

THESE DE DOCTORAT

présentée à

L'UNIVERSITE ABOU BEKR BELKAID TLEMCCEN

pour obtenir le grade de

DOCTEUR EN MATHEMATIQUES

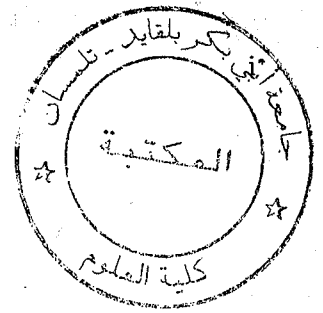
Par

Nawel BENSMAIN

Sujet de la thèse

Prévision par la méthode des sieves d'un processus autorégressif Hilbertien

Soutenue le	2008	devant le jury composé de	
Mr GHOUALI N	Professeur	Président	Université ABB -Tlemcen
Mr MAS A	M.C	Examineur	Université Montpellier 2
Mr CARDOT H	Professeur	Examineur	Université de Bourgogne Dijon
Mr DIB H	Professeur	Examineur	Université ABB-Tlemcen
Mr LAKSACI A	M.C	Examineur	Université Sidi Bel Abbes
Mr MOURID T	Professeur	Directeur de thèse	Université ABB-Tlemcen



Date: 15 MARS 2009
Page: 3537

Remerciements:

J'exprime ma profonde gratitude et ma sincère reconnaissance à Mr MOURID Tahar directeur de cette thèse pour ses conseils judicieux et le suivi qu'il m'a prodigués tout au long de la préparation de cette thèse.

Je tiens à remercier Mr MAS André, Mr CARDOT Hervé pour avoir apporté rigueur et clarté à cette étude en acceptant d'examiner ce travail.

J'exprime mes remerciements à Mr GHOUALI Noureddine qui m'a fait l'honneur de présider ce jury malgré ses préoccupations

Je remercie aussi Mr DIB Hacène et Mr LAKSACI Ali pour l'intérêt qu'ils ont apporté à ce travail en acceptant de l'examiner.

Je tiens à exprimer mes plus sincères remerciements et reconnaissance à mes parents et à mon mari Mr DESSADAT Mustapha qui n'ont cessé de m'aider et de m'encourager pour accomplir ce travail.

Je tiens à remercier enfin tout ceux qui ont contribué à l'élaboration de ce travail.

Résumé: Dans cette thèse nous considérons la prédiction d'un processus à temps continu admettant une représentation autorégressive AR(1) dans l'espace des fonctions continue sur $[0, 1]$ par la méthode des sieves.

La prédiction d'un processus aléatoire à temps continu sur un intervalle par une modélisation autorégressive fonctionnelle a suscité beaucoup de travaux théoriques et appliqué. Etant donné un processus réel $Y = (Y(t), t \in \mathbb{R})$ observé sur $[0, T]$, il s'agit de prévoir le comportement de Y sur tout l'intervalle $[T, T + \delta]$, $\delta > 0$. En posant $X_n(t) := Y(t + n\delta)$, $t \in [0, \delta]$ nous générons une suite de v.a. $X = (X_n, n \in \mathbb{Z})$ à temps discret à valeurs dans un espace fonctionnel approprié et régie par une équation du type autorégressif.

Ainsi la prédiction du processus Y sur $[T, T + \delta]$ est ramenée à la prévision de la v.a. X_n comme élément fonctionnel et Bosq et Mourid ont introduit les processus autorégressifs à valeurs dans un espace de Hilbert pour résoudre ce type de prévision. Nous nous intéressons aux processus autorégressifs hilbertiens d'ordre 1 $ARH(1)$. et à l'estimation de cette classe de processus par la méthode sieves.

Le paramètre d'intérêt est l'opérateur d'autorégression d'un processus ARH (1) et son estimation par la méthode des sieves. Nous donnons les résultats de convergence presque sûre de l'estimateur sieves pour le cas d'un opérateur à noyau de convolution et une forme explicite de l'estimateur sieves dans le cas Gaussien. Dans le cas d'un opérateur de Hilbert Schmidt, nous donnons des conditions générales pour l'existence et la convergence presque sûre de l'estimateur sieves. Nous donnons une vitesse de la dimension m_n de la sieve pour l'opérateur à noyau $m_n = O(n^{\frac{1}{3}-\eta})$, $\eta > 0$ au lieu d'une forme logarithmique $m_n = o(\log(n))$ obtenue auparavant par d'autres auteurs. Nous donnons enfin un résultat de convergence du prédicteur sieves pour la norme de $L^2[0, 1]$. Nous présentons des études de simulations d'un processus AR en utilisant le logiciel R version 2.4.1 et Bibliothèque far de Damon en considérant un opérateur ρ à noyau de convolution K avec simulations des trajectoires du processus et leurs prédicteurs sieves. Les résultats obtenus corroborent les résultats théoriques. Pour des applications nous avons considéré la prévision du phénomène météorologique "El Nino", durant l'année 2006 en utilisant des observations mensuelles de la période 1950-2005. Pour pouvoir comparer avec les résultats de sa prévision durant l'année 1986 qui existent dans la littérature nous avons aussi présenté sa prévision sieves de l'année 1986. De même, la prévision de la température en 1939 du Château de Nottingham sur la base de données mensuelles de 1920 à 1938 et la prévision de la température moyenne en 2004 à Alger basée sur la période 1970-2003 sont présentées. Les résultats sur les erreurs de

prévision sont très faibles comparativement aux erreurs obtenues par d'autres méthodes de prévision.

Title: Sieves prediction of A Functional Autoregressive Processes

Abstract: In this thesis we consider the prediction of continuous time stochastic process admitting a functional autoregressive representation AR(1) in the space of continuous functions $C_{[0,\delta]}$ based on sieves approach. The prediction of a continuous time random process on an interval arises in many applications. When a sample path of a real process $Y = (Y(t), t \in \mathbb{R})$ on the interval $[0, T]$ is available, we predict the behaviour of Y on the entire interval $[T, T + \delta], \delta > 0$. Considering $X_n(t) := Y(t + n\delta), t \in [0, \delta]$, we generate a discrete time function space valued random process $X = (X_n, n \in \mathbb{Z})$ whose time evolution is assumed ruled by an autoregressive equation.

Hence a predictor of the function space valued r.v. X_{n+1} provides a predictor of the process Y on the entire interval $[T, T + \delta]$. Bosq and Mourid have introduced a class of functional autoregressive random processes with values in a real separable Hilbert space (ARH) to solve this kind of prevision. we are mainly concerned with the sieve estimation of the parameter operator of a first order Hilbertian autoregressive process. The parameter of interest is the operator of an hilbertian autoregressive process ARH(1) and its estimation by the method of sieves. We give the results of the almost sure convergence of sieves estimate when the operator is of a kernel type and an explicit form of the sieves in Gaussian case afterwards we give general conditions of existence and almost sure convergence of sieves estimate when the parameter is a Hilbert Schmidt operator. In the kernel operator type the almost sure convergence is obtained under polynomial growth dimension $m_n = O(n^{\frac{1}{3}-\eta})$, $\eta > 0$ improving the logarithmic growth $m_n = o(\log(n))$ obtained earlier by others authors. At last, we give a result of convergence of sieves predictor in the norm of $L^2[0, 1]$. We present studies simulations of an AR process using R version 2.4.1 and The far Package of Damon by considering an integral kernel operator ρ . Simulations of the trajectories of the process and their sieves predictors are given. The obtained results corroborate the theoretical ones. For applications, we consider the prevision of "El Nino" phenomenon during the year 2006 using monthly observations of the period 1950-2005. To compare the results of prevision of the year 1986 obtained in the literature, we use its sieves prediction of 1986.

Also, the prevision of the historical data of temperature in Nottingham Palace during 1939 on a monthly data basis from 1920 to 1938 and the prevision of the mean temperature in Algiers during the year 2004 based on the period of 1970-2003 are presented. The values of quadratic errors and relative errors are very weak compared to those obtained by others methods of prediction.

Mots clés : Processus autorégressif fonctionnel- Estimation Sieves-Prédicteur sieves-Simulations-Série "El Nino"

Classifications AMS: 62A10;62M10;46E15

Key words :Functional autoregressive process- -Sieves Estimation-Sieve's predictor-Simulations-"El Nino" series.

Table des Matières

Introduction	1
1 PROCESSUS AUTOREGRESSIFS BANACHIQUES. EXEMPLES D'ESTIMATION PAR LES SIEVES	5
1.1 Processus autorégressif Banachique ARB	5
1.1.1 Représentation ARB(1) de processus réels à temps continu	7
1.1.2 Processus d'Ornstein-Uhlenbeck	7
1.1.3 Construction d'un processus ARC(1) ($B = C_{[0,1]}$)	8
1.1.4 Classe de processus ARB(1) dans $L^2_{[0,1]}$	10
1.1.5 Processus avec saisonnalité	11
1.2 Loi Forte des Grands Nombres (LFGN) pour les processus ARB	11
1.3 Processus Autorégressifs Hilbertiens d'ordre un	12
1.3.1 Définition d'un ARH(1)	12
1.4 Exemples d'estimation par les "sieves"	12
1.4.1 Estimation par les "sieves" pour des observations dépendantes	12
1.4.2 Paramètres dans l'espace l^2 et L^2	17
1.4.3 Paramètres dans l'espace $C_{[0,1]}$	21
1.4.4 Opérateurs comme paramètres	25
1.4.5 Régression non paramétrique	27
2 ESTIMATEUR SIEVES D'UN PROCESSUS AUTOREGRESSIF HILBERTIEN	40
2.1 Introduction	40
2.2 Définitions et notations	41
2.3 Estimation Sieves de l'opérateur d'un ARH(1)	42
2.3.1 Le paramètre ρ est un opérateur à noyau	42
2.3.2 Le paramètre ρ est un opérateur de Hilbert-Schmidt	46
2.3.3 Prédiction	57
3 SIMULATIONS	59
3.1 Simulation de processus ARH(1)	59
3.1.1 Introduction	59
3.1.2 Définitions	60

3.1.3	Calcul des Prédicteur sieves	61
3.1.4	Exemples	76

Bibliographie		87
----------------------	--	-----------

Introduction

L'interprétation d'un processus stochastique à temps continu comme une variable aléatoire à valeurs dans un espace fonctionnel approprié est considéré naturellement dans plusieurs domaines des probabilités et statistiques. Dans ce travail nous étudions la prédiction d'un processus aléatoire à temps continu sur un intervalle par une modélisation autorégressive fonctionnelle. Comme modèle classique les processus autorégressifs hilbertiens et banachiques ont suscité beaucoup de travaux théoriques et appliqués dernièrement, et s'est avérée d'un grand intérêt ([7], [14], [20]). Plus précisément soit $Y = (Y(t), t \in \mathbb{R})$ un processus stochastique réel observé sur $[0, T]$, il s'agit de la prédiction du processus Y sur tout l'intervalle $[T, T + \delta]$, $\delta > 0$. En posant $X_n(t) := Y(t + n\delta)$, $t \in [0, \delta]$ nous générons une suite de v.a. $X = (X_n, n \in \mathbb{Z})$ à temps discret à valeurs dans un espace fonctionnel approprié et dont l'évolution temporelle est supposée régie par une équation du type autorégressif. Si le processus Y est à trajectoires continues l'espace fonctionnel naturel associé sera l'espace $C_{[0, \delta]}$ des fonctions continues sur $[0, \delta]$. Ainsi la prédiction du processus Y sur $[T, T + \delta]$ est ramenée à la prévision de la v.a. X_n comme élément de $C_{[0, \delta]}$. Pour des applications voir ([7] ch. 9 et les références citées): nous pouvons citer : la prédiction du phénomène météorologique "El Nino", la prévision du trafic routier, la consommation de l'énergie électrique d'une région, l'électrocardiogramme, la prévision du nombre annuel de passagers dans les chemins de Fer Français et la prédiction des niveaux de pollution atmosphérique.

Bosq et Mourid [8] ont introduit les processus autorégressifs à valeurs dans un espace de Hilbert pour résoudre ce problème de prévision. Nous présentons la définition de ces processus à valeurs dans un espace de Banach séparable noté ARB et à valeurs dans un espace de Hilbert noté ARH . Dans ce travail nous nous intéressons particulièrement aux processus autorégressifs hilbertiens d'ordre 1 $ARH(1)$.

Soit (B, β_B) un espace de Banach séparable muni de la tribu borélienne et de la norme $\|\cdot\|$. Considérons une suite $(\varepsilon_n, n \in Z)$ de v.a à valeurs dans B , définies sur l'espace probabilisé (Ω, A, P) indépendantes et de même loi telles que $0 < \sigma^2 = E \|\varepsilon_n\|^2 < +\infty$ et $E\varepsilon_n = 0$ où l'espérance est prise au sens de Bochner. Nous dirons qu'une telle suite est un B - bruit blanc. Soit ρ un opérateur linéaire borné de B dans B et on note $\|\rho\|_L$ la norme d'opérateurs linéaires bornés et $m \in B$.

On dit qu'une suite de v.a. $(X_n, n \in Z)$ définie sur (Ω, A, P) et à valeurs dans (B, β_B) est un processus autorégressif d'ordre 1 noté $ARB(1)$, si

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

Les principales propriétés de cette classe sont établies par Bosq [7] où l'on trouve aussi des exemples de processus à temps continu admettant une représentation autorégressive $ARB(1)$.

Le premier chapitre contient des préliminaires et des définitions des processus autorégressifs hilbertiens et banachiques ARH et ARB . Pour illustration, nous citons des exemples de représentations ARB de processus réel à temps continu [7]. En suivant Grenander [13], Shen X. Wong W.H [22] et White et Wooldrige [23] nous présentons des exemples d'estimations de paramètres de dimension infini par la méthode des sieves. En définissant le cadre général de la méthodologie de l'estimation par la méthode des sieves nous donnons en particulier l'estimation de paramètres de type moyenne de loi de probabilités gaussienne en dimension infinie, opérateurs de covariance et de fonction de régression .

Le second chapitre est consacré à l'estimation du paramètre l'opérateur d'autorégression d'un processus autorégressif hilbertien d'ordre 1 par la méthode des sieves. Dans le cas d'un opérateur à noyau de convolution, nous obtenons des résultats sur l'existence de l'estimateur sieves et leur convergence presque sûre et avec une forme explicite de l'estimateur sieves dans le cas Gaussien . Dans le cas où le paramètre est un opérateur de Hilbert Schmidt, sous des conditions générales similaires à [12] nous obtenons l'existence et la convergence presque sûre de l'estimateur sieves en norme de Hilbert-Schmidt. Nous donnons une vitesse de la dimension m_n de la sieve pour l'opérateur à noyau $m_n = O(n^{\frac{1}{3}-\eta})$, $\eta > 0$ au lieu d'une forme logarithmique $m_n = o(\log(n))$ obtenue auparavant par d'autres auteurs Nous donnons enfin un résultat de convergence du prédicteur sieves pour la norme de $L^2[0, 1]$ [18]

Dans le troisième chapitre nous présentons des études de simulations d'un processus AR en utilisant le logiciel R version 2.4.1. [Bibliothèque far] [10]. Nous nous inspirons

de l'exemple de simulation d'un processus $ARC(1)$ donnée par Damons [10] qui prend en considération le travail de Pumo [19]. Leur travail permet de simuler un bruit blanc dans l'espace $C[0, 1]$ en utilisant les fonctions propres

$$(e_j(t) = \sqrt{2} \sin [(j - 1/2) \pi t], , j = 1, 2, \dots)$$

de l'opérateur de covariance du mouvement brownien. Pour simuler les trajectoires de notre processus AR et les prédicteurs sieves considérés nous devons spécifier le modèle. Plus précisément, nous considérons un opérateur ρ à noyau de convolution K paire périodique de période 1 sur R de norme L^2 inférieure à 1. Les exemples de noyaux K considérés sont :

$$K(t) = \frac{1}{10^2} \cos(t), \quad K(t) = \frac{1}{10^3} t^2, \quad K(t) = \frac{1}{10^2} t^2 \cos(t), \quad t \in [0, 1]$$

Nous simulons $n + 1$ observations d'un processus $ARH(1)$ qui seront nécessaires pour calculer les prédicteurs sieves et qui seront comparés à la $n + 1$ eme trajectoire du processus. Les résultats obtenus sont très satisfaisants et corroborent les résultats théoriques. Les erreurs quadratiques et les erreurs relatives RMAE de prédiction sont très faibles et de l'ordre de quelques pourcents surtout dans le cas des séries réelles . Pour des applications réelles nous avons considéré la prévision du phénomène météorologique "El Nino". Pour sa prévision de l'année 2006 on utilise les observations mensuelles de la période 1950-2005. Pour pouvoir comparer avec les résultats de sa prévision durant l'année 1986 qui existent dans la littérature [2] nous avons aussi présenté sa prévision sieves de l'année 1986. De même, la prévision de la température en 1939 du Château de Nottingham sur la base de données mensuelles de 1920 à 1938 et la prévision de la température moyenne en 2004 à Alger basé sur la période 1970-2003 sont présentés. En général les résultats sur les erreurs quadratiques et les erreurs relatives RMAE de prévision sont très faibles comparativement aux erreurs quadratiques et relatives RMAE obtenues par d'autres méthodes de prévision.

Publications

1. Bensmain, N., Mourid, T. Estimateur sieves de l'opérateur d'un processus ARH (1) . Comptes Rendus Acad. Sci. Paris t. 332 Serie 1. 2001
2. Mourid, T., Bensmain, N. Sieves estimator of the operator of a functional autoregressive process. Statistics and Probability Letters 76,93-108. 2006

3. Bensmain,N., Mourid,T. Prédiction des processus AR Hilbertien via la méthode des sieves. Simulations et exemples . soumis 2008

Chapitre 1

PROCESSUS AUTOREGRESSIFS BANACHIQUES. EXEMPLES D'ESTIMATION PAR LES SIEVES

La classe des processus autorégressive Hilbertiens et Banachique d'ordre un, noté $ARB(1)$, est une généralisation naturelle en dimension infinie des processus autorégressif à valeurs dans \mathbb{R}^k . Dans ce chapitre nous donnons les définitions et les résultats les plus importants concernant cette classe de processus contenus dans [7].

Nous indiquons ensuite deux exemples de processus réels à temps continu qui admettent une représentation autorégressive $ARB(1)$. Une loi forte des grands nombres est présentée ensuite. En fin nous présentons la méthode d'estimation sieves en dimension infinie de Grenander avec des exemples de la littérature.

1.1 Processus autorégressif Banachique ARB

Soit (B, β_B) un espace de Banach séparable muni de la tribu borélienne et de sa norme $\|\cdot\|$. Considérons une suite $(\varepsilon_n, n \in \mathbb{Z})$ de v.a. définies sur un espace probabilisé (Ω, A, P) et à valeurs dans B , indépendantes et de même loi telles que $0 < \sigma^2 = E \|\varepsilon_n\|^2 < +\infty$ et $E\varepsilon_n = 0$ où l'espérance est prise au sens de Bochner. Nous dirons que la suite

$(\varepsilon_n, n \in Z)$ est un B -bruit blanc.

Soit ρ un opérateur linéaire borné de B dans B et on note $\|\rho\|_L$ la norme d'opérateurs linéaires bornés et soit $m \in B$.

Une suite de v.a. $(X_n, n \in Z)$ définie sur (Ω, A, P) et à valeurs dans (B, β_B) est un processus autorégressif d'ordre 1 noté ARB(1), si elle vérifie

$$X_n - m = \rho(X_{n-1} - m) + \varepsilon_n, \quad n \in Z \quad (1.1)$$

Le résultat suivant donne une condition pour l'existence de tel processus

Lemme 1.1.1 [7]: Si $\sum_{j \geq 1} \|\rho^j\| < +\infty$ alors pour tout $i \in Z$ la série

$$\sum_{j \geq 1} \rho^j \varepsilon_{i-j} \quad (1.2)$$

converge presque sûrement et dans L_B^2 .

Preuve. Nous avons

$$\begin{aligned} \Delta_{n,m} &: = E \left\| \sum_{j=n}^m \rho^j \varepsilon_{i-j} \right\|^2 \\ &\leq E \left(\sum_{j=n}^m \|\rho^j \varepsilon_{i-j}\| \right)^2 \\ &\leq E \left(\sum_{j=n}^m \|\rho^j\| \|\varepsilon_{i-j}\| \right)^2 \\ &\leq E \left(\sum_{j=n}^m \sum_{l=n}^m \|\rho^j\| \|\varepsilon_{i-j}\| \|\rho^l\| \|\varepsilon_{i-l}\| \right) \\ &\leq \sum_{j=n}^m \sum_{l=n}^m \|\rho^j\| \|\rho^l\| E (\|\varepsilon_{i-j}\| \|\varepsilon_{i-l}\|) \\ &\leq \sigma^2 \left(\sum_{j=n}^m \|\rho^j\| \right)^2 \end{aligned}$$

Sous la condition du lemme le dernier majorant tend vers 0 si $n, m \rightarrow +\infty$ et donc la suite est de Cauchy dans L_B^2 . La convergence presque sûre découle d'un résultat de Geoffroy (1959)(cf [7]).

Pour $m \in B$ et (ε_i) est un B -bruit blanc on pose

$$X_i = m + \sum_{j \geq 0} \rho^j \varepsilon_{i-j}, \quad i \in Z$$

Par le lemme précédent la série converge presque sûrement et dans L_B^2 .

Alors le processus $(X_n, n \in Z)$ est strictement stationnaire et on a

$$X_i - m = \rho(X_{i-1} - m) + \varepsilon_i, \quad i \in Z \quad (1.3)$$

et ε_i est indépendant de $(X_j, j < i)$,

Par suite $(X_n, n \in Z)$ est un processus autorégressif d'ordre 1, ARB(1) à valeurs dans B .

1.1.1 Représentation ARB(1) de processus réels à temps continu

Nous indiquons une classe de processus réels à temps continu admettant une représentation ARB(1). Cette classe contient notamment le processus d'Ornstein-Uhlenbeck.

Les résultats des deux sections qui suivent sont dus à Bosq [7] Chap. 2

1.1.2 Processus d'Ornstein-Uhlenbeck

On considère le processus réel d'Ornstein-Uhlenbeck

$$\xi_t = \int_{-\infty}^t \exp(-c(t-u)) dW_u, \quad t \in R$$

où (W_u) est un processus de Wiener et c une constante strictement positive.

On choisit $B = C_{[0,1]}$ et pose

$$X_n(t) = \xi_{n+t} \quad 0 \leq t \leq 1, \quad n \in Z$$

La version choisie de ξ_t étant supposée à trajectoires continues, on définit ainsi des v.a X_n à valeurs dans $C_{[0,1]}$. et $X_n(t), t \in [0, 1]$ représente le morceau de trajectoire du processus ξ_t sur l'intervalle $[n, n+1]$.

D'autre part,

$$\begin{aligned}
E(\xi_{n+t}/\xi_s, s \leq n) &= E\left(\int_{\infty}^{n+t} e^{-c(t-u)} dW(u) / \xi_s, s \leq n\right) \\
&= E\left(\int_{\infty}^n e^{-c(t-u)} dW(u) + \int_n^{n+t} e^{-c(t-u)} dW(u) / \xi_s, s \leq n\right) \\
&= E\left(\int_{\infty}^n e^{-c(t-u)} dW(u) / \xi_s, s \leq n\right) \\
&= \int_{\infty}^n e^{-c(t-u)} dW(u) \\
&= e^{-ct} \xi_n \quad 0 \leq t \leq 1
\end{aligned}$$

Ce qui amène à poser:

$$\rho f(t) = e^{-ct} f(1)$$

et

$$\varepsilon_n(t) = \int_n^{n+t} e^{-c(n+t-u)} dW(u) = \int_0^t e^{-c(t+v)} dW(n+v)$$

Alors

$$\begin{aligned}
\varepsilon_n(t) &= \int_{\infty}^{n+t} e^{-c(n+t-u)} dW(u) - e^{-ct} \int_{\infty}^n e^{-c(n-u)} dW(u) \\
&= \xi_{n+t} - (\rho X_{n-1})(t) \quad 0 \leq t \leq 1, n \in \mathbb{Z}
\end{aligned}$$

Enfin comme le processus de Wiener (W_u) est à accroissements indépendants, alors la suite (ε_n) est un bruit blanc.

$$\text{D'autre part } \|\rho^n\| = \sup_{\|f\|=1} \|\rho^n f\| = e^{-c(n-1)}.$$

Donc

$$\sum_{n \geq 0} \|\rho^n\| < \infty$$

et par suite par le lemme nous déduisons que le processus (X_n) est un ARB(1).

