Statistical Inference for Stochastic Processes **1:** 85–103, 1998.

Asymptotic Expansion of *M*-Estimator Over Wiener Space ⋆

YUJI SAKAMOTO and NAKAHIRO YOSHIDA
Nagoya University and University of Tokyo, Japan

Abstract. In this paper we consider an *M*-estimator defined as a solution of a given estimating function. Sufficient conditions of existence of an *M*-estimator and its stochastic expansion are presented. In the case where the underlying probability space is a Wiener space and the leading term of the stochastic expansion is a martingale, asymptotic expansions of its distribution function are obtained with the aid of Malliavin calculus. Applications to a stationary ergodic diffusion model are also discussed.

Mathematics Subject Classifications (1991): 62E20, 62M05, 60H07.

Key words: asymptotic expansion, *M*-estimator, Malliavin calculus, diffusion process, martingale.

1. Introduction

Let $\Theta \subset \boldsymbol{R}$ be a parameter space, and $(\mathfrak{X}^n, \mathfrak{U}^n, \mu_\theta^n)$ a probability space for every $n \in \boldsymbol{N}^1$ and every $\theta \in \Theta$. Denote by θ_0 the true value of an unknown parameter in Θ. In this article we consider an *M*-estimator $\hat{\theta}$ defined as a solution of an estimating equation $g_n(X_n, \theta) = 0$, where g_n is a given estmating function and X_n is a random variable over $(\mathfrak{X}^n, \mathfrak{U}^n, \mu_\theta^n)$ for every $n \in \boldsymbol{N}$.

Under some regularity conditions, the probability that a solution of the estimating equation $g_n(X_n, \theta) = 0$ exists is close to 1 and an *M*-estimator $\hat{\theta}_n$ is well defined. Furthermore, when $r_n^{-1}(\hat{\theta}_n - \theta_0)$ has an asymptotic distribution for some positive sequence $\{r_n\}$, $n \in \boldsymbol{N}$ tending to zero, we may obtain a stochastic expansion $r_n^{-1}(\hat{\theta}_n - \theta_0) = M_n + r_n N_n + R_n$, where M_n and N_n are random variables, and R_n is a remainder term which is small in a certain sense. If we have an asymptotic expansion of the distribution of $Y_n = M_n + r_n N_n$, it is easy to derive an asymptotic expansion of the distribution of $r_n^{-1}(\hat{\theta}_n - \theta_0)$ with the help of the well-known *Delta* method. We will present a sufficient condition under which an *M*-estimator $\hat{\theta}_n$ is well defined and it has a stochastic expansion as above.

In the case where M_n is a terminal random variable of a continuous martingale, the martingale central limit theorem shows that if the quadratic variation of M_n converges

⋆ The research was supported in part by Grant-in-Aid for Encouragement of Young Scientists from the Ministry of Education, Science and Culture, and by the ISM Cooperative Research Program (97-ISM.CRP-A15), (98-ISM.CRP-A7).

in probability to a constant and N_n has an asymptotic distribution, the distribution of Y_n converges weakly to a normal distribution. As for asymptotic expansions of a martingale, Mykland [3] obtained asymptotic expansions of the expectation $E[f(M_n)]$ for a class C^2-function f, and Mykland [4] extended these results to higher order ones. On the other hand, the cases where f is not regular are treated by Yoshida [8–10] in the light of Malliavin calculus: asymptotic expansions of $P(Y_n \in B)$ for any Borel set B are presented by Yoshida [8], and asymptotic ones of $E[f(Y_n)]$ for any measurable function and the local density $(2\pi)^{-1}\int_{\boldsymbol{R}} \mathrm{e}^{-\mathrm{i}ux} E[\mathrm{e}^{\mathrm{i}uY_n}\psi_n]\,\mathrm{d}u$ for a truncation functional ψ_n are also obtained by Yoshida [9]. We will apply Theorem 2 of Yoshida [8] to the normalized *M*-estimator $r_n^{-1}(\hat{\theta}_n - \theta_0)$.

The next section presents main results about stochastic expansions and asymptotic expansions of *M*-estimator. As an application of them, we treat *M*-estimators based on a stationary ergodic diffusion process in Section 3. The proofs of main results with their preliminary lemmas are given in Section 4.

2. Asymptotic Expansions

First we consider the existence of an *M*-estimator and its stochastic expansion over a general probability space. Suppose that a parameter space Θ is a bounded interval in $\boldsymbol{R}$. The true value of an unknown parameter in Θ is denoted by θ_0. For every $\theta \in \Theta$ and every $n \in \boldsymbol{N}$, let $(\mathfrak{X}^n, \mathfrak{U}^n)$ be a measurable space, and let X_n^θ be an $\mathfrak{X}^n$-valued random variable defined on some probability space. The probability measure induced by X_n^θ is denoted by μ_θ^n. We abbreviate to X_n a random variable X_n^θ evaluated at $\theta = \theta_0$. Furthermore, suppose that an estimating function g_n is a real-valued measurable function on $\mathfrak{X}^n \times \Theta$ such that $g_n(x, \cdot) \in C^3(\Theta \to \boldsymbol{R}^1)$ for any $x \in \mathfrak{X}^n$. The differential operator $\partial/\partial\theta$ is denoted by δ. Then, *M*-estimator $\hat{\theta}_n$ is defined and its stochastic expansion is given as follows.

THEOREM 1. *Assume that for any* $\theta_0 \in \Theta^\circ \equiv \mathrm{Int}(\Theta)$, *there exist* $p_1 > 1$, $p_2 > 1$, $p_3 > 1$, $\gamma > 0$ *with* $2/3 + \max(1/p_2, 1/3p_3) < \gamma - 1/p_1$ *satisfying the following conditions*:

[C1] $\sup_n \|r_n g_n(X_n, \theta_0)\|_{p_1} < \infty$;

[C2] *there exists an open interval* $\tilde{\Theta} \subset \Theta^\circ$ *including* θ_0 *and* $I_{\theta_0}(\theta) \in \boldsymbol{R}^1$ *such that*

$$\inf_{\theta \in \tilde{\Theta}} I_{\theta_0}(\theta) > 0, \qquad \sup_{n \in \boldsymbol{N}, \theta \in \Theta^\circ} \|r_n^{-\gamma}(r_n^2 \delta g_n(X_n, \theta) + I_{\theta_0}(\theta))\|_{p_2} < \infty;$$

[C3] *there exists a function* $L_{\theta_0}(\theta) \in \boldsymbol{R}^1$ *such that*

$$\sup_{n \in \boldsymbol{N},\, \theta \in \Theta^\circ} \|r_n^{-\gamma}(r_n^2 \delta^2 g_n(X_n, \theta) + L_{\theta_0}(\theta))\|_{p_2} < \infty;$$

[C4] $\sup_{n\in\boldsymbol{N}} \| \sup_{\theta\in\Theta^\circ} |r_n^2\delta^3 g_n(X_n,\theta)| \|_{p_3} < \infty$.

Then

$$P[(\exists_1\hat{\theta}_n \in \tilde{\Theta}, g_n(X_n,\hat{\theta}_n) = 0) \quad \textit{and} \quad (|\hat{\theta}_n - \theta_0| \leqslant r_n^\gamma)] = 1 - o(r_n). \tag{2.1}$$

Moreover for some extension $\hat{\theta}_n\colon \mathfrak{X}^n \to \Theta$, there exist constants $C > 0$ and $\epsilon > 0$ such that

$$P[|R| \leqslant C r_n^{\epsilon+1}] = 1 - o(r_n), \tag{2.2}$$

where

$$R = r_n^{-1}(\hat{\theta}_n - \theta_0) \; - \; (I_{\theta_0}^{-1}(\theta_0)Z_{n,1} + r_n I_{\theta_0}^{-2}(\theta_0)Z_{n,1}Z_{n,2} - \\ -\tfrac{1}{2}r_n I_{\theta_0}^{-3}(\theta_0)L_{\theta_0}L_{\theta_0}(\theta_0)Z_{n,1}^2), \tag{2.3}$$

$$Z_{n,1} = r_n g_n(X_n,\theta_0), \qquad Z_{n,2} = r_n^{-1}(r_n^2\delta g_n(X_n,\theta_0) + I_{\theta_0}(\theta_0)). \tag{2.4}$$

Proof. See Section 4. □

Remark 1. If we replace [C2] with the condition that there exists a function $I_{\theta_0}(\theta) \in \boldsymbol{R}^1$ such that

$$I_{\theta_0}(\theta_0) > 0, \qquad \sup_{n\in\boldsymbol{N},\theta\in\Theta^\circ} \|r_n^{-\gamma}(r_n^2\delta g_n(X_n,\theta) + I_{\theta_0}(\theta))\|_{p_2} < \infty,$$

then we see from the continuity of $I_{\theta_0}(\theta)$, which is shown in Lemma 1, that there exists a neighborhood $\tilde{\Theta}$ of θ_0 such that

$$\inf_{\theta\in\tilde{\Theta}} I_{\theta_0}(\theta) > 0.$$

Then we can obtain the same result as above.

