JOURNAL OF MULTIVARIATE ANALYSIS 41, 220–242 (1992)

Estimation for Diffusion Processes from Discrete Observation*

NAKAHIRO YOSHIDA

The Institute of Statistical Mathematics, 4-6-7 Minami-Azabu, Minato-ku, Tokyo 106, Japan

Communicated by the Editors

The maximum likelihood estimation of the unknown parameter of a diffusion process based on an approximate likelihood given by the discrete observation is treated when the diffusion coefficients are unknown and the condition for "rapidly increasing experimental design" is broken. The asymptotic normality of the joint distribution of the maximum likelihood estimator of the unknown parameter in the drift term and an estimator of the diffusion coefficient matrix is proved. We prove the weak convergence of the likelihood ratio random field, which serves to show the asymptotic behavior of the likelihood ratio tests with restrictions. © 1992 Academic Press, Inc.

1. INTRODUCTION

In this article we treat the following stochastic differential equation

$$\begin{aligned} dX_t &= a(X_t, \theta)\, dt + b(X_t)\, \sigma\, dW_t \\ X_0 &= x_0, \end{aligned} \tag{1.1}$$

where $\theta \in \bar{\Theta}$, Θ is a bounded convex domain in $\mathbf{R}^m$, and $\sigma \in \mathbf{R}^k \otimes \mathbf{R}^r$ are unknown parameters, a is an $\mathbf{R}^d$-valued function defined on $\mathbf{R}^d \times \bar{\Theta}$, b is an $\mathbf{R}^d \otimes \mathbf{R}^k$-valued function defined on $\mathbf{R}^d$ and W is an r-dimensional standard Wiener process. It is assumed that the observation from realization consists of X_{t_i}, $t_i = ih$, $h > 0$, $i = 0, 1, ..., N$. We estimate θ from the discrete data when $S = \sigma\sigma'$ is unknown. If S is known and the continuous observation is given, the likelihood function of θ is

$$\exp\left\{ \int_0^{Nh} a'\bar{B}\, dX_t - \frac{1}{2} \int_0^{Nh} a'\bar{B}a\, dt \right\},$$

Received July 30, 1990; revised June 6, 1991.

AMS classification numbers: 62F12, 62M05.

Key words and phrases: diffusion process, discrete observation, diffusion coefficient, likelihood ratio, maximum likelihood estimator.

* The research was supported in part by Grant-in-Aid for Encouragement of Young Scientists from the Ministry of Education, Science and Culture.

0047-259X/92 $5.00

where $\bar{B} = b\sigma[\sigma' b' b\sigma]^{-2} \sigma' b'$ and $[\sigma' b' b\sigma]^{-2}$ denotes the square of the Moore–Penrose generalized inverse of $\sigma' b' b\sigma$, see Liptser and Shiryayev [11]. Since $\bar{B}$ is invariant for σ satisfying $S = \sigma\sigma'$, we may write $\bar{B} = \bar{B}(x, S)$. The maximal likelihood estimator (MLE) is defined based on this formula. In our case, the data are discrete and we have to approximate the likelihood function to get the MLE. One of the approximations is to use the function

$$Q_{h,N}(S, \theta) = \exp\left\{ \sum_{i=1}^{N} a'_{i-1}(\theta)\, \bar{B}_{i-1}(S)\, \bar{\Delta}_i - \frac{h}{2} \sum_{i=1}^{N} a'_{i-1}(\theta)\, \bar{B}_{i-1}(S)\, a_{i-1}(\theta) \right\}, \tag{1.2}$$

where $a_i(\theta) = a(X_{t_i}, \theta)$, $\bar{B}_i(S) = \bar{B}(X_{t_i}, S)$, $\bar{\Delta}_i = X_{t_i} - X_{t_{i-1}}$. The approximate MLE is given maximizing (1.2) in θ. For this purpose, first we estimate S with some statistic $\hat{S}_0$ and then find the MLE, $\hat{\theta}_0$ say, substituting $\hat{S}_0$ for S in $Q_{h,N}(S, \theta)$. It is shown that $\hat{\theta}_0$ is a consistent estimator of θ. Next, using $\hat{\theta}_0$ we construct a better estimator $\hat{S}$ for S. Finally, we show that the MLE $\hat{\theta}$ for $Q_{h,N}$ with $\hat{S}$ is an efficient estimator for θ.

The weak convergence of the likelihood ratio random field is proved as in Ibragimov and Has'minskii [4–6], Inagaki and Ogata [7], and Kutoyants [8–10]. This enables us, e.g., to derive the asymptotic properties of likelihood ratio tests with restriction. We only assume a usual condition for consistency while Kutoyants assumed a condition which involves the Laplace transforms of some functionals.

The estimation for diffusion processes by discrete observation has been studied by several authors, see Prakasa Rao [13] and its references. Prakasa Rao [12, 13] treats this problem and shows that the least square estimator is asymptotically normal and efficient under the assumption $hN^{1/2} \to 0$, the condition for "rapidly increasing experimental design" [13]. Here we show this holds for the MLE even when the condition is broken, i.e., in the case $h^3 N = o(1)$. Florens-Zmirou [3] discussed the estimation problem with discrete observation for a one-dimensional diffusion

$$dX_t = V_0(X_t, \theta)\, dt + \sigma\, dW_t.$$

There it is shown that under $h^3 N \to 0$ an approximate maximum likelihood estimator $\hat{\theta}$ of θ has an asymptotic normal distribution and, for a quadratic variation type estimator $\hat{\sigma}^2$ of σ^2, $N^{1/2} h^{1/2} (\hat{\sigma}^2 - \sigma^2)$ converges in distribution to a normal distribution. This model is particular as the unknown σ does not affect maximizing the likelihood function for θ. Including this model, we can prove the convergence of the joint distribution of an approximate maximum likelihood estimator and an estimator of S under $h^3 N \to 0$.

Another approach for this problem is to use the MLE corresponding to the rigorous transition probability of the diffusion process from $t=t_{i-1}$ to $t=t_i$. Dacunha-Castelle and Florens-Zmirou [2] mention this approach for a one-dimensional diffusion process, whose transition probability function is written explicitly with an expectation of a functional of a Brownian bridge associated with the process. However, it does not seem easy in practice to calculate and maximize the likelihood function derived from this transition probability function, while it is theoretically important.

The plan of this article is as follows. In Section 2 we prepare notations and assumptions used later on. Section 3 presents our main results. Proof of these results are given in Section 4.

2. Notations and Assumptions

In this section we state notations and assumptions used later on. Let θ_0, σ_0, S_0 denote the true values of θ, σ, S, respectively. Suppose that $\theta_0 \in \Theta$. Define as follows:

- for matrix A, $|A|^2$ is the sum of squares of the elements of A,
- $B(x)=(b'b)^{-1}b'(x)$, $B_{i-1}=B(X_{t_{i-1}})$,
- C is a generic positive constant independent of h, N and other variables in some cases,
- $\partial_i=\partial/\partial x^i$, $\partial=(\partial_1, ..., \partial_d)$, $\delta_i=\partial/\partial\theta^i$, $\delta=(\delta_1, ..., \delta_m)$,
- $\Delta\bar{B}_i(S_2, S_1)=\bar{B}_i(S_2)-\bar{B}_i(S_1)$,
- $\Delta a_i(\theta_2, \theta_1)=a_i(\theta_2)-a_i(\theta_1)$,
- $\Delta_i(\theta)=X_{t_i}-X_{t_{i-1}}-ha(X_{t_{i-1}}, \theta)$,
- $L=\frac{1}{2}\sum_{i,j=1}^d v^{ij}\partial_i\partial_j+\sum_{i=1}^d a^i\partial_i$, $v^{ij}=[bSb']^{ij}$, for $\theta=\theta_0$ and $S=S_0$,
- the diffusion process X is ergodic with invariant measure ν for $\theta=\theta_0$,
- $Y(S, \theta)=\int_{\mathbf{R}^d} a'(x, \theta)\,\bar{B}(x, S)\{a(x, \theta_0)-\frac{1}{2}a(x, \theta)\}\,\nu(dx)$.

The following conditions are assumed in this article.

(1) There exists a constant L such that

$$|a(x, \theta_0)|+|b(x)| \leqslant L(1+|x|).$$

(2) There exists a constant L such that

$$|a(x, \theta_0)-a(y, \theta_0)|+|b(x)-b(y)| \leqslant L|x-y|.$$

(3) $\inf_x \det(b'b)(x)>0$.

