Available online at www.sciencedirect.com
ScienceDirect
C. R. Acad. Sci. Paris, Ser. I 345 (2007) 101–105
COMPTES RENDUS MATHEMATIQUE
http://france.elsevier.com/direct/CRASS1/

Statistics

Nonparametric trend coefficient estimation for multidimensional diffusions

Annamaria Bianchi

Department of Mathematics, University of Milan, via Saldini 50, 20133 Milano, Italy

Received 10 February 2006; accepted after revision 27 April 2007

Available online 5 July 2007

Presented by Paul Deheuvels

Abstract

We consider the problem of the density and drift estimation by the observation of a trajectory of an $\mathbb{R}^d$ dimensional homogeneous diffusion process with a unique invariant density. We construct estimators of the kernel type and study the mean-square and almost sure uniform asymptotic behavior for these estimators. Finally, we give a class of processes satisfying our assumptions. ***To cite this article: A. Bianchi, C. R. Acad. Sci. Paris, Ser. I 345 (2007).***

Résumé

Estimation non-paramétrique du terme de dérive dans un processus de diffusion multidimensionnel. On considère le problème de l'estimation de la densité et du terme de dérive par l'observation d'une trajectoire d'un processus de diffusion homogène en dimension d ayant une densité invariante unique. On construit les estimateurs par la méthode des noyaux, puis on en étudie le comportement asymptotique en L^2 et presque sûr. Finalement, on donne à titre d'exemple une classe de processus qui satisfont nos hypothèses. ***Pour citer cet article : A. Bianchi, C. R. Acad. Sci. Paris, Ser. I 345 (2007).***

Version française abrégée

On étudie un processus de diffusion $\{X_t\}$ homogène en dimension d et défini par (1). On suppose que ce processus admet une densité invariante unique $f(\cdot)$. Si la valeur initiale X_0 a pour densité $f(\cdot)$ alors le processus $\{X_t\}$ est strictement stationnaire. L'objectif de cette Note est d'estimer le terme de dérive de $\{X_t\}$ à partir de l'observation d'une trajectoire.

Pour estimer $f(\cdot)$ on utilise l'estimateur classique de la densité par noyaux (2). Ensuite on définit l'estimateur $f'_{i,T}$ $(i = 1, \ldots, d)$ pour les dérivées de la densité (3). Enfin à partir de l'équation forward de Kolmogorov, on en déduit un estimateur pour le terme de dérive (4). Enfin, on établit, sous des hypothèses convenables, le comportement asymptotique en L^2 et presque sûr uniforme de ces estimateurs.

E-mail address: abianchi@mat.unimi.it.

1631-073X/$ – see front matter
doi:10.1016/j.crma.2007.05.012

1. Trend coefficient estimation

Let $\{X_t\}$ be an $\mathbb{R}^d$-dimensional homogeneous diffusion process defined by

$$\mathrm{d}X_t = S(X_t)\,\mathrm{d}t + \mathrm{d}W_t, \quad X_0, \quad 0 \leqslant t \leqslant T, \tag{1}$$

where $\{W_t, t \geqslant 0\}$ is a standard d-dimensional Wiener process, $S : \mathbb{R}^d \to \mathbb{R}^d$ is an (unknown) drift coefficient and X_0 is an initial value of X_t, which is assumed to be not dependent on the Wiener process. In our setup, the diffusion coefficient $\sigma(\cdot)^2$ is supposed to be the identity matrix; in the general case, it is identifiable using the quadratic variation of the process. Moreover, recall that we can reduce a multidimensional diffusion process to form (1), under suitable conditions on σ (see [1]). Finally we suppose that $S(\cdot)$ belongs to the class of functions $\mathcal{S}_d$ satisfying:

(A1) S is a Lipschitz function;
(A2) there exist constants $M_0 \geqslant 0$ and $r > 0$ such that

$$\left(S(x), \frac{x}{\|x\|}\right) \leqslant -\frac{r}{\|x\|^p}, \quad 0 \leqslant p < 1,\ \|x\| \geqslant M_0,$$

where $(\cdot\,,\cdot)$ denotes the scalar product in $\mathbb{R}^d$;
(A3) the potential conditions hold

$$\frac{\partial S_j}{\partial x_k} = \frac{\partial S_k}{\partial x_j}, \quad \forall j, k = 1, \dots, d;$$

(A4) S is twice continuously differentiable, with the first and second partial derivatives satisfying the linear growth condition.