Remark 2. When $\tilde{\Theta} = \Theta^\circ$, we obtain the global consistency of the M-estimator $\hat{\theta}_n$. For example, consider the maximum likelihood estimator of the family of diffusion processes defined by the stochastic differential equation:

$$\mathrm{d}X_t = \theta b(X_t)\,\mathrm{d}t + \mathrm{d}w_t.$$

Then g_n is given by

$$g_T((X_t)_{0\leqslant t\leqslant T},\theta) = \int_0^T b(X_t)\,\mathrm{d}X_t - \theta\int_0^T b(X_t)^2\,\mathrm{d}t.$$

If $\int_0^T b(X_t)^2\,\mathrm{d}t/T$ converges in probability to a positive constant V (independent of θ) as $T \to \infty$ under θ_0, then $I_{\theta_0}(\theta) = V$ and we obtain the global non-degeneracy. Moreover, if this family is parameterized as $\theta = \theta(u)$, then

$$I_{u_0}(u) = V[\ddot{\theta}(u)(\theta(u) - \theta(u_0)) + (\dot{\theta}(u))^2].$$

For instance, in the case that $\theta = \sqrt{u}$, $0 < a_1 < u < a_2$, one has

$$I_{u_0}(u) = \frac{V}{4} u^{-3/2} u_0^{1/2},$$

and the global non-degeneracy still holds true.

Next, we consider an M-estimator over Wiener space and present an asymptotic expansion of its distribution function. For each $n \in \boldsymbol{N}$, let (W^n, H^n, P^n) be a (partial) r-dimensional Wiener space: $W^n = W^{(n,1)} \times W^{(n,2)}$, and $P^n = P^{(n,1)} \otimes P^{(n,2)}$, where $(W^{(n,1)}, H^n, P^{(n,1)})$ is a usual Wiener space and $(W^{(n,2)}, \boldsymbol{B}^{(n,2)}, P^{(n,2)})$ is a probability space. Let $\boldsymbol{D}^n_{p,s}$ be the Sobolev space of Wiener functionals on W^n. Although $\boldsymbol{D}^n_{p,s}$ is equipped with a Sobolev norm depending on n, we denote it by $\|\cdot\|_{p,s}$ briefly. Denote by σ_F the Malliavin covariance of a Wiener functional F on W^n. In the following, we suppose that $X_n^\theta \colon W^n \to \mathfrak{X}^n$ is a random variable on (W^n, H^n, P^n) for every $\theta \in \Theta$ and every $n \in \boldsymbol{N}$. As defined in Theorem 1, μ^n_θ is a probability measure on $(\mathfrak{X}^n, \mathfrak{U}^n)$ induced from P^n by X_n^θ, and X_n is a random variable X_n^θ evaluated at $\theta = \theta_0$. Then the M-estimator $\hat{\theta}_n$ for X_n defined in Theorem 1 can be regarded as a Wiener functional $\hat{\theta}_n \circ X_n \colon W^n \to \Theta$. If the leading term $Z_{n,1}$ of the stochastic expansion in Theorem 1 is a terminal random variable of a continuous martingale, we can apply the result of Yoshida [8]. For a martingale $\{M(t)\}_{0 \leqslant t \leqslant T}$, the bracket $\langle M \rangle$ denotes the predictable quadratic variation process of M: moreover we will abbreviate $M(T)$ to M and $\langle M \rangle(T)$ to $\langle M \rangle$, for simplicity.

THEOREM 2. *For any $\theta_0 \in \Theta$, let $\hat{\theta}_n$, $Z_{n,1}$ and $Z_{n,2}$ be Wiener functionals satisfying* (2.1)–(2.4). *Assume that the following conditions hold:*

[A1] $Z_{n,1}, Z_{n,2} \in \cap_{p>1} \boldsymbol{D}^n_{p,4}$ *for any $n \in \boldsymbol{N}$;*
[A2] (i) *there exist continuous martingales $\{Z'_{n,1}(t)\}$ with $Z'_{n,1} \in \cap_{p>1} \boldsymbol{D}^n_{p,4}$ and $\langle Z'_{n,1} \rangle \in \cap_{p>1} \boldsymbol{D}^n_{p,3}$;*
(ii) *for some positive constant τ,*

$$\sup_n \|Z'_{n,1}\|_{p,4} + \sup_n \|r_n^{-1}((\tau I_{\theta_0}(\theta_0))^{-2} \langle Z'_{n,1} \rangle - 1)\|_{p,3} + \\ + \sup_n \|Z_{n,2}\|_{p,4} + \sup_n \|r_n^{-1}(Z_{n,1} - Z'_{n,1})\|_{p,4} < \infty;$$

[A3] *there exists a random vector (Z, ξ, ζ, ρ) such that*

$$((\tau I_{\theta_0}(\theta_0))^{-1} Z'_{n,1},\ r_n^{-1}((\tau I_{\theta_0}(\theta_0))^{-2} \langle Z'_{n,1} \rangle - 1), Z_{n,2}, r_n^{-1}(Z_{n,1} - Z'_{n,1})) \\ \Rightarrow (Z, \xi, \zeta, \rho) \quad \text{in law},$$

and $\partial_z^2(E[\xi | Z = z]\phi(z))$ is bounded integrable, where $\phi(z)$ is the standard normal density.
[A4] *there exists a constant $c > 0$ such that* $\lim_{n\to\infty} P(\sigma_{Z'_{n,1}} < c) = 0$.

Then for any $p > 1$, there exists a constant $C > 0$ and a positive sequence $\epsilon_n = o(r_n)$ such that

$$\Delta_n \leqslant C(1 + \log^+(r_n^{-1}))P(\sigma_{Z'_{n,1}} < c)^{1/p} + \epsilon_n, \tag{2.5}$$

where

$$\Delta_n = \sup_x \left| P[(\tau r_n)^{-1}(\hat{\theta}_n - \theta_0) \leqslant x] - \int_{-\infty}^{x} p_n(z)\,\mathrm{d}z \right|, \tag{2.6}$$

and

$$\begin{aligned} p_n(z) \;=\; & \phi(z) + \tfrac{1}{2} r_n \partial_z^2 (E[\xi | Z = z]\phi(z)) - \\ & - r_n I_{\theta_0}^{-1}(\theta_0)\partial_z \left(E\left[z\zeta - \tfrac{1}{2}\tau L_{\theta_0}(\theta_0)z^2 + \tau^{-1}\rho\,|Z = z\right]\phi(z)\right). \end{aligned} \tag{2.7}$$

Proof. This follows immediately from Theorem 2 of Yoshida [8] by means of the *Delta* method.

3. Diffusion Process

As in the previous section, suppose that the parameter space Θ is a bounded interval in $\boldsymbol{R}$, and denote by θ_0 the true value of the unknown parameter in Θ. Let b be a real-valued function on $\boldsymbol{R} \times \Theta$. In this section, we consider a one-dimensional stationary ergodic diffusion process $X^\theta = (X_t^\theta : t \in \boldsymbol{R}_+)$ defined by

$$\mathrm{d}X_t = b_\theta(X_t)\,\mathrm{d}t + \mathrm{d}w_t,$$

with stationary distribution ν_θ given by

$$\nu_\theta(\mathrm{d}x) = \frac{n_\theta(x)}{\int_{-\infty}^{\infty} n_\theta(u)\,\mathrm{d}u}\,\mathrm{d}x,$$

where

$$n_\theta(x) = \exp\left(2\int_0^x b_\theta(u)\,\mathrm{d}u\right).$$

Assume that $b_\theta \in C^{4,4}(\boldsymbol{R} \times \Theta)$, $\sup_x \partial_x b_\theta(x) < 0$, $\partial_x b_\theta(x)$ is bounded in x for each θ and there exist positive constants m_1 and C_1 such that

$$\sup_{\theta\in\Theta} |\delta^i \partial_x^j b_\theta(x)| \leqslant C_1(1 + |x|)^{m_1}, \tag{3.1}$$

for any $x \in \boldsymbol{R}$ and for any $i, j = 0, \ldots, 4$. It is then immediately shown that for any $p > 1$

$$\sup_{t\in\boldsymbol{R}_+} \|X_t^\theta\|_p < \infty. \tag{3.2}$$

Note that in this model the random variables X_t^θ are defined on the same Wiener space, while in Section 2 the underlying probability spaces (W^n, H^n, P^n) of X_n^θ may be distinct from each other. For simplicity, we omit the symbol θ_0 from the notations of functions of θ when they are evaluated at $\theta = \theta_0$, for example, $X = X^{\theta_0}$, $\nu = \nu_{\theta_0}$.