(4) For each $p>0$, $\sup_t E|X_t|^p<\infty$

(5) The function $S \to \bar{B}(x, S)$ is Hölder continuous in a

neighborhood U of S_0 in $\mathbf{S}_+^k$: the totality of $k \times k$ symmetric nonnegative definite matrices endowed with the relative topology, that is, there exist $\alpha > 0$ and $C > 0$ such that

$$|\bar{B}(x, S_2) - \bar{B}(x, S_1)| \leqslant C(1 + |x|^C) |S_2 - S_1|^\alpha$$

for any S_1, S_2 in U and for all $x \in \mathbf{R}^d$.

(6) $a(x, \theta)$ is twice differentiable in $\theta \in \bar{\Theta}$ and $|\delta a(x, \theta)| + |\delta^2 a(x, \theta)| \leqslant C(1 + |x|^C)$.

(7) The function $\theta \to Y(S_0, \theta)$ has its unique maximum at $\theta = \theta_0$ in $\bar{\Theta}$.

(8) The functions a, δa, b, and $\bar{B}$ are smooth in x and their derivatives are of polynomial growth order in x uniformly in θ or S.

(9) $\Phi = \int_{\mathbf{R}^d} \delta a'(x, \theta_0) \bar{B}(x, S_0) \delta a(x, \theta_0) \nu(dx)$ is positive definite.

Remarks. To verify the condition (5), the following is sufficient. Equip the space of matrices with the Euclidean metric. Let $\pi: \mathbf{R}^k \otimes \mathbf{R}^r \to \mathbf{R}^k \otimes \mathbf{R}^k$ be a projection $\pi(\sigma) = \sigma\sigma'$. Then (5) holds if there exists a Hölder continuous local section ϕ from a neighborhood U of S_0 to $\mathbf{R}^k \otimes \mathbf{R}^r$: $\pi \circ \phi(S) = S$, $S \in U$, and the mapping $\sigma \to \eta(x, \sigma) = b\sigma[\sigma' b' b \sigma]^{-2} \sigma' b'$ is Hölder continuous on $\phi(U)$, i.e., there exist $\alpha > 0$ and $C > 0$ such that

$$|\eta(x, \sigma_2) - \eta(x, \sigma_1)| \leqslant C(1 + |x|^C) |\sigma_2 - \sigma_1|^\alpha.$$

For example, when $k = 2$, $r = 1$, $\sigma = (u, v)'$, $b = I$, we may take $\phi(S) = (S_{11}^{1/2}, S_{22}^{1/2})'$ for $S = (S_{ij})$. Then,

$$\eta(x, \sigma) = (u^2 + v^2)^{-2} \begin{pmatrix} u^2 & uv \\ uv & v^2 \end{pmatrix}$$

and it is smooth when $u^2 + v^2 \neq 0$. When $k \leqslant r$, $b = I$ and $S_0 \in \mathbf{R}^k \otimes \mathbf{R}^k$ is positive definite, let ϕ be the mapping $S \to \sigma \in \mathbf{R}^k \otimes \mathbf{R}^r$ such that for $i < j$ the (i, j)-elements of σ are zero. Then ϕ is a rational function of the elements of S and smooth. Since S_0 is regular, η is also smooth. In the sequel, we assume $\alpha = 1$ in (5) for simplicity as the argument is the same for arbitrary α.

3. Main Results

Let the estimator $\hat{S}_0$ of S be defined by

$$\hat{S}_0 = (hN)^{-1} \sum_{i=1}^{N} B_{i-1} \bar{\Delta}_i \bar{\Delta}_i' B_{i-1}'.$$

To start with, we consider the consistency of $\hat{S}_0$. The following proposition is more or less well known.

PROPOSITION 1. *If Conditions* (1–4) *hold,* $E|\hat{S}_0 - S_0| \leqslant C(h^{1/2} + N^{-1/2})$.

Next, let $\hat{\theta}_0$ be the MLE of θ with respect to $Q_{h,N}(\hat{S}_0, \theta)$. Then we have

PROPOSITION 2. *When* $h \to 0$, $N \to \infty$, *and* $hN \to \infty$, $\hat{\theta}_0 \to^p \theta_0$. *Moreover, if* $h^3 N \to 0$, $(hN)^{1/4}(\hat{\theta}_0 - \theta_0) \to^p 0$.

Now we can improve the convergence rate of $\hat{S}_0$ with a corrected residual sum of squares. Put

$$\hat{S}_1 = (hN)^{-1} \sum_{i=1}^{N} B_{i-1} \Delta_i(\hat{\theta}_0)\, \Delta'_i(\hat{\theta}_0)\, B'_{i-1},$$

$$\tilde{S}_1 = (hN)^{-1} \sum_{i=1}^{N} B_{i-1} \Delta_i(\theta_0)\, \Delta'_i(\theta_0)\, B'_{i-1},$$

and

$$\hat{S} = \hat{S}_1 - \frac{h}{2} N^{-1} \sum_{i=1}^{N} [U_{i-1} + V_{i-1} + V'_{i-1}],$$

where

$$U_{i-1} = B_{i-1} F(X_{t_{i-1}}, \hat{S}_0, \hat{\theta}_0)\, B'_{i-1},$$
$$V_{i-1} = B_{i-1}\, \partial a(X_{t_{i-1}}, \hat{\theta}_0)\, b(X_{t_{i-1}})\, \hat{S}_0,$$
$$F(x, S, \theta) = [F_{\lambda\mu}],\ F_{\lambda\mu} = L_{S,\theta}([bSb']_{\lambda\mu}),\ 1 \leqslant \lambda, \mu \leqslant d,$$

where $L_{S,\theta}$ is the generator corresponding to S and θ. Moreover, let

$$\begin{aligned}\tilde{S} = \tilde{S}_1 &- (hN)^{-1} \sum_{i=1}^{N} \frac{h^2}{2} B_{i-1} F(X_{t_{i-1}})\, B'_{i-1} \\ &- (hN)^{-1} \sum_{i=1}^{N} \int_{t_{i-1}}^{t_i} \int_{t_{i-1}}^{t} \sigma_0 \phi'(u)\, du\, dt \\ &- (hN)^{-1} \sum_{i=1}^{N} \int_{t_{i-1}}^{t_i} \int_{t_{i-1}}^{t} \phi(u)\, \sigma'_0\, du\, dt,\end{aligned}$$

where $F = F(\cdot, S_0, \theta_0)$, $\phi(u) = B_{i-1} \partial a(X_u, \theta_0)\, b(X_u)\, \sigma_0$ for $t_{i-1} < u \leqslant t_i$. With the consistent estimators $\hat{\theta}_0$ and $\hat{S}_0$, we can obtain a better estimator of S. The following lemma shows that the residual sum of squares generally needs to be corrected if $h^3 N = o(1)$.

PROPOSITION 3. $\tilde{S} = S_0 + w + w' + \rho$, *where*

$$w = (hN)^{-1} \sum_{i=1}^{N} \int_{t_{i-1}}^{t_i} \sigma_0 \, dW_t \left(\int_{t_{i-1}}^{t} \sigma_0 \, dW_t \right)',$$

$$N^{1/2} E|\rho| \leqslant C(h^2 + h^{1/2} + h^{3/2} N^{1/2} + h^2 N^{1/2}).$$

Moreover, if $h \to 0$, $hN \to \infty$, *and* $h^3 N = o(1)$, *then*

$$N^{1/2}(\hat{S} - S_0) \to^d H,$$

where $H = (H_{pq}) \in \mathbf{R}^k \otimes \mathbf{R}^k$ *is a multivariate normal random matrix with mean zero and covariances* $\operatorname{cov}(H_{pq}, H_{st}) = S_{ps} S_{qt} + S_{pt} S_{qs}$, $S_0 = (S_{ps})$. *If* $h \to 0$, $hN \to \infty$, *and* $hN^{1/2} = o(1)$, $N^{1/2}(\hat{S}_1 - S_0) \to^d H$.

Thus the correction by U_{i-1} and V_{i-1} is not necessary if $hN^{1/2} = o(1)$. Due to this proposition we can show the asymptotic properties of the MLE in the case $h^3 N = o(1)$. Let

$$Z_{h,N}(S, u) = Q_{h,N}(S, \theta_0 + (hN)^{-1/2} u) / Q_{h,N}(S, \theta_0)$$

and

$$B_{c,h,N} = \{u \in \mathbf{R}^m;\ |u| \leqslant c,\ \theta_0 + (hN)^{-1/2} u \in \bar{\Theta}\}$$

for $c > 0$. We will show the weak convergence of the random field $Z_{h,N}(\hat{S}, \cdot)$. For this purpose, we have to prepare three propositions.

