Sankhyā : The Indian Journal of Statistics
1986, Volume 48, Series A, Pt. 2, pp. 181-187.

BERRY-ESSEEN BOUND FOR THE MAXIMUM LIKELIHOOD ESTIMATOR IN THE ORNSTEIN-UHLENBECK PROCESS

By ARUP BOSE

Indian Statistical Institute

SUMMARY. We prove that the distribution of the normalized maximum likelihood estimate of the drift parameter in the Ornstein-Uhlenbeck process converges to the normal distribution with an error rate $O(T^{-1/2})$.

1. Introduction

The study of the rate of convergence of an estimator often boils down to studying rate of convergence to normality for martingales or their perturbations. This is usually carried out by appealing to Skorokhod's embedding theorem.

This is the approach taken by Mishra and Prakasa Rao (1985) in dealing with maximum likelihood estimate (m.l.e.) for processes of the form

$$dX(t) = -\theta a(X(t))dt + b(X(t))dW(t),\ X(0) = 0,\ t \geqslant 0,\ \theta > 0 \text{ and } W(\cdot)$$

is a Brownian motion. With a condition on the growth of $\int_0^T \frac{a^2(X(t))}{b^2(X(t))}\, dt$, (which, incidentally is hard to check) they obtain error bounds for the normal approximation of the normalized m.l.e. When applied to Ornstein-Uhlenbeck (O-U) process, this yields the rate $T^{-1/5}$.

However, the O-U process being a natural continuous time analogue of the first order discrete autoregressive process with i.i.d. $N(0, 1)$ errors, one is led to believe that the above rate can be sharpened.

The normalized m.l.e. is a ratio of two processes. The numerator is a martingale which converges to a normal variable and the denominator is the corresponding associated increasing process which converges a.s. to a positive constant.

Mishra and Prakasa Rao (1985) use simple Markov inequalities to tackle the denominator. The numerator is embedded in a Brownian motion by

---

AMS (1980) *subject classification :* Primary 62F12, 62M05; Secondary, 60F05, 60J70.

Key words and phrases : Stochastic differential equation, Ito's formula, Burkholders' inequality, Esseen's lemma.

Kunita-Watanabe theorem and then Lemma 3.2 of Hall and Heyde (1980) is invoked. These two together limit the rate obtainable to $T^{-1/5}$. For the O-U process, Burkholder's inequality can be used for the denominator (after applying Ito's formula) to yield the better rate $T^{-1/4+\varepsilon}$, $\varepsilon > 0$. However, as long as we use embedding technique, the rate cannot be better than $T^{-1/4}$.

Hence we take an alternative approach. By extending an argument in Liptser and Shiryayev (1978) (henceforth referred as LS), we obtain the characteristic function of the numerator for suitable values of the argument. This allows us to use Esseen's lemma yielding the rate $O(T^{-1/2})$ for the numerator. The denominator is linked with the numerator via Ito's formula. This helps us to get the final result.

This result opens up the possibility of obtaining faster rates of convergence for general linear diffusions. It also shows that the embedding technique might not lead to the strongest possible results.

2. Preliminaries

Let $X(t)$, $t \geqslant 0$ be a diffusion process satisfying the stochastic differential equation

$$dX(t) = -\theta X(t)dt + dW(t),\ t \geqslant 0,\ X(0) = 0.$$

Here $W(\cdot)$ is a standard Brownian motion and $\theta > 0$ is the unknown parameter. It is well known that the solution $X(t)$ is a continuous Gaussian process.

Let $C[0, T]$ = The space of real valued continuous function on $[0, T]$.

M_θ^T = The measure generated by $X(t)$, $0 \leqslant t \leqslant T$ on $C[0, T]$

M_w^T = The measure generated by $W(t)$, $0 \leqslant t \leqslant T$ on $C[0, T]$.

It is well known that $M_\theta^T \ll M_w^T$ and the Radon-Nikodym derivative (likelihood function) can be explicitly computed. This in turn yields that the m.l.e. θ_T of θ based on the "observation" $X(t)$, $0 \leqslant t \leqslant T$ satisfies

$$\theta_T - \theta = \int_0^T X(t)dW(t) \left(\int_0^T X^2(t)dt \right)^{-1}.$$

The details can be seen from LS.