1.1.3 Construction d'un processus ARC(1) ($B = C_{[0,1]}$).

Rappelons la construction du processus de Wiener par le théorème de Karhunen-Loève. Soit $(\xi_t, t \in T)$ un processus stochastique du 2^{eme} ordre telle que $E(|\xi_t|^2) < \infty$ pour tout $t \in T$ et $(\zeta_r, a \leq r \leq b)$ un processus stochastique de 2^{eme} ordre de moyenne nulle et de fonction de covariance K continu tel que a et b sont finis. Soit $(\varphi_j, j = 1, 2, \dots)$

une base orthonormale de l'espace engendré par les fonctions propres correspondantes aux valeurs propres non nulles de l'opérateur intégral associé à K .

$$(Af)(s) = \int_a^b K(s,t)f(t)dt$$

Dans ces conditions le théorème de Karhunen-Loève donne:

$$\zeta(t) = \sum_{j=1}^{+\infty} \xi_j \varphi_j(t), \quad t \in [a,b]$$

où $\xi_j = \int_a^b \zeta(t) \varphi_j(t) dt$, sont des variables aléatoires orthogonales de moyenne nulle et $E[|\xi_j|^2] = \lambda_j$. Cette série converge dans L^2 , et uniformément sur $[a,b]$.

Si le processus est Gaussien alors le vecteur aléatoire (ξ_1, \dots, ξ_r) est gaussien dans \mathbb{R}^r . Dans le cas d'un processus de Wiener ($W(t)$, $t \geq 0$) on a:

$$K(s,t) = \min(s,t)$$

Les vecteurs et les valeurs propres de l'opérateur de covariance A sont:

$$\begin{aligned} \varphi_i(t) &= \sqrt{2} \sin \left[\left(i - \frac{1}{2} \right) \pi t \right] \\ \mu_i &= \frac{1}{\left(i - \frac{1}{2} \right)^2 \pi^2} \end{aligned}$$

En prenant $\xi_j^* = \frac{\xi_j}{\sqrt{\mu_j}}$ où ξ_j sont donnés par le développement en série de Karhunen-Loève, et on a :

$$W(t) = \sqrt{2} \sum_{j=1}^{\infty} \xi_j^* \frac{\sin \left(j - \frac{1}{2} \right) \pi t}{\left(j - \frac{1}{2} \right) \pi}$$

qui est le développement de Karhunen-Loève du processus de Wiener. Pour un processus de Wiener sur un intervalle $[0, B]$, nous trouvons,

$$\varphi_i(t) = \sqrt{\frac{2}{B}} \sin \left[\left(i - \frac{1}{2} \right) \pi \frac{t}{B} \right], \quad v_i = \frac{1}{\left(i - \frac{1}{2} \right)^2 \pi^2}$$

et donc

$$W(t) = \sqrt{\frac{2}{B}} \sum_{j=1}^{\infty} \xi_j^* \frac{\sin \left[\left(j - \frac{1}{2} \right) \pi \frac{t}{B} \right]}{\left(j - \frac{1}{2} \right) \frac{\pi}{B}}$$

Si $B = n + 1$ on aura une version du processus de Wiener sur $[0, n + 1]$ donnée par:

$$W_u = \sqrt{\frac{2}{n+1}} \sum_{j=1}^{+\infty} Y_j \frac{\sin \left[\left(j - \frac{1}{2} \right) \pi \frac{u}{n+1} \right]}{\left(j - \frac{1}{2} \right) \frac{\pi}{n+1}}, \quad u \in [0, n + 1]$$

où les (Y_j) sont des variables aléatoires i.i.d normales réduites.

Posons pour tout $w \in \Omega$ et $i \in N$:

$$\begin{aligned} \varepsilon_i & : [0, 1] \longrightarrow \mathbb{R} \\ s & \longmapsto \varepsilon_i(s) := W_{i+s} - W_i \end{aligned}$$

alors (ε_n) est un bruit blanc.

Nous définissons maintenant l'opérateur ρ . Puisque nous resterons dans le sous espace engendré par les vecteurs φ_i , il suffit de définir $\rho(\varphi_i)$. Nous prenons

$$\rho[\varphi_i](s) = \beta_i \varphi_i(s)$$

où (β_i) est une suite de réels

Et posons

$$X_i = \sum_{j=0}^{\infty} \rho^j \varepsilon_{i-j}.$$

nous obtenons un processus (X_n) qui est un *ARC* (1).

1.1.4 Classe de processus *ARB*(1) dans $L^2_{[0,1]}$

Soit $(Z_t, t \in R)$ un processus réel du second ordre, centré et à accroissements indépendants et strictement stationnaires. On suppose que nous avons une version de (Z_t) à trajectoires localement de carrés intégrables. Alors en posant:

$$\varepsilon_n(t) = Z_{n+t} - Z_n, \quad 0 \leq t \leq 1, \quad n \in Z$$

on définit un bruit blanc dans $L^2_{[0,1]} = H$.

Soit ρ un opérateur linéaire sur H , intégral de noyau K défini par

$$(\rho f)(t) = \int_0^1 K(s, t) f(s) ds, \quad 0 \leq t \leq 1, \quad f \in L^2_{[0,1]}$$

où

$$0 \leq \int_0^1 \int_0^1 K^2(s, t) dt ds < 1$$

Alors $\|\rho\| < 1$ et on définit un *ARH* (1) en posant:

$$X_n(t) = \sum_{j \geq 0} \rho^j (\varepsilon_{n-j})(t), \quad 0 \leq t \leq 1$$

(Z_t) peut être un processus de Wiener ou un processus de Poisson centré.

1.1.5 Processus avec saisonnalité

Considérons un processus réel de la forme

$$\eta_t = m(t) + \xi_t, \quad t \in \mathbb{R}$$

où (ξ_t) est un processus centré, à trajectoires continues et admettant une représentation $ARB(1)$ où $B = C_{[0,1]}$.

On suppose que m est une fonction continue, non aléatoire, de période h et non constante. Dans ces conditions η_t admet une représentation $ARB(1)$ avec

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

et $EX_n = m$, X_n est donc stationnaire alors que (η_t) ne l'est pas.

Conclusion . Ces exemples montrent l'intérêt de cette classe de processus où des processus réels usuels admettant une représentation autorégressive dans des espaces fonctionnels bien choisis cf [7].

1.2 Loi Forte des Grands Nombres (LFGN) pour les processus ARB

Posons

$$S_n = \sum_{i=1}^n X_i, \quad \bar{X}_n = \frac{S_n}{n}, \quad \bar{\varepsilon}_n = \frac{\sum_{i=1}^n \varepsilon_i}{n} \quad \text{et} \quad R = \sum_{j \geq 0} \|\rho^j\| < \infty,$$

Le résultat suivant établit la loi forte de grands nombres pour $(X_n) \sim ARB(1)$

Théorème 1.2.1 [7]. (LFGN) Soit (X_n) un processus ARB de moyenne m , alors

$$\bar{X}_n \rightarrow m \text{ p.s} \tag{1.4}$$

Pour la convergence en moyenne d'ordre r dans un espace de Banach on a le théorème suivant

Théorème 1.2.2 [17] . *Sous certaines conditions (cf [17])*

Alors pour tout $r \geq 1$:

$$\lim_{n \rightarrow +\infty} E \|\bar{X}_n - m\|^r = 0$$

1.3 Processus Autorégressifs Hilbertiens d'ordre un

1.3.1 Définition d'un ARH(1)

Nous donnerons ici la définition d'un *ARH(1)* que l'on utilisera dans cette section avec deux hypothèses supplémentaires par rapport à la définition d'un *ARB(1)*.

Considérons un espace de Hilbert réel et séparable muni de la norme $\|\cdot\|$. Un processus (X_n) est un *ARH(1)* centré à valeurs dans H si

$$X_i = \rho X_{i-1} + \varepsilon_i \quad i = 0, \pm 1, \pm 2, \dots$$

où (ε_i) est un H - bruit blanc et ρ est un opérateur compact et symétrique sur H tel que $\exists j_0 \geq 1 / \|\rho^{j_0}\|_L < 1$. (cf [7] chap. 2)

1.4 Exemples d'estimation par les "sieves"

1.4.1 Estimation par les "sieves" pour des observations dépendantes

Quand le paramètre à estimer prend ses valeurs dans un espace de paramètre Θ de dimension infinie, les méthodes traditionnelles d'inférence statistique ne s'appliquent pas

directement sans des modifications substantielles. Grenander [13] a introduit la méthode des "sieves" pour remédier à l'estimation dans ce type de problèmes.

Une "sieve" notée par $\Theta(\mu)$ est une famille de sous ensembles paramétrique de Θ indexés par μ appelée dimension de la "sieve". L'approche de Grenander est basée sur la maximisation de la vraisemblance sur les sous ensembles $\Theta(\mu)$ qui sont de dimension finie de telle sorte que le maximum de vraisemblance local existe.

D'autre part, quand la dimension de $\Theta(\mu)$ croît avec la taille de l'échantillon le but étant d'obtenir la convergence de l'estimateur du maximum de vraisemblance local sur $\Theta(\mu)$ vers la vraie valeur du paramètre.

Plus précisément on a

Définition 1.4.1 Soit (Θ, d) un espace paramétrique muni d'une distance d . Soit $(\Theta(\mu))_\mu$ une famille de sous ensembles de Θ . La famille $(\Theta(\mu))_\mu$ est appelée "sieve" si

- 1) $\Theta(\mu)$ est compact.
- 2) $(\Theta(\mu))_\mu$ est une suite croissante
- 3) $\bigcup_{\mu} \Theta(\mu)$ est dense dans Θ .

Les résultats ci dessous sont tirés de [23] et illustrent la méthode des sieves dans des applications . En particulier cet exemple montre comment étendre les résultats obtenus par la méthode des "sieves" dans le cas "i i d" au cas non "i i d" . On note par $B(\Theta)$ la tribu borélienne engendrée par les ouverts de l'ensemble Θ et la tribu borelienne de \mathbb{R} par $\bar{B} = B(\bar{\mathbb{R}})$.

Théorème 1.4.1 [23] Soit (Ω, F, P) un espace de probabilité complet et soit (Θ, d) un espace métrique complet séparable et (Θ_n) une sieve de Θ .

Soit $Q_n : \Omega \times \Theta_n \rightarrow \bar{\mathbb{R}}$ une application $F \otimes B(\Theta_n)/B(\bar{\mathbb{R}})$ -mesurable. On suppose que pour tout $\omega \in \Omega$, $Q_n(\omega, \cdot)$ est semi continue inférieurement sur Θ_n $n=1,2,\dots$ Alors: il existe

$$\hat{\theta}_n : \Omega \rightarrow \Theta_n, \quad F/B(\Theta_n) - \text{mesurable}$$

telle que:

$$Q_n(\omega, \hat{\theta}_n(\omega)) = \inf_{\theta \in \Theta_n} Q_n(\omega, \theta). \text{ Pour tout } n$$

Ce résultat implique :

$$P(\omega, \sup_{\theta \in \Theta_n} |Q_n(\omega, \theta) - \bar{Q}_n(\theta)| > \varepsilon) \rightarrow 0 \text{ quand } n \rightarrow +\infty.$$

Théorème 1.4.2 [23] Soit (Ω, F, P) un espace de probabilité complet. (Θ, d) un espace métrique. $\{\Theta_n\}$ une sieve de Θ . Soit

$$\{s_{nt} : \Omega \times \Theta_n \rightarrow \bar{\mathbb{R}}, n, t = 1, 2, \dots\} \text{ et } \{m_{nt} : \Omega \times \Theta_n \rightarrow \bar{\mathbb{R}}, n, t = 1, 2, \dots\}$$

deux vecteurs de fonctions. Pour tout $\theta \in \Theta_n$ s_{nt} et m_{nt} sont F/\bar{B} mesurables. On suppose qu'il existe une suite $\{d_n : \Theta_n \rightarrow \mathbb{R}^+\}$ et une constante $\lambda > 0$ telle que pour tout $\theta \in \Theta_n$:

$$|s_{nt}(\theta^0) - s_{nt}(\theta)| < m_{nt}(\theta) d(\theta^0, \theta)^\lambda$$

pour tout $\theta^0 \in \eta_n(\theta)$ où

$$\eta_n(\theta) = \{\theta^0 \in \Theta_n, d(\theta^0, \theta) < d_n(\theta)\}$$

soit $M_n \geq \sup_{\theta \in \Theta_n} \sum_{t=1}^n E(m_{nt}(\theta))$

On suppose en plus qu'il existe des fonctions:

$$\gamma_n^s : \Theta_n \times \mathbb{R}^+ \rightarrow \mathbb{R}^+, \gamma_n^m : \Theta_n \times \mathbb{R}^+ \rightarrow \mathbb{R}^+$$

telle que pour tout $\theta \in \Theta_n$

$$P \left[\left| \sum_{t=1}^n s_{nt}(\theta) - E(s_{nt}(\theta)) \right| > \Delta \right] \leq \gamma_n^s(\theta, \Delta)$$

$$P \left[\left| \sum_{t=1}^n m_{nt}(\theta) - E(m_{nt}(\theta)) \right| > \Delta \right] \leq \gamma_n^m(\theta, \Delta), n = 1, 2, \dots$$

On définit:

$$\Gamma_n^s(\Delta) = \sup_{\theta \in \Theta_n} \gamma_n^s(\theta, \Delta), \quad \Gamma_n^m(\Delta) = \sup_{\theta \in \Theta_n} \gamma_n^m(\theta, \Delta); \quad \Delta \in \mathbb{R}^+, n = 1, 2, \dots$$

Soit $G_n(\varepsilon)$ le plus petit nombre d'ensembles ouverts de rayon ε qui recouvrent Θ_n

$$H_n(\varepsilon) = \log G_n(\varepsilon)$$

l'entropie métrique de Θ_n . Soit

$$\{a_n\} = O(M_n \inf_{\theta \in \Theta_n} d_n(\theta)^\lambda)$$

et on suppose que pour tout $\varepsilon > 0$:

$$G_n([\varepsilon a_n / 6M_n]^{1/\lambda}) = O(\min[\Gamma_n^s(M_n)^{-1}, \Gamma_n^m(\varepsilon a_n / 3)^{-1}])$$

Alors:

pour tout $\varepsilon > 0$

$$P(\sup_{\theta \in \Theta_n} |S_n(\theta) - \bar{S}_n(\theta)| > \varepsilon a_n) \rightarrow 0 \text{ quand } n \rightarrow +\infty$$

$$\text{où } S_n(\theta) = \sum_{t=1}^n s_{nt}(\theta), \quad \bar{S}_n(\theta) = E(S_n(\theta))$$

Le résultat implique en particulier :

$$P(\omega, \sup_{\theta \in \Theta_n} |Q_n(\omega, \theta) - \bar{Q}_n(\theta)| > \varepsilon) \rightarrow 0 \text{ quand } n \rightarrow +\infty.$$

Corollaire 1.4.1 *Sous les hypothèses du théorème 1.4.1. On suppose qu'il existe une fonction $\bar{Q} : \Theta \rightarrow \mathbb{R}$ telle que pour tout $\varepsilon > 0$*

$$P[\omega, \sup_{\theta \in \Theta_n} |Q_n(\omega, \theta) - \bar{Q}(\theta)| > \varepsilon] \rightarrow 0 \text{ quand } n \rightarrow +\infty$$

\bar{Q} a un unique minimum fini en $\theta_0 \in \Theta$ et \bar{Q} est continue en θ_0 . Soit $\{\Theta_n\}$ est une suite croissante et $\bigcup_n \Theta_n$ est dense dans Θ . Alors:

$$d(\hat{\theta}_n, \theta_0) \rightarrow 0 \text{ en } P \text{ probabilité}$$

Ce résultat est quelque peu similaire au résultat de [12].

Ici on n'exige pas que P soit un produit infini de mesures de probabilités appropriée à un processus de variables aléatoires indépendantes identiquement distribuées.

Une version de l'inégalité de Bernstein appropriée à cette application est la suivante:

Théorème 1.4.3 [23] *Pour tout $n = 1, 2, \dots$. Soit $\{X_{nt}\}$ un processus stationnaire tel que*

$$|X_{nt}| < D_n < +\infty, E(X_{nt}) = 0$$

ou bien $\varphi(k) = \varphi_0 \rho^k$ ou $\alpha(k) = \alpha_0 \rho^k \alpha_0$, $\varphi_0 > 0$, $0 < \rho < 1$, $k \geq 1$. Alors:

Ils existent des constantes $0 < c_1, c_2 < +\infty$ non dépendantes de n telles que

$$P \left[\left| \sum_{t=1}^n X_{nt} \right| > \Delta \right] \leq c_1 \exp \left[-c_2 \Delta n^{-1/2} / D_n \right]$$

Ces résultats précédents s'appliquent pour obtenir la convergence et la vitesse de convergence des estimateurs "sieves".

Dans le modèle suivant et pour des raisons de commodité, on considère un processus φ ou α -mélangeant stationnaire (Y_t) sur l'intervalle $\Pi = [0, 1]$.

L'intérêt est l'estimation sieve de la fonction de régression $E(Y_t/Y_{t-1}) = \theta_0(Y_{t-1})$ où θ_0 représente la fonction inconnue. On suppose ici que $\theta_0 \in \Theta = C([0, 1])$ et on munit Θ de la norme du sup $\|\cdot\|$.

Par conséquent, (Θ, d) est un espace métrique séparable avec

$$d(\theta_1, \theta_2) = \|\theta_1 - \theta_2\|$$

Une variété de choix pour Θ_n est possible. Si on considère la méthode des moindres carrés l'estimateur est donc

$$\hat{\theta}_n = \arg \min_{\theta \in \Theta_n} n^{-1} \sum_{t=1}^n (Y_t - \theta(Y_{t-1}))^2$$

où $\{\Theta_n\}$ une suite sieve de Θ compacts choisis convenablement.

Le résultat suivant donne la convergence en probabilité de l'estimateur sieve par des moindres carrés.

Proposition 1.4.1 [23] *Soit (Ω, F, P) un espace de probabilité complet et soit $\{Y_t : \Omega \rightarrow \Pi\}$ un processus stationnaire α ou (φ) -mélangeant avec $\varphi(k) = \varphi_0 \rho^k$ ou $\alpha(k) = \alpha_0 \rho^k$, $\alpha_0, \varphi_0 > 0$, $0 < \rho < 1$, $k > 0$,*

On suppose que

$$E(Y_t/Y_{t-1}) = \theta_0(Y_{t-1}), \quad \theta_0 \in \Theta$$

Et soit

$$d(\theta_1, \theta_2) = \|\theta_1 - \theta_2\|$$

Où $\|\cdot\|$ la norme du sup. Soit

$$\{\psi_j : \Pi \rightarrow [-1, 1], j = 1, 2, \dots\}$$

une suite de fonctions continues linéairement indépendantes. Soit $\{q_n\}$ une suite d'entiers positifs et soit $\{B_n\}$ une suite de nombres réels positifs. On définit les "sieves"

$$\Theta_n = \{\theta : \Pi \rightarrow \mathbb{R}, \theta = \sum_{j=1}^{q_n} \beta_j \psi_j, |\beta_j| \leq B_n, j = 1, \dots, q_n\}$$

Telles que

$$\Theta = \overline{\bigcup_{n=1}^{+\infty} \Theta_n}.$$

On suppose que θ_0 est la solution unique du problème

$$\min_{\theta \in \Theta} E([Y_t - \theta(Y_{t-1})]^2)$$

et on définit $\hat{\theta}_n$ comme solution mesurable du problème

$$\min_{\theta \in \Theta_n} n^{-1} \sum_{t=1}^n (Y_t - \theta(Y_{t-1}))^2 / 2.$$

Si pour $0 < B_0 < +\infty$, on choisit $B_n = B_0(\log n)^{1/2}$, $B_n q_n \geq 1$ et on prend

$$q_n = O(n^{1/6} / \log n)$$

Alors:

$$\|\hat{\theta}_n - \theta_0\| \rightarrow 0 \text{ en } P \text{ probabilité}$$

1.4.2 Paramètres dans l'espace l^2 et L^2

Les exemples traités par la suite sont dans [13] [p.p. 420, 426, 444].

Nous considérons séparément les cas de $\Theta = l^2$ et $\Theta = L^2$.

Soit $\Theta = l^2$

$$x = \alpha + y; \quad \alpha \in \Theta$$

où x et y sont des variables aléatoires à valeurs dans l^2 , y a une loi gaussienne sur Θ de moyenne nulle et d'opérateur de covariance R .

Nous avons le résultat suivant : pour que les probabilités P_α et P_{α_0} soient équivalentes il faut que

$$\alpha - \alpha_0 \in \text{Im}(R^{\frac{1}{2}})$$

Il est donc naturel de supposer que $\alpha \in \text{Im}(R^{\frac{1}{2}})$. L'opérateur de covariance R est à trace.

On introduit ses valeurs propres λ_ν et vecteurs propres Φ_ν .

Le développement de Karhunen-Loève de y est sous la forme

$$y = \sum_{\nu=1}^{+\infty} \sqrt{\lambda_\nu} y_\nu \Phi_\nu$$

où $\{\Phi_\nu\}$ sont des vecteurs orthogonaux dans Θ .

Posons $(x, \Phi_\nu) = x_\nu$ les observations où (\cdot, \cdot) est le produit scalaire.

La dérivée de Radon Nicodym est donnée par l'expression

$$\frac{P_\alpha(dx)}{P_0(dx)} = \prod_{\nu=1}^{\infty} \exp \left[\frac{2\alpha_\nu x_\nu - \alpha_\nu^2}{2\lambda_\nu} \right] \quad (1.5)$$

Si on maximise la vraisemblance (1.5) par rapport au paramètre α , on obtient l'estimateur du maximum de vraisemblance

$$\alpha^* = x_\nu$$

Si on dispose d'un échantillon d'observations indépendantes identiquement distribuées $x^{(1)}, \dots, x^{(n)} \in l^2$. alors

$$\alpha^* = \frac{x^{(1)} + \dots + x^{(n)}}{n} = \bar{x}$$

D'après le chapitre 4 section 6 [13] l'estimateur \bar{x} est consistant dans l^2 .

Exemple

Soit $X = (x_1, \dots, x_n)$ où x est donné par l'expression

$$x_\nu = (1 + \alpha_\nu)\xi_\nu$$

et (ξ_ν) sont iid de distribution exponentielle sur \mathbb{R} de moyenne égale à 1.

Si

$$\alpha = (\alpha_1, \dots, \alpha_n) \in l^2$$

alors on applique le théorème de Kakutani [9] pour $f(x) = e^{-x}$. La condition suivante

$$\frac{1}{8} \left[\int_0^{\infty} x^2 e^{-x} dx - 1 \right] = \frac{1}{8} < \infty$$

est vérifiée par conséquent les probabilités P_α et P_{α_0} sont équivalentes se réduit à $\alpha \in l^2$.

