ESTIMATION ADAPTATIVE DE LA DENSITÉ SPECTRALE D'UN PROCESSUS GAUSSIEN FAIBLEMENT OU FORTEMENT DÉPENDANT

PHILIPPE SOULIER

Résumé. Cette note présente une méthode d'estimation non paramétrique de la densité spectrale d'un processus gaussien fractionnaire, qui s'écrit $f(x) = |1 - e^{ix}|^{2d} f^*(x)$, où $-1/2 < d < 1/2$ et f^* est strictement positive. Nous montrons que la vitesse d'estimation ne dépend pas de d mais seulement de la régularité de f^* et donc est la même en forte et en faible dépendance. L'estimateur de régression sur le log-périodogramme réalise les vitesses minimax sur certaines classes fonctionnelles et possède des propriétés d'adaptativité lorsque la règularité de f^* est inconnue.

Abstract. This note presents an estimator of the spectral density f of a fractional Gaussian process. For such a process, f writes $f(x) = |1 - e^{ix}|^{2d} f^*(x)$, where $-1/2 < d < 1/2$ and f^* is positive. The rate of convergence of an estimator of f is shown not to depend on d but only on the smoothness of f^*, and thus is the same for a long range and a short range dependent process. The log-periodogram estimator is shown to achieve the best possible rate of convergence when the smoothness of f^* is known, and to have adaptivity property when this smoothness is unknown.

1. Introduction

L'objet de cette note est de proposer une méthode d'estimation de la densité spectrale f d'un processus gaussien stationnaire $(X_t)_{t\in\mathbb{Z}}$, valable aussi bien en faible qu'enforte dépendance. La modélisation fractionnaire permet de prendre en compte ces deux aspects :

$$f(x) = |1 - e^{ix}|^{-2d} f^*(x), \tag{1}$$

où $d \in]-1/2, 1/2[$ et f^* est une fonction strictement positive et bornée sur $[-\pi, \pi]$. L'hypothèse de non annulation de f est naturelle et correspond à l'inversibilité du processus $(X_t)_{t\in\mathbb{Z}}$. Cette modélisation est dite fractionnaire puisqu'elle correspond au cas où il existe un processus stationnaire $(\epsilon_t)_{t\in\mathbb{Z}}$ de densité spectrale f^* tel que

$$X_t = (1 - B)^{-d} \epsilon_t = \sum_{j=0}^{\infty} \gamma_j \epsilon_{t-j},$$

où B est l'opérateur de retard et $\gamma_j = (-1)^j \Gamma(d + j)/(\Gamma(d) j!)$. Un modèle fractionnaire paramétrique courant est le modèle ARFIMA(p, d, q). Il a pour densité spectrale $f(x) = |1 - e^{ix}|^{-2d} |P(e^{ix})/Q(e^{ix})|^2$ où P et Q sont des polynômes sans racines communes et sans racines dans un disque $D(0, e^{\beta'})$ pour un réel $\beta' > 0$. $\log(f^*)$ admet alors un développement sur la base de cosinus $\log(f^*) = \sum_{j=0}^{\infty} \theta_j h_j$ avec $h_0(x) = 1/\sqrt{2\pi}$, $h_j(x) = \cos(jx)/\sqrt{\pi}$, $j \geq 1$ qui vérifie, pour tout $\beta < \beta'$,

$$\sum_{j=0}^{\infty} \theta_j^2 e^{2\beta j} < \infty. \tag{2}$$

Lorsque $d = 0$, on retrouve le cadre usuel de la faible dépendance, et lorsque $d > 0$, on obtient un processus dit fortement dépendant dont la fonction d'autocovariance n'est pas sommable. Dans le cas $d < 0$, la fonction d'autocovariance est sommable, mais certaines propriétés du processus sont semblables à celles des processus à longue portée. Dans le cas d'un modèle paramétrique, les estimateurs du maximum de vraisemblance ou du minimum de contraste de Whittle sont $\sqrt{n}$-consistants et asymptotiquement efficace au sens de Fisher. Cependant les modèles paramétriques sont très sensibles à la connaissance a priori de l'ordre du modèle, et ce plus encore dans le cas de la longue portée. Pour éviter ce type de problème, il peut être avantageux de considérer un modèle semi-paramétrique. La fonction f^* appartient à une classe fonctionnelle de dimension infinie. Une classe qui étend naturellement la classe des modèles ARFIMA est la classe définie à partir de l'équation (2), pour tous réels $\beta > 0$, $\delta \in [0, 1/2[$ et $L > 0$, par

$$\mathcal{A}(\beta, L, \delta) = \Big\{ f : f(x) = |1 - e^{ix}|^{-2d} \exp\{\sum_{j=0}^{\infty} \theta_j h_j(x)\},\ |d| \leq \delta,\ \sum_{j=0}^{\infty} \theta_j^2 e^{2\beta j} \leq L^2 \Big\}.$$

Par convention, $\delta = 0$ correspond au cas de la faible dépendance. D'une façon générale, la régularité de la fonction f^* a un rôle fondamental. Dans le cas $d = 0$, elle détermine la rapidité de la décroissance de la fonction d'autocovariance du processus. Dans le cas $d \neq 0$, la décroissance de la fonction d'autocovariance est déterminée par d, mais les propriétés d'estimation de f sont encore déterminées par f^*. En particulier, il a été montré par Giraitis, Robinson et Samarov (1997) et par Iouditsky, Moulines et Soulier (1999) que la vitesse de convergence d'un estimateur de d ne dépend que de la régularité de f^*. L'estimation de f doit donc tenir compte de cette régularité, et comme elle est en générale inconnue, doit avoir des propriétés d'adaptivité par rapport à cette régularité. Le problème de l'estimation adaptative de la densité spectrale a été étudié entre autres par Golubev (1993), Efromovich (1998) et Comte (1999), pour des processus gaussiens faiblement dépendant, et relativement au risque quadratique sur $L^2([-\pi, \pi], dx)$. Ce critère n'est pas adapté lorsque l'on envisage simultanément la faible et la forte dépendance puisque dans ce dernier cas, la densité spectrale n'est pas de carré intégrable si $d > 1/4$. Le critère de risque quadratique logarithmique est pertinent pour cette modélisation. Soit $K := K(n)$ une suite strictement croissante d'entiers et soit $y_k = (2k-1)\pi/(2K)$, $1 \leq k \leq K$. On d'efinit pour tout vecteur $u = (u_1, \cdots, u_K)^T$ de $\mathbb{R}^K$ la norme $\|u\|_n^2 = 2\pi K^{-1} \sum_{k=1}^K u_k^2$, et l'on identifie une fonction ϕ au vecteur $(\phi(y_1), \cdots, \phi(y_K))^T$. En appliquant l'inégalité multi-dimensionnelle de Van Trees (cf. Gill et Levit, 1995), on obtient la borne inférieure minimax suivante.

Théorème 1. *Soient $\beta > 0$, $L > 0$ et $\delta \in [0, 1/2[$.*

$$\liminf_n \inf_{\hat{l}_n} \sup_{f \in \mathcal{A}(\beta, L, \delta)} \frac{n}{\log(n)} \mathbb{E}_f [\|\hat{l}_n - l\|_n^2] \geq 2\pi/\beta,$$

où l'infimum porte sur tous les estimateurs $\hat{l}_n$ de $l = \log(f)$ construits sur une observation $X_1, \cdots, X_n$ d'un processus stationnaire $(X_t)_{t \in \mathbb{Z}}$ de densité spectrale f.

Pour montrer que cette borne inférieure, qui ne dépend pas de δ, est valable en faible et en forte dépendance, il est nécessaire de construire un estimateur de la densité spectrale qui réalise la vitesse minimax lorsque la régularité β est connue. L'estimateur de régression sur le log-périodogramme introduit dans la section 2 possède cette propriété. La régularité β n'étant pas connue en pratique, il est nécessaire d'étudier les propriétés d'adaptivité de l'estimateur par rapport à cette régularité inconnue.

