SUT Journal of Mathematics
Vol. ??, No. ? (????), ??–??

Asymptotic expansion and information criteria

Masayuki Uchida and Nakahiro Yoshida

(Received ??? ??, ????)

Abstract. For statistical models including continuous time stochastic processes, two types of information criteria based on the expected Kullback-Leibler information are proposed. The information criteria are applied to the evaluation of various types of statistical models and they are generally different from the results proposed in Uchida and Yoshida [33], which are based on the estimated Kullback-Leibler information. As an example, we present two information criteria for ergodic diffusion processes.

AMS 2000 *Mathematics Subject Classification.* 62B10, 62E20, 62M05, 60H07.

Key words and phrases. Diffusion process, Malliavin calculus, median unbiasedness, M-estimators.

§1. Introduction

AIC is a criterion for statistical model selection proposed by Akaike [2, 3] from aspects of prediction. It evaluates the goodness of fit of a statistical model by the expected Kullback-Leibler divergence between the predicted distribution and the true distribution. The expected Kullback-Leibler divergence is, however, unobservable, and we need to estimate it from the data. In i.i.d. case, it is natural to replace this expected value by the sample mean of the log-likelihood function with the maximum likelihood estimator plugged in the unknown parameter, that is, the maximum log-likelihood divided by the total number of the data. Akaike observed that this intuitive estimator had a bias and proposed that it be corrected by the dimension of the parameter in the same asymptotic argument as in testing hypotheses. His derivation relied on the first order asymptotic theory, and required assumptions: (i) the data are independent random samples from an unknown distribution, (ii) estimation is done by the maximum likelihood method, and (iii) the parametric family of distributions includes the true model. A generalization was soon later done by Takeuchi [32]. He considered a possibly misspecified case instead of the assumption (iii), and obtained a modified criterion called TIC (Takeuchi's information criterion) under the assumptions (i) and (ii). Also, Konishi and Kitagawa [15] recently proposed generalized information criteria GIC under the assumption (i), replacing the assumption (ii) by functional-type estimators. We should note that the correction term in each information criterion is determined so as to adjust the *expectation*-bias between the estimator-plugged log-likelihood and the expected Kullback-Leibler divergence. For more details of the information criteria and related topics, see Barron et al. [4], Burman and Nolan [6], Burnham and Anderson [7], Hall [11], Hurvich and Tsai [12, 13], Konishi and Kitagawa [16], Knight [14], Laud and Ibrahim [19], Portnoy [23], Shibata [28, 29], Shimodaira [30, 31], Yang and Barron [35].

Though it is true that the expectation-unbiasedness is an easily tractable unbiasedness, from decision theoretic aspects, it does not seem to have a firm ground for selecting only it. The mean unbiasedness corresponds to a quadratic loss, and it can be extended in a natural way to other loss functions, such as absolute loss, L^p-loss, etc. In fact, as we will discuss it later as an illustrative example, it is possible to construct a median-unbiased information criterion; more generally, we will obtain 'f-unbiased' information criterion. All existing criteria including AIC, TIC and GIC modify an estimator of the expected Kullback-Leibler divergence to cancel the expectation-bias in the second order. This fact implies that the higher order asymptotic theory (asymptotic expansion) can clarify such phenomena as it was successfully applied to prove the advantages of the bootstrap method. The first order asymptotic theory to the second order terms in our language was sufficient to obtain the expectation-unbiased criteria because of a particular cancellation. However, we will treat a general unbiasedness, and in this situation, it is necessary to consider the second order approximation to the distribution of the error of an estimator of the expected Kullback-Leibler divergence. The first aim of this paper is to formulate the model selection problem in the light of the higher order asymptotic theory in a unified way, and to show several possibilities other than the usual expectation-unbiased criteria, with the median-unbiased information criterion (MUIC) as a byproduct.

It is well known that the asymptotic expansion is an indispensable tool to develop the higher order statistical inference theory. For functionals of independent observations, Bhattacharya and Ghosh [5] guaranteed the validity of the expansion, and for dependent data, Götze and Hipp [9, 10] gave a valid asymptotic expansion of the distribution of an additive functional of a discrete-time process under the geometrically strong mixing condition and a conditional type of Cramér's condition. Recently, with the Malliavin calculus, under geometrically strong mixing condition, Kusuoka and Yoshida [18] studied a valid asymptotic expansion of the distribution of an additive functional of a continuous time ϵ-Markov process with finite autoregression including Markov type semimartingales and time series models with discrete time parameter embedded in continuous time. Sakamoto and Yoshida [25, 27] applied this result to asymptotic expansions of estimators for diffusion processes. Celebrated Mykland's work [20, 21, 22] also treated higher order statistical inference from a martingale approach.

Recently, with the development of the statistical inference for stochastic processes, the problem of the model selection for stochastic processes has been becoming important both in theory and in applications to neural networks, engineering, economics and mathematical finance, etc. Thus, the second aim of this paper is to provide information criteria for stochastic processes: ϵ-Markov mixing processes with continuous time parameter including diffusion processes with jumps and also nonlinear time series models with a discrete time parameter. Consequently, our result validates the use of traditional information criteria even for stochastic processes for which mathematical validation did not necessarily exist. Moreover, it enables us to extend familiar criteria to more general criteria, such as MUIC, which are applied to stochastic processes.

The organization of this paper is as follows. In Section 2, we derive asymptotic expansions of distributions of estimators for mixing processes with a continuous time parameter under the geometrically strong mixing condition. Our main results are stated in Section 3. Two kinds of information criteria are proposed. The general results for M-estimators are also presented. In Section 4, the proposed information criteria are applied to diffusion processes. Section 5 is devoted to prove the results in Sections 2 and 3.

§2. Asymptotic expansion

2.1. Preliminaries

Let $(\Omega, \mathcal{F}, P)$ be a probability space and $(\mathcal{X}_T, \mathcal{A}_T)$ be a measurable space for each $T > 0$. Let $\boldsymbol{X}_T$ denote a $\mathcal{X}_T$-valued random variable with an unknown distribution $Q_T(\cdot) = P(\boldsymbol{X}_T^{-1}(\cdot))$ having a probability density function $q_T(\cdot)$ with respect to a reference measure. Let $\hat{\theta}_T : (\mathcal{X}_T, \mathcal{A}_T) \to \Theta$ be a measurable function, where $\Theta \subset \mathbf{R}^p$. The Borel σ-field of $\mathbf{R}^p$ is denoted by $\mathcal{B}^p$. Estimation is done within a parametric family of distributions $\{P_{T,\theta}(\cdot); \theta \in \Theta\}$ with densities $\{f_T(\cdot, \theta); \theta \in \Theta\}$, which may or may not contain $q_T(\cdot)$. The predictive density function $f_T(z, \hat{\theta}_T)$ for a future observation $\boldsymbol{X}_T(\tilde{\omega}) = z$ (for $\tilde{\omega} \in \Omega$) can be constructed by replacing the unknown parameter vector θ by $\hat{\theta}_T$.

As a model selection criterion, it is possible to use the concept of selecting a model based on minimizing the Kullback-Leibler information, where the

Kullback-Leibler information is defined by

$$
\begin{aligned}
(2.1)\qquad & I\{Q_T; P_{T,\hat{\theta}_T}\} \\
:= & \int_\Omega \log q_T(\boldsymbol{X}_T(\tilde{\omega}))P(d\tilde{\omega}) - \int_\Omega l_T(\boldsymbol{X}_T(\tilde{\omega}), \hat{\theta}_T(\boldsymbol{X}_T(\omega)))P(d\tilde{\omega}),
\end{aligned}
$$

and $l_T(x,\theta) = \log f_T(x,\theta)$. The first term on the right-hand side of (2.1) does not depend on the statistical model and only the second term may be taken into account. Uchida and Yoshida [33] proposed two information criteria as asymptotically unbiased estimators of the expected log likelihood $\int_\Omega l_T(\boldsymbol{X}_T(\tilde{\omega}), \hat{\theta}_T(\boldsymbol{X}_T(\omega)))P(d\tilde{\omega})$.

In this paper, we consider information criteria as asymptotically unbiased estimators of the mean of the expected log likelihood, which means $\int_\Omega\int_\Omega l_T(\boldsymbol{X}_T(\tilde{\omega}), \hat{\theta}_T(\boldsymbol{X}_T(\omega)))P(d\tilde{\omega})P(d\omega)$, from the point of view of selecting a statistical model based on minimizing the expected Kullback-Leibler information, which is $\int_\Omega I\{Q_T; P_{T,\hat{\theta}_T(\omega)}\}dP(\omega)$, over a set of competing models. A simple estimator of the mean of the expected log likelihood is given by $l_T(\boldsymbol{X}_T(\omega), \hat{\theta}_T(\boldsymbol{X}_T(\omega)))$. Let

$$
\begin{aligned}
S_T &= l_T(\boldsymbol{X}_T(\omega), \hat{\theta}_T(\boldsymbol{X}_T(\omega))) - \int_\Omega\int_\Omega l_T(\boldsymbol{X}_T(\tilde{\omega}), \hat{\theta}_T(\boldsymbol{X}_T(\omega)))P(d\tilde{\omega})P(d\omega), \\
\bar{S}_T^* &= r_T S_T - r_T b(\hat{\theta}_T(\boldsymbol{X}_T(\omega))),
\end{aligned}
$$

where $r_T = 1/\sqrt{T}$ and b is an $\mathbf{R}$-valued function defined on $\mathbf{R}^p$.

In order to obtain the second order asymptotic expansion of the distribution of $\bar{S}_T^*$ heuristically, we assume that there exists a parameter $\theta_0 \in \Theta$ such that

$$
(2.2)\qquad {r_T}^{-1}(\hat{\theta}_T - \theta_0) = \bar{\zeta}_T^{(0)} + r_T Q_1(\bar{\zeta}_T^{(0)}, \bar{\zeta}_T^{(1)}) + o_p(r_T)
$$

for functionals $\bar{\zeta}_T^{(0)}$, $\bar{\zeta}_T^{(1)}$ satisfying conditions put later and Q_1 is a polynomial with coefficient bounded as $T \to \infty$. Set $\bar{Z}_T^{(0)} = r_T Z_T^{(0)}$, $\bar{Z}_T^{(1)} = r_T Z_T^{(1)}$, where

$$
\begin{aligned}
Z_T^{(0)} &= l_T(\boldsymbol{X}_T(\omega), \theta_0) - \int_\Omega l_T(\boldsymbol{X}_T(\tilde{\omega}), \theta_0)P(d\tilde{\omega}), \\
Z_T^{(1)} &= \partial_\theta l_T(\boldsymbol{X}_T(\omega), \theta_0) - \int_\Omega \partial_\theta l_T(\boldsymbol{X}_T(\tilde{\omega}), \theta_0)P(d\tilde{\omega}), \quad \partial_\theta = \frac{\partial}{\partial\theta}.
\end{aligned}
$$

We expand $l_T(\boldsymbol{X}_T(\omega), \hat{\theta}_T(\boldsymbol{X}_T(\omega)))$ and $\displaystyle\int_\Omega\int_\Omega l_T(\boldsymbol{X}_T(\tilde{\omega}), \hat{\theta}_T(\boldsymbol{X}_T(\omega)))P(d\tilde{\omega})P(d\omega)$ in a Taylor series around θ_0 and substitute (2.2) in the resulting expansion. Under regularity conditions, by a central limit theorem and a law of large numbers, one has stochastic expansion as follows:

$$
S_T \;=\; Z_T^{(0)} + \bar{Z}_T^{(1)\prime}\bar{\zeta}_T^{(0)}
$$

$$
\begin{aligned}
&+\left\{r_T\int_\Omega \partial_\theta l_T(\boldsymbol{X}_T(\tilde\omega),\theta_0)P(d\tilde\omega)\right\}'\\
&\times\left\{\bar\zeta_T^{(0)}+r_T\left(Q_1(\bar\zeta_T^{(0)},\bar\zeta_T^{(1)})-\int_\Omega Q_1(\bar\zeta_T^{(0)},\bar\zeta_T^{(1)})P(d\omega)\right)\right\}\\
&+\frac{1}{2}\bar\zeta_T^{(0)\prime}\left\{r_T^2\int_\Omega(\partial_\theta)^2\, l_T(\boldsymbol{X}_T(\tilde\omega),\theta_0)P(d\tilde\omega)\right\}\bar\zeta_T^{(0)}\\
&-\frac{1}{2}\int_\Omega\bar\zeta_T^{(0)\prime}\left\{r_T^2\int_\Omega(\partial_\theta)^2\, l_T(\boldsymbol{X}_T(\tilde\omega),\theta_0)P(d\tilde\omega)\right\}\bar\zeta_T^{(0)}P(d\omega)+o_p(1),
\end{aligned}
$$

where A' indicates the transpose of A. Thus we may define the random variable R_T^* by

$$
\begin{aligned}
(2.3)\quad &\bar S_T^*\\
=\quad &\bar Z_T^{(0)}+a_T'\bar\zeta_T^{(0)}\\
&+r_T\left(\bar Z_T^{(1)\prime}\bar\zeta_T^{(0)}+P_{\mathbf{R}1^\perp}[a_T'Q_1(\bar\zeta_T^{(0)},\bar\zeta_T^{(1)})]+\frac{1}{2}P_{\mathbf{R}1^\perp}[\bar\zeta_T^{(0)\prime}b_T\bar\zeta_T^{(0)}]-b(\theta_0)\right)\\
&+R_T^*,
\end{aligned}
$$

where $a_T=r_T^2\int\partial_\theta l_T(\boldsymbol{X}_T(\tilde\omega),\theta_0)P(d\tilde\omega)$, $b_T=r_T^2\int(\partial_\theta)^2\, l_T(\boldsymbol{X}_T(\tilde\omega),\theta_0)P(d\tilde\omega)$ and $P_{\mathbf{R}1^\perp}$ is the projection to the space orthogonal to the space $\mathbf{R}1$ in $L^2(P)$, e.g., $P_{\mathbf{R}1^\perp}[a_T'Q_1(\bar\zeta_T^{(0)},\bar\zeta_T^{(1)})]=a_T'Q_1(\bar\zeta_T^{(0)},\bar\zeta_T^{(1)})-\int_\Omega a_T'Q_1(\bar\zeta_T^{(0)},\bar\zeta_T^{(1)})P(d\omega)$.

