ANNALES DE L'I. H. P., SECTION B

ÉLISABETH GASSIAT

Estimation semi-paramétrique d'un modèle autorégressif stationnaire multiindice non nécessairement causal

Annales de l'I. H. P., section B, tome 26, nº 1 (1990), p. 181-205
<http://www.numdam.org/item?id=AIHPB_1990__26_1_181_0>

Ann. Inst. Henri Poincaré,

Vol. 26, n° 1, 1990, p. 181-205. *Probabilités et Statistiques*

Estimation semi-paramétrique d'un modèle autorégressif stationnaire multiindice non nécessairement causal

par

Élisabeth GASSIAT

Université de Paris-Sud, U.A. n° 743, C.N.R.S., Statistique appliquée, Mathématiques, bat. n° 425, 91405 Orsay Cedex, France

RÉSUMÉ. — Nous étudions le problème d'estimation du paramètre θ d'un modèle de processus autorégressif stationnaire multiindice dans le cas où l'on ne fait sur θ aucune hypothèse. La loi du bruit est supposée inconnue (déconvolution aveugle). En particulier on ne suppose pas que le bruit blanc associé soit l'innovation (modèle non nécessairement causal). Nous proposons une méthode d'estimation qui résoud complètement le problème si la loi du bruit blanc n'est pas gaussienne (dans ce cas le problème est sans solution). Nous étudions la consistance et la loi asymptotique des estimateurs ainsi que leur robustesse. Nous proposons une estimation de l'ordre du modèle.

Mots clés : Champs linéaires, estimation semi-paramétrique, déconvolution, système non causal, minimum de contraste.

ABSTRACT. — We study the problem of the estimation of the parameter θ of a model of an autoregressive process of order p with multiindexation, when there is no assumption on θ. The law of the associated white noise is unknown (blind deconvolution). In particular, we don't make the assumption that this white noise is the innovation of the process. We propose an estimation method that solves completely the problem if the law of the noise is not gaussian (otherway the problem has no solution).

Classification A.M.S. : 62 M 10, 62 G 05, 86 A 15, 62 M 09, 93 E 12, 62 F 35.

We study consistency and asymptotic law of these estimates, and their robustness. We propose an estimate for the order of the model.

I. INTRODUCTION

Nous considérons un processus stochastique $Y=(Y_t,\ t\in\mathbb{Z}^d)$ stationnaire autorégressif d'ordre $p=(p_1,\ \dots,\ p_d)$ satisfaisant la relation :

$$\text{(1)} \qquad \sum_{k=0}^{p} b_k Y_{t-k}=X_t, \qquad t\in\mathbb{Z}^d,$$

où $X=(X_t,\ t\in\mathbb{Z}^d)$ est une suite de variables aléatoires réelles i.i.d. dont la loi est inconnue et qui n'est pas nécessairement l'innovation de Y.

NOTATIONS. — Pour un élément t de $\mathbb{Z}^d$, nous utiliserons alternativement la notation t ou $(t_1,\ \dots,\ t_d)$.

Une série indexée par $\mathbb{Z}^d$ sera notée en majuscule pour les processus, en minuscule pour les filtres :

ainsi Z désigne le processus (Z_t), $t\in\mathbb{Z}^d$, et a désigne le filtre (a_t), $t\in\mathbb{Z}^d$.

De façon générale, on notera $Z=a*U$ le processus défini par :

$$Z_t=\sum_{k\in\mathbb{Z}^d} a_k U_{t-k}.$$

Les hypothèses suivantes seront supposées vérifiées :

(M0) La loi commune F des X_t n'est pas gaussienne.

(M1) Le polynôme P_b n'a pas de racines sur le tore $|z_1|=\dots=|z_d|=1$.

P_b étant le polynôme défini sur $\mathbb{C}^d$ par :

$$P_b(Z_1,\ \dots,\ z_d)=\sum_{k=0}^{p} b_k z_1^{k_1}\dots z_d^{k}.$$

Remarques sur les hypothèses. — Il est connu que, sans aucune connaissance sur s, le problème d'identification de s n'a pas de solution lorsque la loi F est gaussienne. En effet, si $Y=s*X$ et $Z=u*X$ pour deux filtres distincts s et u, les processus Y et Z ne diffèrent que par leur fonction d'autocovariance, ce qui ne suffit pas à déterminer s.

- (M1) est la condition sous laquelle l'équation (1) possède une solution stationnaire.

La relation (1) s'inverse alors en :

$$\text{(2)} \qquad Y_t = \sum_{k \in \mathbb{Z}^d} s_{t-k} X_k, \qquad t \in \mathbb{Z}^d,$$

et le filtre $(s_k,\ k \in \mathbb{Z}^d)$ n'est pas nécessairement causal, c'est-à-dire n'a pas nécessairement de direction où il est porté par une demi-droite.

On observe Y_t, $t \in T_n$, où $T_n = \{t \in \mathbb{Z}^d / \forall\, i \leqq d,\ 1 \leqq t_i \leqq n\}$.

Nous nous intéressons au problème de l'estimation du paramètre $b = (b_k)$. L'étude des modèles autorégressifs fait l'objet d'une abondante littérature ([1], [4], [9], [16] par exemple). Cependant le cas essentiellement étudié est celui où le modèle est écrit en fonction de l'innovation du processus. Autrement dit, dans ces études, on suppose toujours que les racines du polynome $\Sigma\, b_k z^k$ sont situées à l'extérieur du disque unité, ce qui est une restriction essentielle sur le modèle. En revanche, le cas général où l'on ne fait pas cette hypothèse de causalité a été peu étudié théoriquement, bien qu'il soit important pour de nombreuses applications (géophysique, télécommunications, imagerie). Les premiers résultats théoriques précis sont dus d'une part à Ruget et coll. [2] qui proposent un algorithme stochastique permettant l'identification du filtre $s = (s_k)$, d'autre part à Rosenblatt et coll. [14] qui proposent une méthode d'estimation de la fonction de transfert du système.

Nous développons ici une méthode d'estimation proposée initialement par Donoho [6]. Nous considérons des estimateurs de minimum de contraste : nous caractérisons les contrastes utilisables, et étudions la consistance, la loi et l'efficacité asymptotique des estimateurs. Ceci permet de préciser le rôle de la causalité.

Nous étudions ensuite deux types de robustesse de la méthode : lorsque l'on a sous-estimé l'ordre du modèle (nous proposons un estimateur de l'ordre du modèle), et lorsque les observations sont bruitées. Enfin, la méthode d'estimation développée par Ruget et coll. étant apparentée à la notre par l'utilisation d'un contraste, nous l'interprétons à partir de nos résultats.

II. DESCRIPTION DE LA MÉTHODE, CONSISTANCE DE L'ESTIMATEUR

Pour résoudre le problème, nous considérons les processus stationnaires construits sur les Y_t : $X(c) = (X(c)_t)$ avec $X(c)_t = \Sigma_k\, c_k Y_{t-k}$, où c varie dans l'espace $\mathbb{R}^{p(d+1)}$, et l'on cherche à restituer $X(b) = X$. Nous proposons d'estimer b en minimisant une fonctionnelle $J(c)$ qui atteint son minimum unique au point $c = b$.

Rappelons tout d'abord quelques résultats relatifs aux estimateurs de minimum de contraste [5], et comment ils s'appliquent à notre problème.

Considérons un modèle statistique $(\Omega, \mathscr{A}, (P_\theta)_{\theta \in \mathbb{K}}$. Une fonction de contraste de ce modèle relativement à θ est une fonction réelle $K(\theta, \alpha)$ de la variable α qui a un minimum strict pour $\alpha = \theta$. Lorsque les observations sont Y_t, $t \in T_n$, et si $\mathbb{F}_t$ est la filtration définie par : $\mathbb{F}_t = \sigma(\{Y_u / \forall i \leqq d, u_i \leqq t_i\})$, un contraste relativement à θ et à K est une fonction réelle $U : (\alpha, t, \omega) \to U(\alpha, t, \omega)$ qui vérifie les propriétés :

(*a*) Pour tout $(\alpha, t) \in \mathbb{K} \times \mathbb{Z}^d$, la variable aléatoire $U_t(\alpha) : \omega \to U_t(\alpha, \omega)$ est F_t-mesurable.

(*b*) $U_t(\alpha) \to K(\theta, \alpha)$ en probabilité sous P_θ pour $|t| = \min_i t_i \to \infty$. Un estimateur $\hat{\theta}_t$ du minimum de contraste est donné par :

$$U_t(\hat{\theta}_t) = \inf \{U_t(\alpha), \alpha \in \mathbb{K}\}.$$

On a alors le résultat :

THÉORÈME 1. — $\hat{\theta}_t$ *est un estimateur consistant de* θ (i.e. $\hat{\theta}_t \to \theta$ *en probabilité sous* P_θ) si les conditions suivantes sont vérifiées :

(C1) : $\alpha \to K(\theta, \alpha)$ *et* $\alpha \to U_t(\alpha, \omega)$ *sont continues.*

(C2) : *La convergence de* $U_t(\alpha)$ *vers* $K(\theta, \alpha)$ *est uniforme sur* $\mathbb{K}$.

Démonstration. — Soit $E_t(\varepsilon) = \{\sup_{\alpha \in} |U_t(\alpha) - K(\theta, \alpha)| < \varepsilon\}$. $P_\theta(E_t(\varepsilon)) \to 1$ pour $|t| \to \infty$ par (C2). $K(\theta, \alpha)$ est localement continûment inversible autour de $\alpha = \theta$ à cause de l'extrémalité stricte de $K(\theta, \theta)$.

Donc : $\forall \varepsilon > 0$, $\exists \eta > 0$ tel que $|K(\theta, \alpha) - K(\theta, \beta)| \leqq \eta \Rightarrow \|\alpha - \beta\| \leqq \varepsilon$.

Pour $\varepsilon > 0$ et sur $E_t(\eta/2)$ on a :

$$K(\hat{\theta}_t, \theta) - \eta/2 \leqq U_t(\hat{\theta}_t) \leqq U_t(\theta) \leqq K(\theta, \theta) + \eta/2$$
$$|K(\hat{\theta}_t, \theta) - K(\theta, \theta)| \leqq \eta,$$

donc

$$\|\theta - \hat{\theta}_t\| \leqq \varepsilon \quad \text{et} \quad P(\|\theta - \hat{\theta}_t\| \leqq \varepsilon) \geqq P(E_t(\eta/2)).$$

Remarque. — Lorsque $\mathbb{K}$ est compact, (C2) est équivalente, par le théorème de Prokhorov (*voir* [3]) à une condition sur les oscillations de $U_t(\alpha)$:

$$\forall \varepsilon > 0, \quad \forall \eta > 0, \quad \exists \delta > 0$$

tel que

$$P(\sup_{\|\alpha - \beta\| < \delta} |U_t(\alpha) - U_t(\beta)| > \eta) \leqq \varepsilon.$$

Dans notre problème, $\theta = b$; nous appellerons fonctions d'objectif J et objectifs J_n les opposés de fonctions de contraste K et les opposés de contrastes $U_{(n, \ldots, n)}$, et nos estimateurs seront donc estimateurs de maximum d'objectif.

