Nondegeneracy of Random Field and Estimation of Diffusion

Masayuki Uchida and Nakahiro Yoshida

1Graduate School of Engineering Science, Osaka University
2Japan Science and Technology Agency, CREST
3Graduate School of Mathematical Sciences, University of Tokyo
4The Institute of Statistical Mathematics

December 9, 2012

Abstract. We construct a quasi likelihood analysis for diffusions under the high-frequency sampling over a finite time interval. For this, we prove a polynomial type large deviation inequality for the quasi likelihood random field. Then it becomes crucial to prove nondegeneracy of a key index $\chi_0$. By nature of the sampling setting, $\chi_0$ is random. This makes it difficult to apply a naive sufficient condition, and requires a new machinery. In order to establish a quasi likelihood analysis, we need quantitative estimate of the nondegeneracy of $\chi_0$. The existence of a nondegenerate local section of a certain tensor bundle associated with the statistical random field solves this problem.

Key words and phrases: asymptotic mixed normality, Bayes type estimator, convergence of moments, discrete time observation, maximum likelihood type estimator, polynomial type large deviation inequality.

AMS 2000 subject classifications: Primary 62F15, 62M05; Secondary 60J60.

1 Introduction

In this paper, we consider estimation for a stochastic regression model specified by the stochastic integral equation

$$Y_t = Y_0 + \int_0^t b_s ds + \int_0^t \sigma(X_s, \theta) dw_s, \quad t \in [0, T],$$

where $w$ is an $r$-dimensional standard Wiener process on a stochastic basis $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t \in [0, T]}, P)$, $b$ and $X$ are progressively measurable processes with values in $\mathbb{R}^m$ and $\mathbb{R}^d$, respectively, $\sigma$ is an $\mathbb{R}^m \otimes \mathbb{R}^r$-valued function defined on $\mathbb{R}^d \times \Theta$, and $\Theta$ is a bounded domain in $\mathbb{R}^p$. As a special case, if an argument of $X_t$ is $t$, then the volatility in the model (1) is time dependent. Furthermore, if we set $b_t = b(Y_t, t)$ and $X_t = (Y_t, t)$, then $Y$ can be a time-inhomogeneous diffusion process. Of course, the stochastic volatility model like (1) is quite commonly used in finance and econometrics. The data set consists of discrete observations $Z_n = (X_{t_k}, Y_{t_k})_{0 \leq k \leq n}$ with $t_k = kh$ for $h = h_n = T/n$. The process $b$ is completely unobservable and unknown. The asymptotics will be considered for $n \to \infty$, that is, $Z_n$ forms high frequency data.

Asymptotic theory of parametric estimation for the unknown parameter $\theta$ in the volatility of the stochastic differential equation based on high frequency data has been developed. Among many studies
in a long history, we refer the reader to Plakasa Rao (1983,1988), Yoshida (1992,2005), Kessler (1997) under ergodicity, Shimizu and Yoshida (2006), Shimizu (2006), Ogihara and Yoshida (2009) for jump diffusion processes, Sorensen and Uchida (2003), Uchida (2003, 2004, 2008) for perturbed diffusions, Dohnal (1987), Genon-Catalot and Jacod (1993, 1994), Gobet (2001) for the fixed interval case. The limit distribution of the score function becomes a mixture of normal distributions over a finite time interval (LAMN), and a normal distribution over the infinite time interval (LAN) by the averaging effect. In this article, we will consider the LAMN (i.e., locally asymptotically mixed normal) quasi likelihood experiment associated with the sampling scheme over a finite time interval.

A highlight of asymptotic decision theory is the likelihood analysis, the basic frame and functions of which were established by Le Cam, Hájek, Ibragimov and Has’minskii and others. The theory of Ibragimov and Has’minskii provides convergence of likelihood ratio random field on a function space with certain estimates for the tail probability and consequently convergence of moments of the estimator appearing in the likelihood analysis. It was Yury Kutoyants who found this methodology was effective for semimartingales, proving the wide applicability to various stochastic models. See Kutoyants (1984, 1994, 1998, 2004) for more information.

Limiting distribution of the estimator is indispensable, however, it is far from sufficient to develop the elementary statistical theory. It is clear if we consider a problem of model selection, for example. The basic correction term by Akaike was introduced to make the Kullback-Leibler divergence between the predictive distribution and the true distribution asymptotically unbiased. Obviously, it is necessary to validate the existence of moments of the standardized estimator because the bias is described with it. The asymptotic distribution cannot provide sufficient information there. It is also the case in the prediction theory. Furthermore, the same kind of questions inevitably arise in the theory of higher-order statistical inference. Large deviation type estimates enable valid treatments of the higher-order terms in the stochastic expansion of a statistic, and such estimates can be obtained by precise probabilistic estimate of the decay of the accompanying statistical random field.

The quasi likelihood analysis has been developing for stochastic processes. Here the quasi likelihood analysis means a system that gives asymptotic behavior of the quasi likelihood random field, its (polynomial type) large deviation estimate, limit theorems for the quasi maximum likelihood estimator and the quasi Bayesian estimator, and convergence of moments of these estimators. Yoshida (2005, 2011) gave a polynomial type large deviation inequality in the locally asymptotically quadratic (LAQ) setting to carry out the Ibragimov-Has’minskii-Kutoyants scheme for stochastic processes. As a corollary, the quasi-likelihood analysis for ergodic diffusion processes under sampling was presented. The simultaneous and adaptive Bayesian estimators were defined there. See Le Cam (1986), Le Cam and Yang (1990) for the fundamental notions of statistical experiments and approximation.

The polynomial type large deviation inequality works in various settings. Uchida (2010) considered a model selection problem for discretely observed ergodic multi-dimensional diffusion processes and proposed a contrast-based information criterion. The difficulties are in existence of moments, and besides, in handling the exact likelihood function, that has no explicit expression. The polynomial type large deviation inequality and the Malliavin calculus were effectively used. The asymptotic results can be fairly complicated if jumps with heavy tail are involved; even convergence rate of the estimator can differ from the standard one. Masuda (2010) obtained a polynomial type large deviation estimate for the random field associated with a general self-weighted least absolute deviation (SLAD) in the parameter estimation of sampled Ornstein-Uhlenbeck process driven by a heavy-tailed symmetric Lévy process with positive activity index, and clarified asymptotic behavior of the estimator including convergence of moments. A quasi likelihood analysis was constructed by Ogihara and Yoshida (2009) for a nonlinear sampled diffusion process with jumps with the aid of the polynomial type large deviation
In this section, we will present the main results in statistical context. Even apart from statistical results presented here, this paper. Since such nondegeneracy argument is universal, the authors hope this part has its own section of a certain tensor bundle related to the statistical random field. This is the second aim of our model can easily break it. In order to establish a quasi likelihood analysis, we need quantitative estimate of the nondegeneracy of $\chi_0$. This problem is solved by the existence of a nondegenerate local section of a certain tensor bundle related to the statistical random field. This is the second aim of this paper. Since such nondegeneracy argument is universal, the authors hope this part has its own interest even apart from statistical results presented here.

## 2 Quasi likelihood analysis for diffusion and the limit theorems

In this section, we will present the main results in statistical context.

Suppose that $\Theta$ is a bounded domain in $\mathbb{R}^p$ with a locally Lipschitz boundary, which means that $\Theta$ has the strong local Lipschitz condition, see Adams (1975) and Adams and Fournier (2003). $\theta^*$ denotes the true value of $\theta$.

Let $C^{k,l}_*(\mathbb{R}^d \times \Theta; \mathbb{R}^m)$ denote the space of all functions $f$ satisfying the following conditions: (i) $f(x, \theta)$ is an $\mathbb{R}^m$-valued function on $\mathbb{R}^d \times \Theta$, (ii) $f(x, \theta)$ is continuously differentiable with respect to $x$ up to order $k$ for all $\theta$, and their derivatives up to order $k$ are of polynomial growth in $x$ uniformly in $\theta$. Moreover, for $|\nu| = 0, 1, \ldots, k$, $\partial^\nu_x f(x, \theta)$ is of polynomial growth in $x$ uniformly in $\theta$. Here $n = (n_1, \ldots, n_d)$ and $\nu = (\nu_1, \ldots, \nu_p)$ are multi-indices, $p = \text{dim}(\Theta)$, $|n| = n_1 + \ldots + n_d$, $|\nu| = \nu_1 + \ldots + \nu_p$, $\partial^\nu\partial^n f(x, \theta)$ is continuously differentiable with respect to $\theta$ up to order $l$ for all $x$. We denote by $\rightarrow^\nu\rightarrow^n$ and $\rightarrow^d(\mathcal{F})$ the convergence in probability and the $\mathcal{F}$-stable convergence in distribution, respectively. For matrices $A$ and $B$ of the same size, we write $A^{\otimes 2} = A A^\ast$ and $A[B] = \text{Tr}(A B^\ast)$, where $\ast$ means the transpose. Set $S(x, \theta) = \sigma(x, \theta)^{\otimes 2}$ and $\Delta_k Y = Y_{t_k} - Y_{t_{k-1}}$. We assume that the function $\sigma$ admits a continuous extension over $\mathbb{R}^d \times \Theta$, and denote it by $\sigma$. Let

$$Q(x, \theta, \theta^*) = \text{Tr} \left( S(x, \theta)^{-1} S(x, \theta^*) - I_d \right) - \log \det \left( S(x, \theta)^{-1} S(x, \theta^*) \right).$$

We consider the following conditions.

[A1] (i) $\sup_{0 \leq t \leq T} \| b_t \|_p < \infty$ for all $p > 1$.

(ii) $\sigma \in C^{2,4}_*(\mathbb{R}^d \times \Theta; \mathbb{R}^m \otimes \mathbb{R}^r)$ and $\inf_{x, \theta} \det S(x, \theta) > 0$.

[A2] The process $X$ admits a representation

$$X_t = X_0 + \int_0^t \tilde{b}_s ds + \int_0^t a_s d\tilde{w}_s + \int_0^t \tilde{a}_s d\tilde{w}_s,$$

where

(i) $\tilde{b}$, $a$ and $\tilde{a}$ are progressively measurable processes taking values in $\mathbb{R}^d$, $\mathbb{R}^d \otimes \mathbb{R}^r$ and $\mathbb{R}^d \otimes \mathbb{R}^t$, respectively, satisfying

$$\| X_0 \|_p + \sup_{t \in [0,T]} (\| \tilde{b}_t \|_p + \| a_t \|_p + \| \tilde{a}_t \|_p) < \infty$$

3
for every $p > 1$, and $\tilde{w}$ is an $r_1$-dimensional Wiener process independent of $w$, 

(ii) there is a stopping time $\tau$ such that $\text{ess.sup}_{\omega \in \Omega} \tau < T$, $a_\tau^{\otimes 2} + \tilde{a}_\tau^{\otimes 2}$ is bounded, nondegenerate uniformly in $\omega \in \Omega$ and that $a$ and $\tilde{a}$ are right-continuous at $t = \tau$.

We say that a function $f$ admits a $C^j$-supporting function at $(x_0,\theta_0)$ if there exist a function $g$ on a neighborhood $V(x_0,\theta_0) \subset \mathbb{R}^d \times \tilde{\Theta}$ of $(x_0,\theta_0)$ and $\zeta_0 \in \mathbb{R}^d$, $|\zeta_0| = 1$, such that the partial derivatives $\partial_j^ig$ ($j = 0, ..., J$) exists for each $\theta$ near $\theta_0$ and continuous in $(x, \theta)$ and that $|f(x, \theta)| \geq |g(P_{\zeta_0}x, \theta)|$ for $(x, \theta) \in V(x_0,\theta_0)$, where $P_{\zeta_0}$ is the projection on $\mathbb{R}\xi_0$. Let $c_j(x,\theta) = (j!)^{-1}(\partial_j^ig)(P_{\zeta_0}x,\theta)|_{\xi_0}^{\otimes j}$. $[g$ and $c_j$ depend on $(x_0,\theta_0)]$

[A3] $\text{supp}\mathcal{L}\{X_t\}$ is compact, and for some open neighborhood $U$ of $\text{supp}\mathcal{L}\{X_t\}$, there exist a function $f : U \times \tilde{\Theta} \to \mathbb{R}$ and a constant $\rho \in (0,\infty)$ satisfying the following conditions.

(i) $Q(x,\theta,\theta^*)|\theta - \theta^*|^{-2} \geq |f(x,\theta)|^\rho$ for all $(x,\theta) \in U \times (\tilde{\Theta}\{\theta^*\})$.

(ii) $f$ admits a $C^j$-supporting function for each $(x_0,\theta) \in U \times \tilde{\Theta}$ with 

$$\max_{j=0,\ldots,J-1} |c_j(x_0,\theta)| > 0.$$ 

Condition [A1] is for regularity, [A2] is for the nondegeneracy of the process $X$. The stopping time $\tau$ is often taken as $\tau = 0$. The compactness of the support of $\mathcal{L}\{X_t\}$ can be relaxed if we assume stronger global nondegeneracy; a stronger condition will be inevitable in general because degeneracy can occur unless we assume the compactness of the support. Condition [A3] is for the nondegeneracy of the quasi likelihood random field to which the nondegeneracy of $X$ can be conveyed thanks to the condition.

Since the exact transition density is not available, the inference is carried out by a quasi likelihood function. Let 

$$\mathbb{H}_n(\theta) = -\frac{nm}{2} \log(2\pi h) - \frac{1}{2} \sum_{k=1}^n \{\log \det S(X_{t_{k-1}}, \theta) + h^{-1}S^{-1}(X_{t_{k-1}}, \theta)[(\Delta Y)^{\otimes 2}]\}.$$ 

Then the maximum likelihood type estimator $\hat{\theta}_n$ is any estimator that satisfies 

$$\mathbb{H}_n(\hat{\theta}_n) = \sup_{\theta \in \tilde{\Theta}} \mathbb{H}_n(\theta).$$  \hspace{1cm} (3) 

The Bayes type estimator $\tilde{\theta}_n$ for a prior density $\pi : \Theta \to \mathbb{R}_+$ with respect to the quadratic loss is defined by 

$$\tilde{\theta}_n = \left( \int_{\Theta} \exp(\mathbb{H}_n(\theta))\pi(\theta)d\theta \right)^{-1} \int_{\Theta} \theta \exp(\mathbb{H}_n(\theta))\pi(\theta)d\theta.$$  \hspace{1cm} (4) 

We assume that $\pi$ is continuous and $0 < \inf_{\theta \in \Theta} \pi(\theta) \leq \sup_{\theta \in \Theta} \pi(\theta) < \infty$.

Let $\Gamma(\theta^*) = (\Gamma_{ij}(\theta^*))_{i,j=1,\ldots,p}$ with

$$\Gamma_{ij}(\theta^*) = \frac{1}{2T} \int_0^T \text{Tr} \left((\partial_\theta S)S^{-1}(\partial_\theta S)S^{-1}(X_t, \theta^*)\right) dt$$

and let $\zeta$ be a $p$-dimensional standard normal random variable independent of $\Gamma(\theta^*)$. Here are restricted versions of the main results in this article. We will give proof of these results in Section 5.
Theorem 1. Suppose that Conditions [A1]-[A3] are fulfilled. Then \( \sqrt{n}(\hat{\theta}_n - \theta^*) \to^{d_\mathcal{F}} \Gamma(\theta^*)^{-1/2} \) and
\[
E\left[f(\sqrt{n}(\hat{\theta}_n - \theta^*))\right] \to E\left[f(\Gamma(\theta^*)^{-1/2})\right]
\]
as \( n \to \infty \) for all continuous functions \( f \) of at most polynomial growth.

