

50376
1982
33

N° d'ordre : 967

50376
1982
33

THÈSE

PRÉSENTÉE A

L'UNIVERSITÉ DES SCIENCES ET TECHNIQUES DE LILLE I

POUR OBTENIR

LE GRADE DE DOCTEUR DE 3ÈME CYCLE

SPÉCIALITÉ : MATHÉMATIQUES APPLIQUÉES

PAR

MOURID Tahar



SUR L'IDENTIFICATION D'UN PROCESSUS AUTOREGRESSIF



MEMBRES DU JURY : P. POUZET, *PRÉSIDENT*

D. BOSQ, *RAPPORTEUR*

P. JACOB

S. MARQUET

} *EXAMINATEURS*

SOUTENUE LE 8 JUIN 1982

A mes parents.

Je remercie très vivement Monsieur le Professeur Pierre POUZET de me faire l'honneur de présider le jury de cette thèse.

Je tiens à exprimer ma profonde reconnaissance à Monsieur le Professeur Denis BOSQ qui m'a proposé le sujet de cette thèse, qui a su m'initier à la recherche avec compétence et efficacité en me conseillant et me corrigeant durant ce travail.

Je remercie également très vivement Monsieur le Professeur Pierre JACOB qui a accepté de juger ce travail, et avec qui j'ai eu de fructueuses discussions sur ce sujet.

Je remercie également très vivement le Professeur Simone MARQUET qui a accepté de faire partie du jury.

Je remercie très vivement Monsieur VAN INGELANDT pour l'aide remarquable et efficace apportée pour la réalisation du travail informatique.

Je tiens aussi à remercier l'équipe des Probabilistes et Statisticiens de l'U.E.R. de Mathématiques de Lille I, en particulier Michel CARBON.

Je louerai tout spécialement la frappe remarquable et diligente d'Arlette Lengaigne. Je remercie enfin toutes les personnes qui ont participé à la réalisation matérielle de cette thèse, pour la rapidité, la compétence et la gentillesse dont elles ont fait preuve.

INTRODUCTION.-

<u>CHAPITRE I - GENERALITES SUR LES PROCESSUS AR et ARMA.</u>	1
§ 1 - Processus stochastique : préliminaires et notations.	1
§ 2 - Généralités sur les processus AR.	4
§ 3 - Détermination de l'ordre.	24
<u>CHAPITRE II - ETUDE D'UN ESTIMATEUR DE L'ORDRE.</u>	38
§ 1 - Construction d'un estimateur de l'ordre.	38
§ 2 - Loi limite de l'estimateur.	39
§ 3 - Convergence en probabilité.	59
§ 4 - Autres développements :	61
- Détermination de l'ordre d'un AR multidimensionnel gaussien.	
- Détermination de l'ordre d'ARMA(p,q) gaussien.	
§ 5 - Prédiction et étude du risque quadratique asymptotique.	62
5.1. - Position du problème.	62
5.2. - Risques quadratiques asymptotiques.	64
<u>CHAPITRE III - PROCESSUS ARIMA(p,d,q).</u>	79
§ 1 - Notations.	79
§ 2 - Estimation des paramètres.	80
§ 3 - Détermination du degré de différentiation.	83
- Convergence de la procédure.	
§ 4 - Etude de cas particuliers :	88
- Marche aléatoire : ARIMA (0,1,0).	
- ARIMA (0,2,0).	
- ARIMA (1,2,0).	
§ 5 - Simulation.	90
§ 6 - Conclusion.	98

	<u>Pages</u>
<u>CHAPITRE IV - SIMULATION.</u>	98
§ 1 - Statistiques $S_N(k)$.	98
§ 2 - Critères de Schwarz et d'Akaike.	101
§ 3 - Influence de la borne K .	102
§ 4 - Etude de la stabilité de la sélection : $S_N(k)$.	102
§ 5 - Etude de processus réels : séries C, E, F du livre de Box-Jenkins.	102
§ 6 - Remarques et conclusion.	103
Conclusion générale.	104
Tableaux.	106
<u>BIBLIOGRAPHIE.</u>	128

INTRODUCTION. -

Depuis longtemps les statisticiens et praticiens s'intéressent aux modèles autorégressifs, moyenne mobile, ARMA, ARIMA etc ... pour d'écrire les séries chronologiques observées.

Un des premiers problèmes naturels qui se pose est alors l'estimation des paramètres des modèles cités ci-dessus.

Un nombre considérable d'articles ont paru traitant de l'estimation des paramètres de ces modèles.

Dans ce travail on s'intéresse plus précisément à l'identification d'un ARIMA(p,d,q) au vu des données. Dans le chapitre I on fait une synthèse des résultats connus sur l'estimation des paramètres inconnus d'un processus autorégressif, en insistant surtout sur les principales méthodes de détermination de l'ordre d'un autorégressif ([1], [2], [3] [4], [5], [6], [18], [19], [31] ...).

Dans le chapitre II à partir d'une statistique proposée par Shibata dans [39], on construit un estimateur de l'ordre d'un processus autorégressif. On donne la loi limite de cet estimateur, des résultats sur sa convergence, ainsi qu'une extension de l'estimateur aux processus multidimensionnels et aux processus mixtes ARMA. On termine ce chapitre par une étude du risque quadratique asymptotique. La démarche effectuée dans ce chapitre est analogue à celle utilisée par Shibata pour son étude du critère d'Akaike [38].

Dans le chapitre III on aborde l'estimation du paramètre d d'un processus ARIMA(p,d,q) de la manière suivante : on remarque que l'équation polynomiale associée au processus admet d racines de module égal à l'unité. L'idée naturelle est alors d'estimer les racines de cette équation pour évaluer d . On donne des résultats sur la convergence d'un estimateur du degré d .

On illustre cette technique par quelques exemples et en effectuant quelques simulations.

Enfin dans le chapitre IV, on effectue des simulations montrant le comportement de l'estimateur proposée au chapitre II.

CHAPITRE I

GENERALITES SUR LES PROCESSUS AUTOREGRESSIFS
ET ARMA .

I - PROCESSUS STOCHASTIQUES.-

- Préliminaires et Notations.

Définition.- Un processus stochastique réel $X_T = (X_t, t \in T)$ est une famille de variables aléatoires définies sur le même espace probabilisé (Ω, \mathcal{A}, P) à valeurs dans $(\mathbb{R}, \mathcal{B}_{\mathbb{R}})$.

- (Ω, \mathcal{A}, P) est appelé espace de base, où \mathcal{A} est une tribu et P est une probabilité.

- $(\mathbb{R}, \mathcal{B}_{\mathbb{R}})$ est l'espace des états où $\mathcal{B}_{\mathbb{R}}$ est la tribu borélienne.

- T est l'ensemble des temps. Dans la suite on supposera que $T = \mathbb{Z}$.

- Pour ω fixé, $\in \Omega$, l'application : $t \rightarrow X_t(\omega)$ est une réalisation du processus au point ω , ou trajectoire du point ω .

- Pour $t \in T$, l'application : $\omega \rightarrow X_t(\omega)$ est l'état du processus à l'instant t . Le processus $(X_t, t \in T)$ peut être considéré comme une variable aléatoire X_T à valeurs dans $(\mathbb{R}^T, \mathcal{B}_{\mathbb{R}^T})$:

$$X_T : (\Omega, \mathcal{A}, P) \longrightarrow (\mathbb{R}^T, \mathcal{B}_{\mathbb{R}^T})$$
$$\omega \longmapsto (X_t(\omega))_{t \in T}$$

où $\mathcal{B}_{\mathbb{R}^T}$ est la tribu borélienne de \mathbb{R}^T c'est-à-dire la plus petite tribu contenant tous les cylindres $S_n \times \mathbb{R}^{T-T_n}$, où T_n est une partie finie quelconque de T à n éléments et S_n une partie borélienne du produit fini \mathbb{R}^{T_n} .

Remarque : La tribu $\mathcal{B}_{\mathbb{R}^T}$ est aussi la tribu engendrée par les projections canoniques :

$$\begin{aligned} \Pi_t : (\mathbb{R}^T, \mathcal{B}_{\mathbb{R}^T}) &\longrightarrow (\mathbb{R}, \mathcal{B}_{\mathbb{R}}) \\ (X_s, s \in T) &\longrightarrow X_t . \end{aligned}$$

Loi du processus : La loi $L(X_T)$ du processus $X_T = (X_t, t \in T)$ (ou distribution de X_T) est une probabilité P_{X_T} sur $\mathcal{B}_{\mathbb{R}^T}$ définie par :

$$P_{X_T}(S) = P(X_T^{-1}(S)) \quad \text{pour } S \in \mathcal{B}_{\mathbb{R}^T} .$$

Remarque : Le processus canonique $(\Pi_t, t \in T)$ défini sur $(\mathbb{R}^T, \mathcal{B}_{\mathbb{R}^T}, P_{X_T})$ a même loi que le processus $X_T = (X_t, t \in T)$.

Loi de dimension finie : Les lois de dimension finie du processus X_T sont les lois des vecteurs aléatoires $(X_{t_1}, \dots, X_{t_k})$, $k \geq 1$ et $t_1, \dots, t_k \in T$.

Le grand théorème de Kolmogorov montre que la loi du processus $X_T = (X_t, t \in T)$ est entièrement déterminée par les lois de dimension finie du processus (cf. [26] p. 92-93).

Processus Gaussien : Le processus réel X_T est dit gaussien si ses lois de dimension finie sont gaussiennes (i.e. : si toute combinaison linéaire finie de X_t , $t \in T$ est une v.a. réelle gaussienne). La loi d'un tel processus est entièrement déterminée par la donnée des fonctions moyenne et covariance :

$$m(t) = E X_t, \quad t \in T$$

$$C(s,t) = \text{cov}(X_s, X_t), \quad s, t \in T.$$

Stationnarité :

1 - Un processus $(X_t, t \in \mathbb{Z})$ est strictement stationnaire si, pour toute partie finie $\{t_1, \dots, t_n\}$ de T et tout $s > 0$,

$$L(X_{t_1+s}, \dots, X_{t_n+s}) = L(X_{t_1}, \dots, X_{t_n}), \quad \text{où } L(Y) \text{ désigne loi de } Y.$$

2 - Un concept plus faible que la stationnarité stricte est :

Un processus réel X_T où tous les moments $E X_t^2$ existent, est faiblement stationnaire, ou stationnaire du 2^o ordre dans \mathbb{Z} si sa covariance dans $\mathbb{Z} \times \mathbb{Z}$ définie par : $C(s,t) = E(X_s - E X_s)(X_t - E X_t)$ ne dépend que de la différence $s-t$ de ses arguments.

Remarque :

a) Il est clair que pour les processus gaussiens réels les deux notions coïncident.

b) Le concept dans (1) implique le concept dans (2). La réciproque est généralement fautive. (cf. [8-1] p. I.2).

II - GENERALITES SUR LES PROCESSUS AUTOREGRESSIFS.-

Définitions et résultats généraux. (cf. [4] et [8-1]) .

Dans toute la suite $X_T = (X_t, t \in \mathbb{Z})$ désigne un processus réel,
et centré).

2.1. - Définition 0.- On dit que le processus $(\varepsilon_t, t \in \mathbb{Z})$
est un bruit blanc faible si :

$$E \varepsilon_t = 0$$

$$\text{et} \quad E \varepsilon_t \varepsilon_s = \delta_{st} \sigma^2$$

où δ_{st} est le symbole de Kronecker et $\sigma^2 > 0$.

Le processus $(\varepsilon_t, t \in \mathbb{Z})$ est dit un bruit blanc fort si les
variables aléatoires sont centrées, indépendantes de même loi et de
variance $\sigma^2 > 0$.

Définition 1.- Le processus $(X_t, t \in \mathbb{Z})$ est un autogréssif
d'ordre k si il vérifie , pour $k \geq 1$:

$$\begin{cases} X_t = a_1 X_{t-1} + a_2 X_{t-2} + \dots + a_k X_{t-k} + \varepsilon_t , & t \in \mathbb{Z} \\ a_k \neq 0 \end{cases} \quad (1)$$

où les a_1, \dots, a_k sont des nombres réels, les variables aléatoires
 $\varepsilon_t, t \in \mathbb{Z}$ constituent un bruit blanc faible et sont telles que
 $E \varepsilon_t X_s = 0$ pour $s < t$.

On notera souvent un autorégressif d'ordre k par : AR(k) .

Remarque : La condition $E \varepsilon_t X_s = 0$ pour $s < t$ implique
l'unicité de la décomposition dans (1).

Définition 2.- On appelle équation polynômiale associée à un AR(k), l'équation :

$$\Pi(z) = z^k - \sum_{i=1}^k a_i z^{k-i} = 0 . \quad (2)$$

On a les théorèmes suivants :

Théorème 1.- Une condition nécessaire et suffisante pour qu'il existe un processus autorégressif faiblement stationnaire vérifiant :

$$X_t = a_1 X_{t-1} + \dots + a_k X_{t-k} + \varepsilon_t$$

est que les racines de l'équation polynômiale associée :

$$\Pi(z) = 0$$

soient de module strictement inférieur à 1.

Un théorème de représentation d'un autorégressif faiblement stationnaire :

Théorème 2.- Si toutes les racines de l'équation polynômiale sont en module strictement inférieur à 1, alors X_t s'écrit :

$$X_t = \sum_{r=0}^{\infty} \delta_r \varepsilon_{t-r} \quad , \quad t \in \mathbb{Z} \quad (3)$$

où $\delta_0 = 1$, la convergence est en moyenne quadratique et où les δ_r sont les coefficients du développement en série de Taylor de

$$\left(1 - \sum_{i=1}^k a_i z^i\right)^{-1} = \sum_{r=0}^{\infty} \delta_r z^r .$$

De plus X_t est non corrélée avec : $\varepsilon_{t+1} , \varepsilon_{t+2} , \dots$.

Remarques :

1 - La décomposition dans (3) du processus $(X_t, t \in \mathbb{Z})$ n'est autre que la représentation de Wold d'un processus régulier (cf. [10] chap. VII, 4).

2 - Si les v.a. ε_t sont indépendantes de même loi (i.e. bruit blanc fort) alors le processus $(X_t, t \in \mathbb{Z})$ est strictement stationnaire.

Loi du processus :

Théorème 3.- Soit $(X_t, t \in \mathbb{Z})$ un processus AR(k) faiblement stationnaire, tel que les racines de (2) soient en module strictement inférieur à 1 et tel que les $\varepsilon_t, t \in \mathbb{Z}$ soient des v.a. gaussiennes indépendantes $N(0, \sigma^2)$, alors le processus $(X_t, t \in \mathbb{Z})$ est gaussien.

Les lois de dimension finie sont des lois normales de moyenne nulle, et de covariances :

$$c(h) = \sum_{i=1}^k c_i x_i^h$$

où les c_i sont des constantes et les x_i sont les racines distinctes de (2).

Remarque : Le fait d'avoir des racines multiples ne changeant que l'expression des covariances $c(h)$ du processus $(X_t, t \in \mathbb{Z})$.

2.2. - Les fonctions autocovariance et autocorrélation.

Soit $(X_t, t \in \mathbb{Z})$ un processus réel stationnaire du 2° ordre (non dégénéré).

Définition.- La fonction d'autocovariance du processus $(X_t, t \in \mathbb{Z})$ est définie par :

$$R(h) = \text{cov}(X_t, X_{t+h}) \quad \text{où } h \in \mathbb{Z} .$$

La fonction d'autocorrélation est définie par :

$$\rho(h) = \frac{R(h)}{R(0)} , \quad h \in \mathbb{Z} .$$

Propriétés :

$$\text{i) } R(0) = \sigma_X^2 ; \quad R(h) \leq R(0) ; \quad R(h) = R(-h)$$

la fonction $R(h)$ est de type positive i.e. :

Pour tout $t_1, \dots, t_n \in \mathbb{Z}$ et pour les réels a_1, \dots, a_n :

$$\sum_{r=1}^n \sum_{s=1}^n a_r a_s R(t_r - t_s) \geq 0 .$$

$$\text{ii) } \rho(0) = 1 , \quad |\rho(h)| \leq 1 , \quad \rho(h) = \rho(-h)$$

la fonction $\rho(h)$ est de type positif.

De l'équation aux différences (1) on déduit les équations de Yule-Walker sur les covariances.

Théorème 4.- L'autocovariance d'un processus $AR(k)$ vérifie les équations :

$$(4) \quad \left\{ \begin{array}{l} \sum_{i=1}^k a_i R(h-i) = R(h) \quad , \quad h = 1, 2, \dots \\ \sum_{i=1}^k a_i R(i) + \sigma_\epsilon^2 = R(0) \quad . \end{array} \right.$$

De même on établit pour les autocorrélations :

$$\rho(h) - \sum_{i=1}^k a_i \rho(h-i) = 0 \quad , \quad h = 1, 2, \dots \quad (5)$$

En écrivant (5) pour $h = 1, 2, \dots, k$ et en tenant compte de la parité de $\rho(h)$. On obtient le système :

$$\begin{bmatrix} \rho(1) \\ \rho(2) \\ \vdots \\ \vdots \\ \vdots \\ \rho(k) \end{bmatrix} = \begin{bmatrix} 1 & \rho(1) & \rho(2) & \dots & \rho(k-1) \\ \rho(1) & 1 & \rho(1) & \dots & \rho(k-2) \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \vdots & \vdots & \vdots & \vdots & \rho(1) \\ \rho(k-1) & \dots & \dots & \rho(1) & 1 \end{bmatrix} \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ \vdots \\ \vdots \\ a_k \end{bmatrix} \quad (6)$$

Ce système permet par inversion (la matrice étant définie positive, cf. plus loin) d'obtenir les a_i en fonction de $\rho(1), \dots, \rho(k)$.

2.3. - La fonction d'autocorrélation partielle.

Définition.- $(X_t, t \in \mathbb{Z})$ étant un processus stationnaire du 2° ordre, on appelle fonction d'autocorrélation partielle la fonction :

$$r(h) = \frac{\text{cov}(X_t - X_t^*, X_{t-h} - X_{t-h}^*)}{\text{Var}(X_t - X_t^*)}, \quad h \geq 2 \quad (7)$$

où X_t^* (resp. X_{t-h}^*) désigne la régression affine de X_t (resp. X_{t-h}) sur $X_{t-1}, \dots, X_{t-h+1}$.

Remarques.-

1 - $r(h)$ s'interprète comme le coefficient de corrélation de X_t, X_{t-h} quand on a supprimé "l'influence" linéaire de $X_{t-1}, \dots, X_{t-h+1}$ sur X_t et X_{t-h} .

2 - L'expression (7) montre que $r(h)$ est le coefficient de régression de $X_t - X_t^*$ sur $X_{t-h} - X_{t-h}^*$, si $\text{Var}(X_t - X_t^*) \neq 0$.

La suite des autocorrélations partielles d'un processus AR(k) possède une caractéristique importante :

Propriété : Si $(X_t, t \in \mathbb{Z})$ est un AR(k) , alors :

$$r(k) = a_k$$

et $r(p) = 0$ pour $p > k$ (8)

où a_k désigne le dernier coefficient de l'autorégressif AR(k) .

Remarques.-

1 - Du système (6) on tire que $r(k)$ ou a_k est fonction de $\rho(1), \dots, \rho(k)$, et inversement par récurrence on montre que $\rho(k)$ est fonction de $r(1), \dots, r(k)$. Ceci montre que la connaissance de $\rho(k)$ est équivalente à la connaissance de $r(k)$.

2 - Cette propriété de la suite $(r(h), h \geq 2)$ sera utilisée pour identifier un autorégressif observé.

2.4. - Construction de la suite des autocorrélations partielles.

- Soit $(X_t, t \in \mathbb{Z})$ un processus réel centré stationnaire du 2° ordre.

- On suppose que : $R(0) = \sigma_X^2 = 1$ et que la suite $(\rho(h), h \geq 1)$ est telle que :

$$\sum_{i=1}^n \sum_{j=1}^n \delta_i \delta_j \rho(t_i - t_j) > 0$$

(i.e. $\rho(h)$ définie positive) pour tous les réels δ_i non tous nuls, et $t_1, \dots, t_n \in \mathbb{Z}$

et avec $\rho(h) = E(x_t x_{t+h})$ pour tout t et h dans \mathbb{Z} . Dans ce cas on a : $\rho(h) = R(h)$.

- En projetant X_t sur l'e.v. engendré par $\{X_{t-1}, \dots, X_{t-l}, l \geq 1\}$ on a : $X_t = \sum_{j=1}^l a_j(l) X_{t-j} + \varepsilon_t$ où les v.a. ε_t sont centrées et orthogonales à X_{t-1}, \dots, X_{t-l} .

Alors la suite des autocorrélations partielles $(r(h), h \geq 1)$ est déterminée (cf. par ex. [33] ou [29] p. 27-28) en résolvant la suite des équations matricielles suivantes :

$$\boxed{R_\ell a^{(\ell)} = \rho^{(\ell)}, \quad \text{pour } \ell \geq 1} \quad (9)$$

où :

le vecteur $(a^{(\ell)})' = (a_1(\ell), a_2(\ell), \dots, a_\ell(\ell))$

la matrice $R_\ell = (\rho(|i-j|)) ; i, j = 1, \dots, \ell$

le vecteur $\rho^{(\ell)'} = (\rho(1), \dots, \rho(\ell))$.

Et la suite $(r(h), h \geq 1)$ est donnée par :

$$\boxed{r(\ell) = a_\ell(\ell), \quad \text{pour } \ell \geq 1} \quad (10)$$

Durbin (cf. dans [11]) a proposé une résolution récursive de (9), et le schéma de cette solution séquentielle est le suivant :

$$\begin{array}{l}
 \text{(D.1)} \quad r(1) = a_1(1) = \rho(1) \\
 \text{(D.2)} \quad \sigma^2(1) = 1 - r^2(1) \\
 \text{(D.3)} \quad r(\ell+1) = a_{\ell+1}(\ell+1) = (\rho(\ell+1) - \sum_{j=1}^{\ell} a_j(\ell) \rho(\ell+1-j)) / \sigma^2(\ell) \\
 \text{(D)} \left\{ \begin{array}{l} \text{où :} \\ \sigma^2(\ell) = 1 - \sum_{j=1}^{\ell} a_j(\ell) \rho(j) \\ \text{(D.4)} \quad a_j(\ell+1) = a_j(\ell) - r(\ell+1) a_{\ell+1-j}(\ell) \quad , \quad (j = 1, \dots, \ell) \\ \text{(D.5)} \quad \sigma^2(\ell+1) = \sigma^2(\ell) (1 - r^2(\ell+1)) \end{array} \right.
 \end{array}$$

avec les conditions : $\sigma^2(\ell) \neq 0$

$$\text{et} \quad |r^2(\ell)| < 1$$

Remarque : Nous verrons plus loin que ces conditions seront réalisées dans le cas d'un processus autorégressif (cf. Remarque sect. 2.5.)

- Les deux premières équations : (D.1) et (D.2) sont les conditions initiales et le reste des équations D3 - D5 montrent le passage de l'étape (ℓ) à l'étape $(\ell+1)$.

- Le coefficient $a_j(\ell)$ peut être interprété comme le coefficient de X_{t-j} dans la régression linéaire de X_t sur $X_{t-1}, \dots, X_{t-\ell}$ pour $t = 1, 2, \dots, \dots$.

- La quantité $\sigma^2(\ell)$ est la variance résiduelle de cette régression.

2.5. - Autocorrélation partielle d'un processus autorégressif.

Soit $(X_t, t \in \mathbb{Z})$ un processus gaussien stationnaire, centré, ayant respectivement pour fonction de covariance, et de corrélation partielle, $R(h)$, $r(h)$.

On a la propriété importante suivante : (cf. [33])

Théorème 1.- La fonction de covariance $R(h)$ du processus $(X_t, t \in \mathbb{Z})$ est définie positive (i.e. : $\sum_{i=1}^n \sum_{j=1}^n c_i c_j R(i-j) > 0$ avec au moins un $c_{i_0} \neq 0$) si et seulement si :

$$|r(h)| < 1 \quad \text{pour tout } h > 1 .$$

Dans le cas où $(X_t, t \in \mathbb{Z})$ est un autorégressif on a :

Théorème 2.- ([25] p. 27 - 44).

