand, est négatif.

Le reste est simple à vérifier.

Corollaire. — *Sous les hypothèses du théorème* 2, *la chaîne* (Z_t) *est Harris-récurrente.*

Preuve. — En utilisant le lemme 2.2 de [5] on peut montrer que la chaîne (Z_t) est μ-récurrente. Comme (Z_t) est ergodique de probabilité invariante π il s'ensuit, d'après Orey [2], que (Z_t) est π-récurrente et donc Harris-récurrente.

Théorème 3. — *Sous les hypothèses du théorème* 2, (Z_t) *est une chaîne géométriquement ergodique et sa probabilité invariante admet un moment d'ordre* $s'' = \operatorname{Inf}(s', s/2)$.

De plus, il existe $\rho < 1$ *tel que :*

$$\int \pi(dx)\, \|P^n(x, .) - \pi\| = O(\rho^n),$$

où $\|\mu\| = \sup_{|h| < |x|^{s''}} \int \mu(dy)\, h(y)$ *pour toute mesure* μ.

Preuve. — On utilise le résultat de Nummelin-Tuominen [4] pour montrer que tout compact est un petit ensemble.

On choisit ensuite $\varepsilon > 0$ tel que $E|A+B e_t|^{s''} < 1 - \varepsilon < 1$ et on définit $g(x) = |x|^{s''}$.

On obtient :

$$\int P(x, dy) g(y) - (1-\varepsilon) g(y) < 0,$$

pour $|x|$ assez grand.

Les résultats de [4] et [6] permettent alors de conclure.

Remise le 17 juin 1985.

RÉFÉRENCES BIBLIOGRAPHIQUES

[1] M. BHASKARA RAO, T. SUBBA RAO et A. M. WALKER, On the existence of some bilinear time series models, *J.T.S.A.*, 4, 1983, p. 95-110.

[2] S. OREY, *Limit theorems for Markov chains*, New York.

[3] TUAN PHAM DINH, *Bilinear markovian representation and bilinear models*, Technical report n° 161 UMIST, 1983.

[4] E. NUMMELIN et P. TUOMINEN, Geometric ergodicity of Harris recurrent Markov chains with applications to renewal theory, *Stoch. Proc. Appl.*, 12, 1982, p. 187-202.

[5] R. L. TWEEDIE, Sufficient conditions for ergodicity and recurrence of Markov chains, *Stoch. Proc. Appl.*, 3, 1975, p. 385-403.

[6] R. L. TWEEDIE, The existence of moments for stationary Markov chains, *J. Appl. Prob.*, 20, 1983, p. 191-196

U.A. n° 743 C.N.R.S., Université Paris-Sud, Statistique appliquée,
Bât. n° 425, *Mathématique*, 91405 *Orsay*.

C H A P I T R E II

Mixing properties of ARMA processes

A. MOKKADEM

Abstract : Using Markovian method, we obtain a simple sufficient condition for a multivariate ARMA process to be geometricaly completely regular.

1. Introduction

Let $(Y(t))_{t \in \mathbb{Z}}$ be a stationary random process in $\mathbb{R}^{\ell}$; denote by $\mathcal{A}_k$ the σ-algebra generated by $\{Y(t) \; ; \; -\infty < t \leq k\}$ and by $\mathcal{A}^k$ those generated by $\{Y(t) \; ; \; k \leq t < \infty\}$. Following Davydov [2], we define the complete regularity coefficient for $k > 0$ by :

$$\beta(k) = E\,\{\sup_{B \in \mathcal{A}^k} |P(B/\mathcal{A}_o) - P(B)|\}.$$

We shall say that $(Y(t))_{t \in \mathbb{Z}}$ is completely regular, if $\lim_{k \to \infty} \beta(k) = 0$, and that $(Y(t))_{t \in \mathbb{Z}}$ is geometricaly completely regular if there exists ρ, $0 < \rho < 1$ such that $\beta(k) = O(\rho^k)$.

Another mixing coefficient is the strong mixing coefficient defined by :

$$\alpha(k) = \sup\{|P(A \cap B) - P(A).P(B)| \; ; \quad A \in \mathcal{A}_o, \quad B \in \mathcal{A}^k\}.$$

It is easy to see that $\alpha(k) \leq \beta(k)$; so that complete regularity implies strong mixing. The mixing properties are very useful in asymptotic statistic. It is then of interest to obtain conditions for a process to be completely regular or strong mixing.

Ibragimov and Rozanov [4] obtain conditions for linear gaussian process to be strong mixing ; the one dimensional linear process without gaussian hypothesis, is studied by Gorodetskii [3] and Withers [10]. In [7] Tuan D. Pham and Lanh T. Tran show that under Gorodetskii's conditions, the multivariate linear process is completely regular.

In the present paper, using another approach, we obtain a simple condition, different from that of [7], for a multivariate ARMA process to be geometricaly completely regular.

2. Main result

We assume that the process $(Y(t))_{t \in \mathbb{Z}}$ satisfies the ARMA equation :

$$(1) \qquad \sum_{i=0}^{P} B(i)Y(t-i) = \sum_{k=0}^{Q} A(k)\varepsilon(t-k)$$

where $B(i)$ and $A(k)$ are real matrices with respective dimensions $\ell \times \ell$ and $\ell \times r$; $B(0) = Id$; $\varepsilon(t)$ is a sequence of independent identicaly distribued random vectors in $\mathbb{R}^r$ and $E\varepsilon(t) = 0$.

For $z \in \mathbb{C}$ we define the matrices

$$P(z) = \sum_{i=0}^{P} B(i)z^i \quad \text{and} \quad Q(z) = \sum_{k=0}^{Q} A(k)z^k$$

and assume that :

(C) The absolute values of the zeros of the polynome $\det P(z)$ are strictly greater than 1 .

With this condition, the equation (1) has a unique stationary solution :

(1')
$$Y(t) = \sum_{j=0}^{\infty} c(j)\varepsilon(t-j)$$

Our main result is :

THEOREM 1. *If the probability law of* $\varepsilon(t)$ *is equivalent to the Lebesgue measure on* $\mathbb{R}^r$*, then the processus* $Y(t)$ *is geometricaly completely regular.*

Similar result can be written for ARMA processes in $\mathbb{C}^\ell$. Now we state the problem in Markovian form. For this we use the following lemma :

LEMMA 1. (Akaîke [1]) *Let* $k = \mathrm{Max}(P,Q+1)$ *; there exists a Markovian chain* $X(t)$ *in* $\mathbb{R}^k$*, such that :*

(2)
$$X(t) = F.X(t-1) + G\varepsilon(t)$$
$$Y(t) = H\, X(t)$$

where F, G, H *are real matrices and the non zero proper values of* F *are the inverses of the zeros of* $\det P(z)$. *Moreover* $(X(t))_{t\in\mathbb{Z}}$ *is stationary and* $\varepsilon(t)$ *and* $(X(t-1),X(t-2),\dots)$ *are independent.*

The proof of lemma 1 uses equations (1) and (1'). It would be noted that Akaîke in his paper [1], assumes that $\ell = r$ and that the absolute values of the zeros of $\det Q(z)$ are strictly greater than 1 ; by this assumption $\varepsilon(t)$ is the innovation of the process $Y(t)$. However, this assumption is not necessary because in the proof of Akaîke it suffices to take conditional expectation with respect to the σ-algebra generated by $\{\varepsilon(t),\varepsilon(t-1),\ldots\}$ in place of the σ-algebra generated by $\{Y(t),Y(t-1),\ldots\}$.

Clearly, using lemma 1, we see that if $X(t)$ is geometricaly completely regular then theorem 1 is proved.

In lemma 1, the matrices F, G, H obtained from (1) and (1') are of particular form. We donot need this and shall prove the following more general result :

THEOREM 1'. *Let* $(X(t))_{t\in\mathbb{Z}}$ *be a stationary random process in* $\mathbb{R}^k$ *such that :*

$$X(t) = F\,X(t-1) + G_1\varepsilon_1(t)+\ldots+G_r\varepsilon_r(t) \tag{3}$$

and :

i) $G_1,\ldots,G_r$ *are vectors in* $\mathbb{R}^k$ *and* F *is a* k*-dimensional real square matrice ;*

ii) $\varepsilon(t) = (\varepsilon_1(t),\ldots,\varepsilon_r(t))$, $t\in\mathbb{Z}$, *is a sequence of independent identicaly distribued random vectors in* $\mathbb{R}^r$ *and* $\varepsilon(t)$ *and* $\{X(t-1),X(t-2),\ldots\}$ *are independent ;*

iii) *the probability law of* $\varepsilon(t)$ *is equivalente to the Lebesgue measure on* $\mathbb{R}^r$ *and there exists a real number* $s>0$ *such that :*

$$E|\varepsilon(t)|^s<\infty\ ;$$

iv) *the absolute values of the proper values of* F *are strictly smaller than* 1.

Then $X(t)$ *is a geometricaly completely regular process and a geometricaly ergodic Markov chain.*

3. Proof of theorem 1'

To prove theorem 1', we need the following lemma that appears in another form in [5] :

LEMMA 2. *Let* $X(t)$ *be a Markov chain in a subspace* E *of* $\mathbb{R}^k$ *with transition probability* $P(x,.)$ *; assume that :*

i) $\exists n_o \in \mathbb{N}^*$ *such that, for every* x *in* E, $P^{n_o}(x,.) \ll \mu$ *, where* μ *is the Lebesgue measure on* E *;*

ii) *for every compact* K *and measurable set* A *in* E *such that* $\mu(A) > 0$, *there exists an integer* $n_1 \geq n_o$ *such that :*

$$\inf_{x \in K} P^{n_1}(x,A) > 0 \quad ;$$

iii) *there exists a positive function* g, *defined on* E, *and real numbers* $A > 0$, $B > 0$, $M > 0$, $0 < \theta < 1$ *such that :*

$$\theta g(x) \leq g(x) - 1 \qquad \text{for } |x| > M$$

$$E(g(X(t+1)) | X(t) = x) \leq \theta g(x) \qquad \text{for } |x| > M$$

$$B \leq E(g(X(t+1)) | X(t) = x) \leq A \qquad \text{for } |x| \leq M.$$

Then the Markov chain $X(t)$ *is geometricaly ergodic. Moreover, if* $X(t)$ *is a stationary random process, then it is geometricaly completely regular.*

Proof. By i) and ii), $X(t)$ is μ-irreductible and aperiodic. Let π be a subinvariant measure and K a compact in E, then

$$\mu(K) > 0 \Rightarrow 0 < \pi(K) < \infty. \tag{4}$$

To prove (4) we note that $\pi >> \mu$ and then $\pi(K) > 0$.

Moreover, π and μ being σ-finite, we can find A such that :

$$\mu(A) > 0 \quad \text{and} \quad \pi(A) < \infty \;;$$

but by ii) there exists n_1 such that :

$$b = \operatorname*{Inf}_{x \in K} P^{n_1}(x,A) > 0 \;;$$

so that :

$$\pi(A) \geq \int_E \pi(dx) P^{n_1}(x,A) \geq \pi(K).b$$

and then $\pi(K) < \infty$.

Now by (4) we can take compact with non zero μ-measure in the ergodicity criteria of Tweedie [8] : iii) implies that the chain $X(t)$ is ergodic.

We denote also by π its invariant probability.

By (i) and (ii) we obtain that the chain is π-recurrent and that π and μ are equivalent ; consequently the chain $X(t)$ is Harris-recurrent and by (ii) the compacts K such that $\mu(K) > 0$, are small sets as defined by Nummelin and Tuominen [6]. We can then use geometric ergodicity criteria of [6]. The condition iii) implies that the chain $X(t)$ is geometricaly ergodic.

The first part of the lemma is proved.

Assume now that $(X(t))_{t \in \mathbb{Z}}$ is a stationary process. π is then the probability law of the random variable $X(t)$.

But by Nummelin and Tuominen [6], there exists ρ, $0<\rho<1$, such that :

$$\int \pi(dx) \| P^n(x,.) - \pi \| = O(\rho^n)$$

where $\| . \|$ is the total variation.

In anotherway, Davydov [2] shows that :

$$\beta(n) = \int \pi(dx) \| P^n(x,.) - \pi \|$$

Then $(X(t))_{t \in \mathbb{Z}}$ is a geometricaly completely regular random process.

Now, we study the space of values of the process $X(t)$ defined by (3) and its transition probability.

For this, denote by E the real vector space generated by :

$$\{F^p G_k \ ; \ p \geq 0, \ 1 \leq k \leq r\} \ ;$$

and let m be the smaller integer such that $\{F^p G_k \ ; \ 0 \leq p \leq m, \ 1 \leq k \leq r\}$ generate E.

We have then :

Lemma 3. $X(t) \in E$ *almost surely and if* $P(x,.)$ *is the transition probability of the chain* $(X(t))$ *in* E *then* $P^{m+1}(x,.)$ *is equivalente to the Lebesgue measure in* E.

Proof. Let V_{ji} ; $j = 1,\ldots,u$, $i = 1,\ldots,\gamma_j$; be a Jordan basis ; that is to say a basis of $\mathbb{C}^k$ such that :

$$FV_{j1} = \rho_j V_{j1} \quad \text{and for } i > 1, \quad FV_{ji} = \rho_j V_{ji} + V_{j,i-1}$$

where $\rho_1,\ldots,\rho_u$ are the proper values of F (not necessarely distinct).

Let $\mathcal{B} = \{U_1,\ldots,U_v\}$ be a real basis of E. We complete $\mathcal{B}$ in $\mathbb{C}^k$, by using the V_{ji} ordered in the following mean :

$$(j,i) < (j',i') \quad \text{if} \begin{cases} j < j' \\ \text{or} \\ j = j' \quad \text{and} \quad i < i' \end{cases}$$

we obtain then a basis of $\mathbb{C}^k$, $\{U_1,\ldots,U_v,U_{v+1},\ldots,U_k\}$ with the following property :

$$(5) \qquad \begin{cases} \text{for} \quad k \leq v, \quad FU_k \in E \\ \text{for} \quad k > v, \quad FU_k = \rho_k U_k + W \text{ , where } \rho_k \text{ is some proper value} \\ F \quad \text{and} \quad W \in \langle U_1,\ldots,U_{k-1}\rangle . \end{cases}$$

Assume now that :

$$X(t) = \alpha_1(t)U_1+\ldots+\alpha_v(t)U_v + \alpha_{v+1}(t)U_{v+1}+\ldots+\alpha_{v+p}(t)U_{v+p}$$

with $p>0$ and $\alpha_{v+p}(t) \neq 0$ (the $\alpha_k(t)$ are random variables). Clearly $\alpha_{v+p}(t)$ is a stationary process.

But by (3) and (5) :

$$\alpha_{v+p}(t+1) = \rho_{v+p}\alpha_{v+p}(t)$$

and, because $|\rho_{v+p}|<1$, we h e necessarely :

$$\alpha_{v+p}(t) = 0 \quad \text{almost surely ;}$$

so the first part of the lemma is proved.

Now the equation (3) gives :

$$X(t+m) = F^{m+1}X(t-1) + \sum_{k=1}^{r} (\varepsilon_k(t)F^mG_k+\ldots+\varepsilon_k(t+m)G_k)$$

Let $L : \mathbb{R}^{r(m+1)} \to E$ defined by :

$$L(\varepsilon(t),\ldots,\varepsilon(t+m)) = \sum_{k=1}^{r} (\varepsilon_k(t)F^mG_k+\ldots+\varepsilon_k(t+m)G_k)$$

Clearly L is onto and linear.

Then the probability law of $L(\varepsilon(t),\ldots,\varepsilon(t+m))$ is equivalente to the Lebesgue measure in E.

Now, $P^{m+1}(x,.)$ is the law of $L(\varepsilon(t),\ldots,\varepsilon(t+m))+F^{m+1}x$ and then is equivalente to the Lebesgue measure in E.

Note that $P^{m+1}(x,.)$ is obtained by a translation and then it is a continuous function of x. ∎

Now we go to :

<u>Proof of theorem 1'</u>. We apply lemma 2. Clearly, using lemma 3, i) and ii) of lemma 2 are satisfied.

For iii) we take the function $g(x)$ defined as follow.

For $x \in E$, $x = \sum_{j=1}^{v}\sum_{i=1}^{\gamma_j} \alpha_{ji} V_{ji}$, put :

$$g(x) = \left[1 + \sum_{j=1}^{v} \sum_{i=1}^{\gamma_j} |\rho_j|^{(\gamma_j - i)\beta} |\alpha_{ji}|\right]^s$$

where β is such that $|\rho_j|^\beta + |\rho_j| < \xi < 1$ for every j.

Now write :

$G_1\varepsilon_1(t)+\ldots+G_r\varepsilon_r(t) = \sum_{ij} \delta_{ji} V_{ji}$, (the δ_{ji} are fixed linear functions of $\varepsilon_1(t),\ldots,\varepsilon_r(t)$).

We have :

$$E(g(X(t+1))|X(t) = x) = E(g(F(x)+\sum_{ij} \delta_{ji} V_{ji}))$$

$$\leq \left[1+ \sum_{j=1}^{v} (\sum_{i=1}^{\gamma_j-1} |\rho_j|^{(\gamma_j-i)\beta} |\rho_j\alpha_{ji}+\alpha_{j,i+1}|+|\rho_j||\alpha_{j\gamma_j}|)\right]^s$$

$$+ E\left[\sum_{j=1}^{v} (\sum_{i=1}^{\gamma_j-1} |\rho_j|^{(\gamma_j-i)\beta} |\delta_{ji}|+|\delta_{j\gamma_j}|)\right]^s$$

$$\leq \xi^s g(x) + \Delta.$$

Now clearly $\lim_{|x|\to\infty} g(x) = +\infty$.

And then for $\theta < 1$, $\theta > \xi^s$, there exists $M > 0$ such that for $|x| > M$:

$$\xi^s g(x) + \Delta < \theta g(x).$$

The theorem 1' is then proved.

REFERENCES

[1] AKAÏKE, H.- *Markovian representation of stochastic processes.* Ann. Inst. Stat. Math. 26(1974)363-387.

[2] DAVYDOV, YU, A. - *Mixing conditions for Markov chains.* Th. Prob. Appl. 18(1973)312-328.

[3] GORODETSKII, V.V. - *On the strong mixing property for linear sequences.* Th. Prob. Appl. 22(1977)411-413.

[4] IBRAGIMOV, I.A., ROZANOV, Y. - *Processus aléatoires gaussiens.* MIR. Moscou

[5] MOKKADEM, A. - *Sur un modèle autorégressif non linéaire. Ergodicité et ergodicité géométrique.* J.T.S.A., vol.8, n°2(1987) 195-204.

[6] NUMMELIN, E., TUOMINEN, P. - *Geometric ergodicity of Harris-recurrent Markov chains.* Stoch. Proc. Appl.12(1982)187-202.

[7] TUAN D. PHAM, LANH T. TRAN - *Some mixing properties of time series models.* Stoch. Proc. Appl. 19(1986)297-303.

[8] TWEEDIE, R.L. - *Sufficient conditions for ergodicity and recurrence of Markov chains.* Stoch. Proc. Appl. 3(1975)385-403.

[9] TWEEDIE, R.L. - *Criteria for classifying general Markov chains.* Adv. Appl. Prob. 8(1976)737-771.

[10] WITHERS, C.S. - *Conditions for linear processes to be strong mixing.* Z. Wahr. Verw. Gebiete 57(1981)481-494.

C H A P I T R E III

PROPRIETES DE MELANGE DES PROCESSUS AUTOREGRESSIFS POLYNOMIAUX

Abdelkader MOKKADEM

Résumé

On étudie le mélange des processus $(Z_n)_{n\in\mathbb{Z}}$ stationnaires en loi dans $\mathbb{R}^p$ et vérifiant une équation de récurrence :

$$Z_n = \phi(Z_{n-1}, e_n)$$

où ϕ est une application polynomiale de degré un relativement à Z_{n-1} et $(e_n)_{n\in\mathbb{Z}}$ est une suite de variables aléatoires i.i.d. dans $\mathbb{R}^m$. On obtient des conditions suffisantes simples pour que le processus $(Z_n)_{n\in\mathbb{Z}}$ soit géométriquement absolument régulier. Au préalable on établit un théorème de continuité de l'image d'une mesure par une application polynomiale. Les résultats s'appliquent aux processus ARMA et aux processus bilinéaires.

Abstract

We study the mixing of law stationary processes $(Z_n)_{n\in\mathbb{Z}}$ in $\mathbb{R}^p$ satisfying an equation :

$$Z_n = \phi(Z_{n-1}, e_n)$$

where ϕ is a polynomial application with $\deg \phi = 1$ relatively to Z_{n-1} and $(e_n)_{n\in\mathbb{Z}}$ is a sequence of i.i.d. random variables in $\mathbb{R}^m$. We obtain simple sufficient conditions for $(Z_n)_{n\in\mathbb{Z}}$ to be geometricaly absolutely regular. To obtain this conditions, we first prove a continuity theorem for the image of a measure by a polynomial application. The results apply to the ARMA processes and to the bilinear processes.

<u>Mots-Clés</u> : Processus autoregressifs ; processus stationnaires markoviens ; processus bilinéaires ; processus ARMA ; ergodicité géométrique ; régularité absolue géométrique.

Classification AMS 1985 : 60 G 10 ; 60 G 99 ; 60 J 27 ; 62 M 10 .

I - INTRODUCTION

Soit $(X_n)_{n\in\mathbb{Z}}$ un processus stationnaire en loi, à valeurs dans $\mathbb{R}^d$. Notons $\mathfrak{a}_k$ la tribu engendrée par $\{X_k, X_{k-1}, X_{k-2}, \ldots\}$ et $\mathfrak{a}^k$ celle engendrée par $\{X_k, X_{k+1}, X_{k+2}, \ldots\}$. Pour obtenir des résultats asymptotiques sur le processus $(X_n)_{n\in\mathbb{Z}}$ on est souvent amenés à supposer que les tribus $\mathfrak{a}_o$ et $\mathfrak{a}^k$ sont asymptotiquement indépendantes quand k devient grand.

Les hypothèses de mélange sont un des moyens de modéliser cette indépendance asymptotique. De nombreuses notions de mélange ont été introduites. Citons, sans être exhaustif, les mélanges définis à partir des coefficients suivants.

Le coefficient de mélange fort (M. Rosenblatt [31]) :

$$\alpha_k = \sup\{|P(A\cap B) - P(A).P(B)|,\ A\in\mathfrak{a}_o\ ,\ B\in\mathfrak{a}^k\}.$$

Le coefficient de mélange uniforme (I.A. Ibragimov et Y.V. Linnik [13]) :

$$\phi_k = \sup\{|P(B|A) - P(B)|\ ,\ A\in\mathfrak{a}_o\ ,\ B\in\mathfrak{a}^k\}.$$

Le coefficient de régularité absolue (appelé par Y.A. Davydov [5] coefficient de régularité complète) :

$$\beta_k = E\ \sup\{|P(B|\mathfrak{a}_o) - P(B)|\ ,\ B\in\mathfrak{a}^k\}.$$

Le coefficient de ρ-mélange [13] :

$\rho_k = \sup|E(\eta.\nu)|$, le sup étant pris sur les variables aléatoires η et ν vérifiant : η est $\mathfrak{a}_o$-mesurable, ν est $\mathfrak{a}^k$-mesurable,

$$E\eta^2 = E\nu^2 = 1 \qquad \text{et} \qquad E\eta = E\nu = 0\ .$$

Le coefficient de ψ-mélange (J.R. Blum, D.L. Hanson et

L.H. Koopmans [3])

$$\psi_k = \sup\{ \left| \frac{P(A \cap B)}{P(A).P(B)} - 1 \right| , A \in \mathfrak{a}_o , B \in \mathfrak{a}^k \} .$$

Le processus $(X_n)_{n\in\mathbb{Z}}$ sera dit fortement mélangeant (resp. uniformément mélangeant, absolument régulier, ψ-mélangeant, ρ-mélangeant) si la suite α_k (resp. ϕ_k , β_k , ψ_k , ρ_k) converge vers zéro.

Notons qu'à partir des définitions, on peut obtenir les inégalités :

$$\alpha_k \le \beta_k \le \phi_k \le \psi_k \ ; \ \rho_k \le 2\phi_k^{1/2} \ ;$$

de sorte que, par exemple, le mélange uniforme ou la régularité absolue implique le mélange fort.

Sous des hypothèses sur la vitesse de convergence des coefficients de mélange, de nombreux résultats asymptotiques sur le processus $(X_n)_{n\in\mathbb{Z}}$ ont été obtenus..On peut consulter, à titre d'exemples [10] , [13] , [33] , [34] et [40].

Cependant, il n'est pas toujours simple de donner des conditions suffisantes pour qu'un processus soit mélangeant, ni d'avoir la vitesse de convergence du coefficient de mélange. Une première approche de ce problème concerne les processus linéaires. I.A. Ibragimov et Y. Rozanov [14] font une étude systématique du mélange des processus linéaires gaussiens. Pour le cas linéaire non gaussien et unidimensionnel, V.V. Gorodetski [9] et C.S. Withers [39] établissent des conditions suffisantes de mélange fort. T.D. Pham et L.T. Tran [26] montrent que ces conditions impliquent en fait la régularité absolue de $(X_n)_{n\in\mathbb{Z}}$ et celà sans que X_n soit unidimensionnel.

La deuxième approche est une approche markovienne. On suppose que $(X_n)_{n\in\mathbb{Z}}$ est un processus stationnaire markovien. Un tel processus est construit à partir d'une chaîne de Markov homogène $(X'_n)_{n\geq 0}$ qui possède une loi de probabilité invariante π. La loi de X_n est alors π et la loi de $(X_{n_1},\dots,X_{n_k})$ est donnée par la propriété de Markov (voir M. Rosenblatt [30]). La chaîne $(X')_{n\geq 0}$ est dite fortement mélangeante (resp. uniformément mélangeante, absolument régulière, ψ-mélangeante, ρ-mélangeante) si le processus $(X_n)_{n\in\mathbb{Z}}$ est fortement mélangeant (resp. uniformément mélangeant, absolument régulier, ψ-mélangeant, ρ-mélangeant). Des conditions sur la chaîne $(X'_n)_{n\geq 0}$ donnent alors des propriétés de mélange de $(X_n)_{n\in\mathbb{Z}}$. Le premier résultat de ce type est le suivant (J.L. Doob [8]) : si une chaîne de Markov est Doeblin-récurrente alors elle est uniformément mélangeante et son coefficient de mélange uniforme converge vers zéro géométriquement. Dans [5] , Y.A. Davydov fait le lien entre la récurrence au sens de Harris et la régularité absolue. Le mélange fort et d'autres types de mélange sont étudiés par M. Rosenblatt [30] .

Dans le présent article, nous abordons le problème de la manière suivante.

On suppose que le processus stationnaire en loi $(X_n)_{n\in\mathbb{Z}}$ admet une représentation markovienne polynomiale. Nous entendons par là que :

(i) il existe une suite de variables aléatoires i.i.d. $(e_n)_{n\in\mathbb{Z}}$ à valeurs dans $\mathbb{R}^m$ et un processus stationnaire en loi $(Z_n)_{n\in\mathbb{Z}}$ à valeurs dans $\mathbb{R}^p$ tels que e_n et $\{Z_{n-1},Z_{n-2},\dots\}$ soient indépendants ;

(ii) il existe un entier $k>0$, une application polynomiale ϕ de $\mathbb{R}^p\times\mathbb{R}^m$ dans $\mathbb{R}^p$ et une application mesurable ψ de $\mathbb{R}^{p(2k+1)}$ dans

$\mathbb{R}^d$ tels que :

(1.1.) $$Z_n = \phi(Z_{n-1}, e_n)$$

(1.2.) $$X_n = \psi(Z_{n-k}, Z_{n-k+1}, \ldots, Z_{n+k})$$

Il est clair que $(Z_n)_{n\in\mathbb{Z}}$ est un processus stationnaire markovien et sa probabilité de transition $P(z,.)$ est la loi de probabilité de $\phi(z,e_n)$. Comme exemples de processus admettant une représentation markovienne, citons les processus ARMA (H. Akaike [1]) et les processus bilinéaires (T.D. Pham [27]). Observons que si le coefficient de mélange de $(Z_n)_{n\in\mathbb{Z}}$ converge vers zéro, le coefficient de mélange correspondant pour $(X_n)_{n\in\mathbb{Z}}$ converge aussi vers zéro et celà avec la même vitesse ; de sorte que l'étude du mélange de $(X_n)_{n\in\mathbb{Z}}$ se ramène à celle du mélange de $(Z_n)_{n\in\mathbb{Z}}$.

Ce travail est donc consacré à l'étude des processus stationnaires en loi et satisfaisant une équation du type (1.1).

Nous montrons dans la section 4 que, sous des conditions simples sur la loi de probabilité de e_n et sur l'application ϕ , le processus $(Z_n)_{n\in\mathbb{Z}}$ est absolument régulier avec un coefficient de régularité absolue qui tend vers zéro géométriquement.

Pour les processus bilinéaires, on obtient par exemple, le résultat suivant :

Soit $(X_n)_{n\in\mathbb{Z}}$ un processus bilinéaire d'ordre p, q, P, Q et d'équation :

(1.3.) $$X_n = \sum_{j=1}^{p} a_j X_{n-j} + e_n + \sum_{j=1}^{q} b_j e_{n-j} + \sum_{k=1}^{Q} \sum_{j=0}^{P} b_{k_j} X_{n-j-k} e_{n-k} .$$

Soit :

(1.4.) $$Z_n = AZ_{n-1} + BZ_{n-1}\, e_n + Ce_n + De_n^2 + E$$

sa représentation markovienne [27] .

Supposons que :

a) e_n admet une densité de probabilité f telle que l'ensemble $\{f>0\}$ soit un ouvert contenant zéro.

b) $\exists\, s>0$ tel que :

$$E|e_n|^{2s} < \infty \quad \text{et} \quad E\,\|A + Be_n\|^{s} < 1 \; .$$

Alors le processus bilinéaire $(X_n)_{n\in\mathbb{Z}}$ d'équation (1.3) est absolument régulier et son coefficient de régularité absolue tend vers zéro géométriquement.

Notre plan de travail est le suivant.

Dans la section 2, nous établissons un critère de mélange géométrique pour une chaîne de Markov dont l'espace des états est un sous-espace d'un espace topologique séparable et dénombrable à l'infini. Ce critère portera en particulier sur les probabilités de transition $P^k(z,.)$; mais ces probabilités pour le processus $(Z_n)_{n\in\mathbb{Z}}$ défini par l'équation (1.1), apparaissent comme des probabilités images de la loi de $(e_1,\ldots,e_k)$ par des applications polynomiales.

Cela nous conduit à étudier dans la section 3, l'image d'une mesure par une application polynomiale. Nous montrons en particulier le résultat suivant.

Soit $\psi(z,x)$ une application de classe C^1 de $M \times V$ dans W où M et W sont des variétés algébriques (A5) contenues respectivement dans $\mathbb{R}^{n_1}$ et $\mathbb{R}^{n_2}$, V est une variété algébrique lisse (A13) contenue dans $\mathbb{R}^{n_3}$; on suppose que ψ est polynomiale en x .

Soit θ une mesure positive régulière sur V absolument continue par rapport à une mesure lebesguienne sur V (voir J. Dieudonné [7]) et θ_z son image dans W par l'application $\psi(z,.)$.

Alors, si z_o est tel que l'image de V par $\psi(z_o,.)$ contienne un ouvert non vide dans la partie régulière de W , on a :

(1.5) $$\liminf_{z \to z_o} \theta_z(A) \geq \theta_{z_o}(A) \quad \text{pour tout borélien } A \text{ de } W ;$$

de plus, si θ est une mesure finie, alors :

(1.5') $$\lim_{z \to z_o} \theta_z(A) = \theta_{z_o}(A) \quad \text{pour tout borélien } A \text{ de } W .$$

Ce résultat est une extension du théorème de continuité de la translation où ψ est de la forme $\psi(z,x) = z+x$.

Les outils utilisés dans la démonstration sont le théorème de Sard (J. Dieudonné [7]), le théorème de la fibre localement triviale (A26) et le théorème de résolution des singularités (A27).

Indépendamment du présent article, ce résultat de continuité de θ_z possède un intérêt en soi.

