Discretely Observed Diffusions: Approximation of the Continuous-time Score Function

Helle Sørensen
Department of Theoretical Statistics
University of Copenhagen
E-mail: hsoeren@stat.ku.dk

September 1999

Abstract

We discuss parameter estimation for discretely observed, ergodic diffusion processes where the diffusion coefficient does not depend on the parameter. We propose using an approximation of the continuous-time score function as an estimating function. The estimating function can be expressed in simple terms through the drift and the diffusion coefficient and is thus easy to calculate. Simulation studies show that the method performs well.

Key words: continuous-time score function, diffusion process, discrete observations, estimating function

1 Introduction

This paper is about parameter estimation for discretely observed diffusion models with known diffusion function. The idea is to use an approximation of the continuous-time score function as estimating function.

This idea is very much in the spirit of the early work by Le Breton (1976) and Florens-Zmirou (1989). They both studied the usual Riemann-Itô discretization of the continuous-time log-likelihood function, and Florens-Zmirou (1989) showed that the corresponding estimator is inconsistent when the length of the time interval between observations is constant.

More recently, various methods providing consistent estimators have been developed, *e.g.* methods based on approximations of the true, discrete-time likelihood function (Pedersen, 1995; Aït-Sahalia, 1998); methods based on auxillary models (Gallant and Tauchen, 1996; Gourieroux, Monfort and Renault, 1993); and methods based on estimating functions (Bibby and Sørensen, 1995; Hansen and Scheinkman, 1995; Kessler, 1996; Jacobsen, 1998).

The estimating function discussed in this paper is of the simple, explicit type discussed by Hansen and Scheinkman (1995) and Kessler (1996), that is, on the form $\sum_{i=1}^{n} \mathcal{A}_\theta h(X_{t_{i-1}}, \theta)$ where $\mathcal{A}_\theta$ is the generator of the diffusion. Hansen and Scheinkman (1995) focus on identifiability and asymptotic behaviour of the estimating function whereas Kessler (1996) focus on asymptotic behaviour and efficiency of the estimator.

The main contribution of this paper is to recognize that, with a special choice of h, the estimating function can be interpreted as an approximation to the continuous-time score function. The approximating estimating function is unbiased, it is invariant to data transformations, it provides consistent and asymptotically normal estimators, and it can be explicitly expressed in terms of the drift and diffusion coefficient. The estimating function is also — at least in some cases — available for multi-dimensional processes.

The main objection against the method is the need for a completely known diffusion function. In case of a parameter dependent diffusion function the suggested estimating function is still unbiased and can thus in principle be used but there is no longer justification for using it since the continuous-time likelihood function does not exist.

We present the model and the basic assumptions in Section 2, and the estimating function is derived in Section 3. We give the asymptotic results in Section 4, and examples and simulation studies in Section 5. Sections 2–5 discuss one-dimensional diffusion processes exclusively; we study the multi-dimensional case in Section 6.

2 Model and notation

In this section we present the diffusion model, state the assumptions and introduce some notation.

We consider a one-dimensional, time-homogeneous stochastic differential equation

$$dX_t = b(X_t, \theta)\, dt + \sigma(X_t)\, dW_t, \quad X_0 = x_0 \tag{1}$$

where θ is an unknown p-dimensional parameter from the parameter space $\Theta \subseteq \mathbb{R}^p$ and W is a one-dimensional Brownian motion. The functions

$b : \mathbb{R} \times \Theta \to \mathbb{R}$ and $\sigma : \mathbb{R} \to]0, \infty[$ are known, and the derivatives $\partial\sigma/\partial x$ and $\partial^2 b/\partial\theta_j \partial x$ are assumed to exist for all $j = 1, \ldots, p$. Note that σ does not depend on θ.

We assume that, for any θ, (1) has a unique strong solution X, and that the range of X does not depend on θ. Assume furthermore that there exists a unique invariant distribution $\mu_\theta = \mu(x,\theta)dx$ such that a solution to (1) with $X_0 \sim \mu_\theta$ (instead of $X_0 = x_0$) is strictly stationary. Sufficient conditions for these assumptions to hold can be found in Karatzas and Shreve (1991).

The invariant density is given by

$$\mu(x,\theta) = \big(C(\theta)s(x,\theta)\sigma^2(x)\big)^{-1} \tag{2}$$

where $C(\theta)$ is a normalizing constant and $s(\cdot,\theta)$ is the density of the scale measure, *i.e.* $\log s(x,\theta) = -2\int^x b(y,\theta)/\sigma^2(y)\,dy$.

