Ann. Inst. Statist. Math.
Vol. 42, No. 2, 221–251 (1990)

ASYMPTOTIC BEHAVIOR OF *M*-ESTIMATOR AND RELATED RANDOM FIELD FOR DIFFUSION PROCESS

NAKAHIRO YOSHIDA*

Department of Applied Mathematics, Faculty of Engineering Science, Osaka University, Toyonaka, Osaka 560, Japan

(Received September 6, 1988; revised June 30, 1989)

Abstract. The M-estimate which maximizes a positive stochastic process Q is treated for multidimensional diffusion models. The convergence in distribution of the process of ratio of Q's after normalizing is proved. The asymptotic behavior of M-estimates is stated. We present the asymptotic variance in general cases and in estimation by misspecified models.

Key words and phrases: Diffusion process, M-estimator.

1. Introduction

Considering a parametric model of diffusion processes, we estimate the parameter θ from a realization X. It is well-known that the maximum likelihood estimation with respect to the likelihood function is one of the best methods for estimation (see Kutoyants (1984)). However, we can hardly get clean data generated from the model in the strict sense, since it is naturally practical to regard them as contaminated by some noises and misspecification of the true model. One of the purposes of the present paper is to investigate the asymptotic behavior of the maximum likelihood estimators in the case where the true model does not necessarily belong to the observer's model. Lanska (1979) uses the M-estimation for one-dimensional diffusion models and Yoshida (1987) shows the existence of the optimal M-estimation using an influence function scheme. Then it is important to investigate asymptotic properties of robustified M-estimators. Thus, it is also in our scope. Bustos (1982) treats AR-models in such a situation.

We treat multidimensional diffusion processes. Since pathwise representation of estimating equations with Lebesgue integrals is difficult, it is necessary to define the estimating equation with "martingale" parts as

*Now at The Institute of Statistical Mathematics, 4-6-7 Minami-Azabu, Minato-ku, Tokyo 106, Japan.

well as bounded variation parts: our M-estimate $\hat{\theta}_t$ satisfies

$$Q(t, X, \hat{\theta}_t) = \max_{\theta \in \bar{\Theta}} Q(t, X, \theta) ,$$

where the positive stochastic process $Q(t, X, \theta)$ possesses a semimartingale decomposition defined in Section 2. The minimum contrast method is reduced to the M-estimation.

The ratio of likelihood functions, which is the Radon-Nikodym derivative $dP_{\theta 1}/dP_{\theta 2}$, plays an important role in investigating the asymptotic behavior of the maximum likelihood estimator. In the ergodic case, the likelihood ratio normalized by an appropriate function ϕ_T has locally asymptotical normality: for $u \in R^k$,

$$dP_{\theta+\phi_T u}/dP_\theta = \exp\left\{ (u, \tilde{\Delta}_T(\theta)) - \frac{1}{2}(u, I(\theta)u) + \tilde{\varepsilon}_T(u, \theta) \right\},$$

where $\mathscr{L}\{\tilde{\Delta}_T(\theta) | P_\theta\} \to N_k(0, I(\theta))$, $I(\theta)$ is a positive definite matrix and $\tilde{\varepsilon}_T(u, \theta) \to 0$ in P_θ. Then one can show the asymptotic properties of the maximum likelihood estimator not only for i.i.d. models but also for dependent ones, LeCam (1960), Ibragimov and Has'minskii (1972, 1973, 1981), Inagaki and Ogata (1975), Ogata and Inagaki (1977) and Kutoyants (1977, 1978, 1984). In our case it is also useful to consider the ratio of Q's about a particular $\theta_0 \in \Theta$ defined by

$$Z_T(u) = Q(T, X, \theta_0 + \phi_T u)/Q(T, X, \theta_0) .$$

Even if the true model does not belong to some parametric model, the expansion holds:

$$Z_T(u) = \exp\left\{ (u, \Delta_T) - \frac{1}{2}(u, \Gamma_T u) + \rho_T(u) \right\},$$

$$\Phi_T \to \Phi, \quad \Gamma_T \to \Gamma$$

in probability

$$\mathscr{L}\{\Delta_T | P\} \to \mathscr{L}\{\Phi^{1/2} N\}, \quad N \sim N_k(0, I) \quad \text{independent of } (\Phi, \Gamma) ,$$

$$\rho_T(u) \to 0$$

in probability, where Γ and Φ are $k \times k$ positive definite random matrices.

Showing inequalities for the probability of large deviations, one can prove that the random fields $\{Z_T(\cdot)\}$ converge in law to some field $Z(\cdot)$, which enables us to know very naturally the asymptotic properties of maximum likelihood estimators, likelihood ratio test statistics and so on (see Inagaki and Ogata (1975)). But it is sometimes difficult to verify them for stochastic processes involving the calculation of Laplace transforms of functionals on sample spaces. Here, we do not assume the large deviation inequalities as in Kutoyants (1984). Instead weaker conditions are assumed, of course ensuring convergence of estimators in the weaker topology than his. However, the weak convergence of distributions of random fields $\{Z_T(\cdot)\}$ still holds under our assumptions.

In the following two sections we prepare notations and assumptions. Section 4 gives asymptotic properties of M-estimators and convergence of related random fields. As an application, Section 5 presents the asymptotic behavior of maximum likelihood estimators based on misspecified parametric models. The last section is devoted to several examples illustrating our results. Our argument is related to "non-ergodic" statistical inference. For this notion the reader can consult Jeganathan (1982*a*, 1982*b*), Basawa and Scott (1983) and references in it.

2. M-estimation and notations

We discuss the problem of parameter estimation and the estimate of the parameter θ is based on the functional on $C([0, T] \to R^d)$

$$(2.1) \qquad Q(T, X, \theta) = \exp\left\{\int_0^T S_i(X_t, \theta)dX_t^i - \frac{1}{2}\int_0^T R(X_t, \theta)dt\right\},$$

where S_i, R are given functions defined on $R^d \times \bar{\Theta}$, and Θ is a bounded convex domain in R^k. We use the M-estimate $\hat{\theta}$ at which the functional $Q(T, X, \theta)$ attains its maximum in $\bar{\Theta}$.

Let the true process X satisfy the stochastic differential equation

$$(2.2) \qquad \begin{aligned} dX_t^i &= \sum_{\alpha=1}^r V_\alpha^i(X_t)dW_t^\alpha + V_0^i(X_t)dt, \qquad 1 \le i \le d, \\ X_0 &= \eta, \end{aligned}$$

where W^α are independent standard Wiener processes. Even when we consider a parametric model of diffusion processes, in general it is not assumed that it contains the true one. To investigate the asymptotic behavior of the M-estimator when the observed process X is generated by (2.2), which does not necessarily belong to a certain parametric model, it is useful to study that of the ratio of Q's after normalizing about a certain point θ_0:

(2.3) $$Z_T(u) = Z_T(u, \theta_0) = \exp\left\{\int_0^T \Delta S_i dX_t^i - \frac{1}{2}\int_0^T \Delta R dt\right\},$$

where $u \in R^k$, $\Delta S_i = S_i(X, \theta_0 + b^{-1/2}u) - S_i(X, \theta_0)$, $1 \le i \le d$, $\Delta R = R(X, \theta_0 + b^{-1/2}u) - R(X, \theta_0)$, and $b = b(T)$ is a positive divergent function of T. In Section 4 we will study this.

Let us prepare the following notations. The summation sign is abbreviated for repeated indices unless otherwise stated.

(0) ω: an element of the probability space under consideration.

(1) $\partial_i = \partial/\partial x^i$, $\delta_i = \partial/\partial\theta^i$; $\partial = (\partial_1, \ldots, \partial_d)$, $\delta = (\delta_1, \ldots, \delta_k)$.

(2) $V_\alpha = V_\alpha^i \partial_i$, $v^{ij}(x) = V_\alpha^i(x) V_\alpha^j(x)$.

(3) $L^i = v^{ij}\partial_j/2 + V_0^i$.

(4) For a function $f(x, \theta)$ and $u \in R^k$,

$$\Delta f(x, \theta_1, \theta_2) = \Delta f(\theta_1, \theta_2) = f(x, \theta_2) - f(x, \theta_1),$$

$$D^{(1)}f(\theta, u) = f(x, \theta + b^{-1/2}u) - f(x, \theta) - b^{-1/2}u^m \delta_m f(x, \theta),$$

$$D^{(2)}f(\theta, u) = D^{(1)}f(\theta, u) - (2b)^{-1}u^m u^n \delta_m \delta_n f(x, \theta).$$

(5) Generator $L = L^i \partial_i$.

(6) $H = H(\theta) = H(x, \theta) = R(x, \theta)/2 - S_i(x, \theta) V_0^i(x)$.

Remarks 2.1.

(i) Under (2.2),

$$Q(T, X, \theta) = \exp\left\{\int_0^T S_i V_\alpha^i dW_t^\alpha - \int_0^T H dt\right\}.$$

(ii) $L = L^i \partial_i = v^{ij}\partial_i\partial_j/2 + V_0^i \partial_i$.

For simplicity, the argument x and X_t substituted for x of functions are often abbreviated; e.g., $S_i(X_t, \theta)$ is often denoted by S_i or $S_i(\theta)$.

3. Assumptions

This section gives some assumptions used later on. Assume that S_i and R are defined on $R^d \times \bar{\Theta}$, twice differentiable in θ and these derivatives are continuous. Here $\{X_t\}$ satisfies the equation (2.2). In the sequel we assume for the true model (2.2) the existence of $\theta_0 \in \Theta$ satisfying all or some of the following conditions.

CONDITIONS (B).