Theorem 2. Suppose that Conditions [A1]-[A3] are fulfilled. Then \( \sqrt{n}(\tilde{\theta}_n - \theta^*) \to^{d_\mathcal{F}} \Gamma(\theta^*)^{-1/2} \) and
\[
E\left[f(\sqrt{n}(\tilde{\theta}_n - \theta^*))\right] \to E\left[f(\Gamma(\theta^*)^{-1/2})\right]
\]
as \( n \to \infty \) for all continuous functions \( f \) of at most polynomial growth.

We shall consider Condition [A3] (ii). The following simple example suggests degeneracy of the statistical model can easily occur.

Example 1. Let \( X_t \) satisfy the stochastic differential equation
\[
dX_t = (1 + X_t^2) \theta dt, \quad X_0 = 0. \tag{5}
\]
Assume \( \Theta \subset (0, 1/2] \). This model is completely unidentifiable when \( t = 0 \). Now \( S(x, \theta) = (1 + x^2)^2 \theta \) and
\[
Q(x, \theta, \theta^*) |\theta - \theta^*|^{-2} = \left\{ \exp \left(2(\theta^* - \theta) \log(1 + x^2)\right) - 1 - 2(\theta^* - \theta) \log(1 + x^2) \right\} |\theta - \theta^*|^{-2} \\
\geq \left( \log(1 + x^2) \right)^2.
\]
for \((x, \theta) \in U \times \Theta \) and \( U = (-1, 1) \). Set \( f(x, \theta) = \log(1 + x^2) \) and \( \varrho = 2 \). Then
\[
f(x_0, \theta) = \log(1 + x_0^2), \quad \partial_x f(x_0, \theta) = \frac{2x_0}{1 + x_0^2},
\]
\[
\partial_x^2 f(x_0, \theta) = \frac{2}{1 + x_0^2} - \frac{4x_0^2}{(1 + x_0^2)^2}.
\]
Therefore, \( \max_{j=0,1,2} |\partial_x^j f(x_0, \theta)| > 0 \) for each \((x_0, \theta) \in U \times \bar{\Theta} \). This is a rather simple case because we found \( f \) independent of \( \theta \). Thus [A3] (ii) holds. Condition [A2] (ii) is also obvious if we choose \( \tau = 0 \).

Example 2. Consider a trivial model
\[
X_t = \theta (\tau_w - \tau_0), \quad t \in [0, 1].
\]
for a stopping time \( \tau_0 \). Then we should take \( \tau = \tau_0 \). This simple example suggests the necessity of introducing stopping time \( \tau \). Naturally, \( \esssup_{\omega} \tau_0 \) should be less than one for consistent estimation.

Example 3. Consider the stochastic differential equation
\[
dX_t = S(X_t, \theta)^{1/2} dw_t, \quad t \in [0, 1] \\
X_0 = 0,
\]
where
\[
S(x, \theta) = \exp \left( \sin \theta \sin x - \theta^2 \sin^2 x \right),
\]
and \( \theta \in \Theta = (-\pi, \pi) \). This model is completely degenerate at \( t = 0 \). Let \( \theta^* = 0 \). For small neighborhood \( U \) of 0, we have

\[
|S(x, \theta)^{-1}S(x, \theta^*) - 1 - \log\{S(x, \theta)^{-1}S(x, \theta^*)\}||\theta - \theta^*||^{-2} \geq c|f(x, \theta)|^2
\]

for \( (x, \theta) \in U \times \bar{\Theta} \), if we take some \( c > 0 \) independent of \( (x, \theta) \), and

\[
f(x, \theta) = \theta^{-1}(\sin \theta \sin x - \theta^2 \sin^2 x).
\]

With the function \( f \),

\[
f(x_0, \theta) = \frac{\sin \theta}{\theta} \sin x_0 - \theta \sin^2 x_0,
\]

\[
\partial_x f(x_0, \theta) = \frac{\sin \theta}{\theta} \cos x_0 - 2 \theta \sin x_0 \cos x_0,
\]

\[
\frac{1}{2} \partial_x^2 f(x_0, \theta) = -\frac{\sin \theta}{2\theta} \sin x_0 - \theta \cos^2 x_0 + \theta \sin^2 x_0,
\]

the continuous extension being applied at \( \theta = 0 \). Set \( \Theta_1 = \Theta \setminus \Theta_2 \) with \( \Theta_2 = (B(\pi, \rho) \cup B(-\pi, \rho)) \cap \Theta \). Then it is not difficult to fix \( \rho > 0 \) and small \( \epsilon > 0 \) so that \( \partial_x f \) is nondegenerate on \( \Theta_1 \) and so is \( \partial_x^2 f \) on \( \Theta_2 \) in the same time.

A naïve, simple-looking sufficient condition is that \( \inf_x \inf_\theta |f(x, \theta)| > 0 \). However, in this example, \( \inf_x |f(x, \theta)| = 0 \), and the naïve condition does not work.

The later sections will be devoted to the proof of Theorems 1 and 2. Some generalization will be done on the way. The ingredients of the proof of these results are the polynomial type large deviation inequality for the quasi likelihood random field, as well as limit theorems for semimartingales. In Section 3, we recall the polynomial type large deviation inequality for the statistical random field. The aim of the section is to introduce a key random index \( \chi_0 \) associated with \( \mathbb{H}_n \) and to clarify its role for derivation of the large deviation estimate and as a result for establishing the quasi likelihood analysis.

Thus it is necessary to prove the nondegeneracy of a random index \( \chi_0 \). To answer this question, Section 4 is devoted to making a new machinery to induce the nondegeneracy of the statistical random field in a general manner by connecting nondegeneracy of the associated tensor fields over the statistical manifold and the nondegeneracy of the underlying stochastic process.

After laying these foundations, we will return to the proof of Theorems 1 and 2 in Section 5.

3 Polynomial type large deviation inequality and a generalized quasi-likelihood analysis for diffusion

Let \( \mathbb{U}_n = \{ u \in \mathbb{R}^p ; \theta^* + (1/\sqrt{n})u \in \Theta \} \) and \( V_n(r) = \{ u \in \mathbb{U}_n ; r \leq |u| \} \).

We make the following assumption.

[H1] (i) \( E[|X_0|^q] < \infty \) for all \( q > 0 \). For every \( q > 0 \), there exists \( C > 0 \) such that

\[
E[|X_t - X_s|^q] \leq C|t - s|^{q/2}
\]

for all \( t, s \in [0, T] \).
(ii) $\sup_{0 \leq t \leq T} E[|b_t|^q] < \infty$ for all $q > 0$.

(iii) $\sigma \in C^{2,4}_T(\mathbb{R}^d \times \Theta; \mathbb{R}^m \otimes \mathbb{R}^r)$ and $\inf_{x,\theta} \det S(x, \theta) > 0$.

We define the random field $Z_n$ on $U_n$ by

$$Z_n(u) = \exp \left\{ \mathbb{H}_n \left( \theta^* + \frac{1}{\sqrt{n}} u \right) - \mathbb{H}_n(\theta^*) \right\}$$

for $u \in U_n$. Let $\mathbb{Y}_n(\theta) = \frac{1}{n} \{ \mathbb{H}_n(\theta) - \mathbb{H}_n(\theta^*) \}$ and

$$\chi_0 = \inf_{\theta \neq \theta^*} \frac{-\mathbb{Y}(\theta)}{|\theta - \theta^*|^2},$$

where

$$\mathbb{Y}(\theta) = -\frac{1}{2T} \int_0^T \left\{ \log \left( \frac{\det S(X_t, \theta)}{\det S(X_t, \theta^*)} \right) + \text{Tr} \left( S^{-1}(X_t, \theta) S(X_t, \theta^*) - I_d \right) \right\} dt.$$

The following condition is concerning nondegeneracy of the index $\chi_0$.

[H2] For every $L > 0$, there exists $c_L > 0$ such that

$$P[\chi_0 \leq r^{-1}] \leq \frac{c_L}{rL}$$

for all $r > 0$.

**Theorem 3.** Assume [H1] and [H2]. Then, for every $L > 0$, there exists a positive constant $C_L$ such that

$$P \left[ \sup_{u \in V_n(r)} Z_n(u) \geq e^{-r} \right] \leq \frac{C_L}{rL}$$

for all $r > 0$ and $n \in \mathbb{N}$.

Proof of Theorem 3 is given in Section 8. The above theorem clarifies the essential role of the random index $\chi_0$ because it gives the polynomial type large deviation estimate for $Z_n$, from which all tail properties of the estimators are deduced as the theorems below.

In order to obtain the weak convergence of the statistical random field on compact sets, we make the following assumption.

[H1$\dagger$] Conditions [A1] and [A2] (i) hold.

The following theorems generalize Theorems 1 and 2.

**Theorem 4.** Assume [H1$\dagger$] and [H2]. Then, $\sqrt{n}(\hat{\theta}_n - \theta^*) \to^{d_f(\mathcal{F})} \Gamma(\theta^*)^{-1/2} \zeta$ and

$$E \left[ f(\sqrt{n}(\hat{\theta}_n - \theta^*)) \right] \to E \left[ f(\Gamma(\theta^*)^{-1/2} \zeta) \right]$$

as $n \to \infty$ for all continuous functions $f$ of at most polynomial growth.
Theorem 5. Assume [H1*] and [H2]. Then, \( \sqrt{n}(\hat{\theta}_n - \theta^*) \to^{d_{k(F)}} \Gamma(\theta^*)^{-1/2} \zeta \) and

\[
E \left[ f(\sqrt{n}(\hat{\theta}_n - \theta^*)) \right] \to \mathbb{E} \left[ f(\Gamma(\theta^*)^{-1/2} \zeta) \right]
\]
as \( n \to \infty \) for all continuous functions \( f \) of at most polynomial growth.

Proof of Theorems 4 and 5 is given in Section 8.

A sufficient condition for [H2] is that

\[
\inf_{\omega \in \Omega, \theta \in \Theta \setminus \{\theta^*\}} \left\{ \log \left( \frac{\det S(X_t, \theta)}{\det S(X_t, \theta^*)} \right) + \text{Tr} \left( S^{-1}(X_t, \theta)S(X_t, \theta^*) - I_d \right) \right\} |\theta - \theta^*|^2 > 0 \quad \text{a.s.}
\]

Though this kind of condition seems easy to handle and at hand, it is too naïve as it breaks, for example, in a simple model such as (5).

The nondegeneracy condition [H2] of the statistical random field is a key to construction of quasi-likelihood analysis for diffusion. As we saw above, once the nondegeneracy of the index \( \chi_0 \) is established, we can obtain limit theorems for the quasi maximum likelihood estimator and the Bayesian type estimator, and moreover convergence of moments of them.

It should be remarked that the limit theorem for the Bayesian type estimator and convergence of moments of these estimators are new, and that the latter is indispensable to practical applications such as model selection, prediction and theory of asymptotic expansion. We will pursue this nondegeneracy problem for statistical random fields. The question is when Condition [H2] holds. We discuss this problem in Section 4. It involves a new technical aspect.

4 Nondegeneracy of the statistical random field

4.1 Preliminary estimates

Let \( J \in \mathbb{N} \). For \( c = (c_0, c_1, \ldots, c_p) \in \mathbb{R}^{p+1} \), set

\[
p(c, x) = c_0 + c_1 x + \cdots + c_p x^p.
\]

For \( \delta > 0 \), let \( C_\delta = \{c; |c_0| + |c_1| + \cdots + |c_p| \geq \delta\} \). For \( \epsilon \in (0, 1) \), let \( U_\epsilon = \{u = (u_j)_{j=0}^p; \epsilon \leq \inf_j |u_j| \leq \sup_j |u_j| \leq \epsilon^{-1}\} \). Let \( c * u = (c_i u_j) \).

Lemma 1. For any distinct positive numbers \( \{\alpha_i\}_{i=0}^p \) and \( \delta, \epsilon > 0 \), there exist numbers \( L > 0 \) and \( n_0 \in \mathbb{N} \) such that

\[
\inf_{c \in C_\delta} \max_{0 \leq i \leq p} \inf_{u \in U_\epsilon} |p(c * u, n^{-\alpha_i})| \geq n^{-L} \quad \text{for all integers } n \geq n_0.
\]

Proof. Without loss of generality, we may assume \( \alpha_0 > \alpha_1 > \cdots > \alpha_p \). Let \( u_i = (u_{ij})_{j=0}^p = 0 \) for \( j = 0, 1, \ldots, p \). Define a \((p+1) \times (p+1)\)-matrix \( A_n = [a_{ij}]_{i,j=0}^p \) by

\[
a_{ij}^n = u_{ij} n^{-\alpha_i}.
\]

Let \( w = A_n c \), and write \( w = (w_0, \ldots, w_p) \) and \( c = (c_0, \ldots, c_p) \). For \( i = 0, \ldots, p \), the \( w_i \) is a function of \( n, c \) and \( u \); \( w_i = w_i(n, c, u) \). Suppose that \( u_i \in U_\epsilon \). By the fact that \( \det A_n \sim \prod_{i=0}^p u_{ij} n^{-\alpha_i} \) as \( n \to \infty \), there exists a number \( n_0 \in \mathbb{N} \) such that \( |\det A_n| \geq 2^{-1} |\prod_{i=0}^p u_{ij} n^{-\alpha_i}| \) for all \( n \geq n_0 \) and all.
\( \mathbf{u}_i \in \mathcal{U}_\ell \) (\( i = 0, \ldots, p \)). Since \( \mathbf{c} = A_n^{-1} \mathbf{w} \), there exists a constant \( K \) depending only on \( p \) and \( \epsilon \) and it holds that

\[
|c_i| \leq Kn^{L'} \sum_{j=0}^p |w_j(n, \mathbf{c}, \mathbf{u}_j)| \quad \text{for all } n \geq n_0, \mathbf{c} \in \mathbb{R}^{p+1}, \mathbf{u}_j \in \mathcal{U}_\ell \quad (j = 0, \ldots, p),
\]

where \( L' = \sum_{i=0}^p \alpha_i i \). For each \( \mathbf{c} \), the function \( \mathcal{U}_\ell \ni \mathbf{u}_j \mapsto |w_j| = |w_j(n, \mathbf{c}, \mathbf{u}_j)| \) is obviously continuous, and hence \( \inf_{\mathbf{u}_j \in \mathcal{U}_\ell} |w_j| = |w_j(n, \mathbf{c}, \mathbf{u}_j^*)| \) for some \( \mathbf{u}_j^* = \mathbf{u}_j^*(n, \mathbf{c}) \in \mathcal{U}_\ell \). Applying the above inequality to \( \mathbf{c} \in C_\delta \), \( \mathbf{u}_j^* \) (\( j = 0, \ldots, p \)) and \( n \geq n_0 \), we have

\[
\delta \leq \sum_{i=0}^p |c_i| \leq Kn^{L'} \sum_{i=0}^p \sum_{j=0}^p |w_j(n, \mathbf{c}, \mathbf{u}_j^*)| = (p + 1)Kn^{L'} \sum_{j=0}^p \inf_{\mathbf{u}_j \in \mathcal{U}_\ell} |w_j(n, \mathbf{c}, \mathbf{u}_j)|.
\]

The relation \( p(\mathbf{c} + \mathbf{u}, n^{-\alpha}) = w_i(n, \mathbf{c}, \mathbf{u}) \) completes the proof. \( \square \)

Set \( S = \{ \eta \in \mathbb{R}^d, |\eta| = 1 \} \). Let \( \Theta \) be a subset of \( \mathbb{R}^p \). Let \( J \in \mathbb{N} \). Let \( f(x, \theta) \) is a function defined on a neighborhood of \( X_0 \times \Theta \).

