J. Appl. Prob. **21**, 511–520 (1984)
Printed in Israel

ESTIMATION FOR DIFFUSION PROCESSES UNDER MISSPECIFIED MODELS

IAN W. McKEAGUE,* *The Florida State University*

Abstract

The asymptotic behavior of the maximum likelihood estimator of a parameter in the drift term of a stationary ergodic diffusion process is studied under conditions in which the true drift function and true noise function do not coincide with those specified by the parametric model.

STOCHASTIC DIFFERENTIAL EQUATIONS; MAXIMUM LIKELIHOOD ESTIMATION; ROBUSTNESS; ASYMPTOTIC NORMALITY

1. Introduction

Consider the problem of estimating the drift function $b(x)$ of a stationary diffusion process (X_t) given by

$$dX_t = b(X_t)dt + \sigma(X_t)dW_t, \qquad t \geqq 0,$$

where the process is observed over $[0, T]$. The method of maximum likelihood can be used if $b(x)$ is assumed to have a parametric form $f(x, \theta)$, $\theta \in \Theta$. Brown and Hewitt (1975), Kutoyants (1977), Lanska (1979) and Prakasa Rao and Rubin (1981) have shown that the maximum likelihood estimator of θ is consistent and asymptotically normal. Non-parametric methods of estimating b have been developed by Banon (1978) and Geman (1980).

Suppose that a parametric model for the process (X_t) is given by

$$dX_t = f(X_t, \theta)dt + \gamma(X_t)dW_t, \qquad t \geqq 0.$$

This paper studies the asymptotic behavior of the maximum likelihood estimator of θ under departures of the true drift function $b(x)$ or true noise function $\sigma(x)$ from those specified by the parametric model.

The need for such analysis stems from the desirability of using estimators that are robust under small departures from the underlying model. This kind of

Received 19 April 1983; revision received 16 August 1983.
* Postal address: Department of Statistics and Statistical Consulting Center, The Florida State University, Tallahassee, FL 32306, U.S.A.

analysis is familiar in other settings, for example Huber (1967), White (1981) and Berger and Langberg (1981).

2. Maximum likelihood estimation under misspecified models

Let $(X_t,\ t \geqq 0)$ be a stationary, ergodic process which is assumed to be the unique solution of the stochastic differential equation

$$dX_t = b(X_t)dt + \sigma(X_t)dW_t, \qquad t \geqq 0 \tag{2.1}$$

where X_0 is distributed according to the stationary distribution of the process, b and σ are unknown measurable functions and $(W_t, t \geqq 0)$ is a standard Wiener process. Assume that (X_t) has inaccessible boundaries on the state space $(-\infty, \infty)$. Using the notation of Mandl (1968), let

$$B(x) = 2\int_0^x \frac{b(y)}{\sigma^2(y)}\,dy, \quad p(x) = \int_0^x \exp(-B(y))dy,$$

$$m(x) = 2\int_0^x \frac{\exp B(y)}{\sigma^2(y)}\,dy,$$

where the integrals are assumed to exist. Provided that $m(+\infty) < \infty$ and $m(-\infty) > -\infty$, the stationary distribution, denoted ν, has distribution function $M^{-1}m(x)$ where $M = m(+\infty) - m(-\infty)$.

Suppose a parametric model is used to estimate the drift function $b(x)$ by the method of maximum likelihood. Let Θ denote a closed bounded interval. A family of measurable drift functions $\{f(x,\theta), \theta \in \Theta\}$ and a measurable noise function $\gamma(x) > 0$ are provided and inference is based on the model

$$dX_t = f(X_t, \theta)dt + \gamma(X_t)dW_t, \qquad t \geqq 0. \tag{2.2}$$

The process (X_t) is observed over $[0, T]$. Let μ_θ^T and μ^T denote the measures induced on $C[0, T]$ by a process satisfying (2.2) and the process

$$dY_t = \gamma(Y_t)dW_t, \qquad t \geqq 0$$
$$Y_0 = X_0,$$

respectively. Under conditions given in Liptser and Shiryayev (1977), Theorem 7.19, it follows that $\mu_\theta^T \ll \mu^T$ for all $\theta \in \Theta$ and the log likelihood function $l_T(\theta) = \log[d\mu_\theta^T/d\mu^T](X)$ is given by

$$l_T(\theta) = \int_0^T \frac{f(X_t,\theta)}{\gamma^2(X_t)}\,dX_t - \tfrac{1}{2}\int_0^T \left[\frac{f(X_t,\theta)}{\gamma(X_t)}\right]^2 dt, \quad \text{a.e. } (\mu^T). \tag{2.3}$$

A maximum likelihood estimator calculated from $l_T(\theta)$ is denoted $\hat{\theta}_T$.