Si $(X_t, t \in \mathbb{Z})$ est un processus autorégressif, alors : sa fonction de covariance est définie positive.

De ces deux résultats on en déduit une propriété sur la suite $(r(h), h \geq 1)$ d'un autorégressif :

Corollaire.- La suite d'autocorrélation partielle $(r(h), h \geq 1)$ d'un processus autorégressif est telle que :

$$|r(h)| < 1 \quad \text{pour tout } h \geq 1 .$$

Le théorème suivant nous donne des renseignements sur la quantité $\sigma^2(l)$ introduite dans la résolution de Durbin (section 2.4.).

Théorème 3.- ([25] chap. 4, p. 45).

Si $(X_t, t \in \mathbb{Z})$ est un processus autorégressif avec (ε_t) comme bruit blanc associé, alors la quantité $\sigma^2(l)$ est telle que :

$$\sigma^2(l) \geq \sigma_\varepsilon^2 \quad \text{pour } l \geq 1$$

$$\sigma^2(l) \rightarrow \sigma_\varepsilon^2 \quad \text{quand } l \rightarrow +\infty .$$

Et plus précisément si $(X_t, t \in \mathbb{Z})$ est un autorégressif d'ordre k_0
on a :

$$\sigma^2(\ell) = \sigma_\varepsilon^2 \quad \text{pour } \ell \geq k_0$$

où $\sigma_\varepsilon^2 = \text{Var } \varepsilon_t > 0$.

Remarque : Dans le cas d'un processus autorégressif les conditions intervenant dans la résolution de Durbin :

$$|r(h)| < 1$$

et $\sigma^2(\ell) \neq 0$

sont donc bien vérifiées.

2.6. - Ergodicité.

a) Processus mélangeant (cf. [36] p. 103 et 112).

Soit $(X_t, t \in \mathbb{Z})$ un processus strictement stationnaire défini sur (Ω, A, P) .

Soit T la transformation "décalage à gauche" définie sur l'ensemble des suites infinies, par :

$$T(\dots, X_0, \dots) = (\dots, X_1, \dots)$$

T^{-1} désigne la transformation inverse de T .

Définition.- On dit que la suite $(X_t, t \in \mathbb{Z})$ vérifie la condition de mélangeance au sens (1) suivant si :

$$\lim_{j \rightarrow \infty} P(B \cap T^{-j} A) = P(B) \cdot P(A) \quad (1)$$

pour toute paire d'événements A et B .

Remarque : La condition de mélangeance (1) est une forme d'indépendance asymptotique. D'autre part, on montre dans (cf. [36] p. 110) que la condition suivante :

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{j=0}^{n-1} P(B \cap T^{-j} A) = P(A) \cdot P(B) \quad (2)$$

- est une condition nécessaire et suffisante pour que le processus $(X_t, t \in \mathbb{Z})$ soit ergodique.

En notant que la condition (1) implique la condition (2), on en déduit que tout processus mélangeant au sens de (1) est ergodique.

b) Processus linéaires.

b.1 - Si $(Z_t, t \in \mathbb{Z})$ est une suite de v.a. indépendantes, de même loi, de moyenne nulle et de variance σ^2 , alors la suite $(Z_t, t \in \mathbb{Z})$ est mélangeante au sens de (1) (cf. [36] p. 110 - 112).

Alors le processus linéaire :

$$X_t = \sum_{j=1}^{\infty} a_j Z_{t-j}, \quad \text{où} \quad \sum_{j=1}^{\infty} a_j^2 < \infty$$

est mélangeant au sens de (1) et par conséquent est ergodique.

b.2 - Cas d'un processus autorégressif.

Un autorégressif admettant la décomposition de Wold (cf. sect. 2.1. th. 2) est conséquemment ergodique.

c) Processus fortement mélangeant.

Le concept de la mélangeance forte fut introduit dans ([35]) pour pouvoir établir des théorèmes du type central limite.

Rappelons la définition. En appelant $M_{-\infty}^T$ et M_{T+h}^∞ les tribus respectivement engendrées par : $\{X_t, -\infty < t \leq T\}$ et $\{X_t, t \geq T+h, h \geq 0\}$,

On pose :

$$\alpha(h) = \sup_{A \in M_{-\infty}^T, B \in M_{T+h}^\infty} |P(A \cap B) - P(A)P(B)|.$$

La condition de forte mélangeance de Rosenblatt s'écrit :

$$\lim_{h \rightarrow \infty} \alpha(h) = 0.$$

Kolmogorov et Rosanov ([24]) ont obtenu des conditions de régularité sur la densité spectrale d'un processus gaussien stationnaire pour qu'il soit fortement mélangeant. Plus précisément :

Théorème. - Si la densité spectrale $f(\lambda)$ d'un processus stationnaire gaussien est continue et positive ^(*) pour $-\pi \leq \lambda \leq +\pi$ alors : le processus est fortement mélangeant.

Il en résulte qu'un autorégressif stationnaire gaussien est fortement mélangeant.

d) Mélangeance d'un processus linéaire.

$$\text{Soit le processus linéaire } Y_t = \sum_{k=0}^{\infty} g_k Z_{t-k} \quad (3)$$

où la suite $(Z_j, j \in \mathbb{Z})$ est formée de v.a. indépendantes et de densité $p_j(x)$.

Gorodetskii a donné dans [14] des conditions sous lesquelles le processus linéaire est fortement mélangeant. Plus précisément avec les notations :

$$S_i(\delta) = \sum_{j=i}^{\infty} |g_j|^\delta ; \quad \beta(k) = \sum_{i=k}^{\infty} (S_i(\delta))^{1/1+\delta}, \quad \delta < 2$$

$$\beta(k) = \sum_{i=k}^{\infty} \max \{ (S_i(\delta))^{1/1+\delta}, \sqrt{S_i(2) |\text{Log } S_i(2)|} \}, \quad \delta \geq 2.$$

(*) strictement.

$p_i(x)$ étant la densité de la v.a. Z_i , on a :

Théorème. - Si :

- (i) $\int_{-\infty}^{+\infty} |p_i(x) - p_i(x + \alpha)| dx \leq C_1 |\alpha|$;
- (ii) $E |Z_i^\delta| \leq C_2 < \infty$ pour un $\delta > 0$; si $\delta \geq 1$, alors on suppose que $E Z_i = 0$, et si $\delta \geq 2$, que $\text{Var } Z_i = 1$;
- (iii) $g(z) = \sum_{k=0}^{\infty} g_k z^k \neq 0$ pour $|z| \leq 1$;
- (iv) $\beta(0) > \infty$.

Alors $(Y_t, t \in \mathbb{Z})$ satisfait la condition de forte mélangeance.

Remarque : Il est clair que la forte mélangeance est plus faible que la mélangeance au sens (1).

2.7. - Estimation des paramètres d'un autorégressif.

2.7.0. - Construction des estimateurs.

Soit $(X_t, t \in \mathbb{Z})$ un autorégressif d'ordre k gaussien stationnaire et centré, vérifiant :

$$X_t = a_1 X_{t-1} + a_2 X_{t-2} + \dots + a_k X_{t-k} + \varepsilon_t$$

où les paramètres à estimer sont :

$$a_1, a_2, \dots, a_k, \sigma_\varepsilon^2,$$

et où l'ordre k est supposé connu.

Les $(\varepsilon_t, t \in \mathbb{Z})$ constituent une suite de v.a. indépendantes et de même loi : $N(0, \sigma_\varepsilon^2)$. On observe X_1, \dots, X_N .

Les estimateurs les plus utilisés sont les estimateurs des moindres carrés et les estimateurs du maximum de vraisemblance.

On suppose que les variables non observées

$$X_0 = X_{-1} = \dots = X_{-(k-1)} = 0 \quad .$$

- Les premiers estimateurs sont obtenus par la régression de X_t sur X_{t-1}, \dots, X_{t-k} pour $t = 1, \dots, N$, ce qui revient à minimiser la quantité (cf. [30] p. 346) :

$$Q(a_1, \dots, a_k) = \sum_{t=1}^N (X_t - a_1 X_{t-1} - \dots - a_k X_{t-k})^2 \quad .$$

- Pour les seconds estimateurs, on obtient en fait des estimateurs du maximum de vraisemblance approchée (cf. [30] p. 347) par la maximisation du logarithme de la vraisemblance approchée sous l'hypothèse de la normalité des (ε_t) .

Du log de la vraisemblance approchée :

$$L(a_1, \dots, a_k) = -(N-k) \text{Log}(\sigma_\varepsilon \sqrt{2\pi}) - \frac{1}{2 \sigma_\varepsilon^2} \sum_{t=1}^N (X_t - a_1 X_{t-1} - \dots - a_k X_{t-k})^2 \quad ,$$

on en déduit : maximiser $L(a_1, \dots, a_k)$ revient à minimiser son 2ème terme, c'est-à-dire minimiser $Q(a_1, \dots, a_k)$.

On montre alors que les deux estimateurs en question sont identiques et sont donnés par les équations de Yule-Walker, (cf. [30] p. 348-352) :

$$\begin{bmatrix} \hat{R}(1,1) & \dots & \hat{R}(1,k) \\ \vdots & & \vdots \\ \hat{R}(k,1) & \dots & \hat{R}(k,k) \end{bmatrix} \begin{bmatrix} \hat{a}_1(k) \\ \vdots \\ \hat{a}_k(k) \end{bmatrix} = \begin{bmatrix} \hat{R}(0,1) \\ \vdots \\ \hat{R}(0,k) \end{bmatrix} \quad (11)$$

où - $\hat{R}(i,j) = \frac{1}{N} \sum_{t=1}^N X_{t-i} X_{t-j}$, avec $i, j = 0, 1, \dots, k$

- les $\hat{a}_1(k), \dots, \hat{a}_k(k)$ sont les estimateurs de a_1, \dots, a_k .

L'estimateur du maximum de vraisemblance approché de σ_ε^2 est :

$$\hat{\sigma}_\varepsilon^2(k) = \frac{1}{N} \sum_{t=1}^N (X_t - \hat{a}_1(k) X_{t-1} - \dots - \hat{a}_k(k) X_{t-k})^2 \quad (12)$$

Notons que $N \hat{\sigma}_\varepsilon^2(k)$ est juste la somme résiduelle des carrés obtenue par la régression ci-dessus.

2.7.1. - Convergence et loi limite des estimateurs.

A - Convergence en probabilité. Mann et Wald ([27]) établirent les premiers résultats sur la convergence et les lois limites des estimateurs des moindres carrés. Leur résultat principal est :

Théorème 1.- Sous l'hypothèse que les (ε_t) sont indépendantes, centrées, de même loi et telles que : $E \varepsilon_t^4 < \infty$, alors : le vecteur aléatoire :

$$(\sqrt{N} (\hat{a}_1(k) - a_1) , \dots , \sqrt{N} (\hat{a}_k(k) - a_k)) \xrightarrow{L} N(0, \sigma_\varepsilon^2 R_k^{-1})$$

où R_k est la matrice de covariance des X_t :

$$R_k = \begin{bmatrix} R(0) & R(1) & \dots & R(k-1) \\ & R(1) & \dots & \vdots \\ & \vdots & \dots & \vdots \\ & \vdots & \dots & \vdots \\ R(k-1) & \dots & \dots & R(0) \end{bmatrix}$$

Notons que ce théorème entraîne : $\hat{a}_i(k) \rightarrow a_i$ en probabilité pour $i = 1, 2, \dots, k$. Anderson ([4] p. 193 - 200) obtient le même résultat mais sous des conditions plus faibles sur les (ε_t) :

Théorème 2.- Si les ε_t sont indépendantes avec $E \varepsilon_t = 0$ et $E \varepsilon_t^2 = \sigma_\varepsilon^2 > 0$. Si les ε_t sont soit de même loi, soit vérifient $E |\varepsilon_t|^{2+\varepsilon} < m$ avec $t = 1, 2, \dots, \dots$, $\varepsilon > 0$ et $m > 0$. On a alors :

$$\begin{aligned} \hat{a}_i(k) &\xrightarrow{P} a_i \quad \text{pour } i = 1, \dots, k \\ \hat{\sigma}_\varepsilon^2(k) &\xrightarrow{P} \sigma_\varepsilon^2 \quad \text{quand } N \rightarrow +\infty. \end{aligned}$$

Remarques :

- Ces résultats sont évidemment utilisés pour construire des tests et des intervalles de confiance pour les paramètres estimés dans le cas des grands échantillons.

B - Convergence presque sûre. Hannan [16] et Rissanen [34] ont établi la convergence presque sûre des estimateurs du maximum de vraisemblance :

On montre que :

- Sous des hypothèses de régularité sur l'espace des paramètres et sur la fonction de vraisemblance, on a :

$$\hat{a}_i(k) \rightarrow a_i \quad \text{p.s. pour } i = 1, \dots, k$$

$$\hat{\sigma}_\varepsilon^2(k) \rightarrow \sigma_\varepsilon^2 \quad \text{p.s. quand } N \rightarrow +\infty .$$

2.8. - Estimation de l'autocorrélation et de l'autocorrélation partielle.

$(X_t, t \in \mathbb{Z})$ est un processus autorégressif d'ordre k . On observe $X_1 \dots X_N$.

2.8.1. - L'autocorrélation empirique.

Un estimateur naturel de l'autocorrélation $\rho(h)$ est :

$$\hat{\rho}(h) = \frac{\sum_{t=0}^{N-h} (X_t - \bar{X}_N) (X_{t+h} - \bar{X}_N)}{\sum_{t=0}^N (X_t - \bar{X}_N)^2}, \quad h \geq 0 \quad (13)$$

avec
$$\bar{X}_N = \frac{1}{N} \sum_{t=1}^N X_t .$$

Le théorème suivant résume la convergence en probabilité, et la loi limite de l'estimateur $\hat{\rho}(h)$:

Théorème 1.- ([12] p. 240) .

$(X_t, t \in \mathbb{Z})$ admet la représentation :

$$X_t = \sum_{j=-\infty}^{+\infty} \psi_j \varepsilon_{t-j}$$

où $\sum_j |\psi_j| < \infty$ et les v.a. (ε_t) sont centrées, indépendantes, de variance σ^2 et admettant des moments d'ordre 6, avec $E \varepsilon_t^6 = \tau \sigma^6$.

Dans ces conditions, pour h et q fixés, on a :

$$1) \quad N \operatorname{cov}(\hat{\rho}(h), \hat{\rho}(q)) \longrightarrow \sum_{j=-\infty}^{+\infty} [\rho(j) \rho(j-h+q) + \rho(j+q) \rho(j-h) - 2 \rho(q) \rho(j) \rho(j-h) - 2 \rho(h) \rho(j) \rho(j-q) + 2 \rho(h) \rho(q) \rho^2(j)] = \psi_{h,q}$$

$$2) \quad (\sqrt{N}(\hat{\rho}(1) - \rho(1)), \dots, \sqrt{N}(\hat{\rho}(K) - \rho(K))) \xrightarrow{L} N(0, \Sigma), \quad \forall K \text{ fixé,}$$

et où la matrice de covariance Σ a le terme d'indice (h, q) la quantité $\psi_{h,q}$.

Corollaire.- ([12] p. 242) .

Sous les mêmes hypothèses du théorème précédent, on a :

pour $h \geq q > 0$,

$$E(\hat{\rho}(h)) = -\frac{N-h}{N(N-1)} + O(N^{-2})$$

$$\operatorname{cov}(\hat{\rho}(h), \hat{\rho}(q)) = \begin{cases} \frac{N-h}{N^2} + O(N^{-2}) & \text{si } h = q \neq 0 \\ O(N^{-2}) & \text{sinon .} \end{cases}$$

Remarque : Ces résultats montrent que pour N assez grand le biais de $\hat{\rho}(h)$ devient négligeable.

2.8.2. - Autocorrélation partielle empirique.

Pour estimer la suite des autocorrélations partielles $(r(\ell), \ell \geq 2)$ de l'autorégressif d'ordre k , on utilise le fait que $r(\ell)$ est une fonction de $\rho(\ell)$.

En estimant $\rho(\ell)$ par $\hat{\rho}(\ell)$, (cf. (13) ci-dessus), et en se rapportant à la construction de la suite $(r(\ell), \ell \geq 2)$ faite dans la section 2.4., on obtient un estimateur $\hat{r}(\ell)$ de $r(\ell)$ à partir de l'équation 10 ; défini comme suit :

$$\boxed{\hat{r}(\ell) = \hat{a}_\ell(\ell) \quad \text{pour } \ell \geq 1} \quad (14)$$

où l'estimateur $\hat{a}_\ell(\ell)$ est obtenu en résolvant le système (9) à partir des $\hat{\rho}(i)$, $i = 1, \dots$. La résolution de Durbin (cf. sect. 2.4 ou cf. [4] p. 184 - 188) permet alors d'obtenir récursivement la suite $(\hat{r}(\ell), \ell \geq 1)$:

$$\hat{r}(1) = \hat{a}_1(1) = \hat{\rho}(1)$$

$$\hat{r}(\ell+1) = (\hat{\rho}(\ell+1) - \sum_{j=1}^{\ell} \hat{a}_j(\ell) \hat{\rho}(\ell+1-j)) / \hat{\sigma}^2(\ell)$$

$$\text{où } \hat{\sigma}^2(\ell) = 1 - \sum_{j=1}^{\ell} \hat{a}_j(\ell) \hat{\rho}(j) .$$

Notons aussi les équations : (cf. [20] p. 336) :

$$\hat{\sigma}^2(1) = 1 - \hat{r}^2(1)$$

$$\hat{\sigma}^2(\ell+1) = \hat{\sigma}^2(\ell) (1 - \hat{r}^2(\ell+1)) , \quad (15)$$

chaque fois que l'on a : $\hat{\sigma}^2(\ell) \neq 0$

$$\text{et } |\hat{r}^2(\ell)| < 1 \quad (15.a)$$

(cf. Rem. 2 ci-dessus).

Remarques :

1) Si le processus $(X_t, t \in \mathbb{Z})$ est un autorégressif d'ordre k , on a :

$$\hat{r}(k) = \hat{a}_k(k) = \hat{a}_k$$

où \hat{a}_k est l'estimateur du dernier coefficient a_k de l'autorégressif, et

$$\hat{\sigma}_\varepsilon^2(k) = \hat{\sigma}_\varepsilon^2$$

où $\hat{\sigma}_\varepsilon^2 = 1 - \sum_{j=1}^k \hat{a}_j(l) \hat{\rho}(j)$ est l'estimateur de σ_ε^2 défini par les équations de Yule-Walker (cf. Sect. 2.2. th. 4) en supposant $\sigma_X^2 = 1$.

2) Puisque dans le cas d'un autorégressif on a :

$$\sigma^2(l) \neq 0$$

$$|r(l)| < 1 \quad (\text{cf. Rem. Sect. 2.5.})$$

On en déduit, de la convergence presque sûre des estimateurs $\hat{\sigma}^2(l)$ et $\hat{r}(l)$, qu'il existe dans (Ω, \mathcal{A}, P) un sous-ensemble Ω_0 tel que $P(\Omega_0) = 1$, et pour tout $\omega \in \Omega_0$ on ait, pour $l \geq k$:

$$\hat{\sigma}^2(l)(\omega) \neq 0$$

et $\hat{r}^2(l)(\omega) < 1$ pour N assez grand .

III - DETERMINATION DE L'ORDRE.

3.1. - Position du problème.

Soient :

- $(X_t, t \in \mathbb{Z})$, un processus autorégressif d'ordre k_0 , gaussien, stationnaire et centré.

- (X_1, \dots, X_N) un échantillon de taille N du processus $(X_t, t \in \mathbb{Z})$.

Sachant que l'ordre k_0 que l'on veut estimer est majoré par un nombre fixé K , plusieurs auteurs ont proposé des méthodes pour déterminer k_0 , et leur efficacité variant d'une méthode à l'autre.

3.2. - Ajustement de processus autorégressifs d'ordres successifs aux données.

Les observations X_1, \dots, X_N sont faites sur un autorégressif d'ordre k_0 de paramètres $(a_1, \dots, a_{k_0}, \sigma_\varepsilon^2)$. On ajuste aux observations des processus autorégressifs d'ordres successifs k , $k = 0, 1, \dots, K$ où $K \geq k_0$.

On note par :

$$\hat{a}_1(k), \dots, \hat{a}_k(k)$$

les estimateurs des coefficients $a_1(k), \dots, a_k(k)$ de l'autorégressif d'ordre k ajusté aux observations.

Nous définirons : $\hat{a}_i(k) = 0$ pour tout i tel que $k < i \leq K$, et nous écrirons : $\hat{a}(k)' = (\hat{a}_1(k), \dots, \hat{a}_k(k), 0 \dots 0)$ comme vecteur à K dimensions.

Nous noterons aussi : $a' = (a_1, \dots, a_{k_0}, 0, \dots, 0)$ comme vecteur à K dimensions.

On a le résultat suivant (cf. [4] p. 188-200) :

Théorème. - Pour tout $k \geq k_0$, on a :

$$\hat{a}(k) \xrightarrow{P} a$$

$$\hat{\sigma}_\varepsilon^2(k) \xrightarrow{P} \sigma_\varepsilon^2$$

de plus, les estimateurs sont asymptotiquement normaux.

3.3. - Différentes approches.

Plusieurs méthodes ont été proposées pour estimer l'ordre d'un autorégressif observé. Nous en donnerons les principales.

3.3.1. - Tests sur l'ordre. (cf. [4] p. 214-223).

(A) - En utilisant les résultats de la section 2.7.1. :

$$(\sqrt{N}(\hat{a}_1(p) - a_1), \dots, \sqrt{N}(\hat{a}_p(p) - a_p)) \xrightarrow{L} N(0, \sigma_\varepsilon^2 R_p^{-1}) ,$$

la v.a. :

$$\sqrt{N} \frac{\hat{a}_i(p) - a_i}{S \sqrt{-ii}} \text{ est asymptotiquement une v.a. normale } N(0,1) ,$$

pour $i = 1, \dots, p$.

où :

- $S \sqrt{-ii}$ est le $i^{\text{ième}}$ élément de la diagonale de la matrice $S^2 A_N^{-1}$.

- Les éléments de la matrice $S^2 A_N^{-1}$ sont des estimations des éléments correspondants de la matrice $\sigma_\epsilon^2 R_p^{-1}$.

- On suppose que les observations sont faites sur un autorégressif d'ordre p .

On utilise le résultat obtenu ci-dessus pour poser un test d'hypothèses :

" décider si l'ordre est $p-1$ ou p " .

Le test sera donc : tester $H_0 : a_p = 0$ contre $H_1 : a_p \neq 0$,

On rejette l'hypothèse nulle si :

$$\frac{|\hat{a}_p|}{S \sqrt{\bar{a}^{pp}}} > t(\epsilon)$$

au niveau de signification ϵ , pour N assez grand.

Les quantités \hat{a}_p , S et \bar{a}^{pp} sont des estimateurs convergents, si " H_0 est non réalisée", avec

$$\lim \hat{a}_p = a_p \neq 0, \quad \lim S^2 = \sigma_\epsilon^2 \quad \text{en probabilité .}$$

En utilisant ces arguments on peut montrer que le test est convergent, c'est-à-dire : pour une alternative donnée $a_p \neq 0$, la probabilité de rejeter l'hypothèse nulle tend vers 1 quand T croît indéfiniment

ⓑ - Une autre approche a été celle de Quenouille et étendue par Bartlett et Diananda (cf. [5] et [4]) .

- Tester : l'hypothèse nulle H_0 : l'ordre du processus est m
contre l'alternative H_1 : l'ordre du processus est
 p , avec $p > m$.

ou

- Tester l'hypothèse : $a_{m+1} = a_{m+2} = \dots = a_p = 0$.