Remark 1. *(i) For an M-estimator $\hat\theta_T$, as in Sakamoto and Yoshida [24, 25, 27], we can show that for some $E_0>1$, $0<\varepsilon_0<1$ and random variables $\bar\zeta_T^{(0)}$ and $\bar\zeta_T^{(1)}$,*

$$
(2.4)\qquad {r_T}^{-1}(\hat\theta_T-\theta_0)=\bar\zeta_T^{(0)}+r_TQ_1(\bar\zeta_T^{(0)},\bar\zeta_T^{(1)})+R_T,
$$

where $P\left[\hat\theta_T\text{ exists uniquely in }U(\theta_0,r_T^{\varepsilon_0})\text{ and }|R_T|\le r_T^{\varepsilon_0}\right]=1-o(r_T^{E_0})$ and $U(\theta_0,r_T^{\varepsilon_0})$ is the closed ball of radius $r_T^{\varepsilon_0}$ centered at θ_0.

(ii) Moreover, by (i) and some regularity conditions, there exist constants $E>1$ and $\varepsilon>1$ such that

$$
\begin{aligned}
\bar S_T^*\quad=\quad &\bar Z_T^{(0)}+a_T'\bar\zeta_T^{(0)}\\
&+r_T\left(\bar Z_T^{(1)\prime}\bar\zeta_T^{(0)}+P_{\mathbf{R}1^\perp}[a_T'Q_1(\bar\zeta_T^{(0)},\bar\zeta_T^{(1)})]+\frac{1}{2}P_{\mathbf{R}1^\perp}[\bar\zeta_T^{(0)\prime}b_T\bar\zeta_T^{(0)}]-b(\theta_0)\right)\\
&+R_T^*,
\end{aligned}
$$

where $P[|R_T^*|\le r_T^{\varepsilon}]=1-o(r_T^E)$.

(iii) When $\hat\theta_T$ is the maximum likelihood estimator, it follows from $a_T=\mathbf{0}$ and some regularity conditions that there exist constants $E>1$ and $\varepsilon>1$ such that

$$
\bar S_T^*\quad=\quad\bar Z_T^{(0)}+r_T\left(\bar Z_T^{(1)\prime}\bar\zeta_T^{(0)}+\frac{1}{2}P_{\mathbf{R}1^\perp}[\bar\zeta_T^{(0)\prime}b_T\bar\zeta_T^{(0)}]-b(\theta_0)\right)+R_T^*,
$$

where $P[|R_T^*|\le r_T^{\varepsilon}]=1-o(r_T^E)$.

2.2. ϵ-Markov process

We introduce the underlying probabilistic structure of the random variables $\bar{Z}_T^{(0)}$, $\bar{Z}_T^{(1)}$, $\bar{\zeta}_T^{(0)}$ and $\bar{\zeta}_T^{(1)}$ in (2.3). For a probability space $(\Omega, \mathcal{F}, P)$, let $Y = (Y_t)_{t\in\mathbf{R}_+}$ be an $\mathbf{R}^d$-valued càdlàg process defined on Ω, and $X = (X_t)_{t\in\mathbf{R}_+}$ an $\mathbf{R}^r$-valued càdlàg process defined on Ω. Assume that $\mathcal{B}_{[0,t]}^{X,Y}$ is independent of $\mathcal{B}_{[t,\infty]}^{dX}$ for any $t \in \mathbf{R}_+$, where $\mathcal{B}_{[0,t]}^{X,Y} = \sigma[X_u, Y_u : u \in [0,t]] \vee \mathcal{N}, \mathcal{B}_{[t,\infty]}^{dX} = \sigma[X_s - X_u : s,u \in [t,\infty]]$, and $\mathcal{N}$ is the σ-field generated by null sets. For an interval $I \subset \mathbf{R}$, sub σ-fields $\mathcal{B}_I^{dX}$, $\mathcal{B}_I^Y$ and $\mathcal{B}_I$ are defined by $\mathcal{B}_I^{dX} = \sigma[X_t - X_s : s,t \in I \cap \mathbf{R}_+] \vee \mathcal{N}, \mathcal{B}_I^Y = \sigma[Y_t : t \in I \cap \mathbf{R}_+] \vee \mathcal{N}$, and $\mathcal{B}_I = \sigma[X_t - X_s, Y_t : s,t \in I \cap \mathbf{R}_+] \vee \mathcal{N}$, respectively. Suppose that there exists a constant $\epsilon \geq 0$ such that $Y_t \in \mathcal{F}\left(\mathcal{B}_{[s-\epsilon,s]}^Y \vee \mathcal{B}_{[s,t]}^{dX}\right)$ for any $s > 0$ and $t > 0$ satisfying $\epsilon \leq s \leq t$, where for any sub σ-field of $\mathcal{F}$, $\mathcal{F}(\mathcal{A})$ denotes the set of all $\mathcal{A}$-measurable functions. When a process Y meets the above condition, we call Y an ϵ-Markov process driven by X. Here, we assume that for any $T > 0$, $\bar{Z}_T \equiv (\bar{Z}_T^{(0)}, \bar{Z}_T^{(1)}, \bar{\zeta}_T^{(0)}, \bar{\zeta}_T^{(1)})$ in (2.3) is a normalized functional of an additive functionalZ_T, i.e., $\bar{Z}_T = r_T Z_T$ for an $\mathbf{R}^n$-valued process $Z = (Z_t)_{t\in\mathbf{R}_+}$ satisfying $Z_0 \in \mathcal{F}\mathcal{B}_{[0]}$ and $Z_t^s := Z_t - Z_s \in \mathcal{F}\mathcal{B}_{[s,t]}$, for every $s,t \in \mathbf{R}_+, 0 \leq s \leq t$. Note that the dimensions of $\bar{Z}_T^{(0)}$, $\bar{Z}_T^{(1)}$, $\bar{\zeta}_T^{(0)}$ and $\bar{\zeta}_T^{(1)}$ are 1, p, p and q, respectively, i.e., $n = 2p + q + 1$.

2.3. Asymptotic expansion of a functional of (2.3)

Let $\bar{Z}_T^* = (\bar{Z}_T^{(0)*}, \bar{Z}_T^{(1)*}, \bar{\zeta}_T^{(0)*}, \bar{\zeta}_T^{(1)*})$, where $\bar{Z}_T^{(0)*} = \bar{Z}_T^{(0)} + a_T'\bar{\zeta}_T^{(0)}$, $\bar{Z}_T^{(1)*} = \bar{Z}_T^{(1)}$, $\bar{\zeta}_T^{(0)*} = \bar{\zeta}_T^{(0)}$ and $\bar{\zeta}_T^{(1)*} = \bar{\zeta}_T^{(1)}$. It follows from (2.3) that $\bar{S}_T^* = \tilde{S}_T^* + R_T^*$, where

$$\begin{aligned}\tilde{S}_T^* &= \bar{Z}_T^{(0)*} + r_T\Big(\bar{Z}_T^{(1)*\prime}\bar{\zeta}_T^{(0)*} + P_{\mathbf{R}1^\perp}[a_T' Q_1(\bar{\zeta}_T^{(0)*}, \bar{\zeta}_T^{(1)*})] \\ &\qquad + \frac{1}{2}P_{\mathbf{R}1^\perp}[\bar{\zeta}_T^{(0)*\prime} b_T \bar{\zeta}_T^{(0)*}] - b(\theta_0)\Big).\end{aligned}$$

In order to obtain the asymptotic expansion of (2.3), we make the same assumptions as in Kusuoka and Yoshida [18]. [A1] is a mixing condition and [A2] is a moment condition. For details, we can refer [18] and [27].

[$A1$] There exists a positive constant a such that

$$\left\| E[f|\mathcal{B}_{[s-\epsilon,s]}^Y] - E[f] \right\|_{L^1(P)} \leq a^{-1}e^{-a(t-s)}\|f\|_\infty$$

for any $s,t \in \mathbf{R}_+$, $s \leq t$, and for any bounded $\mathcal{B}_{[t,\infty]}^Y$-measurable function f.

[$A2$] For any $S > 0$, $\sup_{t\in\mathbf{R}_+, 0\leq h\leq S}\|Z_{t+h}^t\|_{L^p(P)} < \infty$ for any $p > 1$, and $P[Z_{t+S}^t] = 0$. Moreover, $Z_0 \in \bigcap_{p>1} L^p(P)$ and $P[Z_0] = 0$.

In addition to $[A1]-[A2]$, we make another condition of the regularity of the distribution, which is also assumed in [18]. For each $T>0$, $[u(i),v(i)]$, $j=1,\dots n(T)$ are sub-intervals of $[0,T]$ 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 $\sup_{j,T}\{v(j)-u(j)\}<\infty$ and $\inf_{j,T}\{v(j)-u(j)\}\ge\tau$, where τ is a fixed constant such that $\tau>\epsilon$. Suppose 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, $E[F|\mathcal{B}_{[0,v(j)]}]=E[F|\sigma[\mathcal{Y}_j]]$, a.s. For each $j=1,\dots,n(T)$, a linear operator L_j on $\mathcal{D}(L_j)\subset\bigcap_{p>1}L^p(P)$ is a Malliavin operator over the probability space $(\Omega,\mathcal{B}_{[u(j)-\epsilon,v(j)]},P)$. The Banach space $D^{L_j}_{2,p}$, $p>1$, denotes the completion of $\mathcal{D}(L_j)$ with respect to $\|\cdot\|_{D^{L_j}_{2,p}}$, where $\|F\|_{D^{L_j}_{2,p}}=\|F\|_p+\|L_jF\|_p+\|\Gamma^{1/2}_{L_j}(F,F)\|_p$, $\Gamma_{L_j}(F,G)=L_j(FG)-FL_jG-GL_jF$, and $D^{L_j}_{2,\infty-}=\bigcap_{p\ge2}D^{L_j}_{2,p}$. Suppose that for any $f\in C^\infty_B(\mathbf{R}^{(r+d)m})$ 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 $L_jF=0$. Let $\sigma_F=(\sigma^{i,k}_F)=(\Gamma_{L_j}(F^i,F^k))$ be the Malliavin covariance σ_F of $F\in D^{L_j}_{2,\infty-}(\mathbf{R}^d)\equiv(D^{L_j}_{2,\infty-})^d$, and S_F the determinant of σ_F. Let ψ_j be a truncation functional such that $\psi_j:(\Omega,\mathcal{B}_{[u(j)-\epsilon,v(j)]})\to([0,1],\mathbf{B}([0,1]))$. Assume that $\sup_{j,T}M_j<\infty$. Set $\mathcal{Z}_j=(Z^{u(j)}_{v(j)},\mathcal{Y}_j)$, $S_1[\psi_j;\mathcal{Z}_j]=\{\sigma^{i,k}_{\mathcal{Z}_j},i,k=1,\dots,n+M_j,(S_{\mathcal{Z}_j})^{-1}(S_{\mathcal{Y}_j})^{-(n-1)}\psi_j\}$, and

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

We make the following condition of the regularity of the distribution. For details, see [18] and [27].

$[A3]$ (i) For each $j=1,\dots,n(T)$, there exists a truncation functional ψ_j defined on $(\Omega,\,\mathcal{B}_{[u(j)-\epsilon,v(j)]},P)$ such that $\inf_{j,T}P[\psi_j]>0$;

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

(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 $L^p(P)$.

