

Le livre « *Modélisation Stochastique et Statistique* » est issue des Actes du colloque International *MSS'2004* qui a eu lieu en 2004 à l'Université des Sciences et Technologies *USTHB* Alger. Cette rencontre scientifique de haut niveau avait regroupé des chercheurs universitaires et des experts praticiens du calcul stochastique et statistique et ses applications dans les domaines socio-économique, industriel et environnemental. Par les différents thèmes abordés, cette rencontre a constitué un cadre idéal pour débattre et discuter des développements récents. D'autre part, cette rencontre a offert une occasion de rapprochement entre académiciens et professionnels en matière d'échange d'idées et d'expériences dans le domaine de l'utilisation de l'outil stochastique et statistique en analyse, modélisation, simulation et prospection pour l'aide à la prise de décision. Ce colloque *MSS'2004* a été organisé et soutenu financièrement par les laboratoires, *MSTD* de la faculté de Mathématiques de l'*USTHB* Alger et l'Unité de Recherche *LaMOS* de l'Université de Béjaïa. L'édition 2018 du livre « *Modélisation Stochastique et Statistique* » regroupe les 37 articles de recherche acceptés après expertise internationale.

Pr Djamil AÏSSANI est Directeur de l'Unité de Recherche *LaMOS*, Université de Béjaïa,
Pr Kamal BOUKHETALA est Doyen de la Faculté des Mathématiques de l'*USTHB* Alger.

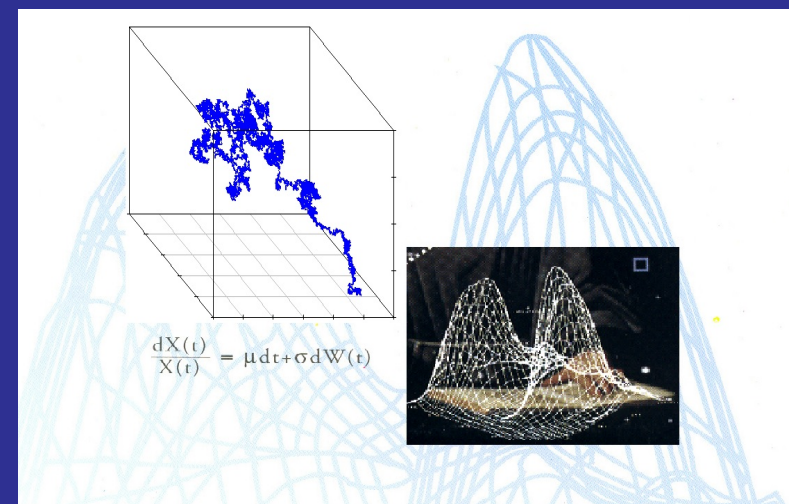
Dépôt légal 2088-2007 :



MSS'2004 — Edition 2018

MSS'2004

Modélisation Stochastique et Statistique

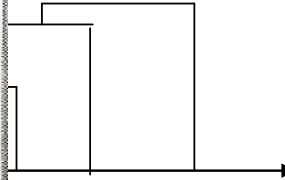
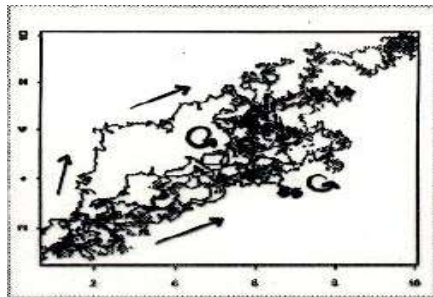


Edité par

Prof. Aissani Djamil , (LaMOS)

Prof. Boukhetala Kamal , (MSTD)

Modélisation Stochastique *et Statistique*



$$\frac{dX(t)}{X(t)} = \mu dt + \sigma dW(t)$$

Calcul Stochastique
Processus Aléatoires
Modèles Econométriques
Analyse de Données
Statistique et Simulation

Kamal BOUKHETALA et Djamil AÏSSANI
Editeurs

Préambule

Le monde d'aujourd'hui évolue dans un environnement de plus en plus complexe et hasardeux, entaché de beaucoup de risques et d'incertitude. La recherche scientifique suit cette évolution avec intérêt et motivation. La modélisation stochastique et statistique, en tant que science moderne, cherche la conception de modèles adéquats pour décrire et simuler des situations réelles ou phénomènes soumis à des sollicitations aléatoires. Les modèles conçus constituent également des standards ou référentiels pour la mesure et le calibrage de l'information statistique observée dans les domaines de l'application.

Cet ouvrage regroupe 37 articles parmi ceux qui avaient été acceptés, présentés et discutés lors de la conférence internationale *Modélisation Stochastique et Statistique (MSS'04)*, USTHB, Alger, 2004. Ces articles ont fait l'objet d'une expertise internationale effectuée après la tenue de *MSS'04*. Nous tenons à remercier vivement l'ensemble des référés pour la qualité de cette expertise et l'effort fourni.

L'ouvrage est organisé en cinq chapitres, selon les thèmes retenus par la conférence, à savoir :

- Calcul Stochastique
- Processus Aléatoire
- Modèles Econométriques
- Analyse de Données et Applications
- Statistiques Computationnelles et Simulation

Ces thèmes indiquent différentes orientations de recherche et constituent un document de référence pour les jeunes chercheurs et les praticiens dans divers domaines de la modélisation stochastique et statistique

Les éditeurs

Modélisation Stochastique et Statistique

Editeurs : Kamal Boukhetala et Djamil Aïssani

Adresse : Département de Probabilités – Statistiques, U.S.T.H.B. Alger (Algérie)

Tél/Fax : (+213) 772 228 630

E-Mail : mss4_usthb@yahoo.fr

Référés

Djamil Aïssani (Béjaia), Kwessi Andam (Accra), Karim Benhenni (Grenoble), Kamal Boukhetala (Alger), Ahmed Bousbaine (Lausanne), Benamar Chouaf (Sidi Bel Abbès), Francs Corset (Grenoble), Azzouz Dermoune (Lille), Jacques Julien (Grenoble), Nikolai Kartashov (Kiev), Zoher Mohdeb (Constantine), Tahar Mourid (Tlemcen), Abdelhakim Necir (Biskra), Elias Ould Said (Lille), France, Cristian Préda (Lille), G.V. Sarma (Delhi), Ali Tahar (Annaba), Jacques Teghem (Mons)

Dépôt légal : 2088-2007

ISBN : 978-9947-0-1750-0

© Publication de la Faculté de Mathématiques, U.S.T.H.B. Alger, 2018

Tous droits de traduction, de reproduction

Et d'adaptation réservés pour tous les pays

SOMMAIRE

| | |
|---|--------|
| <i>I – Calcul Stochastique</i> | 01 |
| Analyse Spectrale des Processus non Stationnaires et leurs Applications en Agronomie. R. SABRE..... | 02 |
| On Asymptotic Behavior of the Extreme Quantile. A. NECIR..... | 06 |
| Une Autre Approche de l’Estimation. A. GUELLIL..... | 15 |
| Optimisation par la méthode adaptée d’un problème linéaire-quadratique convexe. M. BIBI et N. IKHENACHE..... | 27 |
| <i>II – Processus Aléatoires</i> | 33 |
| Sur le Théorème Ergodique pour l’Action d’un Semi-Groupe Libre. F. CHARLOT..... | 34 |
| Conditional Law of Excursions Known to Start at x end at y Without Duality Hypothesis. H. BOUTABIA..... | 46 |
| Inégalités pour les Systèmes de Files d’Attente avec Rappels. M. BOUALEM et A. AÏSSANI..... | 52 |
| Some Inequalities of the Uniform Ergodicity and Strong Stability of Waiting Process. Z. MOUHOUBI and D. AÏSSANI..... | 58 |
| Comportement asymptotique dans l’anneau de Cambridge. B. CHOUAF..... | 64 |
| Strong Approximation on Quantile Function for Strong Mixing and Censored Processes. O. SADKI and E. OULD-SAÏD..... | 70 |
| Sur les Systèmes d’Attente à Serveurs Non Fiables. K. ABBAS et D. AÏSSANI..... | 77 |
| Approximation par la Théorie de l’Information d’une File M/G/1 avec Rappels et Serveur non Fiable. R. SMAÏL et A. AÏSSANI..... | 84 |
| Nombre Moyen de Clients dans Deux Files Parallèles avec Flux d’Arrivée Partage. T. LARDJANE..... | 88 |
| Caractérisation Variationnelle des Processus de Schrodinger avec Potentiel non Borné. A. BENCHETTAH..... | 93 |

| | |
|--|------------|
| La Récurrence des Files d'Attente avec Impatience pour des Equations non Monotones. A. GHERIBALLAH et B. CHOUAF..... | 99 |
| Principes d'Invariance Hölderiens : Cas de Suite non Stationnaire. D. HAMADOUCHE et Y. TALEB..... | 105 |
| III – Modèles économétriques..... | 111 |
| On The Linear Hypothesis Testing in Nonparametric Regression Model. Z. MOHDEB and A. MOKKADEM..... | 112 |
| Analyse Bayésienne d'un Modèle Autorégressif à Erreurs ARCH. H. MERAD et H. GUERBYENNE..... | 118 |
| Sur l'Estimation du Paramètre d'un AR(1) avec Innovations Uniformes. Cas des Echantillons Finis. K. NOUALI et H. FELLAG..... | 124 |
| Etude d'un Modèle Bilinéaire à Coefficients Périodiques. A. BIBI..... | 132 |
| IV – Analyse de données et Applications..... | 138 |
| Méthodologies Statistiques dans l'Industrie. A. BOUSBAINÉ..... | 139 |
| Shannon's Entropy for Network Optimization. D. POMORSKI..... | 145 |
| Détection de Spectres Anormaux en Analyse Multivariée. A. TAHAR , C. BAROUR, S. AÏT-KAKI, R. PALM..... | 155 |
| Sélectivité et Criblage des Facteurs Chimiques sur le Potentiel des Electrodes à Membrane Solide. S. ALI-MOKHNACHE, N. DADDA, A. LARKEM, A. TAHAR et D. MSSADI..... | 162 |
| Discrimination de Populations d'Abeilles Mellifères en Analyse Multivariée. C. BAROUR, A. TAHAR and S. E. RADLOFF..... | 167 |
| Stochastic Regionalization of Hydrogeological Data With Fuzzy-Kriging. B. AZOUZI, J. SCHRÖTER and L. BENAICHATTA..... | 174 |
| Analyse en Composantes Principales sous Contrainte Fonctionnelle. Application à l'Analyse Factorielle d'une Fonction Aléatoire. T. BENCHIKH et A. YOUSFATE..... | 180 |
| Analyse Multi-échelle pour la Détection des Singularités pour des Données Bruitées. Application aux champs Magnétique Terrestre. F. MEZOUEDE et M. EL BAHI..... | 186 |

| | |
|---|-----|
| <i>V – Statistique Computationnelle et Simulation</i> | 193 |
| Choix de la Largeur de la Fenêtre dans l’Estimation de la Densité Spectrale d’un Processus Stationnaire et de l’Opérateur de Régression des Données Fonctionnelles. M. RACHDI..... | 194 |
| Sur le Paramètre de Lissage dans l’Estimation de la Densité de Probabilité par la Méthode du Noyau. S. ADJABI..... | 198 |
| Robust IF Estimation of Multicomponent FM Signals Affected by Heavy-Tailed Noise Using TFD. M. SAHMOUDI and K. ABED-MERAIM..... | 209 |
| Etude des Mesures Prédicatives dans le Modèle de Cox-Dirichlet à Censures Fixes. F. MESSACI et J-P. RAOULT..... | 210 |
| Nouvelle Classe d’Estimateurs à Noyau pour l’Indice des Valeurs Extrêmes : Consistance. D. MERAGHNI et A. NECIR..... | 216 |
| Modèles Epidémiologiques du Sida. F.L. RAHMANI..... | 222 |
| Sur un Algorithme Génétique Spatial pour la Conception d’Echantillonnage Spatial Optimal avec Quantification. S. AIT KACI AZZOU et K. BOUKHETALA..... | 227 |
| On the Use of Artificial Intelligent Methods for Prediction Purposes in Nuclear Research Reactors. H. MAZROU and M. HAMADOUCHE..... | 233 |
| Analyse en Composantes Principales : Approche Neuronale. Application : Analyse d’Images Optiques de Télédétection. S. CHITROUB..... | 239 |
| A Strong Uniform Convergence Rate of Kernel Conditional Quantile Estimator Under Random Censorship. E. OULD-SAÏD..... | 249 |
| <i>Index des Auteurs</i> | 256 |

I - Calcul Stochastique

ANALYSE SPECTRALE DES PROCESSUS NON STATIONNAIRES ET LEURS APPLICATIONS EN AGRONOMIE

Rachid SABRE

Laboratoire MAIS, ENESAD, 26, Docteur Petitjean 21000 Dijon France.

IMB, CNRS, Université de Bourgogne.

r.sabre@enesad.fr

INTRODUCTION

La statistique joue un rôle central dans l'explication et dans la modélisation de nombreux phénomènes agronomiques. Elle est l'outil incontournable pour certains chercheurs permettant de décrire, de traiter et d'analyser des données. La statistique des processus et particulièrement les estimations spectrales trouvent un éventail de domaines d'application assez large incluant entre autre l'agronomie (modélisation de la dépendance entre la composition d'échantillons de sol et leur localisation, en analyse d'image : problèmes de filtrage et de la segmentation de l'image d'un champ agricole, en agroalimentaire : l'étude de la composition aromatique d'un produit alimentaire par la spectrométrie de masse ou chromatographie.

Dans cette conférence nous exposons notre modeste contribution au développement de l'aspect théorique de l'Analyse Spectrale et son application dans les domaines Agronomiques.

En effet, nos travaux concernent principalement les processus non stationnaires que l'on peut séparer en deux parties

1) **Les processus *alpha* stables** : ce travail consiste à estimer la densité spectrale des processus symétriques stables à temps discret. Pour y parvenir, nous avons introduit le noyau de Hosoya. Cependant, dans la pratique, la mesure spectrale peut être composée d'une partie continue et d'un nombre fini de sauts, d'où la nécessité d'estimer la densité d'une mesure spectrale mixte. Ce travail a été réalisé en utilisant la méthode du double noyau. Il trouve ses applications en traitement de signal et en économie dans le cadre d'une nouvelle approche du problème des variations des prix spéculatifs fondée sur les processus stables. On s'est intéressé au cas des processus multidimensionnels, qui sont fréquemment rencontrés dans les applications : traitement de signal, traitement d'image et météorologie. Pour l'estimation de la densité spectrale par échantillonnage périodique, quand le processus est à temps continu, on rencontre le phénomène d'aliasing. Pour le contourner, nous supposons que la densité spectrale est à support compact.

Lorsque l'on estime la densité spectrale d'un signal stationnaire ou non stationnaire on utilise un lissage dépendant d'un paramètre, fonction de la taille de l'échantillon. L'objectif de ce travail est de donner un critère pour le choix du paramètre qui optimise la vitesse de convergence de l'estimateur. Nous avons pu trouver ce critère grâce à la méthode de la validation croisée.

Nous avons étendu ces résultats aux champs aléatoires stationnaires et aux processus et champs symétriques stables.

Dans la pratique, nous commettons souvent des erreurs en observant un processus ou un champ aléatoire stable. L'origine des erreurs peut être la lecture des observations ou l'échelle utilisée. Quand cette erreur est constante, on a besoin de l'estimer afin de la soustraire des

observations. Pour cela, en utilisant les noyaux polynomiaux de type Jackson, nous avons donné un estimateur de cette constante. Nous avons étudié sa vitesse de convergence et sa variation lorsque la densité spectrale est nulle à l'origine (le processus de longue mémoire a été étudié comme exemple).

2) **les processus oscillatoires** : Priestley a défini la densité spectrale évolutive pour les processus non stationnaires de variance finie (du second ordre). Il a également étudié l'estimation de cette densité spectrale. Nous avons commencé par définir de la même manière la densité spectrale évolutive pour un processus non stationnaire symétrique stable. Ce processus a une variance infinie. Nous avons donné un estimateur asymptotiquement sans biais et consistant de la densité spectrale évolutive.

Ensuite, nous avons supposé, comme dans la plus part des cas pratiques, que nous avons des observations manquantes dues soit à l'impossibilité d'observer le processus non stationnaire à certains instants soit à des erreurs commises sur certaines observations que nous serons forcés d'abandonner. Nous avons proposé des estimateurs asymptotiquement sans biais et consistants de la densité spectrale évolutive.

Application à l'agronomie :

Cette présentation ne prétend pas à l'exhaustivité et certains aspects seront traités très rapidement et d'autres passés sous silence. Notre seule excuse est qu'il fallait bien faire des choix pour garder cette présentation aussi claire que possible.

Pour illustrer cette application, nous citons quelques principaux travaux réalisés :

1) Nous avons appliqué certains de nos résultats concernant les estimations dans le domaine de l'analyse spectrale. Il s'agit de l'étude de la fissuration autour de la galerie qui est la modification du sol la plus importante survenant au cours de l'opération de taupage (technique de drainage simplifiée). Cette fissuration influence directement la stabilité et l'efficacité de la galerie. A l'heure actuelle, on ne dispose pas de moyen d'étude évitant l'ouverture de profils cultureux.

Pour éviter cet inconvénient, nous avons utilisé une nouvelle méthode pour étudier les facteurs qui influencent le processus de modification du sol par la charrue taupe. Nous avons placé deux capteurs de forces à jauges de contrainte sur la charrue taupe qui ont permis de mesurer l'interaction sol-outil. Nous avons exploité l'analyse spectrale. Cette méthode a permis d'identifier le processus (signal recueilli par les capteurs) et de filtrer ce signal afin d'évaluer la fissuration sur le profil longitudinal de la galerie. Grâce à la densité spectrale de puissance, nous avons trouvé que la force de traction est aléatoire et gaussienne. Ces résultats sont conformes aux observations notées sur profils cultureux et sur moulages en polyuréthane.

2) Dans le cadre de la modélisation des bio-phénomènes, un projet a été réalisé à l'ENESAD dont le but était de mettre en évidence les molécules responsables de certains arômes. Pour cela nous étudions les pics du signal envoyé par le chromatographe en rapport avec les résultats d'un jury de l'analyse sensorielle.

3) Une modélisation de la résistance du sol en fonction de l'humidité et de la profondeur a été définie. Cette modélisation permet d'étudier l'évolution de l'état structural de sols agricoles par des mesures couplées d'humidité et de « pénétrométrie ».

REFERENCES

- Azzaoui N., Clavier L., Sabre R., (2003)** Path delay model based on stable distribution for the 60GHz indoor channel, à paraître dans *IEEE, Com. Theor.*
- Cambanis S. (1982)** Complex symmetric stable variables and processes, in P.K.SEN, ed, Contributions to Statistics: Essays in Honour of Norman L. Johnson North-Holland. New York, (pp.63-79)
- Chen J.N, Coquille J.C, Douzals J.P, Sabre R. (1997)** Frequency composition of traction and tillage forces on a mole plough, *Soil & Tillage Research*, 44 (1997) 67-79.
- Demesh, N. (1988)** Application of the polynomial kernels to the estimation of the spectra of discrete stable stationary processes, (in russian) Akad.Nauk.Ukrain. S.S.R. Inst.Mat. Preprint 64, pp.12-36
- Dzyadik (1977)**, Introduction à la théorie de l'approximation uniforme par fonctions polynomiales (in russian) Nauka
- Hosoya, Y. (1978)**, Discrete-time stable processes and their certain properties, Ann.Probability 6, pp.94-105.
- Masry E., Cambani S., (1984)** Spectral density estimation for stationary stable processes, Stochastic processes and their applications 18 pp. 1-31.
- Mélard G., Herteleer-de-Shutter (1989)** Contributions to evolutionary spectral theory. Journal of Time Series Analysis. 10, 1, 41-63.
- Neumann M. H. and Von Sachs R. (1997)** Wavelet thresholding in anisotropic function classes and application to adaptive estimation of evolutionary spectra. Ann. Stat. 25, 1, 38-76.
- Perlov M. (1989)** Direct estimation of the spectrum of stationary stochastic processes, Translated from Problemy Peredachi Informatsii 2) vol 25 pp. 3-12.
- Priestley M. (1988)** Non-linear and non-stationary Time Series Analysis. IEEE. Academic Press, New York.
- Rachdi M, Sabre R. (1998)** Le choix de la largeur de fenêtre spectrale pour un champ aléatoire à temps discret. *Journal du traitement du signal*, vol.15, n°6, 569-575.
- Rachdi M, Sabre R. (2000)** Consistent estimates of the mode of the probability density function in nonparametric deconvolution problems. *Statistics & Probability Letters*, 47(2000), 105-114.
- Sabre R. (1994)** Estimation de la densité de la mesure spectrale mixte pour un processus symétrique stable strictement stationnaire. *C. R. Acad. Sc, Paris, t 319, (Série I), 1307-1310.*

Sabre R. (1995) Spectral density estimation for stationary stable random fields. *Applicaciones Mathematicae Journal*, 32, 2, 107-133.

Sabre R. (1999) Estimation of the constant measurement error of stable random field. *The Egy. Statistical journal*, vol.43, N° 2, 117-128.

Sabre R. (2000) Discrete estimation of spectral density for symmetric stable process. *Statistica*, N° 3 (2000), 497-520.

Sabre R. (2002) Aliasing free for stable field, *The Egy. Statistical journal*, vol 6, N1, 53-75.

Sabre R., (2002) Analyse spectrale des processus alpha stables, dans le livre «Ecrits sur les processus aléatoires», *Ed. Hermes, Paris, (2002)*.

Sabre R. (2003) Nonparammetric estimation of the continuous part of mixed measure. À paraitre au journal *Statistica*.

Safi S, Sabre R, Schmitt B. (2000) Estimation de la densité spectrale évolutive et les problèmes liés aux observations manquantes. *C.R.. SFdS, XXXIIe Journées de Statistique*, 685-687.

ON ASYMPTOTIC BEHAVIOR OF THE EXTREME QUANTILES

ABDELHAKIM NECIR

ABSTRACT. We present some results concerning the extreme value theory, namely the domain of attractions and the limiting distributions of maxima of iid random variables. We focus on the extreme quantiles estimate and there applications in reinsurance treaty.

1. DOMAINS OF ATTRACTION

Let X_1, X_2, \dots be independent and identically random variables with common distribution function $F(x) = P(X_1 \leq x)$, $x \in \mathbb{R}$.

Theorem 1. *F is to be in the domain of attraction of non-degenerate distribution G , denoted $F \in D(G)$ if there exist sequences of real numbers (a_n) and (b_n) , with $a_n > 0$ and $b_n \in \mathbb{R}$, such that:*

$$(1) \quad \lim_{n \rightarrow \infty} F^n(a_n x + b_n) = \lim_{n \rightarrow \infty} P \left\{ \frac{\max(X_1, \dots, X_n) - b_n}{a_n} \leq x \right\} = G(x),$$

for all x (see for instance Leadbetter, Lindgren and Rootzén(1983)).

In extreme value theory, similar situation occurs: the limiting distribution G in (1) can only come from a limited class of distribution functions, regardless of the underling distribution function F . It is only the behavior of the tail of the distribution that determines to which domain of attraction F belongs.

That result in extreme value theory, first discovered by Fisher and Tippett (1928) and later proved in complete generality by Gnedenko (1943), is given by the following theorem.

Theorem 2. *Suppose distribution function F satisfies assumption (1) with limiting distribution G . Then G is the same, up to location and*

1991 *Mathematics Subject Classification.* 62G32, 62F40.

Key words and phrases. Extreme value, Hill estimator, Extreme quantiles, Premium calculation, Reinsurance treaty.

scale, as of the following distributions:

$$\text{Type I :} \quad \Lambda(x) := \exp(-e^{-x}) \quad x \in \mathbb{R},$$

$$\text{Type II :} \quad \Phi_\gamma(x) := \begin{cases} 0 & x \leq 0 \\ \exp(-x^{-1/\gamma}) & x > 0 \end{cases} \quad \gamma > 0,$$

$$\text{Type III :} \quad \Phi_\alpha(x) := \begin{cases} \exp(-(-x)^{1/\gamma}) & x < 0 \\ 1 & x \geq 0 \end{cases} \quad \gamma > 0.$$

We will refer to these distribution as the extreme value distribution.

2. GENERAL FORM

Suppose that F in the domain of attraction of an extreme-value distribution, i.e., there exist constants $a_n > 0$ and b_n such that

(1.1)

$$\lim_{n \rightarrow \infty} P \{ a_n^{-1} (X_{n,n} - b_n) \leq x \} = G_\gamma(x) := \exp \left\{ - (1 + \gamma x)^{-1/\gamma} \right\},$$

for all x with $1 + \gamma x > 0$ (notation: $F \in \mathcal{D}(G_\gamma)$). We will use the convention that $G_0(x) := \exp(-e^{-x})$, for $x \in \mathbb{R}$ (see, e.g., Embrechts, Klüppelberg and Mikosch (1997)).

3. SOME CHARACTERIZATIONS

Let

$$x_F^0 := \sup \{ x, F(x) < 1 \} \leq \infty,$$

denote the *upper endpoint* of a distribution F .

Theorem 3. *Necessary and sufficient conditions for a distribution function F to belong to one of the three possible domains of attraction are:*

1- $F \in D(\Lambda)$ if and only if there exists some strictly positive function $g(t)$ such that

$$(2) \quad \lim_{t \uparrow x_F^0} \frac{1 - F(t + xg(t))}{1 - F(t)} = e^{-x},$$

for all real x . Moreover (2) holds for some strictly positive function g , then it also holds for the following choice of g :

$$g(t) = \frac{\int_t^{x_F^0} (1 - F(u)) du}{1 - F(t)}, \quad \text{for } t < x_F^0.$$

2- $F \in D(\Phi_\gamma)$ for some $\gamma > 0$ if and only if $x_F^0 = \infty$ and

$$(3) \quad \lim_{t \rightarrow \infty} \frac{1 - F(tx)}{1 - F(t)} = x^{-1/\gamma}, \quad \text{for each } x > 0.$$

$\exists F \in D(\Psi_\gamma)$ for some $\gamma > 0$ if and only if $x_F^0 < \infty$ and

$$\lim_{h \downarrow 0} \frac{1 - F(x_F^0 - xh)}{1 - F(x_F^0 - h)} = x^{1/\gamma}, \text{ for each } x > 0.$$

Proof. See e.g. Gnedenko(1943). □

4. PARETO TYPE DISTRIBUTIONS (HEAVY TAILED)

A distribution is said to be a Pareto type if $F \in D(\Phi_\gamma)$. Prominent example of heavy-tailed distributions are the lognormal, Pareto and Weibull with shape parameter smaller than 1 ($\gamma < 1$). For example, actuaries believe that lognormal distributions are plausible models for motor insurance., while they feel that Pareto distribution are apt to model fire claim data (see, e.g., Embrechts, Klüppelberg and Mikosch (1997)). The real parameter $\gamma > 0$ is called *the tail index* of F . In the literature there are several estimators for $\gamma > 0$ among them is the well known Hill estimator (see Hill (1975)).

5. HILL ESTIMATOR

For each integer $n \geq 1$, let $X_{1,n} \leq \dots \leq X_{n,n}$ denote the order statistics pertaining to the sample X_1, \dots, X_n . The Hill estimator(Hill (1975)) of the tail index $\gamma > 0$ is defined by

$$\hat{\gamma}^{(H)}(k_n) := k_n^{-1} \sum_{i=1}^{k_n} \log(X_{n-i+1,n}/X_{n-k_n,n}),$$

where the k_n are positive integers satisfying the conditions

$$1 \leq k_n < n, \quad k_n \rightarrow \infty \quad \text{and} \quad k_n/n \rightarrow 0 \quad \text{as} \quad n \rightarrow \infty.$$

Under these conditions, it is well-known (Mason (1982)) that $\hat{\gamma}^{(H)}(k_n)$ is consistent for all $\gamma > 0$.

Several authors have tried to guide the practitioner in choosing k , leading to an adaptive choice \hat{k} such that an estimate of the mean squared error of the Hill estimator is minimized at \hat{k} . This was done by bootstrap methods (see, for instance de Haan et al. (2001)) or by regression diagnostics an a Pareto quantile plot in Beirlant et al. (1996). A somewhat different solution was proposed in Drees et al. (1998).

6. KERNEL-ESTIMATORS

Csörgő, Deheuvels and Mason (1985) [CDM]. The CDM kernel estimators being defined by

$$\hat{\gamma}_{n,h}^{(CDM)} := \sum_{i=1}^{n-1} (i/n) K_h(i/n) \log(X_{n-i+1,n}/X_{n-i,n}),$$

where $K(\cdot)$ be a kernel function fulfilling (CK1) – (CK4) below.

(CK1) $K(x) = 0$, for $x \notin [0, 1)$ and $K(x) \geq 0$, for $x \in [0, 1)$.

(CK2) K is twice continuously differentiable on $(0, 1]$.

(CK3) $K(1) = K'(1) = 0$.

(CK4) $\int_0^1 K(x) dx = 1$.

7. GENERAL KERNEL ESTIMATOR

Groeneboom, Lopuhaä and de Wolf (2003)[GLW]. Selecting $\alpha > 0$ and $h > 0$, let

$$\hat{q}_{n,h}^{(1)} := \sum_{i=1}^{n-1} (i/n)^\alpha K_h(i/n) \log(X_{n-i+1,n}/X_{n-i,n}),$$

where $K_h(u) := h^{-1}K(u/h)$, $u \in (0, 1]$.

For a such function K , since (CK2) holds, we set

$$\hat{q}_{n,h}^{(2)} := \sum_{i=1}^{n-1} \Psi_h(i/n) \log(X_{n-i+1,n}/X_{n-i,n}),$$

where $\Psi_h(u) := (d/du)u^{\alpha+1}K_h(u)$, $u \in (0, 1]$.

We define the estimators $\hat{\gamma}_{n,h}^{(GLW)}$ via

$$\hat{\gamma}_{n,h}^{(GLW)} := \hat{\gamma}_{n,h}^{(CDM)} - 1 + \hat{q}_{n,h}^{(2)}/\hat{q}_{n,h}^{(1)}.$$

We will make use of the following notation. The inverse or quantile function of F is defined by

$$\mathbb{Q}(s) := \inf \{x : F(x) \geq s\}, \quad 0 < s \leq 1,$$

with $\mathbb{Q}(0) := \mathbb{Q}(0+)$.

Assuming differentiability of the function $\mathbb{Q}(\cdot)$ it is convenient to set

$$\phi(s) := -s(d/ds) \log \mathbb{Q}(1-s) \quad \text{for } 0 < s < 1.$$

We require the following additional assumptions upon the distribution F :

(CP1) When $\gamma \geq 0$, $\phi(s) \rightarrow \gamma$ as $s \downarrow 0$.

(CP2) When $\gamma < 0$, for some constant $d > 0$,

$$s^\gamma \phi(s) \rightarrow -d\gamma \text{ as } s \downarrow 0.$$

(CP3) When $\gamma = 0$, for all $s > 0$, $\phi(hs)/\phi(h) \rightarrow 1$ as $h \downarrow 0$.

Theorem A. Assume that $F \in \mathcal{D}(G_\gamma)$ for some $\gamma \in \mathbb{R}$, and let (CP1) – (CP3) hold. Fix an arbitrary $\alpha > 1/2$, and let K be a kernel satisfying conditions (CK1) – (CK4). Whenever $h = h_n$ is such that $h \downarrow 0$ and $(nh)^{-\alpha} \log n = O\left((nh)^{-1/2}\right)$ as $n \rightarrow \infty$, we have

$$(nh)^{1/2} \left(\hat{\gamma}_{n,h}^{(GLW)} - \gamma_h^{(GLW)} \right) \xrightarrow{\mathcal{D}} \mathcal{N} \left(0, \sigma^2(\bar{K}) \right) \text{ as } n \rightarrow \infty,$$

where

$$\sigma^2(\bar{K}) := \int_0^1 (\bar{K}(s))^2 ds.$$

Proof. See Theorem 4.1 of Groeneboom *et al.* (2003).

8. FUNCTIONAL LAW OF THE ITERATED LOGARITHM FOR GLW ESTIMATORS

Theorem (Necir (2003)) Under the assumptions of Theorem 2.2.1, we have wpl

$$\limsup_{n \rightarrow \infty} \pm (nh)^{1/2} \ell_n^{-1/2} \left[\hat{\gamma}_{n,h}^{(GLW)} - \gamma_h^{(GLW)} \right] = 2^{1/2} \sigma(\bar{K}),$$

Moreover, we have wpl

$$\hat{\gamma}_{n,h}^{(GLW)} \rightarrow \gamma \text{ as } n \rightarrow \infty.$$

9. EXTREME QUANTILES

Let $u := u_n \rightarrow \infty$ such that $u > X_{n-\hat{k}_n}$. Let $\tau_n > 0$ such that

$$1 - F(u) = \tau_n, \tau_n \rightarrow 0.$$

We may estimate τ_n by the extreme quantiles methods:

$$\hat{\tau}_n = \left(\hat{k}_n/n \right) \left(X_{n-\hat{k}_n} \right)^{1/\hat{\gamma}^{(H)}(k_n)} u^{-1/\hat{\gamma}^{(H)}(k_n)},$$

when $\hat{\gamma}^{(H)}(k_n)$ is the Hill estimator.

10. REINSURANCE TREATY

10.1. Wang’s premium. From the point of view of the policyholder, an insurance system is a ”mechanism for reducing the adverse financial impact of random events that prevent the fulfilment of reasonable expectations”. This means that an insurance contract can be seen as a risk-exchange between two parties, the insurer and the policyholder. The insurer promises to pay for the financial consequences of the claims produced by the insured risk. In return for this coverage, the policyholder pays a fixed amount, called the premium. Observe that the payments made by the insurer are random, while the payments made by the policyholder are non-random. The pure premium of the insured risk is defined as the expected value of the claim amounts to be paid by the insurer. In practice the insurer will add a risk loading to this

pure premium. The sum of the pure premium and the risk loading is called the net premium. Adding acquisition and administration costs to this net premium, one gets the gross premium.

Let X, X_1, X_2, \dots be a sequence of independent and identically distributed risks with common distribution function $F(x) = P(X \leq x)$, $x \in \mathbb{R}$. Wang (1996) proposes to compute the *net premium* of a risk X as a "distorted" expectation of X :

$$\Pi_p(X) := \int_0^\infty (1 - F(s))^{1/p} ds,$$

for some *loading parameter* $p \geq 1$. This consistent with the practice of adding a safety margin against the ruin.

11. WANG'S PREMIUM OF LOSS

The corresponding *risk-adjusted premium of loss* of a layer from u (*retention level*) to horizon $h = x_F^0$ being as follows:

$$\Pi_{u,p}(X) := \int_u^{x_F^0} (1 - F(s))^{1/p} ds.$$

12. THE MEAN EXCESS FUNCTION

The mean excess function is given by

$$\begin{aligned} m(u) &= E(X - u | X > u) \\ &= \frac{\int_u^\infty (1 - F(x)) dx}{1 - F(u)} \end{aligned}$$

as the expected excess claim size. This function plays a central role in the rating of an excess-of-loss reinsurance in excess of a retention or priority level u , as the corresponding risk premium for the layer from u to infinity is given by (a multiple of)

$$\begin{aligned} E((X - u)_+) &= (1 - F(u)) m(u) \\ &= \int_u^\infty (1 - F(x)) dx. \end{aligned}$$

Here $(x - a)_+ = \max(x - a, 0)$.

13. WANG'S SAFETY LOADING OF LOSS

The safety loading with respect to Wang's premium (notation: $SL(p)$) is

$$\begin{aligned} SL(p) &= \Pi_{u,p}(X) - E((X - u)_+), \\ &= \int_u^\infty (1 - F(s))^{1/p} ds - \int_u^\infty (1 - F(x)) dx > 0, \text{ for } p > 1. \end{aligned}$$

In the sequel we give an adequate $p > 1$ in order to have $SL(p) > 0$.

14. ESTIMATING WANG'S PREMIUM OF LOSS

For heavy tailed distributions we have $x_F^0 = \infty$. Recall that the *risk-adjusted premium of loss of the layer from u (retention level) to ∞* is defined as:

$$\Pi_{u,p}(X) := \int_u^\infty (1 - F(s))^{1/p} ds, \quad p \geq 1.$$

Let F_n be the empirical distribution. Necir and Boukhetala (2004) propose an estimator for $\Pi_{u,p}(X)$ as follows:

$$\begin{aligned} \widehat{\Pi}_{u,p}(X) &= \int_u^\infty (1 - F_n(s))^{1/p} ds \\ &= \int_{Q(1-\widehat{\tau}_n)}^\infty (1 - F_n(s))^{1/p} ds. \\ &= \sum_{i=1}^{[\widehat{\tau}_n]} (i/n)^{1/p} (X_{n-i+1,n} - X_{n-i,n}). \end{aligned}$$

15. CONFIDENCE INTERVAL FOR $\widehat{\Pi}_n$

Let

$$Q(s) = \inf \{t \in R, F(t) \geq s\}, \quad 0 \leq s < 1,$$

denote the generalized inverse or the quantile function pertaining to the distribution function F . We also have a definition, that equivalent to (3), in term of the quantile function, that is

$$(5) \quad \lim_{t \rightarrow 0} \frac{Q(1-tx)}{Q(1-t)} = x^{-\gamma} \quad \text{for } x > 0.$$

(see, e.g., Embrechts, Klüppelberg and Mikosch (1997)). The real parameter γ is called the tail index of F . In the literature there are several estimators for $\gamma > 0$ among them is the well known Hill estimator (see Hill (1975)).

Further, we assume that the distribution tail $1 - F$ is second order regularly varying with first parameter $-1/\gamma$ and second parameter $\rho \leq 0$, that is, there exists a function $A(t) \rightarrow 0$ as $t \rightarrow 0$ which ultimately has constant sign such that the following refinement of (5) holds:

$$(7) \quad \lim_{t \rightarrow 0} (A(t))^{-1} \left\{ \frac{Q(1-tx)}{Q(1-t)} - x^{-\gamma} \right\} = x^{-\gamma} \frac{x^{\rho\gamma} - 1}{\rho\gamma}, \quad \text{for any } x > 0.$$

If $\rho = 0$, interpret $\frac{x^{\rho\gamma} - 1}{\rho\gamma}$ as $\log x$ (see, e.g., de Haan and Stadtmüller(1996)).

Theorem 4. Assume that (7) holds and $s^{1/p}Q(1-s) \rightarrow 0$ as $s \rightarrow 0$, for some $\gamma - 1/2 < 1/p \leq 1$. Let $k = k_n$ is such that $k \rightarrow \infty$, $k/n \rightarrow 0$ and $(k/n)^{1/2} A(k/n) \rightarrow 0$ as $n \rightarrow \infty$. Then,

$$\frac{(k/n)^{-1/p} k^{1/2}}{Q(1-k/n)} \left\{ \widehat{\Pi}_{u,p}(X) - \Pi_{u,p}(X) \right\} \xrightarrow{\mathcal{D}} \mathcal{N}(0, \sigma^2(p, \gamma)), \text{ as } n \rightarrow \infty,$$

where

$$\sigma^2(p, \gamma) := \gamma^2 \begin{cases} p^{-2} \frac{1}{\lambda(\lambda+1)} - p^{-1} \frac{2}{\lambda+1} + 1, & \text{for } 1/p \neq \gamma \\ 2p^{-2} - 2p^{-1} + 1, & \text{for } 1/p = \gamma. \end{cases}$$

with $\lambda := 1/p - \gamma$.

Proof. See Necir and Boukhetala(2004). □

16. CONCLUSION

Our result provides an estimator of premium for the excess of loss reinsurance covers with safety loading. This property does not exist for the usual premium of loss.

REFERENCES

- [1] de Haan, L., Peng, L. and de Vries, C. G.(2001). Using bootstrap method to choose the sample fraction in tail index estimation, *Journal of Mult. Anal.*, **76**, 226-248.
- [2] Drees, H. and Kaufmaan, E.(1998). Selecting the optimal sample fraction in univariate extreme value estimation, *Stoch. Proc. Appl.*, **75**, 149-172.
- [3] Embrechts, P., Klüppelberg, C. and Mikosch, Th.(1997). *Modelling Extremal Events for Insurance and Finance*. Springer Verlag, Berlin Heidelberg New York, .
- [4] Fisher, R. A. and Tippett, L.H.C. (1982). Limiting forms of the frequency distribution of the largest or smallest member of a sample, *Proceedings of the Cambridge Philosophical Society*, **24**, 180-190.
- [5] Gnedenko, B. V (1943). Sur la distribution limite du terme maximum d'une série aléatoire, *Annals of Mathematics*, **44**, 423-453.
- [6] Hill, B.M. (1975). A simple approach to inference about the tail of a distribution. *Ann. Statist.*, **3**, 1163-1174.
- [7] Leadbetter, M. R., Lindgren, G. and Rootzén, H. (1983). *Extremes and related properties of random sequences and processes*. Springer, New York.
- [8] Mason, D. (1982). Laws of the large numbers for sums of extreme values. *Ann. Probab.*, **10**, 754-764.
- [9] Necir, A. and Boukhetala, K. (2004). Estimating the risk-adjustes premium for the largest claims reinsurance covers. *Physica-Veralg, Compostat'2004*.
- [10] Wang, S. (1996). Premium calculation by transforming the layer Premium density. *ASTIN bulletin*, **26**, 71-92.

LABORATORY OF APPLIED MATHEMATICS, UNIVERSITY OF BISKRA, PO BOX
145 RP, 07000 BISKRA.

E-mail address: necirabdelhakim@yahoo.fr

UNE AUTRE APPROCHE DE L'ESTIMATION

Ahmed GUELLIL
Département de Probabilités et Statistiques
Faculté des Mathématiques de l'USTHB
e-mail: aguellil@usthb.dz

Résumé

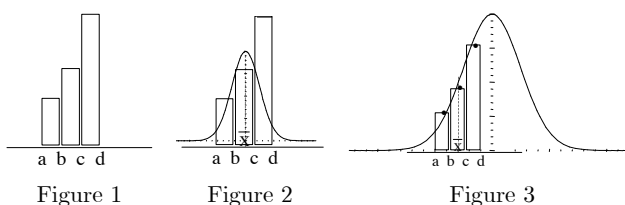
Je considère qu'étant donné un échantillon de taille quelconque, les variations de sa distribution propre ne reproduisent pas obligatoirement même approximativement, toutes les variations de la distribution source. Elles en reproduisent peut être seulement une partie dont l'étendue dépend de la taille de l'échantillon. Cela suggère alors d'aborder l'estimation autrement. Soit un échantillon d'observations aléatoires et indépendantes sur une variable aléatoire X , de densité (ou fonction de masse) f , inconnue, appelée la distribution source de l'échantillon. On se donne une famille de distributions théoriques qui est supposée la contenir. L'objectif est de choisir dans cette famille, au vu des observations, la distribution qui est susceptible d'être la distribution source. La méthode consiste à considérer les distributions conditionnelles, de la famille de distributions théoriques, données. On cherche dans cette nouvelle famille la distribution la plus proche de celle de l'échantillon. Enfin, on détermine la distribution qui lui correspond. Cette dernière est l'estimation de la distribution par la nouvelle approche.

Mots clés: Estimation, Distribution conditionnelle, Courbe image.

1 Introduction

J'ai déjà signalé l'existence de cette approche en 1993 au séminaire de probabilités et statistique de l'ex-institut de mathématiques de l'USTHB à Alger et en 1995 dans [2]. Je reviens dessus avec un peu plus de détails. Toutefois je n'ai pas l'intention de faire une étude comparative détaillée entre l'approche classique et celle-ci, ni d'étaler toutes les possibilités qu'elle offre. Le travail serait trop volumineux et trop compliqué. Mais je vais indiquer seulement à l'aide d'exemples quelques éléments d'appréciation. Le but ici reste de faire découvrir la nouvelle approche.

En pratique étant donné un échantillon, même si sa taille n'est pas à priori suffisamment grande, on assimile les variations de sa distribution propre à ceux de la distribution source. C'est ce qu'on fait, par exemple, lorsqu'on estime la moyenne et la variance de la distribution source respectivement par la moyenne et la variance de l'échantillon. Soit par exemple un échantillon d'une distribution source normale et dont l'histogramme des fréquences a la forme de la figure 1 ci dessous:



La forme de la distribution propre à l'échantillon est visiblement différente de celle de la distribution source. Malgré cela on sait qu'il peut contenir tous les renseignements nécessaires pour la déterminer de façon exacte et avec certitude. Même si c'est le cas, il sera éventuellement rejeté par le test de conformité sous sa forme habituelle. S'il n'est pas rejeté et sa moyenne est comprise entre b et c, dans l'approche classique on lui fait correspondre une distribution normale comme sur la figure 2 ci dessus. Dans ce cas particulier on aura attribué à l'échantillon considéré la propriété de décrire une distribution normale complètement, alors qu'il n'en a ni le même support ni la même forme. Du point de vu de la nouvelle approche, on considère que les variations de la distribution de l'échantillon traduisent seulement les variations de la partie de la distribution source qu'il couvre et le test de conformité, s'il y a lieu de le faire, sera effectué par rapport à cette partie uniquement. Ensuite, en utilisant un calcul déterministe, on déduit des renseignements sur l'ensemble des variations de la distribution source. Ainsi, dans la nouvelle approche on fait correspondre à l'histogramme précédent une distribution normale comme sur la figure 3 précédente. Donc les estimations données par les deux méthodes peuvent être très différentes. On montre que la précision des estimations obtenues à l'aide de la nouvelle approche augmente avec la qualité de la description des variations de la partie de la distribution source qu'il couvre et non nécessairement de toute la distribution comme c'est le cas pour l'estimation classique. De ce fait, la vitesse de convergence des estimations est à priori plus grande avec la nouvelle approche. Quand la taille de l'échantillon tend vers l'infini, il tend à couvrir toutes les parties de la distribution source en décrivant ses variations. Dans ce cas, les deux approches se rejoignent et donnent approximativement les mêmes estimations. Les calculs avec la nouvelle approche sont par contre plus compliqués qu'avec l'approche classique et changent d'une loi à une autre. Mais avec l'utilisation de l'outil informatique on peut effectuer ces calculs aisément. Les applications de la nouvelle approches sont diverses. Par exemple, si

nécessaire on peut restreindre l'étude d'un échantillon sur une partie seulement de son étendue. Si les renseignements qu'il contient sur cette partie au sujet de la distribution source sont suffisants on peut la retrouver. On peut aussi considérer une partition de son étendue et étudier séparément chaque partie. Cela peut servir pour raffiner l'étude de l'échantillon ou pour faire apparaître un mélange éventuel de distributions.

Dans la section 2, on rappelle la définition de la distribution d'un échantillon d'une v.a à valeurs dans \mathbb{R} . Dans la section 3, on introduit la notion de courbe image. Elle est supposée reproduire approximativement les variations de la distribution empirique des observations. Dans la section 4, on expose la nouvelle approche et on donne des exemples pour illustrer la méthode. Dans la section 5, on présente des résultats numériques obtenus par la simulation pour les distributions binomiale et normale.

2 Distribution d'un échantillon d'une variable à valeurs dans \mathbb{R}

Pour appliquer la nouvelle approche, il est nécessaire de mettre en évidence les variations de la distribution de l'échantillon le plus exactement possible. Dans le cas discret, on utilise la fonction:

$$f(x) = \begin{cases} \frac{n_i}{n} & \text{si } x = y_i, i = 1, \dots, k \\ 0 & \text{sinon} \end{cases}$$

où y_i est la $i^{\text{ème}}$ valeur de l'échantillon ordonnée, n_i représente la fréquence absolue de l'observation y_i et $n = \sum_{i=1}^k n_i$.

Dans le cas continue, on se ramène à une distribution discrète mais qui met en évidence les variations de la distribution source sur la partie couverte par l'échantillon. Ce n'est pas le cas de la distribution discrète habituellement considérée et qui affecte aux différentes observations le même poids. Pour déterminer la distribution désirée, on peut se servir de la distribution lissée, définie par :

$$h(x) = \begin{cases} \frac{n_i}{n \times \Delta_i} & \text{si } x \in [a_i, a_{i+1}], i = 1, \dots, k \\ 0 & \text{sinon} \end{cases}$$

où les intervalles $[a_i, a_{i+1}]$ représentent les différentes classes de regroupement et $\Delta_i = a_{i+1} - a_i$.

On remplace chaque classe par son centre ou par la moyenne des observations qu'elle contient. Soient y_1, y_2, \dots, y_k les points ainsi définis. On leur affecte les poids de probabilités respectifs notés $f_i, i = 1, 2, \dots, k$ et définis par:

$$\frac{f_i}{f_j} = \frac{h(y_i)}{h(y_j)},$$

pour tout couple $(i, j) \in \{1, 2, \dots, k\}^2$ et $\sum_{i=1}^k f_i = 1$.

Si les différentes classes ont même amplitude on obtient $f_i = n_i/n$, $i = 1, 2, \dots, k$. En définitif, que ce soit dans le cas discret ou dans le cas continue les variations de la distribution de l'échantillon seront résumées par k couples (y_i, f_i) , $i = 1, 2, \dots, k$. Ils sont supposés nous renseigner sur les variations de la partie de la distribution source couverte par l'échantillon. Les y_i , $i = 1, 2, \dots, k$, représentent le support de la distribution de l'échantillon, comme définie précédemment. On désigne par f^* la distribution discrète ainsi définie et on l'appellera la distribution des observations. On note par P^* la mesure de probabilité correspondante:

$$P^*(B) = \sum_{x \in B} f^*(x),$$

pour tout borélien B de \mathbb{R} .

3 La courbe image

Soit f une distribution donnée. Les renseignements qu'on entend tirer de la distribution d'un échantillon sur celle-ci concernent les $f(y_i)/f(y_j)$, $(i, j) \in \{1, 2, \dots, k\}^2$. Ces rapports sont tout naturellement estimés par les f_i/f_j correspondants où les f_i , $i = 1, 2, \dots, k$, sont les probabilités définies précédemment. Posons :

$$\delta_{i,j} = \left| \frac{f_i}{f_j} - \frac{f(y_i)}{f(y_j)} \right|, \quad (i, j) \in \{1, 2, \dots, k\}^2$$

Lorsque $\delta_{i,j} = 0$ quel que soit le couple $(i, j) \in \{1, 2, \dots, k\}^2$, les points (y_i, f_i) se situent tous sur une courbe semblable à celle du graphe de f . On l'appellera la courbe image de f associée à y_1, y_2, \dots, y_k . Son expression analytique est:

$$g(x) = \frac{f(x)}{f(y_1) + f(y_2) + \dots + f(y_k)}, \quad x \in \mathbb{R}.$$

En fait, en pratique, pour un échantillon quelconque les points (y_i, f_i) ne se situent pas tout à fait sur la courbe image. Mais ils se répartissent au voisinage de celle-ci comme sur la figure 4. La qualité des renseignements qu'ils apportent augmente lorsqu'ils s'approchent de la courbe image. Dans le cas paramétrique, si la distribution source dépend de r paramètres il suffit d'avoir $r + 1$ points distincts (y_i, f_i) , $i = 1, 2, \dots, r + 1$, se situant sur sa courbe image pour la déterminer de façon exacte.

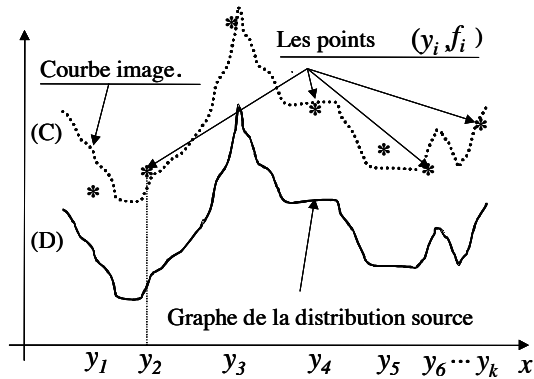


Figure 4

3.1 Exemple

Prenons une distribution source f normale de moyenne $m = 0$ et d'écart type $\sigma = 1$. Prenons les valeurs $y_1 = 1$, $y_2 = 2$ et $y_3 = 3$, ($k = 3$). Déterminons leurs poids conditionnels respectifs correspondants, donnés par:

$$f_1 = g(1), f_2 = g(2) \text{ et } f_3 = g(3).$$

Par définition de la fonction g , on a:

$$\frac{g(2)}{g(1)} = \frac{f(2)}{f(1)} = \frac{e^{-4/2}}{e^{-1/2}}, \quad \frac{g(3)}{g(1)} = \frac{f(3)}{f(1)} = \frac{e^{-9/2}}{e^{-1/2}},$$

avec $g(1) + g(2) + g(3) = 1$. Ce qui implique :

$$\begin{aligned} f_1 &= g(1) = \frac{1}{1 + \frac{e^{-4/2}}{e^{-1/2}} + \frac{e^{-9/2}}{e^{-1/2}}} \approx 0.80551 \\ f_2 &= g(2) = \frac{\frac{e^{-4/2}}{e^{-1/2}}}{1 + \frac{e^{-4/2}}{e^{-1/2}} + \frac{e^{-9/2}}{e^{-1/2}}} \approx 0.17973 \\ f_3 &= g(3) = \frac{\frac{e^{-9/2}}{e^{-1/2}}}{1 + \frac{e^{-4/2}}{e^{-1/2}} + \frac{e^{-9/2}}{e^{-1/2}}} \approx 0.014753 \end{aligned}$$

On obtient ainsi les points $(1, g(1))$, $(2, g(2))$ et $(3, g(3))$. La courbe image de celle de f , associée aux valeurs 1, 2 et 3 passe par ces points comme indiqué sur

la figure suivante:

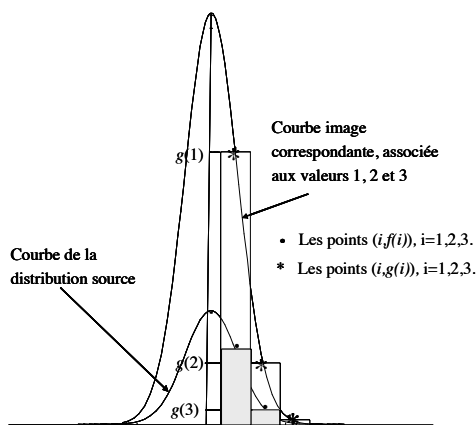


Figure 5

Le graphe a été construit en prenant $m = 0$ et $\sigma = 1$. Notons comment la courbe image a une forme semblable à celle de f , mais ce n'est pas une simple copie de celle-ci. Supposons maintenant que les points 1, 2 et 3 soient les centres des classes auxquelles est réduit un échantillon de f après regroupement. Supposons aussi que l'échantillon contient suffisamment de renseignements sur celle-ci, au sens défini précédemment, de sorte à nous donner sa forme approximative. Si on assimile la distribution de l'échantillon à celle de la distribution source, l'histogramme correspondant attendu sera proche de celui en gris sur la figure 5. Si on l'assimile à la distribution source conditionnelle sachant l'échantillon, l'histogramme attendu sera celui en blanc et qui couvre le premier. La configuration du second par rapport au graphe de la distribution source est plus proche de ce qu'on obtient généralement dans la pratique. Cela étant, ne serait-ce qu'à cause de l'aire de l'histogramme qui doit être égal à 1.

3.2 Remarque

Soit f la distribution binomiale de paramètres $p = 1/2$ et $n = 3$. Soient les tableaux de fréquences suivants représentant des échantillons de cette distribution :

Tableau 3.1.

| x | 0 | 1 | Total |
|-------|-------|--------|--------|
| n_i | n_0 | $3n_0$ | $4n_0$ |
| f_i | $1/4$ | $3/4$ | 1 |

Tableau 3.2.

| x | 2 | 3 | Total |
|-------|--------|-------|--------|
| n_i | $3n_0$ | n_0 | $4n_0$ |
| f_i | $3/4$ | $1/4$ | 1 |

Pour chacun de ces deux tableaux de fréquences les points (y_i, f_i) correspondants se situent tous sur la courbe image relative aux y_i considérés et si $n = 3$ est donné alors ils sont en nombre suffisant pour nous permettre de déterminer analytiquement la distribution source exacte. Pourtant quand on estime le paramètre p par les méthodes habituelles on obtient respectivement $p = 1/4$ et $p = 3/4$. On admet qu'avec des échantillons qui apportent des renseignements différents au sujet de la distribution source on obtient des estimations différentes. On comprend aussi qu'avec des échantillons qui apportent des renseignements insuffisants ou erronés sur la distribution source on obtient des estimations erronées. Mais il y a lieu de s'inquiéter lorsqu'avec des échantillons qui apportent les mêmes renseignements au sujet de la distribution source on obtient des estimations différentes et avec un échantillon qui apporte tous les renseignements nécessaires pour retrouver la distribution source on ne l'obtient pas. Ces problèmes se posent essentiellement lorsque la taille de l'échantillon n'est pas suffisamment grande. Pour comprendre les causes, considérons par exemple l'approche classique de l'estimation dans le cas paramétrique. Dans cette approche on procède comme suit:

- a). On détermine la distribution de l'échantillon.
- b). On se donne une famille de distributions théoriques susceptibles de contenir la distribution source de l'échantillon.
- c). On se donne un procédé mathématique qui permet de choisir la distribution théorique dans la famille en b qui est susceptible d'être la distribution source.

Il n'y a aucune raison pour remettre en question les étapes a) et c) de cette procédure. De même, dans le cas paramétrique la famille des distributions théoriques en b) est une donnée du problème et donc à priori elle ne peut pas être sujet à discussion. Dans la nouvelle approche on garde donc ce schéma mais on procède autrement en faisant apparaître d'autres aspects de la famille des distributions théoriques candidates. En utilisant la nouvelle approche, on montre dans l'exemple 1, de la section suivante, qu'on obtient la valeur exacte de p ($p = 1/2$) dans chacun des tableaux précédents.

3.3 Distribution conditionnelle.

La fonction g n'est pas en général une distribution de probabilité. Mais sa restriction à y_1, y_2, \dots, y_k en est une. C'est la distribution conditionnelle de la distribution source sachant y_1, y_2, \dots, y_k . Les quantités:

$$m(g) = \sum_{i=1}^k y_i g(y_i),$$

$$V(g) = \sum_{i=1}^k (y_i - m(g))^2 g(y_i),$$

représentent, respectivement, la moyenne et la variance conditionnelle de la distribution source sachant y_1, y_2, \dots, y_k .

4 La nouvelle approche

Etant donné un échantillon et une famille de distributions théoriques candidates. La méthode consiste en ce qui suit:

- a'). On détermine la distribution de l'échantillon comme précédemment.
- b'). On détermine les distributions théoriques conditionnelles sachant y_1, y_2, \dots, y_k des distributions théoriques candidates, comme indiqué dans la section précédente. Soit P_{y_1, y_2, \dots, y_k} cette famille.
- c'). On cherche par les méthodes mathématiques usuelles la distribution appartenant à P_{y_1, y_2, \dots, y_k} qui soit la plus proche de la distribution de l'échantillon.

4.1 Exemple 1

Considérons le premier échantillon de la distribution binomiale de paramètres $p = 1/2$ et $n = 3$ donné dans le tableau 3.1. Utilisons la méthode des moments pour déterminer une estimation de p . D'après la nouvelle approche on doit résoudre l'équation :

$$0 \times g(0) + 1 \times g(1) = \bar{X}.$$

Après calcul de \bar{X} et en remplaçant $g(0)$ et $g(1)$ par leurs valeurs en fonction de p , on obtient :

$$\frac{1 \times C_3^1 p^1 (1-p)^2}{C_3^0 p^0 (1-p)^3 + C_3^1 p^1 (1-p)^2} = \frac{3}{4}$$

D'où on tire $p = 1/2$.

Considérons maintenant l'échantillon du tableau 3.2 et utilisons encore la méthode des moments. Dans ce cas, on doit résoudre l'équation:

$$\frac{2 \times C_3^2 p^2 (1-p) + 3 \times C_3^3 p^3 (1-p)^0}{C_3^2 p^2 (1-p)^1 + C_3^3 p^3 (1-p)^0} = \frac{9}{4}$$

D'où on tire $p = 1/2$.

4.2 Exemple 2

Soit x_1, x_2, \dots, x_n un échantillon de taille n d'une distribution f normale de moyenne m inconnue et d'écart type σ donné. On désire estimer m par la méthode du maximum de vraisemblance. Selon cette méthode, du point de vue de la nouvelle approche, la valeur m est celle qui réalise:

$$\max_m \int_{\mathbb{R}} \ln f_{y_1, y_2, \dots, y_k}(y) P^*(dy), \quad (1)$$

où y_1, y_2, \dots, y_k sont les définis précédemment dans la section 2.

Pour simplifier, on suppose que les différentes classes ont la même amplitude. On obtient alors:

$$\frac{\partial}{\partial m} \int_{\mathbb{R}} \ln f_{y_1, y_2, \dots, y_k}(y) P^*(dy) = \frac{1}{n\sigma^2} \left(\sum_{i=1}^k n_i y_i - n A(m, \sigma) \right). \quad (2)$$

Les n_i représentent les fréquences absolues des différentes classes de regroupement et $A(m, \sigma)$ est la moyenne conditionnelle théorique sachant les réalisations y_1, y_2, \dots, y_k .

En égalisant (2) à 0, on obtient :

$$A(m, \sigma) = \bar{X}, \quad (3)$$

\bar{X} étant la moyenne de l'échantillon. La fonction $A(m, \sigma)$ est continue et croissante en m . Par conséquent elle admet une fonction réciproque en m . De (3) on déduit :

$$\hat{m} = A^{-1}(\bar{X}, \sigma).$$

Selon la nouvelle approche l'estimateur de la moyenne d'une loi normale d'écart type σ donné par la méthode du maximum de vraisemblance est \hat{m} et non \bar{X} . Remarquons que la fonction $A^{-1}(\bar{X}, \sigma)$ dépend de la distribution théorique candidate, de \bar{X} et de σ . Lorsque la taille de l'échantillon tend vers l'infini, \bar{X} tend avec une probabilité certaine vers m . Dans le cas de la distribution normale, la fonction $A(\cdot, \cdot)$ étant continue et monotone en m alors avec une probabilité certaine :

$$\lim_{n \rightarrow \infty} A^{-1}(\bar{X}, \sigma) = A^{-1}(m, \sigma).$$

Donc lorsque la taille de l'échantillon est infinie on obtient avec une probabilité certaine : $\bar{X} = \hat{m} = m = A^{-1}(m, \sigma)$. En développant $A(m, \sigma)$ dans (3) on obtient l'équation :

$$(y_1 - \bar{X}) e^{-\frac{(y_1 - m)^2}{2\sigma^2}} + (y_2 - \bar{X}) e^{-\frac{(y_2 - m)^2}{2\sigma^2}} + \dots + (y_k - \bar{X}) e^{-\frac{(y_k - m)^2}{2\sigma^2}} = 0$$

On doit résoudre cette équation en m pour en déterminer une estimation.

4.2.1 Application numérique

Prenons, $y_1 = 1$, $y_2 = 2$ et $y_3 = 3$, avec les fréquences respectives, $n_1 = 3$, $n_2 = 2$ et $n_3 = 1$. On obtient $\bar{X} = 10/6 \approx 1.6667$. Avec la nouvelle approche on remplace σ par sa valeur dans l'équation précédente puis on la résout. Pour $\sigma = 1$, on doit résoudre l'équation

$$\left(1 - \frac{10}{6}\right) e^{-\frac{(1-m)^2}{2}} + \left(2 - \frac{10}{6}\right) e^{-\frac{(2-m)^2}{2}} + \left(3 - \frac{10}{6}\right) e^{-\frac{(3-m)^2}{2}} = 0.$$

Voici les estimations obtenus pour quelques valeurs de σ .

| | | | | | | | | |
|-----------|-------|-------|-------|-------|-------|--------|--------|--------|
| σ | 0.15 | 0.2 | 0.3 | 0.4 | 1 | 2 | 3 | 4 |
| \hat{m} | 1.516 | 1.528 | 1.562 | 1.659 | 1.366 | 0.1823 | -2.791 | -6.447 |

Deux variables normales de même moyenne mais avec des écarts types différents ont des formes différentes. Par conséquent, puisque pour estimer la moyenne on cherche la distribution théorique dont la forme colle le mieux à celle donnée par l'histogramme, l'estimation de celle-ci doit en dépendre. La nouvelle approche reflète bien ce fait là, mais pas l'approche classique. En effet, quelque soit la valeur de l'écart type on trouve la même valeur de \bar{X} .

4.3 Remarque

En plus de ce qui vient d'être dit dans les sections précédentes sur l'intérêt que peut présenter la nouvelle approche, il y a aussi le fait que la distribution de l'échantillon telle qu'elle a été définie s'y prête beaucoup plus pour être une estimation empirique de la distribution source conditionnelle sachant y_1, y_2, \dots, y_k , que de la distribution source elle-même. Par ailleurs, dans tous les cas la distribution de l'échantillon et les distributions théoriques conditionnelles candidates sont discrètes et ont même support. Cela facilite beaucoup leurs comparaisons.

Proposition 1 *Dans le cas d'une distribution normale, si l'écart type est connu, il suffit d'avoir deux points quelconques de la courbe image pour retrouver la distribution source exacte avec certitude.*

Preuve.

Soient la distribution normale de moyenne α et d'écart type σ et deux réels x et y quelconques. Déterminons les probabilités f_1 et f_2 de sorte que les deux points (x, f_1) et (y, f_2) se situent sur la courbe image de cette distribution associée aux réels x et y et calculons \bar{X} . On obtient :

$$\bar{X} = xf_1 + yf_2 = \frac{x \exp\left(\frac{-(x-\alpha)^2}{2\sigma^2}\right) + y \exp\left(\frac{-(y-\alpha)^2}{2\sigma^2}\right)}{\exp\left(\frac{-(x-\alpha)^2}{2\sigma^2}\right) + \exp\left(\frac{-(y-\alpha)^2}{2\sigma^2}\right)}$$

Remarquons que \bar{X} peut prendre n'importe quelle valeur réelle. Pour qu'il soit égal à la moyenne estimée, il est absolument nécessaire de prendre x et y disposés sur l'axe des réels de façon appropriée de part et d'autre de cette moyenne. Pour un échantillon aléatoire donné, ces conditions sont en général remplies avec certitude uniquement lorsqu'il est de taille infinie. Avec la nouvelle approche il suffit seulement que les points (x, f_1) et (y, f_2) se situent sur la courbe image. Pour déterminer une estimation de la moyenne à l'aide de la nouvelle approche on doit résoudre en m l'équation suivante:

$$\left(x - \frac{xe^{-\frac{(x-\alpha)^2}{2\sigma^2}} + ye^{-\frac{(y-\alpha)^2}{2\sigma^2}}}{e^{-\frac{(x-\alpha)^2}{2\sigma^2}} + e^{-\frac{(y-\alpha)^2}{2\sigma^2}}} \right) e^{-\frac{(x-m)^2}{2\sigma^2}} + \left(y - \frac{xe^{-\frac{(x-\alpha)^2}{2\sigma^2}} + ye^{-\frac{(y-\alpha)^2}{2\sigma^2}}}{e^{-\frac{(x-\alpha)^2}{2\sigma^2}} + e^{-\frac{(y-\alpha)^2}{2\sigma^2}}} \right) e^{-\frac{(y-m)^2}{2\sigma^2}} = 0$$

En la résolvant on obtient $m = \alpha$. C'est à dire qu'on obtient exactement la distribution source quelles que soient les valeurs de x et y .

4.4 Les critères de qualité.

Avec des échantillons différents de distributions différentes ayant des caractéristiques différentes, mais qui apportent les mêmes renseignements au sujet de la distribution source, au sens défini plus haut, on obtient exactement les mêmes estimations à l'aide de la nouvelle approche. La qualité des estimations est fonction seulement de la qualité de ces renseignements. C'est un fait nouveau en statistique et qui distingue fondamentalement la nouvelle approche de l'approche classique. Les préoccupations habituelles sur la qualité des estimations ne sont plus les mêmes ici.

5 Simulation

5.1 La loi Binomiale

Soit la distribution binomiale de paramètres $p = 0.6$ et $n = 10$. Pour estimer p à partir d'un échantillon de taille N , On procède de la manière suivante. On simule 10 échantillons de taille N . On estime p par les deux méthodes à partir de chacun ($n = 10$ donné). On note par \hat{p}_{N_j} celui obtenu à partir du $j^{ième}$ échantillon, ($j = 1, \dots, 10$), par la méthode classique. On note par \tilde{p}_{N_j} l'estimation obtenue à partir de la nouvelle approche. On calcul:

$$\hat{p}_N = \frac{\sum_{j=1}^{10} \hat{p}_{N_j}}{10}, \quad \tilde{p}_N = \frac{\sum_{j=1}^{10} \tilde{p}_{N_j}}{10},$$

pour $N = 5, 10$, et on note par $\tilde{\sigma}_N^2$ la variance mesurant la dispersion des \tilde{p}_{N_j} autour de \tilde{p}_N .

On considère ensuite les dix échantillon de tailles N ($N = 5, 10$) et on les mélange pour en faire un seul échantillon de taille $10 \times N$. On utilise ce dernier pour estimer à nouveau le paramètre p par la nouvelle approche. On désigne par \tilde{p}'_N le résultat ($N = 5, 10$). L'estimation correspondante par l'approche classique est toujours égale à la moyenne \hat{p}_N . Les résultats sont donnés dans le tableau 5.1 suivant:

Tableau 5.1

| $\beta(10, 0.6)$ | | | | |
|-------------------------------------|---------------|----------------------|-------------|----------------|
| Estimation de p , $n = 10$ donné. | | | | |
| N | \tilde{p}_N | $\tilde{\sigma}_N^2$ | \hat{p}_N | \tilde{p}'_N |
| 5 | 0.5369 | 0.0046 | 0.5920 | 0.5932 |
| 10 | 0.6314 | 0.0071 | 0.5840 | 0.5848 |

5.2 La loi normale

1. On considère la loi normale de moyenne $m = 15$ et d'écart type $\sigma = 1.2$. On estime la moyenne par les deux méthodes lorsque σ est donné pour des échantillon de tailles 20 et 50. On note par \tilde{m} l'estimation obtenue par la nouvelle approche. On a obtenu les résultats suivants:

Tableau 5.2
 $N(15, 1.2)$
Estimation de m pour $\sigma = 1.2$ donné

| N | \tilde{m} | \bar{X} |
|-----|-------------|-----------|
| 20 | 15.130 | 15.172 |
| 50 | 15.074 | 15.036 |

2. On considère la loi normale centrée réduite. On simule cinq échantillons de taille 10, chacun. On regroupe les observations de chaque échantillon en classes. On estime la moyenne avec les deux méthodes, après regroupement. On obtient les \bar{X}_i et les \tilde{m}_i , $i = 1, 2, \dots, 5$, qui figurent dans le tableau 5.3. On calcule la moyenne des estimations de chaque méthode.

Tableau 5.3
 $N(0, 1)$
Estimation de m , $\sigma = 1$ donné

| Echantillon n° | \bar{X}_i (après regr.) | \tilde{m}_i (après regr.) |
|---|---------------------------|-----------------------------|
| 1 | 0.57574 | 0.36879 |
| 2 | -0.21408 | -0.12190 |
| 3 | -0.57523 | -0.87219 |
| 4 | 0.47778 | 0.54511 |
| 5 | -0.17308 | -0.32718 |
| Moyennes des \bar{X}_i et des \tilde{m}_i | 0.018226 | -0.081474 |

Bibliographie

- [1] A. Borovkov, *Statistique mathématique*, Editions Mir, 1985.
- [2] A. Guellil, *Résolutions de certains paradoxes ou la relativité en mathématique*, Ed. Guellil (ISBN: 9961-9500-0.3), 1995.

Optimisation par la méthode adaptée d'un problème linéaire-quadratique convexe à variables bornées

M.O. BIBI et N. IKHENECHÉ
Université de Béjaïa

Résumé

On se propose de minimiser un problème linéaire-quadratique à variable bornées. La méthode utilisée, dite adaptée, tient compte des spécificités du problème et de ce fait, traite les contraintes telles qu'elles se présentent, c'est à dire, sans chercher à les modifier. Cela évite d'agrandir les dimensions du problème et permet donc d'économiser l'espace mémoire sur ordinateur. En outre, cette méthode présente l'avantage d'arrêter l'algorithme dès qu'une solution suboptimale satisfaisante est obtenue.

Mots clés : Optimisation, méthode adaptée, problème linéaire-quadratique, convexe, variables bornées.

Introduction

Le principe de la méthode adaptée est pratiquement le même que celui de la méthode directe de support R. Gabassov et F.M. Kirillova [4], c'est à dire, qu'au lieu d'utiliser la métrique du simplexe en changeant un seul indice non basique j_0 , on utilisera une autre métrique dite adaptée qui consiste à considérer tous les indices non optimaux en fonction desquels on construit une direction d'amélioration de la fonction objectif et le pas le long de cette direction.

1 Position du Problème et Définitions

Le problème de (PQC) à variable bornées se présente comme suit ¹ :

$$F(x) = \frac{1}{2}x'Dx + c'x \longrightarrow \min, \quad (1)$$

$$Ax = b, \quad (2)$$

$$d^- \leq x \leq d^+, \quad (3)$$

où $D' = D \geq 0$; $\text{rang}A=m$; c, d^-, d^+ et x sont des n -vecteurs; b un m -vecteur; A une matrice d'ordre $m \times n$, avec $\text{rang}A = m < n$.

Soit $I = \{1, 2, \dots, m\}$: l'ensemble d'indices de lignes de A ; $J = \{1, 2, \dots, n\}$: l'ensemble d'indices de colonnes de A ; $J = J_B \cup J_N$ avec $J_B \cap J_N = \phi$, $|J_B| = m$. On peut alors

¹le symbole (') désigne la transposition

fractionner les vecteurs et la matrice de la manière suivante :

$d^- = d^-(J) = (d_j^-, j \in J)$, $d^+ = d^+(J)$, $x = x(J) = (x_j, j \in J)$, $x' = (x'_B | x'_N)$,
 $x_B = x(J_B) = (x_j, j \in J_B)$, $x_N = x(J_N) = (x_j, j \in J_N)$, $c' = (c'_B | c'_N)$, $c_B = c(J_B)$,
 $c_N = c(J_N)$, $A = A(I, J) = (a_{ij}, 1 \leq i \leq m, 1 \leq j \leq n)$; $A = (a_j, j \in J)$,
 $a_j = (a_{1j} \dots a_{mj})'$, $A = (A_B | A_N)$, $A_B = A(I, J_B)$, $A_N = A(I, J_N)$. Voici quelques définitions.

- i) Un vecteur x vérifiant les contraintes (2)-(3) est appelé *plan* du problème (1)-(3).
- ii) Un plan x^0 est dit *optimal* si $F(x^0) = \frac{1}{2}x^{0'}Dx^0 + c'x^0 = \min \frac{1}{2}x'Dx + c'x$, où x est pris parmi tous les vecteurs vérifiant les contraintes (2)-(3).
- iii) Un plan x^ϵ est appelé ϵ -*optimal* ou *suboptimal* si $F(x^\epsilon) - F(x^0) \leq \epsilon$, où ϵ est un nombre arbitraire positif ou nul et x^0 est une solution optimale du problème (1)-(3).
- iv) L'ensemble $J_B \subset J$, $|J_B| = m$, est appelé *support* des contraintes (2)-(3) si $\det A_B \neq 0$.
- v) Le couple $\{x, J_B\}$ est appelé *plan de support* des contraintes.
- vi) Le plan de support est dit *non dégénéré* si $d_j^- < x_j < d_j^+$, $j \in J_B$.

2 Formule d'Accroissement de la Fonction Objectif

Soit $\{x, J_B\}$ un plan de support des contraintes du problème (1)-(3) et considérons un autre plan quelconque $\bar{x} = x + \Delta x$. L'accroissement de la fonction objectif s'écrit alors :

$$F(\bar{x}) - F(x) = g'(x)(\Delta x) + \frac{1}{2}(\Delta x)'D(\Delta x), \quad (4)$$

$g(x) = Dx + c$ est le gradient de la fonction (1).

Par ailleurs, on a $\begin{cases} Ax = b, \\ A\bar{x} = b. \end{cases} \Leftrightarrow A\bar{x} = A(\Delta x + x) = A\Delta x + Ax = b \Rightarrow A\Delta x = 0$.

En posant : $(\Delta x)' = ((\Delta x_B)' | (\Delta x_N)')$, $\Delta x_B = \Delta x(J_B)$, $\Delta x_N = \Delta x(J_N)$, l'égalité $A\Delta x = 0$ peut s'écrire $A_B\Delta x_B + A_N\Delta x_N = 0$; d'où $\Delta x_B = -A_B^{-1}A_N\Delta x_N$. On définit le vecteur des potentiels u et le vecteur des estimations E telle que :

$$u' = g'_B A_B^{-1}, \quad E' = E'(J) = g' - u'A = (E'_B, E'_N),$$

où $E'_B = 0$, $E'_N = g'_N - u'A_N$, et $g' = (g'_B, g'_N)$.

En remplaçant Δx_B par sa valeur, la formule (4) aura la forme suivante :

$$F(\bar{x}) - F(x) = E'_N \Delta x_N + \frac{1}{2}(\Delta x_N)'M(\Delta x_N), \quad (5)$$

où $M = M(J_N, J_N) = Z'DZ$, $Z = Z(J, J_N) = \begin{pmatrix} -A_B^{-1}A_N \\ I_N \end{pmatrix}$, $I_N = I(J_N, J_N)$ est la matrice identité d'ordre (n-m)

3 Critère d'optimalité

Théorème 3.1 Soit $\{x, J_B\}$ un plan de support des contraintes du problème (1) - (3).

Alors les relations :

$$\begin{cases} E_j \geq 0, & \text{si } x_j = d_j^-, \\ E_j \leq 0, & \text{si } x_j = d_j^+, \\ E_j = 0, & \text{si } d_j^- < x_j < d_j^+ \text{ pour } \forall j \in J_N, \end{cases} \quad (6)$$

sont suffisantes pour l'optimalité du plan x .

Ces mêmes relations sont aussi nécessaires dans le cas où le plan de support des contraintes est non dégénéré.

Preuve :

Suffisance. Soit $\{x, J_B\}$ un plan de support des contraintes vérifiant les relations (6). Pour tout plan \bar{x} du problème (1)-(3), la formule d'accroissement (5) donne alors :

$$F(\bar{x}) - F(x) \geq E'_N \Delta x_N \geq \sum_{\substack{j \in J_N \\ E_j \geq 0}} E_j(\bar{x}_j - d_j^-) + \sum_{\substack{j \in J_N \\ E_j \leq 0}} E_j(\bar{x}_j - d_j^+) \geq 0,$$

car la matrice M est semi-définie positive. D'où $F(\bar{x}) \geq F(x)$. Donc le plan x est optimal.

Nécessité. Soit $\{x, J_B\}$ un plan de support optimal non dégénéré du problème (1)-(3) et supposons que les relations (6) ne sont pas vérifiées, i.e., il existe au moins un indice $j_0 \in J_N$, tel que $E_{j_0} < 0$, $x_{j_0} < d_{j_0}^+$ ou $E_{j_0} > 0$, $x_{j_0} > d_{j_0}^-$. On construit un autre plan $\bar{x} = x + \theta l$, où θ est un nombre réel positif et $l = l(J)$ est un vecteur de direction défini comme suit :

$$\begin{cases} l_{j_0} = -\text{sign} E_{j_0}, \\ l_j = 0, j \neq j_0, j \in J_N. \end{cases}$$

D'autre part, on doit avoir $A\bar{x} = Ax + \theta Al = b \Leftrightarrow Al = A_B l_B + A_N l_N = 0$. D'où $l_B = -A_B^{-1} A_N l_N = A_B^{-1} a_{j_0} \text{sign} E_{j_0}$.

En plus le vecteur \bar{x} doit vérifier l'inégalité $d^- \leq x + \theta l \leq d^+$. Soit, en écrivant composante par composante

$$\begin{cases} d_j^- - x_j \leq \theta l_j \leq d_j^+ - x_j, & j \in J_B, \\ d_{j_0}^- - x_{j_0} \leq \theta l_{j_0} \leq d_{j_0}^+ - x_{j_0}. \end{cases}$$

Comme le plan de support est non-dégénéré, on trouve toujours un nombre θ positif assez petit tel que le vecteur \bar{x} sera un plan pour le problème (1)-(3). Pour θ et l ainsi choisis la formule d'accroissement (5) nous donne :

$$\begin{aligned} F(\bar{x}) - F(x) &= \theta E'_N l_N + \frac{1}{2}(\theta)^2 (l_N)' M (l_N) = \varphi(\theta) \\ &= \theta(-|E_{j_0}| + \frac{1}{2}(\theta)(l_N)' M (l_N)) < 0, \end{aligned} \quad (7)$$

ce qui contredit l'optimalité de x . Donc les relations (6) sont forcément vérifiées.

4 Critère de Suboptimalité

Le nombre $\beta(x, J_B)$ appelé *estimation de suboptimalité* estime l'écart qui existe entre la valeur optimale $F(x^0)$ et une autre valeur $F(x)$ d'un plan de support quelconque $\{x, J_B\}$.

Théorème 4.1 Soit $\{x, J_B\}$ un plan de support des contraintes du problème (1)-(3), et ϵ un nombre positif ou nul arbitraire. Si le nombre

$$\beta(x, J_B) = \sum_{j \in J_N, E_j > 0} E_j(x_j - d_j^-) + \sum_{j \in J_N, E_j < 0} E_j(x_j - d_j^+) \leq \epsilon, \quad (8)$$

alors $\{x, J_B\}$ est ϵ -optimal.

5 Méthode de résolution

5.1 Définitions

- a- On appelle support de la fonction objectif (1) l'ensemble des indices $J_S \subset J_N$ tel que $\det M_S = \det M(J_S, J_S) \neq 0$. On posera $J_{NN} = J_N \setminus J_S$.
- b- On appelle support du problème (1)-(3) l'ensemble $J_P = \{J_B, J_S\}$ formé du support des contraintes J_B et de celui de la fonction objectif J_S .
- c- On appelle plan de support du problème (1)-(3) la paire $\{x, J_P\}$ formé du plan x et du support J_P ; il est dit accordé si $E(J_S) = 0$.

5.2 Construction d'une direction d'amélioration adaptée

La méthode directe de support est une méthode d'amélioration basée sur la métrique du simplexe, mais les plans de support admettent encore d'autres métriques pour les directions d'amélioration. Considérons la métrique suivante pour les composantes non basiques de la direction admissible l :

$$d_j^- - x_j \leq l_j \leq d_j^+ - x_j, \quad j \in J_{NN} = J_N \setminus J_S \quad (9)$$

Cette métrique dépend du plan courant x , et de ce fait, elle dite adaptée. Afin de calculer les composantes de la direction admissible d'amélioration l , considérons l'accroissement

$$\Delta F = F(x+l) - F(x) = \sum_{j \in J_N, E_j > 0} E_j l_j + \sum_{j \in J_N, E_j < 0} E_j l_j + \frac{1}{2} l'_N M l_N$$

En tenant compte de la métrique (9), la partie linéaire de ΔF atteint son minimum pour les valeurs des composantes $l(J_{NN})$ suivantes :

$$l_j = \begin{cases} d_j^- - x_j & \text{si } E_j > 0, \\ d_j^+ - x_j & \text{si } E_j < 0, \\ 0 & \text{si } E_j = 0, \end{cases} \quad \text{pour } j \in J_{NN}. \quad (10)$$

Quant aux composantes $l(J_S)$, elles se déduisent à partir de $E_j(x+l) = 0$, $j \in J_S$. Donc $M(J_S, J_S)l(J_S) + M(J_S, J_{NN})l(J_{NN}) = 0$. D'où

$$l(J_S) = -M_S^{-1} M(J_S, J_N) l(J_N). \quad (11)$$

Les composantes $l(J_B)$ seront déduites de $Al = 0$; d'où

$$l(J_B) = -A_B^{-1} (A_S l_S + A_{NN} l_{NN}). \quad (12)$$

5.3 Changement de plan

On construit alors le nouveau plan \bar{x} sous la forme $\bar{x} = x + \theta^0 l$, avec $\theta^0 = \min\{1, \theta_{j_1}, \theta_{j_S}, \theta_F\}$. Les nombre θ_{j_1} , et θ_{j_S} se calculent de façon à ce que les contraintes directes sur le vecteur \bar{x} soient vérifiées :

$$d_j^- - x_j \leq \theta^0 l_j \leq d_j^+ - x_j, \quad j \in J_B, \quad (13)$$

$$d_j^- - x_j \leq \theta^0 l_j \leq d_j^+ - x_j, \quad j \in J_S. \quad (14)$$

Donc $\theta_{j_1} = \min_{j \in J_B} \theta_j$, $\theta_{j_s} = \min_{j \in J_S} \theta_j$, Où

$$\theta_j = \begin{cases} \frac{d_j^+ - x_j}{l_j} & \text{si, } l_j > 0, \\ \frac{d_j^- - x_j}{l_j} & \text{si, } l_j < 0, \\ \infty, & \text{si, } l_j = 0. \end{cases}$$

Quant à θ_F , il se calcul Pour que le passage de x à \bar{x} assure une diminution maximale de la fonction objectif et que $\bar{E}_{j_*} = 0$, où j_* se déduit comme suit : $\theta_F = \sigma_{j_*} = \min \sigma_j$, $j \in J_{NN}$

$$\sigma_j = \begin{cases} -\frac{E_j}{\delta_j}, & \text{si } E_j \delta_j < 0; \\ \infty, & \text{si } E_j \delta_j \geq 0. \end{cases}$$

$$\delta_j = M(j, J_N \setminus J_B).$$

5.4 Changement de support

- Pour $\theta^0 = 1$, on pose $\bar{x} = x + l$ alors, $\bar{J}_B = J_B$, $\bar{J}_S = J_S$. On effectue une nouvelle itération si $\beta(\bar{x}, J_B) > \epsilon$, où ϵ positif où nul donné à l'avance.

Le changement de support s'effectue lorsque $\theta^0 < 1$ et $\beta(\bar{x}, J_B) > \epsilon$. Trois cas peuvent alors se présenter :

- $\theta^0 = \theta_{j_1} \leq 1$: contrairement à la méthode du simplexe, le choix de l'indice j_0 n'est pas unique, ce qui fait la particularité de cette méthode. Lorsque ce cas se réalise pour un indice $j_1 \in J_B$, on a alors $l_{j_1} = -\sum_{j \in J_N} e'_{j_1} A_B^{-1} a_j l_j = -\sum_{j \in J_N} x_{j_1 j} l_j \neq 0$, où e est un vecteur unitaire de dimension m . Il existe alors $j_0 \in J_N$ tel que $x_{j_1 j_0} \neq 0$. Cette dernière condition nous assure par conséquent que $\bar{J}_B = (J_B \setminus j_1) \cup j_0$ est bel et bien un support. Si on peut avoir $x_{j_1 j_0} \neq 0$, avec $j_0 \in J_S$, on posera donc $\bar{J}_B = (J_B \setminus j_1) \cup j_0$, $\bar{J}_S = (J_S \setminus j_0)$.
sinon, on choisira un indice $j_0 \in J_{NN}$ tel que $l_{j_0} \neq 0$ et on posera $\bar{J}_B = (J_B \setminus j_1) \cup j_0$ et $\bar{J}_S = J_S$
- $\theta^0 = \theta_{j_s}$ on pose $\bar{J}_B = J_B$, $\bar{J}_S = J_S \setminus j_s$.
- $\theta^0 = \theta_F$: dans ce cas on aura $\bar{J}_B = J_B$, $\bar{J}_S = J_S \cup j_*$.

5.5 Algorithme de résolution

Algorithme

Début

- 1) Soit un plan de support initial $\{x, J_P\}$ du problème (1)-(3) tel que $J_P = \{J_B, J_S\}$ avec $J_S = \emptyset$ pour plus de facilité
- 2) calculer le vecteur des estimations $E' = E'(J) = (g_j, j \in J_N)' - u' A_N$.
Si $\beta(x, J_B) \leq 0$, alors aller à Fin,
Sinon,
Calculer la direction d'amélioration l .
Calculer le pas $\theta^0 = \min(1, \theta_{j_1}, \theta_{j_s}, \theta_F)$.
Calculer $\bar{x} = x + \theta^0 l$
- 3) Changement de support
 - i) Si $\theta^0 = 1$, alors $\bar{x} = x + l$, $\bar{J}_B = J_B$, $\bar{J}_S = J_S$

si $\beta(\bar{x}, J_B) \leq 0$ alors aller à Fin.
 sinon aller en 2.
 fin si.
 fin si.
 ii) Si $\theta^0 = \theta_{j_1}$, alors :
 si $j_0 \in J_S$ alors, $\bar{J}_B = (J_B \setminus j_1) \cup j_0$, $\bar{J}_S = J_S \setminus j_0$.
 si $j_0 \in J_{NN}$ alors, $\bar{J}_S = J_S$.
 Aller en 2 en partant de $\{\bar{x}, \bar{J}_P\}$.
 fin si.
 iii) Si $\theta^0 = \theta_{j_s}$, alors $\bar{J}_B = J_B$, $\bar{J}_S = J_S \setminus j_s$.
 Aller en 2 en partant de $\{\bar{x}, \bar{J}_P\}$.
 fin si.
 iv) Si $\theta^0 = \theta_F$, alors $\bar{J}_B = J_B$, $\bar{J}_S = J_S \cup j_*$.
 Aller en 2 en partant de $\{\bar{x}, \bar{J}_P\}$.
 fin si.

Fin.

Conclusion

L'avantage de cette méthode par rapport à celle du simplexe [1], est le gain en temps et en espace mémoire qu'elle assure vu qu'elle manipule les bornes des variables telles qu'elles se présentent.

Références

- [1] G.B. Dantzig. *Programming in Linear Structure*, volume 17(1). Econometrica, 1949.
- [2] Kostyukova O.I. et Raketskii V.M. Méthode de diminution de l'estimation de suboptimalité en programmation quadratique. *Actes de l'Académie des Sciences de Biélorussie*, 29(12) :1072–1075, 1985.
- [3] A. Faradji. Algorithmes de minimisation d'une fonctionnelle quadratique. Master's thesis, Université de Tizi-Ouzou, 1998.
- [4] R. Gabassov, O.I. Kostyukova, and F.M. Kirillova. Résolution des problèmes extrémaux linéaire-quadratiques. *Actes de l'académie des sciences (URSS)*, 280(03) :529–533, 1985.
- [5] Bibi M.O. *Résolution d'un problème de programmation quadratique avec un grand nombre de variables.- in Méthode de résolution des problèmes linéaire-quadratiques de commande optimale*. PhD thesis, Université de Minsk, 1985.
- [6] Bibi M.O. *Méthodes adaptées de programmation linéaire*. Cours de postgraduation en recherche opérationnelle, Université de Béjaïa, 2000.

II - Processus Aléatoires

Sur le théorème ergodique pour l'action d'un semi-groupe libre

CHARLOT François - Université de Rouen
Francois.Charlot@univ-rouen.fr

25 février 2005

Table des matières

| | |
|--|-----------|
| 1 Introduction | 1 |
| 2 Le modèle | 3 |
| 3 Théorèmes ergodiques | 6 |
| 3.1 Convergence presque sûr | 6 |
| 3.2 Convergence dans les L^p | 9 |
| 4 Conclusion | 10 |

1 Introduction

Soit G un groupe fini et $(\theta_g, g \in G)$ un groupe indéré par G d'automorphismes d'un espace de probabilité $(\Omega, \mathcal{F}, \mathbb{P})$. Soit $(A_n, n \in \mathbb{N})$ une suite de parties finies de G telle que $\lim_{n \rightarrow \infty} \text{card}(A_n) = \infty$. U étant une variable aléatoire à valeurs réelles définie sur $(\Omega, \mathcal{F}, \mathbb{P})$, à quelles conditions et en quelle sens a-t-on la convergence de la suite $(S_n(U), n \in \mathbb{N})$ définie par

$$S_n(U) = \frac{1}{\text{card}(A_n)} \sum_{g \in A_n} U \circ \theta_g ?$$

La littérature concernant ce problème est abondante lorsque le groupe est commutatif ou moyennable (voir par exemple [18] et [13]). A l'opposé, pour le groupe libre, les résultats sont moins nombreux et, pour la plupart, plus récents. Ils s'intéressent à la convergence, lorsque G est engendré par un

alphabet fini, des S_n lorsque A_n est l'ensemble des mots réduits de longueur n et à des généralisations de ce problème. Le premier d'entre eux est celui de Y. Guivarc'h [12] en 1969 qui démontre la convergence dans $L^2(\Omega, \mathcal{F}, \mathbb{P})$ de $S_{2n}(U)$. A. Nevo et E.M. Stein [15] établissent en 1994, dans un long article, la convergence presque sûre des $S_{2n}(U)$ lorsque $U \in L^2(\Omega, \mathcal{F}, \mathbb{P})$. A. I. Bufetov en 2000 [5] (voir aussi [6]) généralise légèrement le résultat de A. Nevo et E.M. Stein et surtout en donne une superbe et rapide démonstration. Des résultats connexes sur ces groupes sont aussi étudiés : R. I. Grigorchuk [11], A. I. Bufetov [3], [4] et [6], X. Bressaud et V. Maume-Deschamps [2]. Signalons enfin l'étude de ces théorèmes dans un cadre non-commutatif pour des applications à la mécanique quantique : J. P. Conze et N. Dang Ngoc [8], Trent E. Walker [19] et Claire Anantharaman-Dellaroché [1].

La démonstration de A. I. Bufetov ([5] et [6]) se base sur le résultat suivant dû à G. C. Rota [17] en 1962 : si (E, \mathcal{E}, μ) est un espace de probabilité et π un opérateur markovien sur $L^1(E, \mathcal{E}, \mu)$ tel que, $\forall F \in \mathcal{E}$, $\int_E \mu(dx) \pi(x, F) = \mu(F)$, alors, pour toute fonction φ mesurable de E dans \mathbb{R} telle que $\int_E |\varphi(x)| |\ln(|\varphi(x)|)| \mu(dx) < +\infty$, $\pi^n(\pi^*)^n \varphi$ converge μ -p.s. et dans $L^1(E, \mathcal{E}, \mu)$, où π^* est l'adjoint de π par rapport à μ , c'est à dire que, pour tout couple $(f, g) \in L^2(E, \mathcal{E}, \mu) \times L^2(E, \mathcal{E}, \mu)$, $\int_E g(x) \pi(f)(x) \mu(dx) = \int_E f(x) \pi^*(g)(x) \mu(dx)$.

Lorsqu'on se pose le même type de question lorsque G est le semi-groupe libre engendré par un alphabet fini A , on voit que, si on pose $\Psi(U) = \frac{1}{\#A} \sum_{\alpha \in A} U \circ \theta_\alpha$, alors $S_n(U) = \Psi^n(U)$. Le théorème ergodique nous renseigne sur la convergence des $\frac{1}{n} \sum_{k=1}^{k=n} S_k(U)$ mais la convergence des $S_n(U) = \Psi^n(U)$ est une toute autre histoire ! Le résultat est alors du même type que ceux obtenus pour la convergence des itérés d'un noyau markovien récurrent positif apériodique (voir par exemple [16] ou [14]).