2. Régression sur le log-périodogramme

Pour estimer le log-spectre, une technique naturelle est la régression sur le log-périodogramme. Le périodogramme est défini par

$$I_n(x) = \frac{1}{2\pi n}\Big|\sum_{k=1}^{n} h_k X_k e^{ikx}\Big|^2,$$

où (h_k), $1 \leq k \leq n$ est une suite de nombres complexes. Si l'on choisit $h_k = 1$, $1 \leq k \leq n$, on retrouve le périodogramme ordinaire. Soient $x_k = 2k\pi/n$, $1 \leq k < n/2$ les fréquences de Fourier. Il est bien connu que si le processus X_t est un bruit blanc gaussien, lorsque l'on a choisi $h_k = 1$, les variables $I_n(x_k)$, $1 \leq k \leq [n/2]$ sont i.i.d. de loi exponentielle $\Gamma(1,1)$. Dès que le processus (X_t) n'est plus Gaussien, l'indépendance n'est plus vérifiée. Dans le cas d'un processus faiblement dépendant, les variables $I_n(x_k)$ sont tout de même asymptotiquement de loi $\Gamma(1,1)$ et décorrélées. Dans le cas de la longue portée, il a été montré par Hurvich et Beltrao (1993) que ces propriétés asymptotiques ne sont plus vérifiées : les variables $I_n(x_k)$ ne sont asymptotiquement ni décorrélées ni de loi $\Gamma(1,1)$. Une étude précise de la structure de dépendance de la suite $(I_n(x_k))_{1\leq k<n/2}$ a été réalisée par Moulines et Soulier (1997). Pour obtenir des résultats théoriques dans le cadre de l'estimation adaptative, il est nécessaire d'utilser des coefficients h_k particuliers permettant d'obtenir une réduction de corrélation. On utilisera un noyau introduit par Hurvich et Chen (1998) et dont les propriétés de décorrélation ont été obtenues par Hurvich, Moulines et Soulier (1999) :

$$h_t = (1 - e^{2i\pi(t-1/2)/n})/2, \quad 1 \leq t \leq n. \tag{3}$$

Remarquons que pour ce noyau, même dans le cas d'un bruit blanc gaussien, l'indépendance de $I_n(x_k)$ et $I_n(x_l)$ n'est plus vérifiée pour deux fréquances de Fourier adjacentes. Pour retrouver cette indépendance, il est alors nécessaire de ne considérer qu'une fréquence de Fourier sur deux, ce qui entraîne une perte de variance d'un facteur 2. Cette perte de variance peut être partiellement compensée par agrégation. Le périodogramme agrégé est défini, pour un entier $m \geq 1$ (fixé) par

$$\bar{I}_{n,k} = \begin{cases} \sum_{t=m(k-1)+1}^{mk} I_n(x_t), 1 \leq k \leq [n/2m] - 1, & \text{pour le périodogramme ordinaire,} \\ \sum_{t=m(k-1)+1}^{mk} I_n^h(x_{2t}),\ 1 \leq k \leq [n/4m] - 1, & \text{pour le périodogramme fenétré.} \end{cases}$$

Pour unifier les notations, on notera $K = [(n-m)/2m]$ dans le cas du périodogramme ordinaire et $K = [(n-2m)/4m]$ dans le cas du périodogramme fenétré. Notons $\epsilon_{n,k} = \log(\bar{I}_{n,k}/f(y_k)) - \psi(m)$, où ψ est la fonction digamma (cf. Johnson et Kotz, 1970). Si le processus (X_t) est un bruit blanc gaussien (de densité spectrale $1/2\pi$) alors les variables $\epsilon_{n,k}$, $1 \leq k \leq K$ sont indépendantes, centrées et de variance $\psi'(m)$. Par exemple, pour $m = 1$, il est bien connu que $\psi(1) = -\gamma$ (la constante d'Euler) et $\psi'(1) = \pi^2/6$. La régression sur le log-périodogramme est basée sur l'ecriture tautologique suivante :

$$\log(\bar{I}_{n,k}) - \psi(m) = \log(f(y_k)) + \epsilon_{n,k}, \quad 1 \leq k \leq K,$$

On définit un estimateur $\hat{l}_{p,n}$ du log-spectre $l = \log(f)$ en estimant par moindres carrés d et les p premiers coefficients de Fourier de $\log(f^*) = \sum_{j=0}^{\infty} \theta_j h_j$. Posons $g(x) = -2\log(|1 - e^{ix}|)$.

$$(\hat{d}, \hat{\theta}_0, \cdots, \hat{\theta}_{p-1}) = \arg \min_{\bar{d}, \bar{\theta}_0, \bar{\theta}_{p-1}} \sum_{k=1}^{K} \Big(\log(\bar{I}_{n,k}) - \psi(m) - \bar{d}g(y_k) - \sum_{j=0}^{p-1} \bar{\theta}_j h_j(y_k)\Big)^2.$$

On définit alors $\hat{l}_{p,n} = \hat{d}g + \sum_{j=0}^{p} \hat{\theta}_j h_j$. Remarquons que cet estimateur est linéaire par rapport aux "observations" $\log(\bar{I}_{n,k})$ et admet une expression très simple (cf. Moulines et Soulier, 1998). Si l'on note

$\Pi_{p,n}$ le projecteur orthogonal sur le sous espace H_p de $\mathbb{R}^K$ engendré par les *vecteurs* $g, h_0, \cdots, h_{p-1}$, on peut écrire $\hat{l}_{p,n} = \Pi_{p,n} Y_n = \Pi_{p,n} l + \Pi_{p,n} \epsilon_n$. On a alors

$$\mathbb{E}_f [\|\hat{l}_{p,n} - l\|_n^2] = \|l - l_p\|_n^2 + \mathbb{E}_f [\|\Pi_{p,n}\epsilon_n\|_n^2],$$

où $l_p = \Pi_{p,n} l$ est la projection orthogonale de l sur H_p. La quantité $\|l - l_p\|_n^2$ est entièrement déterministe et calculable explicitement. C'est une version discrétisée du carré de la norme L^2 de la projection de l dans le sous espace de $L^2([-\pi, \pi])$ (dont la norme sera notée $\|.\|_2$) orthogonal à l'espace engendré par les *fonctions* $g, h_0, \cdots, h_{p-1}$:

$$\|l - l_p\|_2^2 = \sum_{j=p}^{\infty} \theta_j^2 - \frac{(\sum_{j=p}^{\infty} j^{-1}\theta_j)^2}{\sum_{j=p}^{\infty} j^{-2}} \leq \sum_{j=p}^{\infty} \theta_j^2.$$

La vitesse d'estimation dépend donc de la décroissance des coefficients θ_j. Les variables $\epsilon_{n,k}$ n'étant ni centrées ni décorrélées (même en dépendence faible), pour expliciter leur structure de dépendance, il est nécessaire d'introduire les classes fonctionnelles suivantes. Soient $M > 1$ et $\delta \in [0, 1/2[$.

$$\mathcal{G}(M) = \Big\{ u \in C([-\pi, \pi]),\ u \in D([-\pi, \pi] \setminus 0),\ \frac{\max_{x\in[-\pi,\pi]\setminus 0}\{|u(x)| + |xu'(x)|\}}{\min_{x\in[-\pi,\pi]} |u(x)|} \leq M \Big\},$$

$$\mathcal{L}(M, \delta) = \Big\{ f : f(x) = |1 - e^{ix}|^{-2d} f^*(x),\ f^* \in \mathcal{G}(M),\ |d| < \delta \Big\}.$$

On alors sur cette classe une expression asymptotique de la variance du terme stochastique $\|\Pi_{p,n}\epsilon_n\|_n^2$.

$$\sup_{1\leq p\leq K} \sup_{f\in\mathcal{L}(M,\delta)} \left| \frac{K}{p} \mathbb{E}_f [\|\Pi_{p,n}\epsilon_n\|_n^2] - 2\pi\psi'(m) \right| = o(1),$$

où le terme $o(1)$ est une suite tendant vers 0 et ne dépendant que de M et δ. L'équation de balance biais-variance et l'inclusion $\mathcal{A}(\beta, L, \delta) \subset \mathcal{L}(M, \delta)$ pour $M \geq (1 + \pi L)e^{2L}$ montrent que le choix de $p_n \approx \log(n)/2\beta$ donne un estimateur asymptotiquement minimax sur la classe $\mathcal{A}(L, \beta, \delta)$.

