Ann. Inst. Statist. Math.
Vol. 56, No. 3, 545–597 (2004)

ASYMPTOTIC EXPANSION FORMULAS FOR FUNCTIONALS OF ϵ-MARKOV PROCESSES WITH A MIXING PROPERTY

YUJI SAKAMOTO[1] AND NAKAHIRO YOSHIDA[2]

[1] *Faculty of Human Environment, Hiroshima International University, 555-36 Gakuendai, Kurose-cho, Kamo-Gun, Hiroshima 724-0695, Japan*
[2] *Graduate School of Mathematical Sciences, University of Tokyo, 3-8-1 Komaba, Meguro, Tokyo 153-8914, Japan*

(Received December 18, 2002; revised August 14, 2003)

Abstract. The ϵ-Markov process is a general model of stochastic processes which includes nonlinear time series models, diffusion processes with jumps, and many point processes. With a view to applications to the higher-order statistical inference, we will consider a functional of the ϵ-Markov process admitting a stochastic expansion. Arbitrary order asymptotic expansion of the distribution will be presented under a strong mixing condition. Applying these results, the third order asymptotic expansion of the M-estimator for a general stochastic process will be derived. The Malliavin calculus plays an essential role in this article. We illustrate how to make the Malliavin operator in several concrete examples. We will also show that the third-order expansion formula (Sakamoto and Yoshida (1998, ISM Cooperative Research Report, No. 107, 53–60; 1999, unpublished)) of the maximum likelihood estimator for a diffusion process can be obtained as an example of our result.

Key words and phrases: Asymptotic expansion, ϵ-Markov process, geometric mixing, M-estimator.

1. Introduction

The aim of this article is to provide a rigorous mathematical foundation to the theory of higher-order statistical inference for stochastic processes. In order to handle in a unified way stochastic processes appearing in applied statistics, and in the same time, to develop a theory on a rigid probabilistic basis, we will adopt an ϵ-Markov process as the underlying process based on which other statistics are constructed. Because of the choice of the continuous time, our results can apply to a Markovian semimartingale such as a solution of a stochastic differential equation with jumps. Moreover, logically, they also apply to a discrete time model by embedding it into a continuous time model in a natural way. However, there is a wider difference between discrete time and continuous time than the difference of mere formats, as we will later mention it.

There is an extensive literature on the asymptotic expansion of statistics: Akahira and Takeuchi (1981), Pfanzagl (1982, 1985), Bhattacharya and Rao (1986), Taniguchi (1991), Barndorff-Nielsen and Cox (1994), Ghosh (1994), etc. Bhattacharya and Ghosh (1978) founded a rigorous proof of the asymptotic expansion for a certain form of statistic under i.i.d. setting. Götze and Hipp (1983) presented an asymptotic expansion of the distribution of an additive functional of an approximately Markovian process with a discrete time parameter.

The accuracy of the approximation to the distribution depends on the regularity. The Cramér condition or the conditional type Cramér condition was assumed in those works to ensure the regularity of the distribution. In the independent case, the Cramér condition has a simple form and, for example, if one knows that the underlying random variable has the absolutely continuous distribution part, the condition follows from differentiable approximation to it and integration-by-parts, i.e., from the Riemann-Lebesgue theorem. For dependent observations, the Cramér condition may become more complicated and Götze and Hipp (1983) put a conditional type condition. It is not always a simple matter to verify such a conditional type condition, but Götze and Hipp (1994) did it for Markovian examples by using integration-by-parts formula over Euclidean space.

Here, we will adopt the continuous time, and it causes another difficulty in verifying the regularity. To solve this problem, it is necessary to handle conditional expectations and it involves an infinite-dimensional calculus because random variables are functions over the continuous-time path space. For this purpose, we use the Malliavin calculus, which features the integration-by-parts formula over infinite-dimensional spaces, and replace the conditional type Cramér condition by the nondegeneracy of the Malliavin covariance.

As for continuous-time processes, asymptotic expansions of statistics for the small diffusion model and a general small σ-model were obtained by Yoshida (1992*a*, 1992*b*, 1993, 1996*b*) and Uchida and Yoshida (2004), with the notion of the generalized Wiener functionals. See also Sakamoto and Yoshida (1996) and Dermoune and Kutoyants (1995). For continuous-time martingales, a second order expansion formula of the distribution was proved in Yoshida (1997) with the Malliavin calculus after Mykland's work (1992) on the expansion of smooth functionals without regularity condition inevitably; see Mykland (1993, 1995) for other developments for smooth functionals, and Yoshida (1996*a*, 1999) for a distribution expansion for martingale with jumps. As applications, the expansion of the distribution of the maximum likelihood estimator for an ergodic diffusion process was first presented in Yoshida (1997), and consecutively, that of the M-estimator in Sakamoto and Yoshida (1998*a*).

Recently, it was found in Kusuoka and Yoshida (2000) that another approach ("*local approach*") provides us with an effective solution for geometric-mixing-ϵ-Markov processes, while the martingale approach still has advantages for long memory time series models breaking the geometric mixing condition (Yoshida (1999)). They obtained asymptotic expansions of additive functionals of a geometric mixing, ϵ-Markov processes including time series and diffusion processes with jumps, and also provided an easily verifiable condition on the mixing property of the diffusion process. In order to obtain full generality as we mentioned, they adopted the Malliavin calculus formulated by Bichteler *et al.* (1987). Among other possible formulations of the Malliavin calculus for jump processes, it is a convenience due to a chain rule for the Γ-bilinear form.

When deriving asymptotic expansion of the distribution of a statistic, we often use its stochastic expansion. The simplest example would be an asymptotic expansion of $H(\bar{Y}_T)$, where H is a smooth function and $\bar{Y}_T$ is the sample mean of observations $(Y_t)_{t\in[0,T]}$, i.e., $\bar{Y}_T = (1/T)\int_0^T Y_t dt$. From Taylor's expansion, one has the second order stochastic expansion

$$\sqrt{T}(H(\bar{Y}_T) - H(\mu)) = H'(\mu)\sqrt{T}(\bar{Y}_T - \mu) \tag{1.1}$$

$$+\frac{1}{2\sqrt{T}}H''(\mu)(\sqrt{T}(\bar{Y}_T-\mu))^2+\text{remainder},$$

where $\bar{Y}_T \to \mu$ as $T \to \infty$ in probability, and under some conditions on the convergence rate of the remainder term, the second order asymptotic expansion of the distribution of $\sqrt{T}(H(\bar{Y}_T)-H(\mu))$ follows from this stochastic expansion. The other popular example is the maximum likelihood estimator $\hat{\theta}_T$ for a parameter θ of a probability density $p_T(y;\theta)$ of observations $Y=(Y_t)_{t\in[0,T]}$. It is well-known that $\hat{\theta}_T$ admits the second order stochastic expansion

$$\sqrt{T}(\hat{\theta}_T-\theta)=g^{-1}\dot{\ell}_T(\theta)+\frac{1}{\sqrt{T}}(a_1\dot{\ell}_T^2(\theta)+a_2\dot{\ell}_T(\theta)(\ddot{\ell}_T(\theta)+g))+\text{remainder}, \tag{1.2}$$

where g is the Fisher information, a_1 and a_2 are some constants, ℓ_T is the log-likelihood, i.e., $\ell_T(\theta)=\log p_T(Y,\theta)$, and $\dot{\ell}_T$ and $\ddot{\ell}_T$ are first and second derivatives with respect to θ. This stochastic expansion can be used for deriving the second order asymptotic expansion of the distribution of $\hat{\theta}_T$. In general and as in familiar cases, many of the statistics appearing in inference for stochastic processes have a stochastic expansion taking the form of

$$\text{statistic}=S_T+\text{remainder},$$

where

$$S_T=\bar{Z}_T^{(0)}+\sum_{i=1}^{k}\frac{1}{T^{i/2}}Q_i(\bar{Z}_T^{(0)},\bar{Z}_T^{(1)}), \tag{1.3}$$

and $T>0$ is the terminal time of observations, $\bar{Z}_T=(\bar{Z}_T^{(0)},\bar{Z}_T^{(1)})$ is a vector of functionals, and $\{Q_i\}$ are some polynomials with coefficients being bounded as $T\to\infty$. Therefore, once we have a formula for the asymptotic expansion of S_T, we can easily derive valid higher order asymptotic expansions of the distribution in various statistical models.

Concerning S_T-type of random variables, Bhattacharya and Ghosh (1978) discussed the asymptotic expansion of the maximum likelihood estimator based on i.i.d. observations by using a map which is often referred to as the Bhattacharya-Ghosh map. The functional S_T itself was dealt with by Götze and Hipp (1994) and by Kusuoka and Yoshida (2000), and a program for the derivation of its valid asymptotic expansion was prepared there by using the Bhattacharya-Ghosh map. Since most of statistics have such a stochastic expansion, the program is applicable to many statistical models. However, when higher-order asymptotic properties of a statistic of interest are discussed, it is necessary to perform rather a lot of calculation in order to obtain the asymptotic expansion for each statistic explicitly. The readers will in later sections find that this problem unexpectedly requires a lot of technicalities to settle than computational difficulties.

In this article, we will carry out the program to obtain an *explicit* formula for the k-th order asymptotic expansion of the distribution of S_T and prove the validity in the case where the underlying process of $\bar{Z}_T$ is an ϵ-Markov process with a geometric-mixing property. For this purpose, in Section 5, we will first give a complete description of the approximating density to S_T. Orthogonality between the principal part $\bar{Z}_T^{(0)}$ and the *ancillary* part $\bar{Z}_T^{(1)}$ makes calculations easier. Thus in the second step, we will present a formula under orthogonality. Those results require the non-degeneracy of the covariance of $\bar{Z}_T$. However, in application, we meet examples which have a linear relation among the ancillary variables. Later, we will precisely discuss the M-estimator ("Z-estimator") but the maximum likelihood estimator, a special case, has this degeneracy in its ancillary variables because of the symmetry. It will be shown that our explicit formula is still valid even under such degeneracy if the density formula is interpreted as a Schwartz distribution.

As its typical and useful application, we will discuss the third order asymptotic expansion of an M-estimator in Section 6. After a full investigation into the existence of the M-estimator and the convergence rate of the remainder term of its stochastic expansion, we will present the third order asymptotic expansion of the M-estimator for a general model, by using the results for S_T. We also show that if the expectations of the derivatives of the contrast function satisfy certain relations, which are called the Bartlett identities, coefficients of an asymptotic expansion of a minimum contrast estimator can be represented in terms of the information geometry.

The results in this article have many applications including the model selection problem and the asymptotic expansions of diffusion functionals. In fact, Uchida and Yoshida (1999, 2001) used Theorem 2.1 given in the original manuscript, Sakamoto and Yoshida (1999), of this paper, which is the same one as Theorem 5.1 in this paper, to unify traditional information criteria such as AIC, TIC, GIC, and to present new criteria in the light of the asymptotic expansion for stochastic processes. The manuscript, Sakamoto and Yoshida (1999), is unpublished but already circulated among experts of this field. Third order asymptotic expansions of M-estimators for diffusion processes were also, in Sakamoto and Yoshida (1999), obtained from Theorem 3.4 in Sakamoto and Yoshida (1999) (Theorem 6.4 in this paper). At the end of Section 6, we will show its specific version, the third order asymptotic expansion of the maximum likelihood estimator for the diffusion process, without the proof. Note that before Sakamoto and Yoshida (1999), the third order expansion of MLE for the diffusion process was in Sakamoto and Yoshida (1998*b*) obtained from Kusuoka and Yoshida (2000) under the Bartlett identities and some relations of the derivatives of the likelihood (the assumptions [BI1] $\sim$ [BI4] and [DV1] $\sim$ [DV3] in Section 6 of this paper), while the expansion in Section 6 was derived without such assumptions. The details about the proofs for the results of this paper, which are given in Sections 7 and 8, are also found in Sakamoto and Yoshida (1999) or Sakamoto (1998).

For the readers who are not familiar to the Malliavin calculus, we will in Section 3 explain the integration-by-parts formulas over the Wiener space as well as those over the finite-dimensional space, after the introduction of the ϵ-Markov process which is assumed to be the underlying process of $\bar{Z}_T$. They will in Section 4 be summarized in terms of the Malliavin operator defined by Bichteler *et al.* (1987), and in Section 5 the assumptions for the asymptotic expansion of S_T will be described by the terminology of the Malliavin calculus. See Kutoyants (2004) for inference for ergodic diffusion processes.

2. ϵ-Markov process

In this article, we will consider a class of stochastic processes as a basis on which expansion formulas are validated. We shall begin with examples.

2.1 *Diffusion process*

We denote by $C_{\uparrow}^{\infty}$ the set of smooth functions whose derivatives are of at most polynomial growth, and by C_b^{∞} the set of smooth mappings whose derivatives of order

≥ 1 are bounded. Suppose that $X = (X_t)_{t\in\mathbb{R}_+}$ is a stationary diffusion process satisfying the stochastic differential equation

$$dX_t = V_0(X_t)dt + \sum_{\alpha=1}^{r} V(X_t)_\alpha \circ dw_t^\alpha, \tag{2.1}$$
$$X_0 = x_0,$$

where $V_0 \in C_b^\infty(\mathbb{R}^d;\mathbb{R}^d)$, $V = (V_\alpha) \in C_b^\infty(\mathbb{R}^d;\mathbb{R}^d \otimes \mathbb{R}^r)$, $w = (w^\alpha)$ is an r-dimensional Wiener process, and x_0 is an initial random variable the distribution $\mathcal{L}\{x_0\}$ of which coincides with the stationary distribution of X. The circle $\circ$ means the Stratonovich integral, which is useful to describe Lie algebras.

We consider another stochastic process $Z = (Z_t)_{t\in\mathbb{R}_+}$ defined by

$$Z_t = Z_0 + \int_0^t V_0^*(X_s)ds + \sum_{\alpha=1}^{r} \int_0^t V_\alpha^*(X_s) \circ dw_s^\alpha, \tag{2.2}$$

where Z_0 is a $\sigma[x_0]$ measurable random variable, $V_0^* \in C_\uparrow^\infty(\mathbb{R}^d;\mathbb{R}^n)$ and $V^* = (V_\alpha^*) \in C_\uparrow^\infty(\mathbb{R}^d;\mathbb{R}^n \otimes \mathbb{R}^r)$. Equations (2.1) and (2.2) form a state-space model. Hidden Markov model and stochastic volatility model are examples. The noises in these two equations may be taken differently if we use an expression with degenerate V and V^*. Furthermore, if (2.1) has unknown parameters in its drift, then the log-likelihood function and its derivatives take the form of (2.2) for the true parameter value.

We assume that $E[Z_t] = 0$ for all $t \in \mathbb{R}_+$. It is the case if $E[Z_0] = 0$ and if $E[\tilde{V}_0^*(X_0)] = 0$ for

$$\tilde{V}_0^*(x) = V_0^*(x) + \frac{1}{2}\sum_{\alpha=1}^{r} V_\alpha V_\alpha^*(x).$$

Here we identified the vector V_α with the vector field $\sum_{i=1}^d V_\alpha^i(x)\partial_i$. Even if $c := E[\tilde{V}_0^*(X_0)]$ is not null, we can deform Z_t by subtracting c from the original drift.

2.2 *Non-linear moving-average series*

Let $y = (y_t)_{t\in\mathbb{Z}_+}$ be a non-linear moving-average defined by

$$y_t = h(\xi_{t-m+1}, \cdots, \xi_t), \tag{2.3}$$

where $(\xi_j)_{j\in\mathbb{Z}}$ be an $\mathbb{R}^r$-valued i.i.d. sequence and $h : \mathbb{R}^{rm} \to \mathbb{R}^d$ is a measurable function. In time series models, the asymptotic expansions of many statistics can be derived from the expansion of $T^{-1/2}Z_T$, where

$$Z_t = \sum_{j=1}^{t}(y_j - E[y_j]). \tag{2.4}$$

Note that Götze and Hipp (1994) considered a slightly different case and gave the asymptotic expansion of $T^{-1/2}Z_T$ with applications to stationary ARMA processes and non-linear AR processes.

2.3 *Cluster process*

For $L_1, L_2 > 0$, let

$$E = \mathbb{N} \times \{(t_j)_{j\in\mathbb{N}} : 0 < t_j \le L_1, j \in \mathbb{N}\} \times S^{\mathbb{N}}, \tag{2.5}$$

where S is a set of real-valued functions on $\mathbb{R}$ with supports in $[0, L_2]$, and let μ be a Poisson random measure on $\mathbb{R}\times E$ with compensator ν satisfying $\nu(I\times E) < \infty$ for every bounded interval $I \subset \mathbb{R}$. Denote by $(T_j^{(c)})$ the increasing sequence of the occurrence times of the counting process $\mu([-L_1 - L_2, t] \times E)$ and by $\beta_j = (M_j, (T_{jk}^{(s)})_{k\in\mathbb{N}}, (\tilde{X}_{jk}(\cdot))_{k\in\mathbb{N}})$ a element of E associated with $T_j^{(c)}$ for every $j \in \mathbb{N}$.

For this marked point process $(T_j^{(c)}, \beta_j)$, we consider a process $y = (y_t)_{t\in\mathbb{R}_+}$ defined by

$$y_t = \sum_{j=-\infty}^{\infty} \sum_{k=1}^{M_j} \tilde{X}_{jk}(t - T_{jk}^{(s)} - T_j^{(c)}). \tag{2.6}$$

It is a kind of the cluster process, which is often used for a rainfall model.

When the distribution of y involves an unknown parameter, the asymptotic expansion of the Yule-Walker estimator can be obtained from one of $T^{-1/2}Z_T$, where

$$Z_t = \int_0^t (y_s - E[y_s])ds. \tag{2.7}$$

2.4 *Diffusion process with jumps*

Assume that a stochastic process $(Y_t)_{t\in\mathbb{R}_+}$ is defined as a strong solution of the stochastic integral equation with jumps:

$$Y_T = Y_0 + \int_0^T A(Y_{t-})dt + \int_0^T B(Y_{t-})dw_t + \int_0^T \int_E C(Y_{t-})\tilde{\mu}(dt, dx) \tag{2.8}$$

where $A \in C^\infty(\mathbb{R}^d; \mathbb{R}^d)$, $B \in C^\infty(\mathbb{R}^d; \mathbb{R}^d \otimes \mathbb{R}^m)$, $C \in C^\infty(\mathbb{R}^d \times E; \mathbb{R}^d)$. The process w is an m-dimensional Wiener process, and $\tilde{\mu}$ is a compensated Poisson random measure on $\mathbb{R}_+ \times E$ with intensity $dt \otimes \lambda(dx)$ for an open set E in $\mathbb{R}^b$. λ is the Lebesgue measure on E compensating μ. A set of regularity conditions ensures the existence and the uniqueness of Y.

For this underlying process Y, we consider another process $Z = (Z_t)_{t\in\mathbb{R}_+}$. Z is a process which satisfies the equation:

$$Z_T = Z_0 + \int_0^T A'(Y_{t-})dt + \int_0^T B'(Y_{t-})dw_t + \int_0^T \int_E C'(Y_{t-})\tilde{\mu}(dt, dx),$$

where Z_0 is $\sigma[Y_0]$-measurable, $A' \in C^\infty(\mathbb{R}^d; \mathbb{R}^d)$, $B' \in C^\infty(\mathbb{R}^d; \mathbb{R}^d \otimes \mathbb{R}^m)$, $C' \in C^\infty(\mathbb{R}^d \times E; \mathbb{R}^d)$. Then (Y, Z) forms a state-space model.

2.5 *Definition of the ϵ-Markov process*

In the previous subsections, we viewed examples in our scope. In order to handle those models (and other many models we do not mention in this article) at a time, we will consider a general class of stochastic processes called the ϵ-Markov process.

Let $(\Omega, \mathcal{F}, P)$ be a probability space, $Y = (Y_t)_{t\in\mathbb{R}_+}$ an $\mathbb{R}^d$-valued càdlàg process (or a separable process) defined on Ω, and $X = (X_t)_{t\in\mathbb{R}_+}$ an $\mathbb{R}^r$-valued càdlàg process defined on Ω. Suppose that for any $t \in \mathbb{R}_+$, $\mathcal{B}^{X,Y}_{[0,t]}$ is independent of $\mathcal{B}^{dX}_{[t,\infty]}$, where

$$\mathcal{B}^{X,Y}_{[0,t]} = \sigma[X_u, Y_u : u \in [0,t]] \vee \mathcal{N},$$
$$\mathcal{B}^{dX}_{I} = \sigma[X_t - X_s : s,t \in I \cap \mathbb{R}_+] \vee \mathcal{N} \qquad \text{for} \quad I \subset \mathbb{R},$$

and $\mathcal{N}$ is the σ-field generated by null sets. For $I \subset \mathbb{R}$, define sub σ-fields $\mathcal{B}^Y_I$ and $\mathcal{B}_I$ by

$$\mathcal{B}^Y_I = \sigma[Y_t : t \in I \cap \mathbb{R}_+] \vee \mathcal{N}$$

and

$$\mathcal{B}_I = \sigma[X_t - X_s, Y_t : s,t \in I \cap \mathbb{R}_+] \vee \mathcal{N},$$

respectively. Assume that there exists a positive constant ϵ such that for any $s > 0$ and $t > 0$ satisfying $\epsilon \le s \le t$,

$$Y_t \in \mathcal{F}(\mathcal{B}^Y_{[s-\epsilon,s]} \vee \mathcal{B}^{dX}_{[s,t]}),$$

where for any sub σ-field $\mathcal{A}$ of $\mathcal{F}$, $\mathcal{F}\mathcal{A}$ denotes the set of all $\mathcal{A}$-measurable functions. If a process Y satisfies the above condition, it is called an ϵ-Markov process driven by X.

For processes X and Y, we consider the third process $Z = (Z_t)$ for which asymptotic expansion will be derived. Let $Z = (Z_t)_{t\in\mathbb{R}_+}$ be an $\mathbb{R}^n$-valued process satisfying $Z_0 \in \mathcal{F}\mathcal{B}_{[0]}$ and

$$Z^s_t := Z_t - Z_s \in \mathcal{F}\mathcal{B}_{[s,t]}, \qquad \text{for every} \quad s,t \in \mathbb{R}_+, \ 0 \le s \le t. \tag{2.9}$$

A process Z having the property (2.9) is often called an additive functional of X and Y.

Example. (1) (*diffusion process*) The diffusion process X defined by (2.1) is a Markov (0-Markov) process driven by the Wiener space w and the process Z defined by (2.2) is an additive functional of X and w in the above sense.

(2) (*non-linear moving-average series*) Let $Y = (Y_t)_{t\in\mathbb{Z}_+}$ and $X = (X_t)_{t\in\mathbb{Z}_+}$ be processes defined by $Y_t = (\xi_t, y_t)$ and $X_t = \sum_{j=0}^t \xi_j$ for an i.i.d. sequence $(\xi_t)_{t\in\mathbb{Z}_+}$ and the non-linear time series $(y_t)_{t\in\mathbb{Z}_+}$ defined by (2.3), then Y is an ϵ-Markov process driven by X with $\epsilon = m-2$ and $(Y_t)_{t\in\mathbb{Z}_+}$ and $(X_t)_{t\in\mathbb{Z}_+}$ can be embedded into continuous-time processes $(Y_t)_{t\in\mathbb{R}_+}$ and $(X_t)_{t\in\mathbb{R}_+}$ as $X_t = X_{[t]}$ and $Y_t = Y_{[t]}$ for all $t \in \mathbb{R}_+$. A continuous-time process $(Z_t)_{t\in\mathbb{R}_+}$ defined by $(Z_t)_{t\in\mathbb{Z}_+}$ in (2.4) as X and Y is also an additive functional of X and Y.

(3) (*cluster process*) Let $(y_t)_{t\in\mathbb{R}_+}$ be the cluster process defined by (2.6), and $(E_i)_{i=1}^\infty$ a family of subsets of a configuration space E given by (2.5) such that it determines a measure on E, then $X_t = (\mu([-L_1 - L_2, t] \times E_i))_{i=1}^\infty$ is an ∞-dimensional representation of the Poisson process μ. Putting $Y_t = (X_t, y_t)$, it is seen that $Y = (Y_t)_{t\in\mathbb{R}_+}$ is an ϵ-Markov process driven by X with $\epsilon = L_1 + L_2$. The process $(Z_t)_{t\in\mathbb{R}_+}$ defined by (2.7) is an additive functional of X and Y.

(4) (*diffusion process with jumps*) For the stochastic differential equation with jumps (2.8), the driving process X_t can be taken as $X_t = (w_t, \mu_t(g_i); i \in \mathbb{N})$, where (g_i) is a countable measure-determining family over E. Thus, Y becomes a Markov process (i.e., 0-Markov process) driven by X with independent increments according to our definition. Another process Z is also additive functional of X and Y.

3. IBP-formula and conditional type Cramér condition

It is known that the validity of the asymptotic expansion depends on the regularity of the distribution approximated by the expansion. The most successful condition in independent setting was the Cramér condition. The so-called conditional type Cramér condition played a similar role for discrete-time stochastic processes, as it is seen in Götze and Hipp (1983, 1994).

By the ϵ-Markovian property, it is possible to factorize the characteristic function of $Z_T/\sqrt{T}$ into the product of conditional characteristic functions over small time-intervals. In this way, the estimate of the global characteristic function is reduced to that of the local conditional characteristic functions. The conditioning is usually done by the value of the process Y at the right and left end points of the short time interval. Thus, roughly speaking, when $\epsilon = 0$, it becomes the problem of how to estimate the conditional characteristic function

$$E[e^{iu\cdot(Z_v - Z_u)/\sqrt{T}} \mid Y_u, Y_v]$$

of the normalized increment $(Z_v - Z_u)/\sqrt{T}$ over the interval $[u, v]$.

For continuous-times processes, such a condition is in general hard to check, and it is replaced by a more practical condition, that is, the non-degeneracy of the Malliavin covariance of a certain functional. We will explain how the non-degeneracy of the Malliavin covariance and the induced integration-by-parts formula (IBP-formula) work in examples. For details, we refer the readers to Kusuoka and Yoshida (2000), and Yoshida (2001) in which the readers finds precise description of the derivation and much weaker sufficient conditions for non-degeneracy by support theorems.

Here again, we begin with examples.

3.1 *Diffusion process*

We will again consider the diffusion process in (2.1). Corresponding to the stochastic process (X_t, Z_t), let us consider the stochastic flow $(\bar{X}_t(0,\bar{x}), \bar{x} \in \mathbb{R}^{d+n})$ defined by the enlarged stochastic differential equation

$$\begin{aligned} d\bar{X}_t(0,\bar{x}) &= \bar{V}_0(\bar{X}_t(0,\bar{x}))dt + \sum_{\alpha=1}^{r} \bar{V}_\alpha(\bar{X}_t(0,\bar{x})) \circ dw_t^\alpha, \\ \bar{X}_0(0,\bar{x}) &= \bar{x}, \end{aligned}$$

where

$$\bar{V}_0 = \begin{bmatrix} V_0 \\ V_0^* \end{bmatrix} \quad \text{and} \quad \bar{V} = \begin{bmatrix} V \\ V^* \end{bmatrix}.$$

We will assume the geometrical strong mixing condition for X and the necessary integrability conditions. In order to prove the validity of the Edgeworth expansion of $Z_T/\sqrt{T}$, it is necessary to estimate its characteristic function. Then, we take a sequence of intervals of length one ('reduction intervals') and divide the estimate into that over each subinterval. Thus we may only consider the interval $[0, 1]$ due to the stationarity, and hence it suffices to show that the conditional type Cramér condition (Yoshida (2001), Kusuoka and Yoshida (2000)):

[CD] There exist some point $x_* \in \mathbb{R}^d$ and some positive constants η, η_1, η_2 ($\eta_1 +$

$\eta_2 < 1$) and A such that

$$\sup_{|u|\geq A}\sup_{x\in B(x_*;\eta)} \boldsymbol{E}[|\boldsymbol{E}[e^{iu\cdot\bar{X}_1^{(2)}(0,(x,0))} \mid \bar{X}_1^{(1)}(0,(x,0))]|] \leq \eta_1$$

and

$$P[X_0 \in B(x_*;\eta)] > 1-\eta_2.$$

Here we denoted $\bar{X}$ by $(\bar{X}^{(1)}, \bar{X}^{(2)})$ and $\{x \in \mathbb{R}^d : |x-x_*| < \eta\}$ by $B(x_*;\eta)$.

Similarly to independent cases under the assumption of the existence of absolutely continuous part, in order to prove the sufficiency of [CD], an integration-by-parts formula is applied.