Si $(X_t, t \in \mathbb{Z})$ est un autorégressif stationnaire d'ordre p avec les coefficients :

$$a_1, \dots, a_m, a_{m+1} = 0, \dots, a_p = 0,$$

alors on définit les v.a., pour $i = 1, \dots, p$:

$$\begin{aligned} g_i &= \frac{1}{\sqrt{N}} \sum_{t=1}^N X_{t-i} \varepsilon_t \\ &= \frac{1}{\sqrt{N}} \left[\sum_{t=1}^N X_t X_{t-i} + a_1 \sum_{t=1}^N X_{t-1} X_{t-i} + \dots + a_m \sum_{t=1}^N X_{t-m} X_{t-i} \right], \end{aligned}$$

et on montre que les g_i ont des moyennes nulles et pour matrice de covariances $\sigma_\varepsilon^2 R_p$, d'ordre p , où $R_p = (R(i-j), i, j = 1, \dots, p)$ est la matrice de covariances des X_t .

Alors le vecteur aléatoire (g_1, \dots, g_p) a une loi limite gaussienne grâce à un théorème limite (cf. [4] p. 200, th. 5.5.5.). D'ailleurs la loi limite d'un nombre donné de g_i est la même que celle du même nombre de $\sigma_\varepsilon \cdot X_t$ quand les X_t sont gaussiens.

Alors les v.a. :

$$\begin{aligned} h_j &= \sum_{k=0}^m a_k g_{j-k}, \quad a_0 = 1 \\ &= \frac{1}{\sqrt{N}} \sum_{k=0}^m a_k \sum_{t=1}^N X_{t-i} \varepsilon_t \\ &= \frac{1}{\sqrt{N}} \sum_{k, \ell=0}^m a_k a_\ell \sum_{t=1}^N X_{t-j+k} X_{t-\ell}, \quad j = m+1, \dots, p \end{aligned}$$

ont une loi limite gaussienne comme combinaison linéaire finie des g_i .

Ainsi, pour chaque N , la loi conjointe des h_j a pour moyenne 0 et pour covariance $E h_j h_{j'}$.

Un calcul de $E h_j h_{j'}$ montre que :

$$E h_j h_{j'} = \begin{cases} 0 & \text{si } j > j' \\ \sigma^4 & \text{si } j = j' \end{cases},$$

ce qui implique que, sous l'hypothèse nulle H_0 , les variables aléatoires

$\frac{h_{m+1}}{\sigma_\epsilon^2} = \frac{h_{m+2}}{\sigma_\epsilon^2}, \dots, \frac{h_p}{\sigma_\epsilon^2}$ ont une loi conjointe de moyenne nulle, à matrice de covariances l'identité I ; leur loi limite est normale. On peut

utiliser ces résultats sur les v.a. h_j pour tester l'hypothèse nulle :

$a_{m+1} = \dots = a_p = 0$, dans le cas où les paramètres a_1, \dots, a_m et σ_ϵ^2 sont connus.

En effet, sous H_0 , la variable aléatoire :

$$\frac{1}{\sigma_\epsilon^4} \sum_{j=m+1}^p h_j^2 \tag{16}$$

à une loi limite χ^2 à $(p-m)$ degrés de liberté.

Ce qui suggère quand les paramètres $a_1, \dots, a_m, \sigma_\epsilon^2$ sont inconnus, de les estimer par la méthode du maximum de vraisemblance (par exemple) et donc de proposer comme statistique :

$$\frac{1}{\hat{\sigma}_\epsilon^4(m)} \sum_{j=m+1}^p \hat{h}_j^2 \tag{17}$$

$$\text{où : } \left[\begin{array}{l} \hat{h}_j = \sum_{k=0}^m \hat{a}_k \hat{g}_{j-k} \\ \text{et } \hat{g}_i = \frac{1}{\sqrt{N}} \left[\sum_{t=1}^N X_t X_{t-i} + \hat{a}_1 \sum_{t=1}^N X_{t-1} X_{t-i} + \dots + \hat{a}_m \sum_{t=1}^N X_{t-m} X_{t-i} \right] \end{array} \right.$$

(les estimateurs $\hat{a}_1, \dots, \hat{a}_m, \hat{\sigma}_\varepsilon^2(m)$ sont basés sur l'équation aux différences d'ordre m).

- Du fait que, sous H_0 , les estimateurs $\hat{a}_1, \dots, \hat{a}_m, \hat{\sigma}_\varepsilon^2(m)$ sont des estimateurs convergents, on a le résultat suivant :

Théorème 1.- Si $(X_t, t \in \mathbb{Z})$ vérifie une équation aux différences d'ordre m , sous la condition que les racines du polynôme associé soient dans le disque unité, alors la v.a. (17) :

$$\frac{1}{\hat{\sigma}_\varepsilon^4(m)} \sum_{j=m+1}^p \hat{h}_j^2$$

a une loi limite χ^2 à $(p-m)$ degrés de liberté.

Alors on utilise cette approche pour tester l'ordre du processus autorégressif observé.

On définit le coefficient $a_j(j)$ comme le dernier coefficient de l'autorégressif d'ordre j , et on note $\hat{a}_j(j)$ son estimateur du maximum de vraisemblance.

On a, alors le théorème suivant sur le test de l'ordre :

Théorème 2.- Si $a_{m+1} = \dots = a_p = 0$ alors la v.a.

$$\frac{1}{\hat{\sigma}_\varepsilon^4(m)} \sum_{j=m+1}^p \hat{h}_j^2 - N \sum_{j=m+1}^p \hat{a}_j^2(j)$$

tend en probabilité vers 0.

Remarques :

a) On peut montrer aussi que la v.a. :

$$\frac{1}{\sigma_\varepsilon^4} \sum_{j=m+1}^P h_j^2 - N \sum_{j=m+1}^P \hat{a}_j^2(j) \xrightarrow{P} 0$$

b) Du th. 2 et d'un théorème dans ([7] p. 25) on en déduit que :

$$N \sum_{j=m+1}^P \hat{a}_j^2(j) \xrightarrow{L} \chi_{p-m}^2 .$$

Conclusion : Malgré la rigueur théorique de cette approche sur la détermination de l'ordre du processus observé, son emploi reste très limité dans les applications à cause du nombre de vérifications trop important à faire d'une part, et de la liaison entre l'ordre déterminé et le seuil de signification donné a priori, d'autre part.

3.3.2. - Utilisation de la fonction d'autocorrélation partielle (cf. [9] chap. 6).

L'examen de la fonction d'autocorrélation partielle (f.a.p.) peut déterminer l'ordre du processus observé. La f.a.p. d'un autorégressif d'ordre m vérifie :

$$\begin{cases} r(m) = a_m \\ r(\ell) = 0 \quad \text{pour } \ell \geq m+1 . \end{cases}$$

On estime $r(\ell)$ par la méthode proposée dans [la section 2.8.2, équation (14)].

On montre que si le processus observé est un autorégressif d'ordre k_0 alors :

- les v.a. $\hat{r}(\ell)$, pour $\ell > k_0$, sont asymptotiquement indépendantes et de même loi normale,

- pour N assez grand, on a pour $\ell > k_0$:

$$* E \hat{r}(\ell) \sim 0$$

$$* \text{Var}(\hat{r}(\ell)) \sim \frac{1}{N} .$$

Ainsi l'examen de l'appartenance de l'estimation $\hat{r}(\ell)$ à l'intervalle de confiance : $J = \left[-c_\alpha \sqrt{\frac{1}{N}}, c_\alpha \sqrt{\frac{1}{N}} \right]$ permet d'opter pour la nullité de $r(\ell)$ au niveau de signification $1-\alpha = P[\hat{r}(\ell) \in J]$ et de donner alors l'ordre du processus observé, avec $P(N > c_\alpha) = \frac{\alpha}{2}$ où $L(N) = N(0,1)$.

On remarquera que cette procédure permet de poser le test :

$$H_0 : r(\ell) = 0$$

contre $H_1 : r(\ell) \neq 0 .$

3.3.3. - Critères d'Akaike.

Ⓐ - Critère "F.P.E." (final-predictor-error). (cf. [1], [2])

On suppose que le vrai modèle est un autorégressif d'ordre k_0 .

L'erreur de prédiction dans les modèles autorégressifs est σ_ϵ^2 , quand tous les paramètres sont connus.

On observe X_1, \dots, X_N et on ajuste aux données des modèles autorégressifs d'ordre k pour k variant de 0 à K où $K \geq k_0$.

Alors le critère F.P.E. est bâti sur le principe de choix du modèle qui a le meilleur "pouvoir prédictif".

Pour chaque modèle $AR(k)$ on calcule $\hat{\sigma}_\varepsilon^2(k)$ et on choisit le modèle qui rend minimale la quantité :

$$FPE(k) = \frac{N+k}{N-k} \hat{\sigma}_\varepsilon^2(k) \quad , \quad k = 0, \dots, K \quad (18)$$

c'est-à-dire le point k de quantité $FPE(k)$ minimale, est pris comme une estimation de l'ordre k_0 du processus observé. Une simulation de ce critère qui a été faite, montre que l'ordre k_0 est surestimé dans le cas des grands échantillons.

Ⓑ - Critère "A.I.C." (Akaike-Information-Critere) (cf. [3]).

Ce critère repose sur des concepts d'informations et sur le principe du maximum de vraisemblance. Il a été construit dans un contexte plus général que les processus stochastiques.

Dans un premier temps : on maximise la fonction de vraisemblance séparément pour chaque modèle j en compétition, obtenant ainsi $M_j(X_1, \dots, X_N)$.

Dans un second temps : on choisit le modèle pour lequel la quantité :

$$AIC(k_j) = \text{Log}[M_j(X_1, \dots, X_N)] - k_j \quad (19)$$

est maximale pour k_j variant de 0 à K .

Dans le cas des modèles autorégressifs gaussiens $AR(k)$, le critère AIC devient :

$$AIC(k) = N \text{Log} \hat{\sigma}_\varepsilon^2(k) + 2k \quad , \quad \text{pour } k = 0, \dots, K$$

et on choisit le modèle autorégressif d'ordre k qui rend cette quantité minimale (en fait la 2^{ème} expression est l'opposée de la première (19) mais on a gardé la même appellation AIC(k)).

Shibata dans ([38], Th. 1) étudia la sélection de l'ordre par le critère AIC et montre, en notant cette sélection \hat{k} , que :

$$\lim_{N \rightarrow \infty} \Pr(\hat{k} = k_0) = p_0, \text{ avec } p_0 \neq 1$$

c'est-à-dire que l'estimateur \hat{k} est non convergent. De plus il montre que \hat{k} surestime l'ordre k_0 .

Conclusion.- Malgré le fait que ces critères reposent sur des justifications théoriques pas "tellement solides", et malgré leurs faiblesses théoriques (non convergents), ceux-ci sont les plus couramment utilisés en pratique en raison de la facilité de leur mise en oeuvre sur ordinateur, qui semble donner des résultats satisfaisants.

3.3.4. - Critère de Schwarz (cf. [37]).

Schwarz proposa une approche bayésienne du problème.

Il considère que les observations X_1, \dots, X_N faites sur le processus sont indépendantes (voir remarque en conclusion) identiquement distribuées et possédant, par rapport à une mesure fixée, une densité de la forme :

$$f(x, \theta) = \exp(\theta \cdot T(x) - b(\theta)) \quad (20)$$

(modèle exponentiel)

où le paramètre θ appartient à un sous-ensemble convexe Θ de E_K : espace euclidien de dimension K , et où T est une statistique exhaustive "K dimensionnelle" pour la famille F des lois ayant pour densité $f(x, \theta)$, (20).

Les modèles en "compétition" sont présentés sous formes d'ensembles : $m_j \cap \Theta$, où chaque m_j est une sous-variété linéaire

de E_K , de dimension k_j , $k_j = 0, \dots, K$.

Il étudia le comportement asymptotique des estimateurs de Bayes sous une classe spéciale de probabilité a priori.

Il suppose que la distribution a priori est de la forme :

$$\sum_{j=1}^K \alpha_j \mu_j \quad (21)$$

où :

- α_j est la probabilité a priori du $j^{\text{ième}}$ modèle d'être le vrai modèle,

- μ_j , la distribution a priori conditionnelle (ou a posteriori) de θ étant donné le $j^{\text{ième}}$ modèle, admet une densité (k_j - dimensionnelle) bornée et strictement positive sur tout $m_j \cap \Theta$.

Alors, sous toutes ces hypothèses et sous une hypothèse de régularité sur une fonction de perte (dépendant de θ , et qu'il suppose comprise entre deux bornes fixées positives pour toutes les fausses décisions) la solution de Bayes consiste à choisir le modèle qui est a posteriori le plus probable.

A l'aide de la formule de Bayes, ceci est équivalent à choisir le j qui maximise la quantité :

$$S(Y, N, j) = \text{Log} \int_{m_j \cap \Theta} \alpha_j \exp((Y \circ \theta - b(\theta)) N) d\mu_j(\theta), \quad (22)$$

où $Y = \frac{1}{N} \sum_{i=1}^N T(X_i)$.

Le théorème suivant donne un développement asymptotique de $S(Y, N, j)$.

Théorème.- Pour Y et j fixés et pour N tendant vers ∞ ,

$$S(Y, N, j) = N \sup_{\theta \in m_j \cap \Theta} (Y \circ \theta - b(\theta)) - \frac{1}{2} k_j \text{Log } N + R \quad (23)$$

où le reste $R = R(Y, N, j)$ est borné en N pour Y et j fixés.

Calculons la fonction de vraisemblance de X_1, \dots, X_N du modèle (20) :

$$\begin{aligned} \prod_{i=1}^N f(X_i, \theta) &= \exp(\theta \cdot \sum_{i=1}^N T(X_i) - Nb(\theta)) \\ &= \exp \cdot [(Y \circ \theta - b(\theta)) N] \end{aligned}$$

$$\text{d'où} \quad \text{Log} \prod_{i=1}^N f(X_i, \theta) = (Y \circ \theta - b(\theta)) N$$

Alors, en maximisant la fonction de vraisemblance séparément pour chaque modèle, on obtient ainsi $M_j(X_1, \dots, X_N)$, le critère de Schwarz revient à choisir le modèle pour lequel l'expression :

$$\text{Log } M_j(X_1, \dots, X_N) - \frac{1}{2} k_j \text{Log } N \quad (24)$$

est maximale, où k_j est la dimension du modèle, k_j variant de 0 à K .

Dans le cas des autorégressifs gaussiens, (en calculant la fonction de vraisemblance associée), le critère de Schwarz revient à choisir le modèle autorégressif pour lequel la quantité :

$$S(k) = N \text{Log } \hat{\sigma}_\epsilon^2(k) + k \text{Log } N \quad (25)$$

est minimale, pour $k = 0, 1, \dots, K$, et où $\hat{\sigma}_\epsilon^2(k)$ est l'estimateur du max. de vraisemblance de la variance du bruit blanc, associée à l'autorégressif d'ordre k ajusté aux données X_1, \dots, X_N .

Remarque : Dans le développement asymptotique de la solution de Bayes (23), le premier terme devient juste l'estimateur du maximum de vraisemblance, dans le cas des modèles exponentiels.

Conclusion : L'évaluation des premiers termes du développement asymptotique des estimateurs de Bayes a permis de construire un critère asymptotique pour séparer des modèles de dimension finie, au delà du contexte bayésien puisqu'il ne dépend pas de la classe des lois a priori.

L'hypothèse sur l'indépendance des observations n'est pas très contraignante du fait que les observations, dans le cas des processus autorégressifs, sont "légèrement corrélées" et deviennent asymptotiquement indépendantes. La classe des modèles exponentiels contient la grande majorité des processus.

Le critère de Schwarz (24) diffère de celui d'Akaike (19) par le fait que la dimension k_j est multiplié par $\frac{1}{2} \text{Log } N$, mais la procédure de Schwarz semble donner une estimation convergente de l'ordre (cf. simulation).

3.3.5. - Critère de Hannan. (cf. [18] [19]) .

Hannan proposa un critère pour estimer l'ordre d'un autorégressif observé. Il étudia la statistique :

$$\phi(k) = N \text{Log } \hat{\sigma}_\varepsilon^2(k) + 2 k c \text{Log Log } N \quad (26)$$

où c est une constante supérieure à 1, et $\hat{\sigma}_\varepsilon^2(k)$ l'estimateur du M.V. de $\sigma_\varepsilon^2 = \text{Var } \varepsilon_t$.

Le principe est toujours le même : choisir l'ordre k qui rend minimale $\phi(k)$.

Il montre, en notant cette sélection de l'ordre \hat{k} , que :
([19] , th. 2)

$$\lim_{N \rightarrow \infty} \hat{k} = k_0 \quad \text{p.s.}$$

3.3.6. - Conclusion.

On remarquera que tous ces critères essaient de donner une "formulation mathématique du principe de parcimonie dans la construction des modèles" celui-ci étant fondé sur la minimisation du nombre de paramètres "utiles".

D'autres critères d'estimation de l'ordre ont été proposés mais leur efficacité n'a semble-t-il été testée que sur le plan expérimental, par simulation.

Pour terminer, tous ces critères de sélection entre modèles ont été pour la plus part étendus aux processus Moyenne mobile (M.A) et au processus mixte ARMA ([6], [18], [30] p. 690).

CHAPITRE II

ETUDE D'UN ESTIMATEUR DE L'ORDRE.

I - CONSTRUCTION D'UN ESTIMATEUR DE L'ORDRE.

On considère un processus gaussien, stationnaire centré $(X_t, t \in \mathbb{Z})$ défini sur un espace probabilisé $(\Omega, \mathcal{A}, \text{Pr})$ et vérifiant l'équation aux différences suivante, pour $k \geq 1$:

$$X_t = a_1 X_{t-1} + \dots + a_k X_{t-k} + \varepsilon_t$$

où : - les (ε_t) constituent un bruit blanc de moyenne nulle et de variance $\sigma_\varepsilon^2 > 0$.

- les coefficients a_1, \dots, a_k sont des réels tels que :
 $P(z) = 1 - \sum_{j=1}^k a_j z^j$ soit non nul pour $|z| \leq 1$.

Les estimateurs des paramètres $(a_1, \dots, a_k, \sigma_\varepsilon^2)$ du modèle AR(k) sont les estimateurs du maximum de vraisemblance approchée que l'on obtient à l'aide des équations de Yule-Walker (cf. sect. 2.7.0., chap. I).

Si l'autorégressif observé est d'ordre k_0 , alors les estimateurs $(\hat{a}_1(k), \dots, \hat{a}_k(k), \hat{\sigma}_\varepsilon^2(k))$ des coefficients des AR d'ordre k ajustés, sont convergents et asymptotiquement normaux pour $k \geq k_0$. (cf. sect. 3.2., chap. I).

Le choix de l'ordre :

On observe X_1, \dots, X_N d'un processus autorégressif d'ordre k_0 .

On définit la statistique $S_N(k)$ (cf. [39]) par :

$$S_N(k) = (N + \alpha k N^\beta) \hat{\sigma}^2(k) \quad (27)$$

pour $\alpha > 0$ et $0 \leq \beta < 1$

et k variant de 0 à K , K étant fixé.

Alors on définit un estimateur \hat{k} de l'ordre du processus observé par :

$$\hat{k}(X_1, \dots, X_N) = \text{Min}\{k, 0 \leq k \leq K \mid S_N(k) = \text{Min}\{S_N(j), 0 \leq j \leq K\}\} \quad (28)$$

c'est-à-dire, pour chaque échantillon, \hat{k} prend la valeur du premier entier entre 0 et K réalisant le minimum de $S_N(j)$, pour $j = 0, 1, \dots, K$.

Remarque : La plupart des critères de sélection définis dans les sect. 3.3.3. à 3.3.5. ne prennent pas en compte l'éventualité du minimum réalisé en plusieurs points (ceci nuirait à la définition de leur estimateur) et c'est ce que ce nouvel estimateur essaie de résoudre en adoptant le principe de "parcimonie".

II - LOI LIMITE DE L'ESTIMATEUR.-

Un échantillon X_1, \dots, X_N du processus autorégressif d'ordre k_0 , étant donné, nous lui avons associé un estimateur $\hat{k}(X_1, \dots, X_N)$ de l'ordre k_0 . Nous cherchons, pour K fixé, la loi limite de $\hat{k}(X_1, \dots, X_N)$.

Théorème 1.-

a) Pour tout $\alpha > 0$ et $0 \leq \beta < 1$, on a :

$$\lim_{N \rightarrow \infty} \Pr(\hat{k} = k) = 0, \quad \text{pour } 0 \leq k < k_0.$$

b) Pour tout $\alpha > 0$ et $\beta = 0$, on a :

$$\lim_{N \rightarrow \infty} \Pr(\hat{k} = k) = p(k - k_0) q(K - k), \quad \text{pour } k_0 \leq k \leq K$$

où :

$$\begin{aligned} - p(0) &= q(0) = 1, \\ - p(k - k_0) &= \Pr\left(\bigcap_{i=1}^{k-k_0} (S_i > 0)\right), \\ - q(K-k) &= \Pr\left(\bigcap_{i=1}^{K-k} (S_i \leq 0)\right) \end{aligned}$$

avec $S_i = \sum_{j=1}^i Z_j - i\alpha$, et où les v.a. Z_j sont indépendantes et de même loi : χ_1^2 , pour $j = 1, 2, \dots, K$.

La démonstration de ce théorème repose sur les lemmes suivants :

Lemme 1.- (Un lemme d'analyse combinatoire cf. [40]).

Soient : Z_1, \dots, Z_n des v.a. i.i.d.

$$S_k = Z_1 + \dots + Z_k \quad \text{pour } 1 \leq k \leq n.$$

Si on pose :

$$p(n) = \Pr\left(\bigcap_{i=1}^n (S_i > 0)\right), \quad q(n) = \Pr\left(\bigcap_{i=1}^n (S_i \leq 0)\right)$$

On a alors :

$$p(n) = \sum_n^* \left\{ \prod_{i=1}^n \frac{1}{r_i!} \left(\frac{\alpha_i}{i}\right)^{r_i} \right\},$$

$$q(n) = \sum_n^* \left\{ \prod_{i=1}^n \frac{1}{r_i!} \left(\frac{1-\alpha_i}{i}\right)^{r_i} \right\}$$

où : - $\alpha_i = \Pr(S_i > 0)$,

- la sommation \sum_n^* est étendue à tous les n-uples (r_1, \dots, r_n) d'entiers non négatifs vérifiant $r_1 + 2r_2 + \dots + nr_n = n$.

Lemme 2.- Les variables aléatoires $\hat{a}_i(i)$ définies dans la section 3.2. chap. I sont telles que :

$$\hat{a}_i(i) = \bigcirc_p \left(\frac{1}{\sqrt{N}} \right)$$

pour tout i variant de $k_0 + 1$ à K .

Démonstration : Le vecteur aléatoire $(\sqrt{N} \hat{a}_{k_0+1}(k_0+1), \dots, \sqrt{N} \hat{a}_K(K))$ tend en loi vers une normale $N(0, I)$ où I est la matrice unité (cf. [4] p. 219).

Donc on a :

$$\sqrt{N} \hat{a}_i(i) \xrightarrow{L} N(0, 1) \text{ pour } i \geq k_0 + 1.$$

Soient $\epsilon > 0$, et b_i la limite en loi de la v.a. $\sqrt{N} \hat{a}_i(i)$. La v.a. b_i a pour loi $N(0, 1)$. Alors il existe $\delta = \delta(\epsilon) > 0$ tel que :

$$\Pr(|b_i| > \delta) < \frac{\epsilon}{2}.$$

De la convergence en loi de $\sqrt{N} \hat{a}_i(i)$ vers b_i on tire :

$\forall \delta_1 > 0$, $\exists N(\delta_1)$ tel que :

$$\forall N \geq N(\delta_1) \implies |\Pr(\sqrt{N} |\hat{a}_i(i)| > \delta_1) - \Pr(|b_i| > \delta_1)| < \frac{\varepsilon}{2} .$$

D'où pour $\delta_1 = \delta$ on a :

$$\Pr(\sqrt{N} |\hat{a}_i(i)| > \delta) \leq \Pr(|b_i| > \delta) + \frac{\varepsilon}{2} < \varepsilon$$

dès que N est assez grand.

En conclusion :

pour $\varepsilon > 0$, $\exists \delta > 0$ tel que $\Pr(\sqrt{N} |\hat{a}_i(i)| > \delta) < \varepsilon$ dès que $N > N(\delta)$.

