Local Linear Gaussian Estimation for Nonlinear Stochastic Differential Equations

Mitsunori Iino *  Tohru Ozaki †

January 21, 2000

Abstract

We propose a new estimation method for nonlinear stochastic differential equations based on discrete observations, and evaluate its performance. The dynamics of conditional moments are considered directly and nonlinear functions are approximated locally by linear ones using the local linearization method.

Keywords: discrete time sampling, approximate transition density, local linearization, stochastic differential equations, conditional moments

1 Introduction

We consider statistical inference for one-dimensional nonlinear stochastic differential equations from discrete time observations.

If we know the transition probability functions explicitly, maximizing the likelihood function (which can be calculated as products of the transition probabilities) derives the maximum likelihood estimator. Unfortunately, the transition densities are generally unknown, so some approximations are inevitable to cope with various practical problems. There is a vast amount of literature on this subject, and much progress has taken place recently, for

*DKFTB Pension Research Institute Ltd. 1-6-2 Marunouchi, Chiyoda-ku, Tokyo, 100-0005 Japan. phone: 03-3240-7019, e-mail: iino@ypri.co.jp
†The Institute of Statistical Mathematics, 4-6-7 Minami-Azabu, Minato-ku, Tokyo, 106-0047 Japan. phone: 03-5421-8751, e-mail: ozaki@ism.ac.jp
example, Bibby and Sørensen (1995), Pedersen (1995), Shoji and Ozaki (1997,1998), and the works referred to therein. However, there seems to be still room for improvement. In particular, considering the importance of conditional variance in applied studies (Klebaner, 1998), more research is necessary to develop efficient estimation methods for diffusion functions as well as drift functions.

The purpose of the paper is to propose an efficient estimation method, hereafter referred to as the local linear Gaussian (LLG) method, for general nonlinear stochastic differential equations. Specifically, the procedure consists of the following two parts: 1) on the assumption that transition densities can be approximated by Gaussian densities, the nonlinear differential equations that describe the dynamics of the first two moments are derived; and 2) utilizing the local linearization method (Shoji and Ozaki, 1997), the above nonlinear differential equations are approximated locally by linear ones, the solutions of which provide the estimates of the first two moments. Once the conditional moments are calculated, the parameters can be estimated immediately by maximizing the resulting approximated likelihood.

2 Local Linear Gaussian Estimation

We consider statistical inference for continuous time models given by the following one-dimensional stochastic differential equations,

\[ dX_t = f(X_t)dt + g(X_t)^{\frac{1}{2}}dW_t, \tag{1} \]
\[ X_0 = x_0, \tag{2} \]

where \( W_t \) is a standard Wiener Process, \( \alpha, \beta \) are constant parameters, and \( f(X_t), g(X_t) \) is a twice continuously differentiable function of \( X_t \).
Suppose we want to estimate the parameters $\theta$ based on observations $X_t$ that are observed at the discrete time points,

$$0 = t_0 < t_1 < \ldots < t_n.$$ 

If we know the transition function $p(X_{t_{i-1}}, X_{t_i})$, the likelihood function can be calculated as products of the transition functions. Thus, the parameters $\theta$ are estimated by minimizing the log likelihood function $L(\theta)$ defined as follows.

$$L(\theta) = \sum_{i=1}^{n} p(X_{t_{i-1}}, X_{t_i})$$ 

Unfortunately, the transition function is generally unknown, so we must use some approximation methods to write the log likelihood function explicitly. One reasonable method, at least in a short interval, is to approximate the transition function by a Gaussian density.

Let $m_t, v_t$ denote the conditional mean and the conditional variance of $X_t$ based on $X_{t_{i-1}}$, respectively,

$$m_t = E[X_t | X_{t_{i-1}}],$$

$$v_t = E[(X_t - m_t)^2 | X_{t_{i-1}}].$$

Then the approximate log likelihood function is written as follows.

$$L(\theta) = -\frac{1}{2} \sum_{i=1}^{n} \left\{ \frac{(X_{t_i} - m_{t_i})^2}{v_{t_i}} + \log(2\pi v_{t_i}) \right\} + \log(p(X_0))$$

$$\approx -\frac{1}{2} \sum_{i=1}^{n} \left\{ \frac{(X_{t_i} - m_{t_i})^2}{v_{t_i}} + \log(2\pi v_{t_i}) \right\},$$

where $p(X_0)$ is the probability density of the initial value $X_0$ and parameters can be estimated by maximizing this log likelihood function.

