Statistical Inference for Stochastic Processes **4:** 73–98, 2001.

Information Criteria in Model Selection for Mixing Processes

MASAYUKI UCHIDA[1] and NAKAHIRO YOSHIDA[2]
[1]*Faculty of Mathematics, Kyushu University, Ropponmatsu, Fukuoka 810-8560, Japan*
[2]*Graduate School of Mathematical Sciences, University of Tokyo, 3-8-1 Komaba, Meguro-ku, Tokyo 153-8914, Japan*

Abstract. We present information criteria for statistical model evaluation problems for stochastic processes. The emphasis is put on the use of the asymptotic expansion of the distribution of an estimator based on the conditional Kullback–Leibler divergence for stochastic processes. Asymptotic properties of information criteria and their improvement are discussed. An application to a diffusion process is presented.

AMS Mathematics Subject Classifications (1991): 62B10, 62E20, 62M05, 60H07.

Key words: information criterion, asymptotic expansion, diffusion process, median-unbiasedness, M-estimators, Malliavin calculus.

1. Introduction: Basic Concepts in Model Selection

The purpose of the present article is to provide a new perspective to the model selection problem from aspects of the higher order statistical inference theory. The proposed method enables us to obtain new information criteria in a unified way as well as the traditional criteria, and moreover, to treat a statistical model of mixing processes with continuous time parameter. Here mixing processes include diffusion processes with jumps, mixing point processes, and also nonlinear time series models with discrete time parameter embedded in continuous time.

Since our discussion will be concerned with the fundamentals of the theory of information criteria, in fact, the traditional method will be changed, let us remind the reader the concept of the information criterion. For simplicity, we shall begin with independent observations taking values in $\mathbf{R}^d$. Suppose that $\mathbf{X}_n = \{X_1, \ldots, X_n\}$ are independent random samples from an unknown distribution $G(x)$ with density function $g(x)$. With the information contained in the observations $\mathbf{X}_n$, we choose a parametric model $\{f(x|\theta); \theta \in \Theta\}$ among competing parametric models. The statistical model $\{f(x|\theta); \theta \in \Theta\}$ may or may not contain the true density $g(x)$, but it is expected that its deviation from the statistical model is not so large. We fit the selected parametric model $\{f(x|\theta); \theta \in \Theta\}$ to the real data by replacing the unknown parameter θ with some estimator $\hat{\theta}$, for example the maximum likelihood estimator. Then a future observation z from the true density

$g(\cdot)$ is predicted by the statistical model $f(z|\hat{\theta})$. A basic procedure in the theory of information criteria is to assess the goodness-of-fit of the predicted distribution $f(z|\hat{\theta})$ to the true density $g(z)$ generating real data. The Kullback–Leibler information (Kullback and Leibler (1951)) given the observed data $\mathbf{X}_n$ is defined by

$$I\{g(z);\, f(z|\hat{\theta})\} = E_G\left[\log\frac{g(Z)}{f(Z|\hat{\theta})}\right]$$

and is used for measuring the divergence of $f(z|\hat{\theta})$ from $g(z)$. For a distribution with density, it can be expressed as

$$I\{g(z);\, f(z|\hat{\theta})\} = \int_{\mathbf{R}^d} g(z)\log g(z)\,\mathrm{d}z - \int_{\mathbf{R}^d} g(z)\log f(z|\hat{\theta})\,\mathrm{d}z.$$

The first term does not depend on the statistical model and only the second term is relevant to comparing different models. The second term, which is called the expected log likelihood, depends on the observed data $\mathbf{X}_n$ through $\hat{\theta}$. We choose a statistical model $f(z|\hat{\theta})$ for which the value of the Kullback–Leibler information $I\{g(z);\, f(z|\hat{\theta})\}$ is minimized among competing models. Moreover, it is also possible to choose a model for which the expected Kullback–Leibler information $E_G[I\{g(z);\, f(z|\hat{\theta})\}]$ is minimized. In this case, the information criterion is regarded as an estimator of the expected divergence which is an unknown parameter. Uchida and Yoshida (1999) derived information criteria from this standpoint; the obtained information criteria are different from the criteria we obtain in this paper. Contrarily, the information criterion here is regarded as a predictor of the conditional Kullback–Leibler information. In practice, since one uses $f(z|\hat{\theta})$ which is determined only by the present data $\mathbf{X}_n$, the conditional Kullback–Leibler information is the real divergence between the model used for prediction and the true model. In this sense, the different models should be compared by their conditional divergence, and this is a reason why we here take the conditional divergence. Furthermore, it is also found that, under the expectation-unbiasedness condition, both approaches lead to the same information criteria; it is the case for traditional criteria. However, if we generalize unbiasedness apart from expectation-unbiasedness, the resulting criteria vary according to the approach we take.

Minimizing $I\{g(z);\, f(z|\hat{\theta})\}$ is equivalent to maximizing $\int_{\mathbf{R}^d} g(z)\log f(z|\hat{\theta})\,\mathrm{d}z$. However, since $g(\cdot)$ is unknown, we must estimate $\int_{\mathbf{R}^d} g(z)\log f(z|\hat{\theta})\,\mathrm{d}z$. A simple estimator of $\int_{\mathbf{R}^d} g(z)\log f(z|\hat{\theta})\,\mathrm{d}z$ is given by the (average) log likelihood

$$\int_{\mathbf{R}^d} \log f(z|\hat{\theta})\,\mathrm{d}\hat{G}(z) = \frac{1}{n}\sum_{\alpha=1}^{n}\log f(X_\alpha|\hat{\theta}),$$

which is obtained by replacing the unknown distribution $G(\cdot)$ with the empirical distribution $\hat{G}(\cdot)$. Usually the average log likelihood $\frac{1}{n}\sum_{\alpha=1}^{n}\log f(X_\alpha|\hat{\theta})$ provides an optimistic assessment (overestimation) of the expected log likelihood $\int_{\mathbf{R}^d} g(z) \log f(z|\hat{\theta})\, \mathrm{d}z$, because the same data are used both to estimate the parameter of the model and to evaluate $\int_{\mathbf{R}^d} g(z) \log f(z|\hat{\theta})\, \mathrm{d}z$. We therefore consider the bias correction of the average log likelihood $\frac{1}{n}\sum_{\alpha=1}^{n} \log f(X_\alpha|\hat{\theta})$. The asymptotic bias of the log likelihood in the estimate of the expected log likelihood is given by

$$E_G\left[\frac{1}{n}\sum_{\alpha=1}^{n}\log f(X_\alpha|\hat{\theta}) - \int_{\mathbf{R}^d}\log f(z|\hat{\theta})\,\mathrm{d}G(z)\right] = \frac{1}{n}b(\theta_0) + o\left(\frac{1}{n}\right),$$

where expectation is taken over the true distribution $G(\cdot)$ and a particular $\theta_0 \in \Theta$. If the bias $b(\theta_0)$ can be estimated by an appropriate procedure, then the unbiased estimator of the expected log likelihood is given by

$$\frac{1}{n}\sum_{\alpha=1}^{n}\log f(X_\alpha|\hat{\theta}) - \frac{1}{n}b(\hat{\theta}).$$

Thus, we choose a statistical model for which the value of

$$\mathrm{IC}(\mathbf{X}_n;\hat{\theta}) = \frac{1}{n}\sum_{\alpha=1}^{n}\log f(X_\alpha|\hat{\theta}) - \frac{1}{n}b(\hat{\theta})$$

is maximized among competing models. The bias corrected log likelihood $\mathrm{IC}(\mathbf{X}_n;\hat{\theta})$ is commonly called an ‘information criterion’. For details, see Konishi and Kitagawa (1996).

Akaike’s information criterion AIC (Akaike, 1973, 1974) is a model evaluation-selection tool based on minimizing the Kullback–Leibler divergence between the fitted model and the true model. It can be obtained under the 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.

With the development of various modeling techniques, the construction of criteria capable of evaluating various types of statistical models has been required. Several attempts to construct information criteria which work for various types of statistical models have been made, and the proposed criteria have been examined from theoretical and practical aspects (cf. Shibata, 1980, 1981; Barron, 1986, 1989; Knight, 1989; Hall, 1990; Hurvich and Tsai, 1993, 1995; Burman and Nolan, 1995; Laud and Ibrahim, 1995; Portnoy, 1997; Burnham and Anderson, 1998; Yang and Barron, 1998; Barron et al., 1999, and references therein). In particular, Takeuchi (1976) derived Takeuchi’s information criterion TIC from the assumptions (i) and (ii) in the misspecified model. Konishi and Kitagawa (1996) proposed generalized information criteria GIC under the assumption (i) and with functional-type estimators instead of the assumption (ii).

There seems to be no doubt that the expectation-unbiasedness is a very handy unbiasedness. Nevertheless, from decision theoretic aspects, it does not appear that there exists a decisive basis for approving only expectation-unbiasedness in case of the information criterion. The expectation-unbiasedness corresponds to a quadratic loss, and in a natural way it can be extended to other loss functions, absolute loss, L^p-loss, etc. In fact, as we will discuss later, it is possible to construct a median-unbiased information criterion. All existing criteria including AIC, TIC and GIC modify an estimator of the expected log likelihood to cancel the expectation-bias in the second order. In other words, obtaining the information criteria is equivalent to finding the bias term $b_1(\cdot)$ such that

$$E_G\left[\sqrt{n}\left(\frac{1}{n}\sum_{\alpha=1}^{n}\log f(X_\alpha|\hat{\theta}) - \frac{1}{n}b_1(\hat{\theta}) - \int_{\mathbf{R}^d}\log f(z|\hat{\theta})\,\mathrm{d}G(z)\right)\right] = o\left(\frac{1}{\sqrt{n}}\right).$$

We then obtain the expectation-unbiased information criterion $\frac{1}{n}\sum_{\alpha=1}^{n}\log f(X_\alpha|\hat{\theta}) - \frac{1}{n}b_1(\hat{\theta})$ as an expectation-unbiased estimator of $\int_{\mathbf{R}^d}\log f(z|\hat{\theta})\,\mathrm{d}G(z)$. Though the first order asymptotic theory was sufficient for the derivation of the expectation-unbiased information criteria, for the median-unbiasedness, it is necessary to consider the second order asymptotic expansion of the distribution of the error of an estimator of the expected log likelihood. More precisely, we need to find the bias term $b_2(\cdot)$ such that

$$P\left[\sqrt{n}\left(\frac{1}{n}\sum_{\alpha=1}^{n}\log f(X_\alpha|\hat{\theta}) - \frac{1}{n}b_2(\hat{\theta}) - \int_{\mathbf{R}^d}\log f(z|\hat{\theta})\,\mathrm{d}G(z)\right) > 0\right] = \frac{1}{2} + o\left(\frac{1}{\sqrt{n}}\right)$$

and

$$P\left[\sqrt{n}\left(\frac{1}{n}\sum_{\alpha=1}^{n}\log f(X_\alpha|\hat{\theta}) - \frac{1}{n}b_2(\hat{\theta}) - \int_{\mathbf{R}^d}\log f(z|\hat{\theta})\,\mathrm{d}G(z)\right) < 0\right] = \frac{1}{2} + o\left(\frac{1}{\sqrt{n}}\right).$$

We then obtain the median-unbiased information criterion $\frac{1}{n}\sum_{\alpha=1}^{n}\log f(X_\alpha|\hat{\theta}) - \frac{1}{n}b_2(\hat{\theta})$ as a second order asymptotically median-unbiased estimator of $\int_{\mathbf{R}^d}\log f(z|\hat{\theta})\,\mathrm{d}G(z)$. For details of the second order asymptotically median-unbiased estimator, see Akahira and Takeuchi (1981). The first goal of this article is to formulate the model selection problem in terms of the higher order asymptotic theory in a unified way, and to show other possibilities than the usual expectation-unbiased criteria, with the median-unbiased information criterion.

