Estimateurs cribles des processus autorégressifs Banachiques

Fatiha Rachedi

THÈSE DE DOCTORAT DE L'UNIVERSITÉ PARIS 6

Spécialité : Mathématiques

Option : Statistique

présentée par

Fatiha RACHEDI

pour obtenir le grade de

DOCTEUR DE L'UNIVERSITÉ PARIS 6

Sujet de la thèse

Estimateurs cribles des processus autorégressifs Banachiques

Soutenue le 17 novembre 2005 devant le jury composé de :

M. Marc YOR	Université *Paris* 6	*P*résident
M. Denis BOSQ	Université *Paris 6*	*D*irecteur de recherche
M. Michel BRONIATOWSKI	Université *Paris 6*	*E*xaminateur
M. Herold DEHLING	Université Bochum	*R*apporteur
M. Yury A. KUTOYANTS	Université du Maine	*R*apporteur
M. Besnik PUMO	Institut National d'Horticulture, Angers	Examinateur
M. Mekki TERBECHE	Université *d'Oran*	*E*xaminateur

Remerciements

Tout d'abord, je tiens à remercier trés vivement mon directeur de thèse Monsieur Denis Bosq et à lui exprimer ma profonde gratitude pour son aide et son soutien constant durant ces trois dernières années de travail. Je tiens à dire qu'il a réussi par ses conseils et ses encouragements à developper le plaisir de la recherche.

Je remercie très chaleureusement Messieurs Herold Dehling et Yury A. Kutoyants pour avoir accepté d'être rapporteurs de cette thèse. Je leur suis trés reconnaissante du temps qu'ils ont consacré à l'évaluation de ce travail. J'apprécie à sa juste valeur leur présence dans le jury.

J'adresse mes sincères remerciements à Monsieur Marc Yor d'avoir accepté de présider le jury de cette thèse, à Messieurs Michel Broniatowski, Besnik Pumo et Mekki Terbeche pour l'intérêt qu'ils ont accordé à mon travail et pour avoir accepté de participer au jury.

Mon activité de recherche s'est déroulée au Laboratoire de Statistique Théorique et Appliquée de l'Université Paris VI, dirigé par Paul Deheuvels, je lui témoigne tout mon estime pour l'aide qu'il apporte aux jeunes chercheurs. Je tiens à remercier également tous les membres du laboratoire.

Je veux remercier trés chaleureusement Louise Lamart, Anne Durrande et Pascal Epron qui font preuve chaque jour de leur gentillesse et patience. Je voudrais adresser un salut amical à tous mes collègues du laboratoire et d'ailleurs : Malika Korso Dali-youcef, Karima Hamdani, Noura Yahiaoui, Abdelnasser Menacer, Samia Achiou, Salim Bouzebda, Amor Keziou, Jihen Najar, Fateh Chebana, Jean-Baptiste Aubin, Samuela Leoni, Pierre Ribereau, Céline Turbillon, Rosalba et tant d'autres avec qui j'ai partagé de bons moments.

Je remercie enfin toute ma famille pour leur soutien tout au long de la préparation de cette thèse.

J'ai une pensée émue pour le père d'Imène sans lequel je n'aurais pu commencer cette thèse. Imène a vécu ces neufs derniers mois au jour le jour les doutes et angoisses de la recherche scientifique, j'espère que ce n'était pas trop pesant.

Fatiha RACHEDI

LSTA-Université Paris 6. Le 16 novembre 2005

Table des matières

Introduction Générale

L'interprétation d'un processus comme élément aléatoire à valeurs dans un espace fonctionnel s'est avérée être une approche fructueuse pour aborder certains problèmes d'estimation. Par exemple la théorie de l'estimation dans les processus de diffusion utilise ce genre de technique.

Nous nous intéressons ici aux processus autorégressifs fonctionnels introduit par Bosq (1991). Il s'agit de construire des représentations de processus à temps continu par des processus à temps discret dans un espace fonctionnel ou des espaces de suites. Un tel modèle est de la forme :

$$X_n = \rho(X_{n-1}) + \varepsilon_n, \quad n \in \mathbb{Z} \tag{1}$$

où les X_n sont des variables aléatoires à valeurs dans un espace de Banach B, ρ un opérateur linéaire borné et la suite $(\varepsilon_n, n \in \mathbb{Z})$ est un bruit blanc à valeurs dans B. On le notera ARB(1).

Nous indiquons maintenant comment construire un processus ARB(1) à partir d'un processus à temps continu en prenant l'exemple du processus d'Ornstein-Uhlenbeck. Soit $(\eta(t),\ t \in \mathbb{R})$ un processus à temps continu et à trajectoires continues, on pose

$$X_n(t) = \eta(nh + t), \qquad 0 \leq t \leq h,\ n \in \mathbb{Z}.$$

Ce qui définit un processus à temps discret à valeurs dans $C[0, h]$.
Soit $(\eta(t),\ t \in \mathbb{R})$ un processus d'Ornstein-Uhlenbeck :

$$\eta(t) = \int_{-\infty}^{t} e^{-\theta(t-s)} dW(s), \quad t \in \mathbb{R}$$

où $\theta > 0$ et $W(t)$ un processus de Wiener bilatéral.
On considère l'opérateur $\rho_\theta : C[0, h] \to C[0, h]$ défini par :

$$\rho_\theta(f(t)) = e^{-\theta t} f(h), \qquad 0 \leq t \leq h,\ f \in C[0, h]$$

alors $\rho_\theta \in \mathcal{L}$, $\|\rho_\theta\|_{\mathcal{L}} = 1$, et en général

$$\|\rho_\theta^j\|_{\mathcal{L}} = e^{-\theta(j-1)h}.$$

(X_n) est alors un ARB(1) associé au bruit blanc

$$\varepsilon_n(t) = \int_{nh}^{nh+t} e^{-\theta(nh+t-s)} dW(s), \qquad t \in [0, h], n \in \mathbb{Z}$$

Ce modèle est utilisé pour la prévision des processus à temps continu. La prévision d'un ARB(1) passe par l'estimation de l'opérateur d'autocorrélation ρ définissant la structure autorégressive du processus. Celle ci intervenant après l'estimation des opérateurs de covariance (et de la moyenne dans le cas non centré) : sur ces sujets nous renvoyons aux travaux de Bosq (1991), Pumo (1995), Besse et Cardot (1996), et Dehling et Sharipov (2005), parmi d'autres.

Nous nous intéressons particulièrement à l'estimation de ρ. Ce problème a été abordé par Bosq (1991) où B est un espace de Hilbert et l'opérateur ρ est symétrique compact.

De nombreux travaux ont ensuite été réalisés tant sur le plan théorique qu'appliqué. Pumo (1995) a étudié le problème dans l'espace $C[0,1]$ lorsque ρ est un opérateur intégral. Mourid (1996) a développé la théorie des processus autorégressif d'ordre p. Besse et Cardot (1996) et Cardot (1998) ont utilisé les splines pour la prévision du trafic automobile et du phénomène climatique "El niño". Antoniadis et Sapatinas (2003) ont utilisé les ondelettes pour construire directement un prédicteur. Mas (1999) (2004) a étudié la normalité asymptotique de l'estimateur de ρ. Guillas (2000) (2002) s'est intéressé à la vitesse de convergence de l'estimateur de ρ et a étudié une extension du modèle autorégressif Hilbertien avec variables exogènes, dans Damon et Guillas (2005) il applique ce modèle à des données de pollution par l'ozone.

Une méthode générale pour estimer ρ consiste à utiliser la relation

$$D = \rho C$$

où C et D sont respectivement les opérateurs de covariance et covariance croisée du processus (1). L'estimateur $\hat{\rho}_n$ de ρ s'obtient alors à partir des estimateurs empiriques C_n et D_n de C et D respectivement, mais comme l'opérateur C_n n'est pas inversible en général, on est amené à projeter les observations sur l'espace engendré par les k_n premières vecteurs propres de C_n. Le problème de la détermination du k_n idéal est difficile. Cette méthode a été développée par les auteurs cités précédemment. Pour notre part nous utilisons la méthode des cribles ou "sieves" de Grenander (1981).

Les méthodes classiques d'estimation (maximum de vraisemblance et moindres carrées) s'avèrent inadéquates quand l'espace paramétrique est de dimension infinie, Grenander (1981) a proposé d'estimer le paramètre sur un sous espace de dimension en général finie m, puis d'étudier la consistance de cet estimateur lorsque la dimension m tend vers l'infini avec le nombres d'observations à vitesse convenable. Cette méthode a permis de résoudre de nombreux problèmes. Voici un exemple :

Soit $(\eta(t),\, t \in [-\frac{1}{2}, \frac{1}{2}])$ un processus vérifiant :

$$\eta(t) = \int_{-\frac{1}{2}}^{t} \alpha(s)ds + W(t) \qquad t \in [-\frac{1}{2}, \frac{1}{2}]$$

où $\alpha \in L^2[-\frac{1}{2}, \frac{1}{2}]$ et W est un processus de Wiener de variance 1.

On veut estimer α. On note les coefficients de Fourier de $\eta(t)$, α et $W(t)$ par x_k, α_k et w_k respectivement. On suppose que les coefficients de Fourier de α contenant le sinus sont nuls.

Le processus $(\eta_t, t \in [-\frac{1}{2}, \frac{1}{2}])$ induit une loi de probabilité P_α sur $C([-\frac{1}{2}, \frac{1}{2}])$. La dérivée de Radon Nikodym de P_α par rapport à P_0 est :

$$\frac{P_\alpha(dx)}{P_0(dx)} = \exp(\sum_{k=0}^{+\infty}(\alpha_k x_k - \frac{1}{2}\alpha_k^2)) \tag{2}$$

On maximise chaque terme de la série dans (2). Si $(x_{1,k}, ..., x_{n,k})$ est un échantillon d'observations indépendantes identiquement distribuées de x_k, l'estimateur du maximum de vraisemblance de α est $\hat{\alpha}_k = \dfrac{1}{n}\sum_{i=1}^{n} x_{i,k}$. Mais $x_k = \alpha_k + w_k$ et $\alpha_k \in l^2$, or $w_k \notin l^2$ car $\sum\limits_{k=0}^{+\infty} var(w_k) = \sum\limits_{k=0}^{+\infty} 1 = +\infty$ presque sûrement, donc $x_k \notin l^2$. D'où le maximum de vraisemblance n'existe pas dans l^2.

On peut choisir un crible Θ_m définie par

$$\Theta_m = \left\{\alpha, \quad \alpha \in L^2 \;/\; \sum_{k=0}^{\infty} k^2\alpha_k^2 \leq m\right\}$$

Soit λ le multiplicateur de Lagrange. Le problème est :

$$\left\{\begin{array}{l} \max\limits_{(\alpha_k)} \sum\limits_{k=0}^{\infty}\left(\sum\limits_{i=1}^{n}(\alpha_k x_{i,k} - \frac{1}{2}\alpha_k^2) - \lambda k^2\alpha_k^2\right) \\ \sum\limits_{k=0}^{\infty} k^2\alpha_k^2 \leq m \end{array}\right.$$

On obtient alors

$$\hat{\alpha}_k = \frac{\sum\limits_{i=1}^{n} x_{i,k}}{n+\lambda k^2} \quad \text{tel que} \quad \sum_{k=0}^{\infty} k^2 \frac{(\sum\limits_{i=1}^{n} x_{i,k})^2}{(n+\lambda k^2)^2} = m_n$$

Dans ce travail nous nous intéressons plus précisément au cas où la structure de l'opérateur ρ permet de définir des cribles de dimension finie. Nous avons donc pu utiliser les méthodes classiques d'estimation (maximum de vraisemblance, moindres carrées). Notons que plus généralement il serait possible d'utiliser la méthode des $\phi-$divergences, développée par Broniatowski (2003) et Keziou (2003). Nous espérons revenir ultérieurement sur cette question.

Les premiers travaux sur l'application des cribles aux ARB(1) ont été effectué par Bensmain et Mourid (2001) dans le cas d'un ARB(1) gaussien et lorsque ρ est un opérateur de Hilbert-Schmidt. Ils ont montrer que l'estimateur du maximum de vraisemblance de ρ converge p.s. par rapport à la norme de Hilbert-Schmidt, sous l'hypothèse que la loi induite par la variable aléatoire X_n est absolument continue par rapport à la loi induite par la variable aléatoire ε_n. Ils ont appliqué le résultat trouvé à un ARB(1) pour $B = C[0,1]$ et ρ un opérateur intégral, en utilisant un lemme de Kuelbs (1970) qui permet de plonger avec densité un espace de Banach séparable B dans un espace de Hilbert H. l'estimateur $\hat{\rho}_n$ de ρ est construit à l'aide de la base trigonométrique dans $L^2[0,1]$. Cet estimateur converge p.s. par rapport à la norme 2-intégrale, avec une vitesse de convergence du crible de l'ordre de $n^{\frac{1}{3}-\delta}$, $0 < \delta < \frac{1}{3}$.

Dans le premier chapitre nous rappelons certains résultats de la théorie des processus autorégressifs fonctionnels. Plus précisément sur l'estimation de l'opérateur ρ effectués par Bosq, Pumo (1995), Mas (1999), Guillas (2001), Bensmain et Mourid (2001).

Dans le deuxième chapitre nous nous intéressons à un ARB(1), dans le cas où ρ est un opérateur p-sommable, $p \in]1,+\infty[$, qui est pour $p = 2$ une généralisation naturelle dans un espace de Banach d'un opérateur de Hilbert-Schmidt (Diestel *et al.* (1995)). Comme l'opérateur de covariance ε_n est un opérateur nucléaire défini positif de B^* dans B, où B^* désigne le dual topologique de B, nous définissons la méthode des moindres carrées comme problème d'optimisation en utilisant l'ordre induit par le cône des opérateurs nucléaires définis positifs dans l'espace des opérateurs nucléaires de B^* dans B. La généralisation des problèmes d'optimisation réelle à des problèmes d'optimisation vectorielle est donnée par exemple dans Pallaschke et Rolewicz (1997).

Soit (Θ_m) un crible associé à l'espace des paramètres, qu'on note Θ. Pour $\rho \in \Theta$, soit C_ρ l'opérateur défini par

$$C_\rho(\cdot) = E\left((\cdot, X_1 - \rho X_0)(X_1 - \rho X_0)\right),$$

ou encore

$$C_\rho = C - \rho D^* - D\rho^* + \rho C \rho^*.$$

On note $\mathcal{N}'$ l'espace des opérateurs nucléaires de B^* dans B, muni de la norme nucléaire, et K le cône des opérateurs définis positifs de $\mathcal{N}'$.

L'estimateur crible des moindres carrées de ρ, quand il existe, est solution du problème

$$\overset{\vee}{\inf}_{\rho \in \Theta_m} \left(C_n - \rho D_n^* - D_n \rho^* + \rho C_n \rho^*\right) \tag{3}$$

où $\overset{\vee}{\inf}$ est l'infinimum dans $\mathcal{N}'$ par rapport à l'ordre induit par K. Dans Pallaschke et Rolewicz (1997) nous trouvons la solution de ce type de problèmes en utilisant la dérivée de Fréchet. Si $\hat{\rho}_n$ est solution du problème 3, $\hat{\rho}_n$ vérifie la relation

$$D_n = \hat{\rho}_n C_n.$$

Pour trouver une forme explicite de cet estimateur nous avons considéré ρ un opérateur strictement p-intégral et une base (u_k) dans B. Ce qui a permis d'obtenir une décomposition de l'opérateur ρ en somme infinie d'opérateurs ρ_k de rang 1

$$\rho = \sum_{k \geq 0} \alpha_k \ \rho_k \tag{4}$$

où $(\alpha_k)_k$ est une suite dans ℓ^p. Ainsi si les ρ_k sont connus, $\forall k \geq 0$, l'estimation de ρ revient à l'estimation des $(\alpha_k)_k$ dans ℓ^p. Nous donnons des exemples où les opérateurs ρ_k sont connus. Θ est alors $\left\{ \rho = \sum\limits_{k\geq 0} \alpha_k \ \rho_k \ / \quad (\alpha_k)_k \in \ell^p \right\}$,

La décomposition (4) permet aussi d'utiliser tout simplement les cribles

$$\Theta_m = \{ \ \rho \in \Theta \ / \quad \alpha_k = 0 \ , \ k > m \} \, , \qquad m \geq 0, \ m = m\,(n) \underset{n \to +\infty}{\longrightarrow} +\infty$$

Soit (f_k^*) sont les fonctionnelles de coefficients associée à (u_k). Nous avons alors la

Proposition 0.0.1. *Si* $(C_n \rho_k^* f_k^*, \rho_k^* f_k^*) > 0$ *pour* $k = 0, \ldots, m$; *l'estimateur crible des moindres carrées de* ρ *est l'opérateur* $\widehat{\rho}_{n,m} = \sum\limits_{k=0}^{m} \widehat{\alpha}_k \ \rho_k$

$$\text{où } \widehat{\alpha}_k = \frac{(D_n \rho_k^* f_k^*, f_k^*)}{(C_n \rho_k^* f_k^*, \rho_k^* f_k^*)} \qquad k = 0, \ldots, m.$$

Pour une dimension m du crible et sous des hypothèses de régularités, la vitesse de convergence p.s. de l'estimateur, par rapport à une norme équivalente asymptotiquement à la norme $p-$intégrale, est au moins de l'ordre de $n^{\frac{1}{min(2,p)}} \, (\log n)^{\frac{-\beta}{min(2,p)}}$, $\beta > 1$. Nous montrons aussi la normalité asymptotique de l'estimateur. Enfin nous illustrons ce résultat par une simulation d'un ARB(1) gaussien, avec le logiciel STATISTICA.

Dans le troisième chapitre nous utilisons la méthode des $\phi-$divergences entre des probabilités conditionnelles, dans le but d'estimer ρ. Cette méthode consiste à associer à un espace de paramètres une famille de probabilités qui sont absolument continue par rapport à une loi de probabilité.

Soit φ une fonction convexe de $[0, +\infty]$ dans $[0, +\infty]$. Pour toute mesure de probabilité Q et P tel que Q est absolument continues (a.c.) par rapport à P, la $\phi-$divergence entre Q et P est définie par Rüschendorf (1984) comme

$$\phi(Q, P) = \int \varphi(\frac{dQ}{dP}) dP. \tag{5}$$

Soit maintenant

- Θ l'espace des opérateurs $p-$sommables et la distance associée d est celle déduite de la norme.
- P la loi stationnaire induite sur $(B, \mathcal{B})$ par X_t et P_0 la loi de la variable aléatoire ε_0.
- $\mathcal{B}_t = \sigma(\varepsilon_i, i \leq t)$ la tribu engendré par la suite $(\varepsilon_i, i \leq t)$.
- Pour ρ dans Θ, $P_\rho^{\mathcal{B}_{t-1}}$ désigne la probabilité conditionnelle, par rapport à la tribu $\mathcal{B}_{t-1}$, de la variable aléatoire $(\rho X_{t-1} + \varepsilon_t)$.

Nous supposons que P est absolument continu par rapport à P_0. Dans Bosq et Mourid (1999) dans le cas gaussien et sous certaines conditions, les lois P et P_0 sont équivalentes.

La densité de $P_\rho^{\mathcal{B}_{t-1}}$ par rapport à P_0 est

$$g(x, X_{t-1}, \rho) = \frac{dP_\rho^{\mathcal{B}_{t-1}}(x)}{dP_0(x)}, \qquad x \in B.$$

Alors pour $\rho, \theta \in \Theta$ la $\phi-$divergence entre $P_\rho^{\mathcal{B}_{t-1}}$ et $P_\theta^{\mathcal{B}_{t-1}}$ définie par la fonction φ est :

$$\phi(P_\rho^{\mathcal{B}_{t-1}}, P_\theta^{\mathcal{B}_{t-1}}) = \int \varphi\left(\frac{g(x, X_{t-1}, \rho)}{g(x, X_{t-1}, \theta)}\right) dP_\theta^{\mathcal{B}_{t-1}}(dx)$$

Nous utilisons alors la représentation duale de la $\phi-$divergence proposé par Broniatowski (2003) et Keziou (2003) en estimation paramétrique pour définir l'estimateur du minimum des $\phi-$divergences.

Nous nous limitons ici à l'étude du cas $\varphi(x) = -log(x) + x - 1$, alors l'estimateur dit du minimum de KL-divergence (divergence de Kullback-Leibler), quand il existe, s'écrit sous la forme

$$\arg \inf_{\rho \in \Theta_m} \sup_{\theta \in \Theta_m} \left(\frac{1}{n}\sum_{i=1}^{n} \log(\frac{g(x, X_{t-1}, \rho)}{g(x, X_{t-1}, \theta)})\right).$$

Il est clair que cet estimateur est celui du maximum de vraisemblance. Nous montrons par la suite qu'il converge p.s. vers la vraie valeur du paramètre pour la norme des opérateurs $p-$sommables. La démonstration est basée sur les techniques de Geman et

Hwang (1982), utilisées pour des observations indépendantes et identiquement distribuées, qu'on a adapté au cas autorégressif. Dans le cas d'un ARB(1) gaussien avec ρ un opérateur strictement 2−intégral, nous montrons que si la dimension du crible est de l'ordre de $n^{\frac{1}{3}-\delta}$, $0<\delta<\frac{1}{3}$, on a la convergence p.s. pour la norme 2−intégrale.

Ensuite, nous étudions le cas particulier où ρ est un opérateur 2−nucléaire. Les opérateurs 2−nucléaires sont des opérateurs strictement 2−intégraux, cependant ils admettent une décomposition plus simple.

Enfin nous nous inspirons de l'exemple cité dans Antoniadis et Beder (1989) (p. 81) pour donner l'application suivante

Soit $(X_t,\ t\in\mathbb{Z})$ un ARB(1) gaussien avec $B=C([0,1])$ muni de la topologie de la convergence uniforme. Nous supposons que P_{X_0} et P_{ε_0}, les lois gaussiennes de X_0 et ε_0 respectivement, sont équivalentes et nous considérons P_{ε_0} la mesure de Wiener de fonction de covariance $\gamma(s,t)=\min(s,t)$.

L'estimateur crible de ρ est alors

$$\begin{aligned}\widehat{\rho}_m(f)(s) &= \int_0^1\left(\widehat{\alpha}_0\cdot I_{[0,1]}(s)\cdot+2\sum_{k=1}^{m}\widehat{\alpha}_k\frac{(-1)^k}{k\pi}\sin k\pi s\right)\cdot f(t)\cdot\delta_1(dt)\\ &\qquad +\int_0^1 2f(t)\left(\sum_{k=1}^{m}\widehat{\alpha}_k\ \sin k\pi s\cdot\sin k\pi t\ dt\right)dt,\qquad f\in C([0,1])\end{aligned}$$

où $\widehat{\alpha}_0=\dfrac{\sum_{i=1}^{n}X_{i-1}(1)\ X_i(1)}{\sum_{i=1}^{n}X_{i-1}^2(1)}$,

$$\widehat{\alpha}_k=\frac{\sum_{i=1}^{n}\left((-1)^k X_{i-1}(1)+k\pi\int_0^1 X_{i-1}(t)\ \sin k\pi t\ dt\right)\left((-1)^k X_i(1)+k\pi\int_0^1 X_i(t)\ \sin k\pi t\ dt\right)}{\sum_{i=1}^{n}\left((-1)^k X_{i-1}(1)+k\pi\int_0^1 X_{i-1}(t)\ \sin k\pi t\ dt\right)^2+2n\lambda_n k},$$

$k=1,\ldots,m.$

Nous posons $K_1(s,t)=\alpha_0\cdot I_{[0,1]}(s)\cdot+2\sum_{k\geq 1}\alpha_k\dfrac{(-1)^k}{k\pi}\sin k\pi s$,

$K_2(s,t)=2\sum_{k\geq 1}\alpha_k\ \sin k\pi s\cdot\sin k\pi t,$

et $K = K_1 + K_2$.

L'opérateur ρ est un opérateur à noyau K par rapport à la mesure $\left(\delta_1 + I_{[0,1]}\right) \eta$ où η est la mesure de Lebesgue.

Enfin, dans l'appendice nous faisons quelques rappels des résultats d'analyse fonctionnelle qui sont utilisés dans les preuves de propositions du chapitre 2 et dans la définition de l'estimateur des moindres carrées dans un espace de Banach du chapitre 3.

Travaux et Publications :

- **RACHEDI F. (2005)**. "Vitesse de convergence en norme $p-$intégrale et normalité asymptotique de l'estimateur crible de l'opérateur d'un ARB(1)". C.R. Acad. Sci. Paris, t. 341, Série I, p. 369-374.
- **RACHEDI F. (2004)**. " Vitesse de convergence de l'estimateur crible d'un processus ARB(1) dans le cas d'un opérateur strictement 2-intégral". Annales de l'ISUP, Vol. 48, fasicule 3, p. 87-97.
- **RACHEDI F. et MOURID T. (2003)**. " Estimateur crible de l'opérateur d'un processus ARB(1) ". C.R. Acad. Sci. Paris, t. 336, Série I, p. 605-610.
- **RACHEDI F.** . " Estimateur des moindres carrées dans un espace de Banach de l'opérateur d'un ARB(1)". En préparation.

Chapitre 1

Processus autorégressifs en dimension infinie

1.1 Processus autorégressifs Hilbertiens

Dans ce chapitre nous rappelons certains résultats de la théorie des processus autorégressifs fonctionnels introduits par Bosq (1991). Nous nous intéressons à l'estimation de l'opérateur d'autocorrélation définissant la structure autorégressive du processus. Nous rappelons d'abord les résultats d'estimation dans le cas où cet opérateur est défini sur un espace de Hilbert.

Soit H un espace de Hilbert réel séparable muni du produit scalaire $< \, , \, >$ associé à la norme $\|\cdot\|$, et sa tribu borélienne $\mathcal{B}$. Soient ρ un opérateur linéaire borné sur H tel que

$$\|\rho\|^{j_0} < 1 \text{ pour un certain } j_0 \geq 1$$

et $\varepsilon = (\varepsilon_t, t \in \mathbb{Z})$ une suite de variables aléatoires indépendantes et de même loi à valeurs dans H, telles que

$$0 < E\,\|\varepsilon_t\|^2 = \sigma^2 < +\infty \text{ et } E(\varepsilon_t) = 0$$

Le processus autoregressif hilbertien stationnaire d'ordre 1, noté ARH(1), est l'unique solution stationnaire de l'équation :

$$X_t = \rho(X_{t-1}) + \varepsilon_t, \quad t \in \mathbb{Z} \tag{1.1}$$

Une méthode générale pour estimer ρ consiste à utiliser les opérateurs de covariance et covariance croisée du processus.

L'opérateur de covariance de X_0 est un opérateur symétrique positif et nucléaire de H dans H défini par :

$$C(x) = E[< X_0, x > X_0], \quad x \in H$$

L'opérateur de covariance croisée est :

$$D(x) = E[< X_0, x > X_1], \quad x \in H$$

Les opérateurs C et D vérifient la relation

$$D \;\; = \;\; \rho C. \tag{1.2}$$

Alors, pour estimer ρ au vu des observations $(X_1, ..., X_n)$ on commence par estimer C et D en posant :

$$C_n\left(x\right) = \frac{1}{n}\sum_{i=1}^{n} < X_i, x > X_i$$

et

$$D_n\left(x\right) = \frac{1}{n-1}\sum_{i=1}^{n-1} < X_i, x > X_{i+1}$$

Comme C_n n'est pas inversible en général, on est amené à projeter les observations sur l'espace engendré par les k_n premiers vecteurs propres de C, ou s'ils sont inconnus, sur l'espace engendré par les k_n premiers vecteurs propres de C_n .

Soit alors (λ_j) l'ensemble des valeurs propres de C et H_{k_n} l'espace engendré par $v_1, ..., v_{k_n}$ les premiers vecteurs propres de C, où (k_n) est une suite d'entiers telles que $k_n \leq n$, $n \geq 1$, et $k_n \to \infty$.

