 Vol 28: 617–623, 2001

A Note on Consistent Estimation of Multivariate Parameters in Ergodic Diffusion Models

J. H. VAN ZANTEN

Centre for Mathematics and Computer Science (CWI)

ABSTRACT. Certain aspects of maximum likelihood estimation for ergodic diffusions are studied via recently developed empirical process theory for martingales. This approach enables us to remove some undesirable regularity conditions that usually appear in the statistical literature on ergodic diffusions. In particular, dimension dependent conditions for the existence of a continuous likelihood and for consistency of the maximum likelihood estimator turn out to be unnecessary.

Key words: consistency, continuous likelihood, empirical process methods, ergodic diffusions, maximum likelihood estimation, multivariate parameter.

1. Introduction

We consider the family of stochastic differential equations

$$dX_t = b_\theta(X_t)\,dt + \sigma(X_t)\,dW_t, \tag{1}$$

where θ runs through a compact parameter space $\Theta \subseteq \mathbb{R}^d$. For every $\theta \in \Theta$ we suppose that (1) has a unique stationary, weak solution taking values in a (possibly unbounded) interval $(l, r) \subseteq \mathbb{R}$. In particular, the state space (l, r) is assumed to be independent of the parameter θ. We suppose that $\sigma > 0$ on (l, r). For every $\theta \in \Theta$ we denote by P_θ the law that is generated by the weak solution on the space $C[0, \infty)$ of continuous functions on $[0, \infty)$. We write μ_θ for the marginal law of X_t under P_θ. By the stationary assumption the distribution μ_θ is independent of t. In particular, the law of the initial random variable X_0 is equal to μ_θ. We suppose furthermore that under P_θ, the process X has the ergodic property with invariant measure μ_θ, i.e. for every $f \in L^1(\mu_\theta)$

$$\frac{1}{t}\int_0^t f(X_s)\,ds \to \int f\,d\mu_\theta \quad P_\theta \text{ a.s.} \tag{2}$$

as $t \to \infty$. See e.g. Gihman & Skorohod (1972) or Skorokhod (1989) for conditions on the functions b_θ and σ under which the stochastic differential (1) admits such a stationary, ergodic solution.

Throughout this note, the parameter $\theta_0 \in \Theta$ will play the role of the "true parameter". Denote by $P_{\theta,t}$ the restriction of the measure P_θ to the space $C[0, t]$ of continuous functions on $[0, t]$. It is well-known that under certain conditions, the measures $P_{\theta,t}$ are absolutely continuous with respect to a dominating measure that is independent of θ. Up to a factor involving the initial random variable X_0, the likelihood process is given by

$$L_t(\theta) = \exp\left(\int_0^t \frac{b_\theta(X_s)}{\sigma^2(X_s)}\,dX_s - \frac{1}{2}\int_0^t \frac{b_\theta^2(X_s)}{\sigma^2(X_s)}\,ds\right) \tag{3}$$

(see for instance Liptser & Shiryayev (1977)). The maximum likelihood estimator $\hat{\theta}_t$ is the maximizer of the random map $\theta \mapsto L_t(\theta)$, provided it exists. For the definition of the estimator

$\hat{\theta}_t$ it is not relevant that $L_t(\theta)$ is in fact a likelihood ratio. To ensure that $L_t(\theta)$ is well-defined under P_{θ_0} it is only needed to assume that

$$\int_0^t \frac{b_\theta^2(X_s)}{\sigma^2(X_s)}\, ds < \infty \quad P_{\theta_0} \text{ a.s.} \tag{4}$$

In this paper, we will never need the fact that $L_t(\theta)$ is a likelihood. Instead of conditions for absolute continuity, we shall therefore use condition (4) in the statement of the theorems below.

