Serdica

Mathematical Journal

Сердика

Математическо списание

For further information on
Serdica Mathematical Journal
which is the new series of
Serdica Bulgaricae Mathematicae Publicationes
visit the website of the journal http://www.math.bas.bg/~serdica
or contact: Editorial Office
Serdica Mathematical Journal
Institute of Mathematics and Informatics
Bulgarian Academy of Sciences
Telephone: (+359-2)9792818, FAX:(+359-2)971-36-49
e-mail: serdica@math.bas.bg

Serdica Math. J. **31** (2005), 291–308

Serdica
Mathematical Journal

Bulgarian Academy of Sciences
Institute of Mathematics and Informatics

QUASI-LIKELIHOOD ESTIMATION FOR ORNSTEIN-UHLENBECK DIFFUSION OBSERVED AT RANDOM TIME POINTS

Michel Adès*, Jean-Pierre Dion, Brenda MacGibbon*

Communicated by N. M. Yanev

Abstract. In this paper, we study the quasi-likelihood estimator of the drift parameter θ in the Ornstein-Uhlenbeck diffusion process, when the process is observed at random time points, which are assumed to be unobservable. These time points are arrival times of a Poisson process with known rate. The asymptotic properties of the quasi-likelihood estimator (QLE) of θ, as well as those of its approximations are also elucidated. An extensive simulation study of these estimators is also performed. As a corollary to this work, we obtain the quasi-likelihood estimator iteratively in the deterministic framework with non-equidistant time points.

1. Introduction. The Ornstein-Uhlenbeck process (O-U process) $\{X_t, t \geq 0\}$ arises as a solution to the following stochastic differential equation

$$dX_t = -\theta X_t dt + dW_t \tag{1}$$

2000 *Mathematics Subject Classification:* 60J60, 62M99.

Key words: Diffusion Processes, Ornstein-Uhlenbeck, Quasi-Likelihood, Poisson arrivals.

*The first and third authors greatly appreciate the support of the Naturel Sciences and Engineering Research Council of Canada for this research.

where $\{-\theta X_t\}$ represents the systematic part due to the resistance of the medium and $\{dW_t\}$ represents the random component. It is assumed that these two parts are independent and that $\{W_t\}$ is the standard Brownian motion.

We will concentrate on the estimation of the drift parameter θ, for $\theta > 0$, the so-called ergodic case.

For $t \geq 0$, the solution of (1) is

$$X_t = X_0 \exp(-\theta t) + \int_0^t \exp(-\theta(t-s))dW_s.$$

If $E(X_0^2) < \infty$ and $s < t$, then

$$E(X_t) = E(X_0)\exp(-\theta t),$$

$$\mathrm{Var}(X_t) = (2\theta)^{-1} + (V(X_0) - (2\theta)^{-1})\exp(-2\theta t),$$

$$\rho(s,t) = \mathrm{Cov}(X_s, X_t) = [V(X_0) - (2\theta)^{-1}(1 - \exp(2\theta s))]\exp(-\theta(t-s)).$$

Assume throughout the paper that X_0 is $N(0, (2\theta)^{-1})$. Then $\{X_t\}$ is a stationary zero-mean Gaussian process with $\rho(s,t) = (2\theta)^{-1}\exp(-\theta(t-s))$.

Let $\mathcal{F}_s = \sigma(X_u, u \leq s)$. Then for $s < t$,

$$E(X_t \mid \mathcal{F}_s) = X_s \exp(-\theta(t-s)), \tag{2}$$

$$\mathrm{Var}(X_t \mid \mathcal{F}_s) = (2\theta)^{-1}(1 - \exp(-2\theta(t-s))). \tag{3}$$

Bibby and Sorensen [1] studied the equally spaced case (i.e. $d_i = d$, $i = 0, \ldots, n-1$, where the sample is given by $X_{t_0}, \ldots, X_{t_n}$, and $d_i = t_{i+1} - t_i$) when the underlying diffusion was ergodic and found an estimator θ^* which they showed to be consistent for d fixed. They also made the claim of consistency in the non-ergodic case.

When $d = d(n) \to 0$ as $n \to \infty$, a necessary condition for the consistency of θ^* is that $nd(n) \to 0$ as $n \to \infty$ (see [3] and [6]).

LeBreton [9] studied this problem in the ergodic case with unequally spaced observations in $[0, T]$ and showed that when $\max d_i \to 0$ and then $T \to \infty$, the maximum likelihood estimate (MLE) of θ is a consistent estimator of θ.

There has been a considerable amount of recent research interest centered on the approximation of the continuous time process by a discrete one and its effect on statistical inference about the parameters. We refer the reader to Bibby et al. [2] for an overview of this subject.

Our concern here, however, is a very different one. We consider the sample $S = \{X_{T_0}, X_{T_1}, \ldots, X_{T_n}\}$, where $T_0 = 0$, and $\{D_i = T_{i+1} - T_i, i = 0, \ldots, n-1\}$ are i.i.d. Exponential (λ), and independent of the X_t process. We assume that, as described by the sample, the $\{T_i\}$ are unobservable random variables. In sections 3 and 4, based on S, we find the quasi-likelihood estimator (QLE) for θ and study its asymptotic properties as well as those of some approximations to the QLE. In particular we prove the surprising result that the asymptotic quasi-likelihood estimator (AQLE) $\sum_{i=0}^{n-1} X_{T_i} X_{T_{i+1}} / \sum_{i=0}^{n-1} X_{T_i}^2$, is a consistent estimator for $\lambda/(\lambda+\theta)$ under the sampling scheme S, where the observations are made at random time points $\{T_i\}$, the arrival times of a Poisson process. If λ is known, one has in turn a consistent estimator for θ.

The finite sample properties of these estimators are examined in an extensive simulation study in Section 5. In our development of this problem when the process is sampled at random time points, we also obtain the QLE of θ for the deterministic problem with non-equidistant sampling times in Section 2. The conclusion is presented in section 6. Some proofs are presented in the Appendix.