Ito's formula (See Elliott, 1982) gives

$$2\int_0^T X(s)dW(s) - X^2(T) = 2\theta \int_0^T X^2(t)dt - T.$$

This relation shall be used later.

C shall denote a generic constant (perhaps depending on θ, but not on anything else).

3. The main result

We begin with a few lemmas.

Lemma 3.1 : *For* $Z_1, Z_2 \in C^2$,

Let $$\varphi_T(Z_1, Z_2) = E\, exp(Z_1 \int_0^T X^2(t)dt + Z_2 X^2(T)),$$

Then $\varphi_T(Z_1, Z_2)$ *exists for* $|Z_i| < \delta$, $i = 1, 2$ *for some* $\delta > 0$ *and is given by*

$$\varphi_T(Z_1, Z_2) = exp\left(\frac{\theta T}{2}\right)\left[\frac{2\lambda}{(\lambda-\theta+2Z_2)exp(-\lambda T)+(\lambda+\theta-2Z_2)exp(\lambda T)}\right]^{1/2} \quad \dots \quad (3.1)$$

where $\lambda = (\theta^2 - 2Z_1)^{1/2}$ *and we always choose the principal branch of the square root.*

Proof: First assume that $Z_i = a_i \in \mathcal{R}$, $i = 1, 2$ and a_i are sufficiently small.

Define $\lambda = (\theta^2 - 2a_1)^{1/2}$ and $dX_t^\lambda = \lambda X_t^\lambda\, dt + dW(t)$, $X_0^\lambda = 0$. Also recall that $dX_t^\theta = -\theta X_t^\theta dt + dW(t)$.

Then we have (see LS)

$$\frac{dM_\theta^T}{dM_\lambda^T}(X^\lambda(\cdot)) = \exp\left[(-\theta-\lambda)\int_0^T X_t^\lambda dX_t^\lambda - \left(\frac{\theta^2-\lambda^2}{2}\right)\int_0^T (X_t^\lambda)^2 dt\right]. \quad \dots \quad (3.2)$$

Note that

$$\varphi_T(a_1, a_2) = E_\theta \exp\left(a_1 \int_0^T (X_t^\theta)^2 dt + a_2 (X_T^\theta)^2\right).$$

If we change the measure to that generated by X_t^λ, then by (3.2),

$$\varphi_T(a_1, a_2) = E\, \exp\left[a_1 \int_0^T (X_t^\lambda)^2 dt + a_2 (X_t^\lambda)^2 - (\theta+\lambda)\int_0^T X_t^\lambda dX_t^\lambda - a_1 \int_0^T (X_t^\lambda)^2 dt\right]$$

$$= E\, \exp\left[a_2 (X_T^\lambda)^2 - (\theta+\lambda)\int_0^T X_t^\lambda dX_t^\lambda\right]. \quad \dots \quad (3.3)$$

By Ito's formula,

$$d(X_t^\lambda)^2 = 2X_t^\lambda dX_t^\lambda + dt.$$

Using this in (3.3),

$$\varphi_T(a_1, a_2) = E\exp\Big[(X_T^\lambda)^2\Big(a_2-\frac{\theta+\lambda}{2}\Big)+\frac{T}{2}(\theta+\lambda)\Big]$$

Note that $X_t^\lambda \sim N\left(0, \frac{\exp(2\lambda T)-1}{2\lambda}\right)$. Thus

$$\varphi_T(a_1, a_2) = \exp\Big(\frac{T(\theta+\lambda)}{2}\Big)\Big[\frac{2\lambda}{2\lambda+(\lambda+\theta-2a_2)(\exp(2\lambda T)-1)}\Big]^{1/2}$$

which on simplification yields (3.1) for Z_i's real, around a neighbourhood of zero. Thus, there is no problem of existence of $\varphi_T(Z_1, Z_2)$ around zero in C^2 and since we have shown that the m.g.f. exists, $\varphi_T(Z_1, Z_2)$ is an analytic function. On the other hand (3.1) defines an analytic function in the relevant domain and agrees with $\varphi_T(Z_1, Z_2)$ for Z_1, Z_2 real. This finishes the proof.

