JOURNAL OF MULTIVARIATE ANALYSIS 57, 1–36 (1996)
ARTICLE NO. 0019

Asymptotic Expansions for Perturbed Systems on Wiener Space: Maximum Likelihood Estimators*

NAKAHIRO YOSHIDA

The Institute of Statistical Mathematics, Tokyo, Japan

By means of the Malliavin Calculus, we derive asymptotic expansion of the probability distributions of statistics for systems perturbed by small noises. These results are applied to the problem of the second order asymptotic efficiency of the maximum likelihood estimator.

1. INTRODUCTION

We consider stochastic systems with unknown parameters disturbed by white Gaussian noises or normal random variables. For many such systems the unknown parameters can be estimated consistently by certain statistical estimators when the disturbances become small and the stochastic system tends to the corresponding deterministic one. For instance, the maximum likelihood method and the Bayes method are available for diffusion processes with unknown parameters in their drifts when the diffusion coefficient is small. In this case, the maximum likelihood estimator and the Bayes estimator are consistent and efficient in the first order, e.g., Chapter 3 of Kutoyants [4]. As for higher order properties of estimators, they are known to be second-order efficient in a certain sense. This fact follows from their asymptotic expansions in consideration of a problem of hypothesis testing [12, 14]. Thus asymptotic expansions for estimators play an important role in higher order statistical inference. The purpose of this article is to derive asymptotic expansions for likelihood ratio statistics and maximum likelihood estimators of unknown parameters involved in a system slightly disturbed by white Gaussian noises or normal random variables.

Received March 25, 1994; revised October 1995.

AMS 1991 subject classifications: 62M05, 62F12, 60H10.

Key words and phrases: maximum likelihood estimator, Malliavin calculus, asymptotic expansion.

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

0047-259X/96 $18.00

We formulate this problem as follows. Let $W = \{w;\ w$ is an $\mathbf{R}^r$-valued continuous path on $[0, \infty)$, $w(0) = 0\}$. W is a Fréchet space endowed with sup-norms on compact sets in $[0, \infty)$. Let P be a Wiener measure on $(W, \mathbf{B}(W))$, where $\mathbf{B}(W)$ is the Borel σ-field of W. The probability space (W, P) is referred to as the Wiener space. Let $(S, \mathbf{B}_s)$ be a measurable space. The closure Θ of a bounded convex domain Θ^o in $\mathbf{R}^k$ denotes the space of unknown parameters. We assume that the slightly perturbed system can be represented by an S-valued random element F^ε_θ, $\theta \in \Theta$, $\varepsilon \in (0, 1)$, defined on (W, P). Let P^ε_θ be the probability measure on S induced by F^ε_θ from P. Then we obtain a family of statistical experiments $\{P^\varepsilon_\theta;\ \theta \in \Theta\}$, $\varepsilon \in (0, 1]$, on $(S, \mathbf{B}_s)$. We assume that $P^\varepsilon_{\theta_1}$ and $P^\varepsilon_{\theta_2}$ are mutually dominated for $\theta_1, \theta_2 \in \Theta$. For $\theta_0 \in \Theta^o$ let

$$\lambda_\varepsilon(w, \theta; \theta_1) = \log \frac{dP^\varepsilon_\theta}{dP^\varepsilon_{\theta_0}} (F^\varepsilon_{\theta_1}(w))$$

for $w \in W$, $\theta, \theta_1 \in \Theta$, $\varepsilon \in (0, 1]$($\lambda_\varepsilon(w, \theta; \theta_1)$ depends on θ_0). When $\theta_0 \in \Theta^o$ is the true value, the maximum likelihood estimator $\hat{\theta}_\varepsilon(w; \theta_0)$, if it exists, satisfies

$$\lambda_\varepsilon(w, \hat{\theta}_\varepsilon(w; \theta_0); \theta_0) = \sup_{\theta \in \Theta} \lambda_\varepsilon(w, \theta; \theta_0).$$

Under a set of conditions stated in the next section, which ensures the regularity and entire separation of the statistical experiments $\{P^\varepsilon_\theta;\ \theta \in \Theta\}$, $\varepsilon \in (0, 1)$, we derive the asymptotic expansions for the maximum likelihood estimator and the log likelihood ratio statistic used in the higher order statistical inference. We apply these results to the problem of the second-order efficiency of the maximum likelihood estimator. As in [12], the technique used here is the Malliavin calculus exploited by Watanabe [9–11] and its modification with truncation. This modification enables us to deal with statistical estimators, such as the maximum likelihood estimator, whose existence and regularity cannot be ensured on the whole sample space in general.

The organization of this paper is as follows. The notations and assumptions are stated in Section 2. Sections 3 and 4 give the main results. Examples are presented in Section 5. In Sections 6 and 7, we prove the results stated in Sections 3 and 4. We could reduce the conditions to milder ones for second-order expansions used in Section 5, but we will not pursue this point here.

2. Notations and Assumptions

Let H be the Cameron–Martin subspace of W: the totality of $\mathbf{R}^r$-valued absolutely continuous functions on $[0, \infty)$ with square integrable derivative, endowed with the inner product

$$\langle h_1, h_2 \rangle_H = \int_0^\infty (\dot{h}_{1,t}, \dot{h}_{2,t})\, dt$$

for $h_1, h_2 \in H$. Let D denote the H-derivative.

For Hilbert space E, $\|\cdot\|_p$ denotes the $L^p(E)$-norm of an E-valued Wiener functional, i.e., for each Wiener functional $f\colon W \to E$, $\|f\|_p^p = \int_W |f|_E^p\, P(dw)$. Let L be the Ornstein–Uhlenbeck operator (see Watanabe [10]) and define $\|f\|_{p,s}$ for E-valued Wiener functionals 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 $\|\cdot\|_{p,s}$. It is known that for $n \in \bar{\mathbf{N}} (:= \{0\} \cup \mathbf{N})$, and $p > 1$, the norm $\|\cdot\|_{p,n}$ is equivalent to the norm $\sum_{l=0}^n \|D^l \cdot\|_p$. Let $D^\infty(E)$ be the set of Wiener test functionals of Watanabe [10]: $D^\infty(E) = \bigcap_{s>0} \bigcap_{1<p<\infty} D_p^s(E)$. 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. We suppress $\mathbf{R}$ when $E = \mathbf{R}$. Let us consider a family of E-valued Wiener functionals (or generalized Wiener functionals) $\{F_\varepsilon(w)\}$, $\varepsilon \in (0, 1)$. We will consider the asymptotic expansion taking the form of

$$F_\varepsilon \sim f_0 + \varepsilon f_1 + \cdots$$

as $\varepsilon \downarrow 0$ in $D^\infty(E)$, $D^{-\infty}(E)$, or $\tilde{D}^{-\infty}(E)$. See Watanabe [11] for definition. The generalized mean of $F_\varepsilon(w)$ yields the ordinary asymptotic expansion.

Let $\delta_0 = \partial/\partial_\varepsilon$ and $\delta_i = \partial/\partial\theta^i$, $i = 1, ..., k$, $\theta = (\theta^i)$. For $v = (v_0, v_1, ..., v_k)$ with $v_i \in \bar{\mathbf{N}}$, $i = 0, 1, ..., k$, the differential operator δ_v is defined by $\delta_v = (\delta_0)^{v_0} (\delta_1)^{v_1} \cdots (\delta_k)^{v_k}$. Let $|v| = v_0 + v_1 + \cdots + v_k$. Let $\delta = (\delta_1, ..., \delta_k)$ and for differentiable function f defined on Θ, $\delta^m f$ denotes m-linear form defined by

$$\delta^m f[u_1, ..., u_m] = \sum_{i_1, ..., i_m = 1}^k \delta_{i_1} \cdots \delta_{i_m} f u_1^{i_1} \cdots u_m^{i_m},$$

where $u_j = (u_j^i) \in \mathbf{R}^k$, $i = 1, ..., k, j = 1, ..., m$. Let $\mathbf{M} = \{0\} \times \bar{\mathbf{N}}^k$. Let $\psi\colon \mathbf{R} \to \mathbf{R}$ be a smooth function such that $0 \leqslant \psi(x) \leqslant 1$ for $x \in \mathbf{R}$, $\psi(x) = 1$ for $|x| \leqslant \frac{1}{2}$ and $\psi(x) = 0$ for $|x| \geqslant 1$. Two a.s. equal random variables are identified.

We will construct a Sobolev space of Banach space valued functionals in a similar manner as Kusuoka [2, 3]. Let E be a real Banach space. We say that a strongly measurable map $F\colon (W, \overline{\mathbf{B}(W)}^P) \to (E, \mathbf{B}(E))$ is ray absolutely continuous (RAC) if for any $h \in H$, there exists a strongly measurable map $\tilde{F}_h\colon (W, \overline{\mathbf{B}(W)}^P) \to (E, \mathbf{B}(E))$ such that $\tilde{F}_h(w) = F(w)$ P-a.e. w, and the map $s \to \tilde{F}_h(w+sh)$ is strongly absolutely continuous, i.e., there exists a strongly measurable map $G_h^{F,w}\colon \mathbf{R} \to E$ such that for any $w \in W$ and $a, b \in \mathbf{R}$ $(a<b)$, $\int_a^b \|G_h^{F,w}(s)\|_E\, ds < \infty$ and

$$\tilde{F}_h(w+bh) = \tilde{F}_h(w+ah) + \int_a^b G_h^{F,w}(s)\, ds.$$

Let T^o be a bounded open set in $\mathbf{R}^{k_0}$, and let $T = \bar{T}^o$. Let V be a real separable Hilbert space with a Hilbertian norm $\|\cdot\|_V$. Denote by $C(T \to V)$ the Banach space of continuous maps from T to V equipped with the supremum norm $\|F\|_{C(T\to V)} = \sup_{t\in T} \|F(t)\|_V$. We say that a strongly measurable map $F\colon W \to C(T \to V)$ is strongly stochastically Gâteaux differentiable (SSGD) in H directions if there exists a strongly measurable map $DF\colon W \to C(T \to H \otimes V)$ such that for any $h \in H$.

$$\left\| \frac{1}{s}(F(w+sh) - F(w)) - DF(w)[h] \right\|_{C(T\to V)} \to 0$$

in P as $s \to 0$. $L^p(W \to E)$ denotes the L^p space of strongly measurable maps from W to E satisfying that

$$\|F\|^p_{L^p(W\to E)} := \int_W \|F(w)\|_E^p\, P(dw) < \infty.$$

For $p>1$, let $H_p^0(C(T\to V)) = L^p(W \to C(T \to V))$, and let $H_p^1(C(T\to V)) = \{F \in H_p^0(C(T\to V))\colon$ F is RAC and SSGD in H directions, and $DF \in L^p(W \to C(T \to H\otimes V))\}$. For $F \in H_p^1(C(T\to V))$, put $\|F\|_{H_p^1(C(T\to V))} = \|F\|_{L^p(W\to C(T\to V))} + \|DF\|_{L^p(W\to C(T\to H\otimes V))}$. As in Kusuoka [2, 3], we can prove that the space $H_p^1(C(T\to V))$ is a Banach space with respect to the norm $\|\cdot\|_{H_p^1(C(T\to V))}$. Let T_1^o and T_2^o be open convex bounded sets in $\mathbf{R}^{k_1}$ and $\mathbf{R}^{k_2}$, respectively. Denote $T_i = \bar{T}_i^o$ for $i=1, 2$. Let $T^o = T_1^o \times T_2^o$ and let $\tilde{T}^o$ be a bounded open set satisfying $T^o \subset \tilde{T}^o$ with $k_0 = k_1 + k_2$. Denote the closure of $\tilde{T}^o$ by $\tilde{T}$. It is not difficult to prove that if $g \in H_p^1(C(\tilde{T} \to V))$, then $I(g)(w, \eta) := \int_{T_1} g(w, \xi, \eta)\, d\xi$ belongs to $H_p^1(C(T_2 \to V))$. Consider a map $g\colon W \times \tilde{T} \to V$. We denote by $D_i g$ the derivative of g with respect to $t_i \in T_i$, $i=1, 2$, if for P-a.e. $w \in W$, the map $\tilde{T} \ni t \to g(w, t)$ is differentiable with respect to t_i on $\tilde{T}^o$, and each partial derivative can be extended continuously to $\tilde{T}$. Let D denote the differential operator in the definition of the space $H_p^1(C(T\to V))$. Put $D_0 = D$, and let $H_0 = H$ and $H_i = \mathbf{R}^{k_i}$,

$i = 1, 2$. We say that a functional g: $W \to C(\tilde{T} \to V)$ is smooth if for any $p > 1$, $n \in \bar{\mathbf{N}}$ and $i_1, i_2, ..., i_n \in \{0, 1, 2\}$,

$$D_{i_1} D_{i_2} \cdots D_{i_n} g \in H^1_p(C(\tilde{T} \to H_{i_1} \otimes H_{i_2} \otimes \cdots \otimes H_{i_n} \otimes V)).$$

Consider a functional g: $W \to C(\tilde{T} \to \mathbf{R}^{k_1})$, and assume that $D_1 g(w, t)$ takes values in the set of symmetric matrices for all $t \in \tilde{T}$, P-a.e. w. Given a functional R: $W \to \mathbf{R}$, R is naturally identified with a map taking values in $C(\tilde{T} \to \mathbf{R})$ by $R(w, t) = R(w)$ for any $t \in \tilde{T}$. Let R be smooth. We fix a version of R and g. Assume that for some convex set U in T_1, the following conditions hold for R and g: (1) If $R(w) < 1$, the equation $g(w, \xi, \eta) = 0$ has a root $\hat{\xi}(w, \eta)$ in U for any $\eta \in T_2$; (2) $D_1 g(w, t^*)$ is positive-definite uniformly in $(w, t^*) \in \{w: R(w) < 1\} \times U \times T_2$; (3) For each $h \in H$, there exist RAC versions $\tilde{R}_h(w)$ and $\tilde{g}_h$ such that if $\tilde{R}_h(w) < 1$, the equation $\tilde{g}_h(w, \xi, \eta) = 0$ has a root $\tilde{\hat{\xi}}(w, \eta)$ in U for any $\eta \in T_2$; (4) For each $h \in H$, $D_1 \tilde{g}_h(w, t^*)$ is positive-definite uniformly in $(w, t^*) \in \{w: \tilde{R}_h(w) < 1\} \times U \times T_2$. Furthermore, assume that g is smooth. Then we can prove that, under these conditions, $\psi(3R)\, \hat{\xi}$: $W \to C(T_2 \to \mathbf{R}^{k_1})$ is well defined and smooth [15]. Here, if $T \subset \tilde{T}^o$, the derivatives with respect to the parameter are ordinary derivatives.