Assume that

$$E\left[\frac{b(X_0) - f(X_0,\theta)}{\gamma(X_0)}\right]^2 < \infty$$

for all $\theta \in \Theta$ and, as a function of θ, has a unique minimum at $\theta^* \in \Theta$. The following results describe the asymptotic behavior of $\hat{\theta}_T$ when the observed process satisfies (2.1). The conditions are stated later. g', g'' denote first and second partial derivatives of a function $g(x, \theta)$ with respect to θ.

Theorem 2.1. Under Conditions (C1)–(C3), $\hat{\theta}_T \to \theta^*$ a.s. as $T \to \infty$.

Theorem 2.2. Under Conditions (C1)–(C7), $T^{1/2}(\hat{\theta}_T - \theta^*) \xrightarrow{\mathscr{D}} \mathrm{N}(0, \Sigma)$ where

$$\Sigma = \frac{2M \int_{-\infty}^{\infty} g'(y, \theta^*) \int_{y}^{0} \int_{-\infty}^{s} g'(z, \theta^*) dm(z) dp(s) dm(y)}{\left\{ \int_{-\infty}^{\infty} g''(x, \theta^*) dm(x) \right\}^2}, \tag{2.4}$$

$$g(x, \theta) = \left[\frac{f(x, \theta)}{\gamma(x)} \right]^2 + \sigma^2(x) \frac{\partial}{\partial x} \left(\frac{f(x, \theta)}{\gamma^2(x)} \right). \tag{2.5}$$

Theorem 2.3. In the special case that the drift function has been correctly specified, i.e. $b(x) = f(x, \theta_0)$ for some $\theta_0 \in \Theta$, then (whether or not the noise function has been correctly specified)

(i) $\hat{\theta}_T \to \theta_0$ a.s. as $T \to \infty$ under Conditions (C1)–(C3);

(ii) $T^{1/2}(\hat{\theta}_T - \theta_0) \xrightarrow{\mathscr{D}} \mathrm{N}(0, V)$ under Conditions (C1)–(C9), where

$$V = \frac{\int_{-\infty}^{\infty} \left[\frac{f'(y, \theta_0)}{\gamma(y)} \right]^2 \left[\frac{\sigma(y)}{\gamma(y)} \right]^2 d\nu(y)}{\left\{ \int_{-\infty}^{\infty} \left[\frac{f'(x, \theta_0)}{\gamma(x)} \right]^2 d\nu(x) \right\}^2}. \tag{2.6}$$

Remark. The minimum contrast estimator introduced by Lanska (1979) can be inconsistent under the conditions of Theorem 2.3, whereas the maximum likelihood estimator remains consistent in this case.

Conditions.

(C1) $|f(x, \theta_1) - f(x, \theta_2)| \leqq J(x)\psi(\theta_1 - \theta_2)$, $x \in \mathbb{R}$, $\theta_1, \theta_2 \in \Theta$, where $E[J(X_0)/\gamma(X_0)]^2 < \infty$ and $\lim_{\alpha \to 0} \psi(\alpha) = 0$.

(C2) $E[\sigma(X_0)/\gamma(X_0)]^2 < \infty$ and $|f(x, \theta)| \leqq K(x)$, $x \in \mathbb{R}$, $\theta \in \Theta$, where $E[K(X_0)/\gamma(X_0)]^2 < \infty$.

(C3) $f(x, \theta)$ is continuous in (x, θ) and differentiable with respect to θ. There exists $\alpha > 0$ such that $|f'(x, \theta_1) - f'(x, \theta_2)| \leqq c(x)|\theta_1 - \theta_2|^\alpha$, $x \in \mathbb{R}$, $\theta_1, \theta_2 \in \Theta$, where $E[\sigma(X_0)c(X_0)/\gamma^2(X_0)]^2 < \infty$.

(C4) $f(x, \theta)\gamma^{-2}(x)$ has a continuous first partial derivative with respect to x for each $\theta \in \Theta$.

