



L'association

*femmes
mathématiciennes*

en partenariat avec
la Mission pour la Place des Femmes du CNRS
organise le

15ème Forum des Jeunes Mathématicien-ne-s

Du 26 au 28 novembre 2015, à Lille 1 et Lille 3

Thèmes privilégiés : Probabilités et statistique

Conférences invitées :

Anne Philippe, Université de Nantes
Mylène Maïda, Université Lille 1
Sophie Dabo, Université Lille 3
Anne Gégout-Petit, Université de Lorraine

Site internet : <http://www.femmes-et-maths.fr/>

Soumissions des résumés (avant le 1 octobre), renseignements,
inscriptions : forum15@femmes-et-maths.fr

Avec le soutien financier de l'INSMI, du Ministère de l'Enseignement supérieur, du Service des Droits des femmes, de la Fédération de recherche mathématique du Nord-Pas de Calais, du Labex CEMPI, de l'UFR MIME de l'Université Lille 3, des Laboratoires Painlevé et RIME Lab de l'Université de Lille, du Laboratoire de Mathématiques Pures et Appliquées de l'Université du Littoral et du Laboratoire de Mathématiques et applications de Valenciennes.



CEMPI CENTRE EUROPÉEN
POUR LES MATHÉMATIQUES, LA PHYSIQUE ET
LEURS INTERACTIONS



Fédération de Recherche
Mathématique
du Nord-Pas de Calais

FUNCTIONAL BINARY CHOICE MODELS WITH CHOICE-BASED SAMPLING

AHMED Mohamed Salem

Doctorant

Université Lille, Laboratoire LEM

mohamed-salem.ahmed@etu.univ-lille3.fr

Résumé - *Functional Binary Choice Models have been developed to model a binary response variable as a function of an explanatory random function or to make discrimination and classification of functional data. Several authors have proposed methods allowing the estimation of the parameter function involved in these models, in a context of a random sampling (see Müller and Stadtmüller (2005), Escabias et al (2005)). We propose to generalize this type of models when we have a sample obtained from a Choice-Based Sampling (CBS) process. The interest of the CBS process is that it allows to take into consideration the structure of the population in where the sample is drawn when some choices (values of the response variable) are rarely chosen, unlike the random sample case where all items in the population have the same probability of being chosen. The proposed method is an extension of the Conditional Maximum Likelihood Method proposed by Manski and McFadden (1981). We prove the consistency of the estimate and investigate its finite sample performance on simulated data.*

Mots clés - **Choice-Based Sampling , Functional data, parameter function, Maximum likelihood.**

Conditional Maximum Likelihood Estimator with Explanatory Random Function

We are interested in describing a relation between a discrete random variable Y taking values in $\{0, \dots, M - 1\}$ (in particular take $M = 2$) and a random function $\{X(t), t \in \mathcal{T}\}$ which corresponds to a square integrable stochastic process on the interval \mathcal{T} . We assume that we have a sample (of a given population) provided by a pure choice-based sampling of a population into M subsets (or strata) such that each subsets corresponds to exactly one choice. This problem is known as Choice-Based Sampling. The interest of the choice-based sampling process is that it allows to take into consideration the structure of the population when some choices are rarely chosen, unlike the random sample case where all items in the population have the same probability of being chosen. We are interested in estimating binary choice models in the case of Choice-Based Sampled data with an explanatory random function covariate. Let $Q(j) = P(Y = j)$ be the population share of choice j (1 or 0). And let $H(j)$ be the probability according to which we draw the stratum j . The vector $(H(0), H(1), \dots, H(M-1))$ will be denoted by H and the vector $(Q(0), Q(1), \dots, Q(M-1))$ by Q . Assume that we have a prior information allowing the knowledge of Q and H , then the conditional density of Y

given $X = x$ in the sample at $Y = i$ is

$$g(i|x) = \frac{P(i|x, \theta) H(i)/Q(i)}{\sum_{j=1}^M P(j|x, \theta) H(j)/Q(j)} \quad (1)$$

where $P(i|x, \theta)$ is the probability of $Y = i$ given x with a parameter function θ that replaces the vector of slopes in vectorial case.

We will interested in the particular case $M = 2$ and then

$$P(Y = 1|x, \theta) = \Phi \left(\alpha + \int_{\mathcal{T}} x(t)\theta(t)dt \right)$$

where $\Phi(\cdot)$ is some cumulative distribution function. α is a constant and $\theta(\cdot)$ the parameter function, which assumed to be square integrable on \mathcal{T} .

We assume to have a sample of N independent observations $(Y_n = i_n, \{x_n(t), t \in \mathcal{T}\})$, $n = 1, \dots, N$, so the conditional likelihood function is defined by :

$$L(\alpha, \theta) = \sum_{n=1}^N \log \left(\frac{P(i_n|x_n, \theta) H(i_n)/Q(i_n)}{\sum_{j=1}^M P(j|x_n, \theta) H(j)/Q(j)} \right) \quad (2)$$

When the explanatory variable is real random vector, Manski and McFadden(1981) have maximized (2) to estimate the intercept α and the vector of parameter θ . This method is usually referred to as conditional maximum likelihood estimator.

We aim to use this method to estimate α and the parameter function θ in our context of fonctionnal covariate. Before that, we have to address the difficulty produced by the infinite dimensionality of the explanatory random function by approximating $P(i_n|x_n, \theta)$ with a series of models where the dimension of the explanatory variable is reduced to a positive integer $p = p_N$, where p_N increases as $N \rightarrow \infty$. This dimension reduction has been used in binary choice models of functional covariate in a context of random sample, see for instance Müller and Stadtmüller (2005).

Let $\{\varphi_j, j = 1, 2, \dots\}$ be an orthonormal basis of the functional space $L^2(\mathcal{T})$, usually a Fourier, a Spline basis or a basis constructed by the eigenfunctions of the covariance operator is chosed. We can rewrite $x(t)$ and $\theta(t)$ on the following way

$$x(t) = \sum_{j \geq 1} \varepsilon_j \varphi_j(t), \quad \theta(t) = \sum_{j \geq 1} \theta_j \varphi_j(t)$$

where the real random variables ε_j and the coefficients θ_j are given by

$$\varepsilon_j = \int_{\mathcal{T}} x(t)\varphi_j(t)dt, \quad \text{and } \theta_j = \int_{\mathcal{T}} \theta(t)\varphi_j(t)dt$$

By the orthonormality of this basis, we have

$$\int_{\mathcal{T}} x(t)\theta(t)dt = \sum_{j \geq 1} \theta_j \varepsilon_j.$$

Then, we can write

$$P(Y = 1|X = x, \theta) = \Phi \left(\alpha + \sum_{j \geq 1} \theta_j \varepsilon_j \right) \quad (3)$$

Functional Binary Choice Models With Choice-Based Sampling

Let

$$U_{p_N} = \alpha + \sum_{j=1}^{p_N} \theta_j \varepsilon_j, \quad V_{p_N} = \sum_{j=p_N+1}^{\infty} \theta_j \varepsilon_j$$

We can approximate $\Phi\left(\alpha + \sum_{j \geq 1} \theta_j \varepsilon_j\right)$ by the truncated model $\Phi(U_{p_N})$. We assume the following

$$E \left(\Phi \left(\alpha + \sum_{j \geq 1} \theta_j \varepsilon_j \right) - \Phi(U_{p_N}) \right)^2 \leq 4cE(V_{p_N}^2). \quad (4)$$

where c is a constant such that $\sup_{t \in \mathbb{R}} |\Phi'(t)|^2 \leq c$. The advantage of this approximation is that when we consider eigen-basis, the error of this approximation would be controlled by the eigenvalues associated to this eigen-basis, that is

$$E(V_{p_N}^2) = E \left(\sum_{j=p_N+1}^{\infty} \theta_j \varepsilon_j \right)^2 = \sum_{j=p_N+1}^{\infty} \theta_j^2 E(\varepsilon_j^2) = \sum_{j=p_N+1}^{\infty} \theta_j^2 \lambda_j.$$

Now for some fixed p , we will be interested to estimate the parameters α and $(\theta_1, \dots, \theta_p)$. For that we have to maximize the function obtained by replacing $\Phi\left(\alpha + \sum_{j \geq 1} \theta_j \varepsilon_j\right)$ in (2) by $\Phi(U_{p_N})$. For simplicity, replace the intercept α by θ_0 and we denote $\varepsilon_0 = 1$, then we have to estimate $(\theta_0, \theta_1, \dots, \theta_p)$ by maximizing the following pseudo conditional likelihood function

$$\begin{aligned} \tilde{L}_p(\theta) = & \sum_{n=1}^N i_n \log \frac{\Phi\left(\sum_{j=0}^p \theta_j \varepsilon_j^{(n)}\right) H^*/Q^*}{\Lambda} \\ & + (1 - i_n) \log \frac{\left(1 - \Phi\left(\sum_{j=0}^p \theta_j \varepsilon_j^{(n)}\right)\right) (1 - H^*) / (1 - Q^*)}{\Lambda} \end{aligned} \quad (5)$$

where

$$\Lambda = \Phi\left(\sum_{j=0}^p \theta_j \varepsilon_j^{(n)}\right) H^*/Q^* + \left(1 - \Phi\left(\sum_{j=0}^p \theta_j \varepsilon_j^{(n)}\right)\right) (1 - H^*) / (1 - Q^*),$$

$H^* = H(1)$, $Q^* = Q(1)$ and $\varepsilon_j^{(n)} = \int x_n(t) \varphi_j(t) dt$ for $j = 1, \dots, p$; $n = 1, \dots, N$ and $\varepsilon_0^{(n)} = 1$. The estimator of $(\theta_0, \dots, \theta_p)$ is given by $\hat{\theta}$ that maximises $\tilde{L}_p(\cdot)$, that is $\hat{\alpha} = \hat{\theta}_0$ and $\hat{\theta}(t) = \sum_{j=1}^p \hat{\theta}_j \varphi_j(t)$.

We define the $(p+1)$ matrix

$$\Gamma_p = \left(E \left(\frac{\Phi'^2\left(\sum_{j=0}^p \theta_j \varepsilon_j\right)}{\sigma^2\left(\Phi\left(\sum_{j=0}^p \theta_j \varepsilon_j\right)\right)} - \Psi' \left(\sum_{j=0}^p \theta_j \varepsilon_j \right) \right) \varepsilon_k \varepsilon_l \right)_{0 \leq k, l \leq p}$$

where ε is the generic copie generic of $\varepsilon^{(n)}$, $\sigma(t) = t(1-t)$ and

$$\Psi(t) = \frac{\Phi'(t) (H^*/Q^* - (1 - H^*) / (1 - Q^*))}{\Phi(t) H^*/Q^* + (1 - \Phi(t)) (1 - H^*) / (1 - Q^*)}$$

Under some assumptions, we will show that the estimator $\hat{\theta}$ of θ converges almost surely to θ and satisfies

$$\frac{N(\hat{\theta} - \theta)' \Gamma_{p_N}(\hat{\theta} - \theta) - (p_N + 1)}{\sqrt{2(p_N + 1)}} \rightarrow \mathcal{N}(0, 1).$$

In addition, we show also that

$$\frac{\int \int (\hat{\theta}(t) - \theta(t)) G(s, t) (\hat{\theta}(s) - \theta(s)) ds dt - (p_N + 1)}{\sqrt{2(p_N + 1)}} \rightarrow \mathcal{N}(0, 1)$$

with

$$G(t, s) = E \left(\left(\frac{\Phi'^2 \left(\alpha + \int X(t)\theta(t) dt \right)}{\sigma^2 \left(\Phi \left(\alpha + \int X(t)\theta(t) dt \right) \right)} - \Psi' \left(\alpha + \int X(t)\theta(t) dt \right) \right) X(t)X(s) \right).$$

References

1. Cosslett S.R. Cosslett (1981), Maximum Likelihood Estimation for Choice-based Samples, *Econometrica*, **Vol. 49**, p 1289–1316.
2. Esc M. Escabias, A. M. Aguilera and M. J. Valderrama (2005), Principal Component Estimation of Functional Logistic Regression : Discussion of Two Different Approaches, *Nonparametric Statistics*, **Vol. 16**, p.365–384.
3. Imbens1992 G. W. Imbens (1992), An Efficient Method of Moments Estimator for Discrete Choice Models With Choice-Based Sampling, *Econometrica*, **Vol. 60**, No. 5, p 1187–1214.
4. Manski1981 C.F. Manski, and D. McFadden (1981), Alternative Estimators and Sample Designs for Discrete Choice Analysis, in *Structural Analysis of Discrete Data*, ed. by C. F. Manski and D. McFadden. Cambridge, Massachusetts : M.I.T. Press.
5. Muller2005 H.G. Müller and U. Stadtmüller (2005), Generalized Functional Linear Models, *The Annals of Statistics*, **Vol. 33**, No. 2, p 774–805.

PENALIZED MDF FOR PROTEIN MOVEMENT DETECTION

Hiba Alawieh, Nicolas Wicker

Laboratoire Paul Painlevé, Université Lille 1,
59655 Villeneuve d'Ascq, France.

alawieh.hiba@gmail.com, Nicolas.Wicker@math.univ-lille1.fr

Baydaa Al Ayoubi

Université Libanaise, faculté des sciences 1, département mathématiques appliquées,
Al Hadath, Beyrouth, Liban.

ayoubib@ul.edu.lb

Luc Moulinier

ICube/LBGI, faculté de Médecine, 67 000 Starsbourg, France.

moumou@igbmc.fr

Abstract - *The three-dimensional structure of a given protein can take different conformations depending upon the reaction it undergoes and its substrate/cofactor/partners binding state. Various methods exist to study these conformational changes but only one, called DynDom, is clearly focused on movement detection. An alternative method is proposed, making use of multivariate data analysis, called “penalized Multidimensional Fitting (penalized MDF)” based on penalized movements of points in order to approach the distances between points after movement to the distances given by the reference matrix. The objective is to detect the amino acids that undergo an important movement by fitting the distances of one conformation to the distances of the second one by modifying only the coordinates of the first one. This method is applied to three different proteins.*

Keyword - **Proteins, Conformational changes, movements, Ligand binding, Multidimensional analysis, MDF, Optimization, Penalization**

1 Introduction

Proteins are heteropolymers that can take three dimensional structures. These structures are flexible, highly dynamic, and their biological functions depend intimately on them. They can change their shape, unfold (and refold), and switch to another conformation.

Different methods are used to study the conformational changes of proteins. Here, a novel method called penalized Multidimensional Fitting (penalized MDF) is presented to detect movements by using two conformations of the same protein. This multivariate analysis approach is an adaptation of Multidimensional Fitting method [1,2]. The idea is to compare one protein conformation with another one by modifying the coordinate matrix of the first one, called target matrix, in order to make the distances calculated on the modified target matrix similar to the distances given by the reference matrix that is obtained, in the case of protein movement detection, by computing the distances between its amino acids. Penalization is necessary as it is clear that without it, every transformation would be possible, and then the

solution would reduce to take the modified target matrix equal to the coordinate matrix corresponding to the reference matrix! This of course, would not give any information on which part of the protein has moved. The main work here is to devise a good penalization, and then to apply this method on different proteins.

2 Penalized MDF method

Let $X = \{X_1 | \dots | X_n\}$ be the $n \times p$ target matrix given by the amino acids coordinates of a protein structure and $D = \{d_{ij}\}$ the $n \times n$ reference matrix calculated on another structure of the same protein, this matrix contains the Euclidean distances between the amino acids. Besides, we note $\Delta = \{\delta_{ij}\}$ the distance matrix obtained from X after penalized MDF. The MDF method will allow us to modify the target matrix in order to minimize the difference between the reference matrix and the novel distance matrix computed on the modified target matrix. The idea behind MDF is to minimize the mean square error:

$$\frac{2}{n(n-1)} \sum_{1 \leq i < j \leq n} (d_{ij} - \delta_{ij})^2$$

where $\delta_{ij} = d(f(X_i), f(X_j))$ and $f(X_i) = X_i + L_i$ under some constraints. For all $i \in 1, \dots, n$, the vector $L_i = (l_{i1}, l_{i2}, \dots, l_{ip})$ denotes the displacement for the i^{th} point. Here, no constraint is needed but to avoid unnecessary movements, a penalty term is added to the mean square error leading to the following function to optimize:

$$E = \sum_{1 \leq i < j \leq n} (\|X_i + L_i - X_j - L_j\|_2 - d_{ij})^2 + \lambda \sum_{i=1}^n \text{pen}(L_i)$$

and the optimization problem is: $O : \begin{cases} \min E \\ \text{with } L_i \in \mathbb{R}^p \end{cases}$

Here, no constraint is needed but to avoid unnecessary movements, a penalty term is added leading to the following optimization problem:

$$O : \begin{cases} \min \sum_{1 \leq i < j \leq n} (\|X_i + L_i - X_j - L_j\|_2 - d_{ij})^2 + \lambda \sum_{i=1}^n \text{pen}(l_i) \\ \text{with } l_i \in \mathbb{R}^p \end{cases}$$

The parameter λ is a positive regularization parameter that controls the trade-off between the approximation of the reference matrix by the distance matrix computed on the modified matrix and the use of a parsimonious number of displacements. To have interesting results, it is clear that having a good penalization is important

2.1 choice of penalty function

The three natural possibilities for the choice of penalty function are the ℓ_0 , ℓ_1 and ℓ_2 -norms. First, we want to discard the ℓ_1 -norm, to have a norm invariant by rotation as the structures can have any orientation. The ℓ_2 -norm is used to penalize the displacements of points. Furthermore, to take into account the number of points that move we use the ℓ_0 -norm.

Penalized MDF for Protein Movement Detection

The ℓ_0 -norm is not treated in a first stage to focus on the ℓ_2 -norm. In this stage, the objective function is the following:

$$E = \sum_{1 \leq i < j \leq n} (\|X_i + L_i - X_j - L_j\|_2 - d_{ij})^2 + \lambda \sum_{i=1}^n \|L_i\|_2.$$

In a second stage, the results are transposed to the ℓ_0 -norm in an elastic-net like context.

The combined penalty term is $\sum_{i=1}^n (\gamma \|L_i\|_2 + (1 - \gamma) \|L_i\|_0)$ weighted by a parameter λ , with $\gamma \in [0, 1]$. We call the function $\gamma \|L_i\|_2 + (1 - \gamma) \|L_i\|_0$ the elastic net penalty by analogy with the well-known elastic net. The function to optimize becomes:

$$E = \sum_{1 \leq i < j \leq n} (\|X_i + L_i - X_j - L_j\|_2 - d_{ij})^2 + \lambda \sum_{i=1}^n (\gamma \|L_i\|_2 + (1 - \gamma) \|L_i\|_0)$$

2.2 choice of parameter λ

We have already seen that the value of λ is crucial for obtaining good results. In this section, we want to find the best λ value using ℓ_2 -norm as penalty function (first stage).

Lemma 1 *The solution of problem O is such that there is a fixed point or at least two points moving in different directions provided that $\lambda > 0$.*

Lemma 2 *In one dimension, we suppose that $x_i + l_i \leq x_j + l_j$, $\forall i, j \in 1, \dots, n$ and $i < j$. If for two points i and $i + 1$ moving in opposite directions such that $x_i \leq x_{i+1}$ and by taking $j = i + 1$, the parameter λ is such that $\lambda < n(d_{ij}^0 - d_{ij} - 2\epsilon)$ then $\exists k$, such that $|l_k| > \epsilon$, where d_{ij}^0 is the initial distance computed on the target matrix.*

Lemma 3 *In dimension p , if $\lambda < \frac{n-1}{2}\epsilon$ and $\forall i, j \in 1, \dots, n$, the difference is such that $|\|X_i - X_j\|_2 - d_{ij}| > \epsilon$ then the optimum solution of problem O is given by small displacements that give distances equal to distances given by the reference matrix.*

3 Application

In this section, penalized MDF has been applied to three different proteins to detect important movements in their tridimensional structure. Penalized MDF needs a reference matrix and a coordinate matrix. For two different structures of same protein, the coordinate matrix is given by the C_α coordinates of one structure and the reference matrix is given by the Euclidean distances between the C_α of the second structure. The optimization problem is a nonlinear optimization problem. The Nlopt library has been used to solve it by using DIRECT-L algorithm for global optimization and SBPLX algorithm for local optimization. For the choice of parameter λ , recall that by lemmas 2 and 3 we have seen that the order of magnitude of λ is n times a small gap value, that we take equal to 0.5\AA for the present application. Concerning parameter γ , the following values have been used (0, 0.1, 0.3, 0.5, 0.8, 1) focusing then on 0.5 which gave the best results. To have a threshold to determine if a movement is important, we compare the computed displacement after penalized MDF

with the standard deviation σ_i for each point i . For this, we use the known B-factor (or temperature factor) of each atom i which indicates the true static or dynamic mobility of an atom given by: $B_i = 8\pi^2 d_{mi}^2$, to infer the mean displacement d_{mi} of atom i , $\forall i = 1, \dots, n$. Besides, $d_{mi}^2 = E(\|X_i - \mu_i\|^2)$ with $X_i \rightsquigarrow \mathcal{N}(\mu_i; \sigma_i^2 I_3)$ and μ_i the mean coordinates for each atom i . Then, $d_{mi}^2 = \sigma_i^2 E(\frac{\|X_i - \mu_i\|^2}{\sigma_i^2}) = 3\sigma_i^2$ as $\frac{\|X_i - \mu_i\|^2}{\sigma_i^2} \rightsquigarrow \chi_3^2$. Thus, $\sigma_i = d_{mi}/\sqrt{3}$. We suppose that the value $2\sigma_i$ is high enough to detect important displacements.

Penalized MDF has been applied to three proteins: human estrogen nuclear receptor (ER), Ferrichrome-iron receptor (FhuA), and aspartyl-tRNA synthetase (AspRS). For each proteins, we compare our results with those obtained by DynDom.

3.1 Human estrogen receptor protein

ER is a Nuclear estrogen receptor composed of several functional domains that serve specific roles. The ligand-binding domain (LBD) region corresponds to the ligand binding domain. Many experiments demonstrate that their C-terminal Helix (H12) is more flexible without ligand. Penalized MDF has been used to compare the conformation with and without ligand.

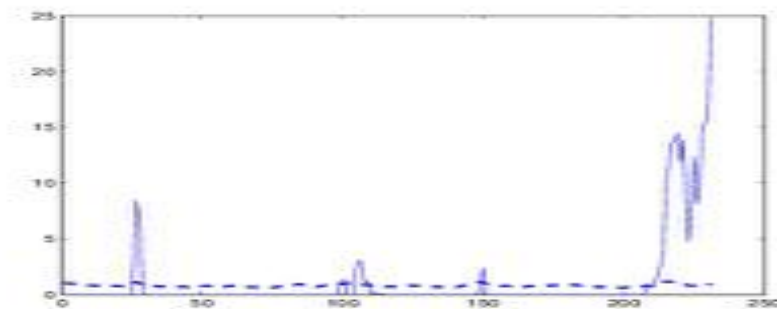


Figure 1: Distances between initial and modified coordinates for each amino acid (solid line) for Human estrogen receptor with $\lambda = n/2$ and $\gamma = 0.5$. The x -axis indicates the amino acids, and the y -axis indicates the distances. The dashed line corresponds to the $2\sigma_i$ threshold. The amino acids 26, 27, 28 and 214 – 231 are considered as important displacements.

Figure 1 indicates that important displacements occur at the amino acids number 26, 27, 28 and at the end of the sequence (the amino acids number 214 to 231) and smaller displacements in others regions.

The dashed line in Figure 1 shows the $2\sigma_i$ threshold for each point. Important displacements are detected at amino acids 26, 27, 28, and 214 to 231. This result is confirmed by Anke and others who note that the position of this helix depends on the presence or absence of a ligand [3]. Concerning positions 26, 27, 28, they correspond to the sequence "SEA" which is apical of helix H3. Sumbayev and others explain the movements of this helix in [4].

3.2 FhuA protein

FhuA is an outer membrane receptor protein of *Escherichia-coli* bacteria. X-ray analysis at 2.7Å resolution reveals two distinct conformations in the presence and absence of ferrichrome

Penalized MDF has been applied to compare the two conformations. Figure ?? depicts

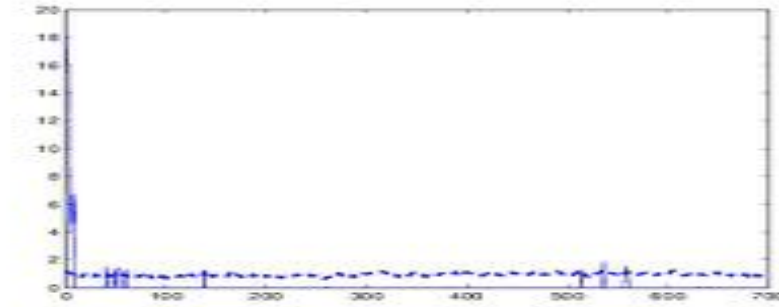


Figure 2: For FhuA, important movements are located in the N-terminus, which is confirmed by the biological literature.

important distances between initial and modified coordinates for the 10 first amino acids of the amino acid sequence of FhuA protein. This result is confirmed by biology [5] et [6]. The N-terminus has moved after ligand binding.

No domains movements are detected by using DynDom for ER and FhuA proteins, whereas using penalized MDF, we succeed in finding the residues that move according to the known biological literature.

4 Conclusion

The purpose of penalized MDF is to modify only the coordinates of amino acids that have significantly moved and fix the others. Penalization term and penalization parameters are crucial in the process of obtaining good results. This involves the choice of a penalty coefficient λ which is related to the minimum displacement.

Penalized MDF has been applied to three different proteins in order to find the residues that were affected by the interaction with other molecules. Comparison of our results with Procrustes analysis, DynDom results and the literature reveals that penalized MDF is an efficient method for protein movement detection and fills a gap in the computational toolbox of molecular biologists. Further research is needed to consider displacements as random variables using a stochastic model to assess how significative displacements are.

References

- [1] Alawieh, H., Wicker, N., Ayoubi, B. and Moulinier, L. Penalized Multidimensional Fitting For Protein Movement Detection, submitted.
- [2] Berge, C., Froloff, N., Kalathur, R.K., Maumy, M., Poch, O., Raffelsberger, W. and Wicker, N. (2010). Multidimensional fitting for multivariate data analysis, *Journal of Computational Biology*, **17**, 723–732.
- [3] Mueller-Fahrnow, A., and Egner, U. (1999) Ligand-binding domain of estrogen receptors, *Current Opinion in Biotechnology*, **10**, 550–556.

- [4] Sumbayev, V.V, Bonefeld-Jorgensen, E.C., Wind, T, Andreasen, P.A. (2005) A novel pesticide-induced conformational state of the oestrogenreceptor ligand-binding domain, detected by conformation-specificpeptide binding, *FEBS Letters*, **579**, 541–548.
- [5] Faraldo-Gomez J.D., Smith G.R., Sansom M.S. (2003) Molecular dynamics simulations of the bacterial outer membrane protein FhuA: a comparative study of the ferrichrome-free and bound states, *Biophysical journal*, **85**, 1406-1420.
- [6] Locher, K.P, Rees B, Koebnik R, Mitschler A, Moulinier L, Rosenbusch J.P, Moras D (1998) Transmembrane Signaling across the Ligand-Gated FhuA Receptor: Crystal Structures of Free and Ferrichrome-Bound States Reveal Allosteric Changes, *Cell press*, **95**, 771–778.

Penalized MDF for Protein Movement Detection

GRANDES DÉVIATIONS DE LA PLUS GRANDE VALEUR PROPRE DES MATRICES DE WIGNER SANS QUEUES GAUSSIENNES.

Fanny Augeri

Institut de Mathématiques de Toulouse
Université Paul Sabatier, Toulouse, France
fanny.augeri@math.univ-toulouse.fr

Résumé - *On s'intéresse aux grandes déviations de matrices de Wigner sans queues gaussiennes, c'est-à-dire telles que les queues de distribution des entrées $\mathbb{P}(|X_{1,1}| > t)$ et $\mathbb{P}(|X_{1,2}| > t)$ se comportent comme e^{-bt^α} et e^{-at^α} respectivement pour $a, b \in (0, +\infty)$ et $\alpha \in (0, 2)$ donnés. On montre pour ce modèle, un principe de grande déviations pour la plus grande valeur propre de vitesse $N^{\alpha/2}$ avec une bonne fonction de taux qui ne dépend que de la queue de distribution des entrées.*

Mots clés - **Grandes déviations, Matrices de Wigner**

1 Introduction

Les premiers résultats de grandes déviations dans le contexte des matrices aléatoires furent obtenus en 1997 grâce au travail de G. Ben Arous et A. Guionnet. Dans un article précurseur [AG97], ils prouvent un principe de grandes déviations pour la mesure spectrale empirique dans le cas des β -ensembles munis d'un potentiel quadratique, de vitesse N^2 et avec une bonne fonction de taux explicite. Dans [AGZ10, p.81], ce résultat a été étendu par les mêmes auteurs, dans le cas des β -ensembles avec un potentiel V tendant vers l'infini plus vite que $\log|x|$. Plus précisément, on considère la mesure de probabilité sur \mathbb{R}^N suivante:

$$P_{V,\beta}^N(d\lambda_1, \dots, d\lambda_N) = \frac{1}{Z_{V,\beta}^N} \prod_{i < j} |\lambda_i - \lambda_j|^\beta \exp\left(-N \sum_{i=1}^N V(\lambda_i)\right) d\lambda_1 \dots d\lambda_N,$$

où $Z_{V,\beta}^N$ est une constante de normalisation, et $\beta > 0$. Le résultat de [AGZ10, p.81] montre que la suite $(\frac{1}{N} \sum_{i=1}^N \delta_{\lambda_i})_{N \in \mathbb{N}}$ satisfait un principe de grandes déviations à vitesse N^2 et avec une bonne fonction de taux explicite faisant intervenir l'entropie non-commutative Voiculescu. Récemment, il a été montré dans [Har12] que la restriction sur la croissance du potentiel V pouvait être allégée, de telle façon que l'on pouvait considérer aussi des potentiels avec une croissance logarithmique.

Dans le cas des ensembles de Wigner déformé, les grandes déviations de la mesure spectrale empirique ont été étudiés, d'abord dans [CDG01] puis dans [GZ02], où un principe de grandes déviations a été établi pour la mesure spectrale empirique d'une somme d'une matrice de Wigner gaussienne et d'une matrice hermitienne déterministe. Pour ce modèle, puisqu'il est impossible de calculer la loi jointe des valeurs propres, la preuve du principe de grandes

déviations repose sur la nature gaussienne des entrées et utilise le mouvement Brownien de Dyson et des outils de calcul stochastique.

En ce qui concerne les grandes déviations des valeurs propres extrémales des matrices de Wigner, le premier résultat a été prouvé dans [BADG01] dans le cas du GOE (Gaussian Orthogonal Ensemble) et a été ensuite étendu dans [AGZ10, p.83] pour les β -ensembles. Le principe de grandes déviations est de vitesse N et avec une fonction de taux explicite. Les grandes déviations des valeurs propres extrémales des ensemble de Wigner déformés ont aussi été étudiés. Dans [Mai07], l'auteur s'intéresse au cas d'une matrice du GOE (respectivement du GUE) perturbée par une matrice symétrique de rang 1 (respectivement hermitienne). Dans [BGGM12], la question des grandes déviations pour la loi jointe des valeurs propres extrémales d'une matrice déterministe diagonale perturbée par une matrice hermitienne de rang fini ayant les vecteurs propres délocalisés est complètement étudiée.

Cependant, tous ces résultats de grandes déviations reposent sur la connaissance de la loi jointe des valeurs propres ou bien sur la nature gaussienne des entrées. Dans [BC12], C. Bordenave et P. Caputo démontrent un principe de grandes déviations pour la mesure spectrale empirique de matrices de Wigner ayant des entrées sans queues gaussiennes, un cas où il n'y a pas de calcul possible de la loi jointe des valeurs propres. Récemment, ce résultat a été étendu au cas des matrices de Wishart dans [GroPR].

2 Résultats

Soient $(X_{i,j})_{i < j}$ une famille de variables aléatoires complexes indépendantes et identiquement distribuées telles que $\mathbb{E}(X_{1,2}) = 0$, $\mathbb{E}|X_{1,2}|^2 = 1$, et $(X_{i,i})_{i \geq 1}$ une famille de variables aléatoires réelles indépendantes et identiquement distribuées, indépendante de $(X_{i,j})_{i < j}$.