First, we consider an estimating function given by

$$g_T(X, \theta) = f_\theta(X_T) - f_\theta(X_0) - \int_0^T A_\theta f_\theta(X_t)\,\mathrm{d}t,$$

where f is a given real-valued function on $\boldsymbol{R} \times \Theta$, and A_θ is a differential operator defined by

$$A_\theta = \tfrac{1}{2}\partial_x^2 + b_\theta(x)\partial_x.$$

This estimating function was treated in Lánska [5] and also in Yoshida [7], and it has an advantage because it is robust, that is, in the sense that g_T is continuous in X with respect to the supremum norm over each compact set. If $\partial_x f_\theta = \delta b_\theta$, it follows from Itô's formula that

$$\begin{aligned} g_T(X, \theta) &= \int_0^T \partial_x f_\theta(X_t)\,\mathrm{d}w_t + \int_0^T (\partial_x f_\theta \cdot b_{\theta_0})(X_t)\,\mathrm{d}t - \\ &\quad - \int_0^T (\partial_x f_\theta \cdot b_\theta)(X_t)\,\mathrm{d}t \\ &= \int_0^T \delta_{b_\theta}(X_t)\,\mathrm{d}X_t - \int_0^T (\delta b_\theta \cdot b_\theta)(X_t)\,\mathrm{d}t, \end{aligned}$$

and hence that the M-estimator coincides with the maximum likelihood estimator in this case. In the following, we will verify the conditions in Theorem 1 and Theorem 2 for this estimating function g_T, and will present an asymptotic expansion of the distribution of the M-estimator corresponding to $g_T(X, \theta) = 0$. Suppose that $f_\theta \in C^{5,4}(\boldsymbol{R} \times \Theta)$ and there exist positive constants m_2 and C_2 such that

$$\sup_{\theta\in\Theta} |\delta^i \partial_x^j f_\theta(x)| \leqslant C_2(1 + |x|^{m_2}), \tag{3.3}$$

for any $x \in \boldsymbol{R}$ and for any $i = 0, \ldots, 4$, $j = 0, \ldots, 5$. Denote $\delta^i(A_\theta f_\theta)$ by $k_{\theta,i}$ for each $i = 0, \ldots, 4$. Then $\sup_\theta |k_{\theta,i}|$ have at most polynomial growth order. Define $I_{\theta_0}(\theta)$ and $L_{\theta_0}(\theta)$ by

$$I_{\theta_0}(\theta) = \nu(k_{\theta,1}) \quad \text{and} \quad L_{\theta_0}(\theta) = \nu(k_{\theta,2}), \tag{3.4}$$

respectively. Let $Z_{T,1}$ and $Z_{T,2}$ be Wiener functionals defined by

$$Z_{T,1} = T^{-1/2} g_T(X, \theta_0) \tag{3.5}$$

and

$$Z_{T,2} = T^{1/2}(T^{-1}\delta g_T(X, \theta_0) + I_{\theta_0}(\theta_0)). \tag{3.6}$$

Assume either that

$$\nu((\partial_x f_{\theta_0})^2) > 0, \qquad \inf_{\theta\in\tilde{\Theta}} I_{\theta_0}(\theta) > 0 \tag{3.7}$$

for an open interval $\tilde{\Theta} \subset \Theta^0$ including θ_0, or that

$$\nu((\partial_x f_{\theta_0})^2) > 0, \qquad I_{\theta_0}(\theta_0) > 0. \tag{3.8}$$

Denote by τ a constant $\sqrt{\nu((\partial_x f_{\theta_0})^2)/I_{\theta_0}(\theta_0)}$. For continuous function $h\colon \boldsymbol{R} \to \boldsymbol{R}$, define $G_h\colon \boldsymbol{R} \to \boldsymbol{R}$ by

$$G_h(x) = -\int_0^x \frac{1}{n(y)} \int_y^\infty 2n(u)h(u)\,\mathrm{d}u\,\mathrm{d}y,$$

if

$$\int_0^\infty n(u)|h(u)|\,\mathrm{d}u < \infty.$$

Let

$$Z_{T,1}(t) = \frac{1}{\sqrt{T}} \int_0^t \partial_x f_{\theta_0}(X_s)\,\mathrm{d}w_s,$$

then we see from Itô's formula that $Z_{T,1} = Z_{T,1}(T)$. Therefore we will regard $Z_{T,1}$ itself as the terminal random variable $Z'_{T,1}$ given in Assumptions [A2]–[A4] of Theorem 2. Since $A_{\theta_0}G_h = h$ for any continuous function h with $\nu(h) = 0$, we have that

$$\begin{aligned} Z_{T,2} &= T^{-1/2}(\delta f_{\theta_0}(X_T) - \delta f_{\theta_0}(X_0)) - T^{-1/2}\int_0^T (k_{\theta_0,1}(X_t) - \nu(k_{\theta_0,1}))\,\mathrm{d}t \\ &= o_p(1) - T^{-1/2}(G_{\bar{k}}(X_T) - G_{\bar{k}}(X_0)) + T^{-1/2}\int_0^T \partial_x G_{\bar{k}}(X_t)\,\mathrm{d}w_t \\ &= o_p(1) + T^{-1/2}\int_0^T \partial_x G_{\bar{k}}(X_t)\,\mathrm{d}w_t, \end{aligned}$$

where $\bar{k} = k_{\theta_0,1} - \nu(k_{\theta_0,1})$. Similarly, we see that

$$\begin{aligned} \sqrt{T}((\tau I_{\theta_0}(\theta_0))^{-2}\langle Z_{T,1}\rangle_T - 1) &= T^{-1/2}\int_0^T \left[\frac{(\partial_x f_{\theta_0})^2(X_t)}{\nu((\partial_x f_{\theta_0})^2)} - 1\right]\mathrm{d}t \\ &= o_p(1) - T^{-1/2}\int_0^T \partial_x G_{\bar{l}}(X_t)\,\mathrm{d}w_t, \end{aligned}$$

where

$$\bar{l}(x) = \frac{(\partial_x f_{\theta_0})^2(x)}{\nu((\partial_x f_{\theta_0})^2)} - 1.$$

It is easy to show from Lemma 6 of Yoshida [10] that $\partial_x G_{\bar{k}}$ and $\partial_x G_{\bar{l}}$ have at most polynomial growth order. Therefore, it follows from the martingale central limit theorem that

$$\begin{aligned}&((\tau I_{\theta_0}(\theta_0))^{-1} Z_{T,1}, \sqrt{T}((\tau I_{\theta_0}(\theta_0))^{-2}\langle Z_{T,1}\rangle_T - 1),\ Z_{T,2})\\ &\quad \Rightarrow N(0,\Sigma) \quad \text{as } T \to \infty,\end{aligned}$$

where $\Sigma = (\Sigma_{i,j})$ is a symmetric matrix given by

$$\Sigma = \begin{bmatrix} 1 & -\dfrac{\nu(\partial_x f_{\theta_0}\cdot\partial_x G_{\bar{l}})}{\sqrt{\nu((\partial_x f_{\theta_0})^2)}} & \dfrac{\nu(\partial_x f_{\theta_0}\cdot\partial_x G_{\bar{k}})}{\sqrt{\nu((\partial_x f_{\theta_0})^2)}} \\ & \nu((\partial_x G_{\bar{l}})^2) & -\nu(\partial_x G_{\bar{l}}\cdot\partial_x G_{\bar{k}}) \\ \text{sym} & & \nu((\partial_x G_{\bar{k}})^2) \end{bmatrix}.$$

This implies that Assumption [A3] of Theorem 2 holds with $Z_{T,1} = Z'_{T,1}$ and $\rho = 0$. Assumptions [A1], [A2] and [A4] can also easily be verified from Lemmas 2, 4, and 5 of Yoshida [10] and from Lemma 7 of Yoshida [10], respectively.