Here we make the first assumption on $\bar{Z}_T$.

Assumption 1. $\bar{Z}_T=(\bar{Z}^{(0)}_T,\bar{Z}^{(1)}_T,\bar{\zeta}^{(0)}_T,\bar{\zeta}^{(1)}_T)$ *satisfies* $[A1]$–$[A3]$.

Define the k-th cumulant $\lambda^{\alpha_1\cdots\alpha_k*}_T$ of $\bar{Z}^*_T$ by

$$\lambda^{\alpha_1\cdots\alpha_k*}_T=i^{-k}\partial_{\alpha_1}\cdots\partial_{\alpha_k}\log E[e^{iu\cdot\bar{Z}^*_T}]|_{u=0},\quad \partial_\alpha=\partial/\partial u^\alpha,$$

and the Hermite polynomial $h_{\alpha_1\cdots\alpha_k}$ by

$$h_{\alpha_1\cdots\alpha_k}(z;\sigma_{\alpha\beta}) = \frac{(-1)^k}{\phi(z;\sigma_{\alpha\beta})}\partial_{\alpha_1}\cdots\partial_{\alpha_k}\phi(z;\sigma_{\alpha\beta}), \quad \partial_\alpha = \partial/\partial z^\alpha,$$

where $\phi(z;\sigma_{\alpha\beta})$ is the density function of the normal distribution with mean 0 and covariance matrix $(\sigma_{\alpha\beta})$. Let Σ_T^* be the covariance matrices $\mathrm{Cov}(\bar{Z}_T^*)$. Then the asymptotic expansions up to the second order of the density of $\bar{Z}_T^*$ itself are formally given by

$$\begin{aligned} p_{T,0}(z) &= \phi(z;\Sigma_T^*), \\ p_{T,1}(z) &= \phi(z;\Sigma_T^*)\left(1+\frac{1}{6}\lambda_T^{\alpha\beta\gamma *}h_{\alpha\beta\gamma}(z;\Sigma_T^*)\right), \end{aligned}$$

where we adopt the Einstein summation convention, and α,β,γ are indices running from 0 to $2p+q$. Divide $\mathrm{Cov}(\bar{Z}_T^*)$ corresponding to the four subvectors $\bar{Z}_T^{(0)*}$, $\bar{Z}_T^{(1)*}$, $\bar{\zeta}_T^{(0)*}$ and $\bar{\zeta}_T^{(1)*}$ of $\bar{Z}_T^*$, i.e.,

$$\Sigma_T^* = \mathrm{Cov}[\bar{Z}_T^{(0)*},\bar{Z}_T^{(1)*},\bar{\zeta}_T^{(0)*},\bar{\zeta}_T^{(1)*}] := \begin{bmatrix} \Sigma_T^{(00)*} & (\Sigma_T^{(10)*})' & (\Sigma_T^{(20)*})' & (\Sigma_T^{(30)*})' \\ \Sigma_T^{(10)*} & \Sigma_T^{(11)*} & \Sigma_T^{(12)*} & \Sigma_T^{(13)*} \\ \Sigma_T^{(20)*} & (\Sigma_T^{(12)*})' & \Sigma_T^{(22)*} & (\Sigma_T^{(32)*})' \\ \Sigma_T^{(30)*} & (\Sigma_T^{(13)*})' & \Sigma_T^{(32)*} & \Sigma_T^{(33)*} \end{bmatrix}.$$

Moreover, we set $\lambda_T^{000*} = E[(\bar{Z}_T^{(0)*})^3]$, $b^* = ((\Sigma_T^{(10)*})',(\Sigma_T^{(20)*})',(\Sigma_T^{(30)*})')'$, $b_1^* = ((\Sigma_T^{(10)*})',(\Sigma_T^{(20)*})')'$, $b_2^* = ((\Sigma_T^{(20)*})',(\Sigma_T^{(30)*})')'$, $b_3^* = (\Sigma_T^{(20)*})$, $S^* = \mathrm{Cov}[\bar{Z}_T^{(1)*},\bar{\zeta}_T^{(0)*},\bar{\zeta}_T^{(1)*}]$, $S_1^* = \mathrm{Cov}[\bar{Z}_T^{(1)*},\bar{\zeta}_T^{(0)*}]$, $S_2^* = \mathrm{Cov}[\bar{\zeta}_T^{(0)*},\bar{\zeta}_T^{(1)*}]$, $S_3^* = \Sigma_T^{(22)*}$. For $M>0$ and $\gamma>0$, the set $\mathcal{E}(M,\gamma)$ of measurable functions from $\mathbf{R}\to\mathbf{R}$ is defined by

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

For any $f\in\mathcal{E}(M,\gamma)$ and $r>0$ and $\hat{\Sigma}^{(00)*}>0$ satisfying $\hat{\Sigma}^{(00)*}>\lim_{T\to\infty}\Sigma_T^{(00)*}$, let

$$\omega(f,r) = \int_{\mathbf{R}} \sup\{|f(x+y)-f(x)| : |y|\le r\}\phi(x;\hat{\Sigma}^{(00)*})dx.$$

We make another assumption.

Assumption 2. *There exist constants $K'>0$ and $\alpha>0$ such that*

$$\sup_{f\in\mathcal{E}(M,\gamma)}\left|E[(f(\bar{S}_T^*)-f(\tilde{S}_T^*))1_{\{|R_T^*|>r_T^{K'}\}}]\right| = o(r_T^\alpha),$$

where 1_A is the indicator function of a set A.

Remark 2. *Suppose that there exist constants* $K' > 1$ *and* $m > 1$ *such that* $P[|R_T^*| \leq r_T^{K'}] = 1 - o(r_T^m)$. *Moreover, suppose that* $\sup_{T>1} ||r_T \bar{S}_T^*||_{L^p} < \infty$ *for some* $p > 1$, *and that* $m(p-1)/p - \gamma > 1$. *Then, we can show that there exists a constant* $\alpha > 1$ *such that* $\sup_{f \in \mathcal{E}(M,\gamma)} \left| E[(f(\bar{S}_T^*) - f(\tilde{S}_T^*))1_{\{|R_T^*| > r_T^{K'}\}}] \right| = o(r_T^{\alpha})$.

The second order asymptotic expansion of a functional of (2.3) is as follows.

Theorem 1. *Let* $M, \gamma > 0$. *Suppose that Assumptions 1 and 2 hold true. Then there exist constants* $\delta > 0$ *and* $\tilde{c} > 0$ *such that for any function* $f \in \mathcal{E}(M,\gamma)$,

$$
\begin{aligned}
E[f(\bar{S}_T^*)] \;=\; & \int_{\mathbf{R}} f(z^{(0)}) \phi(z^{(0)}; \Sigma_T^{(00)*}) dz^{(0)} \\
& + \frac{1}{6} \lambda_T^{000*} \int_{\mathbf{R}} f(z^{(0)}) h_3(z^{(0)}; \Sigma_T^{(00)*}) \phi(z^{(0)}; \Sigma_T^{(00)*}) dz^{(0)} \\
& - r_T \int_{\mathbf{R}} f(z^{(0)}) \partial_{z^{(0)}} \Big[\Big\{ C_T^{(1)}(z^{(0)}) + C_T^{(2)}(z^{(0)}) \\
& + \frac{1}{2} C_T^{(3)}(z^{(0)}) - b(\theta_0) \Big\} \phi(z^{(0)}; \Sigma_T^{(00)*}) \Big] dz^{(0)} \\
+ \; & \rho_T(f),
\end{aligned}
$$

where

$$
\begin{aligned}
\rho_T(f) &= \tilde{c}\omega(f, 2r_T^{K'}) + o(r_T^{((1+\delta)\wedge\alpha)}), \\
C_T^{(1)}(z^{(0)}) &= \frac{(\Sigma_T^{(10)*})' \Sigma_T^{(20)*}}{(\Sigma_T^{(00)*})^2} [(z^{(0)})^2 - \Sigma_T^{(00)*}] + tr \Sigma_T^{(12)*}, \\
C_T^{(2)}(z^{(0)}) &= \int_{\mathbf{R}^{p+q}} a_T' Q_1(z^{(2)}, z^{(3)}) \\
& \quad \times \phi(z^{(2)}, z^{(3)}; b_2^*(\Sigma_T^{(00)*})^{-1} z^{(0)}, S_2^* - b_2^*(\Sigma_T^{(00)*})^{-1}(b_2^*)') dz^{(2)} dz^{(3)} \\
& \quad - \int_{\Omega} a_T' Q_1(\bar{\zeta}_T^{(0)*}, \bar{\zeta}_T^{(1)*}) P(d\omega), \\
C_T^{(3)}(z^{(0)}) &= \frac{(\Sigma_T^{(20)*})' b_T \Sigma_T^{(20)*}}{(\Sigma_T^{(00)*})^2} [(z^{(0)})^2 - \Sigma_T^{(00)*}].
\end{aligned}
$$

Remark 3. *From Theorem 1 with some* $K' > 1$ *and* $\alpha > 1$, *it follows that for adequate measurable functions* f *satisfying that* $f \in \mathcal{E}(M,\gamma)$,

$$
\begin{aligned}
E[f(\bar{S}_T^*)] \;=\; & \int_{\mathbf{R}} f(z^{(0)}) \phi(z^{(0)}; \Sigma_T^{(00)*}) dz^{(0)} \\
& + \frac{1}{6} \lambda_T^{000*} \int_{\mathbf{R}} f(z^{(0)}) h_3(z^{(0)}; \Sigma_T^{(00)*}) \phi(z^{(0)}; \Sigma_T^{(00)*}) dz^{(0)}
\end{aligned}
$$

$$
\begin{aligned}
&-r_T \int_{\mathbf{R}} f(z^{(0)}) \partial_{z^{(0)}} \left[\left\{ C_T^{(1)}(z^{(0)}) + C_T^{(2)}(z^{(0)}) \right.\right. \\
&\left.\left. + \frac{1}{2} C_T^{(3)}(z^{(0)}) \right\} \phi(z^{(0)}; \Sigma_T^{(00)*}) \right] dz^{(0)} \\
&+ r_T b(\theta_0) \int_{\mathbf{R}} f(z^{(0)}) \partial_{z^{(0)}} \left[\phi(z^{(0)}; \Sigma_T^{(00)*}) \right] dz^{(0)} \\
&+ o(r_T).
\end{aligned}
$$

For $M > 0$ and $\gamma > 0$, the set $\mathcal{E}'(M,\gamma)$ of measurable functions $\mathbf{R} \to \mathbf{R}$ is defined by

$$
\begin{aligned}
\mathcal{E}'(M,\gamma) \;\; = \;\; & \left\{ f \in \mathcal{E}(M,\gamma) \;\middle|\; \int_{\mathbf{R}} f(x)\phi(x;\Sigma_T^{(00)*})dx = 0, \right. \\
& \left. \int_{\mathbf{R}} f(x)\partial_x\{\phi(x;\Sigma_T^{(00)*})\}dx \neq 0 \right\}.
\end{aligned}
$$

For suitable measurable functions f satisfying that $f \in \mathcal{E}'(M,\gamma)$, let

$$
IC_f(\boldsymbol{X}_T(\omega)) = r_T l_T(\boldsymbol{X}_T(\omega), \hat{\theta}_T(\boldsymbol{X}_T(\omega))) - r_T b_f(\hat{\theta}_T(\boldsymbol{X}_T(\omega))),
$$

where

$$
\begin{aligned}
b_f(\theta_0) \;\; = \;\; & -\left[\int_{\mathbf{R}} f(z^{(0)}) \partial_{z^{(0)}} \left\{ \phi(z^{(0)}; \Sigma_T^{(00)*}) \right\} dz^{(0)} \right]^{-1} \\
& \times \left[r_T^{-1} \frac{1}{6} \lambda_T^{000*} \int_{\mathbf{R}} f(z^{(0)}) h_3(z^{(0)}; \Sigma_T^{(00)*}) \phi(z^{(0)}; \Sigma_T^{(00)*}) dz^{(0)} \right. \\
& - \int_{\mathbf{R}} f(z^{(0)}) \partial_{z^{(0)}} \left[\left\{ C_T^{(1)}(z^{(0)}) + C_T^{(2)}(z^{(0)}) \right.\right. \\
& \left.\left.\left. + \frac{1}{2} C_T^{(3)}(z^{(0)}) \right\} \phi(z^{(0)}; \Sigma_T^{(00)*}) \right] dz^{(0)} \right].
\end{aligned}
$$

Then, IC_f is the f-unbiased information criterion. In particular, for $f(x) = x$, we obtain the asymptotically expectation-unbiased information criterion. Moreover, for $f(x) = 1_{(-\infty,0)}(x) - \frac{1}{2}$ and $f(x) = 1_{(0,\infty)}(x) - \frac{1}{2}$, we also have the second order asymptotically median-unbiased information criterion. The details will be described in next section.