(C5) $\lim_{x \to \pm\infty} f(x, \theta)\gamma^{-2}(x) \exp B(x) = 0$, $\forall \theta \in \Theta$.

(C6) The partial derivatives g', g'', G', G'' exist and are continuous in (x, θ), where $G(x, \theta)$ is defined by

$$G(x,\theta)=\int_0^x \frac{f(y,\theta)}{\gamma^2(y)}\,dy. \tag{2.7}$$

(C7) $|g''(x,\theta_1)-g''(x,\theta_2)|\leqq U(x)\phi(\theta_1-\theta_2)$, $x\in\mathbb{R}$, $\theta_1,\theta_2\in\Theta$, where $EU(X_0)<\infty$ and $\lim_{\alpha\to 0}\phi(\alpha)=0$.

(C8) $\lim_{x\to\pm\infty} f'(x,\theta_0)\gamma^{-2}(x)[\exp B(x)]\int_0^x f'(s,\theta_0)\gamma^{-2}(s)ds=0$.

(C9) $\lim_{x\to\pm\infty} f''(x,\theta_0)\gamma^{-2}(x)\exp B(x)=0$.

(C10) Condition (C7) with g' in place of g''.

Proof of Theorem 2.1. From (2.1) and (2.3) $l_T(\theta)$ can be written

$$\begin{aligned} l_T(\theta)=&-\tfrac{1}{2}\int_0^T\left[\frac{f(X_t,\theta)-b(X_t)}{\gamma(X_t)}\right]^2 dt+\tfrac{1}{2}\int_0^T\left[\frac{b(X_t)}{\gamma(X_t)}\right]^2 dt\\ &+\int_0^T\frac{f(X_t,\theta)\sigma(X_t)}{\gamma^2(X_t)}\,dW_t. \end{aligned}\tag{2.8}$$

Denote

$$I_T(\theta)=\int_0^T\left[\frac{f(X_t,\theta)-b(X_t)}{\gamma(X_t)}\right]^2 dt.$$

Using (C1) and (C2) it is possible to show that $\{(1/T)I_T(\cdot),\ T\geqq 0\}$ is equicontinuous and uniformly bounded almost surely as a family of functions of θ. By the Arzela–Ascoli theorem this family is relatively compact (almost surely) in the space of continuous functions on Θ provided with the supremum norm. Therefore, by the ergodic theorem

$$\frac{1}{T}I_T(\theta)\xrightarrow{\text{a.s.}}E\left[\frac{f(X_0,\theta)-b(X_0)}{\gamma(X_0)}\right]^2,$$

uniformly in $\theta\in\Theta$ as $T\to\infty$.

Now consider the second term in (2.6). By Condition (C2) and the ergodic theorem

$$\frac{1}{T}\int_0^T\left[\frac{b(X_t)}{\gamma(X_t)}\right]^2 dt\xrightarrow{\text{a.s.}}E\left[\frac{b(X_0)}{\gamma(X_0)}\right]^2\quad\text{as } T\to\infty.$$

Next, using Lemma 4.3 of Prakasa Rao and Rubin (1981) it follows that, under Conditions (C1)–(C3),

$$\frac{1}{T}\int_0^T\frac{f(X_t,\theta)\sigma(X_t)}{\gamma^2(X_t)}\,dW_t\xrightarrow{\text{a.s.}}0,\quad\text{uniformly in }\theta\in\Theta\text{ as }T\to\infty.$$

Thus

$$\frac{1}{T}l_T(\theta)\xrightarrow{\text{a.s.}}-\tfrac{1}{2}E\left[\frac{f(X_0,\theta)-b(X_0)}{\gamma(X_0)}\right]^2+\tfrac{1}{2}E\left[\frac{b(X_0)}{\gamma(X_0)}\right]^2,\tag{2.9}$$

uniformly in $\theta\in\Theta$ as $T\to\infty$. Since the right-hand side of (2.9) has a unique maximum at $\theta^* \in \Theta$ and $\hat{\theta}_T$ maximizes $(1/T)l_T(\theta)$, it is easily proved that $\hat{\theta}_T \to \theta^*$ a.s. as $T \to \infty$.

Proof of Theorem 2.2. The approach used by Prakasa Rao and Rubin (1981) to find the asymptotic distribution of $\hat{\theta}_T$ in the correctly specified case does not extend to the misspecified case. Rather, the proof of this theorem uses the technique, introduced by Lanska (1979), of expressing $l_T(\theta)$ in terms of Lebesgue integrals.

