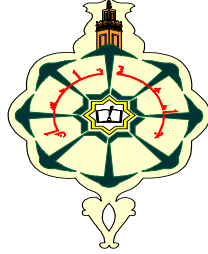


الجمهورية الجزائرية الديمقراطية الشعبية  
وزارة التعليم العالي و البحث العلمي

République Algérienne Démocratique et Populaire  
Ministère de l'Enseignement Supérieur et de la Recherche Scientifique

UNIVERSITE ABOUBEKR BELKAID  
TLEMCEN  
FACULTE DES SCIENCES  
Département de mathématiques



جامعة أبي بكر بلقايد  
تلمسان  
كلية العلوم  
قسم الرياضيات

# THESE DE DOCTORAT

## En Mathématiques

Présentée par

**Wafaa BENYELLES**

Titre de la thèse

**Problèmes d'estimation de la période de la représentation  
autorégressive.**

**Soutenu le : 16 Mai 2013**

**Membres du jury :**

**Président :** Mr. GHOUALI Nouredine

**Examineurs :** Mr. PUMO Besnik

Mr. DJEHICHE Boualem

Mr. LAKSACI Ali

Mr. DIB Hacem

**Invité:** Mr. LABBAS Ahmed

**Rapporteur :** Mr. MOURID Tahar

**Professeur UABBT**

**Professeur Université Angers**

**Professeur KTH Stockholm**

**Professeur Université Sidi Bel abbes**

**Professeur UABBT**

**MCA UABBT**

**Professeur UABBT**

**Titre : Problèmes d'estimation de la période de la représentation autorégressive**

**Résumé:**

Nous considérons un processus à temps continu admettant une représentation autorégressive AR(1) dans l'espace des fonctions continue  $C_{[0,\delta]}$  puis nous nous intéressons à l'estimation de la période de la moyenne du processus. Nous donnons des résultats de convergences de l'estimateur par la méthode de la distance minimale ainsi que sa loi limite. Ensuite, nous présentons des simulations numériques mettant en évidence l'efficacité des estimateurs sur des données simulées et sur des données météorologiques.

**Title: Problems of estimating of the period in autoregressive functional representation**

**Abstract:**

We consider a continuous time process with functional autoregressive representation and we focus on its mean period estimation. We give statistical results on an estimator of functional of the mean determining a minimum distance estimator of the period giving their consistency and limit law. We describe the practical performance of these estimators on numerical simulations and on real data analyzing the cycle of a climatological phenomena.

**العنوان مشاكل التقدير لدورية تمثيل انحدار ذاتي**

**ملخص**

في هذه الأطروحة نعتبر سيرورة ذات وقت مستمر تقبل تمثيل انحدار ذاتي في فضاء الدوال المستمرة على المجال  $[0, \delta]$  ثم نهتم بتقدير او تنبؤ دورية الامل الرياضي للسيرورة. نعطي نتائج المقدره من فعالية و تقارب بطريقة المسافة الأدنى. نقدم بعدها بعض المحاكاة العددية التي تعطي إثبات فعالية المقدرات على بيانات المحاكاة وبيانات الطقس.

# Dédicace

Je dédie ce travail à toutes les personnes chères à mon cœur ; mes oncles tantes, mon beau-père et spécialement ma mère pour l'aide précieuse qu'elle m'a apportée pendant tout mon cursus, ne ménageant ni son temps, ni ses conseils.

Il m'est agréable aussi d'exprimer une profonde gratitude envers feu mes grands-parents qui m'ont couvert de leurs bénédictions.

## **Remerciements**

Tout d'abord, je remercie le bon Dieu, tout puissant de m'avoir accordé la force et l'audace pour dépasser et vaincre toutes les difficultés.

Je ne remercierai pas assez, mon Directeur de thèse, Monsieur Tahar MOURID, pour m'avoir encadré, guidé, conseillé, m'avoir aussi suivi pas à pas dans l'évolution de mon travail, m'avoir également donné le goût de la recherche et l'avoir consolidé en usant de patience et de savoir, me prodiguant de judicieuses remarques et suggestions ainsi que pour l'intérêt soutenu et continu qu'il a manifesté à l'égard de cette recherche. J'en ai tiré un immense profit de ses connaissances et de son savoir-faire. C'est pourquoi, je lui exprime ma profonde reconnaissance et mon grand respect.

J'exprime ma profonde gratitude et mon grand respect aussi à Monsieur le Recteur, Noureddine GHOUALI, de m'avoir fait l'honneur, en acceptant la présidence du jury, sans oublier, d'énoncer ma vive reconnaissance pour l'intérêt et le temps imparti à la lecture de ce travail ainsi qu'aux remarques constructives émises en sa faveur.

J'ai été très honorée par l'acceptation des membres du jury, Messieurs les Professeurs : Monsieur Besnik PUMO, Monsieur Boualem DJEHICHE, Monsieur Ali LAKSACI, Monsieur Hacem DIB et Monsieur Ahmed LABBAS, à siéger à la commission d'examen. Etant mes examinateurs, je les remercie vivement, chacun par son nom, spécialement Monsieur Hacem DIB et Monsieur Ahmed LABBAS, pour l'intérêt qu'ils ont porté à ma recherche et les différentes observations édifiantes, émises de leur part sur le contenu et la forme afin que ce travail soit aussi rigoureux que possible.

Qu'il me soit enfin permis, de remercier tous mes professeurs, qui de près ou de loin, ont contribué à ma formation.

## Résumé

Cette thèse porte sur les problèmes d'estimation de la période de la représentation autorégressive, plus précisément, on s'intéresse à l'estimation de la période de la fonction moyenne d'un processus continu admettant une représentation autorégressive fonctionnelle AR(1). Nous rappelons qu'un processus réel  $Z = (Z(t), t \in \mathbb{R}^+)$  est un processus autorégressif fonctionnel dans un espace fonctionnel B si la suite des v.a.  $(X_n, n \in \mathbb{Z})$  définie par :

$$X_n(t) := Z(n\delta + t) \quad t \in [0, \delta], \text{ pour } \delta > 0$$

et vérifie la relation autorégressive suivante dans l'espace fonctionnel B :

$$X_n - a = \rho(X_{n-1} - a) + \varepsilon_n$$

où  $\rho$  est un opérateur linéaire borné défini sur B,  $a \in B$  et  $(\varepsilon_n)$  un bruit blanc dans B, v.a. i.i.d. du 2ème ordre vérifiant pour tout  $n \in \mathbb{Z}$ :

$$E(\varepsilon_n) = 0 ; \quad 0 < E\|\varepsilon_n\|^2 = \sigma_\varepsilon^2 < \infty$$

Sous une condition sur l'opérateur  $\rho$  la suite  $(X_n, n \in \mathbb{Z})$  est strictement stationnaire dans B et la moyenne  $m(t) = E(Z(t))$  du processus  $Z = (Z(t), t \in \mathbb{R}^+)$  est une fonction périodique de période  $\delta$  pour tout  $t \in [0, \delta]$ .

L'approche de cette estimation est basée sur l'estimation du paramètre  $\delta$  ; période de la représentation autorégressive ; par la méthode de la distance minimale proposée par Millar, W. tout en établissant sa convergence et sa loi limite, de plus afin de corroborer ses résultats, on termine par une série de simulations.

La thèse comporte trois chapitres.

Dans le premier chapitre, on rappelle des résultats généraux d'estimation de cette classe de processus introduit dans l'ouvrage de D.Bosq (2000).

Dans le deuxième chapitre, on aborde l'estimation du paramètre  $\delta$ , utilisant en premier lieu une méthode proposée par Tian pour les suites périodiquement corrélées, dont l'estimateur de la période  $\delta$  est défini par :

$$\delta_N = \text{Arg} \max_{1 \leq d \leq D} (\log V_N(d) - d \cdot \beta(N))$$

$$\text{où } V_N(d) = \frac{d}{G(N)} \sum_{s=1}^{\lfloor G(N)/d \rfloor} |R(sd)|, \quad R(\tau) = \frac{1}{N-\tau} \sum_{t=1}^{N-\tau} (Y(t) - \bar{z}_N)(Y(t+\tau) - \bar{z}_N),$$

$\beta(N) = N^{1/(2+b)}$ ,  $0 < b < 1/2$ ,  $G(N) = O(N)$ ,  $[G(N)]^{-1} = O(\beta(N))$  et  $D > 0$  est une borne supérieure de la période  $\delta$ .

Cependant, dans le cadre des processus autorégressifs fonctionnels nous verrons que cette approche ne s'applique pas. Puis nous proposons en second lieu un estimateur de la période basé sur la méthode de la distance minimale de Millar.W. Nous introduisons la quantité  $g(\delta)$  définie par :

$$g(\delta) = \int_0^{\delta} a(t) d\mu(t)$$

pour  $\delta > 0$  et nous étudions un estimateur défini par :

$$\hat{g}_T = \int_0^T Z(t) d\mu_{[T]}(t)$$

basé sur une observation de la trajectoire du processus  $(Z(t), t \in [0, T])$

où  $\mu_{[T]}$  est la mesure translatée de la mesure  $\mu$  sur  $[0, T]$

Nous définissons un estimateur de la période  $\delta$  en posant :

$$\hat{\delta}_T := \text{Arg} \min_{\delta \in ]0, \Delta[} |\hat{g}_T - g(\delta)|$$

où  $\Delta$  est un nombre strictement positif donné.

Nous obtenons les résultats asymptotiques l'estimateur  $\hat{g}_T$ , sa convergence et une loi limite, puis nous donnons la convergence et une loi limite de l'estimateur de la distance minimale de la distance minimal de  $\delta$  ;  $\hat{\delta}_T$ .

Le troisième chapitre, nous illustrons la performance de l'estimateur par des simulations numériques où nous utiliserons le logiciel R version 2.4.1, plus précisément, la Bibliothèque far du package R.

Ce travail est couronné par la publication de trois articles dans des revues scientifiques internationales (JAS, CRAS, ISUP) :

- Wafaa BENYELLES, Tahar MOURID « On a minimum distance estimate of the period in Functional Autoregressive Processes » Journal of Applied statistics 2012. DOI:10.1080/02664763.2012.668178.
- Wafaa BENYELLES, Tahar MOURID « Estimation de la période d'un processus à temps continu à représentation autorégressive » Annales ISUP Vol. 46 Fasc. 1-2 p. 89-101. 2002.
- Wafaa BENYELLES, Tahar MOURID « Estimation de la période d'un processus à temps continu à représentation autorégressive » Comptes Rendus de l'Académie des sciences (C.R.A.S). Paris, t.333, Série I, p.245-248, Juin 2001.

This article was downloaded by: [tahar mourid]

On: 08 April 2012, At: 13:46

Publisher: Taylor & Francis

Informa Ltd Registered in England and Wales Registered Number: 1072954 Registered office: Mortimer House, 37-41 Mortimer Street, London W1T 3JH, UK



## Journal of Applied Statistics

Publication details, including instructions for authors and subscription information:

<http://www.tandfonline.com/loi/cjas20>

### On a minimum distance estimate of the period in functional autoregressive processes

Wafaa Benyelles<sup>a</sup> & Tahar Mourid<sup>a</sup>

<sup>a</sup> Department of Mathematics, University of Abou Bekr Belkaid, Tlemcen, Algeria

Available online: 13 Mar 2012

To cite this article: Wafaa Benyelles & Tahar Mourid (2012): On a minimum distance estimate of the period in functional autoregressive processes, *Journal of Applied Statistics*, DOI:10.1080/02664763.2012.668178

To link to this article: <http://dx.doi.org/10.1080/02664763.2012.668178>



PLEASE SCROLL DOWN FOR ARTICLE

Full terms and conditions of use: <http://www.tandfonline.com/page/terms-and-conditions>

This article may be used for research, teaching, and private study purposes. Any substantial or systematic reproduction, redistribution, reselling, loan, sub-licensing, systematic supply, or distribution in any form to anyone is expressly forbidden.

The publisher does not give any warranty express or implied or make any representation that the contents will be complete or accurate or up to date. The accuracy of any instructions, formulae, and drug doses should be independently verified with primary sources. The publisher shall not be liable for any loss, actions, claims, proceedings, demand, or costs or damages whatsoever or howsoever caused arising directly or indirectly in connection with or arising out of the use of this material.

# On a minimum distance estimate of the period in functional autoregressive processes<sup>†</sup>

Wafaa Benyelles and Tahar Mourid\*

*Department of Mathematics, University of Abou Bekr Belkaid, Tlemcen, Algeria*

*(Received 3 March 2011; final version received 16 February 2012)*

We consider a continuous time random process with functional autoregressive representation. We state statistical results on a mean functional estimator determining a minimum distance estimator of the period giving consistency and a limit law stated in Mourid and Benyelles [13]. Then we discuss their performance on numerical simulations and on real data analyzing the cycle of a climatic phenomena.

**Keywords:** functional autoregressive process; functional mean estimation; period estimation; minimum distance estimation; functional data; simulation

## 1. Introduction

We deal with functional statistical models where recent progress has been made on developing statistical methodologies for functional data. In many fields such as economy, finance, industry, medicine, meteorology etc., the data are considered as ‘high dimensional’ r.v.’s. Their treatment becomes possible using technical computations and their analysis introduces new tools in probability theory and statistical inference for random processes. Functional statistical models are well adapted to such kind of data and have received a lot of attention, in recent literature. For a review on statistical analysis of functional data with various models we refer particularly to monographies by Bosq [3], Ramsay and Silverman [16,17], Ferraty and Vieu [8] (and the references therein) and Ferraty and Romain [7] for a handbook on statistics in infinite-dimensional spaces. They provide comprehensive discussions on these methods and applications and where among others matters, the prediction of a continuous-time random process plays a central role. Also we mention that various special issues are devoted by high level statistical journals to functional statistical models [4,9,10,18], see also the papers [6,8].

---

\*Corresponding author. Email: ta\_mourid@yahoo.fr

<sup>†</sup>Dedicated to Professor Denis Bosq.



In this paper, we deal with the estimation of the period of the mean in functional autoregressive processes (FAR) which are considered as a natural extension of the classical  $R^d$ -valued autoregressive processes in infinite-dimensional setting. The study of the class of Hilbert or Banach-valued autoregressive processes has been developed by many authors (see [1–3,15]).

First, let us introduce some known facts on this class of processes [3]. Let  $(\Omega, A, P)$  be a complete probability space and  $(B, \|\cdot\|)$  a real separable Banach space endowed with its Borel sets. We denote by  $L(B)$  the Banach algebra of bounded linear operators defined on  $B$  [5]. A Banach strong white noise is a sequence of independent and identically distributed  $B$ -valued random variables (r.v.'s)  $\varepsilon = (\varepsilon_n, n \in \mathbb{Z})$  defined on  $(\Omega, A, P)$  and such that, for all  $n \in \mathbb{Z}$ ,

$$E(\varepsilon_n) = 0; \quad 0 < E\|\varepsilon_n\|^2 = \sigma_\varepsilon^2 < \infty.$$

A sequence  $(X_n, n \in \mathbb{Z})$  of  $B$ -valued r.v.'s defined on  $(\Omega, A, P)$  is said to be a Banach-valued autoregressive process of order one (ARB(1)), if there exists an operator  $\rho$  in  $L(B)$ , a Banach white noise  $\varepsilon = (\varepsilon_n)$  and an element  $a \in B$  such that:

$$X_n - a = \rho(X_{n-1} - a) + \varepsilon_n. \quad (1)$$

Under the condition  $\|\rho^{j_0}\| < 1$ , for some  $j_0 \geq 1$ , Equation (1) has a unique strictly stationary solution in the Banach space  $B$  ([3, Chapter 3]).

A continuous time random process  $Z = (Z(t), t \in \mathbb{R})$  is said to be a FAR process in a function space  $B$ , if the sequence of functional r.v.'s  $(X_n, n \in \mathbb{Z})$  defined by

$$X_n(t) := Z(n\delta + t), \quad t \in [0, \delta] \quad (2)$$

satisfies Equation (1) for a bounded operator  $\rho$ , a  $B$ -white noise  $\varepsilon = (\varepsilon_n, n \in \mathbb{Z})$ , an element  $a \in B$ , and  $\delta > 0$ . The path  $X_n(t, \omega)$  represents the curve  $(Z(t, \omega), t \in \mathbb{R})$  upon the interval  $[n\delta, (n+1)\delta]$  (see [3, Chapter 3] for examples of such processes  $Z = (Z(t), t \in \mathbb{R})$  with such FAR representation). Then the mean  $m(t) = E(Z(t))$  of the process  $Z = (Z(t), t \in \mathbb{R})$  is a  $\delta$ -periodic function and we have  $m(t) = a(t)$  for all  $t \in [0, \delta]$ .

In many applications, the random process  $Z = (Z(t), t \in \mathbb{R})$  has continuous trajectories so natural function space is  $C_{[0,\delta]}$  of the continuous functions defined on  $[0, \delta]$  (the usual function space  $L^p_{[0,\delta]}$ ,  $p \geq 1$  may be used for such representation).

In all statistical problems on the model (1) the knowledge of the true period  $\delta$  is crucial for all parametric estimation and prediction. So when the period is unknown, its estimation arises before any estimation problem. As far as we know, there is no published work on the estimation of the period of this class of functional model. Let us point out that in the signal theory literature, there exists an approach to estimate the period of periodically correlated random processes. This method relies mainly upon the property that the mean and the covariance are periodic functions with the same period. However, in the model (1), the mean function is  $\delta$ -periodic but the covariance function is not. So this method does not work in our setting.

In this paper, we address these issues from a general formulation of a minimum distance estimator of the period  $\delta$  based on the estimation of a mean functional. For the methodology and a review on a minimum distance estimation one may see [11,12]. For convenience, we first present a statistical result established in a previous paper [13] with slight modifications of their proofs. We complete this study by showing the performance of the estimators on numerical simulations and real data examples. There exists a wide scope of applied fields for which our approach could be of interest. Our aim in future work is to extend these theoretical results of period estimation to the nonparametric setting of functional regression models and find examples in Biometrics, Genetics or Environmetrics.

The article is organized as follows. In Section 2, we present the main results on a mean functional estimator giving the consistency and a limit law. We then derive a minimum distance estimator

of the period establishing its convergence and a limit law. Section 3 presents numerical simulations and real data examples (climatic examples) showing the performance of these estimators. Technical proofs are deferred to the appendix.

## 2. Period minimum distance estimator

### 2.1 Mean functional Estimate

Let  $Z = (Z(t), t \in \mathbb{R})$  be a continuous time random process with FAR process in the sense of relation (2) in function space  $C_{[0,\delta]}$  with a  $\delta$ -periodic mean function  $a(t) = E(Z(t))$  and white noise  $\varepsilon$ . For a signed measure  $\mu$  defined on the Borelean class of the interval  $]0, \Delta[$  with a given  $\Delta > 0$ , we define the following functional  $g(\delta)$ ,  $\delta \in ]0, \Delta[$ , of the mean  $a$ :

$$g(\delta) := \frac{1}{\delta} \int_0^\delta a(t) d\mu(t). \quad (3)$$

First, we introduce an estimate of the functional  $g(\delta)$ . When a trajectory  $(Z(t), t \in [0, T])$  of the process  $Z$  is available, we define an estimator of  $g(\delta)$  by setting:

$$\hat{g}_T = \frac{1}{T} \int_0^T Z(t) d\mu_{[T]}(t), \quad (4)$$

where  $\mu_{[T]}$  is a extended measure of  $\mu$  on  $\mathbb{R}^+$  defined by

$$\mu_{[T]}(A) := \sum_{j=0}^{[T]-1} \mu(A \cap [j\delta, (j+1)\delta[-j\delta]), \quad (5)$$

where  $[T]$  is the integer part of  $T \in \mathbb{R}$  and  $A$  is a Borelean set of the interval  $[0, T]$ . The estimator  $\hat{g}_T$  was first introduced in [3] for  $\delta = 1$ .

The following result gives the almost sure and  $L^2$ -convergence of the estimator  $\hat{g}_T$ :

**THEOREM 1** *If the process  $Z = (Z(t), t \in \mathbb{R})$  is an FAR process with mean  $a$  and white noise  $\varepsilon$  in  $C_{[0,\delta]}$ , then we have*

$$\lim_{T \rightarrow \infty} \hat{g}_T = g(\delta)$$

*almost surely and in  $L^2$*

To establish a limit law of the estimator  $\hat{g}_T$  we recall some known facts on covariance operator  $R_{X_0}$  of a FAR sequence  $(X_n)$  defined by Equation (1) [3, Chapter 3]. The covariance operator  $R_{X_0}$  is a kernel symmetric operator on  $L^2_{[0,\delta]}$  where the kernel covariance  $C(s, t)$  and the eigenelements  $(\lambda_j, \Phi_j)_{j \geq 0}$  of  $R_{X_0}$  satisfy the equation:

$$\lambda_j \Phi_j(t) = \int_0^\delta C(s, t) \Phi_j(s) ds, \quad t \in [0, \delta], \quad j = 0, 1, 2, \dots \quad (6)$$

In this work, we make the following condition:

**H.** *The eigenelements of  $R_{X_0}$  are known,  $\lambda_j > 0$  and  $\|\Phi_j\|_{L^2} = 1$  for all  $j \geq 1$ .*

When the eigenelements of  $R_{X_0}$  are unknown, the corresponding eigenelements of the empirical covariance operator are used as estimators [3, Chapter 3] and a study is left for a future work.

The following theorem gives a central limit theorem when the measure  $\mu$  has a density  $\Phi_1$  with respect to the Lebesgue measure.

**THEOREM 2** *Suppose the condition **H** is satisfied and that the process  $Z = (Z(t); t \in \mathbb{R})$  is a FAR process with mean  $a$  and white noise  $\varepsilon$  in  $C_{[0,\delta]}$ . If the trajectories of the process  $Z$  satisfy the Lipschitz condition:*

$$|Z(t, \omega) - Z(s, \omega)| \leq M(\omega) \cdot |t - s| \quad \forall \omega \in \Omega$$

where the rv's  $M > 0$  and  $EM^2 < \infty$ , then we have

$$\frac{1}{\sqrt{T}} \sum_{i=0}^{T-1} \int_0^\delta (Z(i\delta + t) - a(t)) \Phi_1(t) dt \Longrightarrow_{T \rightarrow \infty} N_1,$$

where  $N_1$  is a Gaussian r.v. with zero mean and variance  $\lambda_1$ .

As a consequence, we have a limit law for the estimator  $\hat{g}_T$ .

**THEOREM 3** *Suppose the conditions of Theorem 1 are satisfied. Then we have*

$$\sqrt{T}(\hat{g}_T - g(\delta)) \Longrightarrow_{T \rightarrow \infty} N_2,$$

where  $N_2$  is a Gaussian r.v. with zero mean and variance  $\delta^{-2}\lambda_1$ .

## 2.2 Period estimation

Let  $Z = (Z(t); t \in \mathbb{R})$  be a FAR process in function space  $C_{[0,\delta]}$  with a periodic mean function  $a(t) = E(Z(t))$ . We denote by  $\delta_0$  the true value of the period of the mean  $a$ . With the estimator introduced in the preceding paragraph we propose an estimator of the unknown period  $\delta_0$  following the 'minimum distance method' [6,12,14]. For  $\Delta > 0$ , a minimum distance estimator of  $\delta_0$  is given by

$$\hat{\delta}_T = \text{Arg} \min_{\delta \in ]0, \Delta[} |\hat{g}_T - g(\delta)|, \quad (7)$$

where  $g(\delta)$  and  $\hat{g}_T$  are defined, respectively, by Equations (3) and (4)

Let  $\Theta := ]0, \Delta[$  be the parameter space, and for  $\beta > 0$  we define

$$h(\beta) = \inf_{\delta_0 \in \Theta} \inf_{|\delta - \delta_0| > \beta > 0} |g(\delta) - g(\delta_0)|. \quad (8)$$

The next result gives the uniform convergence in probability of  $\hat{\delta}_T$ .

**THEOREM 4** *Suppose the conditions of Theorem 1 are satisfied and that for all  $\beta > 0$ ,  $h(\beta) > 0$ . Then we have:*

$$P_{\delta_0}(|\hat{\delta}_T - \delta_0| > \beta) \longrightarrow_{T \rightarrow \infty} 0.$$

The convergence is uniform on compact intervals of  $\Theta$ .

*Remark 1* The condition  $h(\beta) > 0$  is an identifiability condition on the parameter. This condition is checked for example when  $\mu$  is the Lebesgue measure and the mean function 'a' keeps a constant sign.

The next result provides a limit law of the estimator  $\hat{\delta}_T$ .

**THEOREM 5** *Suppose the conditions of Theorem 2 are satisfied. Then we have:*

$$\sqrt{T}(\hat{\delta}_T - \delta_0) \implies_{T \rightarrow \infty} N_3,$$

where  $N_3$  is a Gaussian r.v.'s with zero mean and variance  $\sigma^2 = \delta_0^{-2} I^{-2}(\delta_0) \lambda_1$  and where  $\lambda_1$  is the first eigenvalues of  $R_{X_0}$  and  $I(\delta_0) = (1/\delta_0)(a(\delta_0) - g(\delta_0))$ .

*The convergence is uniform on compact intervals of  $\Theta$ .*

**Remark 2** The limit variance in Theorem 5 involves the first eigenvalue  $\lambda_1$  of  $R_{X_0}$ . Concerning its estimation we can proceed as follows: we have a relation between the eigenlements of the covariance operators  $R_{X_0}, R_{\varepsilon_0}$  and operator  $\rho$  in Equation (1). Indeed, let  $(X_n)$  be an AR(1) process with an operator  $\rho$  and white noise  $\varepsilon$ . Denote by  $(\beta_j, \Phi_j)$  the eigenlements of the covariance operator  $R_{\varepsilon_0}$ . If  $\rho$  is a symmetric Hilbert–Schmidt operator with eigenvalues  $(\mu_i)$  and commutes with  $R_{X_0}$ , then the eigenlements of  $R_{X_0}$  are given by  $(\lambda_j = \beta_j/(1 - \mu_j^2), \Phi_j)$  and  $\lambda_1 = \beta_1/(1 - \mu_1^2)$ . (see the Proof in the appendix).

Hence when the white noise  $\varepsilon$  is the Wiener process  $W = (W(t); t \in [0, \delta])$  we know that the eigenlements of the covariance operator  $R_W$  are given by:

$$\beta_j = \left(\frac{\pi}{\delta}\right)^{-2} \left(j - \frac{1}{2}\right)^{-2}; \quad \Phi_j = \sqrt{\frac{2}{\delta}} \sin \left[ \left(j - \frac{1}{2}\right) \frac{\pi t}{\delta} \right]; \quad j = 1, 2, \dots$$

so the relation  $\lambda_1 = \beta_1/(1 - \mu_1^2)$  gives a relation between an estimate of the first eigenvalues of  $R_{X_0}$  and  $\rho$ .

### 3. Numerical simulations and examples

#### Numerical simulations

To investigate numerical performance of the estimators  $\hat{g}_T$  and  $\hat{\delta}_T$  studied in this paper we have conducted a numerical simulations of the model (1) and studied two examples. To simulate the trajectories of the process  $(Z = (Z(t); t \geq 0))$  with FAR representation (1) in function space  $C_{[0,\delta]}$  with a mean function ‘ $a$ ’, we have applied the package ‘far’ developed by Damons and Guillas (Modelization for FAR processes Package: far Version: 0.6-2 License: LGPL-2.1 version 2.4.1 of the R Package). Recall that the Karhunen–Loève expansion of the Wiener process on the interval  $[0, \delta]$ ,  $\delta > 0$ , is given by

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

where  $Y_j^*$  are i.i.d r.v.’s  $N(0, 1)$ , and where we approximate this infinite series by a finite sum. We define a strong white noise  $\varepsilon = (\varepsilon_i(t), i = 0, 1, \dots, t \in [0, \delta])$  from a realization  $\omega_i$  of Wiener process on  $[0, \delta]$  by its increments for a given length  $s > 0$ :

$$\varepsilon_i(t) = \varepsilon_i(t, \omega_i) = W_{t+s}(\omega) - W_t(\omega).$$

To simulate the trajectories of the process  $Z = (Z(t); t \in [0, T])$ , with  $T = n\delta$  (a multiple of the period), we use the relation  $X_n(t) := Z(n\delta + t); t \in [0, \delta]; \delta = 2\pi$ , where  $X_n$  satisfies  $X_n - a = \rho(X_{n-1} - a) + \varepsilon_n; n = 1, 2, \dots$  with initial value  $X_0 = 0$  (or  $X_0 = \varepsilon_0$ ); the mean  $a(t) = \sin t$  and  $\rho$  is a kernel operator with the symmetric kernel  $k(s, t) = \cos(t - s)$ . Eventually others choices of the mean  $a$  and kernel  $k(s, t)$  are possible.

The following graph displays a trajectory of the process  $Z$  on the interval  $[0, T]$ , where  $T = n\delta = 100(2\pi)$  and  $m$  is the number of discretization points in the interval  $[0, 2\pi]$  (Figure 1).

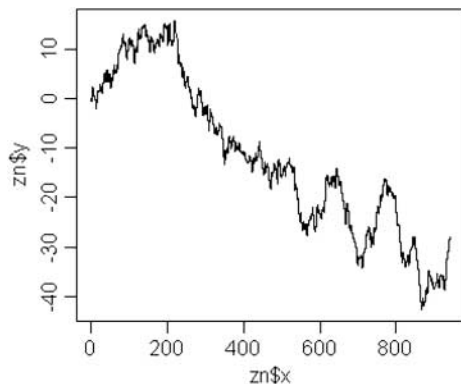


Figure 1. Example of an ARB(1) ( $m = 300$ ,  $n = 150$ ).

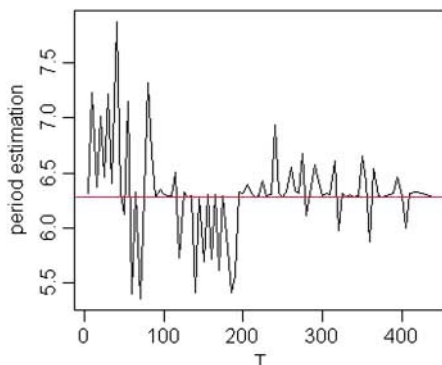


Figure 2. Behavior of the period estimator ( $\mu$  Lebesgue measure).

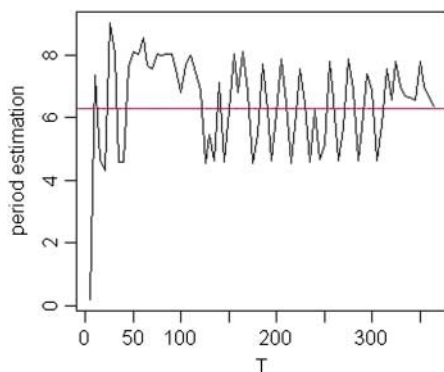


Figure 3. Behavior of the period estimator ( $\mu$  measure with density).

The following graph shows the behavior of the period estimator  $\hat{\delta}_T$  given by Equation (7) in function of  $T$  when the estimator  $\hat{g}_T$  given by Equation (4) is defined by the Lebesgue measure. We observe its convergence to the true period  $\delta_0 = 2\pi$  for large values of  $T$  (Figure 2).

The following graph shows the behavior of the period estimator  $\hat{\delta}_T$  in function of  $T$  when the estimator  $\hat{g}_T$  given by Equation (4) is defined by a measure  $\mu$  having a density  $\Phi_1$  (the first eigenfunction of  $R_{X_0}$ ) with respect to the Lebesgue measure (Figure 3).

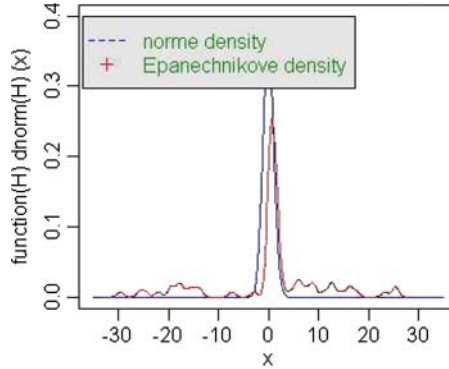


Figure 4. Kernel estimator ( $\mu$  Lebesgue measure).

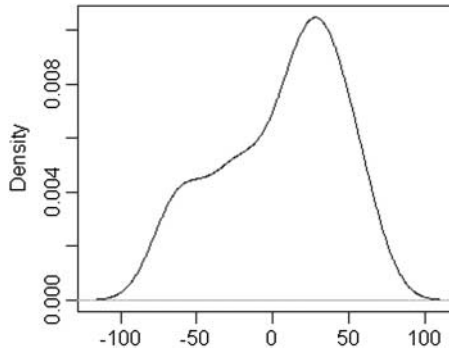


Figure 5. Kernel estimate.  $N = 68$ , bandwidth = 15.23 ( $\mu$  measure with density).

Table 1. Replicates of the period estimator

$n =$ multiple of period	Number of replicates	Mean	Standard error
100	20	6.431	0.3134
100	40	6.360	0.2927
100	60	6.332	0.2956
100	80	6.327	0.3006
150	20	6.290	0.3424
150	40	6.290	0.3370
150	60	6.286	0.3112
150	80	6.285	0.3049

The following graph shows the kernel estimator (with the Epanechnikov kernel) of the limit law of the period estimator  $\hat{\delta}_T$  with the Lebesgue measure in Equation (4) (Figure 4).

The following graph shows the kernel estimator (with the Epanechnikov kernel) of the limit law of the period estimator  $\hat{\delta}_T$  with having a density  $\Phi_1$  (the first eigenfunction of  $R_{X_0}$ ) with respect to the Lebesgue measure in  $\hat{g}_T$  (Figure 5).

