Probab. Theory Relat. Fields 92, 275–311 (1992)

Probability Theory and Related Fields

Asymptotic expansions of maximum likelihood estimators for small diffusions via the theory of Malliavin–Watanabe*

Nakahiro Yoshida

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

Received October 29, 1990; in revised form November 12, 1991

Summary. The asymptotic expansions of the probability distributions of statistics for the small diffusion are derived by means of the Malliavin calculus. From this the second order efficiency of the maximum likelihood estimator is proved.

Mathematics Subject Classifications (1980): 62E20, 62F12, 62M05

1 Introduction

Let X be a diffusion process defined by the stochastic differential equation

$$(1.1)\qquad \begin{aligned} dX_t &= V_0(X_t, \theta)dt + \varepsilon V(X_t)dw_t,\ t\in[0, T],\ \varepsilon\in(0, 1] \\ X_0 &= x_0\ , \end{aligned}$$

where a k-dimensional unknown parameter $\theta\in\Theta$: a bounded convex domain of $\mathbf{R}^k$, T is a fixed value, x_0 is a constant, $V=(V_1, \ldots, V_r)$ is an $\mathbf{R}^d\otimes\mathbf{R}^r$-valued smooth function defined on $\mathbf{R}^d$, V_0 is an $\mathbf{R}^d$-valued smooth function defined on $\mathbf{R}^d\times\Theta$ with bounded x-derivatives and w is an r-dimensional standard Wiener process. The unknown parameter θ requires to be estimated from the observation $\{X_t;\ 0\leqq t\leqq T\}$. We are interested in the asymptotic properties of the maximum likelihood estimator $\hat{\theta}_\varepsilon$ of θ when $\varepsilon\to 0$. The maximum likelihood estimator is consistent, asymptotically normal and efficient in the first order, see Kutoyants [6].

The notions of the second order efficiency of estimators were introduced by R.A. Fisher, C.R. Rao [11], [12], Takeuchi–Akahira, and Ghosh–Subramanyam [2] mainly for independent observations. Akahira–Takeuchi [1] mentions to an autoregressive process. Taniguchi [13], [14] studied for Gaussian ARMA processes. When we consider Takeuchi–Akahira's criterion of the second order efficiency of estimators, the required mathematical tools are the asymptotic expansions of the probability distributions of the estimators and some related statistics.

For the small diffusion we will prove the asymptotic expansions of probability distributions by means of the Malliavin calculus, extending Yoshida [18]. For this

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

theory see Malliavin [9], [10], Watanabe [16], Kusuoka–Strook [5], Ikeda and Watanabe [3], [4]. Watanabe [15] introduced the notion of the generalized Wiener functional, the pull-back of Schwartz distribution under Wiener mappings, and in his celebrated work [17] he exploited the asymptotic expansion of the generalized Wiener functionals in some Sobolev space and derived from this various expansion formulas for heat kernels. In the present paper this method will prove to be useful to show quite directly the asymptotic expansions for statistical estimators.

In the next section, we present the results used later on. Section 3 gives expansion formulas for the bias corrected maximum likelihood estimator and the likelihood ratio statistic. In Section 4 we present non-degeneracy of the Wiener functionals and we will derive the asymptotic expansions of the probability distributions of statistics considering the composite functions of the indicator functions together with those statistical Wiener functionals. In the last section the second order efficiency of the maximum likelihood estimator is proved.

2 Fundamental results

Let (W, P) be the r-dimensional Wiener space and let H be the Cameron-Martin subspace of W endowed with the inner product

$$\langle h_1, h_2 \rangle = \int_0^T \dot{h}_{1,t} \cdot \dot{h}_{2,t} \, dt$$

for $h_1, h_2 \in H$. For a Hilbert space E, $\|\cdot\|_p$ denotes the $L^p(E)$-norm of E-valued Wiener functional, i.e., for Wiener functional $f:(W, P) \to E$

$$\|f\|_p^p = \int_W |f|_E^p \, P(dw) .$$

Let L be the Ornstein-Uhlenbeck operator and define $\|f\|_{p,s}$ for E-valued Wiener functional f, $s \in \mathbf{R}$, $p \in (1, \infty)$ by

$$\|f\|_{p,s} = \|(I-L)^{s/2} f\|_p .$$

The Banach space $D_p^s(E)$ is the completion of the totality $P(E)$ of E-valued polynomials on the Wiener space (W, P) with respect to $\|\cdot\|_{p,s}$. Let $D^\infty(E)$ be the set of Wiener test functionals of Watanabe [16]:

$$D^\infty(E) = \bigcap_{s>0} \bigcap_{1<p<\infty} D_p^s(E) .$$

Then,

$$D^{-\infty}(E) = \bigcup_{s>0} \bigcup_{1<p<\infty} D_p^{-s}(E)$$

and

$$\tilde{D}^{-\infty}(E) = \bigcup_{s>0} \bigcap_{1<p<\infty} D_p^{-s}(E)$$

are the spaces of generalized Wiener functionals. Moreover, let

$$\tilde{D}^{\infty}(E) = \bigcap_{s>0} \bigcup_{1<p<\infty} D_p^s(E) .$$

We suppress $\mathbf{R}$ when $E = \mathbf{R}$. The Fréchet space $S(\mathbf{R}^d)$ is the totality of rapidly decreasing smooth functions on $\mathbf{R}^d$ and $S'(\mathbf{R}^d)$ is its dual. The space C_{2k}, $k = 0, \ \pm 1, \ \pm 2, \dots$ is the completion of $S(\mathbf{R}^d)$ with respect to the norm $\|u\|_{2k} = \sup_x |A^k u(x)|$, where $A = 1 + |x|^2 - \frac{1}{2}\Delta$. We owe the following theorem to S. Watanabe.

Theorem 2.1 *Let $F \in D^{\infty}(\mathbf{R}^d)$ and $\xi \in D^{\infty}$. Let ψ: $\mathbf{R} \to \mathbf{R}$ be a smooth function such that $0 \leqq \psi(x) \leqq 1$ for $x \in \mathbf{R}$, $\psi(x) = 1$ for $|x| \leqq \frac{1}{2}$ and $\psi(x) = 0$ for $|x| \geqq 1$. Suppose that for any $p \in (1, \infty)$, the Malliavin covariance σ_F of F satisfies*

$$E[1_{\{|\xi| \leqq 1\}} (\det \sigma_F)^{-p}] < \infty . \tag{2.1}$$

Then, there exists a linear mapping $T \in S'(\mathbf{R}^d) \to \hat{T} \in \tilde{D}^{-\infty}$ satisfying the following conditions:
(1) if $T \in S(\mathbf{R}^d)$ then $\hat{T} = \psi(\xi) T(F) \in D^{\infty}$,
(2) for $k = 0, 1, \dots$ and $p \in (1, \infty)$ there exists a constant $C(p, k)$ such that

$$\|\hat{T}\|_{p,-2k} \leqq C(p, k) \, \|T\|_{-2k}$$

for $T \in C_{-2k}$. This mapping is uniquely determined.

Proof. Let $T \in S(\mathbf{R}^d)$. Using integration by parts formula, which is applicable by the condition (2.1), we see that for $J \in D^{\infty}$,

$$\begin{aligned} |\langle \psi(\xi) T(F), J \rangle| &= |\langle \psi(\xi) A^k A^{-k} T(F), J \rangle| \\ &= |\langle A^{-k} T(F), \Phi(w; J) \rangle| \leqq C \|T\|_{-2k} \|J\|_{q,2k} \end{aligned}$$

for a smooth functional $\Phi(w; J)$, any $q > 1$ and some $C > 0$. So that for $k = 0, 1, \dots$ and $p > 1$ there exists $C(p, k) > 0$ such that

$$\|\psi(\xi) T(F)\|_{p,-2k} \leqq C(p, k) \|T\|_{-2k}$$

for $T \in S(\mathbf{R}^d)$. Defining $\hat{T}$ as the continuous extension of $\psi(\xi) T(F)$ we have the result. □

If F is nondegenerate in the usual sense of Malliavin,

$${}_{D^{-\infty}}\langle \hat{T}, J \rangle_{D^{\infty}} = {}_{D^{-\infty}}\langle T(F), \psi(\xi) J \rangle_{D^{\infty}}$$

for $J \in D^{\infty}$. Thus $\hat{T}$ is denoted by $\psi(\xi) T \circ F$ or $\psi(\xi) \, T(F)$ if there is no confusion.

Let us consider a family of E-valued Wiener functionals (or generalized Wiener functionals) $\{F_\varepsilon(w)\}$, $\varepsilon \in (0, 1)$. For $k > 0$ if

$$\limsup_{\varepsilon \downarrow 0} \frac{\|F_\varepsilon\|_{p,s}}{\varepsilon^k} < \infty \ ,$$

we say $F_\varepsilon(w) = O(\varepsilon^k)$ in $D_p^s(E)$ as $\varepsilon \downarrow 0$. It is said that $F_\varepsilon(w) \in D^{\infty}(E)$ has the asymptotic expansion

$$F_\varepsilon(w) \sim f_0 + \varepsilon f_1 + \cdots$$

in $D^\infty(E)$ as $\varepsilon\downarrow 0$ with $f_0, f_1, \ldots \in D^\infty(E)$ if for every $p>1$, $s>0$ and $k=1,2,\ldots$

$$F_\varepsilon(w)-(f_0+\varepsilon f_1+\cdots+\varepsilon^{k-1}f_{k-1})=O(\varepsilon^k)$$

in $D_p^s(E)$ as $\varepsilon\downarrow 0$. We say that $F_\varepsilon(w)\in\tilde{D}^\infty(E)$ has the asymptotic expansion

$$F_\varepsilon(w)\sim f_0+\varepsilon f_1+\cdots$$

in $\tilde{D}^\infty(E)$ as $\varepsilon\downarrow 0$ with $f_0, f_1, \ldots \in \tilde{D}^\infty(E)$ if for every $s>0$ and $k=1,2,\ldots$ there exists $p>1$ such that $F_\varepsilon(w), f_0, f_1, \ldots \in D_p^s(E)$ and

$$F_\varepsilon(w)-(f_0+\varepsilon f_1+\cdots+\varepsilon^{k-1}f_{k-1})=O(\varepsilon^k)$$

in $D_p^s(E)$ as $\varepsilon\downarrow 0$. Moreover, we say that $F_\varepsilon(w)\in D^{-\infty}(E)$ has the asymptotic expansion

$$F_\varepsilon(w)\sim f_0+\varepsilon f_1+\cdots$$

in $D^{-\infty}(E)$ as $\varepsilon\downarrow 0$ with $f_0, f_1, \ldots \in D^{-\infty}(E)$ if for every $k=1,2,\ldots$ there exist $p>1$ and $s>0$ such that $F_\varepsilon(w), f_0, f_1, \ldots \in D_p^{-s}(E)$ and

$$F_\varepsilon(w)-(f_0+\varepsilon f_1+\cdots+\varepsilon^{k-1}f_{k-1})=O(\varepsilon^k)$$

in $D_p^{-s}(E)$ as $\varepsilon\downarrow 0$. Similarly, we say that $F_\varepsilon(w)\in\tilde{D}^{-\infty}(E)$ has the asymptotic expansion

$$F_\varepsilon(w)\sim f_0+\varepsilon f_1+\cdots$$

in $\tilde{D}^{-\infty}(E)$ as $\varepsilon\downarrow 0$ with $f_0, f_1, \ldots \in \tilde{D}^{-\infty}(E)$ if for every $k=1,2,\ldots$ there exists $s>0$ such that for every $p>1$ $F_\varepsilon(w), f_0, f_1, \ldots \in D_p^{-s}(E)$ and

$$F_\varepsilon(w)-(f_0+\varepsilon f_1+\cdots+\varepsilon^{k-1}f_{k-1})=O(\varepsilon^k)$$

in $D_p^{-s}(E)$ as $\varepsilon\downarrow 0$. The generalized means of these expansions yield the ordinary asymptotic expansions.

The following theorem will be our fundamental tool.

Theorem 2.2 *Let ψ be a function defined in Theorem 2.1. Let Λ be an index set. Suppose that families $\{F_\varepsilon(w);\ \varepsilon\in(0,1]\}\subset D^\infty(\mathbf{R}^d)$, $\{\xi_\varepsilon(w);\ \varepsilon\in(0,1]\}\subset D^\infty$ and $\{T_\lambda;\ \lambda\in\Lambda\}\subset S'(\mathbf{R}^d)$ satisfy the following conditions.*

1) *For any $p\in(1,\infty)$*

$$\sup_{\varepsilon\in(0,1]} E[1_{\{|\xi_\varepsilon|\leq 1\}}(\det\sigma_{F_\varepsilon})^{-p}]<\infty\ .$$

2) *$\{F_\varepsilon(w);\ \varepsilon\in(0,1]\}$ has the asymptotic expansion*

$$F_\varepsilon(w)\sim f_0+\varepsilon f_1+\cdots \quad \text{in } D^\infty(\mathbf{R}^d) \text{ as } \varepsilon\downarrow 0$$

with $f_i\in D^\infty(\mathbf{R}^d)$.

3) *$\{\xi_\varepsilon(w);\ \varepsilon\in(0,1]\}$ has the asymptotic expansion in D^∞ as $\varepsilon\downarrow 0$.*

4) *For any $n=1,2,\ldots,$*

$$\lim_{\varepsilon\downarrow 0}\varepsilon^{-n}P\left\{|\xi_\varepsilon|>\frac{1}{2}\right\}=0\ .$$

5) *For any* $n = 1, 2, \ldots$, *there exists a nonnegative integer* m *such that* $A^{-m} T_\lambda \in C_b^n(\mathbf{R}^d)$ *for all* $\lambda \in \Lambda$ *and*

$$\sup_{\lambda \in \Lambda} \sum_{|\mathbf{n}| \leqq n} \| \partial^{\mathbf{n}} A^{-m} T_\lambda \|_\infty < \infty ,$$

where $\mathbf{n} = (n_1, \ldots, n_d)$ *is a multi-index,* $|\mathbf{n}| = n_1 + \cdots + n_d$, $\partial^{\mathbf{n}} = \partial_1^{n_1} \ldots \partial_d^{n_d}$, $\partial_i = \partial/\partial x^i$, $i = 1, \ldots, d$. *Then the composite functional* $\psi(\xi_\varepsilon) T_\lambda(F_\varepsilon) \in \tilde{D}^{-\infty}$ *is well-defined and has the asymptotic expansion*

$$\psi(\xi_\varepsilon) T_\lambda(F_\varepsilon) \sim \Phi_{\lambda,0} + \varepsilon \Phi_{\lambda,1} + \cdots \text{ in } \tilde{D}^{-\infty} \text{ as } \varepsilon \downarrow 0$$

uniformly in $\lambda \in \Lambda$ *with* $\Phi_{\lambda,0}, \Phi_{\lambda,1}, \ldots \in \tilde{D}^{-\infty}$ *determined by the formal Taylor expansion*

$$\begin{aligned} T_\lambda(f_0 + [\varepsilon f_1 + \varepsilon^2 f_2 + \cdots]) &= \sum_{\mathbf{n}} \frac{1}{\mathbf{n}!} \partial^{\mathbf{n}} T_\lambda(f_0) [\varepsilon f_1 + \varepsilon^2 f_2 + \cdots]^{\mathbf{n}} \\ &= \Phi_{\lambda,0} + \varepsilon \Phi_{\lambda,1} + \cdots , \end{aligned}$$

where $\mathbf{n}! = n_1! \ldots n_d!$, $a^{\mathbf{n}} = a_1^{n_1} \ldots a_d^{n_d}$ *for* $a \in \mathbf{R}^d$. *In particular,*

$$\Phi_{\lambda,0} = T_\lambda(f_0) ,$$

$$\Phi_{\lambda,1} = \sum_{i=1}^d f_1^i \, \partial_i T_\lambda(f_0) ,$$

$$\Phi_{\lambda,2} = \sum_{i=1}^d f_2^i \, \partial_i T_\lambda(f_0) + \frac{1}{2} \sum_{i,j=1}^d f_1^i f_1^j \, \partial_i \partial_j T_\lambda(f_0), \ldots .$$

Proof. We follow the proof of Theorem 2.3 of Watanabe [17]. Fix $k = 1, 2, \ldots$ arbitrarily. For $T_\lambda \in S'(\mathbf{R}^d)$, there exists a positive integer $m = m(\Lambda, k)$ and $\phi_\lambda(x) \in C_b^k(\mathbf{R}^d)$ such that $T_\lambda = A^m \phi_\lambda$ for all $\lambda \in \Lambda$. For $J \in D^\infty$,

$${}_{D^{-\infty}}\langle \psi(\xi_\varepsilon) T_\lambda(F_\varepsilon), J \rangle_{D^\infty} = E[\phi_\lambda(F_\varepsilon) l_\varepsilon(J)]$$

by the integration by parts formula, where

$$l_\varepsilon(J) = \sum_{i=0}^{2m} \langle P_{\varepsilon,i}, D^i J \rangle_{H \otimes \cdots \otimes H}$$

and $P_{\varepsilon,i}$ is a polynomial in F_ε, $\psi(\xi_\varepsilon)$, $\gamma(\varepsilon) = \sigma_{F_\varepsilon}^{-1}$, L-derivatives of F_ε and their H-derivatives. Here we used the condition 1). From the regularity of ϕ_λ,

$$\phi_\lambda(F_\varepsilon) = \sum_{|\mathbf{n}| \leqq k-1} \frac{1}{\mathbf{n}!} \partial^{\mathbf{n}} \phi_\lambda(f_0) [F_\varepsilon - f_0]^{\mathbf{n}} + V_{k,\lambda,\varepsilon} ,$$

then for $p' \in (1, \infty)$, there exists a constant c_1 independent of $\lambda \in \Lambda$ and $\varepsilon \in (0, 1]$ such that

$$\| V_{k.\lambda,\varepsilon} \|_{p'} \leqq c_1 \varepsilon^k$$

for $\varepsilon \in (0, 1]$. Let q' satisfy $1/p' + 1/q' = 1$ and let $q > q'$. Then

$$\| l_\varepsilon(J) \|_{q'} \leqq c_2 \| J \|_{q, 2m}$$

for all $\varepsilon \in (0, 1]$, $J \in D^{\infty}$, where c_2 is a constant independent of $\lambda \in \Lambda$, $\varepsilon \in (0, 1]$ and $J \in D^{\infty}$. Therefore,

$$\begin{aligned} |E[V_{k,\lambda,\varepsilon} l_\varepsilon(J)]| &\leqq \| V_{k,\varepsilon,\lambda} \|_{p'} \| l_\varepsilon(J) \|_{q'} \\ &\leqq c_1 c_2 \| J \|_{q, 2m} \varepsilon^k \end{aligned}$$

for $\varepsilon \in (0, 1]$ and $J \in D^{\infty}$.
On the other hand, as $[F_\varepsilon - f_0]^{\mathbf{n}} l_\varepsilon(J)$ has the asymptotic expansion,

$$\sum_{|\mathbf{n}| \leqq k-1} \frac{1}{\mathbf{n}!} \partial^{\mathbf{n}} \phi_\lambda(f_0) [F_\varepsilon - f_0]^{\mathbf{n}} l_\varepsilon(J) = Z_{\lambda,0} + \varepsilon Z_{\lambda,1} + \cdots + \varepsilon^{k-1} Z_{\lambda,k-1} + U_{\lambda,\varepsilon,k},$$

where

$$Z_{\lambda,i} = \sum_{j=0}^{2m} \langle Q_{\lambda,i,j}, D^j J \rangle_{H \otimes \cdots \otimes H}, \quad i = 0, \ldots, k-1,$$