PROPOSITION 4. *When* $h \to 0$, $hN \to \infty$, $h^3 N = o(1)$, *for each* $u \in \mathbf{R}^m$,

$$\log Z_{h,N}(\hat{S}, u) = u' \Delta_{h,N} - \tfrac{1}{2} u' \Phi u + \rho_{h,N}(u),$$

where

$$(\Delta_{h,N}, N^{1/2}(\hat{S} - S_0)) \to^d (\Delta, H),$$

$$\Delta \sim N_m(0, \Phi) \qquad \text{independent of } H,$$

$$\Phi = \int_{\mathbf{R}^d} \delta a'(x, \theta_0)\, \bar{B}(x, S_0)\, \delta a(x, \theta_0)\, \nu(dx),$$

and

$$\rho_{h,N}(u) \to^p 0.$$

PROPOSITION 5. *Define*

$$\bar{w}_{h,N}(\delta, c) = \sup |\log Z_{h,N}(\hat{S}, u_2) - \log Z_{h,N}(\hat{S}, u_1)|,$$

where the supremum is taken over u_1, $u_2 \in B_{c,h,N}$ *and* $|u_2 - u_1| \leqslant \delta$. *Then for* $c > 0$ *and* $\eta > 0$,

$$\lim_{\delta \to 0} \limsup_{h,N} P(\bar{w}_{h,N}(\delta, c) > \eta) = 0.$$

PROPOSITION 6. *For* $\varepsilon > 0$,

$$\lim_{c \to \infty} \limsup_{h,N} P(\sup_{|u| \geqslant c} Z_{h,N}(\hat{S}, u) > \varepsilon) = 0.$$

Consider the Banach space $C_0(\mathbf{R}^m)$ of the totality of real-valued continuous functions on $\mathbf{R}^m$ vanishing at the infinity with sup-norm. Let $E = (\mathbf{R}^k \otimes \mathbf{R}^k) \times C_0(\mathbf{R}^m)$ endowed with the product topology and let B be the Borel σ-field of E. Moreover, let $U_{h,N} = \{u; \theta_0 + (hN)^{-1/2} u \in \bar{\Theta}\}$. For $u \in U_{h,N}$, $Z_{h,N}(\hat{S}, u)$ have been defined and extend it to an element of $C_0(\mathbf{R}^m)$ whose maximal points are contained in $U_{h,N}$. Then, from Propositions 4, 5, and 6, we have the following result for the sequence of (E, B)-valued random variables $\{N^{1/2}(\hat{S} - S_0), Z_{h,N}(\hat{S}, \cdot)\}$. See, e.g., Ibragimov and Has'minskii [6] and Kutoyants [10].

THEOREM. *Suppose the conditions stated in Section 2 are satisfied. Then, the sequence* $\{N^{1/2}(\hat{S} - S_0), Z_{h,N}(\hat{S}, \cdot)\}$ *converges to* $\{H, Z(S_0, \cdot)\}$ *in distribution, where*

$$Z(S_0, \cdot) = \exp\{u'\Delta - \tfrac{1}{2} u'\Phi u\},$$

i.e., for any continuous functional f on (E, B),

$$E[f(N^{1/2}(\hat{S} - S_0), Z_{h,N}(\hat{S}, \cdot))] \to E[f(H, Z(S_0, \cdot))]$$

when $h \to 0$, $hN \to \infty$, $h^3 N = o(1)$. *In particular, for the maximum likelihood estimator* $\hat{\theta}$ *corresponding to* $Q_{h,N}(\hat{S}, \cdot)$,

$$\{N^{1/2}(\hat{S} - S_0), (hN)^{1/2}(\hat{\theta} - \theta_0)\} \to^d \{H, \Phi^{-1}\Delta\}.$$

EXAMPLE. Consider the following stochastic differential equation of one-dimension,

$$\ddot{X}_t + \kappa \dot{X}_t + \omega^2 X_t = \sigma \dot{W}_t,$$

where κ, ω, and σ are positive constants, $\omega^2 - \kappa^2/4 > 0$, and $\dot{W}$ is a white Guassian noise with unit variance. Let

$$\mathbf{X}_t = \begin{pmatrix} X_t \\ \dot{X}_t \end{pmatrix}, \qquad K = \begin{pmatrix} 0 & -1 \\ \omega^2 & \kappa \end{pmatrix}, \qquad b = \begin{pmatrix} 0 \\ 1 \end{pmatrix}.$$

Then, it is equivalent to the first-order vector stochastic differential equation

$$d\mathbf{X}_t = -K\mathbf{X}_t\, dt + b\sigma\, dW_t.$$

This two-dimensional diffusion is a degenerate one. The ergodic property is satisfied, which is seen more generally by Arnold and Kliemann [1]. Its invariant measure is a normal distribution on $\mathbf{R}^2$ with mean zero and covariance matrix $\mathrm{diag}(\sigma^2/2\kappa\omega^2, \sigma^2/2\kappa)$. The MLE $(\hat{\kappa}, \hat{\omega}^2, \hat{\sigma}^2)$ is asymptotically normal with covariance matrix $\mathrm{diag}(2\kappa, 2\kappa\omega^2, 2\sigma^4)$.

One of the applications of the weak convergence of the likelihood ratio process is to derive the asymptotic properties of likelihood ratio tests with restriction on parameter spaces. For this purpose, we may confine ourselves to calculating for the limits by means of the weak convergence. See Section 4 of Inagaki and Ogata [7]. The results there also hold for our $Q_{h,N}(\hat{S}, \theta)$ automatically. In particular we obtain the same results about AIC with respect to the approximate likelihood $Q_{h,N}(\hat{S}, \theta)$. In conclusion, it should be noted that we may choose other consistent estimators for $\hat{S}_0$ and $\hat{\theta}_0$.

4. Proofs

We often use the Novikov's moment inequality or Burkholder–Davis–Gundy inequality for martingales without notice.

Proof of Proposition 1. Put $Y_i(t) = B_{i-1}(X_t - X_{t_{i-1}})$, $t_{i-1} \leqslant t \leqslant t_i$. Then

$$Y_i(t) = \int_{t_{i-1}}^{t} f_i(u, \theta_0)\, du + \int_{t_{i-1}}^{t} g_i(u)\, \sigma_0 dW_u,$$

where $f_i(t, \theta) = B_{i-1}a(X_t, \theta)$, $g_i(t) = B_{i-1}b(X_t)$, $t_{i-1} \leqslant t \leqslant t_i$. By Itô's formula,

$$\hat{S}_0 = (hN)^{-1} \sum_{i=1}^{N} Y_i(t_i)\, Y_i'(t_i) = A_1 + A_1' + A_2 + A_2' + A_3,$$

where

$$A_1 = (hN)^{-1} \sum_{i=1}^{N} \int_{t_{i-1}}^{t_i} Y_i(t) f_i'(t, \theta_0)\, dt,$$

$$A_2 = (hN)^{-1} \sum_{i=1}^{N} \int_{t_{i-1}}^{t_i} Y_i(t)[g_i(t)\, \sigma_0 dW_t]',$$

$$A_3 = (hN)^{-1} \sum_{i=1}^{N} \int_{t_{i-1}}^{t_i} g_i(t)\, S_0 g_i'(t)\, dt.$$

It is clear that $E|A_1| \leqslant Ch^{1/2}$. Since

$$E|A_2|^2 = (hN)^{-2} \sum_{i=1}^{N} \int_{t_{i-1}}^{t_i} E|Y_i(t)|^2 \, |g_i(t)\,\sigma_0|^2 \, dt$$

and $E|Y_i(t)|^4 \leqslant Ch^2$, $E|A_2| \leqslant CN^{-1/2}$. Moreover, the inequality

$$E|g_i(t) - I|^2 \leqslant Ch$$

yields that $E|A_3 - S_0| \leqslant Ch^{1/2}$, so that $E|\hat{S}_0 - S_0| \leqslant C(h^{1/2} + N^{-1/2})$. ∎

Let $Y_{h,N}(S, \theta) = (hN)^{-1} \log Q_{h,N}(S, \theta)$. To show the consistency of $\hat{\theta}_0$ (Proposition 2), we will need the following two lemmas.