Table 1 presents the performance of the period estimator  $\hat{\delta}_T$  when we repeat a number of replicates of the trajectories (here  $\delta_0 = 2\pi = 6,283185$ ). We conclude that in all these numerical simulations, each estimate shows a good behavior and is very close to the true value of the parameter. Also when we deal with others choices of the mean function ‘ $a$ ’ and kernel  $k(s, t)$  the simulations give the same conclusion.

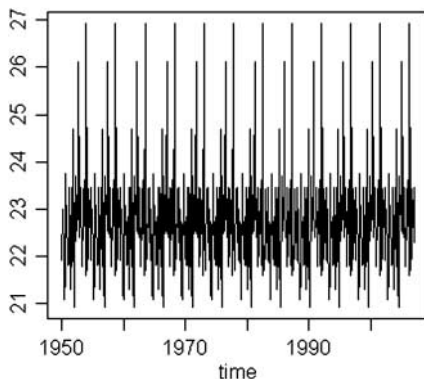


Figure 6. El Niño temperature from 1950 to 2006.

### 3.1 Real data

In this section, we use the period estimate  $\hat{\delta}_T$  to find the true period in some real data. For these climatic time series examples, we know that the true period is  $\delta_0 = 12$  months. So we try to use the estimate  $\hat{\delta}_T$  to recover this true period.

#### 3.1.1 Climatic sequences ENSO

First, we consider the sequence of the climatic phenomenon describing El Niño-Southern Oscillation (ENSO) which is known as a result of the interaction between the atmosphere and the ocean over the tropical Pacific Ocean. The phenomena EL Niño (EN) is an ocean component in ENSO while Southern Oscillation (SO) is an atmosphere component. The monthly mean Nino-3 surface temperature index in (Celsius degree) which provides a contracted description of ENSO are recorded from January 1950 to December 2006 by the National Prediction Environmental Center in United States. Figure 6 displays the time series of mean monthly temperature of this EN index (<http://www.cpc.ncep.noaa.gov/data/indices>)

For the estimation of the true period we take the parameter space  $\Theta := ]0, \Delta[ = ]0, 17[$ . In the definition of  $g$  (cf. (3)), we have to choose the mean function  $a$  and the measure  $\mu$ . Our choice is the natural function  $\cos$  or  $\sin$ . But the  $\sin$  function vanishes in 0 so it is eliminated in view to data. The estimator  $\hat{g}_T$  is given by Equation (4) and the estimator  $\hat{\delta}_T$  by Equation (7).

Case 1:  $a(t) = 23 * \cos(t * \pi/6)$  and  $\mu$  the Lebesgue measure.

This choice of 12 periodic mean function  $a$  is natural for this climatic phenomenon. The coefficient 23 is the mean of recorded monthly temperature (however other choices of coefficient are possible!) and  $\mu$  is the Lebesgue measure.

For a discretization grid 0.01 in  $\Theta := ]0, 17[$  we obtain:  $\hat{\delta}_T = 12.00$ ,

For a discretization grid 0.001 in  $\Theta := ]0, 17[$  we obtain:  $\hat{\delta}_T = 12.003$ .

So the true period 12 is well detected with an error of 3‰

Case 2:  $a(t) = 23 * \cos(t * \pi/6)$  and  $\mu$  has the density  $\Phi_1$  (the first eigenfunction of  $R_{X_0}$ ):

For a discretization grid 0.001 in  $\Theta := ]0, 17[$ , we obtain:  $\hat{\delta}_T = 12.561$ .

If we perturb the mean function choosing  $a$  having not 12-period, with  $\mu$  the Lebesgue measure, we obtain the following:

1. If  $a(t) = 23 * \cos(t * 2\pi/8)$  then  $\hat{\delta}_T = 11.995$ ,
2. If  $a(t) = 23 * \cos(t * 2\pi/13)$  then  $\hat{\delta}_T = 13.002$ .

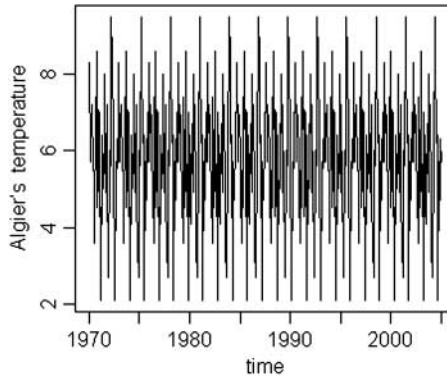


Figure 7. Algiers's temperature.

We observe here also that for the first case the true period 12 is well detected. However, we obtain an over estimation of the true period in the second case. If  $\mu$  has the density  $\Phi_1$  (the first eigenfunction of  $R_{X_0}$ ) the estimator  $\hat{\delta}_T$  has a very slow rate to detect the true period and it is difficult to draw reasonable conclusions.

### 3.1.2 Prediction of Algiers temperature

The following graph displays the series of the mean monthly temperature in Algiers for the period 1970–2004 (Figure 7):

We set  $\Theta := ]0, \Delta[ = ]0, 17[$  and we choose in the definition of  $g$ :  $a(t) = 11.86 * \cos(t * \pi/6)$  and  $\mu$  the Lebesgue measure (where the coefficient 11.86 is taken as the mean of the recorded temperature). The results of the simulation are

For a discretization grid 0.01, we obtain:  $\hat{\delta}_T = 11.99$ ,

For a discretization grid 0.001, we obtain:  $\hat{\delta}_T = 11.994$ .

Hence the true period 12 is well detected with error 6%.

If we perturb the mean function choosing others means  $a$  having not 12 period (with the Lebesgue measure), we obtain:

1. If  $a(t) = 11.86 * \cos(t * 2\pi/8)$ , then  $\hat{\delta}_T = 12.000$ ,
2. If  $a(t) = 11.86 * \cos(t * 2\pi/13)$ , then  $\hat{\delta}_T = 12.990$ .

Hence the true period 12 is also well detected in these cases.

## 4. Conclusion

We point out that in all numerical simulations and examples our period estimator produces a good result detecting well the true period. However, in both numerical simulations and examples, we have large samples of observations so the estimates should possess a good behavior. On the other hand, eventually we could take others choices of the mean  $a$  and kernel  $k(s, t)$  to make more numerical simulations. In this study, we have limited ourselves only to these cases. Also it remains to determine the exact rate of convergence of the estimators which provide an interesting topics in future research. Finally, there exists a wide scope of applied fields for which our approach could be of interest and it is interesting to find examples in Biometrics, Genetics or Environmetrics, etc.



## Acknowledgements

We thank the Editor and two referees for their constructive comments which have helped us to improve the presentation of the paper.

## References

- [1] A. Antoniadis and T. Sapatinas, *Wavelet methods for continuous-time prediction using Hilbert-valued autoregressive processes*, J. Multivariate Anal. 87 (2003), pp. 133–158.
- [2] P. Besse, H. Cardot, and D.B. Stephenson, *Autoregressive forecasting of some functional climatic variations*, Scand. J. Statist. 27(4), pp. 673–687.
- [3] D. Bosq, *Linear Process in Function Spaces. Theory and Application*. Lectures Notes in Statistics Vol. 149. Springer, Berlin, 2000.
- [4] M. Davidian, X. Lin, and J. Wang, *Introduction to Emerging issues in longitudinal and functional data analysis*, Statistica Sinica 14(3) (2004), pp. 613–614.
- [5] N. Dunford and J.T. Schwartz, *Linear Operators*, Wiley, New York, 1958.
- [6] F. Ferraty, A. Laksaci, A. Tadj, and P. Vieu, *Kernel regression with functional response*, Electron. J. Statist. 5 (2011), pp. 159–171.
- [7] F. Ferraty and Y. Romain, *The Oxford Handbook of Functional Data Analysis*, Oxford University Press, Oxford, 2011.
- [8] F. Ferraty and P. Vieu, *Nonparametric Functional Data Analysis. Theory and Practice*, Springer, Berlin, 2006, MR2229687.
- [9] F. Ferraty, P. Vieu, and L. Delsol, *Structural test in regression on functional variables*, J. Multivariate Anal. 102 (2006), pp. 422–447.
- [10] W. González Manteiga and P. Vieu, *Statistics for functional data*, Comput. Statist. Data Anal. 51(10) (2007), pp. 4788–4792.
- [11] Yu. Kutoyants, *Minimum distance parameter estimation for diffusion type observation*, C. R. Acad. Sci. Paris, Ser. I (1991), pp. 637–642.
- [12] P.W. Millar, *A general approach to the optimality of minimum distance estimation*, Trans. Amer. Soc. 286(1) (1984), pp. 377–418.
- [13] T. Mourid and W. Benyelles, *Estimation de la période de la représentation autorégressive d'un processus à temps continu*, Annales de l'I.S.U.P. Volume XXXXVI-Fascicule 1–2 (2002), pp. 89–101.
- [14] C. Parr, *Minimum distance estimation*, Bibliography, Commun. Statist. Theory Math. A 10(12) (1981), pp. 1205–1224.
- [15] B. Pumo, *Prediction of continuous time processes by  $C_{[0,1]}$ -valued autoregressive process*, Stat. Inference Stoch. Process. 1(3) (1998), pp. 297–309.
- [16] J.O. Ramsay and B.W. Silverman, *Applied Functional Data Analysis*, Springer, New York, 2002.
- [17] J.O. Ramsay and B.W. Silverman, *Functional Data Analysis*, 2nd ed. Springer, New York, 2005.
- [18] M. Valderrama, *Introduction to the special issue on modelling functional data in practice*, Comput. Statist. 22 (2007), pp. 331–334.

## Appendix: Proofs

*Proof of Theorem 1* Firstly, we remark for all  $T \in \mathbb{R}^+$  and  $\delta > 0$ , there exists an integer  $n = n_{T,\delta}$  such as:  $n\delta \leq T \leq (n+1)\delta$  and  $n/T \rightarrow 1/\delta$ . From the definition (4) of the estimator  $\hat{g}_T$  we have the following decomposition:

$$\begin{aligned}
 \hat{g}_T &= \frac{1}{T} \int_0^T Z(t) \, d\mu_{[T]}(t) \\
 &= \frac{1}{T} \int_0^{n\delta} Z(t) \, d\mu_{[T]}(t) + \frac{1}{T} \int_{n\delta}^T Z(t) \, d\mu_{[T]}(t) \\
 &:= A_n + B_n.
 \end{aligned} \tag{A1}$$

We treat each term separately. For the first term  $A_n$ ,  $X_j(t) := Z(j\delta + t)$ ,  $t \in [0, \delta]$  and the definition (5) of  $\mu_{[T]}$ , we have:

$$\begin{aligned}
 A_n &:= \frac{1}{T} \int_0^{n\delta} Z(t) \, d\mu_{[T]}(t) \\
 &= \frac{1}{T} \sum_{j=0}^{n-1} \int_{j\delta}^{(j+1)\delta} Z(t) \, d\mu_{[T]}(t) \\
 &= \frac{1}{T} \sum_{j=0}^{n-1} \int_0^\delta Z(\delta j + t) \, d\mu(t) \\
 &= \frac{n}{T} \int_0^\delta \left[ \frac{1}{n} \sum_{j=0}^{n-1} X_j(t) \right] \, d\mu(t). \tag{A2}
 \end{aligned}$$

According to the strong law of large numbers in the Banach space  $(B, \|\cdot\|) = (C_{[0,\delta]}, \|\cdot\|_\infty)$  [3, Theorem 3.7 Chapter 3] we have:

$$\begin{aligned}
 \|\bar{X}_n - a\|_\infty &:= \left\| \frac{1}{n} \sum_{j=0}^{n-1} X_j - a \right\|_\infty \\
 &= \sup_{0 \leq t \leq \delta} \left| \frac{1}{n} \sum_{j=0}^{n-1} X_j(t) - a(t) \right| \rightarrow 0 \quad \text{a.s.}
 \end{aligned}$$

Triangular inequality and almost surely continuity of the trajectories of the process  $Z$  imply

$$\begin{aligned}
 \left\| \frac{1}{n} \sum_{j=0}^{n-1} X_j \right\|_\infty &\leq \frac{1}{n} \sum_{j=0}^{n-1} \|X_j\|_\infty \leq \sup_{0 \leq j \leq (n-1)} \|X_j\|_\infty \\
 &\leq \sup_{0 \leq j \leq (n-1)} \sup_{0 \leq t \leq \delta} |X_j(t)| \\
 &= \sup_{0 \leq j \leq (n-1)} \sup_{0 \leq t \leq \delta} |Z(\delta j + t)| < \infty \quad \text{a.s.}
 \end{aligned}$$

Hence from the dominated convergence theorem and the fact that  $n/T \rightarrow 1/\delta$ , we deduce from Equation (A2):

$$A_n := \frac{1}{T} \int_0^{n\delta} Z(t) \, d\mu_{[T]}(t) = \frac{n}{T} \int_0^\delta \left[ \frac{1}{n} \sum_{j=0}^{n-1} X_j(t) \right] \, d\mu(t) \rightarrow \frac{1}{\delta} \int_0^\delta a(t) \, d\mu(t) \quad \text{a.s.}$$

On the other hand, for the term  $B_n$ , we have

$$\begin{aligned}
 B_n &:= \frac{1}{T} \int_{n\delta}^T Z(t) \, d\mu_{[T]}(t) \\
 &\leq \frac{1}{T} \sup_{n\delta \leq t \leq T} |Z(t)| \cdot \|\mu\|_{TV} \\
 &= \frac{1}{T} \sup_{0 \leq s \leq \delta} |Z(s)| \cdot \|\mu\|_{TV} \\
 &\leq \frac{1}{T} \|X_n\|_\infty \|\mu\|_{TV}, \tag{A3}
 \end{aligned}$$

Downloaded by [taha mourid] at 13:46 08 April 2012

where  $\|\mu\|_{TV}$  denoted the total variation norm of the measure  $\mu$ . The Markov inequality and the stationarity of the process  $(X_n)$  give: for all  $\varepsilon > 0$ ,

$$\begin{aligned} P\left(\left\|\frac{X_n}{T}\right\| > \varepsilon\right) &\leq \frac{1}{T^2\varepsilon^2}E\|X_n\|^2 \\ &= \frac{1}{T^2\varepsilon^2}E\|X_0\|^2. \end{aligned}$$

Then

$$\int_1^\infty P\left(\left\|\frac{X_n}{T}\right\| > \varepsilon\right) dT \leq \frac{E\|X_0\|^2}{\varepsilon^2} \int_1^\infty \frac{1}{T^2} dT < \infty.$$

Hence by Borel–Cantelli lemma, we deduce:

$$\frac{X_n}{T} \xrightarrow[n \rightarrow \infty]{} 0 \quad \text{a.s.}$$

A consequence of Equation (A3), we obtain:

$$B_n := \frac{1}{T} \int_{n\delta}^T Z(t) d\mu_{[T]}(t) \xrightarrow[n \rightarrow \infty]{} 0 \quad \text{a.s.}$$

Finally regrouping the results (A2) and (A3), we obtain from Equation (A1):

$$\lim_{T \rightarrow \infty} \hat{g}_T = \lim_{T \rightarrow \infty} \frac{1}{T} \int_0^T Z(t) d\mu_{[T]}(t) = \frac{1}{\delta} \int_0^\delta a(t) d\mu(t) \quad \text{a.s.}$$

The convergence  $L^2$  is obtained in a similar way using the fact that the convergence  $\|\bar{X}_n - a\|_\infty \rightarrow_{n \rightarrow \infty} 0$  is also true in  $L^2$  [3, Theorem 3.7, Chapter 3]. ■

*Proof of Theorem 2* First, let us verify that the trajectories of the sequence  $(\varepsilon_n, n \in \mathbb{Z})$  satisfy the Lipschitz condition (condition 3 of Theorem 2.8, Chapter 2 in [3]) which gives the central limit theorem in the Banach space  $B = C_{[0,\delta]}$ .

Recall  $X_j(t) := Z(j\delta + t)$  and the process  $(Z(t), t \in R)$  is an AR process with a innovation  $(\varepsilon_n, n \in \mathbb{Z})$ . From the relation (1) we have  $\rho(X_n) = E(X_{n+1}/X_j, j \leq n)$ . So from Equations (1) and (2) we obtain:

$$\varepsilon_n(t) - \varepsilon_n(s) = Z(n\delta + t) - Z(n\delta + s) - E\left(Z(n\delta + t) - \frac{Z(n\delta + s)}{Z(u)}, u \leq n\delta\right).$$

The triangular inequality and under the Lipschitz condition on the trajectories of the process  $(Z(t), t \in R)$  imply

$$\begin{aligned} |\varepsilon_n(t) - \varepsilon_n(s)| &\leq |Z(n\delta + t) - Z(n\delta + s)| + E\left(\frac{|Z(n\delta + t) - Z(n\delta + s)|}{Z(u)}, u \leq n\delta\right) \\ &\leq \left(M + E\left(\frac{M}{Z(u)}, u \leq n\delta\right)\right) \cdot |t - s|. \end{aligned}$$

Hence the sequence  $(\varepsilon_n, n \in \mathbb{Z})$  satisfies condition 3 of Theorem 2.8. in [3]. So the sequence  $(X_n, n \in \mathbb{Z})$  verifies the central limit theorem in  $C_{[0,\delta]}$  :

$$\sqrt{n}(\bar{X}_n - a) \implies_{n \rightarrow \infty} N,$$

where  $N$  is a  $C_{[0,\delta]}$ -valued Gaussian r.v.'s with zero mean and covariance operator  $(I - \rho)^{-1}R_{\varepsilon_0}(I - \rho)^{* -1}$ .

Let  $\mu$  be a linear form in the topological dual  $C_{[0,\delta]}^*$  of  $C_{[0,\delta]}$  which consists of all Borelean measures on  $[0, \delta]$ . By the continuity theorem, we deduce the following central limit theorem in  $\mathbb{R}$ :

$$\mu(\sqrt{n}(\bar{X}_n - a)) \implies_{n \rightarrow \infty} \mu(N) := \int_0^\delta N(t) d\mu(t),$$

where  $\mu(N) := \int_0^\delta N(t) d\mu(t)$  is a real zero mean Gaussian r.v.'s. Set  $N_1 = \mu(N)$ .

But

$$\begin{aligned} \mu(\sqrt{n}(\bar{X}_n - a)) &= \int_0^\delta \sqrt{n}(\bar{X}_n(t) - a(t)) d\mu(t) \\ &= \sqrt{n} \int_0^\delta \left( \frac{1}{n} \sum_{j=0}^{n-1} X_j(t) - a(t) \right) d\mu(t) \\ &= \int_0^\delta \left( \frac{1}{\sqrt{n}} \sum_{j=0}^{n-1} X_j(t) - \sqrt{n}a(t) \right) d\mu(t) \\ &= \int_0^\delta \left( \frac{1}{\sqrt{n}} \sum_{j=0}^{n-1} X_j(t) - \frac{n}{\sqrt{n}}a(t) \right) d\mu(t) \\ &= \frac{1}{\sqrt{n}} \sum_{j=0}^{n-1} \int_0^\delta (X_j(t) - a(t)) d\mu(t). \end{aligned}$$

Therefore

$$\frac{1}{\sqrt{n}} \sum_{j=0}^{n-1} \int_0^\delta (X_j(t) - a(t)) d\mu(t) \implies_{n \rightarrow \infty} N_1 = \mu(N) = \int_0^\delta N(t) d\mu(t).$$

The Banach r.v.  $N$  is Gaussian in  $C_{[0,\delta]}$  and has a zero mean and covariance operator  $(I - \rho)^{-1}R_{\epsilon_0}(I - \rho)^{* -1}$ , so the real r.v.  $N_1 = \mu(N)$  has a zero mean and a variance  $\mu((I - \rho)^{-1}R_{\epsilon_0}(I - \rho)^{* -1}(\mu))$ .

Now by the hypothesis, the measure  $\mu$  has  $\Phi_1$  as a density with respect to the Lebesgue measure on the interval  $[0, \delta]$ . Therefore, we have the following weak convergence:

$$\begin{aligned} \mu(\sqrt{n}(\bar{X}_n - a)) &= \frac{1}{\sqrt{n}} \sum_{j=0}^{n-1} \int_0^\delta (X_j(t) - a(t)) \cdot \Phi_1(t) dt \\ &\implies_{n \rightarrow \infty} \mu(N) = \int_0^\delta N(t) \cdot \Phi_1(t) dt. \end{aligned}$$

Let us compute the variance of the r.v.'s  $N_1 = \mu(N) = \int_0^\delta N(t) \cdot \Phi_1(t) dt$ . First,  $C(s, t)$  is the covariance function of the process  $(X_0(t), t \in [0, \delta])$  so the covariance function of  $(N(t), t \in [0, \delta])$  is, for all  $s, t \in [0, \delta]$ ,

$$\begin{aligned} C(s, t) &= E((X_0(s) - a(s)) \cdot (X_0(t) - a(t))) = C(s, t) \\ &= E(N(s)N(t)). \end{aligned}$$

Downloaded by [tahar mourid] at 13:46 08 April 2012

Then

$$\begin{aligned}
 E\mu^2(N) &= E\left(\int_0^\delta N(t) d\mu(t)\right)^2 = E\left(\int_0^\delta N(t)\Phi_1(t) dt\right)^2 \\
 &= \int_0^\delta \int_0^\delta EN(s)N(t)\Phi_1(s) ds\Phi_1(t) dt \\
 &= \int_0^\delta \left(\int_0^\delta C(t-s)\Phi_1(s) ds\right) \Phi_1(t) dt \\
 &= \int_0^\delta \lambda_1 \Phi_1(t)\Phi_1(t) dt \\
 &= \lambda_1 \int_0^\delta \Phi_1^2(t) dt = \lambda_1.
 \end{aligned}$$

The proof of the theorem is complete. ■

*Proof of Theorem 3* First, we have the following decomposition using the definition of the estimator  $\hat{g}_T$  :

$$\begin{aligned}
 &\frac{1}{\sqrt{T}} \sum_{i=0}^{T-1} \int_0^{\delta_0} Z(i\delta_0 + t)\Phi_1(t) dt \\
 &= \frac{1}{\sqrt{T}} \left( \int_0^{\delta_0} Z(t)\Phi_1(t) dt + \int_0^{\delta_0} Z(\delta_0 + t)\Phi_1(t) dt + \dots + \int_0^{\delta_0} Z((T-1)\delta_0 + t)\Phi_1(t) dt \right) \\
 &= \frac{1}{\sqrt{T}} \int_0^{T\delta_0} Z(t) d\mu_{[T]}(t) \\
 &= \sqrt{T} \left( \frac{1}{T} \int_0^T Z(t) d\mu_{[T]}(t) + \frac{1}{T} \int_T^{T\delta_0} Z(t) d\mu_{[T]}(t) \right) \\
 &= \sqrt{T} \cdot \hat{g}_T + \frac{1}{\sqrt{T}} \int_T^{T\delta_0} Z(t) d\mu_{[T]}(t).
 \end{aligned}$$

In the last line, for the second term we use a similar way given in the proof of the Theorem 1 to show:

$$\left| \frac{1}{\sqrt{T}} \int_T^{T\delta_0} Z(t) d\mu_{[T]}(t) \right| \leq \frac{\|X_{[T\delta_0]}\|_\infty}{[T\delta_0]} \rightarrow_{T \rightarrow \infty} 0 \quad \text{in probability.}$$

Hence

$$\sqrt{T} \cdot \hat{g}_T = \frac{1}{\delta_0 \sqrt{T}} \sum_{i=0}^{T-1} \int_0^{\delta_0} Z(i\delta_0 + t)\Phi_1(t) dt + o_p(1).$$

So

$$\begin{aligned}
 \sqrt{T}(\hat{g}_T - g(\delta_0)) &= \frac{1}{\delta_0 \sqrt{T}} \sum_{i=0}^{T-1} \int_0^{\delta_0} Z(i\delta_0 + t)\Phi_1(t) dt - \sqrt{T} \cdot \frac{1}{\delta_0} \int_0^{\delta_0} a(t)\Phi_1(t) dt + o_p(1) \\
 &= \frac{1}{\delta_0} \cdot \frac{1}{\sqrt{T}} \left[ \sum_{i=0}^{T-1} \int_0^{\delta_0} Z(i\delta_0 + t)\Phi_1(t) dt - T \int_0^{\delta_0} a(t)\Phi_1(t) dt \right] + o_p(1) \\
 &= \frac{1}{\delta_0} \left[ \frac{1}{\sqrt{T}} \sum_{i=0}^{T-1} \int_0^{\delta_0} (Z(i\delta_0 + t) - a(t))\Phi_1(t) dt \right] + o_p(1).
 \end{aligned}$$

Now for the first term, we apply Theorem 2 to deduce it is asymptotically gaussian and then we arrive at

$$\delta_0 \cdot \sqrt{T}(\hat{g}_T - g(\delta_0)) \implies_{T \rightarrow \infty} N_1,$$

where  $N_1$  is a zero mean gaussian r.v.'s with variance  $\lambda_1$  which is the first eigenvalue of the operator  $R_{X_0}$ .

Hence the r.v.'s

$$\sqrt{T}(\hat{g}_T - g(\delta_0)) \implies_{T \rightarrow \infty} N_2,$$

where  $N_2$  is a zero mean gaussian r.v.'s with variance  $\delta_0^{-2}\lambda_1$ .

The proof is complete. ■

*Proof of Theorem 4* Let  $\beta > 0$  and  $\delta_0 \in \Theta$  the true value of the parameter. From the definition of the estimator  $\hat{\delta}_T$  (cf. Equation (7)), we have:

$$|\hat{\delta}_T - \delta_0| > \beta \implies \inf_{|\delta - \delta_0| \leq \beta} |\hat{g}_T - g(\delta)| > \inf_{|\delta - \delta_0| > \beta} |\hat{g}_T - g(\delta)|. \tag{A4}$$

By triangular inequality

$$|\hat{g}_T - g(\delta)| \leq |\hat{g}_T - g(\delta_0)| + |g(\delta_0) - g(\delta)|.$$

Therefore,

$$\inf_{|\delta - \delta_0| \leq \beta} |\hat{g}_T - g(\delta)| \leq |\hat{g}_T - g(\delta_0)|. \tag{A5}$$

Also

$$|g(\delta) - g(\delta_0)| \leq |\hat{g}_T - g(\delta)| + |\hat{g}_T - g(\delta_0)|.$$

Then

$$|\hat{g}_T - g(\delta)| \geq |g(\delta) - g(\delta_0)| - |\hat{g}_T - g(\delta_0)|.$$

Hence, the definition of  $h(\beta)$  (cf. Equation (8)) implies:

$$\begin{aligned} \inf_{|\delta - \delta_0| > \beta} |\hat{g}_T - g(\delta)| &\geq \inf_{|\delta - \delta_0| > \beta} |g(\delta) - g(\delta_0)| - |\hat{g}_T - g(\delta_0)| \\ &\geq h(\beta) - |\hat{g}_T - g(\delta_0)|. \end{aligned} \tag{A6}$$

Using Equations (A4), (A5), (A6), we deduce

$$\begin{aligned} P_{\delta_0}(|\hat{\delta}_T - \delta_0| > \beta) &\leq P_{\delta_0} \left( \inf_{|\delta - \delta_0| \leq \beta} |\hat{g}_T - g(\delta)| > \inf_{|\delta - \delta_0| > \beta} |\hat{g}_T - g(\delta)| \right) \\ &\leq P_{\delta_0} \left( \inf_{|\delta - \delta_0| \leq \beta} |\hat{g}_T - g(\delta)| > h(\beta) - |\hat{g}_T - g(\delta_0)| \right) \\ &\leq P_{\delta_0} \left( |\hat{g}_T - g(\delta_0)| > \frac{1}{2}h(\beta) \right). \end{aligned}$$

Because the condition  $h(\beta) > 0$  and  $\hat{g}_T \xrightarrow{T \rightarrow \infty} g(\delta_0)$  a.s. by Theorem 1, the last bound tends to 0 when  $T \rightarrow \infty$ .

The proof is now complete. ■

*Proof of Theorem 5* For the limit law of the estimator  $\hat{\delta}_T$ , we follow the general theory of minimum distance estimate presented in [9] to have a Taylor's expansion giving the following relation between the estimators  $\hat{\delta}_T$  and  $\hat{g}_T$ :

$$\sqrt{T}(\hat{\delta}_T - \delta_0) = I^{-1}(\delta_0) \cdot \sqrt{T}(\hat{g}_T - g(\delta_0)) + o_p(1), \quad (\text{A7})$$

where the function  $I(\delta)$  is the derivative function of  $g(\delta)$ .

Now we apply Theorem 3 to the first term to deduce

$$I^{-1}(\delta_0) \cdot \sqrt{T}(\hat{g}_T - g(\delta_0)) \implies_{T \rightarrow \infty} N_3,$$

where  $N_3$  is a zero mean Gaussian r.v.'s with variance  $\delta_0^{-2} \cdot I^{-2}(\delta_0) \cdot \lambda_1$ .

So from Equation (A7), we imply the result for the r.v.'s:  $\sqrt{T}(\hat{\delta}_T - \delta_0) \implies_{T \rightarrow \infty} N_3$ . ■

*Proof of Remark 2* Let  $(X_n)$  be an AR process defined (1). From Theorem 3.2 in [3] and the assumptions of the Remark 2 we have:

$$R_{X_0} = \rho R_{X_1 X_0} + R_{\varepsilon_0}, \quad R_{X_0 X_1} = \rho R_{X_0} \quad \text{and} \quad R_{X_1 X_0} = (R_{X_0 X_1})^* = R_{X_0} \rho,$$

where  $R_{X_1 X_0}$  is the cross-covariance operator of the r.v.'s  $X_0$  and  $X_1$ .

So

$$\begin{aligned} R_{X_0} &= \rho R_{X_0} \rho + R_{\varepsilon_0} \\ &= R_{X_0} \rho^2 + R_{\varepsilon_0}. \end{aligned}$$

As the operator  $\rho$  has the eigenelements  $(\mu_i, \Phi_i)$ , we obtain

$$\begin{aligned} R_{X_0}(\Phi_i) &= R_{X_0} \rho(\Phi_i) + R_{\varepsilon_0}(\Phi_i) \\ &= R_{X_0} \rho(\mu_i \Phi_i) + \beta_i \Phi_i \\ &= \mu_i R_{X_0} \rho(\Phi_i) + \beta_i \Phi_i \\ &= \mu_i^2 R_{X_0}(\Phi_i) + \beta_i \Phi_i. \end{aligned}$$

Hence

$$(1 - \mu_i^2) R_{X_0}(\Phi_i) = \beta_i \Phi_i.$$

Finally,

$$R_{X_0}(\Phi_i) = \frac{\beta_i}{(1 - \mu_i^2)} \Phi_i.$$

Therefore the eigenelements of  $R_{X_0}$  are  $(\beta_i / (1 - \mu_i^2), \Phi_i)$ . ■

# Estimation de la période d'un processus à temps continu à représentation autorégressive

Wafaa BENYELLES, Tahar MOURID

Département de mathématiques, faculté des sciences, Université de Tlemcen 13000, Algérie

(Reçu le 22 septembre 2000, accepté après révision le 5 juin 2001)

**Résumé.** Nous considérons l'estimation de la période d'un processus aléatoire à représentation autorégressive. Nous étudions un estimateur de la période de la fonction moyenne basé sur une méthode de la « distance minimale ». Sous des conditions de régularité, nous montrons sa convergence en probabilité et nous établissons une loi limite. © 2001 Académie des sciences/Éditions scientifiques et médicales Elsevier SAS

## *Period estimation of process with autoregressive representation*

**Abstract.** We consider the class of stochastic processes which admit an autoregressive representation. Their mean function is periodic and we construct an estimator of the period based on the "minimum distance". Under regularity conditions we show its consistence and give its asymptotic distribution. © 2001 Académie des sciences/Éditions scientifiques et médicales Elsevier SAS

## 1. Introduction

Soit  $(\Omega, \mathcal{A}, P)$  un espace probabilisé complet et  $(B, \|\cdot\|)$  un espace de Banach réel séparable muni de sa tribu borélienne  $\mathcal{B}$ . On note  $\mathcal{L}(B)$  l'algèbre de Banach des opérateurs linéaires bornés définis sur  $B$  et à valeurs dans  $B$ , munie de la norme usuelle  $\|\cdot\|_1$ , des opérateurs bornés. Un bruit blanc banachique est une suite de variables aléatoires  $(\varepsilon_n, n \in \mathbb{Z})$  indépendantes identiquement distribuées définies sur  $(\Omega, \mathcal{A}, P)$  à valeurs dans  $B$  vérifiant pour tout  $n \in \mathbb{Z}$  :

$$E\varepsilon_n = 0; \quad 0 < E\|\varepsilon_n\|^2 = \sigma_\varepsilon^2 < \infty.$$

Une suite  $(X_n, n \in \mathbb{Z})$  définie sur  $(\Omega, \mathcal{A}, P)$  et à valeurs dans  $B$  est un processus autorégressif banachique d'ordre  $p$ ,  $(ARB(p))$ ,  $p \in \mathbb{N}^*$ , s'il existe des opérateurs  $\rho_1, \rho_2, \dots, \rho_p$  dans  $\mathcal{L}(B)$ , un bruit blanc banachique  $(\varepsilon_n)$  et un élément  $a \in B$  tels que :

$$X_n - a = \rho_1(X_{n-1} - a) + \rho_2(X_{n-2} - a) + \dots + \rho_p(X_{n-p} - a) + \varepsilon_n \quad \text{p.s.} \quad (1)$$

où l'opérateur  $\rho_p$  est différent de l'opérateur nul. Dans [9], chap. I, on trouve une condition sur les opérateurs  $\rho_i$  pour que (1) admette une solution strictement stationnaire dans  $B$ . Nous dirons qu'un

Note présentée par Paul DEHEUVELS.