$Q_{\lambda,i,j}$ is a $H \otimes \cdots \otimes H$-valued polynomial in $\partial^{\mathbf{n}} \phi_\lambda(f_0), f_0, f_1, \ldots, \psi_0, \psi_1, \ldots$ $(\psi(\xi_\varepsilon) \sim \psi_0 + \varepsilon \psi_1 + \cdots)$, $\gamma_0, \gamma_1, \ldots (\gamma(\varepsilon) \sim \gamma_0 + \varepsilon \gamma_1 + \cdots)$, L-derivatives of F_ε and their H-derivatives; and

$$|E[U_{\lambda,\varepsilon,k}]| \leqq c_3 \varepsilon^k \| J \|_{q, 2m},$$

where c_3 is a constant independent of λ, ε and J. Let

$$\Phi_{\lambda,i} = \sum_j D^{*j} Q_{\lambda,i,j}, \tag{2.2}$$

then

$$\begin{aligned} &|{}_{D^{-\infty}}\langle \psi(\xi_\varepsilon) T_\lambda(F_\varepsilon), J \rangle_{D^\infty} - {}_{D^{-\infty}}\langle \textstyle\sum_{i=0}^{k-1} \varepsilon^i \Phi_{\lambda,i}, J \rangle_{D^\infty}| \\ &\leqq |E[U_{\lambda,\varepsilon,k}]| + |E[V_{k,\lambda,\varepsilon} l_\varepsilon(J)]| \\ &\leqq (c_1 c_2 + c_3) \varepsilon^k \| J \|_{q, 2m}. \end{aligned}$$

By the duality,

$$\| \psi(\xi_\varepsilon) T_\lambda(F_\varepsilon) - \sum_{i=0}^{k-1} \varepsilon^i \Phi_{\lambda,i} \|_{p, -2m} \leqq (c_1 c_2 + c_3) \varepsilon^k.$$

Here we take p such that $1/p + 1/q = 1$. $p \in (1, \infty)$ is arbitrary, so that

$$\psi(\xi_\varepsilon) T_\lambda(F_\varepsilon) \sim \Phi_{\lambda,0} + \varepsilon \Phi_{\lambda,1} + \cdots \quad \text{in } \tilde{D}^{-\infty}$$

with $\Phi_{\lambda,0}, \Phi_{\lambda,1}, \ldots \in \tilde{D}^{-\infty}$ uniformly in $\lambda \in \Lambda$. Define the mapping $T \in S'(\mathbf{R}^d) \to \Phi_i(T) \in D^{-\infty}$ as in (2.2). Again by the argument of duality, we see that for positive integer m', there exist $s > 0$, $p \in (1, \infty)$ and $C > 0$ such that $\| \Phi_i(T) \|_{p,-s} \leqq C \| T \|_{-2m'}$ for $T \in C_{-2m'}$. For $\phi \in S(\mathbf{R}^d)$ let $\Phi'_0, \Phi'_2, \ldots$ be the coefficients defined by the (formal) Taylor expansion of $\phi(f_0 + [F_\varepsilon - f_0])$. Then,

$$\begin{aligned} \psi(\xi_\varepsilon) \phi(F_\varepsilon) &\sim \psi(\xi_\varepsilon)(\Phi'_0 + \varepsilon \Phi'_1 + \cdots) \\ &\sim \Phi'_0 + \varepsilon \Phi'_1 + \cdots \end{aligned}$$

in D^{∞}. The last equivalence is by the fact that for any $p \in (1, \infty)$ and any $s > 0$,

$$\| 1 - \psi(\xi_\varepsilon) \|_{p,s} = O(\varepsilon^n)$$

for $n = 1, 2, \ldots$. Therefore, $\Phi_i(\phi) = \Phi'_i$, $i = 0, 1, \ldots$. By continuity we have the result. □

Lemma 2.1 *For $n = 1, 2, \ldots$, there exists a positive integer m such that*

$$\sup_{B \in \mathbf{B}^d} \sum_{|\mathbf{n}| \leqq n} \| \partial^{\mathbf{n}} A^{-m} I_B \|_\infty < \infty ,$$

where $\mathbf{B}^d$ denotes the Borel σ-field of $\mathbf{R}^d$.

Proof. The operator A^{-1} is an integral operator

$$A^{-1} f(x) = \int_{\mathbf{R}^d} A(x, y) f(y) \mathrm{d}y ,$$

with kernel

$$A(x, y) = \int_0^\infty \mathrm{e}^{-t} p_1(t, x, y) \mathrm{d}t ,$$

where $p_1(t, x, y)$ is the transition density of a Brownian motion corresponding to the forward equation

$$\frac{\partial p_1}{\partial t} = \frac{1}{2} \Delta p_1 - |y|^2 p_1 ,$$

see pp. 474–475 of Ikeda and Watanabe [4]. We note that for $B \in \mathbf{B}^d$

$$\partial^{\mathbf{n}} A^{-m} I_B(x) = \int_B \partial_x^{\mathbf{n}} A^{-m} \delta_y(x) \mathrm{d}y .$$

Integrability of $\partial_x^{\mathbf{n}} A^{-m} \delta_y(x)$ follows from a direct estimate using above representation of A^{-1}. Then, we see that

$$\int_{\mathbf{R}^d} | \partial_x^{\mathbf{n}} A^{-m} \delta_y(x) | \mathrm{d}y \leqq \frac{C}{(1 + |x|)^p}$$

for some m and $p > 0$, which completes the proof. □

The composite function of a measurable function and a Wiener functional has a usual meaning.

Lemma 2.2 *For ψ, ξ, F given in Theorem 2.1 and any measurable function $f(x)$ of polynomial growth order,*

$$\psi(\xi) f \circ F = \psi(\xi) f(F)$$

in $\tilde{D}^{-\infty}$.

Proof. For simplicity let $d = 1$. For any measurable function $f(x)$ of polynomial growth order and any $\varepsilon > 0$, there exist $k \in \mathbf{N}$ and $\phi \in S(\mathbf{R}^1)$ such that

$$\sup_{x \in \mathbf{R}^1} |A^{-k} f(x) - \phi(x)| < \varepsilon$$

and hence

$$\| f - A^k \phi \|_{-2k} < \varepsilon .$$

In fact, this follows from the inequality

$$|A^{-1}f(x)| \leqq \left[\int_0^\infty e^{-t}dt \int_{-\infty}^\infty p_1(t,x,y)dy\right]^{1/2}$$

$$\times\left[\int_0^\infty e^{-t}dt \int_{-\infty}^\infty p_1(t,x,y)f^2(y)\,dy\right]^{1/2}$$

$$\leqq C(1+|x|)^{-1}\left[\int_{-\infty}^\infty \frac{1}{\sqrt{2}}e^{-\sqrt{2}|y|}f^2(x-y)\,dy\right]^{1/2}.$$

Therefore, there exist a sequence $\phi_n(x) \in S(\mathbf{R}^1)$ and some $k > 0$ such that $\phi_n \to f$ in C_{-2k}. So we have

$$\psi(\xi)f \circ F = \psi(\xi)f(F)$$

in $\tilde{D}^{-\infty}$ by definition of composite functionals. □

3 Expansion of statistics

Let $X^{\varepsilon,\theta}$ be the solution of the stochastic differential equation (1.1) for θ. Let $P_{\varepsilon,\theta}$ be the induced measure from P on $C([0, T], \mathbf{R}^d)$ by the mapping $w \to X^{\varepsilon,\theta}_\cdot(w)$. The Radon-Nikodym derivative of $P_{\varepsilon,\theta}$ with respect to P_{ε,θ_0} is given by the formula (e.g., Liptser and Shiryayev [8])

$$\Lambda_\varepsilon(\theta; X)\Lambda_\varepsilon(\theta_0; X)^{-1},$$

where

$$\Lambda_\varepsilon(\theta; X) = \exp\left\{\int_0^T \varepsilon^{-2}\, V_0'(VV')^+\,(X_t, \theta)dX_t - \frac{1}{2}\int_0^T \varepsilon^{-2}\, V_0'(VV')^+\, V_0(X_t, \theta)dt\right\}.$$

Here A^+ denotes the Moore-Penrose generalized inverse matrix of matrix A. We assume that $V_0(x, \theta) - V_0(x, \theta_0) \in M\{V(x)\}$: the linear manifold generated by column vectors of $V(x)$, for each x, θ and θ_0. When $\theta_0 \in \Theta$ is true, the maximum likelihood estimator $\hat{\theta}_\varepsilon(w; \theta_0)$ is defined by

$$\Lambda_\varepsilon(\hat{\theta}_\varepsilon(w; \theta_0); X^{\varepsilon,\theta_0}(w)) = \max_{\theta\in\Theta} \Lambda_\varepsilon(\theta; X^{\varepsilon,\theta_0}(w)).$$

Next, we prepare several notations. Fix any $\theta_0 \in \Theta$ and X_t^ε denotes $X_t^{\varepsilon,\theta_0}$. Let X_t^0 be the solution of the ordinary differential equation

$$\frac{dX_t^0}{dt} = V_0(X_t^0, \theta_0), \quad t \in [0, T],$$

$$X_0^0 = x_0.$$

Let an $\mathbf{R}^d \otimes \mathbf{R}^d$-valued process $Y_t^\varepsilon(w)$ be the solution of the stochastic differential equation

$$\mathrm{d}Y_t^\varepsilon = \partial V_0(X_t^{\varepsilon,\theta_0}, \theta_0)\, Y_t^\varepsilon \mathrm{d}t + \varepsilon \sum_{\alpha=1}^{r} \partial V_\alpha(X_t^{\varepsilon,\theta_0})\, Y_t^\varepsilon \,\mathrm{d}w_t^\alpha, \quad t \in [0, T],$$

$$Y_0^\varepsilon = I_d,$$

where $[\partial V_\alpha]^{ij} = \partial_j V_\alpha^i$, $\partial_j = \partial/\partial x^j$, $i, j = 1, \ldots, d$, $\alpha = 0, 1, \ldots r$. Then, $Y_t := Y_t^0$ is a deterministic $\mathbf{R}^d \otimes \mathbf{R}^d$-valued process. For a function $f(x, \theta)$ we abbreviate

$$f_t^\varepsilon(\theta) = f(X_t^\varepsilon, \theta),$$

$$\partial_i f_t^\varepsilon(\theta) = \partial_i f(X_t^\varepsilon, \theta)$$

and similarly $\delta_j f_t^\varepsilon$, $\delta_j \delta_l f_t^\varepsilon(\theta)$, $\partial_i \delta_j f_t^\varepsilon(\theta), \ldots$, where $\delta_j = \partial/\partial\theta^j$. It is known that $\varepsilon \to X_t^{\varepsilon,\theta_0}$ is smooth. In particular, $D_t := \left.\dfrac{\partial X_t^{\varepsilon,\theta_0}}{\partial \varepsilon}\right|_{\varepsilon=0}$ satisfies the stochastic differential equation

$$\mathrm{d}D_t = \partial V_{0,t}^0(\theta_0) D_t \mathrm{d}t + V_t^0 \mathrm{d}w_t, \; t \in [0, T],$$

$$D_0 = 0.$$

Then, D_t is represented by

$$D_t = \int_0^t Y_t Y_s^{-1} V_s^0 \,\mathrm{d}w_s, \quad t \in [0, T].$$

Put

$$\begin{aligned} g_s &= Y_s^{-1} V_s^0 V_s^{0\prime} (V_s^0 V_s^{0\prime})^+ \delta V_{0,s}^0(\theta_0) \\ &= Y_s^{-1} \delta V_{0,s}^0(\theta_0), \end{aligned}$$

where $[\delta a(\theta)]^{jl} = \partial a^j(\theta)/\partial\theta^l$, $j = 1, \ldots, d$, $l = 1, \ldots, k$, for $a(\theta) = (a^1(\theta), \ldots, a^d(\theta))'$. The Fisher information matrix $I(\theta_0) = (I_{ij}(\theta_0))$ is defined by

$$I_{ij}(\theta_0) = \int_0^T \delta_i V_{0,t}^0(\theta_0)' (VV')_t^{+0} \delta_j V_{0,t}^0(\theta_0) \mathrm{d}t$$

for $i, j = 1, \ldots, k$. Let

$$m((i_1, \ldots, i_p)/0) = \int_0^T [\delta_{i_1} \ldots \delta_{i_p} V_0'(VV')^+ V]_t^0(\theta_0) \mathrm{d}w_t,$$

$$m((i_1, \ldots, i_p)/1) = \int_0^T \sum_{l=1}^{d} [\partial_l \{\delta_{i_1} \ldots \delta_{i_p} V_0'(VV')^+ V\}]_t^0(\theta_0) D_t^l \mathrm{d}w_t,$$

$$n((i_1, \ldots, i_p)(j_1, \ldots, j_q)/0) = \int_0^T [\delta_{i_1} \ldots \delta_{i_p} V_0'(VV')^+ \delta_{j_1} \ldots \delta_{j_q} V_0]_t^0(\theta_0) \mathrm{d}t,$$

and

$$\begin{aligned} n((i_1, \ldots, i_p)(j_1, \ldots, j_q)/1) = \int_0^T \sum_{l=1}^{d} [\partial_l \{\delta_{i_1} \ldots \delta_{i_p} V_0'(VV')^+ \\ \delta_{j_1} \ldots \delta_{j_q} V_0\}]_t^0(\theta_0) D_t^l \mathrm{d}t. \end{aligned}$$

Let $\delta^\nu = \delta_1^{\nu_1}\delta_2^{\nu_2}\ldots\delta_k^{\nu_k}$ and let $|\nu| = \nu_1 + \nu_2 + \cdots + \nu_k$ for $\nu = (\nu_1, \nu_2, \ldots, \nu_k)$. In this article we assume the following conditions.
(1) V_0, V and $(VV')^+$ are smooth in (x, θ).
(2) There exists a constant C such that

$$|V_0(x, \theta)| \leqq C(1 + |x|)$$

for all x and θ.
(3) For $|\mathbf{n}| \geqq 1$, $F = V_0, V, (VV')^+$, $\sup_{x,\theta}|\partial^{\mathbf{n}} F| < \infty$.
(4) For $|\nu| \geqq 1$ and $|\mathbf{n}| \geqq 0$, a constant $C_{\nu,\mathbf{n}}$ exists and

$$\sup_\theta |\partial^{\mathbf{n}}\delta^\nu V_0| \leqq C_{\nu,\mathbf{n}}(1 + |x|)^{C_{\nu,\mathbf{n}}},$$

for all x.
(5) $I(\theta)$, $\theta \in \Theta$, are positive definite. For $\theta_0 \in \Theta$, there exists $a_0 > 0$ such that

$$\int_0^T [V_{0,t}^0(\theta) - V_{0,t}^0(\theta_0)]'(VV')_t^{+0}[V_{0,t}^0(\theta) - V_{0,t}^0(\theta_0)]\,\mathrm{d}t \geqq a_0|\theta - \theta_0|^2$$

for $\theta \in \Theta$.

Remark 3.1. To derive the results in this article, we can relax the above conditions. By large-deviation argument, they can be replaced by a certain set of regularity conditions about V_0 and V near the neighborhood of the path of X_t^0.

