On effects of discretization on estimators of drift parameters for diffusion processes*

P. E. Kloeden[1], E. Platen[2], H. Schurz[3] and M. Sørensen[4]

submitted December 6, 1996

[1] School of Computing & Mathematics
Deakin University
Clayton, Vic 3168
Australia
e–mail: kloeden@deakin.edu.au

[2] School of Mathematical Sciences
Australian National University
Canberra, ACT 0200
Australia
e–mail: eckhard@orac.anu.edu.au

[3] Weierstrass Institute for
Applied Analysis and Stochastics
Mohrenstr. 39
D – 10117 Berlin
Germany
e–mail: schurz@wias–berlin.de

[4] Department of Theoretical Statistics
University of Aarhus
DK 8000 Aarhus C
Denmark
e–mail: michael@mi.aau.dk

Preprint No. 292
Berlin 1996

1991 Mathematics Subject Classification. 60H10, 62F12, 65U05.

Keywords. Discrete time sampling, Inference for stochastic processes, Maximum likelihood estimation, Numerical methods, Simulation, Stochastic differential equations, Stochastic Taylor expansions.

*This paper is being published in Journal of Applied Probability, December 1996, and was originally written in 1991 and 1992 (cf. preprint of University of Aarhus). The current presentation more or less coincides with that form appeared in Journal of Applied Probability, except for a few updates concerning references.

SUMMARY. In this paper statistical properties of estimators of drift parameters for diffusion processes are studied by modern numerical methods for stochastic differential equations. This is a particularly useful method for discrete time samples, where estimators can be constructed by making discrete time approximations to the stochastic integrals appearing in the maximum likelihood estimators for continuously observed diffusions. A review is given of the necessary theory for parameter estimation for diffusion processes and for simulation of diffusion processes.
Three examples are studied.

1 Introduction

In recent years significant progress has been made in the theory of numerical simulation of diffusion processes. An extensive presentation of the state of the art can be found in [17], [M], [T] or [A]. In the present paper we propose to use these new simulation methods to study the behavior of estimators in diffusion process models. The availability of fast computers makes this a feasible procedure. Related work has been done by [13], [14] and [2].

Several authors have studied the problem of estimating parameters in the drift coefficient when a diffusion process has been observed continuously, i.e. at all time points in an interval. In this situation an explicit expression for the likelihood function can be given under weak conditions. Reviews can be found in [1] and [19]. A particularly nice theory is obtained if the drift coefficient depends linearly on the parameters. This type of model is common in practice.

In the more realistic situation where a diffusion process has been observed at discrete time points only, an explicit likelihood function is rarely available. This type of data has recently attracted considerable interest; see [22], [4], [12], [28], [5], [6], [8], [26] and [2]. A simple method of obtaining an estimator for discrete time that is often used in practice is to construct, from the available data, an approximation to the estimator found in the theory for continuous observations. This involves, in particular, discrete time approximations to stochastic integrals. It is, in general, difficult to study the quality of such estimators analytically, see [22], [5] and [7], but it can be done easier by numerical simulation.

In section 2 we introduce our three examples and give the continuous time version of the maximum likelihood estimators of the parameters. It is also discussed how to construct approximate maximum likelihood estimators for discrete time observations. In Section 3 we review those aspects of the theory of numerical simulation of diffusion processes that are relevant to our purposes. In particular, a few important simulation schemes are given explicitly. In Section 4, simulation of the diffusion processes considered in the examples of Section 2 is discussed and sample paths of the processes and of the estimators found in Section 2 are simulated. In the appendix a brief summary is given of the theory of maximum likelihood estimation based on continuous observation for a class of stochastic differential equations large enough to cover our examples.

Generalized diffusion processes where discontinuous sample paths are allowed are called diffusion with jumps. These more general processes are of importance in many fields of application. Also for these processes there exists a nice likelihood theory for continuous observation ([31]) and a theory for numerical simulation

([15]), so the ideas outlined in the present paper could be extended to diffusions with jumps.

2 Some diffusion models and their parameter estimators

In this section we give three examples of parametric statistical models defined by stochastic differential equations. For each model the maximum likelihood estimator based on continuous observation of the process in the time interval $[0,T]$ is given. In the appendix a brief summary of the theory of maximum likelihood estimation for a class of stochastic differential equations large enough to cover all our examples can be found.

Example 2.1 *Linear stochastic differential equation.*

Consider the stochastic differential equation

$$d\,X_t = (\theta_1 X_t + \theta_2)dt + c\,dW_t, \tag{2.1}$$

where c is known. The solution is an Ornstein-Uhlenbeck process.

For the model (2.1) we find the maximum likelihood estimators

$$\begin{aligned}\hat{\theta}_{1T} &= \left\{T\int_0^T X_t dX_t - (X_T - X_0)\int_0^T X_t\,dt\right\}/N_T \\ &= \left\{\frac{1}{2}\,T(X_T^2 - X_0^2 - c^2T) - (X_T - X_0)\int_0^T X_t\,dt\right\}/N_T\end{aligned} \tag{2.2}$$

$$\begin{aligned}\hat{\theta}_{2T} &= \left\{(X_T - X_0)\int_0^T X_t^2\,dt - \int_0^T X_t\,dX_t \int_0^T X_t\,dt\right\}/N_T \\ &= \left\{(X_T - X_0)\int_0^T X_t^2\,dt - \frac{1}{2}\,(X_T^2 - X_0^2 - c^2T)\int_0^T X_t\,dt\right\}/N_T\end{aligned} \tag{2.3}$$

where

$$N_T = T\int_0^T X_t^2\,dt - \left(\int_0^T X_t\,dt\right)^2, \tag{2.4}$$

and where we have used Ito's formula. □

Example 2.2 *Duffing-van der Pol oscillator*

The velocity $X_t^{(2)}$ of a Duffing-van der Pol oscillator is determined by the stochastic differential equation

$$dX_t^{(2)} = \left[\alpha X_t^{(1)} - (X_t^{(2)} + (X_t^{(1)})^3)\right]dt + \sigma X_t^{(1)}\,dW_t, \tag{2.5}$$

where

$$X_t^{(1)} = X_0^{(1)} + \int_0^t X_s^{(2)}\,ds \tag{2.6}$$

is the position of the oscillator (displacement).

We assume that σ is known. The maximum likelihood estimator for α is

$$\hat{\alpha}_T = T^{-1}\left\{\int_0^T \left[X_t^{(1)}\right]_+ d\tilde{X}_t + \int_0^T (X_t^{(1)})^2 \, dt\right\}, \tag{2.7}$$

where

$$\tilde{X}_t = X_t^{(2)} + \int_0^t X_s^{(2)} \, ds \tag{2.8}$$

and

$$\left[X_t^{(1)}\right]_+ = \left\{ \begin{array}{ll} (X_t^{(1)})^{-1} & \text{if } X_t^{(1)} \neq 0. \\ 0 & \text{otherwise} \end{array} \right\}. \tag{2.9}$$

Note that $\hat{\alpha}_T = \alpha + T^{-1}\sigma \; W_T$ so that $\hat{\alpha} \sim N(\alpha, T^{-1}\sigma^2)$ exactly and not only asymptotically.

