C. R. Acad. Sci. Paris, Ser. I 334 (2002) 909–912

Statistique/*Statistics*
(Probabilités/*Probability Theory*)

Statistiques d'une saisonnalité perturbée par un processus a représentation autorégressive

Tahar Mourid

Département de mathématiques, Faculté des sciences, Université Abou Bekr Belkaid, Tlamcen 13000, Algérie

Reçu le 1er octobre 2001 ; accepté après révision le 4 mars 2002

Note présentée par Paul Deheuvels.

Résumé Nous considérons des théorèmes limites d'un estimateur fonctionnel d'une saisonnalité $a(\cdot)$ perturbée par un processus réel à temps continu à représentation autorégressive Banach. Nous construisons des régions de confiance pour $a(\cdot)$ à partir d'une loi du logarithme itéré compacte. Un estimateur par projection de $a(\cdot)$ et un estimateur de la dimension sont étudiés dans le cas où la saisonnalité $a(\cdot)$ appartient à un sous espace de dimension finie. ***Pour citer cet article : T. Mourid, C. R. Acad. Sci. Paris, Ser. I 334 (2002) 909–912.*** © 2002 Académie des sciences/Éditions scientifiques et médicales Elsevier SAS

Statistics of seasonality perturbed by time continuous processes with autoregressive representation

Abstract We consider limit theorems for an estimator of a seasonality when it is perturbed by a time continuous process admitting a Banach autoregressive representation. From the compact iterated logarithm law we derive confidence regions for $a(\cdot)$ in the Banach space of continuous functions. When $a(\cdot)$ belongs to a finite dimensional subspace, we study the estimation of $a(\cdot)$ by projection and we estimate the dimension when it is unknown. ***To cite this article: T. Mourid, C. R. Acad. Sci. Paris, Ser. I 334 (2002) 909–912.*** © 2002 Académie des sciences/Éditions scientifiques et médicales Elsevier SAS

1. Introduction

Soit $(\Omega, \mathcal{A}, P)$ un espace probabilisé complet et $(B, \|\cdot\|)$ un espace de Banach réel séparable muni de sa tribu Borelienne $\mathcal{B}$. On note $\mathcal{L}(B)$ l'algèbre de Banach des opérateurs linéaires bornés définis sur B et à valeurs dans B, munie de la norme usuelle $\|\cdot\|_L$ des opérateurs bornés [4]. Un bruit blanc banachique est une suite de variables aléatoires $(\varepsilon_n, n \in \mathbb{Z})$ indépendantes identiquement distribuées définies sur (Ω, A, P) à valeurs dans B vérifiant pour tout $n \in \mathbb{Z}$: $E\varepsilon_n = 0,\ 0 < E\|\varepsilon_n\|^2 = \sigma_\varepsilon^2 < \infty$. Une suite $(X_n,\ n \in \mathbb{Z})$ définie sur (Ω, A, P) et à valeurs dans B est un processus autorégressif banachique d'ordre p (ARB(p)), $p \in \mathbb{N}^*$, s'il existe des opérateurs $\rho_1, \rho_2, \ldots, \rho_p$ dans $\mathcal{L}(B)$, un bruit blanc banachique (ε_n) tels que :

$$X_n = \rho_1 X_{n-1} + \rho_2 X_{n-2} + \cdots + \rho_p X_{n-p} + \varepsilon_n \quad \text{p.s.}, \tag{1}$$

où l'opérateur ρ_p est différent de l'opérateur nul. Dans [8] nous donnons une condition sur les opérateurs ρ_i pour que (1) admette une solution strictement stationnaire dans B. Un processus $\xi = (\xi(t),\ t \in \mathbb{R})$ à temps continu admet une représentation autorégressive d'ordre p dans l'espace de Banach B, s'il existe des opérateurs linéaires bornés $\rho_1, \rho_2, \ldots, \rho_p$ sur B, un bruit blanc $(\varepsilon_n,\ n \in \mathbb{Z})$ à valeurs dans B et un