Consequently, the remaining problem is how to estimate the first two moments, for which we develop an efficient method in the following subsections.
2.1 Dynamics of Conditional Moments

After some calculations, it can be shown that the conditional mean $m_t$ and the conditional variance $v_t$ are the solutions of the following ordinary differential equations (see Appendix 1).

\[
\frac{dm_t}{dt} = E[f(X_t)], \quad (3)
\]

\[
\frac{dv_t}{dt} = 2\text{Cov}(X_t, f(X_t)) + E[g(X_t)], \quad (4)
\]

where the initial conditions are as follows.

\[
m_0 = X_0, \quad v_0 = 0.
\]

Here, it should be noted that these differential equations are not general ones. Because the right-hand sides of (3) and (4) contain the expected values, we need to know the whole probability densities of $X_t$. In other words, in order to estimate the first two moments, we must estimate all the other moments. Needless to say, it is impossible to estimate all moments, so we must use some approximation methods.

2.2 Local Linearization Approach

The basic concept adopted here is to approximate nonlinear functions $f(X_t)$ and $g(X_t)$ locally by linear functions of $X_t$. This method may be considered an extension of the local linearization method, which was proposed by Ozaki (1985) and extended by Ozaki (1992, 1993) and Shoji and Ozaki (1997, 1998) (see also a concise description in Prakasa Rao, 1999).

In order to linearize the above differential equations, we assume that $\frac{\partial f}{\partial X}(X_s), \frac{\partial^2 f}{\partial X^2}(X_s), \ldots$. 

4
\( \frac{\partial^2}{\partial X^2}(X_s), \frac{\partial^2}{\partial X^2}(X_s) \) are constant in a short interval \([s, s+\Delta s]\) for any \(s < t\). Then we can derive the following system of the differential equations (see Appendix 2).

\[
\frac{dY_t}{dt} = a_s + A_s Y_t,
\]  

where

\[
Y_t = (m_t \ v_t)', \\
A_s = \begin{pmatrix}
\frac{\partial f}{\partial X}(X_s) - \frac{\partial f}{\partial X}(X_s)m_s \\
\frac{\partial g}{\partial X}(X_s) - \frac{\partial g}{\partial X}(X_s)m_s \\
\frac{\partial^2 f}{\partial X^2}(X_s) + \frac{1}{2} \frac{\partial^2 g}{\partial X^2}(X_s)
\end{pmatrix}, \\
a_s = \begin{pmatrix}
f(X_s) - \frac{\partial f}{\partial X}(X_s)m_s \\
g(X_s) - \frac{\partial g}{\partial X}(X_s)m_s \\
\frac{1}{2} \frac{\partial^2 f}{\partial X^2}(X_s)
\end{pmatrix}.
\]

Once the linear vector differential equation is derived, we can solve it immediately.

\[
Y_t = \exp(A_s(t-s)) Y_s + A_s^{-1} (\exp(A_s(t-s)) - I) a_s,
\]

where

\[
I = \begin{pmatrix}
1 & 0 \\
0 & 1
\end{pmatrix}, \\
\exp(J) = \sum_{k=1}^{\infty} \frac{J^k}{k!},
\]

and the initial condition at \(s\) is

\[
Y_s = \begin{pmatrix}
X_s \\
0
\end{pmatrix}.
\]

Therefore, the conditional moments at the discrete time point \(t_i\) are given by the following.

\[
\begin{pmatrix}
m_{t_i} \\
v_{t_i}
\end{pmatrix} = \exp(A_{t_i-1}(t_i - t_{i-1})) \begin{pmatrix}
X_{t_i-1} \\
0
\end{pmatrix} + A_{t_i-1}^{-1} (\exp(A_{t_i-1}(t_i - t_{i-1})) - I) a_{t_i-1}. \tag{6}
\]