The function $G(x, \theta)$ defined in (2.7) has a continuous second partial derivative with respect to x for each $\theta \in \Theta$ by Condition (C4). Applying Itô's formula, it follows that

$$G(X_T, \theta) = G(X_0, \theta) + \int_0^T \left[\frac{b(X_t)f(X_t, \theta)}{\gamma^2(X_t)} + \tfrac{1}{2}\sigma^2(X_t)\frac{\partial}{\partial x}\left(\frac{f(X_t, \theta)}{\gamma^2(X_t)}\right)\right] dt + \int_0^T \frac{\sigma(X_t)f(X_t, \theta)}{\gamma^2(X_t)}\, dW_t.$$

Then, using (2.1) and (2.3),

$$\begin{aligned} l_T(\theta) &= \int_0^T \frac{f(X_t, \theta)b(X_t)}{\gamma^2(X_t)}\, dt + \int_0^T \frac{f(X_t, \theta)\sigma(X_t)}{\gamma^2(X_t)}\, dW_t - \tfrac{1}{2}\int_0^T \left[\frac{f(X_t, \theta)}{\gamma(X_t)}\right]^2 dt \\ &= G(X_T, \theta) - G(X_0, \theta) - \tfrac{1}{2}\int_0^T g(X_t, \theta)dt, \end{aligned} \tag{2.10}$$

where g is defined in (2.5). Expand $l'_T(\theta)$ about $\hat{\theta}_T$,

$$l'_T(\theta^*) = l'_T(\hat{\theta}_T) + (\theta^* - \hat{\theta}_T)l''_T(\bar{\theta}_T), \quad \text{where } |\bar{\theta}_T - \theta^*| \leqq |\hat{\theta}_T - \theta^*|.$$

Consider

$$T^{-1/2}l'_T(\theta^*) = T^{-1/2}(G'(X_T, \theta^*) - G'(X_0, \theta^*)) + T^{-1/2}\int_0^T g'(X_t, \theta^*)dt.$$

Using the stationarity of (X_t),

$$T^{-1/2}(G'(X_T, \theta^*) - G'(X_0, \theta^*)) \xrightarrow{\mathrm{P}} 0 \quad \text{as } T \to \infty.$$

Using integration by parts and Condition (C5) it can be shown that

$$Eg(X_0, \theta) = E\left[\frac{f(X_0, \theta) - b(X_0)}{\gamma(X_0)}\right]^2 - E\left[\frac{b(X_0)}{\gamma(X_0)}\right]^2, \tag{2.11}$$

and since the right-hand side of this expression is minimized at $\theta^* \in \Theta$, it follows that $Eg(X_0, \theta)$ is minimized at $\theta^* \in \Theta$ and $Eg'(X_0, \theta^*) = 0$. Then, by Mandl (1968), p. 94,

$$T^{-1/2}\int_0^T g'(X_t, \theta^*) \xrightarrow{\mathcal{D}} \mathrm{N}(0, \Delta),$$

where

$$\Delta = \frac{2}{M} \int_{-\infty}^{\infty} g'(y, \theta^*) \int_y^0 \int_{-\infty}^s g'(z, \theta^*) dm(z) dp(s) dm(y).$$

Thus $T^{-1/2} l_T'(\theta^*) \xrightarrow{\mathscr{D}} \mathrm{N}(0, \Delta)$. By the ergodic theorem and Condition (C6)

$$\frac{1}{T} l_T''(\theta^*) \xrightarrow{\mathrm{P}} Eg''(X_0, \theta^*).$$

Next, using Conditions (C6), (C7) and the fact that $\bar{\theta}_T \to \theta^*$ a.s.

$$\frac{1}{T} (l_T''(\bar{\theta}_T) - l_T''(\theta^*)) \xrightarrow{\mathrm{P}} 0, \quad \text{as } T \to \infty.$$

Thus

$$\frac{1}{T} l_T''(\bar{\theta}_T) \xrightarrow{\mathrm{P}} Eg''(X_0, \theta^*) = \frac{1}{M} \int_{-\infty}^{\infty} g''(x, \theta^*) dm(x),$$

as $T \to \infty$. We conclude that

$$T^{1/2}(\hat{\theta}_T - \theta^*) \xrightarrow{\mathscr{D}} \mathrm{N}(0, \Sigma),$$

where Σ is given in (2.4).