Put

$$G(w, \varepsilon, \theta; \theta_0) = \varepsilon^2 \lambda_\varepsilon(w, \theta; \theta_0).$$

In this paper we consider the following conditions. Conditions (C1) and (C4) (or C5)) are regularity conditions. Condition (C2) ensures the existence of the consistent estimators. By Condition (C3) we confine ourselves to discussing the locally asymptotically normal experiments.

(C1) For each $\theta_0 \in \Theta^o$, the functional $G(\cdot, \cdot, \cdot; \theta_0)$: $W \to C([0, 1] \times \Theta \to \mathbf{R})$ is smooth, where $G(w, 0, \theta; \theta_0) = \lim_{\varepsilon \downarrow 0} G(w, \varepsilon, \theta; \theta_0)$.

(C2) For each $(\theta, \theta_0) \in \Theta \times \Theta^o$, $G(w, 0, \theta; \theta_0)$ is deterministic ($G(0, \theta; \theta_0)$, say), and for each $\theta_0 \in \Theta^o$, there exists $a_0 > 0$ such that $-G(0, \theta; \theta_0) \geqslant a_0 |\theta - \theta_0|^2$ for any $\theta \in \Theta$.

(C3) For $\theta_0 \in \Theta^o$, there exist $h^{(i)} \in H$, $i = 1, 2, ..., k$, such that

$$\delta_0 \delta_i G(w, 0, \theta_0; \theta_0) = \int_0^\infty \dot{h}^{(i)}_s \cdot dw_s$$

and

$$\begin{aligned} \operatorname{cov}(\delta_0 \delta_i G(\cdot, 0, \theta_0; \theta_0), \delta_0 \delta_j G(\cdot, 0, \theta_0; \theta_0)) &\equiv \langle h^{(i)}, h^{(j)} \rangle H \\ &= -\delta_i \delta_j G(0, \theta_0; \theta_0) \end{aligned}$$

for $i, j = 1, ..., k$.

(C4) For $\theta_0 \in \Theta^o$ and any compact set $K \subset \mathbf{R}^k$, there exists $p_0 > 1$ such that

$$\sup_{\substack{\varepsilon \in (0,1) \\ u \in K \\ \theta_0 + \varepsilon u \subset \Theta}} E[\exp\{p_0 \varepsilon^{-2} G(w, \varepsilon, \theta_0 + \varepsilon u; \theta_0)\}] < \infty.$$

Remarks 2.1. (1) For a map $F\colon W \times \Theta \to V$, define $\iota_\theta F\colon W \to V$ by $(\iota_\theta F)(w) = F(w, \theta)$. If $F \in H^1_p(C(\Theta \to V))$, $\iota_\theta F \in D^1_p(V)$. In fact, $\iota_\theta F\colon W \to V$ is RAC and SSGD with $D\iota_\theta F \in L^p(W \to V)$, and, hence, we see that $\iota_\theta F \in D^1_p(V)$ by using the equivalence between Sobolev spaces proved by Sugita [7]. For this map, $\iota_\theta DF = D\iota_\theta F$, where the H-derivative on the RHS is the ordinary one.

(2) $G(0, \theta_0; \theta_0) = \lim_{\varepsilon \downarrow 0} \varepsilon^2 \lambda_\varepsilon(w, \theta_0; \theta_0) = 0$. Therefore from (C2) we see that $\delta_i G(0, \theta_0; \theta_0) = 0$, $1 \leqslant i \leqslant k$, and the bilinear form for the Hessian matrix $(\delta_i \delta_j G(0, \theta_0; \theta_0))_{1 \leqslant i, j \leqslant k}$ is negative definite. The matrix $I(\theta_0) = (I_{ij}) = -\delta^2 G(0, \theta_0; \theta_0)$ is called the Fisher information matrix.

(3) To obtain the results in Sections 3 and 4, it suffices to assume the following weaker condition (C5) in place of (C4):

(C5) For $\theta_0 \in \Theta^o$ and any compact set $K \subset \mathbf{R}^k$, there exist measurable functions φ^u_ε, $u \in K$, $\varepsilon \in (0, 1)$, on S satisfying the following conditions:

(i) $0 \leqslant \varphi^u_\varepsilon(x) \leqslant 1$, $x \in S$.

(ii) $\varphi^u_\varepsilon(F^\varepsilon_{\theta_0}(w)) = 1 - O(\varepsilon^n)$ in D^∞ as $\varepsilon \downarrow 0$ uniformly in $u \in K$ for $n = 1, 2, \ldots$.

(iii) $\varphi^u_\varepsilon(F^\varepsilon_{\theta_0 + \varepsilon u}(w)) = 1 - O(\varepsilon^n)$ in D^∞ as $\varepsilon \downarrow 0$ uniformly in $u \in K$ for $n = 1, 2, \ldots$.

(iv) For some $p_0 > 1$,

$$\sup_{\substack{\varepsilon \in (0,1) \\ u \in K \\ \theta_0 + \varepsilon u \in \Theta}} E[1_{\{\varphi^u_\varepsilon(F^\varepsilon_{\theta_0}(w)) > 0\}} \exp\{p_0 \varepsilon^{-2} G(w, \varepsilon, \theta_0 + \varepsilon u; \theta_0)\}] < \infty.$$

It is clear that if Condition (C4) holds true, Condition (C5) is satisfied for $\varphi^u_\varepsilon \equiv 1$.

3. Asymptotic Expansions for Likelihood Ratio Statistics

In this section we present asymptotic expansions for likelihood ratio statistics. For simplicity denote

$$G_{l;\,i_1,\,i_2,\,\dots,\,i_m}=(\delta_0)^l\,\delta_{i_1}\delta_{i_2}\cdots\delta_{i_m}G(w,0,\theta_0,\theta_0)$$
$$\equiv(\delta_0)^l\,\delta_{i_1}\delta_{i_2}\cdots\delta_{i_m}G(w,\varepsilon,\theta;\theta_0)|_{\varepsilon=0,\,\theta=\theta_0}$$

for $l\in\bar{\mathbf{N}}$, $m\in\mathbf{N}$, and $i_1, i_2, \dots, i_m\in\{1,2,\dots,k\}$. We will use Einstein's rule for repeated indices. For $u=(u^i)\in\mathbf{R}^k$, let

$$f_0^{L,\,u}=G_{1;\,i}u^i-\tfrac{1}{2}I_{ij}u^iu^j,$$
$$f_1^{L,\,u}=\tfrac{1}{2}G_{2;\,i}u^i+\tfrac{1}{2}G_{1;\,i,\,j}u^iu^j+\tfrac{1}{6}\,G_{0;\,i,\,j,\,l}u^iu^ju^l.$$

Moreover, let

$$q^{L,\,u}(x)=E[f_1^{L,\,u}\mid f_0^{L,\,u}=x].$$

The distribution function of the normal distribution $N(\mu,\sigma^2)$ is denoted by $\Phi(x;\mu,\sigma^2)$ and its density by $\phi(x;\mu,\sigma^2)$. The differential operator $\partial/\partial x$ is denoted by ∂. Suppose $\theta_0\in\Theta^o$. Let $\mathbf{B}^1$ denote the Borel σ-field of $\mathbf{R}^1$.

THEOREM 3.1. *Let $u\in\mathbf{R}^k-\{0\}$. Assume that (C1)–(C3) are satisfied. Then the distribution of $\varepsilon^{-2}G(w,\varepsilon,\theta_0+\varepsilon u;\theta_0)$ has the asymptotic expansion*

$$P(\varepsilon^{-2}G(w,\varepsilon,\theta_0+\varepsilon u;\theta_0)\in A)\sim\int_A p_0^{L,\,u}(x)\,dx+\varepsilon\int_A p_1^{L,\,u}(x)\,dx+\cdots$$

as $\varepsilon\downarrow 0$ for $A\in\mathbf{B}^1$, where $p_0^{L,\,u}, p_1^{L,\,u}, \dots$ are integrable smooth functions depending on u. This expansion is uniform in $A\in\mathbf{B}^1$. In particular,

$$p_0^{L,\,u}(x)=\phi(x;\,-\tfrac{1}{2}J,J),$$
$$p_1^{L,\,u}(x)=-\partial\{q^{L,\,u}(x)\,\phi(x;\,-\tfrac{1}{2}J,J)\},$$

where $J=I(\theta_0)[u,u]$. The probability distribution function of $\varepsilon^{-2}G(w,\varepsilon,\theta_0+\varepsilon u;\theta_0)$ has the asymptotic expansion

$$P(\varepsilon^{-2}G(w,\varepsilon,\theta_0+\varepsilon u;\theta_0)\leqslant x)$$
$$\sim\Phi(x;\,-\tfrac{1}{2}J,J)-\varepsilon q^{L,\,u}(x)\,\phi(x;\,-\tfrac{1}{2}J,J)+\cdots$$

as $\varepsilon\downarrow 0$ for $x\in\mathbf{R}$.

The following theorem gives the asymptotic expansion of the distribution of the log likelihood ratio statistic under the contiguous alternative $P^{\varepsilon}_{\theta_0+\varepsilon u}$. As defined in Sections 1 and 2,

$$\lambda_\varepsilon(w,\theta;\theta_0+\varepsilon u)=\log\frac{dP^{\varepsilon}_{\theta}}{dP^{\varepsilon}_{\theta_0}}\,(F^{\varepsilon}_{\theta_0+\varepsilon u}(w))$$

and

$$G(w, \varepsilon, \theta; \theta_0 + \varepsilon u) = \varepsilon^2 \lambda_\varepsilon(w, \theta; \theta_0 + \varepsilon u).$$

Then $\varepsilon^{-2} G(w, \varepsilon, \theta_0 + \varepsilon u; \theta_0 + \varepsilon u)$ is the log likelihood ratio statistic when the true parameter is $\theta_0 + \varepsilon u$.

THEOREM 3.2. *Assume that* (C1)–(C3) *and* (C5) *are satisfied. Let* $u \in \mathbf{R}^k - \{0\}$. *Then the distribution of* $\varepsilon^{-2} G(w, \varepsilon, \theta_0 + \varepsilon u;\ \theta_0 + \varepsilon u)$ *has the asymptotic expansion*

$$P(\varepsilon^{-2} G(w, \varepsilon, \theta_0 + \varepsilon u; \theta_0 + \varepsilon u) \in A) \sim \int_A p_0^{Lc, u}(x)\, dx + \varepsilon \int_A p_1^{Lc, u}(x)\, dx + \cdots$$

as $\varepsilon \downarrow 0$ *for* $A \in \mathbf{B}^1$, *where* $p_0^{Lc, u}, p_1^{Lc, u}, \ldots$ *are integrable smooth functions depending on u. This expansion is uniform in* $A \in \mathbf{B}^1$. *In particular,*

$$p_0^{Lc, u}(x) = \phi(x; \tfrac{1}{2} J, J),$$
$$p_1^{Lc, u}(x) = -\partial\{q^{L, u}(x)\, \phi(x; \tfrac{1}{2} J, J)\} + q^{L, u}(x)\, \phi(x; \tfrac{1}{2} J, J).$$

The probability distribution function of $\varepsilon^{-2} G(w, \varepsilon, \theta_0 + \varepsilon u; \theta_0 + \varepsilon u)$ *has the asymptotic expansion*

$$P(\varepsilon^{-2} G(w, \varepsilon, \theta_0 + \varepsilon u; \theta_0 + \varepsilon u) \leqslant x)$$
$$\sim \Phi(x; \tfrac{1}{2} J, J) + \varepsilon\Big(-q^{L, u}(x)\, \phi(x; \tfrac{1}{2} J, J) + \int_{-\infty}^{x} q^{L, u}(z)\, \phi(z; \tfrac{1}{2} J, J)\, dz\Big) + \cdots$$

as $\varepsilon \downarrow 0$ *for* $x \in \mathbf{R}$.

4. ASYMPTOTIC EXPANSIONS FOR MAXIMUM LIKELIHOOD ESTIMATORS

In this section we present asymptotic expansions for the maximum likelihood estimator. In the higher order statistical asymptotic theory we need bias corrections of maximum likelihood estimators. For smooth function $b(\theta)$ with bounded derivatives on Θ,

$$\hat{\theta}_\varepsilon^*(w; \theta_0) = \hat{\theta}_\varepsilon(w; \theta_0) - \varepsilon^2 b(\hat{\theta}_\varepsilon(w; \theta_0))$$

is called a bias corrected maximum likelihood estimator. Let

$$f_0^i = I^{ij} G_{1;j},$$
$$f_1^i = \tfrac{1}{2} I^{ij} G_{2;j} + I^{ij} G_{1;j,l} f_0^l + \tfrac{1}{2} I^{ij} G_{0;j,l,m}\, f_0^l f_0^m - b(\theta_0)^i$$

for $i = 1, 2, ..., k$. Here $I = (I_{ij}) = I(\theta_0) = -\delta^2 G(0, \theta_0; \theta_0)$ and $I^{-1} = (I^{ij}) = I(\theta_0)^{-1}$. Moreover, let

$$q_b^i(x) = E[f_1^i \mid f_0 = x]$$

for $x \in \mathbf{R}^k$ and $i = 1, 2, ..., k$. We denote by ∂_i the partial differential $\partial/\partial x^i$, or $\partial/\partial y^i$. The density of the k-dimensional normal distribution with mean μ and covariance matrix Σ is denoted by $\phi(x; \mu, \Sigma)$. Then we obtain the following theorem.

THEOREM 4.1. *Suppose Conditions* (C1)–(C3) *hold. Then there exists a consistent maximum likelihood estimator* $\hat{\theta}_\varepsilon(w; \theta_0)$ *for any true value* $\theta_0 \in \Theta^o$. *The distribution of the bias corrected maximum likelihood estimator* $\hat{\theta}_\varepsilon^*(w; \theta_0)$ *has the asymptotic expansion*

$$P(\varepsilon^{-1}(\hat{\theta}_\varepsilon^*(w; \theta_0) - \theta_0) \in A) \sim \int_A p_0(x)\, dx + \varepsilon \int_A p_1(x)\, dx + \cdots$$

as $\varepsilon \downarrow 0$ *for* $A \in \mathbf{B}^k$, *where* $p_i(x), i = 0, 1, ...,$ *are smooth functions. In particular,*

$$p_0(x) = \phi(x; 0, I(\theta_0)^{-1}),$$
$$p_1(x) = -\partial_i \{ q_b^i(x)\, \phi(x; 0, I(\theta_0)^{-1}) \}.$$

This expansion is uniform in $A \in \mathbf{B}^k$.