(BS0) $\sup_{\theta \in \bar{\Theta}} \left|(1/b)\int_0^T [S_i(X_t, \theta) - S_i(X_t, \theta_0)][dX_t^i - V_0^i(X_t)dt]\right| \to 0$ in probability as $T \to \infty$.

(BH1) $C_T := \sup_{\theta \in \bar{\Theta}} \left| (1/b) \int_0^T \delta H(X_t, \theta) dt \right|$ are stochastically bounded, i.e.,

$$\lim_{A \to \infty} \sup_{T \geq 0} P(|C_T| > A) = 0 .$$

CONDITIONS (E).

(EH0) For each $\theta \in \bar{\Theta}$, there exists an r.v. $\tilde{\Gamma}(\theta)$ such that as $T \to \infty$,

$$\frac{1}{b} \int_0^T [H(X_t, \theta) - H(X_t, \theta_0)] dt \to \tilde{\Gamma}(\theta)$$

in probability. Moreover, $\tilde{\Gamma}$ is attains its minimum zero in $\bar{\Theta}$ only at θ_0.

(EH2) There exists a random matrix $\Gamma = (\Gamma_{mn})$ such that as $T \to \infty$,

$$\Gamma_{mnT} := \frac{1}{b} \int_0^T \delta_m \delta_n H(X_t, \theta_0) dt \to \Gamma_{mn} ,$$

in probability and Γ are positive definite a.s. ω.

(ES1) There exists a positive definite random matrix $\Phi = (\Phi_{mn})$ such that, at θ_0 as $T \to \infty$,

$$\Phi_{mnT} := \frac{1}{b} \int_0^T (\delta_m S_i + \partial_i G_m)(\delta_n S_j + \partial_j G_n) v^{ij} dt \to \Phi_{mn}$$

in probability, where the functions G_m are defined in the following condition (SOL). Put

$$\Gamma_T = (\Gamma_{mnT}), \qquad \Phi_T = (\Phi_{mnT}) .$$

CONDITION (SOL). There exist functions $G_m(x)$, $m = 1, \ldots, k$, with continuous differentials $\partial_i \partial_j G_m$ satisfying the partial differential equations

$$LG_m(x) = \delta_m H(x, \theta_0)$$

and

$$b^{-1/2} G_m(X_T) \to 0$$

in probability as $T \to \infty$, $m = 1, \ldots, k$.

CONDITIONS (T).

(TS) For each $\varepsilon > 0$, there exists a random variable $\bar{\delta} = \bar{\delta}(\varepsilon) > 0$ a.s. such that

$$\lim_{T\to\infty} P\left[\sup_{|u|\le \bar{\delta} b^{1/2}} \frac{1}{1+|u|^2} \left| \int_0^T D^{(1)} S_i(X_t, \theta_0, u)[dX_t^i - V_0^i(X_t)dt] \right| > \varepsilon \right] = 0 .$$

(TH) For each $\varepsilon > 0$, there exists a random variable $\bar{\delta} = \bar{\delta}(\varepsilon) > 0$ a.s. such that

$$\lim_{T\to\infty} P\left[\sup_{|u|\le \bar{\delta} b^{1/2}} \frac{1}{1+|u|^2} \left| \int_0^T D^{(2)} H(X_t, \theta_0, u)\, dt \right| > \varepsilon \right] = 0 .$$

Remarks 3.1. (I) Conditions (E) are related to the law of large numbers. If the underlying process X is ergodic, these conditions can be easily verified.

(II) Conditions (B) and (T) are provided for regularity. For sufficient conditions for Conditions (BS0) and (TS) with stochastic integrals with respect to continuous martingales, the following lemmas are useful.

LEMMA 3.1. *Let* $\{f_T(\theta); \theta \in F\}$, $T \ge 0$, *be a family of random fields on* F, *a convex compact in* R^k. *Suppose that there exist constants* p *and* l *such that* $p \ge l > k$, *and for any* θ, θ_1 *and* θ_2,

(1) $E|f_T(\theta_2) - f_T(\theta_1)|^p \le C|\theta_2 - \theta_1|^l$,

(2) $E|f_T(\theta)|^p \le C$,

(3) $f_T(\theta) \to 0$ *in probability*,

where C *is a constant independent of* θ, θ_1, θ_2 *and* T. *Then*,

$$\sup_{\theta \in F} |f_T(\theta)| \to 0$$

in probability.

In fact, (1)-(3) above ensure that $\{f_T(\theta); \theta \in F\}$ converges in distribution to 0 in $C(F)$ (see, e.g., Appendix of Ibragimov and Has'minskii (1981)).

LEMMA 3.2. *If there exists a constant* $p > k$ *such that for all* α, m, n,

(i) $b^{-p} E\left\{\left[\int_0^T \sup_\theta |\delta_m S_i(X_t, \theta) V_\alpha^i(X_t)|^2 dt\right]^{p/2}\right\} \to 0$, *as* $T \to \infty$,

(ii) $\sup_T b^{-p/2} E\left\{\left[\int_0^T \sup_\theta |\delta_m \delta_n S_i(X_t, \theta) V_\alpha^i(X_t)|^2 dt\right]^{p/2}\right\} < \infty$,

then, (BS0) *and* (TS) *hold*.

PROOF. Let K denote generic constants independent of T. Let

$$M_T(\theta)=\int_0^T [S_i(\theta)-S_i(\theta_0)][dX_t^i - V_0^i dt], \qquad \theta \in \bar{\Theta}\,.$$

Then, we have, with C_r-inequalities and Novikov's moment inequalities for stochastic integrals (Novikov (1971)),

$$\begin{aligned} E\left|\frac{1}{b}M_T(\theta_2)-\frac{1}{b}M_T(\theta_1)\right|^p &= b^{-p}E\left|\int_0^T \Delta S_i(\theta_1,\theta_2)[dX_t^i - V_0^i dt]\right|^p \\ &\le Kb^{-p}\sum_\alpha E\left\{\left[\int_0^T |\Delta S_i(\theta_1,\theta_2)V_\alpha^i|^2 dt\right]^{p/2}\right\} \\ &\le Kb^{-p}|\theta_2-\theta_1|^p E\left\{\left[\int_0^T \sup_\theta |\delta S_i(\theta)V_\alpha^i|^2 dt\right]^{p/2}\right\} \\ &\le K|\theta_2-\theta_1|^p \end{aligned}$$

and

$$\begin{aligned} E\left|\frac{1}{b}M_T(\theta)\right|^p = E\left|\frac{1}{b}\int_0^T \Delta S_i(\theta_0,\theta)[dX_t^i - V_0^i dt]\right|^p &\le Kb^{-p}E\left\{\left[\int_0^T \sup_\theta |\delta S_i(\theta)V_\alpha^i|^2 dt\right]^{p/2}\right\} \\ &\to 0\,. \end{aligned}$$

Through Lemma 3.1, (BS0) can be shown.

Next for $\upsilon \in R^k$ putting

$$m_T(\upsilon)=\frac{1}{1+b|\upsilon|^2}\int_0^T D^{(1)}S_i(\theta_0, b^{1/2}\upsilon)[dX_t^i - V_0^i dt]\,,$$

for (TS) it is sufficient to show the convergence in distribution of the random fields $\{m_T(\upsilon);\, |\upsilon|\le\delta\}$ to 0. Let $|\upsilon_1|\le|\upsilon_2|\le\delta$. Then,

$$\begin{aligned} &|(1+b|\upsilon_2|^2)^{-1}D^{(1)}S_i(\theta_0,b^{1/2}\upsilon_2)-(1+b|\upsilon_1|^2)^{-1}D^{(1)}S_i(\theta_0,b^{1/2}\upsilon_1)| \\ &\quad\le |(1+b|\upsilon_2|^2)^{-1} \\ &\qquad\cdot\{S_i(\theta_0+\upsilon_2)-\upsilon_2^m\delta_m S_i(\theta_0)-S_i(\theta_0+\upsilon_1)+\upsilon_1^m\delta_m S_i(\theta_0)\}| \\ &\qquad+|\{(1+b|\upsilon_2|^2)^{-1}-(1+b|\upsilon_1|^2)^{-1}\} \\ &\qquad\cdot\{S_i(\theta_0+\upsilon_1)-S_i(\theta_0)-\upsilon_1^m\delta_m S_i(\theta_0)\}|\,. \end{aligned}$$

The first term of the r.h.s. is not larger than

$$\begin{aligned}(1+b|v_2|^2)^{-1}|v_2-v_1|\,|\delta S_i(\theta_0+\xi)-\delta S_i(\theta_0)|\\ \leq (1+b|v_2|^2)^{-1}|v_2-v_1|\,|\delta\delta S_i(\theta_0+\bar{\xi})|\,|v_2|\\ \leq \frac{1}{2}b^{-1/2}|v_2-v_1|\sup_\theta|\delta\delta S_i(\theta)|\,.\end{aligned}$$

The second term is not larger than

$$\begin{aligned}b\,||v_1|^2-|v_2|^2|(1+b|v_2|^2)^{-1}(1+b|v_1|^2)^{-1}|v_1|^2\sup_\theta|\delta\delta S_i(\theta)|\\ \leq b^{-1/2}|v_2-v_1|\sup_\theta|\delta\delta S_i(\theta)|\end{aligned}$$

since

$$\begin{aligned}&b(|v_1|+|v_2|)(1+b|v_2|^2)^{-1}(1+b|v_1|^2)^{-1}|v_1|^2\\ &\leq b^{-1/2}\frac{b^{1/2}|v_2|}{1+b|v_2|^2}\,\frac{b|v_1|^2}{1+b|v_1|^2}+b^{-1/2}\frac{b|v_2|^2}{1+b|v_2|^2}\,\frac{b^{1/2}|v_1|}{1+b|v_1|^2}\\ &\leq b^{-1/2}\,.\end{aligned}$$

Hence, for $|v_1|, |v_2|\leq\delta$

$$E|m_T(v_2)-m_T(v_1)|^p\leq K|v_2-v_1|^p\,.$$

Moreover,

$$\begin{aligned}E|m_T(v)|^p&\leq KE\left[\int_0^T\left|\frac{v^m v^n}{1+b|v|^2}\,\delta\delta S_i(\theta_0+\tilde{v})\,V_\alpha^i\right|^2dt\right]^{p/2}\\ &\leq Kb^{-p}E\left[\int_0^T\sup_\theta|\delta\delta S_i(\theta)\,V_\alpha^i|^2dt\right]^{p/2}\to 0\,.\end{aligned}$$

Therefore, the weak convergence holds.