2. QLE: The deterministic framework with non-equidistant time points. We consider a stochastic differential equation according to an O-U process given by (1), with $\theta > 0$ and $X_0 \sim N(0, (2\theta)^{-1})$.

We suppose that we have a sample $\{X_{t_0}, X_{t_1}, \ldots, X_{t_n}\}$, where the times $0 = t_0 < t_1 < \cdots < t_n$ are not necessarily equally spaced.

We use the quasi-likelihood approach in the spirit of Heyde [7] to obtain the optimal estimator of the drift parameter θ. More explicitly, let $\mathbf{H}$ be the family of estimating functions $h(n, \theta)$, which consists of zero-mean square integrable martingales

$$h(n,\theta) = \sum_{i=0}^{n-1} b_i u_{i+1}, \tag{4}$$

where b_i is $\mathcal{F}_{t_i}-$ measurable and possibly a function of θ, and

$$u_{i+1} = X_{t_{i+1}} - E\left(X_{t_{i+1}} \mid \mathcal{F}_{t_i}\right). \tag{5}$$

Note that the $\{b_i u_{i+1}\}$ are uncorrelated; in fact they are martingale differences. The family $\mathbf{H}$ is generated by various choices of $\{b_i\}$.

Following Heyde [7], we obtain the optimal weights for b_i as

$$b_i^* = E\left(u'_{i+1} \mid \mathcal{F}_{t_i}\right)/E\left(u_{i+1}^2 \mid \mathcal{F}_{t_i}\right),$$

where u'_{i+1} is the derivative of u_{i+1} with respect to θ.

Therefore, using equations (2) and (3), the optimal b_i is given by

$$b_i^* = -2\theta d_i X_{t_i} e^{-\theta d_i}\left(1 - e^{-2\theta d_i}\right)^{-1}$$

where $d_i = t_{i+1} - t_i$ for $i = 0, 1, \ldots, n-1$.

From $h^*(n,\theta) = \sum_{i=0}^{n-1} b_i^* u_{i+1} = 0$, the optimal estimator θ^* is obtained. It is also called the quasi-likelihood estimator (QLE) of θ and is a solution of the equation

$$\sum_{i=0}^{n-1} d_i X_{t_i} e^{-\theta d_i}\left(1 - e^{-2\theta d_i}\right)^{-1}\left(X_{t_{i+1}} - X_{t_i} e^{-\theta d_i}\right) = 0 \tag{6}$$

In general, there will be no explicit solution for θ^*. However, we can use (6) to compute θ^* iteratively with actual data.

Let us now note the following

Remark 2.1. If the t_i are equally spaced which entails that $d_i = d$ for $i = 0, \ldots, n-1$, then equation (6) admits an explicit solution for θ given by

$$e^{-\theta^* d} = \sum_{i=0}^{n-1} X_{t_i} X_{t_{i+1}} \Big/ \sum_{i=0}^{n-1} X_{t_i}^2,$$

or equivalently

$$\theta^* = -d^{-1} \ln\left(\sum_{i=0}^{n-1} X_{t_i} X_{t_{i+1}} \Big/ \sum_{i=0}^{n-1} X_{t_i}^2\right),$$

provided that $\sum_{i=0}^{n-1} X_{t_i} X_{t_{i+1}}$ is positive, a result originally obtained by Bibby & Sorensen [1].

Remark 2.2. In the case of equidistant sampling times, for $i = 0, 1, \ldots, n-1$, $t_{i+1} - t_i = d$, and $t_0 = 0$ so that $X_{t_i} = X_{id}$. Define $U_i = X_{t_i}$; then the sequence $\{U_i\} = \{X_{t_i}\}$ is a stationary ergodic Markov chain. Therefore, from the ergodic theorem (Karlin & Taylor [8], Theorem 5.6 p. 487), it follows that, as $n \to \infty$

$$\frac{1}{n}\sum_{i=0}^{n-1} X_{t_i}X_{t_{i+1}} \xrightarrow{a.s.} E(X_{t_0}X_{t_1}) = e^{-d\theta}E(X_{t_0}^2)$$

and

$$\frac{1}{n}\sum_{i=0}^{n-1} X_{t_i}^2 \xrightarrow{a.s.} E\left(X_{t_0}^2\right).$$

Therefore $\hat{\theta} \xrightarrow{a.s.} -d^{-1}\ln\left(\frac{E\left(X_{t_0}X_{t_1}\right)}{E\left(X_{t_0}^2\right)}\right) = \theta$, a result derived by Bibby & Sorensen [1] from a more complex argument.

3. QLE: The random time points framework. Now let the sample be $S = \{X_{T_0}, X_{T_1}, \ldots, X_{T_n}\}$ as before, where the unobservable sequence $\{T_i\}$ denotes the arrival time in a homogeneous Poisson process. We assume $T_0 = 0$ and $D_i = T_{i+1} - T_i$ for $0 \leq i \leq n-1$ are i.i.d. Exponential (λ), and independent of the process X_t.

Based on S, we find the QLE estimator for θ and study its asymptotic properties as well as those of some approximations of it.

Let $\mathcal{B}_{T_i}$ be the sigma-field generated by $\{X_{T_0}, X_{T_1}, \ldots, X_{T_i}\}$, and note that the derived process $\{X_{T_i}\}$ is a Markov process (Feller [5], p. 347). Then conditioning on $\mathcal{B}_{T_i}$ will be equivalent to conditioning on X_{T_i}.

We now need the following two results (see Appendix for the proofs)

$$(7) \qquad * E\left(X_{T_{i+1}} \mid X_{T_i}\right) = X_{T_i}\mu(\lambda,\theta),$$

$$(8) \qquad * \operatorname{Var}\left(X_{T_{i+1}} \mid X_{T_i}\right) = X_{T_i}^2\left(\frac{\theta}{\lambda}\right)^2\mu(\lambda,2\theta)(\mu(\lambda,\theta))^2 + \lambda^{-1}\mu(\lambda,2\theta),$$

where $\mu(\lambda, k\theta) = \lambda/(\lambda + k\theta)$ for $k = 1, 2$.