For all $\theta \in \Theta$, the distribution of X is denoted P_θ if $X_0 = x_0$ (as in (1)) and P_θ^μ if $X_0 \sim \mu_\theta$. Under P_θ^μ, all $X_t \sim \mu_\theta$. E_θ^μ is the expectation wrt. P_θ^μ.

The objective is to estimate θ from observations of X at discrete timepoints $t_1 < \cdots < t_n$. Define $t_0 = 0$ and $\Delta_i = t_i - t_{i-1}$ and let θ_0 be the true parameter.

Finally, we need some matrix notation: Vectors in $\mathbb{R}^p$ are considered as $p \times 1$ matrices, and A^T is the transpose of A. For a function $f = (f_1, \ldots, f_q)^T : \mathbb{R} \times \Theta \to \mathbb{R}^q$ we let $f'(x,\theta)$ be the $q \times 1$ matrix of partial derivatives with respect to x and $\dot{f}(x,\theta) = D_\theta f(x,\theta)$ be the $q \times p$ matrix of partial derivatives with respect to θ, *i.e.* $\dot{f}_{jk}(x,\theta) = \frac{\partial}{\partial\theta_k} f_j(x,\theta)$.

3 The estimating function

In this section we derive a simple, unbiased estimating function as an approximation of the continuous-time score function.

First a comment on the model: It is important that σ does not depend on θ. Otherwise the distributions of $\big(X_s\big)_{0\le s\le t}$ corresponding to two different parameter values are typically singular for all $t \ge 0$. If Y is the solution to $dY_t = b(Y_t,\theta)dt + \tilde{\sigma}(Y_t,\theta)dW_t$, then the process $\big(\int^{Y_t} 1/\tilde{\sigma}(y,\theta)dy\big)_{t\ge 0}$ is the solution to (1) with $\sigma \equiv 1$, but this is of no help for estimation purposes since the transformation depends on the (unknown) parameter.

When σ is completely known as we have assumed, it follows from Lipster and Shiryayev (1977) that the likelihood function for a continuous observation $\big(X_s\big)_{0\le s\le t}$ exists and that the corresponding score process S^c is given by

$$S_t^c(\theta) = \int_0^t \frac{\dot{b}(X_s,\theta)}{\sigma^2(X_s)}\, dX_s - \int_0^t \frac{b(X_s,\theta)\dot{b}(X_s,\theta)}{\sigma^2(X_s)}\, ds.$$

Using (1) we find that

$$dS_t^c(\theta) = \frac{\dot{b}(X_t,\theta)}{\sigma(X_t)}\, dW_t. \tag{3}$$

This shows that $S^c(\theta)$ is a local martingale and a genuine martingale if $\mathrm{E}_\theta^\mu \int_0^t \big(\dot{b}_j(X_s,\theta)/\sigma(X_s)\big)^2\, ds < \infty$ for all $t \geq 0$ and all $j = 1, \ldots, p$, *i.e.* if

$$\mathrm{E}_\theta^\mu \left(\frac{\dot{b}_j(X_0,\theta)}{\sigma(X_0)}\right)^2 = \int \left(\frac{\dot{b}_j(x,\theta)}{\sigma(x)}\right)^2 \mu(x,\theta)\, dx < \infty \tag{4}$$

for all $j = 1, \ldots, p$. In particular, $\mathrm{E}_\theta^\mu\, S_t^c = 0$ for all $t \geq 0$ if (4) holds.

If X was observed continuously on the interval $[0, t_n]$ we would estimate θ by solving the equation $S_{t_n}^c(\theta) = 0$. For discrete observations at time-points $t_1, \ldots, t_n$, the idea is to use an approximation of $S_{t_n}^c$ as estimating function.