On suppose que :

$\mathbf{A_1}$: $E\|X_0\|^4 < \infty$,

$\mathbf{A_2}$: $\lambda_j > 0, \ \forall j \geq 1$,

$\mathbf{A_3}$: $P(< X_0, v_j >= 0) = 0,\ \forall j \geq 1.$

Nous distinguons deux cas :

a) Si les (v_j) sont connus, on peut remplacer C_n par :

$$\hat{C}_n = \sum_{j=1}^{\infty} \hat{\lambda}_{jn} < v_j, \cdot > v_j$$

où $\hat{\lambda}_{jn} = \frac{1}{n}\sum_{i=0}^{n} < X_i, v_j >^2$, $j \geq 1$, $n \geq 1$,

avec $\sum_{j=1}^{\infty} \hat{\lambda}_{jn} = \frac{1}{n}\sum_{j=1}^{\infty}\sum_{i=0}^{n} < X_i, v_j >^2 = \frac{1}{n}\sum_{i=0}^{n} \|X_i\|^2 < \infty.$

$\hat{C}_n$ est inversible sur H_{k_n}, alors l'estimateur de ρ est :

$$\hat{\rho}_n(x) = (\pi_{k_n} D_n \hat{C}_n^{-1} \pi_{k_n})(x),\ \ x \in H$$

où $\hat{C}_n^{-1} = \sum_{j=1}^{\infty} \hat{\lambda}_{jn}^{-1} < v_j, \cdot > v_j$ et π_{k_n} est le projecteur orthogonal sur H_{k_n}.

b) Si les (v_j) sont inconnus, H_{k_n} est remplacé par $\tilde{H}_{k_n}$ l'espace engendré par $v_{1n}, ..., v_{k_n n}$ les vecteurs propres de C_n. Dans ce cas on fait les hypothèses suivantes

$\mathbf{B_1}$: $\lambda_1 > \lambda_2 > ... > \lambda_j > ... > 0,$

$\mathbf{B_2}$: $\lambda_{k_n n} > 0,\ \ n \geq 1$ (p.s.).

L'estimateur de ρ s'écrit alors :

$$\hat{\rho}_n(x) = \tilde{\pi}_{k_n} D_n \tilde{C}_n^{-1} \tilde{\pi}_{k_n}(x),\ \ x \in H \tag{1.3}$$

où $\tilde{C}_n^{-1} = \sum_{j=1}^{\infty} \hat{\lambda}_{jn}^{-1} < v_{jn}, \cdot > v_{jn}$ et $\tilde{\pi}_{k_n}$ est le projecteur orthogonal de $\tilde{H}_{k_n}$.

1.2 Convergence et normalité asymptotique de l'estimateur de l'opérateur d'un ARH(1)

Soit $(X_t,\ t \in \mathbb{Z})$ un ARH(1). On considère le cas général où les (v_j) sont inconnus.

Bosq a montré la convergence p.s. de l'estimateur (1.3) en norme linéaire en considérant les notations suivantes

$$a_1 = 2\sqrt{2}(\lambda_1 - \lambda_2)^{-1} \qquad si \quad \lambda_1 \neq \lambda_2$$

et

$$a_j = 2\sqrt{2}max[(\lambda_{j-1} - \lambda_j)^{-1}, (\lambda_j - \lambda_{j+1})^{-1}]$$

Théorème 1.2.1. *On suppose que (A_1), (B_1) et (B_2) sont vérifiées et ρ est un opérateur de Hilbert Schmidt. Alors si pour $\beta > 1$*

$$\lambda_{k_n}^{-1}\sum_{j=1}^{k_n} a_j = O(n^{1/4}(\log n)^{-\beta})$$

on a :

$$\|\hat{\rho}_n - \rho\|_{\mathcal{L}} \longrightarrow 0 \qquad p.s.$$

Si de plus $\|X_0\|$ est bornée, alors

$$P(\|\hat{\rho}_n - \rho\|_{\mathcal{L}} \geq \eta) \leq c_1(\eta) \exp\left(-c_2(\eta) n \lambda_{k_n}^2 (\sum_{j=1}^{k_n} a_j)^{-2}\right),$$

où $\eta > 0$, $n \geq \eta(n)$, et $c_1(\eta)$, $c_2(\eta)$ sont deux constantes posititives.

Ainsi $\dfrac{n\lambda_{k_n}^2}{\log n(\sum_{j=1}^{k_n} a_j)^2} \to 0$ *implique* $\|\hat{\rho}_n - \rho\|_{\mathcal{L}} \to 0 \quad p.s.$

Guillas (2001) a introduit une légère modification sur l'estimateur $\hat{\rho}_n$ afin de donner une vitesse de convergence de $E\|\hat{\rho}_n - \rho\|_{\mathcal{L}}^2$. Il a considéré l'hypothèse suivante

(**H**) : Il existe une suite (a_n) qui satisfait :

$$\exists 0 < \beta < 1,\ 0 < a_n \leq \beta\lambda_{k_n},\ n \in \mathbb{N}.$$

La suite (a_n) permet de mieux contrôler les variations des valeurs propres de C_n . En effet, si ses valeurs propres se rapprochent trop rapidement de 0, il devient difficile de maîtriser le comportement de C_n^{-1} et la vitesse de convergence se dégrade.

Sous l'hypothèse (**H**) on utilise alors l'estimateur :

$$\hat{\rho}_n(x) = \tilde{\pi}_{k_n} D_n \tilde{C}_{n,a}^{-1} \tilde{\pi}_{k_n}(x), \; x \in H, \tag{1.4}$$

où

$$\hat{C}_{n,a} = \sum_{j=1}^{\infty} max(\tilde{\lambda}_{jn}, a_n) < v_{jn}, \cdot > v_{jn}.$$

Théorème 1.2.2. *Guillas (2001) : On suppose que* (A_1), (A_2) *et* (B_1) *sont vérifiées, et qu'il existe* $\alpha > 0$, $0 < \beta < 1$, $\varepsilon < 1/2$ *et* $\gamma \geq 1$ *tel que*

$$\alpha \frac{\lambda_{k_n}^{\gamma}}{n^{\varepsilon}} \leq a_n \leq \beta \lambda_{k_n}$$

alors

$$E\|\hat{\rho}_n - \rho\|_{\mathcal{L}}^2 = O\left(\frac{\Lambda_{k_n}^2}{n^{(1-2\varepsilon)}\lambda_{k_n^{2(1+\gamma)}}}\right) + O(\lambda_{k_n}^2)$$

où $\Lambda_{k_n} = \sup_{j=1,\ldots,k_n} \dfrac{1}{\lambda_j - \lambda_{j+1}}$

Mas (1999) a établit le résultat suivant sur la normalité asymptotique de l'estimateur de ρ :

Théorème 1.2.3. *Sous les hypothèses :*

- (A_1), (B_1) *et* (B_2),
- C_n^{-1} *existe sur* $\tilde{H}_{k_n}$,
- $E\|C^{-1}(\varepsilon_0)\|^2 < \infty$,
- $n\lambda_{k_n}^4 \to \infty$, $n^{-1}\sum_{j=1}^{k_n} a_j \lambda_j^{-2} < \infty$,
- $\lambda_j \lambda_{jn}^{-1}$ *est borné en probabilité pour tout* j

on a :

$$\sqrt{n}(\hat{\rho}_n - \tilde{\pi}^{k_n}\rho) \xrightarrow{\mathcal{D}} N,$$

où $\mathcal{D}$ *désigne la convergence en loi dans l'espace des opérateurs de Hilbert Schmidt et* N *est une variable aléatoire gaussienne à valeurs dans cet espace.*

1.3 Estimateur crible d'un ARH(1)

Dans ce paragraphe nous rappelons les premiers résultats sur l'application des cribles à un ARH(1) effectuée par Bensmain et Mourid (2001). La méthode des cribles de Grenander (1981) consiste à maximiser la fonction de vraisemblance sur des sous espaces de dimension finie et croissant avec la taille de l'échantillon. Cette méthode a permis de résoudre de nombreux problèmes d'estimation nonparamétrique (Grenander (1981), Geman et Hwang (1982), Beder (1988)). Un crible est défini par :

Définition 1.3.1. *Un crible pour l'espace paramétrique* Θ *est une suite de sous ensembles* $\{\Theta_m\}_m$ *de* Θ *telle que* Θ_m *compact,* $\Theta_m \subset \Theta_{m+1}$*, et* $\bigcup_m \Theta_m$ *est dense dans* Θ.

On considére un ARH(1) avec (ε_n) un bruit blanc gaussien et ρ un opérateur de Hilbert-Schmidt.

On note P_ρ la loi stationaire induite sur $(H, \mathcal{B})$ par la variable aléatoire X_n et par P_0 la loi de la variable aléatoire ε_n. On suppose que P_ρ est absolument continue par rapport à P_0. Le modèle est identifiable dans le sens où la famille des lois $(P_\rho, \rho \in \Theta)$ est telle que $P_\rho \neq P_\beta$ si $\rho \neq \beta$.

Bensmain et Mourid (2001) ont montré la convergence p.s. de l'estimateur crible du maximum de vraisemblance en adaptant les techniques de Geman et Hwang (1982) au cas autorégressif.

On considère les notations suivantes :

a) pour $\rho \in \Theta_m$, $\quad B_m(\rho, \varepsilon) := \{\beta \in \Theta_m \ / \ d(\rho, \beta) < \varepsilon\}$.

b) La densité de la probabilité de transition stationnaire du processus (X_n) vérifiant (1), par rapport à P_0 est

$$g(x, y, \rho) = (P_\rho(dx/X_0 = y))/(P_0(dx)) \qquad x, y \in H, \quad \rho \in \Theta.$$

c) L'entropie conditionnelle est :

$$H(\rho, \beta) := E_\rho \ln g(x, y, \beta) = \int \ln(g(x, y, \beta))g(x, y, \rho)dP_0(x) \quad \rho, \beta \in \Theta.$$

d) Pour toute fonction réelle g définie sur un ensemble A et si $B \subseteq A$, on pose :

$g(B) := \sup_{y \in B} g(y)$.

e) $L_n(X_1, ..., X_n; \rho)$ est la vraisemblance conditionnelle.

f) L'ensemble des points de Θ_m où $L_n(x_1, ..., x_n; .)$ atteint son maximum est :

$$M_m^n = \{\rho \in \Theta_m / \quad L_n(\omega; \rho) = L_n(\omega; \Theta_m) := \sup_{\beta \in \Theta_m} L_n(\omega; \beta)\}$$

g) De même le sous ensemble de Θ_m où le maximum de l'entropie conditionnelle est atteint est :

$$A_m = \{\rho \in \Theta_m / \quad H(\rho_0, \rho) = H(\rho_0, \Theta_m) := \sup_{\beta \in \Theta_m} H(\rho_0, \beta)\}$$

où ρ_0 est la vraie valeur du paramètre.

h) Pour tout $C_m \subseteq \Theta$, la notation $C_m \to \rho$ signifie $\sup_{\beta \in C_m} d(\rho, \beta) \to 0$ quand $m \to \infty$.

Dans ce qui suit $m = m_n$ et m_n croit vers l'infini quand $n \to \infty$.

On a alors :

Théorème 1.3.1. *Supposons que le crible $\{\Theta_{m_n}\}$ est choisi tel que*

1) *pour tout n et tout $\rho \in \Theta_{m_n}$, il existe $\varepsilon > 0$ tel que*

$$E_{\rho_0} \ln(g(x, y, B_{m_n}(\rho, \varepsilon)) < \infty$$

2) *$A_{m_n} \to \rho_0$ quand $n \to +\infty$.*

Alors quand $n \to +\infty$ nous avons

$$M_{m_n}^n \to \rho_0 \quad p.s$$

Pour déterminer un ordre de croissance de la dimension m_n nous considérons les conditions suivantes

C1 : Si (ρ_{m_n}) est une suite telle que $\forall n$, $\rho_{m_n} \in \Theta_{m_n}$ et $H(\rho_0, \rho_{m_n}) \to H(\rho_0, \rho_0)$, alors $\rho_{m_n} \to \rho_0$.

C2 : il existe une suite $(\rho_{m_n}) \in \Theta_{m_n}$ telle que $H(\rho_0, \rho_{m_n}) \to H(\rho_0, \rho_0)$.

Pour tout $\delta > 0$ et tout n, on définit les ensembles

$$D_{m_n} = \{\rho \in \Theta_{m_n} \quad / \quad H(\rho_0, \rho) \leq H(\rho_0, \rho_{m_n}) - \delta\}$$

où ρ_{m_n} est la suite définie par **C2**.

Soit l sous ensembles $\Gamma_1, ..., \Gamma_l$ de Θ_{m_n}, on pose

$$\varphi_{m_n} := \sup_k \inf_{t \geq 0} E_{\rho_0} \exp\left[t \ln\left\{\frac{g(x, y, \Gamma_k)}{g(x, y, \rho_{m_n})}\right\}\right]$$

Le théorème suivant donne la convergence presque sûre de l'estimateur crible de l'opérateur ρ avec une condition sur la vitesse de croissance de la suite (m_n).

Théorème 1.3.2. *Soit* $\{\Theta_{m_n}\}$ *un crible vérifiant les conditions* **C1** *et* **C2**. *Supposons que pour tout* $\delta > 0$, *on peut trouver* $\Gamma_1^{m_n}, ..., \Gamma_{l_{m_n}}^{m_n}$ *dans* Θ_{m_n}, $n \geq 1$ *tels que*

$$(i) \quad D_{m_n} \subseteq \bigcup_{k=1}^{l_{m_n}} \Gamma_k^{m_n}$$
$$(ii) \quad \sum_{n=1}^{+\infty} l_{m_n} (\varphi_{m_n})^n < +\infty$$

alors

$$M_{m_n}^n \to \rho_0 \quad p.s.$$

Nous rappelons maintenant quelques résultats sur l'estimation de C et ρ dans le cas d'un processus autoregréssifs dans un espace de Banach.

1.4 Processus autoregressifs banachiques d'ordre 1

Toutes les variables aléatoires considérées dans ce paragraphe sont définies sur un espace de probabilité $(\Omega, \mathcal{A}, \mathbb{P})$. $(B, \mathcal{B})$ est un espace de Banach séparable sur $\mathbb{R}$, muni de sa tribu borélienne $\mathcal{B}$ et de la norme $\| \ \|$, B^* désigne le dual topologique de B et $(\cdot, \cdot)$ le crochet de dualité entre B^* et B.

Soit $\varepsilon = (\varepsilon_n, n \in \mathbf{Z})$ une suite de variables aléatoires définies sur Ω et à valeurs dans B. On dit que :

1. ε est un bruit blanc faible si :
   (a) $E\|\varepsilon_i\|^2 = \sigma^2 < \infty,\ i \in \mathbb{Z}$,
   (b) $E(\varepsilon_i) = 0,\ i \in \mathbb{Z}$,
   (c) C_{ε_i} ne dépend pas de i et $C_{\varepsilon_i,\varepsilon_j} = 0$, $i, j \in \mathbb{Z},\ i \neq j$.
2. ε est une différence de martingale si a), b) sont vérifiés et $E^{\mathcal{B}_{i-1}}(\varepsilon_i) = 0,\ i \in \mathbb{Z}$, où $\mathcal{B}_i = \sigma(X_j, j \leq i) = \sigma(\varepsilon_j, j \leq i)$ la tribu engendré par la suite $(\varepsilon_j, j \leq i)$.
3. ε est un bruit blanc fort si :
   (a) $\varepsilon_t, t \in \mathbb{Z}$, sont indépendants et identiquement distribués,
   (b) $E\|\varepsilon_0\|^2 = \sigma^2 < \infty$,
   (c) $E(\varepsilon_0) = 0$.

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 $\|\cdot\|_{\mathcal{L}}$.

Un processus autorégressif d'ordre 1 dans un espace de Banach est une suite $X = (X_t, t \in \mathbb{Z})$ de variables aléatoires à valeurs dans B telle que :

$$X_t = \rho(X_{t-1}) + \varepsilon_t, \quad t \in \mathbb{Z} \tag{1.5}$$

où $\varepsilon = (\varepsilon_t, t \in \mathbb{Z})$ est un bruit blanc dans B et $\rho \in \mathcal{L}(B)$ est tel que $\|\rho\|_{\mathcal{L}}^{j_0} < 1$ pour un $j_0 \geq 1$. Si ε est un bruit blanc faible, X est un processus autorégressif faiblement stationnaire, on le note WARB(1). Si ε est un bruit blanc fort, X est un processus autorégressif strictement stationnaire, on le note ARB(1). Nous donnons l'exemple suivant d'un processus à temps continu qui admet une représentation autorégressive.

Exemple 1.4.1. *Soit $\eta = (\eta(t),\ t \in \mathbb{R})$ un processus d'Ornstein-Uhlenbeck solution de l'équation différentielle stochastique de Langevin*

$$d\eta(t) \ = \ -\theta\eta(t)dt + dW(t), \tag{1.6}$$

où $\theta > 0$ et $W(t)$ un processus de Wiener bilatéral.

L'unique solution stationnaire de (1.6) s'écrit :

$$\eta(t) = \int_{-\infty}^{t} e^{-\theta(t-s)} dW(s), \quad t \in \mathbb{R}.$$

Soit $B = C[0,h]$, où $h > 0$ est fixé, et

$$X_n(t) = \eta(nh + t), \quad 0 \leq t \leq h, \ n \in \mathbb{Z}.$$

On considère l'opérateur $\rho_\theta : C[0,h] \to C[0,h]$ défini par :

$$\rho_\theta(x)(t) = e^{-\theta t} x(h), \qquad 0 \leq t \leq h,$$

alors $\rho_\theta \in \mathcal{L}$, $\|\rho_\theta\|_{\mathcal{L}} = 1$, et en général

$$\|\rho_\theta^j\|_{\mathcal{L}} = e^{-\theta(j-1)h}.$$

(X_n) est un processus autoégressif Banachique associé au bruit blanc

$$\varepsilon_n(t) = \int_{nh}^{nh+t} e^{-\theta(nh+t-s)} dW(s), \qquad t \in [0,h], \ n \in \mathbb{Z}.$$

Ce processus peut aussi s'écrire sous une autre forme, Mourid (1996) a donné la proposition suivante

Proposition 1.4.1. *Si $\eta = (\eta(t), \ t \in \mathbb{Z})$ est le processus d'Ornstein-Uhlenbeck alors la suite $X_n(t) = \eta(nh + t), \quad 0 \leq t \leq 1, \ n \in \mathbb{Z}$, est un ARB(1) où $B = C[0,1]$, et ρ est défini par :*

$$\rho_\theta(x)(t) = e^{-\theta t} x(t), \qquad 0 \leq t \leq 1,$$

le bruit blanc associé est

$$\varepsilon_n(t) = \int_{(n-1)t}^{nt} e^{-\theta(nt-s)} dW(s), \qquad t \in [0,1], n \in \mathbb{Z}.$$

1.4.1 Opérateurs de covariance dans un espace de Banach

Dans ce paragraphe nous rappelons la définition des opérateurs de covariance et covariance croisée dans un espace de Banach et certains résultats utiles concernant la convergence de leurs estimateurs.

Nous précisons d'abord l'espace auquel apppartiennent ces opérateurs. Soit $(\mathcal{L}', \| \ \|')$ l'espace des opérateurs linéaires et bornés de B^* dans B muni de la norme uniforme :

$$\|\ell\| = \sup_{\|x^*\| \leq 1} \|\ell(x^*)\|, \qquad \ell \in \mathcal{L}'.$$

On dit que $\ell \in \mathcal{L}'$ est nucléaire s'il admet la représentation suiante :

$$\ell(x^*) = \sum_{k=0}^{\infty} (x_k^{**}, x^*) y_k, \qquad x^* \in B^*,$$

où $(x_k^{**})_k \subset B^{**}$ le dual de B^* et $(y_k)_k \subset B$ avec $\displaystyle\sum_{k=0}^{\infty} \|x_k^{**}\| \ \|y_k\| < \infty$ ($\| \ \|$ désigne aussi la norme uniforme dans B^{**}). L'infinimum de $\displaystyle\sum_{k=0}^{\infty} \|x_k^{**}\| \ \|y_k\|$ par rapport à toutes les représentations possibles défini une norme, dite norme nucléaire. L'espace des opérateurs nucléaires de B^* dans B muni de la norme nucléaire, qu'on note $(\mathcal{N}', \| \ \|_{\mathcal{N}'})$, est un espace de Banach.

Les opérateurs de covariance et covariance croisée de X_0 et X_1 centrées, et telles que $E\|X_0\|^2 < \infty$ et $E\|X_1\|^2 < \infty$ sont :

$$C(x^*) = E\left((x^*, X_0) X_0\right), \quad x^* \in B^*.$$

et

$$D(x^*) = E\left((x^*, X_0) X_1\right), \quad x^* \in B^*.$$

Ces opérateurs sont des opérateurs nucléaires de B^* dans B.

Soit $(X_1, ..., X_n)$ des observations du processus vérifiant (1.5). Les opérateurs de covariance empirique et covariance croisée empirique de X_0 sont respectivement définis par :

$$C_n(x^*) = \frac{1}{n}\sum_{i=1}^{n}(x^*, X_i)\, X_i$$

et

$$D_n(x^*) = \frac{1}{n-1}\sum_{i=1}^{n-1}(x^*, X_i)\, X_{i+1}, \qquad x^* \in B^*.$$

Bosq (2002) a montré la convergence de C_n en utilisant les résultats sur la convergence de la moyenne empirique d'un certain WARB(1).
Soit Y_i l'opérateur de B^* dans B défini par :

$$Y_i(x^*) = (x^*, X_i)X_i, \;\; i \in \mathbb{Z}$$

D'après Bosq (2002) (lemme 4.1), $Y = (Y_i, \; i \in \mathbb{Z})$ est un processus autorégressif faiblement stationnaire à valeurs dans $\mathcal{N}'$, noté WAR$\mathcal{N}'$, qui s'écrit

$$Y_i - C = R(Y_i - C) + E_i, \qquad i \in \mathbb{Z}$$

où $R(\ell) = \rho\ell\rho^*$ est un opérateur nucléaire, $\ell \in \mathcal{N}'$ et ρ^* est l'opérateur adjoint de ρ, et E_i est une différence de martingale dans $\mathcal{N}'$ par rapport à $\mathcal{B}_i = \sigma(\varepsilon_j, j \leq i)$, $i \in \mathbb{Z}$, qui s'écrit

$$E_i(x^*) = (x^*, \varepsilon_i)\varepsilon_i + (x^*, \rho X_{i-1})\varepsilon_i + (x^*, \varepsilon_i)X_{i-1} - C_{\varepsilon_0}(x^*) \qquad x^* \in B^*.$$

De plus on utilise le lemme suivant

Lemme 1.4.1. *Si X est un WARB(1) tel que $E\|X_t\|^4 < \infty$ ne dépend pas de t et ε est une différence de martingale qui satisfait*

$$E^{\mathcal{B}_0}((u^*, \varepsilon_k)(v^*, \varepsilon_k)) = E((u^*, \varepsilon_k)(v^*, \varepsilon_k)), \qquad u^*, v^* \in B^*; \;\; k \geq 1$$

alors il existe $c_1 > 0$ et $c_2 \in]0, 1[$ deux constantes telles que

$$Cov((x^*, X_0)(y^*, X_0), (x^*, X_h)(y^*, X_h)) \leq \|x^*\|^2\|y^*\|^2 c_1 c_2^h, , \qquad x^*, y^* \in B^*; \;\; h \geq 1$$

On a alors la convergence de C_n vers C :

Proposition 1.4.2. *Soit X un WARB(1), Si les conditions du lemme précédent sont vérifiées, $\forall x^*, y^* \in B^*$ on a :*

$$\sup_{\|x^*\|\leq 1,\ \|y^*\|\leq 1} E(y^*, (C_n - C)(x^*))^2 = O(\frac{1}{n})$$

et

$$n^{1/2}(\ln n)^{-\beta}(y^*, (C_n - C)(x^*)) \underset{n\to+\infty}{\overset{p.s.}{\longrightarrow}} 0 \qquad \forall\beta > \frac{1}{2}.$$

Dans le cas d'un espace de Hilbert, $(C_n - C)$ convergence en norme nucléaire :

Proposition 1.4.3. *Bosq (2002) : Soit X un ARH(1), tel que $E\|X_t\|^4 < \infty,\ t \in \mathbb{Z}$, on a :*

$$\|C_n - C)\|_{\mathcal{N}} \underset{n\to+\infty}{\overset{p.s.}{\longrightarrow}} 0\,.$$

La loi asymptotique de $(C_n - C)$ est normale sous une condition sur la différence de martingale (E_i).

Proposition 1.4.4. *Bosq (2002) Soit X un ARB(1) tel que $E\|X_t\|^4 < \infty,\ t \in \mathbb{Z}$, et $\underset{i\geq 1}{sup} E\|E_i\|_p^{2+\delta} < \infty$ pour un $\delta > 0$, alors $\forall \ell^* \in \mathcal{N}'^*$,*

$$\ell^*(\sqrt{n}(C_n - C)) \xrightarrow{\mathcal{D}} N \sim \mathcal{N}'\left(0, E(\ell^*(I - R)^{-1}(E_0))^2\right).$$

Si B est un espace de Hilbert alors :

$$\sqrt{n}(C_n - C) \xrightarrow{\mathcal{D}} N \sim \mathcal{N}'\left(0, E((I - R)^{-1}C_{E_0}(I - R^*)^{-1})\right).$$

où $\mathcal{D}$ désigne la convergence en loi dans l'espace des opérateurs de Hilbert-Schmidt.

1.4.2 Equivalence de mesures induites par un ARB(1)

Parmi les résultats utiles de la théorie des ARB(1) il y'a celui de Bosq et Mourid (1999) sur l'equivalence des lois de ε_0 et X_0 dans la cas d'un ARB(1) gaussien. En effet d'après le théorème classique de Feldman-Hajek (Rozanov (1971) (p. 54)), deux mesures gaussiennes sont soit équivalentes, soit orthogonales. Dans le cas de l'équivalence, on peut déterminer la dérivée de Radon-Nikodym de la loi de X_0 par rapport à celle de ε_0, ce qui permet d'utiliser la méthode du maximum de vraisemblance pour estimer ρ.

Bosq et Mourid (1999) ont donné des conditions sur l'équivalence des lois de ε_0 et X_0, en utilisant le résultat de Kuelbs (1970), qui permet de plonger avec densité un espace de Banach séparable B dans un espace de Hilbert H, comme suit

Lemme 1.4.2. *Soit $(B, \|.\|_B)$ un espace de Banach séparable.*

Il existe un produit scalaire noté $< \cdot, \cdot >$ sur B de norme induite, notée $\|.\|$, plus faible que la norme de B.

De plus si on note H le complété de B pour la norme $\|.\|$, alors $\mathcal{A}_B = B \cap \mathcal{A}_H$ où $\mathcal{A}_B$ et $\mathcal{A}_H$ désignent les tribus boréliennes de B et de H respectivement.

Soit P_{X_0} et P_{ε_0} les lois gaussiennes de X_0 et ε_0. Soit H l'espace de Hilbert associé à B par le lemme 1.4.2, on a alors :

Théorème 1.4.1.

$$P_{X_0} \backsim P_{\varepsilon_0} \Leftrightarrow \begin{cases} C^{-1}\rho C\rho^* \, \textit{existe et est de Hilbert-Schmidt} \\ 1 \notin \; \textit{spectre}(C^{-1}\rho C\rho^*) \end{cases}$$

Dans ce cas la densité peut s'écrire pour $x \in H$:

$$f(x) = \frac{dP_{\varepsilon_0}(x)}{dP_{X_0}(x)} = \frac{1}{\prod_{i=1}^{\infty}(1-\lambda_i)^{\frac{1}{2}}} \exp\left(\frac{-1}{2} \sum_{i=1}^{\infty} \left\{ \frac{< x, v_i >_H^2}{1-\lambda_i} - < x, v_i >_H^2 \right\} \right) \; P_{\varepsilon_0} \; p.s.$$

où $(v_i, \lambda_i)_{i\in\mathbb{N}}$ sont les éléments propres de l'opérateur $C^{-1}\rho C\rho^*$.