Soit $X(N)$ la matrice hermitienne de taille $N \times N$ telle que $X(N)_{i,j} = X_{i,j}$ pour tout $1 \leq i \leq j \leq N$. On appelle une telle suite $(X(N))_{N \in \mathbb{N}}$ une matrice de Wigner. Dans la suite, on omettra l'indice N et on écrira X à la place de $X(N)$.

On considère maintenant la matrice normalisée $X_N = X/\sqrt{N}$. On note $\lambda_1 \leq \lambda_2 \leq \dots \leq \lambda_N$ ses valeurs propres rangées dans l'ordre croissant. On définit la mesure spectrale empirique de X_N par

$$\mu_{X_N} = \frac{1}{N} \sum_{i=1}^N \delta_{\lambda_i}.$$

D'après le théorème de Wigner ([Wig58], [AGZ10, Theorem 2.1.21, 2.21]), on sait que presque sûrement,

$$\mu_{X_N} \underset{N \rightarrow +\infty}{\rightsquigarrow} \sigma_{sc},$$

où \rightsquigarrow désigne la convergence faible et σ_{sc} la loi du demi-cercle, qui est définie par

$$\sigma_{sc}(dt) = \frac{1}{2\pi} \mathbf{1}_{t \in [-2,2]} \sqrt{4 - t^2} dx.$$

De plus, si on suppose que $\mathbb{E}|X_{1,1}|^2 < +\infty$ et $\mathbb{E}|X_{1,2}|^4 < +\infty$, alors on a d'après [FK81] et [BY88],

$$\lambda_N \underset{N \rightarrow +\infty}{\longrightarrow} 2 \text{ p.s.}$$

On rappelle qu'une suite de variables aléatoires $(Z_n)_{n \geq 1}$ à valeurs dans un espace topologique \mathcal{X} muni de sa tribu borélienne \mathcal{B} , suit un principe de grande déviations (PGD) à vitesse $v : \mathbb{N} \rightarrow \mathbb{N}$ et avec une bonne fonction de taux $J : \mathcal{X} \rightarrow [0, +\infty]$, si J est semi-continue inférieurement et si v tend vers l'infini et pour tout $B \in \mathcal{B}$,

$$-\inf_{B^\circ} J \leq \liminf_{N \rightarrow +\infty} \frac{1}{v(n)} \log \mathbb{P}(Z_n \in B) \leq \limsup_{n \rightarrow +\infty} \frac{1}{v(n)} \log \mathbb{P}(Z_n \in B) \leq -\inf_{\bar{B}} J,$$

où B° est l'intérieur de B et \bar{B} l'adhérence de B . On rappelle que J est semi-continue inférieurement si ses sous-niveaux $\{x \in \mathcal{X} : J(x) \leq t\}$ sont fermés. De plus, si tout les sous-niveaux de J sont compacts, alors on dit que J est une bonne fonction de taux. On se place sous les mêmes hypothèses que dans [BC12], que l'on énonce ci-dessous.

Hypothèses 1 Soit X une matrice de Wigner telle que $\mathbb{E}X_{1,2}^2 = 0$. Soient $\alpha \in (0, 2)$ et $a, b \in]0, +\infty]$ tels que

$$\begin{aligned} \lim_{t \rightarrow +\infty} -t^\alpha \log \mathbb{P}(|X_{1,1}| > t) &= b, \\ \lim_{t \rightarrow +\infty} -t^\alpha \log \mathbb{P}(|X_{1,2}| > t) &= a. \end{aligned} \tag{1}$$

De plus, on suppose qu'il existe deux mesures de probabilités sur \mathbb{S}^1 , ν_1 et ν_2 , et $t_0 > 0$, tel que pour tout $t \geq t_0$ et tout borélien U de \mathbb{S}^1 ,

$$\begin{aligned} \mathbb{P}(X_{1,1}/|X_{1,1}| \in U, |X_{1,1}| \geq t) &= \nu_1(U) \mathbb{P}(|X_{1,1}| \geq t), \\ \mathbb{P}(X_{1,2}/|X_{1,2}| \in U, |X_{1,2}| \geq t) &= \nu_2(U) \mathbb{P}(|X_{1,2}| \geq t). \end{aligned}$$

Autrement dit, pour tout indices i, j , la valeurs absolue et l'angle de $X_{i,j}$ sont indépendants pour de grandes valeurs de $|X_{i,j}|$.

Sous ces hypothèses, il a été prouvé dans [BC12] que la mesure spectrale empirique de X_N , notée μ_{X_N} , suit un principe de grandes déviations pour la topologie faible. Le PGD est à vitesse $N^{1+\alpha/2}$ et de bonne fonction de taux I définie pour tout $\mu \in \mathcal{M}_1(\mathbb{R})$, où $\mathcal{M}_1(\mathbb{R})$ désigne l'ensemble des mesures de probabilités sur \mathbb{R} , par

$$I(\mu) = \begin{cases} \Phi(\nu) & \text{si } \mu = \sigma_{sc} \boxplus \nu \text{ avec } \nu \in \mathcal{M}_1(\mathbb{R}), \\ +\infty & \text{sinon,} \end{cases}$$

où \boxplus désigne la convolution libre et Φ une bonne fonction de taux (voir [BC12] pour plus de détails).

On montre le résultat de grandes déviations suivant pour la plus grande valeur propre λ_{X_N} de X_N .

Théorème 1 Sous les hypothèses (1), la suite $(\lambda_{X_N})_{N \in \mathbb{N}}$ suit un principe de grandes déviations à vitesse $N^{\alpha/2}$ et de bonne fonction de taux définie pour tout $x \in \mathbb{R}$, par

$$J(x) = \begin{cases} I(x) & \text{si } x > 2, \\ 0 & \text{si } x = 2, \\ +\infty & \text{si } x < 2, \end{cases}$$

où pour tout $x > 2$,

$$I(x) = \inf \left\{ b \sum_{i=1}^{+\infty} |A_{i,i}|^\alpha + \frac{a}{2} \sum_{i \neq j} |A_{i,j}|^\alpha : \lambda(A) = G_{\sigma_{sc}}(x)^{-1}, A \in \mathcal{D} \right\},$$

avec

$$\mathcal{D} = \left\{ A \in \cup_{n \geq 1} H_n(\mathbb{C}) : \forall i, \frac{A_{i,i}}{|A_{i,i}|} \in \text{supp}(v_1), \forall i \neq j, \frac{A_{i,j}}{|A_{i,j}|} \in \text{supp}(v_2) \right\},$$

où $\text{supp}(v_i)$ désigne le support de la mesure v_i pour $i \in \{1, 2\}$, et où $G_{\sigma_{sc}}$ désigne la transformée de Stieltjes de la loi du demi-cercle, c'est-à-dire

$$\forall z \in \mathbb{C} \setminus [-2, 2], G_{\sigma_{sc}}(z) = \int \frac{d\sigma_{sc}(t)}{z-t},$$

avec

$$\sigma_{sc}(dt) = \mathbf{1}_{|t| \leq 2} \frac{1}{2\pi} \sqrt{4-t^2} dt.$$

References

- [AG97] G. Ben Arous and A. Guionnet. Large deviations for Wigner's law and Voiculescu's non-commutative entropy. *Probability theory and related fields*, 108:517–542, 1997.
- [AGZ10] G. W. Anderson, A. Guionnet, and O. Zeitouni. *An introduction to random matrices*, volume 118 of *Cambridge Studies in Advanced Mathematics*. Cambridge University Press, Cambridge, 2010.
- [BADG01] G. Ben Arous, A. Dembo, and A. Guionnet. Aging of spherical spin glasses. *Probab. Theory Related Fields*, 120(1):1–67, 2001.
- [BC12] C. Bordenave and P. Caputo. Large deviation principle for Wigner matrices without gaussian tail. *arXiv:1207.5570. to appear in Annals of Probability*, 2012.
- [BGM12] F. Benaych-Georges, A. Guionnet, and M. Maida. Large deviations of the extreme eigenvalues of random deformations of matrices. *Probab. Theory Related Fields*, 154(3-4):703–751, 2012.
- [BY88] Z. D. Bai and Y. Q. Yin. Necessary and sufficient conditions for almost sure convergence of the largest eigenvalue of a Wigner matrix. *Ann. Probab.*, 16(4):1729–1741, 1988.
- [CDG01] T. Cabanal Duvillard and A. Guionnet. Large deviations upper bounds for the laws of matrix-valued processes and non-communicative entropies. *Ann. Probab.*, 29(3):1205–1261, 2001.
- [FK81] Z. Füredi and J. Komlós. The eigenvalues of random symmetric matrices. *Combinatorica*, 1(3):233–241, 1981.

- [GroPR] B. Groux. Asymptotic freeness for rectangular random matrices and large deviations for sample covariance matrices with sub-gaussian tails. *arXiv*, 1505.05733 [math.PR].
- [GZ02] Alice Guionnet and Ofer Zeitouni. Large deviations asymptotics for spherical integrals. *J. Funct. Anal.*, 188(2):461–515, 2002.
- [Har12] Adrien Hardy. A note on large deviations for 2D Coulomb gas with weakly confining potential. *Electron. Commun. Probab.*, 17:no. 19, 12, 2012.
- [Mai07] M. Maïda. Large deviations for the largest eigenvalue of rank one deformations of Gaussian ensembles. *Electron. J. Probab.*, 12:1131–1150 (electronic), 2007.
- [Wig58] Eugene P. Wigner. On the distribution of the roots of certain symmetric matrices. *Ann. of Math. (2)*, 67:325–327, 1958.

15ème Forum des Jeunes Mathématicien-ne-s, Lille

ESTIMATION DE L'INDICE DE QUEUE POUR DES CHAMPS ALÉATOIRES

Bassène Aladji

Laboratoire LEM, Université Charles-De-Gaulle, Lille3, Maison de la Recherche, domaine universitaire du Pont de Bois, BP 60149, 59653 Villeneuve d'Ascq cedex, France.

aladji.bassene@etu.univ-lille3.fr

Résumé - *Les statistiques d'événements rares ou extrêmes ont été bien développées dans le cas des observations indépendantes et identiquement distribuées (i.i.d). Toutefois, les données peuvent apparaître spatialement dépendantes et à queue lourde. Ce travail s'intéresse au problème d'estimation de l'indice de queue d'une distribution à queue lourde pour un processus spatial lorsque les valeurs les plus importantes sont observées dans des blocs. Cet estimation est appliquée à certaines catégories de processus spatiaux à queue lourde $\{X_{\mathbf{i}} \in \mathbb{R}^d, \mathbf{i} \in \mathbb{Z}^N\}$. Nous nous sommes particulièrement intéressés aux processus spatiaux linéaires causals $\{X_{\mathbf{i}}, \mathbf{i} \in \mathbb{Z}^N\}$, dans le sens où*

$$X_{\mathbf{i}} = \sum_{\mathbf{k} \in \mathbb{Z}^N, \mathbf{k} \geq \mathbf{0}} a_{\mathbf{k}} \epsilon_{\mathbf{i}-\mathbf{k}}, \quad \mathbf{i} \in \mathbb{Z}^N,$$

où les $\{\epsilon_{\mathbf{k}}\}_{\mathbf{k} \in \mathbb{Z}^N}$ sont i.i.d. et les $\{a_{\mathbf{k}} \geq 0, \mathbf{k} \in \mathbb{N}^N\} \in \mathbb{R}^{\infty}$ sont tels que $\sum_{\mathbf{k} \in \mathbb{Z}^N, \mathbf{k} \geq \mathbf{0}} |a_{\mathbf{k}}|^{\delta} < \infty$, où δ est une constante positive. Nous prouvons d'abord la convergence en probabilité de l'estimateur de Hill pour la classe des processus linéaires avant d'étendre le résultat à la classe des champs mélangeants. La normalité asymptotique de l'estimateur de Hill est aussi obtenue pour ce dernier type de processus.

Mots clés - **Processus spatiaux ; M -dépendant ; linéaire causal ; α -mélangeant ; Estimation d'indice de queue ; Consistance.**

1 Introduction

Dans ce travail, nous nous sommes intéressés à l'estimation non paramétrique de l'indice (ou index) de queue d'un processus spatial $\{X_{\mathbf{i}}, \mathbf{i} \in \mathbb{Z}^N\}$ à valeurs dans \mathbb{R}^d et indexé dans \mathbb{Z}^N avec $N \geq 1$. On suppose que le champ aléatoire est défini dans un espace de probabilité (Ω, A, P) , où (\mathbb{R}, d) est l'espace métrique associé à la métrique $d(\cdot, \cdot)$. Nous avons proposé une nouvelle approche de l'estimateur de l'indice de queue d'une distribution dans un cadre spatial.

Soit $X \in \mathbb{R}$ une variable aléatoire telle que : $\forall \lambda > 0, \forall x > 0$,

$$\mathbb{P}(X > x) = x^{-\frac{1}{\gamma}} L(x), \quad (1)$$

est caractérisée par $\gamma > 0$, l'unique paramètre positif inconnu appelé index des valeurs extrêmes. On considère un processus $(X_{\mathbf{i}})_{\mathbf{i} \in \mathbb{Z}^N}$, où \mathbf{i} est un site. Nous considérons des ob-

servations $X_{\mathbf{i}}$ dans un domaine rectangulaire

$$\mathcal{I}_{\mathbf{n}} = \{\mathbf{i} = (i_1, \dots, i_N) \in \mathbb{Z}^N, 1 \leq i_k \leq n_k, k = 1, \dots, N\}, \mathbf{n} = (n_1, \dots, n_N) \in \mathbb{Z}^N,$$

suisant le modèle (1).

Dans le cas non spatial ($N=1$), l'estimateur de l'indice des valeurs extrêmes est construit comme suit (e.g. [3], [5, 6], [1]).

- On définit d'abord une statistique d'ordre associée aux n observations de la sorte $X_{(1),n} \geq X_{(2),n} \geq \dots \geq X_{(n),n}$.
- Ensuite on choisit une séquence intermédiaire k_n telle que $1 \leq k_n \leq n$ et $k_n = o(n)$ quand $n \rightarrow \infty$ où n est la taille de l'échantillon.
- Finalement, on définit l'estimateur de Hill suivant

$$\gamma_n = \frac{1}{k_n} \sum_{i=1}^{k_n} \log \left(\frac{X_{(i),n}}{X_{(k_n+1),n}} \right) \quad (2)$$

Notre objectif est de proposer une nouvelle version de l'estimateur (2) dans le cadre spatial.

1.1 Modèle

Dans la suite, nous supposons que l'ensemble \mathbb{Z}^N , $N \geq 2$ est muni de l'ordre lexicographique (voir [2]), $n_j = n$. Nous traitons parfois l'échantillon spatial comme des tableaux triangulaires, c'est-à-dire les $\{X_{\mathbf{i}}\}_{\mathbf{i} \in \mathcal{I}_{\mathbf{n}}}$ seront écrits $\{X_{i,n}\}$ pour $1 \leq i \leq n^N = \hat{\mathbf{n}}$ (voir [7]) où n^N est la taille de l'échantillon. Nous pouvons identifier chacun des indices $i = 1, \dots, n$ avec un emplacement \mathbf{i} dans l'espace $\mathcal{I}_{\mathbf{n}}$.

Plus généralement, soit g une fonction bijective et continue telle que

$$\begin{aligned} g : \mathbb{N}^N &\longrightarrow \mathbb{N} \\ (i_1, \dots, i_N) &\longrightarrow \tilde{i}. \end{aligned}$$

basée sur les indices spatiaux facilitant ainsi l'identification des sites où les observations extrêmes ont été détectées. Cette fonction transforme les indices spatiaux en indices univariés. Par exemple, lorsque nous avons un réseau régulièrement espacé de dimension 2, où à la fois le nombre n_1 de lignes et le nombre de colonnes n_2 augmentent avec $n_1 * n_2$, le point spatial $\mathbf{i} = (i_1, i_2)$, pour un site $\mathbf{i} \in \mathcal{I}_{\mathbf{n}}$, peut être indexé par $i = n_2(i_1 - 1) + i_2$.

L'ensemble $\mathcal{I}_{\mathbf{n}}$ peut être réécrit comme suit :

$$\{X_{i,\hat{\mathbf{n}}}, i \in J_{\mathbf{n}} = g(\mathcal{I}_{\mathbf{n}})\}.$$

Notons $X_{(1),\hat{\mathbf{n}}} \geq X_{(2),\hat{\mathbf{n}}} \geq \dots \geq X_{(\hat{\mathbf{n}}),\hat{\mathbf{n}}}$ la statistique d'ordre associée aux $n^N = \hat{\mathbf{n}}$ variables $X_{i,\hat{\mathbf{n}}}$ de $J_{\mathbf{n}}$.

Soit $\mathbf{k}_{\mathbf{n}}$ une suite intermédiaire d'éléments de \mathbb{N}^N , dont toutes les composantes sont égales à $k_n \in \mathbb{N}$ telle que $\mathbf{1} \leq \mathbf{k}_{\mathbf{n}} \leq \mathbf{n}$. Nous supposons que $\mathbf{k}_{\mathbf{n}}$ est une séquence intermédiaire spatiale, ce qui signifie que k_n est une séquence d'entiers. Nous supposons que :

$$k = k_{\mathbf{n}} \rightarrow \infty; k_{\mathbf{n}} = o(n) \text{ as } n \rightarrow \infty.$$

Nous pouvons écrire formellement :

$$\mathbf{k}_{\mathbf{n}} \rightarrow \infty; \mathbf{k}_{\mathbf{n}} = o(\mathbf{n}) \text{ as } \mathbf{n} \rightarrow \infty.$$

ESTIMATION DE L'INDICE DE QUEUE POUR DES CHAMPS ALÉATOIRES

Nous définissons alors une version étendue de l'estimateur de l'indice de queue dans le cadre spatial considéré :

$$\gamma_{\mathbf{n}} = \frac{1}{\hat{k}_{\mathbf{n}}} \sum_{i=1}^{\hat{k}_{\mathbf{n}}} \log \left(\frac{X_{(i),n}}{X_{(\hat{k}_{\mathbf{n}}+1),n}} \right)$$

où $\hat{k}_{\mathbf{n}} = k_{\mathbf{n}}^N$.

L'étude de la consistance de l'estimateur de Hill nécessite une mesure de queues lourdes.

1.2 Mesure empirique et consistance de l'indice de queue

L'étude de la consistance de l'estimateur de Hill nécessite la consistance de la mesure de queues lourdes. Pour celà, nous adaptons l'estimateur de la mesure introduit dans [4] au cas spatial. Soit $\mathbb{E} =: \mathbb{R} \setminus \{0\}$, \mathcal{E} la σ -algèbre de Borel générée par des ouverts appartenant à \mathbb{E} , $\mathbf{C}_{\mathbf{K}}^+(\mathbb{E})$ est l'espace des fonction continues non négatives sur \mathbb{E} avec un support compact \mathbf{K} et $\mathbf{M}_+(\mathbb{E})$ est l'espace de la mesure positive de Radon sur \mathbb{E} dotée de la vague topologie. On définit la mesure μ par

$$\mu : \mathcal{E} \rightarrow \mathbb{R}_+,$$

où μ a une densité

$$\mu(dx) = \gamma^{-1} \pi_0 x^{-\gamma^{-1}-1} dx \mathbb{I}_{(0;+\infty)}(x) + \gamma^{-1} (1 - \pi_0) (-x)^{-\gamma^{-1}-1} dx \mathbb{I}_{[-\infty;0)}(x).$$

Pour $x \in \mathbb{E}$ et $A \in \mathcal{E}$, on définit la mesure ε_x par

$$\varepsilon_x(A) = \begin{cases} 1, & \text{si } x \in A \\ 0, & \text{si } x \notin A. \end{cases}$$

Alors, pour $\hat{k} = \hat{k}_{\mathbf{n}} \leq \hat{\mathbf{n}} = n^N$, on définit la mesure empirique de queue par

$$\hat{\mu}_{\mathbf{n}}(\cdot) = \frac{1}{\hat{k}} \sum_{i=1}^{\hat{\mathbf{n}}} \varepsilon_{X_{i,\hat{\mathbf{n}}}}(\cdot).$$

Ce dernier peut être également écrit comme suit

$$\hat{\mu}_{\mathbf{n}}(\cdot) = \frac{1}{\hat{k}} \sum_{i \in \mathcal{I}_{\mathbf{n}}} \varepsilon_{X_i}(\cdot).$$

Nous prouvons la convergence faible de la mesure empirique de queues lourdes $\hat{\mu}_{\mathbf{n}}(\cdot)$ dans le cas spatial c'est-à-dire

$$\hat{\mu}_{\mathbf{n}}(\cdot) \Rightarrow \mu(\cdot).$$

Et sous des conditions classiques des valeurs extrêmes, des propriétés de dépendance du processus, nous obtenons la consistance de l'estimateur proposé, c'est-à-dire

$$\gamma_{\mathbf{n}} \xrightarrow{\mathbb{P}} \gamma.$$

Nous proposons également l'estimation de quantiles associés à l'indice de queue ainsi que leur consistance.

Références

- [1] de Haan, L. and Ferreira, A. (2006), *Extreme Value Theory : An Introduction*. Springer.
- [2] Gao, J. and Lu, Z. and Tjøstheim, D. (2008), *Moment inequalities for spatial processes*. Statistics & Probability Letters, 78(6), 687-697.
- [3] Hill, B. M. (1975), *A simple general approach to inference about the tail of a distribution*. The annals of statistics, 3(5), 1163-1174.
- [4] Resnick, S. (1987), *Extreme values, regular variation, and point processes*. Springer Verlag, **vol 4**.
- [5] Resnick, S. and Stărică, C. (1993), *Consistency of hill's estimator for dependent data*. Technical report, Cornell University Operations Research and Industrial Engineering.
- [6] Resnick, S. and Stărică, C. (1997), *Asymptotic behavior of hill's estimator for autoregressive data*. Communications in statistics. Stochastic models, 13(4),703-721.
- [7] Robinson, P. M. (2011), *Asymptotic theory for nonparametric regression with spatial data*. Journal of Econometrics, 165(1), 5-19.

UNE APPROCHE MARKOVIENNE DU THÉORÈME CENTRAL LIMITE

Claire Delplancke

Institut de Mathématiques de Toulouse
Université Paul Sabatier, Toulouse, France
claire.delplancke@math.univ-toulouse.fr

Résumé - Nous nous intéressons à une nouvelle idée de preuve pour le théorème central limite dans le cas de variables aléatoires réelles $(X_i)_{i \in \mathbb{N}^*}$ i.i.d de densité $\varphi \in L^2(\gamma)$, où γ est la mesure gaussienne standard. Cette idée de preuve repose sur l'utilisation du caractère markovien de la suite

$$Y_n = \frac{1}{\sqrt{n}} \sum_{i=1}^n X_i, \quad n \in \mathbb{N}^*$$

ainsi que de la structure d'espace de Hilbert séparable de $L^2(\gamma)$.

Cette note présente succinctement nos travaux en cours, qui seront prochainement soumis pour publication.

Mots clés - théorème central limite, inégalité à la Berry-Esseen, convergence de processus de Markov, polynômes de Hermite

1 Contexte et démarche

1.1 Contexte

Les inégalités à la Berry-Esseen permettent de quantifier la convergence dans le théorème central limite, en utilisant une quantité qui mesure la distance entre la loi de Y_n et γ . Le premier résultat de ce type est dû à Berry (1941) et Esseen (1942) qui utilisent la distance de Kolmogorov pour quantifier la convergence. Plus récemment, des résultats du même type ont été montrés pour la distance en variation totale ([BC14]), pour l'entropie ([ABBN04]), pour l'information de Fisher ([BCG14]) et pour les distances de Wasserstein ([Rio11]).

Nous présenterons un résultat partiel portant sur une telle inégalité à la Berry-Esseen. Pour toute fonction $f \in L^2(\gamma)$, considérons la quantité

$$\text{Var}_\gamma(f) = \int \left(f - \int f d\gamma \right)^2 d\gamma.$$

C'est la quantité Var_γ qui nous intéresse ici pour quantifier la convergence dans le TCL car elle est bien adaptée à l'espace de Hilbert $L^2(\gamma)$ dans lequel on se place.

1.2 Démarche

Dans un premier temps, la structure de chaîne de Markov inhomogène de $(Y_n)_{n \in \mathbb{N}^*}$ fournit un opérateur de $L^2(\gamma)$ permettant d'exprimer f_{n+1} en fonction de f_n . Dans un second temps,

on étudie les valeurs singulières de cet opérateur.

2 Résultats

Nous proposons une conjecture dans le cas général, que nous sommes en mesure de démontrer pour certaines densités φ polynomiales.

Conjecture 1 Soit $(X_i)_{i \in \mathbb{N}^*}$ des variables aléatoires réelles i.i.d. et pour tout $n \in \mathbb{N}^*$, $Y_n = n^{-1/2} \sum_{i=1}^n X_i$.

On suppose que X_1 admet une densité φ par rapport à γ et que $\varphi \in L^2(\gamma)$, et pour tout $n \in \mathbb{N}^*$ on note f_n la densité de Y_n .

Pour $r \geq 2$ entier, si $\mathbb{E}[X_1^k] = \int x^k d\gamma$ pour $k = 0, \dots, r$, alors $\exists c > 0$ tel que

$$\text{Var}_\gamma(f_n) \leq \frac{c}{n^{r-1}}.$$

L'espace $L^2(\gamma)$ admet comme base hilbertienne l'ensemble des polynômes de Hermite, définis pour tout entier n par

$$H_n(x) = (-1)^n e^{x^2/2} \left(\frac{d}{dx} \right)^n (e^{-x^2/2}).$$

En conséquence, la densité $\varphi \in L^2(\gamma)$ admet une décomposition sur la base des $(H_n)_{n \in \mathbb{N}^*}$. Dans le cas où cette décomposition est finie, c'est-à-dire où φ est polynomiale, et sous certaines conditions sur les coefficients de la décomposition, la conjecture est vérifiée.

Références

- [ABBN04] Shiri Artstein, Keith M Ball, Franck Barthe, and Assaf Naor. On the rate of convergence in the entropic central limit theorem. *Probability theory and related fields*, 129(3) :381–390, 2004.
- [BC14] Vlad Bally and Lucia Caramellino. Asymptotic development for the clt in total variation distance. *arXiv preprint arXiv :1407.0896*, 2014.
- [BCG14] Sergey G Bobkov, Gennadiy P Chistyakov, and Friedrich Götze. Fisher information and the central limit theorem. *Probability Theory and Related Fields*, 159(1-2) :1–59, 2014.
- [Rio11] Emmanuel Rio. Asymptotic constants for minimal distance in the central limit theorem. *Electronic Communications in Probability*, 16 :96–103, 2011.

Une approche markovienne du théorème central limite

SPATIAL STATISTICS IN DISCRETE-CHOICE MODELS

DRWESH Emad-Aldeen

Doctorant

Université Lille, Laboratoire LEM, Domaine du pont de bois, maison de la recherche, 59653

Villeneuve d'ascq cedex

emad-aldeen.drwesh@etu.univ-lille3.fr

Abstract. *Spatial binary models are useful in many areas such as in economic and epidemiology where activities are often located in space. A way that makes the analysis of such spatial activities possible is to find a kind of correlation between some random variables in one location with others at neighboring locations. We propose here to describe and analyze the spatial (geographical) variation in disease (cancer) with respect to some risk factors using spatial binary models containing spatial latent choice variable and/or spatial autoregressive disturbances in a context of sampling data. This problem is known as Choice-Based Sampling (CBS) in discrete choice model. Unlike the random sample where all items in the population have the same probability of being chosen, the Choice-Based Sampling (CBS) in discrete choice model is a type of sampling where the classification of the population into subsets to be sampled is based on the choices or outcomes. In addition to the conditional maximum likelihood estimator (Manski and McFadden, 1981), we adapt the GMM approach (Imbens, 1992) in our context of spatial Choice-Based Sampling. We also provide a GMM estimator based on the generalized residuals (see Gouriéroux et al. (1987)). We present a Monte Carlo experiment to investigate the finite sample performance of these estimation methods. An application to real cancer data in northern France is also provided.*

Keywords. Spatial econometrics, Spatial autocorrelation, Choice-Based Sampling, GMM.

Choice-Based Spatial Modeling

Suppose we have a sample of n observations collected from points or regions located on an irregularly spaced, countable lattice $\mathcal{I} \subset \mathbb{R}^d$, $d \geq 1$; Let $i = (i_1, \dots, i_d) \in \mathbb{R}^d$ denote a site, and assume that all sites in \mathcal{I} are located at distances of at least $d_0 > 0$ for each other; i.e $\forall i, j \in \mathcal{I}$: $\|i - j\| \geq d_0$. Consider a sample of $i \in \mathcal{I}_n$ (\mathcal{I}_n is a finite subset of \mathcal{I} of cardinal n) individuals and let X be a vector of k exogenous discrete or continuous random variables. Our spatial model with latent dependent (unobserved scalar) variable can be written as

$$\begin{aligned} y^* &= \rho W y^* + X\beta + u; \\ u &= \lambda M u + \epsilon; \\ \epsilon & \text{ i.i.d. } \sim N(0, \sigma_\epsilon^2 I_n), \end{aligned} \tag{1}$$

where $y^{*t} = (y_i^*)_{i \in \mathcal{I}_n}$, $\beta \in \mathbb{R}^k$ is unknown, W and M are $n \times n$ known spatial weight matrices with two scalar autoregressive parameters ρ and λ , indicating the degree of spatial dependence and β is a $k \times 1$ vector of coefficients. Consider the binary model, where for individual located at site i , we observe the model

$$y_i = \begin{cases} 1 & \text{if } y_i^* \geq 0 \\ 0 & \text{if } y_i^* < 0. \end{cases} \quad (2)$$

Suppose that the sample that is considered is an endogenously stratified sampling in which the classification of the population into M (here $M = 2$) subsets (or strata) to be sampled is based on the choices or outcomes (for instance $y_i = 0, 1$).

This problem is known as Choice-Based Sampling (CBS) in discrete choice model. CBS is usually used when some choices are rarely chosen, unlike the random sample case where all items in the population have the same probability of being chosen. Let $Q(j)$ be the population share of choice j (1 or 0) (marginal probability of observing an individual choosing alternative j). And let $H(j)$ be the probability according to which we draw the stratum j . Assume that we have knowledge on H and Q , then the conditional density of j given $X = x$ in the sample is

$$g(j|x) = \frac{P(j|x, \theta)H(j)/Q(j)}{\sum_{j'=1}^M P(j'|x, \theta)H(j')/Q(j')} , \quad (3)$$

where $P(j|x, \theta)$ is the probability that one will choose alternative j given x with a finite set of parameters to be estimated θ . In this context, the use of standard Maximum likelihood estimation (MLE) procedure in CBS could lead to an inconsistent (asymptotically biased) estimation. Manski and McFadden (1981), proposed maximizing the conditional density given by equation (3) in a context of independent data. This method is usually referred to as conditional maximum Likelihood estimator (CMLE) using observations $(x_i, y_i = j_i)_{i \in \mathcal{I}_n}$:

$$\begin{aligned} \hat{\theta}_{CMLE} &= \arg \min_{\theta} \sum_{i \in \mathcal{I}_n} \log g(j_i|x_i) \\ &= \arg \min_{\theta} \sum_{i \in \mathcal{I}_n} \log \left(\frac{P(j_i|x_i, \theta)H/Q}{\sum_{l=1}^M P(l_i|x_i, \theta)H/Q} \right). \end{aligned} \quad (4)$$

In addition to the CMLE, we adapt the GMM approach of Imbens (1992) in our context of spatial Choice-Based Sampling. Here, The score function in CMLE will be interpreted as a moment function. That is, the GMM estimator can be defined as

$$\begin{aligned} \hat{\theta}_{GMM} &= \arg \min_{\theta} R_{n, C_n}(H, Q, \theta) \\ &= \arg \min_{\theta} \frac{1}{n} \sum_{i \in \mathcal{I}_n} \psi(H, Q, \theta, j_i, x_i)' \cdot C_n \cdot \frac{1}{n} \sum_{i \in \mathcal{I}_n} \psi(H, Q, \theta, j_i, x_i), \end{aligned} \quad (5)$$

where ψ is the vector of derivatives of the conditional density (equation (3)) at θ , and C_n is a positive definite weight matrix estimated as

$$n \left[\sum_{i \in \mathcal{I}_n} \psi(H, Q, \hat{\theta}, j_i, x_i) \psi(H, Q, \hat{\theta}, j_i, x_i)' \right],$$

where $(\hat{\theta})$ is the solution of $\arg \min_{\theta} R_{n,I}(H, Q, \theta)$.