§3. Information criteria

3.1. Main results

First, we propose an information criterion based on the asymptotically expectation-unbiasedness (AEU) as follows.

Theorem 2 (Information criterion in the sense of AEU). *Suppose that Assumptions 1 and 2 for some $K' > 1$ and $\alpha > 1$ hold true. Let*

$$IC_1(\boldsymbol{X}_T(\omega)) = r_T l_T(\boldsymbol{X}_T(\omega), \hat{\theta}_T(\boldsymbol{X}_T(\omega))) - r_T b_1(\hat{\theta}_T(\boldsymbol{X}_T(\omega))), \tag{3.1}$$

where $b_1(\theta_0) = tr\Sigma_T^{(12)*}$. Then,

$$E\left[IC_1(\boldsymbol{X}_T(\omega)) - r_T \int_\Omega \int_\Omega l_T(\boldsymbol{X}_T(\tilde{\omega}), \hat{\theta}_T(\boldsymbol{X}_T(\omega)))P(d\tilde{\omega})P(d\omega)\right] = o(r_T).$$

Remark 4. *(i) As for Theorem 2, we do not need to suppose [A3] in Assumption 1. (ii) When the data are independent random samples, IC_1 defined by (3.1) can be reduced to AIC, TIC and GIC by maximum likelihood estimator in parametric model, by maximum likelihood estimator in misspecified model, and by functional-type estimator in misspecified model, respectively.*

Next, instead of IC_1, we suggest another information criterion based on the second order asymptotically median-unbiasedness (second order AMU). For details of the second order AMU, see Akahira and Takeuchi [1].

Theorem 3 (Information criterion in the sense of AMU). *Suppose that Assumptions 1 and 2 for some $K' > 1$ and $\alpha > 1$ hold true. Let*

$$IC_2(\boldsymbol{X}_T(\omega)) = r_T l_T(\boldsymbol{X}_T(\omega), \hat{\theta}_T(\boldsymbol{X}_T(\omega))) - r_T b_2(\hat{\theta}_T(\boldsymbol{X}_T(\omega))), \tag{3.2}$$

where

$$\begin{aligned} b_2(\theta_0) &= -\frac{1}{6}{r_T}^{-1}\lambda_T^{000*}\frac{1}{\Sigma_T^{(00)*}} \\ &\quad + \left[tr\Sigma_T^{(12)*} - \frac{(\Sigma_T^{(10)*})'\Sigma_T^{(20)*}}{\Sigma_T^{(00)*}} + C_T^{(2)}(0) - \frac{1}{2}\frac{(\Sigma_T^{(20)*})'b_T\Sigma_T^{(20)*}}{\Sigma_T^{(00)*}}\right]. \end{aligned}$$

Then,

$$\begin{aligned} & P\left[IC_2(\boldsymbol{X}_T(\omega)) - r_T \int_\Omega \int_\Omega l_T(\boldsymbol{X}_T(\tilde{\omega}), \hat{\theta}_T(\boldsymbol{X}_T(\omega)))P(d\tilde{\omega})P(d\omega) > 0\right] \\ = \; & \frac{1}{2} + o\left(r_T\right) \end{aligned}$$

and

$$\begin{aligned} & P\left[IC_2(\boldsymbol{X}_T(\omega)) - r_T \int_\Omega \int_\Omega l_T(\boldsymbol{X}_T(\tilde{\omega}), \hat{\theta}_T(\boldsymbol{X}_T(\omega)))P(d\tilde{\omega})P(d\omega) < 0\right] \\ = \; & \frac{1}{2} + o\left(r_T\right). \end{aligned}$$

3.2. Results for M-estimators

First we review the result of M-estimator given in Sakamoto and Yoshida [27].

Let Θ be an open bounded convex set included in $\mathbf{R}^p$ and T_0 a positive constant. For $T > T_0$, let $(\mathcal{X}_T, \mathcal{A}_T)$ be a measurable space and X_T an $\mathcal{X}_T$-valued random variable on some probability space. For an estimating function $\psi_T : \mathcal{X}_T \times \Theta \to \mathbf{R}^d$, an M-estimator $\hat{\theta}_T$ is defined as a solution of the estimating equation $\psi_T(\theta, X_T) = 0$. Let $\theta_0 \in \Theta$. It follows from assumptions below that θ_0 is the target of M-estimator $\hat{\theta}_T$, and θ_0 is called the (quasi) true value in the parameter space Θ. Set $\psi(\theta) = \psi_T(\theta, X_T)$ and let $\psi_{a;}(\theta)$ be the a-th element of $\psi(\theta)$, and $(\psi_{a;a_1,\ldots,a_k}(\theta))_{a_1,\ldots,a_k=1,\ldots,p}$ the k-th derivatives of $\psi_{a;}(\theta)$ with respect to $\theta_{a_1}, \ldots, \theta_{a_k}$, i.e., $\psi_{a;a_1,\ldots,a_k}(\theta) = \delta_{a_1} \cdots \delta_{a_k} \psi_{a;}(\theta)$, where $\delta_a = \partial/\partial\theta_a$. Let $\bar{\nu}_{a;a_1\cdots a_k}(\theta)$ be a tensor defined on Θ such that it is symmetric in $a_1, \ldots, a_k$. Note that they may depend on T. However, we assume that they are bounded as $T \to \infty$. For $K \in \mathbf{N}$, $q > 1$, $\gamma > 0$ and a positive bounded sequence r_T satisfying that $r_T \to 0$ as $T \to \infty$, we make the following assumptions.
$[\mathrm{C0}]^K$ $\psi \in C^K(\Theta)$ a.s.
$[\mathrm{C1}]_q$ $\sup_{T>T_0} \|r_T \psi_{a;}(\theta_0)\|_q < \infty$ for $a = 1, \ldots, p$.
$[\mathrm{C2}]^K_{q,\gamma}$

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

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

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

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

As in the case that the dimension of θ is one, the second order stochastic expansion of an M-estimator was obtained by Sakamoto and Yoshida [24] under similar conditions to those given above. Let $\bar{\nu}_{a;a_1,\ldots,a_k}(\theta) = r_T^2 \mathrm{E}[\psi_{a;a_1,\ldots,a_k}(\theta)]$.

Theorem 4 (Sakamoto and Yoshida [27]). *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*

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

From this theorem, we see that for any $m > 0$, there exists a subspace $\tilde{\mathcal{X}}_T$ such that $P(\tilde{\mathcal{X}}_T) = 1 - o(r_T^m)$ and that for each observation $X_T \in \tilde{\mathcal{X}}_T$, 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 $\mathcal{X}_T$ will be referred to as the M-estimator of θ_0, and will be also denoted by $\hat{\theta}_T$.

Let $\{\bar{\nu}_{a;}(\theta)\}_{a=1,\ldots,p}$ 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 $Z_{a;} = r_T^{-1}(r_T^2\psi_{a;}(\theta_0) - \bar{\nu}_{a;}(\theta_0))$ and $Z_{a;b} = r_T^{-1}(r_T^2\psi_{a;b}(\theta_0) - \bar{\nu}_{a;b}(\theta_0))$. Moreover, under the condition [C3], set $Z^{a;} = -\bar{\nu}^{a;a'}Z_{a';}$, ${Z^{a;}}_b = -\bar{\nu}^{a;a'}Z_{a';b}$, ${\bar{\nu}^{a;}}_{a_1\cdots a_k} = -\bar{\nu}^{a;a'}\bar{\nu}_{a';a_1\cdots a_k}$, and $\Delta^{a;} = -\bar{\nu}^{a;a'}\Delta_{a';}$. We then obtain the stochastic expansion of M-estimator $\hat{\theta}_T$ and estimate the convergence rate of its remainder term.

Theorem 5 (Sakamoto and Yoshida [27]). *Let $m > 0$ and $\gamma \in (0,1)$. Suppose that $[C0]^3$, $[C1]_{p_1}$, $[C2]^k_{p_2,\gamma}$, $k = 1,2$, $[C3]$, and $[C4]^3_{p_3}$ hold true for some $p_1 > 3m$, $p_2 > \max(p,3m)$, $p_3 > \max(1,m)$ satisfying that $2/3 + \max(m/p_2, m/(3p_3)) < \gamma < 1 - m/p_1$. 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, let R_2^a be defined by*

$$
\begin{aligned}
(3.3)\qquad r_T^{-1}(\hat{\theta}_T - \theta_0)^a &= Z^{a;} + r_T({Z^{a;}}_b Z^{b;} + \frac{1}{2}{\bar{\nu}^{a;}}_{bc}Z^{b;}Z^{c;} + \Delta^{a;}) \\
&\quad + r_T^2 R_2^a.
\end{aligned}
$$

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

$$
P\left[r_T|R_2^a| \le C r_T^{\varepsilon}, a = 1,\ldots,p\right] = 1 - o(r_T^m).
$$

Next, we obtain $Q_1(\bar{\zeta}_T^{(0)*}, \bar{\zeta}_T^{(1)*})$ and $C_T^{(2)}(0)$ explicitly in order to apply Theorem 3 to the case of an M-estimator. Let $\bar{\zeta}_T^{(0)*} = (\bar{\zeta}_{T,1}^{(0)*}, \bar{\zeta}_{T,2}^{(0)*}, \ldots, \bar{\zeta}_{T,p}^{(0)*})'$, $\bar{\zeta}_T^{(1)*} = ((\bar{\zeta}_{T,1}^{(1)*})', (\bar{\zeta}_{T,2}^{(1)*})', \ldots, (\bar{\zeta}_{T,p}^{(1)*})')'$, $\bar{\zeta}_{T,a}^{(0)*} = Z^{a;}$, $\bar{\zeta}_{T,a}^{(1)*} = ({Z^{a;}}_1, {Z^{a;}}_2, \ldots, {Z^{a;}}_p)'$.

From (2.4) and (3.3), we have

$$
\begin{aligned}
(3.4)\qquad Q_1^a(\bar{\zeta}_T^{(0)*}, \bar{\zeta}_T^{(1)*}) &= {Z^{a;}}_b Z^{b;} + \frac{1}{2}{\bar{\nu}^{a;}}_{bc}Z^{b;}Z^{c;} + \Delta^{a;} \\
&= (\bar{\zeta}_{T,a}^{(1)*})'\bar{\zeta}_T^{(0)*} + \frac{1}{2}(\bar{\zeta}_T^{(0)*})'c_{T,a}\bar{\zeta}_T^{(0)*} + \Delta^{a;},
\end{aligned}
$$

where $c_{T,a}$ is the matrix for $a = 1,\ldots,p$, i.e., $c_{T,a} = ({\bar{\nu}^{a;}}_{bc})_{b,c=1,\ldots,p}$.

Let $\mathrm{Cov}[\bar{\zeta}_T^{(1)*}, \bar{Z}_T^{(0)*}] = \left((\Sigma_{T,1}^{(30)*})', \ldots, (\Sigma_{T,p}^{(30)*})'\right)'$, $\Sigma_{T,i}^{(30)*} = \mathrm{Cov}[\bar{\zeta}_{T,i}^{(1)*}, \bar{Z}_T^{(0)*}]$, $\mathrm{Cov}[\bar{\zeta}_T^{(1)*}, \bar{\zeta}_T^{(0)*}] = \left(\Sigma_{T,1}^{(32)*}, \ldots, \Sigma_{T,p}^{(32)*}\right)'$ and $\Sigma_{T,i}^{(32)*} = \mathrm{Cov}[\bar{\zeta}_{T,i}^{(1)*}, \bar{\zeta}_T^{(0)*}]$. From Theorem 1 and (3.4), we have

$$
\begin{aligned}
C_T^{(2)}(z^{(0)}) &= \sum_{i=1}^{p} a_{T,i}\left[C_{T,1,i}^{(2)}(z^{(0)}) + \frac{1}{2}C_{T,2,i}^{(2)}(z^{(0)}) + \Delta^{i;}\right] \\
&\quad - \int_\Omega a_T' Q_1(\bar{\zeta}_T^{(0)*}, \bar{\zeta}_T^{(1)*})P(d\omega),
\end{aligned}
$$

where

$$
\begin{aligned}
C_{T,1,i}^{(2)}(z^{(0)}) &= \int_{\mathbf{R}^{p+q}} (z_i^{(3)})' z^{(2)} \phi(z^{(2)}, z^{(3)}; \mu_3, \Sigma_3) dz^{(2)} dz^{(3)}, \\
C_{T,2,i}^{(2)}(z^{(0)}) &= \int_{\mathbf{R}^{p}} (z^{(2)})' c_{T,i} z^{(2)} \phi(z^{(2)}; \mu_2, \Sigma_2) dz^{(2)}.
\end{aligned}
$$

