

SUR LES MOMENTS DES AUTOCOVARIANCES ÉCHANTILLONNALES
D'UN PROCESSUS ARIMA¹
Roch Roy et Pierre Lefrançois

0. Résumé

Dans cette note, nous obtenons les deux premiers moments asymptotiques des autocovariances échantillonnelles d'un processus ARIMA $(p,1,q)$. Ils sont d'ordre N et N^2 respectivement, où N désigne la longueur de la série, et sont caractérisés par les paramètres du modèle à travers la densité spectrale à la fréquence 0 de la première différence de la série originale.

(Mots clés: Analyse de séries chronologiques; non-stationnarité; autocovariances échantillonnelles.)

1. Introduction

Dans la modélisation de séries chronologiques, les autocovariances et les autocorrélations échantillonnelles jouent un rôle fondamental. Leur comportement, pour une série stationnaire est assez bien connu alors que pour une série non-stationnaire, il l'est beaucoup moins. La plupart des résultats existants ont été développés pour un processus de moyenne mobile intégrée d'ordre $(1,1)$

$$Z_t - Z_{t-1} = a_t - \theta a_{t-1}, \quad t = 1, 2, \dots \quad (1.1)$$

où θ désigne le paramètre de moyenne mobile et $\{a_t\}$ est un bruit blanc.

¹ Recherche subventionnée en partie par le Conseil de recherches en sciences naturelles et en génie du Canada, subvention A-4145.

Wichern [6] ainsi que Box et Jenkins [1] ont étudié les propriétés des autocorrélations r_k d'une série non stationnaire générée du modèle (1.1). En particulier, Wichern a déduit une formule exacte pour l'espérance mathématique des autocovariances échantillonnables c_k et a étudié par simulation le comportement de $E[c_k]/E[c_0]$ comme approximation de $E[r_k]$. Pour le même modèle, Roy [5] a obtenu les deux premiers moments asymptotiques des autocovariances échantillonnables; ils sont d'ordre N et N^2 respectivement, où N désigne la longueur de la série. Les implications de ces résultats pour l'identification du degré de différenciation approprié d'une série non stationnaire furent aussi discutées.

Dans son livre, Fuller [2] section 8.5, présente quelques résultats exacts sur les moments de statistiques reliées aux autocovariances échantillonnables d'une promenade aléatoire, c'est-à-dire dans le cas où $\theta = 0$. En particulier, il donne l'espérance mathématique et la variance de $\sum_{t=1}^N (Z_t - \bar{Z})^2$.

Rao [4] a obtenu la distribution asymptotique de l'estimateur des moindres carrés $\hat{\phi}_N$ de ϕ pour une série générée du modèle

$$Z_t - \phi Z_{t-1} = a_t$$

dans le cas $\phi = 1$. L'estimateur $\hat{\phi}_N$ est l'autocorrélation échantillonnale de délai 1 et contrairement au cas stationnaire, sa distribution asymptotique n'est pas gaussienne.

L'objectif de cette note est de généraliser les résultats de Roy [5] sur les autocovariances échantillonnables au cas général d'un processus ARIMA $(p,1,q)$. Dans la section 2, nous introduisons la notation et présentons quelques résultats préliminaires sur l'approximation d'un processus ARIMA par une promenade aléatoire. Les moments asymptotiques des autocovariances sont obtenus à la section 3. Ils sont caractérisés par les paramètres du modèle à travers la densité spectrale à la fréquence 0 de la première différence de la série originale.

2. Preliminaires

Soit $Z_1, \dots, Z_N$ une série chronologique de longueur N générée d'un processus ARIMA $(p, 1, q)$, c'est-à-dire un processus défini par l'équation

$$\phi(B)\nabla Z_t = \theta(B)a_t, \quad t = 1, 2, \dots \quad (2.1)$$

où, selon la notation de Box et Jenkins (1976), B désigne l'opérateur décalage à gauche, $\nabla = 1 - B$ est l'opérateur différence régulière, $\phi(B) = 1 - \phi_1 B - \dots - \phi_p B^p$ est l'opérateur autorégressif dont toutes les racines sont supposées à l'extérieur du cercle unité, $\theta(B) = 1 - \theta_1 B - \dots - \theta_q B^q$ est l'opérateur de moyenne mobile et $\{a_t\}$ représente un bruit blanc, c'est-à-dire une suite de variables aléatoires indépendantes et identiquement distribuées de moyenne 0 et de même variance σ^2 .

