Ann Inst Stat Math (2009) 61:629–661
DOI 10.1007/s10463-008-0190-4

Third-order asymptotic expansion of M-estimators for diffusion processes

Yuji Sakamoto · Nakahiro Yoshida

Received: 27 December 2004 / Revised: 4 June 2007 / Published online: 18 November 2008

Abstract For an unknown parameter in the drift function of a diffusion process, we consider an M-estimator based on continuously observed data, and obtain its distributional asymptotic expansion up to the third order. Our setting covers the misspecified cases. To represent the coefficients in the asymptotic expansion, we derive some formulas for asymptotic cumulants of stochastic integrals, which are widely applicable to many other problems. Furthermore, asymptotic properties of cumulants of mixing processes will be also studied in a general setting.

Keywords Asymptotic expansion · M-estimator · Diffusion process

1 Introduction

Suppose that we are interested in an unknown parameter $\theta_0 \in \Theta \subset \mathbb{R}^p$, and that we can observe a continuous path $X = (X_t)_{t\in[0,T]}$ of a d-dimensional stationary diffusion process satisfying a stochastic differential equation

$$\mathrm{d}X_t = V_0(X_t)\mathrm{d}t + V(X_t)\mathrm{d}w_t. \tag{1}$$

Y. Sakamoto (✉)
Graduate School of Human Development and Environment, Kobe University, 3-11 Tsurukabuto, Nada, Kobe 657-8501, Japan
e-mail: sakamoto@main.h.kobe-u.ac.jp

N. Yoshida
Graduate School of Mathematical Sciences, The University of Tokyo, Tokyo, Japan

N. Yoshida
Japan Science and Technology Agency, Tokyo, Japan

Here $V_0 = (V_0^i)_{i=1,\dots,d} : \mathbb{R}^d \to \mathbb{R}^d$, $V = (V_j^i)_{i=1,\dots,d,j=1,\dots,r} : \mathbb{R}^d \to \mathbb{R}^d \otimes \mathbb{R}^r$ (whose smoothness conditions are mentioned in Remark 1), and w is an r-dimensional standard Wiener process defined on some probability space $(\Omega, \mathfrak{F}, P)$. We then expect that the observations X have information about the parameter value θ_0.

First, let us discuss the maximum likelihood method just to illustrate a more general estimation scheme we will consider later. To estimate θ_0 based on observations X satisfying (1), we usually model the observation process X in a parametrized d-dimensional stationary diffusion process described by the equation

$$\mathrm{d}X_t = \tilde{V}_0(X_t, \theta)\mathrm{d}t + \tilde{V}(X_t)\mathrm{d}w_t, \quad \theta = (\theta^1, \dots, \theta^p) \in \Theta, \tag{2}$$

where $\tilde{V}_0 = (\tilde{V}_0^i)_{i=1,\dots,d} : \mathbb{R}^d \times \Theta \to \mathbb{R}^d$ and $\tilde{V} = (\tilde{V}_j^i)_{i=1,\dots,d,j=1,\dots,r} : \mathbb{R}^d \to \mathbb{R}^d \otimes \mathbb{R}^r$ are given functions. Note that in this setting the functions V_0 and V in the true process (1) are unknown and θ_0 is unknown target, but the model (2) with $\theta = \theta_0$ does not always coincide with the true model (1). Therefore, the system process (1) in principle has no relations with the parametric model (2) the statistician uses to estimate his/her statistical parameter θ, that is, the misspecified case is in our scope. For model (2), the log-likelihood function ℓ is given by

$$\begin{aligned}\ell(X, \theta) = \log \frac{\mathrm{d}\tilde{\nu}_\theta}{\mathrm{d}\nu_*}(X_0) + &\int_0^T \tilde{V}_0'(\tilde{V}\tilde{V}')^{-1}(X_t, \theta)\mathrm{d}X_t \\ -\frac{1}{2}&\int_0^T \tilde{V}_0'(\tilde{V}\tilde{V}')^{-1}\tilde{V}_0(X_t, \theta)\mathrm{d}t,\end{aligned} \tag{3}$$

where $\tilde{\nu}_\theta$ is a stationary distribution of a diffusion process satisfying (2) and ν_* is a σ-finite measure on $\mathbb{R}^d$ dominating all $\tilde{\nu}_\theta$. By using ℓ, we can compute the maximum likelihood estimator as a solution of the likelihood equation $\delta_a \ell(X, \theta) = 0$, $\delta_a = \frac{\partial}{\partial \theta^a}$, $a = 1, \dots, p$.

More generally, we may use a minimum contrast estimator defined as a solution of the stochastic equation $\delta_a \Psi(X, \theta) = 0$, $a = 1, \dots, p$, where

$$\Psi(X, \theta) = \tilde{A}(X_0, \theta) + \int_0^T \tilde{B}(X_t, \theta)\mathrm{d}X_t + \int_0^T \tilde{C}(X_t, \theta)\mathrm{d}t$$

for given functions $\tilde{A} : \mathbb{R}^d \times \Theta \to \mathbb{R}$, $\tilde{B} : \mathbb{R}^d \times \Theta \to \mathbb{R} \otimes \mathbb{R}^d$, $\tilde{C} : \mathbb{R}^d \times \Theta \to \mathbb{R}$.

On the order hand, it is also possible to consider an estimator defined as a root of

$$H(X, \theta) := h(X_T, \theta) - h(X_0, \theta) - \int_0^T \tilde{\mathscr{A}}_\theta h(X_t, \theta)\mathrm{d}t = 0$$

for a given function $h : \mathbb{R}^d \times \Theta \to \mathbb{R}^p$, where $\tilde{\mathscr{A}}_\theta$ is the generator of (2):

$$\tilde{\mathscr{A}}_\theta = \sum_{i=1}^d \tilde{V}_0^i(x, \theta)\frac{\partial}{\partial x^i} + \frac{1}{2}\sum_{i,j}^d \sum_{k=1}^r \tilde{V}_k^i(x)\tilde{V}_k^j(x)\frac{\partial^2}{\partial x^i x^j}.$$

From Itô's formula, $\delta_a \Psi$ and $H_{a;}$, the a-th element of H, can be rewritten as

$$\delta_a \Psi(X,\theta) = \delta_a \tilde{A}(X_0,\theta) + \int_0^T \delta_a \tilde{B}(X_t,\theta)V(X_t)\mathrm{d}w_t + \int_0^T (\delta_a \tilde{B}(X_t,\theta)V_0(X_t) + \delta_a \tilde{C}(X_t,\theta))\mathrm{d}t$$

and $H_{a;}(X,\theta) = \int_0^T \nabla_x h_{a;}(X_t,\theta)V(X_t)\mathrm{d}w_t + \int_0^T (\mathscr{A} - \tilde{\mathscr{A}_\theta})h_{a;}(X_t,\theta)\mathrm{d}t$, respectively. Here $h_{a;}$ is the a-th element of h, $\nabla_x h_{a;} = (\partial_1 h_{a;},\ldots,\partial_d h_{a;})$, $\partial_i = \partial/\partial x^i$, and $\mathscr{A}$ is the generator of the diffusion process (1).

Unifying the above estimators, we here consider an M-estimator corresponding to a p-dimensional estimating function $\psi = (\psi_{1;},\ldots,\psi_{p;})$ that has a representation [under the true model (1)]

$$\psi_{a;}(X,\theta) = A_{a;}(X_0,\theta) + \int_0^T B_{a;}(X_t,\theta)\mathrm{d}w_t + \int_0^T C_{a;}(X_t,\theta)\mathrm{d}t \tag{4}$$

for some mappings $A_{a;}$, $B_{a;}$ and $C_{a;}$. Note that (4) does not give the definition of the estimating function ψ, but a representation. In an actual situation, ψ becomes the derivative of Ψ (or ℓ), or H itself, given above, and for each estimating function, functions A, B and C in the representation (4) are given by

$$\begin{aligned}
A_{a;}(x,\theta) &= \delta_a \tilde{A}(x,\theta) \quad \text{or} \quad 0,\\
B_{a;}(x,\theta) &= \delta_a \tilde{B}(x,\theta)V(x) \quad \text{or} \quad \nabla_x h_{a;}(x,\theta)V(x),\\
C_{a;}(x,\theta) &= \delta_a \tilde{B}(x,\theta)V_0(x) + \delta_a \tilde{C}(x,\theta) \quad \text{or} \quad (\mathscr{A} - \tilde{\mathscr{A}_\theta})h_{a;}(x,\theta).
\end{aligned} \tag{5}$$

In this article, we consider M-estimators $\hat{\theta}_T$ whose estimating functions have the representation (4). Applying Theorem 6.4 in Sakamoto and Yoshida (2004) or its original version, Sakamoto and Yoshida (1999), with the Hörmander type condition in Kusuoka and Yoshida (2000), we obtain their distributional asymptotic expansions up to the third order.

The theory of the first-order statistical inference for diffusion processes has been well developed. We refer the reader to the text books by Kutoyants (1984, 1994), Prakasa Rao (1999), and Kutoyants (2004). Regarding the Edgeworth expansion and the higher-order statistical inference for ergodic diffusions, the second-order distributional expansion of a martingale with its application to the maximum likelihood estimator is in Yoshida (1997); Edgeworth expansions of M-estimators in Sakamoto and Yoshida (1998a) by the *global approach* (martingale approach).

The aim of the present article is to derive and validate a third-order asymptotic expansion formula for the M-estimator of the diffusion process (*diffusion M formula*). After that, we will make an expansion formula for the maximum likelihood estimator (*diffusion MLE formula*) as a special case of this result. Our guiding principles are the *local approach* (mixing approach) and the Malliavin calculus. See Sakamoto and Yoshida (1999, 2004) for more details.

A third-order diffusion MLE formula was originally obtained in Sakamoto and Yoshida (1998b). It used the Bartlett type identities, as the use of those identities is very common in independent models. However, in this article, we will derive diffusion MLE formula without Bartlett type identities because it is not necessarily easy to prove those identities rigorously for diffusion models. And, again, we can obtain the same third-order diffusion MLE formula as Sakamoto and Yoshida (1998a) if we assume the Bartlett type identities. From a practical point of view, it is meaningful to give explicit expressions to the coefficients appearing in the expansion. For this purpose, we will provide certain cumulant formulas for stochastic integrals.

2 Expansion formulas

To define the M-estimator, we will first show the existence of the solution of the estimating equation, and after that we will present an expansion formula. Finally, we will apply the result to the maximum likelihood estimator.

We denote by ν the stationary distribution of X satisfying (1), and also assume $E|X_0|^k < \infty$ for any $k \geq 1$. Assume that the parameter space Θ is a bounded convex open set in $\mathbb{R}^p$. Fix $\theta_0 \in \Theta$ arbitrarily. For the sake of simplicity, the derivatives of ψ and the functions A, B, C in ψ w.r.t θ are expressed as

$$A_{a;a_1\cdots a_k}(x,\theta) = \delta_{a_1}\cdots\delta_{a_k}A_{a;}(x,\theta), \quad B_{a;a_1\cdots a_k}(x,\theta) = \delta_{a_1}\cdots\delta_{a_k}B_{a;}(x,\theta),$$
$$C_{a;a_1\cdots a_k}(x,\theta) = \delta_{a_1}\cdots\delta_{a_k}C_{a;}(x,\theta), \quad \psi_{a;a_1\cdots a_k}(\theta) = \delta_{a_1}\cdots\delta_{a_k}\psi_{a;}(\theta).$$

For a measurable function $f:\mathbb{R}^d \to \mathbb{R}$, let $G\langle f\rangle$ be a function such that $\mathscr{A}G\langle f\rangle = f - \nu(f)$, and $[f] = -(\nabla G\langle f\rangle)V$, where $\nu(f) = \int_{\mathbb{R}^d} f(x)\nu(\mathrm{d}x)$. Assume that

[DM1] (i) for each $x \in \mathbb{R}^d$ and $a \in \{1,\ldots,p\}$, $A_{a;}(x,\cdot)$, $B_{a;}(x,\cdot)$, $C_{a;}(x,\cdot)$ are of class C^5 on Θ;
(ii) there exist positive constants $C_i, m_i, i = 1, 2, 3$ such that for any $x \in \mathbb{R}^d$, $k = 1,\ldots,5, a, a_k \in \{1,\ldots,p\}$,

$$\sup_{\theta\in\Theta}|A_{a;a_1\cdots a_k}(x,\theta)| \leq C_1(1+|x|)^{m_1},$$
$$\sup_{\theta\in\Theta}|B_{a;a_1\cdots a_k}(x,\theta)| \leq C_2(1+|x|)^{m_2},$$
$$\sup_{\theta\in\Theta}|C_{a;a_1\cdots a_k}(x,\theta)| \leq C_3(1+|x|)^{m_3};$$

[DM2] for $a, b, c, a_1, a_2 \in \{1,\ldots,p\}$, there exist functions $G\langle C_{a;}\rangle(\cdot,\theta_0)$, $G\langle C_{a;a_1}\rangle(\cdot,\theta_0)$, $G\langle C_{a;a_1a_2}\rangle(\cdot,\theta_0)$, $G\langle B^*_{a;}\cdot B^*_{b;}\rangle(\cdot,\theta_0)$, $G\langle B^*_{a;a_1}\cdot B^*_{b;}\rangle(\cdot,\theta_0)$, $G\langle [B^*_{a;}\cdot B^*_{b;}]\cdot B^*_{c;}\rangle(\cdot,\theta_0)$, where $B^*_{a;A} = B_{a;A} + [C_{a;A}]$;
[DM3] (i) for each $a \in \{1,\ldots,p\}$, $\nu(C_{a;}(\cdot,\theta_0)) = 0$;

(ii) for $a, b, c, a_1, a_2 \in \{1, \ldots, p\}$, the functions $C_{a;}$, $C_{a;a_1}$, $C_{a;a_1a_2}$, $B^*_{a;} \cdot B^*_{b;}$, $B^*_{a;a_1} \cdot B^*_{b;}$, $[B^*_{a;} \cdot B^*_{b;}] \cdot B^*_{c;} \in \mathscr{G}$, where

$$\mathscr{G} = \left\{ f : \mathbb{R}^d \times \Theta \to \mathbb{R} \;\middle|\; \exists m > 0, \exists C > 0, \right.$$
$$\left. |G\langle f\rangle(x, \theta_0)| \leq C(1 + |x|)^m, \;\; |[f](x, \theta_0)| \leq C(1 + |x|)^m \right\}.$$

For simplicity, we will hereafter denote $\psi_{a;A}(X, \theta)$ by $\psi_{a;A}(\theta)$. Let

$$Z_{a;} = T^{1/2}(T^{-1}\psi_{a;}(\theta_0) - \bar{\nu}_{a;}(\theta_0)), \quad Z_{a;b} = T^{1/2}(T^{-1}\psi_{a;b}(\theta_0) - \bar{\nu}_{a;b}(\theta_0)),$$
$$Z_{a;bc} = T^{1/2}(T^{-1}\psi_{a;bc}(\theta_0) - \bar{\nu}_{a;bc}(\theta_0)),$$

where $\bar{\nu}_{a;}(\theta) = E[\psi_{a;}(\theta)]/T$, $\bar{\nu}_{a;a_1\cdots a_k}(\theta) = E[\psi_{a;a_1\cdots a_k}(\theta)]/T$. In case the matrix $(\bar{\nu}_{a;b}(\theta_0))_{a,b=1}^p$ is nonsingular, let $(\bar{\nu}^{a;b}) = (\bar{\nu}_{a;b}(\theta_0))^{-1}$, $Z^{a;} = -\bar{\nu}^{a;a'} Z_{a';}$, $Z^{a;}_{\ b} = -\bar{\nu}^{a;a'} Z_{a';b}$, $Z^{a;}_{\ bc} = -\bar{\nu}^{a;a'} Z_{a';bc}$, and $\bar{\nu}^{a;}_{\ bc} = -\bar{\nu}^{a;a'}\bar{\nu}_{a';bc}(\theta_0)$, $\bar{\nu}^{a;}_{\ bcd} = -\bar{\nu}^{a;a'}\bar{\nu}_{a';bcd}(\theta_0)$, and $\Delta^{a;} = -\bar{\nu}^{a;a'}\nu(A_{a';}(\cdot, \theta_0))$. Hereafter, we omit θ_0 in functions of θ when they are evaluated at $\theta = \theta_0$, e.g., $\bar{\nu}_{a;a_1\cdots a_k} = \bar{\nu}_{a;a_1\cdots a_k}(\theta_0)$.

Moreover, we suppose that there exists a positive constant a such that

$$E\left|E\left[f \mid \mathscr{B}^X_{[s]}\right] - E[f]\right| \leq a^{-1}\mathrm{e}^{-a(t-s)}\|f\|_\infty$$

for any $s, t \in \mathbb{R}_+$, $s \leq t$, and for any bounded $\mathscr{B}^X_{[t,\infty)}$-measurable function f, where $\mathscr{B}^X_I = \sigma[X_t \in I \cap \mathbb{R}_+] \vee \mathscr{N}$, $I \subset \mathbb{R}$, $\mathscr{N}$ is the σ-field generated by null sets. Here we say that X has the geometric-mixing property if this condition holds true. Under a very mild condition, the geometric-mixing property of diffusion processes was proved by Kusuoka and Yoshida (2000). See Veretennikov (1987, 1997) for non-degenerate diffusion, Masuda (2004) for Lévy OU process, Meyn and Tweedie (1992, 1993a,b) for discrete or continuous-time Markov process.