We also provide a GMM estimator based on the generalized residuals denoted $\tilde{r}_i, i \in \mathcal{I}_n$ (Gourieroux et al. (1987)) derived from the first order condition for the CMLE by adapting the non CBS spatial GMM (Pinkse and Slade (1998)). The parameter vector θ is then estimated by

$$\hat{\theta}_{GMM_2} = \arg \min_{\theta} r' Z T Z' r, \quad (6)$$

where Z is a matrix of some instruments and T is a positive definite matrix.

Monte Carlo Experiment

We present a Monte Carlo experiment to investigate the finite sample performance of these three estimation methods. This simulation is based on one explanatory variable X , we set the true parameters as $\beta_0 = -0.5, \beta_1 = 1$ and we test several values of ρ and λ . 2000 individuals are generated, two sizes of stratified samples are drawn, namely, $n = 600$ and $n = 300$ and each Monte Carlo experiment is based on 200 replications. Table (1) summarizes the mean and standard deviation of each parameter in logit spatial error model ($\rho = 0$) using GMM Imbense, GMM Pinkse and CML estimation in each Monte Carlo experiment for sample size $n = 300$ and $n = 600$.

λ	n	β_0 Pinkse	β_1 Pinkse	β_0 Imbense	β_1 Imbense	β_0 CMLE	β_1 CMLE
-0,3	$n = 300$	-0,4848 (0,1112)	0,9834 (0,1034)	-0,4865 (0,1131)	0,9854 (0,1055)	-0,4848 (0,1111)	0,9834 (0,1034)
	$n = 600$	-0,4986 (0,0849)	0,9903 (0,0740)	-0,5012 (0,0853)	0,9933 (0,0746)	-0,4987 (0,0849)	0,9904 (0,0740)
0,2	$n = 300$	-0,5110 (0,1143)	1,0040 (0,1171)	-0,5155 (0,1182)	1,0088 (0,1216)	-0,5110 (0,1143)	1,0040 (0,1172)
	$n = 600$	-0,5058 (0,0811)	1,0012 (0,0734)	-0,5087 (0,0825)	1,0041 (0,0757)	-0,5061 (0,0812)	1,0014 (0,0737)
0,4	$n = 300$	-0,4899 (0,1219)	0,9837 (0,1225)	-0,4879 (0,1262)	0,9837 (0,1279)	-0,4408 (0,1219)	0,9766 (0,1225)
	$n = 600$	-0,4967 (0,0897)	0,9966 (0,0810)	-0,4921 (0,0904)	0,9965 (0,0824)	-0,4999 (0,0897)	0,9897 (0,0810)

Table 1: Estimation results for logit spatial lag model using GMM Imbense, GMM Pinkse and CML estimation methods with $n = 300$ and $n = 600$.

Application

Northern France is known to have the highest mortality rate for cancers of the upper aerodigestive tract (UADT), we use the spatial CBS model to identify the impact of several factors on the probability of having a UADT cancer in northern France. The socio-economic determinants of those cancers are well-know (alcohol, tobacco, poor living conditions,...). In this work we would like to study their impact in a spatial context and identify the potential unknown

effects or interactions. We have a sample of 89 of UADT cancer cases collected from 5 hospitals in the region of Nord-Pas-de-Calais and a sample of 348 of controls drawn randomly from the population of this region.

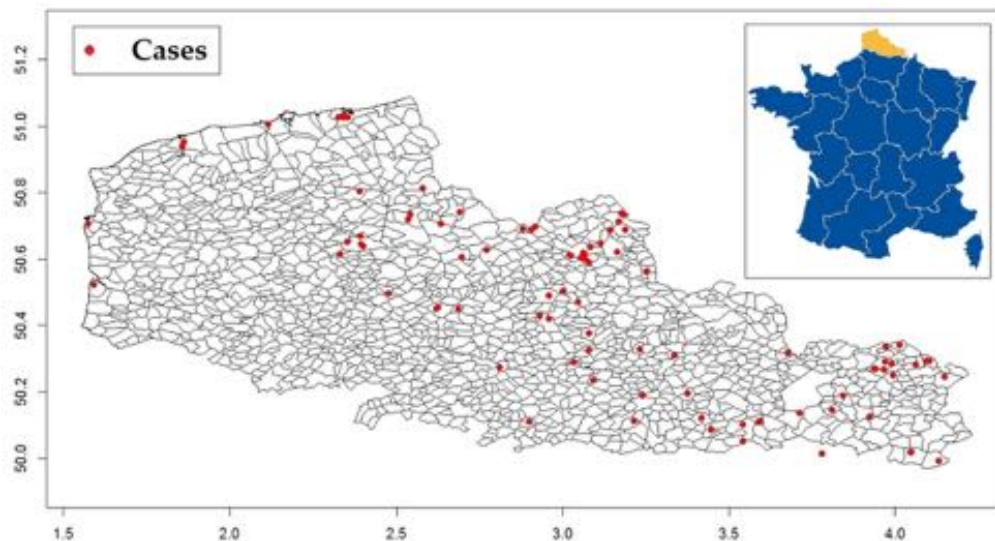


Figure 1: The Nord department in northern France, with 89 of UADT cancer cases.

Bibliography

- [1]Gourieroux, C., Monfort, A., Renault, E. and Trognon, A. (1987), *Generalized residuals*, Journal of econometrics, 34(1), 5-32.
- [2] Imbens, G. W. (1992), *An efficient method of moments estimator for discrete choice models with choice-based sampling*, Econometrica: Journal of the Econometric Society, 1187-1214.
- [3] Klier, T. and D. McMillen (2008), *Clustering of Auto Supplier Plants in the United States*, Journal of Business and Economic Statistics, 26(4), 560-471.
- [4] Manski, C. F. and McFadden, D. (Eds.). (1981), *Structural analysis of discrete data with econometric applications*, (pp. 202-4). Cambridge, MA: MIT Press.
- [5] Pinkse, J. and Slade M.E. (1998), *Contracting in space: An application of spatial statistics to discrete-choice models*, Journal of Econometrics, 85, 125-154.
- [6] Wang, H., Iglesias, E. M., and Wooldridge, J. M. (2013). *Partial maximum likelihood estimation of spatial probit models*. Journal of Econometrics, 172(1), 77-89.

ESTIMATION DE SPECTRES DE SINGULARITÉS PAR DES TECHNIQUES DE GRANDES DÉVIATIONS BASÉES SUR LES ONDELETTES

Céline Esser

Laboratoire Paul Painlevé
Université Lille 1, Villeneuve d'Ascq, France
celine.esser@univ-lille1.fr

Résumé - *La régularité en un point d'une fonction peut être caractérisée par son exposant de Hölder. Si la fonction est très irrégulière, celui-ci peut varier fortement d'un point à l'autre. Dans ce cas, il est naturel d'étudier la distribution des exposants de Hölder : cette information est donnée par le spectre de singularités de la fonction. En général, ce spectre est assez difficile à calculer et on utilise des méthodes pour l'estimer. Plusieurs méthodes basées sur la décomposition en ondelettes ont été proposées ; elles reposent sur une caractérisation de l'exposant de Hölder par les coefficients d'ondelettes.*

Dans cet exposé, nous introduisons une nouvelle technique reposant sur des estimations de grandes déviations basées sur les coefficients d'ondelettes et montrons son efficacité sur plusieurs exemples.

Mots clés - Analyse multifractale, spectre de grandes déviations, ondelettes

1 Introduction

L'objectif de l'analyse multifractale est l'étude de fonctions dont la régularité ponctuelle peut varier d'un point à un autre. Cette régularité est caractérisée par l'exposant de Hölder.

Definition 1 Soient $f \in L_{loc}^{\infty}(\mathbb{R})$, $x_0 \in \mathbb{R}^n$ et $\alpha \geq 0$. La fonction f appartient à l'espace de Hölder $C^{\alpha}(x_0)$ s'il existe un voisinage V de x_0 , une constante $C > 0$ et un polynôme P de degré plus petit que α tels que

$$|f(x) - P(x)| \leq C|x - x_0|^{\alpha} \quad \forall x \in V.$$

L'exposant de Hölder de f en x_0 est défini par

$$h_f(x_0) := \sup \{ \alpha \geq 0 : f \in C^{\alpha}(x_0) \}.$$

L'exposant de Hölder peut varier considérablement d'un point à l'autre. En particulier, les points avec un exposant de Hölder donné peuvent être situés sur des ensembles fractals. Dans ce cas, il est naturel de déterminer leur dimension de Hausdorff.

Definition 2 Le spectre de singularités d'une fonction $f \in L_{loc}^{\infty}(\mathbb{R})$ est la fonction

$$d_f : h \mapsto \dim_{\mathcal{H}} \{ x \in \mathbb{R} : h_f(x) = h \}.$$

Le spectre de singularités de f fournit donc une description géométrique de la répartition des singularités locales de f . Néanmoins, étant donné un signal, le calcul numérique de son spectre s'avère clairement impossible à effectuer directement à partir de sa définition car il implique la détermination successive de plusieurs limites complexes. Par conséquent, on cherche plutôt à l'estimer à partir de quantités qui sont numériquement calculables. Un tel procédé est appelé un formalisme multifractal.

En 1985, Parisi et Frisch proposèrent un premier formalisme multifractal à partir des normes L^p des accroissements du signal [PF85]. Nous présentons ici une variante basée sur une décomposition en ondelettes de la fonction et appelée la *wavelet leaders method* [Jaf04]. L'avantage des ondelettes est qu'elles permettent de définir des méthodes faciles à calculer et relativement stables d'un point de vue numérique. Ces méthodes reposent sur une caractérisation de l'exposant de Hölder via les coefficients d'ondelettes. Néanmoins, l'estimation donnée par la *wavelet leaders method* est obtenue en appliquant une transformée de Legendre et ne permet l'accès qu'à l'enveloppe concave de spectres. Or, il existe des fonctions dont le spectre est non concave. Pour cette raison, nous présentons cet exposé un nouveau formalisme multifractal, appelé *leaders profile method* et basé sur les histogrammes de suprema locaux des coefficients d'ondelettes, qui n'est pas limité aux spectres concaves [BEJ].

2 Ondelettes et exposant de Hölder

Puisque nous sommes intéressés par les propriétés locales de fonctions, nous allons supposer que les fonctions que nous étudions sont 1-périodiques. Nous considérons une mère d'ondelette ψ dans la classe de Schwartz [LM86] et nous notons

$$\psi_{j,k}(x) := \sum_{l \in \mathbb{Z}} \psi(2^j(x-l) - k), \quad j \in \mathbb{N}, k \in \{0, \dots, 2^j - 1\}.$$

Avec la fonction constante égale à 1, les fonctions $2^{j/2}\psi_{j,k}$, $j \in \mathbb{N}$, $k \in \{0, \dots, 2^j - 1\}$ forment une base orthonormée de l'espace des fonctions 1-périodiques de $L^2([0, 1])$ [Dau92, Mal99, Mey90]. Les coefficients d'ondelettes d'une telle fonction f sont définis par

$$c_{j,k} := 2^j \int_0^1 f(x) \psi_{j,k}(x) dx, \quad j \in \mathbb{N}, k \in \{0, \dots, 2^j - 1\}.$$

Remarquons que nous n'avons pas choisi une normalisation L^2 , mais plutôt une normalisation L^∞ qui est plus appropriée pour l'étude de la régularité Hölderienne.

Le coefficient $c_{j,k}$ donne une information sur le comportement de f au voisinage de l'intervalle dyadique

$$\lambda = \lambda(j, k) := \{x \in \mathbb{R} : 2^j x - k \in [0, 1[\} = \left[\frac{k}{2^j}, \frac{k+1}{2^j} \right[, \quad j \in \mathbb{N}, k \in \{0, \dots, 2^j - 1\},$$

près duquel l'ondelette est localisée. Pour tout $j \in \mathbb{N}$, nous dénotons par Λ_j l'ensemble des intervalles dyadiques de $[0, 1[$ de taille 2^{-j} . Afin d'énoncer la propriété qui relie la régularité ponctuelle aux coefficients d'ondelettes, introduisons les coefficients dominants de f , qui peuvent être vus comme des suprema locaux de coefficients d'ondelettes.

Definition 3 Les coefficients dominants de f sont définis par

$$d_{j,k} = d_\lambda := \sup_{\lambda' \subset 3\lambda} |c_{\lambda'}|, \quad \lambda \in \Lambda_j, j \in \mathbb{N}.$$

Si $x \in [0, 1[$, alors pour tout $j \in \mathbb{N}$, il existe un unique $\lambda \in \Lambda_j$ tel que $x \in \lambda$. Nous notons $d_j(x)$ le coefficient dominant associé.

Proposition 1 ([Jaf04]) Si f est uniformément Hölderienne, i.e. s'il existe $\varepsilon > 0$ et $C > 0$ tels que

$$|f(x) - f(y)| \leq C|x - y|^\varepsilon \quad \forall x, y,$$

alors l'exposant de Hölder de f en x est donné par

$$h_f(x) = \liminf_{j \rightarrow \infty} \frac{\log d_j(x)}{\log 2^{-j}}.$$

3 Wavelet leaders method

La *wavelet leaders method* repose sur les valeurs de la fonction d'échelle η_f définie par

$$\eta_f(q) := \liminf_{j \rightarrow +\infty} \frac{\log 2^{-nj} \sum_{\lambda \in \Lambda_j} d_\lambda^q}{\log 2^{-j}},$$

où la somme est prise sur intervalles λ tels que $d_\lambda \neq 0$. Le spectre de singularités de f est alors estimé par la fonction

$$L_f : q \mapsto \inf_q \{hq - \eta_f(q)\} + 1.$$

L'égalité entre d_f et L_f n'est pas toujours vérifiée, mais cette formule est satisfaite pour de nombreuses fonctions et processus aléatoires. De plus, la *wavelet leaders method* donne toujours une borne supérieure pour le spectre, i.e. $d_f \leq L_f$ [Jaf04]. Néanmoins, la limitation de cette méthode est, comme mentionné précédemment, que l'estimation donnée est nécessairement concave.

4 Leaders profile method

La *leaders profile method* repose sur des estimations de grandes déviations basées sur les coefficients dominants. Plus précisément, on considère les fonctions

$$\nu_f^+(h) = \lim_{\varepsilon \rightarrow 0^+} \limsup_{j \rightarrow +\infty} \frac{\log \#\{\lambda \in \Lambda_j : d_\lambda \geq 2^{-(h+\varepsilon)j}\}}{\log 2^j}$$

et

$$\nu_f^-(h) = \lim_{\varepsilon \rightarrow 0^+} \limsup_{j \rightarrow +\infty} \frac{\log \#\{\lambda \in \Lambda_j : d_\lambda \leq 2^{-(h+\varepsilon)j}\}}{\log 2^j}.$$

Approximativement, il existe donc $2^{\nu_f^+(h)j}$ (resp. $2^{\nu_f^-(h)j}$) coefficients dominants d_λ , $\lambda \in \Lambda_j$, plus grands (resp. plus petits) que 2^{-hj} pour les grandes échelles j .

Le spectre multifractal de f est alors estimé par la fonction

$$\nu_f := \min\{\nu_f^+, \nu_f^-\},$$

appelée *profil des coefficients dominants* [BEJ].

Proposition 2 ([BEJ]) *La fonction ν_f ne dépend pas de la base d'ondelettes choisie pour la calculer et on a $d_f \leq \nu_f$.*

De plus, d'un point de vue théorique, la *leaders profile method* est plus efficace que la *wavelet leaders method*. Plus précisément, nous avons le résultat suivant.

Proposition 3 ([EKN]) *Si ν_f prend la valeur $-\infty$ en dehors d'un ensemble compact de $[0, +\infty[$, alors*

$$\eta_f(q) = \inf_{h \in \mathbb{R}} \{hq - \nu_f(h)\} + 1, \quad \forall q \in \mathbb{R}.$$

En particulier, $\nu_f(h) \leq L_f(h)$ pour tout $h \in \mathbb{R}$ et L_f est l'enveloppe concave de ν_f .

5 Illustrations numériques

5.1 Mouvement Brownien fractionnaire

Considérons un mouvement Brownien fractionnaire $\{B_H(t)\}_{t \in [0,1]}$ de paramètre de Hurst $H \in]0, 1[$. Presque sûrement, l'exposant de Hölder de B_H est constant et égal à H [Kah93]. Les méthodes présentées permettent donc d'obtenir une estimation du paramètre du Hurst.

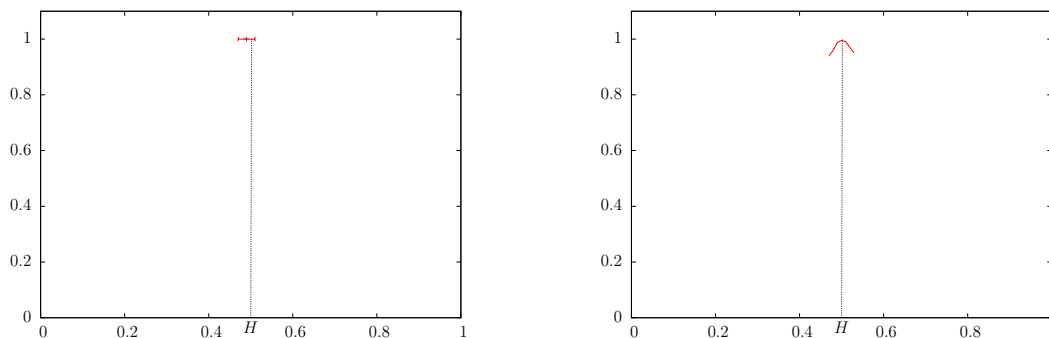


FIGURE 1 – Estimation du spectre d'un mouvement Brownien par la *wavelet leaders method* à gauche et la *leaders profile method* à droite.

5.2 Processus de Lévy

Les propriétés multifractales des processus de Lévy ont été étudiées par Jaffard [Jaf99]. Si le processus X ne possède pas de partie Brownienne, le spectre des singularités est presque sûrement donné par

$$d_X(h) = \beta h, \quad \forall h \in [0, 1/\beta].$$

D'autre part, si le processus a une partie Brownienne, le spectre des singularités est presque sûrement donné par

$$d_X(h) = \begin{cases} \beta h & \text{if } h \in [0, 1/2[, \\ 1 & \text{if } h = 1/2. \end{cases}$$

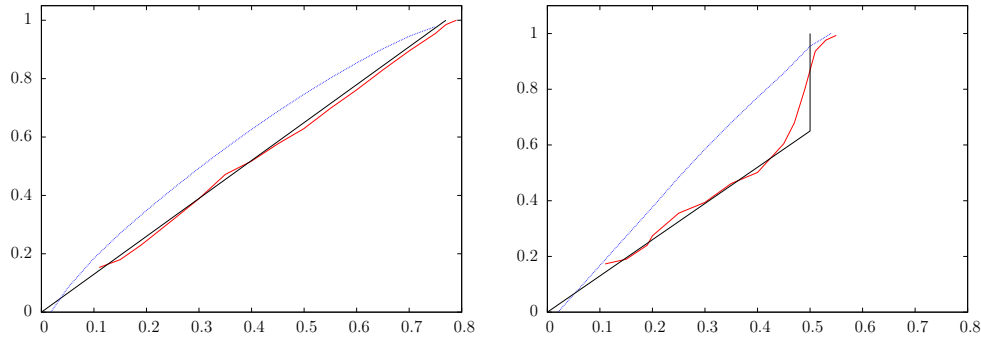


FIGURE 2 – Spectre d'un processus de Lévy sans partie Brownienne à gauche, avec partie Brownienne à droite. En bleu, estimation par la *wavelet leaders method*. En rouge, estimation par la *leaders profile method*.

Références

- [BEJ] F. Bastin, C. Esser, and S. Jaffard. Large deviation spectra based on wavelet leaders. Submitted for publication.
- [Dau92] I. Daubechies. *Ten Lectures on Wavelets*. CBMS-NSF Regional Conference Series in Applied Mathematics, 1992.
- [EKN] C. Esser, T. KleynstSENS, and S. Nicolay. A multifractal formalism for non-concave and non-increasing spectra : the leaders profile method. Submitted for publication.
- [Jaf99] S Jaffard. The multifractal nature of Lévy processes. *Probab. Theory Relat. Fields*, 114 :207–227, 1999.
- [Jaf04] S. Jaffard. Wavelet techniques in multifractal analysis, fractal geometry and applications : A jubilee of Benoit Mandelbrot. *Proceedings of Symposia in Pure Mathematics*, 72 :91–151, 2004.
- [Kah93] J.-P. Kahane. *Some Random Series of Functions*. Cambridge University Press, 1993.
- [LM86] P.G. Lemarié and Y. Meyer. Ondelettes et bases hilbertiennes. *Revista Math.Iberoamericana*, 1 :1–18, 1986.
- [Mal99] S. Mallat. *A Wavelet Tour of Signal Processing*. Academic Press, 1999.
- [Mey90] Y. Meyer. *Ondelettes et opérateurs*. Hermann, 1990.
- [PF85] G. Parisi and U. Frisch. On the singularity structure of fully developed turbulence. *Turbulence and Predictability in Geophysical Fluid Dynamics*, pages 84–87, 1985.

FRÉCHET MEANS ON QUANTUM GRAPHS

Gavra Ioana

Université Paul Sabatier

Toulouse, France

igavra@math.univ-toulouse.fr

Résumé - *There is a lot of data whose structure is organized in the same way as an weighted graph. For example we can view the set of the Netflix movies as the vertex set of a graph. We have an edge between two movies if there is an actor that played in both of them. The weight of the edge is proportional to the number of actors who played in both movies. In studying the geometry of discrete graphs we can get more information about some data structures.*

Let $G = (V, E, W)$ be a weighted graph. V represents the set of vertices, E the set of edges and W the weights of the edges. We define the length of an edge $e_{i,j}$ as $\frac{1}{w_{i,j}}$. The distance d between two vertices is defined in the natural manner as the length of the shortest path between them. We also introduce ν , a probability measure on V . In the above example this can represent the users ratings. We are interested in finding the average film given the ratings at a certain moment and also taking into account the links between different movies. In order to do that we introduce the function U_ν on G :

$$U_\nu(x) = \int d^2(x, y)\nu(y).$$

Let M be the set of global minimums of U_ν . The Frechet mean will be an element of this set. We have chosen the square of d because minimizing the same function for a probability distribution on \mathbb{R} would give us its expectation.

Mots clés - **Quantum Graphs, Simulated Annealing, Entropy**

1 Introduction

The classical way to minimize a function on a discrete space is to use the simulated annealing algorithm. We give a brief presentation of the classical algorithm.

Let U be a function on V , the vertex set of a graph. We want to find the global minimum of U . Let L be a transition matrix on V such that $L(x, x') > 0$ if and only if x' is a neighbor of x . Let T_k be a decreasing sequence of strictly positive numbers such that T_k goes to zero when k goes to infinity. The algorithm goes as follows:

We fix X_0 an initial state.

For $k \geq 1$, suppose $X_k = x$. A potential state $X'_k = x'$ is chosen for X_{k+1} with probability $L(x, x')$.

Now we define X_{k+1} in the following way:

$$X_{k+1} = \begin{cases} X'_k & \text{with probability } p_k \\ X_k & \text{with probability } 1 - p_k \end{cases}$$

where

$$p_k = \exp \frac{-[U(X'_k) - U(X_k)]_+}{T_k}.$$

The way p_k is defined tells us that if $U(X'_k) < U(X_k)$ we accept X'_k with probability 1. It is shown that if the sequence T_k is well chosen than X_k will converge to the set of global minimums of U .

We are interested in finding the minimum value of U_ν in the case where the actual distribution of ν is not available, but we have access to a sequence $(Y_n)_{n \in \mathbb{N}}$, of i.i.d. random variables distributed according to ν . Since we cannot compute the exact value of the function that we want to minimize we cannot apply the classical algorithm.

A possible approach to this problem is to use the homogenization technique: replacing U_ν in the definition of p_k by $U_y(\cdot) = d^2(\cdot, y)$, where y is a value from the sequence Y_n and to change the y we use with an increasing speed.

To resume X_k given by the classical algorithm is a time-inhomogeneous Markov chain with a probability transition of the form:

$$L_{T_k}(x, x') = \exp\left[-\frac{[(U_\nu(x) - U_\nu(x'))_+]}{T_k}\right] L(x, x')$$

and the process given by the homogenization if of the form:

$$L_{T_k}^y(x, x') = \exp\left[-\frac{[(U_y(x) - U_y(x'))_+]}{T_k}\right] L(x, x').$$

Since the exponential is not a linear function it is difficult to verify under what conditions $L_{T_k}^y$ will approach L_{T_k} when y changes at a high speed.

This difficulty doesn't arise in the homogenization of the simulated annealing algorithm on the continuous case, since the exponential is replaced by a gradient. That is why we prefer to work on a quantum graph induced by the initial weighted graph.

1.1 Quantum graph

We can look at G as a collection of segments where we identify the ends of the edges corresponding to the same vertex. From now on we denote Γ the quantum graph that maintains the same geometric structure as G , but includes as points of its space not only the vertices but the points of the edges too. The length of each edge remains the same as before, and the distance d between to points remains the length of the shortest path between them.

We also extend the definition of the function U_ν on Γ . It is obvious that the global minimum of U_ν will no longer be necessarily a vertex.

1.2 Continuous case

We have access to a sequence of random variables $(Y_n)_{n \geq 0}$ and we need a continuous time process. In order to obtain one we will apply a method similar to the one used for the continuous-time random walks.

Let H_y be a Poisson process of parameter 1 and α_t a continuous function. Let h be a function defined by $h(t) = \int_0^t \alpha_s ds$. We define the accelerated process :

$$N_t^\alpha = H_{h(t)}.$$

This process represents the speed at which we will use the sequence $(Y_n)_{n \in \mathbb{N}}$.

The process we will use in the algorithm will be of the form: $dX_t = -\beta_t \nabla U_y(X_t) dt + dB_t$. One of the difficulties that appear is to give a meaning to dB_t . We will name a Brownian motion on Γ a process which on the interior of an edge evolves as a real Brownian motion and every time it gets to a vertex it chooses one of the possible directions with equal probability. Another important point is to choose a 'good' β_t . We will see more about that later.

1.3 The algorithm

Let X_t be a process that starts evolving as a Brownian Motion on the graph, but at each jumping time given by N_t it is drawn towards a new Y_{N_t} proportionally to a function β_t .

Description of the algorithm:

Step 1: At time $t=0$ $X_t=X_0$ a random value on the graph.

Step 2: X_t evolves as a Brownian motion on the graph.

Step 3: At time $t=T_1$: $Y_t=Y_1$, the first value from the sequence Y_n ;
 X_t makes a jump towards Y_t of length proportional to
the distance between them.

Step 4: We repeat Step 2 and 3, going through the sequence Y_n .

2 Technical Details

2.1 Laplacian on the Graph

In order to have an operator which is compatible with the classical Laplacian on a weighted graph and also defined on the interior of the edges, the Laplacian of a function on a quantum graph will be a signed bounded measure. We have used the definitions and notations from [1].

We fix a parametrisation of each edge; $s_e : [0, L_e] \rightarrow e$ such that $s(0)$ is one vertex of the edge and $s(L_e)$ is the other one. In the interest of simplifying the notations, for any given function f we denote :

$$f_e(t) = f(s_e(t)).$$

Definition 1 Let p be a point in the interior of an edge e of length L . Let f be a function on the graph. We say that f is differentiable at p if the quantity: $\left. \frac{d}{dt} f_e(t) \right|_{s_e(t)=p}$ exists.

For all $p \in V$ we denote E_p the set of edges containing p .

Definition 2 Let p be a vertex of the graph Γ and an edge $e \in E_p$. We define the derivative of f at p in the direction e as:

$$D_e f(p) = \lim_{h \rightarrow 0} \frac{f(p) - f(s(h))}{h}$$

if $p = s(0)$ and:

$$D_e f(p) = \lim_{h \rightarrow 0} \frac{f(p) - f(s(L_e - h))}{h}$$

else.

Definition 3 The Laplacian of a smooth enough function f is defined by:

$$\Delta f = f'' \mathbf{1}_{\Gamma \setminus V} + \sum_{p \in V} \sigma_p(f) \delta_p$$

where δ_p is the Dirac measure on p , $\sigma_p(f) = \sum_{e \in E_p} D_e f(p)$.

2.2 Convergence of the algorithm

The stochastic process (X_t, Y_t) given by the algorithm described in the section 1.3 is in fact the inhomogeneous Markov process given by:

$$\mathcal{L}_t f(x, y) = \Delta_x f(x, y) - \beta_t \langle \nabla_x U_y, \nabla_x f(x, y) \rangle + \alpha_t \int f(x, y') - f(x, y) \mu(dy'). \quad (1)$$

For more details see [2].

Here X_t represents the position on the graph at time t and Y_t represents the value of the initial sequence Y_n at the step given by the value of N_t .

We denote m_t the distribution of (X_t, Y_t) at time t , n_t the distribution of X_t and $m_t(y|x)$ the conditional distribution of Y_t given X_t .

Let μ_{β_t} be the distribution of the Gibbs measure with energy U_ν and inverse temperature β_t :

$$\mu_{\beta_t} = \frac{1}{Z_{\beta_t}} \exp(-\beta_t U_\nu(x)),$$

where Z_{β_t} is the normalization factor, ie $Z_{\beta_t} = \int e^{-\beta_t U_\nu(x)} dx$.

If β_t converges to infinity, μ_{β_t} will concentrate on M_ν . We define

$$\alpha_t = t + 1 \text{ and } \beta_t = b \log(t + 1),$$

with b a constant smaller than $\frac{1}{C}$, where C is the maximal depth of a well containing a local but not global minimum of U_ν .

Théorème 1 *Let J_t be the relative entropy of the probability distribution of the stochastic process X_t given by (1) and μ_{β_t} , the Gibbs distribution of inverse temperature β_t :*

$$J_t = \int \log \frac{n_t(x)}{\mu_{\beta_t}(x)} dn_t(x).$$

Then for α_t and β_t defined as above, J_t goes to 0 when t goes to infinity.

The method used to prove this theorem is inspired from [4]. Also an important step in the proof of this theorem is the existence of a Sobolev inequality for μ_{β_t} . To obtain this we have adapted the method given in [3].

3 Further Developments

The minimum of U_ν on the quantum graph is not necessarily on a vertex, therefore it's different from the minimum of U_ν in the discrete case. In order to resolve the initial question it would be interesting to find a viable form of the homogenization method for the discrete case. The main difficulty remains the fact that if Y is a random variable of law ν : $\mathbb{E}_Y[\exp[-\frac{[(U_Y(x)-U_y(x'))_+]}{T}]] \neq \exp[-\frac{[(U_\nu(x)-U_\nu(x'))_+]}{T}]]$.

References

- [1] M. Baker and X. Faber (2005),
Metrized Graphs, Laplacian Operators and Electrical Networks,
American Mathematical Society,
- [2] M. Freidlin and A. Wentzell (1993),
Diffusion Processes on Graphs and the Averaging Principle,
The Annals of Probability,
vol. 21
p. 2215-2245.
- [3] R. Holley, and D. Stroock (1988),
Simulated Annealing via Sobolev Inequalities,
Communications in Mathematical Physics,
vol. 115
p. 553-569.
- [4] Laurent Miclo (1996),
Recuit simulé partiel,
Stochastic Processes and their Applications,
vol. 65
p. 281-298.

ESTIMATION RÉCURSIVE DE LA MEDIAN COVARIATION MATRIX DANS LES ESPACES DE HILBERT

Antoine Godichon-Baggioni

Université de Bourgogne
9 Avenue Alain Savary 21000 Dijon
antoine.godichon@u-bourgogne.fr

Résumé - *La médiane géométrique est souvent utilisée en statistique du fait de sa robustesse. De plus, il est de plus en plus fréquent d'analyser de grands échantillons à valeurs dans des espaces de grandes dimensions. Dans ce contexte, des estimateurs rapides de la médiane ont été introduits par [Cardot et al., 2013], qui consistent en un algorithme de gradient stochastique et son moyenné. On se propose de donner une étude approfondie de ces estimateurs. On définira aussi un nouvel indicateur robuste de dispersion (en lien avec la médiane) appelé Median Covariation Matrix (MCM) avant d'en donner des estimateurs. Cet indicateur peut être très intéressant pour l'Analyse des Composantes Principales. En effet, sous certaines conditions, cette matrice a les mêmes sous-espaces propres que la matrice de covariance, mais est beaucoup moins sensible aux données contaminées.*

Mots clés - Données fonctionnelles, Médiane Géométrique, Estimation récursive

1 Introduction

La médiane géométrique est une généralisation naturelle de la médiane réelle introduite par [Haldane, 1948]. En dimension finie, on peut faire le lien avec le problème de Fermat-Webber qui consiste à minimiser la somme des distances à des points donnés et qui est un problème d'optimisation convexe bien connu. Beaucoup de propriétés de la médiane dans les espaces de Banach sont données par [Kemperman, 1987], telles que son existence, son unicité et sa robustesse. Cette dernière propriété représente l'un des principaux facteurs d'intérêt de la médiane. De plus, [Chakraborty and Chaudhuri, 2014] proposent une étude approfondie des estimateurs pour le cas général des espaces de Banach.