Proof of Theorem 2.3. Suppose that $b(x) = f(x, \theta_0)$, where $\theta_0 \in \Theta$. Then $\theta^* = \theta_0$ and (a) follows directly from Theorem 2.1. The proof of (b) consists in showing that Σ in (2.4) reduces to V given in (2.6). Using integration by parts it can be seen that

$$\int_{-\infty}^s g'(z, \theta_0) dm(z) = 2f'(s, \theta_0) \gamma^{-2}(s) \exp B(s),$$

which gives

$$\int_y^0 \int_{-\infty}^s g'(z, \theta_0) dm(z) dp(s) = 2 \int_y^0 f'(s, \theta_0) \gamma^{-2}(s) ds.$$

Using integration by parts with Condition (C8),

$$\begin{aligned} \int_{-\infty}^{\infty} \sigma^2(y) \frac{\partial}{\partial y} [f'(y, \theta_0) \gamma^{-2}(y)] \int_y^0 f'(s, \theta_0) \gamma^{-2}(s) ds dm(y) \\ = \int_{-\infty}^{\infty} \left[\frac{f'(y, \theta_0)}{\gamma(y)} \right]^2 \left[\frac{\sigma(y)}{\gamma(y)} \right]^2 dm(y) \\ - 2 \int_{-\infty}^{\infty} b(y) f'(y, \theta_0) \gamma^{-2}(y) \int_y^0 f'(s, \theta_0) \gamma^{-2}(s) ds dm(y). \end{aligned}$$

It follows that the numerator of Σ reduces to

$$4M\int_{-\infty}^{\infty}\left[\frac{f'(y,\theta_0)}{\gamma(y)}\right]^2\left[\frac{\sigma(y)}{\gamma(y)}\right]^2 dm(y).$$

Similarly, it can be shown that

$$\int_{-\infty}^{\infty} g''(x,\theta_0)dm(x)=2\int_{-\infty}^{\infty}\left[\frac{f'(x,\theta_0)}{\gamma(x)}\right]^2 dm(x).$$

This completes the proof of the theorem.

Examples.

1. This example presents a misspecified drift function but correctly specified noise function. Suppose that the observed process satisfies

$$dX_t = -X_t dt+\sqrt{2}\,dW_t,$$

where X_0 has an N(0, 1) distribution, the stationary distribution of (X_t). Estimates are calculated from the parametric model

$$dX_t = -\theta X_t^3 dt+\sqrt{2}\,dW_t.$$

The parameter θ^* which minimizes $E(X_0-\theta X_0^3)^2$ is given by

$$\theta^*=\frac{EX_0^4}{EX_0^6}=\frac{1}{5}\,.$$

By Theorem 2.1, $\hat{\theta}_T\to\frac{1}{5}$ a.s. as $T\to\infty$. We also have $g(x,\theta)=\frac{1}{2}\theta^2x^6-3\theta x^2$, so that $g'(x,\theta^*)=\frac{1}{5}x^6-3x^2$. Using repeated integration by parts

$$\int_{-\infty}^{s} g'(z,\theta^*)dm(z)=\int_{-\infty}^{s}(\tfrac{1}{5}z^6-3z^2)\exp(-z^2/2)dz = -(\tfrac{1}{5}s^5+s^3)\exp(-s^2/2)$$

so that

$$\int_y^0\int_{-\infty}^{s} g'(z,\theta^*)dm(z)dp(s)=\tfrac{1}{30}y^6+\tfrac{1}{4}y^4$$

and

$$\begin{aligned}&\int_{-\infty}^{\infty} g'(y,\theta^*)\int_y^0\int_{-\infty}^{s} g'(z,\theta^*)dm(z)dp(s)d\nu(y)\\ &=\int_{-\infty}^{\infty}(\tfrac{1}{5}y^6-3y^2)(\tfrac{1}{30}y^6+\tfrac{1}{4}y^4)d\nu(y)\\ &=\tfrac{1}{150}EX_0^{12}+\tfrac{1}{20}EX_0^{10}-\tfrac{1}{10}EX_0^8-\tfrac{3}{4}EX_0^6\\ &=94.8.\end{aligned}$$

Also

$$\int_{-\infty}^{\infty} g''(x, \theta^*) d\nu(x) = EX_0^6 = 15.$$

Thus $\Sigma = (2)(94.8)/(15)^2 = 0.84$, and by Theorem 2.2 we have $T^{1/2}(\hat{\theta}_T - \frac{1}{5}) \xrightarrow{\mathscr{D}} \mathrm{N}(0, 0.84)$. The asymptotic variance of $\hat{\theta}_T$ is less than in the correctly specified case for which $T^{1/2}(\hat{\theta}_T - 1) \xrightarrow{\mathscr{D}} \mathrm{N}(0,1)$.