3 Numerical Studies

In this section, we conduct numerical studies to evaluate the finite sample performance of the suggested method. In order to make a comparative study, the performance of the
Euler method is also evaluated in a similar way. Approximating the stochastic differential equation (1) by using the Euler method, the conditional moments at discrete time \( t_i \) can be calculated as follows.

\[
\begin{align*}
    m_{ti} &= X_{t_{i-1}} + f(X_{t_{i-1}})(t_i - t_{i-1}) \\
    v_{ti} &= g(X_{t_{i-1}})(t_i - t_{i-1})
\end{align*}
\]

For a simulation of \( \{X_{t_0}, X_{t_1}, \ldots, X_{t_n}\} \), we use the Milstein scheme (strong Taylor approximation of convergence order 1, see Kloeden and Platen, 1992) with time-step 0.01. We set the number of sample points \( n \) to 500. In order to check the influence of discrete sampling interval \( \Delta t = t_i - t_{i-1} \), these experiments are carried out for \( \Delta t = 0.10, 0.15, 0.20 \).

The above simulation is repeated 500 times. For each set of simulated observations, the parameters are estimated using the local linear Gaussian (LLG) method and the Euler method.

We consider two examples of one-dimensional stochastic differential equations below.

**Example 1.**

The first example is the one-dimensional stochastic differential equation known as the logistic model (Gard, 1985).

\[
\begin{align*}
    dX_t &= (\alpha X_t + \beta X_t^2)dt + \gamma \sqrt{X_t}dW_t \\
    X_0 &= x_0
\end{align*}
\]

We assume here that \( x_0 \) is known, namely \( x_0 = -5.0 \) and we want to estimate the parameters \( \alpha, \beta, \gamma \). The true values of the parameters are \( \alpha = 5.0, \beta = -1.0, \gamma = 1.0 \).

The results of the simulations are summarized in Table 1. From an examination of Table 1, the superiority of the LLG method is evident. For all parameters, the LLG method gives better estimates than the Euler method because, for all cases, it has much smaller biases.
and mean squared errors. In particular, when $\Delta t$ is large, the difference is obvious.

**Example 2.**

We now consider the next stochastic differential equation.

\[
\begin{align*}
    dX_t &= \beta X_t^3 dt + \gamma dW_t \\
    X_0 &= x_0
\end{align*}
\]

In the same way as for Example 1 above, we assume here that $x_0 = 0.0$ is known and we want to estimate the parameters $\beta, \gamma$. The true values of the parameters are $\beta = -1.0, \gamma = 1.0$. Table 2 shows almost the same results as for Example 1. We see that the LLG method consistently shows smaller bias and mean squared error. On the other hand, the Euler method shows large bias and mean squared error when $\Delta t$ is large. Therefore, the superiority of the LLG method is again evident.

**4 Conclusion**

We have proposed a new estimation method for nonlinear stochastic differential equations based on discrete observations. There are two distinguishing characteristics. First, we consider the dynamics of the first two conditional moments directly, based on the assumption that Gaussian densities approximate the transition densities. Second, nonlinear differential equations are approximated locally by linear ones using the local linearization method.

We conducted simulation studies to evaluate the finite sample performance, and the proposed method performed very well. Because the proposed method consistently showed smaller biases and mean squared errors, its superiority over the Euler method is evident.
Appendix 1

Taking the expected value in the integral form of the equation (1), we have

\[ \frac{dm_t}{dt} = E[f(X_t)]. \]

By Ito’s formula (see e.g. Karatzas and Shreve, 1991),

\[ dX^2_t = (2X_t f(X_t) + g(X_t))dt + 2X_t \sqrt{g(X_t)} dW_t. \]

Let \( M_t = E[X_t^2] \), then we derive the next equation by taking the expected value of the above equation.

\[ \frac{dM_t}{dt} = E[2X_t f(X_t)] + E[g(X_t)]. \]

Furthermore, from (3),

\[ \frac{dm^2_t}{dt} = 2m_t E[f(X_t)]. \]

Noting \( u_t = M_t - m^2_t \), we have

\[ \frac{du_t}{dt} = 2Cov(X_t, f(X_t)) + E[g(X_t)]. \]