Lemma 3.2 : *For $|t| \leqslant \varepsilon T^{1/2}$, where ε is sufficiently small,*

$$\left| E\,exp\Big(it\Big(\frac{2\theta}{T}\Big)^{1/2}\int_0^T X(s)dW(s)\Big)-exp(-t^2/2)\right| \leqslant C.exp(-t^2/4)|t|^3T^{-1/2}$$

Proof : By Ito's formula,

$$\int_0^T X(s)dW(s) = \theta\int_0^T X^2(t)dt-\frac{T}{2}+\frac{X^2(T)}{2}.$$

Hence,

$$E\exp\Big(it\Big(\frac{2\theta}{T}\Big)^{1/2}\int_0^T X(s)dW(s)\Big) = \varphi_T(Z_1, Z_2)\exp\Big(-\frac{it}{2}(2\theta T)^{1/2}\Big) \quad \text{... (3.4)}$$

where $Z_1 = it\theta\left(\frac{2\theta}{T}\right)^{1/2}$, $Z_2 = \frac{it}{2}\left(\frac{2\theta}{T}\right)^{1/2}$.

Note that (Z_1, Z_2) satisfies the condition of Lemma 3.1 by choosing ε sufficiently small.

Note that $\lambda-\theta = O(|t|T^{-1/2})$, $\lambda+\theta = 2\theta+O(|t|T^{-1/2})$

and $\lambda = \theta\beta_T(t)+O(|t|^3T^{-1/2})$

where $\beta_T(t) = 1-\frac{Z_1}{\theta^2}-\frac{Z_1^2}{2\theta^4}.$

Let $\alpha_T(t)$ denote any function which is $O(|t|T^{-1/2})$. Using these simple estimates,

$$\varphi_T(Z_1, Z_2) = \exp\left(\frac{\theta T}{2}\right)\left[\frac{(1+\alpha_T(t))\exp(T\theta\beta_T(t)+O(|t|^3T^{-1/2}))}{\alpha_T(t)+(2+\alpha_T(t))\exp(2T\theta\beta_T(t)+O(|t^3|T^{-1/2})}\right]^{1/2}$$

Using this in (3.4), the required expectation equals

$$\left[\frac{1+\alpha_T(t)}{\alpha_T(t)\exp\left(-\frac{T\theta}{2}\right)+(1+\alpha_T(t))\exp(\psi_T(t))}\right]^{1/2}$$

where

$$\begin{aligned}\psi_T(t) &= T\theta\beta_T(t)-\theta T+\frac{it}{2}(2\theta T)^{1/2}+O(|t|^3T^{-1/2})\\ &= t^2+O(|t|^3T^{-1/2}).\end{aligned}$$

Thus, the difference to be estimated, is, in absolute value

$$\begin{aligned}&= |\exp(-t^2/2)(1+\alpha_T(t))\exp(O(|t|^3T^{-1/2}))-\exp(-t^2/2)|\\ &\leqslant C.\exp(-t^2/2)|t|^3T^{-1/2}\exp(O(|t|^3T^{-1/2}))\\ &\leqslant C.|t|^3T^{-1/2}\exp(-t^2/4)\end{aligned}$$

choosing ε sufficiently small.

This proves the lemma.

$$\text{Let}\quad Y(T) = \left(\frac{2\theta}{T}\right)^{1/2}\int_0^T X(s)dW(s).$$

Lemma 3.2 and the well known Esseen's lemma immediately yields

Lemma 3.3 : $$\sup_{x\in\mathcal{R}} |P(Y(T)\leqslant x)-\Phi(x)| \leqslant CT^{-1/2}.$$

We now state the main theorem.

Theorem 3.4 : $$\sup_{x\in\mathcal{R}}\left|P\left(\left(\frac{T}{2\theta}\right)^{1/2}(\theta_T-\theta)\leqslant x\right)-\Phi(x)\right| \leqslant C.T^{-1/2}$$

and the bound is uniform over any fixed compact subset of Θ.

