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

M. DUFLO
R. SENOUSSI
A. TOUATI

Propriétés asymptotiques presque sûres de l'estimateur des moindres carrés d'un modèle autorégressif vectoriel

Annales de l'I. H. P., section B, tome 27, nº 1 (1991), p. 1-25
<http://www.numdam.org/item?id=AIHPB_1991__27_1_1_0>

Ann. Inst. Henri Poincaré,

Vol. 27, n° 1, 1991, p. 1-25. *Probabilités et Statistiques*

Propriétés asymptotiques presque sûres de l'estimateur des moindres carrés d'un modèle autorégressif vectoriel

par

M. DUFLO, R. SENOUSSI

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

et

A. TOUATI

École Normale supérieure, Bizerte, Tunisie

RÉSUMÉ. — Pour un modèle autorégressif vectoriel commandable, on prouve la consistance forte de l'estimateur des moindres carrés sauf dans le cas « singulier » suivant : il y a un sous-espace propre de dimension $\geqq 2$ associé à une valeur propre de module >1. Dans le cas singulier, cet estimateur peut être non consistant. Dans le cas régulier, on précise la vitesse de convergence presque sûre; on étudie aussi la suite des prédicteurs et l'estimateur empirique de la covariance.

ABSTRACT. — For a vector valued autoregressive model which is controllable, we prove the strong consistency of the least squares estimator except in the following "singular" case: there exists an eigensubspace of dimension $\geqq 2$ associated to an eigenvalue of modulus >1. In the singular case the consistency may fail. In the regular case, we precise the almost sure rate of convergence; we also study the predictor of this model and the empirical estimator of the covariance.

Classification A.M.S. : 62 M, 60 J.

Annales de l'Institut Henri Poincaré - Probabilités et Statistiques - 0246-0203
Vol. 27/91/01/1/25/$ 4,50/© Gauthier-Villars

I. PRINCIPAUX RÉSULTATS

I.1. Introduction

Malgré la très abondante littérature relative aux modèles linéaires et aux estimateurs des moindres carrés permettant leur identification, certains problèmes extrêmement simples ne sont pas encore résolus. Ainsi, jusqu'à une date récente, restait-il de nombreuses questions ouvertes relatives au modèle autorégressif vectoriel. Lorsqu'un tel modèle est stable – ou causal –, son comportement asymptotique et celui de son estimateur des moindres carrés sont bien connus grâce à des travaux anciens ([1], [2], ...); quelques pas avaient été effectués dans le cas explosif à cette période ([12], [19]).

Les propriétés de convergence presque sûre ont été éclairées dans le cas du modèle unidimensionnel autorégressif d'ordre p par les travaux de Lai, Wei et Chan ([8], [3], [4]); ces auteurs étudient aussi la convergence en loi dans le cas non explosif. Touati ([14], [15]) a étudié le comportement asymptotique en loi du modèle autorégressif vectoriel d'ordre 1 instable ou explosif et décrit le comportement en loi de l'estimateur des moindres carrés dans le cas général.

Nous précisons ici – pour un modèle autorégressif vectoriel – le comportement asymptotique presque sûr des trajectoires. En ce qui concerne la consistance, deux hypothèses apparaissent incontournables : la première, classique, est la commandabilité; la seconde a été dégagée dans le contre-exemple que nous avons bâti dans [5] d'un modèle autorégressif de dimension 2

$$X_{n+1} = 2\,X_n + \varepsilon_{n+1},$$

pour lequel l'estimateur des moindres carrés n'est pas consistant. Nous prouvons la consistance et analysons la vitesse de convergence sous les hypothèses qui apparaissent ainsi comme nécessaires.

I.2. Définitions et notations

On considère ici des suites de vecteurs aléatoires à valeurs dans $\mathbb{C}^d$ muni du produit scalaire hermitien $\langle u, v \rangle = {}^*uv$ et de la norme $\|u\|^2 = \langle u, u \rangle$; u désigne ici soit le vecteur, soit la matrice colonne associée, $^*u$ est sa matrice adjointe. Pour une matrice $d \times d$, A, *A est l'adjointe de A. Si A est hermitienne, on note $\lambda_{\min} A$ et $\lambda_{\max} A$ la plus petite et la plus grande valeur propre.

Soit une filtration $\mathbb{F}$ et un bruit $\varepsilon = (\varepsilon_n)$ adapté à cette filtration, c'est-à-dire une suite de vecteurs aléatoires de dimension d, adaptée à $\mathbb{F}$ et de carré intégrable pour laquelle nous supposons :

$$E(\varepsilon_{n+1} \mid \mathscr{F}_n) = 0 \qquad \text{et} \qquad E(\varepsilon_{n+1} {}^*\varepsilon_{n+1} \mid \mathscr{F}_n) = \Gamma,$$

Γ étant une matrice de covariance $d \times d$ déterministe, la « covariance » du bruit. On fait en outre, dans cet article, sur le bruit l'une *ou* l'autre des hypothèses suivantes.

(BB) le bruit est un « bruit blanc » de covariance Γ adapté à $\mathbb{F}$, c'est-à-dire que les vecteurs aléatoires (ε_n) sont indépendants, de même loi, centrés et de covariance Γ; ε_n est indépendante de $\mathscr{F}_{n-1}$.

(B 2) il existe un $a > 2$ tel que, p.s., $\sup E(\|\varepsilon_{n+1}\|^a \mid \mathscr{F}_n) < \infty$; le bruit « a un moment conditionnel d'ordre > 2 fini ».

La propriété suivante est alors réalisée :

$$n^{-1} \sum_{k=1}^{n} \varepsilon_k {}^*\varepsilon_k \xrightarrow{\text{p.s.}} \Gamma.$$

Il est facile de vérifier que certains de nos résultats n'utilisent que cette loi des grands nombres.

Désormais on donne un *« modèle autorégressif »* d'ordre 1, complexe et de dimension d :

$$X_{n+1} = AX_n + \varepsilon_{n+1},$$

où A est une matrice $d \times d$ et l'état initial est une v. a. X_0 $\mathscr{F}_0$-mesurable quelconque.

On pose, pour Q matrice hermitienne définie positive $\mathscr{F}_0$-mesurable,

$$S_n = Q + \sum_{k=0}^{n} X_k {}^*X_k, \qquad s_n = \sum_{k=0}^{n} \|X_k\|^2,$$
$$f_n = {}^*X_n S_n^{-1} X_n;$$

l'*« estimateur des moindres carrés »* de A, $\hat{A}_n = \left(\sum_{k=1}^{n} X_k {}^*X_{k-1} \right) S_{n-1}^{-1}$ satisfait ${}^*(\hat{A}_n - A) = {}^*\tilde{A}_n = S_{n-1}^{-1} [M_n - Q\, {}^*A]$ avec $M_n = \sum_{k=1}^{n} X_{k-1} {}^*\varepsilon_k$.

On étudie aussi l'erreur de prédiction $\pi_n = \check{A}_n X_n$ et les sommes associées $P_n = \sum_{k=1}^{n} \pi_k {}^*\pi_k$, ainsi que la covariance empirique :

$$\hat{\Gamma}_n = \frac{1}{n} \sum_{k=2}^{n} (X_k - \hat{A}_{k-1} X_{k-1}) {}^*(X_k - \hat{A}_{k-1} X_{k-1}).$$

Les résultats qui suivent seront liés aux modules des valeurs propres de A; on note $\bar{\alpha}$ le plus grand de ces modules et $\underline{\alpha}$ le plus petit. Le modèle est dit « *stable* » si $\bar{\alpha} > 1$, « *explosif* » si $\underline{\alpha} < 1$, « *instable* » si $\bar{\alpha} = \underline{\alpha} = 1$; $\bar{\alpha}$ est le rayon spectral de A, $\underline{\alpha}$ est l'inverse de celui de A^{-1}.

Posons $C_m = \sum_{k=0}^{m} A^k \Gamma {}^*A^k$. Le modèle est « *commandable* » si $\lambda_{\min} C_m$ est >0 pour m assez grand.

On dira que le modèle est « *singulier* » s'il existe une valeur propre de A de module >1 associée à un sous-espace propre de dimension $\geqq 2$. Le modèle est « *régulier* » sinon.

I.3. Énoncé des résultats

Notre résultat général est le suivant. D'après la remarque faite en I.1, on ne peut pas espérer la consistance de l'estimateur des moindres carrés dans le cas singulier. Les notations et hypothèses sont celles qui ont été données en I.2.

THÉORÈME 1. — CAS MIXTE RÉGULIER. — *Soit un modèle autorégressif d'ordre* 1 *vectoriel, commandable et régulier.*

(A) *Le bruit satisfait l'hypothèse* (B2).