The most obvious approximation is obtained by simply replacing the integrals in (3) with the corresponding Riemann and Itô sums,

$$R_n(\theta) = \sum_{i=1}^n \frac{\dot{b}(X_{t_{i-1}},\theta)}{\sigma^2(X_{t_{i-1}})}\big(X_{t_i} - X_{t_{i-1}}\big) - \sum_{i=1}^n \Delta_i \frac{b(X_{t_{i-1}},\theta)\dot{b}(X_{t_{i-1}},\theta)}{\sigma^2(X_{t_{i-1}})}.$$

Note that this would be the score function if the conditional distributions of the increments $X_{t_i} - X_{t_{i-1}}$, given the past, were Gaussian with expectation $\Delta_i b(X_{t_{i-1}},\theta)$ and variance $\Delta_i \sigma^2(X_{t_{i-1}})$. However, usually $\mathrm{E}_\theta^\mu\, R_n(\theta) \neq 0$, and R_n provides inconsistent estimators unless $\sup_{i=1,\ldots,n} \Delta_i \to 0$ (Florens-Zmirou, 1989).

We now propose an unbiased approximation of $S_{t_n}^c$. Let $\mathcal{A}_\theta$ denote the differential operator associated with the infinitesimal generator for X, that is

$$\mathcal{A}_\theta f(x,\theta) = b(x,\theta) f'(x,\theta) + \frac{1}{2}\sigma^2(x) f''(x,\theta)$$

for functions $f : \mathbb{R} \times \Theta \to \mathbb{R}^p$ that are twice continuously differentiable wrt. x.

Recall that μ is the invariant density and assume that the derivatives

$$h^\star = D_\theta \log \mu : \mathbb{R} \times \Theta \to \mathbb{R}^p$$

wrt. the coordinates of θ exist and are twice continuously differentiable wrt. x, such that $\mathcal{A}_\theta h^\star$ is well-defined. The connection between $h^\star$ and S^c is given in the following proposition:

Proposition 3.1 *With respect to P_θ and P_θ^μ, it holds for all $t \geq 0$ that*

$$2\, S_t^c(\theta) = h^\star(X_t, \theta) - h^\star(X_0, \theta) - \int_0^t \mathcal{A}_\theta h^\star(X_s, \theta)\, ds. \tag{5}$$

Proof We show that

$$dh^\star(X_t, \theta) = \mathcal{A}_\theta h^\star(X_t, \theta)\, dt + 2\, dS_t^c(\theta). \tag{6}$$

Then (5) follows immediately since $S_0^c = 0$. Using (2) we easily find the first derivative of $h^\star = D_\theta \log \mu$ in terms of b and σ;

$$h^{\star\prime}(x, \theta) = D_x D_\theta \log \mu(x, \theta) = -D_\theta D_x \log s(x, \theta) = 2\frac{\dot{b}(x, \theta)}{\sigma^2(x)}. \tag{7}$$

Now simply apply Itô's formula on $h^\star$. □

The proposition suggests that we use

$$F_n^\star(\theta) = \frac{1}{2} \sum_{i=1}^n \Delta_i \mathcal{A}_\theta h^\star(X_{t_{i-1}}, \theta),$$

as an approximation to $-S_{t_n}^c$ (since the term $h^\star(X_{t_n}, \theta) - h^\star(X_0, \theta)$ is negligible when n is large) and hence solve the equation $F_n^\star(\theta) = 0$ in order to find an estimator for θ.

The right hand side of (5) with $h^\star$ substituted with an arbitrary function $h \in \mathcal{C}^2(I)$, is a martingale if $\mathrm{E}_\theta^\mu (h'\sigma)^2 < \infty$. Hence,

$$\mathrm{E}_\theta^\mu\, \mathcal{A}_\theta h(X_0, \theta) = 0 \tag{8}$$

if furthermore h and $\mathcal{A}_\theta h$ are in $L^1(\mu_\theta)$. In particular F_n is unbiased, *i.e.* $\mathrm{E}_\theta^\mu\, F_n^\star(\theta) = 0$, if (4) holds and if $h^\star$ and $\mathcal{A}_\theta h^\star$ are in $L^1(\mu_\theta)$.

The moment condition (8) was used by Hansen and Scheinkman (1995) to construct General Method of Moments estimators (their condition C1) and by Kessler (1996) and Jacobsen (1998) to construct unbiased estimating functions. Kessler particularly suggests choosing polynomials h of low degree — regardless of the model. Instead, we suggest the *model-dependent* choice $h = h^\star$. Intuitively, this should be good for small Δ_i's since $F_n \approx -S^c_{t_n}$. Indeed, for $\Delta_i \equiv \Delta$, F_n is small Δ-optimal in the sense of Jacobsen (1998).