LEMMA 3.3. *It is sufficient for Condition* (TH) *that*

$$b^{-1}\int_0^T\sup_\theta|\delta\delta\delta H(X_t,\theta)|dt,\qquad T\geq 0\,,$$

are bounded a.s.

The proof is easy and omitted.

(III) The partial differential equations in Condition (SOL) are elliptic-type equations. The integrability concerning Condition (ES1) puts restrictions on a choice of G_m. If a parametric model in question is correctly specified, Condition (SOL) can be eliminated for the maximum likelihood estimator since $\delta_m H(X, \theta_0) = 0$. A general procedure to get G_m will be seen in an example of Section 6.

(IV) For many "non-ergodic" processes whose paths tend to infinity, Conditions (E) are proven by a version of Doob's martingale convergence theorem for diffusion processes. Keller *et al.* (1984) proved this theorem using convexity properties of coefficients of stochastic differential equations. To remove such convexity assumptions and to extend it to semimartingales are possible (see Appendix). This enables us to prove the asymptotical mixed normality of likelihood ratios of semimartingales within a general framework.

4. Asymptotic behavior of M-estimators

First, we show an in probability convergence theorem for M-estimators.

THEOREM 4.1. *Suppose that* (BS0), (BH1) *and* (EH0) *are satisfied. Then, as* $T \to \infty$,

$$\hat{\theta}_T \to \theta_0$$

in probability.

PROOF. Let

$$Y_T(\theta) := \frac{1}{b} \log Q(T, X, \theta) - \frac{1}{b} \log Q(T, X, \theta_0) \,,$$

$$M_T(\theta) := \int_0^T \Delta S_i(\theta_0, \theta)[dX_t^i - V_0^i dt] \quad \text{and}$$

$$N_T(\theta) := \int_0^T \Delta H_t(\theta_0, \theta) dt \,.$$

Then,

$$Y_T(\theta) = \frac{1}{b} M_T(\theta) - \frac{1}{b} N_T(\theta) \,.$$

For arbitrary $\theta_1, \theta_2 \in \bar{\Theta}$,

$$\left| \frac{1}{b} N_T(\theta_2) - \frac{1}{b} N_T(\theta_1) \right| \leq \left| \frac{1}{b} \int_0^T \delta H_t(\theta_3) dt \right| \cdot |\theta_2 - \theta_1| \tag{4.1}$$
$$\leq C_T(\omega) |\theta_2 - \theta_1| \,,$$

where $\{C_T(\omega)\}$ is a stochastically bounded sequence of nonnegative r.v.'s, independent of θ_1, θ_2 by Condition (BH1). The inequality (4.1) and that $N_T(\theta_0) = 0$ ensure that the family of distributions of $\{N_T(\cdot)/b\}$ on $C(\bar{\Theta})$ with sup-norm is tight. Consequently, the limit $\tilde{\Gamma}(\theta)$ are continuous functions on $\bar{\Theta}$ with probability 1.

Let ε and η be any positive numbers. For any positive δ, one can take a finite subset $\{\theta_n\}$ in Θ such that balls $\{\theta \in \bar{\Theta}; |\theta - \theta_n| \leq \delta\}$ cover $\bar{\Theta}$. For each $\theta \in \bar{\Theta}$, choose $\theta_{n(\theta)}$, which is one of the closest points to θ in $\{\theta_n\}$. When δ is small enough, we have

$$P\left(|C_T| \sup_{\theta \in \bar{\Theta}} |\theta - \theta_{n(\theta)}| > \eta/3 \right) \leq \varepsilon/3$$

and

$$P\left(\sup_{\theta \in \bar{\Theta}} |\tilde{\Gamma}(\theta_{n(\theta)}) - \tilde{\Gamma}(\theta)| > \eta/3 \right) \leq \varepsilon/3$$

by continuity. Since

$$P\left(\sum_n \left| \frac{1}{b} N_T(\theta_n) - \tilde{\Gamma}(\theta_n) \right| > \eta/3 \right) \leq \varepsilon/3, \quad \text{for large } T,$$

$$\begin{aligned} &P\left(\sup_{\theta \in \bar{\Theta}} \left| \frac{1}{b} N_T(\theta) - \tilde{\Gamma}(\theta) \right| > \eta \right) \\ &\leq P\left(\sup_{\theta \in \bar{\Theta}} \left| \frac{1}{b} N_T(\theta) - \frac{1}{b} N_T(\theta_{n(\theta)}) \right| > \eta/3 \right) \\ &\quad + P\left(\sup_{\theta \in \bar{\Theta}} \left| \frac{1}{b} N_T(\theta_{n(\theta)}) - \tilde{\Gamma}(\theta_{n(\theta)}) \right| > \eta/3 \right) \\ &\quad + P\left(\sup_{\theta \in \bar{\Theta}} |\tilde{\Gamma}(\theta_{n(\theta)}) - \tilde{\Gamma}(\theta)| > \eta/3 \right) \\ &\leq \varepsilon \,. \end{aligned} \tag{4.2}$$

Therefore,

$$\sup_{\theta \in \bar{\Theta}} \left| \frac{1}{b} N_T(\theta) - \tilde{\Gamma}(\theta) \right| \to 0 \tag{4.3}$$

in probability.

It is sufficient for the proof to show that for any neighborhood U of θ_0,

$$\lim_{T\to\infty} P\left(\sup_{\theta \in U} Y_T(\theta) - \sup_{\theta \in U^c} Y_T(\theta) > 0 \right) = 1 .$$

Let ε be any positive number. For each ω and T, there exists a $\theta_1(\omega, T) \in U^c$ such that

$$\sup_{\theta \in U^c} Y_T(\theta) < Y_T(\theta_1(\omega, T)) + \varepsilon .$$

Hence,

$$\begin{aligned} P&\left(\sup_{\theta \in U} Y_T(\theta) - \sup_{\theta \in U^c} Y_T(\theta) > 0 \right) \\ &\geq P(-\, Y_T(\theta_1(\omega, T)) > \varepsilon) \\ &\geq P(\tilde{\Gamma}(\theta_1(\omega, T)) > 2\varepsilon) \\ &\quad - P(Y_T(\theta_1(\omega, T)) + \tilde{\Gamma}(\theta_1(\omega, T)) > \varepsilon) \\ &\leq P\left(\inf_{\theta \in U^c} \tilde{\Gamma}(\theta) > 2\varepsilon \right) \\ &\quad - P\left(\sup_{\theta \in \bar{\Theta}} | Y_T(\theta) + \tilde{\Gamma}(\theta)| > \varepsilon \right) . \end{aligned} \tag{4.4}$$

As we have from (BS0) and (4.3)

$$\begin{aligned} &\sup_{\theta \in \bar{\Theta}} | Y_T(\theta) + \tilde{\Gamma}(\theta)| \\ &\quad \leq \sup_{\theta \in \bar{\Theta}} \left| \frac{1}{b} M_T(\theta) \right| + \sup_{\theta \in \bar{\Theta}} \left| \frac{1}{b} N_T(\theta) - \tilde{\Gamma}(\theta) \right| \to 0 \end{aligned}$$

in probability, by (4.4)

$$\begin{aligned} &\liminf_{T} P\left(\sup_{\theta \in U} Y_T(\theta) - \sup_{\theta \in U^c} Y_T(\theta) > 0 \right) \\ &\quad \geq P\left(\inf_{\theta \in U^c} \tilde{\Gamma}(\theta) > 2\varepsilon \right) . \end{aligned}$$

When $\varepsilon \downarrow 0$, the r.h.s. of the above inequality tends to 1 as θ_0 is the unique point at which $\tilde{\Gamma}(\theta)$ is minimum from (EH0). This completes the proof.

The asymptotic behavior of the ratio of Q's after normalizing is given by the following theorem, which is a generalization of locally asymptotic mixed normality appearing in "non-ergodic" statistical inference.

THEOREM 4.2. *Under Conditions* (TS), (TH), (SOL), (EH2) *and* (ES1), *the following expansion holds: for each* $u \in R^k$,

$$\log Z_T(u) = u^m \Delta_{mT} - \frac{1}{2} u^m u^n \Gamma_{mnT} + \rho_T(u)$$

where $\rho_T(u) \to 0$ *in probability,* $(\Phi_T, \Gamma_T) \to (\Phi, \Gamma)$ *in probability,* $\mathscr{L}\{\Delta_T, \Phi_T, \Gamma_T | P\} \to \mathscr{L}\{\Phi^{1/2}N, \Phi, \Gamma\}$ *and* $N \sim N_k(0, I_k)$ *independent of* (Φ, Γ).

