

UNIVERSITÉ PARIS EST MARNE-LA-VALLÉE

DOCUMENT DE SYNTHÈSE PRÉSENTÉ POUR L'OBTENTION DE
L'HABILITATION À DIRIGER LES RECHERCHES

Quelques contributions à la statistique des processus

Arnaud GLOTER

Soutenu le 27 Novembre 2008 devant le jury composé de :

Rapporteurs :	M. Yacine Aït-Sahalia	Université de Princeton
	Mme Fabienne Comte	Université Paris Descartes
	M. Yuri Golubev	Université de Provence
Examineurs :	Mme Valentine Genon-Catalot	Université Paris Descartes
	M. Jean Jacod	Université Paris 6
	M. Damien Lambertson	Université Paris Est Marne-la-Vallée
	Mme Dominique Picard	Université Paris 7

Remerciements

Je remercie vivement Yacine Aït-Sahalia, Fabienne Comte et Yuri Golubev d'avoir accepté d'être rapporteurs de cette habilitation à diriger les recherches. Je tiens également à remercier Valentine Genon-Catalot, Jean Jacod, Damien Lambertson et Dominique Picard qui me font l'honneur d'être membres du jury. Je leur exprime aussi ma gratitude pour l'intérêt qu'ils ont toujours porté à mon travail, ainsi que pour leur soutien bienveillant.

De nombreux travaux présentés dans ce mémoire ont été effectués en collaboration avec Marc Hoffmann. Je le remercie pour toutes nos passionnantes séances de travail, discussions et tous les conseils amicaux qu'il a pu me donner. J'ai aussi eu le plaisir de travailler avec Emmanuel Bacry, Emmanuel Gobet, Jean François Muzy et Michael Sørensen que je remercie pour les enrichissantes discussions que nous avons eues.

Je souhaiterais exprimer ma gratitude aux membres des équipes de recherche qui m'ont chaleureusement accueilli depuis la fin de ma thèse (d'abord à l'université Bordeaux IV, puis à celle de Marne-la-Vallée). Grâce à eux, j'ai pu bénéficier de conditions de travail très agréables durant toutes ces années.

Je pense tout particulièrement au groupe des probabilistes et statisticiens de Marne, ainsi que, parmi les 'bordelais', à Steve Bazen, Sandrine Robert et Guillaume Carlier.

Enfin, je remercie tous mes amis et ma famille.

Table des matières

1	Introduction	3
1.1	La problématique générale	3
1.2	Organisation du mémoire	4
1.3	Liste des travaux	7
2	Estimateurs explicites en statistique des diffusions. Le cas de l'observation de l'intégrale d'une diffusion	9
2.1	Observation discrétisée avec pas fixe pour un processus d'Ornstein-Uhlenbeck	9
2.2	Discrétisation d'une diffusion intégrée et comparaison avec le schéma d'Euler	11
2.3	Estimation jointe des coefficients de dérives et diffusions sous condition d'ergodicité	12
2.4	Estimation de paramètres sous l'hypothèse d'un faible coefficient de diffusion	14
3	Bruit de microstructure	17
3.1	Estimation d'une diffusion gaussienne bruitée : la propriété LAN	17
3.2	Estimateurs efficaces pour une diffusion bruitée	20
3.3	Étude d'un modèle à volatilité stochastique avec bruit de microstructure	22
4	Maximum de vraisemblance par calcul de Malliavin dans des modèles cachés	27
5	Vitesse d'estimation dans des modèles cachés fractionnaires	31
5.1	Le cas d'une volatilité brownienne	31
5.2	Estimation de paramètres dans un modèle à volatilité stochastique fractionnaire	32
5.3	Estimation de l'autosimilarité d'une trajectoire cachée	34
6	Étude statistique de signaux multifractals	39
6.1	Le problème de l'estimation non paramétrique revisité	40
6.2	Modèle de cascades multiplicatives en asymptotique mixte	48

Chapitre 1

Introduction

Le but de ce mémoire est de présenter les travaux que j'ai réalisés en statistique des processus depuis le début de ma thèse.

1.1 La problématique générale

De manière générale, nous nous intéressons à l'estimation d'un paramètre de dimension fini, présent dans un modèle de processus stochastique. Une des originalités de nos travaux est que nous n'observons pas directement le processus dont nous voulons estimer les paramètres. Plusieurs situations pour l'observation reviennent fréquemment :

1. L'observation des valeurs moyennes du processus.
2. L'observation bruitée du processus.

Les motivations pour considérer ces problèmes sont issues des applications telles que la finance, l'économétrie ou des modèles physiques. Du point de vue statistique, elles conduisent à une structure compliquée des observations. En particulier celles ci ne sont jamais markoviennes.

La majeure partie de nos travaux a été réalisée sous l'hypothèse d'observations hautes fréquences. Cette hypothèse permet d'utiliser les propriétés trajectorielles des processus pour construire des estimateurs. De manière générale, nous nous posons aussi la question de l'optimalité des estimateurs que nous obtenons. Même dans le cas où le processus est une diffusion et l'observation est directe, l'étude de l'optimalité est un problème délicat. Dans la situation où le processus est partiellement caché, cette étude est, a priori, encore plus difficile. Cependant, dans certaines situations, nous pouvons calculer l'informations de Fisher du modèle. Dans d'autres cas, nous nous contentons de caractériser la vitesse optimale d'estimation. Ces méthodes sont alors proches de celles employées en statistique non paramétrique (comparaison d'expériences statistiques, propriétés de régularité des signaux, choix adaptatif de fréquences optimales).

Au cours de ces travaux nous nous sommes intéressés à trois grandes classes de processus stochastiques :

1. Les processus de diffusions.
2. Les processus du type mouvements browniens fractionnaires.
3. Les processus à trajectoires multifractales.

Les outils probabilistes utilisés dans ces trois cas sont variés (calcul stochastique, calcul de Malliavin, représentation spectrale, décomposition trajectorielle des processus sur une base). Cependant les résultats que nous obtenons ont une certaine unité. En particulier, les résultats obtenus pour des processus fractionnaires permettent de mieux comprendre ceux sur les diffusions.

Soulignons qu'un travail présenté ([12]) ne concerne pas la statistique des processus. Ce travail est totalement dans le champ de la statistique non paramétrique par méthode de seuillage. Cependant, les outils de probabilité et d'analyse utilisés sont proches de ceux du reste du mémoire.

1.2 Organisation du mémoire

Les travaux présentés sont regroupés en 5 chapitres. Chacune des sections de ces chapitres correspond à un travail publié (numéroté de [1] à [13]). Les chapitres peuvent être lus de manière indépendante. Signalons que pour ne pas alourdir certains énoncés de théorème, nous ne précisons pas toujours toutes les hypothèses nécessaires. En revanche, nous avons essayé d'expliquer les idées des preuves quand celles-ci étaient très techniques dans les papiers publiés.

Le chapitre 2 présente des résultats obtenus pendant ma thèse sur l'estimation des paramètres d'une diffusion X , lorsque l'on observe n moyennes locales sur des intervalles de longueur Δ . L'objectif était de savoir si l'on pouvait obtenir des résultats semblables au cas de l'observation discrétisée ($X_{i\Delta}$). Comme dans le cas d'une observation discrète, les résultats sont très différents suivant la nature du pas Δ , nous avons considéré successivement les situations : Δ fixe, $\Delta = \Delta_n = 1/n$, puis $\Delta_n \rightarrow 0$ avec $n\Delta_n \rightarrow \infty$ ([1]–[3]). Une des difficultés provient du fait que pour l'observation discrète la procédure d'estimation est basé sur le schéma d'Euler. Un tel schéma n'existe pas pour les moyennes de diffusion. Le travail [2] résout le problème de trouver un analogue au schéma d'Euler dans notre cas. Nous présentons aussi dans ce chapitre un travail réalisé plus tard avec Michael Sørensen dans le cadre du réseau Européen DYNSTOCH. Ce travail [4] est assez différent du reste du chapitre. Néanmoins les méthodes utilisées, de type 'contrastes explicites pour les diffusions', nous ont suggéré de le placer dans ce chapitre.

Le chapitre 3 présente des résultats obtenus sur le bruit de microstructure pour des modèles de diffusions. On observe maintenant la diffusion X aux instants i/n et chaque observation est entachée d'une erreur. Une de nos motivations initiales était de comparer avec les travaux de Delattre et Jacod [DJ97] qui considéraient le problème de l'observation avec erreur d'arrondi. Dans les travaux [5]–[6], réalisés avec Jean Jacod en 2001, nous caractérisons la vitesse d'estimation d'un paramètre de diffusion. Notons que le calcul de l'information de Fisher n'est fait que dans le cas plus simple d'une diffusion gaussienne.

Plus tard, le problème de l'estimation non paramétrique de la volatilité, dans une situation proche, a été étudié par Zhang *et al.* ([ZMAS05]). Le dernier travail présenté [7] dans ce chapitre concerne l'estimation de paramètres dans des modèles à volatilité stochastique. Il s'appuie, en parti, sur des résultats anciens de ma thèse, et sur ces méthodes introduites plus récemment dans le champ de l'économétrie pour estimer la volatilité.

Le chapitre 4 contient un travail [8] élaboré courant 2004-2005 avec Emmanuel Gobet. Il s'agissait de voir comment les méthodes introduites par E. Gobet ([Gob01]-[Gob02]) pour étudier la vraisemblance d'une diffusion discrétisée pouvaient s'appliquer dans nos modèles cachés. Nous considérons en détail le cas de l'observation des moyennes locales de la diffusion avec un pas $\Delta_n = 1/n$. Ces résultats précisent donc ceux qui avaient été obtenus pendant ma thèse en calculant l'information de Fisher exacte du modèle. Cependant ce travail est un peu à part, à cause des outils probabilistes utilisés qui sont issus du calcul de Malliavin. Notons aussi que toute une partie des idées de ce papier peuvent s'appliquer pour d'autres types d'observations.

Le chapitre 5 présente deux travaux [10]-[11] élaborés de 2002 à 2004 avec Marc Hoffmann. Il s'agissait de comprendre comment la vitesse d'estimation était modifiée lorsque l'on remplace la diffusion cachée par un processus de type mouvement brownien fractionnaire. Une motivation provient du travail, fait pendant ma thèse, sur l'estimation du paramètre de diffusion dans un modèle à volatilité stochastique. Nous obtenions un estimateur à vitesse $n^{1/4}$ (voir [9]). Ce résultat fut ensuite complété par Marc Hoffmann dans [Hof02], qui montre que cette vitesse est optimale. L'idée est que la régularité trajectorielle du processus semble jouer un rôle dans la vitesse d'estimation quand ce processus est caché. En remplaçant la diffusion cachée par un processus de type mouvement brownien fractionnaire on peut calibrer la régularité du processus par l'indice de Hurst H . Dans [10] nous considérons un modèle à volatilité stochastique où la volatilité est une 'diffusion' dirigée par un brownien fractionnaire : $dV_t = a(V_t, \theta)dW_t^H + \text{'dérive'}$. Nous cherchons à estimer ce paramètre de 'diffusion' θ . La vitesse obtenue est $n^{1/(2+4H)}$. Ceci confirme bien que la régularité de la volatilité joue un rôle dans la précision d'estimation. Remarquons que plus la volatilité est irrégulière, plus le paramètre est facile à estimer. Ceci correspond à l'idée que plus les variations de la volatilité sont importantes, plus elles sont faciles à détecter depuis l'observation du prix de l'actif.

Dans [11], nous complétons cette étude en cherchant à estimer le paramètre de Hurst. Le processus est $X = \theta W_t^H$ et l'observation est discrète et bruitée. On obtient alors encore la même vitesse $n^{1/(2+4H)}$. Cependant le paramètre de Hurst et le paramètre de 'diffusion' θ jouent des rôles très différents. Les travaux [10] et [11] reposent donc sur des idées assez différentes. Précisons que le problème de l'estimation de H dans un modèle à volatilité stochastique a ensuite été traité dans la thèse de Mathieu Rosenbaum ([Ros07]).

Le chapitre 6 présente deux travaux autour des signaux multifractals. C'est en assistant au cours 'Introduction to multifractal analysis' donné par Stéphane Jaffard dans le cadre des 'rencontres mathématiques' du CIRM en mars 2004, que j'ai découvert ce thème. Un peu plus tard, à la lecture du papier de S. Jaffard 'On the

Frisch-Parisi conjecture' [Jaf00], il semble que des problèmes d'estimation non paramétrique apparaissent naturellement dans le cadre de l'analyse multifractale. Le premier travail présenté [12], à été élaboré avec Marc Hoffmann en 2006-2007. Il reprend le problème, classique en statistique non paramétrique, d'estimer une fonction déterministe f dans un modèle de bruit blanc. On sait que la vitesse d'estimation de f dépend de sa régularité, et nous faisons l'hypothèse d'une régularité multifractale pour cette fonction. Nous résolvons alors le problème de la vitesse minimax dans ce problème (ainsi que celui de l'estimation adaptative). Nous montrons que, pour une perte \mathbf{L}^π , l'exposant de la vitesse d'estimation est donné, en fonction du spectre de singularité d de la fonction, par la quantité :

$$\min_H \frac{H + (d - d(H))/\pi}{2H + d}.$$

Un point important est que le résultat classique sur les espaces de Besov (avec 'coude' dans la vitesse [DJ98]) se retrouve comme cas particulier de la formule générale que nous obtenons.

Le dernier travail [13] présenté à été réalisé en 2007-2008, avec Emmanuel Bacry, Marc Hoffmann et Jean François Muzy. Les motivations pour ce travail proviennent de constatations empiriques faites sur des données de fluides turbulents. Ces constatations suggèrent d'étudier le comportement asymptotique de la 'fonction de partition' d'un processus de cascade multiplicative. Nous nous plaçons dans le cadre d'une asymptotique mixte, avec données hautes fréquences et temps total d'observation tendant vers l'infini. Notre travail étend alors les résultats de Resnick *et al* 2003 [RSGW03] qui considéraient le cas d'une observation sur $[0, 1]$.

1.3 Liste des travaux

Articles présentés

- [1] A. Gloter. Parameter estimation for a discrete sampling of an integrated Ornstein-Uhlenbeck process. *Statistics*, 35 :225–243, 2001.
- [2] A. Gloter. Discrete sampling of an integrated diffusion process and parameter estimation of the diffusion coefficient. *ESAIM : Prob. & Stat.*, 4 :205–227, 2000.
- [3] A. Gloter. Parameter Estimation for a discretely observed integrated diffusion process. *Scand. J. Statist.*, 33 :83–104, 2006.
- [4] A. Gloter et M. Sørensen. Estimation for stochastic differential equations with a small diffusion coefficient. *Stochastic. Process. Appl.*, 2008.
- [5] A. Gloter et J. Jacod. Diffusion with measurement errors. I. Local asymptotic normality. *ESAIM : Prob. & Stat.*, 5 :225–242, 2001.
- [6] A. Gloter et J. Jacod. Diffusion with measurement errors. II. Optimal estimator. *ESAIM : Prob. & Stat.*, 5 :243–260, 2001.
- [7] A. Gloter. Efficient estimation of drift parameters in stochastic volatility models. *Finance Stoch.*, 11 :495–519, 2007.
- [8] A. Gloter et E. Gobet. LAMN property for hidden processes : The case of integrated diffusions. *Ann. Inst. H. Poincaré Probab. Statist.*, 44 :104–128, 2008.
- [9] A. Gloter. Estimation du coefficient de diffusion de la volatilité d’un modèle à volatilité stochastique. *C. R. Acad. Sci., Série I*, 330 :243–248, 2000
- [10] A. Gloter et M. Hoffmann Stochastic volatility and fractional Brownian motion. *Stochastic. Process. Appl.*, 113 :143–172, 2004
- [11] A. Gloter et M. Hoffmann Estimation of the Hurst parameter from discrete noisy data. *Ann. Statist.*, 35 :1947–1974, 2007
- [11’] A. Gloter et M. Hoffmann Estimation of the Hurst parameter from discrete noisy data. (version détaillée de [11]) *Document de travail*. 34p., 2007
- [12] A. Gloter et M. Hoffmann Nonparametric reconstruction of a multifractal function from noisy data. *à paraître dans Probab. Theory Related Fields*, 38p., 2008
- [13] E. Bacry, A. Gloter, M. Hoffmann et J.F. Muzy. Multifractal analysis in a mixed asymptotic framework. *Soumis à Annals of applied probability*, 33p., 2008

Remarque : La note au CRAS [9] ci dessus n’est pas doublée d’une publication.

Thèse

- [T] A. Gloter. Estimation des paramètres d'une diffusion cachée : intégrales de processus de diffusions et modèles à volatilité stochastique. *Thèse de l'université de Marne-la-Vallée, rédigée sous la direction de V. Genon-Catalot*, Janvier 2000.

Actes de congrès

- [13] E. Bacry, A. Gloter, M.Hoffmann et J.F. Muzy. Long time behavior for the partition function of multiplicative cascades. *Proceedings of IWAP08 (International Workshop on Applied Probability, Compiègne, France, July 2008)*

Chapitre 2

Estimateurs explicites en statistique des diffusions. Le cas de l'observation de l'intégrale d'une diffusion

Les trois premiers travaux présentés correspondent à la première partie de ma thèse [T]. On étudie le problème de l'estimation des paramètres d'une diffusion dont on observe les moyennes plutôt que les valeurs exactes en certains instants. L'objectif est d'obtenir des estimateurs explicites et d'étudier leur comportement asymptotique. Les différents travaux [1]–[3] correspondent à différents choix pour l'asymptotique du modèle.

Le dernier travail présenté dans cette partie a été réalisé avec Michael Sørensen et concerne l'estimation d'une diffusion sous l'hypothèse d'un coefficient de diffusion asymptotiquement petit. L'objectif de ce travail était en fait de montrer que l'on peut trouver une grande classe d'estimateurs explicites qui sont asymptotiquement gaussiens, en se passant des conditions sur le pas de discrétisation introduites dans le même cadre par Sørensen & Uchida (2003) [SU03].

2.1 Observation discrétisée avec pas fixe pour un processus d'Ornstein-Uhlenbeck

L'intérêt de cet exemple particulier est que l'on peut mener les calculs de façon explicite. Le processus étudié est solution stationnaire de l'équation $dX_t = \mu X_t dt + \sigma dB_t$, et on observe $\bar{X}_i = \frac{1}{\Delta} \int_{i\Delta}^{(i+1)\Delta} X_s ds$ pour $i = 0, \dots, n-1$ et Δ un pas de discrétisation fixe. Dans [1] nous obtenons que le processus des intégrales $(\bar{X}_i)_i$ est un processus gaussien stationnaire ARMA (1,1). Cette structure ARMA (1,1) apparaît

sur la relation suivante que nous obtenons entre deux observations consécutives :

$$\begin{aligned} \bar{X}_{i+1} - e^{\mu\Delta}\bar{X}_i &= \frac{\sigma}{\mu} \int_{i\Delta}^{(i+1)\Delta} \frac{(e^{\mu\Delta} - e^{\mu((i+1)\Delta-s)})}{\Delta} dB_s \\ &\quad + \frac{\sigma}{\mu} \int_{(i+1)\Delta}^{(i+2)\Delta} \frac{(e^{\mu(i+2)\Delta-s} - 1)}{\Delta} dB_s. \end{aligned} \quad (2.1)$$

Nous vérifions que la densité spectrale du processus satisfait les conditions suffisantes pour que l'estimateur de Whittle noté $\hat{\theta}_n = (\hat{\mu}_n, \hat{\sigma}_n^2)$ soit consistant et asymptotiquement normal. On sait qu'il est alors équivalent au maximum de vraisemblance exact (voir Dzhaparidze & Yaglom (1983) [DY83], Dacunha-Castelle & Duflo (1986) [DCD86]). Nous obtenons une expression explicite pour le contraste de Whittle. Si μ est connu, l'estimateur $\hat{\sigma}_n^2$ est lui aussi explicite et sa variance asymptotique est $2\sigma^4$, qui est la même variance asymptotique que celle du maximum de vraisemblance de σ^2 basé sur l'observation des valeurs discrètes $(X_{i\Delta})_{i=0,\dots,n-1}$.

Nous considérons ensuite un estimateur moins spécifique au caractère gaussien des observations et donc a priori plus adapté à une généralisation à d'autres diffusions. Nous utilisons pour cela la méthode introduite par Ryden (1994) [Ryd94] dans le cadre des chaînes de Markov cachées. Nous nous donnons un entier m supérieur ou égal à 1 et découpons les observations $(\bar{X}_0, \dots, \bar{X}_{n-1})$ en paquets de taille m , $(\bar{X}_{km}, \dots, \bar{X}_{km+m-1})$ pour $k = 0, \dots, \lfloor n/m \rfloor - 1$. L'idée est alors d'utiliser comme fonction de contraste la log-vraisemblance qu'auraient les observations si ces paquets étaient indépendants. Le contraste obtenu est alors la somme des log-vraisemblances de chacun des paquets. On note $\tilde{\theta}_n^{(m)} = (\tilde{\mu}_n^{(m)}, \tilde{\sigma}_n^{2(m)})$ l'estimateur résultant de la maximisation de ce contraste. En remarquant que les observations forment une suite α -mélangeante, nous pouvons démontrer la convergence et normalité asymptotique pour $\tilde{\theta}_n^{(m)}$:

$$\sqrt{n/m}(\tilde{\theta}_n^{(m)} - \theta_0) \xrightarrow{n \rightarrow \infty} \mathcal{N}\left(0, I^{(m)}(\theta_0)^{-1} + 2I^{(m)}(\theta_0)^{-1}H^{(m)}(\theta_0)I^{(m)}(\theta_0)^{-1}\right)$$

où $I^{(m)}(\theta_0)$ est la matrice d'information de Fisher correspondant à l'observation d'un bloc de données $(\bar{X}_0, \dots, \bar{X}_{m-1})$ et $H^{(m)}(\theta_0)$ est une matrice explicite qui provient de la dépendance des vraies données.