The asymptotic expansions under contiguous alternatives are important from the statistical viewpoint. For instance, they are useful to calculate the power of a test with the maximum likelihood estimator. The maximum likelihood estimator under the contiguous alternative $P^\varepsilon_{\theta_0 + \varepsilon u}$, $u \in \mathbf{R}^k - \{0\}$, is (roughly speaking) defined by maximizing $\lambda_\varepsilon(w, \theta; \theta_0 + \varepsilon u)$ in $\theta \in \Theta$. Let $\hat{\theta}_\varepsilon(w; \theta_0 + \varepsilon u)$ denote the maximum likelihood estimator under the contiguous alternative $P^\varepsilon_{\theta_0 + \varepsilon u}$. As before the bias corrected maximum likelihood estimator under the contiguous alternative $P^\varepsilon_{\theta_0 + \varepsilon u}$ is defined and denoted by $\hat{\theta}_\varepsilon^*(w; \theta_0 + \varepsilon u)$. Then we obtain the asymptotic expansion for the bias corrected maximum likelihood estimator under contiguous alternatives.

THEOREM 4.2. *Assume that* (C1)–(C3) *and* (C5) *are satisfied. Then the probability distribution of the bias corrected maximum likelihood estimator* $\hat{\theta}_\varepsilon^*(w; \theta_0 + \varepsilon u)$ *under the contiguous alternative* $P^\varepsilon_{\theta_0 + \varepsilon u}$ *has the asymptotic expansion*

$$P(\varepsilon^{-1}(\hat{\theta}_\varepsilon^*(w; \theta_0 + \varepsilon u) - (\theta_0 + \varepsilon u)) \in A) \sim \int_A p_0^{c,u}(y)\, dy + \varepsilon \int_A p_1^{c,u}(y)\, dy + \cdots$$

as $\varepsilon \downarrow 0$, $A \in \mathbf{B}^k$, $u \in \mathbf{R}^k$, *where* $p_0^{c,u}, p_1^{c,u}, \ldots$ *are smooth functions depending on* u. *This expansion is uniform in* $A \in \mathbf{B}^k$ *and* $u \in K$, *where* K *is any compact set in* $\mathbf{R}^k$. *In particular,*

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

$$p_1^{c,u}(y) = -\partial_i(q_b^i(y+u)\,\phi(y; 0, I^{-1})) + q_0^{L,u}(y+u)\,\phi(y; 0, I^{-1}),$$

where

$$q_0^{L,u}(x) = E[f_1^{L,u} \mid f_0 = x].$$

Remark 4.2. By definition

$$q^{L,u}(z) = E[q_0^{L,u}(f_0) \mid f_0^{L,u} = z].$$

If $k=1$ and $u \neq 0$, $q_0^{L,u}(z) = q^{L,u}(Iuz - \frac{1}{2}Iu^2)$.

5. Examples

5.1. *Diffusion Processes Perturbed by Small Noise*

Let $X^{\varepsilon,\theta}$ be a diffusion process defined by the stochastic differential equation

$$dX_t^{\varepsilon,\theta} = V_0(X_t^{\varepsilon,\theta}, \theta)\,dt + \varepsilon V(X_t^{\varepsilon,\theta})\,dw_t, \qquad t \in [0, T],$$
$$X_0^{\varepsilon,\theta} = x_0,$$

$\varepsilon \in (0, 1)$, where θ is a k-dimensional unknown parameter in Θ, $T > 0$ and x_0 are constants, $V = (V_1, \ldots, V_r)$ 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. We consider the parameter estimation problem for θ from observations $\{X_t^{\varepsilon,\theta};\ 0 \leqslant t \leqslant T\}$.

The Radon–Nikodym derivative of P_θ^ε with respect to $P_{\theta_0}^\varepsilon$ is given by the formula (e.g., Liptser and Shiryayev [6]):

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

where

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

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.

We assume the following conditions:

(1) V_0, V and $(VV')^+$ are smooth in (x, θ).

(2) For $\mathbf{n} \in \bar{\mathbf{N}}^k$ with $|\mathbf{n}| \geqslant 1$,

$$\sup_{x,\,\theta} \{|\partial^{\mathbf{n}} V_0| + |\partial^{\mathbf{n}} V| + |\partial^{\mathbf{n}}(VV')^+|\} < \infty.$$

(3) For $|v| \geqslant 1$ and $|\mathbf{n}| \geqslant 0$, a constant $C_{v,\,\mathbf{n}}$ exists and

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

for all x.

(4) For $\theta_0 \in \Theta^o$, there exists $a_0 > 0$ such that

$$\int_0^T [V_0(X_t^{0,\,\theta_0}, \theta) - V_0(X_t^{0,\,\theta_0}, \theta_0)]' \, (VV')^+ \, (X_t^{0,\,\theta_0})$$

$$\times [V_0(X_t^{0,\,\theta_0}, \theta) - V_0(X_t^{0,\,\theta_0}, \theta_0)] \, dt \geqslant a_0 \, |\theta - \theta_0|^2$$

for $\theta \in \Theta$. ($X_t^{0,\,\theta_0}$ is the solution of the differential equation for $\varepsilon = 0$ and $\theta = \theta_0$.)

It is possible to verify the Conditions (C1)–(C3) and (C5). Then we obtain, for example, the asymptotic expansion of the distribution of the bias corrected maximum likelihood estimator under the contiguous alternative $P^{\varepsilon}_{\theta_0 + \varepsilon u}$.

THEOREM 5.1. *The probability distribution of the bias corrected maximum likelihood estimator* $\hat{\theta}^*_{\varepsilon}(w; \theta_0 + \varepsilon u)$ *under the contiguous alternative* $P^{\varepsilon}_{\theta_0 + \varepsilon u}$ *has the asymptotic expansion*

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

as $\varepsilon \downarrow 0$, $A \in \mathbf{B}^k$, $u \in \mathbf{R}^k$, *where* $p_0^{c,\,u}, p_1^{c,\,u}, \ldots$ *are smooth functions. The expansion is uniform in* $A \in \mathbf{B}^k$ *and* $u \in K$, *where* K *is any compact set in* $\mathbf{R}^k$. *In particular,*

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

$$\begin{aligned} p_1^{c,u}(y) = A_{i,j,l}[-y^i y^j y^l - u^l y^i y^j + I^{ij} y^l + I^{ij} u^l]\,\phi(y; 0, I^{-1}) \\ + B_{i,j,l}[-\tfrac{1}{2} y^i y^j y^l - u^i y^j y^l + I^{il} y^j + I^{il} u^j]\,\phi(y; 0, I^{-1}) \\ - b(\theta_0)^j I_{jl} y^l \phi(y; 0, I^{-1}), \end{aligned}$$

where $A_{i,j,l}$ and $B_{i,j,l}$ are constants determined by V_0, V, x_0, and T.

For details see [12].

5.2. *Models with a Discrete Time Parameter*

A Gaussian AR(k) process (X_t^ε) with small noise is defined by

$$\phi_0(B)\, X_t^\varepsilon = \varepsilon e_t, \qquad t = k, k+1, ..., k+n-1,$$

$$X_0^\varepsilon = x_0, ..., \qquad X_{k-1}^\varepsilon = x_{k-1}, \qquad \varepsilon \in [0, 1],$$

where $\phi_0(z) = 1 - \phi_0^1 z - \cdots - \phi_0^k z^k$, $\phi_0^i \in \mathbf{R}$, B is the backward shift operator, $x_0, ..., x_{k-1}$ are constants and $e_t \sim N(0, 1)$ independently. We may construct this AR(k) model on the Wiener space if we take $e_t = w(t-k+1) - w(t-k)$ for $t = k, k+1, ..., k+n-1$. Let $\phi(z) = 1 - \phi^1 z - \cdots - \phi^k z^k$. Then we have

$$\begin{aligned} G(w, \varepsilon, \phi; \phi_0) = -\varepsilon \sum_{t=k}^{k+n-1} (\phi(B) - \phi_0(B))\, X_t^\varepsilon \cdot e_t \\ - \tfrac{1}{2} \sum_{t=k}^{k+n-1} [(\phi(B) - \phi_0(B))\, X_t^\varepsilon]^2. \end{aligned}$$

Let X_t^0 denote the solution for $\varepsilon = 0$ and ϕ_0. Assume that $\sum_{t=k}^{k+n-1} (\phi(B)\, X_t^0)^2 > 0$ for $\phi \neq \phi_0$. It is not difficult to verify Conditions (C1)–(C4).

Example (AR(1) process). Let $X_t^\varepsilon, t = 1, 2, ..., n$, be defined by the difference equation

$$X_t^\varepsilon - \theta_0 X_{t-1}^\varepsilon = \varepsilon e_t,$$

$$X_0^\varepsilon = x_0,$$

where $\varepsilon \in [0, 1]$, x_0 is a constant, $x_0 \neq 0$, and $\{e_t, t = 1, 2, ..., n\}$ is a Gaussian white noise with $E[e_t] = 0$ and $\mathrm{Var}(e_t) = 1$. The experiments generated from this model can be realized on a Wiener space if we take $e_t = w(t) - w(t-1)$. Then

$$G(w, \varepsilon, \theta; \theta_0) = \varepsilon \sum_{t=1}^{n} (\theta - \theta_0) X_{t-1}^{\varepsilon} e_t - \tfrac{1}{2} \sum_{t=1}^{n} (\theta - \theta_0)^2 X_{t-1}^{\varepsilon 2}.$$

We have

$$I(\theta_0) = x_0^2 \sum_{t=1}^{n} \theta_0^{2t-2},$$
$$X_t^0 = x_0 \theta_0^t,$$

and

$$X_t^{\varepsilon} = x_0 \theta_0^t + \varepsilon D_t,$$

where

$$D_0 = 0, D_t = \left(\frac{\partial}{\partial \varepsilon}\right)_0 X_t^{\varepsilon} = \sum_{l=1}^{t} \theta_0^{t-l} e_l, \qquad t = 1, 2, ..., n.$$

Also we have

$$G_{1;\,1} = \sum_{t=1}^{n} X_{t-1}^0 e_t = \sum_{t=1}^{n} x_0 \theta_0^{t-1} e_t,$$
$$G_{2;\,1} = 2 \sum_{t=1}^{n} D_{t-1} e_t = 2 \sum_{t=2}^{n} \sum_{t=1}^{t-1} \theta_0^{t-l-1} e_l e_t,$$
$$G_{1;\,1,\,1} = -2 \sum_{t=1}^{n} X_{t-1}^0 D_{t-1} = -2 \sum_{t=2}^{n} \sum_{t=1}^{t-1} x_0 \theta_0^{2t-l-2} e_l,$$
$$G_{0;\,1,\,1,\,1} = 0.$$

From these equations, we obtain

$$f_0^1 = I^{-1} x_0 \sum_{t=1}^{n} \theta_0^{t-1} e_t,$$
$$f_1^1 = I^{-1} \sum_{t=2}^{n} \sum_{l=1}^{t-1} \theta_0^{t-l-1} e_l e_t - 2I^{-1} x_0 \sum_{t=2}^{n} \sum_{l=1}^{t-1} \theta_0^{2t-l-2} e_l f_0^1 - b(\theta_0)^1,$$
$$f_0^{L,\,u} = x_0 \sum_{t=1}^{n} \theta_0^{t-1} e_t u - \tfrac{1}{2} I u^2,$$
$$f_1^{L,\,u} = \sum_{t=2}^{n} \sum_{l=1}^{t-1} \theta_0^{t-l-1} e_l e_t u - x_0 \sum_{t=2}^{n} \sum_{l=1}^{t-1} \theta_0^{2t-l-2} e_l u^2.$$

Let (a_t), $t=1, 2, ..., n$, be a finite sequence of real numbers. Then

$$E\left[e_t \,\middle|\, \sum_{t=1}^{n} a_t e_t = x \right] = \|a\|^{-2} a_t x,$$

where $\|a\|^2 = \sum_{t=1}^{n} a_t^2$, and

$$E\left[e_s, e_t \,\middle|\, \sum_{t=1}^{n} a_t e_t = x \right] = \delta_{t,s} + \|a\|^{-4} a_s a_t (x^2 - \|a\|^2).$$

Let

$$a_t = x_0 I^{-1} \theta_0^{t-1}.$$

We then have

$$\begin{aligned} q_b^1(x) &= E[f_1^1 \mid f_0^1 = x] \\ &= -x_0^2 I^{-1} \sum_{t=1}^{n} (t-1)\, \theta_0^{2t-3} (x^2 + I^{-1}) - b(\theta_0)^1 \end{aligned}$$

and

$$\begin{aligned} q_0^{L,u}(x) &= E[f_1^{L,u} \mid f_0 = x] \\ &= x_0^2 \sum_{t=1}^{n} (t-1)\, \theta_0^{2t-3} (x^2 u - I^{-1} u - x u^2). \end{aligned}$$

Put

$$c = x_0^2 \sum_{t=1}^{n} (t-1)\, \theta_0^{2t-3}.$$

Then we have the asymptotic expansions:

$$\begin{aligned} & P[\varepsilon^{-1}(\hat{\theta}_\varepsilon^*(w; \theta_0) - \theta_0) \leqslant x] \\ &\qquad \sim \Phi(x; 0, I^{-1}) + \varepsilon[I^{-1} c x^2 + I^{-2} c + b]\, \phi(x; 0, I^{-1}) + \cdots, \\ & P[\varepsilon^{-1}(\hat{\theta}_\varepsilon^*(w; \theta_0 + \varepsilon u) - \theta_0 - \varepsilon u) \leqslant y] \\ &\qquad \sim \Phi(y; 0, I^{-1}) + \varepsilon[I^{-1} c y^2 + I^{-2} c u y + I^{-2} c + b]\, \phi(y; 0, I^{-1}) + \cdots, \\ & P[\varepsilon^{-2} G(w, \varepsilon, \theta_0 + \varepsilon u; \theta_0) \leqslant x] \\ &\qquad \sim \Phi(\bar{x}; 0, J) - \varepsilon c[I^{-2} u^{-1} \bar{x}^2 - I^{-1} u - I^{-1} u \bar{x}]\, \phi(\bar{x}; 0, J) + \cdots, \\ & P[\varepsilon^{-2} G(w, \varepsilon, \theta_0 + \varepsilon u; \theta_0 + \varepsilon u) \leqslant x] \\ &\qquad \sim \Phi(\underline{x}; 0, J) - \varepsilon c[I^{-2} u^{-1} \bar{x}^2 - I^{-1} u]\, \phi(\underline{x}; 0, J) + \cdots, \end{aligned}$$

where $\bar{x} = x + \frac{1}{2} J$, $\underline{x} = x - \frac{1}{2} J$ and $J = I u^2$.

In a similar way we can treat nonlinear time series models such as

$$X_t = f_t(X_0, ..., X_{t-1}, \theta) + g_t(\varepsilon e_t),$$

where f_t are given functions and g_t are transforms of $\mathbf{R}$. AR models are of this type. Another example is estimation for a signal from contaminated observations X_t given by

$$X_t = S_t(\theta) + \varepsilon e_t, \qquad t = 1, 2, ..., n.$$

If for some $a > 0, \sum_t (S_t(\theta) - S_t(\theta_0))^2 \geqslant a\, |\theta - \theta_0|^2$ $(\theta, \theta_0 \in \Theta)$, then we may obtain the asymptotic expansion for the maximum likelihood estimator.