Par conséquent on obtient la dérivée de Radon Nicodym par

$$\frac{P_\alpha(dx)}{P_0(dx)} = \prod_{\nu=1}^{\infty} \frac{1}{1 + \alpha_\nu} \exp \left[\frac{\alpha_\nu x_\nu}{1 + \alpha_\nu} \right] \quad (1.6)$$

Pour chaque facteur dans (1.6), on doit choisir

$$\alpha_{*\nu} = (x_\nu - 1) = (\xi_\nu - 1) + \alpha_\nu \xi_\nu$$

Avec une probabilité égale à 1, la suite $(\alpha_\nu \xi_\nu)$ est dans l^2 et $(\xi_\nu - 1)$ n'est presque sûrement pas dans l^2 , donc $\alpha_* \notin l^2$.

Dans cet exemple une "sieve" est nécessaire bien que l'espace Θ soit le même pour l'exemple précédent. Ce qui montre que l'utilisation de "sieve" ne dépend pas seulement de la topologie de l'espace paramétrique Θ mais dépend aussi de la manière dont la mesure P_α dépend de α dans Θ .

Soit $\Theta = L^2$ et nous considérons l'estimation du paramètre la moyenne d'un processus gaussien. Soit le processus $(X(t), t \in [-\frac{1}{2}, \frac{1}{2}])$ vérifiant:

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

où $\alpha \in L_2[-\frac{1}{2}, \frac{1}{2}]$, W est le processus de Wiener de variance 1. On suppose que les coefficients de Fourier de α par rapport au sinus sont nuls.

On observe $(X(t), t \in [-\frac{1}{2}, \frac{1}{2}])$ et on veut estimer α . Nous avons

$$\begin{aligned} x_k &= \sqrt{2} \int_{-\frac{1}{2}}^{\frac{1}{2}} \cos 2\pi kt X(dt) = \alpha_k + w_k \\ \alpha_k &= \sqrt{2} \int_{-\frac{1}{2}}^{\frac{1}{2}} \cos 2\pi kt \alpha(t) dt \\ w_k &= \sqrt{2} \int_{-\frac{1}{2}}^{\frac{1}{2}} \cos 2\pi kt W(dt) \end{aligned}$$

Le processus $(X_t, t \in [-\frac{1}{2}, \frac{1}{2}])$ induit une loi de probabilité P_α sur $C([-\frac{1}{2}, \frac{1}{2}])$. la dérivée de Radon Nicodým de P_α par rapport à P_0 est (cf [12]):

$$\frac{P_\alpha(dx)}{P_0(dx)} = \exp\left(\sum_{k=0}^{+\infty} (\alpha_k x_k - \frac{1}{2} \alpha_k^2)\right) \quad (1.7)$$

Puisque $\alpha_k \in l^2$, P_α et P_0 sont équivalentes. Pour maximiser (1.7), on maximise chaque terme de la série. Alors le maximum de vraisemblance $\alpha_k^* = x_k$. Mais

$$x_k = \alpha_k + w_k \quad \text{et} \quad \alpha_k \in l^2$$

or

$$w_k \notin l^2 \quad \text{car} \quad \sum_{k=0}^{+\infty} \text{var}^2(w_k) = \sum_{k=0}^{+\infty} 1 = +\infty \quad \text{presque sûrement.}$$

donc $x_k \notin l^2$. D'où le maximum de la vraisemblance n'existe pas dans l^2 .

On peut choisir une "sieve" Θ_m définie par

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

Θ_m est compact, $\Theta_m \subset \Theta_{m+1}$ et $\bigcup \Theta_m = \Theta$.

Soit λ le multiplicateur de Lagrange. Pour obtenir le maximum de vraisemblance, on a à résoudre le problème suivant:

$$\left\{ \begin{array}{l} \max_{\alpha_k} \sum_{k=0}^{\infty} \sum_{i=1}^n (\alpha_k x_k^i - \frac{1}{2} \alpha_k^2) - \lambda k^2 \alpha_k^2 \\ \sum_{k=0}^{\infty} k^2 \alpha_k^2 \leq m \end{array} \right.$$

On obtient alors

$$\hat{\alpha}_m^n(t) = \sum_{k=0}^{\infty} \hat{\alpha}_k \cos(2\pi kt) \quad \text{où}$$

$$\hat{\alpha}_k = \frac{\sum_{i=1}^n x_k^i}{n + \lambda k^2}; \quad \sum_{k=0}^{\infty} k^2 \frac{(\sum_{i=1}^n x_k^i)^2}{(n + \lambda k^2)^2} = m$$

Cet estimateur est consistant voir le théorème 1.4.4 qui sera énoncé ultérieurement.

1.4.3 Paramètres dans l'espace $C_{[0,1]}$

Soit α une fonction continuellement différentiable par morceaux sur l'intervalle $[0, 1]$: $\alpha \in \Theta = PC^1([0, 1])$. Considérons le modèle

$$x_\nu = \alpha\left(\frac{\nu}{n}\right) + z_\nu \quad \nu = 1, 2, \dots, n$$

où z_ν sont iid et suivent la loi normale réduite $N(0,1)$.

Pour estimer la fonction de régression α par le maximum de vraisemblance on a à résoudre le problème suivant

$$\min_{\alpha} \sum_{\nu=1}^n [x_\nu - \alpha\left(\frac{\nu}{n}\right)]^2$$

le suprémum est atteint pour $\alpha_{MV}^* \in \Theta$ donné par

$$\alpha_{MV}^*\left(\frac{\nu}{n}\right) = x_\nu, \quad \alpha_{MV}^* \text{ est pris arbitraire pour le reste dans } [0, 1].$$

Il n'est pas possible de spécifier la fonction α^* dans le reste de l'intervalle $[0,1]$ de sorte qu'on puisse garantir une convergence en probabilité de α_{MV}^* vers α par rapport à la topologie uniforme de Θ . D'où la nécessité de passer par une "sieve".

On introduit une "sieve" dans Θ

$$\Theta(\mu) = \left\{ \alpha / \alpha \in PC^1 \quad |\alpha| \text{ et } |\dot{\alpha}| \leq \frac{1}{\mu} \right\}$$

On montre que cette "sieve" rend l'estimateur du maximum de vraisemblance consistant. La quantité $\sup L(\alpha)$ est atteinte en un point de $\Theta(\mu)$ puisqu'on a à résoudre

$$\min_{\alpha \in \Theta(\mu)} \sum_{\nu=1}^n [x_\nu - \alpha\left(\frac{\nu}{n}\right)]^2$$

et ceci dépend seulement des valeurs de α aux points $\frac{\nu}{n}$ de telle sorte que la fonction de vraisemblance puisse être vue comme une fonction continue sur un compact de \mathbb{R}^n . Il n'est pas important d'avoir les dérivées aux points $\frac{\nu}{n}$ mais entre ces points, on peut rendre α linéaire si μ est suffisamment petit pour un échantillon donné. Si $\beta \in C(0, 1)$ l'espace des fonctions continues sur $[0, 1]$ nous considérons la forme quadratique s :

$$\begin{aligned}
s(\beta) &= \frac{1}{n} \sum_{\nu=1}^n [x_\nu - \beta(\frac{\nu}{n})]^2 - 1 \\
&= \frac{1}{n} \sum_{\nu=1}^n [\alpha(\frac{\nu}{n}) - \beta(\frac{\nu}{n})]^2 + (\frac{1}{n} \sum_{\nu=1}^n z_\nu^2 - 1) + \frac{2}{n} \sum_{\nu=1}^n [\alpha(\frac{\nu}{n}) - \beta(\frac{\nu}{n})] z_\nu \\
&= : s_1 + s_2 + s_3
\end{aligned} \tag{1.8}$$

Quand $n \rightarrow +\infty$ on a la convergence en probabilité

$$\begin{aligned}
s_1 &\rightarrow \int_0^1 [\alpha(x) - \beta(x)]^2 dx = I(\beta) \\
s_2 &\rightarrow 0 \\
s_3 &= \frac{2}{n} \sum_{\nu=1}^n [\alpha(\frac{\nu}{n}) - \beta(\frac{\nu}{n})] z_\nu \\
&\leq 2 \frac{\max}{n} [\alpha(\frac{\nu}{n}) - \beta(\frac{\nu}{n})] \sum_{\nu=1}^n z_\nu \rightarrow 0
\end{aligned} \tag{1.9}$$

Donc

$$s_3 \rightarrow 0 \quad \text{en probabilité}$$

Or

$$s = -\left(\frac{2}{n}\right) \ln L + c$$

Alors (1.9) montre qu'avec une grande probabilité s est proche de $I(\beta)$.

On fait tendre μ vers zéro. Alors avec $\beta = \alpha$ dans (1.8) et (1.9) on voit que

$$\frac{1}{n} \sum_{\nu=1}^n [x_\nu - \alpha(\frac{\nu}{n})]^2 \rightarrow 1 \quad \text{en probabilité quand } n \rightarrow +\infty \tag{1.10}$$

Mais étant donné que α_μ^* est l'EMV dans les sieves $\Theta(\mu)$ alors

$$\begin{aligned}
\frac{1}{n} \sum_{\nu=1}^n [x_\nu - \alpha_\mu^*(\frac{\nu}{n})]^2 &= \frac{1}{n} \sum_{\nu=1}^n [x_\nu - \alpha(\frac{\nu}{n})]^2 + \frac{2}{n} \sum_{\nu=1}^n z_\nu [\alpha(\frac{\nu}{n}) - \alpha_\mu^*(\frac{\nu}{n})] \\
&\quad + \frac{1}{n} \sum_{\nu=1}^n [\alpha(\frac{\nu}{n}) - \alpha_\mu^*(\frac{\nu}{n})]^2.
\end{aligned}$$

Cette quantité n'est pas plus grande que (1.10) par la définition de α_μ^* .

Donc $\frac{1}{n} \sum_{\nu=1}^n z_\nu [\alpha(\frac{\nu}{n}) - \alpha_\mu^*(\frac{\nu}{n})]$ peut être écrite sous la forme

$$\frac{2}{n} \left[\sum_{\nu=1}^{n-1} Z_\nu \left[\alpha\left(\frac{\nu}{n}\right) - \alpha\left(\frac{\nu+1}{n}\right) \right] + Z_n \alpha(1) \right] + 2^{nd} \text{ terme} \quad (1.11)$$

où

$$Z_\nu = z_1 + z_2 + \dots + z_\nu$$

Ainsi le 2^{nd} terme se présente de la même façon sauf que $\alpha_\mu^*(\frac{\nu}{n})$ remplace $\alpha(\frac{\nu}{n})$ et $\alpha \in \Theta(\mu)$. Donc le premier terme dans (1.11) peut être majoré par

$$\frac{c}{n} \sum_{\nu=1}^n E |Z_\nu| + \frac{c}{n} E |Z_n| \quad (1.12)$$

où la constante c dépend seulement de μ . Mais Z_ν suit une loi $N(0, \nu)$ donc (1.12) est majoré par

$$\frac{c}{n} \sum_{\nu=1}^n \sqrt{\nu} + \frac{c}{\sqrt{n}} \quad \text{quand } n \rightarrow +\infty.$$

en utilisant l'inégalité de Cauchy Schwartz.

Par suite

$$\frac{2}{n} \sum_{\nu=1}^n \left[\alpha\left(\frac{\nu}{n}\right) - \alpha_\mu^*\left(\frac{\nu}{n}\right) \right]^2 \rightarrow 0 \text{ quand } n \rightarrow +\infty \quad \text{en probabilité.}$$

Supposons que pour un δ fixé, $\exists x_0 \in [0, 1]$ tel que

$$\alpha(x_0) - \alpha_\mu^*(x_0) > \delta \quad (1.13)$$

puisque

$$|\alpha_\mu^*(x_1) - \alpha_\mu^*(x_2)| \leq \frac{|x_1 - x_2|}{\mu}.$$

car

$$\alpha_\mu^* \in \Theta(\mu).$$

Il s'ensuit que si on considère les ν pour lesquels

$$\left| \alpha\left(\frac{\nu}{n}\right) - \alpha(x_0) \right| < \frac{\delta}{2}.$$

alors

$$\begin{aligned} \alpha\left(\frac{\nu}{n}\right) - \alpha_{\mu}^*\left(\frac{\nu}{n}\right) &= \alpha\left(\frac{\nu}{n}\right) - \alpha(x_0) + \alpha(x_0) - \alpha_{\mu}^*(x_0) + \alpha_{\mu}^*(x_0) - \alpha_{\mu}^*\left(\frac{\nu}{n}\right) \\ &\geq \delta - \frac{\delta}{2} - \frac{1}{\mu} \left| x_0 - \frac{\nu}{n} \right| \\ &= \frac{\delta}{2} - \frac{1}{\mu} \left| x_0 - \frac{\nu}{n} \right|. \end{aligned} \quad (1.14)$$

Pour les mêmes ν , considérons ceux satisfaisant à

$$\left| \frac{\nu}{n} - x_0 \right| < \frac{\mu\delta}{4}. \quad (1.15)$$

Faisons la somme sur ν et notons la par \sum' et on a :

$$\begin{aligned} \frac{1}{n} \sum_{\nu=1}^n [\alpha\left(\frac{\nu}{n}\right) - \alpha_{\mu}^*\left(\frac{\nu}{n}\right)]^2 &\geq \frac{1}{n} \sum' \left[\frac{\delta}{2} - \frac{1}{\mu} \left| x_0 - \frac{\nu}{n} \right| \right]^2 \\ &\geq \frac{\delta^2}{16n} \#\{\nu \text{ dans } \sum'\}. \end{aligned}$$

où $\#$ désigne le cardinal de l'ensemble.

De (1.14) et (1.15) on a

$$\#\{\nu \text{ dans } \sum'\} \geq \delta n \min\left(\frac{\mu}{4}, \frac{1}{\mu}\right) = \delta n K$$

Donc

$$\frac{1}{n} \sum_{\nu=1}^n [\alpha\left(\frac{\nu}{n}\right) - \alpha_{\mu}^*\left(\frac{\nu}{n}\right)]^2 \geq \frac{\delta^2 k}{16}.$$

Comme

$$\frac{1}{n} \sum_{\nu=1}^n [\alpha\left(\frac{\nu}{n}\right) - \alpha_{\mu}^*\left(\frac{\nu}{n}\right)]^2 \rightarrow 0 \text{ en } P \text{ probabilité.}$$

par conséquent on peut rendre cette quantité aussi petite que l'on veut ce qui contredit (1,13). On fait de même pour

$$\alpha(x_0) - \alpha_{\mu}^*(x_0) < -\delta$$

pour aboutir à une contradiction.

Ainsi, en notant aussi que les bornes ne dépendent pas de x_0 , l'estimateur sieve α^* converge uniformément vers α en probabilité, ce qui mène au théorème suivant:

Théorème 1.4.4 [13] *Le maximum de vraisemblance α_{μ}^* dans $\Theta(\mu)$ tend uniformément vers α en probabilité si μ tend vers zéro quand n tend vers $+\infty$.*

1.4.4 Opérateurs comme paramètres

Dans cette partie α est un opérateur sur l^2 et il est représenté par une matrice diagonalisable. On peut noter α par la matrice diagonale notée $(\alpha_1, \alpha_2, \dots)$ où (α_i) des nombres positifs $(\alpha_i) \in l^1$.

En introduisant la dérivé de Radon Nicodym de P_α par rapport à P_β , on a :

$$\beta = (\beta_1, \beta_2, \dots), \quad \beta_\nu > 0 \quad \sum \beta_\nu < +\infty,$$

$$\frac{dP_\alpha}{dP_\beta}(x) = \prod_{i=1}^n \prod_{\nu=1}^{+\infty} \sqrt{\frac{\beta_\nu}{\alpha_\nu}} \exp \left[\frac{1}{2} \left[\frac{1}{\beta_\nu} - \frac{1}{\alpha_\nu} \right] (x_i^\nu)^2 \right]$$

L'estimateur emv est

$$\alpha^* = \left\{ \frac{1}{n} \sum_{i=1}^n (x_i^\nu)^2, \quad \nu = 1, 2, \dots \right\} \in \Theta$$

où Θ est l'espace des opérateurs de Hilbert Schmidt sur l^2 . Notons par (V_1, V_2, \dots) la vraie valeur du paramètre α et considérons α_i^* les coordonnées de α^* qui sont stochastiquement indépendantes.

$$\alpha_i^* = \frac{V_i}{n} \chi_n^2(i).$$

Mais

$$\|\alpha^* - \alpha\| = \max_i |\alpha_i^* - V_i|.$$

ceci est égale en loi à

$$e_n = \max_i V_i \left| \frac{1}{n} \chi_n^2(i) - 1 \right|.$$

qui peut s'écrire aussi

$$e_n = \max(e_n^+, e_n^-) \quad \begin{cases} e_n^+ = \max V_i \left[\frac{1}{n} \chi_n^2(i) - 1 \right] \\ e_n^- = \min V_i \left[\frac{1}{n} \chi_n^2(i) - 1 \right] \end{cases}$$

Considérons la loi de e_n^+ et il en sera de même pour e_n^- . Pour $\varepsilon > 0$

$$P(e_n^+ < \varepsilon) = \prod_{i=1}^{\infty} F_n \left(\frac{n}{V_i} (1 + \varepsilon) \right)$$

où F_n est la fonction de répartition de χ_n^2 . Alors;

$$-\ln P\{e_n^+ < \varepsilon\} = \sum_{i=1}^{\infty} \left\{ -\ln F_n \left[n \frac{1 + \varepsilon}{V_i} \right] \right\}$$

En utilisant l'argument de la convergence dominée et en appliquant l'inégalité de Chebyshev avec le moment d'ordre 1 et

$$E |\chi_n^2 - n| = \frac{n^{\frac{n}{2}} e^{-\frac{n}{2}}}{2^{\frac{n}{2}-1} \Gamma(\frac{n}{2})}$$

et en utilisant l'approximation de Stirling

$$E |\chi_n^2 - n| \sim \sqrt{\frac{n}{\pi}}$$

on obtient une borne asymptotiquement de

$$1 - P(|\chi_n^2 - n| < \frac{\varepsilon}{V_i}) \leq \frac{E |\chi_n^2 - n|}{\varepsilon} V_i.$$

et

$$\frac{E |\chi_n^2 - n|}{V_i} V_i \sim \sqrt{\frac{n V_i}{\pi \varepsilon}}.$$

Donc

$$1 - F_n[n(1 + \frac{\varepsilon}{V_i})] = \frac{1}{\sqrt{n}} O(V_i).$$

Or

$$V_i \rightarrow 0 \text{ quand } i \rightarrow +\infty.$$

Par suite

$$\begin{aligned} \ln F_n[n(1 + \frac{\varepsilon}{V_i})] &\geq 2[F_n[\frac{n}{V_i}(1 + \varepsilon)] - 1] \\ -\ln F_n[n(1 + \frac{\varepsilon}{V_i})] &\leq 2[1 - F_n(\frac{n}{V_i}(1 + \varepsilon))] \\ &= \frac{1}{\sqrt{n}} O(V_i). \end{aligned}$$

De

$$-\ln p(e_n^+ < \varepsilon) \leq \frac{c}{\sqrt{n}} \sum_i V_i = \frac{c}{\sqrt{n}} Tr(\alpha) \rightarrow 0 \text{ quand } n \rightarrow +\infty.$$

donc

$$P(e_n^+ < \varepsilon) \rightarrow 1 \text{ quand } n \rightarrow +\infty$$

Pour l'inégalité inverse, on a pour $\varepsilon > 0$

$$-\ln P(e_n^- > -\varepsilon) = \sum_{i=1}^{+\infty} -\ln[1 - F_n(n(1 - \frac{\varepsilon}{V_i}))].$$

Mais

$$F_n[n(1 - \frac{\varepsilon}{V_i})] = \frac{1}{\sqrt{n}} O(V_i).$$

Comme l'opérateur est à trace

$$-\ln P(e_n^- > -\varepsilon) \leq \frac{1}{\sqrt{n}} c \sum_i V_i = \frac{1}{\sqrt{n}} c \text{Tr}(\alpha) \rightarrow 0.$$

Ceci implique que

$$P(e_n^- > -\varepsilon) \rightarrow 1.$$

Donc

$$e_n \rightarrow 0 \quad \text{en probabilité.}$$

Alors l'estimateur α^* converge en probabilité pour la norme des opérateurs. D'où le théorème:

Théorème 1.4.5 [13] *L'estimateur du maximum de vraisemblance de l'opérateur de covariance converge en probabilité par rapport à la norme des opérateurs, pour tout $\varepsilon > 0$*

$$P\{\|\alpha^* - \alpha\| > \varepsilon\} \rightarrow 0 \text{ quand } n \rightarrow +\infty.$$

1.4.5 Régression non paramétrique

Les résultats et l'exemple suivant sont dans [22]. Nous avons besoin de quelques préliminaires.

Préliminaires:

Définition 1.4.2 *La quantité ε -entropie métrique d'un espace métrique (Θ, d) est le logarithme du nombre minimal de ε -boules qui recouvrent Θ . Si la métrique utilisée est celle de L^2 , on obtient L^2 -entropie métrique.*

Pour définir l'entropie avec "bracketing", considérons

$$F = \{f(\theta, y); \theta \in \Theta\}$$

où f est mesurable définie de $\theta \times X$ dans \mathbb{R} , et $E^2 f(t, Y)$ finie pour tout $\theta \in \Theta$. Si on considère la norme $\|\cdot\|_2$ dans L^2 , alors pour tout $f(s, Y), f(t, Y) \in F$

$$\|f(s, Y) - f(t, Y)\|_2 = (E(f(s, Y) - f(t, Y))^2)^{1/2}.$$

où l'espérance est calculée sous la loi de Y . Pour tout $\varepsilon > 0$, s'il existe

$$S(\varepsilon, k) = \{f_1^l, f_1^u, \dots, f_k^l, f_k^u\} \subset L^2.$$

avec

$$\max_{j \leq k} \|f_j^u - f_j^l\|_2 \leq \varepsilon.$$

tel que pour tout $f \in F$, il existe j avec $f_j^l \leq f \leq f_j^u$ pp.

Définition 1.4.3 $S(\varepsilon, k)$ est appelé " ε -bracketing" recouvrement de F par rapport à $\|\cdot\|_2$.

Soit

$$N_2^B(\varepsilon, F) = \min\{k; S(\varepsilon, k) \text{ est un } \varepsilon\text{-bracketing} \text{ recouvrement de } F\}$$

Le nombre

$$H_2^B = \log N_2^B(\varepsilon, F).$$

est appelé l'entropie "bracketing" métrique L^2 de F .

Les conditions suivantes sont utilisées pour énoncer le théorème suivant.

Condition C1:

pour $A_1 > 0$ et $\alpha > 0$ et $\varepsilon > 0$ assez petit

$$\inf_{\{\rho(\theta, \theta_0) \geq \varepsilon, \theta \in \Theta_n\}} E(l(\theta_0, Y) - l(\theta, Y)) \geq 2A_1 \varepsilon^{2\alpha}.$$

Remarque 1 La quantité $l(\theta_0, y) - l(\theta, y)$ est appelée critère de différence. Pour obtenir la convergence de l'estimateur, l'espérance de cette quantité doit tendre vers zéro. Ainsi, la condition **C1** précise la minoration de cette espérance quand θ s'éloigne de θ_0 . D'autre part, quand θ tend vers θ_0 , cette différence doit tendre vers zéro.

