Sankhyā : The Indian Journal of Statistics
1981, Volume 43, Series A, Pt. 2, pp. 170–189

ASYMPTOTIC THEORY OF ESTIMATION IN NONLINEAR STOCHASTIC DIFFERENTIAL EQUATIONS

By B. L. S. PRAKASA RAO*
Indian Statistical Institute and *Purdue University*
and
HERMAN RUBIN**
Purdue University

SUMMARY. Strong consistency and asymptotic normality of an estimator for parameters in nonlinear stochastic differential equations are investigated by studying families of stochas- tic integrals using Fourier analytic methods.

1. Introduction

The study of inference problems for stochastic processes with both continuous and discrete time parameter is of extreme importance in view of the large number of applications. It has been found that the class of diffusion processes is amenable for statistical analysis among the class of continuous time processes. A survey of the recent work in this area with examples is given in Basawa and Prakasa Rao (1980). Further work on asymptotic theory of estimators for parameters of diffusion processes is discussed in Prakasa Rao (1981a, 1981b) and Lanska (1979).

Dorogovchev (1976) studied weak consistency of least square estimators for parameters of diffusion processes which are solutions of non-linear stochastic differential equations. Asymptotic normality and asymptotic efficiency of these estimators is investigated in Prakasa Rao (1979). Our aim in this paper is to study limiting properties of a process related to least squares estimator and hence to discuss the asymptotic properties of the maximum likelihood estimator derived from the limiting process. We study strong consistency and asymptotic normality of this estimator. Our approach here is entirely different from that of Dorogovchev (1976) and Prakasa Rao (1979). We believe that our techniques for study of families of stochastic integrals is new and is of independent interest.

*Research partially supported by the Office of Naval Research contract NOOO14-75-C 0455 at Purdue University.

**Research partially supported by NSF Grant MCS76-08316.

2. Study of a process related to least squares estimator

Let $\{X(t),\ t \geq 0\}$ be a real-valued stationary ergodic process satisfying the stochastic differential equation

$$dX(t) = f(\theta_0, X(t))dt + d\xi(t),\ X(0) = X_0,\ t \geq 0$$

where $\xi(t)$ is Wiener process with mean zero and variance $\sigma^2 t$, σ^2 known and $E[X_0^2] < \infty$. Suppose $f(\theta, x)$ is a known real-valued function continuous on $\Xi \times R$ where Ξ is a closed interval on the real line and $\theta_0 \in \Xi$ is unknown. Without loss of generality, assume that $\Xi = [-1, 1]$ and $\sigma^2 = 1$.

Suppose the process $\{X(t), 0 \leq t \leq T\}$ is observed at time points t_k, $k = 0, 1, \ldots, n-1$ with $t_0 = 0$ and $t_n = T$. Let

$$Q_n^T(\theta) = \sum_{k=0}^{n-1} \frac{[X(t_{k+1}) - X(t_k) - f(\theta, X(t_k))\Delta t_k]^2}{\Delta t_k}$$

where

$$\Delta t_k = t_{k+1} - t_k,\ 0 \leq k \leq n-1.$$

An estimator $\hat{\theta}_{n,T}$ which minimizes $Q_n^T(\theta)$ over $\theta \in \Xi$ is called a *least squares estimator* of θ. Assume that such an estimator exists. Note that if $\hat{\theta}_{n,T}$ minimizes $Q_n^T(\theta)$, then it minimizes $Q_n^T(\theta) - Q_n^T(\theta_0)$. This estimator is not consistent in general as $n \to \infty$ unless $T \to \infty$ such that the norm of the division tends to zero.

We shall first study the limiting properties of the process

$$\{Q_n^T(\theta) - Q_n^T(\theta_0),\ \theta \in \Xi\} \text{ for fixed } T > 0$$

as the norm of division

$$\Delta_n = \max_{0 \leq k \leq n-1} |t_{k+1} - t_k| \text{ tends to zero.}$$

Let

$$\Delta X_k = X(t_{k+1}) - X(t_k)$$

and

$$\Delta \xi_k = \xi(t_{k+1}) - \xi(t_k),\ 0 \leq k \leq n-1.$$

It is easy to check that

$$Q_n^T(\theta)-Q_n^T(\theta_0)$$

$$=\sum_k [f(\theta_0, X(t_k))-f(\theta, X(t_k))]^2\Delta t_k$$

$$+2\sum_k [f(\theta_0, X(t_k))-f(\theta, X(t_k))]\Delta\xi_k$$

$$+2\sum_k \{f(\theta_0, X(t_k))-f(\theta, X(t_k))\}$$

$$\times \int_{t_k}^{t_{k+1}} \{f(\theta_0, X(t))-f(\theta_0, X(t_k))\}\, dt$$

$$= I_{1n}+2I_{2n}+2I_{3n}. \qquad \ldots \quad (2.0)$$

Assume that the following regularity condition on $f(x, \theta)$ are satisfied.

Assumptions: (A1). $f(\theta, x)$ is continuous in (θ, x) and differentiable with respect to θ. Denote the first partial derivative of f with respect to θ by $f_\theta^{(1)}(\theta, x)$ and the derivative evaluated at θ_0 by $f_\theta^{(1)}(\theta_0, x)$.

(A2). $E[f_\theta^{(1)}(\theta_0, X(0))]^2 < \infty$.

(A3). $f_\theta^{(1)}(\theta, x)$ is Lipschitzian in θ for each x i.e., there exists $\alpha > 0$ such that

$$|f_\theta^{(1)}(\theta, x)-f_\theta^{(1)}(\phi, x)| \leqslant c(x)|\theta-\phi|^\alpha,\ x\in R, \theta, \phi \in \Xi$$

and

$$E[c^2(X(0))] < \infty.$$

(A4). $f(\theta, x)$ satisfies the following conditions:

(i) $|f(\theta, x)| \leqslant L(\theta)(1+|x|),\ \theta\in\Xi,\ x\in R;\ \sup\{L(\theta):\theta\in\Xi\}<\infty.$

(ii) $|f(\theta, x)-f(\theta, y)| \leqslant L(\theta)|x-y|,\ \theta\in\Xi,\ x, y\in R,$

(iii) $|f(\theta, x)-f(\phi, x)| \leqslant J(x)|\phi-\theta|,\ \theta, \phi\in\Xi,\ x\in R$

where $J(\cdot)$ is continuous and $E[J^2(X(0))] < \infty$.