1.5 Estimation de l'opérateur d'un ARC(1)

De nombreux processus à temps continu admettent une représentation autorégressifs dans $C[0,1]$. Pumo (1995) a étudié le modèle (1.5) dans $C[0,1]$, avec ρ un opérateur intégral à noyau, en le considérant naturellement dans $L^2[0,1]$.

Soit (X_n) un processus à valeurs dans $C\,[0,1]$ vérifiant (1.5). On dira que (X_n) est un ARC(1). Soit ρ un opérateur intégral à noyau défini par

$$\rho(f)(t) = \int_0^1 K(s,t)f(s)\mu(ds) \qquad f \in C\,[0,1] \tag{1.7}$$

tel que $\|K\|_{(C\,[0,1])^2} < 1$, donc $||\ \rho\ ||_{\mathcal{L}(\mathcal{C})} < 1$.

La technique est de définir un prolongement ρ' de ρ sur l'espace $H = L^2[0,1]$, muni de la mesure de Lebesgue, et tel que $\|\rho'\|_{\mathcal{L}(L^2)} < 1$.

Ce prolongement s'effectue à travers une base dans $L^2[0,1]$.

Soit $(e_j)_{j\in\mathbb{Z}}$ une base orthonormale dans $L^2[0,1]$. On définit les variables aléatoires suivantes :

$$X'_n := \sum_{j\geq 1}\langle X_n, e_j\rangle e_j, \qquad \varepsilon'_n := \sum_{j\geq 1}\langle \varepsilon_n, e_j\rangle e_j, \quad n \in \mathbb{Z}. \tag{1.8}$$

Ces variables sont à valeurs dans l'espace $L^2[0,1]$ et constituent des prolongements respectifs des variables X_n et $\varepsilon_n, n \in \mathbb{Z}$. Ainsi, pour $H = L^2[0,1]$ on associe, au processus ARC(1) le processus (X'_n) défini par :

$$X'_n = \rho' X'_{n-1} + \varepsilon'_n \qquad t \in \mathbb{Z}. \tag{1.9}$$

Par suite $(X'_n,\ n \in \mathbb{Z})$ est un processus $ARL^2(1)$.

L'estimation de ρ revient alors à celle de l'opérateur ρ'.

Soit $X_1,\dots,X_n$ des observations du processus auxquelles on associe, par (1.9), les observations $X'_1,\dots,X'_n$. Les opérateurs de covariance et de covariance croisée empiriques sont dans ce cas :

$$C_n(\cdot) = \frac{1}{n}\sum_{i=1}^{n} < X'_i, \cdot > X'_i \qquad et \qquad D_n(\cdot) = \frac{1}{n}\sum_{i=1}^{n-1} < X'_i, \cdot > X'_{i+1}.$$

Pour estimer l'opérateur on utilise la méthode proposée par Bosq. Sous les hypothèses (B_1) et (B_2) l'estimateur de ρ est :

$$\hat{\rho}_n(x) = \tilde{\pi}_{k_n} D_n \tilde{C}_n^{-1} \tilde{\pi}_{k_n}(x), \quad x \in C[0,1]$$

où $k_n \to \infty$, $\tilde{C}_n = \tilde{\pi}_{k_n} C_n$, et

$$\tilde{\pi}_{k_n}(x) = \sum_{j=1}^{k_n} < v_{jn}, x > v_{jn}, \quad x \in C[0,1]$$

$(\lambda_{jn}, v_{jn})_{k \geq 1}$ sont les éléments propres de C_n.

Pumo (1995) a montré la convergence p.s. de $\hat{\rho}_n$, en norme linéaire dans $\mathcal{L}(C[0,1])$, en considérant les hypothèses suivantes

$\mathbf{C_1}$: X_0 est borné et satisfait la condition de Hölder

$$|X_0(t) - X_0(s)| \leq M_0 |t - s|^{\alpha},\ 0 \leq s, t \leq 1,$$

où M_0 est une variable aléatoire réelle bornée et $0 < \alpha \leq 1$.

$\mathbf{C_2}$: (a) $v = \sup_{j \geq 1} \|v_j\|_{C([0,1]} < \infty$

et

(b) $\sup_{\|x\|_{C[0,1]} \leq 1} \|\rho(x) - \sum_{j=1}^{k} < \rho(x), v_j > v_j\|_{C[0,1]} \underset{k\to\infty}{\to} 0.$

Pour la convergence de $\hat{\rho}_n$ on a le théorème suivant :

Théorème 1.5.1. *Pumo (1995) : Soit* $(X_n,\ n \in \mathbb{Z})$ *un ARC(1) standard et* ρ *un opérateur à noyau* K *tel que* $\|K_{(C[0,1])^2}\| < 1$*. On suppose que* (B_1), (B_2), (C_1) *et* (C_2) *sont vérifiées. Alors si*

$$\frac{n\lambda_{k_n}^2}{\log n(\sum_{j=1}^{k_n} a_j)^2} \to 0$$

on a :

$$\|\hat{\rho}_n - \rho\|_{\mathcal{L}(C[0,1])} \longrightarrow 0 \qquad p.s.$$

Labbas et Mourid (2002) ont justifié le choix de Pumo (1995), qui consiste à prolonger les (X_i) dans $L^2([0,1])$, à travers le lemme (1.4.2) de Kuelbs (1970). En effet dans le cas général ce lemme permet de transformer le problème de l'estimation de ρ d'un ARB(1) à un ARH(1).

Soit H l'espace de Hilbert associé à B par le lemme (1.4.2). Soit $(e_j)_{j\in\mathbb{Z}}$ une base orthonormale dans H.

Si l'opérateur ρ se prolonge en un opérateur linéaire borné ρ' défini sur H et vérifiant $\|\rho'^{j_0}\|_{\mathcal{L}(H)} < 1$, le processus défini par :

$$X'_n = \rho' X'_{n-1} + \varepsilon'_n \qquad t \in \mathbb{Z}, \tag{1.10}$$

est un ARH(1), où

$X'_n := \sum_{j\geq 0}\langle X_n, e_j\rangle e_j, \qquad \varepsilon'_n := \sum_{j\geq 0}\langle \varepsilon_n, e_j\rangle e_j, \quad n \in \mathbb{Z}.$

L'estimation de ρ revient donc à celle de ρ'.

Bensmain et Mourid (2001) ont étudié ce cas en utilisant la méthode des cribles. Comme (X_n) s'écrit :

$$X_n(t) = \int_0^1 K(t-s)X_{n-1}(s)ds + \varepsilon_n(t), \quad t \in [0,1] \tag{1.11}$$

ou encore

$$X_n(t) = (K * X_{n-1})(t) + \varepsilon_n(t), \tag{1.12}$$

l'estimation de ρ revient donc à celle de son noyau K.

On suppose que K est pair et on considère la base trigonométrique dans $L^2[0,1]$:

$(e_0 = 1_{[0,1]}\,;\, e_{2k}(t) = \sqrt{2}\cos(2\pi kt),\, e_{2k+1}(t) = \sqrt{2}\sin(2\pi kt),\, k \geq 1)$.

On note $a_k(X_n), a_k(X_{n-1}), a_k(\varepsilon_n), a_k(K),\ k \geq 0$, les coefficients de Fourier par rapport au cosinus et par $b_k(X_n),\ b_k(X_{n-1}),\ b_k(\varepsilon_n)$ ceux par rapport au sinus, des fonctions X_n, X_{n-1} et ε_n. On suppose que les variables aléatoires réelles $a_k(\varepsilon_n)$ et $b_k(\varepsilon_n)$ sont gaussiennes indépendantes et de même variance $\sigma_k^2,\ k \geq 0$.

En tenant compte de (1.12) on obtient les relations suivantes pour $k \geq 1$:

$$\begin{cases} a_k(X_n) = (a_k(K)a_k(X_{n-1}))/2 + a_k(\varepsilon_n) \\ b_k(X_n) = (a_k(K)b_k(X_{n-1}))/2 + b_k(\varepsilon_n) \end{cases} \tag{1.13}$$

et pour $k = 0$,

$$a_0(X_n) = a_0(K)a_0(X_{n-1}) + a_0(\varepsilon_n).$$

En notant $x_{n,k} = a_k(X_n),\ x_{n-1,k} = a_k(X_{n-1}),\ \varepsilon_{n,k} = a_k(\varepsilon_n)$ et $c_k = a_k(K)$, la première relation de (1.13) s'écrit alors :

$$x_{n,k} = \frac{1}{2}c_k x_{n-1,k} + \varepsilon_{n,k}, \quad k \geq 1, \tag{1.14}$$

pour $k = 0$ nous avons : $x_{n,0} = c_0 x_{n-1,0} + \varepsilon_{n,0}$. Ainsi on obtient deux processus autorégressifs réels d'ordre 1. L'estimation du noyau K revient donc à estimer ses coefficients de Fourier $c_k = a_k(K),\ k \geq 0$.

On choisit le crible $\{\Theta_{m_n}\}$ sous la forme

$$\Theta_{m_n} = \{K \in L^2 \ /\ K(t) = c_0 1_{[0,1]}(t) + \sum_{k=1}^{m_n} c_k \sqrt{2}\cos(2\pi kt),\ t \in [0,1], \quad \sum_{k=1}^{m_n} k^2 c_k^2 \leq m_n\} \tag{1.15}$$

où K est 1-périodique et $m_n \to +\infty$ quand $n \to +\infty$.

L'estimateur crible du noyau K pour des observations $(X_0, X_1, ..., X_n)$ vérifiant (1.5) est défini dans la proposition suivante

Proposition 1.5.1. *Les coefficients de Fourier de l'estimateur crible $\hat{K}_n$ du noyau K sont :*

$$\hat{c}_0 = \frac{\sum_{i=1}^{n} x_{i,0}x_{i-1,0}}{\sum_{i=1}^{n} x_{i-1,k}^2}, \quad \hat{c}_k = \frac{\sum_{i=1}^{n} x_{i,k}x_{i-1,k}}{\sum_{i=1}^{n} \frac{1}{2}x_{i-1,k}^2 + n2\lambda k}, \qquad k = 1, ..., m_n$$

où $m_n \to +\infty$ quand $n \to +\infty$ et λ vérifie

$$\sum_{k=1}^{m_n} k^2 \left(\frac{\sum_{i=1}^{n} x_{i,k}x_{i-1,k}}{\sum_{i=1}^{n} \frac{1}{2}x_{i-1,k}^2 + n2\lambda k} \right)^2 = m_n$$

Bensmain et Mourid (2001) ont montré la convergence p.s. en norme L^2 de K avec une vitesse de convergence polynomiale de m_n en appliquant le théorème (1.3.2).

Soit K_0 la vraie valeur de K.

Théorème 1.5.2. *Si $m_n = O(n^{\frac{1}{3}-\delta})$ pour $0 < \delta < \frac{1}{3}$, alors*

$$\|\hat{K}_n - K_0\| \xrightarrow[n\to+\infty]{p.s.} 0$$

dans $L^2([0,1], \mathcal{B}_{[0,1]}, \mu)$ où μ est la mesure de Lebesgue.

Chapitre 2

Vitesse de convergence en norme $p-$intégrale et normalité asymptotique de l'estimateur crible de l'opérateur d'un ARB(1)

Ce travail a fait l'objet d'une publication aux Comptes Rendus Mathématique, Académie des Sciences, Paris (Rachedi (2005)), et une aux Annales de l'ISUP (Rachedi (2004)).

2.1 Introduction

Toutes les variables aléatoires considérées dans ce chapitre sont définies sur le même espace de probabilité $(\Omega, \mathcal{A}, \mathbb{P})$. $(B, \mathcal{B})$ est un espace de Banach séparable réel muni de sa tribu borélienne et de sa norme $\| \ \|$. B^* désigne le dual topologique de B et $(\cdot, \cdot)$ le crochet de dualité entre B^* et 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 $\| \cdot \|_{\mathcal{L}}$.

Un processus autorégressif faiblement stationnaire d'ordre 1 dans un espace de Banach ou WARB(1) est une suite $X = (X_t, t \in \mathbb{Z})$ de variables aléatoires à valeurs dans B telle que :

$$X_t = \rho(X_{t-1}) + \varepsilon_t, \quad t \in \mathbb{Z} \tag{2.1}$$

où $\varepsilon = (\varepsilon_t, t \in \mathbb{Z})$ est un bruit blanc faible dans B et $\rho \in \mathcal{L}(B)$ vérifie $\|\rho\|^{j_0} < 1$ pour un $j_0 \geq 1$.

Dans la suite nous utilisons la notation ρX_{t-1} au lieu de $\rho(X_{t-1})$.

Le problème de prévision associé à un ARB(1) est important, il passe par l'estimation de l'opérateur d'autocorrélation ρ définissant la structure autorégressive du processus. Nous montrons la consistance de l'opérateur d'autocorrélation ρ en norme $p-$intégrale, les résultats qui existent sont établis pour ρ un opérateur $2-$intégrale. Notre résultat se distingue ainsi et ouvre un champ plus vaste pour les applications.

Nous définissons la méthode des moindres carrées comme problème d'optimisation dans un espace de Banach. Cette définition est possible lorsque qu'on peut munir l'espace de Banach d'une relation d'ordre. Nous définissons l'estimateur des moindres carrés de ρ en considérant l'ordre partiel induit par le cône des opérateurs définis positifs de l'espace des opérateurs nucléaires. Nous trouvons la solution de ce type de problèmes en utilisant la dérivée de Fréchet. Pour trouver une forme explicite de cet estimateur nous avons considéré ρ un opérateur strictement p-intégral, et une base dans B dite base de shrinking, dont on rappelle la définition par la suite. Nous montrons, sous certaines hypothèses relatives au biais, que la vitesse de convergence est de l'ordre de $n^{\frac{1}{min(2,p)}} (\log n)^{\frac{-\beta}{min(2,p)}}$, $\beta > 1$. La géométrie de l'espace de Banach joue un rôle important dans cette étude. Nous montrons aussi la normalité asymptotique de l'estimateur. Enfin nous illustrons ce résultat par une simulation d'un ARB(1) gaussien, avec le logiciel STATISTICA.

2.2 Rappel

2.2.1 Bases dans un espace de Banach

Nous commençons par quelques définitions qui seront utilisées dans la suite, on peut les trouvées par exemple dans Lindenstrauss et Tzafriri (1977), Diestel (1984), Kadets et Kadets (1991).

a) Une suite $(x_k)_k$ dans B est dite base de Schauder si $\forall x \in B$, $\exists(\gamma_k)$ une suite unique de scalaires tel que $x = \sum_{k\geq 0} \gamma_k x_k$.

b) Soit B un espace de Banach admettant une base de Schauder $(x_k)_k$. Une fonctionnelle de coefficients est une application $x_k^* : B \to \mathbb{R}$ définie pour tout $x = \sum\limits_{k\geq 1} \alpha_k x_k$ dans B par $x_k^*(x) = \alpha_k$, $\forall k$.

c) Si la suite $(x_k^*)_k$ des fonctionnelles de coefficients forme une base dans B^* , $(x_k)_k$ est dite base de **shrinking**. Soit B un espace de Banach tel que B^* admet une base $(x_k^*)_k$ alors $(x_k^*)_k$ est complétement bornée et B admet une base de shrinking

d) Une suite $(y_k^*)_k$ dans B^* est dite **suite basique faible** s'il existe une suite $(x_k)_k$ dans B tels que $(y_k^*, x_l) = \delta_{kl}$. Le couple $(x_k, y_k^*)_k$ est dit système **biorthogonal**.

e) Soit $[y_k^*]_k$ l'espace engendré par $(y_k^*)_k$, on peut écrire $\forall x^* \in [y_k^*]_k$, $x^* = \lim\limits_{k\to+\infty} \sum\limits_{i=1}^{k} (x^*, x_i)\, y_i^*$.

Un système biorthogonal $(x_k^*, x_k)_k$ dans un espace de Banach B , est dit base de **Markushevich** si $[x_k]_k$ est dense dans B et $[x_k^*]_k$ est faiblement dense dans B^*. Si B est un espace de Banach séparable et $\varepsilon > 0$, alors il existe une base de Markushevich $(x_k^*, x_k)_k$ dans B pour laquelle $\|x_k\| \cdot \|x_k^*\| \leq 1 + \varepsilon \quad \forall k$ (Pełczyński (1976)).

2.2.2 Opérateur $p-$sommable

Un opérateur ρ de $\mathcal{L}(B)$ est $p-$sommable, $p \in]1, +\infty[$, s'il existe une constante $C \geq 0$ tel que $\forall x_1, \ldots, x_n$ dans B on a

$$\left(\sum_{i=1}^{n} \|\rho(x_i)\|^p\right)^{\frac{1}{p}} \leq C \cdot \sup_{\|x^*\|\leq 1} \left(\sum_{i=1}^{n} |(x^*, x_i)|^p\right)^{\frac{1}{p}}$$

La plus petite valeur C pour laquelle cette inégalité est vérifiée est notée $\pi_p(\rho)$.

L'ensemble des opérateurs $p-$sommables de B dans B est noté $\Pi_p(B)$. C'est un sous espace vectoriel de $\mathcal{L}(B)$ et π_p définit une norme dans $\Pi_p(B)$ telle que pour tout $\rho \in \Pi_p(B)$ on a : $\|\rho\|_{\mathcal{L}} \leq \pi_p(\rho)$.

L'espace $\Pi_p(B)$ muni de la norme π_p est un espace de Banach (Diestel *et al.* (1995) p. 38).

Si B est un espace de Hilbert, $\Pi_2(B)$ coïncide avec l'espace des opérateurs de Hilbert-Schmidt et π_2 est la norme de Hilbert-Schmidt.

2.2.3 Opérateur strictement $p-$intégral

Un opérateur ρ de $\mathcal{L}(B)$ est strictement $p-$intégral s'il existe un espace de probabilité $(\Lambda, \mathcal{E}, \mu)$ et deux opérateurs linéaires bornés c et d de $L^p(\mu)$ dans B et de B dans $L^\infty(\mu)$ respectivement tel que le diagramme suivant commute :

$$\begin{array}{ccc} B & \xrightarrow{\rho} & B \\ d\downarrow & & \uparrow c \\ L^\infty(\mu) & \xrightarrow{i_p} & L^p(\mu) \end{array} \tag{2.2}$$

où i_p est l'identité formelle de $L^\infty(\mu)$ dans $L^p(\mu)$. On note $\mathcal{J}_p(B)$ l'ensemble des opérateurs strictement $p-$intégraux de B dans B et on associe la norme $p-$intégrale, notée $\|\cdot\|_p$ et définie par $\|\rho\|_p = \inf\limits_{d,c,\mu} \|d\|_{\mathcal{L}} \cdot \|c\|_{\mathcal{L}}$. ρ est $p-$sommable tel que $\pi_p(\rho) = \|\rho\|_p$, on a alors $\|\rho\|_{\mathcal{L}} \leq \|\rho\|_p$.

Soit $(\Gamma, \mathcal{F}, \nu)$ un espace mesurable. Toute fonction $\xi \in L^p(\nu)$ induit un opérateur de multiplication noté M_ξ de $L^\infty(\nu)$ dans $L^p(\nu)$ défini par : $f \to \xi f$ tel que $\|M_\xi\|_{\mathcal{L}} \leq \|\xi\|_{L_p(\nu)}$. On note $\mathcal{M}_p(L^\infty(\nu), L^p(\nu))$ l'ensemble des opérateurs de multiplication de $L^\infty(\nu)$ dans $L^p(\nu)$ et $(\cdot,\cdot)$ le crochet de dualité entre $L^p(\nu)$ et son dual topologique.

ρ est strictement $p-$intégral si et seulement s'il existe un espace mesurable $(\Gamma, \mathcal{F}, \nu)$, deux opérateurs linéaires bornés a et b de $L^p(\nu)$ dans B et de B dans $L^\infty(\nu)$ respectivement, et un opérateur $M_\xi \in \mathcal{M}_p(L^\infty(\nu), L^p(\nu))$ (ou $\xi \in L^p(\nu)$) tel que le diagramme suivant commute :

$$\begin{array}{ccc} B & \xrightarrow{\rho} & B \\ b\downarrow & & \uparrow a \\ L^\infty(\nu) & \xrightarrow{M_\xi} & L^p(\nu) \end{array} \tag{2.3}$$

On note $\mathcal{J}_p(B)$ l'ensemble des opérateurs strictement $p-$intégraux de B dans B et on lui associe la norme $p-$intégrale, notée $\|\cdot\|_p$ et définie par $\|\rho\|_p = \inf\limits_{a,b,\xi} \|b\|_{\mathcal{L}} \cdot \|M_\xi\|_{L^p} \cdot \|a\|_{\mathcal{L}}$ (Diestel *et al.* (1995) p. 111).

2.2.4 Opérateur p-nucléaire

Soit ρ un opérateur strictement $p-$intégral tel que la mesure ν associée au diagramme (2.3) est une mesure de dénombrement sur $\mathbb{N}$, dans ce cas ρ est dit $p-$nucléaire.

On note $\mathcal{D}(\ell^\infty, \ell^p)$ l'ensemble des opérateurs diagonaux de ℓ^∞ dans ℓ^p. ρ est $p-$nucléaire si et seulement s'il existe deux opérateurs linéaires bornés a et b de ℓ^p dans B et de B dans ℓ^∞ respectivement, et un opérateur $D_\alpha \in \mathcal{D}(\ell^\infty, \ell^p)$ (ou $\alpha \in \ell^p$) tel que le diagramme suivant commute :

$$\begin{array}{ccc} B & \xrightarrow{\rho} & B \\ b\downarrow & & \uparrow a \\ \ell^\infty & \xrightarrow{D_\alpha} & \ell^p \end{array} \tag{2.4}$$

la norme p-nucléaire est $\eta_p(\cdot)$ définie par $\eta_p(\rho) = \inf\limits_{b,a,D_\alpha} (\|b\|_{\mathcal{L}} \cdot \|a\|_{\mathcal{L}} \cdot \|D_\alpha\|_{\mathcal{L}})$. L'ensemble des opérateurs $p-$nucléaires de B dans B, qu'on note $\mathcal{N}_p(B)$ est un espace de Banach.

2.3 Estimateur crible des moindres carrées de ρ

Nous considérons maintenant le problème de l'estimation de ρ d'un WARB(1) dans le cas où ρ un opérateur p-sommable ($p > 1$), qui est comme nous l'avons vu une généralisation naturelle d'un opérateur de Hilbert-Schmidt dans un espace de Hilbert.

Soit $(X_t, t \in \mathbb{Z})$ un processus vérifiant (2.1). On pose $\Theta = \Pi_p(B)$, l'espace des opérateurs $p-$sommables, et la distance associée d est celle déduite de la norme. On associe à Θ un crible $(\Theta_m)_m$ (cf. défintion 1.3.1).

2.3.1 Définition de l'estimateur crible des moindres carrées de ρ dans un espace de Banach

Soit $(\mathcal{N}', \| \|_{\mathcal{N}'})$ l'espace des opérateurs nucléaires de B^* dans B, muni de la norme nucléaire. On considère K le cône des opérateurs définis positifs de $\mathcal{N}'$. On muni $\mathcal{N}'$ de la relation d'ordre $\preceq$ définie par : $\forall T_1, T_2 \in K$, $T_1 \preceq T_2$ si et seulement si $T_2 - T_1 \in K$. On note l'infimum dans $\mathcal{N}'$ par rapport à cet ordre par $\overset{\vee}{\inf}$ (voir appendice, paragraphe 3.8). Cet ordre nous permet de définir la méthode des moindres carrées comme problème d'optimisation dans un espace de Banach.

Soit C_ρ l'opérateur défini par

$$C_\rho = E\left((\cdot, X_1 - \rho X_0)(X_1 - \rho X_0)\right), \ \rho \in \Theta,$$

ou encore

$$C_\rho = C - \rho D^* - D\rho^* + \rho C \rho^*, \ \rho \in \Theta.$$

Nous utilisons la généralisation des problèmes d'optimisation réelle à des problèmes d'optimisation vectorielle donnée dans Pallaschke et Rolewicz (1997), pour poser le problème :

$$\overset{\vee}{\inf}_{\rho \in \Theta_m}(C_n - \rho D_n^* - D_n \rho^* + \rho C_n \rho^*). \tag{2.5}$$

On étudie la solution de ce type de problèmes en utilisant la dérivée de Fréchet. Si $\hat{\rho}_{n,m}$ est solution du problème 2.5, d'après la proposition 3.8.1 (appendice), il existe $\ell' \in K^+$, où K^+ l'ensemble des fonctionnelles positives sur K, tel que $\widehat{\rho}_{n,m}$ vérifie :

$$\ell'\left(\lim_{t\downarrow 0}\left(\frac{C_{(\widehat{\rho}_{n,m}+th),n} - C_{\rho,n}}{t}\right)\right) = 0 \tag{2.6}$$

où $h \in \Theta$ et la limite est uniforme sur tout sous-ensemble de la sphère unité de Θ.

D'où

$$\ell'\left(\lim_{t\downarrow 0}\frac{-tD_n h^* - thD_n^* + thC_n\widehat{\rho}_{n,m}^* + t\widehat{\rho}_{n,m}C_n h^* + t^2 hC_n h^*}{t}\right) = 0 \tag{2.7}$$

Alors

$$\ell'\left((\widehat{\rho}_{n,m}C_n - D_n)h^* - h(C_n\widehat{\rho}^*_{n,m} - D_n)\right) = 0 \tag{2.8}$$

Comme ℓ' est linéaire positive et h est quelconque sur tout sous-ensemble de la sphère unité de Θ, alors l'estimateur crible $\widehat{\rho}_{n,m}$ vérifie $\widehat{\rho}_{n,m}C_n = D_n$, c'est la même propriété que celle (1.2) de l'opérateur ρ.

2.4 Décomposition de ρ et existence de l'estimateur crible

Dans le but de trouver une forme explicite de l'estimateur crible nous considérons une représentation "spectrale" de ρ associée à une base dans l'espace de Banach B. Cette décomposition est possible pour un opérateur strictement $p-$intégral, qui est un opérateur $p-$sommable.

Soit ρ un opérateur strictement $p-$intégral tel que $\rho = aM_\xi b$, la décomposition associée au diagramme (2.3). Dans toute la suite $p \in]1, \infty[$ et q est tel que $\dfrac{1}{p} + \dfrac{1}{q} = 1$.

Soit $\left(e^*_{k}, e_{k}\right)_{k\geq 0}$ une base de Markushevich dans $L^p\left(\nu\right)$. La suite $\left(ae_{k}\right)_{k\geq 0}$ est une base dans $Im\left(\rho\right)$ et si a est injective c'est une base de Schauder. Pour obtenir une décomposition de ρ nous supposons que :

H : La suite $\left(ae_{k}\right)_{k\in \mathrm{N}}$ est une base de shrinking dans B.

Notant $\left(f^*_{k}\right)_{k\in \mathrm{N}}$ la suite des fonctionnelles de coefficients associée à $(ae_k)_{k\geq 0}$. Nous avons le lemme :

Lemme 2.4.1. *Sous l'hypothèse* **H***, l'opérateur ρ admet la décomposition :*

$$\rho\left(\cdot\right) = \sum_{k\geq 0} \alpha_k (e^*_{k}, e_k b\left(\cdot\right))\ ae_{k} \tag{2.9}$$

où $\alpha_k = \left(e^*_{k}, \xi\right),\quad \forall k \geq 0,$ et $\xi \in L^p\left(\nu\right)$ est associée à ρ par le diagramme (2.3).

Soit la suite des opérateurs $\rho_k(\cdot) = (e_k^*, e_k b(\cdot))\, a e_k \quad k \geq 0,\ \rho_k$ est un opérateur de rang 1. On pose $\rho_N = \sum\limits_{k=0}^{N} \alpha_k\ \rho_k$, alors ρ_N converge vers $\rho = \sum\limits_{k\geq 0} \alpha_k\ \rho_k$ par rapport à la norme strictement $p-$intégrale donc linéaire.

Donc sous l'hypothèse **H**, l'opérateur ρ s'écrit :

$$\rho = \sum_{k\geq 0} \alpha_k\ \rho_k. \tag{2.10}$$

Nous posons

$$\Theta = \left\{ \rho = \sum_{k\geq 0} \alpha_k\ \rho_k\ / \quad (\alpha_k)_k \in \ell^p \right\},$$

et nous considérons dans Θ la norme $\|\rho\|_p = (\sum\limits_{k\geq 0} |\alpha_k|^p)^{\frac{1}{p}}$, qui est équivalente à la norme $p-$intégrale.