We shall start with integration-by-parts formulas over a finite-dimensional space. Let $\phi(w) = \phi(w;0,I_k)$ be the k-dimensional standard normal density. Denote by $C_\uparrow^\infty(\mathbb{R}^k)$ the set of smooth functions on $\mathbb{R}^k$ all derivatives of which are at most polynomial growth. Let

$$DF = \nabla F$$

for $F \in C_\uparrow^\infty(\mathbb{R}^k)$, and let

$$D^*G(w) = -\operatorname{div} G(w) + \sum_{i=1}^k G^i(w)w^i$$

for $G = (G^i)_{i=1}^k \in C_\uparrow^\infty(\mathbb{R}^k;\mathbb{R}^k)$. The second order differential operator L is defined by $L = -D^*D/2$. Then easy computation yields

$$\int_{\mathbb{R}^k} \langle DF(w), G(w)\rangle_{\mathbb{R}^k}\phi(w)dw = \int_{\mathbb{R}^k} F(w)D^*G(w)\phi(w)dw. \tag{3.1}$$

It is equivalent to so-called Stein's identity.

Now, we let

$$\Gamma(A,B) = \langle DA, DB\rangle_{\mathbb{R}^k}$$

for $A, B \in C_\uparrow^\infty(\mathbb{R}^k)$. Then from (3.1), we obtain

$$\begin{aligned}\sum_{l=1}^k \int_{\mathbb{R}^k} &(\partial_l f)(F(w))\Gamma(F^l(w), B(w))A(w)\phi(w)dw \\ &= \int_{\mathbb{R}^k} f(F(w))D^*(A(w)DB(w))\phi(w)dw\end{aligned} \tag{3.2}$$

for $F \in C_\uparrow^\infty(\mathbb{R}^k;\mathbb{R}^d)$ and $f \in C_\uparrow^\infty(\mathbb{R}^d)$. Moreover, let

$$\sigma_F = (\sigma_F^{lm})_{l,m=1}^d, \qquad \sigma_F^{lm} = \Gamma(F^l, F^m). \tag{3.3}$$

We call σ_F the Malliavin covariance matrix of F. Denote by γ_F^{lm} the (l,m)-element of σ_F^{-1}, and put $\Delta_F = \det\sigma_F$. In addition, for $\mathcal{Z} = (F,G)$, $F \in C_\uparrow^\infty(\mathbb{R}^k;\mathbb{R}^d)$, $G \in C_\uparrow^\infty(\mathbb{R}^k;\mathbb{R}^n)$, let

$$\tilde{\sigma}_{\mathcal{Z}}^{lm} = \sigma_F^{lm} - \sum_{l',m'}^n \Gamma(F^l, G^{l'})\gamma_G^{l'm'}\Gamma(G^{m'}, F^m)$$

and $\tilde{\gamma}_{\mathcal{Z}}^{lm}$ the (l,m)-element of the inverse matrix of $(\tilde{\sigma}_{\mathcal{Z}}^{lm})$. Then, under suitable integrability condition for $\Delta_{\mathcal{Z}}^{-1}$, it follows from (3.2) that for $J \in C_{\uparrow}^{\infty}(\mathbb{R}^k)$ and $g \in C_{\uparrow}^{\infty}(\mathbb{R}^n)$,

$$\begin{aligned}(3.4)\qquad &\int_{\mathbb{R}^k} (\partial_l f)(F(w))J(w)g(G(w))\phi(w)dw \\ &\quad = \int_{\mathbb{R}^k} f(F(w))\Psi_{l,\mathcal{Z}}^*(J)(w)g(G(w))\phi(w)dw,\end{aligned}$$

where

$$\begin{aligned}\Psi_{l,\mathcal{Z}}^*(J) &= \sum_{m}^{d} \Psi_{m,\mathcal{Z}}(\tilde{\gamma}_{\mathcal{Z}}^{lm} J), \\ \Psi_{l,\mathcal{Z}}(J) &= -\Gamma(J, F^l) - 2JLF^l \\ &\quad + \sum_{l',m'}^{n} \{\Gamma(\gamma_G^{l',m'}\Gamma(F^l, G^{l'})J, G^{m'}) + 2\gamma_G^{l',m'}\Gamma(F^l, G^{l'})JLG^{m'}\}.\end{aligned}$$

Note that L is the second order differential operator defined by $L = -D^*D/2$. Formulas (3.2) and (3.4) may be called integration-by-parts formulas.

For diffusion processes, the sample space is a Wiener space, an infinite-dimensional space W^r of continuous functions $w : \mathbb{R}_+ \to \mathbb{R}^r$ with $w_0 = 0$, so that we cannot apply formula (3.4) itself. However, it is possible to construct an infinite-dimensional analog of it over the Wiener space. Namely, it is possible to define the gradient operator D, the divergence operator D^*, the second order differential operator L, the bilinear form Γ, and the Malliavin covariance matrix σ_F of F, and an integration-by-parts formulas exists:

$$\begin{aligned}(3.5)\qquad &\int_{W^r} (\partial_l f)(F(w))J(w)g(G(w))\boldsymbol{P}(dw) \\ &\quad = \int_{W^r} f(F(w))\Psi_{l,\mathcal{Z}}^*(J)(w)g(G(w))\boldsymbol{P}(dw),\end{aligned}$$

where $\boldsymbol{P}$ is a Wiener measure on W^r. See Ikeda and Watanabe (1989) for the definitions and properties of the operators D, D^*, L. A slightly simple version of IBP formula is given there.

In our present case of diffusion, we use (3.5) for

$$\mathcal{Z} = (\bar{X}_1^{(2)}(0,(x,0)), \bar{X}_1^{(1)}(0,(x,0))).$$

For a while, we assume that for some $\gamma > 0$,

$$(3.6)\qquad \sup_{\bar{x}\in B(\bar{x}_*;\gamma)} \boldsymbol{E}[(\det \sigma_{\bar{X}_1(0,\bar{x})})^{-p}] < \infty \qquad (p \in (1,\infty)),$$

where $\bar{x}_* = (x_*, 0)$. Under a usual regularity condition for existence of the solution, it holds that

$$\sup_{\bar{x}\in B(\bar{x}_*;b)} \|D^s\bar{X}_1(0,\bar{x})\|_p < \infty$$

for every $b, p \in (1,\infty)$ and $s \in \mathbb{Z}_+$. Denote by $\varphi : \mathbb{R}_+ \to [0,1]$ a smooth function such that $\varphi(r) = 1$ for $r \in [0,1/2]$ and that $\varphi(r) = 0$ for $r \geq 1$. We will use a truncation functional ψ_j given by

$$\psi_j = \varphi(|X_{2j-1} - x_*|^2/\gamma^2).$$

If the point x_* is in the support of the stationary distribution of the process Y, then we may assume that

$$E[\varphi(|X_1 - x_*|^2/\gamma^2)] > 0.$$

Then we obviously see that

$$\begin{aligned} &E[\psi_j e^{iu\cdot Z_{2j}^{2j-1}} \mid X_{2j-1} = y_0, X_{2j} = y_1] \\ &\quad = \boldsymbol{E}[\hat{\psi}(y_0)e^{iu\cdot \bar{X}_1^{(2)}(0,(y_0,0))} \mid \bar{X}_1^{(1)}(0,(y_0,0)) = y_1], \end{aligned}$$

where $\hat{\psi}(y_0) = \varphi(|y_0 - x_*|^2/\gamma^2)$ and $\bar{X}_t(0,\bar{x}) = (\bar{X}_t^{(1)}(0,\bar{x}), \bar{X}_t^{(2)}(0,\bar{x}))$. By the equivalence of the distributions, we will execute computations over the Wiener space where the flow is constructed.

Let $P^{(1),y_0}$ denote the distribution of $\bar{X}_1^{(1)}(0,(y_0,0))$. The IBP-formula (3.5) applied to $f(F) = e^{iu\cdot F}$ and $g(G) = e^{iv\cdot G}$ with $J = \hat{\psi}$ and $\mathcal{Z} = (F,G) = (\bar{X}_1^{(2)}(0,(y_0,0)), \bar{X}_1^{(1)}(0,(y_0,0)))$ implies that

$$\begin{aligned} &\int \boldsymbol{E}[\hat{\psi}(y_0)e^{iu\cdot \bar{X}_1^{(2)}(0,(y_0,0))} \mid \bar{X}_1^{(1)}(0,(y_0,0)) = y_1] e^{iv\cdot y_1} P^{(1),y_0}(dy_1) \\ &\quad = (iu_l)^{-1} \int \boldsymbol{E}[e^{iu\cdot \bar{X}_1^{(2)}(0,(y_0,0))} \Psi_{l,\mathcal{Z}}^*(\hat{\psi}(y_0)) \mid \bar{X}_1^{(1)}(0,(y_0,0)) = y_1] e^{iv\cdot y_1} P^{(1),y_0}(dy_1). \end{aligned}$$

Therefore, the uniqueness of the Fourier transform leads us to

$$\begin{aligned} &\boldsymbol{E}[\hat{\psi}(y_0)e^{iu\cdot \bar{X}_1^{(2)}(0,(y_0,0))} \mid \bar{X}_1^{(1)}(0,(y_0,0))] \\ &\quad = (iu_l)^{-1} \boldsymbol{E}[e^{iu\cdot \bar{X}_1^{(2)}(0,(y_0,0))} \Psi_{l,\mathcal{Z}}^*(\hat{\psi}(y_0)) \mid \bar{X}_1^{(1)}(0,(y_0,0))] \qquad \boldsymbol{P}-\text{a.s.} \end{aligned}$$

In this way, we have obtained:

$$\boldsymbol{E}\left[\sup_{|u|\geq B} |\boldsymbol{E}[\hat{\psi}(y_0)e^{iu\cdot \bar{X}_1^{(2)}(0,(y_0,0))} \mid \bar{X}_1^{(1)}(0,(y_0,0))]|\right] \leq \frac{C}{B}\boldsymbol{E}[|\Psi_{l,\mathcal{Z}}^*(\hat{\psi}(y_0))|]$$

for some constant $C > 0$.

Generally, it is not easy to prove (3.6) for diffusion processes but a Hörmander type condition is a convenience. For vector fields $V_0, V_1, \ldots, V_r$, let $\Sigma_0 = \{V_1, \ldots, V_r\}$ and $\Sigma_n = \{[V_\alpha, V]; V \in \Sigma_{n-1}, \alpha = 0,1,\ldots,r\}$ for $n \in \mathbb{N}$, where $[,]$ is the Poisson bracket. Moreover, let $Lie[V_0; V_1, \ldots, V_r]$ be the linear manifold spanned by $\cup_{n=0}^\infty \Sigma_n$. The following condition is a sufficient condition for [CD] (Kusuoka and Yoshida (2000), Yoshida (2001)):

[DH] There exists a point $x_* \in \mathbb{R}^d$ for which the following conditions are satisfied: (i) For any $\eta > 0$, $P[X_0 \in B(x_*;\eta)] > 0$. (ii) For $\bar{x}_* = (x_*, 0)$,

$$Lie[\bar{V}_0; \bar{V}_1, \ldots, \bar{V}_r](\bar{x}_*) = \mathbb{R}^{d+n}.$$

The condition [DH] (ii) is called the Hörmander condition. The advantage of [DH] is that it can easily be verified only by differential computations. See Yoshida (2001) for details of this condition and other mild sufficient conditions.

3.2 *Non-linear moving-average series*

Let τ be a positive integer greater than $\epsilon = m-2$, and $I(j) = [u(j), v(j)]_{j\in\mathbb{N}}$ a sequence of intervals defined by $u(j) = 2\tau j$ and $v(j) = 2\tau j + \tau$. As in the case of the diffusion process above, we may reduce the estimation of the characteristic function of $T^{-1/2}Z_T$ to that of the conditional characteristic function

$$E[\exp(iu\cdot(y_{u(j)+1}+\cdots+y_{v(j)})) \mid \mathcal{B}^{dX}_{[u(j)-\epsilon,u(j)]} \vee \mathcal{B}^{dX}_{[v(j)-\epsilon,v(j)]}],$$

where $\mathcal{B}^{dX}_I = \sigma(\xi_t; t\in I)$. Therefore, due to the stationarity, we only have to check the following conditional type Cramér condition: there exists a measurable set $B \in \mathcal{B}_{(2m-2)r}$ such that $P[(\xi_1,\dots,\xi_{2m-2})\in B] > 0$ and that

$$\sup_{|u|\geq A}\ \sup_{(c_1,\dots,c_{2m-2})\in B} |E[\exp(iu\cdot H(c_1,\dots,c_{2m-2},\xi_1,\dots,\xi_N))]| < 1 \tag{3.7}$$

for some $A>0$, where $N = \tau - m + 1$ and

$$\begin{aligned}H(c_1,\dots,c_{2m-2},x_1,\dots,x_N) &= h(c_1,\dots,c_{m-1},x_1) + h(c_2,\dots,c_{m-1},x_1,x_2)+\cdots\\ &\quad + h(x_{N-1},x_N,c_m,\dots,c_{2m-3}) + h(x_N,c_m,\dots,c_{2m-2}).\end{aligned}$$

As before, to verify (3.7), we consider the integration-by-parts formula for functions of $(\xi_i)_{i=1}^N$. Denote by P^{ξ_1} the probability measure induced by ξ_1, and suppose that the decomposition of P^{ξ_1} is given by

$$(1-\lambda)\nu(dx) + \lambda p(x)dx$$

for a constant $\lambda\in(0,1]$, a probability measure ν and a density $p\in C^\infty(\mathbb{R}^r;\mathbb{R}_+)$. Note that ν has a part of the absolutely continuous part of the Lebesgue decomposition for P^{ξ_1}. Define $I = \{x\in\mathbb{R}^r : p(x)>0\}$. Under this assumption, the expectation of a function g of $\xi_1,\dots,\xi_N$ can be rewritten as

$$\int_{\mathfrak{X}^N} g(\pi_1u_1+(1-\pi_1)v_1,\dots,\pi_Nu_N+(1-\pi_N)v_N)d\boldsymbol{P}(\boldsymbol{w}),$$

where $\mathfrak{X} = \{0,1\}\times\mathbb{R}^r\times\mathbb{R}^r$, $\boldsymbol{w} = (\pi,u,v)$, $\pi = (\pi_1,\dots,\pi_N)$, $u=(u_1,\dots,u_N)$, $v = (v_1,\dots,v_N)$, $\boldsymbol{P}(\boldsymbol{w}) = \prod_{i=1}^N B(1;\lambda)(\pi_i)\times p(u_i)du_i\times\nu(dv_i)$ and $B(n;\lambda)$ is the binomial probability measure with the trial number n and the occurrence probability λ, and hence the integration-by-parts formula we need is reduced to that for functions of u (the partial Malliavin calculus):

$$\begin{aligned}&\int_{I^N}(\partial_l f)(F(u))J(u)\boldsymbol{p}(u)du \\ &\qquad = \int_{I^N} F(u)D^*\left(J(u)\sum_{m=1}^k \gamma_{ml}(u)DF^m(u)\right)\boldsymbol{p}(u)du\end{aligned} \tag{3.8}$$

for smooth functions $f:\mathbb{R}^k\to\mathbb{R}$, $F:\mathbb{R}^{rN}\to\mathbb{R}^k$, $J:\mathbb{R}^{rN}\to\mathbb{R}$. Here $\boldsymbol{p}(u) = \prod_{j=1}^n p(u_j)$, D and D^* are differential operators defined by

$$DF(u) = \nabla_u F(u),\qquad D^*G(u) = -\left(\nabla_u + \frac{1}{\boldsymbol{p}(u)}\nabla_u\boldsymbol{p}(u)\right)\cdot G(u)$$

for any $F : \mathbb{R}^{rN} \to \mathbb{R}$ and $G : \mathbb{R}^{rN} \to \mathbb{R}^{rN}$, and γ_F^{lm} is the (l, m)-element of the inverse of the Malliavin covariance σ_F defined by

$$\sigma_F^{lm}(u) = \langle DF^l(u), DF^m(u)\rangle_{\mathbb{R}^{rN}}.$$

Here we assumed a regularity condition for p near the boundary of I. Note that we only need the integration-by-parts formula over the interval I for the estimation of the characteristic function because the integration over the outside of I can be estimated by a constant less than one:

$$|E[e^{it\cdot F(\xi_1,\dots,\xi_N)}]| \le \left|\int_{I^N} e^{it\cdot F(u)}\boldsymbol{p}(u)du\right| + 1 - \lambda^N.$$

As in the previous subsection, in order to apply the integration-by-parts formula, we need the degeneracy of the Malliavin covariance:

$$\boldsymbol{E}[\psi(\det \sigma_F)^{-p}] < \infty$$

for some truncation functional ψ. Note that we can choose τ or N so that this condition is satisfied.

3.3 *Other processes*

For the cluster process, we can construct IBP formulas and verify the conditional type Cramér condition in an analogous way. See Sakamoto and Yoshida (2000) for the detail. The IBP formula for the diffusion process with jump is given in Bichteler *et al.* (1987) and Kusuoka and Yoshida (2000).

4. Malliavin operator

In the previous section, we observed that the integration-by-parts formula played an essential role to deduce the conditional type Cramér conditions. Moreover, we also saw the role of the infinite-dimensional differential operators D and D^* in the diffusion case as well as a finite-dimensional case. In order to treat jump type processes, several possible formulations are nowadays available. Some of them are only for Y_T defined by a stochastic differential equation. From the point of the applicability for other situations, we will adopt the Malliavin calculus formulated by Bichteler *et al.* (1987). Though it is not the most efficient solution to the stochastic differential equation with the purely atomic jump distribution, it still has advantages that it provides IBP-formulas and is relatively easy to handle.

DEFINITION 4.1. (Malliavin operator) Given a probability space $(\Omega, \mathscr{B}, \Pi)$, a linear operator L on $\mathfrak{D}(L) \subset \cap_{p>1} L^p(\Pi)$ into $\cap_{p>1} L^p(\Pi)$ is called a Malliavin operator if the following conditions hold true:

(1) $\mathscr{B}$ is generated by all functions in $\mathfrak{D}(L)$;
(2) If $f \in C^2_\uparrow(\mathbb{R}^d)$ and $F = (F^l)_{l=1}^d$, $F^l \in \mathfrak{D}(L)$, then $f \circ F \in \mathfrak{D}(L)$;
(3) L is self-adjoint in $L^2(\Pi)$, i.e., $E[FLG] = E[GLF]$ for all F, $G \in \mathfrak{D}(L)$;
(4) $L(F^2) \ge 2FLF$ for any $F \in \mathfrak{D}(L)$, i.e., the bilinear operator Γ_L defined by

$$\Gamma_L(F, G) = L(FG) - FLG - GLF$$

is non-negative definite;

(5) If $f \in C_{\uparrow}^2(\mathbb{R}^d)$ and $F = (F^l)_{l=1}^d$, $F^l \in \mathfrak{D}(L)$,

$$L(f \circ F) = \sum_{l=1}^{d} \partial_l f \circ F \cdot LF^l + \frac{1}{2} \sum_{l,m}^{d} \partial_l \partial_m f \circ F \cdot \Gamma_L(F^l, F^m).$$

Let $D_{2,p}^L$, $p \geq 2$, be the completion of $\mathfrak{D}(L)$ with respect to $\|\cdot\|_{D_{2,p}^L}$, where

$$\|F\|_{D_{2,p}^L} = \|F\|_p + \|LF\|_p + \|\Gamma_L^{1/2}(F,F)\|_p,$$

and let $D_{2,\infty-}^L = \cap_{p\geq 2} D_{2,p}^L$. Then L is extended uniquely to an operator L on $D_{2,\infty-}$. For $F \in D_{2,\infty-}^L(\mathbb{R}^d) \equiv (D_{2,\infty-}^L)^d$, the Malliavin covariance σ_F of F is defined by

$$\sigma_F^{lm} = \Gamma_L(F^l, F^m), \qquad l, m = 1, \ldots, d$$

and $\gamma_F = (\gamma_F^{lm})$ denotes the inverse matrix of σ_F. Moreover, the following IBP formulas hold true:

PROPOSITION 4.1. (Theorem 8–18 of Bichteler *et al.* (1987))
(1) *Let* $f \in C_{\uparrow}^2(\mathbb{R}^d)$, $F \in D_{2,\infty-}^L(\mathbb{R}^d)$, *and* G, $\psi \in D_{2,\infty-}^L$. *If* $LG = 0$, $\sigma_F^{lm} \in D_{2,\infty-}^L$ *and* $(\det \sigma_F)^{-1}\psi \in D_{2,\infty-}^L$, *then*

$$E[\partial_l f(F)\psi G] = E[f(F)\mathcal{I}_l^F(\psi)G] \tag{4.1}$$

where

$$\mathcal{I}_l^F(\psi) = -\sum_{m=1}^{d} (2\gamma_F^{lm}\psi LF^m + \Gamma_L(\gamma_F^{lm}\psi, F^m)).$$

(2) *Let* $F \in D_{2,\infty-}^L(\mathbb{R}^d)$, $G \in D_{2,\infty-}^L(\mathbb{R}^n)$, *and* H, $\psi \in D_{2,\infty-}$. *For* $\mathcal{Z} = (F, G)$, *put*

$$\tilde{\sigma}_{\mathcal{Z}}^{lm} = \sigma_F^{lm} - \sum_{l',m'}^{n} \Gamma_L(F^l, G^{l'})\gamma_G^{l'm'}\Gamma_L(G^{m'}, F^m)$$

and $\tilde{\gamma}_{\mathcal{Z}}^{lm}$ *denotes the* (l,m)*-element of the inverse matrix of* $(\tilde{\sigma}_{\mathcal{Z}}^{lm})$. *Suppose that* $LH = 0$, *a.s. Moreover, suppose that* $\sigma_{\mathcal{Z}}^{lm} \in D_{2,\infty-}$, $l, m = 1, \ldots, n$ *and that* $(\det \sigma_{\mathcal{Z}})^{-1}(\det \sigma_G)^{-(d-1)}\psi \in D_{2,\infty-}$. *Then it holds that for* $f \in C_{\uparrow}^2(\mathbb{R}^d)$ *and* $g \in C_{\uparrow}^2(\mathbb{R}^n)$,

$$E[(\partial_l f)(F)\psi g(G)H] = E[f(F)\Psi_{l,\mathcal{Z}}^*(\psi)g(G)H] \tag{4.2}$$

where

$$\Psi_{l,\mathcal{Z}}^*(\psi) = \sum_{m}^{d} \Psi_{m,\mathcal{Z}}(\tilde{\gamma}_{\mathcal{Z}}^{lm}\psi),$$

$$\Psi_{l,\mathcal{Z}}(\psi) = -\Gamma_L(\psi, F^l) - 2\psi LF^l + \sum_{l',m'}^{n} \{\Gamma_L(\gamma_G^{l',m'}\Gamma_L(F^l, G^{l'})\psi, G^{m'}) + 2\gamma_G^{l',m'}\Gamma_L(F^l, G^{l'})\psi LG^{m'}\}.$$

Example. (1) (*diffusion process*) As in Subsection 3.1, for the gradient operator D and the divergence operator defined over the Wiener space, let

$$L = -\frac{1}{2}D^*D.$$

It is easy to show that L is a Malliavin operator over the Wiener space. In this case, the operator L is a kind of second-order differential operator, and it is called the Ornstein-Uhlenbeck operator. It is a symmetric operator compatible with the closed extension, and it is possible to construct the Malliavin calculus starting with L. As we saw in the previous section, the IBP formula (4.2) or (3.5) was applied to verify the conditional type Cramér condition [CD].

(2) (*non-linear moving-average series*) Let $p \in C^\infty(\mathbb{R}^r; \mathbb{R}_+)$ be a positive density and $\boldsymbol{p}(u) = \prod_{j=1}^n p(u_j)$, where $u = (u_1, \dots, u_n)$, $u_j \in \mathbb{R}^r$, $j = 1, \dots, n$. Then $L = -D^*D/2$ is a Malliavin operator, where D and D^* are the gradient and the divergence operator defined in Subsection 3.2. The equation (3.8) is the IBP-formula (4.1) with $\Gamma_L(A, B) = \langle DA, DB\rangle_{\mathbb{R}^{rN}}$, and it was also applied to prove the conditional type Cramér condition (3.7).

5. Stochastic expansion type of functional

In this section, we will present some formulas for the k-th order asymptotic expansion of S_T defined by (1.3) with a functional $\bar{\mathsf{Z}}_T = (\bar{\mathsf{Z}}_T^{(0)}, \bar{\mathsf{Z}}_T^{(1)})$, and will obtain the third order asymptotic expansion in terms of some coefficients in a representation of Q_i by Hermite polynomials. It will be applied to M-estimators for a general statistical model in the next section.

Let $(\Omega, \mathcal{F}, P)$ be a probability space, $Y = (Y_t)_{t\in\mathbb{R}_+}$ an $\mathbb{R}^d$-valued ϵ-Markov process driven by an $\mathbb{R}^r$-valued process X, and Z an $\mathbb{R}^n$-valued additive functional of X and Y as in Subsection 2.5. Moreover, let $\mathsf{C} = (\mathsf{C}_t)_{t\in\mathbb{R}_+}$ be an $\mathbb{R}^n \otimes \mathbb{R}^n$-valued deterministic process such that each element is bounded as $t \to \infty$ and that it converges to non-singular matrix as $t \to \infty$. For these processes, assume that the components $\bar{\mathsf{Z}}_T^{(0)}$ and $\bar{\mathsf{Z}}_T^{(1)}$ of S_T is defined by $\bar{\mathsf{Z}}_T = (\bar{\mathsf{Z}}_T^{(0)}, \bar{\mathsf{Z}}_T^{(1)}) = T^{-1/2}\mathsf{C}_T Z_T$, where $\dim \bar{\mathsf{Z}}_T^{(0)} = p$, and $\dim \bar{\mathsf{Z}}_T^{(1)} = q$. Note that in the stochastic expansions (1.1) and (1.2), C is $H'(\mu)$ and g^{-1}, respectively.

Before going into the asymptotic expansion of S_T, we prepare assumptions about the validity of that of $T^{-1/2}Z_T$. As usual in the asymptotic expansion literature, we adopt a mixing condition:

[A1] There exists a positive constant a such that

$$\|E[f \mid \mathcal{B}^Y_{[s-\epsilon,s]}] - E[f]\|_{\mathbb{L}^1(P)} \le a^{-1}e^{-a(t-s)}\|f\|_\infty$$

for any $s, t \in \mathbb{R}_+$, $s \le t$, and for any bounded $\mathcal{B}^Y_{[t,\infty)}$-measurable function f.

It should be noted that the exponential order of the strong mixing is not necessary: in fact, it is possible to reduce it to a polynomial order (see Lahiri (1993), and Yoshida (2001)). However, the above condition will be assumed for simplicity. The moment condition is also assumed:

[A2] For any $\Delta > 0$, $\sup_{t\in\mathbb{R}_+, 0\le h\le\Delta} \|Z^t_{t+h}\|_{\mathbb{L}^p(P)} < \infty$ for any $p > 1$, and $E[Z^t_{t+\Delta}] = 0$. Moreover, $Z_0 \in \cap_{p>1}\mathbb{L}^p(P)$ and $E[Z_0] = 0$.

In order to estimate the characteristic function of Z_T, we need an IBP formula over sub-intervals of $[0,T]$ as in Subsections 3.1 and 3.2. Let τ be a fixed constant such that $\tau > \epsilon$, where ϵ comes from ϵ-Markov process Y. For each $T > 0$, let $[u(i), v(j)]$, $j = 1, \dots, n(T)$ be sub-intervals of the interval $[0,T]$ such that

$$0 < \epsilon \le u(1) < v(1) \le u(2) < v(2) \le \cdots \le u(n(T)) < v(n(T)) \le T$$

and that $\inf_{j,T}\{v(j) - u(j)\} \ge \tau$, $\sup_{j,T}\{v(j) - u(j)\} < \infty$. Assume that for each interval $J_j = [v(j) - \epsilon, v(j)]$, there exists a finite number of functionals $\mathcal{Y}_j = \{\mathcal{Y}_{j,k}\}_{k=1,\dots,M_j}$ such that $\sigma[\mathcal{Y}_j] \subset \mathcal{B}_{J_j}$ and that for any bounded $\mathcal{B}_{[v(j),\infty)}$-measurable function F, $E[F \mid \mathcal{B}_{[0,v(j)]}] = E[F \mid \sigma[\mathcal{Y}_j]]$, a.s. Then from the Markovian property of Y, the estimate of the characteristic function of Z_T is reduced to that of $E[e^{iu\cdot Z^{u(j)}_{v(j)}} \mid \mathcal{B}_{I_j} \vee \mathcal{B}'_{J_j}]$, where $Z^{u(j)}_{v(j)} = Z_{v(j)} - Z_{u(j)}$, $I_j = [u(j) - \epsilon, u(j)]$, $\mathcal{B}'_{J_j} := \sigma[\mathcal{Y}_j]$, and $\mathcal{B}_I$ is the sub-σ-field defined in Subsection 2.5. ψ_j denotes a truncation functional. Let $C_j = ((Y_t, X_t - X_{u(j)-\epsilon}; t \in I_j), \mathcal{Y}_j)$.