S1631-073X(02)02352-X/FLA

nombre δ strictement positif tels que la suite des variables aléatoires à valeurs dans B définies par :

$$X_n(t) := \xi(n\delta + t), \quad t \in [0, \delta], \tag{2}$$

vérifient la relation (1). Dans [3], Chapitre 6.2 et [9] on trouve des exemples de processus à temps continu admettant une représentation du type (1) et nous citons le processus d'Ornstein–Ulhenbeck. Dans de nombreuses applications, le processus $\xi = (\xi_t,\ t \in \mathbb{R})$ est p.s. à trajectoires continues et l'espace de Banach B naturel est l'espace des fonctions définies et continues sur $[0, \delta]$ noté $\mathrm{C} = \mathrm{C}_{[0,\delta]}$ ou encore les espaces usuels $\mathrm{L}^p_{[0,\delta]}$.

Considérons un processus réel $(\eta(t),\ t \in \mathbb{R})$ à temps continu admettant la représentation suivante :

$$\eta(t) = a(t) + \xi(t), \quad t \in \mathbb{R}, \tag{3}$$

où $a(\cdot)$ est une fonction réelle déterministe continue périodique de période δ et $(\xi(t),\ t \in \mathbb{R})$ est un processus réel centré presque sûrement à trajectoires continues satisfaisant à la représentation autorégressive (1) dans l'espace $\mathcal{C}_{[0,\delta]}$ par l'intermédiaire de la relation (2). Le modèle (3) correspond à la situation où une saisonnalité périodique $a(\cdot)$ est perturbée par un processus aléatoire possédant une structure particulière qui est est supposée ici de la forme (1). Remarquons que le processus $(\eta(t),\ t \in \mathbb{R})$ n'est pas en général stationnaire puisque $E(\eta(t)) = a(t)$. Nous supposons que l'on dispose d'une trajectoire du processus $\eta(t)$ sur un intervalle $[0, T]$, où T est supposé un multiple de la période δ : $T = n\delta$. L'observation de $\eta(t)$ sur n intervalles successifs de longueur δ engendrent n v.a. $Y_1, \ldots, Y_n$ à valeurs dans $\mathcal{C}_{[0,\delta]}$ définies par la relation $Y_i(t) = \eta(ih + t),\ 0 \leqslant t \leqslant \delta$, pour $i = 0, 1, \ldots, n-1$. Pour des problèmes d'estimation de la saisonalité $a(\cdot)$ dans le cadre d'observations i.i.d. on peut citer par exemple [2,5,10,11]. Cependant dans cette étude les v.a. Y_i considérées sont des v.a. dépendantes et du type β-mélangeant [1]. Dans la suite l'espace de Banach séparable B sera l'espace $\mathcal{C}_{[0,\delta]}$ muni de la norme infinie notée $\|\cdot\|_\infty$ et la représentation ARB(p) sera notée ARC(p).

2. Résultats

Des théorèmes limites pour la classe des processus vérifiant (1) sont dans [3], Chapitre 3 pour $p = 1$ et dans [8] pour le cas général $p \geqslant 1$. Nous en déduisons des théorèmes limites pour le processus $(\eta(t),\ t \in \mathbb{R})$ pour le cas $p \geqslant 1$. Notons par $\overline{Y}_n$ la moyenne empirique : $\overline{Y}_n = (1/n)\sum_{i=1}^n Y_i$. Nous avons le résultat suivant.

PROPOSITION 1. –

(i) *Si le processus* $(\xi(t),\ t \in \mathbb{R})$ *admet une représentation* ARC(p) *centrée du type* (1) *alors le processus* $(\eta(t),\ t \in \mathbb{R})$ *admet la même représentation* ARC(p) *du type* (1) *d'espérance* a.