On note Θ^* le dual topologique de Θ, $\Theta^* = \left\{ \rho = \sum\limits_{k\geq 0} \beta_k\ \rho_k\ / \quad (\beta_k)_k \in \ell^q \right\}$.

Ainsi si les opérateurs ρ_k sont connus, $\forall k \geq 0$, l'estimation de ρ revient à l'estimation des $(\alpha_k)_k$ dans ℓ^p. Le fait de considérer que (ρ_k) ou a et b sont connus, n'est pas très restrictif. Voici un exemple :

Exemple 2.4.1. *Soient $B = C([0,1])$ et $\mathcal{M}_p(B)$ l'ensemble des opérateurs de multiplication de B dans $L^p([0,1])$. Toute fonction $\xi \in C([0,1])$ induit un opérateur de multiplication $M_\xi \in \mathcal{M}_p(C([0,1]))$ d'image dans $L^p([0,1])$.*

Soit $\rho \in \mathcal{M}_p(C([0,1]))$ défini par $\xi \in C([0,1])$, ρ intervient dans la représentation de certains processus autorégressif à temps continu, voir par exemple la proposition (1.4.1). Soit μ une mesure de probabilité sur $[0,1]$ de densité $\dfrac{d\mu(t)}{dt} = \left(\dfrac{\xi}{\|\xi\|}\right)^p$, on obtient la décomposition de ρ du diagramme (2.2), où d est l'injection de $C([0,1])$ dans $L^\infty(\mu)$ et c l'application de $L^p(\mu) \to L^p([0,1]) : f \mapsto \xi f$. Donc ρ est strictement $p-$ intégral et $\|\rho\|_p = \|\xi\|_{L^p([0,1])}$.

Les opérateurs a et b dans le diagramme (2.3) sont dans ce cas les injections de $L^p([0,1])$ *dans* $C([0,1])$ *et de* $C([0,1])$ *dans* $L^\infty([0,1])$ *respectivement.*

Nous allons utiliser maintenant des cribles similaires à ceux proposés par Beder (1988) pour un paramètre dans ℓ^2. Soit le crible :

$$\Theta_m = \left\{ \rho = \sum_{k\geq 0} \alpha_k \ \rho_k \ / \ \ \alpha_k = 0 \ , \ k > m \right\}, \qquad m \geq 0, \ m = m(n) \underset{n\to+\infty}{\longrightarrow} +\infty$$

Les variables aléatoires réelles $(f_k^*, \varepsilon_n)_n$ sont indépendantes, de même variance σ_k^2, $\forall k \geq 0$. Nous supposons que :

$$\sigma \ = \ \sum_{k\geq 0} \sigma_k^p < \infty. \tag{2.11}$$

Alors l'estimateur crible de ρ par la méthode des moindres carrées peut être défini, pour des observations $(X_0, X_1, ..., X_n)$, comme solution de l'équation suivante

$$\min_{\rho\in\Theta_m} \sum_{k=0}^{m} (\sum_{i=1}^{n} (f_k^*, X_i - \rho X_{i-1})^2)^{\frac{p}{2}}$$

La solution est donnée dans la proposition :

Proposition 2.4.1. *Si* $(C_n\rho_k^* f_k^*, \rho_k^* f_k^*) > 0$ *pour* $k = 0, \ldots, m$; *l'estimateur crible des moindres carrées de* ρ *est l'opérateur* $\widehat{\rho}_{n,m} = \sum\limits_{k=0}^{m} \widehat{\alpha}_{k,n} \ \rho_k$

$$\text{où } \widehat{\alpha}_{k,n} = \frac{(D_n\rho_k^* f_k^*, f_k^*)}{(C_n\rho_k^* f_k^*, \rho_k^* f_k^*)} \qquad k = 0, \ldots, m.$$

2.5 Convergence de l'estimateur crible

Dans ce paragraphe nous montrons la convergence p.s. de l'estimateur $\widehat{\rho}_{n,m}$, sa normalité asymptotique en utilisant les propriétés des espaces ℓ^p et ceux d'une différence de martingale.

Nous considérons d'abord les hypothèses et les notations suivantes

1. Nous posons $c_k = (C\rho_k^* f_k^*, \rho_k^* f_k^*)$, nous avons alors $c_k \leq \|b^*\|^2 \, \|C\|$, $\forall k \geq 0$.

   Nous supposons que

$$\sigma \;=\; (\sum_{k\geq 0} \frac{\sigma_k^p}{c_k^p \mathbb{I}_{\{c_k>0\}}})^{\frac{1}{p}} < \infty. \tag{2.12}$$

   Cette hypothèse implique (2.11).

2. Soit $F_i = (\cdot, X_{i-1})\,\varepsilon_i$, $i \in \mathbb{Z}$, $E(F_i) = E^{\mathcal{B}_{i-1}}(F_i) = 0$, $i \in \mathbb{Z}$.

3. Nous posons $E_i = \left(\dfrac{(F_i\rho_k^* f_k^*, f_k^*)}{c_k \mathbb{I}_{\{c_k>0\}}}\right)_{k\geq 0}$, suite à l'hypothèse (2.12), E_i est une variable aléatoire à valeurs dans $\left(\ell^p, \|\cdot\|_p\right)$ où $\|\cdot\|_p$ désigne la norme usuelle de ℓ^p, $\forall i \in \mathbb{Z}$.

4. $\overline{E}_n = \dfrac{1}{n}\sum\limits_{i=1}^{n} E_i$ et $C_{E_1}(x^*) = E\left((x^*, E_1)\,E_1\right)$, $x^* \in B^*$.

5. Π_m la projection de Θ sur Θ_m.

Comme F_i est une différence de martingales dans B alors E_i est une différence de martingales dans ℓ^p. Dans la proposition suivante nous donnons la vitesse de convergence p.s. de $\Pi_m\overline{E}_n$ et sa normalité asymptotique, qui vont permettre par la suite de déduire ceux de $\widehat{\rho}_{n,m}$.

Soient les conditions :

(C_1) : Pour $p > 2$, ε_t est une variable aléatoire pré-gaussienne, $t \in \mathbb{Z}$,

(C_2) : $E(\exp \gamma \|E_1\|_p) < \infty$ pour un $\gamma > 0$,

(C_3) : $\|X_t\| \leq c$, où c est une constante,

(C_4) : $\sup\limits_{i\geq 1} E \|E_i\|_p^{2+\delta} < \infty$ pour un $\delta > 0$.

Nous avons alors :

Proposition 2.5.1. *1. Si (C_1) est vérifiée alors pour tout choix de $m = m(n)$ on a :*

$$n^{\frac{1}{min(2,p)}} (\log n)^{-\frac{\beta}{min(2,p)}} \left\|\Pi_m \overline{E}_n\right\|_{max(p,2)} \xrightarrow[n\to+\infty]{p.s.} 0 \quad \forall \beta > 1. \tag{2.13}$$

2. *Si* (C_2) *et* (C_3) *sont vérifiées alors* $\forall\eta > 0$ *:*

$$P\left(\left\|\Pi_m \overline{E}_n\right\|_p > \eta\right) \quad \leq \quad 2\exp\left(-\frac{n\eta^2}{8nl^2 + 4L\eta}\right) \tag{2.14}$$

où $l > 0$ *et* $L > 0$ *sont deux constantes.*

De plus pour $p = 2$ *:*

$$n^{-\frac{1}{2}}\left(\log\log n\right)^{-\frac{1}{2}}\left\|\Pi_m \overline{E}_n\right\|_2 \quad \underset{n\to+\infty}{\overset{p.s.}{\longrightarrow}} \quad 0. \tag{2.15}$$

3. *Si* (C_4) *est vérifiée alors* $\forall u^* \in \ell^q$ *:*

$$\sqrt{n}\ (u^*, \Pi_m \overline{E}_n) \quad \overset{\mathcal{D}}{\longrightarrow} \quad N \sim \mathcal{N}\left(0, E(u^*, E_1)^2\right). \tag{2.16}$$

Dans le cas particulier $p = 2$ *on a :*

$$\sqrt{n}\ \Pi_m \overline{E}_n \quad \overset{\mathcal{D}}{\longrightarrow} \quad N \sim \mathcal{N}\left(0, C_{E_1}\right), \tag{2.17}$$

où $\mathcal{D}$ *désigne la convergence en loi dans* ℓ^2.

Remarque 2.5.1. *La condition* (C_1) *n'est pas nécessaire dans le cas* $p = 2$.

Nous montrons maintenant la convergence p.s. de l'estimateur crible en utilisant la norme de ces coefficient $(\hat{\alpha}_k)$ munie de poids empiriques, cette norme nous donne la possibilité de transformer le problème dans ℓ^p, de plus elle est asymptotiquement équivalente à la norme $p-$intégrale.

Pour définir cette norme nous supposons que $c_k > 0$, pour $k = 0, \ldots, m$ et nous posons :

1. $w_{k,n} = \dfrac{(C_n \rho_k^* f_k^*, \rho_k^* f_k^*)}{c_k}$, $k = 0, \ldots, m\,;$
2. $\lambda_m = \min\limits_{k=0,\ldots,m} c_k$,

Ainsi nous pouvons munir Θ_m de la norme empirique :

$$\|\Pi_m \rho\|_{n,p} \quad = \quad (\sum_{k=0}^{m} w_{k,n}^p |\alpha_k|^p)^{\frac{1}{p}}, \quad \forall \rho \in \Theta.$$

Nous avons alors :

Lemme 2.5.1. *Si*

i) $E\left\|X_t\right\|^4 < \infty, \ t \in \mathbb{Z},$

ii) $n^{1/2} m_n^{1/2} \lambda_{m_n} (\log n)^{-\beta} \underset{n\to+\infty}{\longrightarrow} +\infty, \ \beta > \dfrac{1}{2},$

alors :

$$\forall \rho \in \Theta, \quad n^{1/2} m_n^{1/2} \lambda_{m_n} (\log n)^{-\beta} \left| \|\Pi_{m_n}\rho\|_{n,p}^p - \|\Pi_{m_n}\rho\|_p^p \right| \underset{n\to+\infty}{\overset{p.s.}{\longrightarrow}} 0, \qquad \forall \beta > \frac{1}{2}.$$

Donc $\widehat{\rho}_{n,m}$ converge par rapport à la norme strictement p-intégrale dès qu'il converge par rapport à la norme empirique.

Soit $\rho_0 = \sum_{k\geq 0} \alpha_{0,k}\rho_k$ la vraie valeur de ρ. Nous posons $\rho_m = \Pi_m \rho_0$.

Dans le théorème suivant nous montrons que la vitesse de convergence de $\widehat{\rho}_{n,m}$ est au moins égale à $n^{\frac{1}{min(2,p)}} (\log n)^{-\frac{\beta}{min(2,p)}}$, par rapport à la norme empirique, pour tout choix de $m = m(n)$. Ainsi on peut choisir $m = m(n)$ tel que le biais $\|\mathbb{I} - \Pi_m\rho_0\|_p = \left(\sum_{k>m(n)} \alpha_k^p\right)^{\frac{1}{p}}$ est de l'ordre de $n^{-\frac{1}{min(2,p)}} (\log n)^{\frac{\beta}{min(2,p)}}$.

Théorème 2.5.1. *1. Si (C_1) est vérifiée alors pour tout choix de $m = m(n)$ on a :*

$$n^{\frac{1}{min(2,p)}} (\log n)^{\frac{-\beta}{min(2,p)}} \|\widehat{\rho}_{n,m} - \rho_m\|_{n,max(p,2)} \underset{n\to+\infty}{\overset{p.s.}{\longrightarrow}} 0 \quad \forall \beta > 1. \tag{2.18}$$

2. Si (C_2) et (C_3) sont vérifiées alors pour tout $m = m(n)$ et $\forall \eta > 0$ on a :

$$P\left(\|\widehat{\rho}_{n,m} - \rho_m\|_{n,p} > \eta\right) \quad \leq \quad 2\exp\left(-\frac{n\eta^2}{24nK^2 + 16L\eta}\right), \tag{2.19}$$

où $K > 0$ et $L > 0$ sont deux constantes.

De plus pour $p = 2$:

$$n^{\frac{1}{2}} (\log\log n)^{-\frac{1}{2}} \|\widehat{\rho}_{n,m} - \rho_m\|_{n,max(p,2)} \xrightarrow[n\to+\infty]{p.s.} 0 \,. \tag{2.20}$$

3. *Si (C_4) est vérifiée et si $m_n^{1/2}\lambda_{m_n} (\log n)^{-\beta} \underset{n\to+\infty}{\longrightarrow} +\infty$, $\beta > \frac{1}{2}$, alors $\forall\theta^* \in \Theta^*$:*

$$\sqrt{n}\, (\theta^*, \widehat{\rho}_{n,m} - \rho_m) \xrightarrow{\mathcal{D}} N \sim \mathcal{N}\left(0, E(\zeta^*, E_1)^2\right), \tag{2.21}$$

où ζ^* est la suite dans ℓ^q associée à θ^*.

Dans le cas particulier $p = 2$ on a :

$$\sqrt{n}\, (\widehat{\rho}_{n,m} - \rho_m) \xrightarrow{\mathcal{D}} N \sim \mathcal{N}\left(0, C_{E_1}\right), \tag{2.22}$$

où $\mathcal{D}$ désigne la convergence en loi dans $\mathcal{J}_2(B)$.

Le corollaire suivant donne la vitesse de convergence de l'estimateur dans deux cas particuliers :

Corollaire 2.5.1. *Supposons que :*

i) *C_1 est vérifiée,*

ii) *$n^{1/2}m_n^{1/2}\lambda_{m_n} (\log n)^{-\beta} \underset{n\to+\infty}{\longrightarrow} +\infty$, $\beta > \frac{1}{2}$,*

on a pour :

1. *$\alpha_k \leq k^{-\tau}$, $\tau > 1/p$, si $m_n \simeq \left(\frac{n}{(\log n)^\beta}\right)^{\frac{p}{(\tau p-1)min(2,p)}}$,*

2. *et pour $\alpha_k \leq \omega^k$, $|\omega| < 1$, si $m_n \simeq \frac{c}{min(2,p)} \log\left(\frac{n}{(1-\omega^p)^{\frac{min(2,p)}{p}} (\log n)^\beta}\right)$, $c > 0$,*

   alors $\forall\beta > 1$:

$$\|\widehat{\rho}_{n,m} - \rho_0\|_{max(p,2)} = O\left(\left(\frac{(\log n)^\beta}{n}\right)^{\frac{1}{min(2,p)}}\right).$$

Remarque 2.5.2. *Dans Rachedi (2004), pour le cas $p = 2$, nous avons utiliser la norme de ℓ^2 pour les coéfficients $(\alpha_k)_k$ muni des poids des variances empiriques, cette norme nous a permis de transformer le problème de la convergence de l'estimateur crible à la convergence de la moyenne empirique de la différence de martingales $E_i' = ((F_i\rho_k^* f_k^*, f_k^*))_{k\geq 0}$, dans l'espace de Hilbert ℓ^2 comme suit :*

on munit Θ du produit scalaire :

$\forall \rho, \rho' \in \Theta,\ \langle \rho, \rho' \rangle = \sum_{k\geq 1} c_k^2 \alpha_k \alpha_k'.$

La suite des opérateurs $(\rho_k)_k$ est une base orthogonale dans $\left(\Theta, \|\ \|_p\right)$.

Nous obtenons alors les mêmes résultats de convergence sous la condition (2.11). Cette condition est vérifiée si et seulement si ε_t est pré-gaussienne, c'est à dire qu'il existe une variable aléatoire gaussienne ϑ à valeurs dans B tel que $C_\vartheta = C_{\varepsilon_0}$. D'aprés Ledoux et Talagrand (1991) (p.261), $((f_k^*, \vartheta))_k = (\vartheta_k)_k$ définit une variable aléatoire gaussienne dans ℓ^2 telle que

$$\sum_{k\geq 0} E|\vartheta_k|^2 = c \sum_{k\geq 0} E(\varepsilon_{o,k})^2 < \infty$$

où c est une constante.

Alors $(\varepsilon_{n,k})_k$ dans ℓ^2, est prégaussienne si et seulement si $\sum_{k\geq 1} \sigma_k^2 < \infty$.

2.6 Simulation

Dans ce paragraphe nous donnons deux exemples de l'estimation de ρ pour un ARB(1) gaussien ou $B = \ell^p$ et ρ un opérateur $p-$nucléaire, $p > 1$. Nous représentons d'abord la trajectoire du processus, ensuite nous estimons les coefficients (α_k), en calculant l'ordre de convergence dans ℓ^p et ℓ^2. Nous déduisons que la vitesse de convergence est de l'ordre de $n^{\frac{1}{min(2,p)}} (\log n)^{\frac{-\beta}{min(2,p)}}$, pour un $\beta > 1$. Cette simulation est réalisée avec le logiciel STATISTICA.

Nous donnons d'abord les résultats du paragraphe (2.4) dans le cas particulier où ρ est $p-$nucléaire.

Soit ρ un opérateur $p-$nucléaire, il existe alors $(h_k^*)_{k\in\mathbb{N}}$, $(v_k)_{k\in\mathbb{N}}$ deux suites normées dans la boule unité de B^* et B respectivement, et $\beta = (\beta_k)_{k\in\mathbb{N}} \in \ell^p$, tel que ρ s'ecrit $\rho = \sum\limits_{k\in\mathbb{N}} \beta_k \left(h_k^*, \cdot\right)\ v_k$ (Diestel *et al.* (1995) prop. 5.23). De cette décomposition nous avons la représentation suivante de a et b :

Lemme 2.6.1. *Les opérateurs : a de ℓ^p dans B et b de B dans ℓ^∞ sont définis respectivement par :*

$(c_k)_k \mapsto \sum\limits_{k\in\mathbb{N}} c_k\ v_k$ *et* $x \mapsto ((h_k^*, x))_k$.

Pour obtenir une décomposition de ρ, il suffit de choisir $(e_k)_{k\in\mathbb{N}}$ la base canonique dans ℓ^p (cf. paragraphe 2.4). Nous avons le cas particulier du lemme 2.9 :

Lemme 2.6.2. *Sous l'hypothèse* **H** *l'opérateur $p-$nucléaire ρ admet la décomposition suivante*

$$\rho(\cdot) = \sum_{k\in\mathbb{N}} \alpha_k\ \left(h_k^*, \cdot\right)\ ae_k\,.$$

Remarque 2.6.1. *Si $(f_k^*)_k = (h_k^*)_k$ alors les $(ae_k)_k$ sont les vecteurs propres de ρ associés aux valeurs propres $(\alpha_k)_k$.*

Nous considérons $\Theta = \ell^p$ et le crible $\Theta_m = \left\{\alpha \in \ell^p \,/ \alpha = \sum\limits_{k=0}^{m} \alpha_k e_k,\ \alpha_k = 0,\ k > m\right\}$. Dans ce cas nous avons :

Lemme 2.6.3. *L'estimateur crible de α est la fonction $\widehat{\alpha} = \sum\limits_{k=0}^{m} \widehat{\alpha}_k e_k$*

$$\text{où } \widehat{\alpha}_k = \frac{\sum\limits_{i=1}^{n} \left(h_k^*, X_{i-1}\right)\left(f_k^*, X_i\right)}{\sum\limits_{i=1}^{n} \left(h_k^*, X_{i-1}\right)^2} \qquad ,\quad k = 0, \ldots, m$$

La convergence de $\widehat{\alpha}$ est établie de la même façon que le théorème 2.5.1.

Soit maintenant $(X_t, t \in \mathbb{Z})$ un ARℓ^p gaussien associé à $\rho \in \mathcal{D}(\ell^\infty, \ell^\infty)$ *défini par* $\alpha = (\alpha_k)_{k\in\mathbb{N}} \in \ell^p$. ρ est $p-$ nucléaire d'image dans ℓ^p tel que $\|\rho\|_p = \|\alpha\|_{\ell^p}$. Dans ce cas l'estimation de ρ revient à l'estimation de α dans ℓ^p car ρ s'écrit $\rho(\cdot) = \sum\limits_{k\in\mathbb{N}} \alpha_k \, (e_k, \cdot)\, e_k$, où $(e_k)_{k\in\mathbb{N}}$ est la base canonique de ℓ^p .

On note C_ε de ℓ^q dans ℓ^p l'opérateur de covariance du bruit blanc. D'après le théorème (5.6) dans Vakhania *et al.* (1987), C_ε est l'opérateur de covariance d'une variable aléatoire gaussienne si et seulement s'il est symétrique défini positif et $\sum\limits_{k\geq 0}(C_\varepsilon e_k, e_k)^{\frac{p}{2}} < \infty$. Dans ce cas l'estimateur crible de ρ par la méthode des moindres carrés est solution de l'équation :

$$\min_{\rho\in\Theta_m} \sum_{k=0}^{m}(\sum_{i=1}^{n}(X_i - \rho X_{i-1}, e_k)^2)^{\frac{p}{2}}.$$

L'estimateur crible de α est la suite $\widehat{\alpha}_n = \sum\limits_{k=0}^{m} \widehat{\alpha}_{k,n} e_k$

$$\text{où } \widehat{\alpha}_{k,n} = \frac{\sum\limits_{i=1}^{n} (e_k, X_{i-1})\,(e_k, X_i)}{\sum\limits_{i=1}^{n} (e_k, X_{i-1})^2} \qquad , \quad k = 0, \ldots, m.$$

On remarque que $\widehat{\alpha}_{k,n}$ est l'estimateur du coefficient de correlation du processus AR(1) $(e_k, X_i) = \alpha_k\,(e_k, X_{i-1}) + (e_k, \varepsilon_i)$ pour $k = 0, \ldots, m$.

Exemple 2.6.1. *Soit* (X_t) *un ARB(1) tel que* $\rho = \sum\limits_{k\geq 0} 0.95^{0.95k} e_k$ *et* $\varepsilon_{t,k} \sim \mathcal{N}(0; 0.9^{0.95k})$.

(X_t) *est un ARB(1) dans* $\ell^{1.05}$. *Nous choisissons* $m = 40$ *et nous présentons la trajectoire de ce processus.*

Nous remarquons dans la page suivante, que sur chaque intervalle $[t, 40t]$, $t = 1, ..., 5$ *; la trajectoire varie de forte à de faible amplitudes. Cela est dû aux valeurs décroissante de la suite* (α_k).

Représentation graphique du processus

Tableau des résultats de l'estimation des coefficients de l'opérateur ρ pour

$p = 1.05$, $n = 1000$ **et** $m = 40$

On obtient pour $\beta = 4, 7$, les valeurs suivantes

$\|\hat{\alpha} - \alpha_0\|_{\ell_2} = 0.18$,

$\|\hat{\alpha} - \alpha_0\|_{\ell^{1.05}} = 0.74$.

Pour $n = 1000$ la vitesse de convergence est de $0, 42$ dans ℓ^2, alors que dans $\ell^{1.05}$ elle vaut $0, 19$. Nous déduisons que la vitesse de convergence est meilleur en considérant le processus dans $\ell^{1.05}$ au lieu de ℓ^2.

Exemple 2.6.2. *Soit* (X_t) *un ARB(1) tel que* $\rho = \sum_{k\geq 0} (\frac{1}{k})^{1/2} e_k$ *et* $\varepsilon_{t,k} \sim \mathcal{N}(0; 0.5^{k/8})$.

(X_t) *est un ARB(1) dans* ℓ^4*. Nous choisissons* $m = 40$ *et nous présentons la trajectoire de ce processus.*

Comme l'exemple précédent nous remarquons dans la page suivante, que sur chaque intervalle $[t, 40t]$, $t = 1, ..., 5$ *; la trajectoire varie de forte à de faible amplitudes. Cela est dû aussi aux valeurs décroissante de la suite* (α_k).

Représentation graphique du processus

Tableau des résultats de l'estimation des coefficients de l'opérateur ρ pour

$p = 4$**,** $n = 1000$ **et** $m = 40$

On obtient pour $\beta = 3.5$, $\|\hat{\alpha}-\alpha_0\|_{\ell^4} = 0.240$ et pour $n = 1000$ la vitesse de convergence est de 0.22.

Nous donnons maintenant les démonstrations des lemmes, propositions et le théorème énoncés dans ce chapitre.

2.7 Preuves :

Preuve du lemme 2.4.1 : $\forall x \in B$ nous avons :

$$\begin{aligned}
\rho x &= \sum_{k\in\mathbb{N}} \left(f_k^*, \rho x\right) ae_k = \sum_{k\in\mathbb{N}} \left(f_k^*, aM_\xi bx\right) ae_k = \sum_{k\in\mathbb{N}} \left(e_k^*, M_\xi bx\right) ae_k \\
&= \sum_{k\in\mathbb{N}} \left(e_k^*, \xi bx\right) ae_k = \sum_{k\in\mathbb{N}} \left(e_k^*, \left(\sum_{j\in\mathbb{N}} \left(e_j^*, \xi\right) e_j\right) \cdot bx\right) ae_k \\
&= \sum_{j,k\in\mathbb{N}} \left(e_j^*, \xi\right) \cdot \left(e_k^*, e_j\, bx\right) ae_k
\end{aligned}$$

comme $bx \in L^\infty(\nu)$, alors

$$\rho x = \sum_{k\in\mathbb{N}} \left(e_k^*, \xi\right) \left(e_k^*, e_k\, bx\right) ae_k.$$

Preuve de la proposition 2.4.1 :

$$\begin{aligned}
\min_{\rho\in\Theta_m} \sum_{k=0}^{m} (\sum_{i=1}^{n} \left(f_k^*, X_i - \rho X_{i-1}\right)^2)^{\frac{p}{2}} &= \min_{(\alpha_k)_{k=0,\ldots,m}} \sum_{k=0}^{m} (\sum_{i=1}^{n} \left(\left(f_k^*, X_i\right) - \alpha_k \left(f_k^*, \rho_k X_{i-1}\right)\right)^2)^{\frac{p}{2}} \\
&= \sum_{k=0}^{m} \min_{\alpha_k} (\sum_{i=1}^{n} \left(\left(f_k^*, X_i\right) - \alpha_k \left(f_k^*, \rho_k X_{i-1}\right)\right)^2)^{\frac{p}{2}}.
\end{aligned}$$

L'estimateur des moindres carrés de ρ s'obtient par l'estimateur des moindres carrés des α_k, $k = 0, \ldots, m$.