In order to obtain the optimal b_i^*, we must find:

$$\begin{aligned} * u_{i+1} &= X_{T_{i+1}} - E\left(X_{T_{i+1}} \mid X_{T_i}\right) \\ &= X_{T_{i+1}} - X_{T_i}\mu(\lambda,\theta), \end{aligned}$$

$$* b_i^* = E\left(u_{i+1}' \mid X_{T_i}\right)/E\left(u_{i+1}^2 \mid X_{T_i}\right).$$

In this framework, we have

$$
\begin{aligned}
* \; E\left(u'_{i+1} \mid X_{T_i}\right) &= E\left(\frac{\partial}{\partial\theta}\left(X_{T_{i+1}} - X_{T_i}\mu(\lambda,\theta)\right) \mid X_{T_i}\right) \\
&= \lambda^{-1} X_{T_i}(\mu(\lambda,\theta))^2,
\end{aligned}
$$

$$* \; E\left(u^2_{i+1} \mid X_{T_i}\right) = \operatorname{Var}\left(X_{T_{i+1}} \mid X_{T_i}\right).$$

Therefore, we obtain

$$
\begin{aligned}
b_i^* &= \lambda^{-1} X_{T_i}(\mu(\lambda,\theta))^2 \Big(X^2_{T_i} \left(\frac{\theta}{\lambda}\right)^2 \mu(\lambda, 2\theta)(\mu(\lambda,\theta))^2 \\
&\quad + \lambda^{-1}\mu(\lambda,2\theta)\Big)^{-1}.
\end{aligned}
\tag{9}
$$

$$
\begin{aligned}
h^*(n,\theta) &= \sum_{i=0}^{n-1} b_i^* u_{i+1} \\
&= \lambda \frac{(\mu(\lambda,\theta))^2}{\mu(\lambda,2\theta)} \sum_{i=0}^{n-1} X_{T_i}\left((X_{T_i}\theta\mu(\lambda,\theta))^2 + \lambda\right)^{-1} \\
&\quad \cdot \left(X_{T_{i+1}} - X_{T_i}\mu(\lambda,\theta)\right).
\end{aligned}
\tag{10}
$$

Then, the quasi-likelihood estimator θ^* of θ is the solution of the following equation

$$
\begin{aligned}
h^*(n,\theta) &= \frac{\lambda(\mu(\lambda,\theta))^2}{\mu(\lambda,2\theta)} \sum_{i=0}^{n-1} X_{T_i}\left((X_{T_i}\theta\mu(\lambda,\theta))^2 + \lambda\right)^{-1} \\
&\quad \cdot \left(X_{T_{i+1}} - X_{T_i}\mu(\lambda,\theta)\right) = 0.
\end{aligned}
\tag{11}
$$

Although there is no explicit solution for θ, numerical computation of the estimator is feasible.

4. Asymptotics and approximations to the QLE In this section, we obtain an asymptotic estimator of $\mu(\lambda,\theta)$ by using equation (11) with its quadratic variation, and then showing the consistency of this estimator sampled at random time points.

The quadratic variation associated with equation (11) is given by

$$\langle h^*(n,\theta)\rangle = \frac{\lambda^2(\mu(\lambda,\theta))^4}{(\mu(\lambda,2\theta))^2}\sum_{i=0}^{n-1} X_{T_i}^2\left((X_{T_i}\theta\mu(\lambda,\theta))^2+\lambda\right)^{-2} \cdot E\left(\left(X_{T_{i+1}}-X_{T_i}\mu(\lambda,\theta)\right)^2 \mid X_{T_i}\right) = \frac{(\mu(\lambda,\theta))^4}{\mu(\lambda,2\theta)}\sum_{i=0}^{n-1} X_{T_i}^2\left((X_{T_i}\theta\mu(\lambda,\theta))^2+\lambda\right)^{-1} \quad (12)$$

where $E\left(\left(X_{T_{i+1}}-X_{T_i}\mu(\lambda,\theta)\right)^2 \mid X_{T_i}\right) = E\left(u_{i+1}^2 \mid X_{T_i}\right) = \operatorname{Var}\left(X_{T_{i+1}} \mid X_{T_i}\right)$.

Let:

$$Z_i = C^* X_{T_i}^2\left((X_{T_i}\theta\mu(\lambda,\theta))^2+\lambda\right)^{-1} \quad (13)$$

where $C^* = \dfrac{(\mu(\lambda,\theta))^4}{\mu(\lambda,2\theta)}$.

Note that Z_i is a non-negative continuous random variable with $E(Z_i) = K$ for $0 < K < \infty$, since $E(Z_i) < \dfrac{C^* E\left(X_{T_i}\right)^2}{\lambda} < \infty$.

We shall prove in theorem 1 the ergodicity of $\{X_{T_i}\}$ by using the fact that every stationary mixing process is ergodic. For that, we need the following definition of mixing given by Karlin and Taylor [8].

Definition 4.1. *A stationary process $\{Y_k\}$ is said to be mixing (or strong mixing) if for all sets A and B of k-dimensional real sequences, and $\forall k \geq 1$*

$$\lim_{n\to\infty}\{P(Y_1,Y_2,\dots Y_k)\in A \text{ and } (Y_{n+1},Y_{n+2},\dots,Y_{n+k})\in B\} = P\{(Y_1,Y_2,\dots Y_k)\in A\}\,P\{(Y_1,Y_2,\dots Y_k)\in B\}. \quad (14)$$

Hencerforth, $Y_k = X_{T_k}, k = 0,\dots,n$.

Lemma 4.1. *Let $\{X_t\}$ be an O-U ergodic Markov process and $\{T_k\}, k \geq 0$, a process with non-negative i.i.d. increments, and let us suppose that the two processes are independent. Then:*

a) $Y_k = X_{T_k}$ is $N(0,\sigma^2)$ $\quad\forall k \geq 0$,

b) $\{Y_k\} = \{X_{T_k}\}$ is a stationary process.