Let \( L_\epsilon = T_x \mathbb{R}^d \simeq \mathbb{R}^d \). A cone in the direction of \( \xi \in \mathbb{L} \setminus \{ 0 \} \) is defined by \( C(\xi, a) = \{ \eta \in L; \eta \cdot \xi \geq (1 - a)|\eta||\xi| \} \) for \( a \in (0, 1) \). Let \( S(\xi, \epsilon) = S(\xi, \epsilon) \cup S(-\xi, \epsilon) \) and \( \tilde{S}(\xi, \epsilon) = S(\xi, \epsilon) \cup S(-\xi, \epsilon) \), where \( S(\xi, \epsilon) = C(\xi, \epsilon) \cap S \) and \( S(\xi, \epsilon) = C(\xi, \epsilon) \cap \{ \eta \in \mathbb{R}^d; |\eta| \leq \epsilon \} \). Let \( X_\ell (\ell = 1, \ldots, \tilde{\ell}) \) be subsets of \( X_0 \) and \( X_0 = \bigcup_{\ell} X_\ell \). Let \( \tilde{S}_{x_0}(\xi, \epsilon) = x_0 + \tilde{S}(\xi, \epsilon) \).

\[[N_0^3] \text{ There exist } \xi_{\ell,k} \in S, \epsilon_{\ell,k} > 0, \text{ and } \Theta_{\ell,k} \subset \Theta \text{ for } \ell = 1, \ldots, \tilde{\ell} \text{ and } k = 1, \ldots, \tilde{k}_{\ell} \text{ with } \Theta = \bigcup_k \Theta_{\ell,k} \text{ for each } \ell \text{, and bounded functions } b_{j,\ell,k} : X_\ell \times \Theta_{\ell,k} \to \mathbb{R} \text{ for } j = 0, \ldots, J - 1 \text{ and a bounded function } b_{j,\ell,k} : X_\ell \times \tilde{S}_{x_0}(\xi_{\ell,k}, \epsilon_{\ell,k}) \times \Theta_{\ell,k} \times S \to \mathbb{R} \text{ such that the following conditions are fulfilled.} \]

(i) For each \( \ell = 1, \ldots, \tilde{\ell} \) and \( k = 1, \ldots, \tilde{k}_{\ell} \),

\[
|f(x, \theta)| \geq |G_{\ell,k}(x_0, x, \theta, x - x_0)|
\]

for all \( x_0 \in X_\ell, x \in \tilde{S}_{x_0}(\xi_{\ell,k}, \epsilon_{\ell,k}) \) and \( \theta \in \Theta_{\ell,k}, \) where

\[
G_{\ell,k}(x_0, x, \theta, \xi) = \sum_{j=0}^{J-1} b_{j,\ell,k}(x_0, \theta)(\xi_{\ell,k} \cdot \xi)^j + b_{j,\ell,k}(x_0, \theta, \xi, |\xi|^{-1} \xi)(\xi_{\ell,k} \cdot \xi)^j.
\]

(ii) For each \( \ell = 1, \ldots, \tilde{\ell} \) and \( k = 1, \ldots, \tilde{k}_{\ell} \),

\[
\inf_{(x_0, \theta) \in X_\ell \times \Theta_{\ell,k}} \max_{j=0, \ldots, J-1} |b_{j,\ell,k}(x_0, \theta)| > 0.
\]

It will be shown that \([N_0^3]\) follows from

\[[N_1^3] \text{ (i) } X_0 \text{ and } \Theta \text{ are compact.} \]

\[(ii) \text{ } f \text{ admits a } C^J \text{-supporting function for each } (x_0, \theta) \in X_0 \times \Theta. \]

\[(iii) \text{ For the } C^J \text{-supporting function in (ii) for each } (x_0, \theta) \in X_0 \times \Theta, \text{ max}_{j=0, \ldots, J-1} |c_j(x_0, \theta)| > 0. \]

\footnote{This section is presented independently of the previous sections. In particular, if we assume the compactness of } \Theta, \text{ it corresponds to } \Theta \text{ of other sections.}
Lemma 2. $[N_0^2]$ implies $[N_0^0]$ for some $\{\xi_r\}_r$.

Proof. For each $(x_0, \theta) \in X_0 \times \Theta$, there exist $j(x_0, \theta) \in \{0, \ldots, J - 1\}$ and $\xi(x_0, \theta) \in \mathbb{S}$ such that
\[
c_j(x_0, \theta)(x_0, \theta) \neq 0.
\]
By continuity, there exist an open neighborhood $V(x_0, \theta)$ of $(x_0, \theta)$ such that
\[
\inf \left\{ |c_j(x_0, \theta)(x_0', \theta')| : (x_0', \theta') \in V(x_0, \theta) \right\} > 0
\]
The family $\{V(x_0, \theta)\}_{(x_0, \theta) \in X_0 \times \Theta}$ is an open covering of the compact set $X_0 \times \Theta$, consequently there are $(x_p, \theta_p) \in X_0 \times \Theta$ ($p = 1, \ldots, \bar{p}$) such that $\cup_{p=1}^{\bar{p}} V(x_p, \theta_p) = X_0 \times \Theta$. For each $n \in \mathbb{N}$, consider a family $U = \{U(n, m)\}_{m \in \mathbb{N}}$ of sets each of which is of the form
\[
\left\{(2^{-n}i_1, 2^{-n}(i_1 + 1)) \times \cdots \times (2^{-n}i_{d+p}, 2^{-n}(i_{d+p} + 1))\right\} \cap (X_0 \times \Theta)
\]
for some $i_1, \ldots, i_{d+p} \in \mathbb{Z}$. For each $(x_0, \theta) \in X_0 \times \Theta$, there exists $U(n(x_0, \theta), m(x_0, \theta)) \in U$ such that the closure $\overline{U(n(x_0, \theta), m(x_0, \theta))}$ is included in some $V(x_p, \theta_p)$. The family $\{U(n(x_0, \theta), m(x_0, \theta))\}_{(x_0, \theta) \in X_0 \times \Theta}$ forms an open covering of the compact set $X_0 \times \Theta$, therefore $U$ is already covered by a finite family $\{U(n_i, m_i)\}_{i=1, \ldots, \bar{n}}$ of $U(n, m)$ into some elements of $\{U(n, m)\}_{m \in \mathbb{N}}$. Thus we obtained a partition $\{\xi_{\ell,k}\}_{\ell=1, \ldots, \bar{\ell}, k=1, \ldots, \bar{k}}$ of $X_0 \times \Theta$ such that $X_0 = \bigcup_{\ell=1}^{\bar{\ell}} \xi_{\ell,k}$, $\Theta = \bigcup_{k=1}^{\bar{k}} \Theta_k$, and each $\xi_{\ell,k} \times \Theta_k$ is included in some $V(x_p(\ell,k), \theta_p(\ell,k))$. Set $\xi_{\ell,k} = \xi(x_p(\ell,k), \theta_p(\ell,k))$.

Based on the above construction of the covering, the function $G_{\ell,k}(x_0, x, \theta', \xi)$ is defined in a neighborhood of $(x_0, \theta)$ through the expansion of the support function $g(P_{x_0}, x, \theta')$ in $x$ around $x_0$ of $(x_0', \theta')$ near $(x_0, \theta)$ as follows. Recall $c_j(x, \theta') = (j!)^{-1} \partial_j^{\ell} \{g(P_{x_0}, x, \theta')\}_{0}^{\infty}$. $[c_j(x, \theta')$ depends on $(x_0, \theta)$. ] Let $b_{j,\ell,k}(x_0', \theta') = c_j(x_0', \theta')$ for $j = 0, \ldots, J - 1$ and
\[
b_{j,\ell,k}(x_0', x, \theta', |\xi|^{-1} \xi) = \int_0^1 (1 - s)^{-1} c_j(x_0' + s(x_0, \theta'), \theta') ds.
\]
[In this case, $b_j$ does not depend on $\xi$. ] Then obviously $G_{\ell,k}(x_0', x, \theta', \xi)$ defined by the expansion of $[N_0^0]$ (i) satisfies $G_{\ell,k}(x_0', x, \theta', x - x_0') = g(P_{x_0}, x, \theta')$. We choose $\epsilon(x_0, \theta)$ sufficiently small so that the inequality of $[N_0^0]$ (i) is valid.

Let $E(\xi_0) = \{\xi \in \mathbb{S} : \xi \cdot \xi_0 = 1\}$ for $\xi_0 \in \mathbb{S}$. Set $D(\xi_0, \epsilon) = E(\xi_0) \cap C(\xi_0, \epsilon)$ and $\tilde{D}(\xi_0, \epsilon) = D(\xi_0, \epsilon) \cup D(-\xi_0, \epsilon)$.

Lemma 3. Suppose that $[N_0^0]$ is fulfilled. Then for any distinct positive numbers $\{\alpha_j\}_{j=0}^J$, there exist $L > 0$ and $n_0 \in \mathbb{N}$ such that
\[
\min_{\ell=1, \ldots, \bar{\ell}, k=1, \ldots, \bar{k}} \inf_{(x_0, \theta) \in \xi_{\ell,k}} \max_{i=0, \ldots, J} \inf_{\xi \in \tilde{D}(\xi_0, \epsilon)} |f(x_0 + n^{-\alpha_i} \xi, \theta)| \geq n^{-L}
\]
for all $n \geq n_0$. This estimate is also valid for $|G_{\ell,k}(x_0, \cdot, \cdot, x_0)|$ in place of $|f|$.

Proof. It follows from $[N_0^0]$ (ii) that for some $\delta > 0$,
\[
\min_{(\ell,k) \in \{1, \ldots, \bar{\ell}\} \times \{1, \ldots, \bar{k}\}} \inf_{(x_0, \theta) \in \xi_{\ell,k}} \max_{j=0, \ldots, J - 1} |b_{j,\ell,k}(x_0, \theta)| > \delta.
\]
For \((x_0, \theta) \in X_\ell \times \Theta_{\ell,k}\), for some \(j_1 = j_1(x_0, \theta) \in \{0, \ldots, J - 1\}\), \(|b_{j_1,\ell,k}(x_0, \theta)| > \delta\). Then there exists \(n_1 \in \mathbb{N}\) such that
\[
\sum_{j=j_1}^{J-1} b_{j,\ell,k}(x_0, \theta)n^{-\alpha_i j} + b_{j,\ell,k}(x_0, \theta, |\xi|^j)\xi n^{-\alpha_i j} = b_{j_1,\ell,k}(x_0, \theta)n^{-\alpha_j j_1} \left(1 + \epsilon(n, x, \theta, |\xi|, i, j_1)\right)
\]
for \(n \geq n_1\), where \(|\epsilon(n, x, \theta, |\xi|, i, j, j_1)| \leq 1/2\) and \(n_1\) depends only on \(\delta\), \(J\), \(\alpha_i\) and \(\|b_{j,\ell,k}\|\). Now we can apply Lemma 1 to \(p = j_1\) with \(u_{ij} = 1\) for \(i = 0, \ldots, j_1 - 1\) and \(u_{ij_1} = 1 + \epsilon(n, x, \theta, |\xi|, i, j_1)\).

4.2 Nondegeneracy of the index \(\chi_0\)

Suppose that \(X = (X_t)_{t \in [0, T]}\) is a \(d\)-dimensional separable process. \(\Theta\) is a set in \(\mathbb{R}^p\). \(X_0\) and \(\hat{X}\) are subsets of \(\mathbb{R}^d\) with \(X_0 \subset \hat{X}\), the interior of \(\hat{X}\) in \(\mathbb{R}^d\). Let \(\theta^* \in \Theta\) and let
\[
\chi_0 = \inf_{\theta \in \Theta, \theta \neq \theta^*} \frac{\int_0^T \mathbb{Q}(X_t, \theta)dt}{|\theta - \theta^*|^2}
\]
for a function \(\mathbb{Q} : \hat{X} \times \Theta \rightarrow \mathbb{R}\).

Furthermore, suppose

\[\text{[R]}\] There exist a function \(f : \hat{X} \times \Theta \rightarrow \mathbb{R}\) and a constant \(\varrho \in (0, \infty)\) satisfying
\[
\mathbb{Q}(x, \theta) |\theta - \theta^*|^2 \geq |f(x, \theta)|^\varrho
\]
for all \((x, \theta) \in \hat{X} \times \Theta\), and the function \(f(\cdot, \theta)\) is Lipschitz continuous on \(\hat{X}\) uniformly in \(\theta \in \Theta\).

Here \(f\) and \(\varrho\) possibly depend on \(\theta^*\). An example of \(f\) is \(\mathbb{Q}(x, \theta) |\theta - \theta^*|^2\) itself for \(\varrho = 1\), however we have much more freedom of choice of \(f\) and \(\varrho\). Introducing the subfield \(f\) facilitates application of the result.

We denote by \(\mathcal{P}\) the set of sequences \((a_n)_{n \in \mathbb{N}}\) of nonnegative numbers satisfying the condition that for every \(L > 0\), there exists a number \(C_L\) such that \(a_n \leq C_L/n^L\) for all \(n \in \mathbb{N}\). \(\mathcal{E}\) denotes the set of sequences \((a_n)_{n \in \mathbb{N}}\) of nonnegative numbers such that for some \(c > 0\), \(a_n \leq c^{-1}e^{-cn}\) for all \(n \in \mathbb{N}\).

Now we assume the following conditions for the nondegeneracy of the deterministic field and the variation of the underlying stochastic process. Condition [C] is for estimate of a modulus of continuity of \(X\).

\[\text{[N]}\] There exist \(T_0 \in (0, T)\), subsets \(X_\ell \subset \hat{X}\) \((\ell = 1, \ldots, \ell)\) with \(X_0 \subset \bigcup_\ell X_\ell\) and \(\overline{X_\ell} \subset \hat{X}\) for which the following conditions hold:

\[\text{(i) [N] holds and } \bigcup_{\ell, k} B(x_0, \epsilon, k) \subset \hat{X}\text{.}\]

---

2This section gives a way to the estimate of the key index \(\chi_0\). Since the method is general, we write it for a general stochastic process \(X\), apart from the Itô process \(X\) in Section 3.

3We use the same symbol \(\chi_0\) for the key index as Section 3 since we will apply the nondegeneracy results here to \(\mathbb{Q} = Q(x, \theta^*)/2T\) in Section 5.

4\(B(x, \epsilon)\) is the open ball centered \(x\) with radius \(\epsilon\).
(ii) For each \((\ell, k)\), there exist a positive constants \(c_0\) and distinct positive numbers \(\{\alpha_j := \alpha_j(\ell, k)\}_{j=0}^I\) such that the sequence
\[
1 - P \left[ \bigcup_{\ell=1, \ldots, \tilde{\ell}} \bigcap_{k=1, \ldots, \tilde{k}_k} \bigcup_{\ell_0, \ldots, \ell_0} \bigcup_{j_0, \ldots, j_0} \bigg\{ X_{\ell, k} \in \mathcal{X}_{\ell, k}, X_{\ell_0, k_0} - X_{\ell_0, k_0} = \tilde{S}(\xi_{\ell, k}, \epsilon_{\ell, k}) \right. \\
\left. \quad \text{and } |\xi_{\ell, k} \cdot (X_{\ell_0, k_0} - X_{\ell_0, k_0})| = n^{-\alpha_j} \text{ for some } t \in [0, n^{-c_0}] \bigg\} \bigg\} \right]_{n \in \mathbb{N}}
\]

is in \(\mathcal{P}\).