En évaluant numériquement cette variance asymptotique théorique, il apparaît que, plus m est grand, plus l'estimateur de Ryden s'approche de l'efficacité (ce qui n'est pas étonnant). Il est, au contraire, plus inattendu que quand Δ tend vers 0, la variance asymptotique de $\tilde{\sigma}_n^{2(m)}$ augmente fortement, même si la vraie valeur de μ est connue. L'estimation de σ^2 par cette méthode est donc moins bonne que par $\hat{\sigma}_n^2$ dont la variance asymptotique est indépendante de Δ .

Pour éclairer ce résultat numérique, nous calculons théoriquement un équivalent quand Δ tend vers 0 pour la variance asymptotique de $\tilde{\sigma}_n^{2(m)}$ quand $m = 2$ et μ est connu :

$$\text{var } \tilde{\sigma}_n^{2(m)} \sim_{\Delta \rightarrow 0} \frac{\sigma^2}{4|\mu|\Delta}.$$

Bien que restreint au cas très particulier d'une diffusion de type Ornstein-Uhlenbeck ce travail nous a éclairé sur la possibilité d'estimer efficacement les paramètres d'une diffusion lorsque l'on observe ses moyennes. En particulier le paramètre σ^2 s'estime aussi bien que depuis l'observation discrète des $X_{i\Delta}$. On retrouvera dans le chapitre 4 de ce mémoire, un résultat proche obtenu dans un cadre beaucoup plus général. De plus, ce travail nous a permis de voir que certaines méthodes inspirées des chaînes de Markov cachées à états finis ne sont pas très bonnes quand le pas de discrétisation est petit. Ceci nous a incité, dans les travaux suivants [2]–[3], à utiliser des méthodes plus spécifiques aux propriétés trajectorielles des diffusions dans le cas d'une discrétisation avec pas petit.

2.2 Discrétisation d'une diffusion intégrée et comparaison avec le schéma d'Euler

Dans [2], nous considérons une diffusion unidimensionnelle à valeurs dans un intervalle (l, r) ,

$$dX_t = a(X_t)dB_t + b(X_t)dt,$$

où a et b sont des coefficients a priori quelconques. Pour Δ_n une suite de pas convergent vers 0, on considère les moyennes locales de la diffusion, construites avec ce pas Δ_n :

$$\bar{X}_{i,n} = \frac{1}{\Delta_n} \int_{i\Delta_n}^{(i+1)\Delta_n} X_s ds \quad \text{pour } i = 0, \dots, n-1$$

Notre objectif est d'obtenir des développements du type schéma d'Euler pour le processus $(\bar{X}_{i,n})_{i=0,\dots,n-1}$ et de les comparer au schéma d'Euler classique basé sur les $(X_{i\Delta_n})$. La motivation principale est que l'on sait que l'on peut obtenir des estimateurs explicites basés sur l'observation des $(X_{i\Delta_n})$ en utilisant le schéma d'Euler classique (voir Kessler 1997 [Kes97]).

Une difficulté provient du fait que nous considérons une diffusion à valeurs dans (l, r) avec $-\infty \leq l < r \leq +\infty$. Il nous faut formuler une hypothèse qui assure que la diffusion n'approche pas trop les bords l, r , qui sont en général des points d'explosion pour les fonctions que nous considérerons dans la suite. Pour cela nous introduisons deux fonctions \mathcal{B}_l et \mathcal{B}_r sur (l, r) qui tendent respectivement en l et r vers l'infini (par exemple $\mathcal{B}_l = 1 + (x - l)^{-1}$ et $\mathcal{B}_r = 1 + (r - x)^{-1}$ si les bords sont finis, ou $\mathcal{B}_l(x) = \mathcal{B}_r(x) = 1 + |x|^k$ quand $l = -\infty$ ou $r = \infty$). Nous introduisons l'hypothèse **(R)** suivante :

$$\forall k \geq 0, \exists c > 0, \forall t \geq 0, \quad E \left(\sup_{s \in [t, t+1]} \mathcal{B}_l(X_s)^k \mid \mathcal{G}_t \right) \leq c \mathcal{B}_l^k(X_t) \quad (2.2)$$

$$\forall k \geq 0, \exists c > 0, \forall t \geq 0, \quad E \left(\sup_{s \in [t, t+1]} \mathcal{B}_r(X_s)^k \mid \mathcal{G}_t \right) \leq c \mathcal{B}_r^k(X_t) \quad (2.3)$$

où $\mathcal{G}_t = \sigma(X_s, s \leq t)$.

Nous démontrons que pour une diffusion sur $(-\infty, \infty)$, cette hypothèse est vérifiée si les coefficients a et b satisfont la condition usuelle de croissance sous-linéaire. Vérifier (R) est plus difficile lorsque que l'un des bords est fini. Nous avons en particulier étudié les diffusions suivantes sur $(0, \infty)$, issues de modèles financiers,

$$dX_t = \mu(X_t - m)dt + \sigma X_t^\psi dt, \quad \mu < 0, m > 0 \text{ et } \psi \in [1/2, 1].$$

Pour $\psi = 1$ (diffusion 'GARCH'), on peut ramener problème de la borne 0 à une borne infinie en considérant X^{-1} . Le cas $\psi = 1/2$ (modèle de Cox-Ingersoll-Ross) est plus difficile à traiter. Il faut utiliser le lien entre la loi de ce processus et celle d'un processus de Bessel ([Leb97]) dont on connaît, par ailleurs, la loi du minimum [RY99]. On trouve que si $c_0 = 2|\mu|m/\sigma^2 > 1$ alors la condition (R) est vérifiée avec la restriction que (2.2) n'est vraie que pour $k < c_0 - 1$. Remarquons que si $c_0 < 1$, la borne 0 est atteignable par le processus et donc (R) ne peut être vraie. Nous déduisons le cas $\psi \in (1/2, 1)$ du cas précédent.

Nous montrons alors sous l'hypothèse (R) et des conditions de régularité sur les coefficients a et b le théorème suivant :

Théorème 2.1. *On a,*

$$\bar{X}_{i,n} - X_{i\Delta_n} = a(X_{i\Delta_n})\Delta_n^{1/2}\xi'_{i,n} + e_{i,n} \quad (2.4)$$

$$\bar{X}_{i+1,n} - \bar{X}_{i,n} - b(\bar{X}_{i,n})\Delta_n = a(X_{i\Delta_n})\Delta_n^{1/2}(\xi_{i,n} + \xi'_{i+1,n}) + \varepsilon_{i,n} \quad (2.5)$$

où $\xi_{i,n} = \Delta_n^{-3/2} \int_{i\Delta_n}^{(i+1)\Delta_n} (s - i\Delta_n) dB_s$ et $\xi'_{i+1,n} = \Delta_n^{-3/2} \int_{(i+1)\Delta_n}^{(i+2)\Delta_n} ((i+2)\Delta_n - s) dB_s$ et $e_{i,n}, \varepsilon_{i,n}$ sont des termes de reste d'ordre Δ_n .

Le développement (2.5) est un schéma d'Euler pour $(\bar{X}_{i,n})$ qui est l'approximation pour une diffusion X quelconque de la formule exacte (2.1) obtenue dans le cas d'un processus d'Ornstein-Uhlenbeck. A la différence du schéma d'Euler standard on voit que les innovations $(\xi_{i,n} + \xi'_{i+1,n})$ ne sont pas i.i.d. mais ont la structure d'un processus ARMA(1,1). Nous pouvons déduire de ce résultat un comportement pour la variation quadratique des moyennes d'une diffusion. On montre en effet que si $\Delta_n = T/n$ avec T fixe,

$$\sum_{i=0}^{n-1} (\bar{X}_{i+1,n} - \bar{X}_{i,n})^2 \xrightarrow{n \rightarrow \infty} 2/3 \int_0^T a(X_s)^2 ds.$$

Ce résultat est à rapprocher de ceux de Dellatre et Jacod (1997) [DJ97] qui étudient la variation quadratique du processus X observé avec erreurs d'arrondis.

2.3 Estimation jointe des coefficients de dérives et diffusions sous condition d'ergodicité

Dans ce travail [3], nous reprenons le cadre du travail précédent [2] et faisons l'hypothèse que la diffusion est récurrente positive avec une mesure stationnaire ν_0 . Le temps total d'observation $n\Delta_n$ tend vers l'infini avec n .

Nos premiers résultats sont des théorèmes limites pour des fonctionnelles de $\bar{X}_{i,n}$. Nous établissons, les convergences suivantes pour $f : (l, r) \rightarrow \mathbb{R}$ suffisamment régulières :

$$n^{-1} \sum_{i=0}^{n-1} f(\bar{X}_{i,n}) \xrightarrow[\mathbf{P}]{n \rightarrow \infty} \nu_0(f), \quad (2.6)$$

$$(n\Delta_n)^{-1} \sum_{i=0}^{n-1} f(\bar{X}_{i,n})(\bar{X}_{i+1,n} - \bar{X}_{i,n} - \Delta_n b(\bar{X}_{i,n})) \xrightarrow[\mathbf{P}]{n \rightarrow \infty} \frac{1}{6} \nu_0(f'), \quad (2.7)$$

$$(n\Delta_n)^{-1} \sum_{i=0}^{n-1} f(\bar{X}_{i,n})(\bar{X}_{i+1,n} - \bar{X}_{i,n})^2 \xrightarrow[\mathbf{P}]{n \rightarrow \infty} \frac{2}{3} \nu_0(f). \quad (2.8)$$

Les résultats (2.7)–(2.8) soulignent la différence de comportement asymptotique entre quantités basées sur les moyennes locales et les valeurs exactes ($X_{i\Delta_n}$). En particulier, le fait que la limite dans (2.7) soit différente de zéro si f n'est pas constante provient de la structure non i.i.d des innovations dans le schéma d'Euler (2.5). Pour préciser ces théorèmes limites nous obtenons, sous l'hypothèse $n\Delta_n^2 \rightarrow 0$, des Théorèmes Centraux associés à (2.7) et (2.8) avec des vitesses de convergences respectivement égales à $\sqrt{n\Delta_n}$ et \sqrt{n} . Remarquons que l'hypothèse de décroissance rapide du pas de discrétisation, $n\Delta_n^2 \rightarrow 0$, est classique pour l'obtention de vitesses de convergence [DCFZ86] [Yos92].

Supposons maintenant que les coefficients de diffusion et dérive, $a(x, \sigma)$ et $b(x, \mu)$ dépendent d'un paramètre $\theta = (\mu, \sigma) \in \Theta_1 \times \Theta_2$ où Θ_1 et Θ_2 sont pour simplifier deux compacts de \mathbb{R} . Pour estimer ce paramètre on introduit la fonction de contraste suivante :

$$\begin{aligned} \mathcal{L}_n(\mu, \theta) = & \sum_{i=0}^{n-1} \frac{(\bar{X}_{i+1,n} - \bar{X}_{i,n} - b(\bar{X}_{i,n}, \mu)\Delta_n)^2}{\frac{2}{3}\Delta_n a^2(\bar{X}_{i,n}, \sigma)} + \sum_{i=0}^{n-1} \log a^2(\bar{X}_{i,n}, \sigma) \\ & + \frac{\Delta_n}{2} \sum_{i=0}^{n-1} \frac{\partial}{\partial x} \left(\frac{b(\bar{X}_{i,n}, \mu)}{a^2(\bar{X}_{i,n}, \sigma)} \right) (\bar{X}_{i+1,n} - \bar{X}_{i,n})^2 \frac{3}{2\Delta_n} \end{aligned} \quad (2.9)$$

Ce contraste est obtenu en modifiant le contraste usuel basé sur les $X_{i\Delta_n}$ (voir Kessler (1997) [Kes97]) pour tenir compte de la différence de comportement asymptotique entre fonctionnelles de $(X_{i\Delta})$ et de $(\bar{X}_{i,n})$ mise en évidence par (2.7)–(2.8).

Considérons l'estimateur $(\hat{\mu}_n, \hat{\sigma}_n) = \operatorname{arginf}_{\mu, \sigma} \mathcal{L}_n(\mu, \sigma)$.

Sous la condition d'identifiabilité (minimale) que les fonctions de dérive et diffusion ne sont par ν_0 presque partout égales pour des valeurs différentes des paramètres, on démontre le résultat suivant :

Théorème 2.2. 1) Si $\Delta_n \rightarrow 0$, $n\Delta_n \rightarrow \infty$, l'estimateur est consistant en probabilité.

2) De plus, si $n\Delta_n^2 \rightarrow 0$, on a la convergence en loi de $\left((n\Delta_n)^{\frac{1}{2}}(\hat{\mu}_n - \mu_0), n^{\frac{1}{2}}(\hat{\sigma}_n - \sigma_0) \right)$ vers une loi gaussienne

$$\mathcal{N} \left(0, \left\{ \nu_0 \left(\frac{(\partial_\mu b)^2(\cdot, \mu_0)}{a^2(\cdot, \sigma_0)} \right) \right\}^{-1} \right) \otimes \mathcal{N} \left(0, \frac{9}{16} \left\{ \nu_0 \left(\frac{(\partial_\sigma a)^2(\cdot, \sigma_0)}{a^2(\cdot, \sigma_0)} \right) \right\}^{-1} \right).$$

Remarquons que l'estimation de la dérive est efficace et l'estimation du paramètre de diffusion est à vitesse \sqrt{n} avec une variance asymptotique légèrement supérieure à celle obtenue par Kessler (1997) [Kes97] dans le cas de l'observation directe de X .

Exemple d'application. Pour illustrer la simplicité des estimateurs obtenus par notre méthode, considérons par exemple un modèle C.I.R.

$$dX_t = (\mu X_t + \mu')dt + \sigma \sqrt{X_t}dB_t.$$

Les estimateurs obtenus sont alors entièrement explicites :

$$\begin{bmatrix} \Delta_n \sum_{i=0}^{n-1} \bar{X}_{i,n} & n\Delta_n \\ n\Delta_n & \Delta_n \sum_{i=0}^{n-1} \bar{X}_{i,n}^{-1} \end{bmatrix} \begin{bmatrix} \hat{\mu}_n \\ \hat{\mu}'_n \end{bmatrix} = \begin{bmatrix} \sum_{i=0}^{n-1} \{\bar{X}_{i+1,n} - \bar{X}_{i,n}\} \\ \sum_{i=0}^{n-1} \left\{ \bar{X}_{i,n}^{-1} (\bar{X}_{i+1,n} - \bar{X}_{i,n}) + \frac{1}{4} \bar{X}_{i,n}^{-2} (\bar{X}_{i+1,n} - \bar{X}_{i,n})^2 \right\} \end{bmatrix},$$

et

$$\hat{\sigma}_n^2 = \frac{3}{2} (n\Delta_n)^{-1} \sum_{i=0}^{n-1} \bar{X}_i^{-1} (\bar{X}_{i+1} - \bar{X}_i)^2.$$

De plus, le théorème 2.2 implique le comportement asymptotique suivant :

$$\begin{bmatrix} (n\Delta_n)^{\frac{1}{2}} (\hat{\mu}_n - \mu) \\ (n\Delta_n)^{\frac{1}{2}} (\hat{\mu}'_n - \mu') \\ n^{\frac{1}{2}} (\hat{\sigma}_n^2 - \sigma^2) \end{bmatrix} \xrightarrow[\mathcal{D}]{n \rightarrow \infty} \mathcal{N} \left(0, \begin{bmatrix} 2|\mu| & \sigma^2 - 2\mu' & 0 \\ \sigma^2 - 2\mu' & (2\mu' - \sigma^2) \frac{\mu'}{|\mu|} & 0 \\ 0 & 0 & \frac{9}{4}\sigma^4 \end{bmatrix} \right).$$

2.4 Estimation de paramètres sous l'hypothèse d'un faible coefficient de diffusion

Le dernier travail [4] que nous présentons dans ce chapitre concerne l'obtention d'estimateurs explicites dans un modèle de diffusion avec petit bruit. Considérons, pour $\varepsilon > 0$, la solution de l'EDS

$$\begin{aligned} dX_t^\varepsilon &= b(X_t^\varepsilon, \alpha)dt + \varepsilon \sigma(X_t^\varepsilon, \beta)dW_t, \quad t \in [0, 1], \\ X_0 &= x_0, \end{aligned} \tag{2.10}$$

où W est un mouvement brownien r -dimensionnel. Le paramètre à estimer (α, β) est élément du produit des deux compacts $\bar{\Theta}_\alpha \times \bar{\Theta}_\beta \subset \mathbb{R}^p \times \mathbb{R}^q$. Nous faisons l'hypothèse que le coefficient de diffusion est asymptotiquement petit : $\varepsilon \rightarrow 0$. Si X^ε est observé continûment, seul le problème de l'estimation de la dérive α se pose. Le cas paramétrique est traité dans Kutoyants (1994) ([Kut94]) et des extensions au cas semi paramétrique le sont plus tard par le même auteur ([Kut98], [IK01]).

Si X est observé aux instants k/n pour $k = 0, \dots, n$, Genon-Catalot (1990) [GC90] propose un estimateur de α dans le cas où la diffusion est unidimensionnelle.

Sous la condition, $\varepsilon = O(n^{-1/2})$, reliant le pas discrétisation et la taille du coefficient de diffusion, l'estimateur proposé est asymptotiquement équivalent au maximum de vraisemblance basé sur l'observation continue du processus. Un estimateur des coefficients de dérive et de diffusion basé sur $(X_{k/n})_{k=0,\dots,n}$ est proposé par Sorensen et Uchida (2003) [SU03] sous la condition restrictive

$$(\varepsilon\sqrt{n})^{-1} \rightarrow M > 0. \quad (2.11)$$

Sous des conditions proches, Uchida (2004) [Uch04] considère un estimateur du coefficient de dérive par fonction d'estimation de type martingale approchée.

Dans un travail commun avec M. Sørensen [4] nous proposons un estimateur par fonction de contraste des deux paramètres α, β qui soit asymptotiquement gaussien en remplaçant la condition (2.11) par la condition moins restrictive :

$$\overline{\lim}(\varepsilon n^\rho)^{-1} < \infty \text{ pour un certain } \rho > 0. \quad (2.12)$$

Le contraste que nous introduisons est de la forme

$$\tilde{U}_{\varepsilon,n}(\theta) = \left(\sum_{k=1}^n \{ \log \det \Xi_{k-1}(\beta) + \varepsilon^{-2} n \tilde{P}_k^*(\alpha) \Xi_{k-1}(\beta)^{-1} \tilde{P}_k(\alpha) \} \right) \quad (2.13)$$

avec $\Xi_k(\beta) = [\sigma\sigma^*](X_{k/n}, \beta)$ et $\tilde{P}_k(\alpha) = X_{k/n} - \tilde{\xi}(X_{(k-1)/n}, \alpha)$ où $\tilde{\xi}_{1/n}(x, \alpha)$ est a priori n'importe quelle fonction de prédiction de $X_{i/n}$ quand $X_{(i-1)/n} = x$. Une originalité du travail est que nous exhibons des conditions générales sur $\tilde{\xi}$ pour que l'estimateur résultant du contraste soit consistant et asymptotiquement gaussien. Pour cela notons $\xi_t(x, \alpha)$ le flot correspondant à l'équation différentielle ordinaire obtenue en prenant $\varepsilon = 0$ dans (2.10) et introduisons la condition :

[C] Les différences entre les fonctions $\xi_{1/n}$ et $\tilde{\xi}_{1/n}$ et entre leurs dérivés jusqu'à l'ordre 2 en α , sont contrôlées uniformément (sur un compact contenant la trajectoire de X^0) par $\varepsilon n^{-3/2}$.

Nous montrons alors le théorème suivant :

Théorème 2.3. *Sous les conditions [C] et (2.11), l'estimateur $\hat{\theta}_{\varepsilon,n} = \operatorname{arginf} \tilde{U}_{\varepsilon,n}(\theta)$ est consistant et,*

$$\begin{pmatrix} \varepsilon^{-1}(\hat{\alpha}_{\varepsilon,n} - \alpha_0) \\ \sqrt{n}(\hat{\beta}_{\varepsilon,n} - \beta_0) \end{pmatrix} \rightarrow \mathcal{N}(0, I(\theta_0)^{-1}),$$

où $I(\theta_0)$ est la même matrice d'information que dans Sørensen et Uchida (2003) [SU03].

Remarquons qu'il est toujours possible d'obtenir un contraste (2.13) explicite. Si le flot de l'EDO correspondant à $\varepsilon = 0$ est explicite il suffit de prendre $\tilde{\xi} = \xi$. Sinon, il suffit de considérer une approximation à un pas de schéma d'Euler avec ordre

arbitrairement élevé. Le papier de Sørensen et Uchida [SU03] correspondait en particulier à ce choix avec un schéma d'ordre 1 : $\tilde{\xi}(x, \alpha) = x + b(x, \alpha)/n$. Nous montrons aussi que notre formalisme contient les estimateurs du type fonctions d'estimation martingales de Bibby, Jacobsen et Sørensen (2004) [BJS04]. On peut en effet vérifier que le choix $\tilde{\xi}(x, \alpha) = E_\alpha[X_{i/n} | X_{(i-1)/n} = x]$ satisfait nécessairement la condition [C].