L'estimateur de α_k est obtenu par la régression de (f_k^*, X_i) sur $(f_k^*, \rho_k X_{i-1})$ pour $i = 1, \ldots, n$; du processus réel $(f_k^*, X_t) = \alpha_k (f_k^*, \rho_k X_{t-1}) + (f_k^*, \varepsilon_t)$, $t \in \mathbb{Z}$. ■

Preuve du lemme 2.5.1 :

$\forall \rho \in \Theta$, nous avons :

$$\left|\sum w_{n,k}^p \left|\alpha_k\right|^p - \sum \left|\alpha_k\right|^p\right| = \left|\sum_{k=0}^{m} \alpha_k^p \left(\left(\frac{(C_n \rho_k{}^* f_k^*, \rho_k{}^* f_k^*)}{(C \rho_k{}^* f_k^*, \rho_k{}^* f_k^*)}\right)^p - 1\right)\right|,$$

d'après la proposition 1.4.2 à partir d'un certain rang N, $\dfrac{(C_n\rho_k{}^* f_k^*, \rho_k{}^* f_k^*)}{c_k}$ s'approche de 1, donc soit cette valeur est inférieur à 1 soit supérieur 1, on a alors :

$$\left|\sum w_{n,k}^p \left|\alpha_k\right|^p - \sum \left|\alpha_k\right|^p\right|$$

$$\leq \left|\sum_{k=0}^m |\alpha_k^p| \left(\frac{(C_n\rho_k{}^* f_k^*, \rho_k{}^* f_k^*)}{c_k} - 1\right)\right| + \left|\sum_{k=0}^m |\alpha_k^p| \left(\left(\frac{(C_n\rho_k{}^* f_k^*, \rho_k{}^* f_k^*)}{c_k}\right)^2 - 1\right)\right|$$

$$\leq \left|\sum_{k=0}^m \frac{|\alpha_k|^p}{c_k^2} \left(\left((C_n - C)\,\rho_k{}^* f_k^*, \rho_k{}^* f_k^*\right)^2 + 2c_k\left((C_n - C)\,\rho_k{}^* f_k^*, \rho_k{}^* f_k^*\right)\right)\right|$$

$$+ \left|\sum_{k=0}^m \frac{|\alpha_k|^p}{c_k} \left((C_n - C)\,\rho_k{}^* f_k^*, \rho_k{}^* f_k^*\right)\right|$$

$$\leq \sum_{k=0}^m \frac{|\alpha_k|^p}{c_k^2} \left((C_n - C)\,\rho_k{}^* f_k^*, \rho_k{}^* f_k^*\right)^2 + 3\left|\sum_{k=0}^m \frac{|\alpha_k|^p}{c_k} \left((C_n - C)\,\rho_k{}^* f_k^*, \rho_k{}^* f_k^*\right)\right|.$$

Pour montrer la convergence p.s. vers 0 du deuxième terme de l'inégalité, nous considérons les variables aléatoires réelles :

$$Z_i = \sum_{k=0}^m \left((X_i, \rho_k^* f_k^*)^2 - (C\rho_k^* f_k^*, \rho_k^* f_k^*)\right) \frac{\alpha_k^p}{c_k}\ ,\ i \geq 1.$$

Nous utilisons maintenant le théorème 3.6.1 :

$$E\left(Z_n + \ldots + Z_{n+p-1}\right)^2 = E\left(\sum_{k=0}^m \sum_{i=n}^{n+p-1} \frac{\left((X_i, \rho_k^* f_k^*)^2 - (C\rho_k^* f_k^*, \rho_k^* f_k^*)\right)}{c_k}\ \alpha_k^p\right)^2$$

$$\leq E\sum_{k,h=1}^m \sum_{i,j=n}^{n+p-1} \frac{\left((X_i, \rho_k^* f_k^*)^2 - (C\rho_k^* f_k^*, \rho_k^* f_k^*)\right)}{c_k}\ \alpha_k^p \frac{\left((X_j, \rho_h^* f_h^*)^2 - (C\rho_h^* f_h^*, \rho_h^* f_h^*)\right)}{c_h}\ \alpha_h^p$$

$$\leq \sum_{k,h=1}^m \sum_{i,j=n}^{n+p-1} Cov\left(\frac{(X_i, \rho_k^* f_k^*)^2\,(X_j, \rho_h^* f_h^*)^2}{c_k c_h}\right)\alpha_k^p \alpha_h^p$$

$$\leq \sum_{k,h=1}^m \left(\sum_{i=n}^{n+p-1} Cov\left(\frac{(X_i, \rho_k^* f_k^*)^2\,(X_i, \rho_h^* f_h^*)^2}{c_k c_h}\right)\right)$$

$$+2\sum_{k,h=1}^m \left(\sum_{n\leq j<i\leq n+p-1} Cov\left(\frac{(X_i, \rho_k^* f_k^*)^2\,(X_j, \rho_h^* f_h^*)^2}{c_k c_h}\right)\right)\alpha_k^p \alpha_h^p \qquad (\triangle)$$

d'après le lemme 1.4.1 il existe $c_1 > 0$ et $c_2 \in\,]0,1[$ deux constantes tel que

$$(\triangle) \leq \frac{m}{\lambda_m^2}\left(p\left\|b^*\right\|^4 E\left(\left\|X_0\right\|^4\right)+2\left\|b^*\right\|^4 c_1 \sum_{n \leq j<i \leq n+p-1} c_1^{i-j}\right)\left(\sum_{k \geq 0} \alpha_k^p\right)^2$$

$$\leq \frac{pm}{\lambda_m^2}\left(p\left\|b^*\right\|^4 E\left(\left\|X_0\right\|^4\right)+2\left\|b^*\right\|^4 c_1 \sum_{s=1}^{p-1}(p-s)\, c_2^s\right)\|\alpha\|^{2p}$$

$$\leq \frac{pm}{\lambda_m^2}\left\|b^*\right\|^4\left(E\left(\left\|X_0\right\|^4\right)+\frac{2c_1}{1-c_2}\right) 5\|\alpha\|^{2p},$$

nous avons alors $\forall \rho \in \Theta$

$$n^{1/2} m_n^{1/2} \lambda_{m_n}(\log n)^{-\beta}\left|\sum_{k=0}^{m} \frac{\left(\left(C_n-C\right) \rho_k^* f_k^*, \rho_k^* f_k^*\right)}{c_k} \alpha_k^p\right| \xrightarrow[n \rightarrow+\infty]{p.s.} 0 \qquad \forall \beta>\frac{1}{2}.$$

Concernant le premier terme de l'inégalité, nous considérons les variables aléatoires

$$W_i=\left(\left(\left(X_i, \rho_k^* f_k^*\right)^2-\left(C \rho_k^* f_k^*, \rho_k^* f_k^*\right)\right) \frac{\alpha_k^{\frac{p}{2}}}{c_k}\right)_k$$

dans ℓ^2.

Dans ce cas nous avons :

$$E\left\|W_n+\ldots+W_{n+p-1}\right\|^2=E \sum_{k=0}^{m}\left(\frac{\alpha_k^{\frac{p}{2}}}{c_k} \sum_{i=n}^{n+p-1}\left(X_i, \rho_k^* f_k^*\right)^2-\left(C \rho_k^* f_k^*, \rho_k^* f_k^*\right)\right)^2$$

$$\leq E \sum_{k=0}^{m} \sum_{i,j=n}^{n+p-1}\left(\left(\left(X_i, \rho_k^* f_k^*\right)^2-\left(C \rho_k^* f_k^*, \rho_k^* f_k^*\right)\right)\left(\left(X_j, \rho_k^* f_k^*\right)^2-\left(C \rho_k^* f_k^*, \rho_k^* f_k^*\right)\right)\right) \frac{\alpha_k^p}{c_k^2}$$

$$\leq \sum_{k=0}^{m} \frac{\alpha_k^p}{c_k^2} \sum_{i,j=n}^{n+p-1} Cov\left(\left(X_i, \rho_k^* f_k^*\right)^2\left(X_j, \rho_k^* f_k^*\right)^2\right)$$

$$\leq \sum_{k=0}^{m} \frac{\alpha_k^p}{c_k^2} \sum_{i=n}^{n+p-1} Cov\left(\left(X_i, \rho_k^* f_k^*\right)^2\left(X_i, \rho_k^* f_k^*\right)^2\right)$$

$$+2 \sum_{k=0}^{m} \frac{\alpha_k^p}{c_k^2} \sum_{n \leq j<i \leq n+p-1} Cov\left(\left(X_i, \rho_k^* f_k^*\right)^2\left(X_j, \rho_k^* f_k^*\right)^2\right)$$

$$\leq \sum_{k=0}^{m} \frac{\alpha_k^p}{c_k^2}\left(p\left\|b^*\right\|^4 E\left(\left\|X_0\right\|^4\right)+2 \sum_{n \leq j<i \leq n+p-1} Cov\left(\left(X_i, \rho_k^* f_k^*\right)^2\left(X_j, \rho_k^* f_k^*\right)^2\right)\right) \qquad (\triangle')$$

en continuant de la même manière que précedement on a aussi :

$$(\triangle') \leq \frac{pm}{\lambda_m^2}\left\|b^*\right\|^4\left(E\left(\left\|X_0\right\|^4\right)+\frac{2c_1}{1-c_2}\right) 5\|\alpha\|^p,$$

alors $\forall \rho \in \Theta$

$$n^{1/2}m_n^{1/2}\lambda_{m_n}(\log n)^{-\beta}\sum_{k=0}^{m}((C_n-C)\rho_k^*f_k^*,\rho_k^*f_k^*)^2\frac{\alpha_k^p}{c_k^2}\xrightarrow[n\to+\infty]{p.s.}0\quad\forall\beta>\frac{1}{2}.$$

Finalement $\forall\rho\in\Theta$

$$n^{1/2}m_n^{1/2}\lambda_{m_n}(\log n)^{-\beta}\left|\|\rho\|_{n,p}^p-\|\Pi_m\rho\|_p^p\right|\xrightarrow[n\to+\infty]{p.s.}\qquad\forall\beta>\frac{1}{2}.$$

■

Preuve de la proposition 2.5.1 :

1. Comme l'espace ℓ^p est de type $min(2,p)$, $E\left\|\Pi_m\overline{E}_n\right\|^r\geq E^{\frac{r}{s}}\left(\left\|\Pi_m\overline{E}_n\right\|^s\right)$, pour tout $1\leq s\leq r$, et $E|\varepsilon_k|^p=c_p(E|\varepsilon_k|^2)^{\frac{p}{2}}$ pour $p>2$ ($(\varepsilon_k)_k$ est supposé gaussien dans le cas $p>2$), nous avons alors $\forall n\geq 1$:

$$\begin{aligned}
E\left\|\Pi_m\sum_{i=n}^{n+h-1}E_i\right\|_p^{min(2,p)} &\leq C_p\sum_{i=n}^{n+h-1}E\left(\sum_{k=0}^{m}\left(\frac{|(F_i\rho_k^*f_k^*,f_k^*)|}{c_k}\right)^p\right)^{\frac{min(2,p)}{p}}\\
&\leq C_p\sum_{i=n}^{n+h-1}\left(\sum_{k=0}^{m}E\left(\frac{|(F_i\rho_k^*f_k^*,f_k^*)|}{c_k}\right)^p\right)^{\frac{min(2,p)}{p}}\\
&\leq C_p\sum_{i=n}^{n+h-1}\left(\sum_{k=0}^{m}E|\left(b^*\left(e_k\overline{e}_k\right),X_{i-1}\right)|^p\right.\\
&\qquad\left.E\left(\frac{|(f_k^*,\varepsilon_i)|}{c_k}\right)^p\right)^{\frac{min(2,p)}{p}}\\
&\leq C_p'h\|b^*\|^{min(2,p)}E\|X_0\|^{min(2,p)}\left(\sum_{k=0}^{m}\left(\frac{\sigma_k}{c_k}\right)^p\right)^{min(2,p)}\\
&\leq C_p'h\|b^*\|^{min(2,p)}E(\|X_0\|^{min(2,p)})\sigma^{min(2,p)},
\end{aligned}$$

où C_p et C_p' sont deux constantes strictement positives.

D'après le lemme **3.6.1**, nous avons alors :

$$n^{\frac{1}{min(2,p)}}(\log n)^{-\frac{\beta}{min(2,p)}}\left\|\Pi_m\overline{E}_n\right\|_{max(2,p)}\xrightarrow[n\to+\infty]{p.s.}0\quad\forall\beta>1.\tag{2.23}$$

2. Nous avons montré que $\left\|\Pi_m\overline{E}_n\right\|_p$ converge vers 0, alors à partir d'un certain rang N, $\left\|\Pi_m\overline{E}_n\right\|_p\leq 1$. Nous suivons maintenant les mêmes démarches que dans Bosq (2002) p. 300, nous avons alors :

$$E\left\|\Pi_m\left(E_1+...+E_n\right)\right\|_p \leq \left(E\left\|\Pi_m\left(E_1+...+E_n\right)\right\|_p^2\right)^{\frac{1}{2}}$$

$$\leq \left(E\left\|\Pi_m\left(E_1+...+E_n\right)\right\|_p^{min(2,p)}\right)^{\frac{1}{2}} \leq n^{\frac{1}{2}}c^{\frac{1}{2}}\sigma^{\frac{1}{2}}$$

où $c = \|b^*\| \, E\|X_0\|$.

D'autre part comme $E(\exp\gamma\|E_1\|) < \infty$, pour $\delta > 0$,

$\left\|\Pi_m\left(E_1+...+E_n\right)\right\|_p > n\delta$ implique :

$\left\|\Pi_m\left(E_1+...+E_n\right)\right\|_p - E\left\|\Pi_m\left(E_1+...+E_n\right)\right\|_p \geq n\delta - n^{\frac{1}{2}}c^{\frac{1}{2}}\sigma^{\frac{1}{2}} \geq n^{\frac{1}{2}}\delta$,

d'après Pinelis et Sakhanenko (1985) pour $n \geq \max(4c\frac{\sigma^2}{\delta^2}, N)$, nous avons :

$$P\left(\left\|\Pi_m\left(E_1+...+E_n\right)\right\|_p > n\delta\right) \leq 2\exp\left(-\frac{n\delta^2}{8n\ell^2+4L\delta}\right)$$

où $K > 0$ et $L > 0$ sont deux constantes.

Dans le cas $p = 2$, on utilise la remarque 3 dans Pinelis et Sakhanenko (1985) et des techniques classiques nous avons :

$$P(\max_{1\leq j\leq n}\left\|\Pi_m\left(E_1+...+E_j\right)\right\|_p \geq t) \leq 4\exp\left(-\frac{t^2}{na+bt}\right) \qquad t > 0,$$

où $a, b > 0$.

Si $u_n = (n\ln\ln n)^{1/2}$ et $n_k = N2^k$, $k \geq 0$, nous avons alors :

$$\begin{aligned} P\left(\sup_{n\geq N}\left\|\frac{\Pi_m(E_1+...+E_n)}{u_n}\right\|_p \geq A\right) &= P\left(\sup_{k\geq 0}\max_{n_k\leq n<n_{k+1}}\left\|\frac{\Pi_m(E_1+...+E_n)}{u_n}\right\|_p \geq A\right) \\ &\leq \sum_{k=0}^{\infty} P(\max_{n\leq n_{k+1}}\left\|\frac{\Pi_m(E_1+...+E_n)}{u_n}\right\|_p \geq A) \\ &\leq 4\sum_{k=0}^{\infty}\exp\left(-\frac{A^2u_{u_{n_k}}^2}{a_{n_{k+1}}+bAu_{n_k}}\right). \end{aligned}$$

Pour A telle que $A^2 > 2a + bA$, on a :

$$P\left(\sup_{n\geq N}\left\|\frac{\Pi_m(E_1+...+E_n)}{u_n}\right\|_p \geq A\right) \leq \frac{c}{(\ln N)^d}, \qquad N \geq 2, \quad (c > 0,\ d > 0),$$

nous avons alors le résultat lorsque N tend vers l'infini.

3. Comme (E_i) est une différence de martingale ℓ^p, d'après le théorème centrale limite de Jakubowski (1988), on a la normalité asymptotique de $\sqrt{n}\ (u^*, \Pi_m \overline{E}_n)$.

   Pour $p = 2$, on a le cas particulier où la différence de martingale est à valeurs dans l'espace de Hilbert ℓ^2.

■

Preuve du théorème 2.5.1 :

Pour $k = 0, \ldots, m$; nous avons :

$$\begin{aligned}(\widehat{\alpha}_k - \alpha_{0,k})\,(C_n \rho_k{}^* f_k^*, \rho_k{}^* f_k^*) &= (D_n \rho_k{}^* f_k^*, f_k^*) - \alpha_{0,k}\,(C_n \rho_k{}^* f_k^*, \rho_k{}^* f_k^*) \\ &= \frac{1}{n}\sum_{i=1}^{n} (\rho_k{}^* f_k^*,\ X_{i-1})\,(f_k^*, \varepsilon_i)\end{aligned}$$

Donc :

$$(w_{k,n}\,(\widehat{\alpha}_k - \alpha_{0,k}))_{k=0,\ldots,m} = \Pi_m \overline{E}_n.$$

On déduit alors 2.18, 2.19 et 2.20 à partir de 2.13, 2.14 et 2.15.

Pour 2.21 et 2.22, nous avons :

$$\begin{aligned}\sqrt{n}(\hat{\rho}_{n,m} - \rho_m) &= \sqrt{n}(\sum_{k=0}^{m} w_{k,n}(\hat{\alpha}_k - \alpha_{0,k})\rho_k \\ &- \sum_{k=0}^{m}(w_{k,n} - 1)(\hat{\alpha}_k - \alpha_{0,k})\rho_k)\ , \\ &= \sqrt{n}\sum_{k=0}^{m} \overline{E}_{n,k}\rho_k - \sqrt{n}\sum_{k=0}^{m}(w_{k,n} - 1)(\hat{\alpha}_k - \alpha_{0,k})\rho_k,\end{aligned}$$

comme

$$\begin{aligned}(\hat{\alpha}_k - \alpha_{0,k}) &= \left(\frac{(D_n \rho_{n,k}^* f_k^*, f_k^*)}{(C_n \rho_k^* f_k^*, \rho_k^* f_k^*)} - \frac{(D \rho_{n,k}^* f_k^*, f_k^*)}{(C \rho_k^* f_k^*, \rho_k^* f_k^*)}\right) \\ &= ((D_n - D)\rho_k^* f_k^*, f_k^*)\,(C_n \rho_k^* f_k^*, \rho_k^* f_k^*)^{-1} \\ &\quad + (D\rho_k^* f_k^*, f_k^*)\left((C_n \rho_k^* f_k^*, \rho_k^* f_k^*)^{-1} - (C\rho_k^* f_k^*, \rho_k^* f_k^*)^{-1}\right),\end{aligned}$$

alors $(\hat{\alpha}_k - \alpha_{0,k})$ converge vers 0, donc à partir d'un certain rang N, $(\hat{\alpha}_k - \alpha_{0,k}) \leq 1$. Nous suivons maintenant la même technique utilisée dans la preuve du lemme 2.5.1, nous avons alors :

$$n^{1/2} m_n^{1/2} \lambda_{m_n} (\log n)^{-\beta} \left(\sum_{k=0}^{m} (w_{k,n} - 1)(\hat{\alpha}_k - \alpha_{0,k}) \zeta_k^* \right) \underset{n \to +\infty}{\overset{p.s.}{\longrightarrow}} 0 \quad \forall \beta > \frac{1}{2},$$

où ζ est une suite dans ℓ^q.

Donc si $m_n^{1/2} \lambda_{m_n} (\log n)^{-\beta} \underset{n \to +\infty}{\longrightarrow} 0$, nous déduisons 2.21 et 2.22 à partir de 2.16 et 2.17.

■

Preuve du lemme 2.6.1 :

Soit $a' (\xi_k)_{k \in \mathrm{N}} = \sum_{k \in \mathrm{N}} \xi_k v_k$, $(\xi_k)_{k \in \mathrm{N}} \in \ell^2$ et $b'x = ((h_k^*, x))_k$, $x \in B$, alors

$a' D_\beta b' = \sum_{k \in \mathrm{N}} \beta_k (h_k^*, x) v_k$.

Ainsi on peut choisir $a = a'$ et $b = b'$ et nous avons alors :

$$\begin{aligned} \langle a^* x^*, \xi \rangle = (x^*, a\xi) &= \left(x^*, \sum_{j \in \mathrm{N}} \xi_j v_j \right) \\ &= \sum_{j \in \mathrm{N}} \overline{\xi}_j (x^*, v_j) \end{aligned}$$

comme $a^* \; f_k^* = e_k$ alors $e_k = \left(f_k^*, \; v_j \right)_{j \in \mathrm{N}} \; \forall k \in \mathbb{N}$.

$(e_k)_{k \in \mathrm{N}}$ est la base canonique dans ℓ^p alors $\left(f_k^* \right)_{k \in \mathrm{N}}$ sont les fonctionnelles des coefficients associées à $(v_k)_{k \in \mathrm{N}}$. ■

Preuve du lemme 2.6.2 : D'après le lemme 2.6.1 et la décomposition (2.9) nous avons :

$\rho = \sum_{k \in \mathrm{N}} \alpha_k \; \langle e_k, e_k \; ((h_i^*, \cdot))_i \rangle \; ae_k$

où $(e_k)_{k \in \mathrm{N}}$ est la base canonique dans ℓ^p, alors ρ s'écrit $\forall x \in B$:

$\rho(x) = \sum_{k \in \mathrm{N}} \alpha_k \; (h_k^*, x) \; ae_k$. ■

Preuve du lemme 2.6.3 : Se déduit facilement du lemme 2.6.2 et la proposition 2.4.1. ■

Chapitre 3

Convergence en norme p-sommable de l'estimateur crible de l'opérateur d'un ARB(1)

Ce travail a fait l'objet d'une publication aux Comptes Rendus Mathématique, Académie des Sciences, Paris (Rachedi et Mourid (2003)).

3.1 Introduction et notations

Dans ce chapitre, nous nous intéressons à l'estimation de ρ en utilisant les $\phi-$divergences entre des lois de probabilités conditionnelles induites par le processus ARB(1). Les divergences entre mesures ou $\phi-$divergences ont été introduites par Csiszár (1963) dans le but de définir une méthode générale d'estimation d'un paramètre. Nous rappelons d'abord la définition de la méthode des $\phi-$divergences, nous donnons ensuite quelques exemples de $\phi-$divergences les plus utilisées en théorie de l'estimation.

3.1.1 Divergence entre mesures

Soit φ une fonction convexe de $[0, +\infty]$ dans $[0, +\infty]$. Pour toutes mesures de probabilité (m.p.) Q et P sur $\mathcal{B}$ tel que Q est absolument continues (a.c.) par rapport à P, la $\phi-$divergence entre Q et P est définie par Rüschendorf (1984) comme suit :

Définition 3.1.1. *La $\phi-$divergence entre Q et P est :*

$$\phi(Q,P) = \int \varphi(\frac{dQ}{dP})dP \tag{3.1}$$

Si Q n'est pas a.c. par rapport à P on pose $\phi(Q,P) = +\infty$.

Csiszár (1963) a donné la définition de la $\phi-$divergence entre deux mesures de probabilité d'une famille de probabilités dominées par une mesure $\sigma-$finie.

Pour toute m.p. P, l'application $Q \mapsto \phi(Q,P)$ est convexe nonnégative. La $\phi-$divergence entre Q et P lorsque $Q = P$ est nulle. Si φ est strictement convexe au voisinage de 1, on a la propriété fondamentale :

$$\phi(Q,P) = 0 \text{ si et seulement si } Q = P$$

Exemple 3.1.1. 1. ***Divergence de Kullback-Leibler (KL)*** *: elle est définie par :*

$$KL(Q,P) = \begin{cases} \int \lg(\frac{dQ}{dP})dQ & \textit{si } Q \textit{ est a.c. } P \\ +\infty & \textit{sinon.} \end{cases}$$

Cette divergence correspond à $\varphi(x) = x\lg(x) - x + 1$.

Divergence de Kullback-Leibler modifiée (KL_m)

$$KL_m(Q,P) = \begin{cases} \int -\lg(\frac{dQ}{dP})dP & \textit{si } Q \textit{ est a.c. } P \\ +\infty & \textit{sinon} \end{cases}$$

Dans ce cas $\varphi(x) = x\lg(x) + x - 1$.

2. ***Divergence du χ^2*** *: elle est définie par :*

$$\chi^2(Q,P) = \begin{cases} \int \frac{1}{2}(\frac{dQ}{dP} - 1)^2 dP & \textit{si } Q \textit{ est a.c. } P \\ +\infty & \textit{sinon.} \end{cases}$$

Pour cette divergence $\varphi(x) = \frac{1}{2}(x-1)^2$.

Divergence du χ^2 modifiée (χ^2_m)

$$\chi_m(Q,P) = \begin{cases} \int \frac{1}{2}\frac{(\frac{dQ}{dP} - 1)^2}{\frac{dQ}{dP}} dP & \textit{si } Q \textit{ est a.c. } P \\ +\infty & \textit{sinon} \end{cases}$$

Dans ce cas $\varphi(x) = \frac{1}{2}\dfrac{(x-1)^2}{x}$.

3. ***Distance de Hellinger*** *et* ***distance dans*** L^p**,** $p \geq 1$

$$H(Q,P) = \begin{cases} \int 2(\sqrt{\frac{dQ}{dP}} - 1)^2 dP & \textit{si } Q \textit{ est a.c. } P \\ +\infty & \textit{sinon} \end{cases}$$

et

$$L^p(Q,P) = \begin{cases} \int (\dfrac{dQ}{dP} - 1)^p dP & \textit{si } Q \textit{ est a.c. } P \\ +\infty & \textit{sinon} \end{cases}$$

Pour ces deux $\phi-$*divergences* $\varphi(x) = 2(\sqrt{x} - 1)^2$ *et* $\varphi(x) = |x-1|^p$.

Remarque 3.1.1. *La* $\phi-$*divergence n'est pas nécessairement une distance.*

Nous rappelons maintenant la représentation duale de la $\phi-$divergence proposée par Broniatowski (2003) et Keziou (2003) et utilisée dans l'estimation paramétrique pour définir l'estimateur du minimum des $\phi-$divergences.

On considère d'abord les notations suivantes :

a) L'espace de toutes les probabilités sur $(B, \mathcal{B})$ n'etant pas linéaire, on considère l'espace de toutes les mesures signées finies définies sur $(B, \mathcal{B})$, qu'on note $\mathcal{M}$.

b) Soient $\mathcal{E}$ l'ensemble des fonctions mesurables réelles et bornées définies sur $(B, \mathcal{B})$ et Soit $\mathcal{F}$ une classe de fonctions mesurables réelles définies sur $(B, \mathcal{B})$. On note $< \mathcal{E} \cup \mathcal{F} >$ l'espace linéaire engendré par $\mathcal{E} \cup \mathcal{F}$.

c) Soit le sous espace vectoriel :

$$\mathcal{M}_{\mathcal{F}} = \{Q \in \mathcal{M} \text{ tel que } \int |f| d|Q| < \infty, \ \forall f \in \mathcal{F}\}$$

où $|Q|$ désigne la variation totale de la mesure Q.

On munit $\mathcal{M}_{\mathcal{F}}$ de la $\tau_{\mathcal{F}}$-topologie, c'est la topologie la plus faible qui rend continues les applications $Q \mapsto \int f dQ$, $f \in \mathcal{F}$.

d) Soit ϕ une divergence définie par une fonction φ de $[-\infty, +\infty]$ dans $[0, +\infty]$ tel que $\varphi(1) = 0$.

On note le domaine de φ par

$$D_\varphi = \{x \in [-\infty, +\infty] / \ \varphi(x) < \infty\}$$

e) La transformée de Fenchel-Legendre de la fonction $Q \mapsto \phi(Q, P)$ définie de $(\mathcal{M}_\mathcal{F}, \tau_\mathcal{F})$ dans $\mathbb{R}$ est :

$$\phi^*(f, Q) = \sup_{Q \in \mathcal{M}_\mathcal{F}} \{\int f dQ - \phi(Q, P)\}, \qquad f \in < \mathcal{E} \cup \mathcal{F} >$$

Broniatowski (2003) et Keziou (2003) ont montré que la transformé de Fenchel-Legendre de ϕ^* est ϕ, ensuite ils ont déduit une représentation duale de ϕ, en utilisant la transformée de Fenchel de φ, qu'on note ψ :

$$t \in \mathbb{R} \mapsto \psi(t) = \sup_{h \in \mathbb{R}} th - \varphi(h). \tag{3.2}$$

Soit $\partial\varphi$ la dérivée de φ. On note $(\partial\varphi)^{-1}$ la fonction inverse de $\partial\varphi$ et $Im(\partial\varphi)$ l'ensemble des valeurs de $\partial\varphi$.

On note φ^* la restriction de ψ sur $Im(\partial\varphi)$, alors

$$t \in Im(\partial\varphi) \mapsto \varphi^*(t) = t(\partial\varphi)^{-1}(t) - \varphi((\partial\varphi)^{-1}(t)). \tag{3.3}$$

La représentation duale de ϕ s'exprime en fonction de φ^* comme suit

Théorème 3.1.1. *Keziou (2003) : Soient Q une mesure signée finie et P une mesure de probabilité tels que $\phi(Q, P) < \infty$. On suppose que φ est une fonction strictement convexe et de classe $\mathcal{C}^2$ sur l'intérieur de D_φ. Soit $\mathcal{F}$ une classe de fonction telle que :*

1. *$Q \in \mathcal{M}_\mathcal{F}$,*
2. *$\partial\varphi(\frac{dQ}{dP}) \in \mathcal{F}$,*

3. $\forall f \in \mathcal{F}$, $Imf \subseteq Im(\partial\varphi)$,

alors on a :

$$\phi(Q,P) = \sup_{f\in\mathcal{F}} \left\{ \int f dQ - \int \varphi^*(f) dP \right\}. \tag{3.4}$$

Ce supremum est unique (P-p.s.) et il est atteint au point $f = \partial\varphi(\frac{dQ}{dP})$ *(P-p.s.).*

Nous appliquons par la suite ce théorème à des probabilités associées au processus ARB(1) et qui permettent de définir l'estimateur du minimum des $\phi-$divergences.