We remark that Assumption (A1) guarantees the existence of a unique strong solution to Eq. (1) (see [5]). Assumption (A2) implies the existence of an invariant measure for the process, the convergence of the transition probability to this invariant measure with (sub)exponential rate and the strong mixing property (see [11] and [12]). From now on we assume that the initial value X_0 follows the invariant law, ensuring thus the strong stationarity of the process $\{X_t\}$. Moreover, since the diffusion coefficient $\sigma^2 \equiv I$ is non degenerate, from (A1) we deduce that the invariant measure admits a density $f(\cdot)$ with respect to the Lebesgue measure.

We are now interested in the estimation of the trend coefficient S, by the observation of a trajectory of the process over a time interval $[0, T]$. The method we follow passes through the estimation of f and its derivatives first. It is worthwhile to remark that, while the statistical theory for the scalar case has been very well developed (refer to [7]), there are still few results concerning the multidimensional context.

As far as the invariant density is concerned, we use the standard kernel density estimator defined by

$$f_T(x) = \frac{1}{T h_T^d} \int_0^T K\left(\frac{x - X_t}{h_T}\right) \mathrm{d}t, \quad x \in \mathbb{R}^d, \tag{2}$$

where $h_T \to 0^+$, as $T \to +\infty$ and $K : \mathbb{R}^d \to \mathbb{R}$ is a bounded probability density function such that for $j = 1, \dots, d$

(K1) $\int_{\mathbb{R}^d} K(u_1, \dots, u_d) u_j \,\mathrm{d}u_1 \cdots \mathrm{d}u_d = 0$,
(K2) $\int_{\mathbb{R}^d} \|u\| |u_j| K(u_1, \dots, u_d) \,\mathrm{d}u_1 \cdots \mathrm{d}u_d < +\infty$.

Under Assumptions (A2) and (A3) the density function f admits a gradient $f' = (f_1', \dots, f_d')$, with $f_i' = \frac{\partial f}{\partial x_i}$, $i = 1, \dots, d$. As far as $f'(x)$ is concerned, we consider the estimator $f_T' = (f_{1,T}', \dots, f_{d,T}')$, with components

$$f_{i,T}'(x) = \frac{1}{T h_T^{d+1}} \int_0^T K_i'\left(\frac{x - X_t}{h_T}\right) \mathrm{d}t, \quad x \in \mathbb{R}^d,\ i = 1, \dots, d, \tag{3}$$

where $h_T \to 0^+$ as $T \to +\infty$ and $K(\cdot)$ is a kernel which satisfies conditions (K1) and (K2) and moreover is such that:

(K3) $K_i' = \frac{\partial K}{\partial x_i}$ exists and is continuous everywhere for $i = 1, \dots, d$;
(K4) $\int_{\mathbb{R}^d} |K_i'(u)|\,\mathrm{d}u < +\infty$, $i = 1, \dots, d$;
(K5) $K(\cdot)$ has bounded and continuous second partial derivatives.

We are now able to define the estimator for the trend coefficient. Thanks to Assumption (A3), from the Kolmogorov forward equation we deduce that for $x \in \mathbb{R}^d$ fixed the invariant density f satisfies the relation $f'(x) = 2S(x)f(x)$, $x \in \mathbb{R}^d$ (see [10]). Therefore a natural estimator for $S(\cdot)$ is

$$S_T(x) = \frac{f_T'(x)}{2f_T(x) + \varepsilon_T}, \quad x \in \mathbb{R}^d, \tag{4}$$

with $\varepsilon_T \to 0^+$ as $T \to +\infty$. We have added ε_T at the denominator since for finite values of T, $f_T(x)$ may be very close or even equal to zero. In the next sections, we will choose ε_T and the h_T's depending on the type of convergence.