It should be clear, though, that moment conditions like (8) cannot achieve asymptotically efficient estimators for a given $\Delta > 0$ since each term in the discrete-time score function involves *pairs* of observations. Note that if the observations were iid. μ_θ-distributed, the score function would equal $\sum_{i=1}^n h^\star(X_{t_i}, \theta)$, which is thus optimal for $\Delta \to \infty$. Kessler (1996) discussed this estimating function.

When $\Delta_i \equiv \Delta$, F_n is a *simple* estimating function, *i.e.* a function of the form $\sum_{i=1}^n f(X_{t_{i-1}}, \theta)$ where $\mathrm{E}_\theta^\mu f(X_0, \theta) = 0$ (Kessler, 1996). In general, the Δ_i's can be interpreted as weights compensating for the dependence between observations that are close in time: an observation is given much weight if it is far in time from the previous one, and little weight if it is close in time to the previous one.

Note that (8) holds so that F_n is unbiased even if σ depends on θ. However, the interpretation of F_n as an approximation of minus the continuous-time score function is of course no longer valid and the method will be non-optimal even for small Δ_i's (Jacobsen, 1998).

A nice property of $F_n^\star$ is that it is invariant to transformations of data; the estimator does not change if we observe $\varphi(X_{t_1}), \ldots, \varphi(X_{t_n})$ instead of $X_{t_1}, \ldots, X_{t_n}$. This is not the case for the polynomial martingale estimating functions discussed by Bibby and Sørensen (1995).

To prove the invariance, we need some further notation: For a diffusion process Y satisfying a stochastic differential equation similar to (1), we write μ_Y and $\mathcal{A}_Y$ for the corresponding invariant density and the differential operator, and define $h_Y^\star = D_\theta \log \mu_Y$.

Proposition 3.2 *Let $\varphi : I \to J \subseteq \mathbb{R}$ be a bijection from $\mathcal{C}^2(I)$ with inverse φ^{-1}, and let $Y = \varphi(X)$. Then*

$$\mathcal{A}_Y h_Y^\star(y, \theta) = \mathcal{A}_X h_X^\star\big(\varphi^{-1}(y), \theta\big). \tag{9}$$

Proof By Itô's formula, Y is the solution to

$$dY_t = b_Y(Y_t, \theta)\, dt + \sigma_Y(Y_t)\, dW_t$$

where, with obvious notation,

$$\begin{aligned} b_Y(y, \theta) &= b_X\big(\varphi^{-1}(y), \theta\big)\varphi'\big(\varphi^{-1}(y)\big) + \frac{1}{2}\big(\sigma_X^2 \varphi''\big)\big(\varphi^{-1}(y)\big), \\ \sigma_Y(y) &= \big(\sigma_X \varphi'\big)\big(\varphi^{-1}(y)\big). \end{aligned}$$

One can now either check directly from (10) that (9) holds or argue as follows. The density for the invariant distribution of $Y = \varphi(X)$ is given by $\mu_Y(y,\theta) = \mu_X\big(\varphi^{-1}(y),\theta\big)\big|\big(\varphi^{-1}\big)'(y)\big|$ and thus

$$h_Y^\star(y,\theta) = D_\theta \log \mu_Y(y,\theta) = D_\theta \log \mu_X\big(\varphi^{-1}(y),\theta\big) = h_X^\star\big(\varphi^{-1}(y),\theta\big).$$

Finally, $\mathcal{A}_Y(f\circ\varphi^{-1})(y) = \mathcal{A}_X f\big(\varphi^{-1}(y)\big)$ for all $f \in \mathcal{C}^2(I)$, which concludes the proof. □

In the following we write $f^\star$ for $(\mathcal{A}_\theta h^\star)/2 = (\mathcal{A}_\theta D_\theta \log\mu)/2$. It is important to note that we have an explicit expression for $f^\star$ — and thus for $F_n = \sum_{i=1}^n \Delta_i f^\star(X_{t_{i-1}},\cdot)$ — in terms of b and σ, even if we have no explicit expression for the normalizing constant $C(\theta)$: from (7) we get

$$f^\star = \mathcal{A}_\theta h^\star/2 = \left(\frac{b\dot{b}}{\sigma^2} + \frac{1}{2}\dot{b}' - \frac{\dot{b}\sigma'}{\sigma}\right). \tag{10}$$

4 Asymptotic properties

In this section we state the asymptotic results for F_n. We consider equidistant observations, $t_i = i\Delta$ where Δ does not depend on n, and let $n \to \infty$.