Dans la section 4, nous déterminerons, sous certaines conditions, l'espace des états du processus $(Z_n)_{n\in\mathbb{Z}}$ défini par l'équation (1.1) ; celà nous permet alors d'établir des conditions suffisantes simples pour que ce processus soit géométriquement absolument régulier. En plus des résultats de la section 3, l'outil utilisé dans la section 4 est le théorème ergodique.

La section 5 est consacrée aux processus ARMA vectoriels et aux processus bilinéaires ; la condition que nous obtenons pour les processus ARMA est nettement moins restrictive que celle que nous avons obtenue dans [18] et [19] ; on notera en particulier que la loi de probabilité du bruit e_n qui intervient dans l'équation (1.1) ne sera pas nécessairement absolument continue dans $\mathbb{R}^m$ puisque notre hypothèse sera que e_n est à valeurs dans une variété algébrique qui peut être distincte de $\mathbb{R}^m$ et donc de mesure nulle pour la mesure de Lebesgue sur $\mathbb{R}^m$.

Enfin pour la commodité du lecteur, nous avons regroupé dans un appendice, les notions et résultats de géométrie algébrique réelle que nous utilisons ; les références notées (Ai) renvoient donc à cet appendice.

II - UN CRITERE DE MELANGE GEOMETRIQUE

Pour les différentes notions sur les chaînes de Markov que nous utilisons dans cette partie, on pourra se reporter à [24] , [25] ,[29] et [35] .

Soit $(Z_n)_{n\geq 0}$, une chaîne de Markov à valeurs dans un sous-espace S d'un espace topologique M séparable et dénombrable à l'infini. M est muni d'une mesure positive μ , σ-finie, non triviale, finie sur tout compact et telle que :

(2.1) $\qquad \mu(S) > 0$.

Le théorème qui suit donne un critère d'ergodicité géométrique (voir E. Nummelin et P. Tuominen [24] pour la définition) et de régularité absolue géométrique de la chaîne de Markov :

$$(Z_n)_{n\geq 0} .$$

Quand $(Z_n)_{n\geq 0}$ admet une probabilité stationnaire, on notera encore $(Z_n)_{n\in\mathbb{Z}}$, le processus stationnaire markovien construit avec cette probabilité stationnaire et la probabilité de transition (voir M. Rosenblatt [30]).

Théorème 2.1

Soit $P(z,.)$ *la probabilité de transition de* $(Z_n)_{n\geq 0}$.

Si les hypothèses suivantes sont satisfaites :

(H1) *Pour tout compact* K *de* M *et tout borélien* A *dans* S *de* μ*-mesure nulle, il existe un entier* n_o *tel que :*

$$P^{n_o}(z,A) = 0 \qquad \textit{pour tout } z \textit{ dans } K\cap S .$$

(H2) *Pour tout compact* K *de* M *et tout borélien* A *dans* S *de* μ*-mesure non nulle, il existe un entier* n_1 *tel que :*

$$\operatorname*{Inf}_{z\in K\cap S} P^{n_1}(z,A) > 0 .$$

(H3) *Il existe une fonction* g *positive sur* S *, un compact* K *de* M *et des nombres réels* $A>0$, $B>0$, $0<\rho<1$ *tels que :*

(i) $\mu(K\cap S) > 0$

(ii) $E(g(Z_{n+1})\,|\,Z_n = z) \leq g(z) - 1$ si $z\notin K\cap S$

(iii) $E(g(Z_{n+1})\,|\,Z_n = z) \leq \rho g(z)$ si $z\notin K\cap S$

(iv) $B \leq E(g(Z_{n+1}) | Z_n = z) \leq A$ si $z \in K \cap S$

alors la chaîne $(Z_n)_{n \geq 0}$ *est géométriquement ergodique et géométriquement absolument régulière, de plus le processus stationnaire markovien* $(Z_n)_{n \in \mathbb{Z}}$ *vérifie :*

$$E(g(Z_n)) < \infty \quad .$$

<u>Démonstration</u>

Notons μ_S la restriction de μ à S .

a - La chaîne $(Z_n)_{n\geq 0}$ est μ_S-irréductible grâce à l'hypothèse (H2) ; elle est aussi apériodique. En effet, supposons qu'elle ait une période $d > 1$; il existe alors des sous-ensembles non vide de S , C_1 , $C_2, \ldots, C_d$ tels que (S. OREY [25]) :

(2.2) $\qquad P(z, C_{i+1}) = 1$ si $z \in C_i$ et $1 \leq i \leq d-1$

(2.3) $\qquad P(z, C_1) = 1$ si $z \in C_d$

(2.4) $\qquad \mu_S(S - \cup C_i) = 0$

Un des ensembles C_i est de μ_S-mesure non nulle ; par exemple C_1 .

Prenons $a \in C_1$, $b \in C_2$ et $K = \{a,b\}$.

D'après l'hypothèse (H2) , il existe un entier n_1 tel que :

(2.5) $\qquad P^{n_1}(a, C_1) > 0$ et $P^{n_1}(b, C_1) > 0$.

Mais, puisque la chaîne est de période d , on a alors :

(2.6) $n_1 = m_1 d$ et $n_1 = m_1' d + d - 1$ avec $d > 1$;

ce qui est impossible car n_1 ne peut pas avoir deux restes différents dans sa division euclidienne par d .

b) - $(Z_n)_{n\geq 0}$ étant μ_S-irréductible, il existe une mesure sur S excessive (subinvariant en anglais), σ-finie, non triviale, qu'on notera π et qui domine μ_S (N. Jain et B. Jamison [15]).

Soit K un compact de M ; on a alors :

(2.7) $$\mu_S(K) > 0 \implies 0 < \pi(K \cap S) < \infty .$$

En effet, $\pi(K \cap S) > 0$ est évident puisque π domine μ_S ; d'autre part π et μ_S étant σ-finies, on peut trouver un borélien A de S tel que :

(2.8) $$\mu_S(A) > 0 \quad \text{et} \quad \pi(A) < \infty .$$

On applique l'hypothèse (H2) à K et A :

(2.9) $$\exists\, n_1 \in \mathbb{N} , \text{ tel que } b = \inf_{K\cap S} P^{n_1}(z,A) > 0 ;$$

on a alors :

(2.10) $$\begin{aligned} \pi(A) &\geq \int \pi(dz)\ P^{n_1}(z,A) \\ &\geq \pi\,(K \cap S).b \end{aligned}$$

et donc

(2.11) $$\pi(K \cap S) \leq b^{-1}\ \pi(A) < \infty .$$

L'implication (2.7) nous permet d'utiliser la classe $\mathbb{K}$ des ensembles de la forme $K \cap S$, avec K compact de M et $\mu(K \cap S) > 0$, dans le critère d'ergodicité de R.L. Tweedie [35]. L'hypothèse (H3) implique alors que la chaîne $(Z_n)_{n \in \mathbb{Z}}$ est positive-récurrente.

Notons encore π sa probabilité invariante.

c) - $\pi \sim \mu_S$

Il suffit de montrer que μ_S domine π.

Soit $\varepsilon > 0$ et A tel que $\mu_S(A) = 0$.

Prenons un compact K de M tel que :

(2.12) $\pi(S-K) < \varepsilon$;

(il en existe puisque M est dénombrable à l'infini et π est une probabilité).

D'après l'hypothèse (H1), il existe un entier n_o, tel que :

(2.13) $P^{n_o}(z,A) = 0$ pour tout $z \in K \cap S$;

mais :

$$\pi(A) = \int_{S-K} \pi(dz)\, P^{n_o}(z,A) + \int_{K \cap S} \pi(dz)\, P^{n_o}(z,A) \tag{2.14}$$

et donc :

(2.15) $\pi(A) \leq \pi(S-K) < \varepsilon$

ε étant arbitraire, on conclut que $\pi(A) = 0$.

d) - La chaîne $(Z_n)_{n \geq 0}$ est Harris-récurrente et les ensembles de la classe $\mathcal{K}$ sont des petits ensembles au sens défini par E. Nummelin et P. Tuominen [24] . En effet, d'après le lemme 2.2 de R.L. Tweedie [36] : si $\mu_S(A) > 0$, il existe un borélien $N(A)$ de S de μ_S-mesure nulle tel que :

$$(2.16) \qquad F(z,A) = 1 \quad \text{pour} \quad z \in S - N(A) \ ;$$

où

$$(2.17) \qquad F(z,A) = P(z\ , \bigcup_{n=1}^{\infty} (Z_n \in A)) \ .$$

Il nous faut montrer que $F(z,A) = 1$ pour $z \in N(A)$; soit donc $z \in N(A)$; pour le compact $\{z\}$, il existe un entier n_o tel que :

$$(2.18) \qquad P^{n_o}(z,N(A)) = 0 \ .$$

On a :

$$(2.19) \qquad F(z,A) \geq P^{n_o}(z,A) + \int_{S-A} P^{n_o}(z,dy)\ F(y,A) \ ;$$

la mesure $P^{n_o}(z,dy)$ est nulle sur $N(A)$ et pour $y \notin N(A)$ on a $F(y,A) = 1$; par conséquent :

$$(2.20) \qquad F(z,A) \geq P^{n_o}(z,A) + \int_{S-A} P^{n_o}(z,dy) = 1 \ .$$

La chaîne est donc Harris récurrente.

D'autre part comme $\pi \sim \mu$, l'hypothèse (H2) entraîne que les ensembles de la classe $\mathcal{K}$ sont des petits ensembles.

e) - La chaîne étant μ_S-irréductible, apériodique et Harris-récurrente, on peut appliquer le critère d'ergodicité géométrique de E. Nummelin et P. Tuominen [24] ; l'hypothèse (H3) permet alors de conclure que la chaîne $(Z_n)_{n\geq 0}$ est géométriquement ergodique.

On utilise maintenant les résultats de R.L. Tweedie [36] pour conclure que pour le processus $(Z_n)_{n\in\mathbb{Z}}$,

(2.21) $\qquad E(g(Z_n) < \infty$.

Enfin, E. Nummelin et P. Tuominen montrent également qu'il existe $\xi < 1$ tel que :

(2.22) $$\int \pi(dz)\ \|P^n(z,.) - \pi\| = O(\xi^n)$$

où $\|.\|$ désigne la norme en variation ; mais d'après Y.A. Davydov [5] , le coefficient de régularité absolue de $(Z_n)_{n\in\mathbb{Z}}$ est :

(2.23) $$\beta(n) = \int \nu(dz)\ \|P^n(z,.) - \nu\|$$

où ν et la loi de probabilité de Z_n ; dans notre cas $\nu = \pi$ et on conclut donc que $(Z_n)_{n\in\mathbb{Z}}$ est géométriquement absolument régulier.

Le théorème est démontré $\square$

Dans les hypothèses (H1) et (H2) du théorème 1, les probabilités $P^n(z,.)$ interviennent ; on constate par itération que pour le processus $(Z_n)_{n\in\mathbb{Z}}$ vérifiant (1) , les lois $P^n(z,.)$ sont les images de la loi de probabilité de $(e_1,\dots,e_n)$ par des applications polynomiales paramétrées par z . L'étude de ces lois images fait l'objet de la section qui suit.

III - IMAGE D'UNE MESURE PAR UNE APPLICATION POLYNOMIALE

En général, l'image de $\mathbb{R}^m$ dans $\mathbb{R}^p$ par une application polynomiale est un ensemble semi-algébrique D de dimension plus petite que p (théorème de Tarski-Seidenberg (A22)) ; par conséquent la mesure de D dans $\mathbb{R}^p$ est souvent nulle et donc on ne peut pas toujours utiliser la mesure de Lebesgue dans $\mathbb{R}^p$, comme mesure de référence.

On est ainsi amené à munir D d'une mesure lebesguienne (J. Dieudonné [7]).

La forme la plus générale du problème qui nous intéresse serait donc la suivante : étant donnée une application polynomiale ϕ d'un ensemble semi-algébrique D_1 dans un autre D_2, quelles sont les propriétés de l'image par ϕ d'une mesure θ sur D_1 absolument continue par rapport à une mesure lebesguienne ?

Cependant, si V et W désignent les fermetures pour la topologie de Zariski (A3) de D_1 et D_2, on a :

$$\dim V = \dim D_1 \quad \text{et} \quad \dim W = \dim D_2 ,$$

de plus ϕ est aussi définie sur V et l'image de V par continuité est contenue dans W ; on ne perd donc rien à la généralité en prenant des ensembles algébriques à la place d'ensembles semi-algébriques.

D'autre part un ensemble algébrique est réunion finie de variétés algébriques ; l'image de θ par ϕ peut donc être étudiée en regardant l'image de θ par la restriction de ϕ à chaque variété. Pour ne pas alourdir le texte de la présente section, nous supposerons donc que θ est une mesure définie sur une variété algébrique.

Enfin, les variétés algébriques ont l'avantage sur les ensembles semi-algébriques de posséder la propriété de stationnarité des suites croissantes (A10) et nous serons amenés justement à étudier dans la section 4 une suite d'applications polynomiales dont les images constituent une suite croissante d'ensembles semi-algébriques contenus dans $\mathbb{R}^p$; une telle suite n'a pas de propriété de stationnarité ; par contre si ces ensembles semi algébriques sont irréductibles, la suite de leurs fermetures pour la topologie de Zariski sera stationnaire, c'est-à-dire constante, à partir d'un certain rang.

Nos hypothèses de travail dans cette section seront les suivantes

(3.1) M, V, W sont des variétés algébriques contenues respectivement dans $\mathbb{R}^{n'}$, $\mathbb{R}^{m'}$ et $\mathbb{R}^{p'}$ et de dimensions respectives n, m et p.

(3.2) V est lisse (A13)

(3.3) $\phi(z,x)$ est une application de classe C^1 de $\mathbb{R}^{n'} \times \mathbb{R}^{m'}$ dans $\mathbb{R}^{p'}$, polynomiale en x et vérifiant :

$$\phi(M \times V) \subset W .$$

(3.4) Chaque variété algébrique X est munie d'une mesure régulière μ_X obtenue en munissant le lieu régulier R_X de X (A 12) d'une mesure lebesguienne convenable [7] ; cette mesure est prolongée par zéro au lieu singulier Σ_X de X. Une telle mesure est construite dans [17] .

(3.5) θ est une mesure positive régulière sur V, absolument continue par rapport à μ_V et sa densité a pour ensemble de positivité E. L'ensemble de positivité d'une fonction f mesurable désignant l'ensemble $\{f > 0\}$.

(3.6) Pour $z \in M$, la mesure image dans W de θ par l'application

$\phi_z = \phi(z,.)$ est notée θ_z . Noter que θ_z peut prendre des valeurs infinies).

Notre premier résultat est un théorème d'absolue continuité de θ_z pour z fixé ; pour simplifier l'écriture nous omettrons provisoirement l'indice z ; on suppose donc que ϕ est une application de V dans W et l'image de θ est notée θ'.

Théorème 3.1

Si ϕ *est une application dominante (A 23) de* V *dans* W *, alors* θ' *est absolument continue par rapport à* μ_W *et sa densité a pour ensemble de positivité* $\phi(E)$.

Démonstration

Notons que $m \geq \rho$ car ϕ est dominante (A 23)

Soit Σ_W le lieu singulier de W ; le lieu régulier $R_W = W - \Sigma_W$ est une variété analytique (A 12). On se ramène à R_W de la manière suivante. Tout d'abord par construction de μ_W on a :

(3.7) $\qquad \mu_W(\Sigma_W) = 0$;

d'autre part $\phi^{-1}(\Sigma_W)$ est un fermé de Zariski dans V et :

(3.8) $\qquad \dim \phi^{-1}(\Sigma_W) < \dim V$,

car sinon on aurait $V = \phi^{-1}(\Sigma_W)$ (A 9) et ϕ ne serait pas dominante ; on a par conséquent :

(3.9) $\qquad \mu_V(\phi^{-1}(\Sigma_W)) = 0$

et donc :

(3.10) $$\theta(\phi^{-1}(\Sigma_W)) = \theta'(\Sigma_W) = 0 .$$

Au vu de (3.7), (3.9) et (3.10) on peut restreindre ϕ à $V_1 = V - \phi^{-1}(\Sigma_W)$; V_1 est un ouvert de Zariski et donc en particulier une sous-variété analytique ouverte de V .

Soit :

(3.11) $$G = \{x \in V_1 \ , \ \mathrm{rg}\ d\phi_x < p\} ,$$

G est un fermé de Zariski dans V_1 , de dimension strictement inférieure à m car ϕ est dominante ; on a donc

(3.12) $$\mu_V(G) = 0 ;$$

d'autre part, le théorème de Sard (J. Dieudonné [7]) donne :

(3.13) $$\mu_W(\phi(G)) = 0$$

et donc puisque ϕ est dominante :

(3.14) $$\mu_W(\phi(V) - \phi(V_1 - G)) = 0 .$$

Soit maintenant $x \in V_1 - G$; c'est un point régulier de ϕ ; par conséquent il existe des voisinages ouverts U et U' respectivement de x et $\phi(x)$, un ouvert U'' de $\mathbb{R}^{m-p}$ et un difféomorphisme ψ de U sur $U' \times U''$ tels que le diagramme suivant soit commutatif :

(3.15) $$\begin{tikzcd} U \arrow[rr, "\phi"] \arrow[dr, "\psi"'] & & U' \\ & U' \times U'' \arrow[ur, "\mathrm{proj.}"'] & \end{tikzcd}$$

où proj. désigne la projection (M.Q. Pham [28]).

En procédant ainsi pour chaque x dans $V_1 - G$, on obtient un recouvrement ouvert de $V_1 - G$; comme $V_1 - G$ est une variété analytique, par paracompacité on peut extraire un sous recouvrement dénombrable (F. Warner [37]) que nous noterons $(U_i)_{i \in \mathbb{N}}$, de sorte que chaque U_i vérifie (3.15). Nous noterons ϕ_i la restriction de ϕ à U_i.

Pour établir notre résultat, il suffit de montrer que si A est un borélien de W alors :

(3.16) $$\theta'(A) = 0 \iff \mu_W(A \cap \phi(E)) = 0 .$$

On a :

(3.17) $$\begin{aligned} \theta'(A) = 0 &\iff \theta(\phi^{-1}(A)) = 0 \\ &\iff \theta(\phi^{-1}(A) \cap U_i) = 0 \qquad \forall i \in \mathbb{N} \\ &\iff \mu_V(\phi^{-1}(A) \cap U_i \cap E) = 0 \qquad \forall i \in \mathbb{N} \end{aligned}$$

(la deuxième équivalence provient de (3.9) et (3.12), la troisième des hypothèses sur θ).

Mais :

(3.18) $$\phi^{-1}(A) \cap U_i \cap E = \phi_i^{-1}(A \cap \phi(E_i)) \quad \text{où} \quad E_i = E \cap U_i ,$$

et donc (3.17) devient :

(3.19) $$\begin{aligned} \theta'(A) = 0 &\iff \mu_V(\phi_i^{-1}(A \cap \phi(E_i))) = 0 \qquad \forall i \in \mathbb{N} \\ &\iff \mu_W \otimes \mu(\mathrm{Proj}^{-1}(A \cap \phi(E_i))) = 0 \qquad \forall i \in \mathbb{N} \end{aligned}$$

où μ est la mesure de Lebesgue sur $\mathbb{R}^{m-p}$; on obtient alors :

$$\theta'(A) = 0 \iff \mu_W(A \cap \phi(E_i)) = 0 \qquad \forall i \in \mathbb{N} \tag{3.20}$$

$$\iff \mu_W(A \cap \phi(E)) = 0$$

car d'après (3.14), $\mu_W(\phi(E)) = \mu_W\{\bigcup_i \phi(E_i)\}$.

Le théorème est ainsi démontré.

Remarque 3.1

L'hypothèse ϕ polynomiale sert à montrer (3.9) et (3.12) ; le théorème est donc encore vrai dans le cas où V et W sont des variétés de classe C^∞ , ϕ est de classe C^∞ et l'ensemble des points critiques de ϕ est de μ_V-mesure nulle.

Nous revenons maintenant au cadre que nous avons défini au début de cette section. Pour continuer nous avons besoin du lemme suivant.

Lemme 3.1

On suppose que $m = p$. *Soit* z_o *un point de* M *et* (z_n) *une suite de points de* M *convergeant vers* z_o *; alors il existe une sous-suite* (z'_n) *possédant la propriété suivante*

(3.21) *pour tout point régulier* x_o *de* ϕ_{z_o} *et tout voisinage* U *de* x_o *dans* V *, il existe des voisinages ouverts* $U_o \subset U$ *de* x_o , U'_o *de* $y_o = \phi_{z_o}(x_o)$ *dans* W *et un entier* N *tels que pour tout* $n \geq N$ $\phi_{z'_n}$ *soit un difféomorphisme de* U_o *sur un ouvert de* W *contenant* U'_o .

Démonstration

La difficulté provient du fait que z_o peut être un point singulier de M . Pour résoudre cette difficulté nous utilisons le théorème de résolution des singularités de Hironaka (A 27).

Il existe donc une variété analytique réelle $\hat{M}$ de même dimension que M et une application analytique propre et surjective ℓ de $\hat{M}$ sur M .

Notons $B = \ell^{-1}(z_o)$ la fibre au dessus de z_o ; B est compacte puisque ℓ est propre.

Définissons l'application $\psi(\hat{z},x)$ de $\hat{M}\times V$ dans $\hat{M}\times W$ par :

(3.22) $\qquad \psi(\hat{z},x) = (\hat{z},\phi(\ell(\hat{z}),x))$;

soit $\hat{z}_o$ un point de B , comme $\hat{M}$ est lisse et y_o est un point régulier de W par hypothèse, il s'ensuit que $(\hat{z}_o,y_o)$ est un point régulier de $\hat{M}\times W$ (A 14) ; on peut alors vérifier facilement, par le critère du jacobien que $(\hat{z}_o,x_o)$ est un point régulier de ψ ; en effet ce jacobien est de la forme suivante :

$$J = \left|\begin{array}{cccc:c} 1 & 0 & \cdots & 0 & \\ 0 & 1 & & \vdots & D_{\hat{z}}\,\phi(\ell(.),.))_{(\hat{z}_o,x_o)} \\ \vdots & & \ddots & 0 & \\ 0 & \cdots & 0 & 1 & \\ \hdashline & & 0 & & (D_x\,\phi_{z_o})_{x_o} \end{array}\right|$$

comme par hypothèse $(D_x\,\phi_{z_o})_{x_o} \neq 0$, on voit donc que $J \neq 0$. Il existe donc un voisinage ouvert $U_o \subset U$ de x_o , un voisinage ouvert $\hat{M}_o$ de z_o dans $\hat{M}$ et un voisinage ouvert L_o de $(\hat{z}_o, y_o)$ dans $\hat{M} \times W$ tels que ψ soit un difféomorphisme de $\hat{M}_o \times U_o$ sur L_o . L_o contient un ouvert de la forme $\hat{M}'_o \times U'_o$ où $\hat{M}'_o$ et U'_o sont des ouverts contenant respectivement $\hat{z}_o$ et y_o .

Soit $\hat{z} \in \hat{M}_o$; il est clair que l'ensemble

(3.23) $$U'_{\hat{z}} = \{y \in W \ ; \ (\hat{z}, y) \in L_o\}$$

est un ouvert et que $\phi_{\ell(\hat{z})}$ est un difféomorphisme de U_o sur $U'_{\hat{z}}$; d'autre part si $\hat{z} \in \hat{M}'_o$, alors $U'_{\hat{z}}$ contient U'_o . Nous avons donc montré que :

(3.24) il existe des voisinages ouverts $U_o \subset U$, $\hat{M}'_o$ et U'_o respectivement de x_o dans V , de $\hat{z}_o$ dans $\hat{M}$ et de y_o dans W tels que pour tout $\hat{z} \in \hat{M}'_o$, $\phi_{\ell(\hat{z})}$ soit un difféomorphisme de U_o sur un ouvert de W contenant U'_o .

Soit maintenant une suite z_n convergeant vers z_o dans M ; on peut supposer que la suite z_n est dans un compact K . Choisissons maintenant une suite $\hat{z}_n$ dans $\hat{M}$ telle que :

(3.25) $$\ell(\hat{z}_n) = z_n ,$$

comme $\ell^{-1}(K)$ est compact (ℓ est propre), on peut extraire une sous-suite z'_n telle que $\hat{z}'_n$ soit convergente ; soit $\hat{z}_o$ sa limite. $\hat{z}_o$ est dans B par continuité ; on lui applique le résultat précédent (3.24) et on garde les mêmes notations. A partir d'un rang N , $\hat{z}'_n \in \hat{M}'_o$ et donc

$\phi_{\ell(\hat{z}'_n)} = \phi_{z'_n}$ est un difféomorphisme de U_o sur un ouvert de W contenant

U'_o ; ce qui termine la démonstration.

Nous établissons maintenant la premier résultat de semi continuité.

Lemme 3.2

On suppose que $m = p$. *Soit* (z_n) *une suite de points de* M *de limite* z_o *; si les deux conditions suivantes sont satisfaites :*

(i) ϕ_{z_o} *est dominante,*

(ii) *la suite* (z_n) *possède la propriété* (3.21) *;*

alors pour tout borélien A *de* W :

$$\liminf \theta_{z_n}(A) \geq \theta_{z_o}(A) .$$

Démonstration

a) Soit $x_o \in V$, un point régulier de ϕ_{z_o} ; le point $y_o = \phi_{z_o}(x_o)$ est un point régulier de W et par continuité pour z voisin de z_o , $y = \phi_z(x_o)$ est aussi un point régulier de W . D'autre part, toujours par continuité, pour z voisin de z_o on a :

$$rg(D_x \phi_z)_{x_o} = p$$

et donc, dans un voisinage de z_o , ϕ_z est dominante et d'après le théorème 3.1 θ_z admet une densité que nous noterons f_z . Nous nous placerons constamment dans ce voisinage dans toute la suite de la démonstration.

b) Soit S la fermeture de Zariski de l'ensemble des valeurs singulières de ϕ_{z_o} ; d'après le théorème de Sard algébrique (A 24) :

(3.26) $\dim S < p$ et $\mu_W(S) = 0$.

On peut donc se limiter dans la démonstration aux boréliens contenus dans l'espace $W_1 = W-S$ qui est un ouvert de Zariski lisse dans W .

Maintenant, d'après le théorème de la fibre localement triviale (A 26), il existe un fermé de Zariski Z , de dimension strictement plus petite que p et tel que tout point y_o de W_1-Z possède un voisinage sur lequel les fibres de ϕ_{z_o} sont de cardinal constant fini et égal à : card $\phi_{z_o}^{-1}(y_o)$.

Comme $\mu_W(Z) = 0$, on se place donc dans $W_2 = W_1-Z$ qui est encore lisse. Soit maintenant :

(3.27) $H = \{y \in W_2 \text{ , card } \phi_{z_o}^{-1}(y) = 0\}$;

par définition de θ_{z_o} on a :

(3.28) $\theta_{z_o}(H) = 0$;

encore une fois on peut donc se limiter aux boréliens contenus dans W_2-H qui est un ouvert de W_2 par application du théorème de la fibre localement triviale (et par conséquent une variété analytique).

c) Soit $y \in W_2-H$ et $s = \text{card } \phi_{z_o}^{-1}(y)$; notons $x_1,\ldots,x_s$ les points de $\phi_{z_o}^{-1}(y)$. Ce sont des points réguliers de ϕ_{z_o} grâce à notre construction de W_2 . On peut donc trouver des ouverts disjoints $U_1,\ldots,U_s$ voisinages respectifs de $x_1,\ldots,x_s$ dans V de sorte que ϕ_{z_o} soit un difféomorphisme de U_j sur un voisinage U'_j de y et que :

(3.29) $$\phi_{z_o}^{-1}\left(\bigcap_{j=1}^{s} U'_j\right) \subset \bigcup_{j=1}^{s} U_j \ ;$$

on prendra de plus des U_j relativement compacts.

Maintenant en utilisant deux fois la propriété (3.21) on peut choisir les U_j de telle façon qu'il existe un voisinage ouvert U'_o de y et un entier N tels que :

(3.30) pour $n > N$, ϕ_{z_n} est un difféomorphisme de U_j sur un ouvert contenant U'_o

et ϕ_{z_n} est régulière sur $\bar{U}_j$ pour tout $j = 1,\ldots,s$;

(3.31) ϕ_{z_o} est un difféomorphisme de U_j sur un ouvert contenant U'_o

et ϕ_{z_o} est régulière sur $\bar{U}_j$ pour tout $j = 1,\ldots,s$;

(3.32) $$\phi_{z_o}^{-1}(U'_o) \subset \bigcup_{j=1}^{s} U_j \ .$$

d) Soit maintenant A un borélien contenu dans l'ouvert U'_o construit dans c) ; nous allons montrer que :

(3.33) $$\liminf \theta_{z_n}(A) \geq \theta_{z_o}(A) \ .$$

Notons $\phi_{z_n,j}$ la restriction de ϕ_{z_n} à U_j , $j = 1,\ldots,s$.

D'après la construction faite en c) on a :

(3.34) $$\phi_{z_o}^{-1}(A) = \bigcup_{j=1}^{s} \phi_{z_o,j}^{-1}(A)$$

et

(3.35) $$\phi_{z_n}^{-1}(A) \supset \bigcup_{j=1}^{s} \phi_{z_n,j}^{-1}(A) \quad ;$$

donc

(3.36) $$\theta_{z_o}(A) = \sum_{j=1}^{s} \theta(\phi_{z_o,j}^{-1}(A))$$

et

(3.37) $$\theta_{z_n}(A) \geq \sum_{j=1}^{s} \theta(\phi_{z_n,j}^{-1}(A)) \quad ;$$

ce qui donne :

(3.38) $$\theta_{z_o}(A) = \sum_{j=1}^{s} \int_A h(y)\ f(\phi_{z_o,j}^{-1}(y))\,|J^{-1}(\phi_{z_o,j})|\,d\mu_W$$

et

(3.39) $$\theta_{z_n}(A) \geq \sum_{j=1}^{s} \int_A h(y)\ f(\phi_{z_n,j}^{-1}(y))\,|J^{-1}(\phi_{z_n,j})|\,d\mu_W$$

où J est le jacobien, f la densité de θ et h une fonction positive qui ne dépend que des mesures lebesguiennes choisies sur V et W (plus exactement sur V et la partie régulière de W) .

Notons sur U'_o :

(3.40) $$g_{z_n}(y) = \sum_{j=1}^{s} h(y)\ f(\phi_{z_n,j}^{-1}(y))\,|J^{-1}(\phi_{z_n,j})|\cdot \ ;$$

comme en fait les relations (3.38) et (3.39) sont vraies pour tout borélien A contenu dans U'_o , on a :

(3.41) $$f_{z_o} = g_{z_o} \quad \text{et} \quad f_{z_n} \geq g_{z_n} \quad \text{p.p. sur } U'_o \;;$$

où f_{z_n} est la densité de θ_{z_n}. Pour montrer (3.33) il nous suffira donc de montrer que :

(3.42) $$\lim_n \int_A g_{z_n}(y)\, d\mu_W = \int_A g_{z_o}(y)\, d\mu_W \;;$$

pour celà soit $\varepsilon > 0$; l'espace des fonctions continues sur $\overline{\cup_j U_j}$ étant dense dans l'espace des fonctions absolument intégrables pour μ_V sur $\overline{\cup_j U_j}$, on peut trouver une fonction positive continue $\tilde{f}$ sur $\overline{\cup_j U_j}$ telle que :

(3.43) $$\int_{\overline{\cup_j U_j}} |\tilde{f}-f|\, d\mu_V < \varepsilon \,.$$

on a alors :

(3.44) $$\left|\int_A g_{z_n}(y)d\mu_W - \int_A g_{z_o}(y)d\mu_W\right| \leq \sum_{j=1}^{s} \int_A \left|f(\phi^{-1}_{z_{n,j}}(y)) - \tilde{f}(\phi^{-1}_{z_{n,j}}(y))\right| \left|J^{-1}\right| h(y)\,d\mu_W$$
$$+ \sum_{j=1}^{s} \int_A \left|\tilde{f}(\phi^{-1}_{z_{o,j}}(y)) - f(\phi^{-1}_{z_{o,j}}(y))\right| \left|J^{-1}\right| h(y)\, d\mu_W$$
$$+ \sum_{j=1}^{s} \left|\int_A \tilde{f}(\phi^{-1}_{z_{n,j}}(y)) \left|J^{-1}\right| h(y)d\mu_W - \int_A \tilde{f}(\phi^{-1}_{z_{o,j}}(y)) \left|J^{-1}\right| h(y)d\mu_W\right|$$

les deux premières sommes à droite de (3.44) se ramènent par changement de variable à des intégrales sur V et sont majorées chacune par ε grâce à (3.43) ; la troisième somme converge vers zéro par le théorème de convergence dominée utilisé avec la mesure $h(y)d\mu_W$. En effet, $\tilde{f}$ est bornée sur $\overline{\bigcup_{j=1}^{s} U_j}$ et $\tilde{f}(\phi^{-1}_{z_{n,j}}(y)) \leq \sup \tilde{f}$; d'autre part d'après (3.30) et (3.31) on voit que si z_n est dans un voisinage compact de z_o, $|J|$

est minoré par une constante strictement positive et par conséquent $|J^{-1}|$ est majorée ; ce qui permet de conclure la démonstration de (3.33).

e) Pour passer maintenant à un borélien quelconque A de W_2-H , on commence par recouvrir W_2-H par des ouverts U'_o construit en chaque point $y \in W_2-H$ comme celà a été fait en c). Comme W_2-H est une variété analytique, on peut extraire un sous-recouvrement dénombrable qu'on notera $(U'_{ok})_{k\in\mathbb{N}}$. Tout borélien A de W_2-H peut s'écrire comme réunion disjointe de boréliens

(3.45) $$A = \bigcup_{k\in\mathbb{N}} A_k$$

où $A_k \subset U'_{ok}$ pour tout $k \in \mathbb{N}$; la relation (3.33) valable pour chaque A_k permet alors de conclure que

(3.46) $$\liminf \theta_{z_n}(A) \geq \theta_{z_o}(A) \ ,$$

ce qui termine la démonstration du lemme 3.2.