Let us verify the conditions [C1]–[C4] in Theorem 1. From the Burkholder–Davis–Gundy inequality, we have that for some constant $c_{p_1} > 0$ depending only on $p_1 > 1$,

$$\begin{aligned}\|T^{-1/2} g_T(X,\theta_0)\|_{p_1} &= T^{-1/2}\left\{E\left[\left(\int_0^T \partial_x f_{\theta_0}(X_t)\,\mathrm{d}w_t\right)^{p_1}\right]\right\}^{1/p_1}\\ &\leqslant c_{p_1} T^{-1/2}\left\{E\left[\left(\int_0^T (\partial_x f_{\theta_0}(X_t))^2 \mathrm{d}t\right)^{p_1/2}\right]\right\}^{1/p_1}.\end{aligned}$$

Owing to Jensen's inequality and (3.3), we see that

$$\|T^{-1/2} g_T(X,\theta_0)\|_{p_1} \leqslant c_{p_1}\|\partial_x f_{\theta_0}(X_0)\|_{2p_1} < \infty.$$

Hence [C1] of Theorem 1 holds. In the same fashion, we see that for any $\gamma > 0$ and any $p_2 > 1$ there exists a positive constant $c_{p_2} > 0$ depending only on p_2 such that

$$\left\| T^{\gamma/2-1}\int_0^T \partial_x G_{\bar{k}_\theta(x)}(X_T)\,\mathrm{d}w_t\right\|_{p_2} \leqslant c_{p_2} T^{(\gamma-1)/2}\|\partial_x G_{\bar{k}_\theta(x)}(X_0)\|_{2p_2},$$

where $\bar{k}_\theta(x) = k_{\theta,1}(x) - \nu(k_{\theta,1})$. Therefore we obtain from the stationarity of X that for $\gamma > 0$ and $p_2 > 1$,

$$\begin{aligned}&\|T^{\gamma/2}(T^{-1}\delta g_T(X,\theta) + I_{\theta_0}(\theta))\|_{p_2}\\ &\quad\leqslant 2T^{\gamma/2-1}\|\delta f_\theta(X_0)\|_{p_2} + 2T^{\gamma/2-1}\|G_{\bar{k}_\theta(x)}(X_0)\|_{p_2} +\\ &\qquad + c_{p_2} T^{(\gamma-1)/2}\|\partial_x G_{\bar{k}_\theta(x)}(X_0)\|_{2p_2}.\end{aligned}$$

Applying the same argument as in the proof of Lemma 6 in Yoshida [10] together with (3.1) and (3.3), we see that there exists a positive constant C_3 and a positive integer m_3 such that

$$\sup_{\theta\in\Theta} |G_{\bar{k}_\theta(x)}(x)| \leqslant C_3(1+|x|^{m_3}) \quad \text{and} \quad \sup_{\theta\in\Theta} |\partial_x G_{\bar{k}_\theta(x)}(x)| \leqslant C_3(1+|x|^{m_3}).$$

Thus, it follows that for $\gamma \in (0,1)$, $p_2 > 1$ and $\epsilon > 0$,

$$\sup_{T>\epsilon,\theta\in\Theta^\circ} \|T^{\gamma/2}(T^{-1}\delta g_T(X,\theta) + I_{\theta_0}(\theta))\|_{p_2} < \infty.$$

Combining this and (3.7) (or (3.8)), we see that [C2] of Theorem 1 holds. In exactly the same way, it follows that for $\gamma \in (0,1)$, $p_2 > 1$ and $\epsilon > 0$,

$$\sup_{T>\epsilon,\theta\in\Theta^\circ} \|T^{\gamma/2}(T^{-1}\delta^2 g_T(X,\theta) + L_{\theta_0}(\theta))\|_{p_2} < \infty,$$

and hence that [C3] of Theorem 1 holds. Furthermore, it follows from Sobolev's inequality that for some constant $C_{\Theta^\circ} > 0$

$$\begin{aligned}
&\| \sup_{\theta\in\Theta^\circ} |T^{-1}\delta^3 g_T(X,\theta)| \, \|_{p_3} \\
&\quad\leqslant C_{\Theta^\circ}\left(\int_{\Theta^\circ} \|T^{-1}\delta^3 g_T(X,\theta)\|_{p_3}^{p_3}\,\mathrm{d}\theta + \int_{\Theta^\circ} \|T^{-1}\delta^4 g_T(X,\theta)\|_{p_3}^{p_3}\,\mathrm{d}\theta\right)^{1/p_3} \\
&\quad\leqslant C_{\Theta^\circ}|\Theta^\circ|^{1/p_3}\left(\left(\sup_{\theta\in\Theta^\circ} \|T^{-1}\delta^3 g_T(X,\theta)\|_{p_3}\right)^{p_3} + \right. \\
&\qquad \left. + \left(\sup_{\theta\in\Theta^\circ} \|T^{-1}\delta^4 g_T(X,\theta)\|_{p_3}\right)^{p_3}\right)^{1/p_3}.
\end{aligned}$$

Since

$$\begin{aligned}
\|T^{-1}\delta^i g_T(X,\theta)\|_{p_3} \;\leqslant\; & 2T^{-1}\|\delta^i f_\theta(X_0)\|_{p_3} + 2T^{-1}\|G_{k_{\theta,i}-\nu(k_{\theta,i})}(X_0)\|_{p_3} + \\
& + c_{p_3}T^{-1/2}\|\partial_x G_{k_{\theta,i}-\nu(k_{\theta,i})}(X_0)\|_{p_3} + \nu(k_{\theta,i}),
\end{aligned}$$

for each $i = 3, 4$ and for some constant $c_{p_3} > 0$, it is also seen that for $p_3 > 1$ and $\epsilon > 0$,

$$\sup_{T>\epsilon} \| \sup_{\theta\in\Theta^\circ} |T^{-1}\delta^3 g_T(X,\theta)| \, \|_{p_3} < \infty.$$

Thus we see that the conditions [C1]–[C4] of Theorem 1 holds. Consequently, the verification of the assumptions in Theorems 1 and 2 is completed. In this way, we have:

THEOREM 3. *Assume that (3.1), (3.3) and (3.7) (or (3.8)) hold true. Then there exists an M-estimator $\hat{\theta}_T$ corresponding to the estimating function g_T, and the asymptotic* *expansion of its distribution function is given by*

$$\begin{aligned} P(\sqrt{T}\tau^{-1}(\hat{\theta}_T - \theta_0) \leqslant x) \;=\; & \Phi(x) + \frac{1}{2\sqrt{T}}\Sigma_{1,2}(1-x^2)\phi(x) - \\ & - \frac{1}{\sqrt{T}} I_{\theta_0}^{-1}(\theta_0)\left(\Sigma_{1,3} - \frac{1}{2}\tau L_{\theta_0}(\theta_0)\right) \times \\ & \times x^2\phi(x) + o(1/\sqrt{T}). \end{aligned}$$

We give the results of the numerical studies of this asymptotic expansion in Figures 1 and 2. In the former, X^θ is the diffusion process corresponding to $b_\theta(x) = -\theta x$, the Ornstein–Uhlenbeck process, and in the latter, X^θ is the diffusion process corresponding to $b_\theta(x) = -\theta(x^3 + x)$. In each figure, M-estimator is the maximum likelihood estimator, and asymptotic expansions given above are compared with the true distribution functions obtained by Monte-Carlo simulations (100 000 repetitions).

Next we consider a slightly different estimating function $\tilde{g}_T$ defined by

$$\tilde{g}_T(X,\theta) = -\int_0^T A_\theta f_\theta(X_t)\,\mathrm{d}t.$$

Taking the same steps as for g_T, one can easily show that conditions [C1]–[C4] hold true for this $\tilde{g}_T$ with I_{θ_0} and L_{θ_0} defined by (3.4), and that there exists an M-estimator $\tilde{\theta}_T$ satisfying (2.1)–(2.4). Define $\tilde{Z}_{T,1}$ and $\tilde{Z}_{T,2}$ by

$$\tilde{Z}_{T,1} = T^{-1/2}\tilde{g}_T(X,\theta_0) \quad \text{and} \quad \tilde{Z}_{T,2} = T^{1/2}(T^{-1}\delta\tilde{g}_T(X,\theta_0) + I_{\theta_0}(\theta_0)),$$

respectively. Then we can show that $\tilde{Z}_{T,1}$ and $\tilde{Z}_{T,2} \in \cap_{p>1}\boldsymbol{D}^n_{p,4}$ in the same fashion as for $Z_{T,1}$ and $Z_{T,2}$ given by (3.5) and (3.6), respectively. In this case, $\tilde{Z}_{T,1}$ itself is not a terminal random variable of a martingale, but we see that the terminal random variable $Z_{T,1}$ satisfies assumptions [A2]–[A4] of Theorem 2 for $\tilde{Z}_{T,1}$ and $\tilde{Z}_{T,2}$ as follows. From (3.2), (3.3) and the stationarity of X, it is easy to show that for any $p > 1$ and any $\epsilon > 0$,

$$\sup_{T>\epsilon} \|\sqrt{T}(\tilde{Z}_{T,1} - Z_{T,1})\|_{p,4} = \sup_{T>\epsilon} \|f_{\theta_0}(X_T) - f_{\theta_0}(X_0)\|_{p,4} < \infty.$$