For trajectories where $X^{(1)}$ does not cross the level zero in the time interval $[0, T]$, we can apply Ito's formula to obtain

$$\int_0^T X_t^{(1)^{-1}} \; dX_t^{(2)} = \frac{X_T^{(2)}}{X_T^{(1)}} - \frac{X_0^{(2)}}{X_0^{(1)}} + \int_0^T \left(\frac{X_t^{(2)}}{X_t^{(1)}}\right)^2 \; dt. \tag{2.10}$$

When this relation holds, we obtain from (2.7) that

$$\hat{\alpha}_T = T^{-1}\left\{\frac{X_T^{(2)}}{X_T^{(1)}} - \frac{X_0^{(2)}}{X_0^{(1)}} + \int_0^T \left(\left(\frac{X_t^{(2)}}{X_t^{(1)}}\right)^2 + \frac{X_t^{(2)}}{X_t^{(1)}} + \left(X_t^{(1)}\right)^2\right) dt\right\}. \tag{2.11}$$

□

Example 2.3 *Population model*

A logistic population model of the diffusion branching type is given by

$$dX_t = \alpha X_t(K - X_t)dt + \sigma\sqrt{X_t} \, dW_t, \tag{2.12}$$

where K is the carrying capacity of the environment. The drift coefficient can be written as in (A.1) with $\theta_1 = \alpha K, \theta_2 = -\alpha$.

It is assumed that σ is known.

For the logistic population model (2.12) the maximum likelihood estimators are

$$\begin{aligned} \hat{\theta}_{1,T} &= \left\{(X_T - X_0)\int_0^T X_t^3 \, dt - \int_0^T X_t \, dX_t \int_0^T X_t^2 \, dt\right\} N_T \qquad (2.13) \\ &= \left\{(X_T - X_0)\int_0^T X_t^3 dt - \frac{1}{2}\left(X_T^2 - X_0^2 - \sigma^2 \int_0^T X_t \, dt\right)\int_0^T X_t^2 \, dt\right\}/N_T \\ \hat{\theta}_{1,T} &= \left\{\int_0^T X_t \, dX_t \int_0^T X_t \, dt - (X_T - X_0)\int_0^T X_t^2 \, dt\right\}/N_T \qquad (2.14) \\ &= \left\{\frac{1}{2}\left(X_T^2 - X_0^2 - \sigma^2 \int_0^T X_t \, dt\right)\int_0^T X_t \, dt - (X_T - X_0)\int_0^T X_t^2 \, dt\right\}/N_T \end{aligned}$$

where

$$N_T = \int_0^T X_t \, dt \cdot \int_0^T X_t^3 \, dt - \left(\int_0^T X_t^2 \, dt \right)^2 , \tag{2.15}$$

and where we have used Ito's formula. □

In practice continuous trajectories are usually not observed. Rather the state of the diffusion process is observed at a finite number of times $t_0 < t_1 < ... < t_n$. The exact likelihood function corresponding to such data is the product of transition densities which can only rarely be found explicitly. A simple estimation procedure for such data is to use the continuous time maximum likelihood estimator (A.4) with $t_0 = 0$ and $t_n = T$ and with suitable approximations to the integrals in H_T and S_T given by (A.2) and (A.3). If the spacings between consecutive observation times are small, it seems likely that some of the good properties of the continuous time maximum likelihood estimator are preserved, although its discrete time version will be biased to a certain extent; see [5] and [22].

It is well-known how to approximate the Riemann integrals in S_T and H_T by Riemann sums or quadrature formulae like the trapezoidal formula. Also the stochastic integral in H_T can be approximated by a finite sum. However, it is preferable to replace the stochastic integral by Riemann and Stieltjes integrals when possible, and then to approximate these. This can in most cases be done for a one-dimensional diffusion by direct use of Ito's formula as was done in the examples above. It can, however, be done only in special cases when X is multidimensional.

3 Numerical approximation of stochastic differential equations

3.1 Basic notions

Our final aim in the paper will be to test the practical behaviour and the consistency of the estimators proposed above. For this purpose we will need to generate the solution X of a stochastic differential equation

$$dX_t = a(X_t) \, dt + b(X_t) \, dW_t \tag{3.1}$$

for $t \in [0, T]$ and $X_0 = x_0$. Specifically, we want to simulate the values of trajectories which represent such solution at discrete time instants.

Unfortunately, it is only in rare cases that one knows an explicit solution for a stochastic differential equation. But we shall see that it is possible to simulate approximate solutions Y^Δ at discrete time instants which, for vanishing time step size ($\Delta \to 0$), converge to the exact solution X.

For studying estimators of statistical parameters it is necessary to generate pathwise approximations. Such approximations are called strong approximations. Let us introduce what we mean by the order of strong convergence. We say that a discrete time approximation Y^Δ converges with strong order $\gamma > 0$ at time T if there exist positive constants $K < \infty$ and $\delta_0 < T$ such that for all $\Delta \in (0, \delta_0)$ we have the inequality

$$E(|X_T - Y^\Delta(T)|) \le K\Delta^\gamma, \tag{3.2}$$

where K does not depend on the time step size Δ. If X and Y^Δ are deterministic, then (3.2) represents the usual deterministic criterion. For controlling constant K with increasing time T, see recent contribution [D].

Paper [23] was one of the first to develop numerical schemes solving stochastic differential equations. [33] and [27] obtained general results which allow construction of strong approximations of any desired strong order, as long as enough 'smoothness' of model components is present. Also [3], [32], [25] and [24] studied strong approximations. An extensive discussion of higher order methods for numerical solution of stochastic differential equations is given in [17], where several additional references can also be found. Compare also with [A], [M], [D] and [T].

In the following we will give a short review presenting basic strong numerical schemes which we will apply later to perform numerical simulations. For convenience and ease of autonomous integration, we shall restrict attention to equidistant time Discretization of the interval $[0, T]$ with time points

$$t_n = n\Delta, \tag{3.3}$$

for $n = 1, ..., n_T$ with step size $\Delta = T/n_T$ for some integer n_T. The simplest discrete time approximation is given by the Euler scheme

$$Y_{n+1} = Y_n + a(Y_n)\Delta + b(Y_n)\Delta W_n \tag{3.4}$$

for $n = 0, 1, 2, ...$ with $Y_0^\Delta = X_0$. Here $\Delta W_n = W_{t_{n+1}} - W_{t_n}$ is the current increment of the Wiener process and is normally distributed with mean zero and variance $E((\Delta W_n))^2) = \Delta$. This scheme provides a recursive algorithm for simulating approximate solutions of the stochastic equation (3.1). It turns out that under Lipschitz and growth conditions an a and b, which also ensure the unique existence of the (strong) solution X of (3.1), the Euler approximation converges with strong order $\gamma = 0.5$. This is in contrast to order 1 for deterministic Euler scheme and is a consequence of the difference between deterministic and stochastic calculi.

3.2 Higher order strong approximations

Truncated stochastic Taylor expansions provide a general systematic means of deriving numerical schemes for stochastic differential equations. These are based on the Ito-Taylor formula proposed in [33] or on the Stratonovich-Taylor formula presented in [16]. In both of these formulae functions of an Ito process are represent in terms of multiple stochastic integrals. Here we will not go into the details of stochastic Taylor expansions, but refer to the literature cited above.

The simplest useful scheme which can be obtained by truncation of a stochastic Taylor expansion is the Euler scheme (3.4), which as mentioned above is of strong order of convergence $\gamma = 0.5$. Inclusion of one more term from the stochastic Taylor expansion yields the order $\gamma = 1.0$ strong Taylor scheme which was originally proposed by [23]. In the one-dimensional case it has the Ito version

$$Y_{n+1} = Y_n + a(Y_n)\Delta + b(Y_n)\Delta W_n + \frac{1}{2}b(Y_n)b'(Y_n)\{(\Delta W_n)^2 - \Delta\} \tag{3.5}$$

and the Stratonovich version

$$Y_{n+1} = Y_n + \underline{a}(Y_n)\Delta + b(Y_n)\Delta W_n + \frac{1}{2}b(Y_n)b'(Y_n)(\Delta W_n)^2, \tag{3.6}$$

where $\underline{a} = a - \frac{1}{2}bb'$ is the corrected drift from the Stratonovich stochastic differential equation corresponding to the Ito equation (3.1).