Condition C2:

Pour des constantes $A_2 > 0, \beta > 0$, et pour tout ε petit

$$\sup_{\{\rho(\theta, \theta_0) \leq \varepsilon, \theta \in \Theta_n\}} E(l(\theta_0, Y) - l(\theta, Y))^2 \leq A_2 \varepsilon^{2\beta}$$

et

$$\sup_{\theta \in \Theta_n} E(l(\theta_0, Y) - l(\theta, Y))^2 \leq A_2 n^{2l}$$

Remarque 2 La condition C2 contrôle la distance entre $l(\theta, y)$ et $l(\theta_0, y)$, particulièrement quand $\rho(\theta, \theta_0)$ tend vers zéro.

Condition C3

Soit

$$T_n(\theta, y) = l(\theta, y) - l(\pi_n \theta_0, y)$$

et

$$T_n^{(b, \delta)}(\theta, y) = T_n(\theta, y) I_{\left(\sup_{\{\rho(\theta, \pi_n \theta_0) \leq \delta, \theta \in \Theta_n\}} (T_n(\theta, y) \leq b) \right)}$$

où $I_{\left(\sup_{\{\rho(\theta, \pi_n \theta_0) \leq \delta, \theta \in \Theta_n\}} (T_n(\theta, y) \leq b) \right)}$ est l'indicatrice de $\left(\sup_{\{\rho(\theta, \pi_n \theta_0) \leq \delta, \theta \in \Theta_n\}} (T_n(\theta, y) \leq b) \right)$, et supposons pour les constantes $0 \leq r_0 \leq \frac{1}{2}$, $r_i \geq 0$, $i = 1, 2$. $r \geq 0^+$ et $A_3 > 0$

$$H_2^B(\varepsilon, F_n^{(b, \delta)}) \leq A_3 n^{2r_0} \max\left(\left(\frac{\varepsilon}{b r_1 \delta^{r_2}}\right)^{-r}, 2\right)$$

où

$$H_2^B(\varepsilon, F_n^{(b, \delta)}) \text{ est l'entropie "bracketing" métrique } L^2.$$

de

$$F_n^{(b, \delta)} = \{T_n^{(b, \delta)}(\theta, y), \rho(\theta, \pi_n \theta_0) \leq \delta, \theta \in \Theta_n\}.$$

Remarque 3 La condition C4 complète C3 en fournissant une borne supérieure à la partie tronquée de $G(\delta)$.

Condition C4

Pour $B \geq 0$, il existe

$$g(\delta) = O(\min(\delta^B, 1))$$

pour δ petit positif, et

$$n = O(n^{2\kappa})$$

où κ est un nombre positif avec

$$\kappa \begin{cases} < \frac{s(1-2r_0)}{2(4\alpha-2\beta)} & \text{si } r = 0^+; \\ < \frac{s(1-2r_0)}{2(4\alpha-\beta(2-r)+r_1 r_0)} & \text{si } 0 < r < 2; \\ \leq \frac{1-2r_0+4!}{4} - \frac{\ln \ln n}{2 \ln n} & \text{si } r = 2; \\ \leq \frac{1-2r_0+2r!}{2r} & \text{si } r > 2; \end{cases}$$

tel que

$$P\left(\sup_{\{\rho(\theta, \pi_n \theta_0) \leq \delta, \theta \in \Theta_n\}} T_n(\theta, Y) \geq g(\delta) b_n \right) \leq A_4 a_n$$

pour $A_4 > 0$, où

$$a_n = \begin{cases} o(1/n \ln \ln n) & \text{si } \beta \leq \alpha; \\ o(1/n) & \text{si } \beta > \alpha; \end{cases}$$

Remarque 4 La condition **C4** complète **C3** en fournissant une borne supérieure à la partie tronquée de $G(\delta)$.

$$G(\delta) = \sup\{g(\delta)^{-1}(l(\theta, y) - l(\pi_n \theta_0, y)); \quad \rho(\theta, \pi_n \theta_0) \leq \delta, \quad \theta \in \Theta_n\}$$

Théorème 1.4.6 [22] *On suppose que les conditions **C1**, **C3**, et **C4** sont vérifiées et $\hat{\theta}_n$ satisfait à*

$$L_n(\hat{\theta}_n) \geq \sup_{\theta \in \Theta_n} L_n(\theta) - \eta_n$$

avec $\eta_n = o(n^{-\omega})$, où

$$L_n(\theta) = (1/n) \sum_{i=1}^n \ln p(\theta, Y_i) = 1/n \sum_{i=1}^n l(\theta, Y_i)$$

$$\omega = \begin{cases} \frac{2(1-2r_0)}{2}, & \text{si } r = 0^+ \\ \frac{2(1-2r_0)}{2+r}, & \text{si } 0 < r < 2 \\ \frac{1-2r_0}{2} - \frac{\ln \ln n}{\ln n}, & \text{si } r = 2 \\ \frac{1-2r_0}{r}, & \text{si } r > 2 \end{cases}$$

il est supposé aussi que la condition **C2** est vérifiée pour $0^+ \leq r \leq 2$.

Alors:

$$\rho(\hat{\theta}_n, \theta_0) = O_p(\max(n^{-\tau}, \rho(\pi_n \theta_0, \theta_0), K^{1/2\alpha}(\pi_n \theta_0, \theta_0)))$$

où

$$K(\pi_n \theta_0, \theta_0) = E(l(\theta_0, Y) - l(\pi_n \theta_0, Y))$$

et

$$\tau = \begin{cases} \frac{1-2r_0}{4\alpha - \min(\alpha, \beta)} - \frac{\ln[\max((\beta - r_2) \ln n, 2)]}{2 \ln n} & \text{si } r = 0^+, \\ \frac{1-2r_0}{4\alpha - \min(\alpha, \beta)(2-r)} & \text{si } 0 < r < 2, \\ \frac{1-2r_0}{4\alpha} - \frac{\ln \ln n}{2\alpha \ln n} & \text{si } r = 2, \\ \frac{1-2r_0}{2\alpha r} & \text{si } r > 2, \end{cases}$$

Remarque 5 Le théorème est obtenue par un argument de troncature. Ainsi, la constante de troncature b_n dans la condition **C4** peut être déterminée par des conditions sur les moments. Ceci est illustré dans l'exemple que nous allons traiter. Dans un cas pareil,

les constantes n'affectent pas la vitesse finale. S'il n'existe pas suffisamment de moments, alors κ dans la condition C4 peut ne pas satisfaire l'inégalité exigée. Dans ce cas, la vitesse obtenue est non optimale et elle dépend de κ .

Remarque 6 Pour plusieurs "sieves" en dimension infinie, on peut choisir $r_1 = r_2 = 0$ dans la condition C3.

Pour les problèmes de dimension finie, on a $\alpha = \beta = 1$, $r = 0^+$, $r_0 = 0$, $r_2 = \beta$ alors

$$n^{-\tau} = (1/\sqrt{2})n^{-1/2}.$$

Remarque 7 Si $K(\pi_n\theta_0, \theta_0) = O(n^{-2\alpha\beta})$, où τ est déterminé dans le théorème, alors $K(\pi_n\theta_0, \theta_0)$ n'influe pas sur le calcul de la vitesse.

Remarque 8 Si les constantes A_1 et A_2 dans les conditions C1 et C2 dépendent de n , la vitesse de convergence va être affectée. On peut modifier la démonstration pour obtenir le résultat dans ce cas. Par exemple, si $A_1 = c_1n^{-b_1}$ et $A_2 = c_2n^{b_2}$ pour des constantes positives $c_i, b_i, i = 1, 2$, on parvient aux résultats du théorème en remplaçant r_0 par $r_0 + b_1 + b_2 \max((2-r), 0)/4$ pour $r > 0^+$ et $r_0 + b_1 + b_2/2$ pour $r = 0^+$.

Exemple

Considérons le modèle:

$$Y_i = \theta(X_i) + \varepsilon_i; \quad i = 1, 2, \dots, n$$

Le paramètre à estimer est la fonction $\theta \in \Theta$ où Θ est un espace fonctionnel.

On suppose que X_i et ε_i sont indépendantes, ε_i sont iid. $E(\varepsilon_i) = 0$ et $E(\varepsilon_i^2) = \sigma^2$. On suppose de plus que les X_i sont uniformément distribuées sur $[0,1]$. On distingue trois cas d'espace paramétriques :

1^{er} cas:

$$\Theta = \left\{ \theta \in C^p(0,1]; \quad \|\theta^{(j)}\| < +\infty, j = 0, \dots, p; \quad \left| \theta^{(p)}(x_1) - \theta^{(p)}(x_2) \right| \leq L|x_1 - x_2|^m \right\}$$

où p est un entier, $p + m > 0$ et L une constante inconnue. Si on considère le critère des "moindres-carrés"

$$l(\theta, y) = -(y - \theta)^2$$

Dans ce cas l'estimateur des "moindres-carrés" n'est pas consistant car l'espace paramétrique Θ n'est pas compact. Considérons pour

(i) $p + m > 1/2$ la "sieve"

$$\Theta_n = \left\{ \theta \in \Theta, \quad \theta(x) = \alpha_0 + \sum_{j=1}^{r_n} (\alpha_j \cos(2\pi jx) + \beta_j \sin(2\pi jx)), \alpha_0^2 + \sum_{j=1}^{r_n} j^{2d} (\alpha_j^2 + \beta_j^2) \leq l_n^2 \right\}$$

où d est une constante arbitraire proche de $p + m$ telle que $p + m > d > 1/2$.

On suppose que $E|\varepsilon_i|^\gamma < +\infty$ pour $\gamma > 0$ suffisamment grand. Le choix de l_n, r_n, γ sera fait ultérieurement. Soit

$$\rho(\theta_1, \theta_2) = [E(\theta_1 - \theta_2)^2]^{1/2}$$

D'après la propriété d'approximation de cette "sieve", pour tout $\theta \in \Theta$, il existe $\pi_n \theta \in \Theta_n$, tel que

$$\rho(\pi_n \theta, \theta) \leq \sup_x |\pi_n(\theta)(x) - \theta(x)| \leq O\left(\frac{1}{r_n^{p+m}}\right)$$

Notons que

$$\begin{aligned} l(\theta_0, Y) - l(\theta, Y) &= 2(\theta_0 - \theta)\left(Y - \frac{\theta_0 + \theta}{2}\right) & (1.16) \\ E(l(\theta_0, Y) - l(\theta, Y)) &= E(\theta_0 - \theta)^2 \\ E(l(\theta_0, Y) - l(\theta, Y))^2 &= E\left[2(\theta_0 - \theta)\left(Y - \frac{\theta_0 + \theta}{2}\right)\right]^2 \end{aligned}$$

$$\begin{aligned} E(l(\theta_0, Y) - l(\theta, Y)) &= 4E\left[(\theta_0 - \theta)^2\left(\theta_0 - \varepsilon_i - \frac{\theta_0 + \theta}{2}\right)^2\right] \\ &= 4E\left[(\theta_0 - \theta)^2\left(-\varepsilon_i + \frac{\theta_0 - \theta}{2}\right)^2\right] \\ &= 4E\left[\frac{(\theta_0 - \theta)^2}{2} - \varepsilon_i(\theta_0 - \theta)\right]^2 \\ &= 4E\left[\frac{(\theta_0 - \theta)^4}{4} - E[(\theta_0 - \theta)^3 \varepsilon_i] + 4E(\varepsilon_i^2(\theta_0 - \theta)^2)\right] \\ &= E(\theta_0 - \theta)^4 + 4\sigma^2 E(\theta_0 - \theta)^2 \end{aligned}$$

car ε_i et x_i sont indépendantes. D'après (1.16) la condition **C1** est vérifiée pour $\alpha = 1$.

En utilisant l'inégalité d'interpolation [22]

$$\|\theta - \theta_0\|_\infty \leq \|\theta - \theta_0\|^{(2d-1)/2d} \left\| \theta^{(d)} - \theta_0^{(d)} \right\|_2^{1/2d} \quad (1.17)$$

$$\begin{aligned}
E(\theta - \theta_0)^4 &= E(\theta - \theta_0)^2(\theta - \theta_0)^2 \\
&\leq \sup(\theta - \theta_0)E(\theta - \theta_0)^2 \\
&\leq (\sup(\theta - \theta_0))^2 \rho^2(\theta, \theta_0) \\
&\leq \|\theta - \theta_0\|_\infty^2 \rho^2(\theta, \theta_0)
\end{aligned}$$

et d'après (1.17), ceci est majoré par

$$\|\theta - \theta_0\|_2^{(2d-1)/2d} \|\theta^{(d)} - \theta_0^{(d)}\|^{1/2d} \rho^2(\theta, \theta_0) \leq l_n^{2/2d} (\rho(\theta, \theta_0))^{2(1+(2d-1)/2d)}$$

La condition **C2** est vérifiée si on prend $A_2 = l_n^{2/2d} \varepsilon^{2(2d-1)/2d}$, pour $\varepsilon > 0$. soit $l_n = n^{2\Phi}$, où Φ sera déterminé ultérieurement.

$$\begin{aligned}
(E \sup_{B_\varepsilon(s)} [l(s, Y) - l(t, Y)]^2) &= E \sup_{B_\varepsilon} (s - t)^2 (2Y - (s + t))^2 \\
&\leq O(\varepsilon^2)
\end{aligned}$$

On calcule l'entropie "bracketing" métrique de L^2 utilisée pour la condition **C3**. On définit

$$B_\varepsilon(s) = \{t \in \Theta_n, \|t - s\|_\infty \leq \varepsilon, \rho(t, \pi_n \theta_0) \leq \delta\}$$

Soit $F_n^{(b, \delta)}$ les classes définies dans **C4**, et soit

$$B(\delta) = \{t \in \Theta_n, \rho(t, \pi_n \theta_0) \leq \delta\}$$

D'après le lemme 1 d'Ossiander (voir [22]) on a :

$$H_2^B(\varepsilon, F_n^{(b, \delta)}) \leq H(\varepsilon, B(\delta), \|\cdot\|_\infty)$$

Notons que pour

$$\begin{aligned}
t &= \alpha_0 + \sum_{j=1}^{r_n} (\alpha_j \cos(2\pi j x) + \beta_j \sin(2\pi j x)) \\
\|t\|_\infty &= |\alpha_0| + \sum_{j=1}^{r_n} |\alpha_j| + |\beta_j|
\end{aligned}$$

Par les lemmes 5 et 6 [22] on a

$$H(\varepsilon, B(\delta), \rho) \leq A_3 r_n \ln(\min(\delta, l_n n^{-\frac{d-1}{2}} / \varepsilon))$$

où $A_3 > 0$.

Si $p + m > (1 + \sqrt{5})/4 \simeq 0.809$, on choisit Φ et d tels que

$$2\Phi \leq (d - \frac{1}{2}) - (p + m) / [2(p + m) + 1]$$

Prenons $r_n = n^{2\tau}$. alors

$$\begin{aligned} H_2^B(\varepsilon, F_n^{(b, \delta)}) &\leq H(\varepsilon, B(\delta), \rho) \\ &\leq A_3 n^{2\tau} \ln\left(\frac{\delta}{\varepsilon}\right). \end{aligned}$$

pour $\varepsilon < \delta$ et $A_3 > 0$. Donc **C3** est vérifiée pour $r_0 = r$, $r = 0^+$, $r_1 = 0$ et $r_2 = 1$.

On vérifie la condition **C4** avec

$$b_n = 3l_n^{(1/2d)+1} \text{ et } g(\delta) = \delta^{(2d-1)/2d}$$

$$\begin{aligned} &P\left(\sup_{\{\rho(\theta, \pi_n \theta_0) \leq \delta, \theta \in \Theta_n\}} l(\theta, Y) - l(\pi_n \theta_0, Y) \geq b_n \delta^{(2d-1)/2d}\right) \\ &\leq P\left(\sup_{\{\rho(\theta, \pi_n \theta_0) \leq \delta, \theta \in \Theta_n\}} (\pi_n \theta_0 - \theta)(2Y - (\theta + \pi_n \theta_0)) \geq b_n \delta^{(2d-1)/2d}\right) \\ &\leq P\left(\sup_{\{\rho(\theta, \pi_n \theta_0) \leq \delta, \theta \in \Theta_n\}} \|\pi_n \theta_0 - \theta\| [|2Y - 2\theta_0| + |\theta - \theta_0| + |\pi_n \theta_0 - \theta_0|] \geq b_n \delta^{(2d-1)/2d}\right) \\ &\leq P(\delta^{2d-1}/2d l_n^{1/2d} \sup_{\{\rho(\theta, \pi_n \theta_0) \leq \delta, \theta \in \Theta_n\}} [|2Y - 2\theta_0| + |\theta - \theta_0| + |\pi_n \theta_0 - \theta_0|] \geq b_n \delta^{(2d-1)/2d}) \\ &\leq P(|2Y - 2\theta_0| \geq 3l_n - \sup_{\{\rho(\theta, \pi_n \theta_0) \leq \delta, \theta \in \Theta_n\}} [|\theta - \theta_0| + |\pi_n \theta_0 - \theta_0|]) \\ &\leq P(|2Y - 2\theta_0| \geq 3l_n - 2l_n) \\ &\leq P(|Y - \theta_0| \geq \frac{l_n}{2}) \\ &\leq c \frac{E|\varepsilon_1|^\gamma}{l_n^\gamma} \end{aligned}$$

pour une constante $c > 0$, alors la condition **C4** est vérifiée pour $\gamma > 0$ suffisamment grand, par exemple $\gamma > \max(2, 1/2\Phi)$.

Notons que

$$\begin{aligned} K(\pi_n \theta_0, \theta_0) &= E(l(\theta_0, Y) - l(\pi_n \theta_0, Y)) \\ &= \frac{1}{2} E(\theta_0 - \pi_n \theta_0)^2 \\ &= O\left(\frac{1}{r_n^{2(p+m)}}\right) \end{aligned}$$

Il s'ensuit de la remarque 4 que la vitesse de convergence de l'estimateur de "sieve" est

$$\max(n^{-[1-2(\tau+\Phi)]/2}, n^{-2\tau(p+m)})$$

Pour optimiser cette vitesse, on choisit

$$\tau = \frac{1 - 2\Phi}{2(2(p+m) + 1)}$$

Donc la vitesse de l'estimateur sieve est

$$O_P(\varepsilon_n) \quad \text{avec} \quad \varepsilon_n = n^{-(1-2\Phi)(p+m)/[2(p+m)+1]}$$

Choisissons maintenant l_n qui satisfait à

$$l_n^{1/2d} \varepsilon_n^{2(2d-1)/2d} = O(1)$$

Mais

$$l_n < \min(n^{(2d-1)(p+m)/[(2d+1)(p+m)+1]}, n^{[(2d-1)(2(2(p+m)+1)-2(p+m))/2(2(p+m)+1)]})$$

Si on considère la "sieve"

$$\{\theta \in \Theta_m, \quad \rho(\theta, \theta_0) \leq \varepsilon_n\}$$

on peut vérifier qu'avec cette "sieve", la constante A_2 dans la condition **C2** peut être choisie indépendante de n . En appliquant le théorème, on obtient la vitesse

$$O_P(n^{-(p+m)/[2(p+m)+1]})$$

Dans l'application de ce théorème, la condition sur les moments est déterminée par

$$\tau > 1/2\Phi$$

Considérons le cas $\frac{1}{2} < p+m \leq (1+\sqrt{5})/4$.

Dans ce cas, on a une borne légèrement inférieure à l'entropie métrique

$$H_2^B(\varepsilon; F_n^{(b,\delta)}) \leq A_3 n^{2\tau} \ln(n/\varepsilon).$$

La vitesse de la "sieve" est

$$O_P(n^{-(p+m)/[2(p+m)+1]} (\ln n)^{(p+m)/[2(p+m)+1]})$$

avec

$$\tau = (1 + 0^+)/[2(2(p+m) + 1)]$$

(ii) Pour $p = 0$ et $m \leq \frac{1}{2}$.

On construit une "sieve" différente

$$\Theta_m = \left\{ \theta \in \Theta; \quad \theta(x) = \alpha_0 + \sum_{j=1}^{r_n} (\alpha_j \cos(2\pi jx) + \beta_j \sin(2\pi jx)), \quad \sup |\theta^{(j)}| \leq l_n \right\}$$

où d est une constante arbitrairement proche de m telle que $m > d > 0$.

L'optimisation sur cette "sieve" est plus compliquée que la précédente. La vitesse de convergence est

$$O_p(n^{-\frac{m}{2m+1}} (\ln n)^{-\frac{m}{2m+1}})$$

avec un argument similaire à celui présenté dans le cas précédent avec

$$r_n = n^{2\tau} = n^{\frac{2m}{2m+1}}, \quad l_n < n^{\frac{2dm}{((2d+2)(2m+1)+1)}}, \quad r_0 = \tau, \quad r_1 = 0, \quad r_2 = 1$$

$$\gamma > \frac{(2d+2)(2m+1)+1}{2dm}.$$

Pour obtenir les conditions du théorème, on applique l'inégalité d'interpolation suivante:

$$\|\theta - \theta_0\|_\infty \leq \|\theta - \theta_0\|_2^{2d/2d+1} \|\theta - \theta_0\|_\infty^{1/2d+1}$$

2^{ème} cas

L'espace paramétrique est similaire au cas précédent avec les bornes uniformes de $\|\theta^{(j)}\|_\infty$ pour $j = 0, \dots, p$ et la constante de Lipschitz sont connues.

On suppose que

$$E |\varepsilon_1|^\gamma < +\infty, \quad \text{où} \quad \begin{cases} \gamma > m+p, & \text{si } m+p > \frac{1}{2} \\ \gamma > 2, & \text{si } m+p = \frac{1}{2} \\ \gamma > m+p, & \text{si } m+p < \frac{1}{2} \end{cases}$$

On utilise les mêmes critères de fonction et distance qui mesure la divergence entre deux fonctions. L'optimisation est faite sur l'espace Θ . La condition **C1** est obtenue avec $\alpha = 1$. En appliquant l'inégalité d'interpolation

$$\sup |\theta - \theta_q| \leq c(\rho(\theta, \theta_0))^{\frac{2(m+p)}{2(m+p)+1}}$$

pour c constante, on a la condition **C2** avec $\beta = 1$ et $l = 0$. Suivant un résultat de [8], la condition **C3** est vérifiée avec

$$r_0 = r_1 = r_2 = 0 \text{ et } r = \frac{1}{p+m}.$$

Par analogie au cas précédent, si

$$b_n = n^\tau \text{ et } s = 1 - \frac{2(m+p)}{2(m+p)+1}$$

alors, on a la condition **C4**.

La vitesse de convergence de l'estimateur de régression est:

$$[E(\hat{\theta}_n(x) - \theta_0(x))^2]^{1/2} = \begin{cases} O_p(n^{-\frac{(m+p)}{2(m+p)+1}}); & \text{si } p+m > \frac{1}{2} \\ O_p(n^{-\frac{1}{4}}(\ln n)^{\frac{1}{2}}); & \text{si } p+m = \frac{1}{2} \\ O_p(n^{-\frac{p+m}{2}}); & \text{si } p+m < \frac{1}{2} \end{cases}$$

3^{ème} cas

Considérons une "sieve" construite à partir des approximations par les B-splines.