Dans ce document nous établissons la convergence presque sûre des $\Psi^n(U)$ lorsque U est une variable de $L^2(\Omega, \mathcal{F}, \mathbb{P})$ et dans les $L^p(\Omega, \mathcal{F}, \mathbb{P})$, $p \geq 1$ pour U dans $L^p(\Omega, \mathcal{F}, \mathbb{P})$, dans le cas très particulier où $(\Omega, \mathcal{F}, \mathbb{P})$ est l'espace canonique d'une "chaîne de Markov" récurrente (au sens de Harris) positive et apériodique indéxée par G , objet bien sûr qu'il nous faut définir. Le résultat est alors une application du théorème d'Orey (voir par exemple le livre de D. Revuz [16]) et de sa généralisation que l'on trouve dans le livre de S. P. Meyn et R. L. Tweedie [14], ce qui n'a bien sûr rien de surprenant !

Ce problème m'a été inspiré par les travaux sur la phylogénie des protéines de M. O. Dayhoff, R. V. Eck et C. M. Park (1972, [9] et [7]). En effet, ces auteurs reconstituent des arbres de protéines qui peuvent être vus comme des arbres de mots d'un groupe libre engendré par deux lettres, les

protéines étant alors, du fait des hypothèses faites par les auteurs, un échantillon d'une "chaîne de Markov" indexée par ce groupe. Les statistiques faites alors semblent supposer l'existence de théorèmes ergodiques, les ensembles A_n étant plus compliqués que ceux indiqués précédemment. J'indique en conclusion dans quelle direction je me suis tourné pour répondre le mieux que je pouvais à ce problème.

2 Le modèle

Soit \mathbf{A} un alphabète fini, $\#\mathbf{A} > 1$ où $\#\mathbf{A} = \text{card}(\mathbf{A})$. Les lettres seront écrites en caractères grecs. Soit \mathbf{G} le semi-groupe des mots de longueur finie sur l'alphabète \mathbf{A} (l'opération étant bien sûr la concaténation), et $\mathcal{G} = \mathbf{A}^{\mathbb{N}}$ l'ensemble des mots de longueur infinie. On notera par \emptyset le mot vide. Si $g \in \mathcal{G}$, g_n désigne le mot de longueur n obtenu en ne gardant que les n premières lettres de g . Si $g \in \mathbf{G}$, on note $|g|$ la longueur du mot g , c'est à dire le nombre de lettres composant g . Soit $g \in \mathbf{G}$; si $|g| > 1$, on notera alors $g-1$ le mot obtenu en supprimant la dernière lettre de g c'est à dire, si $g = l\alpha$ avec $l \in \mathbf{G}$ et $\alpha \in \mathbf{A}$, alors $g-1 = l$. On définit de façon récursive $g-n$ par : si $|g| \geq 0$, $g-0 = g$ et si $|g| \geq n+1$, $g-(n+1) = (g-n)-1$. Soit $(g, g') \in \mathbf{G}$; on écrira $g \preceq g'$ si $|g'| \geq |g|$ et s'il existe $n \in \mathbb{N}$, $n \leq |g'|$, tel que $g = g' - n$. On écrira $g \prec g'$ si $g \preceq g'$ et $g \neq g'$. Un sous ensemble $D \subseteq \mathbf{G}$ est un **arbre** si $\#D < \infty$, $\emptyset \in D$ et, $\forall g \in D$ tel que $|g| \neq 0$, $g-1 \in D$; on notera par \mathcal{D} l'ensemble des arbres de \mathbf{G} .

Soit $(\Omega, \mathcal{F}, \mathbb{P})$ un espace de probabilité et (E, \mathcal{E}) un espace mesurable qui sera l'espace des états de nos variables aléatoires. Soit $\mathbf{X} = (X_g, g \in \mathbf{G}) = (X_{\bar{g}}, \bar{g} \in \mathcal{G}, n \in \mathbb{N})$ une famille indexée par \mathbf{G} de variables aléatoires définies sur $(\Omega, \mathcal{F}, \mathbb{P})$ à valeurs dans (E, \mathcal{E}) . On dira que \mathbf{X} est une \mathbf{G} -chaîne et on notera aussi, pour $\bar{g} \in \mathcal{G}$, $\mathbf{X}_{\bar{g}} = (X_{\bar{g}n}, n \in \mathbb{N})$. Une filtration est une famille croissante $(\mathcal{F}_g, g \in \mathbf{G})$ de sous σ -algèbres de \mathcal{F} . La \mathbf{G} -chaîne \mathbf{X} sera dite adaptée à la filtration $(\mathcal{F}_g, g \in \mathbf{G})$ si, pour tout $g \in \mathbf{G}$, X_g est \mathcal{F}_g mesurable et que de plus $\sigma\{X_h/h \neq g\} \subseteq \mathcal{F}_g$.

Définition 1 Soit \mathbf{X} une \mathbf{G} -chaîne adaptée à la filtration $(\mathcal{F}_g, g \in \mathbf{G})$ et Π une transition de probabilité sur (E, \mathcal{E}) . \mathbf{X} est une \mathbf{G} -chaîne de Markov si, pour tout $g \in \mathbf{G}$, pour tout $\alpha \in \mathbf{A}$, et tout $F \in \mathcal{E}$,

$$\mathbb{P}(X_{g\alpha} \in F / \mathcal{F}_g) = \Pi(X_g, F) .$$

Soit $\Omega = E^{\mathbf{G}}$ et pour $g \in \mathbf{G}$, X_g définie par

$$\forall (e_h, h \in \mathbf{G}) \in \Omega, \forall g \in \mathbf{G}, X_g(e_h, h \in \mathbf{G}) = e_g .$$

On note encore bien sûr $X = (X_g, g \in \mathbf{G})$. Soit $\mathcal{F} = \sigma\{X_g, g \in \mathbf{G}\}$ et, pour $g \in \mathbf{G}$, $\mathcal{F}_g = \sigma\{X_h, h \in \mathbf{G}, h \neq g\}$. Pour une probabilité μ sur (E, \mathcal{E}) , il existe une unique probabilité \mathbf{P}_μ sur (Ω, \mathcal{F}) telle que

$$\forall D \in \mathcal{D}, \forall (F_g, g \in D) \in \mathcal{E}^D,$$

$$\mathbf{P}_\mu \left(\bigcap_{g \in D} \{X_g \in F_g\} \right) = \int_{F_\emptyset} \mu(dx_\emptyset) \prod_{g \in D, g \neq \emptyset} \int_{F_g} \Pi(x_{g-1}, dx_g).$$

Pour cette probabilité, X est une \mathbf{G} -chaîne de Markov de probabilité "initiale" μ par rapport à la filtration $(\mathcal{F}_g, g \in \mathbf{G})$. Comme d'ordinaire on notera, pour un point $x \in E$, \mathbf{P}_x à la place de \mathbf{P}_{δ_x} . L'espace

$$(\Omega, \mathcal{F}, (\mathcal{F}_g, g \in \mathbf{G}), \mathbf{P}_x)_{x \in E}$$

sera appelé l'espace canonique de la \mathbf{G} -chaîne de Markov de transition Π .

Pour $g \in \mathbf{G}$, on notera θ_g l'application de Ω dans lui-même définie par

$$\forall h \in \mathbf{G}, \forall \omega \in \Omega, X_h \theta_g(\omega) \stackrel{\text{def}}{=} X_h(\theta_g(\omega)) = X_{gh}(\omega).$$

Par la suite on notera toujours $U\theta_g$ pour $U \circ \theta_g$. $\Theta \stackrel{\text{def}}{=} (\theta_g, g \in \mathbf{G})$ est un semi-groupe libre pour la composition des applications. Dans la suite μ sera toujours une probabilité sur (E, \mathcal{E}) telle que $\mu\Pi = \mu$, c'est-à-dire $\forall F \in \mathcal{E}$, $\int_E \mu(dx) \pi(x, F) = \mu(F)$; alors, pour tout $A \in \mathcal{F}$ et pour tout $g \in \mathbf{G}$, $\mathbf{P}_\mu(\theta_g^{-1}(A)) = \mathbf{P}_\mu(A)$. Autrement dit Θ est un semi-groupe libre de transformations préservant \mathbf{P}_μ .

On notera, pour tout entier p , $\mathcal{X}_p = \sigma\{X_g / |g| \leq p\}$.

Lemme 1 Soit $p \in \mathbf{N}^*$ et $\{U_g, |g| = p\}$ une famille de variables aléatoires positives ou bornées. On a alors, pour tout $x \in E$,

$$\mathbf{E}_x \left(\prod_{|g|=p} U_g \theta_g / \mathcal{X}_p \right) = \prod_{|g|=p} \mathbf{E}_{X_g}(U_g) \mathbf{P}_x - p.s. .$$

Démonstration Soit h tel que $|h| = p$. Pour g tel que $|g| = p$ et $g \neq h$, $U_g \theta_g$ est mesurable par rapport à \mathcal{F}_h . Par les propriétés de l'espérance conditionnelle, on obtient, puisque la tribu \mathcal{X}_p est incluse dans \mathcal{F}_h ,

$$\begin{aligned} \mathbf{E}_x \left(\prod_{|g|=p} U_g \theta_g / \mathcal{X}_p \right) &= \mathbf{E}_x \left(\mathbf{E}_x(U_h \theta_h / \mathcal{F}_h) \prod_{|g|=p, g \neq h} U_g \theta_g / \mathcal{X}_p \right) \\ &= \mathbf{E}_x \left(\mathbf{E}_{X_h}(U_h) \prod_{|g|=p, g \neq h} U_g \theta_g / \mathcal{X}_p \right) = \mathbf{E}_{X_h}(U_h) \mathbf{E}_x \left(\prod_{|g|=p, g \neq h} U_g \theta_g / \mathcal{X}_p \right) \end{aligned}$$

d'où le résultat. ■

On note $D_n = \{g \in \mathcal{G}, |g| \leq n\}$ et $\overline{D}_n = \{g \in \mathcal{G}, |g| = n\}$. On a $\#\overline{D}_n = \#\mathbf{A}^n$ et $\#D_n = \frac{\#\mathbf{A}^{n+1}-1}{\#\mathbf{A}-1}$. Pour une variable aléatoire U sur Ω , on notera

$$\Psi^0(U) = U, \quad \Psi^1(U) = \Psi(U) = \frac{1}{\#\mathbf{A}} \sum_{\alpha \in \mathbf{A}} U\theta_\alpha.$$

Pour tout $n \in \mathbb{N}$, on a

$$\Psi^n(U) = \frac{1}{\#\overline{D}_n} \sum_{g \in \overline{D}_n} U\theta_g.$$

On a bien sûr, $\Psi(1) = 1$, $\Psi(U) \geq 0$ si $U \geq 0$ et $\mathbb{E}_\mu(\Psi(U)) = \mathbb{E}_\mu(U)$ pour tout $U \in L^1(\Omega, \mathcal{F}, \mathbb{P}_\mu)$. Ψ est un opérateur contractant sur $L^p(\Omega, \mathcal{F}, \mathbb{P}_\mu)$, pour tout $p \in [1, \infty]$. C'est donc un opérateur markovien sur $L^1(\Omega, \mathcal{F}, \mathbb{P}_\mu)$. $(\Omega, \mathcal{F}, \mathbb{P}_\mu, \Psi)$ est aussi un processus de Markov au sens de S. R. Foguel [10]. Nous obtenons dans ce papier des théorèmes de convergence pour les suites $(\Psi^n(U), n \in \mathbb{N})$ et $(\frac{1}{\#\overline{D}_n} \sum_{g \in \overline{D}_n} U\theta_g, n \in \mathbb{N})$.

Dans la suite la phrase "un noyau markovien sur (E, \mathcal{E}) , récurrent positif" signifie que le noyau est récurrent positif au sens de Harris. Mais, dans le cas où E est dénombrable, il s'agit de la récurrence au sens "habituel", ce qui est tout à fait suffisant pour comprendre ce qui suit.

Rappelons un résultat que l'on trouve en particulier dans le livre de S. P. Meyn et R. L. Tweedie [14]. Il s'agit d'une extension du théorème classique d'Orey (pour ce dernier théorème, voir aussi D. Revuz [16]).

Théorème 1 *Soit Π un noyau markovien sur (E, \mathcal{E}) , récurrent positif et apériodique, de probabilité invariante μ . Soit f une fonction mesurable de E dans $[1, \infty[$. Si $\mu(f) < \infty$, il existe un ensemble S_f , $\mu(S_f) = 1$, tel que*

$$\forall x \in S_f, \quad \limsup_{n \rightarrow \infty} \{|\Pi^n(x) - \mu(x)| / \mu(x) \mid x \in \mathcal{E}, |x| \leq f\} = 0.$$

En particulier donc, pour toute fonction $g \in L^1(E, \mathcal{E}, \mu)$, la suite $(\Pi^n(g), n \in \mathbb{N})$ converge μ presque partout vers $\mu(g)$.

Le corollaire suivant se déduit facilement du théorème 1 (voir par exemple [16] chapitre 6 exercices de la section 2).

Corollaire 1 *Avec les hypothèses du théorème 1, $\lim_{n \rightarrow \infty} \Pi^n(g) = \mu(g)$ dans $L^p(E, \mathcal{E}, \mu)$ pour toute fonction $g \in L^p(E, \mathcal{E}, \mu)$, $1 \leq p < \infty$.*

3 Théorèmes ergodiques

Dans la suite de cette section le noyau markovien Π est supposé récurrent positif et apériodique.

3.1 Convergence presque sûr

Pour une fonction f mesurable de E dans \mathbb{R} on notera

$$\psi(f) = \frac{1}{\#\mathbf{A}} \sum_{\alpha \in \mathbf{A}} f(X_\alpha) = \Psi(f(X_0)) .$$

et

$$\psi_n(f) = \frac{1}{\#\mathbf{A}^n} \sum_{g \in \mathcal{G}, |g|=n} f(X_g) = \Psi^n(f(X_0)) .$$

Théorème 2 *Pour tout $U \in L^2(\Omega, \mathcal{F}, \mathbf{P}_\mu)$ les suites $(\Psi^n(U), n \in \mathbb{N})$ et $(\frac{1}{\#D_n} \sum_{g \in D_n} U\theta_g, n \in \mathbb{N})$ convergent vers $\mathbf{E}_\mu(U)$, \mathbf{P}_μ -presque sûrement.*

Pour cela, nous commençons d'abord par une "loi des grands nombres".

Proposition 1 *Soit f une fonction de $L^2(E, \mathcal{E}, \mu)$. Alors, $\psi_n(f)$ tend \mathbf{P}_μ -p.s. vers $\mu(f)$ quand n tend vers l'infini.*

Un petit calcul pour débiter.

Lemme 2 *Pour une fonction f de $L^2(E, \mathcal{E}, \mu)$, pour tout entier $p \geq 1$,*

$$\mathbf{E}_\mu (((\psi_p - \psi_{p-1}\Pi)(f))^2) = \frac{1}{\#\overline{D}_p} (\mu(f^2) - \mu((\Pi f)^2)) .$$

Démonstration du lemme

$$\begin{aligned} \mathbf{E}_\mu (((\psi_p - \psi_{p-1}\Pi)(f))^2) &= \mathbf{E}_\mu \left[\left(\frac{1}{\#\overline{D}_{p-1}} \sum_{g \in \overline{D}_{p-1}} \psi(f) \circ \theta_g - \Pi f(X_g) \right)^2 \right] \\ &= \frac{1}{\#\overline{D}_{p-1}^2} \left[\sum_{g \in \overline{D}_{p-1}} \mathbf{E}_\mu [(\psi(f) \circ \theta_g - \Pi f(X_g))^2] \right. \\ &\quad \left. + \sum_{g \neq h} \mathbf{E}_\mu [(\psi(f) \circ \theta_g - \Pi f(X_g)) (\psi(f) \circ \theta_h - \Pi f(X_h))] \right] \end{aligned}$$

Or, pour $g \neq h$, $g \in \overline{D}_{p-1}$ et $h \in \overline{D}_{p-1}$

$$\begin{aligned} \mathbf{E}_\mu [(\psi(f) \circ \theta_g - \Pi f(X_g))(\psi(f) \circ \theta_h - \Pi f(X_h))] \\ = \mathbf{E}_\mu [\mathbf{E}_\mu [(\psi(f) \circ \theta_g - \Pi f(X_g))(\psi(f) \circ \theta_h - \Pi f(X_h)) / \mathcal{X}_{p-1}]] \\ = \mathbf{E}_\mu [\mathbf{E}_{X_g}(\psi(f) - \Pi f(X_0)) \mathbf{E}_{X_h}(\psi(f) - \Pi f(X_0))] = 0 \end{aligned}$$

et donc

$$\begin{aligned} \mathbf{E}_\mu [((\psi_p - \psi_{p-1}\Pi)(f))^2] &= \frac{1}{\#\overline{D}_{p-1}^2} \sum_{g \in \overline{D}_{p-1}} \mathbf{E}_\mu [(\psi(f) \circ \theta_g - \Pi f(X_g))^2] \\ &= \frac{1}{\#\overline{D}_{p-1}} \mathbf{E}_\mu [(\psi(f) - \Pi f(X_0))^2] = \frac{1}{\#\overline{D}_{p-1}} \left[\mathbf{E}_\mu \left((\psi(f))^2 \right) \right. \\ &\quad \left. - 2\mathbf{E}_\mu(\psi(f)\Pi f(X_0)) + (\Pi f(X_0))^2 \right] \\ &= \frac{1}{\#\overline{D}_{p-1}} \left[\mathbf{E}_\mu \left(\left(\frac{1}{\#\mathbf{A}} \sum_{\alpha \in \mathbf{A}} f(X_\alpha) \right)^2 \right) - \mathbf{E}_\mu(\Pi f(X_0)^2) \right] \\ &= \frac{1}{\#\overline{D}_{p-1}} \left[\frac{1}{\#\mathbf{A}^2} \left[\#\mathbf{A} \mathbf{E}_\mu(f(X_\alpha)^2) + \sum_{\alpha \neq \beta} \mathbf{E}_\mu(f(X_\alpha X_\beta)) \right] - \mathbf{E}_\mu(\Pi f(X_0)^2) \right] \\ &= \frac{1}{\#\overline{D}_{p-1}} \left[\left(\frac{1}{\#\mathbf{A}} \mu(f^2) + \frac{\#\mathbf{A}(\#\mathbf{A} - 1)}{\#\mathbf{A}^2} \mathbf{E}_\mu(\Pi f(X_0)^2) \right) - \mathbf{E}_\mu(\Pi f(X_0)^2) \right] \\ &= \frac{1}{\#\overline{D}_p} [\mu(f^2) - \mu((\Pi f)^2)] . \end{aligned}$$

■

Démonstration de la proposition 1.

On a

$$\begin{aligned} \psi_n(f) &= (\psi_n(f) - \psi_{n-1}\Pi(f)) + \psi_{n-1}\Pi(f) \\ &= \sum_{p=1}^{p=n} (\psi_p - \psi_{p-1}\Pi)(\Pi^{n-p}(f)) + \psi_0\Pi^n(f) \end{aligned}$$

avec $\psi_0\Pi^n(f) = \Pi^n(f)(X_0)$, et cette dernière somme est égale à, en posant

$$\Pi^{n-p}(f) = 0 \text{ pour } p > n, \sum_{p=1}^{\infty} (\psi_p - \psi_{p-1}\Pi)(\Pi^{n-p}(f)) + \psi_0\Pi^n(f).$$

D'après le théorème 1, $\lim_{n \rightarrow \infty} \Pi^{n-p}(f)(X_0) = \mu(f) \mathbf{P}_\mu$ -p.s. pour tout $p \in \mathbb{N}$.

Pour obtenir le résultat désiré, il suffit donc de démontrer que

$$\sum_{p=N}^{\infty} \sup_{q \geq p} |(\psi_p - \psi_{p-1}\Pi)(\Pi^{q-p}(f))| < +\infty .$$

En effet, par le théorème de convergence dominée, on aura alors

$$\begin{aligned} & \lim_{n \rightarrow \infty} \left(\sum_{p=1}^{\infty} (\psi_p - \psi_{p-1} \Pi)(\Pi^{n-p}(f)) + \psi_0 \Pi^n(f) \right) \\ &= \sum_{p=1}^{\infty} (\psi_p - \psi_{p-1} \Pi) \left(\lim_{n \rightarrow \infty} \Pi^{n-p}(f) \right) + \lim_{n \rightarrow \infty} \psi_0 \Pi^n(f) \\ &= \sum_{p=1}^{\infty} (\psi_p - \psi_{p-1} \Pi)(\mu(f)) + \mu(f) = \mu(f) \sum_{p=1}^{\infty} (\psi_p - \psi_{p-1} \Pi)(1) + \mu(f) = \mu(f) \end{aligned}$$

Les convergence ayant lieu \mathbf{P}_μ presque sûrement. Or

$$\begin{aligned} & \mathbf{E}_\mu \left(\sup_{q \geq p} |(\psi_p - \psi_{p-1} \Pi)(\Pi^{q-p}(f))| \right) \\ & \leq \left[\mathbf{E}_\mu \left(\left(\sup_{q \geq p} |(\psi_p - \psi_{p-1} \Pi)(\Pi^{q-p}(f))| \right)^2 \right) \right]^{1/2} \\ & \leq \left[\sum_{q=p}^{\infty} \mathbf{E}_\mu \left(|(\psi_p - \psi_{p-1} \Pi)(\Pi^{q-p}(f))|^2 \right) \right]^{1/2} \\ & = \left[\sum_{q=p}^{\infty} \frac{1}{\# \overline{D}_p} \left(\mu((\Pi^{q-p} f)^2) - \mu((\Pi^{q-p+1} f)^2) \right) \right]^{1/2} \leq \frac{1}{\# \mathbf{A}^{p/2}} \mu(f^2) \end{aligned}$$

d'où le résultat. \blacksquare

Démonstration du théorème 2

Soit donc U une variable de $L^2(\Omega, \mathcal{F}, \mathbf{P}_\mu)$. On a $\Psi^n(U) = \Psi^n(Z) + \psi_n(f)$ avec $Z = U - \mathbf{E}_{X_0}(U)$ et $f(x) = \mathbf{E}_x(U)$. On remarque que, pour tout $x \in E$, $\mathbf{E}_x(Z) = 0$. On a, d'après la proposition 1 $\lim_{n \rightarrow \infty} \psi_n(f) = \mu(f) = \mathbf{E}_\mu(U)$

\mathbf{P}_μ -p.s.. Or, $\mathbf{E}_\mu \left[\sum_{n=1}^{\infty} (\Psi^n(Z))^2 \right] < \infty$ ce qui implique que $\lim_{n \rightarrow \infty} \Psi^n(Z) = 0$

\mathbf{P}_μ -p.s. ce qui donne le résultat cherché. En effet

$$\begin{aligned} & \mathbf{E}_\mu \left[\left(\frac{1}{\# \overline{D}_n} \sum_{g \in \overline{D}_n} Z \theta_g \right)^2 \right] = \frac{1}{\# \overline{D}_n^2} \left[\sum_{g \in \overline{D}_n} \mathbf{E}_\mu [(Z \theta_g)^2] + \sum_{g \neq h} \mathbf{E}_\mu (Z \theta_g Z \theta_h) \right] \\ &= \frac{1}{\# \overline{D}_n} \mathbf{E}_\mu (Z^2) + \frac{1}{\# \overline{D}_n^2} \sum_{g \neq h} \mathbf{E}_\mu (\mathbf{E}_\mu (Z \theta_g Z \theta_h / \mathcal{X}_n)) \\ &= \frac{1}{\# A^n} \mathbf{E}_\mu (Z^2) + \frac{1}{\# \overline{D}_n^2} \sum_{g \neq h} \mathbf{E}_\mu (\mathbf{E}_{X_g}(Z) \mathbf{E}_{X_h}(Z)) = \frac{1}{\# A^n} \mathbf{E}_\mu (Z^2) \end{aligned}$$

D'où le résultat. ■ Je termine cette section en donnant au résultat de convergence presque sûre une allure un peu plus markovienne.

Corollaire 2 *Pour tout $U \in L^2(\Omega, \mathcal{F}, \mathbb{P}_\mu)$ et pour μ presque tout $x \in E$ les suites $(\Psi_n(U), n \in \mathbb{N})$ et $(\frac{1}{\#D_n} \sum_{h \in D_n} U\theta_h, n \in \mathbb{N})$ convergent vers $\mathbb{E}_\mu(U)$ \mathbb{P}_x -p.s..*

Ce résultat dépend du lemme suivant, en prenant $Z = \lim_{n \rightarrow \infty} \Psi^n(U)$. C'est un résultat markovien classique.

Lemme 3 *Soit Z une variable aléatoire de $L^1(\Omega, \mathcal{F}, \mathbb{P}_\mu)$ telle que, pour μ presque tout $x \in E$, $\Psi(Z) = Z$ \mathbb{P}_x -p.s.. Alors, pour μ presque tout $x \in E$, $Z = \mathbb{E}_\mu(Z)$ \mathbb{P}_x -p.s..*

Démonstration Soit $f(x) = \mathbb{E}_x(Z)$. On a, pour μ presque tout x ,

$$\begin{aligned} \Pi(f)(x) &= \mathbb{E}_x \left(\frac{1}{\#\mathbf{A}} \sum_{\alpha \in \mathbf{A}} f(X_\alpha) \right) = \mathbb{E}_x \left(\frac{1}{\#\mathbf{A}} \sum_{\alpha \in \mathbf{A}} \mathbb{E}_{X_\alpha}(Z) \right) \\ &= \mathbb{E}_x \left(\mathbb{E}_x \left(\frac{1}{\#\mathbf{A}} \sum_{\alpha \in \mathbf{A}} Z\theta_\alpha \right) / \mathcal{X}_1 \right) = \mathbb{E}_x(Z) = f(x). \end{aligned}$$

La fonction f est donc μ presque partout harmonique pour Π et elle est donc μ -presque partout constante. Mais, pour μ presque tout x ,

$$\psi_n(f) = \mathbb{E}_x \left(\frac{1}{\#D_n} \sum_{g \in D_n} Z\theta_g / \mathcal{X}_n \right) = \mathbb{E}_x(Z / \mathcal{X}_n)$$

converge \mathbb{P}_x -p.s. vers Z d'où le résultat. ■

3.2 Convergence dans les L^p

Théorème 3 *Pour tout $p \in [1, +\infty[$, pour tout $U \in L^p(\Omega, \mathcal{F}, \mathbb{P}_\mu)$ les suites $(\Psi_n(U), n \in \mathbb{N})$ et $(\frac{1}{\#D_n} \sum_{h \in D_n} U\theta_h, n \in \mathbb{N})$ convergent vers $\mathbb{E}_\mu(U)$ dans $L^p(\Omega, \mathcal{F}, \mathbb{P}_\mu)$.*

Démonstration C'est la même démonstration que celle du corollaire 1. Si U est bornée par une constante M , $\Psi^n(U)$ est bornée par la même constante et alors $\Psi^n(U)$ converge \mathbb{P}_μ p.s. et dans $L^p(\Omega, \mathcal{F}, \mathbb{P}_\mu)$ vers $\mathbb{E}_\mu(U)$. Si $U \in L^p$, comme

$$\Psi^n(U) = \Psi^n(U1_{\{U \leq M\}}) + \Psi^n(U1_{\{U > M\}})$$

alors

$$\begin{aligned} & \|\Psi^n(U) - \mathbf{E}_\mu(U)\|_p \\ & \leq \|\Psi^n(U1_{\{U \leq M\}}) - \mathbf{E}_\mu(U1_{\{U \leq M\}})\|_p + |\mathbf{E}_\mu(U1_{\{U > M\}})| + \|\Psi^n(U1_{\{U > M\}})\|_p \\ & \leq \|\Psi^n(U1_{\{U \leq M\}}) - \mathbf{E}_\mu(U1_{\{U \leq M\}})\|_p + 2\|U1_{\{U > M\}}\|_p \end{aligned}$$

(la dernière inégalité provenant de ce que $\|\Psi^n(U)\|_p \leq \|U\|_p$) ce qui prouve la convergence dans L^p . ■

4 Conclusion

Les résultats précédents se généralisent facilement à des noyaux Π possédant plusieurs classes ergodiques. Par ailleurs, dans un travail plus complet¹, ces résultats sont généralisés dans deux directions :

- À chaque étape, chaque point engendre un nombre aléatoire de points et pas seulement un nombre fixe. Les processus de Galton-Watson se trouvent alors introduits.
- Si on met sur chaque état visité une horloge exponentielle de paramètre indépendant de l'état telle que, quand cette horloge se déclenche, le point engendre $\#A$ points de façon markovienne comme précédemment, on peut alors regarder les points visités avant l'instant t . Ici ce sont des processus de Yule qui interviennent.

Dans ces deux cas, on généralise les théorèmes ergodiques précédents.

On peut surtout remarquer que la démonstration de convergence presque-sûre est issue de majorations assez grossières. On est donc tenté de penser qu'il doit exister des théorèmes ergodiques beaucoup plus généraux que ceux démontrés ici. En dernière analyse, il s'agit bel et bien de la convergence de $\Psi^n(U)$, Ψ étant un opérateur markovien.

Remerciements : Je remercie le référé pour ses judicieuses remarques sur le texte ainsi que pour le complément de bibliographie qu'il m'a procuré.

Références

- [1] Claire Anantharaman-Delaroche. On ergodic theorems for free group actions on non-commutative spaces. Preprint, 2004.
- [2] X. Bressaud and V. Maume-Deschamps. Ergodic theorems for group or semigroup actions. Preprint, 2002.

¹On peut l'obtenir en m'écrivant à Francois.Charlot@univ-rouen.fr

- [3] A. I. Bufetov. Operator ergodic theorems for action of free semigroups and groups. *Functional Analysis and Its Applications*, 34(4) : 239–251, 2000.
- [4] A. I. Bufetov. Markov averaging and ergodic theorems for several operators. *Amer. Math. Soc. Transl.*, 202(2) : 39–50, 2001.
- [5] A. I. Bufetov. Markov operators and the Nevo-Stein theorem. Preprint, 2001.
- [6] A. I. Bufetov. Convergence of spherical averages for action of free groups. *Ann. of Math. (2)*, 155(3) : 929–944, 2002.
- [7] D. Cellier, D. Brunhes, F. Charlot, S. Mercier, and M. Rezzouk. Matrice de Dayhoff, matrice Pam, pour la comparaison de séquences protéiques. Preprint, 2000.
- [8] J. P. Conze and N. Dang Ngoc. Ergodic theorems for non-commutative dynamical systems. *Invent. Math.*, 46 : 1–15, 1978.
- [9] M. O. Dayhoff, R. V. Eck, and C. M. Park. A model of evolutionary change in protein. *Atlas of proteins sequences and structure*, 5 : 89–99, 1972.
- [10] Shaul R. Foguel. *The ergodic theory of Markov processes*, volume 21 of *Van Nostrand Mathematical Studies*. Van Nostrand, 2nd edition, 1969.
- [11] R. I. Grigorchuk. Ergodic theorems for action of free groups and free semigroups. *Mathematical Notes*, 65(5) : 654–657, 1999.
- [12] Y. Guivarc’h. Généralisation d’un théorème de Von Neumann. *C. R. Acad. Sci. Paris Sér. I Math.*, 268 : 1020–1023, 1969.
- [13] U. Krengel. *Ergodic theorems*. Number 6 in De Gruyter Studies in Mathematics. Walter de Gruyter and Co., 1985.
- [14] S. P. Meyn and R. L. Tweedie. *Markov Chains and Stochastic Stability*. Communications and Control Engineering Series. Springer Verlag, 2nd edition, 1994.
- [15] A. Nevo and E. Stein. A generalisation of Birkhoff’s pointwise ergodic theorem. *Acta Math.*, 173 : 135–154, 1994.
- [16] A. Revuz. *Markov Chains*, volume 11 of *Math. library*. North Holland, 2nd edition, 1984.
- [17] G. V. Rota. An “alternierende verfahren” for general positive operators. *Bulletin AMS*, 68(5) : 95–102, 1962.
- [18] A. Templeman. *Ergodic theorems for group actions*. Number 78 in Informational and thermodynamical aspect, Mathematics and its Applications. Kluwer Academic Publishers Group, Dordrecht, 1992.

- [19] T. E. Walker. Ergodic theorems for free group actions on Von Neuman algebra. *J. Funct. Anal.*, 150 : 27-47, 1997.

Conditional law of excursions Known to start at x end at y without duality hypothesis

Hacène Boutabia

E-mail: hboutabia@hotmail.com

Abstract

Let B be a continuous additive functional for a standard process $(X_t)_{t \in \mathbb{R}_+}$. In this article we give the excursion laws of $(X_t)_{t \in \mathbb{R}_+}$ conditioned on the strict past and future without duality hypothesis. We study excursions of a general regenerative system and of a regenerative system consisting of the closure of the set of times that the regular points of B are visited. We use the (F_{D_t}) -predictable exit system to bring together the isolated points of M and its perfect part and replace the classical optional exit system. This has been a subject in literature before (e.g. Kaspi [4]) under the classical duality hypothesis.

MSC: 60J25; 60J40; 60J55

Keywords: Standard process; Predictable process; Excursion; Additive functional; Conditional law; Exit measure; Kuznetsov process.

1 Introduction

Let X be a standard process, and let M be a closed random and homogeneous subset of \mathbb{R}_+ . Kaspi [4] construct an additive functional B associated to M and gives, under the classical duality hypothesis, the probability measures allowing the law of excursions associated to B with respect to the σ -algebra $K = \sigma(Z_t : t \in \mathbb{R}_+)$, known to start at x end at y ($Z_t = X_{S_t}$ where $S_t = \inf \{u : B_u > t\}$). The purpose of this paper is to give, without duality, the conditional law $P^{x,y}$ of the excursion straddling an arbitrary random time given the start state x and the finished state y , as a regular probabilities in terms of the (F_{D_t}) -predictable exit measures for M . Excursions associated to B are considered.

In section 2 we introduce our notations and preliminaries Maisonneuve's result, [7] on the strict past conditioning with respect to the filtration (F_{D_t}) . In section 3 we construct the probability measures $P^{x,y}$ which allows the law of the excursion straddling an arbitrary random time, given the initial state x and the final state y . Section 4 deals with excursions associated to a continuous additive functional B . The measures $P^{x,y}$ which governs these excursions are the same as defined in section 3 corresponding to the regenerative set M where contiguous intervals are of the form $]S_{t-}, S_t[$, t is a time of discontinuity of S .

2 Notations and preliminaries

Let $(\Omega, F, F_t, X_t, \theta_t, P^x)$ be a canonical realization for a borel standard semi-group (P_t) with life time ζ , and let M be a closed random and homogeneous subset of $]0, \zeta[$ such that $R = \inf M$ is F^* -measurable; where F^* is the universal completion of the σ -algebra $F^0 = \sigma(X_t : t \in \mathbb{R}_+)$. We assume that the state space E is lusinian, and we note by \mathcal{E} its σ -algebra of borel sets. The cimetry point δ is outside of E .

Let $({}^\circ P^x)_{x \in E \cup \delta}$ be the family of (F_t^D) -predictable exit measures for the process $(X_t^D) = (X_{D_t})$ in the sense of Maisonneuve [6], and let μ be a fixed law on E . Then $({}^\circ P^x)_{x \in E \cup \delta}$ is an universally measurable family of σ -finite measures on (Ω, F^*) , under which the process (X_t) is Markov with respect to (P_t) .

For all $t \in \mathbb{R}_+$, let k_t be the killing operator at t defined by: $k_t(\omega)(s) = \omega(s)$ if $s < t$ and δ if $s \geq t$. If $t \in \mathbb{R}_+$, $\omega, \omega' \in \Omega$, we note by $\omega/t/\omega'$ the trajectory $\bar{\omega} \in \Omega$ such that $\bar{\omega}(s) = \omega(s)$ (or $k_s(\omega)$) if $s < t$ and $\omega'(s-t)$ if $s \geq t$.

Let T be a random time on (Ω, F) such that $T < \inf \{s \geq T : s \in M\}$ on $\{T < \zeta\}$, and let $g = \sup \{s \leq T : s \in M\}$ and $d = \inf \{s > T : s \in M\}$ ($\inf \emptyset = +\infty$). Then with the following notations:

$$A^\omega(\omega') = T(\omega) - g(\omega/g(\omega)/\omega'), \quad C^\omega = \{0 \leq A^\omega < R\}$$

and

$$v(B | A) = \frac{v(B \cap A)}{v(A)} \quad \left(\frac{0}{0} = \frac{\infty}{\infty} = 0 \right)$$

if v is a measure on (Ω, F^*) , we have the basic Maisonneuve's formula [7]:
For almost all $\omega \in \{g < \infty\}$ we have (if δ is non absorbent):

$$P(f(\theta_g) | F_{g-}^D)(\omega) = {}^\circ P^{X_{g-}^D(\omega)}(f | C^\omega) \quad (1)$$

for all function $f \geq 0$ and F^* -measurable, where P is the probability measure defined by $P(f) = \int P^x(f) \mu(dx)$. If we assume that δ is absorbent, then this formula is true on $\{g < \zeta\}$ instead of $\{g < \infty\}$.

Note that if T is an $(F_t^D) = (F_{D_t})$ -stopping time we can replace C^ω by the condition $A^\omega < R$, and if $T \in G^0$ on $\{Y_T \in E\}$ the set C^ω can be replaced by the condition $A^\omega = 0$.

3 The excursion straddling T

For the conditional law of the excursion $e = k_R \circ \theta_g$ straddling T , with respect to $F_{g^-}^D$ and θ_d , we assume that δ is absorbent. In this respect, we consider for $(x, y) \in E \times E$, the measures H^x , H_1^x and $P^{x,y}$ on (Ω, F^*) "defined by":

$$\begin{aligned} H^x &= {}^\circ P^x(k_R \in \cdot \mid X_R = x) \quad ; \quad H_1^x = {}^\circ P^x(\cdot \mid X_R \neq x) \\ \text{and} \quad P^{x,y} &= H_1^x(k_R \in \cdot \mid X_R = y) \quad \text{if } x \neq y. \end{aligned}$$

Since (Ω, F^0) is an U-space, and according to a classical lemma of Doob the measures $P^{x,y}$ can be chosen measurable for the pair (x, y) .

Proposition 1: *Let $x \in E$, $A \in F^*$ such that $0 < {}^\circ P^x(A) < +\infty$, and let the probability measure μ^x defined on (Ω, F^*) by $\mu^x = {}^\circ P^x(\cdot \mid A)$. Then we have:
for almost all $\omega \in \{X_R \neq x\}$*

$$\mu^x(f \circ k_R \mid \theta_R)(\omega) = P^{x, X_R(\omega)}(f \mid (\cdot/\zeta/\theta_R(\omega)) \in A) \quad (2)$$

and for almost all $\omega \in \{X_R = x\}$

$$\mu^x(f \circ k_R \mid \theta_R)(\omega) = H^x(f \mid (\cdot/\zeta/\theta_R(\omega)) \in A) \quad (3)$$

Proof : Follows from the Markov property at time R with an argument of monotone classes, the definitions of $P^{x,y}$ and H^x , and the fact that $\omega = (k_R(\omega)/\zeta \circ k_R(\omega)/\theta_R(\omega))$ for all $\omega \in \Omega$. ■

The following theorem gives the conditional law of the excursion e with respect to $F_{g^-}^D$ and θ_d .

Theorem 1: For all $\omega \in \{g < \infty\}$, let the subset of Ω defined by:

$U_d^\omega = \{\omega' \in \Omega : (\omega'/\zeta(\omega')/\theta_d(\omega')) \in C^\omega\}$. Then we have:

1) For almost all $\omega \in \{X_d \neq X_{g^-}^D; g < \zeta; d - g < \infty\}$

$$P(f(e) \mid F_{g^-}^D, \theta_d)(\omega) = P^{X_{g^-}^D(\omega), X_d(\omega)}(f \mid U_d^\omega) \quad (4)$$

2) For almost all $\omega \in \left\{ X_d = X_{g^-}^D, g < \zeta \right\}$

$$P(f(e) | F_{g^-}^D, \theta_d)(\omega) = H^{X_{g^-}^D(\omega)}(f | U_d^\omega) \quad (5)$$

It follows that if T is an (F_{D_t}) -stopping time such that $T \in G^0$ on $\{Y_T \in E\}$, then the formulas (4) and (5) hold without conditioning by U_d^ω in the right side.

Proof: 1) Follows from formula (1) and formula (2) with $x = X_{g^-}^D(\omega)$ and $A = C^\omega$.

2) The formula (5) is argued in the same manner using the formula (3). ■

Remark 1: Maisonneuve [7] gives several examples where the set C^ω is independent of ω . In these cases theorem 1 implies that the excursion e is conditionally independent of $F_{g^-}^D$ and θ_d given $X_{g^-}^D$ (resp. $X_{g^-}^D$ and X_d) on $\left\{ X_{g^-}^D = X_d; g < \zeta \right\}$ (resp. $\left\{ X_{g^-}^D \neq X_d; g < \zeta; d - g < \infty \right\}$).

Remark 2: Theorem 1 contains results of section 5 of Kaspi [4] under duality hypothesis. In fact if M is perfect, then $X_{g^-}^D = X_{g^-}$ and $F_{g^-}^D = F_{g^-}$. If T is the beginning of the set $\{t \in \mathbb{R}_+ : (X_{t^-}, X_t) \in J\}$ where $J \in \mathcal{E} \otimes \mathcal{E}$, then with the assumption that ${}^\circ P^x(X_{0^-} \neq x) = 0$, the conditions $0 < g(\omega) < T(\omega)$ and $\theta_g(\omega) \in C^\omega$ are equivalent to the following condition: $\omega \in \{(X_{0^-}, X_0) \notin J; 0 < T < R\}$, and the formula (1) becomes:

$$P(f(\theta_g) | F_{g^-}) = {}^\circ P^{X_{g^-}}(f | (X_{0^-}, X_0) \notin J; 0 < T < R) \\ \text{on } \{0 < g < T < \zeta\}$$

According to the same argument used in theorem 1 and the fact that $T = T \circ k_R$ on $\{T < R\}$ and $R = \zeta \circ k_R$, the formula (4) becomes:

$$P(f(e) | F_{g^-}, \theta_d) = P^{X_{g^-}, X_d}(f | (X_{0^-}, X_0) \notin J; 0 < T < \zeta) \\ \text{on } \{X_{g^-} \neq X_d; 0 < g < T < \zeta\}$$

and the formula (5) becomes:

$$P(f(e) | F_{g^-}, \theta_d) = H^{X_{g^-}}(f | (X_{0^-}, X_0) \notin J; 0 < T < \zeta) \\ \text{on } \{X_{g^-} = X_d; 0 < g < T < \zeta; d - g < \infty\}$$

4 Excursions associated to an additive functional

Let (B_t) be a continuous additive functional and let $C = \{x : P^x(R = 0) = 1\}$ its support, where R is the perfect exact terminal time $\inf\{u : B_u > 0\}$. We

associate to the right inverse $S_t = \{u : B_u > t\}$ of (B_t) , the following notations: $Z_t = X_{S_t}$, $\mathcal{M}_t = F_{S_t}$ and $\bar{\theta}_t = \theta_{S_t}$. It is well known that the process $Z = (\Omega, F, \mathcal{M}_t, Z_t, \bar{\theta}_t, P^x)$ is strong Markov with semigroups $(\bar{P}_t) \triangleq (P_{S_t})$ and takes values on $(C, C \cap \mathcal{E}^*)$ (cf. Jacods [3]).

In this section we assume that δ is non absorbent and we consider the random homogeneous set $M = \{t + R \circ \theta_t : t \in \mathbb{R}_+\}$, and its family of (F_{D_t}) -predictable exit measures $({}^0P^x)_{x \in E \cup \{\delta\}}$. If $S_{t-} \neq S_t$, then $D_{S_{t-}} = S_t$. The excursion associated to t is then defined by:

$$e_t = k_R \circ \theta_{S_{t-}} = \begin{cases} X_{S_{t-}+s} & \text{if } s < S_t - S_{t-} \\ \delta & \text{if } s \geq S_t - S_{t-} \end{cases}$$

We note by $(K_t)_{t \in \mathbb{R}_+}$ the filtration, where K_t is the intersection of the P^π -completions of the σ -algebra K_{t+}^0 where π is in the set of all the bounded measures on E ; (K_t^0) is the natural filtration of the process (Z_t) .

For the following lemma we put $K_{0-} = F_0$ by convention.

Lemma 1: *Let T be a (K_t) -stopping time such that $S_{T-} < S_T$ a.s.. Then we have:*

$$F_{(S_{T-})-} = K_{T-}$$

Proof : According to the fact that S_{T-} is not an isolate point of M , we have $F_{(S_{T-})-} = F_{(S_{T-})-}^D$. Since $K_t^0 = \sigma(X_{S_{u-}}^D : u \leq t) \subset F_{(S_{t-})}^D$, then every (K_t) -predictable process is $F_{(S_{t-})}^D$ -predictable, which implies that $K_{T-} \subset F_{(S_{T-})-}$. For the inclusion $F_{(S_{T-})-} \subset K_{T-}$, since $F_0 \subset K_{T-}$ it suffices to prove that: $A = B \cap \{t < S_{T-}\} \in K_{T-}$ for all $B \in F_t$.

Note that $S_{T-} = \sup_{r < T} \{S_r : r \in Q_+\}$; which implies that:

$$\{t < S_{T-}\} = \bigcup_{r \in Q_+} \{r < T\} \cap \{t < S_r\} \text{ and } A = \bigcup_{r \in Q_+} (B \cap \{t < S_r\} \cap \{r < T\})$$

where Q_+ is the set of positive rationals. For all $u \leq t$ we have $X_u = Z_{B_u}$, then $F_t \subset K_{B_t}$ which implies that

$$B \cap \{t < S_r\} = B \cap \{B_t \leq r\} \in K_r$$

and $A \in K_{T-}$. The proof is complete. ■

Remark 3 : For all $t \in \mathbb{R}_+$ the σ -algebras $F_{(S_{t-})-}$ and K_{t-} have the same trace on $\{S_{t-} < S_t\}$.

The following theorem which gives the conditional law of the excursion e_T associated to a (K_t) -stopping time T , with respect to the σ -algebra K generated by K_t ($t \geq 0$) was proved by Kaspi [4] under the duality hypothesis.

Theorem 2 : Let T be a (K_t) –stopping time such that $S_{T^-} \neq S_T$ and $Z_{T^-} \neq Z_T$ on $\{T < \infty\}$. Then we have the following formula:

$$P(f(e_T) | K) = P^{Z_{T^-}, Z_T}(f) \quad \text{on} \quad \{T < \infty\} \quad (6)$$

for all positive and F^* –measurable function f .

Proof : It suffices to prove that:

$$P(f(e_T) U_T \varphi(\bar{\theta}_T) I_{\{T < \infty\}}) = P(P^{Z_{T^-}, Z_T}(f) U_T \varphi(\bar{\theta}_T) I_{\{T < \infty\}}) \quad (7)$$

for all continuous process $U \geq 0$, (K_t) –adapted and for all positive function φ , F^* – measurable. We have $Z_{t^-} = X_{(s_{t^-})^-}^D = X_{(s_{t^-})^-}$ on $\{S_{t^-} < S_t\}$ and $K_{t^-} \cap \{S_{t^-} < S_t\} = F_{(s_{t^-})^-} \cap \{S_{t^-} < S_t\}$, for all $t \in \mathbb{R}_+$. Note that the formula (4) with the stopping time S_{t^-} instead of T and the fact that $S_{t^-} \in G^0$ on $\{S_{t^-} < S_t\}$ implies that:

$$P(f(e_t) | K_{t^-}, \bar{\theta}_t) = P^{Z_{t^-}, Z_t}(f) \quad \text{on} \quad \{S_{t^-} < S_t; Z_{t^-} \neq Z_t\}$$

and by the Markov property for the process Z we obtain that:

$$\begin{aligned} & P\left(f(e_{T_n}) U_{T_n} \varphi(\bar{\theta}_{T_n}) I_{\{S_{T_n^-} \neq S_{T_n}; Z_{T_n^-} \neq Z_{T_n}; T < \infty\}}\right) \\ &= P\left(P^{Z_{T_n^-}, Z_{T_n}}(f) U_{T_n} \varphi(\bar{\theta}_{T_n}) I_{\{S_{T_n^-} \neq S_{T_n}; Z_{T_n^-} \neq Z_{T_n}; T < \infty\}}\right) \end{aligned}$$

where $(T_n)_{n \in \mathbb{N}}$ is the decreasing dyadic approximation of T . The formula (8) follows from the right continuity for the processes (Z_t) , $(\bar{\theta}_t)$, (e_t) , (U_t) and the approximation of the events $\{T < \infty\}$ by $\{S_{T_n^-} \neq S_{T_n}; Z_{T_n^-} \neq Z_{T_n}; T < \infty\}$. ■

REFERENCES

- [1] BOUTABIA H. et MAISONNEUVE B.-Lois conditionnelles des excursions markoviennes, *Sém. Prob.*XXVI 1526, 162-166, 1992.
- [3] DELLACHERIE C., MAISONNEUVE B. et MEYER PA.-*Probabilités et potentiel*, chapitres XVII à XXIV, Hermann, 1992.
- [4] JACOBS.PA-Excursions of a Markov process induced by continuous additive functionals, *Z. Wahrs. Verw. Geb.* 44(1978),325-336.
- [5] KASPI H.-Excursion laws of Markov processes in classical duality, *Ann.Prob.*, 33, (1985), 492-518.
- [6] MAISONNEUVE B.-Système de sortie (F_{D_t}) -prévisibles, *Prob.Th.Fiel* 80 (1989), 395-405.
- [7] MAISONNEUVE B.-Excursions chevauchant un temps aléatoire quelconque, *Astérisque* N°236 , *S.M.F.* (1996).

INEGALITES POUR LES SYSTEMES DE FILES D'ATTENTE AVEC RAPPELS

BOUALEM Mohamed et AISSANI Amar

Oulkhou- Ait Chaffa- Azeffoun

Tizi-Ouzou, 15320, Algérie.

Robertt15dz@yahoo.fr

aaissani@wissal.dz

Section Mss'2004 : Preprocessus Aléatoire

I. Introduction :

Les modèles d'attente avec rappels occupent une situation intermédiaire entre les modèles d'Erlang avec refus et les modèles classiques avec attente *FIFO* (premier arrivé premier servi), qui en constituent les modèles limites dans les cas de faibles et de fortes intensités de rappels.

Les études analytiques de la majorité des modèles de files d'attente se sont avérées très complexes. Les efforts fournis pour simplifier les modèles afin d'obtenir des équations les régissant n'ont pas suffi pour avoir la solution générale, puisque même si on est souvent arrivé à définir les équations du modèle, leur résolution n'a pas toujours été évidente. En effet, dans la majorité des cas, on se retrouve confronté à des systèmes d'équations dont la résolution est complexe, ou possédant des solutions qui ne sont pas facilement interprétables pour que le praticien puisse en bénéficier. Dans bien des cas l'existence et l'unicité de la solution sont difficiles à établir.

Pour pallier aux difficultés rencontrées dans l'obtention de solutions exactes et interprétables pour de nombreux systèmes d'attente, les analystes ont recouru à des méthodes d'approximation.

Ces dernières décennies, Il est apparu dans la littérature des files d'attente des méthodes d'approximation ayant permis d'avoir des estimations quantitatives et qualitatives des caractéristiques des modèles étudiés. Les méthodes de comparaisons stochastiques qui conduisent à des inégalités stochastiques en terme de majoration ou de minoration sont développées dans le but d'avoir des estimations qualitatives.

Dans ce travail, nous étudions l'application de ces méthodes à une classe de modèles de files d'attente avec rappels constants et vacances.

II. ORDRES PARTIELS ET MONOTONIE

1. Ordres partiels :

X et Y deux v.a de fonctions de répartitions F et G respectivement.

-Ordre stochastique :

$$X \leq_{st} Y \text{ lorsque } F(x) \geq G(x), \quad \forall x \in \mathbb{R}.$$

-Ordre convexe :

$$x_+ = \max(0, x).$$

X est inférieure en moyenne de vie résiduelle à Y , on note $X \leq_v Y$, ssi

$$E(X - x)_+ = \int_x^{+\infty} (t - x) dF(t) = \int_x^{+\infty} (1 - F(t)) dt$$

$$\leq \int_x^{+\infty} (1 - G(t)) dt = E(Y - x)_+$$

-Ordre en transformée de Laplace :

$X \leq_L Y$, si pour tout s positif on a :

$$E[\exp\{-sX\}] = \int_0^{+\infty} \exp(-sx) dF(x) \geq \int_0^{+\infty} \exp(-sx) dG(x) = E[\exp\{-sY\}]$$

III. M/G/1 AVEC RAPPELS ET VACANCES

1. Description du modèle

Le taux des arrivées primaires est λ .

Un client qui arrive et trouve le serveur occupé, quitte l'aire du service pour rejoindre un groupe de clients bloqués appelé "orbite".

Après un certain temps aléatoire, il renouvelle sa tentative d'entrer en service, une fois, deux fois,....., jusqu'à ce qu'il le trouve disponible.

Les temps inter rappels suivent une distribution exponentielle de taux μ .

Les temps de service sont supposés d'une loi arbitraire, $B(t)$ ($B(0)=0$),

De plus on suppose que le serveur prend une vacance chaque fois que le système devient vide (service exhaustif).

On veut dire par vacance exhaustive que le serveur ne peut prendre une vacance, seulement lorsqu'il n'y a plus de clients dans le système.

Le flux des arrivées primaires, les intervalles entre les rappels successifs et les temps de service sont mutuellement indépendants.

L'état du système à l'instant t peut être décrit par le processus

$$X(t) = (C(t), N(t), \xi(t))_{t \geq 0}, \text{ où}$$

$$C(t) = \begin{cases} 0 & \text{si le serveur est oisif,} \\ 1 & \text{si le serveur est occupé,} \\ 2 & \text{si le serveur est en vacance.} \end{cases}$$

$N(t)$: le nombre de clients dans l'orbite à l'instant t .

$\xi(t)$: est le temps de service écoulé du client en service si $C(t) = 1$ et est le temps de la vacance écoulé si $C(t) = 2$.

2. Chaîne de Markov incluse :

Soit $\{t_n, n \in \mathbb{N}\}$ une suite d'instants de la complétion d'un service ou bien de la fin d'une vacance propre.

La séquence des vecteurs aléatoires $Y_n = (C(t_n^-), N(t_n^+))$ forme une chaîne de Markov incluse pour $X(t)$.

L'espace d'état est $S = \{1, 2\} \times \mathbb{N}$

Les états de transitions sont donnés par :

$$(i_{n+1}, j_{n+1}) = \begin{cases} (2, X) & \text{si } j_n = 0, \\ (1, j_n - B_n + U) & \text{si } j_n \geq 1. \end{cases}$$

Où X : le nombre de clients qui arrivent durant une vacance.

U : le nombre de clients qui arrivent pendant un temps de service qui se termine à l'instant t_{n+1} .

Et

$$B_n = \begin{cases} 1 & \text{si le client qui quitte le système à } t_{n+1} \text{ provient de l'orbite,} \\ 0 & \text{si le client qui quitte le système à } t_{n+1} \text{ provient de l'extérieur.} \end{cases}$$

.1. La chaîne de Markov incluse est ergodique si et seulement :

$$\lambda \beta_1 (\lambda + \mu)^{-1} < 1.$$

IV. INEGALITES STOCHASTIQUES POUR LE MODELE M/G/1 AVEC RAPPELS ET VACANCES DU SERVEUR

Σ_1 et Σ_2 deux modèles M/G/1 avec rappels et vacances

$V^{(i)}(x)$: distribution du temps de vacance dans Σ_i .

$f_n^{(i)}$: la probabilité du nombre de clients arrivant durant une période de vacance dans Σ_i .

$v^{(i)}$: taux de vacances dans Σ_i .

1. Inégalités préliminaires

On compare les probabilités $\{f_n^{(i)}, i = 1, 2 \text{ et } n \in \mathbb{N}\}$, suivant les ordres partiels ; stochastique, convexe et en transformée de Laplace.

Lemme :

Soient Σ_1 et Σ_2 deux systèmes d'attente M/G/1 avec rappels et vacances, si $\lambda^{(1)} \leq \lambda^{(2)}, V^{(1)} \leq_s V^{(2)}$ alors

$$\left\{ f_n^{(1)} \right\} \leq_s \left\{ f_n^{(2)} \right\} .$$

(s = st , v ou L)

Où ;

$$f_n^{(i)} = P[X = n] = \int_0^{+\infty} \frac{(\lambda^{(i)} x)^n}{n!} \exp\{-\lambda^{(i)} x\} dV^{(i)}(x) .$$

2. Monotonie de la chaîne de Markov incluse

Les probabilités de transitions en un pas de la chaîne de Markov incluse sont

$$P_{nm} = \begin{cases} f_m & \text{si } n = 0, \\ \frac{\mu}{\lambda + \mu} K_{m-n+1} & \text{si } m = n-1 \text{ et } n \geq 1, \\ \frac{\lambda}{\lambda + \mu} K_{m-n} + \frac{\mu}{\lambda + \mu} K_{m-n+1} & \text{si } m \geq n \geq 1, \\ 0 & \text{sinon.} \end{cases}$$

Soit l'opérateur de transition T de la chaîne de Markov incluse. Pour chaque distribution $p = (p_n)_{n \geq 0}$, on associe une distribution $Tp = q = (q_m)_{m \geq 0}$ tel que $q_m = \sum_{n \geq 0} p_n P_{nm}$.

Les deux théorèmes suivants donnent la condition sous laquelle l'opérateur T est monotone par rapport aux ordres stochastique (\leq_{st}) et convexe (\leq_v).

Théorème.1.

Si l'inégalité $V \leq_{st} B$ a lieu, alors l'opérateur de transition T est monotone, par rapport à l'ordre stochastique (\leq_{st}).

C'est-à-dire ;

Pour deux distributions quelconques $p^{(1)}$ et $p^{(2)}$, l'inégalité $p^{(1)} \leq_{st} p^{(2)}$ implique la suivante :

$$Tp^{(1)} \leq_{st} Tp^{(2)} .$$

Théorème.2.

Si l'inégalité $V \geq_v B$ a lieu, alors l'opérateur T est monotone par rapport à l'ordre convexe " \leq_v ". C'est-à-dire ;

Pour deux distributions quelconque $p^{(1)}, p^{(2)}$ l'inégalité $p^{(1)} \leq_v p^{(2)}$ implique la suivante :

$$Tp^{(1)} \leq_v Tp^{(2)}.$$

Σ_1 et Σ_2 deux M/G/1 avec rappels et vacances

$$\lambda^{(1)}, \mu^{(1)}, B^{(1)}(x), V^{(1)}(x) \text{ et } \lambda^{(2)}, \mu^{(2)}, B^{(2)}(x), V^{(2)}(x).$$

$T^{(1)}, T^{(2)}$ les opérateurs de transition associés aux chaînes de Markov incluses

Les deux théorèmes suivants donnent les conditions de comparabilité de ces opérateurs par rapport aux ordres partiels ; stochastique et convexe.

Théorème.3.

Soient Σ_1 et Σ_2 deux systèmes d'attente M/G/1 avec rappels et vacances, si

$$\lambda^{(1)} \leq \lambda^{(2)}, \mu^{(1)} \geq \mu^{(2)}, B^{(1)} \leq_{st} B^{(2)}, V^{(1)} \leq_{st} V^{(2)}$$

alors $T^{(1)} \leq_{st} T^{(2)}$,

c'est-à-dire que pour une distribution quelconque p on a

$$T^{(1)}p \leq_{st} T^{(2)}p.$$

Théorème.4.

Si $\lambda^{(1)} \leq \lambda^{(2)}, \mu^{(1)} \geq \mu^{(2)}, B^{(1)}(x) \leq_v B^{(2)}(x)$ et $V^{(1)}(x) \leq_v V^{(2)}(x)$

alors

$$T^{(1)} \leq T^{(2)}$$

C'est-à-dire que pour une distribution quelconque p on a $T^{(1)}p \leq_v T^{(2)}p$

Les deux théorèmes suivants donnent les conditions de comparabilité des distributions stationnaires du nombre de clients par rapport aux ordres partiels ; stochastique et convexe.

Théorème.5.

On considère deux systèmes de files d'attente avec rappels et vacances ayant les paramètres $\lambda^{(i)}, \mu^{(i)}, B^{(i)}(x)$ et $V^{(i)}(x)$, $i=1,2$ respectivement, et soient $\pi_n^{(1)}, \pi_n^{(2)}$ les distributions stationnaires du nombre de clients dans chaque système, alors si les inégalités suivantes ont lieu

$$\lambda^{(1)} \leq \lambda^{(2)}, \mu^{(1)} \geq \mu^{(2)}, B^{(1)} \leq_s B^{(2)},$$

$$V^{(1)} \leq_s V^{(2)} \text{ et } V^{(2)} \leq_{st} B^{(2)} \text{ (resp. } B^{(2)} \leq_v V^{(2)})$$

alors on a aussi l'inégalité suivante sur les distributions stationnaires:

$$\left\{ \pi_n^{(1)} \right\} \leq_s \left\{ \pi_n^{(2)} \right\}, \quad \text{où } s = st \text{ (ou } v).$$

Théorème.6.

Si pour $M/G/1$ avec rappels et vacances, la distribution de temps de service est NBUE (ou bien NWUE), et si de plus $V^{(1)} \leq_v V^{(2)}$, $B^ \leq_v V^{(2)}$, alors la distribution stationnaire du nombre de clients dans le système est inférieure (respectivement supérieure), par rapport à l'ordre convexe, à la distribution stationnaire de nombre de clients dans le système $M/M/1$ avec rappels et vacances*

V. CONCLUSION

On a établi des conditions, sur les paramètres de $M/G/1$ avec rappels constants et vacances du serveur, qui assurent la monotonie de l'opérateur de transition associé à la chaîne de Markov incluse. Aussi, on a mis en évidence les conditions pour lesquelles les opérateurs de transition ainsi que les distributions stationnaires du nombre de clients dans le système, de deux chaînes de Markov incluses associées à deux systèmes $M/G/1$ avec rappels constants et vacances ayant la même structure, mais avec des paramètres différents, sont comparables au sens des ordres stochastique et convexe.

Enfin, il serait possible de simplifier les démonstrations des théorèmes et lemmes énoncés dans la thèse en utilisant la propriété de la décomposition stochastique prouvée récemment pour ce type de modèle .

Evidemment, on peut étendre ce travail en estimant d'autres mesures de performance, telles que le temps moyen d'attente, le nombre moyen de rappels par client, etc., ou bien, éventuellement, en utilisant d'autres ordres partiels (concave, en transformée de Laplace,...).

Some inequalities of the Uniform Ergodicity and Strong Stability of Waiting Process

Zahir MOUHOUBI and Djamil AÏSSANI¹

L.A.M.O.S

Laboratory of Modelization and Optimization of Systems

Faculty of Sciences and Engineer Sciences, University of Béjaïa, 06000, Algeria

Abstract

In this paper we are interesting about the uniforme ergodicity and strong stability for waiting process. Moreover, we obtain the estimate of the potential of the chain. This process can be represent the $G/G/1$ queuing model.

Keywords: Quantitative estimates, Uniform ergodicity, Stability, Strong Stability, Perturbation, potential of the chain.

1 Introduction

The aim of this paper is to obtain the sufficient conditions for strong stability and uniform ergodicity for the random walk process described by a chain X defined by the recursive relation $X_{n+1} = (X_n + \xi_{n+1})^+$, $n \geq 0$ which taking values in $E = \mathbb{R}^+$, where $(\xi_n)_{n \geq 0}$ is a sequence of independent random variables identically distributed with a common distribution function F . Moreover, we obtain some inequalities of uniform ergodicity and strong stability of this waiting process. In contrast of the inequalities obtained in [4], we suppose that $d \geq 1$ and not only $d = 1$. we note that This example was studied by N.V. Kartashov in [3] and D. Aïssani in [1]. Moreover, all notations used in this paper are introduced in [6] and [5].

For this, setting $h(x) = P(\xi_1 + x \leq 0)$ and $\alpha(dy) = \delta_0(dy)$, where δ_0 is Dirac distribution centred on origine. We consider the distribution

$$P(x, A) = P((x + \xi_1)^+ \in A) = P(X_1 \in A / X_0 = x)$$

It is easy to observe that

$$P(x, A) = P(0 < x + \xi_1 \in A) + h(x).\alpha(A) \quad (1)$$

¹To whom correspondence should be addressed (e-mail : lamos.bejaia@hotmail.com).

Let us put $T(x, A) = P(0 < x + \xi_1 \in A)$ and we consider a distribution $p_n = \alpha T^{n-1}h$, $n \geq 1$. Let us consider $d = GCD\{n \geq 0 : p_n > 0\} \geq 1$ in the rest of this paper. We define the test function v defined as below

$$v = \exp(\gamma x), x \in \mathbb{R}^+ \quad (2)$$

Furthermore, we provide the space $m\mathcal{E}$ of finite measures on the countably generated σ -algebra \mathcal{E} with some norm $\|\cdot\| = \|\cdot\|_v = \|\cdot\|_\gamma$ which have the following form

$$\|\mu\|_\gamma = \|\mu\|_v = \int_0^{+\infty} \exp(\gamma x) |\mu|(dx)$$

The correspondant norms in the space \mathcal{J} of measurable bounded function on E and the space \mathcal{B} of bounded operators have the forme

$$\|f\|_\gamma = \sup_{x \geq 0} \exp(-\gamma x) |f(x)| \quad \forall f \in \mathcal{J}$$

and

$$\|Q\|_\gamma = \sup_{x \geq 0} \exp(-\gamma x) \int_0^{+\infty} |Q|(x, dy) \exp(\gamma y) \quad \forall Q \in \mathcal{B}$$

Let us introduce the stationary projector Π of transition kernel P defined by

$$\Pi = \mathbf{1} \circ \pi = d^{-1} \lambda \sum_{i \geq 0} \sum_{j \geq 0} T^i h \circ \alpha T^j$$

with $\lambda = \lim_{t \rightarrow +\infty} \lambda(t)$, where $\lambda(t)$ is a solution of the renewal equation introduced in [6] and $\mathbf{1}$ is the function identically equals to the unit on the state space E , and π is the unique invariant probability measure of the kernel P .

2 Strong stability and uniform ergodicity estimate

The following theorem gives the sufficient condition of the strong stability of the chain.

Theorem 1. *Let $E\xi_1 < 0$ and $\forall \delta \geq 0$, $E[\exp(\delta\xi_1)] < \infty$. Then, for all γ such that $\rho(\gamma) = E[\exp(\gamma\xi_1)] < 1$, the Markov chain X is aperiodic and strongly v -stable, where v was defined in (2).*

Proof

It is easy to remark that

- 1) We have $Tv(x) = \int_{\mathbb{R}^+} P(0 < x + \xi_1 \in dy) \exp(\gamma y) \leq \rho(\gamma) v(x)$.
- 2) It is obvious from relation (1) that $P = T + h \circ \alpha$ where "o" is the tensorial product of measure and function. Then the condition $T \geq 0$ is hold.
- 3) $p_1 = \alpha h = \int_{\mathbb{R}^+} \alpha(dx)h(x) = h(0) = P(\xi_1 \leq 0) > 0$.
- 4) $\alpha \mathbf{1} = \int_{\mathbb{R}^+} \delta_0(dx) = 1$.
- 5) $\pi h = \int_{\mathbb{R}^+} \pi(dx)P(x + \xi_1 \leq 0) = \pi(\{0\}) > 0$.
- 6) we verify easily that $\|h\|_\gamma \leq 1$ and $\|\alpha\|_\gamma = 1$.
- 7) $\|P\|_\gamma = \|T + h \circ \alpha\|_\gamma \leq \|T\|_\gamma + \|h\|_\gamma \|\alpha\|_\gamma < \rho(\gamma) + 1 < 2 < \infty$

Finally, from theorem 2 in ([2]) leads immediately to the aperiodicity and the strong stability of the chain X . ■

The following theorem gives an evaluation estimates of $\|P^t - \Pi\|_\gamma$.

Theorem 2. *Suppose that conditions of theorem (1) hold. Then, we have*

$$\|P^t - \Pi\|_\gamma \leq \rho^t + \frac{Max(d\pi(\{0\}), M)}{(1 - \rho)^2} + (d - 1)\sigma \quad (3)$$

where M is introduced in [6] and $\sigma = \|\Pi\|_\gamma$.

Proof

Il sufficed to compute λ . Effectively, we have

$$\lambda = (\pi h)(\alpha \mathbf{1}) = 1. \int_E \pi(dx).P(x + \xi_1 \leq 0) = \pi(\{0\})$$

we replace in inequality showed in [6], the values of λ , $\|\alpha\|_\gamma$ and we use an estimation of $\|h\|_\gamma$, the result wanted is proved. ■

Remark 1. It is easy to observe that the estimates obtained by theorem 2 for $d = 1$ does'nt necessary the same as like as the estimates obtained in [4].

Theorem 3. *Suppose that conditions of theorem (1) hold and $d = 1$. Then, we have the estimate of the potential of the process by the following theorem.*

$$\sum_{t \geq 0} \|P^t - \Pi\|_\gamma \leq \frac{1}{1 - \rho} + \frac{\Lambda_0}{(1 - \rho)^2} + \frac{\rho(1 + \rho)}{(1 - \rho)^3} \quad (4)$$

with $\Lambda_0 = \sum_{t \geq 0} |\lambda(t) - \pi(\{0\})| < \infty$.

Proof

The proof follows directly from the inequality obtained in [6], the values of λ , $\|\alpha\|_\gamma$ and an estimation of $\|h\|_\gamma$. ■

Then we have the following corollary.

Corollary 1. *Under the conditions of theorem (3) and for all $\Upsilon < 1$ such that $\rho\Upsilon < 1$, then we have*

$$\sum_{t \geq 0} \|P^t - \Pi\|_\gamma \leq \frac{1}{1 - \rho} + \Lambda(\Upsilon) \frac{\Upsilon}{(\Upsilon - 1)(1 - \rho)^2} + \frac{\rho(1 + \rho)}{(1 - \rho)^3} \quad (5)$$

Where $\Lambda(\Upsilon) = \sup_{t \geq 0} |\lambda(t) - \pi(\{0\})| \Upsilon^t$ and β is a constant introduced in [6].

Proof

From inequality of the corollary in [6], the values of λ , $\|\alpha\|_\gamma$ and an estimate of $\|h\|_\gamma$, the proof follows immediately. ■

Let us consider, the random walk process described by a chain Y satisfying the same recursive relation $Y_{n+1} = (Y_n + \zeta_{n+1})^+$, $n \geq 0$ as the chain X which taking values in $E = \mathbb{R}^+$, where $(\zeta_n)_{n \geq 0}$ is a sequence of independent random variables identically distributed with a common distribution function $G \neq F$. We suppose that Q is the transition kernel of the chain Y . We remark that

$$\varepsilon = \|Q - P\|_\gamma = \sup_{x \geq 0} \exp(-\gamma x) \int_0^{+\infty} |Q(x, dy) - P(x, dy)| \exp(\gamma y)$$

In the same way, we have

$$\sup_{t \geq 0} \|Q^t - P^t\|_\gamma = \sup_{t \geq 0} \sup_{x \geq 0} \exp(-\gamma x) \int_0^{+\infty} |Q^t(x, dy) - P^t(x, dy)| \exp(\gamma y)$$

Consequently, we obtain, by the following theorem, an estimate of strong v -stability of a chain X .

Theorem 4. *Under the same conditions of the theorem (1). Then, for all γ such that $\rho = \rho(\gamma) = E[\exp(\gamma\xi_1)] < 1$ and for all $\Upsilon < 1$ such that $\rho\Upsilon < 1$, we have*

$$\sup_{t \geq 0} \|Q^t - P^t\|_\gamma \leq \hbar \|Q - P\|_\gamma \quad (6)$$

where,

$$\hbar = \frac{1}{(1 - \rho - \epsilon)^2} \left(\rho^t + \frac{\text{Max}(d\pi(\{0\}), M)}{(1 - \rho)^2} + (d - 1)\sigma \right)$$

Proof

The proof follows directly from theorem [5], the values of λ , $\|\alpha\|_\gamma$, an estimation of $\|h\|_\gamma$ and we use the estimate in [5]. Consequently, the result is immediately obtained. ■

References

- [1] D. Aïssani. Perturbation des opérateurs de transition pour l'étude de la stabilité des chaînes de markov. *Cahiers Mathématiques, fascicule N°1*, (04):07–12, Avril 1988.
- [2] N.V. Kartashov. Criteria for uniform ergodicity and strong stability of Markov chains with a common phase space. *Theor. Prob. and Math. Stat.* N° 30, pages 71–89, 1985.
- [3] N.V. Kartashov. Asymptotic expansions and Inequalities in Stability Theorems for Markov Chains with relatively bounded perturbations. *Plenum Publishing Corporation*, pages 509–518, 1988.
- [4] Z. Mouhoubi and D. Aïssani. On the Uniform Ergodicity and Strong Stability Estimates of Waiting Process. *Bulletin of the International Statistical Institute, Volume LX, Book 2, Berlin 2003, pp 97-98.*
- [5] Z. Mouhoubi and D. Aïssani. On the Quantitatives estimates of the Strong Stability for Markov chains. *Proceeding of the XXII-th International Seminar on Stability Problems for Stochastic Models and the Seminar on Statistical Data Analysis. Varna, Bulgaria*, pages 66–67, 2002.

- [6] Z. Mouhoubi and D. Aïssani. On the Quantitative Estimates of the Uniform Ergodicity for Markov chains. *Proceeding of the 8-th International Vilnius Conference on Probability Theory and Mathematical Statistics. Vilnius, Lithuania*, pages 7–8, 2002.

Comportement asymptotique dans l'anneau de Cambridge.

Bénamar CHOUAF
Université Djilali Liabès
Faculté des Sciences
Laboratoire de Mathématiques
22 000 Sidi-Bel-Abbès-Algérie
Tel-Fax: 048 54 43 44
e-mail: bchouaf@univ-sba.dz/chouafb@yahoo.fr

Résumé

On considère N processeurs communiquant sur un canal de transmission fermé, ou anneau de Cambridge, un réseau dont les cellules sont installées de façon circulaire.

Chaque message est assemblé dans un paquet de longueur fixe. Les paquets à envoyer sont produits au hasard par les processeurs et les temps de passages passés par ces paquets sur l'anneau sont également aléatoires.

Les paquets étant expédiés, c'est à dire, les paquets déjà sur l'anneau, ont la priorité sur les paquets en attente.

L'un des principaux problèmes de l'anneau de Cambridge est qu'au même instant, des cellules peuvent être vides, alors que certaines files ne peuvent libérer leurs messages.

On établit les conditions de stabilité de l'anneau grâce à son modèle fluide et on écrit les différentes équations différentielles linéaires simples régies par les limites fluides.

On étudie également le comportement asymptotique de l'anneau en supposant que la génération des paquets suit une loi de Bernoulli de paramètre $\frac{\lambda}{N}$ et que les temps de services des paquets suit une loi uniforme sur $\{0, 1, \dots, N\}$.

Mots clés: *Modèle fluide, limite fluide, réseaux multiclassés, processus stochastiques, stabilité.*

Classification A M S: 60 K 25.

1 Stabilité au sens probabiliste

Les résultats de probabilités obtenus s'appuient en très grande partie sur la méthode de Liapunov. Ce sont les travaux effectués par ce dernier qui ont permis d'obtenir dans le cadre de systèmes régis par des équations différentielles, une méthode d'étude de la stabilité d'ordre mécanique. L'idée repose sur la construction des fonctions dites de Liapunov. la méthode a ensuite été adoptée pour l'étude de la stabilité de modèles stochastiques.

Il a été démontrée que, si $f(X_n)$ est en moyenne strictement décroissante tant que X_n n'a pas atteint F , alors le réseau est stable, où f est une fonctionnelle positive adéquate pour X_n et la partie F . Les premières fonctionnelles jouant ce rôle sont les fonctions de Liapunov. la construction de telles fonctions est très délicate, c'est pourquoi nous utiliserons les limites fluides stochastiques. Elles sont obtenues après renormalisation de suite de modèles dont les charges initiales tendent vers l'infini. On applique ces propriétés à l'étude de la stabilité et du comportement asymptotique dans le modèle que nous étudions qui est l'anneau de Cambridge, un réseau dont les serveurs sont installés de façon circulaire.

2 Anneau de Cambridge

2.1 Présentation du problème

- Ce problème est composé de N files d'attente numérotées de 1 à N réparties autour d'un réseau circulaire composé de N cellules également numérotées de 1 à N .
- Ce réseau circulaire tourne, toujours dans le même sens, en un temps discret. Lorsque à un instant donné, une file d'attente est face à la cellule k , elle est à l'instant suivant face à la cellule $(k - 1)$ ou face à la cellule N si $k = 1$.
- Les variables aléatoires indiquant pour chaque unité de temps et chaque file d'attente le nombre d'arrivées sont indépendantes mais leur loi dépend de la file d'attente.
- Lorsqu'une file est non vide, le client de tête est servi par le serveur situé en face si celui-ci est libre. Le client effectue alors un trajet de durée aléatoire discrète. Toutes les variables indiquant un temps de trajet sont *i.i.d.*.
- Chaque serveur travaillant en général avec différentes files d'attente, celles-ci évoluent de manière indépendantes.
- le but de ce réseau est de servir et transporter les messages de leur file d'attente initiale vers leur file d'arrivée.
- On admettra que certains messages puissent effectuer plusieurs tours avant de rejoindre leur cellule finale.
- Entre deux unités de temps, l'ordre des opérations est le suivant:
 - 1) Tandis que l'anneau tourne d'un cran, les files reçoivent de nouveaux messages.
 - 2) Les messages ayant atteint leur destination, libèrent leur cellule et quitte le réseau définitivement.

- 3) Toute file d'attente non vide place un message dans la cellule qui lui est accessible, si celle ci est libre.
- 4) Une cellule ne peut transporter qu'un seul message à la fois.

3 Stabilité et Comportement asymptotique

Pour $j \in \mathbb{N}$, on notera $[j]$ l'unique entier $l \in \{1, \dots, N\}$ tel que $j = l$ modulo N . Soit $Q_j(n)$ (resp. $L_j(n)$) le nombre de clients en attente à la file j (resp. le temps résiduel d'occupation, par le client éventuellement présent, de la cellule en face de la file j), à l'instant n , c'est-à-dire juste avant le départ des cellules ($L_j(n) = 0$ signifie que la cellule va repartir à vide). Posons $Q(n) = (Q_j(n))_{1 \leq j \leq N}$, et $L(n) = (L_j(n))_{1 \leq j \leq N}$. L'état du système à l'instant n est $X(n) = (Q(n), L(n))$.

À l'instant $n + 1$, les cellules ont avancé d'un cran, et la cellule qui était en face de la file j se retrouve en face de la file $[j + 1]$; la file j a reçu $\delta_j(n)$ nouveaux arrivants, le p^e client de la file j demande un trajet de longueur $\sigma_j(p) \geq 1$; les clients en fin de trajet sont déchargés; les cellules vides chargent un client si la file en vis-à-vis n'est pas vide; les temps résiduels d'occupation sont réactualisés en conséquence. On forme ainsi l'état $X(n + 1)$.

Notons $\delta(n) = (\delta_j(n))_{1 \leq j \leq N}$. On suppose que la suite $(\delta(n))_{n \in \mathbb{N}}$, à valeurs dans \mathbb{N}_+^N , est i.i.d., et que toutes les files reçoivent autant de clients en moyenne par unité de temps; on notera ν la valeur commune des $\mathbb{E}(\delta_j(n))$, $1 \leq j \leq N$. D'autre part, on suppose que les variables $\sigma_j(p)$, $1 \leq j \leq N$ $p \in \mathbb{N}$, sont i.i.d., de moyenne $\frac{1}{\mu}$, et indépendantes des $\delta(n)$, $n \in \mathbb{N}$. On vérifie alors que la suite $X = (X(n))_{n \in \mathbb{N}}$ forme une chaîne de Markov à temps discret, à valeurs dans l'espace (discret) $E = \mathbb{N}^N \times \mathbb{N}^N$. Le théorème suivant est dû à V.Dumas[4]

Théorème 1. [4] *On fait l'hypothèse suivante sur les arrivées de clients dans le système:*

$$\mathbb{P}(\delta(n) = (0, 0, \dots, 0)) > 0.$$

Alors, si $\frac{\nu}{\mu} < 1$, l'anneau de Cambridge est ergodique.

Pour plus de détails sur la stabilité de l'anneau de Cambridge, voir [4] et [3] où les auteurs ont étudié la stabilité en utilisant les limites fluides stochastiques.

Une étude de l'anneau de Cambridge a déjà été effectuée par Coffmann et al[1]. Ils supposent que les intensités d'arrivées sont égales. Ils ont prouvé la stabilité sous l'hypothèse $\nu N < 1$ dans le cas où les messages ne peuvent effectuer plus d'un tour d'anneau. Ils ont également prouvé la stabilité sous l'hypothèse $\frac{\nu}{\mu} < 1$, lorsque le temps de séjour suit une loi géométrique. Enfin des conjectures sur l'asymptotique sont énoncées lorsque N tend vers l'infini et νN reste borné

Le travail présenté traite l'une de ces conjectures. Notre objectif sera de démontrer le résultat suivant.

Théorème 2. *Supposons que $\delta_j(n) \sim \mathcal{B}(\frac{\lambda}{N})$ (loi de Bernoulli) et $\sigma_j(p) \sim \mathcal{U}(1, \dots, N - 1)$ (loi uniforme). Si $\lambda < 2$ alors $\mathbb{E}(Q) = \Theta(\frac{1}{N})$.*

Preuve: Pour tout $j \in \{1, 2, \dots, N\}$ et $n \in \mathbb{N}$ on définit l'événement $V_j(n)$.

$V_j(n)$: "La cellule devant la file j à l'instant n est vide".

Comme le réseau est stable, alors $\mathbb{P}(V_j(n))$ ne dépend pas de n .

Posons $\alpha = \mathbb{P}(V_j(n))$ et Prenons l'événement contraire: $\overline{V_j(n)}$.

$\overline{V_j(n)}$: "La cellule devant la file j à l'instant n est occupée" (c'est-à-dire. le message qui se trouve à l'intérieur à l'instant n n'a pas encore terminé son service).

On définit aussi les événements:

$R_j(n)$ "On a une arrivée à l'instant n dans la file j ".

$\overline{R_j(n)}$ "Pas d'arrivée à l'instant n dans la file j ".

Soit $k \in \mathbb{N}$

a) $k = 0$, alors

$$\begin{aligned} \mathbb{P}(Q_j(n) = 0) &= \mathbb{P}\left[(Q_j(n) = 0) \cap \left((Q_j(n-1) = 0) \cap (Q_j(n-1) = 1)\right)\right] \\ &= \mathbb{P}\left[(Q_j(n) = 0) \cap (Q_j(n-1) = 0)\right] \cup \mathbb{P}\left[(Q_j(n) = 0) \cap (Q_j(n-1) = 1)\right] \end{aligned}$$

et par suite

$$\begin{aligned} \mathbb{P}(Q_j(n) = 0) &= \mathbb{P}(Q_j(n) = 0 / Q_j(n-1) = 0) \mathbb{P}(Q_j(n-1) = 0) \\ &\quad + \mathbb{P}(Q_j(n) = 0 / Q_j(n-1) = 1) \mathbb{P}(Q_j(n-1) = 1) \end{aligned}$$

d'où

$$\begin{aligned} q_0 = \mathbb{P}(Q_j(n) = 0) &= q_0 \left(1 - \frac{\lambda}{N}\right) + q_1 \left(1 - \frac{\lambda}{N}\right) \alpha = \left(1 - \frac{\lambda}{N}\right) [q_0 + \alpha q_1] \\ (1) \quad q_0 &= \left(1 - \frac{\lambda}{N}\right) [q_0 + \alpha q_1] \end{aligned}$$

b) si $k = 1$, alors

$$(2) \quad q_1 = \left(1 - \frac{\lambda}{N}\right) [q_1 + \alpha q_2] + q_1 \frac{\lambda}{N} \alpha.$$

Posons $A = \frac{\frac{\lambda}{N}(1-\alpha)}{\alpha(1-\frac{\lambda}{N})}$, de (2) on peut écrire q_2 sous la forme:

$$(3) \quad q_2 = A q_1$$

c) Soit $k \geq 2$, alors

$$(4) \quad \forall k \geq 2 \quad (q_{k+1} - q_k) = A(q_k - q_{k-1})$$

Ainsi

$$q_{k+1} = A q_k - A q_1 + q_2.$$

D'après (3), $q_2 - A q_1 = 0$ et donc

$$(5) \quad \forall k \geq 1 \quad q_{k+1} = A q_k$$

On définit la fonction f comme suit:

$$\begin{aligned} f : \{1, 2, \dots, N\} &\longrightarrow \{1, 2, \dots, N\} \\ j &\longrightarrow f(j) = \begin{cases} j+1 & \text{si } j < N \\ 1 & \text{si } j = N \end{cases} \end{aligned}$$

Considérons l'événement:

$\bar{V}_j(n)$ "La cellule devant la file j à l'instant n est occupée" (c'est-à-dire. elle contient un message)

Posons $I_{f^i(j)}$ l'événement "Ce message est issue de la file $f^i(j)$ "

Posons $f^0 = I_{\{1,2,\dots,N\}}$ (fonction identité) et pour tout $n \geq 1$: $f^n = f \circ f^{n-1}$

Par suite $f^0 = f^N$

Comme $\sigma \sim \mathcal{U}(1, 2, \dots, N-1)$ alors $i \neq 1$ et $i \neq N$

Soit

$$(6) \quad \alpha = \frac{1}{[1 + (1 - q_o)^{\frac{N-2}{2}}]}$$

D'après (5), $\forall k \geq 2$, on a:

$$q_k = Aq_{k-1} = A^{k-1}q_1$$

Ainsi

$$(7) \quad q_1 = (1 - q_o)(1 - A).$$

De (1), on déduit que

$$q_o = \frac{(1 - \frac{\lambda}{N})\alpha}{\frac{\lambda}{N}}q_1$$

et d'après (7),

$$q_o = \frac{(1 - \frac{\lambda}{N})\alpha(1 - q_o)(1 - A)}{\frac{\lambda}{N}}.$$

Soit

$$(8) \quad q_o[\frac{\lambda}{N} + (1 - \frac{\lambda}{N})\alpha(1 - A)] = (1 - \frac{\lambda}{N})\alpha(1 - A)$$

et

$$(9) \quad q_o = \frac{\alpha - \frac{\lambda}{N}}{\alpha} = 1 - \frac{\lambda}{\alpha}$$

et

$$(10) \quad \alpha = \frac{\frac{\lambda}{N}}{1 - q_o}.$$

D'où;

$$(11) \quad q_o = \frac{1}{1 + \frac{1}{(\frac{1}{\lambda} - \frac{1}{2})N}}$$

Maintenant on peut calculer $\mathbb{E}(Q)$. Soit

$$(12) \quad \mathbb{E}(Q) = \frac{1 - q_o}{1 - A}.$$

D'où d'après (9) et l'expression de $1 - A$;

$$(13) \quad \mathbb{E}(Q) = \frac{(1 - \frac{\lambda}{N})}{\frac{1}{(1 - q_o)} - 1}$$

l'équation (11) implique:

$$1 - q_o = 1 - \frac{1}{1 + \frac{1}{(\frac{1}{\lambda} - \frac{1}{2})^N}} = \frac{\frac{1}{(\frac{1}{\lambda} - \frac{1}{2})^N}}{1 + \frac{1}{(\frac{1}{\lambda} - \frac{1}{2})^N}}$$

$$\frac{1}{1 - q_o} = \frac{1 + \frac{1}{(\frac{1}{\lambda} - \frac{1}{2})^N}}{\frac{1}{(\frac{1}{\lambda} - \frac{1}{2})^N}} = 1 + (\frac{1}{\lambda} - \frac{1}{2})^N$$

D'après (13)

$$\mathbf{E}(Q) = \frac{(1 - \frac{\lambda}{N})}{1 + (\frac{1}{\lambda} - \frac{1}{2})^N - 1} = \frac{(1 - \frac{\lambda}{N})}{(\frac{1}{\lambda} - \frac{1}{2})^N}$$

Soit

$$(14) \quad \mathbf{E}(Q) = \frac{(1 - \frac{\lambda}{N})}{(\frac{1}{\lambda} - \frac{1}{2})^N}$$

Donc

$$\mathbf{E}(Q) = \Theta\left(\frac{1}{N}\right).$$

Références

- [1] E. G. COFFMAN, JR., E. N. GILBERT, A. G. GREENBERG, F. T. T. LEIGHTON, P. ROBERT, AND A. L. STOLYAR (1995), Queues served by a rotating ring, Stochastic Modèles, 11, pp. 371-394.
- [2] E. G. COFFMAN, JR., NABIL KAHALE AND F.T. T. LEIGHTON(1998), Processors-ring communicating: A tight asymptotic bound on packet waiting times,SIAM J.COMPUT, Vol. 27, N^O. 5, pp. 1221-1236.
- [3] J.F.DANTZER(2000), Stabilité des réseaux de files d'attente et limites fluides stochastiques,Thèse de doctorat, Université de Versailles-St-Quentin.
- [4] V.DUMAS(1995), Approches fluides pour la stabilité et l'instabilité de réseaux de files d'attente stochastiques à plusieurs classes de clients, Thèse de doctorat, École polytechnique.

Strong approximation of quantile function for strong mixing and censored processes

Elias Ould-Saïd

Ourida Sadki

L.M.P.A. J. Liouville

Institut de Mathématiques

Univ. du Littoral Côte d'Opale

Univ. des Sci. et Tech. Houari Boumédiène

BP 699, 62228 Calais, France

BP 32 El Alia, 16111, Bab-Ezzouar, Alger, Algérie

e-mail: ouldsaid@lmpa.univ-littoral.fr

e-mail: sadki@hotmail.com

January 27, 2004

Let $(X_i)_{i \geq 1}$ be a sequence of strong-mixing random variables with common unknown absolutely continuous distribution function F subject to random right censoring. Let ξ_p denote the p th ($p \in]0, 1[$) quantile function of the marginal distribution function of the X_i 's which is estimated by a sample quantile ξ_{np} . In this paper, we derive the strong consistency and a Bahadur-type representation for ξ_{np} , the quantile function of the Kaplan-Meier estimator of F for strong-mixing processes.

AMS Subject classification 1991. 62G05, 62G20.

Key words and phrases. Quantile function, Strong-mixing, Kaplan-Meier estimator, Censored data, Bahadur representation, Strong consistency.

1 Introduction.

Consider a sequence of strictly stationary random variables (rv) X_1, X_2, \dots with common unknown absolutely continuous distribution function (df) F . These rv are regarded as the lifetimes of the items under study. In many situations, due to possible withdrawals of the items from the life testing experimentation, the lifetimes may not be directly observable. Instead, we observe only censored lifetimes. That is, assuming that C_1, C_2, \dots, C_n are n censoring rv with common unknown continuous df G , we observe only the n pairs $\{(Z_i, \delta_i), i = 1, 2, \dots, n\}$, where $Z_i = X_i \wedge C_i$ and $\delta_i = \mathbb{I}_{\{X_i \leq C_i\}}$ (where \mathbb{I}_A denotes the indicator function of the set A).

Throughout this paper, to ensure the identifiability of the model, we suppose that the sequences $(X_i)_{i \geq 1}$ and $(C_i)_{i \geq 1}$ are mutually independent, but each of the sequences satisfies some kind of dependence such as α -mixing (for definition see (5) below).

Now, it is known (see Lemma 2 in Cai (2001)) that the Z_i are α -mixing (with an appropriate coefficient α) with common df $H(t) = 1 - (1 - F(t))(1 - G(t))$ (by independence) and the uncensored model is the special case of the censored model with $G = 0$. Consider the following two stochastic processes

$$N_n(t) = \sum_{i=1}^n \mathbb{I}_{\{Z_i \leq t, \delta_i = 1\}} = \sum_{i=1}^n \mathbb{I}_{\{X_i \leq t \wedge C_i\}}$$

(the number of uncensored observations less than or equal to t), and

$$Y_n(t) = \sum_{i=1}^n \mathbb{I}_{\{Z_i \geq t\}}.$$

Then, the well known Kaplan-Meier (1958) estimator (KME) based on the n pairs $\{(Z_i, \delta_i), 1 \leq i \leq n\}$ is given by

$$(1) \quad 1 - \widehat{F}_n(x) = \prod_{s \leq x} \left(1 - \frac{dN_n(s)}{Y_n(s)} \right)$$

where $dN_n(t) = N_n(t) - N_n(t^-)$ and $N_n(t^-) = \lim_{\varepsilon \rightarrow 0^+} N_n(t - \varepsilon)$.

The cumulative hazard function $\Lambda(t)$ is defined by

$$(2) \quad \Lambda(t) = \int_0^t \frac{dF(s)}{1 - F(s^-)}.$$

Clearly, $\Lambda(t) = -\log(1 - F(t))$ in our case.

$$\text{Let } F_*(t) = \mathbb{P}(Z_i \leq t, \delta_i = 1) = \int_0^{+\infty} F(t \wedge s) dG(s) = \int_0^t (1 - G(s)) dF(s).$$

From (2) it can be shown that

$$\Lambda(t) = \int_0^t \frac{dF_*(s)}{\overline{H}(s)}$$

where $\overline{H} = 1 - H$.

Since $N_n(t)$ and $Y_n(t)$ are clearly the empirical estimates of $F_*(t)$ and $\overline{H}(t)$ respectively, the empirical cumulative hazard function is given by

$$\widehat{\Lambda}_n(t) = \int_0^t \frac{dN_n(s)}{Y_n(s)},$$

which is referred to as the Nelson-Aalen estimator of $\Lambda(t)$ in the literature.

In the case where $(X_i)_{i \geq 1}$ and $(C_i)_{i \geq 1}$ are two independent α -sequences, Cai (1998, 2001) established the strong representation and asymptotic normality for (1), with uniform consistency and gave the rates. He also studied a stochastic process generated by (1), which converges weakly to a Gaussian process with a specified covariance structure. For more details on the applications of the KME for dependent data in medicine and agriculture, we refer the reader to the papers by Wei *et al.* (1989), Kohler (1994), Kohler and Symanovski (1995) and Kang and Kohler (1997).

For $0 < p < 1$, the p th quantile of $F(t)$ is defined by

$$(3) \quad F^{-1}(p) = \inf\{x \in \mathbb{R}; F(x) \geq p\}$$

and the sample estimator of $F^{-1}(p)$ is defined by

$$(4) \quad \widehat{F}_n^{-1}(p) = \inf\{x \in \mathbb{R}; \widehat{F}_n(x) \geq p\}$$

where $\widehat{F}_n(x)$ is the KME defined in (1).