Théorème 2. *L'estimateur construit sur le périodogrammme ordinaire (*i.e. $h_k = 1$, $1 \leq k \leq n$*) agrégé est asymptotiquement minimax. Soit* $p_n(\beta) = \log(n)/2\beta$.

$$\limsup_{n\to\infty} \sup_{f\in\mathcal{A}(\beta,L,\delta)} \frac{n}{\log(n)} \mathbb{E}_f [\|\hat{l}_{p_n(\beta),n} - l\|_n^2] \leq 2\pi m\psi'(m)/\beta.$$

Puisque $m\psi'(m)$ tend vers 1 lorsque m tend vers l'infini, ce théorème montre aussi que la borne inférieure du Théorème 1 est exacte.

Corollaire 1. *Soient* $\beta > 0$, $L > 0$ *et* $\delta \in [0, 1/2[$.

$$\lim_{n\to\infty} \inf_{\hat{l}_n} \sup_{f\in\mathcal{A}(\beta,L,\delta)} \frac{n}{\log(n)} \mathbb{E}_f [\|\hat{l}_n - l\|_n^2] = 2\pi/\beta,$$

3. Estimation adaptative

L'estimateur proposé ci-dessus est irréalisable lorsque l'on ne connait pas le rayon d'analyticité β. Il est donc nécessaire de pouvoir faire un choix automatique du nombre de coefficients de Fourier θ_j à estimer. Puisque le problème considéré ici est un problème de régression (avec bruit non i.i.d. mais de variance connue), il est naturel d'utiliser l'approche de selection de modèle de Birgé et Massart (1997) qui consiste à déterminer le nombre $\hat{p}(\kappa)$ qui minimise le contraste pénalisé

$$S_{p,n}(\kappa) = \|Y_n - \hat{l}_{p,n}\|^2 + \kappa \frac{2\pi\psi'(m)p}{K},$$

où $Y_n = (\log(\bar{I}_{n,1}) - \psi(m), \cdots, \log(\bar{I}_{n,K}) - \psi(m))^T$ est le vecteur des observations. Pour $\kappa = 2$, on retrouve la statistique du C_L de Mallows. Moulines et Soulier (1998) ont montré (dans le cas du périodogramme ordinaire) que le choix de $\hat{p}(2)$ est asymptotiquement optimal , *i.e.*

$$\lim_{n\to\infty} \frac{R_n(\hat{p}_{CL})}{\inf_{1\leq p\leq K} R_n(p)} = 1,$$

la convergence ayant lieu en probabilité. Cette optimalité n'est vraie que sous la restriction assez génante que le nombre de coefficients θ_j non nuls est infini, et est asymptotique. Aucun contrôle du risque n'est obtenu pour un n donné. Birgé et Massart ont montré qu'en modifiant la pénalisation, on peut obtenir une borne pour le risque de l'estimateur pénalisé, valable pour tout n dans le cas du bruit i.i.d.

Théorème 3. *On considère l'estimateur de régression sur le log-périodogramme fenétré. Soit* $\kappa > 0$ *et soit* $\hat{p}(\kappa) = \arg\min_{1\leq p\leq K\log^{-3}(K)} S_{n,p}(\kappa)$. *Il existe un choix de* κ *tel que pour tout* $M > 0$ *et tout* $\delta \in [0, 1/2[$, *pour toute densité spectrale* $f \in \mathcal{L}(M, \delta)$, *on ait :*

$$\mathbb{E}_f[\|\hat{l}_{\hat{p}(\kappa),n} - l\|^2] \leq 4 \inf_{1\leq p\leq K/\log^3(n)} \{\|l - l_p\|^2 + \kappa 2\pi\psi'(m)p/K\} + C(M,\delta)/n.$$

Le résultat obtenu n'est valable que pour n au delà d'un certain rang qui dépend de la classe de densités spectrales considérée, mais a l'avantage sur la méthode du C_l d'être valable pour un modèle paramétrique : si le nombre de coefficients θ_j non nuls est fini. le risque de l'estimateur adaptatif est d'ordre n^{-1}. On en déduit un résultat d'adaptivité minimax par rapport à la borne d'analyticité β.

Corollaire 2. *Soient* $\beta^* > \beta_* > 0$ *et* $L^* > 0$.

$$\limsup_{n\to\infty} \sup_{\beta_*<\beta<\beta^*} \sup_{0<L<L^*} \sup_{f\in\mathcal{A}(\beta,L,\delta)} \frac{n}{\log(n)} \mathbb{E}_f[\|\hat{l}_{\hat{p}_\kappa,n} - l\|^2] \leq C(\beta_*\beta^*, L^*).$$

4. Concentration

La preuve du Théorème 3 repose sur une inégalité exponentielle. Pour $f \in \mathcal{L}(M,\delta)$, posons $\epsilon_{n,k} = \log(\bar{I}_{n,k}/f(y_k)) - \psi(m)$.

Proposition 1. *Il existe un réel* $\eta(M,\delta)$ *et un réel* $\lambda_0(M,\delta)$ *et une constante* C *tels que pour tout* $\lambda > \lambda_0(M,\delta)$, *pour tout* $u \in \mathbb{R}^K$ *tel que* $\sum_{k=1}^K u_k^2 = 1$ *et* $\max_{1\leq k\leq K} |u_k| \leq \eta(M,\delta)$,

$$\sup_{f\in\mathcal{L}(M,\delta)} \mathbb{P}_f\left(\sum_{k=1}^K u_k\epsilon_{n,k} > \lambda\right) \leq Ce^{-\lambda}.$$

Cette inégalité entraîne une inégalité de concentration analogue à la proposition 6.1 de Baraud *et al.* (1999). Soit L_q un sous espace de dimension $q \leq K$ de $\mathbb{R}^K$. Pour le réel $\eta(M,\delta)$ définit dans la proposition 1, on pose

$$\bar{Z}(q) = \sup\left\{\sum_{k=1}^K u_k\epsilon_{n,k}\ ; u \in L_q, \sum_{k=1}^K u_k^2 = 1, \max_{1\leq k\leq K} |u_k| \leq \eta(M,\delta)\right\}.$$

Proposition 2. *Il existe une constante universelle* C^* *et une constante* $C(M,\delta)$ *(ne dépendant que de* M *et* δ*) telle que pour tout* $1 \leq q \leq K$,

$$\sup_{f\in\mathcal{L}(M,\delta)} \mathbb{E}_f\left[\left\{\bar{Z}^2(q) - C_1^* q\right\}^+\right] \leq C(M,\delta)e^{-\sqrt{q}/2}.$$

References

[1] Y. Baraud, F. Comte et G. Viennet (1999) Model selection for (auto-)regression with dependent data. *Prépublications de L'Ecole Normale Supérieure* LMENS 99-12.

[2] L. Birgé et P. Massart (1997) From Model Selection to Adaptive Estimation In D. Pollard, I. Torgersen et G. Yang (ed.) Festschrift for Lucien Le Cam. 55-87 New-York : Springer-Verlag.

[3] S. Efromovich (1998) Data-Driven Efficient Estimation of the Spectral Density. *JASA* 92 (442) 762-769.

[4] F. Comte (1999) Adaptive estimation of the spectrum of a stationary Gaussian sequence. *Prépublications du laboratoire de probabilités et modèles aléatoires, Université Paris 6*

[5] R. Gill et B. Levit (1995) Applications of the van Trees inequality: A Bayesian Cramer-Rao bound. *Bernoulli* 1 (1-2) 59-79.

[6] G. K. Golubev (1993) Nonparametric Estimation of Smooth spectral Densities of Gaussian Stationary sequences. *Theory of probability and Applications* 38 (4) 630-639.

[7] C. Hurvich, E. Moulines et Ph. Soulier (1999) The FEXP estimator for non stationary processes. *preprint*

[8] L. Giraitis, P. Robinson et A. Samarov (1998) Adaptive rate-optimal semiparametric estimation of the long memory parameter, *Preprint*

[9] C. Hurvich and K. Beltrao (1993) Asymptotics of the low-frequency ordinates of the periodogram of a long-memory time series. *J. Time Series Anal.*, 14(5):455–472.

