C. R. Acad. Sci. Paris, t. 324, Série I, p. 475-480, 1997
Statistiques/*Statistics*

Estimation non-paramétrique du coefficient de diffusion pour une perte L_p

Marc HOFFMANN

Université de Paris-VII, UFR Mathématiques,
Couloir 45-55, 5^e étage, 2, place Jussieu, 75251 Paris CEDEX 5.

Résumé. On étudie le problème de l'estimation fonctionnelle du coefficient de diffusion, dépendant de la variable d'espace, sous une contrainte de régularité de type Besov et pour une perte L_p. On dispose d'une trajectoire solution discrétisée sur l'intervalle de temps $[0, 1]$. On montre que sous une condition de minoration du temps local du processus observé, la vitesse minimax est $n^{-sp/(1+2s)}$ comme dans les modèles classiques d'estimation de la densité ou de la régression. On propose un estimateur par ondelettes qui atteint la vitesse optimale.

Nonparametric estimation of the diffusion coefficient in L_p-loss

Abstract. *We study the functional estimation of the space dependent diffusion coefficient, for a global L_p-loss over Besov spaces, when the sample path is discretely observed on the time interval $[0, 1]$. We show that under a suitable assumption on the local time of the diffusion process, the minimax rate of convergence is $n^{-sp/(1+2s)}$ as in classical situations, such as density estimation or nonparametric regression. We propose a wavelet estimator achieving the optimal rate of convergence.*

Abridged English Version

This Note investigates the functional estimation in the minimax framework of the diffusion coefficient $\sigma(x)$ in the 1-dimensional model

$$dX_t = b(t, X_t)dt + \sigma(X_t)dW_t, \quad X_0 = x_0, \quad t \in [0, 1], \tag{1}$$

where $(W_t, 0 \le t \le 1)$ is a Wiener process, x_0 is fixed and the functions $\sigma(x)$ and $b(t, x)$ are unknown.

The sample path $(X_t, 0 \le t \le 1)$ is observed at equidistant times i/n, with $i = 0, 1, \cdots, n$. Our aim is to recover σ^2 on a given interval $[m, M]$. We assume that

[H0] *Equation* (1) *admits a strong solution,*

[H1] *The diffusion coefficient σ^2 is Lipschitz continuous and, there exist constants M_0 et M_1 such that $\forall x \in \mathbb{R} : 0 < M_0 \le \sigma^2(x) \le M_1$,*

Note présentée par Paul DEHEUVELS.

[H2] *The function* σ^2 *belongs to the ball of radius* M_2 *of the Besov space* $B_{sp\infty}$ *restricted to the interval* $[m, M]$, *with* $s > 1 + 1/p$ *and* $1 \leq p < \infty$,

[H3] *The drift coefficient* b *is continuous and there exists a constant* M_3 *such that*

$$\forall (t,x) \in [0,1] \times \mathbb{R} \ : \ b^2(t,x) \leq M_3^2(1+x^2).$$

We denote by V_{sp} the functional constraint on the coefficients b and σ^2 given by [H0]-[H3]. In order to achieve global rates of convergence on a fixed domain D, we study the minimax risk *conditionally on the event* $\inf_{x\in[m,M]} L^x \geq \nu > 0$, where $L^x = \lim_{\epsilon\to 0} \frac{1}{2\epsilon}\int_0^1 1_{|X_s - x|\leq\epsilon}\, ds$ is the local time of X at x. We define

$$R_{n,\nu}(\hat{\sigma}^2, V_{sp}) = \sup_{(\sigma^2,b)\in V_{sp}} E_{\sigma,b}\left(\int_{[m,M]} \left|\hat{\sigma}^2(x) - \sigma^2(x)\right|^p dx \;\middle|\; \inf_{x\in[m,M]} L^x \geq \nu \right).$$

THEOREM. – *Let* $\mathcal{E}$ *denote the set of all estimators. Then*

$$C_1\, n^{-sp/(1+2s)} \leq \inf_{\hat{\sigma}^2 \in \mathcal{E}} R_{n,\nu}(\hat{\sigma}^2, V_{sp}) \leq C_2\, n^{-sp/(1+2s)}$$

where C_1 *and* C_2 *are two positive constants depending on* s, p *and* M_i, $i = 0, \cdots, 3$. *Moreover, the estimator given in definition 2 below achieves the optimal rate of convergence.*