1. *L'estimateur des moindres carrés* $(\hat{A}_n)$ *de* A *est fortement consistant. Sa vitesse de convergence presque sûre est toujours de l'ordre suivant :*

$$\| \hat{A}_n - A \| = 0 (\operatorname{Log} n/n)^{1/2}.$$

2. *La suite* $({}^*X_n S_{n-1}^{-1} X_n)$ *converge, p.s., vers une v.a.* $\rho < 1$, *et, p.s. :*

$$\frac{1}{n} \sum_{k=1}^{n} \pi_k {}^*\pi_k \to \rho \Gamma, \qquad \hat{\Gamma}_n \to (1+\rho) \Gamma;$$

$([1 - {}^*X_n S_n^{-1} X_n] \hat{\Gamma}_n)$ *est un estimateur fortement consistant de* Γ.

(B) *Si le bruit satisfait l'hypothèse* (BB), *les résultats ci-dessus restent vrais lorsque* A *n'a pas de valeur propre de module* 1. *Dans le cas général,* *ils restent vrais en probabilité, en particulier* $(\hat{A}_n)$ *et* $([1 - {}^*X_n S_n^{-1} X_n]\hat{\Gamma}_n)$ *sont des estimateurs faiblement consistants de* A *et de* Γ.

THÉORÈME 2. — CAS SINGULIER. — *On suppose le modèle autorégressif commandable et singulier. Le bruit satisfait l'hypothèse* (BB) *ou* (B2).

1. *L'estimateur* $(\hat{A}_n)$ *peut être non consistant.*

2. *On a* $\frac{1}{n} \operatorname{Log} s_n \xrightarrow{\text{p.s.}} 2 \operatorname{Log} \bar{\alpha}$ *et, p.s. :*

$$0 < \liminf \frac{1}{n} \lambda_{\min} S_n \leq \limsup \frac{1}{n} \lambda_{\min} S_n < \infty. \quad \blacksquare$$

On peut dans les cas stable, instable, ou explosif obtenir des résultats plus précis. On désigne par a une fonction réelle croissant vers ∞ et telle que $\int^{\infty} [1/a]$ converge.

THÉORÈME 3. — CAS STABLE ET INSTABLE. — *Le modèle autorégressif est supposé commandable; le bruit satisfait l'hypothèse* (BB) *ou* (B2).

(1) *S'il est stable* $(\bar{\alpha} < 1)$, *on a, p.s. :*

$$S_n/n \to C, \qquad C \text{ inversible}; \qquad \|X_n\| = o(n^{1/2}).$$

Et si $E(\|X_0\|^2) < \infty$:

$$E(\|X_n\|^2) = O(1).$$

(2) *Si* $\bar{\alpha} = 1$ *et si l'ordre le plus grand des zéros de module* 1 *du polynôme minimal de* A *est* ν, *alors, p.s. :*

$$\limsup (n^{2\nu - 1} \operatorname{Log} \operatorname{Log} n)^{-1/2} \|X_n\| \leq \text{Cte} \quad \textit{et} \quad s_n = O(n^{2\nu} \operatorname{Log} \operatorname{Log} n).$$

De plus si $E(\|X_0\|^2) < \infty$:

$$E(\|X_n\|^2) = O(n^{2\nu - 1}).$$

(3) *Si le modèle est non explosif* $(\bar{\alpha} \leq 1)$ *alors, p.s., l'estimateur* $(\hat{A}_n)$ *est fortement consistant et satisfait :*

$$\|\hat{A}_n - A\| = o([a(\operatorname{Log} n)]/n)^{1/2}.$$

(4) *Si* $\bar{\alpha} < 1$, *p.s.,*

$$(f_n) \to 0, \qquad \sum [a(\operatorname{Log} n)]^{-1} \|\pi_n\|^2 < \infty;$$

l'estimateur empirique de la covariance est fortement consistant.

Si le bruit est blanc, p.s. :

$$\limsup (n/\operatorname{Log} \operatorname{Log} n)^{1/2} \|\hat{A}_n - A\| \leq (2 \lambda_{\max} C \lambda_{\max} \Gamma)^{1/2}.$$

(5) *Si* $\bar{\alpha} \leqq 1$, *et si le bruit satisfait* (B2), *alors, p. s.* :

$$\|\hat{A}_n - A\| = O(\text{Log } n/n)^{1/2};$$
$$^*X_n S_n^{-1} X_n \to 0;$$
$$(\text{Log } n)^{-1} (\| S_{n-1}^{1/2}\, ^*(\hat{A}_n - A) \|^2 + \| P_n \|) \leqq \text{Cte}.$$

L'estimateur empirique de la covariance est fortement consistant.

Si en outre le bruit est blanc, on a dans le cas stable, p. s.,

$$\limsup (\text{Log } n)^{-1} \| P_n \| \leqq \text{Cte}.$$

Et dans le cas instable la suite $[(\text{Log } n)^{-1} \| P_n \|]$ *est tendue.*

(6) *Si le bruit satisfait l'hypothèse* (BB), *les résultats donnés en* (5) *restent vrais en probabilité.*

THÉORÈME 4. — CAS EXPLOSIF. — *On suppose ici le modèle autorégressif commandable et* $\bar{\alpha} > 1$, *le bruit satisfaisant l'hypothèse* (BB) *ou* (B2).

(1) *Si l'ordre le plus grand des zéros de module* $\bar{\alpha}$ *du polynôme minimal de* A *est* ν : $[\bar{\alpha}]^{-2n} n^{-2\nu+2} s_n$ *tend vers une v. a. non nulle,*

$$\frac{1}{n} \text{Log } s_n \xrightarrow{\text{p. s.}} 2 \text{ Log } \bar{\alpha};$$

si $E(\| X_0 \|^2) < \infty$,

$$E(\| X_n \|^2) = O(\bar{\alpha}^{2n} n^{2\nu-2}).$$

(2) *Pour* $Z = \sum_{k=1}^{\infty} A^{-k} \varepsilon_k$:

$$A^{-n} X_n \xrightarrow{\text{p. s.}} X_0 + Z = \hat{Z}.$$

La loi de Z *est centrée, de carré intégrable avec une covariance inversible; pour tout état initial, la loi de* $\hat{Z}$ *ne charge aucun hyperplan. En outre* :

$$G_n = A^{-n} S_{n-1}\, ^*A^{-n} \xrightarrow{\text{p. s.}} G = \sum_{k=1}^{\infty} A^{-k} \hat{Z}\, ^*\hat{Z}\, ^*A^{-k}$$

où G *est une matrice hermitienne telle que, pour tout u non nul,* $^*u G u$ *est p. s.* > 0.

3. *Dans le cas explosif régulier,* G *est p. s. inversible et, p. s.* :

$$\lim \frac{1}{n} \text{Log } \lambda_{\min} S_n = 2 \text{ Log } \underline{\alpha};$$
$$\| (\hat{A}_n - A) A^n \| = o[n^{1/2}];$$
$$^*X_n S_{n-1}^{-1} X_n \to \rho = {}^*\hat{Z} G^{-1} \hat{Z};$$
$$n^{-1} \sum_{k=1}^{n} \pi_k\, ^*\pi_k \to \rho \Gamma \qquad \text{et} \qquad \hat{\Gamma}_n \to (1+\rho)\Gamma. \quad \blacksquare$$

Enfin, on peut traduire ce qui précède aux modèles autorégressifs d'ordre p et de dimension d, $AR_d(p)$.

THÉORÈME 5. — *Soit un modèle autorégressif d'ordre p de dimension d :*

$$X_n = A_1 X_{n-1} + \ldots + A_p X_{n-p} + \varepsilon_n,$$

où les $A_1, \ldots, A_p$ *sont des matrices* $d \times d$, *et le bruit* ε *a la covariance* Γ *inversible.*

On note $A(z) = I - A_1 z - \ldots - A_p z^p$. *Le modèle est dit « singulier » s'il existe un z de module* <1 *pour lequel* $A(z)$ *a un noyau de dimension* >1. *Il est* « régulier » *s'il n'est pas singulier, en particulier si* $d=1$ *ou si* $\operatorname{Det} A(z)$ *ne s'annule que pour* $|z| \geqq 1$.

Soit $\hat{A}_n = (\hat{A}_{1,n}, \ldots, \hat{A}_{p,n})$ *l'estimateur des moindres carrés de* $A = (A_1, \ldots, A_p)$.

(1) *Si le modèle est régulier, l'estimateur* $(\hat{A}_n)$ *de* A *est consistant, fortement sous l'hypothèse* (B2) *ou sous* (BB) *lorsque* $A(z)$ *est inversible pour z de module* 1, *faiblement sous* (BB) *dans le cas général.*

(2) *Si* $\operatorname{Det} A(z)$ *ne s'annule pour aucun z de module* <1, *p.s., les résultats relatifs à* $(\hat{A}_n)$ *donnés dans le théorème* 3 *sont valables.*

(3) *Si le plus grand des modules des zéros de* $\operatorname{Det} A(z)$ *est strictement majoré par* $\beta < 1$, *si* A_p *est inversible et si le modèle est régulier, alors, p.s. :*

$$\| \hat{A}_n - A \| = o(n^{1/2} \beta^n). \quad \blacksquare$$

On verra en II.9 comment transcrire plus précisément aux modèles $AR_d(p)$ les résultats des théorèmes 1 à 4 portant sur les modèles $AR_d(1)$.

