

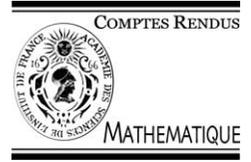


ELSEVIER

Available online at www.sciencedirect.com

SCIENCE @ DIRECT®

C. R. Acad. Sci. Paris, Ser. I 337 (2003) 675–678



Statistique/Probabilités

Estimation adaptative de la densité avec données échantillonnées

Delphine Blanke

L.S.T.A., Université Paris 6, 175, rue du Chevaleret, 75013 Paris, France

Reçu le 14 août 2003 ; accepté après révision le 29 septembre 2003

Présenté par Paul Deheuvels

Résumé

La vitesse d'estimation de la densité en temps continu dépend de la nature des trajectoires observées : plus celles-ci sont « irrégulières », plus les vitesses de convergence sont bonnes. Dans cet esprit, nous donnons une vitesse de convergence presque sûre de l'estimateur à noyau de la densité dans le cas où les données sont délivrées sous une forme discrétisée. Le comportement de l'estimateur dépend de deux paramètres : r_0 , γ_0 liés respectivement à la régularité de la densité estimée et à celle de la trajectoire. Nous proposons un estimateur adaptatif relativement à γ_0 ainsi qu'un estimateur doublement adaptatif (par rapport à r_0 et γ_0). On établit ainsi que la vitesse de convergence obtenue dans le cas r_0 , γ_0 connus est atteinte par ces estimateurs. **Pour citer cet article :** D. Blanke, C. R. Acad. Sci. Paris, Ser. I 337 (2003).

© 2003 Académie des sciences. Publié par Éditions scientifiques et médicales Elsevier SAS. Tous droits réservés.

Abstract

Adaptive estimation of density with sampled observations. In continuous time, rates of convergence for nonparametric density estimators depend on the nature of sample paths: roughly speaking, the more 'irregular' the paths are, the better the rates are. In this framework, we give the pointwise rate of convergence of the kernel density estimator in the case of sampled observations. Behaviour of the estimator depends on two coefficients r_0 , γ_0 respectively linked with regularity of density and regularity of sample paths. We propose an adaptive estimator relatively to γ_0 as well as a doubly adaptive estimator (with respect to r_0 and γ_0). It is shown that the rate of convergence obtained in the case of known r_0 , γ_0 is achieved by such adaptive estimators. **To cite this article:** D. Blanke, C. R. Acad. Sci. Paris, Ser. I 337 (2003).

© 2003 Académie des sciences. Publié par Éditions scientifiques et médicales Elsevier SAS. Tous droits réservés.

1. Introduction

Soit $\{X_t, t \in \mathbb{R}\}$ un processus à temps continu mesurable, à valeurs dans \mathbb{R}^d tel que les X_t ont une densité commune f . L'estimation de f quand le processus est observé sur $[0, T]$, avec $T \rightarrow \infty$, est l'objet de nombreux résultats, voir par exemple [5] et les références incluses. Dans cette Note, nous étudions le comportement presque sûr de l'estimateur à noyau construit sur n observations discrétisées issues de la trajectoire du processus. Plus précisément, on suppose que les observations sont fréquentes et sur une longue durée : $X_{\delta_n}, \dots, X_{n\delta_n}$ avec δ_n tel que $\delta_n \rightarrow 0$ et $n\delta_n \rightarrow +\infty$. Clairement plus les trajectoires sont régulières, plus les variables successives $X_{\delta_n}, X_{2\delta_n}, \dots$

Adresse e-mail : dblank@ccr.jussieu.fr (D. Blanke).

sont localement corrélées : ceci induit un choix du pas de temps δ_n plus grand que si les trajectoires sont de nature plus irrégulières. Ainsi la durée totale d'observation $n\delta_n$ sera fonction de la nature des trajectoires.

Dans une première partie, nous commençons par donner le comportement de l'estimateur lorsque γ_0 est supposé connu : nous donnons les seuils minimaux $\delta_n(\gamma_0)$ tels que pour tous les $\delta_n : \delta_n \geq \delta_n(\gamma_0)$, l'estimateur a une vitesse similaire au cas i.i.d. Pour γ_0 inconnu, nous construisons un estimateur adaptatif atteignant la même vitesse de convergence. Finalement, il est bien connu que les estimateurs dépendent également d'un paramètre, r_0 , lié à la régularité de la densité. Dans une dernière partie, nous étudions et donnons les propriétés d'un estimateur doublement adaptatif (r_0 et γ_0 inconnus).