Theorem 1 *Suppose that there exists an open subset $\tilde{\Theta} \subset \Theta$ such that $\theta_0 \in \tilde{\Theta}$ and that*

$$\inf_{\theta_1, \theta_2 \in \tilde{\Theta}, |x|=1} \left| x' \left(\int_0^1 \nu\left(C_{a;b}(\cdot, \theta_1 + s(\theta_2 - \theta_1))\right) \mathrm{d}s \right) \right| > 0. \tag{6}$$

Moreover, assume that for any $a, b, c = 1, \ldots, p$, $\delta_c \nu(A_{a;b}(\cdot, \theta)) = \nu(A_{a;bc}(\cdot, \theta))$, $\delta_c \nu(C_{a;b}(\cdot, \theta)) = \nu(C_{a;bc}(\cdot, \theta))$. Then, under the conditions [DM1] *and* [DM3] (i)*, for any $m > 0$, $\gamma \in (0, 1)$,*

$$P[(\exists_1 \hat{\theta}_T \in \tilde{\Theta} \text{ such that } \psi(\hat{\theta}_T) = 0) \;\; \text{and} \;\; (|\hat{\theta}_T - \theta_0| < T^{-\gamma/2})] = 1 - o(T^{-m}),$$

where $\exists_1$ stands for unique existence. Furthermore, for any extension of $\hat{\theta}_T$, say $\hat{\theta}_T$, and for any $\beta \in C^2_B(\Theta) := \{f \in C^2(\Theta) \mid f, \partial f, \partial^2 f$ are all bounded$\}$, let $\hat{\theta}^*_T =$

$\hat{\theta}_T - T^{-1}\beta(\hat{\theta}_T)$. *Define* R_3^a *by*

$$\begin{aligned}\sqrt{T}(\hat{\theta}_T^* - \theta_0)^a &= Z^{a;} + \frac{1}{\sqrt{T}}\left(Z^{a;}{}_b Z^{b;} + \frac{1}{2}\bar{\nu}^{a;}{}_{bc} Z^{b;} Z^{c;} + \Delta^{a;} - \beta^a\right)\\ &\quad + \frac{1}{T}\left(\frac{1}{6}\left(\bar{\nu}^{a;}{}_{bcd} + 3\bar{\nu}^{a;}{}_{be}\bar{\nu}^{e;}{}_{cd}\right) Z^{b;} Z^{c;} Z^{d;} + \bar{\nu}^{a;}{}_{bc} Z^{b;} Z^{c;}{}_d Z^{d;}\right.\\ &\quad + \frac{1}{2}\bar{\nu}^{b;}{}_{cd} Z^{a;}{}_b Z^{c;} Z^{d;} + \frac{1}{2} Z^{a;}{}_{bc} Z^{b;} Z^{c;} + Z^{a;}{}_b Z^{b;}{}_c Z^{c;}\\ &\quad \left. - Z^{b;}\delta_b\beta^a + \Delta^{b;}\left(Z^{a;}{}_b + \bar{\nu}^{a;}{}_{bc} Z^{c;}\right)\right) + \frac{1}{T\sqrt{T}} R_3^a. \end{aligned} \tag{7}$$

Then there exist $C > 0$ *and* $\varepsilon > 0$ *such that*

$$P\left[T^{-1/2}|R_3^a| \le CT^{-\varepsilon/2},\ a = 1, \ldots, p\right] = 1 - o(T^{-m/2}). \tag{8}$$

It is possible to choose a measurable version of $\hat{\theta}_T$ by the measurable selection theorem, cf. Pfanzagl (1994): on a certain event described in the proof of the existence of a root $\hat{\theta}_T$, apply the measurable selection theorem to the functional $-|\psi(\theta)|$ to obtain a measurable version of $\hat{\theta}_T$, and next extend it to the whole sample space as a measurable mapping. The above-mentioned theorem ensures the existence of a consistent sequence of M-estimators. On the other hand, it is possible to show the convergence of any sequence of M-estimators with a convergence rate if we apply the polynomial type large deviation inequality (Yoshida 2005).

For the M-estimator $\hat{\theta}_T$ or the bias-corrected version $\hat{\theta}_T^*$ defined in Theorem 1, their distributional asymptotic expansion can be derived from Theorem 6.4 of Sakamoto and Yoshida (2004).

Let $Z_T^{(0)} = T^{1/2}(Z_{1;}, \ldots, Z_{p;})$ and $Z_T^{(1)} = T^{1/2}(\overbrace{Z_{1;1}, \ldots, Z_{p;p}}^{p^2}, \overbrace{Z_{1;11}, \ldots, Z_{p;pp}}^{p^3})$. To designate the dependency of T, write $Z_T^{(0)} = (Z_{1;,T}^{(0)}, \ldots, Z_{p;,T}^{(0)})$ and $Z_T^{(1)} = (Z_{1;1,T}^{(1)}, \ldots, Z_{p;p,T}^{(1)}, Z_{1;11,T}^{(1)}, \ldots, Z_{p;pp,T}^{(1)})$. Then the Stratonovich stochastic differential equations they satisfy are given by

$$\begin{aligned}\mathrm{d}Z_{a;,t}^{(0)} &= B_{a;}(X_t, \theta_0) \circ \mathrm{d}w_t + C_{a;}^*(X_t, \theta_0)\mathrm{d}t,\\ \mathrm{d}Z_{a;b,t}^{(1)} &= B_{a;b}(X_t, \theta_0) \circ \mathrm{d}w_t + C_{a;b}^*(X_t, \theta_0)\mathrm{d}t,\\ \mathrm{d}Z_{a;bc,t}^{(1)} &= B_{a;bc}(X_t, \theta_0) \circ \mathrm{d}w_t + C_{a;bc}^*(X_t, \theta_0)\mathrm{d}t,\end{aligned}$$

where $C_{a;A}^*(x,\theta) = C_{a;A}(x,\theta) - \frac{1}{2}\sum_{j=1}^r \sum_{k=1}^d V_j^k(x)\partial_k B_{a;A}^j(x,\theta) - \nu(C_{a;A}(\cdot,\theta_0))$ for $A = \{\phi, a_1, a_1 a_2\}$, $a_i = 1, \ldots, p$. Note that the d-dimensional diffusion process $X = (X^1, \ldots, X^d)$ defined by (1) satisfies

$$\mathrm{d}X_t^i = V_j^i(X_t) \circ \mathrm{d}w_t^j + \tilde{V}_0^i(X_t)\mathrm{d}t,$$

where V_j^i is the (i, j)-th element of V and V_0^i is the i-th element of V_0, and $\tilde{V}_0^i$ is defined by $\tilde{V}_0^i = V_0^i - \frac{1}{2}\sum_{j=1}^r \sum_{k=1}^d V_j^k \partial_k V_j^i,\ \ i = 1, \ldots, d$. Denote by $B_{a;}^i$ the i-th element of $B_{a;}$ and let $\bar{V}_{0,1} = (\tilde{V}_0^1, \ldots, \tilde{V}_0^d, C_{1;}^*, \ldots, C_{p;}^*)$ and $\bar{V}_{i,1} = (V_i^1, \ldots, V_i^d, B_{1;}^i, \ldots, B_{p;}^i)$, $i = 1, \ldots, r$.

Assume that

[L] for some integer $q_1 \leq p^2 + p^3$, there exists a q_1-dimensional random variable $\dot{Z}_T$ consisting of the elements of $Z_T^{(1)}$ such that
  - (i) $\mathrm{Cov}(T^{-1/2} Z_T^*)$ converges to a positive definite matrix, where $Z_T^* = (Z_T^{(0)}, \dot{Z}_T)$,
  - (ii) $\ddot{Z}_T = L\dot{Z}_T$ for some $q_2 \times q_1$ matrix L, where $\ddot{Z}_T$ is a q_2-dimensional random variable consisting of the other elements of $Z^{(1)}$ than those of $\dot{Z}_T$, and $q_1 + q_2 = p^2 + p^3$,
  - (iii) for some x in $\mathrm{supp}(\nu)$, $\mathrm{Lie}[\bar{V}_0; \bar{V}_1, \ldots, \bar{V}_r](x, 0) = \mathbb{R}^{d+p+q_1}$, where $\bar{V}_0 = (\bar{V}_{0,1}, \dot{C}_1^*, \ldots, \dot{C}_{q_1}^*)$, $\bar{V}_i = (\bar{V}_{i,1}, \dot{B}_1^i, \ldots, \dot{B}_{q_1}^i)$, $i = 1, \ldots, r$, $\dot{C}_j^*$ is the drift of the Stratonovich stochastic differential equation for the j-th element of $\dot{Z}_t$, $\dot{B}_j^i$ is the i-th element of its diffusion coefficient.

Here $\mathrm{Lie}[\bar{V}_0; \bar{V}_1, \ldots, \bar{V}_r]$ denotes the linear manifold spanned by $\bigcup_{n=0}^{\infty} \Sigma_n$, $\Sigma_0 = \{\bar{V}_1, \ldots, \bar{V}_r\}$, $\Sigma_n = \{[\bar{V}_j, V] \mid V \in \Sigma_{n-1}, j = 0, 1, \ldots, r\}$, and $[\bar{V}_j, V]$ is the Lie bracket.

In order to represent coefficients in the expansion formula, we put $(\tilde{\nu}^{a;b}) = (\nu(C_{a;b}))^{-1}$ (evaluated at θ_0), $\tilde{\Delta}^{a;} = -\tilde{\nu}^{a;a'}\nu(A_{a';})$, and $\bar{A}^{a;b} = \tilde{\nu}^{a;a'}\tilde{\nu}^{b;b'}\nu(A_{a';b'})$. For any index sets A, B, C and D, let $B_{a;A}^* = B_{a;A} + [C_{a;A}]$,

$$
\begin{aligned}
&\bar{F}_{a;A,b;B} = \nu(B_{a;A}^* \cdot B_{b;B}^*), \quad \bar{F}_{[a;A,b;B],c;C} = \nu([B_{a;A}^* \cdot B_{b;B}^*] \cdot B_{c;C}^*),\\
&\bar{F}_{[a;A,b;B],[c;C,d;D]} = \nu([B_{a;A}^* \cdot B_{b;B}^*] \cdot [B_{c;C}^* \cdot B_{d;D}^*]),\\
&\bar{F}_{[[a;A,b;B],c;C],d;D} = \nu([[B_{a;A}^* \cdot B_{b;B}^*] \cdot B_{c;C}^*] \cdot B_{d;D}^*).
\end{aligned}
$$

Moreover, the following are also requisite for the formula:

$$
\begin{aligned}
\rho^{ab} &= \tilde{\nu}^{a;a'}\tilde{\nu}^{b;b'}\bar{F}_{a',b'}, \quad (\rho_{ab}) = (\rho^{ab})^{-1},\\
\tilde{\tau}^{ab} &= \tilde{\nu}^{a;a'}\tilde{\nu}^{b;b'}\tau_{a'b'} - \bar{A}^{a;a'}\tilde{\nu}^{b;b'}\bar{F}_{a',b'} - \tilde{\nu}^{a;a'}\bar{A}^{b;b'}\bar{F}_{a',b'},\\
\tau_{ab} &= \mathrm{Cov}[A_{a;}(X_0), A_{b;}(X_0)] - \nu(A_{a;}G\langle C_{b;}\rangle) - \nu(G\langle C_{a;}\rangle A_{b;})\\
&\quad +2\nu(G\langle C_{a;}\rangle\, G\langle C_{b;}\rangle) + E\left[G\langle C_{a;}\rangle(X_T)\int_0^T B_{b;}^*(X_t, \theta)\mathrm{d}w_t\right]\\
&\quad +E\left[\int_0^T B_{a;}^*(X_t, \theta)\mathrm{d}w_t G\langle C_{b;}\rangle(X_T)\right],\\
\mu^{*a;}_{\ \ bc} &= \frac{1}{2}\tilde{\nu}^{a;a'}\left(\tilde{\nu}^{c';c''}\rho_{cc'}\bar{F}_{a';b,c'';} + \tilde{\nu}^{b';b''}\rho_{bb'}\bar{F}_{a';c,b'';} - \nu(C_{a';bc})\right),\\
\eta^{*a;}_{\ \ b,c} &= \tilde{\nu}^{aa'}\left(\tilde{\nu}^{c';c''}\rho_{cc'}\bar{F}_{a';b,c'';} - \nu(C_{a';bc})\right),
\end{aligned}
$$

$$U^{*a;}_{\ \ bcd} = \frac{1}{6}\tilde{\nu}^{a;a'}\left(-\nu(C_{a';bcd}) + \sum_{(bc,d)}^{[3]} \tilde{\nu}^{d';d''}\rho_{dd'}\bar{F}_{a';bc,d'';}\right) + \frac{1}{3}\sum_{(bc,d)}^{[3]} \tilde{\mu}^{*d';}_{\ \ bc}\tilde{\eta}^{*a;}_{\ \ d',d},$$

$$\bar{\lambda}^{*abc} = -\tilde{\nu}^{a;a'}\tilde{\nu}^{b;b'}\tilde{\nu}^{c;c'}\sum_{(ab,c)}^{[3]} \bar{F}_{[a';,b';],c';}$$

$$H^{*abcd} = \tilde{\nu}^{a;a'}\tilde{\nu}^{b;b'}\tilde{\nu}^{c;c'}\tilde{\nu}^{d;d'}\left(\sum_{(a'b',c',d')}^{[6]}\left(\bar{F}_{[[a';,b';],c';],d';} + \bar{F}_{[[a';,b';],d';],c';}\right)\right.$$
$$\left. + \sum_{(a'b',c'd')}^{[3]} \bar{F}_{[a';,b';],[c';,d';]}\right)$$

$$V^{*a;}_{\ \ B,c} = \tilde{\nu}^{a;a'}\tilde{\nu}^{c';c''}\rho_{c'c}\bar{F}_{a';B,c'';}$$

$$\tilde{M}^{*a;}{}_{b,}{}^{c;}{}_{d} = \tilde{\nu}^{a;a'}\tilde{\nu}^{c;c'}\bar{F}_{a';b,c';d} - \tilde{\nu}^{a;a'}\tilde{\nu}^{f;f'}V^{*c;}_{\ \ d,f}\bar{F}_{a';b,f'}$$
$$-\tilde{\nu}^{e;e'}\tilde{\nu}^{c;c'}V^{*a;}_{\ \ b,e}\bar{F}_{c';d,e'} + \tilde{\nu}^{e;e'}\tilde{\nu}^{f;f'}V^{*a;}_{\ \ b,e}V^{*c;}_{\ \ d,f}\bar{F}_{e',f'}$$

and

$$\tilde{N}^{*a;}{}_{,}{}^{b;}{}_{,}{}^{c;}{}_{d} = -\tilde{\nu}^{a;a'}\tilde{\nu}^{b;b'}\tilde{\nu}^{c;c'}\left(\bar{F}_{[a';,b';],c';d} + \bar{F}_{[a';,c';d],b';} + \bar{F}_{[b';,c';d],a';}\right)$$
$$+V^{*c;}_{\ \ d,e}\tilde{\nu}^{a;a'}\tilde{\nu}^{b;b'}\tilde{\nu}^{e;e'}\left(\bar{F}_{[a';,b';],e';} + \bar{F}_{[a';,e';],b';} + \bar{F}_{[b';,e';],a';}\right).$$

Here $\sum_{(ab,c)}^{[3]}$, $\sum_{(ab,c,d)}^{[6]}$, etc. are summations over the indicated number of terms obtained by rearranging the subscripts. For $M > 0$ and $\gamma > 0$, the set $\mathscr{E}(M,\gamma)$ of measurable functions from $\mathbb{R}^p \to \mathbb{R}$ is defined by

$$\mathscr{E}(M,\gamma) = \{f : \mathbb{R}^p \to \mathbb{R}, \text{ measurable}, |f(x)| \le M(1+|x|)^\gamma\},$$

and for $f \in \mathscr{E}(M,\gamma)$, $r > 0$ and a positive definite matrix σ, let

$$\omega(f,r,\sigma) = \int_{\mathbb{R}^p} \sup\{|f(x+y) - f(x)| : |y| \le r\}\phi(x;\sigma)\mathrm{d}x,$$

and let

$$h_{a_1\cdots a_k}(x;\sigma) = \frac{(-1)^k}{\phi(x;\sigma)}\frac{\partial^k}{\partial x^{a_1}\cdots\partial x^{a_k}}\phi(x;\sigma)$$

where $\phi(x;\sigma)$ is the density function of p-dimensional normal distribution $N_p(0,\sigma)$. Hereafter, for a matrix $\sigma = (\sigma^{ab})$, we will often write σ^{ab} to denote σ, for example, $h_{a_1\cdots a_k}(x;\sigma^{ab})$ for $h_{a_1\cdots a_k}(x;\sigma)$.