Nous allons maintenant établir le théorème de semi continuité que nous avons annoncé. Nous restons toujours sous les hypothèses introduites en début de section. On a :

Théorème 3.2

Soit $z_o \in M$ *tel que* ϕ_{z_o} *soit dominante ; alors pour tout borélien* A *dans* W *on a :*

(3.47) $$\liminf_{z\to z_o} \theta_z(A) \geq \theta_{z_o}(A) \ ;$$

de plus si θ *est une mesure finie, alors :*

(3.48) $$\lim_{z \to z_o} \theta_z(A) = \theta_{z_o}(A) .$$

<u>Démonstration</u>

Pour $m = p$, (3.47) est une conséquence des lemmes 3.1 et 3.2 . En effet, soit :

(3.49) $$\gamma = \liminf_{z \to z_o} \theta_z(A) ;$$

il existe une suite (z_n) convergente vers z_o telle que :

(3.50) $$\lim \theta_{z_n}(A) = \gamma ;$$

on utilise le lemme 3.1 , pour extraire une sous-suite (z'_n) possédant la propriété (3.21) ; d'après le lemme 3.2 on a :

(3.51) $$\liminf \theta_{z'_n}(A) \geq \theta_{z_o}(A)$$

ce qui en vertu de (3.50) donne

(3.52) $$\liminf_{z \to z_o} \theta_z(A) = \gamma \geq \theta_{z_o}(A)$$

Supposons maintenant que $m > p$. Soit $F_1 , F_2 , \ldots , F_k$ des polynômes engendrant l'idéal de la variété V ; soit $\phi^1_{z_o} , \phi^2_{z_o} , \ldots , \phi^{p'}_{z_o}$ les composantes de l'application polynomiale ϕ_{z_o} . Si $x = (x_1, \ldots, x_{m'})$ est un point régulier de ϕ_{z_o} dans V , alors la matrice jacobienne :

$$J_{\phi_{z_o}}(x) = \begin{pmatrix} \frac{\partial F_1}{\partial x_1}(x) \ldots\ldots & \frac{\partial F_k}{\partial x_1}(x) & \frac{\partial \phi^1_{z_o}}{\partial x_1}(x) \ldots\ldots & \frac{\partial \phi^{p'}_{z_o}}{\partial x_1}(x) \\ \vdots & & & \\ \frac{\partial F_1}{\partial x_{m'}}(x) & \frac{\partial F_k}{\partial x_{m'}}(x) & \frac{\partial \phi^1_{z_o}}{\partial x_{m'}}(x) \ldots\ldots & \frac{\partial \phi^{p'}_{z_o}}{\partial x_{m'}}(x) \end{pmatrix}$$

est de rang $m'-m+p$ (A 21) ; elle a donc $m'-m+p$ lignes indépendantes. Soient $i_1, i_2, \ldots, i_{m-p}$ les indices des autres lignes ; posons :

(3.53) $$\psi(z,x) = (\phi(z,x), x_{i_1}, x_{i_2}, \ldots, x_{i_{m-p}})$$

et

(3.54) $$\psi_z(x) = (\phi_z(x), x_{i_1}, x_{i_2}, \ldots, x_{i_{m-p}})$$

ψ_{z_o} est une application polynomiale de V dans $W \times \mathbb{R}^{m-p}$ et $W \times \mathbb{R}^{m-p}$ est une variété algébrique de dimension m (A 14).

Comme $\phi_{z_o}(x)$ est un point régulier de W, il s'ensuit que $\psi_{z_o}(x) = (\phi_{z_o}(x), x_{i_1}, \ldots, x_{i_{m-p}})$ est un point régulier de $W \times \mathbb{R}^{m-p}$.
D'autre part la matrice jacobienne de ψ_{z_o} est de la forme :

$$J_{\psi_{z_o}}(x) = \begin{pmatrix} & \vdots & & 0 & & \\ J_{\phi_{z_o}} & \vdots & 1 & & & \\ & \vdots & & 1 & & \\ & \vdots & 0 & & 1 & \\ & \vdots & & & & 1 \end{pmatrix}$$

où les 1 apparaissent sur les lignes $i_1, i_2, \ldots, i_{m-p}$. Par conséquent :

(3.55) $$\operatorname{rg} \psi_{z_o}(x) = m' ,$$

ce qui signifie que x est un point régulier de ψ_{z_o} (A 14) et l'application ψ_{z_o} est par conséquent dominante. Si on note maintenant θ'_z l'image de θ par ψ_z , on a en vertu de (3.52) :

(3.56) $$\liminf \theta'_z(A') \geq \theta'_{z_o}(A')$$

pour tout borélien A' dans $W \times \mathbb{R}^{m-p}$.

Mais il est clair que θ_z est une marginale de θ'_z et en particulier pour tout borélien A dans W on a :

(3.57) $$\theta_z(A) = \theta'_z(A \times \mathbb{R}^{m-p})$$

ce qui donne par (3.56) :

$$\liminf \theta_z(A) \geq \theta_{z_o}(A) \ ;$$

ce qui achève la démonstration de (3.47).

Pour démontrer (3.48), posons $\theta(V) = c$; il est clair que $\theta_z(W) = c$ pour tout $z \in M$. Soit A un borélien de W et A' son complémentaire dans W . (3.47) donne :

(3.58) $$\liminf_{z \to z_o} \theta_z(A') \geq \theta_{z_o}(A')$$

d'où

(3.59) $$\liminf_{z \to z_o} (c - \theta_z(A)) \geq c - \theta_{z_o}(A)$$

et donc

$$\limsup_{z \to z_o} \theta_z(A) \le \theta_{z_o}(A) \tag{3.60}$$

ce qui combiné à (3.47) donne (3.48).

Le théorème 3.2 est donc démontré □

Remarque 3.2

Dans la pratique, on a une application $\phi(z,x)$ définie sur $\mathbb{R}^{n'} \times \mathbb{R}^{m'}$ et on s'intéresse à un voisinage de z_o. On regarde alors la fermeture de Zariski W de $\phi_{z_o}(\mathbb{R}^{m'})$ (ou $\phi_{z_o}(V)$, V variété dans $\mathbb{R}^{m'}$) ; si z_o possède dans une sous variété de $\mathbb{R}^{n'}$ un voisinage ouvert M tel que $\phi_z(\mathbb{R}^{m'}) \subset W$ pour tout $z \in M$, alors on peut obtenir des résultats de semi-continuité (ou de continuité) quand z reste dans M. Si ce n'est pas le cas, la situation peut être pathologique comme le montre l'exemple simple suivant : ϕ est une application de $\mathbb{R} \times \mathbb{R}^2$ dans $\mathbb{R}^2$ définie par :

$$\phi(z,X,Y) = (X,X^2+z) \tag{3.61}$$

les ensembles $\phi_z(\mathbb{R}^2)$ sont des variétés algébriques et constituent une famille de paraboles translatées l'une de l'autre.

Remarque 3.3

L'hypothèse ϕ_{z_o} dominante est importante même si ϕ est polynomiale.

<u>Exemple</u> : soit ϕ l'application polynomiale de $\mathbb{R}^2 \times \mathbb{R}$ dans $\mathbb{R}$ définie par :

(3.62) $\phi(z_1,z_2,x) = (z_1^2+z_2^2-1)x+1$;

l'image de ϕ_z est $\mathbb{R}$ quand $z = (z_1,z_2)$ n'est pas sur le cercle d'équation $z_1^2+z_2^2-1 = 0$ et elle est réduite au point $\{1\}$ sinon ; il est clair qu'on n'a pas de continuité de l'image d'une mesure aux points z du cercle unité.

Cet exemple suggère la question suivante : l'ensemble des points z_o où θ_{z_o} ne vérifie pas (3.47) est-il de μ_M-mesure nulle dans le cas où ϕ est polynomiale en z et x ? La réponse est positive.

Proposition 3.1

Si ϕ est une application polynomiale de $M \times V$ dans W et s'il existe $z_o \in M$ tel que ϕ_{z_o} soit dominante, alors l'ensemble des z pour lesquels ϕ_z n'est pas dominante est contenu dans un ensemble algébrique de dimension strictement inférieure à la dimension de M .

Preuve

Elle est simple. Soit z tel que :

(3.63) $\mathrm{rg}(D\phi_z)_x < p$ pour tout $x \in V$;

comme ϕ est polynomiale, celà signifie que z annule pour chaque $x \in V$ des polynômes $J_{1,x},\dots,J_{s,x}$. Considérons l'ensemble des zéros dans M de ces polynômes indexés par x ; on obtient un ensemble algébrique M_1 dans M . Si $\dim M_1 = \dim M$ on aurait alors $M_1 = M$ et donc pour tout z et tout x :

(3.63) $\mathrm{rg}(D\phi_Z)_x < p$;

celà est impossible puisque ϕ_{z_o} est dominant ; on a par conséquent :

(3.64) $\dim M_1 < \dim M$

ce qui conclut la démonstration.

IV - LE PROCESSSUS AUTOREGRESSIF POLYNOMIAL

4.1. Le modèle et son espace euclidien

On considère un processus $(Z_n)_{n\in\mathbb{Z}}$ stationnaire en loi, à valeurs dans $\mathbb{R}^{p'}$ et vérifiant la relation de récurrence :

(4.1) $Z_{n+1} = \phi(Z_n, e_{n+1})$

où :

(4.2) ϕ est une application polynomiale de $\mathbb{R}^{p'} \times \mathbb{R}^{m'}$ dans $\mathbb{R}^{p'}$, de degré un relativement à Z_n.

(4,3) $(e_n)_{n\in\mathbb{Z}}$ est une suite de variables aléatoires i.i.d. à valeurs dans une variété algébrique lisse V de dimension m contenue dans $\mathbb{R}^{m'}$ et e_n est indépendant de $\{Z_{n-1}, Z_{n-2}, \ldots\}$.

(4.4) La loi de probabilité de e_n est absolument continue par rapport à une mesure lebesguienne μ_V sur V et l'ensemble de positivité E de sa densité f est un ouvert non vide de V contenant le point zéro de $\mathbb{R}^{m'}$; (l'ensemble de positivité de f est l'ensemble $\{f > 0\}$ et est défini à un ensemble de μ_V-mesure nulle près).

Sous ces conditions, il est clair que $(Z_n)_{n\in Z}$ est un processus markovien ;

sa probabilité de transition sera notée $P(z,.)$.

Le modèle (4.1) peut se réécrire :

$$Z_{n+1} = (A + \sum_{\delta} B_{\delta}\, e_{n+1}^{\delta}) Z_n + \sum_{\delta} C_{\delta}\, e_{n+1}^{\delta} + D \tag{4.5}$$

où la sommation sur δ est étendue à un ensemble fini d'éléments non nuls de $\mathbb{N}^{m'}$, A et B_{δ} sont des matrices $p' \times p'$ et C_{δ} et D sont des vecteurs de $\mathbb{R}^{p'}$; si $x = (x_1, \ldots, x_{m'})$ et $\delta = (\delta_1, \ldots, \delta_{m'})$, x^{δ} est donné par $x^{\delta} = x_1^{\delta_1} x_2^{\delta_2} \ldots x_{m'}^{\delta_{m'}}$.

Si on itère l'équation (4.5) on obtient :

$$Z_{n+1} = \prod_{j=0}^{k} (A + \sum_{\delta} B_{\delta} e_{n+1-j}^{\delta}) Z_{n-k} + \sum_{m=0}^{k-1} [\prod_{j=0}^{m} (A + \sum_{\delta} B_{\delta} e_{n+1-j}^{\delta}) L(e_{n-m})] \tag{4.6}$$

$$+ L(e_{n+1})$$

où $L(e_n) = \sum_{\delta} C_{\delta}\, e_n^{\delta} + D$.

Nous ferons les hypothèses suivantes :

$(\mathbf{H}_1)$ $$Z_{n+1} = L(e_{n+1}) + \sum_{m=0}^{\infty} [\prod_{j=0}^{m} (A + \sum_{\delta} B_{\delta}\, e_{n+1-j}^{\delta})\, L(e_{n-m})] \quad \text{p.s.}$$

et

$$\lim_{k} \prod_{j=0}^{k} (A + \sum_{\delta} B_{\delta}\, e_{n+1-j}^{\delta}) z = 0 \quad \text{p.s. pour tout } z \in \mathbb{R}^{p'} .$$

$(\mathbf{H}_2)$ Les valeurs propres de A sont de valeur absolue strictement inférieure à 1 .

Notons que, puisque les e_n sont i.i.d., l'hypothèse $(\mathbf{H}_1)$ implique que

e_n et $\{Z_{n-1}, Z_{n-2}, \ldots\}$ sont indépendants.

L'hypothèse $(\mathcal{H}_2)$ nous servira dans la sous-section 4.2.

Les deux hypothèses $(\mathcal{H}_1)$ et $(\mathcal{H}_2)$ sont d'un usage courant en particulier quand le modèle (4.5) est la représentation markovienne d'un processus ARMA [1] ou d'un processus bilinéaire [27] .

Diverses conditions impliquant $(\mathcal{H}_1)$ ont été établies pour les processus bilinéaires (voir par exemple [21] et [27]) ; certaines d'entre elles peuvent être étendues sans grande difficulté au modèle (4.5). La plus simple est la suivante :

$(\mathcal{H}'_1)$ $\exists s > 0$ tel que :

$$E|L(e_n)|^s < \infty \qquad \text{et} \qquad E\|A + \sum_\delta B_\delta e_n^\delta\|^s < 1$$

où $\|.\|$ désigne la norme d'application linéaire.

En effet, on a dans ce cas :

$$(4.7) \qquad E\|\prod_{j=0}^{k} (A + \sum_\delta B_\delta e_{n+1-j}^\delta)\|^s \leq E(\prod_{j=0}^{k} \|A + \sum_\delta B_\delta e_{n+1-j}\|^s)$$

$$\leq \prod_{j=0}^{k} (E\|A + \sum_\delta B_\delta e_{n+1-j}\|^s)$$

par conséquent :

$$(4.8) \qquad \lim_k \prod_{j=0}^{k} (A + \sum_\delta B_\delta e_{n+1-j}^\delta) z = 0 \qquad \text{p.s.}$$

et la série

(4.9) $$\sum_{m=0}^{\infty} \left(\prod_{j=0}^{m} (A + \sum_{\delta} B_\delta\, e_{n+1-j}) \, L(\varepsilon_{n-m})\right) \text{ converge p.s. ;}$$

d'autre part, puisque (Z_n) est stationnaire,

(4.10) $$\prod_{j=0}^{k} (A + \sum_{\delta} B_\delta\, e^{\delta}_{n+1-j}) \, Z_{n-k} \text{ converge en loi vers zéro ;}$$

on en extrait une sous-suite qui converge p.s. , la conclusion découle alors de (4.6) et (4.9) .

Notons $\langle A , B_\delta ; C_\delta , D\rangle$ le plus petit espace vectoriel invariant par la famille de matrices A , B_δ et contenant la famille de vecteurs C_δ , D ; il est clair d'après (H_1) que Z_n appartient presque sûrement à $\langle A , B_\delta ; C_\delta , D\rangle$. Ce dernier sera appelé espace euclidien du processus $(Z_n)_{n\in\mathbb{Z}}$.

4.2. Le point d'attraction du processus $(Z_n)_{n\in\mathbb{Z}}$.

On définit les applications polynomiales ϕ^k de $\mathbb{R}^{p'} \times V^k$ dans $\mathbb{R}^{p'}$ par :

(4.11) $$\phi^1(z,e) = \phi(z,e)$$

$$\phi^k(z,e_1,\dots,e_k) = \phi(\phi^{k-1}(z,e_1,\dots,e_{k-1}),e_k) \ ;$$

notons alors que (4.6) s'écrit :

(4.12) $$Z_{n+1} = \phi^{k+1}(Z_{n-k}\, , e_{n-k+1}\, , e_{n-k+2}\, ,\dots, e_{n+1}) \ .$$

L'hypothèse (H_2) va nous permettre de définir le point d'attraction de $(Z_n)_{n\in\mathbb{Z}}$ suivant la trajectoire $e_n = 0$.

D'après (4.6) on a pour tout z dans $\mathbb{R}^{p'}$:

(4.13) $$\phi^k(z,0,\dots,0) = A^k z + (I+A+\dots+A^{k-1})D \ ;$$

ce qui donne d'après (H_2) :

(4.14) $$\lim_k \phi^k(z,0,\dots,0) = (I-A)^{-1} D \ .$$

Le point :

(4.15) $$T = (I-A)^{-1} D$$

sera appelé point d'attraction de $(Z_n)_{n\in\mathbb{Z}}$ suivant la trajectoire $e_n = 0$.

La propriété essentielle de T est :

(4.16) $$\phi(T,0) = T \ .$$

4.3. La variété algébrique du processus $(Z_n)_{n\in\mathbb{Z}}$

Soit :

(4.17) $D_k = \phi^k(T,E^k)$ (E est l'ensemble de positivité de la densité de e_n) ;

l'égalité (4.16) entraîne que la suite D_k est une suite croissante d'ensembles ; il est facile de voir que (H_1) implique :

(4.18) $$Z_n \in \overline{\bigcup_k D_k} \quad \text{p.s.}$$

Soit maintenant W_k la fermeture de Zariski de $\phi^k(T,V^k)$; comme V^k est irréductible (A 14) il en est de même de W_k . Mais d'après (4.16) (W_k) est une suite croissante de variétés algébriques et par conséquent stationnaire ; il existe donc un entier k_o et une variété algébrique W tels que :

(4.19) $$W = W_k \qquad \text{pour tout } k \geq k_o .$$

La variété W sera appelée variété algébrique des états du processus $(Z_n)_{n\in\mathbb{Z}}$.

Il est clair que :

(4.20) $$Z_n \in W \qquad \text{p.s.}$$

Notons aussi que d'après (4.16) :

(4.21) $$T \in D_1 \subset W .$$

Comme on le constatera dans la suite la dimension de W sera égale à la dimension de l'espace des états de $(Z_n)_{n\in\mathbb{Z}}$; au sens où cet espace des états contient un ouvert non vide de la partie régulière de W ou si l'on veut sa fermeture de Zariski est W . Il nous a paru, par conséquent, intéressant de donner un critère pour trouver la dimension de W .

Proposition 4.1

Si $W_k = W_{k+1}$ *alors* $W_k = W$.

Démonstration

Notons $H_k = \phi^k(T,V^k)$ (W_k est la fermeture de Zariski de H_k) ; on a alors :

(4.22) $$H_{k+2} = \phi(H_{k+1},V)$$
$$\subset \phi(W_{k+1},V)$$

et comme par hypothèse $W_k = W_{k+1}$, on obtient :

(4.23) $$H_{k+2} \subset \phi(W_k,V) \ ,$$

mais par continuité de ϕ pour la topologie de Zariski, $\phi(W_k,V)$ est contenu dans la fermeture de Zariski de $\phi(H_k,V)$ et comme

(4.24) $$\phi(H_k,V) = H_{k+1}$$

on obtient :

(4.25) $$H_{k+2} \subset W_{k+1}$$

et par conséquent

(4.26) $$W_{k+2} \subset W_{k+1} \ ;$$

la suite (W_n) est donc constante à partir de $n = k$ □

Cette proposition et le critère du jacobien (A 21) donnent un algorithme simple pour déterminer la dimension de W ; dans le cas où $V = \mathbb{R}^m$, si on note $J_k(e_1,\dots,e_k)$ le jacobien de $\phi^k(T,.)$ et n_k la taille du plus grand mineur de J_k qui n'est pas identiquement nul, alors la dimension de W est le premier n_k pour lequel $n_k = n_{k+1}$. Dans le cas où $V \neq \mathbb{R}^m$, on utilise (A 21) , ou alors, ce qui revient au même, on prend un ouvert V_1 de V difféomorphe à un ouvert U de $\mathbb{R}^m$ par une application analytique ϕ_1 (ce qui est toujours possible puisque V est une variété analytique de $\dim m$) . On note ψ^k l'application de U^k dans $\mathbb{R}^{p'}$ définie par :

(4.27) $$\psi^k(x_1,\ldots,x_k) = \phi^k(T,\phi_1(x_1),\ldots,\phi_1(x_k)) \ ;$$

soit n_k la taille du plus grand mineur du jacobien de ψ^k , qui n'est pas identiquement nulle sur U^k , alors la dimension de W est le premier n_k pour lequel $n_k = n_{k+1}$.

Remarquons aussi, que si la dimension de l'espace euclidien est r alors en vertu de la proposition 4.1, $W_r = W$; il est donc inutile de regarder ϕ^k pour $k > r$.

Enfin notons aussi que la suite $H_k = \phi^k(T,V^k)$ est une suite croissante d'ensembles semi-algébriques et n'est pas nécessairement stationnaire.

4.4. L'espace des états du processus $(Z_n)_{n\in\mathbb{Z}}$.

Dans cette partie on va montrer que :

(4.28) $$Z_n \in \bigcup_k D_k \qquad \text{p.s.,}$$

On a tout d'abord le lemme suivant :

Lemme 4.1.

Sous les hypothèses (H_1) *et* (H_2) *on a :*

(4.29) $$P(Z_n \in \bigcup_k D_k) > 0$$

Preuve

a) Il est facile de voir en utilisant la définition des ensembles D_k que pour tout entier q :

(4.30) $$\phi^q(\cup D_k, V^q) \subset W$$

et donc par continuité pour la topologie de Zariski :

(4.31) $$\phi^q(W, V^q) \subset W .$$

b) D'autre part, à partir d'un rang r_o, $\phi^r(T,.)$ est dominante ; comme E^r est un ouvert de V^r (E est l'ensemble de positivité de la densité de e_n), il s'ensuit que $\phi^r(T,E^r)$ contient un ouvert W' de la partie régulière de W ; en particulier W' est contenu dans $\underset{k}{\cup} D_k$.

Soit maintenant $(e_1,\ldots,e_r)$ un point de E^r régulier pour $\phi^r(T,.)$ et tel que

(4.32) $$\phi^r(T,e_1,\ldots,e_r) \in W' ;$$

il est clair que pour $T' \in W$, T' voisin de T :

(4.33) $$\phi^r(T',e_1,\ldots,e_r) \in W'$$

de plus encore par continuité, pour T' voisin de T $(e_1,\ldots,e_r)$ est encore un point régulier de $\phi^r(T',.)$. On peut par conséquent trouver un voisinage W'' de T dans W , de sorte que pour tout T' dans W'' , $\phi^r(T',E^r)$ contient un ouvert de la partie régulière de W qui est contenu dans $\cup D_k$.

c) Soit maintenant K un compact de W tel que

$$(4.34) \qquad P(Z_n \in K) > 0$$

(celà est possible puisque $Z_n \in W$ p.s.)

Grâce à l'hypothèse (H_2) et à (4.31), il existe un entier q tel que :

$$(4.35) \qquad \phi^q(z,0,\dots,0) \in W'' \quad \text{pour tout} \quad z \in K$$

et par conséquent $\phi^{q+r}(z,E^{q+r})$ contient un ouvert de la partie régulière de W qui est contenu dans $\cup D_k$; d'après le théorème 3.1 on a alors :

$$(4.36) \qquad P^{q+r}(z, \cup D_k) > 0 \quad \text{pour tout} \quad z \in K .$$

On a maintenant :

$$(4.37) \qquad P(Z_n \in \cup D_k) = \int P(dz)\, P^{q+r}(z, \cup D_k)$$

$$\geq \int_K P(dz)\, P^{q+r}(z, \cup D_k)$$

ce qui est strictement positif en vertu de (4.36) ; le lemme est donc démontré.

On a maintenant :

<u>Théorème 4.1</u>

Sous les hypothèses (H_1) *et* (H_2) *on a :*

(4.38) $\qquad P(Z_n \in \cup D_k) = 1$.

Démonstration

Notons $(\Omega, \mathcal{A}, P)$ l'espace de probabilité sur lequel sont définis $(Z_n)_{n\in\mathbb{Z}}$ et $(e_n)_{n\in\mathbb{Z}}$. $(e_n)_{n\in\mathbb{Z}}$ étant i.i.d., le processus $(Z_n)_{n\in\mathbb{Z}}$ est ergodique en tant que fonction de $(e_n)_{n\in\mathbb{Z}}$ d'après $(\mathcal{H}_1)$ (voir Y. Rozanov [41] ; dans son ouvrage Y. Rozanov utilise le terme métriquement transitif pour désigner l'ergodicité).

Le processus $(Z'_n) = (Z_{-n})$ qui est fonction de $\varepsilon_n = (e_{-n})$ est également ergodique.

Soit maintenant :

(4.39) $\qquad R_n = 1_{\cup D_k} (Z'_n)$;

le processus $(R_n)_{n\in\mathbb{Z}}$ est encore ergodique et on a :

(4.40) $\qquad q = P(R_n = 1) = P(Z_{-n} \in \cup D_k)$

et

(4.41) $\qquad q = E|R_n| = ER_n$

d'après le lemme 4.1 on a :

(4.42) $\qquad q > 0$;

et par le théorème ergodique :

(4.43) $$\lim \frac{1}{N} \sum_{j=0}^{j=N} R_{-n+j} = q \qquad \text{p.s.}$$

Soit Ω_1 , l'ensemble des ω où la convergence (4.43) a lieu et Ω_2 l'ensemble des ω tels que :

(4.44) $$Z_n(\omega) = \phi(Z_{n-1}(\omega), e_n(\omega)) \quad , \forall n \in \mathbb{Z} \ ;$$

il est clair que :

(4.45) $$P(\Omega_1 \cap \Omega_2) = 1 \ ;$$

puisque pour tout $\omega \in \Omega_1 \cap \Omega_2$, (4.43) est vrai et que q est non nul, alors pour tout $\omega \in \Omega_1 \cap \Omega_2$, il existe un entier j_o tel que $R_{-n+j_o}(\omega) = 1$ et donc :

(4.46) $$Z'_{-n+j_o}(\omega) \in \cup D_k$$

c'est-à-dire

(4.47) $$Z_{n-j_o}(\omega) \in \cup D_k \ .$$

ce qui signifie qu'il existe un entier k tel que

(4.48) $$Z_{n-j_o}(\omega) \in \phi^k(T, E^k)$$

mais d'après (4.44) on a :

(4.49) $$Z_n(\omega) = \phi^{j_o}(Z_{n-j_o}(\omega), e_{n-j_o+1}(\omega), \ldots, e_n(\omega))$$

ce qui prouve que

(4.50) $\quad Z_n(\omega) \in \cup D_k$ pour tout $\omega \in \Omega_1 \cap \Omega_2$.

On conclut avec (4.45).

4.5. Mélange géométrique du processsus $(Z_n)_{n\in\mathbb{Z}}$.

Nous allons montrer maintenant que sous l'hypothèse (H3) du théorème 2.1, $(Z_n)_{n\in\mathbb{Z}}$ est géométriquement absolument régulier ; il nous faudra pour celà vérifier que les hypothèses (H1) et (H2) du même théorème sont satisfaites.

Théorème 4.2

Sous les hypothèses (H_1) , (H_2) *et* (H_3) *le processus* $(Z_n)_{n\in\mathbb{Z}}$ *est géométriquement absolument régulier et*

(4.51) $\quad E\, g(Z_n) < \infty$

(g étant la fonction intervenant dans (H3)*)*

Démonstration

L'espace des états de la chaîne (Z_n) est $\cup D_k$ et on munit W d'une mesure lebesguienne μ_W .

Soit K un compact de W .

1 - Prenons un borélien A contenu dans $\cup D_k$ et de μ_W-mesure nulle ; en opérant comme dans la démonstration du lemme 4.1, on peut trouver un entier q tel que $\phi^q(z,.)$ soit dominante pour tout z dans K . Le théorème 3.1 dit alors que la mesure $P^q(z,.)$ est dominée par μ_W et

donc :

(4.52) $P^q(z,A) = 0$ pour tout z dans K .

L'hypothèse (H1) du théorème 2.1 est ainsi vérifiée.

2 - Soit maintenant un borélien A contenu dans $\cup D_k$ et tel que :

(4.53) $\mu_W(A) > 0$

il existe un entier k tel que :

(4.54) $\mu_W(A \cap D_k) > 0$ et $\phi^k(T,.)$ est dominante.

On va montrer qu'il existe une valeur régulière de $\phi^k(T,.)$ y_o dans D_k , telle que pour tout voisinage ouvert $\mathfrak{U}$ de y_o on ait :

(4.55) $\mu_W(\mathfrak{U} \cap A) > 0$;

pour celà, soit G l'ensemble des points critiques de $\phi^k(T,.)$; supposons que pour tout y dans $D_k - \phi^k(T,G)$ il existe un voisinage ouvert $\mathfrak{U}$ de y tel que :

(4.56) $\mu_W(\mathfrak{U} \cap A) = 0$,

il s'ensuit que pour tout x dans $E^k - G$, il existe un voisinage ouvert $\mathfrak{U}'$ de x tel que :

(4.57) $\mu_W(\phi^k(T,\mathfrak{U}') \cap A) = 0$, On obtient ainsi un recouvrement ouvert de E^k-G dont l'image par $\phi^k(T,.)$ est un recouvrement de $D_k-\phi^k(T,G)$; mais E^k-G est un ouvert de V^k, on peut donc extraire un sous-recouvrement dénombrable (paracompacité). On obtient alors un recouvrement dénombrable $(\mathcal{U}_j)_{j\in\mathbb{N}}$ de $D^k-\phi^k(T,G)$ tel que :

(4.58) $\qquad \mu_W(A \cap \mathcal{U}_j) = 0 \qquad$ pour tout entier j.

Comme $\mu_W(\phi^k(T,G)) = 0$ par le théorème de Sard

on en conclut que :

(4.59) $\qquad \mu_W(A \cap D_k) = 0$,

ce qui contredit (4.53) et prouve l'existence de la valeur régulière y_o cherchée.

Nous allons maintenant montrer qu'il existe un voisinage M_o de T tel que pour tout T' dans M_o :

(4.60) $\qquad \phi^k(T',E^k)$ contient un voisinage ouvert de y_o.

Pour cela, soit $x_o \in E^k$ tel que :

(4.61) $\qquad \phi^k(T,x_o) = y_o$

x_o est un point régulier de $\phi^k(T,.)$; en procédant maintenant comme dans le théorème 3.2 on peut choisir des indices $i_1,i_2,\ldots,i_{km-p}$ tels que l'application :

(4.62) $\psi_T(x) = (\phi^k(T,x), x_{i_1}, \dots, x_{i_{km-p}})$

de V^k dans $W \times \mathbb{R}^{km-p}$ soit régulière en x_o .

