

N° d'ordre : 621

50376
1984
86

50376.
1984.
86.

THÈSE

présentée à

l'UNIVERSITÉ DES SCIENCES ET TECHNIQUES DE LILLE I

pour obtenir le grade de

DOCTEUR ES SCIENCES MATHÉMATIQUES

par

Alain BERLINET



SUR QUELQUES PROBLÈMES D'ESTIMATION FONCTIONNELLE
ET DE STATISTIQUE DES PROCESSUS



030 023119 0

Thèse soutenue le 14 juin 1984 devant la Commission d'Examen

Président et Rapporteur : J. GEFROY (Université de Paris VI)

Directeur de Recherche et Rapporteur : D. BOSQ (Université de Lille I)

Rapporteur : A. MONFORT (I.N.S.E.E.)

Examineurs {
C. BREZINSKI (Université de Lille I)
P. DEHEUVELS (Université de Paris VI)
P. JACOB (Université de Lille I)
B. VAN CUTSEM (Université de Grenoble)

*

* *

Nous remercions tous les membres du jury qui, à une période de l'année chargée par de nombreuses obligations, ont accepté la tâche supplémentaire de juger ce travail.

Monsieur le Professeur Jean Geffroy nous fait le grand honneur de présider le jury de cette thèse. Nous tenons à lui exprimer notre respectueuse gratitude.

Notre travail doit beaucoup au dynamisme incitatif et à la grande disponibilité de Monsieur le Professeur Denis Bosq. Il a guidé notre recherche avec compétence et efficacité. Nous lui en sommes particulièrement reconnaissant.

Les discussions que nous avons eues avec Monsieur le Professeur Alain Monfort au sujet des deux derniers chapitres ont permis d'améliorer la rédaction sur de nombreux points. Nous l'en remercions très vivement.

Nos remerciements vont également à Monsieur le Professeur Claude Brézinski qui nous a donné en troisième cycle le goût de l'Analyse Numérique et les connaissances en Accélération de Convergence qui ont été à la base de la quatrième partie.

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

Nous remercions vivement Monsieur le Professeur Pierre Jacob qui a accepté de participer au jury et dont les conseils nous ont été précieux.

Monsieur le Professeur Bernard Van Cutsem a bien voulu faire partie de la Commission d'examen. Nous lui adressons nos vifs remerciements.

./..

Nous remercions Monsieur le Professeur Johannes Gordesch de l'Université Libre de Berlin pour l'intérêt qu'il a porté à notre travail et Monsieur le Professeur Manfred Deistler de l'Université de Vienne pour ses remarques et commentaires sur le chapitre six.

Que soient remerciés tous les participants au séminaire de Probabilités et Statistique de Lille I qui nous ont aidé, sur le plan mathématique par leurs remarques et leurs suggestions, et sur le plan humain par leur sympathie ou leur amitié.

Mesdames Raymonde Bérat et Arlette Lengaigne ont dactylographié ce travail avec beaucoup de soin et de gentillesse. Qu'elles en soient remerciées, ainsi que toutes les personnes ayant pris part à la réalisation matérielle de cette thèse.

*

* *

P L A N.

INTRODUCTION.

BIBLIOGRAPHIE DE L'INTRODUCTION.

PREMIERE PARTIE :

CHAPITRE 1 : Localisation et homogénéité d'un paramètre.

CHAPITRE 2 : Exhaustivité locale d'une statistique. Application au conditionnement d'estimateurs.

DEUXIEME PARTIE :

CHAPITRE 3 : Convergence des estimateurs splines de la densité.

TROISIEME PARTIE :

CHAPITRE 4 : Approximation forte d'intégrales stochastiques empiriques multidimensionnelles en variables mélangeantes.

QUATRIEME PARTIE :

CHAPITRE 5 : Estimating the degrees of an ARMA model.

Annexe n° 1.

CHAPITRE 6 : Some useful algorithms in time series analysis.

Annexe n° 2.

INTRODUCTION.

Cette thèse comprend quatre parties. Elle est basée sur des articles qui ont été publiés entre 1981 et 1984 et qui sont le résultat de recherches effectuées dans plusieurs directions. Deux grands thèmes ont été abordés : l'estimation fonctionnelle (première et deuxième parties) et la statistique des processus (troisième et quatrième parties). Cette introduction a été rédigée de façon à ce que sa lecture permette de se faire une idée précise des concepts introduits et des développements qui en résultent.

La première partie a fait l'objet des publications [11], [12] et [14]. Elle est consacrée à l'étude de certaines propriétés des paramètres relatives à l'espace des observations, à leurs liens avec les concepts habituels de la statistique et aux estimateurs que l'on peut en déduire. Nous qualifions ces propriétés de *locales* en un sens qui n'a rien à voir avec l'utilisation usuelle de ce vocable : habituellement on nomme *locales* les propriétés d'un paramètre relatives à des "voisinages" dans l'ensemble des lois ou dans l'ensemble dans lequel le paramètre prend ses valeurs. Ces "voisinages" peuvent être des véritables voisinages au sens topologique, ou non (modèles de contamination). Nous nous intéressons ici à des propriétés locales dans un troisième ensemble : celui sur lequel les lois sont définies.

Considérons comme exemple introductif l'ensemble \mathcal{P} des lois de probabilité P sur \mathbb{R} absolument continues par rapport à la mesure de Lebesgue, dont la densité admet une version continue $\psi_1(P) = f_P$ et qui admettent une espérance $\psi_2(P) = \int_{\mathbb{R}} x dP$. Notons $\psi_3(P) = F_P$ la fonction de répartition de P . Les paramètres ψ_1 , ψ_2 et ψ_3 ont un caractère "global" : sous la seule hypothèse que $P \in \mathcal{P}$, la connaissance de la restriction de P à un sous-ensemble E de \mathbb{R} tel que $P(E) < 1$ ne nous permet pas de déterminer les images de P par ψ_1 , ψ_2 ou ψ_3 . Il en va autrement pour les paramètres qui associent à P les réels : $f_P(a)$ ou $F_P(a)$, où a est fixé dans \mathbb{R} : la connaissance de la restriction de P à un ensemble contenant $]-\infty, a[$ entraîne celle de $F_P(a)$ et la connaissance de la restriction de P à un voisinage de a entraîne celle de $f_P(a)$. On a même plus que cette propriété de localisation pour $f_P(a)$ et $F_P(a)$: ces paramètres ont une propriété d'homogénéité : si deux éléments P et Q de \mathcal{P} admettent des probabilités conditionnelles $P(.|U)$ et $Q(.|U)$ par rapport à un voisinage U de a (resp. de $]-\infty, a[$) qui sont égales, alors

$$\frac{f_P(a)}{P(U)} = \frac{f_Q(a)}{Q(U)} \quad (\text{resp.} \quad \frac{F_Q(a)}{P(U)} = \frac{F_Q(a)}{Q(U)})$$

L'incidence en statistique de la localisation est claire : si l'on cherche à estimer la valeur du paramètre pour une loi P inconnue dont on observe un échantillon (X_1, \dots, X_n) , les observations tombant à proximité de l'ensemble localisant le paramètre nous fourniront l'essentiel de l'information apportée par l'échantillon (à titre d'exemple on peut penser à la méthode de la fenêtre mobile pour la densité). Ces propriétés ont été utilisées de façon plus ou moins explicite dans bon nombre de travaux sans pour autant être étudiées de manière détaillée pour elles-mêmes et pour leurs implications. Les premières idées sur ces questions ont été avancées par J. Geffroy et D. Bosq mais n'ont pas fait l'objet, à notre connaissance, d'autre publication que [17].

Soit (E, \mathcal{B}, P) un modèle statistique où E est un espace topologique métrisable, \mathcal{B} sa tribu borélienne et soit ψ un paramètre défini sur P et à valeurs dans un espace vectoriel réel Θ . Soit B un borélien fixé non vide de E , B^c son complémentaire et $\mathcal{O}(B)$ l'ensemble des ouverts contenant B . Pour $U \in \mathcal{O}(B)$ on note P_U l'ensemble des éléments de P qui chargent U . Pour $C \in \mathcal{B}$ et $P \in P$ on note $P|_C$ la restriction de P à C et, si $P(C) > 0$, P_C la probabilité induite par P sur C ; enfin, on note $\mathcal{P}(P, C)$ l'ensemble des éléments Q de P ayant même restriction que P à l'ensemble C .

Soit $m \in \mathbb{R}^+$; on dit que ψ est *homogène de degré m sur B* si pour tout $U \in \mathcal{O}(B)$ on a :

a) $\forall (P, Q) \in P_U^2 \quad (P_U = Q_U \implies [P(U)]^{-m} \psi(P) = [Q(U)]^{-m} \psi(Q))$

b) $\forall P \notin P_U, \psi(P) = 0$.

Pour Θ fixé, on notera $H_m(B)$ l'ensemble des paramètres homogènes de degré m sur B .

La définition de la *localisation* s'obtient à partir de la précédente en remplaçant a) par

$$a') \quad \forall (P, Q) \in \mathcal{P}_U^2 \quad (P|_U = Q|_U \implies \psi(P) = \psi(Q)) .$$

Tout paramètre homogène est localisé, sans que la réciproque soit vraie en général (elle l'est sous certaines conditions pour un paramètre linéaire ; on a alors $m = 1$) ; en dehors de cas très particuliers le degré d'homogénéité est unique. Les notions que nous venons de définir s'appliquent à de nombreux paramètres, ce qui fait l'intérêt de cette étude : sous des hypothèses très générales, la valeur de la fonction de répartition, celle d'une version de la densité ou de ses dérivées en un point x , celle de leurs puissances sont des paramètres homogènes ; la valeur de la fonction frontière (resp. de la régression) en un point a de \mathbb{R} est un paramètre (fortement) localisé par la demi-droite $x = a, y \geq 0$ (resp. la droite $x = a$). Par contre, en général, les quantiles, les moments d'une loi de probabilité, de même que sa fonction de répartition ou sa densité sont des paramètres globaux. La notion de localisation est un cas particulier d'une notion plus générale, extension de la notion de classe déterminante, qui s'exprime en termes fonctionnels : si G est un ensemble de fonctions P -intégrables, on dit que G détermine ψ si l'égalité $\psi(P) = \psi(Q)$ a lieu dès que $(\forall g \in G \int g dP = \int g dQ)$. On peut remarquer que la notion de classe déterminant la convergence s'étend de façon similaire à d'autres paramètres que la loi.

On étudie ensuite les liens entre *localisation*, *homogénéité* et *estimabilité sans biais* lorsque B est un fermé de E et que \mathcal{P} vérifie la condition (C_1) ($\forall U \in \mathcal{O}(B) (P \in \mathcal{P}_U \implies P_U \in \mathcal{P}_U)$). Au besoin, dans de nombreux cas on pourra agrandir l'ensemble \mathcal{P} pour que (C_1) soit vérifiée ; il est clair qu'une mesure μ dominant \mathcal{P} domine également $\mathcal{P}' = \mathcal{P} \cup \{P_U \mid P \in \mathcal{P}_U, U \in \mathcal{O}(B), P \in \mathcal{P}\}$, ainsi que son convexifié. On montre alors qu'un paramètre ψ estimable sans

biais d'ordre ℓ , de noyau T_ℓ , appartient à $H_\ell(B)$ si et seulement si $T_\ell 1_B^\ell$ est un noyau pour ψ . Il en résulte, dans le cas où B est \mathcal{P} -négligeable qu'un paramètre homogène estimable sans biais a un degré d'estimabilité strictement supérieur à son degré d'homogénéité, ce qui constitue un résultat intéressant dans le domaine de la recherche des estimateurs non biaisés. En utilisant conjointement le théorème de Bickel-Lehmann on montre ainsi que lorsque \mathcal{P} est convexe et que $\psi(\alpha P + (1-\alpha) Q)$ est un polynôme en α de degré inférieur ou égal au degré d'homogénéité de ψ , il n'existe pas d'estimateur sans biais du paramètre ψ . Ceci pourra s'appliquer par exemple à la densité et à ses dérivées. Enfin, sous certaines conditions l'estimabilité sans biais d'un paramètre localisé entraîne l'homogénéité de ce paramètre.

La *fonction influence* donne une bonne idée de "l'importance" que va avoir une observation x sur l'estimation d'un paramètre. Sa détermination est un premier pas vers l'étude de l'estimation robuste des paramètres homogènes. Malheureusement elle ne peut avoir lieu que sous des hypothèses très restrictives sur \mathcal{P} , rarement satisfaites en estimation fonctionnelle : on doit supposer que l'on peut définir la dérivée de ψ en tout point P de \mathcal{P} dans la direction de toute mesure de Dirac δ_x , $x \in E$. Nous donnons néanmoins quelques exemples d'application. Dans le cas le plus intéressant (si \mathcal{P} charge tout $U \in \mathcal{O}(B)$ et si $x \notin B$) on a la relation : $I C_{\psi, P}(x) = -m \psi(P)$.

Une idée naturelle est de considérer des statistiques ayant de bonnes propriétés locales afin de définir de nouveaux estimateurs ou d'améliorer des estimateurs connus en utilisant le conditionnement par rapport à ces statistiques. Une statistique sur (E, B) est dite *B-exhaustive* si, pour tout $P \in \mathcal{P}$, elle est exhaustive au sens usuel dans le modèle $(E, B, \mathcal{P}(P, B^c))$. On définit de même les notions locales de B -liberté et B -complétion. C'est surtout la notion de B -exhaustivité qui est utilisée dans la suite. Intuitivement on peut dire qu'une

fois que l'on s'est fixé $Q|_{B^c}$, une statistique B -exhaustive résume l'information que l'on peut avoir sur $Q|_B$, c'est-à-dire ce qui manque pour connaître entièrement la loi Q . Dans le cas de l'observation de n v.a. i.i.d. X_1, \dots, X_n à valeurs dans (E, \mathcal{B}) , nous nous intéresserons aux statistiques qui seront U_n^n -exhaustives dans le modèle $(E, \mathcal{B}, P)^{\otimes n}$, où $(U_n)_{n \in \mathbb{N}}$ est une suite de $\mathcal{O}(B)$ décroissant vers B (ce qui correspond au classique problème du choix de la fenêtre en estimation fonctionnelle). On peut remarquer que la définition que nous venons de donner est identique à celle que l'on obtiendrait en considérant tout élément de P comme un couple $(P|_B, P|_{B^c})$ et en traitant $P|_{B^c}$ comme un paramètre fantôme. Cette considération intuitive manque toutefois de pertinence car $P|_B$ et $P|_{B^c}$ sont en général fortement liés et le qualificatif "fantôme" ne s'applique pas de façon stricte au paramètre $P|_{B^c}$. Toute statistique exhaustive est bien sûr B -exhaustive et les notions habituelles se retrouvent pour $B = E$. Nous particularisons ensuite notre étude à un exemple-clé : dans le cas d'un modèle dominé, une statistique S qui retient les observations tombant dans B et qui censure les autres est B -exhaustive. Une façon classique d'améliorer un estimateur est, d'après le théorème de Rao-Blackwell, de le remplacer par son espérance conditionnelle par rapport à une statistique exhaustive. Nous étudions donc dans la suite les effets du conditionnement par rapport à la statistique S sur certaines familles d'estimateurs.

Les v.a.r. indépendantes X_1, \dots, X_n sont supposées de même loi Q inconnue appartenant à un ensemble \mathcal{P} de probabilités sur $(\mathbb{R}, \mathcal{B}_{\mathbb{R}})$ et $\psi : F \rightarrow \mathbb{R}$ est un paramètre localisé par un fermé $F \neq \mathbb{R}$. On se fixe un ouvert propre U appartenant à $\mathcal{O}(F)$ et une probabilité $P \in \mathcal{P}$ telle que $P(U^c) > 0$. Dans les applications cette probabilité proviendra en général d'une pré-estimation ou d'une information a priori. On se donne un estimateur d'ordre n de la forme

$$T(X_1, \dots, X_n) = \sum_{i=1}^n K_i(X_i)$$
 du paramètre $\psi(Q)$, où les $(K_i)_{1 \leq i \leq n}$ (qui dépendent de n) sont des fonctions de \mathbb{R} dans \mathbb{R} P -intégrables. On note \mathcal{B}_S la tribu

engendrée par S . Le schéma d'estimation ainsi fixé s'applique à de nombreux problèmes ; en particulier à l'estimation, séquentielle ou non, de la densité ou de certains paramètres de régression. On montre qu'une version de

$E_{P^{\otimes n}}^{B_S}(T(X_1, \dots, X_n))$ s'exprime de la même façon que T en remplaçant K_i par

$$\tilde{K}_i = K_i \mathbb{1}_U + m_i(P) \mathbb{1}_{U^c}, \quad \text{où } m_i(P) = [P(U^c)]^{-1} \int_{U^c} K_i dP : \text{ on conserve } K_i$$

sur U et on le remplace par sa valeur moyenne pour la probabilité P à l'extérieur de U . (La méthode nécessitera donc l'utilisation de K_i n'annulant pas les

$m_i(P)$). Des calculs standards permettent de comparer les biais et les variances de $E_{P^{\otimes n}}^{B_S}(T)$ et de préciser la classe des lois pour lesquelles on a diminution

du risque quadratique. Ces considérations sont ensuite appliquées à l'estimation par la méthode du noyau de la densité et de certains paramètres de régression

(de la forme $\psi(Q) = \int_{\mathbb{R}} v^\alpha f_Q(x, v) dv$, où f_Q est une version de la densité

sur \mathbb{R}^2 de n couples observés, indépendants et de même loi Q ; le paramètre

$\int_{\mathbb{R}} f_Q(x, v) dv$ est supposé connu et strictement positif). Nous nous contenterons

ici de détailler la démarche suivie dans le cas de la densité. Les éléments Q

de \mathcal{P} sont supposés admettre une densité par rapport à la mesure de Lebesgue

sur \mathbb{R} , dont une version f_Q est bornée et continue. On cherche à estimer

$\psi(Q) = f_Q(x)$ où x est un réel fixé. Pour $u \in \mathbb{R}$ et $i \in \overline{1, n}$ on pose

$K_i(u) = K(u) = (nh)^{-1} H((u-x)h^{-1})$ où H est λ -intégrable. Le voisinage U

du point x considéré est l'intervalle $]x - h, x + h[$ (le réel strictement

positif h est bien sûr fonction de n . On commence par estimer f_Q par f_P

en utilisant une méthode globale puis on estime $f_Q(x)$ à l'aide de T que l'on

conditionne ensuite par S en ayant fixé $P^{\otimes n}$ comme probabilité sur $(\mathbb{R}^n, \mathcal{B}_{\mathbb{R}^n})$

ce qui permet, grâce aux résultats précédents, d'améliorer l'estimation de

$f_Q(x)$, en tenant compte du fait que f_P est une "bonne estimation" de f_Q

en dehors de U . On pourra également utiliser cette méthode pour une estimation

globale de f_Q lorsque l'on constate des irrégularités locales (bosses) pour la première estimation f_P . Une méthode de préestimation souvent utilisée sera celle de P. Deheuvels et P. Hominal ([19]). On étudie le comportement asymptotique du biais et du risque, sous des hypothèses complémentaires de régularité, ce qui sera utilisé dans un prochain travail et permettra de guider le choix de H et h .

Les problèmes d'*optimalité asymptotique* pour les méthodes à noyau conduisent à minimiser des fonctions des moments du noyau et donc à des problèmes variationnels dans un espace de Riesz. Nous donnons des conditions suffisantes pour qu'un élément d'un tel espace soit solution d'un problème de minimisation qui couvre beaucoup de cas particuliers rencontrés en statistique. Comme application nous retrouvons un résultat récent de L. Devroye qui contient le théorème d'optimalité de Bartlett-Epanechnikov. Nous reviendrons prochainement sur cette question en liaison avec les résultats précédents et ceux de H. Ramlau-Hansen ([22]) en statistique des processus de comptage.

Cette première partie constituera le début d'un travail beaucoup plus important : celui du classement des paramètres et du choix des statistiques à utiliser pour chaque classe (en particulier possibilité de comparer les informations qu'elles apportent donc de définir un "degré d'exhaustivité" utilisable en pratique). Il ne nous semble pas que les propriétés métriques de l'espace des paramètres et la dépendance plus ou moins forte des variables expliquent la totalité des problèmes posés. Intervient également ce que l'on pourrait appeler (en l'ayant correctement définie !) la nature du paramètre, concept qui prendrait en compte les propriétés locales que nous avons définies. Enfin, il nous reste à explorer les problèmes de tests concernant les paramètres homogènes ou localisés.

Dans la deuxième partie nous reprenons des résultats antérieurs ([3] et [4]) sur les *estimateurs splines* de la densité afin d'obtenir leur loi limite. La notion de noyau spline nous avait permis de montrer comment les méthodes splines s'intègrent à la théorie générale des estimateurs de la forme

$\frac{1}{n} \sum_{k=1}^n K_{r(n)}(X_k, t)$ développée par J. Bleuez et D. Bosq ([16]). Les majorations obtenues pour les noyaux permettent d'obtenir des conditions suffisantes de convergence ponctuelle en moyenne quadratique et presque complète, et de convergence uniforme presque complète dans le cas de subdivisions quelconques (la densité est supposée avoir une version suffisamment régulière à support dans $[0,1]$). Dans le cas de subdivisions uniformes, des conditions nécessaires et suffisantes de convergence simple ou uniforme suivant différents modes stochastiques peuvent être données, ainsi qu'un encadrement de la vitesse de convergence. La *loi limite* est donnée dans le cas de splines cubiques interpolant la fonction de répartition empirique aux noeuds d'une subdivision uniforme et ayant une dérivée seconde nulle aux bornes de l'intervalle $[0,1]$.

Les méthodes splines ont connu une vogue rapide en statistique, à cause de leur simplicité de mise en œuvre, de la régularité des courbes obtenues et de la multiplicité des conditions que l'on peut imposer aux solutions. Cet élan semble aujourd'hui stabilisé. Les splines ont pris leur place à côté des méthodes classiques d'estimation (d'inspiration "plus statistique") et ont un champ d'application assez bien cerné : les fonctions à estimer doivent obéir à certaines conditions de régularité et le nombre d'observations doit être suffisamment important pour éviter les phénomènes classiques de sur ou sous-lissage, que l'on utilise des splines d'interpolation, d'ajustement, isotoniques ([23]) ... Leur utilisation peut très bien être envisagée dans la phase d'estimation globale décrite dans la première partie ou dans celle d'estimation spectrale évoquée dans la quatrième partie.

De par leur définition les splines cubiques d'interpolation que nous avons utilisées semblent pouvoir jouer un rôle utile dans l'estimation de certains paramètres qui sont des fonctionnelles simultanées de la fonction de répartition F et de la densité f . Nous étudions actuellement l'estimation consistante de ce type de paramètres.

La troisième partie est consacrée à la fonctionnalisation d'un résultat d'*approximation forte* de Philipp et Pinzur ([2]). Nous montrons que l'on peut construire un *processus de Kiefer* généralisé, indexé par $G \times [1, +\infty[$ où G est un ensemble de fonctions sur $[0,1]^d$, approximant presque sûrement les intégrales stochastiques des éléments de G par rapport au processus empirique associé à des variables aléatoires $(X_n)_{n \geq 1}$ à valeurs dans $[0,1]^d$ fortement mélangeantes et strictement stationnaires. Soit F la fonction de répartition de la loi π de X_1 , F_N la fonction de répartition empirique associée à $(X_n)_{1 \leq n \leq N}$ et $R(s,t) = \tilde{t} (F_N(s) - F(s))$, $s \in [0,1]^d$, $t \in [1, +\infty[$ le processus empirique de $(X_n)_{n \geq 1}$. (\tilde{t} est la partie entière de t). On suppose que le coefficient de forte mélangeance de la suite est $\rho(n) = O(n^{-4-d(1+\varepsilon)})$ ($\varepsilon \in]0, \frac{1}{4}]$). Cette hypothèse peut paraître artificielle et sera probablement affaiblie dans l'avenir, mais est actuellement indispensable. Philipp et Pinzur ont montré que sous ces hypothèses et sans changer sa distribution, on peut redéfinir le processus $R(s,t)$, $s \in [0,1]^d$, $t \in [1, +\infty[$, sur un espace de probabilité (Ω, \mathcal{A}, P) sur lequel il existe un processus de Kiefer $K(s,t)$, $s \in [0,1]^d$, $t \in [1, +\infty[$ vérifiant

$$(1) \quad \sup_{1 \leq t \leq T} \sup_{s \in [0,1]^d} |R(s,t) - K(s,t)| = O(T^{1/2} (\log T)^{-\lambda}) \text{ p.s.}$$

et de fonction de covariance $m(t,t') \Gamma(s,s')$ où

$$\Gamma(s,s') = E(g_1(s) g_1(s')) + \sum_{n=2}^{+\infty} E(g_1(s) g_n(s') + g_n(s) g_1(s'))$$

$$\text{avec} \quad g_n(s) = 1_{\{X_n \leq s\}} - F(s) .$$

λ est une constante strictement positive ne dépendant que de d et ε .

Il semble bien qu'une étude approfondie des techniques utilisées par Philipp et Pinzur et Berkes et Philipp [2] permettra d'obtenir une approximation presque complètement sûre. Nous étendons la définition de R et K à

$G \times [1, +\infty[$, où G est un ensemble de fonctions réelles sur $[0,1]^d$ de façon à ce que la qualité de l'approximation soit conservée. La première extension de ce type a été faite par P. Revesz en 1976. Dans le cas de variables i.i.d. il a montré comment l'on pouvait passer d'une approximation presque sûre dans $[0,1]^2$ par une suite de ponts browniens ou de processus de Kiefer à une approximation presque sûre dans un ensemble d'indicatrices de boréliens de frontière suffisamment régulière. Dans le même cadre i.i.d., Ibero, en s'inspirant de la méthode de Revesz est passé à la classe des fonctions p fois différentiables sur $[0,1]^d$, bornées, ainsi que leurs dérivées partielles jusqu'à l'ordre p ($p > \frac{d}{2}$). Cette technique ne convient pas ici, comme nous le montrons dans le paragraphe VI : la vitesse d'approximation dans (1) ne nous a pas permis d'obtenir un résultat satisfaisant.

Pour $g \in L^1(\pi)$ $R(g,t)$ est l'intégrale stochastique de g par rapport au processus empirique $R(s,t)$:

$$R(g,t) = \int_{[0,1]^d} g(s) dR(s,t) = \sum_{i=1}^{\sim t} G_i(g)$$

en posant $G_i(g) = g(X_i) - \int g dF$.

L'extension que nous proposons est basée sur une seconde méthode due également à Ibero et qui repose sur la représentation suivante de $R(g,t)$ pour $g \in C^d$ (ensemble des fonctions d fois continûment différentiables sur $[0,1]^d$) :

$$R(g,t) = \sum_{p=1}^d \sum_{1 \leq i_1 < \dots < i_p \leq d} (-1)^p \int_{[0,1]^p} \frac{\partial^p g}{\partial x_{i_1} \dots \partial x_{i_p}} (1, \dots, 1, x_{i_1}, \dots, x_{i_p}, 1, \dots, 1) R((1, \dots, 1, x_{i_1}, \dots, x_{i_p}, 1, \dots, 1), t) dx_{i_1} \dots dx_{i_p} .$$

Pour pouvoir définir $K(g,t)$ de la même façon à partir de $K(s,t)$, il suffit donc de montrer que le processus $K(s,t)$, $s \in [0,1]^d$, $t \in [1, +\infty[$ est presque sûrement à trajectoires continues. Pour cela on majore sa fonction de covariance à l'aide d'une inégalité sur les moments de variables

mélangeantes et on conclut par le lemme de Fernique. On peut ainsi définir un processus de Kiefer $K(g,t)$ indexé par $C^d \times [1, +\infty[$, presque sûrement linéaire en g et tel que

$$(2) \quad \sup_{1 \leq t \leq T} \sup_{g \in C_M^d} |R(g,t) - K(g,t)| \leq CM T^{1/2} (\text{Log } T)^{-\lambda} \quad \text{p.s.}$$

C_M^d est le sous-ensemble de C^d constitué des fonctions bornées par M , ainsi que leurs dérivées partielles jusqu'à l'ordre d . La fonction de covariance de $K(g,t)$ est donnée par

$$E(K(g,t) K(g',u)) = \min(t,u) \Lambda(g,g')$$

$$\text{avec} \quad \Lambda(g,g') = E(G_1(g) G_1(g')) + \sum_{n=2}^{+\infty} E(G_1(g) G_n(g') + G_n(g) G_1(g'))$$

On peut alors étendre K aux limites de suites de C_M^d pour différentes topologies. Nous étudions le cas de la convergence uniforme sur $[0,1]^d$ et nous obtenons pour les éléments de $\overline{C_M^d}$ un résultat du même type que (2) mais non uniforme sur la classe de fonctions. Il est sans doute possible d'obtenir (2) sur $\overline{C_M^d}$ en étudiant plus finement les trajectoires de $K(g,t)$, $g \in C_M^d$, $t \in [1, +\infty[$. Sous des hypothèses convenables la mesure empirique peut-être considérée comme une variable aléatoire à valeurs dans un espace à noyau reproduisant H ([5]). On peut donc définir un processus $(R_t)_{t \in [1, +\infty[}$ à valeurs dans H tel que pour tout t , R_t soit le représentant dans H du processus $R(g,t)$, $g \in H$. On appelle H_0 le sous-espace dense de H engendré par les évaluations et on suppose que $H_0 \subset C^d$. Si l'application $g \mapsto K(g,t)$ est presque sûrement bornée sur la boule unité de H_0 , elle est représentée presque sûrement par un élément K_t de H . Ceci permet de définir un processus de Kiefer à valeurs dans H tel que

$$\sup_{t \in [1, T]} \|R_t - K_t\| = O(T^{1/2} (\text{Log } T)^{-\lambda}) \quad \text{p.s.}$$

$$E((g, K_t) (g', K_u)) = \min(t, u) \Lambda(g, g').$$

Cette troisième partie débouche à court terme sur l'élaboration et l'étude de tests fonctionnels et à plus long terme sur un essai d'amélioration des vitesses d'approximation.

La dernière partie a pour sujet l' *estimation des degrés des modèles* ARMA et l'étude de différents *algorithmes* pouvant intervenir efficacement dans le traitement des séries chronologiques uni et multivariées. Elle est constituée principalement des articles [9] et [13] auxquels nous avons adjoint une courte annexe contenant des éléments des prépublications [7] et [10] non repris dans les articles publiés.

Le chapitre cinq est essentiellement consacré au *cas scalaire* ; nous y passons en revue les différentes méthodes d'estimation des degrés d'un processus ARMA directement à partir de la fonction d'autocorrélation empirique (ou à partir d'un estimateur plus sophistiqué de cette fonction) après observation de X_0, X_1, \dots, X_T . Les méthodes que nous décrivons ne nécessitent aucune estimation préalable des coefficients du processus, au contraire des différents critères usuels de choix de modèles (critères d'information d'Akaïke) qui obligent à faire varier les degrés et à estimer à chaque fois tous les coefficients du modèle, ce qui peut représenter beaucoup de calculs inutiles. Dans le cas scalaire les méthodes étudiées sont basées sur le calcul de certains déterminants et reposent sur le théorème fondamental liant la représentation ARMA minimale d'un processus du second ordre et la propriété, pour sa fonction d'autocorrélation, de satisfaire une équation aux différences finies minimale. Les algorithmes utilisés sont ceux capables de détecter le rang minimal et l'ordre minimal d'une telle équation. Nous commençons par l'*epsilon-algorithme* car il est connu et étudié de façon intensive depuis 1956 si bien que les résultats sur cet algorithme nous permettent d'obtenir de façon simple et directe les propriétés des deux autres méthodes envisagées : l'*algorithme* R-S ([20]) et la *méthode du coin* ([1]). Ces méthodes ne doivent pas être vues comme concurrentes et pourront, grâce à leur faible coût, être utilisées simultanément quand un doute subsiste. Néanmoins l'*epsilon-algorithme* nous semble le plus simple des trois et a cet avantage de passer très facilement au cas multivarié. Nous rappelons quelques propriétés bien

connues en analyse numérique dont celles de la transformation de Shanks qui est une transformation non linéaire de suite à suite, qui possède la propriété recherchée pour les suites récurrentes et qui peut être effectuée très simplement à l'aide de l'épsilon-algorithme. Ceci permet d'obtenir le résultat constituant la base du chapitre cinq : le processus du second ordre X admet une *représentation* ARMA(p,q) *minimale* si et seulement si

$$(P) \quad \begin{cases} \varepsilon_{2p}^n(\rho) = 0 & \forall n \geq q - p + 1 \\ \text{et } \varepsilon_{2p}^{q-p}(\rho) \neq 0 \end{cases}$$

où ρ est la fonction d'autocorrélation de X . Le tableau epsilon construit à partir de ρ a donc une structure bien marquée, particulièrement reconnaissable. Pour estimer p et q à partir d'observations X_0, \dots, X_T d'un ARMA inconnu on calcule l'autocorrélation empirique, on lui applique l'épsilon-algorithme, et on a à détecter une variation brusque de l'ordre de grandeur des éléments du tableau epsilon, en passant d'une colonne paire à la suivante (de numéro $2p$) et en passant d'un élément de cette colonne (ε_{2p}^{q-p}) au suivant. La plupart des simulations ont montré que les degrés du processus engendré étaient correctement reconnus, ou figuraient parmi un petit nombre de possibilités à retenir. Bien entendu la méthode n'est pas miraculeuse, quand ρ est mal approximé on ne peut rien conclure à partir du tableau. D'autre part le fait d'obtenir plusieurs possibilités est inhérent au modèle lui-même, en particulier aux ordres de grandeur des coefficients de différents ARMA pouvant "bien approcher" le processus observé. Le fait d'avoir pu sélectionner un petit nombre de possibilités pour lesquelles on fait une étude plus sophistiquée représente un gain important. Nous donnons des exemples et nous discutons les aspects pratiques de la méthode et la possibilité de tests numériques. Notons qu'il est souvent utile de transformer la suite des autocorrélations empiriques avant application de l'algorithme, pour favoriser la détection des sauts dans le tableau ε . Une autre propriété du tableau est exploitée : l'*invariance* de la quantité $I(n) = \sum_{i=0}^{2p-2} (-1)^i \varepsilon_i^n \varepsilon_{i+1}^n$ pour

$n \geq q - p + 1$; elle permet d'étayer la première analyse. En utilisant, sous des hypothèses convenables, les résultats classiques sur les autocorrélations empiriques, on obtient la *convergence* presque sûre des éléments du tableau ϵ vers les éléments du tableau construit à partir de la véritable fonction d'autocorrélation du processus. Il est donc possible de définir une stratégie théorique de sélection assurant la convergence presque sûre de (\hat{p}, \hat{q}) vers (p, q) . On donne enfin le *comportement en loi* des vecteurs $T^{1/2} (\epsilon_{2p}^{n_1}, \dots, \epsilon_{2p}^{n_k})$, $k \in \mathbb{N}^*$, $n_i \geq q - p + 1$, $1 \leq i \leq k$, sous l'hypothèse nulle : $\epsilon_{2p}^n = 0 \quad \forall n \geq q - p + 1$ ce qui conduit à des tests du χ^2 sur les colonnes paires du tableau (on suppose que l'on connaît un majorant des degrés p et q ; ceci n'est pas restrictif en pratique, car on abandonne le modèle ARMA lorsque l'on doit utiliser des degrés trop élevés). Nous terminons ce chapitre 5 en montrant comment les caractérisations données pour l'algorithme RS et la méthode du coin s'obtiennent à partir des résultats obtenus pour l'epsilon-algorithme et en donnant un court aperçu des méthodes multivariées qui seront plus largement évoquées au chapitre six. L'annexe un contient quatre exemples déterministes en double précision et un lemme de simplification des équations aux différences stochastiques.

Dans l'article [13] qui constitue la majeure partie du chapitre six, nous avons repris très brièvement quelques propriétés essentielles de l'epsilon-algorithme scalaire en présentant une nouvelle application aux fonctions de transfert rationnelles (exemple J de Box et Jenkins [18]). Nous avons montré dans le chapitre précédent la possibilité de calculer l'*autocorrélation partielle* grâce à l'algorithme q-d de Rutishauser. Nous montrons ici sans faire appel à la théorie spectrale, comment les algorithmes de calcul des polynômes orthogonaux généralisés (q-d et $\tilde{q-d}$) permettent le calcul des ajustements autorégressifs aux données dans le cas scalaire (le cas multivarié est traité en annexe 2).

La théorie des *modèles ARMA multivariés* pose des problèmes ardues à cause de la multiplicité des représentations. Il est impossible de définir en toute généralité une représentation minimale, comme on l'a fait dans le cas scalaire. On pourra définir des représentations minimales dans certaines classes d'ARMA, chacune de ces classes étant obtenue en imposant des restrictions aux coefficients du modèle. Un processus ARMA donné peut très bien ne pas avoir de représentation dans une telle classe et a fortiori ne pas avoir de représentation minimale. Nous avons défini ici une *minimalité* qui consiste à choisir d'abord q le plus petit possible et ensuite p le plus petit possible, ce qui peut se justifier par l'argument suivant : on veut expliquer à chaque instant le processus davantage par ses propres valeurs passées (observables) que par celles du bruit (non observables). Le reste de notre étude est basé sur cette définition. Nous avons l'intention de revenir très prochainement sur ces questions en liaison avec les propriétés des suites de matrices de covariances car il nous semble indispensable d'envisager un certain nombre "d'autres minimalités" qui s'adapteront aux problèmes particuliers que l'on aura à traiter. On peut très bien imaginer que, tel problème physique s'expliquant par des chocs successifs modélisés par le bruit, l'on préfère minimiser d'abord p et ensuite q . D'autres restrictions sur les paramètres peuvent également intervenir : on peut imposer aux coefficients d'être inversibles, d'être des matrices triangulaires, etc ...

Nous utilisons dans ce cadre multivarié les versions matricielle et vectorielle de l'epsilon-algorithme, pour lesquelles on a, comme dans le cas scalaire, la propriété (P) et l'existence d'invariants. Dans le cas de l'*epsilon matriciel*, et sous certaines hypothèses, la condition est encore nécessaire et suffisante (théorème 10 et annexe 2). Dans le cas de l'*epsilon vectoriel* elle doit être légèrement modifiée et n'est plus qu'une condition nécessaire. Les théorèmes de convergence et les tests statistiques peuvent être étendus au cas multivarié avec les modifications appropriées.

Les différents programmes d'application mis au point lors de cette étude seront intégrés, à côté des méthodes classiques, paramétriques ou non, dans un logiciel de traitement automatique des séries chronologiques, qui, nous l'espérons, pourra rendre service aux praticiens confrontés à des problèmes d'estimation et de prévision.

BIBLIOGRAPHIE DE L'INTRODUCTION.

- [1] BEGUIN, J.M., C. GOURIEROUX et A. MONFORT : Identification of a mixed autoregressive-moving average process : the corner method - Time series, Anderson, O.D. ed., North-Holland, Amsterdam, 1980, 423-436.
- [2] BERKES, I. et PHILIPP, W. : Approximation theorems for independent and weakly dependent random vectors. Ann. of Prob., 7, 1979, p. 29-54.
- [3] BERLINET, A. : Sur les méthodes splines en estimation de la densité. C.R.A.S., t 288, série A, 1979, p. 847-850.
- [4] BERLINET, A. : Espaces autoreproduisants et mesure empirique. Méthodes splines en estimation fonctionnelle. Thèse de 3ème cycle. Université de Lille, 1980.
- [5] BERLINET, A. : Variables aléatoires à valeurs dans les espaces autoreproduisants et mesure empirique. C.R.A.S., t 290, série A, 1980, p. 973-975.
- [6] BERLINET, A. : Convergence des estimateurs splines de la densité. Pub. de l'I.S.U.P., Vol. 26, fasc. 2, 1981.
- [7] BERLINET, A. : Une méthode de détermination des degrés d'un modèle ARMA. Pub. IRMA , Vol. 3, Fasc. 6, Lille 1981. Exposé aux Journées de Statistique de Bruxelles (24-27 mai 1982).
- [8] BERLINET, A. : Degrees of an ARMA model : estimation and tests. Compstat - Toulouse 1982. Part. II. Physica-Verlag.
- [9] BERLINET, A. : Estimating the degrees of an ARMA model. A paraître dans Compstat Lectures, 1984.
- [10] BERLINET, A. : Estimation des degrés d'un ARMA multivarié. Pub. IRMA Lille, vol. 4, fasc. 3, 1982.
- [11] BERLINET, A. : Localisation et homogénéité d'un paramètre. Pub. IRMA Lille, vol. 5, fasc. 1, 1983.
- [12] BERLINET, A. : Exhaustivité locale d'une statistique. Application au conditionnement d'estimateurs. Pub. IRMA, Lille, vol. 5, fasc. 5, 1983.
- [13] BERLINET, A. : Some useful algorithms in time series analysis. Dans "Alternative approaches to time series analysis". Pub. des Facultés Universitaires Saint-Louis, Bruxelles, 1984, p. 95-120.

- [14] BERLINET, A. : Propriétés locales d'un paramètre. Application à l'estimation.
C.R.A.S., t 298 , série I, 1984, p. 345-348.
- [15] BERLINET, A. : Approximation forte d'intégrales stochastiques empiriques multidimensionnelles en variables mélangeantes.
Pub. IRMA Lille, vol. 6, fasc. 3, 1984.
- [16] BLEUEZ, J. et BOSQ, D. : Etude d'une classe d'estimateurs non paramétriques de la densité.
Ann. Inst. Henri Poincaré, 14, n° 4, 1978, p. 479-498.
- [17] BOSQ, D. : Contribution à la théorie de l'estimation fonctionnelle.
Thèse. Pub. de l'I.S.U.P., 1970.
- [18] BOX, G.E.P. et G.M. JENKINS : Time Series Analysis - Forecasting and Control. Holden-Day, San Francisco, 1976.
- [19] DEHEUVELS, P. et HOMINAL, P. : Estimation automatique de la densité.
Revue de Stat. Appl., vol. 28, n° 1, 1980, p. 25-55.
- [20] GRAY, H.L., G.D. KELLEY and D.D. MAC INTIRE : A new approach to ARMA modeling
Comm. in Stat. Simul. and Comp., B7, 1978, 1-77.
- [21] PHILIPP, W. et PINZUR, L. : Almost sure approximation theorems for the multivariate empirical process, Z.f.W., 54, 1980, p. 1-13.
- [22] RAMLAU-HANSEN, H. : The choice of a kernel function in the graduation of counting process intensities.
Scand. Actuarial J., 1983, p. 165-182.
- [23] WEGMAN, E.J. : Splines in statistics, JASA, 78, n° 382, 1983, p. 351-365.

PREMIERE PARTIE.

CHAPITRE 1

LOCALISATION ET HOMOGENEITE D'UN PARAMETRE

Résumé : De nombreux exemples de la statistique amènent à considérer des propriétés de paramètres que l'on peut qualifier de "locales dans l'espace des observations". Le but de ce travail est de définir et d'étudier de telles notions de localisation et leurs liens avec les concepts habituels de la statistique ; on abordera ici la linéarité, l'estimabilité sans biais et la fonction influence.

CLASSIFICATION AMS (MOS) : 62G99

MOTS CLÉS : localisation, homogénéité, non-paramétrique, estimabilité sans biais, fonction influence.

Abstract : This paper deals with local properties of parameters. We define the localization and the homogeneity of a parameter in the sample space and we study their links with usual concepts such as linearity, unbiased estimation and influence curve.

P L A N.

	<u>Pages</u>
I - <u>INTRODUCTION.</u>	1
II - <u>LOCALISATION ET HOMOGENEITE.</u>	2
II.1. - Notations.	
II.2. - Définition d'un paramètre homogène.	
II.3. - Exemples.	
II.4. - Degrés d'homogénéité.	
II.5. - Unicité du degré d'homogénéité.	
II.6. - Remarque sur la définition de l'homogénéité.	
II.7. - Définition de la localisation.	
II.8. - Remarque - Vocabulaire.	
II.9. - Exemples.	
II.10. - Remarques.	
II.11. - Classe fonctionnelle déterminant un paramètre.	
II.12. - Localisation forte. Liens entre homogénéité et localisation.	
II.13. - Cas d'un paramètre linéaire.	
III - <u>HOMOGENEITE, LOCALISATION ET ESTIMABILITE SANS BIAIS.</u>	18
III.1. - Relations vérifiées par un paramètre homogène et estimable sans biais.	
III.2. - Paramètre localisé estimable sans biais.	
IV - <u>FONCTION INFLUENCE D'UN PARAMETRE HOMOGENE.</u>	22
IV.1. - Cas où il existe $U \in O(F)$ tel que $P \notin \mathcal{D}_U$.	
IV.2. - Cas où $P \in \mathcal{D}_U$ pour tout $U \in O(F)$.	
<u>REFERENCES.</u>	25

I - INTRODUCTION.

Lorsque l'on évoque des propriétés locales d'un paramètre défini sur un ensemble de mesures \mathcal{P} et à valeurs dans un espace topologique \mathcal{M} , le mot "locales" a trait, en général, soit à l'ensemble \mathcal{P} et aux voisinages définis dans \mathcal{P} pour une certaine topologie, soit à l'espace \mathcal{M} . Nous nous intéresserons ici à des propriétés locales dans un troisième ensemble : celui sur lequel les lois de \mathcal{P} sont définies.

Considérons comme exemple introductif l'ensemble \mathcal{P} des lois de probabilité P sur \mathbb{R} absolument continues par rapport à la mesure de Lebesgue, dont la densité admet une version continue $\psi_1(P) = f_P$ et qui admettent une espérance $\psi_2(P) = \int_{\mathbb{R}} x dP$. Notons $\psi_3(P) = F_P$ la fonction de répartition de P . Les paramètres ψ_1 , ψ_2 et ψ_3 ont un caractère "global" : sous la seule hypothèse que $P \in \mathcal{P}$, la connaissance de la restriction de P à un sous-ensemble E de \mathbb{R} tel que $P(E) < 1$ ne nous permet pas de déterminer les images de P par ψ_1 , ψ_2 ou ψ_3 . Il en va autrement pour les paramètres qui associent à P les réels : $f_P(a)$ ou $F_P(a)$, où a est fixé dans \mathbb{R} : la connaissance de la restriction de P à un ensemble contenant $]-\infty, a[$ entraîne celle de $F_P(a)$ et la connaissance de la restriction de P à un voisinage de a entraîne celle de $f_P(a)$. On a même plus que cette propriété de localisation pour $f_P(a)$ et $F_P(a)$: ces paramètres ont une propriété d'homogénéité : si deux éléments P et Q de \mathcal{P} admettent des probabilités conditionnelles $P(\cdot|U)$ et $Q(\cdot|U)$ par rapport à un voisinage U de $\{a\}$ (resp. de $]-\infty, a[$) qui sont égales, alors

$$\frac{f_P(a)}{P(U)} = \frac{f_Q(a)}{Q(U)} \quad (\text{resp.} \quad \frac{F_P(a)}{P(U)} = \frac{F_Q(a)}{Q(U)})$$

L'incidence en statistique de la localisation est claire : si l'on cherche à estimer la valeur du paramètre pour une loi P inconnue dont on observe un échantillon (X_1, \dots, X_n) , les observations tombant dans ou "à proximité" de l'ensemble localisant le paramètre nous fourniront l'essentiel de l'information apportée par l'échantillon (à titre d'exemple on peut penser à la méthode de la fenêtre mobile pour la densité). Les notions quelque peu intuitives auxquelles nous venons de faire allusion dans cette introduction demandent bien sûr des définitions précises et des raisonnements rigoureux afin de déduire et d'étudier des classes de paramètres que l'on pourra estimer de façon analogue ; en particulier, avec les mêmes vitesses de convergence. En ce qui concerne ce dernier point, qui ne sera pas abordé ici, il nous semble que les propriétés de localisation constituent le troisième facteur à prendre en considération après les propriétés métriques de \mathbb{Q} et celles de dépendance plus ou moins forte des variables observées.

Les premières idées sur ces questions ont été avancées par J. Geffroy et D. Bosq mais n'ont pas fait l'objet, à notre connaissance, d'autre publication que la thèse de D. Bosq [1970], dont nous reprendrons quelques exemples et l'essentiel des définitions.

II - LOCALISATION ET HOMOGENEITE.-

II.1. - Notations.

Soit Ω un espace topologique métrisable, \mathcal{B}_Ω sa tribu borélienne.

\mathcal{P} l'ensemble des mesures de probabilité sur $(\Omega, \mathcal{B}_\Omega)$.

\mathcal{P}' l'espace vectoriel réel des mesures signées bornées sur $(\Omega, \mathcal{B}_\Omega)$.

\mathcal{D} un sous-ensemble non vide de \mathcal{P} , Θ un espace vectoriel réel et $\psi : \mathcal{D} \rightarrow \Theta$ un paramètre.

B un borélien fixé et non vide de Ω et $\mathcal{O}(B)$ l'ensemble des ouverts contenant B .

Pour $U \in \mathcal{O}(B)$ on note $\mathcal{O}_U(B)$ l'ensemble des ouverts contenant B et inclus dans U et \mathcal{D}_U l'ensemble des éléments P de \mathcal{D} tels que $P(U) > 0$.

Pour $C \in \mathcal{B}_\Omega$ et $P \in \mathcal{P}$ on note $P|_C$ la mesure bornée positive ou nulle sur Ω , restriction de P à C ($\forall A \in \mathcal{B}_\Omega$ $P|_C(A) = P(C \cap A)$)

si $P(C) > 0$ on note P_C la probabilité induite par P sur C (probabilité conditionnelle par rapport à C :

$$\forall A \in \mathcal{B}_\Omega \quad P_C(A) = \frac{P(A \cap C)}{P(C)} .$$

II.2. - Définition d'un paramètre homogène.

Soit $m \in \mathbb{R}^+$; on dit que ψ est un paramètre homogène de degré m sur B si et seulement si pour tout $U \in \mathcal{O}(B)$ on a :

$$a) \quad \forall (P, Q) \in \mathcal{D}_U^2 \quad (P_U = Q_U \implies \frac{\psi(P)}{[P(U)]^m} = \frac{\psi(Q)}{[Q(U)]^m})$$

$$b) \quad \forall P \notin \mathcal{D}_U \quad \psi(P) = 0 .$$

On notera $H_m(B)$ l'ensemble des paramètres homogènes de degré m sur B et à valeurs dans Θ fixé.

II.3. - Exemples.