LEMMA 1. *For any $c > 0$, if $h \to 0$, $N \to \infty$, $v \to 0$, and $hN \to \infty$, then*

$$\sup_{\theta \in \bar{\Theta}, |M| \leqslant c} |Y_{h,N}(S_0 + vM, \theta) - Y_{h,N}(S_0, \theta)| \to^p 0,$$

where M is $k \times k$ symmetric matrix such that $S_0 + vM \in \mathbf{S}_+^k$.

Proof. Put $S_1 = S_0 + vM$ and $\eta_{h,N,v}(M, \theta) = Y_{h,N}(S_1, \theta) - Y_{h,N}(S_0, \theta)$. Then, $\eta_{h,N,v}(M, \theta) = A_1(M, \theta) + A_2(M, \theta) + A_3(M, \theta)$, where

$$A_1(M, \theta) = (hN)^{-1} \sum_{i=1}^{N} \int_{t_{i-1}}^{t_i} a'_{i-1}(\theta)\, \Delta\bar{B}_{i-1}(S_1, S_0)\, b(X_t)\, \sigma_0 dW_t,$$

$$A_2(M, \theta) = (hN)^{-1} \sum_{i=1}^{N} \int_{t_{i-1}}^{t_i} a'_{i-1}(\theta)\, \Delta\bar{B}_{i-1}(S_1, S_0)\, a(X_t, \theta_0)\, dt,$$

$$A_3(M, \theta) = -\frac{1}{2}(hN)^{-1} \sum_{i=1}^{N} ha'_{i-1}(\theta)\, \Delta\bar{B}_{i-1}(S_1, S_0)\, a'_{i-1}(\theta).$$

Let $p > (m + k(k+1)/2)/2$. By Burkholder–Davis–Gundy inequality, we have

$$E|A_1(M, \theta)|^{2p} \leqslant C(hN)^{-p} |v|^{2p} |M|^{2p}.$$

Since $E|A_2(M, \theta)|^{2p} \leqslant C|v|^{2p} |M|^{2p}$ and $E|A_3(M, \theta)|^{2p} \leqslant C|v|^{2p} |M|^{2p}$, we obtain

$$E|\eta_{h,N,v}(M, \theta)|^{2p} \leqslant C \max\{(hN)^{-p}, 1\}\, |v|^{2p} |M|^{2p}$$

and similarly

$$E|\eta_{h,N,v}(M_2, \theta_2) - \eta_{h,N,v}(M_1, \theta_1)|^{2p}$$
$$\leqslant C \max\{(hN)^{-p}, 1\} \max\{|v|^{2p}, 1\} |(M_2, \theta_2) - (M_1, \theta_1)|^{2p}.$$

From these inequalities we have $\sup_{\theta\in\bar{\Theta},|M|\leqslant c}|\eta_{h,N,v}(M,\theta)| \to^p 0$. See Appendix of Ibragimov and Has'minskii [6] or Lemma 3.1 of Yoshida [14]. ∎

The last argument will be used repeatedly in the sequel.

LEMMA 2. *When $h\to 0$, $N\to\infty$, and $hN\to\infty$,*

$$\sup_{\theta\in\bar{\Theta}} |Y_{h,N}(S_0,\theta) - Y(S_0,\theta)| \to^p 0.$$

Proof. Let $p>m/2$. As in the proof of Lemma 1, one has

$$E|Y_{h,N}(S_0,\theta)|^{2p} \leqslant C,$$

$$E|Y_{h,N}(S_0,\theta_2) - Y_{h,N}(S_0,\theta_1)|^{2p} \leqslant C|\theta_2-\theta_1|^{2p},$$

for θ, θ_1, $\theta_2\in\bar{\Theta}$. Therefore, the family of distributions of $Y_{h,N}(S_0,\cdot)$ on the Banach space $C(\bar{\Theta})$ with sup-norm is tight. Since $Y(S_0,\cdot)$ is a point of $C(\bar{\Theta})$, it suffices to show that for each $\theta\in\bar{\Theta}$, $Y_{h,N}(S_0,\theta)$ converges to $Y(S_0,\theta)$ in probability. Ergodic property ensures that $Y_{h,N}(S_0,\theta)\to^p Y(S_0,\theta)$. ∎

Proof of Proposition 2. From Lemmas 1 and 2, for any $c>0$,

$$\sup_{\substack{\theta\in\bar{\Theta}\\|M|\leqslant c}} |Y_{h,N}(S_0+h^{1/2}M,\theta) - Y(S_0,\theta)| \to^p 0.$$

Therefore, for any $c>0$ and $\varepsilon>0$,

$$\limsup_{h,N} P\{\sup_{\theta\in\bar{\Theta}}|Y_{h,N}(\hat{S}_0,\theta) - Y(S_0,\theta)|>\varepsilon\}$$

$$\leqslant \limsup_{h,N} P\{|\hat{S}_0 - S_0|>h^{1/2}c\}$$

$$+\limsup_{h,N} P\{\sup_{\substack{\theta\in\bar{\Theta}\\|M|\leqslant c}} |Y_{h,N}(S_0+h^{1/2}M,\theta) - Y(S_0,\theta)|>\varepsilon\}$$

$$\leqslant \limsup_{h,N} P\{|\hat{S}_0 - S_0|>h^{1/2}c\}.$$

Since c is arbitrary, the left-hand side equals zero, from Proposition 1. It is easy to show the consistency of $\hat{\theta}_0$, using Condition (7).

Next, we shall show the second assertion. There exists a sequence $\varepsilon_{h,N}$ such that $\varepsilon_{h,N}\to 0$ and $P[|\hat{\theta}_0-\theta_0|>\varepsilon_{h,N}]\to 0$ by consistency of $\hat{\theta}_0$. By Proposition 1, it suffices to show that for any $\delta>0$, $P[A_{h,N}]\to 1$, where

$$A_{h,N} = \{\sup_{\substack{(hN)^{-1/4}\delta\leqslant|\theta-\theta_0|\leqslant\varepsilon_{h,N}\\|M|\leqslant(hN)^{-1/4}}} [Y_{h,N}(S_0+M,\theta) - Y_{h,N}(S_0+M,\theta_0)]<0\}.$$

We have that

$$(hN)^{1/2}[Y_{h,N}(S_0+(hN)^{-1/4}M,\theta)-Y_{h,N}(S_0+(hN)^{-1/4}M,\theta_0)]$$
$$=B_1(M,\theta)+B_2(M,\theta)+B_3(M,\theta)+B_4(M,\theta),$$

where

$$B_1(M,\theta)=(hN)^{-1/2}\sum_{i=1}^{N}\int_{t_{i-1}}^{t_i}\Delta a'_{i-1}(\theta,\theta_0)$$
$$\times\bar{B}_{i-1}(S_0+(hN)^{-1/4}M)\,b(X_t)\,\sigma_0\,dW_t,$$

$$B_2(M,\theta)=(hN)^{-1/2}\sum_{i=1}^{N}\int_{t_{i-1}}^{t_i}\Delta a'_{i-1}(\theta,\theta_0)$$
$$\times\bar{B}_{i-1}(S_0+(hN)^{-1/4}M)$$
$$\times\int_{t_{i-1}}^{t}\partial a(X_s,\theta_0)\,b(X_s)\,\sigma_0\,dW_s\,dt,$$

$$B_3(M,\theta)=(hN)^{-1/2}\sum_{i=1}^{N}\int_{t_{i-1}}^{t_i}\Delta a'_{i-1}(\theta,\theta_0)$$
$$\times\bar{B}_{i-1}(S_0+(hN)^{-1/4}M)\int_{t_{i-1}}^{t}F(X_s)\,ds\,dt,$$

$$B_4(M,\theta)=-\frac{1}{2}(hN)^{-1/2}h\sum_{i=1}^{N}\Delta a'_{i-1}(\theta,\theta_0)$$
$$\times\bar{B}_{i-1}(S_0+(hN)^{-1/4}M)\,\Delta a_{i-1}(\theta,\theta_0)$$

for some function F. Using Burkholder's inequalities for continuous and discrete martingales, we have moment inequalities for B_1 and B_2, and

$$\sup_{\substack{|\theta-\theta_0|\leq\varepsilon_{h,N}\\|M|\leq c}}|B_i(M,\theta)|\to^{P}0,\qquad i=1,2,$$

for $c>0$. We will not go into detail here. We see that $\sup_{\theta\in\Theta,|M|\leq c}|B_3(M,\theta)|\to^{P}0$ for $c>0$ from $h^3N\to 0$. Finally, using positive definiteness of Fisher information, for $B_4(M,\theta)$ we have $P[A_{h,N}]\to 1$ and this completes the proof. ∎