3.2 Estimateur crible du minimum des $\phi-$divergences

Soit $X = (X_t, t \in \mathbb{Z})$ un ARB(1) associé à $\rho \in \Pi_p(B)$, l'espace des opérateurs $p-$sommables, qui vérifie $\|\rho\|_{\mathcal{L}}^{j_0} < 1$ pour un $j_0 \geq 1$, et $\varepsilon = (\varepsilon_t, t \in \mathbb{Z})$ un bruit blanc fort dans B. Soient maintenant

- Θ l'espace des opérateurs $p-$sommables et la distance associée d est celle déduite de la norme. ρ_0 est la vraie valeur du paramètre.
- P la loi stationnaire induite sur $(B, \mathcal{B})$ par X_t et P_0 la loi de la variable aléatoire ε_0.
- $\mathcal{B}_t = \sigma(\varepsilon_i, i \leq t)$ la tribu engendré par la suite $(\varepsilon_i, i \leq t)$.
  $E^{\mathcal{B}_{t-1}}(X_t) = \rho(X_{t-1})$ et $\varepsilon_t = X_t - E^{\mathcal{B}_{t-1}}(X_t)$ est dite l'innovation du processus X.
- Pour ρ dans Θ, $P_\rho^{\mathcal{B}_{t-1}}$ désigne la probabilité conditionnelle, par rapport à la tribu $\mathcal{B}_{t-1}$, de la variable aléatoire $(\rho X_{t-1} + \varepsilon_t)$.
- La densité de $P_\rho^{\mathcal{B}_{t-1}}$ par rapport à P_0 est

$$g(x, X_{t-1}, \rho) = \frac{dP_\rho^{\mathcal{B}_{t-1}}(x)}{dP_0(x)}, \qquad x \in B.$$

- φ une fonction convexe de $[0, +\infty]$ dans $[0, +\infty]$.

Nous supposons que P est a.c. par rapport à P_0. D'après Bosq et Mourid (1999) dans le cas gaussien et sous certaines conditions, les lois P et P_0 sont équivalentes. Ainsi pour toutes mesures de probabilité conditionnelle Q et P sur $\mathcal{B}$ a.c. par rapport à P_0, la $\phi-$divergence entre Q et P est

$$\phi(Q,P) = \int \varphi\left(\frac{dQ}{dP_0}/\frac{dP}{dP_0}\right) dP \tag{3.5}$$

Alors pour $\rho, \theta \in \Theta$ la $\phi-$divergence entre $P_\rho^{\mathcal{B}_{t-1}}$ et $P_\theta^{\mathcal{B}_{t-1}}$ définie par la fonction φ est :

$$\phi(P_\rho^{\mathcal{B}_{t-1}}, P_\theta^{\mathcal{B}_{t-1}}) = \int \varphi\left(\frac{g(x, X_{t-1}, \rho)}{g(x, X_{t-1}, \theta)}\right) dP_\theta^{\mathcal{B}_{t-1}}(dx)$$

Nous notons par la suite $\phi(P_\rho^{\mathcal{B}_{t-1}}, P_\theta^{\mathcal{B}_{t-1}})$ par $\phi(\rho,\theta)$, $\rho, \theta \in \Theta$.

Nous utilisons la représentation duale de ϕ du théorème 3.1.1 pour déterminer l'estimateur crible du minimum des $\phi-$divergence.

On suppose maintenant que φ est une fonction strictement convexe et de classe $\mathcal{C}^2$ sur l'intérieur de D_φ. Soit $\mathcal{F}$ une classe de fonction telles que :

1. $\partial\varphi(\dfrac{g(x, X_{t-1}, \rho)}{g(x, X_{t-1}, \theta)}) \in \mathcal{F}$, $\forall \theta \in \Theta$,
2. $\forall f \in \mathcal{F}$, $Imf \subseteq Im(\partial\varphi)$,

D'après le théorème 3.1.1 on a :

$$\phi(\rho,\theta) = \sup_{f\in\mathcal{F}} \left\{ \int f(x) g(x, X_{t-1}, \rho) dP_0(x) - \int \varphi^*(f(x)) g(x, X_{t-1}, \theta) dP_0(x) \right\},$$

ce supremum est unique (P_0-p.s.) et il est atteint au point $f = \partial\varphi(\dfrac{g(\cdot, X_{t-1}, \rho)}{g(\cdot, X_{t-1}, \theta)})$ (P_0-p.s.).

On peut se restreindre alors à chercher la solution dans la classe

$$\mathcal{F} = \left\{ x \mapsto \partial\varphi(\frac{g(x, X_{t-1}, \rho)}{g(x, X_{t-1}, \beta)}, \ \beta \in \Theta_m \right\}.$$

On a alors :

$$\begin{aligned}\phi(\rho,\theta) &= \sup_{\beta\in\Theta_m}\left\{\int \partial\varphi(\frac{g(x,X_{t-1},\rho)}{g(x,X_{t-1},\theta)})g(x,X_{t-1},\rho)dP_0(x)\right. \\ &\left. -\int \varphi^*\left(\partial\varphi(\frac{g(x,X_{t-1},\rho)}{g(x,X_{t-1},\beta)})\right)g(x,X_{t-1},\theta)dP_0(x)\ \right\}\end{aligned}$$

On pose

$$m(x,X_{t-1},\rho,\beta)=\int \partial\varphi(\frac{g(x,X_{t-1},\rho)}{g(x,X_{t-1},\theta)})g(x,X_{t-1},\rho)dP_0(x)-\varphi^*(\partial\varphi(\frac{g(x,X_{t-1},\rho)}{g(x,X_{t-1},\beta)})),$$

ou encore

$$\begin{aligned}m(x,X_{t-1},\rho,\beta) &= \int \partial\varphi(\frac{g(x,X_{t-1},\rho)}{g(x,X_{t-1},\theta)})g(x,X_{t-1},\rho)dP_0(x) \\ &-\partial\varphi\left(\frac{g(x,X_{t-1},\rho)}{g(x,X_{t-1},\beta)}\right)\frac{g(x,X_{t-1},\rho)}{g(x,X_{t-1},\beta)} \\ &-\varphi\left(\frac{g(x,X_{t-1},\rho)}{g(x,X_{t-1},\beta)}\right)\end{aligned}$$

ϕ s'écrit alors :

$$\phi(\rho,\theta)=\sup_{\beta\in\Theta}\int m(x,X_{t-1},\rho,\beta)g(x,X_{t-1},\theta)dP_0(x).$$

ou encore

$$\phi(\rho,\theta)=\sup_{\beta\in\Theta}\int m(x,X_{t-1},\rho,\beta)dP_\theta^{\mathcal{B}_{t-1}}(x).$$

Comme $P_{\rho_0}^{\mathcal{B}_{t-1}}$ est la loi de X_t, on a alors :

$$\phi(\rho,\rho_0)=\sup_{\theta\in\Theta}E\left(m(X_t,X_{t-1},\rho,\theta)\right), \tag{3.6}$$

où E désigne l'espérance par rapport à la loi conditionnelle $P_{\rho_0}^{\mathcal{B}_{t-1}}$.

Alors le minimum des $\phi-$divergences, s'il existe, est solution de

$$\inf_{\rho\in\Theta_m}\sup_{\beta\in\Theta_m}E(m(X_t,X_{t-1},\rho,\beta)).$$

Un estimateur de ce minimum des $\phi-$divergences est donné dans le lemme suivant

Lemme 3.2.1. *L'estimateur crible du minimum des ϕ-divergences de ρ est :*

$$\widehat{\rho}_{m,n} = \arg \inf_{\rho\in\Theta_m} \sup_{\beta\in\Theta_m} \left(\frac{1}{n}\sum_{i=1}^{n} m(X_i, X_{i-1}, \rho, \beta) \right) \tag{3.7}$$

où $(X_1, ..., X_n)$ sont des observations.

Nous nous limitons ici à l'étude du cas $\varphi(x) = -log(x) + x - 1$, alors l'estimateur dit du minimum de KL-divergence (divergence de Kullback-Leibler), quand il existe, s'écrit sous la forme

$$\arg \inf_{\rho\in\Theta_m} \sup_{\theta\in\Theta_m} \left(\frac{1}{n}\sum_{i=1}^{n} \log(\frac{g(X_i, X_{i-1}, \rho)}{g(X_i, X_{i-1}, \theta)}) \right).$$

Il est clair que cet estimateur est celui du maximum de vraisemblance. Nous montrons par la suite qu'il converge p.s. vers la vraie valeur du paramètre pour la norme des opérateurs $p-$sommables.

3.3 Convergence de l'estimateur crible du maximum de vraisemblance de ρ

Nous montrons la convergence p.s. de l'estimateur crible, determiné dans le paragraphe précédant, en suivant les techniques de Geman et Hwang (1982), utilisées pour des observations indépendantes et identiquement distribuées, qu'on a adapté au cas autorégressif. Dans le cas d'un ARB(1) gaussien avec ρ un opérateur strictement $2-$intégral, nous montrons que si $n^{\frac{1}{3}-\delta}$, $0 < \delta < \frac{1}{3}$, on a la convergence p.s. pour la norme $2-$intégrale.

Ensuite, nous étudions le cas particulier où ρ est un opérateur $2-$nucléaire. Les opérateurs $2-$nucléaires sont des opérateurs strictement $2-$intégraux, cependant ils admettent une décomposition plus simple.

Nous introduisons d'abord les notations et les définitions suivantes :

1. L'entropie conditionnelle est

$$H(\rho, \theta) := E_\rho \ln g(X_t, X_{t-1}, \theta) = \int \ln(g(x, X_{t-1}, \theta))g(x, X_{t-1}, \rho)dP_0(x).$$

Comme

$$\phi(\rho_0, \rho) = \int \ln\left(\frac{g(x, X_{t-1}, \rho_0)}{g(x, X_{t-1}, \rho)}\right) g(x, X_{t-1}, \rho_0) dP_0(x),$$

alors

$$\phi(\rho_0, \rho) = H(\rho_0, \rho_0) - H(\rho_0, \rho).$$

L'ensemble des points de Θ_m où $H(\rho_0, \theta)$ atteint son maximum est noté :

$$A_m = \{\rho \in \Theta_m / \quad H(\rho_0, \rho) = H(\rho_0, \Theta_m) := \sup_{\theta \in \Theta_m} H(\rho_0, \theta)\}.$$

2. $L_n(x_1, ..., x_n; \rho)$ est la fonction de vraisemblance conditionnelle.

   L'ensemble des points de Θ_m où $L_n(x_1, ..., x_n; .)$ atteint son maximum est noté :

$$M_m^n = \{\rho \in \Theta_m / \quad L_n(\omega; \rho) = L_n(\omega; \Theta_m) := \sup_{\theta \in \Theta_m} L_n(\omega; \theta)\}.$$

3. $B_m(\rho, \varepsilon) := \{\theta \in \Theta_m \ / \ d(\rho, \theta) < \varepsilon\}$, $\rho \in \Theta_m$,
4. pour une fonction f réelle, nous posons $f(A) := \sup_{y \in A} f(y)$.

Le résultat suivant établit la convergence p.s. de l'estimateur crible.

Théorème 3.3.1. *Si le crible, $\{\Theta_m\}$ est choisie telle que*

1. *Pour tout n et tout $\rho \in \Theta_m$, il existe $\varepsilon > 0$ tel que $E\left(\ln(g(X, X_{t-1}, B_m(\rho, \varepsilon)))\right) < \infty$,*
2. *$A_m \underset{n \to +\infty}{\longrightarrow} \rho_0$,*

alors

$$\sup_{\rho \in M_m^n} \pi_p\left(\rho, \rho_0\right) \underset{m,n \to +\infty}{\longrightarrow} 0 \quad p.s.$$

Pour déterminer un ordre de croissance de la dimension m nous considérons les conditions et notations suivantes :

C$_1$: *Si (ρ_m) est une suite telle que $\forall n$, $\rho_m \in \Theta_m$ et $H(\rho_0, \rho_m) \to H(\rho_0, \rho_0)$ alors $\rho_m \to \rho_0$.*

$\mathbf{C}_2$: *Il existe une suite* $(\rho_m \in \Theta_m)$ *telle que* $H(\rho_0, \rho_m) \to H(\rho_0, \rho_0)$.

Pour tout $\delta > 0$ *et tout* n, *on définit les ensembles :*

$D_m = \{\rho \in \Theta_m \ / \ H(\rho_0, \rho) \leq H(\rho_0, \rho_m) - \delta\}$ *où* ρ_m *est la suite définie par* $\mathbf{C}_2$.

Soit l *sous ensembles* $\Gamma_1, ..., \Gamma_l$ *de* Θ_m *et notons*

$$\varphi_m := \sup_k \inf_{t \geq 0} E_{\rho_0} \exp \left[t \ln \left\{ \frac{g(x, X_{t-1}, \Gamma_k)}{g(x, X_{t-1}, \rho_m)} \right\} \right].$$

Le théorème suivant donne la convergence p.s. de l'estimateur crible avec une condition sur m_n.

Théorème 3.3.2. *Soit* $\{\Theta_m\}$ *un crible vérifiant la condition* C_1 *et* C_2. *Si pour tout* $\delta > 0$, *on peut trouver* $\Gamma_1^m, ..., \Gamma_{l_m}^m$ *dans* Θ_m, $n = 1, 2, ...$ *tels que :*

(*i*) $D_m \subseteq \bigcup_{k=1}^{l_m} \Gamma_k^m$

(*ii*) $\sum_{n=1}^{+\infty} l_{m_n} (\varphi_{m_n})^n < +\infty$, $m_n \underset{n \to +\infty}{\longrightarrow} +\infty$

alors

$$\sup_{\rho \in M_{m_n}^n} \pi_p(\rho, \rho_0) \underset{n \to +\infty}{\longrightarrow} 0 \quad p.s.$$

La condition ii) du théorème est la condition sur m_n, à partir de laquelle on peut déduir une vitesse de convergence du crible. Dans le paragraphe suivant nous déterminons cette vitesse pour un opérateur strictement 2−intégrale d'un ARB(1) gaussien.

3.4 ρ strictement 2-intégral

Soit ρ un opérateur strictement 2−intégral tel que $\rho = aM_\xi b$, où a et b sont deux opérateurs de $L^p(\nu)$ dans B et de B dans $L^\infty(\nu)$ respectivement, ν étant une mesure, et M_ξ un opérateur de multiplication défini par $\xi \in L^p(\nu)$. Si la suite $(ae_k)_{k \in \mathbb{N}}$ est une base de shrinking, où $(e_k)_{k \in \mathbb{N}}$ est une base orthonormale dans $L^2(\nu)$, ρ admet alors la décomposition suivante

$$\rho(\cdot) = \sum_{k\geq 0} \alpha_k \langle e_k, e_k b(\cdot)\rangle \; ae_k \tag{3.8}$$

où $\alpha_k = \langle e_k, \xi\rangle$, $\forall k \geq 0$, (cf. chapitre 2, paragraphe 2.4).

Nous posons $\lambda_k = \dfrac{1}{\left\|f_k^*\right\|^2}$ et $u_k^* = \dfrac{f_k^*}{\left\|f_k^*\right\|}$, $\forall k \in \mathbb{N}$ où $\left(f_k^*\right)_{k\in\mathbb{N}}$ est la suite des fonctionnelles de coefficients associée à $(ae_k)_{k\in\mathbb{N}}$. Nous avons en plus des résultats du paragraphe 2.4 du chapitre 2, le lemme suivant :

Lemme 3.4.1. *Les $\left(u_k^*\right)_{k\in\mathbb{N}}$ sont des vecteurs propres normés de l'opérateur aa^* de B^* dans B associés aux valeurs propres $(\lambda_k)_{k\in\mathbb{N}}$.*

Comme dans le chapitre 2 nous considérons que a et b sont connus, donc l'estimation de ρ revient à l'estimation des coéfficients $(\alpha_k)_k$.

Soit $(X_n)_n$ un processus ARB(1) gaussien. Alors pour tout $k \in \mathbb{N}$ les variables aléatoires réelles $(f_k^*, \varepsilon_n)_n$ sont gaussiennes indépendantes et de même variance σ_k^2. Dans ce cas la densité de transition $g(x, y, \rho)$ du processus (X_n) s'écrit :

$$g(X_n, X_{n-1}, \rho) = \exp[-\sum_{k=0}^{+\infty} \frac{1}{2\sigma_k^2}((f_k^*, X_n - \rho X_{n-1})^2 - (f_k^*, X_n)^2)]\,.$$

Alors la fonction de vraisemblance s'écrit :

$$\begin{aligned} L(X_1, \ldots, X_n, \rho) &= \textstyle\prod_{i=1}^n g(X_i, X_{i-1}, \rho) \\ &= \prod_{i=1}^n \exp\left[-\sum_{k\in\mathbb{N}} \tfrac{1}{2\sigma_k^2} (f_k^*, X_i - \rho X_{i-1})^2 - (f_k^*, X_i)^2\right] \\ &= \prod_{i=1}^n \exp\left[-\sum_{k\in\mathbb{N}} \tfrac{1}{2\sigma_k^2} \left((f_k^*, \rho X_{i-1})^2 - 2(f_k^*, \rho X_{i-1})(f_k^*, X_i)\right)\right] \end{aligned}$$

d'aprés (3.8) $\rho(\cdot) = \sum\limits_{j\in\mathbb{N}} \alpha_j \; \langle e_j, e_j\, b(\cdot)\rangle \; ae_j$, alors

$$L(X_1, \ldots, X_n, \rho) = \exp\sum_{k\in\mathbb{N}} -\frac{1}{2\sigma_k^2}\left[\sum_{i=1}^n \alpha_k^2 \langle e_k, e_k\, bX_{i-1}\rangle^2 - 2\sum_{i=1}^n \alpha_k \langle e_k, e_k\, bX_{i-1}\rangle (f_k^*, X_i)\right] \tag{3.9}$$

où la série converge puisque les coefficients $(\alpha_k)_k$ sont dans ℓ^2.

Dans ce qui suit nous allons utiliser des cribles similaires à ceux utilisés dans Geman et Hwang (1982) pour un paramètre dans ℓ^2.

Soit $\Theta = L^2(\nu)$. Nous considérons le crible

$$\Theta_m^* = \left\{ \xi \in L^2(\nu) / \xi = \sum_{k=0}^{m} \alpha_k e_k, \ \sum_{k=0}^{m} k^2 |\alpha_k|^2 \leq m \right\}.$$

Proposition 3.4.1. *L'estimateur crible de ξ est la fonction $\widehat{\xi}_m = \sum_{k=0}^{m} \widehat{\alpha}_k e_k$*

$$\text{où } \widehat{\alpha}_k = \frac{\sum_{i=1}^{n} \langle e_k, e_k b x_{i-1} \rangle (f_k^*, x_i)}{\sum_{i=1}^{n} \langle e_k, e_k\, b x_{i-1} \rangle^2 + 2n\lambda k} \qquad k = 0, \ldots, m$$

$$\text{et tel que } \lambda \text{ vérifie } \sum_{k=0}^{m} k^2 \left(\frac{\sum_{i=1}^{n} \langle e_k, e_k\, b x_{i-1} \rangle (f_k^*, x_i)}{\sum_{i=1}^{n} \langle e_k, e_k\, b x_{i-1} \rangle^2 + 2n\lambda k} \right)^2 = m.$$

En appliquant le théorème 3.3.2, nous montrons la convergence p.s. dans $L^2(\nu)$ de $\widehat{\xi}_m$ vers ξ_0, la fonction associée à ρ_0.

Théorème 3.4.1. *Si $m_n = O(n^{\frac{1}{3}-\delta})$ pour $0 < \delta < \frac{1}{3}$, alors*

$$\|\widehat{\xi}_{m_n} - \xi_0\|_{L^2(\nu)} \xrightarrow[n \to +\infty]{p.s.} 0$$

3.4.1 Estimateur de l'opérateur d'un ARB(1) gaussien dans le cas de probabilités équivalentes

Notons P_{X_0} et P_{ε_0} les lois gaussiennes de X_0 et ε_0 , C_{X_0} et C_{ε_0} les opérateurs de covariance respectifs, $\mathcal{H}_{X_0}$ et $\mathcal{H}_{\varepsilon_0}$ les espaces de Hilbert à noyau autoreproduisant associés à P_{X_0} et P_{ε_0}. Nous supposons que P_{X_0} et P_{ε_0} sont équivalentes ($P_{X_0} \sim P_{\varepsilon_0}$). D'aprés Kuo (1975) il existe deux applications j_{X_0} et j_{ε_0} continues et compactes de $\mathcal{H}_{X_0}$ et de $\mathcal{H}_{\varepsilon_0}$ dans B respectivement telles que $C_{X_0} = j_{X_0}\, j_{X_0}^*$ et $C_{\varepsilon_0} = j_{\varepsilon_0}\, j_{\varepsilon_0}^*$, nous avons :

$$P_{X_0} \sim P_{\varepsilon_0} \Longleftrightarrow \left\{ \begin{array}{l} \quad \exists T : \mathcal{H}_{\varepsilon_0} \longrightarrow \mathcal{H}_{X_0} \text{ un isomorphisme linéaire tel que :} \\ i)\ j^*_{X_0} = T j^*_{\varepsilon_0} \\ ii)\ S = T^*T - Id_{\mathcal{H}_{\varepsilon_0}} \text{ est un opérateur de Hilbert-Schmidt de } \mathcal{L}\left(\mathcal{H}_{\varepsilon_0}\right). \end{array} \right.$$

Soit $(\lambda_k, w_k)_{k\in\mathbb{N}}$ les éléments propres de l'opérateur S. $\forall k \in \mathbb{N}$, $\exists!$ w^*_k tel que $j^*_{\varepsilon_0} w^*_k = w_k$. $(w_k)_{k\in\mathbb{N}}$ est une base orthonormale dans $\mathcal{H}_{\varepsilon_0}$. Ces propriétés nous ont permis de déduire le lemme :

Lemme 3.4.2. *L'opérateur ρ admet la décomposition "spectrale" suivante :*

$$\rho = \sum_{k\in\mathbb{N}} \alpha_k \left(w^*_k, \cdot\right)\ j_{\varepsilon_0} w_k$$

où $\dfrac{\alpha_k \overline{\alpha_k}}{1 - \alpha_k \overline{\alpha_k}} = \lambda_k\ ,\forall k \in \mathbb{N}$.

Nous déduisons alors que ρ est un opérateur 2-nucléaire.

Lemme 3.4.3. *Si C_{ε_0} est injective alors $(w^*_k, j_{\varepsilon_0} w_k)_{k\in\mathbb{N}}$ est une base de Markushevich et $(j_{\varepsilon_0} w_k)_{k\in\mathbb{N}}$ est une base de shrinking.*

Exemple : Dans cet exemple nous traitons l'estimation de ρ du lemme 3.4.2 dans le cas où ε_0 est un processus de Wiener. Nous nous inspirons de l'exemple cité dans Antoniadis et Beder (1989) (p. 81).

Soient $B = C\left([0,1]\right)$ muni de la topologie de la convergence uniforme et P_0 la mesure de Wiener de fonction de covariance $\gamma\left(s,t\right) = \min\left(s,t\right)$.

Nous avons :

– $\mathcal{H}_{\varepsilon_0} = \left\{ f \text{ absolument continue } / \int\limits_0^1 \left(f'(t)\right)^2 dt < +\infty \right\}$,

– $B^*\left([0,1]\right)$ est identifié à $M\left([0,1]\right)$ l'espace des mesures régulières de Borel sur $\mathcal{B}_{[0,1]}$,

– $$j^*_{\varepsilon_0}\left(\mu\right)\left(s\right) = C_{\varepsilon_0}\left(\mu\right)\left(s\right) = \int\limits_0^1 \min\left(t,s\right)\ \mu(dt)$$
$$= \int\limits_0^s \mu\left([t,1]\right) dt \quad \mu \in M\left([0,1]\right),\ \ s \in [0,1]\,,$$

– $\forall t \in [0,1]\,,\quad w_0(t) = t \quad$ et $\quad w_k(t) = \dfrac{\sqrt{2}}{k\pi} \sin k\pi t \quad k \neq 0,$

- $w_0^* = \delta_1$ la mesure de Dirac en 1

  et $w_k^*(dt) = (-1)^k \sqrt{2}\delta_1(dt) + \sqrt{2}k\pi \sin k\pi t(dt),\ k \geq 1.$

L'estimateur crible de ρ est alors

$$\widehat{\rho}_m\left(f\right)\left(s\right) = \sum_{k=0}^{m} \widehat{\alpha}_k \left\langle w_k^*, \cdot \right\rangle \ w_k$$

où $\widehat{\alpha}_0 = \dfrac{\sum\limits_{i=1}^{n} x_{i-1}(1)\ x_i(1)}{\sum\limits_{i=1}^{n} x_{i-1}^2(1)}$,

$$\widehat{\alpha}_k = \frac{\sum\limits_{i=1}^{n}\left(\left(-1\right)^k x_{i-1}(1) + k\pi \int\limits_0^1 x_{i-1}(t)\ \sin k\pi t\ dt\right)\left(\left(-1\right)^k x_i(1) + k\pi \int\limits_0^1 x_i(t)\ \sin k\pi t\ dt\right)}{\sum\limits_{i=1}^{n}\left(\left(-1\right)^k x_{i-1}(1) + k\pi \int\limits_0^1 x_{i-1}(t)\ \sin k\pi t\ dt\right)^2 + 2n\lambda_n k},$$

$k = 1, \ldots, m.$

ou encore :

$$\begin{aligned}\widehat{\rho}_m\left(f\right)\left(s\right) &= \int\limits_0^1 \left(\widehat{\alpha}_0 \cdot I_{[0,1]}(s) \cdot +2\sum_{k=1}^{m} \widehat{\alpha}_k \frac{\left(-1\right)^k}{k\pi} \sin k\pi s\right) \cdot f(t) \cdot \delta_1(dt) \\ &\quad + \int\limits_0^1 2f(t)\left(\sum_{k=1}^{m} \widehat{\alpha}_k\ \sin k\pi s \cdot \sin k\pi t\ dt\right) dt.\end{aligned}$$

Nous posons :

$K_1\left(s,t\right) = \alpha_0 \cdot I_{[0,1]}(s) \cdot +2\sum\limits_{k\geq 1} \alpha_k \dfrac{\left(-1\right)^k}{k\pi} \sin k\pi s,$

$K_2\left(s,t\right) = 2\sum\limits_{k\geq 1} \alpha_k\ \sin k\pi s \cdot \sin k\pi t,$

et $K = K_1 + K_2$.

L'opérateur ρ est un opérateur à noyau K par rapport à la mesure $\left(\delta_1 + I_{[0,1]}\right)\eta$ où η est la mesure de Lebesgue.

3.5 Preuves

Preuve du théoème 3.3.1 :

$L(X_1, X_2, ..., X_n; \rho) = \prod_{i=1}^{n} g(X_i, X_{i-1}, \rho)$. Comme $g(x, y, .)$ est continue sur l'espace métrique (Θ, d), la vraisemblance $L(X_1, X_2, ..., X_n; .)$ et l'entropie $H(\rho_0, .)$ sont continues sur Θ_{m_n}, et de l'hypothèse (1) de finitude : $E \ln(g(x, y, B_{m_n}(\rho, \tau)) < \infty$, les ensembles $M_{m_n}^n$ et A_{m_n} sont compacts et non vides p.s. .

Notant $B_{m_n}(A_{m_n}, 1/m_n)$ un $(1/m_n)$-voisinage de A_{m_n} au sens de la distance d. Il suffit de montrer que $M_m^n \subset B_{m_n}(A_{m_n}, 1/m_n)$ presque sûrement pour n assez grand. Par la condition (2) nous aurons le théorème.