2. Estimation avec paramètres connus

Nous utilisons un estimateur à noyau classique : $\hat{f}_n(x) = \frac{1}{nh_n^d} \sum_{i=1}^n K\left(\frac{x-X_{i\delta_n}}{h_n}\right)$ où h_n représente la fenêtre : $h_n \rightarrow 0, nh_n^d \rightarrow \infty$ et K est un noyau usuel à support compact E . Dans la suite, nous supposons que pour $s \neq t$, la densité jointe $f_{(X_s, X_t)}$ de (X_s, X_t) existe et on note $g_u = f_{(X_0, X_u)} - f \otimes f, u > 0$. Les hypothèses portant sur le processus $\{X_t, t \in \mathbb{R}\}$ sont les suivantes :

Hypothèses 1 (H1).

- (i) f est bornée et $f_{(X_s, X_t)} = f_{(X_0, X_{t-s})}$ pour $t > s$;
- (ii) $\{X_t, t \in \mathbb{R}\}$ est un processus fortement mélangeant de coefficient α_X (voir e.g. [10]) tel que $\exists 0 < \rho < 1, \alpha_0 > 0 : \forall u \in]0, +\infty[, \alpha_X(u) \leq \alpha_0 \rho^u$;
- (iii) $\exists u_0 > 0 : \forall u \in [u_0, +\infty[, \|g_u\|_\infty \leq \pi(u)$, pour une fonction $\pi(u)$ bornée et décroissante à l'infini telle que $\int_{u_1}^{+\infty} \pi(u) du < +\infty, u_1 > 0$;
- (iv) $\exists \gamma_0 \in]0, +\infty[: f_{(X_0, X_u)}(y, z) \leq M(y, z)u^{-\gamma_0}, (y, z, u) \in \mathcal{V}_x^2 \times]0, u_0[$ avec $M(\cdot, \cdot)$ continue en (x, x) et \mathcal{V}_x représentant un voisinage de x .

Les conditions H1(i)–(iii) sont classiques en estimation non paramétrique. La condition H1(iv) est typique du temps continu : elle est reliée à la nature des trajectoires et à la dimension d . Des commentaires et des exemples précis de processus satisfaisant à ces hypothèses peuvent se trouver dans [1–6], [9] et [12]. Notamment le cas $\gamma_0 < 1$ correspond au cas suroptimal, voir par exemple [8,6] et le cas $\gamma_0 > 1$ intervient surtout pour des dimensions $d > 1$. Pour l'étude du biais, les hypothèses habituelles sur f et K sont nécessaires : pour $r \geq 1$, on note

$$\mathcal{C}_K^{(r)} \left\{ \begin{array}{l} \int_E u_1^{\alpha_1} \dots u_d^{\alpha_d} K(\mathbf{u}) du_1 du_2 \dots du_d = 0, \quad \int_E |u_1|^{\beta_1} \dots |u_d|^{\beta_d} \|u\| |K(\mathbf{u})| du_1 \dots du_d < +\infty, \\ \forall \alpha_i: \alpha_1 + \dots + \alpha_d = k, k = 1, \dots, r-1, \quad \forall \beta_i: \beta_1 + \dots + \beta_d = r \end{array} \right.$$

et

$$\mathcal{C}_f^{(r)} \left\{ \begin{array}{l} f \text{ possède } (r-1) \text{ dérivées partielles continues, } f^{(r)} \text{ existe et} \\ \left| \frac{\partial^{(r)} f}{\partial x_1^{j_1} \dots \partial x_d^{j_d}}(x) - \frac{\partial^{(r)} f}{\partial y_1^{j_1} \dots \partial y_d^{j_d}}(y) \right| \leq \ell \|x - y\|^\lambda, \quad j_1 + \dots + j_d = r, (\ell, \lambda) \in]0, +\infty[\times]0, 1]. \end{array} \right.$$

Hypothèses 2 (H2). f satisfait à la condition $\mathcal{C}_f^{(r_0)}$; K satisfait à la condition $\mathcal{C}_K^{(r_0)}$.

Nous avons alors le

Théorème 2.1. (a) Sous H1(i)–(iii) et H2, si $\delta_n \equiv \delta > 0$ et $h_n = c \left(\frac{\ln n}{n}\right)^{1/(2r_0+d)}$, $c > 0$, alors

$$\limsup_{n \rightarrow +\infty} \left(\frac{n}{\ln n}\right)^{r_0/(2r_0+d)} |\hat{f}_n(x) - f(x)| \leq C_1 \quad p.s. \tag{1}$$

avec

$$C_1 = 2^{3/2} c^{-d/2} f^{1/2}(x) \|K\|_2 + c^{r_0} |b_{r_0}(x)| \tag{2}$$

et $b_{r_0}(x)$ est le terme de biais classique, fonction de $f^{(r_0)}$ et K .