Let

$$l_\varepsilon(\theta; \theta_0) = \log \Lambda_\varepsilon(\theta; X^{\varepsilon,\theta_0}).$$

For matrix $A = (A^{ij})$, let $|A|^2 = \sum_{i,j}|A^{ij}|^2$. Choose γ so that $0 < \gamma < \frac{1}{2}$. Let $m_0, n_0 \in \mathbf{N}$ satisfy $m_0 > \frac{k}{2n_0} + 2$. The Sobolev space $W^{m_0,2n_0}(\Theta)$ is the Banach space endowed with the norm

$$\|u\|_{W^{m_0,2n_0}(\Theta)} = \left(\sum_{|\mathbf{n}| \leqq m_0} \|\delta^{\mathbf{n}} u\|_{L_{2n_0}(\Theta)}^{2n_0}\right)^{1/2n_0},$$

where $L_p(\Theta)$ denotes the L_p space on Θ with respect to the Lebesgue measure. Then, by Sobolev's lemma, the inclusion $W^{m_0,2n_0}(\Theta) \subset C_b^2(\Theta)$ is continuous, that is, there exists a positive constant $C(m_0, 2n_0, 2, \Theta)$ such that

$$\|u\|_{C_b^2(\Theta)} \leqq C(m_0, 2n_0, 2, \Theta)\|u\|_{W^{m_0,2n_0}(\Theta)}$$

for $u \in W^{m_0,2n_0}(\Theta)$, where

$$\|u\|_{C_b^2(\Theta)} = \sum_{|\mathbf{n}| \leqq 2} \sup_{\theta \in \Theta} |\delta^{\mathbf{n}} u(\theta)|.$$

For $\theta_0 \in \Theta$ let

$$Q_1^{jl}(w, \varepsilon, \theta) = \int_0^T \delta_j \bar{V}_{0,t}^\varepsilon(\theta)'(VV')_t^{+\varepsilon}\delta_l \bar{V}_{0,t}^\varepsilon(\theta)\mathrm{d}t,$$

$j, l = 1, \ldots, k$, where

$$\delta\bar{V}_{0,t}^\varepsilon(\theta) = \int_0^1 \delta V_{0,t}^\varepsilon(\theta_0 + u(\theta - \theta_0))\mathrm{d}u$$

and let $Q_1(w, \varepsilon, \theta) = (Q_1^{jl}(w, \varepsilon, \theta))$. Moreover, let

$$Q_2(w, \varepsilon, \theta) = \int_0^T [V_{0,t}^{\varepsilon}(\theta) - V_{0,t}^{\varepsilon}(\theta_0)]'(VV')_t^{+\varepsilon}[V_{0,t}^{\varepsilon}(\theta) - V_{0,t}^{\varepsilon}(\theta_0)]\,dt\,.$$

When $\varepsilon = 0$, $Q_1(w, 0, \theta)$ and $Q_2(w, 0, \theta)$ are deterministic and denoted by $Q_1(0, \theta)$ and $Q_2(0, \theta)$, respectively. For $\eta \in (0, 1)$, define

$$f_1^{jl}(w, \varepsilon, \eta, \theta) = Q_1^{jl}(w, \eta\varepsilon, \theta) - Q_1^{jl}(0, \theta)\,,$$

$$f_2(w, \varepsilon, \eta, \theta) = Q_2(w, \eta\varepsilon, \theta) - Q_2(0, \theta)$$

and

$$f_3(w, \varepsilon, \eta, \theta) = \varepsilon^{1-2\gamma} \int_0^T [V_{0,t}^{\eta\varepsilon}(\theta) - V_{0,t}^{\eta\varepsilon}(\theta_0)]'(VV')_t^{+\eta\varepsilon} V_t^{\eta\varepsilon}\,dw_t\,.$$

For $\theta_0 \in \Theta$ and $c > 0$ let

$$r_{\varepsilon,\eta}^c(w) = c \sum_{j,l=1}^{k} \| f_1^{jl}(w, \varepsilon, \eta, \cdot)\|_{W^{m_0,2n_0}(\Theta)}^{2n_0} + c \sum_{i=2}^{3} \| f_i(w, \varepsilon, \eta, \cdot)\|_{W^{m_0,2n_0}(\Theta)}^{2n_0}\,.$$

For each $\varepsilon \in (0, 1]$, $c > 0$ and $w \in W$, the function $\eta \to r_{\varepsilon,\eta}^c(w)$, $\eta \in (0, 1)$, is smooth. Let

$$R_\varepsilon^c(w) = \| r_{\varepsilon,\cdot}^c(w)\|_{W^{1,2}((0,1))}^2\,.$$

There exists a positive constant $C(1, 2, 0, (0, 1))$ such that

$$\|u\|_{C_b((0,1))} \leqq C(1, 2, 0, (0, 1))\|u\|_{W^{1,2}((0,1))}$$

for $u \in W^{1,2}((0, 1))$, where $\|u\|_{C_b((0,1))} = \sup_{\eta \in (0,1)} |u(\eta)|$. Let

$$C_0 = C(m_0, 2n_0, 2, \Theta)\, C(1, 2, 0, (0, 1))^{1/2n_0}\,.$$

There exists $d_1 > 0$ such that $\{\theta;\, |\theta - \theta_0| \leqq d_1\} \subset \Theta$,

$$\sup_{|\theta - \theta_0| \leqq d_1} |Q_1(0, \theta) - I(\theta_0)| < \frac{1}{2}\lambda_1$$

and

$$\sup_{|\theta - \theta_0| \leqq d_1} \left|\frac{1}{2}\delta^2 Q_2(0, \theta) - I(\theta_0)\right| < \frac{1}{4}\lambda_1\,,$$

where $\lambda_1 = \inf_{|\xi|=1} \xi' I(\theta_0)\xi$, the minimum eigenvalue of $I(\theta_0)$.

Lemma 3.1 *Let c_0 satisfy $C_0 c_0^{-p_1} < \min\{\frac{1}{8k}\lambda_1, \frac{1}{4}a_0 d_1^2\}$, where $p_1 = \frac{1}{2n_0}$. Suppose $c > c_0$. Then,*
(1) If $\varepsilon^\gamma < d_1$ and $R_\varepsilon^c(w) < 1$, a maximum likelihood estimate $\hat{\theta}_{\varepsilon'}(w; \theta_0)$ exists and is in $\{\theta;\, |\theta - \theta_0| < \varepsilon'^\gamma\}$ for $\varepsilon' \leqq \varepsilon$.
(2) If $\varepsilon^\gamma < d_1$ and $R_\varepsilon^c(w) < 1$,

$$\sup_{\substack{|\xi|=1 \\ |\theta - \theta_0| \leqq d_1}} \xi'\delta^2 \varepsilon'^2 l_{\varepsilon'}(\theta; \theta_0)\xi \leqq -\frac{1}{2}\lambda_1\,,$$

for $\varepsilon' \leqq \varepsilon$.
(3) If $\varepsilon^\gamma < d_1$ and $R_\varepsilon^c(w) < 1$, the maximum likelihood estimate $\hat{\theta}_{\varepsilon'}(w; \theta_0)$ is a unique solution in $\{\theta;\, |\theta - \theta_0| < d_1\}$ of the equation $\delta l_{\varepsilon'}(\theta; \theta_0) = 0$ for $\varepsilon' \leqq \varepsilon$.

(4) *Let* $\psi_\varepsilon(w) = \psi(R_\varepsilon^{2c}(w))$ *if* $\varepsilon^\gamma < d_1$ *and let* $\psi_\varepsilon(w) = 0$ *if* $\varepsilon^\gamma \geqq d_1$. *Then,* $0 \leqq \psi_\varepsilon(w) \leqq 1$ *and* $\psi_\varepsilon(w) \in D^\infty$.
(5) $\hat{\theta}_\varepsilon(w; \theta_0)$ *can be extended to a functional on* W *and* $\psi_\varepsilon(w)\hat{\theta}_\varepsilon(w; \theta_0) \in D^\infty(\mathbf{R}^k)$.
(6) *For any* $n = 1, 2, \ldots, \psi_\varepsilon(w) = 1 - O(\varepsilon^n)$ *in* D^∞ *as* $\varepsilon \downarrow 0$.

Proof. (1) Suppose $\varepsilon^\gamma < d_1$ and $R_\varepsilon^c(w) < 1$. By definition of $R_\varepsilon^c(w)$ and Sobolev's inequality, we have

$$\sup_{\varepsilon' \leqq \varepsilon} \| Q_1^{jl}(w, \varepsilon', \cdot) - Q_1^{jl}(0, \cdot) \|_{C_b^2(\Theta)} \leqq C_0 c^{-p_1},$$

$$\sup_{\varepsilon' \leqq \varepsilon} \| Q_2(w, \varepsilon', \cdot) - Q_2(0, \cdot) \|_{C_b^2(\Theta)} \leqq C_0 c^{-p_1}$$

and

$$\sup_{\varepsilon' \leqq \varepsilon} \left\| \varepsilon'^{1-2\gamma} \int_0^T [V_{0,t}^{\varepsilon'}(\cdot) - V_{0,t}^{\varepsilon'}(\theta_0)]'(VV')_t^{+\varepsilon'} V_t^{\varepsilon'} \, dw_t \right\|_{C_b^2(\Theta)} \leqq C_0 c^{-p_1}.$$

Since $c > c_0$, from Condition (5), we see that

$$\begin{aligned}
\inf_{\substack{d_1 < |\theta - \theta_0| \\ \varepsilon' \leqq \varepsilon}} \varepsilon'^{-2\gamma} Q_2(w, \varepsilon', \theta) &\geqq \inf_{\substack{d_1 < |\theta - \theta_0| \\ \varepsilon' \leqq \varepsilon}} \varepsilon'^{-2\gamma} [Q_2(0, \theta) - C_0 c^{-p_1}] \\
&\geqq \varepsilon^{-2\gamma} [a_0 d_1^2 - C_0 c^{-p_1}] \\
&\geqq \frac{3}{4} a_0 d_1^2
\end{aligned}$$

and

$$\begin{aligned}
\inf_{\substack{\varepsilon'^\gamma \leqq |\theta - \theta_0| \leqq d_1 \\ \varepsilon' \leqq \varepsilon}} \varepsilon'^{-2\gamma} Q_2(w, \varepsilon', \theta) &= \inf_{\substack{\varepsilon'^\gamma \leqq |\theta - \theta_0| \leqq d_1 \\ \varepsilon' \leqq \varepsilon}} \varepsilon'^{-2\gamma} (\theta - \theta_0)' Q_1(w, \varepsilon', \theta)(\theta - \theta_0) \\
&\geqq \inf_{\substack{|\xi| = 1, \varepsilon' \leqq \varepsilon \\ \varepsilon'^\gamma \leqq |\theta - \theta_0| \leqq d_1}} \xi' Q_1(w, \varepsilon', \theta) \xi \\
&\geqq \inf_{\substack{|\xi| = 1 \\ \varepsilon'^\gamma \leqq |\theta - \theta_0| \leqq d_1}} \xi' Q_1(0, \theta) \xi - k C_0 c^{-p_1} \\
&\geqq \frac{1}{2} \lambda_1 - k C_0 c^{-p_1} \\
&\geqq \frac{3}{8} \lambda_1 .
\end{aligned}$$

Therefore,

$$\inf_{\substack{\varepsilon'^\gamma \leqq |\theta - \theta_0| \\ \varepsilon' \leqq \varepsilon}} \varepsilon'^{-2\gamma} Q_2(w, \varepsilon', \theta) \geqq \min\left\{ \frac{3}{4} a_0 d_1^2, \frac{3}{8} \lambda_1 \right\} =: C_1 .$$

Then,

$$\sup_{\substack{\varepsilon'^{\gamma} \leqq |\theta-\theta_0| \\ \theta\in\Theta,\, \varepsilon' \leqq \varepsilon}} \varepsilon'^{2(1-\gamma)}[l_{\varepsilon'}(\theta;\theta_0) - l_{\varepsilon'}(\theta_0;\theta_0)] \leqq C_0 c^{-p_1} - \frac{1}{2}C_1$$

$$< \max\left\{-\frac{1}{8}a_0 d_1^2,\ -\frac{1}{16}\lambda_1\right\} < 0\,.$$

Therefore, if $\varepsilon^{\gamma} < d_1$ and $R_\varepsilon^c(w) < 1$, there exists a maximum likelihood estimate $\hat{\theta}_{\varepsilon'}(w;\theta_0) \in \{\theta;\ |\theta-\theta_0| < \varepsilon'^{\gamma}\}$ for $\varepsilon' \leqq \varepsilon$. This proves (1).

(2) If $\varepsilon^{\gamma} < d_1$ and $R_\varepsilon^c(w) < 1$,

$$\sup_{\substack{|\theta-\theta_0| \leqq d_1 \\ \varepsilon' \leqq \varepsilon}} |\varepsilon'^2 \delta^2 l_{\varepsilon'}(\theta;\theta_0) + I(\theta_0)|$$

$$\leqq 2C_0 c^{-p_1} + \sup_{|\theta-\theta_0|\leqq d_1} \left| -\frac{1}{2}\delta^2 Q_2(0,\theta) + I(\theta_0)\right|$$

$$\leqq \frac{1}{2}\lambda_1\,.$$

Consequently,

$$\sup_{\substack{|\xi|=1,\, \varepsilon'\leqq\varepsilon \\ |\theta-\theta_0|\leqq d_1}} \xi'\delta^2\varepsilon'^2 l_{\varepsilon'}(\theta;\theta_0)\xi \leqq -\frac{1}{2}\lambda_1\,.$$

(3) Since $\theta \to l_{\varepsilon'}(\theta;\theta_0)$ is smooth, $\hat{\theta}_{\varepsilon'}(w;\theta_0)$ is a root of the estimating equation $\delta l_{\varepsilon'}(\theta;\theta_0) = 0$. From (2) we see that this root is a unique solution in $\{|\theta-\theta_0| \leqq d_1\}$.

(4) is easy.

(5) Let $\hat{\theta}_\varepsilon(w;\theta_0) = \theta_0$ if $\varepsilon^{\gamma} \geqq d_1$ or $\varepsilon^{\gamma} < d_1$ and $R_\varepsilon^c(w) \geqq 1$. Suppose $\varepsilon^{\gamma} < d_1$. For $h \in H$, if $R_\varepsilon^c(w) < 1$, there exists $t_0(w,h) > 0$ such that $R_\varepsilon^c(w+th) < 1$ for $|t| < t_0(w,h)$. Then, $|\hat{\theta}_\varepsilon(w+th;\theta_0) - \theta_0| < \varepsilon^{\gamma}$ and

$$\delta l_\varepsilon(\hat{\theta}_\varepsilon(w+th;\theta_0);\theta_0)[w+th] = 0\,.$$

By the uniform non-degeneracy of the bilinear forms $\varepsilon^2\delta^2 l_\varepsilon(\theta;\theta_0)$ and the Taylor formula, we have

$$\begin{aligned}
|\hat{\theta}_\varepsilon(w+th;\theta_0) - \hat{\theta}_\varepsilon(w;\theta_0)| &\leqq 2\lambda_1^{-1}|\varepsilon^2\delta l_\varepsilon(\hat{\theta}_\varepsilon(w+th;\theta_0);\theta_0)[w]| \\
&= 2\lambda_1^{-1}|\varepsilon^2\delta l_\varepsilon(\hat{\theta}_\varepsilon(w+th;\theta_0);\theta_0)[w] \\
&\quad - \varepsilon^2\delta l_\varepsilon(\hat{\theta}_\varepsilon(w+th;\theta_0);\theta_0)[w+th]| \\
&\leqq 2\lambda_1^{-1}\sup_{\theta\in\Theta}|\varepsilon^2\delta l_\varepsilon(\theta;\theta_0)[w] - \varepsilon^2\delta l_\varepsilon(\theta;\theta_0)[w+th]|\,. \qquad (3.1)
\end{aligned}$$

Let H_1 be any bounded set in $(H, \|\cdot\|_H)$. We define the process $X_s^* = \sup_{u\leqq s}|X_u|$ for process X. From the stochastic differential equation (1.1), we have for $h \in H_1$,

$$\begin{aligned}
X_t^\varepsilon(w+h) - X_t^0 = \int_0^t \varepsilon V(X_s^0)\dot{h}_s\,ds &+ \int_0^t \{[V_0(X_s^\varepsilon(w+h),\theta_0) - V_0(X_s^0,\theta_0)] \\
&+ \varepsilon[V(X_s^\varepsilon(w+h)) - V(X_s^0)]\dot{h}_s\}\,ds \\
&+ \int_0^t \varepsilon V(X_s^\varepsilon(w+h))\,dw_s\,.
\end{aligned}$$

By Lemma 3.2 (1) below we have

$$\sup_{\substack{\varepsilon\in(0,1)\\ h\in H_1}} \|X_T^{\varepsilon *}(w+h)\|_{2p} < \infty \tag{3.2}$$

for $p \geqq 1$.

Next, let $Z_s^\varepsilon(t,\eta,h) = X_s^\varepsilon(w+t\eta h) - X_s^\varepsilon(w)$ for $s\in[0,T]$, $\varepsilon\in(0,1)$, $t,\eta\in(0,1)$ and $h\in H$. Then, $Z_s^\varepsilon(t,\eta,h)$ satisfies

$$\begin{aligned} Z_s^\varepsilon(t,\eta,h) = &\int_0^s \varepsilon t\eta V(X_u^\varepsilon(w+t\eta h))\dot{h}_u \mathrm{d}u \\ &+ \int_0^s [V_0(X_u^\varepsilon(w+t\eta h),\theta_0) - V_0(X_u^\varepsilon(w),\theta_0)]\,\mathrm{d}u \\ &+ \varepsilon\int_0^s [V(X_u^\varepsilon(w+t\eta h)) - V(X_u^\varepsilon(w))]\,\mathrm{d}w_u\,. \end{aligned}$$

Using Lemma 3.2 (1) below we have for $p\in\mathbf{N}$, $\geqq 2$,

$$\|Z_T^{\varepsilon *}(t,\eta,h)\|_p \leqq C_1|t\varepsilon\eta|$$

for some $C_1>0$ depending on Conditions (2), (3), d, r, p, x_0, T and $\|h\|_H$. Moreover let

$$W_s^\varepsilon(t,\eta,h) = \frac{\partial}{\partial\eta} X_s^\varepsilon(w+t\eta h)\,,$$

$s\in[0,T]$, $\varepsilon\in(0,1)$, $t,\eta\in(0,1)$ and $h\in H$. Then, $W_s^\varepsilon(t,\eta,h)$ satisfies the stochastic integral equation

$$\begin{aligned} W_s^\varepsilon(t,\eta,h) = &\int_0^s t\varepsilon V(X_u^\varepsilon(w+t\eta h))\dot{h}_u\,\mathrm{d}u \\ &+ \int_0^s \Bigg[\sum_i \partial_i V_0(X_u^\varepsilon(w+t\eta h),\theta_0)[W_u^\varepsilon(t,\eta,h)]^i \\ &+ t\varepsilon\eta \sum_i \partial_i V(X_u^\varepsilon(w+t\eta h))[W_u^\varepsilon(t,\eta,h)]^i \dot{h}_u\Bigg]\mathrm{d}u \\ &+ \varepsilon\int_0^s \sum_i \partial_i V(X_u^\varepsilon(w+t\eta h))[W_u^\varepsilon(t,\eta,h)]^i\,\mathrm{d}w_u\,. \end{aligned}$$

Again by Lemma 3.2 (1) and (3.2) we have

$$\|W_T^{\varepsilon *}(t,\eta,h)\|_p \leqq C_2 t\varepsilon$$

for $p\in\mathbf{N}$, $p\geqq 2$, where C_2 depends on Conditions (2), (3), d, r, p, x_0, T and $\|h\|_H$. After all we obtain

$$\sup_{\eta\in(0,1)} E \sup_{0\leqq s\leqq T}\left|\left(\frac{\partial}{\partial\eta}\right)^l (X_s^\varepsilon(w+t\eta h) - X_s^\varepsilon(w))\right|^p \to 0 \quad \text{as } t\to 0$$

for $l = 0, 1$, $h \in H$, each ε and all $p \geqq 2$. From this we obtain, using the Burkhölder-Davis-Gundy inequality and (4),

$$\sup_{\substack{\eta \in (0,1) \\ \theta \in \Theta}} E\left|\left(\frac{\partial}{\partial \eta}\right)^{\nu_0} \delta^{\nu}(\delta l_\varepsilon(\theta; \theta_0)[w + t\eta h] - \delta l_\varepsilon(\theta; \theta_0)[w])\right|^p \to 0$$

as $t \to 0$ for $\nu_0, \nu_1, \ldots, \nu_k \geqq 0$, $\nu_0 + \nu_1 + \cdots + \nu_k \leqq 1$, $h \in H$, each ε and all $p > 1$. Consequently,

$$E\left(\int_{(0,1) \times \Theta} \left|\left(\frac{\partial}{\partial \eta}\right)^{\nu_0} \delta^{\nu}(\delta l_\varepsilon(\theta; \theta_0)[w + t\eta h] - \delta l_\varepsilon(\theta; \theta_0)[w])\right|^p d\eta d\theta\right) \to 0$$

as $t \to 0$ for $\nu_0, \nu_1, \ldots, \nu_k \geqq 0$, $\nu_0 + \nu_1 + \cdots + \nu_k \leqq 1$, $h \in H$, each ε and all $p > 1$. By Sobolev's lemma, we see that there exists a sequence $\{t_q\} \downarrow 0$, $q = 1, 2, \ldots$, such that

$$\sup_{\substack{\eta \in (0,1) \\ \theta \in \Theta}} |\delta l_\varepsilon(\theta; \theta_0)[w + t_q \eta h] - \delta l_\varepsilon(\theta; \theta_0)[w]| \to 0, \quad \text{a.s.}$$

as $q \to \infty$ for $h \in H$, each ε. Consequently, we see that the right hand side of (3.1) tends to zero a.s. as $t \to 0$ and obtain

$$\lim_{t \to 0} \hat{\theta}_\varepsilon(w + th; \theta_0) = \hat{\theta}_\varepsilon(w; \theta_0), \; a.s.$$

if $R_\varepsilon^c(w) < 1$.