Nous étudierons ici une classe très générale d'objectifs, ceux qui vérifient :

(O1) : J est une fonctionnelle définie sur un sous-ensemble V des probabilités sur $\mathbb{R}$. On suppose que pour tout c, $F(c)$, loi de $X(c)_t$ est dans V, ainsi que $F_n(c)$, loi empirique de $X(c)_t$ pour tout n. On identifie alors la fonction J du paramètre c à $J(F(c))$ et J_n à $J(F_n(c))$.

(O2) : J et J_n sont invariants par changement d'échelle, c'est-à-dire $J(\lambda c)=J(c)$ et $J_n(\lambda c)=J_n(c)$ pour tout réel non nul λ.

L'invariance par changement d'échelle est une condition naturelle dans la mesure où l'échelle du paramètre ne peut pas être identifiée, puisque si l'on multiplie S par λ et si l'on divise les X_t par λ, le processus observé Y ne change pas.

Si V est un ensemble de lois stable par convolution par une série sommable et changement d'échelle sur lequel J est défini, on a le résultat suivant :

THÉORÈME 2. — *J est une fonction d'objectif pour le modèle défini par* (1) *si et seulement si :*

Quelle que soit la loi F *dans* V, *quelle que soit la série sommable* (a_i), $i\in\mathbb{Z}$, *si les* (X_i), $i\in\mathbb{Z}$ *sont des variables aléatoires indépendantes de loi* F, *alors :*

(O3) $J(F_{\Sigma_i a_i X_i}) \leqq J(F_X)$, *avec égalité si et seulement si tous les* a_i *sont nuls sauf un.* (*On note* F_Z *la loi de la variable aléatoire* Z.)

Ce théorème est immédiat dès que l'on remarque que $F(c)=F_k(c*s)_k X_{-k}$. Il donne une caractérisation simple des fonctions d'objectif, caractérisation soulignant l'extrémalité de la loi gaussienne. On peut en effet définir une relation d'ordre partiel sur V, ordre pour lequel la gaussienne est le plus petit élément, et (O3) signifie alors que J est respecte l'ordre ainsi défini. Pour des détails et commentaires sur cette caractérisation, *voir* [6].

A cause de la condition d'invariance (O2), on choisit pour le paramètre des conditions de normalisation, et l'on définit deux espaces de paramètre :

$$\mathbb{T}=\{B\in\mathbb{R}^p/\Sigma_k b_k^2=1\},\qquad p=(p_1+1)\dots(p_d+1)$$

et

$$\mathbb{U}=\{B\in\mathbb{R}^p/b_{(0,\dots,0)}=1\}.$$

$b\in\mathbb{U}$ est la normalisation usuelle pour les modèles ARMA, $b\in\mathbb{T}$ est une normalisation par appartenance à un compact.

On a alors le résultat suivant :

THÉORÈME 2 bis. — *Considérant* J *et* J_n *qui satisfont les conditions* (O1), (O2), (O3), *et* (C1), (C2), *les estimateurs* $\hat{\hat{b}}_n$ *et* $\hat{b}_n$ *maximisant* $J_n(c)$ *respectivement sur* $\mathbb{T}$ *et* $\mathbb{U}$ *sont consistants.*

Exemples. — On peut montrer que les objectifs suivants vérifient (O1), (O2) et (O3) :

(E1) *Cumulants standardisés :*

$$J(F_X) = \left| \frac{C_m(X)}{(C_2(X))^{m/2}} \right|,$$

$m > 2$, que l'on définit sur l'espace des variables ayant des moments jusqu'à l'ordre m, où

$$C_m(X) = \left(-i \frac{d}{dt} \right)^m \operatorname{Log} E(\exp itX) \big|_{t=0}$$

(cumulant d'ordre m de X).

On démontre aisément que $U(\Sigma a_i X_i) = \left| \frac{\Sigma a_i^m}{(\Sigma a_i^2)^{m/2}} \right|$ et donc que U vérifie (O3) dès que le cumulant standardisé de X est non nul.

(E2) *Néguentropie :* $J(F_X) = K(X, \varphi)$ où K désigne l'information de Kullback et φ est la gaussienne centrée de même variance que X; on définit J sur l'espace des variables ayant un moment d'ordre 2 et dont la loi possède une densité d'entropie finie.

On peut montrer que (O3) est vérifiée sous l'hypothèse suivante :

(NEG) : $\exists p > 0$ tel que $|\varphi| \in L_p$, φ étant la fonction caractéristique de la variable X.

La vérification de (O3) requiert d'une part la décroissance par convolution finie de l'entropie, d'autre part un passage à la limite dont la démonstration est technique et que nous n'expliciterons pas pour ne pas alourdir l'article.

(E3) *Information de Fisher :* $J(F_X) = I(X) = \int (p')^2/p \, dx$, où p désigne la densité de la loi X. On peut montrer que (O3) est vérifiée sous l'hypothèse suivante (même remarque de démonstration que ci-dessus) :

(FISH) : $\exists \varepsilon > 0$ et $\exists p > 0$ tels que $t^\varepsilon \varphi$ soit élément de L_p.

(E4) *Métrique idéale :* Si μ_s est une métrique idéale d'ordre s, on peut définir : $J(F_X) = \mu_s(X, G)$, où G est une variable aléatoire gaussienne et où $s > 2$. (*Voir* [23] pour définition et propriétés des métriques idéales.)

Un exemple simple de métrique idéale est le suivant :

$$\mu_s(X, Y) = \sup_{f \in D(s, K)} |E(f(X) - f(Y))|,$$

où $s = m + \alpha$, $m \in \mathbb{N}$ et $\alpha \in]0, 1]$, et où $D(s, K)$ est l'ensemble des fonctions réelles f continues m fois dérivables dont la dérivée d'ordre m vérifie : $|f^{(m)}(x) - f^{(m)}(y)| \leqq K |x-y|^\alpha$.

Il s'agit ensuite de définir une suite de fonctions aléatoires J_n vérifiant (C1) et (C2). On peut choisir $J_n(c) = J(F_n(c))$. Dans le cas des cumulants, on fait ce choix. Il est ensuite facile de montrer que si X possède des moments finis jusqu'à l'ordre m, alors J_n converge p.s. uniformément vers J.

Pour les 3 derniers exemples, un tel choix n'est pas possible. En effet, J fait intervenir la loi inconnue de X. On notera $J(c) = J(p, c)$, p désignant la densité de la loi de X.

On peut alors envisager une procédure de type « adaptative » : Poser $J_n(c) = J(p_n, F_n(c))$, où p_n est une estimation régulière de la densité de X obtenue à l'aide d'un estimateur préliminaire de b.

Remarquons que nous considérons ici des processus de variance finie. Les résultats s'étendent en fait au cas où le processus n'a pas de moment d'ordre 2, nous étudions ce cas dans un autre travail [7].

III. COMPORTEMENT ASYMPTOTIQUE

III.1. Étude générale

Dans ce paragraphe, on suppose que J et J_n vérifient les conditions suffisantes de consistance pour $\hat{\hat{b}}_n$ et $\hat{b}_n$. Les résultats qui suivent valent pour tout $d > 0$.

Nous utiliserons les hypothéses suivantes :

(A0) : $J(c)$ et $J_n(c)$ sont deux fois continûment différentiables.

(A1) : $J_n(c) \xrightarrow{\text{p. s.}} J(c)$

$$\operatorname{Grad} J_n(c) \xrightarrow{\text{p. s.}} \operatorname{Grad} J(c)$$

$$D^2 J_n(c) \xrightarrow{\text{p. s.}} D^2 J(c) = A(c), \qquad A(b) = A.$$

(A2) : $n^{d/2} \operatorname{Grad} J_n(b) \xrightarrow{\mathscr{D}} N(0, \Lambda)$, où $\xrightarrow{\mathscr{D}}$ désigne la convergence en loi et $N(0, \Lambda)$ la gaussienne centrée de covariance Λ.

(A3) : $\exists\, r > 0\ tq\,(D^2 J_n(b+\alpha) - D^2 J_n(b)) \xrightarrow{P} 0$ uniformément pour $|\alpha| \leqq r$.

Remarque. — Ces hypothèses n'imposent rien sur la régularité de F, ni même sur l'existence d'une densité. Par exemple, lorsque J est fonction de φ-moments de F (en particulier lorsque J est un cumulant standardisé), F peut être quelconque sous réserve que les φ-moments intervenant soient finis.

LEMME 1. — *Si* (A0) *et* (A1), *alors* $\operatorname{Ker} A = \{\lambda b / \lambda \in \mathbb{R}\}$.

Démonstration. — Soit $\mathcal{D}=\{\lambda b/\lambda\in\mathbb{R}\}$. Pour $u\in\mathbb{R}$ et $c\notin\mathcal{D}$, soit $h(c)(u)$ la fonction donnée par $h(c)(u)=\mathrm{J}(b+uc)$. $h(c)'(u)=\langle \operatorname{Grad}\mathrm{J}(b+uc), c\rangle$, $h(c)''(u)={}^tc.\mathrm{A}(b+uc).c$. $h(c)$ a un maximum strict en $u=0$, donc: $h(c)'(0)=0$ et $h(c)''(0)\neq 0$, soit ${}^tc.\mathrm{A}.c\neq 0$ pour $c\notin\mathcal{D}$.

Prenant maintenant c dans $\mathcal{D}$: $h(\lambda b)$ est constante grace à (02), et $h(c)''$ est identiquement nulle, dont ${}^tc.\mathrm{A}.c=0$.

Pour les théorèmes qui vont suivre, nous aurons besoin de notations particulières :

A est diagonalisable dans une base orthonormée. Prenons $b/\|b\|$ comme premier vecteur de cette base. Dans cette base, soit D (resp. Δ) la transformée de A (resp. Λ), soit $\hat{c}_n$ le transformé de $\hat{b}_n$.

Si M est une matrice de $\mathbb{R}^{p^2}$, on notera (M) la matrice de $\mathbb{R}^{(p-1)^2}$ obtenue en retirant à M sa première ligne et sa première colonne. Pour un vecteur m, on notera (m) le vecteur obtenu en retirant à m sa première coordonnée.

