数理解析研究所講究録
916 巻 1995 年 112-130

A NOTE ON THE CONJECTURE THAT THIRD-ORDER EFFICIENCY IMPLIES FOURTH-ORDER EFFICIENCY

YUTAKA KANO* (狩野 裕)

Abstract. Takeuchi [33] and Pfanzagl [23] proved that *any* first-order efficient estimators are second-order efficient. Many authors e.g., Ghosh [12], have conjectured that *any* third-order efficient estimators are fourth-order efficient. Based on concentration probability of estimators about a true parameter, this paper gives a positive answer to the conjecture in a curved exponential family with multi-structural parameters. It is seen that choice of bias-correction factors is critical.

1. Introduction

Let $\mathbf{u} \in \mathbb{R}^q$ be a parameter vector of interest and let an open set $\Omega(\subset \mathbb{R}^q)$ be a parameter space of $\mathbf{u}$. Let $\mathbf{u}_0 \in \Omega$ be an arbitrarily fixed (inner) point of Ω. For every $\mathbf{u} \in \mathcal{N}(\mathbf{u}_0)$, a neighborhood of $\mathbf{u}_0$, assume that random p-vectors $\mathbf{x}_1, \ldots, \mathbf{x}_n$ are independent and identically distributed according to a (continuous) curved exponential family with density

$$\exp\{\theta(\mathbf{u})'\mathbf{x} - \psi(\theta(\mathbf{u}))\}\,\mu(d\mathbf{x}), \tag{1.1}$$

where $\mu(\cdot)$ is a carrier measure on $\mathbb{R}^p$ (see Amari [5], Section 4.1) and $\theta(\cdot)$ is an $\mathbb{R}^p$-valued measurable function from $\Omega(\subset \mathbb{R}^q)$ smooth in $\mathbf{u} \in \mathcal{N}(\mathbf{u}_0)$. The real-valued function, $\psi(\theta)$, is defined as a normalizing constant $\psi(\theta) = \log \int_{\mathbb{R}^p} \exp(\theta'\mathbf{x})\mu(d\mathbf{x})$. Write $\Theta_1 = \frac{\mathrm{d}}{\mathrm{d}\mathbf{u}'}\theta(\mathbf{u})$, $\Psi_{10} = \frac{\mathrm{d}\psi}{\mathrm{d}\theta}$ and $\Psi_{11} = \frac{\mathrm{d}^2\psi}{\mathrm{d}\theta \mathrm{d}\theta'}$. It is well-known that

$$E[\mathbf{x}_i] = \Psi_{10}(\theta(\mathbf{u}))\ (= \eta(\mathbf{u}),\ \text{say}) \quad \text{and} \quad \mathrm{Cov}[\mathbf{x}_i, \mathbf{x}_i] = \Psi_{11}(\theta(\mathbf{u}))\ (= \Psi_{11},\ \text{say})$$

and that the Fisher information matrix is expressed as

$$i_{\mathbf{u}} = \Theta_1' \Psi_{11} \Theta_1.$$

Assume that Ψ_{11} and $i_{\mathbf{u}}$ are positive definite.

Let $\bar{\mathbf{x}} = \frac{1}{n}\sum_{i=1}^n \mathbf{x}_i$ and then $E[\bar{\mathbf{x}}] = \eta(\mathbf{u})$. Let $\mathbf{g}(\cdot)$, $\mathbf{b}_g^{(1)}(\cdot)$ and $\mathbf{b}_g^{(2)}(\cdot)$ be measurable functions from $\mathbb{R}^p$ to $\mathbb{R}^q$, smooth in $\mathcal{N}(\eta(\mathbf{u}_0))$ a neighborhood of $\eta(\mathbf{u}_0)$. The statistics

*Institute of Mathematics, University of Tsukuba, Ibaraki 305, Japan

$\mathbf{b}_g^{(i)}(\bar{\mathbf{x}})$'s are bias-correction factors, possibly to be zero. For given $\mathbf{b}_g^{(i)}(\bar{\mathbf{x}})$'s, consider a class of possibly bias-corrected Fisher consistent estimators for $\mathbf{u}$, defined as

$$\mathcal{F} = \left\{ \mathbf{g}^*(\bar{\mathbf{x}}) = \mathbf{g}(\bar{\mathbf{x}}) - \tfrac{1}{n}\mathbf{b}_g^{(1)}(\bar{\mathbf{x}}) - \tfrac{1}{n^2}\mathbf{b}_g^{(2)}(\bar{\mathbf{x}}) \middle| \, \mathbf{g}(\eta(\mathbf{u})) = \mathbf{u} \;\text{ for every }\; \mathbf{u} \in \mathcal{N}(\mathbf{u}_0) \right\}.$$

We say that $\mathbf{g}^*(\bar{\mathbf{x}})$ is *first*-order (or *second*-order) bias corrected if $E[\mathbf{g}^*(\bar{\mathbf{x}}) - \mathbf{u}] = O\left(\frac{1}{n^2}\right)$ (or $O\left(\frac{1}{n^3}\right)$). The bias-correction factors are not unique, and the class $\mathcal{F}$ with one correction-factor may be different from that with another factor. Discussing fourth-order efficiency here, we will deal with four different bias-correction factors.

We assume that the MLE $\hat{\mathbf{u}}(\bar{\mathbf{x}})$ that maximizes the likelihood function

$$L(\mathbf{u}) = \prod_{i=1}^{n} \exp\{\theta(\mathbf{u})'\mathbf{x}_i - \psi(\theta(\mathbf{u}))\} = \exp\{n(\theta(\mathbf{u})'\bar{\mathbf{x}} - \psi(\theta(\mathbf{u}))\} \tag{1.2}$$

exists and that it is Fisher-consistent and smooth in $\mathcal{N}(\eta(\mathbf{u}_0))$ a neighborhood of $E[\bar{\mathbf{x}}]$, with a large probability for large n. The assumption is met under certain regularity conditions including strong identifiability (see Kano [16]).

Let $\mathcal{C}$ be a class of all Borel convex sets of R^q *including the origin* and let $\mathcal{C}_0$ be a class of all Borel convex sets of R^q *symmetric about the origin*. The class $\mathcal{C}_0$ is a subclass of $\mathcal{C}$. We expand concentration probability of $\mathbf{g}^*(\bar{\mathbf{x}})$ about the true parameter vector $\mathbf{u}$ as

$$\mathrm{P}_{\theta(\mathbf{u})}\{\sqrt{n}(\mathbf{g}^*(\bar{\mathbf{x}}) - \mathbf{u}) \in C\} = A_1 + \frac{1}{\sqrt{n}}A_2 + \frac{1}{n}A_3 + \frac{1}{n\sqrt{n}}A_4 + \cdots, \tag{1.3}$$

where C belongs to $\mathcal{C}$ or to $\mathcal{C}_0$. The coefficients A_k's depend on $\mathbf{u}$, $\mathbf{g}^*(\cdot)$ and C. The expansion is made with the help of Edgeworth expansions.