Application à un modèle financier. Pour voir si l'hypothèse $\varepsilon \rightarrow 0$ est réaliste en finance, nous avons appliqué notre méthode à un modèle à deux facteurs, log-prix d'actif / taux d'intérêt, avec possibilité de corrélations entre les mouvements browniens (voir Longstaff et Schwartz 1995 [LS95], Overhaus *et al* 2006 [OBB⁺06]).

$$\begin{aligned} dY_t &= (R_t + \mu_1)dt + \varepsilon\kappa_1 dW_t^1, & Y_0 &= y_0 \in \mathbb{R}, \\ dR_t &= \mu_2(m - R_t)dt + \varepsilon\kappa_2\sqrt{R_t}(\rho dW_t^1 + (1 - \rho^2)^{\frac{1}{2}}dW_t^2), & R_0 &= r_0 > 0. \end{aligned}$$

Nous calibrons ce modèle sur des valeurs plausibles pour des taux d'intérêt observés sur 25 ans avec des données mensuelles (voir par exemple Chan *et al* 1992 [CAKLS92]) et un actif sans prime de risque. Il apparaît que l'estimation des paramètres de diffusion est bonne (voir Table 2.1). Par contre l'estimation des paramètres de dérive est plus biaisée. En particulier la valeur de $\mu_1 = -(1/2)\varepsilon^2\kappa_1^2$ est trop petite pour être correctement estimée.

$\hat{\mu}_1$	-0.065 (0.53)	$\hat{\mu}_2$	10.02 (4.9)	\hat{m}	2.06 (0.56)
$\hat{\kappa}_1^2$	24.9 (2.1)	$\hat{\kappa}_2^2$	437 (33)	$\hat{\rho}$	0.013 (0.06)

TABLE 2.1 – moyenne (ecart type) de l'estimateur obtenu par Monte Carlo (les vraies valeurs des paramètres sont $\mu_1 = -0.125$, $\mu_2 = 5.7$, $m = 2$, $\kappa_1^2 = 25$, $\kappa_2^2 = 450$, $\rho = 0$)

Chapitre 3

Bruit de microstructure

Ce chapitre concerne l'estimation de paramètres dans des modèles de diffusions observées avec du bruit. Dans les deux premiers travaux présentés [5]–[6], l'objectif était d'évaluer, de manière théorique, la quantité d'information perdue à cause de ce bruit. Pour cela, nous calculons l'information de Fisher du modèle au travers d'une propriété L.A.N. et présentons ensuite des estimateurs explicites optimaux.

Les modèles de diffusions bruitées sont largement utilisés en finance pour modéliser des prix d'actifs. Le bruit y est alors appelé 'bruit de microstructure de marché' et il modélise que, à très haute fréquence, on ne peut plus négliger dans les variations des prix des phénomènes comme l'illiquidité du marché ou l'écart entre l'offre et la demande. Ce modèle permet, en particulier, de reproduire deux faits empiriques observés sur les données : à très hautes fréquences, les retours sont corrélés négativement et la volatilité est sur-estimée. De nombreux travaux économétriques ont alors considéré le problème d'estimer la volatilité de manière robuste aux bruits. En particulier L. Zhang, P. Mykland et Y. Aït-Sahalia (2005) [ZMAS05], puis Zhang (2006) [Zha06] introduisent des méthodes d'estimations non paramétrique, dont nos résultats (paramétriques) [5] permettent de caractériser l'optimalité en vitesse.

Dans le dernier travail [7] de ce chapitre, nous poursuivons les travaux qui avaient été initiés dans ma thèse sur l'estimation de paramètres dans un modèle à volatilité stochastique. En particulier, nous faisons le lien avec ces développements récents, obtenus en économétrie, autour de l'estimation de volatilité intégrée dans le cadre de bruit de microstructure.

3.1 Estimation d'une diffusion gaussienne bruitée : la propriété LAN

Dans ce travail en commun avec J. Jacod [5] nous considérons une diffusion gaussienne,

$$X_t = \int_0^t \sqrt{c(s, \theta)} dB_s$$

où $\theta \in \Theta$ est un paramètre à estimer. Les observations sont de la forme $X_{i/n} + \sqrt{\rho_n}U_{i,n}$ pour $i = 0, \dots, n$, et pour chaque n les $U_{i,n}$ sont i.i.d. de loi $\mathcal{N}(0, 1)$, indépendants du processus X . La suite ρ_n est une suite bornée qui calibre la taille des erreurs d'observations dans ce modèle.

Notons par p_θ^n la densité sur \mathbb{R}^{n+1} de la loi du vecteur des observations

$$\mathcal{O}^{(n)} = (X_{i/n} + \sqrt{\rho_n}U_{i,n})_{i=0, \dots, n}.$$

Notre résultat principal est le suivant.

Théorème 3.1. *Supposons que $c(\cdot, \cdot) \geq \varepsilon > 0$, $c, \dot{c}, \ddot{c} \in \mathcal{C}(\mathbb{R}^2)$, $\dot{c}(\cdot, \theta)$ est α -Hölder avec $\alpha > 1/4$. Alors le modèle satisfait la propriété LAN autour de θ avec une certaine vitesse u_n et information I_θ :*

$$\forall h, \quad \ln \frac{p_{\theta+u_n h}^{(n)}(\mathcal{O}^{(n)})}{p_\theta^{(n)}} \xrightarrow[\text{loi}]{n \rightarrow \infty} hU \sqrt{I_\theta} - \frac{h^2}{2} I_\theta \quad (3.1)$$

où U suit une loi $\mathcal{N}(0, 1)$. Les vitesses et informations sont données par :

- Cas 1, $n\rho_n \rightarrow 0$: $u_n = n^{1/2}$, $I_\theta = 1/2 \int_0^1 \frac{\dot{c}(s, \theta)^2}{c(s, \theta)^2} ds$,
- Cas 2, $n\rho_n \rightarrow l > 0$: $u_n = n^{1/2}$, $I_\theta = \int_0^1 \frac{\dot{c}(s, \theta)^2 (2 + c(s, \theta)/l)}{2\sqrt{l}c(s, \theta)^{3/2} (4 + c(s, \theta)/l)^{3/2}} ds$,
- Cas 3, $n\rho_n \rightarrow \infty$: $u_n = \left(\frac{\rho_n}{n}\right)^{1/4}$, $I_\theta = 1/8 \int_0^1 \frac{\dot{c}(s, \theta)^2}{c(s, \theta)^{3/2}} ds$.

Interprétation : Rappelons que, de manière générale, une telle propriété LAN permet de caractériser que la vitesse d'estimation optimale dans le modèle est u_n^{-1} et la variance asymptotique associée est $1/I_\theta$ (voir Jegannathan (82, 83) [Jeg82], [Jeg83]).

Le cas 1 dans le théorème 3.1 correspond à la situation où ρ_n tend suffisamment vite vers 0 pour que l'effet du bruit soit négligeable. La vitesse est donc classiquement \sqrt{n} et l'information est la même que dans le cas où les valeurs $X_{i/n}$ sont observées sans bruit [GCJ93]. Dans le cas 2, le bruit a une taille suffisante pour que l'information de Fisher soit diminuée mais la vitesse reste \sqrt{n} . Le cas 3 est le plus intéressant : la présence du bruit modifie la vitesse d'estimation. En particulier si le bruit est de taille constante $\rho_n = \rho$, une situation classique en finance, alors la vitesse d'estimation optimale est $n^{1/4}$.

Idée de la preuve : le cas homogène. Les observations sont gaussiennes, mais le fait qu'elles soient non markoviennes et non stationnaires rend difficile l'étude de la vraisemblance. Notons que la suite des observations forme une chaîne de Markov cachée dont la chaîne de Markov sous-jacente est $(X_{i/n})_{i=0, \dots, n}$. Cependant, les

résultats existants sur l'étude de la vraisemblance des chaînes de Markov cachées (voir Leroux 92 [Ler92], Bickel et Ritov 96,98 [BR96] [BRR98], Douc et Matias 2001 [DM01]), n'ont aucune chance de s'étendre ici. En effet ceux-ci reposent sur la propriété d'oubli de la condition initiale dans le problème de filtrage de la composante cachée par la composante observée. Cette propriété d'oubli est, en général, obtenue comme conséquence d'une propriété d'ergodicité de la chaîne sous-jacente. Or notre asymptotique de pas de discrétisation en $1/n$ exclut cette propriété d'ergodicité. De plus les travaux cités ci-dessus conduisent toujours à une vitesse \sqrt{n} . Ils n'incluent donc pas des situations où l'information contenue dans la composante observée est infiniment plus faible que celle contenue dans la chaîne sous-jacente.

Pour simplifier l'étude nous avons d'abord considéré le cas *homogène* où la fonction $c(s, \theta)$ ne dépend pas de s . Les accroissements entre les observations forment alors la série stationnaire,

$$R_i = \sqrt{c(\theta)}(B_{i/n} - B_{(i-1)/n}) + \sqrt{\rho_n}(U_{i,n} - U_{i-1,n}), \quad i = 1, \dots, n.$$

Il est facile de voir que l'on peut transformer, par une application linéaire ne dépendant pas de θ , ces données en un vecteur gaussien $Z = (Z_1, \dots, Z_n)$, dont les composantes sont indépendantes et

$$\text{var}(Z_i) = \frac{c(\theta)}{n} + 2\rho_n(1 - \cos(\frac{i\pi}{n+1})).$$

Les premières composantes du vecteur Z sont très informatives sur le paramètre car la part du bruit y est réduite. Par exemple $\text{var}(Z_1) = \frac{c(\theta)}{n} + O(\frac{\rho_n}{n^2})$ et si l'on observait n copies i.i.d. de variables distribuées comme Z_1 il serait possible d'estimer θ à la vitesse \sqrt{n} . Au contraire les dernières composantes de Z contiennent très peu d'information sur θ . Prenons par exemple $\rho_n = 1$ alors $\text{var}(Z_n) \approx \frac{c(\theta)}{n} + 4$ et il est facile de voir qu'il serait impossible de construire un estimateur consistant de θ depuis n copies indépendantes de Z_n . Une étude détaillée de la vraisemblance du vecteur Z nous permet de montrer le théorème 3.1 dans le cas où c ne dépend pas de s .

Un problème apparaissant dans la résolution du cas homogène est que les variables d'intérêts dans l'étude Z_1, Z_2, \dots dépendent de toutes les observations et ne sont donc pas du tout 'locales' : on a par exemple $Z_1 = (\frac{2}{n})^{1/2} \sum_{j=1}^n \sin(\frac{\pi j}{n+1}) R_j$. Il est donc impossible de passer au cas non homogène en appliquant directement l'idée naturelle de 'geler' localement en temps le coefficient de diffusion $c(s, \theta)$.

Le cas général : introduction d'un 'sur' et 'sous' modèle. Pour résoudre le problème dans le cas général, nous avons introduit deux modèles plus simples à étudier pour lesquels nous savons montrer la propriété LAN et nous en déduisons ensuite le théorème 3.1.

La principale difficulté dans l'étude de la vraisemblance étant l'absence de propriété de Markov, notre idée a été de la 'forcer' à réapparaître. Pour cela on se donne un entier k_n et on décide que toutes les k_n observations bruitées, on observe aussi la

valeur exacte de la diffusion sous-jacente. On appelle 'sur-modèle', le modèle statistique constitué des observations $\mathcal{O}^{(n),\text{sur}}$ ainsi obtenues :

$$\mathcal{O}^{(n),\text{sur}} = \mathcal{O}^{(n)} \cup \{X_{mk_n/n}, m = 0, \dots, l_n := \lfloor n/k_n \rfloor\}.$$

Nous pouvons découper l'observation $\mathcal{O}^{(n),\text{sur}}$ en l_n blocs de données :

$$\mathbb{B}_m = \{X_{\frac{mk_n}{n}}, X_{\frac{mk_n+1}{n}} + \rho_n U_{mk_n+1,n}, \dots, X_{\frac{mk_n+k_n-1}{n}} + \rho_n U_{mk_n+k_n-1,n}\}$$

pour $m = 0, \dots, l_n - 1$. Un avantage du 'sur-modèle' est que en utilisant la propriété de Markov de la diffusion aux instants mk_n/n , la log-vraisemblance de $\mathcal{O}^{(n),\text{sur}}$ a une structure additive :

$$\ln \frac{p_{\theta+u_n h}^{(n),\text{sur}}(\mathcal{O}^{(n),\text{sur}})}{p_{\theta}^{(n),\text{sur}}} = \sum_{m=0}^{l_n-1} \log \text{vraisemblance du bloc } \mathbb{B}_m$$

On peut alors montrer que sous la condition $k_n/n \rightarrow 0$, et si c est assez régulière, on peut remplacer la log-vraisemblance de chaque bloc par une log-vraisemblance provenant d'un modèle homogène. On arrive alors à montrer sous la condition $l_n = o(u_n^{-2})$ que la log-vraisemblance du 'sur-modèle' satisfait une propriété LAN exactement du même type que (3.1). Remarquons que la condition $l_n = o(u_n^{-2})$ est inévitable. Elle assure que les l_n observations rajoutées ne suffisent pas, à elles seules, à estimer le paramètre à une vitesse supérieur à u_n^{-1} et ne modifient donc pas la vitesse du problème.

Un 'sous-modèle' est obtenu en enlevant des données au vrai modèle tout en cherchant à simplifier la structure de la vraisemblance résultante. Il est facile de voir que si l'on supprime dans le vecteur (R_1, \dots, R_n) les variables R_{mk_n} pour $m = 0, \dots, l_n$ on obtient un modèle statistique dont les données forment des blocs indépendants. Sous les mêmes conditions $k_n/n \rightarrow 0$ et $l_n = o(u_n^{-2})$ on peut alors montrer que ce modèle, moins informatif, satisfait la propriété LAN (3.1).

Il est alors très naturel que notre modèle initial, 'coincé' entre ces deux modèles, vérifie aussi la propriété LAN. C'est une conséquence du résultat général suivant que nous montrons.

Proposition 3.2. *Soit une suite de modèles statistiques $(\Omega_n, \mathcal{F}_0^n, \mathcal{F}_1^n, \mathcal{F}_2^n, (P_{\theta}^n))$, avec $\mathcal{F}_0^n \subset \mathcal{F}_1^n \subset \mathcal{F}_2^n$ et notons $Z_{\theta/\zeta}^{n,i} = (dP_{\theta}^n/dP_{\zeta}^n)|_{\mathcal{F}_i^n}$. Supposons que $Z_{\zeta_n/\theta}^{n,i}$ converge en loi sous P_{θ}^n vers une limite Y vérifiant $0 < Y < \infty$ et $E(Y) = 1$, pour $i = 0$ et pour $i = 2$ (la suite ζ_n étant quelconque). Alors cette convergence a aussi lieu pour $i = 1$.*

3.2 Estimateurs efficaces pour une diffusion bruitée

Dans ce travail avec J. Jacod [6] nous poursuivons l'étude de l'inférence d'une diffusion bruitée en considérant un modèle général pour le processus X ,

$$dX_t = b_t dt + \sqrt{c(t, X_t, \theta)} dW_t, \quad \mathcal{L}(X_0) = \eta. \quad (3.2)$$

Notre observation est toujours de la forme $\mathcal{O}^n = (X_{\frac{i}{n}} + \rho_n U_{i,n})$ où les bruits sont, à n fixé, i.i.d. de loi $\mathcal{N}(0, 1)$. Ces bruits sont définis sur un seul espace probabilisé comme des accroissements d'un brownien auxiliaire $U_{i,n} = \sqrt{n}(W'_{(i+1)/n} - W'_{i/n})$, avec W' indépendant de W .

Nos hypothèses sur la diffusion (3.2) sont :

Hypothèse (HR) : $(\theta, t, x) \mapsto c(\theta, t, x)$ est une fonction de $\Theta \times [0, 1] \times \mathbb{R} \rightarrow (0, \infty)$, qui est deux fois continûment différentiable en θ et une fois en x . Les fonctions c, \dot{c}, \ddot{c} sont deux fois continûment différentiables en x et une fois en t . Le processus de dérive $b = b_t(\omega)$ est optionnel et borné sur les événements du type $\{\sup_{t \in [0, 1]} |X_t| \leq q\}$.

Sous ces hypothèses la loi de X est uniquement déterminée par (3.2), notons cette loi P_θ . Nous faisons les hypothèses d'identifiabilité suivantes :

Hypothèse (HI) : (i) Si $\xi \neq \theta$ on a $P_\theta(c(t, X_t, \xi) - c(t, X_t, \theta)) = 0$, $\forall t \in [0, 1]$
(ii) On a $P_\theta(\dot{c}(t, X_t, \theta) = 0, \forall t \in [0, 1])$.

Plaçons nous dans un des trois cas du Théorème 3.1 avec le choix de la vitesse correspondante :

$$\begin{cases} \text{Cas 1, } n\rho_n \rightarrow 0 : & \text{on prend } u_n = 1/\sqrt{n} \\ \text{Cas 2, } n\rho_n \rightarrow l > 0 : & \text{on prend } u_n = 1/\sqrt{n} \\ \text{Cas 3, } n\rho_n \rightarrow \infty : & \text{on prend } u_n = (\rho_n/n)^{1/4} \end{cases} .$$

Alors nous avons le résultat suivant.

Théorème 3.3. *Supposons (HR) et (HI). On peut construire un estimateur $\hat{\theta}_n$, qui converge en probabilité vers θ sous P_θ . De plus, si $\theta \in \overset{\circ}{\Theta}$, la suite $u_n^{-1}(\hat{\theta}_n - \theta)$ converge en loi sous P_θ , vers une variable de loi mélange gaussien, $\mathcal{N}(0, 1/I_\theta)$ où I_θ a la même expression que dans le théorème 3.1 sauf que l'on remplace les quantités déterministes $c(s, \theta)$ (resp. $\dot{c}(s, \theta)$) par les variables aléatoires $c(s, \theta, X_s)$ (resp. $\dot{c}(s, \theta, X_s)$). De plus, cette convergence en loi est stable par rapport à la tribu engendrée par W et W' (voir [Jac97], [Rén63] pour la définition de convergence stable).*

Dans le cas 1, notre estimateur a le même comportement asymptotique que celui de Genon-Catalot et Jacod (93) [GCJ93] dans le cas où il n'y a pas de bruit. On sait, par exemple par les travaux de Dohal [Doh87], que cet estimateur est optimal et donc le notre l'est aussi dans le cas 1. Dans les cas 2 et 3, une comparaison avec le Théorème 3.1 nous permet de dire que notre estimateur est, au moins dans le cas gaussien, optimal.

L'estimateur que nous construisons dans le théorème 3.3 est obtenu en minimisant une fonction de contraste qui est essentiellement inspirée de la vraisemblance exacte du 'sur-modèle' dans le cas gaussien. Remarquons que les estimateurs classiques basés sur la variation quadratique ne sont pas employables en présence de bruit, même dans le cas 1. En effet, nous vérifions, que si ρ_n n'est pas négligeable devant $n^{-3/2}$ les estimateurs de Genon-Catalot et Jacod [GCJ93] ne convergent alors pas à vitesse \sqrt{n} .

3.3 Étude d'un modèle à volatilité stochastique avec bruit de microstructure

Dans ce travail [7] nous considérons un processus de diffusion V , qui modélise la volatilité d'un actif financier Y solution de $dY_t = \sqrt{V_t}dW_t + \text{'drift'}$. Nous supposons que V est solution d'une équation différentielle stochastique $dV_t = b(V_t, \theta)d\mathcal{B}_t + a(V_t)dt$, qui satisfait les mêmes conditions de régularités que dans les travaux [2]–[3] du chapitre 2. De manière générale, on sait qu'il est impossible d'observer exactement la volatilité d'un actif et de nombreuses méthodes ont été proposées en économétrie pour l'estimer ([BZ02], [BNS04], [ZMAS05], [Zha06], [CD06], [BNHLS08]). Pour rester très général vis à vis de ces méthodes, nous faisons l'hypothèse suivante sur l'observation :

Pour un certain pas $\Delta = \Delta_n \rightarrow 0$, nous pouvons estimer les volatilités moyennes $\bar{V}_i = \Delta_n^{-1} \int_{i\Delta_n}^{(i+1)\Delta_n} V_s ds$ par certaines variables observées \hat{V}_i , $i = 0, \dots, N_n$. De plus la variable \hat{V}_i est mesurable par rapport à la tribu $\mathcal{G}_{i\Delta_n} = \sigma(V_s, Y_s, s \leq (i+1)\Delta_n) \vee \tilde{\mathcal{G}}_{(i+1)\Delta_n}$ où $(\tilde{\mathcal{G}}_t)$ est une filtration engendrée par des variables indépendantes des processus (qui contient, par exemple, des bruits de microstructure).

Résultats sur l'estimation du paramètre

Nous faisons d'abord les hypothèses suivantes sur la qualité de l'estimation de la volatilité :

(B) Il existe une suite $b_n \xrightarrow{n \rightarrow \infty} 0$ telle que :

$$\forall i = 0, \dots, N_n - 1, \quad \left| E \left[\hat{V}_i - \bar{V}_i \mid \mathcal{G}_{i\Delta_n} \right] \right| \leq b_n c(1 + V_{i\Delta_n}^c).$$

(V) Il existe une suite $v_n \xrightarrow{n \rightarrow \infty} 0$ telle que pour tout $p \geq 1$, il existe $c(p)$ avec :

$$\forall i = 0, \dots, N_n - 1, \quad E \left[\left| \hat{V}_i - \bar{V}_i \right|^{2p} \mid \mathcal{G}_{i\Delta_n} \right] \leq v_n^p c(p)(1 + V_{i\Delta_n}^{c(p)}).$$

Sous les mêmes hypothèses de régularité, d'ergodicité et d'identifiabilité sur la diffusion que dans les travaux [2]–[3], nous montrons le résultat suivant.

Théorème 3.4. *Supposons $\Delta_n \rightarrow 0$, $N_n \Delta_n \rightarrow \infty$, $b_n = o(\Delta_n)$, $v_n = o(N_n \Delta_n^2)$ alors on peut construire un estimateur $\hat{\theta}_n$, basé sur les $(\hat{V}_i)_{i=0, \dots, N}$, qui converge en probabilité vers la vraie valeur θ_0 du paramètre.*

Si on suppose de plus, $N_n \Delta_n^3 \rightarrow 0$, $b_n = o(\Delta_n^{1/2} N_n^{-1/2})$, $v_n = o(\Delta_n)$, alors :

$$\sqrt{N_n \Delta_n} (\hat{\theta}_n - \theta_0) \xrightarrow[\mathcal{L}]{n \rightarrow \infty} \mathcal{N}(0, I(\theta_0)^{-1}). \quad (3.3)$$

La matrice $I(\theta_0)$ est la matrice d'information de Fisher du modèle où $(V_t)_{t \leq N_n \Delta_n}$ est directement observé.