Recently, together with the development of the statistical inference for stochastic processes, the problem of the model selection for stochastic processes has been important both in theory and in applications to natural sciences, neural networks, engineering, economics, etc. It is known that the asymptotic expansion is, mainly in independent cases, a promising tool to investigate problems in the higher order statistical inference. Battacharya and Ghosh (1978) established the validity of the asymptotic expansion for a functional of an independent sequence. Götze and Hipp (1983, 1994) gave a valid asymptotic expansion of the distribution of a functional of a discrete-time process under a mixing condition and a conditional type of Cramér condition. Furthermore, for an ϵ-Markov process, Kusuoka and Yoshida (2000) presented a valid asymptotic expansion of the distribution of an additive functional.

With asymptotic expansions, we are now able to reformulate the model selection problem in the light of the higher order asymptotic theory, and to extend objects of consideration to more general stochastic models with continuous time parameter. Thus, the second goal of this article is to propose information criteria which work (i) for stochastic processes with continuous-time parameter, (ii) for various estimators, for example M-estimator, and (iii) for misspecified cases.

The organization of the article is as follows. In Section 2, we state our main results. By using the asymptotic expansion of the distribution of an estimator based on the conditional Kullback–Leibler divergence for mixing processes with continuous-time parameter, a general theory is developed and two information criteria are proposed. In Section 3, the information criteria proposed are applied to diffusion processes. Section 4 presents proofs of the results.

2. Asymptotic Expansion and Information Criteria

Let $(\mathcal{X}_T, \mathcal{A}_T)$ be a measurable space for each $T > 0$. Given a probability space $(\Omega, \mathcal{F}, P)$, let $\mathbf{X}_T$ denote an $\mathcal{X}_T$-valued random variable with an unknown distribution $Q_T(\cdot) = P(\mathbf{X}_T^{-1}(\cdot))$ having a probability density function $q_T(\cdot)$ with respect to a reference measure. Let $\hat{\theta}_T\colon (\mathcal{X}_T, \mathcal{A}_T) \to \Theta$ 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 the 'future' observation $\mathbf{X}_T(\omega') = z$ (for $\omega' \in \Omega$) can be constructed by replacing the unknown parameter θ with $\hat{\theta}_T$.

As a fundamental basis for information criteria, we use the concept of model selection based on minimizing the Kullback–Leibler information $I\{Q_T; P_{T,\hat{\theta}_T}\}$, where

$$\begin{aligned} I\{Q_T; P_{T,\hat{\theta}_T}\} &:= \int_{\mathcal{X}_T} \log q_T(z) Q_T(\mathrm{d}z) - \int_{\mathcal{X}_T} \log f_T\big(z, \hat{\theta}_T(\mathbf{X}_T(\omega))\big) Q_T(\mathrm{d}z) \\ &= \int_{\Omega} \log q_T(\mathbf{X}_T(\omega')) P(\mathrm{d}\omega') - \\ &\quad - \int_{\Omega} l_T\big(\mathbf{X}_T(\omega'), \hat{\theta}_T(\mathbf{X}_T(\omega))\big) P(\mathrm{d}\omega'), \end{aligned} \tag{1}$$

and $l_T(x, \theta) = \log f_T(x, \theta)$.

The first term on the right-hand side of (1) does not depend on the statistical model and only the second term may be taken into account. A simple estimator of the expected log likelihood, that is, $\int_\Omega l_T\big(\mathbf{X}_T(\omega'), \hat\theta_T(\mathbf{X}_T(\omega))\big)P(\mathrm{d}\omega')$, is given by $l_T(\mathbf{X}_T(\omega), \hat\theta_T(\mathbf{X}_T(\omega)))$. Let

$$
\begin{aligned}
\Delta_T &:= l_T(\mathbf{X}_T(\omega), \hat\theta_T(\mathbf{X}_T(\omega))) - \int_\Omega l_T(\mathbf{X}_T(\omega'), \hat\theta_T(\mathbf{X}_T(\omega)))P(\mathrm{d}\omega'),\\
\Delta_T^* &:= \Delta_T - b(\hat\theta_T(\mathbf{X}_T(\omega))),\\
\bar\Delta_T &:= r_T\Delta_T,\\
\bar\Delta_T^* &:= r_T\Delta_T^* = \bar\Delta_T - r_T b(\hat\theta_T(\mathbf{X}_T(\omega))),
\end{aligned}
$$

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

In this section, we consider the second order asymptotic expansion of the distribution of $\bar\Delta_T^*$. First of all, in order to explain the ideas heuristically, we assume that there exists a parameter $\theta_0 \in \Theta$ such that

$$
{r_T}^{-1}(\hat\theta_T - \theta_0) = \bar\zeta_T^{(0)} + o_p(1) \tag{2}
$$

for a functional $\bar\zeta_T^{(0)}$ satisfying conditions put later.

Set

$$
\begin{aligned}
Z_T^{(0)} &:= l_T(\mathbf{X}_T(\omega), \theta_0) - \int_\Omega l_T(\mathbf{X}_T(\omega'), \theta_0)P(\mathrm{d}\omega'),\\
\bar Z_T^{(0)} &:= r_T Z_T^{(0)},\\
Z_T^{(1)} &:= \partial_\theta l_T(\mathbf{X}_T(\omega), \theta_0) - \int_\Omega \partial_\theta l_T(\mathbf{X}_T(\omega'), \theta_0)P(\mathrm{d}\omega'), \qquad \partial_\theta = \frac{\partial}{\partial\theta},\\
\bar Z_T^{(1)} &:= r_T Z_T^{(1)}.
\end{aligned}
$$

Intuitively, expanding $l_T(\mathbf{X}_T(\omega), \hat\theta_T(\mathbf{X}_T(\omega)))$ and $\int_\Omega l_T(\mathbf{X}_T(\omega'), \hat\theta_T(\mathbf{X}_T(\omega)))$ $P(\mathrm{d}\omega')$ in a Taylor series around θ_0 and substituting (2) into the resulting expansion, we obtain stochastic expansions as follows:

$$
\begin{aligned}
\Delta_T &= l_T(\mathbf{X}_T(\omega), \theta_0) - \int_\Omega l_T(\mathbf{X}_T(\omega'), \theta_0)P(\mathrm{d}\omega') + \Bigg\{\partial_\theta l_T(\mathbf{X}_T(\omega), \theta_0) -\\
&\quad - \int_\Omega \partial_\theta l_T(\mathbf{X}_T(\omega'), \theta_0)P(\mathrm{d}\omega')\Bigg\}'(\hat\theta_T(\mathbf{X}_T(\omega)) - \theta_0) +\\
&\quad + \frac{1}{2}(\hat\theta_T(\mathbf{X}_T(\omega)) - \theta_0)'\Bigg\{(\partial_\theta)^2 l_T(\mathbf{X}_T(\omega), \theta_0) -\\
&\quad - \int_\Omega (\partial_\theta)^2 l_T(\mathbf{X}_T(\omega'), \theta_0)P(\mathrm{d}\omega')\Bigg\}(\hat\theta_T(\mathbf{X}_T(\omega)) - \theta_0) + o_p(1)\\
&= Z_T^{(0)} + \bar Z_T^{(1)\prime}\bar\zeta_T^{(0)} + o_p(1),
\end{aligned}
$$

where A' denotes the transposition of A for $A \in \mathbf{R}^p$. Here we assumed a central limit theorem and a law of large numbers.

Let R_T^* be the remainder term in the expansion $\bar{\Delta}_T^*$:

$$\bar{\Delta}_T^* = \bar{Z}_T^{(0)} + r_T(\bar{Z}_T^{(1)\prime}\bar{\zeta}_T^{(0)} - b(\theta_0)) + R_T^*. \tag{3}$$

Remark. (i) For an M-estimator $\hat{\theta}_T$, in the same way as Sakamoto and Yoshida (1998, 1999), it is possible to show that for some $E_0 > 1$, $0 < \varepsilon_0 < 1$ and r.v. $\bar{\zeta}_T^{(0)}$, $r_T{}^{-1}(\hat{\theta}_T - \theta_0) = \bar{\zeta}_T^{(0)} + R_T$, where $P[\hat{\theta}_T$ exists uniquely in $U(\theta_0, r_T^{\varepsilon_0})$ and $|R_T| \leqslant r_T^{\varepsilon_0}] = 1 - o(r_T^{E_0})$ and $U(\theta_0, r_T^{\varepsilon_0})$ is the closed ball of radius $r_T^{\varepsilon_0}$ centered at θ_0.

(ii) From (i) and some regularity conditions, we can also show that there exist constants $E > 1$ and $\varepsilon > 1$ such that $P[|R_T^*| \leqslant r_T^{\varepsilon}] = 1 - o(r_T^E)$ for R_T^* in (3).