Proof of Proposition 3. Let

$$\Phi_{1,i}=\int_{t_{i-1}}^{t_i} B_{i-1}b(X_t)\,\sigma_0\,dW_t,$$

$$\Phi_{2,i}=\int_{t_{i-1}}^{t_i}\left[\int_{t_{i-1}}^{t}\phi(u)\,dW_u\right]dt,$$

$$\rho_{1,i}=\int_{t_{i-1}}^{t_i}\left[\int_{t_{i-1}}^{t} B_{i-1}La(X_u, S_0, \theta_0)\,du\right]dt,$$

where La is the vector of the elements of a operated by L. Then,

$$\begin{aligned}\tilde{S}_1 &= (hN)^{-1}\sum_{i=1}^{N}(\Phi_{1,i}+\Phi_{2,i}+\rho_{1,i})(\Phi_{1,i}+\Phi_{2,i}+\rho_{1,i})' \\ &= (hN)^{-1}\sum_{i=1}^{N}\Phi_{1,i}\Phi'_{1,i}+(hN)^{-1}\sum_{i=1}^{N}\Phi_{1,i}\Phi'_{2,i} \\ &\quad+(hN)^{-1}\sum_{i=1}^{N}\Phi_{2,i}\Phi'_{1,i}+\rho_2,\end{aligned} \tag{1}$$

where

$$\begin{aligned}\rho_2 &= (hN)^{-1}\sum_{i=1}^{N}\{\rho_{1,i}(\Phi_{1,i}+\Phi_{2,i}+\rho_{1,i})' \\ &\quad+(\Phi_{1,i}+\Phi_{2,i})\,\rho'_{1,i}\}+(hN)^{-1}\sum_{i=1}^{N}\Phi_{2,i}\Phi'_{2,i},\end{aligned}$$

and $E|\rho_2|\leqslant Ch^{3/2}+Ch^2$. We have

$$(hN)^{-1}\sum_{i=1}^{N}\Phi_{1,i}\Phi'_{1,i}=\Phi_3+\Phi_4+\Phi_5,$$

where

$$\Phi_3=\Phi'_4=(hN)^{-1}\sum_{i=1}^{N}\int_{t_{i-1}}^{t_i}\Psi_{i,t}\,dW_t\left(\int_{t_{i-1}}^{t}\Psi_{i,u}\,dW_u\right)',$$

$$\Phi_5=(hN)^{-1}\sum_{i=1}^{N}\int_{t_{i-1}}^{t_i}\Psi_{i,t}\Psi'_{i,t}\,dt,$$

and

$$\Psi_{i,t}=B_{i-1}b(X_t)\,\sigma_0.$$

It is easy to show that $E|\Phi_3 - w|^2 \leqslant ChN^{-1}$ and $E|\Phi_4 - w'|^2 \leqslant ChN^{-1}$. On the other hand,

$$\begin{aligned}\Phi_5 &= (hN)^{-1} \sum_{i=1}^{N} \int_{t_{i-1}}^{t_i} B_{i-1} b(X_{t_{i-1}})\, S_0 b'(X_{t_{i-1}})\, B'_{i-1}\, dt \\ &\quad + (hN)^{-1} \sum_{i=1}^{N} \int_{t_{i-1}}^{t_i} B_{i-1} \left[\int_{t_{i-1}}^{t} F_1(X_u) \cdot dW_u \right] B'_{i-1}\, dt \\ &\quad + (hN)^{-1} \sum_{i=1}^{N} \int_{t_{i-1}}^{t_i} B_{i-1} \left[\int_{t_{i-1}}^{t} F(X_u)\, du \right] B'_{i-1}\, dt, \\ &=: S_0 + \Phi_6 + \Phi_7,\end{aligned}$$

where F_1 is a $d \times dr$ blocked matrix whose (p, q)-block is $\partial[bS_0b']_{pq} b\sigma_0$, $\cdot$ denotes the inner product of dW and each block; and F is an $\mathbf{R}^d \otimes \mathbf{R}^d$-valued function. We see that

$$E|\Phi_6|^2 \leqslant ChN^{-1}$$

and

$$E\left|\Phi_7 - (hN)^{-1} \sum_{i=1}^{N} \frac{h^2}{2} B_{i-1} F(X_{t_{i-1}})\, B'_{i-1}\right| \leqslant Ch^{3/2}.$$

Therefore,

$$\begin{aligned}&E\Big|(hN)^{-1} \sum_{i=1}^{N} \Phi_{1,i} \Phi'_{1,i} - S_0 - w - w' \\ &\qquad - (hN)^{-1} \sum_{i=1}^{N} \frac{h^2}{2} B_{i-1} F(X_{t_{i-1}})\, B'_{i-1}\Big| \\ &\qquad \leqslant C(h^{1/2} N^{-1/2} + h^{3/2}).\end{aligned} \tag{2}$$

Next, let

$$\Psi_1 = (hN)^{-1} \sum_{i=1}^{N} \left(\int_{t_{i-1}}^{t_i} \sigma_0\, dW_t \right) \left(\int_{t_{i-1}}^{t_i} \left[\int_{t_{i-1}}^{t} \phi(u)\, dW_u \right] dt \right)'.$$

We have

$$E\left|(hN)^{-1} \sum_{i=1}^{N} \Phi_{1,i} \Phi'_{2,i} - \Psi_1\right| \leqslant Ch^{3/2}$$

with $\Psi_1 = \Phi_8 + \Phi_9$, where

$$\Phi_8 = (hN)^{-1} \sum_{i=1}^{N} \int_{t_{i-1}}^{t_i} \sigma_0 \, dW_s \left(\int_{t_{i-1}}^{s} \left[\int_{t_{i-1}}^{t} \phi_u \, dW_u \right] dt \right)'$$

$$\Phi_9 = (hN)^{-1} \sum_{i=1}^{N} \int_{t_{i-1}}^{t_i} \left(\int_{t_{i-1}}^{t} \sigma_0 \, dW_u \right) \left[\int_{t_{i-1}}^{t} \phi_u \, dW_u \right]' dt.$$

Then, $E|\Phi_8|^2 \leqslant Ch^2 N^{-1}$. On the other hand, $\Phi_9 = \Phi_{10} + \Phi_{11} + \Phi_{12}$, where

$$\Phi_{10} = (hN)^{-1} \sum_{i=1}^{N} \int_{t_{i-1}}^{t_i} \left(\int_{t_{i-1}}^{t} \sigma_0 \, dW_v \left[\int_{t_{i-1}}^{v} \phi_u \, dW_u \right]' \right) dt,$$

$$\Phi_{11} = (hN)^{-1} \sum_{i=1}^{N} \int_{t_{i-1}}^{t_i} \int_{t_{i-1}}^{t} \left(\int_{t_{i-1}}^{u} \sigma_0 \, dW_v \right) [\phi_u \, dW_u]' \, dt,$$

$$\Phi_{12} = (hN)^{-1} \sum_{i=1}^{N} \int_{t_{i-1}}^{t_i} \int_{t_{i-1}}^{t} \sigma_0 \phi_u' \, du \, dt.$$

Then, $E|\Phi_{10}|^2 \leqslant Ch^2 N^{-1}$, $E|\Phi_{11}|^2 \leqslant Ch^2 N^{-1}$, and hence

$$E\left|(hN)^{-1} \sum_{i=1}^{N} \Phi_{1,i} \Phi_{2,i}' - (hN)^{-1} \sum_{i=1}^{N} \int_{t_{i-1}}^{t_i} \int_{t_{i-1}}^{t} \sigma_0 \phi_u' \, du \, dt\right| \leqslant C(hN^{-1/2} + h^{3/2}). \tag{3}$$

From (1–3),

$$\begin{aligned} E\Big|\tilde{S}_1 - S_0 - w - w' &- (hN)^{-1} \sum_{i=1}^{N} \frac{h^2}{2} B_{i-1} F(X_{t_{i-1}}) B_{i-1}' \\ &- (hN)^{-1} \sum_{i=1}^{N} \int_{t_{i-1}}^{t_i} \int_{t_{i-1}}^{t} \sigma_0 \phi_u' \, du \, dt \\ &- (hN)^{-1} \sum_{i=1}^{N} \int_{t_{i-1}}^{t_i} \int_{t_{i-1}}^{t} \phi_u \sigma_0' \, du \, dt\Big| \\ &\leqslant C(hN^{-1/2} + h^{1/2} N^{-1/2} + h^{3/2} + h^2). \end{aligned} \tag{4}$$