Here, $\mu_2 = \Sigma_T^{(20)*}(\Sigma_T^{(00)*})^{-1} z^{(0)}$ and $\Sigma_2 = \Sigma_T^{(22)*} - (\Sigma_T^{(00)*})^{-1}\Sigma_T^{(20)*}(\Sigma_T^{(20)*})'$ and

$$
\begin{aligned}
\mu_3 &= \begin{bmatrix} \Sigma_T^{(20)*} \\ \Sigma_T^{(30)*} \end{bmatrix} (\Sigma_T^{(00)*})^{-1} z^{(0)}, \\
\Sigma_3 &= \begin{bmatrix} \Sigma_T^{(22)*} - (\Sigma_T^{(00)*})^{-1}\Sigma_T^{(20)*}(\Sigma_T^{(20)*})' & (\Sigma_T^{(32)*})' - (\Sigma_T^{(00)*})^{-1}\Sigma_T^{(20)*}(\Sigma_T^{(30)*})' \\ \Sigma_T^{(32)*} - (\Sigma_T^{(00)*})^{-1}\Sigma_T^{(30)*}(\Sigma_T^{(20)*})' & \Sigma_T^{(33)*} - (\Sigma_T^{(00)*})^{-1}\Sigma_T^{(30)*}(\Sigma_T^{(30)*})' \end{bmatrix}.
\end{aligned}
$$

Since it follows that

$$
\begin{aligned}
C_T^{(2)}(z^{(0)}) &= \frac{(\sum_{i=1}^{p} a_{T,i}\Sigma_{T,i}^{(30)*})'\Sigma_T^{(20)*}}{(\Sigma_T^{(00)*})^2}[(z^{(0)})^2 - \Sigma_T^{(00)*}] \\
&\quad + \frac{1}{2}\frac{(\Sigma_T^{(20)*})'(\sum_{i=1}^{p} a_{T,i} c_{T,i})\Sigma_T^{(20)*}}{(\Sigma_T^{(00)*})^2}[(z^{(0)})^2 - \Sigma_T^{(00)*}],
\end{aligned}
$$

One has

$$
\begin{aligned}
C_T^{(2)}(0) &= -\frac{(\sum_{i=1}^{p} a_{T,i}\Sigma_{T,i}^{(30)*})'\Sigma_T^{(20)*}}{(\Sigma_T^{(00)*})} \\
&\quad - \frac{1}{2}\frac{(\Sigma_T^{(20)*})'(\sum_{i=1}^{p} a_{T,i} c_{T,i})\Sigma_T^{(20)*}}{(\Sigma_T^{(00)*})}.
\end{aligned} \tag{3.5}
$$

§4. Applications to ergodic diffusions

We consider applications of the results presented in Section 2 to ergodic diffusion processes.

Let $\boldsymbol{X}_T = \{X_t; 0 \leq t \leq T\}$ be a d-dimensional diffusion process defined by the stochastic differential equation (true model)

$$\begin{array}{rcl} dX_t & = & V_0(X_t)dt + V(X_t)dw_t, \quad t \in [0,T], \\ X_0 & = & x_0, \end{array} \tag{4.1}$$

where X_0 is the initial random variable (r.v.), $V = (V_1, \cdots, V_r)$ is an $\mathbf{R}^d \otimes \mathbf{R}^r$ valued smooth function defined on $\mathbf{R}^d$, V_0 is an $\mathbf{R}^d$-valued smooth function defined on $\mathbf{R}^d$ with bounded x-derivatives and w is an r-dimensional standard Wiener process. We assume that X_t is a stationary, strong mixing diffusion process and X_0 obeys the stationary distribution ν satisfying $\nu(|x|^p) < \infty$ for any $p > 1$.

Consider a d-dimensional diffusion model defined by the stochastic differential equation

$$\begin{array}{rcl} dX_t & = & \tilde{V}_0(X_t,\theta)dt + \tilde{V}(X_t)d\tilde{w}_t, \quad t \in [0,T], \\ X_0 & = & x_0, \end{array} \tag{4.2}$$

where θ is a p-dimensional unknown parameter in Θ, X_0 is the initial r.v., $\tilde{V} = (\tilde{V}_1, \cdots, \tilde{V}_{\tilde{r}})$ is an $\mathbf{R}^d \otimes \mathbf{R}^{\tilde{r}}$ valued smooth function defined on $\mathbf{R}^d$, $\tilde{V}_0$ is an $\mathbf{R}^d$-valued smooth function defined on $\mathbf{R}^d \times \Theta$ and $\tilde{w}$ is an $\tilde{r}$-dimensional standard Wiener process. The unknown parameter θ needs to be estimated from the observation $\boldsymbol{X}_T = \{X_t; 0 \leq t \leq T\}$.

Let $\boldsymbol{X}_T^\theta$ be the solution of the stochastic differential equation (4.2) for θ. We assume that $\boldsymbol{X}_T^\theta$ is a stationary, strong mixing diffusion process with a stationary distribution ν_θ. Since the likelihood function of θ is defined by

$$\begin{array}{rcl} L_T(\boldsymbol{X}_T^\theta, \theta) & := & \dfrac{d\nu_\theta(X_0)}{dx} \exp\left\{ \displaystyle\int_0^T \tilde{V}_0'(\tilde{V}\tilde{V}')^{-1}(X_t,\theta)dX_t \right. \\ & & \left. -\dfrac{1}{2}\displaystyle\int_0^T \tilde{V}_0'(\tilde{V}\tilde{V}')^{-1}\tilde{V}_0(X_t,\theta)dt \right\}, \end{array}$$

the log likelihood function is given by

$$l_T(\boldsymbol{X}_T^\theta, \theta) = \tilde{a}(X_0,\theta) + \int_0^T \tilde{b}(X_t,\theta)dX_t + \int_0^T \tilde{c}(X_t,\theta)dt, \tag{4.3}$$

where we set that $\tilde{a}(x,\theta) = \log \frac{d\nu_\theta(x)}{dx}$, $\tilde{b}(x,\theta) = \tilde{V}_0'(\tilde{V}\tilde{V}')^{-1}(x,\theta)$ and $\tilde{c}(x,\theta) = -\frac{1}{2}\tilde{V}_0'(\tilde{V}\tilde{V}')^{-1}\tilde{V}_0(x,\theta)$.

From (4.1) and (4.3), the log likelihood function under the true model is given by

$$l_T(\boldsymbol{X}_T, \theta) = a(X_0,\theta) + \int_0^T b(X_t,\theta)dw_t + \int_0^T c(X_t,\theta)dt, \tag{4.4}$$

where $a(x,\theta)=\tilde{a}(x,\theta)$, $b(x,\theta)=\tilde{b}(x,\theta)V(x)$ and $c(x,\theta)=\tilde{c}(x,\theta)+\tilde{b}(x,\theta)V_0(x)$.

Define a functional Ψ_T by

$$\Psi_T(\boldsymbol{X}_T^\theta,\theta):=\tilde{A}(X_0,\theta)+\int_0^T\tilde{B}(X_t,\theta)dX_t+\int_0^T\tilde{C}(X_t,\theta)dt, \tag{4.5}$$

where $\tilde{A}$, $\tilde{B}$ and $\tilde{C}$ are given functions. From (4.1) and (4.5), the functional Ψ_T under the true model is given by

$$\Psi_T(\boldsymbol{X}_T,\theta)=A(X_0,\theta)+\int_0^T B(X_t,\theta)dw_t+\int_0^T C(X_t,\theta)dt,$$

where $A(x,\theta)=\tilde{A}(x,\theta)$, $B(x,\theta)=\tilde{B}(x,\theta)V(x)$ and $C(x,\theta)=\tilde{C}(x,\theta)+\tilde{B}(x,\theta)V_0(x)$.

Let $\hat{\theta}_T$ be the M-estimator defined as a solution of the estimating equation

$$\partial_\theta\Psi_T(\boldsymbol{X}_T,\theta)=0. \tag{4.6}$$

Definition 1. *Define the quasi true parameter θ_0 as a solution of the equation*

$$\int_{\mathbf{R}^d}\partial_\theta C(x,\theta)\nu(dx)=0.$$

Under regularity conditions in Theorem 5, we can validate the following argument.

Since it follows from the definition of $\hat{\theta}_T$ that

$$r_T^{-1}(\hat{\theta}_T-\theta_0)^a \;=\; Z^{a;}+r_T(Z^{a;}{}_bZ^{b;}+\frac{1}{2}\bar{\nu}^{a;}{}_{bc}Z^{b;}Z^{c;}+\Delta^{a;})+o_p(r_T),$$

we define

$$\begin{aligned}
\bar{\zeta}_T^{(0)} &= (Z^{1;},Z^{2;},\ldots,Z^{p;})' \\
&= -\nu\left((\partial_\theta)^2C(\cdot,\theta_0)\right)^{-1} \\
&\quad\times\left[r_T\int_0^T\partial_\theta B(X_t,\theta_0)dw_t+r_T\int_0^T\partial_\theta C(X_t,\theta_0)dt\right],
\end{aligned} \tag{4.7}$$

$$\begin{aligned}
\bar{\zeta}_{T,i}^{(1)} &= (Z^{i;}{}_1,Z^{i;}{}_2,\ldots,Z^{i;}{}_p)' \\
&= -\left[r_T\int_0^T(\partial_\theta)^2B(X_t,\theta_0)dw_t\right. \\
&\quad\left.+r_T\int_0^T\left\{(\partial_\theta)^2C(X_t,\theta_0)-\nu\left((\partial_\theta)^2C(\cdot,\theta_0)\right)\right\}dt\right] \\
&\quad\times\nu\left((\partial_\theta)^2C(\cdot,\theta_0)\right)^{-1}_{\cdot i},
\end{aligned} \tag{4.8}$$

where $\nu_{\cdot i}^{-1}$ is the i-th column vector of ν^{-1}.

$\bar{Z}_T^{(0)}$ and $\bar{Z}_T^{(1)}$ are given by

$$\begin{aligned}\bar{Z}_T^{(0)} &= r_T\left[\alpha_1(X_0,\theta_0) + \int_0^T b(X_t,\theta_0)dw_t \right.\\ &\quad \left. + \int_0^T \{c(X_t,\theta_0) - \nu\left(c(\cdot,\theta_0)\right)\}dt\right],\end{aligned}\tag{4.9}$$

$$\begin{aligned}\bar{Z}_T^{(1)} &= r_T\left[\alpha_2(X_0,\theta_0) + \int_0^T \partial_\theta b(X_t,\theta_0)dw_t \right.\\ &\quad \left. + \int_0^T \left\{\partial_\theta c(X_t,\theta_0) - \nu\left(\partial_\theta c(\cdot,\theta_0)\right)\right\} dt\right],\end{aligned}\tag{4.10}$$

where $\alpha_1(X_0,\theta_0) = a(X_0,\theta_0) - \nu\left(a(\cdot,\theta_0)\right)$ and $\alpha_2(X_0,\theta_0) = \partial_\theta a(X_0,\theta_0) - \nu\left(\partial_\theta a(\cdot,\theta_0)\right)$.

For functions f satisfying $\nu(f) = 0$, denote by G_f the Green function such that $\mathcal{A}G_f = f$, where $V_0 = (V_0^i)$, $V = (V_j^i)$ and

$$\mathcal{A} = \sum_{i=1}^d V_0^i \partial_i + \frac{1}{2}\sum_{i,j}^d \sum_{\alpha=1}^r V_\alpha^i V_\alpha^j \partial_i\partial_j, \quad \partial_i = \frac{\partial}{\partial x^i}.$$

From Itô's formula, we see

$$G_f(X_T) - G_f(X_0) = \int_0^T \partial G_f(X_t)V(X_t)dw_t + \int_0^T f(X_t)dt.\tag{4.11}$$

Define f_i $(i = 0,1,2,3)$ by $f_0(x) = c(x,\theta_0) - \nu(c(\cdot,\theta_0))$, $f_1(x) = \partial_\theta c(x,\theta_0) - \nu\left(\partial_\theta c(\cdot,\theta_0)\right)$, $f_2(x) = \partial_\theta C(x,\theta_0)$, $f_3(x) = (\partial_\theta)^2 C(x,\theta_0) - \nu\left((\partial_\theta)^2 C(\cdot,\theta_0)\right)$.