If $R_\varepsilon^c(w) \geqq 1$, $R_\varepsilon^{2c}(w) > 1$ and hence for any $h \in H$ and some $t_1(w, h) > 0$, $\psi_\varepsilon(w + th) = 0$ if $|t| < t_1(w, h)$. Therefore, $\lim_{t \to 0} \psi_\varepsilon(w + th)\hat{\theta}_\varepsilon(w + th; \theta_0) = \psi_\varepsilon(w)\, \hat{\theta}_\varepsilon(w; \theta_0)$ for $h \in H$, $w \in W$, $\varepsilon \in (0, 1)$. Functionals $\psi_\varepsilon(w)\hat{\theta}_\varepsilon(w; \theta_0)$ are bounded and in $\bigcap_{p > 1} L_p(W, P)$. Next, we calculate H-derivatives of these functionals. For $R_\varepsilon^c(w) < 1$, $h \in H$ and $|t| < t_0(w, h)$,

$$\begin{aligned} &\varepsilon^2 \delta l_\varepsilon(\hat{\theta}_\varepsilon(w + th; \theta_0); \theta_0)[w + th] - \varepsilon^2 \delta l_\varepsilon(\hat{\theta}_\varepsilon(w; \theta_0); \theta_0)[w + th] \\ &= -\varepsilon^2 \delta l_\varepsilon(\hat{\theta}_\varepsilon(w; \theta_0); \theta_0)[w + th] + \varepsilon^2 \delta l_\varepsilon(\hat{\theta}_\varepsilon(w; \theta_0); \theta_0)[w] . \end{aligned}$$

For $i = 1, \ldots, k$, there exist $\tilde{\theta}$ and $\tilde{t}$ such that

$$|\tilde{\theta} - \hat{\theta}_\varepsilon(w; \theta_0)| < |\hat{\theta}_\varepsilon(w + th; \theta_0) - \hat{\theta}_\varepsilon(w; \theta_0)| ,$$

$|\tilde{t}| < |t|$ and

$$\begin{aligned} &\varepsilon^2 \delta \delta_i l_\varepsilon(\tilde{\theta}; \theta_0)[w + th](\hat{\theta}_\varepsilon(w + th; \theta_0) - \hat{\theta}_\varepsilon(w; \theta_0)) \\ &= -t(D_h \varepsilon^2 \delta_i l_\varepsilon)(\hat{\theta}_\varepsilon(w; \theta_0); \theta_0)[w + \tilde{t}h] . \end{aligned}$$

Dividing both sides by t and $t \to 0$ we have

$$D_h \hat{\theta}_\varepsilon(w; \theta_0) = -\,[\varepsilon^2 \delta^2 l_\varepsilon(\hat{\theta}_\varepsilon(w; \theta_0); \theta_0)[w]]^{-1} (D_h \varepsilon^2 \delta l_\varepsilon)(\hat{\theta}_\varepsilon(w; \theta_0); \theta_0)[w]$$

for $h \in H$ if $\varepsilon^\gamma < d_1$ and $R_\varepsilon^c(w) < 1$. Then H-derivative of $\psi_\varepsilon(w)\hat{\theta}_\varepsilon(w; \theta_0)$ exists and

$$D_h[\psi_\varepsilon(w)\hat{\theta}_\varepsilon(w; \theta_0)] = [D_h \psi_\varepsilon(w)]\hat{\theta}_\varepsilon(w; \theta_0) + \psi_\varepsilon(w) D_h \hat{\theta}_\varepsilon(w; \theta_0) .$$

When $R_\varepsilon^c(w) \geqq 1$, $R_\varepsilon^{2c}(w) > 1$ and $\psi_\varepsilon(w + th) = 0$ if $|t| < t_1(w, h)$. Therefore, $D_h[\psi_\varepsilon(w)\hat{\theta}_\varepsilon(w; \theta_0)]$ exists and equals zero. Thus we see that $\psi_\varepsilon(w)\hat{\theta}_\varepsilon(w; \theta_0)$ is H-differentiable on whole W. The H-derivative is in $L_p(W, P)$, $p > 1$. In fact, this follows from uniform non-degeneracy of $\varepsilon^2\delta^2 l_\varepsilon(\theta; \theta_0)$ and the integrability of $\sup_{\theta\in\Theta} |D\varepsilon^2\delta l_\varepsilon(\theta; \theta_0)|_{HS}$, which is obtained by estimating the Sobolev norm

$$E[\| |D\varepsilon^2\delta l_\varepsilon(\cdot\,; \theta_0)[w]|_{HS}^2 \|_{W^{m,n}(\Theta)}^p] \,.$$

For a representation of $D\delta l_\varepsilon(\theta; \theta_0)$ see Section 4. Similarly we can verify existence and integrability of higher order H-derivatives, which completes the proof of (5).
(6) From the following Lemma 3.2 (3) for any $n \in \mathbf{N}$ and $c_1 > 0$, $P(R_\varepsilon^c(w) > c_1) = O(\varepsilon^n)$ as $\varepsilon \downarrow 0$. From this fact we can show (6) in view of the chain rule for H-derivatives. This completes the proof. □

Lemma 3.2. (1) *Let $\xi_\theta^\varepsilon(t)$, $t \in [0, T]$, $\varepsilon \in (0, 1)$, $\theta \in \Theta$ (an index set), be d-dimensional nonanticipative processes given by*

$$\xi_\theta^\varepsilon(t) = \eta_\theta^\varepsilon(t) + \int_0^t \alpha_\theta^\varepsilon(s)\psi(s)\,\mathrm{d}s + \varepsilon\int_0^t \beta_\theta^\varepsilon(s)\,\mathrm{d}w_s\,,$$

where η_θ^ε is an $\mathbf{R}^d$-valued continuous nonanticipative process, $\alpha_\theta^\varepsilon$ is an $\mathbf{R}^d \otimes \mathbf{R}^m$-valued nonanticipative process, β_θ^ε is an $\mathbf{R}^d \otimes \mathbf{R}^r$-valued nonanticipative process, ψ is an $\mathbf{R}^m$-valued function satisfying

$$\int_0^T |\psi(s)|^2\,\mathrm{d}s < \infty$$

and w is an r-dimensional Wiener process. Suppose that there exist positive constants K_i, $i = 1, 2, 3$, such that

$$|\alpha_\theta^\varepsilon(s)| \leqq K_1|\xi_\theta^\varepsilon(s)|$$

and

$$|\beta_\theta^\varepsilon(s)| \leqq K_2(|\xi_\theta^\varepsilon(s)| + K_3)\,,$$

a.s. for any $\varepsilon \in (0, 1)$, $\theta \in \Theta$, $0 \leqq s \leqq T$. Assume that for all $n \in \mathbf{N}$, $\varepsilon \in (0, 1)$ and $\theta \in \Theta$,

$$E\left[\sup_{0\leqq t\leqq T} |\eta_\theta^\varepsilon(t)|^n\right] < \infty\,.$$

Then, for any $n \geqq 2$, there exists positive constant $C_n = C_n(T, d, r, K_1, K_2)$ such that

$$\sup_{\theta\in\Theta} E\left(\sup_{0\leqq t\leqq T} |\xi_\theta^\varepsilon(t)|^n\right) \leqq C_n \exp\left\{C_n\left(\int_0^T |\psi(s)|^2\,\mathrm{d}s\right)^{\frac{n}{2}}\right\} \times\left(E\left(\sup_{0\leqq t\leqq T} |\eta_\theta^\varepsilon(t)|^n\right) + K_3^n\varepsilon^n\right)$$

for $\varepsilon \in (0, 1)$.

(2) *Under Conditions* (1), (2) *and* (3), *for* $n \in \mathbf{N}$ *there exist positive constants* c_1, c_2 *independent of* ε *such that*

$$\sup_{\eta \in (0,1)} E\left(\sup_{0 \leqq t \leqq T} |X_t^{\eta\varepsilon} - X_t^0|^n \right) \leqq c_1 \varepsilon^n$$

and

$$\sup_{\eta \in (0,1)} E\left(\sup_{0 \leqq t \leqq T} \left| \frac{\partial}{\partial \eta} X_t^{\eta\varepsilon} \right|^n \right) \leqq c_2 \varepsilon^n$$

for $\varepsilon \in (0, 1)$.
(3) *For* $n \in \mathbf{N}$, $c > 0$ *and* $a > 0$,

$$P(R_\varepsilon^c(w) > a) = O(\varepsilon^n) .$$

Proof. (1) Let

$$\sigma_N = \inf\{t \geqq 0; |\xi_\theta^\varepsilon(t)| \geqq N\} \wedge T$$

for $N \in \mathbf{N}$. For some $C_0 > 0$

$$|\xi_\theta^\varepsilon(t)|^2 \leqq C_0 \left(|\eta_\theta^\varepsilon(t)|^2 + K_0^2 \int_0^t |\alpha_\theta^\varepsilon(s)|^2 \, ds + \left| \varepsilon \int_0^t \beta_\theta^\varepsilon(s) \, dw_s \right|^2 \right),$$

where $K_0 = (\int_0^T |\psi(s)|^2 \, ds)^{1/2}$. By the Burkhölder-Davis-Gundy inequality

$$E\left(\left| \int_0^{\cdot} \beta_\theta^\varepsilon(s) \, dw_s \right|_{t \wedge \sigma_N}^{*n} \right) \leqq C_1(n, T, K_2, d, r) \left(K_3^n + \int_0^t E|\xi_\theta^\varepsilon(s \wedge \sigma_N)|^n \, ds \right),$$

$t \in [0, T]$, $\varepsilon \in (0, 1)$, $\theta \in \Theta$, for some $C_1(n, T, K_2, d, r) > 0$. Using assumptions we have

$$E(|\xi_\theta^\varepsilon|_{t \wedge \sigma_N}^{*n}) \leqq C_2(1 + K_0^n) \left(E(|\eta_\theta^\varepsilon|_T^{*n}) + \int_0^t E(|\xi_\theta^\varepsilon|_{s \wedge \sigma_N}^{*n}) \, ds + K_3^n \varepsilon^n \right),$$

$t \in [0, T]$, $\varepsilon \in (0, 1)$, $\theta \in \Theta$, for some $C_2 = C_2(n, T, K_1, K_2, d, r) > 0$. By Gronwall's lemma we have

$$E(|\xi_\theta^\varepsilon|_{T \wedge \sigma_N}^{*n}) \leqq e^{C_2(1 + K_0^n) T} C_2(1 + K_0^n) [E(|\eta_\theta^\varepsilon|_T^{*n}) + K_3^n \varepsilon^n] ,$$

$\varepsilon \in (0, 1)$, $\theta \in \Theta$. Letting $N \to \infty$ we have the result.
(2) We know that $\left(X_t^{\eta\varepsilon} - X_t^0, \frac{\partial}{\partial \eta} X^{\eta\varepsilon} \right)$ satisfy the stochastic differential equation

$$\begin{cases} d(X_t^{\eta\varepsilon} - X_t^0) = [V_0(X_t^{\eta\varepsilon}, \theta_0) - V_0(X_t^0, \theta_0)] \, dt + \eta\varepsilon V(X_t^{\eta\varepsilon}) \, dw_t \\ d \dfrac{\partial}{\partial \eta} X_t^{\eta\varepsilon} = \partial V_0(X_t^{\eta\varepsilon}, \theta_0) \dfrac{\partial}{\partial \eta} X_t^{\eta\varepsilon} \, dt + \varepsilon \left[\eta \partial V(X_t^{\eta\varepsilon}) \dfrac{\partial}{\partial \eta} X_t^{\eta\varepsilon} + V(X_t^{\eta\varepsilon}) \right] dw_t \\ X_0^{\eta\varepsilon} - X_0^0 = 0 \\ \dfrac{\partial}{\partial \eta} X_0^{\eta\varepsilon} = 0 . \end{cases}$$

Since $\partial V_0(\cdot, \theta_0)$ and ∂V are bounded and $V_0(\cdot, \theta_0)$ and V are of linear growth order, we can apply (1) to this case and obtain the result.

(3) It suffices to show that for any $m \in \mathbf{N}$ there exists $n \in \mathbf{N}$ such that $E[|R_\varepsilon^c(w)|^n] = O(\varepsilon^m)$ as $\varepsilon \downarrow 0$. We can estimate the norms of the parts of $R_\varepsilon^c(w)$ corresponding to f_1^{jl} and f_2 by (2). For the part corresponding to f_3, we obtain a similar estimate using (2) and the Burkhölder-Davis-Gundy inequality, which completes the proof. □

In the context of the higher order statistical asymptotic theory we need to modify the maximum likelihood estimators to get efficient estimators. We call an estimator $\hat{\theta}_\varepsilon^*$ a bias corrected maximum likelihood estimator if

$$\hat{\theta}_\varepsilon^* = \hat{\theta}_\varepsilon - \varepsilon^2 b(\hat{\theta}_\varepsilon),$$

where $b(\theta)$ is a bounded smooth function. Then, $\psi_\varepsilon(w)\hat{\theta}_\varepsilon^*$ is well-defined. The following lemma gives its expansion formula.

Lemma 3.3 *$\psi_\varepsilon(w)\varepsilon^{-1}(\hat{\theta}_\varepsilon^*(w;\theta_0) - \theta_0) \in D^\infty(\mathbf{R}^k)$ has the asymptotic expansion*

$$\psi_\varepsilon(w)\frac{\hat{\theta}_\varepsilon^*(w;\theta_0) - \theta_0}{\varepsilon} \sim f_0 + \varepsilon f_1 + \cdots \text{ in } D^\infty(\mathbf{R}^k) \quad \text{as } \varepsilon \downarrow 0$$

with $f_0, f_1, \ldots \in D^\infty(\mathbf{R}^k)$. In particular,

$$f_0 = I(\theta_0)^{-1}B,$$

$$f_1 = -b(\theta_0) + \tfrac{1}{2}I^{-1}(\theta_0)\Gamma + \tfrac{1}{2}I^{-1}(\theta_0)QI^{-1}(\theta_0)B, \ldots,$$

where $B = (B^i)$, $\Gamma = (\Gamma^i)$ and $Q = (Q_{i,j})$ are defined as follows.

$$B^i = m((i)/0), \quad i = 1, \ldots, k,$$

$$\Gamma^i = 2m((i)/1), \quad i = 1, \ldots, k,$$

$$Q_{i,j} = \sum_{m=1}^{k} [I(\theta_0)^{-1}B]^m N_{i,j,m} + 2\Delta_{i,j}, \quad i, j = 1, \ldots, k,$$

where

$$N_{i,j,m} = -[n((ji)(m)/0) + n((im)(j)/0) + n((mj)(i)/0)], \quad i, j, m = 1, \ldots, k,$$

$$\Delta_{i,j} = m((ji)/0) - n((j)(i)/1), \quad i, j = 1, \ldots, k.$$

Proof. Suppose $\varepsilon_0^\gamma < d_1$ and $R_{\varepsilon_0}^c(w) < 1$. Let $F(\theta, \varepsilon, w) = \varepsilon^2 \delta l_\varepsilon(\theta; \theta_0)[w]$. The mapping $(\varepsilon, \theta) \to F(\theta, \varepsilon, w)$ is smooth and $A(\theta, \varepsilon, w) := \delta F(\theta, \varepsilon, w)$ is non-singular for $\varepsilon \leqq \varepsilon_0$ and θ satisfying $|\theta - \theta_0| < d_1$ uniformly. Moreover, by Lemma 3.2 (1), it can be proved that for a.s. $w \in W$ and $\varepsilon \geqq 0$,

$$\lim_{u \to 0} \sup_{\theta \in \Theta} |F(\theta, \varepsilon + u, w) - F(\theta, \varepsilon, w)| = 0.$$

Therefore, the mapping $\varepsilon \to \hat{\theta}_\varepsilon(w; \theta_0)$ is continuous. Hence, we have

$$\frac{\partial}{\partial \varepsilon}\hat{\theta}_\varepsilon(w;\theta_0) = -A(\hat{\theta}_\varepsilon(w;\theta_0), \varepsilon, w)^{-1}\frac{\partial F}{\partial \varepsilon}(\hat{\theta}_\varepsilon(w;\theta_0), \varepsilon, w) \tag{3.3}$$

for $\varepsilon \leqq \varepsilon_0$. In particular,

$$f_0 = \lim_{\varepsilon \downarrow 0} \frac{\partial}{\partial \varepsilon} \hat{\theta}^*_\varepsilon(w; \theta_0) = I(\theta_0)^{-1} B .$$