2.1. ϵ-MARKOV MODEL

We describe the underlying probabilistic structure of the random variables $\bar{Z}_T^{(0)}$, $\bar{Z}_T^{(1)}$ and $\bar{\zeta}_T^{(0)}$ in (3). Let $(\Omega, \mathcal{F}, P)$ be a probability space, $Y = (Y_t)_{t \in \mathbf{R}_+}$ an $\mathbf{R}^{d_2}$-valued càdlàg process defined on Ω, and $X = (X_t)_{t \in \mathbf{R}_+}$ an $\mathbf{R}^{d_1}$-valued càdlàg process defined on Ω. Suppose that for any $t \in \mathbf{R}_+$, $\mathcal{B}_{[0,t]}^{X,Y}$ is independent of $\mathcal{B}_{[t,\infty)}^{dX}$, where

$$\mathcal{B}_{[0,t]}^{X,Y} = \sigma[X_u, Y_u : u \in [0,t]] \vee \mathcal{N}, \quad \mathcal{B}_{[t,\infty)}^{dX} = \sigma[X_s - X_u : s, u \in [t, \infty)],$$

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

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

and

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

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

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

where $\mathcal{F}(\mathcal{A})$ denotes the set of all $\mathcal{A}$-measurable functions for sub σ-field $\mathcal{A}$ of $\mathcal{F}$. If the process Y satisfies the above condition, it is called an ϵ-Markov process driven by X. Moreover, assume that for any $T > 0$, $\bar{Z}_T \equiv (\bar{Z}_T^{(0)}, \bar{Z}_T^{(1)}, \bar{\zeta}_T^{(0)})$ in (3) is a normalized functional of an additive functional Z_T, that is, $\bar{Z}_T = r_T Z_T$ for an $\mathbf{R}^{d_3}$-valued process $Z = (Z_t)_{t \in \mathbf{R}_+}$ satisfying $Z_0 \in \mathcal{F}\mathcal{B}_{[0]}$ and

$$Z_t^s := Z_t - Z_s \in \mathcal{F}\mathcal{B}_{[s,t]}, \quad \text{for every} \quad s, t \in \mathbf{R}_+, 0 \leqslant s \leqslant t,$$

where the dimension of $\bar{Z}_T^{(0)}$ is one, and both $\bar{Z}_T^{(1)}$ and $\bar{\zeta}_T^{(0)}$ are p-dimensional functionals, that is $d_3 = 2p + 1$.

Let $P[f]$ be the expectation of f with respect to P and $P[f|\mathcal{B}]$ denotes the conditional expectation of f given a sub-σ-field $\mathcal{B}$ of $\mathcal{A}$ with respect to P. In order to present the asymptotic expansion of $\bar{\Delta}_T^*$, we will assume the following conditions given in Kusuoka and Yoshida (2000).

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

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

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

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

For illustrations of the ϵ-Markov model, here are two examples in Kusuoka and Yoshida (2000) as follows:

EXAMPLE 1. Let $\{Y_n\}_{n\in\mathbf{Z}+}$ denote an $\mathbf{R}^{d_2}$-valued m-Markov chain (non-linear time series model) satisfying the stochastic equation

$$Y_n = S_n(Y_{n-1}, \dots, Y_{n-m}, \xi_n), \qquad n \geqslant m,$$

where $\{\xi_n\}_{n\geqslant m}$ is an $\mathbf{R}^{d_1}$-valued independent random sequence and independent of $\{Y_n\}_{n=0}^{m-1}$. Define $Z_n = \sum_{j=1}^n f_j(Y_j, \xi_j)$ and $X_n = \sum_{j=1}^n \xi_j$. Obviously, the process $\{X_n, Y_n, Z_n\}_{n\in\mathbf{Z}_+}$ can be embedded into a process $\{X_t, Y_t, Z_t\}_{t\in\mathbf{R}_+}$ with continuous time parameter as $X_t = X_{[t]}$, $Y_t = Y_{[t]}$ and $Z_t = Z_{[t]}$. Then Y is an $(m-1)$-Markov process driven by the process X with independent increments.

EXAMPLE 2. Let $A \in C^\infty(\mathbf{R}^{d_2}; \mathbf{R}^{d_2})$, $A' \in C^\infty(\mathbf{R}^{d_2}; \mathbf{R}^{d_3})$, $B \in C^\infty(\mathbf{R}^{d_2}; \mathbf{R}^{d_2} \otimes \mathbf{R}^r)$, $B' \in C^\infty(\mathbf{R}^{d_2}; \mathbf{R}^{d_3}\otimes\mathbf{R}^r)$, $C \in C^\infty(\mathbf{R}^{d_2}\times E; \mathbf{R}^{d_2})$ and $C' \in C^\infty(\mathbf{R}^{d_2}\times E; \mathbf{R}^{d_3})$, where E is an open set in $\mathbf{R}^b$. Suppose that $\{Y_t, Z_t\}_{t\in\mathbf{R}_+}$ is a stochastic process defined as a strong solution of the stochastic integral equation with jumps:

$$\begin{aligned} Y_t &= Y_0 + A(Y_-) * t + B(Y_-) * w_t + C(Y_-) * \tilde{\mu}_t, \\ Z_t &= Z_0 + A'(Y_-) * t + B'(Y_-) * w_t + C'(Y_-) * \tilde{\mu}_t, \end{aligned}$$

where Z_0 is $\sigma[Y_0]$-measurable, w is an r-dimensional Wiener process, and $\tilde{\mu}$ is a compensated Poisson random measure on $\mathbf{R}_+ \times E$ with intensity $\mathrm{d}t \otimes \lambda(\mathrm{d}x)$, λ being the Lebesgue measure on E. Under usual regularity conditions, it is possible to regard (Y_t, Z_t) as smooth functionals over the canonical space $\Omega = \{(y_0, w, \mu)\}$, where μ is the integer-valued random measure on $\mathbf{R}_+ \times E$. Let $\mathcal{F}$ be the σ-field generated by the canonical maps on Ω. In this case, the process X_t may be taken as $X_t = (w_t, \mu_t(g_i); i \in \mathbf{N})$, where (g_i) denotes a countable measure determining family over E. Then Y is a Markov process, that is, $\epsilon = 0$, driven by X with independent increments. For more details, see III.6 and IV.10 of Bichteler *et al.* (1987) and Example 2 of Kusuoka and Yoshida (2000).

2.2. MALLIAVIN CALCULUS

We will make use of the nondegeneracy of the Malliavin covariance instead of the conditional type Cramér condition to ensure the regularity of distributions. Taking account of semimartingales with jumps, we here adopted the formulation of the Malliavin calculus by Bichteler et al. (1987).

Let $C^2_\uparrow(\mathbf{R}^n)$ be the set of all functions f of class $C^2(\mathbf{R}^n)$ such that f and all of its derivatives have polynomial growth. Given a probability space $(\Omega, \mathcal{B}, \Pi)$, a linear operator $\mathcal{L}$ on $\mathcal{D}(\mathcal{L}) \subset \cap_{p>1} L^p(\Pi)$ into $\cap_{p>1} L^p(\Pi)$ is called a Malliavin operator if the following conditions are satisfied:

(1) $\mathcal{B}$ is generated by $\mathcal{D}(\mathcal{L})$.
(2) For $f \in C^2_\uparrow(\mathbf{R}^n)$, $n \in \mathbf{N}$, and $F \in \mathcal{D}(\mathcal{L})^n$, $f \circ F \in \mathcal{D}(\mathcal{L})$.
(3) For any $F, G \in \mathcal{D}(\mathcal{L})$, $\Pi[F\mathcal{L}G] = \Pi[G\mathcal{L}F]$.
(4) The bilinear operator $\Gamma_\mathcal{L}$ on $\mathcal{D}(\mathcal{L}) \times \mathcal{D}(\mathcal{L})$ associated with $\mathcal{L}$ by $\Gamma_\mathcal{L}(F, G) = \mathcal{L}(FG) - F\mathcal{L}G - G\mathcal{L}F$ is nonnegative definite. In other words, for $F \in \mathcal{D}(\mathcal{L})$, $\mathcal{L}(F^2) \geqslant 2F\mathcal{L}F$.
(5) For $F = (F^1, \cdots, F^n) \in \mathcal{D}(\mathcal{L})^n$, $n \in \mathbf{N}$, and $f \in C^2_\uparrow(\mathbf{R}^n)$,

$$\mathcal{L}(f \circ F) = \sum_{i=1}^n (\partial_i f \circ F)\mathcal{L}F^i + \frac{1}{2}\sum_{i,j=1}^n (\partial_i \partial_j f \circ F)\Gamma_\mathcal{L}(F^i, F^j).$$

Fix a Malliavin operator $(\mathcal{L}, \mathcal{D}(\mathcal{L}))$. For $p \geqslant 2$, define $\|F\|_{D_{2,p}} = \|F\|_p + \|\mathcal{L}F\|_p + \|\Gamma^{1/2}_\mathcal{L}(F, F)\|_p$. Let $D_{2,p}$ denote the completion of $\mathcal{D}(\mathcal{L})$ with respect to $\|\cdot\|_{D_{2,p}}$. Then $(D_{2,p}, \|\cdot\|_{D_{2,p}})$ is a Banach space. The existence of a Malliavin operator leads us to the existence of an integration-by-parts setting (IBPS). Let $D_{2,\infty-} = \cap_{p \geqslant 2} D_{2,p}$. For $F \in D_{2,\infty-}(\mathbf{R}^n) \equiv (D_{2,\infty-})^n$, the Malliavin covariance matrix σ_F of F is defined by $\sigma_F = (\sigma_F^{ij}) = (\Gamma_\mathcal{L}(F^i, F^j))$ for $i, j = 1, \ldots, n$. From Theorem 8–18 of Bichteler et al. (1987) p.107, we have the IBP formula with truncation. For more details of IBP formula with truncation, see Propositions 1 and 2 of Kusuoka and Yoshida (2000).