D'où le résultat :

$$\hat{a}_i(i) = O_p\left(\frac{1}{\sqrt{N}}\right) .$$

Remarque : Le δ choisi est assez grand.

Lemme 3.- Supposons que $\beta = 0$ et $\alpha > 0$. On a alors :

$$\lim_{N \rightarrow \infty} \Pr(S_N(k_1) \neq S_N(k_2)) = 1$$

pour tout k_1, k_2 supérieurs à k_0 , k_1 distinct de k_2 .

Démonstration : Soient k_1 et $k_2 \geq k_0$, avec $k_1 \neq k_2$.

On suppose que $k_1 < k_2 \leq K$. Le vecteur aléatoire

$$(\sqrt{N} \hat{a}_{k_0+1}(k_0+1) , \dots , \sqrt{N} \hat{a}_K(K))$$

tend en loi vers une v.a. normale $N(0, I)$ où I est la matrice unité (cf. [4] p. 219) .

Par conséquent le vecteur aléatoire

$$(\sqrt{N} \hat{a}_{k_1+1}(k_1+1), \dots, \sqrt{N} \hat{a}_{k_2}(k_2))$$

tend en loi vers une v.a. normale $N(0, I')$.

Par continuité on a :

$$(1) \quad (N \hat{a}_{k_1+1}^2(k_1+1), \dots, N \hat{a}_{k_2}^2(k_2)) \xrightarrow{L} (\chi_1^2, \dots, \chi_1^2)$$

où le χ_1^2 est un chi-deux à 1 d.d.l.

Des relations de Durbin (cf. sect. 2.8.2. chap. I) on obtient :

$$\begin{aligned} \frac{S_N(k_2)}{S_N(k_1)} &= \frac{N + \alpha k_2}{N + \alpha k_1} \cdot \frac{\hat{\sigma}^2(k_2)}{\hat{\sigma}^2(k_1)} = \\ &= \frac{N + \alpha k_2}{N + \alpha k_1} \cdot \left(\prod_{i=k_1+1}^{k_2} (1 - \hat{a}_i^2(i)) \right) \end{aligned}$$

Donc

$$\Pr\left(\frac{S_N(k_2)}{S_N(k_1)} = 1\right) = \Pr\left(\prod_{i=k_1+1}^{k_2} (1 - \hat{a}_i^2(i)) = \frac{N + \alpha k_1}{N + \alpha k_2}\right)$$

D'autre part :

$$\begin{aligned} \prod_{i=k_1+1}^{k_2} (1 - \hat{a}_i^2(i)) &= 1 - \sum_{i=k_1+1}^{k_2} \hat{a}_i^2(i) + \sum_{i \neq j} \hat{a}_i^2(i) \hat{a}_j^2(j) - \sum_{i \neq j \neq k} \hat{a}_i^2(i) \hat{a}_j^2(j) \hat{a}_k^2(k) \dots \\ \dots (-1)^{(k_2 - k_1)} &\hat{a}_{k_1+1}^2(k_1+1) \hat{a}_{k_1+2}^2(k_1+2) \times \dots \times \hat{a}_{k_2}^2(k_2) . \end{aligned}$$

D'où l'événement :

$$\sum_{i=k_1+1}^{k_2} \hat{a}_i^2(i) - \sum_{i \neq j}^{k_2} \hat{a}_i^2(i) \hat{a}_j^2(j) + \dots - (-1)^{k_2-k_1} \cdot \hat{a}_{k_1+1}^2(k_1+1) \times \dots \times \hat{a}_{k_2}^2(k_2) = \frac{\alpha \cdot (k_2-k_1)}{N + \alpha k_2} = C_N$$

ou encore :

$$(2) \quad \sum_{i=k_1+1}^{k_2} N \hat{a}_i^2(i) - \sum_{i \neq j}^{k_2} N \hat{a}_i^2(i) \hat{a}_j^2(j) + \dots - (-1)^{k_2-k_1} \cdot N \hat{a}_{k_1+1}^2(k_1+1) \times \dots \times \hat{a}_{k_2}^2(k_2) = N C_N$$

De (1) il en résulte par continuité que :

$$N \sum_{i=k_1+1}^{k_2} \hat{a}_i^2(i) \xrightarrow{L} \chi_1^2 + \dots + \chi_1^2 = \chi_{k_2-k_1}^2$$

(les v.a. χ_1^2 sont indépendantes).

On montre que les termes dans (2) à partir du second convergent vers 0 en probabilité (cf. lemme 3 i.e. $\hat{a}_i(i) = O_p\left(\frac{1}{\sqrt{N}}\right)$).

(2') $\left\{ \begin{array}{l} \text{On en déduit que la variable aléatoire, membre} \\ \text{de droite dans (2), tend en loi vers une v.a. } \chi_{k_2-k_1}^2 \text{ qui possède} \\ \text{une loi à densité (cf. [7] p. 28).} \end{array} \right.$

On pose le membre de droite dans (2) égale à :

$$\varphi(N \hat{a}_{k_1+1}^2(k_1+1), \dots, N \hat{a}_{k_2}^2(k_2)) = \sum_{i=k_1+1}^{k_2} N \hat{a}_i^2(i) - \sum_{i \neq j}^{k_2} N \hat{a}_i^2(i) \hat{a}_j^2(j) \dots \dots \dots - (-1)^{k_2-k_1} N \hat{a}_{k_1+1}^2(k_1+1) \times \dots \times \hat{a}_{k_2}^2(k_2)$$

et
$${}^N C_N = \frac{N \cdot \alpha \cdot (k_2 - k_1)}{N + \alpha k_2}$$

qui tend vers $\alpha \cdot (k_2 - k_1)$ quand N tend vers l'infini.

Montrons par l'absurde que :

(3)
$$\Pr(S_N(k_1) = S_N(k_2)) = \Pr(\psi(N \hat{a}_{k_1+1}^2(k_1+1), \dots, N \hat{a}_{k_2}^2(k_2))) = {}^N C_N \rightarrow 0$$

Supposons pour cela que (3) ne tende pas vers 0. Cela entraînerait qu'il existe une partie infinie N_0 de N et un nombre $\delta > 0$ tels que :

$$\Pr(\psi(N \hat{a}_{k_1+1}^2(k_1+1), \dots, N \hat{a}_{k_2}^2(k_2))) = {}^N C_N \geq \delta > 0 \text{ pour } N \in N_0 .$$

Or, il existe un voisinage I de $\alpha \cdot (k_2 - k_1)$ dans \mathbb{R}^+ tel que :

$$\Pr(\psi(\chi_1^2, \dots, \chi_1^2) \in I) = \frac{\delta}{2} .$$

Avec pour N assez grand dans N_0 ,

$$\Pr(\psi(N \hat{a}_{k_1+1}^2(k_1+1), \dots, N \hat{a}_{k_2}^2(k_2))) \in I \geq \delta$$

ce qui, grâce à (2') est contradictoire.

D'où le résultat.

Remarque : Le lemme précédent montre que, pour N assez grand, les valeurs de $S_N(k)$, sur $\{k_0, \dots, K\}$ sont distinctes en probabilité.

Soit ξ_k la v.a. indicatrice de l'événement : "k est un minimum local de $S_N(k)$ sur l'ensemble $\{k_0, \dots, K\}$ ".

Alors $v = \sum_{k=k_0}^K \xi_k$ est le nombre de points de l'ensemble

$\{k_0, \dots, K\}$ réalisant un minimum local de $S_N(k)$. On a le résultat suivant :

Corollaire. -

$$\lim_{N \rightarrow \infty} \Pr(v = 1) = 1 .$$

Démonstration : Il suffit de remarquer que :

$$\{v = 1\} = \left\{ \sum_{k=k_0}^K \xi_k = 1 \right\} = \{ \xi_{j_0} = 1, \xi_j = 0, j \neq j_0 \} \text{ donc :}$$

$$\{v = 1\} \supseteq \{S_N(k_1) \neq S_N(k_2) ; k_1 \text{ et } k_2 \geq k_0 ; k_1 \neq k_2\} .$$

d'où , du lemme 2, on a :

$$P(v = 1) \longrightarrow 1 \quad \text{quand } N \rightarrow \infty .$$

Démonstration du théorème 1 : De la section 2.8.2. on a

l'équation (15) :

$$\hat{\sigma}^2(\ell+1) = \hat{\sigma}^2(\ell) (1 - \hat{a}_\ell^2(\ell))$$

définie pour $\hat{\sigma}^2(\ell) \neq 0$ et $|\hat{a}_\ell^2(\ell)| < 1$ et $\ell \geq 1$ (cf. Rem. p.23)

On peut écrire encore :

$$\hat{\sigma}^2(k) = \hat{\sigma}^2(m) \cdot \prod_{i=m+1}^k (1 - \hat{a}_i^2(i)) \quad (29)$$

pour $k > m$.

A - Démonstration du (a) .

Soit k un entier tel que $0 \leq k < k_0$.

1°) De (29) on a :

$$\hat{\sigma}^2(k_0) = \hat{\sigma}^2(k) \prod_{i=k+1}^{k_0} (1 - \hat{a}_i^2(i))$$

ou encore :

$$\frac{\hat{\sigma}^2(k)}{\hat{\sigma}^2(k_0)} = \frac{1}{\prod_{i=k+1}^{k_0} (1-\hat{a}_i^2(i))} \quad (\text{A.1})$$

Puisque : $\forall i \geq 1$, $|\hat{a}_i(i)| < 1$ alors $0 < 1 - \hat{a}_i^2(i) \leq 1$ ce qui implique que :

$$\prod_{i=k+1}^{k_0} (1-\hat{a}_i^2(i)) \leq 1 - \hat{a}_{k_0}^2(k_0) .$$

Donc de (A.1) on a :

$$\frac{\hat{\sigma}^2(k)}{\hat{\sigma}^2(k_0)} = \frac{1}{\prod_{i=k+1}^{k_0} (1-\hat{a}_i^2(i))} \geq \frac{1}{1-\hat{a}_{k_0}^2(k_0)} \quad (\text{A.2})$$

Comme : $\lim_{N \rightarrow \infty} \hat{a}_{k_0}(k_0) = a_{k_0}$ en probabilité

et $a_{k_0} \neq 0$, le processus autorégressif observé étant d'ordre k_0 ,

alors :

$$\frac{1}{1-\hat{a}_{k_0}^2(k_0)} \longrightarrow \frac{1}{1-a_{k_0}^2} \quad \text{en probabilité} \quad (\text{A.3})$$

où :

$$\frac{1}{1-a_{k_0}^2} > 1$$

Posons : $Y_N = \frac{1}{1-\hat{a}_{k_0}^2(k_0)}$ et $a = \frac{1}{1-a_{k_0}^2}$.

De (A.3) on a :

$\forall \epsilon > 0$, $\exists \delta = \delta(\epsilon)$ et un $N(\epsilon)$ tels que :

$$\forall N > N(\epsilon) , \Pr(|Y_N - a| > \delta) < \epsilon$$

ce qui implique :

$$\Pr(Y_N < a - \delta) < \epsilon$$

dès que $N > N(\epsilon)$.

En choisissant $\delta < \frac{a-1}{2}$ on a :

$$\Pr(Y_N < 1 + \delta) < \Pr(Y_N < a - \delta) < \epsilon$$

pour $N > N(\epsilon)$.

D'où de (A.2) on déduit à partir de $N > N(\epsilon)$:

$$\Pr\left(\frac{\hat{\sigma}^2(k)}{\hat{\sigma}^2(k_0)} < 1 + \delta\right) \leq \Pr\left(\frac{1}{1 - \hat{a}_{k_0}^2(k_0)} < 1 + \delta\right) < \epsilon . \quad (\text{A.4})$$

2°) De la définition (28) de l'estimateur \hat{k} , l'événement $\{\hat{k} = k\}$ s'écrit :

$$\{\hat{k} = k\} = \left\{ \bigcap_{0 \leq m < k} (S_N(k) < S_N(m)) \right\} \cap \left\{ \bigcap_{k \leq m \leq K} (S_N(k) \leq S_N(m)) \right\} \quad (30)$$

d'où :

$$\begin{aligned} \Pr(\hat{k} = k) &\leq \Pr\left(\bigcap_{k \leq m \leq K} (S_N(k) \leq S_N(m))\right) \\ &\leq \Pr(S_N(k) \leq S_N(k_0)) \end{aligned} \quad (\text{A.5})$$

car $k < k_0 \leq K$.

De la définition (27) de $S_N(k)$, on obtient :

$$\begin{aligned}
 \Pr(S_N(k) \leq S_N(k_0)) &= \Pr((N + \alpha k N^\beta) \hat{\sigma}^2(k) \leq (N + \alpha k_0 N^\beta) \hat{\sigma}^2(k_0)) \\
 &= \Pr\left(\frac{\hat{\sigma}^2(k)}{\hat{\sigma}^2(k_0)} \leq \frac{N + \alpha k_0 N^\beta}{N + \alpha k N^\beta}\right) \\
 &= \Pr\left(\frac{\hat{\sigma}^2(k)}{\hat{\sigma}^2(k_0)} \leq \frac{N + \alpha k N^\beta + \alpha(k_0 - k) N^\beta}{N + \alpha k N^\beta}\right) \\
 &= \Pr\left(\frac{\hat{\sigma}^2(k)}{\hat{\sigma}^2(k_0)} \leq 1 + \frac{\alpha(k_0 - k) N^\beta}{N + \alpha k N^\beta}\right) \tag{A.6}
 \end{aligned}$$

Comme $k_0 - k > 0$ et $0 \leq \beta < 1$, on a pour le nombre δ choisi dans 1°) :

$$\frac{\alpha(k_0 - k) N^\beta}{N + \alpha k N^\beta} < \delta$$

dès que $N > N_1(\delta)$.

En groupant les résultats partiels (A.4), (A.5) et (A.6)

on en déduit :

$$\begin{aligned}
 \Pr(\hat{k} = k) &\leq \Pr\left(\frac{\hat{\sigma}^2(k)}{\hat{\sigma}^2(k_0)} \leq 1 + \frac{\alpha(k_0 - k) N^\beta}{N + \alpha k N^\beta}\right) \\
 &\leq \Pr\left(\frac{\hat{\sigma}^2(k)}{\hat{\sigma}^2(k_0)} < 1 + \delta\right)
 \end{aligned}$$

$$< \varepsilon$$

dès que $N \geq \text{Max}(N(\varepsilon), N_1(\delta))$.

Conclusion : Comme ε est quelconque. On a le résultat :

$$\lim_{N \rightarrow \infty} \Pr(\hat{k} = k) = 0 \quad \text{pour} \quad 0 \leq k < k_0 .$$

B - Démonstration du (b).

Soit k un entier vérifiant $k_0 \leq k \leq K$.

De (30) on tire :

$$\Pr(\hat{k} = k) \leq \Pr\left(\bigcap_{k_0 \leq m \leq K} (S_N(k) \leq S_N(m))\right) \quad (\text{B.1})$$

Cette démonstration se fera en cinq points.

1 - Nous allons montrer que l'on a :

$$\begin{aligned} & (\forall \varepsilon_1 > 0) , (\exists N(\varepsilon_1) > 0) \text{ tels que :} \\ & (\forall N > N(\varepsilon_1)) , \left| \Pr(\hat{k} = k) - \Pr\left(\bigcap_{k_0 \leq m \leq K} (S_N(k) \leq S_N(m))\right) \right| < \varepsilon_1 \end{aligned} \quad (\text{B.2})$$

Il est évident qu'à partir de (B.1) on a :

$$\Pr(\hat{k} = k) \leq \Pr\left(\bigcap_{k_0 \leq m \leq K} (S_N(k) \leq S_N(m))\right) + \varepsilon_1 \quad (\text{B.3})$$

Il nous suffit donc de montrer l'autre inégalité.

Considérons l'événement :

$$E_N = \{v = 1\} .$$

Dire que l'événement : E_N est réalisé , c'est dire que :

" le minimum de la restriction de $S_N(k)$ à $\{k_0, \dots, K\}$ est réalisé en un point unique " .

Alors, si $\underline{E_N}$ est réalisé, on a l'implication suivante :

$$\left[\begin{array}{l} \hat{k} \geq k_0 \\ \text{et } S_N(k) \leq S_N(m), k_0 \leq m \leq K \end{array} \right] \implies \hat{k} = k .$$

Notons par $\Pr^*(.) = \Pr(.|E_N)$ la probabilité conditionnelle étant donné l'événement E_N .

On a alors :

$$\Pr^*(\hat{k} \geq k_0, S_N(k) \leq S_N(m), k_0 \leq m \leq K) \leq \Pr^*(\hat{k} = k) \quad (\text{B.4})$$

Considérons les événements :

$$\left\{ \begin{array}{l} A_N = \{\hat{k} \geq k_0\}, \\ B_N = \{S_N(k) \leq S_N(m), k_0 \leq m \leq K\} = \left\{ \bigcap_{k_0 \leq m \leq K} (S_N(k) \leq S_N(m)) \right\}. \end{array} \right.$$

De l'égalité :

$$B_N - A_N^c = B_N \cap A_N, \text{ où } A_N^c \text{ désigne le complémentaire}$$

de A_N , on déduit que :

$$\Pr(B_N) - \Pr(A_N^c) = \Pr(B_N \cap A_N) .$$

Du résultat (a) du théorème on a :

$$(\forall \epsilon'_1 > 0), (\exists N(\epsilon'_1) > 0), (\forall N > N(\epsilon'_1)), \Pr(\hat{k} < k_0) = \sum_{i=0}^{k_0-1} \Pr(\hat{k} = i) < \epsilon'_1$$

et par suite, pour $N > N(\epsilon'_1)$ on a :

$$\Pr(A_N^c) = \Pr(\hat{k} < k_0) < \epsilon'_1 .$$

On obtient ainsi :

$$\Pr(B_N) - \varepsilon_1' \leq \Pr(B_N) - \Pr(A_N^c) = \Pr(B_N \cap A_N) \quad (B.5)$$

dès que $N > N(\varepsilon_1')$.

D'autre part :

$$\begin{aligned} \Pr(B_N \cap A_N) &= \Pr(B_N \cap A_N \mid E_N) \Pr(E_N) + \Pr(B_N \cap A_N \mid E_N^c) \Pr(E_N^c) \\ &\leq \Pr(B_N \cap A_N \mid E_N) + \Pr(E_N^c) \\ &= \Pr^*(B_N \cap A_N) + \Pr(E_N^c) \end{aligned}$$

Le corollaire du lemme 2 nous donne :

$$\forall \varepsilon_1'' > 0, (\exists N(\varepsilon_1'') > 0), (\forall N > N(\varepsilon_1'')) , \Pr(E_N^c) < \varepsilon_1''$$

et par conséquent :

$$(\forall N > N(\varepsilon_1'')) , \Pr(B_N \cap A_N) - \varepsilon_1'' \leq \Pr^*(B_N \cap A_N) \quad (B.6)$$

Ainsi en groupant (B.4), (B.5) et (B.6) on a :

$$\Pr(B_N) - \varepsilon_1' - \varepsilon_1'' \leq \Pr(B_N \cap A_N) - \varepsilon_1'' \leq \Pr^*(B_N \cap A_N) \leq \Pr^*(\hat{k} = k)$$

dès que $N > \text{Max}(N(\varepsilon_1'), N(\varepsilon_1''))$.

En posant $\frac{\varepsilon_1}{2} = \varepsilon_1' - \varepsilon_1''$ et $N(\varepsilon_1) = \text{Max}(N(\varepsilon_1'), N(\varepsilon_1''))$ on obtient :

$$(\forall N > N(\varepsilon_1)) , \Pr(B_N) - \frac{\varepsilon_1}{2} \leq \Pr^*(\hat{k} = k) . \quad (B.7)$$

Du corollaire du lemme 2 et de la relation :

$$\Pr(\hat{k} = k) = \Pr^*(\hat{k} = k) P(E_N) + \Pr(\hat{k} = k \mid E_N^c) P(E_N^c)$$

On en déduit que, pour N assez grand :

$$\Pr^*(k = k) \leq \Pr(\hat{k} = k) + \frac{\varepsilon_1}{2} \quad (\text{B.7'})$$

D'où de (B.7) : $\Pr(B_N) - \varepsilon_1 \leq \Pr(\hat{k} = k)$, pour N assez grand.

- Finalement en groupant avec (B.3), il en résulte (B.2).

2 - De la définition (27) de $S_N(k)$ et pour $\beta = 0$ on a :

$$\Pr(S_N(k) \leq S_N(m), k_0 \leq m \leq K) = \Pr\left(\frac{\hat{\sigma}^2(k)}{\hat{\sigma}^2(m)} \leq \frac{N + \alpha m}{N + \alpha k}, k_0 \leq m, k \leq K\right) \quad (\text{B.8})$$

Selon que $m < k$ ou $m > k$ on a :

$$\frac{\hat{\sigma}^2(k)}{\hat{\sigma}^2(m)} = \begin{cases} \prod_{i=m+1}^k (1 - \hat{a}_i^2(i)) & \text{pour } m < k, \\ \left[\prod_{i=k+1}^m (1 - \hat{a}_i^2(i)) \right]^{-1} & \text{pour } k < m. \end{cases}$$

De (B.8) et de ce qui précède, il résulte :

$$\Pr(S_N(k) \leq S_N(m), k_0 \leq m \leq K) = \Pr \left[\begin{array}{l} \prod_{i=m+1}^k (1 - \hat{a}_i^2(i)) \leq \frac{N + \alpha m}{N + \alpha k}, k_0 \leq m < k ; \\ \prod_{i=k+1}^m (1 - \hat{a}_i^2(i)) \geq \frac{N + \alpha k}{N + \alpha m}, k < m \leq K \end{array} \right]$$

(B.9)

3 - En écrivant :

$$\frac{N + \alpha m}{N + \alpha k} = 1 + \frac{\alpha(m-k)}{N + \alpha k}, \text{ si } k_0 \leq m < k,$$

et
$$\frac{N + \alpha k}{N + \alpha m} = 1 + \frac{\alpha(k-m)}{N + \alpha m}, \text{ si } k < m \leq K$$

En prenant la deuxième probabilité dans (B.9) et en passant aux logarithmes on obtient :

$$\Pr \left[\prod_{i=m+1}^k (1 - \hat{a}_i^2(i)) \leq 1 + \frac{\alpha(m-k)}{N + \alpha k}, k_0 \leq m < k, \prod_{i=k+1}^m (1 - \hat{a}_i^2(i)) \geq 1 + \frac{\alpha(k-m)}{N + \alpha m}, k < m \leq K \right]$$

$$= \Pr \left[\text{Log} \prod_{i=m+1}^k (1 - \hat{a}_i^2(i)) \leq \text{Log} \left(1 + \frac{\alpha(m-k)}{N + \alpha k} \right), k_0 \leq m < k, \text{Log} \prod_{i=k+1}^m (1 - \hat{a}_i^2(i)) \geq \text{Log} \left(1 + \frac{\alpha(k-m)}{N + \alpha m} \right), k < m \leq K \right] \quad (\text{B.10})$$

Un développement limité de la fonction logarithme donne :

$$\text{Log}(1 + u) = u + \mathcal{O}(u^2)$$

où $-1 < u < 0$ et $\mathcal{O}(u^2) \rightarrow 0$ quand $u \rightarrow 0$.