Il existe alors un voisinage M_o de T tel que pour tout T' dans M_o :

(4.63) $\psi_{T'}(E^k)$ contient un voisinage ouvert de

$$\tilde{y}_o = (y_o , x_{i_1,o}, \dots, x_{i_{km-p},o}) ;$$

En effet, si ce n'était pas le cas, on pourrait trouver une suite T_n convergeant vers T et telle que $\psi_{T_n}(E^k)$ ne contienne pas de voisinage ouvert de $\tilde{y}_o$; mais le lemme 3.1 appliqué à la suite T_n et au voisinage ouvert $U = E^k$ de x_o permet d'extraire une sous-suite (T'_n) pour laquelle $\psi_{T'_n}(E^k)$ contient un voisinage ouvert de $\tilde{y}_o$; ce qui est contradictoire. Maintenant (4.60) se déduit en composant ψ_T avec la projection sur W .

On conclut la démonstration de la manière suivante. Grâce à (H_2) , on peut trouver un entier k' tel que :

(4.64) $\phi^{k'}(z,0,\dots,0) \in M_o$ pour tout z dans K

et donc avec (4.60) :

(4.65) $\phi^{k+k'}(z,E^{k+k'})$ contient un voisinage ouvert de y_o pour

tout z dans K , ce qui donne en utilisant (4.55) et le théorème 3.1 :

(4.66) $P^{k+k'}(z,A) > 0$ pour tout z dans K

et le théorème 3.2 de continuité :

(4.67) $$\inf_{z\in K} P^{k+k'}(z,A) > 0$$

et donc (H2) est vérifiée □

Remarque 4.1

Nous n'avons pas donné de conditions entraînant (H3) pour garder toute sa généralité au résultat ; cependant on peut vérifier que la condition (H_1') entraîne (H3) avec $g(x) = |x|^s+1$ (il suffit de procéder comme dans [20]).

Remarque 4.2

Il faut noter, d'après la démonstration du théorème 2.1, que la loi de probabilité de Z_n est équivalente à la restriction d'une mesure lebesguienne sur $\cup D_k$ et que le processus $(Z_n)_{n\in\mathbb{Z}}$ est Harris-récurrent.

V - PROCESSUS ARMA VECTORIELS ET PROCESSUS BILINEAIRES

5.1. Processus ARMA vectoriels

Soit $(Y_n)_{n\in\mathbb{Z}}$ un processus stationnaire en loi, à valeurs dans $\mathbb{R}^\ell$, satisfaisant l'équation ARMA :

(5.1) $$\sum_{i=0}^{P} B(i)\, Y_{n-i} = \sum_{k=0}^{Q} A(k)\, e_{n-k}$$

où $B(i)$ et $A(k)$ sont des matrices réelles de dimensions respectives

$\ell \times \ell$ et $\ell \times r$; $B(0) = Id$; $(e_n)_{n\in\mathbb{Z}}$ est une suite de variables aléatoires i.i.d. à valeurs dans $\mathbb{R}^r$.

Pour tout $z \in \mathbb{C}$ on définit les matrices :

(5.2) $$P(z) = \sum_{i=0}^{P} B(i)\, z^{P-i} \text{ et } Q(z) = \sum_{k=0}^{Q} A(k)\, z^{Q-k} ;$$

on fait maintenant les hypothèses suivantes :

(5.3) $E|e_n| < \infty$

(5.4) $Ee_n = 0$

(5.5) $\det P(z) \neq 0$ pour $|z| \geq 1$

Sous ces hypothèses Y_n s'écrit :

(5.6) $$Y_n = \sum_{j=0}^{\infty} c_j\, e_{n-j} ;$$

en utilisant (5.1) et (5.6) H. Akaike [1] montre qu'il existe un processus stationnaire markovien $(Z_n)_{n\in\mathbb{Z}}$ à valeurs dans $\mathbb{R}^k$ avec $k = \mathrm{Max}(P,Q+1)$ et tel que :

(5.7) $$Z_n = F\, Z_{n-1} + G\, e_n$$

(5.8) $$Y_n = H\, Z_n$$

où F , G , H sont des matrices réelles et :

(5.9) les valeurs propres de F sont des zéros de $\det P(z)$.

En fait dans sa démonstration H. Akaike suppose que les zéros de $\det Q(z)$

sont en valeur absolue strictement inférieurs à 1 ; ce qui signifie que (e_n) est l'innovation de (Y_n). Cependant cette hypothèse n'est pas nécessaire ; il suffit, dans sa démonstration, de prendre les espérances conditionnellement à la tribu engendrée par $\{e_n, e_{n-1}, \ldots\}$ au lieu de celle engendrée par $\{Y_n, Y_{n-1}, \ldots\}$.

D'après (5.8) le coefficient de mélange de $(Y_n)_{n\in\mathbb{Z}}$ a les mêmes propriétés que celui de $(Z_n)_{n\in\mathbb{Z}}$; les résultats de la section 4 s'appliquent aux processus $(Z_n)_{n\in\mathbb{Z}}$. Les conditions ($\mathbb{H}$1) et ($\mathbb{H}$2) sont immédiatement vérifiées d'après (5.3) et (5.9) ; la condition (H3) est aussi vérifiée grâce à (5.3) et (5.9) en prenant la fonction g définie dans [18] et [19].

On a donc :

Théorème 5.1

Si la loi de e_n *est dominée par la mesure lebesguienne d'une variété algébrique lisse* V *dans* $\mathbb{R}^r$ *et si l'ensemble de positivité de sa densité est un ouvert de* V *contenant zéro, alors* $(Y_n)_{n\in\mathbb{Z}}$ *est géométriquement absolument régulier.*

Ce résultat améliore nettement celui que nous avions obtenu dans [18] et [19] où on supposait que e_n avait une densité strictement positive dans $\mathbb{R}^r$; dans le théorème 5.1, si $V = \mathbb{R}^r$ il suffit alors que l'ensemble de positivité de la densité de e_n soit un voisinage ouvert de 0 dans $\mathbb{R}^r$ et si $V \neq \mathbb{R}^r$, e_n n'a même pas de densité par rapport à la mesure de Lebesgue sur $\mathbb{R}^r$.

5.2. Processus bilinéaires

Soit $(Y_n)_{n\in\mathbb{Z}}$ un processus stationnaire en loi, réel satisfaisant l'équation bilinéaire :

$$(5.10) \qquad Y_n = \sum_{j=1}^{p} a_j\, Y_{n-j} + e_n + \sum_{j=1}^{q} b_j\, e_{n-j} + \sum_{k=1}^{Q} \sum_{j=0}^{P} b_{k_j}\, Y_{n-j-k}\, e_{n-k}$$

où $(e_n)_{n\in\mathbb{Z}}$ est une suite de variables aléatoires réelle i.i.d. et $Ee_n^2 < \infty$.

D.T. Pham [27] montre qu'il existe un processus stationnaire $(Z_n)_{n\in\mathbb{Z}}$ à valeurs dans $\mathbb{R}^r$ avec $r = \mathrm{Max}(p, P+q, P+Q)$ tel que :

$$(5.11) \qquad Z_n = (A+Be_n)Z_{n-1} + Ce_n + De_n^2 + F$$

et

$$(5.12) \qquad Y_n = H\, Z_{n+m-1}$$

où A , B , H sont des matrices, C , D , F des vecteurs et m l'indice du modèle.

Ici également les résultats de la section 4 s'appliquent ; on obtient :

Théorème 5.2

Supposons que :

(i) e_n *admet une densité de probabilité dont l'ensemble de positivité est un voisinage ouvert de zéro ;*

(ii) *les valeurs propres de* A *sont en valeur absolue strictement inférieures à* 1 ,

(iii) $\exists s > 0$ *tel que* $E|e_n|^{2s} < \infty$ *et* $E\,\|A+Be_n\|^s < 1$

alors le processus $(Y_n)_{n\in\mathbb{Z}}$ *est géométriquement absolument régulier ; de plus*

$$E|Y_n|^s < \infty$$

Nous donnons maintenant deux exemples de variété algébrique des états pour des processus $(Z_n)_{n\in\mathbb{Z}}$ bilinéaires vérifiant une équation (5.11) de la forme :

$$Z_n = (A+Be_n)Z_{n-1} + Ce_n \;; \tag{5.13}$$

on suppose que les hypothèses $(\mathcal{H}1)$ et $(\mathcal{H}2)$ sont satisfaites et que e_n admet une densité de probabilité positive sur tout $\mathbb{R}$ ce qui est le meilleur des cas.

Soit $\langle A\,,\,B\,;\,C\rangle$ l'espace euclidien de $(Z_n)_{n\in\mathbb{Z}}$ et W sa variété algébrique des états ; notons $\langle A\,;\,C\rangle$ l'espace invariant par A et engendré par C .

D'après (5.13) il est clair que le point d'attraction est le point 0 .

Notre premier exemple concerne le cas où $W = \langle A,B;C\rangle$.

<u>Exemple 1</u>

<u>Proposition 5.1</u>

Si $\langle A;C\rangle = \langle A,B;C\rangle$ *alors* $W = \langle A,B;C\rangle$.

Démonstration

Elle est simple ; soit $k = \dim \langle A;C\rangle$, $\{C, AC, \dots, A^{k-1}C\}$ est alors une base de $\langle A;C\rangle$; l'écriture de la matrice $D\phi^k(0,.)$ dans la base $\{C, AC, \dots, A^{k-1}C\}$ et au point $(0,0,\dots,0)$ de $\mathbb{R}^k$ montre alors facilement que :

$$(5.14) \qquad \mathrm{rg}(D\phi^k(0,.))_o = k$$

ϕ étant défini à partir de (5.13) par

$$(5.15) \qquad Z_n = \phi(Z_{n-1}, e_n)$$

et ϕ^k à partir de (4.11).

(5.14) implique que $\dim W = k$, ce qui donne nécessairement :

$$(5.16) \qquad W = \langle A;C\rangle = \langle A,B;C\rangle \, .$$

Nous donnons maintenant un exemple simple où W n'est pas égale à l'espace euclidien de $(Z_n)_{n\in\mathbb{Z}}$.

Exemple 2

Prenons :

$$(5.17) \quad A = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} \qquad B = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix} \qquad C = \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}$$

il est facile de vérifier que :

(5.18) $$\langle A,B;C\rangle = \mathbb{R}^3$$

et que $\{C,AC,BC\}$ est une base.

Si on regarde $\phi^2(0,.)$ et $\phi^3(0,.)$ on constate que :

(5.19) $$\text{Max rg } D\phi^2(0,.) = \text{Max rg } D\phi^3(0,.) = 2$$

ce qui en vertu de la proposition 4.1 signifie que :

(5.20) $$\dim W = 2 \ ;$$

il nous suffit donc de trouver l'image de $\phi^2(0,.)$; on a :

(5.21) $$\phi^2(0,e_1,e_2) = ACe_1 + BCe_1e_2 + Ce_2$$

ce qui veut dire que l'image $\phi^2(0,\mathbb{R}^2)$ est la surface d'équation

(5.22) $$Z - XY = 0$$

dans le repère $\{C,AC,BC\}$.

Cette surface est la variété algébrique des états mais c'est aussi dans ce cas particulier l'espace des états.

VI - CONCLUSION

En guise de conclusion on notera que les résultats des sections 4 et 5 ont des conséquences importantes pour les estimateurs empiriques des paramètres du processus bilinéaire (5.10) ; si les conditions du théorème 5.2

sont satisfaites avec $s > 2$ alors le processus $(Y_n)_{n \in \mathbb{Z}}$ est de carré intégrable (il admet des moments d'ordre s) et sa fonction covariance est celle d'un processus ARMA dont la partie autorégressive a pour coefficient les coefficients $a_1, \ldots, a_p$ de l'équation (5.10) [27] .

Les estimateurs empiriques des covariances de $(Y_n)_{n \in \mathbb{Z}}$ sont

(i) consistants par ergodicité de $(Y_n)_{n \in \mathbb{Z}}$.

(ii) convergents en moyenne d'ordre p pour tout $p < \frac{s}{2}$ avec une vitesse $N^{-p/2}$ grâce à une majoration de R. Yokama [40].

(iii) asymptotiquement gaussiens (pour s suffisamment grand) puisque le processus est géométriquement mélangeant (on peut voir P. Hall et C.C. Heyde [10] par exemple).

On peut ensuite estimer $a_1, \ldots, a_p$ par la méthode de Yule-Walker ; ainsi que le montrent R. Azèncott et D. Dacunha-Castelle [2] les estimateurs obtenus ont alors les mêmes propriétés que les estimateurs des covariances. Enfin, pour terminer, on notera que sous les hypothèses des théorèmes 5.1 et 5.2 les processus ARMA et bilinéaires sont Harris-récurrents.

-APPENDICE-

Les développements importants de la géométrie algébrique complexe ont longtemps contrasté avec le peu de résultats obtenus en géométrie algébrique réelle ; celà est dû au fait que des résultats essentiels en géométrie algébrique complexe comme le Nullstellensatz, la connexité d'un ensemble algébrique irréductible, la correspondance entre les points de $\mathbb{C}^n$ et les idéaux maximaux de $\mathbb{C}[X_1,\ldots,X_n]$ sont faux dans le cas réel. Les variétés algébriques réelles que nous définissons plus bas sont généralement étudiées en tant que "ensemble des points réels d'un schéma algébrique sur $\mathbb{R}$" [6] , en fait d'un schéma algébrique affine puisque nous nous limiterons dans notre exposé à des variétés algébriques affines. En dehors des travaux pionniers de J. Nash [23] et H. Whitney [38] ce n'est que très récemment qu'on assiste à un développement systématique de la géométrie algébrique réelle ; (pour une bibliographie on peut regarder [16] , [42]).

Notre objectif dans l'exposé qui suit est de présenter dans un cadre élémentaire des résultats qu'on peut trouver, de manière souvent implicite et dans un langage parfois abstrait, dans la bibliographie que nous avons cité. Il n'y a donc aucun apport original. Nous nous inspirerons beaucoup du livre de D. Mumford [22] qui traite de géométrie projective complexe et nos premiers énoncés de (A1) jusqu'à (A 12) proviennent de Th. Bröcker [4] .

A1

Un sous-ensemble A de $\mathbb{R}^n$ est appelé ensemble algébrique s'il existe des polynômes $F_1, F_2, \ldots, F_r$ dans $\mathbb{R}[X_1,\ldots,X_n]$ tels que :

$$A = \{x \in \mathbb{R}^n, \; F_1(x) = F_2(x) = \ldots = F_r(x) = 0\} \;;$$

En fait, et c'est là encore une différence avec la géométrie algébrique complexe, A peut être défini avec un seul polynôme puisque :

$$F_1(x) = F_2(x) = \ldots = F_r(x) = 0 \Longleftrightarrow F_1^2(x) + F_2^2(x) + \ldots + F_r^2(x) = 0$$

A2

L'ensemble des polynômes qui s'annulent sur A constituent un idéal de $\mathbb{R}[X_1,\ldots,X_n]$ appelé idéal de A (cet idéal n'est pas nécessairement celui engendré par les polynômes $F_1, F_2, \ldots, F_r$ qui ont permis de définir A)

A3

Les ensembles algébriques de $\mathbb{R}^n$ sont les fermés d'une topologie sur $\mathbb{R}^n$ appelée topologie de Zariski. Cette topologie est moins fine que la topologie classique sur $\mathbb{R}^n$; en particulier tout ouvert (resp. fermé) pour la topologie de Zariski est ouvert (resp. fermé) pour la topologie classique.

A 4

Un ensemble algébrique A dans $\mathbb{R}^n$ est dit irréductible si pour tous ensembles algébriques A_1 et A_2 tels que :

$$A = A_1 \cup A_2$$

on a nécessairement $A = A_1$ ou $A = A_2$.

A 5

A est irréductible si et seulement si son idéal est premier.

Dans ce cas on dira que A est une variété algébrique.

(En fait il s'agit d'une variété algébrique affine ; une variété algébrique plus générale étant définie par la propriété d'être localement une variété algébrique affine).

A 6

Soit $\{F_1, F_2, \ldots, F_k\}$ une partie génératrice de l'idéal I d'une variété algébrique V dans $\mathbb{R}^n$; on appelle rang de V (ou codimension dans $\mathbb{R}^n$) le nombre entier

(A 6.1) $$\rho(V) = \underset{x \in V}{\text{Max}} \; \text{rg} \left(\frac{\partial F_i}{\partial x_j}\right)_{\substack{1 \leq i \leq k \\ 1 \leq j \leq n}}$$

$\rho(V)$ ne dépend pas de la partie génératrice de I choisie et on a :

(A 6.2) $$\rho(V) = \underset{F_1, F_2, \ldots, F_r \in I}{\text{Sup}} \; \underset{x \in V}{\text{Max}} \; \text{rg} \left(\frac{\partial F_i}{\partial x_j}\right)_{\substack{1 \leq i \leq r \\ 1 \leq j \leq n}}$$

A 7

Soit V une variété algébrique dans $\mathbb{R}^n$; la dimension de V est définie par :

$$\dim V = n - \rho(V)$$

A 8

Tout ensemble algébrique A est réunion finie de variétés algébriques

$V_1, \ldots, V_k$. La dimension de A est alors :

$$\dim A = \max_{i=1,\ldots,k} \dim V_i .$$

A 9

Si $V \subset W$ sont deux variétés algébriques ; alors :

$$\dim V = \dim W \text{ si et seulement si } V = W .$$

Comme corollaire de ce résultat on a :

A 10 - <u>Stationnarité</u>

Toute suite croissance de variétés algébriques est stationnaire (c'est-à-dire constante à partir d'un certain rang).

A 11

Soit $\{F_1, F_2, \ldots, F_k\}$ une partie génératrice de l'idéal I d'une variété algébrique V dans $\mathbb{R}^n$ et $\rho(V)$ le rang de V ; l'ensemble des points x dans V qui vérifient :

$$\mathrm{rg} \left(\frac{\partial F_i}{\partial x_j}\right)_{\substack{1 \leq i \leq k \\ 1 \leq j \leq n}} < \rho(V)$$

est appelé lieu singulier de V et sera noté Σ_V ; les points de Σ_V sont appelés points singuliers de V . Σ_V est un ensemble algébrique et :

$$\dim \Sigma_V < \dim V .$$

A 12

L'ensemble $R_V = V - \Sigma_V$ est appelé lieu régulier de V . Les points de R_V

sont appelés points réguliers de V .

R_V est une variété analytique réelle dont la dimension est $\dim V$.

A 13

On appelle ouvert lisse de V tout ouvert de la variété analytique R_V .

Si $\Sigma_V = \phi$ on dit que V est une variété algébrique lisse.

A 14

Soient V_1 une variété algébrique de dimension m dans $\mathbb{R}^M$ et V_2 une variété algébrique de dimension n dans $\mathbb{R}^N$, alors $W = V_1 \times V_2$ est une variété algébrique de dimension $m+n$ dans $\mathbb{R}^{M+N}$; de plus :

$$R_{V_1} \times R_{V_2} \subset R_W .$$

Preuve

Notons $X = (X_1,\ldots,X_M)$ et $Y = (Y_1,\ldots,Y_N)$.

Soit I_1 l'idéal de V_1 dans $\mathbb{R}[X]$ et I_2 celui de V_2 dans $\mathbb{R}[Y]$; il est clair que $V_1 \times V_2$ est l'ensemble des zéros de l'idéal engendré par $I_1 \cup I_2$ dans $\mathbb{R}[X,Y]$. C'est donc un ensemble algébrique. Sa dimension est $m+n$ (voir [6]).

Il nous faut montrer qu'il est irréductible.

Soit $\mathfrak{P}$ son idéal dans $\mathbb{R}[X,Y]$; supposons que

(A 14.1) $\qquad f(X,Y)\; g(X,Y) \in \mathfrak{P}$

il nous faut montrer que

(A 14.2) $\quad f(X,Y) \in \mathfrak{P}$ ou $g(X,Y) \in \mathfrak{P}$.

Notons $\mathfrak{a}$ l'idéal engendré dans $\mathbb{R}[X]$ par les polynômes $f(X,y)$ avec y dans V_2 et $\mathfrak{b}$ celui engendré par les polynômes $g(X,y)$. On a alors :

(A 14.3) $\quad \mathfrak{a}\mathfrak{b} \subset I_1$;

en effet, supposons que ce ne soit pas le cas ; on aurait alors y et y' dans V_2 tels que :

(A 14.4) $\quad f(X,y)\ g(X,y') \notin I_1$

et donc un $x \in V_1$ tel que

(A 14.5) $\quad f(x,y)\ g(x,y') \neq 0$

d'où :

$$f(x,y) \neq 0$$

celà signifie que :

(A 14.6) $\quad f(x,Y) \notin I_2$;

mais

(A 14.7) $\quad f(x,Y)\ g(x,Y) \in I_2$ d'après (A 14.1)

et puisque I_2 est premier on a nécessairement :

(A 14.8) $\quad g(x,Y) \in I_2$

et donc

(A 14.9) $\quad g(x,y') = 0$

ce qui contredit (A 14.5) ; (A 14.3) est par conséquent vrai. Mais comme I_1 est premier, l'un des deux idéaux $\mathfrak{a}$ ou $\mathfrak{b}$ est donc contenu dans I_1 ; disons $\mathfrak{a}$. Mais alors d'après la définition de $\mathfrak{a}$ on a

(A 14.10) $\quad f(x,y) = 0$ si $x \in V_1$ et $y \in V_2$

c'est-à-dire :

(A 14.11) $\quad f(X,Y) \in \mathfrak{P}$

et donc $\mathfrak{P}$ est premier et W est irréductible.

Pour montrer que

$$R_{V_1} \times R_{V_2} \subset R_W$$

prenons $x \in R_{V_1}$, $y \in R_{V_2}$, $\phi_1,\dots,\phi_k$ une partie génératrice de I_1, $\psi_1,\dots,\psi_r$ une partie génératrice de I_2 ; on a alors :

$$\mathrm{rg}\ \left(\frac{\partial\phi_i}{\partial x_j}\right)(x) = M - m$$

et

$$\mathrm{rg}\ \left(\frac{\partial\psi_i}{\partial y_j}\right)(y) = N - n$$

Si on note $(z_1,\dots,z_{N+M}) = (x_1,\dots,x_M,y_1,\dots,y_N)$ et $\Phi = (\phi_1,\dots,\phi_k,\psi_1,\dots,\psi_r)$ il est alors facile de voir que

$$\mathrm{rg}\ (\frac{\partial \Phi_i}{\partial z_j})(z) = N+M-(n+m)$$

Comme $\dim W = n+m$, en utilisant (A 6.1) et (A 6.2) on voit que $z = (x,y)$ est un point régulier de W , ce qui conclut la démonstration □

La propriété (A 12) a une généralisation ; introduisons d'abord la notion d'ensemble semi algébrique.

A 15

Un sous ensemble A de $\mathbb{R}^n$ est dit semi algébrique s'il existe une famille finie de polynômes f_{ij} , g_{ij} dans $\mathbb{R}[X_1,\dots,X_n]$ tels que :

$$A = \bigcup_i \{x \in \mathbb{R}^n,\ f_{ij}(x) > 0\ \ \forall j \ \text{ et } \ g_{ij}(x) = 0 \ \ \forall j\}$$

A 16 [6]

La dimension d'un ensemble semi algébrique est la dimension de sa fermeture pour la topologie de Zariski.

Si cette fermeture est irréductible, l'ensemble semi algébrique est dit irréductible.

A 17 Théorème de stratification [17]

Tout ensemble semi algébrique A est réunion disjointe finie de variétés analytiques connexes $A_1,\dots,A_s$; de plus :

$$\dim A = \max_{i=1,\dots,s} \dim A_i \quad \square$$

Nous passons maintenant à la notion d'application polynomiale.

A 18

Soit $V \subset \mathbb{R}^M$ et $W \subset \mathbb{R}^N$ des variétés algébriques et ϕ une application polynomiale de $\mathbb{R}^M$ dans $\mathbb{R}^N$ (i.e. ses composantes sont des polynômes) ; on dit que ϕ est une application polynomiale de V dans W si $\phi(V) \subset W$.

A 19

Une application polynomiale de V dans W est continue pour la topologie de Zariski.

En effet, un fermé F dans W est défini par une équation polynomiale :

$$F = \{y \in W \ , \ G(y) = 0\} \ ;$$

son image réciproque est alors :

$$\phi^{-1}(F) = \{x \in V \ , \ G(\phi(x)) = 0\}$$

c'est donc un fermé de Zariski dans V .

A 20

Un point x dans V est dit point régulier de ϕ si $x \in R_V$, $\phi(x) \in R_W$ et la restriction de ϕ à R_V est de rang $\dim W$ au point x . Les points non réguliers de ϕ sont appelés points critiques de ϕ , et leurs images sont appelées valeurs singulières de ϕ .

A 21 Critère de jacobien [22]

Supposons que $\dim V = m$, $\dim W = n$, $x \in R_V$ et $\phi(x) \in R_W$; soit $F_1,\ldots,F_k$ une partie génératrice de l'idéal de V ; alors x est un point régulier de ϕ si et seulement si :

$$\mathrm{rg}\ \left(\left(\frac{\partial F_i}{\partial x_j}\right) \vdots \frac{\partial \phi_i}{\partial x_j}\right) = M - m + n$$

Ce critère est utile parce que ϕ est donnée sur tout $\mathbb{R}^M$ et V est définie par des équations. Quand $V = \mathbb{R}^M$ la partie $\left(\frac{\partial F_i}{\partial x_j}\right)$ n'a pas lieu d'être.

A 22 Théorème de Tarski-Seidenberg [32]

L'image d'une variété algébrique par une application polynomiale est un ensemble semi algébrique.

A 23

Soit ϕ une application polynomiale de la variété algébrique V dans la variété algébrique W . On dit que ϕ est dominante si W est la fermeture de Zariski de $\phi(V)$.

Il revient au même de dire que $\phi(V)$ contient un ouvert non vide de R_W ou que ϕ admet un point régulier ; celà se voit facilement en utilisant (A 16) , (A 17) et (A 22) .

Notons qu'on a alors nécessairement :

$$\dim V \geq \dim W .$$

Soit G l'ensemble des points critiques de ϕ et y un point de $W-\phi(G)$, alors $\phi^{-1}(y)$ est un ensemble algébrique de dimension

$$\dim V - \dim W .$$

Que ce soit un ensemble algébrique est évident ; pour la dimension on peut restreindre ϕ à $R_V - \{\phi^{-1}(\Sigma_W) \cup G\}$ qui est une sous-variété analytique de R_V ; l'image de cette sous-variété par ϕ est dans R_W ; on applique par conséquent le résultat connu pour les variétés analytiques (J. Dieudonné [7]) : $\phi^{-1}(y)$ est une variété analytique de dimension

$$\dim V - \dim W$$

on conclut en utilisant (A 12).

Notons comme conséquence que si $\dim V = \dim W$ alors $\operatorname{card} \phi^{-1}(y) < \infty$.

A 24

Soit V une variété algébrique lisse, ϕ une application polynomiale dominante de V dans une variété algébrique W ; notons G l'ensemble des points critiques de ϕ. Alors $\phi(G)$ est contenu dans un ensemble algébrique de dimension strictement inférieure à $\dim W$.

Pour la démonstration, on note que G est un ensemble algébrique puisqu'il apparaît comme réunion de $\phi^{-1}(\Sigma_W)$ et de l'ensemble des zéros d'une famille de polynômes (des mineurs du jacobien). Sa dimension est strictement inférieure à $\dim V$ car ϕ est dominante. $\phi(G)$ est un semi algébrique (A 22) soit L sa fermeture de Zariski. Si $L = W$, L contiendrait un ouvert de R_W et $\phi(G)$ également d'après (A 16) et (A 17), mais celà contredirait le théorème de Sard pour les variétés analytiques [7] .

Comme conséquence on a :

A 25

Si ϕ est une application polynomiale dominante d'une variété algébrique lisse V dans une variété algébrique W alors :

$$\dim W = \max_{x \in V} \operatorname{rg} (D\phi)(x) \ .$$

A 26 Théorème de la fibre localement triviale [11]

Soit ϕ une application polynomiale dominante d'une variété algébrique V dans une variété algébrique W ; alors il existe un ensemble algébrique Z dans W de dimension strictement inférieure à $\dim W$ tel que tout point y dans $W-Z$ possède un voisinage ouvert U de sorte que pour tout $x \in U$, $\phi^{-1}(x)$ et $\phi^{-1}(y)$ soient homéomorphes.

En particulier, quand $\dim V = \dim W$ et y est une valeur régulière de ϕ alors il existe un voisinage ouvert U de y dans W tel que pour tout x dans U on ait :

$$\operatorname{card} \phi^{-1}(x) = \operatorname{card} \phi^{-1}(y) < \infty$$

(il suffit d'utiliser (A 23)).

A 27 Résolution des singularités [12]

Soit V une variété algébrique dans $\mathbb{R}^M$ de dimension m . Alors il existe une variété analytique $\hat{V}$ de dimension m et une application analytique ℓ de $\hat{V}$ dans $\mathbb{R}^M$ telle que :

(i) $\ell(\hat{V}) = V$

(ii) ℓ est une application propre de $\hat{V}$ sur V (ie. l'image réciproque de tout compact est compacte).

-REFERENCES-

[1] H. AKAIKE, *Markovian representation of stochastic processes*, Ann. Inst. Stat. Math., 26 (1974), 363-387.

[2] R. AZENCOTT et D. DACUNHA-CASTELLE, *Séries d'observations irrégulières*, Masson, Paris (1984).

[3] J.R. BLUM, D.L. HANSON, L.H. KOOPMANS, *On the strong law of large numbers for a class of stochastics processes*, Z. Wahr. Verw. Gebiete, 2 (1963), 1-11.

[4] Th. BRÖCKER, *Differentiable germs and catastrophes*, Cambridge University Press (1975).

[5] Y.A. DAVYDOV, *Mixing conditions for Markow chains*, Th. Prob. Appl., 28 (1973), 313-328.

[6] H. DELFS and M. KNEBUSH, *Semi algebraic topology over a real closed fields I et II*, Math. Zeit., n° 177, 107-129 et n° 178, 175-213 (1981).

[7] J. DIEUDONNE, *Eléments d'analyse tome III*, Gauthier-Villars, Paris (1970).

[8] J.L. DOOB, *Stochastic processes*, Wiley, New York (1953).

[9] V.V. GORODETSKI, *On the strong mixing property for linear sequences*, Th. Prob. Appl., 22, 411-413 (1977).

[10] P. HALL and C.C. HEYDE, *Martingale limit theory and its application*, London Academic (1980).

[11] R.M. HARDT, *Semi algebraic local triviality in semi algebraic mappings*, Ann. Journ. Math., 102, 291-302.

[12] H. HIRONAKA, *Resolution of singulariries of an algebraic variety, I-II*, Ann. Math., 79, 109-326 (1964).

[13] I.A. IBRAGIMOV and Y.V. LINNIK, *Independant and stationary sequences of random variables*, Walth-Noordhoof publishing Gröningen (1974).

[14] I.A. IBRAGIMOV et Y. ROZANOV, *Processus aléatoires gaussiens*, MIR, Moscou (1974).

[15] N. JAIN and B. JAMISON, *Contributions to Doeblin's theory of Markov processes*, Z. Wahr. verw. Gebiete, 8, 19-40 (1967).

[16] T.Y. LAM, *An introduction to real algebra*, Rocky Mountain Journ. Math., 14, 4 (1984).

[17] S. LOJASIEWICZ, *Ensembles semi analytiques*, multigraphie de l'I.H.E.S., Bures/Yvette (1965).

[18] A. MOKKADEM, *Sur le mélange d'un processus ARMA vectoriel*, CRAS, série I, t. 303, 519-521 (1986).

[19] A. MOKKADEM, *Mixing properties of ARMA processes*, soumis à Stoch.Proc.Appl.

[20] A. MOKKADEM, *Sur un modèle autorégressif non linéaire, ergodicité et ergodicité géométrique*, J.T.S.A., 8, 195-204 (1987).

[21] A. MOKKADEM, *Conditions suffisantes d'existence et d'ergodicité géométrique des modèles bilinéaires*, CRAS, série I, t. 301, 375-377 (1985).