Appendix 2

By Ito’s formula, for \( s < t \)

\[ f(X_t) = f(X_s) + \int_s^t \frac{\partial f}{\partial X}(X_u) dX_u + \frac{1}{2} \int_s^t \frac{\partial^2 f}{\partial X^2}(X_u) g(X_u) du. \]

In order to linearize the above equation, we assume that \( \frac{\partial f}{\partial X}(X_s) \) and \( \frac{\partial^2 f}{\partial X^2}(X_s) \) are constant in a short interval \([s, s + \Delta s]\) for any \( s < t \). Then we have

\[ f(X_t) \approx f(X_s) + \frac{\partial f}{\partial X}(X_s)(X_t - X_s) + \frac{1}{2} \frac{\partial^2 f}{\partial X^2}(X_s) \int_s^t g(X_u) du. \]
Taking the expectation, and noting $v_t = E[\int_s^t g(X_u) du]$, we derive

$$E[f(X_t)] \approx f(X_s) + \frac{\partial f}{\partial X}(X_s)(m_t - m_s) + \frac{1}{2} \frac{\partial^2 f}{\partial X^2}(X_s)v_t. \quad (7)$$

In a similar way,

$$E[g(X_t)] \approx g(X_s) + \frac{\partial f}{\partial X}(X_s)(m_t - m_s) + \frac{1}{2} \frac{\partial^2 f}{\partial X^2}(X_s)v_t. \quad (8)$$

Furthermore, using the properties of the Ito integral and the assumption that $\frac{\partial f}{\partial X}(X_s)$ is constant in the short interval $[s, s + \Delta s)$, we derive the following equation.

$$\text{Cov}(X_t, f(X_t)) = E[\int_s^t \frac{\partial f}{\partial X}(X_u)g(X_u)du]$$

$$\approx \frac{\partial f}{\partial X}(X_s)E[\int_s^t g(X_u)du]$$

$$= \frac{\partial f}{\partial X}(X_s)v_t. \quad (9)$$

Substituting (7) into (3) as well as (8),(9) into (4), we have (5).
References


Table 1. Mean, bias, standard deviation (SD), and mean squared error (MSE) of the estimator. The true parameter values are $\alpha=5$, $\beta=-1$, $\gamma=1$.

(a) Local Linear Gaussian Method

<table>
<thead>
<tr>
<th>$\Delta t$</th>
<th>$\alpha$</th>
<th>$\beta$</th>
<th>$\gamma$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.10</td>
<td>Mean</td>
<td>5.1272</td>
<td>-1.0273</td>
</tr>
<tr>
<td></td>
<td>Bias</td>
<td>0.1272</td>
<td>-0.0273</td>
</tr>
<tr>
<td></td>
<td>SD</td>
<td>0.5893</td>
<td>0.1171</td>
</tr>
<tr>
<td></td>
<td>MSE</td>
<td>0.3628</td>
<td>0.0144</td>
</tr>
<tr>
<td>0.15</td>
<td>Mean</td>
<td>5.0838</td>
<td>-1.0205</td>
</tr>
<tr>
<td></td>
<td>Bias</td>
<td>0.0838</td>
<td>-0.0205</td>
</tr>
<tr>
<td></td>
<td>SD</td>
<td>0.5666</td>
<td>0.1136</td>
</tr>
<tr>
<td></td>
<td>MSE</td>
<td>0.3274</td>
<td>0.0133</td>
</tr>
<tr>
<td>0.20</td>
<td>Mean</td>
<td>4.9507</td>
<td>-0.9941</td>
</tr>
<tr>
<td></td>
<td>Bias</td>
<td>-0.0493</td>
<td>0.0059</td>
</tr>
<tr>
<td></td>
<td>SD</td>
<td>0.5152</td>
<td>0.1038</td>
</tr>
<tr>
<td></td>
<td>MSE</td>
<td>0.2673</td>
<td>0.0108</td>
</tr>
</tbody>
</table>