P r o o f.

a) $\varphi(s) = E\left(e^{isY_k}\right) = E\left(E\left(e^{isY_k} \mid T_k\right)\right)$, but $E\left(e^{isY_k} \mid T_k = t_k\right) = E\left(e^{isX_{t_k}} \mid T_k = t_k\right) = e^{-\sigma^2 \frac{s^2}{2}}$. Then $E\left(e^{isY_k}\right) = e^{-\sigma^2 \frac{s^2}{2}}$, thus proving $Y_k \sim N(0, \sigma^2)$, where σ^2 is the variance of X_0.

b) The distribution of (Y_k, Y_l) is the same as that of $(Y_{k+m}, Y_{l+m}) \forall m \geq 1$. Indeed, $E\left(e^{i(s_1 Y_k + s_2 Y_l)}\right) = E\left(E\left(e^{i(s_1 Y_1 + s_2 Y_l)} \mid T_k, T_l\right)\right)$.

But,

$$
\begin{aligned}
E&\left(e^{i(s_1 Y_k + s_2 Y_l)} \mid T_k = t_k, T_l = t_l\right) \\
&= E\left(e^{i(s_1 X_{t_k} + s_2 X_{t_l})} \mid T_k = t_k, T_l = t_l\right) \\
&= E\left(e^{i(s_1 X_{t_k} + s_2 X_{t_l})}\right) \\
&= e^{-\frac{1}{2}\left((s_1^2 + s_2^2)\sigma^2 + 2\sigma^2 s_1 s_2 \rho_{t_k t_l}\right)},
\end{aligned}
$$

where $\rho_{t_k t_l} = e^{-\theta(t_l - t_k)}$.

Thus $E\left(e^{i(s_1 Y_k + s_2 Y_l)}\right) = E\left(e^{-\frac{1}{2}\left((s_1^2 + s_2^2)\sigma^2 + 2\sigma^2 s_1 s_2 \rho_{T_k T_l}\right)}\right)$.

The last expression depends only on the distribution of $(T_l - T_k)$ which is the same as that of $(T_{l+m} - T_{k+m})$, since the increments $(T_{j+1} - T_j)$ are i.i.d. This proves $(Y_k, Y_l) \overset{d}{\sim} (Y_{k+m}, Y_{l+m}), \forall m \geq 1$. Taking into account the Markovian nature of $\{Y_k\}, (Y_{k_1}, Y_{k_2}, \ldots, Y_{k_j}) \overset{d}{\sim} (Y_{k_1+m}, \ldots, Y_{k_j+m})$ and $\{Y_k\}$ is thus a stationary process. □

We now prove the following theorem.

Theorem 4.1. *Let $\{X_t\}$ be an O-U stationary ergodic Markov process and $\{T_k\}$ a process with non-negative i.i.d. increments, where the two processes are independent. Then $\{X_{T_k}\}$ is a stationary Markov ergodic process.*

P r o o f. From Lemma 4.1, $\{X_{T_k}\}$ is a Markov stationary process, with $X_{T_k} \sim N(0, \sigma^2)$.

To prove the ergodicity of $\{X_{T_k}\}$, we verify that (14) is true using the characteristic function. Indeed, for $k = 1$ and as seen in the proof of Lemma 4.1,

$$E\left(e^{i(s_1 Y_1 + s_2 Y_{n+1})}\right) = e^{-\frac{s_1^2 \sigma^2}{2}} e^{-\frac{s_2^2 \sigma^2}{2}} E\left(e^{-\sigma^2 s_1 s_2 \rho_{T_{n+1}, T_1}}\right),$$

where

$$\rho_{T_{n+1}, T_1} = e^{-\theta(T_{n+1} - T_1)}. \tag{15}$$

Now, $T_n \to \infty$ a.s. entails that $\rho_{T_{n+1},T_1} \to 0$ a.s. and $e^{-\sigma^2 s_1 s_2 \rho_{T_{n+1},T_1}} \to 1$ a.s.

However, as this sequence is bounded, $E\left(e^{-\sigma^2 s_1 s_2 \rho_{T_{n+1},T_1}}\right) \to 1$. This proves that the joint distribution of (Y_1, Y_{n+1}) converges to the product of two $N(0,\sigma^2)$, thus proving that equation (14) holds for $k = 1$.

For $k = 2$, we have:

$$\begin{aligned} \varphi &= E\left(e^{i(s_1Y_1+s_2Y_2+s_3Y_{n+1}+s_4Y_{n+2})}\right) \\ &= E\left(E\left(e^{i(s_1Y_1+s_2Y_2+s_3Y_{n+1}+s_4Y_{n+2})} \mid T_1, T_2, T_{n+1}, T_{n+2}\right)\right). \end{aligned}$$

But, as $E\Big(e^{i(s_1Y_1+s_2Y_2+s_3Y_{n+1}+s_4Y_{n+2})} \mid T_1 = t_1, T_2 = t_2, T_{n+1} = t_{n+1}, T_{n+2} = t_{n+2}\Big)$ is the characteristic function of a multivariate normal distribution with zero mean and σ^2 variance, therefore:

$$\begin{aligned} \varphi = E\Big(&e^{-\frac{\sigma^2}{2}(s_1^2+s_2^2+s_3^2+s_4^2+2\rho_{12}s_1s_2+2\rho_{13}s_1s_3+2\rho_{14}s_1s_4} \\ &{}^{+2\rho_{23}s_2s_3+2\rho_{24}s_2s_4+2\rho_{34}s_3s_4)}\Big). \end{aligned}$$

$$\text{where} \begin{cases} \rho_{12} = e^{-\theta(T_2-T_1)}, & \rho_{13} = e^{-\theta(T_{n+1}-T_1)}, \\ \rho_{14} = e^{-\theta(T_{n+2}-T_1)}, & \rho_{23} = e^{-\theta(T_{n+1}-T_2)}, \\ \rho_{24} = e^{-\theta(T_{n+2}-T_2)}, & \rho_{34} = e^{-\theta(T_{n+2}-T_{n+1})}. \end{cases}$$

Note that $T_n \to \infty$ a.s. entails $\{\rho_{13}, \rho_{14}, \rho_{23}, \rho_{24} \to 0\}$ a.s., and $\rho_{34} \xrightarrow{d} e^{-\theta U}$, where $U \overset{d}{\sim} (T_4 - T_3)$, say, and independent from ρ_{12} since $\forall n, \rho_{34}$ is so as well.