If we introduce the differential operator

$$L^1 = \sum_{k=1}^{d} b^k \frac{\partial}{\partial x^k}, \tag{3.7}$$

then we write the d-dimensional version of the Milstein scheme in its Stratonovich form as

$$Y_{n+1} = Y_n + \underline{a}(Y_n)\Delta + b(Y_n)\Delta W_n + \frac{1}{2}L^1 b(Y_n)(\Delta W_n)^2 \tag{3.8}$$

with $\underline{a} = a - \frac{1}{2}L^1 b$. The Milstein schemes (3.5), (3.6) and (3.8) converge with strong order $\gamma = 1.0$ under sufficient regularity of the coefficients. These are e.g. Lipschitz and linear growth conditions on $\underline{a}$, $\frac{\partial}{\partial x^k}\underline{a}$, b, $\frac{\partial}{\partial x^k}b$ and $\left(\frac{\partial}{\partial x^k}\right)^2 b$.

By including further terms from Ito-Taylor expansion we can achieve a strong order $\gamma = 1.5$ with the following scheme under further assumptions. For the 1-dimensional case the order $\gamma = 1.5$ strong Taylor scheme is given by

$$\begin{aligned} Y_{n+1} = Y_n \quad &+ \quad a\Delta + b\Delta W_n + \frac{1}{2}bb'\{(\Delta W_n)^2 - \Delta\} \\ &+ \quad ba'\Delta Z_n + \frac{1}{2}\{aa' + \frac{1}{2}b^2 a''\}\Delta^2 \\ &+ \quad \{ab' + \frac{1}{2}b^2 b''\}\{\Delta W_n \Delta - \Delta Z_n\} \\ &+ \quad \frac{1}{2}b(bb')'\{\frac{1}{3}(\Delta W_n)^2 - \Delta\}\Delta W_n \end{aligned} \tag{3.9}$$

where the coefficients $a, b, \ldots$ are taken at Y_n. Here an additional random variable ΔZ_n is required to represent the double stochastic integral

$$\Delta Z_n = \int_{t_n}^{t_{n+1}} \int_{t_n}^{s_2} dW_{s_1}\, ds_2, \tag{3.10}$$

which is Gaussian distributed with mean $E(\Delta Z_n) = 0$, variance $E((\Delta Z_n)^2) = \frac{1}{3}\Delta^3$ and with covariance $E(\Delta Z_n \Delta W_n) = \frac{1}{2}\Delta^2$. The pairs $(\Delta W_n, \Delta Z_n)$ are mutually independent for $n = 1, 2, \ldots$.

We note that a Stratonovich version of (3.9) has a little advantage since it already involves most of the terms appearing in the following order $\gamma = 2.0$ strong Stratonovich-Taylor scheme in the d-dimensional form

$$\begin{aligned} Y_{n+1} = Y_n &+ \underline{a}\Delta + \frac{1}{2}\underline{L}^0\, \underline{a}\Delta^2 + b\Delta W_n + L^1\underline{a}\Delta\, Z_n \\ &+ \underline{L}^0 b\ \{\Delta W_n \Delta - \Delta Z_n\} + \tfrac{1}{2}L^1 b\,(\Delta W_n)^2 \\ &+ \tfrac{1}{3!}L^1 L^1 b(\Delta W_n)^3 + \tfrac{1}{4!}L^1 L^1 L^1 b(\Delta W_n)^4 \\ &+ \underline{L}^0 L^1 b J_{(0,1,1),n} + L^1 \underline{L}^0 b J_{(1,0,1),n} \\ &+ L^1 L^1 \underline{a} J_{(1,1,0),n}, \end{aligned} \tag{3.11}$$

where

$$\underline{L}^0 = \sum_{k=1}^{d} \underline{a}^k \frac{\partial}{\partial x^k}. \tag{3.12}$$

Here we use the multiple Stratonovich integrals

$$J_{(j_1,j_2,j_3),n} = \int_{t_n}^{t_{n+1}} \int_{t_n}^{s_3} \int_{t_n}^{s_2} \circ\, dW_{s_1}^{j_1} \circ dW_{s_2}^{j_2} \circ dW_{s_3}^{j_3} \tag{3.13}$$

$j_1, j_2, j_3 \in \{0,1\}$ with $\circ dW_s^0 = ds$ and $\circ dW_s^1 = \circ dW_s$. The last symbol denotes Stratonovich integration with respect to W.

The random variables $J_{(j_1,j_2,j_3),n}$ can be approximated as accurately as needed by a method described in Kloeden, Platen and Wright (1992) which uses series expansions of the Wiener process. We remark that the Stratonovich version (3.11) of an order $\gamma = 2.0$ strong Taylor scheme is more convenient than its counterpart derived from the Ito-Taylor expansion. In many practical situations it turns out that the schemes become much simpler for specific equations as we shall see in our examples.

For an easy and reliable implementation it is convenient to have schemes which avoid derivatives of the coefficients as Runge-Kutta schemes do in the deterministic case. However, it must be emphasized that they cannot be taken as simple heuristic stochastic adaptations of the deterministic Runge-Kutta schemes as it can be seen from Rümelin (1982). A systematic treatment of this problem can be found in Kloeden and Platen (1992), and here we describe only an explicit order $\gamma = 1.0$ strong scheme which we will use in our simulations. In the d-dimensional case it has the form

$$Y_{n+1} = Y_n + \underline{a}(Y_n)\Delta + \tfrac{1}{2}\Big\{b\Big(Y_n + \underline{a}(Y_n)\Delta + b(Y_n)\Delta W_n\Big) + b(Y_n)\Big\}\Delta W_n. \tag{3.14}$$

Finally, we remark that it might be necessary to simulate multivariate stochastic dynamic systems involving coordinates with extremely different time scales. Such stochastic differential equation systems are considered by Kloeden and Platen (1992) who call them stiff stochastic differential equations. The stiffness of a systems of equations causes numerical instabilities for the schemes described above, which can be avoided by applying implicit strong schemes; see Kloeden and Platen (1992). For more recent analysis of implicit methods, see [E], [D].

4 Case studies

In the following we will generate approximate solutions for the stochastic differential equations given in our Examples 2.1, 2.1 and 2.3 using the numerical schemes described above. From these simulated sample paths we will estimate the parameters by the approximate maximum likelihood estimators described in Section 2. We approximate time integrals by the trapezoidal formula. Other quadrature formulae could be used as well. One can not give a simple rule for how to choose an appropriate numerical scheme and a corresponding step size for the different examples. One could try to implement higher order schemes which, of course, produce a smaller systematic error on finite time-intervals. The choice of the step size depends greatly on the computer time available. Also with other schemes and sufficiently small step size one obtains results similar to those described in the following. There is also some hint to prefer implicit trapezoidal or midpoint rules, cf. [D], to a certain extent.

Example 2.3 *(continued).* For the equation (2.1) we simulate, in $[0, T]$ with $T = 2000$, the corresponding Ornstein-Uhlenbeck process with the parameters $\theta_1 = -1.0, \theta_2 = 1.0$ and $c = 1.0$ starting $X_0 = 0$. Equation (2.1) has additive noise, which implies several simplifications in higher order schemes. Therefore, we choose the order 2.0 strong Taylor scheme (3.11), which in this example has the form

$$Y_{n+1} = Y_n + \{\theta_1 Y_n + \theta_2\} \Delta + \tfrac{1}{2}\theta_1 (\theta_1 Y_n + \theta_2) \Delta^2 + c\Delta W_n + c\theta_1 \Delta Z_n \tag{4.1}$$

with time step size $\Delta = 2^{-4}$. We note that $\underline{a} = a$ and that in (4.1) we do not need to generate the multiple integrals $J_{(1,1,0),n}$, $J_{(1,0,1),n}$ and $J_{(0,1,1),n}$ contained in (3.11) because the corresponding terms vanish. This saves a lot of computer time compared with the implementation of (3.11) for a general stochastic differential equation. A simulated trajectory of (2.1) is shown in Figure 1.