(b) Si $r_0 > \frac{d}{\max(1, \gamma_0)}$ et si de plus la condition H1(iv) est satisfaite, alors pour tout $\delta_n \rightarrow 0$ tel que $\delta_n/\delta_n^*(\gamma_0) \rightarrow \infty$, le résultat (1) reste vrai en ayant posé : pour $\gamma_0 < 1$: $\delta_n^*(\gamma_0) = d_1 h_n^d$, pour $\gamma_0 = 1$: $\delta_n^*(\gamma_0) = d_2 h_n^d \ln(1/h_n)$, pour $\gamma_0 > 1$: $\delta_n^*(\gamma_0) = d_3 h_n^{d/\gamma_0}$ avec $d_i > 0$, $i = 1, 2, 3$ constantes quelconques.

Les vitesses pour l'échantillonnage à pas fixe (Théorème 2.1(a)) sont similaires au cas i.i.d. Pour la partie (b), on voit que δ_n est une fonction croissante de γ_0 : le cas $\gamma_0 < 1$ (cas suroptimal) apparaît déjà dans [6], et l'erreur quadratique est traitée par [4]. Notons également que la valeur des constantes d_i (intervenant dans le seuil $\delta_n^*(\gamma_0)$) ne joue pas de rôle dans le comportement asymptotique de l'estimateur (de part la condition $\delta_n/\delta_n^*(\gamma_0) \rightarrow \infty$).

3. Estimation adaptative : r_0 connu, γ_0 inconnu

Nous utilisons ici une méthodologie similaire à celle initiée dans [11] mais dans un contexte différent. Soit Γ_n la grille dénotant les valeurs testées pour γ_0 :

$$\Gamma_n = \{\tilde{\gamma}_0, \gamma_{1,n}, \gamma_{2,n}, \dots, \gamma_{N_n,n}, \gamma_\infty\} \tag{3}$$

avec $\gamma_{0,n} = \tilde{\gamma}_0$ ($0 < \tilde{\gamma}_0 < 1$), $\gamma_{j+1,n} - \gamma_{j,n} = \tau_{j,n}$, $j = 1, \dots, N_n - 1$, et γ_∞ représentant le cas de l'échantillonnage à pas fixe. L'estimateur de γ_0 , $\gamma_{0,n}^* := \gamma_0^*$, est défini par

$$\gamma_0^* = \min \left\{ \gamma_1 \in \Gamma_n : \forall \gamma_2 \in \Gamma_n, \gamma_2 \geq \gamma_1, \left(\frac{n}{\ln n}\right)^{r_0/(2r_0+d)} |\hat{f}_{r_0, \gamma_2}(x) - \hat{f}_{r_0, \gamma_1}(x)| \leq \eta \right\}, \tag{4}$$

où η est une constante strictement positive et

$$\hat{f}_{r, \gamma}(x) := \frac{1}{n h_n^d(r)} \sum_{i=1}^n K(r) \left(\frac{x - X_{i \delta_h(r, \gamma)}}{h_n(r)} \right), \tag{5}$$

où $h_n(r) = c \left(\frac{\ln n}{n}\right)^{1/(2r+d)}$, $K(r)$ satisfait $\mathcal{C}_K^{(r)}$ et $\delta_h(r, \gamma)$ vérifie

$$\begin{cases} \text{si } \gamma < 1 : & \delta_h(r, \gamma) = d_1 h_n^d(r), & \text{si } \gamma = 1 : & \delta_h(r, \gamma) = d_2 h_n^d(r) \ln(1/h_n(r)), \\ \text{si } \gamma > 1 : & \delta_h(r, \gamma) = d_3 h_n^{d/\gamma}(r), & \text{si } \gamma = \gamma_\infty : & \delta_h(r, \gamma) = \delta. \end{cases} \tag{6}$$

L'estimateur adaptatif pour f est finalement donné par $\hat{f}_{r_0, \gamma_0^*}(x) = \frac{1}{n h_n^d(r_0)} \sum_{i=1}^n K(r_0) \left(\frac{x - X_{i \delta_h(r_0, \gamma_0^*)}}{h_n(r_0)} \right)$.

Nous obtenons ainsi la même vitesse de convergence que dans le cas non adaptatif :

Théorème 3.1. Si H1 et H2 sont satisfaites alors, $\limsup_{n \rightarrow \infty} \left(\frac{n}{\ln n}\right)^{r_0/(2r_0+d)} |\hat{f}_{r_0, \gamma_0^*}(x) - f(x)| \leq \eta + C_1$ p.s. avec C_1, η donnés par (2), (4).