In the independent framework with no censoring, the properties of estimator (4) (where \widehat{F}_n is replaced by the empirical df F_n) have been extensively studied (see e.g. Scorg'ò (1983) or Shorack and Wellner (1986)). Under ϕ -mixing condition (for the definition see Doukhan (1996)), the Bahadur representation

was obtained by Sen (1972) and the extension to the α -mixing case was obtained by Yoshihara (1995). The strong approximation of the quantile process defined by $\rho_n(s) = n^{1/2}f(F^{-1}(s))\left(F^{-1}(s) - F_n^{-1}(s)\right)$, where f is density function and F_n^{-1} is sample quantile function and F_n , was obtained by Fotopoulos *et al.* (1994) and was later improved by Yu (1996).

For a censored model with $(X_i)_{i \geq 1}$ and $(C_i)_{i \geq 1}$ independent and identically distributed (iid) sequences and mutually independent, Padgett and Thombs (1989) stated the strong consistency and asymptotic normality for a smooth estimator of (3). Sander (1975) obtained some asymptotic properties, and Csörgö (1983) and Cheng (1984) discussed strong approximation results with some applications for (4).

In this paper, we establish the pointwise consistency and give a strong representation of Kaplan-Meier quantile function $\hat{F}_n^{-1}(\cdot)$ under a strong-mixing hypothesis.

In order to define the strong mixing coefficient, we introduce the following definitions. Denote by \mathcal{F}_1^k the σ -algebra generated by $(X_1), \dots, (X_k)$ and \mathcal{F}_{k+n}^∞ that generated by $(X_{k+n}), \dots$. Now, for any k , set

$$(5) \quad \alpha(n) = \sup_{A \in \mathcal{F}_1^k} \sup_{B \in \mathcal{F}_{k+n}^\infty} \{|\mathbb{P}(A \cap B) - \mathbb{P}(A)\mathbb{P}(B)|\} ; n = 1, 2, \dots$$

The process $(X_n)_{n \geq 1}$ is said to be strongly mixing if

$$(6) \quad \lim_{n \rightarrow \infty} \alpha(n) = 0.$$

This condition specifies an asymptotic of independence between the past and future of $(X_n)_{n \geq 1}$.

There exists many processes fulfilling the strong mixing property. We quote here the usual ARMA processes which are geometrically strongly mixing, *i.e.*, there exists $\rho \in (0, 1)$ and $a > 0$ such that, for any $n \geq 1$, $\alpha(n) \leq a\rho^n$ (see, e.g., Jones (1978)). The threshold models, the EXPAR models (see Ozaki (1979)), the simple ARCH models (see, Engle (1982), Masry and Tjøstheim (1995, 1997)) and their extensions (see, Diebolt and Guégan (1993)) and the bilinear Markovian models are geometrically strongly mixing under some general ergodicity conditions. Auestad and Tjøstheim (1990) provided excellent discussions on the role of α -mixing for model identification in nonlinear times series analysis. Chen and Tsay (1993) showed that the functional autoregressive process is geometrically ergodic under certain conditions.

Now introduce some assumptions that are used to state our results gathered below for easy reference.

A1. $(X_i)_{i \geq 1}$ is a sequence of stationary α -mixing random variables with continuous df F and mixing coefficient $\alpha_1(n)$.

A2. $(C_i)_{i \geq 1}$ is a sequence of stationary α -mixing random variables with continuous df G and mixing coefficient $\alpha_2(n)$. Moreover the censoring times are independent of $(X_i)_{i \geq 1}$.

A3. $\alpha(n) = O(n^{-\nu})$ for some $\nu > 3$, with $\alpha(n) = \max(\alpha_1(n), \alpha_2(n))$.

REMARK 1. Assumption **A3** is needed in order to use the strong representation of the Kaplan-Meier estimator obtained by Cai (1998) which comes from Dhompongsa's (1984) result. We point out that we do not require the censoring times C_n to be mutually independent, which is done by Lecoutre and Ould-Saïd (1995a, b), Cai (1998) and Cai and Roussas (1998).

For any df $W(\cdot)$, define the (possibly infinite) time $\tau_W = \sup\{x; W(x) < 1\}$, and let $\tau < \tau_H$. It is clear by independence that $\tau_H = \tau_F \wedge \tau_G$.

Let $g(x) = \int_0^x (\overline{H}(u))^{-2} dF_*(u)$ and for positive reals z and x , and $\delta \in \{0, 1\}$ let

$$\xi(z, \delta, x) = g(z \wedge x) - \mathbb{I}_{\{z \leq x, \delta=1\}} / \overline{H}(z).$$

It can be shown that $\mathbb{E}[\xi(z_i, \delta_i, x)] = 0$ and $cov(\xi(z_i, \delta_i, s), \xi(z_i, \delta_i, t)) = g(s \wedge t)$.

Now, we are in position to give our main results. The first result gives the strong consistency of \hat{F}_n^{-1} .

Theorem 1 *Let $0 < p < 1$. Under **A1-A3**, assuming that F is continuous and that $F^{-1}(p)$ is the unique solution of $F(x) = p$, then*

$$\hat{F}_n^{-1}(p) \longrightarrow F^{-1}(p) \text{ a.s.}$$

The strong representation of the sample quantile is given in the following

Theorem 2 Let $0 < p < 1$. Under **A1-A3**, assuming that F is Lipschitz continuous and continuously differentiable at $F^{-1}(p)$ with $F'(F^{-1}(p)) = f(F^{-1}(p)) > 0$, then

$$\begin{aligned}\hat{F}_n^{-1}(p) - F^{-1}(p) &= \frac{p - \hat{F}_n^{-1}(F^{-1}(p))}{f(F^{-1}(p))} + o(a_n) \text{ a.s.} \\ &= o_{\mathbb{P}}(n^{-1/2})\end{aligned}$$

where $a_n = \sqrt{\log \log n/n}$. In addition, if F is twice continuously differentiable at $F^{-1}(p)$, then

$$\hat{F}_n^{-1}(p) - F^{-1}(p) = \frac{p - \hat{F}_n^{-1}(F^{-1}(p))}{f(F^{-1}(p))} + O\left(\frac{\log^{\beta-\lambda} n}{\sqrt{n}}\right) \text{ a.s. where } 0 < \beta < 1 \text{ and } \lambda > 0.$$

REMARK 2. In the α -mixing case, we can not achieve the same rate as in the iid case. The main reason is that our approach utilizes the strong approximation of Kaplan-Meier estimator which is expressed as a mean of random variables with negligible remainder term of order $\log^{-\lambda} n/n^{1/2}$. This is not as sharp as in iid case, for which the order is $\log n/n$ (see, Lo, Mack and Wang (1989)).

References

- [1] BOSQ, D., *Nonparametric Statistics for Stochastic Processes. Estimation and Prediction*. Lect. Notes in Statist., **110**, Springer Verlag. 1998.
- [2] CAI, Z., 1998. Asymptotic properties of Kaplan-Meier estimator for censored dependent data, *Statist. Probab. Lett.*, **37**, 381-389.
- [3] CAI, Z., 2001. Estimating a distribution function for censored times series data, *J. Multivariate Anal.*, **78**, 299-318.
- [4] CAI, Z., ROUSSAS, G.G., 1998. Kaplan-Meier estimator under association, *J. Multivariate Ana.* **67**, 318-348.
- [5] CARBON, M., 1988. Inégalités de grande déviation dans les processus. Application à l'estimation fonctionnelle. Thesis Univ. Paris VI.

- [6] CHENG, K.F., 1984. On almost sure representation for quantiles of the product limit estimator with applications, **46**, 426-443.
- [7] ENGLE, R.F., 1982. Autoregressive conditional heteroscedasticity with estimates of the variance of united Kingdom inflation, *Econometrics*, **50**, 4, 987-1007.
- [8] FOTOPOULOS, S., AHN, S.K., CHO, S., 1994. Strong approximation of the quantile processes and its applications under strong mixing properties, *J. Multivariate Anal.*, **51**, 17-47.
- [9] KAPLAN, E.M., MEIER, P., 1958. Nonparametric estimation from incomplete observations, *J. Amer. Statist. Assoc.*, **53**, 457-481.
- [10] LECOUTRE, J.P., OULD-SAÏD, E., 1995a. Hazard rate estimation for strong-mixing and censored processes. *J. Nonparametric Statist.*, **5**, 83-89.
- [11] LECOUTRE, J.P., OULD-SAÏD, E., 1995b. Convergence of the conditional Kaplan-Meier estimate under strong mixing, *J. Statist. Plann. Inf.*, **44**, 359-369.
- [12] MASRY, E., TJOSTHEIM, D., 1995. Nonparametric estimation and identification of nonlinear ARCH time series: Strong convergence and asymptotic normality. *Econ. Theory*, **11**, 258-289.
- [13] PADGETT, W.T., THOMBS, L.A., 1989. A smooth nonparametric quantile estimation from right censored data., *Statist. Probab. Lett.*, **7**, 113-121.
- [14] SCÖRGÖ, S., *Quantile process and their applications*, SIAM, Philadelphia PA, 1983.
- [15] SHORACK, G.R., WELLNER, J.A. *Empirical processes with applications to statistics*, Wiley, 1986.
- [16] YU, H., 1996. A note on strong approximation for quantile processes of strong mixing sequences, *Statist. Probab. lett.*, **30**, 1-7.

Sur les Systèmes d'Attente à Serveurs Non Fiables

K. Abbas et D. Aïssani

L. A. M. O. S.

Laboratoire de Modélisation et d'Optimisation des Systèmes

Université de Béjaïa (Algérie)

Introduction

Ces dernières années, les gestionnaires ont pris conscience que la mise en place de systèmes faiblement fiables ne correspond nullement aux objectifs de rendement maximum, propres à tout esprit rationnel. C'est pourquoi, lors de la conception de tout système complexe, il est nécessaire de se soucier du problème de la fiabilité de son fonctionnement.

En théorie des files d'attente, nous nous intéressons également à l'étude de l'influence de la non-fiabilité des serveurs sur les caractéristiques du système. Ainsi, les modèles mathématiques des systèmes et réseaux de files d'attente, les plus élaborés sont justement ceux qui tiennent compte de la possibilité pour les serveurs de tomber en panne. En effet, la fiabilité des serveurs influe beaucoup sur les caractéristiques du système. Par exemple, plus les indices de fiabilité des serveurs sont bas, plus le nombre de clients dans la file est élevé et plus la durée d'attente de chaque client dans la file est longue. Concernant la nature de ces interruptions, on distingue généralement les cas suivants :

- **Pannes de nature conservatrices**

Dès que la panne se produit, le service est interrompu, mais le client reste auprès du serveur et attend que ce dernier soit réparé. Après la réparation, le service reprend là où il a été interrompu (la partie de service déjà acquise est donc conservée).

- **Pannes de nature non conservatrices**

Dans ce cas, la partie de service déjà acquise est détruite. Après la réparation de la panne, le service reprend à zéro.

- **Pannes avec perte définitive de client**

Dès que la panne se produit, le client quitte le système pour de bon.

- **Pannes avec perte momentanée de client**

Dès que la panne se produit, le client quitte le serveur et entre en orbite. Par la suite, son comportement ne diffère en rien de celui des clients qui se trouvent déjà en orbite.

L'objectif de cette communication est de présenter quelques modèles de systèmes de files d'attente à serveurs non fiables ayant un intérêt pratique et qui ont été très peu étudiés par les spécialistes. Dans le premier paragraphe, nous réalisons une synthèse des résultats connus sur les systèmes d'attente à serveurs non fiables. Le deuxième paragraphe, concerne les systèmes les plus classiques.

1 Synthèse bibliographique

Les systèmes de files d'attente dont les serveurs sont sujets à des pannes et réparations ont été étudiés par plusieurs chercheurs [2, 4, 3, 9, 11]. White et Christie [18] ont étudié un système de files d'attente avec pannes et leur relation avec le modèle prioritaire. Depuis ce travail, plusieurs travaux ont été élaborés. Avi-Itzhak. et al. [2] et Gaver [9] ont considéré le système d'attente $M/G/1$, et le système $M^X/G/1$, avec serveur sujet à des pannes, sous différentes disciplines de service. Mitrani et al. [14], Neuts et Lucantoni [13], Fakinos [7] ont réalisé une extension détaillée des résultats obtenus pour les systèmes à un seul serveur aux systèmes à N -serveurs. Une autre extension de ces résultats a été élaborée par Kulkarni et Choi [11] au modèle avec rappels, dont ils ont étudié le système de files d'attente $M/G/1$ avec rappels et serveur sujet à des pannes et réparations. Madan [12], a considéré le système de files d'attente $M/M^X/1$ dont le processus de service est sujet à des interruptions, comme il a obtenu la fonction génératrice de la longueur de la file.

Dans tous ces modèles, on supposait que les durées de réparation des serveurs sont des variables aléatoires poissonniennes. Néanmoins, dans les situations réelles, les pannes poissonniennes sont très limitées. Pour cela, plusieurs auteurs [8, 6, 15] ont considéré les périodes de réparation comme étant des variables aléatoires générales, A. Fedegruen et L. Green [8], ont étudié le système $M/G/1$ à serveur non fiable, dont la période d'activité, où le serveur est normalement opérationnel, a une distribution de type phase, et la période d'inactivité, où le serveur est non opérationnel, a une distribution générale. Dans ce travail, ils ont obtenu des bornes et des approximations pour le temps moyen d'attente, la distribution du nombre de clients dans le système à l'état stationnaire et la probabilité de retard. Bruneel [4] a considéré le modèle $Geo^X/D/1$ à serveur sujet à des pannes et réparations, sous l'hypothèse que la période d'activité a une fonction génératrice rationnelle. Il a obtenu aussi la fonction génératrice de la longueur de la file à un instant arbitraire. De même, Bruneel [3], a également considéré un système à plusieurs serveurs à temps discret dans lequel les serveurs disponibles dans un intervalle de temps sont indépendants et identiquement distribués. Il a déterminé la fonction génératrice de la longueur de la file. D. S. Lee [6], a considéré le système de files d'attente $Geo^X/D/1$ pour lequel les périodes de service interrompues suivent un processus semi-markovien.

Lors de l'étude de tous ces modèles, on supposait que le processus d'arrivée des clients est poissonien. Cependant, en pratique, il est très difficile de décrire en toute clarté les trafics de corrélation, qui ont des effets significatifs sur les mesures de performance (la période d'attente et la probabilité de blocage) par un processus de Poisson. A cet effet, N. Bhat [17] a étudié le système de files d'attente $SPP/G/1$ dont le processus d'arrivée des clients est un processus de Poisson. Ainsi, une approximation de la durée d'attente a été réalisée par ce dernier. Takine et Sengupta [16], ont étudié un système de files d'attente à un seul serveur sujet à des pannes, où les clients arrivent au système suivant un Processus d'Arrivée Markovien (MAP). Sous des hypothèses supplémentaires, Bong Dae Choi et Jinmin Chung [5], ont étudié le système de files d'attente $MAP/G/1$, dont le serveur est sujet à des pannes et réparations, alors que les arrivées des clients suivent un Processus d'Arrivée Markovien (MAP), et les périodes de réparations sont indépendantes et identiquement

distribuées selon une loi générale. Ce dernier modèle est une généralisation du modèle $M/G/1$ avec serveur sujet à des pannes poissonniennes et réparations générales. Donc, une étude mathématique unifiée et une généralisation des travaux classiques précédents de la distribution de la longueur de la file d'attente ont été réalisées, et ceci en utilisant la notion de la transformée de Laplace étendue.

Dans [10], K. Jau-Chuan a étudié la commande optimale de la file d'attente $M/G/1$ avec serveur non fiable et deux types de vacances.

2 Système $M/M/1$ à serveur non fiable [1]

Pour le système à serveur non fiable le plus simple, $M/M/1$, le flux des arrivées est poissonien, de paramètre λ , et la durée de service est exponentielle, de paramètre μ . La capacité d'attente est illimitée et il y a un seul serveur sujet à des pannes.

Soit α la durée de bon fonctionnement du serveur et β , sa durée de réparation.

On suppose que α et β sont des variables aléatoires indépendantes entre elles, indépendantes des inter-arrivées ξ et des durées de service τ . De plus, ces variables aléatoires sont des lois exponentielles de paramètres θ et r respectivement.

On peut décrire l'état du système par le processus aléatoire markovien,

$$S(t) = \{X(t), e(t); t \geq 0\},$$

où $X(t)$ est "le nombre de clients présents dans le système à l'instant t ($t \geq 0$)", et

$$e(t) = \begin{cases} 0 & \text{Si le serveur est en bon état,} \\ 1 & \text{Si le serveur est en panne.} \end{cases}$$

Il est aisé de vérifier que les probabilités stationnaires,

$$\lim_{t \rightarrow +\infty} P_{0j}(t) = P_{0j}, \quad j \geq 0; \quad \lim_{t \rightarrow +\infty} P_{1j}(t) = P_{1j}, \quad j \geq 1,$$

sont solutions du système d'équations algébriques suivant :

$$\begin{cases} (\lambda + \theta)P_{00} = \mu P_{01} + rP_{10}, \\ (\lambda + \mu + \theta)P_{0j} = \mu P_{0j+1} + \lambda P_{0j-1} + rP_{1j} & j \geq 1, \\ (\lambda + r)P_{10} = \theta P_{00}, \\ (\lambda + r)P_{1j} = \theta P_{0j} + \lambda P_{1j-1} & j \geq 1. \end{cases}$$

La condition d'ergodicité d'un tel système est de la forme : $\rho = \lambda(1 + \theta/r) < 1$.

3 Système $M/M/1/0$ à serveur non fiable [1]

Le système de files d'attente $M/M/1/0$ à serveur non fiable est un cas particulier du système d'attente $M/M/1$ non fiable. Ici, il n'y a aucun client dans la file.

L'étude de ce système est subdivisée en deux parties. Dans la première, on suppose que le serveur ne peut tomber en panne lorsqu'il est actif (pannes passives), et dans la deuxième partie, on suppose que le serveur peut tomber en panne lorsqu'il est actif (pannes actives).

3.1 Pannes passives

Les équations d'états sont de la forme suivante :

$$\begin{cases} \lambda P_0 = \mu P_1 + r P_2; \\ (\lambda + \theta) P_1 = \lambda P_0; \\ \theta P_1 = r P_2; \end{cases}$$

Alors, on obtient, $P_1 = \frac{\lambda}{\mu + \theta} P_0$, $P_2 = \frac{\lambda \theta}{r(\mu + \theta)} P_0$ et $P_0 = [1 + \frac{\lambda}{\mu + \theta} + \frac{\lambda \theta}{r(\mu + \theta)}]^{-1}$.

3.2 Pannes actives

Dans le cas où le serveur peut tomber en panne lorsqu'il est interactif, mais avec un taux de panne $\theta' < \theta$, nous aurons :

$$\begin{cases} (\lambda + \theta') P_0 = \mu P_1 + r P_2; \\ (\mu + \theta) P_1 = \lambda P_0; \\ r P_2 = \theta P_1 + \theta' P_0. \end{cases}$$

Ceci implique que : $P_1 = \frac{\lambda}{\mu + \theta} P_0$, $P_2 = \frac{1}{r} (\theta' + \frac{\lambda \theta}{\mu + \theta}) P_0$ et $P_0 = [1 + \frac{\lambda}{\mu + \theta} + \frac{\theta'}{r} + \frac{\lambda \theta}{r(\mu + \theta)}]^{-1}$

4 Système $M/G/1$ à serveur non fiable [1]

Considérons un système de file d'attente constitué d'un seul serveur sujet à des pannes aléatoires qui peuvent survenir aussi bien lorsque le serveur est libre, que lorsqu'il est occupé par le service d'un certain client. Ainsi, si à un instant t fixé le serveur s'avère être libre et en bon état, alors, la probabilité pour qu'une panne se produise durant un petit intervalle de temps $(t, t + \Delta t)$ de durée Δt vaut : $\theta_0 \Delta t + o(\Delta t)$ ($\theta_0 > 0$).

Ceci signifie que, lorsque le serveur est libre, la durée de bon fonctionnement du serveur est de loi exponentielle de fonction de répartition D_0 , telle que : $D_0(x) = 1 - e^{-\theta_0 x}$.

Si, par contre, à l'instant t , le serveur est occupé par le service d'un certain client, alors, la probabilité de panne durant $(t, t + \Delta t)$ est égale à : $\theta_1 \Delta t + o(\Delta t)$ ($\theta_1 > 0$).

La "réparation" de la panne débute immédiatement après la panne et se déroule indépendamment du processus des arrivées.

La durée de réparation est une variable aléatoire de fonction de répartition $R_0(x)$, de paramètre r_0 , (respectivement $R_1(x)$, de paramètre r_1).

Supposons qu'un client se présente à un instant fixé arbitrairement. Si à cet instant le serveur est disponible, alors le service débute immédiatement à l'instant $(t+0)$ et sa durée

est une variable aléatoire de loi arbitraire de fonction de répartition $H(x)$.

Si, par contre, à l'instant t le service est bloqué (le serveur est en panne ou occupé), alors le client rejoint la file d'attente et attend son tour. Admettons maintenant, que durant le service d'un certain client C , le serveur tombe en panne. Le service de client est ainsi interrompu durant un certain temps qui sera consacré à la réparation de la panne.

Maintenant, à chaque client, en plus de la durée de service τ , nous pouvons associer une variable aléatoire γ que nous appelons "durée de séjour" du client auprès de serveur. Cette variable aléatoire est égale à la durée de temps qui s'écoule entre l'instant où le serveur entame le service du client, jusqu'à l'instant où ce dernier quitte le système, son service étant achevé.

- Ainsi, pour les pannes de nature conservatrices, γ est constituée de la durée factice de service τ , à laquelle s'ajoutent les durées de réparation des pannes systématiques du serveur, survenues durant le temps τ .

Notons par,

$$h(s) = \mathbb{E}[e^{-s\tau}] = \int_0^{+\infty} e^{-sx} dP(\tau < x) \quad \text{et} \quad g(s) = \mathbb{E}[e^{-s\gamma}] = \int_0^{+\infty} e^{-sx} dP(\gamma < x);$$

Les transformées de Laplace-Stiltjes des fonctions de répartition des variables aléatoires τ et γ respectivement. On montre dans [9] que,

$$g(s) = h(s + \theta_1 - \theta_1(1 - r_1)).$$

- Pour les pannes de nature non conservatrices, il est nécessaire de tenir compte des durées potentielles de service interrompues par les pannes. Et par conséquent,

$$g(s) = h(s + \theta_1) \left[1 - \frac{\theta_1}{s + \theta_1} (1 - h(s + \theta_1)r_1(s)) \right]^{-1}.$$

- Pour l'autre type de pannes (pannes avec perte de client), la variable aléatoire γ admet une interprétation quelque peu différente et nous obtenons,

$$g(s) = h(s + \theta_1) + \frac{\theta_1}{s + \theta_1} r_1(s) [1 - h(s + \theta_1)].$$

La condition d'ergodicité d'un tel système est de la forme : $\rho = \lambda \int_0^{+\infty} x dH(\gamma < x) < 1$.

Conclusion

Dans cette communication, nous avons fait une synthèse des résultats connus sur les systèmes d'attente à serveurs non fiables. Cette revue montre les difficultés d'analyse des modèles d'attente avec serveurs non fiables.

Références

- [1] A. Aïssani and D. Aïssani. *Fiabilité des systèmes. Systèmes de files d'attente non fiables*. Bordj-El-Bahri, Alger, ENITA edition, 1986.

- [2] B. Avi-Itzhak and P. Naor. Some Queueing Problems with the service station subject to breakdown. *Oper. Res.*, 11 :3 :303–320, 1963.
- [3] H. Bruneel. A general model for the behaviour of infinite buffers with periodic service opportunities. *Euro. J. Oper. Res.*, 16 :98–106, 1984.
- [4] H. Bruneel and K. G. Byung. Discrete-Time Models for Communication Systems Including ATM. *Kluwer Academic Publishers*, 1993.
- [5] B. D. Choi and J. Chung. MAP/G/1 queue with a single server subject to breakdowns and repairs. *Korea Advanced Institute of Science and Technology and Center for Applied Mathematics*, pages 1–33, 1995.
- [6] L. Duan-Shin. Analysis of a single server queue with semi-Markovian service interruption. *Queueing Systems*, 27 :153–178, 1997.
- [7] D. Fakinos. The $M/G/1/k$ loss system with server subject to breakdowns. *Journal of Applied Probability*, 20 :706–712, 1983.
- [8] A. Federgruen and L. Green. Queueing system with Service Interruptions II. *Naval Research Logistics*, 35 :345–358, 1988.
- [9] D. P. Gaver. A waiting Line with Interrupted Service, including Priorities. *Roy. Stat. Soc. J*, B25 :73–90, 1962.
- [10] K. Jau-Chuan. Two Thresholds of an $M/G/1$ Queueing System with Server Breakdowns and Two Vacation Types. *Sanmin Rd. , Taichung 404, Taiwan, R. O. C. , National Taichung Institute of Technology*, Sec. 3(129), 1999.
- [11] V. G. Kulkarni and B. D. Choi. Retrial Queues with server subject to breakdowns and repairs. *Queueing Systems*, 7 :191–208, 1990.
- [12] K. C. Madan. A single channel queue with bulk service subject to interruptions. *Microelectron. Reliab.*, 29 :5 :813–818, 1989.
- [13] F. Marcel, Neuts, and D. M. Lucantoni. A Markovien queue with N servers subject to breakdowns and repairs. *Mgmt. Sci.*, 25 :849–861, 1979.
- [14] L. L. Mitrani and B. Avi-Itzhak. A many server queue with service interruptions. *Oper. Res.*, 16 :628–638, 1968.
- [15] T. Takine and B. Sengupta. A single server queue with service interruption. *Queueing System*, 26 :285–300, 1997.
- [16] T. Takine and H. Toshiharu. The workload in the MAP/G/1 queue with state-dependent queue : Its application to a queue with preemptive resume priority. *Commun. Statist.-Stochastic Models*, 10 :1 :183–204, 1994.
- [17] N. B. Vasanthakumar. Switched Poisson Process/G/1 queue with service interruptions. *Computers and Ops. Res.*, 19 :8 :751–756, 1992.
- [18] H. C. White and L. S. Christie. Queueing with preemptive priorities or with breakdown. *Oper. Res.*, 6 :79–95, 1958.

Approximation par la Théorie de l'Information d'une File M/G/1 avec Rappels et Serveur non Fiable

Smail R., Aissani A.,
Université des Sciences & Technologie (USTHB),
BP 32 El Alia, Bab-Ez Zouar, 16111,
Algérie.
E-mail: aaissaniwissal.dz

Mots-clés

Files d'attente, Pannes, Approximation, Théorie de l'information, Entropie.

Résumé

Ce travail est consacré à l'étude de la file M/G/1 avec rappels lorsque le serveur est sujet à des pannes aléatoires. La théorie des files d'attente avec rappels développée ces dernières années fournit un cadre conventionnel pour résoudre de tels modèles. Les mesures de performance de tels systèmes sont disponibles sous forme explicite. Ici nous donnons des approximations plus simples obtenues à l'aide du formalisme du maximum d'entropie. Ces approximations sont obtenues à partir de l'information disponible sur les distributions de probabilité des temps de service et de réparation, ainsi que de la distribution stationnaire des états du système.

1. Introduction

Les files avec rappels [2,4,6] ont été utilisées dans plusieurs applications concrètes, comme la modélisation des réseaux cellulaires mobiles numériques (Digital Cellular Mobile Networks) [10] et les réseaux locaux à topologie étoilée [9]. Cependant la plupart des modèles étudiés négligent le processus de pannes. On considère une file M/G/1 avec rappels lorsque le serveur est sujet à des pannes aléatoires. Ce modèle a été étudié auparavant en utilisant diverses méthodes analytiques (chaînes de Markov incluse, variable supplémentaire, approximation de diffusion ...), et ses mesures de performance sont disponibles sous forme explicite [1,8]. Le lecteur intéressé peut se référer à d'autres variantes de notre modèle [3,11]. Les solutions sont obtenues en termes de transformées de Laplace et fonctions génératrices. Pour les praticiens, il est intéressant d'obtenir des estimations plus simples. Dans cet article, on utilise la théorie de l'information pour obtenir des approximations basées sur l'information disponibles concernant les distributions de probabilité des temps de service et de réparation.

La section suivante est consacrée à la description du modèle. Dans la section 3 on présente le formalisme du maximum d'entropie. La section 4 concerne l'approximation de premier ordre pour laquelle une solution explicite peut-être obtenue. Enfin, dans la section 5, on discute l'approximation de second ordre. On conclut cette étude par quelques exemples numériques montrant la qualité des approximations obtenues.

2. File M/G/1 avec Rappels et Serveur non Fiable

Les clients arrivent dans la file suivant un processus de Poisson de taux $\lambda > 0$. Le serveur est sujet à deux types d'interruptions dues à des pannes. Les pannes indépendantes (ou passives) se produisent lorsque le serveur est oisif et forment un processus de Poisson de taux η ; les pannes actives forment aussi un processus de Poisson de taux μ . Les durées des interruptions indépendantes et actives sont des variables aléatoires D_i et D_b de distributions de probabilité $G(x)$ et $H(x)$, et transformées de Laplace-Stieltjes $\tilde{G}(s)$ et $\tilde{H}(s)$. Si un client arrivant trouve le serveur disponible, son service commence immédiatement durant un temps S suivant une distribution arbitraire $F(x)$ de transformée de Laplace-Stieltjes $\tilde{F}(s)$. Autrement, si le client arrivant trouve le serveur occupé, il rejoint un groupe de rappels (appelé aussi « orbite », qui est une sorte de file) et répète sa demande après un temps aléatoire distribué exponentiellement de taux θ . Finalement, le client dont le service est interrompu par une panne active doit décider, soit de rejoindre l'orbite (avec une probabilité c , appelée aussi « facteur de recouvrement ») ou quitter le système (avec une probabilité $1-c$). On suppose également l'indépendance mutuelle des variables aléatoires introduites. Le système considéré peut-être décrit par un processus stochastique discret $\{Q(t), X(t)\}$ où $Q(t)$ est le nombre de clients dans l'orbite au temps t et

$X(t)=0$ si le serveur est opérationnel et libre.
 $=1$ si le serveur est opérationnel et occupé.
 $=2$ si le serveur est en panne (indépendante).
 $=3$ si le serveur est en panne (active).

Il est connu que la distribution stationnaire des du système [1,8]

$$P_{ij} = \lim_{t \rightarrow \infty} P(X(t)=i, Q(t)=j),$$

$i \in \{0,1,2,3\}$, $j \in \{0,1,2,\dots\}$ existe si

$$\rho = \lambda \frac{1 - \tilde{F}(\mu)}{\mu} \left(1 + \mu (E(D_i) + \frac{c}{\mu}) \right) < 1.$$

3. Formalisme du Maximum d'Entropie

Soit Q un système à espace des états discrets $S = \{S_n\}$ tel que l'information disponible sur Q impose un certain nombre de contraintes sur la distribution de probabilité $P = \{p(S_n)\}$. On suppose que ces contraintes prennent la forme de valeurs moyennes de m fonctions $\{f_k(S_n)\}_{k=1,m}$, où $m < \text{card}(S)$. Le principe du maximum d'entropie (PME) [5,7] stipule que, parmi toutes les distributions satisfaisant les contraintes fournies par l'information donnée, la distribution la moins préjudiciable est celle qui maximise la fonction d'entropie du système:

$$H(p) = - \sum_{S_n \in S} p(S_n) \log p(S_n) \quad (1)$$

soumise aux contraintes:

$$\sum_{S_n \in S} p(S_n) = 1 \quad (2)$$

$$\sum_{S_n \in S} f_k(S_n) p(S_n) = \langle f_k \rangle, k=1, \dots, m \quad (3)$$

où $\langle f_k \rangle$ sont les valeurs moyennes prescrites, définies sur l'ensemble des fonctions $\{f_k(S_n)\}_{k=1,m}$. Le problème de maximisation d'entropie (1) sous les contraintes (2) et (3) peut-être résolu en utilisant la méthode des multiplicateurs de Lagrange, ce qui conduit à la solution:

$$p(S_n) = \frac{1}{Z} \exp \left(- \sum_{k=1}^m \beta_k f_k(S_n) \right) \quad (4)$$

où $\{\beta_k\}_{k=1,m}$ sont les multiplicateurs de Lagrange déterminés à partir de l'ensemble des contraintes (3). La fonction Z est connue en physique statistique sous le nom de "fonction de partition", et est donnée par:

$$Z = \exp(\beta_0) = \sum_{S_n \in S} \exp \left(- \sum_{k=1}^m \beta_k f_k(S_n) \right) \quad (5)$$

où β_0 est le multiplicateur de Lagrange correspondant à la contrainte de normalisation (2).

4. Approximation de premier ordre

Revenons maintenant à notre file M/G/1 avec rappels et serveur sujet à des pannes décrite dans la section 2. Dans [1], on a trouvé des formules explicites pour les moments partiels

$$M_i^k = \sum_{j=0}^{\infty} j^k P_{ij}, \text{ et qui peuvent être utilisés comme}$$

information a priori. Si l'unique information disponible est donnée par $M_i^0, M_i^1, i=0,1,2,3$, alors selon le PME, l'approximation de la distribution de probabilités de l'état du système est:

$$\hat{P}_{in} = \varphi_i \frac{\alpha_i}{\alpha_i + \beta_i} \left(\frac{\beta_i}{\alpha_i + \beta_i} \right)^n, i=0,1,2,3 \text{ et } n \geq 0.$$

où les constantes $\varphi_i, \alpha_i, \beta_i$ sont données en annexe.

$$\varphi_0 = \frac{K}{A}, \alpha_0 = \theta(1-\rho), \beta_0 = \lambda(\rho + \eta E(D_i));$$

$$\varphi_1 = \frac{\alpha_1}{A\mu}, \alpha_1 = (1-\eta K)(1-\tilde{F}(\mu)), \beta_1 = A\mu M_1^1;$$

$$\varphi_2 = \eta(M_0^0 E(D_i)), \alpha_2 = M_0^0 E(D_i),$$

$$\beta_2 = \frac{\lambda}{2} M_0^0 E(D_i^2) + M_0^1 E(D_i);$$

$$\varphi_3 = \mu M_1^0 E(D_b), \alpha_3 = M_1^0 E(D_b),$$

$$\beta_3 = M_1^0 (c E(D_b) + \frac{\lambda}{2} E(D_b^2)) + M_1^1 E(D_b)$$

$$K = \frac{1-\rho}{\lambda(1+\eta E(D_i)) + \eta(1-\rho)}$$

$$A = K(1+\eta E(D_i)) + (1+\tilde{F}(\mu))(1-\eta K) \left(E(D_b) + \frac{1}{\mu} \right)$$

5. Approximation de Second Ordre

Dans cette section, on résout le problème avec une information supplémentaire sur le moment de second ordre M_i^2 donné dans [1]. Dans ce cas, il n'est pas possible d'avoir une solution explicite des paramètres de Lagrange. Cependant, on peut obtenir une solution numérique. Dans notre cas, la maximisation de la fonction d'entropie

$$H(P_i) = - \sum_{n=0}^{\infty} P_{in} \log(P_{in}), i \in \{0,1,2,3\}$$

soumise aux contraintes des moments partiels $M_i^k, i \in \{0,1,2,3\}, k \in \{0,1,2\}$, mène à une solution de la forme

$$\hat{P}_{in} = z_i^{-1} \exp(-\beta_i \{n - \beta_2^i n^2\})$$

$$Z_i = \frac{1}{M_i^0} \sum_{n=0}^{\infty} \exp(-\beta_i \{n - \beta_2^i n^2\})$$

où β_1^i et β_2^i sont les multiplicateurs de Lagrange correspondants respectivement aux contraintes M_i^1 et M_i^2 $i \in \{0, 1, 2, 3\}$.

En utilisant les méthodes standards d'analyse numérique la détermination des paramètres β_1^i et β_2^i se réduit à trouver le minimum d'une fonction potentielle :

$$F_i(\beta_1^i, \beta_2^i) = \log \sum_{k=0}^N \exp \left\{ - \sum_{j=1}^2 \beta_j^i \left(k^j - \frac{M_i^j}{M_i^0} \right) \right\},$$

$i \in \{0, 1, 2, 3\}$, où N est la taille tronquée de l'orbite.

6. Tests numériques

Dans cette section on montre la validité de la méthode proposée à l'aide d'expériences numériques typiques, présentées dans les tables 1 et 2. Des comparaisons relatives ont été effectuées entre les approximations du maximum d'entropie et les valeurs exactes de la file M/M/1 avec rappels et pannes pour différentes distributions du temps de réparation : (i) Exponentielle, (ii) Gamma (iii) Hyperexponentielle. Pour chacun de ces choix, on fait varier les paramètres d'entrée. On utilise les données de base de [4] dans le but de tester la précision des calculs. Des expériences similaires ont été réalisées dans le cas d'une distribution du temps de service hyperexponentielle. Pour des distributions du temps de réparation indépendante et active exponentielles de taux η_1 et μ_1 respectivement, la table 1 montre que l'approximation de premier ordre est décroissante (voir graphe 1). Toutes nos expériences montrent que la performance de l'approximation est améliorée par l'information supplémentaire donnée par le moment de second ordre (voir graphe 2). Cependant, la performance de l'approximation de second ordre se dégrade lorsque le taux de rappels θ décroît. Cela est aussi indiqué dans la dernière ligne des tables où nous avons calculé l'entropie de Shannon qui diminue avec le nombre de moments utilisés comme information a priori.

7. Conclusion

A travers cette étude, nous avons présenté quelques approximations des distributions d'états du système par la théorie de l'information dans une file M/G/1 avec rappels et serveur non fiable.

Puisque l'entropie maximum correspond au désordre maximum dans le système, alors les solutions obtenues sont les moins biaisées parmi

toutes les solutions qui satisfont les contraintes du système.

Pour les autres distributions de performance, comme le temps d'attente, temps résiduel,...etc., l'approche entropique donne aussi des distributions de probabilités approchées. La généralisation de nos résultats sous une telle forme à des files avec rappels plus complexes, présente des difficultés, vu qu'on ne connaît pas en général les moments de la distribution d'états stable. Néanmoins, cette étude pourrait être étendue à des files d'attente avec pannes plus complexes : rappels ou d'inter-arrivées de lois générales, cas de plusieurs serveurs ou de phénomènes particuliers (arrivées négatives,...). Il serait nécessaire dans ce cas d'exploiter d'autres sources d'informations, fournies par exemple par des techniques asymptotiques.

8. Références

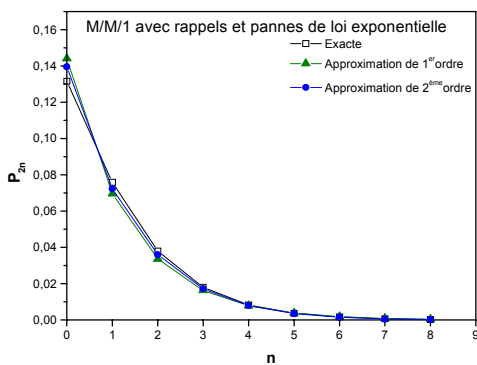
- [1] Aissani A. & Artalejo J.R. (1998), "On the single server retrial queue subject to breakdowns", *Queueing Syst.* 30, 309-321.
- [2] Artalejo, J.R. (1999), "A classified bibliography of research on retrial queues: Progress in 1990-1999", (*TOP*), vol. 7, N°2, pp.187-211.
- [3] Artalejo J.R., Gomez-Corral A. (1998), Unreliable retrial queues due to service interruptions arising from facsimile networks, *Belgian J. of Stat., Operat. research and Computer Sci.* vol.38 (1).
- [4] Falin G.I., (1990), A survey of retrial queues, *Queueing systems* 7, 127-167.
- [5] Falin G.I., Martin Diaz M., Artalejo J.R., (1994), Information theoretic approximations for the M/G/1 retrial queue, *Acta Informatica*, 559-571.
- [6] Falin G.I., Templeton J.G.C., (1997), *Retrial queues*, Chapman and Hall, London.
- [7] Kouvatsos D.D., Tabet-Aouel N., (1989), A maximum entropy priority approximation for a stable G/G/1 queue, *Acta Informatica* 27, 247-286.
- [8] Kulkarni & Choi V.G., Choi B.D. (1990), Retrial queue with server subject to breakdown and repairs, *Queueing systems*, 7, 191-208.
- [9] Janssens G. K. (1997), "The quasi-random input queueing system with repeated attempts as a model for a collision-avoidance star local area network", *IEEE Trans. on Commun.*; vol. 45, N°3, 361-364.
- [10] Sun Jong Kwon & al, (2001), "Performance analysis of CDPD Sleep Mode for power conservation in Mobile End Systems", *IEIC Trans. Commun.*, vol. E84, N°10.
- [11] Wang J., Cao J. (2001), Reliability analysis of the retrial queue with server breakdowns and repairs, *Queueing Systems*, 38, 363-380.

Table 1. Comparaison entre les résultats exacts et approximatifs d'une file $M/M/1$ avec rappels et pannes tels que $\lambda = 2, \nu = 8, \theta = 12, \eta = 0.2, \mu = 0.4, \eta_1 = 3, \mu_1 = 6$ et $c = 0.5$

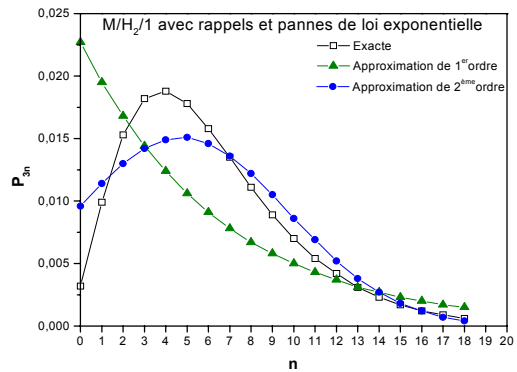
| Distribution du temps de réparation Exponentielle | | | | | | | | | | | | |
|---|----------|----------|----------|----------|------------------|------------------|------------------|------------------|------------------|------------------|------------------|------------------|
| n | P_{0n} | P_{1n} | P_{2n} | P_{3n} | \hat{P}_{0n}^1 | \hat{P}_{1n}^1 | \hat{P}_{2n}^1 | \hat{P}_{3n}^1 | \hat{P}_{0n}^2 | \hat{P}_{1n}^2 | \hat{P}_{2n}^2 | \hat{P}_{3n}^2 |
| 0 | 0.3290 | 0.0969 | 0.1316 | 0.0242 | 0.3298 | 0.0927 | 0.1441 | 0.0434 | 0.3325 | 0.0909 | 0.1396 | 0.0294 |
| 1 | 0.0583 | 0.0550 | 0.0759 | 0.0440 | 0.0697 | 0.0454 | 0.0696 | 0.0278 | 0.0668 | 0.0464 | 0.0723 | 0.0304 |
| 2 | 0.0191 | 0.0288 | 0.0380 | 0.0319 | 0.0147 | 0.0223 | 0.0336 | 0.0179 | 0.0144 | 0.0232 | 0.0360 | 0.0258 |
| 3 | 0.0070 | 0.0145 | 0.0180 | 0.0188 | 0.0031 | 0.0109 | 0.0163 | 0.0115 | 0.0033 | 0.0113 | 0.0172 | 0.0179 |
| 4 | 0.0027 | 0.0071 | 0.0083 | 0.0101 | 0.0007 | 0.0054 | 0.0079 | 0.0074 | 0.0008 | 0.0054 | 0.0079 | 0.0103 |
| 5 | 0.0010 | 0.0034 | 0.0037 | 0.0051 | 0.0001 | 0.0026 | 0.0038 | 0.0047 | 0.0002 | 0.0025 | 0.0035 | 0.0048 |
| 6 | 0.0004 | 0.0016 | 0.0017 | 0.0025 | 0.0000 | 0.0013 | 0.0018 | 0.0030 | 0.0001 | 0.0012 | 0.0015 | 0.0019 |
| 7 | 0.0001 | 0.0007 | 0.0007 | 0.0012 | 0.0000 | 0.0006 | 0.0009 | 0.0020 | 0.0000 | 0.0005 | 0.0006 | 0.0006 |
| 8 | 0.0000 | 0.0003 | 0.0003 | 0.0005 | 0.0000 | 0.0003 | 0.0004 | 0.0013 | 0.0000 | 0.0002 | 0.0002 | 0.0002 |
| Entropie de Shannon | 2.3589 | | | 2.4016 | | | | 2.3885 | | | | |

Table 2. Comparaison entre les résultats exacts et approximatifs d'une file $M/H_2/1$ avec rappels et pannes tels que, $\lambda = 1, E(S) = 0.25, \theta = 4, \eta = 0.5, \mu = 2, E(D_i) = E(D_o) = 0.25$ et $c = 0.8$.

| Distribution du temps de réparation Hyperexponentielle $f(x)=0.5e^{-2x}+4.5e^{-6x}$ | | | | | | | | | | | | |
|---|----------|----------|----------|----------|------------------|------------------|------------------|------------------|------------------|------------------|------------------|------------------|
| n | P_{0n} | P_{1n} | P_{2n} | P_{3n} | \hat{P}_{0n}^1 | \hat{P}_{1n}^1 | \hat{P}_{2n}^1 | \hat{P}_{3n}^1 | \hat{P}_{0n}^2 | \hat{P}_{1n}^2 | \hat{P}_{2n}^2 | \hat{P}_{3n}^2 |
| 0 | 0.5195 | 0.1179 | 0.0494 | 0.0090 | 0.5473 | 0.1068 | 0.0534 | 0.0318 | 0.5502 | 0.1073 | 0.0538 | 0.0165 |
| 1 | 0.0912 | 0.0574 | 0.0198 | 0.0423 | 0.0870 | 0.0429 | 0.0183 | 0.0205 | 0.0837 | 0.0426 | 0.0180 | 0.0237 |
| 2 | 0.0264 | 0.0258 | 0.0074 | 0.0291 | 0.0138 | 0.0172 | 0.0063 | 0.0132 | 0.0138 | 0.0170 | 0.0062 | 0.0233 |
| 3 | 0.0086 | 0.0115 | 0.0027 | 0.0157 | 0.0022 | 0.0069 | 0.0022 | 0.0085 | 0.0025 | 0.0069 | 0.0022 | 0.0157 |
| 4 | 0.0030 | 0.0050 | 0.0010 | 0.0078 | 0.0003 | 0.0028 | 0.0007 | 0.0055 | 0.0005 | 0.0028 | 0.0008 | 0.0072 |
| 5 | 0.0011 | 0.0022 | 0.0004 | 0.0037 | 0.0001 | 0.0011 | 0.0003 | 0.0035 | 0.0001 | 0.0012 | 0.0003 | 0.0023 |
| 6 | 0.0004 | 0.0009 | 0.0001 | 0.0017 | 0.0000 | 0.0004 | 0.0001 | 0.0023 | 0.0000 | 0.0005 | 0.0001 | 0.0019 |
| Entropie de Shannon | 1.7558 | | | 1.7889 | | | | 1.7712 | | | | |



Graph 1. distributions exacte et approximatifs d'états du système $M/M/1$ avec rappels et pannes tels que, $\lambda=2, \nu=8, \theta=12, \eta=2, \mu=4, \eta_1=3, \mu_1=6$ et $c=0.5$.



Graph 2. distributions exacte et approximatifs d'états du système $M/H_2/1$ avec rappels et pannes tels que, $\lambda=1, E(S)=2/3, \theta=1, \eta=0.5, \mu=1, \eta_1=2, \mu_1=3$ et $c=0.5$.

**NOMBRE MOYEN DE CLIENTS DANS DEUX FILES PARALLELES AVEC FLUX
D'ARRIVEE PARTAGE**

Tayeb LARDJANE *

Résumé: On considère deux files d'attente parallèles de discipline d'arrivées et de service exponentielle. On suppose qu'en plus des sources d'arrivée de chacune des deux files existe une autre source à partir de laquelle chacune des deux files peut recevoir des clients si sa taille est plus petite que l'autre. On donne alors dans un cas particulier la taille moyenne en régime stationnaire de chacune des deux files.

Mots clé: files parallèles, processus de régénération, période d'activité.

Abstract: We consider two markovian parallel queues. We suppose that a third source of customers exist and can finish each one of the queues in case of its size is small than an other. We give in a particular case the stationary mean size of each queue.

Key words: parallel queues, regenerative process , busy period.

* Université des sciences et de la technologie Houari Boumedienne Alger

I. Introduction:

On considère deux files d'attente markoviennes parallèles numérotées 1 et 2 de taux respectifs d'arrivée λ_1 et λ_2 et de taux de service μ_1 et μ_2 . On suppose que chacune des deux files et dans le cas où sa taille est plus petite que l'autre peut recevoir des clients avec un taux supplémentaire λ . Le but de cette étude est de donner en régime stationnaire le nombre moyen de clients dans chaque file.

II. Equations de Chapman-Kolmogorov :

Pour tout instant t, soit X_t et Y_t les tailles des files 1 et 2 respectivement à l'instant t. On étudie alors le comportement de la file 1 et il en découlera de même pour la file 2.

Pour tout n, soit $p_n(t) = \text{Prob}(X_t = n)$. On a alors:

$$p_n(t+h) = p_n(t)(1 - \lambda_1 h - \mu_1 h - \lambda \text{Prob}(X_t \leq Y_t)) + p_{n+1}(t)(\mu_1 h) + p_{n-1}(t)(\lambda_1 + \lambda \text{Prob}(X_t \leq Y_t))h$$

C'est alors une équation du type file M/M/1 sauf que $\text{Prob}(X_t \leq Y_t)$ est inconnu.. On utilisera alors la théorie des processus de régénération pour étudier son approximation pour t assez grand.

III. Processus de régénération:

Théorème:

On suppose $\lambda_1 = \lambda_2$ et $\mu_1 = \mu_2$, pour tout t soit $Z_t = X_t - Y_t$; alors $(Z_t)_t$ est un processus de régénération d'instant

de régénération la suite $(T_n)_n$ définie par : $T_0 = 0, T_n = \inf\{t > s; X_t = Y_t\}$ ou $s = \inf\{u > T_{n-1} \mid X_u \neq Y_u\}$

Preuve:

On définit $S_n = \sum_{i=1}^{i=n} T_i$

- a) $(T_n)_n$ est bien un processus de renouvellement(en éliminant le cas $X_t = Y_t = 0$) car avec l'égalité des taux des deux files, l'occurrence de l'égalité des tailles des deux files entraîne la remise à l'état initial du processus décrivant le système.
- b) L'égalité $E[f(Z_{S_m+T_1}, \dots, Z_{S_m+T_{n_1}}) / Z_u; u \leq S_m] = E[f(Z_{t_1}, \dots, Z_{t_n})]$ est assurée par le caractère markovien des deux files d'attente.

Soit alors $A =]-\infty, 0[$ et on définit $K(t,A) = P(S_1 > t, Z_t \in A)$, on veut alors étudier le comportement asymptotique de $f(t) = P(Z_t \in A)$. En supposant que le processus de renouvellement $(T_n)_n$ est récurrent apériodique alors

$$\lim_{t \rightarrow \infty} f(t) = \frac{1}{\beta} \int_0^{+\infty} K(s, A) ds \quad \text{ou}$$

$$\beta = E(T_1)$$

Calcul de K(t, A):

Par définition $K(t,A) = P(S_1 > t, X_t - Y_t < 0)$; au fait sur l'événement $\{S_1 > t\}$ le calcul de $K(t,A)$ se réduit au calcul de $P(X_t - Y_t < 0)$ qui se fera dans ce cas comme si les deux files

(M/M/1) sont complètement indépendantes avec comme taux d'arrivée $\lambda + \lambda_1$ et λ_1 pour les files 1 et 2 respectivement et les taux de service sont inchangés.

D'où $P(X_t < Y_t) = \sum_i P(X_t < Y_t / Y_t = i)P(Y_t = i) = \sum_i P(X_t < i)P(Y_t = i)$

$$\sum_i \left(\sum_{j=0}^{i-1} e^{-\left((\lambda+\lambda_1)^{\frac{1}{2}} - (\mu_1)^{\frac{1}{2}}\right)^2 t} \times \frac{\mu_1 \binom{-j-\frac{1}{2}}{-j-\frac{1}{2}} \frac{1}{2}}{(4\pi t)^{\frac{1}{2}} \lambda^{\frac{-j-\frac{1}{2}}{2}}} + \frac{e^{-\left((\lambda+\lambda_1)^{\frac{1}{2}} - (\mu_1)^{\frac{1}{2}}\right)^2 t} \mu^{\binom{-j+\frac{1}{2}}{-j+\frac{1}{2}}}}{\lambda^{\frac{-j+\frac{3}{2}}{2}}} + e^{-(\lambda+\lambda_1+\mu)t} \left(1 - \frac{\lambda + \lambda_1}{\mu_1}\right) \left(\frac{\lambda + \lambda_1}{\mu_1}\right)^i \times B \right)$$

$$\times e^{-(\sqrt{\lambda_1} - \sqrt{\mu_1})^2 t} \times \frac{1}{\sqrt{4\pi t}} \frac{\mu_1 \binom{-i-\frac{1}{2}}{-i-\frac{1}{2}} \frac{1}{2}}{\lambda_1 \binom{-i+\frac{1}{2}}{-i+\frac{1}{2}} \frac{1}{2}} + e^{-(\sqrt{\lambda} - \sqrt{\mu})^2 t} \frac{\mu_1 \binom{-i+\frac{1}{2}}{-i+\frac{1}{2}} \frac{1}{2}}{\lambda_1 \binom{-i+\frac{3}{2}}{-i+\frac{3}{2}} \frac{1}{2}} + e^{-(\lambda + \mu)t} \left(1 - \frac{\lambda_1}{\mu_1}\right) \left(\frac{\lambda_1}{\mu_1}\right)^i \times C$$

ou:

$$B = \sum_{l=j+2}^{\infty} \left[\left(\frac{\mu_1}{\lambda_1 + \mu_1} \right)^{\frac{l}{2}} \sum_{k=0}^{\infty} \frac{\left(\sqrt{\mu_1 (\lambda + \lambda_1) t} \right)^{l+2k}}{k!(l+k)!} \right] \text{ et}$$

$$C = \sum_{l=i+2}^{\infty} \left[\left(\frac{\mu_1}{\lambda_1} \right)^{\frac{l}{2}} \sum_{k=0}^{\infty} \frac{\left(\sqrt{\lambda_1 \mu_1 t} \right)^{l+2k}}{k!(l+k)!} \right]$$

On a supposé que $X_0 = Y_0 = 1$ (pour des commodités de calculs on ne fera pas de différence avec le cas $X_0 = Y_0 = 0$) et que dans ce cas le taux d'arrivée dans la file 1 est $\lambda_1 + \lambda$ et le taux d'arrivée dans la file 2 est λ_1 .

Calcul de β :

La durée d'une période de renouvellement est la convoleraï d'une exponentielle négative de taux $\lambda + 2\lambda_1 + 2\mu_1$ et d'une variable aléatoire Z ayant pour distribution la distribution de la période d'activité d'un serveur d'une file M/M/1 de taux d'arrivée $\lambda_1 + \mu_1$ et de taux de service $\lambda + \lambda_1 + \mu_1$ (sa taille à l'instant t est $Y_t - X_t$)

Calcul de $\lim_{t \rightarrow \infty} E(X_t)$:

Les équations de Chapman-Kolmogorov écrites en régime stationnaire s'écrivent:

$p_{n+1}(\lambda_1 + \mu_1 + \lambda\delta) = \mu_1 p_{n+1} + (\lambda_1 + \lambda\delta)p_{n-1}$ et qui donne alors :

$$\lim_{t \rightarrow \infty} E(X_t) = \frac{\lambda_1 + \lambda\delta}{\mu_1 - (\lambda_1 + \lambda\delta)} \text{ ou } \delta = \lim_{t \rightarrow \infty} P(X_t \leq Y_t)$$

Conclusions :

-Il y a lieu de relever le caractère restrictif de certaines hypothèses telles que les hypothèses d'égalités des taux des deux files. Cette hypothèse qui fait presque de la suite $(T_n)_n$ un processus de renouvellement peut être levée dans une étude ultérieure en considérant les deux processus de renouvellement qu'elle induit.

- Les conditions d'ergodicité du système sont supposées satisfaites, il suffit par exemple de supposer $\max(\lambda + \lambda_1, \lambda + \lambda_2) < \min(\mu_1, \mu_2)$.

- Si on pose $A(t) = P(X_t \leq Y_t)$ on obtient :

$$A(t) = \int_0^t P(X_t \leq Y_t / S_1 = x) dF(x) + \int_t^\infty P(X_t \leq Y_t / S_1 = x) dF(x)$$

$$= \int_0^t A(t-x) dF(x) + \int_t^\infty P(X_t \leq Y_t / S_1 = x) dF(x)$$

qui est une équation de renouvellement facilement soluble dans le cas où la distribution F de S_1 est facilement calculable.

Bibliographie:

1. Cox D.R Renewal theory 1962
2. Karlin S. A first course in stochastic processes, Academic Press, N.Y. 1974
3. Kleinrock L. Queuing systems, volumes 1& 2 , John Willey & Sons, 1974.
4. Sheldon M. Ross; Introduction to probability models, Academic press, Inc 1985.
5. Cinlar, E. Stochastic processe, Englewood cliffs, N. J (1975)

CARACTÉRISATION VARIATIONNELLE DES PROCESSUS DE SCHRÖDINGER AVEC POTENTIEL NON BORNÉ.

A. BENCHETTAH

Université d'Annaba, Faculté des sciences,
Département de Mathématiques,
BP 12, Annaba 23000 Email: abenchettah@hotmail.com

Abstract. Etant donné une classe de processus de Schrödinger de naissance et de mort avec des potentiels non bornés ou singuliers. Une variante du résultat de Jamison est établie sans les hypothèses de continuité et de la stricte positivité partout de q , à l'aide des méthodes de la théorie des grandes déviations. En effet, nous associons les données de Jamison $(q; P_a; P_b)$ avec la projection de Csiszar Q^2 d'une mesure de référence R^2 sur $E(P_a; P_b)$, un ensemble de mesure de probabilités de marginales $P_a; P_b$. Nous donnons des conditions suffisantes d'existence d'une solution du système de Schrödinger pour une certaine fonction q donnée explicitement vérifiant les hypothèses du résultat précédent et nous montrons que cette fonction q correspond à une densité de quasi-transition au sens de Dynkin. Nous généralisons alors la caractérisation variationnelle des processus de Schrödinger introduite par Wakolbinger pour une plus grande classe de potentiel.

AMS classification [2000]: 49L20, 60J60, 93E20

Mots clés: Equations différentielles stochastiques, contrôle optimale stochastique, processus de Markov, réciproques et de Schrödinger, entropie relative, projection de Csiszar.

1. Introduction

Dans ses papiers remarquables de 1931/1932 [18], [19], Schrödinger a résolu le problème suivant: "sachant la position d'une particule brownienne dans un espace euclidien M aux instants a et b , $a < b$, quelle est la probabilité pour cette particule d'être dans un certain domaine aux instants intermédiaires?"

Une généralisation de ce problème a été amené au concept de pont de Schrödinger et étudié comme problème d'optimisation convexe par différentes voies:

- la théorie des processus réciproques: Bernstein, Beurling, Fortet et Jamison;
- la théorie de l'information et les statistiques où l'on utilise le concept de l'entropie: Kullback et Csiszar [4], [5];
- la théorie des grandes déviations pour des distributions empiriques sur l'espace des trajectoires: Dawson [6], Föllmer [11], Nagasawa [14], [15] et Wakolbinger [22];
- l'approche variationnelle (problème de contrôle optimal stochastique et problème de calcul de variation stochastique): Benchettah [1], [2], Blaquière [3], Nagasawa, Wakolbinger, Yasue.

Le cas traité par Schrödinger est le suivant:

$- = -_0 C^0[a; b]; R^3$ modélisant les trajectoires d'une particule brownienne, $F = M$ la tribu borélienne sur $-_0, 1$ la mesure de Wiener et E l'ensemble des mesures de probabilité sur $(-_0; M)$ de marginales P_a, P_b .

L'existence d'une solution du problème de Schrödinger a été ramenée par Schrödinger à l'existence d'une solution d'un système non linéaire de deux équations intégrales, dit système de Schrödinger noté (S). Notons que la donnée de la mesure $\mathbb{1}$ est liée à la densité de probabilité de transition du mouvement brownien. Cependant le problème de Schrödinger peut être formulé avec une densité de transition noncomplète $q(s;x;t;y)$, $a \cdot s < t \cdot b$. Aussi nous avons deux approches: l'une consiste à discuter l'existence en terme de q (basée sur la théorie des processus réciproques), et l'autre en terme de $\mathbb{1}$ (basée sur l'approche géométrique de Csiszar) qui sont liées.

2. Système de Schrödinger pour un processus de naissance et de mort

Théorème 1. Soient M un espace métrique \mathcal{M} compact, \mathcal{F} une tribu borélienne de M , $\mathbb{P}_a, \mathbb{P}_b$ deux mesures de probabilité sur \mathcal{F} , \mathbb{E} un ensemble de probabilités \mathbb{P} sur $\mathcal{F} - \mathcal{F}$ tel que $\mathbb{P}(\cdot; M) = \mathbb{P}_a(\cdot)$; $\mathbb{P}(M; \cdot) = \mathbb{P}_b(\cdot)$ et une fonction borélienne $q: M \times M \rightarrow \mathbb{R}$ minorée au dessus de zéro $(\mathbb{P}_a - \mathbb{P}_b)$ j.p.s. et $(\mathbb{P}_a - \mathbb{P}_b)$ j.intégrable. Alors il existe une paire de mesures $(\mathbb{Q}^{\mathbb{P}}; \mathcal{M})$ sur $\mathcal{F} - \mathcal{F}$ telle que

- (a) $\mathbb{Q}^{\mathbb{P}}$ est une probabilité et \mathcal{M} est une mesure produit;
- (b) $\mathbb{Q}^{\mathbb{P}} \in \mathbb{E}$;
- (c) $\frac{d\mathbb{Q}^{\mathbb{P}}}{d\mathcal{M}} = q$;
- (d) $H(\mathbb{Q}^{\mathbb{P}}; \mathbb{R}^{\mathbb{P}}) \cdot H(\mathbb{P}; \mathbb{R}^{\mathbb{P}}) \in \mathbb{E}$ où $\mathbb{R}^{\mathbb{P}}$ est définie par:

$$(1) \quad d\mathbb{R}^{\mathbb{P}} = q d(\mathbb{P}_a - \mathbb{P}_b) = q d(\mathbb{P}_a - \mathbb{P}_b);$$

(e)

$$(2) \quad d\mathcal{M} = \int d(\mathbb{P}_a - \mathbb{P}_b) = \int q d(\mathbb{P}_a - \mathbb{P}_b)$$

où $\log \int \in L^1(\mathbb{P}_a)$ et $\log \int \in L^1(\mathbb{P}_b)$;

(f) Si \mathbb{P}_a et \mathbb{P}_b sont absolument continues resp. par rapport à μ_a et μ_b deux mesures \mathcal{M} ..nies et $\log \frac{d\mathbb{P}_a}{d\mu_a} \in L^1(\mathbb{P}_a)$ et $\log \frac{d\mathbb{P}_b}{d\mu_b} \in L^1(\mathbb{P}_b)$, alors

$$(3) \quad d\mathbb{Q}^{\mathbb{P}} = \mu_a \otimes \mu_b d\mu_a d\mu_b;$$

La démonstration de ce théorème repose sur le lemme suivant:

Lemme 1. Si \mathbb{R} est une probabilité sur $\mathcal{F} - \mathcal{F}$ telle que $\mathbb{R} \in \mathbb{E}$ avec $H(\mathbb{P}; \mathbb{R}) < 1$: Alors $\mathbb{Q} \in \mathbb{E}$ telle que $H(\mathbb{Q}; \mathbb{R}) \cdot H(\mathbb{P}; \mathbb{R}) \in \mathbb{E}$:

De plus si \mathbb{P} et \mathbb{R} sont équivalentes à une mesure produit $\mathbb{P} \otimes \mathbb{R}$ sur $\mathcal{F} - \mathcal{F}$; alors $\frac{d\mathbb{Q}}{d\mathbb{R}}$ est définie par: $\frac{d\mathbb{Q}}{d\mathbb{R}}(x; y) = \int \mathbb{A}(x) \circ (y); 0 \cdot \mathbb{A}(x) < 1 \otimes \mu_j$ p.s; $0 \cdot \int (y) < 1$ j.p.s:

Théorème 2. Soient \mathbb{P}_a et \mathbb{P}_b vérifiant les hypothèses (f) du théorème 1 et la fonction q comme dans le théorème 1. Alors $\mathbb{Q}^{\mathbb{P}}$ donnée par

$$(4) \quad d\mathbb{Q}^{\mathbb{P}} = \mu_a \otimes \mu_b d\mu_a d\mu_b$$

vérifie $H(\mathbb{Q}^{\mathbb{P}}; \mathbb{R}) \cdot H(\mathbb{P}; \mathbb{R}) \in \mathbb{E}$ et \mathbb{R} donnée par:

$$(5) \quad d\mathbb{R}(x; y) = f(x) g(y) q(x; y) d\mu_a(x) d\mu_b(y)$$

avec $\log f \in L^1(\mathbb{P}_a)$ et $\log g \in L^1(\mathbb{P}_b)$.

Supposons que M , l'espace métrique \mathcal{M} compact, soit complet (donc séparable), \mathcal{F} sa tribu borélienne et $f \gg (t); a \cdot t < 1$ g un processus de Markov continu défini sur un espace probabilisé $(\cdot; \mathcal{F}; P)$ à valeurs dans M . Posons: $P(s; x; t; B) = P_{s;x}(\gg(t) \in B)$.

Supposons que la fonction de transition $P(s; x; t; B)$ admette $p(s; x; t; y)$ comme densité par rapport à μ_y (mesure \mathcal{M} définie dans \mathcal{F}). Soient \mathcal{E} un ouvert non vide, relativement compact dans M , la fonction $c: [a; b] \times \mathcal{E} \rightarrow \mathbb{R}$ mesurable et le domaine $D = \{f(s; x) \in [a; b] \times \mathcal{E} : |c(s; x)| < 1\}$. Considérons les hypothèses suivantes:

(H₁) $\frac{dP_a}{d\mu_a} = \mathbb{1}_{K_a}$ et $\frac{dP_b}{d\mu_b} = \mathbb{1}_{K_b}$ sont continues à supports compacts K_a et K_b tels que $K_a \cap K_b = \emptyset$;

(H₂) $p(a; \cdot; b; \cdot)$ est C^0 sur $\mathcal{E} - \mathcal{E}$ et strictement positive sur $K_a - K_b$;

(H₃) c est continue et continue sur $[a; b] \times \mathcal{E}$;

(H₄) $\int_a^b \exp(\int_a^r c(r; \gg(r)) dr) \int_{\mathcal{E}} dP_{ax} \mathbb{1}_{K_a}(x) \mu_a(dx) < 1$; où

$T = \inf\{t > a : |c(t; \gg(t))| = 1\}$ s'il existe et 1 sinon; est mesurable.

Théorème 3. Soient les hypothèses (H₁)-(H₄) vérifiées. Alors la fonction q donnée par

$$(6) \quad q(x; y) = E_{ax} \exp\left(\int_a^b c(r; \gg(r)) dr\right) \mathbb{1}_{\{b=y\}} p(a; x; b; y),$$

$x; y \in M$; vérifie les hypothèses du théorème 1.

Corollaire 1. Supposons que les hypothèses (H₁)-(H₄) soient vérifiées. Alors il existe une solution non négative $(\mathbb{1}_a; \mathbb{1}_b)$ du système de Schrödinger

$$(7) \quad \begin{cases} \mathbb{1}_a(x) = \int_a^b \mathbb{1}_a(x) q(x; y) \mathbb{1}_b(y) dy \\ \mathbb{1}_b(y) = \int_a^b \mathbb{1}_b(y) q(x; y) \mathbb{1}_a(x) dx \end{cases}$$

Soient M un espace métrique \mathcal{M} compact complet, $c^0: [a; 1] \times \mathcal{E} \rightarrow \mathbb{R}$ mesurable, $D^0 = \{f(s; x) \in [a; 1] \times \mathcal{E} : |c^0(s; x)| < 1\}$, $\mathbb{1}_s = \inf\{t > s : |c^0(t; \gg(t))| = 1\}$ g s'il existe et 1 sinon et $N_t^s = \mathcal{F}_t^s(\gg(u); a \cdot s \leq u \cdot t < 1)$.

Considérons les hypothèses suivantes:

(h₁) $f \mathbb{1}_s(\cdot) > 2 N_t^s; a \cdot s < t < 1$ (vérifiée si c^0 est assez régulière);

(h₂) c^0 est continue sur D^0 ;

(h₃) $E_{s;x} \exp(\int_s^t c^0(r; \gg(r)) dr) \mathbb{1}_{\{s > t\}} < 1; a \cdot s < t < 1; x \in M$;

(h₄) $c^0(s; x) = c(s; x), (s; x) \in [a; b] \times \mathcal{E}$;

Proposition 1. Sous les hypothèses (h₁)-(h₄), la fonction

$$q(s; x; t; y) = E_{s;x} \exp\left(\int_s^t c(r; \gg(r)) dr\right) \mathbb{1}_{\{T_s > t\}} \mathbb{1}_{\{y\}} p(s; x; t; y);$$

est une densité de quasi-transition; $a \cdot s < t < 1; x, y \in M$; où $T_s = \inf\{t > s : |c(t; \gg(t))| = 1\}$ g s'il existe et 1 sinon.

3. Pont de Schrödinger sur les processus de naissance et de mort

Considérons le cas où M est l'espace euclidien \mathbb{R}^n muni de sa tribu borélienne \mathcal{B} et $f \gg (t); a \cdot t < 1$ g un processus de Markov continu défini sur un espace probabilisé $(\cdot; \mathcal{F}; P)$ à valeurs dans \mathbb{R}^n .

Posons $F = \mathbb{R}^3 \times (t, a \cdot t \cdot b)$ notée N et définissons la probabilité R sur $(\cdot; N)$ par

$$(8) \quad dR = \frac{1}{\int_{\mathbb{R}^3} \int_a^b \exp\left(-\int_a^b c(r; \cdot)(r) dr\right) dP_a} \exp\left(-\int_a^b c(r; \cdot)(r) dr\right) dP_a;$$

où $P_a(\cdot) = \int_{\mathbb{R}^3} P_{ax}(\cdot) \otimes_a(x) dx$ est une probabilité sur N : Ainsi, la mesure de distribution jointe R aux instants a et b du processus $\cdot(t)$ par rapport à R , est donnée par:

$$(9) \quad dR = \frac{1}{\int_{\mathbb{R}^3} \int_a^b \exp\left(-\int_a^b c(r; \cdot)(r) dr\right) dP_a} \otimes_a(x) q(x; y) dx dy$$

avec q donnée par (6). Aussi, rappelons que si P_a et P_b sont absolument continues par rapport à la mesure de Lebesgue dans \mathbb{R}^n , et $\log \frac{dP_a}{dx} \in L^1(P_a)$, $\log \frac{dP_b}{dy} \in L^1(P_b)$ alors

$$(10) \quad dQ^\pi(x; y) = \int_a^b(x) q(x; y) \int_b^b(y) dx dy$$

est la projection de Csiszar de R sur

$$E \stackrel{\text{def}}{=} \{ P \text{ prob. sur } B \otimes B : P(\cdot; \mathbb{R}^n) = P_a(\cdot); P(\mathbb{R}^n; \cdot) = P_b(\cdot) \} \text{g, i.e., } H(Q^\pi; R) \cdot H(P; R); \text{ } \int_{P \in E} dP$$

Soient $(\cdot; M)$ l'espace mesurable où $\cdot = C^0([a; b]; \mathbb{R}^n)$; M sa tribu borélienne et le processus $X(t)$ défini par $X(t; x(\cdot)) = x(t)$, $x(\cdot) \in \cdot, t \in [a; b]$. Soient μ_{ax}^\cdot et μ_a^\cdot les mesures de distribution sur l'espace des trajectoires du processus \cdot par rapport à P_{ax} et P_a , i.e., $\mu_{ax}^\cdot(M) = P_{ax} f! : \cdot \in M \text{g}$ et $\mu_a^\cdot(M) = P_a f! : \cdot \in M \text{g}$; $M \in M$. Ainsi la mesure de distribution sur $(\cdot; M)$ de \cdot par rapport à R est donnée par $d^1 = \frac{\exp\left(-\int_a^b c(r; X(r)) dr\right) \int_b^b d^1 \mu_a^\cdot}{\int_{\mathbb{R}^3} \int_a^b \exp\left(-\int_a^b c(r; \cdot)(r) dr\right) dP_a}$

Le problème est le suivant: trouver μ^π , mesure de probabilité sur M , si elle existe, qui minimise l'entropie relative $H(\mu^\pi; \cdot)$ sur E définie par:

$$E = \{ \mu^\pi \text{ prob. sur } M : \mu^\pi[X(a) \in \cdot] = P_a(\cdot); \mu^\pi[X(b) \in \cdot] = P_b(\cdot) \} \text{g}$$

D'après la formule de multiplication suivant e

$$\frac{d^\mu}{d^1}(X) = \frac{dP}{dR}(X(a); X(b)) \frac{d^\mu(X(b))}{d^1(X(b))}(X); \quad \mu \in E;$$

le problème se réduit à un problème de type du théorème1: i.e., $\min_{P \in E} H(P; R)$.

D'après ce qui précède, nous obtenons alors: $\mu^\pi(\cdot) = \int_{\mathbb{R}^3} \int_a^b \mu^\pi(\cdot) dQ^\pi(x; y)$; avec

$$dQ^\pi(x; y) = \int_a^b(x) q(x; y) \int_b^b(y) dx dy$$

$$dR(x; y) = \frac{\int_{\mathbb{R}^3} \int_a^b \otimes_a(x) q(x; y) dx dy}{\int_{\mathbb{R}^3} \int_a^b \exp\left(-\int_a^b c(r; \cdot)(r) dr\right) dP_a}$$

$$d^1(\cdot) = \frac{\exp\left(-\int_a^b c(r; X(r)) dr\right) \int_b^b d^1 \mu_a^\cdot}{\int_{\mathbb{R}^3} \int_a^b \exp\left(-\int_a^b c(r; \cdot)(r) dr\right) dP_a}$$

On en déduit que:

$$(11) \quad d^{\mu^\pi}(\cdot) = \frac{\int_b^b(X(b))}{\int_a^b(X(a))} \exp\left(-\int_a^b c(r; X(r)) dr\right) \int_b^b d^1 \mu_a^\cdot$$

où $P_a(X(a)) = \mathbb{P}_a(X(a)) = P_a(X(a))$. Notons que $P_a(X(a)) > 0$; P_a i p:s:, et donc P_a i p:s:, puisque $\log P_a \in L^1(P_a)$. De plus posons

$$(12) \quad d^{\circ_{ax}}(\cdot) = \frac{P_b(X(b))}{P_a(X(a))} \exp \int_a^b c(r; X(r)) dr \quad \mathbb{P}_b < \tau d^{\circ_{ax}}$$

où $P_a(\cdot) = \int_a^b P_a(x) dx$, on trouve alors

$$(13) \quad \mathbb{P}^{\circ_{ax}}(\cdot) = \int_a^b \mathbb{P}_a(x) dx:$$

4. Caractérisation variationnelle pour le cas c non borné:

Considérons le cas où le processus de Markov $X(t)$ est un processus de Wiener $W(t); t \in [a; b]$, à valeurs dans \mathbb{R}^n de distribution initiale P_a . Ainsi P_a et P_{ax} seront remplacées par P_a^W et P_{ax}^W et nous garderons les mêmes notations pour P_a ; $\mathbb{P}^{\circ_{ax}}$ et $\mathbb{P}^{\circ_{ax}}$. Utilisant le résultat de la transformation du drift de Girsanov, nous obtenons, d'après (11) où P_a est remplacé par P_a^W ; $\mathbb{P}^{\circ_{ax}} \ll P_a^W$ et donc il existe $(w^{\circ_{ax}}(t); \tilde{A}^{\circ_{ax}}(t)); a \leq t \leq b$ tel que:

$$(14) \quad X(t) = X(a) + \int_a^t \tilde{A}^{\circ_{ax}}(r) dr + w^{\circ_{ax}}(t); t \in [a; b], \mathbb{P}^{\circ_{ax}} \ll P_a^W;$$

$$(15) \quad \int_a^b (\tilde{A}^{\circ_{ax}}(t))^2 dt < 1 = 1,$$

$$(16) \quad \frac{d^{\circ_{ax}}}{d^{\circ_{ax}}}(\cdot) = \exp \int_a^b \tilde{A}^{\circ_{ax}}(t) dX(t) - \frac{1}{2} \int_a^b (\tilde{A}^{\circ_{ax}}(t))^2 dt \quad \mathbb{P}^{\circ_{ax}} \ll P_a^W;$$

Nous allons généraliser un résultat de Wakolbinger[22] à une plus grande classe de potentiels c.

Soit A la classe des processus stochastiques nonanticipatifs $\tilde{A}(t)$ à valeurs dans \mathbb{R}^n par rapport à $(\mathbb{P}_a; M; N_t; \mathbb{P}_a)$, $t \in [a; b]$, vérifiant

(H₅): (i) $\mathbb{P}_a \ll \mathbb{P}_b < Tg = 1$;

(ii) les marginales de \mathbb{P}_a aux instants a et b sont P_a et P_b , avec $\log \frac{dP_a}{dP_b} \in L^1(P_a)$ et $\log \frac{dP_b}{dP_a} \in L^1(P_b)$;

(iii) $E_{\mathbb{P}_a} \int_a^b |\tilde{A}(t)|^2 dt < 1$;

(iv) $X(t) \mid X(a) \mid \mathbb{P}_a$ est un mouvement brownien standard sur [a; b] par rapport à \mathbb{P}_a ;

(v) $J(a; b; \tilde{A}) = E_{\mathbb{P}_a} \int_a^b k \tilde{A}(r) k^2 + c(r; X(r)) dr$ est défini.

(H₆): $E_{\mathbb{P}_a} \int_a^b c(r; X(r)) dr < 1$; (cette hypothèse est vérifiée si c est borné).

Théorème 4. Soient (H₁)-(H₆) vérifiées. Alors $\mathbb{P}^{\circ_{ax}} = (\tilde{A}^{\circ_{ax}}(t); \mathbb{P}^{\circ_{ax}}) \in A$, $t \in [a; b]$, tel que: $1 < J(a; b; \tilde{A}^{\circ_{ax}}) = \min_{A \ni \tilde{A}} J(a; b; \tilde{A}) < 1$; où $\mathbb{P}^{\circ_{ax}}$ est la projection de Csiszar

La récurrence des files d'attente avec impatience pour des équations non monotones

Abdalkader Gheriballah

Département de Mathématiques

Faculté des Sciences

Université de Sidi Bel Abbès

22000 Algérie

E-mail: gheribaek@yahoo.fr

Bénamar Chouaf

Département de Mathématiques

Faculté des Sciences

Université de Sidi Bel Abbès

22000 Algérie

E-mail: chouafb@yahoo.fr

Résumé: Dans notre travail, nous montrons que les files d'attente avec impatience pour des équations non monotones sont irréductibles périodiques sous la condition $b - a > 0$ et nous montrons que les chaînes sont récurrentes au sens de Harris.

0 Introduction

Nous supposons que des clients arrivent dans un système de files d'attente composé d'une seule station avec un seul serveur. La discipline de service est supposée ordinaire premier arrivé premier servi et que le serveur n'intérompt pas de lui même le service.

Les clients arrivent suivant les temps $A_1, A_1 + A_2, \dots, A_1 + A_2 + \dots + A_n$, et réclament des temps de services $(B_n)_n$ et attendent des temps $(W_n)_n$. En supposant que les suites des variables aléatoires $(A_n)_n, (B_n)_n$ sont indépendantes et que les $(C_n)_n$ qui représentent les temps d'impatience sont aussi indépendantes. Dans leur article Baccelli F. Poyer et Hebuterne G. [1] ont montré l'irréductibilité et la récurrence dans le cas où $b - a < 0$. Nous montrons que le système est irréductible, et qu'il est stable sous les conditions citées en ([1]) dans le cas $b - a > 0$.

1 Notations et Définitions

Soient

- α, β deux mesures de probabilité sur $(\mathbb{R}_+, \mathcal{B}_{\mathbb{R}_+})$

- γ une mesure de probabilité sur $(\overline{\mathbb{R}}_+, \mathcal{B}_{\overline{\mathbb{R}}_+})$

$(\Omega, \mathcal{A}, \mathbb{P})$ espace de probabilité tel que

$$\Omega = (\mathbb{R}_+ \times \mathbb{R}_+ \times \overline{\mathbb{R}}_+)^{\mathbb{N}}$$

$$\mathcal{A} = (\mathcal{B}_{\mathbb{R}_+} \otimes \mathcal{B}_{\mathbb{R}_+} \otimes \mathcal{B}_{\overline{\mathbb{R}}_+})^{\otimes \mathbb{N}}$$

$$\mathbb{P} = (\alpha \otimes \beta \otimes \gamma)^{\otimes \mathbb{N}}$$

$(A_n, B_n, C_n; n \geq 1)$ est la famille des coordonnées de l'espace Ω .

$\mathcal{A}_0 = \{\emptyset, \Omega\}; \forall n \geq 1 \quad \mathcal{A}_n = \sigma(A_p, B_p, C_p; 1 \leq p \leq n)$

A_n est l'inter-arrivé entre le client "n" et le client "n-1"

B_n est le temps de service du client "n-1"

C_n est le temps d'impatience du client "n-1".

On suppose que les trois suites sont indépendantes et que chaque suite est formée par des variables aléatoires indépendantes équidistribuées. Le client "0" arrive à l'instant "0". Le client "n" arrive à l'instant $\sum_{i=1}^n A_i$.

Soit W_n le temps d'attente du client "n" entre son arrivé et le début d'être servi. $\min(W_n, C_{n+1})$ est le temps réel de son attente et le client quitte le système à l'instant $\min(W_n, C_{n+1})$. Alors on a l'équation de récurrence suivante

$$W_{n+1} = \begin{cases} (W_n + B_{n+1} - A_{n+1})_+ & \text{si } W_n < C_{n+1} \\ (W_n - A_{n+1})_+ & \text{si } C_{n+1} \leq W_n \end{cases} \quad (1)$$

Pour tout $W_0 = x \in \mathbb{R}_+$, on définit le processus stochastique W^x sur $(\Omega, \mathcal{A}, \mathbb{P})$ par

$$W^x(\omega) = (W_n^x(\omega); n \in \mathbb{N}) = (W_n(x, \omega); n \in \mathbb{N})$$

Théorème 1.1. *La suite des temps d'attente $(W_n(x, \cdot))_n$ est une chaîne de Markov par rapport aux $(\mathcal{A}_n)_n$ à valeurs dans $(\mathbb{R}_+, \mathcal{B}_{\mathbb{R}_+})$ homogène de transition*

$$P(x, F) = (1 - \gamma(x)) \int_{\mathbb{R}_+ \times \mathbb{R}_+} \mathbf{1}_F(x + b - a)_+ \alpha(da) \beta(db) + \gamma(x) \int_{\mathbb{R}_+} \mathbf{1}_F((x - a)_+) \alpha(da).$$

donc $(W_n^x, n \geq 0)$ est la coordonnée de cette espace. Si $\mathcal{F}_n = \sigma(W_p; p \leq n)$, $W = (W_n; n \geq 0)$ est une chaîne de Markov par rapport aux $(\mathcal{F}_n; n \geq 0)$ tel que pour tout $x \in \mathbb{R}_+$, $\mathbb{P}_x(W_0 = x) = 1$. Soient

$$a = \text{ess. sup } A_1 = \inf\{t \geq 0 \mid \alpha([0, t]) = 1\}$$

$$b = \text{ess. inf } B_1 = \sup\{t \geq 0 \mid \beta([t, +\infty[) = 1\}.$$

$$c = \text{ess. inf } C_1 = \sup\{t \geq 0 \mid \gamma([t, +\infty[) = 1\}$$

Proposition 1.1 ([1]). *Si $b - a < 0$, alors pour tout $x \in \mathbb{R}_+$,*

$$\mathbb{P}_x\left(\bigcup_{n \geq 0} \{W_n = 0\}\right) > 0.$$

Proposition 1.2 ([1]). *Si, $\mathbb{P}(C_1 = \infty)\mathbb{E}(B_1) > \mathbb{E}(A_1)$, la chaîne $(W_n; n \geq 0)$ est transiente.*

Théorème 1.2 ([1]). *Sous les conditions, $b - a < 0$ et $\mathbb{P}(C_1 = \infty)\mathbb{E}(B_1) < \mathbb{E}(A_1)$, la chaîne $(W_n)_n$ est récurrente positive au sens de Harris.*

2 Résultats

Proposition 2.3. *Si $b - a \geq 0$, alors l'ensemble $[(c - a)_+, +\infty[$ est un ensemble absorbant et il existe un temps d'arrêt τ tel que*

$$(i) \quad \forall k \geq 1 \quad \mathbb{E}(\tau^k) < +\infty$$

$$(ii) \quad W_\tau \geq c \quad \mathbb{P} \text{ p.s}$$

$$(iii) \quad \forall n \in \mathbb{N} \quad W_{\tau+n} \geq (c - a)_+ \quad \mathbb{P} \text{ p.s}$$

Preuve: Pour $x \geq (c - a)_+$, on a

$$W_1 = \begin{cases} (x + B_1 - A_1)_+ & \text{si } x < C_1 \\ (x - A_1)_+ & \text{si } C_1 \leq x \end{cases} \geq \begin{cases} (x + b - a)_+ & \text{si } x < C_1 \\ (x - a)_+ & \text{si } C_1 \leq x \end{cases} \geq (c - a)_+$$

donc $P(x, [(c - a)_+, \infty]) = \mathbb{P}_x(W_1 \in [(c - a)_+, \infty]) = 1$.

Soit

$$\tau(\omega) = \inf\{n \geq 0 : \sum_{i=1}^n (B_i(\omega) - A_i(\omega)) \geq c\}$$

comme $B_i - A_i \geq b - a \geq 0$, τ satisfait à $\mathbb{E}(\tau^k) < +\infty \quad \forall k \geq 1$.

Il est clair que

si $W_\tau > C_\tau$, $W_{\tau+1} = (W_\tau - A_\tau)_+ \geq (C_\tau - A_\tau) \geq (c - a)_+$, sinon

$$W_{\tau+1} = \left(\sum_{i=0}^{\tau} (B_i - A_i) + B_{\tau+1} - A_{\tau+1}\right) \geq (c + b - a) \geq (c - a)_+.$$

Proposition 2.4. *Si $b - a > 0$, alors*

$$\forall x \in \mathbb{R}_+ \quad \mathbb{P}_x\left(\bigcup_{n \geq 0} \{(c - a)_+ \leq W_n \leq c + b - a\}\right) > 0.$$

Preuve: Si on prend $\varepsilon < \min(a, b - a)$; $N_1 = \left\lceil \frac{(x - (c + b - a))_+}{a - \varepsilon} \right\rceil$; $N_2 = \left\lceil \frac{c}{b - a} \right\rceil$
 et $N = \max(N_1, N_2)$, alors

$$F = \bigcap_{i=1}^{N+1} \{C_i \leq c + \varepsilon, A_i \geq a - \varepsilon\} \subset \{W_{N+1} \leq c + b - a\},$$

ce qui implique que

$$\mathbb{P}_x(W_{N+1} \leq c + b - a) \geq \mathbb{P}(F) = (\mathbb{P}(\{C_1 \leq c + \varepsilon, A_1 \geq a - \varepsilon\}))^{N+1} > 0.$$

Proposition 2.5. *Si $c + b - 2a < 0$, alors pour tout $x \in \mathbb{R}_+$,*

$$\mathbb{P}_x\left(\bigcup_{n \geq 0} \{W_n = 0\}\right) > 0.$$

Preuve: Pour $\varepsilon < a$ et $N_2 = \left\lceil \frac{(c - x + \varepsilon)_+}{b - a + \varepsilon} \right\rceil$, on a

$$F = \bigcap_{i=1}^{N+1} \{B_i \leq b + \varepsilon, C_i \leq c + \varepsilon, A_i \geq a - \varepsilon\} \subset \{W_{N+1} = 0\},$$

ce qui implique que

$$\mathbb{P}_x(W_{N+1} = 0) \geq \mathbb{P}(F) = (\mathbb{P}(\{B_1 \leq b + \varepsilon, C_1 \leq c + \varepsilon, A_1 \geq a - \varepsilon\}))^{N+1} > 0.$$

Théorème 2.3. *Si $c + b - 2a < 0$, alors la chaîne $(W_n^x; n \geq 0)$ est irréductible périodique de période $d = \left\lceil \frac{c}{b - a} \right\rceil + 1$.*

Preuve: Le résultat est une conséquence des Propositions (2.3) et (2.5)

Proposition 2.6. *Sous les conditions $c + b - 2a < 0$ et $\mathbb{P}(C_1 = \infty)\mathbb{E}(B_1) < \mathbb{E}(A_1)$, la chaîne $(W_n^x; n \geq 0)$ est récurrente positive au sens de Harris.*

Preuve: La démonstration est identique à celle de la Proposition (1.1)

Théorème 2.4. *Si $\mathbb{P}(C_1 = \infty)\mathbb{E}(B_1) < \mathbb{E}(A_1)$, alors pour tout $x \in \mathbb{R}_+$,*

$$\mathbb{P}_x(\limsup\{W_n \in [(c - a)_+, c + b - a]\}) = 1.$$

Preuve: Soient $D = [(c - a)_+, c + b - a]$ et $T_D = \inf\{n \geq 0; W_n \in D\}$. D'après la Proposition (2.4), il suffit de montrer que

$$\sup_{x \in D} \mathbb{E}(T_D / W_0 = x) < \infty.$$

Pour cela (voir [10]), il faut montrer qu'il existe, $\varepsilon > 0$ et $M < \infty$ tels que

$$\mathbb{E}(W_1/W_0 = x) \leq x - \varepsilon \quad \forall x \in D^c$$

et

$$\mathbb{E}(W_1/W_0 = x) \leq M \quad \forall x \in D$$

On a

$$\mathbb{E}(W_1/W_0 = x) = \int_0^x \alpha(t) dt + (1 - \gamma(x)) \int_x^\infty (1 - \beta(t - x))\alpha(t) dt$$

$$\mathbb{E}(W_1/W_0 = x) \leq x - (\mathbb{E}(A) - \mathbb{E}(B)(1 - \gamma(x))) + \int_x^\infty (1 - \alpha(t)) dt$$

Pour $x \leq c + b - a$, $\mathbb{E}(W_1/W_0 = x) \leq c + b - a + \mathbb{E}(B)$, et si $\mathbb{P}(C_1 = \infty)\mathbb{E}(B_1) < \mathbb{E}(A_1)$, alors il existe un $x_1 \leq x$ et $\varepsilon > 0$, tels que $\mathbb{E}(W_1/W_0 = x_1) < \varepsilon$, ce qui complète la démonstration.

Remarque 2.1. 1) Si $\min(b - a, c - a) > 0$, la chaîne ne se vide jamais, mais elle récurre dans l'ensemble $[c - a, c + b - a]$.

2) Si $c - a < 0$, la chaîne peut ne pas être vide.

Références

- [1] Bacelli F., Boyer P., Hebuterne G.: Single-server queues with impatient customers. *Adv. Appl. Prob.* 16, 887-905 (1984).
- [2] Charlot F., Ghidouche M., Hamami M.: Irréductibilité et récurrence au sens de Harris des temps d'attente des GI/G/q. *Z. Wahrscheinlichkeitstheorie verw. Gebiete* 43 (1978), 187-203.
- [3] Charlot F., Pujolle J.: Recurrence in single server queues with impatient customers. *Ann. Inst. Henri Poincaré, vol. XIV, n° 4* (1978), 399-410.
- [4] Charlot F., Chouaf B., Guellil A.: Sur la stabilité et la récurrence des chaînes de Markov et de systèmes de files d'attente. *Ann. sc. Clermont Ferrand II n° 93* (1989), 13-47.
- [5] Charlot F., Chouaf B.: Régénération et intégrabilité pour des réseaux ouverts. *C.R. Acad. Sci. Paris, t. 330, Série I, p. 893-897, 2000.*
- [6] Gheriballah A.: Sur la stabilité et la récurrence des systèmes de files d'attente avec impatience. *Thèse de magister. Université de S.B.A (1996).*

- [7] Gheriballah A., Chouaf B.: Recherche des ensembles minimum pour une cascade de files d'attente avec impatience. *Matematicki-Vesnik*, vol.53, No.3-4, pp. 61-70 (2001).
- [8] Nafdi A.: Irréductibilité, petits ensembles et régénération des réseaux de Jackson généralisés, *Thèse de l'université de Rouen*, (1997).
- [9] Numelin E. : General Irreducible Markov Chains and Non-Negative Operators (Cambridge University Press) Cambridge (1984)
- [10] Laslett G.M., Pollard D.B., Tweddie R.L.: Techniques for establishing ergodic and recurrence properties of continuous valued Markov chains. *Naval Res. Logist. Quart* (1978), 25,455-472.
- [11] Sigman K., Sheller-Wolf A.: New bounds for expected delay in FIFO GI/GI/c queues, *Queueing Sys. Th. Appl.* 26 (1997), 169-186.
- [12] Revuz D.: Markov Chains. North Holland, (1975).

Principes d'invariance hölderiens : Cas de suite non stationnaire

Djamel HAMADOUCHE, djhamad@yahoo.fr
 Youcef TALEB, yoataleb@yahoo.fr

Département de Mathématiques, Faculté des Sciences,
 Université M.MAMMERI, Tizi-ouzou, Algérie

Soit $(X_j)_{j \geq 1}$ une suite de variables aléatoires indépendantes, non nécessairement de même loi, centrées et de variances $EX_j^2 = \sigma_j^2$. On note par ξ_n les lignes polygonales obtenues par interpolation linéaire aux points $\left(\frac{j}{n}, \frac{S_j}{n}\right)$ où $S_j = \sum_{k=1}^j X_k$ et $s_n^2 = \sigma_1^2 + \sigma_2^2 + \dots + \sigma_n^2$.

Lorsque les variables aléatoires sont indépendantes et de même loi avec $EX_j^2 = \sigma^2$, le principe d'invariance de Donsker-Prohorov établit la convergence de ξ_n vers le mouvement brownien dans l'espace $C[0,1]$ des fonctions continues sur $[0,1]$ muni de la norme du supremum.

Le principe d'invariance dans les espaces de Hölder, notés H_α^0 , pour les variables aléatoires indépendantes, centrées et de même loi, a été d'abord démontré par Lamperti.