[C] There exist positive constants \(\beta_0\) such that the sequence
\[
1 - P \left[ \sup_{s, t \in [0, T]: X_s \in \mathcal{X}, t \in [s, s+n^{-1}]} |X_t - X_s| \geq n^{-\beta_0} \right] \right] \right]_{n \in \mathbb{N}}
\]

is in \(\mathcal{P}\).

**Proposition 1.** Under \([\mathcal{N}_0]\), \([R]\) and \([C]\), Condition \([H2]\) holds true.\(^5\)

**Proof.** Let
\[
\Omega_{n, \ell, k, j, s} = \left\{ X_{\ell, k} \in \mathcal{X}_{\ell, k} \bigg\{ X_{\ell_0, k_0} - X_{\ell_0, k_0} \in \tilde{S}(\xi_{\ell, k}, \epsilon_{\ell, k}) \right. \\
\left. \quad \text{and } |\xi_{\ell, k} \cdot (X_{\ell_0, k_0} - X_{\ell_0, k_0})| = n^{-\alpha_j} \text{ for some } t \in [0, n^{-c_0}] \bigg\} \bigg\}.
\]

In what follows, we consider sufficiently large \(n\). For \(\omega \in \bigcap_{j=0, \ldots, J} \Omega_{n, \ell, k, j, s}\) and \(\theta \in \Theta_{\ell, k}\), there are random times \(\tau_j = \tau_j(\omega, n, \ell, k, s) \in (s, s+n^{-c_0}]\) such that
\[
X_{\tau_j} - X_s \in \tilde{S}(\xi_{\ell, k}, \epsilon_{\ell, k}) \quad \text{and } |\xi_{\ell, k} \cdot (X_{\tau_j} - X_s)| = n^{-\alpha_j}.
\]

Then Lemma 3 yields
\[
\max_{j=0, \ldots, J} |f(X_{\tau_j}, \theta)| \geq \min_{\ell=1, \ldots, \tilde{\ell}} \inf_{k=1, \ldots, \tilde{k}_k} \left\{ \int_0^T |f(X_{\ell, k}, \theta)|^2 dt \leq n^{-L} \right\}
\]

for \(n \geq n_0\), where \(L > 0\) and \(n_0\) are depending only on \(\{\alpha_j\}_{j=0}^I\) and independent of \(\omega \in \bigcap_{j=0, \ldots, J} \Omega_{n, \ell, k, j, s}\) and \(\theta \in \Theta_{\ell, k}\).

Take \(\kappa \in \mathbb{N}\) such that \(\kappa > L_0[\beta_0(1 + \beta_1)]^{-1}\), and let \(L' > \kappa + L_0\). We have
\[
P[|\chi_0| \leq n^{-L'}] = P \left[ \inf_{\theta \in \Theta} \int_0^T \frac{Q(X_{\ell, k}, \theta)}{|\theta - \theta|^2} dt \leq n^{-L'} \right]
\]

\[
\leq \sum_{\ell=1}^{\tilde{\ell}} P \left[ \left\{ \inf_{\theta \in \Theta_{\ell, k}} \int_0^{T} |f(X_{\ell, k}, \theta)|^2 dt \leq n^{-L'} \right\} \bigg\} \bigg\} \bigg\} \right]_{\tilde{k}_k} \bigg\} \bigg\} + a_n
\]

\[
\leq \sum_{\ell=1}^{\tilde{\ell}} P \left[ \left\{ \inf_{\theta \in \Theta_{\ell, k}} \int_0^{T} |f(X_{\ell, k}, \theta)|^2 dt \leq n^{-L'} \right\} \bigg\} \bigg\} \right]_{\tilde{k}_k} \bigg\} \bigg\} \bigg\} \bigg\} \bigg\} \bigg\} \bigg\} \bigg\} \bigg\} \bigg\} \bigg\} \bigg\} + a_n
\]

\(^5\)Of course, “\(\chi_0\)” is the one in this section.
with \((a_n)_{n \in \mathbb{N}} \in \mathcal{P}\).

Thanks to [C], we have

\[
\begin{align*}
P\left[ \inf_{\theta \in \Theta, k} \int_0^T |f(X_t, \theta)|^\theta \, dt \leq n^{-L'} \right] \ &= \ P\left[ \bigcup_{s \in [0, T]} \bigcap_{j=0, \ldots, J} \bigcup_{\Omega_{n, \ell, k, j, s}} \Omega_{n, \ell, k, j, s} \right] \\
&\leq P\left[ \inf_{\theta \in \Theta, k} \max_{j=0, \ldots, J} |f(X_{\tau_j}, \theta)|^\theta \left( n^{-\kappa} - \sup_{\theta \in \Theta, k} \max_{j=0, \ldots, J} \int_{\tau_j}^{\tau_j + n^{-\kappa}} |f(X_t, \theta)|^\theta - |f(X_{\tau_j}, \theta)|^\theta \right) \, dt \right] \\
&\leq b_n
\end{align*}
\]

for some \((b_n)_{n \in \mathbb{N}} \in \mathcal{P}\). Indeed, \(n^{-L_0 - \kappa} - An^{-\kappa - \kappa \beta_0(\theta \wedge 1)} > n^{-L'}\) as \(n\) becomes large for every \(A > 0\), and

\[
\sup_{\theta \in \Theta, k} \int_{\tau_j}^{\tau_j + n^{-\kappa}} \left| f(X_t, \theta)^\theta - f(X_{\tau_j}, \theta)^\theta \right| \, dt \leq n^{-\kappa - \kappa \beta_0(\theta \wedge 1)}
\]

on the event \(\{ \sup_{s, t \in [0, T]} |X_t - X_s| \leq n^{-\kappa \beta_0(\theta \wedge 1)} \}\), where \(\tilde{X}'\) is a suitable neighborhood of \(\overline{\cup_{t} X_t}\). Therefore, we obtain \(P[\chi_0 \leq n^{-L}] = \epsilon_n\) for some \((\epsilon_n)_{n \in \mathbb{N}} \in \mathcal{P}\). This gives [H2].

\[\Box\]

**Remark 1.** In Proposition \([\mathbb{N}]\), if the sequences in \([N_0]\)(ii) and [C] are in \(\mathcal{E}\), then we obtain \(P[\chi_0 \leq r^{-1}] \leq c^{-1} e^{-cr^2}\) \((r > 0)\) for some \(c > 0\). The same remark is also for Proposition \([\mathbb{R}]\).

**Remark 2.** If we strengthen \([N_0]\) (ii) by replacing \(\tilde{S}(\xi_{\ell, k}, \epsilon_{\ell, k})\) by \(S(\xi_{\ell, k}, \epsilon_{\ell, k})\), then the inequality in Lemma \([\mathbb{L}]\) but with \(D(\xi_{\ell, k}, \epsilon_{\ell, k})\) for \(D(\xi_{\ell, k}, \epsilon_{\ell, k})\) is still sufficient to prove the same result as Proposition \([\mathbb{R}]\). This formulation will work for nondegenerate diffusions. However the original one is worth stating because it is easy to give an example such that the process \(X\) moves toward \(\xi_{\ell, k}\) or \(-\xi_{\ell, k}\) with probability \(1/2\).

When the process \(X\) varies in any direction, it finds a nondegenerating direction \(\xi\) locally uniformly in \((x, \theta)\).

\[\mathbb{N}\]

(i) \(X_0\) and \(\Theta\) are compact.

(ii) \(f\) admits a \(C^1\)-supporting function for each \((x_0, \theta) \in X_0 \times \Theta\) with \(\max_{j=0, \ldots, J-1} |c_j(x_0, \theta)| > 0\).

(iii) There exists a stopping time \(\tau\) satisfying \(\text{ess.sup}_{\omega, \tau} \tau < T\) and \(\text{supp} \mathcal{L}\{X_\tau\} \subset (X_0)^0\) and there exist a positive constant \(c_0\) and distinct positive numbers \(\{\alpha_j\}_{j=0, \ldots, J}\) such that \(\min_j \alpha_j > c_0/2\) and that for every \(\xi \in S\) and \(\epsilon > 0\), the sequence

\[
\left(1 - P\left[ \bigcup_{s \in [\tau, \tau + n^{-\alpha_j}]} \bigcap_{j=0, \ldots, J} \bigcup_{t \in [0, n^{-\alpha_j}]} \left\{ X_{s+t} - X_s \in \tilde{S}(\xi, \epsilon) \right\} \right] \right)_{n \in \mathbb{N}}
\]

is in \(\mathcal{P}\).

**Proposition 2.** Under \([\mathbb{N}]\), \([\mathbb{R}]\) and \([C]\), Condition \([H2]\) holds.
Proof. Conditions $[N_1]$ (i) and (ii) imply $[N_1^0]$, and hence $[N_0^0]$ by Lemma 2. Then $\xi_{\ell,k}, \epsilon_{\ell,k}, X_\ell$ and $\Theta_{\ell,k}$ are given by $[N_0^0]$. By observing the proof of Lemma 2, we may take sufficiently small $X_\ell$ and $\epsilon_{\ell,k}$. Choose small $X$ as a neighborhood of $X_0$.

For $\xi = \xi_{\ell,k}$ and $\epsilon = \epsilon_{\ell,k}$, we apply $[N_1]$(iii) to ensure $[N_0](ii)$ as follows. By the representation of $b_j(x_0, \theta, \xi)$ in the proof of Lemma 2, we can replace $X_\ell$ by an open set $\bar{X}_\ell$ ($X_\ell \subset \bar{X}_\ell \subset X$) in $[N_2^0]$. Denote by $(\alpha_n)_{n \in \mathbb{N}}$ a generic element of $\mathcal{P}$. It changes from line to line. Let

$$B(n, \ell, k, j, s, t) = \left\{ X_{s+t} - X_s \in \bar{S}(\xi_{\ell,k}, \epsilon_{\ell,k}) \text{ and } |\xi_{\ell,k} \cdot (X_{s+t} - X_s) - \epsilon_{\ell,k}| = n^{-\alpha_j} \right\}.$$  

By $[N_1]$ (iii),

$$P\left[ \bigcap_{\ell=1, \ldots, \hat{k}=1, \ldots, \hat{k} \in [\tau, \tau+n^{-\epsilon_0}]} \bigcup_{\ell, k, s \in [\tau, \tau+n^{-\epsilon_0}]} B(n, \ell, k, j, s, t) \right] = 1 - \alpha_n.$$  

Let $A_\ell = \{ X_\tau \in X \}$ for $\ell = 1, \ldots, \hat{\ell}$. Since

$$\left( \bigcup_{\ell} A_\ell \right) \cap \left( \bigcup_{\ell, k, j, s} B(n, \ell, k, j, s, t) \right) \subset \bigcup_{\ell, k, j, s} \left( A_\ell \cap B(n, \ell, k, j, s, t) \right)$$

$$\subset \left[ \bigcup_{\ell, k, j, s} \left( \{ X_s \in \bar{X}_\ell \} \cap B(n, \ell, k, j, s, t) \right) \right] \bigcup \left[ \bigcup_{\ell, s} \{ X_\tau \in X_\ell, X_s \notin \bar{X}_\ell \} \right],$$

we have

$$1 - \alpha_n \leq P\left[ \bigcup_{\ell=1, \ldots, \hat{k}=1, \ldots, \hat{k} \in [\tau, \tau+n^{-\epsilon_0}]} \bigcup_{\ell, k, j, s} \left\{ X_s \in \bar{X}_\ell, X_{s+t} - X_s \in \bar{S}(\xi_{\ell,k}, \epsilon_{\ell,k}) \text{ and } |\xi_{\ell,k} \cdot (X_{s+t} - X_s) - \epsilon_{\ell,k}| = n^{-\alpha_j} \right\} \right].$$

This inequality implies $[N_0](ii)$. Consequently, we can apply Proposition 1 to conclude $[H2]$.  

**Remark 3.** (i) Corresponding to Remark 2 if we assume Condition $[N_1]$ (iii) with $\bar{S}(\xi_{\ell,k}, \epsilon_{\ell,k})$ replaced by $S(\xi_{\ell,k}, \epsilon_{\ell,k})$, then the inequality in Lemma 3 with $D(\xi_{\ell,k}, \epsilon_{\ell,k})$ for $\bar{D}(\xi_{\ell,k}, \epsilon_{\ell,k})$ is still sufficient to prove the same result as Proposition 2. Note that Lemma 3 was implicitly used in the proof of Proposition 2 through Proposition 1 (ii) For Proposition 2 we can eliminate the Lipschitz continuity condition in $[R]$ because we can replace $f$ by the supporting function in the last part of the proof of Proposition 1.

**5 Proof of Theorems 1 and 2**

It suffices to verify $[H1^2]$ and $[H2]$ due to Theorems 1 and 5. We take $\Theta$ for “$\Theta$” in $[N_1]$. Condition $[H1^2]$ is obviously satisfied under the assumptions.

We shall show $[H2]$. For this, we will apply Proposition 2. Let $X_0$ and $\bar{X}$ be compact sets in $\mathbb{R}^d$ such that supp$(X_\tau) \subset X_0 \subset X_0 \subset X \subset U$. Condition $[C]$ is easily verified for $\beta_0 \in (0, 1/2)$. Conditions $[R]$, $[N_1](i)$, (ii) are obvious.
Take numbers \( c_0, c_1, \alpha_j \) \((j = 0, ..., J)\) such that \( \alpha_0 > \cdots > \alpha_J > c_1/2 > c_0/2. \) Let \( T(\xi_0, \epsilon) = C(\xi_0; \epsilon) \cap \{\xi \in \mathbb{L}; 3 < \xi \cdot \xi_0 < 4\} \) for \( \epsilon > 0, \) \( T(\xi_0, \epsilon, \eta) = \{\eta \xi; \xi \in T(\xi_0, \epsilon)\} \) for \( \eta > 0. \) Let \( s(n, k) = \tau + kn^{-c_1} \) for \( k = 1, ..., k(n), \) \( k(n) = [n^{c_1-c_0}] \). Let \( s(n, k, -1) = s(n, k), s(n, k, 0) = s(n, k) + n^{-2\alpha_0} \) and \( s(n, k, j) = s(n, k, j - 1) + n^{-2\alpha_j} \) for \( j = 1, ..., J. \) Obviously, \( s(n, k) \) and \( s(n, k, j) \) are stopping times. We may assume that \( s(n, k, J) \leq s(n, k + 1) \).