4. Estimation double adaptative : r_0 et γ_0 inconnus

L'estimation adaptative relativement à la régularité de la densité (r_0) est maintenant un problème classique, voir par exemple [7] (cas i.i.d.) et [9] (estimateur par projection). Ici la difficulté réside dans le fait que deux paramètres (de nature différentes) sont inconnus : on introduit donc une nouvelle grille Δ_n correspondant aux valeurs testées pour $r_0 : \Delta_n = \{1, 2, \dots, \kappa_n\}$ et l'estimation suit avec

$$r_0^* = \max \left\{ r_1 \in \Delta_n : \forall r_2 \in \Delta_n, r_2 \leq r_1, \left(\frac{n}{\ln n} \right)^{r_2/(2r_2+d)} \left| \hat{f}_{r_2, \gamma_\infty}(x) - \hat{f}_{r_1, \gamma_\infty}(x) \right| \leq \hat{\eta}_a(r_2) \right\},$$

$$\gamma_0^* = \min \left\{ \gamma_1 \in \Gamma_n : \forall \gamma_2 \in \Gamma_n, \gamma_2 \geq \gamma_1, \left(\frac{n}{\ln n} \right)^{r_0^*/(2r_0^*+d)} \left| \hat{f}_{r_0^*, \gamma_2}(x) - \hat{f}_{r_0^*, \gamma_1}(x) \right| \leq \zeta \right\}$$

où $\Gamma_n, \hat{f}_{(r, \gamma)}$ sont définis par (3), (5), ζ est une constante strictement positive et $\hat{\eta}_a(r)$ représente un estimateur préliminaire pour $aC_1, a > 0$, dont la consistance s'établit grâce aux hypothèses H1 et H2. Finalement, $\hat{f}_{r_0^*, \gamma_0^*}(x) = \frac{1}{nh_n^d(r_0^*)} \sum_{i=1}^n K_{(r_0^*)} \left(\frac{x - X_i \delta_n(r_0^*, \gamma_0^*)}{h_n(r_0^*)} \right)$ avec $h_n(r_0^*) = c \left(\frac{\ln n}{n} \right)^{1/(2r_0^*+d)}$, $\delta_n(r_0^*, \gamma_0^*)$ comme dans (6) et $K_{(r_0^*)}$ satisfaisant à $C_K^{(r_0^*)}$.

Théorème 4.1. *Sous H1 et H2, si $\kappa_n = \mathcal{O}\left(\frac{\ln n}{(\ln \ln n)^{1+\varepsilon}}\right)$, $\varepsilon > 0$, on obtient pour tout $a > 2$,*

$$\limsup_{n \rightarrow \infty} \left(\frac{n}{\ln n} \right)^{r_0/(2r_0+d)} \left| \hat{f}_{r_0^*, \gamma_0^*}(x) - f(x) \right| \leq (a+1)C_1 + \zeta \text{ p.s.}$$

Les démonstrations de ces résultats figurent dans [2] et font appel à des techniques classiques liées à l'estimation adaptative et au temps continu. Le cas où la trajectoire est observée globalement est traité dans [1].

Références

- [1] D. Blanke, Sample paths adaptive density estimator, 2003, 29 p., submitted for publication.
- [2] D. Blanke, Optimal adaptive sampling schemes for density estimation, 2003, 29 p., submitted for publication.
- [3] D. Blanke, D. Bosq, A family of minimax rates for density estimators in continuous time, *Stochastic Anal. Appl.* 18 (6) (2000) 871–900.
- [4] D. Blanke, B. Pumo, Optimal sampling for density estimation in continuous time, *J. Time Ser. Anal.* 24 (1) (2003) 1–24.
- [5] D. Bosq, *Nonparametric Statistics for Stochastic Processes. Estimation and Prediction*, 2nd edition, in: *Lecture Notes in Statist.*, Vol. 110, Springer-Verlag, New York, 1998.
- [6] D. Bosq, Parametric rates of nonparametric estimators and predictors for continuous time processes, *Ann. Statist.* 25 (3) (1997) 982–1000.
- [7] C. Butucea, Exact adaptive pointwise estimation on Sobolev classes of densities, *ESAIM Probability and Statistics* 5 (2001) 1–31.
- [8] J.V. Castellana, M.R. Leadbetter, On smoothed probability density estimation for stationary processes, *Stochastic Process. Appl.* 21 (1986) 179–193.
- [9] F. Comte, F. Merlevède, Density estimation for a class of continuous time or discretely observed processes, 2002, submitted for publication.
- [10] P. Doukhan, *Mixing: Properties and Examples*, in: *Lecture Notes in Statist.*, Vol. 85, Springer-Verlag, 1994.
- [11] O.V. Lepskii, On a problem of adaptive estimation in Gaussian white noise, *Theory Probab. Appl.* 35 (1990) 454–466.
- [12] M. Sköld, The asymptotic variance of the continuous-time kernel estimator with applications to bandwidth selection, *Statist. Inference Stochastic Process.* 4 (1) (2001) 99–117.