Soit

$$\Theta_m = \left\{ \theta \in \Theta, \quad \theta = \sum_{i=1}^{r_n+p+1} a_i \Phi_i, \quad \max_{i=1, \dots, r_n+p+1} |a_i| \leq l_n. \right\}$$

où $(\Phi_1, \dots, \Phi_{r_n+p+1})$ sont les B-splines d'ordre $p+m$ sur $[a, b]$, avec $[x_i, x_{i+p+1}]$ les supports de Φ_i et $(a = x_1, \dots, x_{r_n+p+1})$ est la répartition de $[a, b]$

L'approximation de l'erreur de cette "sieve" est de l'ordre de $O(r_n^{-(p+m)})$ qui est une conséquence d'un résultat de Schumaker (voir [22])

Notons que pour $\theta \in \Theta_n$ nous avons l'estimateur sieves

$$\|\theta\|_2^2 = \left[\frac{c}{(r_n + p + 1)^2} \right] \sum_{i=1}^n a_i^2; \quad \text{pour } c \text{ constante positive.}$$

et

$$\|\theta\|_\infty \leq c \max_{i=1, \dots, r_n+p+1} |a_i|; \quad \text{pour une constante } c.$$

en appliquant une technique similaire à celle du lemme 5 [22]. On suppose que les conditions **C1**, **C3**, et **C4** sont vérifiées et $\hat{\theta}_n$ satisfait à

$$L_n(\hat{\theta}_n) \geq \sup_{\theta \in \Theta_n} L_n(\theta) - \eta_n$$

avec $\eta_n = o(n^{-\omega})$, où

$$L_n(\theta) = (1/n) \sum_{i=1}^n \ln p(\theta, Y_i) = 1/n \sum_{i=1}^n l(\theta, Y_i)$$

$$\omega = \begin{cases} \frac{2(1-2r_0)}{2}, & \text{si } r = 0^+ \\ \frac{2(1-2r_0)}{2+r}, & \text{si } 0 < r < 2 \\ \frac{1-2r_0}{2} - \frac{\ln \ln n}{\ln n}, & \text{si } r = 2 \\ \frac{1-2r_0}{r}, & \text{si } r > 2 \end{cases}$$

il est supposé aussi que la condition **C2** est vérifiée pour $0^+ \leq r \leq 2$.

Alors:

$$\rho(\hat{\theta}_n, \theta_0) = O_p(\max(n^{-\tau}, \rho(\pi_n \theta_0, \theta_0), K^{1/2\alpha}(\pi_n \theta_0, \theta_0)))$$

où

$$K(\pi_n \theta_0, \theta_0) = E(l(\theta_0, Y) - l(\pi_n \theta_0, Y))$$

et

$$\tau = \begin{cases} \frac{1-2r_0}{4\alpha - \min(\alpha, \beta)} - \frac{\ln[\max((\beta-r_2)\ln n, 2)]}{2\ln n} & \text{si } r = 0^+, \\ \frac{1-2r_0}{4\alpha - \min(\alpha, \beta)(2-r)} & \text{si } 0 < r < 2, \\ \frac{1-2r_0}{4\alpha} - \frac{\ln \ln n}{2\alpha \ln n} & \text{si } r = 2, \\ \frac{1-2r_0}{2\alpha r} & \text{si } r > 2, \end{cases}$$

avec quelques modifications, on obtient

$$\begin{aligned} H_2^B(\cdot, F_n^{(b, \delta)}) &\leq H(\varepsilon, B(\delta), \|\cdot\|_\infty) \\ &\leq c(r_n + p + 1) \ln \frac{\delta}{\varepsilon}; \quad \text{pour } 0 < \varepsilon < \delta \text{ et } c > 0 \end{aligned}$$

Après quelques calculs, on arrive à la vitesse de convergence de l'estimateur sieve :

$$O_p\left(n^{-\frac{(m+p)}{2(m+p)+1}}\right).$$

Différentes bases de fonctions peuvent mener à la même vitesse optimale. Il est intéressant de noter que pour le cas $p + m \leq \frac{1}{2}$, l'estimateur basé sur l'approximation du 2^{ème} cas n'atteint pas la meilleure vitesse de convergence tandis que les estimateurs du 1^{er} et 3^{ème} cas l'atteignent.

Chapitre 2

ESTIMATEUR SIEVES D'UN PROCESSUS AUTOREGRESSIF HILBERTIEN

2.1 Introduction

Soit (H, B_H) un espace de Hilbert séparable muni de sa tribu Borélienne. Le produit scalaire de H est noté par $\langle \cdot, \cdot \rangle$.

On considère une suite $(\varepsilon_t, t \in Z)$ de variables aléatoires à valeurs dans H , définies sur l'espace probabilisé (Ω, \mathcal{A}, P) , indépendantes et de même loi et telles que

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

Nous dirons qu'une telle suite est un H -bruit blanc. Soit d'autre part, ρ un opérateur linéaire borné de H dans H tel que $\exists j_0 \geq 1 / \|\rho^{j_0}\|_L < 1$.

On dira que $(X_t, t \in Z)$ est un processus autorégressif à valeurs dans H , associé à ρ et à (ε_t) et d'ordre 1 si

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

L'estimateur du maximum de vraisemblance n'est pas consistant quand le paramètre à estimer ρ appartient à un espace de dimension infinie.

Nous appliquons la méthode des "sieves" pour estimer le paramètre ρ . Comme nous l'avons

introduit, la méthode des "sieves" consiste à maximiser la fonction de vraisemblance sur des sous espaces de l'espace paramétrique appelés "sieves" de dimension finie qui croit avec la taille de l'échantillon .

2.2 Définitions et notations

Soit Θ l'espace des opérateurs linéaires bornés de norme inférieure à 1, muni d'une métrique d , (H, \mathcal{B}) est l'espace de Hilbert muni d'une mesure λ σ - finie . Soit P_ρ la loi stationnaire induite par la variable aléatoire X_n satisfaisant (2.1) et P_0 la loi de ε_0 . La mesure P_0 sera la mesure de probabilité définie Sur (H, \mathcal{B}) par rapport à laquelle les dérivées de Radon Nikodym sont évaluées. Sous la condition de ρ est de Hilbert Schmidt et commute avec l'opérateur de covariance de la variable aléatoire X_0 , les mesures de probabilité P_ρ et P_0 sont équivalentes ([8], proposition 4), cette condition sera considérée par la suite. Le modèle (2.1) est identifiable dans le sens où la famille des loi de probabilité $\{P_\rho, \rho \in \Theta\}$ est telle que $P_\rho \neq P_\beta$ si $\rho \neq \beta$ où Θ est l'espace paramétrique.

On a les notations et les définitions suivantes:

(a) pour $\rho \in \Theta_m$

$$B_m(\rho, \varepsilon) = \{\beta \mid \beta \in \Theta_m; \quad d(\rho, \beta) < \varepsilon\}$$

(b) La densité de probabilité stationnaire de transition du processus (X_n) satisfaisant (2.1) est notée par

$$g(x, y, \rho) = P_\rho(dx \mid X_0 = y) / P_0(dx) \quad \text{où } x, y \in H \text{ et } \rho \in \Theta$$

on définit l'entropie conditionnelle par

$$H(\rho, \beta) = E_\rho \ln g(X_0, y, \beta) = \int \ln g(x, y, \beta) g(x, y, \rho) dP_\rho(x)$$

(c) Pour toute extension de fonction à valeurs réelles g sur $A \subset \Theta$, et tout $B \subseteq A$ on pose

$$g(B) = \sup_{\beta \in B} g(\beta)$$

(d) $L_n((x_1, \dots, x_n), \rho)$ est la fonction de vraisemblance conditionnelle évaluée au point $\omega = (x_1, \dots, x_n)$

(e) L'ensemble des estimateurs du maximum de vraisemblance sur Θ_m pour un échantillon de taille n est défini par

$$M_m^n = \left\{ \rho \in \Theta_m, \quad L_n(\omega, \rho) = \sup_{\beta \in \Theta_m} L_n(\omega, \beta) \right\}$$

L'ensemble du maximum de l'entropie conditionnelle est noté par

$$A_m = \{ \rho \in \Theta_m, \quad H(\rho_0, \rho) = H(\rho_0, \Theta_m) \}$$

où ρ_0 est la vraie valeur du paramètre.

(f) Quand $C_m \subseteq A$, la notation $C_m \rightarrow \rho$ signifie $\sup_{\beta \in C_m} d(\rho, \beta) \rightarrow 0$ quand m croît. par la suite, $m = m_n$ et m_n est suite croissante vers l'infini avec n .

2.3 Estimation Sieves de l'opérateur d'un ARH (1)

Notre but est d'obtenir une convergence presque sûre de l'estimateur sieves de l'opérateur ρ dans différentes situations. Dans un premier cas, nous considérons un opérateur à noyau et on montre que le maximum de vraisemblance est atteint pour une sieve particulière (2.6) et nous donnons une forme explicite de l'estimateur sieves dans le cas Gaussien. Dans le cas d'un opérateur de Hilbert-Schmidt, nous donnons une condition générale similaire à celle dans [12] pour l'existence et la convergence presque sûre de l'estimateur sieves. Les convergences sont établies dans la norme de $L^2([0, 1])$ et nous donnons un résultat qui fournit une vitesse de croissance de m_n la taille de la sieve Θ_{m_n} pour l'opérateur à noyau, cette vitesse est de forme polynomiale $m_n = O(n^{\frac{1}{3}-\eta})$, $\eta > 0$ au lieu de la forme logarithmique obtenue auparavant par ([7] chp8), [17], et [19] avec des formes spéciales de valeurs propres de l'opérateur de covariance.

2.3.1 Le paramètre ρ est un opérateur à noyau

Soit $(X_n)_n = (X_n(t), t \in [0, 1])_n$ un processus satisfaisant (2.1) dans l'espace des fonctions de carré intégrable $L^2([0, 1])$, et (ε_n) est un bruit blanc gaussien. La relation (2.1) s'écrit alors dans $L^2([0, 1])$ comme:

$$X_n(t) = \rho X_{n-1}(t) + \varepsilon_n(t), \quad t \in [0, 1] \quad (2.2)$$

où on suppose que l'opérateur ρ est un opérateur de Hilbert Schmidt à noyau de convolution défini par

$$\rho(f)(t) = \int_0^1 K(t-x)f(x)dx$$

avec $K \in L^2[0, 1]$, $\|K\|_2 < 1$, et K périodique de période 1. L'équation (2.2) devient

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

ou

$$X_n(t) = (K * X_{n-1})(t) + \varepsilon_n(t) \quad (2.3)$$

L'estimation de l'opérateur ρ passe par l'estimation de son noyau K . Dans ce cas, l'espace Θ sera l'espace des fonctions $L^2[0, 1]$. Nous considérons la base trigonométrique de sinus et cosinus dans $L^2[0, 1]$:

$$(\phi_0 = 1_{[0,1]}, \phi_{2k}(t) = \sqrt{2} \cos(2\pi kt), \phi_{2k+1}(t) = \sqrt{2} \sin((2k+1)\pi t), k \geq 1)$$

Nous notons respectivement, par

$$a_k(X_n), a_k(X_{n-1}), a_k(\varepsilon_n), a_k(K)$$

les coefficients de Fourier par rapport à la fonction cosinus et par

$$b_k(X_n), b_k(X_{n-1}), b_k(\varepsilon_n), b_k(K)$$

les coefficients de Fourier par rapport à la fonction sinus, des fonctions $X_n, X_{n-1}, \varepsilon_n, K$. Les variables aléatoires $a_k(\varepsilon_n)$ et $b_k(\varepsilon_n)$ sont gaussiennes indépendantes et supposées de même variance.

D'après (2.3), nous avons les relations suivantes:

$$\begin{aligned} a_k(X_n) &= (a_k(K)a_k(X_{n-1}) - b_k(K)b_k(X_{n-1}))/2 + a_k(\varepsilon_n) \\ b_k(X_n) &= (a_k(K)b_k(X_{n-1}) + b_k(K)a_k(X_{n-1}))/2 + b_k(\varepsilon_n) \end{aligned} \quad (2.4)$$

Pour $k = 0$, on a

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

Dans cette partie, nous considérons l'hypothèse suivante:

$H.:$ $K \in L^2([0, 1])$, $\|K\|_2 < 1$, et K paire périodique de période 1.

d'où, $b_k(K) = 0$

Par conséquent, en notant

$$x_{n,k} = a_k(X_n), x_{n-1,k} = a_k(X_{n-1}), \varepsilon_{n,k} = a_k(\varepsilon_n), c_k = a_k(K)$$

la première relation dans (2.4) mène à:

$$\begin{aligned} x_{n,0} &= c_0 x_{n-1,0} + \varepsilon_{n,0} \\ x_{n,k} &= \frac{1}{2} c_k x_{n-1,k} + \varepsilon_{n,k} \end{aligned} \quad (2.5)$$

on obtient deux processus autorégressifs réels de premier ordre. L'estimation du noyau K passe par l'estimation de ses coefficients de Fourier.

On choisit une sieve Θ_{m_n} définie par

$$\Theta_{m_n} = \left\{ K \in L^2 / K(t) = c_0 1_{[0,1]}(t) + \sum_{k=1}^{m_n} c_k \sqrt{2} \cos(2\pi kt), t \in [0,1], \sum_{k=1}^{m_n} k^2 c_k^2 \leq m_n \right\} \quad (2.6)$$

où K est périodique de période 1 et $m_n \rightarrow +\infty$ quand $n \rightarrow +\infty$.

Le résultat suivant donne l'estimateur sieves \hat{K} du noyau K pour des observations (X_1, \dots, X_n) satisfaisant (2.3).

Théorème 2.3.1 [18] *Sous l'hypothèse H , l'estimateur sieve \hat{K} du noyau K est donné par*

$$\hat{K}(t) = \hat{c}_0 1_{[0,1]}(t) + \sum_{k=1}^{m_n} \hat{c}_k \sqrt{2} \cos(2\pi kt), t \in [0,1]$$

où

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

où m_n une suite croissante vers l'infini et $\lambda > 0$ tel que $\sum_{k=0}^{m_n} k^2 \hat{c}_k^2 = m_n$.

Preuve

Puisque les variables aléatoires réelles (ε_n) sont gaussiennes et indépendantes de moyenne nulle et de variance σ_k^2 , La densité de transition stationnaire du processus (X_n) défini par (2.2) par rapport à la loi gaussienne P_0 de ε_0 est donnée par

$$\begin{aligned}
g(X_n, X_{n-1}, K) &= \exp\left(\sum_{k=0}^{+\infty} \frac{1}{2\sigma_k^2} [x_{n,k}^2 - (x_{n,k} - \frac{1}{2}c_k x_{n-1,k})^2]\right) \\
&= \exp\left(\sum_{k=0}^{+\infty} \frac{1}{2\sigma_k^2} [x_{n,k}c_k x_{n-1,k} - \frac{1}{4}c_k^2 x_{n-1,k}^2]\right)
\end{aligned}$$

la dernière série converge puisque les coefficients de Fourier sont dans l^2 .

D'où la vraisemblance conditionnelle $L(X_1, X_2, \dots, X_n; K)$ par rapport à P_0 la loi de ε_0 est donnée par

$$\begin{aligned}
L(X_1, X_2, \dots, X_n; K) &= \prod_{i=1}^n g(X_i, X_{i-1}, K) \\
&= \exp \sum_{i=1}^n \left(\sum_{k=1}^{+\infty} \frac{1}{2\sigma_k^2} [x_{i,k}c_k x_{i-1,k} - \frac{1}{4}c_k^2 x_{i-1,k}^2] \right) \quad (2.7)
\end{aligned}$$

On doit maximiser l'expression (2.7) sur la sieve $\{\Theta_{m_n}\}$ donnée par (2.6)

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

Soit λ le multiplicateur de Lagrange, nous aurons à résoudre le problème suivant:

$$\max_{\Theta_{m_n}} \sum_k \sum_{i=1}^n [x_{i,k}c_k x_{i-1,k} - \frac{1}{4}c_k^2 x_{i-1,k}^2] - \lambda k^2 c_k^2$$

En dérivant par rapport à c_k , $k > 0$, on obtient les formules annoncées au théorème. Pour $k = 0$, nous utilisons la relation

$$x_{n,0} = c_0 x_{n-1,0} + \varepsilon_{n,0}$$

et l'expression de l'estimateur \hat{c}_0 est obtenu par des calculs simples. ■

Remarque 1

(a) Si nous considérons les coefficients $b_k(K)$ non nuls, les equations (2.4) permettent le calcul du maximum de vraisemblance mais la détermination de l'estimateur du maximum de vraisemblance sous la contrainte de la sieve (2.6) pour (a_k) et (b_k) est plus

difficile à écrire. En effet, en utilisant la base sinus et cosinus pour estimer un noyau K , nous arrivons à une diagonalisation par bloque 2×2 .

(b) Si nous considérons une base réelle orthonormale (e_k) dans $L^2[0, 1]$, les coefficients de Fourier satisfont la relation suivante:

$$x_{n,k} = \sum_{i=1}^{+\infty} d_{k,i} x_{n-1,i} + \varepsilon_{n,k}$$

où $(d_{k,i})$ sont les coefficients de Fourier de la fonction $\int_0^1 K(\cdot - s)e_k(s)ds$ par rapport à la base (e_k) . Donc on choisit une sieve similaire à (2.6) mais avec la condition $\sum_{i,k=1}^{m_n} k^2 d_{k,i}^2 \leq m_n$, nous obtenons par suite un système de m_n^2 inconnus à résoudre, d'où le choix d'une base arbitraire ne permet pas la diagonalisation de l'opérateur à noyau ρ . D'autre part, nous pensons que le choix d'une base exponentielle dans $L^2[0, 1]$ permet sous certaines conditions sur les densités des parties réelles et imaginaires des variables aléatoires d'obtenir les coefficients de Fourier.

(c) La convergence presque sûre dans l'espace $(L^2[0, 1], dx)$ pour l'estimateur \hat{K} est donnée dans le théorème 2.3.3.

2.3.2 Le paramètre ρ est un opérateur de Hilbert-Schmidt

Rappelons qu'un opérateur linéaire ρ sur un espace de Hilbert H muni de la norme $\|\cdot\|$ et du produit scalaire $\langle \cdot, \cdot \rangle$ est de Hilbert-Schmidt si

$$\rho(x) = \sum_j \lambda_j \langle x, e_j \rangle f_j, \quad x \in H$$

où (e_j) et (f_j) sont des bases orthonormales dans H et (λ_j) une suite réelle telle que $\sum_j \lambda_j^2 < \infty$.

L'espace \mathcal{S} des opérateurs de Hilbert-Schmidt est un espace de Hilbert muni de la norme $\|\rho\|_{\mathcal{S}} = (\sum_j \lambda_j^2)^{\frac{1}{2}}$. Soit Θ l'espace des opérateurs de Hilbert-Schmidt, la norme $\|\cdot\|_{\mathcal{S}}$ induit une métrique d sur \mathcal{S} en posant $d(\rho, \beta) = \|\rho - \beta\|_{\mathcal{S}}$. Nous considérons la relation (2.1) avec (ε_n) un bruit blanc, ρ_0 est la vraie valeur du paramètre. Nous étudions la convergence presque sûre de l'estimateur sieves de l'opérateur ρ , nous suivons [12]. Nous gardons les notations du paragraphe précédent. Le théorème suivant établit une convergence presque sûre de l'estimateur sieves dans la norme $\|\cdot\|_{\mathcal{S}}$. Dans cette partie, le processus n'est pas nécessairement Gaussien.

Le théorème suivant donne la convergence p.s. de l'estimateur sieve $\hat{\rho}_n$ de ρ_0 .

Théorème 2.3.2 [18] *Supposons que la sieve Θ_{m_n} est choisie telle que:*

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

$$E_{\rho_0} \ln g(X_0, y, B_{m_n}(\rho, \varepsilon)) < \infty$$

(2) *$A_m \rightarrow \rho_0$ quand $m \rightarrow +\infty$ Si $m, n \rightarrow +\infty$ alors*

$$M_m^n \rightarrow \rho_0 \quad p.s \quad d'o\grave{u} \quad \sup_{\hat{\rho}_n \in M_m^n} \|\hat{\rho}_n - \rho_0\|_S \rightarrow 0 \quad p.s$$

Remarque 2

La condition (1) est utilisée pour obtenir des moments finis. Comme dans [12] nous pouvons avoir une condition suffisante pour la condition (2). En effet la condition $H(\rho_0, \rho_m) \rightarrow H(\rho_0, \rho_0)$ implique en général $\rho_{m_n} \rightarrow \rho_0$. Et si la sieve Θ_m est choisie telle que

$$\exists \rho_m \in \Theta_m \quad \text{et} \quad H(\rho_0, \rho_m) \rightarrow H(\rho_0, \rho_0), \quad \text{alors on a} \quad A_m \rightarrow \rho_0 \quad \text{quand} \quad n \rightarrow +\infty$$

Preuve

Rappelons que la densité de probabilité de transition du processus (X_n) est notée par $g(x, y, \rho)$. d'où la vraisemblance conditionnelle est

$$L(\omega, \rho) = L(X_1, \dots, X_n; \rho) = \prod_{i=1}^n g(X_i, X_{i-1}, \rho)$$

à cause de la continuité de $g(x, y, \cdot)$ sur l'espace métrique (Θ, d) , la vraisemblance $L(X_1, \dots, X_n; \cdot)$ et l'entropie $H(\rho_0, \cdot)$ sont continues sur Θ_m . alors de l'hypothèse (1): $E_{\rho_0} \ln g(X_0, y, B_{m_n}(\rho, \varepsilon)) < \infty$, les ensembles $M_{m_n}^n$ et A_m définis précédemment sont compacts et presque sûrement non vides.

Notons par $B_m(A_m, \frac{1}{m})$ un $\frac{1}{m}$ voisinage de A_m dans le sens de la distance d . Il suffit de montrer que $M_m^n \subset B_m(A_m, \frac{1}{m})$ presque sûrement pour n assez grand et par l'hypothèse (2) on a le théorème.

Puisque l'ensemble

$$C_m = \Theta_m - B_m(A_m, \frac{1}{m})$$

est compact, on peut trouver un nombre fini de boules $B_m(\rho_m^k, \varepsilon_m^k)$ qui recouvre C_m où $\rho_m^k \in C_m$, et $\varepsilon_m^k > 0$ pour $k = 1, \dots, l_m$.

Pour β_m fixé, $\beta_m \in A_m$, on a.

$$\begin{aligned} P(M_m^n \subset B_m(A_m, \frac{1}{m})) &\leq P \left\{ \sup_{\rho \in \Theta_m - B_m(A_m, \frac{1}{m})} L_n(\omega, \rho) \geq L_n(\omega, \beta_m) \right\} \\ &\leq P \left\{ \sup_{k=1, \dots, l_m} L_n(\omega, B_m(\rho_m^k, \varepsilon_m^k)) \geq L_n(\omega, \beta_m) \right\} \quad (2.8) \\ &\leq P \left\{ \sup_{k=1, \dots, l_m} \frac{L_n(\omega, B_m(\rho_m^k, \varepsilon_m^k))}{L_n(\omega, \beta_m)} \geq 1 \right\} \end{aligned}$$

Soit

$$\rho_m^k \in \Theta_m - B_m(A_m, \frac{1}{m}), \quad \text{et} \quad a_{\rho_m^k} > 0$$

tel que

$$H(\rho_0, \rho_m^k) - H(\rho_0, \Theta_m) < -a_{\rho_m^k}$$

Soit

$$A = E_{\rho_0}(\ln(g(X_i, X_{i-1}, B_m(\rho_m^k, \varepsilon_m^k)))) - E_{\rho_0}(\ln(g(X_i, X_{i-1}, \beta_m)))$$

De la définition de A_m et pour $\beta_m \in A_m$, on a

$$A = E_{\rho_0} \ln \frac{g(X_i, X_{i-1}, B_m(\rho_m^k, \varepsilon_m^k))}{g(X_i, X_{i-1}, \beta_m)}$$

et

$$E_{\rho_0} \ln \frac{g(X_i, X_{i-1}, B_m(\rho_m^k, \varepsilon_m^k))}{g(X_i, X_{i-1}, \beta_m)} < -a_{\rho_m^k} \quad (2.9)$$

pour ε_m^k suffisamment petit.