An estimator in $\mathcal{F}$ is said to be first-order efficient (1OE) for the class $\mathcal{C}$ (or $\mathcal{C}_0$) iff it minimizes A_1 among all the estimators in $\mathcal{F}$ for every $C \in \mathcal{C}$ (or $C \in \mathcal{C}_0$). A 1OE estimator is said to be second-order efficient (2OE) for the class $\mathcal{C}$ (or $\mathcal{C}_0$) iff it minimizes A_2 among all the 1OE estimators in $\mathcal{F}$ for every $C \in \mathcal{C}$ (or $C \in \mathcal{C}_0$); 3OE and 4OE for the class $\mathcal{C}$ (or $\mathcal{C}_0$) are defined consecutively in the same manner.

Whether an estimator in $\mathcal{F}$ is 1OE does not depend on the bias-corrected factors; and the factor $\mathbf{b}_g^{(2)}(\bar{\mathbf{x}})$ does not influence upon 2OE and 3OE.

The limiting distribution of $\sqrt{n}(\mathbf{g}^*(\bar{\mathbf{x}}) - \mathbf{u})$ (or equivalently $\sqrt{n}(\mathbf{g}(\bar{\mathbf{x}}) - \mathbf{u})$) is multivariate normal with mean zero and covariance matrix $G_1\Psi_{11}G_1'$, where $G_1 = \frac{d\mathbf{g}(\bar{\mathbf{x}})}{d\bar{\mathbf{x}}'}|_{\bar{\mathbf{x}}=\eta(\mathbf{u})}$, and hence

$$A_1 = \int_C N_q(\mathbf{x}|0, G_1\Psi_{11}G_1')d\mathbf{x}.$$

It is known that

$$G_1\Psi_{11}G_1' \geq i_{\mathbf{u}}^{-1}$$

and that the equality holds if and only if $G_1 = i_{\mathbf{u}}^{-1}\Theta_1' = \frac{\mathrm{d}\hat{\mathbf{u}}(\bar{\mathbf{x}})}{\mathrm{d}\bar{\mathbf{x}}'}|_{\bar{\mathbf{x}}=\eta(\mathbf{u})}$, the matrix of the first-order derivatives of the MLE $\hat{\mathbf{u}}(\bar{\mathbf{x}})$ (see e.g., Takeuchi [33] example 6.5). As a result, we have

$$\int_C N_q(\mathbf{x}|0, G_1\Psi_{11}G_1')d\mathbf{x} \leq \int_C N_q(\mathbf{x}|0, i_{\mathbf{u}}^{-1})d\mathbf{x} \qquad \text{for every } \ C \in \mathcal{C}.$$

In other words, the MLE and estimators whose first-order derivatives are identical with the MLE are all 1OE for the class $\mathcal{C}$.

Second-order efficiency has been discussed for a class of first-order bias-corrected estimators $\mathbf{g}^*(\bar{\mathbf{x}})$ which meet

$$E[\mathbf{g}^*(\bar{\mathbf{x}}) - \mathbf{u}] = o\left(\tfrac{1}{n}\right).$$

The 2OE of the (bias-corrected) MLE was established by showing that the second term A_2 of the MLE attains the lower bound, which is derived by applying Neymann-Pearson's lemma to a certain testing problem (Pfanzagl [22]; Akahira [1]). After the 2OE of the MLE, Takeuchi [33] and Pfanzagl [23] found that the term A_2 is identical for *any* 1OE estimators, so that 1OE implies 2OE and that the notion of 2OE cannot distinguish 1OE estimators. The basic results on 2OE have been extended to more general and complicated models by many authors including Akahira and Takeuchi [2,3], Hosoya [15], Taniguchi [34] and Yoshida [36,37].

Consider Takeuchi's and Pfanzagl's surprising result via an alternative criterion, quadratic loss:

$$(1.4) \qquad E[\{\sqrt{n}(\mathbf{g}^*(\bar{\mathbf{x}}) - \mathbf{u})\}\{\sqrt{n}(\mathbf{g}^*(\bar{\mathbf{x}}) - \mathbf{u})\}'] = B_1 + \frac{1}{\sqrt{n}}B_2 + \frac{1}{n}B_3 + \frac{1}{n\sqrt{n}}B_4 + \cdots,$$

Higher-order efficiency based on the criterion is defined in the same way as that on concentration probability in (1.3). Since $B_1 = G_1\Psi_{11}G_1'$, the asymptotic covariance matrix of $\mathbf{g}(\bar{\mathbf{x}})$, we make the same conclusion on 1OE as before. We know that $B_2 = 0$ for any estimators in $\mathcal{F}$, which means that 1OE implies 2OE. Notice that $B_2 = 0$ whether or not bias-correction is made and hence that 2OE holds for the class of 1OE estimators even without bias-correction.

Results on 3OE similar to those on 1OE hold under first-order bias-correction. That is, the MLE and estimators whose second-order derivatives coincide with those of the

MLE are 3OE (e.g., Pfanzagl and Wefelmeyer [25]; Akahira and Takeuchi [3]; Amari [4]; Taniguchi [35]). Rao [26,27,28], Ghosh and Subramanyam [13], Efron [10] and Eguchi [11] have obtained 3OE of the MLE on the basis of quadratic loss or loss of information.

Here is an interesting question; does it hold that 3OE implies 4OE? On the issue, Ghosh [12](page 64) mentioned "given that 1OE implies 2OE, it is natural to conjecture that 3OE implies 4OE. The proof of that must be very messy."

The statement is known to be true if one takes quadratic loss as a criterion, because $B_4 = 0$. This fact would convince us that the conjecture be true under concentration probability as well as under quadratic loss.

In this paper, we distinguish between asymptotic efficiency for the class $\mathcal{C}$ and for the class $\mathcal{C}_0$. The idea was taken by Akahira and Takeuchi [3] to study 3OE of estimators in models with general density functions. We will show that 3OE implies 4OE for $\mathcal{C}$ or for $\mathcal{C}_0$, depending on a choice of bias-correction factors.

In Section 2, we give preliminary results on matrix derivatives and multivariate Hermite polynomials. Section 3 describes main results on asymptotic efficiency up to the fourth-order under several types of bias-correction factors. Proofs are given in Section 4.

The derivation of this paper is so-called formal, and we do not directly prove validity of the Edgeworth expansion (1.3). According to general theory, the formal expansion is actually valid because (i) the continuity of the curved exponential family (1.2) ensures the Cramer condition: $\limsup_{||\mathbf{s}||\to\infty} E[e^{i\mathbf{s}'(\mathbf{x}-\eta(\mathbf{u}))}] < 1$; (ii) the smoothness of $\psi(\theta)$ and $\theta(\mathbf{u})$ is assumed, so that the moments of any orders exist; (iii) the class $\mathcal{C}$ of all measurable convex sets meets a boundary condition: $\sup_{C\in\mathcal{C}} \int_{(\partial C)^\epsilon} N_q(\mathbf{x}|0, I_q)d\mathbf{x} = O(\epsilon)$ $(\epsilon \downarrow 0)$. For details, see Bhattacharya and Denker [9](Remark 1.4.1 and Theorem 2.1).