II.3.1. - $\Omega = \mathbb{R}^p$, $\Theta = \mathbb{R}$, \mathcal{D} : ensemble des mesures de probabilité P absolument continues par rapport à $\lambda^{\otimes p}$ (λ mesure de Lebesgue sur \mathbb{R}) et à densité f_P continue sur un voisinage d'un point

a de \mathbb{R}^p . $\psi(P) = [f_p(a)]^m$ ($m \in \mathbb{R}_*^+$). ψ est un paramètre homogène de degré m sur le fermé $\{a\}$.

II.3.2. - $\Omega = \mathbb{Q} = \mathbb{R}$, \mathcal{D} : ensemble des mesures de probabilité P , absolument continues par rapport à λ , à densité f_p de classe C^k ($k \geq 1$) sur un voisinage de a . $\psi(P) = [f_p^{(k)}(a)]^m$ ($m \in \mathbb{R}_*^+$). ψ est homogène de degré m sur le fermé $\{a\}$.

II.3.3. - $\mathbb{Q} = \mathbb{R}$, $\mathcal{D} = \mathcal{P}$, $\psi(P) = [P(B)]^m$ ($m \in \mathbb{R}^{+*}$).

ψ est homogène de degré m sur le borélien B .

Si $\Omega = \mathbb{R}$, $B =]-\infty, a[$ et $m = 1$ on retrouve l'exemple de la fonction de répartition donné en introduction : $\psi(P) = F_p(a)$.

II.3.4. - Tout paramètre est homogène de tout degré sur Ω .

II.3.5. - Si ψ est homogène sur B , il l'est aussi sur tout borélien contenant B (en particulier sur \bar{B} ; dans la plupart des exemples B sera un fermé de Ω).

II.4. - Degrés d'homogénéité.

Proposition 1. - Soit $\psi \in H_m(B)$, \mathbb{Q}' un espace vectoriel réel et Ψ une application de \mathbb{Q} dans \mathbb{Q}' telle que $\Psi(0) = 0$.

Si $m = 0$, $\Psi \circ \psi$ est homogène de degré 0 sur B .

Si $m \neq 0$ et si Ψ est positivement homogène de degré $r \in \mathbb{R}^+$ ($\forall t \in \mathbb{R}_+^* \quad \forall x \in \mathbb{Q} \quad \Psi(tx) = t^r \Psi(x)$) le paramètre $\Psi \circ \psi$ est homogène de degré mr sur B .

La démonstration est immédiate. Cette proposition montre que pour certaines questions on pourra ramener l'étude des paramètres homogènes à ceux dont le degré est 0 ou 1 : soit $\psi \in H_m(B)$, $m > 0$ et

$E = (e_i)_{i \in I}$ une base vectorielle de \mathbb{Q} ; soit Ψ l'application de \mathbb{Q} dans lui-même ainsi définie :

$$* \Psi(0) = 0$$

$$* \text{ si } x \neq 0 \quad \Psi(x) = \sum_{i \in J_x} |x_i|^{1/m - 1} x_i$$

où J_x est l'ensemble des éléments i de I tels que la composante de x suivant e_i soit non nulle et égale à $x_i e_i$.

Ψ vérifie les hypothèses de la proposition 1 avec $r = \frac{1}{m}$ et le paramètre $\Psi \circ \psi$ est homogène de degré 1 sur B . Cette façon de nous ramener au degré 1 ne nous dispense pas toutefois de l'étude de $H_m(B)$, m quelconque, en raison des problèmes de biais et de vitesses de convergence qui ne sont pas facilement adaptables d'un paramètre réel à l'une de ses puissances. Il est aussi des cas où l'on est amené à poser un problème statistique en termes d'une puissance du paramètre, d'estimer cette puissance et de revenir ensuite au paramètre. Nous proposons l'exemple du maximum de vraisemblance pénalisé (MPLE) qui permet de se rendre compte que l'introduction de l'exposant $m \in \mathbb{R}^+$ n'est pas inutile.

Soit X_1, \dots, X_n un échantillon de v.a.r. (i.i.d.) de même loi P admettant une densité f_P dont une version a sa racine carrée dans l'espace de Sobolev $H^1(\mathbb{R})$.

Le premier MPLE de Good et Gaskins [G.F. de Montricher, R.A. Tapia et J.R. Thompson, 1975] est défini comme étant l'unique solution du problème :

$$P_1 : \left\{ \begin{array}{l} \text{Trouver } f \text{ maximisant } \sum_{i=1}^n \log f(X_i) - \alpha \phi(f) \text{ et} \\ \text{satisfaisant aux contraintes : } \int f = 1, \quad f \geq 0 \\ \sqrt{f} \in H^1(\mathbb{R}) \end{array} \right.$$

où $\alpha > 0$ et $\phi(f) = \int (f')^2 / f$ est l'information de Fisher.

La résolution directe du problème P_1 est très difficile. On lui substitue celle du problème P_2 dont l'unique solution est la racine carrée de la solution de P_1 ($v = \sqrt{f}$).

$$P_2 : \left\{ \begin{array}{l} \text{Trouver } v \text{ maximisant } 2 \sum_{i=1}^n \log v(X_i) - \alpha \|v'\|_{L^2}^2 \\ \text{et satisfaisant aux contraintes :} \\ v \in H^1(\mathbb{R}) ; \|v\|_{L^2} = 1 ; v(X_i) \geq 0 , 1 \leq i \leq n \end{array} \right.$$

Pour $a \in \mathbb{R}$ le problème d'estimation point par point est posé pour $[f_p(a)]^{1/2}$, cas particulier du II.3 a) et les résultats de convergence sont démontrés d'abord pour la solution v_n de P_2 .

II.5. - Unicité du degré d'homogénéité.

En dehors de cas très particuliers un paramètre homogène a un degré unique :

Proposition 2. - Soit $\psi \in H_m(B)$ tel qu'il existe $U \in \mathcal{O}(B)$ et $(P, Q) \in \mathcal{D}_U^2$ vérifiant $P_U = Q_U$ et $\psi(P) \neq \psi(Q)$. Alors m est l'unique degré d'homogénéité de ψ sur B .

Démonstration : Supposons que $\psi \in H_r(B)$. On a alors

$$\frac{\psi(P)}{[P(U)]^r} = \frac{\psi(Q)}{[Q(U)]^r} \quad \text{et} \quad \frac{\psi(P)}{[P(U)]^m} = \frac{\psi(Q)}{[Q(U)]^m}$$

Comme $\psi(P)$ et $\psi(Q)$ ne peuvent pas être nuls, cela entraîne

$$[P(U)]^{m-r} = [Q(U)]^{m-r} .$$

Si on avait $m \neq r$ on aurait $P(U) = Q(U)$ et par conséquent $\psi(P) = \psi(Q)$ ce qui est contraire à l'hypothèse.

II.6. - Remarque sur la définition de l'homogénéité.

Si l'on suppose que la condition suivante est vérifiée :

$$(C_1) \quad \left\{ \begin{array}{l} \forall U \in \mathcal{O}(B) \\ (P \in \mathcal{D}_U \implies P_U \in \mathcal{D}_U) \end{array} \right.$$

le a) de la définition II.2. équivaut à

$$(a') : \quad \forall P \in \mathcal{D}_U \quad \psi(P_U) = \frac{\psi(P)}{[P(U)]^m} .$$

Dans le cas $m = 1$ et B fermé on retrouve ainsi la définition de l'homogénéité donnée par D. Bosq.

La condition (C_1) , que l'on supposera satisfaite dans le paragraphe III, est forte et manque dans certains cas de réalisme (nous la remplacerons dans un prochain travail par une condition beaucoup plus large). S'il existe un élément P de \mathcal{D} et une famille $(U_k)_{k \in \mathbb{N}}$ de voisinages décroissants de F tels que $P(U_{k+1}) < P(U_k)$, $\forall k \in \mathbb{N}$, la condition (C_1) entraîne que \mathcal{D} n'est inclus dans aucun espace vectoriel réel de dimension finie.

II.7. - Définition de la localisation.

On dit que ψ est localisé par le borélien B si et seulement si pour tout $U \in \mathcal{O}(B)$ on a :

- a) $\forall (P, Q) \in \mathcal{D}_U^2 \quad (P|_U = Q|_U \implies \psi(P) = \psi(Q)) .$
 b) $\forall P \notin \mathcal{D}_U \quad \psi(P) = 0 .$

II.8. - Remarque - Vocabulaire.

Nous avons repris la définition donnée par D. Bosq en lui ajoutant la condition b). Il est clair que l'on réduit ainsi l'ensemble de paramètres localisés par B mais cela nous permettra de gagner en concision dans la suite et ne restreint pas la généralité :

Si ψ ne satisfait qu'à la condition a), il prend une valeur fixe a_0 sur l'ensemble des éléments de \mathcal{D} qui ne chargent pas un élément de $\mathcal{O}(B)$; en effet,

$$\text{si } P(U) = 0 \quad (U \in \mathcal{O}(B))$$

$$\text{et } Q(V) = 0 \quad (V \in \mathcal{O}(B))$$

$$P|_{U \cap V} = Q|_{U \cap V} = 0 \quad \text{et } (U \cap V \in \mathcal{O}(B)) \quad \text{donc } \psi(P) = \psi(Q) .$$

Il est alors immédiat que le paramètre $(\psi - a_0)$ vérifie aussi bien a) que b).

Nous adopterons le vocabulaire fixé par D. Bosq :

ψ est dit local s'il est localisé par un singleton ;

ψ est dit μ -local s'il est localisé par un borélien de μ -mesure nulle (μ mesure σ -finie sur $(\Omega, \mathcal{B}_\Omega)$) ;

ψ est dit global s'il n'est localisé que par des boréliens denses dans Ω .

II.9. - Exemples.

II.9.1. - $\Omega = \mathbb{R} = \mathbb{Q}$. \mathcal{D} : ensemble des mesures de probabilité P à fonction de répartition F_P continue et strictement croissante.

$\psi(P) = q_\alpha(P) = F_P^{-1}(\alpha)$: quantile d'ordre α ($\alpha \in]0, 1[$). ψ est un paramètre global.

II.9.2. - \mathbb{Q} : ensemble des mesures positives bornées sur Ω .
 $\mathcal{D} = \mathcal{P}$ $\psi(P) = \lambda P|_B$ (λ fixé dans \mathbb{R}^{+*}). ψ est localisé par B .

II.9.3. - $\mathbb{Q} = \mathcal{P}$ $\mathcal{D} = \{P \in \mathcal{P} \mid P(B) > 0\}$

$\psi(P) = P_B$. ψ est localisé par B .

II.9.4. - Exemple II.3. 1°) ψ est local : il est localisé par $\{a\}$. Il est également λ^P -local.

II.9.5. - Moyenne α -tronquée.

$$\mathbb{Q} = \mathbb{R}.$$

\mathcal{D}_1 : ensemble des probabilités P sur \mathbb{R} à fonction de répartition F_P strictement croissante, dérivable et telle que

$$F_P(a) \geq 1 - c \quad \text{où } a \in \mathbb{R} \text{ et } c \in [0, \frac{1}{2}[$$

\mathcal{D}_2 : ensemble des éléments de \mathcal{D}_1 tels que $\psi_2(P) = \int_{-\infty}^{+\infty} x \, dF_P(x)$ existe et soit fini. ψ_2 est un paramètre global.

Sur \mathcal{D}_1 la moyenne α -tronquée $\psi_1(P) = (1-2\alpha)^{-1} \int_{\alpha}^{1-\alpha} F_P^{-1}(x) \, dx$ ("moyenne" prise entre les quantiles d'ordre α et $(1-\alpha)$) où $\alpha \in [c, \frac{1}{2}[$ définit un paramètre ψ_1 qui est localisé par $]-\infty, a]$:

$$\int_{\alpha}^{1-\alpha} F_P^{-1}(x) \, dx = \int_{F_P^{-1}(\alpha)}^{F_P^{-1}(1-\alpha)} x F_P'(x) \, dx$$

$$1 - \alpha \leq 1 - c \leq F_P(a) \quad \text{donc } F_P^{-1}(1-\alpha) \in]-\infty, a]$$

si, pour $U \supset]-\infty, a]$, on a $P|_U = Q|_U$ les fonctions F_P et F_Q coïncident sur $]-\infty, a]$ donc $\psi_1(P) = \psi_1(Q)$.

II.10. - Remarques.

II.10.1. - Une différence importante entre la définition de l'homogénéité (en particulier de degré 0) et celle de la localisation est due à celle qui existe entre la notion de probabilité induite et la notion de restriction d'une probabilité :

$$\forall (P, Q) \in \mathcal{D}_U^2 \quad (P|_U = Q|_U \implies P_U = Q_U)$$

la réciproque n'étant vraie que si $P(U) = Q(U)$.

Proposition 3. - Soit $U \in \mathcal{O}(F)$ et $P \in \mathcal{P}$ tels que $P(U) > 0$.
L'ensemble des probabilités Q telles que $Q_U = P_U$ est l'ensemble des probabilités $Q_\lambda = \lambda P_U + (1-\lambda) Q'$, $\lambda \in]0, 1]$, $Q' \in \mathcal{P} - \mathcal{D}_U$.

Démonstration :

$$* \text{ Soit } \lambda \in]0, 1] \text{ et } Q' \in \mathcal{P} - \mathcal{D}_U \quad (Q_\lambda)_U = \frac{Q_\lambda|_U}{Q_\lambda(U)} = P_U$$

$$* \text{ Réciproquement soit } Q \text{ telle que } Q(U) > 0 \text{ et } Q_U = P_U$$

$$\text{alors } Q = Q|_U + Q|_{U^c}$$

$$\text{si } Q(U^c) > 0 \quad Q = Q(U) \frac{Q|_U}{Q(U)} + Q(U^c) \frac{Q|_{U^c}}{Q(U^c)}$$

$$Q = Q(U) P_U + [1 - Q(U)] Q_{U^c}$$

$$\text{si } Q(U^c) = 0 \quad Q = P_U .$$

On peut encore remarquer que lorsque $0 < P(U) < 1$ les probabilités P_U et P_{U^c} ont une position extrême sur la droite Δ de \mathcal{P}' qui les contient : $\Delta - \{P_{U^c}, P_U\}$ est l'ensemble des mesures bornées à signe μ , de masse totale 1, telles que :

$$\mu(U^c) \neq 0$$

$$\mu(U) \neq 0$$

$$\frac{\mu|_U}{\mu(U)} = P_U$$

$$\frac{\mu|_{U^c}}{\mu(U^c)} = P_{U^c}$$

et le segment $[P_{U^c}, P_U]$ est l'ensemble des probabilités Q telles que $Q_U = P_U$ et $Q_{U^c} = P_{U^c}$.

II.10.2. - Soit $\psi \in \mathcal{H}_m(B)$. On suppose qu'il existe $(P, Q) \in \mathcal{D}^2$, $U \in \mathcal{O}(B)$, $\lambda \in [0, 1]$ tels que $P|_{U^c} = 0$ $Q|_U = 0$ et $[\lambda P + (1-\lambda) Q] \in \mathcal{D}$.

$$\text{Alors } \psi(\lambda P + (1-\lambda) Q) = \lambda^m \psi(P) .$$

II.11. - Classe fonctionnelle déterminant un paramètre.

II.11.1. - Il est évident que la partie a) de la définition de la localisation (II.7) peut s'écrire de la façon suivante :

$$\forall (P, Q) \in \mathcal{D}_U^2 \quad (\forall A \in \mathcal{B}_\Omega, \int 1_{A \cap U} dP = \int 1_{A \cap U} dQ) \\ \implies (\psi(P) = \psi(Q)) .$$

ce qui montre que la notion introduite est un cas particulier d'une notion plus générale, extension de la notion de classe déterminante, et qui s'exprime en termes fonctionnels.

Soit I l'ensemble des fonctions réelles intégrables par rapport à tout élément de \mathcal{D} et G un sous-ensemble non vide de I .

On dit que G détermine le paramètre ψ si et seulement si pour tout couple $(P, Q) \in \mathcal{D}^2$ on a :

$$(\forall g \in G \int g dP = \int g dQ) \implies (\psi(P) = \psi(Q)) .$$

II.11.2. - Exemples.

a) Un paramètre localisé par le borélien B est déterminé par $(1_{A \cap U})_{A \in \mathcal{B}_\Omega, U \in \mathcal{O}(F)}$.

b) Sur un espace métrique la classe des fonctions réelles uniformément continues bornées détermine la loi. Plus généralement les indicatrices d'une classe déterminante déterminent la loi.

c) Soit ψ un paramètre à valeurs dans un e.v.t.l.c. Θ tel qu'il existe Ψ ("pseudo-transposée" de ψ) : $\Theta' \times \Omega \rightarrow \mathbb{R}$ vérifiant

$$\forall P \in \mathcal{D} \quad \forall e' \in \Theta' \quad (\psi(P), e') = \int_\Omega \Psi(e', x) dP(x) .$$

Pour toute famille $(e'_i)_{i \in I}$ faiblement totale dans Θ' la famille $(\Psi(e'_i, \cdot))_{i \in I}$ détermine le paramètre ψ . Un cas particulier est la classe F étudiée par M. Carbon [1982] avec $\Theta = L^2(\mu)$.

Pour fixer les idées on peut penser à la méthode des fonctions orthogonales pour la densité $\psi(P) = f_P = \frac{dP}{d\mu}$. On a alors $\forall e' \in L^2(\mu) \quad \forall x \in \Omega \quad \Psi(e', x) = e'(x)$ et toute base orthonormale de $L^2(\mu)$ détermine ψ .

On peut noter que la notion de classe déterminant la convergence peut s'étendre de la même façon à d'autres paramètres que la loi. Nous reviendrons sur ces questions dans un prochain travail.

II.12. - Localisation forte. Liens entre homogénéité et localisation.

II.12.1. - Définition. - Un paramètre homogène de degré zéro sur B est dit fortement localisé par B .

D'après la remarque II.10.1., un tel paramètre est évidemment localisé par B (la réciproque étant fautive comme le prouve l'exemple suivant : $\Omega = \Theta = \mathbb{R}$, $\mathcal{D} = \mathcal{P}$, $\psi(P) = P(\{a\})$, a fixé dans \mathbb{R} . ψ est local (et λ -local) car il est localisé par le fermé $\{a\}$ mais il n'est pas fortement localisé par $\{a\}$; $\psi \in H_1(\{a\})$. Un autre exemple est fourni par II.3.1.).

On a en fait une propriété générale :

Proposition 4. - Soit $m \in \mathbb{R}^+$. Tout paramètre homogène de degré m sur B est localisé par B .

Démonstration : Soit $\psi \in H_m(B)$, $U \in \mathcal{O}(B)$ et $(P, Q) \in \mathcal{D}^2$
tels que $P|_U = Q|_U$

si $P(U) > 0$ alors $Q(U) > 0$ et $P_U = Q_U$

donc
$$\frac{\psi(P)}{[P(U)]^m} = \frac{\psi(Q)}{[Q(U)]^m} \quad \text{et} \quad \psi(P) = \psi(Q)$$

si $P(U) = 0$ alors $Q(U) = 0$ et $\psi(P) = \psi(Q) = 0$.

II.12.2. - Exemples de paramètres fortement localisés.

a) Proposition 5.- Soit B un sous-ensemble borélien propre de Ω et $\mathcal{D} = \{P \mid P \in \mathcal{P} \text{ et } 0 < P(B) < 1\}$.

$$\psi : \mathcal{D} \longrightarrow \mathcal{P}$$

$$P \longrightarrow \psi(P) = P_B$$

est fortement localisé par B .

Démonstration : Soit $U \in \mathcal{O}(B)$ $\forall R \in \mathcal{D} \quad R(U) \geq R(B) > 0$.

Soit $(P, Q) \in \mathcal{D}^2$ tel que $P_U = Q_U$ et soit $\Omega_1 \subset B$

on a :

$$\frac{P(B)}{P(U)} = \frac{Q(B)}{Q(U)}$$

$$P_B(\Omega_1) = \frac{P(\Omega_1)}{P(B)} = \frac{P(\Omega_1)}{P(U)} \times \frac{P(U)}{P(B)}$$

$$Q_B(\Omega_1) = \frac{Q(\Omega_1)}{Q(B)} = \frac{Q(\Omega_1)}{Q(U)} \times \frac{Q(U)}{Q(B)}$$

donc $\psi(P) = \psi(Q)$.

b) Si \mathcal{D} est l'ensemble S_B des probabilités à support inclus dans B , tout paramètre est homogène sur B , de degré quelconque (en particulier fortement localisé). On s'aperçoit ainsi que les notions

d'homogénéité et de localisation sont fortement liées, pour un paramètre fixé, à la famille \mathcal{D} et au borélien B considérés.

c) Fonction frontière.

$\Omega = \mathbb{R}^2$ $\Theta = \mathbb{R}^{+*}$ \mathcal{D} : ensemble des mesures de probabilité P sur $(\mathbb{R}^2, \mathcal{B}_{\mathbb{R}^2})$ qui ont un support S_P compact de la forme $S_P = \{(x,y) \mid 0 \leq x \leq 1 \text{ et } 0 \leq y \leq s_P(x)\}$ où s_P est une application continue strictement positive de $[0,1]$ dans \mathbb{R} . S_P est entièrement déterminé par s_P (fonction frontière). $\psi(P) = s_P(a)$, où a est un réel fixé dans $[0,1]$.

ψ est fortement localisé par la demi-droite fermée $F \begin{cases} x = a \\ y \geq 0 \end{cases}$.

Soit $U \in \mathcal{O}(F)$. \mathcal{D}_U est égal à \mathcal{D} : si on avait $P(U) = 0$ on aurait $U \subset S_P^c$ ce qui n'est pas puisque $(a,0) \in U \cap S_P$.

Soit $(P,Q) \in \mathcal{D}^2$ tel que $P_U = Q_U$.

$P|_U$ et $Q|_U$ ont le même support et

$$S_P \cap U = S_{P|_U} = S_{Q|_U} = S_Q \cap U$$

ce qui entraîne $\psi(P) = \psi(Q)$.

De la même façon on montre que ψ_1 , défini sur le sous-ensemble \mathcal{D}_1 des éléments de \mathcal{D} tels que $s_P''(a)$ existe, par $\psi_1(P) = s_P''(a) [1 + (s_P'(a))^2]^{-3/2}$ (courbure algébrique de la frontière au point d'abscisse a), est fortement localisé par F .

d) Régression.

Soit $a \in \mathbb{R}$, $\Omega = \mathbb{R}^2$, \mathcal{D} : ensemble des lois de probabilité P absolument continues par rapport à $\lambda^{\otimes 2}$ et à densité admettant une version f continue telle que :

$\int_{\mathbb{R}} y f(a,y) dy$ existe et que $\int_{\mathbb{R}} f(a,y) dy > 0$.

$$\psi(P) = \left(\int_{\mathbb{R}} f(a,y) dy \right)^{-1} \int_{\mathbb{R}} y f(a,y) dy$$

ψ (fonction de régression de y en x au point a) est fortement localisé par la droite d'équation $x = a$.

II.13. - Cas d'un paramètre linéaire.

Lorsque \mathcal{D} est un espace vectoriel sur \mathbb{R} et que \mathcal{D} est convexe, on dit qu'un paramètre ψ est linéaire si et seulement si

$$\forall (P,Q) \in \mathcal{D}^2 \quad \forall \lambda \in [0,1] \quad \psi(\lambda P + (1-\lambda) Q) = \lambda \psi(P) + (1-\lambda) \psi(Q).$$

Il paraît évident, qu'en général, un paramètre linéaire ne peut être homogène sur un borélien propre de Ω que si son degré d'homogénéité est 1 :

II.13.1. - Proposition 6. - Soit $\psi \in H_m(B)$ ($m \neq 1$) un paramètre linéaire sur \mathcal{D} convexe. On suppose en outre que :

$$(C_2) \quad \left\{ \begin{array}{l} \exists V \in \mathcal{O}(B) \quad , \quad \exists Q \in \mathcal{D} \quad \text{tels que} \quad Q(V) = 0. \end{array} \right.$$

Alors ψ est identiquement nul.

Démonstration : Soit $P \in \mathcal{D}$ et $U \in \mathcal{O}_V(B)$. Si $P(U) = 0$ $\psi(P) = 0$. Supposons $P(U) > 0$. Soit $Q \in \mathcal{D}$ tel que $Q(U) = 0$ (donc $\psi(Q) = 0$). Pour tout $\lambda \in]0,1]$, $[\lambda P + (1-\lambda) Q]$ est un élément de \mathcal{D} dont la probabilité conditionnelle par rapport à U est égale à P_U . On a donc

$$\forall \lambda \in]0,1] \quad \frac{\psi(\lambda P + (1-\lambda) Q)}{\lambda^m [P(U)]^m} = \frac{\psi(P)}{[P(U)]^m}$$

$$\forall \lambda \in]0, 1] \quad \lambda \psi(P) + (1-\lambda) \psi(Q) = \lambda^m \psi(P)$$

$$\forall \lambda \in]0, 1] \quad (\lambda^m - \lambda) \psi(P) = 0$$

ce qui entraîne $\psi(P) = 0$.

Proposition 7.- On suppose que \mathcal{D} est convexe et que la condition (C_2) est vérifiée. Un paramètre linéaire ψ appartient à $H_1(B)$ si et seulement si il est localisé par B .

Démonstration : Il suffit de montrer, sous ces hypothèses, et pour $m = 1$ la réciproque de la proposition 4 (II.12.).

Soit $V \in \mathcal{O}(B)$ et $Q \in \mathcal{D}$ tels que $Q(V) = 0$.

Soit $P \in \mathcal{D}$ et $U \in \mathcal{O}_V(B)$.

Si $P(U) = 0$ alors $\psi(P) = 0$.

Si $P(U) > 0$ et si Q' est telle que $P_U = Q'_U$ les probabilités $(Q'(U)P + (1-Q'(U))Q)$ et $(P(U)Q' + (1-P(U))Q)$ ont même restriction à U donc les valeurs prises par le paramètre pour ces deux éléments de \mathcal{D} sont égales et

$$Q'(U) \psi(P) = P(U) \psi(Q') .$$

II.13.2. - Paramètre intégral.

On dira qu'un paramètre $\psi : \mathcal{D} \rightarrow \mathbb{H}$ (Banach) est intégral si ψ admet un estimateur fortement sans biais d'ordre 1 ; autrement dit il existe un noyau K (non unique en général) : $\Omega \rightarrow \mathbb{H}$ \mathcal{D} -intégrable et tel que $\forall P \in \mathcal{D} \quad \int_{\Omega} K dP = \psi(P)$.

Lorsque \mathcal{D} est convexe un paramètre intégral est évidemment linéaire.

Proposition 8. - On suppose que la condition (C_1) (II.6.) est vérifiée pour $B = F$, fermé de Ω . Un paramètre intégral ψ de noyau K appartient à $H_1(F)$ si et seulement si

$$\int_{F^c} K dP = 0, \quad \forall P \in \mathcal{D}$$

(Autrement dit $K \mathbb{1}_F$ est un noyau pour ψ).

Démonstration : C.N. - Soit $(U_k)_{k \in \mathbb{N}}$ une suite de $\mathcal{O}(F)$ tendant en décroissant vers F et soit $P \in \mathcal{D}$.

S'il existe $k \in \mathbb{N}$ tel que $P(U_k) = 0$ alors $P(F) = 0$ et

$$\psi(P) = 0 = \int_F K dP = \int_{F^c} K dP$$

si $\forall k \in \mathbb{N} \quad P(U_k) > 0$:

$$\forall k \in \mathbb{N} \quad \psi(P_{U_k}) = \int_{\Omega} K dP_{U_k} = \frac{1}{P(U_k)} \int_{U_k} K dP = \frac{\psi(P)}{P(U_k)}$$

donc $\forall k \in \mathbb{N} \quad \psi(P) = \int_{U_k} K dP$.

Il résulte du théorème de convergence dominée que

$$\psi(P) = \lim_{k \rightarrow +\infty} \int_{U_k} K dP = \int_F K dP \quad \text{donc} \quad \int_{F^c} K dP = 0$$

C.S. -

$$\forall P \in \mathcal{D} \quad \psi(P) = \int_F K dP. \quad \text{Soit } U \in \mathcal{O}(F)$$

si $P(U) = 0$ $P(F) = 0$ et $\psi(P) = 0$

si $P(U) > 0$ $\psi(P_U) = \int_F K dP_U = \frac{1}{P(U)} \int_F K dP = \frac{\psi(P)}{P(U)}$.

III - HOMOGENEITE, LOCALISATION ET ESTIMABILITE SANS BIAIS.

Dans tout ce paragraphe nous supposons que la condition (C_1) est vérifiée pour un borélien fermé F de Ω .

III.1. - Relations vérifiées par un paramètre homogène et estimable sans biais.

Soit $\psi \in H_m(F)$ un paramètre estimable d'ordre l et de noyau T_l . $\forall P \in \mathcal{D}$ $\psi(P) = \int_{\Omega} T_l dP^{\otimes l}$.

Soit $(U_k)_{k \in \mathbb{N}}$ une suite de $\mathcal{O}(F)$ tendant en décroissant vers F et soit $P \in \mathcal{D}$.

a) S'il existe k_0 tel que $P(U_{k_0}) = 0$ alors $\psi(P) = 0$

b) Sinon $\forall k \in \mathbb{N}$ $\psi(P_{U_k}) = \frac{\psi(P)}{[P(U_k)]^m} = \int_{U_k} T_l dP_{U_k}^{\otimes l}$

d'où

$$(1) \quad \forall k \in \mathbb{N} \quad \psi(P) = [P(U_k)]^{m-l} \int_{U_k} T_l dP^{\otimes l}.$$

On en déduit, par application du théorème de convergence dominée que si $m \neq 0$ et $l \neq 0$

$$[P(F)]^l \psi(P) = [P(F)]^m \int_F T_l dP^{\otimes l}$$

(relation qui ne présente un intérêt que si $P(F) > 0$) en particulier si $F = \{a\}$ on a, si $m \geq l$ ou si $m < l$ et $P(\{a\}) > 0$,

$$\psi(P) = [P(\{a\})]^m \times C, \quad C \text{ constante de } \mathbb{R}.$$

La démonstration de la proposition 8 s'étend immédiatement aux paramètres estimables d'ordre l quelconque, grâce à l'égalité (1).

Proposition 9.- Un paramètre estimable d'ordre ℓ , de noyau T_ℓ , appartient à $H_\ell(F)$ si et seulement si

$$\forall P \in \mathcal{D} \quad \int_{(F^\ell)^c} T_\ell dP^{\otimes \ell} = 0$$

(on a alors $\forall P \in \mathcal{D} \quad \psi(P) = \int_{F^\ell} T_\ell dP^{\otimes \ell}$).

On en déduit les corollaires suivants :

Corollaire 1.- Un paramètre estimable sans biais, homogène de degré m sur un fermé F de P -mesure nulle pour tout élément P de \mathcal{D} , non constant, est de degré d'estimabilité strictement supérieur à m . (Le degré d'estimabilité d'un paramètre ψ est le plus petit entier n pour lequel ψ admet un estimateur sans biais d'ordre n).

Grâce au théorème de Bickel - Lehmann on a :

Corollaire 2.- Si $\psi \in H_m(F)$, si $P(F) = 0$ pour tout élément P de \mathcal{D} , si ψ n'est pas constant, si \mathcal{D} est convexe et si $\psi(\alpha P + (1-\alpha)Q)$ est un polynôme en α de degré inférieur ou égal à m , ψ n'est pas estimable sans biais.

Le corollaire 2 pourra s'appliquer par exemple à la densité et à ses dérivées.

III.2. - Paramètre localisé estimable sans biais.

Nous avons vu que la proposition 4 (II.12) admettait en ce qui concerne les paramètres linéaires, et sous certaines conditions une réciproque (proposition 7 (II.13)). Il en est de même pour les paramètres estimables sans biais :

Proposition 10.- Soit ψ un paramètre estimable sans biais d'ordre m , de noyau T_m , localisé par F .

On suppose que \mathcal{D} est dominé par μ , mesure positive σ -finie sur Ω et que, pour tout $U \in \mathcal{O}(F)$ et toute probabilité $P \in \mathcal{D}$ tels que $0 < P(U) < 1$ la famille $\mathcal{D}_m(U, P) = \{Q^{\otimes m} \mid Q \in \mathcal{D} \text{ et } Q|_U = P|_U\}$ est $\mu^{\otimes m}$ -complète (autrement dit

$$\left(\int_{\Omega^m} T_m dQ^{\otimes m} = 0, \forall Q^{\otimes m} \in \mathcal{D}_m(U, P) \right) \implies (T = 0 \text{ } \mu^{\otimes m} \text{ p.p.}) .$$

Alors ψ est homogène de degré m .

Démonstration : On utilise la condition N et S de la proposition 9.

Soit $(U_k)_{k \in \mathbb{N}}$ une suite de $\mathcal{O}(F)$, $\uparrow F$ et soit $P \in \mathcal{D}$.

Si $P(F) = 1$ on a $P^{\otimes m}((F^m)^c) = 0$ donc

$$\int_{(F^m)^c} T_m dP^{\otimes m} = 0$$

Si $P(F) \neq 1$ comme $\lim_{k \rightarrow +\infty} P(U_k) = P(F)$ on a, pour k

assez grand $P(U_k) < 1$.

Si $P(U_k) = 0$ $\psi(P) = 0$ et $P(F) = 0$

donc

$$\int_{(F^m)^c} T_m dP^{\otimes m} = 0$$

Supposons que $\forall k \geq k_1 \quad 0 < P(U_k) < 1$.

Soit $k \geq k_1$ et $Q \in \mathcal{D}_m(U_k, P)$

On a

$$\int_{U_k^m} T_m dP^{\otimes m} + \int_{(U_k^m)^c} T_m dP^{\otimes m} = \psi(P)$$

et une égalité analogue pour $\psi(Q)$.

$$\left. \begin{array}{l} F \text{ localise } \psi \\ P|_{U_k} = Q|_{U_k} \end{array} \right\} \implies \psi(P) = \psi(Q)$$

$$\text{Par suite } \alpha(P) = \int_{(U_k^m)^c} T_m dP^{\otimes m} = \int_{(U_k^m)^c} T_m dQ^{\otimes m}, \quad \forall Q \in \mathcal{D}_m(U_k, P)$$

$$\text{donc } \forall Q \in \mathcal{D}_m(U_k, P) \quad \int_{\Omega^m} (1_{(U_k^m)^c} T_m - \alpha(P)) dQ^{\otimes m} = 0 \quad \text{donc } 1_{(U_k^m)^c} T_m = \alpha(P)$$

sur Ω_0 tel que $\mu^{\otimes m}(\Omega_0^c) = 0$.

Comme $P(U_k) > 0$ et que $P \ll \mu$, $\mu(U_k) > 0$ et $\mu^{\otimes m}(U_k^m) > 0$
donc $\mu^{\otimes m}(U_k^m \cap \Omega_0) > 0$.

$U_k^m \cap \Omega_0$ est donc non vide et

$$\forall \omega \in U_k^m \cap \Omega_0 \quad \alpha(P) = 1_{(U_k^m)^c} T_m(\omega) = 0.$$

Par conséquent $\int_{(U_k^m)^c} T_m dP^{\otimes m} = 0$ et le théorème de convergence dominée entraîne $\int_{(F^m)^c} T_m dP^{\otimes m} = 0$.

Un cas particulier important est celui où ψ est μ -local ($\mu(F) = 0$). On renforce alors la conclusion du théorème 4 de D. Bosq en appliquant la proposition 10 et le corollaire 1 :

Corollaire 3.- Soit ψ un paramètre non constant, localisé par F de μ -mesure nulle (μ mesure positive σ -finie sur Ω dominant \mathcal{D}). On suppose qu'il existe $m \in \mathbb{N}^*$ tel que $\forall U \in \mathcal{O}(F)$, $\forall P \in \mathcal{D}$ vérifiant $0 < P(U) < 1$, $\mathcal{D}_m(U, P)$ est $\mu^{\otimes m}$ -complète.

Alors, si ψ est estimable sans biais, il est de degré strictement supérieur à m .

Démonstration : Si ψ était estimable sans biais d'ordre m il serait d'après la proposition 10 homogène de degré m et le corollaire 1 entraîne une contradiction.

IV - FONCTION INFLUENCE D'UN PARAMETRE HOMOGENE.-

La fonction influence, introduite par F.R. Hampel [1974], donne dans beaucoup de cas une bonne idée de "l'importance" que va avoir une observation x sur l'estimation d'un paramètre. Sa détermination est un premier pas vers l'étude de l'estimation robuste des paramètres homogènes. On suppose que \mathcal{D} contient chacun des segments $[P, \delta_x]$ de P , $P \in \mathcal{D}$, $x \in \Omega$. La fonction influence de ψ au point P de \mathcal{D} est alors définie par

$$I C_{\psi, P}(x) = \lim_{s \rightarrow 0} [\psi((1-s)P + s\delta_x) - \psi(P)] s^{-1}$$

c'est la dérivée de ψ au point P dans la direction de δ_x (dérivée au sens de Gâteaux).

Soit ψ un paramètre homogène de degré m sur F fermé de Ω , $x \in \Omega$ et $P \in \mathcal{D}$.

IV.1. - Cas où il existe $U \in \mathcal{O}(F)$ tel que $P \notin \mathcal{D}_U$.

On a alors $\psi(P) = 0$.

si $x \notin F$ il existe $V \in \mathcal{O}(F)$ tel que $x \notin V$

$$P \notin \mathcal{D}_U \cap V, \delta_x \notin \mathcal{D}_U \cap V \text{ donc } I C_{\psi, P}(x) = 0$$

si $x \in F$ $((1-s)P + s\delta_x)_U = (\delta_x)_U = \delta_x$

$$\text{donc } \psi((1-s)P + s\delta_x) = s^m \psi(\delta_x), \quad \forall s \in]0, 1[.$$

$$\text{si } m \in [0, 1[\quad I C_{\psi, P}(x) = \lim_{s \downarrow 0} \frac{\psi(\delta_x)}{s^{1-m}} \quad \left\{ \begin{array}{l} 0 \quad \text{si } \psi(\delta_x) = 0 \\ \text{point à l'infini de la} \\ \text{demi-droite } [0, \psi(\delta_x)) \\ \text{sinon .} \end{array} \right.$$

$$\text{si } m = 1 \quad I C_{\psi, P}(x) = \psi(\delta_x)$$

$$\text{si } m > 1 \quad I C_{\psi, P}(x) = 0 .$$

IV.2. - Cas où $P \in \mathcal{D}_U$ pour tout $U \in \mathcal{O}(F)$.

Ce cas est évidemment le plus intéressant.

Si l'on suppose $x \in F$ on ne peut pas faire de calculs sans hypothèses supplémentaires. On peut cependant remarquer que, dans beaucoup d'exemples courants, le borélien F est de P -probabilité nulle, $\forall P \in \mathcal{D}$.

IV.2.1. - si $x \notin F$ $\exists V \in \mathcal{O}(F)$ tel que $x \notin V$

$$((1-s)P + s\delta_x)_V = P_V$$

$$\text{et } \psi((1-s)P + s\delta_x) = [(1-s)P(V)]^m \frac{\psi(P)}{[P(V)]^m} = (1-s)^m \psi(P)$$

$$\text{d'où } I C_{\psi, P}(x) = -m \psi(P) .$$

Il est clair que l'on obtient également ce résultat pour la dérivée en P dans la direction de toute probabilité Q ne chargeant pas un élément V de $\mathcal{O}(F)$ lorsque le segment $[P, Q]$ est inclus dans \mathcal{D} .

IV.2.2. - si $x \in F$ (exemples)

$$\text{a) } \underline{\psi(P) = [P(F)]^m}, \quad m > 0, \quad \mathcal{D} = \{P \mid P \in \mathcal{P} \text{ et } P(F) > 0\}$$

$$\Theta =]0, 1[, \quad \psi \in H_m(F)$$

$$\begin{aligned}
 I C_{\psi, P}(x) &= \lim_{s \rightarrow 0} \left([(1-s) P(F) + s]^m - [P(F)]^m \right) s^{-1} \\
 &= \frac{d}{ds} [(1-s) P(F) + s]^m \Big|_{s=0} \\
 &= m [(1-s) P(F) + s]^{m-1} (1-P(F)) \Big|_{s=0}
 \end{aligned}$$

$$I C_{\psi, P}(x) = m \psi(P) \frac{P(F^c)}{P(F)} .$$

La quantité $m \frac{P(F^c)}{P(F)}$, influence relative d'une observation tombant dans F par rapport à la valeur du paramètre tend vers $+\infty$ si $P(F)$ tend vers 0 et vers 0 si $P(F)$ tend vers 1.

$$b) \quad \underline{\psi(P) = P|_F} , \quad \mathcal{D} = \mathcal{P} , \quad \mathcal{Q} = \mathcal{P}'$$

$$\psi \text{ est linéaire et } I C_{\psi, P}(x) = \delta_x - P|_F$$

$$c) \quad \underline{\psi(P) = P_F}$$

$$\mathcal{D} = \{P \mid P \in \mathcal{P} \text{ et } P(F) > 0\} \quad \mathcal{Q} = \mathcal{P}'$$

$$\begin{aligned}
 I C_{\psi, P}(x) &= \lim_{s \rightarrow 0} \left(\frac{(1-s) P|_F + s \delta_x - P|_F}{(1-s) P(F) + s} \right) s^{-1} \\
 &= \lim_{s \rightarrow 0} \frac{P(F) \delta_x - P|_F}{((1-s) P(F) + s) P(F)} = \frac{\delta_x - P_F}{P(F)}
 \end{aligned}$$

$$I C_{\psi, P}(x) = \frac{\delta_x - \psi(P)}{P(F)} .$$

R E F E R E N C E S.

- BOSQ, D. - Contribution à la théorie de l'estimation fonctionnelle.
Publications de l'I.S.U.P., Vol. XIX, fasc. 2, 1970.
- CARBON, M. - Sur l'estimation asymptotique d'une classe de paramètres
fonctionnels pour un processus stationnaire.
Thèse de 3ème cycle, Université de Lille I, 1982.
- HAMPEL, F.R. - The influence curve and its role in robust estimation.
JASA, vol. 69, 1974, 383-393.
- de MONTRICHER, G.F. - Non parametric maximum likelihood estimation of
TAPIA R.A. and probability densities by penalty function methods.
THOMPSON J.R. Ann. of Stat., Vol. 3, 1975, 1329-1348.

CHAPITRE 2

EXHAUSTIVITE LOCALE D'UNE STATISTIQUE.

APPLICATION AU CONDITIONNEMENT D'ESTIMATEURS.

RÉSUMÉ : Après avoir défini la notion de propriété locale d'une statistique, nous étudions les effets du conditionnement d'un estimateur par une statistique localement exhaustive et, plus particulièrement, le comportement du risque quadratique. Ces considérations sont appliquées à l'estimation locale de la densité et de la régression par la méthode du noyau. Dans la dernière partie nous donnons des conditions suffisantes d'optimalité pour certaines classes de noyaux.

ABSTRACT : We define the local properties of a statistic and study the effect of taking the conditional expectation of an estimator with respect to a locally sufficient statistic. More precisely, the behaviour of the quadratic risk is pointed out and we look at applications to density and regression estimation by kernel method. In the last part we give sufficient conditions for optimality of some classes of kernels.

CLASSIFICATION AMS (MOS) 1980 - 62B05 - 62G05 - 49B34 -

MOTS CLÉS : Localisation, exhaustivité, non-paramétrique, densité, régression, noyaux optimaux.

KEY WORDS : Local properties, sufficiency, non-parametric, density, regression, optimal kernels.

P L A N.

- I - INTRODUCTION.
- II - PROPRIETES LOCALES D'UNE STATISTIQUE.
 - II.1. - Notations - Définitions.
 - II.2. - Remarques.
 - II.3. - Exemples.
- III - CONDITIONNEMENT PAR RAPPORT A LA STATISTIQUE S.
 - III.1. - Généralités.
 - III.2. - Propriétés de l'estimateur conditionné.
 - III.3. - Forme de l'estimateur conditionné.
 - III.4. - Remarques.
- IV - RISQUE QUADRATIQUE DE L'ESTIMATEUR CONDITIONNE.
- V - APPLICATION A L'ESTIMATION DE LA DENSITE PAR LA METHODE DU NOYAU.
 - V.1. - Limite du biais et de la variance.
 - V.2. - Etude du comportement asymptotique du biais et du risque quadratique sous des hypothèses complémentaires de régularité.
- VI - APPLICATION A L'ESTIMATION DE PARAMETRES DE REGRESSION PAR LA METHODE DU NOYAU.
- VII - QUELQUES REMARQUES SUR L'AMELIORATION D'ESTIMATEURS.
- VIII - PROBLEMES D'OPTIMALITE POUR LES NOYAUX.
 - VIII.1. - Hypothèses.
 - VIII.2. - Proposition 11.
 - VIII.3. - Remarques.
 - VIII.4. - Application au lemme 5.18 de Devroye [1984].

R E F E R E N C E S.

I - INTRODUCTION.-

Nous avons étudié récemment [Berlinet, 1983] des propriétés des paramètres statistiques que nous avons qualifiées de locales dans l'espace des observations. Une idée naturelle est de considérer des statistiques ayant de bonnes propriétés locales correspondantes afin de définir de nouveaux estimateurs ou, au moins, d'améliorer des estimateurs connus en utilisant le conditionnement par rapport à ces statistiques. C'est le but du présent travail dans lequel nous particulariserons les résultats des parties II, III et IV à l'estimation de la densité et de la régression par la méthode du noyau. Dans la dernière partie nous reviendrons sur la démonstration de l'optimalité de certains noyaux, dont celui de Bartlett - Epanechnikov.

II - PROPRIETES LOCALES D'UNE STATISTIQUE.-

II.1. - Notations - Définitions.

Soit (E, \mathcal{B}, P) un modèle statistique où E est un espace topologique et \mathcal{B} sa tribu borélienne. Soit (F, \mathcal{C}) un espace mesurable et S une statistique définie sur (E, \mathcal{B}) et à valeurs dans (F, \mathcal{C}) .

Soit B un borélien quelconque de E . On notera B^c son complémentaire dans E et pour $P \in \mathcal{P}$, $P|_B$ désignera la restriction de P à B .

On dit que S est B-exhaustive (respectivement B-libre en moyenne, B-libre, B-complète, B-quasi-complète) si et seulement si $\forall P \in \mathcal{P}$, S est exhaustive (resp. libre en moyenne, libre, complète, quasi-complète)

dans le modèle $(E, \mathcal{B}, \mathcal{P}(P, B))$ où

$$\mathcal{P}(P, B) = \{Q \mid Q \in \mathcal{P} \text{ et } Q|_{B^c} = P|_{B^c}\} .$$

C'est surtout la notion de B -exhaustivité qui sera utilisée dans la suite. Intuitivement on peut dire qu'une fois que l'on s'est fixé $Q|_{B^c}$, une statistique B -exhaustive résume l'information que l'on peut avoir sur $Q|_B$, c'est-à-dire ce qui manque pour connaître entièrement la loi Q .

Soit F un fermé de E et $\mathcal{O}(F)$ l'ensemble des ouverts contenant F . Dans l'étude de l'estimation des paramètres homogènes ou localisés par F nous nous intéresserons plus particulièrement aux familles de statistiques $(S_{U_n})_{n \in \mathbb{N}}$ où $(U_n)_{n \in \mathbb{N}}$ est une suite de $\mathcal{O}(F)$ décroissant vers F et où $\forall n \in \mathbb{N} S_{U_n}$ est U_n -exhaustive (ce qui correspond au classique problème du choix de la fenêtre en estimation fonctionnelle). Plus précisément, dans le cas de l'observation d'un échantillon X_1, \dots, X_n nous étudierons des statistiques S_n qui seront U_n^n -exhaustives dans le modèle $(E, \mathcal{B}, \mathcal{P})^{\otimes n}$.

II.2. - Remarques.

a) Les définitions précédentes sont identiques à celle que l'on obtiendrait en considérant tout élément de \mathcal{P} comme un couple $(P|_B, P|_{B^c})$ et en traitant $P|_{B^c}$ comme un paramètre fantôme. Cette considération intuitive ne serait toutefois pas exacte en général car $P|_B$ et $P|_{B^c}$ sont fortement liés ; dans beaucoup de problèmes d'estimation on a d'ailleurs correspondance biunivoque entre $\{P|_B \mid P \in \mathcal{P}\}$ et $\{P|_{B^c} \mid P \in \mathcal{P}\}$, par exemple lorsque B est un pavé ouvert de \mathbb{R}^n ; le qualificatif "fantôme" ne s'applique pas alors de façon stricte au paramètre $P|_{B^c}$.

b) Les notions habituelles d'exhaustivité, liberté et complétion se retrouvent pour $B = E$ puisque $\forall P \in \mathcal{P} \quad P(P, E) = P$.

D'autre part toute statistique est trivialement N -exhaustive et N -libre, pour tout ensemble N P -négligeable (si $P(B) = 0 \quad P(P, B) = \{P\}$).

c) Si $C \subset B \quad P(P, C) \subset P(P, B)$ donc toute statistique B -exhaustive (resp. B -libre) est C -exhaustive (resp. C -libre). On utilise en particulier cette remarque dans le modèle $(E, \mathcal{B}, P)^{\otimes n} \quad (n \in \mathbb{N}^*) : B^n \subset \left((B^c)^n \right)^c$ donc $P(P^{\otimes n}, B^n) \subset P(P^{\otimes n}, \left((B^c)^n \right)^c)$. On a aussi l'inclusion $P(P^{\otimes n}, B^n) \subset [P(P, B)]^{\otimes n}$ (si $A \subset P$ on note $A^{\otimes n}$ l'ensemble des $P^{\otimes n}$, $P \in A$). En effet : pour $n = 1$ on a égalité ;

$$\text{pour } n \geq 2 \text{ si } Q \in P \text{ et si } Q \Big|_{(B^n)^c} = P \Big|_{(B^n)^c}$$

$$\forall A \subset B^c \quad Q^{\otimes n}(A^n) = P^{\otimes n}(A^n) \text{ donc } Q(A) = P(A) \text{ et } Q \Big|_{B^c} = P \Big|_{B^c} .$$

d) Une statistique P -sommable S est B -libre en moyenne si et seulement si $S|_B$ l'est.

e) Soit F un fermé de $E \quad P(P, F) \subset \bigcap_{U \in \mathcal{O}(F)} P(P, U)$.