[22] D. MUMFORD, *Algebraic geometry I, Complex projective varieties*, Springer Verlag, Berlin (1976).

[23] J. NASH, *Real algebraic manifolds*, Ann. Math., 56, 405-421 (1952).

[24] E. NUMMELIN and P. TUOMINEN, *Geometric ergodicity of Harris recurrent Markov chains*, Stoch. Proc. Appl., 12, 187-202 (1982).

[25] **S. OREY**, *Limit theorems for Markov chain transition probabilities*, Van Nostrand, London (1971).

[26] **T.D. PHAM and L.T. TRAN**, *Some mixing properties of time series models*, Stoch. Proc. Appl., 19, 297-303 (1986).

[27] **T.D. PHAM**, *Bilinear markovian representation and bilinear models*, Stoch. Proc. Appl., 20, 295-306 (1985).

[28] **M.Q. PHAM**, *Introduction à la géométrie des variétés différentiables*, Dunod, Paris (1969).

[29] **D. REVUZ**, *Markov chains*, North Holland, Amsterdam (1984).

[30] **M. ROSENBLATT**, *Markov processes. Structure and asymptotic behaviour*, Springer, Berlin (1971).

[31] **M. ROSENBLATT**, *A central limit theorem and a strong mixing condition*, Proc. Nat. Acad. Sc. U.S.A., 42, 43-47 (1956).

[32] **A. SEIDENBERG**, *A new decision method for elementary algebra*, Ann. Math., 2, 365-374 (1952).

[33] **H. TAKAHATA**, L_∞ *bounds for asymptotic normality of weakly dependent summands using Stein's methods*, Ann. Prob., 9, 676-683 (1981).

[34] **A.N. TIKHOMIROV**, *On the convergence rate in the central limit theorem for weakly dependent random variables*, Th. Prob. Appl., 25, 790-809 (1980).

[35] **R.L. TWEEDIE**, *Sufficient conditions for ergodicity and recurrence of Markov chains*, Stoch. Proc. Appl., 3, 385-403 (1975).

[36] **R.L. TWEEDIE**, *The existence of moments for stationnary Markov chains*, J. Appl. Prob., 20, 191-196 (1983).

[37] **F.W. WARNER**, *Foundations of differentiable manifolds and Lie groups*, Singer, MIT (1971).

[38] **H. WHITNEY**, *Elementary structure of real algebraic varieties*, Ann. Math., 66, 545-556 (1957).

[39] **C.S. WITHERS**, *Conditions for linear processes to be strong mixing*, Z. Wahr. verw. gebiete, 57, 481-494 (1981).

[40] **R. YOKOYAMA**, *Moments bounds for stationnary mixing sequences*, Z. Wahr. verw. gebiete, 52, 45-57 (1980).

[41] **Y. ROZANOV**, *Stationary random processes*, Holden Day Series (1967).

[42] **E. BECKER**, *On the real spectrum of a ring and its applications to semi algebraic geometry*, Bull. A.M.S., 15, 19-60 (1986).

C H A P I T R E IV

STUDY OF RISKS OF KERNEL ESTIMATORS

Abdelkader MOKKADEM

Abstract.

Let $(X_n)_{n\in\mathbb{Z}}$ be a strictly stationary strong mixing sequence of $\mathbb{R}^d$-valued random variables, $f(x)$ the probability density function of X_n and $f_N(x)$ a kernel estimator of $f(x)$ based on a sample of size N. In this paper, we use Fourier method to study the rates of convergence of the risks $E||x^k f_N(x)-x^k f(x)||_q^p$ for $p>0$ and $0<q\leq\infty$. The rates obtained are the same as in the i.i.d. case. Similar risks for kernel estimators of the conditional density, the regression function and other functionals are also investigated.

Key words.

Kernel estimators ; uniform risk ; mean risk ; regression function ; strong mixing ; Fourier transform.

AMS 1980 subject classifications : Primary 62G05, 62H12.

1. INTRODUCTION

Let $(X_n)_{n\in\mathbb{Z}}$ be a strictly stationary strong mixing sequence of random variables with values in $\mathbb{R}^d$ and $f(x)$ the probability density function of X_n. Following Rosenblatt [13] and Parzen [11], $f(x)$ can be estimated by :

$$f_N(x) = \frac{1}{Nh_N^d} \sum_{j=1}^{N} K\left(\frac{x-X_j}{h_N}\right)$$

where K is a positive bounded kernel and h_N is a sequence of positive real numbers.

In the present paper, the Fourier method is used to give an evaluation of the risks

$$E\|x^k f_N(x)-x^k f(x)\|_q^p, \qquad p>0, \quad 0<q\leq\infty$$

where $\|\cdot\|_q$ is the usual norm in $L^q(\mathbb{R}^d)$. Our main result is that under some conditions on f, K and the mixing of (X_n) (see section 3) we have :

1 - $$E \sup_{\mathbb{R}^d} |x^k f_N(x)-x^k f(x)|^p = 0\left(\frac{1}{N^{p/2}h_N^{dp}} + h_N^{2p}\right)$$

2 - for $2\leq q<\infty$ and $q' = \frac{q}{q-1}$:

$$E\left(\int_{\mathbb{R}^d} |x^k f_N(x)-x^k f(x)|^q\, dx\right)^{p/q} = 0\left(\frac{1}{N^{p/2}h_N^{dp/q'}} + h_N^{2p}\right)$$

3 - for $0<q<2$:

$$E\left(\int_{\mathbb{R}^d} |x^k f_N(x)-x^k f(x)|^q dx\right)^{p/q} = 0\left(\frac{1}{N^{p/2}h_N^{dp/2}} + h_N^{2p}\right)$$

Except a recent work of Doukhan-Leon-Portal [5] on the risk $E \sup_S |f_N(x)-f(x)|$ where S is compact, it seems that the risk $E\|f_N-f\|_\infty^p$ have not been studied under mixing conditions. However the risk $E\|f_N-f\|_2$ have been extensively studied and $E\|f_N-f\|_1$ is evaluated in the Rosenblatt's work [15] ; it is also obtained with a great precision by Devroye-Penrod [4] and Devroye-Györfi [3] in the one dimensional case when (X_n) are i.i.d. and f has a compact support ; but we do not know results for $E\|x^k f_N(x)-x^k f(x)\|_q^p$ with general p and q under mixing conditions.

Our regularity hypothesis on f are stated in the distribution's language (see section 2). We prove, in proposition 1 (section 3), that they are equivalent to the condition introduced by Devroye-Györfi [3] for $q = 1$:

$$\sup_{a>0} \int |(f * \phi_a)''| < \infty \qquad \text{where } \phi \text{ is a positive density with}$$

compact support.

The distribution approach clarifies somewhat, the Devroye-Györfi's condition. Our choice of the strong mixing property is motivated by the following consideration. We first recall the definition of strong mixing, Rosenblatt [14]. Let $\mathfrak{a}_0$ be the σ-algebra generated by $\{X_j\ ,\ j \le 0\}$, $\mathfrak{a}^k$ those generated by $\{X_j\ ,\ j \ge k\}$ for $k \ge 0$ and :

$$\alpha(k) = \sup\{|P(A \cap B)-P(A).P(B)| \ ;\ A \in \mathfrak{a}_0,\ B \in \mathfrak{a}^k\}$$

The process (X_n) is strong mixing if $\lim_{k\to\infty} \alpha(k) = 0$; $\alpha(k)$ are called the strong mixing coefficient of (X_n). In recent works, the strong mixing property have been proved for many processes : linear process (Gorodetskii [6], Whithers [19], Pham-Tran [12]) ARMA (Mokkadem [9]) and some non linear auto-regressive processes (Mokkadem [8]). It seems then of interest to estimate, under strong mixing conditions, the density and other functionals such that regression and conditional density.

We study in section 4, kernel estimators of this functionals.

In section 2 we give some notations and hypothesis.

Some of our results have been announced in [10].

2. NOTATIONS AND HYPOTHESIS

2.1.

We need the following notations.

The Fourier transform of a complex valued function $F(x)$ absolutely integrable on $\mathbb{R}^d$ is the complex valued function :

$$\hat{F}(s) = \left(\frac{1}{\sqrt{2\pi}}\right)^d \int_{\mathbb{R}^d} e^{is.x} F(x)dx$$

If $F(x)$ is vector valued, $F(x) = (F_1(x),\dots,F_n(x))$, his Fourier transform will be $\hat{F}(s) = (\hat{F}_1(s),\dots,\hat{F}_n(s))$. For $k = (k_1,\dots,k_d)$ and $\nu = (\nu_1,\dots,\nu_d)$ elements of $\mathbb{N}^d$, we put :

$$[\nu] = \nu_1+\dots+\nu_d$$

$$D^\nu F = \frac{\partial^{\nu_1+\dots+\nu_d}}{\partial x_1^{\nu_1}\dots\partial x_d^{\nu_d}} F$$

$$C_k^\nu = C_{k_1}^{\nu_1}\dots C_{k_d}^{\nu_d}$$

$$x^\nu = x_1^{\nu_1}\dots x_d^{\nu_d}$$

We shall say that $\nu \le k$ if and only if $\nu_j \le k_j$ for all $j = 1,\dots,d$.

2.2.

Let $C_c^\infty(\mathbb{R}^d)$ be the space of C^∞ functions with compact support and T a distribution on $\mathbb{R}^d$.

For $q \in [1,+\infty]$ and $q' = \frac{q}{q-1}$ we put :

$$||T||_q = \sup\{|T(\phi)|, \quad \phi \in C_c^\infty(\mathbb{R}^d), \ ||\phi||_{q'} \le 1\}$$

where $||\phi||_{q'}$ is the usual norm of ϕ in $L^{q'}(\mathbb{R}^d)$.

In the sequel $T(\phi)$ will be noted $\langle T,\phi\rangle$. Clearly, for $1 < q \le \infty$, $1 \le q' < \infty$

and $C_c^\infty(\mathbb{R}^d)$ is dense in $L^{q'}(\mathbb{R}^d)$. Then $||T||_q$ is finite if and only if T is an element of the dual space of $L^{q'}(\mathbb{R}^d)$; i.e. T is an element of $L^q(\mathbb{R}^d)$; in this case $||T||_q$ is the usual norm in $L^q(\mathbb{R}^d)$.

Now for $q = 1$, $q' = \infty$ and $C_c^\infty(\mathbb{R}^d)$ is dense for the uniform norm in the space $C_c(\mathbb{R}^d)$ of continuous functions with compact support ; the dual of $C_c(\mathbb{R}^d)$ is the space of bounded regular measures on $\mathbb{R}^d$. So that $||T||_1 < \infty$ if and only if T is a bounded regular measure and then $||T||_1$ is the variation norm of the measure T.

For more on Distribution and Measure Theory see Schwartz [17] and Rudin [16].

2.3.

We shall use the following hypothesis, where $\alpha(j)$ are the strong mixing coefficient of (X_n), $k \in \mathbb{N}^d$, $p>0$, $1\le q \le \infty$ and F is a function defined on $\mathbb{R}^d$.

$\mathfrak{M}(b,p)$: for some $\delta > 0$, $E|X_n|^{b(\beta+\delta)} < \infty$ and $\sum_{j=0}^{\infty} (j+1)^{\beta/2\,-1}\,\alpha(j)^{\delta/\beta+\delta} < \infty$, with $\beta = \mathrm{Max}(p,2)$.

$\mathfrak{I}(k)$: if $\nu \in \mathbb{N}^d$ and $\nu \le k$ then $x^\nu F(x)$ and $D^\nu(\hat{F})$ are absolutely integrable.

(Here, because $x^\nu F(x)$ is absolutely integrable $D^\nu\hat{F}$ is well defined as an usual derivative).

In the following hypothesis, the derivatives are distributions derivatives ; to distinguish them from the usual derivatives, they will be noted $\bar{D}^\nu$.

$\mathfrak{D}(k,q)$: 1. $x^\nu F(x) \in L^q(\mathbb{R}^d)$ for $\nu \in \mathbb{N}^d$, $\nu \le k$.

2. $||\bar{D}^\nu(x^k F(x))||_q < \infty$ for $\nu \in \mathbb{N}^d$, $[\nu] = 2$

3. If $k \neq 0$, then $||\bar{D}^\nu(x^{k-\beta}F(x))||_q < \infty$ for $\beta \in \mathbb{N}^d$, $\nu \in \mathbb{N}^d$, $\beta \le k$, $[\beta] = 1$, $[\nu] = 1$.

We shall see later that the condition $||\bar{D}^\nu F||_q < \infty$ is equivalent to the condition used by Györfi-Devroye for $q = 1$: $\sup_{a>0}\int |D^\nu(F * \phi_a)|^q < \infty$ where $\phi_a(x) = \frac{1}{a^d}\,\phi(\frac{x}{a})$ and ϕ is a density of class $C^{[\nu]}$ with compact support in $\mathbb{R}^d$.

2.4.

The density $f(x)$ is estimated by

$$f_N(x) = \frac{1}{N\, h_N^d} \sum_{j=1}^{N} K\left(\frac{x-X_j}{h_N}\right)$$

where h_N is a sequence of positive real numbers and K is a positive bounded kernel such that :

(1) $$\lim h_N = 0 \ ; \quad \int K(x)\,dx = 1 \ ; \quad \int |x|^3 K(x)\,dx < \infty$$

(2) $$\int x_j K(x)\,dx = 0 \quad \text{and} \quad \int x_j^2 K(x)\,dx = \tau^2 \qquad \text{for } j = 1,\dots,d$$

(3) $$\int x_i x_j K(x)\,dx = 0 \qquad \text{for } i \neq j$$

We go now to our main results.

3. RISKS FOR THE DENSITY ESTIMATORS

3.1.

We state the main results ; their proofs are deferred to the end of the present section.

Theorem 1. *Let* $0 < p < \infty$, $2 \le q \le \infty$, $q' = \frac{q}{q-1}$, *and* $k \in \mathbb{N}^d$. *Assume that :*

i) $(X_n)_{n\in\mathbb{Z}}$ *satisfies* $\mathfrak{M}([k],p)$

ii) f *satisfies* $\mathfrak{D}(k,q)$;

iii) K *satisfies* $\mathfrak{J}(k)$.

Then :

$$E\|x^k f_N(x) - x^k f(x)\|_q^p = O\left(\frac{1}{N^{p/2} h_N^{dp/q'}} + h_N^{2p}\right)$$

Remark 1.

Looking at i) and ii), we note that for $k = 0$, there is no condition on the existence of the moments of (X_n).

For $1 \leq q < 2$ the rate obtained is the same as for $q = 2$:

<u>Theorem 2</u>. *Let* $0 < p < \infty$, $1 \leq q < 2$, $k \in \mathbb{N}^d$ *and* $k' = k+e+1 = (k_1+e+1, \ldots, k_d+e+1)$ *where* e *is the integer part of* $\frac{d}{q} \cdot \frac{2-q}{2}$; *assume that :*

i) $(X_n)_{n \in \mathbb{Z}}$ *satisfies* $\mathfrak{M}([k]+e+1,p)$;

ii) f *satisfies* $\mathfrak{D}(k,q)$;

iii) K *satisfies* $\mathfrak{J}(k')$.

Then :

$$E\| x^k f_N(x) - x^k f(x) \|_q^p = 0\left(\frac{1}{N^{p/2} h_N^{dp/2}} + h_N^{2p}\right)$$

<u>Remark 2</u>.

For $k = 0$ and $p = d = q = 1$, the hypothesis i) implies that $\int x^2 f(x)\,dx < \infty$ and then $\int f^{1/2} < \infty$ (see [3]) ; this last condition is necessary to obtain the rate of theorem 2 [3].

The following theorem will be obtained as consequence of theorems 1 and 2.

<u>Theorem 3</u>. *Let* $0 < p < \infty$, $0 < q < 1$, $k \in \mathbb{N}^d$, $k' = k+e+1$ *and* $k'' = k+e'+1$, *where* e *is the integer part of* $\frac{d}{q} \cdot \frac{2-q}{2}$ *and* e' *the integer part of* $\frac{d}{q}(1-q)$. *Assume that :*

i) $(X_n)_{n \in \mathbb{Z}}$ *satisfies* $\mathfrak{M}([k]+e+1,p)$;

ii) f *satisfies* $\mathfrak{D}(k'',1)$;

iii) K *satisfies* $\mathfrak{J}(k')$.

Then :

$$E\| x^k f_N(x) - x^k f(x) \|_q^p = 0\left(\frac{1}{N^{pq/2} h_N^{dpq/2}} + h_N^{2pq}\right)$$

<u>Remark 3</u>.

Here, also, the rate of convergence is the same as for $q = 2$. The number q appears in the formula because, for $q < 1$, $\| f \|_q$ is defined by $\int |f|^q$ while for $q \geq 1$ it is $(\int |f|^q)^{1/q}$.

Now we go to the proofs.

3.2. Proofs of theorems 1, 2, 3.

We need some lemmas. The following is due to Yokoyama [20] and will be used to bound the quantity $E||x^k f_N(x) - x^k E f_N(x)||_q^p$.

Lemma 1. Yokoyama [20]. *Let* $q > 0$ *and* $(Z_n)_{n\in\mathbb{Z}}$ *a strictly stationary strong mixing sequence of random variables with values in* $\mathbb{C}^r$, *such that* $E\, Z_n = 0$ *and* $E|Z_n|^{\beta+\delta} < \infty$ *for some* $\delta > 0$ *and* $\beta = \mathrm{Max}(q,2)$.

If the strong mixing coefficients $\alpha(j)$ *of* (Z_n) *satisfy*

$$\sum_{j=1}^{\infty} (j+1)^{\beta/2-1} \quad \alpha(j)^{\delta/\beta+\delta} < \infty$$

Then

$$E|Z_1+\ldots+Z_N|^q \leq C\ N^{q/2}$$

where the constant C *depends only on* $\alpha(j)$, $E|Z_1|^{\beta+\delta}$ *and* r.

In the lemma 2 below, Δ is the Laplacian operator for distributions and $K_h(x) = \frac{1}{h^d} K(\frac{x}{h})$.

Lemma 2. *Let* $1 \leq q \leq \infty$ *and* $\beta \in \mathbb{N}^d, [\beta] = 1$ *; assume that for some* $r \geq 1$, $F \in L^r(\mathbb{R}^d)$. *Then*

i) $||K_h * F - F||_q \leq h^2 \frac{\tau^2}{2} \sum_{[\nu]=2} ||\overline{D}^\nu F||_q$;

ii) $\underset{h\to 0}{\mathrm{Lim\ Inf}}\ h^{-2} ||K_h * F - F||_q \geq \frac{\tau^2}{2} \quad ||\Delta F||_q$;

iii) $||(x^\beta K_h) * F||_q \leq h^2\ \tau^2 \sum_{[\nu]=1} ||\overline{D}^\nu F||_q$;

iv) $\underset{h\to 0}{\mathrm{Lim\ Inf}}\ h^{-2} ||(x^\beta K_h) * F||_q \geq \tau^2 ||\overline{D}^\beta F||_q$.

Proof.

i) Let $\phi \in C_c^\infty(\mathbb{R}^d)$ with $||\phi||_{q'} \le 1$, $q' = \frac{q}{q-1}$.

(4) $$<(K_h * F - F), \phi> = \int_{\mathbb{R}^d} \int_{\mathbb{R}^d} K_h(x) F(y) \phi(x+y)\, dxdy - \int_{\mathbb{R}^d} F(y)\phi(y)\, dy$$

$$= \iint K_h(x) F(y) (\phi(x+y) - \phi(y))\, dxdy$$

By a Taylor development of ϕ and using (2) we obtain :

(5) $$<(K_h * F - F), \phi> = \sum_{[\nu]=2} \iint K_h(x) F(y) x^\nu [\int_0^1 (1-t) D^\nu \phi(y+tx)\, dt] dxdy$$

and then :

(6) $$|<(K_h * F - F), \phi>| \le \sum_{[\nu]=2} \int_{\mathbb{R}^d} \int_0^1 |x^\nu| K_h(x)\, (1-t)\, |\int_{\mathbb{R}^d} F(y) D^\nu \phi(y+tx)\, dy|\, dxdt$$

Now :

(7) $$\int F(y) D^\nu \phi(y+tx)\, dy = (-1)^\nu <\bar{D}^\nu F, \phi_{tx}>$$

where $\phi_{tx}(y) = \phi(y+tx)$

clearly $||\phi_{tx}||_{q'} = ||\phi||_{q'} \le 1$

so that :

(8) $$|<\bar{D}^\nu F, \phi_{tx}>| \le ||\bar{D}^\nu F||_q$$

by (6) and (8) we conclude (i).

ii) In (4) we write the third order development of ϕ and use condition (2) ; we obtain :

$$(9) \qquad <(K_h * F - F),\phi> = h^2 \frac{\tau^2}{2} \int F(y)\Delta\phi(y)dy +$$

$$+ \sum_{[\nu]=3} \iint K_h(x)F(y)x^\nu [\int_0^1 \frac{(1-t)^2}{2} D^\nu \phi(y+tx)dt]dxdy$$

Now by (1) and because $\phi \in C_c^\infty(\mathbb{R}^d)$, the second term in the right of (9) is bounded by $C\,h^3$ so that :

$$(10) \qquad |<(K * F - F),\phi>| \geq h^2 \frac{\tau^2}{2} |<\Delta F,\phi>| - C\,h^3$$

Then :

$$(11) \qquad \liminf_{h\to 0} h^{-2}|<(K_h * F - F),\phi>| \geq \frac{\tau^2}{2}|<\Delta F,\phi>|.$$

This implies that :

$$\liminf_{h\to 0} h^{-2}|| K_h * F - F||_q \geq \frac{\tau^2}{2} |<\Delta F,\phi>|$$

for any $\phi \in C_c^\infty(\mathbb{R}^d)$ such that $||\phi||_{q'} \leq 1$. We obtain then ii).

iii), iv) :

$$<(x^\beta K_h) * F,\phi> = \iint x^\beta K_h(x)F(y)\phi(x+y)dxdy$$

but by (2), for $[\beta] = 1$:

$$\iint x^\beta K_h(x)F(y)\phi(y)dxdy = \int x^\beta K_h(x)dx \int F(y)\phi(y)dy = 0.$$

Then we obtain :

$$<(x^\beta K_h) * F,\phi> = \iint x^\beta K_h(x)F(y)(\phi(x+y)-\phi(y))dxdy$$

For iii) we write the first order development of $[\phi(x+y)-\phi(y)]$, for iv) the second order development, and we carry on the proof like in i) and ii).

Now we can prove the following lemma on the rate of convergence of the bias of our estimators.

Lemma 3. *Let* $1 \le q \le \infty$ *; assume that for* $\nu \in \mathbb{N}^d$, $\nu \le k$, $x^\nu K(x)$ *is absolutely integrable and that* f *satisfies* $\mathcal{D}(k,q)$. *Then :*

$$\| x^k (f * K_h) - x^k f \|_q = O(h^2).$$

Proof. We first write $x^k (f * K_h)$ as a sum of convolution products. For this we put :

$$x = (x-y)+y$$

by calculation we obtain :

$$x^k = \sum_{\nu \le k} C_k^\nu (x-y)^\nu y^{k-\nu} \tag{12}$$

This implies that :

$$x^k (f * K_h) = \sum_{\nu \le k} C_k^\nu [(x^\nu K_h) * (x^{k-\nu} f)] \tag{13}$$

Now for $[\nu] \ge 2$ we have :

$$\begin{aligned} \|(x^\nu K_h) * (x^{k-\nu} f)\|_q &\le \| x^\nu K_h \|_1 \| x^{k-\nu} f \|_q \\ &\le C\, h^{[\nu]} \end{aligned} \tag{14}$$

And by lemma 2 :

$$\|(x^\nu K_h) * (x^{k-\nu} f)\|_q = O(h^2) \qquad \text{for } [\nu] = 1$$

and

$$\| K_h * (x^k f) - x^k f \|_q = O(h^2)$$

The lemma 3 is then proved.

Remark 4.

The results ii) and iv) of lemma 2 give necessary conditions to obtain a rate h^2 for the bias. In the case $k = 0$ and $d = 1$ we see that a necessary and sufficient condition to obtain a rate h^2 for the bias is : $\| \bar{D}^2 f \|_q < \infty$.

A greater rate h^s, $s > 2$ can be obtained with kernels satisfying $\int x^\nu K(x)\,dx = 0$ for $[\nu] \le s-1$; the condition on f will be $\| \bar{D}^s f \|_q < \infty$.

Note also, that the bound obtained for the bias is valid for any stationary sequence (X_n) without mixing condition.

We give now the proof of the theorems 1, 2, 3.

Proof of theorem 1.

$$(15)\qquad || x^k f_N(x) - x^k f(x) ||_q^p \le 2^{p-1} [|| x^k f_N(x) - x^k Ef_N(x) ||_q^p + || x^k Ef_N(x) - x^k f(x) ||_q^p]$$

The second term in the right of (15) is $O(h^{2p})$ by lemma 3.

To bound the first term we use the following result (see Weil [18]) :

Let $2 \le q \le \infty$, $q' = \frac{q}{q-1}$ and $F \in L^{q'}(\mathbb{R}^d) \cap L^1(\mathbb{R}^d)$, then :

$$|| \hat{F} ||_q \le || F ||_{q'}$$

This result is well known for $q = \infty$; for $q = 2$ we have the Parseval equality ; for $2 < q < \infty$ it is known as Titchmarsh Theorem.

Now we calculate the Fourier transform of $G = x^k f_N(x) - x^k Ef_N(x)$.

We have :

$$(16)\qquad \hat{f}_N(s) = \frac{1}{N} \sum_{j=1}^{N} e^{isX_j} \hat{K}(h_N s)$$

$$(17)\qquad (x^k f_N(x))^\wedge = (i)^{-[k]} D^k \hat{f}_N$$

$$(18)\qquad (x^k Ef_N(x))^\wedge = (i)^{-[k]} D^k ((\sqrt{2\pi})^d . \hat{f}(s) . \hat{K}(h_N s))$$

(Note that, by $\mathfrak{M}(k,p)$, $x^\nu f(x)$ is absolutely integrable for $\nu \le k$ and then the usual derivative $D^\nu \hat{f}$ is a well defined bounded function).

(16),(17) and (18) give :

$$(19)\qquad (x^k f_N(x))^\wedge = (i)^{-[k]} \sum_{\nu \le k} C_k^\nu \, D^\nu \left(\frac{1}{N} \sum e^{isX_j}\right) D^{k-\nu} (\hat{K}(h_N s))$$

$$(x^k Ef_N(x))^\wedge = (i)^{-[k]} \sum_{\nu\le k} C_k^\nu (\sqrt{2\pi})^d D^\nu(\hat{f}(s)) D^{k-\nu}(\hat{K}(h_N s)) \tag{20}$$

$D^{k-\nu}\hat{K}$ being absolutely integrable, it is easy to see that $\hat{G} = (x^k f_N(x) - x^k Ef_N(x)^\wedge \in L^1(\mathbb{R}^d) \cap L^\infty(\mathbb{R}^d)$ and then $\hat{G} \in L^1(\mathbb{R}^d) \cap L^{q'}(\mathbb{R}^d)$ because $1 \le q' < \infty$. We can apply the Titchmarsh theorem to $\hat{G}$.

Clearly, because $\hat{G} \in L^1(\mathbb{R}^d)$:

$$\hat{\hat{G}} = G \tag{21}$$

Then :

$$\|G\|_q^p \le \|\hat{G}\|_{q'}^p \quad . \tag{22}$$

It remains to bound the expectation of the term in the right of (22).

$$E\|G\|_q^p \le C \sum_{\nu\le k} C_k^\nu E\left(\int \left|D^\nu\left(\frac{1}{N}\sum e^{isX_j}\right) - (\sqrt{2\pi})^d D^\nu\hat{f}(s)\right|^{q'} \left|D^{k-\nu}(\hat{K}(h_N s))\right|^{q'} ds\right)^{p/q'} \tag{23}$$

(the constant C depends only on k and p).

We have two cases.

1. If $p/q' \ge 1$; we use a convexity inequality and obtain :

$$E\|G\|_q^p \le C \sum_{\nu\le k} C_k^\nu A_\nu^{p/q'} E\int \left|\frac{1}{N}\sum (D^\nu e^{isX_j} - (\sqrt{2\pi})^d D^\nu\hat{f})\right|^p A_\nu^{-1} \left|D^{k-\nu}(\hat{K}(h_N s))\right|^{q'} ds \tag{24}$$

with $A_\nu = \int |D^{k-\nu}(\hat{K}(h_N s))|^{q'} ds$.

Now :

$E(D^\nu e^{isX_j}) = (\sqrt{2\pi})^d D^\nu\hat{f}(s)$ and for $\delta > 0$ and $\beta = \mathrm{Max}(p,2)$

$E|D^\nu e^{isX_j}|^{\beta+\delta} = E|X_j^\nu e^{isX_j}|^{\beta+\delta} \le E|X_j|^{[k](\beta+\delta)}$

Note also that the mixing coefficients of the processus $(X_j^\nu e^{isX_j})$ are smaller than those of (X_j). We can then, by the $\mathfrak{M}(k,p)$ hypothesis, apply lemma 1. We obtain

$$(25) \qquad E\left|\sum (D^\nu e^{isX_j} - (\sqrt{2\pi})^d D^\nu \hat{f}(s))\right|^p \le C\, N^{p/2}$$

where C depends only on d, the strong mixing coefficients of (X_j) and $E|X_j|^{[k](\beta+\delta)}$.

So that (24) becomes :

$$(26) \qquad E\|G\|_q^p \le \frac{C'}{N^{p/2}} \sum_{\nu\le k} C_k^\nu A_\nu^{p/q'}$$

but
$$A_\nu^{p/q'} = \left[h_N^{q'[k-\nu]-d} \int |D^{k-\nu}\hat{K}(s)|^{q'} ds \right]^{p/q'}$$

$$\le C\, h_N^{-\frac{dp}{q'}} \|D^{k-\nu}\hat{K}\|_{q'}^p ,$$

where $\|D^{k-\nu}\hat{K}\|_{q'} < \infty$ because, by the $\mathcal{J}(k)$ hypothesis, $D^{k-\nu}\hat{K} \in L^1(\mathbb{R}^d) \cap L^\infty(\mathbb{R}^d)$.

2. If $p/q' < 1$, we use first the inequality $E|Z|^{p/q'} \le (E|Z|)^{p/q'}$ and the proof continues like for $p/q' \ge 1$.

Proof of theorem 2.

Here also the bias $\|x^k Ef_N(x) - x^k f(x)\|_q^p$ is $O(h^{2p})$ by lemma 3. But, because $1 \le q < 2$, the Titchmarsh theorem cannot be applied to bound $G = x^k f_N(x) - x^k Ef_N(x)$. However we can use the presence of the factor x^k.

$$(27) \qquad E\|G\|_q^p \le C\left[E\left(\int_{|x|\le 1} |G|^q \right)^{p/q} + E\left(\int_{|x|\ge 1} |G|^q \right)^{p/q} \right]$$

For the first term in the right of (27) we have :

$$(28)\qquad E\left(\int_{|x|\le 1}|G|^q\right)^{p/q} \le E\left[\left(\int_{|x|\le 1}|G|^2\right)^{p/2}\cdot\left(\int_{|x|\le 1}dx\right)^{p/q.\frac{2-q}{2}}\right]$$

$$\le C\ E||G||_2^p$$

For the second term, we take $r>0$, to be defined in the sequel, and we write :

$$(29)\qquad E\left(\int_{|x|\ge 1}|G|^q\right)^{p/q} = E\left(\int_{|x|\ge 1}|x|^r|G|^q|x|^{-r}dx\right)^{p/q}$$

$$\le E\left[\left(\int_{|x|\ge 1}|x|^{2r/q}|G|^2dx\right)^{p/2}\left(\int_{|x|\ge 1}|x|^{-r.\frac{2}{2-q}}\right)^{\frac{p}{q}.\frac{2-q}{2}}\right]$$

Now we want :

1. $|x|^{-r\frac{2}{2-q}}$ integrable on $|x|\ge 1$; this is true if $r>d.\frac{2-q}{2}$.

2. r/q is an integer.

The smaller r satisfying this two conditions is

$r = q(e+1)$ where e is the integer part of $\frac{d}{q}.\frac{2-q}{2}$.