(ii) $\overline{Y}_n$ *est un estimateur sans biais de a et* $\overline{Y}_n \underset{n\to\infty}{\longrightarrow} a$ *p.s.*

(iii) *S'il existe une v.a.r. positive* M *de carrée intégrable telle que*

$$\bigl|\varepsilon_0(\omega, s) - \varepsilon_0(\omega, t)\bigr| \leqslant M(\omega)|s - t|, \quad \omega \in \Omega,\ (s,t) \in [0,1]^2,$$

alors

$$\sqrt{n}\bigl(\overline{Y}_n - a\bigr) \underset{n\to\infty}{\Longrightarrow} \mathcal{N}\bigl(0, (I - \rho_1 - \rho_2 - \cdots - \rho_p)^{-1} R_{\varepsilon_0} (I - \rho_1 - \rho_2 - \cdots - \rho_p)^{*-1}\bigr),$$

où R_{ε_0} *est l'opérateur de covariance de* ε_0.

(iv) *Sous l'hypothèse de* (iii) *et si* $E\|\varepsilon_0\|_\infty^{2+\alpha} < \infty$, *pour* $\alpha > 0$, *alors* $(Y_n,\ n \in \mathbb{Z}))$ *vérifie la loi du logarithme itéré compacte dans* $\mathcal{C}_{[0,\delta]}$:

$$\lim_{n\to\infty} d\left(\frac{\overline{Y}_n - a}{\sqrt{(2\log\log n)/n}}, (I - \rho_1 - \cdots - \rho_p)^{-1}K\right) = 0 \quad p.s.,$$

$$C\left\{\left(\frac{\overline{Y}_n - a}{\sqrt{(2\log\log n)/n}}\right)_{n\geqslant 3}\right\} = (I - \rho_1 - \cdots - \rho_p)^{-1}K \quad p.s.,$$

où K est la boule unité de l'espace autoreproduisant associé au noyau de covariance de ε_0, $d(x,E) := \inf_{y\in E}\|x-y\|_\infty$ si $x\in\mathcal{C}_{[0,\delta]}$, $E\subset\mathcal{C}_{[0,\delta]}$, et $C\{(x_n)\}$ désigne l'ensemble des points d'adhérence dans $\mathcal{C}_{[0,\delta]}$ de la suite (x_n). En particulier nous avons

$$\limsup_{n\to\infty}\left\|\frac{\overline{X}_n-a}{\sqrt{(2\log\log n)/n}}\right\|_\infty=\sup_{x\in K}\left\|(I-\rho_1-\cdots-\rho_p)^{-1}x\right\|_\infty\quad p.s.$$

Remarque 1. – La condition (iii) est une hypothèse classique dans la théorie des théorèmes centrale limite et loi du logarithme itéré dans des espaces généraux [6,7]. Cependant dans les applications nous pouvons s'assurer de cette régularitér sur les résidus observés : $\hat\varepsilon_i=X_i-\rho_1X_{i-1}-\rho_2X_{i-2}-\cdots-\rho_pX_{i-p}$, $i=1,\dots,n$.

Dans cette partie nous considérons l'espace de Banach $(\mathcal{C}_{[0,\delta]},\ \|\cdot\|_\infty)$ muni d'une base de Schauder notée $(e_i,\ i\in\mathbb{N})$. Désignons par $B(x_0,r)$ la boule de centre x_0 et de rayon $r>0$ dans $(\mathcal{C}_{[0,\delta]},\ \|\cdot\|_\infty)$ et par P_k le projecteur sur le sous espace E_k engendré par les fonctions $(e_1,\dots,e_k)$, $k\in\mathbb{N}^*$. Nous ferons une étude où nous supposons que la saisonnalité $a(\cdot)$ appartient à un sous espace noté E_{k_0} de $\mathcal{C}_{[0,\delta]}$ de dimension finie k_0 connu. Nous introduisons donc la condition suivante :

H_1 : $a=\sum_{i=1}^{k_0}f_i(a)e_i$, où f_i, $i=1,\dots,k_0$, sont des fonctionnelles sur $\mathrm{C}_{[0,\delta]}$ déterminées de façon unique telles que $f_{k_0}(a)\neq0$.

Le modèle (3) englobe alors certains modèles de régression classiques [10]. En effet sous la condition H_1 le modèle (3) s'interprète comme une régression sur les fonctions $(e_1,\dots,e_{k_0})$ avec une perturbation ξ du type ARC(p) : $\eta(t)=\sum_{i=1}^{k_0}f_i(a)e_i(t)+\xi(t)$.