Si toutes les probabilités de P sont régulières (en particulier si E est un espace métrique) on a

$$P(P, F) = \bigcap_{U \in \mathcal{O}(F)} P(P, U) .$$

En effet : supposons que $Q \Big|_{F^c} \neq P \Big|_{F^c}$, il existe $A \subset F^c$ tel que $Q(A) \neq P(A)$. Il existe donc un fermé $H \subset A$ tel que $P(H) \neq Q(H)$. H^c est alors un élément de $\mathcal{O}(F)$ tel que $Q \Big|_{H^c} \neq P \Big|_{H^c}$ donc

$$[P(P, F)]^c \subset \bigcup_{U \in \mathcal{O}(F)} [P(P, U)]^c .$$

II.3. - Exemples.

a) Soit X_1, \dots, X_n un échantillon de v.a.r. (i.i.d.) de fonction de répartition F . Soit F_n la f.d.r. empirique associée à X_1, \dots, X_n . Dans le modèle statistique $(\mathbb{R}, \mathcal{B}_{\mathbb{R}}, P)^{\otimes n}$ où P est l'ensemble des lois sur $(\mathbb{R}, \mathcal{B}_{\mathbb{R}})$, la statistique $F_n(y) - F_n(x)$, où $x < y$, suit la loi binômiale $B(n, F(y) - F(x^+))$; elle est donc $]x, y[$ - libre.

b) Soit X_1, \dots, X_n , n v.a. i.i.d. à valeurs dans (E, \mathcal{B}) . Le modèle considéré est $(E, \mathcal{B}, P)^{\otimes n}$ où P est dominé par une mesure positive σ -finie. Soit B un borélien propre de E et $x_0 \in B^c$.

Proposition 1.- La statistique $S = (S_1, \dots, S_n)$ à valeurs dans $(E, \mathcal{B})^{\otimes n}$ définie par

$$\begin{cases} S_i = X_i & \text{si } X_i \in B \\ S_i = x_0 & \text{sinon} \end{cases}$$

est B^n -exhaustive dans le modèle $(E, \mathcal{B}, P)^{\otimes n}$.

Démonstration : On doit montrer que $\forall P \in \mathcal{P}$ S est exhaustive dans le modèle $(E^n, \mathcal{B}^{\otimes n}, P(P^{\otimes n}, B^n))$ qui est bien sûr dominé puisque $P(P^{\otimes n}, B^n) \subset [P(P, B)]^{\otimes n} \subset P^{\otimes n}$ (cf. remarque II.2.c.)). Supposons que la loi de (X_1, \dots, X_n) soit $Q^{\otimes n} \in P(P^{\otimes n}, B^n)$. On a alors $Q|_{B^c} = P|_{B^c}$.

Soit μ une mesure positive σ -finie dominant P , f_Q une version de $\frac{dQ}{d\mu}$ et f_P une version de $\frac{dP}{d\mu}$ coïncidant avec f_Q sur B^c .

La vraisemblance au point (X_1, \dots, X_n) s'écrit :

$$L_Q(X_1, \dots, X_n) = \prod_{i=1}^n (f_Q(S_i) 1_{B(S_i)} + 1_{B^c(S_i)}) \\ \times \prod_{i=1}^n (f_P(X_i) 1_{B^c(X_i)} + 1_{B(X_i)}) .$$

L'exhaustivité de S dans le modèle $(E^n, B^{\otimes n}, P(P^{\otimes n}, B^n))$ (ainsi que dans le modèle $(E^n, B^{\otimes n}, (P(P, B))^{\otimes n})$ résulte du théorème de factorisation.

Une façon classique d'améliorer un estimateur est, d'après le théorème de Rao-Blackwell, de le remplacer par son espérance conditionnelle par rapport à une statistique exhaustive. Nous allons donc étudier dans ce qui suit les effets du conditionnement par rapport à la statistique S définie dans la proposition 1 sur certaines familles d'estimateurs.

III - CONDITIONNEMENT PAR RAPPORT A LA STATISTIQUE S.-

III.1. - Généralités.

On considère n v.a.r. X_1, \dots, X_n définies sur un même espace probabilisé (Ω, A, Pr) indépendantes, de même loi Q inconnue appartenant à un ensemble \mathcal{P} de probabilités sur $(\mathbb{R}, B_{\mathbb{R}})$ et $\psi : \mathcal{P} \rightarrow \mathbb{R}$ un paramètre réel localisé [Berlinet, 1983, p 7] par un fermé F ($F \neq \mathbb{R}$). On se fixe un ouvert propre U appartenant à $O(F)$ et une probabilité $P \in \mathcal{P}$ telle que $P(U^c) > 0$. Dans les applications cette probabilité P proviendra en général d'une pré-estimation ou d'une information a priori (on sait par exemple que $Q|_{U^c}$ est "proche", en un sens à définir, de $P|_{U^c}$).

On se donne un estimateur d'ordre n de la forme
$$T(X_1, \dots, X_n) = \sum_{i=1}^n K_i(X_i)$$
 du paramètre $\psi(Q)$, où les $(K_i)_{1 \leq i \leq n}$ (qui dépendent de n) sont des fonctions de \mathbb{R} dans \mathbb{R} \mathcal{P} -intégrables. On note B_S la tribu engendrée par S et pour $i \in \overline{1, n}$, S_i celle engendrée par S_i .

Le schéma d'estimation ainsi fixé s'appliquera à de nombreux problèmes ; en particulier à l'estimation, séquentielle ou non, de la densité ou de certains paramètres de régression, que nous détaillerons dans la suite.

Proposition 2.- La tribu \mathcal{B}_S est U^n -exhaustive dans le modèle $(\mathbb{R}^n, \mathcal{B}_{\mathbb{R}^n}, P^{\otimes n})$ et $\sum_{i=1}^n (K_i(X_i) 1_{X_i \in U} + m_i(P) 1_{X_i \in U^c})$ est une version (dont on fera choix dans la suite) de $E_{P^{\otimes n}}^{\mathcal{B}_S}(T(X_1, \dots, X_n))$.

$m_i(P)$ désigne la valeur moyenne de K_i sur U^c pour la probabilité P :

$m_i(P) = [P(U^c)]^{-1} \int_{U^c} K_i dP$. Cette quantité sera notée $m(P)$ quand elle sera indépendante de $i \in \overline{1, n}$.

La première partie est fournie par la proposition 1. Le calcul de l'espérance conditionnelle utilise le lemme suivant :

Lemme 1.- Soit $n \geq 2$, $(\Omega_i, \mathcal{A}_i)_{i \in \overline{1, n}}$ et $(\Lambda_i, \mathcal{B}_i)_{i \in \overline{1, n}}$ $2n$ espaces mesurables et pour tout $i \in \overline{1, n}$ soit P_i une probabilité sur $(\Omega_i, \mathcal{A}_i)$. On note $P = \otimes_{i=1}^n P_i$. Soit $k \in \overline{1, n}$ et $g : (\Omega_k, \mathcal{A}_k, P_k) \rightarrow (\mathbb{C}, \mathcal{B}_{\mathbb{C}})$ une application intégrable. Pour tout $i \in \overline{1, n}$ soit $s_i : (\Omega_i, \mathcal{A}_i) \rightarrow (\Lambda_i, \mathcal{B}_i)$ une fonction mesurable et soit S l'application :

$$\left(\prod_{i=1}^n \Omega_i, \otimes_{i=1}^n \mathcal{A}_i, P \right) \longrightarrow \left(\prod_{i=1}^n \Lambda_i, \otimes_{i=1}^n \mathcal{B}_i \right)$$

$$(x_1, \dots, x_n) \longmapsto (s_1(x_1), \dots, s_n(x_n))$$

On note π_k la $k^{\text{ème}}$ projection : $\prod_{i=1}^n \Omega_i \rightarrow \Omega_k$ et S_k la tribu engendrée par s_k .

On a alors $E_{P_k}^{S_k}(g) \circ \pi_k = E_P^{\mathcal{B}_S}(g \circ \pi_k)$ P. p.s. .

Démonstration : Il est clair que $E_{P_k}^{S_k}(g) \circ \pi_k$ est \mathcal{B}_S -mesurable.

Il reste donc à vérifier que

$$\forall A \in \mathcal{B}_S \quad \int_A E_{P_k}^{S_k}(g) \circ \pi_k dP = \int_A g \circ \pi_k dP .$$

Or la tribu \mathcal{B}_S est engendrée par les ensembles de la forme $S^{-1}(B_1 \times B_2 \dots \times B_n) = \prod_{i=1}^n s_i^{-1}(B_i)$; il suffit donc de faire cette vérification sur de tels ensembles. Le théorème de Fubini entraîne que :

$$\int \prod_{i=1}^n \mathbb{1}_{s_i^{-1}(B_i)}(x_i) (E_{P_k}^{S_k}(g) \circ \pi_k)(x_1, \dots, x_n) dP(x_1, \dots, x_n) \\ = \prod_{\substack{i \in \overline{1, n} \\ i \neq k}} P_i(s_i^{-1}(B_i)) \times \int_{s_k^{-1}(B_k)} E_{P_k}^{S_k}(g)(x_k) dP_k(x_k)$$

et

$$\int \prod_{i=1}^n \mathbb{1}_{s_i^{-1}(B_i)}(x_i) (g \circ \pi_k)(x_1, \dots, x_n) dP(x_1, \dots, x_n) \\ = \prod_{\substack{i \in \overline{1, n} \\ i \neq k}} P_i(s_i^{-1}(B_i)) \int_{s_k^{-1}(B_k)} g(x_k) dP_k(x_k)$$

ce qui permet de conclure par définition de $E_{P_k}^{S_k}(g)$.

Démonstration de la proposition 2 : Le lemme 1 va nous permettre de calculer $E_n^{\mathcal{B}_S}(T(X_1, \dots, X_n))$ ce qui nous permettra ultérieurement $\prod_{i=1}^n P_i$

d'étendre facilement certains résultats au cas où les variables ne sont plus équidistribuées.

Soit $i \in \overline{1, n}$. Notons P_0 la probabilité $\prod_{j=1}^n P_j$ (pour $j \in \overline{1, n}$ $P_j(U) > 0$).

Si $n \geq 2$ le lemme 1 entraîne

$$E_{P_0}^{\mathcal{B}_S}(K_i(x_i)) = E_{P_i}^{S_i}(K)(x_i) P_0 \text{ p.s.}$$

et cette relation est évidente pour $n = 1$.

De la définition de S_i (proposition 1) il résulte que l'on a les équivalences

$$x \in U \iff S_i(x) \in U$$

$$x \in U^c \iff S_i(x) = x_0$$

D'autre part

$$K_i = K_i \cdot 1_U + K_i \cdot 1_{U^c} = (K_i \circ S_i) (1_U \circ S_i) + K_i \cdot 1_{U^c}$$

Donc
$$E_{P_i}^{S_i}(K_i) = (K_i \circ S_i) (1_U \circ S_i) + E_{P_i}^{S_i}(K_i \cdot 1_{U^c}) .$$

Sur U^c l'application S_i est constante, donc aussi $E_{P_i}^{S_i}(K_i)$ qui vaut donc $[P_i(U^c)]^{-1} \int_{U^c} K_i dP_i .$

On en déduit que

$$E_{P_0}^{B_S}(T(X_1, \dots, X_n)) = \sum_{i=1}^n (K_i(X_i) \cdot 1_{X_i \in U} + m_i(P_i) \cdot 1_{X_i \in U^c})$$

et la proposition 2.

III.2. - Propriétés de l'estimateur conditionné.

Soit $L : P \times \mathbb{R} \longrightarrow \mathbb{R}^+$ une fonction de perte qui est

$$(Q, r) \longmapsto L(Q, r)$$

telle que $\forall Q \in P \quad L(Q, \cdot)$ est convexe.

Le risque associé à l'estimateur T d'ordre n , lorsque la loi des observations est Q sera noté $R(Q, T) = \int L(Q, T(x_1, \dots, x_n)) dQ^{\otimes n}(x_1, \dots, x_n)$. Comme S est U^n -exhaustive dans le modèle $(\mathbb{R}^n, B_{\mathbb{R}^n}, P^{\otimes n})$, la proposition suivante résulte du théorème de Rao-Blackwell :

Proposition 3. - $E_{P^{\otimes n}}^{B_S}(T)$ est préférable à T dans le modèle $(\mathbb{R}^n, B_{\mathbb{R}^n}, P(P^{\otimes n}, U^n))$. Autrement dit

$$(Q \in P \text{ et } Q^{\otimes n} \Big|_{(U^n)}^c = P^{\otimes n} \Big|_{(U^n)}^c) \implies R(Q, E_{P^{\otimes n}}^{B_S}(T)) \leq R(Q, T)$$

Le paragraphe suivant sera consacré à l'étude du risque quadratique (associé à $L(Q, r) = (\psi(Q) - r)^2$).

III.3. - Forme de l'estimateur conditionné.

Il est facile de constater que $E_{P \otimes n}^{B_S}(T)$ admet une expression identique à celle de T , en remplaçant K_i par $\tilde{K}_i = K_i \mathbb{1}_U + m_i(P) \mathbb{1}_{U^c}$: on conserve K_i sur U et on le remplace par sa valeur moyenne pour la probabilité P à l'extérieur de U . Dans le cas, par exemple, de l'estimation d'une version continue de la densité $f_Q = \frac{dP}{d\mu}$ au point x on a, pour la méthode du noyau, $K_i(u) = K(u) = (nh)^{-1} H((u-x) h^{-1})$ où H est une fonction réelle $\{\lambda\} \cup \mathcal{P}$ -intégrable (λ désigne la mesure de Lebesgue sur \mathbb{R}). Si on choisit $U =]x-h, x+h[$ on a

$$\begin{aligned} \tilde{K}(u) &= (nh)^{-1} H((u-x) h^{-1}) \mathbb{1}_{]-1,1[}((u-x) h^{-1}) \\ &+ [nP(U^c)]^{-1} \int_{]-1,1[}^c H(y) f_P(x+hy) d\lambda(y) \mathbb{1}_{]-1,1[}^c((u-x) h^{-1}). \end{aligned}$$

On a donc un estimateur de la même forme que T en remplaçant H par

$$\tilde{H} = H \mathbb{1}_{]-1,1[} + h [nP(U^c)]^{-1} \left(\int_{]-1,1[}^c H(y) f_P(x+hy) d\lambda(y) \right) \mathbb{1}_{]-1,1[}^c.$$

Si l'intégrale $\int_{]-1,1[}^c H(y) f_P(x+hy) d\lambda(y)$ n'est pas nulle \tilde{H} n'est pas intégrable par rapport aux mesures positives infinies sur $]-1,1[$. On pourrait, en globalisant, reprocher à cet estimateur de ne pas être strict (i.e. de ne pas être une densité) mais dans le contexte que l'on s'est fixé, une globalisation n'a aucun sens.

Dans le cas de la méthode des fonctions orthogonales, si $f_Q \in L^2(\mu)$ et si l'on a fait choix de versions $(e_j)_{j \in \mathbb{N}}$ continues d'une base orthonormale de $L^2(\mu)$ on a :

$$K(u) = n^{-1} \sum_{j=1}^{r(n)} e_j(u) e_j(x)$$

et

$$\tilde{K}(u) = n^{-1} \sum_{j=1}^{r(n)} E_j(u) E_j(x)$$

avec $E_j(u) = \int_U e_j(u) + (e_j, f_P) \int_{U^c} 1_{U^c}(u)$ mais les $(E_j)_{j \in \mathbb{N}}$ ne constituent pas en général une base orthonormale de $L^2(\mu)$.

III.4. - Remarques.

a) si K_i est P p.s. constant sur U^c on a $\tilde{K}_i = K_i$ et $E_{P \otimes n}^{B_S}(T) = T$ (cas de la fenêtre mobile avec noyau à support inclus dans U). Si $\forall i \in \overline{1, n} \quad K_i = \lambda \int_{U^c} 1_{U^c}$ P p.s. les 2 estimateurs s'écrivent $\sum_{i=1}^n K_i(X_i) + \lambda \hat{P}_n(U^c)$ où \hat{P}_n est la mesure empirique associée à (X_1, \dots, X_n) . De façon évidente $E_{P \otimes n}^{B_S}(T)$ et T coïncident également lorsque toutes les observations tombent dans U .

b) Lorsque $Q = P$, le biais de l'estimateur conditionné est égal à celui de T .

c) Si la loi empirique \hat{P}_n est élément de \mathcal{P} (ce qui ne sera pas le cas dans la plupart des problèmes concrets qui seront envisagés), si elle charge U^c et que l'on remplace P par \hat{P}_n on a

$$m_i(\hat{P}_n) = \left(\sum_{j=1}^n \int_{U^c} 1_{U^c}(X_j) \right)^{-1} \sum_{X_j \in U^c} K_i(X_j)$$

D'où

$$\sum_{i=1}^n m_i(\hat{P}_n) \int_{U^c} 1_{U^c}(X_i) = \left(\sum_{j=1}^n \int_{U^c} 1_{U^c}(X_j) \right)^{-1} \sum_{j=1}^n \sum_{i=1}^n K_i(X_j) \int_{U^c} 1_{U^c}(X_i) \int_{U^c} 1_{U^c}(X_j)$$

et dans le cas où $\forall i \in \overline{1, n} \quad K_i = K$ on a

$$\sum_{i=1}^n m_i(\hat{P}_n) 1_{U^c}(X_i) = \sum_{j=1}^n K(X_j) 1_{U^c}(X_j)$$

et $T(X_1, \dots, X_n)$ est une version de $E_{P_n}^{\mathcal{B}_n} (T(X_1, \dots, X_n))$.

Dans le cas général $T(X_1, \dots, X_n)$ est une version de $E_{\prod_{i=1}^n \delta_{X_i}}^{\mathcal{B}_n} (T(X_1, \dots, X_n))$.

d) Le fait que \tilde{K}_i soit constant en dehors de U entraîne que les hypothèses usuelles faites en estimation fonctionnelle pour obtenir des conditions (nécessaires ou suffisantes) de convergence ayant un large champ d'application ne sont pas satisfaites par l'estimateur conditionné : ces hypothèses comportent en général des conditions d'intégrabilité par rapport à une mesure positive μ dominant la loi des $(X_i)_{1 \leq i \leq n}$ et vérifiant $\mu(U^c) = +\infty$.

IV - RISQUE QUADRATIQUE DE L'ESTIMATEUR CONDITIONNE.

On conserve les notations du III.2 avec $L(Q,r) = (\psi(Q) - r)^2$ et on suppose que $\forall i \in \overline{1,n}$ K_i est de carré P -intégrable.

Proposition 4.- Le risque quadratique de l'estimateur conditionné est donné par :

$$R = R(Q, E_{P \otimes n}^{B_S}(T)) = \sum_{i=1}^n \left\{ \int_U K_i^2 dQ - \left(\int_U K_i dQ \right)^2 + m_i(P) Q(U^c) \times \right. \\ \left. \left[m_i(P) Q(U) - 2 \int_U K_i dQ \right] \right\} \\ + \left\{ \sum_{i=1}^n \int_U K_i dQ + Q(U^c) \sum_{i=1}^n m_i(P) - \psi(Q) \right\}^2$$

Démonstration : Nous noterons E l'espérance de $E_{P \otimes n}^{B_S}(T)$

lorsque la loi commune des observations est Q et V sa variance. Les notations E_o , V_o et R_o seront relatives à T .

De l'indépendance des variables $(X_i)_{i \in \overline{1,n}}$, il résulte que

$$V = \sum_{i=1}^n \sigma^2(K_i(X_i) \mathbb{1}_{X_i \in U} + m_i(P) \mathbb{1}_{X_i \in U^c}) \\ V = \sum_{i=1}^n \left\{ \int_U K_i^2 dQ - \left(\int_U K_i dQ \right)^2 + [m_i(P)]^2 Q(U^c) - [m_i(P) Q(U^c)]^2 \right. \\ \left. - 2 Q(U^c) m_i(P) \int_U K_i dQ \right\} .$$

L'expression de R donnée dans la proposition résulte de l'égalité $R = V + [E - \psi(Q)]^2$.

Corollaire 1.- Dans le cas particulier où $\forall i \in \overline{1,n}$ $K_i = K$

on a

$$n^{-1} E = \int_U K dQ + Q(U^c) m(P)$$

$$\text{et } n^{-1} V = \int_U K^2 dQ - \left(\int_U K dQ \right)^2 + m(P) Q(U^c) \left[m(P) Q(U) - 2 \int_U K dQ \right]$$

et l'on en déduit aisément R .

Proposition 5 : Comparaison des biais et des risques quadratiques.

$$E_o - E = Q(U^c) \sum_{i=1}^n (m_i(Q) - m_i(P))$$

et
$$V_o - V = \sum_{i=1}^n \left\{ \int_{U^c} K_i^2 dQ - [m_i(P)]^2 Q(U^c) \right.$$

$$\left. + Q(U^c) [m_i(P) - m_i(Q)] [Q(U^c) (m_i(P) + m_i(Q)) + 2 \int_U K_i dQ] \right.$$

D'où l'on déduit $(R_o - R)$ qui est égal à

$$[V_o - V + (E_o - E) (E_o + E - 2 \psi(Q))].$$

Démonstration :

$$E_o - E = \sum_{i=1}^n \left(\int K_i dQ - \int_U K_i dQ - Q(U^c) m_i(P) \right)$$

et
$$\int_{U^c} K_i dQ = Q(U^c) m_i(Q) .$$

$$V_o - V = \sum_{i=1}^n \left\{ \int K_i^2 dQ - \left(\int K_i dQ \right)^2 \right\} - V$$

$$V_o - V = \sum_{i=1}^n \left\{ \int_{U^c} K_i^2 dQ - \left(\int_{U^c} K_i dQ \right)^2 - 2 \int_U K_i dQ \int_{U^c} K_i dQ \right.$$

$$\left. - [m_i(P)]^2 Q(U^c) + [m_i(P) Q(U^c)]^2 + 2 Q(U^c) m_i(P) \int_U K_i dQ \right\}$$

$$V_o - V = \sum_{i=1}^n \left\{ \int_{U^c} K_i^2 dQ - [Q(U^c) m_i(Q)]^2 - [m_i(P)]^2 Q(U^c) \right.$$

$$\left. + [m_i(P) Q(U^c)]^2 + 2 Q(U^c) \int_U K_i dQ (m_i(P) - m_i(Q)) \right\}$$

D'où l'expression de $(V_o - V)$ donnée dans la proposition.

Les expressions trouvées suggèrent l'utilisation de fonctions des quantités $(m_i(P) - m_i(Q))_{i \in \overline{1, n}}$ comme "mesure de déviation" entre probabilités pour le problème d'estimation considéré. Il n'est pas facile de déterminer en toute généralité l'ensemble des éléments de \mathcal{P} pour lesquels on a diminution du risque quadratique mais les corollaires suivants permettent d'en préciser un sous-ensemble qui contient strictement l'ensemble des éléments pour lesquels l'application du théorème de Rao-Blackwell (proposition 3) nous garantissait une diminution du risque.

Corollaire 2.-

$$V_0 - V \geq Q(U^c) \sum_{i=1}^n (m_i(Q) - m_i(P)) (Q(U) (m_i(Q) + m_i(P)) + 2 \int_U K_i dQ) .$$

Démonstration : De l'inégalité de Schwarz il résulte que

$$\forall i \in \overline{1, n} \left(\int_{U^c} K_i \mathbb{1}_{U^c} dQ \right)^2 \leq \left(\int_{U^c} K_i^2 dQ \right) Q(U^c)$$

$$\text{donc } \int_{U^c} K_i^2 dQ - Q(U^c) [m_i(P)]^2 \geq Q(U^c) [(m_i(Q))^2 - (m_i(P))^2]$$

on en déduit la minoration de $(V_0 - V)$ et une minoration correspondante de $(R_0 - R)$.

Corollaire 3.- Si $\forall i \in \overline{1, n} m_i(Q) = m_i(P)$ on a :

$$E_0 - E = 0$$

$$\text{et } R_0 - R = V_0 - V = \sum_{i=1}^n \left\{ \int_{U^c} K_i^2 dQ - Q(U^c) [m_i(Q)]^2 \right\} \geq 0$$

Sous l'hypothèse du corollaire 3 on a donc $R_0 \geq R$ et la quantité $(R_0 - R)$ ne varie pas si on remplace la probabilité P choisie au départ par une autre P' telle que $\forall i \in \overline{1, n} m_i(P') = m_i(P)$. L'amélioration est stricte $(R_0 > R)$ dès que l'on n'a pas égalité Q -presque sûre entre $K_i \Big|_{U^c}$ et $\lambda \mathbb{1}_{U^c}$ ($\lambda \in \mathbb{R}$), pour tout $i \in \overline{1, n}$.

V - APPLICATION A L'ESTIMATION DE LA DENSITE PAR LA METHODE DU NOYAU.

V.1. - Limite du biais et de la variance.

Les éléments Q de \mathcal{P} sont supposés admettre une densité par rapport à la mesure de Lebesgue λ , dont une version f_Q est bornée et continue au voisinage d'un réel fixé x . On cherche à estimer $\psi(Q) = f_Q(x)$.

Pour $u \in \mathbb{R}$ et $i \in \overline{1, n}$ on pose $K_i(u) = K(u) = (nh)^{-1} H((u-x) h^{-1})$ où H est une fonction λ -intégrable.

Le voisinage U du point x considéré est l'intervalle $]x-h, x+h[$ (h est bien sûr fonction de n). Les notations sont les mêmes que précédemment.

La méthode de conditionnement a pour but une amélioration de l'estimation à n fixé ; il est néanmoins important d'étudier les comportements asymptotiques du biais et de la variance : ils nous fourniront des indications sur le choix de H et h .

Proposition 6. : Limite du biais.

Si l'on suppose que h tend vers 0 avec n^{-1} on a

$$\lim_{n \rightarrow +\infty} E = f_Q(x) \int_{]-1,1[} H \, d\lambda + f_P(x) \int_{]-1,1[}^c H \, d\lambda$$

Démonstration :

$$E = \int_{]-1,1[} H(y) f_Q(x+hy) \, d\lambda(y) + \frac{Q(U^c)}{P(U^c)} \int_{]-1,1[}^c H(y) f_P(x+hy) \, d\lambda(y) .$$

et sous les hypothèses faites les intégrales du membre de droite sont des fonctions de h continues en 0.

Si P provient d'une préestimation et que l'on a $f_Q(x) = f_P(x)$ on a pour l'estimateur conditionné un biais asymptotique égal à $f_Q(x) \left(\int_{\mathbb{R}} H d\lambda - 1 \right)$. Il semble donc indispensable d'imposer à H la condition $\int_{\mathbb{R}} H d\lambda = 1$ et l'estimateur T est alors asymptotiquement sans biais alors que pour l'estimateur conditionné on a :

Corollaire 4.- Si $\lim_{n \rightarrow +\infty} h = 0$ et si $\int_{\mathbb{R}} H d\lambda = 1$ on a $\lim_{n \rightarrow +\infty} E - f_Q(x) = (f_P(x) - f_Q(x)) \int_{]-1,1[} H d\lambda$.

Dans le cas où $f_P(x) \neq f_Q(x)$ le biais asymptotique de $E_{P \otimes n}^{B_S}(T)$ sera strictement plus petit que $|f_P(x) - f_Q(x)|$ si et seulement si $\left| \int_{]-1,1[} H d\lambda \right| < 1$. On notera encore que le choix de H tel que $\int_{]-1,1[} H d\lambda < 0$ entraîne que $\left[\lim_{n \rightarrow +\infty} E - f_Q(x) \right]$ et $\left[f_P(x) - f_Q(x) \right]$ sont de signes contraires, ce qui peut présenter un intérêt si on veut construire des intervalles de confiance pour $E_{P \otimes n}^{B_S}(T)$ (et des "bandes de confiance" pour le graphe de la densité). Il est clair que le fait d'imposer à H de vérifier $\int_{]-1,1[} H d\lambda = 0$ entraîne que $E_{P \otimes n}^{B_S}(T)$ est asymptotiquement sans biais ; la première condition imposée à H s'écrit alors

$$\int_{]-1,1[} H d\lambda = 1 .$$

Proposition 7 : Comportement de la variance.

Si H^2 est également λ -intégrable, si $\lim_{n \rightarrow +\infty} h = \lim_{n \rightarrow +\infty} (nh)^{-1} = 0$ on a $\lim_{n \rightarrow +\infty} V = 0$ et par conséquent $\lim_{n \rightarrow +\infty} R = \left(\lim_{n \rightarrow +\infty} E - f_Q(x) \right)^2$. Plus précisément $V = (nh)^{-1} f_Q(x) \int_{]-1,1[} H^2 d\lambda + o((nh)^{-1})$.

Démonstration : Du corollaire 1 il résulte que

$$\begin{aligned}
 v = & (nh)^{-1} \int_{-1,1} H^2(y) f_Q(x+hy) d\lambda(y) - \frac{1}{n} \left(\int_{-1,1} H(y) f_Q(x+hy) d\lambda(y) \right)^2 \\
 & + \frac{1}{n} \frac{Q(U^c)}{P(U^c)} \int_{-1,1} H(y) f_P(x+hy) d\lambda(y) \left[\frac{Q(U)}{P(U^c)} \int_{-1,1} H(y) f_P(x+hy) d\lambda(y) \right. \\
 & \left. - 2 \int_{-1,1} H(y) f_Q(x+hy) d\lambda(y) \right]
 \end{aligned}$$

et la conclusion s'en déduit aisément.

V.2. - Etude du comportement asymptotique du biais et du risque quadratique sous des hypothèses complémentaires de régularité.

On suppose que les versions $(f_{Q_1})_{Q_1 \in \mathcal{P}}$ des densités définies précédemment admettent une dérivée d'ordre 4 en x et que la fonction H vérifie les conditions suivantes :

$$(C_1) \quad \left\{ \begin{array}{l} * H \text{ est paire ;} \\ * y^i H \text{ et } y^i H^2 \text{ sont } \lambda\text{-intégrables pour } i \in \overline{0,4} . \end{array} \right.$$

On pose, pour $k \in \overline{0,4}$ et $\ell \in \overline{1,2}$,

$$I(k, \ell) = \int_{-1,1} y^k H^\ell(y) d\lambda(y) \quad \text{et} \quad J(k, \ell) = \int_{-1,1}^c y^k H^\ell(y) d\lambda(y)$$

On note $r = \frac{Q(U^c)}{P(U^c)}$ et $I = f_Q(x) I(0,1) + r f_P(x) J(0,1)$.

Proposition 8.- Sous les hypothèses précédentes

$$\begin{aligned} [E - f_Q(x)]^2 &= [I - f_Q(x)]^2 \\ &+ h^2 [I - f_Q(x)] [f_Q''(x) I(2,1) + r f_P''(x) J(2,1)] \\ &+ \frac{h^4}{4} [(f_Q''(x) I(2,1) + r f_P''(x) J(2,1))^2 \\ &\quad + \frac{1}{3} (I - f_Q(x)) (f_Q^{(4)}(x) I(4,1) + r f_P^{(4)}(x) J(4,1))] \\ &+ h^4 o(1) . \end{aligned}$$

Démonstration : La formule de Taylor-Young à l'ordre 4 donne, compte tenu des conditions (C₁) : pour tout Q₁ ∈ P

$$\begin{aligned} \int_{]-1,1[} H(y) f_{Q_1}(x+hy) d\lambda(y) &= f_{Q_1}(x) + \frac{h^2}{2} f_{Q_1}''(x) \int_{]-1,1[} y^2 H(y) d\lambda(y) \\ &+ \frac{h^4}{24} f_{Q_1}^{(4)}(x) \int_{]-1,1[} y^4 H(y) d\lambda(y) + h^4 o(1) \end{aligned}$$

et une expression analogue en intégrant sur]-1,1]^c .

L'expression de E donnée dans la démonstration de la proposition 6 permet alors d'écrire :

$$\begin{aligned} E &= f_Q(x) I(0,1) + r f_P(x) J(0,1) \\ &+ \frac{h^2}{2} (f_Q''(x) I(2,1) + r f_P''(x) J(2,1)) \\ &+ \frac{h^4}{24} (f_Q^{(4)}(x) I(4,1) + r f_P^{(4)}(x) J(4,1)) + h^4 o(1) \end{aligned}$$

$$\begin{aligned} [E - f_Q(x)]^2 &= I^2 + \frac{h^4}{4} (f_Q''(x) I(2,1) + r f_P''(x) J(2,1))^2 \\ &+ h^2 I(f_Q''(x) I(2,1) + r f_P''(x) J(2,1)) \\ &+ \frac{h^4}{12} I(f_Q^{(4)}(x) I(4,1) + r f_P^{(4)}(x) J(4,1)) \\ &- 2 f_Q(x) \times E + [f_Q(x)]^2 + h^4 o(1) \end{aligned}$$

D'où la proposition 8 en réduisant suivant les puissances de h.

Comme on l'a déjà vu le fait d'imposer à H les conditions

$$(C_2) \left\{ \begin{array}{l} \int_{-1,1} H d\lambda = 1 \\ \text{et } \int_{-1,1}^c H d\lambda = 0 \end{array} \right. \quad \text{permet d'obtenir un biais}$$

asymptotique nul.

Corollaire 5.- Si les hypothèses de la proposition 8 et les conditions (C_2) sont satisfaites on a

$$[E - f_Q(x)]^2 = \frac{h^4}{4} (f_Q''(x) I(2,1) + r f_P''(x) J(2,1))^2 + h^4 o(1)$$

$$\text{et } R = \frac{f_Q(x)}{nh} I(0,2) + \frac{h^4}{4} (f_Q''(x) I(2,1) + r f_P''(x) J(2,1))^2 + o\left(\frac{1}{nh}\right) + h^4 o(1) .$$

Nous reviendrons dans un prochain travail sur l'utilisation de ce corollaire pour le choix de H et h . Lorsque f_P devra être obtenue par préestimation nous utiliserons en général la méthode automatique de Deheuvels et Hominal [1980]. L'expression obtenue pour R conduit à une condition que l'on rencontre toujours dans ce type de problèmes : minimiser $I(0,2) = \int_{-1,1} H^2 d\lambda$ sous certaines contraintes. L'étude du second terme montre que si la préestimation peut nous renseigner (p.s.) sur le signe de $f_Q''(x)$ autrement dit si $f_P''(x)$ et $f_Q''(x)$ ont le même signe, on aura intérêt à choisir un noyau H tel que $J(2,1)$ soit négatif.

On trouvera dans le paragraphe VII un théorème qui permet de résoudre facilement certains problèmes de minimisation pour les noyaux.

VI - APPLICATION A L'ESTIMATION DE PARAMETRES DE REGRESSION PAR LA METHODE DU NOYAU.-

Il est possible de reprendre ici une grande partie de ce qui a été dit précédemment. Nous allons donc brièvement pour le problème d'estimation, indiquer la forme des estimateurs et les comportements du biais et de la variance (analogues des propositions 6 et 7). Une étude asymptotique plus précise sera faite ultérieurement.

Soit $\alpha \in \mathbb{N}^*$ et x un réel fixé. On suppose dans ce paragraphe que $E = \mathbb{R}^2$, que les éléments Q de P admettent une densité par rapport à $\lambda^{\otimes 2}$ dont une version $f_Q(u,v)$ est continue sur un rectangle ouvert contenant $\{x\} \times \mathbb{R}$ et est majorée sur ce rectangle par une fonction $g_1(v)$ admettant des moments d'ordre α et 2α par rapport à λ .

On suppose que $g(x) = \int_{\mathbb{R}} f_Q(x,v) dv$ (on utilisera la notation dv au lieu de $d\lambda(v)$) est un nombre réel connu strictement positif et on cherche à estimer $\psi(Q) = \int_{\mathbb{R}} v^\alpha f_Q(x,v) dv$ à partir de l'observation dans \mathbb{R}^2 de n couples $(X_1, Y_1), (X_2, Y_2), \dots, (X_n, Y_n)$ indépendants et de même loi Q .

Lorsque $\alpha = 1$ $\psi(Q) [g(x)]^{-1}$ est la valeur en x d'une version continue de la régression $E(Y|X)$ de Y par rapport à X .

Pour $(u,v) \in \mathbb{R}^2$ et $i \in \overline{1, n}$ on pose

$$K_i(u,v) = K(u,v) = v^\alpha (nh)^{-1} H((u-x) h^{-1})$$

où H est une fonction paire intégrable par rapport à λ et aux marges dont les densités sur \mathbb{R} sont dans $\left\{ \int f_Q(.,v) dv \mid Q \in P \right\}$.

Le voisinage U du point x considéré est $]x-h, x+h[\times \mathbb{R}$.

On a alors (avec les notations des paragraphes précédents) :

$$T((X_1, Y_1), \dots, (X_n, Y_n)) = (nh)^{-1} \sum_{i=1}^n Y_i^\alpha H((X_i - x) h^{-1})$$

$$m(P) = [nP(U^c)]^{-1} \int_{-1,1}^c H(y) \left(\int v^\alpha f_P(x + hy, v) dv \right) dy$$

par application du théorème de Fubini.

Or

$$\begin{aligned} \tilde{K}(u, v) &= v^\alpha (nh)^{-1} H((u-x) h^{-1}) \int_{-1,1}^c ((u-x) h^{-1}) \\ &+ m(P) \int_{-1,1}^c ((u-x) h^{-1}) \end{aligned}$$

et $E_{P_n}^{BS}(T)$ s'en déduit aisément.

Proposition 9 : Limite du biais.

Si l'on suppose que h tend vers 0 avec n^{-1} on a :

$$\lim_{n \rightarrow +\infty} E = \psi(Q) \int_{-1,1} H d\lambda + \psi(P) \int_{-1,1}^c H d\lambda .$$

Démonstration :

$$n \int_U K dQ = \int_{-1,1} H(y) \left(\int v^\alpha f_Q(x + hy, v) dv \right) dy .$$

Sous les hypothèses faites on a $\lim_{n \rightarrow +\infty} n m(P) = \psi(P) \int_{-1,1}^c H d\lambda$

et $\lim_{n \rightarrow +\infty} n \int_U K dQ = \psi(Q) \int_{-1,1} H d\lambda .$

La conclusion résulte du corollaire 1.

Les conséquences que l'on en tire sur H sont identiques à celles faites au paragraphe V : on choisira H de façon à ce que $\int H d\lambda = 1$ pour qu'une bonne préestimation nous garantisse un biais asymptotique nul ; on aura sous cette condition sur H :

$$\lim_{n \rightarrow +\infty} E - \psi(Q) = (\psi(P) - \psi(Q)) \int_{-1,1}^c H d\lambda .$$

Nous renvoyons au V.1.

Proposition 10 : Comportement de la variance.

Si H^2 satisfait aux mêmes conditions d'intégrabilité que H , si $\lim_{n \rightarrow +\infty} h = \lim_{n \rightarrow +\infty} (nh)^{-1} = 0$ on a

$$V = (nh)^{-1} \left(\int_{-1,1}^c H d\lambda \right) \left(\int v^{2\alpha} f_Q(x,v) dv \right) = o((nh)^{-1})$$

et donc $\lim_{n \rightarrow +\infty} R = (E - \psi(Q))^2$.

Démonstration : Les considérations précédentes et le corollaire 1 montrent que le comportement de V est gouverné par celui de $n \int_U K^2 dQ$. La proposition en résulte aisément.

VII - QUELQUES REMARQUES SUR L'AMELIORATION D'ESTIMATEURS.-

Nous faisons dans ce paragraphe quelques remarques très simples sur la modification d'estimateurs dans certains problèmes de statistique fonctionnelle, lorsque l'on veut prendre en compte des informations que l'on a, a priori, sur la fonction à estimer.

On note ψ un paramètre défini sur un ensemble de lois \mathcal{P} et à valeurs dans un espace de Riesz E (espace vectoriel réel ordonné réticulé) et pour lequel nous supposons défini un estimateur T . Nous mesurons la précision de l'estimation par une fonction croissante de $L(|\psi - T|)$ où L est une forme linéaire positive sur E . Dans les applications, L sera très souvent définie par une intégrale à cause des théorèmes de représentation des formes linéaires positives. Dans les paragraphes précédents nous avons $L(|\psi - T|) = \int |\psi - T| d\delta_x$; à la fin de celui-ci nous considérerons un exemple où $E = L_1(\mathbb{R})$ et $L(f) = \int f d\lambda$.

Nous nous posons la question suivante : si nous modifions T en T^* , quelle condition simple peut nous assurer que $L(|\psi - T^*|) \leq L(|\psi - T|)$?

Les valeurs absolues ne posent pas de problèmes dans E puisqu'on peut les exprimer à l'aide des parties positives et négatives :

avec les notations usuelles on a $h = h^+ - h^-$

$$\text{et } |h| = h^+ + h^- = 2h^+ - h = 2h^- + h.$$

L étant linéaire et croissante ces égalités entraînent de façon évidente que :

si $(L(h) \leq L(g) \text{ et } h^- \leq g^-)$

ou si $(L(h) \geq L(g) \text{ et } h^+ \leq g^+)$ on a $L(|h|) \leq L(|g|)$

si $L(h) = L(g)$ on a

$$L(|h|) \leq L(|g|) \iff L(h^+) \leq L(g^+) \iff L(h^-) \leq L(g^-)$$

Dans le cas où E est un sous-espace de l'espace des (classes pour l'égalité λ -presque sûre de) fonctions boréliennes sur \mathbb{R} on utilise les implications :

$$(\forall x \in \mathbb{R} \quad (h(x) < 0 \implies g(x) \leq h(x))) \implies h^- \leq g^-$$

$$(\forall x \in \mathbb{R} \quad (h(x) > 0 \implies h(x) \leq g(x))) \implies h^+ \leq g^+ .$$

Application à la densité dans $E = L_1(\mathbb{R})$:

Toute valeur de ψ (que nous conviendrons de noter encore ψ) est une fonction borélienne non négative sur \mathbb{R} et telle que $\int \psi \, d\lambda = 1$.

Nous supposons que $\int T \, d\lambda = \int T^* \, d\lambda = 1$ (avec les mêmes conventions de notation).

Si l'une des deux propositions suivantes

$$(P_1) \quad \forall x \in \mathbb{R} \quad (\psi(x) < T^*(x) \implies T^*(x) \leq T(x))$$

$$(P_2) \quad \forall x \in \mathbb{R} \quad (\psi(x) > T^*(x) \implies T^*(x) \geq T(x))$$

est vraie on aura $\int |\psi - T^*| \, d\lambda \leq \int |\psi - T| \, d\lambda$

La proposition (P_1) est vraie en particulier pour

$$T^* = \left(\int_A T \, d\lambda \right)^{-1} T \, 1_A \quad \text{où } A = \{x \in \mathbb{R} \mid T(x) > 0\} \quad (\text{on a donc } \int_A T \, d\lambda \geq 1) :$$

$$\forall x \in \mathbb{R} \quad (\psi(x) < T^*(x) \implies x \in A \implies T^*(x) \leq T(x)) .$$

La proposition (P₂) est vraie dans le cas suivant : on suppose que T est une densité, qu'il existe un sous-ensemble A de R tel que $\int_A T d\lambda > 0$ et $\text{supp } \varphi \subset A$ et on pose encore $T^* = \left(\int_A T d\lambda \right)^{-1} T \mathbb{1}_A$:

$$\forall x \in \mathbb{R} \quad (\varphi(x) > T^*(x) \implies x \in A \implies T^*(x) \geq T(x)) .$$

On trouvera ces deux derniers exemples dans le livre de Devroye et Györfi [1984] (théorèmes 8.3 et 8.4).

VIII - PROBLEMES D'OPTIMALITE POUR LES NOYAUX.

De nombreux domaines de la statistique conduisent à des problèmes de minimisation avec contraintes. Par exemple des préoccupations asymptotiques identiques à celles du paragraphe V conduisent invariablement à des problèmes du type Bartlett [1963] - Epanechnikov [1969] :

(P) minimiser $\int H^2 d\lambda$ dans l'ensemble des densités sur R ayant un moment d'ordre 2 fixé.

En général le traitement rigoureux de ce type de problèmes par des méthodes variationnelles, quand il est possible, est assez long. Nous donnons dans ce paragraphe une proposition dont la démonstration est très simple et qui permet de résoudre une classe importante de tels problèmes ; elle donne des conditions suffisantes pour qu'un élément soit solution d'un problème de minimisation. Son application au problème (P) est immédiate. Dans un corollaire nous montrons qu'elle entraîne un lemme de Devroye qui lui-même donne la solution de (P) comme cas particulier.

VIII.1. - Hypothèses.

Soit E un espace Riesz et H un sous-ensemble non vide du cône C des éléments ≥ 0 de E .

On suppose définie sur une partie G de $E \times E$ contenant $H \times H$ une application B à valeurs dans \mathbf{R} , positive ou nulle sur $C \times C \cap G$ et vérifiant les propriétés suivantes :

- * si $B(x,y)$ existe alors $B(y,x)$ et $B(x,-y)$ existent
et $B(y,x) = B(x,y) = -B(x,-y)$
- * si $B(x,y)$ et $B(x,z)$ existent alors $B(x, y + z)$ existe
et $B(x, y + z) = B(x,y) + B(x,z)$
- * $\forall (x,x) \in G \quad B(x,x) \geq 0$
- * si $B(x^+, x^-)$ existe, il est nul

Lemme 2.- Soit $(x,y) \in E \times E$ tel que

$B(x,x)$, $B(y,y)$ et $B(x,y)$ existent .

Alors $B(x + y, x + y)$ existe et

$$B(x + y, x + y) = B(x,x) + B(y,y) + 2B(x,y) .$$

Démonstration : $B(x,y)$ et $B(y,y)$ existent donc $B(x + y, y)$ existe et vaut $B(x,y) + B(y,y)$. De même $B(y + x, x)$ existe et vaut $B(y,x) + B(x,x)$. D'où le lemme par addition.

VIII.2. - Proposition 11. -

On suppose qu'il existe $x_0 \in E$ tel que

$$\left\{ \begin{array}{l} * \forall h \in H \quad B(x_0, h) \text{ existe et est indépendant de } h. \\ * \quad x_0^+ \in H. \end{array} \right.$$

Alors
$$B(x_0^+, x_0^+) = \inf_{h \in H} B(h, h)$$

et, si $(B(x, x) = 0 \implies x = 0)$ x_0^+ est l'unique élément de H ayant cette propriété de minimisation.

Démonstration :

Soit $h \in H$. $B(h, h)$, $B(x_0^+, h)$ et $B(x_0^+, x_0^+)$ existent donc $B(-x_0^+, h)$ et $B(-x_0^+, -x_0^+)$ aussi et (lemme 2) :

$$B(h - x_0^+, h - x_0^+) = B(h, h) + B(x_0^+, x_0^+) - 2B(x_0^+, h)$$

d'où
$$B(h, h) = B(h - x_0^+, h - x_0^+) + B(x_0^+, x_0^+) + 2B(x_0^+, h - x_0^+)$$

Or $B(x_0^+, h - x_0^+) = B(x_0^+, h - x_0^+) - B(x_0^+, h - x_0^+)$ à cause des propriétés de x_0^+ .

Pour tout $k \in H$ $B(x_0^+, k)$ et $B(x_0^+, k)$ existent donc $B(x_0^-, k)$ existe.

En particulier $B(x_0^-, x_0^+) = 0$.

D'autre part x_0^- et h étant dans C on a $B(x_0^-, h) \geq 0$
donc $B(x_0^+, h - x_0^+) = B(x_0^-, h - x_0^+) = B(x_0^-, h) - B(x_0^-, x_0^+) \geq 0$

on en déduit
$$B(h, h) \geq B(h - x_0^+, h - x_0^+) + B(x_0^+, x_0^+)$$

et la proposition s'ensuit .

VIII.3. - Remarques :

a) L'application B "ressemble beaucoup" à une forme bilinéaire et il convient d'illustrer par un exemple l'utilité du lemme 1 et le fait que l'on ait pris tant de précautions : dans certaines applications B(f,g) comportera des intégrales du type $\int fg d\lambda$ avec f et g fonctions boréliennes sur R. On a alors $B(f_1 - f_1, g_1) = 0$ sans que l'on puisse écrire, en général, $B(f_1 - f_1, g_1) = B(f_1, g_1) - B(f_1, g_1)$ puisque $B(f_1, g_1)$ peut très bien ne pas exister.

b) La première condition de la proposition 11 peut sembler forte mais nous verrons dans les applications qu'elle ne l'est pas et qu'elle traduit tout simplement les contraintes.

VIII.4. - Application au lemme 5.18 de Devroye [1984].

Corollaire 6.- (avec des notations claires)

$$\text{Pour tout } p > 0 \quad \inf_K \left(\int K^2 \right)^p \int |x|^p K \geq \left(\frac{p+1}{2p+1} \right)^p \frac{1}{2p+1} \quad \text{où}$$

l'infimum est pris sur l'ensemble des densités sur R par rapport à λ .

L'infimum est atteint pour

$$K_p(x) = \frac{p+1}{2p} (1 - |x|^p)^+$$

et toute autre densité réalisant cet infimum est λ pp égale à K_p .

Pour $p = 2$, ce résultat est dû à Epanechnikov et Bartlett.

Démonstration : Comme le remarque Devroye, K_p est une densité qui vérifie $\int |x|^p K_p = \frac{1}{2p+1}$ et $\int K_p^2 = \frac{p+1}{2p+1}$.

Il remarque en outre que la quantité à minimiser est invariante par changement d'échelle. Il suffit donc de minimiser $\int K^2$ dans l'ensemble

H des densités K de carré intégrable qui vérifient $\int |x|^p K = \frac{1}{2p+1}$.

E est l'ensemble des (classes de) fonctions réelles boréliennes et,

si fg est λ -intégrable, on pose $B(f,g) = \int fg d\lambda$ (donc

$B(f,f) = 0 \implies f = 0$ λ p.p.), et on constate que la (classe de la)

fonction $H_p(x) = \frac{p+1}{2p} (1 - |x|^p)$ satisfait les hypothèses de la proposition 11.

Nous montrerons dans un prochain travail que la proposition 11 a d'autres applications dans les problèmes de minimisation, y compris lorsque l'on n'impose pas la positivité de K .

REFERENCES.

- BARTLETT, M.S. - Statistical estimation of density functions.
Sankhya A, 25, 1963, 245-254.
- BERLINET, A. - Localisation et homogénéité d'un paramètre.
Pub. IRMA - Lille, Vol. V, fasc. 1, 1983.
- DEVROYE, L. et - Non-parametric density estimation : the L_1 view.
GYORFI, L. Wiley, 1984 (à paraître).
- DEHEUVELS, P. et - Estimation automatique de la densité.
HOMINAL, P. Revue de Stat. Appl., Vol. 28, n° 1, 25-55.
- EPANECHNIKOV, V.A. - Non-parametric estimation of a multivariate
probability density. Theor. Prob. Appl. 14, 1969,
153-158.

DEUXIEME PARTIE.

CHAPITRE 3

CONVERGENCE DES ESTIMATEURS

SPLINES DE LA DENSITE.