Expression (3) can be used to calculate the estimator $\hat{\theta}_t$ from the observations $(X_s)_{s\leqslant t}$. To study the (asymptotic) properties of the estimator, another expression is more useful. Take logarithms in (3) and use (1) to see that under P_{θ_0}, maximizing $\theta \mapsto L_t(\theta)$ is equivalent to maximizing $\theta \mapsto l_t(\theta)$, where

$$l_t(\theta) = \int_0^t \frac{b_\theta(X_s) - b_{\theta_0}(X_s)}{\sigma(X_s)}\, dW_s - \frac{1}{2}\int_0^t \frac{(b_\theta(X_s) - b_{\theta_0}(X_s))^2}{\sigma^2(X_s)}\, ds \tag{5}$$

and W is a standard Brownian motion under P_{θ_0}. So the maximum likelihood estimator $\hat{\theta}_t$ is also the maximizer of the log-likelihood $\theta \mapsto l_t(\theta)$, provided it exists.

In this paper we consider two aspects of maximum likelihod estimation in the model (1). First we question the existence of a continuous version of the log-likelihood (5), viewed as a function in θ. If such a continuous version exists, the maximizer $\hat{\theta}_t$ of that version also exists, by compactness of the parameter space. We then investigate the consistency of the maximum likelihood estimator $\hat{\theta}_t$. The problem of maximum likelihood estimation in non-linear ergodic diffusion models has been studied by several authors, see Kutoyants (1977), Lánská (1979), Prakasa Rao & Rubin (1981), Basu (1983), Kutoyants (1984), Yoshida (1990) and Prakasa Rao (1999), to mention just a few. We reconsider this classical subject in order to apply new empirical process results that were developed recently by entropy methods. This will allow us to analyse the model (1) more effectively than in the past and to improve several known results. We use empirical process theory in a similar fashion as for instance Van der Vaart & Wellner (1996) in the case of i.i.d. observations, Van de Geer (1995) for point processes and Nishiyama (1999) for small noise diffusion models.

In section 2 we consider the continuity of the log-likelihood (5). The stochastic integral in (5) is only defined up to indistinguishability. Therefore, since the parameter space Θ is uncountable, different versions of the log-likelihood may not even be almost surely equal. In particular, we see that a statement like "the log-likelihood is continuous" does not make sense. The correct way to formulate the problem is to ask whether or not there exists a continuous version of the random map $\theta \mapsto l_t(\theta)$, i.e. a random map l_t that is continuous and such that for every $\theta \in \Theta$ the relation (5) holds almost surely. In the literature, two main approaches to this problem are pursued. In Lánská (1979) for instance, Itô's formula is used to see that with

$$f_\theta = \int^x \frac{b_\theta(y) - b_{\theta_0}(y)}{\sigma^2(y)}\, dy$$

and

$$g_\theta = b_{\theta_0} f'_\theta + \frac{1}{2}\sigma^2 f''_\theta + \frac{1}{2}\frac{(b_\theta - b_{\theta_0})^2}{\sigma^2}$$

(a prime denotes differentiation with respect to x) the random map

$$\theta \mapsto f_\theta(X_t) - f_\theta(X_0) - \int_0^t g_\theta(X_s)\, ds$$

is a version of the log-likelihood. It can then simply be assumed that this version is continuous in θ. A drawback here is that one needs to assume that the functions b_θ and σ are differentiable with respect to x in order to have the continuity in θ. Other authors use Kolmogorov's continuity theorem (see for instance Ibragimov & Has'minskii (1981), app. 1) to conclude that there exists a continuous log-likelihood. The drawback of this method is that the resulting conditions on the log-likelihood depend on the dimension d of the parameter space (see e.g. Kutoyants (1984)). In connection with smoothness in θ, both assumptions of smoothness in x and dimension dependent conditions are undesirable and it turns out that we can do without them. Instead, we prove in section 2 that l_t admits a continuous version if the pointwise Hölder condition (6) is satisfied.