Mais

$$\begin{aligned} \ln L(\omega, B_m(\rho_m^k, \varepsilon_m^k)) - \ln L(\omega, \beta_m) &= \ln \sup_{\beta \in B_m(\rho_m^k, \varepsilon_m^k)} \prod_{i=1}^n g(X_i, X_{i-1}, \beta) - \ln \prod_{i=1}^n g(X_i, X_{i-1}, \beta_m) \\ &= B \end{aligned}$$

et

$$\begin{aligned} B &\leq \ln \prod_{i=1}^n g(X_i, X_{i-1}, B_m(\rho_m^k, \varepsilon_m^k)) - \ln g(X_i, X_{i-1}, \beta_m) \\ &= \sum_{i=1}^n \left(\ln \left(\frac{g(X_i, X_{i-1}, B_m(\rho_m^k, \varepsilon_m^k))}{g(X_i, X_{i-1}, \beta_m)} \right) \right) \end{aligned}$$

de plus, conditionné par X_{i-1} les lois des variables aléatoires $g(X_i, X_{i-1}, B_m(\rho_m^k, \varepsilon_m^k))$ (et $g(X_i, X_{i-1}, \beta_m)$) sont les images de g par translation des lois de ε_i qui sont i.i.d. Donc par la loi forte des grands nombres des variables aléatoires i.i.d et par (2.9) nous obtenons pour n assez grand.

$$\frac{1}{n} \sum_{i=1}^n \left\{ \ln \left(\frac{g(X_i, X_{i-1}, B_m(\rho_m^k, \varepsilon_m^k))}{g(X_i, X_{i-1}, \beta_m)} \right) \right\} \approx E_{\rho_0} \ln \left(\frac{g(X_i, X_{i-1}, B_m(\rho_m^k, \varepsilon_m^k))}{g(X_i, X_{i-1}, \beta_m)} \right) < -a_{\rho_m^k}$$

par conséquent, presque sûrement pour m grand:

$$\frac{L_n(\omega, B_m(\rho_m^k, \varepsilon_m^k))}{L_n(\omega, \beta_m)} < \exp(-na_{\rho_m^k})$$

En choisissant $a = \min_{k=1, \dots, l_m} a_{\rho_m^k}$, nous déduisons presque sûrement pour m grand

$$\sup_{k=1, \dots, l_m} \frac{L_n(\omega, B_m(\rho_m^k, \varepsilon_m^k))}{L_n(\omega, \beta_m)} < \exp(-na_{\rho_m^k})$$

Il existe N_m tel que pour tout $m \geq N_m$:

$$P \left\{ \sup_{k=1, \dots, l_m} \frac{L_n(\omega, B_m(\rho_m^k, \varepsilon_m^k))}{L_n(\omega, \beta_m)} \geq 1 \right\} \leq \frac{1}{m^2}$$

En fin, de (2.8) on a

$$P(M_m \subset B_m(A_m, \frac{1}{m})) \leq \frac{1}{m^2}$$

la démonstration s'achève par l'application du lemme de Borel-Cantelli ■

Nous considérons les notations suivantes avec les notations et les définitions considérées au paragraphe précédent.

C1 (a) Si (ρ_m) est une suite telle que

$$\forall n, \rho_m \in \Theta_m \text{ et } H(\rho_0, \rho_m) \rightarrow H(\rho_0, \rho_0) \quad \text{alors} \quad \rho_m \rightarrow \rho_0$$

(b) Il existe une suite $\rho_m \in \Theta_m$ telle que $H(\rho_0, \rho_m) \rightarrow H(\rho_0, \rho_0)$.

Pour tout $\delta > 0$ et tout m , nous définissons l'ensemble

$$D_m = \{\rho \in \Theta_m \mid H(\rho_0, \rho) \leq H(\rho_0, \rho_m) - \delta\}$$

où ρ_m est la suite dans C1 (b).

Etant donné l ensembles $\Gamma_1, \dots, \Gamma_l$ dans Θ_m , nous posons

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

Le corollaire suivant assure sous d'autres conditions la convergence presque sûre de l'estimateur sieve de l'opérateur ρ pour la norme $\|\cdot\|_S$.

Corollaire 2.3.1 [18] *Supposons que $\{\Theta_m\}$ est choisie telle que la condition C1 est vérifiée, et que pour tout $\delta > 0$, on peut trouver $\Gamma_1, \dots, \Gamma_l$ dans Θ_m , $n = 1, 2, \dots$ tels que*

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

$$(ii) \sum_{n=1}^{+\infty} l_{m_n} (\varphi_{m_n})^n < +\infty.$$

Alors

$$M_{m_n}^n \rightarrow \rho_0 \quad p.s. \quad \text{où} \quad \sup_{\hat{\rho}_n \in M_{m_n}^n} \|\hat{\rho}_n - \rho_0\|_S \rightarrow 0 \quad p.s.$$

Preuve

Fixons δ , on veut montrer que

$$P(D_{m_n} \cap M_{m_n}^n \neq \emptyset) = 0 \tag{2.10}$$

Si (2.10) est vérifié, alors avec une probabilité égale à 1

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

pour n suffisamment grand.

Puisque δ est arbitraire et

$$H(\rho_0, \rho_m) \rightarrow H(\rho_0, \rho_0)$$

Par la condition C1(b)

$$\liminf_{n \rightarrow +\infty} \inf_{\rho \in M_{m_n}^n} H(\rho_0, \rho) \geq H(\rho_0, \rho_0) \quad ps$$

Sachant que

$$H(\rho_0, \rho) \leq H(\rho_0, \rho_0)$$

$$\lim_{n \rightarrow +\infty} \sup_{\rho \in M_{m_n}^n} |H(\rho_0, \rho) - H(\rho_0, \rho_0)| = 0 \quad ps \quad (2.11)$$

et si on fixe $\varepsilon > 0$ et pour tout n on choisit $\beta_n \in M_{m_n}^n$ tel que

$$\frac{d(\rho_0, \beta_n)}{1 + d(\rho_0, \beta_n)} > \sup_{\rho \in M_{m_n}^n} \frac{d(\rho_0, \rho)}{1 + d(\rho_0, \rho)} - \varepsilon$$

la condition C1(a) combinée avec (2.11) impliquent

$$d(\rho_0, \beta_n) \rightarrow 0 \quad ps$$

Donc

$$\limsup \sup_{\rho \in M_{m_n}^n} \frac{d(\rho_0, \rho)}{1 + d(\rho_0, \rho)} \leq \varepsilon \quad ps$$

puisque ε est arbitraire, on a donc $M_{m_n}^n \rightarrow \rho_0$ ps.

Il nous reste à démontrer (2.10).

Fixons m, n . Alors

$$\begin{aligned} (D_{m_n} \cap M_{m_n}^n \neq \emptyset) &\subseteq \left\{ \sup_{\rho \in D_m} L_n((x_1, \dots, x_n), \rho) \geq L_n((x_1, \dots, x_n), \rho_{m_n}) \right\} \\ &\subseteq \bigcup_{k=1}^{l_m} \left\{ \sup_{\rho \in \Gamma_k^n} \prod_{i=1}^n g(X_i, X_{i-1}, \rho) \geq \prod_{i=1}^n g(X_i, X_{i-1}, \rho_{m_n}) \right\} \\ &\subseteq \bigcup_{k=1}^{l_m} \left\{ \left(\prod_{i=1}^n g(X_i, X_{i-1}, \Gamma_k^{m_n}) \right) \geq \prod_{i=1}^n g(X_i, X_{i-1}, \rho_{m_n}) \right\} \end{aligned}$$

notons

$$P(D_{m_n} \cap M_{m_n}^n \neq \emptyset) = \pi$$

On a

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

pour t_1, \dots, t_k arbitraire, positifs.

D'où

$$\pi \leq l_m (\varphi_m)^n.$$

Donc par le lemme de Borel Cantelli, on obtient (2.10). ■

Une application du Corollaire 2.3.1 donne la convergence presque sûre de l'estimateur sieves du noyau K_0 de ρ_0 considéré dans le Théorème 2.3.1 et fournit une forme polynomiale de la taille de croissance m_n de la sieve (2.6).

Théorème 2.3.3 [18] *Si $m_n \rightarrow +\infty$ telle que $m_n = O(n^{\frac{1}{3}-\varepsilon})$ pour $\varepsilon > 0$. Alors*

$$\|\hat{K} - K\|_2 \rightarrow 0 \text{ ps}$$

quand $n \rightarrow +\infty$ et $\|\cdot\|_2$ est la norme dans $(L^2([0, 1], dx))$.

Preuve

La démonstration consiste à vérifier les conditions du corollaire 2.3.1. Notons que les constantes dans la suite peuvent changer d'une ligne à autre. La sieve considérée Θ_m est donnée par (2.6) et l'existence de l'estimateur sieves \hat{K} associé est explicitement obtenu par le théorème 2.3.1. D'où l'ensemble M_{m_n} est non vide pour tout n . K_0 est Le noyau du vrai paramètre ρ_0 .

Vérifions la condition C 1 (a). De la définition de l'entropie conditionnelle H (voir le paragraphe précédent), la formule de la densité de transition (cf preuve du Théorème

2.3.1) et si nous notons par $c_k, c_{k,0}$ les coefficients de Fourier de K_{m_n} et K_0 respectivement, soit

$$C = H(K_0, K_{m_n}) - H(K_0, K_0)$$

$$\begin{aligned} C &= \sum_{k=0}^{+\infty} \frac{1}{\sigma_k^2} E \left[x_{n,k} c_{k,m_n} x_{n-1,k} - \frac{1}{2} c_{k,m_n}^2 x_{n-1,k}^2 - x_{n,k} c_{k,0} x_{n-1,k} + \frac{1}{2} c_{k,0}^2 x_{n-1,k}^2 \right] \\ &= \sum_k \frac{1}{2\sigma_k^2} E \left[E \left[\begin{array}{l} [x_{n,k}(c_k x_{n-1,k} - c_{k,0} x_{n-1,k}) + \\ \frac{1}{4}(c_{k,0}^2 x_{n-1,k}^2 + c_{k,m_n} x_{n-1,k})^2 + \\ \frac{1}{2} c_{k,m_n} x_{n-1,k} (c_{k,0} x_{n-1,k} - c_{k,m_n} x_{n-1,k})] \end{array} \right] / x_{n-1,k} \right] \\ &= \sum_k \frac{1}{8\sigma_k^2} E (c_{k,0} x_{n-1,k} - c_{k,m_n} x_{n-1,k})^2 \end{aligned} \quad (2.12)$$

Par conséquent, si

$$H(K_0, K_{m_n}) - H(K_0, K_0) \rightarrow 0, \text{ quand } n \rightarrow +\infty$$

alors

$$c_{k,m_n} \rightarrow c_{k,0} \text{ quand } n \rightarrow +\infty$$

D'où

$$K_{m_n} \rightarrow K_0 \text{ dans la norme de } L^2 \text{ quand } n \rightarrow +\infty.$$

Pour la condition C 1(b), nous utilisons la densité de $\bigcup \Theta_{m_n}$ dans l'espace des fonctions paires, périodiques de période 1 de L^2 pour déduire que

$$\exists K_{m_n} \in \Theta_{m_n} \text{ tel que } K_{m_n} \rightarrow K_0 \text{ la norme de } L^2$$

d'où $c_{k,m_n} \rightarrow c_{k,0}$, par suite nous obtenons $H(K_0, K_{m_n}) - H(K_0, K_0) \rightarrow 0$. d'où la condition C 1 (b) vérifiée.

Montrons maintenant que la condition (i) du corollaire 2.3.1 est vérifiée. soit $K \in \Theta_{m_n}$ la sieve définie dans (2.6), par la contrainte dans Θ_{m_n} , les coefficients de Fourier sont tels que

$$|c_k| \leq \frac{\sqrt{m_n}}{k} \quad \text{pour } k \neq 0, \quad \text{et } c_0 \leq \sqrt{m_n}.$$

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, où $[x]$ est le plus petit entier inférieur ou égal à x .

Soit I_k l'ensemble des points qui sont les extrémités de ces intervalles. Pour $k = 0$, I_0

s'obtient de la subdivision de l'intervalle $[-\sqrt{m_n}, \sqrt{m_n}]$ en $[m_n^2 + 1]$ intervalles de même longueur. Si l est la longueur de ces intervalles, où

$$l \leq \frac{2}{m_n^{3/2}}$$

Pour $\varepsilon > 0$, nous associons à chaque collection

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

un ensemble

$$\hat{\Gamma}^{m_n}(\{b_k\}) = \{K \in \Theta_{m_n}, \quad K(t) = \sum_{k=0}^m c_k e_k, \quad |c_k - b_k| \leq 2m^{-1.5}, \quad k = 0, 1, \dots, [m^{1+\varepsilon}]\}$$

Puisque $|c_k| \leq \frac{\sqrt{m_n}}{k}$, choisissons

$$c_k = -\frac{\sqrt{m_n}}{k} + \frac{m_n}{k^2} p, \quad p = 0, 1, \dots \quad (2.13)$$

Alors

$$\begin{aligned} |c_k| &\leq \frac{\sqrt{m_n}}{k} \implies -\frac{\sqrt{m_n}}{k} + \frac{m_n}{k^2} p \leq \frac{\sqrt{m_n}}{k} \\ \implies p &\leq \frac{2k\sqrt{m_n}}{m_n} \leq 2\left(\frac{m_n+1}{m_n}\right)m_n \\ \implies p &\leq cm_n \\ \implies p^{(1+m_n^{1+\varepsilon})} &\leq cm_n^{cm_n^{1+\varepsilon}} \end{aligned}$$

où c une constante positive. Le nombre de fonctions K dont les c_k sont de la forme (2.13)

est inférieur à $p^{(1+m_n^{1+\varepsilon})} \leq cm_n^{cm_n^{1+\varepsilon}}$. D'où $\bigcup_{k=1}^{l_{m_n}} \hat{\Gamma}_k^{m_n}$ recouvre Θ_{m_n} , où

$$l_{m_n} \leq (cm_n)^{cm_n^{1+\varepsilon}} \quad (2.14)$$

Nous définissons

$$\Gamma_k^m = \hat{\Gamma}_k^m \cap D_m, \quad k = 1, \dots, l_m$$

D'après ce qui précède

$$\bigcup_{k=1}^{l_{m_n}} \hat{\Gamma}_k^{m_n} \text{ recouvre } \Theta_{m_n}$$

d'où la condition (i) du corollaire 2.3.1.

Vérifions la condition (ii) du corollaire 2.3.1. Fixons $k \in \mathbb{N}$, et posons

$$\varphi(t) = E_{K_0} \left(\exp \left[t \ln \left\{ \frac{g(X_n, X_{n-1}, \Gamma_k^{m_n})}{g(X_n, X_{n-1}, K_{m_n})} \right\} \right] \right)$$

alors

$$\begin{aligned} \varphi(0) &= 1 \\ \varphi'(0) &= E_{K_0} \left(\left[\ln \left\{ \frac{g(X_n, X_{n-1}, \Gamma_k^{m_n})}{g(X_n, X_{n-1}, K_{m_n})} \right\} \right] \right) \end{aligned}$$

où

$$g(X_n, X_{n-1}, \Gamma_k^{m_n}) = \sup_{\beta \in \Gamma_k^{m_n}} g(X_n, X_{n-1}, \beta)$$

our K fixé dans $\Gamma_k^{m_n}$, soit

$$\begin{aligned} A &= E_{K_0} \ln g(X_n, X_{n-1}, \Gamma_k^{m_n}) - E \ln g(X_n, X_{n-1}, K) \\ A &= E_{K_0} \sup_{\beta \in \Gamma_k^{m_n}} \sum_{k=0}^{+\infty} \frac{1}{2\sigma_k^2} [(c_k x_{n-1,k} - b_k x_{n-1,k})(c_k x_{n-1,k} + b_k x_{n-1,k} - 2x_{n,k})] \end{aligned}$$

avec (b_k) les coefficients de Fourier par rapport à la fonction cosinus de la fonction $\beta \in \Gamma_k^{m_n}$

Or $\{\hat{\Gamma}_k^{m_n}\}_k$ recouvre Θ_{m_n} . D'où si $K, \beta \in \hat{\Gamma}_k^{m_n}$, alors

$$\begin{aligned} \sup_{\Gamma_k^{m_n}} |K(x) - \beta(x)| &\leq \sum_{k=0}^m |c_k - b_k| \\ &\leq \frac{2}{m^{1/2}} \end{aligned}$$

Donc

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

où c désigne une constante. Donc de la définition de D_{m_n} et de la condition (i) du corollaire 2.3.1, nous avons pour $\delta > 0$

$$\begin{aligned}
\dot{\varphi}(0) &\leq H(K_0, K) - H(K_0, K_{m_n}) + \frac{c}{m_n^{\varepsilon/2}} \\
&\leq cm_n^{-\varepsilon/2} - \delta
\end{aligned}$$

A l'aide d'un résultat de Hwang (cf [12]) on a

$$\varphi''(t) \leq cm_n^2 \quad \text{pour } t \in [0, 1/cm_n]$$

Donc par le développement de Taylor on a

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

Alors

$$\varphi\left(\frac{1}{m_n^2}\right) \leq 1 - \frac{\delta}{cm_n^2} \quad (2.15)$$

De la définition de φ_{m_n}

$$\varphi_m = \sup_k \inf_{t \geq 0} \varphi(t) = \sup_k \inf_{t \geq 0} E_{K_0} \left(\exp \left[t \ln \left\{ \frac{g(X_n, X_{n-1}, \Gamma_k^{m_n})}{g(X_n, X_{n-1}, K_{m_n})} \right\} \right] \right)$$

et des inégalités (2.14) et (2.15), nous pouvons déduire pour m_n suffisamment grand que

$$l_{m_n}(\varphi_{m_n})^n \leq cm_n^{cm_n^{1+\varepsilon}} \left(1 - \frac{\delta}{cm_n^2}\right)^n$$

et si $m_n = (n^{\frac{1}{3}-\varepsilon})$, la série $\sum_n l_m(\varphi_m)^n$ est convergente [12] d'où le résultat de Corollaire 2.3.1 que nous pouvons appliquer afin d'obtenir le résultat

$$M_{m_n}^n \rightarrow K_0 \text{ ps.}$$

et puisque $\hat{K} \in M_{m_n}^n$, et nous considérons la norme de L^2 sur la sieve Θ_{m_n} et sous la condition $m_n = O(n^{\frac{1}{3}-\varepsilon})$, nous arrivons à avoir le résultat du Théorème 2.3.3 ■

2.3.3 Prédiction

Le modèle (2.1) permet la prédiction d'un processus stochastique à temps continu $\xi = (\xi(t), t \in \mathbb{R})$ sur un intervalle entier comme cela a été expliqué en introduction. De la relation (2.1) il est facile de voir que le meilleur prédicteur probabliste de la variable aléatoire X_{n+1} en effet $\rho(X_n) = E(X_{n+1} | \sigma(X_1, \dots, X_n))$. Ainsi, le prédicteur naturel de X_{n+1} aura la forme $\hat{\rho}(X_n)$ où $\hat{\rho}$ est l'estimateur sieves de ρ . Les résultats de la convergence presque sûre de l'estimateur sieve assure la convergence du prédicteur en norme de $H = L^2([0, 1])$. Nous avons le théorème suivant:

Théorème 2.3.4 [18] (1) *Sous les conditions du théorème 2.3.2, nous avons quand $n \rightarrow +\infty$:*

$$\|\hat{\rho}(X_n) - \rho(X_n)\|_2 \rightarrow 0 \text{ en probabilité}$$

(2) *Si ρ est un opérateur à noyau et $m_n = O(n^{\frac{1}{3}-\varepsilon})$ pour $\eta > 0$, alors*

$$\|\hat{\rho}(X_n) - \rho(X_n)\|_2 \rightarrow 0 \text{ en probabilité quand } n \rightarrow +\infty$$

Preuve

(1) Pour $\alpha > 0$ et $\beta > 0$, l'inégalité de Markov et la finitude de $E \|X_0\|^2$ impliquent

$$\begin{aligned} P(\|\hat{\rho}(X_n) - \rho(X_n)\|_2 \geq \alpha) &\leq P(\|X_n\| \|\hat{\rho} - \rho\|_S \geq \alpha) \\ &\leq P\left(\|\hat{\rho} - \rho\|_S \geq \frac{\alpha}{\beta}\right) + P(\|X_n\| \geq \beta) \\ &\leq P\left(\|\hat{\rho} - \rho\|_S \geq \frac{\alpha}{\beta}\right) + \frac{E \|X_0\|^2}{\beta^2} \end{aligned}$$

Choisissant $\beta = \left(2E \|X_0\|^2 / \xi\right)^{1/2}$ nous déduisons que

$$P(\|\hat{\rho}(X_n) - \rho(X_n)\|_2 \geq \alpha) \leq P\left(\|\hat{\rho} - \rho\|_S \geq \frac{\alpha}{\beta}\right) + \frac{\xi}{2}$$

Mais de Théorème 2.3.2 on a

$$\|\hat{\rho} - \rho\|_S \rightarrow 0 \text{ p.s.}$$

Donc, la dernière probabilité est inférieure ou égale à ξ quand n est suffisamment grand. Puisque ξ est arbitraire nous arrivons au résultat.

(2) la démonstration est similaire à (1). ■

Remarque. Les derniers résultats de convergence sont établis dans la norme de $L^2([0, 1])$ sous la croissance polynomiale de la taille $m_n = O(n^{\frac{1}{3}-\epsilon})$ de la sieve (2.6) sans conditions de mélange sur le processus et seulement avec condition sur le moment d'ordre 2. Tandis que, dans [17] et [19] les résultats de convergence des estimateurs sont donnés sous la forme logarithmiques de croissance $m_n = O(\log(n))$, avec condition de mélange fort et des formes spéciales de valeurs propres de l'opérateur de covariance et condition sur le moment d'ordre quatre sur le processus. Dans ([7] page 221) la convergence presque sûre dans la norme des opérateurs est obtenue sans condition de mélange mais avec $m_n = O(\log(n))$ et avec le moment d'ordre quatre.

Chapitre 3

SIMULATIONS

3.1 Simulation de processus ARH(1)

3.1.1 Introduction

Pour simuler des trajectoires de processus AR fonctionnels et le calcul des prédicteurs *sieves*, nous utilisons le travail exécuté par [19] en utilisant le logiciel R version 2.4.1. et la bibliothèque *far* développée par J.Damons et S.Guillas [10]. Nous nous inspirons de l'exemple de simulation d'un processus *ARC*(1) [10] et [19] qui permet de simuler un bruit blanc dans l'espace $C[0, 1]$ en utilisant les fonctions propres

$$\left\{ e_j(t) = \sqrt{2} \sin[(j - 1/2) \pi t], j = 1, 2, \dots, h \right\}$$

de l'opérateur de covariance du mouvement brownien. Pour simuler notre processus et les prédicteurs *sieves* nous spécifions le modèle particulier. Plus précisément, nous considérons un opérateur ρ à noyau de convolution K paire périodique de norme L^2 inférieure à 1. Les exemples de fonction K considérés sont :

$$K(t) = \frac{1}{10^2} \cos(t), \quad K(t) = \frac{1}{10^3} t^2, \quad K(t) = \frac{1}{10^2} t^2 \cos(t)$$

Nous simulons $n + 1$ observations d'un processus *ARH*(1) qui seront nécessaires pour calculer les prédicteurs *sieves* et qui seront comparés à la $n + 1$ eme observation. Les résultats obtenus sont très satisfaisants et corroborent les résultats théoriques. Les erreurs quadratiques et les erreurs relatives RMAE de prédiction sont très faibles et de

l'ordre de quelques pourcents dans le cas des séries réelles considérées. Pour des applications réelles nous avons considéré la prévision du phénomène météorologique "El Nino". Pour sa prévision de l'année 2006 on utilise les observations mensuelles de la période 1950-2005. Pour pouvoir comparer avec les résultats de sa prévision durant l'année 1986 qui existent dans la littérature [2] nous avons aussi présenté sa prévision sieves de l'année 1986. De même, la prévision de la température en 1939 du Château de Nottingham sur la base de données mensuelles de 1920 à 1938 et la prévision de la température moyenne en 2004 à Alger basé sur la période 1970-2003 sont présentés. En général les résultats sur les erreurs quadratiques et les erreurs relatives RMAE de prévision sont très faibles comparativement aux erreurs quadratiques et relatives RMAE obtenues par d'autres méthodes de prévision.