Differentiate (3.3) once more we get $\left(\frac{\partial}{\partial \varepsilon}\right)^2 \hat{\theta}^*_\varepsilon(w; \theta_0)$ and f_1. Similarly, we have $\left(\frac{\partial}{\partial \varepsilon}\right)^i \hat{\theta}^*_\varepsilon(w; \theta_0)$, $i = 3, 4, \ldots$ and their limits in terms of A, F, B, their derivatives with respect to θ and ε, and $\hat{\theta}_\varepsilon(w; \theta_0)$. Therefore, $\hat{\theta}^*_\varepsilon(w; \theta_0)$ is smooth on $[0, \varepsilon_0)$. Then we have the expansion

$$\begin{aligned}\hat{\theta}^*_\varepsilon(w; \theta_0) = \theta_0 &+ \frac{\varepsilon}{1!}(\delta_0)_0 \hat{\theta}^*_\cdot(w; \theta_0) + \frac{\varepsilon^2}{2!}(\delta_0)_0^2 \hat{\theta}^*_\cdot(w; \theta_0) + \cdots \\ &+ \frac{\varepsilon^{j-1}}{(j-1)!}(\delta_0)_0^j \hat{\theta}^*_\cdot(w; \theta_0) \\ &+ \varepsilon^j \int_0^1 \frac{1}{(j-1)!}(1-s)^{j-1}(\delta_0)_{s\varepsilon}^j \hat{\theta}^*_\cdot(w; \theta_0)\, ds\end{aligned}$$

for $\varepsilon \leqq \varepsilon_0$, where $(\delta_0)_\varepsilon$ denotes the ε-derivative at ε. Hence,

$$\begin{aligned}\psi_\varepsilon(w)\hat{\theta}^*_\varepsilon(w; \theta_0) &= \psi_\varepsilon(w)\theta_0 + \psi_\varepsilon(w)\frac{\varepsilon}{1!}(\delta_0)_0 \hat{\theta}^*_\cdot(w; \theta_0) + \psi_\varepsilon(w)\frac{\varepsilon^2}{2!}(\delta_0)_0^2 \hat{\theta}^*_\cdot(w; \theta_0) \\ &+ \cdots + \psi_\varepsilon(w)\varepsilon^j \int_0^1 \frac{1}{(j-1)!}(1-s)^{j-1}(\delta_0)_{s\varepsilon}^j \hat{\theta}^*_\cdot(w; \theta_0)\, ds .\end{aligned}$$

This expansion holds on W. The H-derivatives of the integrands of the residual term take the form of $(1-s)^{j-1} G(\hat{\theta}_{s\varepsilon}(w; \theta_0), w)$, where $G(\theta, w)$ is an $H \otimes \cdots \otimes H$-valued random field on Θ. Estimating the norm

$$E[\| |G(\cdot, w)|_{HS}^2 \|^p_{W^{m_0, 2n_0}(\Theta)}] ,$$

$p > 1$, before substituting $\hat{\theta}_{s\varepsilon}(w; \theta_0)$ for θ, we can show

$$\begin{aligned}\psi_\varepsilon(w)\varepsilon^{-1}(\hat{\theta}^*_\varepsilon(w; \theta_0) - \theta_0) &\sim \psi_\varepsilon(w)\frac{1}{1!}(\delta_0)_0 \hat{\theta}^*_\cdot(w; \theta_0) + \psi_\varepsilon(w)\frac{\varepsilon}{2!}(\delta_0)_0^2 \hat{\theta}^*_\cdot(w; \theta_0) \\ &+ \cdots \sim \frac{1}{1!}(\delta_0)_0 \hat{\theta}^*_\cdot(w; \theta_0) + \frac{\varepsilon}{2!}(\delta_0)_0^2 \hat{\theta}^*_\cdot(w; \theta_0) + \cdots\end{aligned}$$

in $D^\infty(\mathbf{R}^k)$ as $\varepsilon \downarrow 0$ since $\psi_\varepsilon(w) = 1 - O(\varepsilon^n)$ in D^∞ as $\varepsilon \downarrow 0$ for $n \in \mathbf{N}$. This completes the proof. □

Remarks 3.2. (1) Formally, we may write

$$B^i = \lim_{\varepsilon \downarrow 0} \frac{\partial F^i}{\partial \varepsilon}(\hat{\theta}_\varepsilon(w; \theta_0), \varepsilon, w), \quad i = 1, \ldots, k,$$

$$\Gamma^i = \lim_{\varepsilon \downarrow 0} \frac{\partial^2 F^i}{\partial \varepsilon^2}(\hat{\theta}_\varepsilon(w; \theta_0), \varepsilon, w), \quad i = 1, \ldots, k,$$

$$Q_{i,j} = \lim_{\varepsilon \downarrow 0} \frac{d}{d\varepsilon} [A(\hat{\theta}_\varepsilon(w; \theta_0), \varepsilon, w)]^{i,j}$$

$$+ \lim_{\varepsilon \downarrow 0} 2 \frac{\partial^2 F^i}{\partial\varepsilon\partial\theta^j} (\hat{\theta}_\varepsilon(w; \theta_0), \varepsilon, w), \quad i, j = 1, \ldots, k,$$

$$N_{i,j,m} = \lim_{\varepsilon \downarrow 0} \delta_m \delta_j F^i(\hat{\theta}_\varepsilon(w; \theta_0), \varepsilon, w), \quad i, j, m = 1, \ldots, k,$$

$$\Delta_{i,j} = \lim_{\varepsilon \downarrow 0} \frac{\partial^2 F^i}{\partial\varepsilon\partial\theta^j} (\hat{\theta}_\varepsilon(w; \theta_0), \varepsilon, w), \quad i, j = 1, \ldots, k.$$

(2) Kutoyants [7] proved an expansion formula for the maximum likelihood estimator of one dimension. We will need a smooth truncation to apply Malliavin's calculus.
For $h \in \mathbf{R}^k$ we denote the log likelihood ratio

$$l_{\varepsilon,h}(w; \theta_0) = l_\varepsilon(\theta_0 + \varepsilon h; \theta_0) - l_\varepsilon(\theta_0; \theta_0).$$

Then,

$$l_{\varepsilon,h}(w; \theta_0) = \int_0^T \varepsilon^{-1} [V_{0,t}^\varepsilon(\theta_0 + \varepsilon h) - V_{0,t}^\varepsilon(\theta_0)]'(VV')_t^{+\varepsilon} V_t^\varepsilon \mathrm{d}w_t$$

$$-\frac{1}{2}\int_0^T \varepsilon^{-2} [V_{0,t}^\varepsilon(\theta_0 + \varepsilon h) - V_{0,t}^\varepsilon(\theta_0)]'(VV')_t^{+\varepsilon} [V_{0,t}^\varepsilon(\theta_0 + \varepsilon h) - V_{0,t}^\varepsilon(\theta_0)] \, \mathrm{d}t.$$

The following lemma gives the expansion of the likelihood ratio, which is rather easy to show and the proof is omitted.

Lemma 3.4 *The* log *likelihood ratio* $l_{\varepsilon,h}(w; \theta_0) \in D^\infty$ *and has the asymptotic expansion*

$$l_{\varepsilon,h}(w; \theta_0) \sim f_0^L + \varepsilon f_1^L + \cdots \text{ in } D^\infty \quad \text{as } \varepsilon \downarrow 0,$$

with $f_0^L, f_1^L, \ldots \in D^\infty$. *In particular,*

$$f_0^L = h'B - \tfrac{1}{2} h' I(\theta_0) h,$$

$$f_1^L = \sum_{i=1}^k h^i m((i)/1) + \frac{1}{2} \sum_{i,j=1}^k h^i h^j m((ij)/0)$$

$$- \frac{1}{2} \sum_{i,j=1}^k h^i h^j n((i)(j)/1) - \frac{1}{2} \sum_{i,j,m=1}^k h^i h^j h^m n((ij)(m)/0), \ldots.$$

This expansion holds uniformly in any compact set of $\mathbf{R}^k$.

4 Asymptotic expansions of probability

First, we know that for $h \in H$

$$D_h X_t^{\varepsilon,\theta_0} = \varepsilon \int_0^t Y_t^\varepsilon Y_s^{\varepsilon-1} V_s^\varepsilon \dot{h}_s \, \mathrm{d}s.$$

For $h \in H$, the H-derivative of $F(\theta, \varepsilon, w)$ in the direction of h is

$$\begin{aligned} D_h F(\theta, \varepsilon, w) &= \varepsilon \int_0^T [\delta V_0'(VV')^+ V]_t^\varepsilon(\theta) \dot{h}_t \, dt \\ &\quad + \varepsilon \int_0^T \sum_l [\partial_l \{\delta V_0'(VV')^+ V\}]_t^\varepsilon(\theta) D_h X_t^{\varepsilon, \theta_0, l} \, dw_t \\ &\quad - \int_0^T \sum_l [\partial_l \{\delta V_0'(VV')^+ (V_0(\cdot, \theta) - V_0(\cdot, \theta_0))\}]_t^\varepsilon(\theta) D_h X_t^{\varepsilon, \theta_0, l} \, dt \\ &= \varepsilon \int_0^T [\delta V_0'(VV')^+ V]_t^\varepsilon(\theta) \dot{h}_t \, dt + \varepsilon^2 \int_0^T K_{1,s}^\varepsilon(\theta) \dot{h}_s \, ds \\ &\quad - \varepsilon \sum_j (\theta - \theta_0)^j \int_0^T K_{2,j,s}^\varepsilon(\theta) \dot{h}_s \, ds , \end{aligned}$$

where

$$K_{1,s}^\varepsilon(\theta) = \sum_{m,l} \int_s^T [\partial_l \{\delta V_0'(VV')^+ V\}]_t^\varepsilon(\theta) [Y_t^\varepsilon]^{lm} dw_t [Y_s^{\varepsilon -1} V_s^\varepsilon]^{m\cdot}$$

and

$$\begin{aligned} K_{2,j,s}^\varepsilon(\theta) = \sum_l \int_s^T [\partial_l \{\delta V_0(\cdot, \theta)'(VV')^+ \int_0^1 \delta_j V_0 \\ (\cdot, \theta_0 + u(\theta - \theta_0)) \, du\}]_t^\varepsilon [Y_t^\varepsilon]^{l\cdot} dt \, Y_s^{\varepsilon -1} V_s^\varepsilon . \end{aligned}$$

Let $\bar{\psi}_\varepsilon(w) = \psi(R_\varepsilon^{3c}(w))$ for some fixed $c > c_0$. Then, if $\bar{\psi}_\varepsilon(w) > 0$,

$$D_h \psi_\varepsilon(w) \hat{\theta}_\varepsilon(w; \theta_0) = -A(\hat{\theta}_\varepsilon(w; \theta_0), \varepsilon, w)^{-1} D_h F(\hat{\theta}_\varepsilon(w; \theta_0), \varepsilon, w) .$$

Therefore, the Malliavin covariance of $\psi_\varepsilon(w)\varepsilon^{-1}(\hat{\theta}_\varepsilon(w; \theta_0) - \theta_0)$ is

$$\begin{aligned} \sigma_{\psi_\varepsilon \varepsilon^{-1} \hat{\theta}_\varepsilon(w; \theta_0)} &= \langle D\psi_\varepsilon \varepsilon^{-1} \hat{\theta}_\varepsilon(w; \theta_0), D\psi_\varepsilon \varepsilon^{-1} \hat{\theta}_\varepsilon(w; \theta_0) \rangle_H \\ &= A(\hat{\theta}_\varepsilon(w; \theta_0), \varepsilon, w)^{-1} \int_0^T \{[\delta V_0'(VV')^+ V]_s^\varepsilon(\hat{\theta}_\varepsilon(w; \theta_0)) \\ &\quad + \varepsilon K_{1,s}^\varepsilon(\hat{\theta}_\varepsilon(w; \theta_0)) - \sum_j [\hat{\theta}_\varepsilon(w; \theta_0) - \theta_0]^j K_{2,j,s}^\varepsilon(\hat{\theta}_\varepsilon(w; \theta_0))\} \\ &\quad \{\ldots\}' \, ds A(\hat{\theta}_\varepsilon(w; \theta_0), \varepsilon, w)^{-1} , \end{aligned}$$

and this can be denoted by

$$A(\hat{\theta}_\varepsilon(w; \theta_0), \varepsilon, w)^{-1} \{I^\varepsilon(w) + \varepsilon R_1^\varepsilon(w) + R_2^\varepsilon(w)(\hat{\theta}_\varepsilon(w; \theta_0) - \theta_0)\} A(\hat{\theta}_\varepsilon(w; \theta_0), \varepsilon, w)^{-1} ,$$

where

$$I^\varepsilon(w) = \int_0^T [\delta V_0'(VV')^+ \delta V_0]_s^\varepsilon(\hat{\theta}_\varepsilon(w; \theta_0)) \, ds ,$$

$R_2^\varepsilon(w)$ is a 3-liner form and $R_2^\varepsilon(w)(x) = R_2^\varepsilon(w)[x, \cdot, \cdot]$.

The Malliavin covariance of $\psi_\varepsilon \varepsilon^{-1}(\hat{\theta}_\varepsilon^*(w;\theta_0)-\theta_0)$ is given by

$$\sigma_{\psi_\varepsilon \varepsilon^{-1}\hat{\theta}_\varepsilon^*(w;\theta_0)} = (I_k - \varepsilon^2 \delta b(\hat{\theta}_\varepsilon(w;\theta_0)))\sigma_{\psi_\varepsilon \varepsilon^{-1}\hat{\theta}_\varepsilon(w;\theta_0)}(I_k - \varepsilon^2 \delta b(\hat{\theta}_\varepsilon(w;\theta_0)))'$$

if $\bar{\psi}_\varepsilon(w) > 0$.

Let

$$f_0(w,\varepsilon,\theta) = Q_0(w,\varepsilon,\theta) - Q_0(0,\theta)\,,$$

where

$$Q_0(w,\varepsilon,\theta) = \int_0^T [\delta V_0'(VV')^+ \delta V_0]_s^\varepsilon(\theta)\,ds\,.$$

Define

$$R_0^{c',\varepsilon}(w) = c' \|\,|f_0(w,\varepsilon,\cdot)|^2\,\|_{W^{m_1,2n_1}(\Theta)}^{2n_1}$$

for $c' > 0$, $m_1, n_1 \in \mathbf{N}$, $m_1 > \dfrac{k}{2n_1}$.

Let

$$\xi_\varepsilon(w) = |\varepsilon^{1/2} R_1^\varepsilon(w)|^2 + |\varepsilon^{\gamma/2} R_2^\varepsilon(w)|^2 + R_0^{c',\varepsilon}(w) + R_\varepsilon^{3c}(w)\,.$$

Then, $\xi_\varepsilon(w) \in D^\infty$, since we can replace $\hat{\theta}_\varepsilon(w;\theta_0)$ in the right hand side of the representation of $\sigma_{\psi_\varepsilon \varepsilon^{-1}\hat{\theta}_\varepsilon(w;\theta_0)}$ by $\psi_\varepsilon \hat{\theta}_\varepsilon(w;\theta_0)$.

Lemma 4.1 *For some $\varepsilon_0 > 0$ and $c' > 0$,*

$$\sup_{\varepsilon<\varepsilon_0} E[1_{\{|\xi_\varepsilon| \leqq 1\}} (\det \sigma_{\psi_\varepsilon \varepsilon^{-1}\hat{\theta}_\varepsilon^*(w;\theta_0)})^{-p}] < \infty$$

for all $p > 1$.

Proof. If $|\xi_\varepsilon| \leqq 1$, $|\hat{\theta}_\varepsilon(w;\theta_0) - \theta_0| < \varepsilon^\gamma$ and

$$\begin{aligned} |I^\varepsilon(w) - I(\theta_0)| &\leqq |f_0(w,\varepsilon,\hat{\theta}_\varepsilon(w;\theta_0))| + |Q_0(0,\hat{\theta}_\varepsilon(w;\theta_0)) - I(\theta_0)| \\ &\leqq C_0' c'^{-1/4n_1} + \sup_{|\theta-\theta_0|<\varepsilon^\gamma} |Q_0(0,\theta) - Q_0(0,\theta_0)|\,, \end{aligned}$$

where C_0' stems from the Sobolev inequality for Θ. Moreover, if $|\xi_\varepsilon| \leqq 1$,

$$|\varepsilon R_1^\varepsilon(w)| \leqq \varepsilon^{1/2}$$

and

$$|R_2^\varepsilon(w)(\hat{\theta}_\varepsilon(w;\theta_0) - \theta_0)| < \varepsilon^\gamma |R_2^\varepsilon(w)| \leqq \varepsilon^{\frac{1}{2}\gamma}\,.$$

From the proof of Lemma 3.1 (2), we see that if $\varepsilon^\gamma < d_1$ and $R_\varepsilon^c(w) < 1$, the operator norm of $A(\hat{\theta}_\varepsilon(w;\theta_0),\varepsilon,w) \leqq \lambda_k + \frac{1}{2}\lambda_1$, where λ_k is the maximum eigenvalue of $I(\theta_0)$. Thus there exist small ε_0, large c' and some $c'' > 0$ such that

$$\det \sigma_{\psi_\varepsilon \varepsilon^{-1}\hat{\theta}_\varepsilon^*(w;\theta_0)} \geqq c''$$

for $\varepsilon < \varepsilon_0$ if $\xi_\varepsilon(w) \leqq 1$. This completes the proof. $\square$

Lemma 4.2 *$\xi_\varepsilon(w)$ has an asymptotic expansion in D^∞ and for $n \in \mathbf{N}$ and $c_1 > 0$,*

$$P[|\xi_\varepsilon| > c_1] = O(\varepsilon^n)$$

as $\varepsilon \downarrow 0$.

Proof. From Lemma 3.2 (2) we see that for $n \in \mathbf{N}$ and $c_1 > 0$

$$P[R_0^{c',\varepsilon}(w) > c_1] = O(\varepsilon^n)$$

as $\varepsilon \downarrow 0$. Applying Lemma 3.2 (1) to $Y_t^\varepsilon - Y_t$ and $Y_t^{\varepsilon-1} - Y_t^{-1}$ we have

$$E\left[\sup_{0 \leqq t \leqq T} |Y_t^\varepsilon|^n\right] + E\left[\sup_{0 \leqq t \leqq T} |Y_t^{\varepsilon-1}|^n\right] < \infty$$

for all $n \in \mathbf{N}$. For $n \in \mathbf{N}$, $c_1 > 0$ and $\alpha > 0$

$$P\left[\varepsilon^\alpha \sup_{\substack{s \in [0,T] \\ \theta \in \Theta}} |K_{2,j,s}^\varepsilon(\theta)| > c_1\right] = O(\varepsilon^n)$$

as $\varepsilon \downarrow 0$ for $j = 1, \ldots, k$. Applying the Burkhölder-Davis-Gundy inequality to

$$\int_s^T [\partial_l\{\delta V_0'(VV')^+ V\}]_t^\varepsilon(\theta)[Y_t^\varepsilon]^{lm} \mathrm{d}w_t = \int_0^T [\partial_l\{\delta V_0'(VV')^+ V\}]_t^\varepsilon(\theta)[Y_t^\varepsilon]^{lm} \mathrm{d}w_t$$

$$- \int_0^s [\partial_l\{\delta V_0'(VV')^+ V\}]_t^\varepsilon(\theta)[Y_t^\varepsilon]^{lm} \mathrm{d}w_t$$

and its derivatives in θ, we have

$$\sup_{\varepsilon \in (0,1)} E\left[\sup_{s \in [0,T]} \|K_{1,s}^\varepsilon(\cdot)\|_{W^{1,2m}(\Theta)}^n\right] < \infty$$

for $l, m, n \in \mathbf{N}$. By Sobolev's lemma we see that for any $\alpha > 0$, $n \in \mathbf{N}$ and $c_1 > 0$

$$P\left[\varepsilon^\alpha \sup_{s \in [0,T]} \sup_{\theta \in \Theta} |K_{1,s}^\varepsilon(\theta)| > c_1\right] = O(\varepsilon^n)$$

as $\varepsilon \downarrow 0$. Since

$$P[|\xi_\varepsilon| > c_1] \leqq P[|\xi_\varepsilon| > c_1, R_\varepsilon^c(w) < 1] + P[R_\varepsilon^c(w) \geqq 1]$$

and $|\hat{\theta}_\varepsilon(w; \theta_0) - \theta_0| < \varepsilon^\gamma$ if $R_\varepsilon^c(w) < 1$, it is not difficult to show this lemma by Lemma 3.2 (3). □

Let $\phi(x; \mu, \Sigma)$ be the probability density function of the k-dimensional normal distribution with mean μ and covariance matrix Σ. Let $I^{-1}(\theta_0) = I^{-1} = (I^{ij})$. Then, we have the following theorem.