2. Preliminary Results

2.1. Matrix derivatives and symmetric tensor

Let $\mathbf{x} = [x_1, \dots, x_p]'$ and $A(\mathbf{x}) = [a_{ij}(\mathbf{x})]$ be an $a_1 \times a_2$ matrix-valued smooth function. Put $\frac{\partial A(\mathbf{x})}{\partial x_k} = \left[\frac{\partial a_{ij}(\mathbf{x})}{\partial x_k}\right]$. Define $\frac{\mathrm{d}A(\mathbf{x})}{\mathrm{d}\mathbf{x}'} = \frac{\mathrm{d}}{\mathrm{d}\mathbf{x}'}A(\mathbf{x}) = \left[\frac{\partial A(\mathbf{x})}{\partial x_1}, \dots, \frac{\partial A(\mathbf{x})}{\partial x_p}\right]$, and $\frac{\mathrm{d}}{\mathrm{d}\mathbf{x}}A(\mathbf{x}) = \left(\frac{\mathrm{d}}{\mathrm{d}\mathbf{x}'}A(\mathbf{x})'\right)'$. The higher-order (matrix) derivatives are defined inductively by $\left(\frac{\mathrm{d}}{\mathrm{d}\mathbf{x}'}\right)^{<k>} = \frac{\mathrm{d}}{\mathrm{d}\mathbf{x}'}\left(\frac{\mathrm{d}}{\mathrm{d}\mathbf{x}'}\right)^{<k-1>}$ with $\left(\frac{\mathrm{d}}{\mathrm{d}\mathbf{x}'}\right)^{<1>} = \frac{\mathrm{d}}{\mathrm{d}\mathbf{x}'}$. For such matrix derivatives, see Kano [17,18] (cf. Bentler and Lee [8]; Magnus and Neudecker [21]). Here we simply note some formulas on the matrix differentiation. The proof is due to direct calculation (see Kano [18]). Let $\mathbf{x}$ be a p-vector, let $A(\mathbf{x})$ and $B(\mathbf{x})$ be of order $a_1 \times a_2$ and $b_1 \times b_2$ and suppose that the following matrix products in the LHS are permissible. We then have

$$\begin{aligned}
\frac{\mathrm{d}}{\mathrm{d}\mathbf{x}'}[A(\mathbf{x})B(\mathbf{x})] =& \frac{\mathrm{d}A(\mathbf{x})}{\mathrm{d}\mathbf{x}'}(I_p \otimes B(\mathbf{x})) + A(\mathbf{x})\frac{\mathrm{d}B(\mathbf{x})}{\mathrm{d}\mathbf{x}'};\\
\frac{\mathrm{d}}{\mathrm{d}\mathbf{x}'}[A(\mathbf{x}) \otimes B(\mathbf{x})] =& \left(\frac{\mathrm{d}A(\mathbf{x})}{\mathrm{d}\mathbf{x}'} \otimes B(\mathbf{x})\right) + \left(A(\mathbf{x}) \otimes \frac{\mathrm{d}B(\mathbf{x})}{\mathrm{d}\mathbf{x}'}\right)(K_{a_2 p} \otimes I_{b_2});\\
\frac{\mathrm{d}}{\mathrm{d}\mathbf{t}'}[AB(\mathbf{x}(\mathbf{t}))C] =& A\frac{\mathrm{d}B(\mathbf{x})}{\mathrm{d}\mathbf{x}'}\left(\frac{\mathrm{d}\mathbf{x}(\mathbf{t})}{\mathrm{d}\mathbf{t}'} \otimes C\right),
\end{aligned}$$

where $\mathbf{x}(\mathbf{t})$ is an $\mathbb{R}^p$-valued function of $\mathbf{t}$ and A and C are constant matrices in the last formula. Here the matrix K_{ab} of order ab is a commutation matrix defined by the relation:

$$K_{ab}(\mathbf{b} \otimes \mathbf{a}) = \mathbf{a} \otimes \mathbf{b}$$

for arbitrary a-vector $\mathbf{a}$ and b-vector $\mathbf{b}$. See Henderson and Searle [14] and Magnus and Neudecker [21](page 47) for commutation matrices.

Let us simply write $A^{<k>} = \overbrace{A \otimes \cdots \otimes A}^{k\text{-fold}}$, k-fold right Kronecker (tensor) product of the same matrix A. Note $A^{<0>} = 1$. The symmetric tensor for the Kronecker product of p-vectors $\mathbf{a}_1, \ldots, \mathbf{a}_k$ is denoted by $N_{p^{<k>}}$, which operates as

$$N_{p^{<k>}}(\mathbf{a}_1 \otimes \cdots \otimes \mathbf{a}_k) = \sum \left(\mathbf{a}_{\sigma(1)} \otimes \cdots \otimes \mathbf{a}_{\sigma(k)}\right)/k!,$$

where the summation runs over all permutations $(\sigma(1), \ldots, \sigma(k))$ of $(1, \ldots, k)$. See e.g. Satake [30](Section 5.3) and Sternberg [31](Section 1.3) for the symmetric tensor. Let $\mathrm{g}(\mathbf{x})$ be an $\mathbb{R}^q$-valued analytic function defined on a neighborhood of $\mathbf{x}_0 \in \mathbb{R}^p$. The Taylor series of $\mathrm{g}(\mathbf{x})$ about $\mathbf{x}_0$ is expressible in the form:

$$\mathrm{g}(\mathbf{x}) = \sum_{k=0}^{\infty} \frac{G_k}{k!}(\mathbf{x} - \mathbf{x}_0)^{<k>} \quad \text{with} \quad G_k = \left(\frac{\mathrm{d}}{\mathrm{d}\mathbf{x}'}\right)^{<k>} \mathrm{g}(\mathbf{x})\Bigg|_{\mathbf{x}=\mathbf{x}_0}$$

The matrix of the derivatives, G_k, has an important property:

$$G_k N_{p^{<k>}} = G_k. \tag{2.1}$$

Define

$$\Psi_{k\ell} = \Psi_{k\ell}(\theta(\mathbf{u})) = \left(\frac{\mathrm{d}}{\mathrm{d}\theta}\right)^{<k>} \left(\frac{\mathrm{d}}{\mathrm{d}\theta'}\right)^{<\ell>} \psi(\theta)\Bigg|_{\theta=\theta(\mathbf{u})}. \tag{2.2}$$

2.2 Multivariate Hermite polynomials

Let $\varphi(\mathbf{s}) = e^{-\mathbf{s}'\Psi\mathbf{s}/2}$ with $\mathbf{s}$ a p-vector and Ψ $p \times p$ positive definite. The *covariant* Hermite polynomials are defined as

$$H_k(\mathbf{s}|\Psi) = \left[\left(-\frac{\mathrm{d}}{\mathrm{d}\mathbf{s}}\right)^{<k>} \varphi(\mathbf{s})\right] / \varphi(\mathbf{s}) \qquad (k \in \mathbb{N}),$$

and the *contravariant* Hermite polynomials are

$$\widetilde{H}_k(\mathbf{s}|\Psi) = (\Psi^{-1})^{<k>} H_k(\mathbf{s}|\Psi) \qquad (k \in \mathbb{N}).$$

The first few polynomials are

$$H_1(\mathbf{s}|\Psi) = \Psi\mathbf{s}, \qquad H_2(\mathbf{s}|\Psi) = (\Psi\mathbf{s})^{<2>} - \mathrm{vec}(\Psi);$$
$$H_3(\mathbf{s}|\Psi) = (\Psi\mathbf{s})^{<3>} - 3N_{p^{<3>}}(\mathrm{vec}(\Psi) \otimes \Psi\mathbf{s}),$$

and

$$\widetilde{H}_1(\mathbf{s}|\Psi) = \mathbf{s}, \qquad \widetilde{H}_2(\mathbf{s}|\Psi) = \mathbf{s}^{<2>} - \mathrm{vec}(\Psi^{-1});$$
$$\widetilde{H}_3(\mathbf{s}|\Psi) = \mathbf{s}^{<3>} - 3N_{p^{<3>}}(\mathrm{vec}(\Psi^{-1}) \otimes \mathbf{s}).$$