De manière générale, beaucoup d'estimateurs de la médiane existent dans la littérature, et un des plus utilisés en dimension finie est celui introduit par [Vardi and Zhang, 2000], qui consiste à résoudre le problème de Fermat-Webber généré par l'échantillon à l'aide de l'algorithme de Weizfeld. Cependant, cet algorithme peut être difficile à compiler lorsque l'on a de grands échantillons à valeurs dans des espaces de grandes dimensions. C'est pourquoi on s'intéressera aux algorithmes introduits par [Cardot et al., 2013], qui consistent en un algorithme de gradient stochastique et son moyenné.

De plus, on s'intéressera à la Median Covariation Matrix (MCM) étudiée par [Kraus and Panaretos, 2012]. C'est un indicateur de dispersion robuste en lien avec la médiane, qui peut être utilisé pour l'ACP robuste ([Cardot and Godichon, 2015]). En effet, sous certaines conditions, si la distri-

bution des données que l'on veut étudier est symétrique, alors la MCM a les mêmes espaces propres que la matrice de covariance. On introduira donc un algorithme capable d'estimer cet indicateur quelles que soient les tailles de l'échantillon et de l'espace. Cet algorithme consiste à estimer simultanément la médiane et la MCM à l'aide d'algorithmes de gradient stochastiques et de leurs moyennés.

Dans ce papier, on définira d'abord la médiane et on donnera quelques hypothèses avant de rappeler les algorithmes introduits par [Cardot et al., 2013]. On donnera aussi quelques résultats non asymptotiques dûs à [Cardot et al., 2015] et [Godichon, 2015]. Finalement, on définira la Median Covariation Matrix et on donnera les algorithmes récursifs introduits par [Cardot and Godichon, 2015] avant d'en donner les vitesses de convergence.

2 Estimation de la médiane géométrique

2.1 Définition et hypothèses

On considère une variable aléatoire X à valeurs dans un espace de Hilbert séparable H (pas nécessairement de dimension fini). On note $\langle \cdot, \cdot \rangle$ le produit scalaire et $\|\cdot\|$ la norme associée. La médiane géométrique m de X est définie par

$$m := \arg \min_{h \in H} \mathbb{E} [\|X - h\| - \|X\|].$$

On introduit maintenant deux hypothèses :

(A1) X n'est pas concentrée sur une droite : pour tout $h \in H$, il existe $h' \in H$ tel que $\langle h, h' \rangle$ et

$$\text{Var}(\langle X, h' \rangle) > 0.$$

(A2) X n'est pas concentrée autour de points isolés : il existe une constante positive C telle que pour tout $h \in H$,

$$\mathbb{E} \left[\frac{1}{\|X - h\|^2} \right] \leq C.$$

L'hypothèse **(A1)** assure l'unicité de la médiane ([Kemperman, 1987]), et l'hypothèse **(A2)** permet de donner quelques propriétés sur la fonction que l'on veut minimiser.

2.2 Les algorithmes

Soit G la fonction que l'on veut minimiser. Elle est définie pour tout $h \in H$ par

$$G(h) := \mathbb{E} [\|X - h\| - \|X\|].$$

Sous les hypothèses, G est convexe et est deux fois différentiable. Son gradient est défini pour tout $h \in H$ par

$$\nabla G(h) = -\mathbb{E} \left[\frac{X - h}{\|X - h\|} \right].$$

Cela légitimise l'utilisation d'un algorithme de gradient stochastique (ou de Robbins-Monro). On considère maintenant des variables aléatoires indépendantes X_1, \dots, X_n, \dots de mêmes lois

que X . On rappelle maintenant l'algorithme de type Robbins-Monro introduit par [Cardot et al., 2013] et défini de manière récursive par

$$m_{n+1} = m_n + \gamma_n \frac{X_{n+1} - m_n}{\|X_{n+1} - m_n\|}, \quad (1)$$

avec m_1 borné. La suite de pas (γ_n) est une suite de réels positifs, décroissante, et vérifie les conditions usuelles suivantes (voir [Duflo, 1997] par exemple)

$$\sum_{n \geq 1} \gamma_n = +\infty, \quad \sum_{n \geq 1} \gamma_n^2 < \infty. \quad (2)$$

La version moyennée de l'algorithme est définie récursivement par

$$\bar{m}_{n+1} = \bar{m}_n + \frac{1}{n+1} (m_{n+1} - \bar{m}_n), \quad (3)$$

avec $\bar{m}_0 = 0$, ce qui peut s'écrire $\bar{m}_n = \frac{1}{n} \sum_{k=1}^n m_k$.

2.3 Vitesses de convergence

La forte consistance de ces algorithmes est donnée par [Cardot et al., 2013]. On considère maintenant une suite de pas de la forme $\gamma_n := c_\gamma n^{-\alpha}$, avec $c_\gamma > 0$ et $\alpha \in (1/2, 1)$. Le théorème suivant donne alors les vitesses de convergence de l'algorithme de type Robbins-Monro.

Théorème 1 *On suppose que les hypothèses (A1) et (A2) sont vérifiées. Il existe une constante positive C' telle que pour tout $n \geq 1$,*

$$\mathbb{E} \left[\|m_n - m\|^2 \right] \leq \frac{C'}{n^\alpha}.$$

Plus généralement, pour tout entier $p \geq 2$, il existe une constante positive K_p telle que pour tout $n \geq 1$,

$$\mathbb{E} \left[\|m_n - m\|^{2p} \right] \leq \frac{K_p}{n^{p\alpha}}.$$

Grâce aux vitesses précédentes, on peut obtenir les vitesses pour l'algorithme moyenné.

Théorème 2 *On suppose que les hypothèses (A1) et (A2) sont vérifiées. Il existe une constante positive C'' telle que pour tout $n \geq 1$,*

$$\mathbb{E} \left[\|\bar{m}_n - m\|^2 \right] \leq \frac{C''}{n}.$$

Plus généralement, pour tout entier $p \geq 1$, il existe une constante positive $K_{p'}$ telle que pour tout $n \geq 1$,

$$\mathbb{E} \left[\|\bar{m}_n - m\|^{2p} \right] \leq \frac{K_{p'}}{n^p}.$$

3 Estimation de la Median Covariation Matrix

3.1 Définition et hypothèses

On considère maintenant un espace de Hilbert séparable H et l'espace des opérateurs linéaires de H dans H , noté $\mathcal{S}(H)$. Soit $(e_j)_{j \in J}$ une base de H , on équipe $\mathcal{S}(H)$ avec le produit scalaire suivant : soient $A, B \in \mathcal{S}(H)$,

$$\langle A, B \rangle_F = \sum_{j \in J} \langle A(e_j), B(e_j) \rangle.$$

Alors, $\mathcal{S}(H)$ est aussi un espace de Hilbert séparable et la norme associée au produit scalaire précédent, notée $\|\cdot\|_F$, est la norme de Hilbert-Schmidt (ou Froebenius).

Soit X une variable aléatoire à valeurs dans H , la Median Covariation Matrix Γ_m de X est définie par

$$\Gamma_m := \arg \min_{V \in \mathcal{S}(H)} \mathbb{E} [\| (X - m)^T(X - m) - V \|_F - \| (X - m)^T(X - m) \|_F], \quad (4)$$

où m est la médiane de X . La Median Covariation Matrix Γ_m peut être vue comme la médiane géométrique de la variable aléatoire $(X - m)^T(X - m)$, et est donc robuste. De la même façon que pour la médiane, on introduit maintenant deux hypothèses :

(A3) Pour tout $V \in \mathcal{S}(H)$, il existe $V' \in \mathcal{S}(H)$ tel que $\langle V, V' \rangle_H = 0$ et

$$\text{Var} (\langle (X - m)^T(X - m), V' \rangle_F) > 0.$$

(A4) Il existe une constante positive C telle que pour tout $h \in H$ et $V \in \mathcal{S}(H)$,

$$\mathbb{E} \left[\frac{1}{\| (X - h)^T(X - h) - V \|_F^2} \right] \leq C.$$

Remarquons que l'hypothèse **(A4)** implique l'hypothèse **(A2)**. Sous les hypothèses **(A1)** et **(A3)**, la Median Covariation Matrix est bien définie et est unique.

3.2 Les algorithmes

Dans le cas où la médiane est connue, les algorithmes et leurs propriétés asymptotiques sont analogues à ceux pour l'estimation de la médiane. On suppose maintenant que m n'est pas connue et pour tout $h \in H$, soit G_h la fonction définie pour tout $V \in \mathcal{S}(H)$ par

$$G_h(V) := \mathbb{E} [\| (X - h)^T(X - h) - V \|_F - \| (X - h)^T(X - h) \|_F].$$

Ces fonctions sont différentiables et leurs gradients sont définis pour tout $V \in \mathcal{S}(H)$ par

$$\nabla G_h(V) = -\mathbb{E} \left[\frac{(X - h)^T(X - h) - V}{\| (X - h)^T(X - h) - V \|_F} \right].$$

On peut maintenant introduire un algorithme de gradient stochastique et son moyenné. Soient X_1, \dots, X_n, \dots des variables aléatoires indépendantes de mêmes lois que X . L'algorithme de

type Robbins-Monro et son moyenné (V_n et \bar{V}_n) sont définis récursivement par

$$m_{n+1} = m_n + \gamma_n \frac{X_{n+1} - m_n}{\|X_{n+1} - m_n\|}, \quad (5)$$

$$\bar{m}_{n+1} = \bar{m}_n + \frac{1}{n+1} (m_{n+1} - \bar{m}_n), \quad (6)$$

$$V_{n+1} = V_n + \gamma_n \frac{(X_{n+1} - \bar{m}_n)^T (X_{n+1} - \bar{m}_n) - V_n}{\|(X_{n+1} - \bar{m}_n)^T (X_{n+1} - \bar{m}_n) - V_n\|_F}, \quad (7)$$

$$\bar{V}_{n+1} = \bar{V}_n + \frac{1}{n+1} (V_{n+1} - \bar{V}_n), \quad (8)$$

avec m_1 et V_1 bornés, et $\bar{m}_0 = 0$, $\bar{V}_0 = 0$.

3.3 Vitesses de convergence

On considère maintenant une suite de pas de la forme $\gamma_n := c_\gamma n^{-\alpha}$, avec $c_\gamma > 0$ et $\alpha \in (1/2, 1)$. On a alors les vitesses de convergence suivantes pour l'algorithme de type Robbins-Monro.

Théorème 3 *On suppose que les hypothèses (A1) à (A4) sont vérifiées. Il existe une constante positive K telle que pour tout $n \geq 1$,*

$$\mathbb{E} \left[\|V_n - \Gamma_m\|_F^2 \right] \leq \frac{K}{n^\alpha}.$$

De plus, pour tout $\beta \in (\alpha, 2\alpha)$, il existe une constante positive K_β telle que pour tout $n \geq 1$,

$$\mathbb{E} \left[\|V_n - \Gamma_m\|_F^4 \right] \leq \frac{K_\beta}{n^\beta}.$$

Finallement, le théorème suivant donne la vitesse de convergence en moyenne quadratique de l'algorithme moyenné.

Théorème 4 *On suppose que les hypothèses (A1) à (A4) sont vérifiées. Il existe une constante positive K telle que pour tout $n \geq 1$,*

$$\mathbb{E} \left[\|\bar{V}_n - \Gamma_m\|_F^2 \right] \leq \frac{K'}{n}.$$

Références

- [Cardot et al., 2015] Cardot, H., Cénac, P., and Godichon, A. (2015). Online estimation of the geometric median in Hilbert spaces : non asymptotic confidence balls. Technical report, arXiv :1501.06930.
- [Cardot et al., 2013] Cardot, H., Cénac, P., and Zitt, P.-A. (2013). Efficient and fast estimation of the geometric median in Hilbert spaces with an averaged stochastic gradient algorithm. *Bernoulli*, 19 :18–43.
- [Cardot and Godichon, 2015] Cardot, H. and Godichon, A. (2015). Robust principal components analysis based on the median covariation matrix. *arXiv :1504.02852*.
- [Chakraborty and Chaudhuri, 2014] Chakraborty, A. and Chaudhuri, P. (2014). The spatial distribution in infinite dimensional spaces and related quantiles and depths. *The Annals of Statistics*, 42 :1203–1231.

- [Duflo, 1997] Duflo, M. (1997). *Random iterative models*, volume 34 of *Applications of Mathematics (New York)*. Springer-Verlag, Berlin. Translated from the 1990 French original by Stephen S. Wilson and revised by the author.
- [Godichon, 2015] Godichon, A. (2015). Estimating geometric median in hilbert spaces with stochastic gradient algorithms ; l^p and almost sure rates of convergence. *arXiv :1504.02267*.
- [Haldane, 1948] Haldane, J. B. S. (1948). Note on the median of a multivariate distribution. *Biometrika*, 35(3-4) :414–417.
- [Kemperman, 1987] Kemperman, J. H. B. (1987). The median of a finite measure on a Banach space. In *Statistical data analysis based on the L_1 -norm and related methods (Neuchâtel, 1987)*, pages 217–230. North-Holland, Amsterdam.
- [Kraus and Panaretos, 2012] Kraus, D. and Panaretos, V. M. (2012). Dispersion operators and resistant second-order functional data analysis. *Biometrika*, 99 :813–832.
- [Vardi and Zhang, 2000] Vardi, Y. and Zhang, C.-H. (2000). The multivariate L_1 -median and associated data depth. *Proc. Natl. Acad. Sci. USA*, 97(4) :1423–1426.

Inégalité de concentration et vitesse de la convergence de marche aléatoire en potentiel aléatoire

LE Thi Thu Hien

Univeristé Paris-Est Créteil Val de Marne

Laboratoire d'Analyse et Matheématiques Appliquées

Thi-Thu-Hien.vo@u-pec.fr

Résumé- *Nous nous intéressons à la marche aléatoire dans un potentiel aléatoire non-négatif sur \mathbb{Z}^d . Nous étudions l'inégalité de concentration et la vitesse de convergence de l'exposant de Lyapunov pour ce modèle.*

Mots clés marche aléatoire, potentiel aléatoire, exposant de Lyapunov; vitesse de convergence; inégalité de concentration;

1 Introduction

Nous présentons un modèle de marche aléatoire en potentiel aléatoire i.i.d positif pour lequel [4] a obtenu un résultat de grandes déviations. Pour ce type de modèle en milieux aléatoires proche de la percolation de premier passage ou de polymère en milieux aléatoires, les techniques utilisées marient souvent propriété de Markov forte et théories ergodiques sous-additifs.

Au cours de cet exposé, nous présenterons tout d'abord le modèle et les résultats de Zerner [4] décrions ensuite un TCL en dimension 1 puis une inégalité de concentration et la vitesse de la convergence (prouvées dans [3]) qui constituent un premier pas vers un TCL en dimension supérieure.

1.1 Modèle

Soit $S_n, n \in \mathbb{N}$ une marche aléatoire simple sur \mathbb{Z}^d . Notons P_x et E_x la loi de la marche et l'espérance, respectivement, de la marche aléatoire simple partant de $x \in \mathbb{Z}^d$. Indépendamment de la marche aléatoire, nous nous donnons une famille de variables aléatoires i.i.d $V(x, \omega) \geq 0, x \in \mathbb{Z}^d$, les *potentiels*, sur un espace $(\Omega, \mathcal{F}, \mathbb{P})$. Dans [4], Zerner s'intéresse au comportement asymptotique de

$$a(x, y, \omega) := -\ln e(x, y, \omega)$$

où
$$e(x, y, \omega) := E_x \left(\exp \left(- \sum_{m=0}^{H(y)-1} V(S_m, \omega) \right) \right), H(y) < \infty$$

avec
$$H(y) := \inf \{ n \geq 0 : S_n = y \},$$

En ajoutant un état "cimetière", $e(x, y, \omega)$ peut s'interpréter comme la probabilité q'une marche avec "cimetière". La propriété de Markov forte permet d'obtenir une décroissance exponentielle justifiant l'introduction de $a(\cdot, \cdot, \omega)$. Cette quantité vérifie une inégalité triangulaire et en utilisant le théorème sous-additif de Kingman, Zerner [4] a montré l'existence d'*exposant de Lyapunov (quenched)* $\alpha(x)$.

Théorème A ([4]) Soit $\mathbb{E}V(0) < \infty$. Alors, il existe une norme non aléatoire sur \mathbb{R}^d qui ne dépend que de la loi du potentiel et de la dimension d et qui est notée $\alpha(\cdot)$ telle que \mathbb{P} -p.s et dans $L^1(\mathbb{P})$, pour tout $x \in \mathbb{Z}^d$:

$$\lim_{n \rightarrow \infty} \frac{1}{n} a(0, nx, \omega) = \lim_{n \rightarrow \infty} \frac{1}{n} \mathbb{E}[a(0, nx)] = \inf_{n \in \mathbb{N}} \frac{1}{n} \mathbb{E}[a(0, nx)] = \alpha(x).$$

1.2 Inégalité de concentration et vitesses de la convergence

L'étude du théorème de centrale limite se pose alors naturellement pour notre modèle. Pour la dimension $d = 1$, le TCL est prouvé dans [4],

$$\frac{a(0, n) - \mathbb{E}a(0, n)}{\sigma\sqrt{n}} \rightarrow N(0, 1) \quad \text{en distribution} \quad (1)$$

En dimension plus grande, pour obtenir la bonne normalisation, on définit *l'exposant critique* χ de tel sorte que $|x|^\chi$ est intuitivement l'ordre de grandeur de la fluctuation de $a(0, x)$ par rapport à sa moyenne. Pour les modèles où le comportement asymptotique est gaussien, χ vaut $1/2$ (comme (1)). Cependant pour quelques modèles reliés, on s'attend à montrer que χ satisfasse la relation du KPZ (avec même $\chi = \frac{1}{3}$ quand $d = 2$) (voir [1] pour percolation de premier passage et [2] pour plusieurs modèles des aléatoires métriques). Comme un premier pas vers une résolution de ce problème, nous donnons ici une inégalité de concentration et la vitesse de la convergence des exposants de Lyapunov.

Hypothèse(1) $V(x, \omega), x \in \mathbb{Z}^d$ sont bornés supérieurement par $M > 0$: $0 \leq V(x, \omega) \leq M$ pour tout $x \in \mathbb{Z}^d$ et pour tout $\omega \in \Omega$. De plus, pour $d = 2$, nous supposons qu'il existe $\epsilon > 0$ tel que $V(x, \omega) \geq \epsilon > 0$.

Théorème 1 *Sous l'hypothèse (1), il existe des constantes positives C_1, C_2, C_3 telles que pour tout $x \in \mathbb{Z}^d, |x| > C_3$, et pour tout $t > 0$:*

$$\mathbb{P}\left(\left|\frac{a(0, x, \omega) - \mathbb{E}a(0, x, \omega)}{\sqrt{|x|}}\right| \geq t\right) \leq C_1 \exp(-C_2 t).$$

Théorème 2 *Sous l'hypothèse (1), pour tous les $n \geq 2$, pour les directions de coordonnées $e_i, 1 \leq i \leq d$ de \mathbb{Z}^d , il existe une constante positive C_4 telle que :*

$$\alpha(ne_i) \leq \mathbb{E}a(0, ne_i, \omega) \leq n\alpha(e_i) + C_4 n^{\frac{1}{2}} \ln |n|.$$

Références

- [1] Sourav Chatterjee. The universal relation between scaling exponents in first-passage percolation. *Ann. of Math. (2)*, 177(2) :663–697, 2013.
- [2] Ivan Corwin. The Kardar–Parisi–Zhang equation and universality class. *Random matrices : Theory and applications*, 1(01) :1130001, 2012.
- [3] Thi Thu Hien Lê. *Exposants de Lyapunov et potentiel aléatoire*. Université de Bretagne Occidentale, 2015. Thesis (Ph.D.).
- [4] Martin P. W. Zerner. Directional decay of the Green's function for a random non-negative potential on \mathbf{Z}^d . *Ann. Appl. Probab.*, 8(1) :246–280, 1998.

ESTIMATION DE LA COURBE D'ACTUALISATION PAR KRIGEAGE SOUS CONTRAINTES

Areski Cousin[†], Hassan Maatouk[‡] et Didier Rullière[†]

([†]) Université de Lyon, Université Lyon 1, ISFA, LSAF, 50 avenue Tony Garnier, 69366 Lyon, France

([‡]) Mines de Saint-Etienne, 158 Cours Fauriel, 42023 Saint-Étienne, France

hassan.maatouk@mines-stetienne.fr

Résumé - *La construction de structures par terme est au coeur de l'évaluation financière et de la gestion du risque voir e.g. [1], [2], [3], [4] et [5]. Une structure par terme est une courbe qui décrit l'évolution d'une grandeur économique ou financière comme une fonction de la maturité ou horizon de temps. Des exemples typiques sont la structure par terme des taux d'intérêt sans risque, la structure par terme d'obligations, la structure par terme de probabilités de défaut et la structure par terme de volatilités implicites de rendements d'actifs financiers. En pratique, les cotations des marchés des produits financiers sous-jacents sont utilisées et fournissent une information partielle sur les structures par terme considérées. De plus, cette information est plus au moins fiable en fonction de la liquidité de la maturité des marchés en question. Le problème est d'obtenir une courbe continue en la maturité à partir de ces informations.*

Mots clés - **facteur d'actualisation, probabilités de défaut, krigeage, contraintes de monotonie**

1 Interpolation sous contraintes linéaires

Dans cet article, on considère le cas où la vraie fonction f vérifie des contraintes linéaires de type égalité de la forme suivante :

$$A \cdot f(\mathbf{X}) = \mathbf{b}, \quad f(\mathbf{X}) = \left(f(x^{(1)}), \dots, f(x^{(n)}) \right)^\top, \quad (1)$$

où A est une matrice de dimension $m \times n$, $m, n \geq 1$ et où $\mathbf{b} \in \mathbb{R}^m$ (A est supposée de rang m). Ce type de contraintes généralise les contraintes d'interpolation étudiées jusqu'ici. Dans le cadre de la Régression par Processus Gaussien, on considère Y un PG de moyenne μ et de fonction de covariance K . La loi conditionnelle de Y sachant (1) est :

$$Y(\mathbf{x}) \mid A \cdot Y(\mathbf{X}) = \mathbf{b} \sim \mathcal{N}(\eta(\mathbf{x}), \tau^2(\mathbf{x})),$$

où

$$\begin{cases} \eta(\mathbf{x}) = \mu(\mathbf{x}) + (A\mathbf{k}(\mathbf{x}))^\top (A\mathbb{K}A^\top)^{-1} (\mathbf{b} - A\boldsymbol{\mu}) \\ \tau^2(\mathbf{x}) = K(\mathbf{x}, \mathbf{x}) - (A\mathbf{k}(\mathbf{x}))^\top (A\mathbb{K}A^\top)^{-1} A\mathbf{k}(\mathbf{x}) \end{cases}$$

areski.cousin@univ-lyon1.fr

hassan.maatouk@mines-stetienne.fr

didier.rulliere@univ-lyon1.fr

avec $\boldsymbol{\mu} = \mu(\mathbf{X}) = (\mu(x^{(1)}), \dots, \mu(x^{(n)}))^\top \in \mathbb{R}^n$, $\mathbf{k}(\mathbf{x})$ le vecteur de covariance entre $Y(\mathbf{x})$ et $Y(\mathbf{X})$ et \mathbb{K} la matrice de covariance de $Y(\mathbf{X})$. Notons que dans le cas où A est une matrice identité, les contraintes linéaires sont des conditions d'interpolation usuelle.

2 Modèle d'interpolation monotone

Dans l'application envisagée dans cet article, la vraie fonction est une courbe d'actualisation ou une probabilité de défaut. Comme fonction de la maturité (T), elle est *a priori* décroissante, de valeur égale à 1 en $T = 0$ et tend vers 0 avec T .

Dans cette section, la variable x est supposée dans $[0, 1]$ et le PG Y centré ($\mu = 0$). On utilise l'approximation suivante (cf. [7]) :

$$Y^N(x) = Y(0) + \sum_{j=0}^N Y'(u_j) \phi_j(x) = \eta + \sum_{j=0}^N \xi_j \phi_j(x), \quad (2)$$

où $\boldsymbol{\xi} = (\eta, \xi_0, \dots, \xi_N)^\top \in \mathbb{R}^{N+2}$ est un vecteur gaussien centré de matrice de covariance Γ^N et où $(\phi_j)_{0 \leq j \leq N}$ sont déterministes :

$$\phi_j(x) = \int_0^x h_j(u) du, \quad x \in [0, 1],$$

où $h_j(x) = h\left(\frac{x-u_j}{\delta}\right) := \left(1 - \frac{|x-u_j|}{\delta}\right) \mathbf{1}_{\frac{|x-u_j|}{\delta} \leq 1}$ est la fonction chapeau centrée sur le $j^{\text{ème}}$ noeud $u_j := j\delta$, $j = 0, \dots, N$, de la subdivision uniforme de pas $\delta = 1/N$. Dans ce cas, la monotonie décroissante du processus gaussien Y^N est *équivalente* à la négativité de tous les coefficients ξ_j , $j = 0, \dots, N$. De plus, la matrice de covariance Γ^N du vecteur gaussien $\boldsymbol{\xi}$ peut s'écrire sous la forme :

$$\Gamma^N = \begin{bmatrix} K(0, 0) & \frac{\partial K}{\partial x'}(0, u_j) \\ \frac{\partial K}{\partial x}(u_i, 0) & \frac{\partial^2 K}{\partial x \partial x'}(u_i, u_j) \end{bmatrix}_{0 \leq i, j \leq N},$$

où $K = \sigma^2 C_\theta$ est la fonction de covariance du processus initial Y , avec σ un paramètre d'échelle, C_θ la fonction de corrélation où θ un paramètre de portée.

2.1 Simulation des trajectoires

En appliquant les conditions linéaires définies dans (1) au modèle proposé $Y^N(x)$ (voir (2)), on obtient :

$$A \cdot Y^N(\mathbf{X}) = A \cdot \left(\eta + \sum_{j=0}^N \xi_j \phi_j(\mathbf{X}) \right) = (A \cdot H) \cdot \boldsymbol{\xi} = \mathbf{b}, \quad (3)$$

où la matrice H de taille $m \times (N + 2)$ est définie par :

$$H_{i,j} = \begin{cases} 1, & \text{pour } i = 1, \dots, n \text{ et } j = 1, \\ \phi_{j-2}(x^{(i)}) & \text{pour } i = 1, \dots, n \text{ et } j = 2, \dots, N + 2. \end{cases}$$

La simulation du processus gaussien conditionnel sous contraintes de monotonie (décroissante) équivaut à simuler un vecteur gaussien $\boldsymbol{\xi}$ tronqué à l'espace correspondant aux conditions linéaires de type égalité et aux contraintes de monotonie suivantes :

$$\begin{cases} (A \cdot H) \cdot \boldsymbol{\xi} = \mathbf{b} & \text{conditions linéaires de type égalité} \\ \boldsymbol{\xi} \in C_\xi & \text{contraintes de monotonie} \end{cases}$$

où $C_\xi = \{\boldsymbol{\xi} \in \mathbb{R}^{N+2} : \xi_j \leq 0, j = 0, \dots, N\}$. Dans ce cas, plusieurs algorithmes peuvent être utilisés voir e.g. [6] et [9].

2.2 Courbe la plus probable

En utilisant le modèle défini dans (2), l'estimateur (MAP) est égal à :

$$M_K(x | A, \mathbf{b}) = \nu + \sum_{j=0}^N \nu_j \phi_j(x), \quad (4)$$

où $\boldsymbol{\nu} = (\nu, \nu_0, \dots, \nu_N)^\top \in \mathbb{R}^{N+2}$ est la solution du problème d'optimisation quadratique suivant :

$$\boldsymbol{\nu} = \arg \min_{c \in I_\xi \cap C_\xi} \left(\frac{1}{2} c^\top (\Gamma^N)^{-1} c \right),$$

avec Γ^N la matrice de covariance du vecteur gaussien $\boldsymbol{\xi}$. Le vecteur $\boldsymbol{\nu}$ est vu comme le mode du vecteur gaussien $\boldsymbol{\xi}$ tronqué à $I_\xi \cap C_\xi$, où I_ξ représente les conditions linéaires de type égalité $(A \cdot H) \cdot \boldsymbol{\xi} = \mathbf{b}$.

L'avantage d'un tel estimateur est de vérifier à la fois les conditions linéaires de type égalité définies dans (3) et la monotonie (décroissante). De plus, il ne dépend pas du paramètre de variance. Pour $N = +\infty$, c'est aussi la fonction spline d'interpolation sous contraintes défini par :

$$h_{opt} = \arg \min_{h \in H \cap C \cap I} \|h\|_H^2,$$

où H est le Reproducing Kernel Hilbert Space (RKHS) associé au noyau de covariance K , C est l'espace des fonctions décroissantes et $I = \{f : A \cdot f(\mathbf{X}) = \mathbf{b}\}$.

3 Étude empirique (Euribor data)

Le modèle proposé a été appliqué à des données réelles.

Les données. La fonction à estimer est définie sur l'intervalle $[0, 40]$ de maturités exprimées en années. Elle démarre de 1 et décroît vers zéro à l'infini. En notant $\mathbf{X} = (1, \dots, 40)^\top$ (de taille 40), les conditions linéaires de type égalité sont définies par une matrice réelle A de taille 14×40 :

$$A \cdot Y^N(\mathbf{X}) = \mathbf{b}, \quad (5)$$

où $\mathbf{b} = (1, \dots, 1)^\top \in \mathbb{R}^{14}$. Dans ce cas, on a $m = 14$ et $n = 40$. En tenant compte du fait que le taux d'actualisation démarre de 1, l'ensemble des contraintes linéaires de type égalité est :

$$\begin{pmatrix} 1 & \phi_0(0) & \dots & \phi_N(0) \\ & A \cdot H & & \end{pmatrix} \boldsymbol{\xi} = \begin{pmatrix} 1 \\ \mathbf{b} \end{pmatrix}. \quad (6)$$

Dans ce cas, la simulation du processus gaussien Y^N conditionnellement à (6) et à la contrainte de monotonie (décroissante) est *équivalente* à simuler le vecteur gaussien $\boldsymbol{\xi}$ tronqué au sous-espace convexe $I_\xi \cap C_\xi$ où I_ξ est donnée par l'équation (6) et $C_\xi := \{\boldsymbol{\xi} \in \mathbb{R}^{N+2} : \xi_j \leq 0, j = 0, \dots, N\}$.

Simulations. On considère les données Euribor au 02/06/2010.

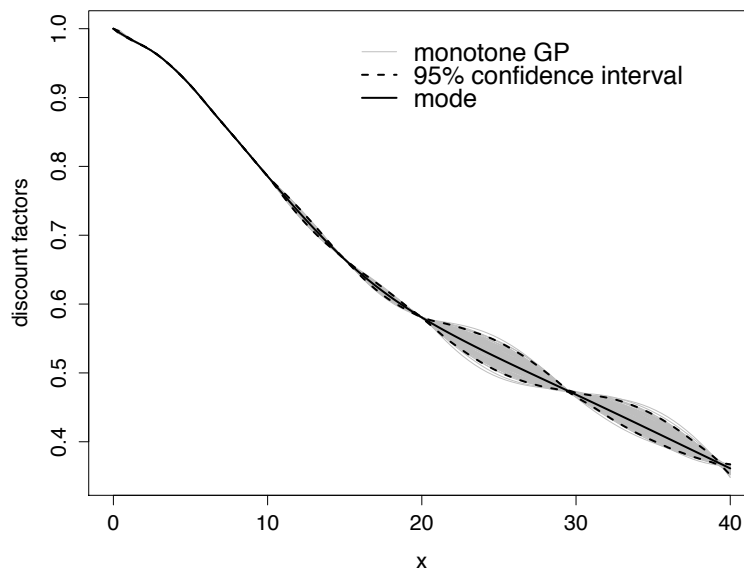


FIGURE 1 – Simulations d'un processus gaussien (trajectoires grises) conditionnellement décroissant en utilisant le noyau Matérn 3/2 avec les paramètres $(\hat{\sigma}^2, \hat{\theta}) = (0.8, 47.9)$ qui sont estimés par un technique de Validation Croisée adaptée aux contraintes inégalité voir [8].