In section 3 we discuss the consistency of the maximum likelihood estimator, which is the (or rather some) point where the continuous version of the likelihood attains its maximum. In the case of a multivariate parameter, consistency is usually derived under certain smoothness conditions in θ that depend on the dimension d of the parameter space (see Basu, 1983; Kutoyants, 1984; Prakasa Rao, 1999). This is typically done by using conventional weak convergence theory as can be found for instance in Ibragimov & Has'minskii (1981, app. 1). This procedure turns out to be too crude however and the dependence on the dimension that it causes turns out to be superfluous. Using finer methods, we show in section 3 that it suffices to require the Hölder condition (6) again, as well as a natural indentifiability assumption.

2. Existence of a continuous likelihood

In this section we prove that there exists a continuous version of the random map (5) if the following Hölder condition is satisfied:

There exist a number $\alpha > 0$ and a function $b \in L^2(\mu_{\theta_0})$ such that

$$\left|\frac{b_\theta(x) - b_\psi(x)}{\sigma(x)}\right| \leqslant |\theta - \psi|^\alpha b(x) \tag{6}$$

for all $\theta, \psi \in \Theta$ and $x \in \mathbb{R}$.

Theorem 1

Suppose that (4) *holds for all* $\theta \in \Theta$ *and* $t \geqslant 0$. *If the Hölder condition* (6) *is satisfied, then the random map* l_t *given by* (5) *admits a continuous version*.

Proof. Let $t \geqslant 0$ be fixed. It follows readily from the Hölder condition that the Lebesgue integral in (5) is continuous in θ, so we only have to show that there exists a continous version of the random map

$$\theta \mapsto Z(\theta) = \int_0^t \left(\frac{b_\theta - b_{\theta_0}}{\sigma}\right)(X_s)\, dW_s. \tag{7}$$

Note that by Fubini's theorem and the stationarity of X it holds that

$$E_{\theta_0}(Z(\theta) - Z(\psi))^2 = E_{\theta_0}\int_0^t \left(\frac{b_\theta - b_\psi}{\sigma}\right)^2 (X_s)\, ds = t\left\|\frac{b_\theta - b_\psi}{\sigma}\right\|^2_{L^2(\mu_{\theta_0})}.$$

Using assumption (6) we thus find that

$$E_{\theta_0}(Z(\theta) - Z(\psi))^2 \leqslant t|\theta - \psi|^{2\alpha}\|b\|^2_{L^2(\mu_{\theta_0})}$$

for every $\theta, \psi \in \Theta$. In particular, the random map Z is continuous in probability, i.e. if $\theta_n \to \theta$, then $Z(\theta_n) \to Z(\theta)$ in probability.

Now let Θ^* be a countable, dense subset of Θ. We want to show that with probability 1, the random map Z is uniformly continuous on Θ^*. To that end, we view Z as a collection of endpoints of continuous local martingales that are indexed by a parameter (cf. (7)). For such random maps, Nishiyama (1998) gives sufficient conditions for uniform continuity. In the present case th. 2.4.3 of Nishiyama (1998) states that if d is a metric on Θ such that

$$\|Z\| := \sup_{\theta \neq \psi} \frac{1}{d(\theta, \psi)} \sqrt{\int_0^t \left(\frac{b_\theta - b_{\theta_0}}{\sigma}\right)^2 (X_s)\, ds} < \infty, \qquad P_{\theta_0} \text{ a.s.} \tag{8}$$

and

$$\int_0^1 \sqrt{\log N(\varepsilon, \Theta, d)}\, d\varepsilon < \infty, \tag{9}$$

then with probability 1 the random map Z is uniformly d-continuous on the countable subset Θ^* (here $N(\varepsilon, \Theta, d)$ denotes the minimal number of balls of d-radius ε needed to cover Θ). We apply this result with the metric d given by $d(\theta, \psi) = |\theta - \psi|^{\alpha \wedge 1}$. Then by the boundedness of Θ and condition (6) there exists a constant $C > 0$ such that the quantity $\|Z\|$ in (8) satisfies

$$\|Z\| \leqslant C \sqrt{\int_0^t b^2(X_s)\, ds}. \tag{10}$$

Since $b \in L^2(\mu_{\theta_0})$ it holds in particular that $\|Z\| < \infty$, P_{θ_0}-almost surely. As for (9), use the well-known fact that this condition is always satisfied if Θ is a bounded subset of Euclidean space and d is the Euclidean distance to some power $0 < \beta \leqslant 1$. So by the cited theorem of Nishiyama (1998) the random map Z is indeed uniformly continuous on Θ^* (with respect to the metric d and therefore also with respect to the Euclidean metric).