On déduit donc de ce théorème que notre estimateur a une variance asymptotique minimale, à la différence de ceux qui avaient été introduits plus tôt dans Genon-Catalot *et al* (1999, 2000) [GCJL99], [GCJL00]. Notre estimateur est basé sur une approximation de la vraisemblance exacte de l'observation continue, donnée par la formule de Girsanov. L'étude des propriétés asymptotiques de l'estimateur repose sur les résultats qui avaient été obtenus dans [2]–[3]. Cependant la condition sur le pas $N_n \Delta_n^3 \rightarrow 0$ est moins restrictive que dans ces travaux. Cela est dû au fait que l'on estime que le paramètre de dérive, et la forme du contraste utilisé est donc plus simple. Remarquons qu'il est crucial d'obtenir la condition la moins restrictive possible sur la vitesse de décroissance vers 0 du pas Δ_n . En effet, celui ci apparaît aussi, en général, dans le contrôle du biais b_n et de la variance v_n d'estimation de la volatilité.

Étude de différents choix d'estimateurs pour la volatilité

Nous étudions d'abord le cas de la volatilité réalisée. Si on observe le processus de prix Y en certains instants t_j , $j = 0, \dots, n$, on utilise alors

$$\hat{V}_i = \frac{1}{\Delta_n} \sum_{j \in B_i} (Y_{t_{j+1}} - Y_{t_j})^2, \quad \text{où } B_i = \{j \mid t_j \in [i\Delta_n, (i+1)\Delta_n]\}.$$

Il est connu que si $\delta_n = \sup_{j=0, \dots, n} |t_{j+1} - t_j|$ est très petit devant Δ_n , la volatilité réalisée \hat{V}_i est un bon estimateur de la volatilité intégrée. Dans notre travail, nous établissons que si $n\delta_n^2 \rightarrow 0$, on peut choisir Δ_n tel que toutes les hypothèses du Théorème 3.4 soient vérifiées.

Un problème est que la volatilité réalisée est sensible aux bruits de microstructure qui apparaissent pour les données très hautes fréquences. Nous avons alors étudié la possibilité d'utiliser le 'Two-Scales-Realized-Volatility' (TSRV) introduit par Zhang *et al* (05) [ZMAS05]. Supposons, maintenant, que nos observations soient $X_{t_j} = Y_{t_j} + \epsilon_{j,n}$ où les $\epsilon_{j,n}$ sont i.i.d. avec des moments de tout ordre. En suivant l'idée de Zhang *et al.* (2005) [ZMAS05] on introduit la variation quadratique calculée avec la plus haute fréquence possible,

$$\hat{\epsilon}_i^2 = \frac{1}{2(\#B_i - 1)} \sum_{\substack{j \in B_i \\ j+1 \in B_i}} (X_{t_{j+1}} - X_{t_j})^2 \quad (3.4)$$

et pour $M \geq 1$, on considère une moyenne de variations quadratiques calculées avec une fréquence plus basse,

$$[X, X]_i^{\text{avg}, M} = \frac{1}{M} \sum_{\substack{j \in B_i \\ j+M \in B_i}} (X_{t_{j+M}} - X_{t_j})^2. \quad (3.5)$$

L'idée de Zhang *et al.* (2005) [ZMAS05] est que (3.4) est essentiellement un estimateur du bruit de microstructure plutôt que de la volatilité. Au contraire, en diminuant

la fréquence d'échantillonnage dans (3.5), la contribution du bruit est plus réduite. Pour obtenir le TSRV on compense les effets des bruits dans ces deux estimateurs en posant :

$$\hat{V}_i = \frac{1}{\Delta_n} \left[[X, X]_i^{\text{avg}, M} - \frac{2(\#B_i - M)}{M} \hat{\epsilon}_i^2 \right]. \quad (3.6)$$

Il est montré dans Zhang *et al* (2005) que si le processus de volatilité $s \mapsto V_s$ est déterministe, Δ_n fixe, et $M = M_n$ calibré de manière optimale, \hat{V}_i converge vers la volatilité intégrée à vitesse $n^{1/6}$ où n est le nombre d'observations bruitées utilisées. Cette vitesse est très lente, mais ce n'est pas étonnant car dans le problème paramétrique associé (qui correspond au cas particulier $V_t = \theta, \forall t$) nous avons montré que la meilleure vitesse possible était $n^{1/4}$ ([5]–[6]). Remarquons que cette vitesse $n^{1/6}$ n'est pas la meilleure possible dans le problème non paramétrique. En effet, Zhang (2006) [Zha06] trouve un estimateur à vitesse $n^{1/4}$ qui est donc optimal. Un avantage avec l'estimateur non paramétrique (3.6) est qu'il est plus simple à mettre en œuvre et à étudier.

Dans notre travail [7], la volatilité V n'est plus déterministe et le pas Δ_n tend vers 0. Cependant en suivant essentiellement la preuve de [ZMAS05] nous montrons la proposition suivante.

Proposition 3.5. *Supposons $n\delta_n^{11/9} \rightarrow 0$, et choisissons $M_n \sim \delta_n^{-2/3}$ et Δ_n telle que $\delta_n^{5/12} n^{1/4} = o(\Delta_n)$, $\Delta_n = o(n^{-1/2} \delta_n^{-1/2})$, alors les conditions du théorème 3.4 sont vérifiées. Il est donc possible d'estimer efficacement le paramètre de dérive de la volatilité en présence de bruits de microstructure.*

La condition entre le nombre d'observations et le pas de discrétisation, $n\delta_n^{11/9} \rightarrow 0$, est évidemment extrêmement restrictive. Elle résulte du fait que chacune des volatilités intégrées \bar{V}_i est estimée à vitesse très lente en présence de bruit.

Quel estimateur choisir sur un échantillon fini ?

Au vu des conditions de la Proposition 3.5, il n'est pas clair que sur un échantillon fini, l'usage du TSRV soit toujours préférable à celui de la volatilité réalisée. Par méthode de Monte Carlo, nous avons étudié numériquement le comportement de l'estimateur sur un modèle de diffusion GARCH. La volatilité est solution de $dV_t = (\alpha V_t + \beta)dt + \sigma V_t dB_t$ avec les valeurs $\alpha = -10$, $\beta = 4$, $\sigma = 2.5$ qui sont plausibles pour des données financières si l'unité de temps est exprimée en années (voir [ABM05]). Le temps total d'observation est 10 ans. Nous considérons 3 cas possibles pour l'observation discrète du processus de prix.

- Cas A : Le nombre d'observations est $n = 4 \cdot 10^6$. Ceci correspond à des données 'ultra haute fréquence' avec une observation de prix toutes les 10 secondes environ.
- Cas B : Le nombre d'observations est $n = 5 \cdot 10^5$.
- Cas C : Le nombre d'observations est $n = 1.33 \cdot 10^5$. Ce qui correspond à une donnée toutes les 5 minutes environ.

3.3. MODÈLE À VOLATILITÉ STOCHASTIQUE ET BRUIT DE MICROSTRUCTURE 25

La table 3.1 montre les résultats obtenus pour notre estimateur basé sur la volatilité réalisée, en absence de bruit de microstructure. L'estimation est très bonne dans les cas A-B, et est légèrement biaisée dans le cas C. La table 3.2 contient les résultats lorsque l'on ajoute, dans le cas A, des bruits de microstructure gaussiens d'écart-type $2 \cdot 10^{-3}$. On voit que la variance de l'estimateur basé sur la volatilité réalisée augmente, alors que l'estimateur basé sur le TSRV est insensible à la présence du bruit. Cependant l'estimation de β est fortement biaisée, même en absence de bruit, lorsque l'on utilise le TSRV (un problème de biais analogue est discuté en détail dans [ASMZ05]).

TABLE 3.1 – Sans bruit de microstructure, $\alpha = -10$, $\beta = 4$, $\sigma = 2.5$

Cas A	$\hat{\alpha}_n$	-10.14 (1.84)	$\hat{\beta}_n$	4.02 (0.53)
Cas B	$\hat{\alpha}_n$	-10.06 (3.98)	$\hat{\beta}_n$	3.98 (0.53)
Cas C	$\hat{\alpha}_n$	-9.02 (1.84)	$\hat{\beta}_n$	3.59 (0.57)

TABLE 3.2 – Cas A, avec bruit de microstructure, $\alpha = -10$, $\beta = 4$, $\sigma = 2.5$.

Basé sur TSRV	(sans bruit)	$\hat{\alpha}_n$	-9.99 (1.94)	$\hat{\beta}_n$	3.45 (0.47)
	(avec bruit)	$\hat{\alpha}_n$	-9.90 (2.15)	$\hat{\beta}_n$	3.42 (0.55)
Basé sur V.R.	(sans bruit)	$\hat{\alpha}_n$	-10.14 (1.84)	$\hat{\beta}_n$	4.02 (0.53)
	(avec bruit)	$\hat{\alpha}_n$	-9.98 (3.22)	$\hat{\beta}_n$	3.97 (1.03)

Chapitre 4

Maximum de vraisemblance par calcul de Malliavin dans des modèles cachés

L'étude de la vraisemblance exacte dans des modèles de diffusions discrétisées est rendue difficile par le fait que la densité de transition d'une diffusion n'est pas explicite. En 1987, Dohnal [Doh87] étudie la vraisemblance de l'observation de $(X_{i/n}^\theta)_{i=0,\dots,n}$ où X^θ est une diffusion unidimensionnelle,

$$dX_t^\theta = a(X_t^\theta, \theta)dB_t + b(X_t^\theta)dt, \quad X_0^\theta = x_0. \quad (4.1)$$

Il démontre que la vraisemblance satisfait un développement asymptotique du type LAMN. L'approximation polynômiale de cette vraisemblance, ainsi que l'étude asymptotique pour une diffusion avec bords, est réalisée par Aït-Sahalia (2002) [AS02]. Ces travaux s'appuient sur une expression exacte de la transition de X comme espérance d'une fonctionnelle d'un pont brownien. Un problème est que cette expression exacte, qui avait été trouvée dans Dacunha–Castelle et Florens-Zmirou (1986) [DCFZ86], n'existe que dans le cas unidimensionnel (même si des possibilités d'approximations existent encore dans le cas multidimensionnel [AS08]).

La propriété LAMN dans le cadre d'une diffusion multidimensionnelle à été établie par Emmanuel Gobet en 2001 [Gob01] [Gob02]. Un point crucial de ce travail est l'obtention d'une nouvelle expression, basée sur le calcul de Malliavin, pour la transition du processus X .

Dans le travail [8], réalisé avec Emmanuel Gobet, l'objectif était de savoir si cette nouvelle méthode pouvait s'étendre à la situation plus difficile d'une observation non markovienne. Bien que de nombreuses idées de ce travail soient générales, nous nous sommes concentrés sur l'exemple de l'observation d'intégrales de diffusions. Cet exemple nous apparaissait, a priori, plus simple que le cas d'une diffusion bruitée ou qu'un modèle à volatilité stochastique.

Une expression générale pour la 'fonction score' du modèle.

La première étape de notre travail a été de généraliser à des observations non markoviennes les expressions de la vraisemblance obtenues par calcul Malliavin dans [Gob01]. Notons $U^\theta = (U_i^\theta)_{i=0,\dots,n-1}$ nos observations de la trajectoire $(X_t^\theta)_{t \in [0,1]}$. Le vecteur U^θ est une fonctionnelle de Wiener que nous supposons régulière au sens de Malliavin (voir [Nua95]), et régulière par rapport au paramètre θ . Notons $K(\theta)$ la matrice de Malliavin de U^θ :

$$K(\theta) = \left[\left\langle D.U_i^\theta, D.U_j^\theta \right\rangle \right]_{0 \leq i, j \leq n-1}.$$

Théorème 4.1. *Sous la condition de non dégénérescence $\det(K(\theta))^{-1} \in \cap_p \mathbf{L}^p$ on a :*

$$\frac{\dot{p}}{p}(u_0, \dots, u_{n-1}, \theta) = E \left[\delta \left(\sum_{0 \leq i, j \leq n-1} \frac{\partial U_i^\theta}{\partial \theta} K(\theta)_{i,j}^{-1} D.U_j^\theta \right) \mid (U_j^\theta)_j = (u_j)_j \right], \quad (4.2)$$

où δ est l'opérateur de Skorohod.

Remarquons que la formule (4.2) n'est pas l'immédiate extension des formules de E. Gobet sur la densité de transition. En effet, si on choisit $U^\theta = X_t^\theta$, notre formule donne

$$\frac{\dot{p}_t}{p_t}(x_0, y, \theta) = E \left[\delta \left(\frac{\partial X_t^\theta}{\partial \theta} \left\langle D.X_t^\theta, D.X_t^\theta \right\rangle^{-1} D.X_t^\theta \right) \mid X_t = y \right],$$

qui n'est pas la formule de [Gob01]. En fait (4.2) étend la formule usuelle pour le score d'un vecteur gaussien. En particulier si U^θ est un élément du premier chaos de Wiener (et est donc gaussien), le membre de droite de (4.2) se calcule explicitement et on retrouve la formule usuelle pour les vecteurs gaussiens. On voit que la matrice de Malliavin joue exactement le même rôle que la matrice de covariance dans le cas gaussien. En particulier, si les observations ont une structure de dépendance complexe, la matrice de Malliavin est typiquement difficile à évaluer. Le membre de droite de (4.2) reste alors très dur à étudier.

Propriété LAMN pour des intégrales de diffusions.

Spécifions maintenant notre observation : on considère μ une mesure positive sur $[0, 1]$ dont le support n'est pas réduit à $\{0, 1\}$. Et on suppose que notre observation est

$$\mathcal{O}^{n,\theta} = (\bar{X}_i^\theta)_{i=0,\dots,n-1} = \left(\int_0^1 X_{\frac{s+i}{n}}^\theta \mu(ds) \right)_{i=0,\dots,n-1}.$$

Notre résultat est le suivant.

Théorème 4.2. *Supposons $a : \mathbb{R} \times \Theta \rightarrow \mathbb{R}$, $b : \mathbb{R} \rightarrow \mathbb{R}$ de classe \mathcal{C}^4 bornées ainsi que toutes leurs dérivées. Supposons de plus, $a(x, \theta) \geq \underline{a} > 0$ pour tout x, θ . Alors, la vraisemblance du modèle satisfait une propriété LAMN avec vitesse \sqrt{n} et*

information aléatoire $I_\theta = 2 \int_0^1 \left(\frac{\dot{a}}{a} \right)^2 (X_s, \theta) ds$.

C'est à dire, si l'on note $p^{n,\theta}$ la densité de $\mathcal{O}^{n,\theta}$ sur \mathbb{R}^n :

$$\forall h \in \mathbb{R}, \quad \ln \frac{dp^{n,\theta+h/\sqrt{n}}}{dp^{n,\theta}}(\mathcal{O}^{n,\theta}) = hN_n - \frac{1}{2}h^2I_n + o_{\mathbf{P}}(1),$$

avec $I_n \xrightarrow[\mathbf{P}]{n \rightarrow \infty} I_\theta$ et $N_n \xrightarrow[\text{loi stable}]{n \rightarrow \infty} U \times (I_\theta)^{1/2}$, où U suit une loi $\mathcal{N}(0,1)$.

Remarquons que ce résultat nous apprend que l'information contenue dans le modèle ne dépend pas de la mesure μ et est en particulier égale à l'information contenue dans les observations ponctuelles $(X_{i/n})_{i=0,\dots,n}$. En fait, on obtient aussi comme sous produit de la preuve du théorème 4.2 le résultat surprenant suivant : Si on observe à la fois les moyennes locales $(\bar{X}_i^\theta)_{i=0,\dots,n-1}$ et les observations ponctuelles alors l'information du modèle est doublée. Cela veut dire qu'il n'y a pas de redondance entre l'observation des moyennes locales et des valeurs exactes.

L'étude de la log-vraisemblance est essentiellement obtenue en intégrant la formule (4.2) entre θ et $\theta + h/\sqrt{n}$. Cependant on ne cherche pas à appliquer directement cette formule au vecteur $\mathcal{O}^{n,\theta}$ car l'expression résultante est alors inextricable. On utilise d'abord la même astuce que dans le travail [5] en rajoutant des observations ponctuelles de la diffusion à certains instants. Ceci permet de définir un 'sur-modèle' dont l'observation est, pour $k = k_n$ entier,

$$\mathcal{O}^{n,\theta,\text{sur}} = \mathcal{O}^{n,\theta} \cup \{X_{mk/n}, m = 0, \dots, l_n := \lfloor n/k \rfloor\}$$

En utilisant la propriété de Markov de X aux instants mk/n , la log-vraisemblance du sur-modèle est alors la somme de log-vraisemblances de blocs de taille k_n pour lesquelles on applique la formule (4.2). Si $k_n/n \rightarrow 0$, on montre que dans chaque bloc, on peut approximer les données par des éléments du premier chaos. Pour cela on contrôle l'erreur faite en norme Sobolev sur l'espace de Wiener, avec un ordre suffisamment élevé, pour pouvoir faire la substitution dans le membre de droite de (4.2). Il est en particulier nécessaire de contrôler finement en n les moments de l'inverse de la matrice de Malliavin d'un bloc de données. On utilise ensuite que l'on sait faire les calculs, de manière exacte, pour des variables du premier chaos.

Notons que pour bien évaluer les termes d'erreurs, on a aussi besoin de majorations et minorations sur la densité du vecteur des données. Le même problème apparaissait dans [Gob01]–[Gob02]. Dans notre travail, nous montrons la proposition suivante :

Proposition 4.3. *Sous les mêmes hypothèses que celles du théorème 4.2, le vecteur $(X_1^\theta, \int_0^1 X_s^\theta d\mu(s))$ admet une densité $p_{x_0}(u, v)$ qui satisfait :*

$$c_1^{-1}e^{-c_1(u^2+v^2)} \leq p_{x_0}(u, v, \theta) \leq c_2^{-1}e^{-c_2(u^2+v^2)}. \quad (4.3)$$

Les constantes c_1 and c_2 dépendent seulement des bornes sur les fonctions a , a^{-1} , b et leurs dérivées.

Pour prouver la proposition 4.3, nous nous inspirons des méthodes développées par Hirsch et Song (1997, 1998) [HS97]-[HS98] pour montrer la positivité des densités de certaines fonctionnelles de Wiener. L'idée est d'utiliser un développement de ces fonctionnelles autour de leur 'squelette'. Remarquons que la borne inférieure dans (4.3) ne se déduit pas, a priori, des méthodes sur les 'variables uniformément elliptiques' introduites par Kohatsu-Higa (2003) [KH03] dans un contexte proche.

Chapitre 5

Vitesse d'estimation dans des modèles cachés fractionnaires

Le premier travail que nous présentons très succinctement est une note au CRAS [9] rédigée durant ma thèse. Elle concerne l'estimation du paramètre de diffusion pour un modèle à volatilité stochastique. Les résultats sur la vitesse d'estimation que nous obtenions, précisés par un travail ultérieur de Marc Hoffmann [Hof02], nous ont servis de motivations pour l'étude de processus cachés fractionnaires. Les deux travaux [10]–[11] que nous présentons ensuite ont été réalisés avec Marc Hoffmann. Nous considérons respectivement l'estimation d'un paramètre de 'type diffusion' et de Hurst pour un processus caché fractionnaire. Soulignons que les difficultés techniques rencontrées dans le cadre des processus fractionnaires d'indice $H \in (1/2, 1)$ sont très différentes (et plus délicates) que dans le cas des diffusions.

5.1 Le cas d'une volatilité brownienne

Considérons le modèle à volatilité stochastique,

$$dY_t = \sqrt{V_t}dW_t \tag{5.1}$$

$$dV_t = b(V_t)dt + \theta a(X_t)dB_t, \tag{5.2}$$

où B et W sont deux mouvements browniens indépendants. On observe la discrétisation $(Y_{i/n})_{i=0,\dots,n}$. Le résultat que nous obtenions dans [9] est le suivant :

Proposition 5.1. *Sous des conditions de régularité sur a et b analogues à celles de [2], on peut construire un estimateur $\hat{\theta}_n$ tel que $n^{1/4}(\hat{\theta}_n - \theta)$ soit tendue sous P_θ .*

Dans un papier publié en (2002) Marc Hoffmann [Hof02] montre que cette vitesse $n^{1/4}$ est la meilleure possible. Remarquons qu'il y a un lien très fort avec le modèle

de diffusion bruitée étudiée dans [5]–[6]. En effet, notre observation est équivalente à

$$\begin{aligned}\mathcal{O}_i^n &= \log \left(n(Y_{(i+1)/n} - Y_{i/n})^2 \right) = \log \left(\sqrt{n} \int_{i/n}^{(i+1)/n} \sqrt{V_s} dW_s \right)^2 \\ &\stackrel{\text{loi}}{=} \log \left(n \int_{i/n}^{(i+1)/n} V_s ds \right) + \log \mathcal{N}(0, 1)^2.\end{aligned}$$

On voit donc que la discrétisation d'un modèle à volatilité stochastique est formellement très proche de l'observation de la volatilité avec un bruit (multiplicatif) de taille 1. La vraisemblance est donc encore plus compliquée que dans le cas d'un bruit additif, pour lequel on ne savait déjà calculer l'information de Fisher que dans des cas particuliers. Pour ces raisons, l'étude de l'optimalité dans [Hof02] repose sur des méthodes non paramétriques qui utilisent la régularité du signal cachée V et ne donne que des résultats sur la vitesse. Pour mieux comprendre ce résultat, avec Marc Hoffmann, nous avons considéré un modèle à volatilité stochastique avec des trajectoires cachées dont la régularité est différente du cas brownien.