Assumption 3. *For f_i (i=0,1,2,3), there exist $G_{f_i} \in C_\uparrow^\infty(\mathbf{R}^d)$ such that $\mathcal{A}G_{f_i} = f_i$.*

From (4.7), (4.8), (4.9), (4.10) and (4.11), we obtain

$$\begin{aligned}
\bar{Z}_T^{(0)} &= r_T\int_0^T \{b(X_t,\theta_0) - \partial G_{f_0}(X_t)V(X_t)\}dw_t + O_p\left(r_T\right),\\
\bar{Z}_T^{(1)} &= r_T\int_0^T \left\{\partial_\theta b(X_t,\theta_0) - \partial G_{f_1}(X_t)V(X_t)\right\}dw_t + O_p\left(r_T\right),\\
\bar{\zeta}_T^{(0)} &= r_T\int_0^T -\nu\left((\partial_\theta)^2 C(\cdot,\theta_0)\right)^{-1}\left\{\partial_\theta B(X_t,\theta_0) - \partial G_{f_2}(X_t)V(X_t)\right\}dw_t\\
&\quad + O_p\left(r_T\right),\\
\bar{\zeta}_{T,i}^{(1)} &= -\left[r_T\int_0^T \left\{(\partial_\theta)^2 B(X_t,\theta_0) - \partial G_{f_3}(X_t)V(X_t)\right\}dw_t\right]\nu\left((\partial_\theta)^2 C(\cdot,\theta_0)\right)_{\cdot i}^{-1}\\
&\quad + O_p\left(r_T\right).
\end{aligned}$$

Set $\xi^{(0)}(x) = b(x,\theta_0) - \partial G_{f_0}(x)V(x)$, $\xi^{(1)}(x) = \partial_\theta b(x,\theta_0) - \partial G_{f_1}(x)V(x)$, $\zeta^{(0)}(x) = -\nu\left((\partial_\theta)^2 C(\cdot,\theta_0)\right)^{-1}\{\partial_\theta B(x,\theta_0) - \partial G_{f_2}(x)V(x)\}$, and $\zeta_i^{(1)}(x) = -\{(\partial_\theta)^2 B(X_t,\theta_0) - \partial G_{f_3}(X_t)V(X_t)\}^{\cdot,a,\cdot}\,\nu\left((\partial_\theta)^2 C(\cdot,\theta_0)\right)^{-1}_{a,i}$. By Itô's formula,

$$\mathrm{Cov}(\bar{Z}_T^{(0)}, \bar{Z}_T^{(0)}) \;=\; E\left[{r_T}^2\int_0^T \xi^{(0)}(X_t)\xi^{(0)}(X_t)'dt\right] + o(1).$$

Since it follows from stationary that

$$\Sigma_T^{(00)} \;=\; \nu(\xi^{(0)}(\cdot)\xi^{(0)}(\cdot)') + o(1),$$

one has

$$\begin{aligned}\Sigma_T^{(00)*} &= \nu(\xi^{(0)}(\cdot)\xi^{(0)}(\cdot)') + a_T'\nu(\zeta^{(0)}(\cdot)\zeta^{(0)}(\cdot)')a_T \\ &\quad +2a_T'\nu(\zeta^{(0)}(\cdot)\xi^{(0)}(\cdot)') + o(1).\end{aligned} \tag{4.12}$$

In the same way, we obtain that

$$\Sigma_T^{(12)*} = \nu(\xi^{(1)}(\cdot)\zeta^{(0)}(\cdot)') + o(1), \tag{4.13}$$
$$\Sigma_T^{(10)*} = \nu(\xi^{(1)}(\cdot)\xi^{(0)}(\cdot)') + \nu(\xi^{(1)}(\cdot)\zeta^{(0)}(\cdot)')a_T + o(1), \tag{4.14}$$
$$\Sigma_T^{(20)*} = \nu(\zeta^{(0)}(\cdot)\xi^{(0)}(\cdot)') + \nu(\zeta^{(0)}(\cdot)\zeta^{(0)}(\cdot)')a_T + o(1), \tag{4.15}$$
$$\Sigma_{T,i}^{(30)*} = \nu(\zeta_i^{(1)}(\cdot)\xi^{(0)}(\cdot)') + \nu(\zeta_i^{(1)}(\cdot)\zeta^{(0)}(\cdot)')a_T + o(1). \tag{4.16}$$

Next, we compute λ_T^{000*}.

$$\begin{aligned}\lambda_T^{000*} &= E[(\bar{Z}_T^{(0)*})^3] \\ &= E[(\bar{Z}_T^{(0)})^3] + 3E[(\bar{Z}_T^{(0)})^2(a_T'\bar{\zeta}_T^{(0)})] \\ &\quad +3E[(\bar{Z}_T^{(0)})(a_T'\bar{\zeta}_T^{(0)})^2] + E[(a_T'\bar{\zeta}_T^{(0)})^3],\end{aligned}$$

where

$$\begin{aligned}\bar{Z}_T^{(0)} &= r_T\int_0^T \xi^{(0)}(X_t)dw_t + r_T\{G_{f_0}(X_T) - G_{f_0}(X_0)\} + r_T\alpha_1(X_0,\theta_0), \\ \bar{\zeta}_T^{(0)} &= r_T\int_0^T \zeta^{(0)}(X_t)dw_t - r_T\nu\Big((\partial_\theta)^2 C(\cdot,\theta_0)\Big)^{-1}\{G_{f_2}(X_T) - G_{f_2}(X_0)\}.\end{aligned}$$

Define that $f_4(x) = \|\xi^{(0)}(x)\|^2 - \nu(\|\xi^{(0)}(\cdot)\|^2)$, $f_5(x) = \xi^{(0)}(X_t)\zeta^{(0)}(X_t)'a_T - \nu(\xi^{(0)}(\cdot)\zeta^{(0)}(\cdot)'a_T)$, $f_6(x) = a_T'\zeta^{(0)}(X_t)\zeta^{(0)}(X_t)'a_T - \nu(a_T'\zeta^{(0)}(\cdot)\zeta^{(0)}(\cdot)'a_T)$.

Assumption 4. *For f_i (i=4,5,6), there exist $G_{f_i} \in C_\uparrow^\infty(\mathbf{R}^d)$ such that $\mathcal{A}G_{f_i} = f_i$.*

By Itô's formula,

$$
\begin{aligned}
& E[(\bar{Z}_T^{(0)})^3] \\
= \;& {r_T}^3 E\left[\left(\int_0^T \xi^{(0)}(X_t)dw_t\right)^3\right] + o\left(r_T\right) \\
= \;& 3{r_T}^3 E\left[\int_0^T \left(\int_0^t \xi^{(0)}(X_u)dw_u\right) \|\xi^{(0)}(X_t)\|^2 dt\right] + o\left(r_T\right) \\
= \;& 3{r_T}^3 E\left[\left(\int_0^T \xi^{(0)}(X_t)dw_t\right) \cdot \left(\int_0^T [\|\xi^{(0)}(X_t)\|^2 - \nu(\|\xi^{(0)}(\cdot)\|^2)]dt\right)\right] + o\left(r_T\right) \\
= \;& 3{r_T}^3 E\left[\left(\int_0^T \xi^{(0)}(X_t)dw_t\right) \cdot \left(\int_0^T f_4(X_t)dt\right)\right] + o\left(r_T\right).
\end{aligned}
$$

It follows from (4.11) that

$$
E[(\bar{Z}_T^{(0)})^3] \;=\; -3r_T E\left[{r_T}^2 \int_0^T \xi^{(0)}(X_t)\left\{\partial G_{f_4}(X_t)V(X_t)\right\}' dt\right] + o\left(r_T\right).
$$

Similarly,

$$
\begin{aligned}
E[(\bar{Z}_T^{(0)})^2(a_T'\bar{\zeta}_T^{(0)})] \;=\; & -r_T E\left[{r_T}^2 \int_0^T a_T'\zeta^{(0)}(X_t)\left\{\partial G_{f_4}(X_t)V(X_t)\right\}' dt\right] \\
& -2r_T E\left[{r_T}^2 \int_0^T \xi^{(0)}(X_t)\left\{\partial G_{f_5}(X_t)V(X_t)\right\}' dt\right] \\
& +o\left(r_T\right), \\
E[(\bar{Z}_T^{(0)})(a_T'\bar{\zeta}_T^{(0)})^2] \;=\; & -2r_T E\left[{r_T}^2 \int_0^T a_T'\zeta^{(0)}(X_t)\left\{\partial G_{f_5}(X_t)V(X_t)\right\}' dt\right] \\
& -r_T E\left[{r_T}^2 \int_0^T \xi^{(0)}(X_t)\left\{\partial G_{f_6}(X_t)V(X_t)\right\}' dt\right] \\
& +o\left(r_T\right), \\
E[(a_T'\bar{\zeta}_T^{(0)})^3] \;=\; & -3r_T E\left[{r_T}^2 \int_0^T a_T'\zeta^{(0)}(X_t)\left\{\partial G_{f_6}(X_t)V(X_t)\right\}' dt\right] \\
& +o\left(r_T\right).
\end{aligned}
$$

Finally, by stationary, we obtain

$$
\begin{aligned}
(4.17)\;\; & \lambda_T^{000*} \\
= \;& -3r_T\nu(\xi^{(0)}(\cdot)\{\partial G_{f_4}(\cdot)V(\cdot)\}') - r_T\nu(a_T'\zeta^{(0)}(\cdot)\{\partial G_{f_4}(\cdot)V(\cdot)\}') \\
& -2r_T\nu(\xi^{(0)}(\cdot)\{\partial G_{f_5}(\cdot)V(\cdot)\}') - 2r_T\nu(a_T'\zeta^{(0)}(\cdot)\{\partial G_{f_5}(\cdot)V(\cdot)\}') \\
& -r_T\nu(\xi^{(0)}(\cdot)\{\partial G_{f_6}(\cdot)V(\cdot)\}') - 3r_T\nu(a_T'\zeta^{(0)}(\cdot)\{\partial G_{f_6}(\cdot)V(\cdot)\}') + o\left(r_T\right).
\end{aligned}
$$

From (4.3), (4.6) and (4.13), one has the information criterion based on the AEU which works for M-estimators.

Theorem 6 (Information criterion in the sense of AEU). *Consider the models (4.1) and (4.2). Suppose that Assumptions 1 and 2 for some $K' > 1$ and $\alpha > 1$ hold true. Moreover, suppose that Assumption 3 for f_1 and f_2 holds true. Then*

$$\begin{aligned} IC_1(\boldsymbol{X}_T(\omega)) &= r_T\left[\tilde{a}(X_0,\hat{\theta}_T) + \int_0^T \tilde{b}(X_t,\theta)dX_t\Big|_{\theta=\hat{\theta}_T} + \int_0^T \tilde{c}(X_t,\hat{\theta}_T)dt\right] \\ &\quad - r_T b_1(\hat{\theta}_T), \end{aligned}$$

where $b_1(\theta_0) = tr\left(\nu(\xi^{(1)}(\cdot)\zeta^{(0)}(\cdot)')\right)$ and $\int_0^T \tilde{b}(X_t,\theta)dX_t\Big|_{\theta=\hat{\theta}_T}$ means substituting $\theta = \hat{\theta}_T$ for the random field $\int_0^T \tilde{b}(X_t,\theta)dX_t$.

From (3.5), (4.3), (4.6), (4.12), (4.13), (4.14), (4.15), (4.16) and (4.17), we obtain the information criterion based on the second order AMU which works for M-estimators.

Theorem 7 (Information criterion in the sense of AMU). *Consider the models (4.1) and (4.2). Suppose that Assumptions 1 and 2 for some $K' > 1$ and $\alpha > 1$ hold true. Moreover, suppose that Assumptions 3 and 4 hold true. Then*

$$\begin{aligned} IC_2(\boldsymbol{X}_T(\omega)) &= r_T\left[\tilde{a}(X_0,\hat{\theta}_T) + \int_0^T \tilde{b}(X_t,\theta)dX_t\Big|_{\theta=\hat{\theta}_T} + \int_0^T \tilde{c}(X_t,\hat{\theta}_T)dt\right] \\ &\quad - r_T b_2(\hat{\theta}_T), \end{aligned}$$

where

$$\begin{aligned} b_2(\theta_0) &= -\frac{1}{6}{r_T}^{-1}\lambda_T^{000*}\frac{1}{\Sigma_T^{(00)*}} \\ &\quad + \left[tr\Sigma_T^{(12)*} - \frac{(\Sigma_T^{(10)*})'\Sigma_T^{(20)*}}{\Sigma_T^{(00)*}} - \frac{1}{2}\frac{(\Sigma_T^{(20)*})' b_T \Sigma_T^{(20)*}}{\Sigma_T^{(00)*}}\right. \\ &\quad \left. - \frac{(\sum_{i=1}^p a_{T,i}\Sigma_{T,i}^{(30)*})'\Sigma_T^{(20)*}}{(\Sigma_T^{(00)*})} - \frac{1}{2}\frac{(\Sigma_T^{(20)*})'(\sum_{i=1}^p a_{T,i}c_{T,i})\Sigma_T^{(20)*}}{(\Sigma_T^{(00)*})}\right]. \end{aligned}$$

Remark 5. *As a sufficient condition for [A1] in Assumption 1, we can refer Veretennikov [34] and Kusuoka and Yoshida [18]. For [A3], see Kusuoka and Yoshida [18], which is used the relation between the Hörmander condition and the regularity of distributions, and see also Yoshida [36] applying the support theorem.*

§5. Proofs

In order to prove Theorem 1, we need two propositions below. For this reason, we introduce some notation. Define functions $\tilde{P}_{T,r}(u)$ by the formal Taylor expansion:

$$\exp\left(\sum_{r=2}^{\infty} r!^{-1}\epsilon^{r-2}\chi_{T,r}(u)\right) = \exp\left(\frac{1}{2}\chi_{T,2}(u)\right) + \sum_{r=1}^{\infty}\epsilon^r T^{-r/2}\tilde{P}_{T,r}(u), \tag{5.1}$$

where

$$\chi_{T,r}(u) = \left(\frac{d}{d\epsilon}\right)_0^r \log E[\exp(i\epsilon u\cdot \bar{Z}_T^*)].$$

Let $\hat{\Psi}_{T,k}(u)$ be the k-th partial sum of the right-hand side of (5.1) with $\epsilon = 1$:

$$\hat{\Psi}_{T,k}(u) = \exp\left(\frac{1}{2}\chi_{T,2}(u)\right) + \sum_{r=1}^{k} T^{-r/2}\tilde{P}_{T,r}(u).$$

For $T > 0$ and $k \in \boldsymbol{N}$, a signed measure $\hat{\Psi}_{T,k}$ is defined as the Fourier inversion of $\hat{\Psi}_{T,k}(u)$.