L'ensemble $C_{m_n} := \Theta_{m_n} - B_{m_n}(A_{m_n}, 1/m_n)$ étant compact on peut le recouvrir par un nombre fini de boules $B_{m_n}(\rho_{m_n}^k, \tau_{m_n}^k)$ où $\rho_{m_n}^k \in C_{m_n}$, $\tau_{m_n}^k > 0$ pour $k = \overline{1, \ell_{m_n}}$. Pour $\theta_{m_n} \in A_{m_n}$ fixé nous avons

$$
\begin{aligned}
P\left\{M_m^n \not\subset B_{m_n}(A_{m_n}, 1/m_n)\right\} &\leq P\left\{\sup_{\rho \in \Theta_m - B_{m_n}(A_{m_n}, 1/m_n)} L_n(w, \rho) \geq L_n(w, \theta_{m_n})\right\} \\
&\leq P\left\{\sup_{k=1,\dots,\ell_{m_n}} L_n(w, B_{m_n}(\rho_{m_n}^k, \tau_{m_n}^k)) \geq L_n(w, \theta_{m_n})\right\} \\
&\leq P\left\{\sup_{k=1,\dots,\ell_{m_n}} \frac{L_n(w, B_{m_n}(\rho_{m_n}^k, \tau_{m_n}^k))}{L_n(w, \theta_{m_n})} \geq 1\right\}.
\end{aligned}
$$

Soit $\rho_{m_n}^k \in \Theta_{m_n} - B_{m_n}(A_{m_n}, 1/m_n)$ et $\delta_{\rho_{m_n}^k} > 0$ tel que $H(\rho_0, \rho_{m_n}^k) - H(\rho_0, \Theta_m) < -\delta_{\rho_{m_n}^k}$.

Par la définition de A_{m_n} et pour $\theta_{m_n} \in A_{m_n}$ nous avons :

$$
\begin{aligned}
E_{\rho_0} \ln(g(X_i, X_{i-1}, B_{m_n}(\rho_{m_n}^k, \tau_{m_n}^k)) - E_{\rho_0} \ln(g(X_i, X_{i-1}, \theta_{m_n})) = \\
E_{\rho_0} \ln\left\{\frac{g(X_i, X_{i-1}, B_{m_n}(\rho_{m_n}^k, \tau_{m_n}^k))}{g(X_i, X_{i-1}, \theta_{m_n})}\right\} < -\delta_{\rho_{m_n}^k}
\end{aligned}
$$

pour $\tau_{m_n}^k$ assez petit.

Or

$$\begin{aligned}
\ln L_n(\omega; B_{m_n}(\rho^k_{m_n}, \tau^k_{m_n}) - \ln L_n(\omega; \theta_{m_n}) &= \ln \sup_{\theta \in B_{m_n}(\rho^k_{m_n}, \varepsilon^k_{m_n})} \prod_{i=1}^{n} g(X_i, X_{i-1}, \theta) \\
&\quad - \ln \prod_{i=1}^{n} g(X_i, X_{i-1}, \theta_{m_n}) \\
&\leq \ln \prod_{i=1}^{n} g(X_i, X_{i-1}, B_{m_n}(\rho^k_{m_n}, \tau^k_{m_n})) \\
&\quad - \ln \prod_{i=1}^{n} g(X_i, X_{i-1}, \theta_{m_n}) \\
&= \sum_{1}^{n} (\ln\{\frac{g(X_i, X_{i-1}, B_{m_n}(\rho^k_{m_n}, \tau^k_{m_n}))}{g(X_i, X_{i-1}, \theta_{m_n})}\}
\end{aligned}$$

D'autre part, conditionellement à X_{i-1} les lois des variables aléatoires réelles $g(X_i, X_{i-1}, B_{m_n}(\rho^k_{m}$ (et $g(X_i, X_{i-1}, \theta_{m_n})$) sont les images par g des translatées des lois de ε_i qui sont indépendantes identiquement distribuées. Donc par la loi forte des grands nombres pour des variables aléatoires i.i.d. nous avons pour n assez grand :

$$\frac{1}{n} \sum_{1}^{n} (\ln\{\frac{g(X_i, X_{i-1}, B_{m_n}(\rho^k_{m_n}, \tau^k_{m_n}))}{g(X_i, X_{i-1}, \theta_{m_n})}\} = E_{\rho_0} \ln\{\frac{g(X_1, X_0, B_{m_n}(\rho^k_{m_n}, \tau^k_{m_n}))}{g(X_1, X_0, \theta_{m_n})}\} < -\delta_{\rho^k_{m_n}}. \quad p.s.$$

Par conséquent presque sûrement pour n assez grand :

$$\frac{L_n(w, B_{m_n}(\rho^k_{m_n}, \tau^k_{m_n}))}{L_n(w, \theta_{m_n})} < \exp(-n\delta_{\rho^k_{m_n}})$$

Pour $\delta = \min_{k=\overline{1,\ell_{m_n}}} \delta_{\rho^k_{m_n}}$ et n assez grand, nous avons alors p.s. :

$$\sup_{k=\overline{1,\ell_{m_n}}} \frac{L_n(w, B_{m_n}(\rho^k_{m_n}, \tau^k_{m_n}))}{L_n(w, \theta_{m_n})} < \exp(-n\delta)$$

Donc il existe un entier N_{m_n} tel que pour tout $n \geq N_{m_n}$:

$$P\left\{ \sup_{k=\overline{1,\ell_{m_n}}} \frac{L_n(w, B_{m_n}(\rho^k_{m_n}, \tau^k_{m_n}))}{L_n(w, \theta_{m_n})} \geq 1 \right\} \leq \frac{1}{m_n^2}$$

Et finalement de (3.10) nous avons

$$P\left\{M_m^n \not\subset B_{m_n}(A_{m_n}, 1/m_n)\right\} \leq \frac{1}{m_n^2}$$

Ensuite il suffit d'appliquer le lemme de Borel-Cantelli.

Preuve du théoème 3.3.2 : Pour montrer le résultat demandé il suffit de montrer que pour $\delta > 0$ nous avons p.s. pour n assez grand :

$$D_{m_n} \cap M_{m_n}^n = \emptyset \tag{3.10}$$

En effet, si (3.10) est vérifié, nous avons presque sûrement pour n assez grand :

$$\inf_{\rho \in M_{m_n}^n} H(\rho_0, \rho) \geq H(\rho_0, \rho_{m_n}) - \delta$$

Par la condition $\mathbf{C_2}$, nous avons $H(\rho_0, \rho_{m_n}) \rightarrow H(\rho_0, \rho_0)$. Comme δ est arbitaire ceci implique que

$\liminf \inf_{\rho \in M_{m_n}^n} H(\rho_0, \rho) \geq H(\rho_0, \rho_0) \quad$ p.s. ,

de plus $H(\rho_0, \rho) \leq H(\rho_0, \rho_0)$ alors :

$$\lim_{n \to +\infty} \sup_{\rho \in M_{m_n}^n} |H(\rho_0, \rho) - H(\rho_0, \rho_0)| = 0 \quad ps \tag{3.11}$$

Soit $\varepsilon > 0$. Par la définition de $M_{m_n}^n$, pour tout n choisissant $\theta_n \in M_{m_n}^n$ tel que

$$\frac{d(\rho_0, \theta_n)}{1 + d(\rho_0, \theta_n)} > \sup_{\rho \in M_{m_n}^n} \frac{d(\rho_0, \rho)}{1 + d(\rho_0, \rho)} - \varepsilon \quad p.s.$$

Par le résultat (3.11) et la condition $\mathbf{C_1}$ nous obtenons alors $d(\rho_0, \theta_n) \rightarrow 0 \quad p.s.$

Par suite

$$\limsup_n \sup_{\rho \in M_{m_n}^n} \frac{d(\rho_0, \rho)}{1 + d(\rho_0, \rho)} \leq \epsilon \quad p.s.$$

Comme ϵ est quelconque nous avons alors :

$$\limsup_n \sup_{\rho \in M_{m_n}^n} d(\rho_0, \rho) = 0 \qquad p.s.$$

donc

$$M_{m_n}^n \to \rho_0 \qquad p.s.$$

Montrons maintenant (3.10). D'aprés la condition (i), pour n et m fixé on a :

$$\begin{aligned}(D_{m_n} \cap M_{m_n}^n \neq \emptyset) \quad &\subseteq \left\{ \sup_{\rho \in D_m} L_n(X_1, ..., X_n, \rho) \geq L_n((X_1, ..., X_n), \rho_{m_n}) \right\} \\ &\subseteq \bigcup_{k=1}^{l_{m_n}} \left\{ \sup_{\rho \in \Gamma_k^n} \prod_{i=1}^n g(X_i, X_{i-1}, \rho) \geq \prod_{i=1}^n g(X_i, X_{i-1}, \rho_{m_n}) \right\} \\ &\subseteq \bigcup_{k=1}^{l_{m_n}} \left\{ \prod_{i=1}^n g(X_i, X_{i-1}, \Gamma_k^{m_n}) \geq \prod_{i=1}^n g(X_i, X_{i-1}, \rho_{m_n}) \right\}\end{aligned}$$

Alors la probabilité de l'ensemble $\left\{D_{m_n} \cap M_{m_n}^n \neq \emptyset\right\}$, qu'on note π, vérifie :

$$\begin{aligned}\pi \quad &\leq \sum_{k=1}^{l_{m_n}} P\left(\prod_{i=1}^n g(X_i, X_{i-1}, \Gamma_k^{m_n}) \geq \prod_{i=1}^n g(X_i, X_{i-1}, \rho_{m_n}) \right) \\ &= \sum_{k=1}^{l_{m_n}} P(\exp \sum_{i=1}^n [t_k \ln \left\{ \frac{g(X_i, X_{i-1}, \Gamma_k^{m_n})}{g(X_i, X_{i-1}, \rho_{m_n})} \right\}] \geq 1) \\ &\leq \sum_{k=1}^{l_{m_n}} \left(E_{\rho_0} \exp[t_k \ln \left\{ \frac{g(X_i, X_{i-1}, \Gamma_k^{m})}{g(X_i, X_{i-1}, \rho_{m_n})} \right\}] \right)^n\end{aligned}$$

Pour tout $t_1, ..., t_k$ arbitraires positifs et conditionnelemment à X_{i-1} les lois des variables aléatoires réelles $g(X_i, X_{i-1}, \Gamma_k^m)$ (et $g(X_i, X_{i-1}, \rho_{m_n})$) sont les translatées des images par g des lois de ε_i qui sont indépendantes identiquement distribuées. Finalement

$$\pi \leq l_{m_n} (\varphi_{m_n})^n.$$

Nous déduisons alors le résultat (3.11) par la condition (ii) et le lemme de Borel Cantelli.

Preuve du lemme 3.4.1 : $\forall i, j \in \mathbb{N} \quad \left(f_j^*, ae_i\right) = \left(a^* f_j^*, e_i\right) = \delta_{ij}$, alors

$a^* f_j^* = e_j \ \forall j \in \mathbb{N}$.

Donc $a^* u_j^* = \sqrt{\lambda_j}\, e_j \ \forall j \in \mathbb{N}$.

Ainsi on obtient

$\left(a^* u_j^*, e_i\right) = \left(a^* u_j^*, a^* u_i^*\right) = \left(a a^* u_j^*, u_i^*\right) = \lambda_j\, \delta_{ij}\ , \forall j, i \in \mathbb{N}.$

■

Preuve de la proposition 3.4.1 :

Le problème est un problème d'optimisation sous contrainte. En notant alors par λ le multiplicateur de Lagrange, celà revient à résoudre :

$$\max_{\Theta_{m_n}} \sum_{k=0}^{m} \sum_{i=1}^{n} [\alpha_k \langle e_k, e_k\, bx_{i-1}\rangle (f_k^*, x_i) - \frac{1}{2}\alpha_k^2 \langle e_k, e_k\, bx_{i-1}\rangle^2] - \lambda k^2 \alpha_k^2$$

Par dérivation par rapport à α_k, $k \in \mathbb{N}$, nous obtenons l'expression de $\widehat{\alpha}_k$ ennoncée, pour $k = 1, ..., m$. ■

Preuve du théoème 3.4.1 :

La preuve consiste à vérifier les conditions du théorème (3.3.2). Rappellons que l'existence de l'estimateur crible $\widehat{\xi}$ découle de la proposition (3.4.1). Ainsi l'ensemble M_{m_n} est non vide.

Pour la condition $\mathbf{C_1}$, par la définition de l'entropie H et en notant $\alpha_{k,0}$ les coéfficients de la fonction ξ_0, nous avons pour $\xi \in \Theta$:

$$\begin{aligned}
H(\xi_0, \xi) - H(\xi_0, \xi_0) &= \sum_{k=0}^{+\infty} \frac{1}{\sigma_k^2} E[\alpha_k \langle e_k, e_k\, bX_{n-1}\rangle (f_k^*, X_n) - \frac{1}{2}\alpha_k^2 \langle e_k, e_k\, bX_{n-1}\rangle^2 \\
&\quad -\alpha_{k,0} \langle e_k, e_k\, bX_{n-1}\rangle (f_k^*, X_n) + \frac{1}{2}\alpha_{k,0}^2 \langle e_k, e_k\, bX_{n-1}\rangle^2] \\
&= \sum_k \frac{1}{\sigma_k^2} E[E[(f_k^*, X_n) (\alpha_k \langle e_k, e_k\, bX_{n-1}\rangle - \alpha_{k,0} \langle e_k, e_k\, bX_{n-1}\rangle) \\
&\quad + \frac{1}{2}(\alpha_{k,0} \langle e_k, e_k\, bX_{n-1}\rangle - \alpha_k \langle e_k, e_k\, bX_{n-1}\rangle)^2 \\
&\quad - \frac{1}{2}\alpha_k \langle e_k, e_k\, bX_{n-1}\rangle (\alpha_k \langle e_k, e_k\, bX_{n-1}\rangle \\
&\quad -\alpha_{k,0} \langle e_k, e_k\, bX_{n-1}\rangle)]/ \langle e_k, e_k\, bX_{n-1}\rangle] \\
&= \sum_k \frac{1}{4\sigma_k^2} E(\alpha_{k,0} \langle e_k, e_k\, bX_{n-1}\rangle - \alpha_k \langle e_k, e_k\, bX_{n-1}\rangle)^2
\end{aligned}$$

Par conséquent quand $n \to \infty$, si $H(\xi_0, \xi) - H(\xi_0, \xi_0) \to 0$, $\alpha_k \to \alpha_{k,0}$.
D'où $\xi \to \xi_0$ dans $L^2(\nu)$.

Pour $\mathbf{C}_2$, comme $\bigcup \Theta^*_{m_n}$ est dense dans $\Theta = L^2(\nu)$, nous avons :

pour tout $\xi_0 \in \Theta$, il existe $\xi_{m_n} \in \Theta^*_{m_n}$ tel que $\xi_{m_n} \to \xi_0$ dans $L^2(\nu)$.

Donc $\alpha_{k,m_n} \to \alpha_{k,0}$.

Nous déduisons alors que $H(\xi_0, \xi_{m_n}) - H(\xi_0, \xi_0) \to 0$.

Vérifions maintenant la condition (i) du théorème 3.3.2.

Pour $\xi \in \Theta^*_{m_n}$, vu la contrainte, ces coefficients $(\alpha_k)_k$ vérifient :

$$|\alpha_k| \leq \frac{\sqrt{m_n}}{k} \qquad \text{pour } k \neq 0 \text{ et } \alpha_0 \leq \sqrt{m_n}.$$

Pour $k \neq 0$, on subdivise l'intervalle $[-\frac{\sqrt{m_n}}{k}, \frac{\sqrt{m_n}}{k}]$ en $[m_n^2/k + 1]$ intervalles de même longueur, [ ] désignant la partie entière.

Soit I_k l'ensemble des extrémités de ces intervalles.

Pour $k = 0$, I_0 s'obtient de la subdivision de $\left[-\sqrt{m_n}, \sqrt{m_n}\right]$ en $[m_n^2 + 1]$ intervalles de même longueur.

Notons l la longueur de ces intervalles. Nous avons alors $l \leq \dfrac{2}{m_n^{3/2}}$.

Pour tout $\varepsilon > 0$, on associe à tout ensemble :

$$\{b_k \quad / \quad b_k \in I_k, \quad k = 0, 1, ..., [m_n^{1+\varepsilon}]\}$$

l'ensemble

$$\hat{\Gamma}^{m_n}(\{b_k\}) = \{\xi \in \Theta_{m_n} \quad / \quad |\alpha_k - b_k| \leq 2m_n^{-1.5}, \quad k = 0, 1, ..., [m_n^{1+\varepsilon}]\}$$

Comme $|\alpha_k| \leq \dfrac{\sqrt{m_n}}{k}$, on choisit

$$\alpha_k = -\frac{\sqrt{m_n}}{k} + \frac{m_n}{k^2}p, \quad p = 0, 1, ... \tag{3.12}$$

Donc

$$\begin{aligned} |\alpha_k| &\leq \frac{\sqrt{m_n}}{k} \Rightarrow -\frac{\sqrt{m_n}}{k} + \frac{m_n}{k^2}p \leq \frac{\sqrt{m_n}}{k} \\ &\Rightarrow p \leq \frac{2k\sqrt{m_n}}{m_n} \leq 2(\frac{m_n+1}{m_n})m_n \\ &\Rightarrow p \leq cm_n \\ &\Rightarrow p^{(1+m_n^{1+\varepsilon})} \leq cm^{cm_n^{1+\varepsilon}} \end{aligned}$$

où c désigne une constante positive. Le nombre de fonctions ξ dont les coefficients α_k sont de la forme (3.12), ne dépasse pas $p^{(1+m_n^{1+\varepsilon})} \leq cm_n^{cm_n^{1+\varepsilon}}$.

D'après ce qui précède, $\bigcup_{i=1}^{l_{m_n}} \hat{\Gamma}_i^{m_n}$ recouvre Θ_{m_n} où

$$l_{m_n} \leq (cm_n)^{cm_n^{1+\varepsilon}} \tag{3.13}$$

Finalement en posant

$\Gamma_i^{m_n} = \hat{\Gamma}_i^{m_n} \cap D_{m_n}, \quad i = 1, ..., l_{m_n}$,

$(\Gamma_i^{m_n})_i$ recouvre D_{m_n}.

D'où la condition (i) du Thórème (3.3.2).

Enfin pour la condition (ii), soit k fixé, nous posons :

$$\varphi(t) := E(\exp[t \ln \left\{ \frac{g(X_n, X_{n-1}, \Gamma_k^{m_n})}{g(X_n, X_{n-1,}, \phi_{m_n})} \right\}])$$

Alors $\varphi(0) = 0$, $\varphi'(0) = E \ln \left\{ \frac{g(X_n, X_{n-1}, \Gamma_k^{m_n})}{g(X_n, X_{n-1}, \phi_{m_n})} \right\}$.

Or $\varphi(\Gamma_k^{m_n}) := \sup_{\beta \in \Gamma_k^{m_n}} \varphi(\beta)$.

Donc pour ξ fixée dans $\Gamma_k^{m_n}$, en posant :

$$A := E \ln g(X_n, X_{n-1}, \Gamma_k^{m_n}) - E \ln g(X_n, X_{n-1}, \phi)$$

nous avons

$A = E \sup_{\beta \in \Gamma_k^{m_n}} \sum_{k=0}^{+\infty} \frac{1}{2\sigma_k^2} [(\alpha_k \langle e_k, e_k \, bX_{n-1} \rangle - b_k \langle e_k, e_k \, bX_{n-1} \rangle)(\alpha_k \langle e_k, e_k \, bX_{n-1} \rangle$

$+b_k \langle e_k, e_k\, bX_{n-1}\rangle - 2\,(f_k^*, X_n))]$

où (b_k) sont les coefficients de $\beta \in \Gamma_k^{m_n} \subset \Theta_{m_n}$ dans la base $(e_k)_k$.

Comme $\{\hat{\Gamma}_k^{m_n}\}_k$ recouvre Θ_m, pour $\xi, \beta \in \Gamma_k^{m_n}$ nous avons :

$$\begin{aligned}\sup|\xi(x) - \beta(x)| &\leq \sum_{k=0}^{m_n} |\alpha_k - \beta_k| \\ &\leq \frac{2}{m_n^{1/2}}.\end{aligned}$$

Donc

$$\begin{aligned}A &\leq \frac{c(m_n^{1+\varepsilon})^{1/2} m_n}{m_n^{1/2} m_n^{1+\varepsilon}} \\ &\leq \frac{c}{m_n^{\varepsilon/2}}\end{aligned}$$

où c désigne une constante.

Finalement

$$\begin{aligned}\varphi'(0) &\leq H(\phi_0, \phi) - H(\phi_0, \phi_{m_n}) + \frac{c}{m_n^{\varepsilon/2}} \\ &\leq cm_n^{-\varepsilon/2} - \delta\end{aligned}$$

D'autre part, par un résultat de Geman et Hwang (1982), pour $t \in [0, 1/cm_n]$, nous avons :

$$\varphi''(t) \leq cm_n^2.$$

Par un développement de $\varphi(t)$ nous obtenons donc :

$$\begin{aligned}\varphi(t) &\leq 1 + t(cm_n^{-\varepsilon/2} - \delta) + t^2 cm_n^2 \\ \varphi(\frac{1}{m_n^2}) &\leq 1 + \frac{c}{m_n^{2+\varepsilon/2}} - \frac{\delta}{m_n^2} + \frac{c}{m_n^2} \\ &\leq 1 + \frac{c}{m_n^2} - \frac{\delta}{m_n^2}\end{aligned}$$

ou encore

$$\varphi(\frac{1}{m_n^2}) \leq 1 - \frac{\delta}{cm_n^2}, \quad c > 0 \tag{3.14}$$

Or

$$\varphi_{m_n} := \sup_k \inf_{t \geq 0} E_{\rho_0}[\exp[t \ln \left\{ \frac{g(X_n, X_{n-1}, \Gamma_k^{m_n})}{g(X_n, X_{n-1}, \phi_{m_n})} \right\}]]$$

Donc pour m_n suffisamment grand et d'après (3.13) et (3.14) nous obtenons :

$$l_{m_n}(\varphi_{m_n})^n \leq cm_n^{cm_n^{1+\varepsilon}}(1 - \frac{\delta}{cm_n^2})^n,$$

la série $\sum_n m_n^{cm_n^{1+\varepsilon}}(1 - \frac{\delta}{cm_n^2})^n$ est convergente si $m_n = (n^{\frac{1}{3}-\delta})$.

Par conséquent $M_{m_n}^n \to \xi_0$ p.s.

Comme $\widehat{\xi} \in M_{m_n}^n$ nous avons alors la convergence p.s. par rapport à la norme de $L^2(\nu)$.

■

Preuve du lemme 3.4.2 : Comme $j_{\varepsilon_0} S j_{\varepsilon_0}^* = j_{\varepsilon_0} T^* T j_{\varepsilon_0}^* - j_{\varepsilon_0} j_{\varepsilon_0}^* = C_{X_0} - C_{\varepsilon_0}$. D'aprés (Bosq (2000) p. 149) $C_{X_0} - C_{\varepsilon_0} = \rho C_{X_0} \rho^*$ alors $j_{\varepsilon_0} S j_{\varepsilon_0}^* = \rho C_{X_0} \rho^*$.
L'opérateur S admet la décomposition unique :

$$\sum_{k \in \mathbb{N}} \lambda_k \langle w_k, \cdot \rangle \; w_k,$$

alors $\rho C_{X_0} \rho^*$ admet la décomposition :

$$\sum_{k \in \mathbb{N}} \lambda_k (\cdot, j_{\varepsilon_0} w_k) \quad j_{\varepsilon_0} w_k.$$

On déduit que $\rho C_{X_0} \rho^*$ de B^* dans B est 2−nucléaire.

Alors $\rho C_{X_0} \rho^* = cM_\lambda d$,
où $c\left((\xi_k)_k\right) = \sum_{k \in \mathbb{N}} \xi_k \; j_{\varepsilon_0} w_k$, $(\xi_k)_k \in \ell^2$ et $dx^* = \left((x^*, j_{\varepsilon_0} w_k)\right)_k$, $x^* \in B^*$.

$\forall x^* \in B^*,\ c^*x^* = ((x^*, j_{\varepsilon_0} w_k))_k$, ainsi le diagramme suivant commute

$$\begin{array}{ccc} B^* & \overset{\rho C_{X_0}\rho^*}{\longrightarrow} & B \\ c^* \downarrow & & \uparrow c \\ \ell^2 & \overset{M_\lambda}{\longrightarrow} & \ell^2 \end{array} \tag{3.15}$$

D'aprés le lemme (2.6.2), $\rho = \sum\limits_{k\in\mathbb{N}} \alpha_k \left(h_k^*, \cdot\right) v_k = aM_\alpha b.$

Donc $\rho C_{X_0}\rho^* = aM_\alpha bC_{X_0} b^* M_{\overline{\alpha}} a^*$. tel que le diagramme suivant commute

$$\begin{array}{ccccccc} B^* & \overset{\rho^*}{\longrightarrow} & B^* & \overset{C_{X_0}}{\longrightarrow} & B & \overset{\rho}{\longrightarrow} & B \\ a^* \downarrow & & \uparrow b^* & & b\downarrow & & \uparrow a \\ \ell^2 & \overset{M_{\overline{\alpha}}}{\longrightarrow} & \ell^1 & & \ell^\infty & \overset{M_\alpha}{\longrightarrow} & \ell^2 \end{array} \tag{3.16}$$

Nous supposons que $a = c$ et nous montrons qu'il existe $\alpha = (\alpha_k)_{k\in\mathbb{N}}$ tel que $\rho C_{X_0}\rho^* = aM_\lambda a^*$, c'est è dire $M_\lambda = M_\alpha bC_{X_0} b^* M_{\overline{\alpha}}$.

Soit $\xi = (\xi_k)_k, \eta = (\eta_k)_k \in \ell^2$, comme $(\xi, bx) = \sum \xi_k (x, h_k^*) = \left(x, \sum \overline{\xi}_k h_k^*\right) \forall x \in B$, nous avons :

$$\begin{aligned} \langle M_\lambda \xi, \eta\rangle &= \langle M_\alpha bC_{X_0} b^* M_{\overline{\alpha}}\xi, \eta\rangle = \langle C_{X_0} b^* M_{\overline{\alpha}}\xi, b^* M_{\overline{\alpha}}\eta\rangle \\ &= \sum_{k\in\mathbb{N}} (1+\lambda_k)\left(b^* M_{\overline{\alpha}}\xi, j_{\varepsilon_0} w_k\right)\left(j_{\varepsilon_0} w_k, b^* M_{\overline{\alpha}}\eta\right) \\ &= \sum_{k\in\mathbb{N}} (1+\lambda_k)\left(M_{\overline{\alpha}}\xi, bj_{\varepsilon_0} w_k\right)\left(bj_{\varepsilon_0} w_k, M_{\overline{\alpha}}\eta\right) \\ &= \sum_{j,s\in\mathbb{N}} \overline{\alpha}_j \alpha_s \xi_j \overline{\eta}_s \sum_{k\in\mathbb{N}} (1+\lambda_k)\left(h_s^*, j_{\varepsilon_0} w_k\right)\left(j_{\varepsilon_0} w_k, h_j^*\right) \\ &= \sum_{j,s\in\mathbb{N}} \overline{\alpha}_j \alpha_s \xi_j \overline{\eta}_s \left(C_{X_0} h_s^*, h_j^*\right) \\ &= \sum_{s\in\mathbb{N}} \lambda_s \xi_s \overline{\eta}_s \end{aligned}$$

comme cette égalité est vérifiée $\forall (\xi_k)_k$ et $\forall (\eta_k)_k$ alors $\overline{\alpha}_j\alpha_s \left(C_{X_0} h_s^*, h_j^*\right) = \lambda_s \delta_{js}$, $\forall i, k \in \mathbb{N}$, cela veut dire que $(h_k^*)_{k\in\mathbb{N}}$ sont colinéaires aux $(w_k^*)_{k\in\mathbb{N}}$, c'est à dire $\exists (\gamma_k)_k$ une suite de scalaires bornée telle que

$v_k^* = \gamma_k w_k^*$, $\forall k \in \mathbb{N}$.