Then,

$$e^{-\frac{\sigma^2}{2}\left(\sum\limits_i s_i^2+2\sum\limits_{i<j}\sum \rho_{ij}s_{ij}\right)} \xrightarrow{d} e^{-\frac{\sigma^2}{2}\left(\sum\limits_{i=1}^{4} s_i^2+2\rho_{12}s_1s_2+2s_3s_4e^{-\theta U}\right)}.$$

As this sequence is positive and bounded by 1, the sequence of first moments converges.

Therefore,

$$\varphi \xrightarrow{d} E\left(e^{-\frac{\sigma^2}{2}(s_1^2+s_2^2+2\rho_{12}s_1s_2)}\right) E\left(e^{-\frac{\sigma^2}{2}(s_3^2+s_4^2+2s_3s_4e^{-\theta U})}\right).$$

which is the product of the characteristic function of (Y_1, Y_2) with that of (Y_3, Y_4), thus proving the mixing property. □

Remark 4.1. Note that the results in Theorem 4.1 are true not only for Poisson arrivals, but also under the general condition of non-negative i.i.d. increments.

We now state and prove the following lemma.

Lemma 4.2. *As* $n \to \infty, \langle h^*(n,\theta)\rangle \to \infty$ *a.s., and thus* $h^*(n,\theta)/\langle h^*(n,\theta)\rangle \to 0$ *a.s. as* $n \to \infty, \forall\theta$.

Proof. It is sufficient to prove that the quadratic variation $\langle h^*(n,\theta)\rangle$ diverges to infinity, since the second part of this lemma is a consequence of the Strong Law of Large Numbers (SLLN) for zero mean square integrable martingales (Davidson [4], Theorem 20.10).

Since $\{X_{T_i}\}$ is an ergodic process by Theorem 4.1, and $\langle h^*(n,\theta)\rangle = \sum_{i=0}^{n-1} Z_i$ from equations (12) and (13), then $\frac{\langle h^*(n,\theta)\rangle}{n} = \frac{\sum_{i=0}^{n-1} Z_i}{n} \to K, 0 < K < \infty$. Thus $\langle h^*(n,\theta)\rangle \to \infty$ a.s. □

We state without proof, the following corollary which is a consequence of Lemma 4.2.

Corollary 4.1.

$$\frac{\sum_{i=0}^{n-1} X_{T_i} X_{T_{i+1}} \left(X_{T_i}^2 \theta^2 \mu^2(\lambda,\theta) + \lambda\right)^{-1}}{\sum_{i=0}^{n-1} X_{T_i}^2 \left(X_{T_i}^2 \theta^2 \mu^2(\lambda,\theta) + \lambda\right)^{-1}} \xrightarrow{a.s.} \mu(\lambda,\theta). \tag{16}$$

Inspired by this result and the ergodicity of $\{X_{T_i}\}$, we can show very easily that

$$\frac{1}{n}\sum_{i=0}^{n-1} X_{T_{i+1}} X_{T_i} \xrightarrow{a.s.} E\left(X_{T_1} X_{T_0}\right) = \mu(\lambda,\theta) E(X_{T_0}^2),$$

and $\frac{1}{n}\sum_{i=0}^{n-1} X_{T_i}^2 \xrightarrow{a.s.} E\left(X_{T_0}^2\right)$.

Therefore, we have:

$$\frac{\sum_{i=0}^{n-1} X_{T_i} X_{T_{i+1}}}{\sum_{i=0}^{n-1} X_{T_i}^2} \xrightarrow{a.s.} \mu(\lambda, \theta), \tag{17}$$

which proves the consistency of this estimator.

The result in (17) is obviously an approximate solution to (11), and hencerforth this estimator will be denoted by AQLE. Indeed, if we define

$$h^+(n, \theta) = \sum_{i=0}^{n-1} X_{T_i} \left(X_{T_{i+1}} - \mu(\lambda, \theta) X_{T_i} \right)$$

as an estimating equation, then $h^+(n, \theta) = 0$, yields $\dfrac{\sum_{i=0}^{n-1} X_{T_i} X_{t_{i+1}}}{\sum_{i=0}^{n=1} X_{T_i}^2}$ as an estimator for $\mu(\lambda, \theta)$.

Thus we have two competing estimators for $\mu(\lambda, \theta)$: the quasi-likelihood estimator (QLE) of $\mu(\lambda, \theta)$ given as the solution to the equation (11), and an approximation (AQLE) to it, given by $\sum_{i=0}^{n-1} X_{T_i} X_{T_{i+1}} / \sum_{i=0}^{n-1} X_{T_i}^2$, which is consistent.

We conjecture that the QLE is also consistent and that both estimators are asymptotically normal. A simulation study is performed in the next section in which the bias of each estimator is compared in an attempt to determine whether the QLE is also consistent. Histograms are constructed for different sample sizes in order to study the limiting distributions.

5. Simulation study. From sections 3 and 4, we have two estimators: the QLE given as a solution to equation (11) which is optimal, and the AQLE defined by equation (17) which was shown to be consistent. The asymptotic distribution of neither of these estimators has yet been studied theoretically, although we conjecture that each is asymptotically normal. Here we propose an extensive numerical simulation study in order to attempt to answer the following questions:

(i) What is the asymptotic distribution of each estimator?

(ii) How is the bias related to $\mu(\lambda, \theta)$?