To estimate the parameters θ_1 and θ_2 we apply the estimators given by the last expressions in (2.2) and (2.3) and obtain, after approximation of the relevant time integrals by the trapezoidal formula, the approximate estimators

$$\hat{\theta}^{\Delta}_{1,T} = \frac{1}{2N^{\Delta}_T} \left\{ T\left(X_T^2 - X_0^2 - c^2 T\right) - (X_T - X_0) \sum_{n=0}^{n_T - 1} \left(X_{t_{n+1}} + X_{t_n}\right) \Delta \right\} \tag{4.2}$$

and

$$\begin{aligned} \hat{\theta}^{\Delta}_{2,T} &= \frac{1}{2N^{\Delta}_T} \Big\{ (X_T - X_0) \sum_{n=0}^{n_T - 1} \left(X^2_{t_{n+1}} + X^2_{t_n}\right) \Delta - \tfrac{1}{2}\left(X_T^2 - X_0^2 - c^2 T\right) \times \\ &\qquad \times \sum_{n=0}^{n_T - 1} \left(X_{t_{n+1}} + X_{t_n}\right) \Delta \Big\} \end{aligned} \tag{4.3}$$

with

$$N^{\Delta}_T = \frac{T}{2} \sum_{n=0}^{n_T - 1} \left(X^2_{t_{n+1}} + X^2_{t_n}\right) \Delta - \left(\frac{1}{2} \sum_{n=0}^{n_T - 1} \left(X_{t_{n+1}} + X_{t_n}\right) \Delta \right)^2 \tag{4.4}$$

Figures 2 and 3 show that the approximate estimators $\hat{\theta}^{\Delta}_{1,t}$ and $\hat{\theta}^{\Delta}_{2,t}$ converge after initial oscillations to the corresponding values of θ_1 and θ_2, respectively. If one uses the Euler scheme, one obtains a similar behaviour of the estimator, but for too large step sizes, e.g. $\Delta > 2^{-2}$ one observes that the estimators are not as close to the exact parameters as in the above figures.

Example 2.2 *(continued).* For the Duffing-van der Pol equation (2.5) we simulate, in $[0, T]$ with $T = 50$, an approximate realization with the parameters $\alpha = 1.0$ and $\sigma = 0.1$ starting at $(X_0^{(1)}, X_0^{(2)}) = (-1.8, 0)$. To avoid simulation of multiple stochastic integrals we use the Milstein scheme (3.8) which for (2.5) has the form

$$Y^{(2)}_{n+1} = Y^{(2)}_n + \left\{ \theta Y^{(1)}_n - \left(Y^{(2)}_n + \left(Y^{(1)}_n\right)^3 \right) \right\} \Delta + \sigma Y^{(1)}_n \Delta W_n \tag{4.5}$$

using $Y^{(1)}_{n+1} = Y^{(1)}_n + Y^{(2)}_n \Delta$. The time step size is $\Delta = 10^{-2}$. Note that, considered as a two-dimensional process $(X_t^{(1)}, X_t^{(2)})$, the Duffing-van der Pol oscillator is of the type (3.1). Also here $\underline{a} = a$, and we emphasize that in this example the Milstein scheme is identical to the Euler scheme because the term $L^1 b(Y_n)$ vanishes. A

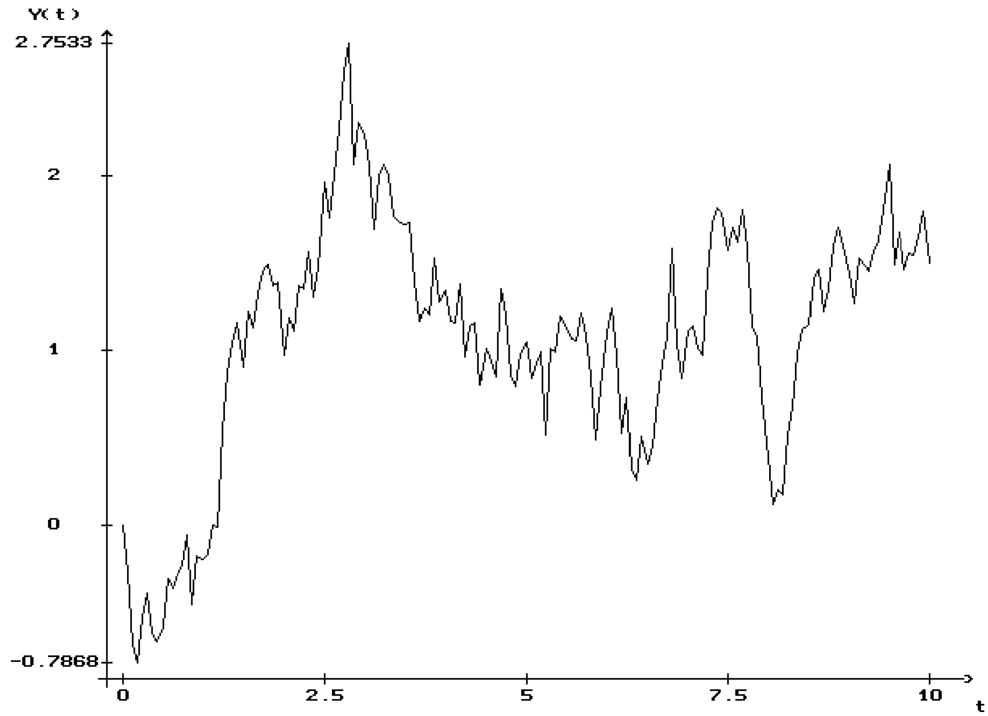

Figure 1. A simulated trajectory for the model (2.1) with $\theta_1 = -1, \theta_2 = 1, c = 1$ and $X_0 = 0$.

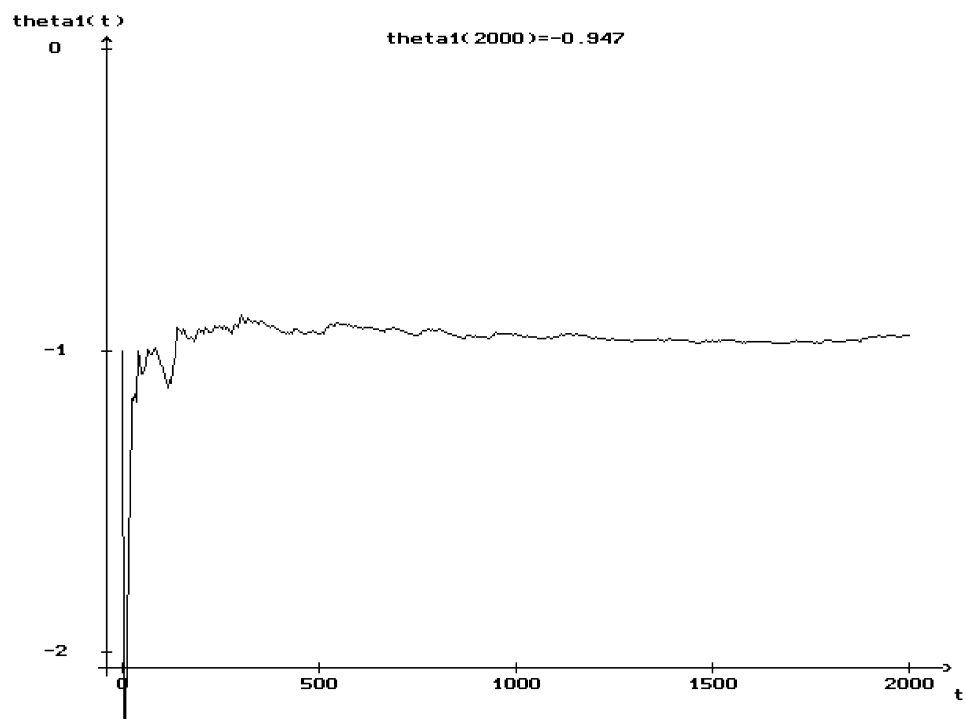

Figure 2. A simulated trajectory for the estimator $\hat{\theta}^{\Delta}_{1,t}$ of Example 2.1. The true value is $\theta_1 = -1$.

typical simulated trajectory of the Duffing-van der Pol oscillator is shown in Figure 4.