5.2 Estimation de paramètres dans un modèle à volatilité stochastique fractionnaire

Considérons le modèle à volatilité stochastique

$$dY_t = \sigma_t dW_t, \quad \sigma_t = \Phi(\theta, W_t^H), \quad (5.3)$$

où W^H est un mouvement brownien fractionnaire d'indice de Hurst $H \in (1/2, 1)$, et θ est élément d'un certain intervalle compact Θ . Le paramètre H , supposé connu, permet de calibrer l'indice de Hölder des trajectoires de la volatilité, ce qui permettra de caractériser le lien entre la vitesse d'estimation et la régularité de la volatilité. La propriété de mémoire longue du mouvement brownien fractionnaire est aussi une justification du choix de la volatilité dans le modèle (5.3) (voir [CR96], [CR98], [BCdL98], [Rob01], [CR02]).

Énonçons précisément les résultats obtenus avec Marc Hoffmann dans [10]. Nous faisons les hypothèses suivantes.

Hypothèse A. La fonction $\Phi : \Theta \times \mathbb{R} \rightarrow \mathbb{R}$ est élément de \mathcal{C}_P^3 (i.e. admet des dérivées d'ordre 3 qui sont au plus à croissance polynomiale).

Hypothèse B. Pour tout $\theta \in \Theta$, $x \rightsquigarrow \Phi(\theta, x)^2$ est bijective de \mathbb{R} vers un certain ouvert $\mathcal{X} \subset \mathbb{R}$; $(x, \theta) \rightsquigarrow a(\theta, x) := \partial_x \Phi^2(\theta, (\Phi^2)^{-1}(\theta, x))$ est \mathcal{C}_P^3 et $\forall x \in \mathcal{X}, \forall \theta \in \Theta, a(\theta, x) > 0$.

Hypothèse C. On a $\forall x \in \mathcal{X}, \forall \theta \in \Theta, \partial_\theta a(\theta, x)/a(\theta, x) > 0$.

Le résultat d'estimation est :

Théorème 5.2. *Soit $H \in (\frac{1}{2}, 1)$ et supposons les hypothèses A, B et C. Alors il existe un estimateur $\hat{\theta}_n$ basé sur $(Y_{i/n})_{i=0, \dots, n}$ tel que*

$$n^{\frac{1}{2+4H}} (\hat{\theta}_n - \theta)$$

soit une suite tendue sous \mathbb{P}_θ (uniformément en θ).

Remarquons que l'hypothèse A est une simple hypothèse de régularité et que l'hypothèse B permet de définir une fonction a qui joue le rôle d'un coefficient de diffusion. On peut en effet montrer que $V = \sigma^2$ est solution de $dV_t = a(\theta, V_t)dW_t^H$ où l'intégrale stochastique par rapport à W^H est définie trajectoriellement (en utilisant $H > 1/2$). L'hypothèse C est alors une hypothèse d'identifiabilité. Essentiellement ce jeu d'hypothèses exclut que θ ait le rôle d'un paramètre de dérive que l'on ne pourrait pas estimer (comme par exemple dans $\Phi(\theta, W_t^H) = \theta + W_t^H$).

L'estimateur est basé sur l'étude de la discrétisation d'une trajectoire de mouvement brownien fractionnaire. On introduit $\bar{V}_i = N \int_{i/N}^{(i+1)/N} V_s ds$, $i = 0, \dots, N-1$ et on considère la différence généralisée

$$\Delta_d \bar{V}_i := \sum_{j=0}^p d_j \bar{V}_{i+j}, \text{ pour } d = (d_0, \dots, d_p).$$

En suivant Istas et Lang (1997) [IL97], on appelle 'ordre' de la différence $d = (d_0, \dots, d_p) \in \mathbb{R}^{p+1}$ l'entier $m(d)$ tel que :

$$\text{pour } k = 0, \dots, m(d) - 1 : \sum_{i=0}^p d_i i^k = 0, \quad \text{et } \sum_{i=0}^p d_i i^{m(d)} \neq 0.$$

Alors on peut démontrer le résultat suivant :

Proposition 5.3. *Supposons $H > 1/2$ et $m(d) \geq 2$. Il existe une constante $\kappa_{d,H}$ telle que pour toutes fonctions h de classe \mathcal{C}_p^1 :*

$$N^{-2H-1} \sum_{i=0}^{N-p-1} (\Delta_d \bar{V}_i)^2 h(\bar{V}_i) = \kappa_{d,H} \int_0^1 a(V_s, \theta)^2 h(V_s) ds + o_{L^1}(N^{-1/2}). \quad (5.4)$$

De nombreux résultats de convergences existent dans des cadres proches de la proposition 5.3 (voir [BM83], [GL89], [IL97], [Coe01]). La difficulté nouvelle provient ici de la présence de la fonction de poids $h(\bar{V}_i)$. De plus, l'absence de structure de martingale pour le mouvement brownien fractionnaire rend impossible l'emploi des méthodes usuelles dans le cadre des diffusions.

Pour démontrer (5.4) nous nous inspirons de la construction de l'intégrale stochastique, par décomposition sur des bases d'ondelettes, donnée dans [CKR93]. En effet, on sait étudier les quantités $S_N(h) := N^{-1} \sum_{i=0}^{N-p-1} \left\{ \frac{(\Delta_d \bar{V}_i)^2}{N^{-2H}} - \kappa_{d,H}^2 \right\} h_{i/n}$, quand h est déterministe. Pour étendre à un processus stochastique $t \rightarrow h_t \simeq h(V_t)$, on

développe ce processus sur une base (déterministe) de $\mathbf{L}^2([0, 1])$. On utilise ensuite la linéarité de l'opérateur $h \mapsto S_N(h)$. La régularité de type Besov $B_{2,\infty}^H$ de la trajectoire du mouvement brownien fractionnaire nous permet alors de démontrer (5.4) dans le cas $H > 1/2$.

Nous montrons ensuite que la vitesse $n^{1/(2+4H)}$ est optimale, sous l'hypothèse suivante.

Hypothèse D. Sur un compact $K \subset \mathbb{R}$, $\inf_{x \in K, \theta \in \Theta} \Phi(\theta, x) > 0$ De plus, $\inf_{x \in \mathbb{R}, \theta \in \Theta} \partial_x \Phi(\theta, x) > 0$.

Théorème 5.4. Soit $H \in [\frac{1}{2}, 1)$. Supposons A et D. Alors la vitesse $n^{1/(4H+2)}$ est optimale au sens minimax suivant. Pour tout ouvert non vide $U \subset \Theta$, il existe $C > 0$ telle que,

$$\liminf_{n \rightarrow \infty} \inf_F \sup_{\theta \in U} \mathbb{P}_\theta \{v_n^{-1} |F - \theta| \geq C\} > 0,$$

où l'infimum est sur tous les estimateurs.

La preuve du théorème 5.4 repose sur une extension de la méthode introduite dans [Hof02], et la difficulté principale nouvelle réside dans le contrôle d'un changement de loi par une formule de Girsanov sur le mouvement brownien fractionnaire (voir Decreusefond et Üstünel, 1999 [DÜ99]). Nous présenterons plus en détail dans le paragraphe suivant une méthode générale pour obtenir des bornes inférieures, qui contient essentiellement le théorème 5.4.

5.3 Estimation de l'autosimilarité d'une trajectoire cachée

L'objectif du travail [11], réalisé en commun avec Marc Hoffmann, était d'étudier l'estimation du paramètre de Hurst dans un modèle proche du paragraphe précédent. Nous simplifions le modèle (5.3) en supposant que le processus caché est un mouvement brownien fractionnaire et notre observation $(Y_i)_{i=0,\dots,n}$ en est une discrétisation bruitée :

$$V_i = \sigma W_t^H, \quad Y_i = V_{i/n} + a(V_{i/n}) \xi_i^n, \quad i = 0, \dots, n. \quad (5.5)$$

Le travail précédent correspondait essentiellement à l'estimation du paramètre σ pour H connu, et nous nous intéressons maintenant à l'estimation de H avec σ inconnu.

Hypothèses et résultats

Considérons un ensemble de paramètres de la forme $\mathcal{D} := [H_-, H_+] \times [\sigma_-, \sigma_+] \subset (\frac{1}{2}, 1) \times (0, +\infty)$.

On suppose que le processus V et les bruits sont définis sur un même espace de probabilité et nous faisons l'hypothèse suivante sur la structure des bruits.

Hypothèse A. (i) La fonction $x \rightsquigarrow a(x)$ est bornée et continûment différentiable avec une dérivée bornée.

(ii) Pour tout n , il existe une certaine filtration $\mathcal{F}^n = (\mathcal{F}_t^n, t \geq 0)$ telle que le processus V est adapté et le terme de bruit ξ_i^n à l'instant i/n est $\mathcal{F}_{(i+1)/n}^n$ -mesurable. De plus :

$$\mathbb{E}_{H,\sigma}^n \{\xi_i^n \mid \mathcal{F}_{i/n}^n\} = 0, \quad \mathbb{E}_{H,\sigma}^n \{(\xi_i^n)^2 \mid \mathcal{F}_{i/n}^n\} = 1,$$

et $\sup_{(H,\sigma) \in \mathcal{D}} \sup_{i,n} \mathbb{E}_{H,\sigma}^n \{(\xi_i^n)^4\} < +\infty$.

Cette hypothèse contient en particulier le cas où les bruits sont i.i.d. et indépendants du processus.

Théorème 5.5. Sous l'hypothèse A, on peut construire un estimateur \widehat{H}_n dont la vitesse est $n^{1/(2+4H)}$,

$$\limsup_{c \rightarrow \infty} \limsup_{n \rightarrow \infty} \sup_{(H,\sigma) \in \mathcal{D}} \mathbb{P}_{H,\sigma}^n \{n^{1/(2+4H)} |\widehat{H}_n - H| \geq c\} = 0.$$

Remarquons que la vitesse dépend du paramètre que l'on cherche à estimer, ce qui est une situation inhabituelle en statistique paramétrique. On peut montrer que cette vitesse est optimale, sous les conditions supplémentaires suivantes.

Hypothèse B. (i) On a $\inf_x a(x) > 0$.

(ii) Conditionnellement à V , les variables ξ_i^n sont indépendantes, absolument continues avec une densité, $x \rightsquigarrow \exp(-v_{i,n}(x))$, de classe \mathcal{C}^2 . Cette densité, ainsi que ses dérivées, sont négligeables devant $1/x^2$ à l'infini. De plus,

$$\sup_{i,n} \mathbb{E} \left\{ \left(\frac{d}{dx} v_{i,n}(\xi_i^n) \right)^2 (1 + |\xi_i^n|^2) \right\} < +\infty, \quad (5.6)$$

et les fonctions $x \rightsquigarrow \frac{d^2}{dx^2} v_{i,n}(x)$ sont Lipschitz, avec une constante de Lipschitz indépendante de i, n .

Théorème 5.6. Sous les hypothèses A et B et pour tout ensemble de paramètres \mathcal{D} non vide,

$$\liminf_{n \rightarrow \infty} \inf_F \sup_{(H,\sigma) \in \mathcal{D}} \mathbb{P}_{H,\sigma}^n \{n^{1/(2+4H)} |F - H| \geq c\} > 0,$$

l'infimum étant pris sur tous les estimateurs possibles.

L'hypothèse B n'est pas uniquement technique. En particulier (5.6) assure que le modèle statistique de translation/dilatation unidimensionnel $\theta \rightarrow \theta + a(\theta)\xi_i^n$ a une information de Fisher finie. Ceci exclut des situations où la position de la trajectoire cachée $V_{i/n}$ serait trop facile à estimer depuis l'observation Y_i (comme par exemple pour des bruits ξ_i^n de loi discrète...).

Construction de l'estimateur

Le paramètre H est lié à la propriété d'invariance d'échelle du fBM donnée par la relation $(W_{\lambda t}^H)_t = (\lambda^H W_t^H)_t$. Trajectoriellement, cette propriété se retrouve, par exemple, sur le développement dans une base multi-échelle. Soient $d_{j,k}, j \geq 0, k \leq 2^j$ les coefficients dans une base d'ondelettes du signal $(V_t)_{t \in [0,1]}$ et introduisons l'énergie contenue dans le niveau j :

$$Q_j = \sum_{k=0}^{2^j} (d_{j,k})^2.$$

Alors, la propriété d'invariance d'échelle implique une décroissance du type $Q_j \sim 2^{j2H}$ ([TK92], [CKR93], [AGF95], [AV98]). Ceci suggère d'approcher H en fonction des énergies du signal par $\mathcal{H}_j = -\frac{1}{2} \log \frac{Q_{j+1}}{Q_j}$. Cette quantité n'est pas observée, mais si $\widehat{Q}_{j,n}$ est un estimateur non paramétrique de la fonctionnelle quadratique du signal Q_j , on peut poser

$$\widehat{\mathcal{H}}_{j,n} = -\frac{1}{2} \log \frac{\widehat{Q}_{j+1,n}}{\widehat{Q}_{j,n}}. \quad (5.7)$$

L'étude de la quantité $\widehat{\mathcal{H}}_{j,n}$ est alors inspirée des méthodes non paramétriques de Gayraud et Tribouley (1999) [GT99] [Tri00] pour l'estimation adaptative d'une fonctionnelle quadratique. En effet on peut écrire

$$\widehat{\mathcal{H}}_{j,n} - H = \mathcal{H}_j - H + \widehat{\mathcal{H}}_{j,n} - \mathcal{H}_j$$

où $\mathcal{H}_j - H$ ne dépend que du signal et joue le rôle du biais en statistique non paramétrique. Au contraire, $\widehat{\mathcal{H}}_{j,n} - \mathcal{H}_j$ provient de l'erreur d'estimation du signal et joue le rôle d'un terme de variance. On peut alors déterminer la vitesse d'estimation en cherchant à équilibrer ces deux termes. Cependant la situation est inversée par rapport au cas classique : le terme de biais ne dépend pas de la régularité du signal et décroît en $2^{-j/2}$, alors que le terme de variance dépend du niveau de régularité du signal. En effet, l'erreur faite entre $\widehat{Q}_{j,n}$ et Q_j est d'ordre $2^{j/2}/n$, et par (5.7) on voit que le terme de variance est de l'ordre

$$\widehat{\mathcal{H}}_{j,n} - \mathcal{H}_j \simeq \frac{2^{j/2} n^{-1}}{\widehat{Q}_{j,n}} \simeq \frac{2^{j/2} n^{-1}}{2^{-2jH}}, \quad (5.8)$$

où la régularité du signal intervient dans la vitesse de décroissance du dénominateur. L'équilibre entre termes de biais et variance est réalisé pour un niveau $j = j_n(H)$ tel que $2^{j_n(H)} \simeq n^{1/(1+2H)}$ qui conduit à la vitesse souhaitée. Comme ce niveau critique dépend de la régularité du signal que l'on ne connaît pas, il faut recourir à une méthode adaptative. Pour cela on définit un indice j^* en cherchant à minorer la décroissance (a priori inconnue) du niveau d'énergie $\widehat{Q}_{j,n}$ qui apparaît au dénominateur dans le terme de variance (5.8). On pose donc

$$j^* = \sup \left\{ j \mid \widehat{Q}_{j,n} \geq t_{j,n} \right\}.$$

Si le niveau $t_{j,n}$ est bien choisi on peut montrer que cet indice aléatoire j^* est toujours comparable à l'indice optimale.

Lemme 5.7. *Si on choisit $t_{j,n} = 2^j n^{-1}$, alors pour tout $\varepsilon > 0$, il existe $L_\varepsilon \geq 0$ telle que*

$$\mathbb{P}_{\sigma,H}^n(j^* \geq j_n(H) - L_\varepsilon) \geq 1 - \varepsilon.$$

On peut alors déduire que l'estimateur $\hat{H}_n = \hat{\mathcal{H}}_{j^*,n}$ vérifie les conclusions du théorème 5.5.

Idée de la preuve de la borne inférieure

Pour montrer la borne inférieure nous nous sommes inspirés des liens classiques entre vitesse d'estimation et problème de test sur des familles à deux points (voir par exemple Le Cam [LeC73], Donoho et Nussbaum [DN90], Donoho et Liu [DL91]).

Pour f une fonction de $[0,1]$ dans \mathbb{R} , notons $\mathbb{P}(\cdot \mid f)$ la loi de l'observation $(Y_i)_{i=0,\dots,n}$ conditionnellement à $V = f$. De plus rappelons que la distance en variation totale entre deux mesures de probabilité,

$$\|\mu - \mu'\|_{TV} = \sup_{\|h\|_\infty \leq 1} \left| \int h \, d(\mu - \mu') \right|,$$

est maximale quand les mesures sont étrangères. Cette distance est alors égale à 2.

La première étape pour démontrer le théorème 5.6 est de caractériser la précision maximale à laquelle la trajectoire cachée peut être reconstruite depuis l'observation. On peut démontrer le lemme suivant.

Lemme 5.8. *Supposons les hypothèses A et B. Pour tout $\varepsilon > 0$, il existe $M > 0$ tel que, $n\|f - g\|_2^2 \leq M$ et les contrôles sur la norme Hölder $1/2$, $\|f\|_{\mathcal{H}^{1/2}}^2 \vee \|g\|_{\mathcal{H}^{1/2}}^2 \leq M$, impliquent :*

$$\|\mathbb{P}(\cdot \mid f) - \mathbb{P}(\cdot \mid g)\|_{TV} < 2 - \varepsilon$$

Ce lemme montre que depuis l'observation, il est impossible de discerner exactement deux signaux cachés dont la distance en norme \mathbf{L}^2 est de l'ordre de $1/\sqrt{n}$.

Pour montrer que la vitesse maximale d'estimation est $n^{\frac{1}{2+4H}}$, on considère un problème de test entre deux points de l'espace des paramètres H_0 et H_1 tels que $H_1 \simeq H_0 + n^{-\frac{1}{2+4H}}$. L'idée est de construire un 'couplage' entre deux mouvements browniens fractionnaires d'indice respectif H_0 et H_1 tel que, trajectoriellement, les deux processus soient à distance $1/\sqrt{n}$. Par le lemme 5.8, il est impossible de savoir lequel des deux processus a été observé. Il en résulte qu'il est impossible d'estimer si H_0 ou H_1 est la vraie valeur du paramètre.

Cependant un tel couplage est trop contraignant à réaliser (et est sans doute impossible). On relâche donc une contrainte dans la construction en s'autorisant à remplacer un des deux browniens fractionnaires par un processus dont la loi est équivalente.

Rigoureusement, nous montrons la proposition suivante (et démontrons ensuite qu'elle est suffisante avec le lemme 5.8 pour obtenir le théorème 5.6).

Proposition 5.9. Soit $H_0 \in (1/2, 1)$, $\sigma_0 > 0$ et posons $H_1 = H_0 + I^{-1}n^{-\frac{1}{2+4H}}$, $\sigma_1 = \sigma_0(1 + I^{-1}(\log n)n^{-\frac{1}{2+4H}})$.

Pour I suffisamment grand, il existe une suite d'espaces de probabilité $(\mathcal{X}^n, \mathfrak{X}^n, \mathbf{P}^n)$ sur lesquels on peut construire deux processus stochastiques $(\xi_t^{i,n})_{t \in [0,1]}$, $i = 0, 1$ tels que :

(i) Pour $1/2 \leq \alpha < H_0$, les suites des normes Hölder, $\|\xi^{0,n}\|_{\mathcal{H}^\alpha}$ et $\|\xi^{1,n}\|_{\mathcal{H}^\alpha}$, sont tendues sous \mathbf{P}^n .

(ii) Définissons, pour $i = 0, 1$, les lois sur \mathbb{R}^{n+1} : $P^{i,n} = \int_{\mathcal{X}^n} \mathbf{P}^n(d\omega) \mathbb{P}^n(\cdot \mid \xi^{i,n}(\omega))$. On a alors :

$$\lim_{n \rightarrow \infty} \|P^{i,n} - \mathbb{P}_{H_i, \sigma_i}^n\|_{TV} = 0, \quad i = 0, 1.$$

(iii) Il existe une transformation mesurable $T^n : \mathcal{X}^n \mapsto \mathcal{X}^n$ telle que la suite $n\|\xi^{1,n}(\omega) - \xi^{0,n}(T^n(\omega))\|_2^2$ soit tendue sous \mathbf{P}^n .

(iv) Si n est suffisamment grand, la mesure de probabilité \mathbf{P}^n et sa mesure image $T^n \mathbf{P}^n$ sont équivalents sur $(\mathcal{X}^n, \mathfrak{X}^n)$. De plus, pour un $c^* \in (0, 2)$, on a

$$\|\mathbf{P}^n - T^n \mathbf{P}^n\|_{TV} \leq 2 - c^* < 2$$

pourvu que n soit assez grand.

Les processus $\xi^{0,n}$ et $\xi^{1,n}$ sont des approximations des browniens fractionnaires d'indice H_0 et H_1 . Le point (ii) signifie que l'observation bruitée de ces deux processus est statistiquement équivalente au sens de Le Cam [LC86] à l'observation bruitée des browniens fractionnaires. Le point (iii) signifie que l'on peut transformer chaque trajectoire d'un des processus en l'autre avec une erreur d'ordre $1/\sqrt{n}$. Et le point (iv) signifie que cette transformation T^n laisse 'invariante' la loi des processus. Les conditions (iii) et (iv) sont essentiellement contradictoires. L'équilibre entre elles deux détermine la vitesse d'estimation du problème.