3.1.2 Définitions

- Simulation de bruit blanc

Pumo [18] utilise la décomposition de Karhunen Loève du mouvement brownien pour simuler les trajectoires d'un bruit blanc sur $[0,1]$. Le bruit blanc est défini comme suit:

$$\varepsilon_t(s) = W_{t+s} - W_t, \quad s \in [0, 1]$$

où

$$W_u = \sum_{j=1}^{\infty} \sqrt{2n+2} Y_j^* \frac{\sin \left[(j-1/2) \frac{\pi u}{n+1} \right]}{\pi (j-1/2)}, \quad u \in [0, n+1]$$

Les Y_j^* sont des variables aléatoires i.i.d $N(0, 1)$.

Pour la programmation, nous itérerons n fois un mouvement brownien sur l'intervalle $[0,1]$ par la formule suivante:

$$W_u = \sum_{j=1}^{\infty} \sqrt{2} Y_j^* \frac{\sin [(j-1/2) \pi u]}{\pi (j-1/2)}$$

Pour la simulation, les sommes infinies sont approximées par des sommes finies.

- L'opérateur ρ

Nous considérons le cas d'un opérateur ρ à noyau K paire, périodique de période 1 de norme L^2 inférieure à 1 défini par :

$$\rho(f)(t) = \int_0^1 K(t-s)f(s)ds$$

Nous prenons les noyaux suivants : $K(t) = \frac{1}{10^2} \cos(t)$, $K(t) = \frac{1}{10^3} t^2$ et enfin $K(t) = \frac{1}{10^2} t^2 \cos(t)$

Pour le calcul de l'intégrale de convolution, nous utilisons la fonction "convolve" du logiciel R version 2.4.1.

- Valeur initiale de X_0

Pour la simulation d'un ARH(1) on prend comme valeur initiale la valeur $X_0 = 0$ et (ε_n) est un bruit blanc défini précédemment où les trajectoires sont calculées en $2m$ points de discrétisation (ici le nombre m est différent de m_n qui est la dimension de la sieves). Nous effectuons ensuite un changement de base pour ce bruit blanc en utilisant la fonction BaseK2BaseC de [10] pour avoir l'écriture du bruit blanc (ε_n) dans la base canonique et en posant $X_1 = \varepsilon_1$.

De la relation autorégressive (1.1) chap. 1 et en notant (X_i^j) $j = 1, \dots, m$ les coordonnées de X_i dans la base canonique $i = 1, \dots, n$, nous avons:

$$X_{i+1}^j = \text{convolve}(10^{-2}K, X_i^j) + \varepsilon_{i+1}^j \quad j = 1, \dots, m$$

Cette formule nous permet de générer un échantillon d'observations de taille n du processus.

3.1.3 Calcul des Prédicteur sieves

La fonction sieve.prédic

- Description

Nous utilisons les n observations du processus pour prédire la $(n+1)^{\text{ème}}$ trajectoire du processus .

A partir des observations $X_i(t_j)$ ($i = 1 : n, j = 1 : m$), on calcule le prédicteur *sieves* $\hat{X}_{n+1}(t_j)$, $j = 1 : m$, aux points discrets t_j .

$$\hat{X}_{n+1}(t_j) = \hat{\rho}(X_n)(t_j) = \int_0^1 \hat{K}(t_j - s) X_n(s) ds$$

La fonction "*sieve.prédic*" calcule le prédicteur *sieves* $\hat{X}_{n+1}(t_j)$, en utilisant la formule du Théorème 2.3.1 chap .2 sur les coefficients de Fourier associés à la fonction \hat{K} :

$$\hat{K}(t) = \sum_{k=1}^{m_n} \frac{\sum_{i=2}^n x_{i,k} x_{i-1,k}}{\sum_{i=1}^n \frac{1}{2} x_{i-1,k}^2 + 2n\lambda k} \sqrt{2} \cos 2\pi kt,$$

où m_n est la taille de la *sieve*. Pour les simulations on prend $m_n = 5$, et λ est le multiplicateur de Lagrange qui vérifie

$$\sum_{k=1}^{m_n} k^2 c_k^2 = m_n \quad (3.1)$$

Différentes valeurs de λ vérifiant (3.1) sont possibles. Nous allons considérer quelques valeurs de λ et nous vérifions empiriquement l'effet de ces dernières sur les erreurs quadratiques et les erreurs relatives RMAE sur des exemples de simulation.

Nous présentons ci dessous les simulations d'un processus ARH(1) sur des intervalles successifs et le prédicteur *sieves* de la dernière trajectoire du processus simulé.

La qualité d'un prédicteur \widehat{X}_{n+1} calculé aux points t_j , $j = 1, \dots, m$, pour prévoir X_{n+1} est mesurée par l'erreur quadratique et l'erreur relative absolue RMAE définie par

$$RMAE = \frac{1}{m} \sum_{i=1}^m \frac{|\widehat{X}_{n+1}(t_j) - X_{n+1}(t_j)|}{|X_{n+1}(t_j)|}$$

Pour le noyau nous avons pris $K(t) = \frac{1}{10^2} \cos(t)$. La figure (3.1) présente la simulation de $n=80$ trajectoires successifs d'un processus ARH(1), chaque trajectoire est calculé en $m=30$ points de discrétisation de l'intervalle $[0,1]$.

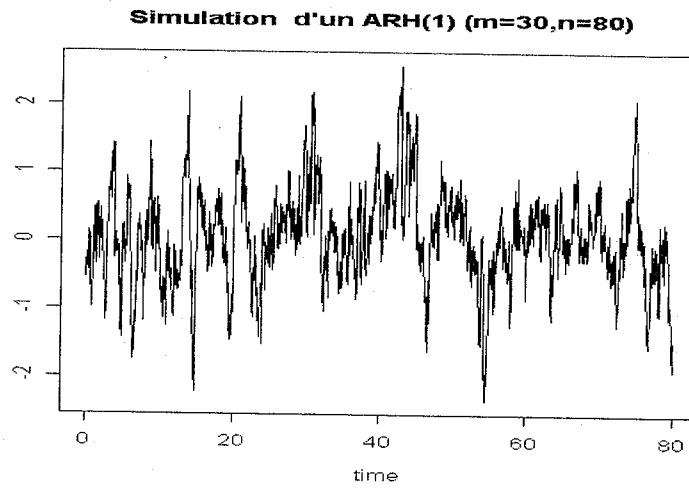


Figure 3.1:

Les figures (3.2), (3.3), (3.4) représentent le prédicteur sieves (trait pointillé rouge) de la 80^{ème} trajectoire (trait noir) pour $m=30, 25, 15$ points de dicrétisation et $\lambda = 5$.

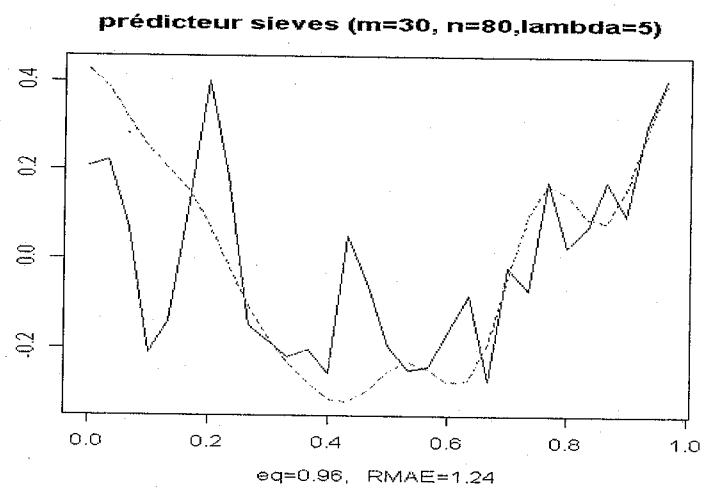


Figure 3.2:

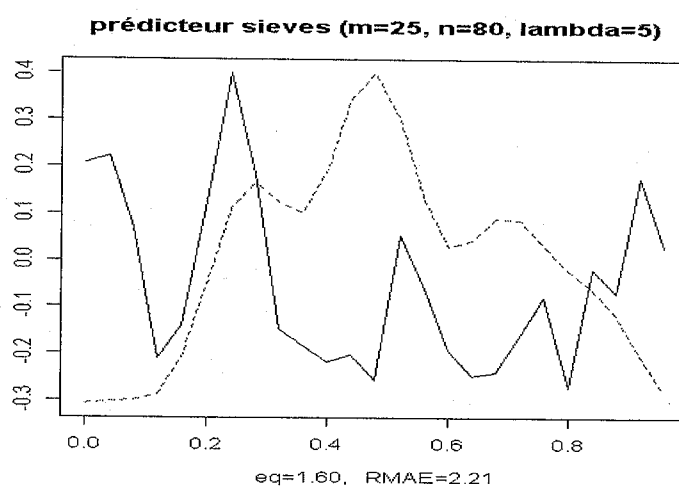


Figure 3.3:

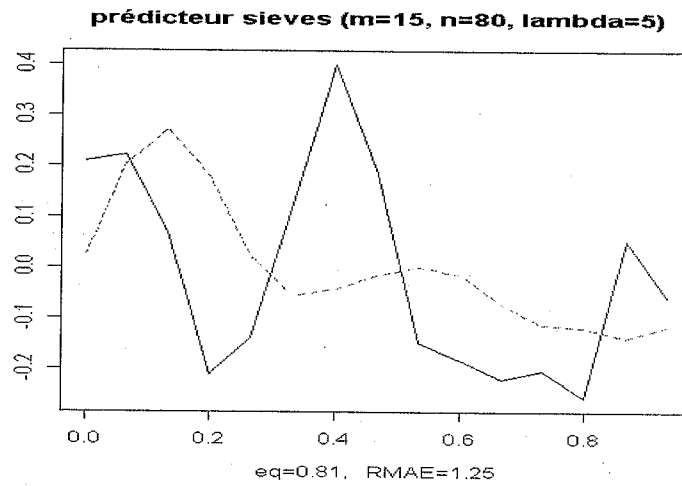


Figure 3.4:

Dans les exemples qui suivent le calcul du prédicteur sieves X_{50} est basé sur la simulation d'un processus fonctionnel avec m fixé et n fixé. $m=15$, $n=50$ et λ variable. Dans les graphes suivants nous présentons l'observation X_{50} et son prédicteur sieves \hat{X}_{50} , la valeur de l'erreur quadratique (eq) ainsi que l'erreur relative $RMAE$ pour les valeurs de $\lambda = 5, 7, 12, 15$.

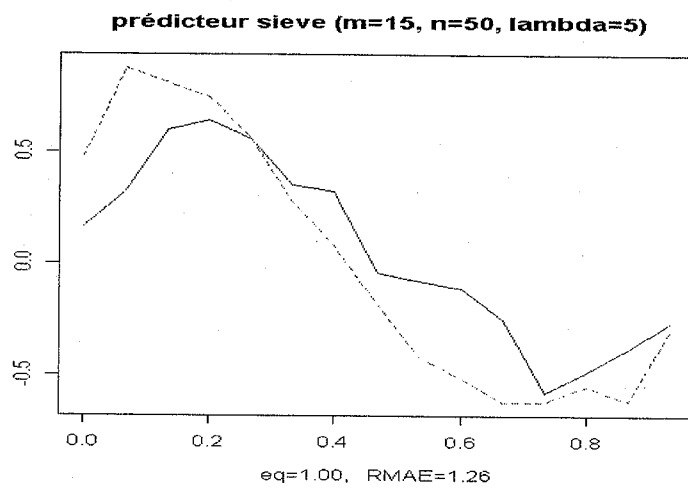


Figure 3.5:

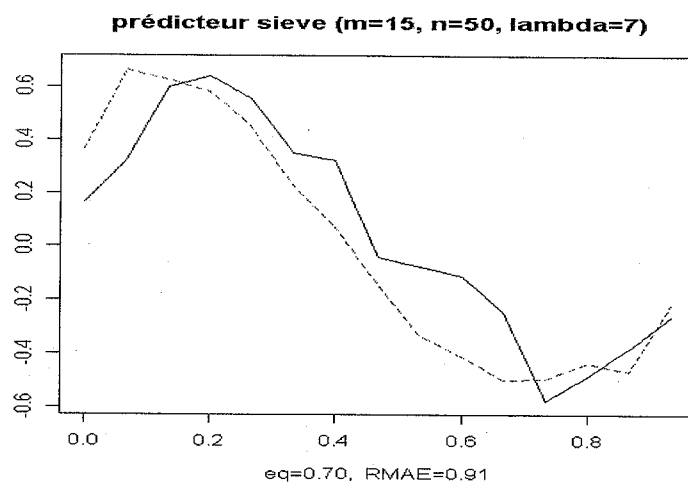


Figure 3.6:

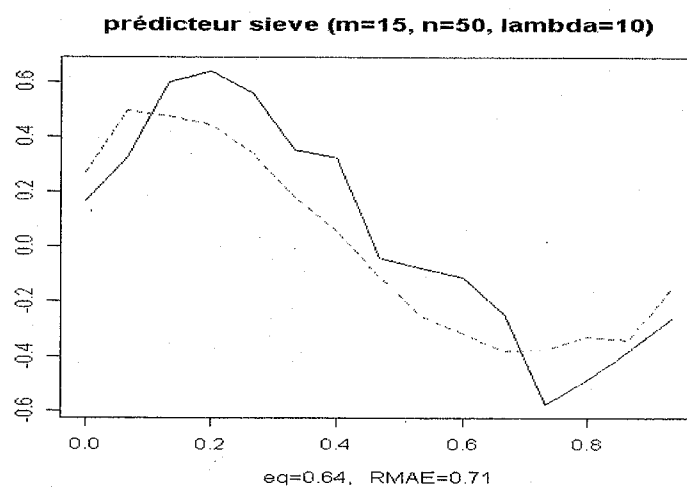


Figure 3.7:

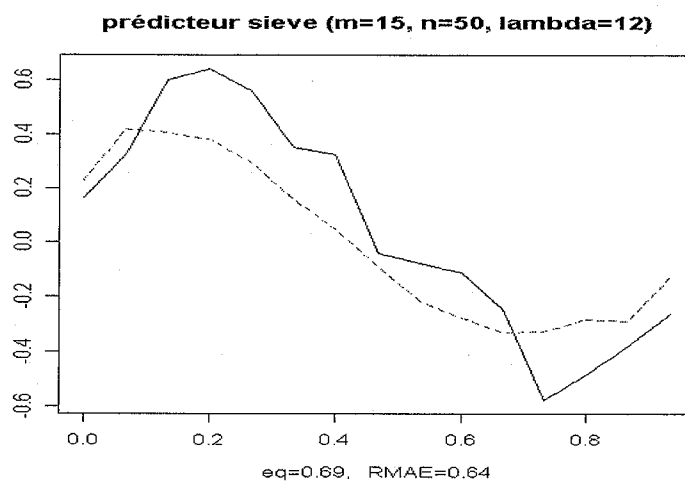


Figure 3.8:

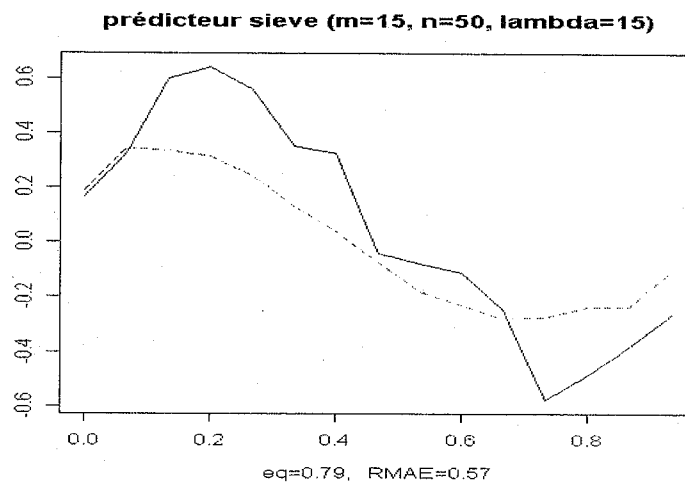


Figure 3.9:

Pour mesurer l'effet de discrétisation des observations m et le multiplicateur de lagrange λ sur le calcul du prédicteur sieves, nous présentons dans le Tableau 1 la moyenne des erreurs quadratiques pour 200 réplifications du processus avec $n = 120$ (fixé) et pour différentes valeurs de

$m = 10, 12, 15, 25, 30, 40$ (points de discrétisation) et $\lambda = 5, 7, 10, 15, 20, 25, 30, 35, 50$.

Tableau 1

$m \backslash \lambda$	5	7	10	15	20	25	30	35	50
10	2.22	2.08	2.13	2.10	1.99	2.17	2.03	2.10	2.09
12	2.41	2.29	2.25	2.13	2.34	2.18	2.15	2.31	2.26
15	2.86	2.69	2.59	2.52	2.68	2.54	2.58	2.50	2.52
25	5.46	4.72	4.34	3.83	3.37	3.71	3.36	3.26	3.18
30	7.43	7.42	6.16	4.79	4.16	4.26	4.16	4.03	3.88
40	14.93	12.17	9.60	8.28	7.48	6.68	6.09	6.31	5.04

Nous remarquons que la moyenne des erreurs quadratiques décroît avec λ , tandis qu'elle croît avec m .

Dans le Tableau 2, nous donnons la moyenne des erreurs quadratiques pour 200 réplifications du processus pour différentes valeurs de $n = 15, 35, 70, 120, 130$ pour $m = 25$ et $\lambda = 5$ fixés.

Tableau 2

n	15	35	70	120	130
EQ moyen	15.21	10.60	6.45	5.48	4.84

En général nous remarquons là aussi que l'erreur quadratique croit avec λ et m et n fixés (Tableau 1) et elle décroît avec n pour λ et m fixés. (Tableau 2).

Dans les exemples qui suivent nous considérons le noyau

$$K(t) = \frac{1}{10^3} t^2$$

La figure (3.10) représente la simulation d'un ARH(1) ($m=50$, $n=70$), les figures (3.11) et (3.12) et (3.13) représentent le prédicteur sieves pour $m = 50, 40, 30$ et $n=70$ fixé.

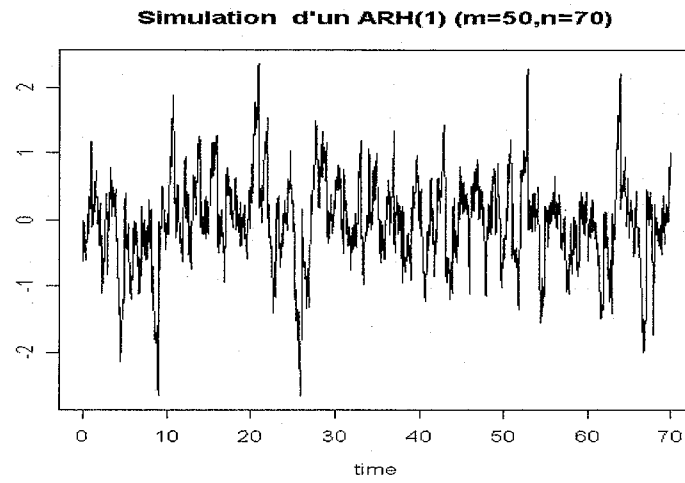


Figure 3.10:

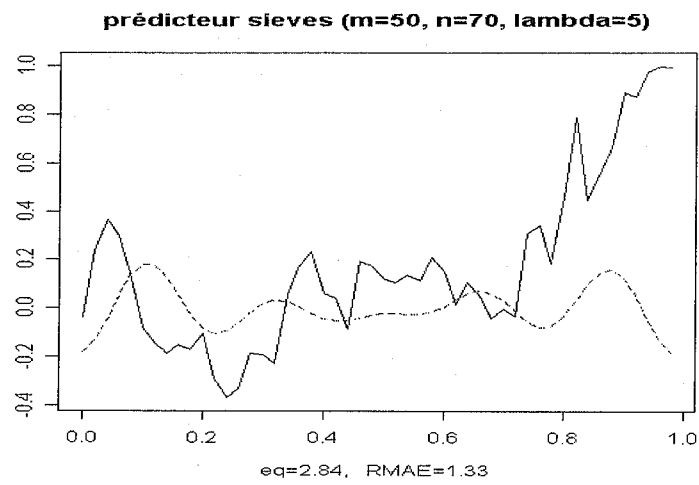


Figure 3.11:

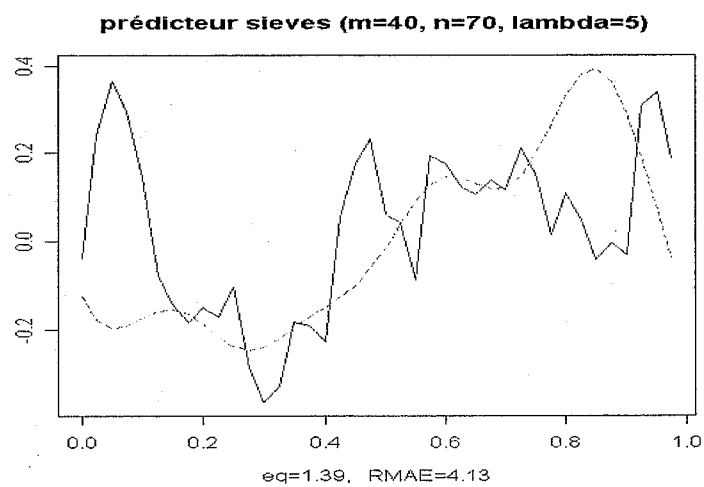


Figure 3.12:

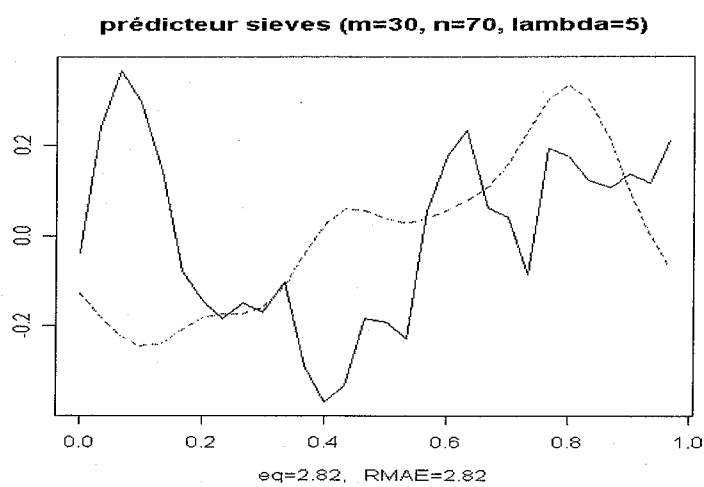


Figure 3.13:

Pour les simulations suivantes nous avons considéré le noyau

$$K(t) = \frac{1}{10^2} t^2 \cos t$$

La figure (3.14) représente la simulation d'un ARH(1) . Les figures (3.15), (3.16) et (3.17) présentent le prédicteurs sieves pour la dernière observation X_{60} , et $n=60, 25, 15$ et $\lambda = 5$ fixé.