To estimate the parameter α we apply the estimator (2.11) and obtain, after approximating the relevant time integrals by the trapezoidal formula, the approximate estimator

$$\hat{\alpha}_T^\Delta = T^{-1}\left\{\frac{X_T^{(2)}}{X_T^{(1)}} - \frac{X_0^{(2)}}{X_0^{(1)}} + \right. \tag{4.6}$$

$$\left. +\frac{1}{2}\sum_{n=0}^{n_T-1}\left[\left[\frac{X_{t_{n+1}}^{(2)}}{X_{t_{n+1}}^{(1)}}\right]^2 + \left[\frac{X_{t_n}^{(2)}}{X_{t_n}^{(1)}}\right]^2 + \frac{X_{t_{n+1}}^{(2)}}{X_{t_{n+1}}^{(1)}} + \frac{X_{t_n}^{(2)}}{X_{t_n}^{(1)}} + \left(X_{t_{n+1}}^{(1)}\right)^2 + \left(X_{t_n}^{(1)}\right)^2\right]\Delta\right\}.$$

This estimator can, of course, only be used for trajectories (or parts of trajectories) where $X_{t_n} \neq 0,\ n = 0, 1, \ldots, n_T$. A trajectory of $\hat{\alpha}_T^\Delta$ is shown in Figure 5. We note that the estimator converges quickly to the exact parameter value.

Example 2.3 *(continued).* For the stochastic population model (2.12) we simulate, in $[0, T]$ with $T = 2000$, the corresponding approximate population growth process with the parameters $\theta_1 = 2.0, \theta_2 = -2.0$ and $\sigma = 0.5$ starting at $X_0 = 0.5$. The fact that the derivatives of $\sqrt{Y_n}$ are getting large for small values of Y_n can cause numerical instability. Therefore we use the derivative free explicit strong scheme of order $\gamma = 1.0$ given by (3.14), which for the population model (2.12) has the form

$$Y_{n+1} = Y_n + \left\{\theta_1 Y_n + \theta_2 Y_n^2 - \tfrac{1}{4}\sigma^2\right\}\Delta \tag{4.7}$$

$$+ \tfrac{1}{2}\sigma\left\{\left[\left(Y_n + \left(\theta_1 Y_n + \theta_2 Y_n^2 - \tfrac{1}{4}\sigma^2\right)\Delta + \sigma\sqrt{(Y_n)^+}\Delta W_n\right)^+\right]^{\frac{1}{2}} + \sqrt{(Y_n)^+}\right\}\Delta W_n$$

with time step size $\Delta = 2^{-6}$. Here x^+ denotes the positive part of x. In contrast to Example 2.1 and 2.2, the Ito equation differs here from its Stratonovich version. We emphasize that for the approximation the expression $Y_n + (\theta_1 Y_n + \theta_2 Y_n^2 - \frac{1}{4}\sigma^2)\Delta + \sigma\sqrt{Y_n}\Delta W_n$ can become negative. In such a case we set this expression equal to zero. A simulated trajectory is shown in Figure 6.

To estimate the parameters θ_1 and θ_2 we apply the estimators given by the last expressions in (2.13) and (2.14) and obtain, after approximation of the relevant time integrals by the trapezoidal formula,

$$\hat{\theta}_{1,T}^\Delta = \frac{1}{2N_T^\Delta}\left\{(X_T - X_0)\sum_{n=0}^{n_T-1}\left(X_{t_{n+1}}^3 + X_{t_n}^3\right)\Delta\right.$$

$$\left. -\frac{1}{2}\left\{X_T^2 - X_0^2 - \frac{\sigma^2}{2}\sum_{n=0}^{n_T-1}\left(X_{t_{n+1}} + X_{t_n}\right)\Delta\right\}\sum_{n=0}^{n_T-1}\left(X_{t_{n+1}}^2 + X_{t_n}^2\right)\Delta\right\} \tag{4.8}$$

and

$$\hat{\theta}_{2,T}^\Delta = \frac{1}{2N_T^\Delta}\left\{\frac{1}{2}\left(X_T^2 - X_0^2 - \frac{\sigma^2}{2}\sum_{n=0}^{n_T-1}\left(X_{t_{n+1}} + X_{t_n}\right)\Delta\right)\sum_{n=0}^{n_T-1}\left(X_{t_{n+1}} + X_{t_n}\right)\Delta\right.$$

$$\left. -(X_t - X_0)\sum_{n=0}^{n_T-1}\left(X_{t_{n+1}}^2 + X_{t_n}^2\right)\Delta\right\} \tag{4.9}$$

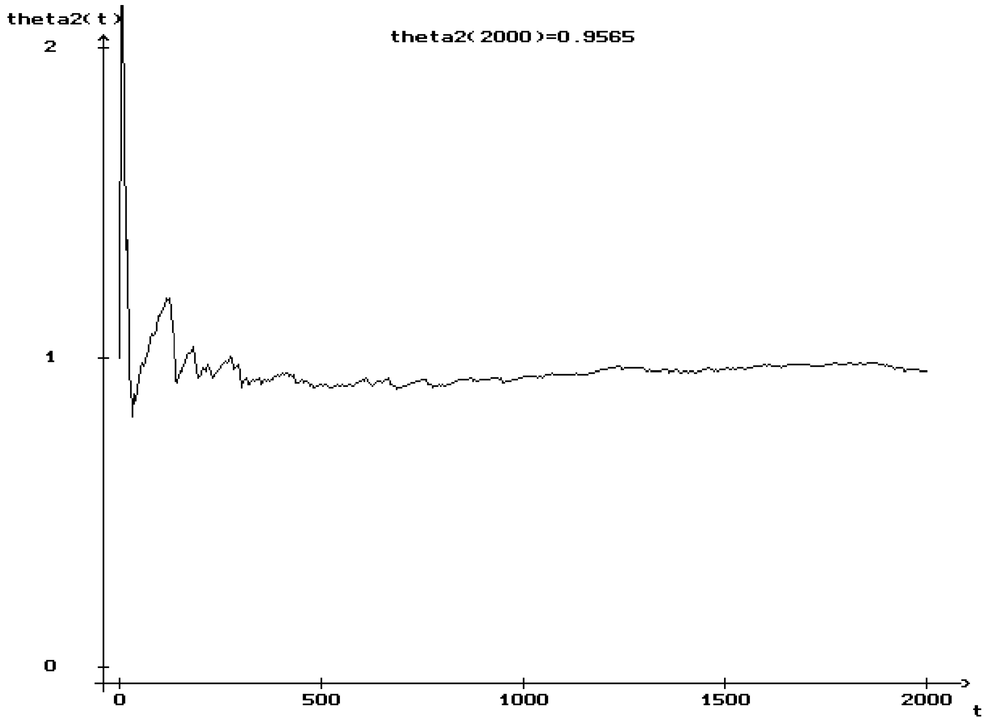

Figure 3. A simulated trajectory for the estimator $\hat{\theta}^{\Delta}_{2,t}$ of Example 2.1. The true value is $\theta_2 = 1$.