1. Introduction

On observe, aux instants i/n, $i = 0, 1, \cdots, n$, le processus de diffusion uni-dimensionnel, solution de

$$dX_t = b(t, X_t)dt + \sigma(X_t)dW_t, \quad X_0 = x_0, \quad t \in [0,1] \tag{1}$$

où $(W_t, 0 \leq t \leq 1)$ est un processus de Wiener, x_0 est fixé et les fonctions $\sigma(x)$ et $b(t,x)$ sont inconnues. On désigne par $(\mathcal{F}_t)$ la filtration naturelle de X. On étudie le problème de l'estimation minimax du coefficient de diffusion σ^2 pour une perte L_p sur un intervalle $[m, M]$ fixé. On suppose que:

[H0] *L'équation (1) admet une solution forte,*

[H1] *La fonction* σ^2 *est Lipschitzienne et il existe* M_0 *et* M_1 *tels que* $\forall x \in \mathbb{R} : 0 < M_0 \leq \sigma^2(x) \leq M_1$.

[H2] *Sur* $[m, M]$, σ^2 *appartient à la boule de rayon* M_2 *de l'espace de Besov* $B_{sp\infty}$ *restreint à l'intervalle* $[m, M]$, *avec* $s > 1 + 1/p$ *et* $1 \leq p < \infty$,

[H3] *Le coefficient de dérive* b *est continu et il existe une constante* M_3 *telle que*

$$\forall (t,x) \in [0,1] \times \mathbb{R} \quad b^2(t,x) \leq M_3^2(1+x^2).$$

Dans toute la suite, on désignera par V_{sp} la contrainte fonctionnelle sur le couple (b, σ^2) imposée par les conditions [H0]-[H3]. La difficulté majeure du modèle défini par (1) est liée au fait que l'on ne peut estimer σ^2 que sur un domaine suffisamment visité par le processus d'observation X. Cette notion est fournie asymptotiquement par le temps local de X :

$$L^x = \lim_{\epsilon\to 0} \frac{1}{2\epsilon}\int_0^1 1_{|X_s - x|\leq\epsilon}\, ds.$$

Puisque l'application $x \to L^x$ est *p.s.* continue, nous pouvons définir $L_{[m,M]} = \inf_{x\in[m,M]} L^x$.

DÉFINITION 1. – *Le risque L_p sous la contrainte V_{sp} d'un estimateur $\hat\sigma^2$ au niveau $\nu > 0$ est*

$$R_{n,\nu}(\hat\sigma^2, V_{sp}) = \sup_{(\sigma^2,b)\in V_{sp}} E_{\sigma,b}\left(\int_{[m,M]} |\hat\sigma^2(x) - \sigma^2(x)|^p dx \,\middle|\, L_{[m,M]} \geq \nu\right),$$

où $P_{\sigma,b}$ désigne la loi de X solution de (1).

Dans cette Note, nous montrons que la vitesse minimax associée au risque $R_{n,\nu}$ est $n^{-sp/(1+2s)}$. Nous construisons un estimateur à l'aide d'un noyau d'ondelettes sur l'intervalle $[m, M]$ qui permet d'atteindre la vitesse optimale. Les résultats obtenus sont comparables aux résultats classiques de l'estimation non-paramétrique linéaire (*voir* [5], [6]).

2. Borne inférieure

PROPOSITION 1. – *Soit $\mathcal{E}$ l'ensemble de tous les estimateurs. Sous* [H0]-[H3] *et pour $p \in [1, +\infty[$, il existe une constante C_1 dépendant de s, p, M_i, $i = 0, 1, 2$ et ν telle que*

$$\inf_{\hat\sigma^2\in\mathcal{E}} R_{n,\nu}(\hat\sigma^2, V_{sp}) \geq C_1\, n^{-sp/(1+2s)}.$$

La démonstration de la proposition 1 utilise un schéma classique (*voir* [5] et [6]). On évalue le risque minimax sur un cube paramétrique $C_j(\gamma)$ à 2^{2^j} éléments inclus dans V_{sp} :

$$C_j(\gamma) = \{\sigma^2 = G + \gamma \sum_k \epsilon_k \psi_{jk},\ \epsilon_k = \pm 1\},$$

où j et γ varient avec l'asymptotique n, G est une constante telle que $M_0 < G < M_1$, ψ est une ondelette à support compact et $\psi_{jk} = 2^{j/2}\psi(2^j x - k)$. On associe une composante de dérive $b = 0$ aux points de $C_j(\gamma)$. Soient σ_-^2 et σ_+^2 deux points génériques de $C_j(\gamma)$ qui ne diffèrent que sur une composante ψ_{jk}. Le point crucial de la preuve est de montrer l'existence de $\lambda > 0$ et $p_0 > 0$, indépendants de n, tels que