S0764-4442(01)02030-4/FLA

© 2001 Académie des sciences/Éditions scientifiques et médicales Elsevier SAS. Tous droits réservés



processus  $\mathcal{Z} = (\mathcal{Z}(t), t \in \mathbb{R})$  à temps continu admet une représentation autorégressive banachique d'ordre  $p$  dans un espace de Banach  $B$ , s'il existe des opérateurs linéaires bornés  $\rho_1, \rho_2, \dots, \rho_p$  sur  $B$ ,  $a \in B$ , une suite  $(\varepsilon_n, n \in \mathbb{Z})$  de variables aléatoires indépendantes identiquement distribuées définies sur  $\Omega$  à valeurs dans  $B$  et un nombre  $\delta$  strictement positif tels que la suite des variables aléatoires à valeurs dans  $B$  définies par :

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

vérifient la relation (1). Dans de nombreuses applications, le processus  $\mathbb{Z} = (\mathbb{Z}_t, t \in \mathbb{R})$  est p.s. à trajectoires continues et l'espace de Banach  $B$  naturel est l'espace des fonctions définies et continues sur  $[0, \delta]$ , noté  $C = C_{[0, \delta]}$  ou encore parfois l'espace  $L^2_{[0, \delta]}$ . Dans [3], chap. 3, [9], chap. 3, on trouve des exemples de processus à temps continu admettant une représentation autorégressive dans des espaces appropriés. De plus, si  $\mathcal{Z}$  admet une représentation autorégressive ARB( $p$ ), alors sa fonction moyenne est une fonction périodique de période  $\delta$  et coïncide avec  $a$  sur  $[0, \delta]$  ([9], p. 64). Un problème naturel en statistique des processus est l'estimation de la fonction moyenne et en particulier dans ce cadre se pose le problème de l'estimation de la période  $\delta$ . Dans [7] l'estimation de la période a été résolue par une méthode de Tian qui repose des propriétés de la fonction covariance pour des processus périodiquement corrélés. Cette classe est caractérisée par le fait que la fonction moyenne et la fonction covariance sont des fonctions périodiques et de même période. Nous indiquerons que dans le cas de processus  $\mathbb{Z}$  à représentation autorégressive Banach que cette méthode n'est pas applicable. Nous proposons un estimateur de la période  $\delta$  du type « distance minimale » [8,6,10]. La méthode est basée sur l'étude d'un estimateur d'une intégrale de la fonction moyenne du processus (convergence presque sûre et dans  $L^2$  et loi limite) et puis nous en déduisons un estimateur de la période par la minimisation de l'écart absolu entre l'estimateur et la fonction à estimer. Nous établissons par la suite la convergence en probabilité et une loi limite pour l'estimateur de la distance minimale étudiée. Les résultats sont obtenus uniformément sur l'espace paramétrique.

## 2. Estimation de la période

Rappelons qu'un processus stochastique  $(\mathcal{Z}(t), t \in \mathbb{Z})$  réel défini sur un espace de probabilité  $(\Omega, \mathcal{A}, P)$  et ayant un moment d'ordre deux fini est appelé processus périodiquement corrélé si la fonction de covariance  $R$  et la fonction moyenne  $m$  sont périodiques de période  $\delta$  :  $R(u, v) = E[\mathcal{Z}(u) \cdot \mathcal{Z}(v)] = R(u + \delta, v + \delta)$  et  $m(u) = E[\mathcal{Z}(u)] = m(u + \delta), \forall (u, v) \in \mathbb{Z} \times \mathbb{Z}$  tel que  $(u + \delta, v + \delta) \in \mathbb{Z} \times \mathbb{Z}$ . Dans la suite, cette classe est notée PC. Dans [7] une synthèse de la méthode de Tian pour l'estimation de la période  $\delta$  pour un processus PC est proposée. La méthode est basée sur des propriétés de la fonction de covariance qui est un moyen usuel de recherche de périodicité en théorie du signal. Nous examinons l'application de la méthode de Tian aux processus ARB( $p$ ). Comme nous l'avons signalé en introduction un processus à temps continu  $(\mathcal{Z}(t), t \in \mathbb{R})$  à représentation autorégressive Banach possède une moyenne périodique de période  $\delta$ . En revanche, la fonction de covariance ne l'est pas. En effet, si  $(\mathcal{Z}(t))$  est à représentation ARC( $l$ ) dans  $C_{[0, \delta]}$  de moyenne  $a$  et de bruit blanc  $\varepsilon = (\varepsilon_n)$ , un calcul montre que pour tout  $t \in [0, \delta]$  :  $C(t + \delta, t + \delta) = E[\mathcal{Z}(t + \delta) - a(t)][\mathcal{Z}(t + \delta) - a(t)] = C(t, t) + c$  avec  $c$  strictement positif. Donc la fonction de covariance  $C(s, t)$  n'est pas périodique. Par suite les processus à représentation ARB( $p$ ) ne sont pas en général élément de la classe PC et par conséquent la méthode de Tian ne s'applique pas ici.

Soit  $\mathcal{Z} = (\mathcal{Z}(t), t \in \mathbb{R})$  un processus réel presque sûrement à trajectoires continues. On suppose que le processus  $\mathcal{Z}$  admet une représentation autorégressive ARC( $p$ ) de moyenne  $a$  et de bruit blanc  $\varepsilon$  dans  $C_{[0, \delta]}$  où  $\delta$  un réel strictement positif. Pour estimer la période  $\delta$  nous utilisons un estimateur qui a été introduit et étudié dans [3,9]. Posons pour  $\delta > 0$ :

$$g(\delta) := \frac{1}{\delta} \int_0^\delta a(t) d\mu(t),$$

## Estimation de la période d'un processus à temps continu à représentation autorégressive

où  $\mu$  est une mesure à signe bornée dans  $C_{[0,\delta]}^*$  (le dual topologique de  $C_{[0,\delta]}$ ). Si nous observons une trajectoire du processus  $Z$  sur  $[0, T] : (Z(t), t \in [0, T])$ , alors un estimateur naturel de la fonction  $g$  est le suivant :

$$\hat{g}_T = \frac{1}{T} \int_0^T Z(t) \cdot d\mu_{[T]}(t), \quad (3)$$

où  $[T]$  est la partie entière de  $T$  et pour tout  $n \in \mathbb{Z}$ , et pour tout borélien  $A$ ,  $A \in B_{[0,n\delta]}$ , la mesure  $\mu_n$  est définie par (translatée de  $\mu$ ) :

$$\mu_n(A) := \sum_{j=0}^{n-1} \mu(A \cap [j\delta, (j+1)\delta] - j\delta).$$

L'étude de la convergence et loi limite de cet estimateur a été faite pour  $p = 1$  et  $\delta = 1$  dans [3], chap. 6.7. Dans la première partie de cette Note nous reprenons l'étude de la convergence presque sûre et dans  $L^2$  ensuite une loi limite dans  $\mathbb{R}^k$  de  $(\hat{g}_{T,i})$  associée à des mesures  $\mu_i$ ,  $i = 1, \dots, k$ , pour  $p \geq 1$  et  $\delta > 0$ . Le résultat suivant établit la convergence presque sûre et dans  $L^2$  de  $\hat{g}_T$  :

**THÉORÈME 1.** – *Si le processus  $Z$  admet une représentation autorégressive  $ARC(p)$  de moyenne  $a$  et de bruit blanc  $\varepsilon$  dans  $C_{[0,\delta]}$ , alors, pour tout  $\delta > 0$ ,*

$$\lim_{T \rightarrow \infty} \hat{g}_T \rightarrow g(\delta)$$

*presque sûrement et dans  $L^2$ .*

La suite  $ARC(p) : X = (X_n)$  associée au processus  $Z$  par (2) admet un opérateur de covariance  $R_{X_0}$  qui est dans ce cas un opérateur à noyau sur  $L_{[0,\delta]}^2$  de noyau la fonction de covariance  $C(s, t)$ . Les éléments propres de  $R_{X_0}$  sont notés  $(\lambda_j, \Phi_j)$  et vérifient :

$$\lambda_j \Phi_j(t) = \int_0^\delta C(s, t) \Phi_j(s) ds, \quad t \in [0, \delta], \quad j = 1, 2, \dots$$

On suppose que les éléments propres de  $R_{X_0}$  sont connus et que pour tout  $j \geq 1$ , les  $\lambda_j$  sont strictement positifs et  $\|\Phi_j\|_{L^2} = 1$  (cf. remarque 1 ci-dessous). Le résultat suivant établit une loi limite dans  $\mathbb{R}^k$  pour un vecteur aléatoire associé à un vecteur de  $\hat{g}_{T,i}$  construit avec des mesures  $\mu_i$  absolument continues par rapport à la mesure de Lebesgue de densité  $\Phi_i$  pour  $i = 1, \dots, k$ .

**THÉORÈME 2.** – *Supposons que le processus  $Z$  admet une représentation autorégressive  $ARC(p)$  de moyenne  $a$  et de bruit blanc  $\varepsilon$  dans  $C_{[0,\delta]}$ , et que les trajectoires de  $Z$  vérifient presque sûrement la condition de Lipschitz :*

$$|Z(t, w) - Z(s, w)| \leq M(w) \cdot |t - s|,$$

*où  $M$  est une variable aléatoire réelle positive vérifiant  $EM^2 < \infty$ . Alors pour tout  $k \in \mathbb{N}^*$ , le vecteur aléatoire*

$$\left( \frac{1}{\sqrt{T}} \sum_{i=0}^{T-1} \int_0^\delta ((Z(i\delta + t) - a(t))) \Phi_j(t) dt; 1 \leq j \leq k \right) \Rightarrow \mathcal{N}_k$$

*quand  $T$  tend vers l'infini où  $\mathcal{N}_k$  est une variable aléatoire gaussienne centrée dans  $\mathbb{R}^k$  de matrice de covariance diagonale :*

$$C_{\mathcal{N}_k} = \text{diag}\{\lambda_1, \dots, \lambda_k\}.$$

Pour estimer la période  $\delta$  nous utilisons la méthode de la « distance minimale » introduite dans [8]. Nous proposons un estimateur de  $\delta$  défini par :

$$\hat{\delta}_T = \text{Arg} \min_{\delta \in ]0, \Delta[} |\hat{g}_T - g(\delta)|, \quad (4)$$

où  $\hat{g}_T$  et  $g(\delta)$  sont définis respectivement par (3) et (4) et où  $\mu$  est la mesure de Lebesgue et  $\Delta$  est un nombre strictement positif donné.

Nous posons  $\Theta := ]0, \Delta[$  un intervalle ouvert de  $\mathbb{R}$  et pour tout  $\beta > 0$

$$h(\beta) := \inf_{\delta_0 \in \Theta} \inf_{|\delta - \delta_0| > \beta > 0} |g(\delta) - g(\delta_0)|. \quad (5)$$

Nous imposons par la suite la condition  $h(\beta) > 0$  qui est une condition d'identifiabilité du paramètre. Nous notons  $\delta_0$  la vraie valeur du paramètre. Le résultat suivant établit la convergence uniforme de l'estimateur :

**THÉORÈME 3.** – *Sous les conditions du théorème 1 et si tout  $\beta > 0 : h(\beta) > 0$ , alors nous avons :*

$$P_{\delta_0}(|\hat{\delta}_T - \delta_0| > \beta) \rightarrow 0 \quad \text{quand } T \rightarrow \infty.$$

La convergence est uniforme sur  $\Theta$ .

Le résultat suivant fournit une loi limite de l'estimateur  $\hat{\delta}_T$ .

**THÉORÈME 4.** – *Sous les conditions du théorème 2 et si tout  $\beta > 0 : h(\beta) > 0$ , alors :*

$$\sqrt{T}(\hat{\delta}_T - \delta_0) \Rightarrow \mathcal{N} \quad \text{quand } T \rightarrow \infty.$$

où  $\mathcal{N}$  est une v.a. gaussienne centrée de variance  $\delta_0^{-2} I^{-2}(\delta_0) \lambda_1$  et  $I(\cdot)$  est la dérivée de la fonction  $g(\cdot)$ . La convergence est uniforme sur  $\Theta$ .

*Remarque 1.* – Une étude similaire peut être menée en supposant que les éléments propres  $(\lambda_j, \Phi_j)$  de l'opérateur de covariance  $R_{X_0}$  sont inconnus. Des estimateurs naturels des éléments propres de  $R_{X_0}$  sont les éléments propres de l'opérateur de covariance empirique (cf. [9], chap. V).

### Références bibliographiques

- [1] Antoniadis A., Parametric estimation for the mean of a Gaussian process, J. Multivariate Anal. 26 (1988) 1–15.
- [2] Billingsley P., Convergence of Probability Measures, Willey, New York, 1968.
- [3] Bosq D., Linear Processes in Function Spaces Theory and Applications, Lecture Notes in Statistics, Vol. 149, Springer, 2000.
- [4] Denford N., Schwartz J.T., Linear Operators, Willey, New York, 1958.
- [5] Hoffman-Jorgensen J., Pisier G., The law of large numbers and the central theorem in Banach spaces, Ann. Probab. 4 (1976) 587–599.
- [6] Kutoyants Yu., Minimum distance parameter estimation for diffusion type observation, C. R. Acad. Sci. Paris, Série I (1991) 637–642.
- [7] Martin D.E.K., Estimation of the minimal period of periodically correlated sequences, Ph.D. Thesis, University of Maryland, 1982.
- [8] Millar P.W., A general approach to the optimality of minimum distance estimation, Trans. Amer. Math. Soc. 286 (1) (1984) 377–418.
- [9] Mourid T., Contribution à la statistique des processus autorégressifs à temps continu, Thèse d'État, Université Paris-6, 1995.
- [10] Parr C., Minimum distance estimation. Bibliography, Comm. Statist. Theory Meth. A 10 (12) (1981) 1205–1224.

# ANNALES DE L'I.S.U.P.

Publications de l'Institut de Statistique de l'Université de Paris

## Sommaire

J.-P. FABRE – P. JACOB

Convergence uniforme presque sûre de l'estimateur à noyau  
de la densité de l'intensité d'un processus ponctuel ..... 3

J. G. DE GOOIJER – A. GANNOUN – I. LARRAMENDY

Nonparametric regression with serially correlated errors ..... 17

L. GARDES

Estimating the support of a poisson process via the  
Faber-Schauder-basis and extreme values ..... 43

A. TOMA

Influence functions of eigenvalues and eigenvectors  
of a covariance matrix ..... 73

T. MOURID – W. BENYELLES

Estimation de la période de la représentation autorégressive  
d'un processus à temps continu ..... 89

*Pub. Inst. Stat. Univ. Paris*  
 XXXXVI, fasc. 1-2, 2002, 89 à 101

## Estimation de la période de la représentation autorégressive d'un processus à temps continu.

**Tahar MOURID et Wafaa BENYELLES** Département De Mathématiques  
 Faculté des Sciences Université Abou Bekr Belkaid TLEMCEM 13000 AL-  
 GERIE.

**Résumé.** Nous considérons l'estimation de la période d'un processus aléatoire à représentation autorégressive. Nous étudions un estimateur de la période de la fonction moyenne basé sur une méthode de la "distance minimale". Sous des conditions de régularité, nous montrons sa convergence en probabilité et nous établissons une loi limite.

Key words and phrases: Processus autoregressifs Banachiques- estimation de la période- distance minimale.

AMS Classification: 62J05 62M10 46E15

**1. Introduction.** Les variables aléatoires à valeurs dans des espaces fonctionnels sont considérées dans de nombreux domaines de la théorie des probabilités et de la statistique des processus aléatoires. L'interprétation d'un processus aléatoire comme une variable aléatoire à valeurs dans un espace de fonctions est très utile en théorie limite et inférence statistique des processus [3]. Dans ce travail nous considérons la classe des processus à temps continu à représentation autorégressive et nous intéressons en particulier au problème de l'estimation de la période de la représentation. La classe des processus autorégressifs à valeurs dans un espace de dimension infinie a trouvé récemment un grand intérêt aussi bien en théorie que dans les applications ([3] et les références citées, [9]). Par ailleurs nous trouvons (cf. [3] chap. 3 et [9] chap. 3) des exemples de processus à temps continu admettant une représentation autorégressive dans des espaces de fonctions appropriés.

Soit  $(\Omega, \mathcal{A}, P)$  un espace probabilisé complet et  $(B, \|\cdot\|)$  un espace de Banach réel séparable muni de sa tribu Borelienne  $\mathcal{B}$ . Nous notons  $\mathcal{L}(B)$  l'algèbre de Banach des opérateurs linéaires bornés définis sur  $B$  et à valeurs dans  $B$ , munie de la norme usuelle  $\|\cdot\|_L$  des opérateurs bornés. Un bruit blanc banachique est une suite de variables aléatoires  $(\varepsilon_n, n \in \mathbb{Z})$  indépendantes identiquement distribuées définies sur  $(\Omega, \mathcal{A}, P)$  à valeurs dans  $B$  vérifiant

pour tout  $n \in \mathbb{Z}$

$$E\varepsilon_n = 0 \quad ; \quad 0 < E\|\varepsilon_n\|^2 = \sigma_\varepsilon^2 < \infty$$

Une suite  $(X_n, n \in \mathbb{Z})$  définie sur  $(\Omega, A, P)$  et à valeurs dans  $B$  est un processus autorégressif banachique d'ordre  $p$  ( $ARB(p)$ )  $p \in \mathbb{N}^*$ , s'il existe des opérateurs  $\rho_1, \rho_2, \dots, \rho_p$  dans  $\mathcal{L}(B)$ , un bruit blanc banachique  $(\varepsilon_n)$  et un élément  $a \in B$  tels que :

$$X_n - a = \rho_1(X_{n-1} - a) + \rho_2(X_{n-2} - a) + \dots + \rho_p(X_{n-p} - a) + \varepsilon_n \quad p.s. \quad (1)$$

où l'opérateur  $\rho_p$  est différent de l'opérateur nul. Sous une condition sur les opérateurs  $\rho_i$  l'équation (1) admet une solution strictement stationnaire dans  $B$  ([9] chap.1 Prop. 3.1). Nous dirons qu'un processus à temps continu  $\mathcal{Z} = (\mathcal{Z}(t), t \in \mathbb{R})$  admet une représentation autorégressive banachique d'ordre  $p$  dans un espace de Banach de fonctions noté  $B$ , s'il existe des opérateurs linéaires bornés  $\rho_1, \rho_2, \dots, \rho_p$  sur  $B$ ,  $a \in B$ , une suite  $(\varepsilon_n, n \in \mathbb{Z})$  de variables aléatoires indépendantes identiquement distribuées définies sur  $(\Omega, A, P)$  à valeurs dans  $B$  et un nombre  $\delta$  strictement positif tels que la suite des variables aléatoires à valeurs dans  $B$  définies par :

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

vérifient la relation (1). Dans de nombreuses applications, le processus  $\mathcal{Z} = (\mathcal{Z}_n, n \in \mathbb{Z})$  est *p.s.* à trajectoires continues et l'espace de Banach  $B$  naturel est l'espace des fonctions définies et continues sur  $[0, \delta]$  noté  $C = C_{[0, \delta]}$  ou encore les espaces usuels  $L^p_{[0, \delta]}$ ,  $p \geq 1$ . De plus si le processus  $\mathcal{Z}$  admet une représentation autorégressive  $ARB(p)$ , alors sa fonction moyenne est une fonction périodique de période  $\delta$  et coïncide avec  $a$  sur l'intervalle  $[0, \delta]$  ([9] p.64). Un problème naturel dans la statistique de cette classe de processus est l'estimation de la fonction moyenne et en particulier l'estimation de sa période  $\delta$ . Dans [7] l'estimation de la période a été résolue par une méthode de Tian qui repose sur des propriétés de la fonction covariance pour la classe des processus périodiquement corrélés. Cette classe est caractérisée par le fait que la fonction moyenne et la fonction covariance sont des fonctions périodiques et de même période. Nous indiquerons pour le cas de processus  $\mathcal{Z}$  à représentation autorégressive Banach que cette méthode n'est pas applicable. Nous proposons un estimateur de la période  $\delta$  du type

"distance minimale" ([8] [6] [10]). La méthode est basée sur l'étude d'un estimateur d'une intégrale de la fonction moyenne du processus (convergence presque sûre, dans  $L^2$  et une loi limite) puis nous déduisons un estimateur de la période par la minimisation de l'écart absolue entre l'estimateur étudié et la fonction à estimer. Nous établissons par la suite la convergence en probabilité et une loi limite pour l'estimateur de la distance minimale étudié. Les résultats sont obtenus uniformément sur l'espace paramétrique.

Le plan du papier est comme suit. La section 2 contient une étude d'un estimateur d'une intégrale de la fonction moyenne. La section 3 est consacré à l'estimation de la période. La section 4 contient les preuves des résultats.

**2. Estimateur d'une intégrale de la fonction moyenne.** Soit  $\mathcal{Z} = (\mathcal{Z}(t), t \in \mathbf{R})$  un processus réel presque sûrement à trajectoires continues. On suppose que le processus  $\mathcal{Z}$  admet une représentation autorégressive  $ARC(p)$  de moyenne  $a$  et de bruit blanc  $\varepsilon$  dans  $C_{[0,\delta]}$  où  $\delta$  un réel strictement positif. Pour estimer la période  $\delta$  nous utilisons un estimateur qui a été introduit et étudié dans [3] [9]. Posons pour  $\delta > 0$  :

$$g(\delta) := \frac{1}{\delta} \int_0^{\delta} a(t) d\mu(t)$$

où  $\mu$  est une mesure à signe bornée dans  $C_{[0,\delta]}^*$  (le dual topologique de  $C_{[0,\delta]}$ ). Si nous observons une trajectoire du processus  $\mathcal{Z}$  sur  $[0, T]$  :  $(\mathcal{Z}(t), t \in [0, T])$ , alors un estimateur naturel de la fonction  $g$  est le suivant :

$$\hat{g}_T = \frac{1}{T} \int_0^T \mathcal{Z}(t) d\mu_{[T]}(t) \quad (3)$$

où  $[T]$  est la partie entière de  $T$  et pour tout  $n \in \mathbf{Z}$ , et pour tout borélien  $A$ ,  $A \in B_{[0,n\delta]}$ , la mesure  $\mu_n$  est définie par (translatée de  $\mu$ ):

$$\mu_n(A) := \sum_{j=0}^{n-1} \mu(A \cap [j\delta, (j+1)\delta] - j\delta)$$

L'étude de la convergence et loi limite de cet estimateur a été faite pour  $p = 1$  et  $\delta = 1$  dans [3] Chap. 6.7. Dans la première partie de travail nous étudions la convergence presque sûre et dans  $L^2$  de l'estimateur  $\hat{g}_T$  ensuite nous établissons une loi limite dans  $\mathbf{R}^k$  du vecteur aléatoire  $(\hat{g}_{T,i}, i = 1, \dots, k)$

associé à des mesures  $\mu_i$ ,  $i = 1, \dots, k$ , pour  $p \geq 1$  et  $\delta > 0$ . Le résultat suivant fournit la convergence presque sûre et dans  $L^2$  de  $\hat{g}_T$ :

**Théorème 1.** *Si le processus  $\mathcal{Z}$  admet une représentation autorégressive ARC(p) de moyenne  $a$  et de bruit blanc  $\varepsilon$  dans  $C_{[0,\delta]}$  où  $\delta > 0$ , alors*

$$\lim_{T \rightarrow \infty} \hat{g}_T = g(\delta)$$

*presque sûrement et dans  $L^2$ .*

La suite ARC(p) :  $X = (X_n)$  associée au processus  $\mathcal{Z}$  par (2) admet un opérateur de covariance  $R_{X_0}$  qui est dans ce cas un opérateur à noyau sur  $L^2_{[0,\delta]}$  de noyau la fonction de covariance  $C(s, t)$ . Les éléments propres de  $R_{X_0}$  sont notés  $(\lambda_j, \Phi_j)$  et vérifient :

$$\lambda_j \Phi_j(t) = \int_0^\delta C(s, t) \Phi_j(s) ds, \quad t \in [0, \delta], \quad j = 1, 2, \dots$$

On suppose que les éléments propres de  $R_{X_0}$  sont connus et que pour tout  $j \geq 1$  les  $\lambda_j$  sont strictement positifs et  $\|\Phi_j\|_{L^2} = 1$  (cf. Remarque 1 ci dessous). Le résultat suivant établit une loi limite dans  $\mathbb{R}^k$  pour un vecteur aléatoire associé à un vecteur de  $\hat{g}_T$ , construit avec des mesures  $\mu_i$  absolument continue par rapport à la mesure de Lebesgue de densité  $\Phi_i$  pour  $i = 1, \dots, k$ .

**Théorème 2.** *Supposons que le processus  $\mathcal{Z}$  admet une représentation autorégressive ARC(p) de moyenne  $a$  et de bruit blanc  $\varepsilon$  dans  $C_{[0,\delta]}$ , et que les trajectoires de  $\mathcal{Z}$  vérifie presque sûrement la condition de Lipschitz suivante:*

$$|\mathcal{Z}(t, \omega) - \mathcal{Z}(s, \omega)| \leq M(\omega) \cdot |t - s|$$

*où  $M$  est une variable aléatoire réelle positive vérifiant  $EM^2 < \infty$ .*

*Alors pour tout  $k \in \mathbb{N}^*$ , le vecteur aléatoire*

$$\left( \frac{1}{\sqrt{T}} \sum_{i=0}^{T-1} \int_0^\delta ((\mathcal{Z}(i\delta + t) - a(t))) \Phi_j(t) dt; 1 \leq j \leq k \right) \implies \mathcal{N}_k$$



quand  $T$  tend vers l'infini où  $N_k$  est une variable aléatoire Gaussienne centrée dans  $\mathbf{R}^k$  de matrice de covariance diagonale :

$$C_{N_k} = \text{diag} \{ \lambda_1, \dots, \lambda_k \}$$

**3. Estimation de la période.** Rappelons qu'un processus stochastique  $(Z(t), t \in \mathbb{T})$  réel défini sur un espace de probabilité  $(\Omega, \mathcal{A}, P)$  et ayant un moment d'ordre deux fini est appelé processus périodiquement corrélé si la fonction de covariance  $R$  et la fonction moyenne  $m$  sont périodiques de période  $\delta$  :  $R(u, v) = E[Z(u) \cdot Z(v)] = R(u + \delta, v + \delta)$  et  $m(u) = E[Z(u)] = m(u + \delta)$ ,  $\forall (u, v) \in \mathbb{Z} \times \mathbb{Z}$  tel que  $(u + \delta, v + \delta) \in \mathbb{Z} \times \mathbb{Z}$ . Dans la suite, cette classe est noté PC. Dans [7] une synthèse de la méthode de Tian pour l'estimation de la période  $\delta$  pour un processus PC est proposée. La méthode est basée sur des propriétés de la fonction de covariance qui est un moyen usuel de recherche de périodicité en théorie du signal. Nous examinons l'application de la méthode de Tian aux processus  $ARB(p)$ . Comme nous l'avons signalé en introduction un processus à temps continu  $(Z(t), t \in \mathbf{R})$  à représentation autorégressive banachique possède une moyenne périodique de période  $\delta$ . Par contre la fonction de covariance ne l'est pas. En effet si  $(Z(t))$  est à représentation  $ARC(1)$  dans  $C_{[0, \delta]}$  de moyenne  $a$  et de bruit blanc  $\varepsilon = (\varepsilon_n)$ , un calcul montre que pour tout  $t \in [0, \delta]$  :  $C(t + \delta, t + \delta) = E[Z(t + \delta) - a(t)][Z(t + \delta) - a(t)] = E[\rho Z(t) + \varepsilon_1(t) - a(t)][Z(t + \delta) - a(t)] = C(t, t) + c$  où  $c$  est strictement positif. Donc la fonction de covariance  $C(s, t)$  n'est pas périodique. Par suite les processus à représentation  $ARB(p)$  ne sont pas en général élément de la classe PC et par conséquent la méthode de Tian ne s'applique pas ici. En reprenant les notations de la Section 2 nous proposons un estimateur de la période  $\delta$  basé sur la méthode de la "distance minimale" ([8]). L'estimateur de  $\delta$  est défini par:

$$\hat{\delta}_T = \text{Arg} \min_{\delta \in ]0, \Delta[} |\hat{g}_T - g(\delta)| \quad (4)$$

où  $\hat{g}_T$  et  $g(\delta)$  sont définis respectivement par (3) et (4) et où  $\mu$  est la mesure de Lebesgue et  $\Delta$  est un nombre strictement positif donné.

Nous posons  $\Theta := ]0, \Delta[$  un intervalle ouvert de  $\mathbf{R}$  et pour tout  $\beta > 0$

$$h(\beta) := \inf_{\delta_0 \in \Theta} \inf_{|\delta - \delta_0| > \beta > 0} |g(\delta) - g(\delta_0)| \quad (5)$$

Nous imposons par la suite la condition :  $h(\beta) > 0$  qui est une condition d'identifiabilité du paramètre. Cette condition est vérifiée par exemple dans

le cas où  $\mu$  est la mesure de Lebesgue est la fonction moyenne  $a$  garde un signe constant. Nous notons  $\delta_0$  la vraie valeur du paramètre. Le résultat suivant établit la convergence en probabilité de l'estimateur uniforme sur  $\Theta$ .

**Théorème 3.** *Sous les conditions du Théorème 1 et si tout  $\beta > 0$  :  $h(\beta) > 0$ , alors nous avons :*

$$P_{\delta_0} \left( \left| \hat{\delta}_T - \delta_0 \right| > \beta \right) \rightarrow 0 \quad \text{quand } T \rightarrow \infty$$

*La convergence est uniforme sur  $\Theta$ .*

Le résultat suivant fournit une loi limite de l'estimateur  $\hat{\delta}_T$ .

**Théorème 4.** *Sous les conditions du Théorème 2 et si tout  $\beta > 0$  :  $h(\beta) > 0$ , alors:*

$$\sqrt{T} (\hat{\delta}_T - \delta_0) \Rightarrow \mathcal{N} \quad \text{quand } T \rightarrow \infty$$

où  $\mathcal{N}$  est une v.a. Gaussienne centrée de variance  $\delta_0^{-2} I^{-2}(\delta_0) \lambda_1$  et  $I(\cdot)$  est la dérivée de la fonction  $g(\cdot)$ . La convergence est uniforme sur  $\Theta$ .

**Remarque 1.** *Une étude similaire peut être menée en supposant cette fois que les éléments propres  $(\lambda_j, \Phi_j)$  de l'opérateur de covariance  $R_{X_0}$  sont inconnus. Des estimateurs naturels des éléments propres de  $R_{X_0}$  sont les éléments propres de l'opérateur de covariance empirique associé (cf. [9] Chap. V).*

**4. Preuves. Preuve du Théorème 1.** Pour tout  $T \in \mathbb{R}$  et  $\delta > 0$ , il existe un entier  $n = n_{T,\delta}$  tel que :  $n\delta \leq T \leq (n+1)\delta$ . Par suite nous avons  $\frac{n}{T} \rightarrow \frac{1}{\delta}$ . Considérons la décomposition suivante :

$$\begin{aligned} \frac{1}{T} \int_0^T \mathcal{Z}(t) d\mu_{[T]}(t) &= \frac{1}{T} \int_0^{n\delta} \mathcal{Z}(t) d\mu_{[T]}(t) + \frac{1}{T} \int_{n\delta}^T \mathcal{Z}(t) d\mu_{[T]}(t) \\ &=: A_n + B_n \end{aligned}$$

D'une part, de  $X_j(t) := \mathcal{Z}(j\delta + t)$ ,  $t \in [0, \delta]$  et de la définition de  $\mu_{[T]}$ , nous obtenons :

$$\begin{aligned} A_n &:= \frac{1}{T} \int_0^{n\delta} \mathcal{Z}(t) d\mu_{[T]}(t) \\ &= \frac{1}{T} \sum_{j=0}^{n-1} \int_{j\delta}^{(j+1)\delta} \mathcal{Z}(t) d\mu_{[T]}(t) \end{aligned}$$

$$\begin{aligned}
&= \frac{1}{T} \sum_{j=0}^{n-1} \int_0^\delta \mathcal{Z}(\delta j + t) d\mu(t) \\
&= \frac{n}{T} \int_0^\delta \left[ \frac{1}{n} \sum_{j=0}^{n-1} X_j(t) \right] d\mu(t) \tag{6}
\end{aligned}$$

D'après le Th. 1.1 Chap. II de [9], appliqué à l'espace  $(B, \|\cdot\|) = (C_{[0,\delta]}, \|\cdot\|_\infty)$  nous avons :

$$\begin{aligned}
\left\| \frac{1}{n} \sum_{j=0}^{n-1} X_j - a \right\|_\infty &:= \sup_{0 \leq t \leq \delta} \left| \frac{1}{n} \sum_{j=0}^{n-1} X_j(t) - a(t) \right| \\
&= \|\bar{X}_n - a\|_\infty \rightarrow 0 \quad p.s.
\end{aligned}$$

quand  $n \rightarrow \infty$  et où  $\bar{X}_n = \frac{1}{n} \sum_{j=0}^{n-1} X_j$ .