Theorem 4.1 *The probability distribution of the bias corrected maximum likelihood estimator $\hat{\theta}_\varepsilon^*(w; \theta_0)$ has the asymptotic expansion*

$$P\left[\frac{\hat{\theta}_\varepsilon^*(w; \theta_0) - \theta_0}{\varepsilon} \in A\right] \sim \int_A p_0(x)\,\mathrm{d}x + \varepsilon \int_A p_1(x)\,\mathrm{d}x + \cdots \text{ as } \varepsilon \downarrow 0, \quad A \in \mathbf{B}^k.$$

The expansion is uniform in Borel sets $A \in \mathbf{B}^k$. In particular,

$$p_0(x) = \phi(x; 0, I^{-1}),$$

$$p_1(x) = \left[\sum_{i,j,l} I^{ij} A_{i,j,l} x^l + \sum_{i,j,l} I^{ij} B_{i,l,j} x^l - \sum_{j,l} b^j(\theta_0) I_{jl} x^l - \sum_{i,j,l} A_{i,j,l} x^i x^j x^l - \sum_{i,j,l} \frac{1}{2} B_{i,j,l} x^i x^j x^l \right] \phi(x; 0, I^{-1}), \ldots,$$

where

$$A_{i,j,n} = \frac{1}{2} \int_0^T \int_0^t \sum_{l,m=1}^d [\partial_l \{\delta_i V_0'(VV')^+ \delta_j V_0\}]_t^0(\theta_0) Y_t^{lm} g_s^{mn} \, \mathrm{d}s \mathrm{d}t$$

and

$$B_{i,j,l} = n((ij)(l)/0).$$

Proof. Let

$$A_{i,j,n}^* = \int_0^T \int_0^t \sum_{l,m=1}^d [\partial_l \{\delta_i V_0'(VV')^+ V\}]_t^0(\theta_0) [V'(VV')^+ \delta_j V_0]_t^0(\theta_0) Y_t^{lm} g_s^{mn} \, \mathrm{d}s \mathrm{d}t$$

and

$$A_{i,j,n}^{**} = \int_0^T \int_0^t \sum_{l,m=1}^d [\delta_i V_0'(VV')^+ V]_t^0(\theta_0) [\partial_l \{V'(VV')^+ \delta_j V_0\}]_t^0(\theta_0) Y_t^{lm} g_s^{mn} \, \mathrm{d}s \mathrm{d}t .$$

Then, $A_{i,j,n}^* = A_{j,i,n}^{**}$ and

$$A_{i,j,n} = \tfrac{1}{2}(A_{i,j,n}^* + A_{i,j,n}^{**}) .$$

It suffices to verify the conditions of Theorem 2.2 for

$$F_\varepsilon = \psi_\varepsilon \varepsilon^{-1} (\hat{\theta}_\varepsilon^*(w; \theta_0) - \theta_0)$$

and ξ_ε given above. The condition 2) is presented by Lemma 3.3. The condition 1) follows from Lemma 4.1. Conditions 3) and 4) are by Lemma 4.2. Lemma 2.1 gave 5). Therefore, for $A \in \mathbf{B}^k$, we have the asymptotic expansion

$$\psi(\xi_\varepsilon) I_A(\psi_\varepsilon \varepsilon^{-1} (\hat{\theta}_\varepsilon^*(w; \theta_0) - \theta_0)) \sim \Phi_{A,0} + \varepsilon \Phi_{A,1} + \cdots \text{ in } \tilde{D}^{-\infty} \quad \text{as } \varepsilon \downarrow 0$$

uniformly in $A \in \mathbf{B}^k$, where

$$\Phi_{A,0} = I_A(f_0),$$

$$\Phi_{A,1} = \sum_{i=1}^k f_1^i \partial_i I_A(f_0), \ldots .$$

Hence, by Lemmas 4.2, and 2.2,

$$P\left[\frac{\hat{\theta}_\varepsilon^*(w; \theta_0) - \theta_0}{\varepsilon} \in A \right] \sim E\left[\psi(\xi_\varepsilon) I_A\left(\psi_\varepsilon \frac{\hat{\theta}_\varepsilon^*(w; \theta_0) - \theta_0}{\varepsilon} \right) \right]$$

$$= E\left[\psi(\xi_\varepsilon) I_A \circ \left(\psi_\varepsilon \frac{\hat{\theta}_\varepsilon^*(w; \theta_0) - \theta_0}{\varepsilon} \right) \right]$$

$$\sim E[\Phi_{A,0}] + \varepsilon E[\Phi_{A,1}] + \cdots$$

as $\varepsilon \downarrow 0$, uniformly in $A \in \mathbf{B}^k$. Therefore, the rest is to calculate $E[\Phi_{A,i}]$, $i = 0, 1, \ldots$. From the regularity of f_0 and integration by parts formula,

$$E[\Phi_{A,i}] = E[G_i(w) I_A \circ f_0] = E[G_i(w) I_A(f_0)]$$

$$= \int_A E[G_i(w) | f_0 = x] p_{f_0}(x) \, dx$$

for some smooth functional $G_i(w)$. Therefore, each term is represented by an integration of a smooth function. We will only determine p_0 and p_1. p_0 is trivial. Considering $A = \{y = (y^i);\ y^i > x^i, i = 1, \ldots, k\}$, $x = (x^i) \in \mathbf{R}^k$, we see

$$p_1(x) = (-1)^k \partial_1 \cdots \partial_k E\left[\sum_i f_1^i \partial_i I_A(f_0) \right]$$

$$(4.1) \quad = (-1)^k \sum_i \partial_i \cdot \partial_1 \cdots \partial_{i-1} \partial_{i+1} \cdots \partial_k E[f_1^i H_{x_1}(f_0^1) \cdots \delta_{x_i}(f_0^i) \ldots H_{x_k}(f_0^k)]$$

$$= -\sum_i \partial_i E[f_1^i \delta_x(f_0)] = -\sum_i \partial_i \{ E[f_1^i | f_0 = x] p_{f_0}(x) \},$$

where $H_x(y) = 1_{[x,\infty)}(y)$ for $x, y \in \mathbf{R}$. Let

$$a_t = I^{-1} \delta V^0_{0,t}(\theta_0)' (VV')^{+0}_t V^0_t$$

and

$$\gamma^i_{l,t} = [\partial_l \{ \delta_i V_0' (VV')^+ V \}]^0_t (\theta_0).$$

Then, by Lemma 4.3 below we have

$$E[m((i)/1) | f_0 = x] = E\left[\int_0^T \sum_l \gamma^i_{l,t} D^l_t \, dw_t | f_0 = x \right]$$

$$= \sum_l E\left[\int_0^T \gamma^i_{l,t} \int_0^t [Y_t]^{l\cdot} Y_s^{-1} V^0_s \, dw_s \, dw_t \,\middle|\, \int_0^T a_t \, dw_t = x \right]$$

$$= \sum_l \mathrm{Tr}\Bigg(\int_0^T \int_0^t [\delta V_0' (VV')^+ V]^0_t (\theta_0)$$

$$[\partial_l \{ V' (VV')^+ \delta_i V_0 \}]^0_t (\theta_0) [Y_t]^{l\cdot} g_s \, ds \, dt (xx' - I^{-1}) \Bigg)$$

$$= \sum_{j,l} A^*_{i,j,l} (x^j x^l - I^{jl}),$$

$$E[m((ji)/0) | f_0 = x] = E\left[\int_0^T [\delta_j \delta_i V_0' (VV')^+ V]^0_t (\theta_0) \, dw_t \,\middle|\, \int_0^T a_t \, dw_t = x \right]$$

$$= \sum_l \int_0^T [\delta_j \delta_i V_0' (VV')^+ \delta_l V_0]^0_t (\theta_0) dt x^l$$

$$= \sum_l B_{i,j,l} x^l$$

and

$$E[n((i)(j)/1)|f_0 = x] = E\left[\int_0^T \sum_l [\partial_l\{\delta_i V_0'(VV')^+\delta_j V_0\}]_t^0(\theta_0)D_t^l \mathrm{d}t \,|\, f_0 = x\right]$$

$$= E\left[\int_0^T \sum_l [\partial_l\{\delta_i V_0'(VV')^+\delta_j V_0\}]_t^0(\theta_0)[Y_t]^{l\cdot}\left(\int_0^t Y_s^{-1} V_s^0 \mathrm{d}w_s\right)\mathrm{d}t \,\middle|\, \int_0^T a_t \mathrm{d}w_t = x\right]$$

$$= \int_0^T \int_0^t \sum_l [\partial_l\{\delta_i V_0'(VV')^+\delta_j V_0\}]_t^0(\theta_0)[Y_t]^{l\cdot} g_s \mathrm{d}s\mathrm{d}t x$$

$$= \sum_l 2A_{i,j,l}x^l .$$

Hence, the conditional expectations appearing in the right hand side of (4.1) are

$$E[f_1^i \,|\, f_0 = x] = -b^i(\theta_0) + \tfrac{1}{2}E[[I^{-1}\Gamma]^i|f_0 = x] + \tfrac{1}{2}E[[I^{-1}QI^{-1}B]^i|\,f_0 = x]$$

$$= -b^i(\theta_0) + \sum_j I^{ij}E[m((j)/1)|\,f_0 = x]$$

$$\text{(4.2)} \qquad -\tfrac{1}{2}\sum_{j,l,m} I^{ij}x^m x^l[n((jl)(m)/0) + n((lm)(j)/0) + n((jm)(l)/0)]$$

$$+\sum_{j,l} I^{ij}x^l E[m((jl)/0) - n((j)(l)/1)|\,f_0 = x]$$

$$= -\sum_{j,l,m} I^{im}A^{**}_{m,j,l}x^j x^l - \sum_{j,l,m}\frac{1}{2}I^{im}B_{l,j,m}x^j x^l$$

$$-b^i(\theta_0) - \sum_{j,l,m} I^{im}I^{jl}A^{**}_{j,m,l} .$$

Then, the function $p_1(x)$ can be derived from (4.1) and (4.2). This completes the proof. □

Lemma 4.3 *Let w be an r-dimensional Wiener process and let functions a_t, b_t, c_t on $[0,T]$ be deterministic. Let $\Sigma = \int_0^T a_t a_t' \mathrm{d}t$.*

(1) *Let a_t be $\mathbf{R}^k \otimes \mathbf{R}^r$-valued and let b_t be $\mathbf{R}^r$-valued. Then,*

$$E\left[\int_0^T b_t' \mathrm{d}w_t \,\middle|\, \int_0^T a_t \mathrm{d}w_t = x\right] = x'\Sigma^{-1}\int_0^T a_t b_t \mathrm{d}t .$$

(2) *Let a_t, b_t and c_t be $\mathbf{R}^k \otimes \mathbf{R}^r$, $\mathbf{R}^m \otimes \mathbf{R}^r$ and $\mathbf{R}^m \otimes \mathbf{R}^r$-valued, respectively. Then,*

$$E\left[\int_0^T \left(\int_0^t b_s \mathrm{d}w_s\right)' c_t \mathrm{d}w_t \,\middle|\, \int_0^T a_t \mathrm{d}w_t = x\right] = \mathrm{Tr}\int_0^T\int_0^t \Sigma^{-1}a_t c_t' b_s a_s' \Sigma^{-1}(xx' - \Sigma)\,\mathrm{d}s\,\mathrm{d}t .$$

Here Tr *stands for the trace.*

To prove the optimality of the maximum likelihood estimator we need the asymptotic expansion of the likelihood ratio process. As the maximum likelihood estimator we can show the non-degeneracy of the Malliavin covariance of the log likelihood ratio and obtain the following theorem.

Theorem 4.2 *Let $h \in \mathbf{R}^k$ and $h \neq 0$. The probability distribution of the log likelihood ratio $l_{\varepsilon,h}(w;\theta_0)$ has the asymptotic expansion*

$$P[l_{\varepsilon,h}(w;\theta_0) \in A] \sim \int_A p_0^L(x)\,\mathrm{d}x + \varepsilon \int_A p_1^L(x)\,\mathrm{d}x + \cdots, \quad \text{as } \varepsilon \downarrow 0,\ A \in \mathbf{B}^1 .$$

The expansion is uniform in $A \in \mathbf{B}^1$. In particular,

$$p_0^L(x) = \phi(\bar{x}; 0, J),$$

$$p_1^L(x) = \left[\sum_{i,j,l} A_{i,j,l} h^i h^j h^l\right] J^{-3}[\bar{x}^3 - J\bar{x}^2 - 3J\bar{x} + J^2]\phi(\bar{x}; 0, J)$$

$$+ \frac{1}{2}\left[\sum_{i,j,l} B_{i,j,l} h^i h^j h^l\right] J^{-2}[\bar{x}^2 - J\bar{x} - J]\,\phi(\bar{x}; 0, J),$$

where $J = h'I(\theta_0)h$ and $\bar{x} = x + \frac{1}{2}J$. The probability distribution function of $l_{\varepsilon,h}(w;\theta_0)$ has the asymptotic expansion

$$P[l_{\varepsilon,h}(w;\theta_0) \leqq x] \sim \Phi(\bar{x}; 0, J) - \varepsilon\left\{\left[\sum_{i,j,l} A_{i,j,l} h^i h^j h^l\right] J^{-2}[\bar{x}^2 - J\bar{x} - J]\right.$$

$$\left. + \frac{1}{2}\left[\sum_{i,j,l} B_{i,j,l} h^i h^j h^l\right] J^{-1}[\bar{x} - J]\right\}\phi(\bar{x}; 0, J) + \cdots,$$

where $\Phi(x; \mu, \sigma^2)$ is the probability distribution function of the one-dimensional normal distribution with mean μ and variance σ^2.

Proof. Constructing an appropriate smooth truncation functional the non-degeneracy of the Malliavin covariance is proved in the sense of Theorem 2.2 and we have the composite functional with a tempered distribution I_A, the asymptotic expansion of the generalized functional and the expansion for distribution

$$P[l_{\varepsilon,h}(w;\theta_0) \in A] \sim E[\Phi_{A,0}^L] + \varepsilon E[\Phi_{A,1}^L] + \cdots$$

as $\varepsilon \downarrow 0$, $A \in \mathbf{B}^1$, where $\Phi_{A,0}^L = I_A(f_0^L)$, $\Phi_{A,1}^L = f_1^L \partial I_A(f_0^L), \ldots$. The expectation

$$E[\Phi_{A,1}^L] = \int_A p_1^L(x)\,\mathrm{d}x,$$

where

$$p_1^L(x) = -\partial\{E[f_1^L | f_0^L = x]p_{f_0^L}(x)\}.$$

Let

$$\bar{a}_t = h'\delta V_{0,t}^0(\theta_0)'(VV')_t^{+0} V_t^0 .$$

Then, by Lemma 4.3 we have

$$E[m((i)/1)|f_0^L = x] = E\left[\int_0^T \sum_l [\partial\{\delta_i V_0'(VV')^+ V\}]_t^0(\theta_0) D_t^l \, \mathrm{d}w_t \middle| \int_0^T \bar{a}_t \mathrm{d}w_t = \bar{x}\right]$$

$$= \int_0^T \int_0^t h'[\delta V_0'(VV')^+ V]_t^0(\theta_0)[\partial\{V'(VV')^+ \delta_i V_0\}]_t^0(\theta_0) Y_t g_s h \mathrm{d}s\mathrm{d}t$$

$$\cdot J^{-2}[\bar{x}^2 - J]$$

$$= \sum_{j,l} A_{i,j,l}^* h^j h^l J^{-2}(\bar{x}^2 - J),$$

$$E[m((ij)/0)| f_0^L = x] = E\left[\int_0^T [\delta_i\delta_j V_0'(VV')^+ V]_t^0(\theta_0) \mathrm{d}w_t \middle| \int_0^T \bar{a}_t \mathrm{d}w_t = \bar{x}\right]$$

$$= \int_0^T h'[\delta V_0'(VV')^+ V]_t^0(\theta_0)[\delta_i\delta_j V_0'(VV')^+ V]_t^0(\theta_0)' \mathrm{d}t J^{-1}\bar{x}$$

$$= \int_0^T h'[\delta V_0'(VV')^+ \delta_i\delta_j V_0]_t^0(\theta_0) \mathrm{d}t J^{-1}\bar{x}$$

$$= \sum_l B_{i,j,l} h^l J^{-1}\bar{x}$$

and

$$E[n((i)(j)/1)| f_0^L = x] = E\left[\int_0^T \sum_l [\partial_l\{\delta_i V_0'(VV')^+ \delta_j V_0\}]_t^0(\theta_0) D_t^l \mathrm{d}t \middle| \int_0^T \bar{a}_t \mathrm{d}w_t = \bar{x}\right]$$

$$= \int_0^T \int_0^t [\partial\{\delta_i V_0'(VV')^+ \delta_j V_0\}]_t^0(\theta_0) Y_t g_s \mathrm{d}s\mathrm{d}t J^{-1} h\bar{x}$$

$$= \sum_l 2A_{i,j,l} h^l J^{-1}\bar{x}.$$

Then, from Lemma 3.4 we have the conditional expectation

$$E[f_1^L | f_0^L = x] = \sum_i h^i E[m((i)/1)| f_0^L = x] + \frac{1}{2}\sum_{i,j} h^i h^j E[m((ij)/0)| f_0^L = x]$$

$$- \frac{1}{2}\sum_{i,j} h^i h^j E[n((i)(j)/1)| f_0^L = x]$$

$$- \frac{1}{2}\sum_{i,j,m} h^i h^j h^m n((ij)(m)/0)$$

$$= \left[\sum_{i,j,l} A_{i,j,l} h^i h^j h^l\right] J^{-2}[\bar{x}^2 - J\bar{x} - J]$$

$$+ \frac{1}{2}\left[\sum_{i,j,l} B_{i,j,l} h^i h^j h^l\right] J^{-1}[\bar{x} - J].$$

From this equation we obtain $p_1^L(x)$, which completes the proof. □

The statistical theory of the asymptotic efficiency is deeply related to the problem of testing statistical hypotheses. To calculate the power of test statistics requires the asymptotic expansions of distribution functions under contiguous alternatives. The following lemma enables us to transform the expansion of the composite functional into those under contiguous alternatives in the space of the generalized Wiener functionals.