5.3. *Second-Order Asymptotic Efficiency of Maximum Likelihood Estimators*

We return to the general model defined in Section 1. For simplicity, let $k = 1$.

An estimator T_ε is said to be second–order asymptotically median unbiased (second-order AMU) if for any $\theta_0 \in \Theta^o$ and any $c > 0$,

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

and

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

See Akahira and Takeuchi [1].

Given a second-order AMU estimator T_ε, if

$$\lim_{\varepsilon \downarrow 0} \varepsilon^{-1}\, |P^\varepsilon_{\theta_0}[\varepsilon^{-1}(T_\varepsilon - \theta_0) \leqslant u] - G_0(u, \theta_0) - \varepsilon G_1(u, \theta_0)| = 0,$$

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

Now consider testing hypothesis $H^+: \theta = \theta_0 + \varepsilon u$ against $K: \theta = \theta_0$, where u is any positive number. Let $c_\varepsilon = \frac{1}{2} J + \varepsilon p + q_\varepsilon$, where

$$p = q^{L,\,u}(\tfrac{1}{2} J) - \sqrt{2\pi J} \int_{-\infty}^{(1/2)\,J} q^{L,\,u}(z)\, \phi(z; \tfrac{1}{2} J, J)\, dz$$

and

$$q_\varepsilon = \varepsilon^{3/2} + \phi(0; 0, J)^{-1}\, |P^\varepsilon_{\theta_0 + \varepsilon u}[\varepsilon^{-1}(T_\varepsilon - \theta_0 - \varepsilon u) \leqslant 0] - \tfrac{1}{2}|.$$

Hence, $q_\varepsilon = o(\varepsilon)$. From Theorem 3.2, we see

$$\begin{aligned}
P[\varepsilon^{-2}G(w, \varepsilon, \theta_0+\varepsilu; \theta_0+\varepsilon u) \leqslant c_\varepsilon]
&= \tfrac{1}{2} + \varepsilon\phi(0; 0, J)\, p + \varepsilon \left\{ -q^{L,u}(\tfrac{1}{2}J)\, \phi(0; 0, J) \right. \\
&\quad \left. + \int_{-\infty}^{(1/2)J} q^{L,u}(z)\, \phi(z; \tfrac{1}{2}J, J)\, dz \right\} \\
&\quad + q_\varepsilon \phi(0; 0, J) + O(\varepsilon^2).
\end{aligned}$$

Then we have

$$P[\varepsilon^{-2}G(w, \varepsilon, \theta_0+\varepsilon u; \theta_0+\varepsilon u) \leqslant c_\varepsilon] > P^\varepsilon_{\theta_0+\varepsilon u}[\varepsilon^{-1}(T_\varepsilon - \theta_0 - \varepsilon u) \leqslant 0],$$

and by Neyman–Pearson's lemma

$$P[\varepsilon^{-2}G(w, \varepsilon, \theta_0+\varepsilon u; \theta_0) \leqslant c_\varepsilon] \geqslant P^\varepsilon_{\theta_0}[\varepsilon^{-1}(T_\varepsilon - \theta_0) \leqslant u]$$

for small ε. Therefore, by Theorem 3.1 for $u > 0$,

$$\begin{aligned}
\liminf_{\varepsilon \downarrow 0} \varepsilon^{-1} &\left\{ \Phi(\tfrac{1}{2}J; -\tfrac{1}{2}J, J) - \varepsilon e^{-(1/2)J} \int_{-\infty}^{(1/2)J} q^{L,u}(z)\, \phi(z; \tfrac{1}{2}J, J)\, dz \right. \\
&\left. - P^\varepsilon_{\theta_0}[\varepsilon^{-1}(T_\varepsilon - \theta_0) \leqslant u] \right\} \geqslant 0.
\end{aligned}$$

Similarly, for $u < 0$ we consider testing hypothesis $H^-: \theta = \theta_0 + \varepsilon u$ against $K: \theta = \theta_0$. Define c_ε as above with the same p and

$$q_\varepsilon = \varepsilon^{3/2} + \phi(0; 0, J)^{-1}\, |P^\varepsilon_{\theta_0+\varepsilon u}[\varepsilon^{-1}(T_\varepsilon - \theta_0 - \varepsilon u) \geqslant 0] - \tfrac{1}{2}|.$$

Then, again by Theorem 3.2, we see

$$P[\varepsilon^{-2}G(w, \varepsilon, \theta_0+\varepsilon u; \theta_0+\varepsilon u) \leqslant c_\varepsilon] > P^\varepsilon_{\theta_0+\varepsilon u}[\varepsilon^{-1}(T_\varepsilon - \theta_0 - \varepsilon u) > 0]$$

for small ε, and by Neyman–Pearson's lemma

$$P[\varepsilon^{-2}G(w, \varepsilon, \theta_0+\varepsilon u; \theta_0) \leqslant c_\varepsilon] \geqslant P^\varepsilon_{\theta_0}[\varepsilon^{-1}(T_\varepsilon - \theta_0) > u],$$

or equivalently

$$1 - P[\varepsilon^{-2}G(w, \varepsilon, \theta_0+\varepsilon u; \theta_0) \leqslant c_\varepsilon] \leqslant P^\varepsilon_{\theta_0}[\varepsilon^{-1}(T_\varepsilon - \theta_0) \leqslant u]$$

for small ε. Consequently, we have for $u<0$

$$\limsup_{\varepsilon \downarrow 0} \varepsilon^{-1} \left\{ \Phi(-\tfrac{1}{2}J; \tfrac{1}{2}J, J) + \varepsilon e^{-(1/2)J} \int_{-\infty}^{(1/2)J} q^{L,u}(z)\, \phi(z; \tfrac{1}{2}J, J)\, dz \right.$$
$$\left. - P^{\varepsilon}_{\theta_0}[\varepsilon^{-1}(T_\varepsilon - \theta_0) \leqslant u] \right\} \leqslant 0.$$

In this sense

$$\Phi(\tfrac{1}{2}J; -\tfrac{1}{2}J, J) - \varepsilon e^{-(1/2)J} \int_{-\infty}^{(1/2)J} q^{L,u}(z)\, \phi(z; \tfrac{1}{2}J, J)\, dz$$

for $u>0$, and

$$\Phi(-\tfrac{1}{2}J; \tfrac{1}{2}J, J) + \varepsilon e^{-(1/2)J} \int_{-\infty}^{(1/2)J} q^{L,u}(z)\, \phi(z; \tfrac{1}{2}J, J)\, dz$$

for $u<0$ are called the bounds of second-order distributions. An AMU estimator attaining these bounds for any $u>0$ and $u<0$ is said to be second-order efficient.

PROPOSITION 5.1. *Let* $\dim(\Theta)=1$ *and assume Conditions* (C1)–(C3) *and* (C5) (*or* (C4)). *Then the second-order AMU bias corrected maximum likelihood estimator is second-order efficient.*

Proof. From Theorem 4.2 we have

$$P(\varepsilon^{-1}(\hat{\theta}^{*}_{\varepsilon}(w; \theta_0 + \varepsilon u) - (\theta_0 + \varepsilon u)) \leqslant 0)$$
$$\sim \tfrac{1}{2} + \varepsilon \left[-q^1_b(u)\, \phi(0; 0, I^{-1}) + \int_{-\infty}^{0} q^{L,u}_0(y+u)\, \phi(y; 0, I^{-1})\, dy \right] + \cdots.$$

Therefore, for the second-order AMU bias-corrected maximum likelihood estimator $\hat{\theta}^{*}_{\varepsilon}(w; \theta_0)$,

$$-q^1_b(u)\, \phi(0; 0, I^{-1}) + \int_{-\infty}^{0} q^{L,u}_0(y+u)\, \phi(y; 0, I^{-1})\, dy = 0$$

and

$$q^1_b(u)\, \phi(0; 0, I^{-1}) + \int_{0}^{\infty} q^{L,u}_0(y+u)\, \phi(y; 0, I^{-1})\, dy = 0.$$

From Theorem 4.1, it is not difficult to show this bias-corrected maximum likelihood estimator attains the bounds of the second-order distributions.

EXAMPLE. For the diffusion process of Section 5.1 with $k=1$, the bias corrected maximum likelihood estimator corresponding to

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

is second-order AMU and therefore second-order efficient.

EXAMPLE. Consider the AR(1) model in Section 5.2. If

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

the bias-corrected maximum likelihood estimator is second-order AMU and therefore second-order asymptotically efficient, which is the consequence of Proposition 5.1 or is proved by comparing the bounds of the second-order distributions with the expansions above for the bias-corrected maximum likelihood estimator.

6. PROOF OF THEOREMS 3.1 AND 3.2

To show Theorem 3.1, we prepare two lemmas.

LEMMA 6.1. *Assume that* (C1) *and* (C2) *are satisfied. For any compact set* $K \subset \mathbf{R}^k$, $\varepsilon^{-2} G(w, \varepsilon, \theta_0 + \varepsilon u; \theta_0)$ *has the asymptotic expansion*

$$\varepsilon^{-2} G(w, \varepsilon, \theta_0 + \varepsilon u; \theta_0) \sim f_0^{L,u} + \varepsilon f_1^{L,u} + \varepsilon^2 f_2^{L,u} + \cdots$$

in D^∞ *as* $\varepsilon \downarrow 0$ *uniformly in* $u \in K$ *with* $f_0^{L,u}, f_1^{L,u}, \ldots \in D^\infty$.

Proof. We can prove this lemma from (C1) and the Taylor expansion using $G(w, \varepsilon, \theta_0; \theta_0) = 0$ and $\delta G(0, \theta_0; \theta_0) = 0$.

The sequence $(P^\varepsilon_{\theta_0 + \varepsilon u})$ is contiguous to the sequence $(P^\varepsilon_{\theta_0})$ (Le Cam [5]). From (C3) the Malliavin covariance of $\delta_0 \delta G(w, 0, \theta_0; \theta_0)[u]$ is $I(\theta_0)[u, u]$. From Lemma 6.1 we see the Malliavin covariance $\sigma_G(\varepsilon)$ of $\varepsilon^{-2} G(w, \varepsilon, \theta_0 + \varepsilon u; \theta_0)$ has the asymptotic expansion

$$\sigma_G(\varepsilon) \sim J + \varepsilon g_1 + \varepsilon^2 g_2 + \cdots$$

in D^∞ as $\varepsilon \downarrow 0$ with $g_1, g_2, \ldots \in D^\infty$, where $J = I(\theta_0)[u, u]$. The Wiener functional $\psi_\varepsilon^{L,u}(w) = \psi(c\, |\sigma_G(\varepsilon) - J|^2)$, $c = (J/2)^{-2}$, can be used as truncation for nondegeneracy of the Malliavin covariance. Thus applying Theorem 4.1 of [13] (or refer to Proposition 6.2 of Takanobu and Watanabe [8], Yoshida [12, 14]), we have the following lemma.

LEMMA 6.2. *Assume that* (C1)–(C3) *are satisfied. For* $u \in \mathbf{R}^k - \{0\}$, *the generalized Wiener functional*

$$\psi_\varepsilon^{L,u}(w)\, I_A(\varepsilon^{-2}G(w, \varepsilon, \theta_0 + \varepsilon u; \theta_0)),$$

$A \in \mathbf{B}^1$, *is well defined and has the asymptotic expansion*

$$\psi_\varepsilon^{L,u}(w)\, I_A(\varepsilon^{-2}G(w, \varepsilon, \theta_0 + \varepsilon u; \theta_0)) \sim \Phi_{A,0}^{L,u} + \varepsilon \Phi_{A,1}^{L,u} + \cdots$$

in $\tilde{D}^{-\infty}$ *as* $\varepsilon \downarrow 0$ *uniformly in* $A \in \mathbf{B}^1$ *with* $\Phi_{A,0}^{L,u}, \Phi_{A,1}^{L,u}, \ldots \in \tilde{D}^{-\infty}$ *determined by the formal Taylor expansion*

$$\begin{aligned} & I_A(f_0^{L,u} + \varepsilon[f_1^{L,u} + \varepsilon f_2^{L,u} + \cdots]) \\ &\quad = \sum_n \frac{1}{n!} \partial^n I_A(f_0^{L,u})[\varepsilon f_1^{L,u} + \varepsilon^2 f_2^{L,u} + \cdots]^n \\ &\quad = \Phi_{A,0}^{L,u} + \varepsilon \Phi_{A,1}^{L,u} + \cdots, \end{aligned}$$

where $\partial = \partial/\partial x$. *In particular,*

$$\begin{aligned} \Phi_{A,0}^{L,u} &= I_A(f_0^{L,u}), \\ \Phi_{A,1}^{L,u} &= f_1^{L,u} \partial I_A(f_0^{L,u}). \end{aligned}$$

Proof of Theorem 3.1. From Lemma 6.2 and the fact that $\psi_\varepsilon^{L,u}(w) = 1 - O(\varepsilon^n)$ in D^∞ as $\varepsilon \downarrow 0$ for any $n \in \mathbf{N}$, we have the asymptotic expansion

$$P(\varepsilon^{-2}G(w, \varepsilon, \theta_0 + \varepsilon u; \theta_0) \in A) \sim E[\Phi_{A,0}^{L,u}] + \varepsilon E[\Phi_{A,1}^{L,u}] + \cdots$$

as $\varepsilon \downarrow 0$. Using the integration-by-parts formula in the Malliavin calculus we see that

$$E[\Phi_{A,i}^{L,u}] = E[\tilde{\Phi}_{A,i}^{L,u} I_A(f_0^{L,u})]$$

for some $\tilde{\Phi}_{A,i}^{L,u} \in D^\infty$, $i = 0, 1, \ldots$ Consequently, each term in the asymptotic expansion is represented by an integration of some smooth function $p_i^{L,u}(x)$ over A. We only have to calculate $p_0^{L,u}$ and $p_1^{L,u}$. $p_0^{L,u}$ is easy. As above integration-by-parts yields

$$\begin{aligned} E[\Phi_{A,1}^{L,u}] &= E[f_1^{L,u} \partial I_A(f_0^{L,u})] \\ &= E[\Psi(w; f_1^{L,u})\, I_A(f_0^{L,u})] = \int_A p_1^{L,u}(x)\, dx, \end{aligned}$$

where $\Psi(w; f_1^{L,u})$ is a certain smooth Wiener functional and $p_1^{L,u}(x) = E[\Psi(w; f_1^{L,u}) \mid f_0^{L,u} = x]\, p_0^{L,u}(x)$ is a smooth function. If $A = (x, \infty)$,

$$E[f_1^{L,u} \partial I_A(f_0^{L,u})] = E[f_1^{L,u} \delta_x(f_0^{L,u})] = E[f_1^{L,u} \mid f_0^{L,u} = x]\, p_0^{L,u}(x).$$

Therefore,

$$p_1^{L,u}(x) = -\partial \{ E[f_1^{L,u} \mid f_0^{L,u} = x]\, p_0^{L,u}(x) \}.$$

Next, we discuss asymptotic expansions under contiguous alternatives. Condition (C5) is assumed.