By using these notations, we obtain the third-order *diffusion M formula*:

Theorem 2 *Let $M, \gamma > 0$, and $\hat{\rho} > (\rho^{ab})$. Assume that* [L], [DM2], [DM3](ii) *and the conditions in Theorem 1 hold true. For any $\beta \in C_B^2(\Theta)$ and $\hat{\theta}_T$ defined in Theorem 1, let $\hat{\theta}_T^* = \hat{\theta}_T - \beta(\hat{\theta}_T)/T$. Moreover, assume that the diffusion process X given (1) has the geometrically strong mixing property. Then there exist positive constants $c, \tilde{C}, \tilde{\varepsilon}$ such that for any $f \in \mathscr{E}(M, \gamma)$*

$$\left| E[f(\sqrt{T}(\hat{\theta}_T^* - \theta_0))] - \int \mathrm{d}y^{(0)} f(y^{(0)}) q_{T,2}(y^{(0)}) \right| \leq c\omega(f, \tilde{C}T^{-(\tilde{\varepsilon}+2)/2}, \hat{\rho}^{ab}) + o(T^{-1}), \tag{9}$$

where

$$\begin{aligned} q_{T,2}(y^{(0)}) = \phi(y^{(0)}; \rho^{ab}) \Bigg(& 1 + \frac{1}{6\sqrt{T}} c^{abc} h_{abc}(y^{(0)}; \rho^{ab}) \\ & + \frac{1}{\sqrt{T}} \left(\tilde{\mu}^{a;}{}_{cd} \rho^{cd} - \tilde{\beta}^{a} \right) h_a(y^{(0)}; \rho^{ab}) + \frac{1}{2T} A^{*ab} h_{ab}(y^{(0)}; \rho^{ab}) \\ & + \frac{1}{24T} c^{abcd} h_{abcd}(y^{(0)}; \rho^{ab}) + \frac{1}{72T} c^{abc} c^{def} h_{abcdef}(y^{(0)}; \rho^{ab}) \Bigg), \end{aligned}$$

$$\begin{aligned} \tilde{\beta}^{a;} &= \beta^{a;} - \tilde{\Delta}^{a;}, \quad c^{abc} = \bar{\lambda}^{*abc} + 6\tilde{\mu}^{*c;}{}_{a'b'} \rho^{a'a} \rho^{b'b}, \\ A^{*ab} &= \tilde{\tau}^{ab} + 2\left(\bar{\lambda}^{*acd} + \tilde{\mu}^{*a;}{}_{c'd'} \rho^{c'c} \rho^{d'd} \right) \tilde{\mu}^{*b;}{}_{cd} + 2\delta^{c}_{c'} \tilde{N}^{*a;}{}_{,}{}^{c';}{}_{,}{}^{b;}{}_{c} + \rho^{cc'} \tilde{M}^{*b;}{}_{c,}{}^{a;}{}_{c'} \\ & \quad + 2\left(\left(\tilde{\Delta}^{c;} \tilde{\eta}^{*a;}{}_{c,b'} - \delta_{b'} \beta^{a} \right) + \delta^{a_1}_{b_1} \tilde{M}^{*a;}{}_{a_1,}{}^{b_1;}{}_{b'} + 3U^{*a;}{}_{cdb'} \rho^{cd} \right) \rho^{b'b} \\ & \quad + (\tilde{\mu}^{*a;}{}_{cd} \rho^{cd} - \tilde{\beta}^{a}) \left(\tilde{\mu}^{*b;}{}_{ef} \rho^{ef} - \tilde{\beta}^{b} \right), \\ c^{abcd} &= H^{*abcd} + 4c^{abc} \left(\tilde{\mu}^{*d;}{}_{ef} \rho^{ef} - \tilde{\beta}^{d} \right) + 24 \left(\bar{\lambda}^{*abe} + 2\tilde{\mu}^{*a;}{}_{b'e'} \rho^{b'b} \rho^{e'e} \right) \tilde{\mu}^{*c;}{}_{d'e} \rho^{d'd} \\ & \quad + 12 \left(\rho^{bb'} \rho^{dd'} \tilde{M}^{*c;}{}_{d',}{}^{a;}{}_{b'} + \tilde{N}^{*a;}{}_{,}{}^{b;}{}_{,}{}^{c;}{}_{d'} \rho^{dd'} \right) + 24U^{*a;}{}_{b'c'd'} \rho^{b'b} \rho^{c'c} \rho^{d'd}. \end{aligned}$$

Remark 1 In Theorem 2, it is implicitly assumed that V, V_0 are of class C_b^∞, the set of smooth functions whose derivatives of positive order are bounded, and $B_{a;A}(\cdot, \theta_0)$, $C_{a;A}(\cdot, \theta_0)$, $|A| \leq 2$, are C^∞-functions on $\mathbb{R}^d$ with all derivatives having at most polynomial growth order for Condition (iii) in [L], while Condition [DM2] in Theorem 2 requires that $G\langle C_{a;a_1 \cdots a_k} \rangle(\cdot, \theta_0) \in C^2(\mathbb{R}^d)$ ($k = 0, 1, 2$; $a, a_1, \ldots, a_k = 1, \ldots, p$), etc. [We only consider a classical solution to the Poisson equation, not a weak solution in the distribution theory.] Under the ellipticity assumption for VV', it is known that the smoothness of function f is transferred to the solution $G\langle f \rangle$ of the Poisson equation. In one-dimensional case, $G\langle \cdot \rangle$ is just a duple integral operator and has an explicit expression. Then it is easy to see the smoothness of $G\langle f \rangle$. See Yoshida (1997), where the growth rate is also presented. On the other hand, we should note that introducing the Poisson equation here is only for convenience of giving a closed form of the coefficients in the asymptotic expansion formula. The existence of those coefficients can be verified by the mixing assumption, without the assumptions of Poisson equations if we do not require closed forms given by Theorem 2. Also, it is possible to construct a solution to the Poisson equation for a zero-mean function as an integral of the semigroup under the mixing condition [see Theorem 3 of Kusuoka and Yoshida (2000), also Pardoux and Veretennikov (2001, 2003) for more explicit presentation]. In robust estimation, the estimating function is often constructed by giving a function G for $G\langle f\rangle$. In a standard case of the maximum likelihood estimator for a correctly specified model, it is possible to replace the second-order coefficient expressed through a Poisson equation by a consistent empirical estimator, so that the existence of the coefficient is sufficient in practice up to the second order under studentization if necessary.

Remark 2 Condition [L](iii), which is referred to as Hörmander type condition, ensures the non-degeneracy of the distribution. It requires only differentiation of coefficient vector fields, and is practically convenient. Instead of this condition, we can use other mild conditions which guarantee local degeneracy of the Malliavin covariance. If the Malliavin covariance is nondegenerate at a skeleton in the support of the process, then the local degeneracy in the vicinity follows. See Yoshida (2004) for details. The inifinite differentiability assumption can be relaxed under a stronger nondegeneracy condition.

The asymptotic expansion of the maximum likelihood estimator can be easily derived from this result, for the misspecified or specified case. Here we confine ourselves to the specified case for the sake of simplicity. Suppose that observation X satisfies (2) with $\theta = \theta_0$ and that the estimating function ψ is the derivative of the log-likelihood function (3). For the key functions A, B, C, we then have

$$A_{a;}(x,\theta) = \frac{\partial}{\partial\theta^a}\left(\log\frac{\mathrm{d}\nu_\theta}{\mathrm{d}\nu_*}(x)\right), \quad B_{a;}(x,\theta) = \frac{\partial}{\partial\theta^a}\left(\tilde{V}_0'(\tilde{V}\tilde{V}')^{-1}\tilde{V}(x,\theta)\right),$$
$$C_{a;}(x,\theta) = \frac{\partial}{\partial\theta^a}\left(\tilde{V}_0'(\tilde{V}\tilde{V}')^{-1}(x,\theta)(\tilde{V}_0(x,\theta_0) - \frac{1}{2}\tilde{V}_0(x,\theta))\right).$$

For the *diffusion MLE formula*, we put

$$\check{F}_{A_1,A_2} = \nu(B_{A_1}\cdot B_{A_2}), \quad \check{F}_{A_1,[A_2,A_3]} = \nu(B_{A_1}\cdot[B_{A_2}\cdot B_{A_3}]),$$
$$\check{F}_{[A_1,A_2],[A_3,A_4]} = \nu([B_{A_1}\cdot B_{A_2}]\cdot[B_{A_3}\cdot B_{A_4}]),$$
$$\check{F}_{[[A_1,A_2],A_3],A_4} = \nu([[B_{A_1}\cdot B_{A_2}]\cdot B_{A_3}]\cdot B_{A_4}),$$

where B_A's are evaluated at $\theta=\theta_0$. By using these $\check{F}$, we define ρ_{ab}, ρ^{ab}, $\tilde{\Gamma}^{(\alpha)}_{ab,c}$, $\check{\mu}^a$, and $\tilde{\eta}^{*a}_{b,c}$ by $\rho_{ab} = \check{F}_{a,b}$, $(\rho^{ab}) = (\rho_{ab})^{-1}$,

$$\tilde{\Gamma}^{(\alpha)}_{ab,c} = \check{F}_{ab,c} - \check{F}_{[a,b],c} + \frac{1-\alpha}{2}\sum_{(ab,c)}^{[3]}\check{F}_{[a,b],c},$$
$$\check{\mu}^a = -\frac{1}{2}\rho^{aa'}\rho^{bc}\tilde{\Gamma}^{(-1)}_{bc,a'}, \quad \text{and} \quad \tilde{\eta}^{*a}_{b,c} = -\rho^{aa'}\left(\tilde{\Gamma}^{(1)}_{a'c,b} + \tilde{\Gamma}^{(-1)}_{bc,a'}\right).$$

Moreover, let

$$\tilde{\Delta}^a = \rho^{aa'}\nu\left(\frac{\partial}{\partial\theta^{a'}}\frac{\mathrm{d}\nu_\theta}{\mathrm{d}x}\bigg|_{\theta=\theta_0}\right), \quad \zeta_{ab} = \nu\left(\frac{\partial^2}{\partial\theta^a\partial\theta^b}\log\frac{\mathrm{d}\nu_\theta}{\mathrm{d}x}\bigg|_{\theta=\theta_0}\right),$$
$$\tau_{ab} = \mathrm{Cov}\left[\frac{\partial}{\partial\theta^a}\log\frac{\mathrm{d}\nu_\theta}{\mathrm{d}x}\bigg|_{\theta=\theta_0}(X_0), \frac{\partial}{\partial\theta^b}\log\frac{\mathrm{d}\nu_\theta}{\mathrm{d}x}\bigg|_{\theta=\theta_0}(X_0)\right].$$

Let $h^{a_1\cdots a_k}(x;\sigma) = \sigma^{a_1b_1}\cdots\sigma^{a_kb_k}h_{b_1\cdots b_k}(x;\sigma)$ for a positive definite matrix $\sigma = (\sigma^{ab})$.

With these notations, the *diffusion MLE formula* is obtained as follows:

Theorem 3 *Let M, $\gamma > 0$, and $\hat{\rho} > (\rho^{ab})$. Assume the same conditions as in Theorem 2 for the diffusion process X and the estimating function $\psi = \partial\ell/\partial\theta$. For any $\beta \in C_B^2(\Theta)$ and the extended $\hat{\theta}_T$, let $\hat{\theta}_T^* = \hat{\theta}_T - \beta(\hat{\theta}_T)/T$. Then there exist positive constants $c, \tilde{C}, \tilde{\varepsilon}$ such that for any $f \in \mathscr{E}(M,\gamma)$*

$$\left|E[f(\sqrt{T}(\hat{\theta}_T^* - \theta_0))] - \int \mathrm{d}y^{(0)} f(y^{(0)})q_{T,2}(y^{(0)})\right| \le c\omega(f, \tilde{C}T^{-(\tilde{\varepsilon}+2)/2}, \hat{\rho}^{ab}) + o(T^{-1}), \tag{10}$$

where

$$\begin{aligned}
q_{T,2}(y^{(0)}) &= \phi(y^{(0)};\rho^{ab})\Bigg(1 + \frac{1}{6\sqrt{T}}c^*_{abc}h^{abc}(y^{(0)};\rho^{ab}) \\
&\quad + \frac{1}{\sqrt{T}}\rho_{aa'}(\check{\mu}^{a'} - \tilde{\beta}^{a'})h^a(y^{(0)};\rho^{ab}) + \frac{1}{2T}A^*_{ab}h^{ab}(y^{(0)};\rho^{ab}) \\
&\quad + \frac{1}{24T}c^*_{abcd}h^{abcd}(y^{(0)};\rho^{ab}) + \frac{1}{72T}c^*_{abc}c^*_{def}h^{abcdef}(y^{(0)};\rho^{ab})\Bigg), \\
c^*_{abc} &= -3\tilde{\Gamma}^{(-1/3)}_{ab,c}, \quad \tilde{\beta}^a = \beta^a - \Delta^a, \\
A^*_{ab} &= \tau_{ab} + 2\zeta_{ab} - \rho^{cd}\Big(\check{F}_{bcd,a} + \check{F}_{ab,cd} - \check{F}_{ac,bd} - \check{F}_{[a,c],[b,d]} + 2\check{F}_{[ab,c],d} \\
&\quad + 2\check{F}_{[ac,b],d} + 4\check{F}_{[b,d],ac} + \check{F}_{[cd,b],a} + 2\check{F}_{[[b,c],a],d} + 2\check{F}_{[[b,c],d],a}\Big) \\
&\quad + \rho^{cd}\rho^{ef}\Bigg(\frac{1}{2}\tilde{\Gamma}^{(-1)}_{ce,b}\tilde{\Gamma}^{(-1)}_{df,a} - \tilde{\Gamma}^{(1)}_{ac,e}\tilde{\Gamma}^{(1)}_{bd,f} + \tilde{\Gamma}^{(-1)}_{cd,e}\left(\tilde{\Gamma}^{(1)}_{ab,f} + \tilde{\Gamma}^{(-1)}_{fb,a}\right) \\
&\quad + \tilde{\Gamma}^{(-1)}_{ce,a}\left(\tilde{\Gamma}^{(1)}_{bd,f} + \tilde{\Gamma}^{(-1)}_{bd,f}\right)\Bigg) \\
&\quad + \rho_{aa'}\rho_{bb'}(\check{\mu}^{a'} - \tilde{\beta}^{a'})(\check{\mu}^{b'} - \tilde{\beta}^{b'}) + 2\rho_{aa'}\left(\Delta^c\tilde{\eta}^{*a'}_{c,b} - \delta_b\beta^{a'}\right), \\
c^*_{abcd} &= -12\left(\check{F}_{[[a,b],c],d} + \check{F}_{[a,b],cd} + \check{F}_{[ab,c],d}\right) + 3\check{F}_{[a,b],[c,d]} - 4\check{F}_{abc,d} \\
&\quad + 12\tilde{\Gamma}^{(-1/3)}_{ab,c}\rho_{dd'}(\tilde{\beta}^{d'} - \check{\mu}^{d'}) + 12\rho^{ef}\left(\tilde{\Gamma}^{(-1)}_{ab,e} + \tilde{\Gamma}^{(1)}_{ae,b}\right)\tilde{\Gamma}^{(-1)}_{cf,d}.
\end{aligned}$$

Remark 3 In Theorem 3, the representation of the coefficients in the expansion are obtained without the Bartlett type identities [BI1]–[BI4], [DV1]–[DV3] in Sakamoto and Yoshida (2004). If one assumes those identities, the representation will become the same one as Sakamoto and Yoshida (1998b).

Remark 4 In case one applies this third-order diffusion MLE formula to the Ornstein-Uhlenbeck process, it turns out that Condition (i) of [L] is not satisfied due to the complete linearity of this exceptional model. However, Sakamoto and Yoshida showed in 2000 that even in such a case, the third-order diffusion formula of Theorem 3 still holds true as mentioned in Uchida and Yoshida (2001). See Sakamoto and Yoshida (2003) and Remark 5 in Sakamoto and Yoshida (2004).

3 Proofs of theorems in Sect. 2

3.1 Cumulants of a mixing process

In this section, we will study the cumulants of a mixing process. The results will be applied to stochastic integrals in the next section which leads to the proof of Theorem 2.

Let $(\tilde{\Omega}, \tilde{\mathscr{F}}, \tilde{P})$ be a probability space with sub σ-fields $\tilde{\mathscr{F}}_I$, where I is any interval in $\mathbb{R}$, satisfying that $\tilde{\mathscr{F}}_I \subset \tilde{\mathscr{F}}_J$ if $I \subset J$. Assume that for $p > 1$. $q > 1$ with $1/p + 1/q < 1$, there exist $\mathsf{a} > 0$ and $\mathsf{b} > 0$ such that

$$|\mathrm{Cov}(F, G)| \leq \mathsf{a}\mathrm{e}^{-\mathsf{b}(t-s)} \|F\|_p \|G\|_q, \tag{11}$$

for any $s, t \in \mathbb{R}$, $s \leq t$, and for any $F \in \mathscr{F}\tilde{\mathscr{F}}_{(-\infty,s]} \cap L_p(\tilde{\Omega})$ and $G \in \mathscr{F}\tilde{\mathscr{F}}_{[t,\infty)} \cap L_q(\tilde{\Omega})$. This inequality is often referred to as covariance inequality, from which the cumulants of measurable functions w.r.t. $\mathscr{F}_{I_i}$ for some intervals I_i can be estimated by the maximal gap of intervals I_i.

Lemma 1 *Let $\varepsilon \geq 0$, $p > 2$, $m, k \in \mathbb{N}$ with $2 \leq k < p$ and $k \vee 3 \leq m$, and let m_1, m_2 and m_3 be positive integers satisfying $m_1 + m_2 + m_3 \leq m$. Suppose that $\{t_i\}_{i=1,\ldots,m}$ is a real valued sequence such that $t_1 = \cdots = t_{m_1} \leq t_{m_1+1} \leq \cdots \leq t_{m_1+m_2} \leq t_{m_1+m_2+1} = \cdots = t_{m_1+m_2+m_3}$ and that $\{G_i\}_{i=1,\ldots,m}$ is a sequence of $\mathbb{R}$-valued random variables such that $G_i \in \mathscr{F}\tilde{\mathscr{F}}_{(-\infty,t_i]} \cap L_p(\tilde{\Omega})$ for $i = 1, \ldots, m_1$, $G_i \in \mathscr{F}\tilde{\mathscr{F}}_{[t_i-\varepsilon,t_i]} \cap L_p(\tilde{\Omega})$ for $i = m_1+1, \ldots, m_1+m_2$, and $G_i \in \mathscr{F}\tilde{\mathscr{F}}_{[t_i-\varepsilon,\infty)} \cap L_p(\tilde{\Omega})$ for $i = m_1 + m_2 + 1, \ldots, m_1 + m_2 + m_3$. Then there exist positive constants b and c depending only on p and k such that for any finite subsequence $i_1 < i_2 < \cdots < i_k$ of $\{1, \ldots, m\}$,*

$$|\mathrm{Cum}[G_{i_1}, \cdots, G_{i_k}]| \leq c\, \mathrm{e}^{-\mathsf{b}((g-\varepsilon)\vee 0)} \prod_{j=1}^{k} \|G_{i_j}\|_p,$$

where $g = \max\{t_{i_{j+1}} - t_{i_j} \mid j = 1, \ldots, k-1\}$.