Assume that for each $j = 1, \dots, n(T)$, there exist a probability space $(\Omega_j, \mathcal{F}_j, \boldsymbol{P}_j)$ and a random vector $(\hat{\psi}_j, \hat{Z}^{u(j)}_{v(j)}, \hat{C}_j)$ satisfying

$$\mathcal{L}\{(\hat{\psi}_j, \hat{Z}^{u(j)}_{v(j)}, \hat{C}_j) \mid \boldsymbol{P}_j\} = \mathcal{L}\{(\psi_j, Z^{u(j)}_{v(j)}, C_j) \mid P\}.$$

The random vector $(\hat{\psi}_j, \hat{Z}^{u(j)}_{v(j)}, \hat{C}_j)$ is called a *distributional equivalent* of $(\psi_j, Z^{u(j)}_{v(j)}, C_j)$. (The present formulation by the distributional equivalents, which is originally given in Yoshida (2001), is essentially the same as the one in Kusuoka and Yoshida (2000)). In the sequel, we will identify the distributional equivalents with their originals, and express them with the same notation (without "hats"), however, we should realize that the operations in the Malliavin calculus are done for distributional equivalents over $(\Omega_j, \mathcal{F}_j, \boldsymbol{P}_j)$.

We also assume that a Malliavin operator L_j is given over the probability space $(\Omega_j, \mathcal{F}_j, \boldsymbol{P}_j)$. Denote by $D^{L_j}_{2,p}$ the Banach space induced by L_j, and put $D_{2,\infty-} = \cap_{p\ge 2} D_{2,p}$. In addition, suppose that for any $f \in C^\infty_b(\mathbb{R}^{(r+d)m})$ and any $u_0, u_1, \dots, u_m$ satisfying $u(j) - \epsilon \le u_0 \le u_1 \le \cdots \le u_m \le u(j)$, the functional $F = f(X_{u_k} - X_{u_{k-1}}, Y_{u_k} : 1 \le k \le m) \in D^{L_j}_{2,\infty-}$ and $L_j F = 0$. This assumption ensures that L_j does not act on any $\mathcal{B}_{I_j}$-measurable functionals. The Malliavin covariance σ_F of $F \in D^{L_j}_{2,\infty-}(\mathbb{R}^{d'}) \equiv (D^{L_j}_{2,\infty-})^{d'}$ is defined by $\sigma_F = (\sigma^{i,k}_F) = (\Gamma_{L_j}(F^i, F^k))$, and the determinant of σ_F is denoted by Δ_F. Let $\mathcal{Z}_j = (Z^{u(j)}_{v(j)}, \mathcal{Y}_j)$, $S^*_1[\psi_j; \mathcal{Z}_j] = \{\sigma^{i,k}_{\mathcal{Z}_j}, i,k = 1, \dots, n + M_j; (\Delta_{\mathcal{Z}_j})^{-1}(\Delta_{\mathcal{Y}_j})^{-(n-1)}\psi_j\}$, and

$$S_{1,j} = \{(\Delta_{\mathcal{Z}_j})^{-1}(\Delta_{\mathcal{Y}_j})^{-(n-1)}\psi_j, \sigma^{kl}_{\mathcal{Z}_j}, L_j\mathcal{Z}_{j,k}, \Gamma_{L_j}(\sigma^{kl}_{\mathcal{Z}_j}, \mathcal{Z}_{j,m}),$$
$$\Gamma_{L_j}((\Delta_{\mathcal{Z}_j})^{-1}(\Delta_{\mathcal{Y}_j})^{-(n-1)}\psi_j, \mathcal{Z}_{j,l})\}.$$

Assume that $\sup_{j,T} M_j < \infty$. By using the terminology and notation above, we consider the following conditions of non-degeneracy:

[A3] (i) $\inf_{j,T} E[\psi_j] > 0$;

(ii) $\liminf_{T\to\infty} n(T)/T > 0$;

(iii) For each $j = 1, \dots, n(T)$, $\mathcal{Z}_j \in (D^{L_j}_{2,\infty-})^{n+M_j}$, $S^*_1[\psi_j; \mathcal{Z}_j] \subset D^{L_j}_{2,\infty-}$, and for any $p' > 1$, $\bigcup_{j=1,\dots,n(T),T>0} S_{1,j}$ is bounded in $\mathbb{L}^{p'}(P)$.

Condition [A3] may seem complicated for some readers. However, it is rather standard and easy to verify. For example, the non-degeneracy of the Lie algebra spanned by the coefficient vector fields, i.e., Hörmander type condition, is sufficient for diffusion processes; see Kusuoka and Yoshida (2000) and Condition [L] in Section 6 of this article. The boundedness of $S_{1,j}$ easily follows from usual regularity conditions for coefficients appearing in the stochastic differential equations: a questioning reader will find that it is a routine procedure if he/she consults a textbook by Bichteler *et al.* (1987). Moreover, with the help of support theorems, it is possible to verify [A3] under much weaker conditions (Yoshida (2001)). It is also the case for jump processes.

In this article, we prefer [A3] in the present form because more precise description would increase the volume of this article too large. However, we emphasizes that Condition [A3] is already very concrete, and indeed, several authors checked this condition for various models to derive expansions.

The conditions $\mathcal{Z}_j \in (D^{L_j}_{2,\infty-})^{n+M_j}$ and $S^*_1[\psi_j; \mathcal{Z}_j] \subset D^{L_j}_{2,\infty-}$ in (iii) of [A3] ensure that the IBP formula (4.2) hold true for the Malliavin operator L_j and the truncation functional ψ_i. The condition that $\cup_{j=1,\dots,n(T),T>0} S_{1,j}$ is bounded in $\mathbb{L}^{p'}(P)$ ensures that $\Psi^*_{l,\mathcal{Z}}(\psi)$ is bounded in $\mathbb{L}^p(P)$. This assumption [A3] is suitable to the case where the underlying processes X and Y have the Markovian properties. However, there are some cases where the simpler IBP formula (4.1) can be applied as the moving-average process in Subsection 3.2. For such cases, we can obtain the same results in this section under the condition in Theorem 1 of Kusuoka and Yoshida (2000). Our main application in this article is the diffusion process, therefore we here adopt this setting.

In order to obtain the asymptotic expansion of S_T, let us prepare some notations. Denote the covariance matrix $\mathrm{Cov}(\bar{\mathsf{Z}}_T)$, $\mathrm{Cov}(\bar{\mathsf{Z}}^{(0)}_T)$, and $\mathrm{Cov}(\bar{\mathsf{Z}}^{(1)}_T)$ by $\bar{g} = (\bar{g}^{\alpha\beta})_{\alpha,\beta=1,\dots,n}$, $g = (g^{ab})_{a,b=1,\dots,p}$, and $\tilde{g} = (\tilde{g}^{\kappa\mu})_{\kappa,\mu=p+1,\dots,p+q}$, respectively. Assume that Cov $(T^{-1/2}Z_T)$ converges to a positive definite matrix; hence for sufficiently large T, the matrix $\bar{g}$ is a positive definite matrix because C_T converges to a non-singular matrix. Note that $\bar{g}$, g, and $\tilde{g}$ may depend on T. Define the j-th cumulant $\lambda^{\alpha_1\cdots\alpha_j}$ of $\bar{\mathsf{Z}}_T$ by

$$\lambda^{\alpha_1\cdots\alpha_j} = i^{-j}\partial^{\alpha_1}\cdots\partial^{\alpha_j}\log E[e^{iu\cdot\bar{\mathsf{Z}}_T}]\,|_{u=0}, \qquad \partial^\alpha = \frac{\partial}{\partial u_\alpha}.$$

Under the assumption [A2], $\lambda^a = 0$, $a = 1,\dots,n$. For any positive definite matrix $\sigma = (\sigma^{\alpha\beta})$, the Hermite polynomials $h_{\alpha_1\cdots\alpha_j}$ and their contravariant representation $h^{\alpha_1\cdots\alpha_j}$ are defined by

$$h_{\alpha_1\cdots\alpha_j}(z;\sigma) = \frac{(-1)^j}{\phi(z;\sigma)}\partial_{\alpha_1}\cdots\partial_{\alpha_j}\phi(z;\sigma), \qquad \partial_\alpha = \frac{\partial}{\partial z^\alpha},$$

and

$$h^{\alpha_1\cdots\alpha_j}(z;\sigma) = \sigma^{\alpha_1\beta_1}\cdots\sigma^{\alpha_j\beta_j}h_{\beta_1\cdots\beta_j}(z;\sigma),$$

respectively, where $\phi(z;\sigma)$ is the density function of the normal distribution with mean 0 and covariance matrix σ. For $M > 0$ and $\gamma > 0$, the set $\mathcal{E}(M,\gamma)$ of measurable functions from $\mathbb{R}^p \to \mathbb{R}$ is defined by

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

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

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

It is easy to show that the formal asymptotic expansion of the density of $\bar{Z}_T$ is given by

$$p_{T,k}(z) = \sum_{j=0}^{k} T^{-j/2}\Xi_{T,j}(z)\phi(z;\bar{g}), \tag{5.1}$$

where $A_m = \alpha_1 \cdots \alpha_m$, $\bar{\lambda}^{A_m} = T^{(m-2)/2}\lambda^{A_m}$, $\Xi_{T,0}(x) = 1$, and for $j \ge 1$,

$$\Xi_{T,j}(z) = \sum_{m=1}^{j} \frac{1}{m!} \sum_{\substack{k_1+\cdots+k_m=j \\ k_1\ge 1,\dots,k_m\ge 1}} \frac{\bar{\lambda}^{A_{k_1+2}} \dots \bar{\lambda}^{A_{k_m+2}}}{(k_1+2)!\cdots(k_m+2)!} h_{A_{k_1+2}\cdots A_{k_m+2}}(z;\bar{g}).$$

See Bhattacharya and Rao (1986) and Sakamoto (1998). In particular,

$$\begin{aligned}
\Xi_{T,1}(z) &= \frac{1}{6}\bar{\lambda}^{\alpha\beta\gamma}h_{\alpha\beta\gamma}(z;\bar{g}),\\
\Xi_{T,2}(z) &= \frac{1}{24}\bar{\lambda}^{\alpha\beta\gamma\delta}h_{\alpha\beta\gamma\delta}(z;\bar{g}) + \frac{1}{72}\bar{\lambda}^{\alpha\beta\gamma}\bar{\lambda}^{\delta\epsilon\sigma}h_{\alpha\beta\gamma\delta\epsilon\sigma}(z;\bar{g}),\\
\Xi_{T,3}(z) &= \frac{1}{120}\bar{\lambda}^{\alpha\beta\gamma\delta\epsilon}h_{\alpha\beta\gamma\delta\epsilon}(z;\bar{g}) + \frac{1}{144}\bar{\lambda}^{\alpha_1\alpha_2\alpha_3}\bar{\lambda}^{\beta_1\beta_2\beta_3\beta_4}h_{\alpha_1\alpha_2\alpha_3\beta_1\beta_2\beta_3\beta_4}(z;\bar{g})\\
&\quad + \frac{1}{1296}\bar{\lambda}^{\alpha_1\alpha_2\alpha_3}\bar{\lambda}^{\beta_1\beta_2\beta_3}\bar{\lambda}^{\gamma_1\gamma_2\gamma_3}h_{\alpha_1\alpha_2\alpha_3\beta_1\beta_2\beta_3\gamma_1\gamma_2\gamma_3}(z;\bar{g}).
\end{aligned}$$

In the above expansion, we adopt the Einstein summation convention, and $\alpha,\beta,\dots,\sigma$ are indices running from 1 to $n = p+q$. In the sequel, we will often use the convention, under which the Greek characters $\alpha,\beta,\dots$ are indices running from 1 to n, the Roman $a,b,\dots$ running from 1 to p, and the lower ordered Greek $\kappa,\lambda,\dots$ running from $p+1$ to n.

Under Conditions [A1], [A2], [A3], the k-th order asymptotic expansion of $Z_T/\sqrt{T}$ was derived by Kusuoka and Yoshida (2000), and they showed in Theorem 5 that the asymptotic expansion of S_T can be derived from $p_{T,k}$ and that it is a expansion of the Edgeworth-type. We here present an *explicit* formula for the k-th order asymptotic expansion of S_T by using the Taylor expansion of $f(S_T)$ around $S_T = Z_T^{(0)}$.

THEOREM 5.1. *Let M, γ, K be positive numbers and let $\hat{g}$ be a positive definite matrix satisfying $\hat{g} > \lim_{T\to\infty} g$. Suppose that Conditions* [A1], [A2], [A3] *hold true. Then for any $k \in \mathbb{N}$, there exist constants $\delta > 0$ and $c > 0$ such that for any $f \in \mathcal{E}(M,\gamma)$,*

$$\left|E[f(S_T)] - \int_{\mathbb{R}^p} dy^{(0)} f(y^{(0)}) q_{T,k}(y^{(0)})\right| \le c\omega(f, T^{-K}, \hat{g}) + \epsilon_T, \tag{5.2}$$

where $y = (y^{(0)}, y^{(1)})$, $\dim(y^{(0)}) = p$, $\dim(y^{(1)}) = q$, Q_i^a *denote the a-th element of* Q_i,

$$q_{T,k}(y^{(0)}) = \int_{\mathbb{R}^q} \phi(y;\bar{g})dy^{(1)} + \sum_{m=1}^{k} T^{-m/2}\Bigg(\int_{\mathbb{R}^q} \Xi_{T,m}(y)\phi(y;\bar{g})dy^{(1)}$$
$$+ \sum_{\substack{s+l=m\\ s\geq 0, l\geq 1}} \sum_{j=1}^{l} \frac{(-1)^j}{j!} \sum_{l_1+\cdots+l_j=l} \partial_{a_1}\cdots\partial_{a_j}\Big(\int_{\mathbb{R}^q} Q_{l_1}^{a_1}(y)\cdots Q_{l_j}^{a_j}(y)\Xi_{T,s}(y)\phi(y;\bar{g})dy^{(1)}\Big)\Bigg)$$

and $\epsilon_T = o(T^{-((k+\delta)/2\wedge K)})$.

Remark 1. In some cases, each of (Q_i) may possibly be a polynomial only in $\bar{Z}_T^{(0)}$ (independent of $\bar{Z}_T^{(1)}$). In this case, the expansion $q_{T,k}$ becomes a simple one which is determined by the Edgeworth expansion of $\bar{Z}_T^{(0)}$ and (Q_i), i.e.,

$$q_{T,k}(y^{(0)}) = \phi(y;g) + \sum_{m=1}^{k} T^{-m/2}\Bigg(\Xi_{T,m}^{(0)}(y^{(0)})\phi(y^{(0)};g)$$
$$+ \sum_{\substack{s+l=m\\ s\geq 0, l\geq 1}} \sum_{j=1}^{l} \frac{(-1)^j}{j!} \sum_{l_1+\cdots+l_j=l} \partial_{a_1}\cdots\partial_{a_j}\Big(Q_{l_1}^{a_1}(y^{(0)})\cdots Q_{l_j}^{a_j}(y^{(0)})\Xi_{T,s}^{(0)}(y^{(0)})\phi(y^{(0)};g)\Big)\Bigg),$$

where $(\Xi_{T,m}^{(0)})$ are polynomials appearing in the Edgeworth expansion of $\bar{Z}_T^{(0)}$, and they are defined by the same formula as that for $\Xi_{T,m}$. It is easy to show that the same result as Theorem 5.1 holds true with this expansion under Conditions [A1], [A2] and [A3] for $\bar{Z}_T^{(0)}$, while Theorem 5.1 suppose the conditions for $\bar{Z}_T = (\bar{Z}_T^{(0)}, \bar{Z}_T^{(1)})$.

When we need a more explicit representation of an asymptotic expansion for a statistic of interest, we only have to make the calculations, differentiations and integrations, in this expansion for the polynomials Q_i corresponding to the statistic. Those calculations are absolutely elementary but rather complicated at least for higher order terms, hence we prepare another representation of $q_{T,k}$, which may require slightly less calculations, in an ordinary way.

Let $\widetilde{Z}_T^{(0)} = \bar{Z}_T^{(0)}$ and $\widetilde{Z}_T^{(1)} = \bar{Z}_T^{(1)} - \bar{\Sigma}_{21}\bar{\Sigma}_{11}^{-1}\bar{Z}_T^{(0)}$, where $\bar{\Sigma}_{21} = \mathrm{Cov}(\bar{Z}_T^{(1)}, \bar{Z}_T^{(0)})$ and $\bar{\Sigma}_{11} = \mathrm{Cov}(\bar{Z}_T^{(0)})(= g)$. Then the covariance matrix $\widetilde{\Sigma}$ of $\widetilde{Z}_T = (\widetilde{Z}_T^{(0)}, \widetilde{Z}_T^{(1)})$ is given by

$$\widetilde{\Sigma} = \begin{bmatrix} \bar{\Sigma}_{11} & 0 \\ 0 & \bar{\Sigma}_{22,1} \end{bmatrix},$$

where $\bar{\Sigma}_{22,1} = \bar{\Sigma}_{22} - \bar{\Sigma}_{21}\bar{\Sigma}_{11}^{-1}\bar{\Sigma}_{12}$ and $\bar{\Sigma}_{22} = \mathrm{Cov}(\bar{Z}_T^{(1)})(= \bar{g})$. In terms of $\widetilde{Z}_T$, the functional S_T is rewritten as

$$S_T = \widetilde{Z}_T^{(0)} + \sum_{i=1}^{k} \frac{1}{T^{i/2}}\tilde{Q}(\widetilde{Z}_T^{(0)}, \widetilde{Z}_T^{(1)}), \tag{5.3}$$

where $\tilde{Q}$ is a polynomial satisfying $\tilde{Q}(z^{(0)}, z^{(1)}) = Q(z^{(0)}, z^{(1)} + \bar{\Sigma}_{21}\bar{\Sigma}_{11}^{-1}z^{(0)})$. For any index set $A = \{\alpha_1, \ldots, \alpha_m\}$, denote the cumulant of $\widetilde{Z}_T^{\alpha_1}, \ldots, \widetilde{Z}_T^{\alpha_m}$ by $\tilde{\lambda}^A$. Let $\tilde{\Xi}_{T,j}$ be a function taking the same form as $\Xi_{T,j}$ with $\widetilde{\Sigma}$ and $\tilde{\lambda}^A$ in place of $\bar{g}$ and $\bar{\lambda}^A$.

COROLLARY 5.1. *Let M, γ, $K > 0$ and $\hat{g}$ be a positive definite matrix satisfying $\hat{g} > \lim_{T\to\infty} g$. Suppose that Conditions* [A1], [A2], [A3] *hold true. Then for any $k \in \mathbb{N}$, there exist constants $\delta > 0$ and $c > 0$ such that for any $f \in \mathcal{E}(M, \gamma)$,*

$$\left|E[f(S_T)] - \int_{\mathbb{R}^p} dy^{(0)} f(y^{(0)}) q_{T,k}(y^{(0)})\right| \le c\omega(f, T^{-K}, \hat{g}) + \epsilon_T, \tag{5.4}$$

where $y = (y^{(0)}, y^{(1)})$, $\dim(y^{(0)}) = p$, $\dim(y^{(1)}) = q$, $\tilde{Q}_i^a$ denote the a-th element of $\tilde{Q}_i$,

$$\begin{aligned} q_{T,k}(y^{(0)}) = &\int_{\mathbb{R}^q} \phi(y; \widetilde{\Sigma}) dy^{(1)} + \sum_{m=1}^{k} T^{-m/2} \Bigg(\int_{\mathbb{R}^q} \widetilde{\Xi}_{T,m}(y) \phi(y; \widetilde{\Sigma}) dy^{(1)} \\ &+ \sum_{\substack{s+l=m \\ s\ge 0, l\ge 1}} \sum_{j=1}^{l} \frac{(-1)^j}{j!} \sum_{l_1+\cdots+l_j=l} \partial_{a_1} \cdots \partial_{a_j} \left(\int_{\mathbb{R}^q} \tilde{Q}_{l_1}^{a_1}(y) \cdots \tilde{Q}_{l_j}^{a_j}(y) \widetilde{\Xi}_{T,s}(y) \phi(y; \widetilde{\Sigma}) dy^{(1)} \right) \Bigg) \end{aligned}$$

and $\epsilon_T = o(T^{-((k+\delta)/2 \wedge K)})$.

In Theorem 5.1 and Corollary 5.1, we assumed the non-degeneracy of the covariance matrix of $T^{-1/2} Z_T$; however, it is necessary to generalize the results to the case where $\mathrm{Cov}(T^{-1/2} Z_T^{(0)})$ is regular but there is a linear relation between the elements of $Z_T^{(1)}$, and where $\mathrm{Cov}(T^{-1/2} Z_T)$ is degenerate. The maximum likelihood estimator as an M-estimator is the case as we will discuss it later. More precisely, in the case of the MLE or the minimum contrast estimator, $Z_T^{(1)}$ consists of the second and the higher order derivatives of the log-likelihood function, and therefore the elements of $Z_T^{(1)}$ have a linear relation due to the exchangeability of the differential operator. If $\bar{g}$ is degenerate, the Hermit polynomials $h_{\alpha_1 \cdots \alpha_k}(z; \bar{g})$ does not make sense as it is. However, it is still possible to interpret each $p_{T,k}(z)$ as a Schwartz distribution, and to prove the validity of the formula for $q_{T,2}$ given in Theorem 5.1.

For $z = (z^{(0)}, z^{(1)}) \in \mathbb{R}^{p+q}$, put $\mathring{z} := z^{(0)}$, $(\dot{z}, \ddot{z}) := z^{(1)}$, and $z^* := (\mathring{z}, \dot{z})$, where $\dot{z} \in \mathbb{R}^{q_1}$, $\ddot{z} \in \mathbb{R}^{q_2}$ and $q_1 + q_2 = q$. Since if $q_2 = 0$, then the following arguments result in the preceding case, we will suppose $q_2 > 0$. By using these notations, we will consider the case where the following conditions hold:

[H0] C_T is a block-diagonal matrix given by

$$\mathsf{C}_T = \begin{bmatrix} \mathsf{C}_T^{(0)} & 0 \\ 0 & \mathsf{C}_T^{(1)} \end{bmatrix},$$

where $\mathsf{C}_T^{(0)}$ is a $p \times p$ matrix converging to a non-singular matrix and $\mathsf{C}_T^{(1)}$ is a $q \times q$ matrix which may be singular;

[H1] $\mathrm{Cov}(T^{-1/2} Z_T^*)$ converges to a positive definite matrix;

[H2] for some matrix $L = (L_\beta^\alpha)_{p+q_1+1 \le \alpha \le p+q, p+1 \le \beta \le p+q_1}$, $\ddot{Z}_T = L \dot{Z}_T$ a.s.

As we will see in the next section, the coefficient matrix C_T for the minimum contrast estimator has the structure given by [H0] and [H2].

For any $(p + q_1) \times (p + q_1)$ positive matrix σ^* and any $q \times q_1$ matrix $\tilde{M}$, let

$$\bar{\sigma} = \begin{bmatrix} \sigma_{11}^* & \sigma_{12}^* \tilde{M}' \\ \tilde{M} \sigma_{21}^* & \tilde{M} \sigma_{22}^* \tilde{M}' \end{bmatrix},$$

where σ_{11}^*, σ_{12}^*, σ_{21}^* and σ_{22}^* are $p \times p$, $p \times q_1$, $q_1 \times p$, and $q_1 \times q_1$ matrices, respectively, giving the decomposition of σ^*:

$$\sigma^* = \begin{bmatrix} \sigma_{11}^* & \sigma_{12}^* \\ \sigma_{21}^* & \sigma_{22}^* \end{bmatrix}.$$

For any index set $A = \alpha_1 \cdots \alpha_m$ ($\alpha_1, \ldots, \alpha_m \in \{1, \ldots, p+q\}$) and any $z^{(0)} \in \mathbb{R}^p$, we define a Schwartz distribution "$h_A(z;\bar{\sigma})\phi(z;\bar{\sigma})$" so that for any $f \in \mathcal{S}(\mathbb{R}^q)$,

$$\text{(5.5)} \qquad \int_{\mathbb{R}^q} f(z^{(1)}) h_A(z;\bar{\sigma})\phi(z;\bar{\sigma}) dz^{(1)} \\ \stackrel{\text{def}}{=} \int_{\mathbb{R}^{q_1}} (\partial_{A^{(1)}} f)(\tilde{M}\dot{z})(-1)^{|A^{(0)}|}\partial_{A^{(0)}}\phi(z^*;\sigma^*) d\dot{z},$$

where $A^{(0)}$ and $A^{(1)}$ be parts of a decomposition of A into two index sets such that $A^{(0)}$ and $A^{(1)}$ consist of indices running from 1 to p and from $p+1$ to $p+q$, respectively. We can easily extend this definition to the case where $f \in C_\uparrow^\infty(\mathbb{R}^q)$, i.e., we also define "$h_A(z;\bar{\sigma})\phi(z;\bar{\sigma})$" by (5.5) for any $f \in C_\uparrow^\infty(\mathbb{R}^q)$. If $|A| = 0$,

$$\text{(5.6)} \qquad \int_{\mathbb{R}^q} f(z^{(1)})\phi(z;\bar{\sigma}) dz^{(1)} \stackrel{\text{def}}{=} \int_{\mathbb{R}^{q_1}} f(\tilde{M}\dot{z})\phi(z^*;\sigma^*) d\dot{z}.$$

Under Conditions [H0], [H1] and [H2], we see that $\bar{g}$ and $\widetilde{\Sigma}$ have the same structure as that of $\bar{\sigma}$ above, regarding the covariance matrices of $(\bar{Z}_T^{(0)}, T^{-1/2}\dot{Z}_T)$ and $(\bar{Z}_T^{(0)}, T^{-1/2}\dot{Z}_T - \mathrm{Cov}(T^{-1/2}\dot{Z}_T, \bar{Z}^{(0)})\bar{\Sigma}_{11}^{-1}\bar{Z}^{(0)})$ as σ^*, respectively. Therefore, we can interpret $q_{T,k}$ of Theorem 5.1 and Corollary 5.1 as (5.5) and (5.6), and can extend Theorem 5.1 and Corollary 5.1 as follows.

THEOREM 5.2. *Assume that* (i) *Conditions* [A1], [A2] *and* [A3] *for* Z_T^* *hold and* (ii) [H0], [H1] *and* [H2] *are satisfied. Then the inequalities in Theorem* 5.1 *and Corollary* 5.1 *hold with the same* $q_{T,k}$ *interpreted as* (5.5) *and* (5.6).