Let \( \epsilon_0 > 0. \) Let

\[
A^X(\xi, \epsilon_0, n, k, j) = \left\{ X_u - X_s(n, k, j - 1) \in T(\xi, \epsilon_0, n^{-\alpha_j}) \text{ for some } u \in (s(n, k, j - 1), s(n, k, j)] \right\}
\]

\[
\cap \left\{ \sup_{u \in (s(n, k, j - 1), s(n, k, j)]} |P^\xi(\mathbb{X}_u - X_s(n, k, j - 1))| \leq \epsilon_0 n^{-\alpha_j} \right\}
\]

for a process \( \mathbb{X} \), where \( P^\xi : \mathbb{L} \rightarrow \mathbb{L} \) is the orthogonal projection on \( \mathbb{L} \) to the subspace orthogonal to \( \xi \). We write \( X_t \) for \( t \geq \tau \) as \( X_t = X_\tau + M_t + R_t \), where

\[
M_t = a_\tau(w_t - w_\tau) + b_\tau(\hat{w}_t - \hat{w}_\tau)
\]

and

\[
R_t = \int_\tau^t (a_s - a_\tau)dw_s + \int_\tau^t (\tilde{a}_s - \tilde{a}_\tau)d\tilde{w}_s + \int_\tau^t \tilde{b}_sds.
\]

The process \( (M_{t-\tau})_{t \in [\tau, \tau + n^{-c_0}]} \) has the same law on \( C([0, n^{-c_0}]; \mathbb{R}^d) \) as \( ((a^{\otimes^2}_\tau + \tilde{a}^{\otimes^2}_\tau)^{1/2} B_t)_{t \in [0, n^{-c_0}]} \), where \( B_t \) is a \( d \)-dimensional standard Wiener process independent of \( \mathcal{F}_\tau \). Using the scaling property and independency between increments of the Wiener process, and also a classical result of the distribution of its absolute deviation or a support theorem, it is easy to see

\[
q := \text{ess.inf}_{\omega, \xi \in \mathbb{S}, n, k, j} P[A^M(\xi, \epsilon_0, n, k, j)|\mathcal{F}_{s(n, k, j - 1)}] > 0.
\]

It should be noted that the uniform (in \( \omega \)) boundedness and the uniform (in \( \omega \)) nondegeneracy of the matrix \( (a^{\otimes^2}_\tau + \tilde{a}^{\otimes^2}_\tau)^{1/2} \) was used to control random linear transform of the Brownian motion \( B_t \).

For any \( \epsilon_1, \epsilon_2 > 0, \) there exists \( n_0 \in \mathbb{N} \) such that

\[
\text{ess.sup}_{\omega} \sup_{n \geq n_0 \atop k = 1, ..., k(n)} \sup_{j = 0, ..., J} P\left[ \left| n^{\alpha_j}(R_t - R_{s(n, k, j - 1)}) \right| > \epsilon_1 |\mathcal{F}_{s(n, k, j - 1)} \right] < \epsilon_2.
\]

Indeed, Lenglart’s inequality gives uniform estimates for stochastic integrals with the aid of the right-continuity of \( a \) and \( \tilde{a} \) as well as the \( L^p \)-boundedness provided in [A2]. For the integral of \( b \), the Hölder inequality with \( L^p \)-estimate for \( \tilde{b}_t \) yields the estimate. In order to check \([N_1](iii)\), we consider arbitrary \( \xi \in \mathbb{S} \) and \( \epsilon > 0. \) We choose positive constants \( \epsilon_0, \epsilon_1 \) and \( \epsilon_2 \) such that \( \epsilon_1 << \epsilon_0 << \epsilon \) and \( \epsilon_2 < q \). Then

\[
\text{ess.inf}_{\omega} \inf_{n \geq n_0 \atop k = 1, ..., k(n)} \sup_{j = 0, ..., J} P[A(\xi, \epsilon, n, k, j)|\mathcal{F}_{s(n, k, j - 1)}] > q - \epsilon_2 =: q' > 0,
\]

where \( A(\xi, \epsilon, n, k, j) \) is the event defined in the same way as \( A^X \) for \( X = X \) and \( \epsilon_0 = \epsilon \), with \( T(\xi_0, \epsilon) \) replaced by \( C(\xi_0; \epsilon) \cap \{\xi \in \mathbb{L}; 2 < \xi \cdot \xi_0 < 5\} \).
Let $\epsilon_3 > 0$. Let

$$B(n, k, j - 1) = \left\{ n^{\alpha_j} | X_{s(n,k)} - X_{s(n,k,j-1)} | < \epsilon_3 \right\}.$$ 

We see that for $\epsilon_4 > 0$, there exists $n_1 \geq n_0$ such that

$$\text{ess.sup}_\omega \sup_{n_1 \geq n_1} \sup_{j=0}^{\min(j, n_0) - 1} P(B(n, k, j - 1)^c | \mathcal{F}_{s(n,k)}) < \epsilon_4.$$ 

Here the ordering $\alpha_0 > \cdots > \alpha_J$ was used.

Since

$$P\left[ B(n, k, j - 1) \cap A(\xi, \epsilon, n, k, j) | \mathcal{F}_{s(n,k,j-1)} \right] \geq P\left[ A(\xi, \epsilon, n, k, j) | \mathcal{F}_{s(n,k,j-1)} \right] - 1_B(n, k, j - 1)^c,$$

we have

$$P\left[ \bigcap_{j=0}^{J' - 1} \left( B(n, k, j - 1) \cap A(\xi, \epsilon, n, k, j) \right) | \mathcal{F}_{s(n,k)} \right] \geq q'P\left[ \bigcap_{j=0}^{J' - 1} \left( B(n, k, j - 1) \cap A(\xi, \epsilon, n, k, j) \right) | \mathcal{F}_{s(n,k)} \right] - \epsilon_4$$

for all $n \geq n_1$, $k$, $j$ and a.s. $\omega$. Let

$$C(\xi, \epsilon, n, k) = \bigcap_{j=0}^{J} \left( B(n, k, j - 1) \cap A(\xi, \epsilon, n, k, j) \right).$$

We use the above inequality repeatedly to obtain

$$\text{ess.inf}_\omega \inf_{n \geq n_1} P(\bigcap_{k=1}^{\min(j, n_0)} C(\xi, \epsilon, n, k)^c) \geq (q')^{J+1} - \epsilon_4 \sum_{j=0}^{J-1} (q')^j \geq q''$$

with some positive constant $q''$ if we take a sufficiently small $\epsilon_4$ for $q'$. Similarly by conditioning,

$$P\left[ \bigcap_{k=1}^{\min(n_0)} C(\xi, \epsilon, n, k)^c \right] = P\left[ \bigcap_{k=1}^{\min(n_0)} C(\xi, \epsilon, n, k)^c P[C(\xi, \epsilon, n, k)^c | \mathcal{F}_{s(n,k(n))}] \right] \leq (1-q'')P\left[ \bigcap_{k=1}^{\min(n_0)-1} C(\xi, \epsilon, n, k)^c \right] \leq \cdots \leq (1-q'')^{k(n)}.$$
for $n \geq n_1$.

We choose a sufficiently small $\epsilon_3$. Now it is easy to see that for large $n$, $X_t - X_{s(n,k)} \in S(\xi, \epsilon)$ and $|\xi \cdot (X_t - X_{s(n,k)})| = n^{-\alpha_j}$ for some $t = t(j) \in (s(n, k), s(n, k) + n^{-c_1}]$ for every $j$ on the event $C(\xi, \epsilon, n, k)$. Therefore,

$$
P \left[ \bigcup_{k=1}^{n} \bigcap_{j=0}^{J} \bigcup_{t \in [0, n^{-c_1}]} \left\{ X_{s(n,k)+t} - X_{s(n,k)} \in S(\xi, \epsilon) \right. \right.
$$

$$
\left. \left. \text{and} \ |\xi \cdot (X_{s(n,k)+t} - X_{s(n,k)})| = n^{-\alpha_j} \right\} \right] 
\geq P \left[ \bigcup_{k=1}^{n} C(\xi, \epsilon, n, k) \right] 
\geq 1 - C(1 - q^m)^k(n) 
$$

for $n \in \mathbb{N}$ and for some constant $C$. This inequality implies $[N_1](iii)$. \hfill \Box

6 Examples and simulation results

As an example, we consider the one-dimensional diffusion process

$$
dX_t = X_t dt + \exp\{\theta \sin^2 X_t\} dw_t, \quad t \in [0, 1], \quad X_0 = 0, 
$$

(10)

where $\theta \in (-\pi, \pi)$.

For the simulations, in order to get the maximum likelihood type estimator, we used the MATHEMATICA 6.0, concretely, ”FindMinimum” with an initial value. We examine the asymptotic behaviour of the estimators, which are the maximum likelihood type estimator $\hat{\theta}_n$ obtained by using ”FindMinimum” with the initial value $\theta_0 = 0.5$ the Bayes type estimator $\tilde{\theta}_n$ with respect to the uniform prior $\pi(\theta) = 1/(2\pi)$ and the maximum likelihood type estimator $\hat{\theta}_n^{(1)}$ obtained by using ”FindMinimum” with initial value being the Bayes type estimator $\tilde{\theta}_n$, through the simulations, which were done for each $h_n = 1/50, 1/250, 1/500$. For the true model (10) with $\theta^* = 1$, 10000 independent sample paths are generated by the Milstein scheme, and the means and the standard deviations of the estimators are computed and shown in Table 1 below.

In Table 1, even if $h_n = 1/50$, the Bayes estimator $\tilde{\theta}_n$ has good performance, but both the maximum likelihood type estimators $\hat{\theta}_n$ and $\hat{\theta}_n^{(1)}$ have biases. In case that $h_n = 1/250$, all three estimators are unbiased and they have good behaviors. In this example, it is better to use the Bayes estimator than the maximum likelihood type estimators.

Table 1: The mean and standard deviation (s.d.) of the three kinds of estimators for 10000 independent simulated sample paths with $\theta^* = 1$.

| $h_n$ | $\hat{\theta}_n$ with $\theta_0 = 0.5$ mean | s.d. | $\hat{\theta}_n^{(1)}$ mean | s.d. | $\hat{\theta}_n$ with $\theta_0 = \tilde{\theta}_n$ mean | s.d. |
|---|---|---|---|---|---|---|
| 1/50 | 0.89850 | 0.55160 | 0.96473 | 0.48914 | 0.89850 | 0.55160 |
| 1/250 | 0.97816 | 0.23723 | 0.99392 | 0.23112 | 0.97816 | 0.23723 |
| 1/500 | 0.99145 | 0.16041 | 0.99969 | 0.15874 | 0.99145 | 0.16041 |
As another example, we consider the one-dimensional diffusion process
\[
dX_t = X_t dt + \exp\{\sin \theta \sin X_t - \theta^2 \sin^2 X_t\} dw_t, \quad t \in [0, 1], \quad X_0 = 0,
\]
where \( \theta \in (-\pi, \pi) \).

For the true model (11) with \( \theta^* = 0 \), simulations were done in the same way as the previous example. The means and the standard deviations of the maximum likelihood type estimator \( \hat{\theta}_n \) with the initial value \( \theta_0 = 0.5 \), the Bayes type estimator \( \tilde{\theta}_n \) with respect to the uniform prior \( \pi(\theta) = 1/(2\pi) \) and the maximum likelihood type estimator \( \hat{\theta}_n^{(1)} \) with the initial value \( \theta_0 = \hat{\theta}_n \) are computed and shown in Table 2 below.

In Table 2, the maximum likelihood type estimator \( \hat{\theta}_n \) with \( \theta_0 = 0.5 \) has a bias in all cases, while the Bayes type estimator \( \tilde{\theta}_n \) and the maximum likelihood estimator \( \hat{\theta}_n^{(1)} \) with \( \theta_0 = \hat{\theta}_n \) have good behaviors in all cases. Furthermore, we see that the standard deviation of the Bayes estimator \( \tilde{\theta}_n \) is smaller than the one of \( \hat{\theta}_n^{(1)} \) in all cases.

Table 2: The mean and standard deviation (s.d.) of the estimators for 10000 independent simulated sample paths with \( \theta^* = 0 \).

| \( h_n \) | \( \theta_n \) with \( \theta_0 = 0.5 \) | \( \theta_n \) | \( \hat{\theta}_n^{(1)} \) with \( \theta_0 = \theta_n \) |
|---|---|---|---|
| | mean | s.d. | mean | s.d. |
| 1/50 | 0.07702 | 0.39184 | 0.00363 | 0.43532 |
| 1/250 | 0.05046 | 0.23677 | 0.00317 | 0.26283 |
| 1/500 | 0.04230 | 0.20471 | 0.00071 | 0.16614 |

7 A geometric criterion

Apart from analytic criteria by derivatives, we shall consider the following condition in the spirits of Lemma [3] and Remarks [2] and [3].

[A3'] \( \text{supp} \mathcal{L}\{X_t\} \) is compact, there exists a function \( f : U \times \Theta \rightarrow \mathbb{R} \) for some open neighborhood \( U \) of \( \text{supp} \mathcal{L}\{X_t\} \) and the following conditions are satisfied.

(i) For some \( \theta \in (0, \infty) \), \( Q(x, \theta, \theta^*)|\theta - \theta^*|^{-2} \geq |f(x, \theta)|^\theta \) for all \( (x, \theta) \in U \times (\Theta \setminus \{\theta^*\}) \).

(ii) For each \( x_0 \in U \), there exist a neighborhood \( V \) in \( U \) of \( x_0 \) and a covering \( \{\Theta_k\}_{k=1,\ldots,k} \) of \( \Theta \) such that for each \( k = 1, \ldots, K \), there exist \( \xi_0 \in S, J \in \mathbb{N} \), some positive numbers \( M, c, \epsilon_0, K_j \) such that \( j = 1, \ldots, J \) and some functions \( \Psi_j : P_{\xi_0}^{\perp} V \times \Theta_k \rightarrow \mathbb{R} \) such that

(a) the function \( P_{\xi_0}^{\perp} V \ni y \mapsto \Psi(y, \theta) \in \mathbb{R} \) is \( M \)-Lipschitz continuous for all \( \theta \in \Theta_k 

(b) for \( (x, \theta) \in V \times \Theta_k \),

\[
|f(x, \theta)| \geq c \prod_{j=1}^{J} (|\xi_0 \cdot x - \Psi_j(P_{\xi_0}^{\perp} x, \theta)| \wedge \epsilon_0)^{K_j}.
\]

In [A3'], \( k \) may depend on \( x_0 \). Note that \( \{x \in V : f(x, \theta) = 0\} \subset \bigcup_{j=1}^{J} \{x \in V : \xi_0 \cdot x = \Psi_j(P_{\xi_0}^{\perp} x, \theta)\} \) under [A3'](ii), that is, the graph of the functions \( \Psi_j \) covers locally the null set of \( f \).
Theorem 6. Suppose that Conditions [A1], [A2] and [A3'] are satisfied. Then the same results as Theorems 4 and 2 hold true.

Proof. We consider an open ball \( B(x_0, \varepsilon_{x_0}) \subset V \) for each \( x_0 \in U \) and the covering \( \{ B(x_0, \varepsilon_{x_0}/2) \}_{x_0 \in U} \) of \( \text{supp}\mathcal{L}\{X_0\} \). By compactness, we obtain a finite number of balls \( V \), and as a result, we have an open neighborhood \( \mathcal{X}_0 \) of \( \text{supp}\mathcal{L}\{X_0\} \) and we may assume that for some \( \epsilon' > 0 \), every \( B(x, \epsilon') \) \((x \in \mathcal{X}_0)\) can find a \( V \supset B(x, \epsilon') \) among them. Call these \( V \)'s \( \mathcal{X}_\ell \) \((\ell = 1, \ldots, \ell)\). Each \( \mathcal{X}_\ell \) has a partition \( \{\Theta_{\ell,k}\}_{k=1, \ldots, \kappa} \).

Let \( \alpha_0 > \alpha_1 > \cdots > \alpha_J > 0 \), and \( \epsilon > 0 \). We consider a sufficiently large \( L \) and sufficiently large \( n \)'s. Fix \( V = \mathcal{X}_\ell \) and a \( \Theta_{\ell,k} \), for which we have \( \xi_0 = \xi_{\ell,k} \) and \( \Psi_j \) depending on \( (\ell,k) \). For \( x^* \in \mathcal{X}_\ell \), let \( \zeta_i = x^* + n^{-\alpha_i}D(\xi_0, \epsilon) \) for \( i = 0, 1, \ldots, J \), and denote \( n^{-L} \)-neighborhood of \( \zeta_i \) by \( \zeta_i \). Moreover denote by \( G_j \) the graph of \( (x, \Psi_j(\cdot, \theta)) \) in \( S_{\ell \cdot}^\ast(\xi_0, \epsilon) \).