Il en résulte, en posant :

$$\begin{cases} \Delta_1 = m-k & \text{pour } m < k \\ \Delta_2 = k-m & \text{pour } m > k \end{cases},$$

$$\text{Log} \left(1 + \frac{\alpha \Delta_1}{N + \alpha k} \right) = \frac{\alpha \Delta_1}{N} + \mathcal{O} \left(\frac{1}{N^2} \right)$$

$$\text{Log} \left(1 + \frac{\alpha \Delta_2}{N + \alpha m} \right) = \frac{\alpha \Delta_2}{N} + \mathcal{O} \left(\frac{1}{N^2} \right) . \quad (\text{B.11})$$

de la continuité à droite de la fonction de répartition et des équations (B.10) et (B.11) il en résulte :

$$(\forall \varepsilon_2 > 0), (\exists N(\varepsilon_2) > 0), (\forall N > N(\varepsilon_2)),$$

$$\left| \Pr \left[\prod_{i=m+1}^k (1 - \hat{a}_i^2(i)) \leq \frac{N + \alpha m}{N + \alpha k}, k_0 \leq m < k; \prod_{i=k+1}^m (1 - \hat{a}_i^2(i)) \geq \frac{N + \alpha k}{N + \alpha m}, k < m \leq K \right] \right.$$

$$\left. - \Pr \left[\text{Log} \prod_{i=m+1}^k (1 - \hat{a}_i^2(i)) \leq \frac{\alpha \Delta_1}{N}, k_0 \leq m < k; \text{Log} \prod_{i=k+1}^m (1 - \hat{a}_i^2(i)) \geq \frac{\alpha \Delta_2}{N}, k < m \leq K \right] \right| \leq \varepsilon_2$$

(B.12)

4 - La deuxième probabilité dans l'inégalité (B.12) est

égale à :

$$\Pr \left[\sum_{i=m+1}^k N \text{Log}(1 - \hat{a}_i^2(i)) \leq \alpha \Delta_1, k_0 \leq m < k; \sum_{i=k+1}^m N \text{Log}(1 - \hat{a}_i^2(i)) > \alpha \Delta_2, k < m \leq K \right].$$

D'autre part, un développement limité d'ordre 2 de la fonction

$\text{Log}(1-x)$, pour $0 < x < 1$, donne :

$$\text{Log}(1-x) = -x + \frac{1}{2} (-x)^2 \frac{1}{(1-\gamma x)^2}, \text{ où le nombre } \gamma \text{ vérifie :}$$

$$0 < \gamma < 1$$

d'où :

$$\text{Log}(1 - \hat{a}_i^2(i)) = -\hat{a}_i^2(i) - \frac{1}{2} \frac{\hat{a}_i^4(i)}{(1 - \gamma \hat{a}_i^2(i))^2}$$

où γ est une v.a. vérifiant $0 < \gamma < 1$,

ou encore :

$$N \text{Log}(1 - \hat{a}_i^2(i)) = -N \hat{a}_i^2(i) - \frac{1}{2} \frac{N \hat{a}_i^4(i)}{(1 - \gamma \hat{a}_i^2(i))^2}, \quad 0 < \gamma < 1.$$

De la démonstration du lemme 3 on a , pour $i \geq k_0 + 1$ à K :

$$N \hat{a}_i^2(i) \xrightarrow{L} \chi_1^2 .$$

Donc :

$$N^2 \hat{a}_i^4(i) \xrightarrow{L} Z$$

où Z est une v.a. carrée d'un χ_1^2 .

Par conséquent

$$N \hat{a}_i^4(i) \longrightarrow 0 \text{ en probabilité .}$$

D'où la différence $N \text{Log}(1 - \hat{a}_i^2(i)) - (-N \hat{a}_i^2(i))$ tend vers 0 en probabilité.

Il en résulte :

$$\left(\forall \varepsilon_3 > 0 \right) , \left(\exists N(\varepsilon_3) > 0 \right) , \left(\forall N > N(\varepsilon_3) \right) ,$$

$$\left| \Pr \left[\sum_{i=m+1}^k N \text{Log}(1 - \hat{a}_i^2(i)) \leq \alpha \Delta_1 , k_0 \leq m < k ; \sum_{i=k+1}^m N \text{Log}(1 - \hat{a}_i^2(i)) > \alpha \Delta_2 , \right. \right.$$

$$\left. \left. \begin{array}{l} k < m \leq K \\ - \Pr \left[\sum_{i=m+1}^k N \hat{a}_i^2(i) > \alpha(k-m) , k_0 \leq m < k ; \sum_{i=k+1}^m N \hat{a}_i^2(i) < \alpha(m-k) , \right. \right. \\ \left. \left. \begin{array}{l} k < m \leq K \end{array} \right] \right] \leq \varepsilon_3 \end{array} \right|$$

(B.13)

5 - Puisque :

$$\sqrt{N} \hat{a}_i(i) \xrightarrow{L} N(0,1) \text{ pour } i = k_0 + 1, \dots, K$$

(cf. dem. du lemme 3).

on déduit :

$$N \hat{a}_i^2(i) \xrightarrow{L} \chi_1^2 .$$

Les v.a. $\sqrt{N} \hat{a}_i(i)$ étant asymptotiquement indépendantes, on a alors :

$$\sum_{i=k_0+1}^K N \hat{a}_i^2(i) \xrightarrow{L} \chi_1^2 + \dots + \chi_1^2 = \chi_{K-k_0}^2 .$$

Soient : Z_1, \dots, Z_{K-k_0} des v.a. indépendantes et de même

loi : χ_1^2 .

$$\text{Posons } S_i = (Z_1 - \alpha) + \dots + (Z_i - \alpha) = \sum_{j=1}^i Z_j - i\alpha .$$

Il en résulte :

$$\forall \varepsilon_4 > 0 , \quad (\exists N(\varepsilon_4) > 0) , \quad (\forall N > N(\varepsilon_4)) ,$$

$$\left| \Pr \left[\sum_{i=m+1}^k N \hat{a}_i^2(i) \geq \alpha(k-m) , \quad k_0 \leq m < k ; \quad \sum_{i=k+1}^m N \hat{a}_i^2(i) < \alpha(m-k) , \right. \right.$$

$$\left. \left. \begin{array}{l} k < m \leq K \end{array} \right] \right.$$

$$\left. - \Pr \left[S_m > 0 , \quad 1 \leq m \leq k-k_0 ; \quad S_m - S_{k-k_0} \leq 0 , \quad k-k_0 < m \leq K-k_0 \right] \right| < \varepsilon_4$$

(B.14)

Comme les v.a. Z_i sont indépendantes, les deux événements :

$$\{ S_m > 0 , \quad 1 \leq m \leq k - k_0 \} ,$$

et

$$\{ S_m - S_{k-k_0} \leq 0 , \quad k - k_0 < m \leq K - k_0 \} ,$$

sont indépendants.

On pose (cf. lemme 1) :

$$\Pr(S_m > 0, 1 \leq m \leq k - k_0) = \Pr\left(\bigcap_{m=1}^{k-k_0} (S_m > 0)\right) = p(k - k_0)$$

$$\Pr(S_m - S_{k-k_0} \leq 0, k - k_0 < m \leq K - k_0) =$$

$$\Pr\left(\bigcap_{m=k-k_0+1}^{K-k_0} (S_m - S_{k-k_0} \leq 0)\right) = q(K - k)$$

(les Z_i étant de même loi).

Conclusion : Des inégalités (B.2), (B.9), (B.12), (B.13)

et (B.14), il en résulte pour $\varepsilon > 0$, et pour

$N > N(\varepsilon) = \text{Max}(N(\varepsilon_1), N(\varepsilon_2), N(\varepsilon_3), N(\varepsilon_4))$:

$$|\Pr(\hat{k} = k) - p(k - k_0) q(K - k)| < \varepsilon$$

d'où le résultat :

$$\lim_{N \rightarrow \infty} \Pr(\hat{k} = k) = p(k - k_0) q(K - k).$$

Remarque : La loi limite de \hat{k} ne dépend que de k_0 et K et non des coefficients de l'autorégressif.

III - CONVERGENCE EN PROBABILITE.-

Soit \hat{k} la sélection de l'ordre obtenue par la statistique $S_N(k) = (N + \alpha(N)k) \hat{\sigma}^2(k)$, suivant la définition (28).

Théorème 2.-

$$\lim_{N \rightarrow \infty} \Pr(\hat{k} = k_0) = 1 \iff \alpha(N) \text{ croit indéfiniment}$$

et $\alpha(N) = o(N)$.

Démonstration : On reprend les étapes de la partie (B) de la démonstration du théorème 1 avec d'autres événements.

1 - On établit de la même façon que dans (B)-1 :

$$\forall \varepsilon_1 > 0, \left| \Pr(\hat{k} = k_0) - \Pr(S_N(k_0) \leq S_N(m), k_0 \leq m \leq K) \right| < \varepsilon_1$$

dès que N est assez grand.

2 - En écrivant l'événement $\{S_N(k_0) \leq S_N(m), k_0 \leq m \leq K\}$ à l'aide de $S_N(k) = (N + \alpha(N)k) \hat{\sigma}^2(k)$, et en utilisant les arguments de la partie 3. (B) on obtient :

$\forall \varepsilon_2 > 0$ et pour N assez grand ,

$$\left| \Pr(S_N(k_0) \leq S_N(m), k_0 \leq m \leq K) - \Pr\left(\sum_{i=k_0+1}^m N \text{Log}(1 - \hat{a}_i^2(i)) > \Delta_2 \alpha(N), k_0 \leq m \leq K\right) \right| < \varepsilon_2$$

où $\Delta_2 = k_0 - m$.

3 - De la même façon que dans 4.(B) on a :

$\forall \varepsilon_3 > 0$ et pour N assez grand ,

$$\left| \Pr\left(\sum_{i=k_0+1}^m N \text{Log}(1 - \hat{a}_i^2(i)) \geq \Delta_2 \alpha(N), k_0 \leq m \leq K\right) - \Pr\left(\sum_{i=k_0+1}^m N \hat{a}_i^2(i) \geq \Delta_2 \alpha(N), k_0 \leq m \leq K\right) \right| < \varepsilon_3 .$$

4 - Posons :

$$S_i = \sum_{j=1}^i Z_j - i\alpha(N) \quad \text{où } Z_j \text{ sont des v.a.i.i.d. de loi } \chi_1^2 .$$

on a alors :

$\forall \varepsilon_4 > 0$ et pour N assez grand (cf. p. 57) :

$$\left| \Pr\left(\sum_{i=k_0+1}^m N \hat{a}_i^2(i) < (m-k_0) \alpha(N) , k_0 \leq m \leq K \right) - \Pr\left(\chi_{m-k_0}^2 \leq (m-k_0) \alpha(N) , k_0 \leq m \leq K \right) \right| < \varepsilon_4$$

Conclusion : En groupant les résultats partiels dans 1 à 4, il en résulte :

$$(\forall \varepsilon > 0) ,$$

$$\left| \Pr(\hat{k} = k_0) - \Pr\left(\chi_{m-k_0}^2 < (m-k_0) \alpha(N) , k_0 \leq m \leq K \right) \right| < \varepsilon$$

dès que N est assez grand .

Comme $\alpha(N)$ croît vers l'infini et $(m-k_0) > 0$ on a :

$$\lim_{N \rightarrow \infty} \Pr\left(\chi_{m-k_0}^2 < (m-k_0) \alpha(N) , k_0 \leq m \leq K \right) = 1$$

d'où :

$$\lim_{N \rightarrow \infty} \Pr(\hat{k} = k_0) = 1 .$$

Corollaire.- Si $\alpha(N) = \alpha N^\beta$, $\alpha > 0$ et $0 < \beta < 1$

ou : $\alpha(N) = \alpha \text{ Log } N$, $\alpha > 0$

alors $\lim_{N \rightarrow \infty} \Pr(\hat{k} = k_0) = 1$.

Remarque : Ces deux exemples ont été traités en simulation.

- Si la sélection \hat{k} est convergente alors \hat{k} est asymptotiquement sans biais. En effet : $\lim E(\hat{k}) = \lim \sum_{i=0}^K i \Pr(\hat{k} = i) = k_0 \lim \Pr(\hat{k} = k_0) = k_0$.

IV - AUTRES DEVELOPPEMENTS.-

Le critère de sélection de l'ordre étudié peut être étendu aux autorégressifs multidimensionnels et aux processus ARMA gaussiens.

a) Autorégressif multidimensionnel gaussien.

Les vecteurs aléatoires $X(n)$ et $\varepsilon(n)$ de dimension p sont tels que :

$$\sum_{j=0}^k A(j) X(n-j) = \varepsilon_n$$

avec :

- $A(0) = I_p$ (identité) ,
- $E \varepsilon(n) = 0$, $E(\varepsilon(m) \varepsilon'(n)) = \delta_{mn} G$, $G > 0$ (définie positive).
- $\det\{\sum_{j=0}^k A(j) z^j\} \neq 0$ pour $|z| \leq 1$.

On considère alors la statistique :

$$S_N(k) = (N + \alpha k N^\beta) \det \hat{G}_k , \quad \alpha > 0 , \quad 0 < \beta < 1 ,$$

où les éléments de la matrice \hat{G}_k sont les estimateurs du maximum de vraisemblance approchés des éléments de la matrice résiduelle G .

On note cette sélection de l'ordre \hat{k} . Cette procédure devrait fournir également un estimateur convergent de l'ordre du processus observé.

b) Processus ARMA Gaussien.

Le processus $(X_t, t \in \mathbb{Z})$ vérifie :

$$\sum_{i=0}^p a_i X_{t-i} = \sum_{j=0}^q b_j \varepsilon_{t-j},$$

avec $a_0 = b_0 = 1$, $E \varepsilon(n) = 0$, $E \varepsilon(n) \varepsilon(m) = \delta_{nm} \sigma^2$,

$$g(z) = \sum_{i=0}^p a_i z^i \neq 0 \text{ pour } |z| \leq 1,$$

$$h(z) = \sum_{i=0}^q b_i z^i \neq 0 \text{ pour } |z| \leq 1.$$

On définit de la même manière la statistique

$$S_N(p,q) = (N + \alpha(p+q) N^\beta) \hat{\sigma}^2(p,q), \quad \alpha > 0, \quad 0 < \beta < 1$$

où $\hat{\sigma}^2(p,q)$ est l'estimateur du maximum de vraisemblance approché de σ^2 . Cette sélection (\hat{p}, \hat{q}) devrait également fournir un estimateur convergent de l'ordre (p,q) du processus observé.

V - PREDICTION ET ETUDE DU RISQUE QUADRATIQUE ASYMPTOTIQUE.

5.1. - Position du problème.

Soit $(X_t, t \in \mathbb{Z})$ un processus autorégressif d'ordre k vérifiant les conditions habituelles (cf. sect. 1 chap. II). On considère un autre processus $(Y_t, t \in \mathbb{Z})$ centré, stationnaire, indépendant de $(X_t, t \in \mathbb{Z})$ et vérifiant la même équation :

$$Y_t = a_1 Y_{t-1} + a_2 Y_{t-2} + \dots + a_k Y_{t-k} + \eta_t, \quad t \in \mathbb{Z} \quad (5.0)$$

(où (η_t) constitue un bruit blanc de variance σ_ε^2).

On fera la prédiction du processus $(Y_t, t \in \mathbb{Z})$ en observant le processus $(X_t, t \in \mathbb{Z})$. Les processus $(X_t, t \in \mathbb{Z})$ et $(Y_t, t \in \mathbb{Z})$ ont même fonction de covariance (cf. rem. p. 76).

On envisagera trois cas pour la construction du meilleur prédicteur linéaire \hat{Y}_t de Y_t au sens de l'erreur quadratique moyenne.

a) Si les paramètres a_i et l'ordre k sont connus, le prédicteur \hat{Y}_t s'écrit :

$$\hat{Y}_t = \sum_{i=1}^k a_i Y_{t-i} \quad (5.1)$$

b) Si l'ordre k est connu et les paramètres a_i sont à estimer à partir des observations faites sur le processus $(X_t, t \in \mathbb{Z})$; le prédicteur \hat{Y}_t devient alors :

$$\hat{Y}_t(k) = \sum_{i=1}^k \hat{a}_i(k) Y_{t-i} \quad (5.2)$$

c) Si l'ordre k (que l'on estimera par une sélection \hat{k}) et les paramètres a_i sont à estimer, alors le prédicteur \hat{Y}_t s'écrit :

$$\hat{Y}_t(\hat{k}) = \sum_{i=1}^{\hat{k}} \hat{a}_i(\hat{k}) Y_{t-i} \quad (5.3)$$

Evidemment c'est le dernier prédicteur qui nous intéresse sachant que l'ordre k et les paramètres a_i sont généralement inconnus.

En désignant, (cf. Sect. 3.2. chap. I) par :

$$\hat{a}(k)' = (\hat{a}_1(k), \dots, \hat{a}_k(k), 0 \dots 0)$$

et
$$\hat{a}(K)' = (\hat{a}_1(K), \dots, \hat{a}_K(K)) ,$$

les vecteurs $(K \times 1)$ des estimateurs des coefficients des autorégressifs respectivement $AR(k)$ et $AR(K)$, on montre que (cf. [3]) $\hat{a}(k)$

est la projection de $\hat{a}(K)$ sur le sous-espace Euclidien de dimension k muni de la norme :

$$||x||^2 = \sum_{i=1}^K \sum_{j=1}^K x_i x_j \hat{R}(i,j) ,$$

et du produit scalaire : $\langle x,y \rangle = x' \hat{R} y$,

où : $\hat{R} = \{\hat{R}(i,j) , 1 \leq i , j \leq K\}$

est la matrice de covariance des observations.

Alors la distance entre $\hat{a}(k)$ et $\hat{a}(\ell)$ ($\ell \leq k$) est donnée par :

$$||\hat{a}(k) - \hat{a}(\ell)||^2 = \hat{\sigma}^2(\ell) - \hat{\sigma}^2(k) .$$

5.2. - Risques quadratiques asymptotiques.

Les erreurs quadratiques des prédicteurs définies ci-dessus sont respectivement :

$$a) \quad E(\hat{Y}_t - Y_t)^2 = E(\eta_t^2) = \sigma_\varepsilon^2$$

$$\begin{aligned} b) \quad E_x E_y (\hat{Y}_t(k) - Y_t)^2 &= E_x E_y \left[\sum_{i=1}^k (\hat{a}_i(k) - a_i) Y_{t-i} \eta_t \right]^2 \\ &= E_x E_y \left[\sum_{i=1}^k (\hat{a}_i(k) - a_i) Y_{t-i} \right]^2 + \sigma_\varepsilon^2 \\ &= E_x \left[(\hat{a}(k) - a)' R (\hat{a}(k) - a) \right] + \sigma_\varepsilon^2 \end{aligned}$$

$$c) \quad E_x E_y (\hat{Y}_t(\hat{k}) - Y_t)^2 = E_x \left[(\hat{a}(\hat{k}) - a)' R (\hat{a}(\hat{k}) - a) \right] + \sigma_\varepsilon^2$$

où : E_x et E_y sont les espérances par rapport à (X_t) et (Y_t) respectivement, et

$$R = (R(i,j) = \text{cov}(X_i, X_j) , 1 \leq i , j \leq K)$$

est la matrice de covariance de (X_t) et de (Y_t) .

On définit la fonction de perte :

$$L(a, \hat{a}(k)) = (\hat{a}(k) - a)' R (\hat{a}(k) - a)$$

forme quadratique qui exprime la déviation des estimateurs $\hat{a}(k)$.

En adoptant le risque du prédicteur (5.1) nul, $E_X(L(a, \hat{a}(k)))$ et $E_X(L(a, \hat{a}(\hat{k})))$ seront considérés comme les risques associés aux prédicteurs (5.2) et (5.3).

Remarque : Comme la matrice de covariance empirique \hat{R} tend en probabilité vers R , la distance $||\hat{a}(k) - a||^2 = (\hat{a}(k) - a)' \hat{R}(\hat{a}(k) - a)$ est asymptotiquement équivalente à la fonction de perte $L(a, \hat{a}(k))$ (i.e. : le rapport tend vers 1 en probabilité).

Dans la suite on supposera que le processus observé $(X_t, t \in \mathbb{Z})$ est d'ordre k_0 et de paramètres $(a_1, \dots, a_{k_0}, \sigma_\varepsilon^2)$. On notera dorénavant E_X par E .

Les deux lemmes suivants sont dûs à Shibata [38]. On posera $a = (a_1, \dots, a_{k_0}, 0, \dots, 0)$.

Lemme 1.- Si $k \geq k_0$

$$\lim_{N \rightarrow \infty} N E [L(a, \hat{a}(k))] = \lim_{N \rightarrow \infty} N E \{ ||\hat{a}(k) - a||^2 \} = k \sigma^2 .$$

Lemme 2.- On a :

$$||\hat{a}(k) - a||^2 - ||\hat{a}(k_0) - a||^2 = \begin{cases} ||\hat{a}(k_0) - \hat{a}(k)||^2 + 2\{ \langle a - \hat{a}(k_0), \hat{a}(k_0) - \hat{a}(k) \rangle \}, & 0 \leq k \leq k_0, \\ ||\hat{a}(k_0) - \hat{a}(k)||^2 & , k_0 < k \leq K. \end{cases}$$

Lemme 3 [40].- Avec les mêmes notations du lemme 1 (§ 2, chap. II),

on a :

$$E(S_n I_{(S_1 > 0, \dots, S_n > 0)}) = \sum_n^* \left\{ \prod_{j=1}^n \frac{1}{r_j!} \left(\frac{\alpha_j}{j}\right)^{r_j} \right\} \left[\sum_{j=1}^n \frac{r_j}{\alpha_j} E\{S_j I_{(S_j > 0)}\} \right]$$

On considère la différence normalisée des risques des prédicteurs $\hat{Y}_t(\hat{k})$ et $\hat{Y}_t(k_0)$:

$$N E [L(a, \hat{a}(\hat{k})) - L(a, \hat{a}(k_0))]]$$

qui est une quantité positive indiquant la croissance du risque quand on estime l'ordre k_0 par une sélection \hat{k} .

Elle pourra jouer le rôle d'une mesure d'efficacité pour de telles sélection \hat{k} .

On considère la statistique $S_N(k) = (N + \alpha k N^\beta) \hat{\sigma}^2(k)$ avec $\alpha > 0$ et $0 \leq \beta < 1$.

On notera \hat{k} la sélection de l'ordre minimisant la statistique $S_N(k)$, suivant la définition (28).

On a le résultat suivant sur la croissance du risque :

Théorème 3.- Si $\beta = 0$ et $\alpha > 0$, alors

$$(5.4) \quad \lim_{N \rightarrow \infty} N [E L(a, \hat{a}(\hat{k})) - L(a, \hat{a}(k_0))] = \begin{cases} \left(\sum_{i=1}^{K-k_0} e_i q(K-k_0-i) \sigma_\epsilon^2 \right), & 0 \leq k_0 < K \\ 0 & , k_0 = K \end{cases}$$

le risque asymptotique est :

$$(5.5) \quad \lim_{N \rightarrow \infty} N E [L(a, \hat{a}(k))] = \begin{cases} \left(\sum_{i=1}^{K-k_0} e_i q(K-k_0-i) + k_0 \right) \sigma_\varepsilon^2 & , 0 \leq k_0 \leq K \\ K \sigma_\varepsilon^2 & , k_0 = K \end{cases}$$

où les $q(1), \dots, q(K)$ sont définies dans le lemme 1, § 2, et les e_i sont telles que :

$$e_n = \sum_n^* \left[\prod_{i=1}^n \frac{1}{r_i!} \left\{ \frac{\Pr(\chi_i^2 > i\alpha)}{i} \right\}^{r_i} \right] \left[\sum_{i=1}^n \left\{ i r_i \frac{\Pr(\chi_{i+2}^2 > i\alpha)}{\Pr(\chi_i^2 > i\alpha)} \right\} \right]$$

avec $n = 1, \dots, K$.

Remarque : La croissance du risque ne dépend asymptotiquement que de σ_ε^2 et de $K - k_0$; e_i et $q(i)$ étant des constantes.

Démonstration : Du lemme 1 il vient :

$$\lim_{N \rightarrow \infty} N E(L(a, \hat{a}(k_0))) = k_0 \sigma_\varepsilon^2$$

alors on obtient le résultat (5.5) du (5.4). En effet :

$$\begin{aligned} \lim_{N \rightarrow \infty} N E(L(a, \hat{a}(k))) &= \lim_{N \rightarrow \infty} N E(L(a, \hat{a}(k)) - L(a, \hat{a}(k_0))) + k_0 \sigma_\varepsilon^2 \\ &= \sum_{i=1}^{K-k_0} e_i q(K-k_0-i) \sigma_\varepsilon^2 + k_0 \sigma_\varepsilon^2 \end{aligned}$$

d'où le résultat (5.5).

Reste à montrer le résultat (5.4).