[10] A. Iouditsky, E. Moulines et Ph. Soulier (1999) Adaptive estimation of the Hurst coefficient. *to appear in Ann. Statist.*

[11] N.L. Johnson et S. Kotz (1970) Continuous univariate distributions I. New York, Wiley.

[12] E. Moulines et Ph. Soulier (1997) Broad band semi-parametric estimation of the long range coefficient. *to appear in Ann. Statist.*

[13] E. Moulines et Ph. Soulier (1998) Data-driven order selectioin for long range dependent time series. *to appear in J. Time Series Analysis*

Appendix A. Preuve du Théorème 3

Le schéma de la preuve du Théorème 3 est désormais bien établi. Nous suivons ici Baraud *et al.* (1999). Pour alléger les notations, toutes les dépendances en n sont sous-entendues, et nous notons $\hat{p} = \hat{p}_\kappa$, et $\mathrm{pen}(p) = 2\pi\kappa\psi'(m)p/K$. Rappelons encore que nous notons l_p la projection orthogonale de l (identifié à un vecteur de $\mathbb{R}^K$) sur le sous espace engendré par les vecteurs $g, h_0, \cdots, h_{p-1}$ et $\hat{l}_p$ est l'estimateur de l obtenu en estimant par moindres carrés les p premiers coefficients de Fourier de l^*. Par définition, on a, pour tout $1 \leq p \leq K$,

$$\|Y_n - \hat{l}_{\hat{p}}\|^2 + \mathrm{pen}(\hat{p}) \leq \|Y_n - \hat{l}_p\|^2 + \mathrm{pen}(p) \leq \|Y_n - l_p\|^2 + \mathrm{pen}(p).$$

Par ailleurs, $Y_n = l + \epsilon_n$, d'où

$$\begin{aligned}
\|l - \hat{l}_{\hat{p}}\|^2 &\leq \|l - l_p\|^2 + 2 < \epsilon_n, l_p - \hat{l}_{\hat{p}} > + \mathrm{pen}(p) - \mathrm{pen}(\hat{p}),\\
&= \|l - l_p\|^2 + 2 < \epsilon_n, l - l_{\hat{p}} > + 2 < \epsilon_n, l_{\hat{p}} - \hat{l}_{\hat{p}} > + 2 < \epsilon_n, l_p - l > + \mathrm{pen}(p) - \mathrm{pen}(\hat{p})\\
&\leq \|l - l_p\|^2 + 2\|l_{\hat{p}} - \hat{l}_{\hat{p}}\| < \hat{u}, \epsilon_n > + 2Z_2(\hat{p}) - 2Z_2(p) + \mathrm{pen}(p) - \mathrm{pen}(\hat{p}),
\end{aligned}$$

où l'on a posé $Z_2(p) =< \epsilon_n, l - l_p >$ et $\hat{u} = \|l_{\hat{p}} - \hat{l}_{\hat{p}}\|^{-1}(l_{\hat{p}} - \hat{l}_{\hat{p}})$. Soit $\alpha > 0$. On utilise maintenant l'argument $2ab \leq \alpha a^2 + \alpha^{-1}b^2$ pour majorer $2\|l_{\hat{p}} - \hat{l}_{\hat{p}}\| < \hat{u}, \epsilon_n >$.

$$\begin{aligned}
\|l - \hat{l}_{\hat{p}}\|^2 &\leq \|l - l_p\|^2 + \alpha\|l_{\hat{p}} - \hat{l}_{\hat{p}}\|^2 + \alpha^{-1} < \hat{u}, \epsilon_n >^2 + 2Z_2(\hat{p}) - 2Z_2(p) + \mathrm{pen}(p) - \mathrm{pen}(\hat{p})\\
&= \|l - l_p\|^2 + \alpha\|l - \hat{l}_{\hat{p}}\|^2 - \alpha\|l - l_{\hat{p}}\|^2 + \alpha^{-1} < \hat{u}, \epsilon_n >^2 + 2Z_2(\hat{p}) - 2Z_2(p) + \mathrm{pen}(p) - \mathrm{pen}(\hat{p}).
\end{aligned}$$

D'où finalement,

$$(1 - \alpha)\|l - \hat{l}_{\hat{p}}\|^2 \leq \|l - l_p\|^2 + \mathrm{pen}(p) + \alpha^{-1} < \hat{u}, \epsilon_n >^2 + 2Z_2(\hat{p}) - \alpha\|l - l_{\hat{p}}\|^2 - 2Z_2(p) - \mathrm{pen}(\hat{p}). \tag{4}$$

Dans le cas d'un bruit i.i.d. centré, il resterait à appliquer une inégalité de concentration pour majorer $\mathbb{E}[< \hat{u}, \epsilon_n >^2]$ et à remarquer que $\mathbb{E}[Z_2(p)] = 0$. Ces deux propriétés ne sont pas vérifiées ici et il faut introduire quelques modifications techniques par rapport au schéma usuel. Remarquons tout d'abord que l'on peut écrire

$$< \hat{u}, \epsilon_n >= \sum_{q=1}^{K \log^{-3}(K)} \sqrt{2\pi/K} < \hat{u}_q, \epsilon_n > \mathbf{1}_{\{\hat{p}=q\}},$$

où l'on a posé

$$\hat{u}_q = \frac{\sqrt{2\pi}}{\|l_q - \hat{l}_q\|\sqrt{K}}(l_q - \hat{l}_q).$$

Pour appliquer la proposition 2, il faut montrer que pour $1 \leq q \leq K \log^{-3}(K)$, et pour n assez grand, les coordonnés de $\hat{u}_q$ sont majorées par $\eta(M, \delta)$. Nous reprenons les notations de Moulines et Soulier (1998). Notons h_{-1} le vecteur unitaire othogonal á $h_0, \cdots, h_{q-1}$ et colinéaire à g. Les projections sur l'espace engendrés par les vecteurs $g, h_0, \cdots, h_{q-1}$ s'écrivent alors

$$l_q = \sum_{j=-1}^{q-1} < h_j, l > h_j, \quad \hat{l}_q = \sum_{j=-1}^{q-1} < h_j, Y_n > h_j,$$

et donc $\hat{u}_q = \sum_{j=-1}^{q-1} \beta_j h_j$, avec

$$\beta_j = \sqrt{2\pi/K}\Big(\sum_{j=-1}^{q-1} < h_j, \epsilon_n >^2\Big)^{-1/2} < \epsilon_n, h_j > .$$

On a clairement $\sum_{j=-1}^{q-1} \beta_j^2 = 1$ et donc $|\beta_j| \leq 1$ pour tout $j = -1, \cdots, q-1$. Par ailleurs, on voit facilement que $\max_{1 \leq k \leq K} |h_{-1}(y_k)| \leq C\sqrt{q}\log(K)$, d'où

$$\max_{1 \leq q \leq K \log^(-3)(K)} \max_{1 \leq k \leq K} |u_q(y_k)| = o(1).$$

Ce qui assure que pour n assez grand, et pour $q \leq K \log^{-3}(K)$, les coorconnées de $\hat{u}_q$ sont bien majorées par $\eta(M, \delta)$. On a alors

$$< \hat{u}, \epsilon_n >^2 \leq \frac{2\pi}{K} \bar{Z}^2(q).$$

En appliquant la Proposition 2, on a,

$$\sup_{f \in \mathcal{L}(M,\delta)} \mathbb{E}_f \left[\{ \frac{2\pi}{K} \bar{Z}^2(q) - \frac{2\pi C_1^* q}{K} \}^+ \right] \leq c(M, \delta) K^{-1} e^{-\sqrt{q}/2}.$$

En sommant pour q compris entre 1 et $K \log^{-3}(K)$, on obtient

$$\sup_{f \in \mathcal{L}(M,\delta)} \mathbb{E}_f \left[\{ < \hat{u}, \epsilon_n >^2 - 2\pi C_1^* \hat{p}/K \}^+ \right] \leq c(M, \delta) K^{-1}.$$

Pour majorer $\mathbb{E}[Z_2(\hat{p})]$ et $\mathbb{E}[Z_2(p)]$ (terme qui serait nul dans le cas d'un bruit centré), nous utilisons les lemmes suivants.