Let $\mathbf{z}$ have a multivariate normal distribution $N_p(0, \Psi)$. We then have $E[e^{i\mathbf{s}'\mathbf{z}}] = \varphi(\mathbf{s})$ and

$$\begin{aligned} E[e^{i\mathbf{s}'\mathbf{z}}\mathbf{z}^{<k>}\mathbf{z}^{<\ell>\prime}] &= i^{k+\ell}\left(-\frac{\mathrm{d}}{\mathrm{d}\mathbf{s}'}\right)^{<\ell>}\left(-\frac{\mathrm{d}}{\mathrm{d}\mathbf{s}}\right)^{<k>}\varphi(\mathbf{s}) \\ &= i^{k+\ell}\left(-\frac{\mathrm{d}}{\mathrm{d}\mathbf{s}'}\right)^{<\ell>}[H_k(\mathbf{s}|\Psi)\varphi(\mathbf{s})] \end{aligned}$$

for $k, \ell = 0, 1, 2, \ldots$ According to Barndorff-Nielsen and Cox [7](Eq. (5.67)) or Takemura [32](page 239), we know that $E[e^{i\mathbf{s}'\mathbf{z}}H_k(\mathbf{z}|\Psi^{-1})] = (i\mathbf{s})^{<k>}\varphi(\mathbf{s})$, and hence

$$E[e^{i\mathbf{s}'\mathbf{z}}\widetilde{H}_k(\mathbf{z}|\Psi^{-1})] = (i\Psi\mathbf{s})^{<k>}\varphi(\mathbf{s}). \tag{2.3}$$

We further note that

$$E[e^{i\mathbf{s}'\mathbf{z}}\widetilde{H}_k(\mathbf{z}|\Psi^{-1})\mathbf{z}^{<\ell>\prime}] = i^{k+\ell}\left(-\frac{\mathrm{d}}{\mathrm{d}\mathbf{s}'}\right)^{<\ell>}\left[(\Psi\mathbf{s})^{<k>}\varphi(\mathbf{s})\right]. \tag{2.4}$$

Recall that $\mathbf{x}_1, \ldots, \mathbf{x}_n$ are an i.i.d. sample from the curved exponential family (1.1). Let us replace $\mathbf{z}$ with $\mathbf{z}_n = \sqrt{n}(\bar{\mathbf{x}} - \eta(\mathbf{u}))$, and then the formulas above still hold provided that we substitute Ψ_{11} for Ψ and add $o(1)$; for example, the corresponding result to (2.3) is

$$E[e^{i\mathbf{s}'\mathbf{z}_n}\widetilde{H}_k(\mathbf{z}_n|\Psi_{11}^{-1})] = (i\Psi_{11}\mathbf{s})^{<k>}\varphi(\mathbf{s}) + o(1) \tag{2.5}$$

with $\varphi(\mathbf{s}) = e^{-\mathbf{s}'\Psi_{11}\mathbf{s}/2}$.

3. Main Results

We begin with second-order efficiency. Let $\mathbf{z}_n = \sqrt{n}(\bar{\mathbf{x}} - \eta(\mathbf{u}))$, $G_k = G_k(\eta(\mathbf{u})) = \left(\frac{\mathrm{d}}{\mathrm{d}\bar{\mathbf{x}}'}\right)^{<k>} \mathbf{g}(\bar{\mathbf{x}})\Big|_{\bar{\mathbf{x}}=\eta(\mathbf{u})}$ and $\bar{G}_k = \bar{G}_k(\eta(\mathbf{u})) = \left(\frac{\mathrm{d}}{\mathrm{d}\bar{\mathbf{x}}'}\right)^{<k>} \hat{\mathbf{u}}(\bar{\mathbf{x}})\Big|_{\bar{\mathbf{x}}=\eta(\mathbf{u})}$ for $k = 0, 1, 2, \ldots$ We expand $\mathbf{g}(\bar{\mathbf{x}})$ in $\mathcal{F}$ and the MLE $\hat{\mathbf{u}}(\bar{\mathbf{x}})$ stochastically in the form:

$$\begin{aligned} \mathbf{g}(\bar{\mathbf{x}}) &= \mathbf{u} + \frac{1}{\sqrt{n}}G_1\mathbf{z}_n + \frac{1}{2n}G_2\mathbf{z}_n^{<2>} + \frac{1}{6n\sqrt{n}}G_3\mathbf{z}_n^{<3>} + \cdots, \\ \hat{\mathbf{u}}(\bar{\mathbf{x}}) &= \mathbf{u} + \frac{1}{\sqrt{n}}\bar{G}_1\mathbf{z}_n + \frac{1}{2n}\bar{G}_2\mathbf{z}_n^{<2>} + \frac{1}{6n\sqrt{n}}\bar{G}_3\mathbf{z}_n^{<3>} + \cdots. \end{aligned} \tag{3.1}$$

The first-order bias is given as $E[\mathbf{g}(\bar{\mathbf{x}}) - \mathbf{u}] = \frac{1}{2n}G_2\Psi_{20} + O\left(\frac{1}{n^2}\right)$ with $\Psi_{20} = \mathrm{vec}(\Psi_{11}) = \left(\frac{\mathrm{d}}{\mathrm{d}\theta}\right)^{<2>}\psi(\theta)|_{\theta=\eta(\mathbf{u})}$.

To discuss 2OE, we take $\mathbf{b}_g^{(1)}(\bar{\mathbf{x}}) = \frac{1}{2}\widehat{G_2\Psi_{20}}$ and $\mathbf{b}_g^{(2)}(\bar{\mathbf{x}})$ to be unspecified. Then,

$$\begin{aligned} \mathbf{g}^*(\bar{\mathbf{x}}) &= \mathbf{g}(\bar{\mathbf{x}}) - \frac{1}{2n}\widehat{G_2\Psi_{20}} - \frac{1}{n^2}\mathbf{b}_g^{(2)}(\bar{\mathbf{x}}), \\ \hat{\mathbf{u}}^*(\bar{\mathbf{x}}) &= \hat{\mathbf{u}}(\bar{\mathbf{x}}) - \frac{1}{2n}\widehat{\bar{G}_2\Psi_{20}} - \frac{1}{n^2}\mathbf{b}_{\hat{u}}^{(2)}(\bar{\mathbf{x}}), \end{aligned} \tag{3.2}$$

and hence

$$E[\sqrt{n}(\mathbf{g}^*(\bar{\mathbf{x}}) - \hat{\mathbf{u}}^*(\bar{\mathbf{x}}))] = O\left(\frac{1}{n\sqrt{n}}\right). \tag{3.3}$$

Here we do not necessarily specify how to estimate $G_2\Psi_{20}$.

Takeuchi [33](page 185) and Pfanzagl [23](Section 7) showed the following theorem. See also Ghosh [12](Section 6.4).