Or presque sûrement les trajectoires du processus  $\mathcal{Z}$  sont continues, donc :

$$\begin{aligned}
\left\| \frac{1}{n} \sum_{j=0}^{n-1} X_j \right\|_\infty &\leq \sup_{0 \leq j \leq (n-1)} \|X_j\|_\infty \\
&\leq \sup_{0 \leq j \leq (n-1)} \sup_{0 \leq t \leq \delta} |X_j(t)| \\
&= \sup_{0 \leq j \leq (n-1)} \sup_{0 \leq t \leq \delta} |\mathcal{Z}(\delta j + t)| < \infty \quad p.s.
\end{aligned}$$

Par conséquent, par le théorème de la convergence dominée et du fait que  $\frac{n}{T} \rightarrow \frac{1}{\delta}$ , nous déduisons de (6) que quand  $n \rightarrow \infty$  :

$$A_n := \frac{1}{T} \int_0^{n\delta} \mathcal{Z}(t) d\mu_{[T]}(t) \rightarrow \frac{1}{\delta} \int_0^\delta a(t) d\mu(t) \quad p.s.$$

D'autre part,

$$\begin{aligned}
B_n &:= \frac{1}{T} \int_{n\delta}^T \mathcal{Z}(t) d\mu_{[T]}(t) \\
&\leq \frac{1}{T} \sup_{n\delta \leq t \leq T} |\mathcal{Z}(t)| \cdot \|\mu\| = \frac{1}{T} \sup_{0 \leq s \leq \delta} |X_n(s)| \cdot \|\mu\| \\
&\leq \frac{1}{T} \|\mu\| \|X_n\| \tag{7}
\end{aligned}$$

où  $\|\mu\|$  désigne la variation totale de  $\mu$ . Par l'inégalité de Markov et la stationarité de  $(X_n)$  nous avons :

$$\begin{aligned} P\left(\left\|\frac{X_n}{T}\right\| > \varepsilon\right) &\leq \frac{1}{T^2\varepsilon^2} E \|X_n\|^2 \\ &= \frac{1}{T^2\varepsilon^2} E \|X_0\|^2 \end{aligned}$$

Par suite

$$\int_1^\infty P\left(\left\|\frac{X_n}{T}\right\| > \varepsilon\right) dT \leq \frac{E \|X_0\|^2}{\varepsilon^2} \int_1^\infty \frac{1}{T^2} dT < \infty$$

Par le lemme de Borel-Cantelli, on obtient donc quand  $n \rightarrow \infty$  :

$$\frac{X_n}{T} \rightarrow 0 \quad p.s.$$

Par conséquent de (7), nous obtenons :

$$B_n := \frac{1}{T} \int_{n\delta}^T \mathcal{Z}(t) d\mu_{[T]}(t) \rightarrow 0 \quad p.s.$$

quand  $T$  tend vers l'infini.

D'où en regroupant les convergence de  $A_n$  et  $B_n$  vers 0, nous obtenons le résultat demandé :

$$\lim_{T \rightarrow \infty} \frac{1}{T} \int_0^T \mathcal{Z}(t) d\mu_{[T]}(t) = \frac{1}{\delta} \int_0^\delta a(t) d\mu(t) \quad p.s.$$

De manière similaire nous obtenons la convergence dans  $L^2$  en utilisant cette fois  $\|\bar{X}_n - a\| \rightarrow 0$  dans  $L^2$  ( cf.[9] Remarque 7.2 Chap. V )

**Preuve du Théorème 2.** Comme  $X_j(t) := \mathcal{Z}(j\delta + t)$  et la suite  $(X_n, n \in \mathbb{Z})$  est un processus  $ARC(p)$  d'innovation  $(\varepsilon_n, n \in \mathbb{Z})$ , alors:

$$\varepsilon_n(t) - \varepsilon_n(s) = \mathcal{Z}(n\delta + t) - \mathcal{Z}(n\delta + s) - E(\mathcal{Z}(n\delta + t) - \mathcal{Z}(n\delta + s) / \mathcal{Z}(u), u \leq n\delta)$$

Par conséquent, de l'hypothèse sur le processus  $\mathcal{Z}$ , on a:

$$\begin{aligned} |\varepsilon_n(t) - \varepsilon_n(s)| &\leq |\mathcal{Z}(n\delta + t) - \mathcal{Z}(n\delta + s)| \\ &\quad + E(|\mathcal{Z}(n\delta + t) - \mathcal{Z}(n\delta + s)| / \mathcal{Z}(u), u \leq n\delta) \\ &\leq (M + E(M/\mathcal{Z}(u), u \leq n\delta)) \cdot |t - s| \end{aligned}$$

Donc la suite  $(\varepsilon_n, n \in \mathbf{Z})$  vérifie la condition d) du Corollaire 5.11. Chap. II dans [9]. Par conséquent par le même Corollaire la suite  $ARB(p)$   $(X_n, n \in \mathbf{Z})$  vérifie le théorème central limite dans  $B = C_{[0,\delta]}$  :

$$\sqrt{n}(\bar{X}_n - a) \Longrightarrow \mathcal{N}$$

où  $\mathcal{N}$  est une Gaussienne centrée dans  $C_{[0,\delta]}$  d'opérateur de covariance  $(I - \rho_1 - \dots - \rho_p)^{-1} R_{\varepsilon_0} (I - \rho_1 - \dots - \rho_p)^{-1*}$ .

Considérons la forme linéaire  $x^* := \mu \in C_{[0,\delta]}^*$  le dual topologique de  $C_{[0,\delta]}$ . Par suite par le théorème de continuité, nous avons le TCL dans  $\mathbf{R}$  suivant :

$$\mu(\sqrt{n}(\bar{X}_n - a)) \Longrightarrow \mu(\mathcal{N})$$

D'autre part,

$$\begin{aligned} \mu(\sqrt{n}(\bar{X}_n - a)) &= \int_0^\delta \sqrt{n}(\bar{X}_n(t) - a(t)) d\mu(t) \\ &= \sqrt{n} \int_0^\delta \left( \frac{1}{n} \sum_{j=0}^{n-1} X_j(t) - a(t) \right) d\mu(t) \\ &= \int_0^\delta \left( \frac{1}{\sqrt{n}} \sum_{j=0}^{n-1} X_j(t) - \sqrt{n}a(t) \right) d\mu(t) \\ &= \int_0^\delta \left( \frac{1}{\sqrt{n}} \sum_{j=0}^{n-1} X_j(t) - \frac{n}{\sqrt{n}}a(t) \right) d\mu(t) \\ &= \frac{1}{\sqrt{n}} \sum_{j=0}^{n-1} \int_0^\delta (X_j(t) - a(t)) d\mu(t) \end{aligned}$$

Donc

$$\frac{1}{\sqrt{n}} \sum_{j=0}^{n-1} \int_0^\delta (X_j(t) - a(t)) d\mu(t) \Longrightarrow \mu(\mathcal{N}) = \int_0^\delta \mathcal{N}(t) d\mu(t)$$

où  $\mu(\mathcal{N})$  est une variable aléatoire réelle Gaussienne centrée de variance  $\mu\left((I - \rho_1 - \dots - \rho_p)^{-1} R_{\varepsilon_0} (I - \rho_1 - \dots - \rho_p)^{-1*}(\mu)\right)$ . Considérons maintenant  $\mu_1, \dots, \mu_k$ ,  $k$  formes linéaires dans  $C_{[0,\delta]}^*$ ,  $k \geq 1$ , de densités respectives  $\Phi_1, \dots, \Phi_k$

par rapport à la mesure de Lebesgue sur  $[0, \delta]$ . D'après ce qui précède, nous aurons pour tout  $i = 1, \dots, k$ :

$$\mu_i(\sqrt{n}(\bar{X}_n - a)) = \frac{1}{\sqrt{n}} \sum_{j=0}^{n-1} \int_0^\delta (X_j(t) - a(t)) \cdot \Phi_i(t) dt \implies \mu_i(\mathcal{N}) = \int_0^\delta \mathcal{N}(t) \cdot \Phi_i(t) dt$$

Notons  $\nu_n$  et  $\nu$  les lois respectives de  $\sqrt{n}(\bar{X}_n - a)$  et de  $\mathcal{N}$ . Par la convergence faible de  $\nu_n$  vers  $\nu$  et le théorème de continuité, nous déduisons la convergence des lois de dimensions finies dans  $\mathbf{R}^k$  :  $\forall \mu_1, \dots, \mu_k \in C_{[0, \delta]}^*$ ,

$$(\mu_1(\sqrt{n}(\bar{X}_n - a)), \dots, \mu_k(\sqrt{n}(\bar{X}_n - a))) \implies (\mu_1(\mathcal{N}), \dots, \mu_k(\mathcal{N}))$$

où le vecteur  $(\mu_1(\mathcal{N}), \dots, \mu_k(\mathcal{N}))$  est Gaussien centré dans  $\mathbf{R}^k$ .

Calculons la matrice de covariance de la loi limite. On a pour  $s, t \in [0, \delta]$  :

$$\begin{aligned} C(s, t) &:= E(\mathcal{Z}(s) - a(s)) \cdot (\mathcal{Z}(t) - a(t)) \\ &= E(X_0(s) - a(s)) \cdot (X_0(t) - a(t)) \\ &= E(\mathcal{N}(s)\mathcal{N}(t)) \end{aligned}$$

Donc, pour  $i, j = 1, \dots, k$ , on a :

$$\begin{aligned} E\mu_i(\mathcal{N})\mu_j(\mathcal{N}) &= E\left(\int_0^\delta \mathcal{N}(t) d\mu_i(t) \int_0^\delta \mathcal{N}(t) d\mu_j(t)\right) \\ &= \int_0^\delta \int_0^\delta E\mathcal{N}(s)\mathcal{N}(t) d\mu_i(s) d\mu_j(t) \\ &= \int_0^\delta \left(\int_0^\delta C(s, t) \Phi_i(s) ds\right) \Phi_j(t) dt \\ &= \int_0^\delta \lambda_i \Phi_i(t) \Phi_j(t) dt \\ &= \lambda_i \delta_{ij} \end{aligned}$$

où  $\delta_{ij}$  est le symbole de Kronecker. D'où le théorème.

**Preuve du Théorème 3.** Soit  $\beta > 0$ . De la définition de  $\hat{\delta}_T$  (cf. (4)) et pour  $\delta_0 \in \Theta$  nous avons :

$$|\hat{\delta}_T - \delta_0| > \beta \implies \inf_{|\delta - \delta_0| \leq \beta} |\hat{g}_T - g(\delta)| > \inf_{|\delta - \delta_0| > \beta} |\hat{g}_T - g(\delta)| \quad (8)$$

D'une part

$$|\hat{g}_T - g(\delta)| \leq |\hat{g}_T - g(\delta_0)| + |g(\delta_0) - g(\delta)|$$

Donc

$$\inf_{|\delta - \delta_0| \leq \beta} |\hat{g}_T - g(\delta)| \leq |\hat{g}_T - g(\delta_0)| \quad (9)$$

D'autre part

$$|g(\delta) - g(\delta_0)| \leq |\hat{g}_T - g(\delta)| + |\hat{g}_T - g(\delta_0)|$$

Par suite

$$|\hat{g}_T - g(\delta)| \geq |g(\delta) - g(\delta_0)| - |\hat{g}_T - g(\delta_0)|$$

Et alors de la définition de  $h(\beta)$  (cf. (5) ) nous obtenons :

$$\begin{aligned} \inf_{|\delta - \delta_0| > \beta} |\hat{g}_T - g(\delta)| &\geq \inf_{|\delta - \delta_0| > \beta} |g(\delta) - g(\delta_0)| - |\hat{g}_T - g(\delta_0)| \\ &\geq h(\beta) - |\hat{g}_T - g(\delta_0)| \end{aligned} \quad (10)$$

Par conséquent de (8) (9) (10) nous en déduisons que :

$$\begin{aligned} P_{\delta_0} \left( |\hat{\delta}_T - \delta_0| > \beta \right) &\leq P_{\delta_0} \left( \inf_{|\delta - \delta_0| \leq \beta} |\hat{g}_T - g(\delta)| > \inf_{|\delta - \delta_0| > \beta} |\hat{g}_T - g(\delta)| \right) \\ &\leq P_{\delta_0} \left( \inf_{|\delta - \delta_0| > \beta} |\hat{g}_T - g(\delta)| > h(\beta) - |\hat{g}_T - g(\delta_0)| \right) \\ &\leq P_{\delta_0} \left( |\hat{g}_T - g(\delta_0)| > \frac{1}{2} h(\beta) \right) \end{aligned}$$

Par hypothèse  $h(\beta) > 0$  et par le Théorème 1 :  $\hat{g}_T \rightarrow g(\delta)$  presque sûrement. Donc le dernier majorant tend vers 0 quand  $N \rightarrow \infty$ . D'où le résultat.

**Preuve du Théorème 4.** Pour obtenir la loi limite de l'estimateur nous utilisons les arguments généraux de Millar dans [8] ( Sect. XV ). Nous avons

$$\sqrt{T} (\hat{\delta}_T - \delta_0) = I^{-1}(\delta_0) \cdot \sqrt{T} (\hat{g}_T - g(\delta_0)) + o_p(1) \quad (11)$$

où  $I(\cdot)$  est la dérivée de  $g(\cdot)$  par rapport à  $\delta$ . D'autre part :

$$\begin{aligned} \frac{1}{\sqrt{T}} \sum_{i=0}^{T-1} \int_0^\delta \mathcal{Z}(i\delta + t) dt &= \frac{1}{\sqrt{T}} \left( \int_0^\delta \mathcal{Z}(t) dt + \int_0^\delta \mathcal{Z}(\delta + t) dt + \dots + \int_0^\delta \mathcal{Z}((T-1)\delta + t) dt \right) \\ &= \frac{1}{\sqrt{T}} \int_0^{T\delta} \mathcal{Z}(t) dt \\ &= \sqrt{T} \left( \frac{1}{T} \int_0^{T\delta} \mathcal{Z}(t) dt \right) = \sqrt{T} \left( \frac{1}{T} \int_0^T \mathcal{Z}(t) dt + \frac{1}{T} \int_T^{T\delta} \mathcal{Z}(t) dt \right) \\ &= \sqrt{T} \cdot \hat{g}_T + \frac{1}{\sqrt{T}} \int_T^{T\delta} \mathcal{Z}(t) dt \end{aligned}$$

par définition de  $\hat{g}_T$ .

On montre comme dans le calcul fait pour  $B_n$  dans la preuve de Théorème 1 que :

$$\left| \frac{1}{\sqrt{T}} \int_T^{T\delta} \mathcal{Z}(t) dt \right| \leq \frac{\|X_{[T\delta]}\|}{[T\delta]} \rightarrow 0$$

en probabilité quand  $T \rightarrow \infty$ . Par conséquent

$$\sqrt{T} \hat{g}_T = \frac{1}{\delta \sqrt{T}} \sum_{i=0}^{T-1} \int_0^\delta \mathcal{Z}(i\delta + t) dt + o_p(1)$$

Par suite

$$\begin{aligned} \sqrt{T} (\hat{g}_T - g(\delta)) &= \frac{1}{\delta \sqrt{T}} \sum_{i=0}^{T-1} \int_0^\delta \mathcal{Z}(i\delta + t) dt - \sqrt{T} \cdot \frac{1}{\delta} \int_0^\delta a(t) dt + o_p(1) \\ &= \frac{1}{\delta} \cdot \frac{1}{\sqrt{T}} \left[ \sum_{i=0}^{T-1} \int_0^\delta \mathcal{Z}(i\delta + t) dt - T \int_0^\delta a(t) dt \right] + o_p(1) \\ &= \frac{1}{\delta} \cdot \left[ \frac{1}{\sqrt{T}} \sum_{i=0}^{T-1} \int_0^\delta (\mathcal{Z}(i\delta + t) - a(t)) dt \right] + o_p(1) \end{aligned}$$

Donc en appliquant le Théorème 3 pour  $k = 1$  et en prenant  $\mu_1$  la mesure de Lebesgue nous obtenons :

$$\delta_0 \cdot \sqrt{T} (\hat{g}_T - g(\delta_0)) \implies \mathcal{N}_1$$



où  $\mathcal{N}_1$  est v.a. Gaussienne centrée de variance  $\lambda_1$ . Par suite de la relation (11) nous obtenons :

$$\delta_0 \cdot I(\delta_0) \cdot \sqrt{T} (\hat{\delta}_T - \delta_0) \implies \mathcal{N}_1$$

D'où le Théorème.

*Nous remercions le referee anonyme pour sa lecture attentive du manuscrit et L'Editeur pour ses remarques.*

## References

- [1] A. Antoniadis (1988). Parametric estimation for the mean of a Gaussian process. Jour. Multi. Anal. 26 pp. 1-15.
- [2] P. Billingsley (1968). " Convergence of probability measures. Willey. New-York.
- [3] D. Bosq (2000). Linear Processes in Function spaces. Theory and Applications. Lecture Notes in Statistics 149. Springer.
- [4] N. Denford et J.T.Schwartz (1958). Linear operators, Willey. New-York.
- [5] J. Hoffman-Jorgensen et G.Pisier (1976). The law of large numbers and the central theorem in Banach spaces, Ann.Proba., n<sup>o</sup>4, p. 587-599.
- [6] Yu. Kutoyants (1991). Minimum distance parameter estimation for diffusion type observation. C. R. Acad. Sci. Paris, Ser. I. p. 637-642.
- [7] D.E.K.Martin (1982). Estimation of the minimal period of periodically correlated sequences. Phd. Thesis. University of Maryland 1982.
- [8] P.W. Millar (1984). A general approach to the optimality of minimum distance estimation. Trans.Amer. Soc. 286,1, p. 377-418.
- [9] T. Mourid (1995). Contribution à la statistique des processus autorégressifs à temps continu. Thèse d'Etat, Univ. Paris 6.
- [10] C.Parr (1981). Minimum distance estimation. Bibliography Commun. Statist. Theory Meth. A10(12)1205-1224.



---

THESE DE DOCTORAT

PRESENTEE A

L'UNIVERSITE ABOU BEKR BELKAID TLEMCCEN

POUR L'OBTENTION DU DIPLOME  
DOCTEUR ES SCIENCES MATHÉMATIQUES

Par

Wafaa BENYELLES

Sujet

**Problèmes d'estimation de la période de la représentation  
autorégressive**

**Soutenu en Mai 2013 à Tlemcen devant le jury :**

Mr. GHOUALI Nouredine	Président	Professeur UABB Tlemcen
Mr. PUMO Besnik	Examineur	Professeur Université Angers
Mr. DJEHICHE Boualem	Examineur	Professeur KTH Stockholm
Mr. LAKSACI Ali	Examineur	Professeur Université Sidi Bel Abbes
Mr. DIB Hacem	Examineur	Professeur UABB Tlemcen
Mr. LABBAS Ahmed	Invite	MCA UABB Tlemcen
Mr. MOURID Tahar	Rapporteur	Professeur UABB Tlemcen

# Table des matière

<b>Introduction</b>	<b>1</b>
<b>1 Processus autorégressifs fonctionnels</b>	<b>7</b>
1.1 Introduction	7
1.2 Préliminaires et définitions	8
1.3 Processus d'Ornstein-Uhlenbeck	12
1.4 Construction d'un modèle ARC(1) ( $B = C_{[0,1]}$ ).	14
1.5 Une classe d'ARB(1) à valeurs dans $L^2_{[0,1]}$	16
1.6 Processus avec saisonnalité	16
1.7 Théorèmes limites des ARH(1)	17
1.7.1 Loi forte des grands nombres(LFGN)	17
1.7.2 Loi des grands nombres en moyenne d'ordre $\alpha > 0$	19
1.7.3 Convergence en moyenne d'ordre $r > 0$ , dans un Banach	20
1.7.4 Théorème central limite(TCL)	20
<b>2 Estimation de la période de la représentation AR</b>	<b>23</b>
2.1 Méthode de Tian	23
2.1.1 Les suites périodiquement corrélées	23
2.1.2 Méthode de Tian associée au PPC	24
2.1.3 Les processus à représentation autorégressive	25
2.2 Méthode de la distance minimale	26
2.2.1 Notations -Hypothèses	26
2.3 Convergence presque sûre et $L^2$	29
2.4 Théorème central limite	34
2.5 Estimateur de la période	39
2.6 Loi limite	40
<b>3 Simulations et exemples</b>	<b>43</b>
3.1 Simulations	43
3.1.1 Introduction	43
3.1.2 Simulation du comportement de l'estimateur de la période	51
3.2 Exemples	62

<b>TABLE DES MATIÈRE</b>	<b>3</b>
3.2.1 Série climatologique ENSO. . . . .	62
3.2.2 Prévision de la température à Alger . . . . .	64
<b>4 Conclusion</b>	<b>67</b>

# Introduction

Dans ce travail, nous nous intéressons à quelques problèmes statistiques sur la classe des processus autorégressifs fonctionnels d'ordre 1. Cette classe de processus est bien introduite dans l'ouvrage de référence de D. Bosq [12]: *Linear Processes in Function spaces. Theory and Applications*. Lecture Notes in Statistics, 149. Springer 2000. Nous rappelons qu'un processus réel  $Z = (Z(t), t \in \mathbb{R})$  est un processus autorégressif fonctionnel dans un espace fonctionnel  $B$  si la suite des v.a.  $(X_n, n \in \mathbb{Z})$  définie par :

$$X_n(t) := Z(n\delta + t) \quad t \in [0, \delta], \text{ pour } \delta > 0 \quad (1)$$

vérifie la relation autorégressive suivante dans l'espace fonctionnel  $B$ :

$$X_n - a = \rho(X_{n-1} - a) + \varepsilon_n$$

où  $\rho$  est un opérateur linéaire borné défini sur  $B$ ,  $a \in B$  et  $(\varepsilon_n)$  un bruit blanc dans  $B$  : v.a. i.i.d. du 2<sup>ème</sup> ordre vérifiant pour tout  $n \in \mathbb{Z}$  :

$$E(\varepsilon_n) = 0 \quad ; \quad 0 < E \|\varepsilon_n\|^2 = \sigma_\varepsilon^2 < \infty$$

Sous une condition sur l'opérateur  $\rho$  la suite  $(X_n, n \in \mathbb{Z})$  est strictement stationnaire dans  $B$  et la moyenne  $m(t)$  qui est la moyenne du processus  $Z = (Z(t), t \in \mathbb{R}^+)$ , est une fonction périodique de période  $\delta$ . Dans plusieurs applications, le processus  $Z = (Z(t), t \in \mathbb{R}^+)$  est à trajectoires continues et l'espace fonctionnel  $B$  sera l'espace  $C_{[0, \delta]}$  des fonctions continues sur  $[0, \delta]$ . Cette approche considérée par D. Bosq et T. Mourid dans leurs premiers travaux (cf [13] [12]), a été motivée par des considérations de prédiction de processus à temps continu : en observant une trajectoire d'un processus à temps continu sur  $n$  intervalles successifs de même longueur  $\delta > 0$ , il s'agit de construire un prédicteur de la trajectoire du processus sur l'intervalle  $[n\delta, (n+1)\delta]$ . Cela a conduit de façon naturelle au problème classique de la prédiction fonctionnelle d'un élément fonctionnel  $X_{n+1}$  à partir d'un échantillon fonctionnel  $X_1, \dots, X_n$ . Cette approche

s'est révélée très fructueuse et a donné lieu à de nombreux travaux. Plusieurs auteurs se sont intéressés au cadre fonctionnel des observations en statistique. Nous citons entre autres : Mourier E., Grenander [22] , Rozanov, Antoniadis [3], Antoniadis-Beder, Dauxois-Pousse-Romain. Dans le cadre de processus AR fonctionnel ou modèle de regression fonctionnelle nous pouvons citer les auteurs : Lodges, D. Bosq [12 ], Ramsay et Silverman [31, 32], Ferraty and Vieu [18, 20], Ferraty and Romain [17] Pumo [30], Besse-Cardot [11] Merlevede F. Bensmain-Mourid ([4][5]) Allam-Mourid ( [1][2]) Guillas A. [23], Labbas A.

Nous complétons en citant certaines applications réalisées à l'aide de cette modélisation dans différents domaines:

Consommation d'électricité [14] - Prévion du nombre annuel de voyageurs à la SNCF- Prévion d'un trafic routier[10]- Etude des electrocardiogrammes [9].

Dans les travaux cités plus haut, les problèmes d'estimation et de prévision dans cette classe de processus supposaient que la période d'échantillonnage  $\delta > 0$  est connue. Dans notre travail, nous étudions plus précisément l'estimation de ce paramètre  $\delta$  appelé aussi période de la représentation autorégressive. Nous rappelons qu'en théorie du signal, il existe une méthode d'estimation de la période proposée par Tian (voir [26]) pour les suites périodiquement corrélées et nous discutons son application dans le cas de processus autorégressifs fonctionnels.

Par la suite, nous proposons un estimateur du paramètre  $\delta$  basé sur la méthode de la distance minimale de Millar.W, [27]. Nous étudions d'abord un estimateur d'une fonctionnelle de la moyenne  $a$ , en établissant sa convergence et une loi limite puis nous terminons par la convergence et une loi limite de l'estimateur de la distance minimale de la période  $\delta$ . Nous illustrons la performance de l' estimateur de la période par des simulations numériques et sur des données réelles : données climatiques d'El -Niño et température de la ville d'Alger.

La thèse comporte trois chapitres.

Dans le Chapitre 1, nous introduisons la classe des processus autorégressifs fonctionnels en suivant le monographe de D. Bosq [12]. Nous rappelons leurs principales propriétés et théorèmes limites

Dans le Chapitre 2, nous abordons l'estimation de la période d'un processus autorégressif fonctionnel. Nous rappelons d'abord une méthode d'estimation de la période proposée par Tian (voir thèse [26]). Cette méthode s'applique aux suites périodiquement corrélées  $(Y(t), t \in N)$ . Ces suites possèdent la propriété suivante : leur moyenne et leur fonction de covariance sont des fonctions périodiques

de même période (on note leur période commune  $\delta > 0$ ). La méthode de Tian consiste à introduire la fonction autocovariance empirique :

$$\hat{R}(\tau) = \frac{1}{N - \tau} \sum_{t=1}^{N-\tau} (Y(t) - \bar{z}_N) (Y(t + \tau) - \bar{z}_N)$$

où

$$\bar{z}_N = \frac{1}{N} \sum_{t=1}^N Y(t)$$

Il propose un estimateur de la période  $\delta$  défini par:

$$\hat{\delta}_N = \text{Arg} \max_{1 \leq d \leq D} (\log V_N(d) - d \cdot \beta(N))$$

où

$$V_N(d) = \frac{d}{G(N)} \sum_{s=1}^{\lfloor G(N)/d \rfloor} \left| \hat{R}(sd) \right|$$

$\beta(N) = N^{-1/(2+b)}$ ,  $0 < b < 1/2$ ,  $G(N) = O(N)$ ,  
 $[G(N)]^{-1} = O(\beta(N))$  et  $D > 0$  est une borne supérieure de la période  $\delta$ .

Cependant, dans le cadre des processus autorégressifs fonctionnels nous verrons que cette approche ne s'applique pas.

Nous proposons un estimateur de la période basé sur la méthode de distance minimale de Millar.W, [27].

Nous introduisons la quantité  $g(\delta)$  définie par :

$$g(\delta) := \frac{1}{\delta} \int_0^\delta a(t) d\mu(t)$$

pour  $\delta > 0$  et nous étudions un estimateur défini par :

$$\hat{g}_T = \frac{1}{T} \int_0^T Z(t) d\mu_{[T]}(t)$$

basé sur une observation de la trajectoire du processus

$(Z(t), t \in [0, T])$  où  $\mu_{[T]}$  est la mesure translatée de la mesure  $\mu$  sur  $[0, T]$ .

Nous définissons un estimateur de la période  $\delta$  en posant :

$$\hat{\delta}_T := \text{Arg} \min_{\delta \in ]0, \Delta[} |\hat{g}_T - g(\delta)|$$

où  $\Delta$  est un nombre strictement positif donné .



Nous obtenons les résultats asymptotiques suivants sur l'estimateur  $\hat{g}_T$ :

**Convergence.** Si le processus  $Z$  admet une représentation autorégressive ARC(1) de moyenne  $a$  et de bruit blanc  $\varepsilon$  dans  $C[0, \delta]$  où  $\delta > 0$ , alors

$$\lim_{T \rightarrow \infty} \hat{g}_T = g(\delta) \text{ p.s. et dans } L^2$$

**Loi Limite.** Sous des conditions, on a:

$$\sqrt{T}(\hat{g}_T - g(\delta_0)) \Longrightarrow_{T \rightarrow \infty} N_2$$

où  $N_2$  est une variable aléatoire centrée Gaussienne avec une variance  $\delta_0^{-2} \lambda_1$ .

Pour l'estimateur de la distance minimale  $\hat{\delta}_T$  de la vraie période notée  $\delta_0$  nous obtenons:

**Loi Limite.** Sous une condition d'identifiabilité, nous avons:

$$\sqrt{T}(\hat{\delta}_T - \delta_0) \Longrightarrow_{T \rightarrow \infty} N$$

où  $N$  est une v.a. Gaussienne centrée de variance  $\delta_0^{-2} \cdot I^{-2}(\delta_0) \cdot \lambda_1$ ,

$$I(\delta_0) = \frac{1}{\delta_0} (a(\delta_0) - g(\delta_0))$$

et  $\lambda_1$  est la première valeur propre de  $R_{X_0}$ . La convergence est uniforme sur tout compact de  $\Theta = ]0, \Delta[$ .

Le Chapitre 3 est consacré à la simulation d'un processus autorégressif hilbertien d'ordre 1 où nous utiliserons le logiciel R version 2.4.1, plus précisément, la Bibliothèque `far` du package `R`. Les simulations montrent que l'estimateur converge vers la vraie valeur de la période pour le choix de la fonction périodique  $a(t) = \cos t, \sin t$  ou  $a(t) = t$  une fonction linéaire à représentation périodique. Les simulations sont aussi menées dans le cas où la mesure  $\mu$  est ou de Lebesgue ou à densité. Des représentations graphiques sur la loi limite sont faites en utilisant l'estimateur à noyau d'Epanechnikov. Nous appliquons la méthode de la distance minimale aux données réelles de la série climatologique décrivant le phénomène El Niño (ENSO) pendant la période de 1950-2006 et à la prévision de la température d'Alger pour la période de 1970- 2004. Les simulations montrent que notre estimateur converge vers la vraie période pour les deux phénomènes même si on applique quelques "perturbations" sur certain choix de paramètres.

**Publications**

- Wafaa BENYELLES, Tahar MOURID « On a minimum distance estimate of the period in Functional Autoregressive Processes » Journal of Applied statistics 2012. DOI:10.1080/02664763.2012.668178.
- Wafaa BENYELLES, Tahar MOURID « Estimation de la période d'un processus à temps continu à représentation autorégressive » Annales ISUP Vol. 46 Fasc. 1-2 p. 89-101. 2002.
- Wafaa BENYELLES, Tahar MOURID « Estimation de la période d'un processus à temps continu à représentation autorégressive » Comptes Rendus de l'Académie des sciences (C.R.A.S). Paris, t.333, Série I, p.245-248, Juin 2001.

# Chapitre 1

## Processus autorégressifs fonctionnels

### 1.1 Introduction

Soit un processus  $Z = (Z(t), t \in \mathbb{R}^+)$  à temps continu dont on observe la trajectoire sur des intervalles successifs, d'une longueur donnée  $\delta$ , non nécessairement disjoints et on s'intéresse à prévoir l'évolution globale future du processus sur un intervalle de même longueur. Ceci est le problème de prédiction classique d'un élément fonctionnel  $X_{n+1}$ , à partir des éléments  $X_1, X_2, X_3, \dots, X_n$  (morceaux de trajectoire).

Ce type de modèle est utilisé dans des études statistiques où des relations linéaires sont envisagées entre des observations considérées comme des éléments aléatoires à valeur dans un espace Banach (trafic routier, consommation d'électricité, nombre de voyageurs,...). Nous présentons la définition d'un autorégressif d'ordre 1 à valeur dans un espace Banach ARB(1)[12] qui fournit un modèle mathématique décrivant des relations entre des observations et ouvrant un champ très vaste pour les applications.

## 1.2 Préliminaires et définitions

Soit  $(\Omega, \mathcal{A}, P)$  un espace probabilisé complet et  $(B, \|\cdot\|)$  un espace de Banach réel séparable muni de sa tribu Borelienne  $\mathcal{B}$ . On note  $L(B)$  l'algèbre de Banach des opérateurs linéaires bornés définis sur  $B$  à valeur dans  $B$ , muni de la norme usuelle  $\|\cdot\|_L$  [12] (cf. Nous utilisons les définitions et notations du chap III). On note aussi par  $L_H^2(\Omega, A, P)$  l'espaces des variables aléatoires à valeurs dans  $H$  dont  $E\|\cdot\|^2 < \infty$ . Une suite de variables aléatoires définies sur  $(\Omega, A, P)$  à valeurs dans  $(B, \mathcal{B})$  est une suite d'applications mesurables par rapport aux tribus  $\mathcal{A}$  et  $\mathcal{B}$ . Un bruit blanc Banachique, est une suite de variables aléatoires  $(\varepsilon_n, n \in \mathbb{Z})$  indépendantes identiquement distribuées (i.i.d) définies sur  $\Omega$  à valeur dans  $B$  et telles que pour tout  $n \in \mathbb{Z}$  :

$$E(\varepsilon_n) = 0 \quad ; \quad 0 < E\|\varepsilon_n\|^2 = \sigma_\varepsilon^2 < \infty$$

On pose la définition suivante:

**Définition 1.1** Une suite  $(X_n, n \in \mathbb{Z})$  définie sur  $\Omega$  à valeur dans  $B$  est un processus autorégressif d'ordre 1, ARB(1), s'il existe un bruit blanc Banachique  $(\varepsilon_n)$ , un opérateur  $\rho$  de  $L(B)$  et un élément  $a \in B$  tels que:

$$X_n - a = \rho(X_{n-1} - a) + \varepsilon_n \quad p.s \quad (1.1)$$

où  $\rho \neq 0$  et  $0$  est l'opérateur nul.