In Corollary 5.1, the orthogonality between $\widetilde{Z}_T^{(0)}$ and $\widetilde{Z}_T^{(1)}$ reduces the expansion $q_{T,k}$ to a simpler form. For any index set K whose elements run from $p+1$ to $p+q$, any positive definite matrix σ_{22}^* and any $q \times q_1$ matrix $\tilde{M}$, let $h_K(z^{(1)}; \tilde{M}\sigma_{22}^*\tilde{M}')\phi(z^{(1)}; \tilde{M}\sigma_{22}^*\tilde{M}')$ be a Schwartz distribution defined by

$$\text{(5.7)} \qquad \int_{\mathbb{R}^q} f(z^{(1)}) h_K(z^{(1)}; \tilde{M}\sigma_{22}^*\tilde{M}')\phi(z^{(1)}; \tilde{M}\sigma_{22}^*\tilde{M}') dz^{(1)} \\ = \int_{\mathbb{R}^{q_1}} (\partial_K f)(\tilde{M}\dot{z})\phi(\dot{z};\sigma_{22}^*) d\dot{z}, \qquad \forall f \in \mathcal{S}(\mathbb{R}^q).$$

Then, if $\sigma_{12}^* = \sigma_{21}^{*\prime} = 0$, the Schwartz distribution $h_A(z;\bar{\sigma})\phi(z;\bar{\sigma})$ defined above is reduced as

$$h_A(z;\bar{\sigma})\phi(z;\bar{\sigma}) = h_{A^{(0)}}(z^{(0)};\sigma_{11}^*)\phi(z^{(0)};\sigma_{11}^*) \\ \times h_{A^{(1)}}(z^{(1)}; \tilde{M}\sigma_{22}^*\tilde{M}')\phi(z^{(1)}; \tilde{M}\sigma_{22}^*\tilde{M}').$$

For any $q \times q$ positive definite matrix σ and any index sets K, L consisting indices running from $p+1$ to $p+q$,

$$\int h_K(y^{(1)};\sigma)h^L(y^{(1)};\sigma)\phi(y^{(1)};\sigma)dy^{(1)} = \begin{cases} (K)! & \text{if} \quad (K) = (L) \\ 0 & \text{otherwise,} \end{cases}$$

where $(K) \in \mathbb{Z}_+^q$ is the multi index corresponding to K, and

$$(K)! = \prod_{j=1}^{q} \kappa_j! \quad \text{for} \quad (K) = (\kappa_1, \dots, \kappa_q).$$

For polynomials $\tilde{Q}_l^a(y)$ in (5.3), define functions $\pi_{l,K}^a(y^{(0)})$ and $\pi_l^{a,K}(y^{(0)})$ by

$$\pi_{l,K}^a(y^{(0)}) = \frac{1}{(K)!}\int_{\mathbb{R}^q} \tilde{Q}_l^a(y)h_K(y^{(1)};\bar{\Sigma}_{22,1})\phi(y^{(1)};\bar{\Sigma}_{22,1})dy^{(1)}, \tag{5.8}$$

$$\pi_l^{a,K}(y^{(0)}) = \bar{\Sigma}_{22,1}^{K,L}\pi_{l,L}^a(y^{(0)}) \tag{5.9}$$

where $K \in \{\phi, \kappa_1, \kappa_1\kappa_2, \dots\}$, $\kappa_1, \kappa_2, \dots \in \{p+1, \dots, p+q\}$, $L \in \{\phi, \mu_1, \mu_1\mu_2, \dots\}$, $\mu_1, \mu_2, \dots \in \{p+1, \dots, p+q\}$, and $\bar{\Sigma}_{22,1}^{K,L} = \bar{\Sigma}_{22,1}^{\kappa_1\mu_1} \cdots \bar{\Sigma}_{22,1}^{\kappa_m\mu_m}$, $\bar{\Sigma}_{22,1} = (\bar{\Sigma}_{22,1}^{\kappa\mu})$. Note that $\pi_{l,K}^a(y^{(0)})$ and $\pi_l^{a,K}(y^{(0)})$ are polynomial in $y^{(0)}$, which is easily shown from (5.7) and (5.8). Moreover, we will use the contravariant and covariant representation of $\pi_{l,K}^a$ and $\pi_l^{a,K}$;

$$\pi_{l,K}^a(y^{(0)}) = \sum_A q_{l,K}^{a,A}h_A(y^{(0)};g) = \sum_A q_{l,K,A}^a h^A(y^{(0)};g) \tag{5.10}$$

and

$$\pi_l^{a,K}(y^{(0)}) = \sum_A q_l^{a,K,A}h_A(y^{(0)};g) = \sum_A q_{l,A}^{a,K} h^A(y^{(0)};g) \tag{5.11}$$

where $A \in \{\phi, a_1, a_1a_2, \dots\}$, $a_1, a_2, \dots \in \{1, \dots, p\}$, $h^A(y^{(0)};g)$ is the contravariant representation of $h_A(y^{(0)};g)$, i.e., $h^A(y^{(0)};g) = g^{a_1b_1}\cdots g^{a_mb_m}\, h_{b_1\cdots b_m}(y^{(0)};g)$, $A = a_1\cdots a_m$, and $q_k^{a,K,A}$ and $q_{k,A}^{a,K}$ are supposed to be symmetric in indices A. The multiplication of the polynomial becomes an action on the algebra of the polynomials, and we express this action by

$$h_Ah_B = \sum_C C_{A,B}^C h_C.$$

Obviously, $C_{A,\phi}^C = C_{\phi,A}^C = \delta_A^C \ (= \delta_{a_1}^{c_1}\cdots\delta_{a_k}^{c_k})$.

By using these constants $q_{l,K}^{a,A}$, $q_{l,K,A}^a$, $q_l^{a,K,A}$, $q_{l,A}^{a,K}$, and $C_{A,B}^C$, we can reduce $q_{T,k}$ to a simpler form. In the following sections, we will consider the third order asymptotic expansions of statistics, therefore we here present a reduced form of $q_{T,k}$ up to third order, say, $q_{T,2}$.

THEOREM 5.3. *Suppose that* [H1] *and* [H2] *hold true. Then* $q_{T,2}$ *defined in Corollary* 5.1 *has the following representation*:

$$q_{T,2}(y^{(0)}) = \phi(y^{(0)};g)\Bigg(1 + \frac{1}{\sqrt{T}}\{\bar{\Lambda}_1(y^{(0)}) + \bar{\Lambda}_2(y^{(0)})\} + \frac{1}{T}\{\bar{\Lambda}_3(y^{(0)}) + \bar{\Lambda}_4(y^{(0)}) + \bar{\Lambda}_5(y^{(0)}) + \bar{\Lambda}_6(y^{(0)})\}\Bigg),$$

where

$$\bar{\Lambda}_1(y^{(0)}) = \frac{1}{6}\tilde{\lambda}^{abc}h_{abc}(y^{(0)};g), \qquad \bar{\Lambda}_2(y^{(0)}) = \sum_A q_{1,\phi}^{a,A}h_{aA}(y^{(0)};g),$$

$$\bar{\Lambda}_3(y^{(0)}) = \frac{1}{24}\tilde{\lambda}^{abcd}h_{abcd}(y^{(0)};g) + \frac{1}{72}\tilde{\lambda}^{abc}\tilde{\lambda}^{def}h_{abcdef}(y^{(0)};g),$$

$$\begin{aligned}\bar{\Lambda}_4(y^{(0)}) &= \frac{1}{6}\sum_{A,C}\tilde{\lambda}^{a'b'c'}q_{1,\phi}^{a,A}C_{a'b'c',A}^{C}h_{aC}(y^{(0)};g) + \frac{1}{2}\sum_{A,C}\tilde{\lambda}^{a'b'\kappa}q_{1,\kappa}^{a,A}C_{a'b',A}^{C}h_{aC}(y^{(0)};g)\\ &\quad + \frac{1}{2}\sum_{A,C}(\kappa\lambda)!\tilde{\lambda}^{a'\kappa\lambda}q_{1,\kappa\lambda}^{a,A}C_{a',A}^{C}h_{aC}(y^{(0)};g)\\ &\quad + \frac{1}{6}\sum_{A}(\kappa\lambda\mu)!\tilde{\lambda}^{\kappa\lambda\mu}q_{1,\kappa\lambda\mu}^{a,A}h_{aA}(y^{(0)};g),\end{aligned}$$

$$\bar{\Lambda}_5(y^{(0)}) = \sum_A q_{2,\phi}^{a,A}h_{aA}(y^{(0)};g),$$

and

$$\bar{\Lambda}_6(y^{(0)}) = \frac{1}{2}\sum_{K,A,B,C}(K)!q_1^{a,K,A}q_{1,K}^{b,B}C_{A,B}^{C}h_{abC}(y^{(0)};g).$$

Note that the coefficients multiplying the Hermite polynomials in $\bar{\Lambda}_1$, $\bar{\Lambda}_3$ and $\bar{\Lambda}_4$ depend on the cumulants of $\widetilde{Z}_T$, and that those in $\bar{\Lambda}_2$, $\bar{\Lambda}_4$, $\bar{\Lambda}_5$ and $\bar{\Lambda}_6$ depend on $\tilde{Q}_1$, $\tilde{Q}_2$.

The polynomial $\pi_{l,K}^a$ defined by (5.8) can be obtained as follows. Let v be a $q\times q$ symmetric matrix satisfying that $\bar{\Sigma}_{22,1}+v$ is a positive definite matrix. Then $h^K(y^{(1)},\bar{\Sigma}_{22,1}+v)$ is well-defined, and there exists a polynomial $\pi_{l,K}^{\bar{\Sigma}_{22,1}+v,a}$ in $y^{(0)}$ uniquely such that it is symmetric in indices K and

$$\tilde{Q}_l^a(y) = \sum_K \pi_{l,K}^{\bar{\Sigma}_{22,1}+v,a}(y^{(0)})h^K(y^{(1)},\bar{\Sigma}_{22,1}+v).$$

Since

$$\begin{aligned}\lim_{|v|\to 0}\int_{\mathbb{R}^q} f(z^{(1)})h_K(z^{(1)};\bar{\Sigma}_{22,1}+v)\phi(z^{(1)};\bar{\Sigma}_{22,1}+v)dz^{(1)}\\ = \int_{\mathbb{R}^q} f(z^{(1)})h_K(z^{(1)};\bar{\Sigma}_{22,1})\phi(z^{(1)};\bar{\Sigma}_{22,1})dz^{(1)},\end{aligned}$$

for any $f\in C_\uparrow^\infty(\mathbb{R}^q)$, we can obtain the polynomial $\pi_{l,K}^a$ as

$$\pi_{l,K}^a(y^{(0)}) = \lim_{|v|\to 0}\pi_{l,K}^{\bar{\Sigma}_{22,1}+v,a}(y^{(0)}). \tag{5.12}$$

In the next section, we will use (5.12) to obtain an asymptotic expansion of an M-estimator.

6. Asymptotic expansion of the M-estimators

The purpose of this section is to present the third order asymptotic expansion of the distribution of M-estimators for a general statistical model. Before the theorem for the asymptotic expansion, we will precisely discuss the existence of M-estimators and the convergence rate of the remainder of their stochastic expansions. After that, we will present asymptotic expansions of M-estimators. In the case where the moments of derivatives of the estimating function have some relations, so-called the Bartlett identities, it will be seen that the coefficients of the expansion becomes the well-known ones for the maximum likelihood estimator in i.i.d. or time series models. At the end of this section, we will show the application to the maximum likelihood estimator for the diffusion process, which is one of the results in Sakamoto and Yoshida (1999).

Let Θ be an open bounded convex set included in $\mathbb{R}^p$ and T_0 a positive constant. For each $T > T_0$ and each $\theta \in \Theta$, let $(\mathfrak{X}_T, \mathfrak{A}_T)$ be a measurable space and X_T^θ an $\mathfrak{X}_T$-valued random variable on some probability space. For an estimating function $\psi_T : \mathfrak{X}_T \times \Theta \to \mathbb{R}^p$, an M-estimator $\hat{\theta}_T$ is defined as a solution of estimating equation $\psi_T(\theta, X_T^{\theta_0}) = 0$. In what follows, θ_0 denotes the true value in the parameter space Θ, and $\psi_T(\theta, X_T^{\theta_0})$ is abbreviated to $\psi(\theta)$, if there is no confusion. We also omit θ_0 in functions of θ when they are evaluated at θ_0. For example, $X_T = X_T^{\theta_0}$ and $\psi = \psi(\theta_0)$. Moreover $\psi_{a;}(\theta)$ denotes the a-th element of $\psi(\theta)$, and $(\psi_{a;a_1\cdots a_k}(\theta))_{a_1,\ldots,a_k=1,\ldots,p}$ denote the k-th derivatives of $\psi_{a;}(\theta)$ with respect to $\theta_{a_1}, \ldots, \theta_{a_k}$, i.e., $\psi_{a;a_1\cdots a_k}(\theta) = \delta_{a_1} \cdots \delta_{a_k} \psi_{a;}(\theta)$, where $\delta_a = \partial/\partial\theta^a$.

Let r_T be a positive bounded sequence tending to 0 as $T \to \infty$, and for $K \in \mathbb{N}$ and $a, a_1, \ldots, a_K = 1, \ldots, p$, let $\bar{\nu}_{a;a_1\cdots a_K}(\theta)$ be a tensor defined on Θ such that it is symmetric in $a_1, \ldots, a_K$. Each of them may depend on T, but is supposed to be bounded as $T \to \infty$. In order to show the existence of M-estimators and their stochastic expansion, we will assume the following conditions for $K \in \mathbb{N}$, $q > 1$ and $\gamma > 0$:

[C0]^K $\psi \in C^K(\Theta)$ a.s.;

[C1]$_q$ $\sup_{T>T_0} \|r_T \psi_{a;}(\theta_0)\|_q < \infty$ for $a = 1, \ldots, p$;

[C2]$^K_{q,\gamma}$ $\sup_{T>T_0, \theta\in\Theta} \|r_T^{-\gamma}(r_T^2 \psi_{a;a_1\cdots a_K}(\theta) - \bar{\nu}_{a;a_1\cdots a_K}(\theta))\|_q < \infty$;

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

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

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

Note that the tensors $\bar{\nu}_{a;a_1\cdots a_K}$ will be supposed to be the expectations of $r_T^2 \psi_{a;a_1\cdots a_K}$ when the asymptotic expansion will be considered. As for the case where $\theta \in \mathbb{R}^1$, the existence and the second order stochastic expansion of M-estimator were discussed by Sakamoto and Yoshida (1998a) under similar conditions to those given above.

THEOREM 6.1. *Let $m > 0$ and $\gamma \in (0,1)$. Suppose that* [C0]^2, [C1]$_{p_1}$, [C2]$^k_{p_2,\gamma}$, *$k = 1, 2$, and* [C3] *hold true for some $p_1 > m$, $p_2 > \max(p, m)$ and $p_3 > 1$ with $m/p_2 < \gamma < 1 - m/p_1$. Moreover, assume that $\delta_c \bar{\nu}_{a;b}(\theta) = \bar{\nu}_{a;bc}(\theta)$. Then*

$$P[(\exists_1 \hat{\theta}_T \in \tilde{\Theta} \text{ such that } \psi(\hat{\theta}_T) = 0) \text{ and } (|\hat{\theta}_T - \theta_0| < r_T^\gamma)] = 1 - o(r_T^m). \tag{6.1}$$

From this theorem, we see that for any $m > 0$, there exists a subspace $\tilde{\mathfrak{X}}_T$ such that $P(\tilde{\mathfrak{X}}_T) = 1 - o(r_T^m)$ and that for each observation $X_T \in \tilde{\mathfrak{X}}_T$, the M-estimator $\hat{\theta}_T$ for θ_0 can be defined as a solution of the estimating equation $\psi(\hat{\theta}_T) = 0$. In the sequel, any extension of $\hat{\theta}_T$ defined on the whole of the sample space $\mathfrak{X}_T$ will be referred to as the M-estimator of θ_0, and will be also denoted by $\hat{\theta}_T$. If we replace non-degeneracy condition [C3] with

$$\inf_{|x|=1} |x'\bar{\nu}_{a;b}(\theta_0)| > 0,$$

we obtain a similar result: under the same assumptions as in Theorem 6.1 except for [C3], (6.1) holds true for some open subset $\tilde{\Theta} \subset \Theta$.

This result is concerning a Cramér type consistency. In order to obtain a stronger consistency result, we need additional conditions; Condition [C3] ensures only local identifiability. Because constructing a consistent estimator is one of the subjects for 0-th-order asymptotics and it is also rather routine (and it is not our main problem here), we only consider this weak consistency here.

Let $\{\bar{\nu}_{a;}(\theta)\}_{a=1,\dots,p}$ be tensors defined on Θ such that they may be depend on T but $\sup_{T,\theta} \Delta_{a;}(\theta) < \infty$, where $\Delta_{a;}(\theta) = r_T^{-2}\bar{\nu}_{a;}(\theta)$. Put $Z_{a;} = r_T^{-1}(r_T^2\psi_{a;}(\theta_0) - \bar{\nu}_{a;}(\theta_0))$, $Z_{a;b} = r_T^{-1}(r_T^2\psi_{a;b}(\theta_0) - \bar{\nu}_{a;b}(\theta_0))$, and $Z_{a;bc} = r_T^{-1}(r_T^2\psi_{a;bc}(\theta_0) - \bar{\nu}_{a;bc}(\theta_0))$. Moreover, under Condition [C3], set $Z^{a;} = -\bar{\nu}^{a;a'}Z_{a';}$, $Z^{a;}{}_b = -\bar{\nu}^{a;a'}Z_{a';b}$, $Z^{a;}{}_{bc} = -\bar{\nu}^{a;a'}Z_{a';bc}$, $\bar{\nu}^{a;}{}_{a_1\cdots a_k} = -\bar{\nu}^{a;a'}\bar{\nu}_{a';a_1\cdots a_k}$, and $\Delta^{a;} = -\bar{\nu}^{a;a'}\Delta_{a';}$. We then obtain the stochastic expansion of a bias-corrected M-estimator $\hat{\theta}_T^*$ defined by

$$\hat{\theta}_T^* = \hat{\theta}_T - r_T^2\beta(\hat{\theta}_T)$$

for some bounded function β, and estimate the convergence rate of its remainder term.

THEOREM 6.2. *Let $m > 0$ and $\gamma \in (0,1)$. Suppose that $[C0]^4$, $[C1]_{p_1}$, $[C2]^k_{p_2,\gamma}$, $k = 1,2,3$, [C3], and $[C4]^4_{p_3}$ hold true for some $p_1 > 4m$, $p_2 > \max(p, 4m)$, $p_3 > m$ with $3/4 + \max\{m/p_2, m/(4p_3)\} < \gamma < 1 - m/p_1$. In addition, suppose that for the tensors $\bar{\nu}_{a;b}$ and $\bar{\nu}_{a;bc}$ in $[C2]^1_{p_2,\gamma}$ and $[C2]^2_{p_2,\gamma}$, $\delta_c\bar{\nu}_{a;b}(\theta) = \bar{\nu}_{a;bc}(\theta)$. Then there exists an M-estimator for θ_0. Moreover, for any extension $\hat{\theta}_T$ of the M-estimator and any $\beta \in C_b^2(\Theta)$, let R_3^a be defined by*

$$\begin{aligned}
(6.2)\qquad r_T^{-1}(\hat{\theta}_T^* - \theta_0)^a &= Z^{a;} + r_T\left(Z^{a;}{}_bZ^{b;} + \frac{1}{2}\bar{\nu}^{a;}{}_{bc}Z^{b;}Z^{c;} + \Delta^{a;} - \beta^a\right)\\
&\quad + r_T^2\Bigg(\frac{1}{6}(\bar{\nu}^{a;}{}_{bcd} + 3\bar{\nu}^{a;}{}_{be}\bar{\nu}^{e;}{}_{cd})Z^{b;}Z^{c;}Z^{d;} + \bar{\nu}^{a;}{}_{bc}Z^{b;}Z^{c;}{}_dZ^{d;}\\
&\quad + \frac{1}{2}\bar{\nu}^{b;}{}_{cd}Z^{a;}{}_bZ^{c;}Z^{d;} + \frac{1}{2}Z^{a;}{}_{bc}Z^{b;}Z^{c;} + Z^{a;}{}_bZ^{b;}{}_cZ^{c;}\\
&\quad - Z^{b;}\delta_b\beta^a + \Delta^{b;}(Z^{a;}{}_b + \bar{\nu}^{a;}{}_{bc}Z^{c;})\Bigg) + r_T^3R_3^a.
\end{aligned}$$

Then there exist $C > 0$ and $\epsilon > 0$ such that

$$(6.3)\qquad P[r_T|R_3^a| \le Cr_T^\epsilon, a = 1,\dots,p] = 1 - o(r_T^m).$$

Note that this stochastic expansion is a generalization of that of the MLE given by Barndorff-Nielsen and Cox (1994).

Combining this stochastic expansion with the formula of Theorem 5.1, we can derive an asymptotic expansion of M-estimator $\hat{\theta}_T^*$, but it is easier to derive it from that of Corollary 5.1. Therefore we consider another stochastic expansion consisting of orthogonal random variables. Suppose that (i) $(g^{ab}) := (\mathrm{Cov}[Z^{a;}, Z^{b;}])$ is a non-singular matrix and (ii) tensors $\bar{\nu}_{a;}$, $\bar{\nu}_{a;a_1\cdots a_k}$, $k = 1,2,3$, given above are the expectations of the estimating functions $\psi_{a;}$ and $\psi_{a;a_1\cdots a_k}$, i.e., $\bar{\nu}_{a;}(\theta) = E[r_T^2\psi_{a;}(\theta)]$ and $\bar{\nu}_{a;a_1\cdots a_k}(\theta) = E[r_T^2\psi_{a;a_1\cdots a_k}(\theta)]$ for $k = 1,2,3$. For index sets $B_1,\dots,B_k$, let

$$\bar{\nu}_{a_1;B_1,a_2;B_2,\dots,a_k;B_k} = r_T^2 E[\psi_{a_1;B_1}\psi_{a_2;B_2}\cdots\psi_{a_k;B_k}]$$

and

$$\bar{\nu}^{a_1;}{}_{B_1,}{}^{a_2;}{}_{B_2,}\cdots{}^{a_k;}{}_{B_k} = (-1)^k\bar{\nu}^{a_1;a_1'}\cdots\bar{\nu}^{a_k;a_k'}\bar{\nu}_{a_1';B_1,a_2';B_2,\dots,a_k';B_k}.$$

Define $\tilde{Z}^{a;}$ and $\tilde{Z}^{a;}{}_{a_1\cdots a_k}$ by $\tilde{Z}^{a;} = Z^{a;}$ and

$$\tilde{Z}^{a;}{}_{a_1\cdots a_k} = Z^{a;}{}_{a_1\cdots a_k} - V^{a;}{}_{a_1\cdots a_k,b}Z^{b;},$$

where $V^{a;}{}_{a_1\cdots a_k,b} = \mathrm{Cov}[Z^{a;}{}_{a_1\cdots a_k}, Z^{b';}]g_{b'b}$ and $(g_{ab}) = (g^{ab})^{-1}$, then $\tilde{Z}^{a;}$ and $\tilde{Z}^{b;}{}_{b_1\cdots b_k}$ are mutually orthogonal, i.e., $\mathrm{Cov}[\tilde{Z}^{a;}, \tilde{Z}^{b;}{}_{b_1\cdots b_k}] = 0$. For these random variables, we rewrite the stochastic expansion of $r_T^{-1}(\hat{\theta}_T^* - \theta_0)$.

THEOREM 6.3. *Let $m > 0$ and $\gamma \in (0,1)$. Suppose that the same condition as in Theorem 6.2 hold. For a bias-corrected M-estimator $\hat{\theta}_T^*$, let $\tilde{R}_3^a$ be defined by*

$$\begin{aligned} r_T^{-1}(\hat{\theta}_T^* - \theta_0)^a = \tilde{Z}^{a;} &+ r_T(\tilde{Z}^{a;}{}_b\tilde{Z}^{b;} + \tilde{\mu}^{a;}{}_{bc}\tilde{Z}^{b;}\tilde{Z}^{c;} - \tilde{\beta}^a) \\ &+ r_T^2(U^{a;}{}_{bcd}\tilde{Z}^{b;}\tilde{Z}^{c;}\tilde{Z}^{d;} + \tilde{\eta}^{a;}{}_{b,c}\tilde{Z}^{b;}{}_d\tilde{Z}^{c;}\tilde{Z}^{d;} + \tilde{\mu}^{b;}{}_{cd}\tilde{Z}^{a;}{}_b\tilde{Z}^{c;}\tilde{Z}^{d;} \\ &\quad + \frac{1}{2}\tilde{Z}^{a;}{}_{bc}\tilde{Z}^{b;}\tilde{Z}^{c;} + \tilde{Z}^{a;}{}_b\tilde{Z}^{b;}{}_c\tilde{Z}^{c;} - \tilde{Z}^{b;}\delta_b\beta^a \\ &\quad + \Delta^{b;}(\tilde{Z}^{a;}{}_b + \tilde{\eta}^{a;}{}_{b,c}\tilde{Z}^{c;})) + r_T^3\tilde{R}_3^{a;}, \end{aligned}$$

where $\tilde{\beta}^{a;} = \beta^a - \Delta^{a;}$, $\tilde{\mu}^{a;}{}_{bc} = (V^{a;}{}_{b,c} + V^{a;}{}_{c,b} + \bar{\nu}^{a;}{}_{bc})/2$, $\tilde{\eta}^{a;}{}_{b,c} = V^{a;}{}_{b,c} + \bar{\nu}^{a;}{}_{bc}$, and

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

Then there exist $\tilde{C} > 0$ and $\tilde{\epsilon} > 0$ such that

$$P[r_T|\tilde{R}_3^a| \le \tilde{C}r_T^{\tilde{\epsilon}}, a = 1,\dots,p] = 1 - o(r_T^m). \tag{6.4}$$

In the sequel, we set $r_T = T^{-1/2}$. Let

$$Z_T^{(0)} = T^{1/2}(Z_{1;},\dots,Z_{p;})$$

and

$$Z_T^{(1)} = T^{1/2}(\underbrace{Z_{1;1}, \dots, Z_{p;p}}_{p^2}, \underbrace{Z_{1;11}, \dots, Z_{p;pp}}_{p^3}).$$

Suppose that for some integer $q_1 \le p^2 + p^3$, there exists a q_1-dimensional random variable $\dot{Z}_T$ consisting of the elements of $Z^{(1)}$ such that $\mathrm{Cov}(Z_T^*)$ converges to a positive matrix and $\ddot{Z}_T = L\dot{Z}_T$ for some $q_2 \times q_1$ matrix L, where $Z_T^* = (Z^{(1)}, \dot{Z}_T)$, $\ddot{Z}_T$ is a q_2-dimensional random variable consisting of the other elements of $Z^{(1)}$ than those of $\dot{Z}_T$, and $q_1 + q_2 = p^2 + p^3$. Assume that X is an ϵ-Markov process with some driving process and that $Z^* = (Z_T^*)_{T\in\mathbb{R}_+}$ is an additive functional of them. For the definition of ϵ-Markov process, its driving process, and their additive functional, see Subsection 2.5. Put $\tilde{M}^{a_1;}{}_{a_2,}{}^{b_1;}{}_{b_2} = E[\tilde{Z}^{a_1;}{}_{a_2}\tilde{Z}^{b_1;}{}_{b_2}]$, $\tilde{N}^{a;}{}_{,}{}^{b;}{}_{,}{}^{a_1;}{}_{a_2} = T^{1/2}E[\tilde{Z}^{a;}\tilde{Z}^{b;}\tilde{Z}^{a_1;}{}_{a_2}]$, $\bar{\lambda}^{abc} = T^{1/2}\mathrm{Cum}[\tilde{Z}^{a;}, \tilde{Z}^{b;}, \tilde{Z}^{c;}]$, $H^{abcd} = T\mathrm{Cum}[\tilde{Z}^{a;}, \tilde{Z}^{b;}, \tilde{Z}^{c;}, \tilde{Z}^{d;}]$. We then obtain a third order asymptotic expansion of the distributions of M-estimators.

THEOREM 6.4. *Let $\hat{g} > \lim_{T\to\infty} g$, M and γ' be positive constants, γ a constant $\in (0,1)$ and m a positive constant satisfying $m > \gamma' + 2$. Suppose that* $[\mathrm{C0}]^4$, $[\mathrm{C1}]_{p_1}$, $[\mathrm{C2}]^k_{p_2,\gamma}$, $k = 1,2,3$, $[\mathrm{C3}]$, *and* $[\mathrm{C4}]^4_{p_3}$ *hold true for some $p_1 > 4m$, $p_2 > \max(p, 4m)$, $p_3 > m$ with $3/4 + \max\{m/p_2, m/(4p_3)\} < \gamma < 1 - m/p_1$, and that for the tensors $\bar{\nu}_{a;b}$ and $\bar{\nu}_{a;bc}$ in* $[\mathrm{C2}]^1_{p_2,\gamma}$ *and* $[\mathrm{C2}]^2_{p_2,\gamma}$, *$\delta_c\bar{\nu}_{a;b}(\theta) = \bar{\nu}_{a;bc}(\theta)$. Moreover, suppose that Conditions* [A1], [A2], [A3] *for Z_T^* hold true. Then there exist constants $c > 0$, $\tilde{C} > 0$, $\tilde{\epsilon} > 0$ and such that for any function $f \in \mathcal{E}(M, \gamma')$,*

$$\left|E[f(\sqrt{T}(\hat{\theta}_T^* - \theta))] - \int dy^{(0)} f(y^{(0)}) q_{T,2}(y^{(0)})\right| \le c\omega(f, \tilde{C}T^{-(\tilde{\epsilon}+2)/2}, \hat{g}) + o(T^{-1}), \tag{6.5}$$

where

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

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

If the estimating function ψ satisfies $\psi(\theta) = \nabla_\theta \rho(\theta)$ for some function ρ, then M-estimator $\hat{\theta}_T$ becomes a minimum contrast estimator corresponding to the contrast function ρ. In this case, $Z_{a;a_1\cdots a_k}$ is symmetric in all indices $a, a_1, \ldots, a_k$ and $\mathrm{Cov}(Z_T^{(1)})$ is singular, but the conditions given above Theorem 6.4 are fulfilled. Therefore the result of Theorem 6.4 comprehends the asymptotic expansion of the minimum contrast estimator.