Commentaires. — (*a*) Ce travail généralise ainsi en dimension d et sous l'hypothèse (BB) celui de Lai et Wei [8] dans le cadre des modèles $AR_1(p)$ avec l'hypothèse (B2); même dans ce cadre, les vitesses de convergence presque sûre sont précisées. L'étude des convergences en loi des mêmes estimateurs est effectuée, pour les modèles $AR_1(p)$ instables dans [3] et [4], et pour les modèles $AR_d(1)$ généraux dans [14] et [15].

(*b*) En ce qui concerne les erreurs de prédiction, nous généralisons l'étude effectuée par Wei [18] pour des modèles $AR_1(p)$ stables ou instables ou $AR_1(1)$ explosifs.

II. DÉMONSTRATIONS

II.1. Rappels et complément

Considérons un modèle de régression complexe de dimension d adapté à la filtration $\mathbb{F}$:

$$X_{n+1} = {}^*R\,\Phi_n + \varepsilon_{n+1};$$

(Φ_n) est une suite adaptée à $\mathbb{F}$ de carré intégrable, R une matrice $d \times d$ complexe. Posant, pour Q hermitienne définie positive :

$$S_n = Q + \sum_{k=0}^{n} \Phi_k\,{}^*\Phi_k, \qquad s_n = \operatorname{tr} S_n,$$

$$f_n = {}^*X_n\,S_n^{-1}\,X_n, \qquad \varphi_n = \sum_{k=1}^{n} f_k,$$

l'estimateur des moindres carrés $(\hat{R}_n)$ de R, l'erreur de prédiction $\pi_n = {}^*[\hat{R}_n - R]\,\Phi_n$, et la covariance empirique

$$\hat{\Gamma}_n = \frac{1}{n} \sum_{k=2}^{n} [X_k - {}^*\hat{R}_{k-1} X_{k-1}]\,{}^*[X_k - {}^*\hat{R}_{k-1} X_{k-1}],$$

satisfont la proposition suivante.

PROPOSITION A. — (1) *Sous l'hypothèse* (BB) *ou* (B2), *si* (s_n) *tend, p.s. vers* ∞, *on a, presque sûrement, pour toute fonction a croissant vers l'ifini telle que* $\int^{\infty} [1/a]$ *converge :*

$$\| S_{n-1}^{1/2} (\hat{R}_n - R) \| = o\,[a(\operatorname{Log} s_{n-1})]^{1/2};$$
$$\sum [a(\operatorname{Log} s_n)]^{-1} (1 - f_n) \| \pi_n \|^2 < \infty.$$

(2) *Sous l'hypothèse* (B2), *pour tout* $u \in \mathbb{C}^d$, *posant :*

$$T_n(u) = \| S_{n-1}^{1/2} (\hat{R}_n - R)\,u \|^2 + \sum_{k=1}^{n} (1 - f_k)\,|\langle u, \pi_k \rangle|^2,$$

on a, p.s., $\lim [\varphi_n]^{-1}\,T_n(u) \to {}^*u\,\Gamma\,u(1+\rho)$ *où* ρ *est une v.a. finie, nulle si* $\varphi_n \to \infty$.

(3) *Sous* (BB) *le résultat de* (2) *reste vrai si* (f_n) *tend, p.s., vers* $f_\infty > 0$. *Il est aussi valable pour la convergence en probabilité lorsque la suite de processus* $(nf_{[nt]};\ 0 \leqq t \leqq 1)$ *converge en loi vers un processus continu non nul.*

(4) *Si* $n^{-1} \sum_{k=1}^{n} \pi_k {}^*\pi_k$ *tend p.s. [ou en probabilité] vers une matrice aléatoire* P *finie, alors* $(\hat{\Gamma}_n)$ *tend p.s. [ou en probabilité] vers* $\Gamma + \mathrm{P}$.

Démonstration. — La partie (1) est prouvée en [11] et [5]. La partie (4) résulte d'une modification élémentaire de [5]. La preuve de (2) et de (3) repose sur la relation suivante, prouvée en [7] et [18] :

$$\mathrm{T}_n(u) - \mathrm{T}_1(u) = \sum_{k=1}^{n-1} f_k \,|\langle \varepsilon_{k+1}, u \rangle|^2 + \zeta_n,$$

où (ζ_n) est une suite qui converge toujours p.s., la limite étant nulle si $\varphi_\infty = \lim \varphi_n = \infty$. Supposant $\varphi_\infty = \infty$, on conclut sous (B2) par une propriété classique du terme de droite (*cf.* [7]); sous (BB), la loi des grands nombres donne le même résultat lorsque (f_n) tend vers $f_\infty > 0$. Sous les hypothèses faites dans la seconde partie de (3) :

$$[\varphi_n]^{-1} \sum_{k=1}^{n} f_k (|\langle \varepsilon_{k+1}, u \rangle|^2 - {}^*u \Gamma u) \xrightarrow{\mathrm{P}} 0;$$

on le voit en considérant la somme comme une intégrale stochastique et en appliquant les résultats de [6]. ■

II.2. Commandabilité et excitation

On a :

$$\mathrm{X}_{n+1} = \mathrm{A}\mathrm{X}_n + \varepsilon_{n+1}$$
$$\mathrm{X}_{n+2} = \mathrm{A}^2 \mathrm{X}_n + \varepsilon_{n+2} + \mathrm{A}\, \varepsilon_{n+1}$$
$$\mathrm{X}_{n+m} = \mathrm{A}^m \mathrm{X}_n + \eta_{n,\, n+m}$$

où $\eta_{n,\, n+m} = \varepsilon_{n+m} + \mathrm{A}\, \varepsilon_{n+m-1} + \ldots + \mathrm{A}^{m-1} \varepsilon_{n+1}$ satisfait :

$$\mathrm{E}(\eta_{n,\, n+m} {}^*\eta_{n,\, n+m} \,|\, \mathscr{F}_n) = \mathrm{C}_{m-1}$$

en posant

$$\mathrm{C}_m = \sum_{k=0}^{m} \mathrm{A}^k \Gamma {}^*\mathrm{A}^k.$$

La matrice hermitienne C_n est inversible si, et seulement si, le rang de $(\Gamma^{1/2}, \mathrm{A}\Gamma^{1/2}, \ldots, \mathrm{A}^n \Gamma^{1/2})$ vaut d. Si δ est le degré du polynôme minimal de A, il est donc équivalent de dire que $\lim_{n \to \infty} \lambda_{\min} \mathrm{C}_n > 0$, ou que $\mathrm{C}_{\delta-1}$ est inversible, ou que C_{d-1} est inversible. Le modèle autorégressif est alors « commandable ».

Soit $r<\delta$ et $Y_n = X_{n\delta+r}$:

$$Y_{n+1} = A^{\delta} Y_n + \chi_{n+1} \qquad \text{où} \quad E(\chi_{n+1}{}^*\chi_{n+1} \mid \mathscr{F}_{n\delta+r}) = C_{\delta-1}.$$

Si le modèle initial est commandable, (Y_n) est un nouveau modèle autorégressif dont le bruit (χ_n) a une covariance inversible.

Proposition B. — (1) *Si le modèle est* « stable » *c'est-à-dire si le rayon spectral de* A *est* <1, *alors la série* $\sum_{k=1}^{\infty} A^k \varepsilon_k$ *converge p.s. et en moyenne quadratique vers une v. a.* L *de carré intégrable. Si le modèle est stable et commandable et si* A *est inversible, alors, pour tout vecteur aléatoire* Z *supposé* $\mathscr{F}_0$*-mesurable et de dimension d,* $Z+L$ *ne charge aucun hyperplan.*

(2) *Si le modèle est commandable, alors, la suite* (X_n) *est* « fortement excitante » *c'est-à-dire qu'il existe une constante* $\rho>0$ *telle que, presque sûrement,* $\liminf \frac{1}{n} \lambda_{\min} S_n \geq \rho$.

La partie (1) de cette proposition est un résultat classique dû à Paul Lévy lorsque le bruit est blanc. Dans le cas où le bruit est un accroissement de martingale ayant un moment conditionnel d'ordre >2, elle est prouvée dans [9]. La partie (2) est connue ([10], [16]).

II.3. Décomposition

Étant donnés r entiers $d(1), \ldots, d(r)$ de somme d, notant $b(j)$ la somme des j premiers, $b(0)=0$. Pour $u \in \mathbb{C}^d$ on peut envisager la décomposition $u=[u(1), \ldots, u(r)]$ où $u(j)$ est le vecteur de $\mathbb{C}^{d(j)}$ formé par les composantes d'indices allant de $b(j-1)+1$ à $b(j)$ de u. Si $A(j)$ est, pour $1 \leq j \leq r$, une matrice $d(j) \times d(j)$, alors on note $\mathrm{Diag}[A(1), \ldots, A(r)]$ la matrice formée en plaçant successivement les blocs $A(j)$ sur la diagonale, les autres termes étant nuls.