La motivation de la définition précédente repose sur la remarque suivante:

**Remarque 1.1** 1] Soit  $Z=(Z(t), t \in \mathbb{R})$  un processus à temps continu. Nous dirons qu'il admet une représentation Banachique d'ordre 1 dans un espace de Banach  $B$  s'il existe un bruit blanc Banachique  $(\varepsilon_n, n \in \mathbb{Z})$  (v.a i.i.d), un opérateur  $\rho$  de  $L(B)$  et un nombre  $\delta$  strictement positif telle que la suite des variables aléatoires.

$$X_n := Z(n\delta + t) \quad t \in [0, \delta] \quad (1.2)$$

vérifie la relation (1.1)

2] Dans les applications, le processus  $Z$  est p.s à trajectoire continue et l'espace de Banach  $B$  est généralement  $C_{[0, \delta]}$  l'espace des fonctions continues sur  $[0, \delta]$  ou  $L_H^2(\Omega, A, P)$ .

Dans le but d'étudier l'existence d'un ARB(1)  $X = (X_n, n \in \mathbb{Z})$ , on introduit les conditions suivantes:

(C0) Il existe un entier  $j_0 \geq 1$  tel que  $\|\rho^{j_0}\| < 1$ .

(C1) Il existe  $a > 0$  et  $0 < b < 1$  tel que  $\|\rho^j\| < ab^j$ ,  $j \geq 0$ .

**Lemme 1** (*[12], lemma 3.1*) les conditions suivantes **(C0)** et **(C1)** sont équivalentes.

**Preuve.** Il est clair que (C1) implique (C0), il reste la preuve du sens inverse.

On a  $0 < \|\rho^{j_0}\| < 1$ . Pour  $j \geq j_0$  et en utilisant la division euclidienne de  $j$  par  $j_0$  on a:

$$j = j_0 q + r \quad (1.3)$$

où  $q \geq 1$  et  $0 \leq r < j_0$

On a :

$$\|\rho^j\| \leq \|\rho^{j_0}\|^q \|\rho^r\|$$

et puisque  $q > \frac{j}{j_0} - 1$  et  $0 < \|\rho^{j_0}\| < 1$  il s'en suit que:

$$\|\rho^j\| < a_1 b^j, \quad j \geq j_0$$

où  $a_1 = a = \|\rho^{j_0}\|^{-1} \max_{0 \leq r < j_0} \|\rho^r\|$  et  $b = \|\rho^{j_0}\|^{1/j_0} < 1$ .

Pour  $0 \leq j < j_0$  et avec l'hypothèse que:  $0 < \|\rho^{j_0}\| < 1$ ; on a :

$$\|\rho^j\| \leq \beta b^j$$

avec  $b = \|\rho^{j_0}\|^{1/j_0} < 1$  et  $\frac{\alpha}{\|\rho^{j_0}\|} \|\rho^{j_0}\|^{j_0} = \beta > 0$ .

En effet, soit  $\alpha = \max_{0 \leq j < j_0} (1, \|\rho\|, \|\rho^2\|, \dots, \|\rho^{j_0-1}\|) = \max_{0 \leq j \leq j_0-1} (\|\rho^j\|)$

$$\|\rho^j\| \leq \alpha$$

Soit  $0 < b = \|\rho^{j_0}\|^{1/j_0} < 1$

$$\left\| \left( \frac{\rho}{b} \right)^j \right\| = \frac{\|\rho^j\|}{b^j} \leq \frac{\alpha}{b^j}$$

Or comme  $\min_{0 \leq j \leq j_0-1} (1, b, b^2, \dots, b^{j_0-1}) = b^{j_0-1}$  alors

$$\left\| \left( \frac{\rho}{b} \right)^j \right\| \leq \frac{\alpha}{b^{j_0-1}} = \frac{\alpha}{\|\rho^{j_0}\|} \|\rho^{j_0}\|^{j_0} = \beta$$

$$\|\rho^j\| \leq \beta b^j$$

Conclusion  $\|\rho^j\| < a b^j$ ,  $j \geq 0$  où  $a = \max(\beta, a_1)$  et  $b = \|\rho^{j_0}\|^{1/j_0}$

Si  $\|\rho^{j_0}\| = 0 \implies \rho^{j_0} = 0$  on a :

$j < j_0$  on pose  $j_0 = \min_i (\|\rho^i\| = 0)$  ie pour  $k \leq j_0 - 1$  on a  $\alpha_k = \|\rho^k\| > 0$

soit  $\alpha = \max_{0 \leq k < j_0} (\alpha_k)$  on a:

$$\|\rho^j\| \leq \alpha \leq ab^j$$

il suffit de prendre  $b = \frac{1}{2}$  et  $a = 2^{j_0} \alpha$ . ■

Ce lemme élémentaire montre la condition "naturelle"  $\sum_{j=0}^{\infty} \|\rho^j\|^2 < \infty$  qui assure l'existence de  $(X_n)_n$  sous la forme d'une série qui converge dans  $L_H^2(P)$  (cf. Théorème 1 ci dessous, valable dès que  $\|\rho^{j_0}\| < 1$  pour un  $j_0 \geq 1$ ).

Finalement, on remarque que **(C0)** (ou **(C1)**) n'implique pas que  $\|\rho\| < 1$  contrairement au cas de la dimension 1 (voir exemple du processus d'Ornstein-Uhlenbeck).

**Définition 1.2** Soient  $X, Y \in L_H^2(P)$  tels que  $EX = EY = 0$

- 1)  $X$  et  $Y$  sont dit faiblement orthogonaux si  $E\langle X, Y \rangle = 0$
- 2)  $X$  et  $Y$  sont dit orthogonaux si pour tout  $x, y \in H$   $E(\langle x, X \rangle \langle y, Y \rangle) = 0$

**Théorème 1** [12] (cf. theorem 3.1) Si **(C0)** est vérifié alors on a une solution stationnaire unique donnée par:

$$X_n = a + \sum_{j=0}^{\infty} \rho^j(\varepsilon_{n-j}), \quad n \in \mathbb{Z} \quad (1.4)$$

la série converge dans  $L_H^2(\Omega, A, P)$  et presque sûrement. En outre  $(\varepsilon_n)_n$  est un processus d'innovation de  $(X_n - a)_n$

**Preuve.** On peut supposer que  $a=0$ . De l'orthogonalité des  $(\varepsilon_n)$

$$\Delta_m^{m'} := \left\| \sum_{j=m}^{m'} \rho^j(\varepsilon_{n-j}) \right\|_{L_H^2(P)}^2 = \sum_{j=m}^{m'} \|\rho^j(\varepsilon_{n-j})\|_{L_H^2(P)}^2$$

pour  $1 \leq m < m'$ . D' autre part

$$\begin{aligned} \|\rho^j(\varepsilon_{n-j})\|_{L_H^2(P)}^2 &= E \langle \rho^j(\varepsilon_{n-j}), \rho^j(\varepsilon_{n-j}) \rangle \\ &\leq \sigma^2 \|\rho^j\|^2 \end{aligned}$$

De plus , on a:

$$\Delta_m^{m'} \leq \sigma^2 \sum_{j=m}^{m'} \|\rho^j\|^2 \longrightarrow 0 \quad \text{quand } m \text{ et } m' \longrightarrow \infty$$

Ainsi, par le critère de Cauchy, il s'en suit que la série converge dans  $L^2_H(P)$

Dans ce cas si  $E \left( \sum_{j=0}^{\infty} \|\rho^j\| \|\varepsilon_{n-j}\| \right)^2 < \infty$ , il s'en suit que  $\sum_{j=0}^{\infty} \|\rho^j\| \|\varepsilon_{n-j}\| < \infty$  la série (1.4) converge presque sûrement aussi.

Considérons maintenant le processus stationnaire

$$Y_n = \sum_{j=0}^{\infty} \rho^j(\varepsilon_{n-j}), \quad n \in \mathbb{Z}$$

A l'aide de la bornitude de  $\rho$  nous voyons que

$$\begin{aligned} Y_n - \rho(Y_{n-1}) &= \sum_{j=0}^{\infty} \rho^j(\varepsilon_{n-j}) - \sum_{j=0}^{\infty} \rho^{j+1}(\varepsilon_{n-1-j}) \\ &= \varepsilon_n, \quad n \in \mathbb{Z} \end{aligned}$$

qui signifie que  $(Y_n)$  est aussi une solution de 1.1.

Inversement, soit  $(X_n)$  une solution stationnaire. Une déduction simple donne

$$X_n = \sum_{j=0}^k \rho^j(\varepsilon_{n-j}) + \rho^{k+1}(X_{n-k-1}), \quad k \geq 1 \quad (1.5)$$

Mais

$$E \left\| X_n - \sum_{j=0}^k \rho^j(\varepsilon_{n-j}) \right\|^2 \leq \|\rho^{k+1}\|^2 E \|X_{n-k-1}\|^2$$

par stationnarité,  $E \|X_{n-k-1}\|^2$  est constante et en utilisant le lemme 1 on a :  $\|\rho^{k+1}\|^2 \rightarrow 0$  quand  $k \rightarrow \infty$ .

Par conséquent

$$X_n = \sum_{j=0}^{\infty} \rho^j(\varepsilon_{n-j}), \quad n \in \mathbb{Z}$$

Ceci prouve l'unicité

Il reste à montrer que  $\varepsilon$  est l'innovation de  $X$ :

Premièrement  $\varepsilon_n = X_n - \rho(X_{n-1}) \in \mathcal{M}_n$  ainsi  $\mathcal{M}_n$  est un LCS (sous espace clos engendré par  $(X_m, m < n)$ )

Deuxièmement  $\varepsilon_n$  est orthogonal à  $\rho^j(\varepsilon_{m-j})$ ,  $j \geq 0$ ,  $m < n$   $\varepsilon_n$  est orthogonal à  $X_m = \sum_{j=0}^{\infty} \rho^j(\varepsilon_{m-j})$ ,  $m < n$ . ■

**Exemple 1** Soit l'espace de Hilbert  $H = L^2([0, 1], B_{[0,1]}, \lambda)$ ,  $\lambda$  étant la mesure de Lebesgue et  $\rho = l_K$  un opérateur à noyau  $K$  tel que:

$$\int_{[0,1]^2} K^2(s, t) ds dt < 1$$

on définit le bruit blanc  $(\varepsilon_n^{(\varphi)})$  par:

$$\varepsilon_n^{(\varphi)}(t) = \int_n^{n+t} \varphi(n+t-s) dW(s), \quad 0 \leq t \leq 1, \quad n \in \mathbb{Z}$$

où  $W$  est le processus de Wiener,  $\varphi \in H$  et  $\int_0^1 \varphi^2(u) du > 0$ . Les conditions du Théorème 1 sont alors satisfaites et ainsi on obtient le processus ARH(1)

$$X_n = \sum_{j=0}^{\infty} l_K^j(\varepsilon_{n-j}^{(\varphi)}), \quad n \in \mathbb{Z}$$

### 1.3 Processus d'Ornstein-Uhlenbeck

On considère le processus réel d'Ornstein-Uhlenbeck

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

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

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

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

La version choisie de  $\xi_t$  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  $\xi_t$  sur l'intervalle  $[n, n+1]$ .



D'autre part,

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

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

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

Enfin comme  $(W(u))$  est à accroissements indépendants,  $(\varepsilon_n)$  est un bruit blanc et d'autre part

$$\|\rho^n\| = \sup_{\|f\|=1} \|\rho^n f\| = e^{-c(n-1)}$$

Donc

$$\sum_{n \geq 0} \|\rho^n\| < \infty \quad \text{et} \quad (X_n) \quad \text{est un ARB}(1)$$

## 1.4 Construction d'un modèle ARC(1) ( $B = C_{[0,1]}$ ).

Nous donnons le théorème de Karhunen-Loève et une application pour construire un processus de Wiener.

Soit  $(\zeta_r, a \leq r \leq b)$ ,  $a$  et  $b$  sont finis, un processus stochastique de 2<sup>eme</sup> ordre de moyenne nulle et de fonction de covariance  $K$  continu. Soit  $(\varphi_j, j = 1, 2, \dots)$  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 tout vecteur aléatoire  $(\xi_1, \dots, \xi_r)$  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  $\xi_j$  sont donnés par le développement en série de Karhunen-Loève, on obtient:

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

Pour un processus de Wiener sur  $[0, T]$ , nous trouvons,

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

et donc

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

Si  $T = n + 1$  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  $i \in N$  :

$$\begin{aligned} \varepsilon_i & : [0, 1] \times \Omega \longmapsto \mathbb{R} \\ (\nu, w) & \longmapsto W_{i+\nu}(w) - W_i(w) \end{aligned}$$

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

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

avec  $(\beta_i, \varphi_i)$  les valeurs et vecteurs propres associées à et posons

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

Ainsi  $(X_n)$  est un *ARC* (1).

## 1.5 Une classe d'ARB(1) à valeurs dans $L^2_{[0,1]}$

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

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

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

Soit  $\rho$  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.

## 1.6 Processus avec saisonnalité

Considérons un processus réel de la forme

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

où  $(\xi_t)$  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  $\eta_t$  admet une représentation ARB(1) avec:

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

et

$$EX_n = m$$

$X_n$  est donc stationnaire alors que  $(\eta_t)$  ne l'est pas.

**Conclusion:**

Ces exemples montrent l'intérêt de cette classe de processus où des processus réels usuels admettent une représentation autorégressive dans des espaces fonctionnels bien choisis.

## 1.7 Théorèmes limites des ARH(1)

On suppose que le processus  $(X_n, n \in \mathbb{Z})$  vérifie (1.1) chap1 et sous la condition (C0) (ou C1). On pose pour tout  $n \in \mathbb{N}$ ,

$$S_n = \sum_{i=1}^n X_i, \quad \overline{X}_n = \frac{S_n}{n}, \quad \overline{\varepsilon}_n = \frac{\sum_{i=1}^n \varepsilon_i}{n}$$

### 1.7.1 Loi forte des grands nombres(LFGN)

Nous avons le théorème sur la loi forte des grands nombres de  $(X_n)$  dans  $B$ .

**Théorème 2** ([12]) *Si les variables aléatoires  $(\varepsilon_n, n \in \mathbb{Z})$  sont centrées indépendantes identiquement distribuées,  $E \|\varepsilon_0\|^2 < \infty$  et si la condition (C0) est vérifiée, alors la suite  $(X_n, n \in \mathbb{Z})$  vérifie la loi forte des grands nombres dans  $B$  :*

$$\overline{X}_n \xrightarrow[n \rightarrow \infty]{} a \quad p.s$$

**Définition 1.3** *Considérons un espace de Hilbert réel et séparable muni de la norme  $\|\cdot\|$ . Un ARH(1) est une suite de variables aléatoires à valeurs dans  $H$  telle que*

$$X_n - a = \rho(X_{n-1} - a) + \varepsilon_n \quad n \in \mathbb{Z} \quad (1.6)$$

où  $(\varepsilon_n)_n$  est un  $H$  - bruit blanc ,  $a \in H$  et  $\rho$  est un opérateur compact et symétrique sur  $H$  tel que  $\|\rho\|_L < 1$ .

Nous avons le lemme suivant:

**Lemme 2** [12] *Si  $X$  est un ARH(1) de moyenne nulle, alors:*

$$|E \langle X_0, X_h \rangle| \leq \|\rho^h\| E \|X_0\|^2, \quad h > 0$$

**Preuve.** De la relation

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

On en déduit que

$$\langle X_0, X_h \rangle = \sum_{j=0}^{h-1} \langle X_0, \rho^j (\varepsilon_{h-j}) \rangle + \langle X_0, \rho^h(X_0) \rangle, \quad h > 0$$

Comme  $\varepsilon$  est le processus d'innovation, on a :

$$E \langle X_0, X_h \rangle = E \langle X_0, \rho^h(X_0) \rangle$$

Ce qui implique notre résultat

$$|E \langle X_0, X_h \rangle| \leq \|\rho^h\| E \|X_0\|^2, \quad h > 0$$

■

Nous avons le théorème suivant:

**Théorème 3** [12] *(cf. théorème 3.7) Soit  $X$  un ARH(1), alors quand  $n \rightarrow \infty$ ,*

$$E \left\| \frac{S_n}{n} \right\|^2 = O\left(\frac{1}{n}\right)$$

De plus pour tout  $\beta > \frac{1}{2}$

$$\frac{n^{1/4}}{(\log n)^\beta} \frac{S_n}{n} \rightarrow 0 \quad p.s$$

**Preuve.** Soit l'identité

$$E \|X_n + \dots + X_{n+p-1}\|^2 = \sum_{n \leq j, j' \leq n+p-1} E \langle X_j, X_{j'} \rangle$$

Utilisant la stationnarité et le lemme (2) on obtient:

$$E \|X_n + \dots + X_{n+p-1}\|^2 \leq 2p E \|X_0\|^2 \sum_{h=0}^{p-1} \|\rho^h\| = O(p)$$

D'où le premier résultat. Le corollaire 2.3[12] implique le deuxième résultat de convergence p.s. ■

**Théorème 4** [12] (cf.theorem 3.9) Soit  $X = (X_n, n \in \mathbb{Z})$  un ARH(1) avec la condition **(C0)**. Alors, il existe  $\alpha_0 > 0$  et  $\beta_0 > 0$ , qui dépend seulement de  $\rho$  et  $P_{\varepsilon_0}$  tel que:

$$P \left( \left\| \frac{S_n}{n} - a \right\| \geq \eta \right) \leq 4 \exp \left( -\frac{n\eta^2}{\alpha_0 + \beta_0\eta} \right), \quad \eta > 0$$

**Preuve.** On peut supposer que  $a = 0$  et on note que  $(1 - \rho)^{-1} = \sum_{j \geq 0} \rho^j$  existe et est borné. On considère la décomposition

$$\frac{S_n}{n} = (1 - \rho)^{-1} \bar{\varepsilon}_n + \Delta_n, \quad n \geq 1$$

où  $\bar{\varepsilon}_n = \frac{1}{n} \sum_{i=1}^n \varepsilon_i$  et

$$\Delta_n = (1 - \rho)^{-1} \frac{\rho^n(\varepsilon_1) + \dots + \rho(\varepsilon_n)}{n} + \frac{(\rho + \dots + \rho^n) X_0}{n}$$

Pour tout  $\eta > 0$  on a

$$P \left( \left\| \frac{S_n}{n} \right\| \geq \eta \right) \leq P \left( \left\| (1 - \rho)^{-1} \right\| \left\| \bar{\varepsilon}_n \right\| \geq \frac{\eta}{2} \right) + P \left( \left\| \Delta_n \right\| \geq \frac{\eta}{2} \right)$$

Par la suite, on utilise l'inégalité du Théorème 2.5 (cf Bosq[12]), ■

### 1.7.2 Loi des grands nombres en moyenne d'ordre $\alpha > 0$

L'étude des probabilités et des sommes de variables aléatoires indépendantes dans un espace de Banach est fortement liée à la géométrie de cet espace. Ceci amène à considérer le type ou le cotype d'un espace de Banach pour établir des résultats de convergence et de vitesse.

**Théorème 5** [12] On suppose que la condition **(C0)** est vérifiée et que les variables aléatoires  $(\varepsilon_n, n \in \mathbb{Z})$  sont centrées iid et de second ordre. Alors:

Si  $B$  est un espace de Banach de type  $\alpha$ ,  $1 < \alpha \leq 2$ , on a:

$$E \left\| \bar{X}_n - a \right\|^\alpha = c_1 \left( \frac{R}{n} \right)^\alpha (nE \|\varepsilon_0\|^\alpha + E \|X_0 - a\|^\alpha)$$

où  $c_1$  est une constante et  $R := \sum_{i=0}^{\infty} \|\rho^i\|$

### 1.7.3 Convergence en moyenne d'ordre $r > 0$ , dans un Banach

**Théorème 6** [12] *Sous les conditions du théorème 5 chap 1.7.2, alors pour tout  $r > 0$  on a :*

$$\lim_{n \rightarrow \infty} E \left\| \bar{X}_n - a \right\|^r = 0$$

### 1.7.4 Théorème central limite(TCL)

Nous dirons que la suite  $(X_n, n \in \mathbb{N})$  vérifie le théorème central limite dans  $B$ , si la suite  $\left(\frac{S_n - na}{\sqrt{n}}, n \in \mathbb{N}\right)$  converge en loi dans  $B$ .

**Théorème 7** [12] *On suppose que les variables aléatoires  $(\varepsilon_n)$  sont centrées indépendantes et identiquement distribuées,  $E \|\varepsilon_0\| < \infty$  et que la condition **(C0)** est réalisée. Alors le processus  $(X_n, n \in \mathbb{N})$  un ARB(1) vérifie le TCL dans  $B$ , si et seulement si, le processus  $(\varepsilon_n, n \in \mathbb{N})$  le vérifie. Dans ce cas :*

$$\sqrt{n} (\bar{X}_n - a) \Longrightarrow \mathcal{N}, \quad \text{quand } n \longrightarrow \infty$$

où la variable aléatoire  $\mathcal{N}$  est Gaussienne centrée dans  $B$ , d'opérateur de covariance

$$(1 - \rho)^{-1} R_{\varepsilon_0} (1 - \rho)^{* - 1}$$

et  $R_{\varepsilon_0}$  est l'opérateur de covariance de  $\varepsilon_0$ .

**Théorème 8** [12] (cf .theorem3.10) *Un processus ARH(1)  $X$  associé à un bruit blanc et un opérateur  $\rho$  qui satisfait la condition **(C0)**, dans ce cas :*

$$\sqrt{n} \left( \frac{S_n}{\sqrt{n}} - a \right) \rightarrow \mathcal{N}$$

où la variable aléatoire  $\mathcal{N}$  est Gaussienne centrée dans  $B$ , d'opérateur de covariance

$$(1 - \rho)^{-1} R_{\varepsilon_0} (1 - \rho)^{* - 1}$$

**Preuve.** On pose  $a=0$  et on considère la décomposition

$$\frac{S_n}{n} = (1 - \rho)^{-1} \bar{\varepsilon}_n + \Delta_n, \quad n \geq 1$$

$$\sqrt{n} \frac{S_n}{n} = (1 - \rho)^{-1} \sqrt{n} \bar{\varepsilon}_n + \sqrt{n} \Delta_n, \quad n \geq 1$$

où  $\bar{\varepsilon}_n = \frac{1}{n} \sum_{i=1}^n \varepsilon_i$  et

$$\Delta_n = (1 - \rho)^{-1} \frac{\rho^n(\varepsilon_1) + \dots + \rho(\varepsilon_n)}{n} + \frac{(\rho + \dots + \rho^n) X_0}{n}$$



on montre dans [12]

$$\|\sqrt{n}\Delta_n\| \leq \|(1-\rho)^{-1}\| \frac{1}{\sqrt{n}} \sum_{j=1}^n \|\rho^{n-j+1}\| \|\varepsilon_j\| + \frac{\|X_0\|}{\sqrt{n}} \sum_{j \geq 1} \|\rho^j\|$$

$$\|\sqrt{n}\Delta_n\| \longrightarrow 0 \quad p.s$$

la continuité de  $(1-\rho)^{-1}$  implique que

$$(1-\rho)^{-1} \sqrt{n} \bar{\varepsilon}_n \rightharpoonup \mathcal{N}(0, (1-\rho)^{-1} R_{\varepsilon_0} (1-\rho)^{* -1})$$

d'où le résultat ■

Nous utiliserons la définition suivante:

**Définition 1.4** [12] (cf .definition 2.3) Soit  $B$  un espace de Banach et  $p \in ]1, 2]$  (resp.  $q \in [2, \infty[$ ) un nombre réel.

$B$  est de **type  $p$**  (resp. de **cotype  $q$** ) s'il existe des constantes strictement positives  $c$  (resp.  $c'$ ) telles que pour toute suite  $X_1, \dots, X_n$  de  $B$ -variables aléatoires indépendantes vérifiant  $E \|X_i\|^p < \infty$  (resp.  $E \|X_i\|^q < \infty$ ),  $EX_i = 0, 1 \leq i \leq n$  et les inégalités suivantes (1.7) (resp. 1.8):

$$E \left\| \sum_{i=1}^n X_i \right\|^p \leq c \sum_{i=1}^n E \|X_i\|^p \quad (1.7)$$

$$\left( \text{resp. } E \left\| \sum_{i=1}^n X_i \right\|^q \geq c' \sum_{i=1}^n E \|X_i\|^q \right) \quad (1.8)$$

**Corollaire 1** [12] Si les variables aléatoires  $(\varepsilon_n)$  sont centrées indépendantes et identiquement distribuées du second ordre, alors chacune des conditions suivantes est suffisante pour que l'ARB(1)  $(X_n, n \in \mathbb{N})$  vérifie le TCL dans  $B$ :

- a)-  $B$  est un espace de Banach de type 2.
- b)-  $B$  est un espace de Banach de cotype 2 et  $\varepsilon_0$  est prégaussienne.
- c)-  $B$  est un espace  $C_{[0,1]}$  et  $\varepsilon_0$  est sous gaussienne.
- d)-  $B$  est un espace  $C_{[0,1]}$  (ou  $B$  est l'espace  $C_{[0,\delta]}$ ) et il existe une variable aléatoire réelle positive  $M$  de carré intégrable telle que:

$$|\varepsilon_0(\omega, s) - \varepsilon_0(\omega, t)| \leq M(\omega) |t - s|$$

**Corollaire 2** [12] Soit  $(X_i, i \geq 1)$  une suite de variables aléatoires indépendantes et identiquement distribuées dans  $B$  tel que:  $\|X_i\| \leq b$ ,  $E \|X_i\|^2 = \sigma^2$  et  $EX_i = 0$

Supposant que  $B$  est un espace de Banach séparable de type  $p$ . Alors, si  $1 < p \leq 2$  on a:

$$\|\overline{X_n}\| = O\left(\frac{1}{n^{1-\frac{1}{p}}}\right) \quad p.s \quad (1.9)$$

**Théorème 9** [12] Soit  $X = (X_i, i \geq 1)$  une suite de variables aléatoires indépendantes et identiquement distribuées dans  $B$ .

- 1) Si  $B$  est de type 2 et  $E \|X_i\|^2 < \infty$ , alors  $X$  satisfait le TCL
- 2) Si  $B$  est de cotype 2 et si  $X_1$  est prégaussien (i.e il existe  $Y$  une variable aléatoire Gaussienne à valeur dans  $B$  telle que:  $C_Y = C_{X_1}$ ), alors  $X$  satisfait le TCL
- 3) Si  $B = C_{[0,1]}$  et

$$|X_1(t) - X_1(s)| \leq M_1 |t - s|, \quad 0 \leq s, t \leq 1$$

où  $M_1$  est une positive variable aléatoire telle que:  $EM_1^2 < \infty$ , alors  $X$  satisfait le TCL .

# Chapitre 2

## Estimation de la période de la représentation AR

### 2.1 Méthode de Tian

#### 2.1.1 Les suites périodiquement corrélées

Les processus périodiquement corrélés (PPC) forment une classe de processus qui sont généralement non stationnaires, mais qui possèdent plusieurs propriétés des processus stationnaires. Ils sont généralement utilisés comme modèle dans la météorologie, la radio physique et la science de communication, etc...

**Définition 2.1** *Un processus stochastique  $\{Z(t), t \in \mathbf{I}\}$  réel défini sur un espace de probabilité  $(\Omega, \mathcal{A}, P)$  et ayant un moment d'ordre deux fini est appelé Processus Périodiquement Corrélé (ou PPC) avec une période minimale  $\delta$  si:  $\delta$  est le plus petit nombre positif tel que:*

$$\begin{aligned} E[(Z(u) - m)(Z(v) - m)] &= R(u, v) \\ &= R(u + \delta, v + \delta) \end{aligned}$$

$$E[Z(u)] = m(u) = m(u + \delta)$$

*Ceci:  $\forall (u, v) \in \mathbf{I} \times \mathbf{I}$  tel que  $(u + \delta, v + \delta) \in \mathbf{I} \times \mathbf{I}$*

**Définition 2.2** *Si  $\mathbf{I} = \mathbb{Z}$  le processus  $(Z(t), t \in \mathbf{I})$  est une suite périodiquement corrélée, notée PPC.*

### 2.1.2 Méthode de Tian associée au PPC

La fonction autocovariance est un moyen usuel dans la recherche des périodicités dans un signal. Tian (voir thèse [26] ) a proposé d'utiliser la fonction autocovariance empirique:

$$\hat{R}(\tau) = \frac{1}{N - \tau} \sum_{t=1}^{N-\tau} (Z(t) - \bar{z}_N) (Z(t + \tau) - \bar{z}_N)$$

pour estimer la période  $\delta$  de  $m(t)$ , et où:

$$\bar{z}_N = \bar{z}_N(1, 1) = \frac{1}{N} \sum_{t=1}^N Z(t)$$

avec

$$\bar{z}_N(i, d) = \frac{1}{h(N-i, d) + 1} \sum_{j=1}^{h(N-i, d)+1} Z(i+j, d), \quad 1 \leq i, d \leq m$$

et  $h(k, l) = \lfloor \frac{k}{l} \rfloor$  ( $\lfloor \cdot \rfloor$  partie entière) et  $N$  nombre d'observations.

Nous remarquerons que pour la suite PC de période  $\delta$ , on a les propriétés suivantes [26]:

$$\lim_{n \rightarrow \infty} E(\bar{z}_N) = \bar{m}$$

$$\lim_{n \rightarrow \infty} E(\hat{R}(\tau)) = R_m(\tau) + \bar{R}(\tau)$$

Où

$$\bar{m} = \frac{1}{\delta} \sum_{t=1}^{\delta} m(t)$$

$$R_m(\tau) = \frac{1}{\delta} \sum_{t=1}^{\delta} [m(t) - \bar{m}] [m(t + \tau) - \bar{m}] \quad (2.1)$$

$$\bar{R}(\tau) = \frac{1}{\delta} \sum_{t=1}^{\delta} cov(Z(t), Z(t + \tau)) \quad (2.2)$$

la fonction dans (2.1) est périodique, de même période  $\delta$  que  $m(t)$ . Pour une large classe de processus la quantité (2.2) décroît exponentiellement quand  $\tau$  croît. Ainsi, pour ces processus, et pour  $n$  assez grand,  $E(\hat{R}(\tau))$  peut être considérée comme une fonction périodique qui a la même période que  $m(t)$ , aussi on a si:

$$\begin{aligned} \left| \frac{1}{\delta} \sum_{t=1}^{\delta} [m(t) - \bar{m}] [m(t + \tau) - \bar{m}] \right| &\leq \frac{1}{\delta} \sum_{t=1}^{\delta} [m(t) - \bar{m}]^2 \\ &= R_m(0) \end{aligned}$$

alors  $|R_m(\tau)|$  atteint son maximum en  $\tau = k\delta$ ,  $k = 1, 2, \dots$

La méthode d'estimation de la période de Tian est basée sur cette maximisation. Son estimateur est défini par:

$$\hat{\delta}_N = \text{Arg} \max_{1 \leq d \leq D} (\log V_N(d) - d \cdot \beta(N))$$

où  $D$  est la borne supérieure de la période  $\delta$ ,  $\beta(N) = N^{-1/(2+b)}$ ,

$$0 < b < 1/2 \text{ et } V_N(d) = \frac{d}{G(N)} \sum_{s=1}^{[G(N)/d]} \left| \widehat{R}(sd) \right|$$

La suite des entiers  $G(N)$ , satisfait aux conditions suivantes:

$$G(N) = O(N) \quad \text{et} \quad [G(N)]^{-1} = O(\beta(N))$$

Sous ces hypothèses Tian [26] a obtenu:

**Théorème 10** [26] *Pour un  $\varepsilon > 0$  et  $1 \leq G(N) = O(N)$  On a:*

$$\max_{1 \leq \tau \leq G(N)} \left| \widehat{R}(\tau) - [R_m(\tau) + \overline{R}(\tau)] \right| = O \left[ \frac{1}{\sqrt{N}} (G(N) \log N)^{2/\lambda} (\log \log N)^{\frac{2(\lambda+\varepsilon)}{\lambda}} \right]$$

**Corollaire 3** [26] *Pour toute suite gaussienne  $\{Z(t)\}$  PPC et pour toute suite  $(G(N))$  telle que  $G(N) = O(N)$  alors:*

$$\forall \varepsilon > 0, \max_{1 \leq \tau \leq G(N)} \left| \widehat{R}(\tau) - [R_m(\tau) + \overline{R}(\tau)] \right| = O \left[ N^{-\frac{1}{2+\varepsilon}} \right] \text{ p.s}$$

En utilisant le théorème et le corollaire, Tian montre que si  $\delta$  est la période minimale de la fonction moyenne  $m(t)$  d'une suite  $\{Z(t)\}$  PPC vérifiant les hypothèses précédentes, alors:

$\hat{\delta}_N$  est un estimateur qui converge fortement vers  $\delta$

$$\hat{\delta}_N \longrightarrow \delta \text{ p.s quand } N \longrightarrow \infty$$

### 2.1.3 Les processus à représentation autorégressive

Dans cette partie, nous examinerons l'application de la méthode de Tian aux processus ARH(1). Soit le processus  $(Z(t), t \in \mathbb{R}^+)$  admettant une représentation autorégressive hilbertienne d'ordre 1 [12]. D'après le lemme 3.1 (cf thèse de Mr Mourid) le processus  $(Z(t), t \in \mathbb{R}^+)$  possède une moyenne périodique de période  $\delta$ . Pour la périodicité de la fonction de covariance regardons le cas d'un processus  $Z(t)$  ARH(1) de moyenne  $a(\cdot)$ .