Sur la Figure 1, on choisit $N = 60$ et en simulant le processus gaussien Y^N conditionnellement monotone (décroissante). Remarquons que les trajectoires simulées (courbes en gris) sont décroissantes sur le domaine tout entier. La courbe noire représente le mode du processus gaussien conditionnel qui respecte à la fois la contrainte de monotonie décroissante et les contraintes linéaires de type égalité (3). Les courbes en tirets noirs représentent un intervalle de confiance à 95% calculé par simulation.

4 Conclusions

Dans cet article, une courbe d'actualisation est estimée en utilisant la Régression par Processus Gaussien sous contraintes. Comme nous avons pu le voir, l'adaptation de la méthode au cas de contraintes linéaires de type égalité quelconque ne pose pas de difficulté. Pour

étudier le modèle proposé, un exemple numérique en dimension 1 est inclus. Le résultat numérique montre que les simulations du processus gaussien conditionnel respectent à la fois les contraintes linéaires de type égalité et la monotonie décroissante par rapport à la maturité.

Références

- [1] Areski Cousin, Hassan Maatouk, and Didier Rullière. Kriging of financial term-structures. September 2015.
- [2] C. P. Fries. Curves and term structure models : Definition, calibration and application of rate curves and term structure models. DZ Bank AG ; LMU Munich, Department of Mathematics, 2013.
- [3] Y. Iwashita. Piecewise polynomial interpolations. Open Gamma Technical Report, 2013.
- [4] C. Kenyon and R. Stamm. *Discounting, Libor, CVA and Funding : Interest Rate and Credit Pricing*. Palgrave Macmillan, 2012.
- [5] F. Le Floch. Stable interpolation for the yield curve. Calypso Technology, 2013.
- [6] H. Maatouk and X. Bay. A New Rejection Sampling Method for Truncated Multivariate Gaussian Random Variables Restricted to Convex Sets. <https://hal.archives-ouvertes.fr/hal-01063978>, September 2014.
- [7] H. Maatouk and X. Bay. Gaussian Process Emulators for Computer Experiments with Inequality Constraints. hal-01096751, December 2014.
- [8] H. Maatouk, O. Roustant, and Y. Richet. Cross-Validation Estimations of Hyper-Parameters of Gaussian Processes with Inequality Constraints. *Procedia Environmental Sciences*, 27 :38 – 44, 2015. Spatial Statistics conference 2015.
- [9] C. P. Robert. Simulation of truncated normal variables. *Statistics and Computing*, 5(2), 1995.

ECHANTILLONAGE DANS UNE POPULATION BRANCHANTE STRUCTURÉE

Aline Marguet

Centre de Mathématiques Appliquées, École Polytechnique
Route de Saclay, Palaiseau, France
aline.marguet@cmap.polytechnique.fr

Résumé - *On s'intéresse à l'évolution d'une population structurée par un trait donné, la taille par exemple. La population est décrite par un processus de Markov branchant à valeurs mesures indexé par un arbre. Le long de chaque branche, on suit l'évolution d'une caractéristique propre à chaque individu. Cette caractéristique évolue suivant un processus de Markov. L'histoire de vie d'un individu dépend de l'évolution de son trait. Autrement dit, la durée de vie, le nombre de descendants et le trait à la naissance des descendants sont des variables aléatoires qui dépendent du trait de l'individu. On observe dans cette situation un biais à l'échantillonnage dû au fait que chaque individu n'a pas la même capacité à produire des descendants. En effet, la descendance d'un individu très fécond constitue une grande proportion de la population totale à un temps fixé. Ainsi, si l'on prend un individu uniformément au hasard dans la population au temps t , son trait n'est pas le trait moyen d'un individu après un temps t mais un trait "biaisé". C'est ce biais que l'on cherche à caractériser. La dynamique du trait dans la population peut être résumée par un processus de Markov auxiliaire généralement inhomogène en temps. Ce processus nous permet de décrire le trait d'un individu uniformément échantillonné à un temps donné.*

Mots clés - **Processus de branchement, formule "Many-to-one", reproduction biaisée par la taille, lignée ancestrale**

1 Introduction

Such processes have been first studied in a discrete framework by Lyons and al. [LPP95]. In particular, they introduced the notion of sized-biased tree and used it to prove the well-known Kesten-Stigum theorem. An extension to the multitype case has been done by Kurtz and al. in [KLPP97] and to an arbitrary state space by Athreya in [Ath00].

One important step in understanding the dynamic of the branching process is finding a so-called many-to-one formula. It is a decomposition of the process in two parts : a trait-related part and a size-related part. To this end, Hardy and Harris developed the notion of spine for branching processes and proved a many-to-one formula usually referred as the spine decomposition [HH06, HH09]. But as this decomposition relies on properties of the eigenlements of the mean operator of the process, there are several examples where it can not be used : in Kimmel's infection model [BT10] because the eigenvector goes to zero as the number of parasite goes to zero or in varying environment. Moreover, the spine decomposition is a weighted many-to-one formula : it does not give the exact behaviour of the population

but of a biased population. Bansaye and al. developed an other approach in the neutral case [BDMT11]. They proved a non-weighted many-to-one formula and used it to prove asymptotic results on the population. This many-to-one formula has been extended by Cloez in the case of a varying division rate to a weighted decomposition using eigenelements [Clo11]. In this talk, we will present our many-to-one formula. This formula is an extension to the formula proved in [BDMT11] in the neutral case. We will introduce an auxiliary process as in [Clo11]. This process is a Markov process which does not involve eigenelements but it is an time-inhomogeneous process.

1.1 Description of the process

First, we will give some notations and a brief description of the model as follow :

- At $t = 0$, there is a unique individual in the population.
- His feature $(X_t^\emptyset)_{t \geq 0}$ evolves as a Markov process with generator \mathcal{G} . Let $\Phi(t, s, x)$ be the stochastic flow corresponding to the evolution of the feature.
- Each individual $u \in \mathcal{U}$ dies at rate $B(X_t^u)$, i.e.

$$\mathbb{P} \left(\beta(u) \in dt \mid \beta(u) > t, (X_s^u)_{s \leq t} \right) = B(X_t^u) dt$$

where $\beta(u)$ is the time of death of the individual u ,

- At its death, it is replaced by A_u offspring, where A_u is a random number with distribution $(p_k(X_{\beta(u)}^u))_{k \geq 0}$. We denote by $m(x) = \sum_{k \geq 0} k p_k(x)$ the mean of A_u and $G_{X_{\beta(u)}^u}(k) = \mathbb{P}(A_u \leq k)$ its cumulative distribution function. Let $\alpha(u)$ be the random variable representing the time of birth of u .
- $(F_j^{(k)}(x, \theta), j = 1 \dots k)$ is the random vector with distribution $P^{(k)}(x, dx_1, \dots, dx_k)$ which represents the traits of the k offspring of an individual of trait x . We denote by $P_j^{(k)}(x, dy) = P^{(k)}(x, \mathcal{X}^{j-1} \times dy \times \mathcal{X}^{k-j})$ its j th marginal distribution and by $m(x, dy) = \sum_{k \geq 1} p_k(x) \sum_{j=1}^k P_j^{(k)}(x, dy)$ the mean distribution of the number of children with trait dy of an individual of trait x .
- Conditionally to the feature of their ancestor, the children evolve following the same dynamic independently from each other.

We denote by $\mathcal{M}_P(\mathcal{X})$ the set of punctual measures on \mathcal{X} . Following N. Fournier and S. Méléard [FM04], we will define :

$$Z_t = \sum_{u \in V_t} \delta_{(X_t^u)},$$

where V_t denotes the individual alive at time t . $(Z_t, t \geq 0)$ is a càdlàg measure-valued Markov process in $\mathbb{D}(\mathbb{R}_+, \mathcal{M}_P(\mathcal{X}))$.

Assumptions. For the entire talk, we will assume the following :

1. The function B is locally bounded.
2. There exists $\bar{m} \geq 0$ such that for all $x \in \mathcal{X}$, $m(x) \leq \bar{m}$.

3. For all $x \in \mathcal{X}$ and $T > 0$, there exists a compact set $K_{x,T} \in \mathcal{X}$ such that for all $t < T$:

$$\mathbb{P}(\Phi(t, 0, x) \in K_{x,T}) = 1.$$

After a short explanation about the existence and uniqueness of the process as a solution of a stochastic differential equation, we will give an equation describing the evolution of the process :

Theorem 1 For every function $(t, x) \mapsto f(t, x)$ in $\mathcal{C}_b^{1,0}(\mathbb{R}_+ \times \mathcal{X}, \mathbb{R}) \cap D_{\mathcal{G}}$, where $D_{\mathcal{G}}$ is the domain of \mathcal{G} , we have :

$$\begin{aligned} \langle Z_t, f \rangle = & f(0, x_0) + \int_0^t \int_{\mathcal{X}} \left(\mathcal{G}f(s, x) + \frac{\partial f}{\partial s}(s, x) \right) Z_s(dx) ds + M_t^f + \int_0^t \int_E \mathbf{1}_{u \in V_{s^-}} \mathbf{1}_{\theta \leq B(X_{s^-}^u)} \\ & \mathbf{1}_{\left\{ G_{X_{s^-}^u}^{-1}(t) = k \right\}} \left(\sum_{j=1}^k f\left(s, F_j^{(k)}(X_{s^-}^u, r)\right) - f\left(s, X_{s^-}^u\right) \right) M(ds, du, d\theta, dl, dr), \end{aligned}$$

where

- $E = \mathcal{U} \times \mathbb{R}_+ \times [0, 1] \times [0, 1]$.
- M is a Poisson point measure on $\mathbb{R}^+ \times \mathcal{U} \times \mathbb{R}^+ \times [0, 1] \times [0, 1]$ with intensity $ds \otimes n(du) \otimes d\theta \otimes dl \otimes dr$,
- $(M_t^f)_{t \geq 0}$ is a \mathcal{F}_t -martingale where \mathcal{F}_t is the filtration associated with the process up to time t .

2 Results about sampling at a fixed time

2.1 The many-to-one formula

A first step in finding the distribution of the process of an uniformly chosen individual at time t consists in the characterization of the dynamic of a typical individual via an auxiliary process. This process take into account the bias present at sampling. It depends on the capacity of an individual to have a lot of offspring. Let $m_t(x) = \mathbb{E}_x(N_t)$ denote the average number of individual in the population after a time t starting from an individual of type x . In the talk, we will introduce the equation linking the dynamic of the entire population and a typical individual following the auxiliary process which is called a many-to-one formula.

Theorem 2 For all $t \in \mathbb{R}$ and $s \leq t$:

$$\mathbb{E}_x \left[\sum_{u \in V_t} f(X_s^u) \right] = m_t(x) \mathbb{E}_x \left[f(Y_s^{(t)}) \right], \quad (1)$$

with $(Y_s^{(t)})_{s \leq t}$ a time-inhomogeneous Markov process with infinitesimal generator $\mathcal{A}_s^{(t)}$ given for f in its domain by :

$$\frac{\mathcal{G}(m_{t-s}f)(x) - f(x)\mathcal{G}(m_{t-s})(x)}{m_{t-s}(x)} + B(x) \left(\int_{\mathcal{X}} (f(y) - f(x)) \frac{m_{t-s}(y)}{m_{t-s}(x)} m(x, dy) \right). \quad (2)$$

2.2 Examples

We will end the talk by giving two examples where we have an explicit formula for the generator of the auxiliary process : the Transmission Control Protocol and the binary cell division.

Références

- [Ath00] K. B Athreya. Change of measures for markov chains and the llogl theorem for branching processes. *Bernoulli*, 2000.
- [Ban13] V. Bansaye. Ancestral lineage and limit theorems for branching markov chains. *ArXiv preprint*, 2013.
- [BDMT11] V. Bansaye, J.-F. Delmas, L. Marsalle, and V.C. Tran. Limit theorems for markov processes indexed by continuous time galton–watson trees. *The Annals of Applied Probability*, 2011.
- [BT10] V. Bansaye and V.C. Tran. Branching Feller diffusion for cell division with parasite infection. *ArXiv preprint*, 2010.
- [Clo11] B. Cloez. Limit theorems for some branching measure-valued processes. *ArXiv preprint*, 2011.
- [FM04] N. Fournier and S. Méléard. A microscopic probabilistic description of a locally regulated population and macroscopic approximations. *The Annals of Applied Probability*, 2004.
- [HH06] R. Hardy and S.C. Harris. A new formulation of the spine approach to branching diffusions. 2006.
- [HH09] R. Hardy and S.C. Harris. A spine approach to branching diffusions with applications to l p-convergence of martingales. 2009.
- [KLPP97] T. Kurtz, R. Lyons, R. Pemantle, and Y. Peres. A conceptual proof of the kestenstigm theorem for multi-type branching processes. 1997.
- [LPP95] R. Lyons, R. Pemantle, and Y. Peres. Conceptual proofs of $l \log l$ criteria for mean behavior of branching processes. *The Annals of Probability*, 1995.

Echantillonnage dans une population branchante structurée

MESURE SÉROLOGIQUE DE LA TRANSMISSION DU PALUDISME À DIEUMO DU SÉNÉGAL PAR UN MODÈLE MARKOVIEEN

Niass Oumy^{1,2}, Diongue Abdou Kâ¹, Philippe Saint-Pierre³,
Faye Michel Matar², Toure Aïssatou²

¹LERSTAD : Laboratoire d'Etudes et de Recherches en Statistiques et Développement,
UFR Sciences Appliquées et Technologie, UGB, BP 234, Saint-Louis-Sénégal

²Unité d'Immunologie, Institut Pasteur de Dakar, Dakar-Sénégal

³Laboratoire de Statistique Théorique et Appliquée (LSTA), UPMC, Paris
niass_oumy@yahoo.fr

Résumé

Dans les études médicales les modèles multi-états sont souvent utilisés pour évaluer l'évolution des maladies. Nous utilisons le modèle Markovien pour étudier la capacité des marqueurs sérologiques à évaluer le niveau de la transmission du paludisme. Nous considérons un modèle à deux états définissant l'activité des anticorps. Les résultats montrent que le modèle de Markov homogène est restrictif toutefois du point de vue clinique, les résultats confirment les résultats épidémiologiques précédents. Ceci montre l'applicabilité de ce type de modèle en sérologie.

Keywords : Modèle multi-états, processus de Markov, Paludisme, sérologie, *P. falciparum*.

1 Introduction

En épidémiologie la mesure de l'intensité de transmission est importante pour mesurer l'impact de la maladie en santé publique, en particulier dans le cas du paludisme pour évaluer l'effet des stratégies de lutte contre ce fléau. Les mesures standards utilisées dans ce sens sont l'EIR (taux d'inoculation entomologique : nombre de moustiques infectés par personne-année) et la prévalence parasitaire. Ces mesures sont devenues coûteuses en temps et insensibles du fait de la diminution du nombre de moustique infecté et du contact homme-vecteurs. Les marqueurs sérologiques ont été proposés comme outil alternatif d'évaluation du niveau de transmission en zone de

faible transmission par le biais d'un modèle de Markov homogène en transversal [2,4,5]. L'objectif de cette étude est d'évaluer l'applicabilité de ce type de modèle en sérologie du paludisme en zone holo-endémique aussi étudier l'impact de la maladie et les interventions de lutte sur cette mesure d'autant que la plupart des études faites dans ce sens utilisent des modèles sans covariables. D'abord un modèle sans covariable est ajusté sur cinq transversaux et sur le longitudinal (combinaison des transversaux). Ensuite on compare les résultats obtenus avec la mesure standard (EIR). Et enfin on incorpore les covariables dans le modèle en longitudinale.

2 Modèle de Markov homogène

Considérons un modèle constitué de k états ($E = 1, 2, \dots, k$) sur lequel les sujets se déplacent indépendamment suivant un processus de Markov à temps continu. Le modèle de Markov est entièrement défini par ses probabilités de transition :

$$P[X(t) = j/X(k) = x_k, k \leq s] = P[X(t) = j/X(s) = x_s], \quad t \geq s \quad (1)$$

Il peut aussi être caractérisé par ses intensités de transition :

$$\lambda_{ij}(t) = \lim_{\Delta t \rightarrow 0} \frac{p_{ij}(t, t + \Delta t)}{\Delta t}, \quad j \neq i \quad (2)$$

$$\lambda_{ii}(t) = - \sum_{i \neq j} \lambda_{ij}(t), \quad i = 1, 2, \dots, k \quad (3)$$

Le modèle est dit homogène si l'intensité de transition $\lambda_{ij}(t) = \lambda_{ij}$ est constant. Dans ce cas les probabilités de transition sont stationnaires et s'écrivent comme :

$$P(s, t) = P(0, t) = P(t) = \exp(\Lambda \times t), \quad (4)$$

Dans ce cas P_{ii} , $i \in E$ l'équation 4 donne,

$$\begin{aligned} P_{ii}(t) &= \exp(\lambda_{ii} \times t) \\ &= \exp(- \sum_{i \neq j} \lambda_{ij} \times t) \end{aligned}$$

2.1 Intégration des covariables

Dans plusieurs domaines principalement en biologie, il est particulièrement important d'étudier l'effet de certains facteurs. Dans ce cas ci on incorpore les covariables dans le modèle comme dans le modèle de Cox en supposant la proportionnalité des intensités de transition.

Supposons qu'on dispose un vecteur de covariables Z , les intensités de transitions s'écrivent sous la forme :

$$\lambda_{ij} = \lambda_{ij_0} \exp\{\beta_{ij}^T Z(t)\},$$

avec β_{ij} le vecteur des coefficients de régression associés à la transition de i vers j .

2.2 Estimation et ajustement du modèle

Pour l'estimation des paramètres on utilise la méthode du maximum de vraisemblance. Soient $\{X_t, t \in T\}$ un processus de Markov homogène à espace d'états fini $S = \{1, 2, \dots, k\}$ et un échantillon $\{1, 2, \dots, n\}$ composé de n individus. Supposons que pour chaque individu l ($l = 1, 2, \dots, n$), nous disposons des temps successifs de prélèvement, $T_0^l < T_1^l < T_2^l < \dots < T_{n_l}^l$ et les différents états occupés $x_j^l = X(T_j^l)$.

La contribution de l'individu l à la vraisemblance est donnée par :

$$L_l = P_0(x_0^l) \prod_{i=1}^{n_l} P_{x_{i-1}^l, x_i^l}(T_i^l - T_{i-1}^l)$$

avec

$$P_{x_{i-1}^l, x_i^l}(T_i^l - T_{i-1}^l) = \exp\{\Lambda \times (T_i^l - T_{i-1}^l)\}$$

La fonction de vraisemblance du modèle est donnée par le produit des contributions de n individus. La vraisemblance ne dépend que des intensités de transition. L'algorithme du quasi-Newton est utilisé pour l'optimisation de la vraisemblance et les tests de Wald et du rapport de vraisemblance pour l'ajustement du modèle. Pour vérifier l'hypothèse d'homogénéité des intensités de transition, un modèle homogène par période [1] est ajusté en comparaison avec le modèle homogène.

3 Application à la sérologie du Paludisme

Nous appliquons la méthode précédemment décrite aux données sérologique prélevées chez les habitants d'un village du Sénégal, Dielmo holo-endémique du paludisme par le biais d'un projet multidisciplinaire qui est mené depuis 1993 dans cette localité. Nous utilisons un modèle à deux états (figure 1) correspondant à la réactivité des anticorps contre les antigènes d'un extrait brute de schizontes (*Pfsh07/03*) du *P. falciparum* définis par les biologiste.

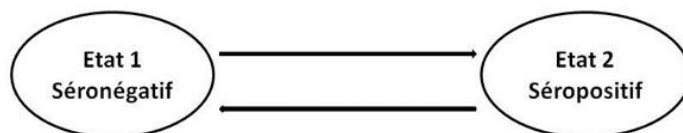


FIG. 1 – Modèle à deux états

Pour chaque transversal, le modèle homogène est appliqué. On s'intéresse particulièrement à l'intensité de transition de l'état séronégatif (état 1) vers

Transition	Modèle deux périodes ($\tau_1 = 15$)		
	Coefficients de régression	Estimation (sd)	p-value
1 -> 2	β_{12}^*	0.452 (0.164)	0.037
2 -> 1	β_{21}^*	-0.789 (0.168)	0.054

TAB. 1 – Estimation des coefficients de régression, l'écart-type (sd) et le maximum entre les p-values obtenues par le test de Wald et le test du rapport de vraisemblance pour le modèle homogène par période

l'état séropositif (état 2) qui représente la force de conversion des anticorps. Pour chaque période cette force est estimée et est corrélée avec la mesure standard de la transmission (EIR). Bien qu'une corrélation de 91% entre l'intensité de transition (λ_{12}) et l'EIR ait été observée, l'hypothèse d'homogénéité des intensités est restrictive. En effet dans la réalité le système immunitaire des enfants et celui des adultes ne réagissent pas de la même façon devant un antigène. Le modèle homogène par période est ajusté afin de vérifier la validité de l'hypothèse d'homogénéité. Un modèle avec deux périodes est utilisé : les intensités de transition sont constantes sur $[0; \tau_1[$ et sur $[\tau_1; \infty[$. Le tableau 1 donne les estimations des coefficients de régression dans un modèle homogène par période (deux périodes, $\tau_1 = 15ans$).

Covariates	Modèle multivarié		
	Coefficients	Estimation (sd)	p-value
Accès clinique	$\hat{\beta}_{12}$	0.12(1.2 - 10 ⁻¹)	< 0.001
	$\hat{\beta}_{21}$	0.94(6.5 - 10 ⁻²)	0.91
Intervention	$\hat{\beta}_{12}$	-0.91(4.2 - 10 ⁻¹)	0.02
	$\hat{\beta}_{21}$	4.41(1.17)	0.96

TAB. 2 – Estimation des coefficients de régression, l'écart-type (sd) et le maximum entre les p-values obtenues par le test de Wald et le test du rapport de vraisemblance pour le modèle de Markov homogène

Tous les coefficients de régression sont statistiquement significatifs. Le coefficient β_{12}^* est positif montrant que le risque est accéléré pour la transition 1 -> 2 dans l'intervalle de temps $[15, \infty[$. Par contre le risque diminue pour la transition 2 -> 1 dans cet même intervalle. Le test du rapport de vraisemblance est utilisé pour la comparaison de modèle et au vue des résultats, il semble que le modèle de Markov homogène et restrictif dans le cas de la sérologie du paludisme. Toutefois qu'il a montré des résultats intéressants : la maladie augmente le risque de conversion des anticorps (le

risque est multiplié par $\exp(0.12)$) alors que les stratégies de lutte le diminuent (tab 2). Ce résultat va à l'encontre des résultats séro-épidémiologiques du paludisme [3].

En supposant que les anticorps restent un certain nombre de temps avant de passer d'un état à un autre et que les durées de séjours sont connues, on peut tester l'hypothèse de Markov en considérant le temps de séjour dans l'état avant de transiter. Ainsi on utilise un modèle semi-Markovien qui intègre la modélisation des temps de séjour. Si les paramètres du temps de séjour sont significatifs alors le modèle homogène est restrictif.

Références

- [1] Alioum A. et Commenges D. Mkvpci : a computer program for markov models with piecewise constant intensities and covariates. *Comput Methods Programs Biomed*, 64(2) :109–119, 2001.
- [2] Corran P, Coleman P, Riley E, Drakeley C. Serology : a robust indicator of malaria transmission intensity? *Trends Parasitol*, 23 :575–582, 2007.
- [3] Diop, F., Richard, V., Diouf, B., Sokhna, C., Diagne, N., Trape, J. F., Faye, M. M., Tall, A., Diop, G., Balde, A. T. Dramatic declines in seropositivity as determined with crude extracts of plasmodium falciparum schizonts between 2000 and 2010 in dielmo and ndiop, senegal. *Malaria Journal*, 13 :83, 2014.
- [4] Drakeley CJ, et al. Estimating medium- and long-term trends in malaria transmission by using serological markers of malaria exposure. *Proc Natl Acad Sci USA*, 102 :5108–5113, 2005.
- [5] Rosas-Aguirre, A. Llanos-Cuentas, A. Speybroeck, N. Cook, J. Contreras-Mancilla, J. Soto, V. Gamboa, D. Pozo, E. Ponce, O. J. Pereira, M. O. Soares, I. S. Theisen, M. D'Alessandro, U. Erhart, A. Assessing malaria transmission in a low endemicity area of north-western peru. *Malaria Journal*, 12 :339, 2013.

TEST DE LA VRAISEMBLANCE EMPIRIQUE POUR UN MODÈLE NON-LINÉAIRE AVEC CHANGE-POINTS.

Zahraa SALLOUM

Université Claude Bernard- Lyon 1

43, boulevard du 11 novembre 1918, Bâtiment Braconnier, Bureau 243.

salloum@math.univ-lyon1.fr

Résumé - On propose une méthode basée sur la vraisemblance empirique pour tester la présence de changement dans les paramètres d'un modèle de régression non-linéaire. Tout d'abord, on teste l'hypothèse nulle de non-changement dans les paramètres de régression, contre l'hypothèse alternative d'un seul changement dans ces paramètres. Sous l'hypothèse nulle, on prouve la consistance et la vitesse de convergence des estimateurs des paramètres de régression. La loi asymptotique de la statistique de test sous l'hypothèse nulle obtenue nous permet de trouver la valeur critique asymptotique. D'autre part, on prouve que la puissance asymptotique de la statistique de test proposée est égale à 1.

Mots clés - Change-point, Modèle paramétrique non-linéaire, Test de la vraisemblance empirique, Comportement asymptotique.

1 Introduction

Un modèle de régression paramétrique "classique" a la forme :

$$Y_i = f(\mathbf{X}_i; \beta) + \varepsilon_i, \quad i = 1, \dots, n. \quad (1)$$

Pour chaque observation i , Y_i désigne la variable réponse. On considère par $\mathbf{X}_i \in \mathbb{R}^p$ des vecteurs aléatoires i.i.d, de fonction de répartition $H(\mathbf{x})$, avec $\mathbf{x} \in \Upsilon$, $\Upsilon \subseteq \mathbb{R}^p$. L'erreur est désignée par ε_i .

La suite de vecteurs aléatoires $(\mathbf{X}_i, \varepsilon_i)_{1 \leq i \leq n}$ est i.i.d, avec la même distribution de $(\mathbf{X}, \varepsilon)$. Pour tout i , ε_i est indépendante de \mathbf{X}_i . La fonction de régression $f : \Upsilon \times \Gamma \rightarrow \mathbb{R}$, avec $\Upsilon \subseteq \mathbb{R}^p$ et $\Gamma \subseteq \mathbb{R}^d$, est connue sauf le paramètre $\beta = (\beta_1, \dots, \beta_d)$. L'ensemble de paramètres Γ est supposé compact.

Dans la suite, pour $\mathbf{x} \in \Upsilon$ et $\beta \in \Gamma$, on utilise la notation $\dot{\mathbf{f}}(\mathbf{x}, \beta) \equiv \partial f(\mathbf{x}, \beta) / \partial \beta$ et $\ddot{\mathbf{f}}(\mathbf{x}, \beta) \equiv \partial^2 f(\mathbf{x}, \beta) / \partial \beta^2$. On considère le problème du change-point dans un modèle non-linéaire général, par la méthode de vraisemblance empirique. On généralise le travail de [3] où le modèle linéaire était considéré. Une des difficultés qui présente dans l'étude du modèle non-linéaire par rapport au modèle linéaire est lié au fait que les fonctions scores correspondantes pour trouver la statistique de test dépendent des paramètres de régression, et surtout, la forme analytique de ces dérivées est inconnue. D'autre part, pour les modèles linéaires, des nombreuses preuves sont basées sur la convexité de la fonction de régression par rapport au paramètre de régression. Alors, la valeur extrême de la fonction convexe est atteint sur les bornes de l'intervalle considéré. Ces deux facteurs produisent des difficultés dans l'étude théorique de la statistique de test associée dans l'étude des modèles non-linéaire.

2 Suppositions

En ce qui concerne la variable aléatoire ε , on fait la supposition suivante :

(A1) $\mathbb{E}[\varepsilon_i] = 0$ et $\mathbb{E}[\varepsilon_i^2] < \infty$, pour tout $i = 1, \dots, n$.

Soit $\eta > 0$ quelconque et le voisinage de β^0 : $\mathcal{V}_\eta(\beta^0) = \{\beta \in \Gamma; \|\beta - \beta^0\| \leq \eta\}$.

La fonction de régression $f : \Upsilon \times \Gamma \rightarrow \mathbb{R}$ et le vecteur aléatoire \mathbf{X} satisfont les conditions suivantes :

(A2) $\forall \mathbf{x} \in \Upsilon$ et pour $\beta \in \Gamma$, la fonction $f(\mathbf{x}, \beta)$ est trois fois dérivable en β et continue en Υ .

(A3) $(|\frac{\partial^2 f(\mathbf{x}, \beta)}{\partial \beta_j \partial \beta_k}|)_{1 \leq j, k \leq d}$ et $(|\frac{\partial^3 f(\mathbf{x}, \beta)}{\partial \beta_j \partial \beta_k \partial \beta_l}|)_{1 \leq j, k, l \leq d}$ sont bornées, $\forall \mathbf{x} \in \Upsilon$ et $\beta \in \mathcal{V}_\eta(\beta^0)$ et $\eta \rightarrow 0$.

(A4) $\mathbb{E}[\|\dot{\mathbf{f}}(\mathbf{X}, \beta)\|_1] < \infty$, $\mathbb{E}[\|\dot{\mathbf{f}}(\mathbf{X}, \beta)\dot{\mathbf{f}}^t(\mathbf{X}, \beta)\|_1] < \infty$ et $\mathbb{E}[\|\frac{\partial^2 f(\mathbf{X}, \beta)}{\partial \beta_j \partial \beta_k}\|] < \infty$, $\forall 1 \leq j, k \leq d$ et $\beta \in \mathcal{V}_\eta(\beta^0)$ et $\eta \rightarrow 0$.

Les suppositions (A3) et (A4) sont des conditions standard pour des modèles non-linéaires, voir par exemple le livre de [4]. On remarque que la supposition (A4) est plus faible que celle employée dans le papier de [1], dont la méthode des moindres carrés a été utilisé pour tester l'hypothèse de non-changement.

Notons que, les suppositions du papier de [1] sont : $\mathbb{E}[\varepsilon_i^{2s}] < \infty$, $\sup_\beta \mathbb{E}[f(\mathbf{X}, \beta)]^{2s} < \infty$, $\sup_\beta \mathbb{E}[\|\dot{\mathbf{f}}(\mathbf{X}, \beta)\|]^{2s} < \infty$ et pour tout $1 \leq j, k \leq d$, $\sup_\beta \mathbb{E}[\|\frac{\partial^2 f(\mathbf{X}, \beta)}{\partial \beta_j \partial \beta_k}\|^s] < \infty$, pour $s > 2$.