On a alors les résultats suivants :

THÉORÈME 3. — *Sous* (A0)...(A3) :

$$n^{d/2}((\hat{c}_n))\xrightarrow{\mathcal{D}}\mathrm{N}(0,\Sigma),\qquad \textit{avec}\quad \Sigma=(\mathrm{D})^{-1}(\Delta)(\mathrm{D}).$$

THÉORÈME 4. — *Sous* (A0)...(A3) :

$$n^{d/2}((\hat{b}_n)-(b))\xrightarrow{\lambda}\mathrm{N}(0,\Sigma^*),\qquad \textit{avec}\quad \Sigma^*=(\mathrm{A})^{-1}(\Lambda)(\mathrm{A}).$$

Pour démontrer ces deux théorèmes, on écrit :

$$\operatorname{Grad}\mathrm{J}_n(c)=\operatorname{Grad}\mathrm{J}_n(b)+{}^t(c-b)\int_0^1 \mathrm{D}^2\,\mathrm{J}_n(b+u(c-b))\,du\,(\mathrm{T})$$

$\operatorname{Grad}\mathrm{J}_n(c)=0$ pour $c=\hat{\hat{b}}_n$ et pour $c=\hat{b}_n$, car grace à (O2), un extremum de J_n sur $\mathbb{T}$ ou sur $\mathbb{U}$ est un extremum sur tout l'espace. Le théorème 4 est une conséquence immédiate de (T), (A2) et du lemme 1.

Démonstration du théorème 3. — Réécrivons (T) dans la base choisie diagonalisant A :

(1) $$n^{d/2}\,{}^t(\hat{c}_n-c).\mathrm{D}_n=\mathrm{E}_n,\qquad \text{où}\quad c=(1,0,\ldots,0),$$

soit b écrit dans la nouvelle base.

(2) $$\mathrm{D}_n\xrightarrow{\mathrm{P}}\operatorname{Diag}(\lambda_i),\qquad \lambda_0=0\quad\text{et}\quad\lambda_j\neq 0\quad\text{pour } j\neq 0.$$

(3) $$\mathrm{E}_n\xrightarrow{\mathcal{D}}\mathrm{N}(0,\Delta)$$

(4) $$\Sigma_j\,\hat{c}_{nj}^2=1$$

(2) : par (A1) et (A3) on obtient facilement que

$$\int_0^1 \mathrm{D}^2\,\mathrm{J}_n(b+u(\hat{b}_n-b))\,du\xrightarrow{\mathrm{P}}\mathrm{A}.$$

Par changement de base on obtient le résultat.
(3) : (A2) écrit dans la nouvelle base.
(4) : car le changement de base est orthonormé.
On remarque que (4) peut se réécrire en : $(\hat{c}_0^n - 1)(\hat{c}_0^n + 1) + \sum_{j \neq 0} \hat{c}_j^{n2} = 0$.

Notons

$$\begin{gathered} v_n = n^{d/2}(\hat{c}^n - c) \\ \mathrm{A}(n, \mathrm{M}) = \{\omega / \| v_n \| > \mathrm{M}\} \\ \mathrm{B}(n, \mathrm{M}) = \{\| \mathrm{J}_n \| > \mathrm{M}\} \\ \mathrm{C}(n, \varepsilon) = \{\sup_{i,j} |(\mathrm{D}_n)_{i,j}| > \varepsilon\} \\ \mathrm{D}(n, \beta) = \{|\hat{c}_0^n - 1| > \beta\} \end{gathered}$$

On sait que $\forall \varepsilon, \beta \lim_{\mathrm{M} \to \infty} \overline{\lim_n} \mathrm{P}(\mathrm{B}(n, \mathrm{M})) = 0$, $\lim_n \mathrm{P}(\mathrm{C}(n, \varepsilon)) = 0$

$$\lim_n \mathrm{P}(\mathrm{D}(n, \beta)) = 0.$$

On en déduit facilement que $\forall \alpha > 0$, $\forall \varepsilon > 0$ et $\forall \beta > 0$,

$$\lim_{\mathrm{M}} \overline{\lim} \, \mathrm{P}(\mathscr{A}(n, \mathrm{M}, \alpha, \varepsilon, \beta)) = \lim_{\mathrm{M}} \overline{\lim} \, \mathrm{P}(\mathrm{A}(n, \mathrm{M}))$$

où $\mathscr{A}(n, \mathrm{M}, \alpha, \varepsilon, \beta) = \mathrm{A}(n, \mathrm{M}) \cap \mathrm{B}^c(n, \alpha\mathrm{M}) \cap \mathrm{C}^c(n, \varepsilon) \cap \mathrm{D}^c(n, \beta)$.

On fixe $\beta_0 < 2/p+1$, $\varepsilon_0 = 1/2(p+1) . \mathrm{Sup}_j |\lambda_j|$, $\alpha_0 = 1/4(p+1) . \inf_j |\lambda_j|$.

Soit $\omega \in \mathscr{A}(n, \mathrm{M}, \alpha_0, \varepsilon_0, \beta_0)$, j la coordonnée telle que $|v_{nj}| = \sup_i |v_{nu}|$. On a $|v_{nj}| \geqq \mathrm{M}/p+1$.

- Si $j \neq 0$ on écrit : $\sum_{i \neq j} v_{nu} \mathrm{D}_{nij} = -v_{nj} \mathrm{D}_{njj} + \mathrm{U}_{nj}$.

Dans $\mathrm{B}^c(n, \alpha\mathrm{M})$, $\|\mathrm{U}_n\| < \alpha\mathrm{M}$ donc $|\mathrm{U}_{nj}| < \alpha\mathrm{M}$

$$\left|\sum_{i \neq j} v_{ni} \mathrm{D}_{nij}\right| \geqq |v_{nj} \mathrm{D}_{njj}| - \alpha\mathrm{M} \geqq |v_{nj}|(|\lambda_j| - \varepsilon) - \alpha\mathrm{M}$$

car on est dans $\mathrm{C}^c(n, \varepsilon_0)$. D'autre part : $\left|\sum_{i \neq j} v_{ni} \mathrm{D}_{nij}\right| \leqq p . |v_{nj}| \varepsilon$.

On a donc :

$$|v_{nj}| . |\lambda_j|/2 \leqq |v_{nj}|(|\lambda_j| - (p+1)\varepsilon) \leqq \alpha\mathrm{M} \leqq |\lambda_j| . \mathrm{M}/4(p+1).$$

D'où $|v_{nj}| \leqq \mathrm{M}/2(p+1)$, ce qui est impossible.

- Si $j = 0$ on écrit $\sum_{j \neq 0} u_{ni}^2 = -u_{n0} . n^{d/2}(\hat{c}_0^n + 1)$.

Donc $p|u_{0n}|^2 \geqq |u_{0n}| . n^{d/2}(2 - \beta_0)$, ce qui donne $\beta_0 \geqq 2/p+1$, ce qui est contradictoire.

Donc $\mathrm{P}(\mathscr{A}(n, \mathrm{M}, \alpha_0, \varepsilon_0, \beta_0)) = 0$, donc $\overline{\lim_{\mathrm{M}}} \lim_n \mathrm{P}(\mathrm{A}(n, \mathrm{M})) = 0$, et $n^{d/2}(\hat{c}^n - c)$ est tendu. Le théorème s'en déduit immédiatement.

On donne maintenant un lemme très simple, qui sera utile plus tard pour simplifier quelques expressions de matrices.

LEMME 2. — *Si* (A0)...(A3), *alors* $b \in \operatorname{Ker} \Lambda$.

Démonstration. — Faire le produit scalaire de (T) avec B, multiplier par $n^{d/2}$, appliquer le lemme 1 et le théorème 1.

III.2. Cas particulier important : cas où J est fonction de φ-moments de F

Dans cette partie, on suppose que J et J_n s'écrivent :

$$J(c) = L(E\varphi_1(X(c)_0); \ldots; E\varphi_q(X(c)_0))$$
$$J_n(c) = L(1/n^d \Sigma_{t=(1,\ldots,1)}^{(n,\ldots,n)} \varphi_1(X(c)_t); \ldots; 1/n^d \Sigma \varphi_q(X(c)_t))$$

L et $\varphi = (\varphi_1, \ldots, \varphi_q)$ étant des fonctions réelles.

Remarque. — Dans le cas des cumulants, $J = |K|$ et $J_n = |K_n|$ où K et K_n sont de la forme précédente, mais il est très facile de voir que tous les résultats qui suivent sont encore valables.

Considérons les hypothèses suivantes :

(H1) φ et L sont deux fois continûment dérivables.

(H2) Les espérances des fonctions suivantes sont finies :

$$\varphi(X(c)_0),\quad \operatorname{Var}\varphi(X_0),\quad \varphi_k''(X(c)_0^\alpha.(X(c)_0),\quad X_0^\alpha \psi^2(X_0)\,(\alpha=0,1,2),$$
$$\varphi_k'(X(c)_0).(X(c)_0)^\beta,\quad \varphi_k(X_0)\psi(X_0)X_0^\beta\,(\beta=0,1),$$

et, si $d>1$ tout produit de 4 termes formés à partir des termes suivants :

$$X_0^\beta \psi(X_0) \quad \text{et} \quad \varphi_k(X_0),\quad \beta=0,1,\quad k=0,\ldots,q.$$

(H3) $E\psi(X_0)=0$ et $EX_0\psi(X_0)=0$.

(H4) $d>1$ et (Y_t), $t \in \mathbb{Z}^d$ est un champ fortement mélangeant.

Où ψ est donnée par :

$$\psi(u) = \sum_{k=1}^{q} \varphi_k'(u) D_k L(E\varphi(X(c)_0).$$

Remarque. — Pour les cumulants, (H2) est une hypothèse sur les moments de X_0.

Sous (H1), on a :

$$\operatorname{Grad} J_n(c)_i = 1/n^d \Sigma \psi_n(X(c)_t) Y_{t-i}$$
$$D^2 J_n(c)_{i,j} = 1/n^d \Sigma \psi_n'(X(c)_t) Y_{t-i} Y_{t-j} + 1/n^{2d} \Sigma \gamma_n(X(c)_t, X(c)_u) Y_{t-i} Y_{u-j}$$

avec

$$\psi_n(u) = \Sigma_1^q \varphi_k'(u) D^k L(E_n \varphi(X(C))$$
$$\gamma_n(u,v) = \Sigma_1^q \varphi_k'(u) \varphi_1'(v) D^k D^l L(E_n \varphi(X(C)).$$

ψ_n et γ_n ne sont pas indicées par c pour plus de lisibilité lorsque cela ne porte pas à confusion. E_n désigne l'espérance sous la loi empirique.

Par ergodicité, on obtient alors immédiatement (A1), (A0) est donnée par (H1), et pour (A3), la démarche est analogue à la consistance. En particulier, (A3) est vraie pour les cumulants, un calcul des oscillations le montrant facilement.