Théorème (Lamperti [4]) Soit $(X_j)_{j \geq 1}$ une suite de variables aléatoires indépendantes et de même loi, centrées, réduites. On suppose qu'il existe $\gamma > 2$ tel que $EX_j^\gamma < \infty$. On pose pour tout $n \in \mathbb{N}^*$, $0 \leq j < n$:

$$\xi_n(t, \omega) = \frac{1}{\sqrt{n}} \left[\sum_{0 \leq k \leq j} X_k(\omega) + (nt - j)X_{j+1}(\omega) \right], \quad \frac{j}{n} \leq t < \frac{j+1}{n}.$$

Alors la suite de lois de ξ_n converge étroitement vers la mesure de Wiener P_W dans H_α^0 pour tout $\alpha < \frac{1}{2} - \frac{1}{\gamma}$.

Ce résultat est étendu par Hamadouche [2] pour une suite strictement stationnaire de variables aléatoires α -mélangeantes ou associées. Par ailleurs, il a établi que le processus de Donsker-Prohorov lissé par convolution, pour les variables aléatoires indépendantes et de même loi, converge aussi vers le mouvement brownien. Le résultat pour les variables aléatoires dépendantes est également établi pour le lissage par convolution.

On se propose d'étendre le théorème de Lamperti, aussi bien pour le lissage polygonal que pour le lissage par convolution, au cas de variables aléatoires indépendantes, centrées non de même loi. La preuve de nos deux résultats repose essentiellement sur les deux résultats suivants

Proposition 1 (Hamadouche [2]) La convergence en loi dans H_α^0 d'une suite de processus $(\xi_n)_{n \geq 1}$ équivaut à l'équitension sur H_α^0 de la suite des lois P_n des éléments aléatoires ξ_n et à la convergence des lois fini-dimensionnelles de ξ_n .

Corollaire 1 (Lamperti [4]) Soit $(\xi_n)_{n \geq 1}$ une suite de processus nuls en zéro et vérifiant pour des constantes $\gamma > 0$, $\delta > 0$ et $c > 0$

$$E|\xi_n(t) - \xi_n(s)|^\gamma \leq c|t-s|^{1+\delta}.$$

Alors la suite des lois P_n des processus ξ_n est équitendue dans H_α^0 pour $0 < \alpha < \frac{\delta}{\gamma}$.

Pour plus de flexibilité dans la vérification de conditions de moments, on donne le théorème ci-dessous sur lequel repose les démonstrations de nos principes d'invariance.

Théorème 2 (Hamadouche [2]) Soit $(\xi_n)_{n \geq 1}$ une suite de processus à trajectoires dans H_α^0 vérifiant les conditions suivantes

a) Il existe des constantes $a > 0$, $b > 0$, $c > 0$ et une suite de nombres positifs $a_n \downarrow 0$ telles que

$$E|\xi_n(t) - \xi_n(s)|^a \leq c|t-s|^b, \quad (1)$$

pour tout n et tous s, t tels que $|t-s| \geq a_n$;

b) $\forall \varepsilon > 0$,

$$\lim_{n \rightarrow \infty} P\{w_\alpha(\xi_n, a_n) > \varepsilon\} = 0. \quad (2)$$

Alors pour tout $\alpha < a^{-1}(\min(a, b) - 1)$, la suite $(\xi_n)_{n \geq 1}$ est équitendue dans H_α^0 .

La première extension du théorème de Lamperti pour les variables aléatoires indépendantes non identiquement distribuées s'énonce comme suit:

Théorème 3 Soit $(X_j)_{j \geq 1}$ une suite de variables aléatoires indépendantes, centrées, non de même loi. On suppose qu'il existe $\gamma > 2$, $m > 0$ et $M > 0$ telles que

$$m \leq E|X_j|^2 \text{ et } E|X_j|^\gamma \leq M < \infty \quad (4)$$

$EX_j^\gamma < \infty$. On pose pour tout $n \in \mathbb{N}^*$, $0 \leq j < n$:

$$\xi_n(t, \omega) = \frac{1}{s_n} \left[\sum_{k=1}^{k=j} X_k(\omega) + (nt - j)X_{j+1}(\omega) \right], \quad \frac{j}{n} \leq t < \frac{j+1}{n} \quad (5)$$

Alors la suite de lois de ξ_n converge étroitement vers la mesure de Wiener P_W dans H_α^0 pour tout $\alpha < \frac{1}{2} - \frac{1}{\gamma}$.

Preuve :

Equitension des lois P_n :

D'après le corollaire 1, il suffit de montrer que sous les

hypothèses du théorème 3, on a $E|\xi_n(t) - \xi_n(s)|^\gamma \leq K|t-s|^\gamma$ avec $1 + \delta = \frac{\gamma}{2}$.

Cas 1 : $\frac{1}{n} \leq s \leq t \leq \frac{j+1}{n}$

$$E|\xi_n(t) - \xi_n(s)|^\gamma \leq E\left|\left(\frac{1}{n}\right)^{\frac{1}{2}}(t-s)X_{j+1}\right|^\gamma \leq m^{-\frac{\gamma}{2}} M |t-s|^{\frac{\gamma}{2}} = K|t-s|^{\frac{\gamma}{2}} \quad (6)$$

Cas 2 : $\frac{1}{n} \leq s \leq t \leq \frac{j+1}{n}$

L'inégalité triangulaire et l'inégalité de Jensen nous donnent

$$E|\xi_n(t) - \xi_n(s)|^\gamma \leq 3^\gamma \left(E|\xi_n(t) - \xi_n\left(\frac{j+k}{n}\right)|^\gamma + E|\xi_n\left(\frac{j+k}{n}\right) - \xi_n\left(\frac{j}{n}\right)|^\gamma + E|\xi_n\left(\frac{j}{n}\right) - \xi_n(s)|^\gamma \right).$$

Le premier terme et le troisième terme du second membre de l'inégalité se traitent de la même manière que le cas précédent et par conséquent

$$E|\xi_n(t) - \xi_n\left(\frac{j+k}{n}\right)|^\gamma \leq K_1 |t-s|^{1+\delta} \quad (7)$$

$$E|\xi_n\left(\frac{j}{n}\right) - \xi_n(s)|^\gamma \leq K_3 |t-s|^{1+\delta} \quad (8)$$

Pour le terme du milieu, l'application de l'inégalité de Marcinkiewicz-Zygmund implique

$$E|\xi_n\left(\frac{j+k}{n}\right) - \xi_n\left(\frac{j}{n}\right)|^\gamma \leq \left(\frac{1}{nm}\right)^{\frac{\gamma}{2}} C_\gamma \left(\sum_{i=1}^k E|X_i|^2\right)^{\frac{\gamma}{2}} \leq C_\gamma \left(\frac{1}{m}\right)^{\frac{\gamma}{2}} \left(\frac{k}{n}\right)^{\frac{\gamma}{2}} = K_2 |t-s|^{1+\delta}. \quad (9)$$

Les inégalités (7), (8) et (9) nous donnent finalement

$$E|\xi_n(t) - \xi_n(s)|^\gamma \leq K|t-s|^{1+\delta}$$

Le corollaire 1 permet de conclure que la suite des lois $P_n = P_{\xi_n}^{\xi_n^{-1}}$ du processus ξ_n est équitendue dans H_α^0 pour $0 < \alpha < \frac{\delta}{2} = \frac{1}{2} - \frac{1}{\gamma}$.

Convergence des lois fini-dimensionnelles

Montrons d'abord que pour un point donné s , $\xi_n(s) \xrightarrow{L} W_s$. De la définition de ξ_n , on déduit $\left|\xi_n(s) - \frac{S_{[ns]}}{s_n}\right| = (ns - [ns]) \frac{1}{s_n} X_{[ns]+1} \leq \frac{1}{s_n} X_{[ns]+1}$.

Il suffit de montrer que $\frac{1}{s_n} X_{[ns]+1} \xrightarrow{P} 0$ et $\frac{S_{[ns]}}{s_n} \xrightarrow{L} W_s$.

Pour $\varepsilon > 0$, l'inégalité de Bienaymé-Tchebychev implique $P\left(\frac{1}{s_n} X_{[ns]+1} \geq \varepsilon\right) \leq \frac{1}{nm} \frac{M^2}{\varepsilon^2} \xrightarrow{n \rightarrow \infty} 0$ et donc $\frac{1}{s_n} X_{[ns]+1} \xrightarrow{P} 0$. Par conséquent, avec le

théorème de Lindeberg, on montre aisément que $\frac{S_{[ns]}}{s_n} \xrightarrow{L} W_s$.

Considérons maintenant deux points s et t avec $s < t$. Les deux membres du couple $\left(\frac{S_{[ns]}}{s_n}, \frac{S_{[nt]} - S_{[ns]}}{s_n}\right)$ sont indépendants et ce grâce à l'indépendance des variables (X_j) . Comme R est séparable, alors

$\left(\frac{S_{[ns]}}{s_n}, \frac{S_{[nt]} - S_{[ns]}}{s_n}\right) \xrightarrow{L} (W_s, W_t - W_s)$. De plus

$$\left|(\xi_n(t) - \xi_n(s)) - \left(\frac{S_{[nt]} - S_{[ns]}}{s_n}\right)\right| \leq \left|\frac{1}{s_n} X_{[nt]+1}\right| + \left|\frac{1}{s_n} X_{[ns]+1}\right| \xrightarrow{P} 0.$$

Par conséquent $(\xi_n(t) - \xi_n(s)) \xrightarrow{L} (W_t - W_s)$.

Comme les variables (X_j) sont indépendantes alors

$(\xi_n(s), \xi_n(t) - \xi_n(s)) \xrightarrow{L} (W_s, W_t - W_s)$ puisque la fonction $(x, y) \rightarrow (x, x - y)$ est continue. On peut faire la même démonstration pour un ensemble de trois points ou plus ce qui revient à dire que les lois finidimensionnelles de ξ_n convergent vers celles du mouvement brownien W ce qui achève la démonstration du théorème 3.

On considère le processus de sommes partielles normalisé de Donsker-Prohorov

$$\xi_n(t) = \frac{1}{s_n} S_{[nt]}(t), \quad t \in [0, 1], \quad (10)$$

où $[nt]$ désigne la partie entière de (nt) .

Selon les besoins, on utilisera aussi les notations suivantes de ξ_n ci-dessous,

$$\xi_n(t) = \frac{1}{s_n} \sum_{k=1}^n S_i 1_{\left[\frac{i}{n}, \frac{i+1}{n}\right]}(t) \quad \text{ou} \quad \xi_n(t) = \frac{1}{s_n} \sum_{k=1}^n X_k 1_{\left[\frac{k}{n}, 1\right]}(t). \quad (11)$$

Soit K une densité de probabilité sur \mathbb{R} telle que:

$$\int_{\mathbb{R}} |u| K(u) du < \infty, \quad (12)$$

$$|K(x) - K(y)| \leq \alpha(K) |x - y|, \quad x, y \in [-1, 1], \quad (13)$$

et $(b_n)_{n \geq 1}$ une suite de réels positifs tendant vers zéro et vérifiant

$$\frac{1}{b_n} = O\left(n^{\frac{\tau}{2}}\right), \quad 0 < \tau < \frac{1}{2}. \quad (14)$$

On définit la suite $(K_n)_{n \geq 1}$ de noyaux de convolution par

$$K_n(t) = \frac{1}{b_n} K\left(\frac{t}{b_n}\right) \quad t \in \mathbb{R}. \quad (15)$$

Soit le processus lissé défini par

$$\zeta_n(t) = (\xi_n * K_n)(t) - (\xi_n * K_n)(0), \quad t \in [0, 1]. \quad (16)$$

Le terme $(\xi_n * K_n)(0)$ assure la nullité de ζ_n en zéro.

Théorème 4 Soit $(X_j)_{j \geq 1}$ une suite de variables aléatoires centrées, indépendantes non identiquement distribuées vérifiant (3). On suppose que les noyaux de convolution K_n vérifient (12), (13) et (15). Alors la suite de processus de sommes partielles lissés ζ_n définis par (17) converge faiblement vers le mouvement brownien W dans H_α^0 pour tout $\alpha < \frac{1}{2} - \max\left(\tau, \frac{1}{\gamma}\right)$.

Preuve

Les hypothèses (13) et (14) impliquent que toutes les trajectoires de ζ_n sont dans H_α^0 pour tout $\alpha < \frac{1}{2}$. On applique la proposition 1 et

le théorème 2 avec $a_n = \frac{1}{n}$.

Equitension:

Pour appliquer le le théorème 2 avec $a_n = \frac{1}{n}$, on doit alors étudier les deux cas: $t-s \geq \frac{1}{n}$ et $t-s < \frac{1}{n}$. Sans perte de généralité, on suppose que $t > s$.

cas 1: $t-s \geq \frac{1}{n}$, on a en introduisant l'espérance par rapport à la densité K_n :

$$E|\zeta_n(t) - \zeta_n(s)|^\gamma = E|(\xi_n * K_n)(t) - (\xi_n * K_n)(s)|^\gamma = E\left|E_{k_n} \left(\frac{1}{s_n} \sum_{k=[n(s-u)]+1}^{[n(t-u)]} X_k \right)^\gamma\right|.$$

En appliquant l'inégalité de Jensen par rapport au noyau K_n et le théorème de Fubini, on obtient

$$E|\zeta_n(t) - \zeta_n(s)|^\gamma \leq E\left(E_{k_n} \left| \frac{1}{s_n} \sum_{k=[n(s-u)]+1}^{[n(t-u)]} X_k \right|^\gamma\right) \leq \int_R E\left| \frac{1}{s_n} \sum_{k=[n(s-u)]+1}^{[n(t-u)]} X_k \right|^\gamma K_n(u) du.$$

Le résultat (5) et l'inégalité de Marcinkiewicz-Zygmund appliquée à l'espérance donnent

$$E|\zeta_n(t) - \zeta_n(s)|^\gamma \leq \int_R \left(\frac{1}{\sqrt{nm}}\right)^\gamma C_\gamma M^{\frac{2}{\gamma}} ([n(t-u)] - [n(s-u)])^{\frac{\gamma}{2}} K_n(u) du \leq C'_\gamma |t-s|^{\frac{\gamma}{2}}.$$

cas 2: $t-s < \frac{1}{n}$, on a en utilisant l'hypothèse (14)

$$E|\zeta_n(t) - \zeta_n(s)|^\gamma \leq \int_R \frac{1}{s_n} \sum_{k=1}^n |X_k| \frac{1}{b_n} \alpha(K) \Big|_{\frac{1}{b_n}}^{t-s} \Big|_{\frac{1}{n}, 1}(u) du \leq \frac{1}{\sqrt{nm}} \frac{1}{b_n} \alpha(K) |t-s|.$$

Finalement on obtient

$$w_\alpha(\zeta_n, \frac{1}{n}) = \sup_{|t-s| \leq \frac{1}{n}} \frac{|\zeta_n(t) - \zeta_n(s)|}{|t-s|^\alpha} \leq \frac{1}{\sqrt{m}} \frac{1}{b_n^{\frac{1}{2}-\alpha}} \alpha(K) \frac{1}{n} \sum_{k=1}^n |X_k|.$$

$$w_\alpha(\zeta_n, \frac{1}{n}) \xrightarrow{P} 0 \text{ si } \frac{1}{\sqrt{m}} \frac{1}{b_n^{\frac{1}{2}-\alpha}} \alpha(K) \frac{1}{n} \sum_{k=1}^n |X_k| \xrightarrow{P} 0.$$

Par l'inégalité de Markov avec l'hypothèse (4) on a

$$P\left[\left(\frac{1}{\sqrt{m}} \frac{1}{b_n^{\frac{1}{2}-\alpha}} \alpha(K) \frac{1}{n} \sum_{k=1}^n |X_k|\right) > \delta\right] \leq \frac{1}{\delta} \frac{1}{\sqrt{m}} \frac{1}{b_n^{\frac{1}{2}-\alpha}} \alpha(K) \frac{nM^\gamma}{n} \xrightarrow{n \rightarrow \infty} 0$$

En utilisant l'hypothèse (15), on déduit

$$\frac{n^{\alpha-\frac{1}{2}}}{b_n^2} = \left(\frac{n^{\frac{1}{2}(\alpha-\frac{1}{2})}}{b_n}\right)^2 \xrightarrow{n \rightarrow \infty} 0 \text{ dès que } \left(\frac{1}{2}-\alpha\right) > \tau \text{ soit } \alpha < \frac{1}{2}-\tau.$$

Le théorème 2 permet de conclure alors que la suite des lois de ζ_n

est équitendue dans H_α^0 pour $a = \gamma$, $b = \frac{\gamma}{2}$, $c = C'_\gamma$ et $a_n = \frac{1}{n}$.

Cette équitension obtenue dans H_α^0 est valable pour tout $\alpha < \frac{1}{2}-\tau$ et $\alpha < \frac{1}{2}-\gamma$ soit $\alpha < \frac{1}{2}-\max(\tau, \frac{1}{\gamma})$.

Convergence des lois fini-dimensionnelles de $\{\zeta_n, n \geq 1\}$:

On montre d'abord que la distance dans R^k de $(\zeta_n(t_1), \dots, \zeta_n(t_k))$ et $(\xi_n(t_1), \dots, \xi_n(t_k))$ tend en probabilité vers zéro. Par le théorème 3, les lois fini-dimensionnelles de ξ_n convergent vers celles du mouvement brownien, il en sera de même alors pour celles de ζ_n . On vérifie alors la convergence vers 0 de $E|\zeta_n(t) - \xi_n(t)|^2$ pour tout $t \in [0, 1]$.

$$E|\zeta_n(t) - \xi_n(t)|^2 = E\left|\int_R \xi_n(t-u)K_n(u)du - \xi_n(t)\right|^2 = E\left|\int_R (\xi_n(t-u)K_n(u)du - \xi_n(t)\int_R K_n(u)du)\right|^2$$

$$= E\left|E_{k_n}\left[\frac{1}{s_n} \sum_{i=[nt]+1}^{[n(t-u)]} X_i\right]\right|^2$$

Par application de l'inégalité de Jensen et le théorème de Fubini, on obtient

$$E|\zeta_n(t) - \xi_n(t)|^2 = \int_R E\left|\frac{1}{s_n} \sum_{i=[nt]+1}^{[n(t-u)]} X_i\right|^2 K_n(u)du$$

En appliquant enfin l'inégalité de Marcinkiewicz-Zygmund avec $q=2$, on aura

$$E|(\xi_n * K_n)(t) - \xi_n(t)|^2 < c \int_R \left[\frac{1}{s_n}\right]^2 \sum_{i=[nt]+1}^{[n(t-u)]} E|X_i|^2 |K_n(u)|^2 du \leq c' \int_R \left(|u| + \frac{2}{n}\right) K_n(u)du$$

Après changement de variable, on déduit

$$E|(\xi_n * K_n)(t) - \xi_n(t)|^2 \leq c'(b_n \int_R |v|K_n(v)dv + \frac{2}{n})$$

Comme $\int_R |v|K_n(v)dv < \infty$ et $b_n \xrightarrow{n \rightarrow \infty} 0$ alors $\xi_n(t) * K_n(t) - \xi_n(t)$ converge vers 0 dans $L^2(\Omega)$ pour tout $t \in [0,1]$ et en particulier pour $t=0$ on a $E|(\xi_n * K_n)(0)|^2 = 0$. Or pour tout t on a

$$E|\zeta_n(t) - \xi_n(t)|^2 \leq 2\left(E|(\xi_n * K_n)(t) - \xi_n(t)|^2 + E|(\xi_n * K_n)(0)|^2\right)$$

Comme les deux espérances du second membre de l'inégalité convergent vers 0, pour tout t alors $\zeta_n(t) - \xi_n(t) \xrightarrow{L^2} 0$ ce qui implique $\zeta_n(t) - \xi_n(t) \xrightarrow{P} 0$ et par conséquent $\sum_{i=1}^k |\zeta_n(t_i) - \xi_n(t_i)|^2 = \|\zeta_n(t) - \xi_n(t)\|_{R^k}^2 \xrightarrow{P} 0$.

Les lois fini-dimensionnelles de ζ_n ont donc la même limite que celles de ξ_n , c'est à dire, elles convergent vers celles du mouvement brownien. On conclut alors que $\zeta_n \xrightarrow{L} W$ dans H_α^0 pour tout $\alpha < \frac{1}{2} - \max\left(\tau, \frac{1}{\gamma}\right)$ et ceci achève la preuve du théorème 4.

Mots clés: Equitension, espace de Hölder, mouvement brownien, principe d'invariance, processus stochastique.

AMS classifications: 60B10, 60F05, 60F17, 62G30.

REFERENCES

- [1] **BILLINGSLEY P.**, Convergence of probability measures, J. Wiley, New York 1968.
- [2] **HAMADOUCHE D.**, Invariance principles in Hölder spaces, Portugaliae Mathematica, vol. 57 Fasc. 2-2000.
- [3] **HAMADOUCHE D., TALEB Y.**, Hölderian version of Donsker-Prohorov's invariance principles, Pub IRMA Lille 2003, Vol. 60 N°XIII,1-15.
- [4] **LAMPERTI J.**, On convergence of stochastic processes, Trans. Amer. Math. Soc. ,104 (1962), 430-435.

III – Modèles économétriques

ON THE LINEAR HYPOTHESIS TESTING IN NONPARAMETRIC REGRESSION MODEL

Zaher MOHDEB¹ and Abdelkader MOKKADEM²

¹*Department of Mathematics*

University Mentouri, Constantine, Algeria

²*Department of Mathematics, Bâtiment Fermat,*

45, Avenue des Etats-Unis,

78035 Versailles Cedex, France

E-mail: mohdeb@wissal.dz, mokkadem@math.uvsq.fr

Abstract

We construct linear hypotheses tests for the regression function in nonparametric regression model in the case of a homoscedastic error structure and a fixed design. More precisely, we test that the regression function f is an element of a finite dimensional vector space E . We assume that the functions satisfy the Hölder condition of order strictly greater than $1/2$ and we obtain the asymptotic weak behavior of the proposed test statistic, then we have the level and the asymptotic power of the test. A simulation study is conducted, for small sample size to demonstrate the performance of the proposed test.

Key words: Nonlinear regression; Nonparametric test.

1 Introduction

We consider the following regression model

$$(1) \quad Y_{i,n} = f(t_{i,n}) + \varepsilon_{i,n}, \quad i = 1, \dots, n,$$

where f is a unknown real function, defined on the interval $[0, 1]$ and $t_{1,n} = 0 < t_{2,n} < \dots < t_{n,n} = 1$, is a fixed sampling of the interval $[0, 1]$. The errors $\varepsilon_{i,n}$ form a triangular array of random variables with expectation zero and finite variance σ^2 , and for any n , the random variables $\varepsilon_{1,n}, \dots, \varepsilon_{n,n}$ are independent.

Let $E_p = \text{span}\{x_1, \dots, x_p\}$, where x_1, \dots, x_p denote p linearly independent functions defined on $[0, 1]$. We want to test:

$$(2) \quad H_0 : f \in E_p \quad \text{against} \quad H_1 : f \notin E_p.$$

The problem of model diagnostics has been the subject of plenty works. Although the considered methods are very varied, a great part of the literature, is based on the spline method: Cox and Coh (1989), Eubank and Spiegelman (1990), Jayasuriya (1996). A method based on a kernel estimate for f is proposed by Müller (1992); Härdle and Mammen (1993) suggested test statistics based on kernel estimation method in combination with bootstrap techniques. The use of the empirical Fourier coefficients of f to construct hypothesis test in the model (1) is studied by Eubank and Spiegelman (1990), Eubank and Hart (1992) and Mohdeb and Mokkadem (2001). Staniswalis and Severini (1991) introduce a test criterion based on a comparison of the parametric MLE and nonparametric MLE. Dette and Munk (1998) introduce a test statistic based on the estimation of the square of the distance of f to E_p with respect to a L^2 -norm. Other related work includes that of Dette (1999), González-Manteiga and Cao (1993) and Stute (1997).

In the present paper, we use a test statistic based on the average of squared residuals and show that it has a parametric asymptotic behavior as the statistic of Dette and Munk (1998) and Mohdeb and Mokkadem (2003).

For small sample size, the Monte Carlo study under H_0 , shows that the distribution of the proposed test statistic is close to the normal distribution and that the test has good power properties.

2 Main result

Let h be a positive density function on the interval $[0, 1]$ and $d\mu = h(t)dt$. We consider linearly independent functions x_1, \dots, x_p and denote by E_p the linear subspace spanned by these functions. We assume that the following conditions are satisfied.

- (C1) The design $\{t_1, \dots, t_n\}$ satisfies $\max_{i=2, \dots, n} \left| \int_{t_{i-1}}^{t_i} h(t) dt - \frac{1}{n} \right| = o\left(\frac{1}{n}\right)$;
- (C2) $h, f, x_k, k = 1, \dots, p$, are Hölder functions of order $\gamma > 1/2$;
- (C3) $\forall n, \varepsilon_{1,n}, \dots, \varepsilon_{n,n}$ are independent and $\exists C \in \mathbb{R}^+$ such that $E(\varepsilon_{i,n}^4) < C, \forall i, n$.

The usual inner product in $L^2(d\mu)$ is denoted by $\langle u, v \rangle = \int_0^1 u(t)v(t)h(t)dt$, $u, v \in L^2(d\mu)$, and L^2 -norm by $\|u\|_2^2 = \int_0^1 u^2(t)h(t)dt$, $u \in L^2(d\mu)$.

We set

$$R^2(f) = \min_{v \in E_p} \|f - v\|_2^2.$$

We use the notations $Y = (Y_1, \dots, Y_n)'$, $f_n = (f(t_1), \dots, f(t_n))'$, $x_{k,n} = (x_k(t_1), \dots, x_k(t_n))'$, $k = 1, \dots, p$, and $X = (x_{1,n}, \dots, x_{p,n})$.

Let $E_{p,n}$ be the subspace of \mathbb{R}^n spanned by $\{x_{1,n}, \dots, x_{p,n}\}$ and $\Pi_n = X(X'X)^{-1}X'$, be the projection matrix on $E_{p,n}$ with respect to the usual inner product in \mathbb{R}^n . $\Pi_n^\perp = I_n - X(X'X)^{-1}X'$, is the projection matrix on the orthogonal of $E_{p,n}$, where I_n is the identity matrix.

The projections of Y and f_n on $E_{p,n}$, are given by $\Pi_n Y = X\hat{\beta}$, and $\Pi_n f_n = X\tilde{\beta}$, where

$$\hat{\beta} = (X'X)^{-1}X'Y, \quad \tilde{\beta} = (X'X)^{-1}X'f_n.$$

Set $\hat{\beta} = (\hat{\beta}_1, \dots, \hat{\beta}_p)$ and we consider the following statistic defined by

$$R_n^2 = \frac{1}{n} \sum_{i=1}^n \left| Y_i - \sum_{k=1}^p \hat{\beta}_k x_k(t_i) \right|^2,$$

which can be written as

$$R_n^2 = \frac{1}{n} Y' \Pi_n^\perp Y.$$

Replacing Y by $f_n + \varepsilon$, and taking the expectation in the previous formula, we obtain

$$E(R_n^2) = \tilde{R}_n^2 + \frac{n-p}{n} \sigma^2,$$

where

$$\tilde{R}_n^2 = \frac{1}{n} f_n' \Pi_n^\perp f_n.$$

This suggests to estimate $R^2(f)$ by $R_n^2 - \frac{n-p}{n} \sigma^2$.

Since σ^2 is unknown, we replace it by the consistent estimator S_ε^2 given by Gasser, Sroka and Jennen-Steinmetz (1986), and defined by

$$S_\varepsilon^2 = \frac{1}{6(n-2)} \sum_{i=2}^{n-1} (Y_{i+1} + Y_{i-1} - 2Y_i)^2.$$

Our test statistic is then

$$\widehat{R}_n^2 = R_n^2 - \frac{n-p}{n} S_\varepsilon^2.$$

We reject the null hypothesis $H_0 : f \in E_p$, if \widehat{R}_n^2 exceeds a critical value.

The following result gives the asymptotic level and power of the test.

Theorem 1 *If the conditions (C1), (C2) and (C3) hold, then*

$$\sqrt{n} (\widehat{R}_n^2 - \widetilde{R}_n^2) \xrightarrow[n \rightarrow +\infty]{d} \mathcal{N} \left(0, \frac{17}{9} \sigma^4 + 4\sigma^2 R^2 \right),$$

where $R^2 = R^2(f)$.

Proposition 1 *Consider the local alternatives of the form $f(t) = g(t) + c_n l(t)$, where g is a function in the subspace E_p , l is a function orthogonal to E_p and c_n is sequence tending to zero, then*

$$\sqrt{n} \widehat{R}_n^2 \xrightarrow[n \rightarrow +\infty]{d} \mathcal{N} \left(\int_0^1 l^2(t) h(t) dt, \frac{17}{9} \sigma^4 \right) \quad \text{if } c_n = n^{-1/4},$$

and

$$\sqrt{n} \widehat{R}_n^2 \xrightarrow[n \rightarrow +\infty]{d} \mathcal{N} \left(0, \frac{17}{9} \sigma^4 \right) \quad \text{if } c_n = o(n^{-1/4}).$$

This means that the test can detect local alternatives converging to the null hypothesis at any rate slower than $n^{-1/4}$.

In Theorem 1, the asymptotic distribution of \widehat{R}_n^2 depends on σ^2 and $R^2(f)$. Under H_0 , $\widetilde{R}_n^2 = 0$ and $R^2(f) = 0$, then $\sqrt{n} \widehat{R}_n^2$ converges in distribution to normal law $\mathcal{N} \left(0, \frac{17}{9} \sigma^4 \right)$. The variance σ^2 is generally unknown in practice; then in order to obtain a test for H_0 , we need a consistent estimator $\widehat{\sigma}^2$ of σ^2 . Then under H_0 , we have the following result as consequence of Theorem 1.

Corollary 1 *If the conditions of Theorem 1 hold and $\widehat{\sigma}^2$ is a consistent estimator of σ^2 , then under the null hypothesis of linearity H_0 we have*

$$\frac{1}{\widehat{\sigma}^2} \left(\frac{9n}{17} \right)^{1/2} \widehat{R}_n^2 \xrightarrow[n \rightarrow +\infty]{d} \mathcal{N} (0, 1).$$

Applying Corollary 1, the hypothesis $H_0 : f \in E_p$ is rejected at the level α if

$$(3) \quad \frac{1}{\widehat{\sigma}^2} \left(\frac{9n}{17} \right)^{1/2} \widehat{R}_n^2 > c_{1-\alpha},$$

where $c_{1-\alpha}$ is the $(1 - \alpha)$ -quantile of the standard normal distribution.

3 Monte Carlo study

In order to investigate the finite sample behavior of the test (3), some simulation studies are conducted and compare it with other procedures proposed in the literature.

We obtain a sample of size n from model $Y_i = f(t_i) + \varepsilon_i$ for $i = 1, \dots, n$, where the ε_i are independent and identically distributed with zero mean and variance σ^2 . As design, we use the regular sequence $t_i = \frac{i-1}{n-1}$, $i = 1, \dots, n$, on the interval $[0, 1]$.

The comparison is extended to alternative approaches given in González-Manteiga and Cao (1993), with four different models,

$$\begin{aligned} f_1(t) &= 1 + 2t + t^2 && \text{model I,} \\ f_2(t) &= 1 + 2t + t^2 + t^3 && \text{model II,} \\ f_3(t) &= 2 + 2 \sin t && \text{model III,} \\ f_4(t) &= 2 + 2 \sin \frac{3\pi t}{2} && \text{model IV,} \end{aligned}$$

and, in every case, we test the null hypothesis that f is a polynomial of degree less than or equal to 2, at the α level of significance. The errors ε_i , $i = 1, \dots, n$ are independent and identically normally distributed with zero mean and variance σ^2 .

We note that the first function is the only belonging to the linear subspace $E_3 = \text{span}\{1, t, t^2\}$, the test may then exhibit the estimated probabilities of rejection close to α . The distance between the remaining functions and E_3 is small.

We compute the empirical rejection probability, with two levels $\alpha = 0.05$ and $\alpha = 0.10$. The experiment is replicated 1000 times with sample size $n = 25$, for each combination of α and σ^2 . The variance σ^2 is estimated by

$$\hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^n |Y_i - \hat{\beta}_1 t_i^2 - \hat{\beta}_2 t_i - \hat{\beta}_3|^2.$$

The simulation results show the empirical probabilities of rejection for the different regression models (I, II, III and IV). We note that the empirical probabilities of rejection are quite low for the regression models II, III and IV which don't belong to the subspace E_3 . It seems reasonable since the distance between each function and the subspace E_3 is small. Observe also that the proportions of rejection decrease whenever the values of the variance σ^2 increase and the results are comparable with those given in González-Manteiga and Cao (1993).

References

- [1] Cox, D. and Koh, E. (1989). A smoothing spline based test of model adequacy in polynomial regression. *Ann. Inst. Statist. Math.*, **41**, 2, 383-400.
- [2] Dette, H. (1999). A consistent test for the functional form of a regression based on a difference of variance estimators. *Ann. Statist.*, **27**, 1012-1040.
- [3] Dette, H., and Munk, A. (1998). Validation of linear regression models. *Ann. Statist.*, **26**, 778-800.
- [4] Eubank, R. L. and Hart, J. D. (1992). Testing goodness-of-fit in regression using nonparametric via order selection criteria. *Ann. Statist.*, **20**, 1412-1425.
- [5] Eubank, R. L. and Spiegelman, C. H. (1990). Testing the goodness-of-fit of a linear model via nonparametric regression techniques. *J. Amer. Statist. Assoc.* **85**, 410, 387-392.
- [6] Gasser, T., Sroka, L. and Jennen-Steinmetz, C. (1986). Residual variance and residual pattern in nonlinear regression. *Biometrika*, **73**, 625-633.
- [7] González-Manteiga, W. and Cao, R. (1993). Testing the hypothesis of a general linear model using nonparametric regression estimation. *Test*, **2**, 161-188.
- [8] Härdle, W. and Mammen, E. (1993). Comparing nonparametric versus parametric regression fits. *Ann. Statist.*, **21**, 1926-1947.
- [9] Jayasuriya, B. R. (1996). Testing for polynomial regression using nonparametric regression techniques. *J. Amer. Statist. Assoc.*, **91**, 1626-1631.
- [10] Mohdeb, Z. and Mokkadem, A. (2001). Testing hypotheses on Fourier coefficients in nonparametric regression model. *Nonparametr. Statist.*, **13**, 605-629.
- [11] Mohdeb, Z. and Mokkadem, A. (2003). Average Squared Residuals Approach For Testing Linear Hypotheses In Nonparametric Regression. Technical report 2003-112, University of Versailles.
- [12] Müller, H. G. (1992). Goodness of fit diagnostics for regression models. *Scand. J. Statist.*, **19**, 157-172.
- [13] Staniswalis, J. G., Severini, T. A. (1991). Diagnostics for assessing regression models. *J. Amer. Statist. Assoc.*, **86**, 415, 684-692.
- [14] Stute, W. (1997). Nonparametric model checks for regression. *Ann. Statist.*, **25**, 613-641.

Analyse bayésienne d'un modèle autorégressif à erreurs ARCH

Hakima Merad et Hafida Guerbyenne

Université des Sciences et de la Technologie Houari Boumediene 16111, Bab Ezzouar, Alger

1. Introduction

Notre objectif est, d'une part, d'estimer les paramètres d'un modèle autorégressif à erreurs ARCH en utilisant l'approche bayésienne et l'algorithme du griddy gibbs sampling que nous avons adapté à notre contexte et, d'autre part, de comparer les estimations avec celles obtenues par la méthode classique du maximum de vraisemblance. Les estimations des différents paramètres, pour cette dernière méthode, ne peuvent pas être obtenues explicitement à partir des équations normales, l'algorithme du score fondé, sur un développement limité de la fonction score à l'ordre un, permet de surmonter ces difficultés d'ordre analytique. A notre connaissance ce travail n'a pas été entrepris auparavant. Lubrano et Bauwens (1998) ont, dans une partie de leurs travaux, considéré un modèle ARCH(1) où $\alpha_0=1$. Nous considérons le cas plus général où α_0 est quelconque et inconnu. Geweke (1988) a considéré un modèle de régression à erreurs ARCH en utilisant l'approche bayésienne et l'échantillonnage d'importance.

2. Le modèle

y_t est un processus autorégressif d'ordre q à erreurs ARCH(p) s'il est solution de l'équation stochastique :

$$\begin{cases} y_t = y_{t-1}\beta_0 + y_{t-2}\beta_1 + \dots + y_{t-q}\beta_{q-1} + \varepsilon_t \\ \varepsilon_t / \psi_{t-1} \rightarrow N(0, h_t) \forall t = 1, \dots, T \\ h_t = \alpha_0 + \alpha_1 \varepsilon_{t-1}^2 + \dots + \alpha_p \varepsilon_{t-p}^2 \text{ avec } \alpha_0 > 0, \alpha_i \geq 0 \forall i = 1, \dots, p \end{cases}$$

où $\beta_0, \beta_1, \dots, \beta_{q-1}$ sont les paramètres de la partie autorégressive du modèle, ψ_{t-1} étant la tribu engendrée par le passé du processus.

3. Les lois a posteriori des différents paramètres du modèle

Théorème

Les lois a posteriori conditionnelles des différents paramètres $\alpha_0, \dots, \alpha_p, \beta_0, \dots, \beta_{q-1}$ d'un modèle AR(q) à erreurs ARCH(p) sont

$$\begin{aligned} f(\alpha_i / \alpha_0, \dots, \alpha_{i-1}, \alpha_{i+1}, \dots, \alpha_p, \beta_0, \dots, \beta_{q-1}, y) &\propto p(\alpha_0, \dots, \alpha_p) * L(\alpha_0, \dots, \alpha_p, \beta_0, \dots, \beta_{q-1} / y) \\ f(\beta_j / \alpha_0, \dots, \alpha_p, \beta_0, \dots, \beta_{j-1}, \beta_{j+1}, \dots, \beta_{q-1}, y) &\propto p(\beta_0, \dots, \beta_{q-1}) * L(\alpha_0, \dots, \alpha_p, \beta_0, \dots, \beta_{q-1} / y) \end{aligned}$$

où $L(\alpha_0, \dots, \alpha_p, \beta_0, \dots, \beta_{q-1} / y)$ est la fonction de vraisemblance conditionnelle dont l'expression est la suivante

$$L(\alpha_0, \dots, \alpha_p, \beta_0, \dots, \beta_{q-1} / y) = \prod_{t=1}^T \frac{1}{\sqrt{2\pi h_t}} \exp \left\{ - \frac{\left(y_t - \sum_{j=1}^q \beta_{j-1} y_{t-j} \right)^2}{2h_t} \right\}$$

où y est le vecteur des observations et h_t est définie par l'expression

$$h_t = \alpha_0 + \sum_{i=1}^p \alpha_i \left(y_t - \sum_{j=1}^q \beta_{j-1} y_{t-j} \right)^2$$

$p(\alpha_0, \alpha_1, \dots, \alpha_p)$ désigne la loi a priori de $(\alpha_0, \dots, \alpha_p)$. Geweke (1988) a considéré les trois lois a priori suivantes

$$\text{i. } p(\alpha_0, \alpha_1, \dots, \alpha_p) \propto \left(\frac{1 - \sum_{i=1}^p \alpha_i}{\alpha_0} \right)^{1/2} J(\alpha_0, \dots, \alpha_p) K(\alpha_0, \dots, \alpha_p)$$

$$\text{ii. } p(\alpha_0, \dots, \alpha_p) \propto J(\alpha_0, \dots, \alpha_p) K(\alpha_0, \dots, \alpha_p)$$

$$\text{iii. } p(\alpha_0, \dots, \alpha_p) \propto J(\alpha_0, \dots, \alpha_p)$$

avec

$$J(\alpha_0, \dots, \alpha_p) = \begin{cases} 1 & \text{si } \alpha_0 > 0 \text{ et } \alpha_i \geq 0 \quad \forall i = 1, \dots, p \\ 0 & \text{sinon} \end{cases}$$

$$K(\alpha_0, \dots, \alpha_p) = \begin{cases} 1 & \text{si les racines du polynôme } \left(1 - \sum_{i=1}^p \alpha_i Z^i \right) \text{ sont en dehors du cercle unité} \\ 0 & \text{sinon} \end{cases}$$

$p(\beta_0, \beta_1, \dots, \beta_q)$ est la loi a priori de $(\beta_0, \beta_1, \dots, \beta_q)$, que nous prendrons proportionnelle à une constante.

Démonstration

Immédiate

4. L'algorithme du Griddy Gibbs sampling

1/ Soit k, m, d_j, d'_j quatre entiers positifs fixés et $\beta_0^{(0)}, \dots, \beta_{q-1}^{(0)}, \alpha_0^{(0)}, \dots, \alpha_p^{(0)}$ les données initiales de $\beta_0, \beta_1, \dots, \beta_{q-1}, \alpha_0, \dots, \alpha_p$

2/ Poser $r = 1, n = 0, j = 0, j' = 0$

2.1/ Générer $u \rightarrow U_{[0,1]}$

2.2/ Soient $[c_1, z_1], \dots, [c_p, z_p], [c'_1, z'_1], \dots, [c'_q, z'_q]$ fixés (les domaines de définition de $\beta_0, \beta_1, \dots, \beta_{q-1}, \alpha_0, \dots, \alpha_p$)

3/ Calculer $h_j = \frac{z_j - c_j}{d_j}$

3.1/ Calculer $f(\alpha_j / \alpha_0^{(n+1)}, \dots, \alpha_{j-1}^{(n+1)}, \alpha_{j+1}^{(n)}, \dots, \alpha_p^{(n)}, \beta_0^{(n)}, \dots, \beta_{q-1}^{(n)}, y)$ en $\alpha_j = c_j + (s_j - 1)h_j$

$\forall s_j = 1, \dots, d_j + 1$

3.2/ Calculer $\int_{c_j}^{c_j + (s_j - 1)h_j} f(\alpha_j / \alpha_0^{(n+1)}, \dots, \alpha_{j-1}^{(n+1)}, \alpha_{j+1}^{(n)}, \dots, \alpha_p^{(n)}, \beta_0^{(n)}, \dots, \beta_{q-1}^{(n)}, y) d\alpha_j \quad \forall s_j = 1, \dots, d_j + 1$ (par

une méthode de calcul d'intégrales approximées, par exemple la méthode des trapèzes ou la méthode de Simpson)

$$3.3/ \text{ Poser } g_j = \left(\begin{array}{l} 0, \int_{c_j}^{c_j+h_j} f(\alpha_j / \alpha_0^{(n+1)}, \dots, \alpha_{j-1}^{(n+1)}, \alpha_{j+1}^{(n)}, \dots, \alpha_p^{(n)}, \beta_0^{(n)}, \dots, \beta_{q-1}^{(n)}, y) d\alpha_j, \dots, \\ \int_{c_j}^{c_j+d_j h_j} f(\alpha_j / \alpha_0^{(n+1)}, \dots, \alpha_{j-1}^{(n+1)}, \alpha_{j+1}^{(n)}, \dots, \alpha_p^{(n)}, \beta_0^{(n)}, \dots, \beta_{q-1}^{(n)}, y) d\alpha_j \end{array} \right)$$

3.4/ Si $\exists b_j \in \{1, \dots, d_j + 1\}$ telle que $|u - g_j(b_j)| < 10^{-4}$ alors $\alpha_j^{(n+1)} = c_j + (b_j - 1)h_j$ et aller en 3.6/

3.6/ sinon $\exists b_j \in \{1, \dots, d_j + 1\}$ telle que $g_j(b_j) < u < g_j(b_j + 1)$ et aller en 3.5/

3.5/ Poser $c_j = c_j + (b_j - 1)h_j$ et $z_j = c_j + b_j h_j$ et aller en 3/

3.6/ Si $j < p$ poser $j = j + 1$ et aller en 3/ sinon on obtient $\alpha_0^{(n+1)}, \dots, \alpha_p^{(n+1)}$ aller en 4/

$$4/ \text{ Calculer } h'_{j'} = \frac{z'_{j'} - c'_{j'}}{d'_{j'}}$$

4.1/ Calculer $f(\beta_{j'} / \alpha_0^{(n+1)}, \dots, \alpha_p^{(n+1)}, \beta_0^{(n+1)}, \dots, \beta_{j'-1}^{(n+1)}, \beta_{j'+1}^{(n)}, \dots, \beta_q^{(n)}, y)$ au points $s'_{j'} = 1, \dots, d'_{j'} + 1$

4.2/ Calculer $\int_{c'_{j'}}^{c'_{j'}+(s'_{j'}-1)h'_{j'}} f(\beta_{j'} / \alpha_0^{(n+1)}, \dots, \alpha_p^{(n+1)}, \beta_0^{(n+1)}, \dots, \beta_{j'-1}^{(n+1)}, \beta_{j'+1}^{(n)}, \dots, \beta_q^{(n)}, y) d\beta_{j'}$

pour $s'_{j'} = 1, \dots, d'_{j'} + 1$ (comme pour les α_j , en utilisant une méthode numérique de calculs d'intégrales)

$$4.3/ \text{ Poser } g'_{j'} = \left(\begin{array}{l} 0, \int_{c'_{j'}}^{c'_{j'}+h'_{j'}} f(\beta_{j'} / \alpha_0^{(n+1)}, \dots, \alpha_p^{(n+1)}, \beta_0^{(n+1)}, \dots, \beta_{j'-1}^{(n+1)}, \beta_{j'+1}^{(n)}, \dots, \beta_q^{(n)}, y) d\beta_{j'}, \dots, \\ \int_{c'_{j'}}^{c'_{j'}+d'_{j'} h'_{j'}} f(\beta_{j'} / \alpha_0^{(n+1)}, \dots, \alpha_p^{(n+1)}, \beta_0^{(n+1)}, \dots, \beta_{j'-1}^{(n+1)}, \beta_{j'+1}^{(n)}, \dots, \beta_q^{(n)}, y) d\beta_{j'} \end{array} \right)$$

4.4/ Si $\exists b'_{j'} \in \{1, \dots, d'_{j'} + 1\}$ telle que $|u - g'_{j'}(b'_{j'})| < 10^{-4}$ alors $\beta_j^{(n+1)} = c'_{j'} + (b'_{j'} - 1)h'_{j'}$ et aller en 4.6/ sinon $\exists b'_{j'} \in \{1, \dots, d'_{j'} + 1\}$ telle que $g'_{j'}(b'_{j'}) < u < g'_{j'}(b'_{j'} + 1)$ et aller en 4.5/

4.5/ Poser $c'_{j'} = c'_{j'} + (b'_{j'} - 1)h'_{j'}$, $z'_{j'} = c'_{j'} + b'_{j'} h'_{j'}$ et aller en 4/

4.6/ Si $j' < q$ poser $j' = j' + 1$ et aller en 4/ sinon aller en 5/

5/ Si $n < k$ poser $n = n + 1$ et aller en 2.1/ sinon poser

$$\alpha_0^{(0)} = \alpha_0^{(k)}, \dots, \alpha_p^{(0)} = \alpha_p^{(k)}, \beta_0^{(0)} = \beta_0^{(k)}, \dots, \beta_{q-1}^{(0)} = \beta_{q-1}^{(k)}$$

et $\alpha_0^{(r,k)} = \alpha_0^{(k)}, \dots, \alpha_p^{(r,k)} = \alpha_p^{(k)}, \beta_0^{(r,k)} = \beta_0^{(k)}, \dots, \beta_{q-1}^{(r,k)} = \beta_{q-1}^{(k)}$ puis aller en 6/

6/ Si $r < m$ poser $r = r + 1$ et aller en 2.1/ sinon aller en 7/

7/ $\hat{f}(\alpha_j / y) = \frac{1}{m} \sum_{l=1}^m f(\alpha_0^{(l,k)}, \dots, \alpha_{j-1}^{(l,k)}, \alpha_{j+1}^{(l,k)}, \dots, \alpha_p^{(l,k)}, \beta_0^{(l,k)}, \dots, \beta_{q-1}^{(l,k)}, y)$ est l'estimateur de la densité a posteriori de $\alpha_j \forall j = 0, \dots, p$

De même $\hat{f}(\beta_{j'} / y) = \frac{1}{m} \sum_{l=1}^m f(\alpha_0^{(l,k)}, \dots, \alpha_p^{(l,k)}, \beta_0^{(l,k)}, \dots, \beta_{j'-1}^{(l,k)}, \beta_{j'+1}^{(l,k)}, \dots, \beta_{q-1}^{(l,k)}, y)$ est l'estimateur de la densité a posteriori de $\beta_{j'} \forall j' = 0, \dots, q - 1$

Applications numériques

Etude par simulation d'un modèle AR(2) à erreurs ARCH(1)

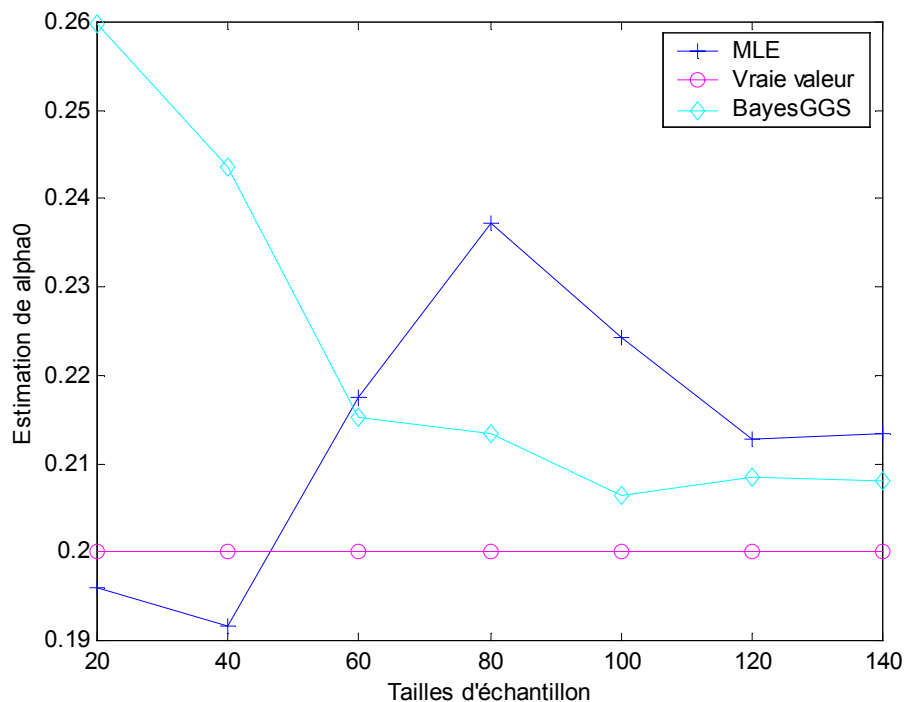
Des données artificielles ont été générées selon le modèle suivant

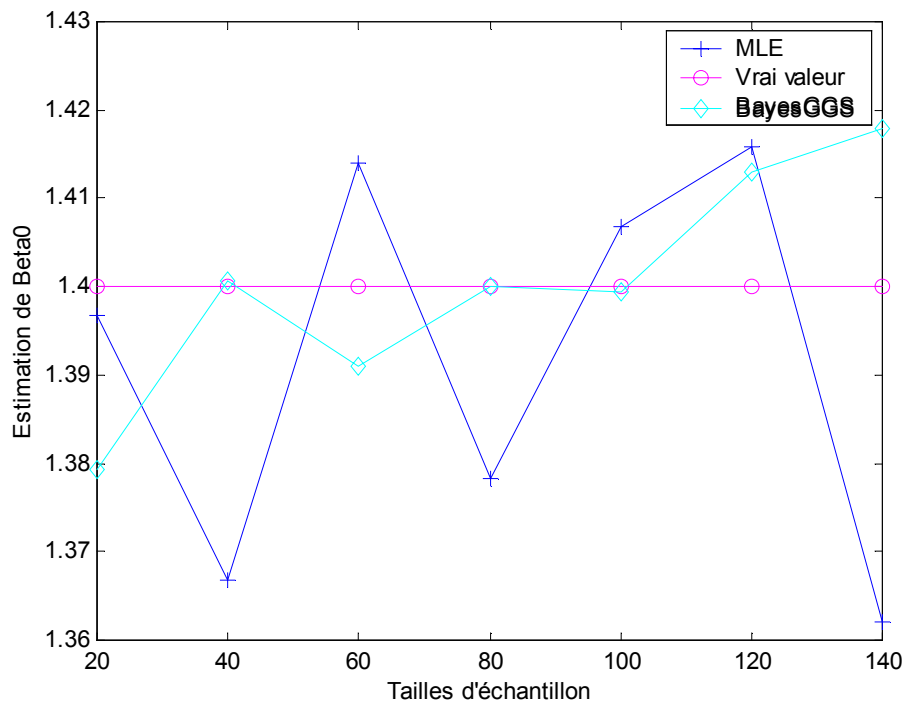
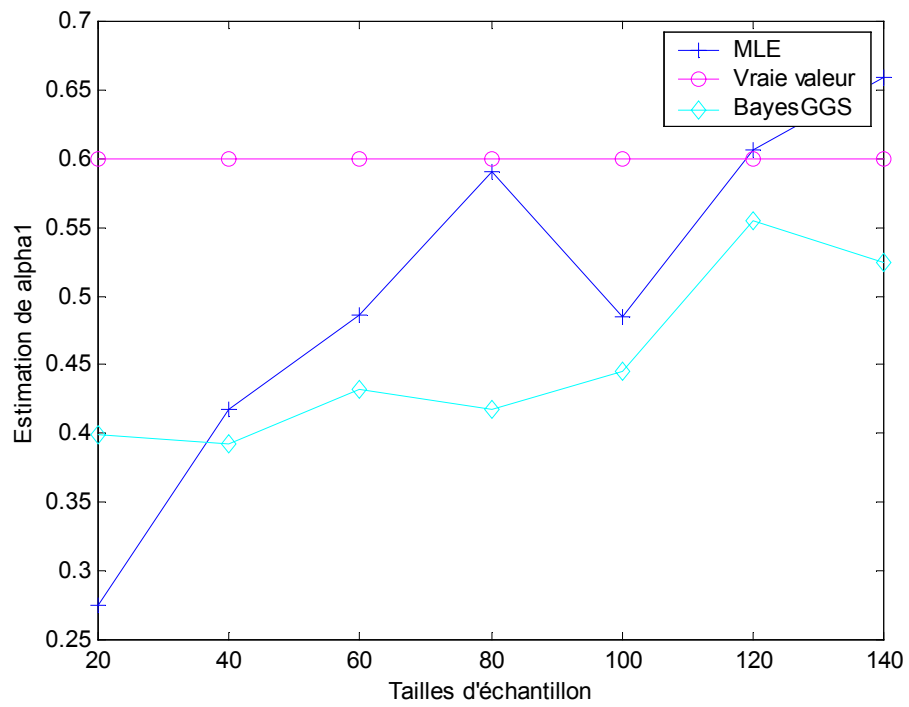
$$\begin{cases} y_t = 1.4y_{t-1} - 0.7y_{t-2} + \varepsilon_t \quad \forall t = 1, \dots, T \\ \varepsilon_t = v_t \sqrt{0.2 + \alpha_1 \varepsilon_{t-1}^2} \quad \forall t = 1, \dots, T \\ v_t \rightarrow N(0,1) \quad \forall t = 1, \dots, T \end{cases}$$

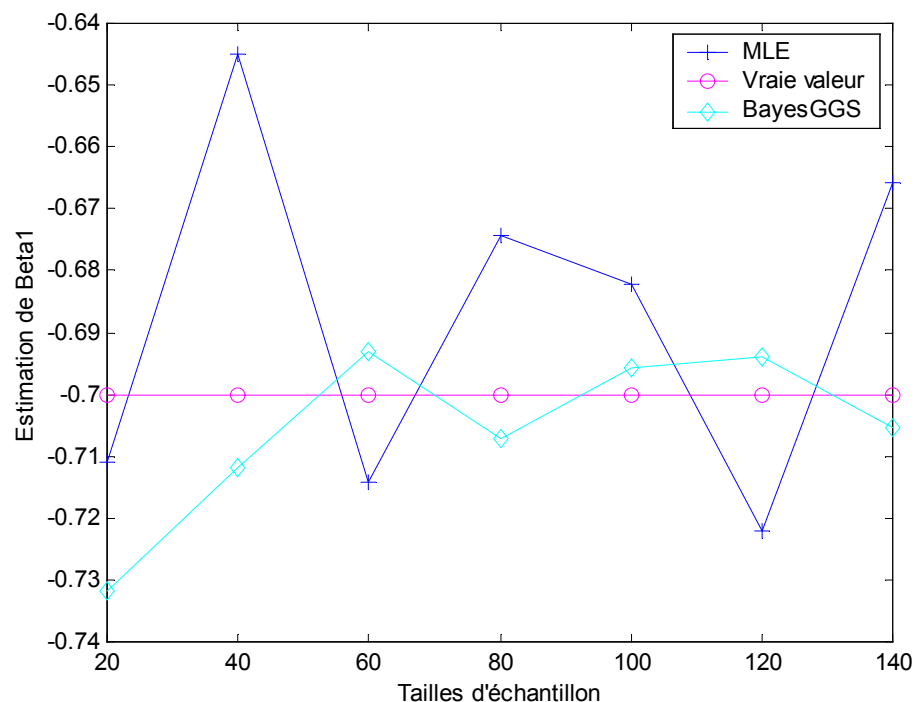
respectivement pour $\alpha_1 = 0.6$ et $\alpha_1 = 0.98$, la deuxième valeur étant proche de la zone de non stationarité dans le but de comparer les modes a posteriori obtenus grâce à l'algorithme du griddy gibbs sampling avec les estimateurs du maximum de vraisemblance moyennant l'algorithme du score. Nous voulons également étudier par simulation l'influence du paramètre α_1 lorsqu'il est proche de la zone de non stationnarité sur l'analyse bayésienne du modèle.

Le programme informatique a été écrit en langage Matlab 5.3 sous environnement Windows. Pour l'algorithme du griddy gibbs sampling nous avons pris un k égal à 15 qui joue un grand rôle dans la vitesse des calculs ; en effet, il est vrai que dans l'algorithme du Griddy Gibbs sampling, nous générons un échantillon de taille m , mais en vérité nous générons un échantillon de taille $m \times k$ seulement nous ne sélectionnons que les k ièmes, c'est à dire que si $k = 15$ et $m = 100$ (comme nous l'avons fait dans ce travail), nous avons généré non pas 100 estimateurs mais 1500 estimateurs (k est un entier représentant la suite de Gibbs). L'algorithme du griddy gibbs sampling est très coûteux contrairement à l'algorithme du score qui met en œuvre des relations linéaires.

Les paramètres sont mieux estimés par l'approche bayésienne, comme l'illustrent les figures suivantes, en particulier, lorsque la taille de l'échantillon est relativement petite, ($T=20, 40, 60$). Nous remarquons l'instabilité des estimations par la méthode du maximum de vraisemblance, moyennant l'algorithme du score, relativement au mode a posteriori. De plus, les modes a posteriori ont un RMSE (root mean square error) inférieur à celui du MLE (maximum likelihood estimator). Lorsque la taille de l'échantillon est grande, le MLE est meilleur. En effet, il est asymptotiquement efficace (in Engle (1982), Crowder (1978)).







Bibliographie

- BAUWENS L. et LUBRANO M., (1998), « Bayesian Inference on GARCH Models Using the Gibbs Sampler », 1, C23-C46.
- CASELLA G. et GEORGE E., (1992), « Explaining the Gibbs Sampler », *American Statistical Association*, 40, 167-174.
- ENGLE R.F., (1982), « Autoregressive Conditional Heteroscedasticity with Estimates of the Variance of U.K. Inflation », *Econometrica*, 50, 987-1008.
- ENGLE R.F., HENDRY D.F. et Trumble D., (1985), « Small Sample Properties of ARCH Estimators and Tests », *Canadian Journal of Economics*, 18, 66-93.
- GELFAND A. E. et SMITH A.F.M., (1990), « Sampling-Based Approches to Calculating Marginal Densitie », *Journal of the American Statistical Association*, 85, 398-409.
- GEMAN S. et GEMAN D., (1984), « Stochastic Relaxation, Gibbs Distributions, and the Bayesian Restoration of Images », *IEEE Transactions on Pattern Analysis and Machines Intelligence*, 6, 721-741.
- GEWEKE J., (1989), « Exact Predictive Densities for Linear Models with ARCH Disturbances », *Journal of Econometrics*, 40, 63-86.
- NAKATSUMA T.,(2000), « Bayesian Analysis of ARMA-GARCH Models : A Markov Chain Sampling Approach », *Journal of Econometrics*, 95, 57-69.
- PANTULA S.G., (1985), « Estimation of Autoregressive Models with ARCH Errors », MS. Department of Statistics, North Carolina State University.
- RIPLEY B.D., (1987), *Stochastic Simulation*, New York : John Wiley.
- RITTER C. et TANNER M.A., (1990), « The Griddy Gibbs Sampler », technical report, University of Rochester.
- Weiss A.A. ,(1986a), « Asymptotic Theory for ARCH Models : Estimation and Testing », *Econometric Theory*, 2, 107-113.

Sur l'estimation du paramètre d'un AR(1) avec innovations uniformes. Cas des échantillons finis.

Karima NOUALI et Hocine FELLAG

Département de mathématiques

Faculté des sciences. Université de Tizi-ouzou

Tizi-Ouzou. 15000. Algérie

k_nouali@hotmail.com

Résumé.

On considère le processus autoregressive d'ordre 1 avec des innovations uniformes. Nous proposons un estimateur modifié de l'estimateur de maximum de vraisemblance obtenu par Bell et Smith(1986) avec lequel nous améliorons l'estimation du paramètre du modèle dans le cas des échantillons finis.

Moyennant des techniques de Monté Carlo, une étude comparative d'estimateurs du paramètre du modèle est établie en présence d'innovation outlier (IO) et d'additive outlier (AO). A travers cette étude comparative, nous montrons l'impact d'un outlier (type AO où IO) sur les méthodes d'estimation du paramètre du modèle.

Mots-Clés: Autorégressive, Estimation, Outlier, Distribution Uniforme.

Classification AMS. 62F11, 62M10.

1 Introduction

On considère le processus autoregressive d'ordre 1 suivant

$$Y_t = \rho Y_{t-1} + \varepsilon_t \quad (1)$$

où les ε_t sont *i.i.d* de distribution $U(0, 1)$. ρ est le paramètre à estimer tel que $0 < \rho < 1$.

Bell et Smith (1986) étudient le problème d'estimation et de test sur le paramètre ρ du modèle (1). Des intervalles de confiance du paramètre ρ ont été proposé par Bickel et Doksum (1977) et Choi (1980).

On suppose ici, que l'état initial Y_0 suit la loi $U(0, 1/(1 - \rho))$.

Alors, on observe le segment

$$Y_1, Y_2, \dots, Y_n \quad n \quad \text{avec } n \text{ fixé}$$

Le processus (Y_t) est stationnaire d'ordre 1 avec

$$E(Y_t) = m = \frac{1}{2(1 - \rho)}, t = 1, 2, \dots$$

2 Les estimateurs du paramètre du modèle

2.1 Méthode du maximum de vraisemblance

L'estimateur de maximum de vraisemblance noté $\hat{\rho}_{MV}$ est défini comme suit

Bell et Smith (1986) montrent les propriétés suivantes de $\hat{\rho}_{MV}$.

Proposition 2.1 (Bell and Smith, 1986).

- 1) Le biais de l'estimateur $\hat{\rho}_{MV}$ est positif.
- 2) L'estimateur $\hat{\rho}_{MV}$ est convergent.

2.2 Méthode des moments.

En utilisant les moments asymptotiques suivants $E(Y_t) = \frac{1}{2(1-\rho)}$

$$V(X_t) = \frac{1}{12(1-\rho)^2}$$

Bell et Smith (1986) obtiennent par la méthode des moments l'estimateur noté $\hat{\rho}_{MM}$ du paramètre ρ défini comme suit

$$\hat{\rho}_{MM} = \max\left\{0, \frac{\bar{Y}^2 - 3S^2}{\bar{Y}^2 + 3S^2}\right\}$$

où

$$\begin{cases} \bar{Y} &= \frac{1}{n} \sum_{i=1}^n Y_i \\ S^2 &= \frac{1}{n} \sum_{i=1}^n (Y_i - \bar{Y})^2 \end{cases}$$

3 Le modèle non contaminé

3.1 Approximation du biais de $\hat{\rho}_{MV}$

En utilisant la méthode d'Anděl(1988) qui consiste à remplacer Y_{t-1} par sa moyenne, on obtient une approximation du biais, de la variance et de l'erreur quadratique moyenne (EQM) de l'estimateur $\hat{\rho}_{MV}$.

Proposition 3.1 Pour tout $n = 2, 3, \dots, \forall \rho$ tel que $0 < \rho < 1$

$$E(\hat{\rho}_{MV} - \rho) \simeq \frac{2(1-\rho)}{n}, \quad V(\hat{\rho}_{MV}) \simeq \frac{2}{n} \sqrt{\frac{n-1}{n+1}} (1-\rho) \text{ et } E(\hat{\rho}_{MV} - \rho)^2 \simeq \frac{8(1-\rho)^2}{n(n+1)}$$

Preuve

En utilisant la méthode d'Anděl, le biais de $\hat{\rho}_{MV}$ is approximé par

$$E(\hat{\rho}_{MV} - \rho) \simeq \frac{1}{m} E(\min_{2 \leq t \leq n} \varepsilon_t)$$

La variable aléatoire $S = \min_{2 \leq t \leq n} \varepsilon_t$ est de distribution Beta(1,n-1).

d'où

$$E(\hat{\rho}_Y - \rho) \simeq \frac{1}{nm} = \frac{2(1-\rho)}{n}$$

On peut écrire l'équation (1) sous cette forme

$$\frac{Y_t}{Y_{t-1}} = \rho + \frac{\varepsilon_t}{Y_{t-1}}$$

alors

où $Z \sim \text{Beta}(1, n - 1)$ et $E(Z^2) = \frac{2}{n(n + 1)}$

d'où

$$E(\hat{\rho}_{MV} - \rho)^2 \simeq \frac{1}{m^2} \frac{2}{n(n + 1)} \simeq \frac{8(1 - \rho)^2}{n(n + 1)}$$

et

$$V(\hat{\rho}_{MV}) \simeq (E(\hat{\rho}_{MV} - \rho)^2 - (E(\hat{\rho}_{MV} - \rho))^2)^{1/2} \simeq \frac{2}{n} \sqrt{\frac{n - 1}{n + 1}} (1 - \rho) \bullet$$

Remarques..

i) Pour un n donné, si $0 < \rho < 1$, Le biais maximal est approximativement $2/n$.

ii) Quand $n \rightarrow +\infty$, le biais, la variance et l'écart quadratique moyen approximatés de $\hat{\rho}_{MV}$ tendent vers zero.

4 Correction du biais approximaté de $\hat{\rho}_{MV}$

Le biais approximaté de $\hat{\rho}_{MV}$ est $E(\hat{\rho}_{MV} - \rho) \simeq \frac{2(1-\rho)}{n}$

On pose $\tilde{\rho}_{MV} = \frac{n\hat{\rho}_{MV} - 2}{n - 2}$

On obtient $E(\tilde{\rho}_{MV} - \rho) \simeq 0$

ce qui signifie que le biais approximaté de $\tilde{\rho}_{MV}$ est nul.

Remarques

(i). l'estimateur $\tilde{\rho}_{MV}$ est un estimateur asymptotiquement sans biais.

(ii). Pour n fini, Son biais est approximativement nul et négative.

Cela est illustré par la table suivante

| n | ρ | $\hat{\rho}_{MV}$ | $\tilde{\rho}_{MV}$ |
|-----|--------|-------------------|---------------------|
| 10 | 0.2 | 0.1407 | -0.0241 |
| | 0.4 | 0.1061 | -0.0174 |
| | 0.6 | 0.0729 | -0.0089 |
| 20 | 0.2 | 0.0739 | -0.0068 |
| | 0.4 | 0.0558 | -0.0046 |
| | 0.6 | 0.0374 | -0.0029 |

Table 2. Les valeurs simulées du biais de $\hat{\rho}_{MV}$ et $\tilde{\rho}_{MV}$.
100.000 répétitions.

5 Comparaison des estimateurs

En comparant l'estimateur $\tilde{\rho}_{MV}$ aux estimateurs $\hat{\rho}_{MV}$ et $\hat{\rho}_{MM}$, on remarque que $\tilde{\rho}_{MV}$ possède un biais et un EQM le plus faible(approximativement nuls) pour des échantillons de tailles modérées ce qui le rend plus recommandé en pratique(voir la table 3).

| ρ | L'estimateur | La moyenne | le biais | EQM |
|--------|---------------------|------------|----------|--------|
| 0.2 | $\hat{\rho}_{MV}$ | 0.2741 | 0.0741 | 0.0099 |
| | $\hat{\rho}_{MM}$ | 0.2389 | 0.0389 | 0.0241 |
| | $\tilde{\rho}_{MV}$ | 0.1934 | -0.0066 | 0.0055 |
| 0.6 | $\hat{\rho}_{MV}$ | 0.6374 | 0.0374 | 0.0025 |
| | $\hat{\rho}_{MM}$ | 0.6343 | 0.0343 | 0.0198 |
| | $\tilde{\rho}_{MV}$ | 0.5971 | -0.0029 | 0.0014 |

Table 3. n=20. 100000 répétitions.

6 Modèle avec contamination IO

Nous supposons que suite à des erreurs qui peuvent survenir, on observe le processus

$$X_t = \rho X_{t-1} + \varepsilon_t + \Delta \delta_{t,k} \quad n \text{ fixé} \quad (2)$$

où

$$\delta_{t,k} = \begin{cases} 1 & \text{si } t = k \\ 0 & \text{si } t \neq k \end{cases}$$

à la place du processus (1). Le processus (2) est appelé innovation outlier (Fox, 1972).

Dans le modèle (2), l'estimateur $\hat{\rho}_{MV}$ s'écrit comme suit

$$\hat{\rho}_{MV}(X) = \min\{X_2/X_1, X_3/X_2, \dots, X_n/X_{n-1}\} \quad (7)$$

où

$$X_t = Y_t \quad \forall t < k \quad \text{et} \quad X_t = Y_t + \rho^{t-k} \Delta \quad \text{pour } k \leq t \leq n$$

Le biais de $\hat{\rho}_{MV}(X)$ peut être approximé de la manière suivante

$$E(\hat{\rho}_{MV}(X) - \rho) \simeq$$

$$E(\min\{\varepsilon_2/m, \varepsilon_3/m, \dots, (\varepsilon_k + \Delta)/m, \varepsilon_{k+1}/(m + \Delta), \dots, \varepsilon_n/(m + \rho^{n-k-1} \Delta)\})$$

On pose $\delta_0 = \frac{-m + \sqrt{m^2 + 4m}}{2}$ et

$$W = \{\varepsilon_2/m, \varepsilon_3/m, \dots, (\varepsilon_k + \Delta)/m, \varepsilon_{k+1}/(m + \Delta), \dots, \varepsilon_n/(m + \rho^{n-k-1} \Delta)\}$$

On note $F_W(\cdot)$, la distribution de la variable W. L'expression de la fonction $F_W(\cdot)$ est assez complexe, pour cela nous proposons de donner l'expression de biais approximé $B_{n,m}(\Delta)$ de l'estimateur $\hat{\rho}_{MV}(X)$ pour $n = 3$.

Par exemple, pour $n = 3$ et $k = 2$, l'expression de l'approximation du biais de $\hat{\rho}_{MV}(X)$

$$(8) \quad B_{3,m}(\Delta) = \begin{cases} \frac{(\Delta^6 + 3m\Delta^5 + m(m-1)\Delta^4 + m^2(m-6)\Delta^3 + 3m^2(1-m)\Delta^2 + 3m^2(1+m)\Delta + 2m^3)}{6m^2(m+\Delta)^2} & \text{si } 0 \leq \Delta < \delta_0 \\ 1 & \text{si } \Delta \geq \delta_0 \end{cases}$$

Remarques

(i). $B_{3,m}(0) = \frac{1}{3m} = \frac{2(1-\rho)}{3}$.

(ii). $\lim_{\Delta \rightarrow +\infty} B_{3,\rho}(\Delta) = 0$.

Maintenant, pour chaque n , on peut utiliser l'expressions de $F_W(\cdot)$ pour obtenir le biais approximé de l'estimateur avec intégration numérique.

Nous illustrons cela par la table 4 dont on présente les valeurs simulées et approximées du biais pour $\rho = 0.5$ et $n = 3, 5, 10$. Les valeurs simulées du biais sont calculées à partir de l'expression de $\hat{\rho}_{MV}(X)$ (formule (7))

| Δ | $n = 3, k = 2$ | | $n = 5, k = 3$ | | $n = 20, k = 18$ | |
|----------|----------------|--------|----------------|--------|------------------|--------|
| | Approx. | Simul. | Approx. | Simul. | Approx. | Simul. |
| 0 | 0.333 | 0.353 | 0.200 | 0.218 | 0.050 | 0.049 |
| 0.2 | 0.365 | 0.373 | 0.214 | 0.233 | 0.051 | 0.051 |
| 0.6 | 0.312 | 0.331 | 0.189 | 0.205 | 0.050 | 0.049 |
| 1.0 | 0.250 | 0.281 | 0.161 | 0.174 | 0.048 | 0.048 |
| 5 | 0.083 | 0.087 | 0.063 | 0.065 | 0.035 | 0.035 |
| 10 | 0.045 | 0.046 | 0.036 | 0.036 | 0.025 | 0.025 |
| 20 | 0.023 | 0.024 | 0.019 | 0.019 | 0.015 | 0.015 |
| 50 | 0.009 | 0.009 | 0.008 | 0.008 | 0.007 | 0.007 |

Table 4. Les valeurs simulées et approximées du biais.
 $\rho = 0.5$, 100000 répétitions.

On peut remarqué que les valeurs simulées et approximées du biais sont pratiquement égales précisément quant Δ et n sont grands. Ce qui nous conduit à dire que l'approximation est valide.

Remarque

On note que la quantité

$$\min\{\varepsilon_2/m, \varepsilon_3/m, \dots, (\varepsilon_k + \Delta)/m, \varepsilon_{k+1}/(m + \Delta), \dots, \varepsilon_n/(m + \rho^{n-k-1}\Delta)\}$$

tend vers zéro quant Δ tend vers l'infini. Ce qui implique que le biais tend vers zéro quant Δ tends vers l'infini.

7 Modèle sous contamination AO

Maintenant, on suppose qu'observe le segment d'observations

$$Z_1, \dots, Z_n \quad n \text{ fixé}$$

du processus

$$Z_t = Y_t \quad \forall t \neq k \quad \text{et} \quad Z_k = Y_k + \Delta$$

à la place du processus du modèle (1).

L'estimateur du maximum de vraisemblance de ρ s'écrit comme suit

Comme dans la section 7, on peut approximer le biais de l'estimateur $\hat{\rho}_{MV}(Z)$ par $E(\hat{\rho}_{MV}(Z) - \rho) \simeq$

$$\frac{1}{m} E(\min\{\varepsilon_2, \varepsilon_3, \dots, \varepsilon_{k-1}, (\varepsilon_k + \Delta), \frac{(\varepsilon_{k+1} - (\frac{2m-1}{2m})\Delta)m}{(m + \Delta)}, \varepsilon_{k+2}, \dots, \varepsilon_n\})$$

On pose

$$T = \min\{\varepsilon_2, \varepsilon_3, \dots, \varepsilon_{k-1}, (\varepsilon_k + \Delta), \frac{(\varepsilon_{k+1} - (\frac{2m-1}{2m})\Delta)m}{(m + \Delta)}, \varepsilon_{k+2}, \dots, \varepsilon_n\}$$

Pour tout n , l'expression du biais approximer est

(10)

$$B_{n,m}(\Delta) = \begin{cases} \frac{1}{2nm^2 I_n} \left\{ \frac{(\Delta(2m-1))^n}{(2\Delta+2m)^{n-1}} \Phi_{m,n,\Delta}^{(1)} - n\Phi_{m,n,\Delta}^{(2)} - (1-\Delta)^{n-3} \Phi_{m,n,\Delta}^{(3)} \right\} & \text{si } 0 \leq \Delta < 1/2 \\ \frac{1}{2m^2} \left\{ -\frac{(\Delta(2m-1))^2}{4(m+d)} + \frac{1}{I_n} \left(\frac{(\Delta(2m+1))^{n-1}}{(2\Delta+2m)^{n-2}} - \Phi_{m,n,\Delta}^{(2)} \right) \right\} & \text{si } 1/2 \leq \Delta < \frac{2m}{2m-1} \\ \frac{m - (2m-1)\Delta}{2m(m+\Delta)} & \text{si } \Delta \geq \frac{2m}{2m-1} \end{cases}$$

avec

$$I_n = (n-1)(n-2)$$

$$\Phi_{m,n,\Delta}^{(1)} = 2n\Delta + 4m(n-1) + n - 2$$

$$\Phi_{m,n,\Delta}^{(2)} = (2m(n-1) - n + 3)\Delta + 2m(2-n)$$

$$\begin{aligned} \Phi_{m,n,\Delta}^{(3)} &= 2m(n-2) + \Delta(m(24n^2 - 8n - 12) + n - 4) \\ &\quad + \Delta^2(m(-10n - 12) + 12 - 4n) + \Delta^3(4m(1-n) - 21n - 12) - \Delta^4(2n - 4) \end{aligned}$$

Remarques.

- (i). $B_{n,m}(0) = 2(1 - \rho)/n$.
- (ii). $\lim_{\Delta \rightarrow +\infty} B_{n,m}(\Delta) = -\rho, \forall n$.
- (iii). $B_{n,m}(\Delta) = B_{n',m}(\Delta), \forall n, n', \forall \Delta \geq \frac{2m}{2m-1}$.
- (iv). $E(\tilde{\rho}_{MV}(Z) - \rho) \simeq -(\rho + \frac{2}{n-2})$.

Application.

La table 5 présente les valeurs approximées du biais de l'estimateur du maximum de vraisemblance obtenues de la formule(10) et les résultats simulés sont obtenus de la formule(9) pour $n = 3, 5, 20$ et $\rho = 0.5$.

| Δ | $n = 3, k = 2$ | | $n = 5, k = 3$ | | $n = 20, k = 15$ | |
|----------|----------------|--------|----------------|--------|------------------|--------|
| | Approx. | Simul. | Approx. | Simul. | Approx. | Simul. |
| 0 | 0.333 | 0.329 | 0.200 | 0.179 | 0.050 | 0.046 |
| 0.2 | 0.304 | 0.301 | 0.190 | 0.169 | 0.046 | 0.038 |
| 0.6 | 0.124 | 0.131 | 0.086 | 0.079 | 0.006 | 0.002 |
| 1.0 | 0.000 | -0.001 | -0.009 | -0.012 | -0.040 | -0.044 |
| 5 | -0.333 | -0.334 | -0.333 | -0.334 | -0.333 | -0.334 |
| 50 | -0.480 | -0.480 | -0.480 | -0.480 | -0.480 | -0.480 |
| 100 | -0.490 | -0.490 | -0.490 | -0.490 | -0.490 | -0.490 |

Table 5. Les valeurs approximées et simulées du biais.
 $\rho = 0.5, 100000$ répétitions.

Comme dans la section 7, on note que les valeurs approximées et simulées du biais sont pratiquement égales précisément quant Δ et n sont grands.

8 Comparaison des estimateurs sous contamination IO et AO

On note par $b_{\hat{\rho}}^{AO}$ et $b_{\hat{\rho}}^{IO}$, le biais de l'estimateur $\hat{\rho}$ sous contamination AO respectivement IO.

A travers la table 6.1 et la table 6.2, nous tirons les résultats suivants

- (i). Quand $\Delta \rightarrow +\infty, b_{\hat{\rho}}^{AO}$ et $b_{\hat{\rho}}^{IO}$ tendent vers $-\rho$.
- (ii). $b_{\hat{\rho}_{MV}}^{IO} > b_{\hat{\rho}_{MV}}^{AO}$; quant Δ augmente de façon significative.

On observe que l'effet d'un IO et d'un AO sur le biais de $\tilde{\rho}_{MV}$ est similaire . Les deux tables (6.1 et 6.2) montrent aussi que l'estimateur de maximum de vraisemblance possède un biais plus stable que celui de $\tilde{\rho}_{MV}$ et de $\hat{\rho}_{MM}$ en présence d'un IO où d'un AO pour de petites valeurs de Δ .

| Δ | $\hat{\rho}_{MV}$ | $\hat{\rho}_{MM}$ | $\tilde{\rho}_{MV}$ |
|----------|-------------------|-------------------|---------------------|
| 0 | 0.0741 | 0.0389 | -0.0066 |
| 0.2 | 0.0720 | 0.0432 | -0.0089 |
| 0.5 | 0.0614 | 0.0174 | -0.0207 |
| 1.0 | 0.0420 | -0.0748 | -0.0423 |
| 2.0 | 0.0021 | -0.1928 | -0.0865 |
| 5 | -0.1191 | -0.2000 | -0.2212 |
| 50 | -0.1572 | -0.2000 | -0.2635 |

Table 6.1. n=20, $\rho = 0.2$
100000 répétitions. AO contamination

| Δ | $\hat{\rho}_{MV}$ | $\hat{\rho}_{MM}$ | $\tilde{\rho}_{MV}$ |
|----------|-------------------|-------------------|---------------------|
| 0 | 0.0741 | 0.0386 | -0.0065 |
| 0.2 | 0.0757 | 0.0432 | -0.0048 |
| 0.5 | 0.0667 | -0.1482 | -0.0147 |
| 1.0 | 0.0449 | -0.2000 | -0.0390 |
| 2.0 | 0.0279 | -0.2000 | -0.0578 |
| 20 | 0.0069 | -0.2000 | -0.0812 |
| 50 | 0.0039 | -0.2000 | -0.0845 |

Table 6.2. n=20, $\rho = 0.2$
100000 répétitions. IO contaminations

9 Conclusion

Dans ce travail, nous obtenons une approximation du biais de l'estimateur de maximum de vraisemblance de paramètre AR(1). Une réduction du biais approximé de ce dernier conduit à l'obtention d'un estimateur du biais approximativement nul pour des échantillons finis. L'étude de simulation établie, nous révèle l'instabilité des méthodes d'estimation étudiées. Ce qui nous conduit à dire qu'un estimateur robuste de paramètre ρ reste à construire.

Références

- Anděl, J. (1988). *On AR(1) processes with exponential white noise*, Communication In Statistics, A, Theory and Methods, **17**, 5, 1481-1495
- Bell, C.B. and Smith, E.P. (1986). *Inference for non-negative autoregressive schemes*, Communication in statistics, theory and methods, **15**(8), 2267-2293.
- Bickel, P. and Doksum, K. (1977). *Mathematical statistics : Basic Ideas and Selected topics*. Wiley : New York.
- Choi, Y.J. (1980). *Kolmogorov-Smirnov test with nuisance parameters in uniform case*. M.S. Thesis, University of Washington.
- Fox, A.J. (1972). *Outliers in time series*, J. Roy. Stat. Soc. **34** (B), 350-363.
- Nouali K et H Fellag (2003). *Approximate bias for first-order autoregressive model with uniform innovations. Small sample case.*, Discussiones Mathematicae. Probability and Statistics **22**, 15-26.

Etude d'un modèle bilinéaire à coefficients périodiques

Abdelouahab Bibi
Département de Mathématiques,
Université Mentouri, Constantine
Algérie.
E-mail: bibi@wissal.dz

Résumé

Les modèles que l'on conviendra ici d'appeler *évolutifs* sont ceux où les coefficients peuvent évoluer au cours du temps. Dans le présent article, nous considérons la sous-classe des processus purement bilinéaires, strictement superdiagonaux et évolutifs notés $BLE(0, 0, p, 1)$, vérifiant $X_t = \sum_{i=2}^p b_i(t)X_{t-i}e_{t-1} + e_t$ où $(e_t)_{t \in \mathbb{Z}}$ constituent une suite de variables aléatoires indépendantes et identiquement distribuées selon la loi normale centrée réduite, et où $\forall i \in \{1, \dots, p\}$, $(b_i(t))_{t \in \mathbb{Z}}$ est une suite réelle dépendant d'un nombre fini de paramètres. On s'intéresse, dans un cadre probabiliste, à l'étude des propriétés théoriques de ces modèles. Cette étude, généralise les résultats obtenus par Guégan (1981) et de Francq (1999).

Mots-clés: Modèle bilinéaire évolutif, Moments d'ordre supérieur, Régularité, Inversibilité.

Classification AMS: 62M10, 60G10.

1 Introduction

Les techniques standards d'analyse des séries chronologiques (non stationnaires en général), reposent sur la propriété fondamentale de linéarité: la série doit posséder une structure linéaire de type autorégressif (AR), moyenne mobile (MA) ou autorégressif moyenne mobile ($ARMA$). A partir des années 70 on trouve un intérêt croissant pour ce genre de modèle surtout dans le domaine de l'économétrie et de finance. De nombreux travaux de recherche ont été développés pour justifier l'absence de stationnarité dans des données économétriques ou financières. Par exemple, dans le domaine de l'économétrie, des données individuelles temporelles (appelées parfois "panels") des recherches montrent que l'hypothèse habituelle d'invariance des coefficients au cours du temps d'un modèle de séries chronologiques est souvent intenable et par conséquent, le recours aux modèles linéaires évolutifs est inévitable. Excellente bibliographie récente sur le sujet, peut-être trouvée dans Bibi et Francq (2003).

Récemment, la classe de modèles bilinéaires évolutifs à connue un intérêt particulier (cf. Subba Rao (1997), Bibi et Oyet (2002), Bibi (2003)). Cette classe de modèles évolutifs vérifient une équation aux différences de type

$$X_t = \sum_{i=1}^p a_i(t)X_{t-1} + \sum_{j=1}^q c_j(t)e_{t-j} + \sum_{i=1}^P \sum_{j=1}^Q b_{ij}(t)X_{t-i}e_{t-j} + e_t$$

où $(a_i(t))_{1 \leq i \leq p}$, $(c_i(t))_{1 \leq i \leq q}$, et $(b_{ij}(t))_{1 \leq i \leq p, 1 \leq j \leq q}$ sont des suites bornées définies sur $\mathbb{Z} = \{0, \pm 1, \pm 2, \dots\}$ à valeurs réelles et où $(e_t)_{t \in \mathbb{Z}}$ constituent une suite de variables aléatoires indépendantes et identiquement distribuées selon la loi normale centrée réduite. De plus, on suppose que e_t et X_s sont indépendants pour tout $s < t$. Ces modèles sont linéaires par rapport aux e_t ou aux X_t séparément, ils sont une extension naturelle des processus *ARMA* évolutifs. Dans le présent article, nous allons hybrider la non stationnarité et la non linéarité en considérant le modèle bilinéaire évolutif "simple" noté *BLE* $(0, 0, p, 1)$ défini par

$$X_t = \sum_{i=2}^p b_i(t) X_{t-i} e_{t-1} + e_t \quad (1.1)$$

où p est un entier naturel supérieur ou égal à 2 connu, et pour chaque $i \in \{2, \dots, p\}$ la suite $(b_i(t))_{t \in \mathbb{Z}}$ varie dans le temps selon deux régimes $b_i(t) = b_{i,1} \mathbb{I}_{\Delta^{(1)}}(t) + b_{i,2} \mathbb{I}_{\Delta^{(2)}}(t)$ où $\Delta^{(1)} = \{0, \pm 2, \pm 4, \dots\}$, $\Delta^{(2)}$ son complémentaire et où $\mathbb{I}_{\Delta^{(i)}}(t)$ désigne une indicatrice traditionnelle: $\mathbb{I}_{\Delta^{(i)}}(t) = 1$ si $t \in \Delta^{(i)}$ et 0 sinon. Le modèle (1.1) peut-être représenté sous la forme vectorielle $X(t) = D(t)X(t-1) + C e_t$ et $X_t = C' X(t)$ où $X(t) = (X_t, X_t, \dots, X_{t-p+1})'_{p \times 1}$, $C = (1, 0, \dots, 0)'_{1 \times p}$, $D(t) = A + B(t)e_{t-1}$, $B(t) = B_1 \mathbb{I}_{\Delta^{(1)}}(t) + B_2 \mathbb{I}_{\Delta^{(2)}}(t)$ avec A et $B(t)$ sont deux matrices appropriées et où l'accent désigne la transposition. L'objet de l'article est l'étude des propriétés théoriques du modèle (1.1). Ces propriétés reposent sur la structure non linéaire des espaces de solutions des équations aux différences et sur le comportement asymptotique de ces solutions. Elles constituent, bien entendu, une extension des propriétés de quelques modèles non linéaires étudiés par, Guégan (1981), Francq (1999), et Terdik (2000). Dans toute la suite, \mathcal{M} désigne l'espace vectoriel fermé engendré par $(X_t)_{t \in \mathbb{Z}}$. Tout élément de \mathcal{M} est alors, ou bien combinaison linéaire d'un certain nombre fini des X_t avec des coefficients constants ou non, ou bien limite (en moyenne quadratique) d'une telle combinaison. Pour $t \in \mathbb{Z}$, soit $\mathcal{M}_t(X)$ le sous espace linéaire de \mathcal{M} engendré par $\{X_s, s \leq t\}$. Notons également $\mathfrak{S}_t(e)$ la tribu engendrée par $\{e_s, s \leq t\}$. Enfin notons pour tout $i \in \{1, 2\}$, $\Gamma_i = A^{\otimes 2} + B_i^{\otimes 2}$ où \otimes désigne le produit de Kronecker, et $\rho(\Gamma)$ désigne la plus grande valeur propre en module de la matrice carrée Γ , et $A^{\otimes 2} = A \otimes A$. Rappelons ici qu'un processus est dit stable si sa variance est bornée.

2 Théorème d'existence de solution

Dans cette section, nous discutons le problème d'existence et d'unicité de solution stable non anticipative, i.e, $\mathfrak{S}_t(e)$ -mesurable de l'équation aux différences (1.1).

Proposition 1 *Une condition nécessaire et suffisante pour que (1.1) ait une unique solution stable non anticipative donnée par (2.1) est que $\lambda = \rho(\Gamma_1 \Gamma_2) = \rho(\Gamma_2 \Gamma_1) < 1$.*

Preuve (cf. Bibi et Moon-Ho (2004)).

Il est important de remarquer que la condition nécessaire et suffisante d'existence et d'unicité de solution stable donnée par la proposition précédente est manifestement basée sur la connaissance des valeurs propres de matrices d'ordre $p^2 \times p^2$. Les calculs deviennent volumineux lorsque p est suffisamment grand, ce qui rend délicat l'utilisation de cette condition. Remarquons tout d'abord que la variance $r_t(0)$ du processus $(X_t)_{t \in \mathbb{Z}}$ satisfait l'équation $r_t(0) = 1 + \sum_{i=2}^p b_i^2(t) r_{t-i}(0)$, cette équation aux différences linéaires d'ordre p admet une solution et une seule si et seulement si les racines du

polynôme $P_t(\lambda) = 1 - \sum_{j=2}^p \lambda^j b_j^2(t)$ sont strictement supérieures à 1 en module, ou encore, les polynômes $P_i(\lambda) = 1 - \sum_{j=2}^p \lambda^j b_{j,i}^2$, $i = 1, 2$ ne s'annulent pas dans le disque unité fermé du plan complexe. Cependant, Berlinet et Francq ont montré que cette condition est équivalente à $\sum_{j=2}^p b_{j,i}^2 < 1$, $i = 1, 2$. D'où la proposition suivante qui donne une condition suffisante très simple

Proposition 2 [Francq (1999)] *La condition*

$$\max \left\{ \sum_{i=2}^p b_{i,1}^2, \sum_{i=2}^p b_{i,2}^2 \right\} < 1 \quad (2.1)$$

est suffisante pour l'existence et l'unicité de solution stable non anticipative satisfaisant à (1.1).

Remarque 1 En posant, $R(t) = (r_t(0), \dots, r_{t-p+1}(0))'$, $A(t) = A_1 \mathbb{I}_{\Delta(1)}(t) + A_2 \mathbb{I}_{\Delta(2)}(t)$ où $(A_i)_{i=1,2}$ sont deux matrices uniquement déterminées par les coefficients $(b_{j,i}^2)_{2 \leq j \leq p, i=1,2}$. Avec ces notations on a $R(t) = A(t)R(t-1) + C$ et $r_t(0) = C'V(t)$. Cependant, la condition (2.1) est équivalente à la condition

$$\max \{ \rho(A_1), \rho(A_2) \} < 1. \quad (2.2)$$

Remarque 2 Sous la condition (2.2), on a

$$r_t(0) = \sum_{k=0}^{\infty} C' \left\{ \prod_{j=0}^{k-1} (A_1 \mathbb{I}_{\Delta(1)}(t-j) + A_2 \mathbb{I}_{\Delta(2)}(t-j)) \right\} C = \begin{cases} r^{(1)}(0), t \in \Delta^{(1)} \\ r^{(2)}(0), t \in \Delta^{(2)} \end{cases}$$

où $r^{(1)}(0) = C'(I - A_1 A_2)^{-1}(I + A_1)C$, $r^{(2)}(0) = C'(I - A_2 A_1)^{-1}(I + A_2)C$.

3 Existence des moments d'ordre supérieur

Dans cette section, nous établissons les conditions nécessaires et suffisantes sous lesquelles les moments d'ordre supérieur existent.

3.1 Cas où $p = 2$.

Dans ce cas le modèle (1.1) s'écrit $X_t = b_2(t)X_{t-2}e_{t-1} + e_t$ puisque le processus $(e_t)_{t \in \mathbb{Z}}$ est gaussien, alors soit $\gamma(k) = E\{e_t^{2k}\} = 1.3.5 \dots (2k-1)$. D'autre part, on vérifie aisément que pour tout $k \geq 0$:

$$E\{X_t^{2k+1}\} = 0 \text{ et que } E\{X_t^{2k}\} = \gamma(k) b_2^{2k}(t) E\{X_{t-2}^{2k}\} + \sum_{i=0}^{k-1} \frac{(2k)!}{(2i)!(2k-2i)!} b_2^{2i}(t) E\{X_{t-2}^{2i}\} \gamma(k-i) \gamma(i).$$

C'est une équation aux différences linéaires du second ordre en $E\{X_t^{2k}\}$, d'où

Proposition 3 soit $(X_t)_{t \in \mathbb{Z}}$ un processus stable généré par le modèle BLE $(0, 0, 2, 1)$, si $\gamma(k) \cdot \max(b_{2,1}^{2k}, b_{2,2}^{2k}) < 1$ alors le processus $(X_t)_{t \in \mathbb{Z}}$ admet des moments d'ordre $2k$ uniformément bornés.

Preuve Par vérification directe.