In the following, we will show that the coefficients of asymptotic expansion of the M-estimator can be represented in terms of the information geometry when the estimating function $\psi_{a;a_1\cdots a_k}$ is symmetric in all indices $a, a_1, \ldots, a_k$ as in the case of the minimum contrast estimator and tensors g_{ab}, $\bar{\nu}_{a;A}$, and $\bar{\nu}_{a_1;A_1,\ldots,a_k;A_k}$ have some relations given below. In what follows, we assume that ψ is symmetric in all indices, and omit semicolon ; in the sequence of indices, e.g., $\bar{\nu}_{aa_1\cdots a_k} = \bar{\nu}_{a;a_1\cdots a_k}$. For the tensors $\bar{\nu}$'s, we suppose that the following identities hold true:

[BI1] $\Delta_a = 0$;

[BI2] $\bar{\nu}_{a,b} + \bar{\nu}_{ab} = 0$;

[BI3] $\bar{\nu}_{a,b,c} + \sum_{(ab,c)}^{[3]} \bar{\nu}_{ab,c} + \bar{\nu}_{abc} = o(\frac{1}{\sqrt{T}})$;

[BI4] $\bar{\nu}_{a,b,c,d} + \sum_{(ab,c,d)}^{[6]} \bar{\nu}_{ab,c,d} + \sum_{(ab,cd)}^{[3]} \bar{\nu}_{ab,cd} + \sum_{(abc,d)}^{[4]} \bar{\nu}_{abc,d} + \bar{\nu}_{abcd} = o(\frac{1}{\sqrt{T}})$.

These are often called Bartlett identities and hold for many of the maximum likelihood estimators. Furthermore, suppose that the following relations hold true:

[DV1] $\delta_a g_{bc} = \bar{\nu}_{ab,c} + \bar{\nu}_{ac,b} + \bar{\nu}_{a,b,c}$;

[DV2] $\delta_a \bar{\nu}_{bc,d} = \bar{\nu}_{abc,d} + \bar{\nu}_{bc,ad} + \bar{\nu}_{a,bc,d}$;

[DV3] $\delta_a \bar{\nu}_{b,c,d} = \bar{\nu}_{ab,c,d} + \bar{\nu}_{b,ac,d} + \bar{\nu}_{b,c,ad} + \bar{\nu}_{a,b,c,d}$.

In the i.i.d. or time series models, these relations are usually assumed, and as a sufficient condition the exchangeability between the differentiation w.r.t θ and the integration w.r.t. the density ν is habitually used. In general, it is not so clear whether these identities hold true or not, but just assuming these identities, we will examine whether our expansion given in Theorem 6.4 can be represented in terms of the information geometry.

Let

$$\Gamma_{abc}^{(\alpha)} = \bar{\nu}_{ab,c} + \frac{1-\alpha}{2}\bar{\nu}_{a,b,c} \quad \text{and} \quad \mu^a = -\frac{1}{2}\Gamma_{bca'}^{(-1)} g^{aa'} g^{bc}.$$

In the case where ψ is the log-likelihood function, $\Gamma_{abc}^{(\alpha)}$ is a coefficient of so-called α-connection. Moreover, put $M_{ab,cd} = E[Z_{ab}Z_{cd}]$, $N_{ab,c,d} = \sqrt{T}E[Z_{ab}Z_cZ_d]$, $L_{abc,d} = E[Z_{abc}Z_d]$, $H_{abcd} = T\mathrm{Cum}[Z_a, Z_b, Z_c, Z_d]$ and

$$\bar{L}_{abcd} = \sum_{(abc,d)}^{[4]} L_{abc,d}, \qquad \bar{N}_{abcd} = \sum_{(ab,c,d)}^{[6]} N_{ab,c,d}.$$

We then have the following asymptotic expansion.

THEOREM 6.5. *Let $\hat{g} > \lim_{T\to\infty} g$, M and γ' be positive constants, γ a constant $\in (0,1)$ and m a positive constant satisfying $m > \gamma'+2$. Suppose that the same conditions as in Theorem 6.4 hold true. Moreover, suppose that Conditions* [BI1], [BI2], [BI3], [BI4] *and Conditions* [DV1], [DV2], [DV3] *hold true. Then there exist constants $c > 0$, $\tilde{C} > 0$,*

$\tilde{\epsilon} > 0$ *such that for any function* $f \in \mathcal{E}(M, \gamma')$,

$$\left| E[f(\sqrt{T}(\hat{\theta}_T^* - \theta))] - \int dy^{(0)} f(y^{(0)}) q_{T,2}(y^{(0)}) \right| \leq c\omega(f, \tilde{C}T^{-(\tilde{\epsilon}+2)/2}, \hat{g}) + o(T^{-1}), \tag{6.6}$$

where

$$q_{T,2}(y^{(0)}) = \phi(y^{(0)}; g^{ab}) \bigg(1 + \frac{1}{6\sqrt{T}} c_{abc} h^{abc} - \frac{1}{\sqrt{T}} g_{ab}(\beta^b - \mu^b) h^a + \frac{1}{2T} A_{ab} h^{ab} + \frac{1}{24T} c_{abcd} h^{abcd} + \frac{1}{72T} c_{abc} c_{def} h^{abcdef} \bigg),$$

$$\begin{aligned}
c_{abc} &= -3\Gamma_{abc}^{(-1/3)}, \\
\mu^a &= -\frac{1}{2} g^{ab} g^{cd} \Gamma_{cdb}^{(-1)}, \\
A_{ab} &= \frac{1}{2} g^{cd} g^{ef} \Gamma_{cea}^{(-1)} \Gamma_{dfb}^{(-1)} + g^{cd}(M_{ac,bd} - \Gamma_{ace}^{(1)} g^{ef} \Gamma_{bdf}^{(1)}) \\
&\quad + g_{ac} g_{bd} (\beta^c - \mu^c)(\beta^d - \mu^d) - 2 g_{bc} \delta_a (\beta^c - \mu^c), \\
c_{abcd} &= -(3H_{abcd} + \bar{L}_{abcd} + 2\bar{N}_{abcd}) + 12\Gamma_{abc}^{(-1/3)} g_{de}(\beta^e - \mu^e) \\
&\quad + 12(\Gamma_{abe}^{(-1)} + \Gamma_{aeb}^{(1)}) \Gamma_{cfd}^{(-1)} g^{ef}.
\end{aligned}$$

Needless to say, this result includes asymptotic expansions of the minimum contrast estimators and the maximum likelihood estimators. In the case where $\hat{\theta}_T$ is the MLE, the representations of the coefficients of the asymptotic expansion in Theorem 6.5 coincides with those given by Taniguchi and Watanabe (1994). They derived them from formal asymptotic expansion and explained certain meanings of the coefficients for the MLE from the viewpoint of information geometry. We here only clarified some sufficient conditions under which the representations for the M-estimator are obtained validly and do not discuss meanings of their coefficients.

In the rest of this section, we will show the application to the maximum likelihood estimator for the diffusion model. The results are due to Sakamoto and Yoshida (1998*b*, 1999), where the third order asymptotic expansion of M-estimators is obtained for the (misspecified or specified) diffusion model.

For any $\theta \in \Theta$, let $X = (X_t)_{t \in \mathbb{R}_+}$ be a d-dimensional stationary diffusion process satisfying

$$dX_t = V_0(X_t, \theta)dt + V(X_t)dw_t, \tag{6.7}$$

with a stationary distribution ν_θ. Here $V_0 = (V_0^i)_{i=1,\dots,d} : \mathbb{R}^d \times \Theta \to \mathbb{R}^d$, $V = (V_j^i)_{i=1,\dots,d, j=1,\dots,r} : \mathbb{R}^d \to \mathbb{R}^d \otimes \mathbb{R}^r$, and w is an r-dimensional standard Wiener process defined on some probability space $(\Omega, \mathfrak{F}, P)$. Assume that (i) (6.7) has a strong solution X, (ii) $E|X_t|^k < \infty$ for any $t \in \mathbb{R}_+$ and $k \geq 1$, (iii) ν_θ is absolutely continuous with respect to the Lebesgue measure, (iv) for any $T > 0$, the log-likelihood function based on $X^T = (X_t)_{t \in [0,T]}$ (w.r.t. some reference measure) is given by

$$\ell(X, \theta) = \log \frac{d\nu_\theta}{dx}(X_0) + \int_0^T V_0'(VV')^{-1}(X_t, \theta) dX_t - \frac{1}{2} \int_0^T V_0'(VV')^{-1} V_0(X_t, \theta) dt.$$

Let

$$A(x,\theta)=\log\frac{d\nu_\theta}{dx}(x),\qquad B(x,\theta)=V_0'(VV')^{-1}V(x,\theta),$$
$$C(x,\theta)=B(x,\theta)\cdot\left(B(x,\theta_0)-\frac{1}{2}B(x,\theta)\right)$$

and denote the derivatives of A, B, C, and ℓ w.r.t θ by

$$A_{a_1\cdots a_k}(x,\theta)=\delta_{a_1}\cdots\delta_{a_k}A(x,\theta),\qquad B_{a_1\cdots a_k}(x,\theta)=\delta_{a_1}\cdots\delta_{a_k}B(x,\theta),$$
$$C_{a_1\cdots a_k}(x,\theta)=\delta_{a_1}\cdots\delta_{a_k}C(x,\theta),\qquad \ell_{a_1\cdots a_k}(\theta)=\delta_{a_1}\cdots\delta_{a_k}\ell(\theta).$$

For any measurable function $f:\mathbb{R}^d\to\mathbb{R}$, let $\check{f}$ be a function such that $\mathcal{A}\check{f}=f-\nu(f)$, and $[f]=-V'\nabla\check{f}$, where $\nu(f)=\int_{\mathbb{R}^d}f(x)\nu(dx)$ and

$$\mathcal{A}=\sum_{i=1}^{d}V_0^i(x,\theta_0)\frac{\partial}{\partial x^i}+\frac{1}{2}\sum_{i,j}^{d}\sum_{k=1}^{r}V_k^i(x)V_k^j(x)\frac{\partial^2}{\partial x^i\partial x^j}.$$

Assume that

[DM1] (i) for each $x\in\mathbb{R}^d$, $A(x,\cdot),B(x,\cdot),C(x,\cdot)\in C^6(\Theta)$;

(ii) there exist positive constants C_i, m_i, $i=1,2,3$ such that for any $x\in\mathbb{R}^d$, $k=1,\dots,6$, $a_k=1,\dots,p$,

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

[DM2] for any $\theta\in\Theta$, $k=1,\dots,5$, there exist functions $\check{C}_{a_1\cdots a_k}$, $a_1,\dots,a_k=1,\dots,p$ such that $\check{C}_{a_1\cdots a_k}\in C^2(\mathbb{R}^d)$ and that

$$\mathcal{A}\check{C}_{a_1\cdots a_k}(x)=C_{a_1\cdots a_k}(x,\theta)-\nu(C_{a_1\cdots a_k}(\cdot,\theta)),$$

[DM3] (i) for each $a=1,\dots,p$, $\nu(C_a(\cdot,\theta_0))=0$;

(ii) there exist positive constants C_i, m_i, $i=4,5$ such that for each $k=1,2,3,4$, and $a_1,\dots,a_k=1,\dots,p$,

$$\sup_{\theta\in\Theta}|\check{C}_{a_1\cdots a_k}(x,\theta)|\le C_4(1+|x|)^{m_4},\qquad \sup_{\theta\in\Theta}|[C]_{a_1\cdots a_k}(x,\theta)|\le C_5(1+|x|)^{m_5}.$$

We then have

THEOREM 6.6. (Sakamoto and Yoshida (1998*b*, 1999)) *Let $\theta_0\in\Theta$. Suppose that there exists a open subset $\tilde{\Theta}\subset\Theta$ such that $\theta_0\in\tilde{\Theta}$ and that the $p\times p$ matrix $(\nu(C_{ab}(\cdot,\theta)))$ is non-singular uniformly in $\theta\in\tilde{\Theta}$. Moreover, assume that for any $\theta\in\Theta$, $a,b,c=1,\dots,p$, $\delta_c\nu_{\theta_0}(A_{ab}(\cdot,\theta))=\nu_{\theta_0}(A_{abc}(\cdot,\theta))$, $\delta_c\nu_{\theta_0}(C_{ab}(\cdot,\theta))=\nu_{\theta_0}(C_{abc}(\cdot,\theta))$. Then, under Conditions* [DM1], [DM2], *and* [DM3], *for any $m>0$, $\gamma\in(0,1)$,*

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

Furthermore, for any extension of $\hat{\theta}_T$, say $\hat{\theta}_T$, and for any $\beta \in C_b^2(\Theta)$, (6.2) and (6.3) hold true with $r_T = 1/\sqrt{T}$ and $\psi_T = (\delta_a \ell)_{a=1}^p$.

For the maximum likelihood estimator given in Theorem 6.6, we can derive asymptotic expansion of its distribution from Theorem 6.4. Suppose that X has the geometric-mixing property, which ensure Condition [A1] in Section 5 with the diffusion process X and the Wiener process w in place of an ϵ-Markov process Y and a driving process X. The geometric-mixing property of diffusion processes, which are not necessarily symmetric, was shown by Kusuoka and Yoshida (2000). See Stroock (1994), Roberts and Tweedie (1996). Let

$$Z_T^{(0)} = T^{1/2}(Z_1, \dots, Z_p) \text{ and } Z_T^{(1)} = T^{1/2}(\underbrace{Z_{11}, \dots, Z_{pp}}_{p^2}, \underbrace{Z_{111}, \dots, Z_{ppp}}_{p^3}),$$

where

$$Z_a = \frac{1}{\sqrt{T}} \delta_a \ell(\theta_0), \qquad Z_{ab} = \frac{1}{\sqrt{T}} (\delta_{ab} \ell(\theta_0) - E_{\theta_0}[\delta_{ab} \ell(\theta_0)])$$
$$Z_{abc} = \frac{1}{\sqrt{T}} (\delta_{abc} \ell(\theta_0) - E_{\theta_0}[\delta_{abc} \ell(\theta_0)]).$$

Denote by $Z_{a,T}^{(0)}$ the a-th element of $Z_T^{(0)}$, and by $Z_{ab,T}^{(1)}$ and $Z_{abc,T}^{(1)}$ the (a,b)-th and the (a,b,c)-th elements of $Z_T^{(1)}$, respectively. Then they satisfy the following Stratonovich stochastic differential equations:

$$dZ_{a,t}^{(0)} = B_a(X_t, \theta_0) \circ dw_t + C_a^*(X_t, \theta_0) dt,$$
$$dZ_{ab,t}^{(1)} = B_{ab}(X_t, \theta_0) \circ dw_t + C_{ab}^*(X_t, \theta_0) dt,$$
$$dZ_{abc,t}^{(1)} = B_{abc}(X_t, \theta_0) \circ dw_t + C_{abc}^*(X_t, \theta_0) dt,$$

where

$$C_A^*(x, \theta) = C_A(x, \theta) - \frac{1}{2} \sum_{j=1}^r \sum_{k=1}^d V_j^k(x) \partial_k B_A^j(x, \theta)$$

for $A = \{a_1, a_1 a_2, a_1 a_2 a_3\}$, $a_i = 1, \dots, p$. Put

$$\tilde{V}_{0,1} = (\tilde{V}_0^1, \dots, \tilde{V}_0^d, C_1^*, \dots, C_p^*)$$

and

$$\tilde{V}_{i,1} = (V_i^1, \dots, V_i^d, B_1^i, \dots, B_p^i), \qquad i = 1, \dots, r,$$

where B_a^i is the i-th element of B_a and

$$\tilde{V}_0^i = V_0^i - \frac{1}{2} \sum_{j=1}^r \sum_{k=1}^d V_j^k \partial_k V_j^i, \qquad i = 1, \dots, d.$$

Assume that

[L] for some integer $q_1 \le p^2 + p^3$, there exists a q_1-dimensional random variable $\dot{Z}_T$ consisting of the elements of $Z^{(1)}$ such that

(i) $\mathrm{Cov}(Z_T^*)$ converges to a positive definite matrix, where $Z_T^* = (Z^{(0)}, \dot{Z}_T)$,

(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 $x \in \mathbb{R}^d$, $\bigcup_{n=0}^{\infty} \Sigma_n(x, 0) = \mathbb{R}^{d+p+q_1}$, where $\dot{C}_j^*$ 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, $\bar{V}_0 = (\bar{V}_{0,1}, \dot{C}_1^*, \dots, \dot{C}_{q_1}^*)$, $\bar{V}_i = (\bar{V}_{i,1}, \dot{B}_1^i, \dots, \dot{B}_{q_1}^i)$, $i = 1, \dots, r$, $\Sigma_0 = \{\bar{V}_1, \dots, \bar{V}_r\}$, $\Sigma_n = \{[\bar{V}_j, V] \mid V \in \Sigma_{n-1}, j = 0, 1, \dots, r\}$, and $[\bar{V}_j, V]$ is the Lie bracket.

Let us prepare some notations. Let

$$\begin{aligned}
&\check{F}_{A_1,A_2} = \nu_{\theta_0}(B_{A_1} \cdot B_{A_2}), \qquad \check{F}_{A_1,[A_2,A_3]} = \nu_{\theta_0}(B_{A_1} \cdot [B_{A_2} \cdot B_{A_3}]),\\
&\check{F}_{[A_1,A_2],[A_3,A_4]} = \nu_{\theta_0}([B_{A_1} \cdot B_{A_2}] \cdot [B_{A_3} \cdot B_{A_4}]),\\
&\check{F}_{[[A_1,A_2],A_3],A_4} = \nu_{\theta_0}([[B_{A_1} \cdot B_{A_2}] \cdot B_{A_3}] \cdot B_{A_4}),
\end{aligned}$$

where B_A's are evaluated at $\theta = \theta_0$. Put $\rho_{ab} = \check{F}_{a,b}$, $(\rho^{ab}) = (\rho_{ab})^{-1}$, $\tilde{\Delta}^a = \rho^{aa'}\nu_{\theta_0}(A_{a'})$, $\tau_{ab} = \mathrm{Cov}[A_a(X_0), A_b(X_0)]$, $\bar{A}^{a;b} = \rho^{aa'}\rho^{bb'}\nu(A_{a'b'})$. Moreover we need the followings:

$$\begin{aligned}
\tilde{\Gamma}^{(\alpha)}_{ab,c} &= \check{F}_{ab,c} - \check{F}_{[a,b],c} + \frac{1-\alpha}{2}\sum_{(ab,c)}^{[3]} \check{F}_{[a,b],c},\\
\check{\mu}^a &= -\frac{1}{2}\rho^{aa'}\rho^{bc}\tilde{\Gamma}^{(-1)}_{bc,a'}, \qquad \tilde{\eta}^{*a}{}_{b,c} = -\rho^{aa'}(\tilde{\Gamma}^{(1)}_{a'c,b} + \tilde{\Gamma}^{(-1)}_{bc,a'}).
\end{aligned}$$

Here $\sum_{(ab,c)}^{[3]}$ is a summation over the indicated number of terms obtained by rearranging the subscripts.

By using these notations, we can obtain the third order asymptotic expansion of MLE $\hat{\theta}_T$ for the diffusion process:

THEOREM 6.7. (Sakamoto and Yoshida (1998*b*, 1999)) *Let M, $\gamma > 0$, and $\hat{\rho} > (\rho^{ab})$. Assume that* [L] *and the conditions in Theorem* 6.6 *hold true. For any $\beta \in C_b^2(\Theta)$ and the $\hat{\theta}_T$ defined in Theorem* 6.6, *let $\hat{\theta}_T^* = \hat{\theta}_T - \beta(\hat{\theta}_T)/T$. Moreover assume that the diffusion process X given* (6.7) *has the geometrically strong mixing property. Then there exist positive constants $c, \tilde{C}, \tilde{\epsilon}$ such that for any $f \in \mathcal{E}(M, \gamma)$*

$$\begin{aligned}
&\left| E[f(\sqrt{T}(\hat{\theta}_T^* - \theta))] - \int dy^{(0)} f(y^{(0)}) q_{T,2}(y^{(0)}) \right| \\
&\qquad \le c\omega(f, \tilde{C}T^{-(\tilde{\epsilon}+2)/2}, \hat{\rho}^{ab}) + o(T^{-1}),
\end{aligned} \tag{6.8}$$

where

$$\begin{aligned}
q_{T,2}(y^{(0)}) = \phi(y^{(0)}; \rho^{ab}) \Bigg(1 &+ \frac{1}{6\sqrt{T}} c^*_{abc} h^{abc}(y^{(0)}; \rho^{ab}) \\
&+ \frac{1}{\sqrt{T}} \rho_{aa'}(\check{\mu}^{a'} - \tilde{\beta}^{a'}) h^a(y^{(0)}; \rho^{ab}) + \frac{1}{2T} A^*_{ab} h^{ab}(y^{(0)}; \rho^{ab}) \\
&+ \frac{1}{24T} c^*_{abcd} h^{abcd}(y^{(0)}; \rho^{ab}) + \frac{1}{72T} c^*_{abc} c^*_{def} h^{abcdef}(y^{(0)}; \rho^{ab}) \Bigg),
\end{aligned}$$

$$
\begin{aligned}
c^*_{abc} &= -3\tilde{\Gamma}^{(-1/3)}_{ab,c}, \qquad \tilde{\beta}^a = \beta^a - \Delta^a, \\
A^*_{ab} &= \tau_{ab} + 2\zeta_{ab} - \rho^{cd}(\check{F}_{bcd,a} + \check{F}_{ab,cd} - \check{F}_{ac,bd} - \check{F}_{[a,c],[b,d]} + 2\check{F}_{[ab,c],d} \\
&\qquad\qquad + 2\check{F}_{[ac,b],d} + 4\check{F}_{[b,d],ac} + \check{F}_{[cd,b],a} + 2\check{F}_{[[b,c],a],d} + 2\check{F}_{[[b,c],d],a}) \\
&\quad + \rho^{cd}\rho^{ef}\Big(\frac{1}{2}\tilde{\Gamma}^{(-1)}_{ce,b}\tilde{\Gamma}^{(-1)}_{df,a} - \tilde{\Gamma}^{(1)}_{ac,e}\tilde{\Gamma}^{(1)}_{bd,f} + \tilde{\Gamma}^{(-1)}_{cd,e}(\tilde{\Gamma}^{(1)}_{ab,f} + \tilde{\Gamma}^{(-1)}_{fb,a}) \\
&\qquad\qquad + \tilde{\Gamma}^{(-1)}_{ce,a}(\tilde{\Gamma}^{(1)}_{bd,f} + \tilde{\Gamma}^{(-1)}_{bd,f})\Big) \\
&\quad + \rho_{aa'}\rho_{bb'}(\check{\mu}^{a'} - \tilde{\beta}^{a'})(\check{\mu}^{b'} - \tilde{\beta}^{b'}) + 2\rho_{aa'}(\Delta^c\tilde{\eta}^{*a'}_{c,b} - \delta_b\beta^{a'}), \\
c^*_{abcd} &= -12(\check{F}_{[[a,b],c],d} + \check{F}_{[a,b],cd} + \check{F}_{[ab,c],d}) + 3\check{F}_{[a,b],[c,d]} - 4\check{F}_{abc,d} \\
&\quad + 12\tilde{\Gamma}^{(-1/3)}_{ab,c}\rho_{dd'}(\tilde{\beta}^{d'} - \check{\mu}^{d'}) + 12\rho^{ef}(\tilde{\Gamma}^{(-1)}_{ab,e} + \tilde{\Gamma}^{(1)}_{ae,b})\tilde{\Gamma}^{(-1)}_{cf,d}.
\end{aligned}
$$

Remark 2. Theorems 6.6 and 6.7 are specific versions of Theorems 4.1 and 4.2 for M-estimators in Sakamoto and Yoshida (1999), respectively.

Remark 3. In Theorem 6.7, the representation of the coefficients in the expansion are obtained without the identities [BI1] $\sim$ [BI4], [DV1] $\sim$ [DV3]. If one assumes those identities, the representation will become slightly simpler.

Remark 4. In Theorem 6.7, it is implicitly assumed that V, V_0, $B_A(\cdot,\theta)$, $C_A(\cdot,\theta) \in C^\infty(\mathbb{R}^d)$, $|A| \leq 3$ for the condition (iii) in [L], while Condition [DM2] in Theorem 6.6 requires that $\check{C}_{a_1\cdots a_k} \in C^2(\mathbb{R}^d)$ for any $\theta \in \Theta$, $k = 1, \ldots, 5, a_1, \ldots, a_k = 1, \ldots, p$.

Remark 5. When one considers the third order asymptotic expansion of the maximum likelihood estimators for the Ornstein-Uhlenbeck process, it is found that the condition (i) of [L] is not fulfilled. It is because of the complete linearity of this exceptional model. However, Sakamoto and Yoshida showed in 2000 that even in such a case, the third-order expansion formula of Theorem 6.7 is still valid (Sakamoto and Yoshida (2003)), as mentioned in Uchida and Yoshida (2001). This fact easily follows from a straightforward application of the perturbation method used in Yoshida (1997). For details, see Sakamoto and Yoshida (2003). In Uchida and Yoshida (1999, 2001), it has already been used for the model selection problems.

Remark 6. Here we adopted a Hörmander type condition for non-degeneracy of the distribution. It is a practical convenience because it involves only differentiation of coefficient vector fields. It is also possible to replace this condition by a condition that ensures local degeneracy of the Malliavin covariance, which is sufficient for our use. If the Malliavin covariance is non-degenerate at a skeleton in the support of the process, then the local degeneracy in the vicinity follows. See Yoshida (2001) for details.

7. Proofs of theorems in Section 5

In this and the next sections, some of the proofs are shortened for saving space. The details were given in Sakamoto and Yoshida (1999) or Sakamoto (1998).

7.1 *Proofs of Theorem* 5.1 *and Corollary* 5.1

In the case where C_T is the identity matrix for any $T > 0$, the asymptotic expansion of $\bar{Z}_T$ was given by Theorem 2 in Kusuoka and Yoshida (2000), but it is easy to extend it to the case treated here. In fact, we have the following proposition.

PROPOSITION 7.1. *Let M, γ, $K > 0$, and $\bar{g}^0$ be a positive definite matrix such that $\bar{g}^0 > \lim_{T\to\infty} \bar{g}$. Suppose that* [A1], [A2] *and* [A3] *are satisfied. Then there exist $\delta > 0$ and $c > 0$ such that for $f \in \mathcal{E}(M,\gamma)$,*

$$\left|E[f(\bar{Z}_T)] - \int_{\mathbb{R}^n} f(z)p_{T,k}(z)dz\right| \le c\omega(f, T^{-K}, \bar{g}^0) + \epsilon_T^{(k)},$$

where $\epsilon_T^{(k)} = o(T^{-(k+\delta)/2})$.

PROOF. Let $\bar{\Sigma} = \mathrm{Cov}(Z_T/\sqrt{T})$, $K' > K$ and $\Sigma^0 = \mathsf{C}^{-1}(\bar{g}^0 + \lim_{T\to\infty} \bar{g})(\mathsf{C}')^{-1}/2$, where $\mathsf{C} = \lim_{T\to\infty} \mathsf{C}_T$. Suppose that T is so large that C_T is non-singular. Since $\bar{\Sigma} = \mathsf{C}_T^{-1}\bar{g}(\mathsf{C}_T')^{-1}$, $\Sigma^0 > \lim_{T\to\infty} \bar{\Sigma}$. Therefore it follows from Theorem 2 in Kusuoka and Yoshida (2000) that there exist $\delta > 0$ and $c' > 0$ such that for $f \in \mathcal{E}(M,\gamma)$,

$$\left|E[f(\bar{Z}_T)] - \int_{\mathbb{R}^n} dz f(\mathsf{C}_T z)\tilde{p}_{T,k}(z)\right| \le c'\omega(f\circ \mathsf{C}_T, T^{-K'}, \Sigma^0) + \epsilon_T^{(k)},$$

where $\epsilon_T^{(k)} = o(T^{-(k+\delta)/2})$, and $\tilde{p}_{T,k}$ is the function taking the same form as $p_{T,k}$ with $\bar{g}$ and $\bar{\lambda}^{a_1\cdots a_m}$ replaced by $\bar{\Sigma}$ and $\bar{\kappa}^{a_1\cdots a_m}$, where

$$\bar{\kappa}^{a_1\cdots a_m} = T^{(m-2)/2}\mathrm{Cum}[Z_T^{a_1}/\sqrt{T}, \dots, Z_T^{a_m}/\sqrt{T}], \qquad m = 2, 3, 4, \dots .$$

Owing to the multilinearity of the cumulant, it follows that for $f \in \mathcal{E}(M,\gamma)$,

$$\int_{\mathbb{R}^n} dz f(\mathsf{C}_T z)\tilde{p}_{T,k}(z) = \int_{\mathbb{R}^n} dz f(z)p_{T,k}(z).$$

Moreover, take T so large that $|\mathsf{C}_T|T^{-K'} < T^{-K}$ and $\mathsf{C}_T\Sigma^0\mathsf{C}_T' < \bar{g}^0$, then

$$\omega(f\circ \mathsf{C}_T, T^{-K'}, \Sigma^0) \le c''\omega(f, T^{-K}, \bar{g}^0),$$

for some positive constant c''. Thus we complete the proof. □

From this proposition, we can prove Theorem 5.1.