(A5). $I(\theta) \equiv E[f(\theta, X(0))-f(\theta_0, X(0))]^2 > 0$ for $\theta\neq\theta_0$.

Remark: Since $E[X^2(0)] < \infty$, assumption A4(i) implies that

$$E[f(\theta, X(0))]^2 < \infty \quad \text{for all } \theta\in\Xi.$$

Since $f(\theta, x)$ is continuous in x and the process X has continuous sample paths with probability one, it follows that

$$I_{1n} \xrightarrow{\text{a.s.}} \int_0^T [f(\theta_0, X(t))-f(\theta, X(t))]^2 \, dt \qquad \dots \quad (2.1)$$

as $\Delta_n \to 0$. Assumption (A4) implies that

$$I_{2n} \xrightarrow{\text{q.m.}} \int_0^T [f(\theta_0, X(t))-f(\theta, X(t))] d\xi(t) \qquad \dots \quad (2.2)$$

as $\Delta_n \to 0$ in view of stationarity of the process X where the last integral is the Ito-stochastic integral.

Let us now estimate I_{3n}. In view of assumption (A4), it can be checked that

$$\left| \int_{t_k}^{t_{k+1}} \{f(\theta_0, X(t))-f(\theta_0, X(t_k))\} \, dt \right|$$

$$\leqslant L(\theta_0)\Delta t_k \sup_{t_k < t < t_{k+1}} |\xi(t)-\xi(t_k)|$$

$$+L^2(\theta_0)\Delta t_k^2 \sup_{t_k < t < t_{k+1}} \{1+|X(t)|\} \qquad \dots \quad (2.3)$$

for $0 \leqslant k \leqslant n-1$. Using assumption (A4) again, we obtain the following inequality :

$$I_{3n} \leqslant C(\theta_0) \left\{ \sum_k \Delta t_k \sup_{t_k < t < t_{k+1}} |\xi(t)-\xi(t_k)| + \sum_k \Delta t_k^2 \right\} |\theta-\theta_0|. \qquad \dots \quad (2.4)$$

Since Ξ is compact, it follows that

$$I_{3n} \leqslant C^*(\theta_0) \left\{ \sum_k \Delta t_k (2\Delta t_k \log_2 1/\Delta t_k)^{1/2} + \sum_k \Delta t_k^2 \right\} \text{ a.s.}$$

whenever Δ_n is sufficiently small by the law of iterated logarithm for Brownian increments (cf. McKean, 1969, p. 14). Therefore

$$I_{3n} = O\left(\sum_k \Delta t_k^{3/2} \log_2^{1/2} 1/\Delta t_k \right) \text{ a.s.} \qquad \dots \quad (2.5)$$

Relations (2.1), (2.2) and (2.5) show that, for any fixed T, $Q_n^T(\theta)-Q_n^T(\theta_0)$ converges in probability to $R_T(\theta)$ as $n \to \infty$ where $R_T(\theta)$ is defined by

$$R_T(\theta) = \int_0^T [f(\theta_0, X(t))-f(\theta, X(t))]^2 \, dt$$

$$+2 \int_0^T [f(\theta_0, X(t))-f(\theta, X(t))]d\xi(t)$$

$$= \int_0^T v^2(\theta, X(t))dt - 2 \int_0^T v(\theta, X(t))d\xi(t) \qquad \dots \quad (2.6)$$

where

$$v(\theta, x) = f(\theta, x) - f(\theta_0, x). \qquad \dots \quad (2.7)$$

We study the limiting properties of the process $\{R_T(\theta), \theta \in \Xi\}$ in the next section.

3. Study of the limiting process related to least squares estimator

Let us now study the properties of the limiting process

$$Z_T(\theta) \equiv \frac{1}{\sqrt{T}} \int_0^T v(\theta, X(t))d\xi(t) \qquad \dots \quad (3.1)$$

as a process in the parameter $\theta \in \Xi = [-1, 1]$ as $T \to \infty$. From the central limit theorem for stochastic integrals (cf. Basawa and Prakasa Rao, 1980), it can be shown that

$$\frac{1}{\sqrt{T}} \int_0^T v(\theta, X(t))d\xi(t) \xrightarrow{\mathcal{L}} N(0, E[v(\theta, X(0))]^2)$$

since the process X is stationary ergodic. In general, finite dimensional distributions of the process $\{Z_T(\theta), \theta \in \Xi\}$ converge to the finite dimensional distributions of the Gaussian process $\{Z(\theta), \theta \in \Xi\}$ with mean zero and covariance function

$$R(\theta_1, \theta_2) = E[v(\theta_1, X(0))v(\theta_2, X(0))].$$

We shall now prove the weak convergence of the process $\{Z_T(\theta), \theta \in \Xi\}$ on $C[-1, 1]$ under uniform norm. It is sufficient to prove that

$$\lim_{T \to \infty} \overline{\lim_{\delta \to 0}} P\left(\sup_{|\theta-\phi|<\delta} |Z_T(\theta)-Z_T(\phi)| > \epsilon \right) = 0. \qquad \dots \quad (3.2)$$

Since $v(\theta, x)$ is differentiable with respect to θ on $[-1, 1]$ by assumption (A1), it is easy to see that there exists a cubic polynomial $g(\theta, x)$ in θ such that

$$g(-1, x) = v(-1, x), g(1, x) = v(1, x)$$

and

$$g_\theta^{(1)}(-1, x) = v_\theta^{(1)}(-1, x), g_\theta^{(1)}(1, x) = v_\theta^{(1)}(1, x).$$

Let

$$h(\theta, x) = v(\theta, x) - g(\theta, x).$$

Then,

$$h(-1, x) = h(1, x) = 0, h_\theta^{(1)}(-1, x) = h_\theta^{(1)}(1, x) = 0.$$

Now

$$Z_T(\theta) = \frac{1}{\sqrt{T}} \int_0^T h(\theta, X(t))d\xi(t) + \frac{1}{\sqrt{T}} \int_0^T g(\theta, X(t))d\xi(t). \qquad \dots \quad (3.3)$$

Since $g(\theta, x)$ is a cubic polynomial in θ with coefficients in x which are linear functions of $v(-1, x)$, $v(1, x)$, $v_\theta^{(1)}(-1, x)$ and $v_\theta^{(1)}(1, x)$, it is easy to check the uniform equi-continuity condition of type (3.2) for,

$$\frac{1}{\sqrt{T}} \int_0^T g(\theta, X(t))d\xi(t).$$

Let us now consider the process

$$W_T(\theta) = \frac{1}{\sqrt{T}} \int_0^T h(\theta, X(t))d\xi(t). \qquad \dots \quad (3.4)$$

Let the Fourier expansion for $h(\theta, x)$ in $L_2([-1, 1])$ be given by

$$h(\theta, x) = \sum_n a_n(x)\, e^{\pi i n\theta}, \quad x \in R. \qquad \dots \quad (3.5)$$

Lemma 3.1 :

$$\int_0^T h(\theta, X(t))d\xi(t) = \sum_n \left\{ \int_0^T a_n(X(t))d\xi(t) \right\} e^{\pi i n\theta} \qquad \dots \quad (3.6)$$

in the sense of convergence in quadratic mean.