2.3. PROCESS WITH FINITE AUTOREGRESSION

In addition to two conditions [A1] and [A2] in Subsection 2.1, we require another condition, which is also assumed in Kusuoka and Yoshida (2000), concerned with the regularity of the distribution. Let τ be a fixed constant such that $\tau > \epsilon$. For each $T > 0$, let $[u(i), v(i)]$, $j = 1, \ldots, n(T)$ be sub-intervals of the interval $[0, T]$ such that

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

and that $\inf_{j,T}\{v(j) - u(j)\} \geqslant \tau$, $\sup_{j,T}\{v(j) - u(j)\} < \infty$. The process considered here is a process with finite autoregression; more precisely, we 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, $P[F|\mathcal{B}_{[0,v(j)]}] = P[F|\sigma[\mathcal{Y}_j]]$, 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 $D^{L_j}_{2,p}$, $p \geqslant 2$, denotes the completion of $\mathcal{D}(L_j)$ with respect to $\|\cdot\|_{D^{L_j}_{2,p}}$. Let $C^\infty_B(\mathbf{R}^n)$ be the set of all functions f of class $C^\infty(\mathbf{R}^n)$ such that f and all of its derivatives are bounded. Suppose that for any $f \in C^\infty_B(\mathbf{R}^{(d_1+d_2)m})$ and any $u_0, u_1, \dots, u_m$ satisfying $u(j) - \epsilon \leqslant u_0 \leqslant u_1 \leqslant \cdots \leqslant u_m \leqslant u(j)$, the functional $F = f(X_{u_k} - X_{u_{k-1}}, Y_{u_k} : 1 \leqslant k \leqslant m) \in D^{L_j}_{2,\infty-}$ and $L_j F = 0$. Let $\sigma_{\mathcal{Z}_j}$ be the Malliavin covariance matrix of $\mathcal{Z}_j = (Z^{u(j)}_{v(j)}, \mathcal{Y}_j)$, and suppose that $Z^{u(j)}_{v(j),l}, \mathcal{Y}_{j,k}, \sigma^{pq}_{\mathcal{Z}_j} \in D^{L_j}_{2,\infty-}$, where $Z^{u(j)}_{v(j)} = (Z^{u(j)}_{v(j),l})_{l=1,\dots,d_3}$ and $\sigma^{pq}_{\mathcal{Z}_j} = \Gamma_{L_j}(\mathcal{Z}^p_j, \mathcal{Z}^q_j)$. Note that d_3 is the dimension of $\bar{Z}_T$, that is, $d_3 = 2p + 1$. Suppose $\sup_{j,T} M_j < \infty$. The measurable function $\psi_j\colon (\Omega, \mathcal{B}_{[u(j)-\epsilon, v(j)]}) \to ([0,1], \mathbf{B}([0,1]))$ denotes a truncation functional. Put $S_1[\psi_j; \mathcal{Z}_j] = \{\sigma^{i,k}_{\mathcal{Z}_j}, i, k = 1, \dots, d_3 + M_j, (\Delta_{\mathcal{Z}_j})^{-d_3}\psi_j\}$, and

$$S_{1,j} = \{(\Delta_{\mathcal{Z}_j})^{-d_3}\psi_j, \sigma^{kl}_{\mathcal{Z}_j}, L_j\mathcal{Z}_{j,k}, \Gamma_{L_j}(\sigma^{kl}_{\mathcal{Z}_j}, \mathcal{Z}_{j,m}), \Gamma_{L_j}((\Delta_{\mathcal{Z}_j})^{-d_3}\psi_j, \mathcal{Z}_{j,l})\}$$

for operator L_j, where $\Delta_{\mathcal{Z}_j} = \det \sigma_{\mathcal{Z}_j}$. By using the terminology and notation above, the regularity condition of the distribution is given as follows:

[$A3$] (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) $\liminf_{T\to\infty} n(T)/T > 0$;
(iii) For each $j = 1, \dots, n(T)$, $\mathcal{Z}_j \in (D^{L_j}_{2,\infty-})^{d_3+M_j}$, $S_1[\psi_j; \mathcal{Z}_j] \subset D^{L_j}_{2,\infty-}$, and for any $p > 1$, $\bigcup_{j=1,\dots,n(T),T>0} S_{1,j}$ is bounded in $L^p(P)$.

With the help of Kusuoka and Yoshida (2000), we see that models in Examples 1 and 2 satisfy the finite autoregression condition. For details, see Examples 1′ and 2′ of Kusuoka and Yoshida (2000).

ASSUMPTION 1. $\bar{Z}_T = (\bar{Z}^{(0)}_T, \bar{Z}^{(1)}_T, \bar{\zeta}^{(0)}_T)$ satisfies conditions [$A1$], [$A2$] and [$A3$].

Let us prepare some notations. Define the k-th cumulant $\lambda^{\alpha_1\cdots\alpha_k}_T$ of $\bar{Z}_T$ by

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

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

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

where $\phi(z; \sigma_{\alpha\beta})$ is the density function of the normal distribution with mean 0 and covariance matrix $(\sigma_{\alpha\beta})$. Denote by Σ_T 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

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

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

$$\Sigma_T = \mathrm{Cov}[\bar{Z}_T^{(0)}, \bar{Z}_T^{(1)}, \bar{\zeta}_T^{(0)}] = \begin{bmatrix} \Sigma_T^{(00)} & (\Sigma_T^{(10)})' & (\Sigma_T^{(20)})' \\ \Sigma_T^{(10)} & \Sigma_T^{(11)} & \Sigma_T^{(12)} \\ \Sigma_T^{(20)} & (\Sigma_T^{(12)})' & \Sigma_T^{(22)} \end{bmatrix} \quad \text{(say)}.$$

The r-th cumulant $\chi_{T,r}(u)$ of $\bar{Z}_T$ is defined by

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

Next, define functions $\tilde{P}_{T,r}(u)$ by the formal Taylor expansion

$$\exp\left(\sum_{r=2}^{\infty} r!^{-1}\epsilon^{r-2}\chi_{T,r}(u)\right) = \exp\left(\frac{1}{2}\chi_{T,2}(u)\right) + \sum_{r=1}^{\infty} \epsilon^r T^{-r/2}\tilde{P}_{T,r}(u). \tag{4}$$

Let $\hat{\Psi}_{T,k}(u)$ be the k-th partial sum of the right-hand side of (4) with $\epsilon = 1$:

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

Finally, for $T > 0$ and $k \in \mathbf{N}$, a signed-measure $\hat{\Psi}_{T,k}$ is defined as the Fourier inversion of $\hat{\Psi}_{T,k}(u)$. In the sequel, we will assume that the second cumulant $\chi_{T,2}(u)$ converges to a negative definite quadratic form $-u'\Sigma u$ as $T \to \infty$. Fix a symmetric matrix $\hat{\Sigma}$ satisfying $\Sigma < \hat{\Sigma}$. For $M > 0$ and $\gamma > 0$, the set $\mathcal{E}(M, \gamma)$ of measurable functions from $\mathbf{R} \to \mathbf{R}$ is defined by

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

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

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

Let $\tilde{\Delta}_T^* = \bar{Z}_T^{(0)} + r_T(\bar{Z}_T^{(1)\prime}\bar{\zeta}_T^{(0)} - b(\theta_0))$. From (3) it follows that $\bar{\Delta}_T^* = \tilde{\Delta}_T^* + R_T^*$. Theorem 5 of Kusuoka and Yoshida (2000), together with the formula of Sakamoto and Yoshida (1999), gives an expansion of $\tilde{\Delta}_T^*$ as follows.

THEOREM 1 (Kusuoka and Yoshida (2000), Sakamoto and Yoshida (1999)). *Let $M, \gamma > 0$. Suppose that Assumption 1 holds true. Then for any $K > 0$,*

(1) *there exist constants* $\delta > 0$ *and* $c > 0$ *such that for any function* $f \in \mathcal{E}(M, \gamma)$,

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

where $\varepsilon_T = o(r_T^{((1+\delta)\wedge K)})$ *depends on* $\mathcal{E}(M, \gamma)$.

(2) *The signed-measure* $\mathrm{d}\Psi_{T,1}$ *has a density* $\mathrm{d}\Psi_{T,1}(z)/\mathrm{d}z = q_{T,1}(z)$ *with*

$$\begin{aligned} q_{T,1}(z^{(0)}) \; = \; & \int_{\mathbf{R}^{2p}} p_{T,1}(z)\,\mathrm{d}z^{(1)}\,\mathrm{d}z^{(2)} \; - \\ & - r_T \partial_{z^{(0)}} \left[\int_{\mathbf{R}^{2p}} \{z^{(1)\prime} z^{(2)} - b(\theta_0)\} \phi(z;\, \Sigma_T)\,\mathrm{d}z^{(1)}\,\mathrm{d}z^{(2)} \right], \end{aligned}$$

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

Remark 2. In Theorem 1, we assumed the non-degeneracy of the covariance matrix of $r_T Z_T$. Even if $\mathrm{Cov}(r_T Z_T)$ 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 Theorem 1. For more details, see Sakamoto and Yoshida (2000).

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

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

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

Remark 3. Suppose that there exist constants $K' > 1$ and $m > 1$ such that $P[|R_T^*| \leqslant r_T^{K'}] = 1 - o(r_T^m)$. Moreover, suppose that $\sup_{T>1} ||r_T \bar{\Delta}_T^*||_{L^p} < \infty$ for some $p > 1$, and that $m(p-1)/p - \gamma > 1$. Then, it is possible to show that there exists a constant $\alpha > 1$ such that $\sup_{f \in \mathcal{E}(M,\gamma)} \left| P[(f(\bar{\Delta}_T^*) - f(\tilde{\Delta}_T^*)) 1_{\{|R_T^*| > r_T^{K'}\}}] \right| = o(r_T^\alpha)$.

We then obtain a second order asymptotic expansion of the distribution of $\bar{\Delta}_T^*$.

THEOREM 2. *Let* $M, \gamma > 0$. *Suppose that Assumptions* 1 *and* 2 *hold true. Then,*

(1) *there exist constants* $\delta > 0$ *and* $\tilde{c} > 0$ *such that for any function* $f \in \mathcal{E}(M, \gamma)$,

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

where $\tilde{\varepsilon}_T = o(r_T^{((1+\delta)\wedge\alpha)})$ depends on $\mathcal{E}(M, \gamma)$.

(2) *The signed-measure* $\mathrm{d}\Psi_{T,1}$ *has the same density* $\mathrm{d}\Psi_{T,1}(z)/\mathrm{d}z = q_{T,1}(z)$ *as in Theorem* 1.

By using Theorem 2, we have an explicit expression for a second order asymptotic expansion of the distribution of $\bar{\Delta}_T^*$.

THEOREM 3. *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 $f \in \mathcal{E}(M, \gamma)$,*

$$\begin{aligned} P[f(\bar{\Delta}_T^*)] \;=\; & \int_{\mathbf{R}} f(z^{(0)})\phi(z^{(0)}; \Sigma_T^{(00)})\,\mathrm{d}z^{(0)} + \\ & + \frac{1}{6}\lambda_T^{000} \int_{\mathbf{R}} f(z^{(0)}) h_3(z^{(0)}; \Sigma_T^{(00)})\phi(z^{(0)}; \Sigma_T^{(00)})\,\mathrm{d}z^{(0)} - \\ & - r_T \int_{\mathbf{R}} f(z^{(0)})\partial_{z^{(0)}} \times \\ & \times \Big[\big\{C_T(z^{(0)}) - b(\theta_0)\big\}\phi(z^{(0)}; \Sigma_T^{(00)})\Big]\,\mathrm{d}z^{(0)} + \rho_T(f), \end{aligned}$$

where

$$\begin{aligned} \rho_T(f) &= \tilde{c}\omega(f, 2r_T^{K'}) + o(r_T^{((1+\delta)\wedge\alpha)}), \\ C_T(z^{(0)}) &= \frac{(\Sigma_T^{(10)})'\Sigma_T^{(20)}}{(\Sigma_T^{(00)})^2}[(z^{(0)})^2 - \Sigma_T^{(00)}] + \mathrm{tr}\Sigma_T^{(12)}. \end{aligned}$$

THEOREM 4. *Suppose that Assumptions* 1 *and* 2 *for some $K' > 1$ and $\alpha > 1$ hold true. Let $b_1(\theta_0) = tr\Sigma_T^{(12)}$. Then*

$$P\left[r_T\Delta_T - r_T b_1(\hat{\theta}_T)\right] = o\,(r_T)\,.$$

Remark 4. There is no need for Theorem 4 to suppose [A3] in Assumption 1.