2. Mean-square asymptotic behavior

In the present section we are interested in studying the asymptotic behavior of the estimators defined above in the mean-square sense. An important property of multidimensional diffusion processes is that, unlike the scalar case, the classical Castellana and Leadbetter condition does not hold anymore. Indeed, the joint density $f_u(x, y)$ of the pair (X_0, X_u), $u > 0$, is not integrable in a neighborhood of $u = 0$, since $f_u(x, y) \geqslant Cu^{-d/2}$, $C > 0$, for $u \in (0, 1]$ (see [9], Theorem 3.1). This implies that the kernel density estimator (2) cannot reach the parametric rate $1/T$ (refer to [4]). In the next proposition we show that this estimator converges with intermediate rates.

Proposition 2.1. *For all $S \in \mathcal{S}_d$, for all kernels $K(\cdot)$ satisfying conditions* (K1) *and* (K2) *and for $h_T = c(\frac{\ln T}{T})^{1/4}$ $(c > 0)$ if $d = 2$, and $h_T = cT^{-1/(d+2)}$ $(c > 0)$ if $d > 2$*

$$\limsup_{T\to+\infty} \Gamma_T^2(d)\mathbb{E}\big[\big(f_T(x) - f(x)\big)^2\big] < +\infty,$$

where $\Gamma_T(d) = (\frac{T}{\ln T})^{1/2}$ if $d = 2$, and $\Gamma_T(d) = T^{2/(d+2)}$ if $d > 2$.

This proposition follows from the results by Blanke and Bosq [3], where intermediate rates for kernel density estimators are provided under slightly weaker conditions than the Castellana and Leadbetter one. These rates depend on the local behavior of the joint density $f_u(x, y)$, when u is small. Using results in [9] and [12], it is easy to see that for all $S \in \mathcal{S}_d$, process (1) satisfies the conditions presented in [3]. From [3], it follows also that the previous rates are sharp, since a lower bound for the quadratic risk can be proved. Moreover, these rates turn out to be minimax in the sense that if the kernel estimator reaches an intermediate rate over a class of processes, than no better estimator exists over that class.

We turn now to the analysis of f_T'. We expect slower rates of convergence compared to the density ones (see [8]). The following proposition shows the asymptotic behavior of f_T':

Proposition 2.2. *For all $S \in \mathcal{S}_d$, for all kernels $K(\cdot)$ satisfying conditions* (K1)–(K4) *and for $h_T = c(\frac{\ln T}{T})^{1/6}$ $(c > 0)$ if $d = 2$, and $h_T = cT^{-1/(d+4)}$ $(c > 0)$ if $d > 2$, we obtain that*

$$\limsup_{T\to+\infty} \Phi_T^2(d)\mathbb{E}\big[\big\| f_T'(x) - f'(x)\big\|^2\big] < +\infty,$$

where $\Phi_T(d) = (\frac{T}{\ln T})^{1/3}$ if $d = 2$, $\Phi_T(d) = T^{2/(d+4)}$ if $d > 2$, and $\|\cdot\|$ is the Euclidean norm in $\mathbb{R}^d$.

In the proof we mainly follow the methods used by Blanke and Bosq [3], and we adapt them to the case of the density derivative.

Using the previous results and a suitable decomposition we find the mean-square asymptotic behavior for the trend coefficient.

Proposition 2.3. *Under the assumptions of Propositions* 2.1 *and* 2.2 *and for* $\varepsilon_T = T^{-1/2}$ *if* $d = 2$, *and* $\varepsilon_T = (T \ln T)^{-2/(d+2)}$ *if* $d > 2$, *we obtain that*

$$\limsup_{T\to+\infty} \Phi_T^2(d)\mathbb{E}\big[\big\|S_T(x) - S(x)\big\|^2\big] < +\infty,$$

where $\Phi_T(d)$ *is defined in Proposition* 2.2.