Proposition 1. *Let $M, \gamma > 0$. Suppose that Assumption 1 holds true. Then for any $K > 0$,*
(i) there exist constants $\delta > 0$ and $c > 0$ such that for any function $f \in \mathcal{E}(M,\gamma)$,

$$|E[f(\tilde{S}_T^*)] - \Psi_{T,1}[f]| \le c\omega(f, r_T^K) + \varepsilon_T,$$

where $\varepsilon_T = o(r_T^{((1+\delta)\wedge K)})$ depends on $\mathcal{E}(M,\gamma)$.
(ii) The sighed-measure $d\Psi_{T,1}$ has a density $d\Psi_{T,1}(z)/dz = q_{T,1}(z)$ with

$$\begin{aligned}
& q_{T,1}(z^{(0)}) \\
= \; & \int_{\mathbf{R}^{2p+q}} p_{T,1}(z)dz^{(1)}dz^{(2)}dz^{(3)} \\
& -r_T\partial_{z^{(0)}}\Bigg[\int_{\mathbf{R}^{2p+q}}\Big\{z^{(1)\prime}z^{(2)} + a_T' Q_1(z^{(2)},z^{(3)}) - \int_\Omega a_T' Q_1(\bar{\zeta}_T^{(0)*},\bar{\zeta}_T^{(1)*})P(d\omega) \\
& +\frac{1}{2}z^{(2)\prime}b_T z^{(2)} - \frac{1}{2}\int_\Omega \bar{\zeta}_T^{(0)*\prime}b_T\bar{\zeta}_T^{(0)*}P(d\omega) - b(\theta_0)\Big\}\phi(z;\Sigma_T^*)dz^{(1)}dz^{(2)}dz^{(3)}\Bigg],
\end{aligned}$$

where $p_{T,1}(z) = \phi(z;\Sigma_T^*)(1+\frac{1}{6}\lambda_T^{\alpha\beta\gamma *}h_{\alpha\beta\gamma}(z;\Sigma_T^*))$ and $\lambda_T^{\alpha\beta\gamma *}$ is the third cumulant of $\bar{Z}_T^*$.

Proof of Proposition 1. Theorem 5 of Kusuoka & Yoshida [18] together with Theorem 5.1 in Sakamoto & Yoshida [27] gives an asymptotic expansion of $\tilde{S}_T^*$ as follows.

For any $K \in \mathbf{N}$, there exist smooth functions $q_{j,1,T} : \mathbf{R} \to \mathbf{R}$ such that

$$q_{0,1,T}(z^{(0)}) = \int_{\mathbf{R}^{2p+q}} \phi(z;\Sigma_T^*)dz^{(1)}dz^{(2)}dz^{(3)},$$

and

$$\begin{aligned} & q_{1,1,T}(z^{(0)}) \\ = & \int_{\mathbf{R}^{2p+q}} \Xi_{T,1}(z)\phi(z;\Sigma_T^*)dz^{(1)}dz^{(2)}dz^{(3)} \\ & -\partial_{z^{(0)}}\left[\int_{\mathbf{R}^{2p+q}} \left\{ z^{(1)\prime}z^{(2)} + a_T' Q_1(z^{(2)}, z^{(3)}) - \int_\Omega a_T' Q_1(\bar{\zeta}_T^{(0)*}, \bar{\zeta}_T^{(1)*})P(d\omega) \right.\right. \\ & \left.\left. +\frac{1}{2} z^{(2)\prime} b_T z^{(2)} - \frac{1}{2}\int_\Omega \bar{\zeta}_T^{(0)*\prime} b_T \bar{\zeta}_T^{(0)*} P(d\omega) - b(\theta_0) \right\} \phi(z;\Sigma_T^*)dz^{(1)}dz^{(2)}dz^{(3)} \right], \end{aligned}$$

where $z = (z^{(0)}, z^{(1)}, z^{(2)}, z^{(3)})$ and $\Xi_{T,1}(z) = r_T^{-1}\frac{1}{6}\lambda^{\alpha\beta\gamma*}h_{\alpha\beta\gamma}(z;\Sigma_T^*)$, and there exist constants $\delta > 0$ and $c > 0$ such that for any $f \in \mathcal{E}(M, \gamma)$,

$$\left| E[f(\tilde{S}_T^*)] - \int_{\mathbf{R}} f(z^{(0)}) \sum_{j=0}^{1} T^{-j/2} q_{j,1,T}(z^{(0)})dz^{(0)} \right| \leq c\omega(f, T^{-K}) + \epsilon_T^{(1)},$$

where $\epsilon_T^{(1)}$ is a sequence of constants independent of f with $\epsilon_T^{(1)} = o(T^{-\frac{1}{2}(1+\delta)\wedge K})$.

It then follows that

$$\begin{aligned} & q_{T,1}(z^{(0)}) \\ := & \sum_{j=0}^{1} T^{-j/2} q_{j,1,T}(z^{(0)}) \\ = & \int_{\mathbf{R}^{2p+q}} p_{T,1}(z)dz^{(1)}dz^{(2)}dz^{(3)} \\ - & r_T\partial_{z^{(0)}}\left[\int_{\mathbf{R}^{2p+q}} \left\{ z^{(1)\prime}z^{(2)} + a_T' Q_1(z^{(2)}, z^{(3)}) - \int_\Omega a_T' Q_1(\bar{\zeta}_T^{(0)*}, \bar{\zeta}_T^{(1)*})P(d\omega) \right.\right. \\ & \left.\left. +\frac{1}{2} z^{(2)\prime} b_T z^{(2)} - \frac{1}{2}\int_\Omega \bar{\zeta}_T^{(0)*\prime} b_T \bar{\zeta}_T^{(0)*} P(d\omega) - b(\theta_0) \right\} \phi(z;\Sigma_T^*)dz^{(1)}dz^{(2)}dz^{(3)} \right], \end{aligned}$$

where $p_{T,1}(z) = \phi(z;\Sigma_T^*)(1 + \frac{1}{6}\lambda_T^{\alpha\beta\gamma*}h_{\alpha\beta\gamma}(z;\Sigma_T^*))$. Consequently, we have the desired result.

Remark 6. *In Proposition 1, we assumed the non-degeneracy of $Cov(r_T Z_T^*)$. However, even if $Cov(r_T Z_T^*)$ is degenerate, it is still possible to interpret each $p_{T,k}(z)$ as a Schwartz distribution, and to prove the validity of the formula for $q_{T,1}$ given in Proposition 1. For more details, see Sakamoto and Yoshida [26].*

Proposition 2. *Let $M, \gamma > 0$. Suppose that Assumptions 1 and 2 hold true. Then,*
(i) there exist constants $\delta > 0$ and $\tilde{c} > 0$ such that for any function $f \in \mathcal{E}(M,\gamma)$,

$$|E[f(\bar{S}_T^*)] - \Psi_{T,1}[f]| \leq \tilde{c}\omega(f, 2r_T^{K'}) + \tilde{\varepsilon}_T,$$

where $\tilde{\varepsilon}_T = o(r_T^{((1+\delta)\wedge\alpha)})$ depends on $\mathcal{E}(M,\gamma)$.
(ii)The sighed-measure $d\Psi_{T,1}$ has the same density $d\Psi_{T,1}(z)/dz = q_{T,1}(z)$ as Proposition 1.

Proof of Proposition 2. In the same way as in the proof of Lemma 6.5 in [25] (Lemma 9.5 in [27]) or Theorem 2 in [33], we can show the result.

Proof of Theorem 1. In order to obtain an explicit expression for the second order asymptotic expansion of the distribution of $\bar{S}_T^*$, we need to compute the following integral.

$$\begin{aligned}
& q_{T,1}(z^{(0)}) \\
= & \int_{\mathbf{R}^{2p+q}} \phi(z;\Sigma_T^*)dz^{(1)}dz^{(2)}dz^{(3)} \\
+ & \frac{1}{6}\int_{\mathbf{R}^{2p+q}} \lambda_T^{\alpha\beta\gamma*}h_{\alpha\beta\gamma}(z;\Sigma_T^*)\phi(z;\Sigma_T^*)dz^{(1)}dz^{(2)}dz^{(3)} \\
- & r_T\partial_{z^{(0)}}\Bigg[\int_{\mathbf{R}^{2p+q}}\Bigg\{z^{(1)\prime}z^{(2)} + a_T'Q_1(z^{(2)},z^{(3)}) - \int_\Omega a_T'Q_1(\bar{\zeta}_T^{(0)*},\bar{\zeta}_T^{(1)*})P(d\omega) \\
& +\frac{1}{2}z^{(2)\prime}b_Tz^{(2)} - \frac{1}{2}\int_\Omega \bar{\zeta}_T^{(0)*\prime}b_T\bar{\zeta}_T^{(0)*}P(d\omega) - b(\theta_0)\Bigg\}\phi(z;\Sigma_T^*)dz^{(1)}dz^{(2)}dz^{(3)}\Bigg] \\
= & \phi(z^{(0)};\Sigma_T^{(00)*}) + \frac{1}{6}(\mathrm{I}) \\
& -r_T\partial_{z^{(0)}}\left[(\mathrm{II}) + (\mathrm{III}) + \frac{1}{2}(\mathrm{IV}) - b(\theta_0)\phi(z^{(0)};\Sigma_T^{(00)*})\right],
\end{aligned}$$

where

$$\begin{aligned}
(\mathrm{I}) &= -\int_{\mathbf{R}^{2p+q}} \lambda_T^{\alpha\beta\gamma*}\partial_\alpha\partial_\beta\partial_\gamma\phi(z;\Sigma_T^*)dz^{(1)}dz^{(2)}dz^{(3)} \\
&= \lambda_T^{000*}h_3(z^{(0)};\Sigma_T^{(00)*})\phi(z^{(0)};\Sigma_T^{(00)*}), \\
(\mathrm{II}) &= \int_{\mathbf{R}^{2p+q}} z^{(1)\prime}z^{(2)}\frac{\phi(z;\Sigma_T^*)}{\int_{\mathbf{R}^{2p+q}}\phi(z;\Sigma_T^*)dz^{(1)}dz^{(2)}dz^{(3)}}dz^{(1)}dz^{(2)}dz^{(3)} \\
&\quad \times\int_{\mathbf{R}^{2p+q}}\phi(z;\Sigma_T^*)dz^{(1)}dz^{(2)}dz^{(3)} \\
&= \int_{\mathbf{R}^{2p}} z^{(1)\prime}z^{(2)}\phi(z^{(1)},z^{(2)};b_1^*(\Sigma_T^{(00)*})^{-1}z^{(0)}, S_1^* - b_1^*(\Sigma_T^{(00)*})^{-1}(b_1^*)') \\
&\quad \times dz^{(1)}dz^{(2)}\phi(z^{(0)};\Sigma_T^{(00)}),
\end{aligned}$$