Le résultat suivant permet d'obtenir des régions de confiances pour la saisonnalité $a(\cdot)$ dans l'espace $\mathcal{C}_{[0,\delta]}$. Nous supposons que les paramètres $\rho_1,\dots,\rho_p$ et $\|R_{\varepsilon_0}\|_L$ du modèle (1) sont connus.

PROPOSITION 2. – *Supposons que la condition* (iv) *de la Proposition* 1 *est vérifiée. Alors il existe* $\Omega_0\subset\Omega$, $P(\Omega_0)=1$ *tel que* $\forall\omega\in\Omega_0$:

(i) $\forall\beta>0,\ \exists N(\beta,\omega),\ \forall n\geqslant N(\beta,\omega)\Rightarrow$

$$a\in B\left(\overline{Y}_n(\omega),\ \sqrt{\frac{2\log\log n}{n}}\left(\beta+\left\|(I-\rho_1-\cdots-\rho_p)^{-1}\right\|\cdot\|R_{\varepsilon_0}\|_L^{1/2}\right)\right).$$

(ii) *Si la saisonnalité a vérifie la condition* H_1*, alors* $P_{k_0}(\overline{Y}_n)$ *est un estimateur sans biais convergeant presque sûrement vers a et*

$$\limsup_{n\to\infty}\sqrt{\frac{n}{2\log\log n}}\left\|\overline{Y}_n(\omega)-P_{k_0}\big(\overline{Y}_n(\omega)\big)\right\|_\infty\leqslant(1+C_1)\left\|(I-\rho_1-\cdots-\rho_p)^{-1}\right\|\|R_{\varepsilon_0}\|_L^{1/2}\quad p.s.,$$

où $C_1=\sup_{k\in\mathbb{N}}\|P_k\|$ *est appelée constante basique de l'espace* $\mathcal{C}_{[0,\delta]}$.

Remarque 2. – (1) Dans le résultat (i) si les opérateurs $\rho_1,\dots,\rho_p$ et $\sigma_{\varepsilon_0}^2$ sont inconnus, en utilisant des estimateurs convergeants : $\hat\rho_{i,n}$ et $\hat\sigma_{\varepsilon_0,n}^2=(1/n)\sum_{i=1}^n\|X_i-\hat\rho_1X_{i-1}-\cdots-\hat\rho_pX_{i-p}\|^2$ et à l'aide de la relation $\|R_{\varepsilon_0}\|_L\leqslant E\|\varepsilon_0\|^2=\sigma_{\varepsilon_0}^2$ nous pourrons considérer les régions de confiance empiriques suivantes : $\forall\beta>0$, $\forall\omega\in\Omega_0$ et pour n assez grand :

$$a\in B\left(\overline{Y}_n(\omega),\ \sqrt{\frac{2\log\log n}{n}}\left(\beta+\left\|(I-\hat\rho_1-\cdots-\hat\rho_p)^{-1}\right\|\hat\sigma_{\varepsilon_0}\right)\right)$$

(une étude est en cours).

(2) Par le résultat (ii) on peut utiliser l'estimateur $P_{k_0}(\overline{Y}_n)$ plutôt que $\overline{Y}_n$ pour estimer la saisonnalité $a(\cdot)$. Cependant l'estimateur $P_{k_0}(\overline{Y}_n)$ nécessite en plus le calcul de ses coefficients dans la base $(e_1,\dots,e_{k_0})$.

Le résultat suivant donne une loi limite pour l'estimateur de projection.