3 Test d'un changement

Dans cette section, on teste l'hypothèse de non-changement dans les paramètres de régression du modèle non-linéaire donné par (1) contre l'hypothèse qu'il y a un seul changement à l'observation k inconnue, dans le paramètre, ce qui implique le modèle suivant

$$Y_i = \begin{cases} f(\mathbf{X}_i; \beta_1) + \varepsilon_i & i = 1, \dots, k \\ f(\mathbf{X}_i; \beta_2) + \varepsilon_i & i = k + 1, \dots, n. \end{cases} \quad (2)$$

3.1 Hypothèses

Sous l'hypothèse H_1 , on suppose que le modèle change à l'observation k , donc le modèle a la forme (2). Alors, l'hypothèse nulle en fonction des paramètres du modèle (2) peut s'écrire

$$H_0 : \beta_1 = \beta_2 = \beta. \quad (3)$$

L'hypothèse alternative suppose la présence d'un seul changement dans les paramètres de régression,

$$H_1 : \beta_1 \neq \beta_2. \quad (4)$$

Sous l'hypothèse nulle H_0 , on désigne par β^0 la vraie valeur (supposé inconnue) du paramètre β . Sous l'hypothèse H_1 , β_1^0 et β_2^0 (aussi inconnus) sont les vrais paramètres de β_1, β_2 . Considérons des notations supplémentaires. Tout d'abord on donne des notations qui seront utilisées sous l'hypothèse nulle H_0 . Soit le vecteur aléatoire, de dimension d , suivant $\mathbf{g}(\mathbf{X}_i, \beta) \equiv \mathbf{g}_i(\beta) \equiv \dot{\mathbf{f}}(\mathbf{X}_i, \beta)[Y_i - f(\mathbf{X}_i, \beta)]$. On remarque que, sous l'hypothèse H_0 , pour tout $i = 1, \dots, n$, on a $\mathbf{g}_i(\beta^0) = \dot{\mathbf{f}}(\mathbf{X}_i, \beta^0)\varepsilon_i$ et $\mathbb{E}[\mathbf{g}_i(\beta^0)] = 0$. On considère également la matrice suivante de dimension $d \times d$: $\mathbf{V} \equiv \mathbb{E}[\dot{\mathbf{f}}(\mathbf{X}_i, \beta^0)\dot{\mathbf{f}}^t(\mathbf{X}_i, \beta^0)]$. Alors, $\text{Var}[\varepsilon_i \dot{\mathbf{f}}(\mathbf{X}_i, \beta^0)] = \sigma^2 \mathbf{V}$. On considère maintenant les deux ensembles d'entiers naturels $I \equiv \{1, \dots, k\}$ et $J \equiv \{k+1, \dots, n\}$. Ces deux ensembles contiennent les indices d'observations des deux segments du modèle (2). Correspondant à ces ensembles, soient les vecteurs de probabilité : (p_1, \dots, p_k) et (q_{k+1}, \dots, q_n) . Ces vecteurs contiennent les probabilités d'observer la variable dépendante y_i (respectivement y_j) pour la variable dépendante Y_i (respectivement Y_j) : pour $i = 1, \dots, k$, $p_i \equiv \mathbb{P}[Y_i = y_i]$ et pour $j = k+1, \dots, n$, $q_j \equiv \mathbb{P}[Y_j = y_j]$. Evidemment, on a que $\sum_{i \in I} p_i = 1$ et $\sum_{j \in J} q_j = 1$.

Test de la vraisemblance empirique pour un modèle non-linéaire avec change-points.

3.2 Statistique de test

Sous l'hypothèse H_0 le profil du rapport de la vraisemblance empirique pour β a la forme

$$\mathcal{R}'_{0,nk}(\beta) = \sup_{(p_1, \dots, p_k)} \sup_{(q_{k+1}, \dots, q_n)} \left\{ \prod_{i \in I} k p_i \prod_{j \in J} (n-k) q_j; \sum_{i \in I} p_i = 1, \sum_{j \in J} q_j = 1, \right. \\ \left. \sum_{i \in I} p_i \mathbf{g}_i(\beta) = \sum_{j \in J} q_j \mathbf{g}_j(\beta) = \mathbf{0}_d \right\}. \quad (5)$$

De même, sous l'hypothèse H_1 , le profil du rapport de vraisemblance empirique pour β_1, β_2 a la forme

$$\mathcal{R}'_{1,nk}(\beta_1, \beta_2) = \sup_{(p_1, \dots, p_k)} \sup_{(q_{k+1}, \dots, q_n)} \left\{ \prod_{i \in I} k p_i \prod_{j \in J} (n-k) q_j, \sum_{i \in I} p_i = 1, \sum_{j \in J} q_j = 1, \right. \\ \left. \sum_{i \in I} p_i \mathbf{g}_i(\beta_1) = \mathbf{0}_d, \sum_{j \in J} q_j \mathbf{g}_j(\beta_2) = \mathbf{0}_d \right\}.$$

Pour tester l'hypothèse H_0 contre H_1 , on utilise une idée similaire au test du maximum de la vraisemblance. On considère alors, le profil du rapport de la vraisemblance empirique suivant $\frac{\mathcal{R}'_{0,nk}(\beta)}{\mathcal{R}'_{1,nk}(\beta_1, \beta_2)}$. Le résultat suivant désigne une généralisation de la version non paramétrique du théorème de Wilks. Plus spécifiquement, sous l'hypothèse H_1 , puisque les observations sont indépendantes, on peut appliquer le théorème de Wilks sur chacun de deux segments. Alors, on prouve que sous l'hypothèse H_1 , le profil du rapport de vraisemblance empirique pour β_1, β_2 a une distribution asymptotique χ^2 . Notons que les preuves de tous les théorèmes énoncés ainsi des lemmes nécessaires pour ces démonstrations sont donnés dans le papier [2].

Théorème 3.1 Sous les suppositions (A1)-(A3), si l'hypothèse H_1 est vraie, alors

$$-2 \log \mathcal{R}'_{1,nk}(\beta_1, \beta_2) \xrightarrow[n \rightarrow \infty]{\mathcal{L}} \chi^2(2d).$$

■

Conséquence de ce théorème, on peut considérer comme statistique de test le terme $-2 \log \mathcal{R}'_{0,nk}(\beta)$. En utilisant la méthode des multiplicateurs de Lagrange, on obtient que

$$p_i = \frac{1}{k + n \boldsymbol{\lambda}_1^t \mathbf{g}_i(\beta)}, \quad q_j = \frac{1}{n - k - n \boldsymbol{\lambda}_2^t \mathbf{g}_j(\beta)}. \quad (6)$$

Alors, la statistique $-2 \log \mathcal{R}'_{nk,0}(\beta)$ devient

$$2 \sum_{i \in I} \log \left[1 + \frac{n}{k} \boldsymbol{\lambda}_1^t \mathbf{g}_i(\beta) \right] + 2 \sum_{j \in J} \log \left[1 - \frac{n}{n-k} \boldsymbol{\lambda}_2^t \mathbf{g}_j(\beta) \right]. \quad (7)$$

Pour avoir un seul paramètre $\boldsymbol{\lambda}$, on se limitera à un cas particulier, quand $\boldsymbol{\lambda}_1$ et $\boldsymbol{\lambda}_2$ satisfont la contrainte : $\mathbf{V}_{1n}(\beta) \boldsymbol{\lambda}_1 = \mathbf{V}_{2n}(\beta) \boldsymbol{\lambda}_2$, avec $\mathbf{V}_{1n}(\beta) \equiv \frac{1}{k} \sum_{i \in I} \dot{\mathbf{g}}_i(\beta)$ et $\mathbf{V}_{2n}(\beta) \equiv \frac{1}{n-k} \sum_{j \in J} \dot{\mathbf{g}}_j(\beta)$, où $\dot{\mathbf{g}}_i(\beta)$ est une matrices de dimension $d \times d$, qui contient les dérivées du vecteur $\mathbf{g}_i(\beta)$ par rapport à β , pour $i = 1, \dots, k$. Pour le vrai paramètre β^0 , ces deux dernières matrices deviennent, respectivement $\mathbf{V}_{1n}^0 \equiv \mathbf{V}_{1n}(\beta^0)$, et $\mathbf{V}_{2n}^0 \equiv \mathbf{V}_{2n}(\beta^0)$. Tenant compte de cette contrainte, et de la relation (6), la statistique de test de la relation (7) devient

$$2 \sum_{i \in I} \log \left[1 + \frac{n}{k} \boldsymbol{\lambda}^t \mathbf{g}_i(\beta) \right] + 2 \sum_{j \in J} \log \left[1 - \frac{n}{n-k} \boldsymbol{\lambda}^t \mathbf{V}_{1n}(\beta) \mathbf{V}_{2n}^{-1}(\beta) \mathbf{g}_j(\beta) \right]. \quad (8)$$

D'autre part, afin que les paramètres appartiennent à un ensemble borné, à la place du paramètre k , on considère $\theta_{nk} \equiv k/n$ et on note la statistique donnée par la relation (8) par $Z_{nk}(\theta_{nk}, \boldsymbol{\lambda}, \boldsymbol{\beta})$.

En résolvant le système suivant

$$\begin{cases} \phi_{1n}(\theta_{nk}, \boldsymbol{\lambda}, \boldsymbol{\beta}) = \frac{\partial Z_{nk}(\theta_{nk}, \boldsymbol{\lambda}, \boldsymbol{\beta})}{\partial \boldsymbol{\lambda}} = \mathbf{0}_d \\ \phi_{2n}(\theta_{nk}, \boldsymbol{\lambda}, \boldsymbol{\beta}) = \frac{\partial Z_{nk}(\theta_{nk}, \boldsymbol{\lambda}, \boldsymbol{\beta})}{\partial \boldsymbol{\beta}} = \mathbf{0}_d, \end{cases} \quad (9)$$

on obtient le couple des solutions $(\widehat{\boldsymbol{\lambda}}(\theta_{nk}), \widehat{\boldsymbol{\beta}}(\theta_{nk}))$ qui maximisent la statistique de test donnée par (8). Remarquons que ces solutions dépendent du change-point k .

3.3 Comportement asymptotique de la statistique de test

Dans cette sous-section, on étudie le comportement asymptotique de la statistique de test obtenue dans la relation (8), d'abord sous l'hypothèse nulle H_0 et ensuite sous l'alternative H_1 .

3.3.1 Comportement asymptotique sous H_0

Afin d'étudier le comportement asymptotique de $\widehat{\boldsymbol{\lambda}}(\theta_{nk})$ et $\widehat{\boldsymbol{\beta}}(\theta_{nk})$, on commence par étudier kp_i , pour $i \in I$, et $(n-k)q_j$, pour $j \in J$, avec les probabilités p_i, q_j données par la relation (6). Plus précisément, on montre que si $\boldsymbol{\beta}$ est dans le voisinage de $\boldsymbol{\beta}^0$, kp_i et $(n-k)q_j$ peuvent encadrer par deux constantes strictement bornées et positives, avec une probabilité proche de 1.

Proposition 3.1 Soit le η -voisinage de $\boldsymbol{\beta}^0$, $\mathcal{V}_\eta(\boldsymbol{\beta}^0) = \{\boldsymbol{\beta} \in \Gamma; \|\boldsymbol{\beta} - \boldsymbol{\beta}^0\| \leq \eta\}$, avec $\eta \rightarrow 0$. Sous l'hypothèse H_0 , considérons que les suppositions (A1)-(A4) sont vérifiées. Alors

(i) Pour $i \in I$, pour tout $\epsilon > 0$, il existe deux constantes positives M_1 et M_2 , tel que, pour tout $\boldsymbol{\beta} \in \mathcal{V}_\eta(\boldsymbol{\beta}^0)$,

$$\mathbb{P}\left[\frac{1}{M_2} \leq \frac{1}{1 + \frac{\boldsymbol{\lambda}^t}{\theta_{nk}} \mathbf{g}_i(\boldsymbol{\beta})} \leq \frac{1}{M_1}\right] \geq 1 - \epsilon. \quad (10)$$

(ii) Pour $j \in J$, pour tout $\epsilon > 0$, il existe deux constantes positives M_3 et $M_4 > 0$, tel que, pour tout $\boldsymbol{\beta} \in \mathcal{V}_\eta(\boldsymbol{\beta}^0)$,

$$\mathbb{P}\left[\frac{1}{M_4} \leq \frac{1}{1 - \frac{\boldsymbol{\lambda}^t}{1 - \theta_{nk}} \mathbf{V}_{1n}(\boldsymbol{\beta})(\mathbf{V}_{2n}(\boldsymbol{\beta}))^{-1} \mathbf{g}_j(\boldsymbol{\beta})} \leq \frac{1}{M_3}\right] \geq 1 - \epsilon. \quad (11)$$

■

Par le résultat suivant, on montre que $\widehat{\boldsymbol{\lambda}}(\theta_{nk})$ et $\widehat{\boldsymbol{\beta}}(\theta_{nk})$, les solutions du système (9), sont des bons estimateurs. Plus précisément, on montre que sous l'hypothèse H_0 , $\|\widehat{\boldsymbol{\lambda}}(\theta_{nk})\|$ converge en probabilité vers 0 quand $n \rightarrow \infty$ et que $\widehat{\boldsymbol{\beta}}(\theta_{nk})$ est un estimateur consistant de $\boldsymbol{\beta}^0$. On obtient également la vitesse de convergence de ces deux estimateurs, ce qui nous permet de proposer une statistique de test simple à la place de $Z_{nk}(\theta_{nk}, \boldsymbol{\lambda}, \boldsymbol{\beta})$.

Théorème 3.2 Sous les suppositions (A1)-(A4), si l'hypothèse H_0 est vraie, alors on a $\widehat{\boldsymbol{\lambda}}(\theta_{nk}) = \min\{\theta_{nk}, 1 - \theta_{nk}\} O_{\mathbb{P}}((n \min\{\theta_{nk}, 1 - \theta_{nk}\})^{-1/2})$ et $\widehat{\boldsymbol{\beta}}(\theta_{nk}) - \boldsymbol{\beta}^0 = O_{\mathbb{P}}((n \min\{\theta_{nk}, 1 - \theta_{nk}\})^{-1/2})$. ■

Remarque 3.1 Tenant compte du Théorème 3.2, sous l'hypothèse H_0 , on peut considérer à la place de la statistique $Z_{nk}(\theta_{nk}, \boldsymbol{\lambda}, \boldsymbol{\beta})$, donnée par la relation (8), la statistique suivante

$$T_{nk}(\theta_{nk}, \boldsymbol{\lambda}, \boldsymbol{\beta}) = 2 \sum_{i \in I} \log\left(1 + \frac{1}{\theta_{nk}} \boldsymbol{\lambda}^t \mathbf{g}_i(\boldsymbol{\beta})\right) + 2 \sum_{j \in J} \log\left(1 - \frac{1}{1 - \theta_{nk}} \boldsymbol{\lambda}^t \mathbf{g}_j(\boldsymbol{\beta})\right). \quad (12)$$

Test de la vraisemblance empirique pour un modèle non-linéaire avec change-points.

Puisque la fonction de régression est non-linéaire, afin que le maximum de la vraisemblance empirique existe toujours, on considère le paramètre $\theta_{nk} \in [\Theta_{1n}, \Theta_{2n}] \subset (0, 1)$, tel que $n\Theta_{1n} \rightarrow \infty$, $n(1-\Theta_{2n}) \rightarrow \infty$, quand $n \rightarrow \infty$. Finalement, la statistique de test pour tester l'hypothèse H_0 contre son alternative H_1 est

$$\tilde{T}_n \equiv \max_{\theta_{nk} \in [\Theta_{1n}, \Theta_{2n}]} T_{nk}(\theta_{nk}, \hat{\lambda}(\theta_{nk}), \hat{\beta}(\theta_{nk})). \quad (13)$$

Alors, on peut considérer comme un estimateur du change-point k^0 , l'estimateur du maximum de la vraisemblance empirique $\tilde{k}_n \equiv n\tilde{\theta}_n \equiv n \min\{\tilde{\theta}_{nk}; \tilde{\theta}_{nk} = \arg \max_{\theta_{nk} \in [\Theta_{1n}, \Theta_{2n}]} T_{nk}(\theta_{nk}, \hat{\lambda}(\theta_{nk}), \hat{\beta}(\theta_{nk}))\}$. On rappelle que $\hat{\lambda}(\theta_{nk})$ et $\hat{\beta}(\theta_{nk})$ sont les solutions du système (9).

Le résultat suivant donne la distribution asymptotique de la statistique de test \tilde{T}_n donnée par la relation (13), sous l'hypothèse nulle de non-changement H_0 . Pour ceci, on considère les fonctions suivantes : $A(x) \equiv (2 \log x)^{1/2}$, $D(x) = 2 \log x + \log \log x$ et $u(n) = \frac{1-\Theta_{1n}\Theta_{2n}}{\Theta_{1n}(1-\Theta_{2n})} \rightarrow \infty$, quand $n \rightarrow \infty$.

Théorème 3.3 Sous les suppositions (A1)-(A4), si l'hypothèse H_0 est vraie, alors pour tout $t \in \mathbb{R}$, on a

$$\lim_{n \rightarrow \infty} \mathbb{P}\{A(\log u(n))(\tilde{T}_n)^{1/2} \leq t + D(\log u(n))\} = \exp(-e^{-t}). \quad (14)$$

■

3.3.2 Comportement asymptotique de T_{nk} et \tilde{T}_{nk} sous H_1

Considérons maintenant que l'hypothèse H_1 est vraie. Tout d'abord, on prouve que le maximum de la statistique T_{nk} converge vers le maximum de sa distribution limite. Alors, on montre que la statistique de test \tilde{T}_n est consistante (à la puissance asymptotique égale à 1). Si k^0 est le vrai temps du changement, on note par $\theta_{n0} = k^0/n$ et on suppose que $\theta_0 \equiv \lim_{n \rightarrow \infty} \theta_{n0}$, où $\theta_0 \in (0, 1)$. Pour $\mathbf{x} \in \Upsilon$ et $e \in \mathcal{R}$, soient $F_{\mathbf{x}}(e)$ la distribution conditionnelle de $\mathbf{g}(\mathbf{X}_i, \beta)$ quand $\mathbf{X}_i = \mathbf{x}$, pour les observations $i \in I$ et $G_{\mathbf{x}}(e)$ la distribution conditionnelle de $\mathbf{g}(\mathbf{X}_j, \beta)$ quand $\mathbf{X}_j = \mathbf{x}$, pour les observations $j \in J$.

Nous rappelons que la fonction de répartition de \mathbf{X} est $H(\mathbf{x})$. Pour $\mathbf{x} \in \Upsilon$ et une valeur fixée de θ , on définit

$$\begin{aligned} dP_{\mathbf{x}}(e) &\equiv (\theta \mathbb{1}_{\{\theta \leq \theta_0\}} + \theta_0 \mathbb{1}_{\{\theta > \theta_0\}}) dF_{\mathbf{x}}(e) + (\theta - \theta_0) \mathbb{1}_{\{\theta > \theta_0\}} dG_{\mathbf{x}}(e), \\ dQ_{\mathbf{x}}(e) &\equiv ((1 - \theta) \mathbb{1}_{\{\theta \geq \theta_0\}} + (1 - \theta_0) \mathbb{1}_{\{\theta < \theta_0\}}) dG_{\mathbf{x}}(e) + (\theta_0 - \theta) \mathbb{1}_{\{\theta < \theta_0\}} dF_{\mathbf{x}}(e), \\ dR_{\mathbf{x}}(e) &\equiv \mathbb{1}_{\{\theta < \theta_0\}} dF_{\mathbf{x}}(e) + \mathbb{1}_{\{\theta > \theta_0\}} dG_{\mathbf{x}}(e). \end{aligned}$$

Comme on a prouvé que sous l'hypothèse H_0 on peut considérer à la place de la statistique de vraisemblance empirique (8) la statistique (12), alors, on définit la statistique suivante

$$\Lambda_{nk}(\theta_{nk}) = \frac{T_{nk}(\theta_{nk}, \tilde{\lambda}(\theta_{nk}), \tilde{\beta}(\theta_{nk}))}{2n}, \Lambda_n(0) = \Lambda_n(1) = 0, \quad (15)$$

avec T_{nk} donné par la relation (12), et $\tilde{\lambda}(\theta_{nk}), \tilde{\beta}(\theta_{nk})$ les solutions du système

$$\left\{ \begin{aligned} \frac{\partial T_{nk}(\theta_{nk}, \lambda, \beta)}{\partial \lambda} &= \sum_{i \in I} \frac{\mathbf{g}_i(\beta)}{\theta_{nk} + \lambda^t \mathbf{g}_i(\beta)} - \sum_{j \in J} \frac{\mathbf{g}_j(\beta)}{1 - \theta_{nk} - \lambda^t \mathbf{g}_j(\beta)} = \mathbf{0}_d, \\ \frac{\partial T_{nk}(\theta_{nk}, \lambda, \beta)}{\partial \beta} &= \sum_{i \in I} \frac{\dot{\mathbf{g}}_i(\beta) \lambda}{\theta_{nk} + \lambda^t \mathbf{g}_i(\beta)} - \sum_{j \in J} \frac{\dot{\mathbf{g}}_j(\beta) \lambda}{1 - \theta_{nk} - \lambda^t \mathbf{g}_j(\beta)} = \mathbf{0}_d. \end{aligned} \right. \quad (16)$$

Par une preuve similaire à celle du théorème 3.2, on a que

$$\begin{aligned} \tilde{\lambda}(\theta_{nk}) &= \min\{\theta_{nk}, 1 - \theta_{nk}\} O_{\mathbb{P}}((n \min\{\theta_{nk}, 1 - \theta_{nk}\})^{-1/2}), \\ \tilde{\beta}(\theta_{nk}) - \beta^0 &= O_{\mathbb{P}}((n \min\{\theta_{nk}, 1 - \theta_{nk}\})^{-1/2}). \end{aligned} \quad (17)$$

Pour un λ et β quelconque, soit la fonction $K : \Upsilon \times \mathbb{R} \times (0, 1)$ définie par $K(\mathbf{x}, e, \theta) = \theta + \lambda \dot{\mathbf{f}}(\mathbf{x}, \beta)[e - f(\mathbf{x}, \beta) + f(\mathbf{x}, \beta^0)]$. Soit également

$$\begin{aligned} \psi(\theta, \lambda, \beta) &= \int_{\Upsilon} \left(\int_{\mathbb{R}} \log K(\mathbf{x}, e, \theta) dP_{\mathbf{x}}(e) + \int_{\mathbb{R}} \log(1 - K(\mathbf{x}, e, \theta)) dQ_{\mathbf{x}}(e) \right) dH(\mathbf{x}) \\ &- \theta \log \theta - (1 - \theta) \log(1 - \theta). \end{aligned} \quad (18)$$

Par le théorème 3.4, on prouve que sous l'hypothèse H_1 , ψ est le processus limite de Λ_{nk} . Alors, pour une valeur fixée de θ , telle que $\theta \in (0, 1)$, on considère par $\tilde{\lambda}(\theta), \tilde{\beta}(\theta)$ la solution du système suivant

$$\begin{cases} \mathbf{z}_1(\theta, \lambda, \beta) = \int_{\Upsilon} \left(\int_{\mathbb{R}} \frac{\mathbf{g}(\mathbf{x}, \beta)}{K(\mathbf{x}, e, \theta)} dP_{\mathbf{x}}(e) - \int_{\mathbb{R}} \frac{\mathbf{g}(\mathbf{x}, \beta)}{1 - K(\mathbf{x}, e, \theta)} dQ_{\mathbf{x}}(e) \right) dH(\mathbf{x}) = \mathbf{0}_d, \\ \mathbf{z}_2(\theta, \lambda, \beta) = \int_{\Upsilon} \left(\int_{\mathbb{R}} \frac{\dot{\mathbf{g}}(\mathbf{x}, \beta) \lambda}{K(\mathbf{x}, e, \theta)} dP_{\mathbf{x}}(e) - \int_{\mathbb{R}} \frac{\dot{\mathbf{g}}(\mathbf{x}, \beta) \lambda}{1 - K(\mathbf{x}, e, \theta)} dQ_{\mathbf{x}}(e) \right) dH(\mathbf{x}) = \mathbf{0}_d, \end{cases} \quad (19)$$

ou,

$$\mathbf{z}_1(\theta, \lambda, \beta) = \frac{\partial \psi(\theta, \lambda, \beta)}{\partial \lambda}, \quad \mathbf{z}_2(\theta, \lambda, \beta) = \frac{\partial \psi(\theta, \lambda, \beta)}{\partial \beta}.$$

Pour la suite, on a besoin des suppositions suivantes :

(A5) La matrice $\begin{pmatrix} -\frac{\partial \mathbf{z}_1(\theta, \lambda, \beta)}{\partial \lambda} & -\frac{\partial \mathbf{z}_1(\theta, \lambda, \beta)}{\partial \beta} \\ -\frac{\partial \mathbf{z}_2(\theta, \lambda, \beta)}{\partial \lambda} & -\frac{\partial \mathbf{z}_2(\theta, \lambda, \beta)}{\partial \beta} \end{pmatrix}$ est inversible pour tout $\theta \in (0, 1)$.

(A6) Les deux intégrales suivantes signifient qu'elles sont appliquées composante par composante à la matrice correspondante.

$$\begin{aligned} \int_{\Upsilon} \int_{\mathbb{R}} \left(\frac{\mathbf{g}(\mathbf{x}, \beta) \mathbf{g}^t(\mathbf{x}, \beta)}{K^2(\mathbf{x}, e, \theta)} + \frac{\mathbf{g}(\mathbf{x}, \beta) \mathbf{g}^t(\mathbf{x}, \beta)}{(1 - K(\mathbf{x}, e, \theta))^2} \right) d(F_{\mathbf{x}}(e) + G_{\mathbf{x}}(e)) < \infty, \\ \int_{\Upsilon} \int_{\mathbb{R}} \left(\frac{\dot{\mathbf{g}}(\mathbf{x}, \beta) \dot{\mathbf{g}}(\mathbf{x}, \beta)}{K^2(\mathbf{x}, e, \theta)} + \frac{\dot{\mathbf{g}}(\mathbf{x}, \beta) \dot{\mathbf{g}}(\mathbf{x}, \beta)}{(1 - K(\mathbf{x}, e, \theta))^2} \right) d(F_{\mathbf{x}}(e) + G_{\mathbf{x}}(e)) < \infty. \end{aligned}$$

Soit $\mathbf{g}_l(\mathbf{x}, \beta)$ la l -ème composante du vecteur $\mathbf{g}(\mathbf{x}, \beta)$. Pour tout $1 \leq j, k \leq d$, on suppose que

$$\int_{\Upsilon} \int_{\mathbb{R}} \left(\frac{1}{K^2(\mathbf{x}, e, \theta)} \frac{\partial^2 \mathbf{g}_l(\mathbf{x}, \beta)}{\partial \beta_j \partial \beta_k} + \frac{1}{(1 - K(\mathbf{x}, e, \theta))^2} \frac{\partial^2 \mathbf{g}_l(\mathbf{x}, \beta)}{\partial \beta_j \partial \beta_k} \right) d(F_{\mathbf{x}}(e) + G_{\mathbf{x}}(e)) < \infty.$$

(A7) Les fonctions $f(\mathbf{x}, \beta)$ et $\dot{\mathbf{f}}(\mathbf{x}, \beta)$ sont équicontinues en β dans Γ .

Remarque 3.2 Une condition suffisante pour que les fonctions $f(\mathbf{x}, \beta)$ et $\dot{\mathbf{f}}(\mathbf{x}, \beta)$ soient équicontinues est qu'elles soient Lipschitziennes par rapport à β dans Γ .

Dans le théorème suivant, on montre que si θ_{nk} converge vers la vraie valeur θ_0 . Alors, le maximum de la statistique de test modifiée converge vers le maximum de sa distribution limite.

Théorème 3.4 Sous l'hypothèse alternative H_1 , si les suppositions (A5)-(A7) sont vérifiées, $\theta_0 \in (0, 1)$ et $\lim_{n \rightarrow \infty} \theta_{nk} = \theta \in [0, 1]$, alors $\Lambda_{nk}(\theta_{nk}) \xrightarrow[n \rightarrow \infty]{p.s.} \psi(\theta, \tilde{\lambda}(\theta), \tilde{\beta}(\theta))$. De plus, $\psi(\theta, \tilde{\lambda}(\theta), \tilde{\beta}(\theta))$ est une fonction strictement croissante en $(0, \theta_0)$ et décroissante en $(\theta_0, 1)$ et $\max_{0 \leq \theta \leq 1} \psi(\theta, \tilde{\lambda}(\theta), \tilde{\beta}(\theta)) = \psi(\theta_0, \tilde{\lambda}(\theta_0), \tilde{\beta}(\theta_0))$. ■

Test de la vraisemblance empirique pour un modèle non-linéaire avec change-points.

Par le théorème suivant, on prouve que la statistique de test \tilde{T}_n donnée par la relation (13) a la puissance asymptotique égale à 1.

Théorème 3.5 Sous les suppositions (A1)-(A7), la puissance du test du rapport du vraisemblance empirique \tilde{T}_n converge vers 1. ■

Références

- [1] Boldea, O., Hall, A.R. (2013), Estimation and inference in untableau nonlinear least squares models. *Journal of Econometrics*, **172**(1), 158-167.
- [2] Ciuperca, G., Salloum, Z. (2015), Empirical likelihood test in a posteriori change-point nonlinear model, *Metrika*, DOI 10.1007/s00184-015-0534-z.
- [3] Liu, Y., Zou, C., Zhang, R. (2008), Empirical likelihood ratio test for a change-point in linear regression model, *Communications in Statistics-Theory and Methods*, **37**, 2551-2563.
- [4] Seber G., Wild C., (2003), *Nonlinear regression* Wiley series in probability and mathematical Statistics, Wiley, Hoboken, NJ.

ESTIMATION PAR ALGORITHME EM POUR UN MODÈLE À FACTEURS ET À ÉQUATION STRUCTURELLE

Tami Myriam, Xavier Bry, Christian Lavergne
IMAG, Universités de Montpellier
Place Eugène Bataillon, Montpellier, France
myriam.tami@umontpellier.fr, xavier.bry@umontpellier.fr,
christian.lavergne@univ-montp3.fr

Résumé - *Nous proposons une nouvelle méthode d'estimation d'un modèle à facteurs et à équation structurelle. Notre méthode est fondée sur la maximisation de la fonction de vraisemblance par algorithme EM. Nous montrerons que cette approche permet les estimations des facteurs en plus de celles des coefficients du modèle. À travers une étude sur données simulées et d'une analyse de sensibilité nous avons évalué les performances de cette méthode. Nous présenterons une application de cette approche sur des données réelles environnementales.*

Mots clés - **Algorithme EM, modèles à équations structurelles, Modèles à facteurs, variables latentes, estimation.**

1 Introduction

Nous nous plaçons dans le contexte des modèles à équations structurelles (*Structural Equation Models, SEM*). On suppose avoir observé plusieurs groupes de variables Y, X^1, \dots, X^p sur les mêmes unités. De chacun de ces groupes est extrait une variable non observable dite latente. Ces variables latentes sont alors liées par des relations de causalité. Le modèle peut aussi être enrichi de variables explicatives également observées. Dans la littérature, deux familles de méthodes existent. L'une, connue sous le nom de PLS (*Partial Least Squares*) et développée par Herman Wold (1975) et ses successeurs, repose sur l'usage de composantes comme variables latentes. La seconde, issue des travaux de Karl Jöreskog (1970), est fondée sur le maximum de vraisemblance et fait appel à des facteurs comme variables latentes. La méthode engendrée porte le nom de LISREL (*LINear Structural RELations*). Ces deux approches ont été comparées dans plusieurs travaux [SS06]. Dans un objectif d'interprétation a priori du modèle, ces derniers ont montré qu'il est préférable d'utiliser un modèle à facteurs plutôt qu'un modèle à composantes, lequel est computationnellement plus efficace mais trop contraint. Nous nous plaçons dans le paradigme de l'estimation par maximum de vraisemblance et nous proposons un algorithme EM ; algorithme bien adapté dans le cas de variables latentes. Contrairement aux méthodes citées plus haut, cette approche a l'avantage d'estimer les facteurs en plus des paramètres du modèle, tout en restant efficace en terme de temps de calcul. Nous présenterons une application de cette méthode sur des données réelles.

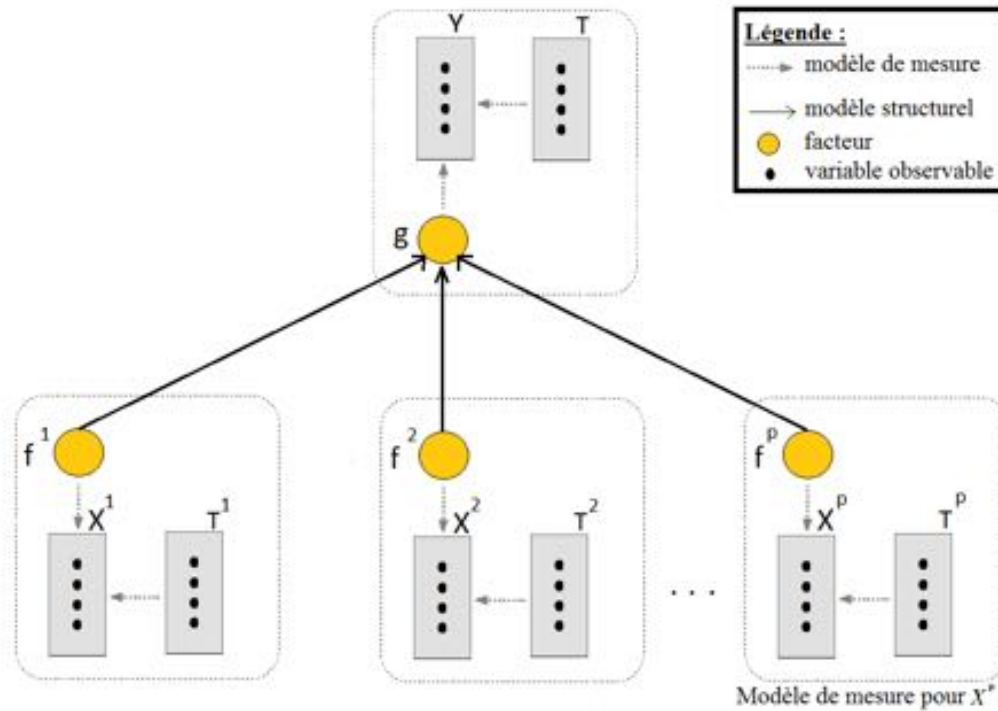


FIGURE 1 – Modèle structurel à p groupes explicatifs et un groupe dépendant

2 Structure générale du modèle

2.1 Notations du modèle à p groupes explicatifs et un groupe dépendant

Nous nous plaçons dans le cadre d'un modèle d'équations structurelles à variables latentes où pour $p \in \llbracket 1, m \rrbracket$ chacun des groupes de variables Y, X^1, \dots, X^p dépend d'une variable latente (respectivement de g, f^1, \dots, f^p). Au niveau du modèle interne (dit aussi structurel), la variable latente g est dépendante de l'ensemble des variables latentes f^1, \dots, f^p (cf. figure 1). Quant aux modèles externes (dit aussi de mesure), chacun forme un groupe de variables $X^m = (X_1^m, \dots, X_{q_m}^m)$ (resp. $Y = (Y_1, \dots, Y_{q_Y})$) comme dépendant du facteur f^m (resp. g). Dépendances pouvant être chacune enrichie par une dépendance supplémentaire aux covariables T^m (resp. T).