L'autocovariance échantillonnale de délai k est définie

$$c_Z^{(N)}(k) = N^{-1} \sum_{t=1}^{N-k} (Z_t - \bar{Z})(Z_{t+k} - \bar{Z}), \quad k = 0, 1, \dots, K,$$

où $\bar{Z} = (Z_1 + \dots + Z_N)/N$ est la moyenne échantillonnale et $0 < K \leq N - 1$ est fixé.

Soit $w_t = \nabla Z_t$. Etant donné une valeur initiale Z_0 , nous pouvons écrire

$$Z_t = Z_0 + w_1 + \dots + w_t, \quad t = 1, 2, \dots \quad (2.2)$$

Les résultats de la section 3 ont été obtenus initialement sous des conditions un peu moins générales par Lefrançois [3] dans son mémoire de maîtrise à l'Université de Montréal. Il représente alors $c_Z^{(N)}(k)$ comme une forme quadratique du vecteur $(w_1, w_2, \dots, w_N)'$ et tire avantage de résultats généraux sur les moments d'une forme quadratique. Quoique son approche est plus élémentaire que celle adoptée ici, ses démonstrations sont beaucoup plus longues et entraînent de nombreuses manipulations algébriques. Dans cet article, nous utilisons une approximation de Z_t qui est implicite dans la démonstration du Théorème 8.5.1 de Fuller [2]. Cette approximation nous permet de présenter une démonstration beaucoup plus concise des mêmes ré-

sultats et nous indique en même temps quelle est la partie de Z_t qui contribue aux résultats asymptotiques.

Le processus w_t est un processus autorégressif moyenne mobile qui satisfait l'équation

$$\phi(B)w_t = \theta(B)a_t. \quad (2.3)$$

Puisque l'on suppose que les racines de $\phi(z) = 1 - \phi_1 z - \dots - \phi_p z^p$ sont à l'extérieur du cercle unité, il existe une solution unique de (2.3) en termes du présent et du passé qui est stationnaire. Cette solution est de la forme

$$w_t = \sum_{j=0}^{\infty} \psi_j a_{t-j} \quad (2.4)$$

et les coefficients ψ_j satisfont l'équation aux différences

$$\phi(B)\psi_j = 0, \quad j > \max\{p-1, q\}$$

(voir par exemple Box et Jenkins [1], p. 96).

D'après la théorie des équations aux différences, si l'on décompose $\phi(B)$ en monômes,

$$\phi(B) = (1 - G_1 B)^{e_1} \dots (1 - G_s B)^{e_s},$$

où $|G_i| < 1$, $i = 1, 2, \dots, s$ et $\sum_{i=1}^s e_i = p$, on peut représenter ψ_j comme une combinaison linéaire de termes de la forme $j^m G_h^j$ avec $0 \leq m \leq e_h - 1$. Puisque chacune des séries $j^m |G_h|^j$ converge si et seulement si $|G_h| < 1$, il s'ensuit que

$$\sum_{j=0}^{\infty} |\psi_j| < \infty.$$

En combinant (2.1) et (2.4), on obtient

$$Z_t = Z_0 + g_0 X_t - D_t + \sum_{j=0}^{\infty} g_{j+1} a_{t-j} \quad (2.5)$$

où

$$x_t = \sum_{j=1}^t a_j, \quad t = 1, 2, \dots,$$

$$g_j = \sum_{i=j}^{\infty} \psi_i, \quad j = 0, 1, \dots,$$

et

$$D_t = \sum_{j=0}^{\infty} g_{j+1} a_{t-j}, \quad t = 0, \pm 1, \dots$$

Le processus D_t est défini en moyenne quadratique et est stationnaire au sens large puisque $\sum_{j=0}^{\infty} |g_j| < \infty$.

En effet, notons d'abord que

$$\sum_{j=0}^{\infty} |g_j| \leq \sum_{j=0}^{\infty} \sum_{k=j}^{\infty} |\psi_k| = \sum_{j=0}^{\infty} (j+1) |\psi_j|.$$

Utilisant la notation "grand o", nous pouvons écrire

$$\psi_j = o(j^{p-1} M^j)$$

où $M = \max\{|G_1|, |G_2|, \dots, |G_s|\} < 1$, et puisque

$$\sum_{j=0}^{\infty} j |\psi_j| \leq \text{cte} \sum_{j=0}^{\infty} j^p M^j < \infty,$$

il en découle que

$$\sum_{j=0}^{\infty} |g_j| < \infty.$$

Dans l'étude des autocovariances échantillonales, nous pouvons négliger, sans perte de généralité, les termes z_0 et $\sum_{j=0}^{\infty} g_{j+1} a_{-j}$ dans la représentation (2.5) puisque les autocovariances sont invariantes par rapport aux translations.