Corollaire 1 Soit $(X_t)_{t \in \mathbb{Z}}$, un processus stable généré par BLE $(0, 0, 2, 1)$, alors les conditions suivantes sont équivalentes

1. $\exists M > 0, \forall t \in \mathbb{Z} : E \{X_t^{2k}\} \leq M$
2. $\gamma(k). \max(b_{2,1}^{2k}, b_{2,2}^{2k}) < 1.$

3.2 Cas où $p > 2$

Remarquons tout d'abord que la représentation (1.1) peut-être transformée en la représentation

$$\begin{cases} \eta(t) = D(t+1)\eta(t-1) + \bar{C}e_t \\ X_t = C'\eta(t-1) + e_t \end{cases} \quad (3.1)$$

où $\bar{C}' = (0, 1, 0 \dots 0)_{1 \times p}$. La représentation (3.1) possède la propriété d'indépendance entre la variable aléatoire $\eta(t)$ et d'autre variables aléatoires d'indices strictement supérieur à t . D'autre part on vérifie aisément que

$$\eta_t^{\otimes k} = D^{\otimes k}(t+1)\eta^{\otimes k}(t-1) + \sum_{l=1}^{k-1} B_{k,l}(t)\eta(t-1)^{\otimes(k-l)}e_t^l + \bar{C}^{\otimes k}e_t^k \quad (3.2)$$

où $B_{k,l}(t)$ sont des matrices aléatoires définies récursivement par

$$\forall l \in \{1, \dots, k-1\} : \begin{cases} B_{k,l}(t) := \bar{C} \otimes B_{k-1,l-1}(t) + D(t+1) \otimes B_{k-1,l}(t) \\ B_{k,0}(t) := D^{\otimes k}(t+1), B_{k,k}(t) := \bar{C}^{\otimes k} \end{cases}$$

Notons $\forall t \in \mathbb{Z}, \forall k \geq 1 : M_k(t) = E \{\eta(t)^{\otimes k}\}$ et $m_k(t) = E \{X_t^k\}$, alors on a

$$m_k(t) = \sum_{j=0, j \text{ pair}}^k \frac{k!}{j!(k-j)!} (\bar{C}')^{\otimes(k-j)} M_{k-j}(t-1)\gamma(j)$$

D'autre part, d'après (3.2) on a $M_k(t) = E \{D^{\otimes k}(t+1)\} M_k(t-1) + \sum_{l=1}^{k-1} E \{B_{k,l}(t)e_t^l\} M_{k-l}(t-1) + \bar{C}^{\otimes k} E \{e_t^k\}$ c'est une équation vectorielle aux différences linéaires de premier ordre en $M_k(t)$, elle admet une solution si et seulement si $\forall t : \rho(E \{D^{\otimes k}(t)\}) < 1$. On vérifie que $D^{\otimes k}(t) = \sum_{l=0}^k D_{k,l}(t)e_{t-1}^l$ où $(D_{k,l}(t))_{0 \leq l \leq k}$ sont des matrices réelles définies récursivement par

$$D_{k,l}(t) = \begin{cases} A^{\otimes k}, l = 0 \\ D_{k-1,l-1}(t) \otimes B(t) + D_{k-1,l}(t) \otimes A, 1 \leq l \leq k-1 \\ B^{\otimes k}(t), k = l. \end{cases}$$

Si on s'intéresse aux moments d'ordre pairs, on a

Corollaire 2 Soit $(X_t)_{t \in \mathbb{Z}}$ un processus stable généré par le modèle (1.1) alors les conditions suivantes sont équivalentes

1. $m_{2k}(t)$ est uniformément bornés,
2. $\rho(E \{D^{\otimes 2k}(t)\}) < 1.$

4 Régularité, inversibilité

4.1 Régularité

On vérifie sans peine que le processus défini par (1.1) est un bruit blanc (au sens de la fonction de covariance), il s'écrit sous la forme $X_t = \xi_t$ où $(\xi_t)_{t \in \mathbb{Z}}$ est le processus des innovations linéaires du processus $(X_t)_{t \in \mathbb{Z}}$ défini par $\xi_t = X_t - P(X_t | \mathcal{M}_{t-1}(X))$ où $P(X_t | \mathcal{M}_{t-1}(X))$ désigne la projection orthogonale de X_t sur $\mathcal{M}_{t-1}(X)$. D'où

Proposition 4 *le processus défini par (1.1) est un processus de bruit blanc, il est régulier, c'est son propre processus d'innovation au sens de Wold-Cramer.*

4.2 Inversibilité

Le concept de l'inversibilité joue un rôle fondamental dans les applications statistiques. Dans la littérature plusieurs définitions ont été données (cf. Francq 1999), ces définitions, et dans le cas stationnaire, reposent sur l'hypothèse de l'ergodicité dûe au fameux théorème de Birkhoff-Neumann. Dans le cas non stationnaire, il existe des classes de processus dans lesquelles on peut formuler des théorèmes de types "ergodiques". La classe qui nous intéresse le plus est la suivante:

Définition 1 *Un processus $(X_t)_{t \in \mathbb{Z}}$ réel, centré, stable, de fonction de covariance $r_t(h) = E\{X_t X_{t+h}\}$ est dit quasi-stationnaire, s'il existe une fonction $r(\cdot)$ telle que pour tout $h \geq 0$, $\lim_{N \rightarrow \infty} \frac{1}{N} \sum_{t=1}^{N-h} r_t(h) = \lim_{N \rightarrow \infty} r_N(h) = r(h)$.*

L'intérêt de cette définition réside dans le fait que les fonctions $r_N(h)$ et $r(h)$ sont de types définies positives de sorte que lorsque le processus est stationnaire au second ordre la fonction $r(\cdot)$ n'est autre que la fonction de covariance du processus. Cependant un problème fondamental s'impose, c'est l'estimation de $r_N(h)$ et $r(h)$. L'estimateur naturel est la fonction de covariance empirique $\hat{r}_N(h) = \frac{1}{N} \sum_{t=1}^{N-h} X_t X_{t+h}$ qui est un estimateur sans biais de $r_N(h)$ et asymptotiquement sans biais de $r(h)$. Remarquons que le processus $(X_t)_{t \in \mathbb{Z}}$ généré par (1.2) vérifiant la condition (2.1), est un processus quasi-stationnaire, en effet

$$r(h) = \lim_{N \rightarrow \infty} \frac{1}{N} \sum_{t=1}^{N-h} C' R_t(h) C = \begin{cases} 0, h \neq 0 \\ \frac{1}{2} \{r^{(1)}(0) + r^{(2)}(0)\}, h = 0. \end{cases} \quad (4.1)$$

Définition 2 *Le modèle (1.1) est dit inversible, si $\forall t \in \mathbb{Z}$, e_t est $\mathfrak{F}_t(X)$ -mesurable où $\mathfrak{F}_t(X)$ est la σ -algèbre engendrée par $\{X_s, s \leq t\}$.*

La proposition suivante donne une condition suffisante d'inversibilité pour notre classe de processus.

Proposition 5 *Soit $(X_t)_{t \in \mathbb{Z}}$ un processus généré par BLE $(0, 0, p, 1)$ admettant des moments d'ordre quatre uniformément bornés. Alors le processus est inversible si $\log \left(\sum_{i=2}^p b_i^2(i)r(0) \right) < 0$ où la quantité $r(0)$ est définie par (4.1).*

Preuve (cf. Bibi et Mon-Ho (2004)).

Corollaire 3 *Si le processus $(X_t)_{t \in \mathbb{Z}}$ généré par $BLE(0, 0, p, 1)$ est inversible alors le processus $(e_t)_{t \in \mathbb{Z}}$ admet la représentation*

$$e_t = \sum_{n \geq 1} (-1)^n \left\{ \prod_{j=0}^{n-1} C'B(t-j)X(t-j-1) \right\} X_{t-n} + X_t$$

où la série converge presque sûrement.

Preuve Conséquence de la proposition précédente.

References

- [1] Bibi, A and Oyet A. (2002). A note on the properties of some time varying bilinear models. *Statistics and probability letters* Vol. 58 (4) pp. 399-411.
- [2] Bibi, A (2002). On the covariance structure of the time-varying bilinear models. To appear in *Stochastic Analysis and Applications* Vol. 21 No.1 pp. 25-60.
- [3] Bibi, A and Francq, C. (2002). Consistent and Asymptotically normal estimators for cyclically time-dependent linear models. *Ann. Inst of Statis. Math.* Vol. 55, pp. 41-68.
- [4] Bibi, A. and Moon-Ho (2004). Properties of Some Bilinear Models With Periodic Regime Switching. Submitted for publication.
- [5] Francq, C. (1999). ARMA models with bilinear innovations. *Stochastic Models*, Vol. 15 pp. 29-52..
- [6] Guégan. D. (1981). Etude d'un modèle non linéaire, le modèle superdiagonal d'ordre 1. *C.R.A.S.* 293. Série I. pp. 95-98.
- [7] Terdik, G. (2000). Bilinear stochastic models and related problems of nonlinear time series, A frequency domain approach. *Lecture Notes in statistics* N124. Springer Verlag.

IV – Analyse de données et Applications

METHODOLOGIES STATISTIQUES DANS L'INDUSTRIE

*Ahmed Bousbaine
Centre de Recherche Nestlé
Vers-chez-les-Blanc - CP 44
CH-1000 Lausanne 26
Suisse*

Résumé

La statistique, en tant que science d'aide à la décision, est un outil très puissant pour le monde industriel. L'utilisation des méthodologies statistiques (plans d'expérience, analyses univariées et multivariées des données, modélisation, optimisation, ...) dans l'industrie, que ce soit au niveau recherche, production et marchés est d'une efficacité indéniable. De plus en plus, les grandes entreprises se dotent de compétences statistiques afin d'améliorer leur productivité et la qualité de leurs produits.

La statistique trouve ses applications dans la recherche pharmaceutique, la médecine, la recherche agro-alimentaire, la finance, l'environnement, les sciences sociales, les assurances ... Cette liste est loin d'être exhaustive. Dans cette présentation, nous nous focaliserons sur les méthodologies statistiques appliquées dans la recherche agroalimentaire ou le Statisticien est amené à collaborer avec des Physiciens, des Chimistes, des Biologistes, des Bioinformaticiens, des Ingénieurs, des Microbiologistes et des Nutritionnistes. Le but est de mettre en évidence certaines facettes du métier de Statisticien.

1. INTRODUCTION

La Statistique, branche des Mathématiques, est de plus en plus nécessaire dans le monde industriel.

Elle se retrouve à tous les niveaux : recherche, production, contrôle de qualité, études consommateurs ... Les entreprises industrielles se rendent compte maintenant (!) de la puissance des outils statistiques dans leurs activités de tous les jours. Ces entreprises engagent donc des Statisticiens pour optimiser leur productivité au sens large du terme.

Cependant, il serait utopique d'espérer que le statisticien, à lui tout seul, puisse résoudre tous les problèmes qui se posent à une entreprise. Le statisticien n'est qu'un maillon de la chaîne. Le résultat découlant d'une analyse statistique est un résultat parmi tant d'autres. Et c'est la mise en commun de tous ces résultats qui permettra de trouver une solution efficace au problème posé. Faut-il rappeler que la statistique n'est pas la décision, mais un outil d'aide à la décision ?!

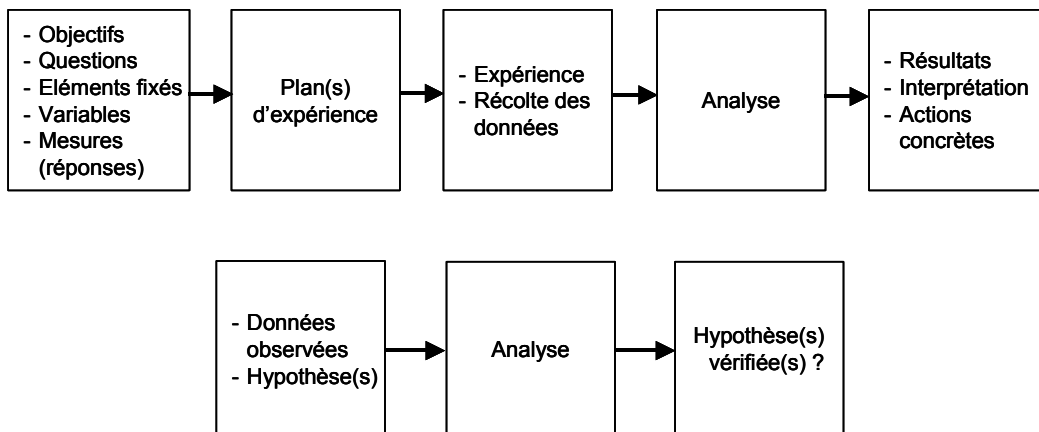
Par ailleurs, il n'y a pas de statistique appliquée sans données. Evidemment, il ne faut pas oublier la statistique théorique, mais tel n'est pas le sujet de cette présentation. Revenons, pour le moment, à la statistique appliquée. Les données sont récoltées dans le but de répondre à certaines questions que l'on se pose. On peut dire aussi que les données récoltées serviront à confirmer ou infirmer une hypothèse que l'on a formulée et qui constitue la base du problème posé.

Le ou les résultats obtenus après analyse de ces données récoltées ne suffisent pas en eux-mêmes à résoudre le problème. Après une analyse statistique, que signifierait un résultat du style : la p-value (d'un certain test que l'on aurait effectué) est égale à 0.03 ?

La réponse est qu'un tel résultat, pris en dehors du contexte du problème, ne signifie rien. Dans le monde industriel, il faut revenir au problème réel et ainsi transformer les résultats obtenus en des actions concrètes. Il va de soi que ces actions sont basées non seulement sur les résultats statistiques, mais aussi sur des informations externes et complémentaires.

2. APPROCHE STATISTIQUE

On peut schématiser l'approche statistique des problèmes industriels de la manière suivante :



3. ANALYSE DU PROBLEME

Dans les 2 schémas ci-dessus, la partie ayant trait aux objectifs, questions, ... ou aux hypothèses est essentielle dans la suite du processus. En effet, c'est à ce moment-là que le statisticien doit déjà être impliqué afin de clarifier certains points tels que :

- Les objectifs sont-ils clairs et mesurables ?
- Quelles sont les questions que l'on se pose ?
- Qu'est-ce que l'on considère comme fixé ?
- Quels sont les paramètres (facteurs) que l'on peut faire varier ?
- Dans quelles limites peut-on les faire varier ?
- Y a-t-il eu des essais préliminaires ?
- Les facteurs que l'on veut faire varier sont-ils tous indépendants ?
- Quelles sont les mesures que l'on aimerait effectuer sur les unités expérimentales ?
- Prévoit-on des répétitions ? Si oui, lesquelles ?

Cette analyse doit se faire avec toutes les personnes impliquées dans le projet.

4. PLANS D'EXPERIENCES

Un plan d'expérience est une technique statistique qui permet d'obtenir le maximum d'information avec un minimum d'effort. Il permet à l'expérimentateur de réaliser ses essais d'une manière structurée et dans un ordre bien défini.

Une fois l'analyse du problème terminée, on passe à l'élaboration du ou des plans d'expériences.

Les plans d'expériences couramment utilisés sont :

- Les plans factoriels complets ou fractionnaires. Ces plans peuvent être construits très facilement. Ce qui est moins facile est de construire des plans fractionnaires où les facteurs n'ont pas le même nombre de niveaux. Notre groupe a travaillé dans ce domaine et a déjà établi un catalogue de plus de 150 plans (dans le cadre d'une thèse de Doctorat).
- Les plans pour surface de réponses et les plans de mélange. Ces plans sont utilisés à des fins d'optimisation. Il est clair que les techniques de modélisation sont utilisées à ce stade (nous y reviendrons plus loin sur la modélisation). En général, ces plans sont relativement simples à établir.
- Les plans d'expériences, avec un minimum d'essais, comportent des facteurs de procédé et des facteurs de mélange. A notre connaissance, bien peu d'investigations ont été faites dans ce domaine.

5. EXPERIENCE ET RECOLTE DES DONNES

L'expérience proprement dite et la récolte des données sont du domaine de l'expérimentateur.

L'expérimentateur doit s'assurer que l'expérience a été menée dans des conditions optimales afin d'éviter que d'innombrables sources de variabilité viennent entacher les données et par conséquent grossir l'erreur expérimentale et ainsi occulter un effet significatif.

Une autre tâche qui revient à l'expérimentateur est de s'assurer que les données récoltées ne comportent pas d'erreurs dues à différentes raisons (mauvaise lecture, mauvais enregistrement des données, données "insensées", ...).

6. ANALYSE STATISTIQUE

Dans les deux schémas du paragraphe 2, il y a aussi la partie analyse. Cette partie analyse est souvent considérée comme la partie technique du processus. C'est la partie où l'on "joue" avec les chiffres. Mais c'est aussi la partie où la méthode (ou la méthodologie) utilisée pour analyser les données doit être choisie avec soin. N'oublions pas que le but est de répondre aux questions que l'on s'est posées. A cette étape du processus, le statisticien va puiser dans le vivier des méthodes existantes, les utiliser directement, en construire des méthodologies, les adapter au problème posé ou carrément développer une nouvelle méthode "sur mesure".

Ce vivier de méthodes ou de méthodologies existantes provient de la recherche théorique en statistique qui est reportée dans les différents types de publications. Un autre outil, très important à cette étape d'analyse, est l'informatique.

En effet, l'utilisation intensive de la statistique dans le monde industriel n'a été possible que grâce à l'implémentation d'un grand nombre de méthodes sous forme de logiciels. Il serait impossible d'essayer d'effectuer une analyse statistique sur un jeu de données comportant des centaines (voire des milliers) de lignes et des dizaines (voire des centaines) de colonnes sans l'utilisation d'un ordinateur.

Cependant, si l'informatique a permis à la statistique d'émerger et de s'épanouir, elle en a fait aussi une pseudo-science où toute personne sachant manipuler un ordinateur s'autoproclame, du jour au lendemain, statisticien !

Il faut bien se rendre compte qu'un jeu de données soumis à un logiciel statistique donnera toujours un résultat. La question est de savoir si ce que l'on a fait est correct d'une part et permet de répondre aux questions posées d'autre part.

Par ailleurs, le statisticien, fort de ses logiciels statistiques qui lui fournissent des résultats rapidement, oublie souvent de prendre le temps de réfléchir aux résultats qu'il a obtenus.

Ce manque d'autocritique peut amener à tirer des conclusions non seulement farfelues, mais qui pourraient s'avérer très dangereuses (par exemple dans le cas de la recherche pharmaceutique ou dans le cas de sécurité des produits alimentaires pour ne citer que ces deux cas).

Il y a un domaine de la statistique qui est beaucoup utilisé dans le monde industriel; il s'agit de la modélisation statistique à côté des méthodes classiques telles que :

- La statistique descriptive (très importante, mais peu utilisée).
- L'analyse de variance (ANOVA).
- La régression.
- Les méthodes multivariées (PCA, PCR, ...).
- Etc...

7. MODELISATION STATISTIQUE

La modélisation statistique est un terme qui veut dire beaucoup de choses.

Probablement, le modèle statistique le plus simple est l'ajustement d'une droite à un jeu de données bivariées (régression linéaire simple). Déjà à ce niveau, avant d'établir une telle relation ($y = a+bx$), on peut se poser beaucoup de questions telles que :

- Cela fait-il sens d'établir une telle relation ?
- Y a-t-il une relation de cause à effet entre y et x ?
- Le modèle, même s'il semble correct, l'est-il vraiment (analyse des résidus, ...) ?
- Les deux coefficients a et b ont-ils une signification ?
- Etc...

Pour des modèles plus compliqués, les mêmes questions se posent et, en plus, on distingue deux cas selon l'objectif que l'on s'est fixé. L'objectif pourrait être tout juste de répondre à la question "comment est liée la réponse y aux variables explicatives x ?" et ainsi d'établir un modèle qui nous permet de faire de la prédiction et/ou de l'optimisation. Un autre objectif, qui va un peu plus loin que le premier, est de répondre à la question "pourquoi ?". Dans ce cas, évidemment, la tâche est plus complexe. En effet, comme il faut donner une interprétation aux différents coefficients,

on ne peut pas se permettre juste d'ajuster le meilleur modèle. C'est la raison pour laquelle, parfois, on se contente d'un modèle de moindre qualité mais qui permet de donner une interprétation du phénomène sous-jacent.

8. INTERPRETATION ET ACTIONS

La partie interprétation des résultats est une partie cruciale. Il ne s'agit pas seulement de présenter des tableaux de chiffres et des graphiques, mais surtout de revenir au problème réel posé et de tirer de ces résultats statistiques des actions concrètes. Cette étape doit se faire de concert avec toutes les personnes impliquées dans le projet. La décision doit se prendre non seulement sur la base des résultats des analyses des données de l'expérience, mais elle doit faire intervenir des informations externes (coût, faisabilité ...). En dehors de toutes ces considérations, il y a un problème que l'on rencontre souvent dans le monde industriel : le cas de plusieurs réponses. A l'étape d'optimisation, il ne s'agit pas d'optimiser sur une réponse, mais sur plusieurs. Ce qui en fait un problème de "multi-réponse optimisation". Pour résoudre ce type de problème, on trouve certaines techniques dans la littérature. Notre avis là-dessus est que ces techniques ne sont pas simples à utiliser d'une part, et qu'elles manquent d'une certaine objectivité d'autre part.

Les discussions entre le statisticien et les autres personnes impliquées dans le projet permettent d'arriver à une solution pratique, compréhensible et applicable.

9. EXEMPLES

Chacune des parties sera illustrée par un ou plusieurs exemples pratiques.

Ces exemples pratiques permettront de mettre en lumière les problèmes rencontrés et la manière dont ils ont été résolus.

10. CONCLUSION

La statistique permet d'approcher les problèmes industriels d'une manière structurée. Cette approche permet aux entreprises d'acquérir un savoir-faire utile à court et long terme et ceci à moindre coût. Le statisticien dans le monde industriel se doit d'acquérir des connaissances multidisciplinaires afin de dialoguer efficacement avec ses collègues ayant des formations différentes de la sienne. Il doit intervenir au début du projet et ainsi il évitera de jouer le rôle de pompier. Il doit à tout moment faire des Mathématiques, mais sans le dire !!!

Le statisticien doit avoir une solide formation de base en mathématiques, en statistique, en probabilité, ainsi que dans les domaines de l'optimisation. Le statisticien doit se tenir au courant des différentes publications afin d'adapter ou de développer de nouvelles méthodologies qui lui permettront de résoudre efficacement les problèmes posés. En plus de ces connaissances scientifiques, travailler dans l'industrie comme statisticien nécessite une grande faculté d'écoute, une bonne communication et un zeste de diplomatie.

11. BIBLIOGRAPHIE

- Bates D.M., Watts D.G. "Nonlinear Regression Analysis & its Applications", Wiley, 1988.
- Box G.E.P., Draper N.R. "Empirical Model-Building and Response Surface", Wiley, 1987.
- Box G.E.P., Hunter W.G., Hunter J.S. "Statistics for Experimenters: An Introduction to Design, Data Analysis, and Model Building", Wiley, 1978.
- Chambers J.M., Cleveland W.S., Kleiner B., Tukey P.A. "Graphical Methods for Data Analysis", Duxbury Press, 1983.
- Cochran W.G., Cox G.M. "Experimental Design", Wiley, 1957.
- Alonso Rodriguez C. "Aportaciones a la Construcción de Diseños Experimentales Factoriales Fraccionarios", (Thèse, Universidad de Alcalá, Madrid), 2002.
- Dey A. "Orthogonal Fractional Factorial Designs", Wiley, 1985.
- Draper N.R., Smith H. "Applied Regression Analysis", Wiley, 1981.
- Khuri A.I. Cornell J.A. "Response surfaces, Designs and Analysis", Marcel Dekker Inc., 1987.
- Mason R.L., Gunst R.F., Hess J.L. "Statistical Design & Analysis of Experiments", Wiley, 1989.
- Searle S.R. "Linear Models for Unbalanced Data", Wiley, 1987.
- Seber G.A.F. "Multivariate Observations", Wiley, 1984.
- Wang C. "Sense and Nonsense of Statistical Inference", Marcel Dekker Inc., 1993.

Shannon's Entropy for Network Optimization

Denis Pomorski

L.A.G.I.S. - UMR CNRS 8021

Building P2 - Cité Scientifique - Université des Sciences et Technologies de Lille

59655 Villeneuve d'Ascq cedex, France

denis.pomorski@univ-lille1.fr

Abstract - This contribution deals with binary detection networks optimization using an entropy-based criterion, from *a priori* probabilities and conditional PDF. The reduction of communication costs is the key focus of distributed detection networks. On the contrary, their performance is lower compared to centralized networks, because the fusion center does not receive all the available information to make the final decision. The optimization of a parallel distributed network with N sensors leads to a set of $2^N + N$ nonlinear equations, that can only be solved in particular cases, assuming statistical independence of the local observations, and for small-sized networks. In order to reduce the number of equations to solve, Shannon's entropy is used to select relevant sensors for the decision process. Once the relevant sensors are selected, assuming conditional independence of the observations, we demonstrate that the optimization of an elementary component of detection consists in applying a variable threshold on the likelihood ratio, which depends on *a posteriori* probabilities. A gradient algorithm is proposed to find this threshold. The optimization results of the elementary component of detection using entropy and Bayes' criteria are compared : the proposed approach is characterized by an interesting property of robustness with respect to rare events, and with respect to events for which *a priori* probabilities are uncertain. In particular, the obtained ROC curve does not recede from the ideal point.

Keywords : Data fusion, Information fusion, binary detection, centralized/distributed networks, optimization, Shannon's entropy.

1 Introduction

Detection theory deals with the problem of discriminating a finite number of situations from N -sensor signals. Within the binary detection framework where H (taking values H_0 or H_1) is the true hypothesis and u_0 (taking values 0 or 1) is the final decision, several networks are usually optimized :

- centralized detection network : all the sensors broadcast their observations to a central detector which elaborates the final decision (figure 1) ;

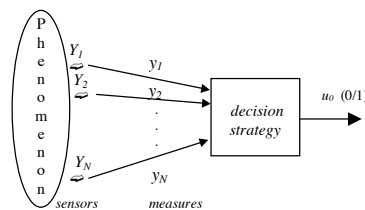


FIG. 1 – A centralized detection network

- parallel distributed detection network : each sensor is associated with a binary detector. Local decisions are broadcasted to a fusion center which combines them to take the final decision (figure 2) [12] [8] ;
- serial distributed detection network : each local detector receives information from one sensor and broadcasts its decision to its following detector. The first detector decision is based on information from a single sensor and the last detector elaborates the final decision (figure 3) [9] [11] [10].

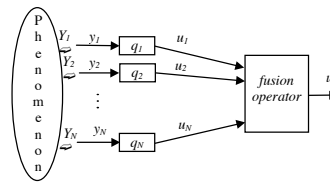


FIG. 2 – A parallel distributed detection network

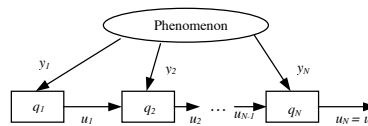


FIG. 3 – A serial distributed detection network

A *priori* probabilities P_i ($i = 0, 1$) and conditional PDF $p(y_1, \dots, y_N/H_i)$ ($i = 0, 1$) for each hypothesis are assumed known. Furthermore, $P_D = P(u_0 = 1/H_1)$ (resp. $P_F = P(u_0 = 1/H_0)$) is the detection (resp. false alarm) probability associated with the whole network.

In the literature, Bayes' or Neyman-Pearson criterion [13] is used to optimize detection networks. In the Bayesian case, a cost is associated with each situation and a risk function is minimized. In the applications where these costs are known and have a precise meaning, the Bayesian approach provides good results. Neyman-Pearson criterion consists in searching the best detection probability for a given false alarm probability.

However, in many instances (in particular when *a priori* probabilities of the hypotheses are uncertain or when one of them is close to zero), it could be interesting to frame the optimization problem by considering the quantity of information relevant to the detection problem. Several studies have considered information-based criteria; Middleton [5] and Gabrielle [3] proposed a centralized detection network based on information loss minimization between the input and the output of the system. Hoballah [4] and Warren [14] extended this result to a parallel distributed detection, considering independent observations, but no comparison with Bayes' criterion was performed, and no examples were presented.

The present work deals with both distributed and centralized network optimization. Let consider for instance the supervision of a complex industrial plant with a large number of sensors which produce digital or/and symbolic information. The optimization of such systems using every sensor is very difficult. In certain cases, sensors introduce redundancy without performance improvement, or they can flood operators with tremendous information. Hence, a sensor selection procedure is introduced (section 2) : among all the available sensors, only those which provide a significant piece of information to the decision process are considered.

Once the relevant sensors are selected, we propose to define (section 3) the elementary component of detection as a *black box* with $p + k$ inputs (p observations and k decisions) and one output u which represents the decision taken by the component.

In section 4, we propose to minimize information loss on H knowing the final decision u_0 . This quantity can be estimated by Shannon's conditional entropy $h(H/u_0)$, which represents the uncertainty on H knowing the decision u_0 [1]. The minimization of $h(H/u_0)$ allows for the determination of a threshold on the likelihood ratio. This threshold is determined from the knowledge of *a posteriori* probabilities associated with the different possible situations. Assuming conditional independence of the observations, the optimization of a complex network can be performed person-by-person.

In section 5, a gradient-based technique is proposed to find the optimal threshold λ^* of an elementary component, and then, P_D^* and P_F^* , from the entropy-based criterion. Finally, a comparison is performed between the entropy-based criterion and Bayes' criterion on a simple detection component. We underline a very interesting robustness property of the entropy-based approach with respect to rare events, and with respect to events for which *a priori* probabilities are uncertain.

2 Selection of relevant sensors

In this section, an entropy-based criterion is used for the selection of a sensor subset that is relevant for a distributed or centralized network. From the notion of multi-dimensional variable coarseness, a top-down hierarchical approach is proposed. It allows to reject the non-informative or redundant sensors for the decision process.

2.1 Problem statement

The goal of selection is to determine a reduced configuration which keeps relevant information from the whole configuration $y = (y_1, \dots, y_N)$. Therefore, it is necessary to determine the sub-vector $(y_{i_1}, \dots, y_{i_p}) \in P(y)$ (where $i_j \in \{1, \dots, N\}$ and $P(y)$ is the power set of $\{y_1, \dots, y_N\}$) such that the uncertainty of H given $(y_{i_1}, \dots, y_{i_p})$ is close to the uncertainty of H given entire measurement vector (y_1, \dots, y_N) , that is :

$$h(H/(y_{i_1}, \dots, y_{i_p})) = E[-\log_2 P(H/(y_{i_1}, \dots, y_{i_p}))] \sim h(H/(y_1, \dots, y_N)) \quad (1)$$

A compromise between performance and complexity of the network is then found.

Furthermore, the developed methods have to be simple in order to be implemented on networks with a large number of sensors.

A first stopping condition of the selection algorithm considers the number of sensors "p" which can be very low. Another stopping condition consists of determining a threshold of maximal conditional entropy h_{\max} above which the amount of information becomes meaningless. The difference $h_{\max} - h(H/(y_1, \dots, y_N))$ corresponds to the error allowed by taking into account p of N sensors.

If there is no constraint on p , it is sufficient to consider $p = N$. If it is impossible to define h_{\max} as a stopping condition, it is sufficient to consider $h_{\max} = h(H/(y_1, \dots, y_N))$.

2.2 Detection and the multi-dimensional variable coarseness

In a centralized detection network, the use of Bayes' or Neyman-Pearson criteria leads to the same optimal detector structure; the entire information produced by the N sensors Y_i ($i = 1, \dots, N$) is broadcast to a decision operator which makes the final decision u_0 by comparing the likelihood ratio of each observation $y = (y_1, \dots, y_N)$ with a threshold λ , that is :

$$\Lambda(y) \underset{u_0=0}{\overset{u_0=1}{\geq}} \lambda \quad \text{where} \quad \Lambda(y) = \frac{p(y/H_1)}{p(y/H_0)} \quad (2)$$

The final decision variable u_0 can thus be considered as coarser than $\Lambda(y)$ because u_0 makes a partition of the observation space more coarsely than $\Lambda(y)$, which itself is also coarser than y . Therefore, the following relation is obtained : $u_0 \succ \Lambda(y) \succ y$ (figure 4) where " \succ " represents the relation "...is coarser than...".

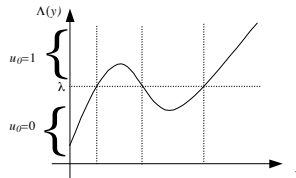


FIG. 4 – u_0 is a variable coarser than $\Lambda(y)$, which itself is coarser than y .

Taking into account both conditional entropy properties and the above relation, the following relation can be stated¹ :

$$h(H/(y_1, \dots, y_N)) = h(H/\Lambda(y_1, \dots, y_N)) \leq h(H/u_0) \leq h(H) \leq \log_2 2 = 1 \quad (3)$$

For a centralized detection network, $h(H/y) = h(H/\Lambda(y))$ since $\Lambda(y)$ is a sufficient statistic. The same reasoning can be applied to the parallel distributed detection problem :

¹The entropy of a variable with n values is less than or equal to $\log_2(n)$.

$$u_0 \succ (u_1, \dots, u_N) \succ (\Lambda(y_1), \dots, \Lambda(y_N)) \succ (y_1, \dots, y_N) \quad (4)$$

and we obtain :

$$h(H/(y_1, \dots, y_N)) \leq h(H/\Lambda(y_1), \dots, \Lambda(y_N)) \leq h(H/u_1, \dots, u_N) \leq h(H/u_0) \leq 1 \quad (5)$$

2.3 What kind of method ?

In practice, the exhaustive calculation of $h(H/(y_{i_1}, \dots, y_{i_p}))$ for each $(y_{i_1}, \dots, y_{i_p}) \in P(y)$ is not suitable for a large number of sensors, and some heuristics have to be used :

- from any sub-vector of p sensors, a first approach consists of applying a transfer procedure : Remove the least informative sensor, then add the most informative sensor among the sensors which are not taken into account. The process is repeated until stabilization.
- a bottom-up hierarchical approach consists of selecting the most informative sensor, then the sensor which, when associated with the first one, provides the most information, and so on... until the stopping condition(s) is (are) met. This approach does not appear to be appropriate for our problem. Indeed, it is possible that the explanatory power of each variable when taken individually, is low, while the explanatory power of the vector S , when considered globally, is large. In that case, an algorithm taking into account the variables "one by one", or not taking into account the whole explanatory power of the multi-dimensional variable S , cannot underline certain relationships.

Therefore, heuristics which allow to reject the non-informative or redundant sensors for the decision process is proposed by using a top-down hierarchical approach.

2.4 Selection of relevant sensors

This method consists of searching for y_{i_j} ($y_{i_j} \in S_{k-1}$) which minimizes $h(H/(S_{k-1} \setminus y_{i_j}))$, with $S_0 = y$ and $S_k = S_{k-1} \setminus y_{i_j}$; where S_0 is the initial sensor vector, and S_k , the sensor vector obtained at step k of the algorithm (figure 5). The algorithm stopping conditions are :

- the current vector is constituted by p sensors which represent the p most relevant sensors for the decision problem or
- the maximal conditional entropy is reached.

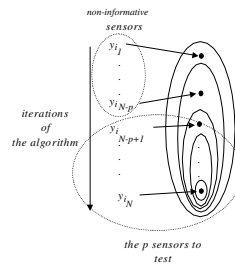


FIG. 5 – Relevant sensor selection using a top-down hierarchical approach

The following greedy algorithm² is proposed :

²With a large number of sensors ($N \rightarrow \infty$), this algorithm does not give a solution in real-time situation.

```

; RELEVANT SENSOR VECTOR SELECTION
; initialisation
0. "p" and/or "h_max" are fixed parameters
1. for i_j := 1 to N      (N=number of potential sensors)
2.   Compute h(H/(y\y_{i_j}))
3. end for
4. Find y_{i_j} which minimizes h(H/(y\y_{i_j}))
5. S ← y\y_{i_j}
; loop
6. Repeat
7.   Compute h(H/(S\y_{i_j})) with y_{i_j} ∈ S
8.   Find y_{i_j} which minimizes h(H/(S\y_{i_j}))
9.   S ← S\y_{i_j}
10. Until [card S = p] or [h(H/(S\y_{i_j})) > h_max]
; Algorithm end
    
```

with $y_{i_j} \in S$ and $S \in P(y)$.

Remark - The quantity $h(H) - h(H/y)$ is the entropy of H explained by y . The index $q(S/y) = \frac{h(H) - h(H/S)}{h(H) - h(H/y)}$ can be defined as a quality measure of the model $H = g(S)$ with respect to the quality of the global model $H = f(y)$. This quantity represents the part of information carried by S on H with respect to the entire information carried by the sensor vector y , and maximizing this quantity is equivalent to choosing h_{\max} , since $h(H)$ and $h(H/y)$ are constant.

To illustrate this approach, let us consider a network containing three sensors with Gaussian conditional probability densities. Under the hypothesis H_0 , these conditional probability density functions are of average 0 and variance 1. Under the hypothesis H_1 , these functions are respectively of averages $m_1 = 1$, $m_2 = 1.5$ and $m_3 = 2$ and of variance 1. The problem consists of selecting the sensor couple which minimizes the conditional entropy. Intuitively, let us notice that the couple (y_2, y_3) seems better than the others, because of the smaller overlap of their Gaussian probability density functions. Figure 6 represents the variations of $h(H)$, $h(H/y_1)$, $h(H/y_2)$, $h(H/y_3)$, $h(H/y_1, y_2)$, $h(H/y_2, y_3)$, $h(H/y_1, y_3)$ and $h(H/y_1, y_2, y_3)$ as functions of P_0 .

The developed approach leads to consider the couple (y_2, y_3) which minimizes $h(H/y_i, y_j)$ ($i, j = 1, 2, 3$) whatever the value taken by the probability P_0 , with a minimal model quality of 94%.

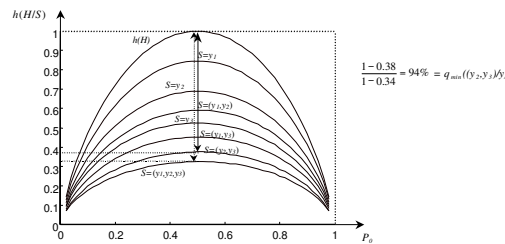


FIG. 6 – Variation of $h(H/S)$ as a function of P_0

Henceforth, we drop one set of indices : (y_1, \dots, y_p) represents the relevant sensor vector. The most relevant sensors being selected, let us define the elementary component of detection.

3 The elementary component of detection

Let us define the elementary component of detection (figure 7) from which all the detection complex networks can be structured [6].

Let $y = (y_1, y_2, \dots, y_p)$ be the measurement vector issued from a sensor vector $Y = (Y_1, Y_2, \dots, Y_p)$. Let (u_1, u_2, \dots, u_k) be a decision vector. From the knowledge of (u_1, u_2, \dots, u_k) and (y_1, y_2, \dots, y_p) , the detection component elaborates a decision $u \in \{0, 1\}$ according to whether hypothesis H_0 or H_1 is decided.

For instance :

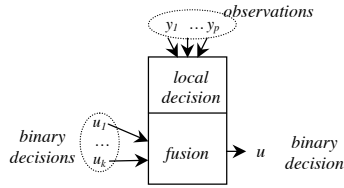


FIG. 7 – The elementary component of a detection network

- a centralized detection network is constituted by only one detection component with $u_0 = u$, $k = 0$ and $N = p$;
- a parallel distributed detection network is constituted by $N + 1$ components : N local detectors and one fusion center ;
- a serial distributed detection network is constituted by N components.

A priori probabilities P_j ($j = 0, 1$) and conditional PDF $p(y_1, \dots, y_p, u_1, \dots, u_k/H_j)$ ($j = 0, 1$) for each hypothesis are assumed known. Furthermore,

- $P_D = P(u = 1/H_1)$ is the detection probability ;
- $P_F = P(u = 1/H_0)$ is the false alarm probability associated with this component.

4 Shannon's conditional entropy of a network

Let us consider a complex distributed network constituted by some elementary components. By definition, the conditional entropy $h(H/u_0)$ can be written as follows :

$$\begin{aligned} h(H/u_0) &= E \left\{ \log \left(\frac{1}{P(H/u_0)} \right) \right\} = - \sum_{i,j \in \{0,1\}} P(u_0 = i, H_j) \cdot \log P(H_j/u_0 = i) \\ &= - \sum_{i \in \{0,1\}} \left[\alpha_i \log \frac{\alpha_i}{\alpha_i + \beta_i} + \beta_i \log \frac{\beta_i}{\alpha_i + \beta_i} \right] \end{aligned} \quad (6)$$

with :

$$\begin{aligned} \alpha_1 &= P_0 P_F & \beta_1 &= (1 - P_0) P_D \\ \alpha_0 &= P_0 (1 - P_F) & \text{and } \beta_0 &= (1 - P_0) (1 - P_D) \end{aligned} \quad (7)$$

When P_0 is constant, from this property : $f(v) = \alpha \log \frac{\alpha}{\alpha + \beta} + \beta \log \frac{\beta}{\alpha + \beta} \Rightarrow \frac{\partial f(v)}{\partial v} = \frac{\partial \alpha}{\partial v} \log \frac{\alpha}{\alpha + \beta} + \frac{\partial \beta}{\partial v} \log \frac{\beta}{\alpha + \beta}$ and from the well-known relation $\frac{\partial P_D}{\partial P_F} = \lambda$, the derivative of $h(H/u_0)$ with respect to any variable v can be expressed as follows :

$$\frac{\partial h(H/u_0)}{\partial v} = P_0 (C_{10} - C_{00}) \frac{\partial P_F}{\partial v} - (1 - P_0) (C_{01} - C_{11}) \lambda \frac{\partial P_F}{\partial v} \quad (8)$$

where :

$$\begin{aligned} C_{10} &= \log \frac{1}{P(H_0/u=1)} = \log \frac{P_0 P_F + (1 - P_0) P_D}{P_0 P_F} = \log \frac{\alpha_1 + \beta_1}{\alpha_1} \\ C_{00} &= \log \frac{P_0 (1 - P_F) + (1 - P_0) (1 - P_D)}{P_0 (1 - P_F)} = \log \frac{\alpha_0 + \beta_0}{\alpha_0} \\ C_{11} &= \log \frac{P_0 P_F + (1 - P_0) P_D}{(1 - P_0) P_D} = \log \frac{\alpha_1 + \beta_1}{\beta_1} \\ C_{01} &= \log \frac{P_0 (1 - P_F) + (1 - P_0) (1 - P_D)}{(1 - P_0) (1 - P_D)} = \log \frac{\alpha_0 + \beta_0}{\beta_0} \end{aligned} \quad (9)$$

Minimizing $h(H/u_0)$ consists thus of minimizing a mean risk for which the costs are not constant, but depend on the *a posteriori* probabilities. This minimization (equation (8) is equal to 0) consists in comparing the likelihood ratio $\Lambda(y_1, \dots, y_N) = \frac{p(y_1, \dots, y_N/H_1)}{p(y_1, \dots, y_N/H_0)}$ with the optimal threshold λ^* , and the following relation is defined :

$$\Lambda(y_1, \dots, y_N) \underset{u_0=0}{\overset{u_0=1}{\geq}} \frac{P_0}{(1-P_0)} \cdot \frac{(C_{10}^* - C_{00}^*)}{(C_{01}^* - C_{11}^*)} = \lambda^* \quad (10)$$

where C_{ij}^* are the optimal costs defined by $C_{ij}^* = \log \frac{1}{P(H_j/u^*=i)}$ ($i, j = 0, 1$), and u^* is the optimal decision, in conditional entropy minimization sense. This rule is similar to the equation obtained by applying Bayes' or Neyman-Pearson criterion. The costs are not constant. They are functions of *a posteriori* probabilities.

Assuming conditional independence of the observations, the optimization of this network is performed person-by-person : The optimization of elementary detection component k can be performed by taking the derivative of $h(H/u_0)$ with respect to P_{F_k} (where P_{F_k} is the false alarm probability of the k -th component). Minimizing $h(H/u_0)$ consists of setting :

$$\lambda_k = \frac{P_0}{1-P_0} \frac{(C_{10} - C_{00}) \frac{\partial P_F}{\partial P_{F_k}}}{(C_{01} - C_{11}) \frac{\partial P_D}{\partial P_{D_k}}} \quad (11)$$

This structure is identical to that obtained by using Bayes' or Neyman-Pearson criterion. The derivative calculation leads to consider all the monotonous fusion functions³ [2]. In order to avoid this combinatorial problem, we introduced in [7] a tree-based representation for the fusion rule.

For notation simplifications, let us consider that index "k" is implicit.

An exhaustive search could be proposed to find the optimal threshold λ^* of a detection component, and thus, P_D^* and P_F^* using relations (12) :

$$P_D = P(u = 1/H_1) = \int_{\lambda}^{+\infty} p(\Lambda(y_1, \dots, y_p, u_1, \dots, u_k)/H_1) \cdot d\Lambda(y_1, \dots, y_p, u_1, \dots, u_k) \quad (12)$$

$$P_F = P(u = 1/H_0) = \int_{\lambda}^{+\infty} p(\Lambda(y_1, \dots, y_p, u_1, \dots, u_k)/H_0) \cdot d\Lambda(y_1, \dots, y_p, u_1, \dots, u_k)$$

It consists in testing all the possible values of the threshold λ and the optimal value minimizes $h(H/u)$.

This exhaustive search could be used if the minimal and maximal bounds of λ are known from a preprocessing of the conditional PDF $p(y_1, \dots, y_p, u_1, \dots, u_k/H_j)$ ($j = 0, 1$). In most cases, a gradient-based technique is favorite.

5 Optimizing an elementary component with a gradient-based technique

Let us consider $v = \lambda$, where λ is any threshold. From equations (6), (7) and (9), we can write the following relation :

$$\frac{\partial h(H/u)}{\partial \lambda} = \frac{\partial P_F}{\partial \lambda} \cdot [P_0(C_{10} - C_{00}) + \lambda(1 - P_0)(C_{11} - C_{01})] \quad (13)$$

Let us recall that :

- from (12), we have $\frac{\partial P_F}{\partial \lambda} \leq 0 \quad \forall \lambda$;

- $P_F \leq P_D$. This inequality implies that $0 \leq \frac{P_F}{P_D} \leq 1 \leq \frac{1-P_F}{1-P_D}$

and from $\begin{cases} C_{11} = \log(1 + \frac{P_0 \cdot P_F}{(1-P_0) \cdot P_D}) \\ C_{01} = \log(1 + \frac{P_0 \cdot (1-P_F)}{(1-P_0) \cdot (1-P_D)}) \end{cases}$ we have $C_{11} - C_{01} \leq 0$

For a given P_0 , $\frac{\partial h(H/u)}{\partial \lambda}$ is an increasing function of λ , and it is equal to zero for λ^* . Hence, $h(H/u)$ is a decreasing function of λ when $\lambda < \lambda^*$, and an increasing function of λ otherwise.

³A fusion function f is monotonous if : $\forall u^* = (u_1^*, \dots, u_p^*)$ and $\forall u^{*'} = (u_1^{*'}, \dots, u_p^{*'})$,

if $\forall i \in \{1, \dots, p\} \quad u_i^* \geq u_i^{*'}$ then $f(u^*) \geq f(u^{*'})$

where :

- $u^* = (u_1^*, \dots, u_p^*)$ is a particular-decision vector.

- $u^{*' } = (u_1^{*' }, \dots, u_p^{*' })$ is another particular-decision vector.

Furthermore, knowing a given threshold λ , detection and false alarm probabilities can be computed using (12). From these and the *a priori* probability P_0 , we can compute a new threshold λ' . Let us note v^k the estimation of the variable v at the step k ($k \geq 0$) of an algorithm.

From (13) at the step k , we obtain :

$$\frac{\partial h(H/u)}{\partial \lambda^k} = \frac{\partial P_F}{\partial \lambda^k} \cdot [P_0(C_{10}^k - C_{00}^k) + \lambda^k(1 - P_0)(C_{11}^k - C_{01}^k)] \quad (14)$$

λ^k can be computed from the following equation :

$$\lambda^k = \frac{P_0}{1 - P_0} \cdot \frac{C_{10}^{k-1} - C_{00}^{k-1}}{C_{01}^{k-1} - C_{11}^{k-1}} \quad (15)$$

And equation (15) implies :

$$P_0 \cdot (C_{10}^{k-1} - C_{00}^{k-1}) = \lambda^k \cdot (1 - P_0) \cdot (C_{01}^{k-1} - C_{11}^{k-1}) \quad (16)$$

Hence

$$\frac{\partial h(H/u)}{\partial \lambda^{k-1}} = \frac{\partial P_F}{\partial \lambda^{k-1}} (1 - P_0)(C_{01}^{k-1} - C_{11}^{k-1})(\lambda^k - \lambda^{k-1}) \quad (17)$$

and

$$\lambda^k = \lambda^{k-1} + \frac{\partial h(H/u)}{\partial \lambda^{k-1}} \cdot \left(\frac{\partial P_F}{\partial \lambda^{k-1}} \cdot (1 - P_0)(C_{01}^{k-1} - C_{11}^{k-1}) \right)^{-1} \quad (18)$$

A recurrence equation $\lambda^k = \lambda^{k-1} - \tau^{k-1} \cdot \frac{\partial h(H/u)}{\partial \lambda^{k-1}}$ is obtained, and a gradient-based technique can be proposed. The stop condition of this algorithm is $|\lambda^k - \lambda^{k-1}| < \varepsilon$, where ε is a precision parameter defined by the user. The algorithm is the following one :

```

; Entropy-based optimization of a detection component
; using a gradient-based technique
; initialisation
0.  $P_0$  is known;  $\lambda^0$  and  $\varepsilon$  are given;  $k \leftarrow 0$ 
; loop
1. do
2.     Compute  $P_D$  and  $P_F$  (with (12))
3.     Compute  $C_{ij}$  ( $i, j = 0, 1$ ) (with (7) and (9))
4.      $k \leftarrow k + 1$ 
5.     Compute  $\lambda^k = \frac{P_0}{(1-P_0)} \cdot \frac{(C_{10}-C_{00})}{(C_{01}-C_{11})}$ 
6. until ( $|\lambda^k - \lambda^{k-1}| < \varepsilon$ )
7.  $\lambda^* = \lambda^k$ ;  $P_D^* = P_D$ ;  $P_F^* = P_F$ 
; End
    
```

From any λ^0 , this gradient-based technique makes the λ -series converge to the optimal solution.

As an example, let us consider the optimization of one detector associated with only one sensor for which the hypothesis probability densities are Gaussian. The standard deviation is equal to 1 for each hypothesis and the average is 0 (*resp.* 1) for H_0 (*resp.* H_1) hypothesis : $p(y/H_i) = N_i(i, 1)$ where y represents the sensor output, *i.e.* the input of the detection component. For Bayes' criterion, the costs are $C_{00} = C_{11} = 0$ and $C_{01} = C_{10} = 1$. Four parameters are used for the result comparison : P_0 , P_F , P_D and the prediction error probability which is defined by :

$$P_{error} = P_0 \cdot P_F + (1 - P_0) \cdot (1 - P_D) \quad (19)$$

Figure 8 represents the error probability as function of P_0 . Figure 9 represents the ROC curve.

By definition, the prediction error probability is minimized by Bayes' criterion (which uses constant costs). The entropy-based optimization does not allow for better results.

In [6], we demonstrated the following properties :

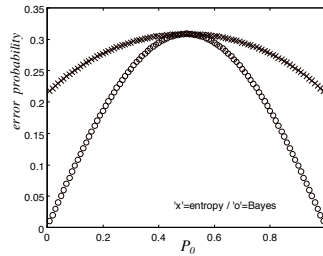


FIG. 8 – Error probability as a function of $P_0 - p(y/H_i) = N_i(i, 1)$

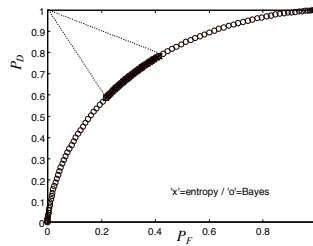


FIG. 9 – ROC curve as a function of $P_0 - p(y/H_i) = N_i(i, 1)$

- In the particular case when the hypotheses are equiprobable (*i.e.* $P_0 \simeq 0.5$), both criteria provide similar results : the same likelihood threshold, P_F , P_D , conditional entropy. The prediction error probability is maximal.
- When $P_0 \simeq 0$ or 1, the entropy-based criterion gives a poor solution in terms of prediction error probability. However, the obtained false alarm and detection probabilities are close to the ideal point ($P_F = 0, P_D = 1$). We demonstrated that the ROC curve is limited to a small arc using the entropy-based criterion, because the prediction error probability lies between two bounds α and β (figure 10).

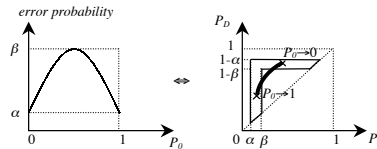


FIG. 10 – From error-probability curve to ROC curve

- Finally, in general, the closer α is to β (the prediction error probability tends to a constant), the more the ROC curve tends to an unique point.

Hence, the entropy-based criterion allows for the decrease of potential value set of P_F and P_D : it is then a robust criterion with respect to P_0 . In particular, when P_0 is close to 0 or 1, the ROC curve is close to the ideal point ($P_F = 0, P_D = 1$). It is not the case for the Bayesian criterion : when an event is rare (*e.g.* faults), Bayes' criterion neglects this event. The entropy-based criterion takes into account this event in a significant way, at the expense of prediction error probability.

6 Conclusion

The reduction of communication costs is the key focus of distributed detection networks. On the contrary, their performance is lower compared to centralized networks, because the fusion center does not receive all the available

information to make the final decision. The optimization of a parallel distributed network with N sensors leads to a set of $2^N + N$ nonlinear equations, that can only be solved in particular cases, assuming statistical independence of the local observations, and for small-sized networks. In order to reduce the number of equations to solve, Shannon's entropy is used to select relevant sensors for the decision process.

Once the relevant sensors are selected, we demonstrated that the optimization of a detection component with entropy-based criterion is equivalent to applying a variable threshold on the likelihood ratio, which depends on *a posteriori* probabilities. We proposed a gradient algorithm in order to find this threshold. We underlined an interesting robustness property with respect to rare events, and with respect to events for which *a priori* probabilities are uncertain (*e.g.* in a distributed detection network). In particular, the ROC curve does not recede from the ideal point ($P_F = 0$, $P_D = 1$) when P_0 is close to 0 or 1 : if a rare event (*e.g.* fault detection) is considered, Bayes' criterion neglects this event. For the entropy-based criterion, this event is taken into account in a significant way, at the expense of prediction error probability. The optimization of complex networks (centralized, parallel and serial distributed networks with N sensors, $N > 1$) was performed in [2]. Finally, these optimization techniques were applied to the distributed quantization problem : we considered a compromise between the broadcast bandwidth of the fusion center and quality of the decision system, in order to approximate centralized detection performance [7].

Références

- [1] R. C. Conant. Laws of information which govern systems. *IEEE Trans. on Systems, Man and Cybernetics*, 6(4) :240–255, April 1976.
- [2] C. Desrousseaux and D. Pomorski. Optimisation entropique des systèmes de détection distribuée. *Traitement du Signal*, 16(4) :275–293, September 1999.
- [3] T. L. Gabrielle. Information criteria for threshold determination. *IEEE Trans. on Information Theory*, 6 :484–486, October 1966.
- [4] I. Y. Hoballah and P. K. Varshney. Distributed bayesian signal detection. *IEEE Trans. on Information Theory*, 35(5) :995–1000, September 1989.
- [5] D. Middleton. *Statistical communication theory*. McGraw-Hill, New-York, 1960.
- [6] D. Pomorski. Entropy-based optimisation for binary detection networks. In *3rd International Conference on Information Fusion*, page vol.II, Paris, France, July 2000.
- [7] D. Pomorski and C. Desrousseaux. Improving performance of distributed detection networks : an entropy-based optimization. *Signal Processing*, 81 :2479–2491, 2001.
- [8] F. A. Sadjadi. Hypotheses testing in a distributed environment. *IEEE Trans. on Aerospace and Electronic Systems*, 22(2) :134–137, March 1986.
- [9] P. F. Swaszek. On the performance of serial networks in distributed detection. *IEEE Trans. on Aerospace and Electronic Systems*, 29(1) :254–260, January 1993.
- [10] Z. B. Tang. *Optimization of detection networks*. PhD thesis, University of Connecticut, 1990.
- [11] Z. B. Tang, K. R. Pattipati, and D. L. Kleinman. Optimization of detection networks : Part i - tandem structures.
- [12] R. R. Tenney and N. R. Sandell. Detection with distributed sensors. *IEEE Trans. on Aerospace and Electronic Systems*, 17(4) :98–101, July 1981.
- [13] P. K. Varshney. *Distributed detection and data fusion*. Springer Verlag, New-York, 1997.
- [14] D. Warren, P. Willet, and R. Rampertab. Shannon's information in decentralized signal detection. In *23rd Conference on Info. Sciences and Systems*, Baltimore, Maryland, March 1989.

DETECTION DE SPECTRES ANORMAUX EN ANALYSE MULTIVARIEE

TAHAR A. ^{*}(¹), BAROUR C. (¹), AÏT-KAKI S. (²), PALM R. (³)

⁽¹⁾ Laboratoire de statistique appliquée – Faculté des sciences – Annaba

⁽²⁾ Centre Universitaire d'El-Tarf

⁽³⁾ Unité de statistique et d'informatique, Faculté Universitaire de Gembloux (Belgique)

*Correspondance : e-mail : pr_tahar_ali@hotmail.com

1. Introduction

Le dosage chimique par spectrophotométrie de réflexion diffuse dans le proche infrarouge repose sur la relation qui existe entre le spectre obtenu pour un échantillon de produit donné et la teneur de ce produit en un élément particulier. L'expression mathématique de cette relation est l'équation de calibrage qui est calculée à partir des spectres d'une série d'échantillons du produit donné, pour lesquels les teneurs ont été déterminées par la voie analytique traditionnelle. Le calcul de l'équation de calibrage doit cependant se faire sur la base d'informations relatives à un ensemble homogène d'échantillons du produit auquel on s'intéresse. Il y a donc lieu d'écarter les échantillons ayant un spectre d'une allure fort différente du spectre moyen ainsi que les échantillons qui ont une teneur en protéines nettement différente de la teneur d'autres échantillons ayant un spectre à peu près identique. Les échantillons à éliminer représentent en fait des individus anormaux. L'objectif de cette note est de proposer aux praticiens une procédure concrète de recherche de ces individus anormaux. Nous présenterons d'abord la collecte des données de base utilisées (paragraphe 2), puis quelques informations concernant le calcul des équations de calibrage (paragraphe 3). Ensuite, nous passerons à l'analyse des distances de MAHALANOBIS, déterminées uniquement à partir des données spectrales (paragraphe 4), et à l'analyse des distances de COOK, déterminées après l'ajustement d'une équation de régression (paragraphe 5). Nous tirerons alors quelques conclusions (paragraphe 6).

2. Collecte des données

L'étude a été réalisée dans le cadre du calcul des équations de calibrage relatives aux teneurs en protéines du blé tendre au cours de trois années consécutives (de 1984 à 1986). Pour chacune des trois années, 150 échantillons de blé ont été prélevés aléatoirement dans les remorques lors de la moisson dans les environs de Gembloux (Belgique). Les teneurs en protéines, en pour cent de matière sèche, ont été déterminées par la méthode chimique de KJELDAHL. Les mêmes échantillons ont été analysés par un spectrophotomètre à infrarouge proche (MII-6250 de Pacifique Scientifique), après avoir été préparés de manière identique (broyage par cyclotec de 1 mm). Les densités optiques (ou régresseurs), DO, ont été obtenues par pas de 2 nm entre 1.100 et 2.500 nm de longueurs d'onde (λ_i). Soit pour chaque année une matrice de données explicatives de dimensions 150×700 et un vecteur de teneurs en protéines (variable à expliquer) de dimensions 150×1 .

3. Calcul des équations de calibrage

Les problèmes rencontrés lors du calcul des équations de calibrage résultent essentiellement, d'une part, de l'influence de nombreux facteurs sur les densités optiques relatives aux

différentes longueurs d'onde et, d'autre part, du nombre élevé de longueurs d'onde pour lesquelles les densités optiques sont disponibles, ce qui conduit à de fortes corrélations entre les réponses aux longueurs d'onde voisines. Les techniques de calcul des équations de calibrage sont nombreuses. Elles se différencient essentiellement par les transformations appliquées aux spectres de longueurs d'onde prises en considération et par les techniques d'ajustement des équations de calibrage. L'incidence des deux premiers facteurs sur la qualité des équations de calibrage a fait l'objet d'une étude antérieure (DAGNELIE *et al.*, 1991). Dans cette étude, nous avons considéré quatre transformations mathématiques des données spectrales : le lissage par le calcul des moyennes mobiles, la détermination des dérivées premières et des dérivées secondes et le calcul des coefficients de FOURIER. Dans le cas des trois premières transformations, les équations de calibrage ont été obtenues par régression multiple pas à pas (DAGNELIE, 1982 ; DRAPER et SMITH, 1982). Les réponses aux différentes longueurs d'onde étant les variables explicatives et les teneurs en protéines représentent la variable à expliquer. Toutefois, avant le recours à l'algorithme de régression pas à pas, nous avons procédé à une réduction préliminaire du nombre de variables explicatives potentielles. Deux procédures ont été envisagées. Dans le premier cas, les variables explicatives à soumettre à la régression multiple ont été choisies de manière systématique, en ne retenant que celles correspondant à des longueurs d'onde séparées par une distance fixée (TAHAR, 1989). Dans le second cas le choix des variables explicatives potentielles a été basé sur la corrélation entre réponses aux longueurs d'onde voisines (TAHAR, 1989). En ce qui concerne les coefficients de FOURIER, la procédure de réduction préliminaire du nombre de variables explicatives a simplement consisté à ne retenir que les variables constituées des p premiers couples de coefficients, et l'équation de calibrage a été obtenue par régression multiple sur ces p couples de variables, sans autre procédure de sélection de variables. Pour plus d'informations concernant les méthodes comparées, le lecteur consultera le texte publié antérieurement (DAGNELIE *et al.*, 199).

4. Analyse des distances de MAHALANOBIS

La recherche de spectres anormaux repose sur l'étude de la distance de MAHALANOBIS, D_i , mesurée entre chacun des spectres relatifs à une année donnée et le spectre moyen. Une définition rigoureuse et une justification de cette distance, couramment utilisée en analyse statistique multivariée, est donnée par BARNETT et LEWIS (1978), DAGNELIE (1982), HAWKINS (1980) et TAHAR (1989). En considérant que les spectres proviennent d'une distribution normale à p dimensions, on peut montrer que la distribution d'échantillonnage des distances D_i possède une distribution approximativement normale. Par conséquent, la recherche de spectres anormaux peut se faire par la recherche des distances anormales. Cette recherche de spectres anormaux (ou distances anormales) a été réalisée, d'une part, par l'intermédiaire du test T^2 de FERGUSSON (1961) et, d'autre part, par l'examen des diagrammes de dispersion linéaires des distances en fonction des scores normaux correspondants. En effet, l'examen de la partie supérieure du graphique permet de détecter la présence d'éventuelles valeurs anormales, s'écartant de façon importante de la première bissectrice (DAGNELIE, 1982 ; DRAPER et SMITH, 1981 ; WEISBERG, 1985). C'est ainsi que nous avons considéré un seul individu anormal pour les données de 1984 (l'individu 6) quatre individus anormaux pour les données de 1985 (les individus 2, 4, 16 et 74) et enfin, deux individus anormaux pour les données de 1986 (les individus 2 et 4). A titre d'exemple, les figures 1, 2 et 3, représentant les diagrammes de dispersion des distances centrées

réduites, calculées sur les 32 premiers coefficients de FOURIER, en fonction des scores
normaux pour

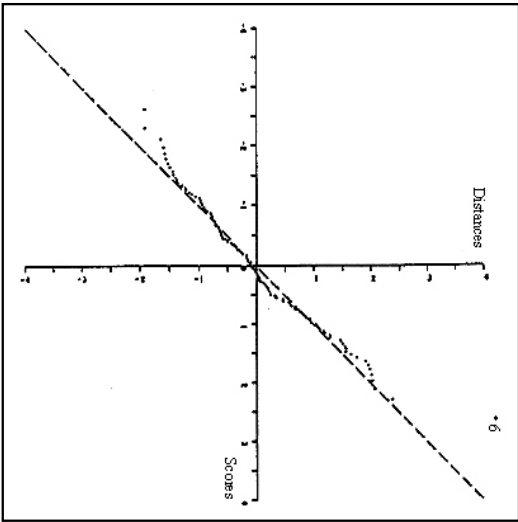


Figure 1 : Diagramme de dispersion des distances centrées réduites de MAHALANOBIS, calculées sur les 32 premiers coefficients de FOURIER, en fonction des scores normaux correspondants (année 1984).

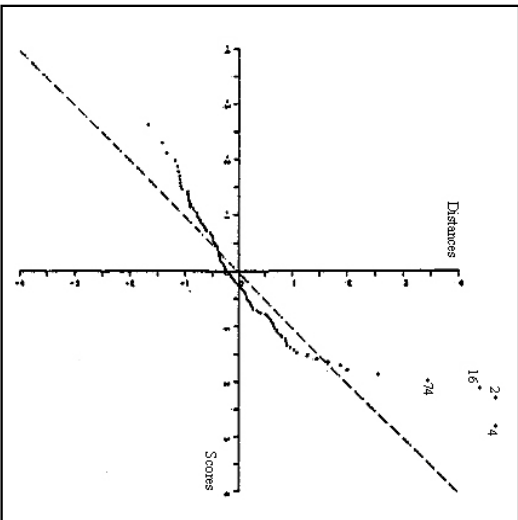


Figure 2 : Diagramme de dispersion des distances centrées réduites de MAHALANOBIS, calculées sur les 32 premiers coefficients de FOURIER, en fonction des scores normaux correspondants (année 1985).

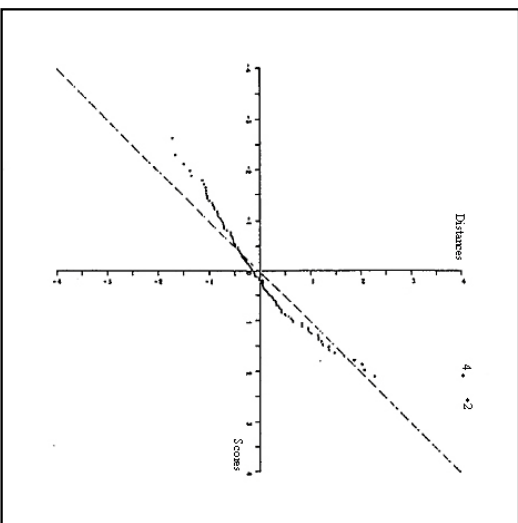


Figure 3 : Diagramme de dispersion des distances centrées réduites de MAHALANOBIS, calculées sur les 32 premiers coefficients de FOURIER, en fonction des scores normaux correspondants (année 1986).

chacune des trois années 1984, 1985 et 1986, montrent bien le caractère anormal des individus cités précédemment (TAHAR, 1989). Le T^2 de FERGUSSON (1961) appliqué aux différents individus donne des résultats qui conduisent, chaque fois, aux mêmes conclusions que celles tirées à partir de l'analyse des diagrammes de dispersion des distances en fonction des scores normaux correspondants. Le tableau 1 reprend les numéros des individus aberrants, les distances de MAHALANOBIS (D_i) et les valeurs correspondantes de T^2_{obs} du test T^2 de FERGUSSON (1961).

Tableau 1 : Numéros des individus et valeurs de D_i et de T^2_{obs} des spectres anormaux pour les données de chacune des trois années 1984, 1985 et 1986. (1)

| 1984 | | 1985 | | 1986 | |
|-------------------|-------------|-------------------|-------------|-------------------|-------------|
| Individu et D_i | T^2_{obs} | Individu et D_i | T^2_{obs} | Individu et D_i | T^2_{obs} |
| 6 (4,23) | 100,10* | 4 (4,70) | 146,16* | 2 (5,75) | 146,78* |
| - | - | 2 (4,69) | 145,97* | 4 (5,72) | 145,86* |
| - | - | 16 (4,39) | 135,80* | - | - |
| - | - | 74 (3,44) | 106,35* | - | - |

(1) La valeur figurant entre parenthèses représente D_i

* L'astérisque indique que les individus ont une valeur de T^2_{obs} supérieure à T^2_{α} (avec $T^2_{\alpha} = 59,13$ avec α risque de première espèce) (TAHAR, 1989).

5. Analyse des distances de COOK

La détection des observations influentes, en régression linéaire, a été considérée par un grand nombre de statisticiens. En particulier ANDREWS et PREGIBON (1978), CHATTERJEE et HADI (1986), COOK et WEISBERG (1980), HOAGLIN et WELSH (1978), HOCKING (1983), PALM (1988) et TAHAR (1989) ont considéré le problème des observations qui sont influentes par l'examen de leurs effets sur l'estimation des paramètres. Parmi les paramètres statistiques définis pour mesurer l'influence ou l'importance de chacune des observations sur l'estimation des coefficients de régression, la distance de COOK, C_i , est la plus couramment utilisée (COOK, 1977). Ce paramètre est calculé, pour chaque individu de l'échantillon, une fois que l'on dispose d'une relation (équation de calibrage) entre la variable à expliquer (teneurs en protéines) et les variables explicatives (densités optiques). Pour chaque observation le logiciel SAS (X, 1985) fournit directement la valeur prédite (\hat{y}_i), le résidu (e_i) et les valeurs de la distance de COOK (C_i), de l'effet de levier (h_i) de la matrice de projection \mathbf{H} et du résidu standardisé (r_i). La détermination des données influentes a été faite par l'analyse minutieuse des valeurs des trois paramètres C_i , h_i et r_i des cinq individus les plus influents, au sens de la distance de COOK, de chaque matrice de données. On a considéré comme influent tout individu qui présente, pour l'ensemble des équations de calibrage de la majorité des matrices de données étudiées, des valeurs de C_i particulièrement élevées. Cette méthode, qui n'est pas basée sur le résultat d'un test statistique, a permis de mettre en évidence le caractère influent des individus 71 et 11 pour les données de 1984, et des individus 15 et 141 pour les données de 1985. Le caractère influent de ces individus est dû, en grande partie, à l'information combinée apportée par les données spectrales et les teneurs en protéines. Enfin, pour les données de 1986 aucun individu influent n'a été déterminé par la distance de COOK. Par ailleurs, la règle empirique, proposée par COOK (1977) et WEISBERG (1985), ne conduit à aucune donnée influente pour les trois années lorsqu'on compare la distance de COOK, C_i , avec la valeur relative de la variable F de SNEDECOR à p et n-p degrés de liberté pour $\alpha = 0,5$, p représentant le nombre de variables explicatives

présentes dans l'équation de régression et n le nombre d'individus à partir desquels le modèle a été calculé. OBENCHAIN (1977) a, d'ailleurs, signalé que cette règle ne correspond pas à un test statistique rigoureux.

6- conclusion

L'élimination des données anormales au sens de la distance généralisée de MAHALANOBIS n'a été effectuée que pour les observations qui ont été mises en évidence dans la majorité des matrices de données, alors que dans le cas où la recherche de données anormales ne se fera que sur une seule matrice, tous les individus anormaux qui seront signalés par le test T^2 de FERGUSSON (1961) doivent être éliminés. Concernant la détection des données influentes à l'aide de la distance de COOK, la décision de considérer une donnée influente a été avant tout basée sur la convergence des résultats obtenus pour les différentes matrices de données relatives à un même groupe d'individus. Cependant, lorsqu'on ne dispose que d'une seule matrice de données et d'une seule équation de régression, nous suggérons le recours à l'utilisation de la règle empirique présentée par COOK (1977) et WEISBERG (1985) pour décider du rejet de données influentes.

Références bibliographiques

- ANDREWS D. et PREGIBON D. (1978). *Finding outliers that matter*. J. Roy. Stat. Soc., Ser. B, 40(2), 85-93.
- BARNETT V. et LEWIS T. (1978). *Outliers in statistical data*. John Wiley, New York, 365p.
- CHATTER JEE S. et HADI A. S. (1986). *Influential observations , High leverage points, and outliers in linear regression*. Stat. Sci., Vol 1(3), 379-416.
- COOK R.D. (1977). *Detection of influential observation in linear regression*. Technometrics, 19(1), 15-18.
- COOK R.D. et WEISBERG S. (1980). *Finding influential cases in regression*. Technometrics, 22(4), 495-508.
- DAGNELIE P. (1982). *Analyse statistique à plusieurs variables*. Presses Agronomiques, Gembloux, 362 p.
- DAGNELIE P., DARDENNE P., PALM R. et TAHAR A. (1991). *Comparaison de techniques de calcul des équations de calibrage en analyse des protéines du blé par spectrophotométrie de réflexion dans le proche infrarouge*. Bull. Rech. Agron. Gembloux, 26(3), 323-336.
- DRAPER N.R. et SMITH H. (1981). *Applied regression analysis*. Wiley, New York, 709 p.
- FERGUSSON T.S. (1961). *Rules for rejection of outliers*. Rev. Inst.Statist., 29, 29-43.
- HAWKINS D.M. (1980). *Identification of outliers*. Chapman et Hall, London, 188 p.
- HOAGLIN D.C. et WELSCH R. E. (1978). *The hat matrix in regression and ANOVA*. Amer. Statist., 32, 17-22.
- HOCKING R.R. (1983). *Developments in linear regression methodology : 1959-1982*.

Technometrics, 25(3), 219-249.

OBENCHAIN R.L. (1977). Letter to the editor. *Technometrics* , 19(3), 348-351.

PALM R. (1988). Diagnostics de validation des équations de régression linéaire. *Notes de Statistique et d'Information, Gembloux*, 88(1), 27 p.

TAHAR A. (1989). Comparaison de techniques de calcul des équations de calibrage en analyse par spectrophotométrie de réflexion dans le proche infrarouge. Thèse de Doct. *Sci. Agron., Fac.Sci. Agron., Gembloux, Belgique*, 290 p.

WEISBERG S. (1985). Applied linear regression. *Wiley, New York, U.S.A.*, 324 p.

X (1985). SAS User's guide : Statistics, version 5 edition. *Cary. SAS Institute INC*, 956 p.

**SELECTIVITE ET CRIBLAGE DES FACTEURS CHIMIQUES SUR LE
POTENTIEL DES ELECTRODES A MEMBRANE SOLIDE.**

**Salima ALI – MOKHNACHE, Nouredine DADDA, Ali LARKEM, Ali TAHAR,
Djelloul MSSADI*.**

Laboratoire de Sécurité Environnementale et Alimentaire,

Université BADJI Mokhtar – Annaba, 23000, Annaba. e – mail : messkher@yahoo.fr

INTRODUCTION

Lorsqu'une électrode est mise au contact d'une solution contenant l'ion i (d'activité a_i) auquel elle est spécifiquement sensible, il se développe une différence de potentiel donnée par l'équation connue de Nernst. Nicolsky et Eisenman [1] ont proposé une relation plus générale qui tient compte de l'influence d'un deuxième ion j , relation qui introduit un facteur de sélectivité noté K_{ij} , qui peut être déterminé par la méthode des solutions mixtes en mesurant le potentiel d'électrode :

- dans une solution pure de l'ion primaire i , soit E ;
- et dans une série de solutions de i et j d'activités variables a'_i et a'_j , soit E' .

On en déduit pour les ions monovalents:

$$K_{ij} a'_j = \{ \exp [(E - E')F / RT] \}^{\alpha} (a_i - a'_i) \quad (1)$$

où F est le faraday, R la constante universelle des gaz parfaits et T la température absolue. L'exposant α est égal à $(+1)$ pour les anions, et à (-1) pour les cations.

Les variations en fonction de a'_j du premier membre de (1) est une droite de pente K_{ij}

La prévision, à partir de quelques données préliminaires, de la sélectivité d'une électrode (caractérisée par K_{ij}) c'est – à – dire de son aptitude à distinguer entre l'ion primaire pour lequel elle a été conçue et les ions étrangers d'une même solution susceptibles d'interférer sur son fonctionnement, est importante.

Pour réaliser un tel objectif nous avons utilisé la méthode du simplex. Elle consiste à optimiser une fonction linéaire de plusieurs variables soumises à un ensemble de contraintes également linéaires, qui limitent l'intervalle de calcul, et permettent ainsi d'accélérer la convergence du processus itératif.

Nous nous intéresserons, dans une première partie, à l'influence des interférences sur le potentiel d'électrode par la détermination des coefficients de sélectivité.

La deuxième partie sera consacrée à l'étude de l'interaction de facteurs chimiques (concentration (C), pH, force ionique μ) sur la valeur du potentiel d'électrode à membrane solide E en suivant un plan factoriel complet (P F C) [2].

I – DETERMINATION DE K_{ij} PAR LA METHODE DE SOLUTIONS MIXTES ET DU SIMPLEX

Nous avons déterminé les interférences possibles en comparant les valeurs du potentiel de l'électrode plongée:

- dans une solution contenant l'ion à doser seul à des concentrations fixes (de 10^{-6} à 10^{-1} mol/l)
- dans une solution contenant l'ion à doser et l'ion interférant, ce dernier pris toujours à la même concentration de 10^{-3} mol/l.

Les coefficients de sélectivité en ont été déduits en utilisant l'ordinateur.

Nous reproduisons ci-après quelques valeurs obtenues pour l'électrode aux fluorures F^- et pour l'électrode spécifique aux ions Pb^{2+} .

Electrode aux fluorures $K_{F/X^{n-}}$ X^{n-} : Cl^- ($2,15 \times 10^{-3}$) ; SO_4^{2-} ($4,60 \times 10^{-2}$) ; $C_2O_4^{2-}$ (2420,00) ; $S_2O_8^{2-}$ (9,60).

Electrode spécifique aux ions Pb^{+2} $K_{Pb^{2+}/X^{n+}}$ X^{n+} : Cu^{2+} (1,00) ; Na^+ ($6,30 \times 10^{-3}$) ; Fe^{3+} (7,9) ; Hg^{2+} (100,00).

Les valeurs des coefficients de sélectivité montrent en général une assez grande sélectivité de l'électrode aux ions fluorures; les anions chlorures (Cl^-) et sulfates (SO_4^{2-}) ont une faible influence sur l'électrode spécifique aux ions fluorures, alors qu'il est impossible de doser les ions fluorures en présence des anions oxalates $C_2O_4^{2-}$ et thiosulfates $S_2O_8^{2-}$.

Pour l'électrode spécifique aux ions Pb^{2+} , nous avons constaté une grande sélectivité pour les ions Zn^{2+} et Co^{2+} ; les ions Hg^{2+} peuvent intervenir dans la détection des ions Pb^{2+} .

Le calcul de K_{ij} par la méthode du simplex conduit aux mêmes valeurs déterminées précédemment à 0,005 près dans le cas le plus défavorable. Le nombre d'itérations dépend de la nature des paires d'ions considérés et de la valeur adoptée pour initialiser les calculs.

II CRIBLAGE DES FACTEURS CHIMIQUES SUR LE POTENTIEL DES ELECTRODES A MEMBRANE SOLIDE [2]

Nous avons étudié l'influence de la force ionique μ ($= X_1$), de la concentration C ($= X_2$) et du pH ($= X_3$) sur le potentiel d'électrode E , en tenant compte des interactions possibles entre ces facteurs chimiques. Le modèle de réponse expérimental a été choisi de la forme :

$$Y = a_0 + a_1 X_1 + a_2 X_2 + a_3 X_3 + a_{12} X_1 X_2 + a_{13} X_1 X_3 + a_{23} X_2 X_3 + a_{123} X_1 X_2 X_3 + e \quad (2)$$

Où Y est la réponse; a_i : les coefficients principaux ; X_1, X_2, X_3 les facteurs qui peuvent influencer sur la réponse ; e : une variable englobant tous les facteurs connus non contrôlés et les facteurs inconnus non contrôlés (c'est aussi l'erreur expérimentale).

Le mode opératoire consiste à mesurer le potentiel engendré par une électrode à membrane solide couplée à une électrode de référence, d'une solution dont la force ionique est maintenue constante par un ajout d'un sel de fond, et dont le pH est ajusté à la valeur désirée par addition d'une solution tampon.

Pour faciliter la construction du plan factoriel complet (qui nécessite $2^3 = 8$ essais) nous avons respecté la convention proposée par Yates [2] pour la notation des plans factoriels et pour l'ordre des essais :

1°/ le niveau haut est noté (+1) et le niveau bas (-1) ;

2°/ l'ordre des essais est obtenu en alternant les signes positives et négatives

de un à un pour le premier facteur,

de deux à deux pour le deuxième,

et, enfin, de quatre à quatre pour le troisième facteur.

On a répété, au moins, deux fois chaque essai afin de calculer l'écart-type de la moyenne de la répétition.

Cet écart-type, assimilé à l'écart-type expérimental, servira à interpréter statistiquement les valeurs des coefficients du modèle ; il est noté S_e .

A partir des réponses Y_i collectées, le calcul des coefficients a_i et a_0 se fait [2] selon les relations :

$$a_j = \frac{\sum_{j=1}^N Y_{i+} - \sum_{j=1}^N Y_{i-}}{N} \quad \text{et} \quad a_0 = \frac{\sum_{j=1}^N Y_j}{N} \quad (3)$$

Dans ces formules, la quantité de $\sum_{j=1}^N Y_{i+}$ représente la somme des réponses Y_i pour lesquelles le facteur X_i est au niveau +1 ; $\sum_{j=1}^N Y_{i-}$ est la somme Y_i pour lesquelles le facteur X_i est au niveau -1.

Nous avons obtenu les valeurs suivantes :

$$a_0 = 47,125 \quad ; \quad a_1 = 27,125 \quad ; \quad a_2 = 40,625 \quad ; \quad a_3 = 38,375$$

La matrice carrée d'ordre 8 construite à partir du plan d'expérience permet, en outre, de calculer les interactions en appliquant la première relation (3) dans laquelle i désigne l'une quelconque des 8 colonnes du tableau.

Tableau I – Matrice du modèle pour Plan Factoriel Complet à 2³.

| Effets principaux | | | | Effets d'interaction | | | |
|-------------------|----------------|----------------|----------------|-------------------------------|-------------------------------|-------------------------------|--|
| X ₀ | X ₁ | X ₂ | X ₃ | X ₁ X ₂ | X ₁ X ₃ | X ₂ X ₃ | X ₁ X ₂ X ₃ |
| a ₀ | a ₁ | a ₂ | a ₃ | a ₁₂ | a ₁₃ | a ₂₃ | a ₁₂₃ |
| + 1 | - 1 | - 1 | - 1 | + 1 | + 1 | + 1 | - 1 |
| + 1 | + 1 | - 1 | - 1 | - 1 | - 1 | + 1 | + 1 |
| + 1 | - 1 | + 1 | - 1 | - 1 | + 1 | - 1 | + 1 |
| + 1 | + 1 | + 1 | - 1 | + 1 | - 1 | - 1 | - 1 |
| + 1 | - 1 | - 1 | + 1 | + 1 | - 1 | - 1 | + 1 |
| + 1 | + 1 | - 1 | + 1 | - 1 | + 1 | - 1 | - 1 |
| + 1 | - 1 | + 1 | + 1 | - 1 | - 1 | + 1 | - 1 |
| + 1 | + 1 | + 1 | + 1 | + 1 | + 1 | + 1 | + 1 |

Finalement on trouve :

$$a_{12} = - 23,375 \quad ; \quad a_{13} = - 21,310 \quad ; \quad a_{23} = 0,375 \quad ; \quad a_{123} = 2,875$$

En prenant pour limite de signification des coefficients principaux et d'interaction : $|a_i| > 2 \text{ Se}$, on élimine du modèle tous les coefficients inférieurs en valeur absolue à 21,4, qui sont considérés comme statistiquement égaux à zéro.

On arrive en définitive au modèle théorique simplifié.

$$\hat{Y} = 47,12 + 27,12 X_1 + 40,62 X_2 + 38,38 X_3 - 23,38 X_1 X_2 \quad (4)$$

En ce qui concerne la force ionique et la concentration, leurs coefficients principaux sont positifs. Apparemment on pourrait les fixer tous les deux à leurs niveaux haut en vue d'avoir une réponse élevée. Mais leur interaction est négative : c'est quand ils sont de signes opposés que cette interaction est la plus intéressante, ce qui arrive dans deux situations diamétralement très différentes.

Une autre méthode d'interprétation consiste à supposer que, si les coefficients des facteurs et des interactions n'ont pas d'influence, il se répartissent selon une loi Normale. Tout coefficient aberrant est alors un coefficient significatif.

Le diagramme des scores normaux pour la force ionique présente un coefficient de corrélation $R = 0,884$ inférieur au test de corrélation pour la normalité, pour $N = 8$ et un niveau de signification $\alpha = 0,05$ soit, $R_C = 0,9029$.

CONCLUSION

Les électrodes spécifiques permettent de disposer d'une méthode analytique simple, surtout sensible et relativement sélective, cependant elles doivent être testées selon des critères bien définis.

Nous avons utilisé des électrodes à membrane solide disponibles sur le marché et nous nous sommes intéressés particulièrement à leur sélectivité.

Nous avons déterminé les coefficients de sélectivité par la méthode des solutions mixtes, puis nous avons confirmé ces valeurs par la méthode du Simplex.

Dans une deuxième partie nous avons essayé d'étudier l'interaction des facteurs : pH, force ionique et concentration sur la valeur du potentiel d'électrode à membrane solide en suivant un plan factoriel complet.

Le diagramme des n-scores confirme l'influence de la force ionique.

REFERENCES BIBLIOGRAPHIQUES

- 1 – S. ALI-MOKHNACHE, Thèse de Doctorat ès-Sciences, Université de Annaba, 16 mars 1993.
- 2 – M. FEINBERG ; La Validation des Méthodes d'Analyses, Edition Masson, Paris, 1996.

Discrimination de populations d'abeilles mellifères en analyse multivariée

Choukri BAROUR ^{1*}, Ali TAHAR ¹, Sarah Elizabeth RADLOFF ²

¹ Laboratoire de statistique appliquée, Faculté des Sciences,
Université de Annaba (Algérie)

² Department of Statistics, Rhodes University, Grahamstown,
6140 South Africa

1. Introduction

Les Abeilles étaient parmi les premiers insectes qui ont été étudiés morphométriquement. Ces études sont arrivées de la Russie au début des années 1900 (Alpatov, 1929). D'un point de vue statistique le plus grand inconvénient de ces études morphométriques était le nombre restreint d'observations utilisées. Une percée est venue en 1916 où Chochlov a effectué la première étude morphométrique en utilisant un nombre adéquat d'Abeilles pour l'analyse statistique. Entre 1924 et 1929 Michailov (1926) et Alpatov (1929) ont étudié l'influence de différents facteurs environnementaux sur la variation des Abeilles basée sur une grande collection d'Abeilles de plusieurs endroits de l'Union Soviétique.

Jusqu'à 1964, l'analyse morphométrique d'Abeille s'est fondée seulement sur des statistiques et des méthodes univariées. Dupraw (1965) était le premier qui a utilisé les méthodes multivariées dans l'analyse biométrique d'Abeilles. Des méthodes multivariées ont été également utilisées pour distinguer les différents écotypes dans une race (Louis & Lefebvre, 1968 ; Tomassone & Fresnaye, 1971). Ruttner *et al.* (1978) ont présenté plus de caractères que ceux qui ont été utilisés par Alpatov, Goetze et Dupraw, et ont établi un ensemble standard de 40 variables biométriques utilisées dans la classification des Abeilles, appliquant un procédé d'analyse discriminante. Les caractères qui sont exigés dans cette analyse varient selon la région d'étude.

L'introduction de l'analyse multivariée à la morphométrie d'Abeilles était une percée importante qui a permis l'utilisation d'autres techniques multivariées: l'analyse en composantes principales (ACP) et l'analyse factorielle, sont utilisées pour détecter les groupes morphologiques des colonies au sein des populations (Ruttner *et al.*, 1978 ; Ruttner, 1988). L'analyse discriminante est utilisée pour confirmer la séparation des groupes, pour déterminer les variables les plus discriminantes et pour calculer les pourcentages des colonies correctement classifiées

(Ruttner, 1988; Daly, 1992). Des phénogrammes ou dendrogrammes sont utilisés pour caractériser les distances (distance généralisée de MAHALANOBIS et d'autres) entre les groupes morphologiques (Tomassone et Fresnaye, 1971; Cornuet *et al.*, 1975; Daly, 1992). L'analyse de la variance multivariée (MANOVA) est utilisée pour mettre en évidence des différences multivariées significatives entre les groupes morphologiques (Rinderer *et al.*, 1990).

Dans ce contexte, nous étudierons les populations d'Abeilles domestiques grâce aux techniques biométriques et aux analyses statistiques multivariées qui permettent d'établir les caractéristiques morphologiques et de tenter de discriminer les populations de la race locale en essayant de répondre aux questions suivantes : Peut-on discerner une différenciation en populations locales ? Trouve-t-on des indices d'influence génétique de races étrangères ayant été importées en Algérie ?

2. Matériel et Méthodes

2.1. Echantillonnage

Les stations échantillonnées ont été sélectionnées en fonction des étages et sous-étages bioclimatiques. Chaque étage bioclimatique est caractérisé par un type de végétation (flore) et par des conditions climatiques différentes. Les stations ainsi que les différents sites d'échantillonnage qui ont fait l'objet de cette étude sont présentés dans le tableau (1). Les prélèvements des échantillons d'Abeilles ouvrières, ont été effectués en différentes saisons (Radloff, 1996 ; Radloff & Hepburn, 1999 ; Radloff *et al.*, 1996, 1998) durant l'année 2001 et 2002. Au niveau de chaque rucher, on a pris en considération d'une façon aléatoire 2 à 5 ruches (colonies).

Pour notre travail, nous avons pris en considération 7 stations avec 12 (ruchers) échantillonnés soit un total de 39 colonies ; avec 20 Abeilles individus ou ouvrières par ruche à l'exception de celle de Selawa (Guelma) où seulement 12 ouvrières par colonie ont été prélevées. Ruttner (1988) suggère que 10 à 20 Abeilles serait un échantillon standard suffisant pour une analyse biométrique.

*Correspondance :
E-mail : apis_mellifera_ch@hotmail.com

Tableau 1: Coordonnées géographiques, nombre d'Abeilles ouvrières et nombre de ruches (colonies) par site d'échantillonnage (Barour, 2003).

| Stations | Sites | Coordonnées | Nbr. de ruches | Effectif |
|-------------------|--------------|------------------|----------------|------------|
| Annaba | Berrahel | 36°50' N 7°26' E | 5 | 100 |
| | Serraïdi | 36°55' N 7°39' E | 4 | 80 |
| El Taref | Aïn Essel | 36°47' N 8°22' E | 3 | 60 |
| | Drean | 36°41' N 7°44' E | 3 | 60 |
| Souk Ahras | Tawra | 36°09' N 8°01' E | 3 | 60 |
| | Souk Ahras | 36°16' N 7°57' E | 3 | 60 |
| Guelma | Selawa | 36°23' N 7°15' E | 3 | 36 |
| | Mewna | 36°28' N 7°25' E | 3 | 60 |
| Jijel | Bni Metrane | 36°48' N 5°47' E | 3 | 60 |
| Skikda | El Hadayeke | 36°50' N 6°53' E | 2 | 40 |
| Ghardaïa | Bni Isguen 1 | 32°27' N 3°41' E | 5 | 100 |
| | Bni Isguen 2 | 32°25' N 3°39' E | 2 | 40 |
| Total | | | 39 | 756 |

2.2. Biométrie

2.2.1. Caractères morphologiques et méthodes de mesures

Les mesures, effectuées sur les ouvrières, portent sur les 25 caractères habituels, décrit en détail par Fresnay (1981) et Ruttner (1988) : longueur et largeur de l'aile antérieure et postérieure, longueur de la nervure alaire A et B, nombre de crochets alaires, pilosité, coloration, tomentum, longueur du 2^{ème} et 3^{ème} sternite, distance entre les miroirs à cire de la 3^{ème} sternite, longueur et largeur du miroir à cire du 3^{ème} sternite, longueur du fémur, du tibia, du métatarse et la largeur du métatarse et longueur du flagelle antennaire, indice cubital, longueur/largeur de L'aile antérieure, du miroir à cire et du métatarse et longueur totale de la patte postérieure. Tous ces caractères biométriques ont fait l'objet de mesures précises réalisées sous une loupe binoculaire (Zeiss) et un microscope (Olympus) équipés d'un micromètre oculaire étalonné avec tous les jeux d'objectifs de ces derniers. Approximativement 20.280 mesures morphométriques ont été prises sur un échantillon de 756 Abeilles pour cette étude.

2.3. Méthodes statistiques

L'objectif de cette étude est de savoir comment se structurent les colonies d'Abeilles étudiées et de déterminer les populations qui se rapprochent les unes des autres du point de vue morphologique afin de

discerner d'éventuels groupes de populations d'Abeilles ou « écotypes ». Pour cela, nous avons appliqué tout d'abord, l'analyse de la variance multivariée (MANOVA) (Palm, 2000) pour comparer toutes les stations entre elles, puis nous avons étudié également la structure des colonies d'Abeilles à l'aide de l'analyse en composantes principales (ACP) (Dagnelie, 1982 ; Palm, 1998) suivi d'une analyse hiérarchique (Bouroche & Saporta, 1980), puis nous avons effectué un reclassement des 756 Abeilles dans les 12 stations prélevées en appliquant l'analyse discriminante linéaire (Palm, 1999). Enfin, nous avons terminé par le test de LEVENE (Dagnelie, 1982) pour vérifier l'homogénéité de la variance inter-coloniale et intra-coloniale entre les groupes morphologiques. Tous les calculs ont été effectués à l'aide du logiciel MINITAB d'analyse et de traitement statistique des données version 13.31 pour WINDOWS (X, 2000).

3. Résultats

3.1. MANOVA

En ce qui concerne la comparaison inter-sites pour l'ensemble des vecteurs de moyennes des 25 variables biométriques entre les 12 ruchers échantillonnés, les trois tests (*Wilk's*, *Lawley-Hotelling* et *Pillai's*) de la MANOVA conduisent clairement et indiscutablement au rejet de l'hypothèse nulle (H_0) d'égalité des vecteurs de moyennes, pour l'ensemble des 12 sites (ruchers) échantillonnés et des différences très

hautement significatives ont été signalées (Tab. 2) (Barour, 2003).

Tableau 2 : Tests multivariés utilisés pour tester l'égalité des vecteurs de moyennes entre les 12 ruchers.

| Critère | Test statistique | F | DL | P |
|------------------|------------------|--------|-------------|-----------|
| Wilk's | 0,007 | 16,781 | (275; 7303) | 0,000 *** |
| Lawley-Hotelling | 8,25725 | 21,564 | (275; 7900) | 0,000 *** |
| Pillai's | 3,4487 | 13,336 | (275; 8030) | 0,000 *** |

* ($p \leq 0,05$) ; ** ($p \leq 0,01$) ; *** ($p \leq 0,001$) ; ns ($p > 0,05$) ; F de Fisher ; DL : degré de liberté ; P : probabilité.