Soit P une matrice $d \times d$ inversible. Le modèle transformé par P s'écrit $PX_{n+1} = PAP^{-1} PX_n + P\varepsilon_{n+1}$ où, notant $y^{\tau} = Py$ pour chaque vecteur et $A^{\tau} = PAP^{-1}$: $X^{\tau}_{n+1} = A^{\tau} X^{\tau}_n + \varepsilon^{\tau}_{n+1}$.

Considérons ce nouveau modèle autorégressif (X^{τ}). Soit $\hat{A}^{\tau}_n$ l'estimateur des moindres carrés de A^{τ}. Les expressions sont les mêmes que pour le modèle initial

$$Q^\tau = PQ\,{}^*P, \qquad S_n^\tau = PS_n\,{}^*P = Q^\tau + \sum_{k=0}^{n} X_k^\tau\,{}^*X_k^\tau,$$

$$M_n^\tau = PM_n\,{}^*P;$$

$${}^*\check{A}_n^\tau = (S_{n-1}^\tau)^{-1}(M_n^\tau - Q^\tau\,{}^*A^\tau) = {}^*(P\,\check{A}_n\,P^{-1});$$

$$\| S_{n-1}^\tau\,{}^*\check{A}_n^\tau \| = \| PS_{n-1}\,\check{A}_n\,{}^*P \|; \qquad {}^*X_n^\tau (S_n^\tau)^{-1} X_n^\tau = f_n^\tau = f_n.$$

Les suites suivantes sont, p. s., du même ordre, c'est-à-dire que leur rapport, a, p. s., une limite inférieure >0 et une limite supérieure $<\infty$:

$$(\lambda_{\min} S_n) \qquad \text{et} \qquad (\lambda_{\min} S_n^\tau),$$

ou

$$(\lambda_{\max} S_n) \qquad \text{et} \qquad (\lambda_{\max} S_n^\tau).$$

Donc $\hat{A}_n^\tau \xrightarrow{\text{p. s.}} A^\tau$ équivaut à $\hat{A}_n \xrightarrow{\text{p. s.}} A$ et les résultats relatifs aux vitesses de convergence ou aux suites (f_n) que nous formulerons seront les mêmes pour X et pour X^τ.

Notons que notre transformation est un artifice mathématique destiné à analyser le modèle; il ne s'agit pas nécessairement d'une transformation observable.

On peut en particulier prendre la décomposition de Jordan de A. On appelle « *bloc de Jordan* » soit un scalaire, soit une matrice carrée Δ d'ordre >1, dont les termes de la diagonale sont égaux à $\lambda \in \mathbb{C}$, tous les termes situés immédiatement au-dessous étant égaux à 1, les autres termes étant nuls; si une telle matrice est de dimension $q \times q$, ses polynômes caractéristique et minimal sont $(x-\lambda)^q$. On peut choisir une matrice de changement de base P telle que PAP^{-1} soit une matrice formée par r blocs de Jordan carrés $\Delta(j)$, $1 \leqq j \leqq r$, placés sur la diagonale et des O ailleurs et pour chaque j, $\Delta(j)$ de dimension $d(j) \times d(j)$. Alors

$$PAP^{-1} = \mathrm{Diag}[\Delta(j);\ 1 \leqq j \leqq r] = A^\tau.$$

Le système $X_{n+1}^\tau = A^\tau X_n^\tau + \varepsilon_{n+1}^\tau$ se décompose en r « *composantes* » :

$$X_{n+1}^\tau(j) = \Delta(j)\,X_n^\tau(j) + \varepsilon_{n+1}^\tau(j)$$

où chaque vecteur y est représenté par $(y(j);\ 1 \leqq j \leqq r)$, $y(j)$ de même dimension que $\Delta(j)$.

Supposons le modèle commandable. Alors chacun des sous-modèles

$$X_{n+1}^\tau(j) = \Delta(j)\,X_n^\tau(j) + \varepsilon_{n+1}^\tau(j)$$

est commandable. En effet,

$$\mathrm{PC}_n{}^*\mathrm{P}=\sum_{k=0}^{n}[\mathrm{PA}^k\mathrm{P}^{-1}][\mathrm{P}\Gamma^*\mathrm{P}][{}^*\mathrm{P}^{-1}\,{}^*\mathrm{A}^k\,{}^*\mathrm{P}]=\sum_{k=0}^{n}(\mathrm{A}^\tau)^k\mathrm{P}\Gamma{}^*\mathrm{P}({}^*\mathrm{A}^\tau)^k.$$

On a $\mathrm{E}(\varepsilon_{n+1}^\tau(j)\,{}^*\varepsilon_{n+1}^\tau(j)\,|\,\mathscr{F}_n)=\Gamma^*(j)$, où $\Gamma^*(j)=\mathrm{P}(j)\Gamma^*\mathrm{P}(j)$ avec $\mathrm{P}(j)$, matrice $d(j)\times d$, telle que P soit la matrice formée par les blocs $\mathrm{P}(1),\ldots,\mathrm{P}(r)$ placés les uns au-dessus des autres. Soit $n\geqq d$ et $u\in\mathbb{C}^d$ tel que $u(j)\neq 0$ et $u(i)=0$ si $i\neq j$;

$$0<{}^*u\mathrm{PC}_n{}^*\mathrm{P}u={}^*u(j)\,\mathrm{C}_n^\tau(j)\,u(j),$$

avec $\mathrm{C}_n^\tau(j)=\sum_{k=0}^{n}(\Delta(j))^k\Gamma^\tau(j)({}^*\Delta(j))^k$. Ce qui établit la commandabilité du sous-modèle.

On connaît dans l'analyse des erreurs de prédiction l'importance de la suite (f_n), $f_n={}^*\mathrm{X}_n\mathrm{S}_n^{-1}\mathrm{X}_n$. Le lemme technique suivant sera utile pour déduire des comportements globaux du modèle de l'analyse des composantes.

LEMME C. — *Soit* $f_n={}^*\mathrm{X}_n\mathrm{S}_n^{-1}\mathrm{X}_n$. *On suppose* $\lambda_{\min}\mathrm{S}_n\to\infty$.

(1) *La convergence p.s. vers* 0 *de* (f_n) *implique celle de*

$$\alpha_n=(\sup_{k\leqq n}{}^*\mathrm{X}_k\mathrm{S}_n^{-1}\mathrm{X}_k).$$

(2) *On suppose que, dans le modèle autorégressif* $\mathrm{A}=\mathrm{Diag}(\mathrm{A}_1,\mathrm{A}_2)$ *pour* A_1 *et* A_2 *de dimensions* $p\times p$ *et* $(d-p)\times(d-p)$. *On considère les sous-modèles :*

$$\mathrm{X}_{n+1}(j)=\mathrm{A}_j\mathrm{X}_n(j)+\varepsilon_{n+1}(j)\qquad pour\quad j=1,2,$$

auxquels on associe $\mathrm{S}_n(j), f_n(j)$ *ou* $\alpha_n(j)$ *définis comme pour le modèle initial.*

Si les hypothèses suivantes sont vérifiées

(*a*) $\alpha_n(1)\to 0$, *p.s.* [*ou en probabilité*]

(*b*) *pour une suite* $(\Gamma_n(2))$ *de matrices aléatoires de dimension* $(d-p)\times(d-p)$ *inversibles, on a p.s. :*

$$\sum_{k=0}^{n}\|[\Gamma_n(2)]^{-1}\mathrm{X}_k(2)\|=O(1)\qquad et\qquad [\Gamma_n(2)]^{-1}\mathrm{S}_n(2)\,{}^*[\Gamma_n(2)]^{-1}\to\mathrm{I},$$

alors, posant $\Gamma_n=\mathrm{Diag}(\mathrm{S}_n^{1/2}(1),\Gamma_n(2))$, *on a p.s.* [*ou en probabilité*]

$$\Gamma_n^{-1}\mathrm{S}_n{}^*\Gamma_n^{-1}\to\mathrm{I},$$
$$(f_n-\|[\Gamma_n(2)]^{-1}\mathrm{X}_n(2)\|^2)\to 0.$$

Démonstration. — (1) Soit

$$\alpha_n=\sup_{1\leqq k\leqq n}{}^*\mathrm{X}_k\mathrm{S}_n^{-1}\mathrm{X}_k\qquad\text{et}\qquad\alpha_{\mathrm{N},n}=\sup_{\mathrm{N}\leqq k\leqq n}{}^*\mathrm{X}_k\mathrm{S}_n^{-1}\mathrm{X}_k.$$

On a

$$f_n \leqq \alpha_{N,n} \leqq \sup \{ \alpha_{N,n-1}, f_n \}.$$

Cela signifie que $\alpha_{N,n}$ ne croît qu'aux indices d'échelle de la suite $(f_n)_{n \geqq N}$ et que, pour $n \geqq N$, $\alpha_{N,n} \leqq f_N^{\#} = \sup_{n \geqq N} f_n$.