Alors $\alpha_i \overline{\alpha}_k \gamma_i \overline{\gamma}_k (1+\lambda_k) = \lambda_k \delta_{ik}$, $\forall k \in \mathbb{N}$,

d'où $\lambda_k = \dfrac{\gamma_k\alpha_k\overline{\gamma_k\alpha_k}}{1-\gamma_k\alpha_k\overline{\gamma_k\alpha_k}}$, $\forall k \in \mathbb{N}$.

Donc $\rho = \sum\limits_{k\in\mathbb{N}} \gamma_k \alpha_k \left(w_k^*, \cdot\right)\ j_{\varepsilon_0} w_k = \sum\limits_{k\in\mathbb{N}} \alpha'_k \cdot \left(w_k^*, \cdot\right)\ j_{\varepsilon_0} w_k$, comme $(\gamma_k)_k$ est bornée, nous avons :

$$\alpha' = (\alpha'_k)_{k\in\mathbb{N}} \in \ell^2 \text{ tel que } \lambda_k = \frac{\alpha'_k \overline{\alpha'}_k}{1-\alpha'_k\overline{\alpha'}_k} \ , \forall k \in \mathbb{N} \ ,$$

et comme $(\lambda_k)_{k\in\mathbb{N}} \in \ell_c^2 = \left\{(\alpha_k)_{k\in\mathbb{N}} \,/\, (1+\alpha_k) > 0, \ k \geq 0\right\}$, $1+\lambda_k = \dfrac{1}{1-\alpha'_k\overline{\alpha'}_k} > 0,$

nous avons alors $|\,\alpha'_k|^2 < 1$, $\forall k \in \mathbb{N}$. ■

Preuve du lemme 3.4.3 : Soit $x^* \in \overline{\sigma}\left\{w_k^*, k \geq 0\right\}$ l'espace fermé engendré par $(w_k^*)_{k\in\mathbb{N}}$, alors $\exists\,(\alpha_k)_{k\in\mathbb{N}}$ une suite de scalaires telle que $x^* = \sum\limits_{k\in\mathbb{N}} \alpha_k w_k^*$.

D'une part $j_{\varepsilon_0}^* x^* = \sum\limits_{k\in\mathbb{N}} \alpha_k j_{\varepsilon_0}^* w_k^* = \sum\limits_{k\in\mathbb{N}} \alpha_k w_k$

et comme $\mathcal{H}_\mu$ est un Hilbert $j_{\varepsilon_0}^* x^*$ s'écrit $\sum\limits_{k\in\mathbb{N}} \left(j_{\varepsilon_0}^* x^*, w_k\right) w_k$ d'une façon unique, alors $\alpha_k = (x^*, j_{\varepsilon_0} w_k)\,, \forall k \in \mathbb{N}$. Donc $(w_k^*)_{k\in\mathbb{N}}$ forme une base de Schauder de son enveloppe linéaire dans B^*.

De même on montre que $(j_{\varepsilon_0} w_k)_{k\in\mathbb{N}}$ forme une base de Schauder de son enveloppe linéaire dans B :

soit $y \in \overline{\sigma}\left\{j_{\varepsilon_0} w_k, k \in \mathbb{N}\right\}$, alors $\exists\,(\gamma_k)_{k\in\mathbb{N}}$ une suites de scalaires tel que $y = \sum\limits_{k\in\mathbb{N}} \gamma_k j_{\varepsilon_0} w_k$ donc $(w_k^*, y) = \sum\limits_{k\in\mathbb{N}} \gamma_j \left(w_k^*, j_{\varepsilon_0} w_j\right) = \gamma_k, \forall k \in \mathbb{N}.$

Nous supposons que C_{ε_0} est injective, comme $C_{\varepsilon_0} = j_{\varepsilon_0} j_{\varepsilon_0}^*$, alors $\forall x^* \in B^*$:

$$C_{\varepsilon_0} x^* = j_{\varepsilon_0} j_{\varepsilon_0}^* x^* = j_{\varepsilon_0} \left(\sum_{k\in\mathbb{N}} \left(j_{\varepsilon_0}^* x^*, w_k\right) w_k\right) = \sum_{k\in\mathbb{N}} \left(j_{\varepsilon_0}^* x^*, w_k\right) j_{\varepsilon_0} w_k$$

de plus $C_{\varepsilon_0} w_k^* = j_{\varepsilon_0} w_k \ \forall k \in \mathbb{N}$, alors

$$C_{\varepsilon_0} x^* = \sum_{k\in\mathbb{N}} \left(j_{\varepsilon_0}^* x^*, w_k\right) C_{\varepsilon_0} w_k^* = C_{\varepsilon_0} \left(\sum_{k\in\mathbb{N}} \left(j_{\varepsilon_0}^* x^*, w_k\right) w_k^*\right),$$

nous avons donc

$$x^* = \sum_{k\in\mathbb{N}} \left(j_{\varepsilon_0}^* x^*, w_k\right) w_k^* = \sum_{k\in\mathbb{N}} \left(x^*, j_{\varepsilon_0} w_k\right) w_k^*$$

Ainsi $(w_k^*)_{k\in\mathbb{N}}$ forme une base dans B^*, nous déduisons que $(j_{\varepsilon_0} w_k)_{k\in\mathbb{N}}$ est une base de shrinking. Comme $(w_k^*)_{k\in\mathbb{N}}$ et $(j_{\varepsilon_0} w_k)_{k\in\mathbb{N}}$ sont biorthogonales alors $(w_k^*, j_{\varepsilon_0} w_k)_{k\in\mathbb{N}}$ est une base de Markushevich. ■

Perspectives

L'interprétation d'un processus à temps continu comme élément aléatoire à valeurs dans un espace fonctionnel s'est avérée être une approche fructueuse pour aborder des problèmes dans des espaces de dimension infinie. Les résultats établis dans ce travail ouvrent un champ plus vaste pour les applications. Nous avons montré la consistance de l'opérateur d'autocoprrélation ρ en norme $p-$intégrale avec $p \neq 2$, un résultats qui se distingue ainsi de ceux qui existent. Nous avons aussi généraliser la méthode des moindres carrés comme problème d'estimation dans un espace de Banach, ce qui permet de résoudre les problèmes d'estimation de paramétres banachiques. Néanmoins dans ce travail, plusieurs questions demeurent :

- Dans quel cas la solution du problème d'optimisation vectorielle 2.5 existe.
- Si les opérateurs a et b du diagramme 2.3 sont inconnus, comment peut-on estimer l'opérateur ρ ?
- Est-il possible d'obtenir une décomposition de ρ sans que l'hypothèse **(H)** (paragraphe 2.4) soit verifiée.
- Enfin, sous quelles conditions l'estimateur crible du minimum des ϕ-divergences de ρ convergence dans le cas général.

Appendice

3.6 Sur l'ordre de grandeur des fonctions sommables

Nous rappelons dans ce paragraphe les résultats sur l'ordre de grandeur des fonctions sommables donnés par Gál et Koksma (1948).

Théorème 3.6.1. *Soient $(X_i, i \geq 1)$ une suite de variables aléatoires centrées à valeurs dans un espace de Banach séparable B et $p \geq 1$.*
Supposons qu'il existe deux suites réelles $(\phi(n,m))_{n,m\geq 0}$ et $(\varphi(m))_{m\geq 0}$ telles que :

i) $E\left\|X_n + \ldots + X_{n+m-1}\right\|^p \leq \phi(n,m), \qquad n \geq 1,\ h \geq 1,$

ii) $\phi(1,m)$ *est non décroissante,*

iii) *il existe une suite $(\psi(k), k \geq 0)$ à valeurs entières positives, croissante vers ∞ telle que :*
$\left\|X_1 + \ldots + X_{\psi(k)}\right\| = o\left((\phi(1,\psi(k)))^{1/p}\varphi(\psi(k))\right) \quad p.s.$

iv) *Soient d'abord les notations suivantes :*

a) $\Omega(h)$ une suite positive croissante, à valeurs entières pour $h = 0, 1, ..., H$, telle que $\Omega(0) = 1$ et $\Omega(H) \geq H$, et $L = L_H$ l'entier positif tel que $\Omega(L) < H \leq \Omega(L+1)$.

b) Posons, pour $K = 0$, $\mu_{L-\lambda} = 0$, $\lambda = 0, ..., L$;
et pour $1 \leq K < H$, soit $\mu_L \geq 0$ l'entier tel que $\mu_L\Omega(L) \leq K < (\mu_L + 1)\Omega(L)$.

c) Soit $K_0 = 0$ et $K_\lambda = \sum_{i=0}^{\lambda-1} \mu_{L-i}\Omega(L-i)$, si $\lambda \geq 1$,
choisissons l'entier $\mu_{L-\lambda}$ tel que $\mu_{L-\lambda}\Omega(L-\lambda) \leq K - K_\lambda < (\mu_{L-\lambda}+1)\Omega(L-\lambda)$, où $1 \leq \lambda \leq L$.

La représentation unique $K=\sum_{\lambda=0}^{L}\mu_{L-\lambda}\Omega(L-\lambda)$ *est dite représentation canonique.*

d) T_λ *l'ensemble des entiers* K_λ *qui figurent dans la représentation canonique d'au moins un* K $(0\le K<H)$. *Soit enfin* κ *un entier positif donné et considérons les* K $(0\le K<H)$ *tels que* $K_\lambda=\kappa$. *On désigne le plus grand des* $\mu_{L-\lambda}\in K_\lambda$ *par* $\mu_{L-\lambda}(\kappa)$.

e) $\Lambda_k=1$ *pour* $p=1$, *et*

$$\Lambda_k(K)=\sum_{\lambda=1}^{L(\Delta)}\left\{\frac{\phi(0,\psi(k)+K_\lambda)^{\frac{1}{p}}\varphi(\psi(k)+K_\lambda)}{\phi(0,\psi(k)+K)^{\frac{1}{p}}\varphi(\psi(k)+K)}\right\}^{\frac{p}{p-1}},$$

pour $p>1$.

f) $\Delta=\Delta(k)=\psi(k+1)-\psi(k)$.

g) $D_k=\max\Lambda_k(K)$ *pour* $0\le K<\Delta(k)$ *et*

$$R_k=\sum_{\lambda=0}^{L(\Delta)}\sum_{T_\lambda}\sum_{\mu=1}^{\mu_{L-\lambda}}\frac{\phi(\psi(k)+K_\lambda,\mu\Omega_k(L-\lambda))}{\phi(0,\psi(k)+K_\lambda+\mu\Omega_k(L-\lambda))\varphi(\psi(k)+K_\lambda+\mu\Omega_k(L-\lambda))^p}.$$

Supposons maintenant que $\sum_{k\ge1}R_k$ *converge.*

On conclut alors :

$$\|X_1+\ldots+X_n\|=o\left(\phi(1,n)^{\frac{1}{p}}\varphi(n)\right)\quad p.s.$$

Remarque : On peut remplacer la condition iii) par l'hypothèse que $\sum_{k\ge1}\varphi(\psi(k))^{-p}$ converge.

Dans le cas particulier où $\phi(n,m)=m^\gamma$, $0<\gamma\le2$, on prendra $\psi(k)=2^k$ et $\chi(h)=2^h$, on a alors le lemme suivant

Lemme 3.6.1. *Soit* $(X_i,i\ge1)$ *une suite de variables aléatoires centrées à valeurs dans un espace de Banach separable* B. *Soit* $p\in]1,\infty[$,

Supposons que $E\|X_1 + \ldots + X_{n+m-1}\|^p \leq cm^\gamma$,
$n \geq 1,\ m \geq 1,$ *où* $c > 0$ *et* $0 < \gamma < 2$ *sont des constantes. Alors* $\forall \beta > 1$,

$$n^{(1-\gamma)/p}\,(\log n)^{-\beta/p}\,\|\overline{X}_n\| \underset{n\to+\infty}{\xrightarrow{p.s.}} 0\,. \qquad \blacksquare$$

3.7 Type et cotype d'un espace de Banach

Soient B un espace de Banach et $p \in]1,2]$ ($q \in [2,\infty[$ respectivement). B est de type p (respectivement cotype q) s'il existe une constante strictement positive c_p (respectivement c_q) telle que pour toute suite $X_1, ..., X_n$ de variables aléatoires centrées indépendantes à valeurs dans B qui vérifient $E\|X_i\|^p < \infty$ (respectivement $E\|X_i\|^q < \infty$), on a :

$$E\|\sum_{i=1}^{n} X_i\|^p \leq c_p \sum_{i=1}^{n} E\|X_i\|^p$$

(respectivement

$$E\|\sum_{i=1}^{n} X_i\|^q \geq c_q \sum_{i=1}^{n} E\|X_i\|^q),$$

(Ledoux et Talagrand (1991)).

Les espaces ℓ^p, $p \in]1,\infty[$ sont de type $min(2,p)$.

Un espace de Hilbert est de type et cotype 2.

Un espace de Banach de type et cotype 2 est isommorphe à un espace de Hilbert.

3.8 Optimisation dans un espace de Banach ordonné

Soient Λ un espace de Banach et Γ un espace muni d'une relation d'ordre $\preceq$, on dit que Γ est un espace ordonné. On rappel les définitions et les propriétés suivantes, qu' on peut trouver dans Pallaschke et Rolewicz (1997).

1. Soient F une fonction de Λ dans Γ et $y_0 \in \Lambda$:

   y_0 est solution du problème $\inf\limits_{y\in\Lambda} F(y)$, si $F(y_0) \preceq F(y),\ \forall y \in \Lambda$.

2. Soit K un cône dans Γ, K est convexe si $\forall x, y \in K, \forall \alpha, \beta \geq 0$, alors $\alpha x + \beta y \in K$.

   Le cône K induit un ordre partiel sur Γ défini par $x \leq_K y$ si et seulement si $x - y \in K$.

   K est un cône avec vertex si $\forall x, y \in (K\backslash\{0\}), \forall \alpha, \beta > 0$, alors $\alpha x + \beta y \in K^0$).

   K est un cône pointu s'il ne contient pas une droite.

3. On dit qu'une fonctionnelle linéaire continue t^* définie sur Γ est uniformément positive s'il existe une constante C tel que $\forall y \in K$, $\|y\| \leq C(t^*, y)$.

   Si de plus le cône qui induit un ordre sur Γ est un cône avec vertex, il existe une fonctionnelle linéaire uniformément positive t^* et $c < C$, tel que y_0 est le minimum de la fonction réelle $f(y) = (t^*, F(y)) + c\|F(y) - F(y_0)\|$.

4. Soit A un ensemble de Γ et K un cône pointu. Un point y_0 est le minimum de Pareto, par rapport à l'ordre induit par K, de l'ensemble A si

   $$(y_0 - K) \cap K = \{y_0\} \tag{3.17}$$

   y_0 est dit un point $K-$efficace.

5. Soit A un ensemble de Γ. L'ensemble polaire de A, noté $A^{\preceq}$, est $\bigcap\limits_{x\in A}\{y \ :\ x \preceq y\}$.

   Le conjugué polaire de A, noté $A^{\succeq}$, est $\bigcap\limits_{y\in A}\{x \ :\ x \preceq y\}$.

6. Γ est un espace vectoriel lattice si $\forall A \subset \Gamma$ les ensembles $A^{\preceq}$ et $A^{\succeq}$ sont tous les deux soit l'espace tout entier soit de la forme $A^{\preceq} = p_A^- + K$ (respectivement $A^{\succeq} = p_A^+ + K$), où K est un cône.

Le point p_A^- est dit l'infinimum de A et p_A^+ le supremum de A.

Remarque : En général cet inf et sup n'appartienne pas à $\overline{A}$, ils sont nommés utopia minimum et utopia supremum respectivement, on les notes $\overset{\vee}{\inf}$ et $\overset{\vee}{\sup}$ respectivement.

7. Soient $\mathcal{C}$ une classe de fonctions de Λ dans Γ et $\varepsilon > 0$. Une fonction $\psi \in \mathcal{C}$ est dite $\mathcal{C}_\varepsilon$-gradient local de F au point y_0 s'il existe U_ε un voisinage de y_0 tel que $\forall y \in U_\varepsilon$:

$$\|(F(y) - F(y_0) - (\psi(y) - \psi(y_0))\| \leq \varepsilon \|y - y_0\|$$

Si ψ est $\mathcal{C}_\varepsilon$-gradient local de F au point y_0, $\forall \varepsilon > 0$, ψ est dite $\mathcal{C}_{0+}$-gradient de F au point y_0, il est nommé aussi $\mathcal{C}$-gradient de Fréchet.

L'ensemble des $\mathcal{C}_{0+}$-gradient de F au point y_0 est dit $\mathcal{C}$-differentielle de Fréchet, on le note $d_{\mathcal{C}}F|_{y_o}$.

Soit $\mathcal{C}$ une classe de fonctions positivement homogénes (i.e. $\alpha\mathcal{C} = \mathcal{C}$, $\forall \alpha \geq 0$) de Λ dans Γ. Soit $\psi \in \mathcal{C}$ et $h \in \Lambda$, on pose $\chi(h) = \psi(y_o + h) - \psi(y_0)$, $\chi(h)$ est un $\mathcal{C}$-gradient de Fréchet de $F(y)$ au point y_0 si et seulement si la limite suivante :

$$\lim_{t \downarrow 0} \frac{F(y_0 + th) - F(y_0)}{t} = \chi(h) \tag{3.18}$$

est uniforme sur tout sous-ensemble de la sphère unité de Λ.

χ est une fonction linéaire de Λ dans Γ, c'est à dire un élément de $\mathcal{L}(\Lambda, \Gamma)$, qu'on note par $(\partial F_{|y_0})^*$

Soit K^+ l'ensemble des fonctionnelles positives sur K. D'aprés la proposition (10.2.2) dans Pallaschke et Rolewicz (1997) (p. 514) :

Proposition 3.8.1. *On suppose que F est continuement differentiable. Si y_0 est solution de $\overset{\vee}{\inf}_{y \in \Lambda} F(y)$ alors il existe $\chi \in K^+$ tel que $(\partial F|y_0)^*(\chi) = 0$.*

Dans le cas Γ de dimension finie, c'est le résultat du théorème de F. John (1948) et Kuhn et Tucker (1951).

Bibliographie

Antoniadis, A. et Beder, J. H. (1989). Joint estimation of the mean and the covariance of a Banach valued Gaussian vector. *Statistics*, **20**(1), 77–93.

Antoniadis, A. et Sapatinas, T. (2003). Wavelet methods for continuous-time prediction using Hilbert-valued autoregressive processes. *J. Multivariate Anal.*, **87**(1), 133–158.

Beder, J. H. (1987). A sieve estimator for the mean of a Gaussian process. *Ann. Statist.*, **15**(1), 59–78.

Beder, J. H. (1988). A sieve estimator for the covariance of a Gaussian process. *Ann. Statist.*, **16**(2), 648–660.

Bensmain, N. et Mourid, T. (2001). Estimateur "sieve" de l'opérateur d'un processus ARH(1). *C. R. Acad. Sci. Paris Sér. I Math.*, **332**(11), 1015–1018.

Besse, P. C. et Cardot, H. (1996). Approximation spline de la prévision d'un processus fonctionnel autorégressif d'ordre 1. *Canad. J. Statist.*, **24**(4), 467–487.

Bosq, D. et Mourid, T. (1999). On the equivalence of the measures induced by Banach valued Gaussian autoregressive processes. *Stochastic Anal. Appl.*, **17**(2), 137–144.

Bosq, D. (1991). mean and covariance operator of autoregressive processes in Banach spaces. *Stat. Inference Stoch. Process.*

Bosq, D. (1998). *Nonparametric statistics for stochastic processes*, volume 110 of *Lec-*

ture Notes in Statistics. Springer-Verlag, New York, second edition. Estimation and prediction.

Bosq, D. (2000). *Linear processes in function spaces*, volume 149 of *Lecture Notes in Statistics*. Springer-Verlag, New York. Theory and applications.

Bosq, D. (2002). Estimation of mean and covariance operator of autoregressive processes in Banach spaces. *Stat. Inference Stoch. Process.*, **5**(3), 287–306.

Broniatowski, M. (2003). Estimation of the Kullback-Leibler divergence. *Math. Methods Statist.*, **12**(4), 391–409 (2004).

Cardot, H. (1998). Convergence du lissage spline de la prévision des processus autorégressifs fonctionnels. *C. R. Acad. Sci. Paris Sér. I Math.*, **326**(6), 755–758.

Csiszár, I. (1963). *Eine informationstheoretische Ungleichung und ihre Anwendung auf den Beweis der Ergodizität von Markoffschen Ketten*, volume 8.

Damon, J. et Guillas, S. (2005). Estimation and simulation of autoregressive Hilbertian processes with exogenous variables. *Stat. Inference Stoch. Process.*, **8**(2), 185–204.

Dehling, H. et Sharipov, O. S. (2005). Estimation of mean and covariance operator for Banach space valued autoregressive processes with dependent innovations. *Stat. Inference Stoch. Process.*, **8**(2), 137–149.

Dehling, H. (1983). Limit theorems for sums of weakly dependent Banach space valued random variables. *Z. Washrsch. Verw. Geb.*, **63**, 393–432.

Diestel, J., Jarchow, H., et Tonge, A. (1995). *Absolutely summing operators*, volume 43 of *Cambridge Studies in Advanced Mathematics*. Cambridge University Press, Cambridge.

Diestel, J. (1984). *Sequences and series in Banach spaces*, volume 92 of *Graduate Texts in Mathematics*. Springer-Verlag, New York.

Gál, I. S. et Koksma, J. F. (1948). Sur l'ordre de grandeur des fonctions sommables. *Comptes Rendus*, **227**, 1321–1323.

Gál, I. S. (1949). Sur les moyennes arithmétiques des suites de fonctions orthogonales. *Ann. Inst. Fourier Grenoble*, **1**, 53–59 (1950).

Geman, S. et Hwang, C.-R. (1982). Nonparametric maximum likelihood estimation by the method of sieves. *Ann. Statist.*, **10**(2), 401–414.

Grenander, U. (1981). *Abstract inference.* John Wiley & Sons Inc., New York. Wiley Series in Probability and Mathematical Statistics.

Guillas, S. (2000). Non-causalité et discrétisation fonctionnelle, théorèmes limites pour un processus ARHX(1). *C. R. Acad. Sci. Paris Sér. I Math.*, **331**(1), 91–94.

Guillas, S. (2001). Rates of convergence of autocorrelation estimates for autoregressive Hilbertian processes. *Statist. Probab. Lett.*, **55**(3), 281–291.

Guillas, S. (2002). Doubly stochastic Hilbertian processes. *J. Appl. Probab.*, **39**(3), 566–580.

Jakubowski, A. (1988). Tightness criteria for random measures with application to the principle of conditioning in Hilbert spaces. *Probab. Math. Statist.*, **9**(1), 95–114.

Kadets, V. M. et Kadets, M. I. (1991). *Rearrangements of series in Banach spaces*, volume 86 of *Translations of Mathematical Monographs.* American Mathematical Society, Providence, RI. Translated from the Russian by Harold H. McFaden.

Keziou, A. (2003). Dual representation of ϕ-divergences and applications. *C. R. Math. Acad. Sci. Paris*, **336**(10), 857–862.

Kuelbs, J. (1970). Gaussian measures on a Banach space. *J. Functional Analysis*, **5**, 354–367.

Kuo, H. H. (1975). *Gaussian measures in Banach spaces.* Springer-Verlag, Berlin. Lecture Notes in Mathematics, Vol. 463.

Labbas, A. et Mourid, T. (2002). Estimation et prévision d'un processus autorégressif Banach. *C. R. Math. Acad. Sci. Paris*, **335**(9), 767–772.

Ledoux, M. et Talagrand, M. (1991). *Probability in Banach spaces*, volume 23 of *Ergebnisse der Mathematik und ihrer Grenzgebiete (3) [Results in Mathematics and Related Areas (3)].* Springer-Verlag, Berlin. Isoperimetry and processes.

Lindenstrauss, J. et Tzafriri, L. (1977). *Classical Banach spaces. I.* Springer-Verlag, Berlin. Sequence spaces, Ergebnisse der Mathematik und ihrer Grenzgebiete, Vol. 92.

Lipster, R. S. et Shiryayev, A. N. (1972). *Statistics of conditionally Gaussian random sequences.* Univ. California Press, Berkeley, Calif.

Liptser, R. S. et Shiryaev, A. N. (2001a). *Statistics of random processes. I*, volume 5 of *Applications of Mathematics (New York).* Springer-Verlag, Berlin, expanded edition. General theory, Translated from the 1974 Russian original by A. B. Aries, Stochastic Modelling and Applied Probability.

Liptser, R. S. et Shiryaev, A. N. (2001b). *Statistics of random processes. II*, volume 6 of *Applications of Mathematics (New York).* Springer-Verlag, Berlin, expanded edition. Applications, Translated from the 1974 Russian original by A. B. Aries, Stochastic Modelling and Applied Probability.

Marion, J.-M. et Pumo, B. (2004). Comparaison des modèles ARH(1) et ARHD(1) sur des données physiologiques. *Ann. I.S.U.P.*, **48**(3), 29–38.

Mas, A. (1999). Normalité asymptotique de l'estimateur empirique de l'opérateur d'autocorrélation d'un processus ARH(1). *C. R. Acad. Sci. Paris Sér. I Math.*, **329**(10), 899–902.

Mas, A. (2004). Un nouveau TCL pour les opérateurs de covariance dans le modèle ARH(1). *Ann. I.S.U.P.*, **48**(3), 49–61.

Mourid, T. (1996). Représentation autorégressive dans un espace de Banach de processus réels à temps continu et équivalence des lois. *C. R. Acad. Sci. Paris Sér. I Math.*, **322**(12), 1219–1224.

Pallaschke, D. et Rolewicz, S. (1997). *Foundations of mathematical optimization*, volume 388 of *Mathematics and its Applications*. Kluwer Academic Publishers Group, Dordrecht. Convex analysis without linearity.

Pełczyński, A. (1976). All separable Banach spaces admit for every $\varepsilon > 0$ fundamental total and bounded by $1 + \varepsilon$ biorthogonal sequences. *Studia Math.*, **55**(3), 295–304.

Pinelis, I. F. et Sakhanenko, I. (1985). Remarks on inequalities for large deviations probabilities. *Theor. Probab. Appl.*, **30**(1), 143–148.

Pumo, B. (1995). Les processus autorégressifs à valeurs dans $C_{[0,\delta]}$. Estimation de processus discrétisés. *C. R. Acad. Sci. Paris Sér. I Math.*, **320**(4), 497–500.

Pumo, B. (1999). Prediction of continuous time processes by C[0,1]-valued autoregressive processes. *Stat. Inference Stoch. Process.*, **3**, 1–13.

Rachedi, F. et Mourid, T. (2003). Estimateur crible de l'opérateur d'un processus ARB(1). *C. R. Math. Acad. Sci. Paris*, **336**(7), 605–610.

Rachedi, F. (2004). Vitesse de convergence de l'estimateur crible d'un ARB(1). *Ann. I.S.U.P.*, **48**(3), 87–96.

Rachedi, F. (2005). Vitesse de convergence en norme p-intégrale et normalité asymptotique de l'estimateur crible de l'opérateur d'un ARB(1). *C. R. Math. Acad. Sci. Paris*, **341**, 369–374.

Rozanov, J. A. (1971). Infinite-dimensional Gaussian distributions. pages iii+161. Translated from the Russian by G. Biriuk.

Rüschendorf, L. (1984). *On the minimum discrimination information theorem.* Number suppl. 1. Recent results in estimation theory and related topics.

Rüschendorf, L. (1987). *Projections of probability measures*, volume 18.

Vakhania, N. N., Tarieladze, V. I., et Chobanyan, S. A. (1987). *Probability distributions on Banach spaces*, volume 14 of *Mathematics and its Applications (Soviet Series)*. D. Reidel Publishing Co., Dordrecht. Translated from the Russian and with a preface by Wojbor A. Woyczynski.