(b) Euler Method

<table>
<thead>
<tr>
<th>$\Delta t$</th>
<th>$\alpha$</th>
<th>$\beta$</th>
<th>$\gamma$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.10</td>
<td>Mean</td>
<td>4.0590</td>
<td>-0.8122</td>
</tr>
<tr>
<td></td>
<td>Bias</td>
<td>-0.9410</td>
<td>0.1878</td>
</tr>
<tr>
<td></td>
<td>SD</td>
<td>0.3507</td>
<td>0.0694</td>
</tr>
<tr>
<td></td>
<td>MSE</td>
<td>1.0081</td>
<td>0.0401</td>
</tr>
<tr>
<td>0.15</td>
<td>Mean</td>
<td>3.6396</td>
<td>-0.7286</td>
</tr>
<tr>
<td></td>
<td>Bias</td>
<td>-1.3604</td>
<td>0.2714</td>
</tr>
<tr>
<td></td>
<td>SD</td>
<td>0.2714</td>
<td>0.0543</td>
</tr>
<tr>
<td></td>
<td>MSE</td>
<td>1.9241</td>
<td>0.0766</td>
</tr>
<tr>
<td>0.20</td>
<td>Mean</td>
<td>3.2492</td>
<td>-0.6497</td>
</tr>
<tr>
<td></td>
<td>Bias</td>
<td>-1.7508</td>
<td>0.3503</td>
</tr>
<tr>
<td></td>
<td>SD</td>
<td>0.1999</td>
<td>0.0402</td>
</tr>
<tr>
<td></td>
<td>MSE</td>
<td>3.1050</td>
<td>0.1243</td>
</tr>
</tbody>
</table>
Table 2. Mean, bias, standard deviation(SD), and mean squared error(MSE) of the estimator. The true parameter values are $\beta = -1$, $\gamma = 1$.

(a) Local Linear Gaussian Method

<table>
<thead>
<tr>
<th>$\Delta t$</th>
<th>$\beta$</th>
<th>$\gamma$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.10</td>
<td>-1.0333</td>
<td>1.0060</td>
</tr>
<tr>
<td></td>
<td>-0.0333</td>
<td>0.0060</td>
</tr>
<tr>
<td></td>
<td>0.2123</td>
<td>0.0334</td>
</tr>
<tr>
<td></td>
<td>0.0461</td>
<td>0.0012</td>
</tr>
<tr>
<td>0.15</td>
<td>-1.0012</td>
<td>1.0066</td>
</tr>
<tr>
<td></td>
<td>-0.0012</td>
<td>0.0066</td>
</tr>
<tr>
<td></td>
<td>0.1578</td>
<td>0.0342</td>
</tr>
<tr>
<td></td>
<td>0.0249</td>
<td>0.0012</td>
</tr>
<tr>
<td>0.20</td>
<td>-0.9521</td>
<td>1.0060</td>
</tr>
<tr>
<td></td>
<td>0.0479</td>
<td>0.0060</td>
</tr>
<tr>
<td></td>
<td>0.1443</td>
<td>0.0352</td>
</tr>
<tr>
<td></td>
<td>0.0231</td>
<td>0.0013</td>
</tr>
</tbody>
</table>

(b) Euler Method

<table>
<thead>
<tr>
<th>$\Delta t$</th>
<th>$\beta$</th>
<th>$\gamma$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.10</td>
<td>-0.9201</td>
<td>0.9420</td>
</tr>
<tr>
<td></td>
<td>0.0799</td>
<td>-0.0580</td>
</tr>
<tr>
<td></td>
<td>0.1897</td>
<td>0.0285</td>
</tr>
<tr>
<td></td>
<td>0.0423</td>
<td>0.0042</td>
</tr>
<tr>
<td>0.15</td>
<td>-0.8554</td>
<td>0.9167</td>
</tr>
<tr>
<td></td>
<td>0.1446</td>
<td>-0.0833</td>
</tr>
<tr>
<td></td>
<td>0.1259</td>
<td>0.0275</td>
</tr>
<tr>
<td></td>
<td>0.0367</td>
<td>0.0077</td>
</tr>
<tr>
<td>0.20</td>
<td>-0.8032</td>
<td>0.8936</td>
</tr>
<tr>
<td></td>
<td>0.1968</td>
<td>-0.1064</td>
</tr>
<tr>
<td></td>
<td>0.1139</td>
<td>0.0268</td>
</tr>
<tr>
<td></td>
<td>0.0517</td>
<td>0.0120</td>
</tr>
</tbody>
</table>