PROOF. By definition,

$$\begin{aligned} Z_T(u) &= Q(T, X, \theta_0 + b^{-1/2}u) / Q(T, X, \theta_0) \\ &= \exp\left\{ \int_0^T \Delta S_i(\theta_0, \theta_0 + b^{-1/2}u)[dX_t^i - V_0^i dt] \right. \\ &\qquad \left. - \int_0^T \Delta H_t(\theta_0, \theta_0 + b^{-1/2}u)\, dt \right\}. \end{aligned}$$

Therefore, by using Ito's formula, we obtain

$$\begin{aligned} \log Z_T(u) &= b^{-1/2} u^m \int_0^T \delta_m S_i(\theta_0)[dX_t^i - V_0^i dt] + \int_0^T D^{(1)} S_i(\theta_0, u)[dX_t^i - V_0^i]dt \\ &\quad - b^{-1/2} u^m \int_0^T \delta_m H(\theta_0)\, dt - (2b)^{-1} u^m u^n \int_0^T \delta_m \delta_n H(\theta_0)\, dt \\ &\quad - \int_0^T D^{(2)} H(\theta_0, u)\, dt \\ &= b^{-1/2} u^m \int_0^T \delta_m S_i(\theta_0)[dX_t^i - V_0^i dt] - b^{-1/2} u^m G_m(X_T) \\ &\quad + b^{-1/2} u^m G_m(X_0) + b^{-1/2} u^m \int_0^T \partial_i G_m(X_t)[dX_t^i - V_0^i dt] \\ &\quad - (2b)^{-1} u^m u^n \int_0^T \delta_m \delta_n H(\theta_0)\, dt + \int_0^T D^{(1)} S_i(\theta_0, u)[dX_t^i - V_0^i]dt \\ &\quad - \int_0^T D^{(2)} H(\theta_0, u)\, dt \qquad (\because \text{(SOL)}) \end{aligned}$$

$$= u^m \Delta_{mT} - \frac{1}{2} u^m u^n \Gamma_{mnT} + \rho_T(u)$$

where

$$\Delta_{mT} = b^{-1/2} \int_0^T [\delta_m S_i(\theta_0) + \partial_i G_m(X_t)][dX_t^i - V_0^i dt] ,$$

$$\Gamma_{mnT} = b^{-1} \int_0^T \delta_m \delta_n H(\theta_0) dt ,$$

$$\rho_T(u) = -b^{-1/2} u^m G_m(X_T) + b^{-1/2} u^m G_m(X_0) + \int_0^T D^{(1)} S_i(\theta_0, u)[dX_t^i - V_0^i dt] - \int_0^T D^{(2)} H(\theta_0, u) dt .$$

For any positive ε, let $\bar{\delta} = \bar{\delta}(\varepsilon(1 + |u|^2)^{-1})$, then since

$$P\left(\left|\int_0^T D^{(2)} H(\theta_0, u) dt\right| > \varepsilon\right) \le P(|u| > \bar{\delta} b^{1/2}) + P\left(\sup_{|\xi| \le \bar{\delta} b^{1/2}} \frac{1}{1 + |\xi|^2} \left|\int_0^T D^{(2)} H(\theta_0, \xi) dt\right| > \frac{\varepsilon}{1 + |u|^2}\right),$$

by (TH)

$$\int_0^T D^{(2)} H(\theta_0, u) dt \to 0$$

in probability as $T \to \infty$. Similarly, by (TS)

$$\int_0^T D^{(1)} S_i(\theta_0, u)[dX_t^i - V_0^i dt] \to 0$$

in probability as $T \to \infty$. Therefore, Conditions (TS), (TH) and (SOL) imply that for each u,

$$\rho_T(u) \to 0$$

in probability. On the other hand, (EH2) and (ES1) yield that

$$(\Phi_T, \Gamma_T) \to (\Phi, \Gamma)$$

in probability. By a martingale CLT and the stability of the weak convergence, we have $\mathscr{L}\{\Delta_T, \Phi_T, \Gamma_T\} \to \mathscr{L}\{\Phi^{1/2} N, \Phi, \Gamma\}$ and $N \sim N_k(0, I_k)$

independent of (Φ, Γ) (see Aldous and Eagleson (1978), Feigin (1985) and Yoshida (1987)).

It is easy to show the convergence of finite dimensional marginals of $\{Z_T(\cdot)\}$.

We now proceed to the distribution of M-estimators. Define

$$B_c = \{u \in R^k;\ |u| \le c\}\ ,$$
$$B_{cT} = \{u \in R^k;\ |u| \le c,\ \theta_0 + b^{-1/2}u \in \bar{\Theta}\} \quad \text{and}$$
$$w_T(\delta, c) = \sup_{\substack{|u_2 - u_1| \le \delta \\ u_1, u_2 \in B_{cT}}} |\log Z_T(u_2) - \log Z_T(u_1)|\ .$$

Then, for large T, $B_{cT} = B_c$.

LEMMA 4.1. *Suppose that* (TS), (TH), (SOL), (EH2) *and* (ES1) *are satisfied. Then, for each* $\varepsilon > 0$ *and* $c > 0$, *as* $\delta \to 0$,

$$\overline{\lim_{T\to\infty}}\ P(w_T(\delta, c) > \varepsilon) \to 0\ .$$

PROOF. From Theorem 4.2,

$$|\log Z_T(u_2) - \log Z_T(u_1)|$$
$$\le |u_2 - u_1|\,|\Delta_T| + B|u_2 - u_1|\,|\Gamma_T| + \sum_{i=1}^{2} |\rho_T(u_i)|$$

where B is a constant. Therefore,

$$P(w_T(\delta, c) > \varepsilon)$$
$$\le P(|\Delta_T| > \varepsilon/4\delta) + P(|\Gamma_T| > \varepsilon/4B\delta)$$
$$+ 2P\left(\sup_{u \in B_{cT}} |\rho_T(u)| > \varepsilon/4\right)\ .$$

Using (TH) for $\bar{\delta} = \bar{\delta}(\varepsilon(1 + c^2)^{-1}/16)$,

$$P\left(\sup_{|u| \le c} \left|\int_0^T D^{(2)}H(\theta_0, u)dt\right| > \varepsilon/16\right)$$
$$\le P(c > \bar{\delta} b^{1/2})$$
$$+ P\left(\sup_{|u| \le \bar{\delta} b^{1/2}} (1 + |u|^2)^{-1} \left|\int_0^T D^{(2)}H(\theta_0, u)dt\right| > \frac{\varepsilon}{16(1 + c^2)}\right)$$

$$\to 0$$

in probability. Similarly, by (TS)

$$P\left(\sup_{|u|\le c}\left|\int_0^T D^{(1)}S_i(\theta_0,u)[dX_t^i - V_0^i dt]\right| > \varepsilon/16\right) \to 0$$

in probability. Since

$$\lim_{T\to\infty} P\left(\sup_{u\in B_{cT}} |\rho_T(u)| > \varepsilon/4\right) = 0$$

by the above and (SOL),

$$\overline{\lim_{T\to\infty}}\, P(w_T(\delta,c)>\varepsilon)$$
$$\le \overline{\lim_{T\to\infty}}\, P(|\Delta_T| > \varepsilon/4\delta) + \overline{\lim_{T\to\infty}}\, P(|\Gamma_T| > \varepsilon/4B\delta)\,.$$

The sequences of the r.v.'s Δ_T, Γ_T converge in distribution, and so

$$\lim_{\delta\to 0}\overline{\lim_{T\to\infty}}\, P(w_T(\delta,c)>\varepsilon) = 0\,.$$

LEMMA 4.2. *Suppose that Conditions* (TS), (TH), (SOL), (EH2) *and* (ES1) *hold. Then, for* $c>0$, *as* $N\to\infty$,

$$\overline{\lim_{T\to\infty}}\, P\left(\sup_{u\in B_{cT}} |\log Z_T(u)| > N\right) \to 0\,.$$

PROOF. As in the proof of Lemma 4.1, it is easy to show the negligibility of $\rho_T(u)$ by (TS), (TH) and (SOL). Conditions (EH2) and (ES1) ensure the convergence of Δ_T and Γ_T in distribution, which concludes the proof.

LEMMA 4.3. *Under Conditions of Section* 3 *for each* $\varepsilon>0$,

$$\lim_{c\to\infty}\overline{\lim_{T\to\infty}}\, P\left(\sup_{|u|\ge c} Z_T(u) > \varepsilon\right) = 0\,.$$

PROOF. Since Γ is positive definite for a.s. ω, for any $\varepsilon_1>0$ there exists an $\varepsilon_2>0$ such that $P\{A(\varepsilon_2)\}\ge 1-\varepsilon_1$, where

$$A(\varepsilon_2) = \left\{\omega;\ \varepsilon_2|u|^2 \le \frac{1}{4}u^m u^n \Gamma_{mn} \text{ for all } u\in R^k\right\}.$$

Let

$$r_T(u) = \frac{1}{2} u^m u^n (\Gamma_{mn} - \Gamma_{mnT}) + \rho_T(u) .$$

Then,

$$\begin{aligned}(1 + |u|^2)^{-1} |r_T(u)| &\le \frac{1}{2} \sum_{mn} |\Gamma_{mn} - \Gamma_{mnT}| + \frac{1}{2} b^{-1/2} \sum_m |G_m(X_T)| \\ &+ \frac{1}{2} b^{-1/2} \sum_m |G_m(X_0)| \\ &+ (1 + |u|^2)^{-1} \left| \int_0^T D^{(1)} S_i(\theta_0, u)[dX_t^i - V_0^i dt] \right| \\ &+ (1 + |u|^2)^{-1} \left| \int_0^T D^{(2)} H(\theta_0, u) dt \right| .\end{aligned}$$