This proves the first assertion of the proposition. Next, we shall show that $\tilde{S}_1$ can be replaced by $\hat{S}_1$. For $\varepsilon > 0$ and $u \in \mathbf{R}^m$, put

$$\xi(u) = N^{1/2} (hN)^{-1} \sum_{i=1}^{N} h B_{i-1} \{ a(X_{t_{i-1}}, \theta_0 + \varepsilon u) - a(X_{t_{i-1}}, \theta_0) \} \Delta'(\theta_1) B_{i-1}',$$

where θ_1 denotes $\theta_0 + \varepsilon u$ or θ_0. Let $p > m/2$. Then

$$E|\xi(u)|^{2p} \leqslant C\varepsilon^{2p}|u|^{2p}\{(hN^{1/2})^{2p} + h^p\}.$$

Similarly,

$$E|\xi(u) - \xi(v)|^{2p} \leqslant C\varepsilon^{2p}|u - v|^{2p}\{(hN^{1/2})^{2p} + h^p\}.$$

Therefore, when $h^3N = o(1)$, $h \to 0$ and $\varepsilon = (hN)^{-1/4} \to 0$, $\sup_{|u| \leqslant 1}|\xi(u)| \to^p 0$. Moreover, if $hN \to \infty$, from the consistency of $\hat{\theta}_0$, by Proposition 2, we see that $P\{|\hat{\theta}_0 - \theta| > \varepsilon\} \to 0$. This implies that $N^{1/2}(\hat{S}_1 - \tilde{S}_1) \to^p 0$. Using Propositions 1 and 2, it is easy to show that the last three terms in the left-hand side of (4) can be replaced by U_{i-1}, V'_{i-1} and so on. The asymptotic distribution of $w + w'$ is trivial by the fact about the Wishart distribution. ▮

The following two lemmas serve to show Proposition 4, i.e., an analog of local asymptotic normality of experiments.

LEMMA 3. *Let W be a bounded set of $\mathbf{R}^k \otimes \mathbf{R}^k$. For $c > 0$, when $h \to 0$, $N \to \infty$, $\varepsilon \to 0$, and $\varepsilon = o((hN)^{-1/2})$,*

$$\sup_{\substack{u \in B_{c,h,N} \\ M \in W}} |\log Z_{h,N}(S_0 + \varepsilon M, u) - \log Z_{h,N}(S_0, u)| \to^p 0.$$

In particular, if $\varepsilon = N^{-1/2}$, this is satisfied.

Proof. Put $S = S_0 + \varepsilon M$, $\theta = \theta_0 + (hN)^{-1/2} u$ and

$$\xi(M, u) = \log Z_{h,N}(S, u) - \log Z_{h,N}(S_0, u).$$

Then,

$$\begin{aligned}\xi(M, u) = &\sum_{i=1}^{N} \Delta a'_{i-1}(\theta, \theta_0)\, \Delta\bar{B}_{i-1}(S, S_0)\, \bar{\Delta}_i \\ &-\frac{1}{2}\sum_{i=1}^{N} h\Delta a'_{i-1}(\theta, \theta_0)\, \Delta\bar{B}_{i-1}(S, S_0)\; a_{i-1}(\theta) \\ &-\frac{1}{2}\sum_{i=1}^{N} ha'_{i-1}(\theta_0)\, \Delta\bar{B}_{i-1}(S, S_0)\, \Delta a_{i-1}(\theta, \theta_0).\end{aligned}$$

The second and the third terms in the right-hand side converges to zero in probability uniformly in u, and M for boundedness. Similarly, the bounded variational part of the first term of the right-hand side converges to zero in probability uniformly in u and M. Its martingale part is $m(M, u) = \sum_{i=1}^{N} \int_{t_{i-1}}^{t_i} F_2(t, M, u)\, dW_t$, where

$$F_2(t, M, u) = \Delta a'_{i-1}(\theta, \theta_0)\, \Delta \bar{B}_{i-1}(S, S_0)\, b(X_t)\, \sigma_0$$

for $t \in [t_{i-1}, t_i)$. For p such that $2p > \max\{m + k^2, 2\}$,

$$E|m(M, u)|^{2p} \leq C\varepsilon^{2p} \to 0,$$

$$E|m(M_2, u_2) - m(M_1, u_1)|^{2p} \leq C\varepsilon^{2p}\{|M_2 - M_1|^{2p} + |u_2 - u_1|^{2p}\}.$$

Therefore, $m(M, u)$ converges to zero in probability uniformly in $u \in B_{c,h,N}$ and $M \in W$. ∎

Define as follows:

$$\theta = \theta_0 + (hN)^{-1/2} u,$$

$$\Delta_{h,N}(S) = (hN)^{-1/2} \sum_{i=1}^{N} \delta a'_{i-1}(\theta_0)\, \bar{B}_{i-1}(S) \int_{t_{i-1}}^{t_i} b(X_t)\, \sigma_0\, dW_t,$$

$$r_{h,N}(S, u) = \tfrac{1}{2}\{u'\Phi u - \Phi(S, u)\} + r_3(S, u),$$

$$\Phi(S, u) = \sum_{i=1}^{N} h\, \Delta a'_{i-1}(\theta, \theta_0)\, \bar{B}_{i-1}(S)\, \Delta a_{i-1}(\theta, \theta_0),$$

$$r_3(S, u) = r_4(S, u) + r_5(S, u) + r_6(S, u),$$

$$r_4(S, u) = \sum_{i=1}^{N} [\Delta a'_{i-1}(\theta, \theta_0) - (hN)^{-1/2} u'\delta a'_{i-1}(\theta_0)] \times \bar{B}_{i-1}(S) \int_{t_{i-1}}^{t_i} b(X_t)\, \sigma_0\, dW_t,$$

$$r_5(S, u) = (hN)^{-1/2} \sum_{i=1}^{N} u'\delta a'_{i-1}(\theta_0)\, \bar{B}_{i-1}(S) \times \int_{t_{i-1}}^{t_i} [a(X_t, \theta_0) - a(X_{t_{i-1}}, \theta_0)]\, dt,$$

$$r_6(S, u) = \sum_{i=1}^{N} [\Delta a'_{i-1}(\theta, \theta_0) - (hN)^{-1/2} u'\delta a'_{i-1}(\theta_0)]\, \bar{B}_{i-1}(S) \times \int_{t_{i-1}}^{t_i} [a(X_t, \theta_0) - a(X_{t_{i-1}}, \theta_0)]\, dt.$$

LEMMA 4. *When $h \to 0$, $hN \to \infty$, $h^3 N = o(1)$, for each $u \in \mathbf{R}^m$,*

$$\log Z_{h,N}(S_0, u) = u'\Delta_{h,N} - \tfrac{1}{2} u'\Phi u + \rho_{h,N}(u),$$

where

$$(\Delta_{h,N}, N^{1/2}(\hat{S}-S_0)) \to^d (\Delta, H),$$

$$\Delta \sim N_m(0, \Phi) \qquad \textit{independent of } H$$

and

$$\rho_{h,N}(u) \to^p 0.$$

Proof. Setting $\theta = \theta_0 + (hN)^{-1/2} u$, we have

$$\log Z_{h,N}(S_0, u) = u'\Delta_{h,N} - \tfrac{1}{2}\Phi_{13} + \rho_3,$$

where $\Delta_{h,N} = \Delta_{h,N}(S_0)$, $\Phi_{13} = \Phi(S_0, u)$, and $\rho_i = r_i(S_0, u)$, $i = 3, 4, 5, 6$. By the property of the Moore–Penrose generalized inverse matrix,

$$\bar{B}(x, S_0)\, b(x)\, \sigma_0\sigma_0' b'(x)\, \bar{B}(x, S_0) = \bar{B}(x, S_0).$$