Theorem 1. *Let $\mathbf{g}^*(\bar{\mathbf{x}}) \in \mathcal{F}$ and the MLE $\hat{\mathbf{u}}^*(\bar{\mathbf{x}}) \in \mathcal{F}$ be bias-corrected as in (3.2). Assume that $\mathbf{g}^*(\bar{\mathbf{x}})$ is 1OE. Then,*

$$P_{\theta(\mathbf{u})}[\sqrt{n}(\mathbf{g}^*(\bar{\mathbf{x}}) - \mathbf{u}) \in C] = P_{\theta(\mathbf{u})}[\sqrt{n}(\hat{\mathbf{u}}^*(\bar{\mathbf{x}}) - \mathbf{u}) \in C] + o\left(\frac{1}{\sqrt{n}}\right)$$

for every $C \in \mathcal{C}$ and $\mathbf{u} \in \mathcal{N}(\mathbf{u}_0)$, that is, under the bias-correction in (3.2), any 1OE estimators are 2OE for the class $\mathcal{C}$.

Can we say something about 2OE, without bias-correction? The following theorem gives a certain answer to the question.

Theorem 2. *Let* $\hat{\mathbf{u}}(\bar{\mathbf{x}})$ *be the MLE. Assume that a Fisher consistent estimator* $\mathbf{g}(\bar{\mathbf{x}})$ *is 1OE. Then,*

$$P_{\theta(\mathbf{u})}[\sqrt{n}(\mathbf{g}(\bar{\mathbf{x}})-\mathbf{u}) \in C] = P_{\theta(\mathbf{u})}[\sqrt{n}(\hat{\mathbf{u}}(\bar{\mathbf{x}})-\mathbf{u}) \in C] + o\left(\tfrac{1}{\sqrt{n}}\right)$$

for every $C \in \mathcal{C}_0$ *and* $\mathbf{u} \in \mathcal{N}(\mathbf{u}_0)$, *that is, any 1OE estimators are 2OE for the class* $\mathcal{C}_0$.

Some authors must have noticed the result of Theorem 2, but no explicit statement regarding 2OE for the class of no bias-corrected estimators has appeared. The next theorem on 3OE is well-known (e.g., Pfanzagl and Wefelmeyer [25] Theorem 1; Akahira and Takeuchi [3] Theorem 5.1.6; Ghosh [12] Chapter 6).

Theorem 3. *Let* $\mathbf{g}^*(\bar{\mathbf{x}}) \in \mathcal{F}$ *and the MLE* $\hat{\mathbf{u}}^*(\bar{\mathbf{x}}) \in \mathcal{F}$ *be bias-corrected as in (3.2). Assume that* $\mathbf{g}^*(\bar{\mathbf{x}})$ *is 1OE. Then,*

$$P_{\theta(\mathbf{u})}[\sqrt{n}(\mathbf{g}^*(\bar{\mathbf{x}})-\mathbf{u}) \in C] = P_{\theta(\mathbf{u})}[\sqrt{n}(\hat{\mathbf{u}}^*(\bar{\mathbf{x}})-\mathbf{u}) \in C] - \frac{1}{n}\Delta^2_{\mathbf{u},\mathbf{g}^*,C} + o\left(\tfrac{1}{n}\right)$$

for every $C \in \mathcal{C}_0$ *and* $\mathbf{u} \in \mathcal{N}(\mathbf{u}_0)$, *where* $\Delta^2_{\mathbf{u},\mathbf{g}^*,C}$ *is nonnegative. Further,* $\Delta^2_{\mathbf{u},\mathbf{g}^*,C} = 0$ *for any* $C \in \mathcal{C}_0$ *if and only if* $G_2 = \bar{G}_2$, *that is, under the bias-correction in (3.2), a 1OE estimator is 3OE for the class* $\mathcal{C}_0$ *if and only if* $G_2 = \bar{G}_2$.

The actual form of $\Delta^2_{\mathbf{u},\mathbf{g}^*,C}$ will be given in (4.13) in the proof. According to Theorem 3, we see that 3OE of $\mathbf{g}^*(\bar{\mathbf{x}})$ means $G_1 = \bar{G}_1$ and $G_2 = \bar{G}_2$.

Now we are in a position to state some results on 4OE. The distribution of $\sqrt{n}(\mathbf{b}_g^{(1)} - \frac{1}{2}G_2\Psi_{20}) = \sqrt{n}(\frac{1}{2}\widehat{G_2\Psi_{20}} - \frac{1}{2}G_2\Psi_{20})$ contributes to the terms of order $O\left(\frac{1}{n\sqrt{n}}\right)$ of concentration probability, and hence how to estimate $G_2\Psi_{20}$ is quite important to study 4OE.

Consider the following different types of bias-corrections (or different choices of $\mathbf{b}_g^{(1)}$ and $\mathbf{b}_g^{(2)}$):

Case (I): first-order bias-correction only

$$\mathbf{g}^*(\bar{\mathbf{x}}) = \mathbf{g}(\bar{\mathbf{x}}) - \tfrac{1}{2n}\widehat{G_2\Psi_{20}};$$

Case (II): second-order bias-correction (1), the historical one (see e.g., Rao et.al. [29])

$$\mathbf{g}^*(\bar{\mathbf{x}}) = \mathbf{g}(\bar{\mathbf{x}}) - \tfrac{1}{2n}G_2(\eta(\mathbf{g}(\bar{\mathbf{x}})))\Psi_{20}(\theta(\mathbf{g}(\bar{\mathbf{x}}))) - \tfrac{1}{n^2}\mathbf{b}_g^{(2)}(\bar{\mathbf{x}})$$

so that $E[\mathbf{g}^*(\bar{\mathbf{x}}) - \mathbf{u}] = O\left(\frac{1}{n^3}\right)$;

Case (III): second-order bias-correction (2), suggested by Kano [17,19]

$$\mathbf{g}^*(\bar{\mathbf{x}}) = \mathbf{g}(\bar{\mathbf{x}}) - \tfrac{1}{2n}G_2(\bar{\mathbf{x}})\Psi_{20}(\theta(\mathbf{g}(\bar{\mathbf{x}}))) - \tfrac{1}{n^2}\mathbf{b}_g^{(2)}(\bar{\mathbf{x}})$$

so that $E[\mathbf{g}^*(\bar{\mathbf{x}}) - \mathbf{u}] = O\left(\frac{1}{n^3}\right)$;

Case (IV): second-order bias-correction (3), based on Amari [5](Eq. (4.27))

$$\mathbf{g}^*(\bar{\mathbf{x}}) = \mathbf{g}(\bar{\mathbf{x}}) - \tfrac{1}{2n}G_2(\bar{\mathbf{x}})\Psi_{20}(\Psi_{10}^{-1}(\bar{\mathbf{x}})) - \tfrac{1}{n^2}\mathbf{b}_g^{(2)}(\bar{\mathbf{x}})$$

so that $E[\mathbf{g}^*(\bar{\mathbf{x}}) - \mathbf{u}] = O\left(\frac{1}{n^3}\right)$.

Here $\theta = \Psi_{10}^{-1}(\eta)$ is the inverse function of $\eta = \Psi_{10}(\theta)$. For Cases (II)-(IV),

$$E[\sqrt{n}(\mathbf{g}^*(\bar{\mathbf{x}}) - \hat{\mathbf{u}}^*(\bar{\mathbf{x}}))] = O\left(\frac{1}{n^2\sqrt{n}}\right), \tag{3.4}$$

whereas (3.3) holds for Case (I).

We now state the main theorem.