We claim that there is no \( G_j \) that intersects with two different \( \zeta_i \)'s. Indeed, if any \( G_j \) intersected with \( \zeta_{i_1} \) and \( \zeta_{i_2} \) for \( i_1 < i_2 \), then there are \( x_k \in x^* + n^{-\alpha_k}D(\xi_0, \epsilon) \) \((k = 1, 2)\) such that

\[
2n^{-L} + |\Psi_j(P_{\xi_0}^\perp x_{1, \theta}) - \Psi_j(P_{\xi_0}^\perp x_{2, \theta})| > 2 \frac{n^{-\alpha_{i_2}}}{\alpha_{i_2}}.
\]

On the other hand, by the \( M \)-Lipschitz continuity of \( \Psi_j(\cdot, \theta) \), we have

\[
|\Psi_j(P_{\xi_0}^\perp x_{1, \theta}) - \Psi_j(P_{\xi_0}^\perp x_{2, \theta})| \leq 2M \sqrt{\epsilon n^{-\alpha_{i_2}}},
\]

which contradicts to \( (12) \) if we make \( \epsilon \) sufficiently small, and proved the claim. Therefore, there is at least one \( \zeta_i \) that does not intersect with any \( G_j \). Thus

\[
\inf_{(x^*, \theta) \in \mathcal{X}_\ell \times \Theta_{\ell,k}} \max_{i=0, 1, \ldots, J} \min_{\xi \in D(\xi_{\ell,k}, \epsilon)} \left| \xi_0 \cdot (x^* + n^{-\alpha_i} \xi) - \Psi_j(P_{\xi_0}^\perp (x^* + n^{-\alpha_i} \xi), \theta) \right| \geq n^{-L}
\]

for large \( n \) for every \( \ell = 1, \ldots, \ell \) and \( k = 1, \ldots, \kappa_\ell \). Consequently, taking large \( L' \), we obtain

\[
\min_{\ell=1, \ldots, \ell} \inf_{k=1, \ldots, \kappa_\ell} \max_{(x^*, \theta) \in \mathcal{X}_\ell \times \Theta_{\ell,k}} \inf_{i=0, 1, \ldots, J} \xi \in D(\xi_{\ell,k}, \epsilon) \left| f(x^* + n^{-\alpha_i} \xi, \theta) \right| \geq cn^{-L'}
\]

for large \( n \). This is Lemma 3 with \( D(\xi_{\ell,k}, \epsilon) \) replaced by \( D(\xi_{\ell,k}, \epsilon) \). Due to Remarks 2 and 3 we can prove the theorem in the same way as Theorems 4 and 2.

8 Proof of Theorems 3, 4 and 5

For the limit of \( Z_n \) given in (6), we define

\[
Z(u) = \exp \left( \Gamma(\theta^*)^{1/2} \zeta[u] - \frac{1}{2} \Gamma(\theta^*)[u, u] \right).
\]

Then the standardized quasi Bayesian estimator \( \tilde{u}_n = \sqrt{n} (\hat{\theta}_n - \theta^*) \) is written by

\[
\tilde{u}_n = \left( \int_{\mathbb{R}_n} Z_n(u) \pi(\theta^* + (1/\sqrt{n})u)du \right)^{-1} \int_{\mathbb{R}_n} uZ_n(u) \pi(\theta^* + (1/\sqrt{n})u)du.
\]

As we will prove later, the limit of \( \tilde{u}_n \) should be

\[
\tilde{u} = \left( \int_{\mathbb{R}_p} Z(u)du \right)^{-1} \int_{\mathbb{R}_p} uZ(u)du = \Gamma(\theta^*)^{-1/2} \zeta.
\]
However, even existence of the integrals requires more rigorous treatment.

In order to prove Theorems\[4\] and \[3\], we will first prepare several lemmas. The results for the quasi maximum likelihood estimator are proved at the same time with common machinery.

**Lemma 4.** Let \(f \in C_1^{1,1}(\mathbb{R}^d \times \Theta; \mathbb{R})\). Assume \([H1]\). Then, for every \(q > 0\),

\[
\sup_{n \in \mathbb{N}} E \left[ \left( \sup_{\theta \in \Theta} \sqrt{n} \left| \frac{1}{n} \sum_{k=1}^{n} f(X_{t_{k-1}}, \theta) - \frac{1}{T} \int_0^T f(X_t, \theta) dt \right| \right]^q \right] < \infty.
\]

**Proof.** Let

\[
F_n(\theta) = \sum_{k=1}^{n} \int_{t_{k-1}}^{t_k} \{ f(X_{t_{k-1}}, \theta) - f(X_t, \theta) \} \, dt.
\]

An easy estimate together with \([H1]\)-(i) implies that for every \(q > p\),

\[
\sup_{\theta \in \Theta} \left\| \sqrt{n} F_n(\theta) \right\|_q^q \leq \frac{n^{3q/2}}{n^q} \sum_{k=1}^{n} \int_{t_{k-1}}^{t_k} \sup_{\theta \in \Theta} \left\| f(X_{t_{k-1}}, \theta) - f(X_t, \theta) \right\|_q^q \, dt < C.
\]

Therefore,

\[
\sup_{n \in \mathbb{N}} \sup_{\theta \in \Theta} \left\| \sqrt{n} F_n(\theta) \right\|_q < \infty.
\]

Moreover, in a similar way,

\[
\sup_{n \in \mathbb{N}} \sup_{\theta \in \Theta} \left\| \sqrt{n} \partial_{\theta} F_n(\theta) \right\|_q < \infty.
\]

It follows from the Sobolev inequality that

\[
E \left[ \sup_{\theta \in \Theta} \left\| \sqrt{n} F_n(\theta) \right\|_q^q \right] \leq E \left[ C \int_{\Theta} \left\{ \left\| \sqrt{n} F_n(\theta) \right\|_q^q + \left\| \sqrt{n} \partial_{\theta} F_n(\theta) \right\|_q^q \right\} \, d\theta \right] \leq C_\Theta \left\{ \sup_{\theta \in \Theta} E \left[ \left\| \sqrt{n} F_n(\theta) \right\|_q^q \right] + \sup_{\theta \in \Theta} E \left[ \left\| \sqrt{n} \partial_{\theta} F_n(\theta) \right\|_q^q \right] \right\},
\]

where \(q > p\). Thus, one has that for every \(q > 0\),

\[
\sup_{n \in \mathbb{N}} \left\| \sup_{\theta \in \Theta} \left\| \sqrt{n} F_n(\theta) \right\|_q \right\|_q < \infty.
\]

This completes the proof. \(\square\)

**Lemma 5.** Let \(f \in C_1^{0,1}(\mathbb{R}^d \times \Theta; \mathbb{R}^m \otimes \mathbb{R}^m)\). Assume \([H1]\). Then, for every \(q > 0\),

\[
\sup_{n \in \mathbb{N}} E \left[ \sup_{\theta \in \Theta} \left( \sqrt{n} \left| \sum_{k=1}^{n} f(X_{t_{k-1}}, \theta) \left\{ \left[ (\Delta_k Y)^{\otimes 2} \right] - \left[ (\sigma(X_{t_{k-1}}, \theta^*) \Delta_k w)^{\otimes 2} \right] \right\} \right) \right]^q \right] < \infty.
\]

**Proof.** Noting that

\[
\Delta_k Y = \int_{t_{k-1}}^{t_k} b_t \, dt + \int_{t_{k-1}}^{t_k} \{ \sigma(X_t, \theta^*) - \sigma(X_{t_{k-1}}, \theta^*) \} \, dw_t + \sigma(X_{t_{k-1}}, \theta^*) \Delta_k w,
\]

\[
\left\| \int_{t_{k-1}}^{t_k} b_t \, dt \right\|_q^q + \left\| \int_{t_{k-1}}^{t_k} \{ \sigma(X_t, \theta^*) - \sigma(X_{t_{k-1}}, \theta^*) \} \, dw_t \right\|_q^q \leq Cn^{-q},
\]

20
and \[ \left\| \sigma(X_{tk-1}, \theta^*) \Delta_k w \right\|_q^q \leq Cn^{-q/2} \] for every \( q > p \lor 2 \), one has that

\[
\sup_{n \in \mathbb{N}} \sup_{\theta \in \Theta} \left\| \sqrt{n} \sum_{k=1}^n f(X_{tk-1}, \theta) \left\{ \left[ (\Delta_k Y)^{\odot 2} \right] - \left[ (\sigma(X_{tk-1}, \theta^*) \Delta_k w)^{\odot 2} \right] \right\} \right\|_q < \infty,
\]

with \( ab - cd = (a - c)b + (b - d)c \). Similarly,

\[
\sup_{n \in \mathbb{N}} \sup_{\theta \in \Theta} \left\| \sqrt{n} \sum_{k=1}^n \partial \sigma f(X_{tk-1}, \theta) \left\{ \left[ (\Delta_k Y)^{\odot 2} \right] - \left[ (\sigma(X_{tk-1}, \theta^*) \Delta_k w)^{\odot 2} \right] \right\} \right\|_q < \infty.
\]

By using the Sobolev inequality, we obtain the desired inequality.

Note that for \( u \in \mathbb{U}_n \),

\[
\mathbb{Z}_n(u) = \exp \left( \Delta_n[u] - \frac{1}{2} \Gamma(\theta^*)[u, u] + r_n(u) \right),
\]

where

\[
\Delta_n[u] = \frac{1}{\sqrt{n}} \partial \sigma \mathbb{H}_n(\theta^*)[u]
\]

\[
= -\frac{1}{2\sqrt{n}} \sum_{k=1}^n \left\{ (\partial \sigma \log \det S(X_{tk-1}, \theta^*))[u] + h^{-1}(\partial \sigma S^{-1})(X_{tk-1}, \theta^*)[u, (\Delta_k Y)^{\odot 2}] \right\},
\]

\[
\Gamma_n(\theta)[u, u] = -\frac{1}{n} \partial^2 \mathbb{H}_n(\theta)[u, u]
\]

\[
= \frac{1}{2n} \sum_{k=1}^n \left\{ (\partial^2 \sigma \log \det S(X_{tk-1}, \theta))[u^{\odot 2}] + h^{-1}(\partial^2 \sigma S^{-1})(X_{tk-1}, \theta)[u^{\odot 2}, (\Delta_k Y)^{\odot 2}] \right\},
\]

\[
\Gamma(\theta^*)[u, u] = \frac{1}{2T} \int_0^T \text{Tr} \left( (\partial \sigma S)S^{-1}(\partial \sigma S)S^{-1}(X_t, \theta^*)[u^{\odot 2}] \right) dt,
\]

\[
r_n(u) = \int_0^1 (1 - s) \left\{ \Gamma(\theta^*)[u, u] - \Gamma_n(\theta^* + s(1/\sqrt{n})u)[u, u] \right\} ds.
\]

**Lemma 6.** Assume \([H1]\). Then, for every \( q > 0 \),

(i) \( \sup_{n \in \mathbb{N}} E \left[ \left\| \Delta_n \right\|^q \right] < \infty. \)

(ii) \( \sup_{n \in \mathbb{N}} E \left[ \left( \sup_{\theta \in \Theta} \sqrt{n} \left| \mathbb{Y}_n(\theta) - \mathbb{Y}(\theta) \right| \right)^q \right] < \infty. \)

**Proof.** (i) Since \( \Delta_n = M_n + R_n \), where

\[
M_n = -\frac{1}{2\sqrt{n}} \sum_{k=1}^n \left\{ (\partial \sigma \log \det S(X_{tk-1}, \theta^*)) + h^{-1}(\partial \sigma S^{-1})(X_{tk-1}, \theta^*) \right\}[\sigma(X_{tk-1}, \theta^*) \Delta_k w]^\odot 2, \]

\[
R_n = -\frac{\sqrt{n}}{2T} \sum_{k=1}^n (\partial \sigma S^{-1}(X_{tk-1}, \theta^*))[(\Delta_k Y)^\odot 2 - \sigma(X_{tk-1}, \theta^*) \Delta_k w]^\odot 2, \]


Lemma 5 yields that for every $q > 1$, $\sup_{n \in \mathbb{N}} ||R_n||_q < \infty$. Moreover, $\sqrt{n}M_n$ is the terminal value of a discrete-time martingale with respect to $(\mathcal{F}_{t_k})_{k=0,1,\ldots,n}$ and it follows from the Burkholder inequality that $\sup_{n \in \mathbb{N}} ||M_n||_q < \infty$. Thus, one has that $\sup_{n \in \mathbb{N}} ||\Delta_n||_q < \infty$ for every $q > 1$.

(ii) Note that

$$\mathcal{Y}_n(\theta) = \mathcal{Y}_n^\uparrow(\theta) + M_n^\uparrow(\theta) + R_n^\uparrow(\theta),$$

where

$$\mathcal{Y}_n^\uparrow(\theta) = -\frac{1}{2n} \sum_{k=1}^{n} \left\{ \log \left( \frac{\det S(X_{t_{k-1}}, \theta)}{\det S(X_{t_{k-1}}, \theta^*)} \right) + \text{Tr} \left( S^{-1}(X_{t_{k-1}}, \theta) S(X_{t_{k-1}}, \theta^*) - I_d \right) \right\},$$

$$M_n^\uparrow(\theta) = -\frac{1}{2n} \sum_{k=1}^{n} (S^{-1}(X_{t_{k-1}}, \theta) - S^{-1}(X_{t_{k-1}}, \theta^*)) |h^{-1}(\sigma(X_{t_{k-1}}, \theta^*) \Delta_k w)\|^2 - S(X_{t_{k-1}}, \theta^*),$$

$$R_n^\uparrow(\theta) = -\frac{1}{2n} \sum_{k=1}^{n} (S^{-1}(X_{t_{k-1}}, \theta) - S^{-1}(X_{t_{k-1}}, \theta^*)) |h^{-1}(\Delta_k Y)\|^2 - h^{-1}(\sigma(X_{t_{k-1}}, \theta^*) \Delta_k w)\|^2].$$

By Lemma 5 for every $q > 1$,

$$\sup_{n \in \mathbb{N}} ||\sup_{\theta \in \Theta} |\sqrt{n}R_n^\uparrow(\theta)||_q < \infty.$$ 

Burkholder’s inequality yields that

$$\sup_{n \in \mathbb{N}} ||\sqrt{n}M_n^\uparrow(\theta)||_q < \infty \quad \text{and} \quad \sup_{n \in \mathbb{N}} ||\sqrt{n}M_n^\uparrow(\theta)||_q < \infty$$

for all $q > 1$. Hence, Sobolev’s inequality implies that

$$\sup_{n \in \mathbb{N}} ||\sup_{\theta \in \Theta} |\sqrt{n}M_n^\uparrow(\theta)||_q < \infty$$

for every $q > p$. Furthermore, it follows from Lemma 4 that

$$\sup_{n \in \mathbb{N}} ||\sup_{\theta \in \Theta} |\sqrt{n}\mathcal{Y}_n(\theta) - \mathcal{Y}(\theta)||_q < \infty$$

for every $q > 1$. This completes the proof. \qed

**Lemma 7.** Assume [H1]. Then, for every $q > 0$,

(i) $\sup_{n \in \mathbb{N}} E \left[ \left( \sqrt{n} \left| \Gamma_n(\theta^*) - \Gamma(\theta^*) \right| \right)^q \right] < \infty$.