Proof This assertion is more or less known, but here we provide a proof for selfcontainedness. Let $\mu_{\tilde{A}} = E[G_{\alpha_1} \cdots G_{\alpha_{k'}}]$ for any index set $\tilde{A} = \{\alpha_1, \ldots, \alpha_{k'}\} \subset \{1, \ldots, m\}$. Fix a subsequence $i_1 < \cdots < i_k$ of $\{1, \ldots, m\}$ arbitrarily. Then it follows from Hölder's inequality that for any disjoint decomposition of $\{i_1, \ldots, i_k\}$ into $A_1, \ldots, A_s$,

$$|\mathrm{Cum}[G_{i_1}, \cdots, G_{i_k}]| \le c_1(k) \prod_{j=1}^{k} \|G_{i_j}\|_p, \tag{12}$$

where $c_1(k) = \sum_{i=1}^{k} (i-1)! N_i^k$ and N_i^k is the number of the decompositions of $\{1, \ldots, k\}$ into i parts. Let n be an index in $\{1, \ldots, k-1\}$ such that $g = t_{i_{n+1}} - t_{i_n}$, and let $A^{(1)} = \{i_1, \ldots, i_n\}$ and $A^{(2)} = \{i_{n+1}, \ldots, i_k\}$. In the case where $g \ge \varepsilon$, one can easily show from (11) and Hölder's inequality that for any disjoint decompositions of $\{i_1, \ldots, i_k\}$ into $A_1, \ldots, A_j$ and for any $i = 1, \ldots, j$,

$$\left|\mu_{A_1^{(1)}} \mu_{A_1^{(2)}} \cdots \mu_{A_{i-1}^{(1)}} \mu_{A_{i-1}^{(2)}} \left(\mu_{A_i} - \mu_{A_i^{(1)}} \mu_{A_i^{(2)}}\right) \mu_{A_{i+1}} \cdots \mu_{A_j}\right| \le \mathsf{a} e^{-\mathsf{b}(g-\varepsilon)} \prod_{j=1}^{k} \|G_{i_j}\|_p,$$

for some positive constants a and b depending on p and k, where $A_i^{(1)} = A_i \cap A^{(1)}$ and $A_i^{(2)} = A_i \cap A^{(2)}$. Note that if $A_i^{(1)} = \phi$ or $A_i^{(2)} = \phi$, then $\mu_{A_i} - \mu_{A_i^{(1)}} \mu_{A_i^{(1)}} = 0$. Therefore, we obtain that

$$\begin{aligned}
|\mathrm{Cum}[G_{i_1}, \cdots, G_{i_k}]| = \sum_{j=1}^{k} (-1)^{j-1} (j-1)! \sum_{\{1,\ldots,k\}/j} &\left(\left(\mu_{A_1} - \mu_{A_1^{(1)}} \mu_{A_1^{(2)}}\right) \right.\\
&\times \mu_{A_2} \cdots \mu_{A_j} + \cdots \\
&+ \mu_{A_1^{(1)}} \mu_{A_1^{(2)}} \cdots \mu_{A_{i-1}^{(1)}} \mu_{A_{i-1}^{(2)}} \left(\mu_{A_i} - \mu_{A_i^{(1)}} \mu_{A_i^{(2)}}\right) \\
&\times \mu_{A_{i+1}} \cdots \mu_{A_j} + \cdots + \mu_{A_1^{(1)}} \mu_{A_1^{(2)}} \cdots \mu_{A_{j-1}^{(1)}} \mu_{A_{j-1}^{(2)}} \\
&\left. \times \left(\mu_{A_j} - \mu_{A_j^{(1)}} \mu_{A_j^{(2)}}\right) \right) \\
\le c_1(k)\, c_1' e^{-\mathsf{b}(g-\varepsilon)} \prod_{j=1}^{k} \|G_{i_j}\|_p,
\end{aligned} \tag{13}$$

for some positive constant c_1' depending on p and k. Here $\sum_{\{1,\ldots,k\}/j}$ stands for the summation over all decompositions of $\{1, \ldots, k\}$ into j disjoint nonempty parts $A_1, \ldots, A_j$. Combining this with (12), we obtain the desired result. □

The cumulants of a process whose increments are measurable w.r.t. $\mathscr{F}_{I_i}$ for disjoint intervals I_i are evaluated by this lemma.

Proposition 1 *Let $p > 2$, $k \in \mathbb{N}$ with $2 \le k < p$, $F_0 \in \mathscr{F}\tilde{\mathscr{F}}_{[0]} \cap L_p(\tilde{\Omega} : \mathbb{R}^d)$, and let $G = (G_t)_{t \in \mathbb{R}_+}$ and $H = (H_t)_{t \in \mathbb{R}_+}$ be $\mathbb{R}^d$-valued processes such that $G_t - G_s \in \mathscr{F}\tilde{\mathscr{F}}_{[s,t]} \cap L_p(\tilde{\Omega} : \mathbb{R}^d)$ for any $s \le t$, $H_t \in \mathscr{F}\tilde{\mathscr{F}}_{[t,\infty)} \cap L_p(\tilde{\Omega} : \mathbb{R}^d)$ for any $t \ge 0$. Suppose that $\sup_{s<t} E|(G_t - G_s)/\sqrt{t-s}|^p := \beta_p < \infty$, $\sup_t E|H_t|^p < \infty$. Denote $G_T/\sqrt{T}$ and $(F_0 + G_T + H_T)/\sqrt{T}$ by $\bar{G}_T$ and $\bar{\psi}_T$, respectively. Then, for any $T > 0$ and any index set $\{a_1, \ldots, a_k\}$, $a_i = 1, \ldots, d$,*

$$\mathrm{Cum}\left[\bar{\psi}_T^{a_1}, \ldots, \bar{\psi}_T^{a_k}\right] = \mathrm{Cum}\left[\bar{G}_T^{a_1}, \ldots, \bar{G}_T^{a_k}\right] + T^{-k/2} R_T^{a_1 \cdots a_k}, \tag{14}$$

where $\bar{\psi}_T^a$ and $\bar{G}_T^a$ are a-th elements of $\bar{\psi}_T$ and $\bar{G}_T$, respectively, and $R_T^{a_1 \cdots a_k}$ is some constant bounded as $T \to \infty$. Furthermore, there exists a positive constant $c(p, k)$ depending only on p and k such that

$$\left|\mathrm{Cum}\left[\bar{G}_T^{a_1}, \ldots, \bar{G}_T^{a_k}\right]\right| \le c(p,k) \beta_p^{k/p} T^{-(k-2)/2}.$$

In particular, R_T^{ab} and R_T^{abc} satisfy

$$\begin{aligned}
R_T^{ab} &= \mathrm{Cov}[F_0^a, F_0^b] + \mathrm{Cov}[H_T^a, H_T^b] + \sum_{(a,b)}^{[2]} (\mathrm{Cov}[F_0^a, G_T^b] \\
&\quad + \mathrm{Cov}[G_T^a, H_T^b]) + O(\mathrm{e}^{-\mathsf{b}T}), \\
R_T^{abc} &= \mathrm{Cum}[F_0^a, F_0^b, F_0^c] + \mathrm{Cum}[H_T^a, H_T^b, H_T^c] + \sum_{(ab,c)}^{[3]} \mathrm{Cum}[F_0^a, F_0^b, G_T^c] \\
&\quad + \sum_{(a,bc)}^{[3]} \mathrm{Cum}[F_0^a, G_T^b, G_T^c] + \sum_{(ab,c)}^{[3]} (\mathrm{Cum}[G_T^a, G_T^b, H_T^c] \\
&\quad + \mathrm{Cum}[G_T^a, H_T^b, H_T^c]) + O(T \mathrm{e}^{-\mathsf{b}T/2}),
\end{aligned}$$

for some $\mathsf{b} > 0$, where a, b, $c = 1, \ldots, d$, and F_0^a, G_T^a, and H_T^a are a-th elements of F_0, G_T and H_T, respectively.

Proof (a) First, we consider the inequality

$$\left|\mathrm{Cum}\left[\bar{G}_T^{a_1}, \ldots, \bar{G}_T^{a_k}\right]\right| \le c(p,k) \beta_p^{k/p} T^{-(k-2)/2}.$$

This estimation is fairly familiar; however, we present here the proof to facilitate the understanding of the second half (b). Let $(\Delta G_i)_{i=0,1,\ldots,[T]+1}$ be a sequence of random variables defined by

$$\Delta G_i = \begin{cases} G_0 & (i = 0) \\ G_i - G_{i-1} & (1 \le i \le [T]) \\ G_T - G_{[T]} & (i = [T] + 1) \end{cases}.$$

Note that $\Delta G_i \in \mathscr{F}\tilde{\mathscr{F}}_{[i-1,i]}$, $i = 0, \ldots, [T]+1$. From the multilinearity of the cumulant, we have

$$
\begin{aligned}
&\left|\mathrm{Cum}\left[\bar{G}_T^{a_1}, \ldots, \bar{G}_T^{a_k}\right]\right| \\
&\le \left| T^{-k/2} \sum_{i_1=0}^{[T]+1} \cdots \sum_{i_k=0}^{[T]+1} \mathrm{Cum}\left[\Delta G_{i_1}^{a_1}, \ldots, \Delta G_{i_k}^{a_k}\right]\right| \\
&\le T^{-k/2} \sum_{m=1}^{[T]+1} \sum_{M=m}^{[T]+1} \sum_{\{i_1,\ldots,i_k\}\in \mathscr{I}_1^k(m,M)} \left|\mathrm{Cum}\left[\Delta G_{i_1}^{a_1}, \ldots, \Delta G_{i_k}^{a_k}\right]\right|,
\end{aligned}
$$

where $\mathscr{I}_1^k(m, M) = \{\{i_1, \ldots, i_k\} \mid \min(i_1, \ldots, i_k) = m, \max(i_1, \ldots, i_k) = M\}$. Note that the number of elements of $\mathscr{I}_1^k(m, M)$ is equal to $N_1^k(M-m)$, where $N_1^k(x) = (x+1)^k - 2x^k + (x-1)^k$ if $x \ge 1$, and $N_1^k(x) = 1$ if $x = 0$. Applying Lemma 1 with $\varepsilon = 1$, we see that there exist positive constants b and c_2 depending only on p and k such that for any index set $\{i_1, \ldots, i_k\} \in \mathscr{I}_1^k(m, M)$,

$$
\left|\mathrm{Cum}\left[\Delta G_{i_1}^{a_1}, \ldots, \Delta G_{i_k}^{a_k}\right]\right| \le c_2\, \mathrm{e}^{-\mathsf{b}((g-1)\vee 0)} \beta_p^{k/p},
$$

where $g = \max\{i_{(j+1)} - i_{(j)} \mid j = 1, \ldots, k-1, \}$, $\{i_{(1)}, \ldots i_{(k)}\}$ is an index set satisfying $\{i_{(1)}, \ldots i_{(k)}\} = \{i_1, \ldots, i_k\}$ and $i_{(1)} \le \cdots \le i_{(k)}$. Since $g \ge (M-m)/(k-1)$, we obtain that

$$
\begin{aligned}
&\sum_{m=1}^{[T]+1} \sum_{M=m}^{[T]+1} \sum_{\{i_1,\ldots,i_k\}\in \mathscr{I}_1^k(m,M)} \left|\mathrm{Cum}\left[\Delta G_{i_1}^{a_1}, \ldots, \Delta G_{i_k}^{a_k}\right]\right| \\
&\le c_2\, \beta_p^k \sum_{m=1}^{[T]+1} \sum_{M=m}^{[T]+1} N_1^k(M-m) \mathrm{e}^{-\mathsf{b}(((M-m)/(k-1)-1)\vee 0)} \\
&\le c_2\, \beta_p^k \sum_{m=1}^{[T]+1} \left(\sum_{M=m}^{m+k-2} N_1^k(M-m) + \sum_{M=m+k-1}^{[T]+1} N_1^k(M-m) \mathrm{e}^{-\mathsf{b}((M-m)/(k-1)-1)} \right) \\
&\le c_2\, \beta_p^k \sum_{m=1}^{[T]+1} \left(\sum_{M'=0}^{k-2} N_1^k(M') + \sum_{M'=k-1}^{[T]-m+1} N_1^k(M') \mathrm{e}^{\mathsf{b}} \mathrm{e}^{-\mathsf{b}M'/(k-1)} \right) \le c_2\, c_3\, \beta_p^{k/p} T
\end{aligned}
$$

for some positive constant c_3 depending only on k. Thus we have that

$$
\mathrm{Cum}\left[\bar{G}_T^{a_1}, \ldots, \bar{G}_T^{a_k}\right] \le c(p,k) \beta_p^{k/p} T^{-(k-2)/2}.
$$

(b) Next, we will consider the remainder term $R_T^{a_1\cdots a_k}$. Let $A=\{a_1,\ldots,a_k\}$, and for any disjoint decomposition $A_1\cup A_2=A$, let

$$\begin{aligned}
\mathrm{Cum}^{F,G}_{A_1,A_2}&=\mathrm{Cum}\left[F_0^{a_{11}},\ldots,F_0^{a_{1m}},G_T^{a_{21}},\ldots,G_T^{a_{2,k-m}}\right]\\
\mathrm{Cum}^{F,H}_{A_1,A_2}&=\mathrm{Cum}\left[F_0^{a_{11}},\ldots,F_0^{a_{1m}},H_T^{a_{21}},\ldots,H_T^{a_{2,k-m}}\right]\\
\mathrm{Cum}^{G,H}_{A_1,A_2}&=\mathrm{Cum}\left[G_T^{a_{11}},\ldots,G_T^{a_{1m}},H_T^{a_{21}},\ldots,H_T^{a_{2,k-m}}\right],
\end{aligned}$$

where $A_1=\{a_{11},\ldots,a_{1m}\}$ and $A_2=\{a_{21},\ldots,a_{2,k-m}\}$. Moreover, for any disjoint decomposition $A_1\cup A_2\cup A_3=A$, let

$$\mathrm{Cum}^{F,G,H}_{A_1,A_2,A_3}=\mathrm{Cum}\left[F_0^{a_{11}},\ldots,F_0^{a_{1m}},G_T^{a_{21}},\ldots,G_T^{a_{2l}},H_T^{a_{31}},\ldots,H_T^{a_{3,k-m-l}}\right],$$

where $A_1=\{a_{11},\ldots,a_{1m}\}$, $A_2=\{a_{21},\ldots,a_{2l}\}$ and $A_3=\{a_{31},\ldots,a_{3,k-l-m}\}$. Then it follows from the multilinearity of the cumulant that

$$\mathrm{Cum}\left[\bar\psi_T^{a_1},\ldots,\bar\psi_T^{a_k}\right]=\mathrm{Cum}\left[\bar G_T^{a_1},\ldots,\bar G_T^{a_k}\right]+T^{-k/2}R_T^{a_1\cdots a_k},$$

where

$$\begin{aligned}
R_T^{a_1\cdots a_k}&=\mathrm{Cum}\left[F_0^{a_1},\ldots,F_0^{a_k}\right]+\mathrm{Cum}\left[H_T^{a_1},\ldots,H_T^{a_k}\right]\\
&\quad+\sum_{A/2}\Big(\mathrm{Cum}^{F,G}_{A_1,A_2}+\mathrm{Cum}^{F,G}_{A_2,A_1}+\mathrm{Cum}^{F,H}_{A_1,A_2}+\mathrm{Cum}^{F,H}_{A_2,A_1}\\
&\quad+\mathrm{Cum}^{G,H}_{A_1,A_2}+\mathrm{Cum}^{G,H}_{A_2,A_1}\Big)\\
&\quad+\sum_{A/3}\Big(\mathrm{Cum}^{F,G,H}_{A_1,A_2,A_3}+\mathrm{Cum}^{F,G,H}_{A_1,A_3,A_2}+\mathrm{Cum}^{F,G,H}_{A_2,A_1,A_3}+\mathrm{Cum}^{F,G,H}_{A_2,A_3,A_1}\\
&\quad+\mathrm{Cum}^{F,G,H}_{A_3,A_1,A_2}+\mathrm{Cum}^{F,G,H}_{A_3,A_2,A_1}\Big).
\end{aligned}\tag{15}$$