On envisagera deux cas : $\begin{cases} \hat{k} < k_0 & \text{et} \\ k_0 < \hat{k} < K & , k_0 \neq K . \end{cases}$

1^{er} cas : $\hat{k} = k < k_0$. Notons par A l'événement $\{\hat{k} = k\}$.

Du lemme 2 et de l'inégalité de Schwarz on obtient :

$$\begin{aligned}
 & E\{|\hat{a}(k) - a|^2 - |\hat{a}(k_0) - a|^2\} \cdot I_A \\
 (5.6) \quad & \leq E\{|\hat{a}(k_0) - \hat{a}(k)|^2 \cdot I_A\} + 2E\langle a - a(k_0), \hat{a}(k_0) - \hat{a}(k) \rangle \cdot I_A \\
 & \leq E\{|\hat{a}(k_0) - \hat{a}(k)|^2 \cdot I_A\} + 2[E|\hat{a}(k_0) - a|^2 \cdot E\{|\hat{a}(k_0) - \hat{a}(k)|^2 \cdot I_A\}]^{1/2} .
 \end{aligned}$$

Or du lemme 1 :

$$\lim_{N \rightarrow \infty} N E |\hat{a}(k_0) - a|^2 = k_0 \sigma^2 \implies E |\hat{a}(k_0) - a|^2 = O\left(\frac{1}{N}\right) .$$

D'autre part, de la définition de \hat{k} il en résulte :

$$\{\hat{k} = k, k < k_0\} \implies \{S_N(k) \leq S_N(k_0)\} .$$

En utilisant la définition de $S_N(k)$, pour $\beta = 0$ et $\alpha > 0$, on obtient :

$$\begin{aligned}
 S_N(k) \leq S_N(k_0) & \implies (N + \alpha k) \hat{\sigma}^2(k) \leq (N + \alpha k_0) \hat{\sigma}^2(k_0) \\
 \implies \frac{\hat{\sigma}^2(k)}{\hat{\sigma}^2(k_0)} & \leq \frac{N + \alpha k}{N + \alpha k_0} = 1 + \frac{\alpha(k - k_0)}{N + \alpha k_0} \\
 \implies \frac{\hat{\sigma}^2(k) - \hat{\sigma}^2(k_0)}{\hat{\sigma}^2(k_0)} & \leq \frac{\alpha(k - k_0)}{N + \alpha k_0}
 \end{aligned}$$

ou encore :

$$A = \{\hat{k} = k\} \implies \hat{\sigma}^2(k) - \hat{\sigma}^2(k_0) \leq \left(\frac{\alpha(k - k_0)}{N + \alpha k_0} \right) \hat{\sigma}^2(k_0) .$$

En utilisant :

$$\text{pr} - \lim \hat{\sigma}^2(k_0) = \sigma_\varepsilon^2 \implies \hat{\sigma}^2(k_0) = \sigma_\varepsilon^2 + o_p(1) , \text{ où } o_p(1) \rightarrow 0 \text{ en proba.}$$

$$\text{et } \|\hat{a}(k) - \hat{a}(\ell)\|^2 = \hat{\sigma}^2(\ell) - \hat{\sigma}^2(k) , \ell \leq k \text{ (cf. sect. 5.1 chap. II).}$$

Il en résulte (l'estimateur $\hat{\sigma}^2(k_0)$ est asymptotiquement sans biais) :

$$\begin{aligned} E\{\|\hat{a}(k_0) - \hat{a}(k)\|^2 \cdot I_A\} &= E[(\hat{\sigma}^2(k) - \hat{\sigma}^2(k_0)) \cdot I_A] \\ &\leq \left(\frac{\alpha(k - k_0)}{N + \alpha k_0} \right) (\sigma_\varepsilon^2 + o(1)) \cdot E(I_A) \\ &\leq \left(\frac{\alpha(k - k_0)}{N + \alpha k_0} \right) (\sigma_\varepsilon^2 + o(1)) \text{Pr}(A) . \end{aligned}$$

Du théorème 1, § 2, il vient, pour $k < k_0$

$$\lim_{N \rightarrow \infty} \text{Pr}(A) = \lim_{N \rightarrow \infty} \text{Pr}(\hat{k} = k) = 0$$

donc :

$$E\{\|\hat{a}(k_0) - \hat{a}(k)\|^2 \cdot I_A\} \text{ tend vers } 0 \text{ avec } \frac{1}{N} \text{ quand } N \rightarrow \infty .$$

Ainsi le nombre de droite de (5.6) est un $O\left(\frac{1}{N}\right)$.

En conclusion, pour $\hat{k} = k < k_0$, et en utilisant la remarque de la p. 65,

on a :

$$\lim_{N \rightarrow \infty} N E[L(a, \hat{a}(k)) - L(a, \hat{a}(k_0))] = \lim_{N \rightarrow \infty} N E[\|\hat{a}(k) - a\|^2 - \|\hat{a}(k_0) - a\|^2] = 0 .$$

2^{ème} cas : $k_0 < \hat{k} < K$ pour $k_0 \neq K$.

Du lemme 2 on a :

$$||\hat{a}(\hat{k}) - a||^2 - ||\hat{a}(k_0) - a||^2 = ||\hat{a}(\hat{k}) - \hat{a}(k_0)||^2 = \hat{\sigma}^2(k_0) - \hat{\sigma}^2(\hat{k}) .$$

De l'égalité :

$$\hat{\sigma}^2(\hat{k}) = \hat{\sigma}^2(k_0) \prod_{i=k_0+1}^{\hat{k}} (1 - \hat{a}_i^2(i))$$

on obtient :

$$\hat{\sigma}^2(k_0) - \hat{\sigma}^2(\hat{k}) = \hat{\sigma}^2(k_0) \left[1 - \prod_{i=k_0+1}^{\hat{k}} (1 - \hat{a}_i^2(i)) \right]$$

D'où, en utilisant $\lim \hat{\sigma}^2(k_0) = \sigma_\epsilon^2$ en probabilité, on a :

$$\begin{aligned} \lim_{N \rightarrow \infty} N E [L(a, \hat{a}(\hat{k})) - L(a, \hat{a}(k_0))] &= \lim_{N \rightarrow \infty} N E \{ ||\hat{a}(\hat{k}) - a||^2 - ||\hat{a}(k_0) - a||^2 \} \\ &= \lim_{N \rightarrow \infty} N E \hat{\sigma}^2(k_0) \left[1 - \prod_{i=k_0+1}^{\hat{k}} (1 - \hat{a}_i^2(i)) \right] \\ &= \lim_{N \rightarrow \infty} \sigma_\epsilon^2 \cdot N E \left[1 - \prod_{i=k_0+1}^{\hat{k}} (1 - \hat{a}_i^2(i)) \right] \end{aligned}$$

Or :

$$\begin{aligned} \left[1 - \prod_{i=k_0+1}^{\hat{k}} (1 - \hat{a}_i^2(i)) \right] &= \left[1 - \left(1 - \sum_{i=k_0+1}^{\hat{k}} \hat{a}_i^2(i) + \right. \right. \\ &\quad \left. \left. + \sum_{\substack{i,j=k_0+1 \\ i \neq j}}^{\hat{k}} \hat{a}_i^2(i) \hat{a}_j^2(j) - \dots \right) \right] \\ &= \left[\sum_{i=k_0+1}^{\hat{k}} \hat{a}_i^2(i) - \sum_{\substack{i,j=k_0+1 \\ i \neq j}}^{\hat{k}} \hat{a}_i^2(i) \hat{a}_j^2(j) + \dots \right] \end{aligned}$$

d'où :

$$\lim N E \left[1 - \prod_{i=k_0+1}^{\hat{k}} (1 - \hat{a}_i^2(i)) \right] = \lim N E \left[\sum_{i=k_0+1}^{\hat{k}} \hat{a}_i^2(i) - \sum_{\substack{i,j=k_0+1 \\ i \neq j}}^{\hat{k}} \hat{a}_i^2(i) \hat{a}_j^2(j) + \dots \right]$$

Montrons que : $\lim_{N \rightarrow \infty} N E \sum_{\substack{i,j=k_0+1 \\ i \neq j}}^{\hat{k}} \hat{a}_i^2(i) \hat{a}_j^2(j) = 0$. Le reste

des termes dans la somme ci-dessus disparaissent de la même manière.

On a vu que les v.a. : $\sqrt{N} \hat{a}_i(i)$, $i = k_0 + 1, \dots, K$ sont asymptotiques indépendantes et de loi normale $N(0,1)$.

De plus : $\hat{a}_i(i) = O_p\left(\frac{1}{\sqrt{N}}\right)$ (cf. lemme 3 , § 2 chap. II) ,

pour $i \geq k_0 + 1$ à K , d'où $\hat{a}_i^2(i) \hat{a}_j^2(j) = O_p\left(\frac{1}{N^2}\right)$, d'où le résultat .

($\hat{k} \leq K$, la somme $\sum_{i=k_0+1}^{\hat{k}}$ est finie, les v.a. $\sqrt{N} \hat{a}_i(i)$ sont uniformément bornées en probabilité).(cf. [7] p. 32).

Ainsi il en résulte :

$$\begin{aligned} \lim N E [L(a, \hat{a}(\hat{k})) - L(a, \hat{a}(k_0))] &= \sigma_\epsilon^2 \lim N E \left[\sum_{i=k_0+1}^{\hat{k}} \hat{a}_i^2(i) \right] \\ &= \sigma_\epsilon^2 \lim N E \left[\sum_{k=k_0+1}^K \sum_{i=k_0+1}^k \hat{a}_i^2(i) \cdot I_A \right] \\ &= \sigma_\epsilon^2 \left(\sum_{k=k_0+1}^K \lim E \left\{ N \sum_{i=k_0+1}^k \hat{a}_i^2(i) \cdot I_A \right\} \right) \end{aligned}$$

où $A = \{\hat{k} = k\}$.

Pour avoir le résultat (5.4) du théorème il suffit de montrer que :

$$\lim_{N \rightarrow \infty} E \left\{ N \sum_{i=k_0+1}^k \hat{a}_i^2(i) \cdot I_A \right\} = e_{k-k_0} \cdot q(K-k), \quad (k_0 < k \leq K).$$

Soit :

$$S_i = (Z_1^{-\alpha}) + \dots + (Z_i^{-\alpha}) = \sum_{j=1}^i Z_j - i\alpha,$$

où les Z_i sont des v.a. indépendantes et de même loi : χ_1^2 .

De la démonstration du théorème 1, § 2, point 5, on a :

$$\sum_{i=k_0+1}^k N \hat{a}_i^2(i) \xrightarrow{L} \sum_{i=1}^{k-k_0} Z_i = S_{k-k_0} + \alpha(k-k_0)$$

et de la page 58 on a :

$$\begin{aligned} \lim E(I_A) &= \lim \Pr(\hat{k}=k) = \Pr(S_m > 0, 1 \leq m \leq k-k_0, S_m - S_{k-k_0} \leq 0, k-k_0 < m \leq K-k_0) \\ &= E(I_{(S_1 > 0, \dots, S_{k-k_0} > 0)} \cap (S_m - S_{k-k_0} \leq 0, k-k_0 < m \leq K-k_0)) \end{aligned}$$

d'où, en notant l'indépendance des deux événements E_1, E_2 associés à

$I_{E_1} \cap E_2$, et l'intégrabilité uniforme des $\sum_{i=k_0+1}^K N \hat{a}_i^2(i)$, (cf. [7] p. 32):

$$\begin{aligned} (5.7) \quad \lim E \left\{ \sum_{i=k_0+1}^k N \hat{a}_i^2(i) \cdot I_A \right\} &= E \left[(S_{k-k_0} + \alpha(k-k_0)) \cdot I_{(S_1 > 0, \dots, S_{k-k_0} > 0)} \cap (S_m - S_{k-k_0} \leq 0, k-k_0 < m \leq K-k_0) \right] \\ &= E \left[(S_{k-k_0} + \alpha(k-k_0)) \cdot I_{(S_1 > 0, \dots, S_{k-k_0} > 0)} \right] \cdot E \left[I_{(S_m - S_{k-k_0} \leq 0, k-k_0 < m \leq K-k_0)} \right] \\ &= E \left[(S_{k-k_0} + \alpha(k-k_0)) \cdot I_{(S_1 > 0, \dots, S_{k-k_0} > 0)} \right] \cdot \Pr(S_m - S_{k-k_0} \leq 0, k-k_0 < m \leq K-k_0) \\ &= E \left[(S_{k-k_0} + \alpha(k-k_0)) \cdot I_{(S_1 > 0, \dots, S_{k-k_0} > 0)} \right] \cdot q(K-k). \end{aligned}$$

Reste à évaluer $E[(S_{k-k_0} + \alpha(k-k_0)) I_{(S_1 > 0, \dots, S_{k-k_0} > 0)}]$.

$$(5.7') \quad = E[S_{k-k_0} \cdot I_{(S_1 > 0, \dots, S_{k-k_0} > 0)}] + \alpha(k-k_0) \Pr(S_1 > 0, \dots, S_{k-k_0} > 0).$$

Du lemme 3 on a :

$$(5.8) \quad E[S_{k-k_0} \cdot I_{(S_1 > 0, \dots, S_{k-k_0} > 0)}] = \sum_{k-k_0}^* \left[\prod_{i=1}^{k-k_0} \frac{1}{r_i!} \left(\frac{\alpha_i}{i}\right)^{r_i} \right] \left[\sum_{i=1}^{k-k_0} \frac{r_i}{\alpha_i} E(S_i \cdot I_{(S_i > 0)}) \right]$$

où $\alpha_i = \Pr(S_i > 0)$.

Calcul de $E(S_i \cdot I_{(S_i > 0)})$, où $S_i = \sum_{j=1}^i Z_j - i\alpha$, Z_j de loi χ_1^2 .

La v.a. $\sum_{j=1}^i Z_j = \chi_i^2$ a pour densité :

$$f(x) = \frac{1}{2^{i/2} \Gamma(\frac{i}{2})} x^{(i/2)-1} e^{-(x/2)}. \quad \text{Il vient :}$$

$$\begin{aligned} E(S_i \cdot I_{(S_i > 0)}) &= \int_{i\alpha}^{\infty} (x-i\alpha) f(x) dx \\ &= \int_{i\alpha}^{\infty} \frac{1}{2^{i/2} \Gamma(\frac{i}{2})} x^{i/2} e^{-x/2} dx - i\alpha \Pr(\chi_i^2 > i\alpha). \end{aligned}$$

En utilisant $\Gamma(x+1) = x \Gamma(x)$ on obtient :

$$\begin{aligned} \int_{i\alpha}^{\infty} \frac{1}{2^{i/2} \Gamma(\frac{i}{2})} x^{i/2} e^{-x/2} dx &= \int_{i\alpha}^{\infty} \frac{1}{2^{i/2} \frac{2}{i} \Gamma(\frac{1}{2} i+1)} x^{((i+2)/2)-1} e^{-x/2} dx \\ &= \int_{i\alpha}^{\infty} i \frac{1}{2^{(i+2)/2} \Gamma(\frac{1}{2} (i+2))} x^{((i+2)/2)-1} e^{-x/2} dx \\ &= i \Pr(\chi_{i+2}^2 > i\alpha) \end{aligned}$$

d'où :

$$E(S_i \cdot I_{(S_i > 0)}) = i \Pr(\chi_{i+2}^2 > i\alpha) - i\alpha \Pr(\chi_i^2 > i\alpha) .$$

Par conséquent, en réécrivant (5.8) avec $\alpha_i = \Pr(S_i > 0) = \Pr(\chi_i^2 > i\alpha)$ on obtient :

$$\begin{aligned} E(S_{k-k_0} \cdot I_{(S_1 > 0, \dots, S_{k-k_0} > 0)}) &= \sum_{k-k_0}^* \left[\prod_{i=1}^{k-k_0} \frac{1}{r_i!} \left\{ \frac{\Pr(\chi_i^2 > i\alpha)}{i} \right\}^{r_i} \right] \\ &\quad \left[\sum_{i=1}^{k-k_0} \left(\frac{r_i}{\Pr(\chi_i^2 > i\alpha)} \right) (i \Pr(\chi_{i+2}^2 > i\alpha) - i\alpha \Pr(\chi_i^2 > i\alpha)) \right] \\ &= \sum_{k-k_0}^* \left[\prod_{i=1}^{k-k_0} \frac{1}{r_i!} \left\{ \frac{\Pr(\chi_i^2 > i\alpha)}{i} \right\}^{r_i} \right] \left[\sum_{i=1}^{k-k_0} i r_i \left(\frac{\Pr(\chi_{i+2}^2 > i\alpha) - \alpha \Pr(\chi_i^2 > i\alpha)}{\Pr(\chi_i^2 > i\alpha)} \right) \right] \\ &= \sum_{k-k_0}^* \left[\prod_{i=1}^{k-k_0} \frac{1}{r_i!} \left\{ \frac{\Pr(\chi_i^2 > i\alpha)}{i} \right\}^{r_i} \right] \left[\left(\sum_{i=1}^{k-k_0} i r_i \left(\frac{\Pr(\chi_{i+2}^2 > i\alpha)}{\Pr(\chi_i^2 > i\alpha)} \right) - \sum_{i=1}^{k-k_0} i r_i \alpha \right) \right] \end{aligned}$$

Or les entiers r_i (cf. lemme 1 § 2) sont tels que :

$$\sum_{i=1}^{k-k_0} i r_i = k - k_0$$

et de la définition de $p(n)$ (cf. lemme 1 § 2) :

$$p(n) = \Pr(S_1 > 0, \dots, S_n > 0) = \sum_n^* \left\{ \prod_{i=1}^n \frac{1}{r_i!} \left(\frac{\Pr(S_i > 0)}{i} \right)^{r_i} \right\}$$

il vient :

$$(5.9) \quad E(S_{k-k_0} \cdot I_{(S_1 > 0, \dots, S_{k-k_0} > 0)}) = e_{k-k_0} - \alpha(k-k_0) \Pr(S_1 > 0, \dots, S_{k-k_0} > 0)$$

où

$$e_{k-k_0} = \sum_{k-k_0}^* \prod_{i=1}^{k-k_0} \frac{1}{r_i!} \left\{ \frac{\Pr(\chi_i^2 > i\alpha)}{i} \right\}^{r_i} \left[\sum_{i=1}^{k-k_0} \left\{ i r_i \frac{\Pr(\chi_{i+2}^2 > i\alpha)}{\Pr(\chi_i^2 > i\alpha)} \right\} \right]$$

Conclusion : En regroupant les résultats (5.7), (5.7'),

et (5.9) il en résulte :

$$\lim E \left\{ \sum_{i=k_0+1}^k N \hat{a}_i^2(i) I_A \right\} = e_{k-k_0} \cdot q(K-k)$$

or ; (cf. p. 71) :

$$\lim N E \left[L(a, \hat{a}(k)) - L(a, \hat{a}(k_0)) \right] = \sigma_\varepsilon^2 \left(\sum_{k=k_0+1}^K \lim E \left\{ \sum_{i=k_0+1}^k N \hat{a}_i^2(i) I_A \right\} \right)$$

d'où le résultat désiré.

Remarques :-Du résultat obtenu on déduit des bornes de la croissance du risque ; les quantités e_i et $q(n)$ étant inférieures à 1 :

$$0 \leq \lim N E \left[L(a, \hat{a}(k)) - L(a, \hat{a}(k_0)) \right] \leq (K - k_0) \sigma_\varepsilon^2 .$$

- La borne inférieure est atteinte si k_0 est connu et $\hat{k} = k_0$.

- La borne supérieure est atteinte si on prend la sélection triviale $\hat{k} = K$.

- Nous ignorons la limite dans (5.4) dans le cas où l'estimateur \hat{k} est convergent.

- L'introduction du processus auxiliaire $(X_t, t \in \mathbb{Z})$ dans la prédiction du processus $(Y_t, t \in \mathbb{Z})$ pourra se justifier de la manière suivante où (Y_t) vérifie (5.0).

- Soit $(Y_t, t \in \mathbb{Z})$ un processus réel centré faiblement stationnaire ayant pour fonction de covariance $R(h)$. Le meilleur prédicteur linéaire \hat{Y}_t de Y_t basé sur k variables $(Y_{t-1}, \dots, Y_{t-k})$ est :

$$\hat{Y}_t = \sum_{i=1}^k a_i Y_{t-i}$$

au sens que l'erreur de prédiction :

$$\phi = E(Y_t - \hat{Y}_t)^2$$

est minimale.

En annulant les dérivées partielles de ϕ par rapport aux a_i on obtient les équations normales :

$$\begin{pmatrix} R(0) & \dots & R(k-1) \\ \vdots & & \vdots \\ R(k-1) & \dots & R(0) \end{pmatrix} \begin{pmatrix} a_1 \\ \vdots \\ a_k \end{pmatrix} = \begin{pmatrix} R(1) \\ \vdots \\ R(k) \end{pmatrix}$$

Ces équations représentent les équations de Yule-Walker (cf. sect. 2.2. chap. I).

Par conséquent, les coefficients a_1, \dots, a_k peuvent être considérés comme les paramètres d'une équation aux différences stochastiques

d'ordre k vérifiée par un processus $(X_t, t \in \mathbb{Z})$ centré faiblement stationnaire ayant même fonction de covariance que le processus $(Y_t, t \in \mathbb{Z}) : R(h) = E X_t X_{t+h} = E Y_t Y_{t+h}$. Par suite $(X_t, t \in \mathbb{Z})$ est un autorégressif d'ordre k vérifiant :

$$X_t = a_1 X_{t-1} + \dots + a_k X_{t-k} + \varepsilon_t, \quad t \in \mathbb{Z}$$

où (ε_t) est un bruit blanc de moyenne nulle et de variance $\sigma^2 = R(0) - \sum_{i=1}^k a_i R(i)$.

Une observation de taille N du processus $(X_t, t \in \mathbb{Z})$ permet de construire des estimateurs \hat{a}_i et donc de donner un estimateur du prédicteur \hat{Y}_t qu'on notera de la même façon :

$$\hat{Y}_t = \sum_{i=1}^k \hat{a}_i Y_{t-i}.$$

Ainsi l'estimation des coefficients a_i du prédicteur \hat{Y}_t ne nécessite que la connaissance de la fonction de covariance du processus $(Y_t, t \in \mathbb{Z})$ et en introduisant un processus auxiliaire $(X_t, t \in \mathbb{Z})$ ayant même fonction de covariance que $(Y_t, t \in \mathbb{Z})$, on ne modifie en rien l'expression du prédicteur \hat{Y}_t (les deux perturbations agissant sur les deux processus ont même variance).

Ceci pourrait aussi se voir en considérant la représentation spectrale d'un processus faiblement stationnaire. La connaissance de la fonction de covariance permet une représentation du processus.

En effet :

$$X_t = \int_{-\pi}^{+\pi} e^{it\lambda} dZ(\lambda)$$

où le processus $Z(\lambda)$ est à accroissements orthogonaux et :

$$E|Z(\lambda)|^2 = F(\lambda)$$

où la répartition spectrale $F(\lambda)$ du processus $(X_t, t \in \mathbb{Z})$ est liée

à la fonction de covariance par :

$$R(h) = \int_{-\pi}^{+\pi} e^{it\lambda} dF(\lambda)$$

Ainsi l'évolution d'un processus faiblement stationnaire est déterminé par la donnée de sa fonction de covariance.

CHAPITRE III

PROCESSUS ARIMA.

I - NOTATIONS.-

Soit (Ω, \mathcal{A}, P) un espace probabilisé et $(X_t, t \in \mathbb{Z})$ un processus réel gaussien. On désignera par B l'opérateur retard :

$$\forall t \in \mathbb{Z} \quad , \quad B(X_t) = X_{t-1} \quad ,$$

par ∇ : l'opérateur différence :

$$\nabla = I - B \quad , \quad \nabla X_t = X_t - X_{t-1} = (1-B) X_t \quad ,$$

et par ∇^d , $d \geq 1$, l'opérateur :

$$\nabla^d X_t = \nabla^{d-1} X_t - \nabla^{d-1} X_{t-1} = (1-B)^d X_t \quad .$$

Le processus $(X_t, t \in \mathbb{Z})$ est un ARIMA (p, d, q) (autoregressive integrated moving average process) si les différences d'ordre d : $\nabla^d X_t$, vérifient l'équation aux différences suivantes :

$$A(B) \nabla^d X_t = B(B) e_t \quad , \quad t \in \mathbb{Z}$$

où :

$$A(B) = (1 - a_1 B - \dots - a_p B^p) \quad p \geq 0 \quad , \quad a_p \neq 0$$

$$B(B) = (1 - b_1 B - \dots - b_q B^q) \quad q \geq 0 \quad , \quad a_q \neq 0$$

Les deux polynômes $A(B)$ et $B(B)$ ont les racines à l'extérieur du cercle unité, et n'ont pas de racines communes.