(ii) $\sup_{n \in \mathbb{N}} E \left[ \left( \frac{1}{n} \sup_{\theta \in \Theta} \left| \partial_\theta \mathbb{H}_n(\theta) \right| \right)^q \right] < \infty$.

**Proof.** (i) Note that

$$\Gamma_n(\theta^*)[u, u] = \tilde{\Gamma}_n[u, u] + \tilde{M}_n[u, u] + \tilde{R}_n[u, u],$$

where

$$\tilde{\Gamma}_n[u, u] = \frac{1}{2n} \sum_{k=1}^{n} \left\{ \partial_\theta^2 \log \det S(X_{t_{k-1}}, \theta^*)[u \otimes^2] + \partial_\theta^2 S^{-1}(X_{t_{k-1}}, \theta^*)[u \otimes^2, S(X_{t_{k-1}}, \theta^*)] \right\},$$

$$\tilde{M}_n[u, u] = \frac{1}{2n} \sum_{k=1}^{n} \partial_\theta^2 S^{-1}(X_{t_{k-1}}, \theta^*)[u \otimes^2, h^{-1}(\sigma(X_{t_{k-1}}, \theta^*) \Delta_k w)\otimes^2 - S(X_{t_{k-1}}, \theta^*)],$$

$$\tilde{R}_n[u, u] = \frac{1}{2n} \sum_{k=1}^{n} \partial_\theta^2 S^{-1}(X_{t_{k-1}}, \theta^*)[u \otimes^2, h^{-1}(\Delta_k Y)\otimes^2 - h^{-1}(\sigma(X_{t_{k-1}}, \theta^*) \Delta_k w)\otimes^2].$$
By Lemmas 4 and 5, for every $q > 1$, 
\[
\sup_{n \in \mathbb{N}} \|\sqrt{n}(\hat{\Gamma}_n - \Gamma(\theta^*))\|_q < \infty, \quad \text{and} \quad \sup_{n \in \mathbb{N}} \|\sqrt{n}\hat{R}_n\|_q < \infty.
\]
Using Burkholder’s inequality and Sobolev’s inequality, one has
\[
\sup_{n \in \mathbb{N}} \|\sqrt{n}\hat{M}_n\|_q < \infty
\]
for $q > p$, and the desired inequality is obtained.

(ii) Since
\[
\frac{1}{n} \partial^{j}_{\theta} \mathbb{H}_n(\theta) = -\frac{1}{2n} \sum_{k=1}^{n} \left\{ \partial^{j}_{\theta} \log \det S(X_{t_{k-1}}, \theta) + h^{-1} \partial^{j}_{\theta} S^{-1}(X_{t_{k-1}}, \theta)(\Delta_k Y)^{\otimes 2} \right\}
\]
for $j=3,4$, it is easy to show that
\[
\sup_{n \in \mathbb{N}} \|\sup_{\theta \in \Theta} n^{-1} \partial^{3}_{\theta} \mathbb{H}_n(\theta)\|_q < \infty
\]
for $q > p$, which completes the proof. \hfill \Box

**Proof of Theorem 3.** It is enough to check the regularity conditions $[A_{1''}], [A_{2}], [A_{3}], [A_{4'}], [A_{5}]$ and $[A_{6}]$ in Theorem 2 of Yoshida (2005). It follows from $[H2]$ that $[A_{2}], [A_{3}]$ with $\rho = 2$ and $[A_{5}]$ are satisfied for every $L > 0$. We can take appropriate parameters satisfying $[A_{4'}]$, and Lemma 7 implies $[A_{1''}]$ for every $L > 0$. Lemma 6 yields $[A_{6}]$ for every $L > 0$. This completes the proof. \hfill \Box

We denote $B(R) = \{ u \in \mathbb{R}^p : |u| \leq R \}$.

**Lemma 8.** Assume $[H1]$. Then, for every $q > p$ and $R > 0$, there exists $C_0 > 0$ such that
\[
\sup_{n \in \mathbb{N}} E(||\log Z_n(u)||^q) \leq C_0 |u|^q
\]
for all $u \in B(R)$.

**Proof.** $\log Z_n(u)$ has a decomposition
\[
\log Z_n(u) = \Delta_n[u] - \frac{1}{2} \Gamma_n(\theta^*)[u, u] + \tilde{r}_n(u),
\]
where
\[
\tilde{r}_n(u) = \frac{1}{2n^{3/2}} \int_{0}^{1} (1-s)^2 \partial^{3}_{\theta} \mathbb{H}_n(\theta^* + s(1/\sqrt{n})u)[u^{\otimes 3}]ds.
\]
By Lemma 6, $\sup_{n \in \mathbb{N}} ||\Delta_n[u]||^q_2 < C_1 |u|^q$. Lemma 7 yields that $\sup_{n \in \mathbb{N}} ||\tilde{r}_n(u)||^q_2 < C_2 |u|^{3q}$. Moreover, $\sup_{n \in \mathbb{N}} ||\Gamma_n(\theta^*)[u, u]||^q_2 < C_3 |u|^{2q}$. Thus, noting that $|u|^{2q} + |u|^{3q} \leq C_4 |u|^q$ for all $u \in B(R)$, we obtain the desired inequality. \hfill \Box

**Remark 4.** By using Lemma 2 in Yoshida (2005), Lemma 8 implies that for every $R > 0$,
\[
\sup_{n \in \mathbb{N}} E \left[ \left( \int_{|u| \leq R} Z_n(u)du \right)^{-1} \right] < \infty.
\]
Consequently, one has that
\[
\sup_{n \in \mathbb{N}} E \left[ \left( \int_{U_n} Z_n(u) du \right)^{-1} \right] < \infty.
\]

**Lemma 9.** Assume \([H1^2]\). Then, for every \(R > 0\), \(Z_n(u) \to_{d,(\mathcal{F})} Z(u)\) in \(C(B(R))\) as \(n \to \infty\).

**Proof.** Let \(\sigma_t^* = \sigma(X_t, \theta^*)\), \(\Delta_k \tilde{X} = \sigma_{k-1}^* \Delta_k w\) and
\[
H(y, u) = -\frac{1}{2\sqrt{n}} \left\{ (\partial_y \log \det S(X_{t_{k-1}}, \theta^*))[u] + h^{-1}(\partial_y S^{-1}(X_{t_{k-1}}, \theta^*)[u, y^{\otimes 2}]) \right\} \\
- \frac{1}{4n} \left\{ (\partial_y^2 \log \det S(X_{t_{k-1}}, \theta^*))[u^{\otimes 2}] + h^{-1}(\partial_y^2 S^{-1}(X_{t_{k-1}}, \theta^*)[u^{\otimes 2}, y^{\otimes 2}] \right\}
\]
for \(y \in \mathbb{R}^d\). Note that
\[
\log Z_n(u) = \sum_{k=1}^n \left\{ H(\Delta_k Y, u) - H(\Delta_k \tilde{X}, u) \right\} + \sum_{k=1}^n H(\Delta_k \tilde{X}, u) + \tilde{r}_n(u)
\]
\[
= \sum_{k=1}^n \{ J_{1,k}(u) + J_{2,k}(u) \} + \sum_{k=1}^n H(\Delta_k \tilde{X}, u) + \tilde{r}_n(u),
\]
where
\[
J_{1,k}(u) = -\frac{\sqrt{n}}{2T} (\partial_y S^{-1})(X_{t_{k-1}}, \theta^*)[u, (\Delta_k Y)^{\otimes 2} - (\Delta_k \tilde{X})^{\otimes 2}],
\]
\[
J_{2,k}(u) = -\frac{1}{4T} (\partial_y^2 S^{-1})(X_{t_{k-1}}, \theta^*)[u^{\otimes 2}, (\Delta_k Y)^{\otimes 2} - (\Delta_k \tilde{X})^{\otimes 2}].
\]

Since it follows from Lemma 7 that \(\|\sqrt{n}\tilde{r}_n(u)\|_2^2 < \infty\), one has that \(\tilde{r}_n(u) \to^p 0\) as \(n \to \infty\). Lemma 5 yields that \(\sum_{k=1}^n J_{2,k}(u) \to^p 0\) as \(n \to \infty\).

An easy calculation yields that
\[
(\Delta_k Y)^* - (\Delta_k \tilde{X})(\Delta_k \tilde{X})^* \\
= \left( \int_{t_{k-1}}^{t_k} b_1 dt + \int_{t_{k-1}}^{t_k} (\sigma_t^* - \sigma_{t_{k-1}}^*) dw_t \right) \left( \int_{t_{k-1}}^{t_k} b_1 dt + \int_{t_{k-1}}^{t_k} (\sigma_t^* - \sigma_{t_{k-1}}^*) dw_t \right)^* \\
+ \left( \int_{t_{k-1}}^{t_k} \sigma_{t_{k-1}}^* dw_t \right) \left( \int_{t_{k-1}}^{t_k} b_1 dt + \int_{t_{k-1}}^{t_k} (\sigma_t^* - \sigma_{t_{k-1}}^*) dw_t \right)^* \\
+ \left( \int_{t_{k-1}}^{t_k} b_1 dt \right) \left( \int_{t_{k-1}}^{t_k} \sigma_{t_{k-1}}^* dw_t \right) \left( \int_{t_{k-1}}^{t_k} \sigma_{t_{k-1}}^* dw_t \right)^*.
\]

Let \(f_t = (\partial_y S^{-1})(X_{t_{k-1}}, \theta^*)[u]\) and \(f_t = f_{t_{k-1}} \) if \(t \in [t_{k-1}, t_k)\). Furthermore, setting
\[
H_{1,k} = -\frac{\sqrt{n}}{2T} f_{t_{k-1}} \left[ \left( \int_{t_{k-1}}^{t_k} b_1 dt + \int_{t_{k-1}}^{t_k} (\sigma_t^* - \sigma_{t_{k-1}}^*) dw_t \right)^{\otimes 2} \right],
\]
\[
H_{2,k} = -\frac{\sqrt{n}}{2T} f_{t_{k-1}} \left[ \left( \int_{t_{k-1}}^{t_k} b_1 dt \right) \left( \int_{t_{k-1}}^{t_k} \sigma_{t_{k-1}}^* dw_t \right)^* \right],
\]
\[
H_{3,k} = -\frac{\sqrt{n}}{2T} f_{t_{k-1}} \left[ \left( \int_{t_{k-1}}^{t_k} \sigma_{t_{k-1}}^* dw_t \right) \left( \int_{t_{k-1}}^{t_k} (\sigma_t^* - \sigma_{t_{k-1}}^*) dw_t \right)^* \right],
\]
24
one has that $J_{1,k}(u) = H_{1,k} + 2H_{2,k} + 2H_{3,k}$. Since
\[
\left\| \sum_{k=1}^{n} H_{1,k} \right\|_1 \leq C \sqrt{n} \sum_{k=1}^{n} \left\| \int_{t_{k-1}}^{t_k} b_t dt + \int_{t_{k-1}}^{t_k} (\sigma_t^* - \sigma_{t_{k-1}}^*) dw_t \right\|_2 \leq C_1 h,
\]
we obtain $\sum_{k=1}^{n} H_{1,k} \to^p 0$. In the decomposition
\[
\sum_{k=1}^{n} H_{2,k} = -\frac{\sqrt{n}}{2T} \sum_{k=1}^{n} f_{t_{k-1}} \left[ \int_{t_{k-1}}^{t_k} (b_t - b_{t_{k-1}}) dt \left( \int_{t_{k-1}}^{t_k} \sigma_{t_{k-1}}^* dw_t \right)^* \right] \\
- \frac{\sqrt{n}}{2T} \sum_{k=1}^{n} f_{t_{k-1}} \left[ b_{t_{k-1}} \left( \int_{t_{k-1}}^{t_k} \sigma_{t_{k-1}}^* dw_t \right)^* \right],
\]
the first term on the right-hand side is $o_p(1)$ by standard estimates, and the second term is $O(n^{-1/2})$ in $L^2$-norm. Thus $\sum_{k=1}^{n} H_{2,k} \to^p 0$.

In order to obtain that $\sum_{k=1}^{n} H_{3,k} \to^p 0$ as $n \to \infty$, it is enough to show that as $n \to \infty$,
\[
\sum_{k=1}^{n} E[\sum_{i=1}^{n} \sigma_{i,t_{k-1}}^* dw_t | F_{t_{k-1}}] \to^p 0, \quad \sum_{k=1}^{n} E[(H_{3,k})^2 | F_{t_{k-1}}] \to^p 0
\]
because of Lemma 9 in Genon-Catalot and Jacod (1993). Since
\[
E[(H_{3,k})^2 | F_{t_{k-1}}] \leq C_1 n E \left[ \left( \int_{t_{k-1}}^{t_k} \sigma_{i,t_{k-1}}^* dw_t \right)^4 | F_{t_{k-1}} \right]^{1/2} E \left[ \left( \int_{t_{k-1}}^{t_k} (\sigma_t^* - \sigma_{t_{k-1}}^*) dw_t \right)^4 | F_{t_{k-1}} \right]^{1/2} \leq C_2 h^2,
\]
one has $\sum_{k=1}^{n} E[(H_{3,k})^2 | F_{t_{k-1}}] \to^p 0$. Since Itô’s formula implies that
\[
\sigma_t^* - \sigma_{t_{k-1}}^* = \int_{t_{k-1}}^{t} \left( \sum_{i=1}^{d} \partial_{x_i} \sigma(X_s, \theta^*) \tilde{b}_s^i + \sum_{i,j=1}^{d} \frac{1}{2} \partial_{x_i} \partial_{x_j} \sigma(X_s, \theta^*)((a_s a^{*})^{ij} + (\tilde{a}_s \tilde{a}^{*})^{ij}) \right) ds \\
+ \int_{t_{k-1}}^{t} \sum_{i=1}^{d} \sum_{j=1}^{r} \partial_{x_i} \sigma(X_s, \theta^*) a^{ij}_s dw_s + \int_{t_{k-1}}^{t} \sum_{i=1}^{d} \sum_{j=1}^{r} \partial_{x_i} \sigma(X_s, \theta^*) \tilde{a}^{ij}_s d\tilde{w}_s,
\]
one has
\[
E[\sum_{k=1}^{n} \sigma_{i,t_{k-1}}^* dw_t | F_{t_{k-1}}] \leq C_1 \sqrt{n} E \left[ \left( \int_{t_{k-1}}^{t_k} (\sigma_t^* - \sigma_{t_{k-1}}^*) dw_t \right) \right] | F_{t_{k-1}} | \leq C_2 h^{3/2}
\]
and $\sum_{k=1}^{n} E[\sum_{i=1}^{n} \sigma_{i,t_{k-1}}^* dw_t | F_{t_{k-1}}] \to^p 0$ as $n \to \infty$. Therefore $\sum_{k=1}^{n} H_{3,k} \to^p 0$, and hence $\sum_{k=1}^{n} J_{1,k}(u) \to^p 0$.