To conclude the proof, define the new random map Z' on Θ by taking it equal to Z on Θ^* and then extending it continuously to all of Θ. Then Z' is continuous by construction and the continuity in probability of Z implies that Z' is a version of Z.

3. Consistency of the MLE

Let us assume condition (6) and denote by l_t the continuous version of the log-likelihood (5) (see theorem 1). Since the parameter space Θ is compact and l_t is continuous, the maximizer of l_t (the maximum likelihood estimator) exists. In fact, it is already enough for the consistency result that we are going to prove in this section to require the estimator $\hat{\theta}_t$ satisfies

$$\frac{1}{t} l_t(\hat{\theta}_t) \geqslant \sup_{\theta \in \Theta} \frac{1}{t} l_t(\theta) - o_P(1). \tag{11}$$

Of course, we cannot expect to estimate the true parameter θ_0 consistently if we cannot identify it properly. We will assure the identifiability of the parameter by demanding that the function

$$\theta \mapsto \left\| \frac{b_\theta - b_{\theta_0}}{\sigma} \right\|_{L^2(\mu_{\theta_0})} \tag{12}$$

only has a zero at θ_0.

Theorem 2

Suppose that (4) *holds for all* $\theta \in \Theta$ *and* $t \geqslant 0$. *If the Hölder condition* (6) *is satisfied, the estimator* $\hat{\theta}_t$ *satisfies* (11) *and the function* (12) *only has a zero at* θ_0, *then*

$$\hat{\theta}_t \xrightarrow{P_{\theta_0}} \theta_0$$

as $t \to \infty$.

Proof. The law of large numbers for martingales and the ergodic property (2) tell us that with increasing t, the random map l_t/t approaches the limit map

$$l(\theta) = -\frac{1}{2}\left\| \frac{b_\theta - b_{\theta_0}}{\sigma} \right\|^2_{L^2(\mu_{\theta_0})}.$$

We might therefore expect that for large t the estimator $\hat{\theta}_t$ is close to the maximizer of l, which is the true parameter θ_0. To make this reasoning precise we apply corol. 3.2.3 of Van der Vaart & Wellner (1996). Using the Hölder condition (6) and the identifiability assumption we see that l is a continuous function that has a unique maximum at θ_0. The cited result of Van der Vaart & Wellner (1996) therefore implies that to prove the consistency, it suffices to show that

$$\sup_{\theta \in \Theta} \left| \frac{1}{t} l_t(\theta) - l(\theta) \right| \to 0 \tag{13}$$

in (outer) probability as $t \to \infty$. Both l_t and l are continuous in θ so we have established (13) as soon as we have proved that for the countable, dense subset $\Theta^* \subseteq \Theta$ it holds that

$$\sup_{\theta \in \Theta^*} \left| \frac{1}{t} \int_0^t \frac{b_\theta(X_s) - b_{\theta_0}(X_s)}{\sigma(X_s)}\, dW_s \right| \xrightarrow{P_{\theta_0}} 0 \tag{14}$$

and

$$\sup_{\theta \in \Theta^*} \left| \frac{1}{t} \int_0^t \frac{(b_\theta(X_s) - b_{\theta_0}(X_s))^2}{\sigma^2(X_s)}\, ds - \left\| \frac{b_\theta - b_{\theta_0}}{\sigma} \right\|^2_{L^2(\mu_{\theta_0})} \right| \xrightarrow{P_{\theta_0}} 0. \tag{15}$$