Montrons que la fonction de covariance  $C(t,s)$  n'est pas périodique. Nous avons:

$$C(s, t) := E((Z(s) - a(s)) (Z(t) - a(t)))$$

$$X_n(t) = \rho X_{n-1}(t) + \varepsilon_n(t)$$

Or  $Z(n\delta + t) = X_n(t)$  par suite  $Z(n\delta + t) = \rho Z((n-1)\delta + t) + \varepsilon_n(t)$ . Si  $n=1$ , on a:

$$\begin{aligned} C(t + \delta, t + \delta) &= E((Z(t + \delta) - a(t + \delta)) \cdot (Z(t + \delta) - a(t + \delta))) \\ &= E(\rho Z(t) + \varepsilon_1(t) - a(t)) \cdot (\rho Z(t) + \varepsilon_1(t) - a(t)) \\ &= E((\rho - 1) Z(t) + \varepsilon_1(t) + Z(t) - a(t)) \cdot \\ &\quad ((\rho - 1) Z(t) + \varepsilon_1(t) + Z(t) - a(t)) \\ &= E(Z(t) - a(t)) (Z(t) - a(t)) + \\ &\quad 2E((\rho - 1) Z(t) + \varepsilon_1(t)) (Z(t) - a(t)) + E((\rho - 1) Z(t) + \varepsilon_1(t))^2 \\ &= C(t, t) + A \end{aligned}$$

où  $A = 2E((\rho - 1) Z(t) + \varepsilon_1(t)) (Z(t) - a(t)) + E((\rho - 1) Z(t) + \varepsilon_1(t))^2$

Donc la fonction covariance  $C(s, t)$  n'est pas périodique et la méthode de Tian ne peut s'appliquer à cette classe de processus.

## 2.2 Méthode de la distance minimale

### 2.2.1 Notations -Hypothèses

Nous utilisons les notations et définition du paragraphe 1.2 (chap I)

Soit  $Z = (Z(t), t \in \mathbb{R})$  un processus réel presque sûrement à trajectoires continues. On suppose que le processus  $Z$  admette une représentation autorégressive d'ordre 1  $ARC(1)$  de moyenne  $a$  et de bruit blanc  $\varepsilon$  dans l'espace  $C = C_{[0, \delta]}$  avec  $\delta$  un réel strictement positif. Rappelons qu'une suite  $(X_n, n \in \mathbb{Z})$  définie sur  $\Omega$  à valeur dans  $C$  est un processus autorégressif d'ordre 1,  $ARC(1)$ , s'il existe un bruit blanc banachique  $(\varepsilon_n)$  dans  $C$ , un opérateur  $\rho$ , dans  $L(C)$  et un élément  $a \in C$  tels que:

$$X_n - a = \rho (X_{n-1} - a) + \varepsilon_n \quad p.s$$

On pose:

$$X_n := Z(n\delta + t) \quad t \in [0, \delta] \quad (2.3)$$

et  $Z=(Z(t), t \in \mathbb{R})$  un processus AR(1) à temps continu. On observe la trajectoire du processus  $(Z(t), t \in [0, T])$  et la figure suivante donne une idée sur la décomposition de la trajectoire de  $Z$  en  $X_1, X_2, \dots, X_n, \dots$

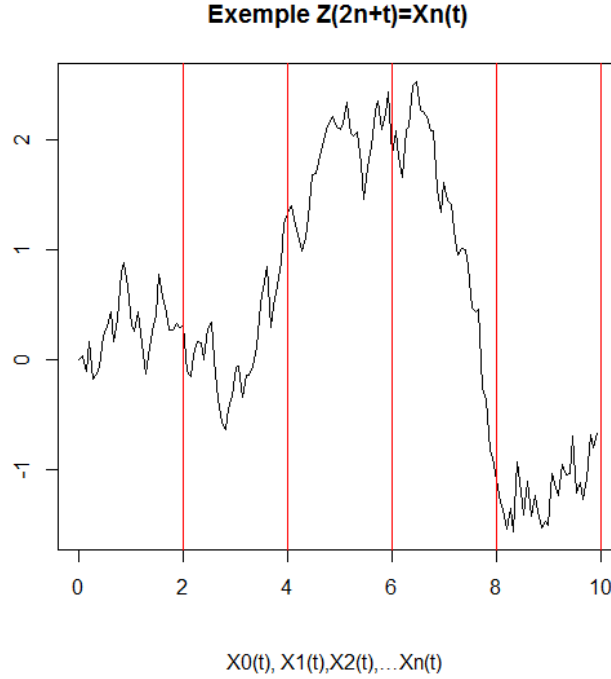


Figure 1: Graphe de la représentation de  $Z$

Nous rappelons la méthode de l'estimateur par la distance minimale de P.W.Millard[27]. Pour  $\delta \in \Theta$  un ouvert de  $\mathbb{R}^d$ , soit  $\xi_T(\delta)$  une variable aléatoire à valeurs dans un espace de Hilbert  $(H, \|\cdot\|_H)$  (la variable aléatoire  $\xi_T$  est définie à partir d'une suite de variables aléatoires indépendantes identiquement distribuées). Le paramètre  $\delta$  est à estimer, Millar propose l'estimateur de la distance du minimum défini par:

$$\left\| \xi_T \left( \hat{\delta}_T \right) \right\|_H = \inf_{\delta \in \Theta} \left\| \xi_T (\delta) \right\|_H$$

$$i.e \quad \hat{\delta}_T = Arg \inf_{\delta \in \Theta} \left\| \xi_T (\delta) \right\|_H$$

On suppose que ce minimum est unique dans  $\Theta$ .

Pour  $\delta_0$  fixé dans  $\Theta$ , on fait les hypothèses suivantes:

- H1 Identifiabilité

$$\forall \varepsilon > 0, \forall \beta > 0, \exists c > 0 / P \left( \inf_{\|\delta - \delta_0\| > \beta} \|\xi_T(\delta) - \xi_T(\delta_0)\| > c \right) > 1 - \varepsilon$$

- H2 Bornitude

La v.a.  $\sqrt{T} \xi_T(\delta)$  est bornée en probabilité.

Dans les applications, cette hypothèse est remplacée par une condition plus forte:

H'2: La v.a.  $\sqrt{T} \xi_T(\delta)$  converge en loi, dans  $H$ , vers une v.a.  $Y$  quand  $T \rightarrow +\infty$ .

- H3 Différentiabilité

$\exists \eta(\delta_0) = (\eta_1, \dots, \eta_d)$ ,  $\eta_i \in H$ , tel que  $\xi_T(\delta) = \xi_T(\delta_0) + T_1(\delta - \delta_0) + \|\delta - \delta_0\| \zeta(\|\delta - \delta_0\|, \cdot)$

où pour tout  $\omega$ ,  $\zeta(\cdot, \omega)$  est définie sur  $\mathbb{R}^+$ , telle que

$$\forall \alpha > 0, \forall \varepsilon > 0, \exists c > 0 / P \left( \sup_{\|\delta\| < \varepsilon} \|\zeta(\|\delta\|, \omega)\| < c \right) > 1 - \alpha$$

et les  $\eta_i$  sont linéairement indépendants dans  $H$  et où  $T_1$  est l'application linéaire de  $\mathbb{R}^d$  dans  $B_\eta$ , définie par:

$$T_1(x) = \sum_{i=1}^d x_i \eta_i$$

$B_\eta$  est le sous espace engendré par  $\eta = (\eta_1, \dots, \eta_d)$

Sous ces conditions Millar montre le résultat suivant ;

**Théorème 11** [27] *Sous les hypothèses H1, H2, H3, nous avons:*

$$1] P \left( \hat{\delta}_T \text{ existe et est unique} \right) \longrightarrow 1 \quad T \rightarrow \infty$$

$$2] \xi_T \left( \hat{\delta}_T \right) = (I - \Pi) \cdot \xi_T(\delta_0) + o_p \left( \sqrt{T} \right) \quad T \rightarrow \infty$$

$$3] \hat{\delta}_T - \delta_0 = T_1^{-1} \circ \Pi \xi_T(\delta_0) + o_p \left( \sqrt{T} \right) \quad T \rightarrow \infty$$



4] Si de plus on suppose que H 2 est vraie, alors:

$$\sqrt{T} \left( \xi_T \left( \hat{\delta}_T \right) - \xi_T \left( \delta \right) \right) \implies -\Pi \circ Y$$

et

$$\sqrt{T} \left( \hat{\delta}_T - \delta_0 \right) \implies -T_1^{-1} \circ \Pi \circ Y$$

où  $T_1$  est l'application linéaire de  $\mathbb{R}^d$  dans  $B_\eta$ , définie par:

$$T_1(x) = \sum_{i=1}^d x_i \eta_i$$

$B_\eta$  est le sous espace engendré par  $\eta = (\eta_1, \dots, \eta_d)$ .

$\Pi$  est le projecteur orthogonal sur  $B_\eta$ .

**Remarque 2.1** les vecteurs  $\eta_i$  étant linéairement indépendants, l'application  $T_1$  est continue bijective et  $T_1^{-1}$  est linéaire et continue [27].

## 2.3 Convergence presque sûre et $L^2$

On définit l'estimateur suivant:

$$\hat{g}_T = \frac{1}{T} \int_0^T Z(t) d\mu_{[T]}(t) \quad (2.4)$$

un estimateur naturel de la fonction  $g$ , posons pour  $\delta > 0$ :

$$g(\delta) := \frac{1}{\delta} \int_0^\delta a(t) d\mu(t) \quad (2.5)$$

où  $\mu$  une mesure à signe borné dans  $C_{[0,\delta]}^*$ (dual),  $[T]$  est la partie entière de  $T$  et pour tout  $n \in \mathbb{N}$ , et pour tout borélien  $A$ ,  $A \in [0, n\delta]$  la mesure  $\mu_n$  est définie par (translaté de  $\mu$ ):

$$\mu_n(A) := \sum_{j=0}^{n-1} \mu(A \cap [j\delta, (j+1)\delta[ - j\delta)$$

**Théorème 12** Si le processus  $Z$  admet une représentation autorégressive  $ARC(1)$  de moyenne  $a$  et de bruit blanc  $\varepsilon$  dans  $C[0, \delta]$  où  $\delta > 0$ , alors

$$\lim_{T \rightarrow \infty} \hat{g}_T = g(\delta)$$

presque sûrement

**Preuve.** Pour tout  $T \in \mathbb{R}^+$  et  $\delta > 0$ , il existe un entier  $n = n_{T,\delta}$  tel que :  $n\delta \leq T \leq (n+1)\delta$ . Par suite nous avons  $\frac{n}{T} \rightarrow \frac{1}{\delta}$ . De la définition (2.4) de l'estimateur  $\hat{g}_T$  on a la décomposition suivante :

$$\begin{aligned}\hat{g}_T &= \frac{1}{T} \int_0^T Z(t) d\mu_{[T]}(t) = \frac{1}{T} \int_0^{n\delta} Z(t) d\mu_{[T]}(t) + \frac{1}{T} \int_{n\delta}^T Z(t) d\mu_{[T]}(t) \\ &= : A_n + B_n\end{aligned}$$

Prenant chaque terme séparément:

Pour  $A_n$ , de  $X_j(t) := Z(j\delta + t)$ ,  $t \in [0, \delta]$  et de la définition de  $\mu_{[T]}$ , nous obtenons:

$$\begin{aligned}A_n &: = \frac{1}{T} \int_0^{n\delta} Z(t) d\mu_{[T]}(t) \\ &= \frac{1}{T} \sum_{j=0}^{n-1} \int_{j\delta}^{(j+1)\delta} Z(t) d\mu_{[T]}(t) \\ &= \frac{1}{T} \sum_{j=0}^{n-1} \int_0^{\delta} Z(\delta j + t) d\mu(t) \\ &= \frac{n}{T} \int_0^{\delta} \left[ \frac{1}{n} \sum_{j=0}^{n-1} X_j(t) \right] d\mu(t)\end{aligned}\tag{2.6}$$

D'après le théorème 2 Chap.1.7.1, appliqué à l'espace  $(B, \|\cdot\|) = (C_{[0, \delta]}, \|\cdot\|_{\infty})$  nous avons:

$$\begin{aligned}\left\| \frac{1}{n} \sum_{j=0}^{n-1} X_j - a \right\|_{\infty} &: = \sup_{0 \leq t \leq \delta} \left| \frac{1}{n} \sum_{j=0}^{n-1} X_j(t) - a(t) \right| \\ &= \|\bar{X}_n - a\|_{\infty} \rightarrow 0 \text{ p.s.}\end{aligned}$$

quand  $n \rightarrow \infty$  et où  $\bar{X}_n = \frac{1}{n} \sum_{j=0}^{n-1} X_j$ .

Pour la majoration par une fonction  $L_{\mu}^1$ , on a pour  $\forall n \geq n_0$

$$\begin{aligned}\left| \frac{1}{n} \sum_{j=0}^{n-1} X_j(t) \right| &\leq \left| \frac{1}{n} \sum_{j=0}^{n-1} X_j(t) - a(t) + a(t) \right| \\ &\leq \left| \frac{1}{n} \sum_{j=0}^{n-1} X_j(t) - a(t) \right| + |a(t)| \\ &\leq C + |a(t)| \quad := l(t)\end{aligned}$$

où  $l(t)$  est une fonction  $L^1_\mu$ .

Par conséquent, par le théorème de la convergence dominée et du fait que  $\frac{n}{T} \rightarrow \frac{1}{\delta}$ , nous déduisons de (2.6) que quand  $n \rightarrow \infty$  :

$$A_n := \frac{1}{T} \int_0^{n\delta} Z(t) d\mu_{[T]}(t) \rightarrow \frac{1}{\delta} \int_0^\delta a(t) d\mu(t) \quad p.s$$

D'autre part, pour  $B_n$ , on a:

$$\begin{aligned} B_n &: = \frac{1}{T} \int_{n\delta}^T Z(t) d\mu_{[T]}(t) & (2.7) \\ &\leq \frac{1}{T} \sup_{n\delta \leq t \leq T} |Z(t)| \cdot \|\mu\| \\ &\leq \frac{1}{T} \sup_{0 \leq s \leq \delta} |Z(s + n\delta)| \cdot \|\mu\| \\ &\leq \frac{1}{T} \|\mu\| \|X_n\| \end{aligned}$$

où  $\|\mu\|$  désigne la variation totale de  $\mu$ . Par l'inégalité de Markov et la stationnarité de  $(X_n)$  nous avons:

$$\begin{aligned} P\left(\left\|\frac{X_n}{T}\right\| > \varepsilon\right) &\leq \frac{1}{T^2 \varepsilon^2} E \|X_n\|^2 \\ &= \frac{1}{T^2 \varepsilon^2} E \|X_0\|^2 \end{aligned}$$

Par suite

$$\int_1^\infty P\left(\left\|\frac{X_n}{T}\right\| > \varepsilon\right) dT \leq \frac{E \|X_0\|^2}{\varepsilon^2} \int_1^\infty \frac{1}{T^2} dT < \infty$$

Par le lemme de Borel-Cantelli, on obtient donc quand  $n \rightarrow \infty$ :

$$\frac{X_n}{T} \rightarrow 0 \quad p.s$$

Par conséquent de (2.7), nous obtenons:

$$B_n := \frac{1}{T} \int_{n\delta}^T Z(t) d\mu_{[T]}(t) \rightarrow 0 \quad p.s$$

quand  $T \rightarrow \infty$

D'où en regroupant les convergences de  $A_n$  et  $B_n$ , nous obtenons le résultat demandé:

$$\lim_{T \rightarrow \infty} \hat{g}_T = \lim_{T \rightarrow \infty} \frac{1}{T} \int_0^T Z(t) d\mu_{[T]}(t) = \frac{1}{\delta} \int_0^\delta a(t) d\mu(t) \quad p.s$$

■

**Corollaire 2.1** *Si le processus  $Z$  admet une représentation autorégressive ARC(1) de moyenne  $a$  et de bruit blanc  $\varepsilon$  dans  $C[0, \delta]$  où  $\delta > 0$ , alors*

$$\hat{g}_T \xrightarrow[T \rightarrow \infty]{L^2} g(\delta)$$

**Preuve.** Pour tout  $T \in \mathbb{R}^+$  et  $\delta > 0$ , il existe un entier  $n = n_{T,\delta}$  tel que :  $n\delta \leq T \leq (n+1)\delta$ . Par suite nous avons  $\frac{n}{T} \rightarrow \frac{1}{\delta}$ .

$$\begin{aligned} \frac{1}{T} \int_0^T Z(t) d\mu_{[T]}(t) &= \frac{1}{T} \int_0^{n\delta} Z(t) d\mu_{[T]}(t) + \frac{1}{T} \int_{n\delta}^T Z(t) d\mu_{[T]}(t) \\ &=: A_n + B_n \end{aligned}$$

D'une part, de  $X_j(t) := Z(j\delta + t)$ ,  $t \in [0, \delta]$  et de la définition de  $\mu_{[T]}$ , nous obtenons:

$$\begin{aligned} A_n &: = \frac{1}{T} \int_0^{n\delta} Z(t) d\mu_{[T]}(t) & (2.8) \\ &= \frac{1}{T} \sum_{j=0}^{n-1} \int_{j\delta}^{(j+1)\delta} Z(t) d\mu_{[T]}(t) \\ &= \frac{1}{T} \sum_{j=0}^{n-1} \int_0^\delta Z(\delta j + t) d\mu(t) \\ &= \frac{n}{T} \int_0^\delta \left[ \frac{1}{n} \sum_{j=0}^{n-1} X_j(t) \right] d\mu(t) \end{aligned}$$

D'après le théorème 6 Chap.1.7.3, appliqué à l'espace  $(B, \|\cdot\|) = (C_{[0, \delta]}, \|\cdot\|_\infty)$  nous avons:

$$\left\| \frac{1}{n} \sum_{j=0}^{n-1} X_j - a \right\| = \|\bar{X}_n - a\| \rightarrow 0 \text{ dans } L^2$$

Nous déduisons de (2.8) que quand  $n \rightarrow \infty$  :

$$A_n := \frac{1}{T} \int_0^{n\delta} Z(t) d\mu_{[T]}(t) \rightarrow \frac{1}{\delta} \int_0^\delta a(t) d\mu(t) \quad \text{dans } L^2$$

D'autre part,

$$\begin{aligned} B_n &: = \frac{1}{T} \int_{n\delta}^T Z(t) d\mu_{[T]}(t) & (2.9) \\ &\leq \frac{1}{T} \sup_{n\delta \leq t \leq T} |Z(t)| \cdot \|\mu\| \\ &\leq \frac{1}{T} \sup_{0 \leq s \leq \delta} |Z(s + n\delta)| \cdot \|\mu\| \\ &\leq \frac{1}{T} \|\mu\| \|X_n\| \end{aligned}$$

Or  $\frac{X_n}{T} \rightarrow 0$  dans  $L^2$ ,

Par conséquent de (2.7) , nous obtenons:

$$B_n := \frac{1}{T} \int_{n\delta}^T Z(t) d\mu_{[T]}(t) \rightarrow 0 \quad \text{dans } L^2$$

quand  $T \rightarrow \infty$

D'où en regroupant les convergences de  $A_n$  et  $B_n$ , nous obtenons le résultat demandé:

$$\lim_{T \rightarrow \infty} \frac{1}{T} \int_0^T Z(t) d\mu_{[T]}(t) = \frac{1}{\delta} \int_0^\delta a(t) d\mu(t) \quad \text{dans } L^2$$

■

## 2.4 Théorème central limite

Soit  $X = (X_n)$  un processus ARC(1) associé à un processus  $Z$  par (2.3), il admet un opérateur de covariance  $R_{X_0}$  qui est un opérateur à noyau sur  $L^2_{[0, \delta]}$  de noyau de covariance  $C(s, t)$ .

$$C(s, t) = E(Z(s) - a(s))(Z(t) - a(t))$$

Les éléments propres de  $R_{X_0}$  sont notés  $(\lambda_j, \Phi_j)$  et vérifient:

$$\lambda_j \Phi_j(t) = \int_0^\delta C(s, t) \Phi_j(s) ds, \quad t \in [0, \delta], \quad j = 0, 1, 2, \dots$$

On fait l'hypothèse suivante:

**H:** On suppose que les éléments propres de  $R_{X_0}$  sont connus et que pour tout  $j \geq 1$  les  $\lambda_j$  sont strictement positifs et  $\|\Phi_j\|_{L^2} = 1$ .

**Théorème 13** [12] *Supposons que l'hypothèse H est satisfaite et que le processus  $Z$  admet une représentation autorégressive ARC(1) de moyenne  $a$  et de bruit blanc  $\varepsilon$  dans  $C_{[0, \delta]}$  et que les trajectoires de  $Z$  vérifient presque sûrement la condition de Lipschitz suivante:*

$$|Z(t, \omega) - Z(s, \omega)| \leq M(\omega) \cdot |t - s|$$

où  $M$  est une variable aléatoire réelle positive vérifiant  $EM^2 < \infty$ . Alors on a:

$$\left( \frac{1}{\sqrt{T}} \sum_{i=0}^{T-1} \int_0^\delta (Z(i\delta + t) - a(t)) \Phi_1(t) dt \right) \Rightarrow N_1$$

quand  $T$  tend vers l'infini où  $N_1$  est une variable aléatoire Gaussienne centrée dans  $\mathbb{R}$  de variance  $\lambda_1$ .

**Preuve.** Comme  $X_j(t) := Z(j\delta + t)$  et la suite  $(X_n, n \in \mathbb{Z})$  est un processus ARC(1) d'innovation  $(\varepsilon_n, n \in \mathbb{Z})$ , alors:

$$\varepsilon_n(t) - \varepsilon_n(s) = \mathcal{Z}(n\delta + t) - \mathcal{Z}(n\delta + s) - E(\mathcal{Z}(n\delta + t) - \mathcal{Z}(n\delta + s) / \mathcal{Z}(u), u \leq n\delta)$$

Par conséquent, de l'hypothèse sur le processus  $Z$ , on a:

$$\begin{aligned} |\varepsilon_n(t) - \varepsilon_n(s)| &\leq |\mathcal{Z}(n\delta + t) - \mathcal{Z}(n\delta + s)| + E(\mathcal{Z}(n\delta + t) - \mathcal{Z}(n\delta + s) / \mathcal{Z}(u), u \leq n\delta) \\ &\leq (M + E(M / \mathcal{Z}(u), u \leq n\delta)) \cdot |t - s| \end{aligned}$$

Donc la suite  $(\varepsilon_n, n \in \mathbb{Z})$  vérifie la condition d) du corollaire 1 Chap.1. Par conséquent par le même corollaire la suite  $ARB(1)$   $(X_n, n \in \mathbb{Z})$  vérifie le théorème central limite dans  $B = C_{[0,\delta]}$  :

$$\sqrt{n}(\bar{X}_n - a) \Longrightarrow N$$

où  $N$  est une Gaussienne centrée dans  $C_{[0,\delta]}$  d'opérateur de covariance

$$(I - \rho_1)^{-1} R_{\varepsilon_0} (I - \rho_1)^{* -1}.$$

Considérons la forme linéaire  $x^* := \mu \in C_{[0,\delta]}$  le dual topologique de  $C_{[0,\delta]}$ . Par la suite, par le théorème de continuité, nous avons le TCL dans  $\mathbb{R}$  suivant:

$$\mu(\sqrt{n}(\bar{X}_n - a)) \Longrightarrow \mu(N)$$

$$\text{où } \mu(N) := \int_0^\delta N(t) d\mu(t)$$

D'autre part,

$$\begin{aligned} \mu(\sqrt{n}(\bar{X}_n - a)) &= \int_0^\delta \sqrt{n}(\bar{X}_n(t) - a(t)) d\mu(t) \\ &= \sqrt{n} \int_0^\delta \left( \frac{1}{n} \sum_{j=0}^{n-1} X_j(t) - a(t) \right) d\mu(t) \\ &= \int_0^\delta \left( \frac{1}{\sqrt{n}} \sum_{j=0}^{n-1} X_j(t) - \sqrt{n} a(t) \right) d\mu(t) \\ &= \int_0^\delta \left( \frac{1}{\sqrt{n}} \sum_{j=0}^{n-1} X_j(t) - \frac{n}{\sqrt{n}} a(t) \right) d\mu(t) \\ &= \frac{1}{\sqrt{n}} \sum_{j=0}^{n-1} \int_0^\delta (X_j(t) - a(t)) d\mu(t) \end{aligned}$$

Donc

$$\frac{1}{\sqrt{n}} \sum_{j=0}^{n-1} \int_0^\delta (X_j(t) - a(t)) d\mu(t) \Longrightarrow \mu(N) = \int_0^\delta N(t) d\mu(t)$$

où  $\mu(N)$  est une variable aléatoire réelle Gaussienne centrée de variance

$$\mu \left( (I - \rho_1)^{-1} R_{\varepsilon_0} (I - \rho_1)^{* -1} \cdot (\mu) \right)$$

Considérons maintenant  $\mu_1$  formes linéaires dans  $C_{[0,\delta]}^*$ , de densités  $\Phi_1$ , par rapport à la mesure de Lebesgue sur  $[0, \delta]$ . D'après ce qui précède, nous aurons :

$$\mu_1 \left( \sqrt{n} (\bar{X}_n - a) \right) = \frac{1}{\sqrt{n}} \sum_{j=0}^{n-1} \int_0^\delta (X_j(t) - a(t)) \cdot \Phi_1(t) dt \implies \mu_1(N) = \int_0^\delta N(t) \cdot \Phi_1(t) dt$$

Notons  $\nu_n$  et  $\nu$  les lois respectives de  $\sqrt{n} (\bar{X}_n - a)$  et de  $N$ . Par convergence faible de  $\nu_n$  vers  $\nu$  et le théorème de la continuité, nous déduisons la convergence des lois de dimensions finies dans  $\mathbb{R}$  :  $\forall \mu_1, \in C_{[0,\delta]}^*$ ,

$$\mu_1 \left( \sqrt{n} (\bar{X}_n - a) \right) \implies \mu_1(N)$$

où  $\mu_1(N)$  est une Gaussienne centrée dans  $\mathbb{R}$ .

Calculons la matrice de covariance de la loi limite. On a pour  $s, t \in [0, \delta]$  :

$$\begin{aligned} C(s, t) & : = E \left( (Z(s) - a(s)) \cdot (Z(t) - a(t)) \right) \\ & = E \left( (X_0(s) - a(s)) \cdot (X_0(t) - a(t)) \right) \\ & = E \left( N(s)N(t) \right) \end{aligned}$$

Donc, pour  $i, j$  on a:

$$\begin{aligned} E \mu_i(N) \mu_j(N) & = E \left( \int_0^\delta N(t) d\mu_i(t) \int_0^\delta N(t) d\mu_j(t) \right) \\ & = \int_0^\delta \int_0^\delta E N(s)N(t) d\mu_i(s) d\mu_j(t) \\ & = \int_0^\delta \left( \int_0^\delta C(s, t) \Phi_i(s) ds \right) \Phi_j(t) dt \\ & = \int_0^\delta \lambda_i \Phi_i(t) \Phi_j(t) dt \\ & = \lambda_i \delta_{ij} \end{aligned}$$

où  $\delta_{ij}$  est le symbole de Kronecker.  
de plus



$$\begin{aligned}
E\mu^2(N) &= E \left( \int_0^\delta N(t) d\mu(t) \right)^2 = E \left( \int_0^\delta N(t) \Phi_1(t) dt \right)^2 \\
&= \int_0^\delta \int_0^\delta E N(s)N(t) \Phi_1(s) ds \Phi_1(t) dt \\
&= \int_0^\delta \left( \int_0^\delta C(t-s) \Phi_1(s) ds \right) \Phi_1(t) dt \\
&= \int_0^\delta \lambda_1 \Phi_1(t) \Phi_1(t) dt \\
&= \lambda_1 \int_0^\delta \Phi_1^2(t) dt = \lambda_1
\end{aligned}$$

D'où le théorème. ■

**Théorème 14** *Supposant que les conditions du théorème 12 soient satisfaites . Alors on a :*

$$\sqrt{T}(\hat{g}_T - g(\delta_0)) \implies_{T \rightarrow \infty} N_2$$

où  $N_2$  est une variable aléatoire centrée Gaussienne avec une variance  $\delta_0^{-2}\lambda_1$ .

**Preuve.** Premièrement, on a la décomposition suivante et utilisant la définition de l'estimateur  $\hat{g}_T$  :

$$\begin{aligned}
\frac{1}{\sqrt{T}} \sum_{i=0}^{T-1} \int_0^{\delta_0} Z(i\delta_0 + t) \Phi_1(t) dt &= \frac{1}{\sqrt{T}} \left( \int_0^{\delta_0} Z(t) \Phi_1(t) dt + \int_0^{\delta_0} Z(\delta_0 + t) \Phi_1(t) dt + \right. \\
&\quad \left. \dots + \int_0^{\delta_0} Z((T-1)\delta_0 + t) \Phi_1(t) dt \right) \\
&= \frac{1}{\sqrt{T}} \int_0^{T\delta_0} Z(t) d\mu_{[T]}(t) \\
&= \sqrt{T} \left( \frac{1}{T} \int_0^T Z(t) d\mu_{[T]}(t) + \frac{1}{T} \int_T^{T\delta_0} Z(t) d\mu_{[T]}(t) \right) \\
&= \sqrt{T} \cdot \hat{g}_T + \frac{1}{\sqrt{T}} \int_T^{T\delta_0} Z(t) d\mu_{[T]}(t)
\end{aligned}$$

Pour le second terme de la dernière ligne on va utiliser la même méthode que la preuve du théorème 12 pour démontrer :

$$\left| \frac{1}{\sqrt{T}} \int_T^{T\delta_0} Z(t) d\mu_{[T]}(t) \right| \leq \frac{\|X_{[T\delta_0]}\|_\infty}{[T\delta_0]} \xrightarrow{T \rightarrow \infty} 0 \text{ en probabilité}$$

Ainsi

$$\sqrt{T} \cdot \hat{g}_T = \frac{1}{\delta_0 \sqrt{T}} \sum_{i=0}^{T-1} \int_0^{\delta_0} Z(i\delta_0 + t) \Phi_1(t) dt + o_p(1)$$

Donc

$$\begin{aligned} \sqrt{T}(\hat{g}_T - g(\delta_0)) &= \frac{1}{\delta_0 \sqrt{T}} \sum_{i=0}^{T-1} \int_0^{\delta_0} Z(i\delta_0 + t) \Phi_1(t) dt - \sqrt{T} \cdot \frac{1}{\delta_0} \int_0^{\delta_0} a(t) \Phi_1(t) dt + o_p(1) \\ &= \frac{1}{\delta_0} \cdot \frac{1}{\sqrt{T}} \left[ \sum_{i=0}^{T-1} \int_0^{\delta_0} Z(i\delta_0 + t) \Phi_1(t) dt - T \int_0^{\delta_0} a(t) \Phi_1(t) dt \right] + o_p(1) \\ &= \frac{1}{\delta_0} \left[ \frac{1}{\sqrt{T}} \sum_{i=0}^{T-1} \int_0^{\delta_0} (Z(i\delta_0 + t) - a(t)) \Phi_1(t) dt \right] + o_p(1) \end{aligned}$$

Or pour le premier terme on applique le Théorème 13, on arrive au résultat :

$$\delta_0 \cdot \sqrt{T}(\hat{g}_T - g(\delta_0)) \implies_{T \rightarrow \infty} N_1$$

où  $N_1$  est une variable aléatoire gaussienne centrée de variance  $\lambda_1$  qui est la première valeur propre de l'opérateur  $R_{X_0}$ .

Ainsi

$$\sqrt{T}(\hat{g}_T - g(\delta_0)) \implies_{T \rightarrow \infty} N_2$$

où  $N_2$  est une variable aléatoire gaussienne centrée de variance  $\delta_0^{-2} \lambda_1$ . ■

## 2.5 Estimateur de la période

Nous proposons, dans cette partie, un estimateur de la période  $\delta$  basé sur la méthode de la "distance minimale" . L'estimateur de  $\delta$  est défini par:

$$\hat{\delta}_T = \text{Arg} \min_{\delta \in ]0, \Delta[} |\hat{g}_T - g(\delta)| \quad (2.10)$$

On suppose que le minimum est unique sur  $]0, \Delta[$

avec  $\hat{g}_T$  et  $g(\delta)$  définis respectivement par

$$\hat{g}_T = \frac{1}{T} \int_0^T Z(t) d\mu_{[T]}(t)$$

et

$$g(\delta) := \frac{1}{\delta} \int_0^\delta a(t) d\mu(t)$$

et où  $\mu$  est la mesure de Lebesgue et  $\Delta$  est un nombre strictement positif donné. Nous posons  $\Theta := ]0, \Delta[$  un intervalle ouvert de  $\mathbb{R}$  et pour  $\beta > 0$

$$h(\beta) = \inf_{\delta_0 \in \Theta} \inf_{|\delta - \delta_0| > \beta > 0} |g(\delta) - g(\delta_0)| \quad (2.11)$$

Nous imposons par la suite la condition  $h(\beta) > 0$  qui est une condition d'identifiabilité du paramètre. Nous notons  $\delta_0$  la vraie valeur du paramètre.