Proof : Note that

$$\left(\frac{T}{2\theta}\right)^{1/2}(\theta_T-\theta) = \frac{Y(T)}{2\theta T^{-1}\int_0^T X^2(t)dt}$$

(2.1) yields $2\theta T^{-1}\int_0^T X^2(t)dt = 1-T^{-1/2}\left(\frac{2}{\theta}\right)^{1/2}\frac{Y(T)}{T}-\frac{X^2(T)}{T}$.

Let $B_1 = \{|Y(T)| \geq \delta \log T\}$

$B_2 = \{T^{-1/4}X^2(T) \geq \delta \log T\}$ where δ is large.

By lemma 3.3,

$$|P(B_1)-P(|N(0,1)| \geq \delta \log T)| \leq C.T^{-1/2}.$$

Using simple approximation for the tails of a normal distribution,

$$P(|N(0,1)| \geq \delta \log T) = O(T^{-1/2}).$$

Thus,

$$P(B_1) = O(T^{-1/2}). \qquad \ldots \quad (3.5)$$

Note that $X(T) \sim N\left(0, \frac{1-\exp(-2\theta T)}{2\theta}\right)$. The variance being bounded in T, simple Markov inequality gives

$$P(B_2) = O(T^{-1/2}).$$

On $B_1^c \cap B_2^c$, $\left| \left(\frac{T}{2\theta}\right)^{1/2}(\theta_T-\theta) \right| \leq C.\log T$ for some C.

Also by a simple binomial expansion of the denominator, on $B_1^c \cap B_2^c$,

$$\left(\frac{T}{2\theta}\right)^{1/2}(\theta_T-\theta) = Y(T)+T^{-1/2}\left(\frac{2}{\theta}\right)^{1/2} Y^2(T)+o(T^{-1/2}).$$

$$= Y(T)+T^{-1/2}aY^2(T)+o(T^{-1/2}), \text{ say.}$$

Now note that for any $|u| \leq C \log T$,

$$x+T^{-1/2}ax^2 \leq u \text{ iff}$$

$$\left(x+\frac{T^{1/2}}{2a}\right)^2 \leq \frac{uT^{1/2}}{a}+\frac{T}{4a^2}$$

iff $a_2(T, u) \leq x \leq a_1(T, u)$ where

$$a_1(T, u) = \left(\frac{uT^{1/2}}{a}+\frac{T}{4a^2}\right)^{1/2}-\frac{T^{1/2}}{2a}$$

$$= u+O(u^2/T)$$

$$= u+O(T^{-1}(\log T)^2))$$

and

$$a_2(T, u) = -\left(\frac{uT^{1/2}}{a}+\frac{T}{4a^2}\right)^{1/2}-\frac{T^{1/2}}{2a} \leq -C.T^{-1/2}.$$

But as in (3.5), $P(Y(T) \leqslant -C.T^{1/2}) = O(T^{-1/2})$,

and
$$|P(Y(T) \leqslant x+O(T^{-1}(\log T)^2)-\Phi(x)|$$
$$\leqslant C.T^{-1/2}+|\Phi(x)-\Phi(x+O(T^{-1}(\log T)^2))|$$
$$\leqslant C.T^{-1/2} \text{ uniformly over } x.$$

This proves the theorem completely.

Acknowledgements. The author is grateful to Prof. G. J. Babu of ISI, Calcutta, for his help and suggestions. Careful reading by the referee improved the presentation and is gratefully acknowledged.

References

Elliott, R. J. (1982): *Stochastic Calculus and Applications*, Springer Verlag, New York.

Hall, P. and Heyde, C. C. (1980): *Martingale Limit Theory and its Application*, Academic Press, New York.

Liptser, R. S. and Shriyayev, A. N. (1978): *Statistics of Random Processes* II, Springer Verlag, New York.

Mishra, M. N. and Prakasa Rao, B. L. S. (1985): On the Berry-Esseen theorem for maximum likelihood estimator for linear homogeneous diffusion processes. *Sankhyā*, 47, Ser. A, Pt. 3, 392-398.

Paper received : March, 1985.

Revised : August, 1985.