Hence, $\Delta_{h,N} \to^d \Delta$. From the representations of w in Proposition 3 and $\Delta_{h,N}$, $(N^{1/2}(\hat{S}-S_0), \Delta_{h,N})$ proves to be asymptotically equivalent to a discrete-time martingale array, converging in distribution to a multivariate normal random vector (H, Δ). The independence of H and Δ is shown as follows: Consider a probability space with filtration $(\Omega, H, P; H_t, t \geqslant 0)$. Suppose that F_t and $\tilde{F}_t$ are (H_t)-adaptive processes, that $\tilde{F}_t$ is $H_{t_{i-1}}$-measurable for $t \in [t_{i-1}, t_i)$, and that $E|F_t - \tilde{F}_t|^2 \leqslant Ch$. Then, for a d-dimensional Wiener process W,

$$\begin{aligned}
E\Bigg[& h^{-1}N^{-1/2}\left\{\sum_{i=1}^{N}\int_{t_{i-1}}^{t_i}\left(\int_{t_{i-1}}^{t} dW_u^p\right) dW_t^q\right\} \\
& \cdot (hN)^{-1/2}\left\{\sum_{i=1}^{N}\int_{t_{i-1}}^{t_i} F_t\, dW_t^s\right\}\Bigg] \\
&= h^{-3/2}N^{-1}\,\delta_{qs}\sum_{i=1}^{N}\int_{t_{i-1}}^{t_i} E\left(\int_{t_{i-1}}^{t} dW_u^p F_t\right) dt \\
&= h^{-3/2}N^{-1}\,\delta_{qs}\sum_{i=1}^{N}\int_{t_{i-1}}^{t_i} E\left(\int_{t_{i-1}}^{t} dW_u^p [F_t - \tilde{F}_t]\right) dt \\
&= O(h^{1/2}), \qquad 1 \leqslant p, q, s \leqslant d.
\end{aligned}$$

We note that for $p > 0$, $\sup_t E|\bar{B}(X_t, S_0)|^p < \infty$, since we can choose some S_1 near S_0 such that $[\sigma_1' b' b \sigma_1]^- = \sigma_1' S_1^{-1}(b'b)^{-1} S_1^{-1}\sigma_1$ $(S_1 = \sigma_1\sigma_1')$ is bounded in x. Therefore, the asymptotic covariance of $N^{1/2}(\hat{S}-S_0)$ and

$\Delta_{h,N}$ is zero. Moreover, it is obvious that $E|\rho_4|^2 \leqslant C(hN)^{-1}$ and $\rho_6 \to^p 0$. We can write

$$\rho_5 = (hN)^{-1/2} \sum_{i=1}^{N} u' \, \delta a'_{i-1}(\theta_0) \, \bar{B}_{i-1}(S_0)$$
$$\times \int_{t_{i-1}}^{t_i} \int_{t_{i-1}}^{t} [\partial a \cdot b(X_u) \, \sigma_0 \, dW_u + F_3(X_u) \, du] \, dt.$$

The L^2-norm of the term involving a stochastic integral with respect to the Wiener process is $O(h)$. The term of F_3 is $O_p(h^{3/2}N^{1/2})$ and tends to zero. Therefore, ρ_5, hence, ρ_3 tends to zero. From the ergodic property, $\Phi_{13} \to^p u'\Phi u$. Put $\rho_{h,N}(u) = \frac{1}{2}(u'\Phi u - \Phi_{13}) + \rho_3$, then $\rho_{h,N}(u) \to^p 0$. ∎

Proof of Proposition 4. From Proposition 3, for positive number η, there exists $A > 0$ such that

$$P(N^{1/2}|\hat{S} - S_0| > A) < \eta/2$$

for small h and large N. Let $S = S_0 + N^{-1/2}M$. Then by Lemma 3, for $\varepsilon > 0$,

$$\begin{aligned} &P(|\log Z_{h,N}(S, u) - \log Z_{h,N}(S_0, u)| > \varepsilon) \\ &\quad \leqslant P(N^{1/2}|\hat{S} - S_0| > A) \\ &\qquad + P(\sup_{|M| \leqslant A} |\log Z_{h,N}(S, u) - \log Z_{h,N}(S_0, u)| > \varepsilon) \\ &\quad < \eta \end{aligned}$$

when $h \to 0$, $hN \to \infty$, $h^3N = o(1)$. Then we can use Lemma 4. ∎

LEMMA 5. *Let* $w_{h,N}(\delta, c) = \sup |\log Z_{h,N}(S, u_2) - \log Z_{h,N}(S, u_1)|$ *for* $\delta > 0$ *and* $c > 0$, *where the supremum is taken over* u_1, u_2, *and* S *such as* $u_1, u_2 \in B_{c,h,N}$, $|u_2 - u_1| \leqslant \delta$, $N^{1/2}|S - S_0| \leqslant A$. *Then, for any* $\eta > 0$,

$$\lim_{\delta \to 0} \limsup_{h,N} P(w_{h,N}(\delta, c) > \eta) = 0.$$

Proof. Let

$$w'_{h,N}(\delta, c) = \sup |\log Z_{h,N}(S_0, u_2) - \log Z_{h,N}(S_0, u_1)|,$$

where the supremum is taken over u_1, $u_2 \in B_{c,h,N}$, $|u_2 - u_1| \leqslant \delta$. Moreover, let

$$g_{h,N}(\delta, c) = \sup |\log Z_{h,N}(S, u) - \log Z_{h,N}(S_0, u)|,$$

where $S = S_0 + N^{-1/2}M$ and the supremum is taken over $u \in B_{c,h,N}$, $|M| \leqslant A$. From Lemma 3, $g_{h,N}(\delta, c) \to^p 0$. For $\varepsilon > 0$,

$$\begin{aligned}
&\limsup_{h,N} P(w_{h,N}(\delta, c) > \varepsilon) \\
&\quad \leqslant \limsup_{h,N} P\left(w'_{h,N}(\delta, c) > \frac{\varepsilon}{2}\right) + \limsup_{h,N} P\left(2g_{h,N}(\delta, c) > \frac{\varepsilon}{2}\right) \\
&\quad = \limsup_{h,N} P\left(w'_{h,N}(\delta, c) > \frac{\varepsilon}{2}\right).
\end{aligned}$$

Therefore, it suffices to show that

$$\lim_{\delta \to 0} \limsup_{h,N} P(w'_{h,N}(\delta, c) > \varepsilon) = 0.$$

This can be proved as in Lemma 4.1 in Yoshida [14] and the proof is completed. ∎

Proof of Proposition 5. For $\varepsilon > 0$ and $\eta > 0$, from Proposition 3, there exists $A > 0$ such that

$$\begin{aligned}
&\limsup_{h,N} P(\bar{w}_{h,N}(\delta, c) > \eta) \\
&\quad \leqslant \limsup_{h,N} P(w_{h,N}(\delta, c) > \eta) + \limsup_{h,N} P(N^{1/2}|\hat{S} - S_0| > A) \\
&\quad \leqslant \limsup_{h,N} P(w_{h,N}(\delta, c) > \eta) + \varepsilon.
\end{aligned}$$

By Lemma 5,

$$\lim_{\delta \to 0} \limsup_{h,N} P(\bar{w}_{h,N}(\delta, c) > \eta) \leqslant \varepsilon.$$

ε is arbitrary and this completes the proof. ∎

Proposition 3 and the following lemma suffice to show Proposition 6.

LEMMA 6. *If Φ is positive definite, for $\varepsilon > 0$ and $A > 0$,*

$$\lim_{c \to \infty} \limsup_{h,N} P\Big(\sup_{\substack{|u| \geqslant c \\ N^{1/2}|S - S_0| \leqslant A}} Z_{h,N}(S, u) > \varepsilon\Big) = 0.$$

Proof. Since Φ is positive definite, there exists a positive number η such that,

$$\eta |u|^2 \leqslant \tfrac{1}{4} u'\Phi u, \qquad u \in \mathbf{R}^m.$$

As in Lemma 4, we have that when θ_0 and S_0 are true values,

$$\log Z_{h,N}(S, u) = u' \Delta_{h,N}(S) - \tfrac{1}{2}u'\Phi u + r_{h,N}(S, u).$$

Let $p > (m + k(k+1)/2)/2$ and let

$$\bar{r}_i(S, u) = \frac{1}{1+|u|^2} r_i(S, u), \qquad i = 3, 4, 5, 6.$$

Then we have

$$E|\bar{r}_4(S, u)|^{2p} \leqslant C(hN)^{-p},$$
$$E|\bar{r}_4(S_2, u_2) - \bar{r}_4(S_1, u_1)|^{2p} \leqslant C(hN)^{-p}\{|S_2 - S_1| + |u_2 - u_1|\}^{2p}.$$