3.2. Analyse en composantes principales (ACP)

Le tableau ci-dessous donne les valeurs propres, les proportions et les proportions cumulées de la variance expliquée par les composantes. La première composante principale (CP_1) prend en compte (34,90%) de la variabilité. Elle est, de loin, la plus importante, puisque les six composantes suivantes n'expliquent, respectivement, que (13,2), (8,9), (8,6), (6,1), (5,0) et (4,2 %). Nous limiterons notre analyse aux sept premières composantes qui, ensemble, expliquent 80,90% de la variabilité.

L'examen des corrélations des variables avec la première composante (CP_1), montre que les variables biométriques suivantes : longueur et largeur de l'aile antérieure ; longueur et largeur de l'aile postérieure, longueur du flagelle antennaire ; longueur du fémur ; longueur du tibia ; longueur du 3^{ème} sternite ; longueur et largeur du miroir à cire qui sont généralement des caractères de taille, sont positivement corrélées avec la première composante principale.

La deuxième composante principale (CP_2) est corrélée positivement avec la longueur de la nervure B et la pilosité. La troisième (CP_3) quant à elle, présente des corrélations positives avec chacune des quatre variables qui sont : la longueur du 2^{ème} sternite, la distance entre les miroirs à cire, le tomentum et l'indice long./larg. du métatarse. Pour la quatrième composante principale (CP_4), nous avons remarqué

également une corrélation positive avec : La longueur de la nervure A, la longueur et la largeur du métatarse, la coloration et la longueur de la patte postérieure. Pour ce qui est de la cinquième composante principale (CP_5), il existe des corrélations positives avec l'indice cubital et l'indice long./larg. de l'aile antérieure.

En ce qui concerne les deux dernières composantes (CP_6) et (CP_7), nous avons observé une seule corrélation positive pour chacune des composantes ; l'indice long./larg. du miroir à cire pour la (CP_6) et le nombre des crochets alaires pour la (CP_7). La figure (1), présente la projection des points moyens des colonies sur le premier plan d'une ACP effectuée sur 756 abeilles regroupées en 39 colonies. Les colonies sont repérées par le numéro de leur site de prélèvement. Il est aisé de constater une nette différenciation en 3 groupes :

Un premier groupe (site 11 et 12) comprend les colonies d'abeilles originaires de la région sud qui représentent les deux sites de Ghardaïa (Bni Isguen 1 et Bni Isguen 2). Un second groupe (sites 3 à 10) rassemble les colonies d'abeilles du rucher de Mewna (Guelma), les deux ruchers d'Aïn Essel et Drean (El Taref), les deux ruchers de Tawra et Souk-Ahras (Souk-Ahras), le rucher de Serraidi (Annaba) et le rucher de El Hadayeke (Skikda). Les colonies des ruchers de Berrahel (Annaba) et du Selawa (Guelma), s'en individualisent suffisamment pour qu'on soit tenté d'en faire un troisième groupe (sites 1 et 2).

Tableau 3 : Informations relatives aux valeurs propres des sept premières composantes principales.

| | CP_1 | CP_2 | CP_3 | CP_4 | CP_5 | CP_6 | CP_7 |
|---------------|---------|---------|---------|---------|---------|---------|---------|
| Valeur propre | 8,7155 | 3,2889 | 2,237 | 2,1443 | 1,5317 | 1,2482 | 1,0388 |
| Proportion | 34,90 % | 13,20 % | 8,90 % | 8,60 % | 6,10 % | 5,00 % | 4,20 % |
| Cumulatif | 34,90 % | 48,10 % | 57,00 % | 65,60 % | 71,70 % | 76,70 % | 80,90 % |

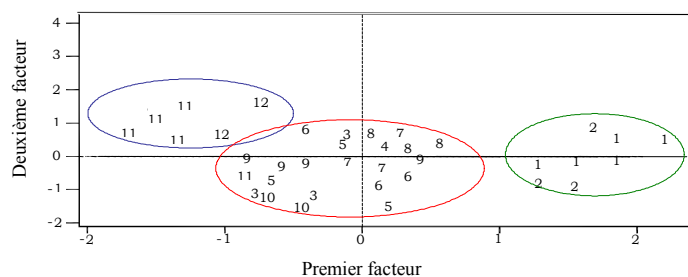


Figure 1 : Projection des points moyens des colonies repérées par leur numéro de site sur le premier plan factoriel d'une analyse en composantes principales.

3.3. Classification hiérarchique ou observations en groupes

Cette analyse révèle l'existence de six groupes assez distincts, qui se succèdent de façon plus évidente. Dans le dendrogramme (fig. 3) nous avons obtenu une partition en 6 classes basée sur les distances EUCLIDIENNES entre les différentes populations d'abeilles des 12 ruchers échantillonnés. Classe 1 : Ruchers de Berrahel (Annaba) et de Selawa (Guelma) ; classe 2 : ruchers d'Aïn Essel (El Taref) et El-Hadayeke (Skikda) ; classe 3 : ruchers de Bni Isguen 1 (Ghardaïa) et Bni Isguen 2 (Ghardaïa) ; classe 4 : ruchers de Souk Ahras (Souk Ahras) et de Serraidi (Annaba) ; classe 5 : ruchers de Bni Metrane (Jijel) et Tawra (Souk Ahras) ; classe 6 : ruchers de Mewna (Guelma) et Drean (El-Taref)

3.3. Discrimination des écotypes

La figure (2) reprend une partie des résultats obtenus par l'analyse discriminante linéaire et qui indique clairement une séparation nette entre les trois groupes morphologiques d'abeilles. Dans cette figure, le premier groupe morphologique représente les colonies d'abeilles de Bni Isguen 1 et 2 provenant de la région sud de Ghardaïa, le deuxième groupe morphologique représente également les colonies d'abeilles de l'ensemble des sites suivants : Serraidi (Annaba), Mewna (Guelma), Souk Ahras et Tawra (Souk Ahras), Aïn Essel et Drean (El-Taref), El-Hadayeke (Skikda) et Bni Metrane (Jijel) ; et le troisième groupe morphologique comprend les colonies d'abeilles du Berrahel (Annaba) et Selawa (Guelma).

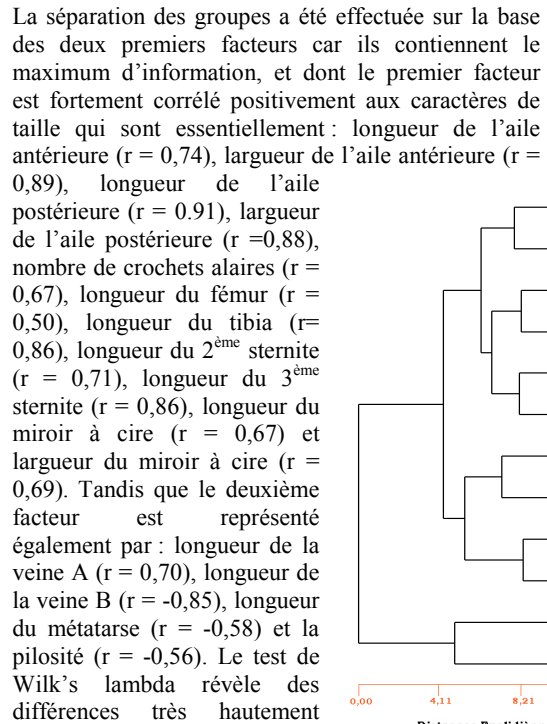
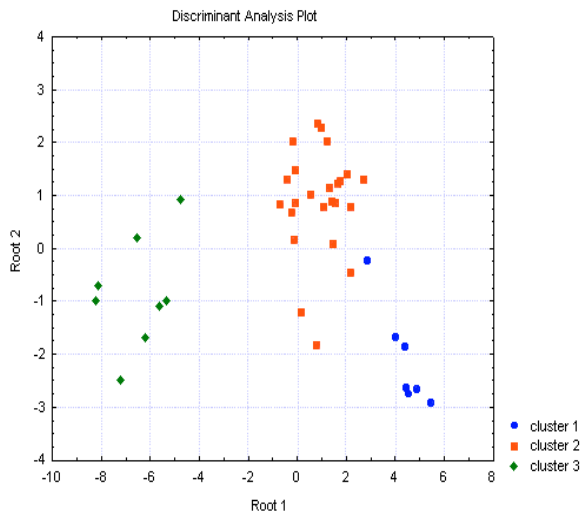


Figure 2 : Discrimination linéaire des trois groupes morphologiques en utilisant les moyennes des 25 caractères morphologiques par colonie.

significatives entre les vecteurs de moyennes des trois groupes morphologiques discriminés ($A = 0,0275$; $F(24, 50) = 10,654$; $P = 0.0000$).

Tableau 4 : Les distances généralisées de MAHALANOBIS (D^2) entre les trois groupes morphologiques obtenus par l'analyse discriminante.

Figure 3 : Classification hiérarchique des différentes populations d'Abeilles étudiées.

| | Groupe morphologique 1 | Groupe morphologique 2 | Groupe morphologique 3 |
|------------------------|------------------------|------------------------|------------------------|
| Groupe morphologique 1 | 0,00 | 20,88 | 119,58 |
| Groupe morphologique 2 | 20,88 | 0,00 | 58,00 |
| Groupe morphologique 3 | 119,58 | 58,00 | 0,00 |

Tableau 5: Matrice de classification des 39 colonies d'abeilles au sein des trois groupes morphologiques obtenus.

| | Pourcentage % | Gr. morphologique 1 | Gr. morphologique 2 | Gr. morphologique 3 |
|---------------------|---------------|---------------------|---------------------|---------------------|
| Gr. morphologique 1 | 85,71 | 6 | 1 | 0 |
| Gr. morphologique 2 | 100 | 0 | 24 | 0 |
| Gr. morphologique 3 | 100 | 0 | 0 | 8 |
| Total | 97,44 | 6 | 25 | 8 |

Dans le tableau (4), on observe les distances généralisées de MAHALANOBIS (D^2), élevées aux carrés, entre les trois groupes morphologiques d'abeilles. On voit immédiatement que la distance la plus faible concerne les deux groupes morphologiques (1) et (2); et que la distance la plus élevée est celle qui sépare le groupe morphologique (3) du groupe morphologique (1). On remarque aussi que la distance au carrés (D^2) qui sépare le groupe morphologique (1) du groupe morphologique (2) est de 20,88 et celle qui sépare le groupe morphologique (1) du groupe morphologique (3), est de 119,68. Enfin, la distance qui sépare le groupe morphologique (2) et le groupe morphologique (3) est de 58,00. Nous avons utilisé la fonction de l'analyse discriminante linéaire pour reclasser les 39 colonies d'abeilles étudiées dans les différents groupes morphologiques.

L'examen du tableau (4) de la matrice de classification montre que dans le groupe morphologique (1), juste une seule colonie d'abeilles du site de Bni Isguen 1 (Ghardaïa) est mal classée dans un autre groupe, qui est le groupe

(3), les 8 colonies d'abeilles qui lui appartiennent sont bien classées représentant un pourcentage de 100%. Le pourcentage total de classification pour l'ensemble des trois groupes morphologiques est de 97,44%.

3.5. Homogénéité de la variance inter-coloniale et intra-coloniale

3.5.1. Variance inter-coloniale.

L'examen du tableau (6) permet de dire que le test de LEVENE est significatif pour le premier facteur relatif aux caractères de taille et qui a été mis en évidence par l'analyse discriminante linéaire ($F = 1,52$; *d.l.* (11, 27); $P = 0,018$). Il en est de même du deuxième facteur (longueur de la veine A, longueur de la veine B, longueur du métatarse et la pilosité) entre les colonies d'abeilles étudiées ($F = 3,00$; *d.l.* (11, 27); $P = 0,009$).

3.5.2. Variance intra-coloniale

| | Variance inter-coloniale | | Variance intra-coloniale | |
|-----------|--------------------------|----------|--------------------------|----------|
| | F | P | F | P |
| Facteur 1 | 1,52 | 0,018 * | 1,89 | 0,03 * |
| Facteur 2 | 3,00 | 0,009 ** | 20,78 | 0,00 *** |

morphologique (2). Les 6 colonies d'abeilles qui restent sont bien classées dans le groupe morphologique d'origine, présentant ainsi un pourcentage de classification de 85,71%. Pour le deuxième groupe morphologique, nous avons un pourcentage de colonies d'abeilles bien classées de 100%, c'est-à-dire aucun mauvais classement n'a été enregistré dans les 24 colonies d'abeilles de ce groupe morphologique. Au niveau du groupe morphologique

Les résultats obtenus par le test de LEVENE ont mis en évidence des différences significatives pour la variation morphologique intra-coloniale pour le premier facteur ($F = 1,89$; *d.l.* (11, 27); $P = 0,03$) et des différences très hautement significatives ($F = 20,78$; *d.l.* (11, 27); $P = 0,00$) pour le deuxième facteur pour chaque rucher analysé (tab. 6).

Tableau 6 : Test de LEVENE de l'homogénéité des Variances inter-coloniales et intra-coloniales des 39 colonies étudiées.

* ($p \leq 0,05$); ** ($p \leq 0,01$); *** ($p \leq 0,001$); ns ($p > 0,05$).

4. Discussion

Au sens large, l'analyse à plusieurs variables ou analyse multidimensionnelle ou analyse multivariée peut être considérée comme formée de l'ensemble des méthodes statistiques qui ont pour objet l'étude des relations existant entre plusieurs variables dépendantes ou interdépendantes (Dagnelie, 1982).

Nous avons voulu mettre en évidence des phénomènes biogéographiques et génétiques chez l'abeille domestique en utilisant uniquement l'analyse multivariée sur plusieurs paramètres biométriques. A l'instar de plusieurs sous-espèces d'abeilles mellifères connues à travers le monde, notre race locale l'*Apis mellifera intermissa* dispose donc de populations régionales (écotypes) différenciables par leur morphologie.

Un résultat important de cette étude concerne la mise en évidence d'une variation graduelle des populations

d'Abeilles le long d'un transect partant du nord-est et allant jusqu'au sud, représenté par la présence de trois groupes morphologiques, qui ont été séparés nettement par l'analyse en composantes principales et l'analyse discriminante linéaire (fig. 1 et 2), et qui peuvent être considérés comme des écotypes adaptés de la race *A. m. intermissa*. Chaque écotype réunit donc un ensemble d'adaptation tout à fait spécifique à son environnement. Evidemment, les facteurs du milieu, comme le climat et la flore, maintiennent une hétérogénéité géographique du matériel génétique, et donc l'existence d'écotypes différenciables sur le plan morphométrique (Radloff, 1996). Il a été possible aussi d'effectuer de façon satisfaisante à l'aide de l'analyse hiérarchique la discrimination et la classification automatique de six classes d'abeilles qui sont considérées comme des sous-populations en cherchant toujours les plus proches voisins sous forme d'un dendrogramme de distances Euclidiennes (fig. 3). Nos résultats des différentes analyses statistiques utilisées dans cette étude biométrique ont mis en évidence la susceptibilité des caractères morphologiques aux effets de l'environnement, c'est à dire que la variation morphologique est bien corrélée avec la variation du milieu, ce qui renforce aussi l'idée que les populations d'Abeilles de la race *intermissa*, sont adaptées aux conditions locales (pluviométrie, température, végétation, climat,...etc.). Dans la mesure où le climat et la végétation au sud sont de type saharien, les Abeilles étant localisées dans les oasis, tandis qu'au nord-est, le climat est moins chaud et plus humide et la végétation est plus riche et géographiquement plus étendue, ce qui permet une apiculture plus répandue encouragée par la qualité mellifère de la région nord-est.

L'examen détaillé des résultats de l'analyse statistique multivariée reflète l'existence d'un écocline pour l'ensemble des échantillons provenant d'une région comprise entre le 32° et le 36° degré de latitude nord et le 3° et le 8° degré longitude Est. Ceci indique l'existence d'une variation biogéographique au sein des populations d'Abeilles interprétée par une hétérogénéité morphologique. En outre, l'ensemble des populations d'Abeilles de : Serraidi, Mewna, Souk Ahras, Tawra, Drean, Aïn Essel, El-Hadayeke et Bni Metrane sont nettement plus grosses que celles de Bni Isguen 1 et Bni Isguen 2 provenant de la région de Ghardaïa, ce qui confirme nettement et clairement la règle biogéographique de Bergmann qui stipule que la taille générale du corps d'Abeille s'accroît progressivement avec la latitude géographique. C'est à dire que les races des climats froids tendent à être plus grosses que les races de la même espèce vivant dans des climats plus chauds (Ruttner, 1988) ; sachant que les populations d'Abeille du nord peuplent des

biotopes où les étages bioclimatiques caractéristiques à ces régions sont en général de type sub-humide frais ou/et tempérés, tandis que l'étage bioclimatique du sud est de type saharien.

Les résultats du test de LEVENE (tab. 6 et 7) ont bien mis en évidence, deux types de variations morphologiques, l'une est inter-coloniale dans un même rucher, et l'autre est intra-coloniale. D'après Radloff & Hepburn (2000) la variation inter-coloniale au sein d'un même rucher s'explique essentiellement par la variance intra-coloniale, d'où nous pouvons également admettre ceci dans la mesure où toutes les 39 colonies d'Abeilles étudiées présentent une variation intra-coloniale. Cette dernière explique un phénomène purement génétique, c'est-à-dire qu'il existe de petites variations morphologiques significatives entre les ouvrières d'Abeilles vivant dans la même colonie, et qui présentent la descendance d'une seule reine. Plusieurs travaux concernant ce type de variation, ont bien mis en évidence ces deux types de variations phénotypiques chez plusieurs races d'Abeilles domestiques africaines, en utilisant toujours le test de LEVENE (Hepburn & Radloff, 1996 ; Radloff & Hepburn, 1997 ; Radloff *et al.*, 2003).

Il est à noter également, que le nombre de colonies prélevées affecte beaucoup plus la variation des valeurs moyennes des caractères morphologiques que le nombre d'Abeilles prélevées par colonie. Cependant, des caractères liés à la taille et à la venation des ailes sont considérablement moins affectés que ceux de la pigmentation de sorte que des groupes de 5 ou 6 colonies par localité avec 10 Abeilles par colonie semblent être adéquats dans l'analyse des données morphologiques des populations d'Abeilles. Les caractères de pigmentation devraient être inclus dans l'analyse statistique avec prudence à moins que les Abeilles soient connues pour être homogènes en couleurs. De plus, toutes combinaisons possibles de (*k*) colonies prélevées d'un total de (*n*) différents sites montrent que l'effet de la taille de l'échantillon des colonies sur les valeurs moyennes des caractères diffère selon les caractères morphométriques utilisés dans l'analyse, et affecte aussi l'homogénéité et/ou l'hétérogénéité de la variation morphométrique des Abeilles à l'intérieur et entre les colonies (Radloff *et al.*, 2003).

Le pouvoir discriminant des divers caractères morphologiques est discuté et la question du choix de ces caractères en fonction de la finalité de l'étude est abordée. De plus, aussi bien pour la collecte des données que pour les analyses statistiques, quelques améliorations doivent être proposées.

Références bibliographiques

- Alpatov W.W., 1929.** Biometrical studies on variation and races of the honey bee *Apis mellifera* L. *Quarterly Review of Biology*, **4**: 1- 58.
- Barour C., 2003.** Etude de la variabilité morphométrique et génétique de populations d'abeilles domestiques *Apis mellifera* L. (Hymenoptera : Apidea) du nord-est et de sud Algériens. Mémoire de Magistère, Université d'Annaba, Algérie.
- Bouroche J.M., & Saporta G., 1980.** L'analyse des données. 1^{ère} édition. *Presses universitaires de France*. Paris, 127p.
- Cornuet J.M., Fresnaye J. & Tassencourt, 1975.** Discrimination et classification d'abeilles à partir de caractères biométrique. *Apidologie*, **9**(1): 145-187.
- Chochlov B.A., 1916.** Investigations on the length of the bee tongue (In Russian). *Ministry of Agriculture, Petrograd*, 17-41.
- Dagnelie P., 1982.** Analyse statistique à plusieurs variables. *Gembloux, presses agronomiques A.S.B.L.*, 362p.
- Daly H.V., 1992.** A statistical and empirical evaluation of some morphometric variables of honey bee classification. In: Ordination in the Study of Morphology, Evolution and Systematics of Insects (Eds. *J.T. Sorensen & R. Footitt*), 127-155, Elsevier, New York, U.S.A.
- Dupraw E., 1965.** Non-Linear taxonomy and the systematics of honeybees. *Systematic Zoology*, **14** : 1-24.
- Fresnaye J., 1981.** Biométrie de l'abeille, 2e éd. Echauffour (Orme), Office Inf Doc Apic, 56 p
- Hepburn H.R., & Radloff S.E., 1996.** Morphometric and pheromonal analyses of *Apis mellifera* L. along a transect from the Sahara to the Pyrenees. *Apidologie*, **27**: 35-45.
- Louis J., Lefebvre J., Moratille R., & Fresnaye J., 1968.** Essai de discrimination de lignées consanguines d'abeilles domestiques (*A. m. mellifica* L.) obtenues par insémination artificielle. *Comptes Rendus, L'Académie de Science, Paris*, 267 Série D : 526-528.
- Michailov A.S., 1926.** On the increasing variability of the honey bee from north to south (In Russian). *Opytnaja Paseca*, **12**: 13-14.
- Palm R., 1998.** L'analyse en composantes principales : principes et applications. *Notes stat. Inform. (Gembloux)* 98/2,31p.
- Palm R., 1999.** L'analyse discriminante décisionnelle : principes et applications. *Notes stat. Inform. (Gembloux)* 99/4, 41p.
- Palm R., 2000.** L'analyse de la variance multivariée et l'analyse canonique discriminante : principes et applications. *Notes stat. Inform. (Gembloux)* 2000/1, 40p.
- Radloff S.E., 1996.** Multivariate analysis of selected honeybee populations in Africa. Ph.D. thesis, Rhodes university, Grahamstown.
- Radloff S.R., & Hepburn H.R., 1997.** Multivariate analysis of honeybees, *Apis mellifera* Linnaeus (Hymenoptera: Apidea), of the Horn of Africa. *African Entomology*, **5**(1): 57-64.
- Radloff S.R., & Hepburn H.R., 1999.** Honeybees, *Apis mellifera* Linnaeus (Hymenoptera : Apidae), of the Drakensberg Mountains in relation to neighboring populations. *African Entomology*, **7**(1): 35-41.
- Radloff S.E., & Hepburn H.R., 2000.** Polpulation structure and morphometric variance of the *Apis mellifera scutellata* group of honeybees in Africa. *Genetics and Molecular Biology*, **23**(2): 305-316.
- Radloff S.R., Hepburn H.R., Robertson M.P., Hille R.V., Davidson Z., & Villet M.H., 1996.** Discriminant analysis of the honeybee populations of southwestern Africa. *African Entomology*, **4**(1): 1-6.
- Radloff S.E., Hepburn H.R. & Fuchs S., 1998.** Ecological and morphological differentiation of the honeybees, *Apis mellifera* Linnaeus (Hymenoptera: Apidea), of west Africa. *African Entomology*, **6**(1): 17-23.
- Radloff S.E., Hepburn H.R., & Bangay L.J., 2003.** Quantitative analysis of intracolony and intercolony morphometric variance in honeybees, *Apis mellifera* and *Apis cerana*. *Apidologie*, (sous presse).
- Rinderer T.E., Daly H.V., Sylvester H.A., Collins A.M., Buco S.M., Hellmich R.L. & Danka R.G., 1990.** Morphometric differences among Africanized and European honey bees and their F1 hybrids (Hymenoptera : Apidae). *Annals of the Entomological Society of America*, **83**(3) : 346-351.
- Ruttner F., 1988.** Biogeography and Taxonomy of honeybees. *Springer Verlag, Berlin*, 284p.
- Ruttner F., Tassencourt L., & Louveaux J., 1978.** Biometrical statistical analysis of the geographic variability of *Apis mellifera* L., *Apidologie*, **9**: 363-381.
- Tomassone R., & Fresnaye, J., 1971.** Étude d'une méthode biométrique et statistique permettant la discrimination et la classification de populations d'abeilles (*Apis mellifica* L.). *Apidologie*, **2**(1): 49- 65.
- X. (2000).** Minitab reference manual, release 13.31 for windows, *PA State College, Minitab*, 1047 p.

Stochastic Regionalization of hydrogeological Data with Fuzzy-Kriging

Dr. Blél Azouzi, Dr. Jürgen Schröter and Lazreg Benaichatta

azouzi@uni-bremen.de/

1 Introduction

The aim of this paper is to present an application of parameter estimation methods and stochastic regionalization procedures for the development of boundary conditions as input for regional hydraulic models. The study example illustrates the application of Fuzzy-Kriging for the enhancement of the standard Kriging error of a Transmissivity map of the Mitidja area (Algeria). The given dataset of measured transmissivities was supplemented by transmissivity range estimates derived from imprecise hydrogeological information. Since in the context of the geostatistic procedures the hydrodynamic parameters are considered to be random variables, mostly very irregularly distributed, they can be regarded as fuzzy data in the sense of the Fuzzy logic, which gives the opportunity to incorporate additional expert knowledge based information to fill in blank areas.

2 Methodology

2.1 Kriging

There are several regionalization techniques in use (for example, an inverse distance method), but this paper focuses on random or stochastic regionalization by Kriging. Kriging is a geostatistical interpolation method of determining the best estimate for each point in a target matrix, based on statistical principles. Kriging is named after a South African engineer, D. G. Krige, who first developed the method (Chiang and Kinzelbach, 1998). Kriging works best with datasets which have regions of densely scattered data and other regions of lightly scattered data. The process of kriging interpolates values for a grid of points based on the weighted grades of points surrounding it (1). The kriging weights are based on the variogram model parameters (Nugget effect, Sill and Range) (fig.1) and the anisotropy factors for the search ellipsoid (Houlding, 2000).

$$Z_o^* = \sum \lambda_i \cdot Z(x_i) \quad (1)$$

Z_o^* : estimated value

$Z(x_i)$: Known values at the location x_i

λ_i : weighting factors dependent on distance and eventually on direction

The advantages of kriging over other estimation methods are:

- the kriging equations minimise the variance of the estimates
- the kriged estimates are conditionally unbiased, that is, they will show no bias provided that mining is carried out according to the block estimates

- the kriging process handles clustered data very well. If the data used for kriging is on a regular grid, then kriging will probably provide no better results than an inverse distance method.

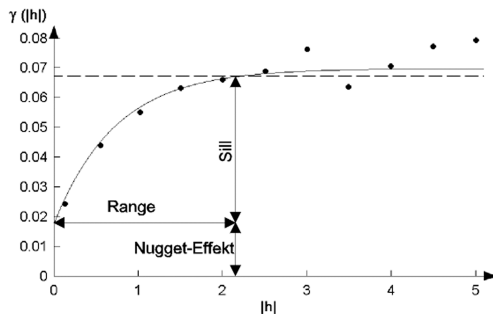


Fig. 1. Variogram model parameters

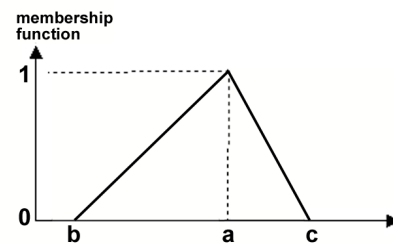


Fig. 2. Parameter of fuzzy number

2.2 Fuzzy Kriging

The fuzzy-logical principle was introduced in 1965 by Lotfi A. Zadeh (Berkley, USA) as an extension of binary mathematics and gradation of bivalence between 0 and 1 (Kosko 1993). Thus a fuzzy set or fuzzy number A is defined with the help of a characteristic function F_A between 0 and 1, which is designated as Membership function (Bothe 1993). Then the set A of the pair $(x, F_A(x))$ with

$A = \{(x; F_A(x)) \mid x \in X, F_A(x) \in R\}$ is a fuzzy set on X (a set of elements x) of the membership function $F_A(x)$,

where:

$$F_A(x) = \begin{cases} 1 & x \in R \\ 0 & \text{sonst} \end{cases} \quad (1)$$

The membership function was characterized by three values estimated subjectively by an expert, **a**. the most likely value, **b**. the lowest possible value and **c**. the highest possible value (fig. 2).

In order to improve the spatial estimation of a variable, experts have the opportunity of adding a number of fuzzy data to real measurement values (crisp data) on the basis of their subjective knowledge by means of this method (Bandemer and Gottwald 1993). The principle of the Fuzzy logic as the means of interpolation of the hydraulic geological parameters was used e.g. by Bardossy et al, 1989 and Piotrowski et al, 1997. Fuzzy Kriging is an extension of the normal Kriging. The computation and the consideration of the spatial structure through the Variogram are exactly the same as for the normal Kriging. Depending on the different types of data input and variograms 3 different Fuzzy Kriging Types defined (Bardossy et al. 1989; Piotrowski et al. 1996):

- fuzzy kriging type 1: with both crisp and fuzzy data and crisp (no fuzzy) variogram,
- fuzzy kriging type 2: with crisp data and fuzzy variogram,

- fuzzy kriging type 3: with both crisp and fuzzy data and fuzzy variogram.

Fuzzy kriging type 1 was selected for the computation logarithmically transformed transmissivities in this study. Since the Fuzzy numbers are defined by the membership function, and the membership function is fixed by α -cuts (a, b and c in figure 2, $\alpha \in [0, 1]$), the fuzzy kriging is performed on the basis of the extension principle of the normal kriging via formed intervals (I_1, \dots, I_n) of the fuzzy variable ($Z(x_1), \dots, Z(x_n)$) between 0 and 1 of the membership function. Thus the Fuzzy variogram (2) and the main fuzzy kriging equation (3) are formed (Dragicevic and Marceau, 2000):

$$2\delta(h, I_1, \dots, I_n) = \{y \in \mathbb{R} \mid \exists (x_1, \dots, x_n) \in (I_1, \dots, I_n) : y = 1/2n(h) \sum [Z(x) - Z(x+h)]^2\} \quad (2)$$

$$Z^*(x, Z(x_1) \dots Z(x_n)) = \sum \lambda_i \cdot Z(x_i) \quad (3)$$

2.3 Description of Study Area

The study area Mitidja, which is the most expanded, sublittoral, Algerian plain and covers approximately 1300 km², stretching 12-18 km southwards to the Atlas, from the river Boudouaou, the eastern borderline, over 100 km westwards to the basin of the river Menaceur over Hadjout, is located in the centre of North Algeria. It is limited by small Sahel from the north and the east (fig.3).

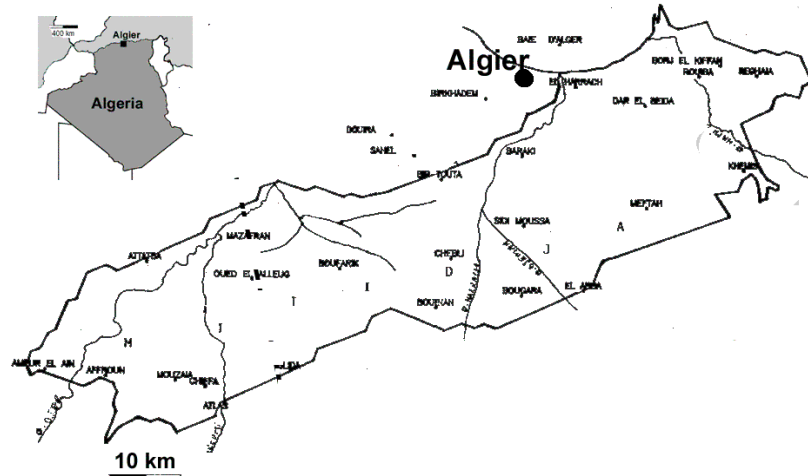


Fig. 3. Location of the study area Mitidja (Northern Algeria).

Hydrogeologically the Mitidja is built up of two important aquifer reservoirs, which are separated by the impermeable marl layer of El Harrach on almost the entire territory with an exception in the east (Bennie, 1983):

- the aquifer Astien, consisting of calcareous sand,
- the overlaying quaternary alluvium, the aquifer Mitidja, consisting of loam and gravel deposits of different thickness and expansion.

The hydrogeological dataset with 90 transmissivity measuring values were collect by the ANRH (Agence Nationale des Ressources Hydriques) from 1968 to 1978, and they are distributed irregularly in Mitidja aquifer.

The sediments deposited by the large rivers Djer, Bouroumi, Chiffa and Mazafran show generally higher Transmissivities around. $23 \text{ E-4 m}^2/\text{s}$ than those deposited at the Piedmont mountains ($T = 33 \text{ E-4} - 77 \text{ E-4 m}^2/\text{s}$). In the eastern part of the area near Harrach Hamiz and in Rouiba, i.e., in the contact zone of the aquifers of Astien and Mitidja, the transmissivity is increased up to $\text{E-2 m}^2/\text{s}$ (Azouzi 1993). It is possible that these transmissivity result from the contact of the two aquifers, since the thickness in this zone is large.

3 Presentation of results

For the application of the Fuzzy Geostatistics during the regionalization of the transmissivity the program FUZZEKS 1.0 developed by Bartels (1997) was used. First an interpolation of the log-transformed measurements was performed by means of the Kriging procedure on a mesh of 150×76 cells of $500 \times 500 \text{ m}$ cells. The regionalized transmissivity is represented in figure 4. The experimental Variogram which was computed only with exact measurement data, increases on a range of 10 km and then is stable and constant (fig. 4). The Variogram model is spherical with a nugget effect = 0.13 and a sill = $0.29 (\text{m}^2/\text{s})^2$. These Variogram parameters are very similar to the Fuzzy Variogram parameters.

High transmissivity values are located in the northeast (Rouiba) and in the zone Mazafran. The high values in the east are probably caused by hydrodynamic exchange in the contact zone between the alluvium aquifer of Mitidja and that of Astien. Around Mazafran the increase of the Transmissivity can be explained by the higher permeability of the aquifer sediments. Particularly in the central part and in the west of the area strongly varying Transmissivity can be found, which result from irregular data distribution. The lack of observations and data concerning the entire investigation area affects directly the quality of the interpolation. The Kriging variance of the regionalised log-T-values close to the measured values ranges below $0.05 (\text{m}^2/\text{s})^2$ and rises to $0.15 (\text{m}^2/\text{s})^2$ in extrapolated external areas.

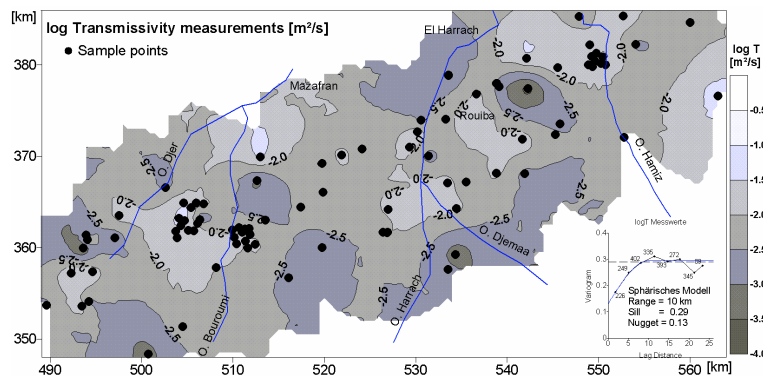


Fig. 4. Contour lines of Log transmissivity estimated by Kriging

In order to improve the regionalization inside the area with lower data density, in points with highest Kriging variance 20 data were added as Fuzzy numbers. For each point 3 parameters

a, b and c of a fuzzy number (fig.2) was estimated on the basis of the membership function and geological and hydraulic geological knowledge of the investigation area.

Since a small number of additional fuzzy data points does not affect the spatial structure of a parameter very much, the new fuzzy variogram (fig.5) is very similar to the Variogram of crisp data (same range = 10 km, nugget effect = 0,1157 and sill = 0,269 (m²/s)²). The new fuzzy contour lines of the transmissivity differ little from the normal Kriging contour lines (using no Fuzzy data) except in the regions where Fuzzy data was used, particularly in the southwest of the investigation area. The Kriging variance is strongly reduced by the introduction of the additional fuzzy data in this area. It decreases in this computation below 0.05 (m²/s)².

A characteristic of this method is the estimation of the lowest and highest possible values in form of a contour lines map. In the context of environmental questions it is an important means for the simulation of the different scenarios of the groundwater dynamics and the material transfer in the groundwater (Azouzi, 2002). The probable deviation of the transmissivity in the Mitidja is represented by the difference map of the estimated lowest and highest possible values in figure 6. Larger local deviations are especially in the southwest area (Fig 6).

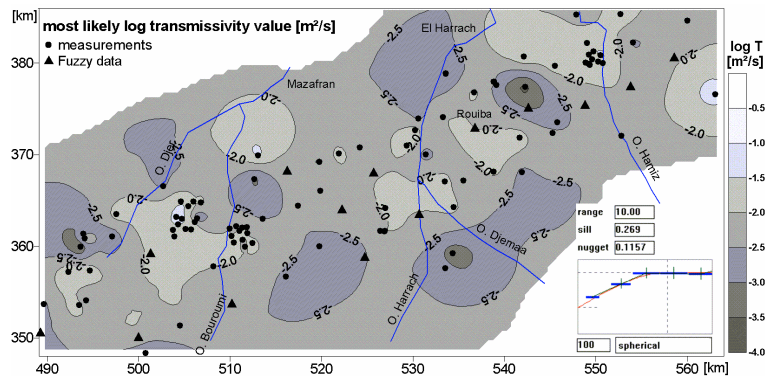


Fig. 5. Contour lines of the most likely transmissivity values estimated by fuzzy kriging

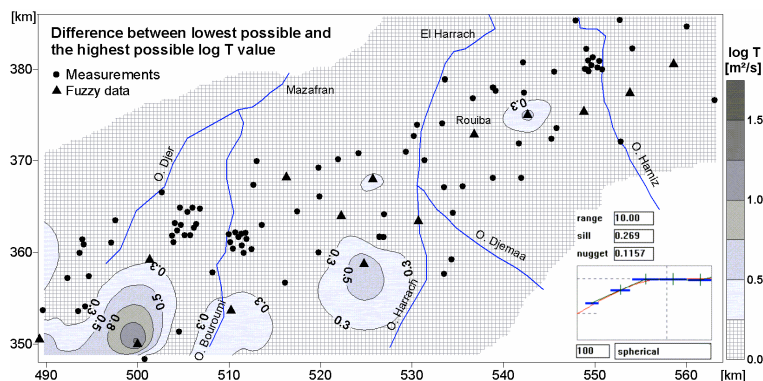


Fig. 6. Difference between lowest possible and the highest possible value of transmissivity

4 conclusions

The principal objective of this study was the reconstruction of a Transmissivity distribution by integration of fuzzy information about geological structure and description of sediment of an aquifer in the north of Algeria. By addition of the hydrogeological knowledge of the Aquifer, in the eastern part of Mitidja, as well as that of Harrach Hamiz and a part of Mazafran, intervals of Transmissivity were estimated from thickness and sediment descriptions in analogy to existing evaluations of this parameter, in order to reach a quality improvement of the implemented estimation in areas of small data density.

The generation and regionalization of hydrogeological data via the study describing the improvement of Transmissivity distribution calculations by supplementation of Fuzzy numbers are shown for the parameter generation as an input for mathematical hydraulic models. This case is an example for the benefit deriving from the combination of expert knowledge combined with existing stochastic tools for the enhanced interpretation of existing data.

References

- Azouzi B (2002) Geostatistische Regionalisierung hydrochemischer Parameter des Grundwasserleiters. Berichte, Fachbereich Geowissenschaften, Univ. Bremen, 202, 108 p.
- Azouzi B (1993) Contribution à la cartographie automatique des grandeurs hydrochimiques Application à la Mitidja. 163 pp. Thèse de Magistère, Ed. INA d'El Harrach, Alger
- Bandemer H, Gottwald F (1993) Einführung in Fuzzy-Methoden. 264 pp, Akademie-Verlage, Berlin
- Bardossy A, Borgardi I, Kelly WE (1989) Geostatistics utilizing imprecise (Fuzzy) information. *Fuzzy-Sets and Systems* 31: 311-327
- Bartels F (1997) Ein Fuzzy-Auswertung- und Krigingsystem für raumbezogene Daten. 92 pp, Dipl. Arbeit, Christian-Albrechts-Universität Kiel
- Bennie P (1983) Schémas d'aménagement des ressources en eau de la région d'Alger et Sebaou. Mission B et F, Vol 4., ANRH. Alger
- Bothe HH (1993) Fuzzy Logik, Einführung in Theorie und Anwendungen. 226 pp, Springer-Verlag Berlin, Heidelberg
- Chiang WH, Kinzelbach (1998) Processing Modflow, A Simulation system for modeling groundwater flow and pollution. 346 pp, User's manual.
- Dragicevic S, Marceau D (2000): A fuzzy set approach for modeling time in GIS. *International Journal of Geographical Information Science*. Vol 14, no. 3, pp 225-245.
- HOULDING, S. W. (2000): Practical geostatistics. Modeling and spatial analysis.- 159 S.; Springer, Berlin.
- Kosko B (1993) Fuzzy-Logisch ,eine neue Art Denkens. 352 pp, Carlsen-Verlag GmbH, Hamburg
- Piotrowski J, Bartels F, Salski A, Schmidt G (1996) Geostatistical regionalization of glacial aquitard thickness in north-western Germany, based on Fuzzy Kriging. *Mathematical Geology* 28(4): 437-452
- Piotrowski J, Bartels F, Salski A, Schmidt G (1997) Regionalisierung der Durchlässigkeitsbeiwerte mit unscharfen (Fuzzy) Zahlen: Der Natur näher? *Grundwasser Zeitschr. der Fachsektion Hydrogeologie* 2(1): 3-10.
- Zadeh LA (1965) Fuzzy sets: Information and control. Vol 8, no. 3, pp 338-353.

Analyse en Composantes principale sous contrainte fonctionnelle. Application a l'Analyse Factorielle d'une Fonction Aléatoire.

T. Benchikh & A. Yousfate
Laboratoire des Mathématiques.

Département des mathématiques. Faculté des Sciences
UDL, Sidi bel Abbès 22000, algérie.

e-mail : tbenchikh@umiv-sba.dz yousfate@umiv-sba.dz

Résumé :

Dans ce travail nous présentons l'A.C.P. " pas à pas" sous contrainte de $\mathcal{B}_1 \otimes \mathcal{B}_2$ -mesurabilité d'un opérateur linéaire continu U associé à un processus de 2^{ième} ordre sous l'aspect d'une optimisation sous contrainte linéaire itérée ou globale. Dans le cas fini, l'ACP sous contrainte linéaire a été traitée par plusieurs auteurs : Tebboune (choix de métrique), Sabatier (variable instrumentales), Dans le cas infini on trouve les travaux de Boudou (sous contrainte de \mathcal{B} -mesurabilité), Tchene (sous contrainte d'orthogonalité ou d'inclusion) Smadhi (A.C.P.C.L. sur variables hilbertiennes), Yousfate (ACP et filtrage), Besse (Transformations splines), Rachedi (A.C.P.C.L. sur les L^p transformés en des espaces de Hilbert),

Mots clés : A.C.P.C.L. , Mesure spectrale, Mesure aléatoire, Projecteur, Processus Stationnaire, Produit tensoriel.

Classification AMS : 47Axx, 47Bxx, 60G12, 60G57, 62H11, 62H25.

1 introduction

Les techniques factorielles permettant une certaine décomposition de l'espace des "résultats" d'observations d'un phénomène étudié ont été développées par Pearson au début du XX^e siècle. Rao, e. 1964, a synthétisé les techniques factorielles connues jusqu'alors dans son article "" en utilisant des espaces de dimension finie. En 1976, Dauxois et Pousse ont réussi à étendre l'A.C.P. aux espaces de Hilbert. Depuis de nombreux résultats concernant la convergence, le lissage, le modèle linéaire, ont été développés. Par ailleurs, nous avons développé l'A.C.P. dans des Banach réflexif. Dans ce travail nous intéressons à l'A.C.P. dans un espace de Hilbert avec la particularité que les solutions sont contenues dans un espace de référence: c'est ce que nous allons appeler A.C.P. sous contrainte linéaire ou A.C.P.C.L.. L'approche de l'A.C.P.C.L. dans des espaces de dimension finie a été développée par Tebboune en 1981 et repris par Sabatier par la suite. Dans un espace de Hilbert l'A.C.P.C.L. a été développée par SMADHI & Yousfate en 1992 ensuite par différents auteurs dont RACHEDI, BENCHIKH et YOUSFATE....

2 Rappels sur l'A.C.P.C.L.

Pour présenter notre étude, nous considérons deux espaces de Hilbert réels et séparables H_1 et H_2 que nous identifions à leurs duaux topologiques respectifs H_1' et H_2' . Nous supposons également que F_1 (resp. F_2) est un sous-espace fermé de H_1 (resp. H_2) et que $U : H_1 \rightarrow H_2$ est un morphisme injectif et continu.

L'A.C.P. de U revient alors à faire la décomposition spectrale de l'opérateur $V = U \circ U^*$ ou de $W = U^* \circ U$. En d'autres termes nous présentons l'A.C.P. comme une solution générale au problème suivant :

$$\begin{cases} \max_{u \in H_2} \|U^*u\| = \max_{u \in H_2} \langle U \circ U^*u, u \rangle \\ \|u\|_{H_2} = 1. \end{cases} \quad (1)$$

et itération sous contrainte d'orthogonalité des arguments des solutions (si les solutions itératives ne parcourent pas tout le spectre de V , nous considérons l'A.C.P. globale (voir ???)

Dans le cadre de l'A.C.P.C.L₂, nous posons le problème itératif de A.C.P.C.L₂ comme suit :

$$\begin{cases} \max_{u \in F_2} \|U^*u\| \\ \|u\|_{H_1} = 1: \end{cases} \quad (2)$$

et itération sous contrainte d'orthonormalité des arguments des solutions.

En notant P_2 le projecteur tel que $P_2(H_2) = F_2$, le problème (2) devient

$$\begin{cases} \max_{u \in H_2} \|U^* \circ P_2u\| \\ \|u\|_{H_2} = 1: \end{cases}$$

Ainsi, nous remarquons que l'A.C.P.C.L₂ de U sous F_2 correspond à l'A.C.P. de $P_2 \circ U$ qui consiste à décomposer l'opérateur $P_2 \circ U \circ U^* \circ P_2 = P_2 \circ V \circ P_2$ ou de $U^* \circ P_2 \circ U$. De la même manière, on montre que l'A.C.P.C.L₁ sous F_1 correspond à l'A.C.P. de $U^* \circ P_1$ qui consiste à décomposer l'opérateur $P_1 \circ W \circ P_1$ ou $U \circ P_1 \circ U^*$.

En général l'A.C.P.C.L₁ et l'A.C.P.C.L₂ sont différentes.

On montre facilement que :

$$\begin{aligned} \Pi_i : \mathcal{L}(H_i) &\longrightarrow \mathcal{L}(H_i) \\ M &\longmapsto P_i \circ M \circ P_i \end{aligned}$$

pour $i = 1, 2$, est un projecteur.

Si on note $\mathcal{A}_+^i = \{M : M \text{ associé à un semi produit scalaire } \}$, on a $\Pi_i(\mathcal{A}_+^i) \subset \mathcal{A}_+^i$.

Ces techniques d'A.C.P.C.L. peuvent être étendue au cas mixte en mettant une contrainte sur H_1 et une contraintes sur H_2 simultanément. On montre, en suivant la même technique que l'A.C.P.C.L. mixte revient à décomposer spectralement l'opérateur $V_{1,2} = P_1 \circ U \circ P_2 \circ U^* \circ P_1$ ou $W_{1,2} = P_2 \circ U^* \circ P_1 \circ U \circ P_2$.

3 Rappels sur l'algèbre opératorielle

En termes opératoriels, si E et F sont des espaces de Hilbert, E' et F' leurs duaux topologiques respectifs et si $U_i \in \mathcal{L}(F', E)$, $i = 1, 2$, on écrit

$$\begin{aligned} U_1 \overset{\wedge}{\otimes} U_2 : \mathcal{L}(F', F) &\longrightarrow \mathcal{L}(E', E) \\ M &\longmapsto U_2 \circ M \circ U_1^* \end{aligned}$$

De la même manière, $U_1^* \overset{\cdot}{\otimes} U_2^*$ est représenté comme un élément de $\mathcal{L}(\mathcal{L}(E, E'), \mathcal{L}(F', F))$. Bien que dans le premier cas $\overset{\cdot}{\otimes}$ est un opérateur de $\mathcal{L}([\mathcal{L}(F', E)]^2, \mathcal{L}(\mathcal{L}(F, F'), \mathcal{L}(E', E)))$ tandis que dans le second cas $\overset{\cdot}{\otimes}$ est un opérateur de $\mathcal{L}([\mathcal{L}(E', F)]^2, \mathcal{L}(\mathcal{L}(E, E'), \mathcal{L}(F', F)))$ on ne fera pas de distinction symbolique dès que l'on détermine la nature des U_i . Par la suite nous ne mentionnerons plus les espaces fonctionnels considérés sauf quand c'est nécessaire.

Dans le calcul opératoire, si $U_i \in \mathcal{L}(E)$, on montre que $U_1 \overset{\cdot}{\otimes} U_2$ est un projecteur dès que U_1 et U_2 sont des projecteurs. Les propriétés algébrique de $\overset{\cdot}{\otimes}$ peuvent être consultées dans [6].

4 Calcule opératoire sur L'A.C.P.C.L.

Dans le cadre de l'A.C.P.C.L., $\Pi_i(M) = P_i \overset{\cdot}{\otimes} P_i(M)$ où $M = W$ (resp. V) dans le cas de l'A.C.P.C.L₁ (resp. A.C.P.C.L₂).

Dans le cas l'A.C.P.C.L. mixte, l'opérateur à décomposer s'écrit $W_{2,1} = (P_2 \overset{\cdot}{\otimes} P_2)[(U \overset{\cdot}{\otimes} U)P_1]$. Cette décomposition est équivalente à celle de l'opérateur $W_{1,2} = (P_1 \overset{\cdot}{\otimes} P_1)[(U^* \overset{\cdot}{\otimes} U^*)P_2]$.

Ainsi l'opérateur d'intérêt pour l'A.C.P. classique est $U \overset{\cdot}{\otimes} UI$ ou $U^* \overset{\cdot}{\otimes} U^*I$.

Pour l'A.C.P.C.L₁ on a $U \overset{\cdot}{\otimes} UP_1$ ou $P_1 \overset{\cdot}{\otimes} P_1(U^* \overset{\cdot}{\otimes} U^*I)$ et pour l'A.C.P.C.L₂ on a $U^* \overset{\cdot}{\otimes} U^*P_2$ ou $P_2 \overset{\cdot}{\otimes} P_2(U \overset{\cdot}{\otimes} UI)$.

Dans le cas où l'on considère une suite P_{1j} (resp. P_{2k}) de projecteurs tels que les P_{1j} (resp. P_{2k}) vérifient $P_{1j_1} \circ P_{1j_2} = 0$ si $j_1 \neq j_2$ (resp. $P_{1k_1} \circ P_{1k_2} = 0$ si $k_1 \neq k_2$). $W_{2,1}$ peut être décomposé comme suit :

$$W_{2,1} = \sum_{j=1}^{n_1} \left(\sum_{k=1}^{n_2} P_{2k} \overset{\cdot}{\otimes} \sum_{k=1}^{n_2} P_{2k} \right) [(U \overset{\cdot}{\otimes} U)P_{1j}]$$

on remarque immédiatement que

$$W_{2,1} = \sum_{j=1}^{n_1} \sum_{k=1}^{n_2} P_{2k} \overset{\cdot}{\otimes} P_{2k} [(U \overset{\cdot}{\otimes} U)P_{1j}]$$

si et seulement si les P_{2k} commutent avec U .

Ainsi si les P_{2k} sont obtenus à partir de la décomposition spectrale de $V = U \circ U^*$, l'A.C.P.C.L. mixte par "paquets" passe de $n_1 * n_2^2$ projections $(P_{2k} \overset{\cdot}{\otimes} P_{2k'})[(U \overset{\cdot}{\otimes} U)P_{1j}]$ à $n_1 * n_2$ projections $P_{2k} \overset{\cdot}{\otimes} P_{2k}[(U \overset{\cdot}{\otimes} U)P_{1j}]$ qui rejoint l'approche de la technique développée par Tebboune dans le cas fini ensuite par Sabatier.

Si ce sont les P_{1j} qui commutent avec U , alors on considère $W_{1,2} = \sum_{k=1}^{n_2} \sum_{j=1}^{n_1} P_{1j} \overset{\cdot}{\otimes} P_{1j} [(U^* \overset{\cdot}{\otimes} U^*)P_{2k}]$ Dont la décomposition spectrale est équivalente à celle de $W_{1,2}$. En général les $P_{2k} \overset{\cdot}{\otimes} P_{2k'}[(U \overset{\cdot}{\otimes} U)P_{1j}]$ (dont on note l'image $F_{2,kk'j}$) ne sont pas nuls même si $k' \neq k$. Pour réduire le nombre de sous espaces de projection nous regroupons, par analyse canonique, les sous-espaces $F_{2,kk'j}$ qui sont jugés proches (voir Dauxois, Romain, Viguier [4]).

5 Comparaison d'A.C.P.C.L.

5.1 comparaison d'opérateurs par l'A.C.P.

Si l'on veut comparer les opérateurs U_1 et $U_2 \in \mathcal{L}(H_2, H_1)$ nous avons la méthode arithmétique qui consiste à décomposer U_1 et U_2 comme suit :

$$U_1 = S + A \quad U_2 = S_A$$

où $S = \frac{U_1 + U_2}{2}$ et $A = \frac{U_1 - U_2}{2}$.

Pour faire l'A.C.P. comparative arithmétique de U_1 et U_2 on considère l'opérateur

$$\frac{1}{2}(U_1 \overset{\cdot}{\otimes} U_2 + U_2 \overset{\cdot}{\otimes} U_1) = S \overset{\cdot}{\otimes} S - A \overset{\cdot}{\otimes} A$$

où $S \overset{\cdot}{\otimes} S$ exprime une A.C.P. de ce qui est commun à U_1 et U_2 et $A \overset{\cdot}{\otimes} A$ exprime une A.C.P. de ce qui est différent entre U_1 et U_2 .

Dans un modèle où A serait "orthogonal" à S , nous pouvons utiliser la complexification suivante de U_1 et U_2

$$\tilde{U}_1 = S + iA \quad \text{et} \quad \tilde{U}_2 = S - iA$$

L'opérateur

$$\begin{aligned} U_1 \overset{\cdot}{\otimes} U_2 &= (S + iA) \overset{\cdot}{\otimes} (S - iA) \\ &= S \overset{\cdot}{\otimes} S - A \overset{\cdot}{\otimes} A + i(S \overset{\cdot}{\otimes} A + A \overset{\cdot}{\otimes} S) \end{aligned}$$

Les termes $S \overset{\cdot}{\otimes} S$, $A \overset{\cdot}{\otimes} A$ et $S \overset{\cdot}{\otimes} A + A \overset{\cdot}{\otimes} S$ appliqués à un opérateur symétrique donnent tous un opérateur symétrique dont on peut faire aisément la décomposition spectrale. Nous avons également la méthode géométrique qui consiste à étudier :

$$U_1 \circ U_2^* \circ U_2 \circ U_1^* = (U_1 \overset{\cdot}{\otimes} U_1)[(U_2 \overset{\cdot}{\otimes} U_2)I]$$

ou

$$U_1 \circ U_1^* \circ U_2 \circ U_2^* = (U_1 \overset{\cdot}{\otimes} U_2)[(U_1 \overset{\cdot}{\otimes} U_2)I]$$

Nous rejetons le premier cas, car il ne représente pas l'opérateur d'une A.C.P., contrairement au second qui représente l'opérateur de covariance de $U_1 \overset{\cdot}{\otimes} U_2$.

Une étude particulière peut être développée si les U_i , $i = 1, 2$ sont des endomorphismes sur le même espace de Hilbert dont on considère qu'il est identifié à son dual topologique. Dans ce cas, nous pouvons introduire un autre cas dans la méthode arithmétique :

$$U_1 = S + A \quad \text{et} \quad U_{*2} = S - A$$

avec $S = \frac{U_1 + U_{*2}}{2}$ et $A = \frac{U_1 - U_{*2}}{2}$. S alors autoadjointe et A antisymétrique. La complexification de U_1 et U_2 se fera alors comme suit :

$$\tilde{U}_1 = S + iA \quad \text{et} \quad \tilde{U}_2 = S - iA$$

5.2 Comparaison d'opérateurs par l'A.C.P.C.L.

Dans la méthode arithmétique réelle, on considère l'opérateur :

$$(P_2 \overset{\cdot}{\otimes} P_2)[(S \overset{\cdot}{\otimes} S - A \overset{\cdot}{\otimes} A)P_1]$$

ou

$$(P_1 \overset{\cdot}{\otimes} P_1) \overset{\cdot}{\otimes} (S^* \overset{\cdot}{\otimes} S^* - A^* \overset{\cdot}{\otimes} A^*)$$

où P_1 et P_2 sont les projecteurs ayant pour image respectives les sous-espaces de contrainte linéaire. L'A.C.P.C.L₁ considère $P_2 = I$ et l'A.C.P.C.L₂ considère $P_1 = I$.

Si A et S peuvent subir des contraintes séparées la méthode arithmétique réelle consiste à étudier les opérateurs $\{(P_{12} \overset{\cdot}{\otimes} P_{12})[(S \overset{\cdot}{\otimes} S)P_{11}]$ et $(P_{22} \overset{\cdot}{\otimes} P_{22})[(A \overset{\cdot}{\otimes} A)P_{21}]\}$ ou $\{P_{11} \overset{\cdot}{\otimes} P_{11}[(S^* \overset{\cdot}{\otimes} S^*)P_{12}]$ et $(P_{21} \overset{\cdot}{\otimes} P_{21})[(A^* \overset{\cdot}{\otimes} A^*)P_{22}]\}$ où les P_{ij} sont les projecteurs associés aux contraintes.

Dans la méthode géométrique réelle, on considère l'opérateur :

$$(\tilde{P} \overset{\cdot}{\otimes} \tilde{P})[(U_1 \overset{\cdot}{\otimes} U_2) \overset{\cdot}{\otimes} (U_1 \overset{\cdot}{\otimes} U_2)P]$$

où $P = P_{11} \overset{\cdot}{\otimes} P_{12}$ et $\tilde{P} = P_{21} \overset{\cdot}{\otimes} P_{22}$.

P est alors un projecteur orthogonal dans $\mathcal{L}(H_1) \cong H_1 \otimes H_1$ et \tilde{P} est alors un projecteur orthogonal dans $\mathcal{L}(H_2) \cong H_2 \otimes H_2$.

6 A.C.P.C.L. sous contrainte fonctionnelle d'un processus stationnaire

Rappelons qu'un processus $(X_t)_{t \in T}$ de second ordre centré admet une A.C.P. associée à l'opérateur

$$\begin{aligned} U : L^2(\Omega) &\longrightarrow L^2(T) \\ h &\longmapsto Uh = \int_{\Omega} X_t(\omega)h(\omega)d\mathbb{P}(\omega) \end{aligned}$$

l'ACP de $(X_t)_{t \in T} \implies$ analyse spectrale de $V = U \circ U^*$ (opérateur de covariance) ou celle de $W = U^* \circ U$ (opérateur d'Escoufier) du type :

$$X_t = \sum_{k=1}^{\infty} \sqrt{\lambda_k} \xi_k f_k(t) \quad m.q.$$

où les ξ_k sont des variables aléatoires normées ne dépendant pas du temps et les $f_k(t)$ des fonctions déterministes normées exprimées en fonction du temps.

Proposition 6.1 (Décomposition de Karhunen-Loève) Pour que $X_t = \sum_{k=1}^{\infty} \sqrt{\lambda_k} \xi_k f_k(t)$ m.q. où les $f_k(t)$ et les ξ_k forment des systèmes orthonormaux respectifs de $L^2(T)$ et de $L^2(\Omega)$, il faut que les f_k soient un système de fonctions propres de l'opérateur de covariance V du processus et les λ_k les valeurs propres correspondantes.

Nous avons également si $(X_t)_{t \in T}$, $T = \mathbb{R}$, est un processus stationnaire au sens large, alors $(X_t)_{t \in T}$ admet la représentation de Cramer :

$$X_t = \int_{-\infty}^{+\infty} e^{it\lambda} dZ(\lambda)$$

où $Z(\lambda)$ est un processus complexe à accroissements orthogonaux.

Faire l'ACPCL sur un tel processus revient à faire une ACP, tout en introduisant une projection sur $L^2(\Omega)$, sur $L^2(T)$ ou sur les deux simultanément.

Soient $(X_t)_{t \in T}$ et $(Y_t)_{t \in T}$ deux processus stationnaire à valeur dans $H = L^2(\Omega, \mathcal{A}, \mathbf{P})$ de mesure aléatoire Z_1 et Z_2 respectivement et soient \mathcal{B}_1 et \mathcal{B}_2 deux sous tribus indépendante de \mathcal{A} . On note \mathcal{C} la sous tribu de \mathcal{A} engendrée par la famille de parties $\{B_1 \cap B_2 : B_1 \in \mathcal{B}_1, B_2 \in \mathcal{B}_2\}$, il existe alors une isométrie \mathcal{I} entre les espaces $\sigma_2(L^2(\mathcal{B}_1), L^2(\mathcal{B}_2))$ et $L^2(\mathcal{C})$ qui, à $y_1 \otimes y_2$ associe $y_1 \cdot y_2$, $\forall (y_1, y_2) \in L^2(\mathcal{B}_1) \times L^2(\mathcal{B}_2)$.

L'ACPCL₁ de (X_t) sous la contrainte de \mathcal{B}_1 -mesurabilité revient à faire l'ACP de $E^{\mathcal{B}_1}(X_t)$ ou $G_1 = L^2(\mathcal{B}_1)$ et $D_1 = E^{\mathcal{B}_1}$ l'espérance conditionnelle.

De même L'ACPCL₁ de (y_t) sous la contrainte de \mathcal{B}_2 -mesurabilité revient à faire l'ACP de $E^{\mathcal{B}_2}(X_t)$ ou $F_1 = L^2(\mathcal{B}_2)$ et $G_1 = E^{\mathcal{B}_2}$ l'espérance conditionnelle.

Si on suppose de $f |X_t|^4 d\mu < \infty$ et $f |Y_t|^4 d\mu < \infty$ alors $U_{x \otimes y} \in L^2(T)$ et

$$[U_{x \otimes y}(h_1 \otimes h_2)](t) = (U_x h_1)(t) \cdot (U_y h_2)(t) = [(U_x h_1) \overset{\mathcal{I}}{\otimes} (U_y h_2)](t) \quad \forall (h_1, h_2) \in H \times H.$$

Et donc l'ACPCL₁ de $U_{x \otimes y}^*$ sous la contrainte de \mathcal{B}_1 -mesurabilité de X_t et \mathcal{B}_2 -mesurabilité

de Y_t revient à l'ACP de $(E^{\mathcal{B}_1} \overset{\mathcal{I}}{\otimes} E^{\mathcal{B}_2}) \circ U_{x \otimes y}^*$

Le schémat de dualité associé :

$$\begin{array}{ccccc} H_2 \otimes H_2 & \xleftarrow{U_{x \otimes y}} & H_1 \otimes H_1 & \xleftarrow{E^{\mathcal{B}_1 \otimes \mathcal{C}}, E^{\mathcal{B}_2}} & H_1 \otimes H_1 \\ \uparrow I & & \downarrow W & & \downarrow W_{1 \otimes 2} \\ H_2 \otimes H_2 & \xrightarrow{U_{x \otimes y}^*} & H_1 \otimes H_1 & \xrightarrow{E^{\mathcal{B}_1 \otimes \mathcal{C}}, E^{\mathcal{B}_2}} & H_1 \otimes H_1 \end{array}$$

Références

- [1] BENCHIKH T., " ACP d'une suite de v.a. sous contrainte de B-mesurabilité " . Deuxièmes Journées de Statistique Fonctionnelle et Opératoirelle. 12-13 Juin 2003.
- [2] BOUDOU A., ROMAIN Y., "Processus hilbertien associé à la convolée de deux mesures aléatoires". C.R. Acad. Sci. Paris, t. 332. Série I, p. 361-364. 2001
- [3] DAUXOIS J. & POUSSE A. , (1976) , "Les analyses factorielles en calcul des probabilité et en statistique : essai d'étude synthétique " . Thèse es-sciences , Toulouse.
- [4] DAUXOIS J., ROMAIN Y., VIGUIER S., " Comparison of Two Factor Subspaces", Journal of Multivariate Analysis, Vol 44, No. 1 (1993), p. 161-178
- [5] DURAND J., "Generalized principal component analysis with respect to instrumental variables via univariate spline transformations". Computational statistics and data analysis no. 16 (1993), p. 423-440. North-Holland.
- [6] ROMAIN Y. (2000), " Etude des somme, différence et rproduit tensoriels fonctionnels de deux ndomorphismes". Publi. Labo. Stat. Probl., 2-00, 1-65. Univ. P. Sabatier, Toulouse.
- [7] SMADHI S. & YOUSFATE A., (1992), "Extension de l'ACPCL à une variable hilbertienne", 3rd islamic countries conference on statistical sciences approaching, Rabat.

F.MEZOUED, M.EL BAH I

Faculté des sciences mathématiques
Département de statistiques

Résumé

La transformée en ondelettes orthogonales (notée O.O) peut être considérée comme un outil pour le débruitage en traitement de signal. Des méthodes non linéaires proposées par Donoho et Jhonstone (1994) consistent à estimer une fonction, à partir des données bruitées, par seuillage des coefficients d'ondelettes empiriques. Contrairement aux méthodes linéaires, ces méthodes ont de bonnes propriétés minimax pour certains espaces fonctionnels (Espace de Besov).

La première partie de cette étude porte sur un signal synthétique perturbé par un processus stationnaire.

Une seconde partie consiste à détecter les singularités et leurs niveaux de régularité en utilisant le formalisme de la transformée en ondelettes continue (Mallat 1992). Elle permet la localisation des singularités et la caractérisation des structures fines contenues dans le signal. Ces méthodes sont testées sur un signal synthétique dans un premier temps, et utilisées pour l'analyse du champs magnétique terrestre après débruitage.

Des résultats satisfaisants ont été obtenus dans ce sens et ont permis de mettre en évidence certains orages magnétiques liés à l'activité solaire

-1- Introduction

En raison du vent solaire provenant du soleil, la terre est frappée par un plasma magnétique, supersonique portant une grande quantité d'énergie cinétique et électromagnétique créant l'activité géomagnétique composée d'orages et de sous orages. Ces derniers se définissent par les variations irrégulières observées dans les enregistrements du champ magnétique terrestre.

Dans la présente étude nous nous intéressons particulièrement à la détection de ces événements.

Les données expérimentales du champs magnétique terrestre sont fournies par le programme Intermagnet avec une résolution de 0.1 nT, celles-ci imposent une modélisation comportant un aléa additif assimilé à un bruit c.à.d

$$X(t_i) = y(t_i) + \varepsilon(t_i) \dots \quad i = 1 \dots n \quad (1)$$

où $X = (X(t_i))_i$ représente les données observées
 $Y = (y(t_i))_i$ représente les données sans bruit
 $\varepsilon = (\varepsilon(t_i))_i$ représente le bruit contenu dans les données observées

Le problème principal qui se pose est que nous ne connaissons ni la fonction $y(t)$ ni la localisation de ses singularités ni le niveau de régularité en ces points. Nous devons estimer la fonction $y(t)$ dans un premier temps et localiser ses singularités et leurs niveaux de régularité dans un deuxième temps. Des méthodes non linéaires ont été proposées par L.DONOH O et M.JOHNSTONE basées sur la décomposition en O.O. Plus précisément nous ferons appel à la méthode du seuillage. Contrairement aux méthodes linéaires, les estimateurs qui en résultent sont optimaux dans le sens suivant :

- L'estimateur $\hat{y}(t)$ est aussi régulier que $y(t)$ c.à.d
 $\exists (\Pi_n) \Pi_n \rightarrow 1$ quand $n \rightarrow +\infty, \exists c \in \mathbb{R}$ telles que $P(\|\hat{y}(t)\|_F \leq c, \|y(t)\|_F \geq \Pi_n) \geq \Pi_n$
 F est une classe fonctionnel de Besov
- L'estimateur $\hat{y}(t)$ est presque minimax pour de nombreux espaces fonctionnels aussi large que l'espace de Besov auquel appartient probablement les données non bruitées.

-2- Réduction du bruit par seuillage des coefficients d'ondelettes :

L'application de la transformée en ondelettes discrète aux données $X(t_i) i=1, n$ maintient le modèle(1) inchangé :

$$W_i = \theta_i + z_i \quad i \in I = \{(j,k) 1 \leq j \leq m \text{ et } 1 \leq k \leq 2^{j-1}\} \dots \quad (2)$$

L.DONOH O et M.JONSTONE ont établi le résultat suivant.

Soit le modèle (2) avec $z_i i \in I$ un bruit blanc gaussien on suppose que $\theta_i i \in I$ est une variable aléatoire de mesure a priori $\mathbb{P}_i \in M_{p,q}^\alpha = \{\mathbb{P} / \|\tau\|_{p,q}^\alpha \leq c\}$ où $\tau = (\tau_i)_i, \tau_i = (E_{\pi_i} |\theta_i|^{p \wedge q})^{1/p \wedge q}$

L'estimateur $\theta^* = (\theta_i^*)_i, i \in I$ qui est de la forme $\theta_i^* = \delta^*(w_i)$ ou δ^* est une fonction non linéaire réalise le risque minimax bayésien B^* défini par : $B^*(n, \Theta_{p,q}^\alpha) = \inf_{\theta^*} \sup_{\theta \in \Theta_{p,q}^\alpha} E_{\pi} E_{\theta} \|\theta^* - \theta\|^2$

De plus, l'utilisation de la fonction non linéaire « seuillage doux » : $\delta_\lambda(w) = \text{sigw}(|w| - \lambda)_+$ avec un seuil λ
 ou bien de la fonction non linéaire « seuillage dur » : $\delta_\mu(w) = w I_{\{|w| \geq \mu\}}$ avec un seuil μ

fournit un estimateur bayésien presque minimax ce qui est exprimé par le théorème suivant :

théorème -1- Considérons le modèle (2) Il existe deux constantes $N(p \wedge q)$ et $M(p \wedge q)$ telles que $B_{\lambda}^*(n, \Theta_{p,q}^\alpha) \leq N(p \wedge q) B^*(n, \Theta_{p,q}^\alpha)$ et $B_{\mu}^*(n, \Theta_{p,q}^\alpha) \leq M(p \wedge q) B^*(n, \Theta_{p,q}^\alpha)$

- a partir des évaluations numériques, il a été constaté que les constantes $N(p \wedge q)$ et $M(p \wedge q)$ sont inférieures a 2.22 et que la constante $N(p \wedge q)$ tends à être plus petite que la constante $M(p \wedge q)$ [re]

- $\theta^* = (\theta_i^*)_i, \theta_i^* = \delta_{\lambda}^*(w_i) i \in I$ avec $\hat{Y} = W^{-1}(\theta^*)$ W^{-1} étant la T.O inverse

2-1-Seuil universel :

Le seuil optimal λ^* est donnée par la valeur du seuil qui minimise le risque $r(\theta^*, \theta) = E(\|\theta^* - \theta\|^2)$

Assyptotiquement, $\lambda^* \approx \sigma (2 \log n)^{1/2}$ σ l'écart type du bruit
 ce seuil est appelé seuil universel

- DONOH O et JHONSTONE ont montré que l'utilisation du seuil universel fournit un estimateur θ^* dont le risque est a un facteur logarithmique près du risque Benchmark (relatif a la sélection idéale des coefficients) ce qui est énoncé par le théorème suivant :

Théorème -2- Soit le modèle (2) l'estimateur $\hat{\theta}^* = \delta(w, \sigma (2 \log n)^{1/2})$ satisfait $r(\hat{\theta}^*, \theta) \leq (2 \log n + 1)(\sigma^2 + \sum_{i=1}^{i=n} (\theta_i^2 \wedge \sigma^2))$

• le théorème suivant affirme que l'estimateur $\hat{\theta}^*$ obtenu atteint asymptotiquement le meilleur risque relativement au risque benchmark

Théorème -3- Soit le modèle (2) l'estimateur $\hat{\theta}^* = \delta(w, \sigma (2 \log n)^{1/2})$ satisfait :

$$\inf_{\theta^*} \sup_{\theta} E(\|\hat{\theta}^* - \theta\|^2) / (2 \log n + 1)(\sigma^2 + \sum_{i=1}^{i=n} (\theta_i^2 \wedge \sigma^2)) \rightarrow 1 \quad \text{qd } n \rightarrow +\infty$$

• l'estimateur $\hat{\theta}^*$ est d'un risque qui est a un facteur logarithmique près du risque minimax .en effet,

Théorème -4- Soit $B_{p,q}^a(c)$ une boule de Besov, l'estimateur \hat{y}_n^u est presque minimax c.a.d.

$$\sup_{\theta} E(\|\hat{y}_n^u - y\|^2) \leq k \log n \inf \sup_y E(\|\hat{y} - y\|^2) \quad \text{avec } k \in \mathbb{R} \text{ et } y \in B_{p,q}^a(c)$$

2-2-Seuil SURE

C.STEIN a proposé un seuil λ^s dit « seuil sure » basé sur la minimisation de l'estimateur sans biais de l'EQM :

$$\text{SURE}(\lambda, w) = 1/n (\|\hat{\theta}_{\lambda^s}^* - w\|^2 + \sigma^2 (n - 2n_0)) \quad \text{ou} \quad n_0 = \text{card} \{ \theta_i / |\theta_i| < \lambda \}$$

l'estimateur $\hat{\theta}^{*s} = \delta(w, \lambda^s)$ est à une constante près du risque minimax :

$$\sup_{\theta} E(\|\hat{\theta}^{*s} - \theta\|^2) \leq c \cdot \inf_{\theta^*} \sup_{\theta} E(\|\hat{\theta}^* - \theta\|^2) \quad \text{avec } c \in \mathbb{R} \text{ et } \theta \in \Theta_{p,q}(c)$$

2-3-Extension au cas d'un bruit corrélé :

W.SILVERMAN et M.JOHNSTONE ont étendu la procédure d'élimination du bruit au cas d'un bruit stationnaire corrélé (AR(1)) en effectuant un seuillage dépendant d'échelle . l'estimateur obtenu est :

$$\hat{\theta}^s = (\hat{\theta}_{jk}^*)_{jk} \quad \hat{\theta}_{jk}^* = \delta^s(w_{jk}, \lambda_j) \quad \lambda_j \text{ étant le seuil appliqué au } j\text{ième niveau, avec } (j,k) \in I$$

• le théorème ci-dessous nous énonce que le risque de l'estimateur \hat{y}_n^s est équivalent au risque minimax

théorème -5- Soit le modèle (1) et soit Ψ une ondelette a r moment nuls et r fois continûment différentiable alors

$$\sup_y E(\|\hat{y}_n^s - y\|^2) \approx \inf_y \sup_y E(\|\hat{y} - y\|^2) \quad , \quad y \in B_{p,q}^a(c) \quad \text{pour tout } p,q \in [1, +\infty[, c \in [0, +\infty[\text{ et } \alpha < r$$

-3- Analyse multi échelle pour la détection des singularités et spécifications des régularités .

la transformée en ondelettes continue W permet d'étudier la régularité d'une fonction f(x) .

MALLAT et HWANG ont montré qu'il ne peut y avoir de singularités sans qu'il y'ait de maxima du module de la transformée en ondelettes, ce qui est énoncé par le théorème suivant :

Théorème -6- Soit n un entier strictement positif, et soit $\Psi(t)$ une ondelette a n moment nuls, a support compact et n fois continûment différentiable alors

Soit f(t) $\in L^1([b, b'])$

• S'il existe une échelle a_0 telle que pour toute échelle $a < a_0$ et $t \in]b, b'[$, $|Wf(a,t)|$ n'admet pas de maximum local , alors pour tout $\varepsilon > 0$ et $\alpha < n$, f(t) est uniformément n-holderienne sur $]b + \varepsilon, b - \varepsilon[$

• si $\Psi(t)$ est la n^{ième} dérivée d'une fonction régulière alors f(t) est uniformément n-holderienne sur $]b - \varepsilon, b' + \varepsilon[$

• Le théorème suivant caractérise la régularité locale d'une singularité et qui est donné par le maximum des pentes des lignes maxima du module de la transformée en ondelettes continue.

Théorème -7- Soit f(t) une distribution tempérée telle que sa transformée en ondelettes continue soit bien définie sur $]b, b'[$ et soit $t_0 \in]b, b'[$

S'il existe une échelle a_0 et une constante c telles que pour $t \in]b, b'[$ [et $a < a_0$, tous les maxima du module de $Wf(a, t)$ appartiennent au cône défini par :

$$|t - t_0| \leq c \cdot a \quad \dots (3)$$

Alors, pour tout point $t_1 \in]b, b'[$ tel que $t_1 \neq t_0$, f(t) est uniformément n-holderienne au voisinage de t_1

soit $\alpha < n$ non entier ,la fonction f(t) est α -holderienne en t_0 si et seulement s'il existe une constante A telle qu'en chaque maxima (a,t) du module de $Wf(a,t)$ a l'intérieur du cône (3) :

$$|Wf(a,t)| \leq A \cdot a^\alpha \quad \dots (4)$$

la propriété (4) est équivalente à : $\log |Wf(a,t)| \leq \alpha \log a + \log A \quad \dots (5)$

cette dernière propriété (5) nous permet d'estimer la régularité α de la fonction f(t) au point t_0 en prenant la pente maximum des lignes droites dans un diagramme log-log

-4-Etude sur un signal synthétique

Le signal synthétique $S(t) = \sum_{i=1}^{i=44} R_i(t-t_i)$ est gènère par la fonction

$$R(t-t_i) = a_i(t-t_i)_+^b \exp\{-c_i(t-t_i)\} \quad \text{où } a_i, b_i \in \mathbb{R} \text{ et } c_i \in \mathbb{R}_+^* \quad i=1, \dots, 44$$

Notons que $R(t-t_i)$ est de régularité b_i en t_i

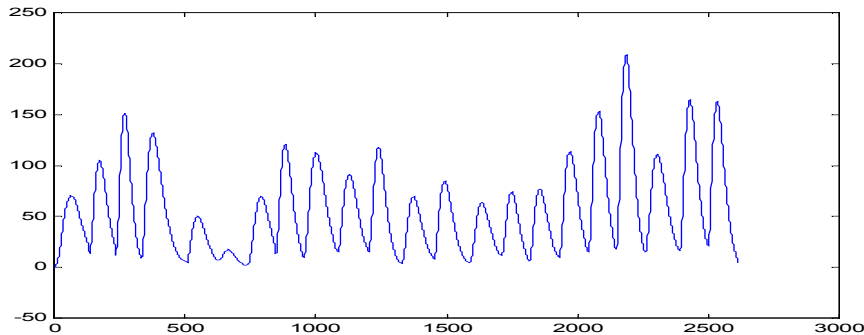


Figure 4-1 : signal synthétique

Une première étude consiste à appliquer les méthodes de réduction de bruit au signal synthétique a lequel nous additionnons différents types de bruit (bb ,AR(1)) et par la suite détecter les singularités présentes dans le signal et spécifier le type d'irrégularité correspondantes .

4-1- Réduction du bruit par seuillage des coefficients d'ondelettes en utilisant les seuils « sure » et « universel » : Les résultats du seuillage (global ou dépendant d'échelles) sont illustrés dans les figures suivantes :

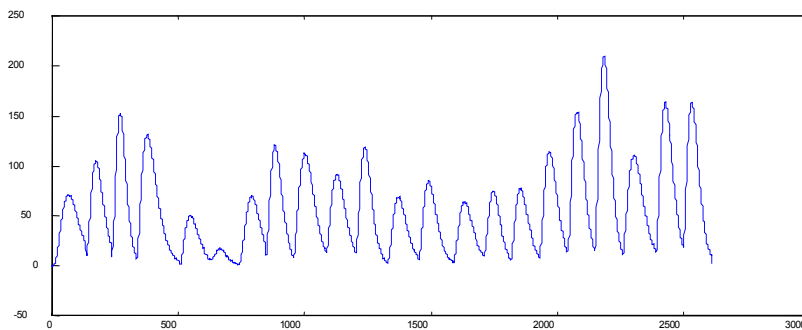


Figure 4.2: Synthétique +bb débruité (sure, global)

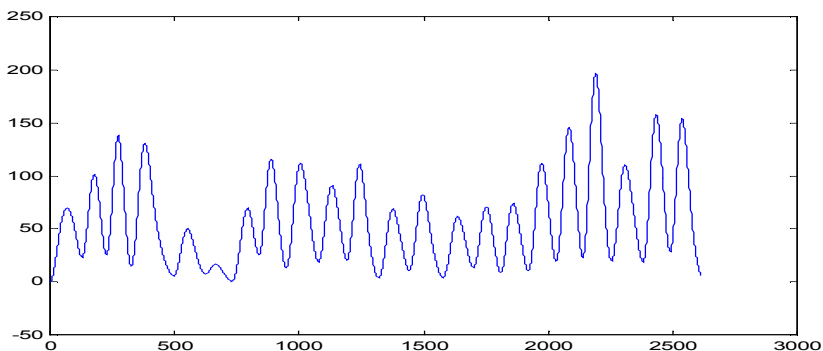


Figure 4.3: Synthétique +bb débruité (universel, global)

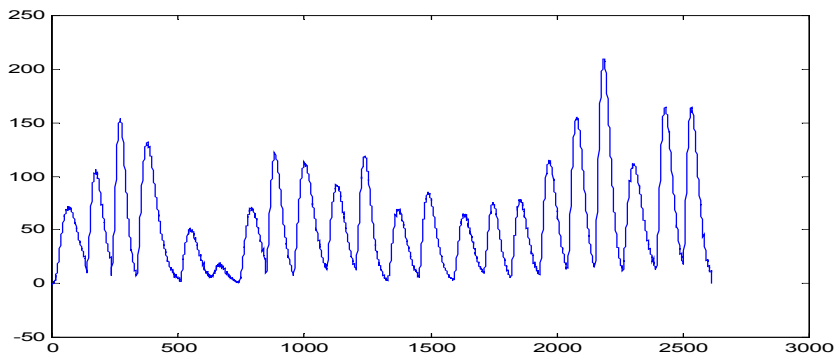


Figure 4.4 :Synthétique+ar(1) ($\varphi=0.2$) après seuillage (sure dépendant d'échelle)

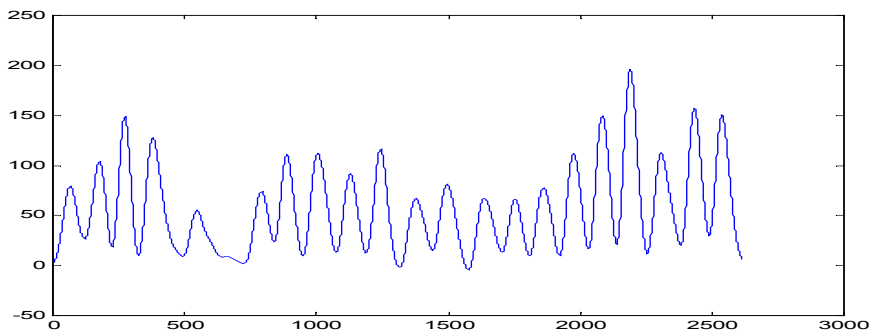


Figure 4.5:Synthétique+ar(1) ($\varphi=0.2$) après seuillage (universel dépendant d'échelle)

4-1-1 Evaluation qualitative :

Dans le cas d'un bruit blanc, nous remarquons qu'une grande partie du bruit a été éliminée (figure 4.3 ,4.4) que ça soit par l'utilisation d'un seuillage sure ou universel. de plus ,l'apparition de certaines fluctuations qui sont dues au bruit dans le résultat du seuillage sure

Quant au cas d'un bruit corrélé tel que AR (1) avec $\varphi=0.2$, un seuillage global n'effectue pas un bon seuillage car on a la présence d'un remarquable taux de bruit et le passage a un seuillage dépendant d'échelle nous permet d'avoir une net amélioration des résultats.

4-1-2 Evaluation qualitative :

L'évaluation quantitative se fait par le biais de l'erreur en moyenne quadratique entre le signal original et le signal obtenu après seuillage et dont les valeurs sont synthétisées dans les tableaux ci-dessous (tab 4-1 et tab4-2)

| paramètre | $\sigma = 1$ | | $\sigma = 2$ | | $\sigma = 4$ | | $\sigma = 6$ | | $\sigma = 8$ | |
|-----------|--------------|------------------|--------------|------------------|--------------|------------------|--------------|------------------|--------------|------------------|
| seuillage | Sure global | Universel global | Sure global | Universel global | Sure global | Universel global | Sure global | Universel global | Sure global | Universel global |
| EQM | 0.082 | 0.194 | 0.095 | 0.210 | 0.099 | 0.287 | 1.04 | 1.87 | 2.11 | 2.64 |

Tab 4-1 : les erreurs en moyenne quadratique du signal synthétique bruité (bruit blanc) après seuillage

| modèle | Synthétique + AR(1) | | | | | | | | | | | |
|-----------|---------------------|-------------|--------------------|--------------------|-----------------|--------------|--------------------|--------------------|-----------------|--------------|--------------------|--------------------|
| paramètre | $\varphi = 0.2$ | | | | $\varphi = 0.5$ | | | | $\varphi = 0.8$ | | | |
| seuillage | Sure global | Univ global | Sure dep d'échelle | Univ dep d'échelle | Sure global | Univ. global | Sure dep d'échelle | Univ dep d'échelle | Sure global | Univ. global | Sure dep d'échelle | Univ dep d'échelle |
| EQM | 8.53 | 9.56 | 5.81 | 5.29 | 19.37 | 20.04 | 17.85 | 18.26 | 32.69 | 33.26 | 30.73 | 31.05 |

| Modèle | Synthétique + AR(1) | | | | | | | |
|-----------|---------------------|-------------|--------------------|--------------------|---------------|--------------|--------------------|--------------------|
| paramètre | $\varphi = 0.95$ | | | | $\varphi = 1$ | | | |
| seuillage | Sure global | Univ global | Sure dep d'échelle | Univ dep d'échelle | Sure global | Univ. global | Sure dep d'échelle | Univ dep d'échelle |
| EQM | 65.48 | 66.10 | 61.13 | 62.92 | 350.64 | 351.24 | 349.90 | 350.54 |

Tab4-2 : les erreurs en moyenne quadratique du signal synthétique bruité (bruit ar (1))après seuillage

Une remarque importante à faire, est que les erreurs obtenues après utilisation d'un seuillage sure sont inférieures a celles obtenus après utilisation d'un seuillage universel.

Dans le cas d'un bruit blanc les erreurs sont très petites et augmentent légèrement avec le paramètre σ .dans le cas du bruit AR (1) , nous constatons que les erreurs obtenues après un seuillage dépendant d'échelle sont inférieures a celles obtenues après un seuillage global (tab4-2) et que celles-ci sont relativement petites lorsque le paramètre φ est petit et atteignent de grandes valeurs lorsque le paramètre φ s'approche de la valeur 1 .

4-2 détections des singularités par le biais de la TOC

4-2-1 analyse en ondelette du cycle synthétique (sans bruit)

Nous étudions donc, la transformée en ondelettes continue du dernier cycle du signal synthétique afin de caractériser la singularité en $t = 2528$ qui correspond au maximum du cycle et dont la régularité b est égale à $\frac{3}{4}$.les lignes maxima qui lui sont associées sont illustrées dans la figure (4.6) suivante :

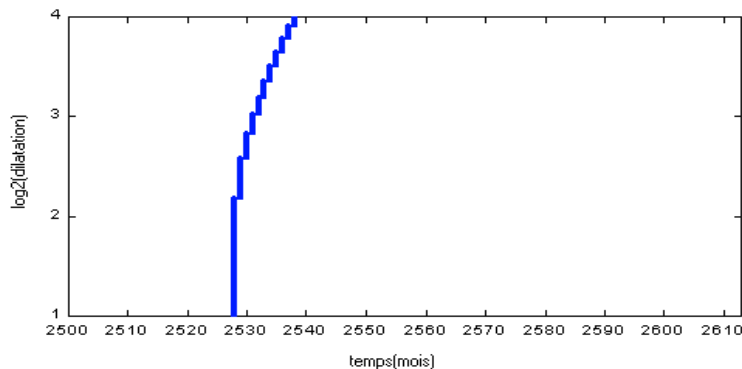


Figure 4.6 : les lignes maxima d'un cycle synthétique

L'étude étant faite sans effet de bord, nous remarquons la présence d'une seule ligne maxima qui atteint une dilatation 2^4 et qui converge vers le point $t = 2528$ et qui correspond à l'occurrence du maximum du cycle. la ligne de pente ou fonction ridge 'figure (4-7) est calculée afin de déterminer la régularité de la singularité correspondante et dont le graphe et le suivant :

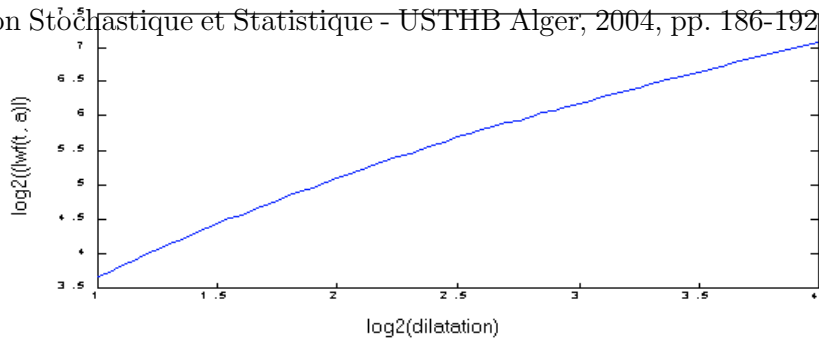


Figure 4.7 : la fonction ridge de la ligne maxima du cycle synthétique ($\alpha = 1.33$)

L'ajustement de la fonction ridge par une ligne droite donne une valeur de pente égale à 1.33 qui coïncide avec la régularité de singularité en $t = 2528$.

4-2-2 analyse en ondelette du cycle synthétique bruité après seuillage

dans le cas du signal synthétique bruité (bb,ar(1)) l'étude sera faite après un seuillage sure ,et dépendant d'échelle dans le cas du bruit ar(1).

Les lignes maxima du cycle pour les différents cas de bruit sont présentées dans la figure ci-dessous :

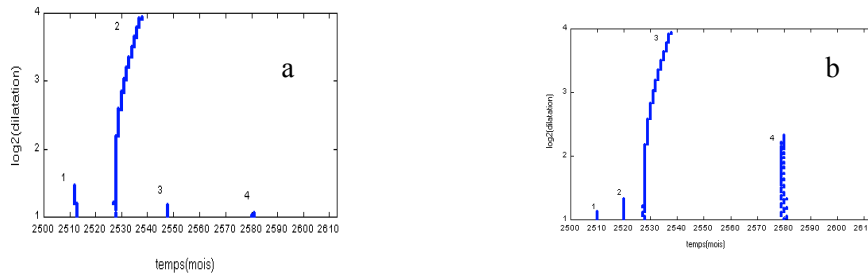


Figure 4-8: les lignes maxima du cycle synthétique après seuillage :

- a- cycle synthétiques bruité avec un bruit blanc
- b- cycle synthétiques bruité avec un bruit ar(1)

Nous constatons l'apparition de certaines lignes générées par le bruit qui ne dépasse pas la dilatation 2^3 , ainsi que la présence de la ligne qui est due au cycle lui-même et qui continue à dominer dans les différents cas de bruits (figure 4-8). En effet, dans le cas d'un bruit blanc, nous remarquons la présence de trois lignes au total qui convergent vers les points $t_1=2528$, $t_2 = 2564$, $t_3 = 2581$, quant au cas du bruit ar(1) nous avons la présence de sept lignes qui convergent vers les instants $t_1=2513$, $t_2 = 2520$, $t_3 = 2528$, $t_4=2544$, $t_5 = 2576$, $t_6 = 2593$ et $t_7 = 2600$.

Quant aux fonctions ridges relatives aux lignes maxima, certaines d'entre elles n'admettent pas un ajustement par une droite et d'autres donnent des valeurs de pentes négatives. En effet, les trois lignes obtenues dans le cas d'un bruit blanc sont de pentes respectives de 1.35, -1.29, -0.65 (figure 4-9). Dans le cas du bruit ar(1), la ligne correspondante au maximum du cycle est de pente égale à 1.37, les autres sont de pentes comprises entre -1.87 et 0.03 (figure 4-10).

Nous constatons que les régularités correspondantes à la singularité en $t = 2528$ obtenues dans les différents cas du bruit sont de valeurs proches et dont les fonctions ridges sont principalement dominées par la partie déterministe du signal.