Conditions (EH2), (SOL), (TS) and (TH) yield that for any $\varepsilon_3 > 0$, there exists $\bar{\delta} = \bar{\delta}(\varepsilon_3/4)$ such that $\lim_{T\to\infty} P(S(T, \bar{\delta})) = 1$, where

$$S(T, \bar{\delta}) = \left\{ \omega; \sup_{|u| \le \bar{\delta} b^{1/2}} (1 + |u|^2)^{-1} |r_T(u)| < \varepsilon_3 \right\} .$$

Let $\varepsilon_3 < \varepsilon_2$. For $\omega \in S(T, \bar{\delta}) \cap A(\varepsilon_2)$, when $|u| \le \bar{\delta} b^{1/2}$,

$$\begin{aligned}\log Z_T(u) &= u^m \Delta_{mT} - \frac{1}{2} u^m u^n \Gamma_{mn} + r_T(u) \\ &\le u^m \Delta_{mT} - \frac{1}{2} u^m u^n \Gamma_{mn} + \varepsilon_3 (1 + |u|^2) \\ &\le |u| \, |\Delta_T| - \varepsilon_2 |u|^2 + \varepsilon_3 .\end{aligned}$$

Then

$$\begin{aligned}&P\left\{ \sup_{r \le |u| \le \bar{\delta} b^{1/2}} Z_T(u) \ge \exp(-\varepsilon_2 r^2/2) \right\} \\ &\quad \le P\{S(T, \bar{\delta})^c\} + P\{A(\varepsilon_2)^c\} \\ &\qquad + P\left\{ \sup_{r \le |u| \le \bar{\delta} b^{1/2}} (|u| \, |\Delta_T| - \varepsilon_2 |u|^2) + \varepsilon_3 \ge -\varepsilon_2 r^2/2 \right\}\end{aligned}$$

$$\leq P\{|\Delta_T| > 2\varepsilon_2 r\} + P\{r|\Delta_T| - \varepsilon_2 r^2 + \varepsilon_3 \geq -\varepsilon_2 r^2/2\} + \varepsilon_1 + o(1)$$
$$\leq 2P\{|\Delta_T| > \varepsilon_2 r/2 - \varepsilon_3/r\} + \varepsilon_1 + o(1) .$$

Let ε and γ be arbitrary positive numbers and let $\varepsilon_1 = \varepsilon/3$. When r is large enough so that

$$\exp(-\varepsilon_2 r^2/2) < \gamma$$

and

$$\overline{\lim_{T\to\infty}} P\{|\Delta_T| > \varepsilon_2 r/2 - \varepsilon_3/r\} < \varepsilon/3 ,$$

then

$$\overline{\lim_{T\to\infty}} P\left\{ \sup_{r\leq|u|\leq\delta b^{1/2}} Z_T(u) \geq \gamma \right\} \leq \varepsilon .$$

Now,

$$\overline{\lim_{T\to\infty}} P\left[\sup_{\delta b^{1/2}\leq|u|} Z_T(u) \geq \gamma \right]$$
$$= \overline{\lim_{T\to\infty}} P\left[\sup_{\delta\leq|h|} Y_T(\theta_0 + h) \geq \frac{1}{b} \log \gamma \right]$$
$$= P\left[\sup_{\delta\leq|h|} (-\tilde{\Gamma}(\theta_0 + h)) \geq 0 \right] = 0 ,$$

in view of the proof of Theorem 4.1 and (EH0). Therefore, one has

$$\overline{\lim_{T\to\infty}} P\left[\sup_{r\leq|u|} Z_T(u) \geq \gamma \right] \leq \varepsilon .$$

This completes the proof (see also Section 84 of Strasser (1985)).

Let $C_0(R^k)$ be the Banach space of continuous functions on R^k vanishing at infinity with sup-norm. We extend $Z_T(u)$ to a function of $C_0(R^k)$ which has its maxima inside $B_T = \{u; \theta_0 + b^{-1/2}u \in \bar{\Theta}\}$, using the same notation. Using the weak convergence of finite dimensional marginals of the random fields $\{Z_T(\cdot); u \in R^k\}$ and Lemmas 4.1–4.3, the convergence in distribution of the sequence can be proved.

THEOREM 4.3. *Under Conditions of Section* 3,

$$Z_T(\cdot) \xrightarrow{\mathscr{D}} Z(\cdot) \quad in \quad C_0(R^k) \quad as \quad T \to \infty ,$$

where

$$\log Z(u) = u^m \Delta_m - \frac{1}{2} u^m u^n \Gamma_{mn} ,$$

$$(\Delta_m) = \Phi^{1/2} N, \qquad (N \text{ is given in Theorem } 4.2) .$$

The weak convergence theorem is available for investigating the asymptotic properties of M-estimators and test statistics related to them. Inagaki and Ogata (1975) mentioned to various applications including likelihood ratio tests, AIC and so on.

Let $\hat{u}_T$ be a point at which $Z_T(u)$ is maximum in $\{u;\, \theta_0 + b^{-1/2} u \in \bar{\Theta}\}$ and $\hat{u}$ be a point at which $Z(u)$ is maximum in R^k. From the weak convergence of the random fields $\{Z_T(\cdot)\}$, one has

$$\hat{u}_T \xrightarrow{\mathscr{D}} \hat{u} = \Gamma^{-1} \Delta .$$

THEOREM 4.4. *Suppose that Conditions in Section* 3 *hold. Then,*

$$\mathscr{L}\{b^{1/2}(\hat{\theta} - \theta_0) | P\} \to \mathscr{L}\{\Gamma^{-1}\Delta\} ,$$

$$\Delta = \Phi^{1/2} N ,$$

$$N \sim N_k(0, I_k) \quad independent\ of \quad (\Phi, \Gamma) .$$

In particular, if Φ and Γ are deterministic, the limit has the normal distribution $N_k(0, \Gamma^{-1}\Phi\Gamma^{-1})$.

If we estimate θ with a correctly specified model, it is possible to eliminate Condition (SOL).

5. Estimation by misspecified model

In parameter estimation it is known that if the parametric model contains the true one, the maximum likelihood method is effective; e.g., in the sense that it is asymptotically minimax in some class of regular estimators. So, if an observer assures himself that his parametric model is correct after consideration, he should estimate unknown parameters maximizing the likelihood induced from it.

As is well-known in the extensive literature of robust estimation, the maximum likelihood estimator is sensitive to contamination of the data and misspecification of the model. The aim of this section is to study how the difference between the true model and the parametric model affects the asymptotic behavior of the maximum likelihood estimator. Let the observer's parametric model be defined by

$$
(5.1)\qquad \begin{aligned} dX_t^i &= \sum_{\beta=1}^{r'} A_\beta^i(X_t)dW^\beta(t) + A_0^i(X_t,\theta)dt, \qquad 1 \le i \le d\,,\\ X_0 &= \eta\,. \end{aligned}
$$

Then the likelihood function based on the observation $\{X_t; 0 \le t \le T\}$ is given by the following formula:

$$
(5.2)\qquad \exp\left\{\int_0^T A_0^i a_{ij} dX_t^j - \frac{1}{2}\int_0^T A_0^i a_{ij} A_0^j dt\right\},
$$

where $(a_{ij}) = (a^{ij})^+$ (the symmetric generalized inverse) and $a^{ij} = \sum_{\beta=1}^{r'} A_\beta^i A_\beta^j$, Liptser and Shiryayev (1977). Since in our words this is $Q(T, X, \theta)$ with

$$
S_i = A_0^j a_{ij} \quad \text{and} \quad R = A_0^i a_{ij} A_0^j\,,
$$

we are ready to answer our problem. Though it is easy to extend it to more general cases, we confine ourselves to the ergodic case for simplicity.

PROPOSITION 5.1. *Let X be ergodic and stationary with invariant distribution $\nu(dx)$. Suppose that there exists a θ_0 at which*

$$
\begin{aligned} \theta \to \int_{R^d}\Big[\frac{1}{2} A_0^i(\theta)a_{ij}A_0^j(\theta) - A_0^i(\theta)a_{ij}V_0^j \\ - \frac{1}{2} A_0^i(\theta_0)a_{ij}A_0^j(\theta_0) + A_0^i(\theta_0)a_{ij}V_0^j\Big]d\nu(x) \end{aligned}
$$

has a unique minimum and Γ with components

$$
\Gamma_{mn} = \int_{R^d} \gamma_{mn}\, d\nu(x)\,,
$$

where

$$
\begin{aligned} \gamma_{mn} &= (\delta_m A_0^i(\theta_0))a_{ij}(\delta_n A_0^j(\theta_0)) \\ &\quad + (\delta_m\delta_n A_0^i(\theta_0))a_{ij}[A_0^j(\theta_0) - V_0^j]\,, \end{aligned}
$$

is positive definite. Moreover, suppose the following conditions:

(1) *V_0^i, V_α^i, a_{ij} and maxima of A_0^i and their derivatives up to the third*

order with respect to θ belong to $\bigcap_{p>1} L^p(R^d, \nu)$.