En outre, la convergence vers 0 de (f_n) implique celle de $(f_N^{\#})$ donc celle de $(\alpha_{N,\infty})$. Or, pour $n \geqq N$, $\alpha_n \leqq \alpha_{N,\infty} + \sup_{1 \leqq k \leqq N} {}^*X_k S_n^{-1} X_k$.

On en déduit facilement la propriété (1).

(2) On a $D_n = \Gamma_n^{-1} S_n {}^*\Gamma_n^{-1} = \begin{bmatrix} I & \Sigma_n \\ {}^*\Sigma_n & H_n \end{bmatrix}$ où

$$\| \Sigma_n \| = \left\| \sum_{k=0}^{n} [S_n(1)]^{-1/2} X_k(1) {}^*X_k(2) {}^*[\Gamma_n(2)]^{-1} \right\|$$

$$\leqq (\alpha_n(1))^{1/2} \sum_{k=0}^{n} \| \Gamma_n(2)]^{-1} X_k(2) \|.$$

Le majorant tend p. s. [en probabilité] vers 0. Comme $H_n \overset{\text{p. s.}}{\to} I$, on a la première propriété annoncée. En outre :

$$f_n = {}^*X_n S_n^{-1} X_n = {}^*X_n {}^*\Gamma_n^{-1} D_n^{-1} \Gamma_n^{-1} X_n;$$

$$| f_n - \| \Gamma_n^{-1} X_n \|^2 | = | f_n - f_n(1) - {}^*X_n(2) {}^*\Gamma_n^{-1}(2) \Gamma_n^{-1}(2) X_n(2) |$$

$$\leqq \| I - D_n^{-1} \| (f_n(1) + \| \Gamma_n^{-1}(2) X_n(2) \|^2).$$

D'où le dernier résultat annoncé dans (2). ■

II.4. Analyse des composantes

Dans cette partie, on étudie un modèle autorégressif de dimension d, $d \geqq 1$, $X_{n+1} = AX_n + \varepsilon_{n+1}$, pour lequel A est un bloc de Jordan associé à la valeur propre de module α, conjuguée de λ.

Comportement asymptotique des trajectoires

Soit (e^i) la base canonique; A étant un bloc de Jordan, on a :

$${}^*A e^1 = \lambda e^1, \qquad \text{et, pour } j > 1, \quad {}^*A e^j = \lambda e^j + e^{j-1}.$$

D'où : $*A^k e^1 = \lambda^k e^1$, et, pour $j>1$ et $\alpha \neq 0$:

$$*A^k e^j = \lambda^k e^j + \ldots + \binom{k}{i} \lambda^{k-i} e^{j-i} + \ldots + \binom{k}{j-1} \lambda^{k-j+1} e^1,$$

en posant $0 = \binom{k}{i}$ si $k<i$.

D'où :

$$\langle X_n, e^j \rangle - \langle X_0, *A^n e^j \rangle = \sum_{k=1}^{n} \langle \varepsilon_k, *A^{n-k} e^j \rangle = T_n(j, 0) + \ldots + T_n(j, j-1)$$

avec

$$T_n(j, i) = \sum_{k=1}^{n} \binom{n-k}{i} \lambda^{n-k-i} \langle \varepsilon_k, e^{j-i} \rangle.$$

Puisque

$$n^{-1} \sum_{k=1}^{n} \| \varepsilon_k \|^2 \xrightarrow{\text{p. s.}} \operatorname{tr} \Gamma, \; n^{-1/2} \sup_{k \leqq n} \| \varepsilon_k \| \xrightarrow{\text{p. s.}} 0;$$

et, p. s. :

$$\begin{aligned} \| X_n \| &= o(n^{1/2}) \quad \text{si} \quad \alpha<1, \\ &= o(n^{d+1/2}) \quad \text{si} \quad \alpha=1, \\ &= o(n^{d-1/2} \alpha^n) \quad \text{si} \quad \alpha>1; \end{aligned}$$

$$E(| T_n(j, i) |^2) = \sum_{k=1}^{n} \binom{n-k}{i}^2 \alpha^{2n-2k-2i} \, {}^*e^{j-i} \Gamma e^{j-i}.$$

Par une inégalité classique des martingales, si $E(\| X_0 \|^2) < \infty$:

$$\begin{aligned} E(\| X_n \|^2) &= O(1) \quad \text{si} \quad \alpha<1, \\ &= O(n^{2d-1}) \quad \text{si} \quad \alpha=1, \\ &= O(\alpha^{2n} n^{2d-2}) \quad \text{si} \quad \alpha>1. \end{aligned}$$

On obtient, p. s.,

$$\begin{aligned} s_N &= O(N) \quad \text{si} \quad \alpha<1, \\ &= O(N^{2d+2}) \quad \text{si} \quad \alpha=1, \\ &= O(\alpha^{2N} N^{2d-1}) \quad \text{si} \quad \alpha>1. \end{aligned}$$

On peut aussi écrire, lorsque $\alpha \neq 0$:

$$A^{-n} X_n = X_0 + \sum_{k=1}^{n} A^{-k} \varepsilon_k.$$

Or il existe une base $(f^1, \ldots, f^d)$ de $\mathbb{C}^d$ telle que :

$$*A^{-1} f^1 = \lambda^{-1} f^1 \qquad \text{et, pour } i>1, \quad *A^{-1} f^i = \lambda^{-1} f^i + f^{i-1}.$$

Par conséquent :

$$\lambda^{-n} \langle X_n, f^1 \rangle = Y_n^1 = \langle X_0, f^1 \rangle + \sum_{k=1}^{n} \lambda^{-k} \langle \varepsilon_k, f^1 \rangle;$$

$$\lambda^{-n} \langle X_n, f^i \rangle + \ldots + \lambda^{-n+i-1} \binom{n}{i-1} \langle X_n, f^1 \rangle = Y_n^i$$

$$= \langle X_0, f^i \rangle + \sum_{k=1}^{n} \left[\lambda^{-k} \langle \varepsilon_k, f^i \rangle + \ldots + \lambda^{-k+i-1} \binom{k}{i-1} \langle \varepsilon_k, f^1 \rangle \right].$$

Analyse des composantes instables

En appliquant la loi du logarithme itéré relative aux v. a. indépendantes lorsque le bruit est blanc, ou le corollaire 6 de [5], s'il a un moment conditionnel d'ordre >2 fini, on obtient, p. s. :

$$\limsup \, (n^{2i-1} \operatorname{Log} \operatorname{Log} n)^{-1/2} \, | Y_n^i | \leqq \text{Cte},$$
$$\limsup \, (n^{2d-1} \operatorname{Log} \operatorname{Log} n)^{-1/2} \, \| X_n \| \leqq \text{Cte}.$$

Analyse des composantes explosives

Dans le cas explosif, les suites (Y_n^i) convergent p. s. vers des variables aléatoires finies Z^i. Donc, p. s.

$$| \lambda^{-n} n^{-i+1} \langle X_n, f^i \rangle | \to | [\lambda^{i-1}/(i-1)\,!] \, Z^1 |,$$

et

$$\alpha^{-n} n^{-d+1} \| X_n \| \to \text{Cte} \, | Z^1 |.$$

Soit $u \neq 0$ de $\mathbb{C}_d$, $u = \sum u^i f^i$. On a

$$\liminf \alpha^{-2n} | \langle X_n, u \rangle |^2 \geqq \| u \|^2 \, | Z^1 |^2$$

avec égalité si $u = f^1$;

$$\limsup \, \alpha^{-2n} n^{-2d+2} | \langle X_n, u \rangle |^2 \leqq \text{Cte} \, \| u \|^2 \, | Z^1 |^2,$$

avec égalité si $u = f^d$.

D'après la proposition B, si le modèle est commandable, la variable aléatoire $Z^1 = \langle X_0, f^1 \rangle + \sum_{k=1}^{\infty} \lambda^{-k} \langle \varepsilon_k, f^1 \rangle$ a une loi diffuse; donc, p.s. $\alpha^{-2n} n^{-2d+2} s_n$ tend vers une v.a. constante $\neq 0$;

$$\liminf \alpha^{-2n} \lambda_{\min} S_n > 0;$$
$$2n \operatorname{Log} \alpha + \mathrm{Cte} \leqq \operatorname{Log} \lambda_{\min} S_n \leqq \operatorname{Log} \lambda_{\max} S_n \leqq 2n \operatorname{Log} \alpha + \mathrm{Cte}.\operatorname{Log} n;$$
$$\lim \frac{1}{n} \operatorname{Log} \lambda_{\min} S_n = \lim \frac{1}{n} \operatorname{Log} \lambda_{\max} S_n = 2 \operatorname{Log} \alpha.$$

II.5. Cas stable ou instable : démonstration du théorème 3

(*a*) En considérant la décomposition de Jordan, on obtient les propriétés de X_n et de s_n données dans les parties (1) et (2).

Il en résulte, $\operatorname{Log} s_n = O(\operatorname{Log} n)$; comme $n = O(\lambda_{\min} S_n)$, la vitesse de convergence de l'estimateur des moindres carrés donnée dans la partie (3) résulte de la partie (1) de la proposition A.