Résumé : On donne des conditions suffisantes de convergence ponctuelle et uniforme pour les estimateurs splines de degré deux de la densité dans le cas général. Dans un cas particulier les conditions données sont nécessaires et suffisantes ; on donne un encadrement de la vitesse de convergence et on montre la normalité asymptotique de l'estimateur.

Abstract : Sufficient conditions for pointwise and uniform convergence of quadratic spline density estimators are given in the general case. In a particular case the given conditions are necessary and sufficient ; upper and lower bounds are given for the rate of convergence and the asymptotic normality of the estimator is proved.

Classification AMS : 62 G 05

Mots-clés : Estimation de la densité.

Fonctions splines.

PLAN

Introduction.

I - ESTIMATION DE LA DENSITE PAR DES FONCTIONS SPLINES.

1. Splines cubiques (cas CL1 et CL2).
2. Noyaux splines.
3. Propriétés des estimateurs splines.

II - CONVERGENCE DES SPLINES ET BIAIS DES ESTIMATEURS ASSOCIES.

1. Expression locale.
2. Convergence des splines.
3. Espérance des estimateurs.

III - CONDITIONS SUFFISANTES DE CONVERGENCE (subdivisions quelconques).

1. Majoration des noyaux splines.
2. Notations.
3. Conditions suffisantes de convergence ponctuelle.
4. Condition suffisante de convergence uniforme presque complète.

IV - ETUDE DE L'ESTIMATEUR DANS LE CAS CL2 AVEC SUBDIVISIONS UNIFORMES.

1. Propriétés du noyau spline de type 2.
2. Conditions nécessaires et suffisantes de convergence simple.
3. Conditions nécessaires et suffisantes de convergence uniforme.
4. Vitesse de convergence.
5. Loi limite de l'estimateur.

Bibliographie.

Introduction.

Soit X une variable aléatoire réelle de loi inconnue, définie sur un espace probabilisé (Ω, \mathcal{A}, P) , et (X_1, \dots, X_n) un échantillon de taille n de X . Le théorème de Glivenko-Cantelli fournit un estimateur presque sûrement uniformément convergent de la fonction de répartition F de la loi de X : c'est la fonction de répartition empirique $F_n(t) = \frac{1}{n} \sum_{i=1}^n 1_{]-\infty, t]}(X_i)$, $t \in \mathbb{R}$.

Dans le cas où la loi de X admet une densité f à support inclus dans $[0, 1]$, on lisse l'estimateur F_n par des splines cubiques et on en déduit un estimateur de f .

Cette méthode a été présentée dans [5] par L. Boneva, D. Kendall et I. Stefanov et la convergence ponctuelle en moyenne quadratique de l'estimateur obtenu a été étudiée (dans le cas CL1 avec subdivisions uniformes) par G. Wahba ([9]).

On obtient ici des conditions de convergence ponctuelle et uniforme suivant différents modes stochastiques, dont la convergence uniforme presque complète, en utilisant les résultats de D. Bosq et J. Bleuez qui ont étudié dans [6] une classe très large d'estimateurs de la densité qui, outre les estimateurs splines, contient les estimateurs obtenus par la méthode du noyau et celle des fonctions orthogonales.

I - ESTIMATION DE LA DENSITE PAR DES FONCTIONS SPLINES.1. Splines cubiques.

Soit $\Delta : 0 = x_0 < x_1 \dots < x_N = 1$ une subdivision de $[0, 1]$ ($N > 1$). Pour toute fonction réelle ϕ définie sur $[0, 1]$

on appellera spline de type 1 (resp. type 2) de ψ et on notera S_{ψ}^1 (resp. S_{ψ}^2) la spline cubique (fonction de classe C^2 sur $[0,1]$ dont la restriction à $[x_i, x_{i+1}]$, $0 \leq i \leq N-1$, est un polynôme de degré 3 au plus) interpolant ψ sur les noeuds $(x_i)_{0 \leq i \leq N}$ avec les conditions

$$\text{CL1} \quad (S_{\psi}^1)'(0) = a \quad \text{et} \quad (S_{\psi}^1)'(1) = b$$

$$(\text{resp. CL2} : (S_{\psi}^2)''(0) = (S_{\psi}^2)''(1) = 0).$$

a est la dérivée en 0 du polynôme P_m de degré m interpolant ψ en x_0, \dots, x_m . b est la dérivée en 1 du polynôme Q_m de degré m interpolant ψ en x_{N-m}, \dots, x_N ($m \in \mathbb{N}^*$, $m \leq N$).

Les conditions imposées entraînent, pour toute fonction ψ , l'existence et l'unicité de S_{ψ}^i , $i \in \{1,2\}$.

2. Noyaux splines.

On obtient un estimateur \hat{f}_n^1 (resp. \hat{f}_n^2) de la densité f en dérivant $\hat{F}_n^1 = S_{F_n}^1$ (resp. $\hat{F}_n^2 = S_{F_n}^2$).

Pour $x \in [0,1]$ on note ψ_x l'application $[0,1] \rightarrow \mathbb{R}$
 $t \mapsto \psi_x(t) = \psi(x+t)$.

Si on pose pour $t \in [0,1]$ et $i \in \{1,2\}$

$$K_{\Delta}^i(x,t) = \frac{d}{dt} S_{\psi_x}^i(t) \quad (\text{noyaux splines})$$

la définition de F_n implique, par linéarité de l'application qui, à une fonction, associe la spline qui l'interpole,

$$\hat{f}_n^i(t) = \frac{1}{n} \sum_{k=1}^n K_{\Delta}^i(x_k, t) \quad i \in \{1,2\}$$

et les estimateurs splines appartiennent à la classe d'estimateurs de la densité étudiée dans [6]. Plus généralement, ces remarques seront

valables pour le lissage d'estimateurs qui sont déjà de la forme

$$\frac{1}{n} \sum_{k=1}^n K_r(x_k, t).$$

3. Propriétés des estimateurs splines.

S_{ψ}^1 et S_{ψ}^2 sont solutions d'un problème de minimisation ([8]).

La dérivée de S_{ψ}^2 est elle-même solution du problème :

$$\text{Trouver } \sigma \in I \text{ telle que } \int_0^1 [\sigma'(t)]^2 dt = \inf_{s \in I} \int_0^1 [s'(t)]^2 dt$$

où I est l'ensemble des fonctions s de $H^1([0,1])$ vérifiant

$$\int_{x_{i-1}}^{x_i} s(t) dt = \psi(x_i) - \psi(x_{i-1}) \quad \text{pour } 1 \leq i \leq N.$$

On en déduit les propriétés des estimateurs correspondants :

Proposition 1.

* \hat{f}_n^1 minimise $\int_0^1 [g'(t)]^2 dt$ parmi toutes les fonctions g de $H^1([0,1])$ dont une primitive interpole F_n sur les $(x_i)_{0 \leq i \leq N}$ et qui vérifient $g(0) = a$ et $g(1) = b$.

* \hat{f}_n^2 minimise $\int_0^1 [g'(t)]^2 dt$ parmi toutes les fonctions g de $H^1([0,1])$ dont une primitive interpole F_n sur les $(x_i)_{0 \leq i \leq N}$.

Les conditions (CL2) conduisent donc à un estimateur meilleur du point de vue lissage et plus simple à calculer pratiquement.

II - CONVERGENCE DES SPLINES ET BIAIS DES ESTIMATEURS ASSOCIES.

Les résultats classiques ([1]) s'appliquent (cas CL2) ou s'adaptent facilement (cas CL1).

1. Expression locale.

Soit S une spline cubique relativement à la subdivision Δ .

On pose $h_j = x_j - x_{j-1}$, $1 \leq j \leq N$

$$\lambda_j = \frac{h_{j+1}}{h_j + h_{j+1}}, \quad \mu_j = 1 - \lambda_j, \quad 1 \leq j \leq N-1$$

$$y_j = S(x_j), \quad M_j = S''(x_j), \quad 0 \leq j \leq N$$

$$d_j = \frac{6}{h_j + h_{j+1}} \left(\frac{y_{j+1} - y_j}{h_{j+1}} - \frac{y_j - y_{j-1}}{h_j} \right), \quad 1 \leq j \leq N-1.$$

On a alors, pour $t \in [x_{j-1}, x_j]$,

$$S(t) = M_{j-1} \frac{(x_j - t)^3}{6h_j} + M_j \frac{(t - x_{j-1})^3}{6h_j} + M_{j-1} \frac{h_j(t - x_j)}{6} \\ - M_j \frac{h_j(t - x_{j-1})}{6} + \frac{y_j(t - x_{j-1}) - y_{j-1}(t - x_j)}{h_j}$$

où les $(M_j)_{0 \leq j \leq N}$ vérifient le système suivant :

$$\begin{aligned} * \text{ dans le cas CL1 :} \\ (S_1) \quad \left\{ \begin{aligned} \mu_j M_{j-1} + 2M_j + \lambda_j M_{j+1} &= d_j, \quad 1 \leq j \leq N-1 \\ 2M_0 + M_1 &= \frac{6}{h_1} \left(\frac{y_1 - y_0}{h_1} - a \right) = d_0 \\ M_{N-1} + 2M_N &= \frac{6}{h_N} \left(b - \frac{y_N - y_{N-1}}{h_N} \right) = d_N \end{aligned} \right. \end{aligned}$$

$$\begin{aligned} * \text{ dans le cas CL2 :} \\ (S_2) \quad \left\{ \begin{aligned} \mu_j M_{j-1} + 2M_j + \lambda_j M_{j+1} &= d_j, \quad 1 \leq j \leq N-1 \\ M_0 = M_N &= 0 \end{aligned} \right. \end{aligned}$$

Lemme 1. - Dans chacun des cas CL1 et CL2, on a :

$$\sup_{0 \leq j \leq N} |M_j| \leq \sup_{0 \leq j \leq N} |d_j|.$$

Démonstration :

Soit M_1 la matrice de (S_1) et M_2 la matrice du système aux $(N-1)$ inconnues $(M_j)_{1 \leq j \leq N-1}$ dans le cas CL2. M_1 et M_2 sont à diagonale dominante. Une telle matrice $M = (m_{ij})_{\substack{1 \leq i \leq p \\ 1 \leq j \leq p}}$ admet un inverse dont la norme (correspondant à la norme du sup sur \mathbb{R}^p) est majorée par $\left[\min_{1 \leq i \leq p} (|m_{ii}| - \sum_{j \neq i} |m_{ij}|) \right]^{-1}$ ([1]). Il en résulte que si l'on munit \mathbb{R}^{N+1} et \mathbb{R}^{N-1} de la norme du sup, les normes correspondantes de M_1^{-1} et M_2^{-1} sont majorées par 1. D'où le lemme.

2. Convergence des splines.

Pour tout $n \in \mathbb{N}^*$ on se donne une subdivision Δ_n de $[0,1]$:
 $0 = x_0^n < x_1^n < \dots < x_N^n = 1$. On pose

$$h_j^n = x_j^n - x_{j-1}^n, \quad h_n = \min_{1 \leq j \leq N_n} h_j^n, \quad H_n = \max_{1 \leq j \leq N_n} h_j^n$$

(on omettra les indices "n" chaque fois que possible).

On suppose que $\lim_{n \rightarrow +\infty} h_n = 0$ et qu'il existe $C_0 \in \mathbb{R}_+^*$

tel que $\forall n \in \mathbb{N}^* \quad \frac{H_n}{h_n} \leq C_0$.

Pour toute fonction réelle g bornée sur $[0,1]$ on note

$$\|g\|_\infty = \sup_{t \in [0,1]} |g(t)|.$$

Proposition 2. - Si F_0 est de classe C^{m+1} et si S_n est, pour $n \in \mathbb{N}^*$, la spline de type 1 interpolant F_0 sur Δ_n , on a

$$\|S_n - F_0\|_\infty = O(H_n^2)$$

et $\|S_n' - F_0'\|_\infty = O(H_n)$.

Démonstration : P'_m est le polynôme d'interpolation de F'_0 en $\theta_1, \dots, \theta_m$ où $\theta_i \in]x_{i-1}, x_i[$, $1 \leq i \leq m$.

F'_0 étant de classe C^m , il existe $\theta \in [0, \theta_m]$ tel que

$$F'_0(0) - P'_m(0) = F'_0(0) - a = \frac{V_{m-1}(0)}{m!} F_0^{(m+1)}(\theta)$$

(expression de l'erreur d'interpolation)

avec

$$V_{m-1}(x) = \prod_{i=1}^m (x - \theta_i) .$$

On obtient une expression analogue pour la différence

$(F'_0(1) - b)$ et on en déduit une majoration de $\sup_{1 \leq j \leq N_n} |d_j|$. Le

lemme 1 et l'expression locale de la spline et de sa dérivée permettent de conclure.

Proposition 3. - Si $F_0 \in H^2([0,1])$ et si S_n est, pour $n \in \mathbb{N}^*$, la spline de type 2 interpolant F_0 sur Δ_n on a

$$\|S_n - F_0\|_{\infty} \leq \frac{1}{2} \|F_0''\|_{L^2} \times H_n^{\frac{3}{2}}$$

et

$$\|S'_n - F'_0\|_{\infty} \leq \|F_0''\|_{L^2} \times H_n^{\frac{1}{2}} .$$

Pour les démonstrations détaillées voir [2].

3. Espérance des estimateurs splines.

De l'expression locale des splines, il résulte la

Proposition 4. - Dans chacun des cas CL1 et CL2, l'espérance de la spline est la spline de même type interpolant F sur les mêmes noeuds ; l'espérance de la dérivée de la spline est la dérivée de la spline de même type interpolant F sur les mêmes noeuds.

Des propositions 2, 3 et 4, il découle la

Proposition 5. - Si F est de classe C^{m+1} (resp. appartient à $H^2([0,1])$) \hat{F}_n^1 (resp. \hat{F}_n^2) est un estimateur uniformément asympto-

tiquement sans biais (E.U.A.S.B.) de F et \hat{f}_n^1 (resp. \hat{f}_n^2) est un E.U.A.S.B. de f .

III - CONDITIONS SUFFISANTES DE CONVERGENCE (SUBDIVISIONS QUELCONQUES).

1. Majoration des noyaux splines.

Proposition 6.- $\sup_{\substack{x \in [0,1] \\ t \in [0,1]}} |K_{\Delta_n}^i(x,t)| \leq \frac{D}{h_n} \quad i \in \{1,2\}$

D est une constante ne dépendant que de C_0 si $i = 2$, de C_0 et m si $i = 1$.

Démonstration : De la définition des noyaux splines, il résulte que $\sup_{1 \leq j \leq N-1} |d_j| \leq \frac{3}{h_n^2}$.

Dans le cas CL1, on a :

$$a = \sum_{v=0}^m \ell'_{0,v}(0) F_n(x_v) \quad \text{où } \ell_{0,v}, \quad 0 \leq v \leq m,$$

est le polynôme de degré m prenant la valeur 1 en x_v et la valeur 0 en x_j , $0 \leq j \leq m$, $j \neq v$

$$\forall t \in [0,1] \quad \ell_{0,v}(t) = \prod_{\substack{0 \leq j \leq m \\ j \neq v}} (t-x_j) \times \prod_{\substack{0 \leq j \leq m \\ j \neq v}} (x_v-x_j)^{-1}.$$

On en tire les majorations :

$$|\ell'_{0,v}(0)| \leq \frac{m! C_0^{m-1}}{h_n}$$

et $|a| \leq \frac{C}{h_n}$ avec $C = (m+1)! C_0^{m-1}$.

De même $|b| \leq \frac{C}{h_n}$

grâce au lemme 1, on déduit alors que

$$\sup_{0 \leq j \leq N} |M_j| \leq \frac{6(1+C)}{h_n^2} \quad \text{dans le cas CL1}$$

et

$$\sup_{0 \leq j \leq N} |M_j| \leq \frac{3}{h_n^2} \quad \text{dans le cas CL2.}$$

On conclut en utilisant l'expression locale du noyau spline.

2. Notations.

On choisit un type de noyaux pour estimer f et on désigne par K_n le noyau spline relatif à la subdivision Δ_n .

$$\text{On pose } \hat{f}_n(t) = \frac{1}{n} \sum_{k=1}^n K_n(X_k, t)$$

$n \in \mathbb{N}^*$

$$f_n(t) = E(\hat{f}_n(t)) = \int_0^1 K_n(x, t) f(x) dx \quad t \in [0, 1]$$

\mathcal{D}_1 est l'ensemble des densités définies sur $[0, 1]$, bornées et telles que $\lim_{n \rightarrow +\infty} f_n(t)$ existe pour tout $t \in [0, 1]$; cette limite est notée $\bar{f}(t)$.

$$\mathcal{D}_2 = \{g \mid g \in \mathcal{D}_1 \text{ et } \lim_{n \rightarrow +\infty} \|f_n - f\|_\infty = 0\}$$

$$\hat{S}_n(t) = \hat{f}_n(t) - f_n(t)$$

$$D_n(t) = |\hat{f}_n(t) - \bar{f}(t)| \quad t \in [0, 1], \quad n \in \mathbb{N}^*$$

$$\delta_n = \|\hat{f}_n - f\|_\infty.$$

Remarque 1. - De la proposition 5, il résulte que

* dans le cas CL1 $C^m([0, 1]) \subset \mathcal{D}_2$

* dans le cas CL2 $H^1([0, 1]) \subset \mathcal{D}_2$.

3. Conditions suffisantes de convergence ponctuelle.

La proposition 6 entraîne que les conditions I et II de [6]

page 487 sont vérifiées avec $\alpha = 2$.

On en déduit :

Proposition 7. - Soit $t \in [0, 1]$ $f \in \mathcal{D}_1$.

Si $\lim_{n \rightarrow +\infty} n h_n^2 = +\infty$ alors $\lim_{n \rightarrow +\infty} E([D_n(t)]^2) = 0$.

Proposition 8. - Soit $t \in [0, 1]$ $f \in \mathcal{D}_1$.

Si pour tout $\beta > 0$, $\sum_{n \in \mathbb{N}^*} \exp(-\beta n h_n^2) < +\infty$ alors

$D_n(t) \rightarrow 0$ presque complètement.
 $n \rightarrow +\infty$

La démonstration de cette dernière proposition est basée sur une inégalité de Hoeffding ([4]) et met en évidence la majoration :

$$P(\hat{S}_n(t) \geq \varepsilon) \leq \exp(-\beta_1 \varepsilon n h_n^2)$$

où β_1 est une constante et ceci pour ε suffisamment petit.

4. Condition suffisante de convergence uniforme.

Proposition 9. - Si $f \in \mathcal{D}_2$ et si $\lim_{n \rightarrow +\infty} (n h_n^2)^{-1} \text{Log } n = 0$

alors $\delta_n \xrightarrow{n \rightarrow +\infty} 0$ presque complètement

et $E\delta_n \xrightarrow{n \rightarrow +\infty} 0$.

Démonstration : On applique le lemme 2, p. 491 de [6]. Pour cela il reste à montrer que le noyau spline vérifie la condition de Lipschitz V p. 488 du même article : de l'expression locale de

$\frac{d}{dt} K_n(x, t)$ et des majorations obtenues dans la démonstration de la proposition 6 pour $\sup_{0 \leq j \leq N} |M_j|$ on déduit que

$$\left| \frac{d}{dt} K_n(x, t) \right| \leq \frac{12(1+C)}{h_n^2} \quad \forall x \in [0, 1] \quad \forall t \in [0, 1]$$

et donc $\sup_{x \in [0, 1]} |K_n(x, t') - K_n(x, t)| \leq \frac{12(C+1)}{h_n^2} |t-t'|$

IV - ETUDE DE L'ESTIMATEUR DANS LE CAS CL2 AVEC SUBDIVISIONS UNIFORMES.

Dans le cas de subdivisions uniformes on peut expliciter les inverses des matrices M_1 et $M_2([1])$ et calculer en fonction du pas h les coefficients M_j^k qui déterminent le noyau spline.

Dans toute la suite la notation K_n désignera le noyau spline de type 2 relativement à Δ_n (de pas h_n).

1. Propriétés du noyau spline de type 2.

Proposition 10.-

$$a) \quad \forall n \in \mathbb{N}^* \quad \int_0^{h_n} K_n(x,0) dx > 1.$$

$$b) \quad \text{il existe } A > 0, \quad B > 0 \text{ tels que pour } n \text{ assez grand on ait } \forall t \in [0,1] \quad \frac{A}{h_n} \leq \int_0^1 [K_n(x,t)]^2 dx \leq \frac{B}{h_n}.$$

Démonstration :

On utilise l'expression locale du noyau spline, en remplaçant les coefficients M_j^k par leur valeur en fonction de h_n (on remarquera que les indices j et k sont des fonctions de n).

Les calculs sont détaillés dans [2].

Remarque 2.- Soit $I_n(t) = h_n \int_0^1 [K_n(x,t)]^2 dx.$

$$\text{On montre que } \lim_{n \rightarrow +\infty} I_n(0) = \lim_{n \rightarrow +\infty} I_n(1) = \sqrt{3}.$$

Pour $t \in]0,1[$ $I_n(t)$ est encadré par A et B , pour n assez grand, mais n 'a pas de limite en général, sauf si l'on choisit des subdivisions particulières (dépendant du point t considéré).

La proposition 10 permet d'appliquer les résultats de [6] :

2. Conditions nécessaires et suffisantes de convergence simple.

Proposition 11. (Convergence en moyenne d'ordre 1 et 2).-

Les 3 conditions suivantes sont équivalentes :

$$a) \lim_{n \rightarrow +\infty} n h_n = +\infty$$

$$b) \forall f \in \mathcal{D}_1 \quad \forall t \in [0,1] \quad \lim_{n \rightarrow +\infty} E[D_n(t)] = 0$$

$$c) \forall f \in \mathcal{D}_1 \quad \forall t \in [0,1] \quad \lim_{n \rightarrow +\infty} E([D_n(t)]^2) = 0.$$

Proposition 12. (Convergence presque complète et en moyenne d'ordre 1 et 2).

Les 3 conditions suivantes sont équivalentes :

$$a) \forall \varepsilon > 0 \quad \sum_{n \in \mathbb{N}^*} \exp[-\varepsilon n h_n] < +\infty.$$

$$b) \forall t \in [0,1] \quad \forall f \in \mathcal{D}_1 \quad D_n(t) \xrightarrow[n \rightarrow +\infty]{} 0 \text{ presque complètement}$$

et $E(D_n(t)) \xrightarrow[n \rightarrow +\infty]{} 0.$

$$c) \forall t \in [0,1] \quad \forall f \in \mathcal{D}_1 \quad D_n(t) \xrightarrow[n \rightarrow +\infty]{} 0 \text{ presque complètement}$$

et $E([D_n(t)]^2) \xrightarrow[n \rightarrow +\infty]{} 0.$

3. Conditions nécessaires et suffisantes de convergence uniforme.

Définition.- On dira que la suite $(z_n)_{n \in \mathbb{N}^*}$ est asymptotiquement concave si et seulement si

$$(C) \quad \left\{ \begin{array}{l} \text{il existe } a_1 > 0, \quad b_1 > 0 \text{ et une fonction concave } g_1 \\ \text{telle que } a_1 z_n \leq g_1(n) \leq b_1 z_n \text{ pour } n \text{ assez grand.} \end{array} \right.$$

Proposition 13. (Equivalence entre convergence simple et convergence uniforme).

Si $(\frac{1}{h_n})_{n \in \mathbb{N}^*}$ vérifie (C) les propositions suivantes sont

équivalentes :

$$a) \lim_{n \rightarrow +\infty} (n h_n)^{-1} \text{Log } n = 0.$$

b) $\forall t \in [0,1] \quad \forall f \in \mathcal{D}_2 \quad D_n(t) \xrightarrow[n \rightarrow +\infty]{} 0$ presque complètement et $E[D_n(t)] \xrightarrow[n \rightarrow +\infty]{} 0$.

$$c) \forall f \in \mathcal{D}_2 \quad \delta_n \xrightarrow[n \rightarrow +\infty]{} 0 \text{ presque complètement et } E(\delta_n) \xrightarrow[n \rightarrow +\infty]{} 0.$$

Remarque 3. - On trouvera dans [7] une condition nécessaire et suffisante de convergence uniforme presque complète en dimension 2.

4. Vitesse de convergence.

Proposition 14. - Si $f \in \mathcal{D}_2$, s'il existe $t_1 \in [0,1]$ et $E > 0$ tels que $\int_0^1 [K_n(x, t_1)]^2 f(x) dx \geq \frac{E}{h_n}$ pour $n \in \mathbb{N}^*$, alors il existe $b_2 > 0$ tel que pour n assez grand

$$\exp[-\rho'' \eta^2 n h_n] \leq P(\delta_n \geq \eta) \leq 2 \exp[-\rho' \eta n h_n]$$

pour $0 < \eta < b_2$ où ρ' et ρ'' sont des constantes positives ne dépendant que de la suite $(h_n)_{n \in \mathbb{N}^*}$ et de f .

De la minoration de $I_n(t)$ (remarque 2) il résulte que l'encadrement de la proposition 14 est valable pour toute densité de \mathcal{D}_2 qui est minorée sur $[0,1]$ par un réel strictement positif.

5. Loi limite de l'estimateur.

Proposition 15. - Si f est un élément de \mathcal{D}_1 (en particulier si $f \in H^1([0,1])$) minoré par $m_0 > 0$ et si $\lim_{n \rightarrow +\infty} n h_n = +\infty$ alors pour tout $t \in [0,1]$, la variable aléatoire $[\sigma(\hat{f}_n(t))]^{-1} (\hat{f}_n(t) - f_n(t))$ converge en loi vers une variable aléatoire normale centrée réduite.

Démonstration : Soit f un élément de \mathcal{D}_1 minoré par $m_0 > 0$ et $t \in [0,1]$. Si l'on pose :

$$\xi_{n,k} = K_n(X_k, t) - f_n(t), \quad n \in \mathbb{N}^* \quad 1 \leq k \leq n$$

$$\text{et } s_n = \sqrt{n} \sigma(K_n(X_1, t))$$

$$\text{on a } [\sigma(\hat{f}_n(t))]^{-1} (\hat{f}_n(t) - f_n(t)) = \frac{1}{s_n} \sum_{k=1}^n \xi_{n,k}.$$

Pour $n \in \mathbb{N}^*$ $\xi_{n,1}, \dots, \xi_{n,n}$ sont des variables aléatoires indépendantes, de moyenne nulle et de variance $\sigma^2(K_n(X_1, t))$.

La proposition découlera donc du théorème central limite de Lindeberg-Feller ([3]) dès que l'on aura montré que la condition

$$(C_1) \quad \forall \varepsilon > 0 \quad \lim_{n \rightarrow +\infty} \frac{1}{s_n^2} \sum_{k=1}^n \int_{|\xi_{n,k}| \geq \varepsilon s_n} \xi_{n,k}^2 dP = 0$$

est vérifiée.

De la proposition 10, il résulte que pour n assez grand

$$\int_0^1 [K_n(x, t)]^2 f(x) dx \geq \frac{m_0 A}{h_n}.$$

D'autre part $\sigma^2(K_n(X_1, t)) = \int_0^1 [K_n(x, t)]^2 f(x) dx - [f_n(t)]^2$
avec $\lim_{n \rightarrow +\infty} f_n(t) = \bar{f}(t)$.

On a donc pour n assez grand $\sigma^2(K_n(X_1, t)) \geq \frac{m_0 A}{2h_n}$.

De la majoration du noyau spline (proposition 6) on tire alors

$$\frac{\sup \xi_{n,k}^2}{s_n^2} \leq \frac{4D^2}{h_n^2} \times \frac{2h_n}{n m_0 A} = \frac{8D^2}{m_0 A} \times \frac{1}{n h_n}.$$

$$\text{Puisque } \int_{|\xi_{n,k}| \geq \varepsilon s_n} \xi_{n,k}^2 dP \leq \sup \xi_{n,k}^2 \frac{\sigma^2(K_n(X_1, t))}{\varepsilon^2 s_n^2}$$

l'hypothèse $\lim_{n \rightarrow +\infty} n h_n = +\infty$ entraîne que la condition (C_1) est vérifiée.

BIBLIOGRAPHIE

- [1] J.H. AHLBERG, E.N. NILSON et J.L. WALSH.
The theory of splines and their applications, Academic Press, 1967.
- [2] A. BERLINET.
Espaces autoreproduisants et mesure empirique.
Méthodes splines en estimation fonctionnelle. Thèse de 3ème cycle,
Lille I, 1980.
- [3] P. BILLINGSLEY.
Convergence of probability measures.
John Wiley and sons, 1968.
- [4] J. BLEUEZ.
Conditions nécessaires et suffisantes de convergence pour une
classe d'estimateurs de la densité, Thèse de 3ème cycle, Lille I,
1976.
- [5] L. BONEVA, D. KENDALL et I. STEFANOV.
Spline transformations. Three new diagnostic aids for the sta-
tistical data-analyst, Journal of the Royal Statistical Society,
Vol. 33, n° 1, 1971.
- [6] D. BOSQ et J. BLEUEZ.
Etude d'une classe d'estimateurs non-paramétriques de la densité,
Annales de l'Institut Henri Poincaré XIV, n° 4, 1978, p. 479-498.
- [7] M. DELABROYE.
Convergence uniforme de certains estimateurs de la densité par
la méthode spline.
C.R. Acad. Sci., 289, Série A, 1979, p. 825-827.
- [8] P.J. LAURENT.
Approximation et Optimisation.
Hermann, 1972.
- [9] G. WAHBA.
Interpolating spline methods for density estimation I,
Equi-spaced knots,
Ann. of Stat., 3, 1975, p. 30-48.

TROISIEME PARTIE.

CHAPITRE 4

APPROXIMATION FORTE D'INTEGRALES STOCHASTIQUES EMPIRIQUES

MULTIDIMENSIONNELLES EN VARIABLES MELANGEANTES.

Résumé : On montre l'existence d'un processus de Kiefer généralisé, indexé par $G \times [1, +\infty[$, où G est un ensemble de fonctions sur $[0,1]^d$, approximant presque sûrement les intégrales stochastiques des éléments de G par rapport au processus empirique associé à des variables aléatoires à valeurs dans $[0,1]^d$ fortement mélangeantes et strictement stationnaires.

Abstract : Let $R(g,t)$, $(g,t) \in G \times [1, +\infty[$, where G is a set of functions on $[0,1]^d$, be the empirical process associated with strong mixing strictly stationary random variables with values in $[0,1]^d$. We give an almost sure approximation of R by a Kiefer process.

CLASSIFICATION AMS (MOS) 60F15.

Mots clés : Approximation forte, forte mélangeance, multivarié, processus empirique, processus de Kiefer.

Key Words : Strong approximation, strong mixing condition, multivariate, empirical process, Kiefer process.

P L A N.

- I - INTRODUCTION.
- II - NOTATIONS ET HYPOTHESES.
- III - APPROXIMATION DES INTEGRALES STOCHASTIQUES DES ELEMENTS DE C^d PAR RAPPORT AU PROCESSUS EMPIRIQUE.
- IV - EXTENSION DU PROCESSUS κ AUX LIMITES UNIFORMES D'ELEMENTS DE C_M^d .
- V - APPROXIMATION FORTE DANS DES ESPACES A NOYAU REPRODUISANT.
- VI - AUTRE MODE D'EXTENSION DU PROCESSUS DE KIEFER.

BIBLIOGRAPHIE.

I - INTRODUCTION.-

Philipp et Pinzur ont publié en 1980 un théorème d'approximation presque sûre du processus empirique multivarié par un processus de Kiefer dans le cas où les variables forment une suite fortement mélangeante. Nous nous proposons ici de fonctionnaliser ce résultat en utilisant une méthode introduite par Ibero, également en 1980, et qu'il avait appliquée à des variables indépendantes. Il s'agit donc de prouver que l'on peut construire un processus de Kiefer généralisé, indexé par $G \times [1, +\infty[$ où G est un ensemble de fonctions, approximant presque sûrement les intégrales stochastiques des éléments de G par rapport au processus empirique associé à des variables fortement mélangeantes. Les résultats donnés ici feront l'objet d'application en statistique fonctionnelle.

II - NOTATIONS ET HYPOTHESES.-

Soit $(X_n)_{n \geq 1}$ une suite strictement stationnaire de vecteurs aléatoires définis sur $(\Omega_1, \mathcal{A}_1, P_1)$, à valeurs dans $[0, 1]^d$, de loi π et de fonction de répartition F . On suppose que π admet une densité par rapport à la mesure de Lebesgue sur $[0, 1]^d$, dont une version f est bornée par A . On note $M_a^b((a, b) \in \mathbb{N}^* \times \mathbb{N}^*)$ la tribu engendrée par les variables $(X_n)_{a \leq n \leq b}$ et

$$\rho(n) = \sup_{k \in \mathbb{N}^*} \sup_{\substack{A \in M_{k, k+n}^k \\ B \in M_{k+n, k+n}^1}} |P(A \cap B) - P(A)P(B)|$$

On suppose que $(X_n)_{n \geq 1}$ est fortement mélangeante ($\rho(n) \xrightarrow[n \rightarrow +\infty]{} 0$) avec

l'hypothèse supplémentaire $\rho(n) = O(n^{-4-d(1+\epsilon)})$ ($\epsilon \in]0, \frac{1}{4}[$). Ceci peut paraître artificiel mais est indispensable pour pouvoir exploiter le résultat de Philipp et Pinzur.

Dans $[0,1]^d$ $(a_1, \dots, a_d) \leq (b_1, \dots, b_d)$ signifiera

$\forall i \in \overline{1, d}$ $a_i \leq b_i$; on notera $s \wedge s'$ la borne inférieure de s et s' .

Pour $(s, s') \in [0,1]^{2d}$ et $n \in \mathbb{N}^*$ on pose

$$g_n(s) = 1_{\{X_n \leq s\}} - F(s)$$

$$\text{et } \Gamma(s, s') = E(g_1(s) g_1(s')) + \sum_{n=2}^{+\infty} E(g_1(s) g_n(s')) + \sum_{n=2}^{+\infty} E(g_1(s') g_n(s))$$

(sous les hypothèses faites ces deux séries sont absolument convergentes).

On définit le processus empirique de $(X_n)_{n \geq 1}$ par

$$R(s, t) = \tilde{t} (F_{\tilde{t}}(s) - F(s)) \quad s \in [0,1]^d \quad t \in [1, +\infty[$$

$$F_{\tilde{t}}(s) = \tilde{t}^{-1} \sum_{n=1}^{\tilde{t}} 1_{\{X_n \leq s\}} \quad (\tilde{t} \text{ est la partie entière de } t).$$

Théorème 1 [d'après Philipp et Pinzur, 1980].

Sous les hypothèses précédentes et sans changer sa distribution, on peut redéfinir le processus empirique $R(s, t)$, $s \in [0,1]^d$, $t \in [1, +\infty[$, sur un espace de probabilité (Ω, \mathcal{A}, P) sur lequel il existe un processus de Kiefer $K(s, t)$, $s \in [0,1]^d$, $t \in [1, +\infty[$ de fonction de covariance $\min(t, t') \times \Gamma(s, s')$ vérifiant

$$\sup_{1 \leq t \leq T} \sup_{s \in [0,1]^d} |R(s, t) - K(s, t)| = O(T^{1/2} (\text{Log } T)^{-\lambda}) \text{ p.s.}$$

où λ est une constante strictement positive ne dépendant que de d et ε .

K est un processus gaussien séparable centré.

Nous allons étendre la définition de R et K à $G \times [1, +\infty[$ où G est un ensemble de fonctions réelles π -intégrables de façon que

$$\sup_{1 \leq t \leq T} \sup_{g \in G} |R(g,t) - K(g,t)| = O(T^{1/2} (\log T)^{-\lambda}) \quad \text{p.s.}$$

La première extension de ce type a été faite par P. Revesz en 1976. Dans le cas de variables i.i.d. il a montré comment l'on pouvait passer d'une approximation presque sûre dans $[0,1]^2$ par une suite de ponts browniens ou de processus de Kiefer à une approximation presque sûre dans un ensemble d'indicateurs de boréliens de frontière suffisamment régulière. Dans le même cadre i.i.d., Ibero, en s'inspirant de la méthode de Revesz est passé à la classe des fonctions p fois différentiables sur $[0,1]^d$, bornées, ainsi que leurs dérivées partielles jusqu'à l'ordre p ($p > \frac{d}{2}$). Nous montrerons dans la dernière partie pourquoi cette technique ne peut pas convenir ici.

Nous utiliserons une autre méthode d'extension, également mise au point par Ibero [5] et utilisant une formule d'intégration par parties pour les intégrales stochastiques.

Pour $g \in L^1(\pi)$ $R(g,t)$ est l'intégrale stochastique de g par rapport au processus empirique $R(s,t)$:

$$R(g,t) = \int_{[0,1]^d} g(s) dR(s,t) = \sum_{i=1}^{\lfloor t \rfloor} (g(X_i) - \int g dF)$$

$R(g,t)$ est de façon évidente linéaire en g ; nous construirons $K(g,t)$ de façon à ce qu'il ait la même propriété, ce qui nous permettra d'obtenir les résultats de la quatrième et de la cinquième partie.

On posera $G_i(g) = g(X_i) - \int g dF$

et
$$\Lambda(g, g') = E(G_1(g) G_1(g')) + \sum_{n=2}^{+\infty} E(G_1(g) G_n(g'))$$

$$+ \sum_{n=2}^{+\infty} E(G_n(g) G_1(g'))$$

(séries absolument convergentes) .

Les v.a.r. $G_i(g)$ sont fortement stationnaires, centrées, ρ -fortement mélangées et linéaires en g .

Si $\|g\|_\infty = \sup_{s \in [0,1]^d} |g(s)| < +\infty$ on a, de façon évidente,

$$|G_i(g)| \leq 2 \|g\|_\infty$$

$$E(|G_i(g)|^2) \leq 2 \|g\|_\infty^2$$

et
$$|R(g, t)| \leq 2\sqrt{t} \|g\|_\infty$$

C_M^d désignera l'ensemble des fonctions de l'ensemble C^d des fonctions d fois continûment différentiables sur $[0,1]^d$ qui sont bornées par M , ainsi que leurs dérivées partielles jusqu'à l'ordre d .

Définition 1 : processus de Kiefer généralisé.

Soit E un ensemble et D un sous-ensemble de \mathbb{R}^+ . Un processus gaussien centré indexé par $E \times D$ sera appelé processus de Kiefer si sa fonction de covariance est de la forme :

$$E(K(\alpha, t) K(\beta, u)) = \min(t, u) \Gamma(\alpha, \beta)$$

Pour t fixé dans $D - \{0\}$, $t^{-1/2} K(\alpha, t)$ est un processus gaussien centré de fonction de covariance $\Gamma(\alpha, \beta)$.

Pour α fixé dans E tel que $\Gamma(\alpha, \alpha) > 0$ le processus $[\Gamma(\alpha, \alpha)]^{-1/2} K(\alpha, t)$ est un processus de Wiener.

III - APPROXIMATION DES INTEGRALES STOCHASTIQUES DES ELEMENTS DE C^d PAR RAPPORT AU PROCESSUS EMPIRIQUE.-

Théorème 2.-

Il existe sur (Ω, \mathcal{A}, P) un processus de Kiefer K indexé par $C^d \times [1, +\infty[$ tel que l'application $g \mapsto K(g, t)$ soit presque sûrement pour tout t une forme linéaire sur C^d bornée sur tout C_M^d et une constante C (indépendante de T et M) tels qu'on ait, presque sûrement, pour tout $T \in]1, +\infty[$,

$$\sup_{1 \leq t \leq T} \sup_{g \in C_M^d} |R(g, t) - K(g, t)| \leq CM T^{1/2} (\text{Log } T)^{-\lambda}$$

La fonction de covariance de $K(g, t)$ est donnée par $E(K(g, t) K(g', u)) = \min(t, u) \Lambda(g, g')$.

La définition de $K(g, t)$ (nous noterons de la même façon les différentes extensions du processus de Kiefer) et la démonstration du théorème 2 reposent sur les lemmes suivants :

Lemme 1 [Ibero, 1980, p. 352].

Pour toute fonction g de C^d on peut écrire :

$$g(\xi_1, \dots, \xi_d) = g(1, \dots, 1)$$

$$+ \sum_{p=1}^d \sum_{1 \leq i_1 < \dots < i_p \leq d} (-1)^p \int_{\xi_{i_1}}^1 \dots \int_{\xi_{i_p}}^1 \frac{\partial^p g}{\partial x_{i_1} \dots \partial x_{i_p}} (1, \dots, 1, x_{i_1}, \dots, x_{i_p}, 1, \dots, 1) dx_{i_1} \dots dx_{i_p}$$

Lemme 2 [Ibero, 1980, p. 354].

Pour toute probabilité ν sur $([0,1]^d, \mathcal{B}([0,1]^d))$ on a, pour toute fonction intégrable ψ par rapport à la mesure de Lebesgue sur $[0,1]^d$:

$$\int_{[0,1]^d} \int_{\xi_{i_1}}^1 \dots \int_{\xi_{i_p}}^1 \psi(1, \dots, 1, x_{i_1}, \dots, x_{i_p}, 1, \dots, 1) dx_{i_1} \dots dx_{i_p} d\nu(\xi)$$

$$= \int_{[0,1]^d} \psi(1, \dots, 1, x_{i_1}, \dots, x_{i_p}, 1, \dots, 1) F_\nu(1, \dots, 1, x_{i_1}, \dots, x_{i_p}, 1, \dots, 1) dx_{i_1} \dots dx_{i_p} .$$

Dans la suite le signe \sum sans autre indication signifiera $\sum_{p=1}^d \sum_{1 \leq i_1 < \dots < i_p \leq d}$

et un d-uple tel que $(1, \dots, 1, x_{i_1}, \dots, x_{i_p}, 1, \dots, 1)$, où l'on ne trouve que des 1 en dehors des places i_1, \dots, i_p , sera noté (***) .

Les lemmes 1 et 2 permettent donc d'écrire pour toute fonction g de C^d :

$$R(g,t) = \int_{[0,1]^d} g(x) dR(x,t) = \sum (-1)^p \int_{[0,1]^p} \frac{\partial^p g}{\partial x_{i_1} \dots \partial x_{i_p}} (***)$$

$$R(***), t) dx_{i_1} \dots dx_{i_p} .$$

Lemme 3.-

Le processus $K(s,t)$, $s \in [0,1]^d$, $t \in [1, +\infty[$ est presque sûrement à trajectoires continues.

Démonstration :

Elle consiste à appliquer une condition suffisante classique de continuité presque sûre des trajectoires d'un processus gaussien ([Fernique p. 48] ou [Neveu p. 93]) à $K'_N(s)$, $s \in [0, 1]^{d+1}$ défini, pour N fixé dans \mathbb{N}^* , par

$$K'_N(s) = K((s_1, \dots, s_d), N s_{d+1} + 1) .$$

Pour cela on majore $\sup_{\substack{|s_i - s'_i| \leq h \\ 1 \leq i \leq d+1}} E(K'(s) - K'(s'))^2$ par le produit d'une constante

et d'une puissance strictement positive de h .

Soit s et s' dans $[0, 1]^d$, u et u' dans $[0, 1]$ tels que

$$\max_{1 \leq i \leq d} |s_i - s'_i| \leq h \text{ et } |u - u'| \leq h$$

$$E(K'(s, u) - K'(s', u'))^2 = (Nu+1) \Gamma(s, s) + (Nu'+1) \Gamma(s', s') - 2 (N \min(u, u') + 1) \times \Gamma(s, s')$$

$$E(K'(s, u) - K'(s', u'))^2 \leq (N+1) (|\Gamma_1(s, s')| + h \times \max(\Gamma(s, s), \Gamma(s', s')))$$

avec $\Gamma_1(s, s') = \Gamma(s, s) + \Gamma(s', s') - 2\Gamma(s, s')$

ou encore

$$\Gamma_1(s, s') = E(g_1(s) - g_1(s'))^2 + 2 \sum_{n=2}^{+\infty} E[(g_1(s) - g_1(s')) (g_n(s) - g_n(s'))]$$

$$E(g_1(s) - g_1(s'))^2 = F(s) - [F(s)]^2 + F(s') - [F(s')]^2 - 2(F(s \wedge s') - F(s) F(s'))$$

Donc

$$E(g_1(s) - g_1(s'))^2 \leq F(s) - F(s \wedge s') + F(s') - F(s \wedge s') \leq 2dAh .$$

D'autre part, il résulte d'une inégalité de Davydov [Deo, 1973] que pour tout $r_1 \in]1, 4 + d(1+\varepsilon)[$, si l'on pose $r_2 = 2 r_1 / (r_1 - 1)$

on a :

$$(1) \quad |E(g_1(s) g_n(s'))| \leq 20 [\rho(n)]^{1/r_1}$$

$$\text{et} \quad |E(g_1(s) - g_1(s')) (g_n(s) - g_n(s'))| \leq 10 [\rho(n)]^{1/r_1}$$

$$\|g_1(s) - g_1(s')\|_{r_2} \|g_n(s) - g_n(s')\|_{r_2}$$

Les valeurs prises par $[g_i(s) - g_i(s')]$ sont 0 ; $1 - [F(s) - F(s')]$; $-1 - [F(s) - F(s')]$ avec les probabilités respectives

$$1 - F(s) - F(s') + 2F(s \wedge s') ; F(s) - F(s \wedge s') ; F(s') - F(s \wedge s')$$

$$\text{donc} \quad E(|g_i(s) - g_i(s')|^{r_2}) \leq 2^{r_2+1} dAh .$$

Par conséquent

$$|\Gamma_1(s, s')| \leq 2dAh + 20 \sum_{n=2}^{+\infty} [\rho(n)]^{1/r_1} \times (2^{r_2+1} dAh)^{2/r_2}$$

$$|\Gamma_1(s, s')| \leq C_1 h^{1-1/r_1} , \quad \text{pour } h < 1 .$$

L'inégalité (1) permet de majorer $\Gamma(s, s)$ par une constante C_2 et d'obtenir

$$E(K'(s, u) - K'(s', u'))^2 \leq (N+1) C_3 h^{1-1/r_1} \quad \text{pour } h < 1$$

ce qui permet de conclure.

Démonstration du théorème 2 :

Le lemme 3 permet de poser presque sûrement, pour $g \in C^d$ et $t \in [1, +\infty[$

$$K(g,t) = \sum (-1)^P \int_{[0,1]^P} \frac{\partial^P g}{\partial x_{i_1} \dots \partial x_{i_p}} (***) K(***) , t) dx_{i_1} \dots dx_{i_p} .$$

Du théorème 1 il résulte qu'il existe une constante C_4 indépendante de T et M telle que presque sûrement, pour $T \in]1, +\infty[$

$$\sup_{1 \leq t \leq T} \sup_{s \in [0,1]^d} |R(s,t) - K(s,t)| \leq C_4 T^{1/2} (\text{Log } T)^{-\lambda}$$

L'expression de $R(g,t)$ donnée à la suite des lemmes 1 et 2 montre qu'il existe une constante C indépendante de T et M telle que presque sûrement, pour $T \in]1, +\infty[$

$$\sup_{1 \leq t \leq T} \sup_{g \in C_M^d} |R(g,t) - K(g,t)| \leq CM T^{1/2} (\text{Log } T)^{-\lambda} .$$

D'autre part si $g \in C_M^d$ on majore $|K(g,t)|$ par $2M\sqrt{t} + |R(g,t) - K(g,t)|$ ce qui prouve que la forme linéaire $g \mapsto K(g,t)$ est bornée sur tout C_M^d .

En écrivant que les intégrales de Riemann définissant $K(g,t)$ sont des limites de sommes finies on montre que $K(g,t)$ est bien un processus gaussien centré. Il reste à calculer sa fonction de covariance. Dans le calcul qui suit on utilisera plusieurs fois le théorème de Fubini appliqué à des intégrales de fonctions mesurables bornées par rapport à des mesures finies, ainsi que les lemmes 1 et 2.

En utilisant les conventions d'écriture précédentes on a :

$$E(K(g,t) K(g',u)) = E\left(\sum \sum (-1)^{P+P'} \int_{[0,1]^P} \int_{[0,1]^{P'}}$$

$$\frac{\partial^P g}{\partial x_{i_1} \dots \partial x_{i_p}} (***) K(***) , t) \frac{\partial^{P'} g'}{\partial x_{j_1} \dots \partial x_{j_{p'}}} (---) K(---) , t) dx_{i_1} \dots dx_{i_p} dx_{j_1} \dots dx_{j_{p'}} .$$

$$E(K(g,t) K(g',u)) = \min(t,u) \sum \sum (-1)^{p+p'} \int_{[0,1]^p} \int_{[0,1]^{p'}}$$

$$\frac{\partial^p g}{\partial x_{i_1} \dots \partial x_{i_p}} (***) \frac{\partial^{p'} g'}{\partial x_{j_1} \dots \partial x_{j_{p'}}} (---) \Gamma(***, ---) dx_{i_1} \dots dx_{i_p} dx_{j_1} \dots dx_{j_{p'}}$$

$$E(K(g,t) K(g',u)) = \min(t,u) \left(\sum \sum (-1)^{p+p'} \int_{[0,1]^p} \int_{[0,1]^{p'}} \right.$$

$$\frac{\partial^p g}{\partial x_{i_1} \dots \partial x_{i_p}} (***) \frac{\partial^{p'} g'}{\partial x_{j_1} \dots \partial x_{j_{p'}}} (---) g_1(***) g_1(---) dx_{i_1} \dots dx_{i_p} dx_{j_1} \dots dx_{j_{p'}} \left. \right)$$

$$+ \min(t,u) E \left(\sum \sum (-1)^{p+p'} \sum_{n=2}^{+\infty} \int_{[0,1]^p} \int_{[0,1]^{p'}} \frac{\partial^p g}{\partial x_{i_1} \dots \partial x_{i_p}} (***) \right.$$

$$\left. \frac{\partial^{p'} g'}{\partial x_{j_1} \dots \partial x_{j_{p'}}} (---) (g_1(***) g_n(---) + g_1(---) g_n(***)) dx_{i_1} \dots dx_{i_p} dx_{j_1} \dots dx_{j_{p'}} \right)$$

Or

$$\int_{[0,1]^p} \frac{\partial^p g}{\partial x_{i_1} \dots \partial x_{i_p}} (***) g_n(***) dx_{i_1} \dots dx_{i_p}$$

$$= \int_{[0,1]^d} g(d\delta_{X_n} - dF) = g(X_n) - \int g dF = G_n(g)$$

$$E(K(g,t) K(g',u)) = \min(t,u) \{E(G_1(g) G_1(g')) + \sum_{n=2}^{+\infty} E(G_1(g) G_n(g') + G_1(g') G_n(g))\}$$

$$E(K(g,t) K(g',u)) = \min(t,u) \Lambda(g, g')$$

IV - EXTENSION DU PROCESSUS K AUX LIMITES UNIFORMES D'ELEMENTS DE C_M^d .-

On notera L_d la réunion pour $M \in \mathbb{R}^+$ des fermetures $\overline{C_M^d}$ des ensembles C_M^d dans $C^0 = C([0,1]^d, \mathbb{R})$, ensemble des applications continues de $[0,1]^d$ dans \mathbb{R} , muni de la topologie de la convergence uniforme.