The proof is based on the following decomposition

$$S_{i,T} - S_i = \frac{1}{2f_T + \varepsilon_T}(f'_{i,T} - \mathbb{E}f'_{i,T}) + \mathbb{E}f'_{i,T}\left(\frac{1}{2f_T + \varepsilon_T} - \frac{1}{2f}\right) + \frac{1}{2f}(\mathbb{E}f'_{i,T} - f'_i), \quad i = 1, \ldots, d, \tag{5}$$

and the fact that $\mathbb{P}(2f_T + \varepsilon_T < \mathbb{E}f_T)$ tends to zero exponentially fast as $T \to +\infty$, thanks to an exponential type inequality (Theorem 1.3 in [4]). Indeed, using the inequality $(a+b+c+d)^2 \leqslant 4(a^2+b^2+c^2+d^2)$ in (5) and taking the expectations, we can use the majoration $\frac{1}{2f_T+\varepsilon_T} \leqslant \frac{1}{\mathbb{E}f_T}$, the remaining part being negligible. Results close to those presented in this section were independently obtained in [6]. However, the approach of this Note may be of interest, since the estimators employed and the methods used in the proofs are different.

3. Almost sure behavior

In this section we study the almost sure uniform convergence of f, f'_T and S_T.

Proposition 3.1. *For all* $S \in \mathcal{S}_d$, *for all kernels* $K(\cdot)$ *with compact support and satisfying conditions* (K1)–(K3), *and for* $h_T = c(\frac{(\log T)^{1+1/\delta}}{T})^{1/(d+2)}$, $c > 0$,

$$\limsup_{T\to+\infty}\left(\frac{T}{(\log T)^{1+1/\delta}}\right)^{2/(d+2)} \sup_{x\in\mathbb{R}^d}\big|f_T(x) - f(x)\big| < +\infty \quad a.s.,$$

where $\delta = (1-p)/(1+p)$, p *being the exponent appearing in condition* (A2).

This proposition is a consequence of a more general result by Blanke [2], where she proves the strong uniform convergence for kernel density estimators in continuous time.

Concerning the estimator of the density derivative, we have the following result:

Proposition 3.2. *For all* $S \in \mathcal{S}_d$, *for all kernels* $K(\cdot)$ *with compact support and satisfying conditions* (K1)–(K5), *and for* $h_T = c(\frac{(\log T)^{1+1/\delta}}{T})^{1/(d+2)}$ $(c > 0)$ *we obtain that*

$$\limsup_{T\to+\infty} \Psi_T(d) \sup_{x\in\mathbb{R}^d}\big\|f'_T(x) - f'(x)\big\| < +\infty \quad a.s.,$$

where $\Psi_T(d) = (\frac{T^{1/3}}{\log T})$ *if* $d = 2$, $\Psi_T(d) = (\frac{T^{2/(d+4)}}{(\log T)^{1/2}})$ *if* $d > 2$.

In order to prove this result we mainly adapt the methods used by Blanke [2] to the estimation of the density derivative. First we prove the pointwise strong consistency by using a Benstein's type inequality together with a slight variation of the Borel–Cantelli lemma for continuous time processes [4], thanks to the uniform continuity of the function $T \mapsto \Psi_T(d) f'_T$. Then we study the uniform strong consistency over an increasing sequence of compact subsets of $\mathbb{R}^d$, using the Lipschitzianity of K'_i, $i = 1, \ldots, d$. Finally we extend the result to the whole $\mathbb{R}^d$ thanks to the continuity and the integrability of $f'(x)$ and the fact that $f'_T(x)$ tends to zero almost surely sufficiently fast as $\|x\| \to +\infty$.

As far as the trend coefficient is concerned, it follows immediately that, for $x \in \mathbb{R}^d$ fixed, $S_T(x)$ converges almost surely to $S(x)$, because S_T is a continuous function of f_T, f'_T and ε_T. By choosing, for example, ε_T as in Proposition 2.3 the convergence rate is $\Psi_T(d)$, which was defined above. We recognize in the supnorm rates obtained in Propositions 3.1 and 3.2 the same rates, up to logarithmic factors, as those presented for the mean-square error.