Table 1. $m = 1000$, $n = 100$, $\lambda = 1$, $\theta = 0.125$, $\mu(\lambda, \theta) = 8.8889 \cdot 10^{-1}$

	Sample mean	Sample standard deviation
QLE	$8.7333 \cdot 10^{-1}$	$5.4224 \cdot 10^{-2}$
AQLE	$8.7154 \cdot 10^{-1}$	$5.5068 \cdot 10^{-2}$

Figure 1. Histograms: AQLE (left), QLE (right)

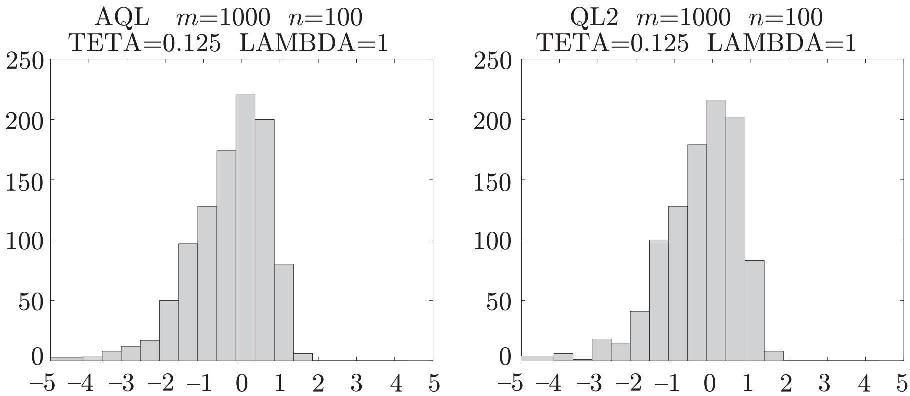


Table 2. $m = 1000$, $n = 500$, $\lambda = 1$, $\theta = 0.125$, $\mu(\lambda, \theta) = 8.8889 \cdot 10^{-1}$

	Sample mean	Sample standard deviation
QLE	$8.8565 \cdot 10^{-1}$	$2.2835 \cdot 10^{-2}$
AQLE	$8.8530 \cdot 10^{-1}$	$2.2884 \cdot 10^{-2}$

Figure 2. Histograms: AQLE (left), QLE (right)

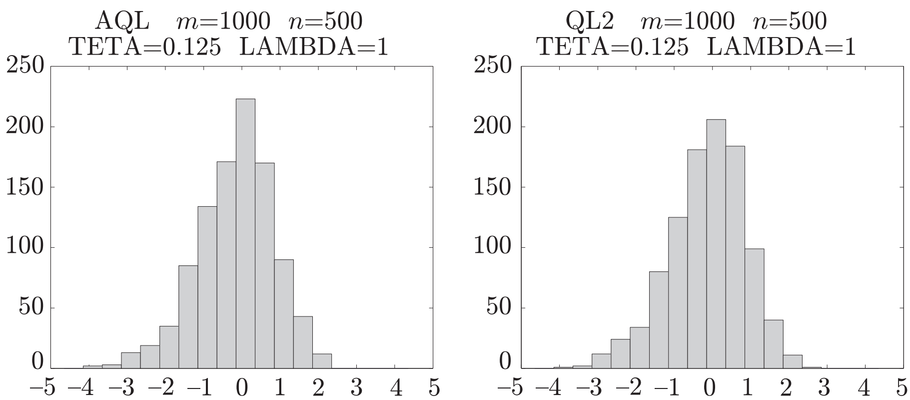


Table 3. $m = 1000$, $n = 5000$, $\lambda = 1$, $\theta = 0.125$, $\mu(\lambda, \theta) = 8.8889 \cdot 10^{-1}$

	Sample mean	Sample standard deviation
QLE	$8.8879 \cdot 10^{-1}$	$6.7259 \cdot 10^{-3}$
AQLE	$8.8877 \cdot 10^{-1}$	$6.7624 \cdot 10^{-3}$

Figure 3. Histograms: AQLE (left), QLE (right)

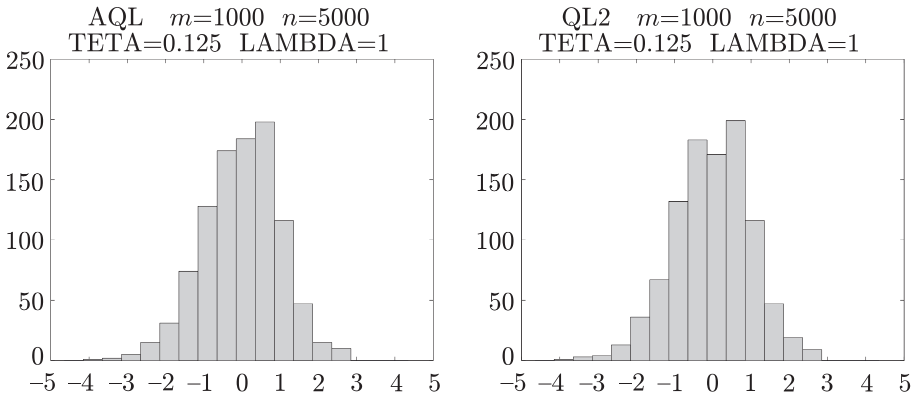


Table 4. $m = 1000$, $n = 100$, $\lambda = 1$, $\theta = 0.5$, $\mu(\lambda, \theta) = 6.6667 \cdot 10^{-1}$

	Sample mean	Sample standard deviation
QLE	$6.5819 \cdot 10^{-1}$	$8.0505 \cdot 10^{-2}$
AQLE	$6.5670 \cdot 10^{-1}$	$8.0835 \cdot 10^{-2}$

Figure 4. Histograms: AQLE (left), QLE (right)