With this r, (29) becomes :

$$(30)\qquad E\left(\int_{|x|\ge 1}|G|^q\right)^{p/q} \le C\sum_{j=1}^{d} E||x_j^{e+1}\ x^kf_N(x) - x_j^{e+1}x^kE\ f_N(x)||_2^p$$

and then using our hypothesis

$$E||G||_q^p = 0\left(\frac{1}{N^{p/2}\ h_N^{dp/2}}\right).$$

The theorem is proved.

Proof of theorem 3.

To bound $G = x^k f_N(x) - x^k Ef_N(x)$ we use the same method as in the proof of theorem 2. While to bound $E\|x^k Ef_N(x) - x^k f(x)\|_q^p$, we write :

$$E\| x^k Ef_N(x) - x^k f(x)\|_q^p \le C\left[E\left(\int_{|x|\le 1} |x^k Ef_N(x) - x^k f(x)|^q dx\right)^p + \right.$$

$$\left. + E\left(\int_{|x|\ge 1} |x^k Ef_N(x) - x^k f(x)|^q dx\right)^p \right]$$

The first term in the right is bounded by

$$E\| x^k Ef_N(x) - x^k f(x)\|_1^{p.q} \cdot \left(\int_{|x|\le 1} dx\right)^{p(1-q)}$$

For the second we introduce $r > 0$ and we write :

$$E\left(\int_{|x|\ge 1} |x^k Ef_N(x) - x^k f(x)|^q dx\right)^p = E\left(\int_{|x|\ge 1} |x|^r |x^k Ef_N(x) - x^k f(x)|^q |x|^{-r} dx\right)^p$$

$$\le E\left[\left(\int_{|x|\ge 1} |x|^{r/q} |x^k Ef_N(x) - x^k f(x)| dx\right)^{p.q} \cdot \left(\int_{|x|\ge 1} |x|^{-r.\frac{1}{1-q}} dx\right)^{p(1-q)}\right]$$

we take $r = q(e'+1)$ where e' is the integer part of $\frac{d}{q} \cdot (1-q)$ and we conclude as in theorem 2.

Remark 5.

It can be seen in the proofs of theorems 1, 2, 3, that the rate of convergence of $E\| x^k f_N(x) - x^k Ef_N(x)\|_q^p$ is valid without any smothing hypothesis on f. The condition $\mathfrak{D}(k,q)$ is used only to bound the bias.

Remark 6.

In theorem 1 the optimal choice of h_N is $h_N = N^{-\frac{q'}{2(2q'+d)}}$ and then the rate of convergence is $N^{-\frac{pq'}{2q'+d}}$; so that if $p > \frac{2q'+d}{q'}$ we have with probability one the convergence of $\| x^k f_N(x) - x^k f(x) \|_q$ to zero.

Similar remark can be down in theorems 2 and 3.

To conclude this section we discuss the Devroye-Györfi's condition.

Proposition 1. *Let* $1 \leq q \leq \infty$, $F \in L^1(\mathbb{R}^d)$, $\nu \in \mathbb{N}^d$ *and* ϕ *a positive density of class* $C^{[\nu]}$ *and with compact support in* $\mathbb{R}^d$. *Put* $\phi_a(x) = \frac{1}{a^d}\phi(\frac{x}{a})$ *; then :*

$$\lim_{a\to 0} \operatorname{Inf} \| D^\nu(\phi_a * F) \|_q = \sup_{a>0} \| D^\nu(\phi_a * F) \|_q = \| \overline{D}^\nu F \|_q \tag{31}$$

Proof.

Clearly $\phi_a * F$ is of class $C^{[\nu]}$ and then

$$D^\nu(\phi_a * F) = \overline{D}^\nu(\phi_a * F) = \phi_a * \overline{D}^\nu F. \tag{32}$$

Let $\Psi \in C_c^\infty(\mathbb{R}^d)$ such that $\|\Psi\|_{q'} \leq 1$.

We have (see [17]) :

$$<\phi_a * \overline{D}^\nu F, \Psi> = \int \phi_a(y) <\overline{D}^\nu F, \Psi_y> dy \tag{33}$$

where $\Psi_y(x) = \Psi(x+y)$.

We have $\|\Psi_y\|_{q'} \leq 1$ and then

$$|<\overline{D}^\nu F, \Psi_y>| \leq \| \overline{D}^\nu F \|_q \tag{34}$$

(33) and (34) give :

(35) $$|<\phi_a * \overline{D}^\nu F, \Psi>| \leq ||\overline{D}^\nu F||_q$$

so that the following inequalities hold always :

(36) $$\lim_{a\to 0} \mathrm{Inf} ||D^\nu(\phi_a * F)||_q \leq \sup_{a>0} ||D^\nu(\phi_a * F)||_q \leq ||\overline{D}^\nu F||_q$$

Clearly, if $\lim\limits_{a\to 0} \mathrm{Inf}||D^\nu(\phi_a * F)||_q = \infty$, we have (31). Let us assume that $\lim\limits_{a\to 0} \mathrm{Inf}||D^\nu(\phi_a * F)||_q = \rho < \infty$. There is a sequence a_j such that :

$$\lim_{a_j\to 0} ||D^\nu(\phi_{a_j} * F)||_q = \rho.$$

Now $\sup\limits_{a_j} ||D^\nu(\phi_{a_j} * F)||_q < \infty$ and then, by the compactness of the unit ball in the weak $*$ topology (see for example Brezis [2]), there exists a subsequence a'_j and S such that :

1. $S \in L^q(\mathbb{R}^d)$ if $q > 1$ or S is a bounded regular measure if $q = 1$.
2. $\lim\limits_{a'_j} D^\nu(\phi_{a'_j} * F) = S$ for the weak $*$ topology in $L^q(\mathbb{R}^d)$ if $q > 1$ and in the space of bounded regular measure if $q = 1$.
3. $||S||_q \leq \rho$.

Particularly, for $\Psi \in C_c^\infty(\mathbb{R}^d)$ we have

(37) $$\lim_{a'_j\to 0} <D^\nu(\phi_{a'_j} * F), \Psi> = <S, \Psi>$$

But as distribution :

$$\lim_{a'_j\to 0} (\phi_{a'_j} * \overline{D}^\nu F) = \overline{D}^\nu F$$

(because $\operatorname{supp} \phi_{a'_j} \subset \operatorname{supp} \phi_{a_o}$ where $a_o = \sup a'_j$; and in this case the convolution is continuous [17]).

This implies that $S = \overline{D}^\nu F$ and then :

(38) $$\|\overline{D}^{\nu}F\|_q \leq \lim_{a\to 0} \mathrm{Inf}\|D^{\nu}(\phi_a * F)\|_q$$

(38) and (36) give (31).

Remark 7.

A profound study of the conditions $\|\overline{D}^{\nu}F\|_q < \infty$ can be found in [17] ; in the one dimensional case, we have :

For $q > 1$ (by theorems III, VII of Chap. II [17]), $\|\overline{D}^2F\|_q < \infty$ if and only if F is C^1, F' is absolutely continuous with almost everywhere derivative $F'' = \overline{D}^2F$ and $F'' \in L^q(\mathbb{R}^d)$.

For $q = 1$ (by theorems II, III of Chap. II [17]), $\|\overline{D}^2F\|_1 < \infty$ if and only if F is absolutely continuous and its almost everywhere derivative F' is a function with bounded variation.

This last result precises exactly the class of functions satisfying the Devroye-Györfi's condition in the one dimensional case.

Remark 8.

Devroye-Penrod [4] and Devroye-Györfi [3] introduce the class $\mathcal{F}$ of functions f satisfying

(i) f is absolutely continuous with a.e. derivative f' ;

(ii) f' is absolutely continuous with a.e. derivative f''.

For such functions $f'' = \overline{D}^2f$ (see [17]).

We go now to the applications.

4. APPLICATIONS

4.1. Regression and conditional density

Let $(Y_n)_{n\in\mathbb{Z}}$ be a sequence of $\mathbb{R}^m$-valued random variables such that $(W_n)_{n\in\mathbb{Z}} = (Y_n, X_n)_{n\in\mathbb{Z}}$ is a stationary strong mixing process. Denote by $M(y,x)$ the density of W_n and put :

$$B(x) = \int y\ M(y,x)\,dy\ ; \qquad r(x) = \frac{B(x)}{f(x)}\ ;$$

$$s(y,x) = \frac{M(y,x)}{f(x)}\ ; \quad g(y,x) = s(y+r(x),x).$$

$B(x)$ and $r(x)$ are vector valued functions ; $r(x)$ is the regression function of Y_n on X_n, $s(y,x)$ is the conditional density of Y_n given $X_n = x$; the function $g(y,x)$ comes from the model :

$$Z_{n+1} = \phi(Z_n) + \varepsilon_{Z_n,n} \qquad \text{(cf. [8])}$$

in this model, if we put $(Z_{n+1},Z_n) = (Y_n,X_n)$ then the probability density of $\varepsilon_{x,n}$ is $s(y+\phi(x),x)$. Let now, H be a positive bounded kernel satisfying (1), (2) and (3) ; we estimate M, B, r, s, g respectively by :

$$M_N(y,x) = \frac{1}{N h_N^{m+d}} \sum_{j=1}^{N} H\left(\frac{y-Y_j}{h_N}\right) K\left(\frac{x-x_j}{h_N}\right)\ ; \quad B_N(x) = \frac{1}{N h_N^{d}} \sum_{j=1}^{N} Y_j\ K\left(\frac{x-x_j}{h_N}\right)$$

$$r_N(x) = \frac{B_N(x)}{f_N(x)}\ ; \qquad s_N(y,x) = \frac{M_N(y,x)}{f_N(x)}\ ; \qquad g_N(y,x) = s_N(y+r_N(x),x).$$

To study r_N, s_N and g_N we need results on B_N. We observe that :

(39) $$EB_N = K_h \star B$$

($K_h \star B$ is the vector valued function obtained when each component of B is convoled with K_h).

(40) $$\hat{B}_N(s) = \frac{1}{N} \sum_{j=0}^{N} Y_j\ e^{isX_j} . \hat{K}(h_N s)$$

(41) $$\widehat{EB}_N(s) = (K_h \star B)^\wedge = \hat{K}(h_N s).\hat{B}(s)$$

(42) $$E\ Y_j\ e^{isX_j} = (\sqrt{2\pi})^d\ \hat{B}(s)$$

So that, the method developped in section 3, can be applied to B_N. We do it for $q \geq 2$.

Proposition 2. *Let* $0 < p < \infty$, $2 \le q \le \infty$, $q' = \frac{q}{q-1}$ *and* $k \in \mathbb{N}^d$. *Assume that :*

i) $(W_n)_{n\in\mathbb{Z}}$ *satisfies* $\mathfrak{M}([k+1], p)$

ii) B *satisfies* $\mathfrak{D}(k,q)$;

iii) K *satisfies* $\mathfrak{J}(k)$.

Then :

$$E\| x^k B_N(x) - x^k B(x)\|_q^p = 0\left(\frac{1}{N^{p/2} h_N^{dp/q'}} + h_N^{2p}\right)$$

We study now r_N , p_N and g_N. For this we introduce the following hypothesis.

(H1) $\exists \gamma > 0$ such that $E \exp \gamma |Y_n| < \infty$

(H2) $\limsup (N h_N^{2d})^{-1} (\operatorname{Log} N)^2 < \infty$ and $\limsup h_N^2 \operatorname{Log} N < \infty$.

(H3) $\limsup (h_N^{m+d+1} N^{1/2})^{-1} < \infty$.

We need also the following lemma :

Lemma 4. *Let* (H1) *be satisfied ; then for* $p > 0$

$$E\| r_N\|_\infty^p = 0(\log N)^p$$

Proof.

Clearly $\| r_N\|_\infty^p \le (\operatorname{Max}_j |Y_j|)^p \quad j = 1,\dots,N.$

Let us put $Y = \gamma \operatorname{Max}_j |Y_j|$;

$$E\, Y^p = E(\log \exp Y)^p$$

(Note that $\exp Y \ge 1$ because $Y \ge 0$).

Let now $h(t) = (\log t)^p$ defined for $t \ge 1$.

h is concave on $t \ge 1$ if $p \le 1$ and on $t \ge e^{p-1}$ if $p > 1$.

If $p \le 1$, by concavity we have :

(43) $$E(Y^p) \le (\log E \exp Y)^p$$

$$\le (\log E(\sum_j \exp \gamma|Y_j|))^p$$

$$\le (\log N + \log E \exp \gamma|Y_1|)^p$$

and then $$E||r_N||_\infty^p = 0(\log N)^p$$

If $p > 1$, we have :

(44) $$E|Y|^p = E\, 1_{Y<p-1} Y^p + E\, 1_{Y\ge p-1} Y^p$$

$$\le (p-1)^p + (\log E \exp Y)^p$$

$$\le (p-1)^p + 0(\log N)^p$$

Now we state the results for r_N and s_N.

Theorem 4. *Let* $0 < p < \infty$ *and* $G \subset \mathbb{R}^d$ *such that* $\operatorname{Inf}_G f > 0$, $\sup_G B < \infty$. *Assume that :*

i) (H1) *and* (H2) *are satisfied ;*

ii) $(W_n)_{n\in\mathbb{Z}}$ *satisfies* $\mathfrak{M}(1,p_1)$ *where* $p_1 = 4p$;

iii) f *and* B *satisfy* $\mathfrak{D}(0,\infty)$;

iv) K *satisfies* $\mathfrak{J}(0)$.

Then

$$E \sup_G |r_N(x) - r(x)|^p = 0\left(\frac{1}{N^{p/2} h_N^{dp}} + h_N^{2p}\right)$$

Theorem 5. *Let* $0 < p < \infty$ *and* $G' \times G \subset \mathbb{R}^m \times \mathbb{R}^d$ *such that* $\operatorname{Inf}_G f > 0$, $\sup_{G'\times G} M < \infty$. *Assume that :*

i) (H3) *is satisfied ;*

ii) $(W_n)_{n\in\mathbb{Z}}$ *satisfies* $\mathfrak{M}(0,p_1)$ *where* $p_1 = (m+1)p$;

iii) M *and* f *satisfy* $\mathfrak{D}(0,\infty)$;

iv) K *satisfies* $\mathfrak{J}(0)$.

Then :

$$E \sup_{G'\times G} |s_N - s|^p = O\left(\frac{1}{N^{p/2} h_N^{(d+m)p}} + h_N^{2p}\right)$$

Proof of theorems 4 and 5.

We write for r_N :

$$(45) \qquad r_N - r = \frac{B_N - B}{f} + \frac{(B_N-B)(f_N-f)}{f^2} + \ldots + \frac{(B_N-B)(f-f_N)}{f^u} + B\,\frac{f-f_N}{f^2} + $$

$$+ \ldots + \frac{B(f-f_N)^{u-1}}{f^2} + r_N \frac{(f-f_N)^u}{f^u}$$

and a similar expansion for s_N with M and M_N in place of B and B_N.

We take $u = 2$ for r_N and $u = m+1$ for s_N. Note that $\sup s_N \le \frac{1}{h_N^m}$.

Next we take expectation of supremum and use Holder inequality. The results are then obtained by application of theorem 1 to $|f_N - f|$ and $|M_N - M|$ and proposition 2 to $|B_N - B|$.

For g_N we have.

Theorem 6. *Let* $0 < p < \infty$ *and* $G' \times G \subset \mathbb{R}^m \times \mathbb{R}^d$ *such that* $\operatorname*{Inf}_G f > 0$ *and* $\sup_{G'\times G} M < \infty$. *Assume that :*

i) M *is a Lipschitz function ;*

ii) (H1), (H2) *and* (H3) *are satisfied ;*

iii) $(W_n)_{n\in\mathbb{Z}}$ *satisfies* $\mathfrak{M}(1,p_1)$ *where* $p_1 = \mathrm{Max}(4p,(m+1)p)$;

iv) f, B *and* M *satisfy* $\mathfrak{D}(0,\infty)$;

v) K *satisfies* $\mathfrak{J}(0)$.

Then :

$$E \sup_{G' \times G} |g_N - g|^p = 0\left(\frac{1}{N^{p/2} h_N^{(m+d)p}} + h_N^{2p}\right)$$

Proof.

Clearly, because $g_N(y,x) = \dfrac{M_N(y+r_N(x),x)}{f_N(x)}$, $g_N - g$ can be written as in (45) with $M_N(y+r_N(x),x)$ and $M(y+r(x),x)$ in place of B_N and B. Here also we take $u = m+1$.

It suffices then to bound $|M_N(y+r_N(x),x) - M(y+r(x),x)|$.

But :

(46) $$|M_N(y+r_N(x),x)-M(y+r(x),x)|^p \le 2^{p-1}[|M_N(y+r_N(x),x)-M(y+r_N(x),x)|^p + |M(y+r_N(x),x)-M(y+r(x),x)|^p]$$

and then

(47) $$E \sup_{G' \times G} |M_N(y+r_N(x),x)-M(y+r(x),x)|^p \le 2^{p-1}[E||M_N-M||_\infty^p + C^p||r_N - r||_\infty^p]$$

where C is the Lipschitz coefficient of M ; we conclude by theorems 1 and 4.

4.2. Others functional estimates

The results obtained can be used to evaluate risks for some functional estimates. As example, we do it for the Matusita's measure of affinity between two density [7]. We keep our notations and denote by $h(y)$ the density of Y_n. The measure of affinity between h and f is defined by :

$$\rho = \int (h(y) f(y))^{1/2} dy \quad \text{when} \quad m = d.$$

$h(y)$ is estimated by :

$$h_N(y) = \frac{1}{N h_N^d} \sum_{j=1}^{N} H\left(\frac{y - Y_j}{h_N}\right)$$

I. Ahmad [1] proposed the following estimate of ρ :

$$\hat{\rho} = \int (h_N(y) f_N(y))^{1/2} dy.$$

Now we evaluate the risk for this estimate. It would be noted that we do not need as in [1] the independance between Y_n and X_n.

Theorem 7. *Let* $0 < a < \infty$ *; assume that* f, h, K *and* H *satisfy the conditions of theorem 2 for* $p = \frac{a}{2}$ *and* $q = 1$. *Then :*

$$E|\hat{\rho}-\rho|^a = 0\left(\frac{1}{N^{a/4} h_N^{da/4}} + h_N^a\right)$$

Proof.

$$E|\hat{\rho}-\rho|^a \leq 2^{a-1} E\left(\int f_N^{1/2} |h_N^{1/2} - h^{1/2}|\right)^a + 2^{a-1} E\left(\int h^{1/2} |f_N^{1/2} - f^{1/2}|\right)^a$$

$$\leq 2^{a-1} E\left[\left(\int f_N\right)^{a/2} \left(\int |h_N - h|\right)^{a/2}\right] + 2^{a-1} E\left[\left(\int h\right)^{a/2} \left(\int |f_N - f|\right)^{a/2}\right]$$

We conclude by theorem 2.

5. CONCLUSION

It would be noted that the method developed in this paper apply in the same line to evaluate the risks for the density derivatives estimates. These derivatives appear as dual of the quantities $x^k f(x)$.

REFERENCES

[1] AHMAD, I. (1980). *Non parametric estimation of the Matusita's measure of affinity between absolutely continuous distributions.* Ann. Inst. Stat. Math. 32, 241-245.

[2] BREZIS, H. (1983). *Analyse fonctionnelle. Théorie et applications.* Masson, Paris.

[3] DEVROYE, L. and GYÖRFI, L. (1985). *Non parametric density estimation. The* L_1 *view.* Wiley, New-York.

[4] DEVROYE, L. and PENROD, C.S. (1984). *Distribution-free lower bounds in density estimation.* Ann. Stat. 12, 1250-1262.

[5] DOUKHAN, P. LEON, J., PORTAL, F. *Principe d'invariance faible pour la mesure empirique d'une suite de variables aléatoires mélangeantes.* (To appear).

[6] GORODETSKII, V.V. (1977). *On the strong mixing property for linear sequences.* Theory Prob. Appl. 22, 411-413.

[7] MATUSITA, K. (1955). *Decision rules based on the distance for the problem of fit, two samples and estimation.* Ann. Math. Stat. 26, 631-640.

[8] MOKKADEM, A. (1985). *Le modèle non linéaire* AR(1) *général. Ergodicité et ergodicité géométrique.* C.R.A.S., t.301, 889-892.

[9] MOKKADEM, A. (1986). *Sur le mélange d'un processus* ARMA *vectoriel.* C.R.A.S., t.303, 519-521.

[10] MOKKADEM, A. (1985). *Etude des risques des estimateurs à noyaux.* C.R.A.S., t.301, 447-450.

[11] PARZEN, E. (1962). *On estimation of a probability density function and the mode.* Ann. Math. Stat. 33, 1065-1076.

[12] PHAM, D. and TRAN, L.T. (1986). *Some mixing properties of time series models.* Stoch. Proc. Appl. 19, 297-303.

[13] ROSENBLATT, M. (1956). *Remarks on some non parametric estimates of a density function.* Ann. Math. Stat. 27, 832-835.

[14] ROSENBLATT, M. (1956). *A central limit theorem and a strong mixing condition.* Proc. Nat. Acad. Sci., U.S.A. 42, 43-47.

[15] ROSENBLATT, M. (1979). *Global measure of deviation of kernel and nearest neighbor density estimates.* Lect. Notes Math. 757, 181-190.

[16] RUDIN, W. (1966). *Real and Complex Analysis.* Mc Graw-Hill, New-York.

[17] SCHWARTZ, L. (1966). *Théorie des distributions*. Hermann, Paris.

[18] WEIL, A. (1965). *L'intégration dans les groupes topologiques et ses applications*. Hermann, Paris.

[19] WHITHERS, C.S. (1981). *Conditions for linear processes to be strong mixing*. Z. Wahr. verw. Gebiete, 57, 481-494.

[20] YOKOYAMA, R. (1980). *Moment bounds for stationary mixing sequences*. Z. Wahr. verw. Gebiete, 52, 45-57.

C H A P I T R E V

ESTIMATION OF THE ENTROPY AND INFORMATION OF ABSOLUTELY CONTINUOUS RANDOM VARIABLES.

Abdelkader Mokkadem

U.A.743 CNRS- Statistique Appliquée
Batiment 425, Mathématiques
Université de Paris-Sud
91405 ORSAY Cedex-FRANCE

Abstract : *We propose an estimation of the entropy and the mutual information of absolutely continuous random vectors and we give an upper bound of the mean risks for the proposed estimators, under strong mixing conditions.*

Key words : Entropy ; information ; kernel estimators ; strong mixing.

1. INTRODUCTION.

Let X be a random vector with values in $\mathbb{R}^m$ and with probability density function $f(x)$. The entropy of X is defined by :

$$H(X) = -\int_{\mathbb{R}^m} f(x)\mathrm{Log}f(x)\, dx \ ;$$

we shall always assume that $f(x)\mathrm{Log}f(x)$ is integrable.

The first estimation of the entropy of a discrete random variable has been proposed and studied by G. P. Basharin [2] and later by K. Hutcheson and L. R. Shenton [7].

For a random variable with density, I. A. Ahmad and P. E. Lin [1] proposed the following approach for $m = 1$. Taking a random sample $X_1,\dots,X_N$ of X and a positive kernel $K(x)$, $f(x)$ can be estimated by:

$$f_N(x) = (1/Nh_N) \sum_{i=1}^{N} K((x-X_i)/h_N)$$

where h_N is a sequence of positive real numbers which converge to zero. Using this, they estimate $H(X)$ by

$$\hat{H}_N = -(1/N) \sum_{i=1}^{N} \mathrm{Log}f(X_i)$$

and study the first and second mean consistency of $\hat{H}_N$. However their proof is incomplete. More precisely they use systematicaly a Taylor expansion of the following form:

$$\mathrm{Log}\ z - \mathrm{Log}\ c = (z-c)/(\theta z-(1-\theta)c) \ , \quad 0 < \theta < 1 \ ;$$

but in their work z and c are functions of N and x and then θ is a random function of N and x. So θ must be integrated and they must study its limit for N tending to infinity.

A second approach is due to B. L. S. Prakasa Rao [9], who proposed the estimator

$$\hat{H}_N = -\int_{-k_N}^{k_N} f_N(x) \mathrm{Log} f_N(x)\ dx \ ,$$

where k_N is an increasing sequence of positive real numbers tending to infinity. There is also the work of L. Birgé [3], who studied a similar estimator where f_N is the empirical density estimator.

Another interesting approach is due to O. Vasicek [13], who uses his result for testing the normality of X. This approach is also used by E. J. Dudewicz and E. C. Vander Meulen [5] for testing the uniformity of X. Unfortunately, the method of O. Vasicek, based on order statistics, is particular to the one dimensional case and can not be used to estimate H(X) in the multidimensional case or to estimate the mutual information of two random variables.

Except in L. Birgé [3], there is nowhere an upper bound of the mean risks for the proposed estimators.

In what follows, we propose another approach based on the following well known result (see [6]). If μ is a probability measure on Ω and the positive function f belongs to $L^q(d\mu)$ for some $q > 0$, then

$$(1/r)\ \mathrm{Log} \int f^r d\mu \ \underset{r \downarrow 0}{\downarrow} \int \mathrm{Log} f\ d\mu \ .$$

Using this we estimate H(X) by

$$(1) \qquad H_N(X) = -(1/r_N)\ \mathrm{Log} \int_{\mathbb{R}^m} f_N^{r_N+1}(x)\ dx \ ,$$

where $r_N = h_N^\gamma$ for some fixed positive real number γ .Under some mixing conditions (see sections 2 and 3), we obtain the following result (Theorem 1) for $p > 0$ and $\gamma = 1$:

$$E|H_N(X) - H(X)|^p = O((1/N^{p/2} h_N^{(m+2)p/2}) + h_N^p) .$$

The same upper bound is obtained (Theorem 2) for the mean risks of the natural estimator

(2) $$\tilde{H}_N(X) = -\int_{\mathbb{R}^m} f_N(x) \operatorname{Log} f_N(x)\, dx .$$

As an application we give, in section 4, an estimator of the mutual information of two random variables .

2. NOTATIONS AND ASSUMPTIONS.

We introduce now some notations and assumptions. Let $(X_n)_{n\in\mathbb{Z}}$ be a strictly stationary sequence of random vectors in $\mathbb{R}^m$ .We denote by $\mathcal{Q}_0$ the σ-algebra generated by $\{X_n , n \leqslant 0\}$ and by $\mathcal{Q}^k$ those generated by $\{X_n , n \geqslant k\}$ for $k > 0$. The strong mixing coefficients of the process (X_n) are defined by

$$\alpha(k) = \sup \{ |P(A\cap B) - P(A).P(B)| , A\in\mathcal{Q}_0 , B\in\mathcal{Q}^k \} ,$$

(M. Rosenblatt [10]) . We always assume that the process (X_n) is strong mixing , i.e. $\lim \alpha(k) = 0$.

Let $f(x)$ be the probability density of X_n , (h_N) be a sequence of positive real numbers which converge to zero and K be a positive bounded kernel on $\mathbb{R}^m$ such that :

(3) $$\int K(x)\, dx = 1 , \int x\, K(x)\, dx = 0 , \int |x|^3 K(x)\, dx < \infty , \int |x|^2 f(x)\, dx < \infty$$

(4) $$\int |x_i|^2 K(x)\, dx = \tau^2, \int x_i x_j\, K(x)\, dx = 0 \quad \text{for } i \neq j \text{ and } x = (x_i)_{1\leqslant i\leqslant m}$$

(5) $\int f(x)(\mathrm{Log} f(x))^2\, dx < \infty$, $\int K(x)(\mathrm{Log} K(x))^2\, dx < \infty$,

$\sup f(x) = M < \infty$, $\sup K(x) = M' < \infty$

(6) $x^{\nu}K(x)$ and $D^{\nu}\hat{K}(s)$ are absolutely integrable for $\nu=(\nu_1,..,\nu_m) \in \mathbb{N}^m$, $[\nu] \leqslant a$; (where $\hat{K}$ is the Fourier transform of K , $[\nu]=\nu_1+...+\nu_m$ and a is the integer part of $(m/2)+1$) .

For the next assumption we need some definition; let $\mathcal{D}(\mathbb{R}^m)$ be the space of C^{∞} functions with compact support in $\mathbb{R}^m$ and T be a distribution on $\mathbb{R}^m$ (see [12]), we put

$\|T\|_1 = \sup \{ |T(\varphi)| ; \varphi \in \mathcal{D}(\mathbb{R}^m) , \|\varphi\|_{\infty} \leqslant 1 \}$

Clearly, $\|T\|_1 < \infty$ if and only if T is a bounded regular measure ; particularly, if $T \in L^1(\mathbb{R}^m)$ then $\|T\|_1$ is the usual L^1-norm .

Now our assumption is

(7) $\|\bar{D}^{\nu}f\|_1 < \infty$ for $\nu \in \mathbb{N}^m$, $[\nu] = 2$; (here and in the sequel the distributional derivatives are noted $\bar{D}^{\nu}f$) .

It is proved in [8] that (7) is equivalent to the condition used by L. Devroye and L. Györfi [4] in the case m = 1 :

$$\sup_{a>0} \int |(f*\varphi_a)''| < \infty$$

where φ is a positive density of class C^2 with compact support and $\varphi_a(x) = (1/a)\varphi(x/a)$.

Following M. Rosenblatt [11] , we estimate f(x) by

(8) $$f_N(x) = (1/Nh_N^m) \sum_{i=1}^{N} K((x-X_i)/h_N)$$

and $H(X_n)$ by $H_N(X)$ and $\tilde{H}_N(X)$ as defined in (1) and (2) .

We shall use also the following assumption with $0 < p < \infty$.

$\mathcal{M}(p)$: For some d>0, $E|X_n|^{a(\beta+d)} < \infty$ and $\sum_{k=0}^{\infty}(k+1)^{(\beta/2)-1}\alpha(k)^{d/(\beta+d)} < \infty$; with $\beta = \mathrm{Max}(p , 2)$ and a defined in (6) .

3. THE MEAN RISKS .

We need the following lemma whose proof is in [8].

Lemma 1 . If $\mathcal{M}(p)$ is satisfied ,$0 < p < \infty$, then

$$E\|f_N - f\|_1^p \leqslant (c_1/N^{p/2}h_N^{mp/2}) + c_2 h_N^{2p}$$

with $$c_2 = \left((\tau^2/2) \sum_{[\nu]=2} \|\bar{D}^{\nu} f\|_1\right)^p$$

and $$c_1 = c_0 \sum_{i=0}^{m} \sum_{j=0}^{a} \begin{pmatrix} j \\ a \end{pmatrix} h_N^{(a-j)p} \|x_i^{a-j}\ K(x)\|_2^p$$

where the constant c_0 depends only on m, p, $\alpha(k)$ and $E|X_n|^{a(\beta+d)}$.

Now we can state our first result ; we take for simplicity $r_N = h_N$.

Theorem 1 . If $\mathcal{M}(2p)$ is satisfied , $0 < p < \infty$, then

$$E|H_N(X) - H(X)|^p = O((1/N^{p/2}h_N^{(m+2)p/2}) + h_N^p) .$$

Proof : we shall write for simplicity H_N , H and $\int f$ instead of $H_N(X)$, $H(X)$ and $\int f(x)\ dx$. We have

(9) $$|H_N - H|^p \leqslant c \left(|H_N + (1/r_N)\mathrm{Log}\int f^{r_N+1}|^p + |(1/r_N)\ \mathrm{Log}\int f^{r_N+1} + H|^p\right).$$

where c is 2^{p-1} if $p \geqslant 1$ and 1 if $p < 1$.

We begin with the first term in the right side of (9) :

(10) $$|H_N + (1/r_N)\mathrm{Log}\int f^{r_N+1}| \leqslant (1/r_N\theta)|\int f_N^{r_N+1} - \int f^{r_N+1}|$$

where

(11) $$\theta \geqslant \mathrm{Min}\ (\int f_N^{r_N+1}\ ,\ \int f^{r_N+1}\ )$$

and then

(12) $$\theta \geqslant \mathrm{Min}\ (\ e^{-r_N\tilde{H}_N}\ ,\ e^{-r_N H}\ )$$

so that

(13) $(1/\theta) \leq e^{r_N\tilde{H}_N} + e^{r_N H}$.

We must now bound $\tilde{H}_N$ and H .