Under suitable regularity conditions, a solution $\hat{\theta}_n$ to $F_n(\theta) = 0$ exists with a P_{θ_0}-probability tending to 1, and $\hat{\theta}_n$ is a consistent, asymptotically normal estimator for θ. The asymptotic distribution of $\hat{\theta}_n$ is given by

$$\sqrt{n}\left(\tilde{\theta}_n - \theta_0\right) \xrightarrow{\mathcal{D}} N\Big(0, A(\theta_0)^{-1}V(\theta_0)\big(A(\theta_0)^{-1}\big)^T\Big)$$

wrt. P_{θ_0} as $n \to \infty$, where $A(\theta_0) = \mathrm{E}^\mu_{\theta_0} \dot{f}^\star(X_0,\theta_0)$ and

$$V(\theta_0) = \mathrm{E}^\mu_{\theta_0} f^\star(X_0,\theta_0) f^\star(X_0,\theta_0)^T + 2\sum_{k=1}^\infty \mathrm{E}^\mu_{\theta_0} f^\star(X_0,\theta_0) f^\star(X_{k\Delta},\theta_0)^T.$$

Conditions that ensure convergence of the sum in are given by Kessler (1996). If (8) holds for each $\dot{h}_j^\star$, then

$$A(\theta_0) = 2\,\mathrm{E}^\mu_{\theta_0}\left(\frac{\dot{b}(X_0,\theta_0)}{\sigma(X_0)}\right)^T\left(\frac{\dot{b}(X_0,\theta_0)}{\sigma(X_0)}\right)$$

and $A(\theta_0)$ is symmetric and positive semidefinite. It must be positive definite. We will not go through the additional regularity conditions here but refer to Kessler (1996) and particularly to Sørensen (1998).

5 Examples

As already mentioned F_n can always be expressed explicitly in terms of b and σ. When b is linear wrt. the parameter we even get explicit estimators. Assume that $b(x,\theta) = b_0(x) + \sum_{j=1}^p b_j(x)\theta_j$ for known functions $b_0, b_1, \ldots, b_p : \mathbb{R} \to \mathbb{R}$ such that the assumptions of Sections 2 and 3 hold. From (10) we easily deduce that the k'th coordinate of $f^\star$ is given by

$$f_k^\star(x,\theta) = \sum_{j=1}^p \frac{b_k(x)b_j(x)}{\sigma^2(x)}\theta_j + \frac{b_0(x)b_k(x)}{\sigma^2(x)} + \frac{1}{2}b_k'(x) - \frac{b_k(x)\sigma'(x)}{\sigma(x)}.$$

It follows that F_n is linear in θ and it is easy to show that the estimating equation has a unique, explicit solution if and only if $b_1, \ldots, b_p$ are linearly independent.

The Ornstein-Uhlenbeck model and the Cox-Ingersoll-Ross model are special cases of this setup. Several authors have studied inference for these models, see *e.g.* Bibby and Sørensen (1995), Gourieroux *et al.* (1993), Kessler (1996), Jacobsen (1998), Overbeck and Rydén (1997), and Pedersen (1995). From now on, we consider equidistant observations, $\Delta_i \equiv \Delta$.

Example *(The Ornstein-Uhlenbeck process)* Let X be the solution to

$$dX_t = \theta X_t\, dt + dW_t, \quad X_0 = x_0,$$

where $\theta < 0$. The estimator is given by $\hat\theta_n = -n/(2\sum_{i=1}^n X_{(i-1)\Delta}^2)$. $h^\star$ is a eigenfunction for $\mathcal{A}_\theta$ so the simple estimating functions corresponding to $f = h^\star$ and $f = f^\star$ are proportional (and hence provide the same estimator). Kessler (1996) showed that, for all Δ, $\hat\theta_n$ has the least asymptotic variance among estimators obtained from simple estimating equations. This also follows from results in Jacobsen (1998). □

Example *(The Cox-Ingersoll-Ross process)* Consider the solution X to

$$dX_t = (\alpha + \beta X_t)\ dt + \sqrt{X_t}\, dW_t, \quad X_0 = x_0.$$

where $\beta < 0$ and $\alpha \geq 1/2$. The estimating function F_n is given by

$$F_n(\alpha,\beta) = \begin{pmatrix} \left(\alpha - \frac{1}{2}\right)\sum_{i=1}^n \frac{1}{X_{(i-1)\Delta}} + n\beta \\ \beta\sum_{i=1}^n X_{(i-1)\Delta} + n\alpha \end{pmatrix}.$$

To see how the estimator performs we have compared it to three other estimators in a simulation study. We have simulated 500 processes on the interval $[0, 500]$ by the Euler scheme with time-step $1/1000$. The number of observations is $n = 500$ and $\Delta = 1$. For each simulation we have calculated four estimators, namely those obtained from F_n, R_n, $H_n = \sum_{i=1}^{n} h^\star(X_{t_{i-1}}, \cdot)$, and the martingale estimating function suggested by Bibby and Sørensen (1995).