Théorème 3.-

Il existe un processus de Kiefer K défini sur (Ω, \mathcal{A}, P) et indexé par $L_d \times [1, +\infty[$ et une constante C' (indépendante de T et M) tels que pour toute fonction g de $\overline{C_M^d}$ on ait, presque sûrement, pour tout $T \in]1, +\infty[$

$$\sup_{1 \leq t \leq T} |R(g,t) - K(g,t)| \leq C' M T^{1/2} (\text{Log } T)^{-\lambda}$$

La fonction de covariance de $K(g,t)$ est donnée par

$$E(K(g,t) K(g',u)) = \min(t,u) \Lambda(g,g') .$$

Démonstration :

Commençons par majorer $|\Lambda(g,g')|$ pour g et g' dans C^0 . On utilise pour cela l'inégalité de Davydov déjà citée et l'inégalité

$$\|G_i(g)\|_\infty \leq 2 \|g\|_\infty$$

On en déduit l'existence d'une constante D_1 indépendante de g et g' telle que

$$|\Lambda(g,g')| \leq D_1 \|g\|_\infty \|g'\|_\infty .$$

Soit $\psi \in L_d$ et $t \in [1, +\infty[$. Il existe $M \in \mathbb{R}^+$ et une suite $(\psi_n)_{n \in \mathbb{N}}$ de C_M^d convergeant uniformément vers ψ . Pour $(m,n) \in \mathbb{N}^2$ la variable aléatoire $K(\psi_n - \psi_m, t)$ est gaussienne, centrée, de variance majorée par $D_1 t \|\psi_n - \psi_m\|_\infty^2$.

Par linéarité presque sûre de K sur C_M^d on en déduit que la suite $(K(\Psi_n, t))_{n \in \mathbb{N}}$ est de Cauchy dans $L^2(P)$ donc converge dans cet espace vers une variable aléatoire gaussienne (éventuellement dégénérée) centrée $K(\Psi, t)$.

Presque sûrement $K(\Psi, t)$ ne dépend pas de la suite $(\Psi_n)_{n \in \mathbb{N}}$ de C_M^d convergeant uniformément vers Ψ . Il suffit, pour s'en rendre compte de considérer, pour

tout autre suite $(\Psi'_n)_{n \in \mathbb{N}}$ de C_M^d tendant uniformément vers Ψ , la

suite mélangée $\Psi''_n = \Psi_n$ si n pair

$\Psi''_n = \Psi'_n$ si n impair.

La linéarité presque sûre pour tout t de $g \mapsto K(g, t)$ sur C^d entraîne que

$K(\Psi, t)$ est un processus gaussien centré sur $L_d \times [1, +\infty[$. Sa fonction de

covariance est donnée par :

$$\begin{aligned} E(K(\Psi_1, t) K(\Psi_2, u)) &= \lim_{n \rightarrow +\infty} \min(t, u) \Lambda(\Psi_{1,n}, \Psi_{2,n}) \\ &= \min(t, u) \Lambda(\Psi_1, \Psi_2) \end{aligned}$$

car la convergence uniforme de $(\Psi_n)_{n \in \mathbb{N}}$ vers Ψ entraîne la convergence de $(G_i(\Psi_n))_{n \in \mathbb{N}}$ vers $G_i(\Psi)$ dans $L_\infty(P)$.

Pour toute suite $(\Psi_n)_{n \in \mathbb{N}}$ de C_M^d convergeant uniformément vers Ψ et telle que $(K(\Psi_n, t))_{n \in \mathbb{N}}$ converge p.s. vers $K(\Psi, t)$ on a

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

$$|R(\Psi, t) - K(\Psi, t)| \leq |R(\Psi, t) - R(\Psi_n, t)| + |R(\Psi_n, t) - K(\Psi_n, t)| + |K(\Psi, t) - K(\Psi_n, t)|$$

$$\forall t \in]1, +\infty[$$

$$\forall n \in \mathbb{N} \quad \sup_{1 \leq t \leq T} |R(\Psi, t) - K(\Psi, t)| \leq 2T \|\Psi - \Psi_n\|_\infty$$

$$+ CM T^{1/2} (\text{Log } T)^{-\lambda}$$

$$+ \sup_{1 \leq t \leq T} |K(\Psi - \Psi_n, t)| \quad \text{p.s.}$$

car presque sûrement, il existe une sous-suite $(\Psi_{m_k})_{k \in \mathbb{N}}$ de $(\Psi_n)_{n \in \mathbb{N}}$ telle que

$$K(\Psi - \Psi_n, t) = \lim_{k \rightarrow +\infty} K(\Psi_{m_k} - \Psi_n, t) = K(\Psi, t) - K(\Psi_n, t)$$

si $\Lambda(\Psi - \Psi_n, \Psi - \Psi_n) = 0$ $K(\Psi - \Psi_n, t) = 0$ p.s.

sinon $[\Lambda(\Psi - \Psi_n, \Psi - \Psi_n)]^{-1/2} K(\Psi - \Psi_n, s, T)$ est un processus de Wiener sur $[\frac{1}{T}, 1]$ et

$$\begin{aligned} \forall \epsilon > 0 \quad P(\sup_{1 \leq t \leq T} [\Lambda(\Psi - \Psi_n, \Psi - \Psi_n)]^{-1/2} |K(\Psi - \Psi_n, t)| \geq \epsilon) \\ \leq 2 \sqrt{\frac{2}{\pi}} \int_{\epsilon}^{+\infty} e^{-x^2/2} dx \leq \sqrt{\frac{8}{\pi}} \frac{e^{-\epsilon^2/2}}{\epsilon} . \end{aligned}$$

Soit $\alpha > \frac{3}{2}$. On peut choisir $(\Psi_n)_{n \in \mathbb{N}}$ de façon à ce que $\|\Psi - \Psi_n\|_{\infty} \leq \frac{1}{(n+1)^\alpha}$.

On a alors

$$\begin{aligned} P(\sup_{1 \leq t \leq T} |K(\Psi - \Psi_T, t)| \geq \frac{\sqrt{D_1}}{T^{\alpha-1}}) \\ \leq P(\sup_{1 \leq t \leq T} |K(\Psi - \Psi_T, t)| \geq [\Lambda(\Psi - \Psi_T, \Psi - \Psi_T)]^{1/2} T^{1/2}) \leq \sqrt{\frac{8}{\pi}} \frac{e^{-T/2}}{\sqrt{T}} \end{aligned}$$

D'où l'on déduit par le lemme de Borel - Cantelli que

$$\sup_{1 \leq t \leq T} |K(\Psi - \Psi_T, t)| = O(T^{1-\alpha}) \quad \text{p.s.}$$

D'où la majoration presque sûre du théorème 3. Il est probable qu'une étude plus fine des processus $K(\Psi - \Psi_n, t)$ permettra de donner une majoration presque sûre de

$$\sup_{1 \leq t \leq T} \sup_{\Psi \in C_M^d} |R(\Psi, t) - K(\Psi, t)| .$$

V - APPROXIMATION FORTE DANS DES ESPACES A NOYAU REPRODUISANT.

On utilisera sans les rappeler des résultats de [1].

Soit H un espace de Hilbert séparable de fonctions réelles définies sur $[0,1]^d$, dont le produit scalaire est noté (\cdot, \cdot) , la norme $\|\cdot\|$ et la tribu borélienne $B(H)$.

On suppose que H admet un noyau reproduisant mesurable $H : ([0,1]^{2d}, B([0,1]^{2d})) \rightarrow (\mathbb{R}, B_{\mathbb{R}})$ tel que la fonction $x \mapsto H(x,x)$ soit dans $L^1(\pi)$. (Toute fonction de H est alors de carré intégrable par rapport à π).

Pour $t \geq 1$ l'application $R_t : (\Omega, A, P) \rightarrow (H, B(H))$

$$\omega \mapsto \sum_{k=1}^t (H(\cdot, X_k(\omega)) - \int H(\cdot, x) d\pi(x))$$

(où $\int H(\cdot, x) d\pi(x)$ est l'intégrale de Bochner par rapport à π de l'application $\begin{matrix} [0,1]^d & \rightarrow & H \\ x & \mapsto & H(\cdot, x) \end{matrix}$) est une variable aléatoire à valeurs dans H et

$$\forall g \in H \quad \forall t \in [1, +\infty[\quad R(g,t) = (R_t, g)$$

On notera H_0 le sous-espace dense de H engendré par les fonctions $(H(\cdot, x))_{x \in [0,1]^d}$ et B_0 sa boule unité. On suppose que $H_0 \subset C^d$ et que les éléments de B_0 ont une norme uniforme majorée par M_1 .

Définition 2 : Processus de Kiefer hilbertien.

Un processus gaussien centré $(K_t)_{t \in D}$, $D \subset \mathbb{R}^+$, à valeurs dans un espace de Hilbert séparable H est appelé un processus de Kiefer hilbertien si le processus (g, K_t) , $(g,t) \in H \times D$ vérifie les conditions de la définition 1.

On suppose que, presque sûrement, il existe un réel C'' (indépendant de T) tel que, pour tout $T \in]1, +\infty[$, la condition suivante soit satisfaite pour le processus de Kiefer K dont l'existence est assurée par le théorème 2 :

$$(I) \quad \sup_{1 \leq t \leq T} \sup_{g \in B_0} |R(g,t) - K(g,t)| \leq C'' T^{1/2} (\text{Log } T)^{-\lambda}.$$

La condition (I) est réalisée en particulier s'il existe $M \in \mathbb{R}^+$ tel que $B_0 \subset C_M^d$ (on peut alors prendre $C'' = CM$ et $M_1 = M$).

Théorème 4.-

Sous la condition (I), il existe un processus de Kiefer hilbertien $(K_t)_{t \in [1, +\infty[}$ défini sur (Ω, \mathcal{A}, P) à valeurs dans H tel que, presque sûrement, pour tout $T \in]1, +\infty[$,

$$\sup_{t \in [1, T]} \|R_t - K_t\| \leq C'' T^{1/2} (\text{Log } T)^{-\lambda}.$$

Le processus $(K_t)_{t \in [1, +\infty[}$ est centré et vérifie :

$$\forall (g, g') \in H^2 \quad E((g, K_t)(g', K_u)) = \min(t, u) \Lambda(g, g').$$

Démonstration :

La démonstration que nous avons donnée dans [1] pour le cas i.i.d. s'adapte aisément :

On a presque sûrement $\forall T \in]1, +\infty[\quad \forall t \in [1, T]$

$$\begin{aligned} \sup_{g \in B_0} |K(g,t)| &\leq \sup_{g \in B_0} (|R(g,t)| + |R(g,t) - K(g,t)|) \\ &\leq 2 M_1 T + C'' T^{1/2} (\text{Log } T)^{-\lambda} \end{aligned}$$

où K est le processus de Kiefer dont l'existence est assurée par le théorème 2.

Pour ω appartenant à un ensemble Ω_0 de P-probabilité 1

l'application $H_0 \rightarrow \mathbb{R}$ est donc une forme linéaire continue sur H_0 ,
 $g \mapsto K^\omega(g, t)$

que l'on peut prolonger sur H grâce au théorème de Hahn - Banach et qui est représentée dans H par un élément $\ell_t(\omega)$ de norme majorée par

$$2 M_1 T + C'' T^{1/2} (\text{Log } T)^{-\lambda} .$$

On pose alors $K_t(\omega) = \ell_t(\omega)$ si $\omega \in \Omega_0$
 $= 0$ sinon

K_t est bien une v.a. à valeurs dans H puisque

$$\forall t \in [1, +\infty[\quad (K_t, H(\cdot, x)) = \int_{\Omega_0} K(H(\cdot, x), t)$$

K_t est Bochner - intégrable, puisque bornée en norme. Pour tout élément g de H , $K(g, t)$ et (g, K_t) sont presque sûrement égales, ce qui montre que $(K_t)_{t \in [1, +\infty[}$ est un processus de Kiefer à valeurs dans H et que

$$E((g, K_t) (g', K_u)) = \min(t, u) \Lambda(g, g') .$$

L'inégalité du théorème s'obtient aisément puisque :

$$\|R_t - K_t\| = \sup_{g \in B_0} |(R_t - K_t, g)|$$

Le théorème 4 est lui aussi susceptible d'améliorations. Il permettrait des applications intéressantes s'il pouvait être démontré sous la simple hypothèse $B_0 \subset C^d$.

VI - AUTRE MODE D'EXTENSION DU PROCESSUS DE KIEFER.

$R(s,t)$ est la fonction de répartition d'une mesure à signe sur $[0,1]^d$, il est donc facile d'exprimer $\int \psi dR$, où ψ est l'indicatrice d'un rectangle $\prod_{i=1}^d [a_i, b_i[$ de $[0,1]^d$, à l'aide des valeurs de R aux sommets de ce rectangle :

$$R(\psi, t) = \sum_{\substack{\epsilon_i \in \{0,1\} \\ 1 \leq i \leq d}} (-1)^{\sum_{i=1}^d \epsilon_i} R(\epsilon_1 a_1 + (1-\epsilon_1) b_1, \dots, \epsilon_d a_d + (1-\epsilon_d) b_d) .$$

Cette définition s'étend aisément par linéarité aux fonctions en escalier sur les pavés.

Une méthode naturelle d'extension de K consiste à définir $K(\psi, t)$ d'une manière analogue et de procéder ensuite par approximations. Cette voie a été suivie par Ibero dans le cas i.i.d. [4]. Nous allons voir que le résultat de Philipp et Pinzur ne permet pas d'obtenir un résultat satisfaisant de cette façon.

Nous nous limiterons ici, pour simplifier, au cas $d = 2$.

Le passage à d quelconque se fait sans difficulté à l'aide des lemmes donnés dans [6].

Notations : On se fixe $M \in \mathbb{R}^{+*}$ et on pose, pour $r \in \mathbb{N}$

$$I_{ij}^r = \{(x,y) \in [0,1]^2 \mid i \leq 2^r x < i+1 \text{ et } j \leq 2^r y < j+1\} \quad i \in \overline{0, 2^r-1} \quad j \in \overline{0, 2^r-1}$$

$$\psi_{ij}^r = 1_{I_{ij}^r} .$$

Pour toute fonction g sur $[0,1]^d$

$$g_r = \sum_{(i,j) \in 0, 2^{r-1}} \frac{1}{2} M 2^{-r} [2^r M^{-1} g(i 2^{-r}, j 2^{-r})] \psi_{ij}^r$$

où le crochet surligné en pointillés désigne la partie entière.

g_r est une fonction définie sur $[0,1]^2$ et en escalier sur les I_{ij}^r .

Si g est une fonction sur $[0,1]^2$ admettant des dérivées partielles en chaque point $(2^{-r} i, 2^{-r} j)$ on pose

$$g_r^* = \sum_{(i,j) \in 0, 2^{r-1}} \frac{1}{2} \{g(i 2^{-r}, j 2^{-r}) + g'_x(i 2^{-r}, j 2^{-r})(x - i 2^{-r}) + g'_y(i 2^{-r}, j 2^{-r})(y - j 2^{-r})\} \psi_{ij}^r$$

g_r^* est une fonction définie sur $[0,1]^2$ et affine sur chaque I_{ij}^r .

L'extension de K se fera en trois étapes : on définira d'abord $K(g,t)$ pour $g \in G_{M,r}$: ensemble des fonctions bornées par M qui sont constantes sur chaque I_{ij}^r , puis pour $g \in P_{M,r}$: ensemble des fonctions dont la restriction à chaque I_{ij}^r est affine et qui est bornée, ainsi que ses dérivées partielles par M . On définira enfin $K(g,t)$ pour $g \in F_M$: ensemble des fonctions continues sur $[0,1]^2$, bornées par M et admettant des dérivées partielles d'ordre 1 bornées par M et lipschitziennes de rapport M . On notera

$$F^* = \bigcup_{M>0} F_M, \quad P^* = \bigcup_{\substack{M>0 \\ r \in \mathbb{N}}} P_{M,r}, \quad G^* = \bigcup_{\substack{M>0 \\ r \in \mathbb{N}}} G_{M,r}$$

Proposition 1.-

Il existe sur (Ω, \mathcal{A}, P) un processus de Kiefer K indexé par $G^* \times [1, +\infty[$ tel que l'application $g \mapsto K(g, t)$ soit pour tout t une forme linéaire sur G^* bornée sur tout $G_{M,r}$ et une constante C (indépendante de T et M) tels qu'on ait presque sûrement pour tout $T \in]1, +\infty[$

$$\sup_{1 \leq t \leq T} \sup_{g \in G_{M,r}} |R(g, t) - K(g, t)| \leq 4 C M 2^{2r} T^{1/2} (\text{Log } T)^{-\lambda} .$$

La fonction de covariance de K est donnée par

$$E(K(g, t) K(g', u)) = \min(t, u) \Lambda(g, g') .$$

Démonstration :

$$\begin{aligned} \text{On pose } K(\psi_{ij}^r, t) &= K(((i+1) 2^{-r}, (j+1) 2^{-r}), t) + K((i 2^{-r}, j 2^{-r}), t) \\ &\quad - K(((i+1) 2^{-r}, j 2^{-r}), t) - K((i 2^{-r}, (j+1) 2^{-r}), t) \end{aligned}$$

et on l'étend par linéarité à $G_{M,r}$. On vérifie que le processus $K(g, t)$ est bien défini pour $g \in G^*$ (si une fonction admet deux représentations différentes, les valeurs de $K(g, t)$ calculées à partir de celles-ci sont égales) et est linéaire par rapport à g .

$K(g, t)$ est un processus gaussien centré sur $G^* \times [1, +\infty[$ et :

$$\begin{aligned} \sup_{1 \leq t \leq T} \sup_{g \in G_{M,r}} |R(g, t) - K(g, t)| &= \\ & \sup_{1 \leq t \leq T} \sup_{g \in G_{M,r}} \left| \sum_{(i,j) \in 0, 2^r-1} \frac{1}{2} \lambda_{ij} (R(\psi_{ij}^r, t) - K(\psi_{ij}^r, t)) \right| \\ & \leq 2^{2r} M C T^{1/2} (\text{Log } T)^{-\lambda} \end{aligned}$$

$|R(g,t)|$ étant bornée par $2 \hat{M}t$ sur $G_{M,r}$, $|K(g,t)|$ y est également borné.

La fonction de covariance est donnée par un calcul très simple :

$$\begin{aligned} E(K(\sum_{(i,j) \in 0, 2^r-1} \lambda_{ij} \varphi_{ij}^r, t) K(\sum_{(k,\ell) \in 0, 2^r-1} \mu_{k\ell} \varphi_{k\ell}^r, u)) \\ = \sum_{(i,j) \in 0, 2^r-1} \sum_{(k,\ell) \in 0, 2^r-1} \lambda_{ij} \mu_{k\ell} E(K(\varphi_{ij}^r, t) K(\varphi_{k\ell}^r, u)) \end{aligned}$$

Or $E(K(\varphi_{ij}^r, t) K(\varphi_{k\ell}^r, u)) = \min(t,u) \Lambda(\varphi_{ij}^r, \varphi_{k\ell}^r)$ ce qui permet de conclure.

On peut alors procéder, avec des modifications adéquates, comme pour les lemmes 7 et 8 de [4] : on approxime toute fonction g de F par une fonction g_r^* , et cette dernière fonction par un élément g_{r+s} de $G_{M,r+s}$. On montre alors simultanément que l'on peut étendre K de G^* à P^* puis de P^* à F^* et que l'on peut majorer convenablement les valeurs de $K(.,t)$ pour les fonctions $(g - g_r^*)$ et $(g_r^* - g_{r+s})$. Les résultats que l'on obtient ainsi à partir de la proposition 1 sont très insuffisants ; c'est le facteur 2^{2r} qui est responsable de cette faiblesse : si l'on veut obtenir une approximation intéressante sur G_M^* (un $O(T^{1/2} (\log T)^{-\nu})$, avec $\nu > 0$) on doit choisir la taille des cellules I_{ij}^r de façon à ce que $2^{2r} = O((\log T)^\mu)$ avec $\mu < \lambda$. Il semble bien que cette condition rende impossible une majoration adéquate des oscillations du processus empirique sur les ensembles de fonctions permettant les approximations successives. Ce qui fait que la méthode employée dans ce paragraphe ne pourrait donner de bons résultats que si l'on disposait d'un théorème analogue au théorème 1 avec une vitesse de l'ordre de $T^{1/2-\nu}$, $\nu > 0$.

R E F E R E N C E S.

- [1] BERLINET, A. - Variables aléatoires à valeurs dans les espaces autoreproduisants et mesure empirique. C.R.A.S., série A, 1980, p. 973-975.
- [2] DEO, CHANDRAKANT, M. - A note on empirical processes of strong mixing sequences. Ann. of Prob. 1, 1973, p. 870-875.
- [3] FERNIQUE, X.M. - Régularité des trajectoires des fonctions aléatoires gaussiennes. Lecture Notes in Math., Springer-Verlag, 1975.
- [4] IBERO, M. - Approximation forte du processus empirique fonctionnel multidimensionnel, Bull. Sc. Math., 2^{ème} série, 103, 1979, p. 409-422.
- [5] IBERO, M. - Approximation d'intégrales stochastiques empiriques multidimensionnelles, Bull. Sc. Math., 2^{ème} série, 104, 1980, p. 1-28.
- [6] IBERO, M. - Intégrales stochastiques, processus empiriques et tests multidimensionnels. Thèse de Doctorat d'Etat. Université de Paris VI, 1980.
- [7] NEVEU, J. - Bases mathématiques du Calcul des Probabilités, Masson, Paris, 1970.
- [8] PHILIPP, W. et PINZUR, L. - Almost sure approximation theorems for the multivariate empirical process, Z.f.W., 54, 1980, p. 1-13.
- [9] REVESZ, P. - On strong approximation of the multidimensional empirical process, Ann. of Prob., 4, 1976, p. 729-743.

QUATRIEME PARTIE.

CHAPITRE 5

ESTIMATING THE DEGREES OF AN ARMA MODEL⁽¹⁾

by

Alain BERLINET

September 1982

AMS 1980 - Subject classification 62 M 10

KEY WORDS : Univariate and Multivariate Time Series

ARMA models

Epsilon - algorithm.

(1) to appear in Compstat Lectures (1984)

SUMMARY

A. Introduction.

B. The scalar case.

B.1. - Notations and assumptions.

B.2. - Fundamental theorem 1 : Characterization of a minimal ARMA (p,q) representation.

B.3. - Statement of the problem.

B.4. - Shanks transformation.

B.4.1.- Remark.

B.4.2.- Definitions.

B.4.3.- Fundamental property.

B.5. - The ϵ -algorithm.

B.5.1.- Theorem 2 : Equivalence between the ϵ -algorithm and the Shanks transformation .

B.5.2.- Theorem 3.

B.5.3.- Estimation of p and q.

B.5.4.- Examples.

B.5.5.- Some remarks on the computation.

B.5.6.- Particular cases.

B.5.7.- Computation and estimation of the partial auto-correlation function.

B.5.8.- Other properties of the ϵ -algorithm applied to the auto-correlation function ρ of a minimal ARMA (p,q) process.

B.5.9.- Convergence theorems and statistical tests on the ϵ -array.

B.6. - The R-S algorithm.

B.6.1. - Definition of the R-S algorithm.

B.6.2. - Properties of the R and S arrays.

B.7. - The corner method.

C. The vector case.

C.1. - Notations and assumptions.

C.1.1. - Multivariate ARMA models.

C.1.2. - Remarks.

C.1.3. - Existence, invertibility and identifiability.

C.2. - Fundamental theorem 16 : characterization of a canonical ARMA (p,q) representation.

C.3. - The matrix ϵ -algorithm.

C.4. - The vector ϵ -algorithm.

References.

ESTIMATING THE DEGREES OF AN ARMA MODEL

by

Alain BERLINET, LILLE ¹⁾.

A. Introduction.

This paper is a review of available methods to estimate the unknown degrees of an ARMA process $X = (X_t)_{t \in \mathbb{Z}}$ directly from the empirical auto-correlation function after observation of X_0, X_1, \dots, X_T . The scalar and the vector cases are considered. The described methods need no preliminary estimation of the process coefficients, unlike different criteria based on residual variance (e.g. Akaike's criteria [Priestley]) which compel us to vary the degrees and to estimate at each time all the coefficients. We shall not deal with these criteria here.

In the scalar case, the methods are determinantal ones : the computed quantities can be expressed by means of determinants and are based on the fundamental theorem (B.2.). We begin by the epsilon-algorithm because it is known and extensively studied since 1956 so that theorems about this algorithm will give us straightforward proofs for properties of the two other methods : the R-S algorithm [Gray, Kelley and Mac Intire] and the corner method [Béguin, Gouriéroux and Monfort]. Of course, these methods can be used simultaneously when a doubt subsists. We shall see the relationships between them and how to extend the algorithms to the vector case. We re-state some definitions and properties about ARMA processes. For generalities about these models and proofs of their properties the reader is referred to Hannan [1970], Anderson [1971] and Priestley [1981].

B. The scalar case.

B.1. - Notations and assumptions.

Let $X = (X_t)_{t \in \mathbb{Z}}$ be a real stationary (wide sense) process ; we suppose without loss of generality that

1) A. Berlinet, U.E.R. de Mathématiques, Université de Lille I,
F - 59655 - VILLENEUVE D'ASCQ CEDEX

$$EX_t = 0, \quad \forall t \in \mathbb{Z}.$$

Let $\gamma = (\gamma_h)_{h \in \mathbb{Z}}$ and $\rho = (\rho_h)_{h \in \mathbb{Z}}$ be respectively the auto-covariance and auto-correlation function of X :

$$\forall (t, h) \in \mathbb{Z}^2 \quad E(X_t X_{t+h}) = \gamma_h \quad (\gamma_0 \neq 0)$$

$$\rho_h = \gamma_h \gamma_0^{-1}.$$

We recall that X is an AutoRegressive Moving Average process of degrees p and q (ARMA (p, q)) if X satisfies the following stochastic difference equation :

$$(C) \quad \left\{ \begin{array}{l} (1) \quad \forall t \in \mathbb{Z} \quad \sum_{i=0}^p \phi_i X_{t-p+i} = \sum_{i=0}^q \theta_i U_{t-q+i} \\ \text{where} \quad \phi_0 \neq 0, \quad \phi_p = 1 \\ \quad \quad \theta_0 \neq 0, \quad \theta_q = 1 \\ U = (U_t)_{t \in \mathbb{Z}} \text{ is a white noise } (EU_t = 0, \quad EU_t U_s = \sigma^2 \delta_{ts}, \quad \sigma > 0) \end{array} \right.$$

(1) can be written as $\phi(B)(X) = \theta(B)(U)$ with

$$\phi(z) = \sum_{i=0}^p \phi_i z^{p-i}$$

$$\theta(z) = \sum_{i=0}^q \theta_i z^{q-i}$$

B is the backward shift operator :

$$B((x_t)_{t \in \mathbb{Z}}) = (x_{t-1})_{t \in \mathbb{Z}}.$$

It is well known that a sufficient condition for the existence and unicity of a second order stationary (wide sense) process satisfying (C) is that the roots of ϕ are of modulus strictly greater than 1. We assume that the roots of θ verify also the preceding condition so that U is the innovation process of X . Then, we have the following expansion for X_t :

$$(2) \quad \forall t \in \mathbb{Z}, \quad X_t = \sum_{j=0}^{\infty} c_j \sum_{\ell=0}^q \theta_{\ell} U_{t-q+\ell-j} = \sum_{i=0}^{\infty} \psi_i U_{t-i}$$

where the c_j 's are the coefficients of the Taylor series expansion of $[\phi(z)]^{-1}$ ($c_0 = 1$).

The representation (1) will be called minimal when p and q are as small as possible i.e. when ϕ and θ have no common root.

Let us remark that under the above assumptions an equation such like (1) can be simplified when ϕ and θ have in common roots of modulus strictly greater than 1 (see annex 1).

B.2. - Fundamental theorem 1 : Characterization of the degrees of a minimal ARMA representation [Béguin et al].

The second order stationary process X admits an ARMA (p, q) minimal representation if and only if the sequence ρ satisfies a difference equation of minimal order p from the minimal rank ($q-p+1$) :

$$\forall n \geq q-p+1 \quad \sum_{i=0}^p \phi_i \rho_{n+i} = 0, \quad \phi_0 \neq 0, \quad \phi_p = 1$$

$$\sum_{i=0}^p \phi_i \rho_{q-p+i} \neq 0.$$

Outline of the proof : this theorem has been first proved in pure cases

$q = 0$ Autoregressive process of degree p : AR(p) (Yule-Walker equations).

$p = 0$ Moving Average of degree q : MA(q) [Ansley, Spivey and Wroblewski].

By calculating $E(X_{t-p-n} \sum_{i=0}^p \phi_i X_{t-p+i})$ necessity is obvious from condition (C)

and representation (2). Sufficiency is derived from the case $p = 0$ applied to

$$\sum_{i=0}^p \phi_i X_{t-p+i} \cdot \blacksquare$$

B.3. - Statement of the problem.

If we dispose of an algorithm which, applied to a sequence satisfying a difference equation, gives us the minimal rank and the minimal order of this equation, we are able to recognize a stationary minimal ARMA (p, q) and to calculate p and q from the sequence ρ , or from a sequence deduced from ρ and satisfying a difference equation of the same minimal order from the same minimal rank.

However in practice we only observe X_0, \dots, X_T and we do not know ρ . A natural strategy is to apply our algorithm to an estimate of ρ , computed from X_0, \dots, X_T and if X is an ARMA (p,q) to estimate, from the results, the degrees p and q . As an estimate of ρ we have used in the following $\hat{\rho}$, the empirical autocorrelation function. This is not a restriction, other estimates can be used (and sometimes give better results) for instance smoothed autocorrelation derived via spectral analysis.

Three algorithms have been used to solve this statistical problem :

- the R and S arrays [Gray et al, 1978]
- the corner method [Béguin et al, 1980]
- the ϵ -algorithm [Berlinet, 1981, 1982 (a)].

Let us now recall some definitions and properties well known in acceleration of convergence. The reader is referred to Brezinski [1977] for the theoretical aspects and proofs of the theorems, and to Brezinski [1978] for the practical ones (FORTRAN programs).

B.4. - Shanks transformation [1955].

For any sequence $s = (s_n)_{n \in \mathbb{Z}}$ of real numbers we note

- * Δs the sequence $(s_{n+1} - s_n)_{n \in \mathbb{Z}}$.
- * $\Delta^k s$ the sequence $\Delta(\Delta^{k-1} s)$, $k \geq 2$.
- * $H_k^n(s)$ the Hankel determinant of s of order k from the rank n :

$$H_k^n(s) = \begin{vmatrix} s_n & s_{n+1} & s_{n+2} & \dots & s_{n+k-1} \\ s_{n+1} & s_{n+2} & \dots & \dots & s_{n+k} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ s_{n+k-1} & \dots & \dots & \dots & s_{n+2k-2} \end{vmatrix} \quad \begin{matrix} k \in \mathbb{N}^* \\ n \in \mathbb{Z} \end{matrix}$$

$$H_0^n(s) = 1, \quad \forall n \in \mathbb{Z}$$

* $S_k^n(\mathbb{Z})$ will be the subset of $\mathbb{R}^{\mathbb{Z}}$ of sequences s such that $H_k^n(\Delta^2 s) \neq 0$.

* $S_k(\mathbb{Z}) = \bigcap_{n \in \mathbb{Z}} S_k^n(\mathbb{Z})$.

B.4.1. Remark.- From a theorem of Widder [1946, p. 129-136] if a sequence s is such that $\forall m \in \mathbb{N} \quad (\Delta^2 s)_{n+m} = \int_{-\infty}^{+\infty} t^m d\alpha(t)$ where α is non-decreasing with infinitely many points of increase then we have $H_k^n(\Delta^2 s) > 0, \forall k \in \mathbb{N}^*$ and s belongs to $S_k^n(\mathbb{Z}), \forall k \in \mathbb{Z}$.

B.4.2. Definitions.-

For $k \in \mathbb{N}, n \in \mathbb{Z}$ let $e_k^n : S_k^n(\mathbb{Z}) \longrightarrow \mathbb{R}$
 $s \longmapsto e_k^n(s) = \left[H_{k+1}^n(s) \right] \left[H_k^n(\Delta^2 s) \right]^{-1}$

We call Shanks transformation of order k the application

$e_k : S_k(\mathbb{Z}) \longrightarrow \mathbb{R}^{\mathbb{Z}}$
 $s \longmapsto (e_k^n(s))_{n \in \mathbb{Z}}$

Shanks transformation has the required property with respect to sequences satisfying difference equations :

B.4.3. Fundamental property [Shanks, 1955].-

Let $S \in \mathbb{R}, k \in \mathbb{N}, N \in \mathbb{Z}$ and $s \in \bigcap_{n > N} S_k^n(\mathbb{Z})$.

We have $e_k^n(s) = S, \forall n > N$, if and only if there exists a sequence $(a_i)_{0 \leq i \leq k}$ of real numbers such that

$$\left\{ \begin{array}{l} \sum_{i=0}^k a_i \neq 0 \\ \text{and} \quad \sum_{i=0}^k a_i (S_{n+i} - S) = 0, \quad \forall n > N. \end{array} \right.$$

Using the definition B.4.2., this transformation needs the computation of Hankel determinants which satisfy the following relationship :

(3) $H_k^{n+1} \times H_k^{n-1} - (H_k^n)^2 = H_{k+1}^{n-1} \times H_{k-1}^{n+1}$.

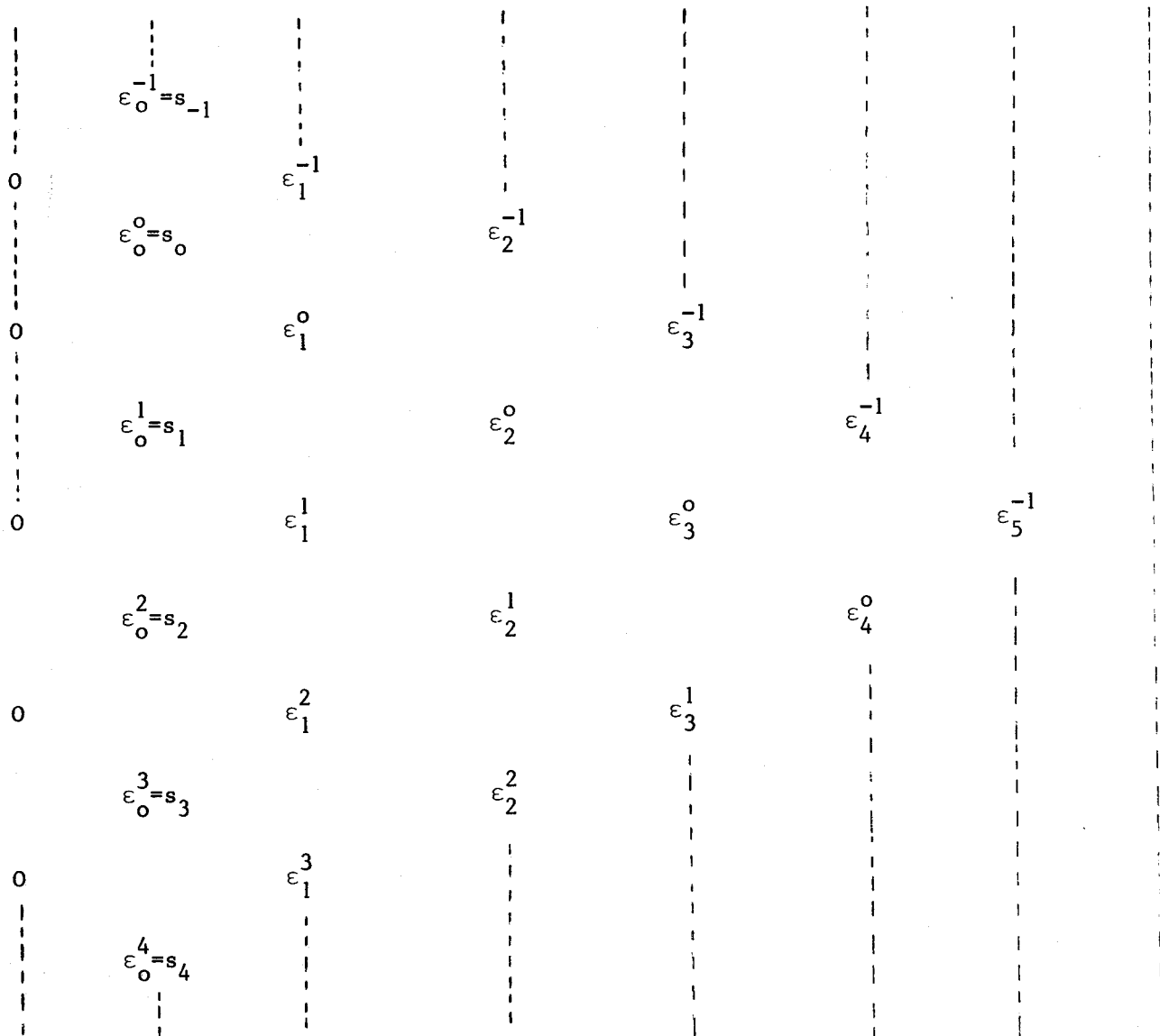
Fortunately an algorithm with a very simple rule permits us to do this transformation without any computation of determinant.

B.5. The ϵ -algorithm.

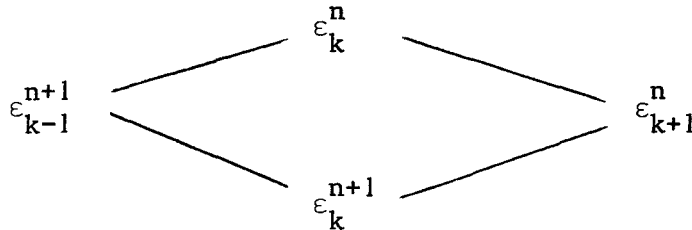
The ϵ -algorithm is a recursive lozenge algorithm due to Wynn [1956] defined as follows : let $s = (s_n)_{n \in \mathbb{Z}}$ be an element of $\mathbb{R}^{\mathbb{Z}}$

$$(D) \quad \begin{cases} \forall n \in \mathbb{Z}, & \epsilon_{-1}^0(s) = 0 & \epsilon_0^n(s) = s_n \\ \forall k \in \mathbb{Z}, \forall n \in \mathbb{Z}, & \epsilon_{k+1}^n(s) = \epsilon_{k-1}^{n+1}(s) + \left[\epsilon_k^{n+1}(s) - \epsilon_k^n(s) \right]^{-1} \end{cases}$$

We shall suppose that for all sequences considered below the quantities defined by (D) will be computable. Usually the ϵ -array is drawn as follows :



The second line of (D) links four numbers situated on the vertices of a lozenge :



B.5.1. Theorem 2 .- Equivalence between the ϵ -algorithm and the Shanks transformation [Wynn, 1956].

Let $k \in \mathbb{N}$, $n \in \mathbb{Z}$

$$\begin{aligned} \epsilon_{2k}^n(s) &= e_k^n(s) = \left[H_{k+1}^n(s) \right] \left[H_k^n(\Delta^2 s) \right]^{-1} \\ \epsilon_{2k+1}^n(s) &= \left[e_k^n(\Delta s) \right]^{-1} = \left[H_k^n(\Delta^3 s) \right] \left[H_{k+1}^n(\Delta s) \right]^{-1} \end{aligned}$$

In the ϵ -array, only columns corresponding to k even will be of interest for us. It is possible to compute only these columns, using the cross-rule of Wynn [1966] :

The following numbers

$$\begin{aligned} N &= \epsilon_k^{n-1} \\ W &= \epsilon_{k-2}^{n+1} & C &= \epsilon_k^n & E &= \epsilon_{k+2}^{n-1} \\ S &= \epsilon_k^{n+1} \end{aligned}$$

are linked by the relationship (using obvious geographical notations) :

$$(C-E)^{-1} + (C-W)^{-1} = (C-S)^{-1} + (C-N)^{-1}.$$

From B.2., B.4.3. and B.5.1., we obtain :

B.5.2. Theorem 3 .- The second order stationary process X admits a minimal ARMA (p,q) representation if and only if

$$\begin{cases} \epsilon_{2p}^n(\rho) = 0 & \forall n \geq q-p+1 \\ \text{and } \epsilon_{2p}^{q-p} \neq 0. \end{cases}$$

A minimal ARMA (p,q) process is therefore characterized by the following even columns in the ε -array built from its auto-correlation function ρ :

column \ line	0	2	4	...	2p
q-p	$\varepsilon_0^{q-p} = \rho_{q-p}$				
q-p+1	$\varepsilon_0^{q-p+1} = \rho_{q-p+1}$	ε_2^{q-p}			
q-p+2	$\varepsilon_0^{q-p+2} = \rho_{q-p+2}$	ε_2^{q-p+1}	ε_4^{q-p}		
...
q	$\varepsilon_0^q = \rho_q$				$\varepsilon_{2p}^{q-p} \neq 0$
q+1	$\varepsilon_0^{q+1} = \rho_{q+1}$				0
...	...				0
...	...				0
...	...				0
...	...				0
...

We can obtain a complete array by setting

$$\varepsilon_{2(p+k)}^n = 0 \quad \forall k \in \mathbb{N} \quad \forall n \geq (q-p+1-k).$$

B.5.3. - Estimation of p and q.

We suppose that we have observed realizations X_0, X_1, \dots, X_T of a minimal ARMA (p,q) process X the degrees, coefficients and mean of which are unknown.

We compute the empirical auto-correlation function $\hat{\rho} = (\hat{\rho}_h)_{\substack{h \in \mathbb{Z} \\ |h| \leq T}}$

$$\hat{\rho}_h = \left[\sum_{t=0}^{T-|h|} (x_t - \bar{x}_T)(x_{t+|h|} - \bar{x}_T) \right] \left[\sum_{t=0}^T (x_t - \bar{x}_T)^2 \right]^{-1}$$

$$\bar{x}_T = (T+1)^{-1} \sum_{t=0}^T x_t$$

and we apply the ϵ -algorithm to $\hat{\rho}$.

To estimate p and q we have to detect a sudden variation in the ϵ -array. We are not compelled to build the whole array from $(\hat{\rho}_h)_{|h| \leq T}$: we have to compute at least ϵ_{2p}^{q-p} and ϵ_{2p}^{q-p+1} ; for that we have to use at least $\hat{\rho}_h$ for $q-p \leq h \leq q+3p+1$ (T is supposed greater than $q+3p+1$. We shall give below convergence theorems for $T \rightarrow +\infty$). A good sequential way to proceed is to compute the SW-NE diagonal from $\epsilon_o^h = \hat{\rho}_h$ and the NW-SE diagonal from $\epsilon_o^{-h} = \hat{\rho}_{-h}$ after the calculation of $\hat{\rho}_h$.

B.5.4. - Examples.

Most of the simulations have shown that the degrees of the process were well recognized, or that a few number of possibilities (with the true values among them) could be selected. Of course the method is not miraculous: when ρ is badly approximated we are unable to conclude anything or we give bad estimates of p and q . To prove the easiness of the method we give verifiable numerical examples: using a pocket computer Casio FX 702 P we have applied the rule (D) to the theoretical and empirical functions of processes, observations of which are given by Priestley [1981, Appendix]. Because of the arithmetic of this computer we have kept six digits after the decimal point for ρ and three for $\hat{\rho}$.

Example 1.- ARMA (2,2), $p = q = 2$.

$$\forall t \in \mathbb{Z} \quad 0,5x_{t-2} + 1,4x_{t-1} + x_t = -0,1 u_{t-2} - 0,2 u_{t-1} + u_t$$

$$u_t \sim N(0,1).$$

ϵ -array built from the theoretical auto-correlation function.

column line	0	2	2p=4	6	8	10
-2	0,801300					
-1	-0,935611	-0,02				
0	1	3,2E-2	6,7E-3			
1	-0,935611	-0,02	-1,6E-3	-6,6E-4		
q=2	0,801300	9,5E-3	-4,6E-4	3,8E-4	-3,4E-5	
3	-0,654005	-5,7E-3	2,6E-6	9,5E-7	2,4E-7	8,7E-7
4	0,514967	3,6E-3	9,4E-7	-2,5E-7	5,8E-7	6,5E-7
5	-0,393948	-2,4E-3	2,5E-7	5,7E-7	6,4E-7	5,8E-7
6	0,294042	1,6E-3	8,6E-7	6,5E-7	5,8E-7	-7,3E-7
7	-0,214682	-1,1E-3	5,3E-7	9,7E-7	2,6E-7	
8	0,153538	7,5E-4	-8,1E-7	-7,3E-7		
9	-0,107615	-5,3E-4	-7,3E-7			
10	0,073887	3,9E-4				
11	-0,049632					

ϵ -array built from the empirical auto-correlation function.

column \ line	0	2	$2\hat{p}=4$	6	8	10
-2	0,778					
-1	-0,929	-2,3E-2				
0	1	3,6E-2	6,9E-3			
1	-0,929	-2,3E-2	-1,4E-3	-1,0E-3		
$\hat{q}=2$	0,778	1,3E-2	-1,0E-3	-1,6E-3	1,0E-4	
3	-0,608	-1,1E-2	2,8E-5	3,3E-4	-5,5E-4	-3,4E-4
4	0,442	9,3E-3	2,6E-4	7,6E-5	-3,2E-4	-3,8E-4
5	-0,294	-7,9E-3	-7,8E-4	-7,3E-4	-4,4E-4	-3,4E-4
6	0,174	4,8E-3	-7,3E-4	-7,9E-4	-1,2E-3	-2,7E-4
7	-0,091	-5,6E-3	-2,6E-4	-1,1E-3	-7,4E-4	
8	0,035	7,0E-3	1,1E-3	2,1E-4		
9	-0,001	-1,1E-2	-3,7E-3			
10	-0,009	-5,6E-3				
11	-0,003					

Example 2.- AR(2) $p = 2$ $q = 0$

$$\forall t \in \mathbb{Z} \quad 0,7X_{t-2} - 0,4X_{t-1} + X_t = U_t$$

$$U_t \sim N(0,1).$$

ϵ -array built from the theoretical auto-correlation function.

column \ line	0	2	2p=4	6	8	10
-4	0,261309					
-3	-0,407043	-0,69				
-2	-0,605868	-0,45	1,5E-5			
-1	0,235294	8,65	2,4E-6	1,5E-5		
q=0	1	0,62	0,20	0,12	0,08	
1	0,235294	8,65	2,4E-6	1,5E-5	2,9E-6	1,2E-5
2	-0,605868	-0,45	1,5E-5	8,6E-6	8,1E-6	6,8E-6
3	-0,407043	-0,69	2,9E-6	8,1E-6	8,9E-6	8,5E-6
4	0,261309	0,42	1,2E-5	6,8E-6	8,5E-6	8,9E-6
5	0,389459	0,29	-4,0E-7	-3,9E-5	2,7E-6	
6	-0,027123	-0,69	-9,9E-6	-3,3E-6		
7	-0,283471	-0,17	1,1E-5			
8	-0,094426	-0,82				
9	0,160669					

ϵ -array built from the empirical auto-correlation function.

column \ line	0	2	$2\hat{p}=4$	6	8	10
-4	0,306					
-3	-0,359	-0,79				
-2	-0,622	-0,42	-8,9E-4			
-1	0,208	17,51	2,7E-3	-9,7E-4		
$\hat{q}=0$	1	0,60	0,19	0,11	0,08	
1	0,208	17,51	2,7E-3	-9,7E-4	-1,8E-2	-1,8E-2
2	-0,622	-0,42	-8,9E-4	4,0E-3	-1,8E-2	-1,8E-2
3	-0,359	-0,79	-1,4E-2	-1,7E-2	-6,4E-3	-1,2E-2
4	0,306	0,33	-1,6E-2	-1,5E-2	-1,2E-2	-1,2E-2
5	0,327	0,31	-2,3E-3	-1,1E-2	-1,2E-2	
6	-0,138	-0,32	-2,3E-2	-1,4E-2		
7	-0,269	-0,18	-6,6E-3			
8	-0,015	0,55				
9	0,160					

Example 3 : MA(2) p = 0 q = 2

$$\forall t \in \mathbb{Z} \quad X_t = 0,2U_{t-2} + 1,1U_{t-1} + U_t$$

$$U_t \sim N(0,1).$$

ϵ -array built from the theoretical autocorrelation function.

column \ line	2p=0	2	4	6
-3	0			
-2	0,088889	-1,9E-2		
-1	0,586667	3,02	0,30	
0	1	0,79	0,54	0,46
1	0,586667	3,02	0,30	
q=2	0,088889	-1,9E-2		
3	0			
4	0			
5	0			
6	0			

ε -array built from the empirical auto-correlation function.

column \ line	$2\hat{p}=0$	2	4	6	8
-3	-0,036				
-2	0,039	-0,049			
-1	0,558	3,54	0,27		
0	1	0,78	0,52	0,43	
$\hat{q}=1$	0,558	3,54	0,27	2,22	0,07
2	0,039	-0,049	-0,046	-0,17	-0,18
3	-0,036	-0,046	-0,049	-0,18	
4	-0,045	-0,035	0,20		
5	-0,112	-0,21			
6	-0,151				

This third example is a typical case of mistake on the estimation of q :

$$\varepsilon_{2p}^{q-p} \times \left[\varepsilon_{2p}^{q-p-1} \right]^{-1} = \rho_2 \times \rho_1^{-1} \neq 0,15.$$

p is well identified but $\hat{q} = q-1$.

A more precise study is needed ; however the identification as a MA(1) of this MA(2) process is not too bad because the coefficient of U_{t-2} is small compared to those of U_{t-1} and U_t .

B.5.5. - A few remarks on the computation.

a) Using the usual arithmetic of the computer numerical results are sometimes better when the ε -algorithm is applied not to $\hat{\rho}$ but to a sequence deduced from $\hat{\rho}$ by RO-transformation :

Definition.- An application from $\mathbb{R}^{\mathbb{Z}}$ to $\mathbb{R}^{\mathbb{Z}}$ is called a RO-transformation if and only if it transforms any sequence satisfying an homogeneous difference equation with constant coefficients of minimal Order N from the minimal Rank R in a sequence having the same property.