$$P_{\sigma_-,0}(\Lambda(\sigma_+^2, \sigma_-^2) > e^{-\lambda}) \geq p_0,$$

où $\Lambda(\sigma_+^2, \sigma_-^2)$ désigne le rapport des vraisemblances des lois induites par σ_+^2 et σ_-^2. Pour cela, on approche la densité de transition du processus X par la densité gaussienne

$$p_t(x, y) = \frac{1}{\sqrt{2\pi t}}\, \frac{1}{\sigma(x)} \exp\left(-\frac{1}{2t}\, \frac{(y - x)^2}{\sigma^2(x)}\right),$$

et on utilise des résultats d'approximation de la densité de transition d'une diffusion en temps petit (*voir* [2]).

3. Borne supérieure

D. Florens-Zmirou a proposé dans [3] l'estimateur

$$\hat\sigma_0^2(x) = \frac{\sum_{i=0}^n 1_{|X_{i/n} - x| \leq h_n} (X_{(i+1)/n} - X_{i/n})^2}{\frac{1}{n}\sum_{i=0}^n 1_{|X_{i/n} - x| \leq h_n}} \tag{2}$$

avec $h_n \to 0$ et $nh_n \to \infty$. Lacoste (*voir* [7]) a étudié la déviation du risque L_2 (pondérée par une approximation du temps local de X) de $\hat\sigma_0^2$. Dans notre cadre, il semble difficile de montrer que $\hat\sigma_0^2$

est optimal pour le risque $R_{n,\nu}$, précisément à cause de l'approximation de L^x au dénominateur de (2) qui possède des mauvaises propriétés de régularité. En effet, la fonction $x \to L^x$ n'appartient pas *p.s.* à $B_{sp\infty}$ pour $s > 1/2$.

3.1. Contruction d'un estimateur optimal pour $R_{n,\nu}$

Nous construisons un estimateur basé sur une approximation de « l'horloge interne » de la diffusion, fournie par la variation quadratique de X, ce qui nous permet d'aborder le problème sous l'angle de la régression. Dans la suite, nous supposerons pour simplifier la présentation que $M - m = 1$ et $b = 0$ (on passe au cas général à l'aide de l'hypothèse [H3]). On écrit

$$(3) \qquad Y_{i/n} = n(X_{(i+1)/n} - X_{i/n})^2 = n\int_{i/n}^{(i+1)/n} \sigma^2(X_s)ds + \varepsilon_{i/n}, \quad i = 1, \cdots, n,$$

où

$$\varepsilon_i = n\left(\int_{i/n}^{(i+1)/n} \sigma(X_s)\, dW_s\right)^2 - n\int_{i/n}^{(i+1)/n} \sigma^2(X_s)\, ds.$$

Les $\varepsilon_{i/n}$ sont des accroissements de martingale de $(\mathcal{F}_{(i+1)/n})$ que l'on peut considérer comme des termes de « bruit » pour l'estimation de $n\int_{i/n}^{(i+1)/n} \sigma^2(X_s)ds \simeq \sigma^2(X_{i/n})$ à partir de l'observation $Y_{i/n}$.

3.1.1. *Construction d'un sous-échantillonnage des observations*

On partitionne le domaine $[m, M]$ en h_n^{-1} « boîtes » identiques $C_\lambda = [m + (\lambda - 1)h_n, m + \lambda h_n[$, $\lambda = 1, \cdots, h_n^{-1}$, où h_n est un seuil à choisir convenablement. Soit $N_{i/n}^\lambda = (\sum_{j \le i} 1_{X_{j/n} \in C_\lambda}) \wedge \lfloor nh_n\nu \rfloor$ le compteur associé à C_λ arrêté lorsque exactement $\lfloor nh_n\nu \rfloor$ observations saturent C_λ. Les propriétés de convergence du temps local empirique vers le temps local (Proposition 3 ci-dessous) assurent que sur l'événement $(L_{[m,M]} \ge \nu)$, le nombre d'observations par boîte est asymptotiquement supérieur ou égal à $\lfloor nh_n\nu \rfloor$.