To prove (14) we define for every $\theta \in \Theta$ and $t > 0$ the process $M^{\theta,t}$ by

$$M_s^{\theta,t} = \frac{1}{\sqrt{t}} \int_0^{st} \frac{b_\theta(X_u) - b_{\theta_0}(X_u)}{\sigma(X_u)}\, dW_u, \qquad s \geqslant 0.$$

Observe that $M^{\theta,t}$ is a martingale with quadratic variation $\langle M^{\theta,t} \rangle$ given by

$$\langle M^{\theta,t} \rangle_s = \frac{1}{t} \int_0^{st} \frac{(b_\theta(X_u) - b_{\theta_0}(X_u))^2}{\sigma^2(X_u)}\, du.$$

Our aim is first to show that as $t \to \infty$, the random maps

$$\theta \mapsto M_1^{\theta,t} \tag{16}$$

have a weak limit in the space $l^\infty(\Theta^*)$ of bounded functions on Θ^* (see sect. 1.5 of Van der Vaart & Wellner (1996) for basic facts about weak convergence in such spaces). By the ergodic property (2) and the central limit theorem for martingales (see e.g. Kutoyants, 1984, p. 78) or more recently Van Zanten (2000a)) we have finite dimensional convergence of the random maps (16) to a zero-mean Gaussian random map. To see that this may be strengthened to weak convergence in $l^\infty(\Theta^*)$ we apply th. 3.4.2 of Nishiyama (1998). That result implies that if d is a metric on Θ such that

$$\|M^t\| := \sup_{\theta \neq \psi} \frac{\sqrt{\langle M^{\theta,t} - M^{\psi,t} \rangle_1}}{d(\theta, \psi)} = O_P(1) \tag{17}$$

and (9) holds, then the weak convergence takes place in $l^\infty(\Theta^*)$. As in the proof of theorem 1 we take $d(\theta, \psi) = |\theta - \psi|^{\alpha \wedge 1}$. The entropy condition (9) is then satisfied again. By the boundedness of Θ and condition (6) there exists a constant $C > 0$ such that

$$\|M^t\| \leqslant C\sqrt{\frac{1}{t}\int_0^t b^2(X_s)\, ds}.$$

So the ergodic property (2) implies that (17) holds. We may thus conclude that the random maps (16) have a weak limit in $l^\infty(\Theta^*)$. By Slutsky's lemma (see Van der Vaart & Wellner (1996, p. 32) this implies that the random maps $\theta \mapsto M_1^{\theta,t}/\sqrt{t}$ converge weakly to 0 in $l^\infty(\Theta^*)$, which is equivalent to (14).

It remains to show that (15) holds. Observe that (15) can be viewed as a uniform version of the ergodic property (2). With

$$\mathscr{F} = \{f_\theta\colon \theta \in \Theta^*\}, \quad \text{where } f_\theta = \left(\frac{b_\theta - b_{\theta_0}}{\sigma}\right)^2$$

it reads as

$$\sup_{f \in \mathscr{F}} \left| \frac{1}{t}\int_0^t f(X_s)\, ds - \int f\, d\mu_{\theta_0} \right| \xrightarrow{P_{\theta_0}} 0. \tag{18}$$

As was observed in the paper Van Zanten (2000b), such a uniform ergodic theorem can be proved by a straightforward modification of the arguments that were used by Dudley (1984) to prove the "classical" uniform law of large numbers for i.i.d. random variables. Sufficient condition for (18) are then formulated in terms of the bracketing numbers of the class $\mathscr{F}$. Recall that given two functions l and u, the bracket $[l, u]$ is the set of all functions f with $l \leqslant f \leqslant u$. If $\|\cdot\|$ is a norm on $\mathscr{F}$, a bracket $[l, u]$ is called an ε-bracket if $\|u - l\| < \varepsilon$. The bracketing number $N_{[\,]}(\varepsilon, \mathscr{F}, \|\cdot\|)$ is the minimum number of ε-brackets needed to cover $\mathscr{F}$. Lem. 3.4 of Van Zanten (2000b) implies that (18) holds if