In the same way as in the discussion for $\mathrm{Cum}[\bar G_T^{a_1},\ldots,\bar G_T^{a_k}]$, we obtain that

$$\begin{aligned}
&\left|\mathrm{Cum}^{F,G}_{A_1,A_2}\right|\\
&\quad\le\sum_{M=0}^{[T]+1}\sum_{\{i_1,\ldots,i_{k-m}\}\in\mathscr{I}_2^{k-m}(M)}\left|\mathrm{Cum}\left[F_0^{a_{11}},\ldots,F_0^{a_{1m}},\Delta G_{i_1}^{a_{21}},\ldots,\Delta G_{i_{k-m}}^{a_{2,k-m}}\right]\right|
\end{aligned}$$

and

$$\left|\mathrm{Cum}^{G,H}_{A_1,A_2}\right|\le\sum_{M=0}^{[T]+1}\sum_{\{i_1,\ldots,i_m\}\in\mathscr{I}_3^{m}(M)}\left|\mathrm{Cum}\left[\Delta G_{i_1}^{a_{11}},\ldots,\Delta G_{i_m}^{a_{1m}},H_T^{a_{21}},\ldots,H_T^{a_{2,k-m}}\right]\right|,$$

where

$$\mathscr{I}_2^l(M) = \{\{i_1, \dots, i_l\} \mid 0 \le i_j \le M,\ j = 1, \dots, l,\ \max(i_1, \dots, i_l) = M\},$$
$$\mathscr{I}_3^l(M) = \{\{i_1, \dots, i_l\} \mid M \le i_j \le [T] + 1,\ j = 1, \dots, l,\ \min(i_1, \dots, i_l) = M\}.$$

The numbers of elements of $\mathscr{I}_2^l(M)$ and $\mathscr{I}_3^l(m)$ are given by $N_2^l(M) = (M+1)^l - M^l$ and $N_3^l(M) = ([T] + 2 - M)^l - ([T] + 1 - M)^l$, respectively. Applying Lemma 1 with $\varepsilon = 1$ again, we obtain that

$$\begin{aligned}\left|\mathrm{Cum}_{A_1,A_2}^{F,G}\right| &\le \sum_{M=0}^{[T]+1} N_2^{k-m}(M) c_2\, \mathrm{e}^{-\mathsf{b}((M/(k-m)-1)\vee 0)} \|F_0\|_p^m \beta_p^{(k-m)/p}\\ &\le c_4(p,k)\, \|F_0\|_p^m \beta_p^{(k-m)/p},\end{aligned} \tag{16}$$

where $c_4(p, k)$ is a positive constant depending only on p and k. Similarly, we have

$$\begin{aligned}\left|\mathrm{Cum}_{A_1,A_2}^{G,H}\right| &\le \sum_{M=0}^{[T]+1} N_3^k(M) c_2 \mathrm{e}^{-\mathsf{b}(([T]+1-M)/k-1)\vee 0)} \beta_p^{k/p} \sup_T \|H_T\|_p^{k-m}\\ &\le c_5(p,k)\beta_p^{k/p} \sup_T \|H_T\|_p^{k-m}\end{aligned} \tag{17}$$

for some positive constant $c_5(p, k)$. Furthermore,

$$\begin{aligned}\left|\mathrm{Cum}_{A_1,A_2,A_3}^{F,G,H}\right| &\le \sum_{i_1=0}^{[T]+1} \cdots \sum_{i_l=0}^{[T]+1} \Big|\mathrm{Cum}\Big[F_0^{a_{11}}, \dots, F_0^{a_{1m}},\\ &\qquad \Delta G_{i_1}^{a_{21}}, \dots, \Delta G_{i_l}^{a_{2l}}, H_T^{a_{31}}, \dots, H_T^{a_{3,k-m-l}}\Big]\Big|\\ &\le ([T]+1)^l c_2\, \mathrm{e}^{-\mathsf{b}((([T]+1)/(l+1)-1)\vee 0)} \|F_0\|_p^k \beta_p^{l/p} \sup_T \|H_T\|_p^{k-l-m}\\ &\le c_6(l,p,k) T^l \mathrm{e}^{-\mathsf{b}T/(l+1)} \|F_0\|_p^k \beta_p^{l/p} \sup_T \|H_T\|_p^{k-l-m}\end{aligned} \tag{18}$$

for some positive constant $c_6(l, p, k)$. Since

$$\left|\mathrm{Cum}_{A_1,A_2}^{F,H}\right| \le c_2\, \mathrm{e}^{-\mathsf{b}T} \|F_0\|_p^m \sup_T \|H_T\|_p^{k-m}, \tag{19}$$

we see that $R_T^{a_1 \cdots a_k} = O(1)$ as $T \to \infty$. From (15), (18) and (19), the representations of R_T^{ab} and R_T^{abc} are obtained. □

Corollary 1 *In addition to the conditions of Proposition 1, suppose that for any $T > 0$, $E[G_T] = 0$. Then, for any $T > 0$ and any index set $\{a_1, \dots, a_k\}$, $a_i = 1, \dots, d$,*

$$\left|E\left[\bar{G}_T^{a_1} \cdots \bar{G}_T^{a_k}\right]\right| \le M\beta_p^{k/p}, \tag{20}$$

where

$$M=\sum_{j=1}^{k}\sum_{\substack{A_1+\cdots+A_j=\{1,\ldots,k\}\\ |A_s|\geq 2}}\prod_{i=1}^{j}c(p,|A_i|),$$

and $c(p,k)$ *and* β_p *are positive constants in the statement of Proposition* 1.

Proof *For any index set* $A=\{i_1,\ldots,i_s\}$, $i_m\in\{1,\ldots,k\}$, *let* $\kappa_A=\mathrm{Cum}[\bar{G}_T^{a_{i_1}},\ldots,\bar{G}_T^{a_{i_s}}]$. *Then*

$$\begin{aligned}\left|E\left[\bar{G}_T^{a_1}\cdots\bar{G}_T^{a_k}\right]\right| &\leq \sum_{j=1}^{k}\sum_{A_1+\cdots+A_j=\{1,\ldots,k\}}|\kappa_{A_1}\cdots\kappa_{A_j}| \\ &\leq \sum_{j=1}^{k}\sum_{\substack{A_1+\cdots+A_j=\{1,\ldots,k\}\\ |A_s|\geq 2}}\Pi_{i=1}^{j}c(p,|A_i|)\beta_p^{|A_i|/p}T^{-|A_i|/2+1} \\ &\leq \beta_p^{k/p}\,T^{-k/2+[k/2]}\sum_{j=1}^{k}\sum_{\substack{A_1+\cdots+A_j=\{1,\ldots,k\}\\ |A_s|\geq 2}}\Pi_{i=1}^{j}c(p,|A_i|)\leq M\beta_p^{k/p}.\end{aligned}$$

□

3.2 Proof of Theorem 1

We will apply Theorem 6.2 of Sakamoto and Yoshida (2004) to prove Theorem 1. We note that θ_0 above does not always play a role of specifying the true model because the statistical model (2) may not include the true equation (1). Therefore, the model considered here is more general than that in Sect. 6 of Sakamoto and Yoshida (2004). But all of the results in Sect. 6 of Sakamoto and Yoshida (2004) still hold true for the model here because the difference between these models is not essential in the proofs.

For convenience of explanation, here we write down the conditions for Theorem 6.2 of Sakamoto and Yoshida (2004): for $K\in\mathbb{N}$, $q>1$ and $\gamma>0$:

[C0]^K $\psi\in C^K(\Theta)$ a.s.;
[C1]$_q$ $\sup_{T>T_0}\left\|r_T\psi_{a;}(\theta_0)\right\|_q<\infty$ for $a=1,\ldots,p$;
[C2]$^K_{q,\gamma}$

$$\sup_{T>T_0,\ \theta\in\Theta}\left\|r_T^{-\gamma}\left(r_T^2\psi_{a;a_1\cdots a_K}(\theta)-\bar{\nu}_{a;a_1\cdots a_K}(\theta)\right)\right\|_q<\infty;$$

[C3] There exists an open set $\tilde{\Theta}$ including θ_0 such that

$$\inf_{T>T_0,\ \theta_1,\theta_2\in\tilde{\Theta},\ |x|=1}\left|x'\left(\int_0^1\bar{\nu}_{a;b}(\theta_1+s(\theta_2-\theta_1))\mathrm{d}s\right)\right|>0;$$

$[\mathrm{C4}]_q^K$ $\sup_{T>T_0} \left\| \sup_{\theta\in\Theta} |r_T^2 \psi_{a;a_1\cdots a_K}(\theta)| \right\|_q < \infty$ for $a, a_j = 1, \ldots, p$, $j = 1, \ldots, K$.

Here $r_T = T^{-1/2}$, $\bar{\nu}_{a;a_1\cdots a_k}(\theta) = E[\psi_{a;a_1\cdots a_k}(\theta)]$. In Theorem 6.2 of Sakamoto and Yoshida (2004), it is assumed that for given $m > 0$ and $\gamma \in (3/4, 1)$, $[\mathrm{C0}]^4$, $[\mathrm{C1}]_{p_1}$, $[\mathrm{C2}]^k_{p_2,\gamma}$, $k = 1, 2, 3$, [C3], and $[\mathrm{C4}]^4_{p_3}$ hold true for some $p_1 > 4m$, $p_2 > \max(p, 4m)$, $p_3 > m$ with $3/4 + \max(m/p_2, m/(4p_3)) < \gamma < 1 - m/p_1$. Moreover, it is also assumed that $\delta_c \bar{\nu}_{a;b}(\theta) = \bar{\nu}_{a;bc}(\theta)$. In the following, we will verify these assumptions under the conditions of Theorem 1.

In general, if a measurable function $f : \mathbb{R}^d \times \Theta \to \mathbb{R}$ satisfies (i) $f(x, \cdot) \in C^2(\Theta)$ for each $x \in \mathbb{R}^d$ and (ii) for any compact set $K \subset \Theta$, there exist positive constants M and m such that $\sum_{j=0}^2 \sup_{\theta\in K} |\delta_\theta^j f(x, \theta)| \le M(1 + |x|)^m$, then $\int_0^T f(X_t, \theta)\mathrm{d}w_t$ is differentiable w.r.t. θ and $\delta_a \int_0^T f(X_t, \theta)\mathrm{d}w_t = \int_0^T \delta_a f(X_t, \theta)\mathrm{d}w_t$. See Kunita (1990). Therefore, we see that under Condition [DM1], Condition $[\mathrm{C0}]^4$ holds true and

$$\psi_{a;a_1\cdots a_k}(\theta) = A_{a;a_1\cdots a_k}(X_0, \theta) + \int_0^T B_{a;a_1\cdots a_k}(X_t, \theta)\mathrm{d}w_t + \int_0^T C_{a;a_1\cdots a_k}(X_t, \theta)\mathrm{d}t$$

for $k = 1, \ldots, 4$. Moreover, we see that for $k = 1, \ldots, 4$,

$$\bar{\nu}_{a;a_1\cdots a_k}(\theta) := \frac{1}{T} E[\psi_{a;a_1\cdots a_k}(\theta)] = \frac{1}{T}\nu(A_{a;a_1\cdots a_k}(\cdot, \theta)) + \nu(C_{a;a_1\cdots a_k}(\cdot, \theta))$$

due to the existence of the moments of X_t, $t \in \mathbb{R}_+$, up to any order. Therefore, the condition in Theorem 1 concerning the differentiability of $\nu(A_{a;b}(\cdot, \theta))$ and $\nu(C_{a;b}(\cdot, \theta))$ w.r.t θ leads $\delta_c \bar{\nu}_{a;b}(\theta) = \bar{\nu}_{a;bc}(\theta)$. Condition [C3] can be easily proved under the condition (6). Hence, if the conditions $[\mathrm{C1}]_{p_1}$, $[\mathrm{C2}]^k_{p_2,\gamma}$, $k = 1, 2, 3$, and $[\mathrm{C4}]^4_{p_3}$ for any $p_1 > 1$, $p_2 > 1$, $p_3 > 1$, $\gamma \in (0, 1)$ are verified, the proof will be completed.

Under Condition [DM1], Burkholder–Davis–Gundy's inequality and Jensen's inequality yield that for any $q > 1$, there exists a positive constant c_q such that for $k = 1, \ldots, 5$, $a_j \in \{1, \ldots, p\}$,

$$\left\| T^{-1/2} \int_0^T B_{a;a_1\cdots a_k}(X_t, \theta) \cdot \mathrm{d}w_t \right\|_q \le c_q \| B_{a;a_1\cdots a_k}(X_0, \theta)\|_q. \tag{21}$$

On the other hand, Condition [DM1] and Corollary 1 lead that for any positive integer $m \ge 2$, $q' > m$

$$\left\| T^{-1/2} \int_0^T \left(C_{a;a_1\cdots a_k}(X_t, \theta) - \nu(C_{a;a_1\cdots a_k}(\cdot, \theta))\right) \mathrm{d}t \right\|_m \le M' \| C_{a;a_1\cdots a_k}(X_0, \theta)\|_{q'}, \tag{22}$$

where M' is a positive constant independent of θ. By using these inequalities with Condition [DM3](i), we see that for any $p_1 > 1$, there exist $c_{p_1} > 0$, $M > 0$, $q' > p_1$

such that

$$\|T^{-1/2}\psi_{a;}(\theta_0)\|_{p_1} \le T^{-1/2}\|A_{a;}(X_0,\theta_0)\|_{p_1} \\ +c_{p_1}\|B_{a;}(X_0,\theta_0)\|_{2p_1} + M\|C_{a;}(X_0,\theta_0)\|_{q'},$$

which prove Condition $[\mathrm{C1}]_{p_1}$.

In the same way, we have that for any $p_2 > 1$, $\gamma \in (0,1)$, $k = 1,2,3$, and $a, a_1, \ldots, a_k = 1, \ldots, p$, there exist $c_{p_2} > 0$, $M' > 0$, $q'' > p_2$ such that

$$\begin{aligned}&\|T^{\gamma/2}(T^{-1}\psi_{a;a_1\cdots a_k}(\theta) - \bar{\nu}_{a;a_1\cdots a_k}(\theta)))\|_{p_2}\\ &\le T^{(\gamma-1)/2}\Big(2T^{-1/2}\|A_{a;a_1\cdots a_k}(X_0,\theta)\|_{p_2} + c_{p_2}\|B_{a;a_1\cdots a_k}(X_0,\theta)\|_{2p_2}\\ &\quad +M'\|C_{a;a_1\cdots a_k}(X_0,\theta)\|_{q''}\Big).\end{aligned}$$

Combining this with Condition [DM1], one can prove Conditions $[\mathrm{C2}]^k_{p_2,\gamma}$, $k=1,2,3$.

Furthermore, we will consider the Condition $[\mathrm{C4}]^4_{p_3}$ for $p_3 > 1$: $\|\sup_{\theta\in\Theta}\psi_{a;A}(\theta)\|_{p_3} < \infty$ for any $a = 1, \ldots, p$, and index set A, $|A| = 4$. Because a continuous version of $\psi_{a;A}(\theta)$, $|A| = 4$, can be chosen owing to [DM1](i), it follows from the GRR inequality that if $\beta + 2p/r < 1$, $\beta > 0$, $r > 0$, and $p = \dim(\Theta)$, there exists a positive constant C_Θ depending p, β, r, and the shape of the boundary of Θ such that

$$\sup_{\theta\in\Theta}|T^{-1}\psi_{a;A}(\theta)| \le C_\Theta \Upsilon(T^{-1}\psi_{a;A})\sup_{\theta\in\Theta}|\theta-\theta_0|^\beta + |T^{-1}\psi_{a;A}(\theta_0)|,$$

where

$$\Upsilon(f) = \left\{\int_\Theta\int_\Theta\left(\frac{|f(\theta_1)-f(\theta_2)|}{|\theta_1-\theta_2|^{\beta+\frac{2p}{r}}}\right)^r \mathrm{d}\theta_1\mathrm{d}\theta_2\right\}^{\frac{1}{r}}$$

for any continuous function $f : \Theta \to \mathbb{R}$. Therefore, putting $\overline{C}_{\Theta,\beta} = C_\Theta \sup_{\theta\in\Theta}|\theta-\theta_0|^\beta$, we have that for any $p_3 > r$,

$$\begin{aligned}\left\|\sup_{\theta\in\Theta}|T^{-1}\psi_{a;A}(\theta)|\right\|_{p_3} &\le \overline{C}_{\Theta,\beta}\Bigg\{T^{-1}\left\|\Upsilon\left(A_{a;A}(X_0,\cdot)\right)\right\|_{p_3}\\ &\quad +T^{-1/2}\left\|\Upsilon\left(T^{-1/2}\int_0^T B_{a;A}(X_t,\cdot)\mathrm{d}w_t\right)\right\|_{p_3}\\ &\quad +\left\|\Upsilon\left(T^{-1}\int_0^T C_{a;A}(X_t,\cdot)\mathrm{d}t\right)\right\|_{p_3}\Bigg\} + \left\|T^{-1}\psi_{a;A}(\theta_0)\right\|_{p_3}.\end{aligned}$$

Applying Burkholder–Davis–Gundy's inequality, we obtain that

$$\left\| \Upsilon \left(T^{-1/2} \int_0^T B_{a;A}(X_t,\cdot)\mathrm{d}w_t \right) \right\|_{p_3}^{p_3} \leq |\Theta|^{\frac{2p_3}{r}-2} \int_\Theta \int_\Theta \frac{(c_{p_3})^{p_3}}{|\theta_1-\theta_2|^{(\beta+\frac{2p}{r})p_3}} \times \| B_{a;A}(X_0,\theta_1) - B_{a;A}(X_0,\theta_2)\|_{p_3}^{p_3} \mathrm{d}\theta_1 \mathrm{d}\theta_2.$$

Since

$$\begin{aligned} \| B_{a;A}(X_0,\theta_1) - B_{a;A}(X_0,\theta_2)\|_{p_3}^{p_3} &\leq \left\| \sum_{a'=1}^{p} \left| \int_0^1 B_{a;Aa'}(X_0;\theta_2+u(\theta_1-\theta_2))\mathrm{d}u \right| \right\|_{p_3}^{p_3} \\ &\quad \times |\theta_1-\theta_2|^{p_3} \\ &\leq C'|\theta_1-\theta_2|^{p_3} \end{aligned}$$

for some constant $C' > 0$, it can be shown that for any $p \geq 1$, $p_3 > r$,

$$\left\| \Upsilon \left(T^{-1/2} \int_0^T B_{a;A}(X_t,\cdot)\mathrm{d}w_t \right) \right\|_{p_3}^{p_3} \leq |\Theta|^{\frac{2p_3}{r}-2} \int_\Theta \int_\Theta \frac{C'(c_{p_3})^{p_3}}{|\theta_1-\theta_2|^{\left(\beta+\frac{2p}{r}\right)p_3-p_3}} \mathrm{d}\theta_1 \mathrm{d}\theta_2 < \infty.$$

In the same fashion, one can show that for any $p_3 > 1$, $\|\Upsilon(A_{a;A}(X_0,\cdot))\|_{p_3} < \infty$ and $\|\Upsilon(T^{-1}\int_0^T C_{a;A}(X_t,\cdot)\mathrm{d}t)\|_{p_3} < \infty$. In this way, Condition $[\mathrm{C4}]_{p_3}^4$ holds true for any $p_3 > 1$ under Conditions [DM1]. Thus the proof is completed. □

3.3 Cumulants of stochastic integrals

Our aim of this section is to derive asymptotic expansions of the cumulants of stochastic integrals in the case where the integrands are functions of a diffusion process with a geometric strong mixing property. For this purpose, we will use the following identities concerned with the moments of stochastic integrals and the Lebesgue integrals of processes, which are not always functions of diffusion processes.