Theorem 4. *Let* $\mathbf{g}^*(\bar{\mathbf{x}}) \in \mathcal{F}$ *and the MLE* $\hat{\mathbf{u}}^*(\bar{\mathbf{x}}) \in \mathcal{F}$ *be bias-corrected as in one of the four cases above. Assume that* $\mathbf{g}^*(\bar{\mathbf{x}})$ *is 3OE. (i) Under Cases (I) and (II),*

$$P_{\theta(\mathbf{u})}[\sqrt{n}(\mathbf{g}^*(\bar{\mathbf{x}}) - \mathbf{u}) \in C] = P_{\theta(\mathbf{u})}[\sqrt{n}(\hat{\mathbf{u}}^*(\bar{\mathbf{x}}) - \mathbf{u}) \in C] + o\left(\frac{1}{n\sqrt{n}}\right) \tag{3.5}$$

for every $C \in \mathcal{C}_0$ *and* $\mathbf{u} \in \mathcal{N}(\mathbf{u}_0)$, *that is, any 3OE estimators are 4OE for the class* $\mathcal{C}_0$. *(ii) Under Cases (III) and (IV), (3.5) holds for the class* $\mathcal{C}$, *that is, any 3OE estimators are 4OE for the class* $\mathcal{C}$.

Kano [17,19] and Amari [6] have established 5OE of the MLE under the bias-correction (III) or (IV) on the basis of quadratic loss, concentration probability and information loss, respectively.

4. Proofs

4.1 Auxiliary results

Differentiation of both sides in the equality $G_k(\Psi_{10}(\theta(\mathbf{u}))) = \bar{G}_k(\Psi_{10}(\theta(\mathbf{u})))$ w.r.t. $\mathbf{u}'$ and use of (2.1) shows that

(4.1a)
$$G_1 = \bar{G}_1 \implies (G_2 - \bar{G}_2)(\Psi_{11}\Theta_1 \otimes I_p) = (G_2 - \bar{G}_2)(I_p \otimes \Psi_{11}\Theta_1) = 0,$$

(4.1b)
$$\begin{aligned} G_2 = \bar{G}_2 \implies (G_3 - \bar{G}_3)(\Psi_{11}\Theta_1 \otimes I_{p^2}) &= (G_2 - \bar{G}_2)(I_p \otimes \Psi_{11}\Theta_1 \otimes I_p) \\ &= (G_2 - \bar{G}_2)(I_{p^2} \otimes \Psi_{11}\Theta_1) \\ &= 0. \end{aligned}$$

These are key relations to our derivation here.

Let $\mathbf{t} \in \mathbb{R}^q$ and $\varphi(\mathbf{s}) = e^{-\mathbf{s}'\Psi_{11}\mathbf{s}/2}$. Recall that $\bar{G}_1 = i_{\mathbf{u}}^{-1}\Theta_1'$. It follows from (2.3) and (4.1a) that

$$\begin{aligned} (G_2 - \bar{G}_2)E[e^{i\mathbf{t}'\bar{G}_1\mathbf{z}_n}(\mathbf{z}_n^{<2>} - \Psi_{20})] &= -(G_2 - \bar{G}_2)(\Psi_{11}\Theta_1 i_{\mathbf{u}}^{-1})^{<2>}\varphi(\bar{G}_1'\mathbf{t}) + o(1) \\ &= o(1) \end{aligned} \tag{4.2a}$$

and hence

$$(4.2b) \qquad (G_2 - \bar{G}_2)E[e^{it'\bar{G}_1 \mathbf{z}_n} \mathbf{z}_n^{<2>}] = (G_2 - \bar{G}_2)\Psi_{20} e^{-\mathbf{t}' i_u^{-1} \mathbf{t}/2} + o(1).$$

Using (2.4), we have

$$E[e^{it'\bar{G}_1 \mathbf{z}_n}(\mathbf{z}_n^{<2>} - \Psi_{20})\mathbf{z}_n'] = i\frac{\mathrm{d}}{\mathrm{ds}'}\left[(\Psi_{11}\mathbf{s})^{<2>}\varphi(\mathbf{s})\right]\Big|_{\mathbf{s}=\bar{G}_1'\mathbf{t}} + o(1)$$

$$(4.3a) \qquad = ie^{-\mathbf{t}' i_u^{-1} \mathbf{t}/2} N_{p<2>}\left[(2\Psi_{11} - \Psi_{11}\bar{G}_1'\mathbf{t}\mathbf{t}'\bar{G}_1\Psi_{11}) \otimes (\Psi_{11}\bar{G}_1'\mathbf{t})\right] + o(1),$$

and hence use of (4.1a) results in

$$(4.3b) \qquad (G_2 - \bar{G}_2)E[e^{it'\bar{G}_1 \mathbf{z}_n}(\mathbf{z}_n^{<2>} - \Psi_{20})\mathbf{z}_n'] = o(1).$$

In a similar manner, we have from (2.4) and (4.1a)

$$\begin{aligned}(G_2 - \bar{G}_2)&E[e^{it'\bar{G}_1 \mathbf{z}_n}(\mathbf{z}_n^{<2>} - \Psi_{20})\mathbf{z}_n^{<2>\prime}] \\ &= (G_2 - \bar{G}_2)\left(\frac{\mathrm{d}}{\mathrm{ds}'}\right)^{<2>}\left[(\Psi_{11}\mathbf{s})^{<2>}\varphi(\mathbf{s})\right]\Big|_{\mathbf{s}=\bar{G}_1'\mathbf{t}} + o(1) \\ &= 2e^{-\mathbf{t}' i_u^{-1} \mathbf{t}/2}(G_2 - \bar{G}_2)\Psi_{11}^{<2>} + o(1),\end{aligned}$$

from which, it follows that

$$(4.4a) \qquad (G_2 - \bar{G}_2)E[e^{it'\bar{G}_1 \mathbf{z}_n}(\mathbf{z}_n^{<2>} - \Psi_{20})\mathbf{z}_n^{<2>\prime}]\bar{G}_2' = o(1)$$

and

$$\begin{aligned}&(G_2 - \bar{G}_2)E[e^{it'\bar{G}_1 \mathbf{z}_n}(\mathbf{z}_n^{<2>} - \Psi_{20})(\mathbf{z}_n^{<2>} - \Psi_{20})'](G_2 - \bar{G}_2)' \\ (4.4b) \qquad &\qquad = 2e^{-\mathbf{t}' i_u^{-1} \mathbf{t}/2}(G_2 - \bar{G}_2)\Psi_{11}^{<2>}(G_2 - \bar{G}_2)' + o(1),\end{aligned}$$

We have used that $\bar{G}_2 = B(\Theta_1' \otimes I_p)N_{p<2>}$ for some B (see Kano [17] Lemma 3.3) and have applied (4.1a) again to get (4.4a).

Using (2.4) and (4.1b) again, we have

$$\begin{aligned}(G_3 - \bar{G}_3)&E[e^{it'\bar{G}_1 \mathbf{z}_n}(\mathbf{z}_n^{<3>} - 3(\Psi_{20} \otimes \mathbf{z}_n)\mathbf{z}_n^{<\ell>\prime}] \\ &= i^{3+\ell}(G_3 - \bar{G}_3)\left(-\frac{\mathrm{d}}{\mathrm{ds}'}\right)^{<\ell>}\left[(\Psi_{11}\mathbf{s})^{<3>}\varphi(\mathbf{s})\right]\Big|_{\mathbf{s}=\bar{G}_1'\mathbf{t}} + o(1) \\ (4.5) \qquad &= o(1)\end{aligned}$$

for $\ell = 0, 1, 2$.