Therefore, for any $\delta > 0$,

$$\sup_{U_1, V} |\bar{r}_4(S, u)| \to^P 0,$$

where $U_1 = \{u; |u| \leqslant \delta(hN)^{1/2}\}$ and $V = \{S; N^{1/2}|S - S_0| \leqslant A\}$. See the proof of Lemma 3.2 of Yoshida [14]. It is obvious that $\sup_{U_1, V} |\bar{r}_6(S, u)| \to^P 0$. Next, we write

$$r_5(S, u) = (hN)^{-1/2} \sum_{i=1}^{N} u' \,\delta a'_{i-1}(\theta_0)\, \bar{B}_{i-1}(S)$$
$$\times \int_{t_{i-1}}^{t_i} \int_{t_{i-1}}^{t} [\partial a \cdot b(X_u)\, \sigma_0\, dW_u + F_3(X_u)\, du]\, dt,$$

and call the terms involving dW and F_3, $r_{51}(S, u)$ and $r_{52}(S, u)$, respectively. Moreover, define $\bar{r}_{51}$ and $\bar{r}_{52}$ such as $\bar{r}_4$. It is easy to show

$$\sup_{U_1, V} |\bar{r}_{52}(S, u)| \to^P 0.$$

As in the estimation for $\bar{r}_4$, we obtain

$$E|\bar{r}_{51}(S_2, u)|^{2p} \leqslant Ch^{2p},$$
$$E|\bar{r}_{51}(S_2, u_2) - \bar{r}_{51}(S_1, u_1)|^{2p} \leqslant Ch^{2p}\{|S_2 - S_1| + |u_2 - u_1|\}^{2p},$$

and hence,

$$\sup_{U_1, V} |\bar{r}_{51}(S, u)| \to^P 0$$

and

$$\sup_{U_1, V} |\bar{r}_3(S, u)| \to^P 0.$$

On the other hand, for any $\varepsilon > 0$, if $\delta > 0$ is small enough,

$$\lim_{h,N} P\left\{\sup_{U_1,V} \frac{1}{1+|u|^2} |u'\Phi u - \Phi(S,u)| < \varepsilon\right\} = 1.$$

So, for any $\varepsilon > 0$,

$$\lim_{h,N} P\{A_{h,N}\} = 1$$

for small $\delta > 0$, where

$$A_{h,N} = \left\{\sup_{U_1,V} \frac{1}{1+|u|^2} |r_{h,N}(S,u)| < \varepsilon\right\}.$$

Let $\varepsilon < \eta$. On the event $A_{h,N}$, if $|u| \leqslant \delta(hN)^{1/2}$, $N^{1/2}|S - S_0| \leqslant A$,

$$\log Z_{h,N}(S,u) \leqslant |u|\,|\Delta_{h,N}(S)| - \eta|u|^2 + \varepsilon.$$

Defining $U_2 = \{u; r \leqslant |u| \leqslant \delta(hN)^{1/2}\}$, where r is a positive number, we have

$$\begin{aligned}
&P\left\{\sup_{U_2,V} Z_{h,N}(S,u) \geqslant \exp\left(-\frac{\eta r^2}{2}\right)\right\}\\
&\quad\leqslant P\{A^c_{h,N}\} + P\left\{\sup_{U_2}(|u| \sup_V |\Delta_{h,N}(S)| - \eta|u|^2) + \varepsilon \geqslant -\frac{\eta r^2}{2}\right\}\\
&\quad\leqslant P\{\sup_V |\Delta_{h,N}(S)| > 2\eta r\}\\
&\qquad + P\left\{r \sup_V |\Delta_{h,N}(S)| - \eta r^2 + \varepsilon \geqslant -\frac{\eta r^2}{2}\right\} + o(1)\\
&\quad\leqslant 2P\left\{\sup_V |\Delta_{h,N}(S)| > \frac{\eta r}{2} - \frac{\varepsilon}{r}\right\} + o(1).
\end{aligned}$$

Let $\xi > 0$ and $\gamma > 0$. For large r, $\exp(-\eta r^2/2) < \gamma$ and

$$\limsup_{h,N} P\left\{\sup_V |\Delta_{h,N}(S)| > \frac{\eta r}{2} - \frac{\varepsilon}{r}\right\} < \frac{\xi}{3}.$$

Then,

$$\limsup_{h,N} P\{\sup_{U_2,V} Z_{h,N}(S,u) \geqslant \gamma\} \leqslant \xi.$$

If we define $U_3 = \{u; |u| \geqslant \delta(hN)^{1/2}\}$ and $H_1 = \{y; |y| \geqslant \delta\}$, for $g > 0$,

$$\begin{aligned}
&\limsup_{h,N} P\{\sup_{U_3, V} Z_{h,N}(S, u) \geqslant \gamma\} \\
&\quad = \limsup_{h,N} P\{\sup_{H_1, V} [Y_{h,N}(S, \theta_0 + y) - Y_{h,N}(S, \theta_0)] \\
&\quad \geqslant (hN)^{-1} \log \gamma\} \\
&\quad \leqslant \limsup_{h,N} P\left\{\sup_{H_1, V} [Y_{h,N}(S, \theta_0 + y) - Y_{h,N}(S, \theta_0)\right. \\
&\quad \left. - Y_{h,N}(S_0, \theta_0 + y) + Y_{h,N}(S_0, \theta_0)] \geqslant \frac{g}{2}\right\} \\
&\quad + \limsup_{h,N} P\left\{\sup_{H_1} [Y_{h,N}(S_0, \theta_0 + y) - Y_{h,N}(S_0, \theta_0)\right. \\
&\quad \left. - Y(S_0, \theta_0 + y) + Y(S_0, \theta_0)] \geqslant \frac{g}{2}\right\} \\
&\quad + \limsup_{h,N} P\{\sup_{H_1} Y(S_0, \theta_0 + y) - Y(S_0, \theta_0) \\
&\quad \geqslant (hN)^{-1} \log \gamma - g\}.
\end{aligned}$$

If g is small, the third term in the right-hand side is zero, the first term tends to zero from Lemma 1 and the second term tends to zero from Lemma 2. Therefore, for $\gamma > 0$ and $\xi > 0$, if r is large,

$$\limsup_{h,N} P\{\sup_{|u| \geqslant r, V} Z_{h,N}(S, u) > \gamma\} \leqslant \xi. \quad \blacksquare$$

Acknowledgments

The author wishes to thank the referees and the editor for their valuable comments.

References

[1] Arnold, L., and Kliemann, W. (1987). On unique ergodicity for degenerate diffusions. *Stochastics* **21** 41–61.

[2] Dacunha-Castelle, D., and Florens-Zmirou, D. (1986). Estimation of the coefficients of a diffusion from discrete observation. *Stochastics* **19** 263–284.

[3] Florens-Zmirou, D. (1989). Approximate discrete-time schemes for statistics of diffusion processes. *Statistics* **20**(4) 547–557.

[4] Ibragimov, I. A., and Has'minskii, R. Z. (1972). Asymptotic behavior of statistical estimators in the smooth case. *Theory Probab. Appl.* **17** 443–460.

[5] IBRAGIMOV, I. A., AND HAS'MINSKII, R. Z. (1973). Asymptotic behavior of some statistical estimators. II. Limit theorem for the a posteriori density and Bayes estimators. *Theory Probab. Appl.* **18** 76–91.

[6] IBRAGIMOV, I. A., AND HAS'MINSKII, R. Z. (1981). *Statistical Estimation.* Springer-Verlag, New York/Berlin.

[7] INAGAKI, N. AND OGATA, Y. (1975). The weak convergence of likelihood ratio random fields and applications. *Ann. Inst. Statist. Math.* **27** 391–419.

[8] KUTOYANTS, YU. A. (1977). Estimation of the drift coefficient parameter of a diffusion in the smooth case. *Theory Probab. Appl.* **22** 399–406.

[9] KUTOYANTS, YU. A. (1978). Estimation of a parameter of a diffusion processes. *Theory Probab. Appl.* **23** 641–649.

[10] KUTOYANTS, YU. A. (1984). *Parameter Estimation for Stochastic Processes.* Translated and edited by B.L.S. Prakasa Rao. Heldermann.

[11] LIPTSER, R. S., AND SHIRYAYEV, A. N. (1977). *Statistics of random processes II,* Springer-Verlag, New York/Berlin.

[12] PRAKASA RAO, B. L. S. (1983). Asymptotic theory for non-linear least square estimator for diffusion processes, *Math. Operationsforsch. Statist. Ser. Statist.,* Vol. 14, pp. 195–209, Berlin.

[13] PRAKASA RAO, B. L. S. (1988). Statistical inference from sampled data for stochastic processes. *Contemporary Mathematics,* Vol. 80, 249–284. Amer. Math. Soc., Providence, RI.

[14] YOSHIDA, N. (1990). Asymptotic behavior of M-estimator and related random field for diffusion process. *Ann. Inst. Statist. Math.* **42**(2) 221–251.