**Théorème 15** *Sous les conditions du Théorème 12 et si pour tout  $\beta > 0$ ,  $h(\beta) > 0$ , alors nous avons:*

$$P_{\delta_0} \left( \left| \hat{\delta}_T - \delta_0 \right| > \beta \right) \longrightarrow 0 \quad \text{quand } T \longrightarrow \infty$$

*La convergence est uniforme sur tout intervalle compact de  $\Theta$ .*

**Preuve.** Soit  $\beta > 0$ . De la définition de  $\hat{\delta}_T$  (2.10) et pour  $\delta_0 \in \Theta$  nous avons :

$$\left| \hat{\delta}_T - \delta_0 \right| > \beta \Rightarrow \inf_{|\delta - \delta_0| \leq \beta} |\hat{g}_T - g(\delta)| > \inf_{|\delta - \delta_0| > \beta} |\hat{g}_T - g(\delta)| \quad (2.12)$$

D'une part

$$|\hat{g}_T - g(\delta)| \leq |\hat{g}_T - g(\delta_0)| + |g(\delta_0) - g(\delta)|$$

Donc

$$\inf_{|\delta - \delta_0| \leq \beta} |\hat{g}_T - g(\delta)| \leq |\hat{g}_T - g(\delta_0)| \quad (2.13)$$

D'autre part

$$|g(\delta) - g(\delta_0)| \leq |\hat{g}_T - g(\delta)| + |\hat{g}_T - g(\delta_0)|$$

Par suite

$$|\hat{g}_T - g(\delta)| \geq |g(\delta) - g(\delta_0)| - |\hat{g}_T - g(\delta_0)|$$

Et alors, de la définition de  $h(\beta)$  (2.11) nous obtenons :

$$\begin{aligned} \inf_{|\delta - \delta_0| > \beta} |\hat{g}_T - g(\delta)| &\geq \inf_{|\delta - \delta_0| > \beta} |g(\delta) - g(\delta_0)| - |\hat{g}_T - g(\delta_0)| \quad (2.14) \\ &\geq h(\beta) - |\hat{g}_T - g(\delta_0)| \end{aligned}$$

Par conséquent de (2.12) (2.13) (2.14) nous en déduisons que :

$$\begin{aligned} P_{\delta_0} \left( \left| \hat{\delta}_T - \delta_0 \right| > \beta \right) &\leq P_{\delta_0} \left( \inf_{|\delta - \delta_0| \leq \beta} |\hat{g}_T - g(\delta)| > \inf_{|\delta - \delta_0| > \beta} |\hat{g}_T - g(\delta)| \right) \\ &\leq P_{\delta_0} \left( \inf_{|\delta - \delta_0| \leq \beta} |\hat{g}_T - g(\delta)| > h(\beta) - |\hat{g}_T - g(\delta_0)| \right) \\ &\leq P_{\delta_0} \left( |\hat{g}_T - g(\delta_0)| > \frac{1}{2} h(\beta) \right) \\ &\leq \frac{4E \|\hat{g}_T - g(\delta_0)\|^2}{h^2(\beta)} \end{aligned}$$

Par hypothèse  $h(\beta) > 0$  et par le théorème 12 :  $\hat{g}_T \rightarrow g(\delta)$  presque sûrement et dans  $L^2$ . Donc le dernier majorant tend vers 0 quand  $N \rightarrow \infty$ . D'où le résultat. ■

**Remarque 2.2** la condition  $h(\beta) > 0$  est la condition d'identifiabilité qui est vérifiée quand  $\mu$  est une mesure de Lebesgue et la moyenne  $a(\cdot)$  garde un signe constant.

## 2.6 Loi limite

le résultat suivant donne la loi limite de l'estimateur de la période.

**Théorème 16** Sous les conditions du théorème 12 et si pour tout  $\beta > 0$ ;  $h(\beta) > 0$ , alors nous avons:

$$\sqrt{T} \left( \hat{\delta}_T - \delta_0 \right) \longrightarrow N \quad \text{quand} \quad T \longrightarrow \infty$$

où  $N$  est une v.a. Gaussienne centrée de variance  $\delta_0^{-2} \cdot I^{-2}(\delta_0) \cdot \lambda_1$  telle que  $I(\cdot)$  est la dérivée de la fonction  $g(\cdot)$  et  $\lambda_1 = \frac{\beta_1}{(1-\mu_1^2)}$ . La convergence est uniforme sur  $\Theta$ .

**Preuve.** On applique le théorème 11 de Millar en vérifiant les hypothèses H1, H2, H3. Dans le cas étudié et avec les notations du chapitre 2.2, l'espace de Hilbert H est l'espace  $L^2([0, \delta], \mu)$  et la variable hilbertienne  $\xi_T(\cdot) = \hat{g}_T - g(\cdot)$ .

L'hypothèse d'identifiabilité H1 découle de la condition sur la fonction  $h(\beta)$ .

L'hypothèse H2 est une conséquence du théorème 14 chap 2.4

L'hypothèse H3 découle de la dérivabilité de  $g(\delta)$  par rapport au paramètre  $\delta$ .

Donc a partir du théorème 11, nous avons:

$$\sqrt{T} \left( \hat{\delta}_T - \delta_0 \right) = I^{-1}(\delta_0) \cdot \sqrt{T}(\hat{g}_T - g(\delta_0)) + o_p(1) \quad (2.15)$$

où  $N_1$  est une v.a. gaussienne centrée de variance  $\lambda_1$ . Par suite de la relation (2.15) nous obtenons:

$$\delta_0 \cdot I(\delta_0) \cdot \sqrt{T} \left( \hat{\delta}_T - \delta_0 \right) \Longrightarrow N_1$$

où la fonction  $I(\delta)$  est la dérivée de la fonction de  $g(\delta)$ .

En appliquant le Théorème 14 on en déduit:

$$I^{-1}(\delta_0) \cdot \sqrt{T}(\hat{g}_T - g(\delta_0)) \Longrightarrow_{T \rightarrow \infty} N_3$$

où  $N_3$  est une v.a. gaussienne centrée de variance  $\delta_0^{-2} \cdot I^{-2}(\delta_0) \cdot \lambda_1$ .

D'où le résultat

$$\sqrt{T} \left( \hat{\delta}_T - \delta_0 \right) \Longrightarrow_{T \rightarrow \infty} N_3$$

.

D'où le théorème ■

Nous terminons ce chapitre par le lemme suivant qui permet une estimation de la valeur propre  $\lambda_1$  de  $R_{X_0}$  qui figure dans la loi limite de  $N$ .

**Lemme 3** Soit  $(X_n)$  un ARH(1) processus et  $R_{X_0}$  l'opérateur de covariance de  $X_0$ . On suppose que l'opérateur de covariance  $R_{\varepsilon_0}$  admet  $(\beta_i, \Phi_i)$  comme éléments propres l'opérateur de covariance du mouvement Brownien  $R_{W_0}$ . Si  $\rho$  est un opérateur de Hilbert-Schmidt symétrique et qui commute avec  $R_{X_0}$  avec comme éléments propres  $(\mu_i, \Psi_i)$ ; alors les éléments propres de  $R_{X_0}$  sont  $\left( \frac{\beta_i}{1-\mu_i^2}, \Phi_i \right)$

**Preuve.** Soit  $(X_n)$  un ARH(1):

$$X_n = \rho X_{n-1} + \varepsilon_n$$

Nous avons de ([12])

$$R_{X_0} = \rho R_{X_1 X_0} + R_{\varepsilon_0}$$

On sait aussi que

$$R_{X_0 X_1} = \rho R_{X_0}$$

Alors

$$R_{X_1 X_0} = (R_{X_0 X_1})^* = R_{X_0} \rho$$

Par les hypothèses du lemme

$$\begin{aligned} R_{X_0} &= \rho R_{X_0} \rho + R_{\varepsilon_0} \\ &= R_{X_0} \rho^2 + R_{\varepsilon_0} \end{aligned}$$

Comme  $(\mu_i, \Phi_i)$  sont les éléments propres de  $\rho$  nous avons:

$$\begin{aligned} R_{X_0} (\Phi_i) &= R_{X_0} \rho(\rho(\Phi_i) + R_{\varepsilon_0}(\Phi_i)) \\ &= R_{X_0} \rho(\mu_i \Phi_i) + \beta_i \Phi_i \\ &= \mu_i R_{X_0} \rho(\Phi_i) + \beta_i \Phi_i \\ &= \mu_i^2 R_{X_0} (\Phi_i) + \beta_i \Phi_i \end{aligned}$$

Donc

$$(1 - \mu_i^2) R_{X_0} (\Phi_i) = \beta_i \Phi_i$$

Finalement

$$R_{X_0} (\Phi_i) = \frac{\beta_i}{(1 - \mu_i^2)} \Phi_i \text{ avec } \mu_i \neq \pm 1$$

Donc les éléments propres de  $R_{X_0}$  sont  $\left( \frac{\beta_i}{(1 - \mu_i^2)}, \Phi_i \right)$  ■

**Remarque 2.3** Une même étude peut être faite si on suppose que les éléments propres  $(\lambda_j, \Phi_j)$  de l'opérateur de covariance  $R_{X_0}$  sont inconnus et  $\lambda_j = \frac{\beta_j}{(1 - \mu_j^2)}$  (lemme 1).

Un estimateur naturel des éléments propres de  $R_{X_0}$  sont les éléments propres de l'opérateur de covariance empirique associée (cf. [12] chap 3).

# Chapitre 3

## Simulations et exemples

### 3.1 Simulations

#### 3.1.1 Introduction

Pour simuler des trajectoires de processus ARH(1) fonctionnel et calculer l'estimateur  $\hat{\delta}_T$  du Théorème 15 du chapitre 2.5, nous utilisons la bibliothèque "far" développée par J. Damons et S. Guillas ( Modelization for Functional AutoRegressive processes Package: far Version: 0.6-2 License: LGPL-2.1 version 2.4.1 du logiciel R) .

Plus précisément, on simule notre processus en utilisant les étapes suivantes dont la première est la simulation du mouvement Brownien sur  $[0, n + 1]$ . Ainsi on peut construire  $n + 1$  observations à partir des définitions suivantes:

- **Simulation du bruit blanc**

Nous utiliserons le développement de Karhunen- Loève sur  $[0, T]$  du mouvement Brownien suivant:

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

où les v.a.  $Y_j^*$  sont i.i.d de loi  $N(0, 1)$ , qui va nous permettre de définir un bruit blanc  $(\varepsilon_n(t), n = 0, 1, \dots, t \in [0, \delta])$  par:

$$\varepsilon_n(\omega) = W_{n+s}(\omega) - W_n(\omega) , s \in [0, \delta]$$

défini sur l'intervalle  $[0, \delta]$ . Pour la simulation, on aura une approximation de la somme infinie par une somme finie comme dans [30]

- **La moyenne  $a(t)$**

Nous choisirons plusieurs types de fonctions périodiques et plusieurs périodes pour pouvoir vérifier si l'estimateur convergera toujours vers la vraie période de l'estimateur

Exemples: Pour  $\delta_0 = 2\pi$  avec  $a(t) = \sin(t)$ ,  $a(t) = \cos(t)$ ,.....  
 Pour  $\delta_0 = \pi$  avec  $a(t) = \sin(2t)$ ,.....  
 Pour  $\delta_0 = 1$  avec  $a(t) = \sin(2\pi t)$ ,.....  
 Pour  $\delta_0 = 2$  avec  $a(t) = t$  ( périodique linéaire),.....

- **L'opérateur  $\rho$**

Nous considérons le cas d'un opérateur  $\rho$  à noyau  $K$  pair, symétrique, périodique de période  $\delta$  et de norme  $L^2$  inférieure à 1, défini par:

$$\rho(f)(t) = \int_0^{\delta} k(t-s)f(s)ds$$

nous prenons par exemple  $K(t) = \frac{1}{10^2} \cos(t)$ .

**Remarques:**

- 1 Dans nos simulations on a aussi pris le programme prédéfini dans le Logiciel R, mais qu'on a adapté car la définition de la décomposition de Karhunen-Loève utilisée était sur  $[0, 1]$
- 2 Valeur initiale de  $X_0 = 0$

- **Simulation de l'ARH(1)**

Pour la simulation d'un ARH(1) on prend comme valeur initiale la valeur  $X_0 = 0$  et  $(\varepsilon_n)$  est le bruit blanc dans la base de Karhunen Loève ensuite, nous effectuons un changement de base pour ce bruit blanc en utilisant la fonction BaseK2BaseC (du Package far) pour avoir l'écriture de  $(\varepsilon_n)$  dans la base canonique et nous posons  $X_1 = \varepsilon_1$ .

Soient  $(X_i^j)$   $j=1, \dots, m$  ( $m$  étant le nombre de point de discrétisation) les coordonnées de  $X_i$  dans la base canonique  $i=1, \dots, n$ .

Soient  $(a_i^j)$   $j=1, \dots, m$  les coordonnées de  $a_i$ ,  $i=1, \dots, n$  où les  $(a_i^j)$  représentent les valeurs de la fonction  $a(t)$  aux points de discrétisation  $m$  quand  $t \in [0, i\delta]$ ; On a:

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



car

$$X_n - a = \rho(X_{n-1} - a) + \varepsilon_n$$

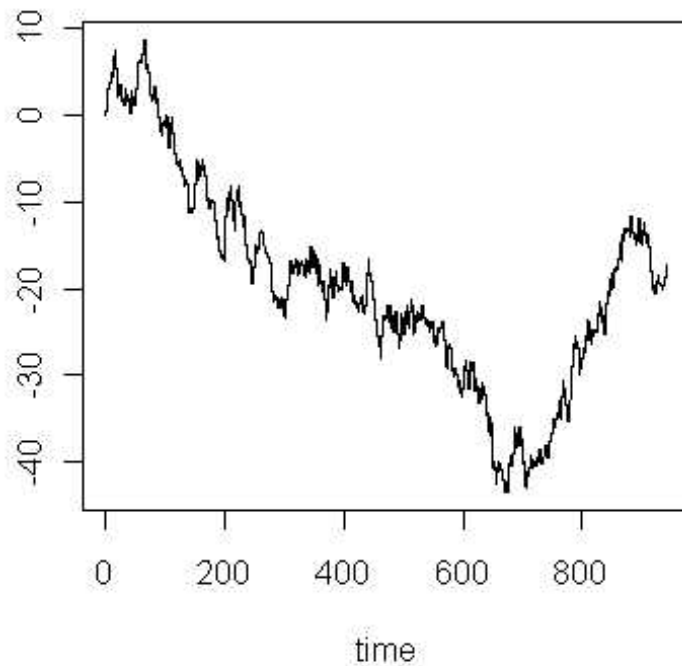
et

$$X_n := Z(n\delta + t) \quad t \in [0, \delta]$$

On choisit comme valeur initiale la valeur  $X_0 = 0$ ,  $X_1 = \varepsilon_1$  et  $a(t) = \sin(t)$  pour simuler un processus ARH(1).

fr

### exemple d'un ARH(1) (m=300,n=150)



2.jpg

Figure 2: Exemple d'un ARH(1) (m=300, n=150)

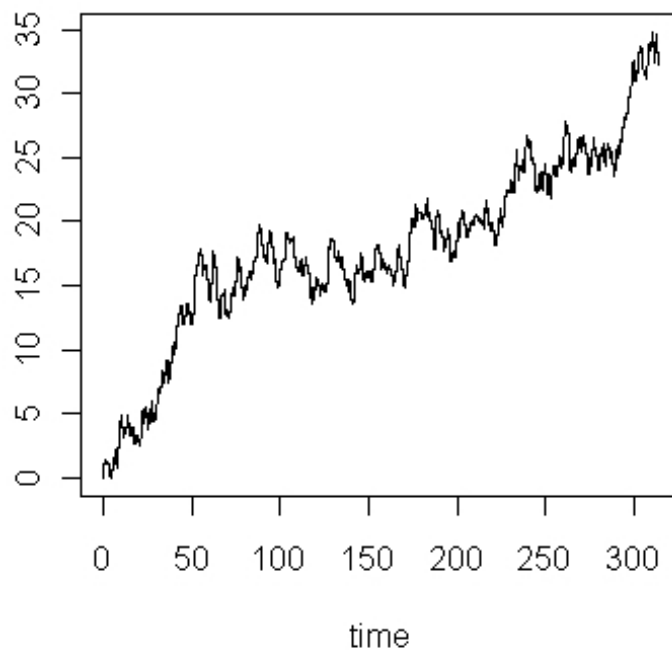
On choisit pour un deuxième exemple comme valeur initiale la valeur

$X_0 = 0$ ,  $X_1 = \varepsilon_1$  et  $a(t) = \sin(t)$  pour simuler un processus ARH(1).

pi

fr

### Exemple d'un ARH(1) (m=300,n=100)

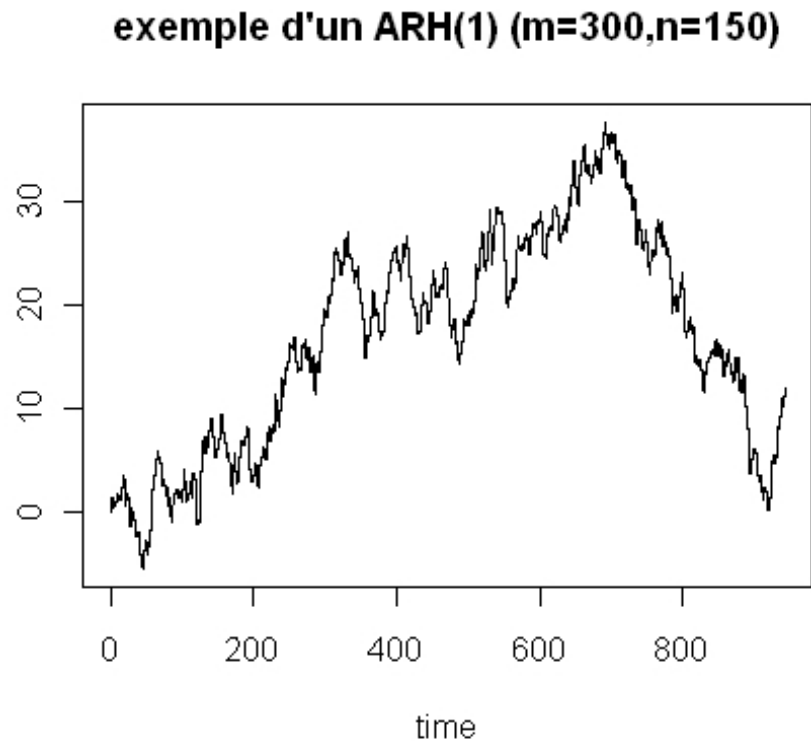


<sup>3.jpg</sup>  
Figure 3: Exemple d'un ARH(1) (m=300, n=100) de priode  $\pi$ .

On choisira  $a(t) = \cos(t)$  pour simuler un processus ARH(1) (avec

les mêmes conditions initiales).

cos



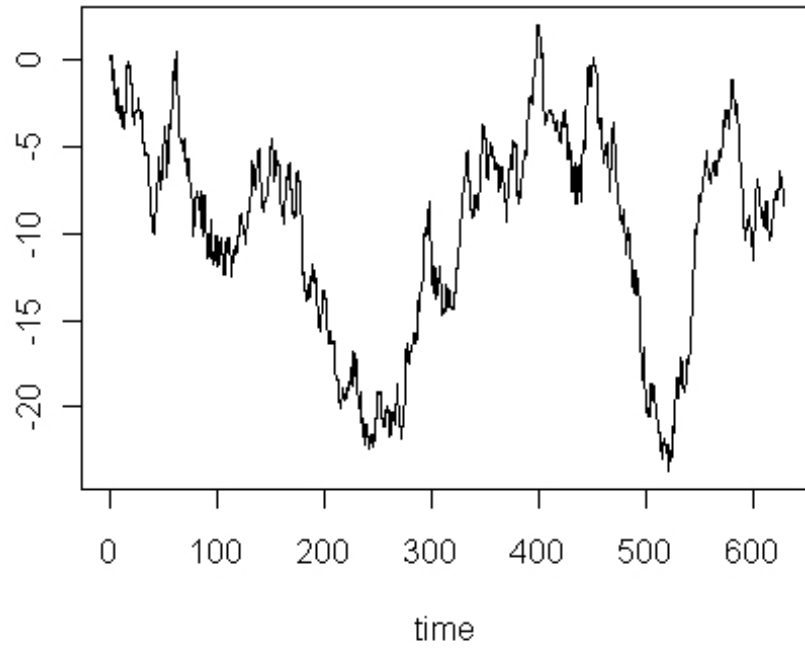
4.jpg  
Figure 4: Exemple d'un ARH(1) (m=300, n=100) avec  $a(t) = \cos(t)$ .

On s'intéressera aussi à la simulation d'un ARH(1), sous les mêmes

conditions, mais en faisant varier  $m$  (le nombre de discrétisation).

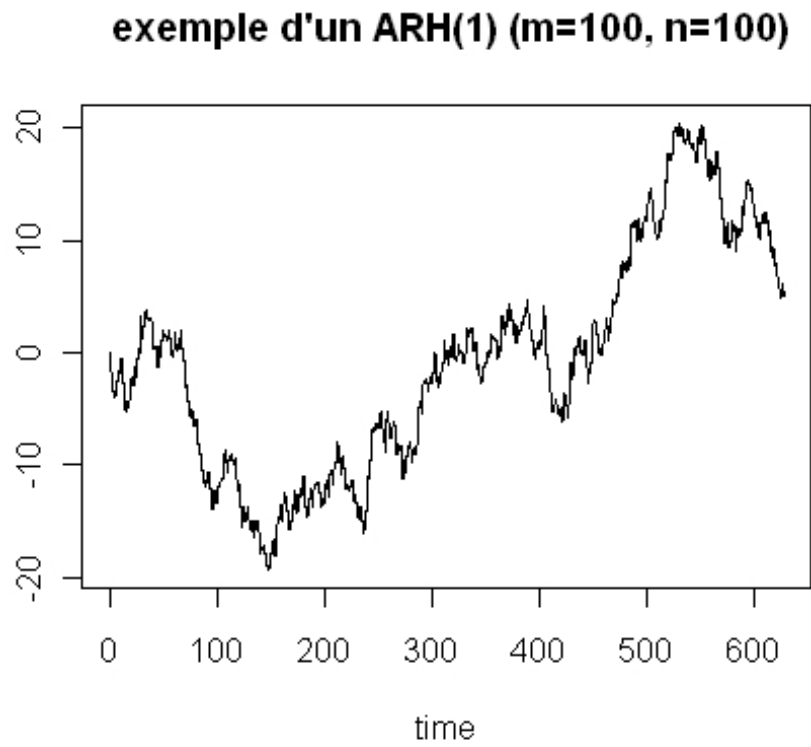
$m=50$

**exemple d'un ARH(1) ( $m=50, n=100$ )**



*5.jpg*

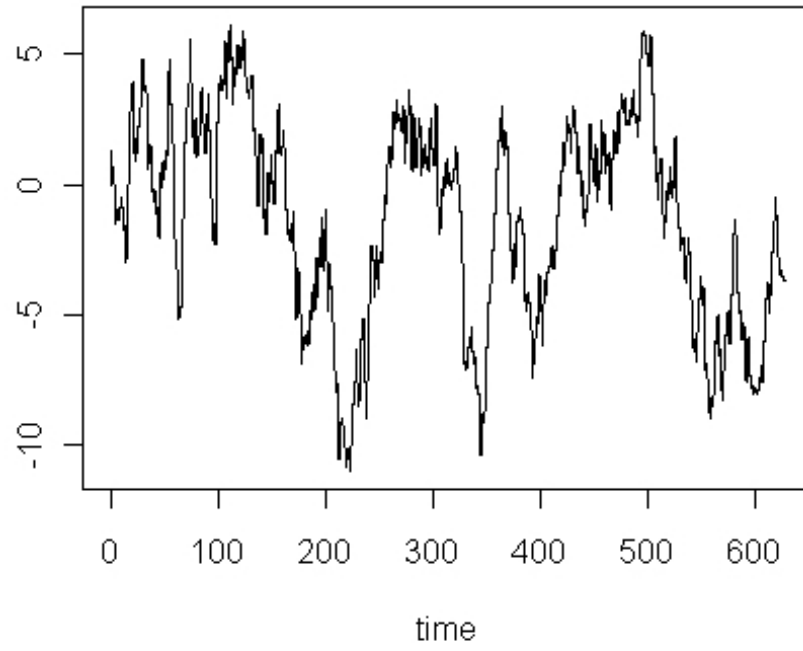
m100



<sup>6.jpg</sup>  
Figure 5: Exemple d'un ARH(1) (m=50 et m=100, n=100) de priode  $2\pi$ .

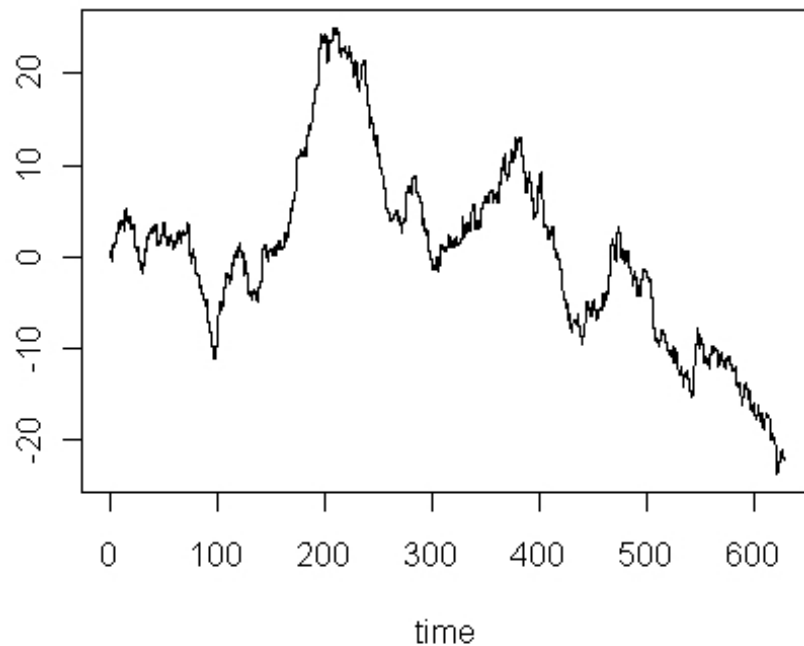
*m150*

**exemple d'un ARH(1) (m=150, n=100)**



*7.jpg*

m200

**exemple d'un ARH(1) (m=200, n=100)**

*s.jpg*  
 Figure 6: Exemple d'un ARH(1) (m=150 et m=200, n=100) de période  $2\pi$ .

### 3.1.2 Simulation du comportement de l'estimateur de la période

#### Description

Nous utiliserons les  $n$  observations du processus pour calculer l'estimateur  $\hat{g}_T$  donné par:

$$\hat{g}_T = \frac{1}{T} \int_0^T Z(t) d\mu_{[T]}(t) \quad (3.1)$$

et pour chaque point de notre discrétisation de l'intervalle on calcule

$$g(\delta) := \frac{1}{\delta} \int_0^\delta a(t) d\mu(t) \quad \text{pour } \delta \in [0, \Delta]$$

où  $\mu$  est définie par la mesure de Lebesgue.

Puis en fixant  $\Delta$  tel que la vraie période  $\delta_0 \in ]0, \Delta[$  on simule  $\hat{\delta}_T$  l'estimateur donné par:

$$\hat{\delta}_T = \text{Arg} \min_{\delta \in ]0, \Delta[} |\hat{g}_T - g(\delta)|$$

Quand  $T$  tend vers l'infini ( $\infty$ ) les simulations montrent que l'estimateur converge vers la vraie valeur de la période  $\delta_0 = 2\pi$  (cette simulation est choisie pour la fonction  $a(t) = \sin(t)$ ) et  $\mu$  mesure de Lebesgue.

### comportement de l'estimateur de delta en fonction de T

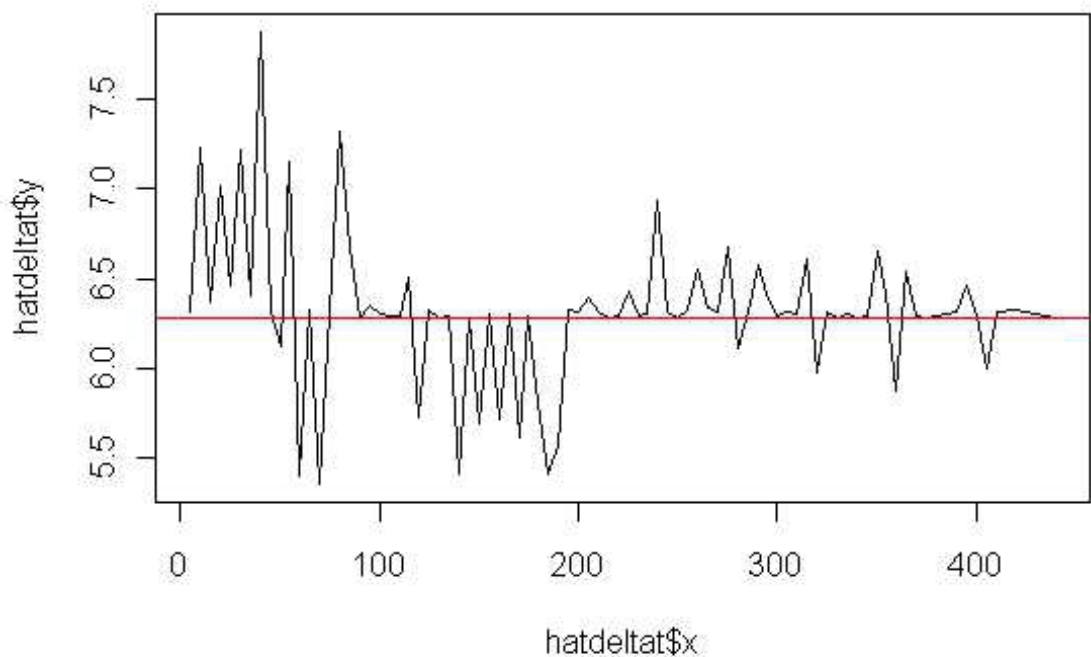


Figure 7: Comportement de l'estimateur de delta en fonction de T (mesure de Lebesgue).

Le graphe suivant montre le comportement de l'estimateur de la période  $\hat{\delta}_T$  en fonction de T, où l'estimateur  $\hat{g}_T$  défini par (3.1) est défini



par la mesure  $\mu$  ayant la densité  $\Phi_1 = \sqrt{\frac{1}{\pi}} \sin \left[ \frac{t}{4} \right]$

### comportement de l'estimateur de delta en fonction de T

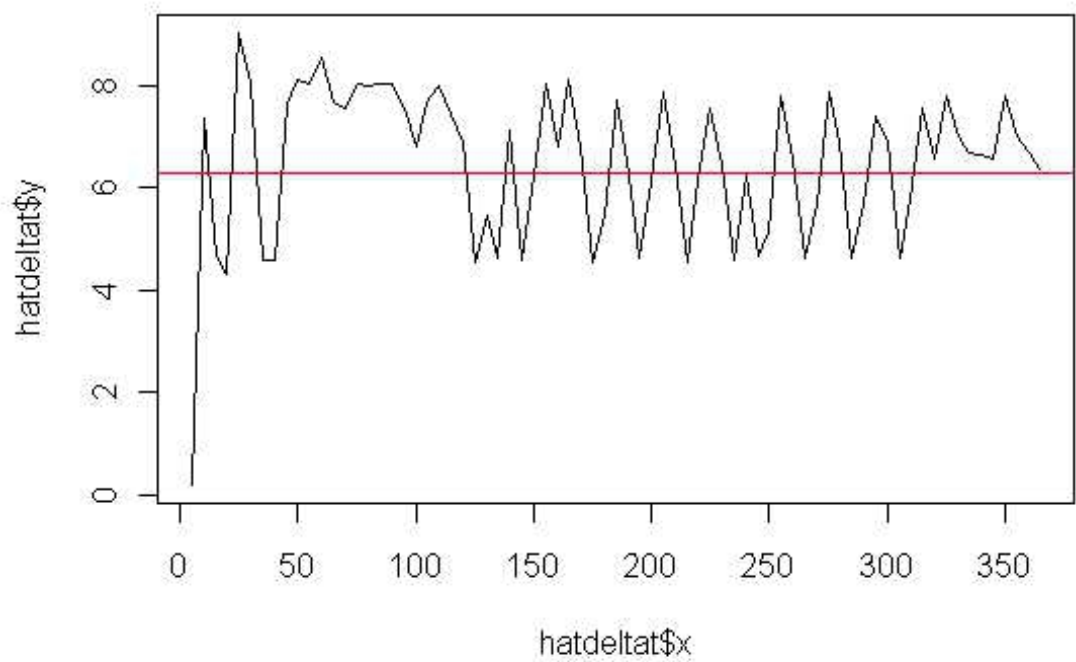


Figure 8: Comportement de l'estimateur de delta en fonction de T (mesure densit).