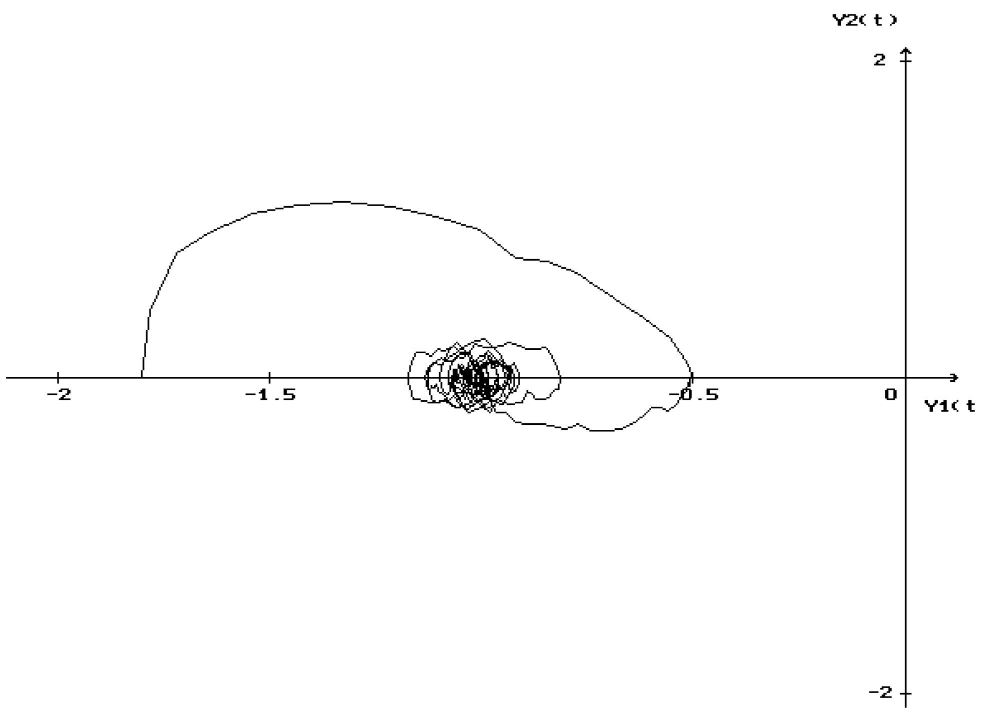

Figure 4. A simulated trajectory for the Duffing-van der Pol oscillator (2.5) with $\alpha = 1,\ \sigma = 0.1$ and $(X_0^{(1)}, X_0^{(2)}) = (-1.8, 0)$.

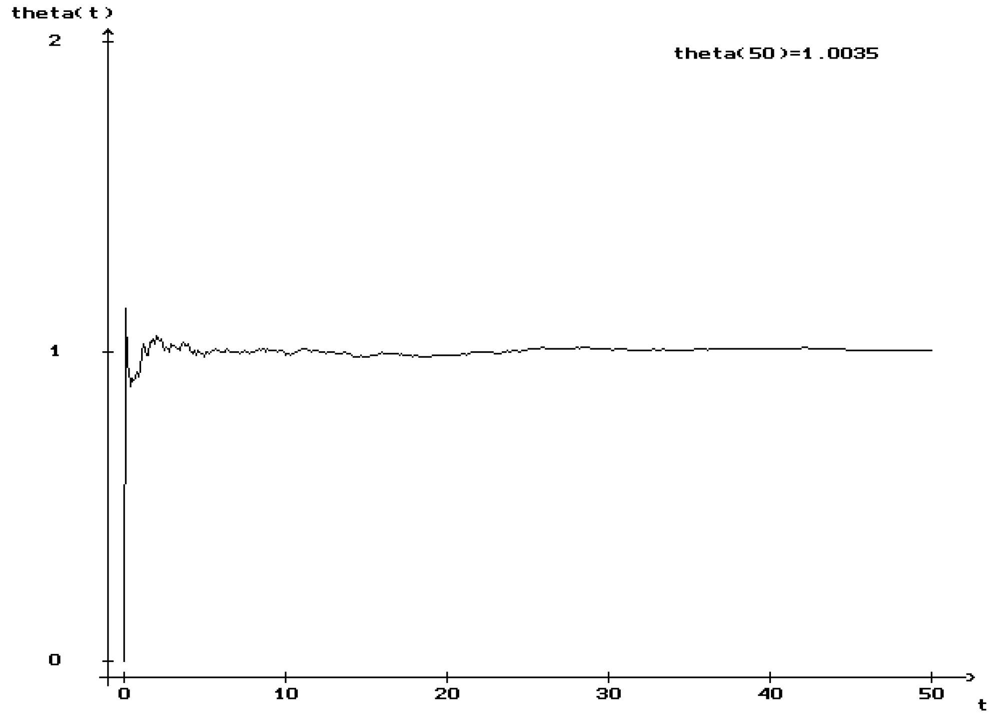

Figure 5. A simulated trajectory for the estimator $\hat{\alpha}_t^\Delta \hat{=} \theta(t)$ of example 2.2. The true value is $\alpha = 1$.

using

$$N_T^\Delta = \frac{1}{4} \sum_{n=0}^{n_T - 1} \left(X_{t_{n+1}} + X_{t_n} \right) \Delta \sum_{n=0}^{n_T - 1} \left(X_{t_{n+1}}^3 + X_{t_n}^3 \right) \Delta - \left\{ \frac{1}{2} \sum_{n=0}^{n_T - 1} \left(X_{t_{n+1}}^2 + X_{t_n}^2 \right) \Delta \right\}^2 . \tag{4.10}$$

Trajectories of $\hat{\theta}_{1,T}^\Delta$ and $\hat{\theta}_{2,T}^\Delta$ are shown in Figure 7 and Figure 8, respectively.

A Appendix: Maximum likelihood estimation

In this appendix we briefly review the theory of maximum likelihood estimation for a class of stochastic differential equations broad enough to cover the examples studied in this paper. We shall consider parametric statistical models for d–dimensional diffusion processes defined by a class of stochastic differential equations of the form

$$dX_t = [A_t(X) + B_t(X)\theta]\, dt + D_t(X)\, dW_t. \tag{A.1}$$

We assume that this equation has a unique strong solution for every θ. In (A.1) A_t, B_t and D_t are functionals depending only on the sample path up to time t. For simplicity we will assume that they are continuous functions of t for all sample paths of X. The statistical parameter θ is k–dimensional, and W is a d_1–dimensional Wiener process, so B is a $d \times k$–matrix and D is a $d \times d_1$–matrix. The vector A is d–dimensional. The functionals A, B and D are assumed to be known (given by the scientific problem under study), while the parameter θ is to be estimated

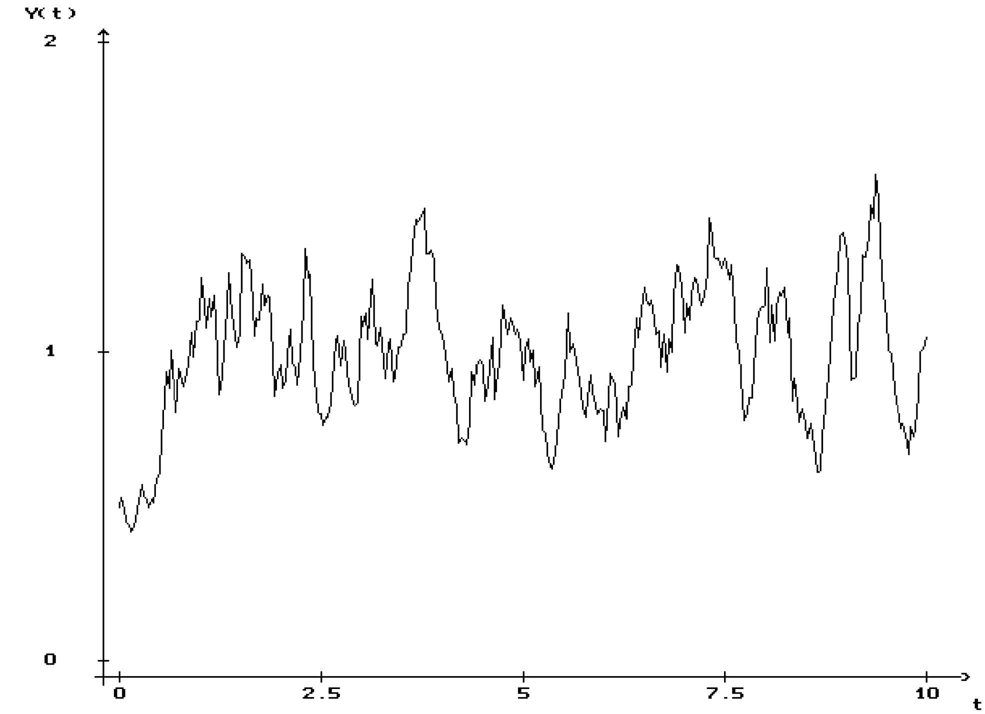

Figure 6. A simulated trajectory for the population model (2.12) with $\theta_1 = 2,\ \theta_2 = -2,\ \sigma = 0.5$ and $X_0 = 0.5$.