Pour construire ce couplage, nous utilisons la représentation du mouvement brownien fractionnaire, sur une base de $\mathbf{L}^2(\mathbb{R})$, donnée par Meyer, Sellan et Taqqu (1999) [MST99]. Un avantage de cette représentation est que les coefficients basses fréquences forment une série temporelle de type ARFIMA et les coefficients hautes fréquences sont i.i.d. Les approximations $\xi^{0,n}$ (resp. $\xi^{1,n}$) sont obtenues en ne retenant qu'un nombre fini de coefficients dans les représentations de $\sigma_0 W^{H_0}$ (resp. $\sigma_1 W^{H_1}$). La transformation T^n consiste alors à envoyer les coefficients basses fréquences de la représentation correspondant à l'indice H_0 sur ceux correspondant à l'indice H_1 . Pour contrôler l'influence de T^n sur la loi du processus, nous devons majorer des distances entre matrices de Toeplitz de grandes tailles. Pour cela nous nous inspirons des méthodes utilisées par Dahlhaus (1989) [Dah89], Fox et Taqqu (1987) [FT87] dans le contexte d'observations de séries temporelles à mémoire longue. Ces calculs sont assez longs et délicats. Ils sont présentés dans la 'version détaillée' du papier [11].

Chapitre 6

Étude statistique de signaux multifractals

Les signaux multifractals apparaissent dans de nombreuses disciplines telles que la turbulence, la finance, l'étude du trafic internet ou celle du génome (voir par exemple [Kol62], [Fri95], [AdCB⁺96], [GBP⁺96], [Man97], [PW99], [BMD00], [YAL01]).

De manière heuristique, une fonction multifractale est une fonction f pour laquelle le comportement asymptotique de la norme l^p des accroissements

$$\lim_{\delta_n \rightarrow 0} \frac{\log \left([T/\delta_n]^{-1} \sum_{k=0}^{[T/\delta_n]} |f((k+1)\delta_n) - f(k\delta_n)|^p \right)^{1/p}}{\log \delta_n} = -s \quad (6.1)$$

dépend du choix de p . Nous renvoyons, par exemple, à Lee *et al* [LLR06], où il est très clair que sur certaines données financières la limite dans (6.1) dépend de p . Étudier des signaux multifractals nous apparaissait, aussi, assez naturel après nos travaux sur le brownien fractionnaire. En effet, si f est la réalisation d'un M.B.f., la limite dans (6.1) vaut presque sûrement l'indice de Hurst (indépendamment de p).

Le premier travail présenté dans ce paragraphe [12], réalisé avec Marc Hoffmann, est cependant très différent des travaux précédents. Il s'agit d'étudier le problème classique de la vitesse d'estimation minimax d'une fonction inconnue appartenant à une certain espace fonctionnel. En utilisant les résultats de S. Jaffard [Jaf00], on résout ce problème sur un espace de fonctions de régularité multifractale prescrite. Un des points important de ce travail est qu'il généralise les résultats bien connus de vitesse minimax sur les espaces de Besov classiques.

Le second travail [13] présenté concerne les processus de cascades multiplicatives, qui sont des processus dont les trajectoires sont multifractales. La motivation de ce travail provient de constatations empiriques, faites par E. Bacry et J.F. Muzy [MBBP08], sur les données de turbulences. En effet sur ces données, le comportement dans (6.1) semble différer selon le temps total d'observation T . Pour comprendre cela, nous étudions la convergence (6.1), pour f réalisation presque sûre d'un processus de cascade, avec $\delta_n \rightarrow 0$ et $T \rightarrow \infty$. Ce travail, en collaboration avec E. Bacry, M.

Hoffmann et J.F. Muzy, étend les résultats de [RSGW03] qui considérait le cas $\delta_n \rightarrow 0$ et $T = 1$. En particulier, nous montrons que la limite de (6.1) dépend effectivement du comportement asymptotique de T .

6.1 Le problème de l'estimation non paramétrique revisité

Considérons le modèle de bruit blanc gaussien :

$$Y_\varepsilon(dx) = f(x)dx + \varepsilon W(dx), \quad (6.2)$$

où $f \in \mathbf{L}^2([0, 1]^d)$ et $W(dx)$ est un champ gaussien sur $[0, 1]^d$ centré, d'intensité donnée par la mesure de Lebesgue. Les observables sont de la forme $Y_\varepsilon(\varphi) := \langle \varphi, f \rangle + \varepsilon \int \varphi(x)W(dx)$, pour toutes fonctions tests φ . Il est bien connu que la vitesse d'estimation de f dans (6.2), sous l'asymptotique $\varepsilon \rightarrow 0$, est représentative de la vitesse d'estimation dans d'autres problèmes non paramétriques : régression, estimation de densité (voir [IH77], [BL96], [Nus96], [GN98]).

On appelle risque minimax sur une classe de fonctions \mathcal{F} , avec perte \mathbf{L}^π ($\pi \geq 1$), la quantité :

$$R(\pi, \mathcal{F}, \varepsilon) = \inf_{\hat{f}_\varepsilon} \sup_{f \in \mathcal{F}} \left(\mathbb{E}_f \left\| \hat{f}_\varepsilon - f \right\|_\pi^\pi \right)^{1/\pi}. \quad (6.3)$$

L'évaluation du risque minimax pour différentes classes de fonctions (régularité Hölderienne, Sobolev, Besov, analytique) à été le sujet de nombreux travaux (voir [IK76], [IH77], [IH78], [Nus85], [GLT96]). Plus tard, le problème d'adaptativité par rapport à la classe \mathcal{F} a aussi été étudié ([GN90], [Lep90], [Lep91], [GN92], [Lep92b], [Lep92a], [Don94], [DJ94a], [DJKP95], [DJKP96], [DJ98]).

Le fait principal est que la vitesse d'estimation dépend de la régularité de la fonction à estimer. Pour illustrer cela, rappelons les résultats de Donoho *et al* (1996) [DJKP96] dans le cadre des boules des espaces de Besov. La norme Besov d'indice $s > 0$, $p > 0$ est, par exemple, définie comme

$$\|f\|_{B_{p,\infty}^s} := \sup_{j \geq -1} 2^{j(s+d(\frac{1}{2}-\frac{1}{p}))} \left(\sum_k |c_{j,k}|^p \right)^{1/p}$$

où $(\psi_{j,k})_{j \geq -1, k}$ est une base d'ondelettes de $[0, 1]^d$ (voir [Coh00]).

Donoho *et al* (1996) [DJKP96] évaluent le risque quand la classe de fonction est une boule $\mathcal{F} = \mathcal{B}_{p,\infty}^s(r)$ de rayon r de l'espace de Besov. Le contrôle asymptotique sur le risque minimax dépend des relations entre π , p et s , et est donné (à un facteur log près) par la décroissance polynômiale :

$$R(\pi, \mathcal{B}_{p,\infty}^s(r), \varepsilon) \asymp_{\varepsilon \rightarrow 0} \varepsilon^{2v(s,p,\pi)}.$$

L'expression de la vitesse est [DJKP96] :

$$v(s, p, \pi) = \begin{cases} \frac{s}{d+2s} & \text{si } \pi \leq p(1+2s), \\ \frac{s - 1/p + 1/\pi}{d+2s - 2/p} & \text{si } \pi \geq p(1+2s). \end{cases}$$

On voit que cette expression présente un 'coude', qui fait que la vitesse devient plus lente si la norme de perte est trop grande par rapport à la norme de contrôle de la régularité. Les deux cas $\pi \leq p(1+2s)$ (resp. $\pi \geq p(1+2s)$) sont parfois appelé cas 'dense' (resp. cas 'lacunaire') à cause de la structure des mesures des problèmes bayesiens associés (voir [DJKP96], [DJ94b]). Le cas $\pi = p(1+2s)$ est souvent appelé 'cas critique'.

Avant d'introduire la classe de régularité qui nous intéresse, nous avons besoin de quelques rappels d'analyse multifractale.

Les fonctions multifractales et la conjecture de Frisch-Parisi

En analyse, la notion de multifractalité est d'abord une notion de régularité locale. Soit $x_0 \in [0, 1]^d$ et $\alpha > 0$, on dit que f est $\mathcal{C}^\alpha(x_0)$ si il existe $c > 0$ et un polynôme P_{x_0} de degré au plus $[\alpha]$ tel que sur un voisinage de x_0 :

$$|f(x) - P_{x_0}(x)| \leq c|x - x_0|^\alpha.$$

L'exposant de Hölder de f en x_0 est

$$h_f(x_0) := \sup \{ \alpha > 0, f \in \mathcal{C}^\alpha(x_0) \},$$

et on considère l'ensemble des points de régularité exactement H :

$$S_f(H) := \{x \in [0, 1]^d, h_f(x) = H\}.$$

Cet ensemble est typiquement de mesure nulle, mais on peut évaluer sa taille en posant pour tout H ,

$$d_f(H) := \begin{cases} \dim S_f(H) & \text{si } S_f(H) \neq \emptyset, \\ -\infty & \text{si } S_f(H) = \emptyset, \end{cases} \quad (6.4)$$

où \dim est la dimension de Hausdorff. La fonction $H \mapsto d(H)$ est appelé *spectre de singularité* de f . Une fonction est considérée multifractale si $d(H) \neq -\infty$ pour au moins deux valeurs de H .

Il est clair que cette notion est délicate à manipuler en pratique. Cependant elle est reliée à la quantité suivante plus exploitable. Pour $p > 0$, on pose

$$s_f(1/p) := \sup \{ s > 0, f \in B_{p,\infty}^s \}. \quad (6.5)$$

La fonction $1/p \mapsto s_f(1/p) \in (0, \infty]$ est bien définie dès que f est globalement Hölder, et est appelé *frontière du domaine de Besov* de f . Notons que les indices $s_f(1/p)$ définis rigoureusement dans (6.5) sont essentiellement les mêmes que la limite apparaissant dans la définition heuristique (6.1). On peut montrer le lemme suivant (voir [Jaf00]).

Lemme 6.1. *La fonction $s_f(\cdot)$ est croissante, concave et vérifie $s'_f(\cdot) \leq d$.*

Il avait été conjecturé par U. Frisch et G. Parisi [FP85] que la régularité locale (6.4) peut être retrouvée depuis des informations globales du type (6.5). La relation proposée (aussi appelé 'formalisme multifractal') est de la forme suivante :

$$d_f(H) = \inf_p \{pH - ps_f(1/p) + d\}. \quad (6.6)$$

Des arguments heuristiques qui permettent de deviner (6.6) peuvent être trouvés dans [FP85], [Jaf97] ou [12]. Il existe des fonctions f qui ne vérifient pas la relation (6.6), cependant Jaffard (2000) [Jaf97] montre que $d_f(H)$ est nécessairement inférieur au membre de droite de (6.6) (à condition de réduire l'ensemble des p sur lequel l'infimum est calculé). Il montre aussi, dans [Jaf00], que la conjecture est vérifiée pour quasi-toutes les fonctions.

C'est à la lecture de ce papier, que nous est venu l'idée du programme suivant. Puisque la régularité locale est reliée à une description en espace fonctionnel, peut-on caractériser la vitesse minimax sur cet espace et la relier à la régularité locale ?

Résultats de vitesse d'estimation

Soit s une fonction vérifiant les conclusions du lemme 6.1 et $s(0) > 0$. Il est naturel d'introduire l'ensemble des fonctions dont le domaine de Besov est inclus dans celui limité par s :

$$\mathcal{M}(s(\cdot), r) := \{f \in \mathbf{L}^2([0, 1]^d), \sup_{p \in (0, +\infty]} \|f\|_{B_{p, \infty}^{s(1/p)}} \leq r\}. \quad (6.7)$$

Bien sur, seul le bord du domaine, c'est à dire les fonctions f telles que $s_f = s$, seront vraiment importantes dans la suite.

Pour $\pi \geq 1$, introduisons l'équation suivante en p :

$$s(1/p) = \frac{d}{2} \left(\frac{\pi}{p} - 1 \right). \quad (6.8)$$

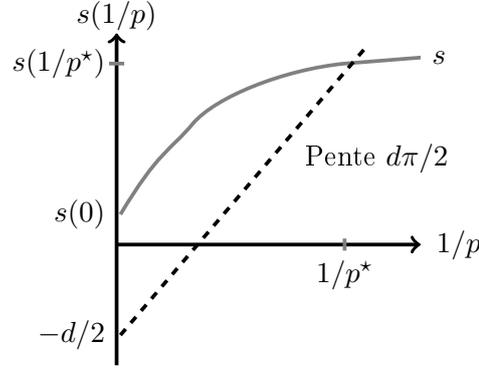
Notons p^* la solution nécessairement unique (si elle existe) à cette équation.

Théorème 6.2. *Si $s'(\infty) < d\pi/2$ (qui est toujours vrai si $\pi > 2$), alors la solution p^* de (6.8) existe.*

- On peut construire un estimateur $\widehat{f}_{\varepsilon, \pi}$ tel que :

$$\sup_{f \in \mathcal{M}(s(\cdot), r)} \left(E_f \left\| \widehat{f}_{\varepsilon, \pi} - f \right\|_{\pi}^{\pi} \right)^{1/\pi} \lesssim \left(\varepsilon \sqrt{\log \frac{1}{\varepsilon}} \right)^{\frac{2s(1/p^*)}{2s(1/p^*)+d}} \left(\log \frac{1}{\varepsilon} \right)^{1/\pi}, \quad (6.9)$$

où \lesssim signifie inférieur à une constante près qui dépend de π , $s(\cdot)$ et r seulement.

FIGURE 6.1 – Régularité effective en fonction de π et s 

– Cas extremal : Si $s'(\cdot) = 0$ sur un voisinage $1/p^*$ alors on a le raffinement :

$$\sup_{f \in \mathcal{M}(s(\cdot), r)} \left(E_f \left\| \widehat{f}_{\varepsilon, \pi} - f \right\|_{\pi}^{\pi} \right)^{1/\pi} \lesssim \left(\varepsilon \sqrt{\log \frac{1}{\varepsilon}} \right)^{\frac{2s(1/p^*)}{2s(1/p^*)+d}}.$$

On voit que $s(1/p^*)$ joue le rôle d'un indice de régularité effectif dans le problème. La figure 6.1 illustre les notations du théorème. Remarquons que lorsque π croît vers ∞ , il est clair sur le dessin que $s(1/p^*)$ décroît jusqu'à $s(0)$, la régularité Hölder globale des fonctions de la classe. Intuitivement cela n'est pas surprenant.

Le résultat suivant précise que la vitesse de (6.9) est optimale à un facteur log près.

Théorème 6.3. *Le risque minimax sur la classe $\mathcal{M}(s(\cdot), r)$ est minoré de la manière suivante :*

– Si $s'(1/p^*) > 0$, alors

$$R(\pi, \mathcal{M}(s(\cdot), r), \varepsilon) \gtrsim \left(\varepsilon \sqrt{\log \frac{1}{\varepsilon}} \right)^{\frac{2s(1/p^*)}{2s(1/p^*)+d}}. \quad (6.10)$$

– Si $s'(1/p^*) = 0$, alors

$$R(\pi, \mathcal{M}(s(\cdot), r), \varepsilon) \gtrsim \varepsilon^{\frac{2s(1/p^*)}{2s(1/p^*)+d}}.$$

On verra que la distinction $s'(1/p^*) = 0$ contre $s'(1/p^*) > 0$ correspond essentiellement à la distinction standard entre cas 'dense' et cas 'lacunaire' dans le cas des espaces de Besov.

On peut maintenant faire le lien avec la régularité locale des fonctions. Notons $\mathcal{FP}(s(\cdot), r)$ le sous ensemble des fonctions de $\mathcal{M}(s(\cdot), r)$ pour lesquelles le spectre de singularité est donné par la conjecture de Frisch-Parisi. Rappelons l'énoncé précis donné par Jaffard (2000) de cette conjecture [Jaf00] :

$$d(H) = \inf_{p \geq p_c} \{Hp - ps(1/p) + d\}, \quad \forall H \in [s(0), d/p_c],$$

où p_c est la solution de $s(1/p_c) = d/p_c$.

Théorème 6.4. *Les bornes sup et inf des théorèmes 6.2-6.3 sont vraies sur $\mathcal{FP}(s(\cdot), r)$. De plus, si $\pi \geq p_c + 2$, l'exposant de la vitesse d'estimation s'exprime de la manière suivante :*

$$\frac{s(1/p^*)}{d + 2s(1/p^*)} = \min_{H \in [s(0), d/p_c]} \frac{H + (d - d(H))/\pi}{2H + d}. \quad (6.11)$$

La formule (6.11) nous permet de voir comment dans la vitesse d'estimation l'impact des points de régularité H est pondéré par leur quantité (mesurée en terme de dimension). En particulier si $\pi \rightarrow \infty$, l'effet des dimensions s'estompe et la vitesse est gouvernée par l'irrégularité la plus forte présente dans le signal : $H = s(0)$.

On retrouve aussi la séparation cas 'dense'/lacunaire'. En effet dans le cas dense, $s'(1/p^*) = 0$, l'infimum dans (6.11) est atteint pour un H tel que $d(H) = d$. Sous la condition $s'(1/p^*) > 0$, il est atteint pour un H tel que $d(H) < d$.

Construction de l'estimateur et explication de la vitesse

L'estimateur utilisé est un estimateur de type seuillage de coefficients sur une base de $\mathbf{L}^2([0, 1]^d)$. Cette base vérifie des propriétés de base inconditionnelle et de 'super concentration', telles que présentées dans Kerkyacharian et Picard (2000) [KP00]. Un point clé de ces propriétés sur la base est qu'elles permettent de contrôler les termes d'erreur non linéaire, en norme \mathbf{L}^π . En effet, elles permettent d'écrire, pour $\pi \neq 2$, la relation de type orthogonalité suivante (voir [DKT98]) :

$$\left\| \sum_{(j,k)} \psi_{j,k} 1_{\{(j,k) \in \Lambda\}} \right\|_\pi^\pi \leq c \sum_{(j,k)} \|\psi_{j,k}\|_\pi^\pi 1_{\{(j,k) \in \Lambda\}} \quad (6.12)$$

où Λ est n'importe quel sous ensemble d'indices. La vitesse de l'estimateur est alors reliée à l'appartenance à des espaces de Besov faibles pour le signal. Ce lien est étudié précisément dans les travaux sur les problèmes de maxiset (voir [KP02], [Riv05]). Il apparaît clairement en comparant le membre de droite de (6.12) à la définition de la norme de l'espace de Besov faible $\ell_{q,\infty}(\pi)$:

$$\forall f = \sum_{j,k} c_{j,k} \psi_{j,k}, \quad \|f\|_{\ell_{q,\infty}(\pi)}^q := \sup_{t>0} t^q \sum_{(j,k)} \|\psi_{j,k}\|_\pi^\pi 1_{\{|c_{j,k}| \geq t\}}.$$

Essentiellement, on peut montrer que la vitesse d'estimation pour un signal dans $\ell_{q,\infty}(\pi)$ est $1 - q/\pi$. Pour voir apparaître la vitesse de notre problème il faut donc chercher pour quelle valeur minimale de q le signal est élément de $\ell_{q,\infty}(\pi)$. A cette fin, rappelons le résultat de [KP00] sur le lien entre espaces forts et faibles :

Proposition 6.5. *(Theorem 6.2 dans [KP00]) Si $1 \leq \pi < \infty$, $s \geq 0$,*

$$p > \frac{d\pi}{2s + d} \implies B_{p,\infty}^s \subset \ell_{\frac{d\pi}{2s+d},\infty}(\pi). \quad (6.13)$$

Par la définition (6.7), on peut appliquer cette proposition à tout couple $(p, s) = (p, s(1/p))$. En cherchant à minimiser l'indice $q = \frac{d\pi}{2s+d}$ résultant, on voit qu'il faut considérer $p^* = \sup\{p \mid p > \frac{d\pi}{2s(1/p)+d}\}$. Ceci conduit immédiatement à l'équation (6.8). Rigoureusement, on ne peut pas appliquer les résultats à ce p critique qui vérifie l'égalité $p = \frac{d\pi}{2s(1/p)+d}$. On remplace alors (6.13) par un résultat proche, sans que cela change la vitesse (à des facteurs log près).

La borne inférieure et les risques Bayésiens

Pour démontrer le théorème 6.3, nous considérons un problème Bayésien, avec une mesure a priori μ sur l'espace $\mathbf{L}^2([0, 1]^d)$. Ce problème Bayésien doit être comparable avec le problème minimax initial. Il faut, en particulier, être capable de construire une mesure qui, avec forte probabilité, charge les fonctions de l'espace $\mathcal{M}(s(\cdot), r)$. Cependant, même la construction d'un seul élément de $\mathcal{M}(s(\cdot), r)$ est un problème compliqué, qui est résolu dans Jaffard [Jaf00]. L'expression trouvée dans [Jaf00] nous semblait difficile à exploiter si $s(\cdot)$ est quelconque. Cependant si s est linéaire, le résultat de S. Jaffard se simplifie, et la construction est essentiellement une "série lacunaire" d'ondelettes à coefficients déterministes.

Considérons donc une fonction linéaire $\tilde{s}(1/p) = \beta + \alpha/p$ avec $\alpha \in [0, d]$ et $\beta > 0$. Si on choisit pour \tilde{s} une fonction tangente à s , on a alors par concavité : $\tilde{s}(\cdot) \geq s(\cdot)$. On déduit donc l'inclusion $\mathcal{M}(\tilde{s}(\cdot), r) \subset \mathcal{M}(s(\cdot), r)$ qui conduit à la comparaison des risques :

$$R(\pi, \mathcal{M}(\tilde{s}(\cdot), r), \varepsilon) \leq R(\pi, \mathcal{M}(s(\cdot), r), \varepsilon). \quad (6.14)$$

Le lemme suivant, simple, nous permet de construire des éléments de $\mathcal{M}(\tilde{s}(\cdot), r)$ comme séries lacunaires :

Lemme 6.6. *Soit $g = \sum_{j \geq 0} \sum_k c_{j,k} \psi_{j,k}$ telle que :*

$$\forall j, k, \quad |c_{j,k}| \leq 2^{-j(\beta+d/2)}$$

et le nombre de coefficients différents de zéro satisfait :

$$\forall j, \quad \text{Card}\{k, c_{j,k} \neq 0\} \leq 2^{j(d-\alpha)}.$$

Alors, on a $\|g\|_{B_{p,\infty}^{\tilde{s}(1/p)}} \leq 1$ pour tout $p > 0$.