Proof: An approximating sum in L_2-norm for

$$\int_0^T h(\theta, X(t))d\xi(t)$$

is

$$A_{1N} = \sum_{j=1}^{N} h(\theta, X(t_{j-1}))\Delta\xi_j$$

and an approximating sum in L_2-norm for $\sum_n \left\{ \int_0^T a_n(X(t))d\xi(t) \right\} e^{\pi in\theta}$ is

$$A_{2NM} = \sum_{|n| \leq M} e^{\pi in\theta} \left(\sum_{j=1}^{N} a_n(X(t_{j-1}))\, \Delta\xi_j \right).$$

It is sufficient to prove that $E|A_{1N}-A_{2NM}|^2 \to 0$ as $N \to \infty$ and $M \to \infty$. Now

$$E|A_{1N}-A_{2NM}|^2 = E\left| \sum_{j=1}^{N} \left\{ h(\theta, X(t_{j-1})) - \sum_{n=-M}^{N} e^{\pi in\theta} a_n(X(t_{j-1})) \right\} \Delta\xi_j \right|^2$$

$$= E\left| \sum_{j=1}^{N} \sum_{|n| > M} a_n(X(t_{j-1}))\, e^{\pi in\theta} \Delta\xi_j \right|^2$$

$$\leq \left[\sum_{|n| > M} \left\{ E\left(\sum_{j=1}^{N} a_n(X(t_{j-1}))\Delta\xi_j \right)^2 \right\}^{\frac{1}{2}} \right]^2$$

by the elementary inequality

$$E\left| \sum_n \lambda_n \gamma_n \right|^2 \leq \left(\sum_n |\lambda_n| (E(\gamma_n^2))^{\frac{1}{2}} \right)^2$$

for any sequence of complex numbers $\{\lambda_n\}$ and any sequence of real valued random variables $\{\gamma_n, n \geq 1\}$. Hence

$$E|A_{1N}-A_{2NM}|^2 \leq \left[\sum_{|n| > M} \left\{ \sum_{j=1}^{N} E(a_n(X(t_{j-1}))^2 \Delta t_j \right\}^{\frac{1}{2}} \right]^2.$$

Since,

$$\sum_{j=1}^{N} E(a_n(X(t_{j-1}))^2\, \Delta t_j \to \int_0^T E\{a_n(X(t)\}^2 dt = T\mu_n \text{ (say)}$$

as $N \to \infty$, it is sufficient to prove that $\sum_n \mu_n^{1/2} < \infty$. This follows from remarks following Lemma 3 of the Appendix under assumption (A3). Let

$$W_n = \frac{1}{\sqrt{T}} \int_0^T a_n(X(t))d\xi(t). \quad \ldots \quad (3.7)$$

Lemma 3.2 : *For every $\epsilon > 0$,*

$$\lim_{\delta \to 0} P\left(\sup_{|\theta-\phi| < \delta} |W_T(\theta) - W_T(\phi)| > \epsilon \right) = 0 \quad \ldots \quad (3.8)$$

for every $T > 0$.

Proof: In view of Lemma 3.1, for any $\varepsilon > 0$,

$$P\left(\sup_{|\theta-\phi|<\delta} |W_T(\theta)-W_T(\phi)| > \varepsilon\right)$$

$$= P\left(\sup_{|\theta-\phi|<\delta} \left|\sum_{n=-\infty}^{\infty} W_n(e^{\pi in\theta}-e^{\pi in\phi})\right| > \varepsilon\right)$$

$$\leqslant P\left(\sup_{|\theta-\phi|<\delta} \sum_{n=-\infty}^{\infty} |W_n|\,|e^{\pi in\theta}-e^{\pi in\phi}| > \varepsilon\right). \quad \ldots \quad (3.9)$$

Let n_0 be chosen so that

$$\sum_{n=n_0}^{\infty} \mu_n^{1/3} < \varepsilon\, 2^{-4/3}. \quad \ldots \quad (3.10)$$

This is possible since $\sum_{n=1}^{\infty} \mu_n^{1/3} < \infty$ by Lemma 3 of the appendix. Inequality (3.9) implies that

$$P\left(\sup_{|\theta-\phi|<\delta} |W_T(\theta)-W_T(\phi)| > \varepsilon\right)$$

$$\leqslant P\left(\sup_{|\theta-\phi|<\delta} \sum_{n=-n_0}^{n_0} |W_n|\,n\,|\theta-\phi| > \frac{\varepsilon}{4\pi}\right) + P\left(\sum_{|n|>n_0} |W_n| > \frac{\varepsilon}{2}\right)$$

$$\leqslant \sum_{n=1}^{n_0} P\left(|W_n| > \frac{\varepsilon}{2\pi n_0\delta}\right) + 2\sum_{n=n_0+1}^{\infty} P(|W_n| > \varepsilon_n)$$

$$\left(\text{Here } \varepsilon_n = \frac{\varepsilon}{2^{4/3}}\,\mu_n^{1/3}\left(\sum_{n=n_0+1}^{\infty} \mu_n^{1/3}\right)^{-1}\right)$$

$$\leqslant \left(\frac{2\pi n_0\delta}{\varepsilon}\right)^2 \sum_{n=1}^{n_0} \mu_n + \sum_{n=n_0+1}^{\infty} \frac{\mu_n}{\varepsilon_n^2}$$