$$N_{[\,]}(\varepsilon, \mathscr{F}, \|\cdot\|_{L^1(\mu_{\theta_0})}) < \infty \tag{19}$$

for every $\varepsilon > 0$. To verify (19) we note that $\mathscr{F}$ is a class of functions that is Lipschitz in a parameter. Indeed, by (6) and the boundedness of Θ, there exists a constant $C > 0$ such that

$$|f_\theta(x) - f_\psi(x)| \leqslant Cd(\theta, \psi) b^2(x)$$

for all $\theta, \psi \in \Theta$ and $x \in \mathbb{R}$. Therefore th. 2.7.11 of Van der Vaart & Wellner (1996) may be applied to conclude that

$$N_{[\,]}(2C_\varepsilon \|b\|^2_{L^2(\mu_{\theta_0})}, \mathscr{F}, \|\cdot\|_{L^1(\mu_{\theta_0})}) \leqslant N(\varepsilon, \Theta, d) < \infty$$

for all $\varepsilon > 0$. Thus (15) holds true and the proof is completed.

References

Basu, A. (1983). Asymptotic theory of estimation in nonlinear stochastic differential equations for the multiparameter case. *Sankhyā Ser. A* **45**, 56–65.

Dudley, R. M. (1984). A course on empirical processes. In *École d'été de probabilités de Saint-Flour XII–1982*, 1–142. Springer-Verlag, Berlin.

Gihman, I. I. & Skorohod, A. V. (1972). *Stochastic differential equations*. Springer-Verlag, New York.

Ibragimov, I. A. & Has'minskii, R. Z. (1981). *Statistical estimation, asymptotic theory*. Springer-Verlag, New York.

Kutoyants, Y. A. (1977). Estimation of the trend parameter of a diffusion process in the smooth case. *Theor. Probab. Appl.* **22**, 399–405.

Kutoyants, Y. A. (1984). *Parameter estimation for stochastic processes*. Heldermann Verlag, Berlin.

Lánská, V. (1979). Minimum contrast estimation in diffusion processes. *J. Appl. Probab.* **16**, 65–75.

Liptser, R. S. & Shiryayev, A. N. (1977). *Statistics of random processes I.* Springer-Verlag, New York.

Nishiyama, Y. (1998). Entropy methods for martingales. PhD thesis, Universiteit Utrecht. Published in 2000 as CWI Tract **128**, CWI Amsterdam.

Nishiyama, Y. (1999). A maximal inequality for continuous martingales and *M*-estimation in a Gaussian white noise model. *Ann. Statist*. **27**, 675–696.

Prakasa Rao, B. L. S. (1999). *Statistical inference for diffusion type processes*. Edward Arnold, London.

Prakasa Rao, B. L. S. & Rubin, H. (1981). Asymptotic theory of estimation in nonlinear stochastic differential equations. *Sankhyā Ser. A* **43**, 170–189.

Skorokhod, A. V. (1989). *Asymptotic methods in the theory of stochastic differential equations*. American Mathematical Society, Providence, RI.

Van de Geer, S. (1995). Exponential inequality for martingales, with application to maximum likelihood estimation for counting processes. *Ann. Statist*. **23**, 1779–1801.

Van der Vaart, A. W. & Wellner, J. A. (1996). *Weak convergence and empirical processes with applications to statistics*. Springer-Verlag, New York.

Van Zanten, J. H. (2000a). A multivariate central limit theorem for continuous local martingales. *Statist. Probab. Lett.* **50**, 229–235.

Van Zanten, J. H. (2000b). On the uniform convergence of the empirical density of an ergodic diffusion. *Statist. Inference Stochastic Process.* to appear.

Yoshida, N. (1990). Asymptotic behavior of *M*-estimator and related random field for diffusion process. *Ann. Inst. Statist. Math.* **42**, 221–251.

Received March 2000, in final form October 2000

J. H. van Zanten, CWI, P.O. Box 94079, 1090 GB Amsterdam, The Netherlands.