4.2 Proofs of the theorems

Theorem 1 is a known result. We will give a proof of Theorem 2 and then note the distinction between the proofs of Theorems 1 and 2.

Since $G_1 = \bar{G}_1$ for any 1OE estimators $\mathbf{g}(\bar{\mathbf{x}})$, the difference is expressed from (3.1) as

$$\sqrt{n}(\mathbf{g}(\bar{\mathbf{x}}) - \hat{\mathbf{u}}(\bar{\mathbf{x}})) = \frac{1}{2\sqrt{n}}(G_2 - \bar{G}_2)\mathbf{z}_n^{<2>} + o_p\left(\tfrac{1}{\sqrt{n}}\right) = O_p\left(\tfrac{1}{\sqrt{n}}\right), \tag{4.6}$$

from which, the characteristic function of standardized $\mathbf{g}(\bar{\mathbf{x}})$ can be evaluated in the form:

$$\begin{aligned} E[e^{i\mathbf{t}'\sqrt{n}(\mathbf{g}(\bar{\mathbf{x}})-\mathbf{u})}] =& E[e^{i\mathbf{t}'\sqrt{n}(\hat{\mathbf{u}}(\bar{\mathbf{x}})-\mathbf{u})} \cdot e^{i\mathbf{t}'\sqrt{n}(\mathbf{g}(\bar{\mathbf{x}})-\hat{\mathbf{u}}(\bar{\mathbf{x}}))}] \\ =& E[e^{i\mathbf{t}'\sqrt{n}(\hat{\mathbf{u}}(\bar{\mathbf{x}})-\mathbf{u})}\{1 + i\mathbf{t}'\sqrt{n}(\mathbf{g}(\bar{\mathbf{x}}) - \hat{\mathbf{u}}(\bar{\mathbf{x}})) + O_p\left(\tfrac{1}{n}\right)\}] \\ =& E[e^{i\mathbf{t}'\sqrt{n}(\hat{\mathbf{u}}(\bar{\mathbf{x}})-\mathbf{u})}] \\ & + (i\mathbf{t})'E[e^{i\mathbf{t}'\sqrt{n}(\hat{\mathbf{u}}(\bar{\mathbf{x}})-\mathbf{u})}\sqrt{n}(\mathbf{g}(\bar{\mathbf{x}}) - \hat{\mathbf{u}}(\bar{\mathbf{x}}))] + o\left(\tfrac{1}{\sqrt{n}}\right), \end{aligned} \tag{4.7}$$

and

$$\begin{aligned} E[e^{i\mathbf{t}'\sqrt{n}(\hat{\mathbf{u}}(\bar{\mathbf{x}})-\mathbf{u})}\sqrt{n}(\mathbf{g}(\bar{\mathbf{x}}) - \hat{\mathbf{u}}(\bar{\mathbf{x}}))] =& \frac{1}{2\sqrt{n}}(i\mathbf{t})'(G_2 - \bar{G}_2)E[e^{i\mathbf{t}'\bar{G}_1\mathbf{z}_n}\mathbf{z}_n^{<2>}] + o\left(\tfrac{1}{\sqrt{n}}\right) \\ =& \frac{e^{-\mathbf{t}'i_{\mathbf{u}}^{-1}\mathbf{t}/2}}{2\sqrt{n}}(i\mathbf{t})'(G_2 - \bar{G}_2)\Psi_{20} + o\left(\tfrac{1}{\sqrt{n}}\right) \end{aligned}$$

in view of (4.2b). Thus, we obtain the difference between the two characteristic functions of $\mathbf{g}(\bar{\mathbf{x}})$ and $\hat{\mathbf{u}}(\bar{\mathbf{x}})$ as

$$E[e^{i\mathbf{t}'\sqrt{n}(\mathbf{g}(\bar{\mathbf{x}})-\mathbf{u})}] - E[e^{i\mathbf{t}'\sqrt{n}(\hat{\mathbf{u}}(\bar{\mathbf{x}})-\mathbf{u})}] = \frac{e^{-\mathbf{t}'i_{\mathbf{u}}^{-1}\mathbf{t}/2}}{2\sqrt{n}}(i\mathbf{t})'(G_2 - \bar{G}_2)\Psi_{20} + o\left(\tfrac{1}{\sqrt{n}}\right).$$

Inversion of the characteristic functions to distribution functions results in

$$\begin{aligned} &\mathrm{P}_{\theta(\mathbf{u})}[\sqrt{n}(\mathbf{g}(\bar{\mathbf{x}}) - \mathbf{u}) \in C] - \mathrm{P}_{\theta(\mathbf{u})}[\sqrt{n}(\hat{\mathbf{u}}(\bar{\mathbf{x}}) - \mathbf{u}) \in C] \\ &\qquad = \frac{1}{2\sqrt{n}}\Psi_{20}'(G_2 - \bar{G}_2)' \int_C \mathbf{x}N_q(\mathbf{x}|0, i_{\mathbf{u}}^{-1})d\mathbf{x} + o\left(\tfrac{1}{\sqrt{n}}\right) \end{aligned} \tag{4.8}$$

for every measurable set $C \subset \mathrm{R}^q$ meeting a certain condition on the boundary ∂C, which is satisfied by convex sets (see Bhattacharya and Denker [9] Remark 1.4.1 and Theorem 2.1). The integrand in (4.8) is an odd function in $\mathbf{x}$. The RHS in (4.8) vanishes when C is symmetric about the origin. As a result, we have

$$\mathrm{P}_{\theta(\mathbf{u})}[\sqrt{n}(\mathbf{g}(\bar{\mathbf{x}}) - \mathbf{u}) \in C] = \mathrm{P}_{\theta(\mathbf{u})}[\sqrt{n}(\hat{\mathbf{u}}(\bar{\mathbf{x}}) - \mathbf{u}) \in C] + o\left(\tfrac{1}{\sqrt{n}}\right)$$

for every $C \in \mathcal{C}_0$. This proves Theorem 2. *Q.E.D.*

When the bias-correction in (3.2) is made, (4.6) becomes

$$\sqrt{n}(\mathbf{g}^*(\bar{\mathbf{x}}) - \hat{\mathbf{u}}^*(\bar{\mathbf{x}})) = \frac{1}{2\sqrt{n}}(G_2 - \bar{G}_2)(\mathbf{z}_n^{<2>} - \Psi_{20}) + o_p\left(\tfrac{1}{\sqrt{n}}\right), \tag{4.9}$$

and hence we get

$$E[e^{i\mathbf{t}'\sqrt{n}(\hat{\mathbf{u}}^*(\bar{\mathbf{x}})-\mathbf{u})}\sqrt{n}(\mathbf{g}^*(\bar{\mathbf{x}}) - \hat{\mathbf{u}}^*(\bar{\mathbf{x}}))] = o\left(\tfrac{1}{\sqrt{n}}\right) \tag{4.10}$$

by the virtue of (4.2a). Consequently, the difference between the characteristic functions is of order $o\left(\frac{1}{\sqrt{n}}\right)$. This proves Theorem 1. *Q.E.D.*

To prove Theorems 3 and 4, we need to investigate the expectation (4.10), a sort of cross terms, up to higher-order. There are useful lemmas to do so. The basic idea of the lemmas were originated in Kano [19], who studies fifth-order efficiency. The proof will be given in Appendix.