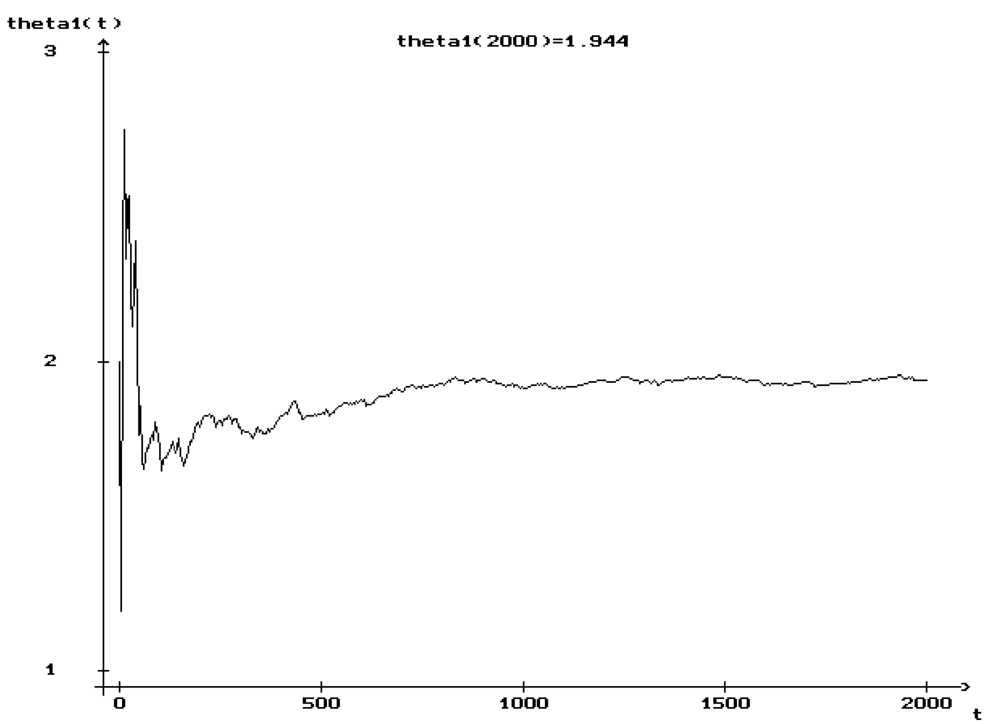

Figure 7. A simulated trajectory for the estimator $\hat{\theta}^{\Delta}_{1,t}$ of Example 2.3. The true value is $\theta_1 = 2$.

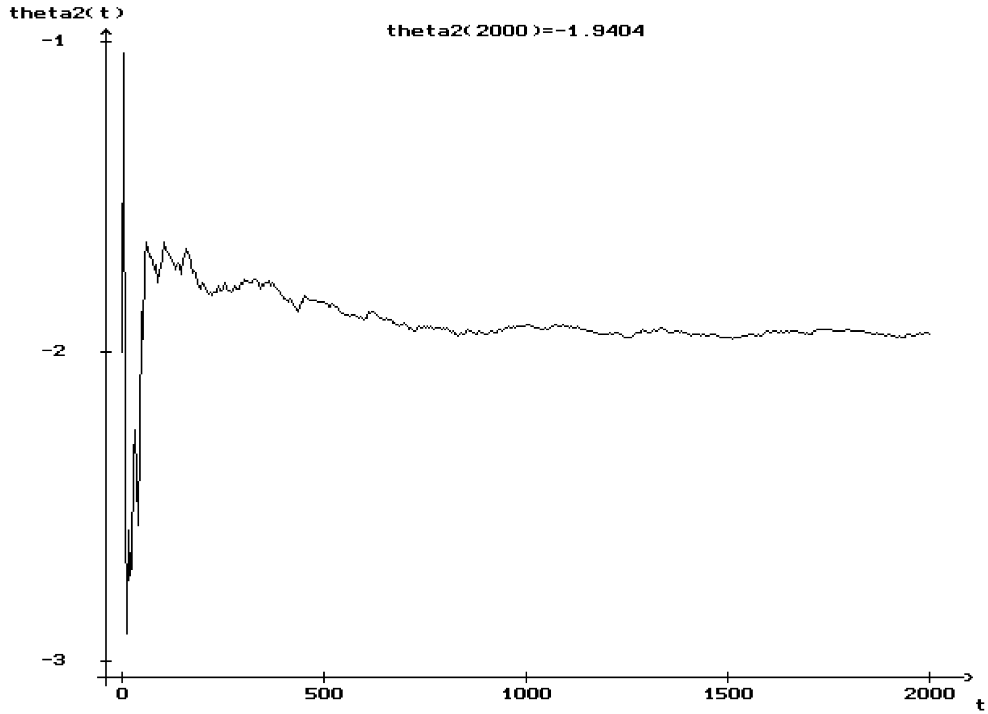

Figure 8. A simulated trajectory for the estimator $\hat{\theta}^{\Delta}_{2,t}$ of Example 2.3. The true value is $\theta_2 = -2$.

from an observed sample path $\{X_t : t \in [0,T]\}$. We assume that the columns of the matrix $B_t(X)$ are linearly independent functions of t. Otherwise the model could be parametrized by a parameter of dimension smaller than k. Models of the type (A.1) are statistically well–behaved because they are exponential families of stochastic processes; see Küchler and Sørensen (1989, 1992).

We assume that $\theta \in \Theta$, an open subset of $\mathbb{R}^k$. Without loss of generality we can assume that $0 \in \Theta$. In most applications the functionals A_t, B_t and D_t depend on X through X_t only, i.e. $A_t(X) = A_t(X_t), B_t(X) = B_t(X_t)$ and $D_t(X) = D_t(X_t)$. In such cases the solution of (A.1) is Markovian. An example where this is not the case is given in Example 2.2, since $A_t(X^{(2)}) = -(X_t^{(2)} + X_t^{(1)})^3$, $B_t(X^{(2)}) = X_t^{(1)}$ and $D_t(X^{(2)}) = \sigma X_t^{(1)}$ depend on the whole sample path of $X^{(2)}$ in $[0,t]$. Note, however, that the pair $(X^{(1)}, X^{(2)})$ is a Markov process.