Ce lemme nous suggère de définir une mesure a priori de la manière suivante. On fixe un niveau j et on choisit au hasard les coefficients $c_{j,k}$ de la fonction f sur ce niveau j en posant :

$$c_{j,k} = \begin{cases} r2^{-j(\beta+d/2)} & \text{avec probabilité } 2^{-j\alpha-1} \\ 0 & \text{avec probabilité } 1 - 2^{-j\alpha-1} \end{cases} .$$

La pente α calibre donc le niveau de lacunarité de la mesure a priori. Pour $\alpha = 0$ la mesure est dense. Notons μ_j la mesure ainsi définie et considérons le risque de Bayes

associé, pour un estimateur \widehat{f} quelconque :

$$\mathcal{E}_{B,j,\pi}(\widehat{f}) := \left(\int_{L^2} E_f [\|\widehat{f} - f\|_{L^\pi}^\pi] \mu_j(df) \right)^{1/\pi}.$$

En nous inspirant des méthodes de [DJ94b], nous montrons les minoration suivantes pour le risque de Bayes.

Proposition 6.7. 1) Si $\alpha \neq 0$, prenons M avec $0 < M < r^{-2} 2 \log 2$ et soit $j = j(\varepsilon)$ satisfaisant

$$M\alpha(j-1)2^{(j-1)(d+2\beta)} \leq \varepsilon^{-2} < M\alpha j 2^{j(d+2\beta)}.$$

Alors, on a :

$$\inf_{\widehat{f}} \mathcal{E}_{B,j(\varepsilon),\pi}(\widehat{f})^\pi \gtrsim \left(\varepsilon \sqrt{\log \frac{1}{\varepsilon}} \right)^{2 \frac{\alpha + \pi\beta}{d+2\beta}}. \quad (6.15)$$

2) Si $\alpha = 0$, prenons $j = j(\varepsilon)$ satisfaisant

$$2^{(j-1)(d+2\beta)} \leq \varepsilon^{-2} < 2^{j(d+2\beta)},$$

on a alors :

$$\inf_{\widehat{f}} \mathcal{E}_{B,j(\varepsilon),\pi}(\widehat{f})^\pi \gtrsim \varepsilon^{2 \frac{\pi\beta}{d+2\beta}}. \quad (6.16)$$

On peut ensuite montrer que le risque minimax sur $\mathcal{M}(\tilde{s}(\cdot), r)$ est comparable au risque Bayésien pour la mesure μ_j . Puis par la comparaison (6.14), on voit que les membres de droite dans (6.15)-(6.16) sont des minoration pour le risque minimax sur $\mathcal{M}(\tilde{s}(\cdot), r)$. Le choix de α et β est libre, pourvu que $\tilde{s}(1/p) = \beta + \alpha/p$ soit tangent à s . Si on optimise ce choix, on voit qu'il faut considérer une tangente en $1/p^*$ et la borne inférieure est alors (6.10). Le théorème 6.3 est ainsi démontré.

Un point intéressant dans la preuve est que le niveau de lacunarité de la mesure bayésienne dépend de la pente de $s(\cdot)$ en l'indice critique $1/p^*$. En particulier la mesure est dense si $\alpha = s'(1/p^*) = 0$.

Un autre point intéressant dans la preuve est que le choix de la tangente optimale nous montre la relation suivante sur la vitesse :

$$\frac{s(1/p^*)}{d + 2s(1/p^*)} = \inf_{\alpha, \beta} \frac{\alpha + \pi\beta}{d + 2\beta},$$

où l'infimum est restreint aux couples α, β définissant une tangente à s . Cet infimum se réexprime aisément en utilisant la transformée de Legendre de $q \rightarrow s(q)$. Or $d(H)$ est reliée par (6.6) à la transformée de Legendre de $p \rightarrow ps(1/p)$ (qui est la réciproque de la précédente). Par quelques calculs standard sur les fonctions concaves on en déduit l'expression (6.11).

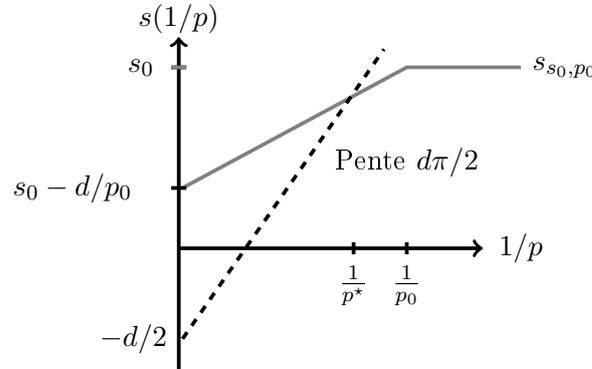
Exemples de domaine de Besov

• *Retour sur les boules de Besov.* En fait le cas classique $\mathcal{F} = \mathcal{B}_{p_0, \infty}^{s_0}(r)$ est contenu dans nos résultats. Il correspond en effet au bord du domaine de Besov défini par la fonction suivante (voir figure 6.2) :

$$s_{s_0, p_0}(1/p) := \begin{cases} s_0 + d(1/p - 1/p_0) & \text{si } 0 \leq 1/p \leq 1/p_0 \\ s_0 & \text{sinon.} \end{cases}$$

On retrouve en particulier que la mesure de Bayes est, soit dense (si $s'(1/p^*) = 0$),

FIGURE 6.2 – Cas d'un seul espace de Besov



soit avec un degré maximum de lacunarité (si $s'(1/p^*) = d$, comme sur le dessin). Le cas 'critique' correspond à $1/p^* = 1/p_0$ et il y alors une infinité de tangentes possibles à la courbe en ce point. Il y a donc une infinité de mesures de Bayes possibles dans la preuve de borne inférieure (si on combine ces mesures, on peut gagner un facteur log dans la borne inférieure, ce qui est connu dans le cas 'critique').

• *L'intersection de deux espaces de Besov.* A titre d'illustration, on peut calculer la vitesse minimax sur l'intersection de deux boules de Besov, $\mathcal{F} = B_{p_1, \infty}^{s_1}(r) \cap B_{p_2, \infty}^{s_2}(r)$. On voit apparaître trois régimes selon les valeurs de π . Si π est au dessus d'un certain seuil la vitesse est donnée par le régime 'lacunaire' de l'un des espaces de Besov, si π est sous un certain seuil la vitesse est celle du régime 'dense' de l'autre espace. Entre ces deux seuils, une nouvelle vitesse apparaît (qui mélange l'information sur les deux espaces).

• *Un signal avec Chirps.* Dans [HKP99], Hall, Kerkyacharian et Picard développent une méthode de seuillage par bloc pour une perte \mathbf{L}^2 . Ils considèrent une classe de fonctions de régularité Hölder global s_2 avec des singularités isolées de type 'chirps'. Ces singularités sont décrites par un niveau Hölder $s_1 < s_2$ et un niveau de portée $\gamma \in [0, 1]$. Les auteurs précisent des conditions pour que les singularités isolées ne contribuent pas et que la vitesse d'estimation soit donc $s_2/(2s_2 + d)$.

En fait on peut identifier un bord de domaine de Besov qui correspond à l'ensemble fonctionnel décrit dans [HKP99]. Notre méthode permet alors de retrouver

les conditions de [HKP99] (qui correspondent à un cas dense). On peut, en plus, étendre les résultats au cas où les singularités isolées ne sont plus négligeables. Notre méthode donne alors de nouvelles vitesses (qui dépendent de s_1, s_2 et γ).

- *Signaux aléatoires multifractals*. Dans [12] nous déterminons la vitesse d'estimation quand f est la réalisation de certains processus aléatoires multifractals (séries lacunaires aléatoires, cascades).

6.2 Modèle de cascades multiplicatives en asymptotique mixte

Ce travail [13], réalisé avec E. Bacry, M. Hoffmann et J.F Muzy, concerne la statistique des processus de cascades. Initialement, ces processus ont été proposés par B. Mandelbrot [Man74] pour modéliser la dissipation d'énergie en turbulence. Plus récemment, il a été montré que ces modèles décrivent bien certains faits empiriques des séries financières (voir [BP03], [MFC97], [BMD00]).

Avant de présenter nos résultats, rappelons quelques définitions sur ces processus.

Définition des cascades multiplicatives

Dans la suite on notera, $r = (r_1, \dots, r_j) \in \{0, 1\}^j$, un j -uplet de longueur j quelconque. On peut considérer r comme une chemin dans un arbre dyadique et pour $i \leq j$, on notera $r|i$ le chemin raccourci :

$$r|i = (r_1, \dots, r_i).$$

Pour $r \in \{0, 1\}^j$ on note $I_r = [\sum_{i=1}^j r_i 2^{-i}, \sum_{i=1}^j r_i 2^{-i} + 2^{-j}]$ l'intervalle de longueur 2^{-j} , obtenu par j divisions successives de $[0, 1]$.

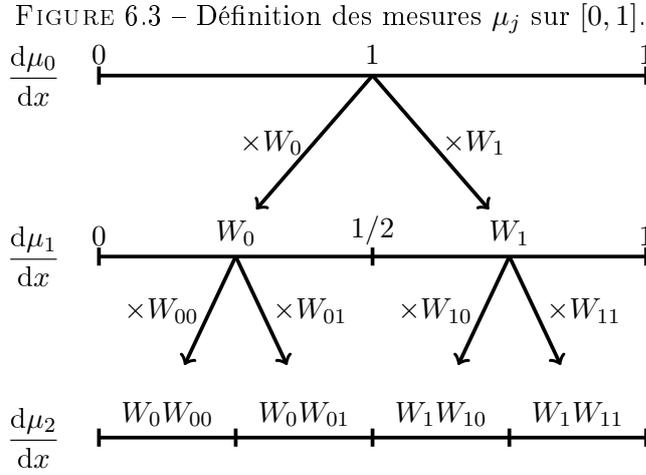
Soit $\{W_r\}_{r \in \{0,1\}^j, j \in \mathbb{N}^*}$ un ensemble de variables i.i.d indexées par tous les r -uplets de longueur finie. On suppose que la loi commune de ces variables vérifie $W \geq 0$, $E[W] = 1$.

Pour tout $j \in \mathbb{N}^*$, on définit une mesure aléatoire μ_j sur $[0, 1]$ dont la dérivée par rapport à la mesure de Lebesgue est constante sur les intervalles dyadiques de longueur 2^{-j} en posant :

$$\frac{d\mu_j}{dx}(x) = \prod_{i=1}^j W_{r|i}, \quad \text{pour } x \in I_r \text{ et tout } r \in \{0, 1\}^j.$$

La figure 6.3 présente comment ces densités sont obtenues par récurrence (en découpant chaque intervalle dyadique en deux et en multipliant par une variable indépendante sur chaque nouvel intervalle). La mesure de cascade μ_∞ est obtenue en passant à la limite $j \rightarrow \infty$. Par exemple, la masse totale est donc :

$$\mu_\infty([0, 1]) = \lim_{j \rightarrow \infty} 2^{-j} \sum_{r \in \{0,1\}^j} \prod_{i=1}^j W_{r|i}.$$



Le processus de cascade est obtenu en intégrant cette mesure :

$$X_t = \mu_\infty([0, t]), \quad \forall t \in [0, 1].$$

Rappelons le résultat de Kahane et Peyrière [KP76] sur l'existence de la mesure limite.

Théorème 6.8 (Kahane et Peyrière). 1) Si $E[W \log_2 W] < 1$ la variable aléatoire $\mu_\infty([0, 1])$ n'est pas presque sûrement nulle.

2) Si on pose $\tau(p) = p - \log_2 E[W^p] - 1$. Pour $p > 1$,

$$\tau(p) > 0 \implies E[\mu_\infty([0, 1])^p] < \infty.$$

Par construction, le processus vérifie des relations d'autosimilarité comme par exemple $X_{1/2} \stackrel{\text{loi}}{=} (W/2)X_1$. On peut alors aisément voir que, si les moments sont finis, ils vérifient la relation exacte :

$$E \left[\left(X_{\frac{k+1}{2^j}} - X_{\frac{k}{2^j}} \right)^p \right] = \left(\frac{1}{2^j} \right)^{\tau(p)+1} E[X_1^p]. \quad (6.17)$$

Ceci explique le comportement multifractal dans le sens où la limite de (6.1) doit dépendre de p au travers de l'exposant $\tau(p)$. Posons

$$S(j, p) = \sum_{k=0}^{2^j-1} \left(X_{\frac{k+1}{2^j}} - X_{\frac{k}{2^j}} \right)^p.$$

On a $E[S(j, p)] = 2^{-j\tau(p)} E[X_1^p]$. Cependant, un des faits les plus importants des processus de cascades est que les moments empiriques n'ont pas le même comportement que leur espérance si $|p|$ est trop grand.

Théorème 6.9 ([Mol96], [OW00], [RSGW03]). *Soit $p \in \mathbb{R}$ alors,*

$$\lim_{j \rightarrow \infty} \frac{\log_2 S(j, p)}{-j} \xrightarrow[p.s.]{} \tau_0(p),$$

où $\tau_0(p)$ est égale à

$$\tau_0(p) = \begin{cases} \tau(p), & \forall p \in (p_0^-, p_0^+) \\ h_0^+ p, & \forall p \geq p_0^+ \\ h_0^- p, & \forall p \leq p_0^- \end{cases} \quad (6.18)$$

avec

$$p_0^+ = \inf\{p \geq 1 \mid p\tau'(p) - \tau(p) \leq 0\} \in (1, \infty], \quad h_0^+ = \tau'(p_0^+), \\ p_0^- = \sup\{p \leq 0 \mid p\tau'(p) - \tau(p) \leq 0\} \in [-\infty, 0), \quad h_0^- = \tau'(p_0^-).$$

Cet effet de linéarisation (6.18) de la loi d'échelle pour $|p|$ grand est classiquement observé sur les données de turbulence (voir par exemple [OW02], [LAC04]).

Cependant, les exposants p_0^\pm où interviennent la linéarisation semble dépendre du rapport entre la longueur totale de la série observée et l' "échelle intégrale" (ou échelle de décorrélation) du phénomène ([MBBP08]). En particulier, il est constaté que la linéarisation apparaît pour des valeurs de $|p|$ plus grandes si le temps d'observation est important.

Un choix d'asymptotique mixte

Pour comprendre cela, dans [13], nous considérons donc un modèle de processus multifractal $(X_t)_{t \geq 0}$ dont l'échelle typique de décorrélation est finie (égale à 1). Le temps total d'observation tend vers $+\infty$.

Pour tout $m \geq 0$, on se donne $(\mu_\infty^{(m)})$ une mesure de cascade sur $[m, m+1]$ de générateur W . Ces cascades sont indépendantes, et on définit un processus $(X_t)_{t \geq 0}$ en intégrant la somme de ces mesures :

$$X_t = \mu_\infty^{(m)}([t, m]) + \sum_{l=0}^{m-1} \mu_\infty^{(l)}([l, l+1]), \quad \forall t \in [m, m+1].$$

Nous considérons les moments empiriques des accroissements, avec pas 2^{-j} , sur un intervalle $[0, N]$, ($N \geq 1$) :

$$S(j, N, p) := \sum_{k=0}^{N2^j-1} \left(X_{\frac{k+1}{2^j}} - X_{\frac{k}{2^j}} \right)^p. \quad (6.19)$$

Par (6.17), on a $E[S(j, N, p)] = N2^{-j\tau(p)} E[X_1^p]$.

Dans [13], nous étudions le comportement de $S(j, N, p)$ dans l'asymptotique mixte $j \rightarrow \infty$ et $N \rightarrow \infty$.

Théorème 6.10. *Supposons que $N \sim_{j \rightarrow \infty} 2^{j\chi}$ avec $\chi > 0$ alors :*

$$\lim_{j \rightarrow \infty} \frac{\log_2 S(j, N, p)}{-j} \xrightarrow[p.s.]{} \tau_\chi(p),$$

où $\tau_\chi(p)$ est égale à

$$\tau_\chi(p) = \begin{cases} \tau(p) - \chi, & \forall p \in (p_\chi^-, p_\chi^+) \\ h_\chi^+ p, & \forall p \geq p_\chi^+ \\ h_\chi^- p, & \forall p \leq p_\chi^- \end{cases}$$

avec

$$p_\chi^+ = \inf\{p \geq 1 \mid p\tau'(p) - \tau(p) \leq 0\} \in (1, \infty], \quad h_\chi^+ = \tau'(p_\chi^+), \\ p_\chi^- = \sup\{p \leq 0 \mid p\tau'(p) - \tau(p) \leq 0\} \in [-\infty, 0), \quad h_\chi^- = \tau'(p_\chi^-).$$

Ce théorème permet de voir que la linéarisation est repoussée pour des valeurs de $|p|$ plus grandes. Ceci confirme les résultats empiriques.

Pour démontrer le théorème 6.10, nous démontrons en fait le résultat plus fort qui est la convergence presque sûr de $2^{-j(\chi+1)}S(j, N, p)$ vers une constante. Nous précisons ensuite cette convergence par l'obtention d'un TCL. La preuve s'inspire de celle du théorème 6.9 dans [RSGW03]. Une des difficultés nouvelles, ici, est que nous avons besoin d'évaluer les moments des mesures μ_j qui apparaissent dans les construction de cascades quand les conditions du théorème 6.8 pour la bornitude de ces moments ne sont plus vraies. Pour exploiter l'indépendance des cascades, qui permet de compenser ces problèmes d'explosion des moments, nous utilisons une inégalité de type interpolation entre L^1 et L^2 pour des sommes de variables indépendantes (voir [vBE65]).

Dans notre travail, nous étendons aussi les résultats de convergence quand on remplace dans (6.19) les accroissements de X par des coefficients dans une base d'ondelettes. Les calculs sont alors plus délicats car la base d'ondelettes n'est pas nécessairement adaptée à la structure dyadique utilisée dans la construction des cascades.

Finalement, nous montrons que l'on peut déduire du théorème 6.10 une caractérisation de type grande déviation de la distribution des accroissements du processus sur $[0, N]$. Pour cela, soit $d(h)$ le spectre multifractal tel que prévu par la conjecture de Frisch-Parisi (6.6), qui s'écrit ici : $d(h) = \inf_{p \in \mathbb{R}} \{ph - \tau(p)\}$.

Théorème 6.11. *Soit $h \in (h_\chi^+, h_\chi^-)$, alors :*

$$\lim_{\varepsilon \rightarrow 0} \lim_j \frac{1}{j} \log \# \left\{ k \in \{0, \dots, N2^j - 1\} \mid 2^{-j(h+\varepsilon)} \leq X_{\frac{k+1}{2^j}} - X_{\frac{k}{2^j}} \leq 2^{-j(h-\varepsilon)} \right\} \\ = \chi + d(h), \quad (6.20)$$

$$\lim_{\varepsilon \rightarrow 0} \overline{\lim}_j \frac{1}{j} \log \# \left\{ k \in \{0, \dots, N2^j - 1\} \mid 2^{-j(h+\varepsilon)} \leq X_{\frac{k+1}{2^j}} - X_{\frac{k}{2^j}} \leq 2^{-j(h-\varepsilon)} \right\} \\ = \chi + d(h). \quad (6.21)$$

Ce résultat permet de dénombrer la quantité de singularités correspondant à un exposant de Hölder h dans le signal. Le théorème 6.11 montre que, en asymptotique mixte, des singularités de régularité plus forte (et plus faible) apparaissent. Elles correspondent aux exposants $h \in [h_\chi^+, h_0^+)$ (resp. $h \in (h_0^-, h_\chi^-]$). Ce résultat permet de comprendre que des estimateurs basés sur la statistique d'ordre des accroissements du processus devraient avoir un comportement dépendant de χ . C'est en particulier le cas des estimateurs utilisés dans l'analyse des valeurs extrêmes. Nous envisageons d'étudier, dans la suite, le comportement asymptotique de ces estimateurs sur des modèles de cascades.

Bibliographie

- [ABM05] T.G. Andersen, T. Bollerslev, and N. Meddahi. Correcting the errors : volatility forecast evaluation using high-frequency data and realized volatilities. *Econometrica*, 73 :279–296, 2005.
- [AdCB⁺96] A. Arneodo, Y. d'Aubenton Cafara, E. Bacry, P.V. Graves, Muzy J.F, and Thermes C. Wavelet based fractal analysis of DNA sequences. *Physica D*, 96(291–320), 1996.
- [AGF95] P. Abry, P. Gonçalvès, and P. Flandrin. Wavelet, spectrum estimation and $1/f$ processes. in A. Antoniadis, G. Oppenheim (Eds), *Wavelets and Statistics, Lectures note in Statistics*, 103 :15–30, 1995.
- [AS02] Y. Aït-Sahalia. Maximum likelihood estimation of discretely sampled diffusions : a closed-form approximation approach. *Econometrica*, 70(1) :223–262, 2002.
- [AS08] Y. Aït-Sahalia. Closed-form likelihood expansions for multivariate diffusions. *Ann. Statist.*, 36(2) :906–937, 2008.
- [ASMZ05] Y. Aït-Sahalia, P. A. Mykland, and L. Zhang. Ultra high frequency volatility estimation with dependent microstructure noise. *Working paper, Princeton University*, 2005.
- [AV98] P. Abry and D. Veitch. Wavelet analysis of long range dependence traffic. *IEEE Trans. Inform. Theory*, 44 :2–15, 1998.
- [BCdL98] F.J. Breidt, N. Crato, and P. de Lima. The detection and estimation of long memory in stochastic volatility. *J. Econom.*, 83 :325–348, 1998.
- [BJS04] B.M. Bibby, M. Jacobsen, and M. Sørensen. Estimating functions for discretely sampled diffusion-type models. *Preprint 2004-4, Department of applied mathematics and statistics, University of Copenhagen. To appear in Ait Sahalia Y. & Hansen L.P. (eds) : Handbook of Financial Econometrics*, 2004.
- [BL96] L.D. Brown and M.G. Low. Asymptotic equivalence of nonparametric regression and white noise. *Ann. Statist.*, 24(6) :2384–2398, 1996.
- [BM83] P. Breuer and P. Major. Central limit theorems for nonlinear functionals of Gaussian fields. *J. Multivariate Anal.*, 13(3) :425–441, 1983.