Le graphe suivant présente l'estimateur à noyau (en utilisant l'Epanechnikov) de la loi limite de l'estimateur  $\hat{\delta}_T$  avec la mesure de Lebesgue qui

définit  $\hat{g}_T$  :

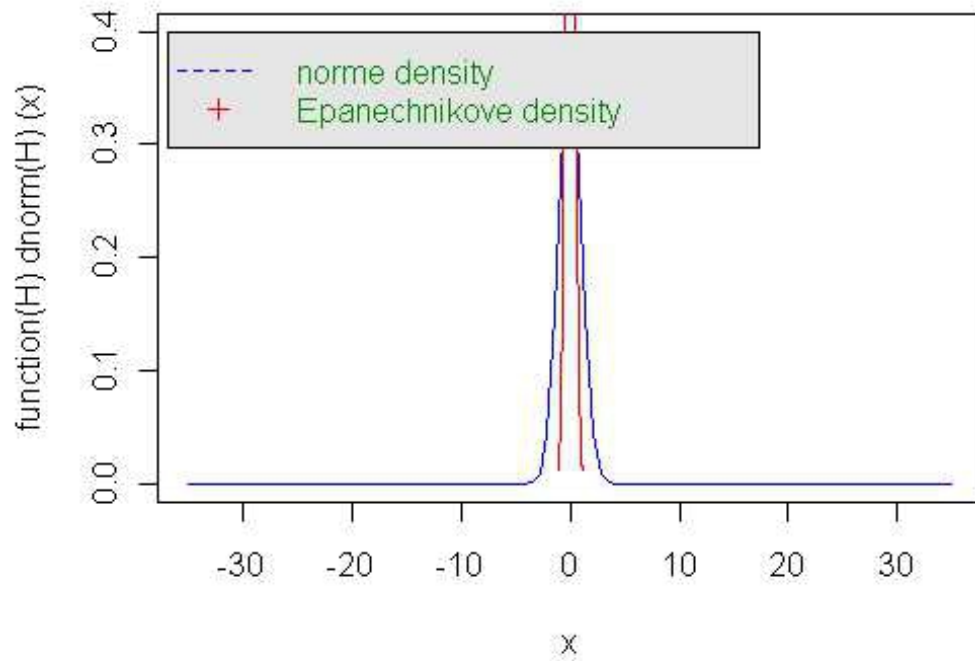


Figure 9: Estimateur noyau de la loi limite.

Le graphe suivant montre l'estimateur à noyau (en utilisant l'Epanechnikov) de la loi limite de l'estimateur  $\hat{\delta}_T$  ayant la densité  $\Phi_1$  définie en  $\hat{g}_T$

:

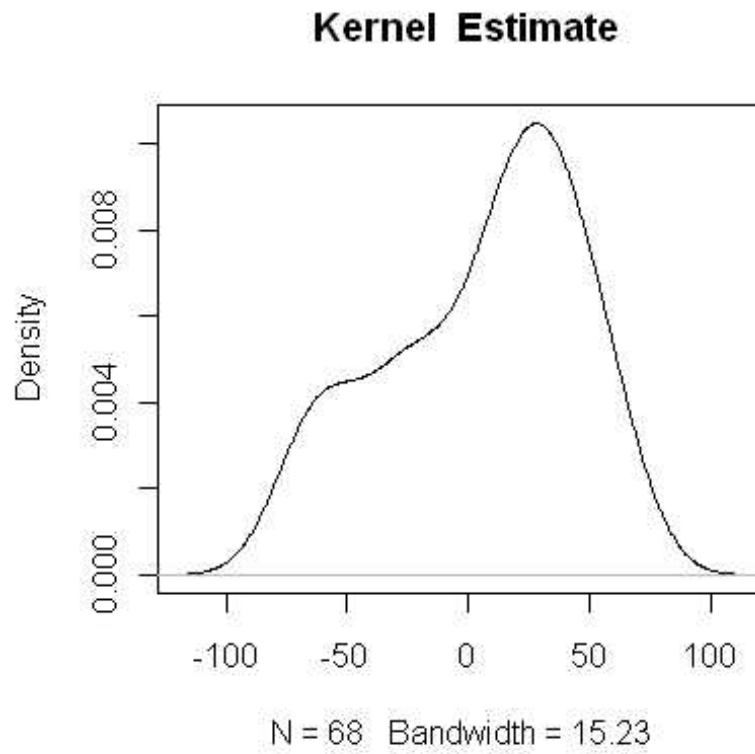
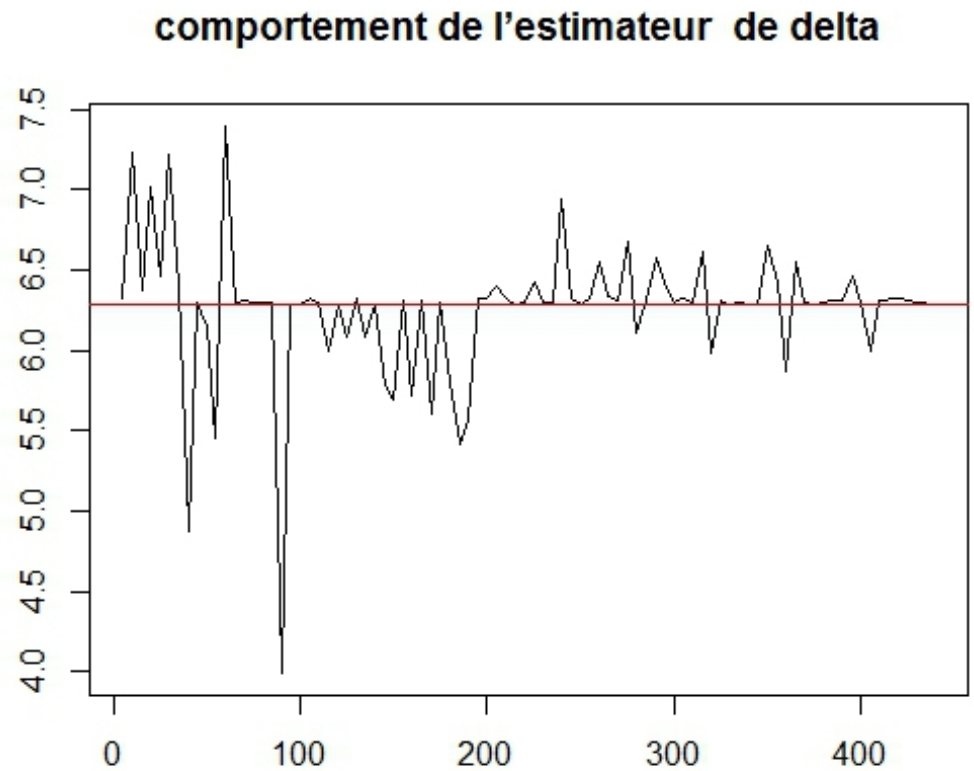


Figure 10: Estimateur noyau de la loi limite avec une mesure densit.

Pour  $a(t) = \cos(t)$ , on peut remarquer le bon comportement de l'estimateur (pour  $T$  tendant vers l'infini ( $\infty$ )) l'estimateur converge

vers la vraie valeur de la période  $\delta_0 = 2\pi$  )

cos



<sup>13.jpg</sup> Figure 11: Comportement de l'estimateur de delta en fonction de T  
( $a(t) = \cos(t)$ )

Le tableau suivant présente les estimations de  $\hat{\delta}_T$  quand le nombre de répliques des trajectoires augmente (ici  $\delta_0 = 2\pi = 6,283185$ ).

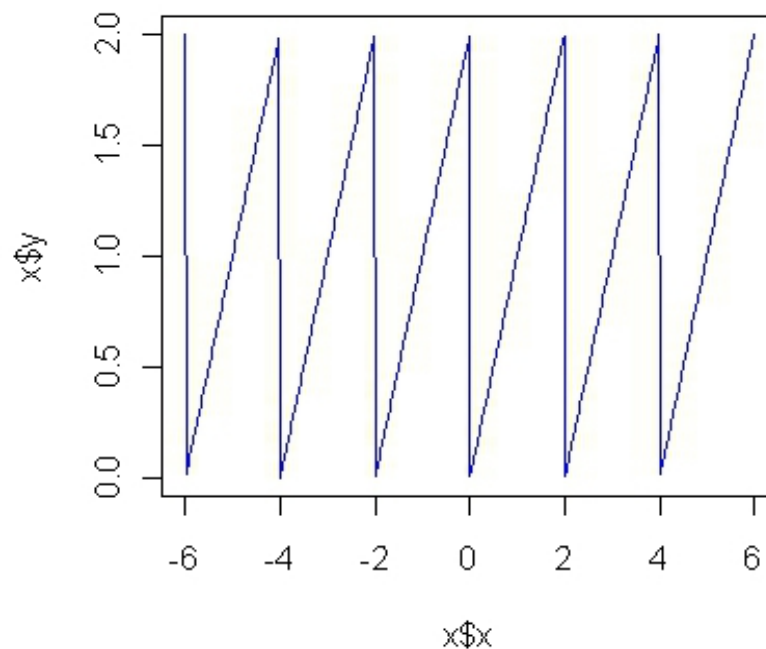
n=multiple de la periode	nombre de répliques	moyenne	l'écart type
100	20	6.431	0.3134
100	40	6.360	0.2927
100	60	6.332	0.2956
100	80	6.327	0.3006
150	20	6.290	0.3424
150	40	6.290	0.3370
150	60	6.286	0.3112
150	80	6.285	0.3049

On choisit, ensuite une fonction  $a(t)$  périodique linéaire :

$$a(t) = t \quad , \quad \delta = 2$$

x

### Fonction linéaire périodique : x



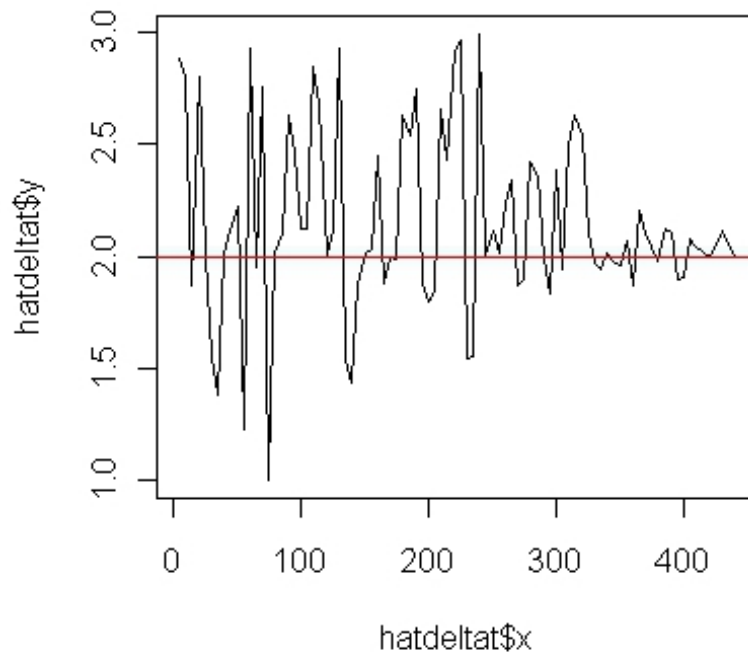
<sup>14.jpg</sup>  
Figure 12: Graphe de la representation priodique de la fonction linéaire.

En réécrivant le programme, le comportement de l'estimateur nous donne le graphe suivant:

lineaire

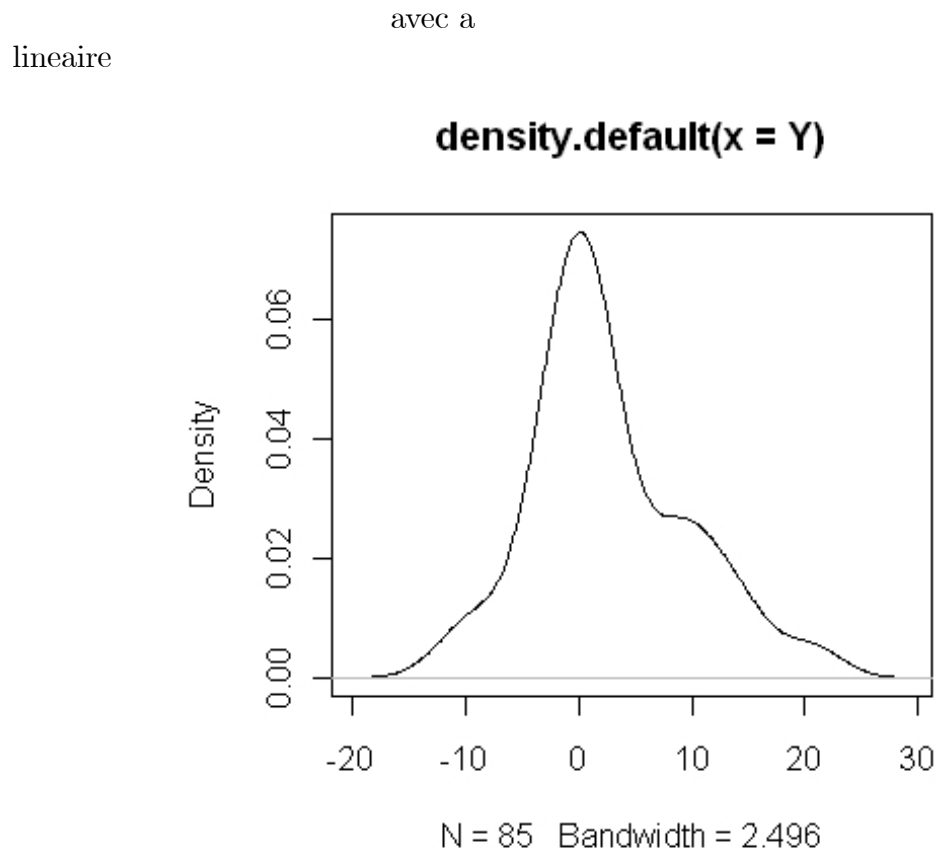
avec a

### comportement de l'estimateur de delta



<sup>15.jpg</sup>  
Figure 13: Comportement de l'estimateur pour  $a(t)=t$ .

Le graphe de l'estimateur à noyau de la densité



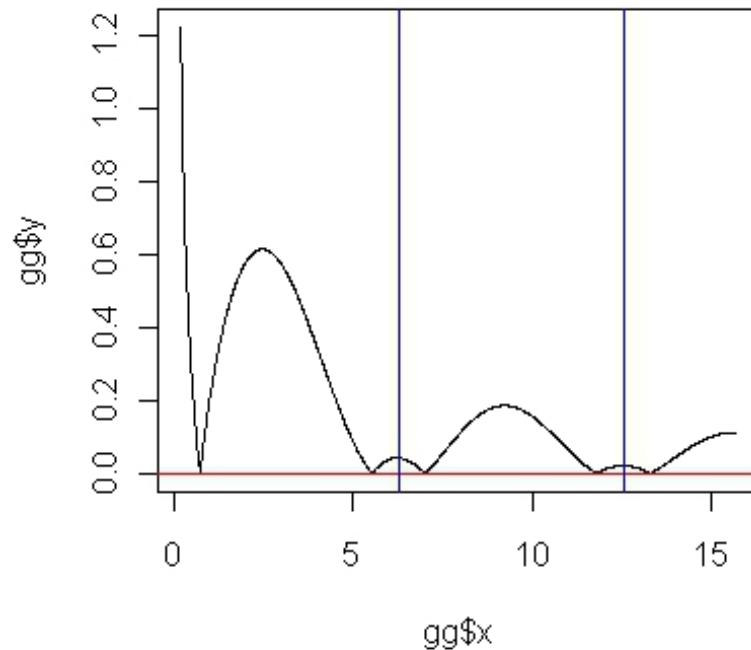
<sup>16.jpg</sup>  
Figure 14: Estimateur noyau de la densité de la loi limite pour  $a(t)=t$ .

En conclusion, toutes les estimations fournies par les simulations donnent une bonne approximation de la valeur de la période  $\delta$  donnée. Si on change la moyenne  $a$  et le noyau  $k(s, t)$  les simulations confirment les mêmes conclusions.

**Remarque 3.1** *Pour mieux prédire le comportement de notre estimateur, nous avons pris un intervalle  $]0, \Delta[$  qui contient plusieurs périodes (des multiples de la période), nous avons observé que notre estimateur  $\hat{\delta}_T$  converge vers la plus grande période (ie le plus grand multiple).*

Ainsi, le graphe suivant nous montre d'une part le comportement de la fonction  $|\hat{g}_T - g(\delta)|$  sur l'intervalle  $]0, \Delta[ = ]0, 5\pi[$  qui contient plusieurs multiples de la période  $\delta = 2\pi$  et d'autre part nous remarquons que le minimum de cette fonction est atteint sur le plus grand multiple de la période  $\hat{\delta}_T = 12.88919$ .

5pi



17.jpg

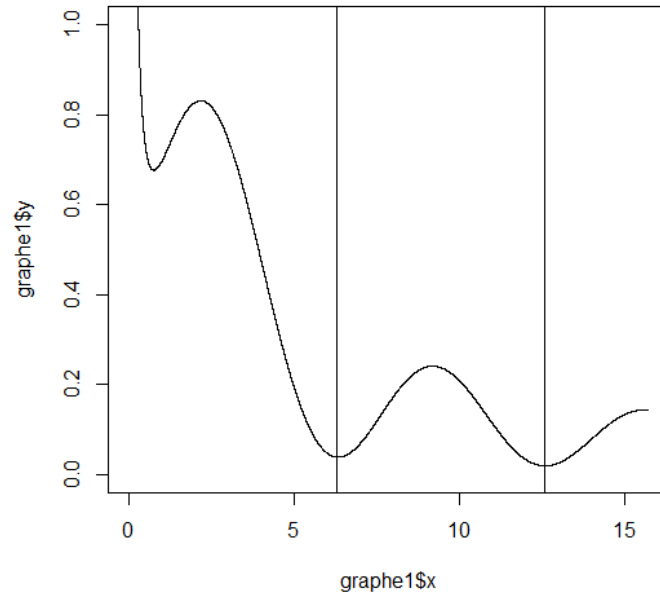
Figure 15

Pour l'instant, le résultat de cette simulation est un constat, nous pensons qu'il sera élucidé durant les prochaines recherches théoriques.

Sur cette deuxième simulation où  $]0, \Delta[ = ]0, 5\pi[$  contient plusieurs multiples de la période  $\delta = 2\pi$ , nous remarquons que le minimum de cette fonction est



atteint sur le plus grand multiple de la période  $\hat{\delta}_T = 12.5712$ .



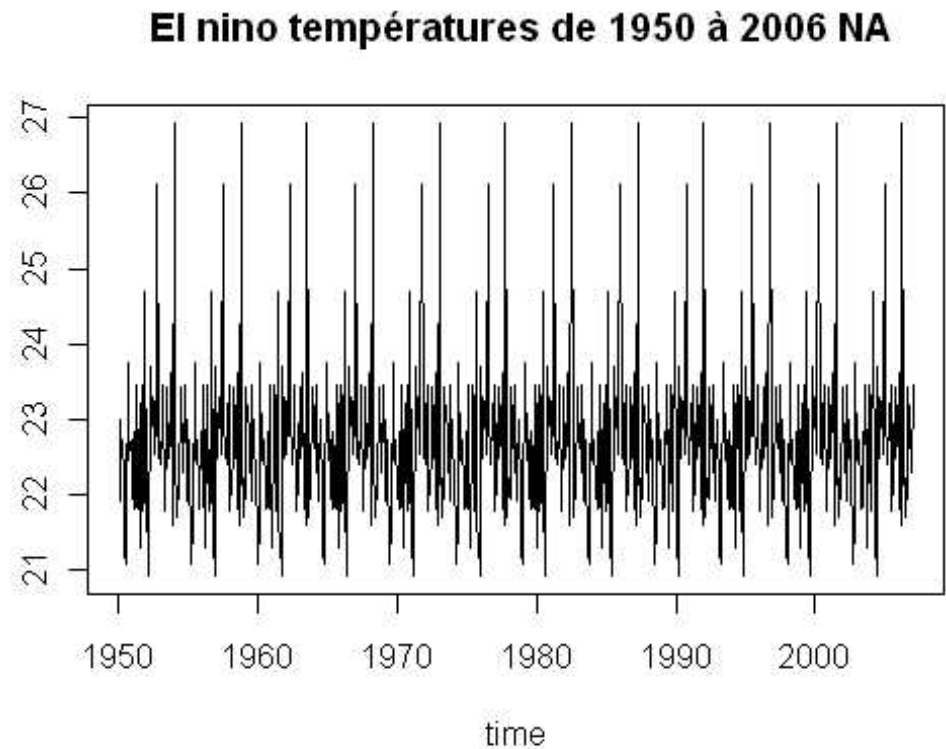
## 3.2 Exemples

### 3.2.1 Série climatologique ENSO.

Nous considérons 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 Niño (EN) est la composante océan dans ENSO tandis que Southern Oscillation (SO) est la composante atmosphère . Un index mesurant la variabilité d'EL Niño est fourni par les températures à la surface de l'océan ramenées à une moyenne observée dans le domaine Niño-1 (5oS - 5oN, 150oW - 90oW). 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 [3]

[11] [28]).

fr nino



19.jpg

Figure 15: Graphe des températures d'El niño

Pour l'estimation de la vraie période, on choisit :  $\Theta := ]0, \Delta[ = ]0, 17[$ . Dans la définition de  $g$  (2.5) on prédéfinera la moyenne  $a$  et la mesure  $\mu$  et naturellement notre choix va se porter sur les fonctions cos ou sin. Mais comme la fonction sin est nulle en 0 elle sera donc éliminée. L'estimateur  $\hat{g}_T$  étant donné par (2.4) et l'estimateur  $\hat{\delta}_T$  par (2.10).

*1<sup>er</sup> cas.*  $a(t) = 23.02 * \cos(t * \frac{\pi}{6})$  et  $\mu$  la mesure de Lebesgue.

Un choix tel que la moyenne  $a$  soit 12-periodique est évident, voire naturel pour un phénomène climatique. Le coefficient 23.02 est la valeur moyenne des températures mensuelles ( Toutefois d'autres choix sont possibles !)

Le résultat des simulations donne:

$$\begin{aligned}\hat{\delta}_T &= 12.00 && \text{Pour une discrétisation de pas } 0.01 \text{ de } ]0, \Delta[ = ]0, 17[ \\ \hat{\delta}_T &= 12.003 && \text{Pour une discrétisation de pas } 0.001 \text{ de } ]0, \Delta[ = ]0, 17[\end{aligned}$$

La vraie période étant 12, on remarque que l'estimateur détecte bien cette valeur avec une erreur de  $3^{\circ}/_{\infty}$

*2<sup>nd</sup> cas.*  $a(t) = 23 * \cos(t * \frac{\pi}{6})$  et  $\mu$  a une densité  $\Phi_1$  (l'élément propre de  $R_{W_0}$ ) :

$$\hat{\delta}_T = 12.561 \quad \text{Pour une discrétisation de pas } 0.001 \text{ de } ]0, \Delta[ = ]0, 17[$$

Si on perturbe la moyenne en choisissant  $a(t)$  une fonction telle que sa période ne soit pas 12 ( la vraie période du phénomène) et  $\mu$  la mesure de Lebesgue, on obtient les résultats suivants:

1. Si  $a(t) = 23 * \cos(t * \frac{2\pi}{8})$  alors  $\hat{\delta}_T = 11.995$ .
2. Si  $a(t) = 23 * \cos(t * \frac{2\pi}{13})$  alors  $\hat{\delta}_T = 13.002$ .

On remarque que l'estimateur détecte bien la vraie période dans le premier cas mais avec une erreur appréciable au second cas. Si  $\mu$  a une densité  $\Phi_1$  ( l'élément propre de  $R_{W_0}$ ) l'estimateur  $\hat{\delta}_T$  converge très lentement vers la vraie période. Ainsi il est difficile de donner des conclusions raisonnables.

### 3.2.2 Prédiction de la température à Alger

Dans cette deuxième partie nous présentons des résultats de prédiction de la température moyenne à Alger pour la période de

1970- 2004.

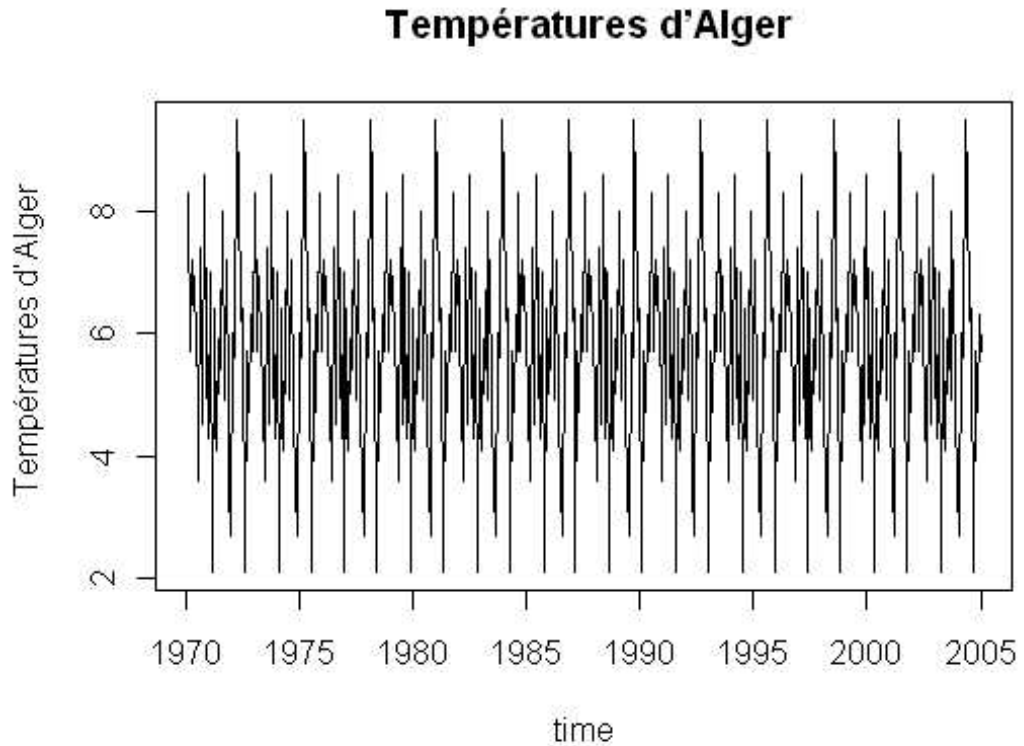


Figure 16: Graphe des températures d'Alger.

On choisit  $\Theta := ]0, \Delta[ = ]0, 17[$  et  $a(t) = 11.86 * \cos(t * \frac{\pi}{6})$  avec  $\mu$  la mesure de Lebesgue (où le coefficient 11.86 est la valeur moyenne des températures d'Alger).

On obtient:

$$\begin{aligned} \hat{\delta}_T &= 11.99 && \text{Pour une discrétisation de pas } 0.01 \text{ de } ]0, \Delta[ = ]0, 17[ \\ \hat{\delta}_T &= 11.994 && \text{Pour une discrétisation de pas } 0.001 \text{ de } ]0, \Delta[ = ]0, 17[ \end{aligned}$$

La vraie période étant 12, on remarque que l'estimateur détecte bien cette valeur avec une erreur de  $6^{\circ}/_{\infty}$ .

Si on perturbe la moyenne en choisissant la  $a(t)$  une fonction telle que sa période ne soit pas 12 ( la vraie période du phénomène) et  $\mu$  la mesure de Lebesgue, on obtient les résultats suivants:

1. Si  $a(t) = 11.86 * \cos(t * \frac{2\pi}{8})$  alors  $\hat{\delta}_T = 12.000$

2. Si  $a(t) = 11.86 * \cos(t * \frac{2\pi}{13})$  alors  $\hat{\delta}_T = 12.990$

On remarque aussi que l'estimateur détecte la vraie période dans ces cas.

**Remarque 3.2** *Pour tester la "robustesse" de l'estimateur ; on applique de "petite perturbation" sur  $\delta$  . On va sous estimer la période  $\delta$  sur un voisinage "unilatéral" à gauche ]8, 12[ et on remarque que*

$$\hat{\delta}_T \longrightarrow 12$$

# Chapitre 4

## Conclusion

Nous concluons, en faisant remarquer que dans toutes les simulations numériques et les exemples étudiés, notre estimateur de la période donne de très bons résultats, en détectant la vraie période. Nous noterons aussi que dans les simulations numériques et exemples donnés, l'estimateur a un bon comportement quand  $T$  devient très grand. D'un autre côté plusieurs choix de la moyenne "a" et du noyau  $k(s,t)$  sont possibles dans d'autres simulations, dans notre étude nous nous sommes limités à quelques exemples.

Pour des recherches futures, nous essayerons de déterminer la vitesse de convergence de notre estimateur et de voir d'autres applications intéressantes dans les domaines : génétique, environnemental,...

# Bibliographie

- [1] Allam, A., Mourid, T. Geometric absolute regularity of Banach space valued autoregressive processes. *Statistics and Probability Letters* 60 (2002) 241-252.
- [2] Allam, A., Mourid, T. Propriétés de mélange des processus autorégressifs banachiques. *C.R. Acad. Sci. Paris, Série I* 332 (2001) 363-368.
- [3] Antoniadis, A. and Sapatinas, T. Wavelet methods for continuous- time prediction using Hilbert-valued autoregressive processes. *J. Multivariate Anal.*, 87, 133-158. 2003.
- [4] Bensmain, N. Mourid, T. Estimateur sieves de l'opérateur d'un processus ARH(1). *Comptes Rendus Acad. Sci. Paris. t 332 Serie I.* 2001.
- [5] Bensmain, N. Mourid, T. Sieves estimator of the operator of a functional autoregressive process. *Statistics and Probability Letters* 76,93-108. 2006.
- [6] Benyelles, W. Mourid, T. On a minimum distance estimate of the period in Functional Autoregressive Processes . *Journal of Applied statistics* . 2012. DOI:10.1080/02664763.2012.668178.
- [7] Benyelles, W. Mourid, T. Estimation de la période d'un processus à temps continu à représentation autorégressive. *Annales ISUP* Vol. 46 Fasc. 1-2 p. 89-101. 2002.
- [8] Benyelles, W. Mourid, T. Estimation de la période d'un processus à temps continu à représentation autorégressive. *Comptes Rendus de l'Académie des sciences (C.R.A.S).* Paris, t.333, Série I, p.245-248, Juin 2001.
- [9] Bernard, P. (1997) . *Analyse de signaux physiologiques* . Mémoire Univ. Cathol. Angers.
- [10] Besse, P. and Cardot, H. Approximation spline de la prévision d'un processus fonctionne autorégressivf d'ordre 1. *Canad. J. of Statistics* 24 467-487. December 1996.



- 
- [11] Besse, P., Cardot, H. and Stephenson, D.B. Autoregressive Forecasting of Some Functional Climatic Variations. *Scandinavian Journal of Statistics* 27 (4) 673–687. December 2000.
- [12] Bosq, D. (2000). *Linear process in function spaces. Theory and Application*. Lectures Notes in Statistics, Vol 149. Springer-Verlag.
- [13] 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.
- [14] Cavallini, A, G.C. Montanari, M. Loggini, O.Lessi, and M.cacciari (1994). Nonparametric prediction of harmonic levels in electrical networks. *Proceed. IEEE ICHPS VI, Bologna, 165-171*.
- [15] Davidian, M., Lin, X. and Wang, J. Introduction to Emerging issues in longitudinal and functional data analysis. *Statistica. Sinica* 14 (2004), no. 3, 613–614.
- [16] Dunford, N. and Schwartz, J.T. (1958). *Linear operators*, Willey New-York.
- [17] Ferraty, F. and Romain, Y. (2011). *The Oxford Handbook of Functional Data Analysis*. Oxford University Press.
- [18] Ferraty, F. and Vieu, P. (2006). *Nonparametric functional data analysis. Theory and Practice*. Springer-Verlag. MR2229687.
- [19] Ferraty, F., Laksaci, A., Tadj, A. and Vieu, P. Kernel regression with functional response . *Electronic Journal of Statistics* Vol. 5 (2011) 159–171
- [20] Ferraty, F., Vieu, P. and Delsol, L. (2011) Structural test in regression on functional variables. *Journal of Multivariate Analysis*. 102, p.422-447.
- [21] González Manteiga, W. and Vieu, P. Statistics for functional data. *Computational Statistics & Data Analysis*, Anal. 51 (2007), no. 10, 4788–4792.
- [22] Grenander, U. *Abstract inference*. ed Wiley, New York 1981.
- [23] Guillas, S. Thèse de Doctorat En Sciences, Univrsité Paris 6.(2001).
- [24] Hoffman- Jorgensen and G.Pisier(1976). The law of large numbers and the central theorem in banach spaces, *Ann.Proba*, n°4, p. 587-599.
- [25] Kutoyants, Yu. (1991). Minimum distance parameter estimation for diffusion type observation. *C. R.Acad.Sci. Paris, Ser.I*. p 637-642.
- [26] Martin D.E.K (1982) *Estimation of the minimal period of periodically correlated sequences*. Phd.Thesis. University of Maryland.

- 
- [27] Millar, P.W. (1984). A general approach to the optimality of minimum distance estimation. *Trans. Amer. Soc.* 286, 1, p. 377-418.
- [28] Mourid, T. and Benyelles, W. Estimation de la période de la représentation autorégressive d'un processus à temps continu, *Annales de l'I.S.U.P.*, Volume XXXVI-Fascicule 1-2; p. 89-101. 2002.
- [29] Parr, C. (1981) Minimum distance estimation. *Bibliography. Commun. Statist. Theory Math.* A10 (12) 1205-1224.
- [30] Pumo, B. Prediction of continuous time processes by  $C_{[0,1]}$  - valued autoregressive process. *Statist. Inf. Stoch. Proc. I* . 139-153. 1998.
- [31] Ramsay, J.O. and Silverman, B.W. *Applied Functional Data Analysis*, Springer-Verlag New York, 2002.
- [32] Ramsay, J.O. and Silverman, B.W. *Functional Data Analysis*, 2nd ed. Springer, New-York, 2005.
- [33] Valderrama, M. (2007). Introduction to the special issue on modelling functional data in practice. *Comput. Statist.* 22, 331-334.