PROOF OF THEOREM 5.1. Put $\Sigma = \lim_{T\to\infty} \bar{g}$ $(= \lim_{T\to\infty} \mathrm{Cov}(\bar{Z}_T))$ and partition it into four blocks, say,

$$\Sigma = \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix}$$

where $\Sigma_{11} = \lim_{T\to\infty} g$ $(= \lim_{T\to\infty} \mathrm{Cov}(\bar{Z}_T^{(0)}))$, $\Sigma_{12} = \lim_{T\to\infty} \mathrm{Cov}(\bar{Z}_T^{(0)}, \bar{Z}_T^{(1)})$, $\Sigma_{21} = \Sigma_{12}'$, and $\Sigma_{22} = \lim_{T\to\infty} \mathrm{Cov}(\bar{Z}_T^{(1)})$. For any $\hat{g} > \Sigma_{11}$ and $\hat{\Sigma}_{22} > \Sigma_{22}$, let $\check{\Sigma}_{11} = (\Sigma_{11} + \hat{g})/2$, $\check{\Sigma}_{22} = (\Sigma_{22} + \hat{\Sigma}_{22})/2$,

$$\check{\Sigma} = \begin{bmatrix} \check{\Sigma}_{11} & \Sigma_{12} \\ \Sigma_{21} & \check{\Sigma}_{22} \end{bmatrix}, \qquad \hat{\Sigma} = \begin{bmatrix} \hat{g} & \Sigma_{12} \\ \Sigma_{21} & \hat{\Sigma}_{22} \end{bmatrix}.$$

Then it holds that $\Sigma < \check{\Sigma} < \hat{\Sigma}$. For this $\check{\Sigma}$ and any $K' > K$, we see from Proposition 7.1 that there exist $c' > 0$ and $\delta > 0$ such that for $f \in \mathcal{E}(M, \gamma)$,

$$\left| E[f(S_T)] - \int_{\mathbb{R}^{p+q}} f(S_T(z)) p_{T,k}(z) dz \right| \le c' \omega(f \circ S_T, T^{-K'}, \check{\Sigma}) + \epsilon_T^{(k)}$$

where $\epsilon_T^{(k)} = o(T^{-(k+\delta)/2})$ and

$$S_T(z) = z^{(0)} + \sum_{i=1}^{k} T^{-i/2} Q_i(z^{(0)}, z^{(1)}), \qquad z = (z^{(0)}, z^{(1)}).$$

Note that δ, c' and $\epsilon_T^{(k)}$ depend not only on k, M, γ and K, but also on positive constants M' and γ' satisfying $\sup_T |S_T(z)| \le M'(1+|z|^{\gamma'})$. In order to obtain $q_{T,k}$, we will rewrite $\int_{\mathbb{R}^{p+q}} f(S_T(z)) p_{T,k}(z) dz$, and will estimate $\omega(f \circ S_T, T^{-K'}, \check{\Sigma})$ later.

First, let us consider the Bhattacharya-Ghosh map $y(z)$, which is defined by

$$y(z) = \begin{bmatrix} y^{(0)}(z) \\ y^{(1)}(z) \end{bmatrix} = \begin{bmatrix} S_T(z) \\ z^{(1)} \end{bmatrix}.$$

Put $\bar{Q}(z) = \sum_{j=1}^{k} T^{-j/2} Q_j(z)$, and take $\alpha > 0$ so that there exists $C_1 > 0$ such that for any $T > 1$ and $|z| < T^\alpha + 1$,

$$|\bar{Q}(z)| \vee \max \left\{ \left| \frac{\partial \bar{Q}^j(z)}{\partial z^i} \right| \,\middle|\, i = 1, \dots, n, \; j = 1, \dots, p \right\} \le C_1 T^{-\alpha}.$$

From Taylor's theorem, we see that for any $z_1, z_2 \in \mathbb{R}^{p+q}$, $y(z_1) - y(z_2) = J(z_1, z_2)(z_1 - z_2)$, where $J(z_1, z_2)$ is a $(p+q) \times (p+q)$ matrix defined by

$$J(z_1, z_2) = \left(\int_0^1 \frac{\partial y^i}{\partial z^j}(z_2 + u(z_1 - z_2)) du \right).$$

If $T > 1$ and $|z| < T^\alpha + 1$,

$$|J(z_1, z_2)| = \left| I_p + \int_0^1 \frac{\partial \bar{Q}}{\partial z^{(0)}}(z_2 + u(z_1 - z_2)) du \right| \ge 1 - C_2 T^{-\alpha}$$

for some constant $C_2 > 0$. Therefore for sufficiently large $T > 0$, the map $y(z)$ is one-to-one on M_T, where $M_T = \{z \in \mathbb{R}^{p+q} \mid |z| < T^\alpha\}$. In the following, we suppose T to be sufficiently large so that $y(z)$ is one-to-one on M_T and denote the inverse map by $z(y)$.

Let ρ be a real-valued function on $\mathbb{R}$ such that (i) $\rho \in C^\infty$, (ii) $0 \le \rho(x) \le 1$, and (iii) $\rho(x) = 1$ if $x \le 0$ and $\rho(x) = 0$ if $x \ge 1$. Put $\rho_T(z) = \rho(|z|^2 - T^{2\alpha} + 1)$, then

$$\sup_T |\partial_{i_1} \cdots \partial_{i_m} \rho_T(z)| \le \text{a polynomial of } |z|.$$

By using the Bhattacharya-Ghosh map $y(z)$, we see that

$$\begin{aligned} \int_{\mathbb{R}^{p+q}} f(S_T(z)) p_{T,k}(z) dz &= \int_{\mathbb{R}^{p+q}} f(S_T(z)) \rho_T(z) p_{T,k}(z) dz + o(T^{-K}) \\ &= \int_{\mathbb{R}^p} f(y^{(0)}) q_{T,k}^*(y^{(0)}) dy^{(0)} + o(T^{-K}), \end{aligned}$$

where

$$q^*_{T,k}(y^{(0)}) = \int_{\mathbb{R}^q} \rho_T(z(y)) p_{T,k}(z(y)) \left| \frac{\partial z}{\partial y} \right| dy^{(1)}.$$

The signed measure $q^*_{T,k}(y^{(0)})dy^{(0)}$ is rewritten as follows. From Taylor's theorem and the integration-by-parts formula, we obtain that for any $h \in C_0^\infty(\mathbb{R}^p)$,

$$\begin{aligned}
&\int_{\mathbb{R}^p} h(y^{(0)}) q^*_{T,k}(y^{(0)}) dy^{(0)} \\
&\quad = \int_{\mathbb{R}^p} h(S_T(z)) \rho_T(z) p_{T,k}(z) dz \\
&\quad = \int_{\mathbb{R}^p} h(z^{(0)}) \int_{\mathbb{R}^q} \rho_T(z) p_{T,k}(z) dz^{(1)} dz^{(0)} \\
&\quad\quad + \int_{\mathbb{R}^p} h(z^{(0)}) \sum_{j=1}^{k} \frac{(-1)^j}{j!} \partial_{a_1} \cdots \partial_{a_j} \left(\int_{\mathbb{R}^q} (\bar{Q}^{a_1} \cdots \bar{Q}^{a_j} \rho_T p_{T,k})(z) dz^{(1)} \right) dz^{(0)} + R_f,
\end{aligned}$$

where

$$\begin{aligned}
R_f = \frac{1}{k!} \int_0^1 (1-u)^k \int_{\mathbb{R}^{p+q}} & (\partial_{a_1} \cdots \partial_{a_{k+1}} h)(z^{(0)} + u\bar{Q}(z)) \\
& \times \bar{Q}^{a_1}(z) \cdots \bar{Q}^{a_{k+1}}(z) \rho_T(z) p_{T,k}(z) dz du.
\end{aligned}$$

For each $u \in [0,1]$, let

$$y_u(z) = \begin{bmatrix} y_u^{(0)} \\ y_u^{(1)} \end{bmatrix} = \begin{bmatrix} z^{(0)} + u\bar{Q}(z) \\ z^{(1)} \end{bmatrix}.$$

Since this map y_u is one-to-one on M_T, we have that

$$\begin{aligned}
R_f = \frac{1}{k!} \int_0^1 (1-u)^k \int_{\mathbb{R}^{p+q}} & (\partial_{a_1} \cdots \partial_{a_{k+1}} h)(y^{(0)}) \bar{Q}^{a_1}(z_u(y)) \cdots \bar{Q}^{a_{k+1}}(z_u(y)) \\
& \times \rho_T(z_u(y)) p_{T,k}(z_u(y)) \left| \frac{\partial z_u}{\partial y} \right| dy du,
\end{aligned}$$

where z_u is the inverse map of y_u. It is easy to show that if $y \in y_u(M_T)$, then

$$\left| \frac{\partial^k z_u^i}{\partial y^{j_1} \cdots \partial y^{j_k}} \right| \leq \text{a polynomial in } |z_u(y)| \leq \text{a polynomial in } |y|$$

and that if $y \in y_u(M_T)$, then

$$|p_{T,k}(z_u(y))| \leq C_3 \phi(y; g^*)$$

for some positive matrix $g^* > \bar{g}$ and a positive constant C_3 independent of T. Therefore it follows from Fubini's theorem and the integration-by-parts formula that

$$R_f = \int_{\mathbb{R}^p} h(y^{(0)}) R_1(y^{(0)}) dy^{(0)},$$

where

$$R_1(y^{(0)}) = \frac{(-1)^{k+1}}{k!}\partial_{a_1}\cdots\partial_{a_{k+1}}\left(\int_{\mathbb{R}^q}\int_0^1 (1-u)^k(\bar{Q}^{a_1}\cdots\bar{Q}^{a_{k+1}}\rho_T p_{T,k})(z_u(y)) \times\left|\frac{\partial z_u}{\partial y}\right|(y)dudy^{(1)}\right).$$

Thus we obtain that

$$q^*_{T,k}(y^{(0)}) = \int_{\mathbb{R}^q}\rho_T(y)p_{T,k}(y)dy^{(1)} + \sum_{j=1}^k \frac{(-1)^j}{j!}\partial_{a_1}\cdots\partial_{a_j}\left(\int_{\mathbb{R}^q}(\bar{Q}^{a_1}\cdots\bar{Q}^{a_j}\rho_T p_{T,k})(y)dy^{(1)}\right) + R_1(y^{(0)}).$$

Note that $\sup_{y^{(0)}}|R_1(y^{(0)})| \le C_4 T^{-(k+1)/2}\phi(y^{(0)};g^{**})$ for some positive matrix $g^{**} > g$ and $C_4 > 0$. Since $\int_{\mathbb{R}^{p+q}} f(y^{(0)})(\rho_T(y)-1)p_{T,k}(y)dy = o(T^{-K})$ and

$$\int_{\mathbb{R}^p} f(y^{(0)})\sum_{j=1}^k \frac{(-1)^j}{j!}\partial_{a_1}\cdots\partial_{a_j}\left(\int_{\mathbb{R}^q}(\bar{Q}^{a_1}\cdots\bar{Q}^{a_j}(\rho_T-1)p_{T,k})(y)dy^{(1)}\right)dy^{(0)} = o(T^{-K})$$

for any $f\in\mathcal{E}(M,\gamma)$, we obtain that for any $f\in\mathcal{E}(M,\gamma)$,

$$\begin{aligned}\int_{\mathbb{R}^{p+q}} & f(S_T(z))p_{T,k}(z)dz \\ &= \int_{\mathbb{R}^{p+q}} f(y^{(0)})p_{T,k}(y)dy \\ &\quad + \int_{\mathbb{R}^p} f(y^{(0)})\sum_{j=1}^k\frac{(-1)^j}{j!}\partial_{a_1}\cdots\partial_{a_j}\left(\int_{\mathbb{R}^q}(\bar{Q}^{a_1}\cdots\bar{Q}^{a_j}p_{T,k})(y)dy^{(1)}\right)dy^{(0)} \\ &\quad + O(T^{-(k+1)/2}) + o(T^{-K}).\end{aligned}$$

Substituting (5.1) and the definition of $\bar{Q}$ into the first and the second terms of the right hand side, one can easily show that

$$\int_{\mathbb{R}^{p+q}} f(S_T(z))p_{T,k}(z)dz = \int_{\mathbb{R}^p} f(y^{(0)})q_{T,k}(y^{(0)})dy^{(0)} + O(T^{-(k+1)/2}) + o(T^{-K}).$$

Finally, we will rewrite $\omega(f\circ S_T, T^{-K'},\check{\Sigma})$. From the definition of ω,

$$\begin{aligned}\omega(f\circ S_T, T^{-K'},\check{\Sigma}) &= \int_{\mathbb{R}^{p+q}}\sup\{|f\circ S_T(z+y) - f\circ S_T(z)| : |y| < T^{-K'}\}1_{\{z\in M_T\}}\phi(z;\check{\Sigma})dz \\ &\quad + o(T^{-K}).\end{aligned}$$

Since $|\partial\bar{Q}^j/\partial z^i| < C_1T^{-\alpha}$ for any $T>1$ and $z\in\mathbb{R}^{p+q}$ with $|z| < T^\alpha + 1$, if $z\in M_T$ and $|y| < T^{-K'}$, then

$$|S_T(z+y) - S_T(z)| \le C_5|y| \le C_5T^{-K'}$$

for some constant $C_5 > 0$. Therefore, taking so large T that $C_5 T^{-K'} < T^{-K}$, we obtain that

$$\omega(f \circ S_T, T^{-K'}, \hat{\Sigma}) \le \omega(f, T^{-K}, \hat{g}) + o(T^{-K}),$$

which completes the proof. □

PROOF OF COROLLARY 5.1. Let

$$M_T = \begin{bmatrix} I & 0 \\ -\Sigma_{21}\Sigma_{11}^{-1} & I \end{bmatrix}.$$

Then $\widetilde{Z}_T = M_T C_T(T^{-1/2} Z_T)$. Applying Proposition 7.1 to $\widetilde{Z}_T$, we obtain the same inequality for $\widetilde{Z}_T$ as in Proposition 7.1. Therefore, in exactly the same way as in Theorem 5.1, we can show the assertion. □

7.2 *Proof of Theorem* 5.2

Let $\breve{Z}_T = (\bar{Z}_T^{(0)}, \dot{Z}_T/\sqrt{T})$ and $\breve{L}$ be a $(p+q) \times (p+q_1)$ matrix defined by

$$\breve{L} = \begin{bmatrix} I_p & 0 \\ 0 & L^{(1)} \end{bmatrix}$$

where I_m is the m-dimensional identity matrix and

$$L^{(1)} = C_T^{(1)} \begin{bmatrix} I_{q_1} \\ L \end{bmatrix}.$$

Then

$$S_T = \breve{Z}_T^{(0)} + \sum_{i=1}^{k} T^{-i/2} \breve{Q}_i(\breve{Z}_T)$$

where $\breve{Z}_T^{(0)} = \bar{Z}_T^{(0)}$ and $\{\breve{Q}_i\}$ are polynomials in $z^* \in \mathbb{R}^{p+q_1}$ defined by $\breve{Q}_i(z^*) = Q_i(\breve{L}z^*)$.

Applying Theorem 5.1 to this functional S_T with $\breve{Z}_T$, we see that (5.2) holds true with

$$\breve{q}_{T,k}(y^{(0)}) = \int_{\mathbb{R}^{q_1}} \phi(y^*; \breve{g}) d\dot{y} + \sum_{m=1}^{k} T^{-m/2} \Bigg(\int_{\mathbb{R}^{q_1}} \breve{\Xi}_{T,m}(y^*) \phi(y^*; \breve{g}) d\dot{y}$$
$$+ \sum_{\substack{s+l=m \\ s \ge 0, l \ge 1}} \sum_{j=1}^{l} \frac{(-1)^j}{j!} \sum_{l_1 + \cdots + l_j = l} \partial_{i_1} \cdots \partial_{i_j} \left(\int_{\mathbb{R}^{q_1}} \breve{Q}_{l_1}^{i_1}(y^*) \cdots \breve{Q}_{l_j}^{i_j}(y^*) \breve{\Xi}_{T,s}(y^*) \phi(y^*; \breve{g}) d\dot{y} \right) \Bigg),$$

where $\breve{g} = \mathrm{Cov}(\breve{Z}_T)$,

$$\breve{\Xi}_{T,j}(y^*) = \sum_{m=1}^{j} \frac{1}{m!} \sum_{\substack{k_1 + \cdots + k_m = j \\ k_1 \ge 1, \ldots, k_m \ge 1}} \frac{\breve{\lambda}^{A^*_{k_1+2}} \cdots \breve{\lambda}^{A^*_{k_m+2}}}{(k_1+2)! \cdots (k_m+2)!} h_{A^*_{k_1+2} \cdots A^*_{k_m+2}}(y^*; \breve{g}),$$

and

$$\breve{\lambda}^{\alpha_1^* \cdots \alpha_j^*} = T^{(j-2)/2} i^{-j} \partial^{\alpha_1^*} \cdots \partial^{\alpha_j^*} \log E[e^{iu^* \cdot \breve{Z}_T}] \,|_{u=0}\,.$$

Note that indices with the symbol $*$, e.g., α_j^*, run from 1 to $p+q_1$, and that index sets with the star $*$, e.g., $A_{k_j}^*$, consist of such indices. For any polynomial p in z, let

$$\eta_p(y^{(0)}) = \bar{\lambda}^{A_{k_1}} \cdots \bar{\lambda}^{A_{k_m}} \int_{\mathbb{R}^q} p(y) h_{A_{k_1} \cdots A_{k_m}}(y;\bar{g})\phi(y;\bar{g})dy^{(1)}.$$

If we can show that

$$\eta_p(y^{(0)}) = \breve{\lambda}^{B_{k_1}^*} \cdots \breve{\lambda}^{B_{k_m}^*} \int_{\mathbb{R}^{q_1}} p(\breve{L}y^*) h_{B_{k_1}^* \cdots B_{k_m}^*}(y^*;\breve{g})\phi(y^*;\breve{g})d\dot{y}, \tag{7.1}$$

we will see that the inequality of Theorem 5.1 holds true with the same $q_{T,k}$ interpreted as (5.5) and (5.6).

First, we see from the definition of $h_{\alpha_1\cdots\alpha_j}(y;\bar{g})\phi(y;\bar{g})$ that for any $f \in C_0^\infty(\mathbb{R}^p)$,

$$\begin{aligned}
&\int_{\mathbb{R}^p} f(y^{(0)})\eta_p(y^{(0)})dy^{(0)} \\
&\quad = \int_{\mathbb{R}^p} f(y^{(0)})\bar{\lambda}^{A_{k_1}} \cdots \bar{\lambda}^{A_{k_m}} \int_{\mathbb{R}^q} p(y)h_{A_{k_1}\cdots A_{k_m}}(y;\bar{g})\phi(y;\bar{g})dy^{(1)}dy^{(0)} \\
&\quad = \bar{\lambda}^{A_{k_1}} \cdots \bar{\lambda}^{A_{k_m}} \int_{\mathbb{R}^p} f(y^{(0)}) \int_{\mathbb{R}^{q_1}} (\partial_{A_{k_1}^{(1)}} \cdots \partial_{A_{k_m}^{(1)}} p)(\breve{L}y^*) \\
&\qquad\qquad \times (-1)^{|A_{k_1}^{(0)}|+\cdots+|A_{k_m}^{(0)}|}\partial_{A_{k_1}^{(0)}} \cdots \partial_{A_{k_m}^{(0)}} \phi(y^*;\breve{g})d\dot{y}dy^{(0)}.
\end{aligned}$$

Put $\bar{f}(y) = f(y^{(0)})$. For index sets $A_{k_j} = \alpha_1 \cdots \alpha_{k_j}$ and $B_{k_j}^* = \beta_1^* \cdots \beta_{k_j}^*$, let $\breve{L}_{B_{k_j}^*}^{A_{k_j}} = \breve{L}_{\beta_1^*}^{\alpha_1} \cdots \breve{L}_{\beta_{k_j}^*}^{\alpha_{k_j}}$. Since

$$\bar{\lambda}^{\alpha_1\cdots\alpha_j} = \breve{L}_{\beta_1^*}^{\alpha_1} \cdots \breve{L}_{\beta_j^*}^{\alpha_j} \breve{\lambda}^{\beta_1^*\cdots\beta_j^*}$$

and

$$\partial_{\beta_1^*} \cdots \partial_{\beta_j^*}(p(\breve{L}y^*)) = \tilde{L}_{\beta_1^*}^{\alpha_1} \cdots \tilde{L}_{\beta_j^*}^{\alpha_j}(\partial_{\alpha_1} \cdots \partial_{\alpha_j} p)(\breve{L}y^*),$$

it follows that

$$\begin{aligned}
&\int_{\mathbb{R}^p} f(y^{(0)})\eta_p(y^{(0)})dy^{(0)} \\
&\quad = \bar{\lambda}^{A_{k_1}} \cdots \bar{\lambda}^{A_{k_m}} \int_{\mathbb{R}^p}\int_{\mathbb{R}^{q_1}} (\partial_{A_{k_1}} \cdots \partial_{A_{k_m}} \bar{f} \cdot p)(\breve{L}y^*)\phi(y^*;\breve{g})d\dot{y}dy^{(0)} \\
&\quad = \breve{L}_{B_{k_1}^*}^{A_{k_1}} \cdots \breve{L}_{B_{k_m}^*}^{A_{k_m}} \breve{\lambda}^{B_{k_1}^*} \cdots \breve{\lambda}^{B_{k_m}^*} \int_{\mathbb{R}^p}\int_{\mathbb{R}^{q_1}} (\partial_{A_{k_1}} \cdots \partial_{A_{k_m}} \bar{f} \cdot p)(\breve{L}y^*)\phi(y^*;\breve{g})d\dot{y}dy^{(0)} \\
&\quad = \breve{\lambda}^{B_{k_1}^*} \cdots \breve{\lambda}^{B_{k_m}^*} \int_{\mathbb{R}^p}\int_{\mathbb{R}^{q_1}} \partial_{B_{k_1}^*} \cdots \partial_{B_{k_m}^*}((\bar{f} \cdot p)(\breve{L}y^*))\phi(y^*;\breve{g})d\dot{y}dy^{(0)} \\
&\quad = \int_{\mathbb{R}^p} f(y^{(0)})\breve{\lambda}^{B_{k_1}^*} \cdots \breve{\lambda}^{B_{k_m}^*} \int_{\mathbb{R}^{q_1}} p(Ly^*)h_{B_{k_1}^*\cdots B_{k_m}^*}(y^*;\breve{g})\phi(y^*;\breve{g})d\dot{y}dy^{(0)},
\end{aligned}$$

which implies (7.1).

Applying Corollary 5.1 to the representation of S_T by the orthogonalized random variable of $\breve{Z}_T$, we can derive another version of $q_{T,k}^*$. In the same way as above we can show, that the inequality of Corollary 5.1 holds true with the same $q_{T,k}$ interpreted as (5.5) and (5.6). □

7.3 *Proof of Theorem* 5.3

By using the definitions (5.8) and (5.10), we obtain from Corollary 5.1 and Theorem 5.2 that

$$\begin{aligned} q_{T,2}(y^{(0)}) = \phi(y^{(0)};g)\Bigg(1 &+ \frac{1}{\sqrt{T}}\{\bar{\Lambda}_1(y^{(0)}) + \bar{\Lambda}_2(y^{(0)})\} \\ &+ \frac{1}{T}\{\bar{\Lambda}_3(y^{(0)}) + \bar{\Lambda}_4(y^{(0)}) + \bar{\Lambda}_5(y^{(0)})\}\Bigg) \\ &+ \frac{1}{2T}\partial_a\partial_b\int_{\mathbb{R}^q} \tilde{Q}_1^a(y)\tilde{Q}_1^b(y)\phi(y;\tilde{\Sigma})dy^{(1)}. \end{aligned}$$

Therefore, it suffices to show that

$$\bar{\Lambda}_6(y^{(0)})\phi(y^{(0)};g) = \frac{1}{2}\partial_a\partial_b\int_{\mathbb{R}^q} \tilde{Q}_1^a(y)\tilde{Q}_1^b(y)\phi(y;\tilde{\Sigma})dy^{(1)}.$$

For any $q\times q$ symmetric matrix v satisfying that $\Sigma_{22,1}+v$ is non-singular, there exist polynomials $\pi_{l,K}^{\Sigma_{22,1}+v,a}$ and $\pi_l^{\Sigma_{22,1}+v,a,K}$ such that

$$\begin{aligned} \tilde{Q}_l^a(y) &= \sum_K \pi_{l,K}^{\Sigma_{22,1}+v,a}(y^{(0)})h^K(y^{(1)};\Sigma_{22,1}+v) \\ &= \sum_K \pi_l^{\Sigma_{22,1}+v,a,K}(y^{(0)})h_K(y^{(1)};\Sigma_{22,1}+v). \end{aligned}$$

For these polynomials, we see that

$$\int_{\mathbb{R}^q} \tilde{Q}_1^a(y)\tilde{Q}_1^b(y)\phi(y^{(1)};\Sigma_{22,1}+v)dy^{(1)} = \sum_K (K)!\pi_1^{\Sigma_{22,1}+v,a,K}(y^{(0)})\pi_{1,K}^{\Sigma_{22,1}+v,b}(y^{(0)}).$$

Thus we have that

$$\int_{\mathbb{R}^q} \tilde{Q}_1^a(y)\tilde{Q}_1^b(y)\phi(y;\tilde{\Sigma})dy^{(1)} = \sum_{K,A,B,C}(K)!q_1^{a,K,A}q_{1,K}^{b,B}C_{A,B}^{C}h_C(y^{(0)};g)\phi(y^{(0)};g),$$

which completes the proof. □

8. Proofs of theorems in Section 6

As in the previous section, some of the proofs in this section are shortened for saving space. See the details in Sakamoto and Yoshida (1999) or Sakamoto (1998).

8.1 *Proof of Theorem* 6.1

First, we prepare a lemma for the proof of Theorem 6.1.

LEMMA 8.1. *Let $m>0$, $\gamma>0$, and $p_2>p$ with $\gamma p_2>m$. Assume that* $[\mathrm{C0}]^2$, $[\mathrm{C2}]^k_{p_2,\gamma}$, $k=1,2$ *hold true, and that $\delta_c\bar{\nu}_{a;b}(\theta)=\bar{\nu}_{a;bc}(\theta)$. Then for any $C_0>0$, a, $b=1,\dots,p$,*

$$P\left[\sup_{\theta\in\tilde{\Theta}}|r_T^2\psi_{a;b}(\theta)-\bar{\nu}_{a;b}(\theta)|>C_0\right] = \frac{1}{C_0^{p_2}}\epsilon_T^{(m)},$$

where $\epsilon_T^{(m)} = o(r_T^m)$ *and it is independent of* C_0. *Furthermore, assume that* [C3] *holds true. Then there exists a constant* $C_1 > 0$ *independent of* T *such that*

$$P\left[\inf_{\substack{\theta_1,\theta_2\in\tilde{\Theta}\\|x|=1}}\left|-r_T^2\left(\int_0^1 \psi_{a;b}(\theta_1+s(\theta_2-\theta_1))ds\right)' x\right| > C_1\right] = 1 - o(r_T^m).$$

PROOF. Since $p_2 > p$, $\delta_c\bar{\nu}_{a;b}(\theta) = \bar{\nu}_{a;bc}(\theta)$ and $[\mathrm{C0}]^2$ hold, it is seen from Sobolev's inequality that there exists a positive constant C_Θ independent of T such that for any $X_T \in \mathfrak{X}_T$ and a, $b = 1, \ldots, p$,

$$\begin{aligned}
\sup_{\theta\in\tilde{\Theta}}|r_T^2\psi_{a;b}(\theta) - \bar{\nu}_{a;b}(\theta)| &\le \sup_{\theta\in\Theta}|\, r_T^2\psi_{a;b}(\theta) - \bar{\nu}_{a;b}(\theta)| \\
&\le C_\Theta\left(\int_\Theta |r_T^2\psi_{a;b}(\theta) - \bar{\nu}_{a;b}(\theta)|^{p_2} d\theta\right. \\
&\qquad \left. + \int_\Theta \left|\sum_{c=1}^{c}(r_T^2\psi_{a;bc}(\theta) - \bar{\nu}_{a;bc}(\theta))\right|^{p_2} d\theta\right)^{1/p_2}.
\end{aligned}$$

Combining this inequality with the conditions $\gamma p_2 > m$, $[\mathrm{C2}]^1_{p_2,\gamma}$ and $[\mathrm{C2}]^2_{p_2,\gamma}$, we can show that for any $C_0 > 0$ and a, $b = 1, \ldots, p$

$$P\left[\sup_{\theta\in\tilde{\Theta}}|r_T^2\psi_{a;b}(\theta) - \bar{\nu}_{a;b}(\theta)| \ge C_0\right] = \frac{1}{C_0^{p_2}}\epsilon_T^{(m)},$$

where $\epsilon_T^{(m)} = o(r_T^m)$ and it is independent of C_0.