LEMMA 6.3. *Assume that* (C1)–(C3) *and* (C5) *are satisfied. Let* $\theta_0 \in \Theta^o$ *and let* K *be any compact set in* $\mathbf{R}^k$. *Then there exist Wiener functionals* $\bar{\varphi}_\varepsilon^K(w) \in D^\infty$, $\varepsilon \in (0, 1)$, *such that the following conditions are satisfied for* $\bar{\varphi}_\varepsilon^K(w)$ *and* $\phi_\varepsilon^{K,u}(w) := \bar{\varphi}_\varepsilon^K(w)\, \varphi_\varepsilon^u(F_{\theta_0}^\varepsilon(w))$, $u \in K$:

(i) $0 \leqslant \phi_\varepsilon^{K,u}(w) \leqslant \bar{\varphi}_\varepsilon^K(w) \leqslant 1$.

(ii) *For any* $n \in \mathbf{N}$, $\bar{\varphi}_\varepsilon^K(w) = 1 - O(\varepsilon^n)$ *in* D^∞ *as* $\varepsilon \downarrow 0$, *and also* $\phi_\varepsilon^{K,u}(w) = 1 - O(\varepsilon^n)$ *in* D^∞ *as* $\varepsilon \downarrow 0$ *uniformly in* $u \in K$.

(iii) *If Wiener functionals* $\zeta_\varepsilon^u(w)$ *depending on* $u \in K$ *satisfy* $E[|\zeta_\varepsilon^u(w) - 1|^p] = O(\varepsilon^n)$ *as* $\varepsilon \downarrow 0$ *uniformly in* $u \in K$ *for any* $p > 1$ *and any* $n \in \mathbf{N}$, *then*

$$E[|\zeta_\varepsilon^u(w) - 1|\, \varphi_\varepsilon^u(F_{\theta_0}^\varepsilon(w)) \exp\{\varepsilon^{-2} G(w, \varepsilon, \theta_0 + \varepsilon u; \theta_0)\}] = O(\varepsilon^n)$$

as $\varepsilon \downarrow 0$ *uniformly in* $u \in K$ *for any* $n \in \mathbf{N}$. *In particular, for any* $n \in \mathbf{N}$, *uniformly in* $u \in K$,

$$E[|\phi_\varepsilon^{K,u}(w) - \varphi_\varepsilon^u(F_{\theta_0}^\varepsilon(w))| \exp\{\varepsilon^{-2} G(w, \varepsilon, \theta_0 + \varepsilon u; \theta_0)\}] = O(\varepsilon^n)$$

as $\varepsilon \downarrow 0$.

(iv) *For any* $p > 1$,

$$\sup_{\substack{\varepsilon \in (0,1) \\ u \in K \\ \theta_0 + \varepsilon u \in \Theta}} E[1_{\{\bar{\varphi}_\varepsilon^K(w) > 0\}} \exp\{p \varepsilon^{-2} G(w, \varepsilon, \theta_0 + \varepsilon u; \theta_0)\}] < \infty.$$

Proof. Let

$$\rho_\varepsilon(u) = \varepsilon^{-2} G(w, \varepsilon, \theta_0 + \varepsilon u; \theta_0) - \delta_0 \delta G(w, 0, \theta_0; \theta_0)[u] + \tfrac{1}{2} I(\theta_0)[u, u].$$

Choose $r > 0$ so that $K \subset B(r)$, where $B(r) = \{u \in \mathbf{R}^k;\ |u| < r\}$. Let $\chi_\varepsilon(w) = \| \rho_\varepsilon(\cdot) \|_{W^{1,2}(B(r))}^2$ ($W^{1,2}(B(r))$ is a Sobolev space. For definition see Section 7.) and let $\bar{\varphi}_\varepsilon^K(w) = \psi_1(\chi_\varepsilon(w))$, where $\psi_1 : \mathbf{R} \to \mathbf{R}$ is a smooth function which satisfies the same conditions as ψ and also satisfies that $\psi_1(x) > 0$ if $|x| < 1$. Thus (i) holds. By the Taylor formula

$$\rho_\varepsilon(u) = \varepsilon \int_0^1 \frac{(1-s)^2}{2} \left(\frac{\partial}{\partial t}\right)^3 \{G(w, t, \theta_0 + tu; \theta_0)\} \big|_{t=s\varepsilon} ds.$$

Therefore it is easy to show (ii) using (C1). If $\bar{\varphi}_\varepsilon^K(w) > 0$, $\chi_\varepsilon(w) < 1$ and by Sobolev's inequality $\sup_{u \in B(r)} |\rho_\varepsilon(u)| < a$ for some $a > 0$. Hence, if $\bar{\varphi}_\varepsilon^K(w) > 0$,

$$\begin{aligned} &\exp\{\varepsilon^{-2} G(w, \varepsilon, \theta_0 + \varepsilon u; \theta_0)\} \\ &\qquad \leqslant \exp\{\delta_0 \delta G(w, 0, \theta_0; \theta_0)[u] - \tfrac{1}{2} I(\theta_0)[u, u] + a\}. \end{aligned}$$

By (C3), $\delta_0 \delta G(w, 0, \theta_0; \theta_0)$ is Gaussian and we obtain (iv). For $q_0 = p_0/(p_0 - 1)$ and any $n \in \mathbf{N}$, $E[|\zeta_\varepsilon^u - 1|^{q_0}] = O(\varepsilon^n)$. Therefore, we obtain the first equation of (iii) using (iv) of (C5) and the Hölder inequality. Taking $\zeta_\varepsilon^u(w) = \bar{\varphi}_\varepsilon^K(w)$, we obtain the second one. ∎

We fix $\bar{\varphi}_\varepsilon^K(w)$ as in the proof of Lemma 6.3.

LEMMA 6.4. *Assume that* (C1)–(C3) *and* (C5) *are satisfied. Let* $\theta_0 \in \Theta^o$ *and let* K *be any compact set in* $\mathbf{R}^k$. *Then, for* $u \in K$,

(1) $\phi_\varepsilon^{K,u}(w) \exp\{\varepsilon^{-2} G(w, \varepsilon, \theta_0 + \varepsilon u; \theta_0)\}$ *has the asymptotic expansion*

$$\phi_\varepsilon^{K,u}(w) \exp\{\varepsilon^{-2} G(w, \varepsilon, \theta_0 + \varepsilon u; \theta_0)\} \sim e^{f_0^{L,u}}(1 + \varepsilon \Psi_1^u + \varepsilon^2 \Psi_2^u + \cdots)$$

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

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

This expansion is uniform in $u \in K$

(2) *Suppose that* $\Phi_\lambda(w, \varepsilon) \in D^{-\infty}$ (*resp.* $\tilde{D}^{-\infty}$), $\varepsilon \in (0, 1)$, $\lambda \in \Lambda$ (*an index set*) *has the asymptotic expansion*

$$\Phi_\lambda(w, \varepsilon) \sim \Phi_{\lambda,0} + \varepsilon \Phi_{\lambda,1} + \cdots$$

in $D^{-\infty}$ (*resp.* $\tilde{D}^{-\infty}$) *as* $\varepsilon \downarrow 0$ *uniformly in* $\lambda \in \Lambda$ *with* $\Phi_{\lambda,0}, \Phi_{\lambda,1}, \ldots \in D^{-\infty}$ (*resp.* $\tilde{D}^{-\infty}$). *Then*

$$\phi_\varepsilon^{K,u}(w) \exp\{\varepsilon^{-2} G(w, \varepsilon, \theta_0 + \varepsilon u; \theta_0)\}\ \Phi_\lambda(w, \varepsilon)$$

has the asymptotic expansion

$$\phi_\varepsilon^{K,u}(w) \exp\{\varepsilon^{-2} G(w, \varepsilon, \theta_0 + \varepsilon u; \theta_0)\}\ \Phi_\lambda(w, \varepsilon) \sim e^{f_0^{L,u}}(\bar{\Phi}_{\lambda,0}^u + \varepsilon \bar{\Phi}_{\lambda,1}^u + \cdots)$$

in $D^{-\infty}$ *(resp.* $\tilde{D}^{-\infty}$*) as* $\varepsilon\downarrow 0$ *uniformly in* $u\in K$ *and* $\lambda\in\Lambda$ *with* $\bar{\Phi}^u_{\lambda,0}, \bar{\Phi}^u_{\lambda,1},\ldots\in D^{-\infty}$ *(resp.* $\tilde{D}^{-\infty}$*) determined by the formal Taylor expansion*

$$(1+\varepsilon\Psi^u_1+\varepsilon^2\Psi^u_2+\cdots)(\Phi_{\lambda,0}+\varepsilon\Phi_{\lambda,1}+\cdots)=\bar{\Phi}^u_{\lambda,0}+\varepsilon\bar{\Phi}^u_{\lambda,1}+\cdots.$$

In particular,

$$\bar{\Phi}^u_{\lambda,0}=\Phi_{\lambda,0}$$
$$\bar{\Phi}^u_{\lambda,1}=\Phi_{\lambda,1}+\Phi_{\lambda,0}\Psi^u_1.$$

(3) *Let* $u\in\mathbf{R}^k-\{0\}$ *with* $u\in K$. *Then*

$$\phi^{K,u}_\varepsilon(w)\exp\{\varepsilon^{-2}G(w,\varepsilon,\theta_0+\varepsilon u;\theta_0)\}\,\psi^{L,u}_\varepsilon(w)\,I_A(\varepsilon^{-2}G(w,\varepsilon,\theta_0+\varepsilon u;\theta_0))$$
$$\sim e^{f^{L,u}_0}[\bar{\Phi}^{L,u}_{A,0}+\varepsilon\bar{\Phi}^{L,u}_{A,1}+\cdots]$$

in $\tilde{D}^{-\infty}$ *as* $\varepsilon\downarrow 0$ *uniformly in* $A\in\mathbf{B}^1$ *with* $\bar{\Phi}^{L,u}_{A,0}, \bar{\Phi}^{L,u}_{A,1},\ldots\in\tilde{D}^{-\infty}$ *determined by the formal Taylor expansion*

$$(1+\varepsilon\Psi^u_1+\varepsilon^2\Psi^u_2+\cdots)(\Phi^{L,u}_{A,0}+\varepsilon\Phi^{L,u}_{A,1}+\cdots)=\bar{\Phi}^{L,u}_{A,0}+\varepsilon\bar{\Phi}^{L,u}_{A,1}+\cdots$$

In particular,

$$\bar{\Phi}^{L,u}_{A,0}=\Phi^{L,u}_{A,0},\qquad \bar{\Phi}^{L,u}_{A,1}=\Phi^{L,u}_{A,1}+f^{L,u}_1\Phi^{L,u}_{A,0}.$$

Proof. Using Lemmas 6.1 and 6.3 (iv), we obtain (1) as in the proof of Lemma 4.5 (2) of [12] (Note that for any $m\in\bar{\mathbf{N}}$, $D^m\phi^{K,u}_\varepsilon(w)=D^m\phi^{K,u}_\varepsilon(w)\cdot I_{\{\bar{\varphi}^K_\varepsilon(w)>0\}}$). (2) follows from (1) and Theorem 2.2 (ii) of Watanabe [11]. Lemma 6.2 and (2) lead us to (3).

Proof of Theorem 3.2. By (iii), (iv) of (C5), Lemmas 6.3, 6.4, and the fact that $\psi^{L,u}_\varepsilon(w)=1-O(\varepsilon^n)$ in D^∞ for any $n\in\mathbf{N}$, we have

$$P(\varepsilon^{-2}G(w,\varepsilon,\theta_0+\varepsilon u;\theta_0+\varepsilon u)\in A)$$
$$=E\left[I_A\left(\log\frac{dP^\varepsilon_{\theta_0+\varepsilon u}}{dP^\varepsilon_{\theta_0}}(F^\varepsilon_{\theta_0+\varepsilon u}(w))\right)\right]$$
$$\sim E\left[\varphi^u_\varepsilon(F^\varepsilon_{\theta_0+\varepsilon u}(w))\,I_A\left(\log\frac{dP^\varepsilon_{\theta_0+\varepsilon u}}{dP^\varepsilon_{\theta_0}}(F^\varepsilon_{\theta_0+\varepsilon u}(w))\right)\right]$$

$$= E\left[\varphi_\varepsilon^u(F_{\theta_0}^\varepsilon(w))\, I_A\left(\log \frac{dP_{\theta_0+\varepsilon u}^\varepsilon}{dP_{\theta_0}^\varepsilon}(F_{\theta_0}^\varepsilon(w))\right)\right.$$

$$\left. \times \exp\{\varepsilon^{-2} G(w, \varepsilon, \theta_0 + \varepsilon u; \theta_0)\}\right]$$

$$\sim E\left[\varphi_\varepsilon^u(F_{\theta_0}^\varepsilon(w))\, \psi_\varepsilon^{L,u}(w)\, I_A\left(\log \frac{dP_{\theta_0+\varepsilon u}^\varepsilon}{dP_{\theta_0}^\varepsilon}(F_{\theta_0}^\varepsilon(w))\right)\right.$$

$$\left. \times \exp\{\varepsilon^{-2} G(w, \varepsilon, \theta_0 + \varepsilon u; \theta_0)\}\right]$$

$$\sim E\left[\phi_\varepsilon^{K,u}(w) \exp\{\varepsilon^{-2} G(w, \varepsilon, \theta_0 + \varepsilon u; \theta_0)\}\, \psi_\varepsilon^{L,u}(w)\, I_A\right.$$

$$\left. \circ\left(\log \frac{dP_{\theta_0+\varepsilon u}^\varepsilon}{dP_{\theta_0}^\varepsilon}(F_{\theta_0}^\varepsilon(w))\right)\right]$$

$$\sim e[e^{f_0^{L,u}} \bar{\Phi}_{A,0}^{L,u}] + \varepsilon E[e^{f_0^{L,u}} \bar{\Phi}_{A,1}^{L,u}] + \cdots.$$

The rest is finding $p_0^{Lc,u}$ and $p_1^{Lc,u}$. We see

$$E[e^{f_0^{L,u}} \bar{\Phi}_{A,0}^{L,u}] = E[e^{f_0^{L,u}} I_A(f_0^{L,u})] = \int_A e^x p_0^{L,u}(x)\, dx = \int_A \phi(x; \tfrac{1}{2} J, J)\, dx$$

and for $A = \{y; y < x\}$:

$$\begin{aligned}
&E[e^{f_0^{L,u}} \bar{\Phi}_{A,1}^{L,u}] \\
&\quad = E[e^{f_0^{L,u}} (\Phi_{A,1}^{L,u} + f_1^{L,u} \Phi_{A,0}^{L,u})] \\
&\quad = E[e^{f_0^{L,u}} (f_1^{L,u} \partial I_A(f_0^{L,u}) + f_1^{L,u} I_A(f_0^{L,u}))] \\
&\quad = -E[e^{f_0^{L,u}} f_1^{L,u} \delta_x(f_0^{L,u})] + E[e^{f_0^{L,u}} f_1^{L,u} I_A(f_0^{L,u})] \\
&\quad = -E[e^{f_0^{L,u}} f_1^{L,u} \mid f_0^{L,u} = x]\, p_0^{L,u}(x) \\
&\qquad + \int_{-\infty}^x e^x E[f_1^{L,u} \mid f_0^{L,u} = x]\, p_0^{L,u}(x)\, dx \\
&\quad = -E[f_1^{L,u} \mid f_0^{L,u} = x]\, \phi(x; \tfrac{1}{2} J, J) \\
&\qquad + \int_{-\infty}^x E[f_1^{L,u} \mid f_0^{L,u} = x]\, \phi(x; \tfrac{1}{2} J, J)\, dx.
\end{aligned}$$

Therefore,

$$\begin{aligned}
p_1^{Lc,u}(x) = &-\partial\{E[f_1^{L,u} \mid f_0^{L,u} = x]\, \phi(x; \tfrac{1}{2} J, J)\} \\
&+ E[f_1^{L,u} \mid f_0^{L,u} = x]\, \phi(x; \tfrac{1}{2} J, J).
\end{aligned}$$

7. Proof of Theorems 4.1. and 4.2.

In general we cannot ensure the existence and smoothness of the maximum likelihood estimators on the whole Wiener space, but it is possible to extend them to smooth Wiener functionals by multiplying a certain truncation functional. We begin with constructing such a functional for the maximum likelihood estimator.