(*b*) Dans le cas stable, $f_n \leqq \| X_n \|^2 / \lambda_{\min} S_n \overset{\text{p. s.}}{\to} 0$. On en déduit le début de la propriété (4).

La propriété $S_n/n \overset{\text{p. s.}}{\to} C$ est classique et résulte de la stabilité lorsque le bruit est blanc. Lorsqu'il s'agit d'un bruit général avec un moment d'ordre >2, hors d'un ensemble N négligeable de trajectoires, $s_n = O(n)$ donc $[S_n/n]$ est bornée, et $\| X_n \|^2 = o(n)$ donc $[(S_{n+1} - S_n)/n]$ tend vers 0. Or :

$$S_{n+1} - S_0 = A(S_n - Q)^* A + \sum_{k=1}^{n+1} \varepsilon_k {}^* \varepsilon_k + M_{n+1} + {}^* M_{n+1},$$

où $M_n = \sum_{k=1}^{n} A X_{k-1} {}^* \varepsilon_k$; pour $u \in \mathbb{C}^d$, ${}^* u M_n u$ est une martingale de processus croissant ${}^* u A (S_{n-1} - Q) {}^* A u {}^* u \Gamma u$. On en déduit, hors d'un ensemble négligeable N_1 de trajectoires :

$$[S_{n+1}/n] - A [S_n/n]^* A \overset{\text{p. s.}}{\to} \Gamma.$$

Hors de l'ensemble négligeable $N \cup N_1$, la suite $[S_n/n]$ est bornée et toute valeur d'adhérence L satisfait :

$$L - AL^*A = \Gamma, \qquad \text{donc} \quad L = \sum_{k=0}^{\infty} A^k \Gamma^* A^k = C.$$

(c) Toujours dans le cas stable, lorsque le bruit est blanc, il y a une loi stationnaire. Prenant X_0 indépendante du bruit et ayant la loi stationnaire, la suite (X_n, ε_{n+1}) est stationnaire. La loi du logarithme itéré de Stout relative aux martingales stationnaire [13] s'applique pour u et v de norme 1 à la martingale

$$N_n = {}^*v\, S_{n-1}\, {}^*(\hat{A}_n - A)\, u + {}^*v\, Q\, {}^*A\, u = \sum_{k=1}^{n} \langle v, X_{k-1} \rangle \langle \varepsilon_k, u \rangle$$

dont la variation quadratique prévisible ${}^*v\,[S_{n-1} - Q]\, v\, {}^*u\, \Gamma\, u$ est, p. s., équivalente à $n\, {}^*v\, C\, v\, {}^*u\, \Gamma\, u$. Presque sûrement, pour LL = Log Log,

$$\limsup\, (2\, n\, \mathrm{LL}\, n)^{-1/2}\, |N_n| = ({}^*v\, C\, v\, {}^*u\, \Gamma\, u)^{1/2}.$$

Pour étendre ce fait à un état initial quelconque, notant X_n^x le modèle autorégressif tel que $X_0 = x$ et X_n celui dont la loi initiale était la loi stationnaire indépendante du bruit, on a $\| X_n^x - X_n \| \leqq \| A^n \| \, \| x - X_0 \|$; on obtient ainsi pour tout état initial :

$$\limsup\, (2\, n\, \mathrm{LL}\, n)^{-1}\, [\lambda_{\min} S_{n-1}]^2\, \| {}^*(\hat{A}_n - A)\, u \|^2 \leqq \lambda_{\max} C\, {}^*u\, \Gamma\, u \leqq \| u \|^2\, \lambda_{\max} C\, \lambda_{\max} \Gamma.$$

D'où la fin de la partie (4).

(d) Si A est inversible de rayon spectral $\leqq 1$ et si le bruit satisfait l'hypothèse (B2), alors (f_n) tend, p. s., vers 0. La preuve en est une adaptation simple de celle donnée en [7] pour les modèles $\mathrm{AR}_1(p)$. Nous ne la reproduisons pas.

Sous l'hypothèse (BB), on prouve dans [15] que le processus $(n f_{[nt]};\ 0 \leqq t \leqq 1)$ converge en loi vers un processus continu non nul; il en résulte que $\sup_{k \leqq n} f_k \geqq \alpha_n$ tend vers 0 en probabilité.

On suppose le modèle décomposé en deux composantes commandables

$$X_{n+1}(j) = A_j X_n(j) + \varepsilon_{n+1}(j) \qquad \text{pour} \quad j = 1, 2$$

et que $A = \mathrm{Diag}[A_1, A_2]$ où A_2 est nilpotente et A_1 est inversible et de rayon spectral $\leqq 1$. On peut appliquer la partie (2) du lemme C. En effet, si $A_2^m = 0$, pour $n \geqq m$:

$$X_n(2) = \sum_{k=0}^{m-1} A_2^k\, \varepsilon_{n-k}(2),$$

et

$$\| X_n(2) \| \leqq \mathrm{Cte}\, (\| \varepsilon_n(2) \| + \ldots + \| \varepsilon_{n-m+1}(2) \|) = o(n^{1/2}).$$

On en déduit $f_n(2) \overset{\text{p. s.}}{\to} 0$ et

$$\sum_{k=1}^{n} \| X_k(2) \| \leqq m \sum_{k=1}^{n} \| \varepsilon_k(2) \| + \text{Cte} = O(n), \quad \text{p. s.}$$

Le lemme C s'applique, et (f_n) tend vers 0, p. s. sous (B2) en probabilité sous (BB). La décomposition de Jordan permet toujours de se ramener à ce cas. D'où le début des parties (5) et (6) du théorème 3, en tenant compte de la proposition A.

Lorsque le bruit est blanc satisfaisant (B2) :

– dans le cas stable, la partie (4) signifie que

$$(\text{Log } n)^{-1} \| S_{n-1}^{1/2} {}^*\tilde{A}_n \| \overset{\text{p. s.}}{\to} 0;$$

– dans le cas instable, la suite $\| S_{n-1}^{1/2} {}^*\tilde{A}_n \|$ converge en loi, d'après [15], et $(\text{Log } n)^{-1} \| S_{n-1}^{1/2} {}^*\tilde{A}_n \| \overset{P}{\to} 0$.

Le théorème 3 est démontré. ■

II.6. Cas explosif et régulier : démonstration du théorème 4

(*a*) La partie (1) résulte facilement de la partie II.4 en considérant la décomposition de Jordan de A.

(*b*) Si le modèle est explosif, A est inversible. On a donc :

$$A^{-n} X_n = X_0 + \sum_{k=1}^{n} A^{-k} \varepsilon_k = X_0 + Z_n = \hat{Z}_n.$$

La covariance de Z_n est $\sum_{k=1}^{n} A^{-k} \Gamma {}^*A^{-k} = K_n$, où :

$$K_{nd} = \sum_{k=1}^{n} A^{-kd} C_{d-1} {}^*A^{-kd}.$$

Donc $K_n \to K$ matrice hermitienne finie, minorée par la matrice inversible $A^{-d} C_{d-1} {}^*A^{-d}$. On a donc $Z_n \to Z$, p. s. et en moyenne quadratique; la covariance de Z est inversible.

En outre la proposition B s'applique en tenant compte du fait que A^{-1} est inversible : la loi de $\hat{Z} = X_0 + Z$ ne charge aucun hyperplan, sa covariance sera notée $\hat{K}$.

On a :

$$G_n - \sum_{k=1}^{n} A^{-k} \hat{Z} * \hat{Z} * A^{-k} - A^{-n} Q * A^{-n}$$
$$= \sum_{k=0}^{n-1} A^{-(n-k)} A^{-k} X_k * X_k * A^{-k} * A^{-(n-k)} - \sum_{k=0}^{n-1} A^{-(n-k)} \hat{Z} * \hat{Z} * A^{-(n-k)}$$
$$= \sum_{k=0}^{n-1} A^{-(n-k)} H_k * A^{-(n-k)}$$

où (H_k) est une suite de matrices hermitiennes qui tend p. s. vers 0. On en déduit facilement que cette expression tend p. s. vers 0.

Pour tout u non nul, $*u \sum_{k=1}^{n} A^{-k} \hat{Z} * \hat{Z} * A^{-k} u$ est une suite croissante et bornée p. s. : G_n converge p. s. et en moyenne d'ordre 1 vers G, matrice aléatoire d'espérance $\sum_{k \geq 1} A^{-k} \hat{K} * A^{-k}$.

Soit u non nul; comme A^{-1} est inversible, le terme constant du polynôme caractéristique de A^{-1} n'est pas nul : I est combinaison linéaire des A^{-k} pour $0<k<d$. Donc $*u G u$ nul implique $\langle u, \hat{Z} \rangle$ nul, ce qui est p. s. impossible.