Since it follows from Theorem 4 that $r_T\Delta_T - r_T b_1(\hat{\theta}_T)$ is asymptotically expectation-unbiased (AEU), we can propose an information criterion based on the asymptotically expectation-bias corrected log likelihood as follows:

Information criterion 1 (in the sense of AEU).

$$\mathrm{IC}_1(\mathbf{X}_T(\omega)) = r_T l_T(\mathbf{X}_T(\omega), \hat{\theta}_T(\mathbf{X}_T(\omega))) - r_T b_1(\hat{\theta}_T(\mathbf{X}_T(\omega))), \tag{5}$$

where $b_1(\theta_0) = \mathrm{tr}\Sigma_T^{(12)}$.

Remark 5. Suppose that the data are independent random samples. Under the assumption that $\hat{\theta}_T$ in (5) is the functional-type estimator in the misspecified model, IC_1 corresponds to GIC. By using the maximum likelihood estimator (MLE) in the misspecified model, IC_1 is equivalent to TIC. Moreover, for the MLE in the correctly specified model, IC_1 corresponds to AIC.

THEOREM 5. *Suppose that Assumptions* 1 *and* 2 *for some $K' > 1$ and $\alpha > 1$ hold true. Let*

$$b_2(\theta_0) \;=\; -\frac{1}{6}{r_T}^{-1}\lambda_T^{000}\frac{1}{\Sigma_T^{(00)}} + \left[\mathrm{tr}\Sigma_T^{(12)} - \frac{(\Sigma_T^{(10)})'\Sigma_T^{(20)}}{\Sigma_T^{(00)}}\right].$$

Then

$$\begin{aligned} P\left[r_T\Delta_T - r_T b_2(\hat{\theta}_T(\mathbf{X}_T(\omega))) > 0\right] &= \tfrac{1}{2} + o\left(r_T\right),\\ P\left[r_T\Delta_T - r_T b_2(\hat{\theta}_T(\mathbf{X}_T(\omega))) < 0\right] &= \tfrac{1}{2} + o\left(r_T\right). \end{aligned}$$

From Theorem 5, it follows that $r_T\Delta_T - r_T b_2(\hat{\theta}_T)$ is the second order asymptotically median-unbiased (second order AMU); see Akahira and Takeuchi (1981). Thus we also propose another information criterion based on the asymptotically median-bias corrected log likelihood as follows:

Information criterion 2 (in the sense of the second order AMU).

$$\mathrm{IC}_2(\mathbf{X}_T(\omega)) = r_T l_T(\mathbf{X}_T(\omega), \hat{\theta}_T(\mathbf{X}_T(\omega))) - r_T b_2(\hat{\theta}_T(\mathbf{X}_T(\omega))), \tag{6}$$

where

$$b_2(\theta_0) = -\frac{1}{6} r_T{}^{-1}\lambda_T^{000}\frac{1}{\Sigma_T^{(00)}} + \left[\mathrm{tr}\Sigma_T^{(12)} - \frac{(\Sigma_T^{(10)})'\Sigma_T^{(20)}}{\Sigma_T^{(00)}}\right].$$

Remark 6. For suitable measurable functions f satisfying two conditions: $\int_{\mathbf{R}} f(x)\phi(x;\,\Sigma_T^{(00)})\,\mathrm{d}x = 0$ and $\int_{\mathbf{R}} f(x)\partial_x\{\phi(x;\,\Sigma_T^{(00)})\}\,\mathrm{d}x \neq 0$, let $b_f(\cdot)$ denote

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

From Theorem 3 and certain regularity conditions it then follows that

$$\mathrm{IC}_f(\mathbf{X}_T(\omega)) = r_T l_T(\mathbf{X}_T(\omega), \hat{\theta}_T(\mathbf{X}_T(\omega))) - r_T b_f(\hat{\theta}_T(\mathbf{X}_T(\omega)))$$

is the f-unbiased information criterion, that is,

$$P\left[f\left(\mathrm{IC}_f(\mathbf{X}_T(\omega)) - r_T\int_{\Omega} l_T(\mathbf{X}_T(\omega'), \hat{\theta}_T(\mathbf{X}_T(\omega)))P(\mathrm{d}\omega')\right)\right] = o(r_T).$$

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.

3. Application to Diffusion Processes

We present an application of the results in Section 2 to diffusion processes.

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

$$\begin{aligned} \mathrm{d}X_t &= V_0(X_t)\,\mathrm{d}t + V(X_t)\,\mathrm{d}w_t, \qquad t \in [0, T], \\ X_0 &= x_0, \end{aligned} \tag{7}$$

where X_0 is the initial random variable (r.v.), $V = (V_1, \dots, V_r)$ 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

$$\begin{aligned} \mathrm{d}X_t &= \tilde{V}_0(X_t, \theta)\,\mathrm{d}t + \tilde{V}(X_t)\,\mathrm{d}\tilde{w}_t, \qquad t \in [0, T], \\ X_0 &= x_0, \end{aligned} \tag{8}$$

where θ is a p-dimensional unknown parameter in Θ, X_0 is the initial r.v., $\tilde{V} = (\tilde{V}_1, \dots, \tilde{V}_{\tilde{r}})$ 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 requires to be estimated from the observation $\mathbf{X}_T = \{X_t; 0 \leqslant t \leqslant T\}$.

Let $\mathbf{X}_T^\theta$ be the solution of the stochastic differential Equation (8) for θ. We assume that $\mathbf{X}_T^\theta$ is a stationary, strong mixing diffusion process with stationary distribution ν_θ. Since the likelihood function of θ is defined by

$$\begin{aligned} L_T(\mathbf{X}_T^\theta, \theta) &:= \frac{\mathrm{d}\nu_\theta(X_0)}{\mathrm{d}x} \exp\left\{ \int_0^T \tilde{V}_0'(\tilde{V}\tilde{V}')^{-1}(X_t, \theta)\,\mathrm{d}X_t - \right. \\ &\quad \left. - \frac{1}{2} \int_0^T \tilde{V}_0'(\tilde{V}\tilde{V}')^{-1}\tilde{V}_0(X_t, \theta)\,\mathrm{d}t \right\}, \end{aligned}$$

the log likelihood function is given by

$$l_T(\mathbf{X}_T^\theta, \theta) = \tilde{a}(X_0, \theta) + \int_0^T \tilde{b}(X_t, \theta)\,\mathrm{d}X_t + \int_0^T \tilde{c}(X_t, \theta)\,\mathrm{d}t, \tag{9}$$

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

From (7) and (9), the log likelihood function under the true model is given by

$$l_T(\mathbf{X}_T, \theta) = a(X_0, \theta) + \int_0^T b(X_t, \theta)\,\mathrm{d}w_t + \int_0^T c(X_t, \theta)\,\mathrm{d}t, \tag{10}$$

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

$$\Psi_T(\mathbf{X}_T^\theta, \theta) := \tilde{A}(X_0, \theta) + \int_0^T \tilde{B}(X_t, \theta)\,\mathrm{d}X_t + \int_0^T \tilde{C}(X_t, \theta)\,\mathrm{d}t, \tag{11}$$

where $\tilde{A}$, $\tilde{B}$, $\tilde{C}$ are given functions. From (7) and (11), the functional Ψ_T under the true model is given by

$$\Psi_T(\mathbf{X}_T, \theta) = A(X_0, \theta) + \int_0^T B(X_t, \theta)\,\mathrm{d}w_t + \int_0^T C(X_t, \theta)\,\mathrm{d}t,$$

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, that is,

$$\hat{\theta}_T := \operatorname*{argmax}_\theta \Psi_T(\mathbf{X}_T, \theta). \tag{12}$$

DEFINITION 1.

$$\theta_0 = \operatorname*{argmax}_\theta \int_{\mathbf{R}^d} C(x, \theta)\nu(\mathrm{d}x).$$

Under certain regularity conditions, we can validate the following argument (cf. Sakamoto and Yoshida, 1999).

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

$$r_T{}^{-1}(\hat{\theta}_T - \theta_0) = -\left[r_T{}^2\,(\partial_\theta)^2\,\Psi_T(\mathbf{X}_T, \theta_0)\right]^{-1} r_T \partial_\theta \Psi_T(\mathbf{X}_T, \theta_0) + o_p(1),$$

so that we define

$$\begin{aligned}\bar{\zeta}_T^{(0)} &= -\nu\left((\partial_\theta)^2 C(\cdot, \theta_0)\right)^{-1} \times \\ &\quad \times \left[r_T \int_0^T \partial_\theta B(X_t, \theta_0)\,\mathrm{d}w_t + r_T \int_0^T \partial_\theta C(X_t, \theta_0)\,\mathrm{d}t\right]. \end{aligned} \tag{13}$$

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

$$\begin{aligned}\bar{Z}_T^{(0)} &:= r_T\,[l_T(\mathbf{X}_T, \theta_0) - P[l_T(\mathbf{X}_T, \theta_0)]] \\ &= r_T\left[\alpha_1(X_0, \theta_0) + \int_0^T b(X_t, \theta_0)\,\mathrm{d}w_t + \right. \\ &\quad \left. + \int_0^T \{c(X_t, \theta_0) - \nu\,(c(\cdot, \theta_0))\}\,\mathrm{d}t\right], \end{aligned} \tag{14}$$

$$\begin{aligned}\bar{Z}_T^{(1)} &:= r_T\,[\partial_\theta l_T(\mathbf{X}_T, \theta_0) - P[\partial_\theta l_T(\mathbf{X}_T, \theta_0)]] \\ &= r_T\left[\alpha_2(X_0, \theta_0) + \int_0^T \partial_\theta b(X_t, \theta_0)\,\mathrm{d}w_t + \right. \\ &\quad \left. + \int_0^T \{\partial_\theta c(X_t, \theta_0) - \nu\,(\partial_\theta c(\cdot, \theta_0))\}\;\mathrm{d}t\right], \end{aligned} \tag{15}$$

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

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

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

From Itô's formula, we see

$$G_f(X_T) - G_f(X_0) = \int_0^T \partial G_f(X_t)V(X_t)\,\mathrm{d}w_t + \int_0^T f(X_t)\,\mathrm{d}t. \tag{16}$$

Define f_i $(i = 0, 1, 2)$ by $f_0(x) = c(x,\theta_0) - \nu(c(\cdot,\theta_0))$, $f_1(x) = \partial_\theta c(x,\theta_0) - \nu(\partial_\theta c(\cdot,\theta_0))$ and $f_2(x) = \partial_\theta C(x,\theta_0)$, respectively. Let $C_\uparrow^\infty(\mathbf{R}^d)$ be the set of all functions f of class $C^\infty(\mathbf{R}^d)$ such that f and all of its derivatives have polynomial growth.