(2) *There exist functions G_m such that $(\partial_i G_m)^2 V^{ij} \in L^1(R^d, \nu)$ and*

$$LG_m = (\delta_m A_0^i(\theta_0)) a_{ij}[A_0^j(\theta_0) - V_0^j] .$$

Then convergence of distributions of random fields $Z_T(\cdot)$ holds and

$$\mathscr{L}\{T^{1/2}(\hat{\theta}_T - \theta_0) | P\} \to N_k(0, \Gamma^{-1}\Phi\Gamma^{-1})$$

where $\Phi = (\Phi_{mn})$ with

$$\Phi_{mn} = \int (\delta_m A_0^p(\theta_0) a_{ip} + \partial_i G_m)(\delta_m A_0^q(\theta_0) a_{jq} + \partial_j G_n) v^{ij} d\nu(x) .$$

If the parametric model contains the true one, (2) above is not necessary as $G_m = 0$, and the maximum likelihood estimator is consistent with asymptotic variance Γ^{-1}, the reciprocal of Fisher information. The integrability in (1) can be weakened.

Thus far we have used Ito stochastic integrals. Another approach is to formulate our problem using Stratonovich's symmetric stochastic integrals. It is convenient when a stochastic integral is regarded as a limit of integrals with respect to smooth stochastic processes (see Wong and Zakai (1965) and McShane (1972, 1974)) and also when one treats diffusion processes on manifolds, as in navigation problems. In ordinary cases the two formulation are equivalent and it does not matter which integrals are used. But this is not the case when the true model is misspecified. If the observer rewrites (5.1) in the Stratonovich form, the estimate would be different from the previous one because their estimating function differs under the true model.

Consider a family of stochastic differential equations

$$(5.1') \quad \begin{aligned} dX_t^i &= \sum_{\beta=1}^{r'} \bar{A}_\beta^i(X_t) \circ dW^\beta(t) + \bar{A}_0^i(X_t, \theta) dt, \qquad 1 \le i \le d , \\ X_0 &= \eta , \end{aligned}$$

where $\circ$ stands for the Stratonovich integral. In accord, let the true process be described by

$$(2.2') \quad \begin{aligned} dX_t^i &= \sum_{\alpha=1}^{r} \bar{V}_\alpha^i(X_t) \circ dW^\alpha(t) + \bar{V}_0^i(X_t, \theta) dt, \qquad 1 \le i \le d , \\ X_0 &= \eta . \end{aligned}$$

For the maximum likelihood estimator corresponding to

$$Q(T, X, \theta) = \exp\left\{\int_0^T \bar{S}_i(X_t, \theta) \circ dX_t^i - \frac{1}{2}\int_0^T \bar{R}(X_t, \theta)\, dt\right\}$$

with

$$\bar{S}_i = \left[\bar{A}_0^j + \frac{1}{2}(\bar{A}_\beta \bar{A}_\beta^j)\right]\bar{a}_{ji}\,,$$

$$\bar{R} = \left[\bar{A}_0^i + \frac{1}{2}(\bar{A}_\beta \bar{A}_\beta^i)\right]\bar{a}_{ij}\left[\bar{A}_0^j + \frac{1}{2}(\bar{A}_\beta \bar{A}_\beta^j)\right]$$

$$+ \bar{A}_\beta^i \bar{A}_\beta\left(\left[\bar{A}_0^j + \frac{1}{2}(\bar{A}_\beta \bar{A}_\beta^j)\right]\bar{a}_{ji}\right),$$

we have a similar result to Proposition 5.1:

$$\mathscr{L}\{T^{1/2}(\hat{\theta}_T - \theta_0)\,|\,P\} \to N_k(0, \bar{\Gamma}^{-1}\bar{\Phi}\bar{\Gamma}^{-1})$$

where

$$\bar{\Phi}_{mn} = \int_{R^d} \phi_i \phi_j \bar{v}_{ij} \bar{\nu}(dx)\,,$$

$$\bar{\Gamma}_{mn} = \int_{R^d} \{(\delta_m \bar{A}_0^i)(\delta_n \bar{A}_0^j)\bar{a}_{ij} - [\bar{L}^i - \bar{L}^i(\theta_0)][\delta_m \delta_n \bar{A}_0^j a_{ij}]\}\bar{\nu}(dx)\,,$$

with θ_0 in place of θ,

$$\bar{L}^i = \frac{1}{2}\bar{V}_\alpha \bar{V}_\alpha^i + \bar{V}_0^i, \qquad \bar{L}^i(\theta) = \frac{1}{2}\bar{A}_\beta \bar{A}_\beta^i + \bar{A}_0^i\,.$$

Here $\bar{v}^{ij}$, $\bar{a}_{ij}$ and $\bar{\nu}$ are naturally defined and ϕ_i are functions satisfying

$$\bar{L}^i \phi_i = \bar{L}^i(\theta_0)(\delta_m \bar{A}_0^j \bar{a}_{ji})\,.$$

For details see Yoshida (1988*b*). In particular, if the parametric model contains the true one, the asymptotic variance is $\bar{\Gamma}^{-1}$ where

$$\bar{\Gamma}_{mn} = \int_{R^d} (\delta_m \bar{A}_0^i)(\delta_n \bar{A}_0^j) a_{ij} \bar{\nu}_{\theta_0}(dx)\,.$$

6. Applications

This section gives several examples to illustrate our results.

Non-ergodic models. Consider the following one-dimensional linear diffusion process

$$dX_t = -\zeta\theta X_t dt + dW_t ,$$

where θ is a positive parameter and ζ is a positive random variable independent of $\sigma\{W(t);\, t \geq 0\}$. We observe (X, ζ) to estimate θ. The maximum likelihood estimate is obtained by maximizing

$$Q(T, X, \theta) = \exp\left\{\int_0^T -\zeta\theta X_t dX_t - \frac{1}{2}\int_0^T \zeta^2\theta^2 X_t^2 dt\right\}.$$

The process X is ergodic for each given ζ with invariant measure $\nu_\zeta(dx)$ whose density is $(\zeta\theta/\pi)^{1/2} \exp(-\zeta\theta x^2)$. If θ_0 is the true value,

$$\frac{1}{T}\int_0^T \Delta H(\theta_0, \theta) dt \to \tilde{\Gamma} = \frac{\zeta}{4\theta_0}(\theta - \theta_0)^2$$

a.s. and consistency holds. The asymptotic distribution of the maximum likelihood estimator is the mixture of normal distributions $N(0, 2\theta_0/\zeta)$ by ζ. In this example the reader may have the property directly through the explicit representation of the maximum likelihood estimator

$$\hat{\theta}_T = -\int_0^T \zeta X_t dX_t \Big/ \int_0^T \zeta^2 X_t^2 dt .$$

If we do not observe the value of ζ and estimate θ e.g., replacing ζ by $\bar{\zeta}$, the mean of random variable ζ, the consistency fails.

Another example of a non-ergodic model is a non-linear diffusion process defined by

$$dX_t = [g(X_t) + \theta f(X_t)]dt + dW_t, \qquad X_0 = x_0 > 0 ,$$

where $\theta > 0$ is a parameter, g and f are uniformly positive functions such that their derivatives up to second order are bounded, $g' + \theta f' \geq 0$ for large x and

$$\lim_{x\to\infty} g(x)/x = 1 \qquad \text{and} \qquad \lim_{x\to\infty} f(x)/x^\alpha = 1 ,$$

for some α, $0 < \alpha < 1$. As $t \to \infty$, $X_t \to \infty$ with probability 1 (see p. 117 of Gihman and Skorohod (1972)). Moreover, from Appendix it is known that there exists a positive random variable K and $X_t/\eta_t \to K$ a.s., where η_t the inverse of the map

$$x \to \int_{x_0}^{x} [g(u) + \theta_0 f(u)]^{-1} du .$$

See also Keller *et al.* (1984), where they show this with the convexity of coefficient functions of stochastic differential equations. In this example,

$$b_T = \int_0^T \eta_t^{2\alpha} dt ,$$

$$\tilde{\Gamma} = \frac{1}{2} (\theta - \theta_0)^2 K^{2\alpha} \quad \text{and}$$

$$\Phi = \Gamma = K^{2\alpha} .$$

Hence, the maximum likelihood estimator has consistency and asymptotic mixed normality with random variance $\Gamma^{-1} = K^{-2\alpha}$.

Robust estimation. Consider a one-dimensional ergodic and stationary diffusion process defined by

$$dX_t = \sigma(X_t) dW_t + f(X_t, \theta) dt ,$$

$$X_0 = x_0 ,$$

where θ is a parameter in a bounded interval Θ of R^1. The coefficients f and σ are smooth, σ is positive and W is a standard Wiener process. Let X be stationary and ergodic with stationary distribution $\nu(dx, \theta)$ when θ is true. To construct the robust estimator, we may use the M-estimate with the estimating equation

$$M(T, \theta) := \int_0^T \psi(X_t, \theta) dt = 0 ,$$

where ψ has to satisfy

$$\int_{-\infty}^{\infty} \psi(x, \theta) \nu(dx, \theta) = 0$$

for Fisher's consistency. The integration of $M(T, \theta)$ with respect to θ corresponds to the logarithm of Q:

$$\log Q(T, X, \theta) = -\int_0^T H(X_t, \theta) dt ,$$

where $H(x, \theta) := \int^{\theta} \psi(x, \theta) d\theta$.