(*c*) On se place ici dans le cas régulier. Alors, le polynôme minimal de A est de degré d; il en va de même pour celui, noté T, de A^{-1}. Soit u un vecteur non nul tel que $R(A^{-1}) u = 0$ pour un polynôme R de degré $<d$; si R est un polynôme de degré le plus petit possible tel que $R(A^{-1}) u = 0$, il divise T (le reste de la division euclidienne de T par R s'annule en effet sur u). Si R est de degré $<d$, le noyau de $R(A^{-1})$ est un sous-espace strict de $\mathbb{R}^d$ et il existe un vecteur $x(R)$ orthogonal à ce noyau. Donc Z n'est p. s. pas dans ce noyau puisque $\langle \hat{Z}, x(R) \rangle$ est p. s. non nul. Or il n'y a, à un facteur multiplicatif près, qu'un nombre fini de diviseurs stricts de T. Par suite, p. s. :

$$\{\hat{Z}, A^{-1} \hat{Z}, \ldots, A^{-(d-1)} \hat{Z}\} \quad \text{est un système libre}$$

et

$$\sum_{k=1}^{d} A^{-k} \hat{Z} * \hat{Z} * A^{-k} \quad \text{est inversible.}$$

On a donc, p. s. :

$$0 < \lambda_{\min} G = \liminf \lambda_{\min} G_n < \limsup \lambda_{\max} G_n = \lambda_{\max} G < \infty$$

et :

$$S_{n-1} = A^n G_n * A^n \qquad \text{et} \qquad S_{n-1}^{-1} = * A^{-n} G_n^{-1} A^{-n}.$$

D'où :

$$\lambda_{\min} G_n^{-1} \| A^{-2n} \| \leqq \lambda_{\max} S_{n-1}^{-1} = [\lambda_{\min} S_{n-1}]^{-1} \leqq \lambda_{\max} G_n^{-1} \| A^{-2n} \|.$$

Or :

$$\frac{1}{n} \operatorname{Log}(\| A^{-2n} \|) \to 2 \operatorname{Log} \rho(A^{-1}) = -2 \operatorname{Log} \underline{\alpha},$$

et

$$\frac{1}{n} \operatorname{Log} \lambda_{\min} S_n \to 2 \operatorname{Log} \underline{\alpha}.$$

On a d'autre part $1 - f_n = (1 + g_n)^{-1}$ avec

$$g_n = {}^*X_n S_{n-1}^{-1} X_n = {}^*X_n {}^*A^{-n} {}^*A^n S_{n-1}^{-1} A^n A^{-n} X_n;$$

(g_n) converge p. s. vers $g_\infty = {}^*\hat{Z} G^{-1} \hat{Z}$; $(f_n) \overset{\text{p. s.}}{\to} f_\infty < 1$.

Enfin :

$$\| {}^*A^n {}^*(\hat{A}_n - A) \| \leqq \left\| {}^*A^n S_{n-1}^{-1} \sum_{k=1}^{n} X_{k-1} {}^*\varepsilon_k \right\| + \| {}^*A^n S_{n-1}^{-1} Q {}^*A \|$$

$$\leqq \| G_n^{-1} \| \left(\sup_{k \leqq n} \| \varepsilon_k \| \left\| \sum_{k=1}^{n} A^{-n} X_{k-1} \right\| + \text{Cte} \| A^{-n} \| \right)$$

$$\leqq \| G_n^{-1} \| \left(o(n^{1/2}) \sup_{0 \leqq j < n} \| A^{-j} X_j \| \sum_{k=1}^{n} \| A \|^{-n+k-1} + \text{Cte} \right) = o(n^{1/2}).$$

Le reste des résultats du théorème 4 résulte immédiatement de la proposition A.

II.7. Étude du cas mixte régulier : démonstration du théorème 1

On prend un modéle autorégressif où A est écrite dans sa base de Jordan. On suppose qu'il est régulier et commandable.

On peut alors considérer la décomposition en deux blocs A (1) et A (2), A (1) de rayon spectral $\leqq 1$ et A (2) explosif : pour $i = 1, 2$,

$$X_{n+1}(i) = A(i) X_n(i) + \varepsilon_{n+1}(i).$$

On pose :

$$S_n = \begin{bmatrix} S_n(1) & S_n(1, 2) \\ S_n(2, 1) & S_n(2) \end{bmatrix}, \qquad Q = \begin{bmatrix} Q(1) & Q(1, 2) \\ Q(2, 1) & Q(2) \end{bmatrix},$$

$$M_n(i) = \sum_{k=1}^{n} X_{k-1}(i) {}^*\varepsilon_k(i);$$

on déduit de l'étude des théorèmes 3 et 4 que p. s. :

$$\| S_{n-1}(1)]^{-1/2} M_n(1) \| = O[\text{Log } n]^{1/2};$$
$$\| [^*A(2)]^n [S_{n-1}(2)]^{-1} M_n(2) \| = o(n^{1/2}),$$

d'où $\| [A(2)]^{-n} M_n(2) \| = o(n^{1/2})$.

Comme, p. s. $[A(2)]^{-n-1} S_n(2)\, ^*[A(2)]^{-n-1} \to G$, G hermitienne définie positive,

$$\sum_{k=0}^{n} \| [A(2)]^{-n-1} X_k(2) \| \leq \sup_k \| [A(2)]^{-k} X_k(2) \| \sum_{k=1}^{\infty} \| [A(2)]^{-k} \| = O(1).$$

Notant $\Gamma_n = \text{Diag}[S_n^{1/2}(1), \Gamma_n(2)]$,

$$\Gamma_{n-1}^{-1} M_n = \begin{bmatrix} [S_{n-1}(1)]^{-1/2} M_n(1) & 0 \\ 0 & \Gamma_{n-1}^{-1}(2) M_n(2) \end{bmatrix}.$$

Le lemme C s'applique avec $\Gamma_n(2) = [A(2)^{n+1}] G^{1/2}$.

Donc, p. s. sous (B2) ou sous (BB) lorsque A(1) a un rayon spectral <1, et en probabilité sous (BB) lorsque A(1) a un rayon spectral égal à 1, (f_n) a la même limite f_∞, $0 < f_\infty < 1$, que $(f_n(2))$ et,

$$\Gamma_n^{-1} S_n\, ^*\Gamma_n^{-1} \xrightarrow{\text{p. s.}} I.$$

Par suite, pour $u \in \mathbb{C}^d$, avec $^*u = (^*u(1), ^*u(2))$:

$$\| ^*u\, ^*\Gamma_{n-1} S_{n-1}^{-1} M_n \|^2 = \| ^*u\, ^*\Gamma_{n-1} S_{n-1}^{-1} \Gamma_{n-1} \Gamma_{n-1}^{-1} M_n \|^2$$
$$\leq C(\| u(1) \|^2 O(\text{Log } n) + \| u(2) \|^2 o(n)),$$

C étant une v. a. finie, p. s. Notons $v = \Gamma_{n-1} u$ et α le plus petit module des valeurs propres de A (2) :

$$\| ^*v\, S_{n-1}^{-1} M_n \|^2 \leq C [\| v(1) \|^2 O(\text{Log } n/n) + \| v(2) \|^2 o(n\alpha^{-n})]$$
$$= \| v \|^2 O(\text{Log } n/n).$$

Comme $^*(\hat{A}_n - A) = S_{n-1}^{-1} [M_n - Q\, ^*A]$, on obtient le théorème 1.

II.8. Cas singulier : démonstration du théorème 2

On se place ici dans le cas singulier.

Il existe une valeur propre μ de module >1 ayant un sous-espace propre de dimension ≥ 2. En considérant v_1 *et* v_2 vecteurs propres normés et orthogonaux dans le sous-espace propre de $^*A$ associé au conjugué λ de μ, on a :

$$\langle X_n, v_i \rangle = \lambda^n \left[\langle X_o, v_i \rangle + \sum_{k=1}^{\infty} \lambda^{-k} \langle \varepsilon_n, v_i \rangle \right] - \sum_{k=1}^{\infty} \lambda^{-k} \langle \varepsilon_{n+k}, v_i \rangle$$
$$= \lambda^n \eta_i - \langle R_n, v_i \rangle.$$

Le vecteur aléatoire $\eta = \bar{\eta}_1 v_1 + \bar{\eta}_2 v_2$ ne charge, p. s., aucun hyperplan. Soit $\zeta = \eta_2 v_1 - \eta_1 v_2$. On a :

$$\langle X_n, \eta \rangle = \lambda^n \| \eta \|^2 - \langle R_n, \eta \rangle, \qquad \langle X_n, \zeta \rangle = - \langle R_n, \zeta \rangle;$$

une adaptation mineure du raisonnement effectué dans [5] en remplaçant 2 par λ, donne :

$$N^{-1} \sum_{n=1}^{N} R_n {}^*R_n \overset{\text{p. s.}}{\to} [|\lambda|^2 - 1]^{-1} \Gamma;$$

$$n^{-1/2} \| R_n \| \overset{\text{p. s.}}{\to} 0;$$

et, posant $P = [\zeta, \eta]$, matrice dont les vecteurs colonnes sont ζ et η et $\Delta_n = \operatorname{Diag}[n^{1/2}, \lambda^n]$:

$$G_n = {}^*\Delta_n^{-1} {}^*PS_n P \Delta_n^{-1} \overset{\text{p. s.}}{\to} G = \begin{bmatrix} G_1 & 0 \\ 0 & G_2 \end{bmatrix},$$

avec

$$G_1 (|\lambda|^2 - 1)^{-1} {}^*\zeta \Gamma \zeta, \qquad G_2 = |\lambda|^2 (|\lambda|^2 - 1)^{-1} \| \eta \|^4.$$

Pour

$$H_n = {}^*\Delta_{n-1}^{-1} {}^*PM_n, \qquad n^{-1/2} H_n \overset{\text{p. s.}}{\to} -\lambda^{-1} \begin{bmatrix} {}^*\zeta \Gamma \\ 0 \end{bmatrix}.$$

Ainsi, $n^{-1} {}^*\zeta S_n \zeta$ a, p. s., une limite non nulle et :

$$\limsup \frac{1}{n} \lambda_{\min} S_n < \infty.$$

La partie (2) du théorème 2 est établie en tenant compte de la commandabilité et de la partie (1) du théorème 4.