On construit la suite des $\lfloor nh_n\nu \rfloor$ premiers temps d'atteinte du processus $(X_{i/n})$, $i = 0, \cdots, n$ dans chaque C_λ :

$$T_1 = 0, \quad \text{et pour } i \ge 2\ :\ T_i = \inf\{j/n > T_{i-1}\ :\ \sum_\lambda (N_{j/n}^\lambda - N_{T_{i-1}}^\lambda) \ge 1\} \wedge 1.$$

Les $T_i, i = 1, \cdots, \lfloor n\nu \rfloor$ constituent une suite croissante de temps d'arrêts de $(\mathcal{F}_t)$. Le processus « de bruit » $\sum_{j \le i} \varepsilon_{T_j}$ est une $(\mathcal{F}_{T_i + 1/n})$-martingale. Considérons le sous-échantillonnage $(X_{T_1}, \cdots, X_{T_{\lfloor n\nu \rfloor}})$ obtenu à partir de la données initiale $(X_0, X_{1/n}, \ldots, X_1)$. On applique la transformée en ondelettes empirique aux observations $(X_{T_i}, Y_{T_i}, i = 1, \cdots, \lfloor n\nu \rfloor)$, avec

$$Y_{T_i} = n(X_{T_i + 1/n} - X_{T_i})^2$$

comme pour une régression classique avec des données équiréparties. L'erreur de localisation des points d'observation X_{T_i} par rapport à un schéma équiréparti est contrôlée par le seuil h_n.

3.1.2. *Transformée en ondelettes empirique*

Soit φ une fonction d'échelle à support compact r-fois différentiable, avec $r > s$ (*voir* [8]). On note $\varphi_{jk}(x) = 2^{j/2}\varphi(2^j x - k)$ et $\alpha_{jk} = \int \sigma^2 \varphi_{jk}$ le coefficient d'ondelette associé a σ^2.

DÉFINITION 2. – L'estimateur de σ^2 au niveau ν est

$$\hat{\sigma}^2 = \sum_{k\in\mathbf{Z}} \hat{\alpha}_{j_n,k}\varphi_{j_n,k} \quad \text{avec} \quad \hat{\alpha}_{j_n,k} = \frac{1}{\lfloor n\nu\rfloor}\sum_{i=0}^{\lfloor n\nu\rfloor} Y_{T_i}\,\varphi_{j_n,k}(x_{T_i}).$$

On note x_{T_i} le point du réseau équiréparti $(m + i/\lfloor n\nu\rfloor, i = 1,\cdots\lfloor n\nu\rfloor)$ défini par

$$x_{T_i} = m + (\lambda_{T_i} - 1)h_n + l_{T_i}/\lfloor n\nu\rfloor,$$

où λ_{T_i} est l'indice de la boîte C_λ dans laquelle tombe X_{T_i} et où $l_{T_i} = \#\{X_{T_j} \in C_{\lambda_{T_i}}, j \le i\}$.

3.1.3. *Résultats*

PROPOSITION 2. – *Soit $p \in [1, +\infty[$. Soit σ^2 l'estimateur de la définition 2 construit au niveau $\nu/2$. Sous* [H0]-[H3], *si $2^{j_n} \simeq n^{1/(1+2s)}$ et $h_n \simeq n^{-s/(1+2s)}$, alors*

$$R_{n,\nu}(\hat{\sigma}^2, V_{sp}) \le C_2\, n^{-sp/(1+2s)}$$

où C_2 est une constante dépendant de φ, s, p, M_i, $i = 0,\cdots,3$ et ν.

PROPOSITION 3. – *Pour $\gamma \ge 2$, il existe C_3, dépendant de s, p, M_0, M_1, M_3 et γ, tel que*

$$\sup_{x\in D} E_{\sigma,0}(|L_n^x - L^x|^\gamma) \le C_3\left\{h_n^{\gamma/2} + \left(\frac{1}{nh_n^2}\right)^\gamma\right\},$$

où $L_n^x = \dfrac{1}{nh_n}\displaystyle\sum_{i=0}^{n}\phi\big(h_n^{-1}(X_{i/n} - x)\big)$ désigne le temps local empirique de X relativement à une fonction régulière et localisée ϕ telle que $\int\phi = 1$.