Lemme 1. *Soient $M > 1$ et $\delta \in [0, 1/2[$.*

$$\sup_{f \in \mathcal{L}(M,\delta} \mathbb{E}_f \left[\left\{ Z_2(\hat{p}) - (\alpha/2) \|l - l_{\hat{p}}\|^2 - \pi \hat{p}/(\alpha K) \right\}^+ \right] \leq c(M, \delta) K^{-1}.$$

Lemme 2. *Soient $M > 1$, $\delta \in [0, 1/2[$ et $\beta > 0$. Pour tout $1 \leq p \leq K$,*

$$\sup_{f \in \mathcal{L}(M,\delta} |\mathbb{E}_f [(Z_2(p))]| \leq \beta \|l - l_p\|^2 + c(M, \delta, \beta) K^{-1}.$$

On peut maintenant récrire (4) en tenant compte des majorations précédentes.

$$\begin{aligned}
(1-\alpha)\mathbb{E}[\|l - \hat{l}_{\hat{p}}\|^2] &\leq 2\|l - l_p\|^2 + \mathrm{pen}(p) \\
&+ \alpha^{-1}\mathbb{E}\big[< \hat{u}, \epsilon_n >^2 - 2\pi C_1^* \hat{p} K^{-1}\big] + \mathbb{E}\big[2Z_2(\hat{p}) - \alpha\|l - l_{\hat{p}}\|^2 - \pi\hat{p}/(\alpha K)\big] \\
&- 2\mathbb{E}[Z_2(p)] + \mathbb{E}\big[-\mathrm{pen}(\hat{p}) + \alpha^{-1}\pi(2C_1^* + 1)\hat{p}K^{-1}\big] \\
&\leq 2\|l - l_p\|^2 + \mathrm{pen}(p) + c(M, \delta)K^{-1} + \mathbb{E}\big[-\mathrm{pen}(\hat{p}) + \alpha^{-1}\pi(2C_1^* + 1)\hat{p}K^{-1}\big].
\end{aligned}$$

Il suffit de choisir $\kappa = (2C_1^* + 1)/(2\alpha\psi'(m))$, et l'on obtient

$$(1-\alpha)\mathbb{E}_f[\|l - \hat{l}_{\hat{p}}\|_n^2] \leq 2\|l - l_p\|_n^2 + \mathrm{pen}(p) + c(M, \delta)K^{-1}.$$

L'inégalité précédente étant vraie pour tout $p = 1, \cdots, K \log^{-3}(K)$, en prenant par exemple $\alpha = 1/2$, on obtient finalement,

$$\mathbb{E}_f[\|l - \hat{l}_{\hat{p}}\|^2] \leq 4 \inf_{1 \leq p \leq K \log^{-3}(K)} \left(\|l - l_p\|^2 + \kappa 2\pi\psi'(m) p K^{-1} \right) + c(M, \delta)K^{-1}.$$

Appendix B. Preuve des Propositions 1 et 2 et des Lemmes 1 et 2

Rappelons tout d'abord la décomposition du log-périodogramme obtenue par Moulines et Soulier (1997) dans le cas du périodogramme ordinaire et par Hurvich, Moulines et Soulier (1999) dans le cas du périodogramme fenétré.

Théorème 4. *Soit $(X_t)_{t\in\mathbb{Z}}$ un processus stationnaire gaussian de densité spectrale $f \in \mathcal{L}(M,\delta)$. On a alors la décomposition*

$$\epsilon_{n,k} = \eta_{n,k} + r_{n,k} \tag{5}$$

où les variables $\eta_{n,k}$ $1 \le k \le K$ sont centrées de variance $\psi'(m)$ et de même loi $\log(Y) - \psi(m)$ où Y suit la loi $\Gamma(m,1)$, et les bornes suivantes

$$\sup_{f\in\mathcal{L}} |\mathbb{E}_f\,(\eta_{n,k},\eta_{n,j})| \le c(M,\delta)p_\delta(k,j),\ \ 1 \le k < j \le K, \tag{6}$$

$$p_\delta(k,j) = \begin{cases} k^{-2\delta}j^{2\delta-2}\log^2(j) & \textit{pour le périodogramme ordinaire,} \\ k^{-1}(j-k)^{-2}(j/k)^{|\delta|} & \textit{pour le périodogramme fenétré,} \end{cases} \tag{7}$$

$$|r_{n,k}| \le c(M,\delta)\log(1+k)k^{-1},\ \ 1 \le k \le K. \tag{8}$$

Preuve de la Proposition 1. Soit $\lambda > 0$ et soit $u \in \mathbb{R}^K$ tel que $\sum_{k=1}^K u_k^2 = 1$. Soit $\rho > 0$. En utilisant la décomposition (5), on a

$$\mathbb{P}_f\Big(\sum_{k=1}^K u_k\epsilon_{n,k} > \lambda\Big) \le \mathbb{P}_f\Big(\sum_{k=1}^K u_k\eta_{n,k} > \lambda/(1+\rho)\Big) + \mathbb{P}_f\Big(\sum_{k=1}^K u_k r_{n,k} > \lambda\rho/(1+\rho)\Big),$$

où l^* est un entier dont le choix sera précisé plus loin. La borne (8) entraîne que pour $\lambda > \lambda_0(M,\delta,\rho)$, $\mathbb{P}_f(\sum_{k=1}^K u_k r_{n,k} > \lambda\rho/(1+\rho)) = 0$. Il reste donc à borner le second terme. En appliquant l'inégalité de Markov, on a

$$\mathbb{P}_f\Big(\sum_{k=1}^K u_k\eta_{n,k} > \lambda\rho/(1+\rho)\Big) \le e^{-\alpha\lambda\rho/(1+\rho)}\mathbb{E}_f\,[\exp\{\alpha\sum_{k=1} K u_k\eta_{n,k}\}].$$

La loi de $\eta_{n,k}$ étant donnée par le théorème 4, on voit facilement que

$$\log(\mathbb{E}_f\,[\exp\{t\eta_{n,k}\}]) = \log(\Gamma(m+t)) - \log(\Gamma(m)) - t\psi(m).$$

La fonction digamma ψ étant concave, on a

$$\log(\mathbb{E}_f\,[\exp\{t\eta_{n,k}\}]) \le \begin{cases} \frac{t^2}{2}\psi'(m) & \text{si } t \ge 0, \\ \frac{t^2}{2}\psi'(m+t) & \text{si } -m < t < 0. \end{cases}$$

Pour un ρ donné, on peut trouver un $\eta \le 1$ tel que si $|t| \le \eta$,

$$\mathbb{E}_f\,[\exp\{t\eta_{n,k}\}] \le e^{t^2\psi'(m)(1+\rho)/2}.$$

Posons $u_n^* = \max_{1\le k\le K}|u_k|$. Si $\alpha l^* u_n^* \le \eta$, on a alors

$$\mathbb{E}_f\,[\exp\{\alpha\sum_{k=1}^{l^*} u_k\eta_{n,k}\}] \le e^{\alpha^2\psi'(m)(1+\rho)/2}.$$

Pour étudier le second terme, il faut utiliser l'inégalité exponentielle suivante, prouvée dans Soulier (1998). On pose, pour $x \in \mathbb{R}^{2m}$, $\phi(x) = \log(|x|^2/2) - \psi(m)$

Proposition 3. *Soient $\xi_1, \cdots, \xi_n$ des vecteurs $2m$-dimensionnels conjointement gaussiens standard et soit Γ la matrice de covariance du vecteur $X = (\xi_1, \cdots, \xi_n)$. Soient $\beta_1, \cdots, \beta_n$ des nombres réels tels que t $\sum_{k=1}^n \beta^2 = 1$. Soient $0 < \rho < m$ et $0 < \epsilon < 1$. Il existe un réel $\eta > 0$ et une constante $C(\epsilon, \rho)$ tels que si $\mathrm{tr}\{(\Gamma - I_u)^2\} < 1 - \epsilon$ et $\max_{1 \le k \le n} |\beta_{n,k}| \ge \eta$, on a*

$$\mathbb{E}[e^{\alpha \sum_{k=1}^n \beta_k \phi(X_k)}] \le C(\epsilon, \rho) e^{\alpha^2 \psi'(m)(1+\rho)/2}. \tag{9}$$