Suppose we have observed the process X in the time interval $[0,T]$. Let P_θ^T denote the probability measure on the set of continuous functions from $[0,T]$ into $\mathbb{R}^d$ corresponding to the solution of (A.1) for the parameter value θ. The likelihood function for our observation $\{X_t : t \in [0,T]\}$ is the Radon–Nikodym derivative $L_T(\theta) = dP_\theta^T / dP_0^T$ provided P_θ^T is dominated by P_0^T for all $\theta \in \Theta$. This is the case if the $d \times d$–matrix $C_t(X) = D_t(X) D_t(X)^*$ is non–singular for almost all $t \in [0,T]$, and $P_\theta^T(S_T < \infty) = 1$ for all $\theta \in \Theta$, where S_T is the $k \times k$–matrix

$$S_T = \int_0^T B_t(X)^* \, C_t(X)^{-1} \, B_t(X) \, dt. \tag{A.2}$$

A star denotes matrix transposition. Define the k-dimensional random vector

$$H_T = \int_0^T B_t(X)^* C_t(X)^{-1} d\tilde{X}_t, \qquad (A.3)$$

$$\tilde{X}_t = X_t - \int_0^t A_s(X)\, ds.$$

Then the likelihood function is given by

$$L_T(\theta) = \exp[\theta^* H_T - \frac{1}{2}\theta^* S_T \theta].$$

The estimator obtained by maximizing $L_T(\theta)$ is

$$\hat{\theta}_t = S_T^{-1} H_T, \qquad (A.4)$$

which exists because S_T is non-singular under the conditions imposed. Note that S_T is the observed information matrix.

In cases where P_θ^T is not dominated by P_0^T we can interpret $L_T(\theta)$ as a quasi-likelihood function, see [11], [9], [30] and [10]. Estimators obtained by maximizing a quasi-likelihood function have many of the optimality properties enjoyed by the maximum likelihood estimator.

Under natural regularity conditions one can apply a central limit theorem for martingales to the score martingale $H_t - S_T\theta$ to prove that the maximum likelihood estimator is consistent and asymptotically normal distributed, see e.g. [31].

References

[1] Basawa, I. V. and Prakasa Rao, B.L.S. (1980): Statistical Inference for Stochastic Processes. Academic Press, London.

[2] Bibby, B.M. and Sørensen, M. (1994): Martingale estimating functions for discretely observed diffusion processes. Research Report No. 277, Dept. Theor. Statist., Aarhus Univ. To appear in Bernoulli.

[3] Clark, J. M. C., (1982): An efficient approximation scheme for a class of stochastic differential equations. Springer Lecture Notes in Control and Inform. Sc. 42, pp 69 - 78.

[4] Dacunha-Castelle, D. and Florens-Zmirou, D (1986): Estimation of the coefficients of a diffusion from discrete observations. Stochastics 19, 263 - 284.

[5] Florens-Zmirou, D. (1989): Approximate discrete-time schemes for statistics of diffusions processes. Statistics 20, 547 - 557.

[6] Florens-Zmirou, D. (1993): On estimating the diffusion coefficient from discrete observations. J. Appl. Probab. 30, 790 - 804.

[7] Genon-Catalot, V. (1990): Maximum contrast estimation for diffusion processes from discrete observations. Statistics 21, 99 - 116.

[8] Genon-Catalot, V. and Jacod, J. (1993): On the estimation of the diffusion coefficient for multidimensional diffusion processes. Ann. Inst. Henri Poincaré 29, 119 - 151.

[9] Godambe, V. P. and Heyde, C. C. (1987): Quasi-likelihood and optimal estimation. Int. Statist. Rev. 55, 231 - 244.

[10] Heyde, C. C. (1994): A quasi-likelihood approach to estimating parameters of diffusion processes. J. Appl. Probab. 31A, in press.

[11] Hutton, J. E. and Nelson, P. I. (1986): Quasi-likelihood estimation for semimartingales. Stoch. Proc. Appl. 22, 245 - 257.

[12] Kasonga, R. A. (1988): The consistency of a non-linear least squares estimator for diffusion processes. Stoch. Proc. Appl. 30, 263 - 275.

[13] Kazimierczyk, P. (1989): Consistent ML estimator for drift parameters of both ergodic and nonergodic diffusions. Springer Lecture Notes in Control an Information Science 136, 318 - 327.

[14] Kazimierczyk, P. (1992): Explicit correction formulae for parametric identification of stochastic differential systems. Math. and Comp. in Simulation.

[15] Kloeden, P. E. and Platen, E. (1989): A survey of numerical methods for stochastic differential equations. Stochastic Hydrol. Hydraul. 3, 155 - 178.

[16] Kloeden, P. E. and Platen, E. (1991): Stratonovich and Ito stochastic Taylor expansions. Math. Nachr. 151, 33 - 50.

[17] Kloeden, P. E. and Platen, E. (1992): The Numerical Solution of Stochastic Differential Equations. Springer Verlag, Berlin.

[A] Kloeden, P. E., Platen, E. and Schurz, H. (1994): Numerical Solution of SDEs through Computer Experiments. Springer Verlag, Berlin.

[18] Kloeden, P. E., Platen, E. and Wright, T. (1992): The approximation of multiple stochastic integrals. J. Stoch. Anal. Appl. 10, 431–441.

[19] Kutoyants, Yu. A. (1984): Parameter Estimation for Stochastic Processes. Heldermann Verlag, Berlin.

[20] Küchler and Sørensen, M. (1989): Exponential families of stochastic processes: A unifying semimartingale approach. Int. Statist. Rev. 57. 123 - 144.

[21] Küchler and Sørensen, M. (1992): Exponential families of stochastic processes with time-continuous likelihood function. Research Report No. 251, Department of Theoretical Statistics, Aarhus University. To appear in Scand. J. Statist.

[22] Le Breton, A. (1979): On continuous and discrete sampling for parameter estimation in diffusion type processes. Mathematical Programming Study 5. 124 -144.

[23] Milstein, G.N. (1974): Approximate integration of stochastic differential equations. Theor. Prob. Appl. 19, 557 - 562.

[M] Milstein, G.N. (1995): Numerical Integration of Stochastic Differential Equations, Kluwer Academic Publishers, Dordrecht.

[24] Newton, N. (1986): An asymptotically efficient difference formula for solving stochastic differential equations. Stochastics 19, 175 - 206.

[25] Pardoux, E. and Talay, D. (1985): Discretization and simulation of stochastic differential equations. Acta Appl. Math. 3, 23 - 47.

[26] Petersen, A. R. (1993): A new approach to maximum likelihood estimation for stochastic differential equation based on discrete observation. Research Report No. 264, Dept. Theor. Statist., Aarhus Univ. To appear in Scand J. Statist.

[27] Platen, E. (1981): An approximation method for a class of Ito processes. Lietuvos Matem. Rink. 21, 121 - 133.

[28] Prakasa Rao, B.L.S. (1988): Statistical inference from sampled data for stochastic processes. Contemp. Math. 80, 249 - 284.

[29] Rümelin, W. (1982): Numerical treatment of stochastic differential equations. SIAM J. Numer. Analysis 19, 604 - 613.

[D] Schurz, H. (1996): Stability, stationarity, and boundedness of some implicit numerical methods or stochastic differential equations. Report, WIAS, Berlin.

[E] Schurz, H. (1996): Asymptotical mean square stability of an equilibrium point of some linear numerical solutions with multiplicative noise. Stoch. Anal. Applic. 14 (3), 313–354.

[30] Sørensen, M. (1990): On quasi likelihood for semimartingales. Stoch. Processes. Appl. 35, 331 - 346.

[31] Sørensen, M. (1991): Likelihood methods for diffusion with jumps. In Prabhu, N. U. and Basawa, I.V. (eds.): Statistical Inference in Stochastic Processes, 67 - 105. Marcel Dekker, New York.

[32] Talay, D. (1983): Résolution trajectorielle et analyse numérique des équations differérentielles stochastiques. Stochastics 9, 275 - 306.

[T] Talay, D. (1995): Simulation of stochastic differential systems. in P. Kree, W. Wedig (Eds.) Probabilistic Methods in Applied Physics, LNP 451, Springer Verlag, Berlin, 54–96.

[33] Wagner, W. and Platen, E. (1978): Approximation of Ito integral equations. Preprint ZIMM, Akad. der Wiss. der DDR, Berlin.

[34] Yoshida, N. (1992): Estimation for diffusion processes from discrete observation. J. Multivariate Anal. 41, 220 - 242.