Le calcul donne alors :

$$A_{i,j}=\Sigma_1^q \mathrm{E}(\varphi_k''(X_0)Y_{-i}Y_{-j})\,D^k L(\mathrm{E}\,\varphi(X_0))$$
$$+\Sigma_1^q \mathrm{E}(\varphi_k'(X_0)Y_{-i})\,\mathrm{E}(\varphi_1'(X_0)Y_{-j})\,D^k D^l L(\mathrm{E}\,\varphi(X_0)).$$

On a le théorème :

THÉORÈME 3. — *Sous* ((H1)...(H3) *si* $d=1$ *ou* (H1)...(H4) *si* $d>1$) *on a :*

$$n^{d/2}\,\mathrm{Grad}\,J_n(b)\xrightarrow{\mathscr{D}} N(0,\Lambda).$$
$$\Lambda_{i,j}=\mathrm{E}(\psi^2(X_0)Y_{-i}Y_{-j})+\Sigma_1^q a_{k,i}\,\mathrm{E}(\varphi_k(X_0)\psi(X_0)Y_{-j})$$
$$+\Sigma_1^q a_{k,j}\,\mathrm{E}(\varphi_k(X_0)\psi(X_0)Y_{-i})+\Sigma_1^q a_{k,i}a_{l,j}\,\mathrm{Cov}(\varphi_k(X_0),\varphi_l(X_0))$$

où $a_{k,i}=\sum_{l=1}^{q} D^k D^l L(\mathrm{E}\,\varphi(X_0))\,.\,\mathrm{E}(\varphi_l'(X_0)Y_{-i})$.

De ce théorème on déduit la loi asymptotique pour $\hat{b}_n$ et $\hat{\hat{b}}_n$.

Remarques sur les hypothèses. — (1) Lorsque $d=1$, on n'a fait aucune hypothèse de régularité sur la loi de X.

(2) La condition de champ mélangeant n'a pas été étudiée pour des processus linéaires. Des conditions de mélange fort ont été données pour des processus linéaires indexés dans $\mathbb{Z}$ par Gorodetski [8] et Mokkadem [13].

Démonstration du théorème :

$$n^{d/2}\,\mathrm{Grad}\,J_n(b)=1/n^{d/2}\,\Sigma\,\psi(X_t)Y_{t-i}$$
$$+\Sigma_{k,l}\,a_{k,l,i}^n\,n^{d/2}(\mathrm{E}_n\varphi_l(X)-\mathrm{E}\,\varphi_l(X))$$
$$\times\left\{D^k D^l L(\mathrm{E}\,\varphi(X))+\int_0^1[D^k D^l L(\mathrm{E}\,\varphi+t(\mathrm{E}_n\varphi-\mathrm{E}\,\varphi))-D^k D^l L(\mathrm{E}\,\varphi)]\,dt\right\}$$

où

- $a_{kli}^n=1/n^d\,\Sigma\,\varphi_k'(X_t)Y_{t-i}\xrightarrow{P}\mathrm{E}\,\varphi_k'(X_0)Y_{-i}$.
- Le reste intégral tend en probabilité vers 0.

Il reste à montrer :

$$\begin{matrix} 1/n^{d/2}\,.\,\Sigma\,\psi(X_t)Y_t \\ \vdots \\ 1/n^{d/2}\,.\,\Sigma\,\psi(X_t)Y_{t-p} \\ n^{d/2}(\mathrm{E}_n\varphi_1(X)-\mathrm{E}\,\varphi_q(X)) \\ \vdots \\ n^{d/2}(\mathrm{E}_n\varphi_q(X)-\mathrm{E}\,\varphi_q(X)) \end{matrix}\xrightarrow{\mathscr{D}} N(0,\zeta).$$

On prend une combinaison linéaire des coordonnées du vecteur étudié multiplié par $n^{d/2}$.

$$Z_u = \Sigma_0^p \alpha_i (\Sigma_{t=(1, \ldots, 1)}^{(u1, \ldots, ud)} \psi(X_t) Y_{t-i}) + \Sigma_1^q \beta_i \Sigma_t (\varphi_i(X_t) - E\varphi_i(X))$$

Z_u est un processus stationnaire centré [par (H3)] sur $\mathbb{Z}^d$.

$$\begin{aligned} EZ_u^2 = {} & \Sigma \alpha_i \alpha_j \Sigma_{t1,t2} E[\psi(X_{t1}) \psi(X_{t2}) Y_{t1-i} Y_{t2-j}] \\ & + \Sigma \beta_i \beta_j \Sigma_{t1,t2} E[(\varphi_i(X_{t1}) - E\varphi_i(X_{t1})) . (\varphi_j(X_{t2}) - E\varphi_j(X_{t2}))] \\ & + \Sigma \alpha_i \beta_j \Sigma_{t1,t2} E[\psi(X_{t1}) Y_{t1-i} . (\varphi_j(X_{t2}) - E\varphi_j(X_{t2}))]. \end{aligned}$$

Lorsque $u = (u_1; \ldots; u_d)$ varie dans $\mathbb{Z}^d$ on obtient :

$$\begin{aligned} EZ_u^2 = {} & \prod_{i=1}^d u_i . [\Sigma \beta_i^2 \operatorname{Var} \varphi_i(X) + \Sigma \alpha_i^2 E(\psi^2(X) Y_{-i}^2) \\ & + \Sigma \alpha_i \beta_j E[\psi(X) Y_{-i} (\varphi_j(X) - E\varphi_j(X))] = \prod_{i=1}^d u_i . (\alpha, \beta) \xi^t (\alpha . \beta). \end{aligned}$$

ξ définie de manière évidente ci-dessus, est donc la variance limite du vecteur étudié lorsque n tend vers l'infini et que $u_i = n^d$.

1. *Lorsque* $d=1$. — On utilise le principe d'invariance de Mac-Leish [12]. Les hypothèses à vérifier sont les suivantes :

Si $\mathbb{F}_n$ est une séquence non-décroissante de sigma-algèbres, et si V_n est un processus tel que :

- Il existe une suite g_k positive tendant vers 0 telle que pour $i > 0$, $k \geqq 0$,

$$\| E^{i-k} V_i \|_2 \leqq g_k$$
$$\| V_i - E^{i+k} V_i \|_2 \leqq g_{k+1}$$

où E^i désigne l'espérance conditionnelle à la tribu $\mathbb{F}_i$.

- Il existe une suite positive non décroissante L_k telle que :

$$\Sigma 1/n L_n < \infty, \qquad L_n - L_{n-1} = O(L_n/n, \qquad g_n = o(1/\sqrt{n} . L_n).$$

- Les V_n sont uniformément intégrables.
- $E^{k-m}[(S_{n+k} - S_k)^2/n] \to \sigma^2$ quand $m, k, n \to \infty$.

Alors $S_n/\sqrt{n}$ tend en loi vers $N(0, \sigma)$, avec $S_n = \sum_{i=1}^n V_i$.

On prend $V_n = \Sigma_0^p \alpha_i \psi(X_n) Y_{n-i} + \Sigma_1^q \beta_i (\varphi_i(X_n) - E\varphi_i(X))$.

$\mathbb{F}_n$: sigma algèbre engendrée par les X_p, $p \leqq n$.

On a $S_n = Z_n$

- $E^{n-k}(V_n) = 0$.
- $\| E^{n+k}(V_n) - V_n \|^2 = \Sigma \alpha_i \alpha_j \sum_{l_1 < -i-k,\, l_2 < -j-k} s_{l_1} . s_{l_2}$

$$\times E(\psi(X_n^2) X_{n-i-l_1} X_{n-j-l_2})$$

on prend $g_{k+1}=(\mathrm{EX}^2 . \mathrm{E}\,\psi^2)^{1/2} . (\Sigma\,|\alpha_i|) . \sum_{l \geqq 0} |s_{-k-l}|$ qui décroît exponentiellement vers 0. En prenant $\mathrm{L}_n = \mathrm{Cte} . n$, les 2 premières hypothèses sont vérifiées.

- Le processus V_n étant stationnaire, les V_n^2 sont uniformément intégrables.
- Le processus étant stationnaire et les s_k décroissant exponentiellement vers 0, il est immédiat que la dernière hypothèse est vérifiée avec $\sigma^2 = (\alpha, \beta)\, \xi^t (\alpha, \beta)$.

On a donc la conclusion.

2. *Lorsque* $d > 1$. — On utilise un théorème de Rosenblatt ([14], p. 77) :

Soit $(\mathrm{Z}_n)\, n \in \mathbb{Z}^d$ un champ centré strictement stationnaire et fortement mélangeant tel que :

$$\mathrm{E}\left|\sum_{n_i=c_i}^{d_i} \mathrm{Z}_{n_1, \ldots, n_d}\right|^2 = h(d_1-c_1; \ldots; d_d-c_d)$$

où $h \to \infty$ quand $d_i - c_i \to \infty$, où $h(\alpha_1; \ldots; \alpha_d) = 0(h(\beta_1; \ldots; \beta_d))$ dès que $\alpha_i \to \infty\ \forall i$ avec $\alpha_i = 0(\beta_i)$ mais $\alpha_j = 0(\beta_j)$ pour un certain j.

Si $\exists \delta > 0$ tel que $\mathrm{E}\,|\Sigma\, \mathrm{Z}_n|^{2+\delta} = O(h(d_1-c_1; \ldots; d_d-c_d)^{1+\delta/2})$ alors $\sum_{ni=1}^{\mathrm{N}} \mathrm{Z}_n$ convenablement normalisé converge en loi vers $\mathrm{N}(0, \mathrm{Id})$ quand $d_i - c_i \to \infty\ \forall i$.