Esquisse de la démonstration de la proposition 2. On a $R_{n,\nu} \le C(B_n + V_n)$, où

$$V_n = E_{\sigma,0}\left(\int_{[m,M]} |\sum_k(\hat{\alpha}_{j_nk} - \alpha_{j_nk})\varphi_{j_nk}|^p 1_{L_{[m,M]}\ge\nu}\right)$$

et

$$B_n = \|\sigma^2 - \sum_k \alpha_{j_nk}\varphi_{j_nk}\|_p^p.$$

Le terme B_n est de l'ordre de $2^{-j_n sp}$ d'après [H2]. On introduit la pénalisation $P_{j,k} = \prod_{\lambda\in C_{j,k}} 1_{L_n^{x_\lambda}\ge\nu}$, avec $C_{j,k} = \{\lambda : C_\lambda \cap S_{jk} \ne \vee\}$, où S_{jk} désigne le support de φ_{jk} et x_λ le point milieu de C_λ. En appliquant successivement le lemme de Meyer (*voir* [8], p. 30) et l'inégalité de Cauchy-Schwarz, on obtient

$$\begin{aligned} V_n &\le C2^{j_n(p/2-1)}\sum_k E_{\sigma,0}\big(|\hat{\alpha}_{j_nk} - \alpha_{j_nk}|^p\, 1_{L_{[m,M]}\ge\nu}\big) \\ &\le C\Bigg[2^{j_n(p/2-1)}\sum_k E_{\sigma,0}\big(|\hat{\alpha}_{j_nk} - \alpha_{j_nk}|^{2p}\, P_{j_nk}\big)^{1/2}\cdots \\ &\qquad \cdots + 2^{j_np/2}\sup_{x\in[m,M]} P_{\sigma,0}(|L_n^x - L^x| \ge \nu/2)\Bigg] \end{aligned}$$

Le dernier terme de l'inégalité est asymptotiquement négligeable d'après l'inégalité de Chebyshev et la Proposition 3. Pour le premier terme, on écrit

$$\hat{\alpha}_{j_n,k} - \alpha_{j_n,k} = Q_1 + Q_2 + Q_3,$$

avec

$$Q_1 = \frac{1}{\lfloor n\nu/2 \rfloor} \sum_{i=1}^{\lfloor n\nu/2 \rfloor} \sigma^2(X_{T_i})\varphi_{j_n k}(x_{T_i}) - \int \sigma^2 \varphi_{j_n k},$$

$$Q_2 = \frac{1}{\lfloor n\nu/2 \rfloor} \sum_{i=1}^{\lfloor n\nu/2 \rfloor} n \int_{T_i}^{T_i+1/n} \left(\sigma^2(X_{T_i}) - \sigma^2(X_s)\right)\varphi_{j_n k}(x_{T_i})ds,$$

$$Q_3 = \frac{1}{\lfloor n\nu/2 \rfloor} \sum_{i=1}^{\lfloor n\nu/2 \rfloor} \varepsilon_{T_i}\varphi_{j_n k}(x_{T_i}).$$

Le terme Q_1 est traité par la régularité de σ^2 et la pénalisation $P_{j,k}$ (en choisissant convenablement ϕ) qui assure que l'on dispose de suffisamment de points X_{T_i} pour procéder à l'approximation. On utilise l'inégalité de Burckholder-Davis-Gundy pour Q_2. Enfin, le terme de bruit Q_3 est traité par une version martingale de l'inégalité de Rosenthal (*voir* [4], p 23.) qui peut s'appliquer ici car les T_i sont des temps d'arrêt de $(\mathcal{F}_t)$.

On montre finalement que V_n est de l'ordre de $(2^{j_n}/n)^{p/2}$. En équilibrant cette quantité avec $2^{-j_n sp}$, on obtient la proposition 2.

Note remise le 24 mai 1996, acceptée après révision le 22 octobre 1996.

Références bibliographiques

[1] **Dacunha-Castelle D. et Florens-Zmirou D., 1986.** Estimation of the coefficients of a diffusion from discrete observations. *Stochastics*, 19, p. 263-284.

[3] **Florens-Zmirou D., 1993.** On estimating the diffusion coefficient from discrete observations. *J. Appl. Probab.*, 30, p. 790-804.

[4] **Hall P. et Heyde C.C., 1980.** *Martingale Limit Theory and its Application.* Academic Press, New York.

[5] **Kerkyacharian et Picard D., 1993.** Density estimation by kernel and wavelets methods : Optimality over Besov spaces. *Statist. Probab. Lett.*, 18, p. 327-336.

[6] **Korostelev A. P. et Tsybakov A. D., 1993.** *Minimax theory of image reconstruction.* Lecture Notes in Statist., vol 82. Springer, Berlin.

[7] **Lacoste V., 1991.** Propriétés de l'erreur quadratique intégrée des estimateurs à noyau du terme de variance d'une diffusion. *C.R. Acad. Sci. Paris*, 313, Série I, p. 317-320.

[8] **Meyer Y., 1990.** *Ondelettes et Opérateurs I.* Hermann, Paris, 1990.