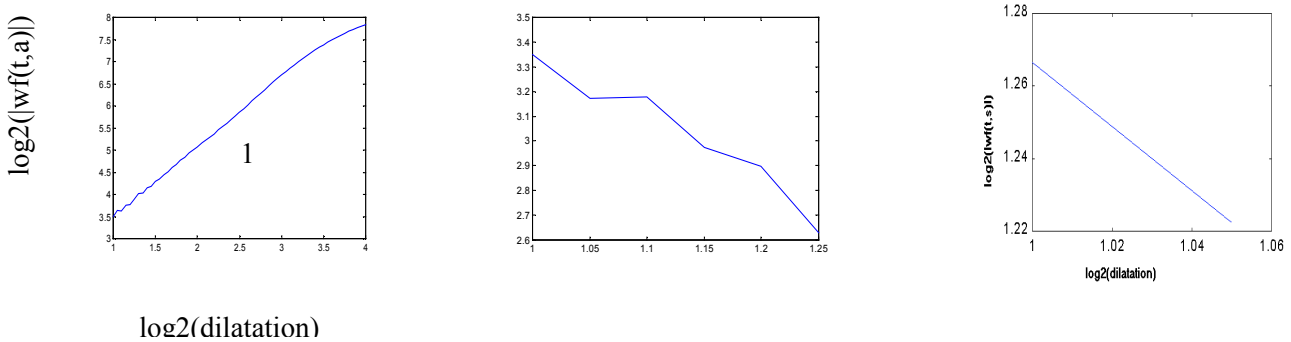


Figure 4-9: les fonctions ridges du cycle synthétique bruité(bb) après seuillage

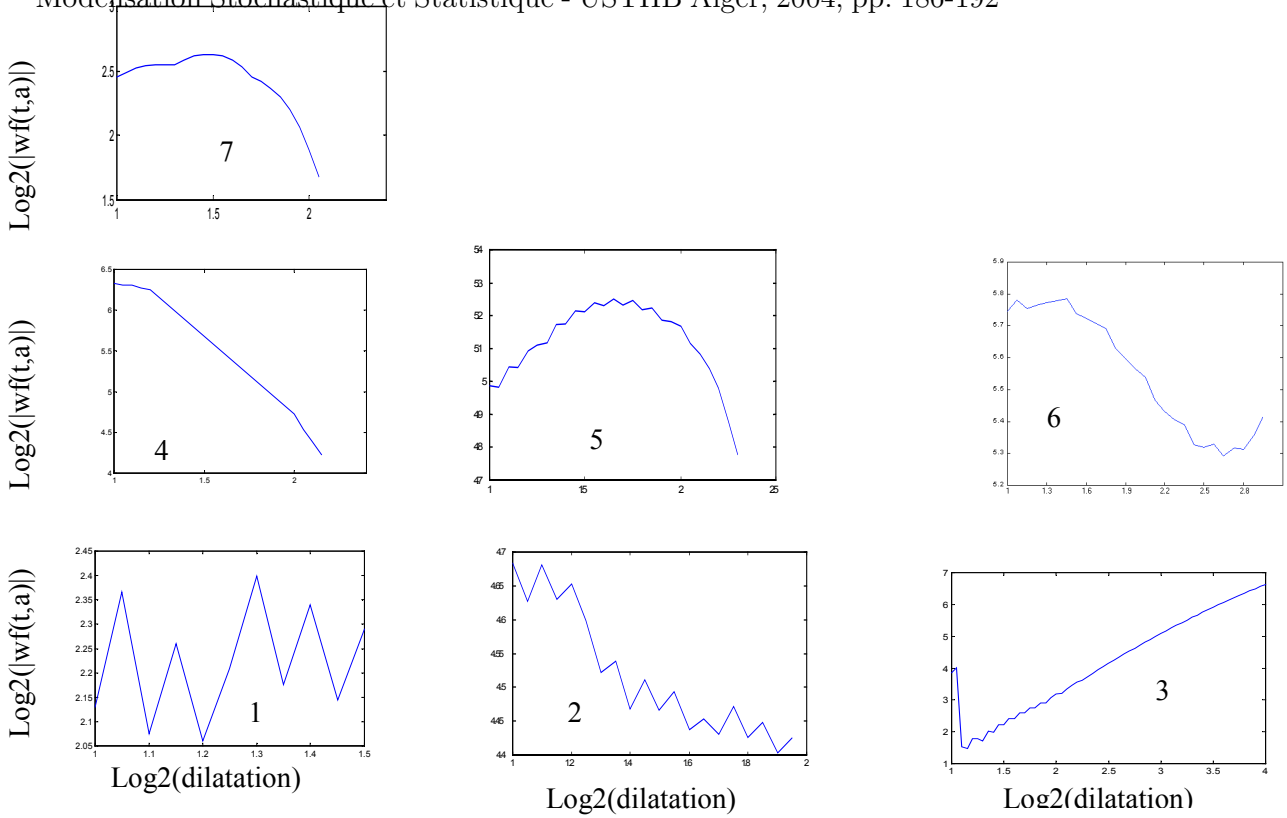


Figure 4-10 : les fonctions ridges du cycle synthétique bruité(ar(1))après seuillage

5- Application au champ magnétique terrestre de l'observatoire CHAMBON-LA- FORET

Dans ce qui suit , nous nous intéresserons à l'étude des valeurs en moyenne journalières de l'observatoire chambon-la-fôret (France). La détection des singularités via le module de la transformée en ondelettes continue sera faite sur la composante z des données magnétiques de l'année 1999 (figure 5-1) après avoir effectué un seuillage sure dépendant d'échelles (figure5-2)

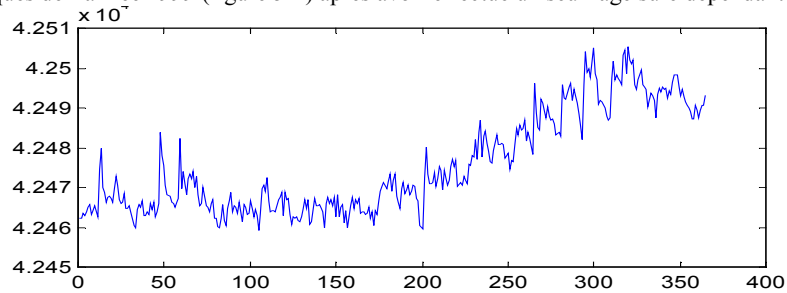


Figure 5.1 : la composante z du champ magnétique de l'observatoire CLF (année 1999)

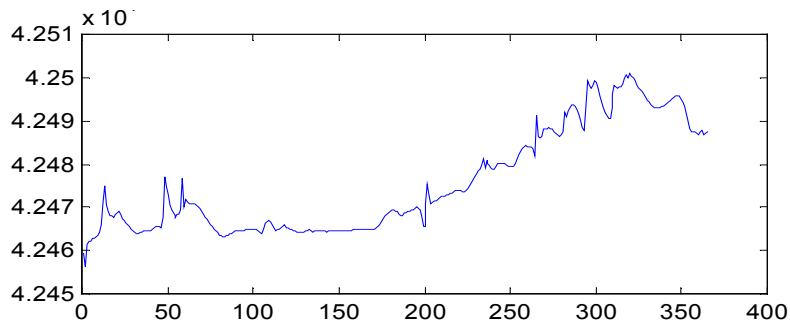


Figure5.2 : la composante z du champ magnétique après seuillage (sure dépendant d'échelle)

Les lignes maxima qui lui sont associées sont présentées dans la figure suivante :

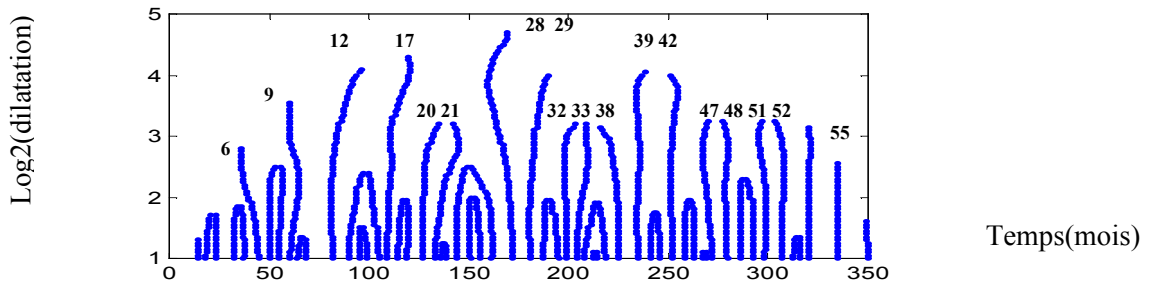


Figure 5.3 : Lignes maxima de la composante z du champ magnétique de Chambon-la-forêt (1999)

Nous ne prendrons en considération que les lignes maxima les plus dominantes pour lesquelles la dilatation $a \geq 3$. Ces 18 lignes convergent vers les singularités du signal considéré. Afin de mettre en évidence les singularités du champ géomagnétique qui sont en rapport avec l'activité solaire, nous calculerons les étendues des différents intervalles qui séparent les lignes maxima et qui sont synthétisées dans le tableau suivant :

| | | | | | | | | | | | | | |
|----------------------|----|----|----|----|----|----|----|----|----|----|----|----|----|
| N de la ligne | 9 | 12 | 17 | 21 | 27 | 29 | 33 | 38 | 42 | 47 | 48 | 52 | 55 |
| Etendue | 27 | 27 | 26 | 29 | 20 | 23 | 21 | 27 | 27 | 27 | 28 | 28 | |

D'après les résultats obtenus ci-dessous, on remarque que mise à part les étendues de 20, 23 et 21 jours les autres sont en moyenne d'étendue égale à 27.37 jours. Par conséquent les singularités qui leur sont associées sont certainement dues au cycle de 27 jours qui est lié à la rotation du soleil autour de son axe.

-5- conclusion

Le travail décrit dans cet exposé nous a permis en premier lieu de déterminer un estimateur d'une fonction sans connaissance a priori sur cette fonction. Ceci nous a induit à révéler les avantages des méthodes de seuillage basées sur la décomposition en ondelettes orthogonales tant sur le plan théorique que sur le plan pratique. En effet, nous avons constaté que l'estimateur obtenu à partir des données bruitées par l'utilisation d'un seuil sûr est d'un risque qui est à une constante près du risque minimax et qu'il lui est asymptotiquement équivalent dans le cas d'un seuillage dépendant d'échelle. Contrairement au seuil universel dont le risque de l'estimateur qui est facteur de $\log n$ près du risque minimax.

À partir de l'étude qui a été faite sur le signal synthétique, nous avons constaté que l'utilisation d'un seuillage global donne de bons résultats dans le cas d'un bruit blanc gaussien, contrairement au cas d'un bruit corrélé tel que AR(1), l'apport d'un seuillage dépendant d'échelle est positif car il permet d'améliorer les résultats en particulier dans le cas d'un bruit stationnaire. En comparant les erreurs calculées à partir de l'utilisation des seuils sûrs et universels, nous concluons que les méthodes basées sur le seuil sûr sont optimales. Par conséquent, nous avons choisi la procédure sûre (dépendant d'échelle) comme méthode de réduction de bruit dans les données réelles du champ géomagnétique.

Quant à l'étude qui consiste à détecter les singularités, nous avons constaté que le formalisme de la T.O.C ne fait aucune hypothèse a priori sur l'instant t_0 de l'occurrence de l'évènement. Cette analyse a donné des résultats encourageants sur la caractérisation des fines structures contenues dans le signal considéré. Nous avons remarqué que les lignes maxima les plus dominantes contiennent plus d'information sur le signal étudié. Concernant l'analyse des données du champ magnétique terrestre de l'observatoire Chambon-la-forêt, le seuillage nous a permis une meilleure compréhension des phénomènes physiques que contient le champ géomagnétique, l'étude étant faite sur une année, nous avons pu mettre certaines singularités liées à l'activité solaire, à savoir celles qui sont dues à la rotation du soleil autour de son axe et dont la période est de 27 jours.

-6- bibliographie

- 1/ F.MEZOUED, « analyse multi échelle pour la détection des singularités . application aux champs magnétique terrestre
- 2/ Y.MEYER, « les ondelettes : algorithmes et applications », édition ARMAND COLIN, 1992
- 3/ D.HATHAWAY,R.M.WILSON, «a synthesis of solar prediction techniques».journal of geophysical research.vol104 n^o = A10.october 1999
- 4/ D.L.DONOHU et I.M. JHONSTONE ,”ideal adaptation by wavelet shrinkage”.departement of statistics, Stanford university,ca.94305-4065,usa
- 5/ D.L.DONOHU et I.M. JHONSTONE, «adapting to unknown smoothness via wavelet shrinkage» departement of statistics, Stanford University. July 1994
- 6/ R.R.COIFMAN and D.L.DONOHU, ” translation-invariant de-noising” Yale university and Stanford university.
- 7/ D.L.DONOHU et I.M. JHONSTONE, “wavelet shrinkage: asymptotia “departement of statistics, Stanford university. July 1995
- 8/ D.L.DONOHU et I.M. JHONSTONE, “minimax estimation via wavelet shrinkage “departement of statistics, Stanford University. July 1995
- 9/ I.M.JHONSTONE et D.L.DONOHU, “wavelet threshold estimations for data with correlated noise” departement of statistics, Stanford University. July 1994
- 10/ D.L.DONOHU , “denoising by soft-thresholding “IEE transactions on information theory.vol41.n^o= 3.1995
- 11/ D.L.DONOHU et I.M. JHONSTONE, “minimax risk over L_p -balls for L_q errors “departement of statistics, Stanford University, ca 94305. july 20 , 1994.
- 12/ S.MALLAT et W. I. HWANG, ”singularity detection and processing with wavelets” . IEE Tran. on information theory, vol.38 .n^o= 2,march 1992

V – Statistique Computationnelle et Simulation

Choix de la largeur de fenêtre dans l'estimation de la densité spectrale d'un processus stationnaire et de l'opérateur de régression des données fonctionnelles.

M. Rachdi

Université de Grenoble, IMAG-LMC UMR CNRS 5523 and LabSAD EA 3698, UFR SHS-BSHM, BP. 47, F38040 Grenoble Cedex 09, France. E-mail: mrachdi@upmf-grenoble.fr

L'objet de cette conférence rentre dans le domaine de la statistique non paramétrique, ainsi que dans le domaine de la statistique des processus. Nous nous intéressons au choix du paramètre de lissage utilisé lors de l'estimation de type noyau d'une fonctionnelle, en particulier dans l'estimation de la densité spectrale d'un processus stationnaires et dans l'estimation de l'opérateur de régression quand la variable explicative est fonctionnelle.

Estimation de la densité spectrale d'un processus stationnaire

L'analyse spectrale des séries temporelles occupe une place importante et originale, tant dans le domaine théorique que dans les applications scientifiques.

Sa structure de base réside en une dépendance des phénomènes étudiés vis-à-vis du temps, concept essentiel dans ces deux domaines. Il y a peu de disciplines qui ne soient confrontées à l'étude de phénomènes évoluant dans le temps, que l'on désire décrire, analyser, interpréter, contrôler ou prévoir.

La physique (ingénierie électrique), l'acoustique, la géophysique, la médecine, la biologie, l'économie, la météorologie constituent quelques exemples de ces disciplines.

Habituellement, lorsque l'on cherche à analyser un ensemble d'observations $\{X_t, t = 0, \dots, n-1\}$, c'est à dire un vecteur aléatoire, on essaie de décomposer celui-ci sur une base bien choisie sous la forme d'une combinaison linéaire d'exponentielles imaginaires

$$X_t = \sum_j \exp(-i\omega_j t) a_j,$$

où les a_j sont des variables aléatoires non corrélées.

Une fois montrée l'existence d'une telle décomposition, il reste à voir comment la déterminer, ou plutôt comment l'estimer à partir des observations et examiner si certaines composantes sont interprétables en terme de tendance et de saisonnalité.

Soit $\{X(t)\}_{t \in \mathbb{T}}$ un processus stationnaire au sens large, harmonisable,

$$X_t = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} \exp(i\omega t) dF(\omega)$$

où $T = \mathbb{Z}, \mathbb{R}$ et

$$F(\omega) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} \exp(-i\omega t) X_t dt$$

est la mesure spectrale.

Pour connaître le comportement de X , on s'intéresse à F . Nous considérons le cas où F est absolument continue par rapport à la mesure de Lebesgue : $dF(\omega) = f(\omega)d\omega$, où f la densité spectrale.

La densité spectrale f est donc la transformée de Fourier de la fonction de covariance. Il est donc naturel de vouloir l'estimer par la transformée de Fourier de la covariance empirique. Il en découle un estimateur, appelé périodogramme, mais celui-ci ne s'avère pas satisfaisant, car il n'est pas consistant en moyenne quadratique. En lissant le périodogramme par un noyau, il est possible d'obtenir des estimateurs consistants en moyenne quadratique. L'inconvénient des estimateurs obtenus par un lissage est qu'ils dépendent d'un paramètre (comme la majorité des estimateurs) qui joue un rôle crucial dans le comportement asymptotique de l'estimateur.

La procédure que nous adoptons est basée sur les idées de validation croisée. Cette méthode a été étudiée et a permis de résoudre le problème que nous soulevons dans l'estimation de la densité de probabilité, de la régression réelle, de la fonction de hasard,

Dans cet exposé, nous allons donc nous intéresser aux aspects théorique et pratique de l'estimation d'une densité spectrale par la méthode du noyau. Dans un premier temps, nous considérons les propriétés asymptotiques de l'estimateur à noyau. Nous montrons que, comme dans la plupart des problèmes d'estimation fonctionnelle, le choix du paramètre de lissage joue un rôle crucial dans le comportement asymptotique de l'estimateur. Dans un second temps, nous abordons, dans le cadre d'un processus à temps discret (avec ou sans hypothèse de mélangeance), le problème de la sélection du paramètre de lissage. Ensuite nous étendons les résultats obtenus au cas des processus à temps continu (toujours avec ou sans hypothèse de mélangeance). Plus précisément, nous proposons (dans chaque cas) un critère de sélection de la largeur de fenêtre basé sur la validation croisée. Des résultats d'optimalité de la fenêtre sélectionnée sont établis. Ces résultats théoriques obtenus concernant le choix de la fenêtre spectrale sont illustrés par des simulations.

Estimation de l'opérateur régression à partir des données fonctionnelles

La statistique sur les données fonctionnelles fait l'objet de quelques récentes études dont les applications couvrent divers domaines (médecine, économétrie, chimiométrie, ...). Un échantillon de données fonctionnelles peut être traité dans plusieurs problèmes statistiques (classification, discrimination, étude longitudinale, prédiction, régression, ...). Donc, aujourd'hui c'est devenu un défi pour les statisticiens de construire des procédures statistiques pour traiter ce type de données. L'auteur peut trouver dans le livre Ce domaine de la statistique moderne est devenu populaire par le bouquin de Ramsay et Sierverman (1997) et l'article de Ferraty et al. (2003) une vue d'ensemble sur les différentes problématiques et les récentes avancées dans ce domaine important de la statistique moderne.

Dans cet exposé, nous nous consacrons à certains problèmes spécifiques faisant intervenir les données fonctionnelles. Il s'agit de la prédiction d'une réponse aléatoire

scalaire Y sachant une variable fonctionnelle X . Autrement dit, la question est d'estimer l'opérateur : $r(\cdot) = E(Y|X = \cdot)$, où X est une variable aléatoire à valeurs dans un espace de dimension infinie. L'estimation de l'opérateur r a été considérée par plusieurs auteurs durant la dernière décennie. La plupart des travaux concernaient les modèles linéaire (r linéaire). Les plus récentes avancées ainsi que plusieurs références sur ce sujet sont données l'article de Cardot et al. (1990). Comme dans le cas non fonctionnel, nous nous confrontons au problème de l'estimation de l'opérateur r , quand ce dernier n'est pas linéaire. Dans le travail de Ferraty et Vieu (2003) une version de lissage par noyau a été utilisée pour construire un estimateur non paramétrique de l'opérateur r . Cet estimateur admet de bonnes propriétés asymptotiques et le problème de dimension est contrôlé sous certaines hypothèses sur la concentration de la loi de X . Afin d'assurer de bonnes performances de l'estimateur sur le plan pratique, le paramètre de lissage doit être bien sélectionné. Le but principal de cette partie de l'exposé est donc de proposer une procédure de choix automatique de ce paramètre, et d'établir son optimalité asymptotique au sens de l'erreur quadratique.

References

- [1] H. Cardot, F. Ferraty and P. Sarda. Linear Functional Model. *Statist. & Prob. Letters*, 45, 11-22, 1999.
- [2] F. Ferraty and P. Vieu. Functional nonparametric statistics: a double infinite dimensional framework. *Recent Advances and Trends in Nonparametric Statistics*, 2003.
- [3] F. Ferraty and P. Vieu. Nonparametric models for functional data, with application in regression, time series prediction and curve discrimination. *Jr. Nonparametric Statistics*, 2004.
- [4] F. Ferraty and P. Vieu. Functional nonparametric statistics in action. *Art of Semi-parametric, Springer Lecture Notes in Statistics*, 2004.
- [5] M. Rachdi. Choix de la largeur de fenêtre spectrale par validation- croisée pour un processus stationnaire à temps continu. *C. R. Acad. Sci., Sér. 1, Math.*, 327(8):777–780, 1998.
- [6] M. Rachdi. Choix optimal de la fenêtre spectrale pour un processus stationnaire à temps discret α -mélangeant. *C. R. Acad. Sci., Sér. 1, Math.*, 327(4):405–408, 1998.
- [7] M. Rachdi et R. Sabre. The optimal choice of the spectral bandwidth for a random field. *Trait. Signal.*, 15(6):569–575, 1999.
- [8] M. Rachdi. Strong consistency with rates of spectral estimation of continuous-time processes: from Periodic and Poisson sampling schemes, *Nonparametric statistics*, ():1-16, 2004.

- [9] F. Charlot and M. Rachdi. On the statistical properties of a stationary process sampled by a stationary point process. *Stat. & Proba. Lett.*, 2004.
- [10] M. Rachdi and P. Vieu. Nonparametric regression for functional data: automatic smoothing parameter selection, *Soumis*, 2004.

Sur le paramètre de lissage dans l'estimation de la densité de probabilité par la méthode du noyau

Smaïl ADJABI

Laboratoire de Modélisation et d'Optimisation des Systèmes

LAMOS, Université de Béjaia, 06000 Béjaia, Algérie

E-mail : adjabi@wissal.dz

Résumé

Ce travail est une contribution sur le problème du choix du paramètre de lissage dans l'estimation de la densité de probabilité par la méthode du noyau. Si le choix du noyau n'est pas un problème dans l'estimation de la densité, il n'en est pas de même pour le choix du paramètre de lissage qui ne dépend que de la taille de l'échantillon. Plusieurs travaux ont montré que les estimateurs peuvent changer dramatiquement pour de petites variations du paramètre de lissage. Actuellement il n'existe pas de choix optimal pour ce paramètre de lissage (le choix optimal qui minimise l'erreur relative globale dépend de la dérivée seconde de la densité inconnue). Dans cette contribution, on propose de combiner premièrement les deux méthodes : celle qui minimise l'erreur relative globale et celle qui l'estime par la méthode du pseudo maximum de vraisemblance et ensuite, celle qui combine l'estimateur du pseudo maximum de vraisemblance et l'approche des plus proches voisins.

Mots clés : Estimation, densité de probabilité, paramètre de lissage, noyau, plus proches voisins, pseudo-maximum de vraisemblance.

Introduction

Soit X_1, X_2, \dots, X_n un n -échantillon issu d'une variable aléatoire X de loi de probabilité f . Quand on a aucun a priori sur la loi de X ou quand les tests du Khi-Deux et de Kolmogorov-Smirnov ne permettent aucun ajustement par une loi connue, on estime alors cette loi de X . Le problème de l'estimation de la densité de probabilité est intéressant pour plusieurs raisons, nous mentionnerons le problème de l'estimation du "hasard", du taux de défaillance $\frac{f(x)}{1-F(x)}$, où F est la fonction de répartition de la variable X , de la courbe de régression de la moyenne. En physique et en météorologie par exemple, on estime la densité de probabilité pour isoler les observations rares ou pour isoler des pics assez rares sur un fond bruité.

Il existe plusieurs méthodes hiérarchisées pour estimer la densité : Approche par histogramme, estimateur de Rosenblatt, estimateur de Parzen. Nous nous n'intéresserons qu'à l'estimateur de Parzen parcequ'il est "populaire" vu sa souplesse d'utilisation et il présente de bonnes propriétés asymptotiques.

Cet estimateur à noyaux de Parzen est fonction du noyau que nous noterons K et du paramètre de lissage appelé aussi fenêtre que nous noterons h . Si le choix du noyau n'est

pas un problème dans l'estimation de la densité, il n'en est pas de même pour le choix du paramètre de lissage qui ne dépend que de la taille n de l'échantillon. Plusieurs travaux ont montrés que les estimateurs peuvent changer dramatiquement pour de petites variations du paramètre de lissage. La valeur du paramètre de lissage h qui minimise l'expression asymptotique de l'erreur relative globale est fonction de la densité inconnue f . La valeur de h est actuellement inconnue bien qu'il existe quelques théories raisonnables pour le choix de h mais ont apparemment très peu de signification pratique.

1 Estimateur de Parzen

L'estimateur de Parzen [4] s'écrit de la forme

$$f_n(x) = \frac{1}{nh} \sum_{i=1}^n K\left(\frac{x - x_i}{h}\right).$$

La fonction K est appelée noyau elle vérifie

$$\int_{-\infty}^{\infty} K(y)dy = 1, \quad \int_{-\infty}^{\infty} |K(y)|dy < \infty, \quad \sup_{-\infty < y < \infty} |K(y)| < \infty \text{ et } \lim_{y \rightarrow 0} |K(y)| = 0.$$

Le paramètre h est appelé paramètre de lissage, c'est une fonction déterministe indépendante des observations (elle ne dépend que de la taille de l'échantillon). Les seules conditions imposées à ce paramètre pour garantir la consistance de l'estimateur $f_n(x)$ de $f(x)$ sont :

$$\lim_{n \rightarrow \infty} h(n) = 0 \quad \text{et} \quad \lim_{n \rightarrow \infty} nh(n) = \infty.$$

2 Propriétés de l'estimateur

2.1 Moments de l'estimateur

On suppose que le noyau K vérifie les conditions suivantes

$$\int K(y)dy = 1, \quad \int yK(y)dy = 0, \quad \int y^2K(y)dy = \sigma_k^2 > 0$$

et le paramètre de lissage h vérifie : $\lim_{n \rightarrow \infty} h(n) = 0$ et $\lim_{n \rightarrow \infty} nh(n) = \infty$.

En faisant un développement de Taylor autour du point x et en faisant tendre n vers l'infini, on obtient

$$\text{Biais}(f_n(x)) = \frac{1}{2}\sigma_k^2 f''(x) + \mathcal{O}(h^4) \quad \text{et la variance} \quad \mathbb{V}(f_n(x)) = \frac{1}{nh} f(x) \int K^2(x)dx + \mathcal{O}(n^{-1}).$$

2.2 Propriétés de l'estimateur

Soit les hypothèses :

$$A_1 : \quad \lim_{n \rightarrow \infty} h(n) = 0 \quad , \quad \lim_{n \rightarrow \infty} nh(n) = \infty.$$

$$A_2 : \quad \int K(y)dy = 1, \quad \int_{-\infty}^{\infty} |K(y)|dy < \infty, \quad \sup_{-\infty < y < \infty} |K(y)| < \infty \text{ et } \lim_{y \rightarrow 0} |K(y)| = 0.$$

f_n est un estimateur consistant en moyenne quadratique

Si le paramètre de lissage h satisfait les hypothèses A_1 et le noyau K satisfait les hypothèses A_2 . Alors l'estimateur f_n est un estimateur consistant en moyenne quadratique. c'est-à-dire

$$\mathbb{E}|f_n(x) - f(x)|^2 \rightarrow 0 \text{ quand } n \rightarrow \infty,$$

en tout point x pour lequel la densité f est continue.

f_n est un estimateur uniformément consistant

Si le paramètre de lissage h satisfait les hypothèses A_1 , si noyau K satisfait les hypothèses A_2 et si la transformée de Fourier de $K(y)$ définie par $\int e^{-iuy}K(y)dy$ est absolument intégrable. Alors

$$\forall \varepsilon > 0, \mathbb{P}\left(\sup_x |f_n(x) - f(x)| < \varepsilon\right) = 1,$$

en tout point x pour lequel la densité f est uniformément continue.

f_n est un estimateur asymptotiquement gaussien

Si le paramètre de lissage h satisfait les hypothèses A_1 , et si noyau K satisfait les hypothèses A_2 . Alors f_n est un estimateur asymptotiquement gaussien. C'est-à-dire

$$\forall c \in \mathbb{R}, \mathbb{P}\left(\frac{f_n(x) - \mathbb{E}f_n(x)}{\sqrt{\mathbb{V}(f_n(x))}} < c\right) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^c e^{-\frac{y^2}{2}} dy.$$

3 Choix du paramètre de lissage

3.1 estimateur optimal

Rosenblatt [5] donne le choix optimal du paramètre de lissage h qui minimise l'erreur relative globale notée IMSE et définie par

$$\int_{-\infty}^{\infty} \mathbb{E}(f_n(x) - f(x))^2 dx = \frac{f(x)}{nh} \int K^2(y)dy + \frac{1}{4}h^4[f''(x)]^2[\int y^2 K(y)dy]^2 + o\left(\frac{1}{nh} + h^4\right). \quad (1)$$

Une approximation de cette erreur relative globale est

$$\int_{-\infty}^{\infty} \mathbb{E}(f_n(x) - f(x))^2 dx \simeq \frac{1}{nh} \int K^2(y)dy + \frac{1}{4}h^4 \int (f''(x))^2 dx [\int y^2 K(y)dy]^2 + o\left(\frac{1}{nh} + h^2\right). \quad (2)$$

Le coefficient optimal h_0 qui minimise l'IMSE est obtenu en dérivant (2) par rapport à h et en annulant la dérivée quand $n \rightarrow \infty$. On obtient

$$h_0 = C_f n^{-\frac{1}{5}}, \quad (3)$$

$$\text{avec } C_f = \frac{2^{\frac{4}{5}} (\int K^2(y)dy)^{\frac{1}{5}}}{(\int (f''(x))^2 dx)^{\frac{1}{5}} (\int y^2 K(y)dy)^{\frac{2}{5}}},$$

où la constante C_f fait intervenir la dérivée seconde la densité inconnue f .

3.2 estimateur de Silverman

Le moyen le plus simple [6] pour choisir le paramètre de lissage h est de choisir une loi pour la densité f . Si on suppose que f est la loi normale et le noyau K est le noyau gaussien, la constante C_f sera alors égale à $10.59S$. L'optimal paramètre de lissage h_0 sera de la forme

$$h_0(n) = 1.059Sn^{-\frac{1}{5}}, \quad S \text{ étant l'écart type de l'échantillon } X_1, X_2, \dots, X_n.$$

Cet estimateur est appelé Silverman's Rule of Thumb.

3.3 Estimateur du pseudo-maximum de vraisemblance

Habbema et Hermans [2] ont introduit une estimation du paramètre de lissage h par la méthode du Pseudo-maximum de vraisemblance (Pseudo-Likelihood) dite "Méthode du Jackknife". La valeur \hat{h} choisie pour h à partir de l'échantillon est la valeur qui maximise

$$PL(h) = \prod_{j=1}^n \frac{1}{n-1} \sum_{\substack{i=1 \\ i \neq j}}^n K\left(\frac{x_i - x_j}{h}\right).$$

Prendre la vraisemblance usuelle (sans considérer $i \neq j$) conduit à $h = 0$ donc à une dégénérescence vers une somme de Dirac sur les points x_i .

Remarque.

Aucun résultat théorique n'a encore montré que l'estimateur de $h(n)$ obtenu cette méthode ne conduit à une bonne vitesse de convergence.

3.4 Estimateur combiné

La procédure que nous utilisons pour estimer h est la suivante : Partant d'un nombre n_1 d'observations tirées au hasard parmi les n observations, on obtient une estimation $\hat{h}(n_1)$ de $h(n_1)$ par le pseudo maximum de vraisemblance et donc à partir de (3) une première estimation de la constante C_f

$$\hat{C}_f(n_1) = \hat{h}(n_1)n_1^{\frac{1}{5}}.$$

On prend ensuite une deuxième valeur $n_2 > n_1$ (par exemple $n_2 = 4n_1$) on relance l'algorithme d'estimation de $h(n_2)$

$$\hat{h}(n_2) = \hat{C}_f(n_1)n_2^{-\frac{1}{5}}.$$

Ce qui donne une deuxième estimation de la constante C_f

$$\hat{C}_f(n_2) = \hat{h}(n_2)n_2^{\frac{1}{5}}.$$

On arrive ainsi après k itérations (jusqu'à ce que la constante C_f ne change pas beaucoup, en général $k = 2$ ou 3) à la valeur

$$\hat{h}(n) = \hat{C}_f(n_k)n^{-\frac{1}{5}} = \hat{h}(n_k)n_k^{\frac{1}{5}}n^{-\frac{1}{5}} = \hat{h}(n_k)\left(\frac{n_k}{n}\right)^{\frac{1}{5}}.$$

3.5 Estimateur par les plus proches voisins

Généralement la procédure précédente d'estimation du paramètre de lissage ne différencie pas les régions de forte densité, ou la fonction densité risque d'être trop lissée (quand h est trop grand) et celles, ou les observations sont moins concentrées et ou peuvent apparaître dans l'estimation si h est trop faible des ruptures ou des pics non significatifs. Pour pallier à cette difficulté, on peut pour chaque observation x_i moduler le paramètre h en fonction de la concentration des observations autour de x_i . On écrira alors

$$f_n(x) = \frac{1}{n} \sum_{i=1}^n \frac{1}{h_i} K\left(\frac{x - x_i}{h_i}\right), \text{ où } h_i = hd(i), \text{ } h \text{ étant l'estimateur du pseudo-likelihood.}$$

$d(i)$ est une fonction d'autant plus faible que la densité dans le voisinage du point est élevé. Fixant à priori un nombre entier $k(n)$, $0 < k(n) < n$, on détermine les $k(n)$ plus proches voisins de x_i parmi x_1, x_2, \dots, x_n . A l'aide de ces seuls $k(n)$ (y compris x_i), on calcule, pour chaque x_i ,

$$\text{la moyenne locale } \bar{X}_{i,k(n)} = \frac{1}{k(n)} \sum_{j=1}^{k(n)} X_j, \text{ la variance } \sigma_{i,k(n)}^2 = \frac{1}{k(n)} \sum_{j=1}^{k(n)} (X_j - \bar{X}_{i,k(n)})^2,$$

$$d(i) = \frac{dd(i)}{\bar{d}}, \text{ avec } dd(i) = \sqrt{\sigma_{i,k(n)}^2} \text{ et } \bar{d} = \sum_{i=1}^{k(n)} dd(i).$$

D'après Loftsgaarden et Quesenberry [3], se basant sur plusieurs travaux empiriques, la valeur de $k(n)$ proche de \sqrt{n} donne d'assez bons résultats.

4 Choix optimal du noyau

En minimisant l'erreur relative globale notée IMSE définie dans (2), on obtient le choix optimal du noyau K sous les conditions

$$\int K(y)dy = 1, \quad K(-y) = K(y), \quad \int y^2 K(y)dy < \infty.$$

On obtient

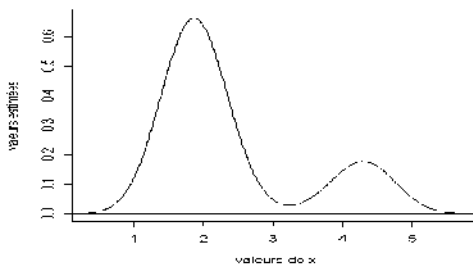
$$K_0(y) = \begin{cases} \frac{3}{4\sqrt{5}}(1 - y^2) & \text{si } |y| \leq \sqrt{5} \\ 0 & \text{si } |y| > \sqrt{5}. \end{cases}$$

Le noyau optimal K_0 est indépendant de la vraie densité f et de la taille de l'échantillon. Epanechnikov a montré dans [1] que le choix du noyau n'est pas important, et qu'il est complètement satisfaisant de choisir la fonction noyau pour la convenance du calcul informatique comme par exemple le noyau gaussien.

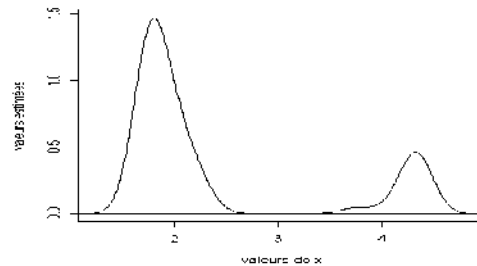
5 Exemple

L'exemple traité concerne des données météorologiques dont les mesures sont des vitesses de vent en hautes altitudes.

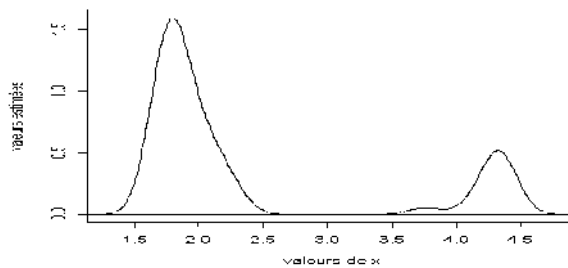
Sur le graphe 1 est représentée l'estimation de la densité de probabilité de ces données dans le cas où le noyau K est le noyau gaussien et où le paramètre de lissage $h = 0.4455$ est estimé par l'estimateur simple. Sur le graphe 2 est représentée l'estimation de la densité de probabilité de ces données dans le cas où le noyau K est le noyau gaussien et où le paramètre de lissage $h = 0.1492$ est estimé par la méthode combinée. Sur le graphe 3 est représentée l'estimation de la densité de probabilité de ces données dans le cas où le noyau K est le noyau gaussien et où le paramètre de lissage $h = 0.1299$ est estimé par la méthode des plus proches voisins.



Graphe 1. h estimé par Rule of Thumb



Graphe 2. h estimé par la méthode combinée



Graphe 3. h estimé par la méthode des plus proches voisins

Références

- [1] V.A. Epanechnikov. Non parametric estimation of a multivariate probability density. *Theory of probability and its applications*, 14 :141, 1969.
- [2] J.D.F. Habbema and J. Hermans. The allock package multigroup discriminant analysis programs based on direct density estimation. *Comp. Stat. Gordesch and P. naeve, physica, Verlag. - Wien*, 1976.
- [3] D.O. Loftsgaarden and C.P. Quesenberry. A non parametric estimate of a multivariate density function. *Ann. Math. Stat.*, 36, 1965.
- [4] E. Parzen. On estimation of probability density function and mode. *Ann. Math. Stat.*, 33 :1065, 1962.
- [5] M. Rosenblatt. Curves estimates. *Ann. Math. Stat.*, 33 :1815, 1962.
- [6] B.W. Silverman. *Density estimation for statistics and data analysis*. Chapman and hall : London, 1986.

Robust IF Estimation of Multicomponent FM Signals Affected by Heavy-Tailed Noise Using TFD

Mohamed SAHMOUDI¹ and Karim ABED-MERAIM²

¹ LSS, CNRS-SUPELEC-UPS, 91192, Gif-sur-Yvette

² TSI, Telecom Paris, 37-39 rue Dareau, 75014 Paris

E-mail: sahmoudi@lss.supelec.fr, abed@tsi.enst.fr

Abstract

We consider the problem of instantaneous frequency (IF) estimation of multicomponent frequency-modulated (FM) signals in heavy-tailed noise environment. The proposed methods are based on the signal time-frequency distribution (TFD). To combat the noise effect two approaches are proposed: (i) in the first one, a new robust time-frequency distribution is developed as a generalization of the robust minimax M-estimates; (ii) while in the second one, a signal pre-processing is performed to mitigate the noise term. Monte-Carlo simulations are presented in order to validate and assess the performances of the proposed methods.

I. INTRODUCTION-PROBLEM STATEMENT

THIS paper is concerned with the analysis of multicomponent frequency-modulated (FM) signals, corrupted by additive heavy-tailed noise. By a multicomponent signal, we mean a signal whose time-frequency representation presents multiple ridges in the time-frequency plane. Analytically, the noisy signal considered in this paper may be defined as,

$$x(t) = \sum_{i=1}^M s_i(t) + z(t) \quad (1)$$

where each component $s_i(t)$, of the form

$$s_i(t) = a_i(t) e^{j\phi_i(t)}, \quad (2)$$

is assumed to have only one ridge, or one continuous curve, in the time-frequency plane. $a_i(t)$ is the amplitude and $\phi_i(t)$ denotes the phase of the i th component of the signal. The probability density function (pdf) of the random noise $z(t)$ is modeled as a heavy-tailed distribution¹. Examples of this kind of distributions include generalized Gaussian and α -stable distributions with $\alpha < 2$:

- **Symmetric α -stable process ($S_{\alpha S}$):** $S_{\alpha S}$ have proved to be effective in modeling many real-life engineering problems such as outliers and impulse signals [10]. The pdf of $S_{\alpha S}$ processes do not have closed form except for the cases $\alpha = 1$ (Cauchy distribution) and $\alpha = 2$ (Gaussian distribution). The $S_{\alpha S}$ pdf is defined by means of its characteristic function $\varphi_{\alpha}(t) = \exp\{j\mu t - \gamma|t|^{\alpha}\}$, where α ($0 < \alpha \leq 2$) is the characteristic exponent, controlling the heaviness of the pdf tail, γ ($\gamma > 0$) is the dispersion, which plays an analogous role to the variance, and μ ($\mu \in \mathbb{R}$) is the location parameter, the symmetry axis of the pdf. Due to their heavy tail, stable distributions do not have finite second or higher-order moments, except for the limiting case of $\alpha = 2$.
- **Generalized Gaussian pdf:** Another way to modelize impulsive noise processes is through the generalized Gaussian pdf given by [9] $G_{\alpha}(x) = A \exp(-b|x|^{\eta})$ where $A = \frac{\Gamma(\frac{\eta}{2})}{\sqrt{\eta} \Gamma(\frac{\eta}{2})}$, $b = |\sigma|^{-\frac{2\eta}{\eta-2}} \left[\frac{\Gamma(\frac{\eta}{2})}{\Gamma(\frac{\eta}{2})} \right]^{\frac{2}{\eta}}$ and $\eta > 0$ ($\Gamma(x)$ being the usual Gamma function). For $\eta = 2$ we have the Gaussian distribution and for $\eta = 1$ we have the Laplacian distribution (which is known to be a good model for impulsive noise [9]).

Our primary interest, in this work, is to estimate the instantaneous frequency of each FM signal $s_i(t)$ of (1), defined as

$$IF_i(t) \triangleq \frac{1}{2\pi} \frac{d\phi_i(t)}{dt} \quad (3)$$

The problem of non-parametric IF estimation for multicomponent non-stationary signals is an important and unresolved issue in signal processing. Time-frequency analysis techniques are used as they reveal the multicomponent nature of such signals. Ideally, for a given FM signal, the TFD is represented as a row of delta functions around the signals instantaneous frequency. This property makes the peak of the TFD a very powerful tool as an IF estimator. However, quadratic TFD of multicomponent signals suffer from the presence of cross-terms, which can obscure the real features of interest in the signal. The properties

¹For complex-valued noise signal, we simply consider that $z(t) = z_r(t) + jz_i(t)$ where $z_r(t)$ and $z_i(t)$ represents two independent heavy-tailed processes with a same pdf function.

of a quadratic TFD are completely determined by its kernel. This kernel should have the shape of a two-dimensional (2-D) low-pass filter to attenuate the cross-terms that exist away from the origin in the ambiguity domain and preserve the auto-terms that concentrate around the origin of this domain [6]. Considerable efforts have been made to define TFD that reduce the effect of cross terms while improving the time-frequency resolution (e.g., [2], [6]). This led to the so-called reduced interference distributions that include the modified B-distribution (MBD), and the signal-dependent optimal time-frequency representation. In this work, we have used the MBD [6] given by:

$$T(t, f) = \int \int_{-\infty}^{+\infty} G_{MBD}^{\sigma}(t') [x(t - t' + \frac{T}{2}) x^*(t - t' - \frac{T}{2})] e^{-j2\pi f t'} dt' dT \quad (4)$$

where $G_{MBD}^{\sigma}(t') = \frac{k_{\sigma}}{\cos(\pi t')^{2\sigma}}$, $0 \leq \sigma \leq 1$ is a real parameter that controls the tradeoff between component's resolution and cross-terms suppression and $k_{\sigma} = \Gamma(2\sigma)/(2^{2\sigma-1})\Gamma^2(\sigma)$ is the normalizing factor. The choice of the MBD, stems from the fact that it presents a good performance in terms of resolution and cross-terms suppression [6]. The effect of noise on the TFD is another consideration that has direct influence on IF estimation. Study of additive Gaussian noise influence on TFD-based IF estimation is an important issue [2], [6], [11]. However, in many practical applications, especially in communications, signals are disturbed by impulsive noise due to the propagation environment or to large errors in collecting and recording the data. These noise processes are commonly modeled by heavy-tailed distribution [10]. Since outliers or impulsive noise have an unusually great influence on standard (least squares, maximum likelihood, TFD-based, ...) IF estimators, robust procedures attempt to modify those schemes. Among the existing robust estimation techniques, one have the class of M-estimation methods. The M-estimation principle can be applied in order to design TFD based methods robust with respect to impulsive noise. In [3], [8], the authors proposed the robust spectrogram and the robust polynomial Wigner-Ville distribution (PWVD), respectively. However, it is known that the spectrogram suffers from low resolution in time-frequency domain, while the PWVD suffers from cross-terms for multicomponent signals. In this paper, we propose three different robust procedures designed to handle the multicomponent FM signal case. The first and the second is based on a preprocessing stage to attenuate the heavy-tailed noise effect, whilst the third one is based on a new robust TFD which is referred to as the *robust modified B-distribution (R-MBD)*. In this latter method we use the modified B-distribution kernel [6] and the M-estimation theory to design the R-MBD. We show that each of the proposed methods can solve problems that existing distributions cannot.

II. ROBUST ESTIMATION

It is recognized that "outliers" which arise from heavy tailed distribution noise or are simply bad points due to measurement errors, have an unusually large influence on the standards estimators like as least squares estimators. Accordingly, robust methods have been developed to modify least squares schemes so that the outliers have much less influence on the final estimates. One of the most satisfying robust procedures is that given by a modification of the principle of maximum likelihood; hence we begin with that approach called M-estimation [5].

A. M-Estimation

Let x_1, \dots, x_N be an i.i.d. random sample that arises from a distribution with density $p(x - \theta)$ of the continuous type, where θ is a location parameter. The logarithm of the likelihood function is

$$\ln L(\theta) = \sum_{i=1}^N \ln p(x_i - \theta) = - \sum_{i=1}^N \rho(x_i - \theta) \quad (5)$$

where $\rho(x) = -\ln p(x)$. In maximum likelihood we wish to maximize $\ln L(\theta)$ or, in terms of the *loss function* ρ , minimize $R(\theta) = \sum \rho(x_i - \theta)$. Suppose that this minimization can be achieved by differentiating and solving $R'(\theta) = 0$; that is, finding the appropriate θ that satisfies $\sum_{i=1}^N \psi(x_i - \theta) = 0$, where $\psi(x) = \rho'(x) = -p'(x)/p(x)$ (ψ is known as the score function). The solution of this equation that minimizes $R(\theta)$ is the maximum-likelihood estimate of θ .

Some typical classroom examples of this process are given by the following distributions, the first of which provides the least squares estimate, the second, the "least absolute values" estimate and the third provides the M-estimate.

- 1) Normal criterion: For a gaussian distribution, the ML estimation leads to

$$\rho(x) = \frac{x^2}{2}; \quad \psi(x) = x \quad (6)$$

$$\sum_{i=1}^N (x_i - \theta) = 0 \quad \text{yields} \quad \hat{\theta} = \bar{x} \triangleq \frac{1}{N} \sum_{i=1}^N x_i$$

- 2) Double exponential criterion: For a modulus distributions, the score function is given by

$$\rho(x) = |x|; \quad \psi(x) = \begin{cases} -1 & x < 0 \\ 1 & x > 0 \end{cases} \quad (7)$$

$$\sum_{i=1}^N \psi(x_i - \theta) = 0 \quad \text{yields} \quad \hat{\theta} = \text{sample median}$$

3) **Huber criterion:** Using a more technical definition of robustness, Huber [5] derived the following robust ρ and ψ :

$$\rho_H(x) = \begin{cases} \frac{x^2}{2}, & |x| \leq K \\ K|x| - \frac{K^2}{2}, & |x| > K \end{cases} \quad ; \quad \psi_H(x) = \begin{cases} \text{sign}(x)K, & |x| > K \\ x, & |x| \leq K. \end{cases} \quad (8)$$

$$\sum_{i=1}^N \psi_H(x_i - \theta) = 0 \quad \text{is solved by numerical methods}$$

These are the ρ and ψ functions associated with a function which is “normal” in the middle with “double exponential” tails. The constant K regulates the degree of robustness; good choices for K are between 1 or 2 times the standard deviation of the observations. The corresponding M-estimator is the minimax solution.

B. Optimal loss function

The maximum likelihood (ML) approach can be used for selection of the appropriate loss function ρ if the pdf $p(z)$ of the noise is known. The ML suggests the best loss function $\rho = -\ln p(z)$ which minimizes the asymptotic variance of the estimation error [5]. However, the ML solution is quite sensitive to deviations from the parametric model and the hypothetical distribution. Even a slight deviation from the hypothesis can result in a strong degradation of the ML estimate. The robust minimax approach is an alternative to the conventional ML that overcomes the ML estimates sensitivity and improves the efficiency in an environment with unknown heavy-tailed pdf. The robust Huber’s minimax estimator over a certain density class \mathcal{P} is defined by

$$\hat{\theta}_{robust} = \text{arg min}_{\theta} \sum_{i=1}^N \tilde{\rho}\{x_i - \theta\} \quad (9)$$

with the loss function

$$\tilde{\rho}(z) = -\ln(\tilde{p}(z)) \quad (10)$$

$\tilde{p}(z)$ can be defined as a solution of the problem

$$\tilde{p}(z) = \text{arg min}_{p \in \mathcal{P}} I(p) \quad (11)$$

where $I(p) = \int (p'(z))^2 / p(z) dz$ denotes the Fisher information. We call the M-estimator robust if the loss function ρ is given according to (11) and (10). This approach consists to consider the worst case (among $p \in \mathcal{P}$) corresponding to the pdf giving the minimum Fisher information value. Solving the worst case would insure robustness (good estimation performance) if the considered signal pdf belongs to \mathcal{P} . It is emphasized that the robustness property of the estimator depends on how the class \mathcal{P} is defined. Thus, in order to obtain the robust minimax estimator, first, an appropriate class \mathcal{P} should be defined, and after that, the loss function ρ is given by (10) and (11).

Note that ρ should be chosen convex with a bounded derivative $\psi = \rho'$ to guarantee the convergence of the optimization in (9) (e.g. Huber function). It is interesting to note that the solutions associated with the absolute-values criterion and Huber criterion are not influenced much by outliers and thus are robust. In the other hand, the least squares estimate \bar{X} of normal criterion is not robust since extreme values do greatly influence \bar{X} [5].

C. β -exponential criterion

To improve the efficiency of IF estimation in an environment with the heavy-tailed pdf, we propose to use the exponential loss function

$$\rho_{\beta}(x) = |x|^{\beta} \quad ; \quad \psi_{\beta}(x) = \beta \text{sign}[x] |x|^{\beta-1} \quad (12)$$

($\text{sign}[x] \triangleq x/|x|$) as a tool of the robust estimation which appears originally as a heuristic idea supported latter by theoretical and experimental studies. In particular, for $\beta = 1$, β -exponential criterion is a double exponential loss function. This latter referred to as “modulus function” in [3], [8] was used to define the robust periodogram and robust PWV distributions. We wish to emphasize that with $\beta < 2$, the β -exponential criterion is relevant to heavy-tailed distribution provided that error pdf is of the form $p(x) = C \exp(-|x|^{\beta})$ which is in the generalized Gaussian heavy-tailed distribution class.

III. HEAVY-TAILED NOISE MITIGATION

It is well known that the conventional TFD are quite sensitive with respect to non-Gaussian noise, in particular, to impulsive noise in which case they produce poor estimation results. In order to get a good estimation performance in this context, we propose in the first procedure a preprocessing of the signal to attenuate the impulsive noise effect before processing the signal TFD. In the second procedure, we use the robust statistics theory to define a new robust MB-distribution.

A. Preprocessing Techniques

The first step consists in reducing the impulsive noise amplitudes in order to improve the SNR. To do so, two solutions might be suggested.

- *β -exponential filter*: We propose here to pass the noisy signal through a nonlinear device that compresses the large amplitudes (i.e., reduces the dynamic range of the noisy signal) before further analysis [1]. The output of the nonlinear device, is expressed as

$$\tilde{x}(t) = \psi_{\beta}[x(t)] = |x(t)|^{\beta} \text{sign}[x(t)]$$

where $0 < \beta \leq 1$ is a real coefficient that controls the amount of compression applied to the input noisy signal $x(t)$.

- *Huber filter*: We use here the Huber criterion to define the Huber filter which truncates in amplitude the ‘large-valued’ observations that represent ‘‘large’’ impulsive noise realizations. For the choice of the truncation constant K , we propose to compute the histogram of observations and choose K such that $[-K, K]$ contains 90 % of the data; then, the output of the Huber linear filter, is expressed as:

$$\tilde{x}(t) = \psi_H[x(t)] = \begin{cases} x(t) & \text{if } |x(t)| \leq K \\ \text{sign}[x(t)]K & \text{if } |x(t)| > K \end{cases}$$

The second step consists in applying a time-frequency analysis presented in the next section to the processed signal $\tilde{x}(t)$ for the IF problem estimation.

B. Robust modified B-distribution (R-MBD)

Let consider the noisy signal (1) in discrete-time $x(kT) = s(kT) + z(kT)$ where T is a sampling period. We will use the MB-distribution kernel (4) in the design of the robust R-MBD to analyze FM signals affected by impulsive noise. Let us consider the optimization problem given by

$$\hat{\mathcal{B}} = \text{arg min}_{\mathcal{B}} \mathcal{J}(kT, f, \mathcal{B}) \quad (13)$$

where

$$\mathcal{J}(kT, f, \mathcal{B}) = \sum_{n=-N/2}^{N/2} w(nT) \rho[c(k, f, n)],$$

$$c(k, f, n) = \mathcal{G}_x(kT, nT) c^{-j2\pi f n T} - \mathcal{B}$$

where $w(nT)$ is a window function, $\mathcal{G}_x(kT, nT)$ being the kernel of the modified B-distribution for the FM signal $x(kT)$ and \mathcal{B} is an estimate of the expectation of the sample average of the quantity $\mathcal{G}_x(kT, nT) c^{-j2\pi f n T}$.

If we choose the loss function $\rho(c) = |c|^{\beta}$, we can show by solving for \mathcal{B} the expression $\frac{d\mathcal{J}(kT, f, \mathcal{B})}{d\mathcal{B}^*} = 0$ that the optimal solution corresponds to the standard MBD:

$$\mathcal{B}_x^s(kT, f) = \frac{\sum_{n=-N/2}^{N/2} w(nT) \mathcal{G}_x(kT, nT) c^{-j\pi f n T}}{\sum_{n=-N/2}^{N/2} w(nT)} \quad (14)$$

Similarly to the standard spectrogram, WVD and PWVD, the standard MB-distribution is not an adequate analysis tool in presence of heavy-tailed noise. In this context, we need to choose the loss function as $\rho(c) = |c|^{\beta}$ with $\beta < 2$ is a parameter to control the exponential loss function degree. The choice of this optimal criterion is well detailed in section II. In this case, we find the optimal solution, labeled the robust modified B-distribution (R-MBD), to be

$$\frac{\partial}{\partial \mathcal{B}^*} \left\{ \sum_{n=-N/2}^{N/2} w(nT) |\mathcal{G}_x(kT, nT) c^{-j\pi f n T} - \mathcal{B}|^{\beta} \right\} = 0 \quad (15)$$

$$\Leftrightarrow \sum_{n=-N/2}^{N/2} w(nT) (\mathcal{G}_x(kT, nT) c^{-j\pi f n T} - \mathcal{B}) |\mathcal{G}_x(kT, nT) c^{-j\pi f n T} - \mathcal{B}|^{\beta-2} = 0 \quad (16)$$

$$\Leftrightarrow \mathcal{B}_x^r(kT, f) = \frac{\sum_{n=-N/2}^{N/2} d(k, f, n) \mathcal{G}_x(kT, nT) c^{-j\pi f n T}}{D_0(kT, f)} \quad (17)$$

$$d(k, f, n) = w(nT) |\mathcal{G}_x(kT, nT) c^{-j\pi f n T} - \mathcal{B}_x^r(kT, f)|^{\beta-2}, \quad (18)$$

$$D_0(kT, f) = \sum_{n=-N/2}^{N/2} d(k, f, n) \quad (19)$$

Since, the quantity $\mathcal{B}_x^r(kT, f)$ appears on the right as well as on the left hand side of equation (17), an iterative procedure is necessary in order to obtain the R-MBD. It was shown in [7] that the above iterative algorithm will converge to a local minimum under a good choice of the initial value.

IV. IF ESTIMATION & COMPONENT SEPARATION

In [2], an algorithm that separates the signal components and estimate their respective IF laws from the signal TFD, has been presented. Here, we propose to apply this algorithm to (i): the TFD of the pre-processed signal in the first procedure and (ii) to the robust MB-distribution of the noisy signal in the second procedure. This proposed components separation algorithm is illustrated in Table I. The first step of the algorithm consists in noise thresholding to remove the undesired 'low' energy peaks in the time-frequency domain. This operation can be written as:

$$T_{th}(t, f) = \begin{cases} T(t, f) & \text{if } T(t, f) > \epsilon \\ 0 & \text{otherwise} \end{cases}$$

where ϵ is a properly chosen threshold (in our simulations we used $\epsilon = 0.01 \max_{(t,f)} T(t, f)$).

A. Components separation procedure

Assuming a 'clean' TFD, the M components IF are estimated, at each time instant t , from the M peak positions of the TFD slice $T_{th}(t, f)$. Let observe that if, at a time instant t_0 , two components are crossing then the number of peaks (at this particular slice $T(t_0, f)$) is smaller than the total number of components M . For practical implementation reasons, we decide that a crossing occurs when the number of peaks is smaller than M over a fixed number of consecutive slices. In this case, we implement the following procedure:

- 1) Choose a particular maximum point location in the slice where the crossing occurs.
- 2) Measure all distances from this point to the peaks locations of the previous slice (with no crossing).
- 3) Select the 2 smallest distances and add them.
- 4) Repeat Steps 1 to 3 for all other maximum point locations in the slice where the crossing occurred.
- 5) From the set of the smallest sums found above, the program selects the smallest value and the points associated to them. This will yield the location where the crossing occurred and the 2 components involved in the crossing.

Then, we use a simple numerical permutation operation of the 2 components involved in the crossing. The details of the proposed separation technique is outlined in Table I.

| |
|---|
| <ol style="list-style-type: none"> 1. Assign an index to each of the M components in an orderly manner. 2. For each time instant t (starting from $t = 1$), find the components frequencies as the peaks positions of the TFD slice $T(t, f)$. 3. Assign a peak to a particular component based on the smallest distance to the peaks of the previous slice $T(t-1, f)$ (the IFs are continuous functions of time). For the special case of a crossing point (see Step 4 how to detect it and its corresponding components), we assign the peak to both crossing components. 4. If at a time instant t a crossing point exists (i.e., number of peaks smaller than the number of components), identify the crossing components using the smallest distance criterion by comparing the distances of the actual peaks to those of the previous slice. 5. Permute the indices of the corresponding crossing components. |
|---|

Table I: Component separation procedure for the proposed algorithm

V. PERFORMANCE EVALUATION & COMPARISON

In this section, we consider a discrete-time multicomponent FM signal consisting of two linear FM components embedded in additive impulsive α -stable noise $z(n)$

$$x(n) = s_1(n) + s_2(n) + z(n) \quad n = 0, 1, \dots, N-1$$

where $s_1(n) = \exp\{j2\pi(a_1n + b_1n^2)\}$ and $s_2(n) = \exp\{j2\pi(a_2n + b_2n^2)\}$. The noise $z(n)$ is chosen with zero location parameter, characteristic exponent $\alpha = 1$ and dispersion equal to $\gamma = 1$. The signals' IF coefficients are given by $a_1 = 0.2$, $b_1 = 0.1 * 10^{-3}$, $a_2 = 0.45$ and $b_2 = -1.5 * 10^{-3}$. In the first step, we perform a pre-processing of the noisy signal to mitigate the impulsive noise using the β -exponential filter in the β -TFD algorithm (with parameter $\beta = 0.1$) and the Huber filter in the Huber-TFD algorithm. In the second step, we put the pre-processed signal $\tilde{x}(n)$ through the proposed algorithm (we chose $\sigma = 0.01$ for the MB-distribution kernel) in order to extract the two respective components. The peaks of the extracted components (in the time-frequency domain) are, then, used to estimate the IFs of

the chirps [2]. We use a simple polynomial fit to obtain estimates of (a_1, b_1) from $IF_1(n)$ and estimates of (a_2, b_2) from $IF_2(n)$.

The same noisy signal $x(n)$ is also put through the R-MBD (with $\beta = 1$) algorithm developed in this paper to validate this method and to compare it with the preprocessing-based methods. The estimation performance is measured by the normalized MSE defined by

$$NMSE = \frac{1}{N_r} \sum_{r=1}^{N_r} \frac{\|\hat{\theta}_r - \theta\|^2}{\|\theta\|^2}$$

where θ is the considered parameter, $\hat{\theta}_r$ is the estimate of θ at the r th experiment and N_r is the number of Monte-Carlo chosen here equal to 200. In Figure I, we display the NMSE of the R-MBD, β -TFD and Huber-TFD versus the sample size.

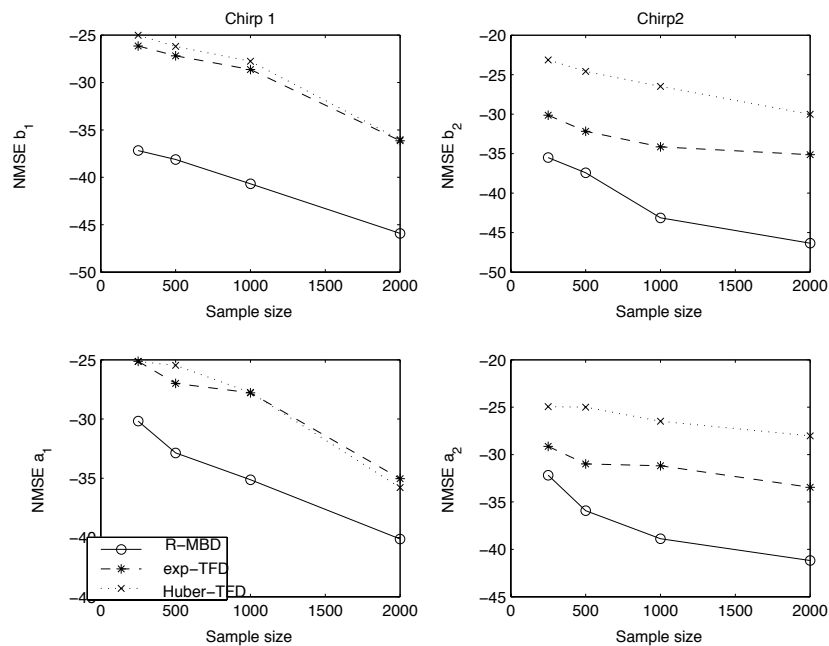


Fig I: The NMSE versus sample size: a comparative study.

These simulations confirm the effectiveness of the proposed algorithms and, at least in this simulation context, the best results in terms of estimation accuracy are obtained by the R-MBD algorithm (which is, on the other hand, the most expensive one) followed by the β -TFD method.

REFERENCES

- [1] B. Barkat and K. Abed-Meraim, "An effective technique for the IF estimation of FM signals in heavy-tailed noise" *Proc. ISSPIT*, Germany, Dec. 2003.
- [2] B. Barkat and K. Abed-Meraim, "Algorithms for Blind Components Separation and Extraction from the Time-Frequency Distribution of their Mixture", to appear in *JASP*, 2004.
- [3] B. Barkat and L.J. Stankovic, "Analysis of polynomial FM signals corrupted by heavy-tailed noise", *Sig. Proc.*, 84, pp: 69-75, (2004).
- [4] A. Francos and M. Porat, "Analysis and synthesis of multicomponent signals using positive TFD", *IEEE Trans. Sig. Proc.*, pp. 493-504, Feb. 1999.
- [5] P. J. Huber, *Robust Statistics*. New York: Wiley, 1981.
- [6] Z. M. Hussain and B. Boashash, "Adaptive Instantaneous Frequency Estimation of multicomponent FM Signals Using Quadratic Time-Frequency Distributions", *IEEE Tr. Sig. Proc.*: 1866-1876, Aug. 2002.
- [7] S. Kaluri, G.R. Arce, "Fast algorithms for weighted myriad algorithm computation by fixed point search", *IEEE Trans. on Sig. Proc.*, Apr. 2000.
- [8] V. Katkovnik, I. Djurovic, L.J. Stankovic, "Robust Time-Frequency Representation", in B. Boashash (Ed), *Time-Frequency Signal Analysis and Processing*, Elsevier, Oxford, 2003.
- [9] S.M. Kay, *Fundamentals of Statistical Signal Processing: Detection Theory*, Prentice-Hall, Englewood Cliffs, NJ, 1998.
- [10] C.L. Nikias and M. Shao, "Signal processing with alpha-stable distributions and applications", *New York, John Wiley & Sons*, 1995.
- [11] S. Peleg and B. Friedlander, "Multicomponent signal analysis using the polynomial-phase transform", *IEEE Tr. On AES*, Jan. 1996.

Etude des mesures prédictives dans le modèle de Cox-Dirichlet à censures fixes

Fatiha MESSACI ^a et Jean-Pierre RAOULT ^b

^a Université Mentouri, Constantine, Algérie

Email : f_messaci@yahoo.fr

^b Université de Marne-la-Vallée, France

Email : raoult@math.univ-mlv.fr

Résumé

Dans cette étude nous étendons au cas d'observations censurées par des constantes le calcul des densités prédictives dans le modèle non paramétrique bayésien de Cox-Dirichlet effectué, pour des observations non censurées, par N. Gouget et J. P. Raoult, dans [3]. Nous calculons les densités restreintes à des sous ensembles des faces, parties caractérisées par la fixation des variables censurées et comportant des égalités et des inégalités strictes entre les observations non censurées.

1 INTRODUCTION

Dans [3], N. Gouget et J. P. Raoult introduisent la notion de mesure régulière par portions sur \mathbb{R}^n et, partant de la fonction de survie de la loi prédictive dans le modèle bayésien de Cox-Dirichlet, calculée dans [6], ils montrent que la mesure qui lui est associée rentre dans ce cadre. Ils en déduisent les densités restreintes de la loi prédictive à des sous ensembles caractérisés par des égalités et des inégalités strictes entre les observations. Pour préciser ce résultat, rappelons les notions et notations suivantes de [3].

Une configuration d'ex æquo est toute suite de nombres entiers strictement positifs $\lambda = (\lambda_1, \dots, \lambda_l)$ telle que $\sum_{j=1}^l \lambda_j = n$. Nous notons $n_j = \sum_{h=1}^j \lambda_h$ (avec $n_0 = 0$). Pour chaque n -permutation $\sigma = (\sigma_1, \dots, \sigma_n)$ et chaque configuration d'ex æquo λ , nous introduisons le sous ensemble de \mathbb{R}^n défini par :

$$\Delta_\sigma^\lambda = \{(t_i)_{1 \leq i \leq n} : \forall j \in \{1, \dots, l\} t_{\sigma_{n_{j-1}+1}} = \dots = t_{\sigma_{n_j}}, \\ \forall j \in \{1, \dots, l-1\} t_{\sigma_{n_j}} < t_{\sigma_{n_{j+1}}}\}.$$

Δ_σ^λ est l'ensemble de toutes les suites dans \mathbb{R}^n qui peuvent être réordonnées en croissant par l'usage de la permutation σ et dans lesquelles il y a exactement l blocs d'ex æquo qui, lorsqu'on les recense en croissant, ont les tailles successives $\lambda_1, \dots, \lambda_l$.

Pour $\lambda = (1, \dots, 1)$, suite de longueur n dont tous les éléments sont égaux à 1, nous obtenons les sous ensembles

$$\Delta_\sigma = \left\{ (t_i)_{1 \leq i \leq n} : \forall j \in \{1, \dots, l\} - t_{\sigma_j} < t_{\sigma_{j+1}} \right\}; \text{ la fermeture de } \Delta_\sigma \text{ est}$$

$$\bar{\Delta}_\sigma = \left\{ (t_i)_{1 \leq i \leq n} : \forall j \in \{1, \dots, n-1\} t_{\sigma_j} \leq t_{\sigma_{j+1}} \right\}.$$

Si σ est la permutation identité, nous obtenons les ensembles

$$\Delta^\lambda = \left\{ (t_i)_{1 \leq i \leq n} : \forall j \in \{1, \dots, l\} t_{n_{j-1}+1} = \dots = t_{n_j}, \right. \\ \left. \forall j \in \{1, \dots, l-1\} t_{n_j} < t_{n_{j+1}} \right\}.$$

Soit A une partie de \mathbb{R}^l et f une application de A dans \mathbb{R} . Soit $s = (s_1, \dots, s_k)$ une suite d'entiers distincts dans $\{1, \dots, l\}$. La k dérivée partielle $\frac{\partial^k f}{\partial x_{s_1} \dots \partial x_{s_k}}$ de f est notée $\partial_s f$.

Une mesure finie sur \mathbb{R}^n est dite régulière par portions si, pour toute n -permutation σ , la restriction de sa fonction de survie, S , à $\bar{\Delta}_\sigma$, notée \hat{S}_σ , admet la n -dérivée continue $\partial_{1, \dots, n} \hat{S}_\sigma$.

La densité de la restriction d'une telle mesure à Δ^λ est donnée dans le théorème 2 de [3]; le passage de Δ^λ à Δ_σ^λ se fait alors par permutation.

Dans [4] F. Messaci étend cette étude au cas où des censures aléatoires à droite, à gauche ou mixtes, elles même régies par un modèle bayésien de Cox-Dirichlet, sont appliquées aux observations. La mesure prédictive obtenue jouit encore de la propriété de régularité par portions et les calculs des densités des restrictions de cette loi aux ensembles Δ_σ^λ en sont déduits, dans le cas de la censure à droite. Dans ce travail nous nous intéressons au cas des censures fixes appliquées à des temps de survie toujours régis par un modèle bayésien de Cox-Dirichlet. De façon précise, nous observons n variables aléatoires positives censurées $X_i = T_i \wedge \tau_i$. τ_i est fixé pour tout i tel que $1 \leq i \leq n$. Les variables aléatoires T_i sont strictement positives indépendantes et il existe une fonction de survie de base G et une suite de nombres strictement positifs $(c_i)_{1 \leq i \leq n}$, appelés contraintes, de sorte que la fonction de survie de T_i soit la puissance G^{c_i} . Par ailleurs la famille des lois de base est munie de l'a priori de Dirichlet défini dans [1] et caractérisé par la mesure α^* , de fonction de survie α .

La loi prédictive des (T_i) a été donnée pour tout $t = (t_i)_{1 \leq i \leq n}$ par M. Ruggiero, dans [6], par l'expression suivante :

$$S(t) = \frac{\Gamma(\alpha(0))}{\Gamma(\alpha(0) + d)} \prod_{i=1}^n \frac{\Gamma(\alpha(t_{\sigma_i}) + d_i^\sigma)}{\Gamma(\alpha(t_{\sigma_i}) + d_{i+1}^\sigma)} \quad (1)$$

où :

- où Γ désigne la fonction gamma eulérienne
- $\alpha(0)$ est la masse totale de la mesure α^* ,
- $\sigma = (\sigma_1, \dots, \sigma_n)$ est une permutation réordonnant en croissant la suite $(t_i)_{1 \leq i \leq n}$, c'est à dire telle que la suite $(t_{\sigma_j})_{1 \leq j \leq n}$ soit croissante,
- $d_i^\sigma = \sum_{h=i}^n c_{\sigma_h}$ avec $d_{n+1}^\sigma = 0$ et $d = d_1^\sigma = \sum_{i=1}^n c_i$.

La loi prédictive des (X_i) est concentrée sur $\prod_{i=1}^n [0, \tau_i]$ et perd la propriété de

régularité par portions. Cependant sur $\prod_{i=1}^n [0, \tau_i]$ l'étude est la même que dans le cas non censuré considéré par N. Gouget et J. P. Raoult dans [3]. Il reste donc à calculer les densités restreintes aux faces, ensembles caractérisés par la fixation des variables censurées. Nous traitons d'abord le cas général d'une mesure régulière par portions censurée par des constantes puis nous appliquons le résultat obtenu au cas particulier de la mesure prédictive dans le modèle bayésien de Cox-Dirichlet.

2 Calcul des densités

Soit $(T_i)_{1 \leq i \leq n}$ un vecteur aléatoire de loi μ régulière par portions, de fonction de survie S . On s'intéresse à la loi du vecteur aléatoire $(T_i \wedge \tau_i)_{1 \leq i \leq n}$ où τ_i est une censure (à droite) fixe pour tout i tel que $1 \leq i \leq n$.

Chacune des faces est caractérisée par l'ensemble \mathcal{C} des indices i qui y sont fixés à la valeur de la censure τ_i . Soit donc

$\mathcal{C} = \{(c_1, \dots, c_b) \subseteq \{1, \dots, n\}, c_1 < c_2 \dots < c_b\}$; la face de dimension $n - b$ associée est $\mathcal{F}_{\mathcal{C}} = \{(x_1, \dots, x_n) / x_i = \tau_i \forall i \in \mathcal{C}, x_i < \tau_i \forall i \notin \mathcal{C}\}$.

En tenant compte du fait que, si $i \notin \mathcal{C}$, alors $x_i < \tau_i$, nous déduisons que $\mathcal{F}_{\mathcal{C}}$ a une intersection vide, avec certains des ensembles Δ_{σ} . Cependant il est élémentaire que pour toute permutation σ telle que $\Delta_{\sigma} \cap \mathcal{F}_{\mathcal{C}}$ n'est pas vide, la restriction de la mesure prédictive à $\Delta_{\sigma} \cap \mathcal{F}_{\mathcal{C}}$ admet, pour densité, le produit par $(-1)^{n-b}$ de la $\{r_1, \dots, r_{n-b}\}$ -dérivée de S_{σ} où $r_j \notin \mathcal{C}$ pour tout j tel que $1 \leq j \leq n - b$. Reste à étudier la restriction au complémentaire, dans $\mathcal{F}_{\mathcal{C}}$, de $\cup_{\sigma} (\Delta_{\sigma} \cap \mathcal{F}_{\mathcal{C}})$, qui se compose de deux types de parties :

1. Des parties pour lesquelles une variable non censurée prend la valeur d'une variable censurée ; elles sont de mesure nulle puisque pour une mesure régulière par portions toute variété linéaire parallèle à un axe de coordonnée a une mesure nulle.

2. Des parties comprenant des ex æquo d'une part entre des valeurs de variables non censurées et (ou) d'autre part entre des valeurs de variables censurées ; ce sont les $\Delta^{\lambda, \delta}$ que nous allons étudier maintenant en nous restreignant aux suites croissantes d'observations puisque les autres cas s'en déduisent par permutations.

Soit $\lambda' = (\lambda'_1, \dots, \lambda'_l)$ la configuration d'ex æquo de la suite croissante d'observations (x_1, \dots, x_n) . On définit la "suite de regroupements" (qui a la structure d'une configuration d'ex æquo) λ à partir de λ' en conservant les effectifs des blocs d'ex æquo de valeurs non censurées mais en remplaçant par $(1, \dots, 1)$ (suite de longueur λ'_h) tout $\lambda'_h \neq 1$ caractérisant des valeurs censurées ex æquo. On pose $\lambda = (\lambda_i)_{1 \leq i \leq l}$ et on introduit δ , suite de 1 et de 0 composée des indicateurs de censure de λ , où 1 est associé à toute observation censurée et 0 est associé à tout bloc d'observations non censurées ; on a $b = \sum_{j=1}^l \delta_j$.

Alors $\Delta^{\lambda, \delta}$ est l'ensemble des suites croissantes (t_1, \dots, t_n) telles que, si on note

$$n_j = \sum_{h=1}^j \lambda_h,$$

- (i) pour tout j ($1 \leq j \leq l$) $t_{n_{j-1}+1} = \dots = t_{n_j}$,
- (ii) pour tout j ($1 \leq j \leq l-1$) tel que $(\delta_j, \delta_{j+1}) \neq (1, 1)$ $t_{n_j} < t_{n_{j+1}}$,
- (iii) pour tout j ($1 \leq j \leq l$) et tout i ($n_{j-1} < i \leq n_j$)
 - si $\delta_j = 0$ alors $t_i < \tau_i$,
 - si $\delta_j = 1$ alors $t_i = \tau_i$.

Remarquons d'une part que la condition (iii) veut dire que $\Delta^{\lambda, \delta}$ est inclus dans \mathcal{F}_C et d'autre part que $\Delta^{\lambda, \delta}$ est en bijection avec un sous ensemble de \mathbb{R}^{l-b} par l'application : $(t_i)_{1 \leq i \leq n} \mapsto (t_{n_j})_{j / \delta_j=0}$, ce qui permet de définir la mesure de Lebesgue sur $\Delta^{\lambda, \delta}$ par transport de la mesure de Lebesgue sur \mathbb{R}^{l-b} .

Grâce à une opération de suppression des censures détaillée dans [5], nous nous ramenons au cas non censuré, ce qui nous permet d'appliquer le théorème 2 de [3]. Ensuite par réintroduction des censures, nous arrivons au résultat suivant.

Théorème 1 Soit μ une mesure régulière par portions sur \mathbb{R}^n , de fonction de survie S , à laquelle on applique n censures (à droite) fixes τ_1, \dots, τ_n . Soit $\lambda = (\lambda_j)_{1 \leq j \leq l}$ une suite de regroupements et $\delta = (\delta_j)_{1 \leq j \leq l}$ une suite d'indicateurs de censure compatible avec λ (i.e. si $\delta_j = 1$, $\lambda_j = 1$); le nombre de censures défini par δ est $b = \sum_{j=1}^l \delta_j$. Notons $A^\lambda = \{0, 1\}^{n-l}$ et $S_\sigma(t_1, \dots, t_n) = S(t_{\sigma^{-1}(1)}, \dots, t_{\sigma^{-1}(n)})$.

Alors la restriction de μ à la portion de face $\Delta^{\lambda, \delta}$ admet pour densité, par rapport à la mesure de Lebesgue, la fonction :

$$g^{\lambda, \delta} = \sum_{a \in A^\lambda} (-1)^{q(a)} \partial_{s(a)} S_\sigma(a), \quad (2)$$

où, si on note $a = ((a_m^j)_{1 \leq m \leq \lambda_j - 1})_{1 \leq j \leq l}$, on obtient, avec $n_j = \sum_{h=1}^j \lambda_h$:

- $q(a) = n - \sum_{j=1}^l \sum_{m=1}^{\lambda_j - 1} a_m^j - b$,

- $s(a) = (s_j)_{1 \leq j \leq k}$, $k = l - b$

avec $s_j = 1 + n_{p_j - 1} + \sum_{m=1}^{\lambda_{p_j} - 1} a_m^{p_j}$, où p_j est le $j^{\text{ième}}$ indice m tel que $\delta_m = 0$

- $(\sigma(a))^{-1} = ((\pi_\nu^j)_{1 \leq \nu \leq \lambda_j})_{1 \leq j \leq l}$ avec

$$\pi_\nu^j = n_{j-1} + 1 + \sum_{u=1}^{\nu-1} a_u^j + (1 - a_\nu^j)(\lambda_j - \nu).$$

Exemple

Soit $\Delta^{\lambda, \delta}$ l'ensemble des suites (t_1, \dots, t_9) telles que

$$t_1 < t_2 < t_3 = t_4 = t_5 < t_6 = t_7 < t_8 = t_9,$$

où l'ensemble des indices de censure est $\mathcal{C} = \{1, 2, 6, 7\}$.

Ici $\lambda' = (1, 1, 3, 2, 2)$, $\lambda = (1, 1, 3, 1, 1, 2)$, $\delta = (1, 1, 0, 1, 1, 0)$ et la densité res-

treinte à $\Delta^{\lambda, \delta}$ est :

$$g^{\lambda, \delta}(\tau_1, \tau_2, x_3, x_3, x_3, \tau_6, \tau_7, x_6, x_6) = (-\partial_{38} S_{125436798} + \partial_{39} S_{125436789} + \partial_{48} S_{124536789} + \partial_{48} S_{123546798} - \partial_{49} S_{124536789} - \partial_{58} S_{123456798} - \partial_{49} S_{123546798} + \partial_{59} S_{123456789}) (\tau_1, \tau_2, x_3, x_3, x_3, \tau_6, \tau_7, x_6, x_6).$$

3 Application au modèle de Cox-Dirichlet

Afin d'effectuer le passage des suites croissantes aux suites quelconques nous introduisons

$\Delta_{\sigma}^{\lambda, \delta}$ qui est l'ensemble des suites (t_1, \dots, t_n) telles que

- (i) (t_{σ_j}) est une suite croissante,
- (ii) pour tout j ($1 \leq j \leq l$) $t_{\sigma_{n_{j-1}+1}} = \dots = t_{\sigma_{n_j}}$,
- iii) pour tout j ($1 \leq j \leq l-1$) tel que $(\delta_j, \delta_{j+1}) \neq (1, 1)$ $t_{\sigma_{n_j}} < t_{\sigma_{n_{j+1}}}$,
- (vi) pour tout j ($1 \leq j \leq l$) et tout i ($n_{j-1} < i \leq n_j$)
 - si $\delta_j = 0$ alors $t_{\sigma_i} < \tau_{\sigma_i}$
 - si $\delta_j = 1$ alors $t_{\sigma_i} = \tau_{\sigma_i}$

Ici μ est la loi prédictive dans le modèle de Cox-Dirichlet de fonction de survie S donnée par la formule (1). L'application de la formule (2) du théorème 1 nécessite d'explicitier les termes $\partial_{s(a)} S_{\sigma(a)}$, ce qui après calculs conduit au résultat suivant.

Théorème 2 Soit μ la loi prédictive de n durées de survie dans le modèle de Cox-Dirichlet censuré, à droite, par les constantes τ_1, \dots, τ_n , avec la suite de contraintes $c = (c_i)_{1 \leq i \leq n}$ et l'a priori de Dirichlet défini par la mesure α^* , de fonction de survie α , qui est continûment dérivable sur \mathbb{R}_+ . Soit $\lambda = (\lambda_j)_{1 \leq j \leq l}$ une suite de regroupements (pour la suite croissante t_{σ_j}) et $\delta = (\delta_j)_{1 \leq j \leq l}$

une suite d'indicateurs de censure compatible avec λ ; soit $b = \sum_{j=1}^l \delta_j$. Nous décomposons, selon λ , la suite de contraintes c en l sous suites successives (c_{ν}^j) ; nous introduisons les sommes partielles supérieures de contraintes $e_j = \sum_{h=j}^l \sum_{\nu=1}^{\lambda_h} c_{\nu}^h$.

Alors la restriction de μ à l'ensemble (supposé non vide) $\Delta_{\sigma}^{\lambda, \delta}$ admet pour densité, par rapport à la mesure de Lebesgue, la fonction :

$$g_{\sigma}^{\lambda, \delta}(t_1, \dots, t_n) = (-1)^k \frac{\Gamma(\alpha(0))}{\Gamma(\alpha(o) + d)} \prod_{j=1}^l B_j(x_j) \prod_{j=1}^k C_j(x_{p_j})$$

où

- x_j est la valeur commune des observations t_i appartenant au $j^{\text{ième}}$ bloc d'ex a quo une fois rangées en ordre croissant,
- $k = l - b$,
- $B_j(x_j) = \frac{\Gamma(\alpha(x_j) + e_j)}{\Gamma(\alpha(x_j) + e_{j+1})}$,
- $\mathcal{P}_{\lambda_j-1}$ désignant l'ensemble de tous les sous ensembles de $\{1, \dots, \lambda_j - 1\}$ et p_j étant le $j^{\text{ième}}$ indice m tel que $\delta_m = 0$,

$$C_j(x_{p_j}) = \alpha'(x_{p_j}) \sum_{M \in \mathcal{P}_{\lambda_{p_j}-1}} (-1)^{Card M} [(\ln \Gamma)'(\alpha(x_{p_j}) + e_{p_j+1} + c_{\lambda_{p_j}}^{p_j} + \sum_{\nu \in M} c_{\nu}^{p_j}) - (\ln \Gamma)'(\alpha(x_{p_j}) + e_{p_j+1} + \sum_{\nu \in M} c_{\nu}^{p_j})].$$

Remarquons que le passage des suites croissantes (sur $\Delta^{\lambda, \delta}$) aux suites quelconques (sur $\Delta_{\sigma}^{\lambda, \delta}$) se fait en appliquant le théorème 1 à la fonction de survie : $(t_1, \dots, t_n) \rightarrow S(t_{(\sigma^{-1})_1}, \dots, t_{(\sigma^{-1})_n})$.

Références

- [1] Ferguson, T. S. (1973). A Bayesian analysis of some nonparametric problems, *Ann.Statist.* **1**, 209-230.
- [2] Gouget, N. (1999). *Statistique semi-paramétrique bayésienne de durées de vie. Résultats théoriques et mise en oeuvre en fiabilité industrielle*. Thèse, Université de Marne la Vallée, Champs-sur-Marne, France.
- [3] Gouget, N. and Raoult, J. P. (1999). Computation of predictive densities in the semi-parametric Bayesian Cox-Dirichlet model. *Non parametric Statistics.* **10**, 307-341.
- [4] Messaci, F. (2002). Mesures prédictives dans le modèle de Cox-Dirichlet à censures aléatoires. *C.R.Acad.Sci.Paris. Ser. I* **335**, 1-4.
- [5] Messaci, F. (2003). *Modèles de statistique semi-paramétrique en fiabilité*. Thèse, Université Mentouri (Constantine), Algérie.
- [6] Ruggiero, M. (1994). Bayesian semiparametric estimation of proportional hazard model, *Journal of econometrics*, **62**, 272-300.

Nouvelle classe d'estimateurs à noyau pour l'indice des valeurs extrêmes: Consistance

Djamel Meraghni^(*) & Abdelhakim Necir^(**)

Laboratoire de Mathématiques Appliquées
Université de Biskra, BP. 145 RP, 07000 Biskra, Algérie

Résumé: Dans la théorie de l'estimation de l'indice des valeurs extrêmes, beaucoup d'estimateurs ont été proposés. Pour les indices positifs, l'estimateur le plus célèbre est celui de Hill (1975) qui est généralisé par Dekkers, Einmahl et de Haan (1989) sous le nom d'estimateur des moments. L'inconvénient de ces estimateurs est le caractère discret de leur comportement. Csörgő, Deheuvels et Mason (1985) lissent l'estimateur de Hill par le biais d'un noyau, mais leurs résultats sont valables seulement dans le cas d'indices positifs. Une nouvelle classe d'estimateurs à noyau est définie par Groeneboom, Lopuhaä et de Wolf (2003). Nous en discutons la consistance.

AMS 2000 Subject Classification. Primary 62G32; secondary 62G05, 62G20.

Mots Clés: Distribution des valeurs extrêmes, indice des valeurs extrêmes, domaine d'attraction, statistiques d'ordre, estimateurs à noyau, convergence en probabilité, fonction des quantiles.

1 Introduction

Soit X une population (variable aléatoire) dont la fonction de répartition F est supposée être dans le domaine d'attraction d'une distribution G_γ des valeurs extrêmes d'indice réel γ : $F \in \mathcal{D}(G_\gamma)$. Autrement dit, il existe deux suites $(a_n)_{n \in \mathbb{N}}$ et $(b_n)_{n \in \mathbb{N}}$ avec $a_n \in \mathbb{R}_+^*$ et $b_n \in \mathbb{R}$, telles que pour tout x :

$$\lim_{n \rightarrow \infty} F^n(a_n x + b_n) = \lim_{n \rightarrow \infty} P\left(\frac{\max(X_1, \dots, X_n) - b_n}{a_n} \leq x\right) = G_\gamma(x). \quad (1.1)$$

La distribution limite G_γ est de l'une des trois formes suivantes:

$$\Lambda(x) = \exp(-e^{-x}) \quad x \in \mathbb{R}, \quad (\text{Gumbel}),$$

$$\Phi_\alpha(x) = \begin{cases} 0 & x < 0 \\ \exp(-x^{-\alpha}) & x \geq 0 \end{cases}, \quad \text{avec } \alpha > 0 \quad (\text{Fréchet}),$$

et

$$\Psi_\alpha(x) = \begin{cases} \exp(-(-x)^\alpha) & x < 0 \\ 1 & x \geq 0 \end{cases}, \quad \text{avec } \alpha > 0 \quad (\text{Weibull}).$$

Un bon choix des suites (a_n) et (b_n) permet de combiner ces trois formes en une seule, appelée distribution généralisée des valeurs extrêmes:

$$G_\gamma(x) = \exp\left(- (1 + \gamma x)^{-1/\gamma}\right), \quad (1.2)$$

pour tout x tel que $1 + \gamma x > 0$ avec $\gamma \in \mathbb{R}$. On convient que

$$G_0(x) = \lim_{\gamma \rightarrow 0} G_\gamma(x) = \exp(-e^{-x}).$$

Le principal sujet d'intérêt de la théorie des valeurs extrêmes était (et l'est toujours) le problème de l'estimation du paramètre réel γ et tout ce qui s'y rapporte. En ce sens et selon les valeurs de γ , plusieurs estimateurs (plus ou moins liés l'un à l'autre) ont été construits et discutés. Nous en citons l'estimateur de Hill (Hill (1975)), l'estimateur de Pickands (Pickands (1975)), l'estimateur à noyau (Csörgő, Deheuvels et Mason (1985)), l'estimateur du maximum de vraisemblance (Smith (1987)) et l'estimateur des moments (Dekkers, Einmahl et de Haan (1989)). Récemment, P. Groeneboom, H.P. Lopuhaä et P.P. de Wolf (2003) proposèrent un nouvel estimateur (qu'on abrégie par l'estimateur GLW) du type estimateur à noyau et présentèrent un certain nombre de résultats concernant son comportement. Un de ces résultats est précisément la consistance.

2 Définition de l'estimateur

L'estimateur GLW est construit sur la base des conditions de von Mises qui sont suffisantes pour que F appartienne au domaine d'attraction d'une distribution des valeurs extrêmes:

$$\lim_{t \uparrow x_F^0} \left(\frac{d}{dt} \frac{1 - F(t)}{F'(t)} \right) = \gamma, \quad (2.1)$$

où $x_F^0 = \sup \{x : F(x) > 1\} \leq \infty$ est le point limite supérieur de F .

L'estimateur est obtenu en remplaçant, dans (2.1), t par $Q(1-s)$ où Q est la fonction des quantiles:

$$Q(s) = F^{-1}(s) = \inf \{x : F(x) \geq s\} \text{ pour } 0 < s < 1.$$

On prélève un échantillon X_1, \dots, X_n de la population X et on note par $X_{(1)}, \dots, X_{(n)}$ les statistiques d'ordre correspondantes. On désigne par F_n et Q_n la fonction de répartition empirique et la fonction empirique des quantiles respectivement.

$$F_n(x) = \frac{1}{n} \sum_{i=1}^n 1(X_i \leq x)$$

$$Q_n(u) = \inf \{x : F_n(x) \geq u\} \text{ pour } 0 < u < 1.$$

On se donne une fonction $K : [0, 1] \rightarrow \mathbb{R}^+$, appelée noyau, avec des propriétés qu'on précisera plus loin. L'estimateur GLW est alors défini comme suit:

$$\hat{\gamma}_{n,h}^{(GLW)} = \hat{\gamma}_{n,h}^{(pos)} - 1 + \frac{\hat{q}_{n,h}^{(2)}}{\hat{q}_{n,h}^{(1)}}, \quad (2.2)$$

où

$$\begin{aligned} \hat{\gamma}_{n,h}^{(pos)} &= - \int_0^h u K_h(u) d \log Q_n(1-u) \\ &= \sum_{i=1}^{n-1} \frac{i}{n} K_h\left(\frac{i}{n}\right) (\log X_{(n-i+1)} - \log X_{(n-i)}) \end{aligned} \quad (2.3)$$

$$\begin{aligned} \hat{q}_{n,h}^{(1)} &= - \int_0^h u^\alpha K_h(u) d \log Q_n(1-u) \\ &= \sum_{i=1}^{n-1} \left(\frac{i}{n}\right)^\alpha K_h\left(\frac{i}{n}\right) (\log X_{(n-i+1)} - \log X_{(n-i)}) \end{aligned} \quad (2.4)$$

$$\begin{aligned} \hat{q}_{n,h}^{(2)} &= - \int_0^h \frac{d}{du} [u^{\alpha+1} K_h(u)] d \log Q_n(1-u) \\ &= \sum_{i=1}^{n-1} \frac{d}{du} [u^{\alpha+1} K_h(\cdot)]_{u=i/n} (\log X_{(n-i+1)} - \log X_{(n-i)}) \end{aligned} \quad (2.5)$$

avec $K_h(u) = K(u/h)/h$. On note que, en réarrangeant les termes et en utilisant le fait que

$$Q_n(1-u) = X_{(n-k)} \text{ pour } k/n \leq u < (k+1)/n, \quad (2.6)$$

On peut aussi écrire:

$$\hat{\gamma}_{n,h}^{(pos)} = \int_0^1 \log Q_n(1-hu) d(uK(u)) \quad (2.7)$$

$$\hat{q}_{n,h}^{(1)} = h^{\alpha-1} \int_0^1 \log Q_n(1-hu) d(u^\alpha K(u)) \quad (2.8)$$

$$\hat{q}_{n,h}^{(2)} = h^{\alpha-1} \int_0^1 \log Q_n(1-hu) d\left(\frac{d}{du} [u^{1+\alpha} K(u)]\right) \quad (2.9)$$

Les conditions imposées au noyau K sont :

- [C1] $K(x) = 0$ pour $x \notin [0, 1[$ et $K(x) \geq 0$ pour $x \in [0, 1[$
- [C2] $K(1) = K'(1) = 0$
- [C3] $\int_0^1 K(x) dx = 1$
- [C4] $K(\cdot)$ et $K'(\cdot)$ sont bornées.
- [C5] $\int_0^1 u^{\alpha-1} K(u) du \neq 0$

3 Consistance

Le résultat principal de cette communication, peut être considéré comme un direct corollaire des deux lemmes (3.1 et 3.2) suivants:

Lemme 3.1 Avec les mêmes hypothèses du théorème, on a $\hat{\gamma}_{n,h}^{(pos)} \rightarrow (\gamma \vee 0)$ en probabilité lorsque $n \rightarrow \infty$.

Lemme 3.2 Avec les mêmes hypothèses du théorème, on a $\hat{q}_{n,h}^{(2)}/\hat{q}_{n,h}^{(1)} \rightarrow 1 + (\gamma \wedge 0)$ en probabilité lorsque $n \rightarrow \infty$.

Théorème (consistance) On suppose que F est dans $\mathcal{D}(G_\gamma)$ (pour $\gamma \in \mathbb{R}$), K est un noyau vérifiant les conditions [C1] – [C5] (pour $\alpha > 0$) et $\hat{\gamma}_{n,h}^{(GLW)}$ est défini par (2.2). Si $h = h_n$ est tel que $h \downarrow 0$ et $nh \rightarrow \infty$, alors $\hat{\gamma}_{n,h}^{(GLW)} \rightarrow \gamma$ en probabilité lorsque $n \rightarrow \infty$.

“Preuves” des Lemmes 3.1 et 3.2:

Ces preuves sont basées sur l’application de deux autres lemmes (3.3 et 3.4), dont les démonstrations se trouvent dans Groeneboom, Lopuhaä et de Wolf (2003). On note tout d’abord que,

$$Q_n(s) \stackrel{\mathcal{D}}{=} Q(\Gamma_n(s)) \text{ et } \Gamma_n(1-s) \stackrel{\mathcal{D}}{=} 1 - \Gamma_n(s), \quad (3.1)$$

où Γ_n est la fonction empirique des quantiles d’un échantillon U_1, \dots, U_n uniforme sur l’intervalle $]0, 1[$.