Ici, $h(\alpha_1; \ldots; \alpha_d) = \lambda\, \Pi\, \alpha_i$, qui vérifie bien les conditions requises. Nous allons vérifier la dernière condition pour $\delta = 2$.

$$\begin{aligned}
\mathrm{EZ}_t^4 = {} & \Sigma\, \alpha_{i1} \ldots \alpha_{i4}\, \Sigma\, \mathrm{E}(\psi(\mathrm{X}_{t1}) \ldots \psi(\mathrm{X}_{t4})\, \mathrm{Y}_{t1-i1} \ldots \mathrm{T}_{t4-i4} \\
& + \Sigma\, \beta_{i1} \ldots \beta_{i4}\, \Sigma\, \mathrm{E}(\varphi_{i1}(\mathrm{X}_{t1}) - \mathrm{E}\,\varphi_{i1}(\mathrm{X})) \ldots (\varphi_{i4}(\mathrm{X}_{t4}) - \mathrm{E}\,\varphi_{i4}(\mathrm{X})) \\
& + 4\, \Sigma\, \alpha_i\, \beta_{j1} \ldots \beta_{j3}\, \Sigma\, \mathrm{E}(\psi(\mathrm{X}_{t1})\, \mathrm{Y}_{t1-i} \\
& \times (\varphi_{j1}(\mathrm{X}_{t2}) - \mathrm{E}\,\varphi_{j1}(\mathrm{X})) \ldots (\varphi_{j3}(\mathrm{X}_{t4}) - \mathrm{E}\,\varphi_{j3}(\mathrm{X})) \\
& + 4\, \Sigma\, \alpha_{i1} \ldots \alpha_{i3}\, \beta_j\, \Sigma\, \mathrm{E}(\psi(\mathrm{X}_{t1})\, \mathrm{Y}_{t1-i1} \ldots \psi(\mathrm{X}_{t3}) \\
& \times \mathrm{Y}_{t3-i3}(\varphi_j(\mathrm{X}_{t4}) - \mathrm{E}\,\varphi_j(\mathrm{X})) + 6\, \Sigma\, \alpha_{i1}\, \alpha_{i2}\, \beta_{j1}\, \beta_{j2}\, \Sigma\, \mathrm{E}(\psi(\mathrm{X}_{t2})\, \mathrm{Y}_{t2-i2} \\
& \times \psi(\mathrm{X}_{t2})\, \mathrm{Y}_{t2-i2}(\varphi_{j1}(\mathrm{X}_{t3}) - \mathrm{E}\,\varphi_{j1}(\mathrm{X})) . (\varphi_{j2}(\mathrm{X}_{t4}) - \mathrm{E}\,\varphi_{j2}(\mathrm{X})).
\end{aligned}$$

Pour chaque terme, on regarde suivant que les t_i sont distincts ou non et en développant les Y sur les X : l'on obtient, sachant (H3) :

- $t_1, \ldots, t_4$ tous distincts : 0
- $t_1 = \ldots = t_4$: $\mathrm{Cte} . \Pi\,(d_i - c_i)$
- deux des t_i distincts : $\leqq \mathrm{Cte} . [\Pi\,(d_i - c_i)]^2$

• 3 des t_i distincts : seul le quatrième terme de la somme n'est pas nul, on obtient des termes de la forme :

$$4\Sigma\,\alpha_{i1}\ldots\alpha_{i3}\,\beta_j.\mathrm{Cte}.\mathrm{E}\,\psi^2(\mathrm{X}_0)$$
$$\times\,\mathrm{EX}_0^2\,\psi(\mathrm{X}_0).(\mathrm{EX}_0\,\varphi_j(\mathrm{X}_0)-\mathrm{EX}_0.\mathrm{E}\,\varphi_j(\mathrm{X}_0))$$
$$\times\,\Sigma\,s_{-i3}\,s_{t-t3-i2}\,s_{t-t4-i1}+s_{-i3}\,s_{t-t3-i1}$$
$$\times\,s_{t-t4-i2}+s_{t-t3-i1}\,s_{t-t3-i2}\,s_{t3-t4-i3}$$

qui est borné par :

$$\mathrm{C1}.|s_{-i3}|.(\Sigma\,|s_k|)^2.\Pi\,(d_i-c_i)+\mathrm{C2}.(\Sigma\,|s_k\,s_{k+i1-i2}|).(\Sigma\,|s_k|).\Pi\,(d_i-c_i).$$

On a donc $\mathrm{EZ}_t^4=O(\Pi\,(d_i-c_i)^2)$.

Maintenant, en appliquant les lemmes 1 et 2, on peut simplifier les expressions de A et Λ pour obtenir :

$$\mathrm{A}_{i,j}=\sum_{k\neq 0} s_{k-i}s_{k-j}\,\mathrm{Var}\,\mathrm{X}_0\,\mathrm{E}\,\psi'(\mathrm{X}_0)+(\sum_{k\neq 0} s_{k-i}).(\sum_{k\neq 0} s_{k-j}).(\mathrm{EX}_0^2)$$
$$\times\{\Sigma\,\mathrm{E}\,\varphi'_m(\mathrm{X}_0)\,\mathrm{E}\,\varphi_1(\mathrm{X}_0)\,\mathrm{D}^m\,\mathrm{D}^l\,\mathrm{L}(\mathrm{E}\,\varphi(\mathrm{X}_0))+\mathrm{E}\,\psi'(\mathrm{X}_0)\}$$
$$\Lambda_{i,j}=\sum_{k\neq 0} s_{k-i}s_{k-j}\,\mathrm{Var}\,\mathrm{X}_0\,\mathrm{E}\,\psi^2(\mathrm{X}_0)+(\sum_{k\neq 0} s_{k-i}).(\sum_{k\neq 0} s_{k-j}).(\mathrm{EX}_0^2)$$
$$\times\{\mathrm{E}\,\psi^2(\mathrm{X}_0)+\Sigma\,\mathrm{D}^{k1}\,\mathrm{D}^{l1}\,\mathrm{L}(\mathrm{E}\,\varphi(\mathrm{X}_0))\mathrm{D}^{k2}\,\mathrm{D}^{l2}\,\mathrm{L}(\mathrm{E}\,\varphi(\mathrm{X}_0)).\mathrm{E}\varphi'_{l1}(\mathrm{X}_0)$$
$$\times\,\mathrm{E}\,\varphi'_{l2}(\mathrm{X}_0).\mathrm{Cov}(\varphi_{k1}(\mathrm{X}_0),\,\varphi_{k2}(\mathrm{X}_0))$$
$$+2\Sigma\,\mathrm{E}\,\varphi'_l(\mathrm{X}_0)\,\mathrm{E}\,(\varphi_k(\mathrm{X}_0)\,\psi(\mathrm{X}_0))\,\mathrm{D}^k\,\mathrm{D}^l\,\mathrm{L}(\mathrm{E}\,\varphi)\}.$$

III.3. Étude de la variance asymptotique

Lorsque X est centré, on a en notant $\mathrm{T}_{i,j}=\sum_{k\neq 0} s_{k-i}\,s_{k-j}$:

$$\mathrm{A}=\mathrm{var}\,\mathrm{X}_0.\mathrm{E}\,\psi'(\mathrm{X}_0).\mathrm{T}$$
$$\Lambda=\mathrm{var}\,\mathrm{X}_0\,\mathrm{E}\,\psi^2(\mathrm{X}_0).\mathrm{T}.$$

Remarque. — Λ et A diagonalisent dans la même base, donc, en regardant la démonstration du théorème 3, on voit que les $((\hat{c}_j^n))$ sont asymptotiquement indépendantes.

On obtient $\Sigma^*=\dfrac{\mathrm{E}\,\psi^2(\mathrm{X}_0)}{\mathrm{Var}\,\mathrm{X}_0\,(\mathrm{E}\,\varphi'(\mathrm{X}_0))^2}.\mathrm{T}^{-1}=\mathrm{V}(\psi,\,\mathrm{X}_0).\mathrm{T}^{-1}$.

On obtient pour Σ^* la borne inférieure donnée par le résultat suivant :

THÉORÈME 6. — *Lorsque la loi de* X_0 *possède une information de Fisher :*

$$\mathrm{V}(\psi,\,\mathrm{X}_0)\geqq\frac{1}{\mathrm{Var}\,\mathrm{X}_0\,\mathrm{I}(\mathrm{X}_0)-1}$$

où I(X) *est l'information de Fisher de la loi de* X.

Démonstration. — Soit

$$\mathrm{A}(\psi,\,\mathrm{X})=\sup_{\mathrm{E}\,\psi=0,\;|\mathrm{E}\,\psi'|<\infty,\;\mathrm{EX}\,\psi=0,\;\mathrm{E}\,\psi^2<\infty}\frac{(\mathrm{E}\,\psi'(\mathrm{X})^2)}{\mathrm{E}\,\psi^2(\mathrm{X}_0)}.$$

Posant $\theta(u)=\psi(u)-\dfrac{\mathrm{EX}_0\,\psi(\mathrm{X}_0)}{\mathrm{Var}\,\mathrm{X}_0}.(u-\mathrm{EX}_0)$, on voit facilement que :

$$\mathrm{A}(\psi, \mathrm{X})=\sup_{\mathrm{E}\,\theta(\mathrm{X})=0,\ |\mathrm{E}\,\theta'(\mathrm{X})|<\infty,\ \mathrm{E}\,\theta^2(\mathrm{X})<\infty}\frac{(\mathrm{E}\,\theta'^2(\mathrm{X}))}{\mathrm{E}\,\theta^2(\mathrm{X})}$$

$$\mathrm{E}\,\theta'(\mathrm{X})=\mathrm{E}\,\psi'(\mathrm{X})-\frac{\mathrm{EX}\,\psi(\mathrm{X})}{\mathrm{Var}\,\mathrm{X}}=-\left(\mathrm{E}\,\psi(\mathrm{X})\frac{f'(\mathrm{X})}{f(\mathrm{X})}+\frac{\mathrm{EX}\,\psi(\mathrm{X})}{\mathrm{Var}\,\mathrm{X}}\right)$$

sous des conditions suffisantes de régularité de la densité f de la loi de X.

$$\mathrm{E}\,\theta'(\mathrm{X})=-\mathrm{E}\left(\psi(\mathrm{X})-\frac{\mathrm{X}-\mathrm{EX}}{\mathrm{Var}\,\mathrm{X}}\mathrm{EX}\,\psi(\mathrm{X})\right)\times\left(f'/f(\mathrm{X})+\frac{\mathrm{X}-\mathrm{EX}}{\mathrm{Var}\,\mathrm{X}}\right)\quad \text{car } \mathrm{E}\,\psi(\mathrm{X})=0.$$

Écrivons l'inégalité de Schwarz :

$$(\mathrm{E}\,\theta'(\mathrm{X}))^2 \leqq (\mathrm{E}\,\theta^2(\mathrm{X})).\mathrm{E}[f'/f(\mathrm{X})+(\mathrm{X}-\mathrm{EX})/\mathrm{Var}\,\mathrm{X}]^2$$

soit

$$\mathrm{A}(\psi, \mathrm{X}) \leqq \mathrm{E}[f'/f(\mathrm{X})+(\mathrm{X}-\mathrm{EX})/\mathrm{Var}\,\mathrm{X}]^2=\mathrm{I}(\mathrm{X})-1/\mathrm{Var}\,\mathrm{X},$$

et cette borne est atteinte uniquement pour :

$$\psi(u)-\frac{u-\mathrm{EX}}{\mathrm{Var}\,\mathrm{X}}.\mathrm{EX}\,\psi(\mathrm{X})=\alpha.\left(\frac{f'}{f}(u)+\frac{u-\mathrm{EX}}{\mathrm{Var}\,\mathrm{X}}\right),$$

soit $\psi=\alpha\dfrac{f'}{f}+\beta\dfrac{u-\mathrm{EX}}{\mathrm{Var}\,\mathrm{X}}$, la contrainte $\mathrm{EX}\,\psi(\mathrm{X})=0$ entrainant $\alpha=\beta$. □

Remarques. — 1. Cette inégalité souligne l'extrémalité de la loi gaussienne, où le problème n'a pas de solution [et où $\mathrm{var}\,\mathrm{X}_0\,\mathrm{I}(\mathrm{X}_0)=1$].