Let $\varDelta$ be a bounded convex domain in $\mathbf{R}^{k_0}$. Let $m, n, j \in \mathbf{N}$ satisfy $m > k_0/2n + j$. Then we know that the Sobolev space $(W^{m,2n}(\varDelta), \|\cdot\|_{W^{m,2n}(\varDelta)})$ is embedded by a compact operator into $C_b^j(\varDelta)$, the totality of continuous functions on $\varDelta$ with bounded continuous derivatives up to the jth order, equipped with the norm

$$\|f\|_{C_b^j(\varDelta)} = \sum_{\nu \in \bar{\mathbf{N}}^{k_0},\, |\nu| \leqslant j} \sup_{\theta \in \varDelta} |\delta_\nu f(\theta)|.$$

Here the multiindex is defined similarly as in Section 2. In particular, for some $C(m, n, \varDelta) > 0$

$$\|f\|_{C_b^j(\varDelta)} \leqslant C(m, n, \varDelta) \|f\|_{W^{m,2n}(\varDelta)}$$

for $f \in W^{m,2n}(\varDelta)$.

Take γ so that $O < \gamma < \frac{1}{2}$. Define $\Theta' = (0, 1) \times \Theta^o$ and F' by

$$F'(w, \varepsilon, (\eta, \theta)) = F(w, \eta\varepsilon, \theta), \qquad (\eta, \theta) \in \Theta'$$

for functional $F(w, \varepsilon, \theta)$ on $W \times [0, 1) \times \Theta^o$. Let $m_0, n_0 \in \mathbf{N}$ satisfy $m_0 > (k+1)/2n_0 + 2$. The derivative operator with respect to η is denoted by δ_η. For $\theta_0 \in \Theta^o$, $c > 0$, and $\varepsilon \in (0, 1)$, let

$$\begin{aligned} R_\varepsilon^c(w) = {} & c\, \|G'(w, \varepsilon, \cdot\,; \theta_0) - G'(0, \cdot\,; \theta_0)\|_{W^{m_0, 2n_0}(\Theta')}^{2n_0} \\ & + c\, \| |\varepsilon^{1-2\gamma} (\delta_0 \delta G)'(w, \varepsilon, (\cdot, \theta_0); \theta_0)|^2 \|_{W^{1,2}((0,1))}^2 \end{aligned}$$

and let $R_0^c(w) = 0$.

Lemma 7.1. *Assume that* (C1) *and* (C2) *are satisfied.*

(1) *For* $l \in \bar{\mathbf{N}}$, $\nu \in \mathbf{M}$ *and* $p > 1$,

$$\sup_{(\eta, \theta) \in \Theta'} \|\delta_\eta^l \delta_\nu G'(w, \varepsilon, (\eta, \theta); \theta_0) - \delta_\eta^l \delta_\nu G'(0, (\eta, \theta); \theta_0)\|_{p,0} \leqslant C_1(l, \nu, p)\, \varepsilon^{l \vee 1},$$

$\varepsilon \in [0, 1)$, *for some constant* $C_1(l, \nu, p)$.

(2) *Let* $m, n \in \mathbf{N}$. *For* $p > 2n$, *there exists a positive constant* $C_2(m, n, p, \Theta')$ *such that*

$$\| \| G'(w, \varepsilon, \cdot; \theta_0) - G'(0, \cdot; \theta_0) \|_{W^{m, 2n}(\Theta')} \|_{p, 0} \leqslant C_2(m, n, p, \Theta') \cdot \varepsilon$$

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

(3) *For* $l \in \bar{\mathbf{N}}$, $v \in \bar{\mathbf{N}}^{k+1}$, *and* $p > 1$,

$$\sup_{\substack{(\eta, \theta) \in \Theta' \\ \varepsilon \in [0, 1)}} \| \delta_\eta^l \delta_v \delta G'(w, \varepsilon, (\eta, \theta); \theta_0) \|_{p, 0} \leqslant C_3(l, v, p)$$

for some constant $C_3(l, v, p) > 0$.

(4) *For* $p > 2$ *there exists constant* $C_4(p, k) > 0$ *such that*

$$\| \| |(\delta_0 \delta G)' (w, \varepsilon, (\cdot, \theta_0); \theta_0)|^2 \|_{W^{1, 2}((0, 1))} \|_{p, 0} \leqslant C_4(p, k) < \infty$$

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

(5) *For* $a > 0$, $c > 0$, *and* $n \in \mathbf{N}$, $P(R_\varepsilon^c(w) > a) = O(\varepsilon^n)$ *as* $\varepsilon \downarrow 0$.

(6) *If* $R_\varepsilon^c(w) < 1$, *then*

$$\| G(w, \varepsilon', \cdot; \theta_0) - G(0, \cdot; \theta_0) \|_{C_b^2(\Theta)} \leqslant C(m_0, n_0, \Theta')\, c^{-p_1}$$

for $0 \leqslant \varepsilon' \leqslant \varepsilon$, $p_1 = 1/2n_0$;

$$\sup_{\eta \in (0, 1)} |\varepsilon^{1-2\gamma} (\delta_0 \delta G)(w, \eta\varepsilon, \theta_0; \theta_0)| \leqslant C(1, 1, (0, 1))^{1/2}\, c^{-1/4};$$

and

$$|\varepsilon'^{-2\gamma} \delta G(w, \varepsilon', \theta_0; \theta_0)| \leqslant C(1, 1, (0, 1))^{1/2}\, c^{-1/4}$$

for $0 \leqslant \varepsilon' \leqslant \varepsilon$.

Proof. (1) For $v \in \mathbf{M}$ and $p > 1$,

$$\begin{aligned}
&\sup_{(\eta, \theta) \in \Theta'} \| \delta_v G'(w, \varepsilon, (\eta, \theta); \theta_0) - \delta_v G'(w, 0, (\eta, \theta); \theta_0) \|_{p, 0} \\
&\quad = \sup_{(\eta, \theta) \in \Theta'} \| \delta_v G(w, \eta\varepsilon, \theta; \theta_0) - \delta_v G(w, 0, \theta; \theta_0) \|_{p, 0} \\
&\quad = \sup_{(\eta, \theta) \in \Theta'} \left\| \eta\varepsilon \int_0^1 (\delta_0 \delta_v G)(w, s\eta\varepsilon, \theta; \theta_0)\, ds \right\|_{p, 0} \\
&\quad \leqslant \sup_{(\eta, \theta) \in \Theta'} \varepsilon \int_0^1 \| (\delta_0 \delta_v G)(w, s\eta\varepsilon, \theta; \theta_0) \|_{p, 0}\, ds \\
&\quad \leqslant C_1(0, v, p)\, \varepsilon,
\end{aligned}$$

$\varepsilon \in [0, 1)$, by (C1). Since

$$\delta_\eta^l \delta_v G'(w, 0, (\eta, \theta); \theta_0) = 0$$

for $l \geqslant 1$,

$$\begin{aligned}
\sup_{(\eta, \theta) \in \Theta'} & \|\delta_\eta^l \delta_v G'(w, \varepsilon, (\eta, \theta); \theta_0) - \delta_\eta^l \delta_v G'(w, 0, (\eta, \theta); \theta_0)\|_{p, 0} \\
&= \sup_{(\eta, \theta) \in \Theta'} \|\delta_\eta^l \delta_v G'(w, \varepsilon, (\eta, \theta); \theta_0)\|_{p, 0} \\
&= \sup_{(\eta, \theta) \in \Theta'} \|\varepsilon^l (\delta_0^l \delta_v G)(w, \eta\varepsilon, \theta; \theta_0)\|_{p, 0} \\
&\leqslant C_1(l, v, p)\, \varepsilon^l,
\end{aligned}$$

$\varepsilon \in [0, 1)$, by ($C1$). This shows (1).

One has (2) form (1), (3) form ($C1$), and (4) from (3), respectively.

(5) For $p > 2n_0$,

$$\begin{aligned}
P(R_\varepsilon^c(w) > a) &= P(c\, \|G'(w, \varepsilon, \cdot; \theta_0) - G'(0, \cdot; \theta_0)\|_{W^{m_0, 2n_0}(\Theta')}^{2n_0} \\
&\quad + c\| |\varepsilon^{1-2\gamma} (\delta_0 \delta G)'(w, \varepsilon, (\cdot, \theta_0); \theta_0)|^2 \|_{W^{1,2}((0,1))}^2 > a) \\
&\leqslant P(\|G'(w, \varepsilon, \cdot; \theta_0) - G'(0, \cdot; \theta_0)\|_{W^{m_0, 2n_0}(\Theta')}^p > (2^{-1} a c^{-1})^{p/2n_0}) \\
&\quad + P(\| |(\delta_0 \delta G)'\, (w, \varepsilon, (\cdot, \theta_0); \theta_0)|^2 \|_{W^{1,2}((0,1))}^{2p} \\
&\quad > (2^{-1} a c^{-1})^p\, \varepsilon^{-4p(1-2\gamma)}) \\
&\leqslant (2^{-1} a c^{-1})^{-p/2n_0}\, C_2(m_0, n_0, p, \Theta')^p\, \varepsilon^p \\
&\quad + (2^{-1} a c^{-1})^{-p}\, C_4(2p, k)^{2p}\, \varepsilon^{4p(1-2\gamma)},
\end{aligned}$$

$\varepsilon \in (0, 1)$. Let $p > \max\{2n_0, n/4(1-2\gamma), n\}$, which completes the proof.

(6) If $R_\varepsilon^c(w) < 1$, by definition we see that

$$\|G'(w, \varepsilon, \cdot; \theta_0) - G'(0, \cdot; \theta_0)\|_{W^{m_0, 2n_0}(\Theta')} \leqslant c^{-1/2n_0}$$

and

$$\| |\varepsilon^{1-2\gamma} (\delta_0 \delta G)'\, (w, \varepsilon, (\cdot, \theta_0); \theta_0)|^2 \|_{W^{1,2}((0,1))} \leqslant c^{-1/2}.$$

By Sobolev's inequality

$$\|G'(w, \varepsilon, \cdot; \theta_0) - G'(0, \cdot; \theta_0)\|_{C_b^2(\Theta')} \leqslant C(m_0, n_0, \Theta')\, c^{-1/2n_0}$$

and

$$\| |\varepsilon^{1-2\gamma}(\delta_0\delta G)'(w, \varepsilon, (\cdot, \theta_0); \theta_0)|^2 \|_{C_b((0,1))} \leqslant C(1, 1, (0, 1))c^{-1/2},$$

from which we have the first two assertions. Finally, since $\delta G(0, \theta_0; \theta_0) = 0$,

$$\begin{aligned}\sup_{0<\varepsilon'\leqslant\varepsilon} |\varepsilon'^{-2\gamma}\delta G(w, \varepsilon', \theta_0; \theta_0)| &= \sup_{0<\eta\leqslant 1} |(\eta\varepsilon)^{-2\gamma}\, \delta G(w, \eta\varepsilon, \theta_0; \theta_0)| \\ &\leqslant \sup_{0<\eta\leqslant 1} (\eta\varepsilon)^{1-2\gamma} \int_0^1 |(\delta_0\delta G)(w, s\eta\varepsilon, \theta_0; \theta_0)|\, ds \\ &\leqslant \sup_{0<\eta\leqslant 1} \int_0^1 |\varepsilon^{1-2\gamma}(\delta_0\delta G)(w, s\eta\varepsilon, \theta_0; \theta_0)|\, ds \\ &\leqslant C(1, 1, (0, 1))^{1/2}\, c^{-1/4}\end{aligned}$$

from the above result. ∎

The functional $R^c_\varepsilon(w) \in D^\infty$ for $c > 0$ and $\varepsilon \in [0, 1)$. Let λ_1 denote the minimum eigenvalue of the positive definite bilinear form $I(\theta_0) = -\delta^2 G(0, \theta_0; \theta_0)$. Choose $c_1 > 0$ and $d_1 > 0$ for $\theta_0 \in \Theta^o$ so that

$$c_1 < \tfrac{1}{4}\lambda_1 \tag{7.1}$$

and

$$|\delta^2 G(0, \theta; \theta_0) - \delta^2 G(0, \theta_0; \theta_0)| \leqslant c_1 \tag{7.2}$$

for $|\theta - \theta_0| \leqslant d_1$. Next, take $c > 0$ large enough so that

$$C(m_0, n_0, \Theta')c^{-p_1} < \min\{\tfrac{1}{4}\lambda_1, a_0 d_1^2\} \tag{7.3}$$

and

$$d_1\, C(1, 1, (0, 1))^{1/2}\, c^{-1/4} < \tfrac{1}{8}\lambda_1, \tag{7.4}$$

where $p_1 = 1/2n_0$. Let $\psi_\varepsilon(w) = 0$ if $\varepsilon^\gamma \geqslant d_1$ and $\psi_\varepsilon(w) = \psi(R^{3c}_\varepsilon(w))$ if $\varepsilon^\gamma < d_1$.