The estimating function H_n is given by

$$H_n(\theta) = 2 \begin{pmatrix} \sum_{i=1}^{n} \log X_{(i-1)\Delta} - n\Psi(2\alpha) + n\log(-2\beta) \\ \sum_{i=1}^{n} X_{(i-1)\Delta} + n\frac{\alpha}{\beta} \end{pmatrix}$$

where Ψ is the Digamma function, $\Psi(x) = \frac{\partial}{\partial x} \log \Gamma(x)$. Note that the second coordinates of F_n and H_n are equivalent and that $H_n(\theta) = 0$ cannot be solved explicitly.

The empirical means and standard errors of the four estimators are listed in Table 1. The true parameter values are $\alpha_0 = 10$ and $\beta_0 = -1$.

Estimating function	$\hat{\alpha}_n$		$\hat{\beta}_n$	
	mean	s.e.	mean	s.e.
F_n	10.1271	0.7218	-1.0126	0.0737
H_n	10.1543	0.7151	-1.0154	0.0729
R_n	6.3691	0.4279	-0.6368	0.0430
Martingale	10.2000	1.1900	-1.0200	0.1200

Table 1: Empirical means and standard errors for 500 realizations of various estimators for (α, β) in the Cox-Ingersoll-Ross model. The number of observations is $n = 500$ and $\Delta = 1$. The true value is $(\alpha_0, \beta_0) = (10, -1)$.

The estimator from R_n is biased (as we knew). F_n and H_n seem to be almost equally good and are both better than the martingale estimating function. □

Finally, we consider an example where the parameter of interest enters as an exponent in the drift function.

Example *(A generalized Cox-Ingersoll-Ross model)* Let X be the solution to

$$dX_t = \left(\alpha + \beta X_t^{\theta}\right) dt + \sqrt{X_t}\, dW_t$$

where $\alpha \geq \frac{1}{2}$ and $\beta < 0$ are known, and $\theta > 0$ is the unknown parameter. X is a Cox-Ingersoll-Ross process if $\theta = 1$; for $\theta \neq 1$ the mean reverting force is stronger or weaker.

From (10) it follows that $f^\star = (\mathcal{A}_\theta h^\star)/2$ is given by

$$f^\star(x,\theta) = \beta x^{\theta-1} \log x \Big(\alpha + \beta x^\theta + \frac{\theta}{2} - \frac{1}{2}\Big) + \frac{1}{2}\beta x^{\theta-1}.$$

The estimating equation must be solved numerically. For comparison we have also considered the simple estimating function corresponding to

$$\tilde{f}(x,\theta) = x - \mathrm{E}_\theta^\mu X_0 = x - \frac{\Gamma\big((2\alpha+1)/\theta\big)}{\Gamma\big(2\alpha/\theta\big)} \left(-\frac{\theta}{2\beta}\right)^{1/\theta} \tag{11}$$

As above, we have simulated 500 processes by means of the Euler scheme; $n = 500$ and $\Delta = 1$. The true value of θ is $\theta_0 = 1.5$ and $\alpha = 2$, $\beta = -1$. The means (standard errors) of the estimators are 1.5028 (0.0508) when using F_n and 1.5001 (0.0590) when using $\sum \tilde{f}(X_{(i-1)\Delta}, \cdot)$.

Both estimators are very precise. The estimator obtained from (11) is closer to the true value but has larger standard error than the estimator obtained from F_n. □

6 Multi-dimensional processes

So far, we have only studied one-dimensional diffusion processes. In this section we discuss to what extend the ideas carry over to the multi-dimensional case.