PROPOSITION 3. – *Sous la condition* H_1 *et la condition* (iii) *Proposition* 1 *nous avons* :

$$\sqrt{n}\big(P_{k_0}(\overline{Y}_n) - a\big) \underset{n\to\infty}{\Longrightarrow} \mathcal{N},$$

où $\mathcal{N}$ *est une v.a. gaussienne centrée dans* $\mathbb{R}^{k_0}$ *de matrice de covariance*

$$P_{k_0}(I - \rho_1 - \cdots - \rho_p)^{-1} R_{\varepsilon_0} (I - \rho_1 - \cdots - \rho_p)^{*-1} P_{k_0}.$$

Nous supposons maintenant que la dimension k_0 du sous espace E_{k_0} est inconnue et que $0 \leqslant k_0 \leqslant K_0$, où K_0 est un entier positif donné. Le résultat (ii) de la Proposition 2 permet de définir un estimateur de la dimension k_0 en supposant que les paramètres $\rho_1, \ldots, \rho_p$ et la norme $\|R_{\varepsilon_0}\|$ sont connus. Posons

$$\hat{k}_n := \min\Big\{k \in \{1, \ldots, K_0\} \mid \\ \big\|\overline{Y}_n - P_k(\overline{Y}_n)\big\|_\infty \leqslant (1 + C_1)\|R_{\varepsilon_0}\|^{1/2}\big\|(I - \rho_1 - \cdots - \rho_p)^{-1}\big\|\sqrt{\frac{2\log\log n}{n}}\Big\}.$$

Remarque 3. – (1) L'existence de $\hat{k}_n$ découle du résultat (ii) Proposition 2 : pour n assez grand $\hat{k}_n$ existe et vérifie $\hat{k}_n \leqslant k_0$. Pour n « petit » nous pouvons prendre $\hat{k}_n = 1$ quand il n'est pas défini.

(2) Quand les paramètres $\rho_1, \ldots, \rho_p$ et $\|R_{\varepsilon_0}\|$ sont inconnus, nous pourrons considérer des estimateurs convergeants des paramètres et en définissant un estimateur $\hat{\hat{k}}_n$ de facon similaire.

Le résultat suivant donne la convergence presque sûre de $\hat{k}_n$:

PROPOSITION 4. – *Sous la condition* H_1 *et la condition* (iv) *Proposition* 1 *nous avons* :

$$\lim_{n\to\infty} \hat{k}_n = k_0 \quad p.s.$$

Le résultat précédent permet de définir dans ce cas un estimateur de la saisonnalité $a(\cdot)$.

COROLLAIRE 5. –

$$P_{\hat{k}_n}(\overline{Y}_n) \underset{n\to\infty}{\longrightarrow} a \quad \textit{en probabilité.}$$

Nous remercions les deux rapporteurs pour leurs suggestions constructives et les améliorations proposées.

Références bibliographiques

[1] A. Allam, T. Mourid, Proprietés de mélange des processus autoregressifs banachiques, C. R. Acad. Sci. Paris, Série I 332 (2001) 363–368.
[2] A. Antoniadis, Parametric estimation for the mean of a Gaussien process by the method of Sieves, J. Multivariate Anal. 26 (1988) 1–15.
[3] D. Bosq, Linear Processes in Function Spaces. Theory and Applications, Lecture Notes in Statist., Vol. 149, Springer, 2000.
[4] N. Dunford, J.T. Schwartz, Linear Operators I, Wiley, 1958.
[5] U. Grenander, Abstract Inference, Wiley, New York, 1981.
[6] J. Kuelbs, A strong convergence theorem for Banach valued random variables, Ann. Probab. 4 (1976) 744–771.
[7] M. Ledoux, M. Talagrand, Probability in Banach Spaces, Isoperimetry and Processes, Springer-Verlag, 1991.
[8] T. Mourid, Processus autorégressifs banachiques d'ordre supérieur, C. R. Acad. Sci. Paris, Série I 317 (1993) 1167–1172.
[9] T. Mourid, Repésentation autorégressive dans un espace de Banach de processus réels à temps continu et équivalence des lois, C. R. Acad. Sci. Paris, Série I 322 (1996) 1219–1224.
[10] E. Parzen, An approach to time series analysis, Ann. Math. Statist. 32 (1961) 951–989.
[11] Ju.A. Rozanov, Infinite-Dimensional Gaussian Distributions, American Mathematical Society, Providence, RI, 1971.