We can also show from (3.1), (3.3), and Lemmas 2, 4, 6 of [10] that for any $p > 1$ and any $\epsilon > 0$,

$$\begin{aligned} \sup_{T>\epsilon} \|\tilde{Z}_{T,2}\|_{p,4} \;=\; & \sup_{T>\epsilon} \|T^{-1/2}\int_0^T \bar{k}(X_t)\,\mathrm{d}t\|_{p,4} \\ =\; & \sup_{T>\epsilon} \|T^{-1/2}(G_{\bar{k}}(X_T) - G_{\bar{k}}(X_0))\|_{p,4} + \\ & + \sup_{T>\epsilon} \|T^{-1/2}\int_0^T \partial_x G_{\bar{k}}(X_t)\,\mathrm{d}w_t\|_{p,4} < \infty. \end{aligned}$$

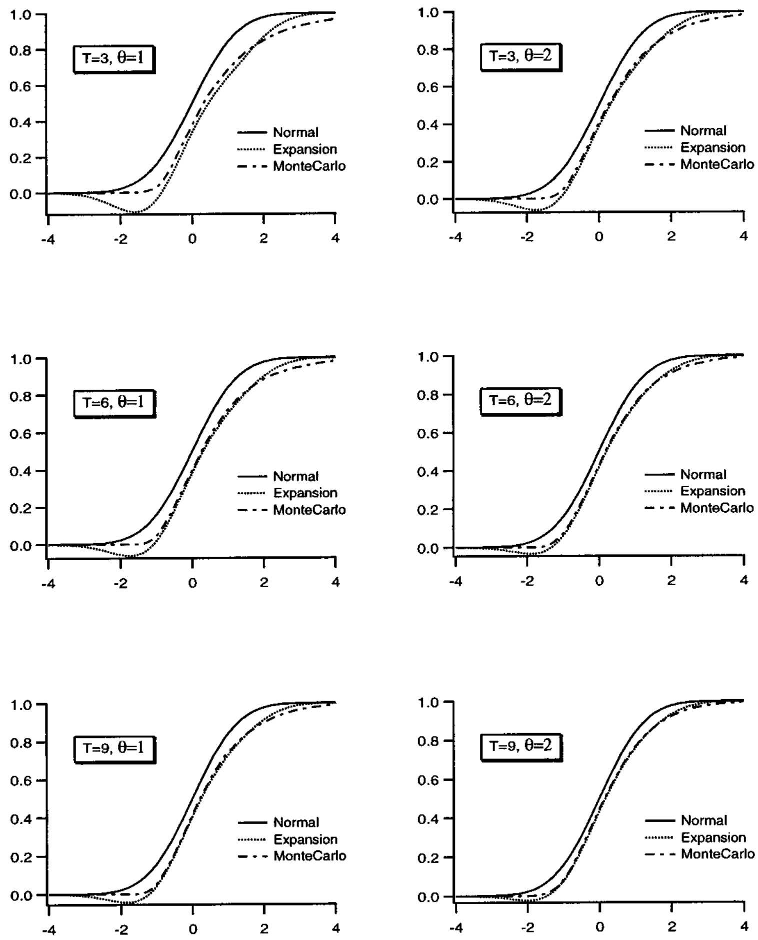

Figure 1. Distribution functions of M-estimator; $b_\theta(x) = -\theta x$.

Therefore it suffices to show [A3] of Theorem 3. Let

$$V_{T,1} = \frac{1}{\sqrt{T}\tau I_{\theta_0}} \int_{T^{1/2}}^{T-T^{1/2}} \partial_x f_{\theta_0}(X_t)\,\mathrm{d}w_t,$$

$$V^*_{T,1} = -\frac{1}{\sqrt{T}} \int_{T^{1/2}}^{T-T^{1/2}} \partial_x G_{\bar{l}}(X_t)\,\mathrm{d}w_t,$$

and

$$V_{T,2} = \frac{1}{\sqrt{T}} \int_{T^{1/2}}^{T-T^{1/2}} \partial_x G_{\bar{k}}(X_t)\,\mathrm{d}w_t.$$

It then follows that the asymptotic distribution of the random vector

$$((\tau I_{\theta_0})^{-1} Z_{T,1}, \sqrt{T}((\tau I_{\theta_0})^{-2}\langle Z_{T,1}\rangle_T - 1), \tilde{Z}_{T,2}, f_{\theta_0}(X_T), f_{\theta_0}(X_0))$$

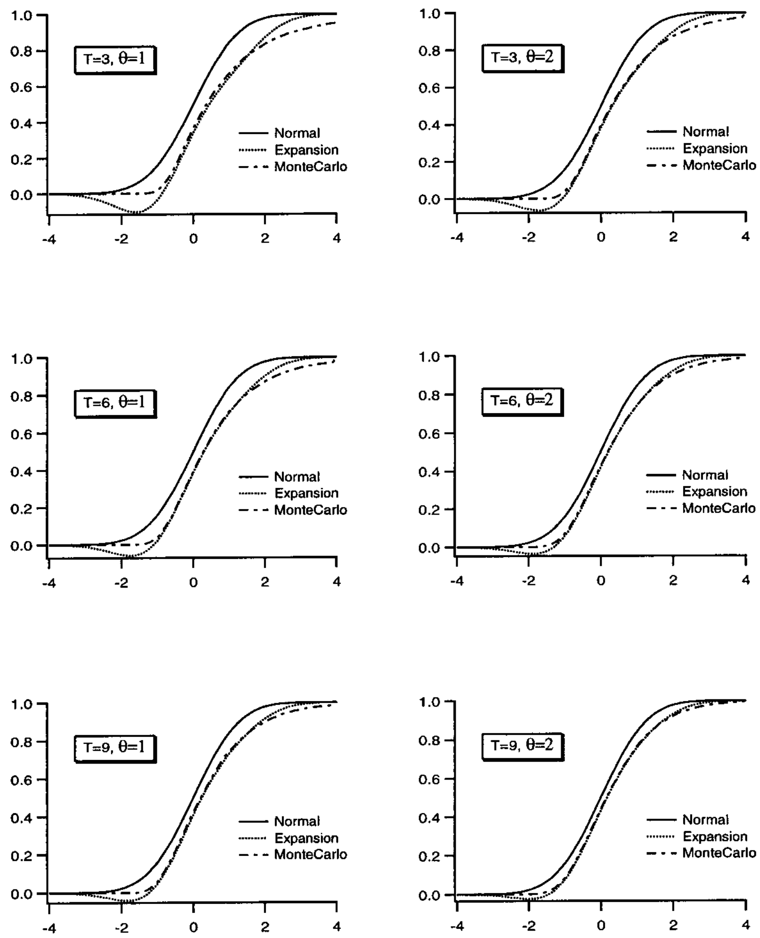

Figure 2. Distribution functions of M-estimator; $b_\theta(x) = -\theta(x^3 + x)$.

is equal to the asymptotic one of $(V_{T,1}, V^*_{T,1}, V_{T,2}, f_{\theta_0}(X_T), f_{\theta_0}(X_0))$. Since X_T has a strong mixing property (cf. Doukhan [1]; Veretennikov [6]; Kusuoka and Yoshida [2]), we obtain from covariance inequality that for any $(u_1, \ldots, u_5) \in \boldsymbol{R}^5$,

$$
\begin{aligned}
&E[\exp(\sqrt{-1}(u_1 V_{T,1} + u_2 V^*_{T,1} + u_3 V_{T,2} + u_4 f_{\theta_0}(X_T) + u_5 f_{\theta_0}(X_0)))] \\
&\quad \to E[\exp(-\tfrac{1}{2}(u_1, u_2, u_3)\Sigma(u_1, u_2, u_3)')]\times \\
&\quad \times E[\exp(\sqrt{-1}\, u_4 f_{\theta_0}(X_0))]E[\exp(\sqrt{-1}\, u_5 f_{\theta_0}(X_0))],
\end{aligned}
$$

as $T \to \infty$. Thus we see from Theorem 2 that the asymptotic expansion of $\tilde{\theta}_T$ coincides with that of $\hat{\theta}_T$.