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

ASSUMPTION 4. For f_0, there exists $G_{f_0} \in C_\uparrow^\infty(\mathbf{R}^d)$ such that $\mathcal{A}G_{f_0} = f_0$. Moreover, for $f_3(x) = \|\xi^{(0)}(x)\|^2 - \nu(\|\xi^{(0)}(\cdot)\|^2)$, there exists $G_{f_3} \in C_\uparrow^\infty(\mathbf{R}^d)$ such that $\mathcal{A}G_{f_3} = f_3$. Here $\xi^{(0)}(x) = b(x,\theta_0) - \partial G_{f_0}(x)V(x)$.

Let $\zeta^{(0)}(x) = -\nu\left((\partial_\theta)^2 C(\cdot,\theta_0)\right)^{-1}\left\{\partial_\theta B(x,\theta_0) - \partial G_{f_2}(x)V(x)\right\}$ and $\xi^{(1)}(x) = \partial_\theta b(x,\theta_0) - \partial G_{f_1}(x)V(x)$. For the ergodic diffusion model, we have the information criterion in the sense of AEU by using Theorem 4.

THEOREM 6. *Suppose that Assumptions* 1 *and* 2 *for some* $K' > 1$ *and* $\alpha > 1$ *hold true. Moreover, suppose that Assumption* 3 *holds true. Then*

$$\begin{aligned}\mathrm{IC}_1(\mathbf{X}_T(\omega)) \;=\; & r_T\left[\tilde{a}(X_0,\hat{\theta}_T) + \int_0^T \tilde{b}(X_t,\theta)\,\mathrm{d}X_t\bigg|_{\theta=\hat{\theta}_T} + \right.\\ & \left. + \int_0^T \tilde{c}(X_t,\hat{\theta}_T)\,\mathrm{d}t\right] - r_T b_1(\hat{\theta}_T),\end{aligned}$$

where $b_1(\theta_0) = tr\left(\nu(\xi^{(1)}(\cdot)\zeta^{(0)}(\cdot)')\right)$ *and* $\int_0^T \tilde{b}(X_t,\theta)\,\mathrm{d}X_t\Big|_{\theta=\hat{\theta}_T}$ *should be read as substituting* $\theta = \hat{\theta}_T$ *for* θ *in the random field* $\int_0^T \tilde{b}(X_t,\theta)\,\mathrm{d}X_t$.

Remark 7. In particular, in the correctly specified and MLE case, we have AIC for the ergodic diffusion model:

$$\begin{aligned} b_1(\theta_0) &= \mathrm{tr}\left[-\nu\left((\partial_\theta)^2 c(\cdot,\theta_0)\right)^{-1}\nu\left(\|\partial_\theta b(\cdot,\theta_0)\|^2\right)\right] \\ &= p \quad \text{(dimension of parameter space).}\end{aligned}$$

From Theorem 5, the information criterion in the sense of AMU for the ergodic diffusion model is given as follows.

THEOREM 7. *Suppose that Assumptions* 1 *and* 2 *for some* $K' > 1$ *and* $\alpha > 1$ *hold true. Moreover, suppose that Assumptions* 3 *and* 4 *are satisfied. Then*

$$\begin{aligned}\mathrm{IC}_2(\mathbf{X}_T(\omega)) &= r_T \left[\tilde{a}(X_0, \hat{\theta}_T) + \int_0^T \tilde{b}(X_t, \theta)\, \mathrm{d}X_t \Big|_{\theta=\hat{\theta}_T} + \right. \\ &\qquad \left. + \int_0^T \tilde{c}(X_t, \hat{\theta}_T)\, \mathrm{d}t \right] - r_T b_2(\hat{\theta}_T),\end{aligned}$$

where

$$\begin{aligned}b_2(\theta_0) &= \left[\frac{\nu\left(\xi^{(0)}(\cdot)\left\{\partial G_{f_3}(\cdot)V(\cdot)\right\}'\right)}{2\nu\left(\|\xi^{(0)}(\cdot)\|^2\right)} + tr\left\{\nu\left(\xi^{(1)}(\cdot)\zeta^{(0)}(\cdot)'\right)\right\} - \right. \\ &\qquad \left. - \frac{\left\{\nu\left(\xi^{(1)}(\cdot)\xi^{(0)}(\cdot)'\right)\right\}'\nu\left(\zeta^{(0)}(\cdot)\xi^{(0)}(\cdot)'\right)}{\nu\left(\|\xi^{(0)}(\cdot)\|^2\right)} \right].\end{aligned}$$

Remark 8. For [A1] in Assumption 1, we can refer Veretennikov (1997) and Kusuoka and Yoshida (2000). For the assurance of [A3], see Kusuoka and Yoshida (2000) using the relation between the Hörmander condition and the regularity of distributions, and Yoshida (2000) applying the support theorem.

4. Proofs

Proof of Theorem 1. From Theorem 5 in Kusuoka and Yoshida (2000), we have the second order asymptotic expansion of $\tilde{\Delta}_T^* = \bar{Z}_T^{(0)} + r_T(\bar{Z}_T^{(1)\prime}\bar{\zeta}_T^{(0)} - b(\theta_0))$ as follows.

Let $M, \gamma, K > 0$. Suppose that Assumption 1 holds true. Then for any $K \in \mathbf{N}$, there exist smooth functions $q_{j,1,T} : \mathbf{R} \to \mathbf{R}$ such that $q_{0,1,T}(z^{(0)}) = \phi(z^{(0)}; \Sigma_T^{(00)})$ and that for some $b > 0$ and $B > 0$,

$$|q_{j,1,T}(z^{(0)})| \leqslant B\mathrm{e}^{-b|z^{(0)}|^2},$$

and there exist constants $\delta > 0$ and $c > 0$ such that

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

for any $f \in \mathcal{E}(M, \gamma)$, where $\epsilon_T^{(1)}$ is a sequence of constants independent of f with $\epsilon_T^{(1)} = o(T^{-\frac{1}{2}(1+\delta)\wedge K})$.

With the aid of Theorem 2.1 in Sakamoto and Yoshida (1999),

$$\begin{aligned}
q_{0,1,T}(z^{(0)}) &= \int_{\mathbf{R}^{2p}} \phi(z;\, \Sigma_T)\, \mathrm{d}z^{(1)}\, \mathrm{d}z^{(2)}, \\
q_{1,1,T}(z^{(0)}) &= \int_{\mathbf{R}^{2p}} \Xi_{T,1}(z)\phi(z;\, \Sigma_T)\, \mathrm{d}z^{(1)}\, \mathrm{d}z^{(2)} - \\
&\quad -\partial_{z^{(0)}} \left[\int_{\mathbf{R}^{2p}} \left\{ z^{(1)\prime} z^{(2)} - b(\theta_0) \right\} \Xi_{T,0}(z)\phi(z;\, \Sigma_T)\, \mathrm{d}z^{(1)}\, \mathrm{d}z^{(2)} \right].
\end{aligned}$$

Here $z = (z^{(0)}, z^{(1)}, z^{(2)})$ and functions $\Xi_{T,j}$, $j = 0, 1$, are defined, with the Einstein summation convention, by $\Xi_{T,0}(z) = 1$ and $\Xi_{T,1}(z) = r_T^{-1} \frac{1}{6} \lambda^{\alpha\beta\gamma} h_{\alpha\beta\gamma}(z;\, \Sigma_T)$, respectively.

We then obtain

$$\begin{aligned}
q_{T,1}(z^{(0)}) &:= \sum_{j=0}^{1} T^{-j/2} q_{j,1,T}(z^{(0)}) \\
&= \int_{\mathbf{R}^{2p}} \phi(z;\, \Sigma_T)\, \mathrm{d}z^{(1)}\, \mathrm{d}z^{(2)} + \\
&\quad + r_T \int_{\mathbf{R}^{2p}} \Xi_{T,1}(z)\phi(z;\, \Sigma_T)\, \mathrm{d}z^{(1)}\, \mathrm{d}z^{(2)} - \\
&\quad - r_T\, \partial_{z^{(0)}} \left[\int_{\mathbf{R}^{2p}} \left\{ z^{(1)\prime} z^{(2)} - b(\theta_0) \right\} \Xi_{T,0}(z)\phi(z;\, \Sigma_T)\, \mathrm{d}z^{(1)}\, \mathrm{d}z^{(2)} \right] \\
&= \int_{\mathbf{R}^{2p}} p_{T,1}(z)\, \mathrm{d}z^{(1)}\, \mathrm{d}z^{(2)} - \\
&\quad - r_T\, \partial_{z^{(0)}} \left[\int_{\mathbf{R}^{2p}} \{ z^{(1)\prime} z^{(2)} - b(\theta_0) \} \phi(z;\, \Sigma_T)\, \mathrm{d}z^{(1)}\, \mathrm{d}z^{(2)} \right],
\end{aligned}$$

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.