2. Si l'on fait sur le modèle la restriction de causalité, on peut comparer aux estimateurs classiques de moindres carrés. Par exemple lorsque $d=1$, $\mathrm{Ex}_0=0$, $s_k=0$ pour $k<0$, la variance asymptotique de tels estimateurs est T^{-1} [14]. Si $\mathrm{var}\,\mathrm{X}_0\,\mathrm{I}(\mathrm{X}_0)\gg 1$, et que l'objectif est bien choisi (au sens où l'inégalité du théorème 6 est presque atteinte), l'estimation par objectif est meilleure que celle par moindres carrés. Cette méthode peut donc être considérée comme une alternative aux méthodes usuelles, dans le cas restrictif causal, à utiliser lorsque l'on est « loin de la situation gaussienne » (au sens où $\mathrm{var}\,\mathrm{X}_0\,\mathrm{I}(\mathrm{X}_0)\gg 1$).

III.4. Problèmes d'efficacité

Lorsque $d=1$ et avec la normalisation $\hat{\theta}\in\mathbb{U}$, on a un modèle paramétré de paramètre $\theta=(c_1, \ldots, c_p)$. On calcule l'information de Fisher du modèle lorsqu'elle existe :

$$\mathrm{I} = \lim_{\infty} 1/n\, \mathrm{I}_n(\theta), \qquad \text{où} \quad \mathrm{I}_n(\theta) = \mathrm{E}(\partial \mathrm{L}_n/\partial c_i \,.\, \partial \mathrm{L}_n/\partial c_j \,.\, 1/\mathrm{L}_n^2),$$

L_n étant la vraisemblance du processus X. On peut en effet montrer que le résultat est le même que si l'on calcule I_n avec la vraisemblance de Y, il s'agit d'un changement de paramètre biunivoque. On rappelle qu'alors, quel que soit l'estimateur du paramètre, sa variance asymptotique Σ quand elle existe au sens décrit en I.3 vérifie :

$\Sigma - \mathrm{I}^{-1}$ est s. d. p. Si $\Sigma = \mathrm{I}^{-1}$, on dit que l'estimateur est efficace.

Sous des conditions de régularité et d'intégrations de la fonction caractéristique de la loi de X, on obtient (*voir* le calcul et les conditions plus loin) :

THÉORÈME 7 :

$$\begin{aligned} \mathrm{I}_{i,j} = {} & \mathrm{Var}\,\mathrm{X}\,.\,\mathrm{I}(\mathrm{X}) \sum_{k \neq 0} s_{k-i} s_{k-j} + \sum_{k \neq 0} s_{k-i} s_{-k-j} \\ & + \{\mathrm{E}[(\mathrm{X}-\mathrm{EX})^2 f'^2/f^2] - 1\} s_{-i} s_{-j} + (\mathrm{EX})^2\, \mathrm{E} f'^2/f^2\, (\Sigma f_k)^2 \\ & + [\mathrm{EX}\,.\,\mathrm{E}(\mathrm{X} f'^2/f^2) - (\mathrm{EX})^2\, \mathrm{E} f'^2/f^2]\, (s_{-i} + s_{-j})\, \Sigma s_k. \end{aligned}$$

Démonstration. — On calcule les vraisemblances à l'aide de leurs fonctions caractéristiques. Si $\varphi_{c,n}$ est la fonction caractéristique de la loi jointe de $(\mathrm{X}_1(c), \ldots, \mathrm{X}_n(c))$, on a :

$$\varphi_{c,n}(t_1, \ldots, t_n) = \prod_{l=-\infty}^{+\infty} \varphi\left(\sum_{k=0}^{n} (c * f)_{k-l}\, t_k\right),$$

où φ est la fonction caractéristique de la loi de X.

Les hypothèses sont du type suivant :

- Régularités, dérivabilités de φ.
- Intégrabilité de fonctions de φ et de ses dérivées et majoration par des fonctions intégrables (ceci pour échanger dérivation et intégration).

Ce sont celles qui permettent d'écrire le formulaire suivant :

$$f(x) = \int \varphi(t) \exp - itx \,.\, 1/2\pi, \qquad \int \varphi'(t) \exp - itx = 2\pi\, ixf(x),$$

$$\int t\, \varphi(t) \exp - itx = 2\pi\, if'(x),$$

$$\int t\, \varphi'(t) \exp - itx = 2\pi \,.\, -(f(x) + xf'(x)),$$

$$\int t^2\, \varphi(t) \exp - itx = -2\pi f''(x),$$

$$\int t^2\, \varphi'(t) \exp - itx = 2\pi \,.\, -i(f'(x) + xf''(x)),$$

$$\int \varphi''(t) \exp - itx = 2\pi . - x^2 f(x),$$

$$\int t \varphi''(t) \exp - itx = -2\pi i . (2xf(x) + x^2 f'(x),$$

$$\int t^2 \varphi''(t) \exp - itx = 2\pi . - (2f(x) + 4xp'(x) + x^2 f''(x)).$$

On écrit alors :

$$\frac{\partial \varphi_{c,n}}{\partial c_i} = \Sigma_l \left(\sum_{k=0}^{n} s_{k-l-i} t_k \right) \varphi' \left(\sum_{k=0}^{n} (c * s)_{k-l} t_k \right) \prod_{m \neq l} \varphi \left(\sum_{k=0}^{n} (c * s)_{k-m} t_k \right)$$

$$\frac{\partial^2 \varphi_{c,n}}{\partial c_i \partial c_j} = \Sigma_l \left(\sum_{k=0}^{n} s_{k-li} t_k \right)$$

$$\times \left(\sum_{k=0}^{n} s_{k-l-j} t_k \right) \varphi'' \left(\sum_{k=0}^{n} (c * s)_{k-l} t_k \right) \prod_{m \neq l} \varphi \left(\sum_{k=0}^{n} (c * s)_{k-m} t_k \right)$$

$$+ \Sigma_l \sum_{m \neq l} \left(\sum_{k=0}^{n} s_{k-li} t_k \right) \left(\sum_{k=0}^{n} s_{k-m-j} t_k \right)$$

$$\times \varphi' \left(\sum_{k=0}^{n} (c * s)_{k-l} t_k \right) \varphi' \left(\sum_{k=0}^{n} (c * s)_{k-m} t_k \right) \prod_{p \neq m \text{ et } l} \varphi \left(\sum_{k=0}^{n} s_{k-p} t_k \right).$$

En écrivant ces relations en $c=b$, en les multipliant par $\exp - i \Sigma t_k x_k$ et en les intégrant à l'aide du formulaire précédent, on obtient

$$\left. \frac{1}{f_{n,c}} . \frac{\partial f_{n,c}}{\partial c_i} \right|_{c=b}$$

et

$$\left. \frac{1}{f_{n,c}^2} . \frac{\partial f_{n,c}}{\partial c_i} . \frac{\partial f_{n,c}}{\partial c_j} - \frac{1}{f_{n,c}} . \frac{\partial^2 f_{n,c}}{\partial c_i \partial c_j} \right|_{c=b}$$

puis on en prend l'espérance pour obtenir :

$$(\mathrm{I}_n)_{i,j} = -(n+1) s_{-i} s_{-j} + \sum_{k \neq 0, = -n}^{n} s_{k-i} s_{-k-j} . (n - |k| + 1)$$

$$+ (n+1) s_{-i} s_{-j} \mathrm{EX}^2 p'^2/p^2$$

$$\times \mathrm{EX}^2 . \mathrm{E}f'^2/f^2 . \sum_{l=0}^{n} \sum_{k \neq 1, =0}^{n} s_{k-l-i} s_{k-l-j}$$

$$+ (\mathrm{EX})^2 \mathrm{E}f'^2/f^2 . \left(\Sigma_{l1, l2} \sum_{k \neq l1 \text{ et } l2, =0}^{n} s_{k-l1-i} s_{k-l2-j} \right.$$

$$\left. - \Sigma_l \sum_{k \neq 1, =0}^{n} s_{k-l-i} s_{k-l-j} \right)$$

$$+\mathrm{EX}.\mathrm{EX}f'^2/f^2.\left[\Sigma_{l1}\sum_{l2\neq l1,\ =0}^{n}(s_{-i}s_{l2-l1-j}+s_{-j}+s_{-j}s_{l2-l1-i})\right].$$

La limite $\mathrm{I}_{i,j}$ s'en déduit aussitôt.

Lorsque X est centré et que le filtre s est causal :

$$\mathrm{I}^{-1}=\frac{1}{\mathrm{Var}\,\mathrm{X}.\mathrm{I}(\mathrm{X})}\mathrm{T}^{-1}.$$

On s'aperçoit alors que la méthode d'estimation par objectif de type linéaire n'est jamais efficace.

Remarque sur un estimateur adaptatif. — Dans le cas où s est causal, l'existence et la construction d'estimateurs adaptatifs efficaces a été complètement résolue par Kreiss [10]. Il semble que la causalité soit cruciale dans ce problème et nous faisons la

Conjecture. — Un estimateur adaptatif efficace existe si et seulement si s est causal.

En effet, supposons que la loi de X_0 est centrée et de la forme : $f=(1-t)g+th$. Calculons l'information de Fisher asymptotique du modèle paramétré par (b, t). Avec un calcul analogue à celui développé précédemment, on obtient facilement que le terme croisé en b_i, t est proportionnel à s_{-i}. Pour que l'adaptation soit possible, suivant les idées de Stein de décorrélation asymptotique du paramètre estimé et du paramètre de nuisance [15], ce terme devrait être nul. Ceci implique alors : $\forall i=1\ldots p$, $s_{-i}=0$, et puisque s a un inverse de longueur p, ceci implique : $\forall i>0$, $s_{-i}=0$.

Un travail ultérieur étudiera ce problème.