The following lemma ensures the existence and the smoothness of the maximum likelihood estimator under truncation.

LEMMA 7.2. *Assume that* (C1) *and* (C2) *are satisfied. Suppose that positive constants* c_1, d_1, *and* c *satisfy* (7.1)–(7.4). *Then*

(1) *For each* $\varepsilon \in [0, 1)$, *the functional* $w \mapsto \psi_\varepsilon(w) \in D^\infty$ *and* $0 \leqslant \psi_\varepsilon(w) \leqslant 1$.

(2) *If* $\varepsilon^\gamma < d_1$ *and* $R^c_\varepsilon(w) < 1$, *there exists a maximum likelihood estimate* $\hat{\theta}_\varepsilon(w; \theta_0)$.

(3) *$\delta^2 G(w, \varepsilon, \theta; \theta_0)$ are uniformly negative definite. More precisely, for $|\theta - \theta_0| \leqslant d_1$,*

$$\sup_{|\xi|=1} \delta^2 G(w, \varepsilon, \theta; \theta_0)[\xi, \xi] \leqslant -\tfrac{1}{2}\lambda_1$$

if $\varepsilon^\gamma \leqslant d_1$ and $R_\varepsilon^c(w) < 1$.

(4) *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 $\delta G(w,\ \varepsilon,\ \theta;\ \theta_0) = 0$.*

(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 $n \in \mathbf{N}$, $\psi_\varepsilon(w) = 1 - O(\varepsilon^n)$ in D^∞ as $\varepsilon \downarrow 0$.*

Proof. (1) is easy to show. We verify (2), (3), and (4). For $\xi \in \mathbf{R}^k$ and $\theta \in \Theta$, $|\theta - \theta_0| \leqslant d_1$,

$$\begin{aligned}
&\delta^2 G(w, \varepsilon, \theta; \theta_0)[\xi, \xi] \\
&\quad \leqslant \delta^2 G(0, \theta_0; \theta_0)[\xi, \xi] + |\delta^2 G(w, \varepsilon, \theta; \theta_0) - \delta^2 G(0, \theta; \theta_0)|\ |\xi|^2 \\
&\qquad + |\delta^2 G(0, \theta; \theta_0) - \delta^2 G(0, \theta_0; \theta_0)|\ |\xi|^2 \\
&\quad \leqslant (-\lambda_1 + C(m_0, n_0, \Theta')\ c^{-p_1} + c_1)\ |\xi|^2 \\
&\quad \leqslant (-1 + \tfrac{1}{4} + \tfrac{1}{4})\ \lambda_1\ |\xi|^2 \\
&\quad = -\tfrac{1}{2}\lambda_1\ |\xi|^2
\end{aligned}$$

from Lemma 7.1(6), (7.1), (7.2), (7.3) if $R_\varepsilon^c(w) < 1$. Then

$$\begin{aligned}
&\sup_{\varepsilon^\gamma \leqslant |\theta - \theta_0| \leqslant d_1} \varepsilon^{-2\gamma} G(w, \varepsilon, \theta; \theta_0) \\
&\quad \leqslant \sup_{\varepsilon^\gamma \leqslant |\theta - \theta_0| \leqslant d_1} \varepsilon^{-2\gamma} \delta G(w, \varepsilon, \theta_0; \theta_0)[\theta - \theta_0] \\
&\qquad + \sup_{\substack{\varepsilon^\gamma \leqslant |\theta - \theta_0| \leqslant d_1 \\ |\tilde{\theta} - \theta_0| \leqslant d_1}} \tfrac{1}{2}\varepsilon^{-2\gamma} \delta^2 G(w, \varepsilon, \tilde{\theta}; \theta_0)[\theta - \theta_0, \theta - \theta_0] \\
&\quad \leqslant d_1\, C(1, 1, (0, 1))^{1/2}\ c^{-1/4} \\
&\qquad + \sup_{\varepsilon^\gamma \leqslant |\theta - \theta_0| \leqslant d_1} \tfrac{1}{2}\varepsilon^{-2\gamma}(-\tfrac{1}{2}\lambda_1\ |\theta - \theta_0|^2) \\
&\quad \leqslant -\tfrac{1}{8}\lambda_1 < 0
\end{aligned}$$

from Lemma 7.1(6) and (7.4) if $R_\varepsilon^c(w)<1$. On the other hand, from (C2) and (7.3),

$$\begin{aligned}
&\sup_{d_1 \leqslant |\theta-\theta_0|} G(w, \varepsilon, \theta; \theta_0)\\
&\quad= \sup_{d_1 \leqslant |\theta-\theta_0|} [G(0, \theta; \theta_0)+(G(w, \varepsilon, \theta; \theta_0)-G(0, \theta; \theta_0))]\\
&\quad\leqslant \sup_{d_1 \leqslant |\theta-\theta_0|} (-a_0\,|\theta-\theta_0|^2)+C(m_0.n_0, \Theta')\, c^{-p_1}\\
&\quad\leqslant -a_0 d_1^2+C(m_0\cdot n_0, \Theta')\, c^{-p_1}\\
&\quad<0
\end{aligned}$$

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

$$\sup_{\varepsilon^\gamma \leqslant |\theta-\theta_0|} G(w, \varepsilon, \theta, \theta_0)<0$$

if $\varepsilon^\gamma \leqslant d_1$ and $R_\varepsilon^c(w)<1$. This shows that a maximum likelihood estimate exists in $|\theta-\theta_0|<\varepsilon^\gamma$. Moreover, for $|\theta-\theta_0| \leqslant d_1$,

$$\sup_{|\xi|=1} \delta^2 G(w, \varepsilon, \theta;\theta_0)[\xi, \xi] \leqslant -\tfrac{1}{2}\lambda_1$$

if $\varepsilon^\gamma \leqslant d_1$ and $R_\varepsilon^c(w)<1$, as shown above, so that $\delta^2 G(w, \varepsilon, \theta;\theta_0)$ are uniformly negative definite and, hence, the maximum likelihood estimate is a unique root in $\{\theta;\ |\theta-\theta_0|<d_1\}$ of the equation $\delta G(w, \varepsilon, \theta; \theta_0)=0$.

For each $h \in H$, constructing versions $\tilde{G}_h$, $\widetilde{\delta G_h}$, etc. and $\tilde{R}_{\varepsilon h}^{3c}$ naturally, we can show (5) [15].

Finally, by Lemma 7.1 (5) and applying chain rules with respect to H-derivatives to $\psi(R_\varepsilon^{3c}(w))$ we obtain (6). ∎

Remark 7.1. We can also prove that if $\varepsilon^\gamma<d_1$ and $R_\varepsilon^c(w)<1$, then for $0 \leqslant \varepsilon' \leqslant \varepsilon$ the maximum likelihood estimate $\hat{\theta}_{\varepsilon'}(w; \theta_0)$ exists in $\{\theta;\ |\theta-\theta_0| \leqslant \varepsilon'^\gamma\}$ and is a unique solution in $\{\theta; |\theta-\theta_0|<d_1\}$ of the equation $\delta G(w, \varepsilon', \theta; \theta_0)=0$. Moreover, for $|\theta-\theta_0| \leqslant d_1$,

$$\sup_{|\xi|=1} \delta^2 G(w, \varepsilon', \theta; \theta_0)[\xi, \xi] \leqslant -\tfrac{1}{2}\lambda_1.$$

For each $\varepsilon \in (0, 1)$, $\psi_\varepsilon(w)\, \hat{\theta}.(w; \theta_0)$ is a $C([0, \varepsilon] \rightarrow \mathbf{R}^k)$-valued smooth functional.

LEMMA 7.3. *Assume that* (C1) *and* (C2) *are satisfied. Then* $\psi_\varepsilon(w)\,\varepsilon^{-1}(\hat{\theta}^*_\varepsilon(w;\theta_0)-\theta_0)\in D^\infty(\mathbf{R}^k)$ *has the asymptotic expansion*

$$\psi_\varepsilon(w)\,\varepsilon^{-1}(\hat{\theta}^*_\varepsilon(w;\theta_0)-\theta_0)\sim f_0+\varepsilon f_1+\cdots$$

in $D^\infty(\mathbf{R}^k)$ *as* $\varepsilon\downarrow 0$ *with* $f_0, f_1, ...\in D^\infty(\mathbf{R}^k)$.

Proof. Suppose that $\varepsilon_0^\gamma<d_1$. Let $R^c_{\varepsilon 0}(w)<1$. Then, for $\varepsilon, \varepsilon'<\varepsilon_0$, there exist $\tilde{\varepsilon}(|\tilde{\varepsilon}-\varepsilon|<|\varepsilon'-\varepsilon|)$ and $\tilde{\theta}_s(|\tilde{\theta}_s-\hat{\theta}_\varepsilon(w;\theta_0)|<|\hat{\theta}_{\varepsilon'}(w;\theta_0)-\hat{\theta}_\varepsilon(w;\theta_0)|)$ for which

$$\begin{aligned}0&=\delta G(w,\varepsilon',\hat{\theta}_{\varepsilon'}(w;\theta_0);\theta_0)-\delta G(w,\varepsilon,\hat{\theta}_\varepsilon(w;\theta_0);\theta_0)\\&=\int_0^1\delta^2G(w,\varepsilon',\tilde{\theta}_s;\theta_0)\,ds[\hat{\theta}_{\varepsilon'}(w;\theta_0)-\hat{\theta}_\varepsilon(w;\theta_0)]\\&\quad+\delta_0\delta G(w,\tilde{\varepsilon},\hat{\theta}_\varepsilon(w;\theta_0);\theta_0)(\varepsilon'-\varepsilon).\end{aligned}$$

From this equation, Remark 7.1 and (C1) we see that if $0\leqslant\varepsilon<\varepsilon_0$, $\hat{\theta}_\varepsilon(w;\theta_0)$ is continuous in ε and differentiable in ε:

$$\delta_0\hat{\theta}_\varepsilon(w;\theta_0)=-[\delta^2G(w,\varepsilon,\hat{\theta}_\varepsilon(w;\theta_0);\theta_0)]^{-1}\,\delta_0\delta G(w,\varepsilon,\hat{\theta}_\varepsilon(w;\theta_0);\theta_0).$$

Since both sides are differentiable,

$$\begin{aligned}&\delta_0^2\hat{\theta}_\varepsilon(w;\theta_0)\\&\quad=2[\delta^2G(w,\varepsilon,\hat{\theta}_\varepsilon(w;\theta_0);\theta_0)]^{-1}\,\delta_0\delta^2G(w,\varepsilon,\hat{\theta}_\varepsilon(w;\theta_0);\theta_0)\\&\qquad\times[\delta^2G(w,\varepsilon,\hat{\theta}_\varepsilon(w;\theta_0);\theta_0)]^{-1}\,\delta_0\delta G(w,\varepsilon,\hat{\theta}_\varepsilon(w;\theta_0);\theta_0)\\&\qquad-[\delta^2G(w,\varepsilon,\hat{\theta}_\varepsilon(w;\theta_0);\theta_0)]^{-1}\,\delta_0^2\delta G(w,\varepsilon,\hat{\theta}_\varepsilon(w;\theta_0);\theta_0)\\&\qquad+[\delta^2G(w,\varepsilon,\hat{\theta}_\varepsilon(w;\theta_0);\theta_0)]^{-1}\,\delta^3G(w,\varepsilon,\hat{\theta}_\varepsilon(w;\theta_0);\theta_0)[\,\cdot\,,\delta_0\hat{\theta}_\varepsilon(w;\theta_0),\\&\qquad\quad\{\delta^2G(w,\varepsilon,\hat{\theta}_\varepsilon(w;\theta_0);\theta_0)\}^{-1}\,\delta_0\delta G(w,\varepsilon,\hat{\theta}_\varepsilon(w;\theta_0);\theta_0)]\end{aligned}$$

if $0\leqslant\varepsilon<\varepsilon_0$. The higher order derivatives with respect to ε also exist and can be calculated in a similar way. We note that if $R^c_{\varepsilon 0}(w)<1$, $\hat{\theta}_\varepsilon(w;\theta_0)$ is the unique solution of the estimating equation and smooth on $[0,\varepsilon_0]$, which is a consequence of the choice of $R^c_{\varepsilon 0}(w)$. See Remark 7.1. Then we have the expansion

$$\begin{aligned}\hat{\theta}_{\varepsilon 0}(w;\theta_0)&=\theta_0+\frac{\varepsilon_0}{1!}(\delta_0)_0\,\hat{\theta}.(w;\theta_0)+\frac{\varepsilon_0^2}{2!}(\delta_0)_0^2\,\hat{\theta}.(w;\theta_0)+\cdots\\&\quad+\frac{\varepsilon_0^{j-1}}{(j-1)!}(\delta_0)_0^{j-1}\,\hat{\theta}.(w;\theta_0)+\varepsilon_0^j\int_0^1\frac{1}{(j-1)!}(1-s)^{j-1}\\&\quad\times(\delta_0)^j_{s\varepsilon_0}\,\hat{\theta}.(w;\theta_0)\,ds.\end{aligned}$$

It is not difficult to show from this equation that

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

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 $n\in\mathbf{N}$. ∎

Let $\sigma_\varepsilon^*(w)=(\sigma_\varepsilon^{*ij}(w))_{i,j=1}^k$, $0<\varepsilon<1$, denote the Malliavin covariance of $\psi_\varepsilon(w)\varepsilon^{-1}[\hat{\theta}_\varepsilon^*(w;\theta_0)-\theta_0]$, that is,

$$\begin{aligned}
\sigma_\varepsilon^{*ij}(w)=\langle D\{\psi_\varepsilon(w)\,\varepsilon^{-1}[\hat{\theta}_\varepsilon^*(w;\theta_0)-\theta_0]^i\},\\
D\{\psi_\varepsilon(w)\,\varepsilon^{-1}[\hat{\theta}_\varepsilon^*(w;\theta_0)-\theta_0]^j\}\rangle_H
\end{aligned}$$

for $i,j=1,2\cdots,k$. Similarly, define $\sigma_\varepsilon(w)$ for $\psi_\varepsilon(w)\,\varepsilon^{-1}[\hat{\theta}_\varepsilon(w;\theta_0)-\theta_0]$.

LEMMA 7.4. *Assume that* (C1)–(C3) *are satisfied. Then for* $\theta_0\in\Theta^o$, $\varepsilon\in(0,1)$, *and* $c>0$ *satisfying* (7.3) *and* (7.4), *there exists a Wiener functional* $\xi_\varepsilon(w)$ *with the following properties*:

(1) $0\leqslant\xi_\varepsilon(w)\in D^\infty$.

(2) *If* $\xi_\varepsilon(w)\leqslant 1$, $R_\varepsilon^{3c}(w)<\frac{1}{2}$.