Lemma 4.4 *Let $\psi(\xi_\varepsilon)T_\lambda(F_\varepsilon)$ be the generalized Wiener functional given in Theorem 2.2. Suppose that $\Lambda^\varepsilon(w) \in \tilde{D}^\infty$ has the asymptotic expansion*

$$\Lambda^\varepsilon \sim \Lambda^0(1 + \varepsilon\Psi_1(w) + \varepsilon^2\Psi_2(w) + \cdots) \quad \text{in } \tilde{D}^\infty \text{ as } \varepsilon \downarrow 0 ,$$

with $\Lambda^0\Psi_1, \Lambda^0\Psi_2 \ldots \in \tilde{D}^\infty$. Then, $\Lambda^\varepsilon(w)\psi(\xi_\varepsilon)T_\lambda(F_\varepsilon)$ has the asymptotic expansion

$$\Lambda^\varepsilon(w)\psi(\xi_\varepsilon)T_\lambda(F_\varepsilon) \sim \Lambda^0[\bar{\Phi}_{\lambda,0} + \varepsilon\bar{\Phi}_{\lambda,1} + \cdots], \quad \text{in } D^{-\infty} \text{ as } \varepsilon \downarrow 0,$$

uniformly in $\lambda \in \Lambda$ with $\Lambda^0\bar{\Phi}_{\lambda,0}, \Lambda^0\bar{\Phi}_{\lambda,1}, \ldots \in D^{-\infty}$ determined by the formal Taylor expansion

$$(1 + \varepsilon\Psi_1 + \varepsilon^2\Psi_2 + \cdots)(\Phi_{\lambda,0} + \varepsilon\Phi_{\lambda,1} + \cdots) = \bar{\Phi}_{\lambda,0} + \varepsilon\bar{\Phi}_{\lambda,1} + \cdots .$$

In particular,

$$\bar{\Phi}_{\lambda,0} = \Phi_{\lambda,0} ,$$

$$\bar{\Phi}_{\lambda,1} = \Phi_{\lambda,1} + \Phi_{\lambda,0}\Psi_1, \ldots .$$

Proof. We can show this lemma by Theorem 2.2 (ii) of Watanabe [17]. □

Lemma 4.5. (1) *For $\theta_0 \in \Theta$ and $h \in \mathbf{R}^k$, there exist functionals ϕ_ε^h, $\varepsilon \in (0, 1)$, on $C([0, T], \mathbf{R}^d)$ satisfying the following conditions for any compact set $K \subset \mathbf{R}^k$.*
(i) $0 \leqq \phi_\varepsilon^h(X) \leqq 1$, $X \in C([0, T], \mathbf{R}^d)$.
(ii) $\phi_\varepsilon^h(X^{\varepsilon,\theta_0}(w)) = 1 - O(\varepsilon^n)$ *in D^∞ as $\varepsilon \downarrow 0$ uniformly in $h \in K$ for $n = 1, 2, \ldots$.*
(iii) $\phi_\varepsilon^h(X^{\varepsilon,\theta_0+\varepsilon h}(w)) = 1 - O(\varepsilon^n)$ *in D^∞ as $\varepsilon \downarrow 0$ uniformly in $h \in K$ for $n = 1, 2, \ldots$.*
(iv) *For all $p > 1$,*

$$\sup_{\substack{\varepsilon\in(0,1)\\ h\in K}} E[1_{\{\phi_\varepsilon^h(X^{\varepsilon,\theta_0}(w)) > 0\}} \exp\{pl_{\varepsilon,h}(w; \theta_0)\}] < \infty .$$

(2) *Let $\theta_0 \in \Theta$ and let K be any compact set of $\mathbf{R}^k$. Then, $\phi_\varepsilon^h(X^{\varepsilon,\theta_0}) \exp\{l_{\varepsilon,h}(w; \theta_0)\}$ has the asymptotic expansion*

$$\phi_\varepsilon^h(X^{\varepsilon,\theta_0})\exp\{l_{\varepsilon,h}(w; \theta_0)\} \sim e^{f_0^L}(1 + \varepsilon\Psi_1 + \varepsilon^2\Psi_2 + \cdots)$$

in D^∞ as $\varepsilon \downarrow 0$ with $\Psi_1, \Psi_2, \ldots \in D^\infty$ determined by the formal Taylor expansion

$$\exp\{\varepsilon f_1^L + \varepsilon^2 f_2^L + \cdots\} = 1 + \varepsilon\Psi_1 + \varepsilon^2\Psi_2 + \cdots .$$

This expansion is uniform in $h \in K$.

Proof. (1) For $X \in C([0, T], \mathbf{R}^d)$, let

$$\phi_\varepsilon^h(X) = \psi(Q_\varepsilon^h(X) - h'I(\theta_0)h) ,$$

where

$$Q_\varepsilon^h(X) = \varepsilon^{-2} \int_0^T [V_0(X_t, \theta_0 + \varepsilon h) - V_0(X_t, \theta_0)]' (VV')^+(X_t)$$
$$\cdot [V_0(X_t, \theta_0 + \varepsilon h) - V_0(X_t, \theta_0)]\, dt .$$

Using Lemma 3.2 (1) for $X_t^{\varepsilon,\theta_0+\varepsilon h} - X_t^{0,\theta_0+\varepsilon h}$ we have

$$\sup_{h\in K} P\left(\sup_{0\leqq t\leqq T} |X_t^{\varepsilon,\theta_0+\varepsilon h} - X_t^{0,\theta_0}| > a_1 \right) = O(\varepsilon^n)$$

as $\varepsilon \downarrow 0$ for any $n \in \mathbf{N}$ and $a_1 > 0$. Then, it is easy to show (i), (ii) and (iii). Let $p > 1$. Then,

$$\begin{aligned} E[1_{\{\phi_\varepsilon^h(X^{\varepsilon,\theta_0})>0\}} \exp\{pl_{\varepsilon,h}(w;\theta_0)\}] &\leqq \Bigg(E\Bigg[\exp\Bigg\{ pq\varepsilon^{-1} \int_0^T [V_0(X_t^{\varepsilon,\theta_0}, \theta_0 + \varepsilon h) \\ &\quad - V_0(X_t^{\varepsilon,\theta_0}, \theta_0)]' \\ &\quad (VV')^+ V(X_t^{\varepsilon,\theta_0})\, dw_t - \tfrac{1}{2} rq Q_\varepsilon^h(X^{\varepsilon,\theta_0}) \Bigg\} \Bigg] \Bigg)^{1/q} \\ &\quad \cdot (E[1_{\{\phi_\varepsilon^h(X^{\varepsilon,\theta_0})>0\}} \exp\{\tfrac{1}{2} rq' Q_\varepsilon^h(X^{\varepsilon,\theta_0})\}])^{1/q'} \\ &\leqq (\exp\{\tfrac{1}{2} rq' [h' I(\theta_0) h + 1]\})^{1/q'}, \end{aligned}$$

where $q > 1$, $q' > 1$, $\frac{1}{q} + \frac{1}{q'} = 1$ and $r = p^2 q$. This proves (iv).
(2) Let $G_\varepsilon = l_{\varepsilon,h}(w;\theta_0)$. Then, from Lemma 3.4

$$G_\varepsilon \sim f_0^L + \varepsilon f_1^L + \cdots$$

in D^∞ uniformly in $h \in K$. We have the expansion

$$(G_\varepsilon - f_0^L)^n = \varepsilon g_{n,1} + \cdots + \varepsilon^{l-1} g_{n,l-1} + \varepsilon^l r_n(w, \varepsilon)$$

with $g_{n,1}, \ldots, g_{n,l-1} \in D^\infty$, $r_n(w,\varepsilon) = O(1)$ in D^∞ as $\varepsilon \downarrow 0$. Using the formula

$$e^x = 1 + \frac{1}{1!} x + \cdots + \frac{1}{(l-1)!} x^{l-1} + R_l(x) ,$$

where $|R_l(x)| \leqq \frac{1}{l!}(e^x + 1)|x|^l$, for $x = G_\varepsilon - f_0^L$ we obtain

$$e^{G_\varepsilon - f_0^L} = 1 + \varepsilon \Psi_1 + \cdots + \varepsilon^{l-1} \Psi_{l-1} + \varepsilon^l r(w, \varepsilon) + R_l(G_\varepsilon - f_0^L) ,$$

where $r(w,\varepsilon) = \sum_{n=1}^{l-1} \frac{1}{n!} r_n(w,\varepsilon)$. Then,

$$\begin{aligned} &\| \phi_\varepsilon^h(X^{\varepsilon,\theta_0}) e^{G_\varepsilon} - e^{f_0^L}(1 + \varepsilon\Psi_1 + \cdots + \varepsilon^{l-1}\Psi_{l-1}) \|_{p,0} \\ &\leqq \| \phi_\varepsilon^h(X^{\varepsilon,\theta_0}) e^{f_0^L} [e^{G_\varepsilon - f_0^L} - (1 + \varepsilon\Psi_1 + \cdots + \varepsilon^{l-1}\Psi_{l-1})] \|_{p,0} \\ &\quad + \| (\phi_\varepsilon^h(X^{\varepsilon,\theta_0}) - 1) e^{f_0^L}(1 + \varepsilon\Psi_1 + \cdots + \varepsilon^{l-1}\Psi_{l-1}) \|_{p,0} \\ &\leqq \| \phi_\varepsilon^h(X^{\varepsilon,\theta_0}) e^{f_0^L} R_l(G_\varepsilon - f_0^L) \|_{p,0} + \varepsilon^l \| \phi_\varepsilon^h(X^{\varepsilon,\theta_0}) e^{f_0^L} r(w,\varepsilon) \|_{p,0} \\ &\quad + \| (\phi_\varepsilon^h(X^{\varepsilon,\theta_0}) - 1) e^{f_0^L}(1 + \varepsilon\Psi_1 + \cdots + \varepsilon^{l-1}\Psi_{l-1}) \|_{p,0} \\ &\leqq \| \phi_\varepsilon^h(X^{\varepsilon,\theta_0}) e^{f_0^L} R_l(G_\varepsilon - f_0^L) \|_{p,0} + O(\varepsilon^l) \end{aligned}$$

uniformly in $h \in K$. The first term in the right hand side is

$$\begin{aligned}
&\| \phi_\varepsilon^h(X^{\varepsilon,\theta_0}) e^{f_0^L} R_l(G_\varepsilon - f_0^L) \|_{p,0} \\
&\leqq \frac{1}{l!} \| \phi_\varepsilon^h(X^{\varepsilon,\theta_0}) e^{G_\varepsilon}(G_\varepsilon - f_0^L)^l \|_{p,0} + \frac{1}{l!} \| \phi_\varepsilon^h(X^{\varepsilon,\theta_0}) e^{f_0^L}(G_\varepsilon - f_0^L)^l \|_{p,0} \\
&\leqq \frac{1}{l!} \| 1_{\{\phi_\varepsilon^h(X^{\varepsilon,\theta_0}) > 0\}} e^{G_\varepsilon} \|_{2p,0} \| |G_\varepsilon - f_0^L|^l \|_{2p,0} + O(\varepsilon^l) \\
&= O(\varepsilon^l)
\end{aligned}$$

as $\varepsilon \downarrow 0$ by (1). We can show that a similar estimation holds for $\| \cdot \|_{p,s}$-norms, $p > 1, s > 0$. This proves (2). □

Using Lemma 4.4 for

$$\begin{aligned}
\Lambda_h^\varepsilon &:= \phi_\varepsilon^h(X^{\varepsilon,\theta_0}) \exp\{l_{\varepsilon,h}(w; \theta_0)\} \\
&\sim \Lambda_h^0(1 + \varepsilon\Psi_1 + \varepsilon^2\Psi_2 + \cdots) \textit{ in } D^\infty \textit{ as } \varepsilon \downarrow 0,
\end{aligned} \tag{4.3}$$

where $\Lambda_h^0 = \exp\{f_0^L\}$, we see

$$\Lambda_h^\varepsilon \psi(\xi_\varepsilon) T_\lambda(F_\varepsilon) \sim \Lambda_h^0[\bar{\Phi}_{h,\lambda,0} + \varepsilon\bar{\Phi}_{h,\lambda,1} + \cdots] \quad \textit{in } D^{-\infty} \textit{ as } \varepsilon \downarrow 0$$

uniformly in $\lambda \in \Lambda$ with $\bar{\Phi}_{h,\lambda,0}, \bar{\Phi}_{h,\lambda,1}, \ldots \in D^{-\infty}$ determined by the formal Taylor expansion

$$\exp\{\varepsilon f_1^L + \varepsilon^2 f_2^L + \cdots\}(\Phi_{\lambda,0} + \varepsilon\Phi_{\lambda,1} + \cdots) = \bar{\Phi}_{h,\lambda,0} + \varepsilon\bar{\Phi}_{h,\lambda,1} + \cdots.$$

In particular,

$$\begin{aligned}
\bar{\Phi}_{h,\lambda,0} &= \Phi_{\lambda,0}, \\
\bar{\Phi}_{h,\lambda,1} &= \Phi_{\lambda,1} + f_1^L \Phi_{\lambda,0}.
\end{aligned}$$

For statistic $S(X)$ and a measurable function f,

$$\begin{aligned}
E_{\theta_0+\varepsilon h}[f(S(X))] &= \int_{C([0,T],\mathbf{R}^d)} f(S(x)) P_{\varepsilon,\theta_0+\varepsilon h}(\mathrm{d}x) \\
&= \int_{C([0,T],\mathbf{R}^d)} f(S(x)) \frac{dP_{\varepsilon,\theta_0+\varepsilon h}}{dP_{\varepsilon,\theta_0}}(x) P_{\varepsilon,\theta_0}(\mathrm{d}x) \\
&= E_{\theta_0}\left[\frac{dP_{\varepsilon,\theta_0+\varepsilon h}}{dP_{\varepsilon,\theta_0}}(X) f(S(X))\right].
\end{aligned}$$

Let K be any compact set of $\mathbf{R}^k$. Since (4.3) is uniform in $h \in K$,

$$\begin{aligned}
P_{\varepsilon,\theta_0+\varepsilon h}[F^\varepsilon[X] \in A] &= E[I_A(F^\varepsilon[X^{\varepsilon,\theta_0+\varepsilon h}])] \\
&\sim E[\phi_\varepsilon^h(X^{\varepsilon,\theta_0+\varepsilon h}) I_A(F^\varepsilon[X^{\varepsilon,\theta_0+\varepsilon h}])] \\
&= E[\phi_\varepsilon^h(X^{\varepsilon,\theta_0}) \exp\{l_{\varepsilon,h}(w; \theta_0)\} \cdot I_A(F^\varepsilon[X^{\varepsilon,\theta_0}])] \\
&\sim E[\phi_\varepsilon^h(X^{\varepsilon,\theta_0}) \exp\{l_{\varepsilon,h}(w; \theta_0)\} \psi(\xi_\varepsilon) I_A(F^\varepsilon[X^{\varepsilon,\theta_0}])] \\
&= E[\phi_\varepsilon^h(X^{\varepsilon,\theta_0}) \exp\{l_{\varepsilon,h}(w; \theta_0)\} \psi(\xi_\varepsilon) I_A \circ (F^\varepsilon[X^{\varepsilon,\theta_0}])] \\
&\sim E[\Lambda_h^0 \bar{\Phi}_{h,A,0}] + \varepsilon E[\Lambda_h^0 \bar{\Phi}_{h,A,1}] + \cdots
\end{aligned}$$

uniformly in K for k-dimensional regular statistic $F^\varepsilon[X]$ with some truncation $\psi(\xi_\varepsilon)$ and $A \in \mathbf{B}^k$, where $\bar{\Phi}_{h,A,i}$ are corresponding to $I_A \circ F^\varepsilon$ for $T_\lambda \circ F^\varepsilon$, i.e.,

$$\bar{\Phi}_{h,A,0} = \Phi_{A,0}$$
$$\bar{\Phi}_{h,A,1} = \Phi_{A,1} + f_1^L \Phi_{A,0}, \dots,$$

where $\Phi_{A,0}, \Phi_{A,1}, \dots \in \tilde{D}^{-\infty}$ are coefficients of the expansion

$$\psi(\xi_\varepsilon) I_A \circ (F^\varepsilon[X^{\varepsilon,\theta_0}]) \sim \Phi_{A,0} + \varepsilon \Phi_{A,1} + \cdots$$

in $\tilde{D}^{-\infty}$ as $\varepsilon \downarrow 0$. From this argument we can obtain the asymptotic expansions of distributions of the bias corrected maximum likelihood estimator and the log likelihood ratio under the contiguous alternative without expanding them again.