THEOREM 4. *Assume that (3.1), (3.3) and (3.7) (or(3.8)) hold true. Then there exists an M-estimator $\tilde{\theta}_T$ corresponding to the estimating function $\tilde{g}_T$, and the asymptotic* *expansion of its distribution function is given by*

$$\begin{aligned} P(\sqrt{T}\tau^{-1}(\tilde{\theta}_T-\theta_0)\leqslant x) \;=\; & \Phi(x)+\frac{1}{2\sqrt{T}}\Sigma_{1,2}(1-x^2)\phi(x)- \\ & -\frac{1}{\sqrt{T}}I_{\theta_0}^{-1}(\theta_0)\left(\Sigma_{1,3}-\frac{1}{2}\tau L_{\theta_0}(\theta_0)\right)x^2\phi(x)+ \\ & +o(1/\sqrt{T}). \end{aligned}$$

4. Proof of Theorems

We first present some preliminary lemmas.

LEMMA 1. *Assume that* [C2]–[C4] *in Theorem 1 hold for some* $\gamma > 0,\ p_2 > 1$ *and* $p_3 > 1$. *Then* $I_{\theta_0} \in C^1(\Theta^\circ)$ *and* $\delta I_{\theta_0}(\theta) = L_{\theta_0}(\theta)$.

Proof. Setting $p = \min(p_2, p_3)$, then for any $\theta_1, \theta_2 \in \Theta^\circ$,

$$\begin{aligned} &|L_{\theta_0}(\theta_1)-L_{\theta_0}(\theta_2)| \\ &\quad\leqslant \sum_{i=1}^{2}\|r_n^2\delta^2 g_n(X_n,\theta_i)+L_{\theta_0}(\theta_i)\|_p+\left\|\int_{\theta_2}^{\theta_1}|r_n^2\delta^3 g_n(X_n,\theta)|\,\mathrm{d}\theta\right\|_p \\ &\quad\leqslant r_n^\gamma\sum_{i=1}^{2}\|r_n^{-\gamma}(r_n^2\delta^2 g_n(X_n,\theta_i))+\ L_{\theta_0}(\theta_i)\|_p+ \\ &\qquad+|\theta_1-\theta_2|\sup_{\theta\in\Theta^\circ}\|r_n^2\delta^3 g_n(X_n,\theta)\|_p. \end{aligned}$$

From [C3] and [C4] in Theorem 1, there exists a constant C independent of θ_1 and θ_2 such that

$$|L_{\theta_0}(\theta_1)-L_{\theta_0}(\theta_2)|\leqslant C|\theta_1-\theta_2|.$$

This implies that $L_{\theta_0}(\theta)$ is continuous on Θ°. Furthermore, for any $\theta_1, \theta_2 \in \Theta^\circ$,

$$\begin{aligned} &\left|I_{\theta_0}(\theta_1)-I_{\theta_0}(\theta_2)-\int_{\theta_2}^{\theta_1}L_{\theta_0}(\theta)\,\mathrm{d}\theta\right| \\ &\quad\leqslant \sum_{i=1}^{2}\|r_n^2\delta g_n(X_n,\theta_i)+I_{\theta_0}(\theta_i)\|_{p_2}+ \\ &\quad+\left\|\int_{\theta_2}^{\theta_1}|r_n^2\delta^2 g_n(X_n,\theta)+L_{\theta_0}(\theta)|\,\mathrm{d}\theta\right\|_{p_2} \\ &\quad\leqslant r_n^\gamma\sum_{i=1}^{2}\|r_n^{-\gamma}(r_n^2\delta g_n(X_n,\theta_i)+I_{\theta_0}(\theta_i))\|_{p_2}+ \\ &\qquad+|\theta_1-\theta_2|r_n^\gamma\sup_{\theta\in\Theta^\circ}\|r_n^{-\gamma}(r_n^2\delta^2 g_n(X_n,\theta)+L_{\theta_0}(\theta))\|_{p_2}. \end{aligned}$$

Therefore it follows from [C2] and [C3] in Theorem 1 that

$$I_{\theta_0}(\theta_1) - I_{\theta_0}(\theta_2) = \int_{\theta_2}^{\theta_1} L_{\theta_0}(\theta)\,\mathrm{d}\theta$$

Since $L_{\theta_0}(\theta)$ is continuous on Θ°, it is seen that $\delta I_{\theta_0}(\theta) = L_{\theta_0}(\theta)$. □

LEMMA 2. *Let* $\gamma > 0$ *and* $p_2 > 1$ *with* $\gamma p_2 > 1$. *Assume that* [C2]–[C4] *in Theorem 1 hold. Then there exists a constant C independent of* n *such that*

$$P[\inf_{\theta\in\tilde{\Theta}}(-r_n^2\delta g_n(X_n,\theta)) > C] = 1 - o(r_n).$$

Proof. From [C2] in Theorem 1, there exists a constant C independent of n such that $\inf_{\theta\in\tilde{\Theta}} I_{\theta_0}(\theta) > 2C$. Therefore it follows that

$$\begin{aligned}
&P\left[\inf_{\theta\in\tilde{\Theta}}(-r_n^2\delta g_n(X_n,\theta)) \leqslant C\right]\\
&\quad\leqslant P\left[\inf_{\theta\in\tilde{\Theta}}(-r_n^2\delta g_n(X_n,\theta) - I_{\theta_0}(\theta)) + \inf_{\theta\in\tilde{\Theta}} I_{\theta_0}(\theta) \leqslant C\right]\\
&\quad\leqslant P\left[\sup_{\theta\in\tilde{\Theta}}|r_n^2\delta g_n(X_n,\theta) + I_{\theta_0}(\theta)| \geqslant C\right].
\end{aligned}$$

From Sobolev's inequality and Lemma 1, it is seen that for any $X_n \in \mathfrak{X}^n$, there exists a constant $C_{\tilde{\Theta}} > 0$ independent of n such that

$$\begin{aligned}
&\sup_{\theta\in\tilde{\Theta}}|r_n^2\delta g_n(X_n,\theta) + I_{\theta_0}(\theta)|\\
&\quad\leqslant C_{\tilde{\Theta}}\left(\int_{\tilde{\Theta}}|r_n^2\delta g_n(X_n,\theta) + I_{\theta_0}(\theta)|^{p_2}\,\mathrm{d}\theta +\right.\\
&\qquad\left. + \int_{\tilde{\Theta}}|r_n^2\delta^2 g_n(X_n,\theta) + L_{\theta_0}(\theta)|^{p_2}\,\mathrm{d}\theta\right)^{1/p_2}.
\end{aligned}$$

This shows that

$$\begin{aligned}
&P\left[\sup_{\theta\in\tilde{\Theta}}|r_n^2\delta g_n(X_n,\theta) + I_{\theta_0}(\theta)| \geqslant C\right]\\
&\quad\leqslant P\left[\int_{\tilde{\Theta}}|r_n^2\delta g_n(X_n,\theta) + I_{\theta_0}(\theta)|^{p_2}\,\mathrm{d}\theta \geqslant \frac{1}{2}\left(\frac{C}{C_{\tilde{\Theta}}}\right)^{p_2}\right] +\\
&\qquad + \left[\int_{\tilde{\Theta}}|r_n^2\delta^2 g_n(X_n,\theta) + L_{\theta_0}(\theta)|^{p_2}\,\mathrm{d}\theta \geqslant \frac{1}{2}\left(\frac{C}{C_{\tilde{\Theta}}}\right)^{p_2}\right]\\
&\quad\leqslant 2\left(\frac{C_{\tilde{\Theta}}}{C}\right)^{p_2}\left(E\left[\int_{\tilde{\Theta}}|r_n^2\delta g_n(X_n,\theta) + I_{\theta_0}(\theta)|^{p_2}\,\mathrm{d}\theta\right] +\right.\\
&\qquad\left. + E\left[\int_{\tilde{\Theta}}|r_n^2\delta^2 g_n(X_n,\theta) + L_{\theta_0}(\theta)|^{p_2}\,\mathrm{d}\theta\right]\right)
\end{aligned}$$

$$\begin{aligned}&\leqslant 2\left(\frac{C_{\tilde{\Theta}}}{C}\right)^{p_2}|\tilde{\Theta}|r_n^{\gamma p_2}\left(\sup_{\theta\in\tilde{\Theta}}\|r_n^{-\gamma}(r_n^2\delta g_n(X_n,\theta)+I_{\theta_0}(\theta))\|_{p_2}^{p_2}+\right.\\&\left.\quad+\sup_{\theta\in\tilde{\Theta}}\|r_n^{-\gamma}(r_n^2\delta^2 g_n(X_n,\theta)+L_{\theta_0}(\theta))\|_{p_2}^{p_2}\right).\end{aligned}$$

Since $\gamma p_2 > 1$, it follows that

$$P\left[\inf_{\theta\in\tilde{\Theta}}(-r_n^2\delta g_n(X_n,\theta))\leqslant C\right]=o(r_n),$$

which completes the proof. □

In the following lemma, $\hat{\theta}_n$ stands for a generic point in Θ° not necessarily the M-estimate.