Lemma 2 *Let* $\{C_{ab}\}_{a,b=1,\ldots,q}$ *be a given sequence, and* $\{f_a\}_{a=1,\ldots,q}$ *be* $\mathbb{R}^r$*-valued bounded progressively measurable processes on a probability space* $(\tilde{\Omega}, \tilde{\mathscr{F}}, \tilde{P})$*. Define* $I_a = \frac{1}{\sqrt{T}} \int_0^T f_a(t)\cdot \mathrm{d}w_t$ *and* $J_{ab} = \frac{1}{T}\int_0^T f_a(t)\cdot f_b(t)\mathrm{d}t - C_{ab}$*, where* $w = (w_t)_{t\in\mathbb{R}_+}$ *is a standard Wiener process on* $(\tilde{\Omega}, \tilde{\mathscr{F}}, \tilde{P})$*. Then*

$$E\left[\exp\left(\varepsilon^a I_a - \frac{1}{2}\varepsilon^a\varepsilon^b J_{ab}\right)\right] = \exp\left(\frac{1}{2}\varepsilon^a\varepsilon^b C_{ab}\right). \tag{23}$$

Moreover,

$$E[I_a I_b - J_{ab}] = C_{ab} \tag{24}$$

$$E[I_a I_b I_c - J_{bc} I_a - J_{ca} I_b - J_{ab} I_c] = 0 \tag{25}$$

$$E\left[I_a I_b I_c I_d - \sum_{(ab,c,d)}^{[6]} J_{ab} I_c I_d + \sum_{(ab,cd)}^{[3]} J_{ab} J_{cd} \right] = \sum_{(ab,cd)}^{[3]} C_{ab} C_{cd}. \tag{26}$$

If there exist the expectations in the left-hand size of equalities above, these equalities hold true for unbounded processes $\{f_a\}$.

Proof Let $Y_t = \frac{1}{\sqrt{T}} \int_0^t \varepsilon^a f_a(s) \cdot \mathrm{d}w_s - \frac{1}{2T} \int_0^t \varepsilon^a \varepsilon^b f_a(s) \cdot f_b(s) \mathrm{d}s$. Then it follows from Ito's formula that $\exp(Y_t) = 1 + \frac{1}{\sqrt{T}} \int_0^t \exp(Y_s) \varepsilon^a f_a(s) \cdot \mathrm{d}w_s$, which shows that (23) holds. Differentiating both side of it w.r.t ε successively and substituting $\varepsilon = 0$ into the results provide other equations. By using the ordinary method, we can extend these results to the case where $\{f_a\}$ are unbounded. □

We apply these identities to the case where integrands are functions of the diffusion process defined as follows. Let $(\Omega, \mathscr{F}, P)$ be a probability space, and $X = (X_t)_{t \in \mathbb{R}_+}$ an $\mathbb{R}^d$-valued stationary diffusion process satisfying

$$\mathrm{d}X_t = V_0(X_t)\mathrm{d}t + V(X_t)\mathrm{d}w_t$$

where $V_0 : \mathbb{R}^d \to \mathbb{R}^d$, $V : \mathbb{R}^d \to \mathbb{R}^d \otimes \mathbb{R}^r$, and $w = (w_t)_{t \in \mathbb{R}_+}$ is an $\mathbb{R}^r$-valued standard Wiener process on $(\Omega, \mathscr{F}, P)$. For any interval $I \subset \mathbb{R}_+$, let $\mathscr{F}_I = \sigma[w_t - w_s, X_t \ : \ s, t \in I]$. Assume that (i) there exist $a > 0$, $b > 0$ such that (11) holds true with $\tilde{\mathscr{F}}_I = \mathscr{F}_I$, and (ii) $E|X_t|^k < \infty$ for any $t \in \mathbb{R}_+$, $k \geq 1$. Note that under this assumption $F_0 = f(X_0)$, $G_T = \int_0^T g(X_t) \cdot \mathrm{d}w_t$ and $H_T = h(X_T)$ satisfy the condition of Proposition 1 if f, g and h are measurable function having at most polynomial growth order. Denote by ν the measure of the stationary distribution of X, and write $\nu(f) = \int_{\mathbb{R}^d} f(x)\nu(\mathrm{d}x)$. Moreover, for a function f: $\mathbb{R}^d \to \mathbb{R}$, let G_f be a function such that $\mathscr{A} G_f = f$, and $[f] = -V \cdot \nabla G_{\bar{f}}$, if they exist, where $\bar{f} = f - \nu(f)$.

Lemma 3 *Let* $\{f_a\}_{a=1,\dots,q}$ *be functions:* $\mathbb{R}^d \to \mathbb{R}^r$ *having at most polynomial growth order, and* $I_a = T^{-1/2} \int_0^T f_a(X_t) \cdot \mathrm{d}w_t$, $J_{ab} = T^{-1} \int_0^T (f_a(X_t) \cdot f_b(X_t) - \nu(f_a \cdot f_b))\mathrm{d}t$. *Then*

$$E[I_a I_b] = \nu(f_a \cdot f_b).$$

Assume that there exist $G_{\overline{f_a \cdot f_b}}$ *and* $[f_a \cdot f_b]$ *having at most polynomial growth order, then*

$$E[J_{ab}I_c] = \frac{1}{\sqrt{T}}\nu([f_a \cdot f_b] \cdot f_c) + O\left(\frac{1}{T\sqrt{T}}\right),$$
$$E[J_{ab}J_{cd}] = \frac{1}{T}\nu([f_a \cdot f_b] \cdot [f_c \cdot f_d]) + O\left(\frac{1}{T^2}\right),$$
$$\mathrm{Cum}[I_a, I_b, I_c] = \frac{1}{\sqrt{T}}\sum_{(ab,c)}^{[3]} \nu([f_a \cdot f_b] \cdot f_c) + O\left(\frac{1}{T\sqrt{T}}\right).$$

Moreover, assume in addition that there exist $G_{\overline{[f_a \cdot f_b]\cdot f_c}}$ *and* $[[f_a \cdot f_b] \cdot f_c]$ *having at most polynomial growth order, then*

$$E[J_{ab}I_cI_d] = \frac{1}{T}\left(\sum_{(c,d)}^{[2]} \nu([[f_a \cdot f_b] \cdot f_c] \cdot f_d) + \nu([f_a \cdot f_b] \cdot [f_c \cdot f_d])\right) + O\left(\frac{1}{T^2}\right),$$
$$\mathrm{Cum}[I_a, I_b, I_c, I_d] = \frac{1}{T}\left(\sum_{(ab,c,d)}^{[6]} (\nu([[f_a \cdot f_b] \cdot f_c] \cdot f_d) + \nu([[f_a \cdot f_b] \cdot f_d] \cdot f_c)) + \sum_{(ab,cd)}^{[3]} \nu([f_a \cdot f_b] \cdot [f_c \cdot f_d])\right) + O\left(\frac{1}{T^2}\right).$$

Proof Since $E|X_t|^k < \infty$ for any $t \in \mathbb{R}_+$ and $k \geq 1$ and f_a has at most polynomial growth order, the stochastic integral I_a is well defined, and I_a and J_{ab} have moments up to any order. Therefore, applying (24) of Lemma 2 with $C_{ab} = \nu(f_a \cdot f_b)$, one has

$$E[I_aI_b] = E[J_{ab}] + \nu(f_a \cdot f_b) = \nu(f_a \cdot f_b).$$

From the existence of $G_{\overline{f_a \cdot f_b}}$ and $[f_a \cdot f_b]$, Itô's formula says that

$$E[J_{ab}I_c] = \frac{1}{\sqrt{T}}E\left[\frac{1}{\sqrt{T}}\int_0^T [f_a \cdot f_b] \cdot \mathrm{d}w_t I_c\right] + \frac{1}{T\sqrt{T}}E\left[(G_{f_a \cdot f_b - \nu(f_a \cdot f_b)}(X_T) - G_{f_a \cdot f_b - \nu(f_a \cdot f_b)}(X_0))\int_0^T f_c(X_t) \cdot \mathrm{d}w_t\right].$$

Because the polynomial growth orders of f_a and $G_{\overline{f_a \cdot f_b}}$ ensure the existence of the moments of $G_{\overline{f_a \cdot f_b}}(X_T)$ and I_a, we can apply (16) and (17) in the proof of Proposition 1 to them and obtain that

$$E\left[(G_{f_a\cdot f_b-\nu(f_a\cdot f_b)}(X_T)-G_{f_a\cdot f_b-\nu(f_a\cdot f_b)}(X_0))\int_0^T f_c(X_t)\cdot \mathrm{d}w_t\right]=O(1) \quad \text{as } T\to\infty.$$

Note that $E[G_{f_a\cdot f_b-\nu(f_a\cdot f_b)}(X_0)\int_0^T f_c(X_t)\cdot \mathrm{d}w_t]=0$. Besides, since the polynomial growth order of $[f_a\cdot f_b]$ implies the existence of the moments up to any order of $\int_0^T[f_a\cdot f_b](X_t)\cdot \mathrm{d}w_t$, we obtain from the first result for $E[I_aI_b]$ that

$$E[J_{ab}I_c]=\frac{1}{\sqrt{T}}\nu([f_a\cdot f_b]\cdot f_c)+O\left(\frac{1}{T\sqrt{T}}\right). \tag{27}$$

In the same way, the expansion for $E[J_{ab}J_{cd}]$ can be obtained. Combining (25) of Lemma 2 and (27), we also obtain the expansion for $\mathrm{Cum}[I_a,I_b,I_c]$. Moreover, it follows from (16) and (17) that

$$E[J_{ab}I_cI_d]=\frac{1}{\sqrt{T}}E\left[\frac{1}{\sqrt{T}}\int_0^T[f_a\cdot f_b]\cdot \mathrm{d}w_tI_cI_d\right]+O\left(\frac{1}{T^2}\right).$$

If there exist $G_{\overline{[f_a\cdot f_b]\cdot f_c}}$ and $[[f_a\cdot f_b]\cdot f_c]$ having at most polynomial growth order, we can apply the result for $E[I_aI_bI_c]$ to the first term in the right-hand side above and obtain the result for $E[J_{ab}I_cI_d]$. Finally, Combining this with the expansion for $E[J_{ab}J_{cd}]$ and (26) of Lemma 2 yields the last result for $\mathrm{Cum}[I_a,I_b,I_c,I_d]$. □

3.4 Proof of Theorem 2

Since we assume that the diffusion process has the geometric-mixing property, Condition [A1] in Sakamoto and Yoshida (2004) holds true for the diffusion process X and the Wiener process w in place of an ε-Markov process Y and a driving process X there. Under the conditions [DM1], we have that for any $\Delta>0$, $p>1$, $q>p$ there exists $M>0$ such that

$$\begin{aligned}\sup_{\substack{t\in\mathbb{R}_+\\0\le h\le\Delta}}\left\|\left(Z^{(0)}_{t+h}-Z^{(0)}_t\right)_{a;}\right\|_p\le\sup_{\substack{t\in\mathbb{R}_+\\0\le h\le\Delta}}&\Big(2\|A_{a;}(X_0,\theta_0)\|_p+c_ph^{1/2}\|B_{a;}(X_0,\theta_0)\|_{2p}\\&+Mh^{1/2}E|C_{a;}(X_0,\theta_0)|^q\Big)<\infty.\end{aligned}$$

In the same way, we see that $\sup_{\substack{t\in\mathbb{R}_+\\0\le h\le\Delta}}\|Z^{(1)}_{t+h}-Z^{(1)}_t\|_p<\infty$. Thus the conditions [DM1] imply the [A2] in Sakamoto and Yoshida (2004) for $Z_T=(Z^{(0)}_T,Z^{(1)}_T)$. As in Theorem 4 of Kusuoka and Yoshida (2000), we see that Condition [L] ensures [A3] in Sakamoto and Yoshida (2004) for Z^*_T. In standard literature such as Ikeda and Watanabe (1989), it is assumed that the coefficients of the stochastic differential equation are in C_b^∞ to prove the non-degeneracy of the Malliavin covariance of the solution. Here we have assumed L^p-boundedness of X_t (or L^p-finiteness of X_0 by stationarity) and the only at most polynomial growth condition for coefficients of Z. However it is sufficient for our purpose. Let $\beta > 0$ and k be a positive integer such that $\beta + 1/k < 1/2$, and let

$$M(f) = \left\{ \int_0^{t_0} \int_0^{t_0} \left(\frac{|f(t) - f(s)|}{|t-s|^{\beta+1/k}} \right)^{2k} \mathrm{d}t\mathrm{d}s \right\}^{\frac{1}{2k}}$$

for $f \in C_B([0, t_0]; \mathbb{R}^d)$. Then $M(X|_{[0,t_0]}) \in \mathbb{D}_\infty$. Let $\psi \in C^\infty(\mathbb{R}; [0, 1])$ such that $\psi(y) = 1$ if $|y| \le 1/2$ and $\psi(y) = 0$ if $|y| \ge 1$, and define ψ_1 by

$$\psi_1 = \psi\left(c^{-1}|X_0 - x|^2\right) \psi\left(cM(X|_{[0,t_0]})\right).$$

Let $\mathscr{X} = (X, Z)$. Then there exists a constant C such that $\sup_{t\in[0,t_0]} |X_t| \le C$ whenever $\psi_1 > 0$. There is a stochastic differential equation whose coefficients are bounded with smooth bounded derivatives and its unique strong solution $\hat{\mathscr{X}}$ constructed on the same probability space as $\mathscr{X}$ satisfies $\mathscr{X}|_{[0,t_0]} = \hat{\mathscr{X}}|_{[0,t_0]}$ whenever $\psi_1 > 0$, therefore, we may assume that all coefficients and their derivatives are bounded when we apply the Malliavin calculus. We choose $c > 0$ sufficiently small so that the uniform nondegeneracy of the Malliavin covariance of $\hat{\mathscr{X}}$ under truncation by ψ_1 holds. See Remark 3, p. 575 of Yoshida (2004). After all, we only have to consider the representation of coefficients of the asymptotic expansion in Theorem 6.4 of Sakamoto and Yoshida (2004).

For convenience of explanation, here we recall the definitions of the coefficients used there. Let $(g^{ab}) := (\mathrm{Cov}[Z^{a;}, Z^{b;}])$, $(g_{ab}) = (g^{ab})^{-1}$, $V^{a;}_{a_1\cdots a_k,b} = \mathrm{Cov}[Z^{a;}_{a_1\cdots a_k}, Z^{b';}]g_{b'b}$, $\tilde{\mu}^{a;}_{bc} = (V^{a;}_{b,c} + V^{a;}_{c,b} + \bar{\nu}^{a;}_{bc})/2$, $\tilde{\eta}^{a;}_{b,c} = V^{a;}_{b,c} + \bar{\nu}^{a;}_{bc}$, and

$$U^{a;}_{bcd} = \frac{1}{6}\left(\bar{\nu}^{a;}_{bcd} + \sum_{(bc,d)}^{[3]} V^{a;}_{bc,d} \right) + \frac{1}{3} \sum_{(bc,d)}^{[3]} \tilde{\mu}^{d';}_{bc} \tilde{\eta}^{a;}_{d',d}.$$

Put $\tilde{M}^{a;\ c;}_{b,\ d} = E[Z^{a;}_{b} Z^{c;}_{d}] - V^{a;}_{b,b'} V^{c;}_{d,d'} g^{b'd'}$, $\tilde{N}^{a;b;c;}_{,\ ,\ d} = T^{1/2} E[Z^{a;} Z^{b;} (Z^{c;}_{d} - V^{c;}_{d,d'} Z^{d';})]$, $\bar{\lambda}^{abc} = T^{1/2}\mathrm{Cum}[Z^{a;}, Z^{b;}, Z^{c;}]$, $H^{abcd} = T\mathrm{Cum}[Z^{a;}, Z^{b;}, Z^{c;}, Z^{d;}]$.