Proof of Theorem 2.

$$\begin{aligned}
|P[f(\bar{\Delta}_T^*)] - \Psi_{T,1}[f]| &\leqslant |P[f(\bar{\Delta}_T^*)] - P[f(\tilde{\Delta}_T^*)]| + |P[f(\tilde{\Delta}_T^*)] - \\
&\quad - \Psi_{T,1}[f]| \\
&\leqslant |P[f(\bar{\Delta}_T^*)] - P[f(\tilde{\Delta}_T^*)]| + c\omega(f, r_T^K) + \varepsilon_T \\
&= \Xi_1 + \Xi_2 + c\omega(f, r_T^K) + \varepsilon_T,
\end{aligned}$$

where

$$\begin{aligned}
\Xi_1 &= |P[(f(\bar{\Delta}_T^*) - f(\tilde{\Delta}_T^*)) 1_{\{|R_T^*| \leqslant r_T^{K'}\}}]|, \\
\Xi_2 &= |P[(f(\bar{\Delta}_T^*) - f(\tilde{\Delta}_T^*)) 1_{\{|R_T^*| > r_T^{K'}\}}]|.
\end{aligned}$$

From the assumption concerned with $\tilde{\Delta}_T^*$, it follows that

$$\begin{aligned}\Xi_1 &\leqslant P[\sup\{|f(\tilde{\Delta}_T^* + y) - f(\tilde{\Delta}_T^*)| : |y| \leqslant r_T^{K'}\}] \\ &\leqslant |\Psi_{T,1}[h]| + c_0\omega(h, r_T^K) + \varepsilon_T,\end{aligned}$$

where $h(x) = \sup\{|f(x+y) - f(x)|; |y| \leqslant r_T^{K'}\}$ and c_0 is a positive constant. Note that for any $\delta > 0$, there exists a constant $C > 0$ such that for any $z \in \mathbf{R}^{2p+1}$,

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

Therefore we have

$$\begin{aligned}|\Psi_{T,1}[h]| &\leqslant \int_{\mathbf{R}} h(x)|q_{T,1}(x)|\,\mathrm{d}x \\ &\leqslant c_1 \int_{\mathbf{R}} h(x)\phi(x; \hat{\Sigma}^{(00)})\,\mathrm{d}x \\ &= c_1\omega(f, r_T^{K'})\end{aligned}$$

for some constant $c_1 > 0$. Since

$$\begin{aligned}&\sup\{|h(x+y) - h(x)|; |y| \leqslant r\} \\ \leqslant\ &\sup\{|h(x+y)| + |h(x)|; |y| \leqslant r\} \\ \leqslant\ &\sup\{|h(x+y)|; |y| \leqslant r\} + |h(x)| \\ \leqslant\ &\sup\{|f(x+y+z) - f(x+y)|; |y| \leqslant r, |z| \leqslant r_T^{K'}\} + |h(x)| \\ \leqslant\ &\sup\{|f(x+y+z) - f(x)| + |f(x+y) - f(x)|; |y| \leqslant r, |z| \leqslant r_T^{K'}\} + \\ &+ |h(x)| \\ \leqslant\ &3\sup\{|f(x+y_1) - f(x)|; |y_1| \leqslant r + r_T^{K'}\},\end{aligned}$$

we obtain

$$\begin{aligned}\omega(h, r_T^K) &= \int_{\mathbf{R}} \sup\{|h(x+y) - h(x)|; |y| \leqslant r_T^K\}\phi(x, \hat{\Sigma}^{(00)})\,\mathrm{d}x \\ &\leqslant 3\int_{\mathbf{R}} \sup\{|f(x+y_1) - f(x)|; |y_1| \leqslant r_T^K + r_T^{K'}\}\phi(x, \hat{\Sigma}^{(00)})\,\mathrm{d}x \\ &= 3\omega(f, r_T^K + r_T^{K'}).\end{aligned}$$

Thus we see that

$$\Xi_1 \leqslant c_1\omega(f, r_T^{K'}) + 3c_0\omega(f, r_T^K + r_T^{K'}) + \varepsilon_T. \tag{17}$$

From (17) and Assumption 2,

$$\Xi_1 + \Xi_2 \leqslant c_1\omega(f, r_T^{K'}) + 3c_0\omega(f, r_T^K + r_T^{K'}) + o(r_T^{(1+\delta)\wedge K}) + o(r_T^\alpha).$$

We then have that for some $\tilde{c} > 0$,

$$\begin{aligned}
|P[f(\bar{\Delta}_T^*)] - \Psi_{T,1}[f]| &\leqslant c_1\omega(f, r_T^{K'}) + 3c_0\omega(f, r_T^K + r_T^{K'}) + \\
&\quad + o(r_T^{(1+\delta)\wedge K}) + o(r_T^{\alpha}) + c\omega(f, r_T^K) + \\
&\quad + o(r_T^{(1+\delta)\wedge K}) \\
&\leqslant \tilde{c}\omega(f, r_T^K + r_T^{K'}) + o(r_T^{(1+\delta)\wedge K\wedge\alpha}) \\
&\leqslant \tilde{c}\omega(f, 2r_T^{K'}) + o(r_T^{(1+\delta)\wedge K\wedge\alpha}),
\end{aligned}$$

which completes the proof.

Proof of Theorem 3. In order to obtain an explicit expression for a second order asymptotic expansion of the distribution of $\bar{\Delta}_T^*$, we need to compute the following integral.

$$\begin{aligned}
q_{T,1}(z^{(0)}) &= \int_{\mathbf{R}^{2p}} \phi(z; \Sigma_T)\, \mathrm{d}z^{(1)}\, \mathrm{d}z^{(2)} + \\
&+ \frac{1}{6}\int_{\mathbf{R}^{2p}} \lambda_T^{\alpha\beta\gamma} h_{\alpha\beta\gamma}(z; \Sigma_T)\phi(z; \Sigma_T)\, \mathrm{d}z^{(1)}\, \mathrm{d}z^{(2)} - \\
&- r_T\partial_{z^{(0)}}\left[\int_{\mathbf{R}^{2p}} \{z^{(1)\prime}z^{(2)} - b(\theta_0)\}\phi(z; \Sigma_T)\mathrm{d}z^{(1)}\, \mathrm{d}z^{(2)}\right] \\
&= \phi(z^{(0)}; \Sigma_T^{(00)}) + \frac{1}{6}(\mathrm{I}) - r_T\partial_{z^{(0)}}\left[(\mathrm{II}) - b(\theta_0)\phi(z^{(0)}; \Sigma_T^{(00)})\right],
\end{aligned}$$

where

$$\begin{aligned}
(\mathrm{I}) &= -\int_{\mathbf{R}^{2p}} \lambda_T^{\alpha\beta\gamma}\partial_\alpha\partial_\beta\partial_\gamma\phi(z; \Sigma_T)\, \mathrm{d}z^{(1)}\, \mathrm{d}z^{(2)} \\
&= -\lambda_T^{000}(\partial_{z^{(0)}})^3\int_{\mathbf{R}^{2p}} \phi(z; \Sigma_T)\, \mathrm{d}z^{(1)}\, \mathrm{d}z^{(2)} \\
&= -\lambda_T^{000}(\partial_{z^{(0)}})^3\phi(z^{(0)}; \Sigma_T^{(00)}) \\
&= \lambda_T^{000}h_3(z^{(0)}; \Sigma_T^{(00)})\phi(z^{(0)}; \Sigma_T^{(00)}), \\
(\mathrm{II}) &= \int_{\mathbf{R}^{2p}} z^{(1)\prime}z^{(2)}\frac{\phi(z; \Sigma_T)}{\int_{\mathbf{R}^{2p}}\phi(z; \Sigma_T)\, \mathrm{d}z^{(1)}\, \mathrm{d}z^{(2)}}\, \mathrm{d}z^{(1)}\, \mathrm{d}z^{(2)} \times \\
&\quad \times \int_{\mathbf{R}^{2p}} \phi(z; \Sigma_T)\, \mathrm{d}z^{(1)}\, \mathrm{d}z^{(2)} \\
&= \int_{\mathbf{R}^{2p}} z^{(1)\prime}z^{(2)}\phi(z^{(1)}, z^{(2)}; \mu, \Sigma)\mathrm{d}z^{(1)}\, \mathrm{d}z^{(2)}\phi(z^{(0)}; \Sigma_T^{(00)}).
\end{aligned}$$

Here $\phi(z^{(1)}, z^{(2)}; \mu, \Sigma)$ is normal with mean μ and covariance matrix Σ, where

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

Hence

$$\begin{aligned}
C_T(z^{(0)}) &:= \int_{\mathbf{R}^{2p}} z^{(1)\prime} z^{(2)} \phi(z^{(1)}, z^{(2)}; \mu, \Sigma) \mathrm{d}z^{(1)}\, \mathrm{d}z^{(2)} \\
&= (\Sigma_T^{(10)}(\Sigma_T^{(00)})^{-1} z^{(0)})'(\Sigma_T^{(20)}(\Sigma_T^{(00)})^{-1} z^{(0)}) + \mathrm{tr}\{\Sigma_T^{(12)} - \\
&\quad - (\Sigma_T^{(00)})^{-1} \Sigma_T^{(10)} (\Sigma_T^{(20)})'\} \\
&= \frac{(\Sigma_T^{(10)})' \Sigma_T^{(20)}}{(\Sigma_T^{(00)})^2} [(z^{(0)})^2 - \Sigma_T^{(00)}] + \mathrm{tr}\Sigma_T^{(12)}.
\end{aligned}$$

From the representation of (I) and $C_T(z^{(0)})$, we obtain

$$\begin{aligned}
q_{T,1}(z^{(0)}) &= \phi(z^{(0)}; \Sigma_T^{(00)}) + \frac{1}{6}\lambda_T^{000} h_3(z^{(0)}; \Sigma_T^{(00)}) \phi(z^{(0)}; \Sigma_T^{(00)}) - \\
&\quad - r_T \partial_{z^{(0)}} \left[\left\{ C_T(z^{(0)}) - b(\theta_0) \right\} \phi(z^{(0)}; \Sigma_T^{(00)}) \right].
\end{aligned}$$

This completes the proof.

Proof of Theorem 4. In Theorem 3, by setting $f(x) = x$ and $b(\cdot) = b_1(\cdot)$, we see that

$$\begin{aligned}
P[\bar{\Delta}_T^*] &= P\left[r_T \Delta_T - r_T b_1(\hat{\theta}_T) \right] \\
&= r_T \left[\mathrm{tr}\Sigma_T^{(12)} - b_1(\theta_0) \right] + o\left(r_T\right) \\
&= o\left(r_T\right),
\end{aligned}$$

which completes the proof.

Proof of Theorem 5. In Theorem 3, by putting $f(x) = 1_{(a,\infty)}(x)$ and $b(\cdot) = b_2(\cdot)$, we obtain

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

In particular, by setting $a = 0$, we have

$$\begin{aligned}
P[\bar{\Delta}_T^* > 0] &= P[r_T \Delta_T - r_T b_2(\hat{\theta}_T) > 0] \\
&= \int_0^{\infty} \phi(z^{(0)}; \Sigma_T^{(00)})\,\mathrm{d}z^{(0)} - \frac{1}{6}\lambda_T^{000} \frac{1}{\Sigma_T^{(00)}} \phi(0; \Sigma_T^{(00)}) + \\
&\quad + r_T \left[\mathrm{tr}\Sigma_T^{(12)} - \frac{(\Sigma_T^{(10)})' \Sigma_T^{(20)}}{\Sigma_T^{(00)}} - b_2(\theta_0) \right] \phi(0; \Sigma_T^{(00)}) + o\,(r_T) \\
&= \frac{1}{2} + o\,(r_T)\,.
\end{aligned}$$

Similarly, by setting $f(x) = 1_{(-\infty,0)}(x)$ and $b(\cdot) = b_2(\cdot)$ in Theorem 3, we obtain

$$\begin{aligned}
P[\bar{\Delta}_T^* < 0] &= P[r_T \Delta_T - r_T b_2(\hat{\theta}_T) < 0] \\
&= \int_{-\infty}^{0} \phi(z^{(0)}; \Sigma_T^{(00)})\,\mathrm{d}z^{(0)} + \frac{1}{6}\lambda_T^{000} \frac{1}{\Sigma_T^{(00)}} \phi(0; \Sigma_T^{(00)}) - \\
&\quad - r_T \left[\mathrm{tr}\Sigma_T^{(12)} - \frac{(\Sigma_T^{(10)})' \Sigma_T^{(20)}}{\Sigma_T^{(00)}} - b_2(\theta_0) \right] \phi(0; \Sigma_T^{(00)}) + o\,(r_T) \\
&= \frac{1}{2} + o\,(r_T)\,.
\end{aligned}$$

This completes the proof.