Let the true realization X be generated by

$$dX_t = \upsilon_1(X_t)dW_t + \upsilon_0(X_t)dt\,,$$
$$X_0 = x_0\,.$$

Let ν be the stationary distribution of the true model, whose density is proportional to

$$2\upsilon_1^{-2}(u)\exp(B(u))$$

where $B(x) = \int_0^x 2\upsilon_1^{-2}(u)\upsilon_0(u)du$. Then θ_0 is a solution of the equation

$$\int_{-\infty}^{\infty} \psi(x,\theta)\nu(dx) = 0\,.$$

Set

$$G(x,\theta) = -\int_0^x \exp(-B(y))dy\int_y^\infty 2\psi(u,\theta)\upsilon_1^{-2}(u)\exp(B(u))du\,,$$
$$G_1(x) = G(x,\theta_0)$$

satisfies the equation

$$\left(\upsilon_0(x)\partial + \frac{1}{2}\upsilon_1(x)\partial^2\right)G_1 = \psi(x,\theta_0)\,.$$

The functions Γ and Φ are given by

$$\Gamma = \int_{-\infty}^{\infty} \delta\psi(x,\theta_0)\nu(dx)$$

and

$$\Phi = \int_{-\infty}^{\infty} \{\partial G_1(x)\}^2\upsilon_1^2(x)\nu(dx)\,.$$

The asymptotic variance of our M-estimator is Φ/Γ^2.

An approach to robust estimation is to construct a bias robust estimator for contaminated stationary distributions. Considering the infinitesimal contamination of $\nu(\cdot,\theta)$ an influence function is defined as

$$IF(x,\psi,\nu(\cdot,\theta)) = \Gamma^{-1}\psi(x,\theta)\,.$$

Our task is then to seek for a function ψ under boundedness of influence functions so that its asymptotic variance is not bad compared to the maximum likelihood estimator even for data free from contamination. One can construct a robust ψ from f and σ as in Yoshida (1988*a*).

Parameter estimation with a one-dimensional diffusion model observing data generated from a two-dimensional diffusion model. Let realization of a process be governed by a second-order stochastic differential equation

$$\ddot{X}_t + a\dot{X}_t + b^2 X_t = \dot{W}_t$$

where a and b are positive constants, $\beta^2 = b^2 - a^2/4 > 0$, and $\dot{W}$ is a white Gaussian noise. Let

$$X_t = \begin{pmatrix} x_{1t} \\ x_{2t} \end{pmatrix} = \begin{pmatrix} X_t \\ \dot{X}_t \end{pmatrix}, \quad K = \begin{pmatrix} 0 & -1 \\ b^2 & a \end{pmatrix} \quad \text{and} \quad B = \begin{pmatrix} 0 \\ 1 \end{pmatrix},$$

it is equivalent to the first-order vector stochastic differential equation

$$(1) \qquad dX_t = -KX_t dt + BdW_t .$$

To illustrate the case where the degrees of the true model and the observer's parametric model are different, we now assume that the observer will fit a one-dimensional diffusion model

$$(2) \qquad dX_t = -\theta X_t dt + dW_t ,$$

to realization, where θ is a positive parameter.

Let

$$C = \begin{pmatrix} \lambda_2 & 1 \\ \lambda_1 & 1 \end{pmatrix}, \quad M(t) = C^{-1} \begin{pmatrix} \exp(-\lambda_1 t) & 0 \\ 0 & \exp(-\lambda_2 t) \end{pmatrix} C$$

and

$$\Sigma(t) = C^{-1} \begin{pmatrix} \dfrac{1-\exp(-2\lambda_1 t)}{2\lambda_1} & \dfrac{1-\exp(-(\lambda_1+\lambda_2)t)}{\lambda_1+\lambda_2} \\ \dfrac{1-\exp(-(\lambda_1+\lambda_2)t)}{\lambda_1+\lambda_2} & \dfrac{1-\exp(-2\lambda_2 t)}{2\lambda_2} \end{pmatrix} C'^{-1},$$

where $\lambda_1 = a/2 + i\beta$ and $\lambda_2 = a/2 - i\beta$ are the eigenvalues of K. Then the transition probability density of the two-dimensional process defined by (1) is given by

$$p_t(x,y) = (2\pi)^{-1}|\Sigma(t)|^{-1/2} \exp\left\{ -\frac{1}{2}(y - M(t)x)'\Sigma(t)^{-1}(y - M(t)x) \right\}.$$

The degenerate diffusion described by (1) is ergodic (see Arnold and Kliemann (1987)), and its invariant measure has a density, the limit of $p_t(x,y)$ as $t \to \infty$,

$$p_\infty(y) = (2\pi)^{-1}|\Sigma|^{-1/2} \exp\left\{ -\frac{1}{2} y'\Sigma^{-1}y \right\},$$

where

$$\Sigma = \begin{pmatrix} 1/2ab^2 & 0 \\ 0 & 1/2a \end{pmatrix}.$$

Our observer may estimate the unknown parameter θ, for example, with the M-estimator corresponding to

$$Q(T, X, \theta) = \exp\left\{ -\frac{1}{2}\int_0^T (\theta^2 x_{1t}^2 - \theta)dt \right\},$$

which is an asymptotically equivalent version of the maximum likelihood estimator derived from (2) if it does generate data. A simple calculation yields that $\theta_0 = ab^2$. To obtain G_1, consider the Neumann function

$$N(x,y) = \int_0^\infty [p_t(x,y) - p_\infty(y)]dt .$$

Then

$$G_1(x) = \int_{-\infty}^\infty \int_{-\infty}^\infty N(x,y)\delta H(y_1, \theta_0)dy_1dy_2$$
$$= c_0 + c_{11}x_1^2 + c_{12}x_1x_2 + c_{22}x_2^2 ,$$

with

$$c_0 = \frac{-a^4 + 2a^2b^2 + 8b^2}{16\beta^2ab^2},$$

$$c_{11} = \frac{a^4 - 3a^2b^2 - 4b^4}{8\beta^2},$$

$c_{12} = -a$ and

$$c_{22} = -1/2\,,$$

satisfies

$$LG_1 = ab^2x_1^2 - 1/2\,,$$

where

$$L = \frac{1}{2}(\partial_2)^2 + x_2\partial_1 - (b^2x_1 + ax_2)\partial_2\,.$$

It is easy to show the degree of concentration

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}[\partial_2 G_1(x)]^2 p_\infty(x)\,dx_1dx_2\left\{\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}x_1^2 p_\infty(x)\,dx_1dx_2\right\}^{-2}$$
$$= 2ab^2(a^2 + b^2)\,.$$

Testing hypotheses. Consider a test $H{:}\,\theta = \theta_1$ against $K{:}\,\theta = \theta_1 + b^{-1/2}u$, $u \in U$, where $U = \{u \in R^k;\, Au = \xi\}$ for fixed fullrank matrix A and vector ξ. The map $f \to \sup_{u\in U} f(u)$ is continuous on $C_0(R^k)$. By the continuity theorem, if $\theta_1 = \theta_0$ and $Q(T, X, \theta)$ is a likelihood function of a correct model, the log likelihood ratio statistic

$$q_T = 2\log\left[\sup_{u\in U} Q(T, X, \theta_1 + b^{-1/2}u)/\,Q(T, X, \theta_1)\right]$$

converges to

$$\sup_{u\in U} 2\log Z(u) = \Delta'\Gamma^{-1}\Delta - (A\Gamma^{-1}\Delta - \xi)'(A\Gamma^{-1}A')^{-1}(A\Gamma^{-1}\Delta - \xi)$$

in distribution under P_{θ_0}. Locally asymptotic mixed normality and contiguity argument yield that under P_{θ_T}, $\theta_T = \theta_0 + b^{-1/2}h$, $h \in U$, q_T converges weakly to

$$(\Delta + \Gamma h)'\Gamma^{-1}(\Delta + \Gamma h)$$
$$- (A\Gamma^{-1}\Delta + Ah - \xi)'(A\Gamma^{-1}A')^{-1}(A\Gamma^{-1}\Delta + Ah - \xi)\,.$$

Constructing likelihood ratio tests of asymptotically size-α and calculating the Pitman power are easy using these results. A similar argument is possible for nonlinear restrictions. For other applications of convergence of the likelihood ratio fields including AIC, see Inagaki and Ogata (1975).

When misspecified models are treated, likelihood ratio tests may have no asymptotic property. Consider diffusion processes with deterministic drifts defined by $dX_t = \theta a_t dt + dW_t$ and testing $H: \theta = 0$ against $K: \theta_T = T^{-1/2}u$, $u \neq 0$. If the true model is $dX_t = b_t dt + dW_t$, then it is easy to give cases where the likelihood ratio statistic $Q(T, X, \theta_T)/Q(T, X, 0)$ converges to 1, ∞ or no value.

How does misspecification of the true model influence the size-α likelihood ratio statistic testing $H: \theta = \theta_0$ against $K: \theta = \theta_1 + b^{-1/2}u$, $u \in U$, a subset of R^k, when $\theta_1 = \theta_0$? Since q_T converges in law to $\sup_{u \in U} 2 \log Z(u)$, $P\left\{ \sup_{u \in U} 2 \log Z(u) > c_\alpha \right\}$ is an error of the first kind, where c_α is determined by $P_{\theta_0}\left\{ \lim_{T \to \infty} q_T > c_\alpha \right\} = \alpha$. This differs from α, of course. As an example, let the true model be ergodic with invariant measure ν satisfying

$$dX_t = [C(X_t, \theta_0) + D(X_t)]dt + dW_t ,$$

where θ is a scalar parameter, and let the parametric model used for testing be

$$dX_t = C(X_t, \theta)dt + dW_t .$$

Here assume that $D(\cdot)$ is orthogonal to the linear span of the family $\{C(\cdot, \theta)\}$ in $L^2(R, \nu)$. Then q_T has asymptotically $\chi^2(1)$-distribution under P_{θ_0}. Let c_α be its α probability point. On the other hand, for the true model q_T converges to $\Gamma^{-1}\Phi N^2$ in distribution, where

$$\Gamma = \int (\delta C(\theta_0))^2 d\nu, \qquad \Phi = \int (\delta C(\theta_0) + \partial G_1)^2 d\nu ,$$

G_1 is a solution of $Lu = -D\delta C(\theta_0)$ and N is a standard normal random variable. The error of the first kind is asymptotically $P\{\chi^2(1) > \Gamma\Phi^{-1}c_\alpha\}$.