Notons que ce modèle peut se généraliser à plusieurs facteurs dans chaque groupe sans difficultés.

2.2 Formulation du modèle à p groupes explicatifs et un groupe dépendant

Introduisons D (resp. D^m) une matrice $r_T \times q_Y$ (resp. $r_m \times q_m$) de coefficients pondérateurs, b (resp. a^m) un vecteur $1 \times q_Y$ (resp. $1 \times q_m$) de coefficients pondérateurs et ε^Y (resp. ε^m) une matrice des erreurs $n \times q_Y$ (resp. $n \times q_m$), associées au groupe de variable Y (resp. X^m). On notera également ε^g la matrice des erreurs associées à g . Le modèle peut alors être formulé

ainsi :

$$\begin{cases} Y & = TD + gb' + \varepsilon^Y \\ \forall m \in \llbracket 1, p \rrbracket, X^m & = T^m D^m + f^m a^{m'} + \varepsilon^m \\ g & = f^1 c^1 + \dots + f^p c^p + \varepsilon^g \end{cases} \quad (1)$$

Nous imposons que les éléments de la première colonne des matrices de covariables T et T^m soient fixés à 1. Ainsi, la première ligne de D et de chaque matrice D^m correspondront aux paramètres de moyenne. On y adjoint, sous contraintes d'identifiabilité, les hypothèses suivantes :

$$\forall m \in \llbracket 1, p \rrbracket, f^m \sim \mathcal{N}(0, 1); \varepsilon_i^m \sim \mathcal{N}(0, \psi_m), \varepsilon_i^Y \sim \mathcal{N}(0, \psi_Y)$$

où $\psi_m = \text{diag}(\sigma_{m,j}^2)_{j \in \llbracket 1, q_m \rrbracket}$ de dimension $q_m \times q_m$ et $\psi_Y = \text{diag}(\sigma_{Y,j}^2)_{j \in \llbracket 1, q_Y \rrbracket}$ de dimension $q_Y \times q_Y$; $\varepsilon_i^g \sim \mathcal{N}(0, 1)$; $g \sim \mathcal{N}(0, (c^1)^2 + \dots + (c^p)^2 + 1)$ et enfin, ε_i^g , mutuellement indépendants des f^m pour tout observation $i \in \llbracket 1, n \rrbracket$ et $\forall m \in \llbracket 1, p \rrbracket$ ε^Y et ε^m sont indépendant. .

3 Estimation par algorithme EM

L'algorithme EM de Dempster, Laird et Rubin (1977) [DLR77] est une procédure générale pour maximiser la vraisemblance. Il est souvent utilisé dans le cas de problèmes à données manquantes. Dans le cadre des modèles d'équations structurelles à variables latentes, les données manquantes correspondent aux facteurs. Afin d'estimer les paramètres du modèle, cet algorithme procède en deux étapes E (pour "Expectation") et M (pour "Maximization"). Nous le présentons avec une restriction à $p = 2$ groupes explicatifs dans les sections suivantes.

3.1 L'algorithme EM dans le cadre d'un modèle du type 2 groupes explicatifs et un groupe dépendant

Nous sommes ici en présence de deux groupes explicatifs et un groupe dépendant. Soit q_Y (resp. q_1, q_2) le nombre de variables étudiées Y_j (resp. X_j^1, X_j^2) et n le nombre d'observations dont on dispose pour chacune des variables. Pour $j \in \llbracket 1, q_m \rrbracket$ ou $j \in \llbracket 1, q_Y \rrbracket$ on écrira $D_{,j}$ et b_j (resp. $D_{,j}^1, D_{,j}^2, a_{,j}^1, a_{,j}^2$) les coefficients pondérateurs associés à Y_j (resp. X_j^1, X_j^2) et ε_{ij}^Y (resp. $\varepsilon_{ij}^1, \varepsilon_{ij}^2$) les erreurs, associées à la variable Y_j (resp. X_j^1, X_j^2). On notera également ε_i^g les erreurs associées à g . Pour i une observation, le modèle peut être formulé selon le système d'équations suivant :

$$\begin{cases} y_i' & = t_i' D + g_i b' + \varepsilon_i^{y'} \\ x_i^{1'} & = t_i^{1'} D^1 + f_i^1 a^{1'} + \varepsilon_i^{1'} \\ x_i^{2'} & = t_i^{2'} D^2 + f_i^2 a^{2'} + \varepsilon_i^{2'} \\ g_i & = f_i^1 c^1 + f_i^2 c^2 + \varepsilon_i^g \end{cases} \quad (2)$$

Où nous faisons les simplifications suivantes $\psi_Y = \sigma_Y^2, \psi_1 = \sigma_1^2, \psi_2 = \sigma_2^2$. On note $\theta = \{D, D^1, D^2, b, a^1, a^2, c^1, c^2, \sigma_Y^2, \sigma_1^2, \sigma_2^2\}$ l'ensemble des paramètres du modèle. Ainsi, la dimension de θ est :

$$K = 5 + q_Y(r_T + 1) + \sum_{m=1}^2 q_m(r_m + 1)$$

Les hypothèses de ce modèle sont du même type que celles du modèle à p groupes explicatifs et un groupe dépendant avec $p = 2$.

Pour $z = (y, x^1, x^2)$ et $h = (g, f^1, f^2)$ la log vraisemblance s'écrit :

$$\begin{aligned} \mathcal{L}(\theta; z, h) = & -\frac{1}{2} \sum_{i=1}^n \{ \ln|\psi_Y| + \ln|\psi_1| + \ln|\psi_2| \\ & + (y_i - D't_i - g_i b)' \psi_Y^{-1} (y_i - D't_i - g_i b) \\ & + (x_i^1 - D^1 t_i^1 - f_i^1 a^1)' \psi_1^{-1} (x_i^1 - D^1 t_i^1 - f_i^1 a^1) \\ & + (x_i^2 - D^2 t_i^2 - f_i^2 a^2)' \psi_2^{-1} (x_i^2 - D^2 t_i^2 - f_i^2 a^2) \\ & + (g_i - c^1 f_i^1 - c^2 f_i^2)^2 + (f_i^1)^2 + (f_i^2)^2 \} + \lambda \end{aligned}$$

Où θ est l'ensemble des paramètres du modèle de dimension K et λ une constante. Cependant, à cause de la simplification du modèle (1), dans notre cas (2), $\theta = \{D, D^1, D^2, b, a^1, a^2, c^1, c^2, \sigma_Y^2, \sigma_1^2, \sigma_2^2\}$. En effet, $\psi_Y = \sigma_Y^2 Id_{q_Y}$, $\psi_1 = \sigma_1^2 Id_{q_1}$ et $\psi_2 = \sigma_2^2 Id_{q_2}$.

Pour maximiser la fonction log-vraisemblance par algorithme EM nous résolvons :

$$\mathbb{E}_z^h \left[\frac{\partial}{\partial \theta} \mathcal{L}(\theta; z, h) \right] = 0 \quad (3)$$

(cf. [Fou02]).

Pour ce faire, nous utilisons pour chaque i une observation :

$$h_i | z_i \sim \mathcal{N} \left(m_i = \begin{pmatrix} m_{1i} \\ m_{2i} \\ m_{3i} \end{pmatrix}, \Sigma_i = \begin{pmatrix} \sigma_{11i} & \sigma_{12i} & \sigma_{13i} \\ \sigma_{21i} & \sigma_{22i} & \sigma_{23i} \\ \sigma_{31i} & \sigma_{32i} & \sigma_{33i} \end{pmatrix} \right)$$

Et nous notons :

$$\begin{aligned} \tilde{\gamma}_i &= \mathbb{E}_{z_i}^{h_i} [g_i^2] = (\mathbb{E}_{z_i}^{h_i} [g_i])^2 + \mathbb{V}_{z_i}^{h_i} [g_i] = m_{1i}^2 + \sigma_{11i}, & \tilde{g}_i &= \mathbb{E}_{z_i}^{h_i} [g_i] = m_{1i}, \\ \tilde{\phi}_i^1 &= \mathbb{E}_{z_i}^{h_i} [(f_i^1)^2] = (\mathbb{E}_{z_i}^{h_i} [f_i^1])^2 + \mathbb{V}_{z_i}^{h_i} [f_i^1] = m_{2i}^2 + \sigma_{22i}, & \tilde{f}_i^1 &= \mathbb{E}_{z_i}^{h_i} [f_i^1] = m_{2i}, \\ \tilde{\phi}_i^2 &= \mathbb{E}_{z_i}^{h_i} [(f_i^2)^2] = (\mathbb{E}_{z_i}^{h_i} [f_i^2])^2 + \mathbb{V}_{z_i}^{h_i} [f_i^2] = m_{3i}^2 + \sigma_{33i}, & \tilde{f}_i^2 &= \mathbb{E}_{z_i}^{h_i} [f_i^2] = m_{3i}. \end{aligned}$$

Nous obtenons des formules solutions explicites de (3) que nous ne présentons pas ici pour ne pas alourdir le document. Cependant vous pouvez y avoir accès sur l'article en preprint [TBL15].

3.2 L'algorithme

Pour estimer les paramètres de θ et les facteurs g, f^m , nous proposons l'algorithme qui suit où nous notons $[t]$ la t^{ieth} -itération de l'algorithme.

1. Initialisation = choix des valeurs initiales des paramètres $\theta^{[0]}$.
2. À l'itération courante $t \geq 1$, jusque satisfaction du critère d'arrêt on procède comme suit :
 - (a) **E-step** : Avec $\theta^{[t-1]}$,
 - i. On calcule explicitement la distribution $h_i | z_i$ pour tout $i \in \llbracket 1, n \rrbracket$.
 - ii. On estime les valeurs des facteurs $\tilde{g}^{[t]}, \tilde{f}^m^{[t]}$, $m \in \{1, 2\}$.

iii. On calcule $\tilde{\gamma}^{[t]}$ and $\tilde{\phi}^m^{[t]}$, $m \in \{1, 2\}$.

(b) **M-step** :

i. On actualise θ à $\theta^{[t]}$ en introduisant $\tilde{g}^{[t]}$, $\tilde{\gamma}^{[t]}$ et $\tilde{f}^m^{[t]}$, $\tilde{\phi}^m^{[t]}$, $m \in \{1, 2\}$ dans les formules solutions de (3) (cf. [TBL15]).

3. Nous utilisons le critère d’arrêt suivant avec ϵ le plus petit possible :

$$\sum_{k=1}^K \frac{|\theta^{*[t+1]}[k] - \theta^{*[t]}[k]|}{\theta^{*[t+1]}[k]} < \epsilon$$

où θ^* est le vecteur de dimension K contenant tous les paramètres scalaires de l’ensemble des paramètres de θ .

4 Performances de l’approche et application

Suite à une analyse de sensibilité, nous montrerons les performances de la méthodes. Diverses simulations seront présentées ainsi qu’une application sur des données réelles de notre méthode d’estimation. Application qui nous permettra de proposer un modèle explicatif.

Références

- [Bac87] F. Bacher. LES MODÈLES STRUCTURAUX EN PSYCHOLOGIE PRÉSENTATION D’UN MODÈLE : LISREL Première partie. *Le Travail Humain*, 50(4) :347–370, January 1987.
- [Bac88] F. Bacher. LES MODÈLES STRUCTURAUX EN PSYCHOLOGIE PRÉSENTATION D’UN MODÈLE : LISREL. *Le Travail Humain*, 51(4) :273–288, January 1988.
- [Bol14] Kenneth A. Bollen. *Structural Equations with Latent Variables*. John Wiley & Sons, August 2014.
- [BRVC12] Xavier Bry, Patrick Redont, Thomas Verron, and Pierre Cazes. THEME-SEER : a multidimensional exploratory technique to analyze a structural model using an extended covariance criterion. *J. Chemometrics*, 26(5) :158–169, May 2012.
- [BTVM13] X. Bry, C. Trottier, T. Verron, and F. Mortier. Supervised component generalized linear regression using a PLS-extension of the Fisher scoring algorithm. *Journal of Multivariate Analysis*, 119 :47–60, August 2013.
- [BV15] Xavier Bry and Thomas Verron. THEME : Thematic model exploration through multiple co-structure maximization. *J. Chemometrics. Accepted but unpublished*, August 2015.
- [DLR77] A. P. Dempster, N. M. Laird, and D. B. Rubin. Maximum Likelihood from Incomplete Data via the EM Algorithm. *Journal of the Royal Statistical Society. Series B (Methodological)*, 39(1) :1–38, January 1977.
- [EVT14] V Esposito Vinzi and L Trinchera. Modèles à équations structurelles, approches basées sur les composantes, 2014.
- [Fou02] Jean-Louis Foulley. Algorithme EM : Théorie et application au modèle mixte. *Journal de la Société française de statistique*, 143(3-4) :57–109, 2002.
- [HT04] Heungsun Hwang and Yoshio Takane. Generalized structured component analysis. *Psychometrika*, 69(1) :81–99, March 2004.
- [Jö70] K. G. Jöreskog. A general method for analysis of covariance structures. *Biometrika*, 57(2) :239–251, January 1970.
- [Jak07] Emmanuel Jakobowicz. *Contributions aux modèles d’équations structurelles à variables latentes*. phdthesis, Conservatoire national des arts et métiers - CNAM, October 2007.
- [Jor03] Michael I. Jordan. *An introduction to probabilistic graphical models*. preparation, 2003.

- [JS82] Karl G. Jöreskog and Dag Sörbom. Recent Developments in Structural Equation Modeling. *Journal of Marketing Research*, 19(4) :404–416, November 1982.
- [Loh13] Jan-Bernd Lohmöller. *Latent Variable Path Modeling with Partial Least Squares*. Springer Science & Business Media, November 2013.
- [MM14] George A. Marcoulides and Irini Moustaki. *Latent Variable and Latent Structure Models*. Psychology Press, April 2014.
- [Sai06] Mohamed Saidane. *Modèles à Facteurs Conditionnellement Hétéroscédastiques et à Structure Markovienne Cachée pour les Séries Financières*. phdthesis, Université Montpellier II - Sciences et Techniques du Languedoc, July 2006.
- [SS06] Valentina Stan and Gilbert Saporta. Une comparaison expérimentale entre les approches PLS et LISREL. In *38 èmes Journées de Statistique, Clamart, France, X*, France, January 2006.
- [TBL14] Myriam Tami, Xavier Bry, and Christian Lavergne. Estimation of structural equation models with factors by EM algorithm. In *JDS 2014 Rennes*, Rennes, France, June 2014.
- [TBL15] Myriam Tami, Xavier Bry, and Christian Lavergne. Em estimation of a structural equation model (preprint). Montpellier, France, September 2015.
- [TT11] Arthur Tenenhaus and Michel Tenenhaus. Regularized Generalized Canonical Correlation Analysis. *Psychometrika*, 76(2) :257–284, March 2011.
- [TVCL05] Michel Tenenhaus, Vincenzo Esposito Vinzi, Yves-Marie Chatelin, and Carlo Lauro. PLS path modeling. *Computational Statistics & Data Analysis*, 48(1) :159–205, January 2005.
- [VCHW10] Vincenzo Esposito Vinzi, Wynne W. Chin, Jörg Henseler, and Huiwen Wang. *Handbook of Partial Least Squares : Concepts, Methods and Applications*. Springer Science & Business Media, March 2010.
- [WK89] L. E. Wangen and B. R. Kowalski. A multiblock partial least squares algorithm for investigating complex chemical systems. *J. Chemometrics*, 3(1) :3–20, January 1989.

MODÉLISATION NON PARAMÉTRIQUE DE LA RÉGRESSION POUR VARIABLES EXPLICATIVES FONCTIONNELLES ET ERREURS AUTOCORRÉLÉES.

Camille Ternynck*, Sophie Dabo-Niang et Serge Guillas

*Université de Lille - Faculté de médecine et CERIM

1 Place de Verdun, 59045 Lille Cedex, France

camille.ternynck@univ-lille2.fr

Résumé - Dans cette présentation, nous introduisons une nouvelle approche basée sur l'estimateur à noyau pour estimer le modèle de régression non linéaire en présence de variables explicatives à valeurs dans un espace fonctionnel. De plus, nous considérons la situation où la variable réponse prend ses valeurs dans \mathbb{R}^d , $d \geq 1$. Par ailleurs, le processus résiduel est considéré stationnaire et autocorrélé. La procédure consiste à pré-blanchir la variable dépendante en se basant sur l'autocorrélation estimée. L'idée principale est de transformer le modèle de régression original de sorte que le terme d'erreur du modèle transformé devienne non corrélé. Nous établissons la convergence de l'estimateur de la régression ainsi que sa normalité asymptotique en considérant des variables explicatives α -mélangeantes. Bien que, dans la littérature sur les méthodes à noyau, il est généralement préférable d'ignorer entièrement la structure de corrélation, nous montrons ici que la fonction d'autocorrélation du processus des erreurs apporte de l'information utile permettant d'améliorer l'estimation de la fonction de régression. Nous appliquons l'estimateur proposé à des données simulées ainsi qu'à des données de concentration en ozone dans l'air. Lorsque le processus des erreurs présente une forte corrélation, nous constatons que notre procédure permet d'améliorer les résultats obtenus avec l'estimateur classique.

Mots clés - Régression à noyau, Séries temporelles, Pré-blanchiment, Données fonctionnelles

1 Introduction

Ce travail concerne l'étude d'un modèle de régression en séries temporelles lorsque les variables explicatives X_t sont fonctionnelles, les variables réponses \mathbf{Y}_t sont à valeurs dans \mathbb{R}^d et les termes d'erreurs $\mathbf{u}_t \in \mathbb{R}^d$ sont autocorrélés. Plus précisément, les variables explicatives X_t appartiennent à l'espace fonctionnel (\mathcal{E}, d) , muni de la semi-métrique d . Notons qu'il n'existe pas de référence sur l'estimation non paramétrique de la fonction de régression pour variables explicatives fonctionnelles lorsque la variable réponse est multivariée. La plupart des travaux sur le sujet considèrent une réponse scalaire bien que certains étudient le cas où la réponse est de nature fonctionnelle. Lorsque les variables explicatives et réponses sont multivariées, [XQP13] proposent un estimateur à noyau de la fonction de régression. Notre modèle de régression est une extension [XQP13] et s'écrit de la manière suivante :

$$\mathbf{Y}_t = \mathbf{m}(X_t) + \mathbf{u}_t, \quad t = 1, \dots, T, \quad (1)$$

où $\mathbf{Y}_t = (Y_{t,1}, \dots, Y_{t,d})' \in \mathbb{R}^d$, $\mathbf{m}(X_t) = (m_1(X_t), \dots, m_d(X_t))'$, les variables explicatives sont fonctionnelles (c'est à dire, X_t peut éventuellement prendre ses valeurs dans un espace de dimension infinie), $\mathbf{u}_t = (u_{t,1}, \dots, u_{t,d})'$. De plus, le processus stationnaire des résidus \mathbf{u}_t est autocorrélé et indépendant de X_t . Le processus $(X_t, \mathbf{Y}_t)_t$ n'est pas nécessairement strictement stationnaire, la stationnarité au second ordre est suffisante.

La particularité de ce travail est de proposer une approche basée sur l'estimateur à noyau qui permet de tenir compte de l'information contenue dans le terme d'erreur. Cette procédure est une généralisation d'un travail existant dans le cadre réel (voir [XLCM03]). En effet, l'estimateur à noyau classique, adapté au cadre des données fonctionnelles dans [FV00, FV06], ignore la structure de corrélation, induisant une perte d'information. Notre objectif est de montrer que l'information contenue dans le terme d'erreur \mathbf{u}_t permet d'améliorer l'estimation de la fonction de régression. L'idée principale est de transformer le modèle de régression original de sorte que le terme d'erreur du modèle transformé devienne non corrélé.

Dans la suite, la méthode d'estimation est introduite et les résultats asymptotiques obtenus sont énoncés. Puis, les résultats numériques issus d'une étude de simulations ainsi que l'application à des données réelles sont présentés.

2 Procédure d'estimation

Soit un échantillon $\{(X_1, \mathbf{Y}_1), \dots, (X_T, \mathbf{Y}_T)\}$, $t = 1, \dots, T$, où X_t est une variable aléatoire qui prend ses valeurs dans un espace semi-métrique (\mathcal{C}, d) de dimension infinie et $\mathbf{Y}_t \in \mathbb{R}^d$ est la variable réponse du modèle de régression non paramétrique (1). On suppose que le processus des résidus $\mathbf{u}_t \in \mathbb{R}^d$ est stationnaire, de moyenne $\mathbf{0}$ et dont les covariances croisées s'écrivent $\gamma_{\mathbf{u}}$. Ce processus admet la représentation linéaire inversible suivante (avec coefficients bornés) :

$$\mathbf{u}_t = \sum_{k=0}^{\infty} \Psi_k \mathbf{e}_{t-k} = \Psi(L) \mathbf{e}_t$$

où $\Psi_0 = I$ (matrice identité), $\Psi(L) = \sum_{k=0}^{\infty} \Psi_k L^k$ est une matrice $d \times d$ avec L l'opérateur retard, les $\mathbf{e}_t \in \mathbb{R}^d$ forment un bruit blanc de moyenne $\mathbb{E}(\mathbf{e}_t) = \mathbf{0}$, la matrice $\mathbb{E}(\mathbf{e}_t \mathbf{e}_t') = \Sigma_{\mathbf{e}}$ est définie positive, $\mathbb{E}(\mathbf{e}_t \mathbf{e}_{t+k}) = \mathbf{0}$ pour $k \neq 0$ et $\mathbb{E}[|e_{t,j}|] < \infty$, $\forall j = 1, \dots, d$. On pose $\Psi(L)^{-1} = \Pi(L) = I - \sum_{k=1}^{\infty} \Pi_k L^k$ avec $\Pi_0 = I$. Ainsi, nous avons la représentation autorégressive inversible suivante :

$$\Pi(L) \mathbf{u}_t = \mathbf{e}_t.$$

Nous considérons une version tronquée de $\Pi(L)$ à l'ordre Q . On applique $\Pi(L)$ à l'équation (1) et on obtient $\Pi(L) \mathbf{Y}_t = \Pi(L) \mathbf{m}(X_t) + \mathbf{e}_t$. Puis, nous écrivons le modèle transformé de la régression $\underline{\mathbf{Y}}_t = \mathbf{m}(X_t) + \mathbf{e}_t$, avec $\underline{\mathbf{Y}}_t = \mathbf{Y}_t - \sum_{k=1}^Q \Pi_k L^k (\mathbf{Y}_t - \mathbf{m}(X_t))$, ainsi le terme d'erreur du modèle transformé devient non corrélé.

La matrice des coefficients $\{\Psi_k\}_{k=0}^{\infty}$ et la fonction de régression $\mathbf{m}(\cdot)$ sont inconnus. Si $\underline{\mathbf{Y}}_t$ était connu, l'estimateur non paramétrique de la régression de $\underline{\mathbf{Y}}_t$ sur X_t serait plus efficace que

l'estimateur à noyau classique. Dans ce travail, on considère un estimateur de type Nadaraya-Watson comme introduit dans [FV04], [Mas05], [DNR09] où pour un échantillon $\{V_t, X_t\}$, l'estimation de la régression de $V_t \in \mathbb{R}$ sur $X_t \in (\mathcal{C}, d(\cdot, \cdot))$ s'écrit

$$\frac{\sum_{t=1}^T V_t K\left(\frac{d(x, X_t)}{h}\right)}{\sum_{s=1}^T K\left(\frac{d(x, X_s)}{h}\right)}, \quad x \in \mathcal{C}$$

où $K(\cdot)$ est une fonction sur $[0, +\infty[$ appelée noyau, $h > 0$ est un paramètre de lissage et $d(\cdot, \cdot)$ est une semi-métrie. Pour $x \in \mathcal{C}$ fixé, on écrit $\hat{m}_j(x)$ l'estimateur lorsque $V_t = Y_{t,j}$ et $\bar{m}_j(x)$ lorsque $V_t = \underline{Y}_{t,j}$. Soit K_0 et K_1 deux noyaux définis sur $[0; +\infty[$, h_0 et h_1 les deux paramètres de lissage correspondants. On considère l'estimateur $\bar{\mathbf{m}}(x) = (\bar{m}_1(x), \dots, \bar{m}_d(x))'$ où

$$\bar{m}_j(x) = \frac{\frac{1}{T\mathbb{E}\left[K_0\left(\frac{d(x, X_1)}{h_0}\right)\right]} \sum_{t=1}^T \underline{Y}_{t,j} K_0\left(\frac{d(x, X_t)}{h_0}\right)}{\frac{1}{T\mathbb{E}\left[K_0\left(\frac{d(x, X_1)}{h_0}\right)\right]} \sum_{s=1}^T K_0\left(\frac{d(x, X_s)}{h_0}\right)}, \quad j = 1, \dots, d$$

En pratique, la matrice des coefficients $\{\Pi_k\}_{k=1}^Q$ est inconnue et donc $\underline{\mathbf{Y}}_t$ n'est pas calculable, ainsi la régression $\underline{\mathbf{Y}}_t = \mathbf{m}(X_t) + \mathbf{e}_t$ et $\bar{\mathbf{m}}(x)$ ne sont pas manipulables. Un estimateur pratique est obtenu en remplaçant $\underline{\mathbf{Y}}_t$ par son approximation basée sur les estimations de $\{\Pi_k\}_{k=1}^Q$. La procédure d'estimation proposée est la suivante et étend celle de [XLCM03].

1. Pour $j = 1, \dots, d$: on obtient un estimateur de \mathbf{m} par la régression de \mathbf{Y}_t sur X_t en considérant le noyau K_0 et le paramètre de lissage h_0 et en supposant les erreurs i.i.d. On écrit cet estimateur $\hat{\mathbf{m}}(X_t)$ et on calcule les résidus estimés $\hat{\mathbf{u}}_t = \mathbf{Y}_t - \hat{\mathbf{m}}(X_t)$, $\hat{\mathbf{u}}_t = (u_{t,1}, \dots, u_{t,d})'$, $\hat{u}_{t,j} = Y_{t,j} - \hat{m}_j(X_t)$.
2. On modélise les erreurs estimées sous la forme d'un processus $VAR(Q)$ et on estime les matrices des coefficients du $VAR(Q)$ de l'autorégression de $\hat{\mathbf{u}}_t$: $\hat{\mathbf{u}}_t = \hat{\Pi}_1 \hat{\mathbf{u}}_{t-1} + \dots + \hat{\Pi}_Q \hat{\mathbf{u}}_{t-Q} + \mathbf{e}_t$, où $\mathbf{e}_t = (e_{t,1}, \dots, e_{t,d})'$ est un vecteur de bruits i.i.d.
3. On approche $\underline{\mathbf{Y}}_t$, $t = 2, \dots, T$, c'est à dire $\hat{\underline{\mathbf{Y}}}_t = \mathbf{Y}_t - \hat{\Pi}_1 (\mathbf{Y}_{t-1} - \hat{\mathbf{m}}(X_{t-1})) - \dots - \hat{\Pi}_Q (\mathbf{Y}_{t-Q} - \hat{\mathbf{m}}(X_{t-Q}))$. L'estimateur proposé de $\mathbf{m}(x)$ est donc obtenu par la régression de $\hat{\underline{\mathbf{Y}}}_t$ sur X_t avec K_1 le noyau correspondant et h_1 le paramètre de lissage. Il en résulte l'estimation suivante de $\tilde{\mathbf{m}}(x) = (\tilde{m}_1(x), \dots, \tilde{m}_d(x))'$:

$$\tilde{m}_j(x) = \frac{\frac{1}{T\mathbb{E}\left[K_1\left(\frac{d(x, X_1)}{h_1}\right)\right]} \sum_{t=2}^T \hat{\underline{Y}}_{t,j} K_1\left(\frac{d(x, X_1)}{h_1}\right)}{\frac{1}{T\mathbb{E}\left[K_1\left(\frac{d(x, X_1)}{h_1}\right)\right]} \sum_{s=2}^T K_1\left(\frac{d(x, X_s)}{h_1}\right)}$$

Dans la suite, pour simplifier, on considère $Q = 1$. On suppose que le processus des erreurs $\{\mathbf{u}_t\}$ est indépendant du processus $\{X_t\}$ et que $\mathbb{E}[\mathbf{e}_t | X_t] = 0$. Nous considérons que le processus $\{X_t, \mathbf{Y}_t\}$ est α -mélangeant, les cas le plus général de variables faiblement dépendantes.

3 Hypothèses et résultats de convergence

Tout d'abord, nous énonçons les hypothèses permettant d'obtenir les résultats de convergence de l'estimateur $\tilde{\mathbf{m}}(x)$.

H1 (*Conditions de régularité*)

1. m_j est une fonction de Lipschitz bornée : $|m_j(u) - m_j(v)| \leq c_3 d(u, v)^\beta$ pour tout $u, v \in (\mathcal{C}, d)$ et $\beta > 0$.
2. On pose $\mathbf{G}_2(u) = \text{Var}[\mathbf{Y}_t | X_t = u]$, $u \in (\mathcal{E}, d)$, la matrice de variance de \mathbf{Y}_t sachant $X_t = u$.
 $\mathbf{G}_2(u)$ est indépendant de t et continu dans un certain voisinage de x .

$$\sup_{\{u: d(x, u) \leq h\}} |\mathbf{G}_2(u) - \mathbf{G}_2(x)| = o(1) \quad \text{as } h \rightarrow 0$$

Supposons que $\mathbb{E}|\mathbf{Y}_t|^\nu < \infty$ et $\mathbb{E}[|\mathbf{e}_t|^\nu] < \infty$ pour un certain $\nu > 2$ et

$$\mathbf{G}_\nu(u) = \mathbb{E}[|\mathbf{Y}_t - \mathbf{m}(x)|^\nu | X_t = u]$$

est continu dans un certain voisinage de x .

3. Pour $t \neq s$ et $u, v \in (\mathcal{E}, d)$,

$$\mathbf{G}(u, v; x) = \mathbb{E}[(\mathbf{Y}_t - \mathbf{m}(x))(\mathbf{Y}_s - \mathbf{m}(x)) | X_t = u, X_s = v],$$

ne dépend pas de t, s et est continu dans un certain voisinage de (x, x) .

H2 (*Conditions sur les noyaux*)

Les noyaux K_i ($i = 0$ ou 1) sont des noyaux bornés, symétriques, non négatifs avec support compact $[0, 1]$ satisfaisant

1. $\int K_i(u) du = 1$ et $c_1 \mathbf{1}_{[0,1]} < K_i < c_2 \mathbf{1}_{[0,1]}$, c_1 et c_2 sont deux constantes finies.
2. Pour $j = 1, 2$, nous avons $I_j(h_i) \rightarrow C_j$ lorsque $h_i \rightarrow 0$, pour une certaine constante positive C_j , avec

$$I_j(h_i) = \frac{1}{\phi(h_i)/h_i} \int_0^1 K_i^j(u) \phi'(h_i v) dv$$

où $\phi(\cdot)$ est définie ci-après.