From the representation (4) of the estimating function $\psi_{a;}$, we see that the constants defined just before Theorem 1 satisfy

$$\bar{\nu}_{a;} = \frac{1}{T}\nu(A_{a;}) + \nu(C_{a;}), \quad \bar{\nu}_{a;A} = \frac{1}{T}\nu(A_{a;A}) + \nu(C_{a;A}),$$

$$\Delta^{a;} = \tilde{\Delta}^{a;} + O(T^{-1}), \quad \bar{\nu}^{a;b} = \tilde{\nu}^{a;b} - \frac{1}{T}\bar{A}^{a;b} + O(T^{-2})$$

for any $a = 1, \ldots, p$ and any index set A with $|A| \leq 4$. It follows from Itô's formula that under Condition [DM2]

$$\psi_{a;A}(\theta_0) - E[\psi_{a;A}(\theta_0)] = A_{a;A}(X_0, \theta_0) - \nu(A_{a;A}(\cdot, \theta_0)) + G\langle C_{a;A}\rangle(X_T, \theta_0)$$
$$-G\langle C_{a;A}\rangle(X_0, \theta_0) + \int_0^T B^*_{a;A}(X_t, \theta_0)\mathrm{d}w_t \tag{28}$$

for $A = \phi, \{a\}, \{a, b\}, a, b \in \{1, \ldots, p\}$. Combining this with Proposition 1 and Lemma 3, we see that under [DM2] and [DM3](ii),

$$\begin{aligned}
&\mathrm{Cov}\left[T^{-1/2}\psi_{a;A}, T^{-1/2}\psi_{b;B}\right] \\
&= \bar{F}_{a;A,b;B} + \frac{1}{T}\Bigg(\mathrm{Cov}[A_{a;A}(X_0), A_{b;B}(X_0)] - \nu(A_{a;A}\, G\langle C_{b;B}\rangle) \\
&\quad -\nu(A_{b;B}\, G\langle C_{a;A}\rangle) \\
&\quad +2\nu(G\langle C_{a;A}\rangle\, G\langle C_{b;B}\rangle) + E\left[\int_0^T B^*_{a;A}\mathrm{d}w_t G\langle C_{b;B}\rangle(X_T)\right] \\
&\quad +E\left[\int_0^T B^*_{b;B}\mathrm{d}w_t G\langle C_{a;A}\rangle(X_T)\right]\Bigg) + o(T^{-1})
\end{aligned} \tag{29}$$

for any $a, b \in \{1, \ldots, p\}$, index sets A, B with $0 \leq |A| \leq 2$, $0 \leq |B| \leq 2$. From this formula, we see that the coefficients defined by the covariances fulfill

$$\begin{aligned}
g^{ab} &= \rho^{ab} + \frac{1}{T}\tilde{\tau}^{ab} + o(T^{-1}), \quad V^{a;}{}_{B,c} = V^{*a;}{}_{B,c} + O(T^{-1}), \\
\tilde{\mu}^{a;}{}_{bc} &= \mu^{*a;}{}_{bc} + O(T^{-1}), \quad \tilde{\eta}^{a;}{}_{b,c} = \eta^{*a;}{}_{b,c} + O(T^{-1}), \\
U^{a;}{}_{bcd} &= U^{*a;}{}_{bcd} + O(T^{-1}), \quad \tilde{M}^{a;}{}_{b,}{}^{c;}{}_{d} = \tilde{M}^{*a;}{}_{b,}{}^{c;}{}_{d} + O(T^{-1}).
\end{aligned}$$

Note that the constants in the RSH's were defined in above Theorem 2.

Moreover, under [DM2] and [DM3](ii), the representation (28) with Proposition 1 and Lemma 3 yields expansions of third and fourth order cumulants of $T^{-1}\psi_{a;A}$, which imply

$$\begin{aligned}
\bar{\lambda}^{abc} &= \bar{\lambda}^{*abc} + O(T^{-1}), \quad \tilde{N}^{a;,b;,c;}{}_{d} = \tilde{N}^{*a;,b;,c;}{}_{d} + O(T^{-1}), \\
H^{abcd} &= H^{*abcd} + O(T^{-1}).
\end{aligned}$$

Since $g_{ab} = \rho_{ab} - \tilde{\tau}_{ab}/T + o(T^{-1})$, we have

$$\phi(z; g^{ab}) = \phi(z; \rho^{ab})\left(1 + \frac{1}{2T}\tilde{\tau}^{ab}h_{ab}(z; \rho^{ab})\right) + o(T^{-1}).$$

Combining these results with Theorem 6.4 in Sakamoto and Yoshida (2004), we can obtain the representation of the coefficient in Theorem 2. □

3.5 Proof of Theorem 3

The inequality (10) with $q_{T,2}$ of Theorem 2 is trivial; however, the translation from $q_{T,2}$ of Theorem 2 into that of Theorem 3 asks routine but thorough and lengthy calculations. Here we make rough sketches of such calculations for one's convenience.

Put $B(x,\theta)=\tilde{V}_0'(\tilde{V}\tilde{V}')^{-1}\tilde{V}(x,\theta), C(x,\theta)=B(x,\theta)\cdot B(x,\theta_0)-\frac{1}{2}B(x,\theta)\cdot B(x,\theta)$, then we see that $B_{a;}(x,\theta)=\delta_a B(x,\theta)$ and $C_{a;}(x,\theta)=\delta_a C(x,\theta)$. These relations lead that $C_a(x,\theta_0)=0$, $G\langle C_a\rangle(x,\theta_0)=0$, $[C_a](x,\theta_0)=0$, $\bar{\nu}(C_{ab})=-\check{F}_{a,b}$,

$$\bar{\nu}(C_{abc})=-\sum_{(ab,c)}^{[3]}\check{F}_{ab,c},\quad\text{and}\quad \bar{\nu}(C_{abcd})=-\sum_{(abc,d)}^{[4]}\check{F}_{abc,d}-\sum_{(ab,cd)}^{[3]}\check{F}_{ab,cd}.$$

Moreover, $\bar{F}$'s in Theorem 2 can also be represented by $\check{F}$'s. In particular,

$$\bar{F}_{a,b}=\check{F}_{a,b},\quad \bar{F}_{ab,c}=\check{F}_{ab,c}-\check{F}_{[a,b],c},\quad \bar{F}_{abc,d}=\check{F}_{abc,d}-\sum_{(ab,c)}^{[3]}\check{F}_{[ab,c],d},$$
$$\bar{F}_{[a,b],c}=\check{F}_{[a,b],c},\quad \bar{F}_{[a,b],[c,d]}=\check{F}_{[a,b],[c,d]},\quad \bar{F}_{[[a,b],c],d}=\check{F}_{[[a,b],c],d}.$$

Therefore, the constants in the expansion of Theorem 2 are expressed as follows;

$$\begin{aligned}
(\tilde{\nu}^{ab}) &= (\nu(C_{ab}))^{-1}=-(\check{F}_{a,b})^{-1},\quad (\rho^{ab})=(\check{F}_{a,b})^{-1},\\
(\rho_{ab}) &= (\check{F}_{a,b}),\quad \tilde{\nu}^{ab}=-\rho^{ab},\\
\bar{A}^{ab} &= \rho^{aa'}\rho^{bb'}\nu(A_{a'b'})=\rho^{aa'}\rho^{bb'}\zeta_{a'b'},\\
\tau_{ab} &= \mathrm{Cov}[A_a(X_0),A_b(X_0)],\\
\tilde{\tau}^{ab} &= \rho^{aa'}\rho^{bb'}\tau_{a'b'}+2\bar{A}^{ab}=\rho^{aa'}\rho^{bb'}(\tau_{a'b'}+2\zeta_{a'b'}),\\
\mu^{*a}{}_{bc} &= -\frac{1}{2}\rho^{aa'}\left(-\bar{F}_{a'b,c}-\bar{F}_{a'c,b}+\sum_{(a'b,c)}^{[3]}\check{F}_{a'b,c}\right)\\
&= -\frac{1}{2}\rho^{aa'}\left(\check{F}_{bc,a'}+\check{F}_{[a',b],c}+\check{F}_{[a',c],b}\right)=-\frac{1}{2}\rho^{aa'}\tilde{\Gamma}^{(-1)}_{bc,a'},\\
\eta^{*a}{}_{b,c} &= -\rho^{aa'}\left(-\bar{F}_{a'b,c}+\sum_{(a'b,c)}^{[3]}\check{F}_{a'b,c}\right)\\
&= -\rho^{aa'}\left(\check{F}_{a'c,b}+\check{F}_{bc,a'}+\check{F}_{[a',b],c}\right)=-\rho^{aa'}\left(\tilde{\Gamma}^{(-1)}_{a'c,b}+\tilde{\Gamma}^{(1)}_{bc,a'}\right),\\
U^{*a}{}_{bcd} &= -\frac{1}{6}\rho^{aa'}\left(\sum_{(a'bc,d)}^{[4]}\check{F}_{a'bc,d}+\sum_{(a'b,cd)}^{[3]}\check{F}_{a'b,cd}-\sum_{(bc,d)}^{[3]}\bar{F}_{a'bc,d}\right)\\
&\quad +\frac{1}{6}\sum_{(bc,d)}^{[3]}\rho^{d'd''}\rho^{aa'}\tilde{\Gamma}^{(-1)}_{bc,d''}\left(\tilde{\Gamma}^{(-1)}_{a'd,d'}+\tilde{\Gamma}^{(1)}_{d'd,a'}\right)
\end{aligned}$$

$$
\begin{aligned}
&= -\frac{1}{6}\rho^{aa'}\left(\check{F}_{bcd,a'} + \sum_{(a'b,cd)}^{[3]}\check{F}_{a'b,cd} + \sum_{(bc,d)}^{[3]}\sum_{(a'b,c)}^{[3]}\check{F}_{[a'b,c],d}\right)\\
&\quad + \frac{1}{6}\sum_{(bc,d)}^{[3]}\rho^{d'd''}\rho^{aa'}\tilde{\Gamma}^{(-1)}_{bc,d''}\left(\tilde{\Gamma}^{(-1)}_{a'd,d'} + \tilde{\Gamma}^{(1)}_{d'd,a'}\right),
\end{aligned}
$$

$$
\bar{\lambda}^{*abc} = \rho^{aa'}\rho^{bb'}\rho^{cc'}\sum_{(ab,c)}^{[3]}\check{F}_{[a',b'],c'},
$$

$$
\begin{aligned}
H^{*abcd} &= \rho^{aa'}\rho^{bb'}\rho^{cc'}\rho^{dd'}\left(\sum_{(a'b',c',d')}^{[6]}\left(\check{F}_{[[a',b'],c'],d'} + \check{F}_{[[a',b'],d'],c'}\right)\right.\\
&\quad \left. + \sum_{(a'b',c'd')}^{[3]}\check{F}_{[a',b'],[c',d']}\right),
\end{aligned}
$$

$$
V^{*a}_{\ \ b,c} = \rho^{aa'}\left(\check{F}_{a'b,c} - \check{F}_{[a',b],c}\right) = \rho^{aa'}\tilde{\Gamma}^{(1)}_{a'b,c},
$$

$$
\begin{aligned}
\tilde{M}^{*a;}{}_{b,}{}^{c;}{}_{d} &= \rho^{aa'}\rho^{cc'}\left(\check{F}_{a'b,c'd} - \check{F}_{[a',b],c'd} - \check{F}_{a'b,[c',d]} + \check{F}_{[a',b],[c',d]}\right)\\
&\quad - \rho^{aa'}\rho^{ff'}\rho^{cc'}\tilde{\Gamma}^{(1)}_{c'd,f}\tilde{\Gamma}^{(1)}_{a'b,f'}\\
&\quad - \rho^{ee'}\rho^{cc'}\rho^{aa'}\tilde{\Gamma}^{(1)}_{a'b,e}\tilde{\Gamma}^{(1)}_{c'd,e'} + \rho^{ef}\rho^{aa'}\tilde{\Gamma}^{(1)}_{a'b,e}\rho^{cc'}\tilde{\Gamma}^{(1)}_{c'd,f}\\
&= \rho^{aa'}\rho^{cc'}\left(\check{F}_{a'b,c'd} - \check{F}_{[a',b],c'd} - \check{F}_{a'b,[c',d]}\right.\\
&\quad \left. + \check{F}_{[a',b],[c',d]} - \rho^{ef}\tilde{\Gamma}^{(1)}_{a'b,e}\tilde{\Gamma}^{(1)}_{c'd,f}\right),
\end{aligned}
$$

and

$$
\begin{aligned}
\tilde{N}^{*a;}{}_{,}{}^{b;}{}_{,}{}^{c;}{}_{d} &= \rho^{aa'}\rho^{bb'}\rho^{cc'}\left(\left(\check{F}_{[a',b'],c'd} - \check{F}_{[a',b'],[c',d]}\right) + \left(\check{F}_{[a',c'd],b'} - \check{F}_{[a',[c',d]],b'}\right)\right.\\
&\quad \left. + \left(\check{F}_{[b',c'd],a'} - \check{F}_{[b',[c',d]],a'}\right) - \rho^{ee'}\tilde{\Gamma}^{(1)}_{c'd,e}\sum_{(a'b',e')}^{[3]}\check{F}_{[a',b'],e'}\right).
\end{aligned}
$$

By using these expressions, we can translate coefficients. First, $c^{abc}h_{abc}$ becomes

$$
\begin{aligned}
c^{abc}h_{abc} &= h_{abc}\rho^{aa'}\rho^{bb'}\rho^{cc'}\left(\sum_{(a'b',c')}^{[3]}\check{F}_{[a',b'],c'} - 3\check{F}_{a'b',c'} - 3\check{F}_{[c',a'],b'} - 3\check{F}_{[c',b'],a'}\right)\\
&= -3h^{abc}\left(\check{F}_{ab,c} + \check{F}_{[a,b],c}\right) = -3h^{abc}\tilde{\Gamma}^{(-1/3)}_{ab,c}.
\end{aligned}
$$

Next, putting $\check{\mu}^a = -\frac{1}{2}\rho^{aa'}\tilde{\Gamma}^{(-1)}_{bc,a'}\rho^{bc}$ $(= \tilde{\mu}^{*a}_{\ bc}\rho^{bc})$, we have

$$
\begin{aligned}
c^{abcd}h_{abcd} = h^{abcd} & \left(\sum_{(ab,c,d)}^{[6]} \left(\check{F}_{[[a,b],c],d} + \check{F}_{[[a,b],d],c} \right) + \sum_{(ab,cd)}^{[3]} \check{F}_{[a,b],[c,d]} \right) \\
& + 12h^{abcd}\tilde{\Gamma}^{(-1/3)}_{ab,c}\rho_{dd'}(\tilde{\beta}^{d'} - \check{\mu}^{d'}) - 12h^{abcd}\rho^{ee'} \\
& \times \left(\sum_{(ab,e')}^{[3]} \check{F}_{[a,b],e'} - \tilde{\Gamma}^{(-1)}_{be',a} \right) \tilde{\Gamma}^{(-1)}_{de,c} \\
& + 12h^{abcd} \Bigg(\check{F}_{ab,cd} - \check{F}_{ab,[c,d]} - \check{F}_{[a,b],cd} + \check{F}_{[a,b],[c,d]} \\
& - \rho^{ef}\tilde{\Gamma}^{(1)}_{ab,e}\tilde{\Gamma}^{(1)}_{cd,f} + \check{F}_{[a,b],cd} - \check{F}_{[a,b],[c,d]} + \check{F}_{[a,cd],b} \\
& - \check{F}_{[a,[c,d]],b} + \check{F}_{[b,cd],a} - \check{F}_{[b,[c,d]],a} - \rho^{ee'}\tilde{\Gamma}^{(1)}_{cd,e} \sum_{(ab,e')}^{[3]} \check{F}_{[a,b],e'} \Bigg) \\
& - 4h^{abcd} \left(\check{F}_{bcd,a} + \sum_{(ab,cd)}^{[3]} \check{F}_{ab,cd} + \sum_{(bc,d)}^{[3]} \sum_{(ab,c)}^{[3]} \check{F}_{[ab,c],d} \right) \\
& + 4h^{abcd} \sum_{(bc,d)}^{[3]} \rho^{ef}\tilde{\Gamma}^{(-1)}_{bc,e} \left(\tilde{\Gamma}^{(-1)}_{ad,f} + \tilde{\Gamma}^{(1)}_{df,a} \right) \\
= h^{abcd} & \Bigg(- \sum_{(ab,c,d)}^{[6]} \Big(\check{F}_{[[a,b],c],d} + \check{F}_{[[a,b],d],c} + \check{F}_{ab,[c,d]} + \check{F}_{[a,b],cd} \\
& + \check{F}_{[ab,c],d} + \check{F}_{[ab,d],c} \Big) \\
& + \sum_{(ab,cd)}^{[3]} \check{F}_{[a,b],[c,d]} - \sum_{(abc,d)}^{[4]} \check{F}_{abc,d} \Bigg) + 12h^{abcd}\tilde{\Gamma}^{(-1/3)}_{ab,c}\rho_{dd'}(\tilde{\beta}^{d'} - \check{\mu}^{d'}) \\
& + 12h^{abcd}\rho^{ef} \left(\tilde{\Gamma}^{(1)}_{bf,a}\tilde{\Gamma}^{(-1)}_{de,c} - \tilde{\Gamma}^{(1)}_{cd,f}\tilde{\Gamma}^{(-1)}_{ab,e} + \tilde{\Gamma}^{(-1)}_{bc,e} \left(\tilde{\Gamma}^{(1)}_{ad,f} + \tilde{\Gamma}^{(-1)}_{df,a} \right) \right) \\
= h^{abcd} & \Bigg(- \sum_{(ab,c,d)}^{[6]} \Big(\check{F}_{[[a,b],c],d} + \check{F}_{[[a,b],d],c} + \check{F}_{ab,[c,d]} + \check{F}_{[a,b],cd} \\
& + \check{F}_{[ab,c],d} + \check{F}_{[ab,d],c} \Big) \\
& + \sum_{(ab,cd)}^{[3]} \check{F}_{[a,b],[c,d]} - \sum_{(abc,d)}^{[4]} \check{F}_{abc,d} \Bigg) + 12h^{abcd}\tilde{\Gamma}^{(-1/3)}_{ab,c}\rho_{dd'}(\tilde{\beta}^{d'} - \check{\mu}^{d'}) \\
& + 12h^{abcd}\rho^{ef} \left(\tilde{\Gamma}^{(-1)}_{ab,e} + \tilde{\Gamma}^{(1)}_{ae,b} \right) \tilde{\Gamma}^{(-1)}_{cf,d}
\end{aligned}
$$

$$= h^{abcd}\left(-12\left(\check{F}_{[[a,b],c],d}+\check{F}_{ab,[c,d]}+\check{F}_{[ab,c],d}\right)+3\check{F}_{[a,b],[c,d]}-4\check{F}_{abc,d}\right.$$
$$\left.+12\tilde{\Gamma}^{(-1/3)}_{ab,c}\rho_{dd'}(\tilde{\beta}^{d'}-\check{\mu}^{d'})+12\rho^{ef}\left(\tilde{\Gamma}^{(-1)}_{ab,e}+\tilde{\Gamma}^{(1)}_{ae,b}\right)\tilde{\Gamma}^{(-1)}_{cf,d}\right)$$