Concerning the uniform convergence of S_T, it can be obtained over compact sets, but in general not over the whole $\mathbb{R}^d$, since the lower bound for f over $\mathbb{R}^d$ is zero. On the other hand, for any compact set $K \subset \mathbb{R}^d$, $\inf_{x\in K} f(x) > 0$. Moreover, in order to apply a modification of the Borel–Cantelli lemma in continuous time (see [4] and [3]), we need the uniform continuity of the function $T \mapsto \sup_{x\in K} S_T(x)$. In order to have this property we decrease the rate of convergence of the estimator.

Proposition 3.3. *Under the assumptions of Propositions* 3.1 *and* 3.2 *and for* $\varepsilon_T = T^{-1/(d+2)}$ *and for any compact set* $K \subset \mathbb{R}^d$*, we have that*

$$\limsup_{T\to+\infty} T^{\frac{1}{d+2}} \sup_{x\in K} \left\| S_T(x) - S(x) \right\| < +\infty \quad a.s.$$

The proof is based on decomposition (5) and the fact that

$$\mathbb{P}\left(\limsup_{T\to+\infty} \sup_{x\in K} \frac{1}{2f_T + \varepsilon_T} < +\infty \right) = 1.$$

The last statement may be easily proved by following the same scheme as the one described for the density derivative.

4. Examples

We can easily verify that the classical Ornstein–Uhlenbeck process

$$\mathrm{d}X_t = BX_t\,\mathrm{d}t + \mathrm{d}W_t,$$

where $B = (b_{ij})_{1\leqslant i,j\leqslant d}$, with $b_{ij} = b_{ji}$, $i \neq j$ and $\det B > 0$, $\operatorname{tr} B < 0$, satisfies our conditions.

Another possible example is the stochastic process described by the following system of equations

$$\mathrm{d}X_{k,t} = -\left[\nabla U(X_{k,t}) + \frac{\theta}{N} \sum_{l=1}^{N} (X_{k,t} - X_{l,t}) \right] \mathrm{d}t + \mathrm{d}W_{k,t}, \quad k = 1, \ldots, N,$$

where $W_1, \ldots, W_N$ are independent d-dimensional Wiener processes and $\theta \geqslant 0$ denotes a constant. The function $U(x)$, $x \in \mathbb{R}^d$, is called potential; we can take, for example, $U(x) = x^2$.

Acknowledgements

We are thankful to Prof. Denis Bosq and Prof. Vincenzo Capasso for their continuous advice and relevant discussions, and to the anonymous referees for the very careful reading and useful suggestions.

References

[1] Y. Ait-Sahalia, Closed-form likelihood expansions for multivariate diffusions, Preprint available under http://www.princeton.edu/~yacine/research.htm, 2004.
[2] D. Blanke, Sample paths adaptive density estimator, Math. Methods Statist. 13 (2) (2004) 123–152.
[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. Bosq, Nonparametric Statistics for Stochastic Processes, Springer-Verlag, New York, 1998.
[5] V. Capasso, D. Bakstein, An Introduction to Continuous-Time Stochastic Processes, Birkhäuser, Boston, 2005.
[6] A. Dalalyan, M. Reiß, Asymptotic statistical equivalence for ergodic diffusions: the multidimensional case, Preprint available under http://www.proba.jussieu.fr/pageperso/dalalyan/, 2005.
[7] Yu.A. Kutoyants, Statistical Inference for Ergodic Diffusion Processes, Springer Series in Statistics, New York, 2004.
[8] A. Lucas, Can we estimate the density's derivative with suroptimal rate?, Statist. Inference Stoch. Process. 1 (1998) 29–41.
[9] Z. Qian, W. Zheng, A representation formula for transition probability densities of diffusions and applications, Stochastic Process. Appl. 111 (2004) 57–76.
[10] H. Risken, The Fokker–Planck Equation, Springer-Verlag, Berlin, 1989.
[11] A.Yu. Veretennikov, Bounds for the mixing rate in the theory of stochastic equations, Theory Probab. Appl. 32 (2) (1987) 273–281.
[12] A.Yu. Veretennikov, On subexponential mixing rate for Markov processes, Theory Probab. Appl. 49 (1) (2005) 110–122.