La partie (1) résulte de l'examen du cas singulier lorsque $d = 2$ comme en [5]. On obtient alors :

$$P^{-1} S_{n-1}^{-1} M_n = P^{-1} {}^*(\hat{A}_n - A) \overset{\text{p. s.}}{\to} -\lambda^{-1} (|\lambda|^2 - 1) [{}^*\zeta \Gamma \zeta]^{-1} \begin{bmatrix} {}^*\zeta \Gamma \\ 0 \end{bmatrix}. \qquad \blacksquare$$

II.9. Modèle $AR_d(p)$

Considérons un modèle $AR_d(p)$, autorégressif d'ordre p et de dimension d :

$$X_n = A_1 X_{n-1} + \ldots + A_p X_{n-p} + \varepsilon_n,$$

ou encore, en utilisant le polynôme matriciel qui à $z \in \mathbb{C}$ associe

$$A(z) = I - A_1 z - \ldots - A_p z^p,$$

$A(R)X = \varepsilon$, R étant l'« opérateur retard ».

On associe au modèle $AR_d(p)$ un modèle $AR_{d \times p}(1)$ en posant :

$$^tY_n = (^tX_n, {}^tX_{n-1}, \ldots, {}^tX_{n-p+1})$$
$$^t\eta_n = (^t\varepsilon_n, 0, 0, \ldots, 0)$$

car, on obtient alors :

$$Y_{n+1} = BY_n + \eta_{n+1}$$

$$B = \begin{bmatrix} A_1 & A_2 & & & A_p \\ I & 0 & \ldots & \ldots & 0 \\ \ldots & \ldots & \ldots & \ldots & \ldots \\ 0 & 0 & \ldots & I & 0 \end{bmatrix},$$

« matrice compagne » du polynôme matriciel $z \to A(z)$.

Supposons la covariance Γ du bruit ε inversible. Alors le modèle autorégressif (Y_n) est commandable. Soit en effet (e_j) une base de $\mathbb{C}^d$ et, pour chaque j, $f_j = (e_j, 0, \ldots, 0)$ dans $\mathbb{C}^{pd}$. On voit facilement que le système de vecteurs $\{B^k f_j;\ 1 \leqq j \leqq d,\ 0 \leqq k < p\}$ est un système libre. La matrice de covariance de η est la matrice de dimension $pd \times pd$, K dont le bloc formé par les p-premières lignes et colonnes est Γ les autres termes étant nuls. Il en résulte que la matrice $[K^{1/2}, BK^{1/2}, \ldots, B^{p-1}K^{1/2}]$ est de rang dp. Dès lors, tous les résultats antérieurs ont leur traduction au modèle $AR_d(p)$.

On vérifie facilement, qu'au signe près, le polynôme caractéristique de B est $z \to \operatorname{Det}[A_p + z A_{p-1} + \ldots + z^{p-1} A_1 - z^p I]$ et que les valeurs propres de B sont éventuellement 0 (lorsque A_p n'est pas inversible), et que ses valeurs propres non nulles sont les z tels que $\operatorname{Det} A(1/z) = 0$, la dimension du sous-espace propre associé à $z \neq 0$ étant alors celle du noyau de $A(1/z)$. Le cas singulier pour Y correspond bien au cas dénommé « singulier » dans l'énoncé du théorème 5.

Le modèle (Y_n) est stable si $\operatorname{Det} A(z)$ ne s'annule que pour des z de module > 1. Il est explosif si A_p est inversible et si $A(z)$ est inversible pour z de module $\geqq 1$.

Tous les résultats obtenus dans le cas autorégressif d'ordre 1 ont donc leur extension rapide. D'où le théorème 5 auquel on peut facilement adjoindre des transcriptions des théorèmes 1-3-4 relatifs aux comportements des trajectoires ou à l'estimation de la covariance.

RÉFÉRENCES

[1] T. W. ANDERSON, On Asymptotic Distributions of Estimates of Parameters of Stochastic Difference Equations, *Ann. Math. Stat.*, Vol. **30**, 1959, p. 676-687.

[2] T. W. ANDERSON et J. TAYLOR, Strong Consistency of Least Squares Estimators in dynamic Models, *Ann. Stat.*, Vol. **7**, 1979, p. 484-489.

[3] N. H. CHAN et C. Z. WEI, Asymptotic Inference for Nearly Nonstationary AR(1) Processes, *Ann. Stat.*, Vol. **15**, 1987, p. 1050-1063.

[4] N. H. CHAN et C. Z. WEI, Limiting Distributions of Least Squares estimates of Instable Autoregressive Processes. *Ann. Stat.*, Vol. **16**, 1986, p. 367-401.

[5] M. DUFLO, R. SENOUSSI et A. TOUATI, Sur la loi des grands nombres pour les martingales vectorielles et l'estimateur des moindres carrés d'un modèle de régression, *Ann. Inst. Henri-Poincaré, Série B*, Vol. **26**, n° 4, 1990.

[6] A. JACUBOWSKY, J. MEMIN et G. PAGES, Convergence en loi des suites d'intégrales stochastiques sur l'espace D^1 de Skorokhod, *Prob. Theor. Rel. Fields*, Vol. **81**, 1989, p. 111-138.

[7] T. L. LAI, et C. Z. WEI, Least Squares Estimates in Stochastic Regression Models with Applications to Identification and Control of Dynamic Systems, *Ann. Math. Stat.*, Vol. **10**, 1982, p. 154-166.

[8] T. L. LAI et C. Z. WEI, Asymptotic Properties of General Autoregressive Models and Strong Consistency of Least Squares Estimate and their Parameters, *J. Mult. Anal.*, Vol. **13**, 1983, p. 1-23.

[9] T. L. LAI et C. Z. WEI, A Note on Martingale Difference Sequences Satisfying the Local Marcinkiewicz-Zygmund Condition, *Bull. Inst. Math. Academia Sinica*, Vol. **11**, 1983, p. 1-13.

[10] T. L. LAI et C. Z. WEI, On the Concept of Excitation in Least Squares Identification and Adaptative Control, *Stochastics*, Vol. **16**, 1986, p. 227-254.

[11] LE BRETON et M. MUSIELA, Consistency Set of Least Squares Estimates in Stochastic Regression Models, Stoch. Diff. Syst., *Lect. Notes Control Inf. Sci.*, Vol. **126**, 1989.

[12] M. M. RAO, Consistency and Limit Distributions of Estimators of Parameters in Explosive Stochastic Difference Equations, *Ann. Math. Stat.*, Vol. **32**, 1961, p. 195-218.

[13] W. STOUT, The Hartman-Wintner Law of Iterated Logarithm, *Ann. Math. Stat.*, Vol. **41**, 1970, p. 2158-2160.

[14] A. TOUATI, Convergence en loi de l'estimateur des moindres carrés d'un modèle autorégressif (cas mixte), *Ann. Inst. Henri-Poincaré, Série B*, Vol. **26**, 1990, p. 549-566.

[15] A. TOUATI, Loi limite de l'estimateur des moindres carrés dans le modèle autorégressif (cas général). *Ann. Inst. Henri-Poincaré, Série B*, Vol. **26**, 1990, p. 549-566.

[16] M. C. VIANO, *Thèse*, Université Paris-Sud, 1988.

[17] C. Z. WEI, Asymptotic Properties of Least Squares Estimates in Stochastic Models, *Ann. Stat.*, Vol. **13**, 1985, p. 1498-1508.

[18] C. Z. WEI, Adaptive Prediction by Least Squares Predictors in Stochastic Regression Models with Applications to Time Series, *Ann. Stat.*, Vol. **15**, 1987, p. 1667-1682.

[19] J. S. WHITE, The Limiting Distribution of the Serial Correlation Coefficient in the Explosive Case, *Ann. Math. Stat.*, Vol. **29**, 1958, p. 1188-1197.

(*Manuscrit reçu le* 9 *janvier* 1990;
corrigé le 21 *mai* 1990.)