Finally, with

$$(I)=h^{ab}(\tau_{ab}+2\zeta_{ab})+h^{ab}\rho^{cd}\left(-4\check{F}_{[a,d],bc}+\check{F}_{[a,d],[b,c]}-2\check{F}_{[ab,c],d}-2\check{F}_{[bc,a],d}\right.$$
$$\left.-\check{F}_{[cd,a],b}-2\check{F}_{[[b,c],a],d}-2\check{F}_{[[b,c],d],a}+\check{F}_{bc,ad}-\check{F}_{ab,cd}-\check{F}_{cdb,a}\right)$$
$$+2h^{ab}\rho_{aa'}\left(\tilde{\Delta}^c\tilde{\eta}^{*a'}{}_{c,b}-\delta_b\beta^{a'}\right)+h^{ab}\rho_{aa'}\rho_{bb'}(\tilde{\beta}^{a'}-\check{\mu}^{a'})(\tilde{\beta}^{b'}-\check{\mu}^{b'}),$$

we obtain

$$A^{*ab}h_{ab}=h_{ab}\rho^{aa'}\rho^{bb'}(\tau_{a'b'}+2\zeta_{a'b'})$$
$$-h_{ab}\left(\rho^{aa'}\rho^{cc'}\rho^{dd'}\sum_{(a'c',d')}^{[3]}\check{F}_{[a',c'],d'}-\frac{1}{2}\tilde{\Gamma}^{(-1)}_{c'd',a'}\rho^{aa'}\rho^{c'c}\rho^{d'd}\right)\tilde{\Gamma}^{(-1)}_{cd,b'}\rho^{bb'}$$
$$+2h_{ab}\delta^c_{c'}\rho^{aa'}\rho^{c'c''}\rho^{bb'}\left(\check{F}_{[a',c''],b'c}-\check{F}_{[a',c''],[b',c]}+\check{F}_{[b'c,a'],c''}\right.$$
$$\left.-\check{F}_{[[b',c],a'],c''}+\check{F}_{[b'c,c''],a'}-\check{F}_{[[b',c],c''],a'}-\rho^{ef}\tilde{\Gamma}^{(1)}_{b'c,e}\sum_{(a'c'',f)}^{[3]}\check{F}_{[a',c''],f}\right)$$
$$+h_{ab}\rho^{cc'}\rho^{bb'}\rho^{aa'}\left(\check{F}_{b'c,a'c'}-\check{F}_{[b',c],a'c'}-\check{F}_{b'c,[a',c']}+\check{F}_{[b',c],[a',c']}\right.$$
$$\left.-\rho^{ef}\tilde{\Gamma}^{(1)}_{b'c,e}\tilde{\Gamma}^{(1)}_{a'c',f}\right)+2h_{ab}\left(\tilde{\Delta}^c\tilde{\eta}^{*a}{}_{c,b'}-\delta_{b'}\beta^a\right)\rho^{b'b}$$
$$+2h_{ab}\rho^{b'b}\delta^{a_1}_{b_1}\rho^{aa'}\rho^{b_1b'_1}\left(\check{F}_{a'a_1,b'_1b'}-\check{F}_{[a',a_1],b'_1b'}-\check{F}_{a'a_1,[b'_1,b']}\right.$$
$$\left.+\check{F}_{[a',a_1],[b'_1,b']}-\rho^{ef}\tilde{\Gamma}^{(1)}_{a'a_1,e}\tilde{\Gamma}^{(1)}_{b'_1b',f}\right)$$
$$+6h_{ab}\rho^{cd}\rho^{b'b}\left(-\frac{1}{6}\rho^{aa'}\left(\check{F}_{cdb',a'}+\sum_{(a'c,db')}^{[3]}\check{F}_{a'c,db'}\right.\right.$$
$$\left.\left.+\sum_{(cd,b')}^{[3]}\sum_{(a'c,d)}^{[3]}\check{F}_{[a'c,d],b'}\right)+\frac{1}{6}\sum_{(cd,b')}^{[3]}\rho^{ef}\rho^{aa'}\tilde{\Gamma}^{(-1)}_{cd,e}\left(\tilde{\Gamma}^{(-1)}_{a'b',f}+\tilde{\Gamma}^{(1)}_{fb',a'}\right)\right)$$
$$+h_{ab}(\tilde{\beta}^a-\check{\mu}^a)(\tilde{\beta}^b-\check{\mu}^b)$$
$$=h^{ab}(\tau_{ab}+2\zeta_{ab})+h^{ab}\rho^{cd}\left(2\check{F}_{[a,d],bc}-2\check{F}_{[a,d],[b,c]}\right.$$
$$+2\check{F}_{[bc,a],d}-2\check{F}_{[[b,c],a],d}+2\check{F}_{[bc,d],a}-2\check{F}_{[[b,c],d],a}$$
$$+\check{F}_{bc,ad}-\check{F}_{[b,c],ad}-\check{F}_{bc,[a,d]}+\check{F}_{[b,c],[a,d]}$$

$$
\begin{aligned}
&+2\check{F}_{ac,db}-2\check{F}_{[a,c],db}-2\check{F}_{ac,[d,b]}+2\check{F}_{[a,c],[d,b]}\\
&-\left(\check{F}_{cdb,a}+\sum_{(ac,db)}^{[3]}\check{F}_{ac,db}+\sum_{(cd,b)}^{[3]}\sum_{(ac,d)}^{[3]}\check{F}_{[ac,d],b}\right)\Bigg)\\
&+h^{ab}\rho^{cd}\rho^{ef}\left(-\left(\sum_{(ad,f)}^{[3]}\check{F}_{[a,d],f}-\frac{1}{2}\tilde{\Gamma}^{(-1)}_{df,a}\right)\tilde{\Gamma}^{(-1)}_{ce,b}-2\tilde{\Gamma}^{(1)}_{bc,e}\sum_{(ad,f)}^{[3]}\check{F}_{[a,d],f}\right.\\
&\left.-\tilde{\Gamma}^{(1)}_{bc,e}\tilde{\Gamma}^{(1)}_{ad,f}-2\tilde{\Gamma}^{(1)}_{ac,e}\tilde{\Gamma}^{(1)}_{db,f}+\sum_{(cd,b)}^{[3]}\tilde{\Gamma}^{(-1)}_{cd,e}\left(\tilde{\Gamma}^{(-1)}_{ab,f}+\tilde{\Gamma}^{(1)}_{fb,a}\right)\right)\\
&+2h_{ab}\left(\tilde{\Delta}^{c}\tilde{\eta}^{*a}_{\ \ c,b'}-\delta_{b'}\beta^{a}\right)\rho^{b'b}+h_{ab}(\tilde{\beta}^{a}-\check{\mu}^{a})(\tilde{\beta}^{b}-\check{\mu}^{b})\\
=&\,(I)+h^{ab}\rho^{cd}\rho^{ef}\left(\frac{1}{2}\tilde{\Gamma}^{(-1)}_{df,a}\tilde{\Gamma}^{(-1)}_{ce,b}-\tilde{\Gamma}^{(1)}_{bc,e}\tilde{\Gamma}^{(1)}_{ad,f}+\tilde{\Gamma}^{(-1)}_{cd,e}\left(\tilde{\Gamma}^{(-1)}_{ab,f}+\tilde{\Gamma}^{(1)}_{fb,a}\right)\right.\\
&-\tilde{\Gamma}^{(-1)}_{ce,b}\sum_{(ad,f)}^{[3]}\check{F}_{[a,d],f}-2\tilde{\Gamma}^{(1)}_{bc,e}\tilde{\Gamma}^{(-1)}_{ad,f}\\
&\left.+\tilde{\Gamma}^{(-1)}_{bd,e}\left(\tilde{\Gamma}^{(-1)}_{ac,f}+\tilde{\Gamma}^{(1)}_{fc,a}\right)+\tilde{\Gamma}^{(-1)}_{bc,e}\left(\tilde{\Gamma}^{(-1)}_{ad,f}+\tilde{\Gamma}^{(1)}_{fd,a}\right)\right)\\
=&\,(I)+h^{ab}\rho^{cd}\rho^{ef}\left(\frac{1}{2}\tilde{\Gamma}^{(-1)}_{df,a}\tilde{\Gamma}^{(-1)}_{ce,b}-\tilde{\Gamma}^{(1)}_{bc,e}\tilde{\Gamma}^{(1)}_{ad,f}+\tilde{\Gamma}^{(-1)}_{cd,e}\left(\tilde{\Gamma}^{(-1)}_{ab,f}+\tilde{\Gamma}^{(1)}_{fb,a}\right)\right.\\
&-\tilde{\Gamma}^{(-1)}_{ce,b}\sum_{(ad,f)}^{[3]}\check{F}_{[a,d],f}-2\tilde{\Gamma}^{(1)}_{bc,e}\tilde{\Gamma}^{(-1)}_{ad,f}\\
&\left.+\tilde{\Gamma}^{(-1)}_{bd,e}\left(\tilde{\Gamma}^{(1)}_{ac,f}+\tilde{\Gamma}^{(-1)}_{fc,a}\right)+\tilde{\Gamma}^{(-1)}_{bc,e}\left(\tilde{\Gamma}^{(1)}_{ad,f}+\tilde{\Gamma}^{(-1)}_{fd,a}\right)\right)\\
=&\,(I)+h^{ab}\rho^{cd}\rho^{ef}\left(\frac{1}{2}\tilde{\Gamma}^{(-1)}_{df,a}\tilde{\Gamma}^{(-1)}_{ce,b}-\tilde{\Gamma}^{(1)}_{bc,e}\tilde{\Gamma}^{(1)}_{ad,f}+\tilde{\Gamma}^{(-1)}_{cd,e}\left(\tilde{\Gamma}^{(-1)}_{ab,f}+\tilde{\Gamma}^{(1)}_{fb,a}\right)\right.\\
&\left.-\tilde{\Gamma}^{(-1)}_{ce,b}\sum_{(ad,f)}^{[3]}\check{F}_{[a,d],f}+\tilde{\Gamma}^{(-1)}_{bd,e}\tilde{\Gamma}^{(-1)}_{fc,a}+\tilde{\Gamma}^{(-1)}_{bc,e}\tilde{\Gamma}^{(-1)}_{fd,a}\right)\\
=&\,(I)+h^{ab}\rho^{cd}\rho^{ef}\left(\frac{1}{2}\tilde{\Gamma}^{(-1)}_{df,a}\tilde{\Gamma}^{(-1)}_{ce,b}-\tilde{\Gamma}^{(1)}_{bc,e}\tilde{\Gamma}^{(1)}_{ad,f}+\tilde{\Gamma}^{(-1)}_{cd,e}\left(\tilde{\Gamma}^{(-1)}_{ab,f}+\tilde{\Gamma}^{(1)}_{fb,a}\right)\right.\\
&\left.+\tilde{\Gamma}^{(-1)}_{bd,e}\tilde{\Gamma}^{(-1)}_{fc,a}+\tilde{\Gamma}^{(1)}_{ad,f}\tilde{\Gamma}^{(-1)}_{ce,b}\right)\\
=&\,(I)+h^{ab}\rho^{cd}\rho^{ef}\left(\frac{1}{2}\tilde{\Gamma}^{(-1)}_{df,a}\tilde{\Gamma}^{(-1)}_{ce,b}-\tilde{\Gamma}^{(1)}_{bc,e}\tilde{\Gamma}^{(1)}_{ad,f}+\tilde{\Gamma}^{(-1)}_{cd,e}\left(\tilde{\Gamma}^{(-1)}_{ab,f}+\tilde{\Gamma}^{(1)}_{fb,a}\right)\right.\\
&\left.+\tilde{\Gamma}^{(-1)}_{ce,b}\left(\tilde{\Gamma}^{(-1)}_{ad,f}+\tilde{\Gamma}^{(1)}_{ad,f}\right)\right).
\end{aligned}
$$

□

Acknowledgments This work was in part supported by JSPS Grants-in-Aid for Scientific Research, Cooperative Research Program of the Institute of Statistical Mathematics, JSPS Japan-Russia Joint Research Projects and Scientific Seminars under the Bilateral Programs, and Twenty-First Century COE Program "Base for New development of Mathematics to Science and Technology", Graduate School of Mathematical Sciences, University of Tokyo.

References

Ikeda, N., Watanabe, S. (1989). Stochastic differential equations and diffusion processes. In *North-Holland mathematical library* (Vol. 24). Amsterdam: North-Holland.

Kunita, H. (1990). Stochastic flows and stochastic differential equations. In *Cambridge studies in advanced mathematics* (Vol. 24). Cambridge: Cambridge University Press.

Kusuoka, S., Yoshida, N. (2000). Malliavin calculus, geometric mixing, and expansion of diffusion functionals. *Probability Theory and Related Fields*, *116*(4), 457–484.

Kutoyants, Y. A. (1984). Parameter estimation for stochastic processes. In *Research and exposition in mathematics* (Vol. 6). Berlin: Heldermann Verlag. (Translated from the Russian and edited by B. L. S. Prakasa Rao).

Kutoyants, Y. (1994). Identification of dynamical systems with small noise. In *Mathematics and its applications* (Vol. 300). Dordrecht: Kluwer.

Kutoyants, Y. A. (2004). *Statistical inference for ergodic diffusion processes*. Springer series in statistics. London: Springer.

Masuda, H. (2004). On multidimensional Ornstein-Uhlenbeck processes driven by a general Lévy process. *Bernoulli*, *10*(1), 97–120.

Meyn, S. P., Tweedie, R. L. (1992). Stability of Markovian processes. I. Criteria for discrete-time chains. *Advances in Applied Probability*, *24*(3), 542–574.

Meyn, S. P., Tweedie, R. L. (1993a). Stability of Markovian processes. II. Continuous-time processes and sampled chains. *Advances in Applied Probability*, *25*(3), 487–517.

Meyn, S. P., Tweedie, R. L. (1993b). Stability of Markovian processes. III. Foster–Lyapunov criteria for continuous-time processes. *Advances in Applied Probability*, *25*(3), 518–548.

Pardoux, E., Veretennikov, A. Y. (2001). On the Poisson equation and diffusion approximation. I. *The Annals of Probability*, *29*(3), 1061–1085.

Pardoux, È., Veretennikov, A. Y. (2003). On Poisson equation and diffusion approximation. II. *The Annals of Probability*, *31*(3), 1166–1192.

Pfanzagl, J. (1994). *Parametric statistical theory*. de Gruyter textbook. Berlin: Walter de Gruyter Co. (With the assistance of R. Hamböker).

Prakasa Rao, B. L. S. (1999). Statistical inference for diffusion type processes. In *Kendall's library of Statistics* (Vol. 8). London: Edward Arnold.

Sakamoto, Y., Yoshida, N. (1998a). Asymptotic expansion of M-estimator over Wiener space. *Statistical Inference for Stochastic Processes*, *1*(1), 85–103.

Sakamoto, Y., Yoshida, N. (1998b). Third order asymptotic expansion for diffusion process. In *Theory of statistical analysis and its applications*. *Cooperative research report* (Vol. 107, pp. 53–60). The Institute of Statistical Mathematics.

Sakamoto, Y., Yoshida, N. (1999). Higher order asymptotic expansion for a functional of a mixing process with applications to diffusion processes. (Unpublished).

Sakamoto, Y., Yoshida, N. (2003). Asymptotic expansion under degeneracy. *Journal of the Japan Statistical Society*, *33*(2), 145–156.

Sakamoto, Y., Yoshida, N. (2004). Asymptotic expansion formulas for functionals of ε-Markov processes with a mixing property. *Annals of the Institute of Statistical Mathematics*, *56*(3), 545–597.

Uchida, M., Yoshida, N. (2001). Information criteria in model selection for mixing processes. *Statistical Inference for Stochastic Processes*, *4*(1), 73–98.

Veretennikov, A. Y. (1987). Bounds for the mixing rate in the theory of stochastic equations. *Theory of Probability and its Applications*, *32*(2), 273–281.

Veretennikov, A. Y. (1997). On polynomial mixing bounds for stochastic differential equations. *Stochastic Processes and their Applications*, *70*(1), 115–127.

Yoshida, N. (1997). Malliavin calculus and asymptotic expansion for martingales. *Probability Theory and Related Fields*, *109*(3), 301–342.

Yoshida, N. (2004). Partial mixing and Edgeworth expansion. *Probability Theory and Related Fields*, *129*(4), 559–624.

Yoshida, N. (2005). Polynomial type large deviation inequalities and quasi-likelihood analysis for stochastic differential equations. (Submitted).