De plus, en vertu des conditions [C2] et [C4], on a pour $\alpha > 0$,

$$\int_0^1 d(u^\alpha K(u)) = 0 \text{ et } \int_0^1 d\left(\frac{d}{du} [u^{\alpha+1} K(u)]\right) = 0. \quad (3.2)$$

Alors, (3.1) et (3.2) nous permettent d'écrire, avec $k = \lfloor nh \rfloor$,

$$\hat{\gamma}_{n,h}^{(pos)} = \int_0^1 (\log Q(1 - \Gamma_n(hu)) - \log Q(1 - U_{(k+1)})) d(u K(u)) \quad (3.3)$$

$$\hat{q}_{n,h}^{(1)} = h^{\alpha-1} \int_0^1 (\log Q(1 - \Gamma_n(hu)) - \log Q(1 - U_{(k+1)})) d(u^\alpha K(u)) \quad (3.4)$$

$$\hat{q}_{n,h}^{(2)} = h^{\alpha-1} \int_0^1 (\log Q(1 - \Gamma_n(hu)) - \log Q(1 - U_{(k+1)})) d\left(\frac{d}{du} [u^{\alpha+1} K(u)]\right). \quad (3.5)$$

Maintenant et selon le signe de l'indice γ , on utilise les résultats des deux lemmes suivants:

Lemme 3.3 On suppose que $F \in \mathcal{D}(G_\gamma)$ avec $x_F^0 > 0$. Alors, pour une fonction positive $a(\cdot)$,

$$\lim_{s \downarrow 0} \frac{\log Q(1 - sy) - \log Q(1 - s)}{a(s)/Q(1 - s)} = \begin{cases} -\log y & \text{si } \gamma \geq 0 \\ \frac{y^{-\gamma} - 1}{\gamma} & \text{si } \gamma < 0 \end{cases} \quad (3.6)$$

pour tout $y > 0$. De plus, pour chaque $\varepsilon > 0$ il existe s_0 tel que, pour $0 < s < s_0$ et $0 < y \leq 1$,

$$(1 - \varepsilon) \frac{1 - y^\varepsilon}{\varepsilon} - \varepsilon < \frac{\log Q(1 - sy) - \log Q(1 - s)}{\log Q(1) - \log Q(1 - s)} < (1 + \varepsilon) \frac{y^{-\varepsilon} - 1}{\varepsilon} + \varepsilon \quad (3.7)$$

au cas où $\gamma \geq 0$, et

$$1 - (1 + \varepsilon) y^{-\gamma - \varepsilon} < \frac{\log Q(1 - sy) - \log Q(1 - s)}{\log Q(1) - \log Q(1 - s)} < 1 - (1 - \varepsilon) y^{-\gamma + \varepsilon} \quad (3.8)$$

au cas où $\gamma < 0$.

Lemme 3.4 Soient h une suite de nombres positifs, avec $h = h_n \rightarrow 0$ et $nh_n \rightarrow \infty$ lorsque $n \rightarrow \infty$, et $L(\cdot)$ une fonction intégrable, bornée et positive sur $]0, 1[$. On définit $k = \lfloor nh \rfloor$ et $\bar{\lambda} = (\lambda \wedge 0)$ pour $\lambda > -1$. Si on désigne par $\Gamma_n(\cdot)$ la fonction empirique des quantiles des variables aléatoires

U_1, \dots, U_n indépendantes et identiquement distribuées (*i.i.d.*) $\mathcal{U}]0, 1[$, alors, pour chaque $\beta > (-1 - \bar{\lambda})$,

$$\int_0^1 \left[\left(\frac{\Gamma_n(hu)}{U_{(k+1)}} \right)^\beta - u^\beta \right] u^\lambda L(u) du \xrightarrow{\mathcal{P}} 0, \text{ lorsque } n \rightarrow \infty. \quad (3.9)$$

Références

CSORGO, S., DEHEUVELS, P. AND MASON, D. (1985) Kernel estimates of the tail index of a distribution. *Ann. Statist.*, **13**, 1050-1077.

DEKKERS, A.L.M., EINMAHL, J.H.J. AND DE HAAN, L. (1989) A moment estimator for the index of an extreme value index. *Ann. Statist.*, **17**, 1833-1855.

GROENEBOOM, P., LOPUHAA, H.P. AND DE WOLF, P.P. (2003). Kernel estimators for the extreme value index. *Ann. Statist.*, **31**, 1956-1995.

HILL, B.M. (1975) A simple general approach to inference about the tail of a distribution. *Ann. Statist.*, **3**, 1163-1174.

NECIR, A. (2003). On the asymptotic behavior of kernel-type estimators of the tail index. Bulletin of the International Statistical Institute 54th Session, Berlin.

PICKANDS, J. (1975) Statistical inference using extreme order statistics. *Ann. Statist.*, **3**, 119-131.

SMITH, R.L. (1987) Estimating tails of probability distributions. *Ann. Statist.*, **15**, 1174-1207.

DE WOLF, P.P. (1999) Estimating the extreme value index, PhD thesis, TU Delft, 1999.

(*) e-mail: djamelmeraghni@Yahoo.fr

(**) e-mail: Necirabdelhakim@Yahoo.fr

Modèles épidémiologiques du SIDA

Fouad Lazhar RAHMANI

Université Mentouri Constantine

Faculté des Sciences-Département de Mathématiques

Avril 2004

Introduction Une population de taille n fixé comprend deux sous-populations notées X et Y . La population X représente les individus susceptibles (ou à risque) et Y est la population des infectieux. On s'intéresse à l'évolution de l'infection déclarée dans la population totale supposée fermée. Ceci définit deux processus stochastiques $(x(t); t > 0)$ et $(y(t); t > 0)$ où à chaque instant t ($t > 0$) : $x(t)$ (resp. $y(t)$) représente la taille de la population X (resp. Y).

La diffusion de l'épidémie aux susceptibles est naturellement dépendante des contacts possibles entre susceptibles et infectieux. Si on suppose que dans un intervalle de temps relativement petit $[t, t + \Delta t]$, un infectieux ne peut avoir au plus qu'un seul contact (cette hypothèse de contamination est réaliste dans le cas par exemple de l'épidémie du SIDA), alors il est légitime de penser que le nombre de contacts est proportionnel au produit $X(t) \cdot Y(t) \cdot \Delta t$. Le nombre d'apparition de nouveaux cas infectés pendant cet intervalle de temps Δt ne peut-être que proportionnel au nombre de contacts possibles et est mesuré par $\alpha \cdot X(t) \cdot Y(t) \cdot \Delta t$ où α est le **taux d'infection**.

Le taux d'infection α dépend de multiples facteurs de contamination et ce taux peut prendre une forme expressive selon le mode de contamination actif.

On présente ci-dessous quelques modèles équivalents à des modèles suggérés dans le cadre de la modélisation mathématique de l'épidémie du SIDA et présenté par ISHEM [1].

0.1 Présentation des modèles épidémiologiques du SIDA

Il s'agit d'une population subdivisée en deux sous-populations X et Y des susceptibles et des infectieux. Ces sous - populations se comportent en general comme deux populations pouvant avoir des dynamiques d'évolution distinctes. Lorsque la population

globale est fermée (ni émigration, ni immigration), alors la dynamique de l'une des deux sous-populations explique entièrement la dynamique de la population totale.

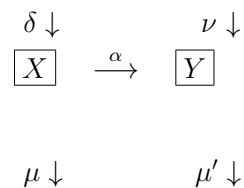
Les modèles épidémiologiques du SIDA

On suppose la population globale composée d'individus susceptibles et d'individus infectieux. On désigne par \mathbf{X} la population des susceptibles et par \mathbf{Y} celle des infectieux. Le caractère complexe de cette dernière population conduit à la subdiviser en différentes classes "homogènes". Cette caractérisation est basée sur la nécessité de contrôler les infectieux.

Entre ces différentes populations, on peut observer une dynamique caractérisée par des mouvements intra et inter populations susceptibles de l'expliquer.

Le modèle simple à une population à deux types

Un modèle simple décrivant la diffusion de l'épidémie dépend des contacts possibles entre susceptibles et infectieux, sous l'hypothèse que dans un intervalle de temps relativement petit $[t, t + \Delta t]$, un infectieux ne peut avoir au plus qu'un seul contact; et de fluctuations intra population résultant d'immigration ou d'émigration. Le modèle simple peut être illustré par le diagramme suivant :



où

α est le taux d'apparition de nouveaux infectieux (séropositifs),

μ est le taux de retrait (ou individus retirés) de la population \mathbf{X}

μ' est le taux de retrait (ou individus retirés) de la population \mathbf{Y}

δ est le taux de naissance dans la population \mathbf{X}

ν est le taux d'immigration dans la population \mathbf{Y} .

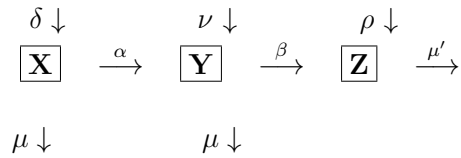
Le nombre d'apparitions de nouveaux cas infectés pendant cet intervalle de temps Δt est alors mesuré par $\alpha.X(t).Y(t).\Delta t + \nu.\Delta t - \mu'.Y(t).\Delta t$.

Le modèle à trois types

Un modèle devant tenir compte d'une politique de santé basée sur le contrôle devant entraîner l'isolement des infectieux (séropositifs et séro-négatifs déclarés), fait que l'apparition de nouveaux infectieux ne devrait résulter que de contacts entre susceptibles et infectieux qui s'ignorent. Ceci conduit à décomposer la population des infectieux en deux classes :

- * les individus infectieux (seropositifs) qui s'ignorent, notée **Y**
- * les individus infectieux (seropositifs et sideens) qui le savent, notée **Z**.

Ce modèle dit à trois types est décrit par le diagramme



où

α est le taux d'apparition de nouveaux infectieux (séropositifs) qui s'ignorent,

β est le taux d'apparition de nouveaux infectieux détectés,

δ est le taux d'immigration dans la population **X**

ν est le taux d'immigration dans la population **Y**

ρ est le taux d'immigration dans la population **Z**

μ est le taux de retrait (ou individus retirés) de la population **X** ou **Y**

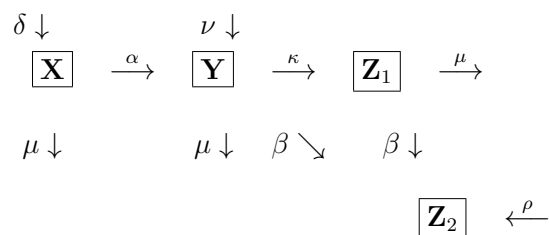
μ' est le taux de retrait (ou individus retirés) de la population **Z**.

Le nombre d'apparitions de nouveaux cas infectés pendant cet intervalle de temps Δt est ici mesuré par $\alpha.X(t).Y(t).\Delta t + \nu.\Delta t - (\mu + \beta).Y(t).\Delta t$.

Dans le modèle à trois types, les individus séropositifs détectés et sideens sont confondus.

Le modèle à quatre types

Comme il est admis qu'un séropositif est considéré comme un individu "sain" (absence de manifestations pathologiques du SIDA), ceci nous amène à envisager une subdivision de la classe des infectieux en deux sous populations : celle des séropositifs et celle des sideens. L'utilité de cette distinction permet de différencier le taux de mort dans chacune de ces sous populations et de mesurer aussi la période moyenne, le taux de passage à l'état de sida avéré, et estimer la taille de la population des sideens; et également de mesurer l'efficacité d'un programme de détection de nouveaux infectieux séropositifs. Si on désigne par **Z₁** la population des séropositifs détectés et par **Z₂** celle des sideens avérés, le modèle précédent dit à quatre types est décrit par le diagramme suivant :



$$\mu' \downarrow$$

où

α est le taux d'apparition de nouveaux infectieux (seropositifs) qui s'ignorent,

κ est le taux d'apparition de nouveaux seropositifs detectes

β est le taux d'apparition de nouveaux sideens

μ est le taux de retrait (ou individus retirés) de la population \mathbf{X} ou \mathbf{Y} ou \mathbf{Z}_1

μ' est le taux de retrait (ou individus retirés) de la population \mathbf{Z}_2

δ est le taux de naissance dans la population \mathbf{X}

ν est le taux d'immigration dans la population \mathbf{Y}

ρ est le taux d'immigration dans la population \mathbf{Z}_2 .

Dans ce modèle, le nombre d'apparitions de nouveaux cas infectés pendant cet intervalle de temps Δt est ici mesuré par

$$\alpha.X(t).Y(t).\Delta t + \nu.\Delta t - (\kappa + \mu + \beta).Y(t).\Delta t.$$

On se propose de développer l'étude de chaque modèle en considérant d'abord la version déterministe de ces modèles et ensuite la version stochastique. Suivant alors Ishem, cette étude sera menée d'abord dans le cadre d'une population fermée et ensuite dans le cadre plus général d'une population ou la taille de la population est sensible a des facteurs externes (mort, émigration et immigration), auquel cas la variable $n(t) = X(t) + Y(t)$ (la taille de la population à l'instant t) est une variable aléatoire dynamique.

0.2 Le modèle simple à deux types

Il s'agit du modèle simple à une seule sous-population d'individus infectieux. Ce modèle est décrit par le diagramme (1).

Le modèle déterministe

Dans un intervalle de temps $[t, t + \Delta t]$ la taille de la population X des susceptibles subit une variation égale

$$X(t + \Delta t) - X(t) = -\alpha.X(t).Y(t).\Delta t + \delta.\Delta t - \mu.X(t).\Delta t$$

et celle des infectieux une variation égale a

$$Y(t + \Delta t) - X(t) = \alpha.X(t).Y(t).\Delta t + \nu.\Delta t - \mu'.Y(t).\Delta t$$

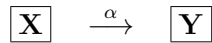
Et par suite, les équations différentielles régissant ce modèle s'écrivent

$$(1) \quad \begin{cases} \frac{dx(t)}{dt} = -\alpha.x(t).y(t) - \mu.x(t) + \delta \\ \frac{dy(t)}{dt} = \alpha.x(t).y(t) - \mu'.y(t) + \nu \end{cases}$$

sous la condition initiale $y(0) = m$ et $x(0) + y(0) = n$. Ce système est non linéaire et on peut, sous la condition qu'à tout instant $t > 0 : x(t) + y(t) = n(t)$, le réécrire sous la forme :

$$\begin{cases} \frac{dx(t)}{dt} = -\alpha \cdot x(t) \cdot (n(t) - x(t)) - \mu \cdot x(t) + \delta \\ \frac{dy(t)}{dt} = \alpha \cdot x(t) \cdot y(t) - \mu' \cdot y(t) + \nu \end{cases}$$

On commence par examiner le cas particulier où la population est isolée (i.e. $\mu = \mu' = \delta = \nu = 0$), et le modèle est illustré par le schéma



Cette hypothèse valide au début de l'épidémie, assure qu'en tout instant $t > 0$, on a $x(t) + y(t) = n(t) = n$. Dans ce cas, le système d'équations différentielles ci-dessus se réduit à un signe près à une seule équation différentielle non linéaire et nous prenons par exemple l'équation :

$$\frac{dy(t)}{dt} = \alpha \cdot [n - y(t)] \cdot y(t).$$

admettant pour solution particulière $y(t) = n$. En considérant ensuite le changement de variable $y(t) = n + \frac{1}{w(t)}$, l'équation (2) se ramène à une équation différentielle linéaire :

$$\frac{dw(t)}{dt} = \alpha \cdot n \cdot w(t) + \alpha,$$

dont les solutions sont fournies par

$$w(t) = c_0 \cdot \exp(n\alpha t) - \frac{1}{n} = \frac{y(0)}{n[y(0) - n]} \cdot \exp(n\alpha t) - \frac{1}{n}.$$

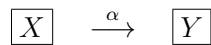
D'où nous déduisons la solution en $y(t)$:

$$y(t) = \frac{n \cdot y(0)}{y(0) - (y(0) - n) \cdot \exp(-n\alpha t)}.$$

Au début de l'épidémie, on peut admettre que $x(0) \approx n$, aussi la solution précédente peut-être approchée par

$$y(t) \approx y(0) \cdot \exp(n\alpha t).$$

Si maintenant on suppose que la population des susceptibles est fermé et celle des infectieux est soumise à des départs seulement (mort ou émigration), situation reflétée par le diagramme



$\downarrow \mu'$

alors le système d'équations différentielles régissant ce modèle devient dans ce cas

$$(1) \quad \begin{cases} \frac{dx(t)}{dt} = -\alpha \cdot x(t) \cdot y(t) \\ \frac{dy(t)}{dt} = \alpha \cdot x(t) \cdot y(t) - \mu' \cdot y(t) \end{cases}$$

sous la condition initiale $x(0) + y(0) = n(0) = n$.

Ce système est non linéaire et peut-être réécrit sous la forme :

$$\begin{cases} \frac{dx(t)}{dt} = -\alpha \cdot x(t) \cdot (n(t) - x(t)) \\ \frac{dy(t)}{dt} = \alpha \cdot (n(t) - y(t)) \cdot y(t) - \mu' \cdot y(t) \end{cases}$$

où $n(t) = x(t) + y(t)$, en tout instant $t > 0$. Ici la variable $n(t)$ est une variable dynamique fonction du temps t . Si on suppose que dans un intervalle de temps $[t, t + dt]$ suffisamment petit, la taille $y(u)$ de la population des infectieux en un instant $u \in [t, t + dt]$ est constante à $y(t)$ alors on peut montrer le résultat suivant :

Proposition 1 *En tout instant $t > 0$, on a*

$$n(t) = n - \mu' \cdot \int_0^t y(u) du.$$

Et en particulier, si t est petit, on a

$$n(t) = n - \mu' t \cdot y(0).$$

L'expression analytique de $n(t)$ ne nous facilite pas pour autant la résolution de notre système d'équations différentielles. Cependant au tout début de l'épidémie, on peut admettre que $x(0) \simeq n$, et dans ce cas, on peut produire une solution explicite du système

$$\begin{cases} x(t) \simeq x(0) \exp \left[\frac{y(0)}{n - \rho} (1 - e^{\alpha(n-\rho)t}) \right] \\ y(t) \simeq y(0) \cdot \exp \{ \alpha \cdot (n - \rho) t \} \end{cases}.$$

Dans la situation où l'épidémie est au delà de ses premiers stades d'évolution, la dynamique de celle-ci peut-être expliqué par l'introduction d'une variable auxiliaire $Z(t)$. La variable auxiliaire Z compte le nombre d'individus retirés de la population des infectieux à tout instant t ($t > 0$). Avec ces notations, on a

$$n(t) = n - z(t)$$

où $z(t)$ est le nombre d'individus retirés de la population Y . Si on note Z la population des infectieux retirés, le système précédent peut être complété par une troisième équation

$$\frac{dz(t)}{dt} = \mu' \cdot y(t).$$

avec la condition initiale $z(0) = 0$, et on note $\rho = \frac{\mu'}{\alpha}$ le taux relatif d'inactivité.
 La première équation montre que la fonction $x(\cdot)$ est une fonction décroissante de t ,
 et la seconde équation qui s'écrit

$$\frac{dy(t)}{dt} = \alpha y(t) \cdot [x(t) - \rho] \quad (2')$$

montre que si $x(0) \leq \rho$, alors la populations des infectieux décroît et nous assistons à une extinction rapide de l'épidémie; et si $x(0) > \rho$, nous assistons au début de la propagation de l'épidémie. Aussi, le nombre minimal de susceptibles permettant la propagation de l'épidémie est $x(0) = \rho$ et ce nombre est appelé seuil de propagation. Le rapport des équation 1) et 3) de ce système, nous fournit une solution en x fonction de z :

$$x(t) = x(0) \cdot \exp\left(-\frac{z(t)}{\rho}\right).$$

D'autre part, compte tenu de la relation $n(t) = n - z(t)$, la dernière équation de ce système devient

$$\frac{dz(t)}{dt} = \mu' \left[n - z(t) - x(0) \exp\left(-\frac{z(t)}{\rho}\right) \right] \quad (3')$$

Proposition 2 Si $\frac{z(t)}{\rho}$ est petit, on obtient pour solution de l'équation (3') :

$$z(t) = \frac{\rho^2}{x(0)} \left[\frac{x(0)}{\rho} - 1 + \lambda \cdot \tanh\left(\frac{1}{2}\lambda\mu't - \phi\right) \right]$$

où $\lambda = \sqrt{\left(\frac{x(0)}{\rho} - 1\right)^2 + \frac{2x(0)y(0)}{\rho^2}}$ et $\phi = \arg \tanh\left[\frac{1}{\lambda}\left(\frac{x(0)}{\rho} - 1\right)\right]$.

A l'aide de cette solution on peut réécrire $\frac{dz(t)}{dt}$ sous la forme :

$$\frac{dz(t)}{dt} = \frac{\mu'\rho^2\lambda^2}{2x(0)} \cdot \frac{1}{\cosh^2\left(\frac{1}{2}\lambda\mu't - \phi\right)} \quad (3')$$

ce qui montre que le graphe de $\frac{dz(t)}{dt}$ est symétrique par rapport à $t = \frac{2\phi}{\lambda\mu'}$. (graphie)
 Et nous déduisons directement de (3) et (3') une solution de $y(t)$:

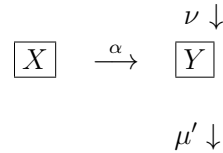
$$y(t) = \frac{\rho^2\lambda^2}{2x(0)} \cdot \frac{1}{\cosh^2\left(\frac{1}{2}\lambda\mu't - \phi\right)},$$

et aussi une solution en $x(t)$:

$$x(t) = x(0) \cdot \exp\left[-\frac{1}{\rho} \frac{\rho^2}{x(0)} \left(\frac{x(0)}{\rho} - 1 + \lambda \cdot \tanh\left(\frac{1}{2}\lambda\mu't - \phi\right)\right)\right].$$

En conclusion, la variable auxillaire apporte une information certaine sur l'évolution de l'épidémie à travers les individus inactifs. Ce résultat est conforté par le fait que les solutions trouvées sont fonctions des taux d'infection et de retrait α et μ' . D'autre part, à la limite quand $t \rightarrow +\infty$, la variable $z(t) \rightarrow \frac{\rho^2}{x(0)} \left(\frac{x(0)}{\rho} - 1 + \lambda \right)$ et si $\frac{2x(0) \cdot y(0)}{\rho^2} \ll \left(\frac{x(0)}{\rho} - 1 \right)^2$ et si l'épidémie se développe (i.e. $x(0) > \rho$), alors $z(t) \rightarrow 2\rho \left(1 - \frac{\rho}{x(0)} \right)$, et si on pose $x(0) = \rho + \sigma$ avec $\frac{\sigma}{\rho}$ petit, alors $z(t) \rightarrow 2\sigma$ et $x(t) \rightarrow \rho - \sigma$ lorsque $t \rightarrow +\infty$. En d'autres termes, la dynamique de l'épidémie est traduite à la limite, par 2σ susceptibles devenus infectieux auxquels correspondent 2σ infectieux retirés.

On considère maintenant la situation où il y a immigration vers la population des infectieux. Le modèle est représenté par le schéma



Les équations régissant ce modèle sont dans ce cas :

$$\begin{cases} \frac{dx(t)}{dt} = -\alpha \cdot x(t) \cdot y(t) \\ \frac{dy(t)}{dt} = \alpha \cdot x(t) \cdot y(t) - \mu' \cdot y(t) + \nu \end{cases}$$

avec la condition initiale $x(0) + y(0) = n(0) = n$. Là également, la taille de la population fluctue en fonction du taux de retrait μ' et du taux d'immigration ν . Et suivant le développement ci-dessus, au début de l'épidémie on peut exhiber des solutions du système et on trouve

$$\begin{cases} x(t) \simeq x(0) \cdot \exp\left(-\frac{\nu}{\alpha}\right) \cdot \exp\left\{\frac{\nu t}{x(0) - \rho} + \frac{\nu}{\alpha} \cdot \exp[\alpha(x(0) - \rho)t]\right\} \\ y(t) \simeq \frac{-\nu}{\alpha(x(0) - \rho)} + \left[y(0) + \frac{\nu}{\alpha(x(0) - \rho)}\right] \cdot \exp\{\alpha(x(0) - \rho)t\} \end{cases} \quad .?$$

Lorsque l'épidémie est à un stade d'évolution avancé, nous introduisons de nouveau la variable auxillaire $Z(t)$. Celle-ci satisfait l'équation

$$\frac{dz(t)}{dt} = \mu' \cdot y(t).$$

avec la condition initiale $z(0) = 0$. Cette fois-ci, on a

$$\frac{dy(t)}{dt} = \alpha y(t) \cdot [x(t) - \rho] + \nu \quad (2')$$

ce qui montre que l'épidémie a tendance à se propager rapidement si $x(0) > \rho$ et d'ailleurs, elle ne risque pas de s'éteindre si $x(0) \leq \rho$ si bien sûr le taux d'immigration ν est important.

Et de la même manière que ci-dessus, en observant cette fois-ci que

$$\frac{dz(t)}{dt} = \mu' \left[n - z(t) + \nu - x(0) \exp\left(-\frac{z(t)}{\rho}\right) \right]$$

et sous l'hypothèse que $\frac{z(t)}{\rho}$ est petit, on obtient comme équation approchée

$$\frac{dz(t)}{dt} = \mu' \cdot \frac{\rho^2 \lambda^2}{2x(0)} \left\{ 1 - \frac{x^2(0)}{\rho^4 \lambda^2} \left[z(t) - \frac{\rho^2}{x(0)} \left(\frac{x(0)}{\rho} - 1 \right) \right]^2 \right\}$$

avec $\lambda = \sqrt{\frac{2x(0)(y(0) + \nu)}{\rho^2} + \left(\frac{x(0)}{\rho} - 1 \right)^2}$ et dont la solution est donnée par

$$z(t) = \frac{\rho^2}{x(0)} \left[\frac{x(0)}{\rho} - 1 + \lambda \cdot \tanh\left(\frac{1}{2}\lambda\mu't - \phi\right) \right]$$

avec la constante ϕ la même que celle de la proposition 1. De même, on dérive de cette solution des solutions approchées de $x(t)$ et $y(t)$, on trouve

$$x(t) = x(0) \cdot \exp\left[-\frac{1}{\rho} \frac{\rho^2}{x(0)} \left(\frac{x(0)}{\rho} - 1 + \lambda \cdot \tanh\left(\frac{1}{2}\lambda\mu't - \phi\right) \right)\right] \text{ et } y(t) = \frac{\rho^2 \lambda^2}{2x(0)} \cdot \frac{1}{\cosh^2\left(\frac{1}{2}\lambda\mu't - \phi\right)}$$

avec λ spécifié ci-dessus.

La dépendance de la solution $x(t)$ par rapport à ν est conforté par la liaison entre les équations en $x(t)$ et $y(t)$.

Sur un Algorithme Génétique Spatial, pour conception d'échantillonnage spatial optimal avec quantification

S. AIT KACI AZZOU et K. BOUKHETALA

*Département de Probabilités et Statistique
Bp.32, El-Alia, USTHB, Bab-Ezzouar, Alger,
Algérie*

[sadoun_ait_kaci@hotmail](mailto:sadoun_ait_kaci@hotmail.com)

kboukhetala@usthb.dz

Key words: Spatial sampling, Genetic Algorithm, Quantization, Localization .

Abstract : A Spatial Genetic Algorithm (SGA) is proposed to design a finite optimal spatial sampling. A geostatistical application is studied for a problem of optimal localization of meteorological stations. The quantization is used to improve the performance of the SGA.

1. Introduction

Optimal sampling of estimation of random quantities is an actual problem for practical and theoretical statistical considerations. For example, the real situations where the observed quantities are accompanied by measurement errors.

Whenever errors are correlated, both of sampling problem and methodologies of global search have been studied. In scalar case, this problem has been discussed by [1] using the heuristics methods, like genetic algorithms and annealing search algorithms. For a finite sampling, we have confirmed in [4] the effect of quantization over convergence speed of used algorithms. For a large sample, [2] has proved that one may reduce the rate convergence of mean quadratic error from n^{-2} to $n^{-1.5}$. In [3] we have generalized the previous studies in

spatial case, by developing an adapted search algorithm. In our work, we propose a generalization of studies of [4] that concern spatial optimal sampling with quantization, by developing, in spatial case, an adequate global search algorithm. The proposed algorithm is a *Spatial Genetic Algorithm* that is powerful heuristic for the solution of hard combinatorial problems. We describe two new crossover operators called: global and local crossover operator. This type of crossover explores the spatial aspect and gives better results than [3]. On the other hand, we show that the quantization of spatially correlated errors improve clearly the quality of results for a small number of sampling points (k), but gives bad results when $k \rightarrow \infty$ which is theoretically proved by [2].

We may conclude that the quantization has a positive effect with the proposed *spatial genetic algorithm* for a small number of sampling points, that presents a practice interest for different problems encounter in hydrological and environmental sciences.

2 - Spatial Genetic Algorithm

The Genetic Algorithms (GA) represent an enough rich and very interesting family of stochastic optimisation algorithms, are founded on mechanisms of the natural selection and genetics. Fields of application are very varied; theory of the picture compression, automatic programming and others. The principle of these algorithms is to effectuate a stochastic research on an important space through "a population" of Pseudo- Solutions. These algorithms are simple and very efficient. Besides, they are not like at hypotheses on the function to optimise, as the continuity and the differentiability. They operated directly one the explored space, after a coding of parameters.

They processes on a population of points, instead of an unique point, use values of the studied function only, without another auxiliary knowledge and rules of transition probabilistic.

In order to well to fear the genetic algorithm dynamics, let's introduce the main elements of the jargon, used in the literature.

- *Individual or Chromosome: a potential solution.*
- *Population: a set of chromosomes or of points of the research space.*
- *Environment: the space of research;*
- *Fitness or Function of assessment: the positive function that one seek to maximize.*

The working of the GA is based on different basis operators. These operators are inspired directly of the natural selection mechanism and of the genetic phenomenon. It consists to evolve the population in order to adapt individuals to the environment. Technically, a new generation generally gets himself at the end of a cycle composed of three main standard operators, namely, operators of reproduction, crossover, and mutation. For a reason of implementation adapted to computers, a chromosomal representation or coding of individuals (feasible solutions) is necessary. We propose in the case of our problem:

- A spatial coding of the solution,
- Two operators of crossover and mutations appropriated.

What permits us to propose a new Genetic Algorithm adapted to the Spatial case, denoted *SGA*.

2.1. Coding of the Solution (CS)

One considers a spatial n - sample, in a space of research, of dimension $L < n$, that is a matrix of real, of the following form:

$$X = \begin{bmatrix} x_{11} & x_{12} & \dots & x_{1L} \\ \cdot & & \dots & \cdot \\ \cdot & & \dots & \cdot \\ x_{n1} & x_{n2} & \dots & x_{nL} \end{bmatrix}$$

Thus, a population of solution in this space is a set of real matrixes $M_{(n \times L)}$. We adept coding relatively to a matrix; it is given for individual i , by a vector of real of size $n * L$:

$$X_i = ((x_{11}, \dots, x_{1k})_i, (x_{21}, \dots, x_{2L})_i, \dots, (x_{n1}, \dots, x_{nL})_i) \cdot$$

2.1.1. Operator of Crossover (OC)

One consider two coded real configurations:

$$X_1 = (x_1, \dots, x_L)_1 \quad \text{and} \quad X_2 = (x_1, \dots, x_L)_2 \cdot$$

A classic crossover, that we call global crossover produced of the new configurations thereafter. The uniform crossover corresponds, here, to interchange at random, a subset of points of \mathbb{R}^L . Without losing generality, we consider an example for the case of the plan, in order to illustrate the process of crossover. For two parents X_p and Y_p , such that:

$$X_p := (x_1, x_2, \dots, x_7) = ((x_1^1, x_1^2), (x_2^1, x_2^2), \dots, (x_7^1, x_7^2))$$

$$Y_p := (x_a, x_b, x_c, x_d, x_e, x_f, x_g) = ((x_a^1, x_a^2), (x_b^1, x_b^2), \dots, (x_g^1, x_g^2)),$$

an uniform crossover products two individuals children, X_e and Y_e , with a appropriate schema. We has a difficulty, when we applied this operators with a low probability for the mutation, in accordance with the theory of the AG. What gives a little changes in the diversity of the individuals population, relatively to the initial population. On the other hand, the defined crossover operator has for simple role to changes indication of components. A first idea was to increase the probability of mutation to maintain a diversity being sufficient and so to preserve the improvement. The works of Jong clearly demonstrate that this method is not efficient. An uncontrolled increase of the mutation rate can damage the performance of the genetic algorithm. Therefore, we introduced a new operator, says local crossover, that acts locally on couples of two individual components to cross.

2.1.2. Operator of Local Crossover (OLC)

From a couple (X_1, X_2) of individuals, we choose a sample (x_i^1, x_i^2) . We proceed as following to define the OLC operator, by generating a new sample $(x_i^1, x_i^2) \in [a, b] \times [a, b]$, with:

$$\begin{cases} x_i^1 = P_E \left(\frac{(b-a)}{pas} \times U \right) \times pas + pas / 2; \\ x_i^2 = x_i^1 + x_i^2 - x_i^1 \end{cases}$$

for

$$\begin{aligned} a &= \max(x_i^1, x_i^2) - \min(x_i^1, x_i^2) \\ b &= \min(x_i^1 + x_i^2, Sup) \end{aligned}$$

The quantities, denoted by pas , P_E et U are respectively the increment of discretization, integer part and a random value, taken from $[0,1]$.

At the time of the research of optimal sample, it can contain two identical samples, because the combinatorial nature of the problem. For it, one introduces an operator of diversity that is defined as following, by increasing or creasing of the gene:
we set

$$\begin{cases} x'_i = x_i \text{ Sgn } pas, \\ y'_i = y_i \text{ Sgn } pas \end{cases}, \quad (x_k, y_k) \in [a_1, b_1] \times [c_1, d_1].$$

- Sgn is the function, defined by:

$$\text{Sgn} = \begin{cases} + & \text{if } (x_i = a_1 \text{ or } y_i = c_1) \\ - & \text{if } (x_i = b_1 \text{ or } y_i = d_1) \\ (+ \text{ or } -) & \text{with probability } \frac{1}{2} \end{cases}$$

Finally, local crossover is for our problem, an adaptation very interesting of the crossover classic multi-site on chromosomes produced by the concatenation of binary code.

2.1.3. Operator of Mutation (OM)

The operator of mutation that we propose to our spatial sampling problem is defined as following:
We operate on a gene $S_k = (x_k, y_k)$, to mute it to a new gene $S'_k = (x'_k, y'_k)$, as following:

$$\begin{aligned} &\begin{cases} x'_k = x_k \text{ Sgn } U \\ y'_k = y_k \text{ Sgn } V \end{cases} \\ \text{such that: } &U \in \begin{cases} P_E \left(\frac{U_{[x_k, b_1]}}{pas} \right) \times pas & \text{if } x_k \geq \frac{a_1 + b_1}{2} \\ P_E \left(\frac{U_{[a_1, x_k]}}{pas} \right) \times pas & \text{if } x_k < \frac{a_1 + b_1}{2} \end{cases} \\ &V \in \begin{cases} U_{[y_k, d_1]} & \text{si } y_k \geq \frac{c_1 + d_1}{2} \\ U_{[c_1, y_k]} & \text{si } y_k < \frac{c_1 + d_1}{2} \end{cases} \end{aligned}$$

and $U_{[a_1, b_1]}$ is a random value from $[a_1, b_1]$.

2.2. Stopping Criteria (SC)

The number of iterations is fixed , after some tests on the fitness function. It satisfies a compromise between constraints of convergence of the population, the CPU time and the precision.

Therefore we are now able to propose the following Spatial Genetic Algorithm :

SGA algorithm:

Begin

Stage 1 - Function of fitness

Stage 2 - Coding: CS

Stage 3 - Generation of the feasible solutions set

Stage 4 -

Do

OC, OLC and OM

until SC.

End

3- Application

The goal is the installation of a rainfall optimal network, that permits to observe a quantity M of rain data taken on a region D with a reduces number of stations k.

We have :

$$D = \left\{ (x, y, z) \mid x \in [540, 800], y \in [150, 400], z \in [456, 1250] \right\}$$

- $s=(x, y, z)$ is a co-ordinate of the rainfall site in the region D , with z is the altitude.

The rainfall annual data, obtained from the domain D, permit to give the following regression model, relatively to the observed data of a site $S_i = (x_i, y_i, z_i)$:

$$Q_{S_i} = \langle \tilde{\alpha}, S_i \rangle + d + m \cdot \mathcal{E}(S_i) \quad (\text{millimetre}).$$

$$\tilde{\alpha} = \begin{pmatrix} a \\ b \\ c \end{pmatrix}, \text{ d and m are some real parameters. The parameters a, b, c and d are estimated by :}$$

$$\hat{a} = 60.05 \cdot 10^{-3} \quad \hat{b} = 0.701 \quad \hat{c} = 0.29 \quad \hat{d} = -94.44.$$

with a value of $R^2 = 0.89$.

The observations errors are modelled by spatial random process \mathcal{E}_s , by taking in consideration the characteristic of variability, uncertainty and heterogeneity of the real rainfall process. We suppose that the spatial random process modelling the errors associated to each site. It is a spatial homogenous Gaussian process, with mean $\mu_{\mathcal{E}}$ and covariance $C(\mathcal{E}_i, \mathcal{E}_j)$. Therefore, to determine the optimal sites comes back to find the optimal sample of

finite size k, relatively to a optimality criteria: $E(\bar{Q} - M)^2$, with:

$$\bar{Q} : \text{ is the mean quantity of rainfall of the } k \text{ sites } \left(\bar{Q} = \frac{1}{k} \sum_{i=1}^k Q_{S_i} \right)$$

3.1 - Mathematical formulation of the Problem

Let $S = \{s_i \in D \subset \mathbb{R}^3, i=1..k\}$, be a set of sample of sites of size $k=|S| \leq n$, where n is the size of the domain D . The optimality criteria is given by :

$$\sigma^2(S^* / \mathcal{E}) = \underset{S \in D}{\text{Min}} \left\{ E \left(\sum_{i=1}^k \frac{Q(s_i)}{k} - M \right)^2 \right\}$$

k : Size of S .

S : Set of sample of size k .

Q_{S_i} : Quantity of rainfall observed on site s_i .

$s_i^* = (x_i^*, y_i^*, z_i^*) \quad i=1, \dots, k$: coordinates of optimal spatial sample, S^*

$\mathcal{E} = (\mathcal{E}(s_i))_{i=1, \dots, k}$: Vector of correlated observations errors in each site, with mean zero and covariance function $\text{Cov}(s_i, s_j)$ of a Gaussian spatial process, of analytical type.

$E(.)$: mathematical mean defined relatively to the law of \mathcal{E} .

3.2. Function of fitness

We have :

$$\sigma^2(S^*; \mathcal{E}) = \text{Min} \left\{ \frac{1}{k^2} \left[(Cst)^2 + m^2 \cdot \sum_{i=1}^k \sum_{j=1}^k \text{Cov}(\mathcal{E}(s_i), \mathcal{E}(s_j)) \right] \right\} \dots\dots\dots (I)$$

such that : $\{s_i \in S : i=1, \dots, k\}$ et $\{s_j \in S : j=1, \dots, k\}$ et

$$Cst = \hat{a} \sum_{i=1}^k x_i + \hat{b} \sum_{i=1}^k y_i + \hat{c} \sum_{i=1}^k z_i + k \cdot \hat{d} - k \cdot M$$

We applied our SGA algorithm for following choice of parameters, that are fixed by the expert:

M : mean quantity of rainfall : 337.9 mm.

- k : number of rainfall stations : 5.
- Selection by ranging (Pressure of selection : 1.4).
- Crossover : OC and OLC are used, with adequate probabilities P_{OC} and P_{OLC} , receptively.
- Mutation: OM is used with probability adequate probability P_{OM} .
- Spatial Covariance $\text{Cov}(s_i, s_j) = \exp(-\theta \|s_i - s_j\|^2)$, with $\theta = 0.05$.

The initial choice of individual has an effect on the precision of solution. A positive effect, when this choice is a bad individual. For $\sigma^2(S^* / \mathcal{E}) = 0,62$, the corresponding optimal spatial configuration is given by the figure 3.

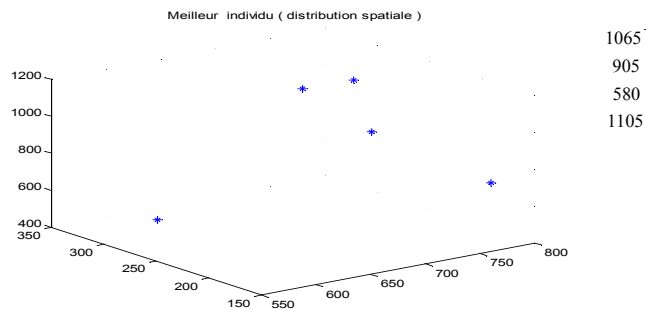


Figure 3: Optimal spatial Localization, obtained by the SGA.

The following table give the obtained results of the SGA, for optimum parameters and relatively to a regular sampling, by a repetition of 80 times of the SGA.

| P_{CG} | P_{CL} | Absolute Error (AE): $ \bar{Q} - M $ | Confi. Interval of (AE) |
|----------|----------|--|-------------------------|
| 0.9 | 0.01 | 4.43, 3.98, 0.52, 1.81, 2.02, 1.4, 0.3, 1.26, 0.76, 0.11, 5.10, 2.41, 3.46, 9.07, 1.81, 0.651, 0.34, 0.41, 2.19, 0.00, 0.94, 0.75, 5.17, 2.80, 7.14, 2.79, 0.19, 1.78, 0.03, 0.19, 8.28, 0.63, 1.44, 7.59, 8.84, 1.42, 2.14, 0.58, 4.09, 2.46, 7.54, 1.42, 4.62, 2.02, 4.14, 0.39, 11.85, 0.11, 0.26, 1.11, 0.56, 1.81, 2.39, 0.84, 2.01, 0.01, 0.98, 1.58, 7.71, 3.28, 1.07, 2.16, 0.07, 0.21, 0.95, 4.56, 0.92, 0.03, 0.43, 1.22, 1.46, 13.74, 7.81, 1.27, 5.51, 1.64, 0.90, 5.31, 1.18, 0.99. | [4.02, 9.96] |

4. Quantization

The use of a quantization operator reduce variability of the estimator of \bar{Q} and improve performance of the SGA, as it is shown by the following table:

| P_{CG} | P_{CL} | Absolute Error (AE): $ \bar{Q} - M $ | Confi. Interval of (AE) |
|----------|----------|---|-------------------------|
| 0.9 | 0.01 | 0.35 2093 1.31 0.22 2.57 0.7 0.56 0.26 0.43 0.61 0.56 2.38 5.58 0.93 5.53 2.70 0.32 0.80 0.441 0.47 6.63 1.95 1.03 1.17 1.67 | [2.43, 5.79] |

5. Implementation

The SGA is implemented by the MATLAB language. The object procedure of this language are used in generation of Gaussian spatial measure errors and by simulation operators *OC*, *OLC* and *OM*.

6. Conclusion and comments

A Spatial Genetic Algorithm is proposed. The obtained results of sampling are improved by quantization of errors, for small size of sample. It justifies itself by the fact that the quantization has an effect on the regularity of function fitness, by creating a configuration of picks, desirable for an algorithm of type genetic. The perspective works will be devoted to improve parameters and operators of Spatial Genetic Algorithm, in order to decrease the size of errors and to obtain realistic confidence intervals more adapted to practical applications.

7. References

- [1] K.Boukhetala, K.Benhenni et S.Benamara (1996). *Optimal Sampling for estimating integral of function from observations with correlated measurement errors*, Compstat'96, Edition UPC, A. Prat and E. Ripoll, pp 161-162.
- [2] K. Benhenni and Stamatis Cambanis (1998). *The effect of quantization on the performance of sampling design*. *IEEE transactions on information theory* Vol 44, N°5.
- [3] K.Boukhetala et Mehassouel Nadjib (2000) . *Optimal Spatial Sampling for an estimating problem, based on correlated observations*, Compstat'2000, edition Statistics Netherlands.
- [4] K.Boukhetala et Habib Fatma Zohra (2002) . *Finite Sampling design with quantization for a pharmacokinetic problem*. The 6th WMSCI, Computer Science, Vol XI, edition IIIS, USA, pp. 167-170.

ON THE USE OF ARTIFICIAL INTELLIGENT METHODS FOR PREDICTION PURPOSES IN NUCLEAR RESEARCH REACTORS

Hakim MAZROU* and Mohamed HAMADOUCHE[‡]

* Laboratoire des Analyses de Sûreté – Centre de Recherche Nucléaire d'Alger (CRNA).
02 Boulevard Frantz – Fanon, B.P. 399, 16000 Alger, Algérie. mazrou_h@comena-dz.org

[‡] Département de mathématiques appliquées, Faculté des Sciences, Université de Blida.

ABSTRACT

This paper reports on the use of Artificial Neural Networks (ANNs) in predicting core safety parameters, such as multiplication factor K_{eff} and fuel powers peaks P_{max} , in Light Water Research Reactors (LWRRs) using a personal computer. The purpose of this emulation, when it's associated with optimal in-core fuel management, is to allow rapid and extensive exploration of potential good configurations of fuel in LWRRs cores.

The key idea of the developed program lies on the use of an adaptive learning rate procedure in a typical back-propagation error model. The performance of the algorithm was improved to the proper setting of the learning rate, which is adjusted according to the minimal error predicted.

Using the 2Dimensional neutronic diffusion code MUDICO-2D for the referenced target values, the neural network development process is reported on a slight modified version of the IAEA 10Mw benchmark LEU core.

Key Words: Artificial Neural Networks, learning rate, back-propagation error, multiplication factor, fuel powers peaks, benchmark LEU core.

1. INTRODUCTION

The current status of emulating the calculations with artificial neural networks (ANN's), in the field of nuclear technology, has been extensively explored in recent years, with some degree of success [1-2], throughout a wide variety of national projects at various specialized laboratories over the world. Thus, for complex tasks like those associated with in-core nuclear fuel management optimization problem where a huge number of possible arrangement exist in the available search space, the use of Artificial Intelligence (AI) is in growing need. And nowadays the automatic resort to such intelligent methods is intensively recommended, that's, in one part, to reduce drastically the computation time usually needed in such optimization process and, in another part, to offer extensive exploration to find potential good configurations.

The feasibility of this application has been demonstrated, that's why, an attempt was made by our laboratory [3-4], to use ANN's to emulate adequately the results given by the developed neutronic computer code MUDICO-2D [5] in small computational system.

This work was part of a project, accomplished with the aim of using this application, as a computational support, to provide preliminary design of an optimal fuel configuration rearrangement in light water research reactors. The goodness of the proposed optimal solution is based on prediction of two parameters of interest in safety view point. It concern, the multiplication factor K_{eff} and the fuel powers peaks P_{max} .

Traditionally, the direct calculation of these safety core parameters is done by solving numerically neutron diffusion equations. However, as the global optimization process takes prohibitive time, the use of an adaptive back-propagation neural network method is proposed in the following paper as an alternative solution to the typical diffusion techniques commonly used to evaluate these parameters.

To reach this purpose, we have developed an application on the IAEA 10Mw benchmark LEU (Low Enriched Uranium) core [6], in which a slight modification was introduced in its initial core arrangement, in order, to ensure geometrical asymmetry in the core and thereafter to obtain the sufficient search space.

2. DESCRIPTION OF THE WORK

In the present work, the investigated global strategy followed was based upon the combination use of the well-known Back-Propagation (BP) algorithm with a special adaptive learning rate procedure, in conjunction with an appropriate selection of configurations for the training data-base. The motivation in using such computational procedure lies in the fact that it will let us to use merely hundreds of configurations rather than thousands, in the learning stage, as it's usually required in such typical calculations to ensure reasonable predictions. Hence, as shown in fig.1., a suitable neural networks development strategy was tested based on executing the following two main calculational stages, in a independent way : **Learning** and **prediction** stage.

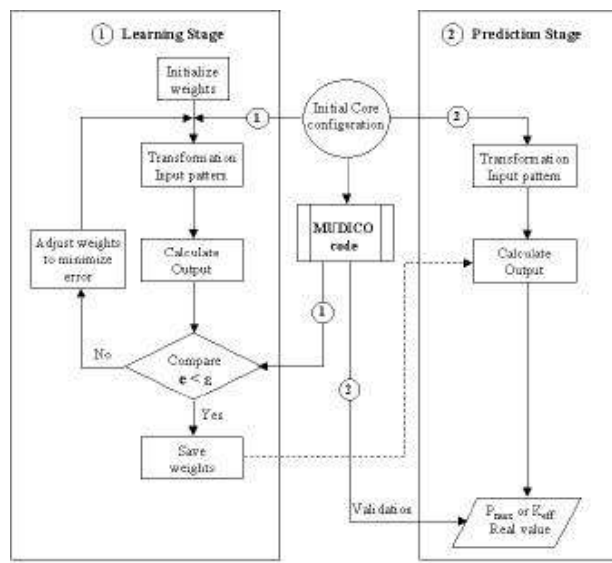


Fig.1. Overall Back-Propagation computational strategy for the core parameter prediction.

The first stage of computational procedure, consist of creating suitable networks by applying appropriate learning rule using a desired database. The information required in the related database will contain a coupled input values with the corresponding target output values. These values are used to train the networks until the error reach a desired value stated at the beginning of the learning process.

The second stage is the prediction one where the weights, from the inter-connected neurons, have been adjusted to the desired error in the previous calculations stage. These weights will be used, in a global computational sequence, to predict the networks outputs when an unseen data will be presented to the developed networks. This is the power of the network approach and one of the justified reasons for using it. This stage is necessary to test the performance of the developed neural network. Therefore, the main steps required in the process of developing neural network are as follow:

2.1. Create data-base for training:

The core of generic 10 Mw IAEA LEU benchmark research reactor was selected to create the needed database for network training. The core is an arrangement of 5x6 elements containing 21 standard MTR-type fuel elements and 4 control fuel element.

A total of 30 positions were considered. In addition, for better representation of various state of the reactor core during its useful life, several configurations have been taken into account and simulated using both standard and control fuel element with different burn-up, ranging from fresh fuel (FF) to 50% (FF, 5%, 10%, 25%, 30%, 45%, 50%) for the first one and from fresh fuel to 30% (FF, 25%, 30%) for the second one.

The whole reactor core is then transformed into a vector form of 30 positions so that it can be presented to the networks as input values to be treated.

Finally, using several established heuristics rules, initial core configurations have been selected according to the following two requirements:

1. Presence of fuel elements that influence both K_{eff} and P_{max} parameters;
2. Random generation of finite set of initial configurations.

The first criteria ensure an orientation in the research, whether the second is random and lets a diversity in the search space.

At the end, a total of 800 configurations were generated, among which 700 were devoted to create the training data-base, whether the remaining are considered to test the performance of the developed networks.

2.2. Construction of networks for training:

Two neural networks have been developed, based on typical three layers (input-hidden-output) model. The first one predict K_{eff} and the second one predicts P_{max} . They contain 30 neurons as input and one neuron as output. As for, the number of hidden neurons was set to 280 neurons, after sensitive calculations.

2.3. Choosing a learning function:

In the present work, the well named back-propagation algorithm with an adaptive learning rate procedure was implemented in the learning stage. This algorithm proceed with discrete jump, in which the step size, is adjusted according to the learning parameter. The choice of an adequate value of this parameter is quite difficult, since it determines the convergence of the computational procedure. If the learning rate is set too high, the algorithm may oscillate and become unstable. If the learning rate is too small, the algorithm will take too long to converge. It is not practical to determine the optimal setting for the learning rate before training, and, in fact, the optimal learning rate changes during the training process, as the algorithm moves across the performance error surface.

The performance of the steepest descent algorithm can be improved if we allow the learning rate to change during the training process. An adaptive learning rate will attempt to keep the learning step size as large as possible while keeping learning stable. The learning rate is made responsive to the complexity of the local error surface.

In this work, we use the *delta-bar-delta* rule [7], which consist of both a weight update and a learning rate update rule. The weight update rule is similar to the steepest descent algorithm with the sole exception that each weight possesses its own learning rate parameter. During the learning stage, the quadratic error E is minimized in the three-layered network according to the following formula:

$$\frac{\partial E}{\partial w_{ij}}=0 \quad (1)$$

Where w_{ij} is the j 'th connection weight of the i 'th neuron. Thus, adjusting the weights (w_{ij}) to minimize E at the $(t+1)$ -th step is performed as follow:

$$w_{ij}^{t+1}=w_{ij}^t-\eta_{ij}^{t+1}\frac{\partial E}{\partial w_{ij}} \quad (2)$$

Where in (2) η_{ij} is the learning parameter which is updated after each correction step according to a suitable Jacob's rule [8]:

$$\eta_{ij}^{t+1}=\eta_{ij}^t+\Delta\eta_{ij}^t \quad (3)$$

$$\Delta\eta_{ij}^t=\begin{cases} \kappa & \text{if } S_{ij}^{t-1}\frac{\partial E_t}{\partial w_{ij}}>0 \\ -\phi\eta_{ij}^t & \text{if } S_{ij}^{t-1}\frac{\partial E_t}{\partial w_{ij}}<0 \\ 0 & \text{otherwise} \end{cases} \quad (4)$$

where:

$$S_{ij}^t=(1-\Theta)\frac{\partial E_t}{\partial w_{ij}}+\Theta S_{ij}^{t-1} \quad (5)$$

In these equations S^t , S^{t-1} is an exponential average of the current and past derivatives with Θ as the base and time as the exponent. κ , ϕ and Θ are small constant and positive values which are respectively 0.05, 0.3, 0.7. These values were fixed by Jacobs during his experiences. The initial value of steps η_j^0 is generally set to 0.1.

2.4. Train the networks:

Supervised training has been carried out for each parameter. To start the training process, all the weights (w_{ij}) have been initialized randomly between [- 0.5, +0.5] and adjusted, through back-propagation algorithm, to minimize the error E for 700 configurations. According to the convergence criteria, training was completed and weights were stored when both normalized error reached a value of 0.0007 for K_{eff} predictions and 0.003 for P_{max} predictions. This result in a fixed training of about 36 and 24 hours, respectively for both K_{eff} and P_{max} parameters, on a small computational system (300Mhz Pentium II PC).

2.5. Testing the networks:

After training, networks were created for each parameter and tested against the unseen data. One hundred configuration were devoted to this purpose. The prediction results for K_{eff} and P_{max} parameters is shown in figure 2 and 3. When the model was recalled, more than 95% of the training patterns were predicted within $\pm 0.7\%$ of the reference value for the K_{eff} , with a maximum and average error of 0.84% and 0.54% respectively. As for P_{max} , the network predicts 95% within $\pm 8.5\%$ of the reference value, with a maximum and average error of 14.0% and 7.8% respectively.

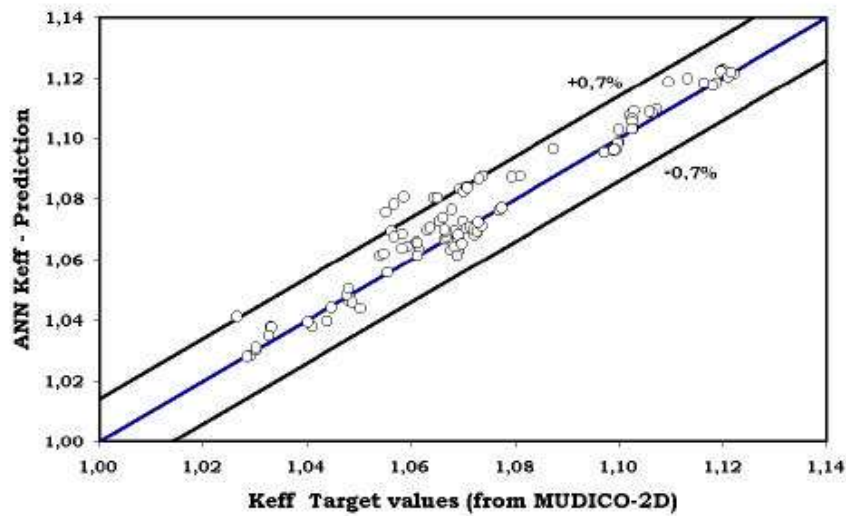


Fig. 2. K_{eff} prediction results.

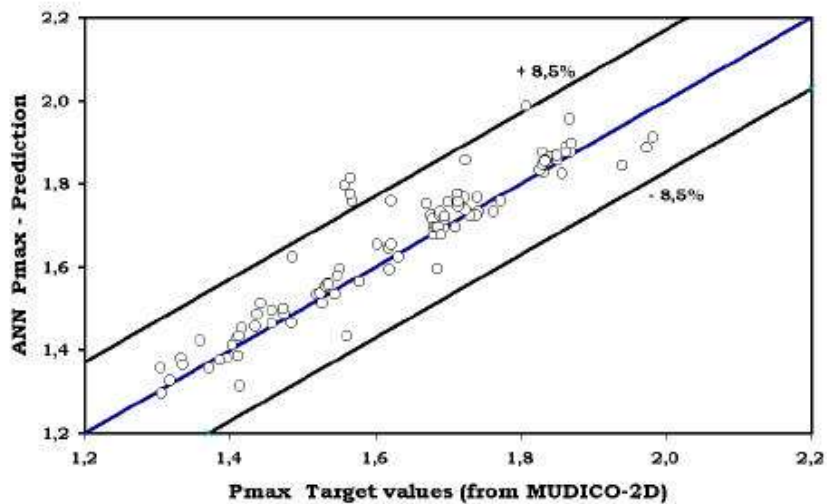


Fig. 3. P_{max} prediction results.

3. CONCLUDING REMARKS

Throughout the performance of the present work by using small computational system, the obtained results have demonstrated the ability to emulate, when needed for special applications like those associated with optimal in-core fuel management, sophisticated and cumbersome computer codes with reasonable error. Moreover, we take advantage and great benefit of using ANN, since it provide us a useful and very fast computational tool.

In addition, with regard to the strategy followed to develop the desired neural networks, the obtained results show reasonable predictions level. However, at this stage, the prediction accuracy is not sufficient yet, to emulate completely the referenced computer code. The loss of accuracy in final predictions is inevitable since we attempted to cover a very large operating period of the research reactors during its useful life.

Thus, one of the proposed way to investigate and explore in greater depth, in order, to enhance the quality of the results, should be oriented towards setting appropriate criteria in selecting core configurations needed to create the necessary database for the training stage. This will be our main future investigation.

REFERENCES

- [1] MILLER L. F., ALGUTIFAN A., and UHRIG R. E., "An Application of Neural Networks and Artificial Intelligence for In-Core Fuel Management ", TRANSACTIONS of American Nuclear Society and the European Nuclear Society, **Vol. 66**, 108-109, Chicago, IL, November 15-20 (1992).
- [2] KIM H. G., CHANG S. H., and LEE B. H., "Pressurized Water Reactor Core Parameter Prediction Using an Artificial Neural Network", Nucl. Sci. Eng.,**113**, 70-76 (1993).
- [3] MAZROU H. and HAMADOUCHE M., "Research Reactor Core Parameter Prediction Using an Adaptive Back-Propagation Neural Network Method", TRANSACTIONS of American Nuclear Society, **Vol. 88**, pp. 312-313, 2003.
- [4] MAZROU H. and HAMADOUCHE M., "Application of Artificial Neural Network for Safety Core Parameters Prediction in LWRRS", Accepted for publication in Progress in Nuclear Energy, 2003.
- [5] IBRAHIM K., MAZROU H., HAMIDOUCHE T. AND BENKHARFIA H., « MUDICO-2D: A Two-Dimensional Multigroupe Diffusion Code for Perturbation Calculation in Light Water Research Reactors ». Proceedings of International Conference on the New Frontiers of Nuclear Technology: Reactor Physics, Safety and High-Performance Computing. **PHYSOR 2002** October 7-10, 2002, Seoul, Korea, **ISBN 0-89448-672-1**, CD-ROM.
- [6] IAEA technical document: "Research Reactor Core Conversion Guidebook", **IAEA TECDOC-643**, 1992.
- [7] JODOUIN J. F. "Les Réseaux Neuromimétiques, Modèles et Applications", Edit. Hermès, Paris (1994).
- [8] JACOBS R. A., "Increased Rates of Convergence Through Learning Rate Adaptation", Neural Networks, **Vol.1**, pp. 295-307, (1988).

Analyse en Composantes Principales : Approche Neuronale Application : Analyse d'Images Optiques de Télédétection

Salim Chitroub

Laboratoire de Traitement du signal et Images
Faculté d'Electronique et d'Informatique, U. S. T. H. B.
B. P. 32, El-Alia, Bab-Ezzouar, 16111, Alger, ALGERIE
Fax: (213) (21) 24-76-07, Email: s_chitroub@hotmail.com

Résumé : L'application de l'Analyse en Composantes principales (ACP) pour l'amélioration et la compression des images de télédétection implique le calcul de la matrice de covariance des images d'entrée et de sa décomposition spectrale pour extraire leurs valeurs propres et vecteurs propres correspondants. Quand la taille de la scène imagée et/ou le nombre d'images d'entrée croît de manière significative, le calcul de la matrice de covariance et sa décomposition spectrale deviennent pratiquement inefficace et imprécis dus aux erreurs d'approximation. Ces déficiences rendent l'ACP inefficace pour cette application. Nous proposons ici un modèle de réseau de neurones qui effectue l'ACP directement des images originales d'entrée sans aucuns calculs non-neuronaux additionnels ou évaluations préliminaires des matrices. Le modèle est soigneusement conçu afin d'être indépendant de l'utilisateur autant que possible. Ceci inclut la conception de la topologie du réseau et son processus d'apprentissage. Son application a été réalisée sur une image de télédétection fournie par le satellite Landsat-Thematic Mapper. Les résultats obtenus prouvent l'efficacité du modèle. Une étude comparative entre le modèle proposé et la méthode statistique de l'ACP est effectuée. L'étude a montré que l'approche neuronale de l'ACP proposée est supérieure à la méthode statistique de l'ACP.

Mots Clefs : Réseaux de neurones ; Traitement d'images ; Télédétection ; Analyse en Composantes Principales ; Compression de données ; Suppression du bruit.

1. Introduction

Les images optiques de télédétection sont sujettes à des déformations et à des perturbations dues aux effets atmosphériques, aux vibrations des capteurs du système d'acquisition, et autres [1]. Ceci se traduit par un bruit additif qui se rajoute au contenu réel du pixel. Ces images sont corrélées et donc posent le problème de redondance de l'information. Si la taille de la scène imagée et/ ou le nombre d'images spectrales croît, l'espace mémoire pour le stockage des données devient important et le coût de traitement augmente [2]. Donc, une représentation optimale, dans laquelle le bruit est supprimé et le nombre des images est réduit, devient une tâche essentielle. L'Analyse en Composantes Principales (ACP) peut être appliquée pour atteindre cet objectif.

L'ACP statistique et ses variantes sont basées sur la décomposition spectrale de la matrice de covariance des images spectrales d'entrée et celle du bruit additif pour les variantes qui font appel à cette matrice dans leur concept [2], [3]. Ceci consiste en premier lieu à estimer la matrice de covariance des images d'entrée, et celle du bruit pour certaines variantes, puis le calcul de leurs vecteurs propres et valeurs propres afin de former la matrice de transformation linéaire. Cependant, dans la pratique, cette manière de mise en œuvre de l'ACP et de ses variantes souffre de la difficulté d'une évaluation précise des vecteurs propres de la matrice de transformation linéaire. En particulier, lorsque les matrices de covariance des images d'entrée et du bruit sont mal conditionnées pour des raisons pratiques, ou lorsque leur évaluation n'est pas précise. L'estimation des matrices de covariance n'est pas une tâche facile, en particulier lorsque la taille de l'image et/ ou le nombre d'images spectrales est important, où nous observons une accumulation des erreurs du calcul. Ces déficiences rendent l'ACP et ses variantes inefficaces pour des applications en

téledétection. Ainsi, pour une application efficace de l'ACP et de ses variantes, une méthode qui détermine avec précision la matrice de transformation linéaire, sans devoir estimer les matrices de covariance et calculer leurs décompositions spectrales, est nécessaire. Ceci peut être réalisé en utilisant les réseaux de neurones qui présentent le caractère d'auto-organisation dans l'apprentissage pour s'adapter à la nature des données [4]. Ceci fait l'objet de notre proposition d'article.

L'organisation de l'article s'établit comme suit. Le paragraphe 2 est consacré à expliquer le principe d'une ACP neuronale et donner un état de l'art sur les modèles de réseaux de neurones de l'ACP développés dans les communautés de réseaux de neurones et de traitement du signal. Le modèle neuronal proposé ici pour une ACP des images de téledétection sera donné en détail dans le paragraphe 3. Les résultats expérimentaux obtenus en utilisant une image multispectrale de téledétection fournie par le satellite Landsat-TM seront exposés et commentés dans le paragraphe suivant. Une étude comparative entre le modèle proposé et la méthode statistique de l'ACP est effectuée.

2. ACP Neuronale : Principe et Modèles

Récemment, dans les communautés des réseaux de neurones et de traitement du signal, un certain nombre de modèles des réseaux de neurones pour l'ACP a été proposé [5], [6], [7]. Dans la plupart des cas, il a été montré que ces modèles ont de meilleures performances et surmontent les inconvénients inhérents aux approches classiques de l'ACP [8]. Cependant, à l'exception du travail développé par Lee et Landgrebe [9] pour la sélection des caractéristiques des données de l'imageur de téledétection AVIRIS, aucune application de réseau neurones pour l'extraction des caractéristiques et la réduction de dimension des images de téledétection n'a été rapportée dans la littérature.

La relation entre l'ACP et les réseaux de neurones a commencé depuis que les neurones ont été organisés individuellement pour identifier les caractéristiques spécifiques de l'environnement, et que cette propriété d'auto-organisation a été liée aux critères statistiques, comme celui de l'ACP. La plupart des travaux effectués dans ce contexte a été inspirée des systèmes biologiques tout en développant ce qu'on appelle l'apprentissage non-supervisé. Les règles d'apprentissage non-supervisé sont prévues pour apprendre à partir des observations (les données d'entrée). Le but est de découvrir les caractéristiques significatives des données d'entrée. Un des pionniers dans l'apprentissage non-supervisé était le psychologue Donald Hebb qui a formulé la règle Hebbienne d'apprentissage [10]. S'il y a deux neurones simultanément activés de chaque côté d'une connexion alors le poids de cette connexion est forcé. Pour bien illustrer la règle Hebbienne d'apprentissage, considérons un neurone simple ayant comme entrée le vecteur $X=(x_1, x_2, \dots, x_p)^T$ et génère comme sortie la réponse y . Les connexions sont orientées de l'entrée vers la sortie du neurone. L'activation est calculée comme étant le produit inhérent du vecteur d'entrée X par le vecteur des poids synaptiques $W=(w_1, w_2, \dots, w_p)^T$, tel que :

$$u = \sum_{i=1}^p w_i x_i \quad (1)$$

La fonction d'activation, $y = f(u)$, calcule la sortie du neurone en fonction du produit inhérent. Hebb [10] a proposé une règle simple pour ajuster les poids synaptiques durant l'apprentissage du réseau : Quand deux neurones sont simultanément excités, la connexion entre eux est forcée. La fonction d'activation pourrait en principe être n'importe quelle fonction croissante du produit inhérent. Selon la règle Hebbienne d'apprentissage, le changement dans les poids synaptiques dans un neurone simple est tel que :

$$\Delta w_i(k) = \beta y(k) x_i(k) \quad (2)$$

où k est l'itération et β est le paramètre d'apprentissage. La mise à jour des w_i prend alors la forme suivante :

$$w(k+1) = w(k) + \beta y(k)x(k) \quad (3)$$

Le problème avec cette forme de mise à jour est qu'elle mène à une croissance exponentielle de w quand le nombre d'échantillons d'apprentissage augmente. La règle d'Hebb est donc numériquement instable [11], [13]. L'algorithme de normalisation proposé par Oja dans [13] pour stabiliser les règles d'Hebb consiste à incorporer un terme de normalisation. Ceci peut être formulé par l'équation itérative suivante :

$$w(k+1) = \frac{w(k) + \beta y(k)x(k)}{\left(\sum (w(k) + \beta y(k)x(k))^2\right)^{1/2}} \quad (4)$$

Pour de petites valeurs de paramètre β , l'équation (4) prend la forme approximative suivante :

$$w(k+1) = w(k) + \beta y(k)(x(k) - y(k)w(k)) \quad (5)$$

L'équation (5) est appelée la règle Hebbienne d'apprentissage et avec une petite modification, elle s'appelle aussi la règle Hebbienne d'apprentissage généralisée [12]. Cette règle s'avère être étroitement liée à l'ACP quand les neurones sont représentés par un modèle linéaire. Oja [11] a prouvé qu'une version normalisée de la règle Hebbienne appliquée sur une simple unité linéaire converge vers le vecteur propre principal de la matrice de covariance des données d'entrée. La normalisation proposée par Oja est non seulement essentielle pour la stabilité, mais constitue également la caractéristique commune dans plusieurs algorithmes neuronaux de l'ACP. Le rapport entre la règle Hebbienne normalisée et l'ACP, peut être développé et servir comme une idée de base pour plusieurs algorithmes neuronaux de l'ACP. Puisque la règle d'Oja peut seulement extraire une seule composante principale (CP), les travaux développés ont étendu cette approche pour extraire plusieurs CP à partir d'un réseau de plusieurs neurones linéaires [13]-[21]. D'une manière générale, les algorithmes neuronaux de l'ACP développés dans la littérature peuvent être divisés en deux catégories [2], [9] : Algorithmes de réestimation et algorithmes de décorrélation. Dans les algorithmes de réestimation, le réseau de neurones a seulement les coefficients synaptiques connectant les neurones d'entrée aux neurones de sortie. La mise à jour de ces coefficients est basée sur la règle Hebbienne d'apprentissage. Avant d'extraire une nouvelle CP, on soustrait les premières CP déjà extraites de l'entrée avant que les données ne soient impliquées dans l'apprentissage. En revanche, dans les algorithmes de décorrélation, le réseau de neurones possède les coefficients synaptiques et les coefficients latéraux. La mise à jour des coefficients synaptiques se fait selon la règle Hebbienne d'apprentissage, tandis que la mise à jour des coefficients latéraux est selon la règle anti-Hebbienne d'apprentissage. Les CP ainsi extraites sont décorrélées par annulation des coefficients latéraux durant l'apprentissage.

3. ACP Neuronale d'Images de Télédétection

La topologie du modèle neuronal proposé ici pour l'ACP des images de télédétection est présenté dans la figure 2. Le modèle effectue l'ACP d'image de télédétection sans aucun calcul non-neuronal additionnel et aucune estimation précédente d'une matrice quelconque. L'extraction des CP se fait de façon séquentielle composante après composante. Les images spectrales d'entrée (vecteur X) sont connectées aux images CP à extraire (vecteur Y) par les coefficients synaptiques (vecteur W). Le modèle utilise ce qu'on appelle les poids latéraux du réseau (vecteur C) qui servent à décorréler la $m^{ième}$ CP au cours de l'extraction avec les $(m-1)$ CP déjà extraites. L'algorithme d'apprentissage du modèle permet d'extraire d'une manière adaptative des images CP directement

à partir des images spectral d'entrée. L'apprentissage du modèle est basé sur le calcul récursif au sens des moindres carrés "*recursive least squares (RLS) computation*" pour extraire la $m^{ième}$ CP sachant que les $(m-1)$ CP ont été déjà extraites. La motivation derrière cette approche est la nécessité d'extraire les CP des données d'entrée dans les situations où le nombre nécessaire des CP à extraire est *a priori* inconnu. Parmi les modèles mentionnés dans le paragraphe 2, les modèles proposés dans [6] et [16] ne diffèrent pas du modèle de la figure 2 et possèdent presque la même architecture. Les deux modèles, cependant, diffèrent significativement en terme de rapidité de convergence due à l'utilisation de l'algorithme RLS par le modèle de la figure 2. Les modèles proposés dans [7] et [21] utilisent également l'algorithme RLS. Cependant, ils n'ont pas de connections latérales entre les neurones de sortie, mais utilisent d'une manière explicite la transformation de déflation pour l'extraction des CP. Autre avantage principal du modèle de la figure 2 est que le paramètre d'apprentissage de l'algorithme de mise à jour n'est pas fixe et dépend des caractéristiques statistiques des données d'entrée. De ce fait, le modèle n'a pas le problème de précision engendré par la rapidité de convergence [12] et détermine d'une manière récursive la variance associée à chaque CP. Son algorithme d'apprentissage est établi de telle sorte que l'exécution du logiciel développé soit totalement indépendante de l'utilisateur. L'apprentissage se fait selon les deux règles suivantes :

$$W_m^{(k+1)} = W_m^{(k)} + \beta_k (Y_m X - Y_m^2 W_m^{(k)}), \quad C_m^{(k+1)} = C_m^{(k)} - \beta_k (Y_m X + Y_m^2 C_m^{(k)}) \quad (6)$$

La première règle, dite adaptative, est la règle d'apprentissage d'Hebb normalisée. Elle sert à l'extraction de toutes les CP. La deuxième règle, dite d'orthogonalité inspirée du principe de déflation de matrices, est la règle d'apprentissage anti-Hebb. Elle n'est pas utilisée dans la construction de la première CP. Elle permet de rendre la $m^{ième}$ CP orthogonale aux $(m-1)$ CP déjà extraites. Le paramètre β_k joue un rôle capital dans la détermination des performances du réseau. En effet, afin de concevoir un algorithme d'apprentissage récursif optimal, l'évolution de β_k doit être cohérente avec la nature des données et l'application envisagée et en même temps ne doit pas conduire à une convergence lente [6], [7]. L'idée est d'intégrer le critère d'optimisation du réseau dans les règles (7) et (8) via un choix optimal de la valeur de β_k . Pour cela, nous avons exprimé β_k en fonction de la taille de l'image, N , et de la variance de la CP en cours de construction, $E(y_m^2)$, telle que : $\beta_k = 1/(N.E(y_m^2))$. La mise en œuvre de cette idée a nécessité quelques considérations pratiques afin de rendre le déroulement de l'algorithme d'apprentissage simple et plus rapide. Le réseau se stabilise lorsque les vecteurs W_i convergent vers les vecteurs propres de la matrice de covariance des images spectrales d'entrée, et les vecteurs C_i convergent vers zéro.

4. Influence de la variance et de la corrélation sur les résultats d'une ACP

Nous notons que les résultats d'une ACP obtenus à partir de la décomposition spectrale de la matrice de covariance des images spectrales sont fortement dépendants des variances des différentes images spectrales. Une légère modification dans les valeurs des variances des images spectrales a pour effet de changer les proportions de l'information portée par chaque image CP ainsi que sa valeur du SNR. Par conséquent, l'interprétation des résultats obtenus est à revoir. En fait, les images spectrales de variances importantes sont pondérées par des coefficients élevés dans la formation des premières images CP. Tandis que les images spectrales de variances relativement faibles participent à la formation des premières images CP avec de faibles coefficients, ainsi qu'à la formation des dernières images CP par des coefficients plus ou moins importants. Cependant, il est parfois intéressant d'équilibrer entre les participations des différentes images spectrales dans la formation des premières images CP, afin de permettre aux images spectrales de variances relativement faibles d'avoir plus de chance d'introduire l'information qu'elles portent dans les premières images CP. Ceci est possible, en mettant toutes les images spectrales possédant la même

moyenne statistique et la même variance statistique. Nous effectuons alors une ACP des images spectrales centrées et réduites. Les images CP obtenues sont appelées images CP standardisées.

Il arrive parfois que l'information contenue dans les dernières images CP, extrêmement bruitées, soit de grand intérêt. Cette information, caractérisée par de petites variances, peut nous donner une idée plus claire sur les structures fines de la scène imagée telles que : les structures géologiques, les réseaux routiers, les fleuves, détection de changement...etc. Elle représente le contraste spectral existant entre les différentes images spectrales. Ainsi, au lieu de chercher des images CP ayant des variances importantes, il est d'intérêt de chercher l'information inhérente à chaque image spectrale. Nous pouvons faire alors une ACP sélective qui utilise seulement un sous-groupe d'images spectrales sélectionnées. La sélection des images spectrales dépend de l'objectif visé. Cet objectif est réalisé en exploitant le contraste spectral existant entre les images spectrales. Le critère de sélection est basé sur la corrélation : une corrélation forte entre les images spectrales implique que le contraste spectral est faible et une corrélation moyenne ou faible implique que le contraste spectral est important. Si le but est de réduire la dimensionalité avec un minimum d'informations contenues dans les dernières images CP bruitées, alors le sous-groupe d'images spectrales fortement corrélées est utilisé comme entrée de l'ACP. L'information commune portée par ces images sélectionnées sera contenue dans la première image CP et le reste des images CP ne contient que du bruit. Si le but est la caractérisation du contraste spectral entre les images spectrales, de sorte que l'information inhérente à chaque image spectrale puisse être obtenue, les sous-groupes choisis doivent contenir des paires d'images spectrales de corrélation moyenne ou faible. Dans ce cas là, seulement la deuxième image CP est d'intérêt puisqu'elle contient le contraste entre les deux images spectrales. Cette image est moins bruitée puisque la corrélation entre les deux images d'entrée est moyenne ou faible. La première image CP contient l'information commune entre les deux images spectrales.

5. Résultats Expérimentaux

L'image multispectrale de télédétection (6 images spectrales) utilisée pour tester et évaluer l'approche neuronale de l'ACP proposée est une image de taille (200x200pixels) provenant du satellite Landsat-TM et couvrant la région Djebel Amour dans le Sud de l'Algérie. La figure 2 montre les six images spectrales originales. Pratiquement, aucune information ne peut être extraite à partir de ces images, qui sont sombres et bruitées, à l'exception de la cinquième image spectrale qui semble plus claire que les autres. L'existence des redondances entre ces images est indiquée par les coefficients de corrélation élevés donnés dans le tableau 1. Le tableau 2 nous donne les valeurs des rapports signal à bruit "*Signal-to-Noise Ratio (SNR)*" dans les images spectrales. Selon ces valeurs, la deuxième image spectrale possède le SNR le plus faible et la cinquième image spectrale le SNR le plus fort. Les six images CP extraites en utilisant l'ACP neuronale sont montrées dans la figure 3. Le tableau 2 nous donne les valeurs des SNR dans les images CP extraites. L'essentiel des informations, contenue dans les images spectrales originales est concentrée maintenant dans la première image CP, qui est une image de très haute qualité. L'information portée par cette image caractérise d'une manière plus précise l'état de surface de la scène observée. Presque toutes les variations de la scène sont visibles dans cette image. La deuxième et troisième image CP sont moins contrastées et le SNR est faible (tableau 2). Néanmoins, quelques détails de la scène sont évidents dans ces deux composantes. Les trois dernières images CP extraites sont dominées par le bruit et l'information portée est négligeable.

Afin d'effectuer une étude comparative, l'ACP statistique est appliquée sur les mêmes données. Les résultats obtenus sont donnés dans la Figure 4. En examinant les matrices de la transformation linéaire de l'ACP neuronale et de l'ACP statistique (tableau 3), nous remarquons que les vecteurs propres constituant ces matrices sont différents, incluant les vecteurs propres correspondant aux premières images CP. Un autre critère qui nous permet de faire ressortir les différences entre les deux méthodes est les valeurs des SNR. Les images spectrales ont des SNR s'étendant de 45,10 à 443,80. Tandis que, pour l'ACP neuronale, la valeur de SNR dans la première image CP est 964,940; ceci correspond à un facteur d'amélioration de 2,2. Pour l'ACP statistique, la valeur du

SNR de la première image CP est proche de celle de l'ACP neuronal (960,95). Cependant, les autres images CP ont des valeurs de SNR qui sont très différentes des valeurs du SNR de l'ACP neuronale. Ceci confirme de nouveau que les résultats de l'ACP neuronale proposée sont différents de ceux obtenus en utilisant l'ACP statistique. Pour montrer quelle méthode de l'ACP est la plus précise, nous avons utilisé un critère de mesure la fiabilité dans la compression d'images sans perte importante d'information. Ceci consiste à reconstituer les images spectrales en utilisant seulement les deux premières images CP. Pour mesurer le degré de similitude entre les images spectrales originales et les images spectrales reconstituées, l'erreur des moindres carrées "Mean-Square-Error (MSE)" est utilisée dans le processus de reconstitution : $MSE = \sqrt{E((x - \hat{x})(x - \hat{x}))}$, où x et \hat{x} sont les images spectrales originales et reconstituées, respectivement. Les figures 5 et 6 nous donnent les valeurs de MSE pour chaque paire d'images spectrale/reconstituée de l'ACP neuronale et statistique, respectivement. Pour l'ACP neuronale, toutes les valeurs du MSE se trouvent dans l'intervalle [2,10 et 3,95], ce qui reflète une forte ressemblance entre les images originales et les images reconstituées. En fait, visuellement, il est difficile de détecter les différences entre les images originales et les images reconstituées. La valeur moyenne du MSE du processus de reconstitution est approximativement égale à 2,987 (figure 5). Cette erreur est acceptable puisque les deux premières images CP représentent 98,87% de la variance totale des images spectrales originales. Ceci signifie que la méthode neuronale représente un facteur de compression égal à 3. Pour l'ACP statistique (figure 6), les valeurs du MSE varient de 2,50 à 4,50 et la valeur moyenne de MSE du processus de reconstitution est égale à 3,50. Étant donné que les deux premières images CP de l'ACP statistique représentent 98,43% de la variance totale des images spectrales originales, nous pouvons dire que le l'ACP neuronale est plus précise que l'ACP statistique puisque sa valeur moyenne correspondante de MSE est plus petite.

6. Conclusion

Le but essentiel du travail présenté ici était l'extraction des CP de l'image de télédétection sans avoir à calculer la matrice de covariance des images spectrales originales. Les réseaux de neurones permettent de réaliser cette idée, mais l'inconvénient majeur reste le temps d'apprentissage qui est généralement important. L'approche neuronale de l'ACP proposée ici ne présente pas cet inconvénient. Ceci est dû à l'algorithme d'apprentissage élaboré qui tient compte d'une manière implicite du critère d'optimisation du réseau. Ce critère est établi en tenant compte de la nature des données et du critère d'arrêt des itérations. L'extraction des images CP désirées est séquentielle et adaptative. L'application du modèle a été réalisée sur une image multispectrale de télédétection fournie par le satellite Landsat-TM. Les résultats obtenus montrent l'efficacité de l'approche dans l'extraction des CP les plus significatives de l'image multispectrale de télédétection. La capacité de compression de l'ACP neuronale proposée est prouvée en reconstituant les images spectrales avec un minimum d'erreurs au sens des moindres carrées. La précision des résultats obtenus est satisfaisante. Une étude comparative entre le modèle neuronal proposé et l'approche statistique de l'ACP est également effectuée. L'étude comparative a prouvé que l'approche neuronale de l'ACP est légèrement supérieure à l'approche statistique de l'ACP standard.

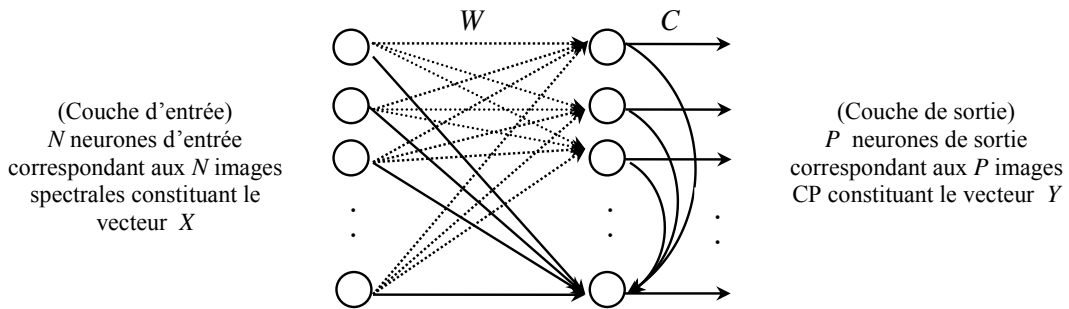


Figure 1. Modèle de réseaux de neurones pour l'ACP d'image de télédétection.

Tableau 1 : Coefficients de corrélation entre les six images spectrales.

| Images spectrales | 1 ^{ère} image | 2 ^{ème} image | 3 ^{ème} image | 4 ^{ème} image | 5 ^{ème} image | 6 ^{ème} image |
|------------------------|------------------------|------------------------|------------------------|------------------------|------------------------|------------------------|
| 1 ^{ère} image | 1.000000 | | | | | |
| 2 ^{ème} image | 0.950527 | 1.000000 | | | | |
| 3 ^{ème} image | 0.914597 | 0.977681 | 1.000000 | | | |
| 4 ^{ème} image | 0.893477 | 0.959884 | 0.981840 | 1.000000 | | |
| 5 ^{ème} image | 0.883223 | 0.937426 | 0.950577 | 0.947707 | 1.000000 | |
| 6 ^{ème} image | 0.872567 | 0.939118 | 0.958085 | 0.943569 | 0.980474 | 1.000000 |

Tableau 2 : SNR dans les images spectrales originales et dans les images CP de l'ACP neuronale et l'ACP statistique.

| SNR dans les images spectrales originales | | | | | |
|---|----------------------------------|----------------------------------|----------------------------------|----------------------------------|----------------------------------|
| 1 ^{ère} image spectrale | 2 ^{ème} image spectrale | 3 ^{ème} image spectrale | 4 ^{ème} image spectrale | 5 ^{ème} image spectrale | 6 ^{ème} image spectrale |
| 45.095375 | 41.679432 | 140.698288 | 142.237122 | 443.795746 | 181.927475 |
| SNR dans les images CP (ACP neuronale) | | | | | |
| 1 ^{ère} image CP | 2 ^{ème} image CP | 3 ^{ème} image CP | 4 ^{ème} image CP | 5 ^{ème} image CP | 6 ^{ème} image CP |
| 964.941357 | 19.277951 | 7.787135 | 6.024435 | 1.486921 | 0.728147 |
| SNR dans les images CP (ACP statistique) | | | | | |
| 1 ^{ère} image CP | 2 ^{ème} image CP | 3 ^{ème} image CP | 4 ^{ème} image CP | 5 ^{ème} image CP | 6 ^{ème} image CP |
| 960.954285 | 18.823294 | 7.699635 | 5.393777 | 1.747629 | 0.815081 |

Tableau 3 : Vecteurs propres des matrices de transformation de l'ACP neuronale et ACP statistique.

| Vecteurs propres de la matrice de transformation de l'ACP neuronale | | | | | |
|---|--------------------------|--------------------------|--------------------------|--------------------------|--------------------------|
| 1 ^{er} vecteur | 2 ^{ème} vecteur | 3 ^{ème} vecteur | 4 ^{ème} vecteur | 5 ^{ème} vecteur | 6 ^{ème} vecteur |
| 0.4259083 | -0.1824951 | -0.5542965 | 0.6200117 | 0.3061889 | -0.0032694 |
| 0.1933921 | 0.2669236 | 0.1411071 | 0.1579317 | -0.1644201 | 0.9052563 |
| 0.3551904 | 0.5163083 | -0.2321867 | 0.0304151 | -0.6713281 | -0.3191141 |
| 0.3623360 | 0.4961509 | -0.1190056 | -0.5033777 | 0.59582036 | -0.0090611 |
| 0.6954251 | -0.5606455 | 0.26883521 | -0.3178275 | -0.1705911 | -0.0005935 |
| 0.2020298 | 0.2615021 | 0.7297256 | 0.4850929 | 0.2107831 | -0.2803287 |
| Vecteurs propres de la matrice de transformation de l'ACP statistique | | | | | |
| 1 ^{er} vecteur | 2 ^{ème} vecteur | 3 ^{ème} vecteur | 4 ^{ème} vecteur | 5 ^{ème} vecteur | 6 ^{ème} vecteur |
| -0.1975441 | 0.3545559 | 0.8035035 | 0.0520671 | 0.3127581 | 0.2985325 |
| -0.2014603 | 0.2840557 | 0.2205537 | 0.1256235 | -0.1960211 | -0.8808395 |
| -0.3752205 | 0.4632445 | -0.1883529 | 0.2040297 | -0.6595953 | 0.3639297 |
| -0.3750777 | 0.4659833 | -0.4524561 | -0.4128843 | 0.5142297 | -0.0505557 |
| -0.6742529 | -0.5485271 | 0.1744883 | -0.4251579 | -0.1824769 | 0.0009841 |
| -0.4293041 | -0.2469559 | -0.1873481 | 0.7672355 | 0.3618885 | 0.0005263 |

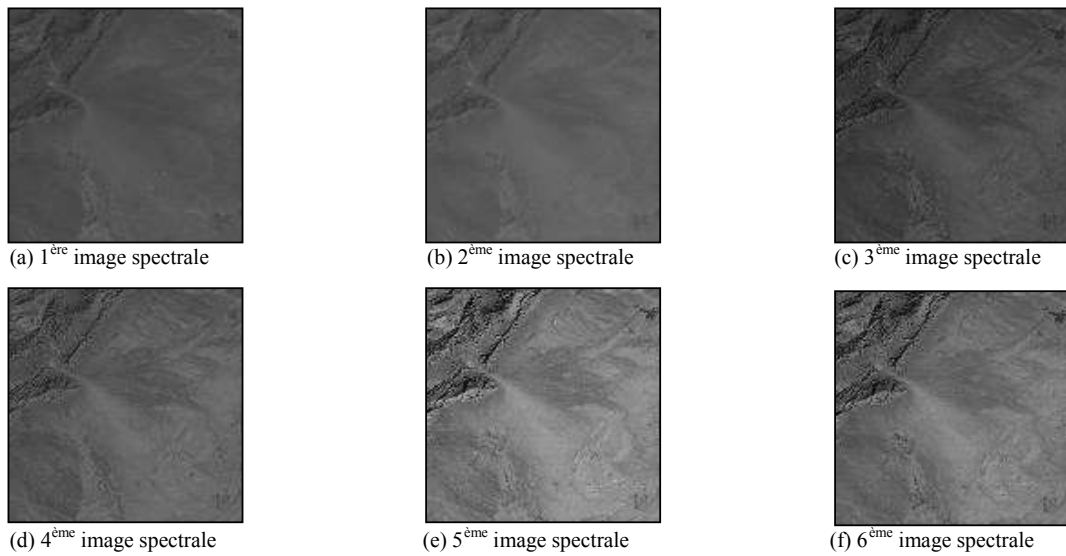


Figure 2. Les six images spectrales originales fournies par le satellite Landsat-TM.

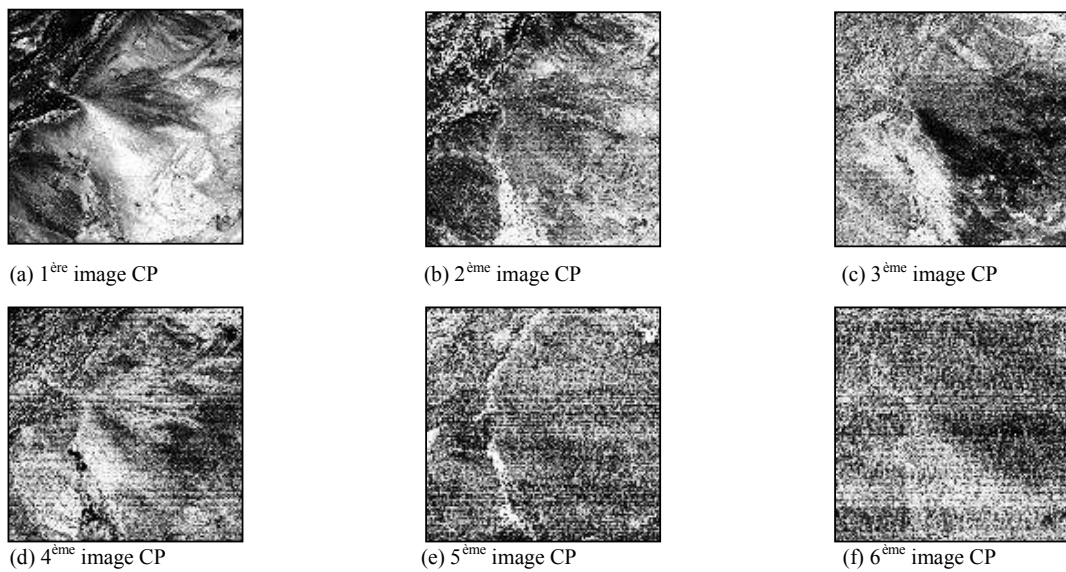


Figure 3. Les six images composantes principales produites par l'ACP neuronale.

Index des Auteurs

INDEX DES AUTEURS

| | |
|--------------------|----------|
| ABBAS K. | 77 |
| ABED-MERAÏM K. | 209 |
| ADJABI S. | 198 |
| AÏSSANI A. | 52, 84 |
| AÏSSANI D. | 58, 77 |
| AIT KACI AZZOUZ S. | 227 |
| AÏT-KAKI S. | 155 |
| ALI-MOKHNACHE S. | 162 |
| BAROUR C. | 155, 167 |
| AZOUZI B. | 174 |
| BENAICHATTA L. | 174 |
| BENCHETTAH A. | 93 |
| BENCHIKH T. | 180 |
| BIBI A. | 132 |
| BIBI M. | 27 |
| BOUALEM M. | 52 |
| BOUKETALA K. | 227 |
| BOUSBAINÉ A. | 139 |
| BOUTABIA H. | 46 |
| CHARLOT F. | 34 |
| CHITROUB S. | 239 |
| CHOUAF B. | 64, 99 |
| DADDA N. | 162 |
| EL BAHÏ M. | 186 |

| | |
|---------------|---------|
| FELLAG H. | 124 |
| GHRIBALLAH A. | 99 |
| GUELLIL A. | 15 |
| GUERBYENNE H. | 118 |
| HAMADOUCHE D. | 105 |
| HAMADOUCHE M. | 233 |
| IKHNACHE N. | 27 |
| LARDJANE T. | 88 |
| LARKEM A. | 162 |
| MAZROU H. | 233 |
| MERAD H. | 118 |
| MERAGHNI D. | 216 |
| MESSACI F. | 210 |
| MEZOUED F. | 186 |
| MOHDEB Z. | 112 |
| MOKKADEM A. | 112 |
| MOUHOUBI Z. | 58 |
| MSSADI D. | 162 |
| NECIR A. | 06, 216 |
| NOUALI K. | 124 |
| OULD-SAÏD E. | 70, 249 |
| PALM R. | 155 |
| POMORSKI D. | 145 |
| RACHDI M. | 194 |
| RADLOFF S.E. | 167 |

| | |
|--------------|---------------|
| RAHMANI F.L. | 222 |
| RAOULT J-P. | 210 |
| SABRE R. | 02 |
| SADKI O. | 70 |
| SAHMOUDI M. | 209 |
| SCHRÖTER J. | 174 |
| SMAIL R. | 84 |
| TAHAR A. | 155, 162, 167 |
| TALEB Y. | 105 |
| YOUSFATE A. | 180 |

USTHB Alger et Lamos Béjaia

Edition_ 2018