Furthermore, from [C3], there exists a constant C_1 independent of T such that

$$\inf_{\substack{\theta_1,\theta_2\in\tilde{\Theta}\\|x|=1}}\left|x'\left(\int_0^1 \bar{\nu}_{a;b}(\theta_1+s(\theta_2-\theta_1))ds\right)\right| > 2C_1.$$

Therefore we see that

$$\begin{aligned}
&P\left[\inf_{\substack{\theta_1,\theta_2\in\tilde{\Theta}\\|x|=1}}\left|-r_T^2\left(\int_0^1 \psi_{a;b}(\theta_1+s(\theta_2-\theta_1))ds\right)' x\right| \le C_1\right] \\
&\qquad \le \sum_{a,b}^{p} P\left[\sup_{\theta\in\tilde{\Theta}}\{|r_T^2\psi_{a;b}(\theta) - \bar{\nu}_{a;b}(\theta)|\} \ge C_1/p^2\right] = o(r_T^m),
\end{aligned}$$

which completes the proof. □

From this lemma, we can show Theorem 6.1.

PROOF OF THEOREM 6.1. Let C be a positive constant satisfying

$$\inf_{\substack{\theta_1,\theta_2\in\tilde{\Theta}\\|x|=1}}\left|x'\left(\int_0^1 \bar{\nu}_{a;b}(\theta_1+s(\theta_2-\theta_1))ds\right)\right| > 2C.$$

By virtue of $[\mathrm{C0}]^2$, there exists $\tilde{T}_0(C) > 0$ such that for any $T > \tilde{T}_0(C)$ and $\delta \in \mathbb{R}^p$ satisfying $|\delta| \le 1$, $\{\theta : |\theta - \theta_0| < r_T^\gamma\} \subset \tilde{\Theta}$ and $|\bar{\nu}_{a;b}(\theta_0 + \delta r_T^\gamma) - \bar{\nu}_{a;b}(\theta_0)| < C/(2p^2)$. For such $C > 0$, let $\mathfrak{X}_{T,0}$ be the subset of $\mathfrak{X}_T$ defined by

$$\mathfrak{X}_{T,0} := \left\{ X_T \in \mathfrak{X}_T \;\middle|\; \inf_{\substack{\theta \in \tilde{\Theta} \\ |x|=1}} \left| -r_T^2 \left(\int_0^1 \psi_{a;b}(\theta_1 + s(\theta_2 - \theta_1)) ds \right)' x \right| > C, \right.$$
$$\left. |r_T^{2-\gamma} \psi(\theta_0)| < C, \sup_{\theta \in \tilde{\Theta}} |r_T^2 \psi_{a;b}(\theta) - \bar{\nu}_{a;b}(\theta)| < \frac{C}{2p^2} \text{ for } a, b = 1, \dots, p \right\},$$

and for any a, $b = 1, \dots, p$, $X_T \in \mathfrak{X}_{T,0}$ and $T > \tilde{T}_0$, let $\hat{I}_{ab}$ be a function on $\{u : |u| \le 1\}$ defined by

$$\hat{I}_{ab}(u) = -r_T^2 \int_0^1 \psi_{a;b}(\theta_0 + r_T^\gamma u \xi) d\xi.$$

Then, for $x \in \mathbb{R}^p$ satisfying $|x| = 1$

$$|(\hat{I}_{ab}(u) + \bar{\nu}_{a;b}(\theta_0))' x| \le \sum_{a,b}^{p} \left(\sup_{\theta \in \tilde{\Theta}} |r_T^2 \psi_{a;b}(\theta) - \bar{\nu}_{a;b}(\theta)| + C/2p^2 \right) < C.$$

Therefore

$$\inf_{|x|=1} |(\hat{I}_{ab}(u))' x| \ge \inf_{|x|=1} (|(-\bar{\nu}_{a;b}(\theta_0))' x| - |(\hat{I}_{ab}(u) + \bar{\nu}_{a;b}(\theta_0))' x|) > C,$$

which implies that matrix $(\hat{I}_{ab}(u))$ is non-singular. Let $H_\gamma(u) = (H_\gamma^a(u))_{a=1,\dots,p}$ be a function on $\{u : |u| \le 1\}$ defined by

$$H_\gamma^a(u) := r_T^{2-\gamma} \hat{I}^{ab}(u) \psi_{b;}(\theta_0),$$

where $(\hat{I}^{ab}(u)) = (\hat{I}_{ab}(u))^{-1}$. We then have that for any $X_T \in \mathfrak{X}_{T,0}$ and any $T \ge \tilde{T}_0$,

$$|H_\gamma(u)| \le C \sup_{|x| \le 1} |(\hat{I}^{ab}(u)) x| = \frac{C}{\inf_{|x|=1} |(\hat{I}_{ab}(u))' x|} \le 1.$$

Therefore it follows from Brouwer's fixed point theorem that if $X_T \in \mathfrak{X}_{T,0}$ and $T \ge \tilde{T}_0$, then there exists a $\hat{u} \in \{u : |u| \le 1\}$ such that $H_\gamma(\hat{u}) = \hat{u}$. Setting $\hat{\theta}_T = \theta_0 + r_T^\gamma \hat{u}$, we have from Taylor's theorem that for each $a = 1, \dots, p$,

$$\psi_{a;}(\hat{\theta}_T) = \psi_{a;}(\theta_0) + \int_0^1 \psi_{a;b}(\theta_0 + r_T^\gamma \hat{u} \xi) d\xi r_T^\gamma \hat{u}^b = r_T^{\gamma-2} \hat{I}_{ab}(\hat{u})(H_\gamma^b(\hat{u}) - \hat{u}^b) = 0.$$

Since $(\int_0^1 \psi_{a;b}(\theta_1 + s(\theta_2 - \theta_1)) ds)$ is non-singular uniformly in $\theta_1, \theta_2 \in \tilde{\Theta}$ for any $X_T \in \mathfrak{X}_{T,0}$, we see that if $X_T \in \mathfrak{X}_{T,0}$ and $T \ge \tilde{T}_0$, then there exists a unique $\hat{\theta}_T \in \tilde{\Theta}$ satisfying $\psi(\hat{\theta}_T) = 0$ and such $\hat{\theta}_T$ lies in the r_T^γ-neighborhood of θ_0. Furthermore, we see from $[\mathrm{C1}]_{p_1}$ and $\gamma < 1 - m/p_1$ that

$$P[|r_T^{2-\gamma} \psi(\theta_0)| \ge C] \le \frac{r_T^{(1-\gamma)p_1}}{C^{p_1}} \|r_T \psi(\theta_0)\|_{p_1}^{p_1} = o(r_T^m).$$

From this and Lemma 8.1, $P[(\mathfrak{X}_{T,0})^c] = o(r_T^m)$. Thus we see that

$$P[(\exists_1 \hat{\theta}_T \in \tilde{\Theta}, \psi(\hat{\theta}_T) = 0) \text{ and } (|\hat{\theta}_T - \theta_0| < r_T^\gamma)] = 1 - o(r_T^m). \qquad \square$$

Note that if the tensors $\bar{\nu}$'s are independent of T, the additional assumption in Theorem 6.1, i.e., $\delta_c \bar{\nu}_{a;b}(\theta) = \bar{\nu}_{a;bc}(\theta)$, can be proved. In fact, we have

LEMMA 8.2. *Assume that* $[\mathrm{C0}]^4$, $[\mathrm{C2}]^k_{p_2,\gamma}$, $k = 1,2,3$, $[\mathrm{C4}]^4_{p_3}$ *hold for some* $\gamma > 0$, $p_2 > 1$ *and* $p_3 > 1$. *Moreover, assume that* $\bar{\nu}_{a;b}(\theta)$, $\bar{\nu}_{a;bc}(\theta)$ *and* $\bar{\nu}_{a;bcd}(\theta)$ *are independent of* T. *Then* $\bar{\nu}_{a;b} \in C^2(\Theta)$, $\bar{\nu}_{a;bc} \in C^1(\Theta)$, *and* $\bar{\nu}_{a;bcd}$ *satisfies Lipschitz's condition. Moreover* $\delta_c \bar{\nu}_{a;b} = \bar{\nu}_{a;bc}$ *and* $\delta_d \bar{\nu}_{a;bc} = \bar{\nu}_{a;bcd}$.

PROOF. In the same way as in the proof of Lemma 1 in Sakamoto and Yoshida (1998*a*), we can prove this lemma. See Sakamoto and Yoshida (1999) or Sakamoto (1998). $\square$

8.2 *Proof of Theorem* 6.2

We have two lemmas for the proof of Theorem 6.2.

LEMMA 8.3. *Let* $T > 0$ *and* K *be a positive integer greater than or equal to* 3. *For any* $k = 1, \ldots, K$, $j = 1, \ldots, k$ *and* $a, a_j = 1, \ldots, p$, *let* $\bar{\nu}_{a;}$ *and* $\bar{\nu}_{a;a_1\cdots a_k}$ *be constants depending on* T *such that a matrix* $(\bar{\nu}_{a;b})$ *is non-singular for each* T *and that* $\bar{\nu}_{a;a_1\cdots a_k}$ *is symmetric in* $a_1, \ldots, a_k$. *Suppose that* $[\mathrm{C0}]^K$ *holds and that for any* $\theta_0 \in \Theta$, *there exists* $\hat{\theta}_T \in \Theta$ *such that* $\psi(\hat{\theta}_T) = 0$. *For any* $\theta_0 \in \Theta$, *define* $Z_{a;}$, $Z_{a;a_1\cdots a_k}$ *and* $\bar{\theta}^{a_1\cdots a_k}$, $k = 1, \ldots, K$, *by*

$$Z_{a;} = r_T^{-1}(r_T^2 \psi_{a;}(\theta_0) - \bar{\nu}_{a;})$$
$$Z_{a;a_1\cdots a_k} = r_T^{-1}(r_T^2 \psi_{a;a_1\cdots a_k}(\theta_0) - \bar{\nu}_{a;a_1\cdots a_k})$$

and

$$\bar{\theta}^{a_1\cdots a_k} = r_T^{-k}(\hat{\theta}_T - \theta_0)^{a_1} \cdots (\hat{\theta}_T - \theta_0)^{a_k},$$

respectively. Then for any $k = 2, \ldots, K-1$,

$$\bar{\theta}^a = -\bar{\nu}^{a;b} Z_{b;} - r_T^{-1} \bar{\nu}^{a;b} \bar{\nu}_{b;} - \sum_{j=1}^{k-2} r_T^j \left(\frac{1}{j!} \bar{\nu}^{a;b} Z_{b;A_j} \bar{\theta}^{A_j} + \frac{1}{(j+1)!} \bar{\nu}^{a;b} \bar{\nu}_{b;A_{j+1}} \bar{\theta}^{A_{j+1}} \right)$$
$$+ r_T^{k-1} \bar{R}^a_{k-1},$$

where $A_j = a_1 \cdots a_j$ *and* $B_j = b_1 \cdots b_j$ *are arbitrary index sets of length* j, $(\bar{\nu}^{a;b})$ *is the inverse matrix of* $(\bar{\nu}_{a;b})$, *and* $\bar{R}^a_k$ *is defined by*

$$\bar{R}^a_{k-1} = -\frac{1}{(k-1)!} \bar{\nu}^{a;b} Z_{b;A_{k-1}} \bar{\theta}^{A_{k-1}}$$
$$-\frac{1}{(k-1)!} \int_0^1 (1-u)^{k-1} \bar{\nu}^{a;b} r_T^2 \psi_{b;A_k}(\theta_0 + u(\hat{\theta}_T - \theta_0)) du \bar{\theta}^{A_k}.$$

PROOF. One can easily obtain the result from the Taylor expansion. See Sakamoto and Yoshida (1999) or Sakamoto (1998). □

LEMMA 8.4. *Suppose that* $[\mathrm{C0}]^4$ *holds true and that for any* $\theta_0 \in \Theta$, *there exists* $\hat{\theta}_T \in \Theta$ *such that* $\psi(\hat{\theta}_T) = 0$. *For* $\bar{\nu}_{a;}$, $Z_{a;}$, $\bar{\nu}_{a;a_1\cdots a_k}$ *and* $Z_{a;a_1\cdots a_k}$, $k = 1,\dots,4$, $a, a_1,\dots,a_k = 1,\dots,p$, *in Lemma 8.3, put* $\Delta_{a;} = r_T^{-2}\bar{\nu}_{a;}$, $\Delta^{a;} = -\bar{\nu}^{a;b}\Delta_{b;}$, $Z^{a;} = -\bar{\nu}^{a;b}Z_{b;}$, $Z^{a;}{}_{a_1\cdots a_k} = -\bar{\nu}^{a;a'}Z_{a';a_1\cdots a_k}$, *and* $\bar{\nu}^{a;}{}_{a_1\cdots a_k} = -\bar{\nu}^{a;a'}\bar{\nu}_{a';a_1\cdots a_k}$. *Then*

$$
\begin{aligned}
\bar{\theta}^a = Z^{a;} &+ r_T\left(Z^{a;}{}_b Z^{b;} + \frac{1}{2}\bar{\nu}^{a;}{}_{bc}Z^{b;}Z^{c;} + \Delta^{a;}\right) \\
&+ r_T^2\bigg(\frac{1}{6}(\bar{\nu}^{a;}{}_{bcd} + 3\bar{\nu}^{a;}{}_{be}\bar{\nu}^{e;}{}_{cd})Z^{b;}Z^{c;}Z^{d;} + \bar{\nu}^{a;}{}_{bc}Z^{b;}Z^{c;}{}_d Z^{d;} + \frac{1}{2}\bar{\nu}^{b;}{}_{cd}Z^{a;}{}_b Z^{c;}Z^{d;} \\
&\qquad + \frac{1}{2}Z^{a;}{}_{bc}Z^{b;}Z^{c;} + Z^{a;}{}_b Z^{b;}{}_c Z^{c;} + \Delta^{b;}Z^{a;}{}_b + \bar{\nu}^{a;}{}_{bc}\Delta^{b;}Z^{c;}\bigg) \\
&+ r_T^3\check{R}_3^a,
\end{aligned}
$$

where

$$
\begin{aligned}
\check{R}_3^a = {}& Q^{a,2}(Z) + r_T Q^{a,1}(Z) + Q_b^{a,2}(Z)\bar{R}_1^b + r_T Q_b^{a,1}(Z)\bar{R}_1^b + r_T Q_{bc}^{a,1}(Z)\bar{R}_1^b\bar{R}_1^c \\
&+ Q_b^{a,1}(Z)\bar{R}_2^b + Q_{bc}^{a,0}(Z)\bar{R}_1^b\bar{R}_1^c + r_T^2 Q^{a,0}(Z) + r_T^2 Q_b^{a,0}(Z)\bar{R}_1^b \\
&+ r_T^2 Q_{bc}^{a,0}(Z)\bar{R}_1^b\bar{R}_1^c + r_T^2 Q_{bcd}^{a,0}(Z)\bar{R}_1^b\bar{R}_1^c\bar{R}_1^d + \bar{R}_3^a.
\end{aligned}
\tag{8.1}
$$

$\{\bar{R}_k^a\}$ *are defined in Lemma 8.3, and the functions* $Q_{a_1\cdots a_k}^{a,m}(Z)$ *are polynomials with degree* m *in* $Z_{a;}$, $Z_{a;b}$, $Z_{a;bc}$ *whose coefficients are constants of order* $O(1)$ *as* $T \to \infty$.

PROOF. Using Lemma 8.3, the similar discussion to one in the proof of Theorem 1 in Sakamoto and Yoshida (1998*a*) gives the result. See Sakamoto and Yoshida (1999) or Sakamoto (1998). □

By using the results obtained above, we prove Theorem 6.2.

PROOF OF THEOREM 6.2. Expanding $\beta^a(\hat{\theta}_T)$ around $\hat{\theta}_T = \theta_0$, we have

$$
R_3^a = \check{R}_3^a - R_{(\beta)}^a,
\tag{8.2}
$$

where

$$
R_{(\beta)}^a = \delta_b\beta^a R_1^a + \int_0^1 (1-u)\delta_b\delta_c\beta^a(\theta_0 + u(\hat{\theta}_T - \theta_0))du\bar{\theta}^b\bar{\theta}^c.
$$

For $\gamma' \in (3/4, \gamma - m/p_2)$ and $\alpha \in (m/p_3, 4\gamma - 3)$, let $\mathfrak{X}_{T,1}$ be a subset of $\mathfrak{X}_T$ defined by

$$
\begin{aligned}
\mathfrak{X}_{T,1} = \bigg\{X_n \in \mathfrak{X}_n \;\bigg|\; & |r_T^2\psi_{a;b} - \bar{\nu}_{a;b}| < r_T^{\gamma'}, |r_T^2\psi_{a;bc} - \bar{\nu}_{a;bc}| < r_T^{\gamma'}, \\
& |r_T^2\psi_{a;bcd} - \bar{\nu}_{a;bcd}| < r_T^{\gamma'}, \sup_{\theta\in\Theta}|r_T^2\psi_{a;bcde}(\theta)| < r_T^{-\alpha}, a,b,c,d,e = 1,\dots,p\bigg\}.
\end{aligned}
$$

Moreover, let $\mathfrak{X}_{T,0}$ be a subset of $\mathfrak{X}_T$ defined in the proof of Theorem 6.1. As we discussed in the proof of Theorem 6.1, we see that for any $X_T \in \mathfrak{X}_{T,0}$ and any index set A_k of length k, $\bar{\theta}^{A_k} < r_T^{k(\gamma-1)}$. From the definitions of $\mathfrak{X}_{T,0}$ and $\mathfrak{X}_{T,1}$, we see that there exists a positive constant $\bar{C}$ such that for any $X_T \in \mathfrak{X}_{T,0} \cap \mathfrak{X}_{T,1}$, $k = 1, 2, 3$, $a = 1, \dots, p$, and $a_1, \dots, a_k = 1, \dots, p$,

$$|Z^{a;}| < \bar{C} r_T^{\gamma-1}, \qquad |Z^{a;}{}_{a_1 \cdots a_k}| < \bar{C} r_T^{\gamma'-1}.$$

Therefore, putting $\epsilon = \min(4\gamma' - 3, 4\gamma - 3 - \alpha)$, we see from the definition of $\bar{R}_3^a$ in Lemma 8.3 that

$$|r_T \bar{R}_3^a| \leq C_3 r_T^{\epsilon}$$

for some positive constant C_3. Moreover, from this inequality and the recurrence formula of $\bar{R}_k^a$, we inductively obtain the inequalities

$$|r_T \bar{R}_2^a| \leq C_2 r_T^{3\gamma'-2} \qquad \text{and} \qquad |\bar{R}_1^a| \leq C_1 r_T^{2\gamma'-2}$$

for some $C_2, C_1 > 0$. Therefore it follows from (8.1) and (8.2) that there exists a constant $C > 0$ such that for any $X_T \in \mathfrak{X}_{T,0} \cap \mathfrak{X}_{T,1}$,

$$|r_T R_3^a| \leq C r_T^{\epsilon}.$$

From Markov's inequality, we have that

$$P[(\mathfrak{X}_{T,1})^c] = o(r_T^m).$$

Thus it is shown that

$$P[r_T R_3^a \leq C r_T^{\epsilon}, a = 1, \dots, p] \geq P[\mathfrak{X}_{T,0} \cap \mathfrak{X}_{T,1}] = 1 - o(r_T^m). \qquad \square$$

8.3 *Proof of Theorem* 6.4

In Theorem 6.3, we showed that the coefficients of the stochastic expansions are given in terms of the moments of $\tilde{Z}$'s. In the following, we will first give a representation of $q_{T,2}$ in Theorem 5.3 in the case where the coefficients $q_{k,K}^{a,A}$ and $q_k^{a,A,K}$ of the covariant and contravariant representation of $\pi_{k,K}^a$ and $\pi_k^{a,K}$ satisfy some conditions; next we will represent $q_{T,2}$ in the case where the polynomials $\tilde{Q}_k$ have the same form as those of the stochastic expansion of the M-estimator. Finally, we will prove the theorem with the aid of the *Delta* method.

PROPOSITION 8.1. *Let p and q be positive integers. For $a = 1, \dots, p$, $A \in \{\phi, a_1, a_1 a_2, \dots\}$, $a_1, a_2, \dots \in \{1, \dots, p\}$, $K \in \{\phi, \kappa_1, \kappa_1\kappa_2, \dots\}$, $\kappa_1, \kappa_2, \dots \in \{p+1, \dots, p+q\}$, and $l = 1, 2$, $\{q_{l,K}^{a,A}\}$ are constants satisfying*

(1) $q_{1,\phi}^{a,A} = 0$, *if* $|A| = 1$ *or* $|A| \geq 3$,
(2) $q_{1,\kappa}^{a,A} = 0$, *if* $|A| = 0$ *or* $|A| \geq 2$,
(3) $q_{1,K}^{a,A} = 0$, *if* $|K| \geq 2$,
(4) $q_{2,\phi}^{a,A} = 0$, *if* $|A| = 0$, *or* $|A| = 2$, *or* $|A| \geq 4$.

Let (g^{ab}) *be a* $p \times p$*-positive matrix,* $\{\tilde\lambda^{\alpha\beta\gamma}\}_{\alpha,\beta,\gamma=1,\dots,p+q}$ *and* $\{\tilde\lambda^{abcd}\}_{a,b,c,d=1,\dots,p}$ *be sequences of constants, and* $q_l^{a,K,A} = g^{K,L} q_{l,K}^{a,A}$. *For these constants and* $T > 0$, *define a function* $q_{T,2}$ *on* $\mathbb{R}^p$ *in Theorem* 5.3. *Then*

$$q_{T,2}(y^{(0)}) = \phi(y^{(0)}; g^{ab}) \left(1 + \frac{1}{6\sqrt{T}} c^{abc} h_{abc} + \frac{1}{\sqrt{T}} q_{1,\phi}^{a,\phi} h_a + \frac{1}{2T} A^{ab} h_{ab} + \frac{1}{24T} c^{abcd} h_{abcd} + \frac{1}{72T} c^{abc} c^{def} h_{abcdef}\right),$$

where

$$c^{abc} = \tilde\lambda^{abc} + 6 q_{1,\phi}^{c,ab},$$

$$A^{ab} = 2(\tilde\lambda^{ace} + q_{1,\phi}^{a,ce}) q_{1,\phi}^{b,df} g_{cd} g_{ef} + (2\tilde\lambda^{ac\kappa} + q_1^{a,\kappa,c}) q_{1,\kappa}^{b,d} g_{cd} + 2 q_{2,\phi}^{a,b} + q_{1,\phi}^{a,\phi} q_{1,\phi}^{b,\phi},$$

$$c^{abcd} = \tilde\lambda^{abcd} + 4 c^{abc} q_{1,\phi}^{d,\phi} + 24(\tilde\lambda^{abe} + 2 q_{1,\phi}^{a,be}) q_{1,\phi}^{c,df} g_{ef} + 12(q_1^{a,\kappa,b} + \tilde\lambda^{ab\kappa}) q_{1,\kappa}^{c,d} + 24 q_{2,\phi}^{a,bcd}.$$

PROOF. From the definition of $\bar\Lambda_i$, $i = 1, \dots, 6$, and the recurrence formulas for the Hermit polynomials, the elementary algebra leads to the result. See Sakamoto and Yoshida (1999) or Sakamoto (1998). □

PROPOSITION 8.2. *Let* p *be a positive integer and* $q = p^2 + p^3$. *For variables* $\{y^{a;}\}_{a=1,\dots,p}$, $\{y^{a;}{}_b\}_{a,b=1,\dots,p}$, *and* $\{y^{a;}{}_{bc}\}_{a,b,c=1,\dots,p}$, *let* $y^{(0)} = (y^{1;}, \dots, y^{p;})$ *and*

$$y^{(1)} = (y^{1;}{}_1, \dots, y^{1;}{}_p, y^{2;}{}_1, \dots, y^{2;}{}_p, \dots, y^{p;}{}_p, y^{1;}{}_{11}, \dots, y^{1;}{}_{1p}, y^{1;}{}_{21}, \dots, y^{p;}{}_{pp}),$$

and $y = (y^{(0)}, y^{(1)})$. *Define* $Q_1^{a;}(y)$ *and* $Q_2^{a;}(y)$ *by*

$$Q_1^{a;}(y) = y^{a;}{}_b y^{b;} + \tilde\mu^{a;}{}_{bc} y^{b;} y^{c;} - \tilde\beta^a$$

and

$$Q_2^{a;}(y) = U^{a;}{}_{bcd} y^{b;} y^{c;} y^{d;} + \tilde\eta^{a;}{}_{b,c} y^{b;}{}_d y^{c;} y^{d;} + \tilde\mu^{b;}{}_{cd} y^{a;}{}_b y^{c;} y^{d;} + \frac{1}{2} y^{a;}{}_{bc} y^{b;} y^{c;} + y^{a;}{}_b y^{b;}{}_c y^{c;} - y^{b;} \dot\beta_b^a + \Delta^{b;} (y^{a;}{}_b + \tilde\eta^{a;}{}_{b,c} y^{c;}),$$

for some constants $\tilde\mu^{a;}{}_{bc}$, $\tilde\beta^a$, $\tilde\eta^{a;}{}_{b,c}$, $U^{a;}{}_{bcd}$, $\dot\beta_b^a$, *and* $\Delta^{b;}$. *Suppose that* $g = (g^{ab})$ *is a* $p \times p$*-positive matrix and* $\bar\Sigma_{22,1} = (\bar\Sigma_{22,1}^{\kappa\mu})$ *is a* $q \times q$*-symmetric matrix defined by* $\bar\Sigma_{22,1} = \tilde M \sigma_{22}^* \tilde M'$ *for some* $q_1 \times q_1$*-positive matrix* σ_{22}^* *and* $q \times q_1$*-matrix* $\tilde M$, *where* $q_1 \le q$. *For these polynomials* Q_1^a, Q_2^a *and* $\bar\Sigma = \mathrm{diag}(g, \bar\Sigma_{22,1})$, *define polynomials* $\pi_{k,K}^a$ *and constants* $q_{k,K}^{a,A}$ *by* (5.8) *and* (5.10), *respectively. Then* $q_{k,K}^{a,A}$ *satisfies* (1)–(4) *in Proposition* 8.1. *Moreover, let* $\{\tilde\lambda^{\alpha\beta\gamma}\}_{\alpha,\beta,\gamma=1,\dots,p+p^2}$ *and* $\{\tilde\lambda^{abcd}\}_{a,b,c,d=1,\dots,p}$ *be sequences of constants, and* $q_l^{a,K,A} = \bar\Sigma_{22,1}^{K,L} q_{l,L}^{a,A}$. *For these constants and* $T > 0$, *define a function* $q_{T,2}$ *on* $\mathbb{R}^p$ *in Theorem* 5.3. *Then*

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

where $\tilde{N}^{a;\,b;\,a_1;}_{\ \ ,\ \ ,\ \ a_2} = \tilde{\lambda}^{ab\kappa}$ *for* $\kappa \in \{p+1,\dots,p+p^2\}$ *corresponding to* $(a_1,a_2) \in \{(1,1),\dots,(p,p)\}$, $\tilde{M}^{a_1;}{}_{a_2,}{}^{b_1;}{}_{b_2} = \bar{\Sigma}^{\kappa\mu}_{22,1}$ *for* κ *and* $\mu \in \{p+1,\dots,p+p^2\}$ *corresponding to* (a_1,a_2) *and* $(b_1,b_2) \in \{(1,1),\dots,(p,p)\}$, *respectively,*