Le théorème 4 entra–ˆine que l'on peut choisir l^* ne dépendant que de M et δ tel aue les hypothèses de la proposition 3 soient satisfaites. On a donc finalement, si $\alpha u_n^* \le \eta(M, \delta)$, en posant $S(u) = \sum_{k=1}^K u_k \eta_{n,k}$,

$$\mathbb{E}_f [e^{\alpha S(u)}] \le c(\rho) e^{\alpha^2 \psi'(m)(1+\rho)/2}. \tag{10}$$

Cette inégalité entraîne une inégalité exponentielle en $e^{-\lambda^2/2s^2}$ pour λ petit. Pour obtenir une inégalité valable pour tout λ, on utilise une technique déjà uttilisée par Laurent et Massart (1998). Posons $v = \psi'(m)(1 + \rho)$. Pour tout $\epsilon > 0$, on a

$$\mathbb{E}_f [e^{\alpha S(u)}] \le c(\rho) e^{v\alpha^2/(2-\epsilon\alpha)}.$$

En appliquant l'inégalité de Markov, on a

$$\mathbb{P}_f(S(u) > \lambda) \le c(\rho) \exp\{-\alpha\lambda + v\alpha^2/(2 - \epsilon\alpha\}.$$

Cette inégalité est optimisée par $\alpha = \lambda/(v + \epsilon\lambda) \le \epsilon$. Pour pouvoir appliquer (10), il faut donc avoir $u_n^* \le \epsilon\eta(M, \delta)$. On a alors, pour tout $\lambda > 0$,

$$\mathbb{P}_f(S(u) > \lambda) \le c(\rho) \exp\{-\lambda^2/(2v + \epsilon\lambda)\}.$$

Le choix de ϵ est arbitraire. On peut choisir par exemple $\epsilon = 1/2(1 + \rho)$. Pour $\lambda > 4v(1 + \rho)$, on a donc

$$\mathbb{P}_f(S(u) > \lambda) \le c(\rho) e^{-\lambda(1+\rho)}.$$

Par ajustement de la constante, l'inégalité est rendue valable pour tout $\lambda > 0$. On a donc finalement pour $\lambda > \lambda_0(M, \delta, \rho)$, et pour tout $u \in \mathbb{R}^K$ tel que $\sum_{k=1}^K u_k^2 = 1$ et $u_n^* \le \eta(M, \delta)$,

$$\mathbb{P}_f(\sum_{k=1}^K u_k \epsilon_{n,k} > \lambda) \le c(\rho) e^{-\lambda}.$$

Remarquons que ρ est arbitraire et qu'on peut par exemple choisir $\rho = 1$. La proposition 1 est prouvée.

Preuve de la Proposition 2. Posons, pour tout vecteur $u \in \mathbb{R}^K$, $Z(u) = \sum_{k=1}^K u_k \epsilon_{n,k}$. Soit $\eta := \eta(M, \delta)$ le réel défini dans la proposition 1. On a alors $\bar{Z}(q) = \sup_{u \in B_q(\eta)} Z(u)$, où $B_q(\eta)$ est l'ensemble des vecteurs u de L_q tels que $\sum_{k=1}^K u_k^2 \le 1$ et $\max_{1 \le k \le K} |u_k| \le \eta$. Soit $\delta_0 \in]0, 1]$. Il existe un recouvrement fini T_k de $B_q(\eta)$ par des boules de rayon $\delta_k = \delta_0 2^{-k}$ de cardinal au plus égal à $(3/\delta_k)^q$ (η ne joue ici aucun rôle). On pose $H_k = \log(|T_k|)$. Tout vecteur $u \in B_q(\eta)$ peut alors s'écrire

$$u = u_0 + \sum_{k \ge 1} (u_k - u_{k-1}),$$

où $u_k \in T_k$ et $\|u_k - u_{k-1}\| \le \delta_k + \delta_{k-1} = \frac{3}{2}\delta_{k-1}$. Soit $x = (x_k)_{n \in \mathbb{N}}$ une suite de réels positifs. Posons $v(x) = (x_0 + \sum_{k \ge 1} 3\delta_{k-1} x_k/2)$.

$$\begin{aligned} \mathbb{P}_f(\bar{Z}(q) > v(x)) &\le \mathbb{P}_f \left(\forall k \in \mathbb{N}, \exists u_k \in T_k, Z(u_0) + \sum_{k \ge 1} Z(u_k - u_{k-1}) > v(x) \right) \\ &\le \sum_{u \in T_0} \mathbb{P}_f(Z(u) > x_0) + \sum_{k \ge 1} \sum_{u \in T_k, v \in T_{k-1}, \|u-v\| \le 3\delta_k/2} \mathbb{P}_f(Z(u - v) > \frac{3}{2}\delta_{k-1} x_k). \end{aligned}$$

En appliquant la Proposition 1, on obtient, sous réserve du choix des x_k,

$$\forall u \in T_0,\ \mathbb{P}_f[Z(u) > x_0] \leq ce^{-x_0},$$
$$\forall u \in T_k, \forall v \in T_{k-1},\ \text{tels que } \|u - v\| \leq 3\delta_{k-1}/2,$$
$$\mathbb{P}_f[Z(u-v) > \frac{3}{2}\delta_{k-1}x_k] \leq ce^{-x_k}.$$

En sommant ces inégalités et en tenant compte du fait que $|T_k| = e^{H_k}$, on a

$$\mathbb{P}_f[\bar{Z}(q) > v(x)] \leq ce^{H_0 - Kx_0} + c\ \sum\nolimits_{k\geq 1} e^{H_k + H_{k-1} - Kx_k}.$$

On choisit maintenant $x_0 = H_0 + \sqrt{q+\xi}$ et $x_k = H_k + H_{k+1} + (k+1)\sqrt{q+\xi}$, et l'on obtient, pour $\sqrt{q+\xi} > \lambda_0$,

$$\mathbb{P}_f[\bar{Z}(q) > v(x)] \leq ce^{-\sqrt{q+\xi}}.$$

Il reste à évaluer $v(x)$. Du fait que $H_k \leq q\log(3/\delta_0) + kq\log(2)$, il vient $v(x) = (A(\delta_0)\sqrt{q} + B(\delta_0)\xi)$, où $A(\delta_0)$ et $B(\delta_0)$ sont des fonctions explicitables. On choisit δ^* qui minimise $A(\delta)$ et l'on pose $C_1^* = A(\delta^*)$ et $B^* = B(\delta^*)$. On obtient alors, si $\sqrt{q+\xi} > \lambda_0$ (λ_0 est défini dans la proposition 1),

$$\mathbb{P}_f\left(\bar{Z}^2(q) - C_1^* q > B^*\xi\right) \leq ce^{-\sqrt{q+\xi}}.$$

Par intégration, on obtient,

$$\begin{aligned}
\mathbb{E}_f\left[\left\{\bar{Z}^2(q) - C_1^* q\right\}^+\right] &= \int_0^\infty \mathbb{P}_f\left(\bar{Z}^2(q) - C_1^* q > v\right) dv \\
&= B^* \int_0^\infty \mathbb{P}_f\left(\bar{Z}^2(q) - C_1^* q > B(\delta^*)\xi\right) d\xi \\
&= B^* \int_0^{(\lambda_0 - q)^+} \mathbb{P}_f\left(\bar{Z}^2(q) - C_1^* q > B^*\xi\right) d\xi \\
&+ B^* \int_{(\lambda_0 - q)^+}^\infty \mathbb{P}_f\left(\bar{Z}^2(q) - C_1^* q > B^*\xi\right) d\xi \\
&\leq c(M,\delta)\mathbf{1}_{\{q \leq \lambda_0\}} + c(M,\delta)e^{-\sqrt{q}}\mathbf{1}_{\{q > \lambda_0\}}.
\end{aligned}$$

λ_0 ne dépendant que de M et δ, on obtient finalement

$$\mathbb{E}_f\left[\left\{\bar{Z}^2(q) - C_1^* q\right\}^+\right] \leq c(M,\delta)e^{-\sqrt{q}/2}.$$

Preuve du Lemme 1. La preuve du Lemme 1 est une adaptation de la preuve de la proposition 6.2 de Baraud *et al.* (1999). Soient $\epsilon > 0$ et $x > 0$ et soit q un entier.