Ainsi, dans la suite, nous utiliserons l'approximation suivante de Z_t :

$$Z_t = g_0 x_t - D_t, \quad t = 1, 2, \dots \quad (2.6)$$

De (2.6), il découle immédiatement que

$$c_Z^{(N)}(k) = g_0^2 c_X^{(N)}(k) - g_0 R^{(N)}(k) + c_D^{(N)}(k) , \quad (2.7)$$

où

$$R^{(N)}(k) = \frac{1}{N} \sum_{t=1}^{N-k} (X_t - \bar{X})(D_{t+k} - \bar{D}) + (D_t - \bar{D})(X_{t+k} - \bar{X}) , \quad k = 0, 1, \dots, K$$

et $c_X^{(N)}(k)$, $c_D^{(N)}(k)$ désignent l'autocovariance échantillonnale de délai k de X_t et D_t respectivement.

Le processus X_t étant une promenade aléatoire, les moments asymptotiques des $c_X^{(N)}(k)$ peuvent être déduits de Roy [5] en posant $\theta = 0$. En fait, nous avons

$$\lim_{N \rightarrow \infty} E[c_X^{(N)}(k)/N] = \sigma^2/6 ,$$

$$\lim_{N \rightarrow \infty} \text{Cov}(c_X^{(N)}(k)/N, c_X^{(N)}(h)/N) = \sigma^4/45 , \quad k, h = 0, 1, \dots, K .$$

De plus, les moments asymptotiques des $c_D^{(N)}(k)$ sont connus puisque D_t est stationnaire.

3. Moments asymptotiques

En tirant avantage des résultats décrits dans la section précédente, nous pouvons démontrer le théorème suivant.

THÉORÈME. Soit $Z_1, \dots, Z_N$ une série chronologique de longueur N générée d'un processus ARIMA $(p, 1, q)$ dont les racines de l'opérateur autorégressif sont toutes à l'extérieur du cercle unité. Soit $w_t = \nabla Z_t$ et posons $E[w_t] = 0$, $\text{Cov}(w_t, w_{t+k}) = \gamma_k$, $t, k = 0, \pm 1, \dots$. Alors

$$\lim_{N \rightarrow \infty} E[c_Z^{(N)}(k)/N] = (1/6)S , \quad (3.1)$$

$$\lim_{N \rightarrow \infty} \text{Cov}(c_Z^{(N)}(k)/N, c_Z^{(N)}(h)/N) = (1/45)S^2, \quad (3.2)$$

$$k, h = 0, 1, \dots, K \quad \text{où} \quad S = \sum_{j=-\infty}^{\infty} \gamma_j.$$

DEMONSTRATION. De (2.7), nous avons

$$\begin{aligned} \lim_{N \rightarrow \infty} E[c_Z^{(N)}(k)/N] &= g_0^2 \lim_{N \rightarrow \infty} E[c_X^{(N)}(k)/N] - g_0 \lim_{N \rightarrow \infty} E[R^{(N)}(k)/N] \\ &\quad + \lim_{N \rightarrow \infty} E[c_D^{(N)}(k)/N]. \end{aligned}$$

Le troisième terme est nul puisque $c_D^{(N)}(k)$ est un estimateur convergent de l'auto-covariance théorique de délai k de D_t . De plus, en appliquant l'inégalité de Cauchy-Schwarz une première fois, on trouve

$$|R^{(N)}(k)/N| \leq \frac{2}{\sqrt{N}} (c_X^{(N)}(0)/N)^{\frac{1}{2}} (c_D^{(N)}(0))^{\frac{1}{2}} \quad (3.3)$$

et en l'appliquant une seconde fois, il découle que

$$E[|R^{(N)}(k)/N|] \leq \frac{2}{\sqrt{N}} (E[c_X^{(N)}(0)/N] E[c_D^{(N)}(0)])^{\frac{1}{2}}.$$

Ainsi, nous avons

$$\lim_{N \rightarrow \infty} E[R^{(N)}(k)/N] = 0$$

et (3.1) est démontré, puisque $\sigma^2 g_0^2 = S$.