(since $E(W_n) = 0$ and $\mathrm{var}(W_n) = \mu_n$)

$$= \frac{(2\pi n_0\delta)^2}{\varepsilon^2} \sum_{n=1}^{n_0} \mu_n + \frac{8}{\varepsilon^2}\left(\sum_{n=n_0+1}^{\infty} \mu_n^{1/3}\right)^3$$

$$= C_{n_0}\frac{\delta^2}{\varepsilon^2} + \frac{8}{\varepsilon^2}\left(\frac{\varepsilon}{2}\right)^3$$

where C_{n_0} depends only on n_0. Choosing δ such that

$$C_{n_0}\frac{\delta^2}{\varepsilon^2} < \varepsilon \quad \text{i.e.} \quad 0 < \delta < \left(\frac{\varepsilon^3}{2C_{n_0}}\right)^{\frac{1}{2}}$$

we have the inequality

$$P\left(\sup_{|\theta-\phi|<\delta} |W_T(\theta)-W_T(\phi)| > \varepsilon\right) \leqslant 2\varepsilon$$

for every $0 < \delta < \left(\frac{\varepsilon^3}{2C_{n_0}}\right)^{\frac{1}{2}}$ and for every $T > 0$. This proves (3.8). In view of Lemma 3.2 and the remarks made earlier, we have the following theorem.

Theorem 3.1 : *The family of stochastic processes* $\{Z_T(\theta),\ \theta \in \Xi\}$ *on* $C[-1, 1]$ *converge in distribution to the Gaussian process with mean zero and covariance function*

$$R(\theta_1, \theta_2) = E[v(\theta_1, X(0))v(\theta_2, X(0))]$$

as $T \to \infty$.

4. Strong consistency

Let us now consider the limiting processes $R_T(\theta)$ defined by (2.6). Suppose there exists an estimator $\hat{\theta}_T$ which minimizes

$$R_T(\theta) \equiv \int_0^T \{f(\theta, X(t))-f(\theta_0, X(t))\}^2 dt$$

$$-2\int [f(\theta, X(t))-f(\theta_0, X(t))]d\xi(t) \qquad \dots \quad (4.1)$$

over $\theta \in \Xi$.

Let μ_θ be the measure generated by the process X on $C[0, T]$ when θ is the true parameter. From the general theory of diffusion processes, the Radon-Nikodym derivative of μ_θ with respect to μ_{θ_0} exists and is given by

$$\frac{d\mu_\theta}{d\mu_{\theta_0}} = \exp\left\{\int_0^T \{f(\theta, X(t))-f(\theta_0, X(t))\}\, d\xi(t)\right.$$

$$\left. -\frac{1}{2}\int_0^T \{f(\theta, X(t))-f(\theta_0, X(t))\}^2 dt\right\}. \qquad \dots \quad (4.2)$$

(cf. Gikhman and Skorokhod (1972), p. 90). Hence

$$\log \frac{d\mu_\theta}{d\mu_{\theta_0}} = -\frac{1}{2} R_T(\theta)$$

which proves that the estimator θ_T is the same as the maximum likelihood estimator $\bar{\theta}_T$ of θ (cf. Basawa and Prakasa Rao (1980)) when the process X is observed over $[0, T]$. Let

$$I_T(\theta) \equiv \int_0^T [f(\theta, X(t))-f(\theta_0, X(t))]^2 dt \qquad \dots \quad (4.3)$$

and W^* be a standard Wiener process. Since the solution of the stochastic differential equation given in Section 2 is stationary ergodic by hypothesis, it follows that $I_T(\theta) \to \infty$ a.s. for $\theta \neq \theta_0$ by (A5) and the process $\{R_T(\theta)\}$ can be identified with the process $\{I_T(\theta)-2W^*(I_T(\theta)\}$. Furthermore

$$I_T(\theta)-2W^*(I_T(\theta)) \to \infty \text{ a.s.} \qquad \dots \quad (4.4)$$

as $T \to \infty$ for any $\theta \neq \theta_0$. Hence θ and θ_0 are pairwise consistent. Note that

$$R_T(\theta) = I_T(\theta)-2\sqrt{T}\, Z_T(\theta), \quad \theta \in \Xi, \quad T > 0 \qquad \dots \quad (4.5)$$

where $I_T(\theta)$ is defined by (4.3) and $Z_T(\theta)$ is given by (3.1). Let

$$Z_T^*(\theta) = \sqrt{T}\, Z_T(\theta). \qquad \dots \quad (4.6)$$

It is obvious that

$$\frac{1}{T} I_T(\theta) \to I(\theta) \text{ a.s.} \quad \text{as } T \to \infty \qquad \dots \quad (4.7)$$

by the ergodic theorem. Note that

$$I_T(\theta)-I_T(\phi) = \int_0^T \{f(X(t), \theta)-f(X(t), \phi)\} \cdot$$

$$\cdot \{f(X(t), \phi)+f(X(t), \theta)-2f(X(t), \theta_0)\} dt$$

and therefore

$$|I_T(\theta)-I_T(\phi)| \leqslant |\theta-\phi| \int_0^T J(X(t)) \cdot \{L(\theta)+L(\phi)+2L(\theta_0)\}\{1+|X(t)|\} dt$$

$$\leqslant C_1 |\theta-\phi| \int_0^T J(X(t)(\{1+|X(t)|\} dt.$$

Since $E[J^2(X(0)] < \infty$ and $E[X^2(0)] < \infty$, it follows that $E[J(X(0))X(0)] < \infty$ and hence, by the ergodic theorem,

$$\frac{1}{T}\int_0^T J(X(t))\{1+|X(t)|\}dt \xrightarrow{a.s.} E[J(X(0))\{1+|X(0)|\}] < \infty \text{ as } T\to\infty.$$

Therefore,

$$|I_T(\theta)-I_T(\phi)| \leqslant C^* T|\theta-\phi| \text{ a.s.} \quad \dots \quad (4.8)$$

as $T\to\infty$ for some constant $C^*>0$. In view of (4.7) it follows that

$$\frac{I_T(\theta)}{T} \xrightarrow{a.s.} I(\theta) \equiv E[f(\theta, X(0))-f(\theta_0, X(0))]^2 \quad \dots \quad (4.9)$$

uniformly in $\theta \in \Xi$ as $T\to\infty$. But $I_T(\theta_0)=0$ and $\underset{T\to\infty}{\lim} \frac{I_T(\theta)}{T} > 0$ a.s. for $\theta \neq \theta_0$ by (A5). Hence, for any $\delta > 0$.

$$\inf_{|\theta-\theta_0|>\delta} \frac{I_T(\theta)}{T} \xrightarrow{a.s.} \lambda \text{ as } T\to\infty \quad \dots \quad (4.10)$$

for some $\lambda > 0$ depending on δ.