For instance $T_1 : s \rightarrow T_1(s) = \alpha s, \quad \alpha \in \mathbb{R}^*$

$T_2 : s \rightarrow T_2(s)$

$\forall n \in \mathbb{Z} (T_2(s))_n = \alpha^n s_n, \quad \alpha \in \mathbb{R}^*$

are RO-transformations.

We have specially used T_2 with $\alpha = -1$: numerical results are better when the autocorrelation changes of sign at each lag ; when we have successive groups of autocorrelations of the same sign, we apply the algorithm to $((-1)^{n\hat{\rho}_n})_{n \in \mathbb{Z}}$. Rounding errors can be important when calculating the inverse of a difference of two numbers close to one another. When the autocorrelations are rapidly small, we can use $((-2)^{n\hat{\rho}_n})_{n \in \mathbb{Z}}$ or $(\alpha \times (-1)^{n\hat{\rho}_n})_{n \in \mathbb{Z}}, \quad \alpha \in \mathbb{R}^*$.

Examples :

Example 1 : Sunspot data.

Many authors have fitted the series of Wolfer's sunspot numbers which can be found for instance in Anderson [1971] for the years 1749 to 1924.

The entire series or subseries are often fitted by AR(2) models. With the ϵ -algorithm applied to $((-1)^{h\hat{\rho}_h})$ we obtain $\hat{p} = 2, \quad \hat{q} = 0$ for the entire series.

We give below the ϵ -arrays built from $(\hat{\rho}_h)_{-1 \leq h \leq 10}$ and $((-1)^{h\hat{\rho}_h})_{-1 \leq h \leq 10}$ for the subseries studied for instance by Box and Jenkins [1976] and Gray et al [1978] for years 1770 to 1869 ; these two arrays illustrate very well our considerations about RO-transformations. Gray et al had already used T_2 with $\alpha = -1$ in their paper.

In the ϵ -array built from $(\hat{\rho}_h)_{-10 \leq h \leq 10}$ no sudden variation can be detected ; we observe only four changes of sign in column 0. The array built from $((-1)^{h\hat{\rho}_h})_{-10 \leq h \leq 10}$ gives us clearly $\hat{p} = 2$ and $\hat{q} = 0$; we observe sixteen changes of sign in column 0.

Example 2 : Given by Gray et al [1978, p. 9].

ARMA (2,1) p = 2 q = 1

$$0,68X_{t-2} - 1,32X_{t-1} + X_t = -0,8 U_{t-1} + U_t$$

$$U_t \sim N(0,1)$$

ε -array built from $(\hat{\rho}_h)_{-2 \leq h \leq 12}$

column \ line	0	2	4	6	8	10
-2	0,064					
-1	0,582	2,7				
0	1	0,79	0,53			
1	0,582	2,7	-1,1	-0,46		
2	0,064	-1,9	-0,29	-5,0	-0,073	
3	-0,347	-0,82	1,7	0,12	0,044	-0,011
4	-0,566	-0,52	-0,19	0,038	0,39	-0,061
5	-0,514	-0,58	0,98	-0,037	-0,070	0,050
6	-0,288	-1,1	-0,32	-0,060	-0,043	-0,023
7	0,025	0,61	0,23	0,043	-0,010	-0,013
8	0,229	0,35	-0,98	-0,22	-0,014	
9	0,303	0,27	-0,17	-4,5		
10	0,240	0,34	0,13			
11	0,071	-0,95				
12	-0,074					

ϵ -array built from $((-1)^{h\hat{p}})_{-2 \leq h \leq 12}$

column \ line	0	2	$2\hat{p}=4$	6	8	10
-2	0,064					
-1	-0,582	-0,12				
0	1	0,21	0,072			
$\hat{q}=1$	-0,582	-0,12	0,027	-0,26		
2	0,064	0,57	-3,2E-3	-6,7E-3	-3,9E-3	
3	0,347	0,13	-6,3E-3	-5,1E-3	-4,6E-3	-3,4E-3
4	-0,566	-0,071	-4,4E-3	-4,6E-3	-5,0E-3	-5,7E-3
5	0,514	0,054	-4,6E-3	-4,4E-3	-5,6E-3	-5,7E-3
6	-0,288	-0,090	-7,6E-3	-6,8E-3	-5,7E-3	-5,6E-3
7	-0,025	7,4	-6,4E-3	-8,2E-3	-0,018	-3,4E-3
8	0,229	0,057	-3,1E-3	-0,012	-6,8E-3	
9	-0,303	-0,034	1,8E-3	8,5E-3		
10	0,240	0,042	4,8E-3			
11	0,071	-0,074				
12	-0,074					

The estimation of p and q from the first array is not possible when from the second the choice $\hat{p} = 2, \hat{q} = 1$ is obvious.

b) When we have a good approximation of the autocorrelation function p and q can be found from a visual examination of the ϵ -array.

Numerical tests can be built easily for instance by computing $\epsilon_{2(k+1)}^n \times \left[\epsilon_{2k}^{n+1} \right]^{-1}$

and $\varepsilon_{2k}^{n+1} \left[\varepsilon_{2k}^n \right]^{-1}$ and by comparing these numbers to 10^{-1} or 10^{-2} or to a smaller number if ρ has been well approximated. We shall deal with statistical tests below.

Sometimes several possibilities can be kept ; then we have to use an other criterion (like Akaike's) to choose between them.

c) To avoid rounding errors it is better to compute the ε -array in the set \mathbb{Q} of rational numbers. It is interesting to vary the number of digits for ρ and $\hat{\rho}$.

d) Singular rules.

When ε_k^{n+1} is close to ε_k^n , the number ε_{k+1}^n is great and subject to important rounding errors which spread in a sub-triangle of the array. To avoid these errors particular rules have been defined by Wynn [1963]. Singular rules have been introduced for the case where two or more successive numbers of a column are strictly equal by Wynn [1963] and Cordellier [1973]. We do not detail these rules here ; the reader is referred to the authors mentioned above.

B.5.6.- Particular cases.

a) Moving average process : minimal ARMA (0,q) or MA(q).

The criterion B.5.2. reduces to

$$\begin{cases} \varepsilon_0^n(\rho) = 0 & \forall n \geq q+1 \\ \text{and } \varepsilon_0^q \neq 0 \end{cases}$$

As seen in the example 3 of B.5.4. this is the usual criterion because

$$\varepsilon_0^k(\rho) = \rho_k.$$

b) Autoregressive process : minimal ARMA (p,0) or AR(p).

The partial autocorrelation function $r = (r_h)_{h \geq 1}$ of the process is generally used to recognize an AR(p).

The functions r and ρ are linked by the following relationships :

$$\forall h \geq 1 \left\{ \begin{array}{l} \begin{pmatrix} \rho_1 \\ \rho_2 \\ \vdots \\ \rho_h \end{pmatrix} \\ r_h = -\phi_{0,h} \end{array} \right. = \begin{pmatrix} 1 & \rho_1 & \rho_2 & \dots & \rho_{h-1} \\ \rho_1 & 1 & \rho_1 & \dots & \rho_{h-2} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \rho_{h-1} & \dots & \dots & \dots & 1 \end{pmatrix} \begin{pmatrix} -\phi_{h-1,h} \\ \vdots \\ \vdots \\ -\phi_{1,h} \\ -\phi_{0,h} \end{pmatrix} \quad \text{(Yule-Walker equations)}$$

so that

$$(3) \quad \forall h \geq 1 \quad r_h = (-1)^{h-1} H_h^{2-h}(\rho) \times [H_h^{1-h}(\rho)]^{-1} .$$

For an AR(p), we have

$$r_p \neq 0 \text{ and } r_h = 0 \text{ if } h > p.$$

The criterion B.5.2. reduces to

$$\left\{ \begin{array}{l} \varepsilon_{2p}^n(\rho) = 0 \quad \forall n \geq -p+1 \\ \text{and } \varepsilon_{2p}^{-p} \neq 0 . \end{array} \right.$$

The numerator of $\varepsilon_{2p}^{-p+1}(\rho)$ and the one of $(-1)^p r_{p+1}$ are both equal to $H_{p+1}^{-p+1} : r_{p+1}$ and $\varepsilon_{2p}^{-p+1}(\rho)$ vanish jointly but the vanishings of ε_{2p}^n , $n \geq -p+1$ are equivalent to those of H_{p+1} , $n \geq -p+1$, which are determinants of the same order when the vanishings of r_h , $h \geq p+1$ are equivalent to those of $H_h^{2-h}(\rho)$, $h \geq p+1$, which are determinants of increasing order.

More precisely we have :

$$\forall h \geq 1, r_h = [\varepsilon_{2(h-1)}^{2-h}(\rho)] \times [\varepsilon_{2(h-1)}^{1-h}(\rho)]^{-1} \times [H_{h-1}^{2-h}(\Delta^2 \rho)] \times [H_{h-1}^{1-h}(\Delta^2 \rho)]^{-1} \times (-1)^{h-1}.$$

However this relationship is not useful to compute r : it is more complicated than (3).

B.5.7. - Computation and estimation of the partial autocorrelation function.

Durbin has given a recursive method to calculate $(\phi_{j,h})_{0 \leq j \leq h-1}$, $h \geq 1$:

$$\phi_{h+1-j,h+1} = \phi_{h-j,h} + \phi_{0,h+1} \phi_{j-1,h}, \quad 1 \leq j \leq h$$

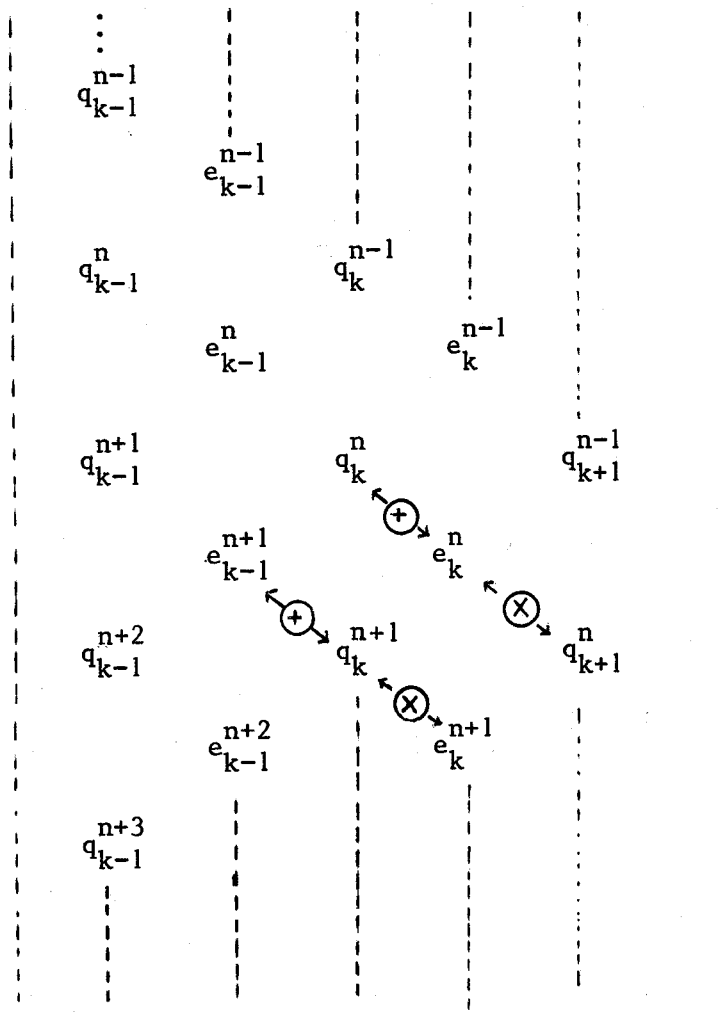
$$\phi_{0,h+1} = -(\rho_{h+1} + \sum_{j=1}^h \phi_{h-j,h} \rho_{h+1-j}) (1 + \sum_{j=1}^h \phi_{h-j,h} \rho_j)^{-1}.$$

When ρ is estimated by $\hat{\rho}$, the Yule-Walker equations give estimates $(\hat{\phi}_{j,h})_{0 \leq j \leq h-1}$ of the parameters of an AR(h) fitting the data. Durbin's algorithm give estimates of $(r_h)_{h \geq 1}$ which are very sensitive to rounding errors and should not be used if the process is "close to nonstationarity", i.e. if ϕ has a root of modulus close to one. The partial autocorrelation function may be evaluated by calculating the least squares estimate of the autoregressive parameter $\phi_{0,h}$ in successive AR(h), $h \geq 1$ [Box and Jenkins, 1976 : program 3 (USES) page 500].

An other algorithm is very efficient to compute r from γ : the quotient-difference algorithm of Rutishauser. This algorithm is closely related to the ε -algorithm through orthogonal polynomials and continued fractions [Brezinski, 1977] and is defined as follows :

<p>q-d algorithm applied to $\gamma = (\gamma_h)_{h \in \mathbb{Z}}$</p>	{	<p>Let $q_1^n = \gamma_{n+1} \gamma_n^{-1}$; $e_0^n = 0$, $n \in \mathbb{Z}$</p> <p>$q_{k+1}^n e_k^n = q_k^{n+1} e_k^{n+1}$ $k \in \mathbb{N}^*$</p> <p>$q_k^{n+1} e_k^n = q_k^{n+1} e_{k-1}^{n+1}$ $n \in \mathbb{Z}$</p>
---	---	--

The quantities e_k^n and q_k^n can be disposed in the following manner :



Theorem 4 .- If the q-d algorithm is applied to the autocovariance function γ of the second order stationary process X we have

$$\forall h \geq 2 \quad r_h = (-1)^{h-1} \prod_{i=1}^h q_i^{1-h} .$$

Proof : The result is obvious from the relationship

$$\prod_{i=1}^k q_i^{n+1} = [H_k^{n+2}(\gamma)] [H_k^{n+1}(\gamma)]^{-1} \quad [\text{Brezinski, 1977, page 261}] .$$

Thus replacing γ by $\hat{\gamma}$, we obtain Yule-Walker estimates $(\hat{r}_h)_{h \geq 1}$ of $(r_h)_{h \geq 1}$. It is easy to see from (3) that r_h is a differentiable function of $(\rho_1, \rho_2, \dots, \rho_h)$ so that, using the assumptions and arguments given below for the ϵ -algorithm, it is possible to get the same convergence theorems (almost sure convergence of \hat{r}_h to r_h and joint asymptotic normality of $\sqrt{T}(\hat{r}_h - r_h)$, $1 \leq h \leq s$).

B.5.8. - Other properties of the ϵ -algorithm applied to the autocorrelation function ρ of a minimal ARMA (p,q) process.

Theorem 5 (Bauer - Wynn invariant).-

If $p \geq 1$, then for $n \geq q-p+1$ we have

$$I(n) = \sum_{i=0}^{2p-2} (-1)^i \epsilon_i^n(\rho) \epsilon_{i+1}^n(\rho) = -p + \phi'(1) [\phi(1)]^{-1}.$$

Proof : The existence of this invariant has been proved by Bauer [1957].

Wynn [1966] has shown that for a sequence $s = (s_n)_{n \in \mathbb{Z}}$ satisfying the minimal difference equation $\sum_{i=0}^p a_i s_{n+i} = 0$, $\forall n > N$ with $\sum_{i=0}^p a_i \neq 0$ we have

$$I(n) = -R'(1) \times [R(1)]^{-1}$$

with

$$R(z) = \sum_{i=0}^p a_i z^i.$$

Here we have $R(z) = \sum_{i=0}^p \phi_i z^i = z^p \phi(\frac{1}{z})$

and $R'(1) \times [R(1)]^{-1} = p - \phi'(1) \times [\phi(1)]^{-1}$. ■

Example : The following arrays give $I(n,k,\rho) = \sum_{i=0}^{2k-2} (-1)^i \epsilon_i^n(\rho) \epsilon_{i+1}^n(\rho)$ and $I(n,k,\hat{\rho})$ for the ARMA (2,2) of B.5.4 (example 1) $\phi(z) = 0,5z^2 + 1,4z + 1$ and the true value of $I(n,2,\rho)$ for $n \geq 1$ is

$$-2 + (2,4/2,9) = -34/29 \# -1,17241.$$

I(n,k,ρ)

$n \backslash k$	1	p=2	3	4	5
-2	-0,46	-0,87	-1,94	-3,25	-4,16
-1	-0,48	-1,13	-2,62	-3,17	-2,76
q-p=0	-0,52	-1,22	-2,17	-3,53	6,28
1	-0,54	-1,1726	-2,73	-2,83	-4,13
2	-0,55	-1,1722	-2,53	6,88	-3,58
3	-0,56	-1,17244	-0,77	-1,78	
4	-0,57	-1,1723	-3,78	-2,14	
5	-0,57	-1,1729	-1,57		
6	-0,58	-1,1724	-10,78		
7	-0,58	-1,1713			
8	-0,59	-1,1727			
9	-0,59				
10	-0,60				

$$I(n, k, \hat{\rho})$$

n \ k	k						
	1	$\hat{p}=2$	3	4	5	6	7
-2	-0,46	-0,88	-1,95	-2,98	-3,08	2,57	-3,34
-1	-0,48	-1,12	-4,67	-2,92	-4,70	-8,04	
$\hat{q}-\hat{p}=0$	-0,52	-1,19	-2,09	-2,30	1,17	-6,97	
1	-0,54	-1,11	-0,99	-1,49	-2,40		
2	-0,56	-1,11	-1,40	-2,97	-2,68		
3	-0,58	-1,14	-17,0	-0,10			
4	-0,60	-1,25	-2,67	-2,20			
5	-0,63	-1,12	-1,36				
6	-0,66	-1,17	-1,58				
7	-0,72	-1,35					
8	-0,97	-1,85					
9	0,125						
10	-1,5						

The mean of $(I(n, 2, \hat{\rho}))_{1 \leq n \leq 6}$ is -1,1795.

This is a good estimation of the true invariant.

Theorem 6. - (Approximation of the roots of ϕ) [Wynn, 1962].

If $p \geq 1$ and if the roots $(\lambda_i)_{1 \leq i \leq p}$ of ϕ are real distincts and such that $|\lambda_1| < |\lambda_2| < \dots < |\lambda_p|$ then $\lim_{n \rightarrow +\infty} \{ [\varepsilon_{2i-1}^n(\rho)]^{-1} \varepsilon_{2i-1}^{n+1}(\rho) \} = \lambda_i$, $1 \leq i \leq p$ and $\lim_{n \rightarrow +\infty} \{ [\varepsilon_{2i-2}^n(\rho)]^{-1} \varepsilon_{2i-2}^{n+1}(\rho) \} = \lambda_i^{-1}$, $1 \leq i \leq p$.

This theorem appears as a generalization of the well-known property that for an ARMA (1,q) with $\phi(z) = \phi_0 z + 1$, we have $\forall n \geq q$, $\rho_{n+1} = -\phi_0 \rho_n$ so that $\forall n \geq q$, $[\varepsilon_0^n(\rho)]^{-1} \times \varepsilon_0^{n+1}(\rho) = -\phi_0 = \lambda_1^{-1}$
 $[\varepsilon_1^n(\rho)]^{-1} \times \varepsilon_1^{n+1}(\rho) = (\rho_{n+1} + \rho_{n+1} \phi_0^{-1}) \times (-\phi_0 \rho_{n+1} - \rho_{n+1}) = -\phi_0^{-1} = \lambda_1$.

It can be used to estimate the degree d of differentiation in a minimal ARIMA (p,q,d) process by counting the roots very close to one. This can be done only when ρ is very well approximated.

B.5.9. - Convergence theorems and statistical tests on the ε -array.

To give convergence theorems we have to make the following assumption :

$$(E) \quad \begin{cases} \forall t \in \mathbb{Z} , & E(U_t | M_{t-1}) = 0 \\ & E(U_t^2 | M_{t-1}) = \sigma^2 \end{cases}$$

where M_t , $t \in \mathbb{Z}$ is the σ -field generated by $(X_s)_{s \leq t}$. (Of course condition (E) is fulfilled when the $(U_t)_{t \in \mathbb{Z}}$ are i.i.d.).

Theorem 7. - Under assumption (E) we have

$$\forall k \in \mathbb{N} , \quad \forall n \in \mathbb{Z} \quad \varepsilon_k^n(\hat{\rho}) \xrightarrow[T \rightarrow +\infty]{} \varepsilon_k^n(\rho) \quad \text{a.s.}$$

Proof : X is a stationary ARMA process so that, under assumption (E) we have

[Hannan and Heyde, 1972, theorem 1]

$$\forall h \in \mathbb{Z} , \quad \hat{\rho}_h \xrightarrow[T \rightarrow +\infty]{} \rho_h \quad \text{a.s.}$$

It is easy to see from its determinantal expression (B.5.1.) that $\varepsilon_k^n(s)$ is a differentiable function of $(s_n, s_{n+1}, \dots, s_{n+2k})$.

The conclusion follows. ■

Corollary.- We have obviously almost sure convergence of $I(n, k, \hat{\rho})$ to $I(n, k, \rho)$ and a similar result for the approximation of the roots of ϕ (under hypothesis of theorem 6).

$$\lim_{n \rightarrow +\infty} \lim_{T \rightarrow +\infty} \{ [\varepsilon_{2i-1}^n(\hat{\rho})]^{-1} \varepsilon_{2i-1}^{n+1}(\hat{\rho}) \} = \lambda_i \quad \text{a.s.} \quad 1 \leq i \leq p$$

$$\lim_{n \rightarrow +\infty} \lim_{T \rightarrow +\infty} \{ [\varepsilon_{2i-2}^n(\hat{\rho})]^{-1} \varepsilon_{2i-2}^{n+1}(\hat{\rho}) \} = \lambda_i^{-1} \quad \text{a.s.} \quad 1 \leq i \leq p.$$

We can obtain strong consistency of \hat{p} and \hat{q} , estimates of p and q . Suppose that K is an integer strictly greater than $3p+q$. Let $\varepsilon(K)$ be the ε -array (even columns) built from $(\rho_h)_{-K \leq h \leq K}$. As $K > 3p+q$ the numbers ε_{2p}^{q-p} and ε_{2p}^{q-p+1} are in $\varepsilon(K)$. We suppose that

$$m = \min |\varepsilon_{2k}^n| > 0 \quad \text{where the min is taken over the numbers } \varepsilon_{2k}^n \text{ of } \varepsilon(K)$$

such that $\begin{cases} k < p \\ \text{or} \\ k = p \text{ and } n \leq q-p. \end{cases}$

Suppose that α is a known number such that

$$0 < \alpha < m.$$

The above conditions are usually verified. For each T we built $\varepsilon(K, T)$, the ε -array (even columns) from $(\hat{\rho}_h)_{-K \leq h \leq K}$.

We define \hat{p} as follows :

$$\left\{ \begin{array}{l} \hat{p} = \min_{\substack{k \\ \varepsilon_{2k}^{K-2k} \in \varepsilon(K, T)}} \{ k \mid |\varepsilon_{2k}^{K-2k}(\hat{\rho})| < \alpha \} \quad \text{if this set is non empty} \\ \hat{p} = 0 \quad \text{otherwise.} \end{array} \right.$$

The set over which the min is taken corresponds to the S.W.-N.E. diagonal from $\varepsilon_o^K(\hat{\rho})$ in $\varepsilon(K, T)$.

\hat{q} is then defined by :

$$\left\{ \begin{array}{l} \hat{q} = \min_{\hat{p} \leq k \leq K - \hat{p} - 1} \{k \mid |\varepsilon_{2\hat{p}}^{k-\hat{p}+1}| < \alpha \text{ and } |\varepsilon_{2\hat{p}}^{k-\hat{p}}| > \alpha\} \text{ if this set is non empty} \\ \\ \hat{q} = 0 \text{ otherwise.} \end{array} \right.$$

The set over which the min is taken corresponds to the column number $2\hat{p}$ in $\varepsilon(K, T)$.

Theorem 8 .- (\hat{p}, \hat{q}) defined above converges almost surely to (p, q) i.e. for T great enough $\hat{p} = p$ a.s. and $\hat{q} = q$ a.s.

Proof : As $\varepsilon_{2k}^n(\hat{\rho}) \rightarrow \varepsilon_{2k}^n(\rho)$ a.s., $0 \leq k \leq p$ we have for T great enough

$$\left. \begin{array}{l} |\varepsilon_{2k}^n(\hat{\rho})| > \alpha \text{ a.s.,} \\ |\varepsilon_{2k}^n(\hat{\rho})| < \alpha \text{ a.s.,} \end{array} \right\} \begin{array}{l} k < p \\ \text{or } k = p \text{ and } n \leq q-p \\ k = p \text{ and } n \geq q-p+1. \end{array}$$

Thus $\hat{p} = p$ a.s. and therefore $\hat{q} = q$ a.s. ■

Theorem 9 .- Under the null hypothesis H_0 :

for $n \geq q-p+1$ $\varepsilon_{2p}^n(\rho) = 0$, the vector

$$V = T^{1/2} (\varepsilon_{2p}^{n_1}(\hat{\rho}), \varepsilon_{2p}^{n_2}(\hat{\rho}), \dots, \varepsilon_{2p}^{n_k}(\hat{\rho})) \text{ where } k \in \mathbb{N}^* \text{ and } n_i \geq q-p+1, 1 \leq i \leq k,$$

converges in distribution to a gaussian random vector of mean 0 and of covariance matrix $J \ddagger J'$

J is the jacobian matrix of $(\varepsilon_{2p}^{n_1}, \dots, \varepsilon_{2p}^{n_k})$ at $(\rho_1, \rho_2, \dots, \rho_{n_k+2p})$;

J' is the transposed matrix of J ;

$$\ddagger = \left(\sum_{r=-\infty}^{+\infty} \psi_{ij}(r) \right)_{\substack{1 \leq i \leq n_k + 2p \\ 1 \leq j \leq n_k + 2p}}$$

$$\psi_{ij}(r) = \rho_r \rho_{r+i-j} + \rho_r \rho_{r+i+j} + 2\rho_r^2 \rho_i \rho_j - 2\rho_r \rho_i \rho_{r+j} - 2\rho_r \rho_j \rho_{r+i} .$$

Proof : Let us first remark that $\sum_{i=1}^{\infty} i \psi_i^2 < +\infty$.

(the ψ_i 's have been defined in (2) of B.1.) :

$[\phi(z)]^{-1}$ is analytic in the disc $D(0, 1+\beta) = \{z \mid |z| < 1+\beta\}$, $\beta > 0$ small enough ; thus the series $\sum_{i=0}^{\infty} \psi_i z^i$, $\sum_{i=0}^{\infty} \psi_i^2 z^i$ and $\sum_{i=0}^{\infty} i \psi_i^2 z^i$ have radius of convergence strictly greater than one.

Therefore the conditions of Hannan and Heyde [1972, theorem 3] are fulfilled and we can conclude that the joint distribution of $T^{1/2}(\hat{\rho}_i - \rho_i)$, $1 \leq i \leq n_k + 2p$ converges to the normal distribution with zero mean vector and covariance matrix \ddagger . As already noticed, $\varepsilon_{2k}^n(\rho)$ is a differentiable function of $(\rho_n, \rho_{n+1}, \dots, \rho_{n+2k})$. The conclusion follows from the theorem 2.9.2. of Mardia, Kent and Bibby. ■

Corollary.- If rank $(J \ddagger J')$ = k the distribution of the real random variable $V(J \ddagger J')^{-1} V'$ converges to $\chi^2(k)$.

The matrix J is computable from minors of Hankel determinants. We don't know if it can be deduced more rapidly from the ε -array. In practice J can be approximated by means of a numerical derivation method. We shall have to estimate $J \ddagger J'$ from $\hat{\rho}$ to test the hypothesis H_0 from the ε -array. Beguin et al have proposed similar tests for the corner method.

B.6. The R-S algorithm.

As the ε -algorithm has been defined by Wynn to avoid the evaluation of determinants in the Schanks transformation, the R-S algorithm has been introduced by Pye and Atchison to compute the G-transformation which has been introduced essentially to evaluate improper integrals of the first order. We shall

not deal with that subject here. This transformation involves the computation of the quantities $[\mathbb{H}_{k-1}^n(\Delta s)]^{-1} \mathbb{H}_k^n(s)$ and $[\mathbb{H}_k^n(s)]^{-1} \mathbb{H}_k^n(\Delta s)$.

The computation of these quantities is the keystone of the paper of Gray et al. It is possible without any evaluation of determinant owing to the following algorithm :

B.6.1.- Definition of the R-S algorithm.

Let $s \in \mathbb{R}^Z$

$$\forall n \in \mathbb{Z} \quad S_0^n(s) = 1 \quad R_1^n(s) = s_n$$

$$\forall k \in \mathbb{N} \quad \forall k \in \mathbb{Z}, \quad R_{k+1}^n(s) = R_k^{n+1}(s) ((S_k^n(s))^{-1} S_k^{n+1}(s) - 1)$$

$$S_k^n(s) = S_{k-1}^{n+1}(s) ((R_k^n(s))^{-1} R_k^{n+1}(s) - 1).$$

B.6.2.- Properties of the R and S arrays.

Theorem 10.-

$$\forall k \in \mathbb{Z} \quad \forall n \in \mathbb{Z} \quad R_k^n(s) = [\mathbb{H}_{k-1}^n(\Delta s)]^{-1} \mathbb{H}_k^n(s)$$

$$S_k^n(s) = [\mathbb{H}_k^n(s)]^{-1} \mathbb{H}_k^n(\Delta s)$$

this theorem has been proved first by Pye and Atchison. Using the theory of orthogonal polynomials, Brezinski [1979] gave an other proof of this result and many relationships between all the algorithms considered in the present paper. We continue to suppose that all the denominators considered are different from zero.

The paper written by Gray et al is much detailed and contains many examples. We are going to show that the three basic theorems of this paper can be easily deduced from our above results.

Theorem 11.-

Let $s \in \mathbb{R}^Z$ and $N \in \mathbb{N}$, $S_k^n(s)$ is a non null constant, as a function of n , for $n > N$ if and only if there exists a sequence $(a_i)_{0 \leq i \leq k}$ of real numbers such that

$$\left\{ \begin{array}{l} \sum_{i=0}^k a_i \neq 0 \\ \text{and } \sum_{i=0}^k a_i s_{n+i} = 0, \quad \forall n > N \end{array} \right.$$

Proof : To use the fundamental property B.5.2. we have only to remark that $e_k^n(s)$ and $R_{k+1}^n(s)$ have the same numerator $H_{k+1}^n(s)$.

We have $e_k^n(s) = 0, \quad \forall n > N$
if and only if $R_{k+1}^n(s) = 0, \quad \forall n > N$.

B.6.1. shows that this last condition is equivalent to $S_k^n(s)$ constant non null, as a function of n for $n > N$. ■

Theorem 12 .- A second order stationary process admits a minimal ARMA(p,q) representation if and only if its autocorrelation function ρ is such that

$$\left\{ \begin{array}{l} S_k^n(\rho) = C_1 \quad \forall n \geq q-p+1 \\ \text{and } S_k^{q-p}(\rho) \neq C_1 \end{array} \right.$$

(C_1 is a non null constant).

Proof : Straightforward from theorem 11 and the characterization of a minimal ARMA (p,q) representation.

Theorem 13 .- The necessary and sufficient condition of theorem 12 can be replaced by

$$\left\{ \begin{array}{l} S_k^n(\rho) = C_2 \quad \forall n \leq -q-p \\ \text{and } S_k^{-q-p+1}(\rho) \neq C_2 \end{array} \right.$$

($C_2 \neq 0$).

Proof : As the autocorrelation ρ is an even function the ε -array built from $(\rho_h)_{-K \leq h \leq K}$ is symmetric. The middle of the column $2k$ is occupied by ε_{2k}^{-k} and is a symmetry center for the column :

$$\varepsilon_{2k}^n = \varepsilon_{2k}^{-n-2k}, \quad \forall n \in \mathbb{Z}.$$

$$\text{The condition } \begin{cases} \varepsilon_{2p}^n(\rho) = 0 & \forall n \geq q-p \\ \varepsilon_{2p}^{q-p+1}(\rho) \neq 0 \end{cases}$$

is equivalent to the condition

$$\begin{cases} \varepsilon_{2p}^n(\rho) = 0 & , \quad \forall n < -q-p-1 \\ \varepsilon_{2p}^{-q-p}(\rho) \neq 0 \end{cases}$$

which is itself equivalent to

$$\text{and } \begin{cases} R_{p+1}^n(\rho) = 0 & \forall n < -q-p-1 \\ R_{p+1}^{-q-p}(\rho) \neq 0. \end{cases}$$

Using B.6.1. we conclude as in the proof of theorem 11 ■

The constants C_1 and C_2 can be calculated easily from the determinantal expression of $S_k^n(\rho)$ and are respectively equal to $(-1)^p \phi(1)$ and to $(-1)^p \phi_0^{-1} \phi(1)$.

The R-S algorithm is a little more complicated than the ε -algorithm and needs more operations. Usually the two methods give the same result (see for instance examples 1 and 2 of B.5.5., also given in Gray et al). When there is a doubt, a comparison between the ε , R and S arrays is useful.

B.7.- The corner method [Béguin et al].

This method is based on the computation of

$$\Delta(i,j) = \begin{vmatrix} \rho_i & \rho_{i-1} & \dots & \rho_{i-j+1} \\ \rho_{i+1} & \rho_i & \dots & \rho_{i-j+2} \\ \vdots & \vdots & \ddots & \vdots \\ \rho_{i+j-1} & \rho_{i+j-2} & \dots & \rho_i \end{vmatrix} \quad \begin{array}{l} i \geq 0 \\ j \geq 1 \end{array}$$

and on the theorem 1 in B.2.

It is easy to see that

$$\Delta(i,j) = (-1)^{j(j-1)/2} H_j^{i-j+1}(\rho).$$

$H_j^{i-j+1}(\rho)$ is the numerator of $\varepsilon_{2(j-1)}^{i-j+1}(\rho)$. Using theorem 3 in B.5.2. we easily deduce the corner characterization of Beguin et al.

Theorem 14 .- A second order stationary process admits a minimal ARMA (p,q) representation if and only if its autocorrelation function ρ is such that

$$\begin{array}{ll} \Delta(i,j) = 0 & \forall i \geq q+1 \quad \text{and} \quad j \geq p+1 \\ \Delta(i,p) \neq 0 & \forall i \geq q \\ \Delta(q,j) \neq 0 & \forall j \geq p. \end{array}$$

Proof : The condition $\begin{cases} \varepsilon_{2p}^n = 0 & \forall n \geq q-p+1 \\ \text{and } \varepsilon_{2p}^{q-p} \neq 0 \end{cases}$

is equivalent to

$$(F) \quad \begin{cases} \Delta(i,p+1) = 0 & \forall i \geq q+1 \\ \text{and } \Delta(q,p+1) \neq 0 \end{cases}$$

so that the condition of the theorem is redundant ; this is to show that we have to detect a corner of zeroes. The given condition can be obtained by using the recurrence relationship between Hankel determinants (B.4.3. (3)) which becomes in terms of $\Delta'(i,j) = H_j^{i-j+1}(\rho)$:

$$(G) \quad \left\{ \Delta'(i,j)\Delta'(i-2,j) - [\Delta'(i-1,j)]^2 = \Delta'(i-1,j+1)\Delta'(i-1,j-1) \quad \forall i, \forall j \geq 1 \right.$$

The quantities $\Delta'(i,j)$ can be disposed in the following way :

j i \	1	2	p	p+1	p+2	
1	$\Delta'(1,1)$	$\Delta'(1,2)$...				
2	$\Delta'(2,1)$	$\Delta'(2,2)$...				
⋮	⋮	⋮				
q			$\Delta'(q,p)$	$\Delta'(q,p+1)$	$\Delta'(q,p+2)$	
q+1			$\Delta'(q+1,p)$	$\Delta'(q+1,p+1)$		
q+2			$\Delta'(q+2,p)$			

Suppose that we have computed this array from the autocorrelation function of a minimal ARMA (p,q) process.

$$\text{Then } \left\{ \begin{array}{l} \Delta'(i,p+1) = 0 \quad \forall i \geq q+1 \\ \text{and } \Delta'(q,q+1) \neq 0 \end{array} \right. .$$

From (G) we have $\Delta'(p,q) \neq 0$ and $\Delta'(q,p+2) \neq 0$.

If $\Delta'(q+1,p) = 0$ we have $\Delta'(q+2,p) = 0$,

$$\Delta'(q+3,p) = 0 \text{ and so on, i.e. } \Delta'(i,p) = 0, \quad \forall i \geq q+1.$$

This contradicts the minimality of p.

Thus $\Delta'(q+1,p) \neq 0$ and $\Delta'(i,p+2) = 0, \quad \forall i \geq q+1$. We have shown that

$$(F) \text{ implies } \left\{ \begin{array}{l} \Delta'(q,p) \neq 0 \\ \Delta'(q,p+2) \neq 0 \\ \text{and } \Delta'(i,p+2) = 0, \quad \forall i \geq q+1. \end{array} \right.$$

Therefore $\Delta'(q,j) \neq 0 \quad \forall j \geq p$
and $\Delta'(i,j) = 0 \quad \forall i \geq q+1, \quad \forall j \geq p+1.$

As $\Delta'(q+1,p) \neq 0$, (G) shows that
 $\Delta'(q+2,p) \neq 0$ and so on, i.e.
 $\Delta'(i,p) \neq 0, \quad \forall i \geq q. \blacksquare$

In their paper, Beguin et al give properties of the Δ -array and statistical testing procedures. This method has been recently used by several authors in their investigations in time series domain.

Liu and Hanssens have used a modification of it to identify parsimonious rational forms of transfer functions. They illustrate their procedure by a simulated example. They also mentioned in their work a forthcoming paper by De Gooijer and Heuts which contains critical remarks and applications of the corner method to examples.

C. The vector case.

The direct estimation of p and q in the multivariate case, from the empirical auto- and cross-covariances has been much less studied than in the univariate case.

Recently Box and Tiao have proposed an extension of the determinantal methods, but their paper deals little with this subject. They say to be "studying sampling properties of estimates of appropriate functions" of matrices of determinants built from $(R(s))_{s \in \mathbb{Z}}$. Our method is based on the matrix and vector ε -algorithms but is no longer a determinantal one because these algorithms are not expressed in terms of determinants. We shall only give below the outlines of this extension to the multivariate case. It is detailed in [Berlinet, 1984] where other algorithms are also considered.

C.1.- Notations and assumptions.

Let n be a fixed integer strictly greater than 1 and $I = \{1, 2, \dots, n\}$.

$\pi_1(\mathbb{Z})$ is the real vector space of processes $(X_t)_{t \in \mathbb{Z}}$ taking their values in \mathbb{R}^n , defined on a probability space (Ω, A, Pr) and such that $\forall t \in \mathbb{Z} \quad EX_t = 0$.

$\pi_2(\mathbb{Z})$ is the subset of $\pi_1(\mathbb{Z})$ of second order stationary (wide sense) processes :

$$\text{if } X \in \pi_2(\mathbb{Z}) \quad X_t = (X_{1,t}; X_{2,t}; \dots; X_{n,t})'$$

$E(X_{k,t} X_{\ell,s})$ depends only on k, ℓ and $(t-s)$ and will be noted $R_{k,\ell}(t-s)$.

(We have $R_{\ell,k}(s-t) = R_{k,\ell}(t-s)$).

For $s \in \mathbb{Z}$ we note $R_X(s) = (R_{ij}(s))_{\substack{i \in I \\ j \in I}}$.

We call multivariate white noise of covariance matrix $\Sigma (\neq 0)$ any element U of $\pi_2(\mathbb{Z})$ such that $E(U_t U_s') = \delta_{ts} \Sigma$.

The components of U are n univariate white noises (possibly correlated) but such that $U_{i,t}$ and $U_{j,s}$ are not correlated if $t \neq s$.

C.1.1.- Multivariate ARMA models.

An element $X=(X_t)_{t \in \mathbb{Z}}$ of $\pi_2(\mathbb{Z})$ is said to admit an ARMA (p,q) representation if there exist matrices $A_0, A_1, \dots, A_p = I_n$

$$B_0, B_1, \dots, B_q = I_n$$

and a multivariate white noise U such that

$$(5) \quad \forall t \in \mathbb{Z} \quad \sum_{i=0}^p A_i X_{t-p+i} = \sum_{i=0}^q B_i U_{t-q+i} .$$

C.1.2.- Remarks.

a) Generally it is impossible to transform equation (5) in a system of n equations, each of them being an univariate ARMA representation of $X_{i,t}$, $i \in I$.

b) If $\forall i A_i = 0$, X is a multivariate moving average $MA(q)$ or $ARMA(0,q)$.

c) If $\forall i B_i = 0$, X is a multivariate autoregressive process $AR(p)$ or $ARMA(p,0)$;

d) If we note $\alpha(z)$ the polynomial matrix $I_n + \sum_{i=0}^{p-1} A_i z^{p-i}$ and $\beta(z)$ the polynomial matrix $I_n + \sum_{i=0}^{q-1} B_i z^{q-i}$ equation (5) can be written as

$$\alpha(B)(X) = \beta(B)(U) .$$

C.1.3.- Existence, invertibility and identifiability.

The problem of identifiability of multivariate ARMA processes (that is the determination of the matrices $(A_i)_{0 \leq i \leq p-1}$ and $(B_i)_{0 \leq i \leq q-1}$ from the covariance matrices of X) is much more complicated than in the univariate case [Hannan, 1969; Deistler, 1980 ; Priestley, 1981].

Theorem 15 .- If $\det \alpha(z) \neq 0$ for $|z| \leq 1$ (existence and stationarity condition)

and $\det \beta(z) \neq 0$ for $|z| \leq 1$ (invertibility condition)

there exists only one element X of $\pi_2(Z)$ verifying (5).

Moreover $\forall t \in \mathbb{Z}$, $X_t = \sum_{s=0}^{\infty} T_s U_{t-s}$ where T_s is the coefficient of z^s in the Taylor series expansion of $T(z) = [\alpha(z)]^{-1} \beta(z)$ ($T_0 = I_n$).

Among all the ARMA representations of the same process we shall call canonical ARMA representation (when it does exist) the one which has the lowest degree p (of autoregression) among those which have the lowest degree q (of moving average) and which, moreover, verifies $\text{rank } A_0 = n$ and: a greatest common left divisor of $\alpha(z)$ and $\beta(z)$ is I_n . Such a canonical model is identifiable (we have taken for A_0 and B_0 a condition slightly stronger than that given by Hannan [1969] : $\text{rank}[A_0 ; B_0] = n$).

We shall suppose in the following that X verifies the above conditions and has a canonical ARMA representation. We have only considered

covariance matrices because in the multivariate case correlation matrices have not the properties of the autocorrelation function in the scalar case : in general the cross-correlations are not symmetric and may attain their maximum absolute value (less than or equal to one) at any lag.

C.2. - Fundamental theorem 16 : characterization of the degrees of a canonical ARMA(p,q) representation.

p and q are the degrees of the canonical ARMA representation of X if and only if the sequence $(R(s))_{s \in \mathbb{Z}}$ of its covariance matrices satisfies a difference equation with constant matrix coefficients of minimal order p from the minimal rank (q-p+1) :

$$\begin{aligned} \forall \ell \geq q-p+1 \quad \sum_{i=0}^p A_i R(\ell+i) &= 0 & \det A_0 &\neq 0 \\ & & A_p &= I_n \\ \sum_{i=0}^p A_i R(q-p+i) &\neq 0 \end{aligned}$$

Proof : As in the univariate case, necessity is obvious by calculating

$$E\left(\sum_{i=0}^p A_i X_{t-p+i} X'_{t-p-\ell}\right).$$

Sufficiency is derived from the multivariate extension of the case p = 0

[Berlinet, 1982 (b)] .

Therefore we are in the same situation that in the univariate case : we have to find an algorithm which gives us the minimal rank and the minimal order of a difference equation satisfied by a matrix sequence. From the part B of this paper, a natural extension is the use of the matrix ε -algorithm : the rule (D) (B.5) can be applied to matrix sequences. An other possibility is the use of the vector ε -algorithm applied to the column vectors of the matrices : we use the rule (D) with the following definition of the "inverse" of a vector of \mathbb{R}^n :

Definition : Let $y \in \mathbb{R}^n$, $y \neq 0$. The inverse of y , y^{-1} , is the vector $y/||y||^2$ where $|| \quad ||$ is the euclidian norm.

However, in this case the theorems on the ε -algorithm we shall use only give us necessity of the vanishing of terms of the ε -array. This differs from the scalar case for which we have given a necessary and sufficient condition.

C.3.- The matrix ε -algorithm.

The following theorem is a consequence of the theory of Padé approximants upon non-commutative algebras [Draux, 1983] and of theorem 16.

Theorem 17.- Let X be an element of $\pi_2(\mathbb{Z})$ admitting a minimal ARMA (p,q) representation. If the matrix ε -algorithm is applied to the sequence R of autocovariance matrices of X then

$$\begin{cases} \varepsilon_{2p}^{\ell}(R) = 0 & \forall \ell \geq q-p+1 \\ \varepsilon_{2p}^{q-p}(R) \neq 0 \end{cases}$$

Theorem 18 states that the above condition is necessary and sufficient in the case $p = 1$.

Theorem 18.- The process X admits a minimal ARMA $(1,q)$ representation if and only if

$$\begin{cases} \varepsilon_2^{\ell}(R) = 0 & \forall \ell \geq q \\ \varepsilon_2^{q-1}(R) \neq 0 \end{cases}$$

Proof : Under the usual assumption that all the quantities involved in the theorem are computable :

$$\varepsilon_2^{\ell}(R) = 0 \iff [\bar{R}(\ell+1) - R(\ell)] [\bar{R}(\ell+1)]^{-1} = I - [\bar{R}(\ell+1) - R(\ell)] [\bar{R}(\ell+2) - R(\ell+1)]^{-1}$$

$$\varepsilon_2^{\ell}(R) = 0 \iff R(\ell) [\bar{R}(\ell+1)]^{-1} = R(\ell+1) [\bar{R}(\ell+2)]^{-1}$$

$$(\varepsilon_2^{\ell}(R) = 0, \forall \ell \geq q) \iff (\text{There exists a constant matrix } A_0 \text{ such that } \forall \ell \geq q, A_0 R(\ell) + R(\ell+1) = 0).$$

We conclude by using theorem 16. ■

We shall come back on the comparison between the two algorithms with numerical examples (real or simulated), on the invariants and on further investigation of the properties of the algorithms studied above in the vector as well as in the scalar case in a next paper.

References.

- ANDERSON, T.W. : The Statistical Analysis of Time Series - Wiley, New-York. 1971
- ANSLEY, C.F., W.A. SPIVEY and W.J. WROBLESKI : On the structure of moving average processes - Journal of Econometrics 6, 1977, 121-134
- BAUER, F.L. : Connections between the q-d algorithm of Rutishauser and the ϵ -algorithm of Wynn - Deutsche Forschungsgemeinschaft Tech. Rep. Ba/106, 1957
- BÉGUIN, J.M., C. GOURIÉROUX and A. MONFORT : Identification of a mixed autoregressive-moving average process : the corner method - Time series, Anderson, O.D., ed., North-Holland, Amsterdam, 1980, 423-436
- BERLINET, A. : Une méthode de détermination des degrés d'un modèle ARMA - Publications IRMA, Vol 3, Fasc 6, Lille, 1981
Exposé aux Journées de Statistique de Bruxelles (24-27 mai 1982)
- BERLINET, A. : Degrees of an ARMA model : estimation and tests. Compstat - Toulouse 1982 (a).
- BERLINET, A. : Estimation des degrés d'un ARMA multivarié - Publications IRMA, Vol. 4, fasc. 3, Lille, 1982 (b).
- BERLINET, A. : Some useful algorithms in time series analysis. In : Alternative approaches to time series analysis. Pub. des Facultés Universitaires Saint-Louis (Belgique), 1984.
- BOX, G.E.P. and G.M. JENKINS : Time Series Analysis - Forecasting and Control - Holden-Day, San Francisco, 1976
- BOX, G.E.P. and G.C. TIAO : Modeling Multiple Time Series with Applications. JASA, Vol 76, Numb. 376, 1981, 802-816
- BRÉZINSKI, C. : Accélération de la convergence en Analyse Numérique - Lecture Notes in Math. N°584, Springer-Verlag, Berlin, 1977
- BRÉZINSKI, C. : Algorithmes d'accélération de la convergence - Etude Numérique. Technip, Paris, 1978
- BRÉZINSKI, C. : Sur le calcul de certains rapports de déterminants - Lecture Notes in Math. n°765, Springer-Verlag, Berlin, 1979, 184-210
- CORDELLIER, F. : Interprétation géométrique d'une étape de l' ϵ -algorithme - Publ. 40, Labo de Calcul, Lille, 1973
- CORDELLIER, F. : Particular rules for the vector ϵ -algorithm - Numer. Math., 27, 1977, 203-207

- DE GOOIJER, J.G. and R.M.J. HEUTS : The Corner Method : An Investigation of an Order Discrimination Procedure for General ARMA Processes - Journal of the Operational Research Society, to appear
- DEISTLER, M. : Parametrization and consistent estimation of ARMA systems - Time Series, Anderson, O.D., ed., North-Holland, Amsterdam, 1980, 373-385
- DRAUX, A. : Publication A.N.O., U.E.R. d'I.E.E.A., Univ. de Lille I, to appear.
- DURBIN, J. : The fitting of time series models - Rev. Inst. Internat. Stat., 28, 1960, 233-244
- GRAY, H.L., G.D. KELLEY and D.D. MAC INTIRE : A new approach to ARMA modeling - Comm. in Stat. Simul. and Comp., B7, 1978, 1-77
- HANNAN, E.J. : The identification of vector mixed autoregressive-moving average systems - Biometrika, 56, 1969, 223-225
- HANNAN, E.J. : Multiple Time Series - Wiley, New-York, 1970
- HANNAN, E.J. and C.C. HEYDE : On limit theorems for quadratic functions of discrete time series - Ann. of Math. Stat. Vol. 43, n°6, 1972, 2058-2066.
- LIU, L.M. and D.M. HANSSENS : Identification of multiple-input transfer function models - Comm. Stat. - Theor. Meth., 11, 1982, 297-314
- MARDIA, K.V., J.T. KENT and J.M. BIBBY : Multivariate Analysis - Academic Press, New-York, 1979
- PRIESTLEY, M.B. : Spectral Analysis and Time Series - Vol 1 and 2 - Academic Press 1981
- PYE, W.C. and T.A. ATCHISON : An algorithm for the computation of higher order G-transformation, SIAM J. Numer. Anal. 10, 1973, 1-7
- SHANKS, D. : Non linear transformations of divergent and slowly convergent series - J. Math. Phys. 34, 1955, 1-42
- WIDDER, D.V. : The Laplace transform - Princeton University Press, 1946
- WYNN, P. : On a device for computing the $e_m(S_n)$ transformation - MTAC, 10, 1956
- WYNN, P. : Acceleration techniques for iterated vector and matrix problems - Math.Comp., 16, 1962, 301-322
- WYNN, P. : Singular rules for certain non linear algorithms - BIT 3, 1963, 175-195
- WYNN, P. : Upon systems of recursions which obtain among the quotients of the Padé table - Num. Math. 8, 1966, 264-269
- WYNN, P. : Upon an invariant associated with the epsilon-algorithm - MRC Technical Summary Report 675, 1966

ANNEXE 1 - (Extrait de Pub. IRMA - Lille - Vol. 3 - Fasc. 6 - 1981).