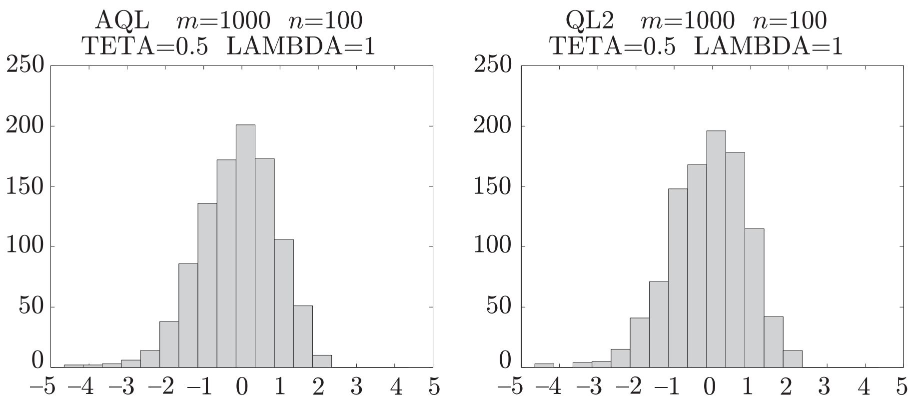


Table 5. $m = 1000$, $n = 500$, $\lambda = 1$, $\theta = 0.5$, $\mu(\lambda, \theta) = 6.6667 \cdot 10^{-1}$

	Sample mean	Sample standard deviation
QLE	$6.6434 \cdot 10^{-1}$	$3.5963 \cdot 10^{-2}$
AQLE	$6.6353 \cdot 10^{-1}$	$3.6152 \cdot 10^{-2}$

Figure 5. Histograms: AQLE (left), QLE (right)

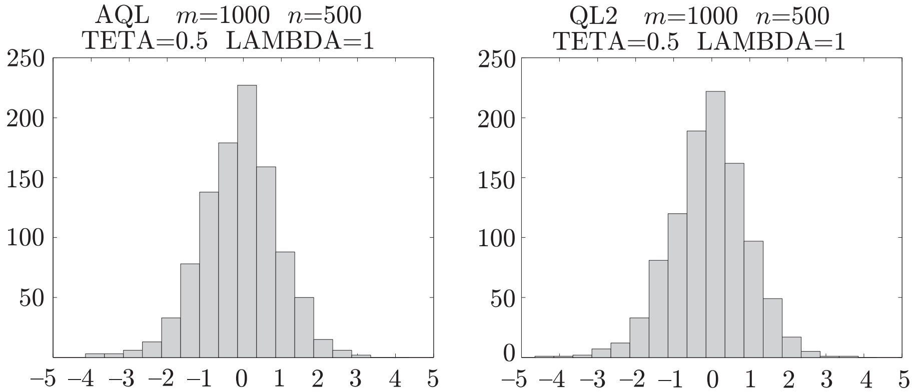
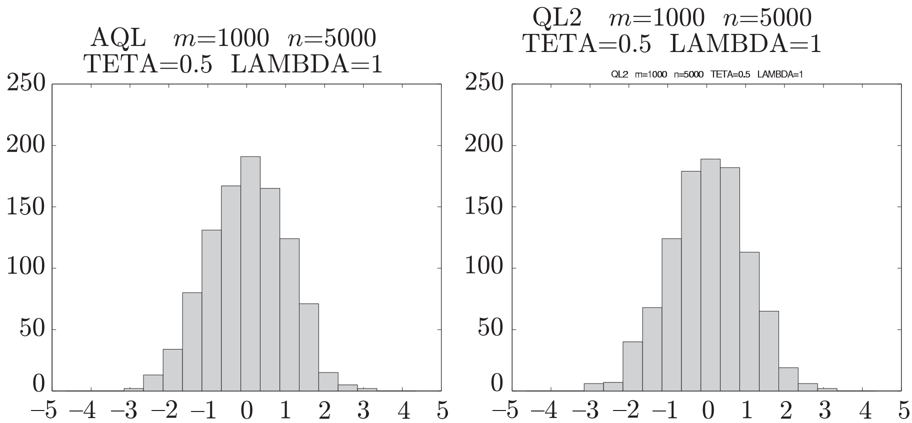


Table 6. $m = 1000$, $n = 5000$, $\lambda = 1$, $\theta = 0.5$, $\mu(\lambda, \theta) = 6.6667 \cdot 10^{-1}$

	Sample mean	Sample standard deviation
QLE	$6.6701 \cdot 10^{-1}$	$1.1365 \cdot 10^{-2}$
AQLE	$6.6698 \cdot 10^{-1}$	$1.1439 \cdot 10^{-2}$

Figure 6. Histograms: AQLE (left), QLE (right)



(iii) How is the bias related to the sample size?

Assuming:

$$\begin{aligned} T_0 &\equiv 0 \\ T_{i+1} - T_i &\sim \text{Exponential}(\lambda) \\ X_0 &\sim \mathcal{N}(0, (2\theta)^{-1}) \\ X_t | X_s &\sim \mathcal{N}\left(X_s e^{(-\theta(t-s))}, (2\theta)^{-1}(1 - e^{(-2\theta(t-s))}\right) \end{aligned}$$

Note that, 48 different cases as follows were simulated:

m (number of samples)	1000
n (size of each sample)	100, 500, 1000, 5000
θ	0.5, 0.25, 0.125
λ	1, 0.5, 0.2, 0.1

The simulation study is too large to include all the results here. We present only the following cases: $m = 1000, \lambda = 1, \theta = .125$ and $n = 100$, 500, 5000 and $m = 1000, \lambda = 1, \theta = .5$. It should be noted that the histograms are centered at 0 by subtracting the true mean.

The main conclusions related to our three questions are as follows:

(i) Both estimators appear to have asymptotically normal distributions, since in all cases the histograms with $n \geq 500$ appear to be normal.

(ii) as $\mu \to 1$, the bias of both the QLE and the AQLE increase. This can be seen by comparing Tables 1, 2 and 3 with Tables 4, 5 and 6 respectively.

(iii) For both values of μ and for nearly all sample sizes, the absolute value of the bias, of the QLE is less than that of the AQLE. The only exception is for $\mu = 6.667 \times 10^{-1}$ and $n = 5000$. In this case the bias of the QLE $=$ .00094 and of the AQLE $= -.00031$. This suggests that the QLE is also consistent.

6. Conclusion. We have studied the quasi-likelihood estimator of the drift parameter θ in the Ornstein-Uhlenbeck diffusion process observed at Poisson arrival times, which are assumed unobservable. An asymptotic version of this estimator, the AQLE, has also been elucidated and shown to be consistent. An extensive simulation study supports our conjectures that the QLE is also consistent and that both estimators are asymptotically normal. As a corollary to this work we obtain the QLE iteratively in the deterministic framework with non-equidistant points.