For this we denote by V′ the transpose of a column vector V and respectively by D and G_N the matrices

$$\mathrm{Var}X_n \quad \text{and} \quad \int(xx')f_N(x)dx - \left(\int xf_N(x)dx\right)\left(\int xf_N(x)dx\right)' .$$

Noting that $\int(xx')K(x) = \tau^2 Id$, it is easy to obtain

(14) $G_N = \tau^2 h_N^{2m} Id + (1/N) \sum_{i=1}^{N} (X_i - \overline{X})(X_i - \overline{X})'$.

Now we use the fact that, when the random vector Z has a fixed covariance matrix , the maximum of H(Z) is attained for Z gaussian . We obtain

(15) $(1/\theta) \leq (2\pi e)^{mr_N/2} \left(|\det G_N|^{r_N/2} + |\det D|^{r_N/2}\right)$.

It is easily seen that $E|\det G_N|^{r_N/2}$ is bounded by $m!(2E|X_n|^2 + h_N^{2m}\tau^2)^{mr_N/2}$.

Now we bound the term $|\int f_N^{r_N+1} - \int f^{r_N+1}|$ in the right side of (10);we have

(16) $$\left|\int f_N^{r_N+1} - \int f^{r_N+1}\right| \leq (r_N+1)\int(f_N^{r_N} + f^{r_N})|f_N - f|$$

$$\leq (r_N+1)[(M'/h_N)^{r_N} + M^{r_N}]\int|f_N - f|$$

$$= d_N \int|f_N - f|$$

where $d_N = (r_N+1)[(M'/h_N^m)^{r_N} + M^{r_N}]$ and satisfies lim $d_N = 1$.

Combining (10) , (15) and (16) we obtain

(17) $|H_N + (1/r_N)\mathrm{Log}\int f^{r_N+1}| \leq (d_N'/r_N)\left(|\det G_N|^{r_N/2} + |\det D|^{r_N/2}\right)\int|f_N - f|$

where

(18) $d_N' = (2\pi e)^{mr_N/2} d_N$.

Now we study the second term in the right side of (9)

$$|(1/r_N)\mathrm{Log}\int f^{r_N+1} - \int f\mathrm{Log}f| \ .$$

Using a Taylor expansion of $e^{r_N \mathrm{Log}f}$, we obtain

$$(19) \quad (1/r_N)\mathrm{Log}\int(e^{r_N\mathrm{Log}f} f) = (1/r_N)\ \mathrm{Log}\ \int[(1+r_N\mathrm{Log}f+r_N^2 e^{r_0\mathrm{Log}f}(\mathrm{Log}f)^2)f]$$

$$\leqslant (1/r_N)\ \left[r_N\int f\mathrm{Log}f + r_N^2\int[e^{r_0\mathrm{Log}f} f(\mathrm{Log}f)^2]\right]$$

where $0 < r_0 < r_N$ (r_0 is function of x) .

Now because $(1/r_N)\mathrm{Log}\int f^{r_N+1}$ is greater than $\int f\mathrm{Log}f$, we obtain

$$(20) \quad |(1/r_N)\mathrm{Log}\int f^{r_N+1} + H| \leqslant r_N\int[f^{r_0+1}(\mathrm{Log}f)^2]$$

$$\leqslant r_N\ M^{r_N}\int f(\mathrm{Log}f)^2 \ .$$

Using now (17) and (20) , (9) becomes

$$(21) \quad |H_N - H|^p \leqslant c\ [(A_N^p/r_N^p)\ (\int|f_N - f|)^p + B_N^p r_N^p\ ]$$

where

$$A_N = d_N'(|\det G_N|^{r_N/2}+|\det D|^{r_N/2})\ ;\ B_N = M^{r_N}\int f(\mathrm{Log}f)^2 \text{ and } \lim B_N = 1 \ .$$

Taking the expectation in (21) and using Schwarz's inequality, we obtain

$$(22) \quad E|H_N - H|^p \leqslant c\ \left[(1/r_N^p)(EA_N^{2p})^{1/2}[E(\int|f_N-f|)^{2p}]^{1/2} + B_N^p r_N^p\ \right] \ .$$

whith $EA_N^{2p} \leqslant 2c^2 d_N'\left(m!(2E|X|^2 + h_N^{2m}\ \tau^2\ )^{pmr_N} + |\det D|^{pr_N}\right)$ when $pmr_N \leqslant 1$.

We apply lemma 1 to conclude .

<u>Remark 1</u> . Note that in (10) , θ is bounded by $(M_N'/h_N)^{r_N} + M^{r_N}$ and then

$$|H_N + (1/r_N)\mathrm{Log}\int f^{r_N+1}| \sim (1/r_N)|\int f_N^{r_N+1} - \int f^{r_N+1}| \ .$$

We can obtain a similar result for $\tilde{H}_N$. For this , we assume that for some fixed $s < 1$, the following conditions are satisfied for $k \in \mathbb{N}^m$, $[k] \leqslant (m(1-s)/s) + 1$.

(23) $\int f^s < \infty$

(24) $\|\overline{D}^\nu (x^k f(x)\|_1 < \infty$ for $\nu \in \mathbb{N}^m$, $[\nu] = 2$

(25) $\|\overline{D}^\nu (x^{k-\beta} f(x))\|_1 < \infty$ for $\beta \in \mathbb{N}^m$, $\nu \in \mathbb{N}^m$, $[\beta] = [\nu] = 1$.

We have then the following lemma .

Lemma 2 . If (24) and (25) are satisfied , then $\int |Ef_N - f|^s = O(h_N^{2s})$.

We can now state our result for $\tilde{H}_N(X)$.

Theorem 2 . If (23),(24),(25) and $\mathcal{M}(2p)$ are satisfied , $0 < p < \infty$, then

$$E|\tilde{H}_N(X) - H(X)|^p = O((N^{p/2} h_N^{(m+2)p/2} + h_N^p) .$$

Proof. We must evaluate $E|\tilde{H}_N - H_N|$; but the same expansion as (19) and (20) gives

$$(26)\quad |H_N - \tilde{H}_N| \leqslant r_N \int f_N^{r_0+1} (\mathrm{Log} f_N)^2$$

$$\leqslant r_N \left(M'/h_N^m\right)^{r_N} \int f_N (\mathrm{Log} f_N)^2$$

and then

$$(27)\quad E|H_N - \tilde{H}_N|^p \leqslant \left(M'/h_N^m\right)^{pr_N} . r_N^p \, E\left(\int f_N (\mathrm{Log} f_N)^2 \right)^p .$$

Now we consider two cases . Case 1 : $p \geqslant 1$. Then

$$(28)\quad E\left(\int f_N (\mathrm{Log} f_N)^2 \right)^p \leqslant E \int f_N (\mathrm{Log} f_N)^{2p}$$

$$\leqslant E\left(\int_{f_N \leqslant R} f_N (\mathrm{Log} f_N)^{2p} + \int f_N^2 \right)$$

where R is such that

(29) $\text{Log } u < u^{1/2p}$ for $u \geqslant R$;

and because $u^{1-s}(\text{Log } u)^{2p}$ is bounded on $[0 , R]$, we obtain

(30) $E\left(\int f_N(\text{Log} f_N)^2 \right)^p \leqslant B.E\int f_N^s + E\int f_N^2$

where

(31) $B = \sup\limits_{0 \leqslant u \leqslant R} u^{1-s}(\text{Log } u)^{2p}$

Now it is well known that $E\int f_N^2$ is bounded . Also

(32) $$E\int f_N^s = \int Ef_N^s$$
$$\leqslant \int (Ef_N)^s$$
$$\leqslant \int |Ef_N - f|^s + \int f^s$$

and we conclude by lemma 2 .

Case 2 : $0 < p < 1$. Then

(33) $E\left(\int f_N(\text{Log} f_N)^2 \right)^p \leqslant \left(E\int f_N(\text{Log} f_N)^2 \right)^p$

and this quantity is bounded as in the case 1 .

<u>Remark 2</u> . The optimal choice of h_N is $N^{-1/(m+4)}$ and the p-mean risk of our estimators is then $N^{-p/(m+4)}$.

4. ESTIMATION OF THE MUTUAL INFORMATION.

Let X and Y be random vectors in $\mathbb{R}^m$ and $\mathbb{R}^q$ respectively, with joint probability density function $g(x,y)$ and marginals $f(x)$ and $l(y)$. The mutual information of X and Y is defined by

$$(34)\quad I(X;Y) = \int_{\mathbb{R}^m\times\mathbb{R}^q} g(x,y)\,\mathrm{Log}\,\frac{g(x,y)}{f(x)l(y)}\,dx\,dy$$

and takes its values in $[0\ ,\ +\infty]$. We assume that

$$(35)\quad I(X;Y) < \infty\ .$$

To estimate $I(X;Y)$, we use the well known relationship

$$(36)\quad I(X;Y) = H(X) + H(Y) - H(X,Y)\ .$$

Let $(X_n,Y_n)_{n\in\mathbb{Z}}$ be a strictly stationary sequence. We denote also by $\mathcal{M}(p)$ the corresponding condition on the strong mixing coefficients of (X_n,Y_n). Let K' be a positive bounded kernel on $\mathbb{R}^q$. We assume that K', g and l satisfy the same properties as K and f ((3),...,(7)). Now we estimate g and l by

$$(37)\quad g_N(x,y)= (1/Nh_N^{m+q})\sum_{i=1}^{N}K(x-X_i)K'(y-Y_i) \text{ and } l_N(y)= (1/Nh_N^{q})\sum_{i=1}^{N}K'(y-Y_i)$$

and we take

$$(38)\quad H_N(Y) = (-1/r_N)\mathrm{Log}\int l_N^{r_N+1} \quad\text{and}\quad H_N(X,Y) = (-1/r_N)\mathrm{Log}\int g_N^{r_N+1}$$

$I(X;Y)$ is then estimated by

$$(39)\quad I_N(X;Y) = H_N(X) + H_N(Y) - H_N(X,Y)\ .$$

As consequence of theorem 1, we obtain

<u>Theorem 3</u>. If the sequence (X_n,Y_n) satisfy $\mathcal{M}(2p)$, $0 < p < \infty$, then

$$E|I_N(X;Y) - I(X;Y)|^p = O(N^{p/2}h_N^{(m+q+2)p/2} + h_N^p)\ .$$

Remark 3 . Based on the well known property that X and Y are independent if and only if $I(X;Y) = 0$, one can construct a consistent test of independence by rejecting the independence of X and Y if $I_N(X;Y) \geqslant \delta$ where δ is some small positive real number .

-References-

[1] I. A. Ahmad and P. E. Lin , *A nonparametric estimation of the entropy for absolutely continuous distributions* , IEEE Trans.Inf.Theory, Vol.I , T.22, 372-375, 1976 .

[2] G. P. Basharin , *On statistical estimate for the entropy of a sequence of independent random variables* , Theory Prob. Appl., Vol.4, 333-336, 1956 .

[3] L. Birgé , *Estimation de l'entropie d'une densité* , exposé du Séminaire d'Orsay, Paris 1985 .

[4] L. Devroye and L. Györfi , *Nonparametric density estimation. The L1 view* , Wiley, New-York, 1985 .

[5] E. J. Dudewicz and E. C. Vander Meulen , *Entropy-base tests of Uniformity* , J.Amer.Stat.Ass., Vol.76, N°376, 967-974, 1981 .

[6] G. H. Hardy, J. E. Littlewood and G. Polya , *Inequalities* , Cambridge University Press , 1967 .

[7] K. Hutcheson and L. R. Shenton , *Some moments of an estimate of Shannon's measure of information* , Comm. Stat., Vol.3, 89-94, 1974 .

[8] A. Mokkadem , Thèse doctorat d'état, Université Paris XI, Orsay, 1987

[9] B. L. S. Prakasa Rao , *Nonparametric Functional Estimation* , Academic Press , 1983 .

[10]M. Rosenblatt , *A central limit theorem and a strong mixing condition,* Proc. Nat. Acad. Sc., U. S. A., 42, 43-47, 1956 .

[11]M. Rosenblatt , *Remark on some non parametric estimates of a density function* , Ann. Math. Stat., 27, 832-835, 1956 .

[12]L. Schwartz , *Theorie des Distributions* , Hermann, Paris, 1966 .

[13]O. Vasicek , *A test for normality based on sample entropy* , J.Roy.Stat.Soc., Ser.B, 38, 54-59, 1976 .

C H A P I T R E VI

ENTROPIE DES PROCESSUS LINEAIRES

A. Mokkadem

- Résumé -

Soit $Y = (Y_n)_{n\in\mathbb{Z}}$ un processus réel stationnaire de représentation spectrale $Y_n = \int_{-\pi}^{\pi} e^{in\lambda}\, dZ_Y$ et $X = (X_n)_{n\in\mathbb{Z}}$ le processus défini par $X_n = \int_{-\pi}^{\pi} e^{in\lambda}\phi(\lambda)dZ_Y$; on montre que l'entropie de X satisfait l'inégalité :

$$H(X) \geq \frac{1}{2\pi}\int_{-\pi}^{\pi} \operatorname{Log}\, |\phi(\lambda)|d\lambda + H(Y) \ ;$$

divers cas d'égalité sont établis. Nous donnons également des applications de ce résultat.

- Abstract -

Let $Y = (Y_n)_{n\in\mathbb{Z}}$ be a real stationary process with spectral representation $Y_n = \int_{-\pi}^{\pi} e^{in\lambda}\, dZ_Y$ and $X = (X_n)_{n\in\mathbb{Z}}$ the process defined by $X_n = \int_{-\pi}^{\pi} e^{in\lambda}\phi(\lambda)dZ_Y$; we prove that the entropy of X satisfies :

$$H(X) \geq \frac{1}{2\pi}\int_{-\pi}^{\pi} \operatorname{Log}\, |\phi(\lambda)|d\lambda + H(Y) \ ;$$

some cases where equality holds are obtained. We give also some applications.

Mots-clés : Entropie d'un processus ; processus symétrique α-stable ; processus harmonisable ; filtre linéaire ; principe du maximum d'entropie.

Classification A.M.S. 1985 : 60 G 25 , 94 A 15.

1 - INTRODUCTION

Soit X une variable aléatoire à valeurs dans $\mathbb{R}^k$, admettant une densité de probabilité $f(x)$ telle que $f(x) \operatorname{Log} f(x)$ soit intégrable ; l'entropie de X est définie par :

$$(1) \qquad h(X) = -\int_{\mathbb{R}^k} f(x) \operatorname{Log} f(x)\, dx \;;$$

(cette définition sera étendue dans la section 2 à toute variable aléatoire admettant un moment absolu d'ordre $s > 0$ fini).

Pour un processus réel, stationnaire en loi, $X = (X_n)_{n\in\mathbb{Z}}$, l'entropie de X est définie par :

$$(2) \qquad H(X) = \lim_n \frac{1}{n+1} h(X_o,\dots,X_n) \;.$$

Considérons maintenant un processus stationnaire en loi, réel et harmonisable (voir section 2) $Y = (Y_n)_{n\in\mathbb{Z}}$ et $X = (X_n)_{n\in\mathbb{Z}}$ filtré de Y par un filtre linéaire réel :

$$(3) \qquad X_n = \int_{-\pi}^{\pi} e^{in\lambda}\, \phi(\lambda)\, dZ_Y(\lambda)$$

où dZ_Y désigne la mesure spectrale de Y, $\phi(\lambda)$ est intégrable par rapport à dZ_Y et vérifie

$$(4) \qquad \phi(\lambda) = \overline{\phi(-\lambda)}$$

cette dernière condition assurant que X est réel.

On établit dans cet article le résultat suivant qui généralise des résultats obtenus par M. Kanter [9] dans le cas où les Y_n sont indépendants et par L.A. Shepp, D. Slepian et A.D. Wyner [15] dans le cas où ϕ est un polynôme trigonométrique :

$$H(X) \geq \frac{1}{2\pi}\int_{-\pi}^{\pi} \operatorname{Log}|\phi(\lambda)|d\lambda + H(Y) \ ; \tag{5}$$

l'égalité dans (5) est obtenue dans différents cas, en particulier quand Y est régulier ou symétrique α-stable, ce qui permet d'écrire l'entropie de Y en fonction de son spectre et de l'entropie de son innovation. Un tel résultat a été établi par C.E. Shannon [14] dans le cas gaussien en faisant directement le calcul de $h(X_o,\ldots,X_n)$ et en passant à la limite ; notre méthode n'utilise pas ce calcul explicite ; elle permet en particulier d'avoir une nouvelle interprétation du premier théorème de Szëgo :

$$\lim_n |\det Q_{n+1}|^{1/n+1} = \exp\left\{\frac{1}{2\pi}\int_{-\pi}^{\pi} \operatorname{Log} f(\lambda)d\lambda\right\} \tag{6}$$

où Q_{n+1} est la matrice des covariances de $(X_o,\ldots,X_n)$ et f la densité spectrale de $X = (X_n)_{n\in\mathbb{Z}}$.

Le plan de l'article est le suivant : la section 2 est constituée de préliminaires qu'on utilise dans la suite ; on établit notre résultat principal dans la section 3 ; des applications et une extension au cas des processus n'ayant pas de représentation spectrale sont données dans la section 4 ; on discute en particulier dans cette section du principe du maximum d'entropie introduit par J.P. Burg [2] , [3] .

Tous les processus que nous considérerons seront supposés centrés.

2 - PRELIMINAIRES ET NOTATIONS

Nous introduisons d'abord quelques propriétés de l'information et de l'entropie de variables aléatoires.

Définition 2.1.

Soient P_1 et P_2 deux mesures de probabilité définies sur un même espace mesurable $(\Omega, \mathcal{A})$; on définit l'information de Kullback de P_1 et P_2 par :

(7) $$K(P_1, P_2) = \sup \sum_i P_1(E_i) \operatorname{Log} \frac{P_1(E_i)}{P_2(E_i)}$$

où le supremum est pris sur toutes les partitions finies (E_i) de Ω .

$K(P_1, P_2)$ est à valeurs dans $[0, +\infty]$; elle est infinie si P_2 ne domine pas P_1 ; quand P_1 et P_2 admettent des densités de probabilité f_1 et f_2 , on a :

(8) $$K(P_1, P_2) = \int f_1 \operatorname{Log} f_1/f_2$$

Définition 2.2.

Soit (X,Y) un couple de variables aléatoires de loi de probabilité P_{XY} ; notons P_X (resp P_Y) la loi de probabilité de X (resp Y) . On définit l'information mutuelle de X et Y par :

(9) $$I(X,Y) = K(P_{XY}, P_X \otimes P_Y) .$$

Les propriétés suivantes sont établies dans M. Pinsker [10] :

(10) si P_{1n} et P_{2n} convergent respectivement vers P_1 et P_2 alors :

$$\liminf K(P_{1n}, P_{2n}) \geq K(P_1, P_2)$$

(11) $I(X,Y) = I(Y,X)$.

(12) $I(X,Y) = 0$ si et seulement si X et Y sont indépendantes.

(13) $I(X,X) = \infty$ si X est absolument continue.

(14) $I(X,Y) \geq I(f(X),Y)$ pour toute application mesurable f .

(15) Si $X = (X_1, X_2, \dots)$ alors :

$$I(X,Y) = \lim_n I((X_1,\dots,X_n),Y)$$

Nous passons maintenant à la définition de l'entropie d'une variable aléatoire ; pour une variable aléatoire X dans $\mathbb{R}^k$ admettant une densité de probabilité f et un moment absolu d'ordre $s > 0$ fini, l'intégrale

(16) $$h(X) = -\int f(x) \operatorname{Log} f(x) dx$$

est bien définie et à valeurs dans $[-\infty, +\infty[$; pour éviter d'avoir à vérifier à chaque étape l'existence d'une densité, nous poserons

(17) $$h(X) = -\infty \quad \text{si } X \text{ n'est pas absolument continue.}$$

Pour justifier (17) et rester dans le cadre du début de la présente section, on pose :

Définition 2.3

Si Q_u est une mesure de probabilité sur $\mathbb{R}^k$ de densité u et si la variable aléatoire X à valeurs dans $\mathbb{R}^k$ a pour loi de probabilité P et vérifie :

(18) $$E_P \{\operatorname{Log} u(X)\} > -\infty,$$

on pose :

(19) $$h(X) = -K(P,Q) - E_P\{\operatorname{Log} u(X)\} .$$

On a ainsi défini l'entropie pour toute variable aléatoire à valeurs dans $\mathbb{R}^k$, satisfaisant à une condition du type (18) ; cette entropie est à valeurs dans $[-\infty, +\infty[$, elle vaut $-\infty$ quand la variable aléatoire n'est pas absolument continue et est égale à l'entropie usuelle (16) sinon. La condition (18) est vérifiée par les variables aléatoires admettant un moment absolu d'ordre $s > 0$ en prenant

$$(20) \qquad u(x) = C\, e^{-|x|^s}$$

mais aussi par exemple par les variables aléatoires X vérifiant $E \operatorname{Log}(1+|X|^2) < \infty$ en prenant :

$$(21) \qquad u(x) = \frac{C}{1+|x|^2}$$

Les propriétés usuelles de l'entropie sont naturellement encore vérifiées; en particulier on a (R.B. Ash [1]) :

(22) $h(X) = h(X+a)$ pour toute constante a .

(23) si A est une application linéaire de $\mathbb{R}^k$ dans $\mathbb{R}^k$:

$$h(AX) = \operatorname{Log}|\det A| + h(A).$$

(24) $h(X,Y) \leq h(X) + h(Y)$ avec égalité si X et Y sont indépendantes.

(25) Si la matrice des covariances de X est fixée le maximum de $h(X)$ est atteint quand X est gaussienne.

Nous définissons maintenant l'entropie conditionnelle de X sachant Y où Y est à valeurs dans un espace quelconque et X est à valeurs dans $\mathbb{R}^k$ et admet une entropie au sens de (19) par :

$$(26) \qquad h(X|Y) = h(X) - I(X,Y) .$$

Les propriétés suivantes proviennent alors de celles de l'information et de l'entropie :

(27) $h(X,X') = h(X') + h(X|X')$.

(28) $h(X|g(X')) \geq h(X|X')$ si g est mesurable.

(29) $h(X+g(X')|X') = h(X|X')$

notons que (28) implique :

$$h(X|X',g(X')) = h(X|X') \quad . \tag{30}$$

Nous aurons besoin dans la suite des propositions suivantes.

__Proposition 2.1__

Soient X et Y deux variables aléatoires indépendantes et à valeurs dans $\mathbb{R}^k$; si X et $X+Y$ admettent une entropie au sens de (19) alors :

$$h(X+Y) \geq h(X) \quad .$$

__Preuve__ :

Si $h(X) = -\infty$ il n'y a rien à démontrer ;

si $h(X) > -\infty$ on a d'après (12) et (19) :

$$h(X+Y|Y) = h(X) \tag{31}$$

et aussi :

$$h(X+Y|Y) = h(X+Y) - I(X+Y,Y) \tag{32}$$

Comme $I(X+Y,Y)$ est positif on conclut.

Proposition 2.2.

Soit (X_n) une suite de variables aléatoires à valeurs dans $\mathbb{R}^k$ convergente en loi vers X et u une densité de probabilité sur $\mathbb{R}^k$; si les variables $\operatorname{Log} u(X_n)$ sont uniformément intégrables, alors :

(33) $$\limsup_n h(X_n) \leq h(X) .$$

Preuve :

C'est une conséquence de (10) et (19)

Nous définissons maintenant l'entropie d'un processus stationnaire.

Définition 2.4.

Soit $X = (X_n)_{n\in\mathbb{Z}}$ un processus réel stationnaire en loi admettant un moment absolu d'ordre $s > 0$ fini ; l'entropie de X est définie par :

(34) $$H(X) = \lim_n \frac{1}{n+1} h(X_o,\dots,X_n)$$

Notons que par stationnarité on a aussi pour tout k,

(35) $$H(X) = \lim_n \frac{1}{n+1} h(X_k,\dots,X_{k+n}) \ ;$$

on peut vérifier (voir M. Kanter [9]) que la suite $\frac{1}{n+1} h(X_o,\dots,X_n)$ est décroissante et que :

(36) $$H(X) = \lim_n h(X_o | X_{-1},\dots,X_{-n}) = h(X_o | X_{-\infty}^{-n})$$

où $X_{-\infty}^{-n} = (X_{-1},\dots,X_{-n},\dots)$

Nous allons maintenant définir des classes de processus que nous utiliserons dans la suite ; pour les propriétés des variables aléatoires symétriques

α-stable ($S\alpha S$) on se refère à S. Cambanis [4] où est définie en particulier la covariation de deux variables $S\alpha S$ X et Y, que nous noterons également $[X,Y]_\alpha$; la norme en covariance sera notée $\|X\|_\alpha = [X,X]_\alpha^{1/\alpha}$.

Dans toute la suite on supposera que $1 < \alpha \leq 2$.

Un processus réel ou complexe $X = (X_t)_{t \in T}$ est dit $S\alpha S$ si pour tout $t_1, \dots, t_k$ les variables $X_{t_1}, \dots, X_{t_k}$ sont conjointement $S\alpha S$; pour $\alpha = 2$, X est alors gaussien.

Soit $\xi(\lambda)$, $\lambda \in [-\pi, \pi]$ un processus $S\alpha S$ à accroissements indépendants, alors l'intégrale $\int_{-\pi}^{\pi} f(\lambda)\, d\xi(\lambda)$ est définie pour toute fonction f dans $L^\alpha(\mu, T)$ où $T = [-\pi, \pi]$ et μ est la mesure sur T définie par :

(37) $$\mu\{[a,b]\} = \|\xi(b) - \xi(a)\|_\alpha^\alpha \quad ; [4], [8], [13].$$

On a de plus :

(33) $$\left[\int_{-\pi}^{\pi} f d\xi \,,\, \int_{-\pi}^{\pi} g d\xi\right]_\alpha = \int_{-\pi}^{\pi} f (g)^{<\alpha-1>} d\mu$$

(39) $$\left\|\int f d\xi\right\|_\alpha = \|f\|_{L^\alpha(\mu,T)}$$

avec $x^{<\beta>} = |x|^{\beta-1}\, \bar{x}$.

On notera $\mathcal{S}_\alpha$ la classe des processus réels, $X = (X_n)_{n \in \mathbb{Z}}$ strictement stationnaires, $S\alpha S$ et qui admettent une représentation :

(40) $$X_n = \int_{-\pi}^{\pi} e^{in\lambda}\, d\xi(\lambda)$$

où $\xi(\lambda)$ est un processus $S\alpha S$ à accroissements indépendants, dont la mesure μ définie par (37) est finie. On dira que μ est la mesure spectrale de X et que $d\xi$ est sa mesure spectrale aléatoire ; on les notera respectivement μ_X et dZ_X .

On considère aussi la classe $\mathfrak{C}$ des processus réels strictement stationnaires et de carré intégrable ; (notons que $\mathcal{S}_2$ est la classe des processus gaussiens et est contenue dans $\mathfrak{C}$).

La théorie spectrale des processus dans $\mathfrak{C}$ est largement développée [11] ; S. Cambanis et A.R. Soltani [6] développent une théorie analogue dans $\mathcal{S}_\alpha$; ils obtiennent en particulier

2.5. Décomposition de Wold :

Tout $X \in \mathcal{S}_\alpha$, s'écrit $X = X_1 + X_2$ où X_1 et X_2 sont indépendants ; si $X_1 \neq 0$, μ_{X_1} admet une densité f par rapport à la mesure de Lebesgue et :

(41) $$\int_{-\pi}^{\pi} \text{Log } f(\lambda) d\lambda > -\infty \; ;$$

si $X_2 \neq 0$ la densité g de la partie absolument continue de μ_{X_2} vérifie :

(42) $$\int_{-\pi}^{\pi} \text{Log } g(\lambda)\, d\lambda = -\infty \; ;$$

X_1 est la partie régulière de X et X_2 sa partie singulière ; X est dit régulier (resp singulier) si $X_2 = 0$ (resp. $X_1 = 0$).

2.6. Moyenne mobile :

Si X est régulier de densité spectrale $f(\lambda)$, il existe une fonction extérieure h dans l'espace de Hardy H^α telle que $|h(\lambda)|^\alpha = f(\lambda)$ et le processus :

(43) $$\eta(\lambda) = \int_0^\lambda (\bar{h})^{-1}\, dZ_X$$

est un processus $S\alpha S$ à accroissements indépendants dont la mesure associée définie par (37) est la mesure de Lebesgue (η est un mouvement α-stable ; pour $\alpha = 2$, η est un mouvement brownien).

Si on pose :

(44) $$W_n = \int_{-\pi}^{\pi} e^{in\lambda} \, d\eta(\lambda)$$

on a alors :

(45) $$X_n = \int_{-\pi}^{\pi} e^{in\lambda} \bar{h}(\lambda) \, dZ_W \; ;$$

notons que la mesure spectrale μ_W de W est la mesure de Lebesgue ; si $\alpha \neq 2$, S. Cambanis et A.R. Soltani [6] montrent que les variables aléatoires $(W_n)_{n\in\mathbb{Z}}$ ne sont pas indépendantes ; mais les covariations $[W_n, W'_n]_\alpha$ sont nulles pour $n \neq n'$ comme on peut le voir en utilisant (38) ; on appellera W , l'innovation de X ; plus généralement si X est non singulier on appellera innovation de X , l'innovation de sa partie régulière.

Pour terminer notons que, si $X \in \mathcal{S}_\alpha$ et $\phi \in L^\alpha(\mu_X, T)$, le processus $Y = (Y_n)_{n\in\mathbb{Z}}$ défini par :

(46) $$Y_n = \int_{-\pi}^{\pi} e^{in\lambda} \, \phi(\lambda) \, dZ_X$$

est dans $\mathcal{S}_\alpha$; de plus :

(47) $$dZ_Y = \phi(\lambda) \, dZ_X \quad \text{et} \quad d\mu_Y = |\phi(\lambda)|^\alpha \, d\mu_X \; ;$$

on dira que Y est obtenu de X par le filtre linéaire ϕ .

3 - ENTROPIE ET FILTRE LINEAIRE

L'objet de cette section est d'évaluer l'entropie d'un processus obtenu par un filtre linéaire ; nous commençons par calculer l'entropie d'un processus singulier dans $\mathcal{C}$ ou $\mathcal{S}_\alpha$.

Proposition 3.1.

Si Y est singulier, alors $H(Y) = -\infty$.

Preuve :

Puisque Y est singulier, on a $Y_o = T(Y_{-\infty}^{-1})$ où T est linéaire, donc :

(48) $$I(Y_o, Y_{-\infty}^{-1}) \geq I(Y_o, T(Y_{-\infty}^{-1})) \geq I(Y_o, Y_o) ,$$

mais

(49) $$H(Y) = h(Y_o) - I(Y_o, Y_{-\infty}^{-1})$$

d'où

(50) $$H(Y) \leq h(Y_o) - I(Y_o, Y_o) ;$$

si Y_o est absolument continue on conclut grâce à (13) ; si Y_o n'est pas absolument continue on sait qu'alors $h(Y_o) = -\infty$ et on conclut encore.

Nous comparons maintenant l'entropie de Y avec l'entropie de sa partie régulière.

Proposition 3.2

Si Y est non singulier et Y_1 (resp Y_2) est sa partie régulière (resp singulière) alors :

(51) $$H(Y_1) \geq H(Y) \ ;$$

de plus on a l'égalité si Y_1 et Y_2 sont indépendants.

Preuve :

On pose $Y_1 = (Y_{1,n})_{n\in\mathbb{Z}}$ et $Y_2 = (Y_{2,n})_{n\in\mathbb{Z}}$

on a en utilisant (28) :

(52) $$\begin{aligned} H(Y_1) &= h(Y_{1,n} | Y_{1,-\infty}^{n-1}) \\ &\geq h(Y_n - Y_{2,n} \mid Y_{1,-\infty}^{n-1}, Y_{-\infty}^{n-1}) \ , \end{aligned}$$

Comme Y_{2n} est fonction de $Y_{-\infty}^{n-1}$ on obtient par (29) :

(53) $$H(Y_1) \geq h(Y_n | Y_{1,-\infty}^{n-1}, Y_{-\infty}^{n-1})$$

de même $Y_{1,-\infty}^{n-1}$ est fonction de $Y_{-\infty}^{n-1}$ et donc par (28) :

(54) $$H(Y_1) \geq h(Y_n | Y_{-\infty}^{n-1}) \ ;$$

ce qui termine la première partie.