We consider a d-dimensional stochastic differential equation

$$dX_t = b(X_t,\theta)\,dt + \sigma(X_t)\,dW_t, \quad X_0 = x_0. \tag{12}$$

The parameter θ is still p-dimensional, $\theta \in \Theta \subseteq \mathbb{R}^p$, but X and W are now d-dimensional. The functions $b : \mathbb{R}^d \times \Theta \to \mathbb{R}^d$ and $\sigma : \mathbb{R}^d \to \mathbb{R}^{d\times d}$ are known, $\sigma(x)$ is regular for all $x \in \mathbb{R}^d$, and $x_0 \in \mathbb{R}^d$.

Let $\dot{b}$ be the $d \times p$ matrix of derivatives; $\dot{b}_{ij} = \partial b_i/\partial\theta_j$, and let $D_i g = \partial g/\partial x_i$ and $D^2_{ij} g = \partial g^2/\partial x_i \partial x_j$ for functions $g : \mathbb{R}^d \times \Theta \to \mathbb{R}$ with $g(\cdot,\theta)$ in $\mathcal{C}^2(\mathbb{R}^d)$. With this notation the analogue of $\mathcal{A}_\theta$ is given by

$$\mathcal{A}_\theta g(x,\theta) = \sum_{i=1}^d b_i(x,\theta) D_i g(x,\theta) + \frac{1}{2}\sum_{i,j=1}^d \big(\sigma(x)\sigma^T(x)\big)_{ij} D^2_{ij} g(x,\theta)$$

and the score process is given by

$$S_t^c(\theta) = \int_0^t \dot{b}^T(X_s,\theta)\Sigma(X_s)\,dX_s - \int_0^t \dot{b}^T(X_s,\theta)\Sigma(X_s)b(X_S,\theta)\,ds,$$

where $\Sigma(x) = \left(\sigma(x)\sigma^T(x)\right)^{-1}$, see Lipster and Shiryayev (1977). Using (12) we find $dS_t^c(\theta) = \dot{b}^T(X_t,\theta)(\sigma^{-1})^T(X_t)\,dW_t$.

Now, similarly to (6) we look for functions $h_1^\star,\dots,h_p^\star : \mathbb{R}^d \times \Theta \to \mathbb{R}$ such that for each k, $dh_k^\star(X_t,\theta) = \mathcal{A}_\theta h_k^\star(X_t,\theta)\,dt + 2\,dS_{k,t}^c(\theta)$. Arguing as above this leads to the equations

$$D_i h_k^\star(x,\theta) = 2\left[\dot{b}^T(x,\theta)\Sigma(x)\right]_{ki} = 2\sum_{r=1}^d \dot{b}_{rk}(x,\theta)\Sigma_{ir}(x), \quad i = 1,\dots,d \tag{13}$$

and thus

$$\mathcal{A}_\theta h_k^\star = 2\sum_{i,r=1}^d \dot{b}_{rk}\Sigma_{ir}b_i + \sum_{j=1}^d \frac{\partial \dot{b}_{jk}}{\partial x_j} + \sum_{i,j,r=1}^d \left(\sigma\sigma^T\right)_{ij}\dot{b}_{rk}\frac{\partial \Sigma_{ir}}{\partial x_j} \tag{14}$$

The equations (13) may, however, not have solutions; differentiation wrt. x_j yields

$$D_{ij}^2 h_k^\star = 2\sum_{r=1}^d \left(\frac{\partial^2 b_r}{\partial\theta_k \partial x_j}\Sigma_{ir} + \frac{\partial b_r}{\partial \theta_k}\frac{\partial \Sigma_{ir}}{\partial x_j}\right),$$

but the right hand side is not necessarily symmetric wrt. i and j, see the example below.

If there *are* solutions, then (14) has expectation zero and the simple estimating function with $f = \left(\mathcal{A}_\theta h_1^\star,\dots,\mathcal{A}_\theta h_p^\star\right)^T$ may be used. Otherwise, the right hand side of (14) is typically biased.

Example *(Homogeneous Gaussian diffusions)* Let B be a 2×2-matrix with eigenvalues with strictly negative parts, and let A be an arbitrary 2×1 matrix. Consider the stochastic differential equation

$$dX_t = (A + BX_t)\,dt + \sigma dW_t, \quad X_0 = x_0$$

where $\sigma > 0$ is known, W is a two-dimensional Brownian motion, and $x_0 \in \mathbb{R}^2$.