It is also possible to calculate the Pitman power for tests with M-estimators.

Acknowledgements

The author is grateful to Professor Nobuo Inagaki for his extensive help and also wishes to thank the referees for their careful reading of the manuscript and their valuable comments.

Appendix

Here we state an a.s. convergence theorem which serves to show convergence of conditional Fisher information in non-ergodic statistical inference.

Let (Ω, F, P) be a complete probability space with filtration (F_t), $t \geq 0$. Let $X = (X(t), F_t)$, $t \geq 0$, be a continuous semimartingale defined on it with canonical decomposition

$$X(t) = X(0) + A(t) + X^c(t)$$

where X^c is the continuous martingale part of X and A is a locally bounded variation process. Set $C_t = \langle X^c \rangle_t$. Consider the following conditions.

(A1) There exists a function b of class $C^1((0,\infty) \times [0,\infty) \to (0,\infty))$ such that $A(t) = \int_0^t b(X(s), s)\,ds$.

(A2) $X(0) = x_0 \geq 0$.

Set

$$G(x,t) = \int_{x_0}^{x} b^{-1}(u,t)\,du\,.$$

(A3) $G(\infty, t) = \infty$, $t \in [0,\infty)$.

The mapping $p\colon (x,t) \to (G(x,t),t)$ is one-to-one and let $p^{-1}(z,t) = (H(z,t),t)$ be its inverse. We borrow the notation of Keller *et al.* (1984), i.e., for a function $f(x,t)$, $\tilde{f}$ is defined by

$$\tilde{f}(z,t) = f(p^{-1}(z,t))\,.$$

(A4) There exists a non-negative measurable function $a(x,t)$ on $(0,\infty) \times [0,\infty)$ such that $dC_t = a^2(X(s),s)\,ds$.

Let $\tilde{B}_1(z,t) = [a/b]^{\sim}(z,t)$, $\tilde{B}_2(z,t) = (1/2)[b_x' a^2/b^2]^{\sim}(z,t)$ and $K(x,s) = G_t'(x,s) - B_2(x,s)$.

(A5) $\lim_{x,s \to \infty} K(x,s) = 0$.

(A6) $\tilde{B}_1$ is bounded and there exists a positive constant δ_0 $(0 < \delta_0 < 1)$ such that

$$\limsup_{x,s \to \infty} \sup_{\delta_0 \leq \delta \leq 1} |\tilde{B}_1(x,s)/\tilde{B}_1(\delta x,s)| < \infty\,,$$

and if $q = (1 + \delta_0)/2$, $\int_0^\infty \tilde{B}_1^2(qs,s)\,dx < \infty$.

(A7) There exists a positive function L such that L is regularly varying with a finite exponent, extremely monotone, $\int^\infty L(x)\,dx < \infty$ and $|\tilde{K}(u,s) - \tilde{K}(v,s)| \leq L(u \wedge v)|u - v|$ for large u, v, s.

(A8) $\int_0^t \tilde{K}(s,s)ds$ converges as $t \to \infty$.

(A9) $\lim_{z,t\to\infty} H(z,t) = \infty$.

Let $h(t)$ be a positive function, $\lim_{t\to\infty} h(t) = \infty$, and let

$$P(h) = \{\Pi; \Pi \text{ be defined on } (0,\infty)\times[0,\infty) \text{ and for any non-negative function } a(t) \text{ such that } r = \lim a(t)/h(t) > 0 \text{ exists, } \lim \Pi(a(t),t)/\Pi(h(t),t) > 0 \text{ exists and this becomes a function of } r\}.$$

If we treat functions which are independent of t, $P(h)$ is properly larger than the class of functions that are regularly varying with finite exponents and extremely monotone.

(A10) $\tilde{b}(x,t)$ belongs to $P(t)$.

THEOREM A.1. *Suppose that* (A1)–(A9) *are satisfied and* $\lim_{x,t\to\infty} b(x,t)/x = \lambda$. *Then* $X(t)/u(t)$ *converges to* $\exp(\lambda W)$ *a.s. on* $\{X(t)\to\infty\}$ *as* $t\to\infty$, *where* W *is a random variable and* $u(t) = H(t,t)$. *Moreover, if* $\lambda = 0$ *and* (A10) *holds,* $[X(t) - u(t)]/\tilde{b}(t,t)$ *converges to* W *a.s. on* $\{X(t)\to\infty\}$ *as* $t\to\infty$.

For its proof and a general version for semimartingales see Yoshida (1987).

Remarks. (1) In (A6), "$\sup_{\delta_0\le\delta\le 1}$" can be replaced by "$\limsup_{\delta\uparrow 1}$" and the second inequality by $\int_0^\infty \sup_{s\ge 0} \tilde{B}_1^2(x,s)dx < \infty$. If $\tilde{B}_1$ has only the space-argument and decreasing in it, (A6) is satisfied.

(2) For (A9) it suffices that

$$\liminf_{t\to\infty} \inf_{x_0\le x} b(x,t) > 0 .$$

REFERENCES

Aldous, D. J. and Eagleson, G. K. (1978). On mixing and stability of limit theorems, *Ann. Probab.*, **6**, 325–331.

Arnold, L. and Kliemann, W. (1987). On unique ergodicity for degenerate diffusions, *Stochastics*, **21**, 41–61.

Basawa, I. V. and Scott, D. J. (1983). Asymptotic optimal inference for non-ergodic models, *Lecture Notes in Statist.*, **17**, Springer, New York.

Bustos, O. (1982). General M-estimates for contaminated pth-order autoregressive processes: consistency and asymptotic normality, *Z. Wahrsch. Verw. Gebiete*, **59**, 491–504.

Feigin, P. D. (1985). Stable convergence of semimartingales, *Stochastic Process. Appl.*, **19**, 125–134.

Gihman, I. I. and Skorohod, A. V. (1972). *Stochastic Differential Equations*, Springer, New York.

Ibragimov, I. A. and Has'minskii, R. Z. (1972). Asymptotic behavior of statistical estimators in the smooth case, *Theory Probab. Appl.*, **17**, 443-460.

Ibragimov, I. A. and Has'minskii, R. Z. (1973). Asymptotic behavior of some statistical estimators II, Limit theorem for the *a posteriori* density and Bayes estimators, *Theory Probab. Appl.*, **18**, 76-91.

Ibragimov, I. A. and Has'minskii, R. Z. (1981). *Statistical Estimation*, Springer, New York.

Inagaki, N. and Ogata, Y. (1975). The weak convergence of likelihood ratio random fields and applications, *Ann. Inst. Statist. Math.*, **27**, 391-419.

Jeganathan, P. (1982*a*). On the asymptotic theory of statistical estimation when the limit of the log-likelihood ratios is mixed normal, *Sankhyā Ser. A*, **44**, 173-212.

Jeganathan, P. (1982*b*). On the convergence of moments of statistical estimators, *Sankhyā Ser. A*, **44**, 213-232.

Keller, G., Kersting, G. and Rösler, U. (1984). On the asymptotic behaviour of solutions of stochastic differential equations, *Z. Wahrsch. Verw. Gebiete*, **68**, 163-189.

Kutoyants, Yu. A. (1977). Estimation of the drift coefficient parameter of a diffusion in the smooth case, *Theory Probab. Appl.*, **22**, 399-406.

Kutoyants, Yu. A. (1978). Estimation of a parameter of a diffusion processes, *Theory Probab. Appl.*, **23**, 641-649.

Kutoyants, Yu. A. (1984). *Parameter Estimation for Stochastic Processes*, (trans. and ed. B. L. S. Prakasa Rao), Heldermann Verlag, Berlin.

Lanska, V. (1979). Minimum contrast estimation in diffusion processes, *J. Appl. Probab.*, **16**, 65-75.

LeCam, L. (1960). Locally asymptotic normal family of distributions, *Univ. Calif. Publ. Statist.*, **3**, 27-98.

Liptser, R. S. and Shiryayev, A. N. (1977). *Statistics of Random Processes*, Springer, New York.

McShane, E. J. (1972). Stochastic differential equations and models of random processes, *Proc. Sixth Berkeley Symp. on Math. Statist. Prob.*, Vol. 3, 263-294, Univ. of California Press, Berkeley.

McShane, E. J. (1974). *Stochastic Calculus and Stochastic Models*, Academic Press, New York.

Novikov, A. A. (1971). On moment inequalities for stochastic integrals, *Theory Probab. Appl.*, **16**, 538-541.

Ogata, Y. and Inagaki, N. (1977). The weak convergence of the likelihood ratio random fields for Markov observations, *Ann. Inst. Statist. Math.*, **29**, 165-187.

Strasser, H. (1985). *Mathematical Theory of Statistics*, Walter de Gruyter, Berlin.

Wong, E. and Zakai, M. (1965). On the convergence of ordinary integrals to stochastic integrals, *Ann. Math. Statist.*, **36**, 1560-1564.

Yoshida, N. (1987). Asymptotic behavior of likelihood ratio in stochastic process (in Japanese), Master Thesis, Osaka University.

Yoshida, N. (1988*a*). Robust *M*-estimators in diffusion processes, *Ann. Inst. Statist. Math.*, **40**, 799-820.

Yoshida, N. (1988*b*). Asymptotic behavior of *M*-estimates in diffusion processes, Research Reports on Statistics, **22**, Osaka Univ.