Lemma 4.1 : *Under the assumptions* (A1)–(A4), *for any* $T_0 > 0$ *and any* $\epsilon > 0$,

$$P\left(\sup_{\theta}\sup_{0<T\leqslant T_0} |Z_T^*(\theta)| > \epsilon\right) \leqslant C_2\frac{T_0}{\epsilon^2} \quad \dots \quad (4.11)$$

for some constant $C_2 > 0$.

Proof: Let $h(\theta, x)$ and $g(\theta, x)$ be defined as in Section 3 and

$$h(\theta, x) = \Sigma\, a_n(x)e^{\pi i n\theta}, \quad \theta\in[-1, 1].$$

Let

$$W_n^* = \int_0^T a_n(X(t))d\xi(t).$$

Since $g(\theta, x)$ is a cubic polynomial in θ with coefficients in x, it is easy to check, by Kolmogorov's inequality, that

$$\sup_{\theta}\sup_{0\leqslant T\leqslant T_0}\left|\int_0^T g(\theta, X(t))d\xi(t)\right| = O_p(T_0^{\frac{1}{2}}) \quad \dots \quad (4.12)$$

using the fact that $|\theta| \leqslant 1$. On the other hand, for any $\varepsilon > 0$,

$$P\left(\sup_{\theta} \sup_{0<T<T_0} \left| \sum_n \left\{ \int_0^T a_n(X(t))d\xi(t) \right\} e^{\pi i n\theta} \right| > \varepsilon \right)$$

$$\leqslant P\left(\sup_{0<T<T_0} \sum_n \left| \int_0^T a_n(X(t))d\xi(t) \right| > \varepsilon \right)$$

$$\leqslant \sum_n P\left(\operatorname{spu}_{0<T<T_0} \left| \int_0^T a_n(X(t))d\xi(t) \right| > \varepsilon_n \right)$$

(where $\Sigma\, \varepsilon_n \leqslant \varepsilon$)

$$\leqslant \sum_n \frac{1}{\varepsilon_n^2} \operatorname{var}\left(\int_0^{T_0} a_n(X(t))d\xi(t) \right)$$

(by Kolmogorov's inequality for martingales)

$$\leqslant \sum_n \frac{1}{\varepsilon_n^2} \int_0^{T_0} E(a_n(X(t)))^2 dt$$

$$= T_0 \sum_n \frac{\mu_n}{\varepsilon_n^2}$$

$$= \frac{T_0}{\varepsilon^2} (\Sigma\, \mu_n^{1/3})^3. \qquad \ldots \quad (4.13)$$

where ε_n is chosen to be $\varepsilon \mu_n^{1/3} \left(\sum_n \mu_n^{1/3} \right)^{-1}$. Note that $M \equiv \Sigma\, \mu_n^{1/3} < \infty$. Hence relations (4.12) and (4.13) together prove that

$$P\left(\sup_{\theta} \sup_{0 \leqslant T \leqslant T_0} | Z_T^*(\theta) | > \varepsilon \right) \leqslant C_2 \frac{T_0}{\varepsilon^2}$$

for some constant $C_2 > 0$ independent of T_0 and ε.

Lemma 4.2 : *For any $\gamma > 1/2$, there exists $H > 0$ such that*

$$\limsup_{T \to \infty} \sup_{\theta} \frac{| Z_T^*(\theta) |}{T^{1/2}(\log T)^{\gamma}} \leqslant H \text{ a.s.} \qquad \ldots \quad (4.14)$$

Proof: Let

$$A_n = \Big[\sup_{2^{n-1} < T < 2^n} \sup_{\theta} |Z_T^*(\theta)| > H' 2^{n/2} n^{\gamma} \Big], \quad n \geqslant 1.$$

Observe that Lemma 4.1 gives the inequality

$$P(A_n) = P\Big[\sup_{0 < T < 2^{n-1}} \sup_{\theta} |Z_T^*(\theta)| > H' 2^{n/2} n^{\gamma} \Big]$$

(by stationarity of the process $X(t)$)

$$\leqslant \frac{C 2^{n-1}}{H'^2 2^n n^{2\gamma}} = \frac{C}{2H'^2} \frac{1}{n^{2\gamma}}.$$

Hence $\sum_{n=1}^{\infty} P(A_n) < \infty$ which implies that $P(A_n$ occurs infinitely often$) = 0$ by Borel-Cantelli Lemma. Therefore $\sup_{\theta} |Z_T^*(\theta)| \leqslant H' 2^{n/2} n^{\gamma}$ for all $2^{n-1} < T \leqslant 2^n$ except for finitely many n with probability one and hence (4.4) holds for suitable $H > 0$ depending on γ.

Theorem 4.1 : *Under the assumptions* (A1)-(A5),

$$\hat{\theta}_T \to \theta_0 \text{ a.s. as } T \to \infty.$$

Proof: Note that

$$R_T(\theta) = I_T(\theta) - 2Z_T^*(\theta)$$

and $R_T(\theta_0) = 0$. Furthermore, for any $\delta > 0$, there exists $\lambda > 0$ depending on δ such that

$$\inf_{|\theta - \theta_0| \geqslant \delta} I_T(\theta) \geqslant T\lambda \quad \text{a.s. as } T \to \infty$$

by (4.10) and with probability one, for any $\gamma > \frac{1}{2}$, there exists $H > 0$ depending on γ such that

$$\sup_{\theta} |Z_T^*(\theta)| \leqslant H T^{1/2} (\log T)^{\gamma} \text{ a.s.}$$

for sufficiently large T. Hence

$$\inf_{|\theta - \theta_0| \geqslant \delta} R_T(\theta) \geqslant \lambda^* T > 0 \quad \text{a.s. as } T \to \infty$$

for some $\lambda^* > 0$ depending on δ and γ. Since $\hat{\theta}_T$ minimizes $R_T(\theta)$ and $R_T(\theta_0) = 0$, it follows that $|\hat{\theta}_T - \theta_0| \leqslant \delta$ a.s. as $T \to \infty$. Hence $\hat{\theta}_T \to \theta_0$ a.s. as $T \to \infty$.