Solutions to all the equations (13) exist if and only if B is symmetric. Let $\alpha_1, \alpha_2, \beta_{11}, \beta_{22}, \beta_{12}$ denote the entries of A and B and let $S_1 = \sum X_{1,(i-1)\Delta}$, $S_2 = \sum X_{2,(i-1)\Delta}$, $S_{11} = \sum X^2_{1,(i-1)\Delta}$, $S_{22} = \sum X^2_{2,(i-1)\Delta}$, and $S_{12} = \sum X_{1,(i-1)\Delta} X_{2,(i-1)\Delta}$; all sums are from 1 to n. Then the estimating equation is given by

$$\frac{1}{\sigma^2}\begin{pmatrix} n & 0 & S_1 & 0 & S_2 \\ 0 & n & 0 & S_2 & S_1 \\ S_1 & 0 & S_{11} & 0 & S_{12} \\ 0 & S_2 & 0 & S_{22} & S_{12} \\ S_2 & S_1 & S_{12} & S_{12} & S_{11}+S_{22} \end{pmatrix}\begin{pmatrix} \alpha_1 \\ \alpha_2 \\ \beta_{11} \\ \beta_{22} \\ \beta_{12} \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ -n/2 \\ -n/2 \\ 0 \end{pmatrix}.$$

We have simulated 500 processes (by exact simulation), each of a length of 500 with $\Delta = 1$ and $\sigma = \sqrt{2}$. The true matrices are

$$A_0 = \begin{pmatrix} 4 \\ 1 \end{pmatrix} \quad \text{and} \quad B_0 = \begin{pmatrix} -2 & 1 \\ 1 & -3 \end{pmatrix}.$$

The means and the standard errors (to the right) are

$$\hat{A}_n = \begin{pmatrix} 4.0349 \\ 1.0035 \end{pmatrix} \qquad \begin{matrix} 0.2904 \\ 0.2891 \end{matrix}$$

and

$$\hat{B}_n = \begin{pmatrix} -2.0155 & 1.0078 \\ 1.0078 & -3.0247 \end{pmatrix} \qquad \begin{matrix} 0.1248 & 0.1177 \\ 0.1177 & 0.1978 \end{matrix}$$

The estimators are good. □

Acknowledgements I am grateful to my supervisor Martin Jacobsen for valuable discussions and suggestions, and to Jens Lund and Martin Richter for comments on an early version of the manuscript.

References

Aït-Sahalia, Y. (1998), Maximum likelihood estimation of discretely sampled diffusions: a closed-form approach, Technical report, Graduate School of Business, University of Chicago. Revised version of Working Paper 467.

Bibby, B. M. and Sørensen, M. (1995), Martingale estimation functions for discretely observed diffusion processes, *Bernoulli* **1**, 17–39.

Florens-Zmirou, D. (1989), Approximate discrete-time schemes for statistics of diffusion processes, *Statistics* **20**, 547–557.

Gallant, A. R. and Tauchen, G. (1996), Which moments to match?, *Econometric Theory* **12**, 657–681.

Gourieroux, C., Monfort, A. and Renault, E. (1993), Indirect inference, *Journal of Applied Econometrics* **8**, 85–118.

Hansen, L. P. and Scheinkman, J. A. (1995), Back to the future: generating moment implications for continuous-time Markov processes, *Econometrica* **63**, 767–804.

Jacobsen, M. (1998), Discretely observed diffusions: classes of estimating functions and small Δ-optimality, Preprint 11, Department of Theoretical Statistics, University of Copenhagen.

Karatzas, I. and Shreve, S. E. (1991), *Brownian Motion and Stochastic Calculus*, Springer-Verlag.

Kessler, M. (1996), Simple and explicit estimating functions for a discretely observed diffusion process, Research report 336, University of Århus. To appear in *Scand. J. Statist.*

Le Breton, A. (1976), On continuous and discrete sampling for parameter estimation in diffusion type processes, *Mathematical Programming Study* **5**, 124–144.

Lipster, R. S. and Shiryayev, A. N. (1977), *Statistics of Random Processes I: General Theory*, Springer-Verlag.

Overbeck, L. and Rydén, T. (1997), Estimation in the Cox-Ingersoll-Ross model, *Econom. Theory* **13**, 430–461.

Pedersen, A. R. (1995), A new approach to maximum likelihood estimation for stochastic differential equations based on discrete observations, *Scand. J. Statist.* **22**, 55–71.

Sørensen, M. (1998), On asymptotics of estimating functions, Preprint 6, Department of Theoretical Statistics, University of Copenhagen. Submitted for publication.
`http://www.math.ku.dk/~michael/estrew.ps`.