$$\mathbb{P}_f\left(< l - l_q, \epsilon_n > > \epsilon\frac{\|l - l_q\|^2 + x^2}{2x}\right) \leq \mathbb{P}_f\left(< v_q, \epsilon_n > > \epsilon\right),$$

où l'on a posé $v_q = \|l - l_q\|^{-1}(l - l_q)$. Pour $q \leq K\log^{-3}(K)$, pour les mêmes raisons que dans la preuve de la proposition 2, on peut appliquer la proposition 1. On a donc

$$\mathbb{P}_f(< v_q, \epsilon_n > > \epsilon) \leq ce^{-\epsilon\sqrt{K/2\pi}}$$

En choisissant $x = \epsilon/\alpha$ et $\epsilon = \sqrt{2\pi(q+\xi)/K}$, on obtient, si $\sqrt{q+\xi} \geq \lambda_0(M,\delta)$,

$$\mathbb{P}_f\left(< l - l_q, \epsilon_n > -(\alpha/2)\|l - l_q\|^2 > \pi(q+\xi)/(\alpha K)\right) \leq ce^{-\sqrt{q+\xi}}.$$

En intégrant comme pour la preuve de la Proposition 2, il vient

$$\mathbb{E}\Big[\Big\{Z_2(q)-(\alpha/2)\|l-l_q\|^2-\pi q/(\alpha K)\Big\}^+\Big]\le c(M,\delta)K^{-1}e^{-\sqrt{q}/2}.$$

En sommant ces inégalités pour $q=1,\cdots,K\log^{-3}(K)$, on obtient

$$\mathbb{E}\Big[\Big\{Z_2(\hat p)-(\alpha/2)\|l-l_{\hat p}\|^2-\pi\hat p/(\alpha K)\Big\}^+\Big]\le c(M,\delta)K^{-1}.$$

Preuve du Lemme 2.

Appendix C. Preuve du Théorème 1

La preuve est donnée dans le cas $\delta>0$. La preuve dans le cas $\delta=0$, *i.e.* dans le cas de la faible dépendance, suit les mêmes lignes, et est simplifiée par l'absence des termes liés à la fonction g. Rappelons que l'on note $<.,.>$ le produit scalaire associé à la norme $\|.\|$. Soit alors $\tilde\gamma_p=\|g-\sum_{j=0}^{p-1}<\alpha_j,g>h_j\|^2$ et $h_{-1}=\tilde\gamma_p^{-1/2}(g-\sum_{j=0}^{p-1}<\alpha_j,g>h_j)$. Les vecteurs h_j, $-1\le j\le p-1$ forment une base orthonormée du sous-espace de $\mathbb{R}^K$ engendré par g, $h_0,\cdots,h_{p-1}$. Soit $(d_n)_{n\ge1}$ une suite de réels strictement positifs et soit $(p_n)_{n\ge1}$ une suite strictement croissante d'entiers vérifiant pour tout n : $p_n\le K$. On pose $\Theta=[-d_n,d_n]^{p_n}$ et on considère la famille paramétrique suivante

$$\mathcal{F}=\{f_\theta,\theta\in\Theta\},$$
$$\log(f_\theta)=l_\theta=\theta_{-1}\tilde\gamma_p^{-1/2}h_{-1}+\sum_{j=1}^{p_n-1}\theta_jh_j.$$

Pour cette famille paramétrique, du fait de l'orthonormalité des *vecteurs* h_j, $-1\le j\le p-1$, si $\hat\theta$ est un estimateur de θ basé sur une observation $X_1,\cdots,X_n$ d'un processus $(X_t)_{t\in\mathbb{Z}}$ de densité spectrale f_θ, on a

$$\|l_\theta-l_{\hat\theta}\|^2=\tilde\gamma_p(\hat\theta_{-1}-\theta_{-1})^2+\sum_{j=1}^{p-1}(\hat\theta_j-\theta_j)^2=(\hat\theta-\theta)^TB^{-1}(\hat\theta-\theta),$$

où $B=\mathrm{Diag}(\tilde\gamma_p^{-1},1,\cdots,1)$ est une matrice diagonale de taille p. Soit λ une densité de classe C^1 sur $[-1,1]$, telle que $\lambda(-1)=\lambda(1)=\lambda'(-1)=\lambda'(1)=0$ et d'information de Fisher finie : $I_\lambda=\int_{-1}^1(\lambda'(x))^2/\lambda(x)dx<\infty$. Soit $\lambda_n(x)=d_n^{-1}\lambda(d_n^{-1}x)$ et soit λ_n la densité d'efinie sur Θ par $\Lambda_n(x)=\prod_{i=1}^{p_n}\lambda(x_i)$. Par application de l'inégalité multidimensionnelle de Van-Trees (cf. Gill et Levit (1995) Théorème 1), on a

$$\begin{aligned}\inf_{\hat\theta}\sup_{\theta\in\Theta}\mathbb{E}_\theta[(\hat\theta-\theta)^TB^{-1}(\hat\theta-\theta)]\ge&\inf_{\hat\theta_n}\int_\Theta\mathbb{E}_\theta[(\hat\theta-\theta)^T(\hat\theta_n-\theta)]\Lambda_n(\theta)d\theta\\ \ge&p^2\left(\int_\Theta\mathrm{tr}\{BI_n(\theta)\}\Lambda_n(\theta)d\theta+(\tilde\gamma_p^{-1}+p-1)d_n^{-2}I_\lambda\right)^{-1}.\end{aligned}$$

où l'infimum est pris sur tous les estimateurs de θ basés sur une observation $X_1,\cdots,X_n$ d'un processus *gaussien* de densité spectrale f_θ et $I_n(\theta)$ est l'information de Fisher de l'observation. Il reste donc à calculer cette information.

Lemme 3. *Si $d_n \log(n) = o(1)$, $d_n p_n = o(1)$ et si il existe un $\epsilon > 0$ tel que $\sqrt{n} d_n^2 p_n^{-1/2+\epsilon} = o(1)$, alors uniformément par rapport à $\theta \in \Theta$*

$$I_n(\theta) = \frac{np_n(1+o(1))}{4\pi}.$$

Il reste à prouver que l'on peut trouver des suites d_n, p_n vérifiant les conditions du Lemme 3 et telles que la famille $\mathcal{F}$ est incluse dans la classe $\mathcal{A}(\beta, L, \delta)$, ce qui revient à vérifier $d_n^2 e^{2\beta p_n}$ est borné. Posons $p_n = \{\log(n) - 4\log\log(n)\}/2\beta$. Pour ce choix, on a $e^{2\beta p_n} = n/\log^2(n)$ et donc il suffit de choisir $d_n = C\log(n)/\sqrt{n}$ pour une constante bien choisie et la famille $\mathcal{F}$ est alors incluse dans la classe $\mathcal{A}(\beta, L, \delta)$. Les deux autres conditions du Lemme 3 sont aussi vérifiées. On peut maintenant conclure la preuve du Théorème 1.

$$\inf_{\hat{f}_n} \sup_{f \in \mathcal{A}(\beta,L,\delta)} \frac{n}{\log(n)} \mathbb{E}_f[\|\hat{l} - l\|^2] \geq \inf_{\hat{\theta}_n} \sup_{\theta \in \Theta} \frac{n}{\log(n)} \mathbb{E}_\theta[(\hat{\theta}_n - \theta)^T B(\hat{\theta}_n - \theta)] \geq 2\pi/\beta(1 + o(1)).$$

Preuve du Lemme 3. Soit $\Sigma(\theta)$ la matrice de covariance de l'observation gaussienne $X_1, \cdots, X_n$ de densité spectrale f_θ. On sait que

$$I_n(\theta) = \frac{1}{2}\left(\mathrm{tr}\left\{\left(\Sigma(\theta)^{-1}\partial_{\theta_i}\Sigma(\theta)\right)^2\right\}\right)_{-1 \leq i,j \leq p-1}.$$