Afin de démontrer (3.2), notons d'abord qu'en utilisant (2.7), nous pouvons écrire

$$c_Z^{(N)}(k) c_Z^{(N)}(h) = g_0^4 c_X^{(N)}(k) c_X^{(N)}(h) + S^{(N)}(h, k)$$

où, dans $S^{(N)}(h, k)$, les termes d'ordre supérieur sont ceux impliquant $R^{(N)}(k)$ et/ou $c_X^{(N)}(k)$. Il est facile de vérifier que

$$\lim_{N \rightarrow \infty} \frac{1}{N^2} E[S^{(N)}(h, k)] = 0$$

en considérant chacun des termes séparément. Par exemple, nous obtenons de (3.3)

$$\frac{1}{N^2} |R^{(N)}(k)R^{(N)}(h)| \leq \frac{4}{N} (c_X^{(N)}(0)/N)c_D^{(N)}(0)$$

d'où il s'ensuit que

$$\lim_{N \rightarrow \infty} \frac{1}{N^2} E[|R^{(N)}(k)R^{(N)}(h)|] \leq \lim_{N \rightarrow \infty} \frac{4}{N} (E[(c_X^{(N)}(0)/N)^2]E[(c_D^{(N)}(0))^2])^{\frac{1}{2}} = 0. \quad (3.4)$$

Tous les autres termes peuvent être traités de façon similaire en utilisant (3.4) si nécessaire. A titre illustratif, nous avons

$$\lim_{N \rightarrow \infty} \frac{1}{N^2} E[|R^{(N)}(k)c_X^{(N)}(h)|] \leq \lim_{N \rightarrow \infty} (E[(R^{(N)}(k)/N)^2]E[(c_X^{(N)}(h)/N)^2])^{\frac{1}{2}} = 0.$$

Finalement, nous obtenons

$$\lim_{N \rightarrow \infty} \frac{1}{N^2} E[c_Z^{(N)}(k)c_Z^{(N)}(h)] = g_0 \lim_{N \rightarrow \infty} \frac{1}{N^2} E[c_X^{(N)}(k)c_X^{(N)}(h)] = \frac{1}{20} g_0^4,$$

et (3.2) s'ensuit. La démonstration est donc complétée.

Les équations (3.1) et (3.2) nous indiquent que les moments asymptotiques de $c_Z^{(N)}(k)/N$ sont indépendants du délai k et sont complètement caractérisés par

$$S = \sum_{j=-\infty}^{\infty} \gamma_j = 2\pi f(0)$$

où $f(\omega)$ désigne la densité spectrale à la fréquence ω de ∇Z_t . De plus, il découle de (3.2) que la corrélation asymptotique entre $c_Z^{(N)}(k)/N$ et $c_Z^{(N)}(h)/N$ est un pour tout h et k .

BIBLIOGRAPHIE

- [1] BOX, G.E.P. and JENKINS, G.M., *Time Series Analysis: Forecasting and Control*, Holden-Day, San Francisco, édition révisée (1976).
- [2] FULLER, W.A., *Introduction to Statistical Time Series*, John Wiley & Sons, New York (1976).

- [3] LEFRANCOIS, P., *Comportement asymptotique de l'autocovariance échantillonnale d'un processus ARIMA (p,1,q)*, Publication no 309, Département d'informatique et de recherche opérationnelle, Université de Montréal, Montréal, Québec (1978).
- [4] RAO, M.M., *Asymptotic distribution of an estimator of the boundary parameter of an unstable process*, Ann. Statist., vol. 6 (1978), pp. 185-190.
- [5] ROY, R., *On the asymptotic behaviour of the sample autocovariance function for an integrated moving average process*, Biometrika, vol. 64 (1977), pp. 419-421.
- [6] WICHERN, D.W., *The behaviour of the sample autocorrelation function for an integrated moving average process*, Biometrika, vol. 60 (1973), pp. 235-239.

Département d'informatique et de
recherche opérationnelle
Université de Montréal
C.P. 6128, Succ. "A"
Montréal, Québec
H3C 3J7

Manuscrit reçu le 5 août 1980.
Révisé le 31 décembre 1980.