Si Y_1 et Y_2 sont indépendants, on a d'après la proposition 2.1, $H(Y) \geq H(Y_1)$, on conclut donc la démonstration.

Pour évaluer l'entropie d'un processus X obtenu avec un filtre linéaire appliqué à un processus Y de $\mathfrak{C}$ ou $\mathfrak{S}_\alpha$, on a besoin de deux résultats préliminaires. Le premier est dû à L.A. Shepp, D. Slepian et A.D. Wyner [15] ; nous en donnons une démonstration élémentaire dans un cas particulier.

Lemme 3.1.

Soit X et Y deux processus réels stationnaires en loi admettant un moment absolu d'ordre $s > 0$ fini et tels que :

(55) $$X_n = \sum_{j=0}^{k} a_j Y_{n+m-j} \quad \text{où les } a_j \text{ sont réels et } m \text{ est fixé ;}$$

posons $\phi(\lambda) = \sum_{j=0}^{k} a_j e^{ij\lambda}$, on a alors :

(56) $$H(X) \geq \frac{1}{2\pi} \int_{-\pi}^{\pi} \mathrm{Log}|\phi(\lambda)|d\lambda + H(Y) .$$

Preuve : (quand $\sum a_j z^j$ n'a pas de zéros dans $|z| < 1$)

Sans rien perdre à la généralité on peut supposer que :

(57) $$X_n = \sum_{j=0}^{k} a_j Y_{n-j} ;$$

on voit alors facilement que :

(58) $$(Y_o, Y_1, \ldots, Y_{k-1}, X_k, X_{k+1}, \ldots, X_{n+k})' = A(Y_o, Y_1, \ldots, Y_{n+k})$$

où la matrice A a pour déterminant a_o^{n+1} .

On obtient en utilisant (23) et (24)

(59) $$h(Y_o, Y_1, \ldots, Y_{k-1}) + h(X_k, X_{k+1}, \ldots, X_{n+k}) \geq \mathrm{Log}|a_o|^{n+1} + h(Y_o, Y_1, \ldots, Y_{n+k})$$

mais il est immédiat que :

(60) $$\lim_n \frac{1}{n+1} h(Y_o, Y_1, \ldots, Y_{k-1}) \leq 0 \quad \text{et} \quad \lim_n \frac{1}{n+1} h(Y_o, Y_1, \ldots, Y_{n+k}) = H(Y)$$

on obtient donc :

(61) $$H(X) \geq \mathrm{Log}|a_o| + h(Y)$$

et d'après le théorème de Jensen (voir W. Rudin [12]), on a

(62) $$\mathrm{Log}|a_o| = \frac{1}{2\pi}\int_{-\pi}^{\pi} \mathrm{Log}\ |\phi(\lambda)|d\lambda \quad \text{où} \quad \phi(\lambda) = \sum_{j=0}^{k} a_j\ e^{ij\lambda} \ ;$$

ce qui conclut la démonstration ∎

Le deuxième résultat est un lemme d'approximation.

Lemme 3.2

Soit μ la mesure spectrale d'un processus réel non singulier de $\mathfrak{C}$ ou $\mathfrak{S}_\alpha$.
Soit $\phi \in L^p(\mu,T)$ avec $p \geq 1$; on suppose que :

(63) $$\phi(\lambda) = \overline{\phi(-\lambda)} \quad \text{pour tout } \lambda \in [-\pi,\pi] \ .$$

et

(64) $$\int_{-\pi}^{\pi} \mathrm{Log}\ |\phi(\lambda)|\ d\lambda > -\infty$$

Alors il existe une suite ϕ_n de polynômes trigonométriques à coefficients réels tels que :

(65) $$\lim \phi_n = \phi \qquad \text{dans } L^p(\mu,T)$$

et

(66) $$\lim \int_{-\pi}^{\pi} \mathrm{Log}\ |\phi_n(\lambda)|\ d\lambda = \int_{-\pi}^{\pi} \mathrm{Log}\ |\phi(\lambda)|\ d\lambda \ .$$

Preuve :

Soit $\mu = \mu_1 + \mu_2$ la décomposition de Lebesgue de μ . Puisque le processus de mesure spectrale μ est non singulier, μ_1 est équivalente à la mesure de Lebesgue.

Faisons d'abord l'approximation dans $L^p(\mu_1,T)$.

Soit $\delta > 0$, il existe alors $\varepsilon > 0$, $\varepsilon < \delta$ tel que pour tout borélien B vérifiant $\mu_1(B) < \varepsilon$ on ait :

(67) $$\int_B |\phi|^p \, d\mu_1 < \delta \quad ; \quad \int_B \big|\mathrm{Log}|\phi(\lambda)|\big| \, d\lambda < \delta$$

$$\int_B |\dot{\phi}| \, d\mu_1 < \delta \quad ; \quad \int_B d\lambda < \delta$$

celà provient du fait que les mesures $|\phi|^p \, d\mu_1$, $\big|\mathrm{Log}|\phi|\big| \, d\lambda$, $|\phi| \, d\mu_1$ et $d\lambda$ sont dominées par $d\mu_1$.

Puisque $\int \mathrm{Log}|\phi| \, d\lambda > -\infty$, la mesure de Lebesgue de l'ensemble N des zéros de ϕ est nulle.

On se place dans le complémentaire de N ; par le théorème de Lusin [12] , on peut trouver un compact K_1 qu'on peut choisir symétrique par rapport à $\lambda = 0$ et une fonction continue ψ_1 tels que :

(68) $$\mu_1(T-K_1) < \varepsilon$$

et

(69) $$\psi_1 = \phi \quad \text{sur} \quad K_1 \, .$$

ψ_1 étant sans zéros sur K_1 , soient :

(70) $$\gamma = \sup_{K_1} \left(\big|\mathrm{Log}|\psi_1|\big|\right) \quad \text{et} \quad \beta = \max\{1, \sup_{K_1}|\psi_1|\} \, .$$

Choisissons maintenant $\varepsilon_2 > 0$ et K_2 un compact dans $T-K_1$ symétrique par rapport à $\lambda = 0$ tels que :

(71) $$\varepsilon_2 < \frac{\delta}{\beta^p} \; ; \; \varepsilon_2 < \frac{\delta}{\gamma}$$

et

(72) $$\mu_1(T-(K_1 \cup K_2)) < \varepsilon_2 .$$

Définissons sur $K = K_1 \cup K_2$ la fonction :

(73) $$\psi_2 = \begin{cases} \psi_1 & \text{sur } K_1 \\ 1 & \text{sur } K_2 \end{cases}$$

On prolonge ψ_2 de la manière suivante.

Si $I =]\lambda_1\ \lambda_2[$ est un intervalle dans le complémentaire de K tel que λ_1 et λ_2 soient dans K, on pose pour tout λ dans I :

(74) $$\psi_2(\lambda) = \left[\frac{\lambda-\lambda_1}{\lambda_2-\lambda_1}|\psi_2(\lambda_2)| + \frac{\lambda_2-\lambda}{\lambda_2-\lambda_1}|\psi_2(\lambda_1)|\right] \exp\left\{i\left[\frac{\lambda-\lambda_1}{\lambda_2-\lambda_1}\operatorname{Arg}\psi_2(\lambda_2) + \frac{\lambda_2-\lambda}{\lambda_2-\lambda_1}\operatorname{Arg}\psi_2(\lambda_1)\right]\right\}$$

On obtient ainsi une fonction continue qui vérifie :

(75) $$\psi_2(\lambda) = \overline{\psi_2(-\lambda)}$$

$$|\psi_2| < \beta \ ; \ |\text{Log } |\psi_2|| < \gamma$$

(ψ_2 n'a pas de zéro sur T).

Il est alors facile de vérifier que :

(76) $$\int_{-\pi}^{\pi} |\phi-\psi_2|^p \, d\mu_1 \leq 8\delta$$

et

(77) $$\left|\int_{-\pi}^{\pi} \text{Log } |\phi| d\lambda - \int_{-\pi}^{\pi} \text{Log } |\psi_2| \, d\lambda\right| \le 3\delta$$

Passons maintenant à $L^p(\mu,T)$; on peut prendre un support de μ_2 : S symétrique par rapport à $\lambda = 0$.

Soit Q un polynôme à coefficients réels tel que :

(78) $$\int_{-\pi}^{\pi} |\phi-Q|^p \, d\mu_2 < \delta \quad ;$$

on peut prendre Q sans zéro sur T en perturbant faiblement ses racines et en gardant des coefficients réels. Notons :

$$\psi_3 = \begin{cases} \psi_2 & \text{sur } T-S \\ Q & \text{sur } S \end{cases}$$

Il est clair que :

(79) $$\int_{-\pi}^{\pi} |\phi-\psi_3|^p \, d\mu \le 9\delta$$

et

(80) $$\left|\int_{-\pi}^{\pi} \text{Log } |\psi_3| \, d\lambda - \int_{-\pi}^{\pi} \text{Log } |\phi| d\lambda\right| \le 3\delta$$

ψ_3 n'a pas de zéros sur T ; on pose :

(81) $$\beta = \sup_T |\psi_3| \quad \text{et} \quad \gamma = \sup_T |\text{Log } |\psi_3||.$$

Soient ε' et ε'' tels que :

(82) $$\varepsilon' < \delta/\gamma \ , \quad \varepsilon'' < \frac{\delta}{4\beta^p}$$

et

(83) $$\mu(B) < \varepsilon'' \Rightarrow \int_B d\lambda < \varepsilon' .$$

Par le théorème de Lusin, il existe un compact K (qu'on prendra symétrique) et une fonction continue ψ' tels que :

(84) $$\mu(T-K) < \varepsilon''$$

et

(85) $$\psi' = \psi_3 \quad \text{sur} \quad K .$$

On prolonge ψ' comme plus haut en une fonction continue ψ qui vérifie

(86) $$|\psi| \leq \beta \quad \text{et} \quad |\mathrm{Log}\ |\psi|| \leq \gamma \ ;$$

il ne reste plus qu'à approcher ψ uniformément par un polynôme trigonométrique à coefficients réels pour conclure la démonstration.

Remarque 3.1

Le lemme est encore vrai pour la norme L^∞ ; si ϕ est une fonction continue sur T, on peut l'approcher uniformément par des polynômes trigonométriques à coefficients réels qui vérifient (66).

Remarque 3.2

La condition $\phi(\lambda) = \overline{\phi(-\lambda)}$ signifie que le filtre est réel ; sans cette condition le lemme est encore vrai mais l'approximation est faite avec des polynômes à coefficients complexes.

On établit maintenant notre résultat principal.

Théorème 3.1

Soient X et Y deux processus dans $\mathfrak{C}$ ou $\mathfrak{S}_\alpha$ tels que :

$$X_n = \int_{-\pi}^{\pi} e^{in\lambda}\,\phi(\lambda)dZ_Y \quad \text{avec} \quad \phi \in L^\alpha(\mu_Y,T)\ .$$

Alors :

(i) $$H(X) \geq \frac{1}{2\pi}\int_{-\pi}^{\pi} \text{Log}\ |\phi(\lambda)|d\lambda + H(Y)$$

(ii) On a l'égalité dans (i) dans les cas suivants :

a) $\phi(\lambda) \neq 0$ μ_Y-presque partout

b) X est singulier

c) Y est régulier

d) $Y \in \mathfrak{S}_\alpha$

Preuve

(i) Si Y est singulier, X l'est aussi et (i) est vérifiée d'après la proposition 3.1 et on a en fait trivialement l'égalité dans (i).

Supposons que Y est non singulier ; on a alors deux possibilités ; si X est non singulier, on a :

(87) $$\int_{-\pi}^{\pi} \text{Log}\ |\phi(\lambda)|\ d\lambda > -\infty\ ;$$

d'après le lemme 3.2 on peut trouver une suite de polynômes trigonométriques à coefficients réels ϕ_k, tels que :

(88) $$\lim \phi_k = \phi \quad \text{dans} \quad L^\alpha(\mu_Y,T) \quad (\text{avec } \alpha = 2 \text{ si } X \text{ et } Y \text{ sont dans } \mathfrak{C})$$

et

(89) $$\lim_k \int_{-\pi}^{\pi} \text{Log}\ |\phi_k(\lambda)|d\lambda = \int_{-\pi}^{\pi} \text{Log}\ |\phi(\lambda)|\ d\lambda\ .$$

Soit :

(90) $$X_n(k) = \int_{-\pi}^{\pi} e^{in\lambda} \phi_k(\lambda)\, dZ_Y$$

on a :

(91) $$\|X_n(k) - X_n\|_\alpha^\alpha = \int |\phi_k(\lambda) - \phi(\lambda)|^\alpha \, d\mu_Y = \delta_k$$

et :

(92) $$\lim \delta_k = 0$$

On peut extraire une sous suite telle que :

(93) $$\Sigma\, \delta_k < \infty$$

de sorte que le processus $X(k)$ converge en loi vers X ; et donc :

(94) $$\lim\inf I(X_o(k), X_{-\infty}^{-1}(k)) \geq I(X_o, X_{-\infty}^{-1})$$

mais :

(95) $$H(X(k)) = h(X_o(k)) - I(X_o(k), X_{-\infty}^{-1}(k))$$

et d'après la proposition 2.2 .

(96) $$\limsup_k h(X_o(k)) \leq h(X_o) \;;$$

il vient par conséquent :

(97) $$\limsup_k H(X(k)) \leq H(X)$$

on conclut alors avec les lemmes 3.1 et 3.2.

Si maintenant X est singulier, on a :

(98) $$\int_{-\pi}^{\pi} \mathrm{Log}\ |\phi(\lambda)|\ d\lambda = -\infty$$

et

(99) $$-\infty = H(X) = -\infty + H(Y)$$

ii) a) si $\phi(\lambda) \neq 0$ μ_Y-presque partout, alors $\psi(\lambda) = \frac{1}{\phi(\lambda)}$ est dans $L^{\alpha}(\mu_X)$ et on a :

(100) $$Y_n = \int_{-\pi}^{\pi} e^{in\lambda}\ \psi(\lambda)\ dZ_X$$

par conséquent d'après (i) :

(101) $$H(Y) \geq -\frac{1}{2\pi}\int_{-\pi}^{\pi} \mathrm{Log}|\phi(\lambda)|d\lambda + H(X) ,$$

on conclut donc.

b) a été démontré dans (i)

c) si Y est régulier, ou bien X est singulier et on a l'égalité d'après b), ou bien X est non singulier et alors $\phi(\lambda) \neq 0$ presque partout pour la mesure de Lebesgue. Comme μ_Y est absolument continue il s'ensuit que $\phi(\lambda) \neq 0$ μ_Y-presque partout ; ce qui nous ramène à a).

d) Si Y est singulier on a évidemment l'égalité. Si Y est non singulier, soit $Y = Y_1 + Y_2$ sa décomposition de Wold. Y_1 et Y_2 étant indépendants on a d'après la proposition 3.2.

(102) $$H(Y) = H(Y_1) .$$

D'autre part, il est clair que si X_1 est la partie régulière de X on a :

(103) $$X_{1n} = \int_{-\pi}^{\pi} e^{in\lambda} \phi(\lambda)\, dZ_{Y_1}$$

et donc :

(104) $$H(X_1) = \frac{1}{2\pi} \int_{-\pi}^{\pi} \mathrm{Log}\, |\phi(\lambda)|\, d\lambda + H(Y_1)$$

or :

(105) $$H(X_1) \geq H(X)$$

d'après la proposition 3.2. ; (102) (104) et (105) impliquent alors :

(106) $$\frac{1}{2\pi} \int_{-\pi}^{\pi} \mathrm{Log}\, |\phi(\lambda)|\, d\lambda + H(Y) \geq H(X)$$

ce qui achève la démonstration.

Remarque 3.3

Dans [15] , L.A. Shepp, D. Slepian et A.D. Wyner affirment qu'on a l'égalité dans (i) quand Y est ergodique ; leur démonstration est malheureusement inexacte. Elle consiste principalement à montrer que le filtre est inversible ; ce qui ne peut être le cas si la mesure μ_Y charge les zéros de ϕ (le seul point qui n'est jamais chargé par μ_Y quand Y est ergodique est le point $\lambda = 0$). Leur principal argument consiste à montrer que si $\xi \in \mathbb{C}$, $|\xi| = 1$ alors la suite de variables aléatoires :

(107) $$c_k = \frac{1}{k} \sum_{j=1}^{k} \xi^{-j}\, Y_{n+j}$$

converge vers une constante presque sûrement. Cela est inexact si $\xi \neq 1$; un contrexemple simple est le suivant. Supposons $\xi = e^{i\lambda}$ avec $\lambda \neq 0$ et prenons

(108) $$Y_n = e^{in\lambda} V \; , \; (Y_n) \text{ ergodique} \; ;$$

dans [11] la loi de V est complètement caractérisée : $|V|$ est constant p.s. et la loi de Arg V est uniforme sur le sous groupe fermé de T engendré par $e^{in\lambda}$. Comme $\lambda \neq 0$, ce sous groupe est toujours de cardinal strictement plus grand que 1 (il est fini si $\lambda/2\pi$ est rationnel et égal à T sinon). V n'est donc pas constant p.s. ; si on regarde maintenant c_k on obtient

(109) $$c_k = \frac{1}{k} \sum_{j=1}^{k} e^{-i\lambda j} e^{i(n+j)\lambda} V$$

$$= e^{in\lambda} V$$

et donc

(110) $$\lim c_k = e^{in\lambda} V$$

ce qui n'est ni constant ni indépendant de n.

4 - APPLICATIONS

4.1. Entropie d'un processus régulier de $\mathbf{C}$ ou $\mathcal{S}_\alpha$

Soit X un processus régulier de densité spectrale f ; soit W son innovation (μ_W est la mesure de Lebesgue) ; on a alors :

(111) $$X_n = \int_{-\pi}^{\pi} e^{in\lambda} h(\lambda) \, dZ_W$$

où

(112) $$|h(\lambda)|^{\alpha} = f(\lambda) \quad \text{avec} \quad \alpha = 2 \quad \text{si} \quad X \in \mathbf{C} \, .$$

On obtient :

$$(113) \qquad H(X) = \frac{1}{2\alpha\pi} \int_{-\pi}^{\pi} \text{Log } f(\lambda)\, d\lambda + H(W) .$$

Dans $\mathcal{C}$, $H(W)$ est maximal si W est gaussien et on obtient donc :

$$(114) \qquad H(X) \leq \frac{1}{4\pi} \int_{-\pi}^{\pi} \text{Log } f(\lambda) d\lambda + \frac{1}{2} \text{Log } 2\pi e.$$

avec égalité si X est gaussien.

Enfin on a aussi le corollaire

Corollaire 4.1

Soit X un processus régulier de $\mathcal{C}$ et W son innovation ; alors :

Si $H(W) > -\infty$ les lois de probabilité marginales de dimension finie de X sont absolument continues ;

Si $H(X) > -\infty$ les lois de probabilité marginales de dimension finie de W sont absolument continues.

4.2. Entropie d'un processus non singulier de $\mathcal{S}_\alpha$.

Si $X \in \mathcal{S}_\alpha$ et X est non singulier, soit X_1 sa partie régulière et f sa densité spectrale. Comme $H(X_1) = H(X)$, on a alors :

$$(115) \qquad H(X) = \frac{1}{2\alpha\pi} \int_{-\pi}^{\pi} \text{Log } f(\lambda) d\lambda + H(W)$$

où W est l'innovation de X_1 ; en particulier si $\alpha = 2$, X est gaussien et $H(W) = \frac{1}{2} \text{Log } 2\pi e$; il faut remarquer alors que dans ce cas :

$$(116) \qquad \frac{1}{n+1} h(X_o,\ldots,X_{n+1}) = \frac{1}{2} \text{Log } 2\pi e \left|\det Q_{n+1}\right|^{1/n+1}$$

où Q_{n+1} est la matrice des covariances de $(X_o,\dots,X_n)$; comme

(117) $$H(X) = \lim_n \frac{1}{n+1} h(X_o,\dots,X_n)$$

on en déduit que :

(118) $$\lim_n |\det Q_{n+1}|^{1/n+1} = \exp\{\frac{1}{2\pi}\int_{-\pi}^{\pi} \text{Log } f(\lambda)\, d\lambda\}$$

ce qui est le premier théorème de Szëgo.

4.3. Processus d'entropie maximale pour une densité spectrale $f(\lambda)$ fixée

D'après la formule (113) il s'agit de trouver l'innovation W pour laquelle $H(W)$ est maximum ; celà est possible dans $\mathfrak{C}$; la seule contrainte sur W est $EW_n^2 = 1$; comme $H(W) \leq h(W_n)$ avec égalité si et seulement si les (W_n) sont indépendants, il suffit donc de prendre des W_n indépendants tels que $h(W_n)$ soit maximum sous la contrainte $EW_n^2 = 1$; celà est réalisé d'après (25) pour W_n gaussienne. Le processus d'entropie maximale est donc le processus gaussien de densité spectrale $f(\lambda)$; en conséquence pour tout X dans $\mathfrak{C}$ de densité spectrale $f(\lambda)$ on a :

(119) $$H(X) \leq \frac{1}{4\pi}\int_{-\pi}^{\pi} \text{Log } f(\lambda)d\lambda + \frac{1}{2}\text{Log } 2\pi e\ .$$

Dans $\mathcal{S}_\alpha$, $\alpha \neq 2$, la situation est différente car les W_n ne peuvent pas être indépendants. Le problème qui reste donc posé est celui de trouver le processus stationnaire $S\alpha S$ de mesure spectrale la mesure de Lebesgue et qui a une entropie maximale ; cependant on peut obtenir une majoration de $H(X)$ en majorant $H(W)$ par $h(W_o)$; $h(W_o)$ est majorée de la manière suivante : pour $1 \leq p < \alpha$ on a :

(120) $$(E|W_o|^p)^{1/p} = \|W_o\|_p = C(p,\alpha)\ \|W_o\|_\alpha$$

où $C(p,\alpha)$ est une constante qui ne dépend que de p et α [5] d'autre part il est bien connu que :

(121) $$h(W_o) \leq \text{Log} \frac{2\Gamma(1/p)\ \|w_o\|_p}{p^{\frac{p-1}{p}}} + \frac{1}{p} ;$$

ce qui permet d'obtenir une majoration de $H(X)$.

4.4. Comparaison de l'erreur de prédiction linéaire et non linéaire

Soit Y dans $\mathcal{S}_\alpha$ ou $\mathcal{C}$ et X tel que :

(122) $$X_n = \int_{-\pi}^{\pi} e^{in\lambda}\ \phi(\lambda)\ dZ_Y ;$$

Soit $\hat{X}_n = f(X_{-\infty}^{n-1})$ un prédicteur de X_n et $p > 0$; on peut montrer alors que :

(123) $$\|X_n-\hat{X}\|_p \geq C(p)\ e^{H(X)} \quad , [15]$$

où $C(p)$ est une constante qui ne dépend que de p .

Si on note $\varepsilon_L(X)$ et $\varepsilon_L(Y)$ les erreurs de prédiction linéaires à un pas respectivement de X et Y on voit alors que :

(124) $$\|\varepsilon_L(X)\|_\alpha = [\exp\{\frac{1}{2\pi}\int_{-\pi}^{\pi} \text{Log}\ |\phi(\lambda)|d\lambda\}].\ \|\varepsilon_L(Y)\|_\alpha \quad ,$$

(pour $\alpha = 2$ voir Y. Rozanov [11] et pour $1 < \alpha \leq 2$ S. Cambanis et A.R. Soltani [6]) ; ce qui donne en utilisant le théorème 3.1 et (123) :

(125) $$\|X_n-\hat{X}\|_p \geq \{C(p)\ .\ \|\varepsilon_L(Y)\|_\alpha^{-1}\ .\ e^{H(Y)}\}\ \|\varepsilon_L(X)\|_\alpha .$$

4.5. Principe du maximum d'entropie

Connaissant les k premières covariations d'un processus X dans $\mathfrak{C}$ ou $\mathfrak{S}_\alpha$, on peut alors se poser deux problèmes :

4.5.1. Quand l'innovation de X est fixée trouver la densité spectrale f qui maximise $H(X)$;

4.5.2. Trouver le processus X pour lequel $H(X)$ est maximal.

Qu'on soit dans $\mathfrak{C}$ ou $\mathfrak{S}_\alpha$ le problème 4.5.1 revient à trouver la fonction positive intégrable f qui maximise

$$(126) \qquad \int_{-\pi}^{\pi} \text{Log } f(\lambda)\, d\lambda \quad ,$$

les k premiers coefficients de Fourier de f étant fixés ; c'est la méthode d'estimation de f par maximum d'entropie (J.P. Burg [2] , [3]) . La solution à ce problème est connue ; la densité spectrale cherchée est celle d'un processus autorégressif dans $\mathfrak{C}$ ou $\mathfrak{S}_\alpha$.

La solution au problème 4.5.2 d'après ce qu'on vient de voir, est un processus autorégressif ; il reste à trouver l'innovation W qui maximise $H(W)$ (la contrainte étant $\|W_o\|_\alpha = 1$) ; comme on l'a vu en 4.3., dans $\mathfrak{C}$ la solution est obtenue en prenant W gaussien ; par contre dans $\mathfrak{S}_\alpha$, $\alpha < 2$, le problème reste entièrement posé puisqu'on ne peut pas choisir $W = (W_n)_{n\in\mathbb{Z}}$ avec des W_n indépendants.

Enfin si on connaît d'autres moments de X , celà imposera des conditions sur les multispectres de X et la solution du problème 4.5.2 dans $\mathfrak{C}$ ne sera plus un processus gaussien ; pour éclairer celà nous prenons l'exemple simple suivant :

4.5.3. Trouver le processus X dans $\mathbb{C}$ dont l'entropie est maximale et les k premières covariances et le moment d'ordre trois sont fixés ; le moment d'ordre trois étant non nul.

La solution est encore un processus autorégressif :

$$(127) \qquad X_n = \int_{-\pi}^{\pi} e^{in\lambda} \phi(\lambda)\, dZ_W$$

où ϕ est l'inverse d'un polynôme et W une innovation ; il reste à maximiser $H(W)$.

La densité bispectrale de X est donnée par :

$$(128) \qquad f(\lambda_1,\lambda_2) = \phi(-\lambda_1-\lambda_2)\ \phi(\lambda_1)\ \phi(\lambda_2)\ h(\lambda_1,\lambda_2) \qquad [16]$$

où h est la densité bispectrale de W .

Comme la condition $E(X_o^3)$ = cste peut être satisfaite avec $W = (W_n)_{n\in\mathbb{Z}}$ où les W_n sont indépendants ; on choisit donc une innovation de ce type ; la relation (128) devient alors :

$$(129) \qquad f(\lambda_1,\lambda_2) = \frac{E\, W_o^3}{(2\pi)^2}\ \phi(-\lambda_1-\lambda_2)\ \phi(\lambda_1)\ \phi(\lambda_2)$$

et la condition $E\, X_o^3$ = cste donne :

$$(130) \qquad \frac{E\, W_o^3}{(2\pi)^2} \int_{-\pi}^{\pi}\int_{-\pi}^{\pi} \phi(-\lambda_1-\lambda_2)\ \phi(\lambda_1)\ \phi(\lambda_2)\ d\lambda_1\, d\lambda_2 = E\, X_o^3$$

et par conséquent :

(131) $$E\, W_o^3 = \text{cste} .$$

Comme on a aussi $E\, W_o^2 = 1$ et que $H(W) = h(W_o)$, le maximum de $H(W)$ est obtenu en prenant $W = (W_n)_{n\in \mathbb{Z}}$ avec des W_n indépendants et de densité de probabilité

(132) $$g(w) = c \exp \{a_1 w + a_2 w^2 + a_3 w^3\} \quad \text{(I. Csiszar [7])}$$

où c , a_1 , a_2 et a_3 sont déterminés par :

(133) $$\int g(w)dw = 1 \ , \quad \int w\, g(w)dw = 0 \ , \quad \int w^2 g(w)dw = 1 \ , \quad \int w^3 g(w)dw = EW_o^3 \ .$$

4.6. Cas où Y n'a pas de représentation spectrale

Si on ne suppose plus que Y admet une représentation spectrale, on obtient le résultat suivant

Théorème 4.1

Soit $Y = (Y_n)_{n\in\mathbb{Z}}$ un processus réel stationnaire en loi vérifiant $E|Y_n| < \infty$; soit $X = (X_n)_{n\in\mathbb{Z}}$ le processus défini par la série convergente en moyenne :

(134) $$X_n = \sum_{-\infty}^{+\infty} a_j\, Y_{n-j} \quad \text{où} \quad \sum_{-\infty}^{+\infty} |a_j| < \infty \ ;$$

alors :

(135) $$H(X) \geq \frac{1}{2\pi} \int_{-\pi}^{\pi} \operatorname{Log} |\phi(\lambda)|\, d\lambda + H(Y)$$

où $$\phi(\lambda) = \sum_{-\infty}^{+\infty} a_j\, e^{ij\lambda}$$

de plus on a l'égalité dans (135) si $\phi(\lambda)$ n'a pas de zéros sur le tore $T = [-\pi,\pi]$.

Preuve

Quand $\int_{-\pi}^{\pi}$ Log $|\phi(\lambda)|d\lambda = -\infty$, le théorème est évident ; sinon $\phi(\lambda)$ étant une fonction continue, on l'approxime alors uniformément par des polynômes trigonométriques $\phi_k(\lambda)$ (remarque 3.1 du Lemme 3.2) ; (135) s'obtient alors par passage à la limite comme dans le théorème 3.1. Si $\phi(\lambda) \neq 0$ sur T , comme $\Sigma|a_j| < \infty$ on a d'après le théorème de Wiener (W. Rudin [12]) :

(136) $$\frac{1}{\phi(\lambda)} = \sum_{-\infty}^{+\infty} b_j e^{ij\lambda} \quad \text{où} \quad \sum_{-\infty}^{+\infty} |b_j| < \infty$$

il est alors facile de voir que :

(137) $$Y_n = \sum_{j=-\infty}^{+\infty} b_j X_{n-j} ;$$

et l'application une nouvelle fois de (135) donne l'égalité voulue.

REFERENCES

[1] R.B. ASH (1967), Information theory. Intersciences publishers.

[2] J.P. BURG (1978), Maximum entropy spectral analysis. In modern spectrum analysis, ed. D.G. Childers, Wiley, New York.

[3] J.P. BURG (1978), A new analysis technique for time series data. In modern spectrum analysis, ed. D.G. Childers, Wiley, New York.

[4] S. CAMBANIS (1982), Complex symmetric stable variables and processes. In : Contribution to Statistics : Essays in Honour of Norman L. Johnson, pp. 63-79. Sen, P.K., Ed New York, North Holland.

[5] S. CAMBANIS and G. MILLER (1981), Linear problems in p th order and stable processes. SIAM J. Appl. Math. 41, pp. 43-69.

[6] S. CAMBANIS and A.R. SOLTANI (1984), Prediction of stable processes : spectral and moving average representations. Z. Wahr. verw. gebiete, 66, pp. 593-612.

[7] I. CSISZAR (1975), I-divergence geometry of probability distributions and minimization problems. Ann. Prob., vol. 3, n° 1, pp. 146-158.

[8] Y. HOSOYA (1982), Harmonizable stable processes. Z. Wahr. verw. gebiete 60, pp. 517-533.

[9] M. KANTER (1979), Lower bound for non linear prediction error in moving average processes, Ann. Prob., 7, pp. 128-138.

[10] M. PINSKER (1964), Information and information stability of random variables and processes. Holden Day Series.

[11] Y. ROZANOV (1967), Stationary random processes, Holden Day Series.

[12] W. RUDIN (1974), Real and complex analysis, Mc Graw Hill.

[13] M. SCHILDER (1970), Some structure theorems for the symmetric stable laws, Ann. Math. Stat., 41, p. 412-421.

[14] C.E. SHANNON (1948), The mathematical theory of communications, Bell System Technical Journal.

[15] L.A. SHEPP, D. SLEPIAN and A.D. WYNER (1980), On prediction of moving average processes, Bell system Technical Journal vol. 59, n° 3, pp. 367-415.

[16] T. SUBBA RAO and M.M. GABR (1984), An introduction to bispectral analysis and bilinear time series models. Springer Verlag.