III.5. Robustesse

III.5.1. *Robustesse à une sous-paramétrisation* (*cas* $d=1$)

On se place dans la situation où l'on a sous-estimé p, c'est-à-dire : $b=(b_1, b_2)$, $b_1\in\mathbb{R}^{p-q}$ et $b_2\in\mathbb{R}^q$, $\|b_2\|=\varepsilon_q$ est petit, et l'on cherche b dans $\mathbb{R}^{p-q}$, c'est-à-dire que l'on maximise $\mathrm{J}_n(p-q, c)$ en c. Soit ω le module de continuité de $\mathrm{J}(p, c)$ en b et ω_{-1} le module de continuité de J^{-1} [fonction inverse de $\mathrm{J}(p, c)$] en b.

On a le résultat suivant :

THÉORÈME 9. — $\forall\delta>0$, $\mathrm{P}[\|c_n-b\|\geqq\omega_{-1}(\delta+\omega(\varepsilon_q))]\to 0$ *quand* $n\to\infty$ *où* c_n *désigne l'estimateur complété par* $q\,0$.

Démonstration. — On reprend les notations de la démonstration de consistance. Sur $\mathrm{E}_n(\delta/2)$, on a :

$$\mathrm{J}(c_n)+\delta/2\geqq\mathrm{J}_n(c_n)\geqq\mathrm{J}_n(b_1, 0)\geqq\mathrm{J}(b_1, 0)-\delta/2\geqq\mathrm{J}(b)-\delta/2-\omega(\varepsilon_q).$$

Donc $|J(c_n)-J(b)| \leqq \delta+\omega(\varepsilon_q)$, et $\|c_n-b\| \leqq \omega_{-1}(\delta+\omega(\varepsilon_q))$.
Donc $P[\|c_n-b\| \leqq \omega_{-1}(\delta+\omega(\varepsilon_q))] \leqq P(E_n(\delta/2))$.

III.5.2. *Robustesse au bruit*

On suppose mainteant que le signal sortant est bruité :
$Y=s*X+\varepsilon$, où (ε) est un bruit indépendant de X.

La question est la suivante : L'estimateur de maximum d'objectif conserve-t-il de bonnes propriétés ?

Nous étudierons seulement un cas particulier du problème, mais relativement représentatif, le cas où J est un cumulant standardisé et où les ε_t sont i. i. d. de loi gaussienne.

On a alors le résultat suivant :

THÉORÈME 6. — *Dans le cadre décrit ci-dessus, on a :* $\forall\delta>0$, $\exists\alpha>0$ (α *ne dépendant pas de* X) *tel que :*

$$|\operatorname{Var}\varepsilon_0/\operatorname{Var}X_0|<\alpha \quad\Rightarrow\quad P(\limsup_n \|b_n-b\| \leqq \delta)=1.$$

Démonstration. — Par un calcul facile on obtient :

$$J(c)=\frac{\Sigma(c*s)_k^m}{(\Sigma(c*s)_k^2+\Sigma_{(0,\ ,0)}^{(p,\ ,p)}c_k^2.\operatorname{Var}\varepsilon_0/\operatorname{Var}X_0)^{m/2}}.h(X_0)$$
$$=g(c,\ \operatorname{Var}\varepsilon_0/\operatorname{Var}X_0).h(X_0)$$

$h(X_0)$ étant le cumulant d'ordre m standardisé de X_0.

On choisit l'estimateur b_n comme réalisant le maximum global de $J_n(c)$ sur $\mathbb{T}$ (il peut y avoir plusieurs choix possibles). Soit $c(\eta)$ une suite sur $\mathbb{T}$ de maximisateurs de $J(c)$ lorsque $\operatorname{Var}\varepsilon_0$ tend vers 0. Étant sur un compact, elle a au moins une valeur d'adhérence c_*. Par passage à la limite, $g(c_*,0)=0$. Or, pour $\operatorname{Var}\varepsilon_0=0$, on sait que J a un maximisateur unique B. Donc $c(\eta)$ a une seule valeur d'adhérence b qui est ainsi limite de la suite. Ceci s'écrit :

(1) $\forall\delta>0$, $\exists\alpha>0$ tel que

$$|\operatorname{Var}\varepsilon_0/\operatorname{Var}X_0|<\alpha \quad\Rightarrow\quad \|c(\eta)-b\|<\delta,\qquad \eta=\operatorname{Var}\varepsilon_0.$$

Il n'y a, pour tout η, qu'un nombre fini de valeurs de $c(\eta)$ (g est une équation polynomiale). α dans (1) peut donc être choisi indépendamment de $c(\eta)$.

D'autre part, rappelons qu'en étudiant $J_n(c)-J_n(c)$, il est facile de montrer que $J_n(c)$ converge presque sûrement vers J(C) uniformément en c. Soit $\Omega=\{\omega/\lim_n \sup_{\mathbb{T}} |J_n(c)-J(c)|\}$, $P(\Omega)=1$.

Soit ω dans Ω, et soit b_* une valeur d'adhérence de $b_n(\omega)$. Il est facile de voir que b_* est un $c(\eta)$ [=maximisateur de $J(c)$], et donc que $\|b_*-b\|<\delta$ dès que $|\operatorname{Var}\varepsilon_0/\operatorname{Var}X_0|<\alpha$.

On en déduit aisément le théorème.

III.6. Identification = estimation de l'ordre p

L'idée est d'utiliser pour estimer p une méthode analogue à la méthode de vraisemblance compensée. (*Voir* [1].)

Si $d > 1$, on supposera $p_1 = \ldots = p_d = p$.

On introduit le paramètre supplémentaire q paramétrisant l'ordre, et l'on pose :

$$A_n(q, c) = n^{d\alpha} J_n(q, c) - \lambda q,$$

où $J_n(q, c)$ est la fonction d'objectif définie lorsque le modèle est d'ordre q.

On choisit alors comme estimateur (q_n, c_n) le couple (q, c) qui maximise $A_n(q, c)$ pour $q \in \{0, \ldots, Q\}$ et c vérifiant la condition de normalisation choisie.

Q peut être arbitrairement grand, mais il doit être fixé.

On a alors :

THÉORÈME 8. — *Pour* $0 < \alpha < 1$ *et* $\lambda > 0$, *sous les conditions* (C1) *et* (C2), *et* (A1)...(A3) :

$$(q_n, c_n) \xrightarrow{P} (p, b)$$

et l'asymptotique pour c_n *est la même que lorsque l'on n'estime pas* p.

Démonstration. — 1. La première partie du théorème découle du fait que, si $\alpha \in]0; 1[$, et si $q > p$, $n^{d\alpha}[J_n(q.c_{*n}) - J_n(q, b_*)] \xrightarrow{P} 0$ où c_{*n} maximise $J_n(q, c)$ à q fixé et où $b_* = (b, 0\ldots 0)$.

En effet, à $q > p$ fixé, tous les résultats précédents sont valables (le modèle peut être considéré comme étant d'ordre q et de paramètre b_*), et en particulier :

$$J_n(q, c_{*n}) - J_n(q, b) = 1/2.{}^t(C_{*n} - b_*)[D^2 J_n(c_{*n}) + O_P(1)](c_{*n} - b_*)$$

et le résultat est alors immédiat sachant que $n^{d/2}(c_{*n} - b_*)$ est tendu. On en déduit immédiatement :

Si $q > p$,

$$A_n(q, c_{*n}) - A_n(p, b_{*n}) \xrightarrow{P} -\lambda(q-p) < 0.$$

Maintenant :

$$\{q_n > p\} \subset \{\exists q \in \{p+1\ldots Q\} / A_n(q, c_{*n}) - A_n(p, b_{*n}) \geqq 0\}$$

donc

$$P(q_n > p) \leqq \sum_{q=p+1}^{Q} P(A_n(q, c_{*n}) - A_n(p, b_{*n}) \geqq 0),$$

et $P(q_n > p) \underset{n \to \infty}{\longrightarrow} 0$. D'autre part, on a de la même façon :

$$P(q_n < p) \leqq \sum_{q=1}^{p-1} P(A_n(q, c_{*n}) - A_n(p, b_{*n}) \geqq 0)$$
$$\leqq \sum_{q=1}^{p-1} P(J_n(q, c_{*n}) \geqq J_n(p, b_{*n}) - \lambda p/n^{d\alpha}\}.$$

Il est immédiat, du fait de l'optimalité stricte de $J(b)$ et du fait que l'optimum est atteint que l'on a, si la normalisation choisie est $\mathbb{T}$ ou $\mathbb{U}$:

$$\sup_c J(q, c) < \sup_b J(p, b).$$

Il est de même, grâce à ceci et à la convergence uniforme de $J_n(q, c)$ vers $J(q, c)$, facile de montrer que, pour $q < p$:

$$\forall \varepsilon > 0, \quad P(J_n(q, c_{*n}) \geqq J_n(p, b_{*n}) - \varepsilon) \to 0 \text{ quand } n \to \infty.$$

On en déduit que $P(q_n < p) \to 0$ quand $n \to \infty$, et donc que $q_n \overset{P}{\to} p$. La convergence p.s. de q_n n'est pas étudiée ici, elle nécessite des calculs beaucoup plus fins.

2. La deuxième partie du théorème découle du fait que $q_n \to p$ en probabilité et que sur $\{q_n = p\}$, l'asymptotique est la même que celle étudiée précédemment.

IV. LES CAS SUR ET SOUS-GAUSSIENS

Dans ce paragraphe, nous reprenons les résultats et la méthode développés par A. Benveniste, M. Goursat et G. Ruget dans leur article : Robust identification of a nonminimum phase system: Blind adjustment of a linear equalizer in data communications [2].

On ne suppose plus que l'inverse de s est fini-dimensionnel. Dans ce cadre les méthodes proposées dans le chapitre précédent ne s'appliquent que si l'on impose à s l'appartenance à un compact fixé, condition non imposée dans le travail de Ruget et coll. qui font en contrepartie des hypothèses beaucoup plus fortes sur les lois. On cherche à construire un système linéaire de filtre θ de telle sorte que le système global $r = \theta * s$ soit l'identité avec un retard éventuel. Le moyen proposé dans [2] est l'utilisation d'une fonctionnelle $V(\theta) = J(r)$ construite à partir de la loi des variables U_t issues du système global : $U_t = \Sigma_k r_k X_{t-k} = \Sigma_k \theta_k Y_{t-k}$, cette fonctionnelle présentant en $\theta = s^{-1}$ un minimum global.

On reconnaît la méthode de type « contraste », identique à celle utilisée dans cet article.

Le point de vue adopté ensuite est celui d'un algorithme stochastique, présentant des avantages de calcul importants par rapport aux méthodes d'estimation par minimum de contraste.

Les auteurs restreignent leur étude au cas où la loi de X est symétrique et, bien entendu, toujours non gaussienne.

Ils étudient un ensemble particulier de contrastes, de la forme : $\mathrm{V}(\theta)=\mathrm{J}(r)=\mathrm{E}\,\psi(\mathrm{U}_t)$, où ψ est une fonction paire pour donner le même poids à un point et à son symétrique puisque la loi initiale est paire; ces contrastes conduisent à un algorithme stochastique dont nous n'étudierons pas les aspects propres.