Theorem 4.3 *The probability distribution of the bias corrected maximum likelihood estimator $\hat{\theta}_\varepsilon^*(w; \theta_0 + \varepsilon h)$ under the contiguous alternative $P_{\varepsilon,\theta_0+\varepsilon h}$ has the asymptotic expansion*

$$P\left[\frac{\hat{\theta}_\varepsilon^*(w; \theta_0 + \varepsilon h) - (\theta_0 + \varepsilon h)}{\varepsilon} \in A\right] \sim \int_A p_0^c(y)\,\mathrm{d}y + \varepsilon \int_A p_1^c(y)\,\mathrm{d}y + \cdots,$$

$$as\ \varepsilon \downarrow 0,\ A \in \mathbf{B}^k,\ h \in \mathbf{R}^k,$$

where $p_0^c, p_1^c, \dots$ are smooth functions. The expansion is uniform in $A \in \mathbf{B}^k$ and $h \in K$. In particular,

$$p_0^c(y) = \phi(y; 0, I^{-1}),$$

$$p_1^c(y) = \sum_{i,j,l} A_{i,j,l}[-y^i y^j y^l - h^l y^i y^j + I^{ij} y^l + I^{ij} h^l]\phi(y; 0, I^{-1})$$
$$+ \sum_{i,j,l} B_{i,j,l}\left[-\frac{1}{2} y^i y^j y^l - h^i y^j y^l + I^{il} y^j + I^{il} h^j\right]\phi(y; 0, I^{-1})$$
$$- \sum_{j,l} b^j(\theta_0) I_{jl} y^l \phi(y; 0, I^{-1}).$$

Proof. First, we note that

$$P\left[\frac{\hat{\theta}_\varepsilon^*(w; \theta_0 + \varepsilon h) - (\theta_0 + \varepsilon h)}{\varepsilon} \in A\right] = P\left[\frac{\hat{\theta}_\varepsilon^*(w; \theta_0 + \varepsilon h) - \theta_0}{\varepsilon} \in A + h\right],$$

where $A + h = \{x \in \mathbf{R}^k;\ x - h \in A\}$. Let $y = x - h$. All what to do is to find p_0^c and p_1^c.

$$E[\Lambda_h^0 \bar{\Phi}_{h,A+h,0}] = \int_{A+h} \Lambda_h^0(x) p_{f_0}(x)\,\mathrm{d}x = \int_{A+h} \phi(y; 0, I^{-1})\mathrm{d}x,$$

where $\Lambda_h^0(x) = \exp\{h'Ix - \frac{1}{2}h'Ih\}$ and $\bar{\Phi}_{h,A,0} = I_A$. We see that

$$E[\Lambda_h^0 f_1^L I_{A+h}(f_0)] = E[\Lambda_h^0(f_0) I_{A+h}(f_0) E[f_1^L | f_0]]$$
$$= \int_{A+h} E[f_1^L | f_0 = x]\phi(y; 0, I^{-1})\mathrm{d}x$$

and for a smooth function $\tilde{p}_1^c(x)$,

$$\sum_{i=1}^{k} E[\Lambda_h^0 f_1^i \partial_i I_{A+h}(f_0)] = \int_{A+h} \tilde{p}_1^c(x)\mathrm{d}x .$$

By Lemma 4.3, we have

$$E[f_1^L | f_0 = x] = \sum_{i,j,l} A^*_{i,j,l} h^i (x^j x^l - I^{jl}) + \sum_{i,j,l} \frac{1}{2} B_{i,j,l} h^i h^j x^l - \sum_{i,j,l} A_{i,j,l} h^i h^j x^l - \frac{1}{2} \sum_{i,j,l} B_{i,j,l} h^i h^j h^l . \tag{4.4}$$

On the other hand, for $B = \{z; z^i > x^i, i = 1, \ldots, k\}$,

$$\begin{aligned} \tilde{p}_1^c(x) &= (-1)^k \partial_1 \ldots \partial_k \sum_i E[\Lambda_h^0 f_1^i \partial_i I_B(f_0)] \\ &= -\sum_i \partial_i E[\Lambda_h^0 f_1^i \delta_x(f_0)] \\ &= -\sum_i \partial_i \{\Lambda_h^0(x) E[f_1^i | f_0 = x] p_{f_0}(x)\} \\ &= \Lambda_h^0(x) p_1(x) - \sum_i [h' I]^i \Lambda_h^0(x) E[f_1^i | f_0 = x] p_{f_0}(x) . \end{aligned} \tag{4.5}$$

Since

$$p_1^c(y) = \tilde{p}_1^c(y + h) + E[f_1^L | f_0 = y + h]\phi(y; 0, I^{-1}) ,$$

(4.4), (4.5), Theorem 4.1 and (4.2) lead to

$$\begin{aligned} p_1^c(y) = \Bigg\{ &\sum_{i,j,l} A_{i,j,l}[-x^i x^j x^l + 2h^i x^j x^l - h^i h^j x^l + I^{ij} x^l] \\ &+ \sum_{i,j,l} B_{i,j,l}\left[-\frac{1}{2} x^i x^j x^l + \frac{1}{2} h^l x^i x^j + \frac{1}{2} h^i h^j x^l + I^{il} x^j - \frac{1}{2} h^i h^j h^l \right] \\ &- \sum_{i,j} b^i I_{ij}(x^j - h^j) \Bigg\} \phi(y; 0, I^{-1}) . \end{aligned}$$

Substituting $y + h$ for x we obtain the result. □

For the likelihood ratio statistic we obtain the following theorem.

Theorem 4.4 *Let $h \in \mathbf{R}^k$ and $h \neq 0$. The probability distribution of the* log *likelihood ratio $l_{\varepsilon,h}(w; \theta_0 + \varepsilon h)$ has the asymptotic expansion*

$$P[l_{\varepsilon,h}(w; \theta_0 + \varepsilon h) \in A] \sim \int_A p_0^{Lc}(x)\mathrm{d}x + \varepsilon \int_A p_1^{Lc}(x)\,\mathrm{d}x + \cdots, \quad \textit{as } \varepsilon \downarrow 0, A \in \mathbf{B}^1 .$$

The expansion is uniform in $A \in \mathbf{B}^1$. *In particular,*

$$p_0^{Lc}(x) = \phi(\underline{x}; 0, J),$$

$$p_1^{Lc}(x) = \left[\sum_{i,j,l} A_{i,j,l} h^i h^j h^l\right] J^{-3}[\underline{x}^3 + 2J\underline{x}^2 - (3J - J^2)\underline{x} - 2J^2]\phi(\underline{x}; 0, J)$$
$$+ \frac{1}{2}\left[\sum_{i,j,l} B_{i,j,l} h^i h^j h^l\right] J^{-2}[\underline{x}^2 + J\underline{x} - J]\phi(\underline{x}; 0, J),$$

where $\underline{x} = x - \frac{1}{2}J$. *The probability distribution function of* $l_{\varepsilon,h}(w; \theta_0 + \varepsilon h)$ *has the asymptotic expansion*

$$P[l_{\varepsilon,h}(w; \theta_0 + \varepsilon h) \leqq x] \sim \Phi(\underline{x}; 0, J) + \varepsilon\left\{\left[\sum_{i,j,l} A_{i,j,l} h^i h^j h^l\right] J^{-2}[-\underline{x}^2 - 2J\underline{x} + J - J^2]\right.$$
$$\left. + \frac{1}{2}\left[\sum_{i,j,l} B_{i,j,l} h^i h^j h^l\right] J^{-1}[-\underline{x} - J]\right\}\phi(\underline{x}; 0, J) + \cdots.$$

Proof. With some truncation functional $\psi_\varepsilon^L(w) \in D^\infty$ we have

$$\psi_\varepsilon^L(w) I_A(l_{\varepsilon,h}(w; \theta_0)) \sim \Phi_{h,A,0}^L + \varepsilon \Phi_{h,A,1}^L + \cdots$$

in $\tilde{D}^{-\infty}$ as $\varepsilon \downarrow 0$ with $\Phi_{h,A,0}^L, \Phi_{h,A,1}^L, \ldots \in \tilde{D}^{-\infty}$. In particular,

$$\Phi_{h,A,0}^L = I_A(f_0^L)$$
$$\Phi_{h,A,1}^L = f_1^L \partial I_A(f_0^L).$$

By the above argument, we obtain the expansion

$$P(l_{\varepsilon,h}(w; \theta_0 + \varepsilon h) \in A) \sim E[\Lambda_h^0 \bar{\Phi}_{h,A,0}^L] + \varepsilon E[\Lambda_h^0 \bar{\Phi}_{h,A,1}^L] + \cdots,$$

where

$$\bar{\Phi}_{h,A,0}^L = I_A(f_0^L)$$
$$\bar{\Phi}_{h,A,1}^L = f_1^L \partial I_A(f_0^L) + f_1^L I_A(f_0^L).$$

Then,

$$E[\Lambda_h^0 \bar{\Phi}_{h,A,0}^L] = E[\Lambda_h^0 I_A(f_0^L)] = E[e^{f_0^L} I_A(f_0^L)]$$
$$= \int_A e^x \phi(\bar{x}; 0, J)\,\mathrm{d}x = \int_A \phi(\underline{x}; 0, J)\mathrm{d}x.$$

Hence, $p_0^{Lc}(x) = \phi(\underline{x}; 0, J)$. On the other hand,

$$E[\Lambda_h^0 f_1^L \partial I_A(f_0^L)] = \int_A q_1(x)\mathrm{d}x,$$

where

$$q_1(x) = -\partial E[\Lambda_h^0 f_1^L \delta_x(f_0^L)]$$
$$= -\partial\{e^x E[f_1^L | f_0^L = x]\phi(\bar{x}; 0, J)\}$$
$$= -\partial\{E[f_1^L | f_0^L = x]\phi(\underline{x}; 0, J)\}$$

$$= \left\{ \left[\sum_{i,j,l} A_{i,j,l} h^i h^j h^l \right] J^{-3} (\underline{x}^3 + J\underline{x}^2 - 3J\underline{x} - J^2) \right.$$
$$\left. + \frac{1}{2} \left[\sum_{i,j,l} B_{i,j,l} h^i h^j h^l \right] J^{-2} (\underline{x}^2 - J) \right\} \phi(\underline{x}; 0, J) .$$

Moreover,

$$E[A_h^0 f_1^L I_A(f_0^L)] = \int_A q_2(x) \mathrm{d}x ,$$

where

$$q_2(x) = e^x E[f_1^L | f_0^L = x] \phi(\bar{x}; 0, J)$$
$$= E[f_1^L | f_0^L = x] \phi(\underline{x}; 0, J)$$
$$= \left\{ \left[\sum_{i,j,l} A_{i,j,l} h^i h^j h^l \right] J^{-3} (J\underline{x}^2 + J^2 \underline{x} - J^2) \right.$$
$$\left. + \frac{1}{2} \left[\sum_{i,j,l} B_{i,j,l} h^i h^j h^l \right] J^{-1} \underline{x} \right\} \phi(\underline{x}; 0, J) .$$

Since

$$p_1^{Lc}(x) = q_1(x) + q_2(x) ,$$

we obtain $p_1^{Lc}(x)$. □

5 Second-order asymptotic efficiency of the maximum likelihood estimator

The notion of second order efficiency of statistical estimators has been introduced by Fisher, Rao [11], [12], Takeuchi–Akahira, Ghosh–Subramanyam [2] and other authors. Here, as an example of statistical applications of our expansion formulas, we adopt Takeuchi–Akahira's criterion by probability of concentration and show that a bias corrected maximum likelihood estimator is second order asymptotically efficient. For simplicity, we only treat a one dimensional parameter θ, though all is ready for the multiparameter case.

Definition 5.1. An estimator T_ε is *second order asymptotically median unbiased* (*second order* AMU) if for any $\theta_0 \in \Theta$ and any $c > 0$

$$\lim_{\varepsilon \downarrow 0} \sup_{\theta \in \Theta : |\theta - \theta_0| < \varepsilon c} \varepsilon^{-1} |P_{\varepsilon,\theta}[T_\varepsilon - \theta \leqq 0] - \tfrac{1}{2}| = 0$$

and

$$\lim_{\varepsilon \downarrow 0} \sup_{\theta \in \Theta : |\theta - \theta_0| < \varepsilon c} \varepsilon^{-1} |P_{\varepsilon,\theta}[T_\varepsilon - \theta \geqq 0] - \tfrac{1}{2}| = 0 .$$

Given a second order AMU estimator T_ε, if

$$\lim_{\varepsilon \downarrow 0} \varepsilon^{-1} | P_{\varepsilon,\theta_0}[\varepsilon^{-1}(T_\varepsilon - \theta_0) \leqq h] - G_0(h, \theta_0) - \varepsilon G_1(h, \theta_0) | = 0 ,$$

then $G_0(h, \theta_0) + \varepsilon G_1(h, \theta_0)$ is called a second order asymptotic distribution of T_ε.

Consider testing hypothesis H^+: $\theta = \theta_0 + \varepsilon h$ against K: $\theta = \theta_0$, where h is any positive number.

Let $c_\varepsilon = \frac{1}{2}J + \varepsilon p + q_\varepsilon$, where $q_\varepsilon = o(\varepsilon)$ is a sequence. From Theorem 4.4, we see

$$P[l_{\varepsilon,h}(w; \theta_0 + \varepsilon h) \leqq c_\varepsilon] = \tfrac{1}{2} + \varepsilon\{p + A_{1,1,1}h^3[J^{-1} - 1] - \tfrac{1}{2}B_{1,1,1}h^3\}\phi(0; 0, J) + q_\varepsilon\phi(0; 0, J) + O(\varepsilon^2)\,.$$

If we take

$$p = -A_{1,1,1}h^3[J^{-1} - 1] + \tfrac{1}{2}B_{1,1,1}h^3$$

and

$$q_\varepsilon = \varepsilon^{3/2} + \phi(0; 0, J)^{-1}|P_{\varepsilon,\theta_0+\varepsilon h}[\varepsilon^{-1}(T_\varepsilon - \theta_0 - \varepsilon h) \leqq 0] - \tfrac{1}{2}|\,,$$

then we see by Neyman–Pearson's lemma

$$P[l_{\varepsilon,h}(w; \theta_0) \leqq c_\varepsilon] \geqq P_{\varepsilon,\theta_0}[\varepsilon^{-1}(T_\varepsilon - \theta_0) \leqq h]$$

for small ε. Therefore, by Theorem 4.2, for $h > 0$,

$$\liminf_{\varepsilon\downarrow 0} \varepsilon^{-1}\{\Phi(J; 0, J) + \varepsilon[A_{1,1,1}h^3 + \tfrac{1}{2}B_{1,1,1}h^3]\phi(J; 0, J) - P_{\varepsilon,\theta_0}[\varepsilon^{-1}(T_\varepsilon - \theta_0) \leqq h]\} \geqq 0\,.$$

Similarly, for $h < 0$ we have

$$\limsup_{\varepsilon\downarrow 0} \varepsilon^{-1}\{\Phi(-J; 0, J) - \varepsilon[A_{1,1,1}h^3 + \tfrac{1}{2}B_{1,1,1}h^3]\phi(J; 0, J) - P_{\varepsilon,\theta_0}[\varepsilon^{-1}(T_\varepsilon - \theta_0) \leqq h]\} \leqq 0\,.$$

In this sense

$$\Phi(J; 0, J) + \varepsilon[A_{1,1,1}h^3 + \tfrac{1}{2}B_{1,1,1}h^3]\phi(J; 0, J)$$

and

$$\Phi(-J; 0, J) - \varepsilon[A_{1,1,1}h^3 + \tfrac{1}{2}B_{1,1,1}h^3]\phi(J; 0, J)$$

are called the bounds of second order distributions. An AMU estimator attaining these bounds for any $h > 0$ and $h < 0$ is called to be second order efficient. The bias corrected maximum likelihood estimator $\hat{\theta}_\varepsilon^*(w; \theta_0)$ with

$$b(\theta_0) = -A_{1,1,1}I(\theta_0)^{-2}$$

is second order AMU by Theorem 4.3 and it is seen that $\hat{\theta}_\varepsilon^*(w; \theta_0)$ attains the bounds of the second order distributions. Therefore, the bias corrected maximum likelihood estimator is second order efficient.

In conclusion, we note that we have already got the asymptotic expansions for various risk functions, e.g., $E[|\hat{\theta}_\varepsilon^*(w; \theta_0) - \theta_0|^p]$, $p \geqq 1$ if we take a tempered distribution $T(x) = |x|^p$ and notice that for any $p \geqq 1$, $\hat{\theta}_\varepsilon^*(w; \theta_0)$ has L^p-moments, see Chap. 3 of Kutoyants [6].

Acknowledgements. The author is grateful to Professor S. Watanabe for his helpful advice and also wishes to thank the referee and the associate editor for their valuable comments.

References

1. Akahira, M., Takeuchi, K.: Asymptotic efficiency of statistical estimators: concepts and higher order asymptotic efficiency. (Lect. Notes Stat., Vol. 7) Berlin Heidelberg New York: Springer 1981
2. Ghosh, J.K., Subramanyam, K.: Second order efficiency of maximum likelihood estimators. Sankhyā, Ser. A **36**, 325–358 (1974)
3. Ikeda, N., Watanabe, S.: An introduction to Malliavin's calculus. In: Itô, K. (ed.) Stochastic analysis. Proc. Taniguchi Intern. Symp. on Stochastic Analysis, Katada and Kyoto 1982, pp. 1–52. Tokyo: Kinokuniya/North-Holland 1984
4. Ikeda, N., Watanabe, S.: Stochastic differential equations and diffusion processes, 2nd edn. Tokyo: North-Holland/Kodansha 1989
5. Kusuoka, S., Strook, D.W.: Applications of the Malliavin calculus, Part I. In: Itô, K. (ed.) Stochastic analysis. Proc. Taniguchi Intern. Symp. on Stochastic Analysis, Katada and Kyoto 1982, pp. 271–306. Tokyo: Kinokuniya/North-Holland 1984
6. Kutoyants, Yu.A.: Parameter estimation for stochastic processes. Prakasa Rao, B.L.S. (ed.) Berlin: Heldermann 1984
7. Kutoyants, Yu.A.: Expansion of a maximum likelihood estimate by diffusion powers. Theory Probab. Appl. **29**, 465–477 (1984)
8. Liptser, R.S., Shiryayev, A.N.: Statistics of random processes I. Berlin Heidelberg New York: Springer 1977
9. Malliavin, P.: Stochastic calculus of variation and hypoelliptic operators. In: Itô, K. (ed.) Proceedings of the International Symposium SDE Kyoto, pp. 195–263. Tokyo: Kinokuniya 1976
10. Malliavin, P.: C^k-hypoellipticity with degeneracy. In: Friedman, A., Pinsky, M. (eds.) Stochastic analysis, pp. 199–214, 327–340. New York: Academic Press 1978
11. Rao, C.R.: Asymptotic efficiency and limiting information. In: Proc. Fourth Berkeley Symp. on Math. Statist. Probab. Vol. 1, pp. 531–545. Berkeley and Los Angeles: University of California Press 1961
12. Rao, C.R.: Criteria of estimation in large samples. Sankhyā **25**, 189–206 (1963)
13. Taniguchi, M.: On the second order asymptotic efficiency of estimators of Gaussian ARMA processes. Ann. Stat. **11**, 157–169 (1983)
14. Taniguchi, M.: Higher order asymptotic theory for time series analysis (Lect. Notes Stat. Vol. 68), Berlin Heidelberg New York: Springer 1991
15. Watanabe, S.: Malliavin's calculus in terms of generalized Wiener functionals. In: (Lect. Notes Control Inf. Sci., Vol. 49) Berlin Heidelberg New York: Springer 1983
16. Watanabe, S.: Lectures on stochastic differential equations and Malliavin calculus. Tata Institute of Fundamental Research. Berlin Heidelberg New York: Springer 1984
17. Watanabe, S.: Analysis of Wiener functionals (Malliavin calculus) and its applications to heat kernels. Ann. Probab. **15**, 1–39 (1987)
18. Yoshida, N.: Asymptotic expansion for small diffusion – an application of Malliavin–Watanabe theory. Research Memorandum Vol. 383. The Institute of Statistical Mathematics, 1990