- [BMD00] E. Bacry, J.F. Muzy, and J. Delour. Modelling fluctuations of financial time series : from cascade processes to stochastic volatility models. *Eur. Phys. J. B.*, 17 :537–548, 2000.
- [BNHLS08] O. E. Barndorff-Nielsen, P. R. Hansen, A. Lunde, and N. Shephard. Designing realised kernels to measure the ex-post variation of equity prices in the presence of noise. *Econometrica*, 2008.
- [BNS04] O. E. Barndorff-Nielsen and N. Shephard. Econometric analysis of realized covariation : high frequency based covariance, regression, and correlation in financial economics. *Econometrica*, 72(3) :885–925, 2004.
- [BP03] J.P. Bouchaud and M. Potters. *Theory of Financial Risk and Derivative Pricing*. Cambridge University Press, Cambridge, 2003.
- [BR96] P. J. Bickel and Y. Ritov. Inference in hidden Markov models. I. Local asymptotic normality in the stationary case. *Bernoulli*, 2(3) :199–228, 1996.
- [BR98] P. J. Bickel, Y. Ritov, and T. Rydén. Asymptotic normality of the maximum-likelihood estimator for general hidden Markov models. *Ann. Statist.*, 26(4) :1614–1635, 1998.
- [BZ02] T. Bollerslev and H. Zhou. Estimating volatility diffusion using conditional moments of integrated volatility. *J. Econom.*, 109 :33–65, 2002.
- [CAKLS92] K.C. Chan, G. Andrew Karolyi, F.A. Longstaff, and A.B. Sanders. An empirical comparison of alternative models of the short-term interest rate. *The journal of finance*, 47(3) :1209–1227, 1992.
- [CD06] V. Corradi and W. Distaso. Semi parametric comparison of stochastic volatility models using realized measures. *Rev. Econom. Stud.*, 73 :635–667, 2006.
- [CKR93] Z. Ciesielski, G. Kerkycharian, and B. Roynette. Quelques espaces fonctionnels associés à des processus gaussiens. *Studia Math.*, 107(2) :171–204, 1993.
- [Coe01] J.F. Coeurjolly. Estimating the parameters of a fractional Brownian motion by discrete variations of its sample paths. *Stat. Inference Stoch. Process.*, 4(2) :199–227, 2001.
- [Coh00] A. Cohen. Wavelet methods in numerical analysis. In *Handbook of numerical analysis, Vol. VII*, Handb. Numer. Anal., VII, pages 417–711. North-Holland, Amsterdam, 2000.
- [CR96] F. Comte and E. Renault. Long memory in continuous-time models. *J. Econom.*, 73 :101–150, 1996.
- [CR98] F. Comte and E. Renault. Long memory in continuous-time stochastic volatility models. *Math. Finance*, 8(4) :291–323, 1998.
- [CR02] N. Crato and B.K. Ray. Semi-parametric smoothing estimators for long-memory processes with added noise. *J. Statist. Plann. Inference*, 105(2) :283–297, 2002.

- [Dah89] R. Dahlhaus. Efficient parameter estimation for self-similar processes. *Ann. Statist.*, 17(4) :1749–1766, 1989.
- [DCD86] D. Dacunha-Castelle and M. Dufflo. *Probability and statistics. Vol. I*. Springer-Verlag, New York, 1986. Translated from the French by David McHale.
- [DCFZ86] D. Dacunha-Castelle and D. Florens-Zmirou. Estimation of the coefficient of a diffusion from discrete observations. *Stochastics*, 19 :263–284, 1986.
- [DJ94a] D.L. Donoho and I.M. Johnstone. Ideal spatial adaptation by wavelet shrinkage. *Biometrika*, 81(3) :425–455, 1994.
- [DJ94b] D.L. Donoho and I.M. Johnstone. Minimax risk over l_p -balls for l_q -error. *Probab. Theory Related Fields*, 99(2) :277–303, 1994.
- [DJ97] S. Delattre and J. Jacod. A central limit theorem for normalized functions of the increments of a diffusion process, in the presence of round-off errors. *Bernoulli*, 3(1) :1–28, 1997.
- [DJ98] D.L. Donoho and I.M. Johnstone. Minimax estimation via wavelet shrinkage. *Ann. Statist.*, 26(3) :879–921, 1998.
- [DJKP95] D.L. Donoho, I.M. Johnstone, G. Kerkyacharian, and D. Picard. Wavelet shrinkage : asymptopia ? *J. Roy. Statist. Soc. Ser. B*, 57(2) :301–369, 1995. With discussion and a reply by the authors.
- [DJKP96] D.L. Donoho, I.M. Johnstone, G. Kerkyacharian, and D. Picard. Density estimation by wavelet thresholding. *Ann. Statist.*, 24(2) :508–539, 1996.
- [DKT98] R. A. DeVore, S. V. Konyagin, and V. N. Temlyakov. Hyperbolic wavelet approximation. *Constr. Approx.*, 14(1) :1–26, 1998.
- [DL91] D.L. Donoho and R.C. Liu. Geometrizing rates of convergence. II, III. *Ann. Statist.*, 19(2) :633–667, 668–701, 1991.
- [DM01] R. Douc and C. Matias. Asymptotics of the maximum likelihood estimator for general hidden Markov models. *Bernoulli*, 7(3) :381–420, 2001.
- [DN90] D.L. Donoho and M. Nussbaum. Minimax quadratic estimation of a quadratic functional. *J. Complexity*, 6(3) :290–323, 1990.
- [Doh87] G. Dohnal. On estimating the diffusion coefficient. *J. Appl. Probab.*, 24(1) :105–114, 1987.
- [Don94] D.L. Donoho. Asymptotic minimax risk for sup-norm loss : solution via optimal recovery. *Probab. Theory Related Fields*, 99(2) :145–170, 1994.
- [DÜ99] L. Decreusefond and A. S. Üstünel. Stochastic analysis of the fractional Brownian motion. *Potential Anal.*, 10(2) :177–214, 1999.
- [DY83] K. O. Dzharidze and A. M. Yaglom. Spectrum parameter estimation in time series analysis. In *Developments in statistics, Vol. 4*, volume 4 of *Develop. in Statist.*, pages 1–96. Academic Press, New York, 1983.

- [FP85] U. Frisch and G. Parisi. On the singularity structure of fully developed turbulence; appendix to fully developed turbulence and intermittency, by u. frisch. In *Proc. Int. Summer School Phys. Enrico Fermi.*, pages 84–88. North-Holland, 1985.
- [Fri95] U. Frisch. *Turbulence*. Cambridge University Press, Cambridge, 1995. The legacy of A. N. Kolmogorov.
- [FT87] R. Fox and M.S. Taqqu. Central limit theorems for quadratic forms in random variables having long-range dependence. *Probab. Theory Related Fields*, 74(2) :213–240, 1987.
- [GBP⁺96] S. Ghashghaie, W. Breymann, J. Peinke, P. Talkner, and Y. Dodge. Turbulent cascades in foreign exchange markets. *Nature*, 381 :767–770, 1996.
- [GC90] V. Genon-Catalot. Maximum contrast estimation for diffusion processes from discrete observations. *Statistics*, 21(1) :99–116, 1990.
- [GCJ93] V. Genon-Catalot and J. Jacod. On the estimation of the diffusion coefficient for multidimensional diffusion processes. *Annales de l’I.H.P., probabilité–Statistiques*, 29 :119–151, 1993.
- [GCJL99] V. Genon-Catalot, T. Jeantheau, and C. Larédo. Parameter estimation for discretely observed stochastic volatility models. *Bernoulli*, 5 :855–872, 1999.
- [GCJL00] V. Genon-Catalot, T. Jeantheau, and C. Larédo. Stochastic volatility as hidden markov models. *Bernoulli*, 6 :1051–1079, 2000.
- [GL89] X. Guyon and J. León. Convergence en loi des H -variations d’un processus gaussien stationnaire sur \mathbf{R} . *Ann. Inst. H. Poincaré Probab. Statist.*, 25(3) :265–282, 1989.
- [GLT96] Y.K. Golubev, B.Y. Levit, and A.B. Tsybakov. Asymptotically efficient estimation of analytic functions in Gaussian noise. *Bernoulli*, 2(2) :167–181, 1996.
- [GN90] G.K. Golubev and M. Nusbaum. Nonparametric estimation of a regression function in L^2 . *Problemy Peredachi Informatsii*, 26(3) :38–49, 1990.
- [GN92] G. K. Golubev and M. Nussbaum. Adaptive spline estimates in a nonparametric regression model. *Teor. Veroyatnost. i Primenen.*, 37(3) :554–561, 1992.
- [GN98] I. Grama and M. Nussbaum. Asymptotic equivalence for nonparametric generalized linear models. *Probab. Theory Related Fields*, 111(2) :167–214, 1998.
- [Gob01] E. Gobet. Local asymptotic mixed normality property for elliptic diffusion : a Malliavin calculus approach. *Bernoulli*, 7(6) :899–912, 2001.
- [Gob02] E. Gobet. LAN property for ergodic diffusions with discrete observations. *Ann. Inst. H. Poincaré Probab. Statist.*, 38(5) :711–737, 2002.

- [GT99] G. Gayraud and K. Tribouley. Wavelet methods to estimate an integrated quadratic functional : adaptivity and asymptotic law. *Statist. Probab. Lett.*, 44(2) :109–122, 1999.
- [HKP99] P. Hall, G. Kerkycharian, and D. Picard. On the minimax optimality of block thresholded wavelet estimators. *Statist. Sinica*, 9(1) :33–49, 1999.
- [Hof02] M. Hoffmann. Rate of convergence for parametric estimation in a stochastic volatility model. *Stochastic Process. Appl.*, 97(1) :147–170, 2002.
- [HS97] F. Hirsch and S. Song. Criteria of positivity for the density of the law of a Wiener functional. *Bull. Sci. Math.*, 121(4) :261–273, 1997.
- [HS98] F. Hirsch and S. Song. Properties of the set of positivity for the density of a regular Wiener functional. *Bull. Sci. Math.*, 122(1) :1–15, 1998.
- [IH77] I. A. Ibragimov and R. Z. Has'minskiĭ. Estimation of infinite-dimensional parameter in Gaussian white noise. *Dokl. Akad. Nauk SSSR*, 236(5) :1053–1055, 1977.
- [IH78] I. A. Ibragimov and R. Z. Has'minskiĭ. Capacity in the transmission of smooth signals. *Dokl. Akad. Nauk SSSR*, 242(1) :32–35, 1978.
- [IK76] I.A. Ibragimov and R.Z. Khas'minskiĭ. Some problems of statistical data processing in communications. In *Proceedings of the IEEE-USSR Joint Workshop on Information Theory (Moscow, 1975)*, pages 86–92. Inst. Electr. Electron. Engrs., New York, 1976.
- [IK01] S. M. Iacus and Yu. A. Kutoyants. Semiparametric hypotheses testing for dynamical systems with small noise. *Math. Methods Statist.*, 10(1) :105–120, 2001.
- [IL97] J. Istas and G. Lang. Quadratic variations and estimation of the local Hölder index of a Gaussian process. *Ann. Inst. H. Poincaré Probab. Statist.*, 33(4) :407–436, 1997.
- [Jac97] J. Jacod. On continuous conditional gaussian martingales and stable convergence in law, séminaire proba XXXI. Springer-Verlag, Berlin. *Lecture Notes in Math*, 1655 :232–246, 1997.
- [Jaf97] S. Jaffard. Multifractal formalism for functions. I. Results valid for all functions. *SIAM J. Math. Anal.*, 28(4) :944–970, 1997.
- [Jaf00] S. Jaffard. On the Frisch-Parisi conjecture. *J. Math. Pures Appl. (9)*, 79(6) :525–552, 2000.
- [Jeg82] P. Jeganathan. On the asymptotic theory of estimation when the limit of the log-likelihood ratios is mixed normal. *Sankhyā Ser. A*, 44(2) :173–212, 1982.
- [Jeg83] P. Jeganathan. Some asymptotic properties of risk functions when the limit of the experiment is mixed normal. *Sankhyā Ser. A*, 45(1) :66–87, 1983.
- [Kes97] M. Kessler. Estimation of an ergodic diffusion from discrete observations. *Scandinavian journal of statistics*, 24(2) :211–229, 1997.

- [KH03] A. Kohatsu-Higa. Lower bounds for densities of uniformly elliptic random variables on Wiener space. *Probab. Theory Related Fields*, 126(3) :421–457, 2003.
- [Kol62] A.N. Kolmogorov. A refinement of previous hypotheses concerning the local structure of turbulence in a viscous incompressible fluid at high Reynolds number. *J. Fluid Mech.*, 13 :82–85, 1962.
- [KP76] J.P. Kahane and J. Peyrière. Sur certaines martingales de benoît mandelbrot. *Adv. in Mathematics*, 22 :131–145, 1976.
- [KP00] G. Kerkyacharian and D. Picard. Thresholding algorithms, maxisets and well-concentrated bases. *Test*, 9(2) :283–344, 2000. With comments, and a rejoinder by the authors.
- [KP02] G. Kerkyacharian and D. Picard. Minimax or maxisets? *Bernoulli*, 8(2) :219–253, 2002.
- [Kut94] Yu. Kutoyants. *Identification of dynamical systems with small noise*, volume 300 of *Mathematics and its Applications*. Kluwer Academic Publishers Group, Dordrecht, 1994.
- [Kut98] Yu. A. Kutoyants. Semiparametric estimation for dynamical systems with small noise. *Math. Methods Statist.*, 7(4) :457–465 (1999), 1998.
- [LAC04] B. Lashermes, P. Abry, and P. Chainais. New insights into the estimation of scaling exponents. *Int. J. of Wavelets, Multiresolution and Inf. Proc.*, 2(497–523), 2004.
- [LC86] L. Le Cam. *Asymptotic methods in statistical decision theory*. Springer Series in Statistics. Springer-Verlag, New York, 1986.
- [Leb97] B. Leblanc. *Modélisation de la volatilité d'un actif financier et applications*. PhD thesis, Université Paris 7, 1997.
- [LeC73] L. LeCam. Convergence of estimates under dimensionality restrictions. *Ann. Statist.*, 1 :38–53, 1973.
- [Lep90] O.V. Lepskii. A problem of adaptive estimation in Gaussian white noise. *Teor. Veroyatnost. i Primenen.*, 35(3) :459–470, 1990.
- [Lep91] O. V. Lepskii. Asymptotically minimax adaptive estimation. I. Upper bounds. Optimally adaptive estimates. *Teor. Veroyatnost. i Primenen.*, 36(4) :645–659, 1991.
- [Lep92a] O. V. Lepskii. On problems of adaptive estimation in white Gaussian noise. In *Topics in nonparametric estimation*, volume 12 of *Adv. Soviet Math.*, pages 87–106. Amer. Math. Soc., Providence, RI, 1992.
- [Lep92b] O.V. Lepskii. Asymptotically minimax adaptive estimation. II. Schemes without optimal adaptation. Adaptive estimates. *Teor. Veroyatnost. i Primenen.*, 37(3) :468–481, 1992.
- [Ler92] G. Leroux. Maximum-likelihood estimation for hidden Markov models. *Stochastic Process. Appl.*, 40(1) :127–143, 1992.

- [LLR06] J.W. Lee, E.L. Lee, and P.E. Rikvold. Multifractal behavior of the korean stock-market index kospi. *Physica A : Statistical Mechanics and its Applications*, 364 :355–361, 2006.
- [LS95] F. Longstaff and E. Schwartz. A simple approach to valuing risky fixed and floating rate debt. *The journal of finance*, 1 :789–819, 1995.
- [Man74] B.B. Mandelbrot. Intermittent turbulence in self-similar cascades : divergence of high moments and dimension of the carrier. *Journal of Fluid Mechanics*, 62 :331–358, 1974.
- [Man97] B.B. Mandelbrot. *Fractals and scaling in finance*. Selected Works of Benoit B. Mandelbrot. Springer-Verlag, New York, 1997. Discontinuity, concentration, risk, Selecta Volume E, With a foreword by R. E. Gomory.
- [MBBP08] J.F. Muzy, E. Bacry, R. Baile, and P. Poggi. Uncovering latent singularities from multifractal scaling law in mixed asymptotic regime. application to turbulence. *submitted to Euro Physics Letters*, 2008.
- [MFC97] B.B. Mandelbrot, A. Fisher, and L. Calvet. A multifractal model of asset returns. Cowles Foundation Discussion Paper, 1164, 1997.
- [Mol96] G.M. Molchan. Scaling exponents and multifractal dimensions for independent random cascades. *Comm. in Math. Phys.*, 179 :681–702, 1996.
- [MST99] Y. Meyer, F. Sellan, and M.S. Taqqu. Wavelets, generalized white noise and fractional integration : the synthesis of fractional Brownian motion. *J. Fourier Anal. Appl.*, 5(5) :465–494, 1999.
- [Nua95] D. Nualart. *The Malliavin calculus and related topics*. Probability and its Applications (New York). Springer-Verlag, New York, 1995.
- [Nus85] M. Nussbaum. Spline smoothing in regression models and asymptotic efficiency in L_2 . *Ann. Statist.*, 13(3) :984–997, 1985.
- [Nus96] M. Nussbaum. Asymptotic equivalence of density estimation and Gaussian white noise. *Ann. Statist.*, 24(6) :2399–2430, 1996.
- [OBB⁺06] M. Overhaus, A. Bermúdez, H. Buehler, A. Ferraris, C. Jordinson, and A. Lamnouar. *Equity hybrid derivatives*. Wiley finance series., 2006.
- [OW00] M. Ossiander and E.C. Waymire. Statistical estimation for multiplicative cascades. *The Ann. of Stat.*, 28 :1533–1560, 2000.
- [OW02] M. Ossiander and E.C. Waymire. On estimation theory for multiplicative cascades. *Sankhyā Ser. A*, 64(2) :323–343, 2002. Selected articles from San Antonio Conference in honour of C. R. Rao (San Antonio, TX, 2000).
- [PW99] K. Park and W. Willinger. *Self-similar network traffic : An overview*. In K.Park and W.Willinger, editors, *Self-similar Network Traffic and Performance Evaluation*. Wiley interscience, 1999.
- [Rén63] A. Rényi. On stable sequences of events. *Sankhyā Ser. A*, 25 :293–302, 1963.

- [Riv05] V. Rivoirard. Bayesian modeling of sparse sequences and maxisets for Bayes rules. *Math. Methods Statist.*, 14(3) :346–376, 2005.
- [Rob01] P.M. Robinson. The memory of stochastic volatility models. *J. Econom.*, 101 :195–218, 2001.
- [Ros07] M. Rosenbaum. *Etude de quelques problèmes d'estimation statistique en finance*. PhD thesis, Université Paris-Est, 2007.
- [RSGW03] S. Resnick, G. Samorodnitsky, A. Gilbert, and W. Willinger. Wavelet analysis of conservative cascades. *Bernoulli*, 9(1) :97–135, 2003.
- [RY99] D. Revuz and M. Yor. *Continuous martingales and Brownian motion*, volume 293 of *Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]*. Springer-Verlag, Berlin, third edition, 1999.
- [Ryd94] T. Ryden. Consistent and asymptotically normal parameter estimates for hidden Markov models. *The annals of statistics*, 22(4) :1884–1895, 1994.
- [SU03] M. Sørensen and M. Uchida. Small-diffusion asymptotics for discretely sampled stochastic differential equations. *Bernoulli*, 9(6) :1051–1069, 2003.
- [TK92] A. H. Tewfik and M. Kim. Correlations structure of the discrete wavelet coefficients of fractional Brownian motion. *IEEE Trans. Inform. Theory*, 38(2, part 2) :904–909, 1992.
- [Tri00] K. Tribouley. Adaptive estimation of integrated functionals. *Math. Methods Statist.*, 9(1) :19–38, 2000.
- [Uch04] M. Uchida. Estimation for discretely observed small diffusions based on approximate martingale estimating functions. *Scand. J. Statist.*, 31(4) :553–566, 2004.
- [vBE65] B. von Bahr and K.G. Esseen. Inequalities for the r th absolute moment of a sum of random variables, $1 \leq r \leq 2$. *Ann. Math. Stat.*, 36 :299–303, 1965.
- [YAL01] Z.G. Yu, V. Anh, and K.S. Lau. Multifractal characterisation of length sequences of coding and noncoding segments in a complete genome. *Physica A*, 301 :351–361, 2001.
- [Yos92] N. Yoshida. Estimation for diffusion processes from discrete observation. *J. Multivariate Anal.*, 41(2) :220–242, 1992.
- [Zha06] L. Zhang. Efficient estimation of stochastic volatility using noisy observations : a multi-scale approach. *Bernoulli*, 12(6) :1019–1043, 2006.
- [ZMAS05] L. Zhang, P. A. Mykland, and Y. Aït-Sahalia. A tale of two time scales : determining integrated volatility with noisy high-frequency data. *J. Amer. Statist. Assoc.*, 100(472) :1394–1411, 2005.