$(e_t, t \in \mathbb{Z})$ est un bruit blanc de moyenne nulle et de variance σ^2 .

On remarquera que le processus $(\nabla^d X_t, t \in \mathbb{Z})$ est un ARMA stationnaire. Enfin, on suppose :

$$X_0 = X_{-1} = \dots = X_{-(p+d-1)} = 0 .$$

II - ESTIMATION DES PARAMETRES.- (cf. [21] [22] [41]) .

On supposera par la suite $q = 0$ et on notera le processus ARIMA $(p,d,0)$ par ARI (p,d) . Alors les différences d'ordre d . $\nabla^d X_t$ satisfont :

$$(1 - a_1 B - \dots - a_p B^p) \nabla^d X_t = e_t$$

ou
$$(1 - a_1 B - \dots - a_p B^p) (1-B)^d X_t = e_t .$$

En développant $(1-B)^d$ et en regroupant les termes, on obtient une équation que l'on notera :

$$(1 - \alpha_1 B - \dots - \alpha_{p+d} B^{p+d}) X_t = e_t$$

Désignons par :

$$\hat{\alpha}_N = (\hat{\alpha}_1, \dots, \hat{\alpha}_{p+d})$$

l'estimateur des moindres carrés obtenu par la régression de X_t sur $X_{t-1}, \dots, X_{t-(p+d)}$, pour $t = 1, \dots, N$.

En posant :

$$\phi_t = (X_{t-1}, \dots, X_{t-(p+d)})'$$

et où les f_i sont les coefficients de B^i dans le développement de $(1-B)^d$ i.e.

$$(1-B)^d = f_0 + f_1 B + \dots + f_d B^d .$$

Remarque : Le lemme précédent établit la relation entre les paramètres a et α .

Convergence presque sûre de l'estimateur.

Théorème 1.- ([21]) .

Soit $(X_t, t \in \mathbb{Z})$ un processus ARI (p,d) gaussien.

Si $\sum_{t=0}^{d-1} E|X_t|^2 < \infty$, alors :

$$\lim_{N \rightarrow \infty} \hat{\alpha}_N = \alpha \quad \text{p.s.}$$

Loi limite de l'estimateur.

Théorème 2.- ([21])

Soit $(X_t, t \in \mathbb{Z})$ un processus ARI (p,d) gaussien.

Si $\sum_{t=0}^{d-1} E|X_t|^2 < \infty$, alors :

$$\sqrt{N} (\hat{\alpha}_N - \alpha) \xrightarrow{L} N(0, \sigma^2 D \Gamma_p^{-1} D') \quad \text{quand } N \rightarrow \infty ,$$

où :

$$\Gamma_p = \begin{bmatrix} R(0) & R(1) & \dots & R(p-1) \\ R(1) & R(0) & \dots & R(p-2) \\ \vdots & \vdots & \ddots & \vdots \\ \vdots & \vdots & \vdots & \vdots \\ R(p-1) & R(p-2) & \dots & R(0) \end{bmatrix}$$

et $R(\tau) = E \nabla^d X(t+\tau) \nabla^d X(t)$.

De plus, l'estimateur $\hat{\alpha}_N$ possède des propriétés asymptotiques optimales.

III - DETERMINATION DU DEGRE DE DIFFERENTIATION.-

Soit $(X_t, t \in \mathbb{Z})$ un processus ARI(p,d). Le processus des différences $(\nabla^d X_t, t \in \mathbb{Z})$ satisfait :

$$(1 - a_1 B - \dots - a_p B^p) \nabla^d X_t = e_t$$

que l'on écrit :

$$(1 - \alpha_1 B - \dots - \alpha_{p+d} B^{p+d}) X_t = e_t$$

ou encore :

$$X_t = \sum_{j=1}^{p+d} \alpha_j X_{t-j} + e_t .$$

Ainsi, le processus $(X_t, t \in \mathbb{Z})$ sera considéré comme un autorégressif non stationnaire d'ordre $p+d$. Son équation polynomiale associée sera :

$$m^{p+d} - \sum_{j=1}^{p+d} \alpha_j m^{p+d-j} = 0 ,$$

admettant comme racines :

$$m_1 = \dots = m_d = 1 .$$

$$|m_j| < 1 , \quad j = d+1 , \dots , p+d .$$

Il en résulte que l'équation polynômiale de ce type de processus admet "d" racines égales à l'unité, les autres étant strictement inférieures à 1.

On observe le processus $(X_t, t \in \mathbb{Z}) : X_1 \dots\dots X_N$.

On utilise la statistique $S_N(k)$ étudiée au chap. II pour estimer l'ordre $p+d$.

Ensuite on construit les estimateurs :

$$\hat{\alpha}_1, \dots, \hat{\alpha}_{p+d}$$

à l'aide de

$$\hat{\alpha}_N = \left(\sum_{t=1}^N \phi_t \phi_t' \right)^{-1} \sum_{t=1}^N \phi_t X_t$$

où $\phi_t = (X_{t-1}, \dots, X_{t-(p+d)})'$.

L'équation polynômiale estimée s'écrit alors :

$$m^{p+d} - \sum_{j=1}^{p+d} \hat{\alpha}_j m^{p+d-j} = 0$$

Une résolution de cette équation fournit $p+d$ racines. Un examen des racines "significativement proche de 1" permet alors de décider du degré de différentiation du processus observé puis de déterminer le nombre p . L'efficacité de cette méthode est étroitement liée aux justifications théoriques nécessaires.

Convergence de la procédure.-

Dans la méthode de détermination du degré de différentiation d , décrite précédemment, on a fait correspondre à chaque observation (X_1, \dots, X_N) de taille N , une estimation du nombre de racines "significativement proche de 1".

$$\text{Désignons par : } \begin{cases} |m_i| = 1 & , \quad i = 1, \dots, d \\ |m_i| < 1 & , \quad i = d+1, \dots, p+d \end{cases}$$

les racines du polynôme associé au processus ARIMA (p,d,0).

Notons par I_i^α la région de confiance : $[m_i - \alpha, m_i + \alpha]$ de la racine m_i pour $i = 1, \dots, p+d$ (dans le cas des racines complexes on prendra comme région de confiance des couronnes), avec α un nombre positif donné.

On choisit α tel que :

$$\alpha < \frac{1 - \sup \{|m_i|, \quad i = d+1, \dots, p+d\}}{2}$$

pour assurer que l'intervalle de la racine de module 1 ne coupe celui de la plus grande racine : $\sup \{|m_i|, \quad i = d+1, \dots, p+d\}$.

On définit l'estimateur :

$$(X_1, \dots, X_N) \longrightarrow \hat{d}^\alpha(X_1, \dots, X_N),$$

où $\hat{d}^\alpha(X_1, \dots, X_N)$ est l'estimation du nombre de racines dans l'intervalle $[1 - \alpha, 1 + \alpha]$ pour l'échantillon X_1, \dots, X_N .

Sous les conditions précédentes, on a l'inclusion :

$$\{\hat{d}^\alpha = d\} \supseteq \left\{ \bigcap_{i=1}^{p+d} (\hat{m}_i \in I_i^\alpha) \right\}$$

D'où, de la convergence des estimateurs \hat{m}_i vers m_i , $\forall i$, il en résulte

Proposition. - $\lim_{N \rightarrow \infty} P(\hat{d}^\alpha = d) = 1.$

Remarques :

1) De la convergence presque sûre des estimateurs \hat{m}_i (cf. Th. 1, § 2), on obtient :

$$\hat{d}^\alpha \longrightarrow d \text{ p.s.}$$

ce qui montre que la procédure de détermination du degré de différentiation d décrite dans le § 3 est convergente presque sûrement.

2) On peut choisir des régions de confiance I_i^α de la forme :

$$I_i^\alpha(k_N) = \left[m_i - \frac{\alpha}{k_N}, m_i + \frac{\alpha}{k_N} \right]$$

où α est une constante et k_N un nombre croissant indéfiniment. Le choix de la suite $(k_N, N \in \mathbb{N})$ est lié à la vitesse d'estimation des racines m_i .

Comme la vitesse d'estimation des coefficients α_i des polynômes associés est de l'ordre de \sqrt{N} (cf. Th. 2, § 2) on pourrait espérer obtenir la vitesse d'estimation des racines au moyen des transformations symétriques des racines.

Mais en tous les cas, cette vitesse ne pourra excéder \sqrt{N} (\sqrt{N} en cas de normalité asymptotique), par suite on pourra choisir : $k_N = O(\sqrt{N})$, par exemple : $k_N = \text{Log } N$ ou $\text{Log Log } N$ etc ...

Bien sûr nous touchons à un problème de robustesse de l'estimateur \hat{d}^α .

Et de la même manière, en notant par $\hat{d}^\alpha(k_N)$ l'estimation du nombre de racines dans l'intervalle $I_i^\alpha(k_N) = \left[1 - \frac{\alpha}{k_N}, 1 + \frac{\alpha}{k_N} \right]$;

on obtient :

$$\lim_{N \rightarrow \infty} \Pr(\widehat{d}^\alpha(k_N) = d) = 1 .$$

Remarque : La convergence presque sûre des estimateurs des racines résulte de celle des $\widehat{\alpha}_i$.

Loi limite des estimateurs des racines. Considérons l'équation :

$$Z^n - a_1 Z^{n-1} + \dots + (-1)^n a_n = 0 ,$$

où les a_i sont des nombres complexes. Le problème est d'obtenir la loi des racines quand les coefficients aléatoires a_i suivent une loi donnée. Dans [13], l'auteur traite ce problème dans un cas particulier :

On suppose que la partie réelle et la partie imaginaire des nombres a_i sont des v.a. indépendantes et de loi normale $N(0, \sigma^2)$, et que la distribution des a_i est donnée par la densité :

$$\left(\frac{1}{\sqrt{2\pi\sigma}}\right)^{2n} \exp\left[-\frac{1}{2\sigma^2} \sum_{i=1}^n a_i \bar{a}_i\right]$$

où \bar{a}_i désigne le conjugué de a_i .

Alors, en notant Z_1, \dots, Z_n les racines de l'équation, la loi conjointe des racines admet pour densité :

$$\left(\frac{1}{\sqrt{2\pi\sigma}}\right)^{2n} \exp\left[-\frac{1}{2\sigma^2} \left\{ \sum_{j=1}^n Z_j \bar{Z}_j + \dots + Z_1 \bar{Z}_1 \times \dots \times Z_n \bar{Z}_n \right\}\right] \\ \times \sum_{p=1}^n \sum_{q=p+1}^n |Z_p - Z_q|^2$$

Conclusion : La loi des racines n'étant pas très simple, il est difficile de construire des intervalles de confiance ou de poser des tests pour les estimateurs des racines.

IV - ETUDE DE CAS PARTICULIERS.

1°) Marche aléatoire : ARIMA(0,1,0) , $d = 1$.

$$\boxed{X_t = X_{t-1} + e_t} , \quad X_0 = 0 .$$

On observe X_1, \dots, X_N .

L'estimateur $\hat{\alpha}_1$ s'écrit :

$$\hat{\alpha}_1 = \left(\sum_{t=1}^N X_{t-1}^2 \right)^{-1} \sum_{t=1}^N X_t X_{t-1}$$

et l'équation polynômiale associée devient :

$$m - \hat{\alpha}_1 = 0$$

d'où la racine $m = \hat{\alpha}_1$. De plus, des théorèmes 1 et 2 on a :

$$\hat{m}_1 \longrightarrow 1 \text{ p.s. } ,$$

et

$$\sqrt{N}(\hat{m}_1 - 1) \rightarrow N(0, V) .$$

2°) ARIMA(0,2,0).

Le processus $(X_t, t \in \mathbb{Z})$ est tel que : $\nabla^2 X_t = e_t$

d'où :

$$\boxed{X_t = 2 X_{t-1} - X_{t-2} + e_t}$$

C.I. : $X_{-1} = X_0 = 0$.

Les estimateurs de $\alpha_1 = 2$ et $\alpha_2 = -1$ sont :

$$\begin{pmatrix} \hat{\alpha}_1 \\ \hat{\alpha}_2 \end{pmatrix} = M^{-1} \begin{pmatrix} \sum_{t=1}^N X_t X_{t-1} \\ \sum_{t=1}^N X_t X_{t-2} \end{pmatrix}$$

où

$$M = \begin{pmatrix} \sum_1^N X_{t-1}^2 & \sum_1^N X_{t-1} X_{t-2} \\ \sum_1^N X_{t-1} X_{t-2} & \sum_1^N X_{t-1}^2 \end{pmatrix}$$

On résout l'équation : $m^2 - \hat{\alpha}_1 m - \hat{\alpha}_2 = 0$, et de plus

$$(\hat{m}_1, \hat{m}_2) \longrightarrow (1,1) \text{ p.s.}$$

3°) ARIMA(1,2,0).

Un processus ARIMA(1,2,0) est tel que :

$$(1 - a_1 B) \nabla^2 X_t = e_t$$

ou
$$X_t = \alpha_1 X_{t-1} + \alpha_2 X_{t-2} + \alpha_3 X_{t-3} + e_t .$$

C.I. : $X_{-2} = X_{-1} = X_0 = 0 .$

On construit les estimateurs $\hat{\alpha}_1$, $\hat{\alpha}_2$ et $\hat{\alpha}_3$:

$$\begin{pmatrix} \hat{\alpha}_1 \\ \hat{\alpha}_2 \\ \hat{\alpha}_3 \end{pmatrix} = M^{-1} \begin{pmatrix} \sum_1^N X_t X_{t-1} \\ \sum_1^N X_t X_{t-2} \\ \sum_1^N X_t X_{t-3} \end{pmatrix}$$

$$M = \begin{pmatrix} \sum x_{t-1}^2 & \sum x_{t-1} x_{t-2} & \sum x_{t-1} x_{t-3} \\ \sum x_{t-1} x_{t-2} & \sum x_{t-1}^2 & \sum x_{t-2} x_{t-3} \\ \sum x_{t-1} x_{t-3} & \sum x_{t-2} x_{t-3} & \sum x_{t-1}^2 \end{pmatrix}$$

L'équation associée sera :

$$m^3 - \hat{\alpha}_1 m^2 - \hat{\alpha}_2 m - \hat{\alpha}_3 = 0 ,$$

et ses racines sont telles que :

$$\begin{cases} (\hat{m}_1, \hat{m}_2) \longrightarrow (1, 1) \text{ p.s.} \\ \hat{m}_3 \longrightarrow a_1 \text{ p.s.} \end{cases}$$

V - SIMULATIONS.

A partir des réalisations des trois types de processus ARIMA décrits au paragraphe précédent on a essayé de déterminer leur degré de différentiation par la méthode exposée au paragraphe 3.

On a utilisé la statistique $S_N(k)$ avec $\alpha = 7$ et $\beta = 0.3$ pour estimer l'ordre $p+d$. Nous avons augmenté le nombre d'observations dans le 2°) et le 3°) pour améliorer l'estimation de l'ordre et des racines. On a pris $I_1^\alpha = [1 - 0,05 ; 1 + 0,05]$ comme région de confiance de la racine 1.

1°) Marche aléatoire : ARIMA(0,1,0).

Données : - $X_0 = 0$

- $\alpha_1 = 1$

- Le nombre d'observations $N = 100$.

Chaque réalisation du processus est faite à partir de 100 observations. On a enregistré les fréquences des ordres de 0 à 7 dans 100 réalisations du processus : (cf. pour les détails chap. IV).

k	0	1	2	3	4	5	6	7
fr	0	100	0	0	0	0	0	0

Ainsi, il en résulte que l'ordre le plus fréquent est 1.

D'où : $p + d = 1$.

Le calcul de $\hat{\alpha}_1$ donne : $\hat{\alpha}_1 = 0.990959$. On en déduit :

$$\hat{m}_1 = 0.990959 .$$

Finalement, la valeur estimée \hat{m}_1 étant "très proche de 1" , on en conclut que $d = 1$.

Par conséquent, $p = 0$. Cela correspond bien à la marche aléatoire ARIMA(0,1,0) simulée.

2°) ARIMA(0,2,0).

a) Données : $X_0 = X_{-1} = 0$; $\alpha_1 = 2$, $\alpha_2 = -1$, $N = 50$.

Le tableau des fréquences des ordres est :

k	0	1	2	3	4	5	6	7
fr	0	0	99	0	1	0	0	0

d'où : $p + d = 2$.

Le calcul de $\hat{\alpha}_1$ et $\hat{\alpha}_2$ pour ces réalisations donne :

$$\hat{\alpha}_1 = 1.94442$$

$$\hat{\alpha}_2 = - 0.944846$$

Une résolution de l'équation associée donne comme racines :

$$\hat{m}_1 = 0.99081$$

$$\hat{m}_2 = 0.95360$$

Les racines étant "très proches de 1" on en déduit $d = 2$, donc

$p = 0$. D'où X_t est un ARIMA(0,2,0) ce qui correspond bien au processus simulé.

b) Pour $N = 200$.

On a obtenu :

k	0	1	2	3	4	5	6	7
fr	0	0	100	0	0	0	0	0

d'où : $p + d = 2$.

Les valeurs de $\hat{\alpha}_1$ et $\hat{\alpha}_2$ pour ces réalisations sont :

$$\left\{ \begin{array}{l} \hat{\alpha}_1 = 1.98418 \\ \hat{\alpha}_2 = -0.984154 \end{array} \right.$$

ce qui donnent pour les racines estimées :

$$\hat{m}_1 = 1.00150$$

$$\hat{m}_2 = 0.98267$$

Donc : $d = 2$ et $p = 0$. Finalement $X(t)$ est un $ARIMA(0,2,0)$.

3°) ARIMA(1,2,0).

Données : $X_{-2} = X_{-1} = X_0 = 0$; $\alpha_1 = 2.5$, $\alpha_2 = -2$; $\alpha_3 = 0.5$

a) N = 50

On a le tableau des fréquences :

k	0	1	2	3	4	5	6	7
fr	0	0	39	56	5	0	0	0

d'où , l'ordre le plus fréquent est 3. Alors $p + d = 3$.

Les valeurs des $\hat{\alpha}_1$, $\hat{\alpha}_2$ et $\hat{\alpha}_3$ sont :

$$\hat{\alpha}_1 = 2.44118$$

$$\hat{\alpha}_2 = -1.91085$$

$$\hat{\alpha}_3 = 0.469387$$

La résolution de l'équation associée a conduit aux racines :

$$\hat{m}_1 = 0.9582$$

$$\hat{m}_2 = 0.9865$$

$$\hat{m}_3 = 0.4967$$

Deux racines étant "très proches de 1" on en déduit $d = 2$, donc $p = 1$.
D'où $X(t)$ est un ARIMA(1,2,0).

b) $N = 80$

On a obtenu les résultats suivants :

k	0	1	2	3	4	5	6	7
fr	0	0	38	60	2	0	0	0

d'où : $p + d = 3$.

Les valeurs des estimateurs :

$$\hat{\alpha}_1 = 2.45877$$

$$\hat{\alpha}_2 = -1.93531$$

$$\hat{\alpha}_3 = 0.476430$$

Les racines associées :

$$\hat{m}_1 = 0.9735$$

$$\hat{m}_2 = 0.9919$$

$$\hat{m}_3 = 0.4934$$

Il en résulte $d = 2$ et $p = 1$. Donc on aboutit à la même conclusion sur le processus $X(t)$.

c) N = 90

On a obtenu les résultats suivants :

k	0	1	2	3	4	5	6	7
fr	0	0	31	67	2	0	0	0

d'où : p + d = 3.

Les valeurs des estimateurs :

$$\hat{\alpha}_1 = 2.46384$$

$$\hat{\alpha}_2 = -1.94483$$

$$\hat{\alpha}_3 = 0.480883$$

les racines associées :

$$\hat{m}_1 = 0.9740$$

$$\hat{m}_2 = 0.9923$$

$$\hat{m}_3 = 0.4976$$

On en déduit : $d = 2$ et $p = 1$. Donc $X(t) \sim \text{ARIMA}(1,2,0)$.

d) N = 100

On a le tableau des fréquences :

k	0	1	2	3	4	5	6	7
fr	0	0	39	60	1	0	0	0

d'où $p + d = 3$. On aboutit à la même conclusion sur $X(t)$ que précédemment.

e) $N = 200$

On a obtenu le résultat suivant :

k	0	1	2	3	4	5	6	7
fr	0	0	62	38	0	0	0	0

Il en résulte : $p + d = 2$. Comme le vrai ordre du processus observé est 3, on en déduit que la statistique sous-estime l'ordre et par conséquent elle est non convergente dans le cas des processus non stationnaire. Mais par contre la précision sur l'estimation des racines augmente quand N croît indéfiniment. Par exemple pour $N = 200$ on a obtenu les estimations suivantes :

$$\left\{ \begin{array}{ll} \hat{\alpha}_1 = 2.47806 & \hat{m}_1 = 0.9854 \\ \hat{\alpha}_2 = -1.96371 & \hat{m}_2 = 0.9997 \\ \hat{\alpha}_3 = 0.485650 & \hat{m}_3 = 0.4930 \end{array} \right.$$

qui sont plus précises que celles obtenues dans a) b) c) et d).

VI - CONCLUSION. -

Nous avons vu que la statistique $S_N(k)$ a tendance à sous estimer l'ordre du processus observé quand la taille de l'échantillon augmente indéfiniment. Par contre pour des petits échantillons, l'estimation de l'ordre est satisfaisante comme semble le confirmer les exemples précédents, et les résultats obtenus sur l'évaluation de d sont convaincants. A notre connaissance, aucune étude sur l'estimation convergente de l'ordre d'un autorégressif non-stationnaire n'a été faite et par conséquent le problème reste ouvert.

CHAPITRE IV

SIMULATION.

I - STATISTIQUE $S_N(k)$.-

Afin d'illustrer les résultats obtenus au chapitre II on a réalisé une simulation sur un processus autorégressif d'ordre 1 :

$$X_t = a_1 X_{t-1} + \varepsilon_t \quad (1)$$

avec $X_0 = 0$, $|a_1| < 1$ et (ε_t) v.a. i.i.d. de loi normale $N(0,1)$.

On construit un échantillon de taille N du processus : en appelant pour chaque t une variable aléatoire ε_t générée par un sous-programme et en utilisant l'équation (1).

Pour chaque échantillon donné, on construit les estimateurs $\hat{a}_1(k)$ et $\hat{\sigma}^2(k)$ (cf. chap. I) et puis on calcule $S_N(k) = (N + \alpha(N)k) \hat{\sigma}^2(k)$ pour "k" variant de 0 à K (on a pris $K = 7$).

Pour le coefficient a_1 on a envisagé les deux cas suivants :

$$a_1 = 0.2$$

$$a_1 = 0.8$$

correspondant a des corrélations, respectivement faible et forte.

A chaque réalisation X_1, \dots, X_N du processus , $S_N(k)$ détermine une estimation de l'ordre comprise entre 0 et 7 . On a simulé 100 échantillons et enregistré les fréquences des ordres obtenus par $S_N(k)$.

Exemple : $\alpha = 5$, $\beta = 0,1$, $a_1 = 0.8$

N = 100

ordres	0	1	2	3	4	5	6	7
fr	0	94	5	0	1	0	0	0

c'est-à-dire parmi les 100 échantillons utilisés 94 ont donné l'ordre 1 , 5 ont donné l'ordre 2 et 1 a donné l'ordre 4.

La précision aurait pu être augmentée en accroissant considérablement le nombre d'échantillons examinés ce qui aurait rendu cette étude trop coûteuse.