5. Asymptotic normality

In addition to the conditions (A1)–(A5) assumed in Section 2, let us suppose that there exists a neighbourhood V_{θ_0} of θ_0 such that

$$\text{(A6)} \qquad |f_\theta^{(1)}(\theta, x)| \leqslant M(\theta)(1+|x|), \quad \theta \in V_{\theta_0}$$

and

$$\sup \{M(\theta) : \theta \in V_{\theta_0}\} = M < \infty.$$

We shall now obtain the asymptotic distribution of $\hat{\theta}_T$ under the conditions (A1)–(A6). Since $\hat{\theta}_T$ is strongly consistent $\hat{\theta}_T \in V_{\theta_0}$ with probability one for large T. Expanding $f(\theta, x)$ in a neighbourhood of θ_0, we have

$$f(\theta, x) = f(\theta_0, x) + (\theta-\theta_0) f_\theta^{(1)}(\tilde{\theta}, x)$$

where $|\tilde{\theta}-\theta_0| \leqslant |\theta-\theta_0|$ and hence

$$I_T(\theta) \equiv \int_0^T \{f(\theta, X(t)) - f(\theta_0, X(t))\}^2 \, dt$$

$$= (\theta-\theta_0)^2 \int_0^T \{f_\theta^{(1)}(\theta_0, X(t))\}^2 dt$$

$$+ (\theta-\theta_0)^2 \int_0^T [\{f_\theta^{(1)}(\tilde{\theta}, X(t))\}^2 - \{f_\theta^{(1)}(\theta_0, X(t))\}^2] dt. \qquad \ldots (5.1)$$

Observe that

$$|\{f_\theta^{(1)}(\tilde{\theta}, x)\}^2 - \{f_\theta^{(1)}(\theta_0, x)\}^2|$$

$$= |f_\theta^{(1)}(\tilde{\theta}, x) - f_\theta^{(1)}(\theta_0, x)| \, |f_\theta^{(1)}(\tilde{\theta}, x) + f_\theta^{(1)}(\theta_0, x)|$$

$$\leqslant 2M |\tilde{\theta}-\theta_0|^\alpha c(x)(1+|x|) \qquad \ldots (5.2)$$

by assumptions (A3) and (A6). Therefore

$$\left| I_T(\theta) - (\theta-\theta_0)^2 \int_0^T \{f_\theta^{(1)}(\theta_0, X(t))\}^2 \, dt \right|$$

$$\leqslant 2M |\theta-\theta_0|^{2+\alpha} \int_0^T c(X(t))(1+|X(t)|) \, dt. \qquad \ldots (5.3)$$

Let us write $\theta-\theta_0 = T^{-1/2}\psi$. Then it follows that

$$\sup_{|\psi| \leqslant A_T} \left| I_T(\theta) - \psi^2 T^{-1} \int_0^T \{f_\theta^{(1)}(\theta_0, X(t))\}^2 \, dt \right| \leqslant M_1 A_T^{2+\alpha} T^{-1-\alpha} \qquad \ldots (5.4)$$

for some constant $M_1 > 0$ by the orogodic theorem since

$$E(c(X(0))(1+|X(0)|)) < \infty.$$

On the other hand, let

$$v_T(\psi, x) = T^{1/2}[f(\theta_0+\psi T^{-1/2}, x)-f(\theta_0, x)-\psi T^{-1/2}f_\theta^{(1)}(\theta_0, x)]$$

for $|\psi| \leqslant A_T$. Then $v_T(\psi, X)$ is differentiable with respect to ψ and the derivative $v_T^{(1)}(\psi, x)$ satisfies

$$v_T^{(1)}(\psi, x)-v_T^{(1)}(\zeta, x) = f_\theta^{(1)}(\theta_0+\psi T^{-1/2}, x)-f_\theta^{(1)}(\theta_0+\zeta T^{-1/2}, x)$$

and hence

$$|v_T^{(1)}(\psi, x)-v_T^{(1)}(\zeta, x)| \leqslant c(x)T^{-\alpha/2}|\psi-\zeta|^\alpha \quad \ldots \quad (5.5)$$

by (A3) for all ψ, ζ in $[-A_T, A_T]$. It can to shown that there exists a polynomial in ψ with coefficients in x viz.

$$g_T(\psi, x) = v_T(A_T, x)P_1\left(\frac{\psi}{A_T}\right)+A_T v_T^{(1)}(A_T, x)P_2\left(\frac{\psi}{A_T}\right)$$

$$+v_T(-A_T, x)P_3\left(\frac{\psi}{A_T}\right)+A_T v_T^{(1)}(-A_T, x)P_4\left(\frac{\psi}{A_T}\right) \quad \ldots \quad (5.6)$$

on $[-A_T, A_T]$ such that

$$g_T(A_T, x) = v_T(A_T, x), g_T(-A_T, x) = v_T(-A_T, x), \quad \ldots \quad (5.7)$$

$$g_T^{(1)}(A_T, x) = v_T^{(1)}(A_T, x) \text{ and } g_T^{(1)}(-A_T, x) = v_T^{(1)}(-A_T, x) \quad \ldots \quad (5.8)$$

where P_i, $1 \leqslant i \leqslant 4$ are polynomials in $\dfrac{\psi}{A_T}$ with constant coefficients. Observing that $v_T(0, x) = v_T^{(1)}(0, x) = 0$, it is easy to check that

$$|g_T^{(1)}(A_T, x)| \leqslant c(x)A_T^\alpha T^{-\alpha/2}, \quad \ldots \quad (5.5)$$

$$|g_T^{(1)}(-A_T, x)| \leqslant c(x)A_T^\alpha T^{-\alpha/2}, \quad \ldots \quad (5.10)$$

$$|g_T(A_T, x)| \leqslant c(x)A_T^{1+\alpha}T^{-\alpha/2}, \quad \ldots \quad (5.11)$$

and

$$|g_T(-A_T, x)| \leqslant c(x)A_T^{1+\alpha}T^{-\alpha/2}. \quad \ldots \quad (5.12)$$