2. Our second example has a correctly specified drift function and a misspecified noise function. The observed process is the same as in the first example but the parametric model is given by

$$dX_t = -\theta X_t dt + \left(\frac{2}{1 + X_t^2}\right)^{1/2} dW_t.$$

Theorem 2.3 yields $T^{1/2}(\hat{\theta}_T - 1) \xrightarrow{\mathscr{D}} \mathrm{N}(0, 2.75)$. The asymptotic variance of $\hat{\theta}_T$ has almost tripled due to the misspecified noise.

3. Discriminating between separate families of drift functions

Let (X_t) satisfy (2.1) and assume throughout this section $\sigma(x) \equiv 1$. Suppose that two parametric models for this process have been suggested. It is required to decide in favor of the model which best fits the observed trajectory $\{X_t, 0 \leqq t \leqq T\}$.

Let $\{f_1(x, \theta) : \theta \in \Theta\}$, $\{f_2(x, \phi) : \phi \in \Phi\}$ be distinct families of drift functions, where Θ, Φ are closed bounded intervals. A reasonable way to compare the goodness of fit of these families to the true drift function $b(x)$ is to estimate the parameter

$$\Delta = E[f_2(X_0, \phi^*) - b(X_0)]^2 - E[f_1(X_0, \theta^*) - b(X_0)]^2.$$

In this section we introduce an estimator for Δ. The noise function is assumed to be correctly specified; that is $\gamma(x) \equiv 1$.

Let $l_T^{(1)}(\theta)$, $l_T^{(2)}(\phi)$ denote the log likelihoods for the two models. Define

$$\hat{\Delta}_T = \frac{2}{T} [l_T^{(1)}(\hat{\theta}_T) - l_T^{(2)}(\hat{\phi}_T)].$$

Given that f_1, f_2 satisfy Conditions (C1)–(C3), it follows from (2.7) in the proof of Theorem 2.1 that $\hat{\Delta}_T \to \Delta$ a.s. as $T \to \infty$. The following result shows that $\hat{\Delta}_T$ is asymptotically normal. Conditions from Section 2 are used interchangeably between the two families of drift functions indexed by Θ and Φ.

Theorem 3.1. Suppose that f_1, f_2 satisfy Conditions (C1)–(C6) and (C10), where g_1, g_2 are given by (2.5) with $f = f_1, f_2$ respectively, $\gamma(x) \equiv 1$. Then $T^{1/2}(\hat{\Delta}_T - \Delta) \xrightarrow{\mathscr{D}} \mathrm{N}(0, \Sigma_2)$, where

$$\Sigma_2 = \frac{2}{M}\int_{-\infty}^{\infty} [g_2(y,\phi^*) - g_1(y,\theta^*) - \Delta]$$
$$\times \int_y^0 \int_{-\infty}^{s} [g_2(z,\phi^*) - g_1(z,\theta^*) - \Delta] dm(z) dp(s) dm(y).$$

Proof. From (2.9) it is clear that $T^{1/2}(\hat{\Delta}_T - \Delta)$ has the same limiting distribution as

$$\begin{aligned} T^{-1/2}\int_0^T [g_2(X_t,\hat{\phi}_T) - g_1(X_t,\hat{\theta}_T) - \Delta]dt &= T^{-1/2}\int_0^T [g_2(X_t,\phi^*) - g_1(X_t,\theta^*) - \Delta]dt \\ &\quad + T^{-1/2}\int_0^T [g_2(X_t,\hat{\phi}_T) - g_2(X_t,\phi^*)]dt \\ &\quad + T^{-1/2}\int_0^T [g_1(X_t,\theta^*) - g_1(X_t,\hat{\theta}_T)]dt \\ &= A_T + B_T + C_T. \end{aligned}$$