Le problème de recherche de conditions suffisantes, portant sur ψ, et relativement générales, sous lesquelles J est un contraste est étudié par les auteurs. Nous en proposerons une interprétation.

Nous nous intéresserons à la restriction de $\mathrm{J}(r)$ à la sphère, l'extension à l^2 tout entier est ensuite simplement un problème d'échelle.

Le théorème porte sur deux classes de lois, sous-gaussiennes et super-gaussiennes, lois ayant une densité de la forme :

$$\mathrm{F}(dx)=\mathrm{K}\,.\exp(-g(x))\,dx,$$

et on a les définitions suivantes :

1. F est sous-gaussienne si g est paire et si $g(x)$ et $g'(x)/x$ sont strictement croissantes sur $\mathbb{R}$, ou bien si F est uniforme sur un segment $[-d;\ +d]$.

2. F est super-gaussienne si g est paire, strictement croissante sur $\mathbb{R}^+$ et $g'(x)/x$ strictement décroissante sur $\mathbb{R}^+$.

Par exemple, la loi de densité $\mathrm{C}.\exp-|x|^{\mathrm{T}}$ est sous-gaussienne si $\gamma > 2$ et super-gaussienne si $\gamma < 2$.

On a alors [2].

THÉORÈME. — *Si* F *est sous-gaussienne* (*resp. super-gaussienne*), *et si* φ *est deux fois dérivable sauf peut être à l'origine avec* $\varphi(0+)\leqq 0$ [*resp.* $\varphi(0+)\geqq 0$] *et* $\varphi''\geqq 0$ *sur* $\mathbb{R}^+$ (*resp.* $\varphi''\leqq 0$), *l'une des deux inégalités étant stricte, alors :* $\dfrac{\partial \mathrm{J}}{\partial \alpha} > 0$ *pour* $\alpha\in\,]0,\pi/4[$, *et* Id *et* $-$Id *à un retard près sont les seuls minima de* J *sur la sphère unité.*

Ceci conduit à des fonctions ψ de la forme :

$$\psi(x)=\alpha x^2+\varepsilon.\gamma|x|+\beta-\delta.\operatorname{Log} h(x),$$

où $\gamma\geqq 0$, $\delta\geqq 0$, $h(x)=\mathrm{K}.\exp(-g(x))$, g étant paire et trois fois dérivable sauf peut-être à l'origine, K normalisant h au sens l_1, et avec :

1. Si F est sous-gaussienne, $\varepsilon=-1$ et $g'''\geqq 0$, l'inégalité étant stricte si $\gamma=0$ et $g'(0+)=0$.

2. Si F est super-gaussienne, $\varepsilon=+1$ et $g'''\leqq 0$, l'inégalité étant stricte si $\gamma=0$ et $g'(0+)=0$.

Le contraste est alors donné par :

$$J(r)=\alpha . E(\Sigma r_k X_{-k})^2+\beta+\varepsilon . \gamma E|\Sigma r_k X_{-k}|-E(\operatorname{Log} h(\Sigma_k r_k X_{-k})).$$

Le terme de variance est constant sur la sphère unité. A un terme constant près, les contrastes considérés sont des combinaisons linéaires positives de $\pm E|Y|$ et $-E(\operatorname{Log} h(Y))$.

On remarque tout d'abord que si J est un contraste pour une loi initiale sous-gaussienne, alors $-J$ est un contraste pour une loi initiale super-gaussienne.

Si F_r désigne la loi de la variable aléatoire $\Sigma_k r_k X_{-k}$, et si H désigne la loi de densité h, et si H_0 désigne la loi super-gaussienne de densité $C.\exp(-|x|)$, on a :

$$J(r)=\delta .(K(F_r, H)-K(F_r, G))+\varepsilon\gamma .(K(F_r; H_0)-K(F_r; G)),$$

où G est la loi gaussienne centrée de même variance que X_0 et K désigne la distance de Kullback.

Ceci appelle l'interprétation suivante :

Soient les ensembles de loi suivants :

$U=\{F$, loi de densité $h(x)=K.\exp(-g(x))$ où g est paire, strictement croissante sur $\mathbb{R}^+$, trois fois dérivable sauf peut-être à l'origine, et $g''' > 0$ sur $R_*^+\}$.

$S=\{F$, loi de densité $h(x)=K.\exp(-g(x))$ où g est paire, strictement croissante sur $\mathbb{R}^+$, trois fois dérivable sauf peut-être à l'origine, et $g''' < 0$ sur $R_*^+\}$.

U est un sous-ensemble des lois sous-gaussiennes et S un sous-ensemble des lois super-gaussiennes.

On a alors :

Si la loi initiale F est dans U (resp. dans S), alors si H est une loi quelconque de U (resp. S), $K(F_r, H)-K(F_r, G)$ est un contraste. On peut de même choisir $J(T)=\underset{H \in U \text{ (resp. } H \in S)}{\operatorname{Min}} K(F_r, H)-K(F_r, G)$. Le fait que de tels J soient des contrastes peut maintenant s'énoncer de la façon suivante :

Filtrer la loi initiale linéairement écarte plus de la classe de lois considérées que de se placer en un point quelconque de la classe, ou encore : la « distance » entre la loi filtrée et la classe initiale est plus grande que la plus grande largeur de la classe, tous les termes d'éloignement étant pris au sens distance de Kullback compensée par l'écart à la gaussienne.

D'autre part, on a vu que l'opposé d'un contraste pour une loi initiale dans une classe était un contraste pour une loi initiale dans l'autre classe. On peut alors donner une autre formulation :

La distance entre les deux classes de lois est plus grande que la distance d'une loi filtrée (la loi initiale étant dans une des deux classes) à une quelconque des deux classes.

Le résultat du théorème est donc, formulé ainsi, une caractérisation de deux classes de loi par rapport à une certaine géométrie et de leur comportement conjoint par filtration linéaire.

Il serait intéressant de trouver une caractérisation portant plus sur les fonctionnelles que sur les lois, et d'obtenir des contrastes opérant sur des classes de lois plus larges. En regardant des contrastes particuliers obtenus par cette méthode, on s'aperçoit qu'ils opèrent pour des classes de lois initiales beaucoup plus vastes; on retrouve en fait des contrastes cités dans le chapitre précédent :

(*a*) En prenant pour H la fonction initiale, on a :

$$J(r)=K(F_r, F)-K(F_r, G).$$

On reconnaît le contraste optimal au sens variance asymptotique pour les estimateurs d'objectif de type linéaire étudiés dans le chapitre précédent.

(*b*) En regardant la famille de lois sous- et super-gaussienne de densité $C.\exp-|x|^{\gamma}$, on a $J(r)=E\,\Sigma_k r_k X_{-k}|^{\gamma}$.

Pour $\gamma=4$, on obtient :

$$E(\Sigma r_k X_{-k})^4=(\Sigma r_k^4)(EX_0^4-3(EX_0^2)^2)-3(\Sigma r_k^2)^2(EX_0^2)^2.$$

On voit donc que $J(r)=E(\Sigma t_k X_{-k})^4$ est un contraste pour des lois initiales ayant un moment d'ordre 4 et vérifiant $EX_0^4<3(EX_0^2)^2$, et que $-E(\Sigma r_k X_{-k})^4$ est un contraste pour des lois dont les moments d'ordre 4 et 2 vérifient l'inégalité contraire.

On retrouve en fait ici le cumulant d'ordre 4.

Rappelons enfin que la valeur absolue des cumulants standardisés est un contraste pour toutes les lois initiales possédant un tel cumulant non nul, mais que ce contraste n'est pas (sauf pour l'ordre 4) de la forme très simple étudiée ici $E\,\psi(Y)$.

RÉFÉRENCES

[1] R. Azencott et D. Dacunha-Castelle, *Séries d'observations irrégulières*, Masson, 1984.

[2] A. Benveniste, M. Goursat et G. Ruget, Robust Identification of a Non Minimum Phase System: Blind Adjustment of a Linear Equalizer in Data Communications, *I.E.E.E. Trans. Aut. Cont.*, vol. **A.C. 25**, 1980, p. 3.

[3] P. Billingsley, *Convergence of Probability Measures*, Wiley, 1986.

[4] G. E. Box et G. M. Jenkins, *Time Series Analysis, Forecasting and Control*, Holden-day, 1976.

[5] D. Dacunha-Castelle et M. Duflo, *Probabilités et statistiques*, t. **2**, Masson, 1983.

[6] D. Donoho, On Minimum Entropy Deconvolution, *Proc. of 2nd Appl. Time Series Symposium*, Tulsa, D.F. Findley, 1980, p. 565-608.

[7] E. Gassiat, *Semi-Parametric Estimation of a Stationary Non Necessary Causal* AR(p) *Process with Infinite Variance*, accepté pour publication dans Journal of Multivariate Analysis.

[8] V. V. GORODETSKI, On the Strong Mixing Property for Linear Sequences, *Theor. prob. appl.*, vol. **22**, 1977, p. 411-413.

[9] X. GUYON, *Champs stationnaires sur* $\mathbb{Z}^2$: *modèles, statistiques et simulations*, prepub., Orsay, 1985.

[10] J. P. KREISS, On Adaptative Estimation in Stationary ARMA Processes, *Ann. Stat.*, vol. **15**, 1, 1987, p. 112-133.

[11] K. S. LII et M. ROSENBLATT, Deconvolution and Estimation of Transfer Function Phase and Coefficients for Non Gaussian Linear Processes, *Ann. stat.*, vol. **10**, 4, 1982, p. 1195-1208.

[12] MAC-LEISH, Invariance Principle for Dependent Variables, *Z. Wahrsch.*, vol. **32**, 1975, p. 165-178.

[13] A. MOKKADEM, Sur le mélange d'un processus ARMA vectoriel, *C.R. Acad. Sci. Paris*, t. **303**, série I, 1986, p. 519-521.

[14] M. ROSENBLATT, *Stationary Sequences and Random Fields*, Birckhauser, Boston, 1985.

[15] C. STEIN, Efficient Non Parametric Testing and Estimation, *Proc. third Berkeley symp. math. stat. prob.*, vol. **1**, University of California press, 1956, p. 187-196.

[16] P. WHITTLE, Gaussian Estimation in Stationary Time Series, *Bull. I.S.I.*, vol. **39**, p. 105-129.

[17] ZOLOTAREV, Ideal Metrics in Problems of Probability Theory and Mathematical Statistics, *Austral. J. Stat.*, vol. **21**, 1979, p. 193-203.

(*Manuscrit reçu le* 21 *février* 1989.)