Furthermore there exists a constant $M_2 > 0$ independent of T such that

$$|g_T^{(1)}(\psi, x) - g_T^{(1)}(\zeta, x)| \leqslant M_2 c(x) A_T^{\alpha-1} T^{-\alpha/2} |\psi - \zeta| \qquad \dots \quad (5.13)$$

for all $\psi, \zeta \in [-A_T, A_T]$. But

$$A_T^{\alpha-1} |\psi - \zeta| \leqslant 2^{1-\alpha} |\psi - \zeta|^\alpha$$

since $|\psi - \zeta| \leqslant 2A_T$. Hence there exists a constant $M_3 > 0$ independent of T such that

$$|g_T^{(1)}(\psi, x) - g_T^{(1)}(\zeta, x)| \leqslant M_3 c(x) T^{-\alpha/2} |\psi - \zeta|^\alpha \qquad \dots \quad (5.14)$$

for all $\psi, \zeta \in [-A_T, A_T]$. Renormalizing, we get that

$$|g_T^{(1)}(\psi^*, x) - g_T^{(1)}(\zeta^*, x)| \leqslant M_3 c(x) A_T^\alpha |\psi^* - \zeta^*| T^{-\alpha/2} \qquad \dots \quad (5.15)$$

for all $\psi^*, \zeta^* \in [-1, 1]$. Let

$$h_T(\psi^*, x) = v_T(\psi^*, x) - g_T(\psi^*, x). \qquad \dots \quad (5.16)$$

Then there exists a constant $M_3^* > 0$ independent of T such that

$$|h_T^{(1)}(\psi^*, x) - h_T^{(1)}(\zeta^*, x)| \leqslant M_3^* c(x) A_T^\alpha |\psi^* - \zeta^*|^\alpha T^{-\alpha/2} \qquad \dots \quad (5.17)$$

for all $\psi^*, \zeta^* \in [-1, 1]$ by relations (5.6) and (5.15). Now, applying Fourier series methods as in Lemma 4.1, it can be shown that for every $\varepsilon > 0$,

$$P\left(\sup_{|\psi^*| \leqslant 1} \left| \int_0^T v_T(\psi^*, X(t)) d\xi(t) \right| > \varepsilon \right) \leqslant \frac{M_4 T}{\varepsilon^2} A_T^{2\alpha} T^{-\alpha} E[c^2(X(0))]$$

and hence

$$P\left(\sup_{|\psi| \leqslant A_T} \left| \int_0^T \{f(\theta_0 + \psi T^{-1/2}, X(t)) - f(\theta_0, X(t)) - \psi T^{-1/2} f_\theta^{(1)}(\theta_0, X(t))\} d\xi(t) \right| > \varepsilon \right)$$

$$\leqslant \frac{M_4}{\varepsilon^2} A_T^{2\alpha} T^{-\alpha} E[c^2(X(0))]. \qquad \dots \quad (5.18)$$

Let us choose $A_T = \log T$. Since

$$\frac{1}{T} \int_0^T \{f_\theta^{(1)}(\theta_0, X(t))\}^2 dt \to I(\theta_0) = E[f_\theta^{(1)}(\theta_0, X(0))]^2 \text{ a.s.}$$

as $T \to \infty$ by the ergodic theorem and

$$\frac{1}{\sqrt{T}} \int_0^T f_\theta^{(1)}(\theta_0, X(t)) d\xi(t) \xrightarrow{\mathcal{L}} N(0, I(\theta_0)) \quad \text{as} \quad T \to \infty$$

by the central limit theorem for stochastic integrals (cf. Basawa and Prakasa Rao (1980)), relations (5.4) and (5.18) imply that the asymptotic distribution of $\hat{\theta}_T$ which minimizes $R_T(\theta)$ given by (2.6) can be obtained from the process

$$\psi^2 I(\theta_0)-2\psi Z, \ -\infty < \psi < \infty \qquad \dots \quad (5.19)$$

where Z is normal with mean 0 and variance $I(\theta_0)$. Since

$$\hat{\psi} = Z/I(\theta_0)$$

minimizes (5.16), it follows that

$$T^{1/2}(\hat{\theta}_T-\theta_0) \xrightarrow{\mathcal{L}} N(0, 1/I(\theta_0)) \text{ as } T \to \infty. \qquad \dots \quad (5.20)$$

This result is obtained under stronger conditions in Prakasa Rao (1979) for the least squares estimator $\hat{\theta}_{n,T}$ as $n \to \infty$ and $T \to \infty$ defined at the beginning of Section 2. Results obtained in this section as well as the earlier sections can be easily extended to the case when σ^2 is unknown.

Appendix

Lemma 1: *Suppose $\phi(u)$ is square integrable on $[-1, 1]$ and $\phi(.)$ is Lipschitz of order α i.e., there exists $c > 0$ such that*

$$|\phi(u)-\phi(v)| \leqslant c|u-v|^{\alpha}. \qquad \dots \quad (1)$$

Let $\phi(u) = \sum_n a_n e^{\pi i n u}$. Then, for any $0 < \gamma < \alpha$, there exists $K_1(\alpha, \gamma) > 0$ such that

$$\sum_n |a_n|^2 n^{2\gamma} \leqslant K_1(\alpha, \gamma)c^2. \qquad \dots \quad (2)$$

Proof: It is easy to check that

$$\int_{-1}^{1} |\phi(u+h)-\phi(u-h)|^2 \, du = 4 \sum_n |a_n|^2 \sin^2 \pi n h. \qquad \dots \quad (3)$$

Since ϕ is Lipschitz satisfying (1), it follows that

$$4 \sum_n |a_n|^2 \sin^2 \pi n h \leqslant 2^{2\alpha+1} c^2 h^{2\alpha} \qquad \dots \quad (4)$$

for all $h \in [0, 1]$. Let $h = 2^{-k}$ and $2^{k-2} < n \leqslant 2^{k-1}$. It is clear that $\sin^2 \pi n h \geqslant \frac{1}{2}$ and relation (4) shows that

$$\sum_{n=2^{k-2}+1}^{2^{k-1}} |a_n|^2 \leqslant 2^{2\alpha} c^2 2^{-2k\alpha} \qquad \dots \quad (5)$$

for any $k \geqslant 2$ and hence for any $0 < \gamma < \alpha$,

$$\sum_{n=2^{k-2}+1}^{2^{k-1}} |a_n|^2 n^{2\gamma} \leqslant 2^{2\alpha} c^2 2^{(2\gamma-2\alpha)k}. \quad \ldots \quad (6)$$

Summing over all $k \geqslant 2$, we obtain that

$$\sum_n |a_n|^2 n^{2\gamma} \leqslant 2^{2\alpha} c^2 (1-2^{(2\gamma-2\alpha)})^{-1}. \quad \ldots \quad (7)$$

Hence there exists a constant $K_1(\alpha, \gamma) > 0$ such that

$$\sum_n |a_n|^2 n^{2\gamma} \leqslant K_1(\alpha, \gamma) c^2 \quad \ldots \quad (8)$$

where c is the Lipschitzian constant given by (1).