On a utilisé deux types de croissances pour $\alpha(N)$:

1^{er} cas : $\alpha(N) = \alpha N^\beta$ avec $\alpha = 2,3,\dots,7$
et $\beta = 0.097 ; 0.1 ; 0.3 ; 0.5$

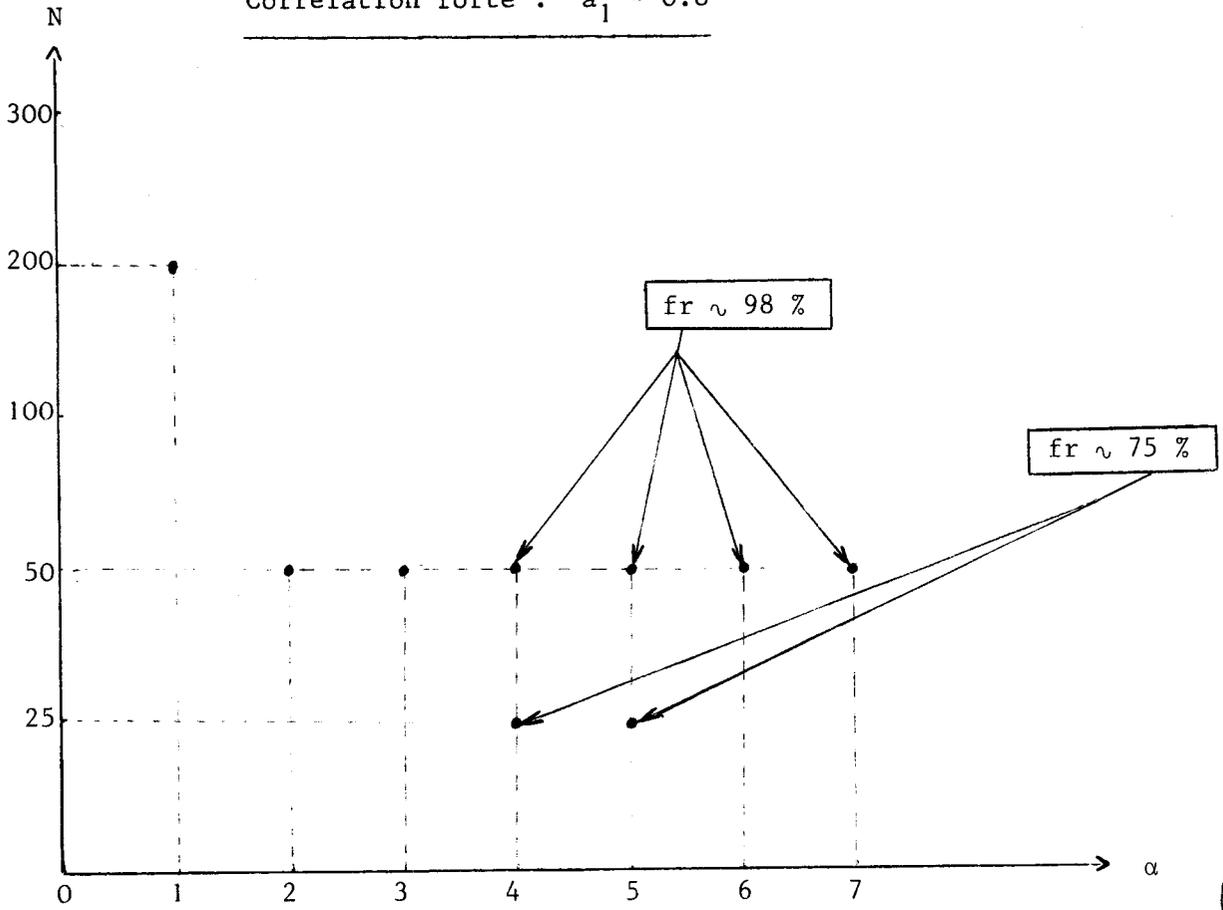
2^{ème} cas : $\alpha(N) = \alpha \text{Log } N$ avec $\alpha = 1,2,\dots,7$

Il a paru intéressant de comparer les résultats obtenus avec ceux que donnent les critères de Schwarz et d'Akaike. On a observé, expérimentalement, que les choix suivants donnent de bons résultats :

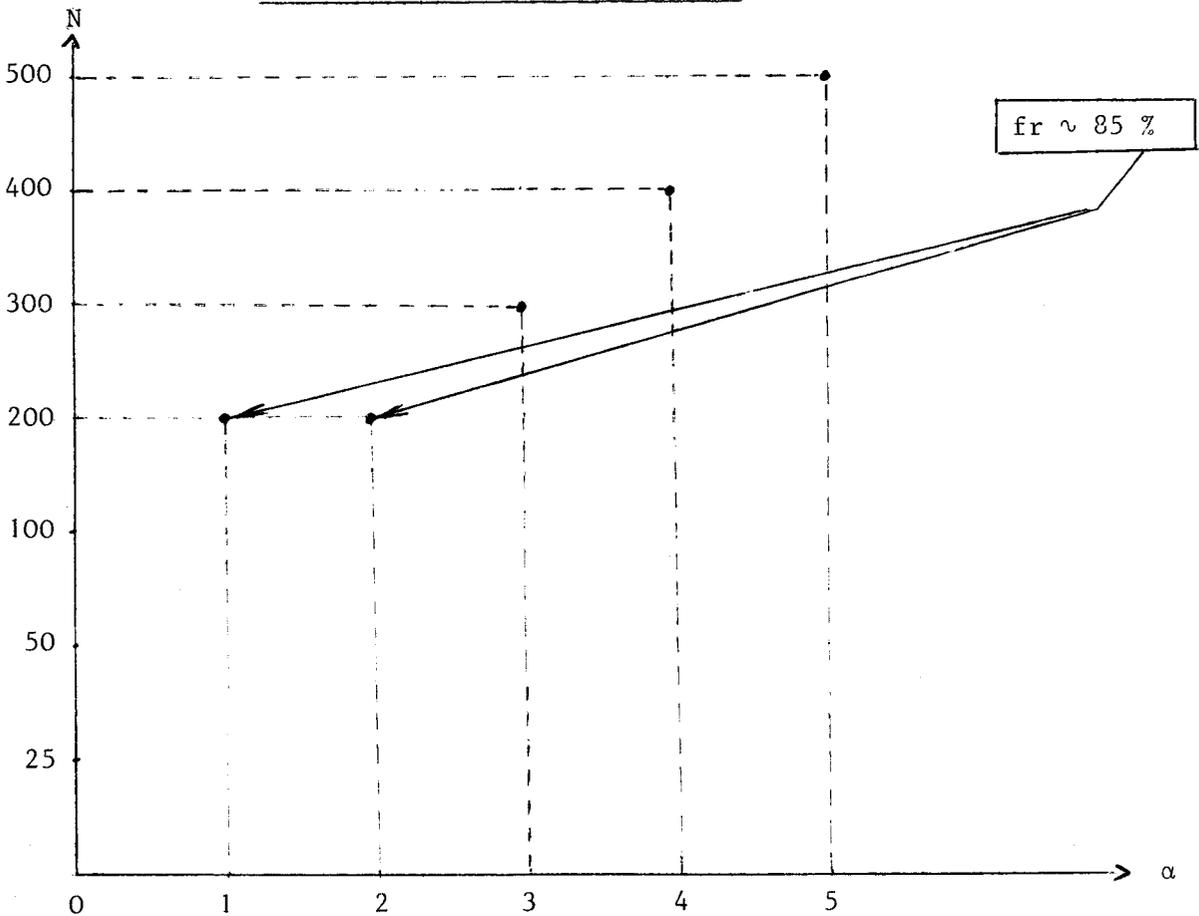
Ⓐ Pour $S_N(k) = (N + \alpha k \text{Log } N) \hat{\sigma}^2(k)$.

Les tableaux suivants indiquent le nombre minimal d'observations, en fonction de α , pour que la fréquence du vrai ordre soit supérieur à 80 % .

Corrélation forte : $a_1 = 0.8$



Corrélation faible : $a_1 = 0.2$



Il en résulte :

- Pour $\underline{a_1 = 0.8}$, la fréquence du vrai ordre est de l'ordre de 96 % , pour un nombre d'observations minimal égal à 50, si $\alpha > 2$.

Pour $N = 25$, la fréquence est de l'ordre de 75 % pour $\alpha \geq 4$.

- Pour $\underline{a_2 = 0.2}$, la fréquence du bon ordre est de l'ordre de 85 % , pour $N = 200$, $\alpha \geq 1$.

ⓑ Pour $S_N(k) = (N + \alpha k N^\beta) \hat{\sigma}^2(k)$.

Si $\underline{a_1 = 0.8}$, la fréquence du bon ordre est de l'ordre de 90 % pour un nombre d'observations minimal égal à 50 si :

$$\left\{ \begin{array}{l} \alpha \geq 2 \quad \text{et} \quad \beta \geq 0.3 \\ \text{ou} \\ \alpha \geq 4 \quad \text{et} \quad \beta \geq 0.1 \end{array} \right.$$

Pour $N = 25$, la fréquence de l'ordre 1 est de l'ordre de 76 % pour $\alpha \geq 3$ et $\beta \geq 0.3$.

Si $\underline{a_1 = 0.2}$, la fréquence du bon ordre est de l'ordre de 80 % pour $N = 300$ et $\alpha = 3$, $\beta = 0.1$

II - CRITERE DE SCHWARZ : $S_N(k) = N \text{Log } \hat{\sigma}^2(k) + k \text{Log } N$.

Pour $N = 200$, on observe que la fréquence du choix du bon ordre est de l'ordre de 86 % pour les deux corrélations faible et forte.

Critère d'Akaike : $AIC(k) = N \text{Log } \hat{\sigma}^2(k) + 2k$.

Malgré que le critère est non convergent (cf. chap. I), la fréquence du bon ordre se stabilise à 75 % , approximativement, pour le deux corrélations à partir de $N \geq 500$.

III - INFLUENCE DE LA BORNE K SUR LES SELECTIONS D'ORDRES.-

L'influence de la borne K est pratiquement nulle sur la sélection de l'ordre par $S_N(k)$ pour certains choix de α et β (par exemple $\alpha > 3$ et $\beta > 0.3$) et ceci produit même pour de petites tailles d'échantillons (ex. $N = 25$). Pour les critères de Schwarz et Akaike cette influence est non négligeable pour $N = 25$.

IV - ETUDE DE LA STABILITE DE LA SELECTION.-

On a procédé à l'expérience suivante sur un autorégressif d'ordre 2 simulé :

- En observant un échantillon de taille $N = 25$ on a déterminé l'ordre correspondant. A l'étape suivante : on augmente l'échantillon précédent de 5 observations et on détermine l'ordre. Ainsi on continue ce processus, en augmentant à chaque étape l'échantillon précédent de 5 observations et en déterminant l'ordre associé. On observe que la sélection du vrai ordre se stabilise à partir de $N \geq 30$ pour le choix de $S_N(k) = (N + \alpha k N^\beta) \hat{\sigma}^2(k)$, avec $\alpha = 6$, $\beta = 0.1$.

V - ETUDE DE PROCESSUS REELS.-

On a essayé de déterminer les ordres des processus autorégressifs ajustés aux données réelles correspondant aux séries C, E, F du livre de BOX-JENKINS (p. 525 et suivantes).

On a obtenu les résultats suivants :

Séries	N	$S_N(k)$	Nature de la série observée
C	100	$\alpha = 7 , \beta = 0.3$	ARIMA(1,1,0)
E	100	$\alpha = 7 , \beta = 0.3$	AR(3)
F	70	$\alpha = 7 , \beta = 0.3$	AR(2)

VI - REMARQUES ET CONCLUSION. -

L'avantage de la statistique $S_N(k)$ est qu'elle est fonction de deux paramètres α et β , qui permettent en les faisant varier d'augmenter la performance de l'estimateur de l'ordre ce qui n'est pas le cas pour les autres critères de sélection.

La sélection de l'ordre par $S_N(k)$ est "très sensible" au coefficient a_{k_0} de l'autorégressif $AR(k_0)$. Quand ce dernier est très faible, par exemple de l'ordre de 0.1, il faut augmenter considérablement le nombre d'observations ($N \geq 500$) pour que le bon ordre soit très fréquent dans la sélection.

Enfin, le temps d'exécution d'une expérience (i.e. : une ligne de fréquence du tableau) est étroitement lié au nombre d'observations.

Il est de l'ordre de 6 mn pour $N = 500$ et 13 mn pour $N = 1000$.

CONCLUSION.-

Dans ce travail, nous avons introduit un nouvel estimateur de l'ordre d'un autorégressif qui contrairement aux autres estimateurs trouvés précédemment, ([1], [3][18] etc ...), tient compte de l'éventualité que le minimum de $S_N(k)$ soit atteint en plusieurs points. Ce point écarté par ces auteurs nous fait franchir une nouvelle étape dans ce type de problèmes. Hélas des nouvelles difficultés techniques induites par cet estimateur, telles que l'unicité du minimum de la statistique $S_N(k)$ à distance finie, ne nous ont pas permis dans un premier temps d'établir la convergence presque sûre.

- Le problème de la vitesse de convergence reste entier. Nous n'avons pu le résoudre à cause de la complexité de la démonstration du théorème 1. Cependant, cette vitesse dépend étroitement du choix des paramètres α et β (pour K fixé). Dans le dernier chapitre, on a mis en évidence, sur des exemples, des choix pour α et β qui donnent le "bon" k_0 pour un minimum d'observations, en remarquant que ceci améliore les procédures antérieures. Le problème de l'optimalité pour α et β reste ouvert.

- Dans la pratique, dans les modèles $AR(k)$, k dépasse très rarement 5. En conséquence, on se limitera au choix de K inférieur à 10.

- Un autre problème est le suivant : la loi limite des racines (chap. III, § 5). Il semble possible de tabuler ces lois limites dans les cas simples, (degré du polynôme ≤ 3), en vue d'obtenir des intervalles de confiance, ou fabriquer des tests.

- Comme il a été signalé dans le chap. II, § IV, nos résultats peuvent certainement s'étendre aux processus multidimensionnels. Mais l'observation étant indexée par \mathbb{R}^D ; amène à des problèmes très difficiles où l'on bute sur la recherche même d'une expression de la vraisemblance des observations donnant des estimateurs adéquats.

- Les résultats établis pour les processus gaussiens peuvent être étendus sans difficultés aux processus non gaussiens en prenant comme estimateurs de moindres carrés (cf. sect. 2.7. chap. I).

- On a considéré dans le chap. III les processus ARIMA(p,d,q) avec $q = 0$. Cependant pour les ARIMA(p,d,q) où $q \neq 0$, on peut les transformer en "ARMA (p+d,q)".

On estime alors les ordres $p+d$ et q (cf. chap. II, § IV) et la suite du traitement reste inchangée.

Finalement, ces résultats encore incomplets appellent bien d'autres développements.

Simulation

Etude de la Statistique $S_n(k) = (N + \alpha k N^\beta) \hat{\sigma}^2(k)$.

$$\alpha > 0, 0 < \beta < 1, k = 0, 1, 2, \dots, K$$

La fréquence de l'ordre sélectionné par $S_n(k)$ dans 100 réalisations d'un processus autorégressif d'ordre 1. $\{x_t = a_1 x_{t-1} + \varepsilon_t, \varepsilon_t \sim N(0,1), \text{id}\}$.

$K = 7$

$$\alpha = 2 : \beta = 0.1$$

N	K a_1	0	1	2	3	4	5	6	7
		25	0.2	13	7	5	2	7	8
	0.8	0	13	3	7	5	7	21	44
50	0.2	29	23	6	6	6	4	11	15
	0.8	0	47	6	4	9	5	15	14
100	0.2	18	38	10	7	10	4	5	8
	0.8	0	59	13	4	7	5	4	8
200	0.2	6	51	22	8	6	2	1	4
	0.8	0	65	11	6	5	6	1	6
300	0.2	4	68	11	6	4	3	0	4
	0.8	0	75	11	5	4	2	0	3
400	0.2								
	0.8								
500	0.2	0	74	11	2	7	2	2	1
	0.8	0	73	12	4	3	4	2	2
1000	0.2	0	78	13	2	3	1	2	1
	0.8	0	77	10	2	4	2	5	0

Simulation

Etude de la Statistique $S_n(k) = (N + \alpha k N^\beta) \hat{\sigma}^2(k)$

$\alpha > 0$, $0 < \beta < 1$, $k = 0, 1, 2, \dots, K$

La fréquence de l'ordre sélectionné par $S_n(k)$ dans 100 réalisations d'un processus autorégressif d'ordre 1 . $\{x_t = a_1 x_{t-1} + \epsilon_t$, $\epsilon_t \sim N(0,1)$ id}.

K = 7

$\alpha = 2$: $\beta = 0.097$

N	K		0	1	2	3	4	5	6	7
	a_1									
25	0.2		13	7	5	2	7	7	14	45
	0.8		0	12	3	7	4	7	21	46
50	0.2									
	0.8									
100	0.2		18	38	9	7	10	5	5	8
	0.8		0	55	13	4	8	6	5	9
200	0.2									
	0.8		0	64	10	7	6	6	1	6
300	0.2		3	68	12	6	4	3	0	4
	0.8		0	73	11	6	4	3	0	3
400	0.2									
	0.8									
500	0.2		0	71	11	2	5	5	4	3
	0.8		0	70	12	5	3	5	3	2
1000	0.2		0	75	12	3	5	2	3	0
	0.8		0	74	11	3	5	2	5	0

B I B L I O G R A P H I E.

- [1] AKAIKE M. (1969)
Fitting autoregressive models for prédiction.
Ann. Inst. Stat. Math., 21, 243-7 .
- [2] AKAIKE M. (1970)
Statistical predictor indentification.
Ann. Inst. Stat. Math., 22, 203-17.
- [3] AKAIKE M. (1973)
Information theory and an extension of the maximum likelihood principle.
In 2nd International Symposium on Information theory, Eds, B.N. Petroo and F. Csaki pp. 267-81. Budapest : Akademia Kiado.
- [4] ANDERSON TW. (1971)
The statistical analysis of time series.
J. Weley - New-York.
- [5] BARTLETT - DIANANDA (1950)
Extensions of Quenouille's test fort autoregressive schemes.
J. Roy. Statist. Soc. Ser. B, 12, 108-115.
- [6] BEGUIN - GOURIEROUX - MONFORT (1979)
Identification of a mixed autoregressive moving average process :
the corner method.
Publication E.N.S.A.E.
- [7] BILLINGSGLEY P. (1968)
Convergence of probability measures.
J. Wiley, New-York.

- [8] BOSQ D.
1) Cours de D.E.A. : 74-75.
2) Cours de D.E.A. : 80-81.
- [9] BOX G. - JENKINS G. (1976)
Time series analysis - Forecasting and control.
Holden - Day.
- [10] DOOB J.L. (1953)
Stochastic processes.
J. Wiley, New-York.
- [11] DUBIN J. (1960)
The fitting of times series model.
Revue de l'Inst. Internat. de Stat. 28, p. 233-244.
- [12] FULLER W.A. (1976)
Introduction to statistical time series.
J. Wiley, New-York.
- [13] GIRSHICK M.A. (1942)
Note on the distribution of a polynomial with random complex
coefficients.
The Ann. of. Math. Statist. p. 235.
- [14] GORODETSKII V.V. (1977)
On the strong mixing property for linear sequences.
Theory prob. appl. , P. 411.
- [15] GUIKHMAN I. - SKOROKHOD A.
Introduction à la théorie des processus aléatoires,
Edition MIR - 1980.

- [16] HANNAN E.J. and DUNSMUIR, W. (1976)
Vector linear time series models.
Advances in Appl. Probability 8, 2, 339-364.
- [17] HANNAN E.J. and HEYDE C.C. (1972)
On the limit theorems for quadratic function of discrete times series.
The A.M.S. Vol. 42 n° 6, p. 2058-2566.
- [18] HANNAN E.J. (1980)
The estimation of the order of an ARMA process.
The Annals of Statist. Vol. 8, n° 5, 1071-1081.
- [19] HANNAN E.J. and QUINN B.G.
The determination of the order of an autoregressive.
J.R. Stat. Soc. B. (1979), 41, n° 2, pp. 190-195.
- [20] HANNAN E.J. (1970)
Multiple time series.
J. Wiley, New-York.
- [21] HIRONAO K. (1980)
Parameter estimation of autoregressive integrated processes
by least squares.
The Annals of statistics, vol. 8, n° 2, 423-425.
- [22] HASZA B.P., FULLER W.A. (1979)
Estimation for autoregressive processes with unit roots.
The Annals of Statist., vol. 7, n° 5, 1106-1120.
- [23] IBRAGIMOV Z.A. (1962)
Some limit. theorems for stationary processes.
Theory of prob. and its appl. vol. VII, n° 4, p. 349-82.

- [24] KOLMOGOROV A.N., ROSANOV Yu (1960)
On strong mixing condition for stationary gaussian processes.
Theory of prob. and its appl. , Vol. V, n° 2, p. 204.
- [25] KROMER R.G. (1969)
Asymptotic properties of the autoregressive spectral estimator.
Technical report n° 13, December 15, 1969.
- [26] LOEVE M. (1955)
Probability theory.
Van-Nanstrand - New-York.
- [27] MANN M.B. , WALD. A. (1943)
On the statistical treatment of linear stochastic difference equation.
Econometrica, Vol. 11, pp. 173-200.
- [28] MOURID T.
Sur l'identification d'un processus autorégressif.
Article en préparation.
- [29] MONFORT A. (1979)
Méthodes de prévision des séries temporelles.
Cours de l'I.N.S.E..
- [30] PRIESTLEY M.B. (1981)
Spectral analysis and time series.
Academic Press (T1 et T2).
- [31] QUIN B.C. (1980)
Order determination for a multivariate autoregression.
J.R. Statist. Soc. B., 42, n° 2, pp. 182.

- [32] QUENOUILLE (1947)
A large sample test for the goodness of fit of autoregressive schemes.
J. Ros., Statist. Soc., Ser. B, 12 p. 123-129.
- [33] RAMSEY F.L. (1974)
Characterization of the partial autocorrelation function.
The annals of Stats. vol. 2, n° 6, 1296-1301.
- [34] RISSANEN J. and CAINES P.E. (1979)
The strong consistency of maximum likelihood estimators for ARMA processes.
The annals of Stat., vol. 7, n° 2, 297-315.
- [35] ROSENBLATT M. (1956)
A central limit theorem and a strong mixing condition.
Proceedings, vol. 42, p. 43-47.
- [36] ROSENBLATT M. (1974)
Random processes.
Springer-Verlag New-York, Berlin.
- [37] SCHWARZ G.
Estimating the dimension of a model.
Ann. Stat. 6, 461-464.
- [38] SHIBATA R. (1976)
Selection of the order of an autoregressive model by Akaike's
Information Criterion.
Brometika , 63, 1, pp. 117-26.
- [39] SHIBATA R. (1980)
Asymptotically efficient selection of the order of the model
for estimating parameters of a linear process.
The Annals of Stat., vol. 8, n° 1, 147-164.

[40] SPTIZER F. (1956)

A combinatorial lemma and its applications to probability theory.
Trans. Am. Math. Soc. 82, 323-39.

[41] WAYNE A. FULLER - HASZA - GOEBEL (1981)

Estimation of the parameters of stochastic difference equations.
The Annals of Stat. vol. 9, n° 3, 531-43.



RÉSUMÉ

Ce travail est essentiellement consacré à l'estimation des paramètres d'un processus autorégressif à partir de l'observation d'un morceau de trajectoire du processus.

A partir du statistique proposée par SHIBATA, on construit un estimateur de l'ordre d'un $AR(p)$. Une loi limite est fournie, on donne un théorème de convergence, et on étudie le risque quadratique asymptotique associé.

On montre que cette procédure peut s'étendre aux AR multidimensionnels, ainsi qu'aux $ARMA(p,q)$.

Comme application, on fabrique un procédé d'identification du degré de différentiation d'un processus $ARIMA(p,d,q)$.

Enfin, des exemples, et des simulations illustrent cette étude.

MOTS CLÉS

- PROCESSUS AUTOREGRESSIF, ARIMA,
- IDENTIFICATION,
- SIMULATION - MODELE.