Soit $\mathcal{B}(x, h)$ une boule de rayon h centrée en $x \in (\mathcal{E}, d)$ et

$$\begin{aligned} F_x^t(h) &= \mathbb{P}[X_t \in \mathcal{B}(x, h)] \\ F_{x,x}^{s,t}(h) &= \mathbb{P}[(X_t, X_s) \in \mathcal{B}(x, h) \times \mathcal{B}(x, h)] \\ F_{x,y}^{s,t}(h) &= \mathbb{P}[(X_t, X_s) \in \mathcal{B}(x, h) \times \mathcal{B}(y, h)] \end{aligned}$$

où $F_x^t(h)$ correspond à la notion de *probabilités de petites boules*. Soient f_k , $k = 1, 2$ et 3 , des fonctions finies non-négatives (uniformément bornées).

H3 (*Distributions*)

1. $F_x^t(h) = \phi(h) f_1(x)$ lorsque $h \rightarrow 0$, où $\phi(0) = 0$, $\phi(h)$ est absolument continue au voisinage de l'origine, $\sup_t f_1(X_t) < \infty$.

2. $\sup_{t \neq s} F_{x,x}^{s,t}(h) \leq \psi_1(h) f_2(x)$ lorsque $h \rightarrow 0$, où $\psi_1(h) \rightarrow 0$ lorsque $h \rightarrow 0$ et $\sup_t f_2(X_t) < \infty$. On suppose que $\psi_1(h)/\phi^2(h)$ est borné et que $\exists \zeta_1 \in (0, 1)$, $0 < F_{x,x}(h) = O(\phi(h)^{1+\zeta_1})$.
3. $\sup_{t \neq s} F_{x,y}^{s,t}(h) \leq \psi_2(h) f_3(x, y)$ lorsque $h \rightarrow 0$, où $\psi_2(h) \rightarrow 0$ lorsque $h \rightarrow 0$ et $\sup_{t,s} f_3(X_t, X_s) < \infty$. On suppose que $\psi_2(h)/\phi^2(h)$ est borné.

H4 (Mélange)

$$\sum_{l=1}^{\infty} l^\delta [\alpha(l)]^{1-2/\nu} < \infty$$

pour un certain $\nu > 2$ et $\delta > 1 - 2/\nu$. (ν est l'ordre du moment dans **H1**(2))

H5 : On pose $h_i \rightarrow 0$, $h_0/h_1 \rightarrow 0$ et $\frac{\log T}{T^{1/2}\phi(h_0)} \rightarrow 0$ lorsque $T \rightarrow \infty$. Soit $\{v_T\}$ une séquence positive d'entiers satisfaisant $v_T \rightarrow \infty$ telle que $v_T = o((T\phi(h_0))^{1/2})$ et $(T/\phi(h_0))^{1/2}\alpha(v_T) \rightarrow 0$, $Th_0^{2\beta} \rightarrow 0$ lorsque $T \rightarrow \infty$.

Sous certaines conditions, nous montrons la normalité asymptotique de $\tilde{\mathbf{m}}(x)$ dont les résultats sont exposés dans le Théorème 1.

Théorème 1 Sous **H1- H5**,

$$(T\phi(h_1))^{1/2}[\tilde{\mathbf{m}}(x) - \mathbf{m}(x) - \mathbf{B}_T(x)] \xrightarrow{L} \mathcal{N}_d(0, \Sigma_x)$$

avec $\mathbf{B}_T(x) = \mathbb{E}[\tilde{\mathbf{m}}(x)] - \mathbf{m}(x)$, $\Sigma_x = \frac{C_2 \mathbf{G}_2(x)}{C_1^2 f_1(x)}$ est la matrice ($d \times d$) de variance asymptotique, $x \in (\mathcal{C}, d)$ quelque soit $f_1(x) > 0$.

Nous obtenons également la convergence en probabilité de l'estimateur $\tilde{\mathbf{m}}(x)$, énoncée dans le théorème suivant.

Théorème 2 Sous **H1-H5**,

$$\lim_{T \rightarrow \infty} \tilde{\mathbf{m}}(x) = \mathbf{m}(x) \quad \text{en probabilité.}$$

Nous avons énoncé les résultats de convergence de l'estimateur proposé. Dans la suite, nous allons étudier son comportement sur des données simulées et réelles.

4 Applications

Nous avons d'abord considéré l'application de notre approche sur des données simulées, en considérant tout d'abord le cas univarié ($d = 1$). Les observations fonctionnelles X_t (avec $t = 1, \dots, T$) sont définies par $X_t(w) = 1 + 10e_{0,t} + 3e_{1,t}w^2 + 4e_{2,t}(1-w)^3$, $w \in [0, 1]$ où $e_{0,t}$, $e_{1,t}$ et $e_{2,t}$ sont i.i.d. de loi $\mathcal{N}(0, 1)$. Nous posons $m(X_t(w)) = \sqrt{|0.5 \int_0^1 X_t(w) dw|}$. Le processus des erreurs u_t est un processus $AR(1)$, tel que $u_t = \epsilon_t + \rho\epsilon_{t-1}$. Différentes valeurs du paramètre ρ sont considérées. Pour chaque cas étudié, le nombre de réplifications est de 200. La sélection des paramètres de lissage est faite par validation croisée. Nous reportons l'efficacité relative, calculée comme le ratio du carré des erreurs de l'estimateur proposé $\tilde{m}(x)$ et de celui de l'estimateur conventionnel $\hat{m}(x)$. Nous constatons que l'amélioration obtenue avec l'estimateur

proposé est importante, jusqu'à 25% en présence de niveaux élevés d'autocorrélation. Les améliorations moyennes de notre estimateur sont toujours positives sauf quand $\rho \leq 0.1$, c'est à dire pour un faible niveau d'autocorrélation. Des résultats similaires sont également obtenus en considérant des données bi-variés et lorsque les observations fonctionnelles présentent de la dépendance.

Dans un second temps, nous avons adapté l'estimateur de la régression à la prédiction, adaptation que nous avons appliquée à des données réelles. Notre objectif est de prédire la concentration en ozone à une certaine date non observée à partir du passé. Nous disposons de la concentration horaire en ozone du 2 Juin au 31 Août 2005 (soit 91 jours) pour différentes stations. Nous nous intéressons aux prédictions horaires en ozone sur une journée. Pour illustrer notre objectif, nous allons prédire les concentrations en ozone du 31 Août à partir des observations faites les 90 jours précédents. Nous montrons que notre méthode permet d'améliorer les prédictions obtenues avec l'estimateur à noyau classique de la régression.

5 Conclusion

Nous montrons que la procédure présentée permet d'améliorer l'estimateur à noyau classique de la fonction de régression, en présence de variables fonctionnelles et d'erreurs autocorrélées. Dans un travail en cours, cette procédure fait l'objet d'une extension plus générale. En effet, nous l'étendons au cadre des modèles à indice fonctionnel.

Références

- [DNR09] Sophie Dabo-Niang and Nouredine Rhomari. Kernel regression estimation in a banach space. *J. Statist. Plann. Inference*, 139(4) :1421–1434, 2009.
- [FV00] F. Ferraty and P. Vieu. Dimension fractale et estimation de la régression dans des espaces vectoriels semi-normés. *C. R. Math. Acad. Sci. Paris*, 330 :403–406, 2000.
- [FV04] F Ferraty and Ph Vieu. Nonparametric models for functional data, with application in regression, time series prediction and curve discrimination. *J. Nonparametr. Stat.*, 16(1-2) :111–125, 2004.
- [FV06] F. Ferraty and P. Vieu. *Nonparametric functional data analysis : theory and practice*. Springer, 2006.
- [Mas05] E. Masry. Nonparametric regression estimation for dependent functional data : asymptotic normality. *Stochastic Process. Appl.*, 115(1) :155–177, 2005.
- [Ros56] M. Rosenblatt. A central limit theorem and a strong mixing condition. *Proceedings of the National Academy of Sciences of the United States of America*, 42(1) :43, 1956.
- [XLCM03] Z. Xiao, O.B. Linton, R.J. Carroll, and E. Mammen. More efficient local polynomial estimation in nonparametric regression with autocorrelated errors. *J. Amer. Statist. Assoc.*, 98(464) :980–992, 2003.
- [XQP13] Dongdong Xiang, Peihua Qiu, and Xiaolong Pu. Nonparametric regression analysis of multivariate longitudinal data. *Statistica Sinica*, 23 :769–789, 2013.

QML ESTIMATION OF A CLASS OF ASYMMETRIC GARCH MODELS WITH EXOGENOUS COVARIATES

THIEU Le Quyen

Pierre and Marie Curie University
4 Place Jussieu, 75005 Paris, French.
thieulequyen1411@gmail.com

Abstract - *The asymptotic distribution of the Gaussian quasi-maximum likelihood estimator (QMLE) is obtained for a wide class of asymmetric GARCH models with exogenous covariates. The true value of the parameter is not restricted to belong to the interior of the parameter space, which allows us to derive tests for the significance of the parameters. In particular, the relevance of the exogenous variables can be assessed. The results are obtained without assuming that the innovations are independent, which allows conditioning on different information sets. Monte Carlo experiments and applications to financial series illustrate the asymptotic results. In particular, an empirical study demonstrates that the realized volatility can be an helpful covariate for predicting squared returns.*

Keywords - **APARCH model augmented with explanatory variables, Boundary of the parameter space, Power-transformed and Threshold GARCH with exogenous covariates.**

1 Introduction

The GARCH-type models are of the form

$$\varepsilon_t = \sigma_t \eta_t, \tag{1}$$

where the squared volatility σ_t^2 is the best predictor of ε_t^2 given a certain information set \mathcal{F}_{t-1} available at time t . For the usual GARCH models, \mathcal{F}_{t-1} is simply the sigma-field generated by the past returns $\{\varepsilon_u, u < t\}$, and the volatility has a parametric form $\sigma_t = \sigma(\varepsilon_u, u < t; \boldsymbol{\theta}_0)$, where $\boldsymbol{\theta}_0$ is a vector of parameters. It is however often the case that some extra information is available, under the form of a vector \boldsymbol{x}_{t-1} of exogenous covariates, such as the daily volume of transactions, or high frequency intraday data, or even series of other returns. It is natural to try to take advantage of the extra information, in order to improve the prediction of the squares. To incorporate the information conveyed by $\{\boldsymbol{x}_u, u < t\}$ into \mathcal{F}_{t-1} , researchers have considered GARCH models augmented with additional explanatory variables, the so-called GARCH-X models, which are of the form $\sigma_t = \sigma(\varepsilon_u, \boldsymbol{x}_u, u < t; \boldsymbol{\vartheta}_0)$, where $\boldsymbol{\vartheta}_0 = (\boldsymbol{\theta}'_0, \boldsymbol{\pi}'_0)'$ is a vector of parameters including a parameter $\boldsymbol{\theta}_0$ specific to the past returns and a parameter $\boldsymbol{\pi}_0$ related to the exogenous covariates.

Questions that seem particularly relevant in the GARCH-X framework are: is it really useful to introduce covariates in the volatility? which covariates should we add to \mathcal{F}_{t-1} ? how many

lagged values should we consider in the GARCH formulation? Each of these questions can be discussed by testing the nullity of certain components of $\boldsymbol{\vartheta}_0$. It is thus of interest to study the behaviour of the estimator $\widehat{\boldsymbol{\vartheta}}_n$ of $\boldsymbol{\vartheta}_0$ when this parameter may stand at the boundary of the parameter space.

2 The model and strict stationarity condition

Let $x^+ = \max(x, 0)$ and $x^- = \max(-x, 0)$. The Asymmetric Power GARCH model augmented by the exogenous variables is defined by

$$\begin{cases} \varepsilon_t = h_t^{1/\delta} \eta_t \\ h_t = \omega_0 + \sum_{i=1}^q [\alpha_{0i+} (\varepsilon_{t-i}^+)^{\delta} + \alpha_{0i-} (\varepsilon_{t-i}^-)^{\delta}] + \sum_{j=1}^p \beta_{0j} h_{t-j} + \boldsymbol{\pi}'_0 \mathbf{x}_{t-1} \end{cases} \quad (2)$$

where $\mathbf{x}_t = (x_{1,t}, \dots, x_{r,t})'$ is a vector of r exogenous covariates which are assumed to be almost surely positive and the coefficients satisfy $\alpha_{0i+} \geq 0$, $\alpha_{0i-} \geq 0$, $\beta_{0j} \geq 0$, $\omega_0 > 0$, $\delta > 0$ and $\boldsymbol{\pi}_0 = (\pi_{01}, \dots, \pi_{0r})' \geq 0$ componentwise. Now assume that

A1: (η_t, \mathbf{x}'_t) is a strictly stationary and ergodic process, and there exists $s > 0$ such that $E|\eta_1|^s < \infty$ and $E\|\mathbf{x}_1\|^s < \infty$.

Lemme 1 *Suppose that A1 is satisfied. A necessary and sufficient condition for the existence of a strictly stationary, non anticipative and ergodic solution to model (2) is $\gamma < 0$, where γ is the top Lyapunov exponent of the sequence of the matrices depending on $(\eta_t^+)^{\delta}$, $(\eta_t^-)^{\delta}$.*

3 QML estimation

3.1 Strong consistency of the QMLE

We consider that δ is fixed. Let $(\varepsilon_1, \dots, \varepsilon_n)$ be a realization of length n of the stationary solution (ε_t) to the APARCH-X model (2), and let $(\mathbf{x}_1, \dots, \mathbf{x}_n)$ be the corresponding observations of the exogenous variables. Given initial values $\varepsilon_{1-q}, \dots, \varepsilon_0, \tilde{\sigma}_{1-p} \geq 0, \dots, \tilde{\sigma}_0 \geq 0, \mathbf{x}_0 \geq 0$, the QMLE of $\boldsymbol{\vartheta}_0$ is defined as any measurable solution $\widehat{\boldsymbol{\vartheta}}_n$ of

$$\widehat{\boldsymbol{\vartheta}}_n = \arg \max_{\boldsymbol{\vartheta} \in \Theta} L_n(\boldsymbol{\vartheta}) = \arg \min_{\boldsymbol{\vartheta} \in \Theta} \tilde{Q}_n(\boldsymbol{\vartheta}) \quad (3)$$

where $L_n(\boldsymbol{\vartheta})$ is the Gaussian quasi-likelihood and $\tilde{Q}_n(\boldsymbol{\vartheta}) = \frac{1}{n} \sum_{t=1}^n \frac{\varepsilon_t^2}{\tilde{\sigma}_t^2} + \ln \tilde{\sigma}_t^2$

Denote by $\mathcal{F}_{t,i}$ the σ -field generated by $\{\eta_{t-j}, j > i, \mathbf{x}_{t-k}, k > 0\}$ (so that $\mathcal{F}_{t-1} = \mathcal{F}_{t,0}$). To show the strong consistency of the QMLE, we need the following assumptions.

A2: $E(\eta_t | \mathcal{F}_{t-1}) = 0$ and $E(\eta_t^2 | \mathcal{F}_{t-1}) = 1$.

A3: $\boldsymbol{\vartheta}_0 \in \Theta$, Θ is compact.

A4: for all $i \geq 1$ (and all $t \in \mathbb{Z}$), the support of the distribution of η_{t-i} given $\mathcal{F}_{t,i}$ is not included in $[0, \infty)$ or in $(-\infty, 0]$ and contains at least three points.

A5: $\gamma < 0$ and $\sum_{j=1}^p \beta_j < 1$ for all $\boldsymbol{\vartheta} \in \Theta$.

A6: there exists $s > 0$, such that $Eh_t^s < \infty$ and $E|\varepsilon_t|^s < \infty$.

A7: if $p > 0$, $\mathcal{B}_{\vartheta_0}(z)$ has no common root with $\mathcal{A}_{\vartheta_{0+}}(z)$ and $\mathcal{A}_{\vartheta_{0-}}(z)$; $\mathcal{A}_{\vartheta_{0+}}(1) + \mathcal{A}_{\vartheta_{0-}}(1) \neq 0$ and $\alpha_{0q+} + \alpha_{0q-} + \beta_{0p} \neq 0$ (with the notation $\alpha_{00+} = \alpha_{00-} = \beta_{00} = 1$).

A8: If \mathbf{d} is a non zero vector of \mathbb{R}^r then $\mathbf{d}'\mathbf{x}_1$ is not degenerated.

Theoreme 1 Under **A1–A8**,

$$\widehat{\vartheta}_n \rightarrow \vartheta_0 \text{ a.s. as } n \rightarrow \infty.$$

3.2 Asymptotic distribution of the QMLE

The asymptotic distribution of the QMLE is studied in the following four cases:

Case A : η_t is independent of \mathcal{F}_{t-1} and all the components of ϑ_0 are strictly positive;

Case B : η_t is independent of \mathcal{F}_{t-1} and at least one component of ϑ_0 is equal to zero;

Case C : η_t is not independent of \mathcal{F}_{t-1} and all the components of ϑ_0 are strictly positive;

Case D : η_t is not independent of \mathcal{F}_{t-1} and at least one component of ϑ_0 is equal to zero.

We assume that

A9: $\mathcal{C} := \lim_{n \rightarrow \infty} \sqrt{n}(\Theta - \vartheta_0) = \prod_{i=1}^d \mathcal{C}_i$, where $\mathcal{C}_i = [0, +\infty)$ when $\vartheta_{0i} = 0$ and $\mathcal{C}_i = \mathbb{R}$ otherwise.

A10: $E\eta_t^4 < \infty$ in Cases A and B, and $E|\eta_t|^{4+\nu} < \infty$ for some $\nu > 0$ in Cases C and D.

A11: $E|\varepsilon_t|^{2\delta} < \infty$ and $E\|\mathbf{x}_t\|^2 < \infty$ in Case B, and $E|\varepsilon_t|^{2\delta+8\delta/\nu} < \infty$ and $E\|\mathbf{x}_t\|^{2+8/\nu} < \infty$ with ν satisfying **A10** in Case D.

A12: in Cases B and D, there exist Holder conjugate numbers p and $q > 1$ such that

$$p^{-1} + q^{-1} = 1 \quad \text{and} \quad E|\varepsilon_t|^{2\delta q} < \infty, \quad E|\varepsilon_t|^{2p} < \infty, \quad E\|\mathbf{x}_t\|^{2q} < \infty.$$

Theoreme 2 Under the assumptions of Theorem 1 and **A9–A12**, as $n \rightarrow \infty$,

$$\sqrt{n}(\widehat{\vartheta}_n - \vartheta_0) \xrightarrow{d} \mathbf{Z}^{\mathcal{C}} \tag{4}$$

where $\mathbf{Z} \sim \mathcal{N}\{0, \mathbf{J}^{-1}\mathbf{I}\mathbf{J}^{-1}\}$, $\mathbf{Z}^{\mathcal{C}} = \arg \inf_{\mathbf{C} \in \mathcal{C}} \|\mathbf{C} - \mathbf{Z}\|_{\mathbf{J}}$ with the norm $\|\mathbf{x}\|_{\mathbf{J}}^2 = \mathbf{x}'\mathbf{J}\mathbf{x}$ and $\mathbf{J} = \frac{4}{\delta^2} E(\mathbf{d}_t \mathbf{d}_t')$, $\mathbf{I} = \frac{4}{\delta^2} E[\{E(\eta_t^4 | \mathcal{F}_{t-1}) - 1\} \mathbf{d}_t \mathbf{d}_t']$, $\mathbf{d}_t = \frac{1}{\sigma_t^\delta} \frac{\partial \sigma_t^\delta(\vartheta_0)}{\partial \vartheta}$

4 Testing

4.1 Testing the nullity of only one coefficient

Let \mathbf{e}_k be the k -th element of the canonical basis of \mathbb{R}^d . We will test the hypothesis that the k -th element of $\boldsymbol{\vartheta}_0$ is equal to zero, assuming the other elements are positive:

$$H_0 : \mathbf{e}'_k \boldsymbol{\vartheta}_0 = 0 \text{ and } \mathbf{e}'_\ell \boldsymbol{\vartheta}_0 > 0 \quad \forall \ell \neq k \quad \text{against} \quad H_1 : \mathbf{e}'_k \boldsymbol{\vartheta}_0 > 0. \quad (5)$$

The Student t -test statistic is defined by $t_n(k) = \sqrt{n} \frac{\mathbf{e}'_k \hat{\boldsymbol{\vartheta}}_n}{\sqrt{\mathbf{e}'_k \hat{\mathbf{J}}_n^{-1} \hat{\mathbf{I}}_n \hat{\mathbf{J}}_n^{-1} \mathbf{e}_k}}$ where $\hat{\mathbf{I}}_n$ and $\hat{\mathbf{J}}_n$ are respectively estimators of the matrices \mathbf{I} and \mathbf{J} .

Corollary 1 *Under the assumptions of Theorem 2, the test of rejection region*

$$\{t_n^2(k) > \chi_1^2(1 - 2\alpha)\}$$

has the asymptotic level α under H_0 and is consistent under H_1 defined in (5).

4.2 Testing the relevance of the covariates

For ARCH-X models, the most interesting tests concern the relevance of the covariates, assuming that the GARCH parameter $\boldsymbol{\theta}_0$ is strictly positive component-wise. Without loss of generality, one can assume that we test that the last $r_0 \geq 1$ exogenous variables are irrelevant and the others are relevant. Let $\boldsymbol{\pi}_0 = (\boldsymbol{\pi}'_{01}, \boldsymbol{\pi}'_{02})'$, $\boldsymbol{\pi}_{01} = (\pi_{01}, \dots, \pi_{0,r-r_0})'$, $\boldsymbol{\pi}_{02} = (\pi_{0,r-r_0+1}, \dots, \pi_{0r})'$. Under the maintained assumption $\boldsymbol{\theta}_0 > 0$ and $\boldsymbol{\pi}_{01} > 0$ element-wise, we are interested in testing

$$H_0^\pi : \boldsymbol{\pi}_{02} = 0_{r_0} \quad \text{against} \quad H_1^\pi : \boldsymbol{\pi}_{02} \neq 0_{r_0}. \quad (6)$$

Let $K = (0_{r_0 \times (d-r_0)} : I_{r_0})$ where I_k denotes the identity matrix of size k and $0_{k_1 \times k_2}$ denotes the null matrix of size $k_1 \times k_2$, with $0_{k_1 \times k_2} = 0_{k_1}$ when $k_2 = 1$. The usual Wald test statistic is defined by $W_n = n \|\hat{\boldsymbol{\pi}}_{2n}\|_{(K \hat{\Sigma} K')^{-1}}^2$ where $\hat{\boldsymbol{\pi}}_{2n} = K \hat{\boldsymbol{\vartheta}}_n$.

Corollary 2 *Under the assumptions of Theorem 2 and H_0^π ,*

$$W_n \xrightarrow{d} W = \|K \mathbf{Z}^c\|_{(K \Sigma K')^{-1}}^2. \quad (7)$$

At the asymptotic level $\alpha \in (0, 1)$, the critical value q_α of the Wald test $\{W_n \geq q_\alpha\}$ can be obtained from the algorithm recently proposed by Pedersen (2015).

5 Numerical illustrations

5.1 Simulation experiments

We simulated the following TARX(1,1) model

$$\begin{cases} \varepsilon_t = \sigma_t \eta_t \\ \sigma_t = \omega_0 + \alpha_{0+}(\varepsilon_{t-1}^+) + \alpha_{0-}(\varepsilon_{t-1}^-) + \beta_0 \sigma_{t-1} + \pi_{01} x_{t-1} + \pi_{02} x_{t-2}. \end{cases} \quad (8)$$

The exogenous variable x_t is assumed to be the stochastic volatility defined by

$$x_t = e^{y_t}, \quad y_t = a y_{t-1} + e_t, \quad (e_t) \text{ iid } \mathcal{N}(t, \infty) \quad (9)$$

Table 1: Relative frequencies (in %) of rejection of the assumptions that the first and second lagged values of the exogenous variable do not appear in the conditional variance

	$H_0^{\pi_{01}} : \pi_{01} = 0$				$H_0^{\pi_{02}} : \pi_{02} = 0$			
	$\alpha = 1\%$		$\alpha = 5\%$		$\alpha = 1\%$		$\alpha = 5\%$	
	$n = 1,000$	$n = 2,000$	$n = 1,000$	$n = 2,000$	$n = 1,000$	$n = 2,000$	$n = 1,000$	$n = 2,000$
A	83.00	99.40	96.00	99.80	66.80	91.00	85.80	98.40
B	99.40	100.00	100.00	100.00	3.60	1.40	7.60	5.20
C	72.80	92.00	88.00	98.00	50.80	77.20	70.60	92.60
D	96.40	98.80	99.20	98.80	2.20	2.00	6.40	5.80

5.2 SP500 with realized range, volume and other indices

In this section, we built a model which aims to explain the volatility of the daily returns of the SP500 index by its past values, the realized range, the volume and the returns of the Nikkei, Nik_t , and of the FTSE, Ft_t . We fitted APARCH-X(1,1) models with $\delta \in \{0.5, 1, 1.5, 2\}$. The model with the largest likelihood is obtained for $\delta = 1$, and is given by

$$\begin{cases} \varepsilon_t = h_t \eta_t \\ h_t = \begin{matrix} 0.018 & + & 0.000 & \varepsilon_{t-1}^+ & + & 0.110 & \varepsilon_{t-1}^- & + & 0.879 & h_{t-1} & + & 4.331 & rr_{t-1} \\ (0.006) & 0.002 & (0.020) & 0.500 & & (0.035) & 0.001 & & (0.020) & 0.000 & & (1.493) & 0.002 \\ + & 0.061 & v_{t-1} & + & 0.000 & Nik_{t-1}^2 & + & 0.000 & Ft_{t-1}^2 \\ (0.026) & 0.010 & & (0.007) & 0.500 & & (0.007) & 0.500 \end{matrix} \end{cases}$$

5.3 US stocks with realized volatility

We are interested in whether the realized volatility is useful to predict the squared returns or not. In order to answer this question, we estimated APARCH-X(1,1) models of the form

$$\begin{cases} \varepsilon_t = h_t^{1/\delta} \eta_t \\ h_t = \omega + \alpha_+ (\varepsilon_{t-1}^+)^{\delta} + \alpha_- (\varepsilon_{t-1}^-)^{\delta} + \beta h_{t-1} + \pi_1 r v_{t-1}^{\delta/2} + \pi_2 r v_{t-2}^{\delta/2}, \end{cases} \quad (10)$$

with $\delta \in \{0.5, 1, 1.5, 2\}$. Table 2 displays the fitted model on each of the 29 stocks.

6 Conclusion

In this paper, we studied the asymptotic behavior of the QMLE for the PGARCH models augmented with exogenous variables when the parameter is not restricted to belong to the interior of the parameter space. The main assumptions on the exogenous variables are the stationarity and non-collinearity with the other explanatory variables of the volatility. We were also able to derive tests for the significance of the exogenous variables.

The asymptotic theory developed in the paper has been applied to simulations and real series. Our empirical results are in accordance with numerous applied studies, and complement them by providing a formal test for the significance of the exogenous variables.

Table 2: APARCH-X(1,1) models (10) fitted by QMLE on daily returns of US stock with two lagged values of realized volatilities as covariates. The estimated standard deviations are displayed into parentheses. For the estimated values of π_1 and π_2 , one star (*) means a p -value $p \in [0.01, 0.05)$ for testing the nullity of the coefficient, two stars (**) means $p \in [0.001, 0.01)$, and three stars (***) means $p < 0.001$. The last column gives the selected value of the power δ .

	ω	α_+	α_-	β	π_1	π_2	δ
AAPL	0.080 (0.029)	0.042 (0.015)	0.055 (0.013)	0.796 (0.056)	0.120* (0.072)	0.000 (0.079)	0.5
ABT	0.046 (0.027)	0.023 (0.024)	0.019 (0.027)	0.661 (0.100)	0.285*** (0.071)	0.000 (0.102)	0.5
AXP	0.027 (0.010)	0.000 (0.019)	0.074 (0.019)	0.809 (0.038)	0.155** (0.061)	0.000 (0.069)	1
BA	0.046 (0.022)	0.027 (0.020)	0.092 (0.023)	0.843 (0.037)	0.084 (0.065)	0.000 (0.071)	2
BAC	0.007 (0.010)	0.009 (0.026)	0.090 (0.029)	0.813 (0.045)	0.151** (0.063)	0.000 (0.083)	1
BMJ	0.000 (0.020)	0.051 (0.018)	0.077 (0.025)	0.880 (0.031)	0.072 (0.106)	0.000 (0.108)	1
BP	0.017 (0.015)	0.010 (0.020)	0.043 (0.017)	0.682 (0.068)	0.191* (0.086)	0.106 (0.092)	0.5
C	0.013 (0.011)	0.019 (0.025)	0.123 (0.028)	0.744 (0.056)	0.178** (0.059)	0.007 (0.077)	1
CAT	0.045 (0.022)	0.000 (0.016)	0.011 (0.017)	0.780 (0.066)	0.183** (0.074)	0.000 (0.087)	0.5
CL	0.127 (0.054)	0.032 (0.021)	0.212 (0.065)	0.424 (0.105)	0.156** (0.057)	0.193** (0.073)	2
CSCO	0.013 (0.015)	0.000 (0.020)	0.054 (0.020)	0.848 (0.034)	0.131 (0.081)	0.000 (0.084)	1
CVX	0.082 (0.030)	0.014 (0.022)	0.068 (0.025)	0.716 (0.071)	0.102* (0.062)	0.088 (0.088)	2
DELL	0.000 (0.006)	0.030 (0.012)	0.055 (0.012)	0.874 (0.034)	0.093 (0.080)	0.000 (0.085)	0.5
DIS	0.042 (0.021)	0.000 (0.016)	0.070 (0.022)	0.807 (0.056)	0.132* (0.063)	0.002 (0.069)	2
EK	0.182 (0.091)	0.071 (0.029)	0.111 (0.036)	0.583 (0.161)	0.224* (0.103)	0.000 (0.155)	0.5
EXC	0.092 (0.033)	0.056 (0.032)	0.157 (0.039)	0.677 (0.060)	0.196** (0.072)	0.000 (0.077)	1.5
F	0.060 (0.046)	0.068 (0.037)	0.075 (0.025)	0.740 (0.050)	0.091 (0.088)	0.103 (0.097)	1
FDX	0.033 (0.018)	0.012 (0.021)	0.025 (0.021)	0.803 (0.070)	0.162* (0.074)	0.000 (0.081)	0.5
GE	0.005 (0.010)	0.000 (0.019)	0.052 (0.022)	0.802 (0.050)	0.180* (0.085)	0.000 (0.076)	1
GM	0.026 (0.023)	0.018 (0.015)	0.047 (0.023)	0.881 (0.031)	0.094 (0.093)	0.000 (0.099)	2
HD	0.007 (0.010)	0.000 (0.015)	0.030 (0.013)	0.850 (0.037)	0.135* (0.070)	0.000 (0.071)	0.5
HNZ	0.007 (0.009)	0.050 (0.018)	0.084 (0.022)	0.840 (0.038)	0.105* (0.059)	0.000 (0.064)	1
HON	0.015 (0.012)	0.000 (0.023)	0.108 (0.019)	0.860 (0.030)	0.092 (0.068)	0.000 (0.065)	1
IBM	0.011 (0.007)	0.000 (0.015)	0.044 (0.016)	0.858 (0.033)	0.116* (0.063)	0.000 (0.070)	0.5
INTC	0.013 (0.009)	0.000 (0.012)	0.031 (0.013)	0.862 (0.029)	0.119* (0.061)	0.000 (0.063)	0.5
JNJ	0.022 (0.011)	0.004 (0.023)	0.176 (0.033)	0.757 (0.049)	0.144** (0.061)	0.000 (0.070)	1.5
KO	0.009 (0.016)	0.010 (0.023)	0.084 (0.028)	0.702 (0.057)	0.248** (0.082)	0.000 (0.093)	1
LLY	0.021 (0.089)	0.071 (0.029)	0.065 (0.036)	0.245 (0.146)	0.108 (0.097)	0.599* (0.316)	0.5
MCD	0.015 (0.010)	0.031 (0.015)	0.043 (0.016)	0.863 (0.029)	0.096 (0.065)	0.000 (0.070)	0.5

References

- [1] Andrews, D.W (2001), Testing when a parameter is on the boundary of the maintained hypothesis. *Econometrica* 69, p. 683-734.
- [2] Bougerol, P. and N. Picard (1992a), Strict stationarity of generalized autoregressive processes. *Annals of Probability* 20, p. 1714-1729.
- [3] Francq, C. and J-M. Zakoian (2010), *GARCH Models: Structure, Statistical Inference and Financial Applications*. John Wiley.
- [4] Han, H. and D. Kristensen (2014), Asymptotic Theory for the QMLE in GARCH-X Models With Stationary and Nonstationary Covariates. *Journal of Business and Economic Statistics* 32, p. 416-429.