I - SIMPLIFICATION DES EQUATIONS AUX DIFFERENCES STOCHASTIQUES.1°) Notations.

On désignera par $\Pi(\mathbb{Z})$ l'espace vectoriel* sur \mathbb{R} des processus réels, centrés, indexés par \mathbb{Z} et définis sur l'espace probabilisé $(\Omega, \mathcal{A}, \Pr)$, par $\Pi_b(\mathbb{Z})$ le sous-espace de $\Pi(\mathbb{Z})$ constitué des processus $(x_t)_{t \in \mathbb{Z}}$ du second ordre bornés en moyenne quadratique (i.e. $\sup_{t \in \mathbb{Z}} E(x_t^2) < +\infty$) et par $\Pi_{st}(\mathbb{Z})$ le sous-ensemble de $\Pi_b(\mathbb{Z})$ constitué des processus stationnaires (au sens large : l'autocovariance $E(x_t x_{t+h})$ ne dépend que de $h \in \mathbb{Z}$).

On notera B l'automorphisme de $\Pi(\mathbb{Z})$ qui, au processus $(x_t)_{t \in \mathbb{Z}}$ associe $B((x_t)_{t \in \mathbb{Z}}) = (x_{t-1})_{t \in \mathbb{Z}}$;

$$B^0 = \text{Id}_{\Pi(\mathbb{Z})}, \quad B^i = B \circ B^{i-1}, \quad i \geq 1.$$

$\mathbb{R}[X]$ désignera l'anneau des polynômes à une indéterminée, à coefficients dans \mathbb{R} .

Pour tout élément P de $\mathbb{R}[X]$ tel que $P(z) = \sum_{i=0}^P a_i z^{P-i}$, $\forall z \in \mathbb{R}$ on notera $P(B)$ l'endomorphisme $\sum_{i=0}^P a_i B^{P-i}$.

2°) Equations aux différences stochastiques.

Soit $u = (u_t)_{t \in \mathbb{Z}} \in \Pi(\mathbb{Z})$. Soit $x = (x_t)_{t \in \mathbb{Z}} \in \Pi(\mathbb{Z})$, tel qu'il existe $\phi_1 \in \mathbb{R}[X]$, $\theta_1 \in \mathbb{R}[X]$ vérifiant $\phi_1(B)(x) = \theta_1(B)(u)$.

Soit $\Phi = \{P \in \mathbb{R}[X] \mid \exists Q \in \mathbb{R}[X] \text{ tel que } P(B)(x) = Q(B)(u)\}$

et $\Theta = \{Q \in \mathbb{R}[X] \mid \exists P \in \mathbb{R}[X] \text{ tel que } P(B)(x) = Q(B)(u)\}$

(*) On identifiera les processus qui sont modification l'un de l'autre.

Φ et Θ sont des idéaux de $\mathbb{R}[X]$, donc

Φ est l'ensemble des multiples d'un polynôme ϕ et Θ l'ensemble des multiples d'un polynôme θ .

Si les variables aléatoires réelles $(u_t)_{t \in \mathbb{Z}}$ sont linéairement indépendantes dans $L^1(\text{Pr})$, à tout polynôme $P \in \Phi$ correspond un unique polynôme $\ell(P) \in \Theta$ tel que $P(B)(x) = \ell(P)(B)(u)$; en effet, la relation $P(B)(x) = Q_1(B)(u) = Q_2(B)(u)$ entraîne la nullité d'une combinaison linéaire finie des $(u_t)_{t \in \mathbb{Z}}$.

On définit ainsi une application linéaire ℓ , surjective (par définition de Θ) mais non nécessairement injective de Φ sur Θ et il existe un unique réel α_0 tel que $\ell(\phi) = \alpha_0 \theta$; en effet, si $\ell(\phi) = Q\theta$ et $\ell(P\phi) = \theta$ on a $\ell(P\phi) = PQ\theta = \theta$ donc $PQ = 1$ et P et Q sont deux constantes inverses l'une de l'autre.

Proposition 1. - Si les processus x et u appartiennent à $\Pi_b(\mathbb{Z})$, les polynômes ϕ et θ n'ont pas de racine complexe de module différent de 1 en commun.

Démonstration : Supposons que $\phi(\lambda) = \theta(\lambda) = 0$ avec $|\lambda| \neq 1$.

$$\phi(z) = (z-\lambda) \phi_1(z) \quad \theta(z) = (z-\lambda) \theta_1(z)$$

$$(B-\lambda I) \phi_1(B)(x) = (B-\lambda I) \theta_1(B)(u) .$$

a) Si $\lambda = 0$ comme B est injectif on a $\phi_1(B)(x) = \theta_1(B)(u)$ ce qui contredit le fait que ϕ et θ soient les générateurs respectifs des idéaux Φ et Θ .

b) Si $\lambda \in \mathbb{R}^*$ supposons que la restriction de $(B-\lambda I)$ à $\Pi_b(\mathbb{Z})$ ne soit pas injective. Il existe $Z \in \Pi_b(\mathbb{Z})$, $Z \neq 0$ tel que $(B-\lambda I) Z = 0$

$$\text{donc } \forall t \in \mathbb{Z} \quad Z_{t-1} = \lambda Z_t$$

$$\text{et } \forall t \in \mathbb{Z} \quad Z_t = \lambda^{-t} Z_0 \quad \text{avec } Z_0 \neq 0$$

si $|\lambda| < 1$ $\lim_{t \rightarrow -\infty} E(Z_t^2) = +\infty$; si $|\lambda| > 1$ $\lim_{t \rightarrow +\infty} E(Z_t^2) = +\infty$,

ce qui est contradictoire avec le fait que $Z \in \Pi_b(\mathbb{Z})$.

On a donc encore $\phi_1(B)(x) = \theta_1(B)(u)$ et on conclut comme au a).

c) Si $\lambda \in \mathbb{C} - \mathbb{R}$

$$\phi_1(z) = (z-\bar{\lambda}) \phi_2(z) \quad \theta_1(z) = (z-\bar{\lambda}) \theta_2(z)$$

$$(B^2 - (2\text{Re}\lambda) B + |\lambda|^2 I) \phi_2(B)(x) = (B^2 - (2\text{Re}\lambda) B + |\lambda|^2 I) \theta_2(B)(u)$$

Supposons qu'il existe $Z \in \Pi_b(\mathbb{Z})$, $Z \neq 0$ tel que

$$(B^2 - (2\text{Re}\lambda) B + |\lambda|^2 I) Z = 0$$

$$\text{alors } \forall t \in \mathbb{Z} \quad Z_{t+2} - \frac{2\text{Re}(\lambda)}{|\lambda|^2} Z_{t+1} + \frac{Z_t}{|\lambda|^2} = 0$$

équation aux différences finies dont l'équation caractéristique a pour solutions $\frac{1}{\lambda}$ et $\frac{1}{\bar{\lambda}}$

$$\text{donc } \forall t \in \mathbb{Z} \quad Z_t = \frac{\alpha}{\lambda^t} + \frac{\beta}{\bar{\lambda}^t}$$

$$Z_0 \text{ et } Z_1 \text{ étant réels on a } Z_0 = (\alpha + \beta) \in \mathbb{R}$$

$$\text{et } (\bar{\lambda}\alpha + \lambda\beta) \in \mathbb{R}$$

$$\text{si } \lambda = x + iy \quad \bar{\lambda}\alpha + \lambda\beta = x(\alpha + \beta) + iy(-\alpha + \beta)$$

$$x \in \mathbb{R}, y \in \mathbb{R}$$

$$\text{donc} \quad -\alpha + \beta = 0 \quad \text{et} \quad \alpha = \beta$$

$$\text{et} \quad \forall t \in \mathbb{Z} \quad Z_t = \frac{Z_0}{2} \frac{(\bar{\lambda}^{-t} + \lambda^t)}{|\lambda|^2} = \frac{\operatorname{Re}(\lambda^t)}{|\lambda|^2} Z_0 \quad (\text{avec } E(Z_0^2) \neq 0)$$

$$\text{si} \quad |\lambda| < 1 \quad \lim_{t \rightarrow +\infty} E(Z_t^2) = +\infty$$

$$\text{si} \quad |\lambda| > 1 \quad \lambda = r e^{i\theta} \quad \operatorname{Re}(\lambda^t) = r^t \cos t\theta$$

$$r \in]1, +\infty[$$

$$\theta \in [0, 2\pi[$$

On peut donc trouver une suite $(t_k)_{k \in \mathbb{Z}}$ d'entiers relatifs tels que $\lim_{k \rightarrow +\infty} E(Z_{t_k}^2) = +\infty$ et on peut conclure de la même façon qu'au b).

Corollaire. - Dans $\Pi_b(\mathbb{Z})$ on peut "simplifier" l'équation

$$\phi(B)(x) = \theta(B)(u) \quad \text{en}$$

$$\phi_1(B)(x) = \theta_1(B)(u) \quad (\text{cas a) et b)}) \quad \text{ou en}$$

$$\phi_2(B)(x) = \theta_2(B)(u) \quad (\text{cas c)})$$

lorsque ϕ et θ ont en commun une racine λ de module différent de 1.

II - EXEMPLES NUMÉRIQUES DÉTERMINISTES.

Exemple 1.-

$$X(1) = 1 \quad X(2) = 0$$

$$\forall N \geq 1 \quad X(n+2) - 0,5 X(n+1) - 3X(n) = 0$$

$X(n)$ vérifie une équation aux différences finies à coefficients constants d'ordre minimal $K = 2$ à partir du rang minimal $N = 1$.

Polynôme caractéristique de l'équation :

$$\lambda^2 - 0,5 \lambda - 3 = (\lambda - 2) (\lambda + 1,5) \quad \lambda_1 = 2$$

$$\lambda_2 = -1,5$$

Invariant de l' ϵ -algorithme :

$$I = -((-0,5) + 2) (-3 - 0,5 + 1)^{-1} = 0,6$$

Calcul de
$$I(n) = \sum_{i=0}^{2K-2} (-1)^i \epsilon_i^n \epsilon_{i+1}^n$$

n	I(n)
1 = N	0,5999997
2	0,6000000
3	0,6000001
4	0,5999997
5	0,6000002
6	0,6000001
7	0,5999996

$$\frac{\epsilon_0^{10}}{\epsilon_0^9} = 1,6 \quad \frac{\epsilon_1^8}{\epsilon_1^9} = 0,9 \quad \frac{\epsilon_2^8}{\epsilon_2^7} = 33,1 \quad \frac{\epsilon_3^6}{\epsilon_3^7} = 3,3 \quad (\text{cf. théorème 6 - Chap. V : "approximation" de } \lambda_1 \text{ et } \lambda_2)$$

Exemple 2.-

$$X(1) = 1 \quad X(2) = 0 \quad X(3) = 3$$

$$\forall n \geq 2 \quad X(n+2) - 3,5 X(n+1) + 3 X(n) = 0$$

Ordre minimal $K = 2$ Rang minimal $N = 2$

Polynôme caractéristique

$$\lambda^2 - 3,5 \lambda + 3 = (\lambda - 2) (\lambda - 1,5) \quad \lambda_1 = 2$$

$$\lambda_2 = 1,5$$

Invariant

$$I = - (- 3,5 + 2) (3 - 3,5 + 1)^{-1} = 3$$

Calcul de $I(n)$

n	I(n)
2 = N	3,000000
3	3,000000
4	2,999999
5	3,000000
6	2,999999
7	3,000001

$$\frac{\epsilon_0^{10}}{\epsilon_0^9} = 2,08$$

$$\frac{\epsilon_1^8}{\epsilon_1} = 2,05$$

$$\frac{\epsilon_2^8}{\epsilon_2^7} = 1,48$$

$$\frac{\epsilon_3^6}{\epsilon_3^7} = 1,49$$

Exemple 3.-

$$X(1) = 1,156488 \quad X(2) = -1,432724 \quad X(3) = 0,2821283 \quad X(4) = -0,7778597$$

(les premières valeurs des exemples 3 et 4 ont été générées par un programme simulant une variable normale centrée réduite).

$$\forall n \geq 3 \quad X(n+2) - 3,5 X(n+1) + 3X(n) = 0$$

Ordre minimal $K = 2$ Rang minimal $N = 3$.

L'équation aux différences est la même qu'à l'exemple 2, donc aussi le polynôme caractéristique, les racines et l'invariant.

Calcul de $I(n)$

n	I(n)
3 = N	3,000000
4	3,000001
5	2,999999
6	3,000000
7	3,000000

$$\frac{\varepsilon_0^{10}}{\varepsilon_0^9} = 2,12$$

$$\frac{\varepsilon_1^8}{\varepsilon_1^9} = 2,08$$

$$\frac{\varepsilon_2^8}{\varepsilon_2^7} = 1,46$$

$$\frac{\varepsilon_3^6}{\varepsilon_3^7} = 1,48$$

Exemple 4.-

$$X(1) = 1,156488 \quad X(2) = -1,432724 \quad X(3) = 0,2821283 \quad X(4) = -0,7778597$$

$$X(5) = -1,007263$$

$$\forall n \geq 3 \quad X(n+3) - 2,3 X(n+2) - 2,1 X(n+1) + 5,4 X(n) = 0$$

Ordre minimal K = 3 Rang minimal N = 3

Polynôme caractéristique :

$$\lambda^3 - 2,3 \lambda^2 - 2,1 \lambda + 5,4 = (\lambda + 1,5) (\lambda - 1,8) (\lambda - 2)$$

$$\lambda_1 = 2$$

$$\lambda_2 = 1,8$$

$$\lambda_3 = -1,5$$

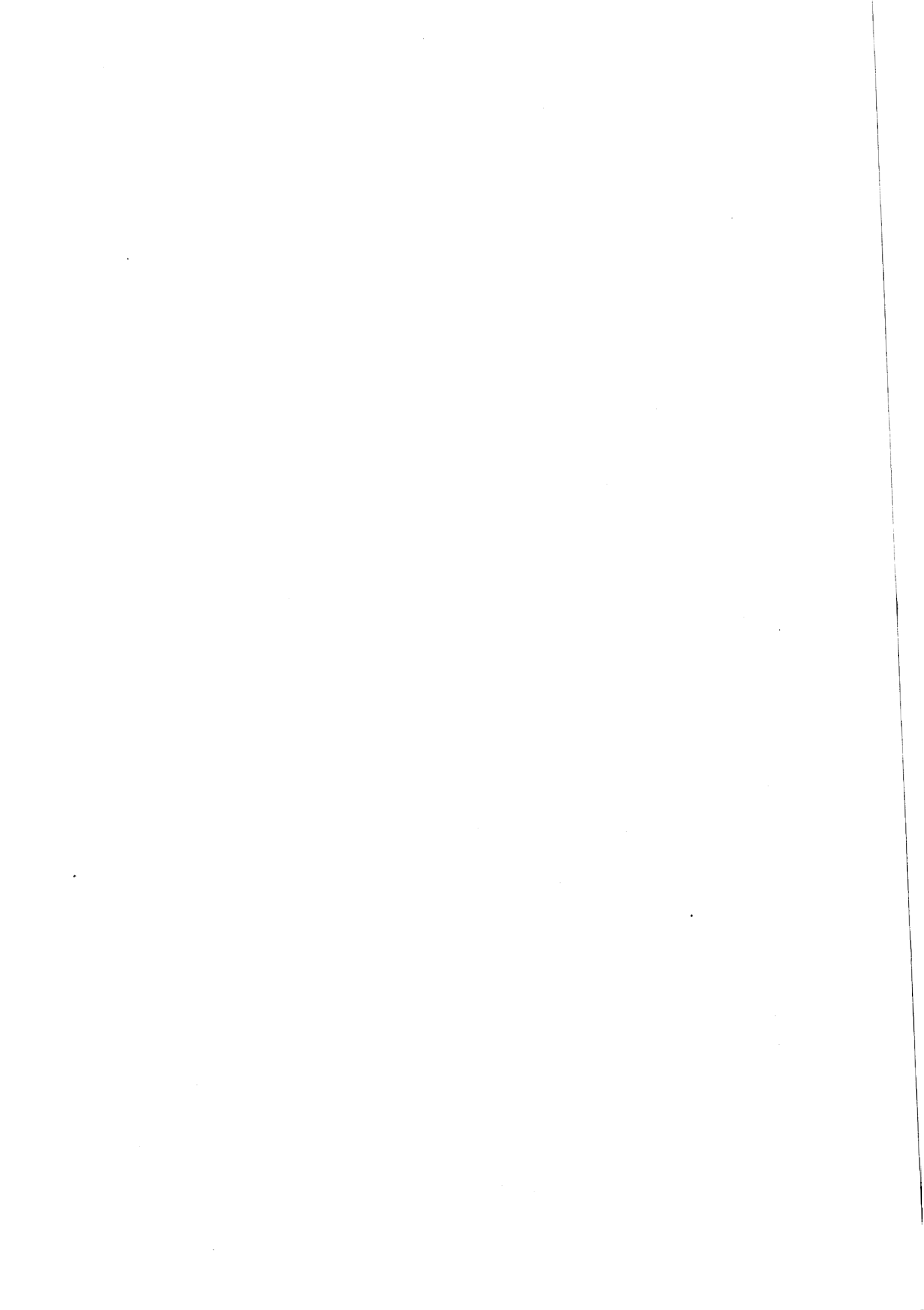
Invariant

$$I = - (-2,1 - 4,6 + 3) (5,4 - 2,1 - 2,3 + 1) - 1 = 1,85$$

Calcul de I(n)

n	I(n)
3 = N	1,850000
4	1,850000
5	1,849998

$$\frac{\epsilon_0^{10}}{\epsilon_9} = 2,41 \quad \frac{\epsilon_1^8}{\epsilon_1} = 2,75 \quad \frac{\epsilon_2^8}{\epsilon_2} = -0,55 \quad \frac{\epsilon_3^6}{\epsilon_3} = 0,19 \quad \frac{\epsilon_4^6}{\epsilon_4} = 1,68 \quad \frac{\epsilon_5^4}{\epsilon_5} = 1,69$$



Alternative approaches to time series analysis.
PUB. des Facultés Universitaires Saint Louis
(Belgique) 1984.

CHAPITRE 6

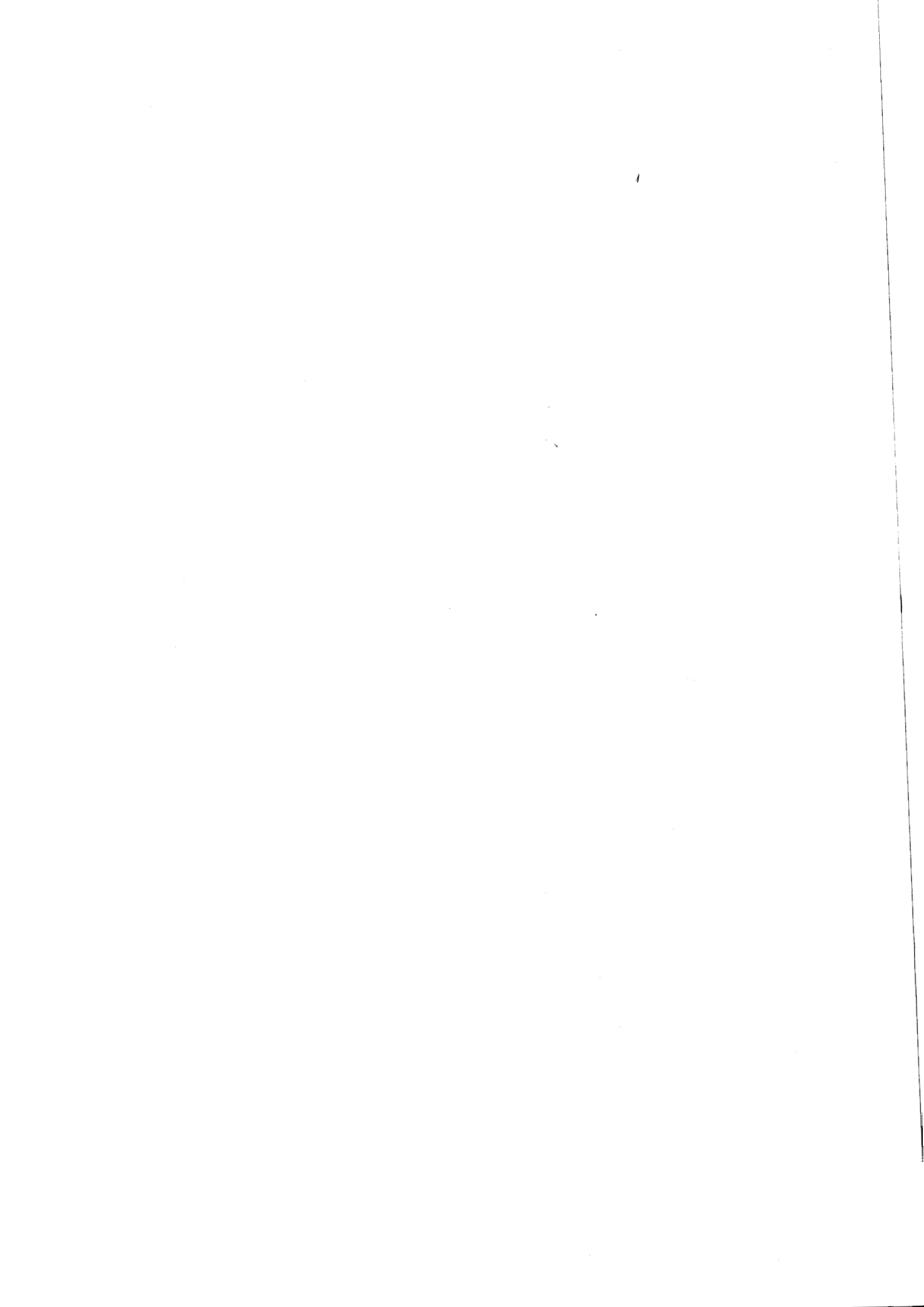
SOME USEFUL ALGORITHMS IN TIME SERIES ANALYSIS

Abstract : In the first part of this paper we re-state some properties of the ϵ -algorithm applied to the autocorrelation function of a univariate ARMA process and how to use it to estimate the unknown degrees of such a process or, more generally, of parsimonious rational forms of transfer functions. We also show how general orthogonal polynomials can give results about the fitting of time series by means of autoregressive processes, without any reference to spectral theory. In the second part we use two extensions of the scalar ϵ -algorithm to estimate the degrees of multivariate ARMA models.

AMS 1980 - Subject classification 62M10

Key words : Univariate and multivariate time series

ARMA models. Epsilon - algorithm.



S U M M A R Y.

A) INTRODUCTION.

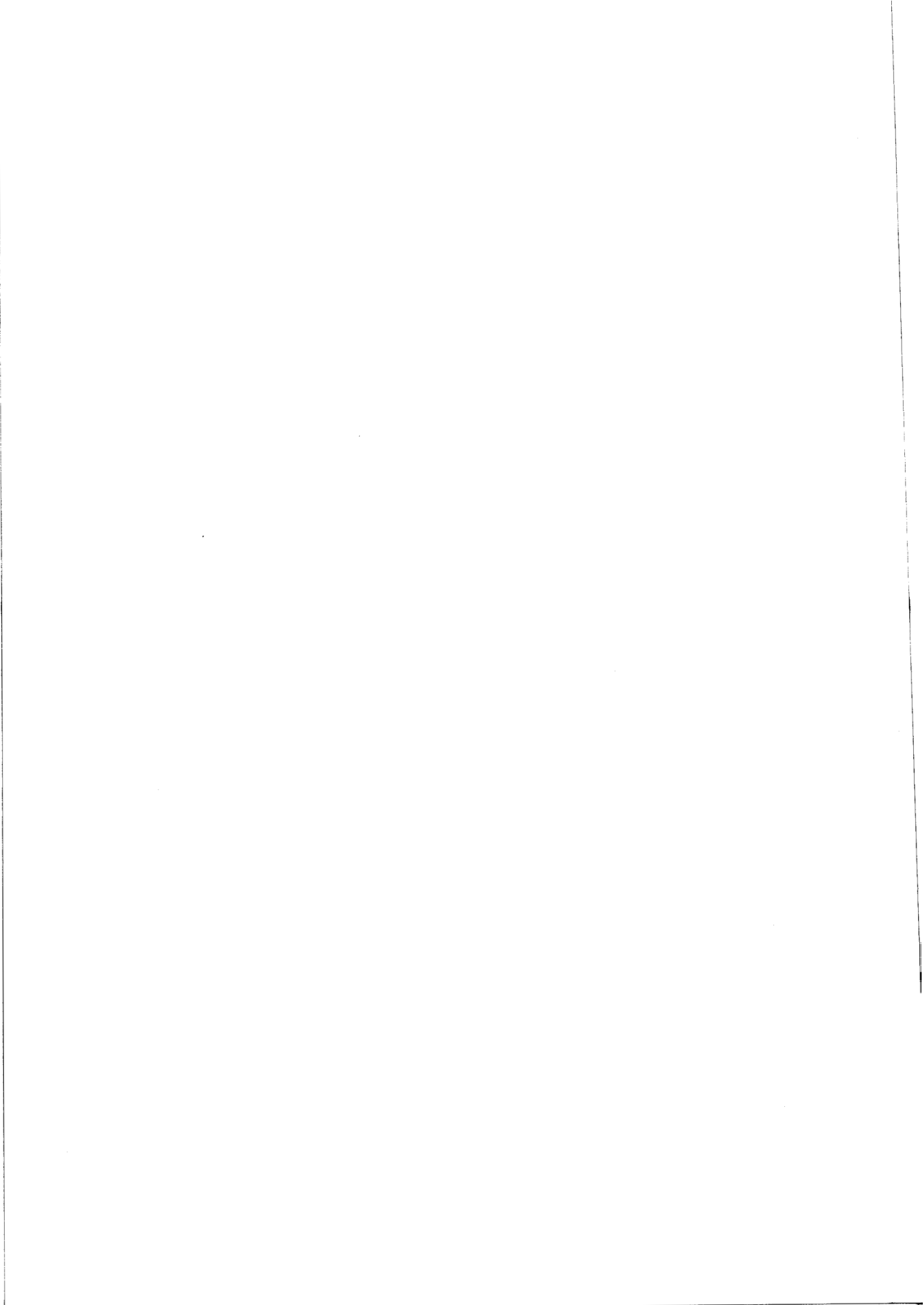
B) UNIVARIATE CASE.

- B.1. - Notations and assumptions.
- B.2. - Fundamental theorem 1 : characterization of the degrees of a minimal ARMA representation.
- B.3. - Statement of the problem.
- B.4. - The ϵ -algorithm.
 - B.4.1. - Theorem 2 : Determinantal form of the ϵ -algorithm.
 - B.4.2. - Theorem 3 : Fundamental property of the ϵ -algorithm.
- B.5. - Estimation of p and q .
- B.6. - Examples.
 - B.6.1. - Example 1 : Sunspot data.
 - B.6.2. - Example 2
 - B.6.3. - Example 3 : Gas furnace data.
- B.7. - Other properties of the ϵ -algorithm applied to the autocorrelation function ρ of a minimal ARMA (p,q) process.
 - B.7.1. - Theorem 5 (Bauer - Wynn invariant).
 - B.7.2. - Other results.
- B.8. - Autoregressions and general orthogonal polynomials.

C) MULTIVARIATE CASE.

- C.1. - Notations and assumptions.
 - C.1.1. - Multivariate ARMA models.
 - C.1.2. - Remarks.
 - C.1.3. - Existence, invertibility, identifiability and minimal representation.
- C.2. - Multivariate moving averages.
 - C.2.1. - Theorem 7.
 - C.2.2. - Remark.
 - C.2.3. - Corollary.
 - C.2.4. - Lemma.
 - C.2.5. - Theorem 8.
- C.3. - Characterization of the degrees of a minimal ARMA representation.
- C.4. - The matrix ϵ -algorithm.
- C.5. - The vector ϵ -algorithm.
- C.6. - Estimation of p and q .
- C.7. - Simulations.
 - C.7.1. - Example 1 : Bivariate minimal ARMA(2,1).
 - C.7.2. - Example 2 : Bivariate AR(1).

REFERENCES.



A - INTRODUCTION.

In this paper we deal with algorithms which are well known in numerical analysis, which are very easy to compute, of low cost and which provide a preliminary estimation procedure for parameters of a model directly from the empirical autocovariance function. In the first part we re-state some properties of the ϵ -algorithm applied to the autocorrelation function of a univariate ARMA process and how to use it to estimate the unknown degrees of such a process or, more generally, of parsimonious rational forms of transfer functions. In this case the ϵ -algorithm can be expressed in terms of determinants. Two other methods have been used to estimate the degrees : the R-S algorithm [Gray, Kelley and Mac Intire] and the corner method [Béguin, Gouriéroux and Monfort] ; these are also determinantal methods. Their links with the ϵ -algorithm have been studied in [Berlinet, 1982(b)]. We also show, without any reference to spectral theory, how general orthogonal polynomials can give us results about the fitting of time series by means of autoregressive processes, and therefore, about the computation of the partial autocorrelation function.

In the second part we use the extension of the characterization of moving averages given by Ansley, Spivey and Wroblewski for univariate processes to characterize minimal ARMA (p,q) representations. Two extensions of the scalar ϵ -algorithm are used to estimate p and q : the vector and the matrix ϵ -algorithm.

We shall use definitions and properties of ARMA processes which are detailed in [Rozanov, 1967], [Hannan, 1970], [Anderson, 1971], [Priestley, 1981], and some results well known in acceleration of convergence ; see [Brézinski, 1977, 1980] for the theoretical aspects and proofs of theorems and [Brézinski, 1978] for the practical ones (FORTRAN programs).

B - UNIVARIATE CASE.

This case has been examined in detail in [Berlinet, 1982 (b)] where proofs and examples can be found. We will simply re-state here the basic properties of the ε -algorithm and give three applications. We will suppose that all the quantities involved in the following are computable. This is a usual assumption when sequence transformations occur.

B.1. - Notations and assumptions.

Let $X = (X_t)_{t \in \mathbb{Z}}$ be a real stationary (wide sense) process ; we suppose without loss of generality that $EX_t = 0, \forall t \in \mathbb{Z}$.

Let $\gamma = (\gamma_h)_{h \in \mathbb{Z}}$ and $\rho = (\rho_h)_{h \in \mathbb{Z}}$ be respectively the autocovariance and autocorrelation function of X :

$$\forall (t,h) \in \mathbb{Z}^2 \quad E(X_t X_{t+h}) = \gamma_h (\gamma_0 \neq 0)$$

$$\rho_h = \gamma_h \gamma_0^{-1} .$$

X is said to be an AutoRegressive Moving Average process of degrees p and q (ARMA (p,q)) if it satisfies the following stochastic difference equation

$$(C) \left\{ \begin{array}{l} (1) \quad \forall t \in \mathbb{Z} \quad \sum_{i=0}^p \phi_i X_{t-p+i} = \sum_{i=0}^q \theta_i U_{t-q+i} \\ \text{where} \quad \phi_0 \neq 0, \quad \phi_p = 1 \\ \quad \quad \theta_0 \neq 0, \quad \theta_q = 1 \\ U = (U_t)_{t \in \mathbb{Z}} \text{ is a white noise } (EU_t = 0, EU_t U_s = \sigma^2 \delta_{ts}, \sigma > 0) \end{array} \right.$$

(1) can be written as $\phi(B)(X) = \theta(B)(U)$ with

$$\phi(z) = \sum_{i=0}^p \phi_i z^{p-i}$$

$$\theta(z) = \sum_{i=0}^q \theta_i z^{q-i}$$

B is the backward shift operator :

$$B((x_t)_{t \in \mathbb{Z}}) = (x_{t-1})_{t \in \mathbb{Z}} .$$

It is well known that a sufficient condition for the existence and unicity of a second order stationary (wide sense) process satisfying (C) is that the roots of ϕ are of modulus strictly greater than 1. We assume that the roots of θ also verify the above condition so that U is the innovation process of X. Then, we have the following expansion for X_t :

$$(2) \quad \forall t \in \mathbb{Z} , \quad X_t = \sum_{j=0}^{\infty} c_j \sum_{\ell=0}^q \theta_{\ell} U_{t-q+\ell-j} = \sum_{i=0}^{\infty} \psi_i U_{t-i}$$

where the c_j 's are the coefficients of the Taylor series expansion of $[\phi(z)]^{-1}$ ($c_0 = 1$).

Representation (1) will be called minimal when p and q are as small as possible i.e. when ϕ and θ have no common root.

It is well to note that under the above assumptions an equation such like (1) can be simplified when ϕ and θ have in common roots of modulus strictly greater than 1.

For any sequence $s = (s_n)_{n \in \mathbb{Z}}$ of real numbers we note

- * Δs the sequence $(s_{n+1} - s_n)_{n \in \mathbb{Z}}$.
- * $\Delta^k s$ the sequence $\Delta(\Delta^{k-1} s)$, $k \geq 2$.
- * $H_k^n(s)$ the Hankel determinant of s of order k from the rank n :

$$H_k^n(s) = \begin{vmatrix} s_n & s_{n+1} & s_{n+2} & \dots & s_{n+k-1} \\ s_{n+1} & s_{n+2} & \dots & \dots & s_{n+k} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ s_{n+k-1} & \dots & \dots & \dots & s_{n+2k-2} \end{vmatrix} \quad \begin{matrix} k \in \mathbb{N}^* \\ n \in \mathbb{Z} \end{matrix}$$

$$H_0^n(s) = 1 , \quad \forall n \in \mathbb{Z} .$$

B.2. - Fundamental theorem 1 : Characterization of the degrees of a minimal ARMA representation [Béguin et al].

The second order stationary process X admits an ARMA (p,q) minimal representation if and only if the sequence ρ satisfies a difference equation of minimal order p from minimal rank $(q-p+1)$:

$$\forall n \geq q-p+1 \quad \sum_{i=0}^p \phi_i \rho_{n+i} = 0 \quad , \quad \phi_0 \neq 0 \quad , \quad \phi_p = 1$$

$$\sum_{i=0}^p \phi_i \rho_{q-p+i} \neq 0 \quad .$$

B.3. - Statement of the problem.

If we dispose of an algorithm which, applied to a sequence satisfying a difference equation, gives us the minimal rank and the minimal order of this equation, we are able to recognize a stationary minimal ARMA (p,q) and to calculate p and q from the sequence ρ , or from a sequence deduced from ρ and satisfying a difference equation of the same minimal order from the same minimal rank.

However, in practice, we only observe X_0, \dots, X_T and we do not know ρ . A natural strategy is to apply our algorithm to an estimate of ρ , computed from X_0, \dots, X_T and if X is an ARMA (p,q) to estimate, from the results, the degrees p and q . As an estimate of ρ we have used in the following $\hat{\rho}$, the empirical autocorrelation function. This is not a restriction, other estimates can be used (and sometimes give better results); for instance, smoothed autocorrelation derived via spectral analysis.

B.4. - The ϵ -algorithm.

The ϵ -algorithm is a recursive lozenge algorithm first defined by Wynn [1956] as follows : let $s = (s_n)_{n \in \mathbb{Z}}$ be an element of $\mathbb{R}^{\mathbb{Z}}$:

$$(D) \begin{cases} \forall n \in \mathbb{Z} , \quad \varepsilon_{-1}^0(s) = 0 & \varepsilon_0^n(s) = s_n \\ \forall k \in \mathbb{Z} , \quad \forall n \in \mathbb{Z} , \quad \varepsilon_{k+1}^n(s) = \varepsilon_{k-1}^{n+1}(s) + [\varepsilon_k^{n+1}(s) - \varepsilon_k^n(s)]^{-1} . \end{cases}$$

B.4.1. - Theorem 2 : Determinantal form of the ε -algorithm
 [Wynn, 1956].

Let $k \in \mathbb{N} , n \in \mathbb{Z}$

$$\varepsilon_{2k}^n(s) = [H_{k+1}^n(s)] [H_k^n(\Delta^2 s)]^{-1}$$

$$\varepsilon_{2k+1}^n(s) = [H_k^n(\Delta^3 s)] [H_{k+1}^n(\Delta s)]^{-1}$$

B.4.2. - Theorem 3 : Fundamental property of the ε -algorithm.

Let $S \in \mathbb{R} , k \in \mathbb{N} , N \in \mathbb{Z}$ and $s = (s_n)_{n \in \mathbb{N}}$.

A necessary and sufficient condition for $\varepsilon_{2k}^n(s) = S , \forall n > N$, is the existence of a family $(a_i)_{0 \leq i \leq k}$ of real numbers such that

$$\begin{cases} \sum_{i=0}^k a_i \neq 0 \\ \text{and} \quad \sum_{i=0}^k a_i (s_{n+i} - S) = 0 , \quad \forall n > N . \end{cases}$$

From B.2. and B.4.2. we get :

Theorem 4.- The second order stationary process X admits a minimal ARMA(p,q) representation if and only if

$$\begin{cases} \varepsilon_{2p}^n(\rho) = 0 & \forall n \geq q-p+1 \\ \text{and} \quad \varepsilon_{2p}^{q-p}(\rho) \neq 0 . \end{cases}$$

In the ϵ -array, only columns corresponding to k being even will be of interest to us. It is possible to compute only these columns, using the cross-rule of Wynn [1966] :

$$[\epsilon_{k+2}^{n-1} - \epsilon_k^n]^{-1} - [\epsilon_k^n - \epsilon_{k-2}^{n+1}]^{-1} = [\epsilon_k^{n+1} - \epsilon_k^n]^{-1} - [\epsilon_k^n - \epsilon_k^{n-1}]^{-1} .$$

A minimal ARMA(p,q) process is therefore characterized by the following even columns in the ϵ -array built from its autocorrelation function ρ :

column \ line	0	2	4	...	2p
q-p	$\epsilon_0^{q-p} = \rho_{q-p}$				
q-p+1	$\epsilon_0^{q-p+1} = \rho_{q-p+1}$	ϵ_2^{q-p}			
q-p+2	$\epsilon_0^{q-p+2} = \rho_{q-p+2}$	ϵ_2^{q-p+1}	ϵ_4^{q-p}		
...	...				
q	$\epsilon_0^q = \rho_q$	-----	-----	-----	$\epsilon_{2p}^{q-p} \neq 0$
q+1	$\epsilon_0^{q+1} = \rho_{q+1}$	-----	-----	-----	0
...	...				0
					0
					0
					...

B.5. - Estimation of p and q.

We suppose that we have observed realizations X_0, X_1, \dots, X_T of a minimal ARMA(p,q) process X the degrees, coefficients and mean of which are unknown. We compute the empirical autocorrelation function $\hat{\rho} = (\hat{\rho}_h)_{h \in \mathbb{Z}, |h| \leq T}$

$$\hat{\rho}_h = \left[\sum_{t=0}^{T-|h|} (X_t - \bar{X}_T)(X_{t+|h|} - \bar{X}_T) \right] \left[\sum_{t=0}^T (X_t - \bar{X}_T)^2 \right]^{-1/2}$$

$$\bar{X}_T = (T+1)^{-1} \sum_{t=0}^T X_t$$

and we apply the ϵ -algorithm to $\hat{\rho}$ or to a sequence deduced from $\hat{\rho}$ by means of some transformation : numerical results

are better when the autocorrelation changes sign at each lag ; when we have successive groups of autocorrelations of the same sign, we apply the algorithm to $((-1)^n \hat{\rho}_n)_{n \in \mathbb{Z}}$. Rounding errors can be serious when calculating the inverse of a difference of two numbers close to one another. When the autocorrelations become small quickly we can use $((-2)^n \hat{\rho}_n)_{n \in \mathbb{Z}}$ or $(\alpha \times (-1)^n \hat{\rho}_n)_{n \in \mathbb{Z}}$, $\alpha \in \mathbb{R}^+$, for instance.

To estimate p and q we have to detect a sudden variation in the ϵ -array. We are not compelled to build the whole array from $(\hat{\rho}_h)_{|h| \leq T}$: we have to compute at least ϵ_{2p}^{q-p} and ϵ_{2p}^{q-p+1} ; for that we have to use at least $\hat{\rho}_h$ for $q-p \leq h \leq q+3p+1$ (It is assumed that T is greater than $q+3p+1$).

A good sequential way to proceed is to compute the SW-NE diagonal from $\epsilon_0^h = \hat{\rho}_h$ and the NW-SE diagonal from $\epsilon_0^{-h} = \hat{\rho}_{-h}$ after the calculation of $\hat{\rho}_h$.

B.6. - Examples.

Most of the simulations have shown that the degrees of the process were correctly estimated, or that a small number of possibilities (with the true values among them) could be selected. Of course the method is not miraculous : when ρ

is badly approximated we are unable to conclude anything or we give bad estimates of p and q . Other examples can be found in [Berlinet, 1982 b].

B.6.1. - Example 1 : Sunspot data.

Many authors have fitted the series of Wolfer's sunspot numbers which can be found for instance in Anderson [1971] for the years 1749 to 1924.

The entire series or subseries are often fitted with AR(2) models. With the ε -algorithm applied to $((-1)^h \hat{\rho}_h)$ we obtain $\hat{p} = 2$, $\hat{q} = 0$ for the entire series. A more refined analysis of the sunspot series, using different models (linear, bilinear and threshold models) and Akaike's information criterion can be found in [Priestley]. Gray et al also deal with this example, using the R-S - algorithm.

Sunspot numbers (1770 to 1869) - ϵ -array built from $((-1)^{\hat{\rho}_h})_{-10 \leq h \leq 10}$

column \ line	0	2	$2\hat{p}=4$	6	8	10	12	14	16
-10	0,410								
-9	-0,330	-0,032							
-8	0,169	0,069	3,9E-3						
-7	0,044	0,29	-2,8E-4	1,4E-3					
-6	-0,212	-0,045	2,4E-3	5,5E-4	3,5E-4				
-5	0,266	0,038	-3,0E-3	3,7E-4	5,2E-4	8,0E-4			
-4	-0,169	-0,088	4,3E-3	1,6E-3	9,1E-4	6,8E-4	6,8E-4		
-3	-0,070	-0,19	2,1E-4	-2,4E-3	-2,3E-3	6,8E-4	6,8E-4	-1,2E-3	
-2	0,428	0,073	-1,4E-3	-2,3E-3	-2,4E-3	1,0E-3	4,3E-4	8,8E-4	1,2E-3
-1	-0,806	-0,073	-2,0E-3	-1,4E-3	-4,5E-5	2,1E-3	-6,7E-3	1,7E-3	-7,2E-5
$\hat{q}=0$	1	0,097	0,019	0,010	6,4E-3	4,8E-3	4,0E-3	3,3E-3	3,0E-3
1	-0,806	-0,073	-2,0E-3	-1,4E-3	-4,5E-5	2,1E-3	-6,7E-3	1,7E-3	-7,2E-5
2	0,428	0,073	-1,4E-3	-2,3E-3	-2,4E-3	1,0E-3	4,3E-4	8,8E-4	1,2E-3
3	-0,070	-0,19	2,1E-4	-2,4E-3	-2,3E-3	6,8E-4	6,8E-4	-1,2E-3	
4	-0,169	-0,088	4,3E-3	1,6E-3	9,1E-4	6,8E-4	6,8E-4		
5	0,266	0,038	-3,0E-3	3,7E-4	5,2E-4	8,0E-4			
6	-0,212	-0,045	2,4E-3	5,5E-4	3,5E-4				
7	0,044	0,29	-2,8E-4	1,4E-3					
8	0,169	0,069	3,9E-3						
9	-0,330	-0,032							
10	0,410								

B.6.2. - Example 2 : Given by Gray et al [1978, p. 9].

ARMA (2,1) p = 2 q = 1

$$0,68X_{t-2} - 1,32X_{t-1} + X_t = -0,8 U_{t-1} + U_t$$

$$U_t \sim N(0,1)$$

ϵ -array built from $(\hat{\rho}_h)_{-2 \leq h \leq 12}$

column line	0	2	4	6	8	10
-2	0,064					
-1	0,582	2,7				
0	1	0,79	0,53			
1	0,582	2,7	-1,1	-0,46		
2	0,064	-1,9	-0,29	-5,0	-0,073	
3	-0,347	-0,82	1,7	0,12	0,044	-0,011
4	-0,566	-0,52	-0,19	0,038	0,39	-0,061
5	-0,514	-0,58	0,98	-0,037	-0,070	0,050
6	-0,288	-1,1	-0,32	-0,060	-0,043	-0,023
7	0,025	0,61	0,23	0,043	-0,010	-0,013
8	0,229	0,35	-0,98	-0,22	-0,014	
9	0,303	0,27	-0,17	-4,5		
10	0,240	0,34	0,13			
11	0,071	-0,95				
12	-0,074					

ϵ -array built from $((-1)^h \hat{\rho}_h)_{-2 \leq h \leq 12}$

column line	0	2	$\hat{2p}=4$	6	8	10
-2	0,064					
-1	-0,582	-0,12				
0	1	0,21	0,072			
$\hat{q}=1$	-0,582	-0,12	0,027	-0,26		
2	0,064	0,57	-3,2E-3	-6,7E-3	-3,9E-3	
3	0,347	0,13	-6,3E-3	-5,1E-3	-4,6E-3	-3,4E-3
4	-0,566	-0,071	-4,4E-3	-4,6E-3	-5,0E-3	-5,7E-3
5	0,514	0,054	-4,6E-3	-4,4E-3	-5,6E-3	-5,7E-3
6	-0,288	-0,090	-7,6E-3	-6,8E-3	-5,7E-3	-5,6E-3
7	-0,025	7,4	-6,4E-3	-8,2E-3	-0,018	-3,4E-3
8	0,229	0,057	-3,1E-3	-0,012	-6,8E-3	
9	-0,303	-0,034	1,8E-3	8,5E-3		
10	0,240	0,042				
11	0,071	-0,074				
12	-0,074					

The estimation of p and q from the first array is not possible when from the second the choice $\hat{p} = 2$, $\hat{q} = 1$ is obvious.

B.6.3. - Example 3 : Gas furnace data [Box and Jenkins Series J]

This example has been dealt with by Liu and Hanssens using the corner method. We give it here for comparison.

The method is straightforward and needs no "prewhitening". We shall use notations and results of Box and Jenkins [1976, p. 378].

Suppose that the transfer function model

$$Y_t = v(B) X_t + N_t \quad v(B) = \sum_{i=0}^{\infty} v_i B^i$$

may be parsimoniously parametrized in the form

$$Y_t = \delta^{-1}(B) \omega(B) X_{t-b} + N_t$$

where $\delta(B) = 1 - \delta_1 B - \delta_2 B^2 \dots - \delta_r B^r$

and $\omega(B) = \omega_0 - \omega_1 B \dots - \omega_s B^s$, $(b, r, s) \in \mathbb{N}^3$.

From the empirical autocovariance function of the input X_t and the empirical cross-covariance function between the input and the output it is possible to get rough estimates \hat{v}_i of the transfer function weights v_i , $0 \leq i \leq K$, by solving a linear system. The weights satisfy a difference equation of minimal order r from minimal rank $b + s - r + 1$ and the first b are zero. For gas furnace data the estimates of the weights are :

i	0	1	2	3	4	5	6	7	8	9	10	11	12
\hat{v}_i	-0,02	0,10	-0,06	-0,53	-0,63	-0,88	-0,52	-0,32	-0,06	0,06	-0,10	-0,06	-0,04

If we apply the ϵ -algorithm to $(-1)^i (\hat{v}_i)_{0 \leq i \leq 13}$ we get the following array (we have to use the singular rule of the algorithm because $\hat{v}_8 = -\hat{v}_9$). Details about particular and singular rules can be found in [Brézinski, 1977, p. 132].

column \ line	0	$\hat{2r=2}$	4	6	8
0	- 0,02				
1	- 0,10	- 0,07			
2	- 0,06	- 0,10	- 0,03		
3	0,53	0,14	0,07	0,08	
4	- 0,63	0,03	0,08	0,06	0,08
5	0,88	0,15	0,09	0,08	0,08
6	- 0,52	0,005	0,04	- 0,02	0,17
7	0,32	0,06	0,03	0,06	0,15
8	- 0,06	- 0,06	- 0,12	- 0,35	0,05
9	- 0,06	- 0,06	- 0,07	- 0,13	
10	- 0,1	- 0,07	- 0,06		
11	0,06	- 0,002			
12	- 0,04				

From this array we can select a small number of models to study more precisely. They must satisfy

$$\left\{ \begin{array}{l} * b \in \{1,3\} \\ * (r = 0 \text{ and } b + s = 6) \text{ or } (r = 1 \text{ and } b + s = 5) \end{array} \right.$$

Box and Jenkins only consider the case $b = 3$ and $r \geq 1$, for which they conclude $r = 1$ and $s = 2$. In that case, we get the same values directly from the array.

B.7. - Other properties of the ϵ -algorithm applied to the autocorrelation function ρ of a minimal ARMA(p,q) process.