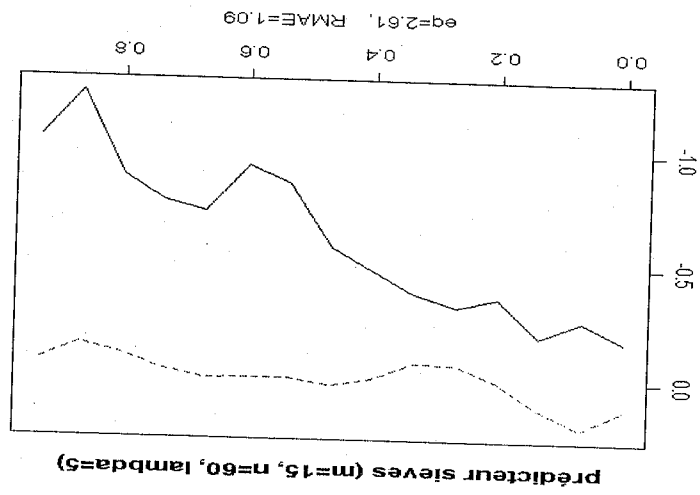


Figure 3.15:

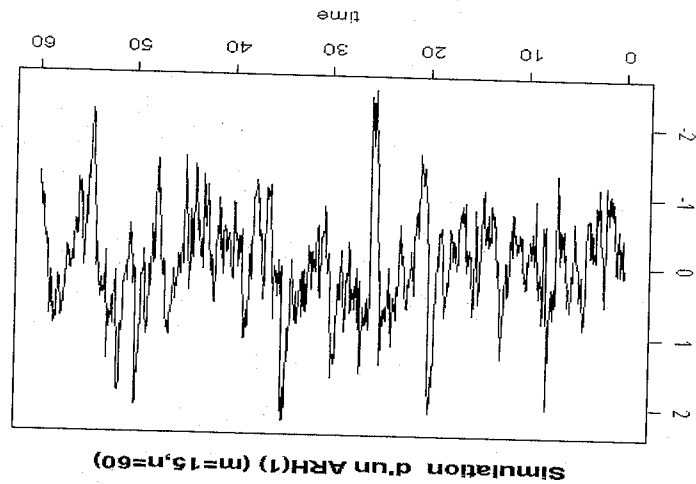


Figure 3.14:

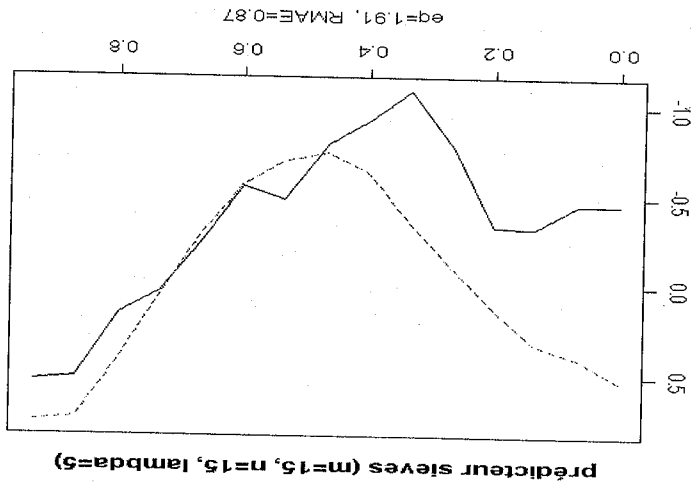


Figure 3.17:

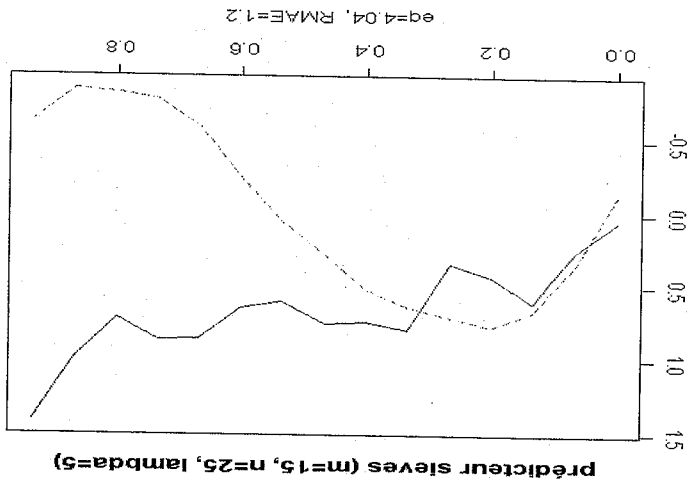


Figure 3.16:

3.1.4 Exemples

Nous utilisons les prédicteurs sives pour la prévision de séries réelles et nous comparons leurs performances avec d'autres méthodes de prédiction.

Série climatologique ENSO.

Nous appliquons la méthode de prédiction sives à la série chronologique décrivant le phénomène climatologique ENSO qui résulte des interactions entre l'atmosphère et l'océan au dessus de l'Océan Pacifique tropical. Le phénomène EL Nino (EN) est la composante océan dans ENSO tandis que Southern Oscillation (SO) est la composante atmosphère. Un index mesurant la variabilité d'EL Nino est fourni par les températures à la surface de l'océan ramenées à une moyenne observée dans le domaine Nino-3 (50S - 50N, 150W - 90W). Des valeurs moyennes mensuelles sont enregistrées depuis le mois de Janvier 1950 à Juin 2007 par des centres nationaux de la prévision environnementale aux Etats-Unis. La série chronologique de cet index montre des variations inter-annuelles marquées et superposées à une composante saisonnière forte. Cette série a été analysée par beaucoup d'auteurs (voir, par exemple, [2]; [14]). Pour pouvoir comparer avec leurs résultats nous utilisons des observations mensuelles durant la période 1950-1985 pour la prédiction de la température durant l'année 1986. Le tableau suivant donne les valeurs de l'erreur $RMAE$. La qualité d'un prédicteur \widehat{X}_{n+1} calculé aux points $t_i, i = 1, \dots, m$, pour prévoir X_{n+1} est mesurée par l'erreur relative absolue $RMAE$ définie par

$$RMAE = \frac{1}{m} \sum_{i=1}^m \frac{|X_{n+1}(t_i) - \widehat{X}_{n+1}(t_i)|}{|X_{n+1}(t_i)|}$$

associées aux différentes méthodes de prévision : Wavelet II et Wavelet III : [1], [2], ARH et ARHD : [15], FAR : [6] et le prédiction BLUP discret de [16]. Pour plus de détails sur ces résultats voir aussi [15].

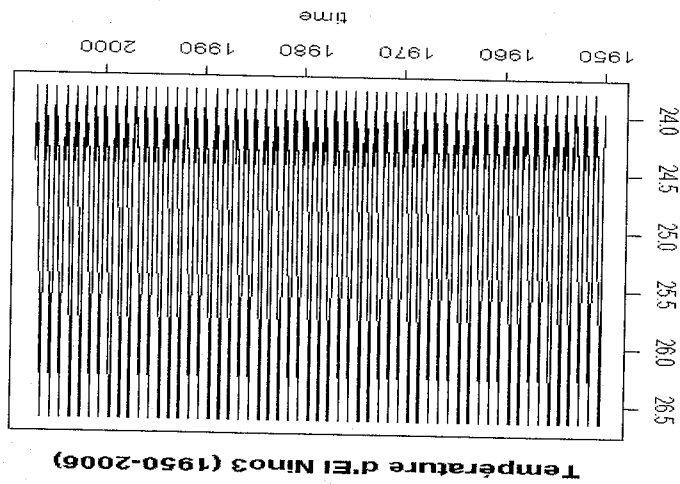
Tableau 3: Erreur *RMAE* de différentes méthodes de prévision de la température à la surface de l'océan de l'année 1986.

Méthodes de prévision	<i>RMAE</i>
Wavelet II	0.89%
Wavelet III	1.20%
ARHD $\alpha = 0.4, \beta = 0.8$	1.33%
ARHD $\alpha = 0.1, \beta = 0.4$	1.25%
FAR	0.89%
SARIMA	3.72%
ARH (I)	1.60%
BLUP discret	1.25%
Predicteur sieves	1.31%

Nous remarquons que le predicteur sieves donne une erreur *RMAE* assez faible. La Figure (3.18) représente la variation de la température (EL Nino3) de 1950 à 2006. La figures (3.19) représente la température à la surface (EL Nino3) de l'année 1986 et sa prévision par le predicteur sieves et l'erreur *RMAE*

Par la suite nous utilisons les observations (un intervalle = 1 année) de la période (1950-2005) pour la prévision de la température à la surface de l'océan de l'année 2006 (figure (3.20)). Dans [6] plusieurs méthodes de prévision ont été utilisées en particulier la méthode de prévision "Climathologie" qui est basée sur l'estimation du cycle moyen d'une année pour la prévision de l'année suivante.

Figure 3.18:



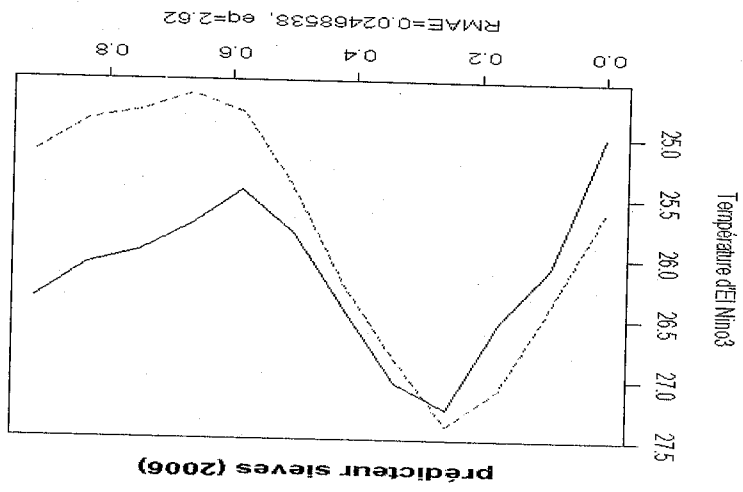


Figure 3.20:

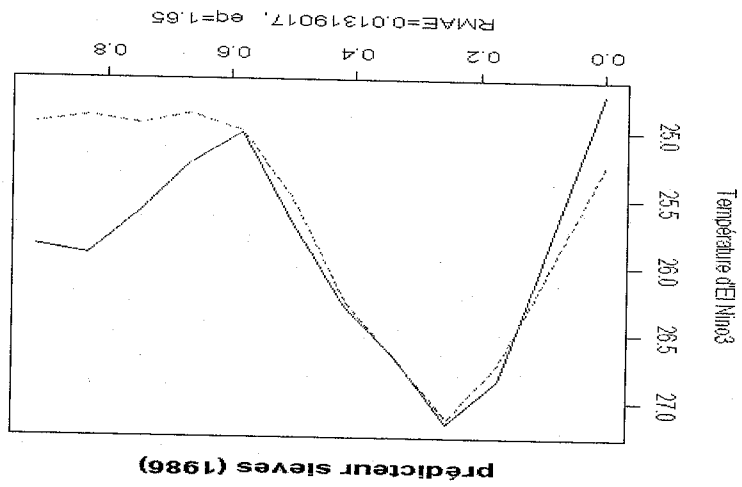


Figure 3.19:

Tableau 4: Erreur *RMAE* de différentes méthodes de prévision de la température à la surface de l'océan de l'année 2006

Méthodes de prévision	<i>RMAE</i>
Climatologie	2.5%
SARIMA	3.7%
Kernel	2.3%
Functional kernel	2.2%
Smooth FAR(1)	2.3%
Smooth FAR(1) with $q=p=12$	2.4%
Local FAR(1)	2.2%
BLUP discret	1.40%
Predicteur sieves	2.46%

Nous remarquons là aussi que le prédicteur sieves donne de bons résultats en erreur *RMAE* comparables aux autres méthodes sauf celle de BLUP discret.

Prévision de la température à Nottingham

Nous appliquons aussi la méthode de prédiction sieves sur la série historique des relevés de la température à Nottingham (température moyenne mensuelle au château de Nottingham de Janvier 1920 et arrêtée en 1939) disponible dans la série Nottem de la bibliothèque de MASSE de S-PLUS.

Tableau 5: Erreur *RMAE* de prévision de différentes méthodes de la température de l'année 1939 à Nottingham .

Méthodes de prévision	RMAE
Wavelet - Kernel	30 %
SS (Spline Smoothing)	28 %
SARIMA	31 %
BLUP cas continu	2.96%
Prédicteur sieves	2.95%

Nous remarquons que le prédicteur sieves donne une *RMAE* faible comparable à celle du BLUP continu de [16]. La Figure(3.21) représente la température à Nottingham de 1920-1939. Figure (3.22) représente la température de l'année 1939 à Nottingham et sa

prévision sieves.

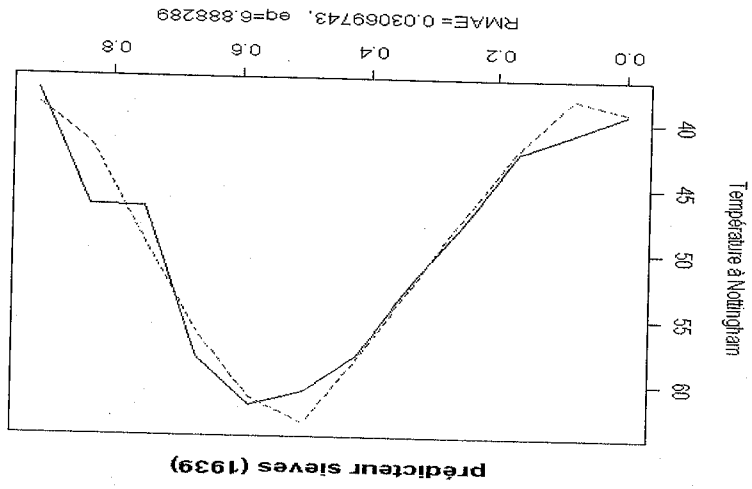


Figure 3.22:

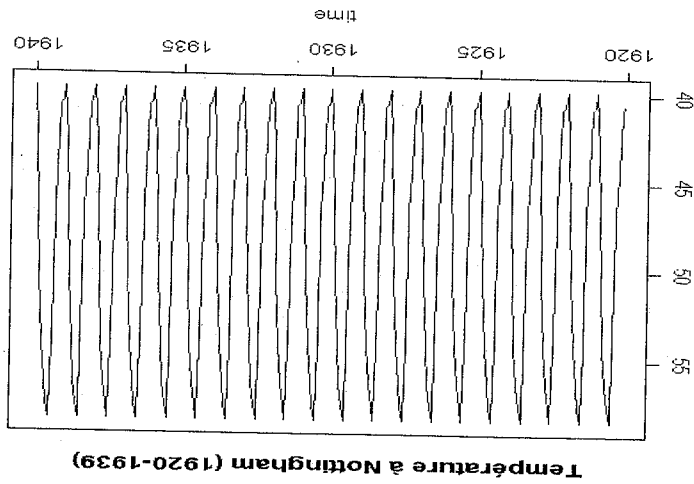


Figure 3.21:

Prévision de la température à Alger

Dans cette troisième partie nous présentons des résultats de prévision de la température moyenne à Alger pour l'année 2004. La Figure (3.23) représente la variation de la température à Alger durant la période 1970-2004. La Figure (3.24) représente la température moyenne de l'année 2004 à Alger et le prédicteur sieves avec les erreurs *RMAE*

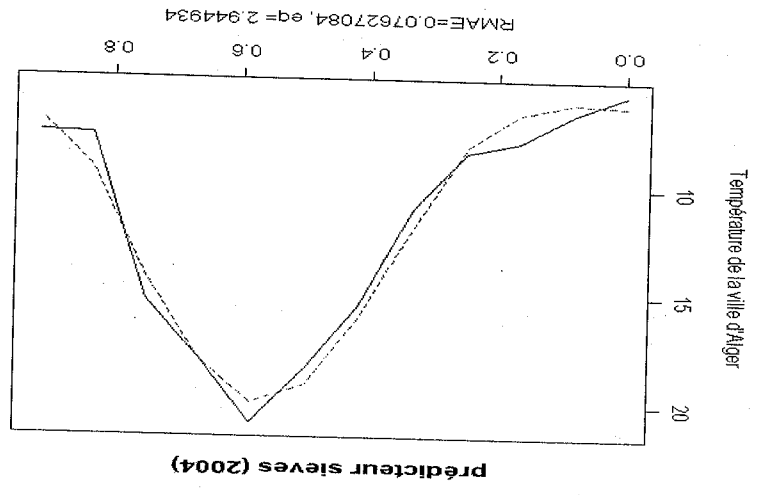


Figure 3.24:

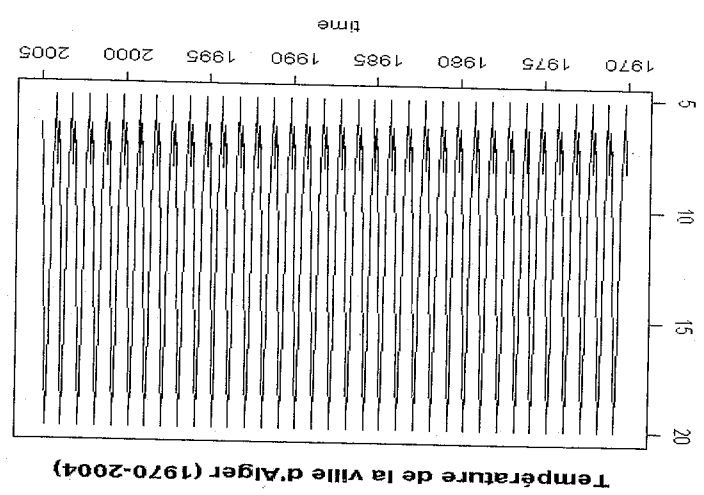


Figure 3.23:

Conclusion

Les simulations montrent un bon comportement asymptotique des prédicteurs sieves en fonction de la taille de l'échantillon et de certains paramètres définissant le pré-

dicteur.

Pour des valeurs de n pas élevées (à partir de $n=30$) nous avons obtenu des erreurs quadratiques et RMAE assez faibles cependant tous les résultats constatés empiriquement nécessitent une étude théorique par la suite. Pour les applications réelles : le phénomène El-Nino - température à Alger et température à Nottingham, nous avons comparé notre méthode de prévision sieves avec d'autres méthodes de prévision disponibles dans la littérature sur la base de leurs erreurs relatives moyenne RMAE. Les prédicteurs sieves donnent en général des erreurs RMAE assez faibles par rapport aux autres méthodes et comparables à celles obtenus dans [15]

Conclusion générale

Nous avons appliqué la méthode d'estimation sieves pour l'estimation du paramètre opérateur dans un processus autorégressif fonctionnel.

Dans le premier cas l'estimation du paramètre l'opérateur d'autorégression d'un processus autorégressif hilbertien $ARH(1)$ a été résolu dans le cas d'un opérateur à noyau de convolution paire périodique. Nous avons obtenu des résultats sur l'existence de l'estimateur sieves et leur convergence presque sûre et avec une forme explicite de l'estimateur sieves dans le cas Gaussien. Dans le cas où le paramètre est un opérateur de Hilbert Schmidt, sous des conditions générales similaires à [12] nous obtenons l'existence et la convergence presque sûre de l'estimateur sieves en norme de Hilbert-Schmidt. Nous donnons une vitesse de la dimension m_n de la sieve pour l'opérateur à noyau $m_n = O(n^{\frac{1}{2}-\eta})$, $\eta > 0$. Nous donnons enfin un résultat de convergence du prédicteur sieves pour la norme de $L^2[0, 1]$. Nous avons présenté des études de simulations d'un processus AR en utilisant le logiciel R version 2.4.1 en suivant l'exemple de simulation d'un processus $ARC(1)$ donnée par Julien Damons et Purno.

Les résultats obtenus corroborent les résultats théoriques et les erreurs quadratiques et erreurs relatives RMAE de prédiction sont très faibles. Des applications réelles ont été considérées telles que la prévision du phénomène météorologique "El Niño", la température au Château de Nottingham et à Alger avec des erreurs quadratiques et erreurs relatives de prévision très faibles comparativement à celles obtenues par d'autres méthodes de prévision. Les résultats de convergence des estimateurs et prédicteurs sieves ont été obtenus sous des hypothèses que le noyau de l'opérateur soit paire périodique. Nous pensons que ces résultats peuvent être généralisés à d'autres type de noyaux.

Le problème des vitesses de convergence des estimateurs et des prédicteurs ainsi que les lois limites n'ont pas été abordés dans ce travail et nous pensons y revenir plus tard.

Bibliographie

- [1] Antoniadis, A. A functional Wavelet-kernel approach for continuous-time Prediction.ccsd-00004891, version 1-10 May 2005.
- [2] Antoniadis, A., Sapatinas, T. Wavelet methods for continuous-time prediction using Hilbert-valued autoregressive processes. *J. Multivariate Anal.*, 87, 133-158. 2003.
- [3] Bahadur, P.R. Rates of convergence of estimates and test statistics. *Ann. Math. Statist.* 38, 303-324. 1967.
- [4] Bensmail, N., Mourid, T. Estimateur sieves de l'opérateur d'un processus ARH (1). *Comptes Rendus Acad. Sci. Paris t. 332 Serie I.* 2001.
- [5] Bensmail, N., Mourid, T. Prédiction des processus AR Hilbertien via la méthode des sieves. Simulations et exemples . soumis 2008
- [6] Besse, P., Cardot, H., Stephenson D.B.. Autoregressive Forecasting of Some Functional Climatic Variations. *Scandinavian Journal of Statistics* 27 (4) 673-687. 1997
- [7] Bosq, D. Linear Processes in Function spaces. Theory and Applications. Lecture Notes in Statistics 149 Springer. 2000.
- [8] Bosq, D.; Mourid, T. On the equivalence of the measure induced by the Banach valued Gaussian Autoregressive processes. *Stochastic. Anal. Appl.* 17(2), 137-144. 1997.
- [9] Chen, X., Shen, X. Sieve extremum estimates for weakly dependent data. *Econometrica* (66), 289-324. 1998.
- [10] Dammou, J., Gillias, S. Modelization for Functional Autoregressive processes. the far Package Version: 0.6-2 License: LGPL-2.1 version 2.4.1 du logiciel R). 2005.

- [11] Dunford, N., Schwartz, D. Linear Operators Part I. John Wiley, New York. 1962
- [12] Geman, S., Hwang, C.R.. Nonparametric maximum likelihood estimation by the method of sieves. Ann. Statist. 10 (2), 400-414. 1982.
- [13] Grenander, U.. Abstract Inference. Wiley, New York. 1981.
- [14] Guillas, S. Thèse de Doctorat En Sciences, Université Paris 6. 2001.
- [15] Mas, A., Pumo, B. The ARHD model. arXiv:math.ST/0.502285 v1 14 Feb 2005.
- [16] Mokhtari, F., Mourid, T. Prévission des processus AR(1) via le prédicteur de Parzan. Exemples et simulations. Les Annales de l'ISUP. LII, fasc. 1-2, 81-90. 2008.
- [17] Mourid, T. Thèse de Doctorat d'Etat en Sciences, Université Paris 6. 1995.
- [18] Mourid, T., Bensmain, N. Sieves estimator of the operator of a functional autoregressive process. Statistics and Probability Letters 76, 93-108. 2006.
- [19] Pumo, B. Thèse de Doctorat, Université Paris 6. 1993.
- [20] Pumo, B. Prediction of continuous time processes by $C(0,1)$ - valued autoregressive process. Statist. Inf. Stoch. Proc. I. 139-153. 1998.
- [21] Ramsay, J.O., Silverman, B.W. Functional Data Analysis. Springer, New York. 1997.
- [22] Shen, X., Wong, W.H. Convergence rate of sieves estimate Ann of Statist (22), 580-615. 1994.
- [23] White, H., Wooldrige, J.M. Some results on sieve estimation with dependant observations. University of California. Econometrica (59), 260-275. 1991.