(3) *There exists positive constant* a_1 *such that if* $\xi_\varepsilon(w)\leqslant 1$,

$$\inf_{\substack{v\in\mathbf{R}^k\\|v|=1}}\sigma_\varepsilon^*(w)[v,v]\geqslant a_1.$$

(4) *for any* $n\in\mathbf{N}$,

$$\lim_{\varepsilon\downarrow 0}\varepsilon^{-n}P(|\xi_\varepsilon(w)|>\tfrac{1}{2})=0$$

Proof. For $\varepsilon<\min\{d_1^{1/\gamma},(2\,|||\delta b|||_\infty)^{-1/2}\}$, $c_2>6c$, and $c_3>0$, let

$$\xi_\varepsilon^{c_2,c_3}(w)=R_\varepsilon^{c_2}(w)+c_3\,\||\sigma(w,\varepsilon,\cdot)-I(\theta_0)|^2\|_{W^{m_1,2n_1}(\{|u|<1\})}^{2n_1},$$

where $\sigma(w,\varepsilon,u)=([\sigma(w,\varepsilon,u)]^{ij})_{i,j=1}^k$ with

$$[\sigma(w,\varepsilon,u)]^{ij}=\langle\varepsilon^{-1}D\delta_iG(w,\varepsilon,\theta_0+\varepsilon^\gamma u;\theta_0),\varepsilon^{-1}D\delta_jG(w,\varepsilon,\theta_0+\varepsilon^\gamma u;\theta_0)\rangle_H$$

for $u \in \mathbf{R}^k$, $|u| \leqslant 1$. Let $\xi_\varepsilon^{c_2, c_3} = 2$ for other ε. Let $m_1, n_1 \in \mathbf{N}$, $m_1 > k/2n_1 + 2$. Define $\xi_\varepsilon(w) = \xi_\varepsilon^{c_2, c_3}(w)$. Assertions (1) and (2) are trivial. If $\xi_\varepsilon^{c_2, c_3} \leqslant 1$ then

$$\sup_{|u| \leqslant 1} |\sigma(w, \varepsilon, u) - I(\theta_0)| \leqslant C(m_1, n_1, \{|u| < 1\})^{1/2}\, c_3^{-1/4n_1},$$

$$\sup_{|u| \leqslant 1} |\delta^2 G(w, \varepsilon, \theta_0 + \varepsilon^\gamma u; \theta_0) - \delta^2 G(0, \theta_0; \theta_0)| \leqslant C(m_0, n_0, \Theta')\, c_2^{-p_1} + c_1,$$

and

$$\sup_{|u| \leqslant 1} |\varepsilon^2 \delta b(\theta_0 + \varepsilon^\gamma u)| \leqslant \tfrac{1}{2}$$

for $|u| \leqslant 1, \varepsilon < \min\{d_1^{1/\gamma}, (2\, |\!|\!|\delta b|\!|\!|_\infty)^{-1/2}\}$. We know that the Malliavin covariance

$$\sigma_\varepsilon(w) = [\delta^2 G(w, \varepsilon, \hat{\theta}_\varepsilon(w; \theta_0); \theta_0)]^{-1}\, \sigma(w, \varepsilon, \varepsilon^{-\gamma}(\hat{\theta}_\varepsilon(w; \theta_0) - \theta_0))$$
$$\times [\delta^2 G(w, \varepsilon, \hat{\theta}_\varepsilon(w; \theta_0); \theta_0)]'^{-1}$$

and

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

if $\xi_\varepsilon(w) \leqslant 1$ since then $\psi_\varepsilon(w) = 1$ and $D\psi_\varepsilon(w) = 0$. Thus if we choose c_2, c_3 large enough, (3) holds.

Finally we show (4). The Malliavin covariance of $\delta_0 \delta G(w, 0, \theta_0; \theta_0)$ is $I(\theta_0)$. We have

$$\varepsilon^{-1} D\delta G(w, \varepsilon, \theta_0 + \varepsilon^\gamma u; \theta_0) = D\delta_0 \delta G(w, 0, \theta_0; \theta_0) + \varepsilon^\gamma r_\varepsilon(u),$$

where $r_\varepsilon(u)$ is given by the Bochner integral

$$r_\varepsilon(u) = \varepsilon^{1-\gamma} \int_0^1 (1 - \eta)(D\delta_0^2 \delta G)(w, \eta\varepsilon, \theta_0 + \varepsilon^\gamma u; \theta_0)\, d\eta$$
$$+ \int_0^1 (D\delta_0 \delta^2 G)(w, 0, \theta_0 + s\varepsilon^\gamma u; \theta_0)\, ds[\,\cdot, u].$$

From this fact we obtain (4) by estimating the Sobolev norm of the difference of Malliavin covariances.

For multiindex $\mathbf{n} = (n_1, ..., n_k)$, let $\mathbf{n}! = n_1! \cdots n_k!$, $a^{\mathbf{n}} = a_1^{n_1} \cdots a_k^{n_k}$ for $a \in \mathbf{R}^k$ and $\partial^{\mathbf{n}} = \partial_1^{n_1} \cdots \partial_k^{n_k}$, where $\partial_i = \partial/\partial x^i$, $i = 1, ..., k$.

LEMMA 7.5. *Assume that* (C1)–(C3) *are satisfied. Then the generalized Wiener functional*

$$\psi(\xi_\varepsilon(w))\, I_A(\psi_\varepsilon(w)\, \varepsilon^{-1}(\hat{\theta}_\varepsilon^*(w; \theta_0) - \theta_0)),$$

$A \in \mathbf{B}^k$, is well defined and has the asymptotic expansion

$$\psi(\xi_\varepsilon(w))\, I_A(\psi_\varepsilon(w)\, \varepsilon^{-1}\, (\hat{\theta}^*_\varepsilon(w; \theta_0) - \theta_0)) \sim \Phi_{A,0} + \varepsilon \Phi_{A,1} + \cdots$$

in $\tilde{D}^{-\infty}$ as $\varepsilon \downarrow 0$ uniformly in $A \in \mathbf{B}^k$ with $\Phi_{A,0}, \Phi_{A,1},... \in \tilde{D}^{-\infty}$ determined by the formal Taylor expansion

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

In particular,

$$\begin{aligned} \Phi_{A,0} &= I_A(f_0), \\ \Phi_{A,1} &= f_1^i \partial_i I_A(f_0). \end{aligned}$$

Proof. We obtain the results from Lemmas 7.3 and 7.4 above and Theorem 4.1 of [13].

LEMMA 7.6. *For $A \in \mathbf{B}^k$,*

$$E[f_1^i \partial_i I_A(f_0)] = \int_A p_1(x)\, dx,$$

where

$$p_1(x) = -\partial_i \{ q_b^i(x)\, \phi(x; 0, I(\theta_0)^{-1}) \}.$$

Proof. Using the integration by parts formula in the Malliavin calculus, for some smooth functional $F(w, f_1)$ we have

$$E[f_1^i \partial_i I_A(f_0)] = E[F(w, f_1)\, I_A(f_0)] = \int_A p_1(x)\, dx,$$

where

$$p_1(x) = E[F(w, f_1) \mid f_0 = x]\, \phi(x; 0, I(\theta_0)^{-1}).$$

To get $p_1(x)$, let $A_x = [x^1, \infty) \times \cdots \times [x^k, \infty)$. Then

$$\begin{aligned} p_1(x) &= (-1)^k\, \partial_1 \cdots \partial_k E[f_1^i \partial_i I_{A_x}(f_0)] \\ &= -\partial_i E[f_1^i \delta_x(f_0)] \\ &= -\partial_i \{ E[f_1^i \mid f_0 = x]\, \phi(x; 0, I(\theta_0)^{-1}) \}. \quad \blacksquare \end{aligned}$$

Proof of Theorem 4.1. We obtain Theorem 4.1 from Lemmas 7.5 and 7.6.

Next, by Lemmas 6.4(2) and 7.5, we have the following.

LEMMA 7.7. *Assume that* (C1)–(C3) *and* (C5) *are satisfied. Let* $\theta_0 \in \Theta^o$ *and let* K *be any compact set in* $\mathbf{R}^k$. *Then*

$$F_\varepsilon^{u,K}(w) := \phi_\varepsilon^{K,u}(w) \exp\{\varepsilon^{-2} G(w, \varepsilon, \theta_0 + \varepsilon u; \theta_0)\}$$
$$\times \psi(\xi_\varepsilon(w))\, I_A(\psi_\varepsilon(w)\, \varepsilon^{-1}(\hat{\theta}_\varepsilon^*(w; \theta_0) - \theta_0))$$

is well defined and has the asymptotic expansion

$$F_\varepsilon^{u,K} \sim e^{f_0^{L,u}}(\bar{\Phi}_{A,0}^u + \varepsilon \bar{\Phi}_{A,1}^u + \cdots),$$

in $\tilde{D}^{-\infty}$ *as* $\varepsilon \downarrow 0$ *uniformly in* $A \in \mathbf{B}^k$ *and* $u \in K$ *with* $\bar{\Phi}_{A,0}^u$, $\bar{\Phi}_{A,1}^u$,... *determined by the formal Taylor expansion*

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

In particular,

$$\bar{\Phi}_{A,0}^u = I_A(f_0), \qquad \bar{\Phi}_{A,1}^u = f_1^i \partial_i I_A(f_0) + f_1^{L,u} I_A(f_0).$$

Proof of Theorem 4.2. The validity of the asymptotic expansion can be proved in a similar fashion as in the proof of Theorem 3.2. From integration by parts we see that each term on the right-hand side of the asymptotic expansion of $E[F_\varepsilon^{u,K}(w)]$ of Lemma 7.7 is represented by an integration of some smooth function on the set A. We determine $p_0^{c,u}$ and $p_1^{c,u}$. Let $A + u = \{x; x - u \in A\}$. We have

$$f_0^{L,u} = I[f_0, u] - \tfrac{1}{2} I[u, u].$$

Using Lemma 7.7 for $A + u$ in place of A, we have

$$\begin{aligned} E[e^{f_0^{L,u}} \bar{\Phi}_{A+u,0}^u] &= E[e^{f_0^{L,u}} I_{A+u}(f_0)] \\ &= \int_{A+u} \exp\{I[x, u] - \tfrac{1}{2} I[u, u]\}\, \phi(x; 0, I^{-1})\, dx \\ &= \int_{A+u} \phi(x - u; 0, I^{-1})\, dx \\ &= \int_A \phi(y; 0, I^{-1})\, dy. \end{aligned}$$

Next,

$$
\begin{aligned}
E[e^{f_0^{L,u}}\bar{\Phi}^u_{A+u,1}]
&= E[e^{f_0^{L,u}} f_1^i \partial_i I_{A+u}(f_0)] + E[e^{f_0^{L,u}} f_1^{L,u} I_{A+u}(f_0)] \\
&= \int_{A+u} -\partial_i E[e^{f_0^{L,u}} f_1^i \delta_x(f_0)]\, dx + E[e^{f_0^{L,u}} f_1^{L,u} I_{A+u}(f_0)] \\
&= \int_{A+u} [-\partial_i(\exp\{I[x,u] - \tfrac{1}{2} I[u,u]\}\ E[f_1^i \mid f_0 = x]\ \phi(x; 0, I^{-1})) \\
&\qquad + \exp\{I[x,u] - \tfrac{1}{2} I[u,u]\}\ E[f_1^{L,u} \mid f_0 = x]\ \phi(x; 0, I^{-1})]\, dx \\
&= \int_{A+u} [-\partial_i(q_b^i(x)\, \phi(x-u; 0, I^{-1})) + q_0^{L,u}(x)\, \phi(x-u; 0, I^{-1})]\, dx \\
&= \int_A [-\partial_i(q_b^i(y+u)\, \phi(y; 0, I^{-1})) + q_0^{L,u}(y+u)\, \phi(y; 0, I^{-1})]\, dy.
\end{aligned}
$$

This completes the proof.

Acknowledgments

I am grateful to the editor, the associate editor, and the referees for their valuable comments.

References

1. Akahira, M., and Takeuchi, K. 1981. "Asymptotic Efficiency of Statistical Estimators: Concepts and Higher Order Asymptotic Efficiency," Lect. Notes Statist., Vol. 7, Springer-Verlag, Berlin/Heidelberg/New York.
2. Kusuoka, S. 1982. Dirichlet forms and diffusion processes on Banach spaces, *J. Fac. Sci. Univ. Tokyo Sect. IA* **29** 79–95.
3. Kusuoka, S. 1982. The nonlinear transformation of Gaussian measure on Banach space and its absolute continuity, I, *J. Fac. Sci. Univ. Tokyo Sect. IA* **29** 567–597.
4. Kutoyants, Yu. A. 1984. "Parameter Estimation for Stochastic Processes" (B. L. S. Prakasa Rao, Ed.), Heldermann, Berlin.
5. Le Cam, L. 1960. Locally asymptotically normal families of distributions, *Univ. California Publ. Statist.* **3** 37–98.
6. Lipster, R. S., and Shiryayev, A. N. 1977. "Statistics of Random Processes I," Springer-Verlag, Berlin/Heidelberg/New York.
7. Sugita, H. 1985. On a characterization of the Sobolev spaces over an abstract Wiener space, *J. Math. Kyoto Univ.* **25** 717–725.
8. Takanobu, S., and Watanabe, S. 1993. Asymptotic expansion formulas of the Schilder type for a class of conditional Wiener functional integrations (K. D. Elworthy and

N. Ikeda, Eds.), Asymptotic problems in probability theory: Wiener functionals and asymptotics, *Proceedings*, *the Taniguchi International Symposium*, *Sanda and Kyoto*, 1990, pp. 194–241, Longman, Harlow, UK.

9. WATANABE, S. 1983. Malliavin's calculus in terms of generalized Wiener functionals, in "Lecture Notes in Control Inform. Sci.," Vol. 49, pp. 284–290, Springer-Verlag, Berlin/Heidelberg/New York.
10. WATANABE, S. 1984. "Lectures on Stochastic Differential Equations and Malliavin Calculus," Tata Institute of Fundamental Research, Springer-Verlag, Berlin/Heidelberg/New York.
11. WATANABE, S. 1987. Analysis of Wiener functionals (Malliavin calculus) and its applications to heat kernels, *Ann. Probab.* **15** 1–39.
12. YOSHIDA, N. 1992. Asymptotic expansions of maximum likelihood estimators for small diffusions via the theory of Malliavin–Watanabe, *Probab. Theory Related Fields* **92** 275–311.
13. YOSHIDA, N. 1992. Asymptotic expansion for statistics related to small diffusion, *J. Japan Statist. Soc.* **22**, No. 2, 139–159.
14. YOSHIDA, N. 1993. Asymptotic expansion of Bayes estimators for small diffusions, *Probab. Theory Related Fields* **95** 429–450.
15. YOSHIDA, N. 1993. "Banach Space Valued Functionals and Smoothness of M-Estimators," Research Memorandum 494, Institute of Statistical Mathematics.