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

PROOF. From the definition of $Q_1^{a;}$, we have that

$$
Q_1^{a;}(y) = \delta^a_{a_1}g^{a_2b}g_{bb'}y^{b';}y^{a_1;}{}_{a_2} + \tilde{\mu}^{a;}{}_{b'c'}g^{b'b}g^{c'c}g_{bb''}g_{cc''}y^{b'';}y^{c'';} - \tilde{\beta}^a,
$$

which implies that

$$
\begin{aligned}
&\pi^{a;}_{1,\kappa}(y^{(0)}) = q^{a,b}_{1,\kappa}h_b(y^{(0)};g), \qquad q^{a,b}_{1,\kappa} = \begin{cases} \delta^a_{a_1}g^{a_2b} & (\kappa = a_1a_2)\\ 0 & (\kappa = a_1a_2a_3),\end{cases}\\
&\pi^{a;}_{1,\phi}(y^{(0)}) = q^{a,bc}_{1,\phi}h_{bc}(y^{(0)};g) + q^{a,\phi}_{1,\phi},\\
&q^{a,bc}_{1,\phi} = \tilde{\mu}^{a;}{}_{b'c'}g^{b'b}g^{c'c}, \qquad q^{a,\phi}_{1,\phi} = \tilde{\mu}^{a;}{}_{bc}g^{bc} - \tilde{\beta}^a,
\end{aligned}
$$

where $\kappa \in \{11,12,\dots,1p,21,\dots,2p,\dots,pp,111,112,\dots,11p,121,\dots,ppp\}$, and that $q^{a,A}_{1,K}$ satisfies the conditions (1), (2), (3) in Proposition 8.1. From (5.9) and (5.11), we have that

$$
\begin{aligned}
&\pi_1^{a;,\kappa}(y^{(0)}) = q_1^{a,\kappa,b}h_b(y^{(0)}), \qquad q_1^{a,\kappa,b} = \begin{cases} \delta^a_{b_1}g^{b_2b}\tilde{M}^{b_1;}{}_{b_2,}{}^{a_1;}{}_{a_2} & (\kappa = a_1a_2)\\ 0 & (\kappa = a_1a_2a_3),\end{cases}\\
&\pi_1^{a;,\phi}(y^{(0)}) = q_1^{a,\phi,bc}h_{bc}(y^{(0)}) + q_1^{a,\phi,\phi},\\
&q_1^{a,\phi,bc} = \tilde{\mu}^{a;}{}_{b'c'}g^{b'b}g^{c'c}, \qquad q_1^{a,\phi,\phi} = \tilde{\mu}^{a;}{}_{bc}g^{bc} - \tilde{\beta}^a.
\end{aligned}
$$

In the same fashion, we see that $\{q^{a,A}_{2,K}\}$ satisfies (4) in Proposition 8.1, and that

$$
\begin{aligned}
\pi^a_{2,\kappa\mu}(y^{(0)}) &= q^{a,b}_{2,\kappa\mu}h_b(y^{(0)})\\
\pi^a_{2,\kappa}(y^{(0)}) &= q^{a,bc}_{2,\kappa}h_{bc}(y^{(0)}) + q^{a,\phi}_{2,\kappa}\\
\pi^a_{2,\phi}(y^{(0)}) &= q^{a,bcd}_{2,\phi}h_{bcd}(y^{(0)}) + q^{a,b}_{2,\phi}h_b(y^{(0)})
\end{aligned}
$$

for some constants $q^{a,b}_{2,\kappa\mu}$, $q^{a,bc}_{2,\kappa}$, $q^{a,\phi}_{2,\kappa}$, $q^{a,bcd}_{2,\phi}$ and $q^{a,b}_{2,\phi}$, where $\kappa,\mu \in \{11,12,\dots,pp,111,112,\dots,ppp\}$. In particular,

$$
\begin{aligned}
q^{a,bcd}_{2,\phi} &= U^{a;}{}_{b'c'd'}g^{b'b}g^{c'c}g^{d'd}\\
q^{a,b}_{2,\phi} &= ((\Delta^{c;}\tilde{\eta}^{a;}{}_{c,b'} - \dot{\beta}^a_{b'}) + \delta^{a_1}_{b_1}\tilde{M}^{a;}{}_{a_1,}{}^{b_1;}{}_{b'} + 3U^{a;}{}_{cdb'}g^{cd})g^{b'b}.
\end{aligned}
$$

By using

$$q_1^{a,\kappa,b} q_{1,\kappa}^{c,d} = g^{bb'} g^{dd'} \tilde{M}^{a;}{}_{b',}{}^{c;}{}_{d'} \quad \text{and} \quad \tilde{\lambda}^{ab\kappa} q_{1,\kappa}^{c,d} = \tilde{N}^{a;}{}_{,}{}^{b;}{}_{,}{}^{c;}{}_{d'} g^{dd'},$$

we can derive the representation of the coefficients c^{abc}, A^{ab}, and c^{abcd} from Proposition 8.1. □

LEMMA 8.5. *Let $m > 0$, $M > 0$, $\gamma > 0$, and r_T be a positive sequence tending to 0 as $T \to \infty$. For any $T > 0$, let S_T and R_T be some $\mathbb{R}^p$-valued random variables, and g_T a positive matrix. Suppose that for any $K > 0$, there exist a constant $c > 0$ and a $p \times p$ positive definite matrix $\hat{g}$ such that for any $f \in \mathcal{E}(M, \gamma)$,*

$$|E[f(S_T)] - \Psi[f, g_T, \Xi_T]| \leq c\omega(f, r_T^K, \hat{g}) + \epsilon_T,$$

where ϵ_T is a sequence of constants independent of f with $\epsilon_T = o(r_T^{(m \wedge K)})$, and

$$\Psi[f, g_T, \Xi_T] = \int_{\mathbb{R}^p} dy f(y) \Xi_T(y) \phi(y; g)$$

for some polynomial Ξ_T with coefficients being bounded as $T \to \infty$. Moreover, suppose that there exist $C' > 0$ and $K' > 0$ such that

$$P[|R_T| \leq C' r_T^{K'}] = 1 - \epsilon_T',$$

where $\epsilon_T' = o(r_T^m)$. If $P(\sup_T |r_T(S_T + R_T)| < \tilde{M}) = 1$ for some constant $\tilde{M} > 0$ and $E|S_T|^q < \infty$ for some $q > 1$ (or if f is bounded), then there exist constants $\tilde{c} > 0$, $\tilde{C} > 0$, and a positive matrix $\breve{g}$ such that

$$|E[f(S_T + R_T)] - \Psi_T[f, g_T, \Xi_T]| \leq \tilde{c}\omega(f, \tilde{C}(r_T^K + r_T^{K'}), \breve{g}) + \tilde{\epsilon}_T,$$

where $\tilde{\epsilon}_T = o(r_T^{(m-\gamma)\wedge(m(q-\gamma)/q)\wedge K})$ (or $o(r_T^{m \wedge K})$).

PROOF. First, we see that

$$|E[f(S_T + R_T)] - \Psi_T[f, g_T, \Xi_T]| \leq c\omega(f, r_T^K, \hat{g}) + \epsilon_T + \Delta_1 + \Delta_2,$$

where

$$\Delta_1 = |E[(f(S_T + R_T) - f(S_T))1_{\{|R_T| \leq C' r_T^{K'}\}}]|,$$
$$\Delta_2 = |E[(f(S_T + R_T) - f(S_T))1_{\{|R_T| > C' r_T^{K'}\}}]|,$$

and 1_A is the indicator function of a set A. From the assumption concerned with S_T, it follows that

$$\Delta_1 \leq |\Psi[h, g_T, \Xi_T]| + c_0\omega(h, r_T^K, \hat{g}) + \epsilon_T,$$

where $h(y) = \sup\{|f(y + z) - f(y)| : |z| \leq C' r_T^{K'}\}$, and c_0 is a positive constant. Note that for any $\delta > 0$ and a positive matrix Σ_T converging to a positive matrix $\bar{\Sigma}$, there exist a constant $C > 0$ and a positive matrix $\hat{\Sigma}$ satisfying $\hat{\Sigma} > \bar{\Sigma}$ such that for any $y \in \mathbb{R}^p$,

$$\sup_T |y|^\delta \phi(y; \Sigma_T) \leq C\phi(y; \hat{\Sigma}).$$

Therefore we have that

$$|\Psi[h, g_T, \Xi_T]| \le \int_{\mathbb{R}^p} dy h(y)|\Xi_T(y)|\phi(y; g_T) \le c_1 \int_{\mathbb{R}^p} dy h(y)\phi(y; \hat{g}_1)$$
$$= c_1\omega(f, C' r_T^{K'}, \hat{g}_1),$$

for some constant $c_1 > 0$ and some positive definite matrix $\hat{g}_1$ satisfying $\hat{g}_1 > \lim_{T\to\infty} g_T$. Since

$$\sup\{|h(y+z_1) - h(y)| : |z_1| < r\} \le 3\sup\{|f(y+z) - f(y)| : |z| < r + C' r_T^{K'}\},$$

we obtain that

$$\omega(h, r_T^K, \hat{g}) \le 3\omega(f, r_T^K + C' r_T^{K'}, \hat{g}).$$

Thus we see that

$$\Delta_1 \le c_1\omega(f, r_T^{K'}, \hat{g}_1) + 3c_0\omega(f, r_T^K + C' r_T^{K'}, \hat{g}) + \epsilon_T.$$

If $P(\sup_T |r_T(S_T + R_T)| < \tilde{M}) = 1$ for some constant $\tilde{M} > 0$ and $\sup_T E|S_T|^q < \infty$, then it follows that

$$\Delta_2 \le M(1 + r_T^{-1}|\tilde{M}|)^\gamma P[|R_T| > C' r_T^{K'}]$$
$$+ \|M(1 + |S_T|)^\gamma\|_{q/\gamma}(P[|R_T| > C' r_T^{K'}])^{(q-\gamma)/q} = o(r_T^{(m-\gamma)\wedge(m(q-\gamma)/q)}).$$

If f is bounded, then we have $\Delta_2 = o(r_T^m)$. Thus the assertions are proved. □

From Lemma 8.5 and Proposition 8.2, we can show Theorem 6.4.

8.4 *Proof of Theorem* 6.5

Let $\tilde{\Theta}$ be an open set in $\mathbb{R}^p$. For k index sets $A_1 = a_{11}\cdots a_{1m_1}, \ldots, A_k = a_{k1}\cdots a_{km_k}$ whose elements a_{ij} run from 1 to p, let $\bar{\nu}_{A_1,\ldots,A_k} : \tilde{\Theta} \to \mathbb{R}$ be a tensor satisfying that the matrix $(\bar{\nu}_{ab})$ is negative definite for each $\theta \in \tilde{\Theta}$, that

$$\bar{\nu}_{A_1,\ldots,A_i,\ldots,A_j,\ldots,A_k}(\theta) = \bar{\nu}_{A_1,\ldots,A_j,\ldots,A_i,\ldots,A_k}(\theta) \quad \text{for} \quad i, j = 1, \ldots, k,$$

and that

$$\bar{\nu}_{A_1,\ldots,A_i,\ldots,A_k}(\theta) = \bar{\nu}_{A_1,\ldots,B_i,\ldots,A_k}(\theta) \quad \text{if} \quad (A_i) = (B_i).$$

Suppose that the matrix $(\bar{\nu}_{ab}(\theta))$, $\theta \in \tilde{\Theta}$, is non-singular, and denote by $\bar{\nu}^{ab}$ the (a,b)-element of the inverse of the matrix $(\bar{\nu}_{ab})$. For these tensors, we will assume the following conditions later.

[BI1] $\bar{\nu}_a(\theta) = 0$;

[BI2] $\bar{\nu}_{a,b}(\theta) + \bar{\nu}_{ab}(\theta) = 0$;

[BI3] $\bar{\nu}_{a,b,c}(\theta) + \sum_{(ab,c)}^{[3]} \bar{\nu}_{ab,c}(\theta) + \bar{\nu}_{abc}(\theta) = o(\frac{1}{\sqrt{T}})$;

[BI4] $\bar{\nu}_{a,b,c,d}(\theta) + \sum_{(ab,c,d)}^{[6]} \bar{\nu}_{ab,c,d}(\theta) + \sum_{(ab,cd)}^{[3]} \bar{\nu}_{ab,cd}(\theta) + \sum_{(abc,d)}^{[4]} \bar{\nu}_{abc,d}(\theta) + \bar{\nu}_{abcd}(\theta) = o(\frac{1}{\sqrt{T}})$;

[DV1] $\delta_a\bar{\nu}_{b,c}(\theta) = \bar{\nu}_{ab,c}(\theta) + \bar{\nu}_{ac,b}(\theta) + \bar{\nu}_{a,b,c}(\theta)$;

[DV2] $\delta_a\bar{\nu}_{bc,d}(\theta) = \bar{\nu}_{abc,d}(\theta) + \bar{\nu}_{bc,ad}(\theta) + \bar{\nu}_{a,bc,d}(\theta)$;

[DV3] $\delta_a\bar{\nu}_{b,c,d}(\theta) = \bar{\nu}_{ab,c,d}(\theta) + \bar{\nu}_{b,ac,d}(\theta) + \bar{\nu}_{b,c,ad}(\theta) + \bar{\nu}_{a,b,c,d}(\theta)$.

LEMMA 8.6. *Define functions* g^{ab}, g_{ab}, $L_{abc,d}$, $M_{ab,cd}$, $N_{ab,c,d}$, H_{abcd}, $\{\Gamma^{(\alpha)}_{abc}\}_{\alpha\in\mathbb{R}}$, *and* μ^a *by* $g^{ab} = \bar\nu^{aa'}\bar\nu^{bb'}\bar\nu_{a',b'}$, $(g_{ab}) = (g^{ab})^{-1}$, $L_{abc,d} = \bar\nu_{abc,d}$, $M_{ab,cd} = \bar\nu_{ab,cd} - r_T^{-2}g_{ab}g_{cd}$, $N_{ab,c,d} = \bar\nu_{ab,c,d} + r_T^{-2}g_{ab}g_{cd}$,

$$H_{abcd} = \bar\nu_{a,b,c,d} - r_T^{-2}\sum_{(ab,cd)}^{[3]} g_{ab}g_{cd}, \qquad \Gamma^{(\alpha)}_{abc} = \bar\nu_{ab,c} + \frac{1-\alpha}{2}\bar\nu_{a,b,c},$$

and $\mu^a = -\frac{1}{2}\Gamma^{(-1)}_{bca'}g^{aa'}g^{bc}$, *respectively. Suppose that* [BI2], [DV1], [DV2] *and* [DV3] *hold true. Then*

$$\begin{aligned} g_{bb'}\delta_a\mu^{b'} &= \frac{1}{2}g^{cd}g^{ef}(\Gamma^{(1)}_{abc} + \Gamma^{(-1)}_{acb})\Gamma^{(-1)}_{efd} + \frac{1}{2}g^{cd}g^{ef}(\Gamma^{(1)}_{ace} + \Gamma^{(-1)}_{aec})\Gamma^{(-1)}_{dfb} \\ &\quad - \frac{1}{2}g^{cd}(L_{acd,b} + M_{ab,cd} + N_{cd,a,b} + N_{ac,d,b} + N_{ad,b,c} + N_{ab,c,d} + H_{abcd}). \end{aligned}$$

PROOF. From the assumptions [DV2], [DV3], and the definitions of H_{abcd}, $L_{abc,d}$, $M_{ab,cd}$ and $N_{ab,c,d}$, we have

$$\delta_a\Gamma^{(-1)}_{bcd} = L_{abc,d} + M_{ad,bc} + N_{bc,a,d} + N_{ab,c,d} + N_{ac,b,d} + N_{ad,b,c} + H_{abcd}.$$

Since $g_{ab} = -\bar\nu_{ab}$ under Condition [BI2], it follows from [BI2] and [DV1] that

$$\delta_a(g^{bc}g_{cd}) = (\delta_a g^{bc})g_{cd} + g^{bc}(\bar\nu_{ac,d} + \bar\nu_{ad,c} + \bar\nu_{a,c,d}) = 0,$$

which implies that

$$\delta_a g^{bc} = -g^{bb'}g^{cc'}(\bar\nu_{ab',c'} + \bar\nu_{ac',b'} + \bar\nu_{a,b',c'}).$$

Combining these results, we obtain the desired result. □

LEMMA 8.7. *Let* g^{ab} *be the function given in Lemma 8.6. Define functions* $\tilde\mu^{a;}{}_{bc}$, $\tilde\eta^{a;}{}_{b,c}$, $U^{a;}{}_{bcd}$ *by*

$$\begin{aligned} \tilde\mu^{a;}{}_{bc}(\theta) &= \frac{1}{2}g^{aa'}(\theta)(\bar\nu_{a'b,c}(\theta) + \bar\nu_{a'c,b}(\theta) + \bar\nu_{a'bc}(\theta)), \\ \tilde\eta^{a;}{}_{b,c}(\theta) &= g^{aa'}(\theta)(\bar\nu_{a'b,c}(\theta) + \bar\nu_{a'bc}(\theta)), \\ U^{a;}{}_{bcd}(\theta) &= \frac{1}{6}g^{aa'}(\theta)\left(\bar\nu_{a'bcd}(\theta) + \sum_{(bc,d)}^{[3]}\bar\nu_{a'bc,d}(\theta)\right) + \frac{1}{3}\sum_{(cd,b)}^{[3]}\tilde\mu^{b'}{}_{cd}(\theta)\tilde\eta^{a}{}_{b',b}(\theta), \end{aligned}$$

respectively. Under Conditions [BI3] *and* [BI4], *it holds that*

$$\begin{aligned} \tilde\mu^{a}{}_{bc} &= -\frac{1}{2}g^{aa'}\Gamma^{(-1)}_{bca'} + o\left(\frac{1}{\sqrt{T}}\right), \\ \tilde\eta^{a}{}_{b,c} &= -g^{aa'}(\Gamma^{(-1)}_{bca'} + \Gamma^{(1)}_{a'cb}) + o\left(\frac{1}{\sqrt{T}}\right), \\ U^{a}{}_{bcd} &= -\frac{1}{6}g^{aa'}\left(H_{a'bcd} + \sum_{(a'b,c,d)}^{[6]} N_{a'b,c,d} + \sum_{(a'b,cd)}^{[3]} M_{a'b,cd} + L_{bcd,a'}\right) \\ &\quad + \frac{1}{6}\sum_{(cd,b)}^{[3]} g^{b'b''}\Gamma^{(-1)}_{cdb''}g^{aa'}(\Gamma^{(-1)}_{b'ba'} + \Gamma^{(1)}_{a'bb'}) + o\left(\frac{1}{\sqrt{T}}\right), \end{aligned}$$

where $\{\Gamma^{(\alpha)}\}_{\alpha\in\mathbb{R}}$, H_{abcd}, $L_{abc,d}$, $M_{ab,cd}$, $N_{ab,c,d}$ *are the functions defined in Lemma* 8.6.

PROOF. It is easy to show this lemma. See Sakamoto and Yoshida (1999) or Sakamoto (1998). □

LEMMA 8.8. *Let* g^{ab}, g_{ab}, H_{abcd}, $L_{abc,d}$, $M_{ab,cd}$, $N_{ab,c,d}$, $\{\Gamma^{(\alpha)}_{abc}\}_{\alpha\in\mathbb{R}}$, *and* μ^a *be functions defined in Lemma* 8.6. *For this* g^{ab}, *let* $\tilde{\mu}^a{}_{bc}$, $\tilde{\eta}^a{}_{b,c}$, *and* $U^a{}_{bcd}$ *be functions defined in Lemma* 8.7. *Define* $\Delta^{a;}$, $\bar{\lambda}^{abd}$, $\bar{\lambda}^{abcd}$, $\tilde{M}^{a_1;}{}_{a_2,}{}^{b_1;}{}_{b_2}$, *and* $\tilde{N}^{a_1;}{}_{,}{}^{a_2;}{}_{,}{}^{b_1;}{}_{b_2}$ *by* $\Delta^{a;} = r_T^{-2}g^{aa'}\bar{\nu}_{a'}$, $\bar{\lambda}^{abc} = g^{aa'}g^{bb'}g^{cc'}\bar{\nu}_{a',b',c'}$,

$$\bar{\lambda}^{abcd} = g^{aa'}g^{bb'}g^{cc'}g^{dd'}H_{a'b'c'd'},$$
$$\tilde{M}^{a_1;}{}_{a_2,}{}^{b_1;}{}_{b_2} = g^{a_1a_1'}g^{b_1b_1'}(M_{a_1'a_2,b_1'b_2} - \bar{\nu}_{a_1'a_2,e}g^{ef}\bar{\nu}_{b_1'b_2,f})$$

and

$$\tilde{N}^{a_1;}{}_{,}{}^{a_2;}{}_{,}{}^{b_1;}{}_{b_2} = g^{a_1a_1'}g^{a_2a_2'}g^{b_1b_1'}(N_{a_1'a_2',b_1',b_2} - \bar{\nu}_{a_1'a_2',e}g^{ef}\bar{\nu}_{b_1',b_2,f}).$$

Moreover, for these functions, let c^{abc}, c^{abcd}, *and* A^{ab} *be functions defined in Proposition* 8.3 *with* $\tilde{\lambda}^{abc} = \bar{\lambda}^{abc}$ *and* $\tilde{\lambda}^{abcd} = \bar{\lambda}^{abcd}$. *Then, under Conditions* [BI1]–[BI4] *and Conditions* [DV1]–[DV3], *it holds that*

$$c^{abc}h_{abc}\phi = -3\Gamma^{(-1/3)}_{abc}h^{abc}\phi + o\left(\frac{1}{\sqrt{T}}\right),$$
$$c^{abcd}h_{abcd}\phi = \Bigg(-3H_{abcd} - 2\sum_{(ab,c,d)}^{[6]} N_{ab,c,d} - \sum_{(abc,d)}^{[4]} L_{abc,d}$$
$$+ 4c_{abc}g_{de}(\mu^e - \tilde{\beta}^e) + 12g^{ef}(\Gamma^{(-1)}_{abe} + \Gamma^{(1)}_{aeb})\Gamma^{(-1)}_{cfd}\Bigg)h^{abcd}\phi + o\left(\frac{1}{\sqrt{T}}\right),$$
$$A^{ab}h_{ab}\phi = \Bigg(g^{cd}\tilde{M}_{ac,bc} + g_{a'a}g_{b'b}(\mu^{a'} - \tilde{\beta}^{a'})(\mu^{b'} - \tilde{\beta}^{b'})$$
$$- 2g_{bc}\delta_a\beta^c + \frac{1}{2}g^{cd}g^{ef}\Gamma^{(-1)}_{cea}\Gamma^{(-1)}_{dfb} + 2g_{bc}\delta_a\mu^c\Bigg)h^{ab}\phi + o\left(\frac{1}{\sqrt{T}}\right).$$

PROOF. From [BI3] and the definitions of $\bar{\lambda}^{abc}$, $\tilde{\mu}^{a;}{}_{bc}$ and $\Gamma^{(\alpha)}_{abc}$, it follows that

$$c^{abc}h_{abc}\phi = -3\Gamma^{(-1/3)}_{abc}h^{abc}\phi + o\left(\frac{1}{\sqrt{T}}\right).$$

In the same fashion, the definitions of the functions and the result of Lemma 8.7, which was proved under Conditions [BI3] and [BI4], yield the second result for $c^{abcd}h_{abcd}\phi$. Furthermore, from Lemma 8.6 and Lemma 8.7, the last result for $A^{ab}h_{ab}\phi$ can be shown. See Sakamoto and Yoshida (1999) or Sakamoto (1998). □

From this lemma, we can show Theorem 6.5.

Acknowledgements

We express our gratitude to the editor and the referees for careful reading and valuable comments. Also, we thank Prof. Yury Kutoyants for encouragement to publish our previous unpublished manuscript. This work was in part supported by the Research Fund for Scientists of the Ministry of Science, Education and Culture, and by Cooperative Research Program of the Institute of Statistical Mathematics.

References

Akahira, M. and Takeuchi, K. (1981). Asymptotic efficiency of statistical estimators, concepts and higher order asymptotic efficiency, *Lecture Notes in Statistics*, **7**, Springer, New York.

Barndorff-Nielsen, O. E. and Cox, D. R. (1994). *Inference and Asymptotics*, Chapman & Hall, London.

Bhattacharya, R. N. and Ghosh, J. K. (1978). On the validity of the formal Edgeworth expansion, *Annals of Statistics*, **6**, 434–451.

Bhattacharya, R. N. and Rao, R. R. (1986). *Normal Approximation and Asymptotic Expansions*, Krieger, Melbourne, Florida.

Bichteler, K., Gravereaux, J. B. and Jacod, J. (1987). *Malliavin Calculus for Processes with Jumps*, Gordon and Breach Science Publisher, New York.

Dermoune, A. and Kutoyants, Yu. A. (1995). Expansion of distribution function of maximum likelihood estimate for misspecified diffusion type observations, *Stochastics and Stochastic Reports*, **52**, 121–145.

Ghosh, J. K. (1994). *Higher Order Asymptotics*, Institute of Mathematical Statistics, California.

Götze, F. and Hipp, C. (1983). Asymptotic expansions for sums of weakly dependent random vectors, *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete*, **64**, 211–239.

Götze, F. and Hipp, C. (1994). Asymptotic distribution of statistics in time series, *Annals of Statistics*, **22**, 211–239.

Ikeda, N. and Watanabe, S. (1989). *Stochastic Differential Equations and Diffusion Processes*, 2nd ed., North-Holland/Kodansha, Tokyo.

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

Kutoyants, Yu. A. (2004). *Statistical Inference for Ergodic Diffusion Processes*, Springer, New York.

Lahiri, S. N. (1993). Refinements in the asymptotic expansions for the sums of weakly dependent random vectors, *Annals of Probability*, **21**, 791–799.

Mykland, P. A. (1992). Asymptotic expansions and bootstrapping distributions for dependent variables, a martingale approach, *Annals of Statistics*, **20**(2), 623–654.

Mykland, P. A. (1993). Asymptotic expansions for martingales, *Annals of Probability*, **21**, 800–818.

Mykland, P. A. (1995). Martingale expansions and second order inference, *Annals of Statistics*, **23**, 707–731.

Pfanzagl, J. (1982). Contributions to a general asymptotic statistical theory, *Lecture Notes in Statistics*, **13**, Springer, New York.

Pfanzagl, J. (1985). Asymptotic expansions for general statistical models, *Lecture Notes in Statistics*, **31**, Springer, Berlin.

Roberts, G. O. and Tweedie, R. L. (1996). Exponential convergence of Langevin distributions and their discrete approximations, *Bernoulli*, **2**(4), 341–363.

Sakamoto, Y. (1998). Higher order asymptotic expansions for functionals of mixing processes, ISM Cooperative Research Report, No. 111.

Sakamoto, Y. and Yoshida, N. (1996). Expansion of perturbed random variables based on generalized Wiener functionals, *Journal of Multivariate Analysis*, **59**(1), 34–59.

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

Sakamoto, Y. and Yoshida, N. (1998*b*). Third order asymptotic expansion for diffusion process, in "Theory of Statistical Analysis and Its Applications", ISM Cooperative Research Report, No. 107, 53–60.

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

Sakamoto, Y. and Yoshida, N. (2000). Asymptotic expansions of the cluster process (in preparation).

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

Stroock, D. W. (1994). *Probability Theory, an Analytic View*, Cambridge University Press, Cambridge.

Taniguchi, M. (1991). Higher order asymptotic theory for time series analysis, *Lecture Notes in Statistics*, **68**, Springer, Berlin.

Taniguchi, M. and Watanabe, Y. (1994). Statistical analysis of curved probability densities, *Journal of Multivariate Analysis*, **48**, 228–248.

Uchida, M. and Yoshida, N. (1999). Asymptotic expansion and information criteria (submitted).

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

Uchida, M. and Yoshida, N. (2004). Information criteria for small diffusions via the theory of Malliavin-Watanabe, *Statistical Inference for Stochastic Processes*, **7**, 35–67.

Yoshida, N. (1992*a*). Asymptotic expansions for small diffusions via the theory of Malliavin-Watanabe, *Probability Theory and Related Fields*, **92**, 275–311.

Yoshida, N. (1992*b*). Asymptotic expansion for statistics related to small diffusions, *Journal of the Japan Statistical Society*, **22**(2), 139–159.

Yoshida, N. (1993). Asymptotic expansion of Bayes estimators for small diffusions, *Probability Theory and Related Fields*, **95**, 429–450.

Yoshida, N. (1996*a*). Asymptotic expansion for martingales with jumps and Malliavin calculus, Research Memo., No. 601, The Institute of Statistical Mathematics, Tokyo.

Yoshida, N. (1996*b*). Asymptotic expansions for perturbed systems on Wiener space: Maximum likelihood estimators, *Journal of Multivariate Analysis*, **57**, 1–36.

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

Yoshida, N. (1999). Asymptotic expansion for martingales with jumps (preprint).

Yoshida, N. (2001). Partial mixing and conditional Edgeworth expansion for diffusions with jumps, *Probability Theory and Related Fields* (to appear).