Remark : A slight variation of the above result is due to Szasz (1922). The proof given above is the same as in Szasz (1922) and is given here for completeness.

Lemma 2 : *Suppose $h(u)$ is square integrable ion $[-1, 1]$ with $h(-1) = h(1) = 0$. Further suppose that $h'(\cdot)$ exists and is Lipschitzian of order α i.e., there exists $c > 0$ such that*

$$|h'(u)-h'(v)| \leqslant c|u-v|^{\alpha}. \quad \ldots \quad (9)$$

Let $h(u) = \sum_n a_n e^{\pi i n u}$. Then, for any $0 < \gamma < \alpha$, there exists $K(\alpha, \gamma) > 0$, $i = 2, 3$ such that

$$\sum_n |a_n|^2 n^{2+2\gamma} \leqslant K_2(\alpha, \gamma) c^2 \quad \ldots \quad (10)$$

and

$$\sum_n |a_n|^{2/3} \leqslant K_3(\alpha, \gamma) c^2. \quad \ldots \quad (11)$$

Proof : Since $h'(u) = \pi i \sum_n n a_n e^{\pi i n u}$, inequality (10) follows from Lemma 1 :

Observe that

$$\begin{aligned} \sum_n |a_n|^{2/3} &\leqslant (\Sigma |a_n|^2 n^{2+2\gamma})^{1/3} (\Sigma n^{-(1+\gamma)})^{2/3} \\ &\leqslant K_2(\alpha, \gamma) c^2 (\Sigma n^{-(1+\gamma)})^{2/3} \\ &= K_3(\alpha, \gamma) c^2. \end{aligned}$$

Lemma 3 : *Let* $h(\theta, x) = \sum_n a_n(x) e^{\pi i n\theta}$ *and suppose there exists* $\alpha > 0$ *such that*

$$|h_\theta^{(1)}(\theta, x) - h_\theta^{(1)}(\phi, x)| \leqslant c(x)|\theta - \phi|^\alpha$$

for all θ, ϕ *in* $[-1, 1]$ *where* $f_\theta^{(1)}$ *denotes the partial derivative of* f *with respect to* θ. *Let* $\{X(t), t \in [0, T]\}$ *be a stochastic process such that*

$$E[h(0, X(t))]^2 < \infty$$

for every $t \in [0, T]$. *Then, for any* $\gamma < \alpha$, *there exists a positive constant* $K_4(\alpha, \gamma)$ *such that*

$$\sum_n \left\{\frac{1}{T} \int_0^T E[a_n^2(X(t))]\, dt\right\}^{1/3} \leqslant K_4(\alpha, \gamma) \left\{\frac{1}{T} \int_0^T E(c^2(X(t))\, dt\right\}^{1/3}$$

Proof: By Lemma 2, it follows that

$$\sum_n |a_n(X(t))|^2 n^{2+2\gamma} \leqslant K_2(\alpha, \gamma) c^2(X(t)) \quad \text{a.s.}$$

for every $t \in [0, T]$. Hence

$$\sum_n E[a_n^2(X(t))] n^{2+2\gamma} \leqslant K_2(\alpha, \gamma) E[c^2(X(t))]$$

for all $t \in [0, T]$. Let

$$\mu_n = \frac{1}{T} \int_0^T E[a_n^2(X(t))] dt.$$

The inequality proved above gives the relation

$$\sum_n \mu_n n^{2+2\gamma} \leqslant K_2(\alpha, \gamma) \frac{1}{T} \int_0^T E[c^2(X(t))] dt$$

and hence

$$\sum_n \mu_n^{1/3} \leqslant (\Sigma \mu_n n^{2+2\gamma})^{1/3} (\Sigma n^{-(1+\gamma)})^{2/3}$$

$$\leqslant K_2^{1/3}(\alpha, \gamma) (\Sigma n^{-(1+\gamma)})^{2/3} \left\{\frac{1}{T} \int_0^T E[c^2(X(t))] dt\right\}^{1/3}$$

$$\leqslant K_4(\alpha, \gamma) \left\{\frac{1}{T} \int_0^T E[c^2(X(t))] dt\right\}^{1/3}.$$

Remark : Analgous argument proves that

$$\Sigma\ \mu_n^{1/2} \leqslant (\Sigma\mu_n n^{2+2\gamma})^{1/2}(\Sigma n^{-2(1+\gamma)})^{1/2}$$
$$< \infty.$$

Acknowledgement. One of the authors (B. L. S. Prakasa Rao) thanks the Departments of Statistics and Mathematics of Purdue University for invitng him to spend the Summer 1979 which made the collaboration possible.

REFERENCES

BASAWA, I. V., PRAKASA RAO, B. L. S. (1980): *Statistical Inference for Stochastic Processes*, Academic Press, London.

DOROGOVCEV, A. JA. (1976): The consistency of an estimate of a parameter of a stochastic differential equation, *Theory of Probability and Math. Statist.* **10**, 73-82.

GIKHMAN, I. I. and SKOROKHOD A. V. (1972): *Stochastic Differential Equations*. Springer-Verlag, Berlin.

LANSKA, V. (1979): Minimum contrast estimation in diffusion processes, *J. Appl. Prob.* **16**, **16**, 65-75.

McKEAN, H. P. (1969): *Stochastic Integrals*, Academic Press, New York.

PRAKASA RAO, B. L. S. (1979): Asymptotic theory for non-linear least squares estimators for diffusion processes, (Preprint), Indian Statistical Institute, New Delhi.

——— (1981a): Maximum probability estimation for diffusion processes. In *Statistics and Probability : Essays in Honor of C. R. Rao*. (G. Kalianpur et al ed.) North Holland, Amsterdam. (In press)

——— (1981b): The Bernstein-von Misses theorem for a class of diffusion processes, *Theory of Random Processes 9* (To appear) (in Russian).

SZASZ, O. (1922): Uber den Kongvergenzexponent der FourierssJchen Reihen, *Munchener Sitzungskerichte*, 135-150.

Paper received : *August, 1980.*