Proof of Theorems 6 and 7. From (13), (14), (15) and (16), we obtain

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

From Itô's formula it follows that

$$\begin{aligned}
\mathrm{Cov}(\bar{Z}_T^{(0)}, \bar{Z}_T^{(0)}) &= P\left[{r_T}^2 \int_0^T \xi^{(0)}(X_t) \xi^{(0)}(X_t)'\,\mathrm{d}t\right] + o(1), \\
\mathrm{Cov}(\bar{Z}_T^{(1)}, \bar{\zeta}_T^{(0)}) &= P\left[{r_T}^2 \int_0^T \xi^{(1)}(X_t) \zeta^{(0)}(X_t)'\,\mathrm{d}t\right] + o(1), \\
\mathrm{Cov}(\bar{Z}_T^{(1)}, \bar{Z}_T^{(0)}) &= P\left[{r_T}^2 \int_0^T \xi^{(1)}(X_t) \xi^{(0)}(X_t)'\,\mathrm{d}t\right] + o(1), \\
\mathrm{Cov}(\bar{\zeta}_T^{(0)}, \bar{Z}_T^{(0)}) &= P\left[{r_T}^2 \int_0^T \zeta^{(0)}(X_t) \xi^{(0)}(X_t)'\,\mathrm{d}t\right] + o(1).
\end{aligned}$$

From stationarity, we obtain

$$\begin{aligned}
\Sigma_T^{(00)} &= \nu(\xi^{(0)}(\cdot)\xi^{(0)}(\cdot)') + o(1), & (18)\\
\Sigma_T^{(12)} &= \nu(\xi^{(1)}(\cdot)\zeta^{(0)}(\cdot)') + o(1), & (19)\\
\Sigma_T^{(10)} &= \nu(\xi^{(1)}(\cdot)\xi^{(0)}(\cdot)') + o(1), & (20)\\
\Sigma_T^{(20)} &= \nu(\zeta^{(0)}(\cdot)\xi^{(0)}(\cdot)') + o(1). & (21)
\end{aligned}$$

Next, we compute λ_T^{000}.

$$\bar{Z}_T^{(0)} = r_T \int_0^T \xi^{(0)}(X_t)\,\mathrm{d}w_t + r_T\left\{G_{f_0}(X_T) - G_{f_0}(X_0)\right\} + r_T\alpha_1(X_0,\theta_0).$$

By using the strong mixing property, we obtain

$$\lambda_T^{000} = P[(\bar{Z}_T^{(0)})^3] = {r_T}^3 P\left[\left(\int_0^T \xi^{(0)}(X_t)\,\mathrm{d}w_t\right)^3\right] + o\,(r_T)\,.$$

From Itô's formula,

$$\begin{aligned}
\lambda_T^{000} &= 3{r_T}^3 \cdot P\left[\int_0^T \left(\int_0^t \xi^{(0)}(X_u)\,\mathrm{d}w_u\right) \|\xi^{(0)}(X_t)\|^2\,\mathrm{d}t\right] + o\,(r_T)\\
&= 3{r_T}^3 \cdot P\left[\int_0^T \left(\int_0^t \xi^{(0)}(X_u)\,\mathrm{d}w_u\right) [\|\xi^{(0)}(X_t)\|^2 - \right.\\
&\quad \left. - \nu(\|\xi^{(0)}(\cdot)\|^2)]\,\mathrm{d}t\right] + o\,(r_T)\\
&= 3{r_T}^3 \cdot P\left[\left(\int_0^T \xi^{(0)}(X_t)\,\mathrm{d}w_t\right) \cdot \left(\int_0^T [\|\xi^{(0)}(X_t)\|^2 - \right.\right.\\
&\quad \left.\left. - \nu(\|\xi^{(0)}(\cdot)\|^2)]\,\mathrm{d}t\right)\right] + o\,(r_T)\,.
\end{aligned}$$

From 16 it follows that

$$\begin{aligned}
\lambda_T^{000} &= 3r_T P\left[\left(r_T \int_0^T \xi^{(0)}(X_t)\,\mathrm{d}w_t\right) \times \right.\\
&\quad \left. \times \left(r_T\left\{G_{f_3}(X_T) - G_{f_3}(X_0)\right\} - r_T \int_0^T \partial G_{f_3}(X_t)V(X_t)\,\mathrm{d}w_t\right)\right] +\\
&\quad + o\,(r_T)\\
&= -3r_T P\left[\left(r_T \int_0^T \xi^{(0)}(X_t)\,\mathrm{d}w_t\right)\left(r_T \int_0^T \partial G_{f_3}(X_t)V(X_t)\,\mathrm{d}w_t\right)\right] +\\
&\quad + o\,(r_T)\\
&= -3r_T P\left[{r_T}^2 \int_0^T \xi^{(0)}(X_t)\left\{\partial G_{f_3}(X_t)V(X_t)\right\}'\,\mathrm{d}t\right] + o\,(r_T)\,.
\end{aligned}$$

Finally, from stationarity,

$$\lambda_T^{000} = -3r_T\nu(\xi^{(0)}(\cdot)\{\partial G_{f_3}(\cdot)V(\cdot)\}') + o\,(r_T)\,. \qquad (22)$$

We then obtain IC_1 from (10), (12) and (19). Moreover, IC_2 is given by (10), (12), (18), (19), (20), (21) and (22). This completes the proof.

Acknowledgements

The authors are grateful to Prof. S. Konishi and Prof. G. Kitagawa for their helpful advice and valuable suggestions. We also wish to thank Prof. Yuji Sakamoto and Dr. H. Shimodaira for their valuable comments and encouragement. Many thanks go to the editor, the associate editor and the referee for their careful reading of this manuscript and valuable comments. This work was in part supported by the Research Fund for Scientists of the Ministry of Science, Education and Culture, and by Cooperative 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, in *Lecture Notes in Statistics*, Vol. 7, Berlin, Heidelberg, New York, Springer.

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

Akaike, H.: 1974, A new look at the statistical model identification, *IEEE Trans. Auto. Control*, **AC-19**, 716–723.

Barron, A.R.: 1986, Entropy and the central limit theorem, *Ann. Probab.* **14**, 336–342.

Barron, A.R.: 1989, Uniformly powerful goodness of fit tests, *Ann. Statist.* **17**, 107–124.

Barron, A.R., Birgé, L. and Massart, P.: 1999, Risk bounds for model selection via penalization, *Probab. Theory Relat. Fields* **113**, 301–413.

Bhattacharya, R.N. and Ghosh, J.K.: 1978, On the validity of the formal Edgeworth expansion, *Ann. Statist.* **6**, 434–451.

Bichteler, K., Gravereaux, J.-B. and Jacod, J.: 1987, *Malliavin Calculus for Processes with Jumps*, New York London Paris Montreux Tokyo, Gordon and Breach Science Publishers.

Burman, P. and Nolan, D.: 1995, A general Akaike-type criterion for model selection in robust regression, *Biometrika* **82**, 877–886.

Burnham, K.P. and Anderson, D.R.: 1998, *Model Selection and Inference*, Berlin, Heidelberg, New York, Springer.

Götze, F. and Hipp, C.: 1983, Asymptotic expansions for sums of weakly dependent random vectors, *Z. Wahrsch. Verw. Gebiete* **64**, 211–239.

Götze, F. and Hipp, C.: 1994, Asymptotic distribution of statistics in time series, *Ann. Statist.* **22**, 2062–2088.

Hall, P.: 1990, Akaike's information criterion and Kullback–Leibler loss for histogram density estimation, *Probab. Theory Related Fields* **85**, 449–467.

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

Hurvich, C.M. and Tsai, C.-L.: 1995, Relative rates of convergence for efficient model selection criteria in linear regression, *Biometrika* **82**, 418–425.

Knight, K.: 1989, Consistency of Akaike's information criterion for infinite variance autoregressive processes, *Ann. Statist.* **17**, 824–840.

Konishi, S. and Kitagawa, G.: 1996, Generalised information criteria in model selection, *Biometrika* **83**, 875–890.

Kullback, S. and Leibler, R.A.: 1951, On information and sufficiency, *Ann. Math. Statist.* **22**, 79–86.

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

Laud, P.W. and Ibrahim, J.G.: 1995, Predictive model selection, *J. R. Statist. Soc.* **B57**, 247–262.

Portnoy, S.: 1997, Local asymptotics for quantile smoothing splines, *Ann. Statist.* **25**, 414–434.

Sakamoto, Y. and Yoshida, N.: 1998, Asymptotic expansions of M-estimator over Wiener space, *Statist. Infer. Stochast. Process.* **1**, 85–103.

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

Sakamoto, Y. and Yoshida, N.: 2000, Asymptotic expansion under degeneracy, (in preparation).

Shibata, R.: 1980, Asymptotically efficient selection of the order of the model for estimating parameters of a linear process, *Ann. Statist.* **8**, 147–164.

Shibata, R.: 1981, An optimal autoregressive spectral estimate, *Ann. Statist.* **9**, 300–306.

Takeuchi, K.: 1976, Distribution of information statistics and criteria for adequacy of models, *Math. Sci.* **153**, 12–18 (in Japanese).

Uchida, M. and Yoshida, N.: 1999, Information criteria in model selection for stochastic processes (II). Research Memorandum **730**, The Institute of the Statistical Mathematics, Tokyo.

Veretennikov, A.Y.: 1997, On polynomial mixing bounds for stochastic differential equations, *Stochast. Process. Appl.* **70**, 115–127.

Yang, Y. and Barron, A.R.: 1998, An asymptotic property of model selection criteria, *IEEE Transact. Info. Theory* **44**, 95–116.

Yoshida, N.: 2000, Partial mixing and conditional Edgeworth expansion for diffusions with jumps, (preprint).