LEMMA 3. *For any sequence $\hat{\theta}_n \in \Theta^\circ$, R defined by (2.3) can be rewritten as*

$$\begin{aligned}R &= I_{\theta_0}^{-1}(\theta_0)(r_nZ_{n,2}R_3-\tfrac{1}{2}r_nL_{\theta_0}(\theta_0)(2I_{\theta_0}^{-1}(\theta_0)Z_{n,1}R_3+R_3^2)\ +\\&\quad+\tfrac{1}{2}r_n^{-1}R_4(\hat{\theta}_n-\theta_0)^2+r_nR_1-r_ng_n(X_n,\hat{\theta}_n)),\end{aligned}$$

where

$$\begin{aligned}R_1 &= \frac{1}{2}\int_0^1(1-t)^2\delta^3 g_n(X_n,\theta_0+t(\hat{\theta}_n-\theta_0))\,\mathrm{d}t(\hat{\theta}_n-\theta_0)^3,\\R_2 &= \tfrac{1}{2}r_n\delta^2 g_n(X_n,\theta_0)(\hat{\theta}_n-\theta_0)^2+r_nR_1,\\R_3 &= I_{\theta_0}^{-1}(\theta_0)(Z_{n,2}(\hat{\theta}_n-\theta_0)+R_2-r_ng_n(X_n,\hat{\theta}_n)),\end{aligned}$$

and

$$R_4=r_n^2\delta^2 g_n(X_n,\theta_0)+L_{\theta_0}(\theta_0).$$

Proof. Since $g_n(X_n,\theta)\in C^3(\Theta^\circ\to\boldsymbol{R})$ for any $X_n\in\mathfrak{X}^n$, it follows from Taylor's formula that

$$\begin{aligned}g_n(X_n,\hat{\theta}_n)-g_n(X_n,\theta_0) &= \delta g_n(X_n,\theta_0)(\hat{\theta}_n-\theta_0)+\\&\quad+\tfrac{1}{2}\delta^2 g_n(X_n,\theta_0)(\hat{\theta}_n-\theta_0)^2+R_1.\end{aligned}\tag{4.9}$$

Since

$$r_n^2\delta g_n(X_n,\theta_0)=r_nZ_{n,2}-I_{\theta_0}(\theta_0),\qquad Z_{n,1}=r_ng_n(X_n,\theta_0),$$

it follows from (4.9) that

$$\begin{aligned}r_ng_n(X_n,\hat{\theta}_n)-Z_{n,1} &= (r_nZ_{n,2}-I_{\theta_0}(\theta_0))r_n^{-1}(\hat{\theta}_n-\theta_0)+\\&\quad+\tfrac{1}{2}r_n\delta^2 g_n(X_n,\theta_0)(\hat{\theta}_n-\theta_0)^2+r_nR_1\\&= -I_{\theta_0}(\theta_0)r_n^{-1}(\hat{\theta}_n-\theta_0)+Z_{n,2}(\hat{\theta}_n-\theta_0)+R_2.\end{aligned}$$

Hence we obtain that

$$r_n^{-1}(\hat{\theta}_n - \theta_0) = I_{\theta_0}^{-1}(\theta_0) Z_{n,1} + R_3. \tag{4.10}$$

Moreover, we have from (4.9) and (4.10) that

$$\begin{aligned} r_n g_n(X_n, \hat{\theta}_n) - Z_{n,1} &= (r_n Z_{n,2} - I_{\theta_0}(\theta_0)) r_n^{-1}(\hat{\theta}_n - \theta_0) + \\ &\quad + \tfrac{1}{2} r_n (R_4 - L_{\theta_0}(\theta_0))(r_n^{-1}(\hat{\theta}_n - \theta_0))^2 + r_n R_1 \\ &= -I_{\theta_0}(\theta_0) r_n^{-1}(\hat{\theta}_n - \theta_0) + r_n Z_{n,2}(I_{\theta_0}^{-1}(\theta_0) Z_{n,1} + R_3) - \\ &\quad - \tfrac{1}{2} r_n L_{\theta_0}(\theta_0)(I_{\theta_0}^{-1}(\theta_0) Z_{n,1} + R_3)^2 + \\ &\quad + \tfrac{1}{2} r_n^{-1} R_4 (\hat{\theta}_n - \theta_0)^2 + r_n R_1 \\ &= -I_{\theta_0}(\theta_0) r_n^{-1}(\hat{\theta}_n - \theta_0) + r_n I_{\theta_0}^{-1}(\theta_0) Z_{n,1} Z_{n,2} - \\ &\quad - \tfrac{1}{2} r_n L_{\theta_0}(\theta_0) I_{\theta_0}^{-2}(\theta_0) Z_{n,1}^2 + \\ &\quad + r_n Z_{n,2} R_3 - \tfrac{1}{2} r_n L_{\theta_0}(\theta_0)(2 I_{\theta_0}^{-1}(\theta_0) Z_{n,1} R_3 + R_3^2) + \\ &\quad + \tfrac{1}{2} r_n^{-1} R_4 (\hat{\theta}_n - \theta_0)^2 + r_n R_1. \end{aligned}$$

Therefore the proof is complete. □

Proof of Theorem 1. For some $C > 0$, let $\mathfrak{X}_0^n$ be the subset of $\mathfrak{X}^n$ defined by

$$\mathfrak{X}_0^n :\, = \{X_n \in \mathfrak{X}^n | \inf_{\theta \in \tilde{\Theta}} (-r_n^2 \delta g_n(X_n, \theta)) > C, \qquad |r_n^{2-\gamma} g_n(X_n, \theta_0)| < C\},$$

and let $\Gamma_\gamma(u)$ be a function on $[-1, 1]$ defined by

$$\Gamma_\gamma(u) :\, = \frac{r_n^{2-\gamma} g_n(X_n, \theta_0)}{\int_0^1 (-r_n^2 \delta g_n(X_n, \theta_0 + r_n^\gamma ut))\, \mathrm{d}t}.$$

Suppose that $n_0 \in \boldsymbol{N}$ satisfies that $\forall n \geqslant n_0$, $(\theta_0 - r_n^\gamma, \theta_0 + r_n^\gamma) \subset \tilde{\Theta}$. Then, for any $X_n \in \mathfrak{X}_0^n$ and any $n \geqslant n_0$, $\Gamma_\gamma(u) \in [-1, 1]$. Therefore it follows from Brouwer's fixed point theorem that if $X_n \in \mathfrak{X}_0^n$ and $n \geqslant n_0$, then there exists a $\hat{u} \in [-1, 1]$ such that $\Gamma_\gamma(\hat{u}) = \hat{u}$. Setting $\hat{\theta}_n = \theta_0 + r_n^\gamma \hat{u}$, we have from Taylor's formula that

$$\begin{aligned} g_n(X_n, \hat{\theta}_n) &= g_n(X_n, \theta_0) + \int_0^1 \delta g_n(X_n, \theta_0 + t(\hat{\theta}_n - \theta_0))\, \mathrm{d}t (\hat{\theta}_n - \theta_0) \\ &= g_n(X_n, \theta_0) - \int_0^1 (-r_n^2 \delta g_n(X_n, \theta_0 + r_n^\gamma \hat{u} t))\, \mathrm{d}t r_n^{\gamma-2} \hat{u} \\ &= r_n^{\gamma-2}(\Gamma_\gamma(\hat{u}) - \hat{u}) \int_0^1 (-r_n^2 \delta g_n(X_n, \theta_0 + r_n^\gamma \hat{u} t))\, \mathrm{d}t = 0. \end{aligned}$$

Since $g_n(X_n, \cdot)$ is monotone on $\tilde{\Theta}$ for any $X_n \in \mathfrak{X}_0^n$, we see that if $X_n \in \mathfrak{X}_0^n$ and $n \geqslant n_0$, then there exists a unique $\hat{\theta}_n \in \tilde{\Theta}$ satisfying $g_n(X_n, \hat{\theta}_n) = 0$ and such $\hat{\theta}_n$ lies in the r_n^γ-neighborhood of θ_0. The proof of the first assertion of Theorem 1 will be complete if we show that

$$P[\mathcal{X}_0^n] = 1 - o(r_n).$$

From [C1] and $\gamma < 1 - 1/p_1$,

$$P[r_n^{2-\gamma}|g_n(X_n,\theta_0)| \geqslant C] \leqslant \frac{r_n^{(1-\gamma)p_1}}{C^{p_1}} \|r_n g_n(X_n,\theta_0)\|_{p_1}^{p_1} = o(r_n).$$