$$
\begin{aligned}
&=: \quad C_T^{(1)}(z^{(0)})\phi(z^{(0)};\Sigma_T^{(00)*}),\\
\text{(III)} \quad &= \quad \int_{\mathbf{R}^{2p+q}} a_T' Q_1(z^{(2)},z^{(3)}) \frac{\phi(z;\Sigma_T^*)}{\int_{\mathbf{R}^{2p+q}}\phi(z;\Sigma_T^*)dz^{(1)}dz^{(2)}dz^{(3)}} dz^{(1)}dz^{(2)}dz^{(3)}\\
&\quad\; \times \int_{\mathbf{R}^{2p+q}} \phi(z;\Sigma_T^*)dz^{(1)}dz^{(2)}dz^{(3)} - \int_\Omega a_T' Q_1(\bar\zeta_T^{(0)},\bar\zeta_T^{(1)})P(d\omega)\phi(z^{(0)};\Sigma_T^{(00)*})\\
&= \quad \int_{\mathbf{R}^{p+q}} a_T' Q_1(z^{(2)},z^{(3)})\phi(z^{(2)},z^{(3)};b_2^*(\Sigma_T^{(00)*})^{-1}z^{(0)},S_2^*-b_2^*(\Sigma_T^{(00)*})^{-1}(b_2^*)')\\
&\quad\; \times dz^{(2)}dz^{(3)}\phi(z^{(0)};\Sigma_T^{(00)}) - \int_\Omega a_T' Q_1(\bar\zeta_T^{(0)},\bar\zeta_T^{(1)})P(d\omega)\phi(z^{(0)};\Sigma_T^{(00)*})\\
&=: \quad C_T^{(2)}(z^{(0)})\phi(z^{(0)};\Sigma_T^{(00)*}),\\
\text{(IV)} \quad &= \quad \int_{\mathbf{R}^{2p+q}} (z^{(2)})' b_T z^{(2)} \frac{\phi(z;\Sigma_T^*)}{\int_{\mathbf{R}^{2p+q}}\phi(z;\Sigma_T^*)dz^{(1)}dz^{(2)}dz^{(3)}} dz^{(1)}dz^{(2)}dz^{(3)}\\
&\quad\; \times \int_{\mathbf{R}^{2p+q}} \phi(z;\Sigma_T^*)dz^{(1)}dz^{(2)}dz^{(3)} - \int_\Omega \bar\zeta_T^{(0)\prime} b_T \bar\zeta_T^{(0)} P(d\omega)\phi(z^{(0)};\Sigma_T^{(00)*})\\
&= \quad \int_{\mathbf{R}^{p}} (z^{(2)})' b_T z^{(2)} \phi(z^{(2)};b_3^*(\Sigma_T^{(00)*})^{-1}z^{(0)},S_3^*-b_3^*(\Sigma_T^{(00)*})^{-1}(b_3^*)')dz^{(2)}\\
&\quad\; \times \phi(z^{(0)};\Sigma_T^{(00)}) - \int_\Omega \bar\zeta_T^{(0)\prime} b_T \bar\zeta_T^{(0)} P(d\omega)\phi(z^{(0)};\Sigma_T^{(00)*})\\
&=: \quad C_T^{(3)}(z^{(0)})\phi(z^{(0)};\Sigma_T^{(00)*}).
\end{aligned}
$$

Since $\phi(z^{(1)},z^{(2)};b_1^*(\Sigma_T^{(00)*})^{-1}z^{(0)},S_1^*-b_1^*(\Sigma_T^{(00)*})^{-1}(b_1^*)')$ is normal with mean μ_1 and covariance matrix Σ_1, where

$$
\begin{aligned}
\mu_1 &= \begin{bmatrix} \Sigma_T^{(10)*} \\ \Sigma_T^{(20)*} \end{bmatrix} (\Sigma_T^{(00)*})^{-1} z^{(0)},\\
\Sigma_1 &= \begin{bmatrix} \Sigma_T^{(11)*} - (\Sigma_T^{(00)*})^{-1}\Sigma_T^{(10)*}(\Sigma_T^{(10)*})' & \Sigma_T^{(12)*} - (\Sigma_T^{(00)*})^{-1}\Sigma_T^{(10)*}(\Sigma_T^{(20)*})' \\ (\Sigma_T^{(12)*})' - (\Sigma_T^{(00)*})^{-1}\Sigma_T^{(20)*}(\Sigma_T^{(10)*})' & \Sigma_T^{(22)*} - (\Sigma_T^{(00)*})^{-1}\Sigma_T^{(20)*}(\Sigma_T^{(20)*})' \end{bmatrix},
\end{aligned}
$$

we have

$$
C_T^{(1)}(z^{(0)}) = \frac{(\Sigma_T^{(10)*})'\Sigma_T^{(20)*}}{(\Sigma_T^{(00)*})^2}[(z^{(0)})^2 - \Sigma_T^{(00)*}] + \mathrm{tr}\Sigma_T^{(12)*}.
$$

Since $\phi(z^{(2)};b_3^*(\Sigma_T^{(00)*})^{-1}z^{(0)},S_3^*-b_3^*(\Sigma_T^{(00)*})^{-1}(b_3^*)')$ is normal with mean μ_2 and covariance matrix Σ_2, where $\mu_2 = \Sigma_T^{(20)*}(\Sigma_T^{(00)*})^{-1}z^{(0)}$ and $\Sigma_2 = \Sigma_T^{(22)*} - (\Sigma_T^{(00)*})^{-1}\Sigma_T^{(20)*}(\Sigma_T^{(20)*})'$, we obtain

$$
C_T^{(3)}(z^{(0)}) = \frac{(\Sigma_T^{(20)*})' b_T \Sigma_T^{(20)*}}{(\Sigma_T^{(00)*})^2}[(z^{(0)})^2 - \Sigma_T^{(00)*}].
$$

This completes the proof.

Proof of Theorem 2. In Theorem 1, setting $f(x) = x$ and $b(\cdot) = b_1(\cdot)$, we see that $E[\bar{S}_T^*] = r_T \left[\mathrm{tr}\Sigma_T^{(12)*} - b_1(\theta_0)\right] + o\,(r_T) = o\,(r_T)$, which completes the proof.

Proof of Theorem 3. In Theorem 1, putting $f(x) = 1_{(0,\infty)}(x) - \frac{1}{2}$ and $b(\cdot) = b_2(\cdot)$, we obtain

$$\begin{aligned}
& P[\bar{S}_T^* > 0] - \frac{1}{2} \\
= \;& -\frac{1}{6}\lambda_T^{000*}\frac{1}{\Sigma_T^{(00)*}}\phi(0;\Sigma_T^{(00)*}) \\
+ \;& r_T\left[\mathrm{tr}\Sigma_T^{(12)*} - \frac{(\Sigma_T^{(10)*})'\Sigma_T^{(20)*}}{\Sigma_T^{(00)*}} + C_T^{(2)}(0) - \frac{1}{2}\frac{(\Sigma_T^{(20)*})'b_T\Sigma_T^{(20)*}}{\Sigma_T^{(00)*}} - b_2(\theta_0)\right] \\
& \times\phi(0;\Sigma_T^{(00)*}) + o\,(r_T) \\
= \;& o\,(r_T)\,.
\end{aligned}$$

Similarly, putting $f(x) = 1_{(-\infty,0)}(x) - \frac{1}{2}$ and $b(\cdot) = b_2(\cdot)$ in Theorem 1, we have $P[\bar{S}_T^* < 0] - \frac{1}{2} = o\,(r_T)$, which completes the proof.

Acknowledgements

The authors wish to thank the referee for his careful reading and valuable comments. This work was in part supported by the Research Fund for Scientists of the Ministry of Education, Culture, Sports, Science and Technology, and by Cooperative Program of the Institute of Statistical Mathematics.

References

[1] Akahira, M. and Takeuchi, K. Asymptotic efficiency of statistical estimators: concepts and higher order asymptotic efficiency. Lecture Notes in Statistics Vol. 7, Berlin Heidelberg New York, Springer. (1981).

[2] Akaike, H. Information theory and an extension of the maximum likelihood principle, *2nd International Symposium in Information Theory*, Petrov, B.N. and Csaki, F. eds., Akademiai Kaido, Budapest, (1973), 267-281.

[3] Akaike, H. A new look at the statistical model identification, *IEEE Trans. Auto. Control*, **AC-19**, (1974), 716-723.

[4] Barron, A.R., Birgé, L. and Massart, P. Risk bounds for model selection via penalization, *Probab. Theory Relat. Fields*, **113**, (1999), 301-413.

[5] Bhattacharya, R.N. and Ghosh, J.K. On the validity of the formal Edgeworth expansion, *Ann. Statist.*, **6**, (1978), 434-451.

[6] Burman, P. and Nolan, D. A general Akaike-type criterion for model selection in robust regression, *Biometrika*, **82**, (1995), 877-886.

[7] Burnham, K.P. and Anderson, D.R. Model selection and Inference, Berlin Heidelberg New York, Springer. (1998).

[8] Gihman, I.I. and Skorohod, A.V. Stochastic differential equations, Berlin, Springer. (1972).

[9] Götze, F. and Hipp, C. Asymptotic expansions for sums of weakly dependent random vectors, *Z. Wahrsh. Verw. Gebiete*, **64**, (1983), 211-239.

[10] Götze, F. and Hipp, C. Asymptotic distribution of statistics in time series, *Ann. Statist.*, **22**, (1994), 2062-2088.

[11] Hall, P. Akaike's information criterion and Kullback-Leibler loss for histogram density estimation, *Probab. Theory Related Fields*, **85**, (1990), 449-467.

[12] Hurvich, C.M. and Tsai, C.-L. A corrected Akaike information criterion for vector autoregressive model selection, *J. Time Series Anal.*, **14**, (1993), 271-279.

[13] Hurvich, C.M. and Tsai, C.-L. Relative rates of convergence for efficient model selection criteria in linear regression, *Biometrika*, **82**, (1995), 418-425.

[14] Knight, K. Consistency of Akaike's information criterion for infinite variance autoregressive processes, *Ann. Statist.*, **17**, (1989), 824-840.

[15] Konishi, S. and Kitagawa, G. Generalised information criteria in model selection, *Biometrika*, **83**, (1996), 875-890.

[16] Konishi, S. and Kitagawa, G. Asymptotic theory for information criteria in model selection —functional approach. C. R. Rao 80th birthday felicitation volume, Part IV. *J. Statist. Plann. Inference*, **114**, (2003), 45–61.

[17] Kullback, S. and Leibler, R.A. On information and sufficiency, *Ann. Math. Statist.*, **22**, (1951), 79-86.

[18] Kusuoka, S. and Yoshida, N. Malliavin calculus, geometric mixing, and expansion of diffusion functionals, *Probab. Theory Related Fields*, **116**, (2000), 457-484.

[19] Laud, P.W. and Ibrahim, J.G. Predictive model selection, *J. R. Statist. Soc.*, **B57**, (1995), 247-262.

[20] Mykland, P.A. Asymptotic expansions and bootstrapping distributions for dependent variables: a martingale approach, *Ann. Statist.*, **20**, (1992), 623-654.

[21] Mykland, P.A. Asymptotic expansions for martingales, *Ann. Probab.*, **21**, (1993), 800-818.

[22] Mykland, P.A. Martingale expansions and second order inference, *Ann. Statist.*, **23**, (1995), 707-731.

[23] Portnoy, S. Local asymptotics for quantile smoothing splines, *Ann. Statist.*, **25**, (1997), 414-434.

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

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

[26] Sakamoto, Y. and Yoshida, N. Asymptotic expansion under degeneracy. *J. Japan Statist. Soc.*, **33**, (2003), 145-156.

[27] Sakamoto, Y. and Yoshida, N. Asymptotic expansion formulas for functionals of ϵ-Markov processes with a mixing property. *Ann. Inst. Statist. Math.* **56**, (2004), 545-597.

[28] Shibata, R. Asymptotically efficient selection of the order of the model for estimating parameters of a linear process, *Ann. Statist.*, **8**, (1980), 147-164.

[29] Shibata, R. An optimal autoregressive spectral estimate, *Ann. Statist.*, **9**, (1981), 300-306.

[30] Shimodaira, H. An application of multiple comparison techniques to model selection, *Ann. Inst. Stat. Math.*, **50**, (1998), 1-13.

[31] Shimodaira, H. Recent developments in model selection theory, *Proc. Inst. Statist. Math.*, **47**, (1999), 3-27. (in Japanese).

[32] Takeuchi, K. Distribution of information statistics and criteria for adequacy of models, *Mathematical Sciences*, **153**, (1976), 12-18. (in Japanese).

[33] Uchida, M. and Yoshida, N. Information criteria in model selection for mixing processes, *Statist. Infer. Stochast. Process.*, **4**, (2001), 73-98.

[34] Veretennikov, A.Y. On polynomial mixing bounds for stochastic differential equations, *Stochastic Process. Appl.*, **70**, (1997), 115-127.

[35] Yang, Y. and Barron, A.R. An asymptotic property of model selection criteria, *IEEE Transactions on Information Theory*, **44**, (1998), 95-116.

[36] Yoshida, N. Partial mixing and Edgeworth expansion. *Probab. Theory Related Fields* **129**, (2004), 559-624.

Masayuki Uchida
Faculty of Mathematics, Kyushu University, Ropponmatsu, Fukuoka 810-8560, Japan

Nakahiro Yoshida
Graduate School of Mathematical Sciences, University of Tokyo, 3-8-1 Komaba, Meguro-ku, Tokyo 153-8914, Japan