Let $\xi_k := \xi_k(u) = H(\Delta_k \hat{X}, u)$ and
\[
\Gamma(t^*) |u, u| = \frac{1}{2T} \int_0^{T} \text{Tr} \left( (\partial_d S) S^{-1}(\partial_d S) S^{-1}(X_s, \theta^*) [u^{\otimes 2}] \right) ds.
\]

---

For a progressively measurable process $b_t$ satisfying $E[\int_0^{T} |b_t|^2 dt] < \infty$, for any $\epsilon > 0$, one can find a right-continuous adapted process $(\beta_t)$ such that $E[\int_0^{T} |b_t - \beta_t|^2 dt] < \epsilon$. Then we may assume the right-continuity of $(b_t)$.
It follows from Theorem 3-2 of Jacod (1997) that
\[ \log Z_n(u) \to \Gamma(\theta^*)^{1/2} \xi[u] - \frac{1}{2} \Gamma(\theta^*)[u, u] \] is \( F \)-stable \quad (14)
if we show that
\[
\sup_{t \in [0,T]} \left| \sum_{k=1}^{\left\lfloor nt/T \right\rfloor} E \left[ \xi_k \mid F_{tk-1} \right] - \frac{1}{2} \Gamma_t(\theta^*)[u, u] \right| \to^p 0, \tag{15}
\]
\[
\sum_{k=1}^{\left\lfloor nt/T \right\rfloor} \left\{ E \left[ \xi_k^2 \mid F_{tk-1} \right] - (E \left[ \xi_k \mid G_{tk-1} \right])^2 \right\} \to^p \Gamma_t(\theta^*)[u, u], \quad \text{for all } t \in [0, T], \tag{16}
\]
\[
\sum_{k=1}^{\left\lfloor nt/T \right\rfloor} E \left[ \xi_k \Delta_k w \mid F_{tk-1} \right] \to^p 0, \quad \text{for all } t \in [0, T], \tag{17}
\]
\[
\sum_{k=1}^{\left\lfloor nt/T \right\rfloor} E \left[ \xi_k \Delta_k N \mid F_{tk-1} \right] \to^p 0, \quad \text{for all } t \in [0, T] \text{ and } N \in \mathcal{M}_b(w^\perp), \tag{18}
\]
\[
\sum_{k=1}^{\left\lfloor nt/T \right\rfloor} E \left[ \xi_k^2 \mid F_{tk-1} \right] \to^p 0, \tag{19}
\]
where \( \mathcal{M}_b(w^\perp) \) is the class of all bounded \( F \)-martingales which is orthogonal to \( w \).

The symbol \( R(r_n) \) denotes a sequence of random variables for which \( \| R(r_n) \|_q \leq C_q r_n \) for every \( q > 1 \), where \( C_q \) depend on neither \( n \) nor other variables. Indeed, one has
\[
E \left[ \xi_k \mid F_{tk-1} \right] = -\frac{1}{2\sqrt{n}} \left\{ (\partial_\theta \log \det S(X_{tk-1}, \theta^*))[u] + (\partial_\theta S^{-1})(X_{tk-1}, \theta^*)[u, S(X_{tk-1}, \theta^*)] \right\}
- \frac{1}{4n} \left\{ (\partial_\theta^2 \log \det S(X_{tk-1}, \theta^*))[u^{\otimes 2}] + (\partial_\theta^2 S^{-1})(X_{tk-1}, \theta^*)[u^{\otimes 2}, S(X_{tk-1}, \theta^*)] \right\},
\]
\[
E \left[ \xi_k^2 \mid F_{tk-1} \right] = \frac{1}{4n} \left\{ (\partial_\theta \log \det S(X_{tk-1}, \theta^*))[u]^2 \right\}
+ 2(\partial_\theta \log \det S(X_{tk-1}, \theta^*))[u] (\partial_\theta S^{-1})(X_{tk-1}, \theta^*)[u, S(X_{tk-1}, \theta^*)]
+ ((\partial_\theta S^{-1})(X_{tk-1}, \theta^*)[u, S(X_{tk-1}, \theta^*)])^2 + 2\text{Tr} \left( (\partial_\theta S^{-1})(\partial_\theta S)S^{-1}(X_{tk-1}, \theta^*)[u^{\otimes 2}] \right) + R(t^{3/2})
\]
\[
\frac{1}{2n} \text{Tr} \left( (\partial_\theta S^{-1})(\partial_\theta S)S^{-1}(X_{tk-1}, \theta^*)[u^{\otimes 2}] \right) + R(t^{3/2}).
\]
\[
E \left[ \xi_k \Delta_k w \mid F_{tk-1} \right] = -\frac{\sqrt{n}}{2T} E \left[ (\partial_\theta S^{-1})(X_{tk-1}, \theta^*)[u, (\sigma(X_{tk-1}, \theta^*)\Delta_k w)^{\otimes 2}] \right] \Delta_k w \mid F_{tk-1}\]
- \frac{1}{4T} E \left[ (\partial_\theta^2 S^{-1})(X_{tk-1}, \theta^*)[u^{\otimes 2}, (\sigma(X_{tk-1}, \theta^*)\Delta_k w)^{\otimes 2}] \right] \Delta_k w \mid F_{tk-1}
= 0,
\]
where in the last estimate, we note that $E[(\Delta_kw)^2] = E[h(\Delta_kw)^2 - h]$. Therefore, (15) and (18) are shown. Furthermore, $E[\xi_k^2 | \mathcal{F}_{t_{k-1}}] = R(h^2)$, which proves (14). Moreover, it is easy to see that the joint convergence for finitely many $u$'s is valid in (14).

In order to show the tightness of the family $\{\log Z_n(u)_{B(R)} : n \in \mathbb{N}\}$ for every $R > 0$, it is enough to prove that $\sup_{n \in \mathbb{N}} E \left[ \left( \sup_{u \in B(R)} |\partial_u \log Z_n(u)| \right)^q \right] < \infty$ for $q > 1$. For details, see Yoshida (1990). In the same way as in the proof of Lemma 6 using Lemma 5 one has that $\sup_{n \in \mathbb{N}} \sup_{u \in B(R)} |\partial_u \log Z_n(u)|_q < \infty$ and $\sup_{n \in \mathbb{N}} \sup_{u \in B(R)} E \left[ |\partial_u \log Z_n(u)|_q \right] < \infty$. This completes the proof.

Proof of Theorem 4 Note that Theorem 4 implies Proposition 2 in Yoshida (2005). Using Lemma 9 together with this fact, one can apply Theorem 5 in Yoshida (2005), which completes the proof.

Proof of Theorem 5 By Theorem 4 and Lemma 9 the regularity conditions of Theorem 6 in Yoshida (2005) are satisfied and one has that

$$\int f(u)Z_n(u)du \rightarrow_d \int f(u)Z(u)du$$

as $n \rightarrow \infty$, for all continuous functions $f$ of at most polynomial growth. It follows from Lemma 2 in Yoshida (2005) that Lemma 8 meets the condition (i) of Theorem 8 in Yoshida (2005). Note that the condition (ii) of Theorem 8 in Yoshida (2005) is satisfied in our setting. This completes the proof.

9 Some remarks on the analytic criteria

In this paper, we derived Theorems 1 and 2 thought $[N_0]$ by checking $[N_1]$. Because of $[N_1]$, we assumed existence of supporting functions, this condition obviously works for the one-dimensional $X$, and also does in a multi-dimensional case if the null set of $f$ is locally a regular submanifold. However, this condition is not completely general.

The supporting function supports $f$ in a neighborhood of a point $(x_0, \theta) \in X \times \Theta$. However, what was necessary in our logic was the existence of a function supporting $f$ in a particular sector depending on the location of $X_s$ of a good increment $X_{t_j} - X_s$, as was seen in the proof of Proposition 1. As a matter of fact, the sectors can be chosen discontinuously though it was done continuously under $[N_1]$. It is because Condition $[N_1]$ is a topological condition and this nature was used in Lemma 2.

On the other hand, Condition $[N_0]$ (or $[N_0']$) does not require continuity of the supporting function $G_{t,k}(x_0, x, \theta, \xi)$ in $(x_0, \theta)$ associated with a sector. Besides, the double sided sector is not necessary for nondegenerate diffusions; of course, there are cases in which it really works effectively. Therefore, Theorems 4 and 5 together with $[N_0']$ are much stronger than Theorems 1 and 2. Though we do not go into details here, a more general condition will be as follows: There exist $t_0 > 0$, $a \in (0, 1)$, $\epsilon > 0$ and a finite subset $\mathcal{E}$ of $S$ such that for any $(x_0, \theta)$, for some $\xi_0 \in \mathcal{E}$, for all $t \in (0, t_0)$,

$$\inf_{\xi \in \mathcal{B}(\xi_0, \epsilon)} |f(x_0 + t\xi, \theta)| > a|f(x_0 + t\xi_0, \theta)|$$

27
and max_{j=0,...,J-1} |c_j(x_0,\theta)| > 0 for each (x_0,\theta), where c_j(x_0,\theta) are given by the derivatives of f and \xi_0.

If the null set of f includes irregular points and if X_\tau hits them, then the criteria like [A3] do not work in general. However, even if the process X starts bad points, if it moves quickly to a good area of regular points, it is possible to apply the idea of our criteria by some modification. As before, let \{\mathcal{X}_\ell\}_{\ell=1,...,J} cover \mathcal{X}_0 and \{\Theta_{\ell,k}\}_{k=1,...,\bar{\ell}} cover \Theta for each \ell. We assume that for each (x_0,\theta) \in \mathcal{X}_\ell \times \Theta_{\ell,k}, there are a function g and \xi_{\ell,k} \in \mathcal{S} such that

(i) \partial_x^j g exist and they are continuous for j = 0,...,J, and max_{j=0,...,J-1} |c_j(x_0,\theta)| > 0 for each (x_0,\theta), where c_j(x_0,\theta) are given by the derivatives of g and \xi_0.

(ii) For every (x_0,\theta) \in (\mathcal{X}_\ell \cap U_n) \times \Theta_{\ell,k}, |f(x,\theta)| \geq |g(P_{\ell,k} x,\theta)| for all (x,\theta) \in B(x_0,n^{-\beta_0}) \times \Theta_{\ell,k}.

Moreover suppose that there is a sequence of stopping times \tau_n such that (1 - P[\tau_n \leq T_0, X_{\tau_n} \in U_n])_{n \in \mathbb{N}} \in \mathcal{P} for some T_0 \in [0,T).

If x_\star is a singular point, we can take U_n = \mathbb{R}^d \setminus B(x_\star, n^{-\beta_1}) for \beta_1 \in (0, \beta_0). Then it is possible to prove [I2] for a nondegenerate diffusion process X, by composing the arguments in the previous sections with \alpha_0 > \alpha_1 > \cdots > \alpha_J > \beta_0. A simple example is f(x_1,x_2,\theta) = x_1(x_1^2 - \theta x_2^4) for x = (x_1,x_2), x_\star = (0,0), and \text{supp}\mathcal{L}\{X_0\} = \{0\} \times [0,1].

References

Adams, R. A. (1975). Sobolev spaces. Pure and Applied Mathematics, Vol. 65. Academic Press, New York-London.

Adams, R. A. and Fournier, J. J. F. (2003). Sobolev spaces. Second edition. Pure and Applied Mathematics (Amsterdam), 140. Elsevier/Academic Press, Amsterdam.

Dohmal, G. (1987). On estimating the diffusion coefficient. J. Appl. Probab. 24, 105–114.

Genon-Catalot, V. and Jacod, J. (1993). On the estimation of the diffusion coefficient for multidimensional diffusion processes. Ann. Inst. Henri Poincaré Probab. Statist. 29, 119–151.

Genon-Catalot, V. and Jacod, J. (1994). Estimation of the diffusion coefficient for diffusion processes: random sampling. Scand. J. Statist. 21 (1994), 193–221.

Gobet, E. (2001). Local asymptotic mixed normality property for elliptic diffusion: a Malliavin calculus approach. Bernoulli, 7, 899–912.

Ibragimov, I. A. and Has’minskii, R. Z. (1972). The asymptotic behavior of certain statistical estimates in the smooth case. I. Investigation of the likelihood ratio. (Russian) Teor. Verojatnost. i Primenen. 17, 469–486.

Ibragimov, I. A. and Has’minskii, R. Z. (1973). Asymptotic behavior of certain statistical estimates. II. Limit theorems for a posteriori density and for Bayesian estimates. (Russian) Teor. Verojatnost. i Primenen. 18, 78–93.

Ibragimov, I. A. and Has’minskii, R. Z. (1981). Statistical estimation. Springer Verlag, New York.

Jacod, J. (1997). On continuous conditional Gaussian martingales and stable convergence in law. Seminaire de Probabilites, XXXI, 232–246, Lecture Notes in Math., 1655, Springer, Berlin.
Kessler, M. (1997). Estimation of an ergodic diffusion from discrete observations. *Scand. J. Statist.* **24**, 211–229.

Kutoyants, Yu. A. (1984). Parameter estimation for stochastic processes. Translated from the Russian and edited by B. L. S. Prakasa Rao. Research and Exposition in Mathematics, 6. Heldermann Verlag, Berlin.

Kutoyants, Yu. A. (1994). *Identification of dynamical systems with small noise.* Kluwer, Dordrecht.

Kutoyants, Yu. A. (1998). Statistical inference for spatial Poisson processes. Lecture Notes in Statistics, 134. Springer-Verlag, New York.

Kutoyants, Yu. A. (2004). *Statistical inference for ergodic diffusion processes.* Springer-Verlag, London.

Le Cam, L. (1986). *Asymptotic methods in statistical decision theory.* Springer-Verlag, New York.

Le Cam, L. and Yang, G. L. (1990). *Asymptotics in statistics.* Springer-Verlag, New York.

Masuda, H. (2010). Approximate self-weighted LAD estimation of discretely observed ergodic Ornstein-Uhlenbeck processes. *Electronic Journal of Statistics* **4**, 525–565.

Ogihara, T. and Yoshida, N. (2009). Quasi-likelihood analysis for the stochastic differential equation with jumps. To appear in *Statistical Inference for Stochastic Processes.*

Prakasa Rao, B. L. S. (1983). Asymptotic theory for nonlinear least squares estimator for diffusion processes. *Math. Operationsforsch. Statist. Ser. Statist.* **14**, 195–209.

Prakasa Rao, B. L. S. (1988). Statistical inference from sampled data for stochastic processes. *Contemp. Math.* **80**, 249–284. Amer. Math. Soc., Providence, RI.

Shimizu, Y. (2006). Density estimation of Levy measures for discretely observed diffusion processes with jumps. *J. Japan Statist. Soc.* **36**, 37–62.

Shimizu, Y. and Yoshida, N. (2006). Estimation of parameters for diffusion processes with jumps from discrete observations. *Statist. Infer. Stoch. Proc.* **9**, 227–277.

Sørensen, M. and Uchida, M. (2003). Small diffusion asymptotics for discretely sampled stochastic differential equations. *Bernoulli* **9**, 1051–1069.

Uchida, M. (2003). Estimation for dynamical systems with small noise from discrete observations. *J. Japan Statist. Soc.* **33**, 157–167.

Uchida, M. (2004). Estimation for discretely observed small diffusions based on approximate martingale estimating functions. *Scand. J. Statist.* **31**, 553–566.

Uchida, M. (2010). Contrast-based information criterion for ergodic diffusion processes from discrete observations. *Ann. Inst. Statist. Math.*, **62**, 161–187.

Yoshida, N. (1990). Asymptotic behavior of M-estimator and related random field for diffusion process. *Ann. Inst. Statist. Math.* **42**, 221–251.
Yoshida, N. (2005). Polynomial type large deviation inequalities and quasi-likelihood analysis for stochastic differential equations. (to appear in *Ann. Inst. Statist. Math.*)

Yoshida, N. (2011). Polynomial type large deviation inequalities and quasi-likelihood analysis for stochastic differential equations. *Ann. Inst. Statist. Math.*, 63, 431–479.