B.7.1. - Theorem 5 (Bauer - Wynn invariant).-

If $p \geq 1$, then for $n \geq q-p+1$ we have

$$I(n) = \sum_{i=0}^{2p-2} (-1)^i \epsilon_i^n(\rho) \epsilon_{i+1}^n(\rho) = -p + \phi'(1) [\phi(1)]^{-1}.$$

Example : The following array gives $I(n,k) = \sum_{i=0}^{2k-2} (-1)^i \epsilon_i^n(\hat{\rho}') \epsilon_{i+1}^n(\hat{\rho}')$

for sunspot data, where $\hat{\rho}'_h = (-1)^h \hat{\rho}_h$:

$k \backslash n$	1	$\hat{p}=2$	3	4	5
-4	- 1,707	- 0,866	1,491	- 1,090	- 3,904
-3	- 0,141	- 0,848	- 0,967	- 2,910	- 2,925
$\hat{q}-\hat{p}=-2$	- 0,347	- 0,875	- 2,033	- 2,944	- 2,626
-1	- 0,446	- 1,125	- 4,491	17,680	- 2,397
0	- 0,554	- 1,152	- 2,000	- 1,681	- 5,312
1	- 0,653	- 1,134	- 1,081	- 2,981	- 2,482
2	- 0,859	- 1,133	- 1,678	0,598	- 1,024
3	0,707	- 1,086	- 1,566	- 1,187	
4	- 0,389	- 1,014	- 1,860	1,015	
5	- 0,556	- 0,964	- 1,031		
6	- 0,828	- 0,963	- 2,131		
7	0,352	- 1,022			
8	- 0,339	- 1,010			
9	- 0,446				
10	- 0,510				

B.7.2. - Other results : (detailed in [Berlinet, 1982 (b)]).

When ρ is very well approximated, the ε -array provides estimates of the roots of ϕ .

Under usual assumptions one can get convergence theorems for the elements of the ε -array, which lead to χ^2 -tests.

B.8. - Autoregressions and general orthogonal polynomials.

To estimate the dimension of a model, a very useful parameter is the partial autocorrelation function : if it vanishes on $[p, +\infty[\cap \mathbb{N}$, the series is an autoregression. We can associate to any autocovariance function of a stationary second order process a functional on $\mathbb{R}[X]$, vector space of polynomials with real coefficients, and general orthogonal polynomials [Akhiezer, 1965 ; Brézinski, 1980 ; Draux, 1981] with respect to this functional which are very useful to compute the partial autocorrelation function $r = (r_h)_{h \in \mathbb{N}}^*$ and to fit data by means of autoregressive processes of increasing degree. In a classical way orthogonal polynomials are considered with respect to the spectral measure of a stationary process. Here no reference to spectral theory is made. Our results explain that the partial autocorrelation function can be computed from the autocovariance function by means of the q-d algorithm of Rustishauser [Berlinet, 1982, a)]. The results given in this paragraph can easily be extended to the multivariate case.

Let $X = (X_t)_{t \in \mathbb{Z}}$ be a real, zero-mean, second order stationary process with autocovariance function $\gamma = (\gamma_h)_{h \in \mathbb{Z}}$.

We define a linear functional c_p on $\mathbb{R}[X]$ by setting $c_p[X^i] = \gamma_{p-i}$, $i \in \mathbb{N}$, where p is a fixed non zero integer.

We assume the existence of the p^{th} orthogonal polynomial with respect to the functional c_p , i.e. the polynomial ψ_p of exact degree p determined

by $\psi_p(0)$ and

$$(3) \quad c_p(x^i \psi_p(x)) = 0, \quad 0 \leq i \leq p-1$$

Setting $\psi_p(0) = 1$ and $\psi_p(x) = 1 + \sum_{i=1}^p a_i x^i$, (3) is equivalent to

$$\gamma_{p-j} + \sum_{i=1}^p a_i \gamma_{p-i-j} = 0, \quad 0 \leq j \leq p-1$$

or

$$(4) \quad \gamma_j + \sum_{i=1}^p a_i \gamma_{j-i} = 0, \quad 1 \leq j \leq p.$$

The equations in (4) are Yule-Walker's for the autoregressive process of degree p with autocovariance equalling γ on $\{0, 1, 2, \dots, p\}$.

The systems of orthogonal polynomials with respect to the functionals c_p and c_{p+1} are adjacent so that they can be computed recursively by applying to the autocovariance function of X the algorithms using the well-known recurrence relationships between them. If we apply the q -d algorithm to $(\gamma_i)_{i \in \mathbb{Z}}$:

$$q_1^n = \gamma_{n+1} \gamma_n^{-1}; \quad e_0^n = 0, \quad n \in \mathbb{Z}$$

$$q_{k+1}^n e_k^n = e_k^{n+1} q_k^{n+1}; \quad q_k^n + e_k^n = q_k^{n+1} + e_{k-1}^{n+1}, \quad k \in \mathbb{N}^*, \quad n \in \mathbb{Z}$$

we have $\forall h \geq 2 \quad r_h = (-1)^{h-1} \prod_{i=1}^h q_i^{1-h}$ [Berlinet, 1982, b]

then the relationships

$$\tilde{q}_1^p = \gamma_{p-1} \gamma_p^{-1}; \quad \tilde{e}_0^p = 0, \quad p \in \mathbb{N}^*$$

$$\tilde{e}_k^0 = e_k^0; \quad \tilde{q}_k^0 = q_k^0, \quad k \in \mathbb{N}^*$$

$$\tilde{q}_{k+1}^{p+1} \tilde{e}_k^{p+1} = \tilde{e}_k^p \tilde{q}_k^p; \quad \tilde{q}_{k+1}^p + \tilde{e}_k^p = \tilde{q}_{k+1}^{p+1} + \tilde{e}_{k+1}^{p+1}, \quad k \in \mathbb{N}^*, \quad p \in \mathbb{N}$$

give us $Q_k^p(x)$, $k \in \mathbb{N}^*$, the unitary orthogonal polynomials with respect to the functional c_p by means of the relationship:

$$Q_{k+1}^p(x) = (x - \frac{\tilde{q}_k^p}{\tilde{q}_{k+1}^p} - \frac{\tilde{e}_k^p}{\tilde{e}_k^p}) Q_k^p(x) - \frac{\tilde{q}_k^p}{\tilde{q}_k^p} \frac{\tilde{e}_k^p}{\tilde{e}_k^p} Q_{k-1}^p(x).$$

Therefore $\psi_p(x) = Q_p^p(x) [Q_p^p(0)]^{-1} = -r_p Q_p^p(x)$

Example : For sunspot data we obtain in this way, from the empirical autocovariances the following estimates for the partial autocorrelations :

p	1	2	3	4	5
\hat{r}_p	0,806	- 0,633	0,079	0,063	0,002

The AR(2) fit is given by

$$\psi_2(x) = 0,63 x^2 - 1,31 x + 1$$

and the associated invariant (cf. B.7.1.) is - 1,12 .

As shown by Brezinski these algorithms are strongly linked with the ϵ -algorithm ; we are currently studying their stability to compare with Durbin's algorithm when coming near non-stationarity.

C - MULTIVARIATE CASE.-

The direct estimation of p and q in the multivariate case, from the empirical auto - and cross - covariances has been far less studied than in the univariate case.

Recently Box and Tiao have proposed an extension of the determinantal methods, but their paper deals little with this subject. They say they are "studying sampling properties of estimates of appropriate functions" of matrices of determinants built from $(R(s))_{s \in \mathbb{Z}}$. Our method is based on the matrix and vector ε -algorithms but is no longer a determinantal one because these algorithms are not defined in terms of determinants.

We give a characterization of minimal ARMA representations using an extension to the vector case of the results of Ansley et al on moving averages. We end by giving some results of simulation. The definitions and properties of multivariate models needed in this part can be found in Rozanov [1967], Hannan [1970] and Priestley [1981].

C.1. - Notations and assumptions.

Let n be a fixed integer strictly greater than 1 and $I = \{1, 2, \dots, n\}$.

For any matrix M , M' will denote the transposed matrix and M^{-1} , when existing, the inverse one. We shall use the same notation for an n dimensional vector and the matrix $n \times 1$ of its coordinates in the canonical basis ; for $\ell \in \mathbb{N}^*$, I_ℓ will denote the unit matrix of order ℓ .

$\pi_1(\mathbb{Z})$ is the real vector space of processes $(X_t)_{t \in \mathbb{Z}}$ taking their values in \mathbb{R}^n , defined on a probability space $(\Omega, \mathcal{A}, \text{Pr})$ and such that $\forall t \in \mathbb{Z} \quad \text{EX}_t = 0$.

$\pi_2(\mathbb{Z})$ is the subset of $\pi_1(\mathbb{Z})$ of second order stationary (wide sense) processes :

$$\text{if } X \in \pi_2(\mathbb{Z}) \quad X_t = (X_{1,t}; X_{2,t}; \dots; X_{n,t})'$$

$E(X_{k,t} X_{\ell,s})$ depends only on k, ℓ and $(t-s)$ and will be denoted by $R_{k,\ell}(t-s)$. (We have $R_{\ell,k}(s-t) = R_{k,\ell}(t-s)$).

$$\text{For } s \in \mathbb{Z}, \text{ let } R_X(s) = (R_{ij}(s))_{\substack{i \in I \\ j \in I}}.$$

We call multivariate white noise of covariance matrix $\Sigma (\neq 0)$ any process ξ of $\pi_2(\mathbb{Z})$ such that $E(\xi_t \xi_s') = \delta_{ts} \Sigma$. The components of ξ are n univariate white noises possibly correlated, but such that $\xi_{i,t}$ and $\xi_{j,s}$ are not correlated if $t \neq s$.

C.1.1. - Multivariate ARMA models.

A process $X = (X_t)_{t \in \mathbb{Z}}$ of $\pi_2(\mathbb{Z})$ is said to have an ARMA(p,q) representation if there exist matrices $A_0, A_1, \dots, A_p = I_n$
 $B_0, B_1, \dots, B_q = I_n$ and a multivariate white noise ξ such that

$$(5) \quad \forall t \in \mathbb{Z} \quad \sum_{i=0}^p A_i X_{t-p+i} = \sum_{i=0}^q B_i \xi_{t-q+i}.$$

The integers p and q (not unique in general) are the degrees of the model.

C.1.2. - Remarks.

a) Generally it is impossible to find from (5) a univariate ARMA representation for each component $X_{i,t}$ of X_t .

b) If $p = 0$, X is a multivariate moving average MA(q) or ARMA(0,q).

c) If $q = 0$, X is a multivariate autoregressive process AR(p) or ARMA(p,0);

d) If $\alpha(z)$ denotes the polynomial matrix $I_n + \sum_{i=0}^{p-1} A_i z^{p-i}$ and $\beta(z)$ the polynomial matrix $I_n + \sum_{i=0}^{q-1} B_i z^{q-i}$ equation (5) can be written as

$$\alpha(B)(X) = \beta(B)(\xi)$$

where B is the usual backward shift operator in $\pi_2(\mathbb{Z})$.

C.1.3. - Existence, invertibility, identifiability and minimal representation.

The problem of identifiability of multivariate ARMA processes (that is the determination of the matrices $(A_i)_{0 \leq i \leq p-1}$ and $(B_i)_{0 \leq i \leq q-1}$ from the covariance matrices of X when p and q are known) is much more difficult than in the univariate case [Hannan, 1969 ; Deistler, 1980 ; Priestley, 1981].

Theorem 6.- A sufficient condition for the existence of a stationary process verifying (5) is that $\det \alpha(z) \neq 0$, $\forall z \in D(0,1)$ (closed unit disc in the complex plane).

Then X has the representation :

$$\forall t \in \mathbb{Z} \quad X_t = \sum_{s=0}^{\infty} T_s \xi_{t-s},$$

where T_s is the coefficient of z^s in the Taylor series expansion of $T(z) = [\alpha(z)]^{-1} \beta(z)$ ($T_0 = I_n$). If $\det \beta(z) \neq 0$, $\forall z \in D(0,1)$ the process is invertible. ξ is the innovation process of X .

For an ARMA process X , consider the class $C(X)$ of ARMA representations of lowest degree p (of autoregression) among those of lowest degree q (of moving average). Representations in $C(X)$ all have the same degrees and explain, as much as possible, X_t by its own past. Hannan's identifiability theorem [1969] implies that at most one representation in $C(X)$ satisfies

$$(6) \begin{cases} * \text{ rank } A_0 = n \\ * \text{ a greatest common left divisor (G C L D) of } \alpha(z) \text{ and } \beta(z) \text{ is } I_n. \end{cases}$$

Definition.- If there exists a representation in $C(X)$ satisfying (6) it will be called minimal ARMA representation of X .

Remarks :

* Such a minimal model is identifiable because we have taken for A_0 , B_0 , $\alpha(z)$ and $\beta(z)$ a condition slightly stronger than the necessary and sufficient condition of unicity of the representation given by Hannan [1969] : $\text{rank } \begin{bmatrix} A_0 \\ B_0 \end{bmatrix} = n$ and $\text{G C L D } (\alpha(z), \beta(z)) = I_n$; when p and q are known the second stage of analysis (estimation of the coefficients) can take place without any ambiguity.

* It must be noticed that a multivariate ARMA system has not always a minimal representation : the MA(1) model $X_t = A \xi_{t-1} + \xi_t$ with $A \neq 0$, $A^2 = 0$ and $E(\xi_t \xi_s') = \delta_{ts} I_n$ has also the AR(1) representation $X_t - A X_{t-1} = \xi_t$, therefore it has no minimal representation (see theorem 8 below).

We shall suppose in the following that X verifies the above conditions and has a minimal ARMA representation. We have only considered covariance matrices because in the multivariate case correlation matrices don't have the properties of the autocorrelation function in the scalar case : in general the cross-correlations are not symmetric and may attain their maximum absolute value (less than or equal to one) at any lag.

C.2. - Multivariate moving averages.

C.2.1. - Theorem 7.- (Extension to the multivariate case of the result of Ansley et al). Let $q \in \mathbb{N}^*$ and Y be a process of $\pi_2(Z)$ with

covariance matrices $R_Y(t)$, $t \in \mathbb{Z}$.

Y is a moving average and q is the lowest possible degree of a MA representation of Y if and only if

$$\begin{cases} R_Y(t) = 0, & \forall t > q \\ \text{and } R_Y(q) \neq 0 \end{cases}$$

Theorem 7 can be proved by means of Wold decomposition of a stationary process (see annex 2).

C.2.2. - Remark.- If, by convention, a white noise is said to be a moving average of degree 0, we see that for $q = 0$ the above characterization is, by definition, that of a white noise.

C.2.3. - Corollary.- The sum of two uncorrelated multivariate moving averages of lowest degrees q_1 and q_2 is a moving average of lowest degree $q \leq \max(q_1, q_2)$ (equality holds if $q_1 \neq q_2$).

Proof :

$$E[X_1(t+s) + X_2(t+s)] [X_1'(t) + X_2'(s)] = R_{X_1}(s) + R_{X_2}(s)$$

and

$$\begin{cases} R_{X_1}(s) = 0 & \forall s > q_1 \\ R_{X_2}(s) = 0 & \forall s > q_2 \end{cases} .$$

C.2.4. - Lemma.- If X has an ARMA(p,q) representation satisfying conditions of theorem 6, $(R(s))_{s \in \mathbb{Z}}$ satisfies a difference equation of order p from rank $(q-p+1)$.

Proof :

$$E\left(\sum_{i=0}^p A_i X_{t-p+i} X'_{t-p-\ell}\right) = \sum_{i=0}^q B_i E(\xi_{t-q+i} X'_{t-p-\ell})$$

thus

$$\sum_{i=0}^p A_i R_X(\ell+i) = 0, \quad \forall \ell \geq q-p+1 .$$

C.2.5. - Theorem 8.- If a moving average Y of lowest degree r has a minimal ARMA(p, q) representation then $p = 0$ and $q = r$.

Proof : The result is obvious if $r = 0$.

Otherwise, let $\sum_{i=0}^p A_i Y_{t-p+i} = \sum_{i=0}^q B_i \xi_{t-q+i}$ be the minimal representation of Y ($q \leq r$).

Suppose that $q < r$ (thus $p > 0$). From lemma C.2.4. we have $A_0 R_Y(\ell) = - \sum_{i=1}^p A_i R_Y(\ell+i)$, $\forall \ell \geq q-p+1$. Since $q-p+1 \leq r-1$

$R_Y(r) = A_0^{-1} (- \sum_{i=1}^p A_i R_Y(r+i)) = 0$. This contradicts the minimality of r .

Thus $q = r$ and $p = 0$.

C.3. - Characterization of the degrees of a minimal ARMA representation.

Theorem 9.- p and q are the degrees of the minimal ARMA representation of X if and only if the sequence $(R(s))_{s \in \mathbb{Z}}$ of covariance matrices of X satisfies a difference equation with constant matrix coefficients of minimal order p from minimal rank $(q-p+1)$:

$$\left\{ \begin{array}{l} \forall \ell \geq q-p+1 \quad \sum_{i=0}^p A_i R(\ell+i) = 0 \\ \text{and} \quad \sum_{i=0}^p A_i R(q-p+i) \neq 0 . \end{array} \right.$$

Proof : Let p and q be the degrees of the minimal ARMA representation of X .

a) If $(R(s))_{s \in \mathbb{Z}}$ satisfies a difference equation of order k from rank r , X has an ARMA ($k, k+r-1$) representation : let

$Y_t = \sum_{i=0}^k A_i X_{t-p+i}$, $t \in \mathbb{Z}$ where the A_i 's are the coefficients of the recurrence satisfied by $(R(s))_{s \in \mathbb{Z}}$.

$$E(Y_t Y'_{t+s}) = \sum_{j=0}^k A_j \sum_{i=0}^k R(-s+j-i) A_i'$$

Thus $Y \in \pi_2(\mathbb{Z})$ and $E(Y_t Y'_{t+s}) = 0 \quad \forall s > k+r-1$

Y is a moving average of degree $(k+r-1)$ and X is an ARMA($k, k+r-1$).

b) If the minimal order of the recurrence satisfied by $(R(s))_{s \in \mathbb{Z}}$ was $p_1 < p$, X would have, by a) and lemma C.2.4., an ARMA $(p_1, p_1 + (q-p+1) - 1)$ representation. This would contradict the minimality of q . Thus p is the minimal order of the recurrence. In the same way $r = q-p+1$ is minimal.

c) If $(R(s))_{s \in \mathbb{Z}}$ satisfies a recurrence of minimal order k from minimal rank r , X has an ARMA $(k, k+r-1)$ representation. The minimality of k, r and q implies

$$\begin{cases} q \leq k+r-1 \\ p \geq k \\ q-p+1 \geq r \end{cases} \quad \text{thus} \quad \begin{cases} p = k \\ \text{and } q = k+r-1 \end{cases}$$

Therefore we are in the same situation as in the univariate case : we have to find an algorithm which gives the minimal rank and minimal order of a difference equation satisfied by a matrix sequence. We shall use two extensions of the scalar ϵ -algorithm : the matrix and the vector versions. In the multivariate case we shall be able to obtain only sufficient conditions of vanishing for the elements of the ϵ -array. The proof of existence of Bauer-Wynn invariants can easily be extended to the multivariate case. For the matrix ϵ -algorithm, because of the non-commutativity of the matrix product, we get two different invariants ([Wynn, 1963]).

C.4. - The matrix ϵ -algorithm.

The rule (D) (B.4) can be applied to matrix sequences because it involves only computation of differences, sums and inverses. As in part B we shall suppose that all the quantities involved in the following are computable.

Theorem 10.- If the matrix ϵ -algorithm is applied to the sequence R of covariance matrices of a minimal ARMA(p,q) process we get :

$$(7) \quad \begin{cases} \epsilon_{2p}^{\ell}(R) = 0, & \forall \ell \geq q-p+1 \\ \epsilon_{2p}^{q-p}(R) \neq 0. \end{cases}$$

Proof : Theorem 10 follows from theorem 9 and the extension to the matrix case of the fundamental property of the ϵ -algorithm applied to a sequence satisfying a difference equation [Draux, 1983]. However sufficiency of (7) (analogous to condition of theorem 4) has not yet been proved for any p .

Theorem 11.- If $p = 1$, condition (7) of theorem 10 implies that X is a minimal ARMA(1,q) process.

Proof :

$$\epsilon_2^{\ell}(R) = R(\ell+1) + \{ [R(\ell+2) - R(\ell+1)]^{-1} - [R(\ell+1) - R(\ell)]^{-1} \}^{-1}$$

$$\epsilon_2^{\ell}(R) = 0 \iff \begin{pmatrix} [R(\ell+1) - R(\ell)] [R(\ell+1)]^{-1} \\ = I - [R(\ell+1) - R(\ell)] [R(\ell+2) - R(\ell+1)]^{-1} \end{pmatrix}$$

$$\epsilon_2^{\ell}(R) = 0 \iff R(\ell) [R(\ell+1)]^{-1} = R(\ell+1) [R(\ell+2)]^{-1}$$

$$(\epsilon_2^{\ell}(R) = 0, \forall \ell \geq q) \iff \left(\begin{array}{l} \text{(There exists a regular constant matrix } A_0 \\ \text{such that } \forall \ell \geq q \quad R(\ell) [R(\ell+1)]^{-1} = -A_0^{-1} \end{array} \right)$$

This last condition is equivalent to $A_0 R(\ell) + R(\ell+1) = 0, \forall \ell \geq q$ and the conclusion follows from theorem 9.

Theorem 12.- If the matrix ϵ -algorithm is applied to the sequence of covariance matrices of a minimal ARMA(p,q) process ($p \geq 1$) the quantities

$$I_1(n) = \epsilon_0^n \epsilon_1^n - \epsilon_2^n \epsilon_1^n + \epsilon_2^n \epsilon_3^n - \epsilon_4^n \epsilon_3^n + \dots + \epsilon_{2p-2}^n \epsilon_{2p-1}^n$$

and
$$I_2(n) = \epsilon_1^n \epsilon_0^n - \epsilon_1^n \epsilon_2^n + \epsilon_3^n \epsilon_2^n - \epsilon_3^n \epsilon_4^n + \dots + \epsilon_{2p-1}^n \epsilon_{2p-2}^n$$

are constant for $n \geq q-p+1$.

C.5. - The vector ϵ -algorithm.

The main deficiency of the matrix ϵ -algorithm is to involve a matrix inversion at each stage of the computation. This leads to an instability which grows with n . Therefore one may prefer (or use simultaneously) the vector ϵ -algorithm which is defined by means of the rule (D) with the following definition of the inverse of an n -dimensional vector :

Definition.- Let $y \in \mathbb{R}^n$, $y \neq 0$. The inverse of y is the vector $y^{-1} = y / \|y\|^2$, where $\| \cdot \|$ is the euclidian norm (y^{-1} is the inverse of y w.r.t. the unit sphere of \mathbb{R}^n).

Let $V_{i,\ell}$ be the i^{th} column vector of $R(\ell)$. If $R = (R(\ell))_{\ell \in \mathbb{Z}}$ satisfies a difference equation, the sequence $V_i = (V_{i,\ell})_{\ell \in \mathbb{Z}}$ satisfies the same equation for any $i \in I$. However the minimal order for each of the sequences V_i , $i \in I$, may be strictly lower than the minimal recurrence order of the sequence R . Nevertheless, the following theorem allows us to select a small number of possibilities for p and q .

Theorem 13.- Let X be a process of $\pi_2(\mathbb{Z})$ with a minimal ARMA(p,q) representation. If we apply the vector ϵ -algorithm to the sequence of vectors $(V_{i,\ell})_{\ell \in \mathbb{Z}}$ then $\epsilon_{2m_i}^{\ell} = 0 \quad \forall \ell \geq q-p+1$,

where m_i is the degree of the minimal polynomial of the matrix

$$\begin{pmatrix} -A_{p-1} & \dots & -A_1 & -A_0 \\ I_n & & & 0 \\ 0 & & & \vdots \\ \vdots & & & \vdots \\ \vdots & & & \vdots \\ 0 & \dots & I_n & 0 \end{pmatrix} \quad \text{for the} \quad \begin{pmatrix} V_{i,q} \\ \vdots \\ V_{i,q-p+2} \\ V_{i,q-p+1} \end{pmatrix}$$

and we have $m_i n^{-1} \leq p \leq m_i$.

Proof : Straightforward from our assumptions and theorem 95 of Brezinski [1977].

Because of the symmetry of the inner product $(.,.)$ in \mathbb{R}^n , we obtain the same invariant as in the univariate case :

Theorem 14.- Under assumptions of theorem 12, if $p \geq 1$, the scalar sequence $I(n) = \sum_{k=0}^{2m_i-2} (-1)^k (\epsilon_k^n, \epsilon_{k+1}^n)$ is constant for $n \geq q-p+1$.

C.6. - Estimation of p and q.

We suppose that X_0, X_1, \dots, X_T are $(T+1)$ observed realizations of a minimal ARMA(p,q) \mathbb{R}^n -valued process the degrees, coefficients and mean of which are unknown. We estimate $R_{ij}(s)$ by the empirical autocovariance :

$$\begin{cases} \hat{R}_{ij}(s) = (T+1)^{-1} \sum_{t=0}^{T-s} (X_{i,t+s} - \bar{X}_i) (X_{j,t} - \bar{X}_j) \\ \hat{R}_{ij}(-s) = \hat{R}_{ji}(s) \\ \text{with } \bar{X}_i = (T+1)^{-1} \sum_{i=0}^T X_{i,t} \end{cases} \quad \begin{matrix} 0 \leq s \leq T \\ i \in I \\ j \in I \end{matrix}$$

Then we apply the matrix ϵ -algorithm to the sequence \hat{R} or the vector ϵ -algorithm to each of the sequences $(\hat{V}_i)_{i \in I}$ where $\hat{V}_{i,\ell}$ is the i^{th} column vector of $\hat{R}(\ell)$ and we have to detect a sudden variation in the so-built arrays.

Theoretical results do not allow us to deduce from the arrays a unique value for (\hat{p}, \hat{q}) but only to select a limited number of possibilities. However, in practice, we get the same results as in the scalar case. Perhaps, it will be possible to prove converses of theorems 10 and 13.

Remarks made in the univariate case about the computation (B.5.) are still available. For the vector ϵ -algorithm particular and singular rules have been studied by Cordellier. Convergence theorems and statistical tests given in the univariate case ([Berlinet, 1982, b]) can be extended, with appropriate modifications, to the multivariate case, using results of Hannan [1976].

Let us remark, however, that the aim of our method is to provide preliminary estimates ; so that statistical tests will be more useful after a more refined study of the series.

C.7. - Simulations.

Simulations often give good results : a small number of possibilities for p and q , with the true one among them, can be selected from the ϵ -arrays. We give two examples. For each of them the starting values were zero, and 200 values of the process were generated with gaussian noise.

C.7.1. - Example 1 : Bivariate minimal ARMA(2,1) process.

$$A_0 = 0,2 \quad I_2 \quad A_1 = \begin{pmatrix} 0,1 & 0,7 \\ -0,8 & -0,1 \end{pmatrix}$$

$$B_0 = \begin{pmatrix} -0,3 & -0,2 \\ 0,4 & -0,6 \end{pmatrix} \quad \Sigma = \begin{pmatrix} 4 & 1 \\ 1 & 1 \end{pmatrix}$$

The vector ϵ -algorithm has been applied to the sequences $(\hat{V}_{1,s})$ and $(\hat{V}_{2,s})$ of column vectors of $\hat{R}(s)$, $-2 \leq s \leq 10$.

In each column of the array one finds on the left hand side results about \hat{V}_1 and on the right hand side results about \hat{V}_2 .

column line	0	2	$2\hat{p} = 4$	6	8
- 2	-10, 1,9 1,3 -12,				
- 1	- 0,5 -12, 11, - 0,23	- 1,7 0,83 0,9 1,5			
0	11, - 1,6 - 1,6 13,	0,02 - 1,3 - 1,5 1,6	0,6 - 0,11 - 0,27 0,71		
$q = 1$	- 0,5 11, -12, - 0,23	1,2 - 0,65 - 0,72 - 1,8	- 0,17 + 0,27 - 0,58 - 0,16	- 0,0009 - 0,08 - 0,09 - 0,007	
2	-10, 1,3 1,9 -12,	0,34 1,4 1,5 - 0,81	- 0,02 - 0,02 0,17 - 0,06	- 0,04 0,01 0,01 0,06	- 0,004 0,04 0,003 - 0,007
3	0,35 -10, 10, 0,008	- 1,3 0,52 0,42 1,4	- 0,04 - 0,03 - 0,06 0,06	- 0,002 - 0,01 0,03 - 0,009	- 0,007 - 0,0009 0,03 - 0,006
4	9, - 1,2 - 1,4 9,9	- 0,29 - 1, - 1,1 0,6	- 0,04 0,06 - 0,08 - 0,01	- 0,007 0,0003 0,04 - 0,005	- 0,002 - 0,01 0,03 - 0,003
5	- 0,33 8,5 - 8,8 0,34	0,97 - 0,63 - 0,45 - 1,	- 0,04 - 0,1 0,02 0,07	0,03 0,0005 0,03 0,05	- 0,003 - 0,006 0,01 0,03
6	- 7,8 1,3 1,2 - 8,7	0,44 0,93 1,1 - 0,6	0,07 0,06 - 0,04 0,07	0,005 - 0,03 - 0,009 0,02	0,001 - 0,004 0,006 - 0,03
7	0,19 - 7,6 7,9 - 0,32	- 1, 0,61 0,53 1,	- 0,08 0,02 0,03 - 0,1	- 0,009 0,005 0,01 0,02	
8	7, - 1,2 - 1,3 7,9	- 0,26 - 0,94 - 1,1 0,69	0,06 - 0,09 0,02 0,12		
9	- 0,38 6,7 - 6,9 - 0,03	0,95 - 0,36 - 0,41 - 1,			
10	- 6,1 0,7 1,4 - 6,7				

C.7.2. - Example 2 - Bivariate AR(1).

$$A_0 = \begin{pmatrix} -0,2 & -0,3 \\ 0,6 & -1,1 \end{pmatrix} \quad \Sigma = \begin{pmatrix} 4 & 1 \\ 1 & 1 \end{pmatrix}$$

The matrix ϵ -algorithm has been applied to the sequence $(-1)^s \hat{R}(s)$,

$- 2 \leq s \leq 10$.

column line	0	$\bar{p} = 2$	4	6	8
- 2	0,22 - 0,16	3,3 8,2			
- 1	- 2,1 - 1,	- 3,4 - 9,3	0,078 - 0,011	0,058 0,014	
$\bar{q} = 0$	5,1 3,5	3,5 10,	1,4 0,61	0,61 0,43	0,87 0,36
1	- 2,1 - 3,4	- 1, - 9,3	0,078 0,058	- 0,011 0,014	- 0,28 - 0,099
2	0,92 3,3	- 0,16 8,2	- 0,17 - 0,054	- 0,051 - 0,006	- 0,12 - 0,029
3	- 0,66 - 3,2	0,71 - 6,9	- 0,11 - 0,004	- 0,035 0,021	- 0,12 - 0,009
4	0,076 2,9	- 1,2 5,8	- 0,15 - 0,008	- 0,013 0,016	- 0,11 - 0,0078
5	- 0,14 - 2,6	1,5 - 4,5	- 0,087 - 0,023	0,017 0,028	0,06 0,14
6	- 0,14 2,2	- 1,6 3,5	0,11 0,088	0,068 0,064	- 3,3 - 1,5
7	0,62 - 1,5	1,8 - 2,5	0,23 0,046	0,07 - 0,01	0,11 0,018
8	- 1,8 1,2	- 1,6 1,6	0,1 - 0,007	- 0,046 - 0,035	0,16 0,038
9	0,19 - 0,97	1,4 - 1,1	0,11 0,064	0,044 0,037	
10	0,25 1,1	- 1,1 0,78			

REFERENCES.

- AKHIEZER, N.I. : The classical moment problem, Oliver and Boyd, Edinburgh, 1965.
- ANDERSON, T.W. : The Statistical Analysis of Time Series - Wiley, New-York, 1971.
- ANSLEY, C.F., W.A. SPIVEY and W.J. WROBLESKI : On the structure of moving average processes - Journal of Econometrics 6, 1977, 121-134.
- BÉGUIN, J.M., C. GOURIÉROUX and A. MONFORT : Identification of a mixed autoregressive-moving average process : the corner method - Time series, Anderson, O.D., ed., North-Holland, Amsterdam, 1980, 423-436.
- BERLINET, A. : Une méthode de détermination des degrés d'un modèle ARMA - Publications IRMA, Vol. 3, Fasc. 6, Lille, 1981 Exposé aux Journées de Statistique de Bruxelles (24-27 mai 1982).
- BERLINET, A. : Degrees of an ARMA model : estimation and tests. Compstat - Toulouse 1982 (a).
- BERLINET, A. : Estimating the degrees of an ARMA model. Prépublication, 1982 (b).
- BOX, G.E.P. and G.M. JENKINS : Time Series Analysis - Forecasting and Control Holden-Day, San Francisco, 1976.
- BOX, G.E.P. and G.C. TIAO : Modeling Multiple - Time Series with Applications. JASA, Vol. 76, Numb. 376, 1981, 802-816.
- BRÉZINSKI, C. : Accélération de la convergence en Analyse Numérique - Lecture Notes in Math. N° 584, Springer-Verlag, Berlin, 1977.
- BRÉZINSKI, C. : Algorithmes d'accélération de la convergence - Etude Numérique. Technip, Paris, 1978.
- BRÉZINSKI, C. : Padé - Type Approximation and General Orthogonal Polynomials, Birkhauser, Basel, 1980.
- CORDELLIER, F. : Particular rules for the vector ϵ -algorithm. Numer. Math. 27, 1977, 203-207.
- DEISTLER, M. : Parametrization and consistent estimation of ARMA systems. Time Series, Anderson, O.D., ed. North-Holland, Amsterdam, 1980, 373-385.
- DRAUX, A. : Polynômes orthogonaux formels. Applications. Thesis, Lille, 1981 to appear in Lecture Notes in Mathematics, Springer, Heidelberg, 1983.
- DRAUX, A. : Polynômes orthogonaux formels dans une algèbre non commutative. Publ. AN092, Université de Lille, 1983 (to appear).

- GRAY, H.L., G.D. KELLEY and D.D. MAC INTIRE : A new approach to ARMA modeling
Comm. in Stat. Simul. and Comp., B7, 1978, 1-77.
- HANNAN, E.J. : The identification of vector mixed autoregressive-moving
average systems - Biometrika, 56, 1969, 223-225.
- HANNAN, E.J. : Multiple Time Series - Wiley, New-York, 1970.
- HANNAN, E.J. : The asymptotic distribution of serial covariances.
The Annals of Statistics 4, 1976, 396-399.
- LIU, L.M. and D.M. HANSSENS : Identification of multiple-input transfer
function models - Comm. Stat. - Theor. Meth. 11, 1982,
297-314.
- PRIESTLEY, M.B. : Spectral Analysis and Time Series - Vol. 1 and 2
Academic Press 1981.
- ROZANOV, YU. A. : Stationary Random Processes.
Holden-Day, San Francisco, 1967.
- WYNN, P. : On a device for computing the $e_m(S)_n$ transformation -
MTAC, 10, 1956.
- WYNN, P. : Continued fractions whose coefficients obey a non commutative
law of multiplication. Arch. Rat. Mech. Anal., 12, 1963,
273-312.
- WYNN, P. : Upon systems of recursions which obtain among the quotients
of the Padé table - Num. Math. 8, 1966, 264-269.

ANNEXE 2

I - MOYENNES MOBILES MULTIVARIEES. - (Extrait de Pub. IRMA - Lille -
Vol. 4 - Fasc. 3 - 1982).

1°) Notations complémentaires.

Soit H_X le sous-espace fermé de $L^2(\text{Pr})$ engendré par
 $(X_{k,t})_{\substack{k \in I \\ t \in \mathbb{Z}}}$; pour $t \in \mathbb{Z}$ soit $H_X(t)$ le sous-espace fermé de $L^2(\text{Pr})$
engendré par $(X_{k,s})_{\substack{k \in I \\ s \leq t}}$ et $D_X(t)$ l'espace d'innovation de X à l'instant
 t (supplémentaire orthogonal de $H_X(t-1)$ dans $H_X(t)$).

On note $S_X = \bigcap_{t \in \mathbb{Z}} H_X(t)$.

Soit $(U_t)_{t \in \mathbb{Z}}$ la famille unitaire associée à X :

$\forall t \in \mathbb{Z}$ U_t est l'unique opérateur unitaire de H_X tel que

$$\forall s \in \mathbb{Z} \quad \forall k \in I \quad U_t X_{k,s} = X_{k,t+s} .$$

2°) Théorème (théorème 7 du chapitre 6). - Soit $q \in \mathbb{N}^*$ et Y un
élément de $\pi_2(\mathbb{Z})$ dont les matrices de covariances sont notées $R_Y(t)$, $t \in \mathbb{Z}$.

Une condition nécessaire et suffisante pour que Y soit une
moyenne mobile de degré minimal q est que

$$\begin{cases} R_Y(t) = 0 & , \quad \forall t > q \\ R_Y(q) \neq 0 & . \end{cases}$$

Remarque. - On peut démontrer ce théorème en utilisant des méthodes
spectrales. Nous donnons ici une démonstration basée sur la décomposition de Wold.

a) Condition suffisante.

Y étant stationnaire, il admet la décomposition unique suivante [Rozanov, 1967] :

$$(1) \quad \forall t \in \mathbb{Z} \quad Y(t) = \eta(t) + \zeta(t)$$

où $\eta(\cdot) = (\eta_k(\cdot))_{k \in I}$ et $\zeta(\cdot) = (\zeta_k(\cdot))_{k \in I}$ sont subordonnés à Y :

$$\forall t \in \mathbb{Z} \quad H_\eta(t) \subset H_Y(t) \quad [H_\eta(t)]^{\perp H_\eta} \subset [H_Y(t)]^{\perp H_Y}$$

$$\text{et} \quad H_\zeta(t) \subset H_Y(t) \quad [H_\zeta(t)]^{\perp H_\zeta} \subset [H_Y(t)]^{\perp H_Y}$$

η et ζ sont mutuellement non corrélés ($E \eta_k(t) \zeta_\ell(s) = 0$, $\forall (k, \ell) \in I^2, \forall (s, t) \in \mathbb{Z}^2$).

η est linéairement régulier ($S_\eta = \{0\}$) et ζ est linéairement singulier ($S_\zeta = H_\zeta$; ce qui entraîne $[H_\zeta(t)]^{\perp H_\zeta} = \{0\}$ et l'inclusion évidente de cet ensemble dans $[H_Y(t)]^{\perp H_Y}$, $\forall t \in \mathbb{Z}$).

On peut remarquer que $\eta \neq 0$; dans l'hypothèse contraire on aurait $Y = \zeta$ et donc $S_Y = H_Y$, or l'hypothèse $R_Y(t) = 0 \quad \forall t > q$ entraîne que $Y_k(t)$ est orthogonal à S_Y et par conséquent que $Y_k(t) = 0 \quad \forall t \in \mathbb{Z} \quad \forall k \in I$ ce qui n'est pas puisque $R_Y(q) \neq 0$.

Comme η est régulier il admet une décomposition de Wold unique

$$(2) \quad \eta_k(t) = \sum_{s=-\infty}^t \sum_{j=1}^m c_{kj}(t-s) \xi_j(s)$$

où $c_{kj}(t) = E(\eta_k(t) \xi_j(0))$ et ξ est un processus fondamental pour η (tel que $\forall s \quad (\xi_j(s))_{1 \leq j \leq m}$ est une base orthonormée de $D_\eta(s)$).

(1) et (2) donnent la décomposition de Wold de Y .

On a $H_\eta(t-1) \oplus D_\eta(t) = H_\eta(t) \subset H_Y(t)$ (η étant régulier $D_\eta(t) \neq \{0\}$).

$\forall k \in I$ $Y_k(t+s)$ est orthogonal à $H_Y(t)$ dès que $s > q$ donc $[\eta_k(t) + \zeta_k(t)]$ est orthogonal à $D_\eta(0)$ pour $t > q$ et $c_{kj}(t) = 0$ pour $t > q$.

Par conséquent :

$$Y_k(t) = \sum_{s=t-q}^t \sum_{j=1}^m c_{kj}(t-s) \xi_j(s) + \zeta_k(t).$$

Comme $H_Y(t) = H_\eta(t) \oplus H_\zeta(t)$

on a $S_Y = S_\eta \oplus S_\zeta$

et donc $S_Y = S_\zeta = H_\zeta$

$\zeta_k(t) \in S_Y$ et $\eta_k(t) = Y_k(t) - \zeta_k(t)$ est orthogonal à S_Y donc

$$E([\zeta_k(t)]^2) = E(\zeta_k(t) Y_k(t)) = 0 \text{ car } Y_k(t) \text{ est orthogonal à } S_Y.$$

Donc $Y = \eta$ et

$$Y_k(t) = \sum_{s=0}^q \sum_{j=1}^m c_{kj}(s) \xi_j(t-s).$$

Soit $e_o(t)$ le vecteur de \mathbb{R}^n défini par

$$e_o(t) = (\xi_j(t))_{1 \leq j \leq n} \quad \text{si } m = n$$

$$e_o(t) = (\xi_1(t), \xi_2(t), \dots, \xi_m(t), 0, \dots, 0)' \quad \text{si } m \neq n.$$

On a alors si $t \neq s$ $E(e_o(t) e_o'(s)) = 0$

$$E(e_o(t) e_o'(t)) = I_n \text{ si } m = n$$

$$= \begin{pmatrix} I_m & 0 \\ 0 & 0 \end{pmatrix} \text{ si } m < n.$$

On pose $C(s) = (c_{kj}(s))_{\substack{k \in I \\ 1 \leq j \leq m}}$.

La matrice $C(0) = E(Y_k(0) \xi_j(0))_{\substack{1 \leq k \leq n \\ 1 \leq j \leq m}}$ est la matrice des

coordonnées de $(Y_k(0))_{1 \leq k \leq n}$ dans la base $(\xi_j(0))_{1 \leq j \leq m}$. Elle est donc de rang m .

Si $m = n$ on pose $C_1(0) = C(0)$ et $C_1(0)$ est une matrice régulière.

Si $m < n$ on peut trouver $(m-n)$ matrices $(n \times 1)$:

V_1, V_2, \dots, V_{m-n} telles que :

$C_1(0) = (C(0) \begin{matrix} \vdots \\ V_1 \\ \vdots \\ V_2 \\ \vdots \\ \dots \\ \vdots \\ V_{n-m} \end{matrix})$ soit une matrice

régulière.

Pour $1 \leq i \leq q$ on pose $C_1(i) = C(i)$ si $m = n$

$$\left. \begin{aligned} (C_1(i))_{kl} &= c_{kl}(i) && \text{si } 1 \leq l \leq m \\ & && k \in I \\ &= 0 && \text{sinon} \end{aligned} \right\} \text{ si } m < n$$

et $B(i) = C_1(i) \times [C_1(0)]^{-1}$,

on a alors $\forall t \in \mathbb{Z} \quad Y(t) = e(t) + \sum_{i=1}^q B(i) e(t-i)$

où $\forall t \in \mathbb{Z} \quad e(t) = C_1(0) e_0(t)$.

e est bien un bruit blanc multivarié :

$$E(e(t) e'(s)) = E(C_1(0) e_0(t) e_0'(s) C_1'(0))$$

$$= 0 \quad \text{si } t \neq s$$

$$= C_1(0) C_1'(0) \quad \text{si } t = s .$$

$$\begin{aligned} \text{D'autre part } C_1(0) C_1'(0) B'(q) &= C_1(0) C_1'(0) [C_1'(0)]^{-1} C_1'(q) \\ &= C_1(0) C_1'(q) \end{aligned}$$

ce produit n'est pas nul car $C(0) C'(q)$ est égal à $R_Y(q) \neq 0$ et Y est une moyenne mobile de degré minimal q .

b) Condition nécessaire.

Si Y est une M.A.(q).

$$\text{On a } \forall t \in \mathbb{Z} \quad Y_t = \xi_t + \sum_{i=1}^q B_i \xi_{t-i},$$

où $\xi = (\xi_t)_{t \in \mathbb{Z}}$ est un bruit blanc multivarié de matrice de variances-covariances $\Sigma \neq 0$.

$$\text{On a alors } R_X(Y) = E(Y_0 Y_t') = 0 \quad \text{si } t > q$$

$$\text{et } R_X(q) = E(\xi_0 \xi_0' B_q') = \Sigma B_q'.$$

Si $\Sigma B_q'$ était nul, d'après la démonstration de la condition suffisante, Y serait nul (impossible puisque $E(Y_t \xi_t') = \Sigma$) ou serait une M.A.(q') avec $q' < q$, ce qui contredirait la minimalité de q . ■

II - COMPLEMENT SUR L'EPSILON-ALGORITHME MATRICIEL.-

La propriété fondamentale de l'epsilon-algorithme scalaire appliqué à une suite satisfaisant une équation aux différences a été étendue, sous certaines conditions, à l'epsilon matriciel (Draux, A., Pub. ANO Lille, 115, Théorèmes 4.3.18, 4.3.19 et 4.3.20). Sous les mêmes conditions on peut donc établir la réciproque du théorème 10 en utilisant le théorème 9.

III - AUTOREGRESSIONS MULTIVARIEES ET POLYNOMES ORTHOGONAUX GENERALISES.-

Supposons qu'il existe un processus autorégressif $(Y_t)_{t \in \mathbb{Z}}$ de degré p dont la suite des matrices de covariance coïncide avec celle de $(X_t)_{t \in \mathbb{Z}}$ sur $\{0, 1, \dots, p\}$. Le processus $(Y_t)_{t \in \mathbb{Z}}$ est tel qu'il existe des matrices M_1, M_2, \dots, M_p ($M_p \neq 0$) pour lesquelles $\det(I_n + \sum_{i=1}^p M_i z^i)$ ne s'annule pas pour $|z| \leq 1$ et un bruit blanc $(\epsilon_t)_{t \in \mathbb{Z}}$ tels que

$$(3) \quad \forall t \in \mathbb{Z} \quad Y_t + \sum_{i=1}^p M_i Y_{t-i} = \epsilon_t .$$

Le processus $(Y_t)_{t \in \mathbb{Z}}$ est l'ajustement autorégressif de degré p au processus $(X_t)_{t \in \mathbb{Z}}$ et la suite Γ de ses matrices de covariance vérifie les équations de Yule-Walker :

$$(4) \quad \Gamma(j) + \sum_{i=1}^p M_i \Gamma(j-i) = 0, \quad 1 \leq j \leq p .$$

En mettant (4) sous la forme

$$\Gamma(p-j) + \sum_{i=1}^p M_i \Gamma(p-i-j) = 0, \quad 0 \leq j \leq p-1 .$$

et en posant $\varphi_p(x) = I_n + \sum_{i=1}^p M_i x^i$, on se rend compte que les équations de Yule-Walker sont équivalentes à

$$(5) \quad c_p(x^j \varphi_p(x)) = 0, \quad 0 \leq j \leq p-1$$

où c_p est l'application linéaire à valeurs dans l'algèbre M_n des matrices carrées réelles d'ordre n et définie sur $M_n[X]$, anneau des polynômes à une indéterminée à coefficients dans M_n , par

$$c_p[x^i] = \Gamma(p-i), \quad i \in \mathbb{N} .$$

Sous l'hypothèse que la matrice

$$N_p = \begin{pmatrix} \Gamma(-p+2) & \Gamma(-p+3) & \dots & \Gamma(0) \\ \Gamma(-p+3) & \dots & \dots & \Gamma(1) \\ \vdots & \vdots & \vdots & \vdots \\ \Gamma(1) & \dots & \dots & \Gamma(p-1) \end{pmatrix}$$

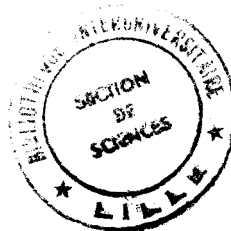
est inversible, il existe un unique polynôme $W_p^{(d)}$ I_n -unitaire vérifiant (5) : c'est le polynôme de degré p orthogonal à droite par rapport à c_p et dont le coefficient du terme de plus haut degré est I_n .

Si on suppose M_p inversible on a alors les relations :

$$M_p^{-1} \psi_p(x) = W_p^{(d)}(x)$$

et
$$\psi_p(x) = [W_p^{(d)}(0)]^{-1} W_p^{(d)}(x)$$

Sous les hypothèses habituelles de calculabilité, les généralisations des algorithmes qd et $\tilde{q}d$ permettent donc la détermination de ψ_p et celle de l'autocorrélation partielle généralisée.



RÉSUMÉ

Ce travail comprend quatre parties.

La première est consacrée à l'étude de certaines propriétés des paramètres relatives à l'espace des observations, aux statistiques adaptées à ces propriétés et aux estimateurs que l'on peut en déduire. Lorsque l'on évoque des propriétés locales d'un paramètre défini sur un ensemble de probabilités \mathcal{P} et à valeurs dans un espace topologique \mathcal{M} , le qualificatif "locales" a trait, en général, soit à l'ensemble \mathcal{P} et aux voisinages définis dans \mathcal{P} pour une certaine topologie, soit à l'espace \mathcal{M} . Nous nous intéressons ici à des propriétés locales dans un troisième ensemble : celui sur lequel les lois de \mathcal{P} sont définies. Les concepts introduits conduisent tout naturellement à étudier les statistiques ayant de bonnes propriétés correspondantes afin de définir de nouveaux estimateurs ou d'améliorer des estimateurs connus en utilisant le conditionnement par rapport à ces statistiques.

Dans la seconde partie nous reprenons des résultats que nous avons obtenus antérieurement sur les estimateurs splines de la densité afin d'établir leur normalité asymptotique dans le cas de subdivisions uniformes.

La troisième partie est consacrée à la fonctionnalisation d'un résultat d'approximation forte de Philipp et Pinzur. Nous montrons que l'on peut construire un processus de Kiefer généralisé, indexé par $G \times [1, +\infty[$ où G est un ensemble de fonctions sur $[0, 1]^d$, approximant presque sûrement les intégrales stochastiques des éléments de G par rapport au processus empirique associé à des variables aléatoires à valeurs dans $[0, 1]^d$ fortement mélangeantes et strictement stationnaires.

Dans la dernière partie nous étudions le problème de l'estimation des degrés d'un modèle ARMA à partir d'un estimateur de la fonction d'autocorrélation. Nous introduisons une méthode basée sur un algorithme d'accélération de convergence : l' ϵ -algorithme. Nous montrons ses liens avec d'autres méthodes et les tests numériques et statistiques qui en découlent. L'estimation de l'autocorrélation partielle et des polynômes d'autorégression est également étudiée. Nous traitons le cas de séries chronologiques uni ou multivariées.

MOTS CLÉS : - LOCALISATION, HOMOGENEITE, NON-PARAMETRIQUE
EXHAUSTIVITE, DENSITE, REGRESSION, NOYAUX OPTIMAUX.
- SPLINES, CONVERGENCES STOCHASTIQUES, LOIS LIMITES.
- APPROXIMATION FORTE, PROCESSUS DE KIEFER, FORTEMENT
MELANGEANT, FORTEMENT STATIONNAIRE.
- SERIES TEMPORELLES, MODELES ARMA, UNIVARIE, MULTIVARIE,
EPSILON-ALGORITHME .