Acknowledgments. Michel Adès wishes to dedicate his contribution to this paper to Peter Caines (McGill University, Canada) in honour of his 60^{th} birthday in 2005. The authors also wish to thank Michael Sorensen for providing them with his reports, René Ferland for fruitfull discussions and Claude Gravel for his programming of the simulations. Finally, the authors are thankful to the referee for a careful reading of the manuscript.

Appendix. Here are the proofs of equations (7) and (8) of section 3.

Let $\{X_t\}, t \geq 0$ be an O-U process, sampled at random time points T_k, where $\{T_k\}, k \geq 0$ is a Poisson process. To simplify the notation, let $\{X_{T_k}\} = \{Y_k\}$. We have the following results.

1) $E\left(Y_{k+1} \mid Y_k\right) = Y_k \mu(\lambda, \theta)$, where $\mu(\lambda, \theta) = \frac{\lambda}{\lambda+\theta}$

P r o o f.

$$E\left(Y_{k+1} \mid Y_k\right) = E\left(E\left(Y_{k+1} \mid Y_k, T_k, T_{k+1}\right)\right).$$

But,

$$\begin{aligned} & E\left(X_{T_{k+1}} \mid X_{T_k}, T_k = t_k, T_{k+1} = t_{k+1}\right) \\ &= E\left(X_{t_{k+1}} \mid X_{t_k}, T_k = t_k, T_{k+1} = t_{k+1}\right) \\ &= X_{t_k} e^{-\theta(t_{k+1}-t_k)}, \end{aligned}$$

by using (2).

Then,

$$\begin{aligned} E\left(X_{T_{k+1}} \mid X_{T_k}\right) &= X_{T_k} E\left(e^{-\theta(T_{k+1}-T_k)}\right) \\ &= X_{T_k} \mu(\lambda, \theta). \end{aligned}$$

Similarly, we show that

$$E\left(Y_{k+m} \mid Y_k\right) = Y_k (\mu(\lambda, \theta))^m, \forall m \geq 1. \qquad \square$$

2) $\operatorname{Var}\left(Y_{k+1} \mid Y_k\right) = Y_k^2 \left(\frac{\theta}{\lambda}\right)^2 (\mu(\lambda, \theta))^2 \mu(\lambda, 2\theta) + \lambda^{-1} \mu(\lambda, 2\theta)$.

P r o o f. It is well known that

$$\begin{aligned} \operatorname{Var}\left(Y_{k+1} \mid Y_k\right) &= \operatorname{Var}\left(E\left(Y_{k+1} \mid Y_k, T_k, T_{k+1}\right)\right) \\ &\quad + E\left(\operatorname{Var}\left(Y_{k+1} \mid Y_k, T_k, T_{k+1}\right)\right). \end{aligned}$$

But,

$$\begin{aligned}\mathrm{Var}\left(E\left(Y_{k+1} \mid Y_k, T_k, T_{k+1}\right)\right) &= \mathrm{Var}\left(Y_k e^{-\theta(T_{k+1}-T_k)} \mid Y_k\right) \\ &= Y_k^2 \frac{pq}{(p+q)^2(p+q+1)},\end{aligned}$$

where $(T_{k+1} - T_k)$ is Beta (p, q) with $p = \frac{\lambda}{\theta}$ and $q = 1$.
Using (3), we have

$$\mathrm{Var}\left(Y_{k+1} \mid Y_k, T_k, T_{k+1}\right) = \sigma^2\left(1 - e^{-2\theta(T_{k+1}-T_k)}\right)$$

Then,

$$\begin{aligned}E\left(\mathrm{Var}\left(Y_{k+1} \mid Y_k, T_k, T_{k+1}\right)\right) &= \sigma^2(1 - \mu(\lambda, 2\theta)) \\ &= \lambda^{-1}\mu(\lambda, 2\theta).\end{aligned}$$

Finally, we have

$$\mathrm{Var}\left(Y_{k+1} \mid Y_k\right) = aY_k^2 + b.$$

where $\begin{cases} a = \left(\frac{\theta}{\lambda}\right)^2 (\mu(\lambda,\theta))^2 \mu(\lambda, 2\theta) \\ b = \lambda^{-1}\mu(\lambda, 2\theta). \end{cases}$ □

REFERENCES

[1] B. M. Bibby, M. Sorensen. Martingale estimation functions for discretely observed diffusion processes. *Bernouilli* **1** (1995), 17–39.

[2] B. M. Bibby, M. Jacobsen, M. Sorensen. Estimating Functions for Discretely Sampled Diffusion-Type Models. Preprint 4, september 2004, Department of Applied Mathematics and Statistics, University of Copenhagen.

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

[4] J. Davidson. Stochastic Limit Theory. Oxford University Press, 2001.

[5] W. Feller. An Introduction to Probability Theory and Its Applications. Volume II, Second Edition, Wiley, 1971.

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

[7] C. C. Heyde. Quasi-Likelihood And Its Application. A General Approach to Optimal Parameter Estimation. Springer, 1997.

[8] S. Karlin, H. M. Taylor. A First Course in Stochastic Processes. Second Edition, Academic Press, 1975.

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

Michel Adès
Université du Québec à Montréal
Département de mathématiques
Case postale 8888
Succursale Centre-ville
Montréal (Québec)
H3C 3P8
e-mail: ades.michel@uqam.ca

Jean-Pierre Dion
Université du Québec à Montréal
Département de mathématiques
Case postale 8888
Succursale Centre-ville
Montréal (Québec)
H3C 3P8
e-mail: dion.jean-pierre@uqam.ca

Brenda MacGibbon
Université du Québec à Montréal
Département de mathématiques
Case postale 8888
Succursale Centre-ville
Montréal (Québec)
H3C 3P8
e-mail: macgibbon.brenda@uqam.ca

Received August 3, 2005
Revised December 15, 2005