Combining this and Lemma 2, we obtain that

$$P[(\mathcal{X}_0^n)^c] \leqslant P\left[\inf_{\theta\in\tilde{\Theta}} (-r_n^2 \delta g_n(X_n,\theta)) \leqslant C\right] + P[r_n^{2-\gamma}|g_n(X_n,\theta_0)| \geqslant C] = o(r_n).$$

Thus we have the first assertion of Theorem 1:

$$P[(\exists_1 \hat{\theta}_n \in \tilde{\Theta}, g_n(X_n,\hat{\theta}_n) = 0) \quad \text{and} \quad (|\hat{\theta}_n - \theta_0| < r_n^\gamma)] = 1 - o(r_n).$$

If $n \geqslant n_0$, then $\hat{\theta}_n$ is well defined as a unique solution of the equation $g_n(X_n,\theta) = 0$ for any $X_n \in \mathcal{X}_0^n$. We extend $\hat{\theta}_n$ to a Θ-valued random variable on the whole of $\mathcal{X}^n$ and denote it by the same symbol $\hat{\theta}_n$.

Let us prove the second half of Theorem 1. For $\gamma' \in (2/3, \gamma)$ and $\alpha \in (0, 3\gamma - 2)$, let $\mathcal{X}_1^n$ be a subset of $\mathcal{X}^n$ defined by

$$\begin{aligned}\mathcal{X}_1^n = \Big\{ & X_n \in \mathcal{X}^n \mid |r_n^2 \delta g_n(X_n,\theta_0) + I_{\theta_0}(\theta_0)| < r_n^{\gamma'}, \\ & |r_n^2 \delta^2 g_n(X_n,\theta_0) + L_{\theta_0}(\theta_0)| < r_n^{\gamma'}, \sup_{\theta\in\Theta^\circ} |r_n^2 \delta^3 g_n(X_n,\theta)| < r_n^{-\alpha} \Big\}.\end{aligned}$$

From the definitions of $\mathcal{X}_0^n$ and $\mathcal{X}_1^n$, if $X_n \in \mathcal{X}_0^n \cap \mathcal{X}_1^n$, then

$$|\hat{\theta}_n - \theta_0| \leqslant r_n^\gamma, \quad |Z_{n,1}| < C r_n^{\gamma-1}, \quad |Z_{n,2}| < r_n^{\gamma'-1}, \quad |R_4| < r_n^{\gamma'}.$$

It is easy to show that for any $X_n \in \mathcal{X}_0^n \cap \mathcal{X}_1^n$,

$$|R_1| \leqslant \tfrac{1}{6} r_n^{3\gamma-2-\alpha}, \quad |R_2| \leqslant \tfrac{1}{2} r_n^{2\gamma-1}(|L_{\theta_0}(\theta_0)| + r_n^{\gamma'}) + \tfrac{1}{6} r_n^{3\gamma-1-\alpha}$$

and

$$|R_3| \leqslant I_{\theta_0}^{-1}(\theta_0)(r_n^{\gamma+\gamma'-1} + \tfrac{1}{2} r_n^{2\gamma-1}(|L_{\theta_0}(\theta_0)| + r_n^{\gamma'}) + \tfrac{1}{6} r_n^{3\gamma-1-\alpha}).$$

Therefore if $X_n \in \mathcal{X}_0^n \cap \mathcal{X}_1^n$, then Lemma 3 implies

$$\begin{aligned} r_n^{-1}|R| \leqslant I_{\theta_0}^{-1}(\theta_0)\,(& r_n^{\gamma'-1}|R_3| + \tfrac{1}{2}\,|\,L_{\theta_0}(\theta_0)|(2 I_{\theta_0}^{-1}(\theta_0) C r_n^{\gamma-1}|R_3| + |R_3|^2) + \\ & + \tfrac{1}{2} r_n^{2\gamma+\gamma'-2} + \tfrac{1}{6} r_n^{3\gamma-2-\alpha}). \end{aligned}$$

Since $2/3 < \gamma' < \gamma < 1$ and $0 < \alpha < 3\gamma - 2$, there exists a constant $C_1 > 0$ such that for any $X_n \in \mathcal{X}_0^n \cap \mathcal{X}_1^n$, $|R_3| \leqslant C_1 r_n^{1/3}$. Moreover, setting $\epsilon = \min(3\gamma -$

$2-\alpha, \gamma'-2/3)$, we see that there exists a constant $C_2 > 0$ such that for any $X_n \in \mathcal{X}_0^n \cap \mathcal{X}_1^n$, $r_n^{-1}|R| \leqslant C_2 r_n^{\epsilon}$.

The proof will be complete if we show that

$$P[\mathcal{X}_0^n \cap \mathcal{X}_1^n] = 1 - o(r_n). \tag{4.11}$$

From Markov's inequality, we see that

$$\begin{aligned}
P[(\mathcal{X}_1^n)^c] \;\leqslant\;& P[|r_n^2 \delta g_n(X_n,\theta_0) + I_{\theta_0}(\theta_0)| \geqslant r_n^{\gamma'}] + \\
&+ P[|r_n^2 \delta^2 g_n(X_n,\theta_0) + L_{\theta_0}(\theta_0)| \geqslant r_n^{\gamma'}] + \\
&+ P\left[\sup_{\theta\in\Theta^\circ} |r_n^2\delta^3 g_n(X_n,\theta)| \geqslant r_n^{-\alpha}\right] \\
\leqslant\;& r_n^{p_2(\gamma-\gamma')} E|r_n^{-\gamma}(r_n^2\delta g_n(X_n,\theta_0) + I_{\theta_0}(\theta_0))|^{p_2} + \\
&+ r_n^{p_2(\gamma-\gamma')} E|r_n^{-\gamma}(r_n^2\delta^2 g_n(X_n,\theta_0) + L_{\theta_0}(\theta_0))|^{p_2} + \\
&+ r_n^{\alpha p_3} E\left|\sup_{\theta\in\Theta^\circ}\left| r_n^2 \delta^3 g_n(X_n,\theta_0)\right|\right|^{p_3}.
\end{aligned}$$

Since $\gamma - 1/p_2 > 2/3$ and $3\gamma - 2 > 1/p_3$, we can choose some $\gamma' \in (2/3, \gamma - 1/p_2)$ and $\alpha \in (1/p_3, 3\gamma - 2)$. For such γ' and α, it is shown that $P[(\mathcal{X}_1^n)^c] = o(r_n)$. As is shown in a previous subsection,

$$P[(\mathcal{X}_0^n)^c] = o(r_n).$$

Therefore the proof is complete.

Acknowledgement

The authors wish to thank the referees and the editor for their valuable comments.

Note

[1]Even if the index set $\boldsymbol{N}$ is replaced by $\boldsymbol{R}_+$, the results of this paper hold true.

References

1. Doukhan, P.: *Mixing-Properties and Examples, Lecture Notes in Statistics* 85, Springer, Berlin, Heidelberg, New York, 1995.
2. Kusuoka, S. and Yoshida, N.: Malliavin calculus, strong mixing, and expansion of diffusion functionals, 1997, Preprint.
3. Mykland, P. A.: Asymptotic expansions and bootstrapping distributions for dependent variables: a martingale approach, *Ann. Statist.* **20**(2) (1992), 623–654.
4. Mykland, P. A.: Embedding and asymptotic expansions for martingales, *Probab. Theory Relat. Fields* **103** (1995), 475–492.
5. Lánska, V.: Minimum contrast estimation in diffusion processes, *J. Appl. Probab.* **16** (1979), 65–75.

6. Veretennikov, A. Yu.: On polynomial mixing bounds for stochastic differential equations, *Stochastic Processes and their Applications* **70** (1997), 115–127.
7. Yoshida, N.: Robust M-estimators in diffusion processes, *Ann. Inst. Statist. Math.* **40**(4) (1988), 799–820.
8. Yoshida, N.: Malliavin calculus and asymptotic expansion for martingale I, II, *Research Memorandum*, The Institute of Statistical Mathematics **504, 517**, 1994a.
9. Yoshida, N.: Asymptotic expansion for local density of martingales, *Research Memorandum*, The Institute of Statistical Mathematics **520**, 1994b.
10. Yoshida, N.: Asymptotic expansions of a functional of a one-dimensional diffusion processes, *Research Memorandum*, The Institute of Statistical Mathematics **560**, 1995.