From (2.11),

$$E[g_2(X_0,\phi^*) - g_1(X_0,\theta^*) - \Delta] = 0$$

so, by Mandl (1968), p. 94, $A_T \xrightarrow{\mathscr{D}} N(0,\Sigma_2)$. Next, consider C_T. Expanding g_1 in a neighborhood around θ^*,

$$C_T = T^{1/2}(\hat{\theta}_T - \theta^*)\cdot\frac{1}{T}\int_0^T g_1'(X_t,\bar{\theta}_T)dt, \quad \text{where } |\bar{\theta}_T - \theta^*| \leqq |\hat{\theta}_T - \theta^*|.$$

But,

$$\frac{1}{T}\int_0^T g_1'(X_t,\bar{\theta}_T)dt = \frac{1}{T}\int_0^T g_1'(X_t,\theta^*)dt + \frac{1}{T}\int_0^T [g_1'(X_t,\bar{\theta}_T) - g_1'(X_t,\theta^*)]dt. \tag{3.1}$$

By the proof of Theorem 2.2, $Eg_1'(X_0,\theta^*) = 0$, so by the ergodic theorem the first term in (3.1) converges to 0 a.s. as $T \to \infty$. From Condition (C10), the second term in (3.1) is bounded above by

$$\phi(\bar{\theta}_T - \theta^*)\cdot\frac{1}{T}\int_0^T U(X_t)dt, \tag{3.2}$$

which converges to 0 a.s. since $|\bar{\theta}_T - \theta^*| \leqq |\hat{\theta}_T - \theta^*|$, $\hat{\theta}_T \to \theta^*$ a.s. and

$$\frac{1}{T}\int_0^T U(X_t)dt \xrightarrow{\text{a.s.}} EU(X_0) \quad \text{as } T \to \infty.$$

Thus $C_T \to 0$ a.s. and similarly $B_T \to 0$ a.s. This completes the proof of the theorem.

Example. Consider the two models

$$dX_t = -\theta X_t dt + dW_t, \tag{3.3}$$

$$dX_t = -\phi X_t^3 dt + dW_t, \tag{3.4}$$

and suppose that the observed process satisfies (3.3) with $\theta = \theta_0 > 0$. Some involved but routine calculations give that $\Delta = 0.2\theta_0$ and $\Sigma_2 = 0.64\theta_0 + 9.36\theta_0^{-3}$. Note that $\Sigma_2 \to \infty$ as $\theta_0 \to 0$. The poor performance of $\hat{\Delta}_T$ for small θ_0 is to be expected since, as $\theta_0 \to 0$, the drift function has less effect on the dynamics of the process so it is harder to discriminate between the two models.

Acknowledgement

The author thanks Loren Cobb for some helpful conversations.

References

BANON, G. (1978) Nonparametric identification for diffusion processes. *SIAM J. Control and Optimization* **16**, 380–395.

BERGER, R. L. AND LANGBERG, N. A. (1981) Linear least squares estimates and nonlinear means. Florida State University Statistics Report M-573.

BROWN, B. M. AND HEWITT, J. I. (1975) Asymptotic likelihood theory for diffusion processes. *J. Appl. Prob.* **12**, 228–238.

GEMAN, S. (1980) An application of the method of sieves: functional estimator for the drift of a diffusion. *Colloq. Math. Soc. János Bolyai* **32**. Nonparametric Statistical Inference, Budapest, Hungary.

HUBER, P. J. (1967) The behavior of maximum likelihood estimates under nonstandard conditions. *Proc. 5th Berkeley Symp. Math. Statist. Prob.* **1**, 221–233.

KUTOYANTS, YU. A. (1977) Estimation of the drift coefficient parameter of a diffusion process in the smooth case. *Theory Prob. Appl.* **22**, 399–406.

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

LIPTSER, R. S. AND SHIRYAYEV, A. N. (1977) *Statistics of Random Processes I.* Springer-Verlag, New York.

MANDL, P. (1968) *Analytical Treatment of One-Dimensional Markov Processes.* Academia, Prague; Springer-Verlag, Berlin.

PRAKASA RAO, B. L. S. AND RUBIN, H. (1981) Asymptotic theory of estimation in non-linear stochastic differential equations. *Sankhyā A* **43**, 170–189.

WHITE, H. (1981) Consequences and detection of misspecified nonlinear regression models. *J. Amer. Statist. Assoc.* **76**, 419–433.