Soit J_n la matrice identité n-dimensionnelle et soit $T_n(\phi)$ la matrice de Toeplitz d'ordre n de la fonction intégrable ϕ : $T_n(\phi)_{u,v} = \int_{-\pi}^{\pi} \phi(x) e^{i(u-v)x} dx$, $1 \leq u, v \leq n$. On alors

$$\Sigma(\theta) = T_n(f_\theta) = 2\pi J_n + T_n(e^{l_\theta} - 1).$$

Posons $e^{l_\theta} - 1 = l_\theta + g_\theta$. Sous réserve de prouver que le rayon spectral de la matrice $T_n(g_\theta)$ est strictement plus petit que 1, on a

$$\Sigma(\theta)^{-1}\partial_{\theta_i}\Sigma(\theta) = \left(2\pi J_n + T_n(e^{l_\theta} - 1)\right)^{-1}(\partial_{\theta_i} T_n(l_\theta) + \partial_{\theta_i} T_n(g_\theta)) = (2\pi)^{-1}\partial_{\theta_i} T_n(l_\theta) + R_{n,i}(\theta).$$

Pour $i = -1, 1, \cdots, p_n - 1$, on a donc

$$I_n(\theta)_{i,i} = (8\pi^2)^{-1}\mathrm{tr}\left\{(\partial_{\theta_i} T_n(l_\theta))^2\right\} + r_{n,i}(\theta).$$

Calculons tout d'abord les coefficients de $T_n(l_\theta)$ et le terme principal de $\mathrm{tr}\{BI_n(\theta)\}$.

$$T_n(l_\theta)_{u,v} = \sqrt{\pi}\theta_{|u-v|}\mathbf{1}_{\{0<|u-v|<p_n\}} + \sqrt{\pi}\theta_{-1}\tilde{\gamma}_{p_n}^{1/2}\alpha_{|u-v|}\mathbf{1}_{\{p_n \leq |u-v|\}} + \sqrt{\pi}\theta_{-1}\tilde{\gamma}_{p_n}^{1/2}\{\alpha_{|u-v|} - < g, h_{|u-v|} >\}\mathbf{1}_{\{0 \leq |u-v| < p_n\}},$$

$$\partial_{\theta_{-1}} T_n(l_\theta)_{u,v} = \sqrt{\pi}\tilde{\gamma}_{p_n}^{1/2}\alpha_{|u-v|}\mathbf{1}_{\{p_n \leq |u-v|\}}, + \sqrt{\pi}\tilde{\gamma}_{p_n}^{1/2}(\alpha_{|u-v|} - < g, h_{|u-v|} >)\mathbf{1}_{\{0 \leq |u-v| < p_n\}},$$

$$\partial_{\theta_i} T_n(l_\theta)_{u,v} = \sqrt{\pi}\mathbf{1}_{\{|u-v|=i\}},\ 1 \leq i \leq p_n - 1,$$

$$\mathrm{tr}\left\{(\partial_{\theta_{-1}} T_n(l_\theta))^2\right\} = 2\pi\tilde{\gamma}_{p_n}\sum_{j=p_n}^{n-1}(n-j)\alpha_j^2 + 2\pi\tilde{\gamma}_{p_n}\sum_{j=0}^{p_n-1}(\alpha_j - < g, h_j >)^2,$$

$$\mathrm{tr}\left\{(\partial_{\theta_i} T_n(l_\theta))^2\right\} = 2\pi(n-i),$$

$$\mathrm{tr}\{BI_n(\theta)\} = (4\pi)^{-1}\left(\sum_{j=p_n}^{n-1}(n-j)\alpha_j^2 + \sum_{j=0}^{p_n-1}(\alpha_j - < g, h_j >)^2 + \sum_{i=1}^{p_n-1}(n-i)\right) + r_n(\theta).$$

Remarquons que pour $j \leq K$, $\alpha_j - < g, h_j > = < g_K^*, h_j >$, où l'on définit our tout $p \geq 1$, $g_p^* = \sum_{j=p}^{\infty} \alpha_j h_j$. On a alors

$$\sum_{j=0}^{p_n-1} (\alpha_j - < g, h_j >)^2 = \sum_{j=0}^{p_n-1} < g_K^*, h_j >^2 \leq \sum_{j=0}^{K-1} < g_K^*, h_j >^2 = \|g_K^*\|^2.$$

La technique de sommation d'Abel permet de montrer que $\|g_K^*\|^2 = O(K^{-1})$. On a finalement

$$\mathrm{tr}\{BI_n(\theta)\} = (4\pi)^{-1} n p_n (1 + o(1)) + r_n.$$

Il reste à prouver que r_n est bien négligeable. Il faut tout d'abord évaluer le rayon spectral de la matrice $T_n(e^{l_\theta} - 1)$. En développant la fonction exponentielle à l'ordre $r \geq 3$, on obtient

$$T_n(e^{l_\theta} - 1) = \sum_{r=1}^{s-1} \frac{1}{r!} T_n(l_\theta^r) + \frac{1}{s!} T_n(h_\theta),$$

avec $|h_\theta| \leq |l_\theta|^s e^{|l_\theta|}$. Pour majorer le rayon spectral d'une matrice de Toeplitz d'ordre n d'une fonction, il suffit de majorer la somme des $n-1$ premiers coefficients de Fourier de la fonction considérée. Remarquons qu'on peut écrire

$$l_\theta = \theta_{-1} g_p^* + \sum_{j=1}^{p-1} (\theta_j - < h_j, g_K^* > \theta_1) h_j.$$

En notant $\beta_j = \theta_j - < h_j, g_K^* > \theta_{-1}$, on obtient, pour $j = 0, \cdots, n-1$

$$\sum_{j=0}^{n-1} |\hat{l}_\theta(j)| \leq C |\theta_{-1}| \sum_{j=p_n}^{n-1} \alpha_j + C \sum_{j=1}^{p-1} |\beta_j| = O(d_n \log(n) + d_n p_n).$$

Du fait que $\int_{-\pi}^{\pi} (g_p^*(x))^2 dx = O(p^{-1})$, en appliquant l'inégalité de Hölder on obtient que pour tout $p \geq 1$ et pour tout $\epsilon > 0$, $\int_{-\pi}^{\pi} (g_p^*(x))^{2s} dx = O(p^{-1+\epsilon})$. En appliquant l'inégalité de Parseval, on obtient , pour $r \geq 2$,

$$\sum_{j=0}^{n-1} |\widehat{l_\theta^r}(j)| \leq \sqrt{n} \left(\int_{-\pi}^{\pi} l_\theta^{2r}(x) dx \right)^{1/2} \leq C \sqrt{n} d_n^r \left(\int_{-\pi}^{\pi} g_p^{2r}(x) dx + O(p_n^{2r}) \right)^{1/2} = O(\sqrt{n} d_n^r p_n^{-1/2+\epsilon} + p_n^r d_n^r).$$

La fonction e^{l_θ} est intgrable et uniformément bornée dans tout $L^p([-\pi, \pi])$, donc la même technique donne pour h_θ,

$$\sum_{j=0}^{n-1} |\widehat{h_\theta}(j)| = O(\sqrt{n} d_n^r p_n^{-1/2+\epsilon} + p_n^r d_n^r).$$

Le rayon spectral ρ_n de la matrice $T_n(e^{l_\theta} - 1)$ est donc $o(1)$ sous les hypothèses du Lemme 3. On peut donc écrire $\Sigma(\theta)^{-1} = (2\pi)^{-1} + \Delta_n$, où Δ est une matrice de rayon spectral $O(\rho_n)$. En utilisant l'inégalité pour des matrices symmétriques $|\mathrm{tr}(A^2 B)| \leq \rho(B)(\mathrm{tr}(A^2)$, où $\rho(B)$ est le rayon spectral de la matrice B, on obtient, avec les notations

$$I_n(\theta)_{i,i} = (8\pi^2)^{-1} \mathrm{tr} \left\{ (\partial_{\theta_i} T_n(l_\theta))^2 \right\} (1 + O(\rho_n^*)).$$

Et l'on a finalement

$$\mathrm{tr}\{BI_n(\theta)\} = (4\pi)^{-1} n p_n (1 + o(1)).$$

Références additionnelles

B. Laurent et P. Massart (1998) Adaptive estimation of a quadrative functional by model selection. Prépublication 98-81de l'Université Paris-Sud.

Ph. Soulier (1998) Some new bounds and a central limit theorem for functions of Gaussian vectors. Prépublication 71 de l'Université d'Evry Val d'Essonne.