Lemma 1. *Assume that $\mathbf{x}_1, \ldots, \mathbf{x}_n$ are an i.i.d. random sample from the curved exponential family (1.1). Let $T_n^{(1)} = T_n^{(1)}(\mathbf{x}_1, \ldots, \mathbf{x}_n)$ and $T_n^{(2)} = T_n^{(2)}(\mathbf{x}_1, \ldots, \mathbf{x}_n)$ be random q- and q'-vectors which may depend on the sample size n but do not depend on the parameter $\mathbf{u} \in \Omega$. Let $\phi = \phi(\mathbf{u}, \mathbf{t}) = E[e^{i\mathbf{t}'\sqrt{n}(T_n^{(1)}-\mathbf{u})}T_n^{(2)}]$. Then, ϕ satisfies the following system of partial differential equations:*

$$\frac{d\phi}{d\mathbf{t}'} = -\{(\mathbf{t}'i_{\mathbf{u}}^{-1}) \otimes \phi\} + C(\mathbf{u}, \mathbf{t}), \tag{4.11}$$

where

$$C(\mathbf{u}, \mathbf{t}) = \frac{i}{\sqrt{n}}\frac{d\phi}{d\mathbf{u}'}i_{\mathbf{u}}^{-1} + iE[e^{i\mathbf{t}'\sqrt{n}(T_n^{(1)}-\mathbf{u})}T_n^{(2)}\{\sqrt{n}(T_n^{(1)} - \mathbf{u}) - \bar{G}_1\mathbf{z}_n\}'].$$

Lemma 2. *Let $\mathbf{t}_k = [t_1, \ldots, t_k, 0, \ldots, 0]'$ and $C(\mathbf{u}, \mathbf{t}) = [\mathbf{c}_1(\mathbf{u}, \mathbf{t}), \ldots, \mathbf{c}_q(\mathbf{u}, \mathbf{t})]$. The solution to (4.11) is expressible as*

$$\phi(\mathbf{u}, \mathbf{t}) = e^{-\mathbf{t}'i_{\mathbf{u}}^{-1}\mathbf{t}/2}\left(\phi(\mathbf{u}, 0) + \sum_{k=1}^{q}\int_0^{t_k}\mathbf{c}_k(\mathbf{u}, \mathbf{t}_k)e^{\mathbf{t}_k'i_{\mathbf{u}}^{-1}\mathbf{t}_k/2}dt_k\right).$$

Here we shall use these lemmas to provide a proof of Theorem 3, which is different and simpler compared to the existing proofs. In a similar way as in (4.7), we have

(4.12)

$$\begin{aligned}
&E[e^{i\mathbf{t}'\sqrt{n}(\mathbf{g}^*(\bar{\mathbf{x}})-\mathbf{u})}] - E[e^{i\mathbf{t}'\sqrt{n}(\hat{\mathbf{u}}^*(\bar{\mathbf{x}})-\mathbf{u})}] \\
&\quad = (i\mathbf{t})'E[e^{i\mathbf{t}'\sqrt{n}(\hat{\mathbf{u}}^*(\bar{\mathbf{x}})-\mathbf{u})}\sqrt{n}(\mathbf{g}^*(\bar{\mathbf{x}}) - \hat{\mathbf{u}}^*(\bar{\mathbf{x}}))] \\
&\quad + \frac{1}{2}(i\mathbf{t})'E[e^{i\mathbf{t}'\sqrt{n}(\hat{\mathbf{u}}^*(\bar{\mathbf{x}})-\mathbf{u})}\sqrt{n}(\mathbf{g}^*(\bar{\mathbf{x}}) - \hat{\mathbf{u}}^*(\bar{\mathbf{x}}))\sqrt{n}(\mathbf{g}^*(\bar{\mathbf{x}}) - \hat{\mathbf{u}}^*(\bar{\mathbf{x}}))'](i\mathbf{t}) \\
&\quad + o\left(\tfrac{1}{n}\right).
\end{aligned}$$

The first term in the RHS in (4.12) is of order $o\left(\frac{1}{n}\right)$ as will be shown later and so, use of (4.9) and (4.4b) results in

$$
\begin{aligned}
E[e^{i\mathbf{t}'\sqrt{n}(\mathbf{g}^*(\bar{\mathbf{x}})-\mathbf{u})}]-&E[e^{i\mathbf{t}'\sqrt{n}(\hat{\mathbf{u}}^*(\bar{\mathbf{x}})-\mathbf{u})}] \\
&= \frac{e^{-\mathbf{t}'i_{\mathbf{u}}^{-1}\mathbf{t}/2}}{4n}(i\mathbf{t})'(G_2-\bar{G}_2)\Psi_{11}^{<2>}(G_2-\bar{G}_2)'(i\mathbf{t})+o\left(\tfrac{1}{n}\right).
\end{aligned}
$$

Inverting the characteristic functions to distribution functions, we have

$$
\mathrm{P}_{\theta(\mathbf{u})}[\sqrt{n}(\mathbf{g}^*(\bar{\mathbf{x}})-\mathbf{u})\in C]=\mathrm{P}_{\theta(\mathbf{u})}[\sqrt{n}(\hat{\mathbf{u}}^*(\bar{\mathbf{x}})-\mathbf{u})\in C]-\frac{1}{n}\Delta^2_{\mathbf{u},\mathbf{g}^*,C}+o\left(\tfrac{1}{n}\right)
$$

for any Borel convex set $C\subset \mathbb{R}^q$, where

$$
\begin{aligned}
\Delta^2_{\mathbf{u},\mathbf{g}^*,C} &= \frac{1}{4}\mathrm{vec}(G_2-\bar{G}_2)' \\
&\cdot\left(\Psi_{11}^{<2>}\otimes\int_C(i_{\mathbf{u}}-i_{\mathbf{u}}\mathbf{x}\mathbf{x}'i_{\mathbf{u}})N_q(\mathbf{x}|0,i_{\mathbf{u}}^{-1})d\mathbf{x}\right)\mathrm{vec}(G_2-\bar{G}_2).
\end{aligned}
\tag{4.13}
$$

If C is symmetric about the origin (i.e., $C\in\mathcal{C}_0$), the integral above is shown to be nonnegative; and it is positive, assuming further that the interior point of C is not empty (see Pfanzagl [24] Lemma 13.2.4). Thus, Theorem 3 follows, provided that

$$
\phi=\phi(\mathbf{u},\mathbf{t})=E[e^{i\mathbf{t}'\sqrt{n}(\hat{\mathbf{u}}^*(\bar{\mathbf{x}})-\mathbf{u})}\sqrt{n}(\mathbf{g}^*(\bar{\mathbf{x}})-\hat{\mathbf{u}}^*(\bar{\mathbf{x}}))]=o\left(\tfrac{1}{n}\right).
$$

Applying Lemma 1 $[T_n^{(1)}=\hat{\mathbf{u}}^*(\bar{\mathbf{x}}),\ T_n^{(2)}=\sqrt{n}(\mathbf{g}^*(\bar{\mathbf{x}})-\hat{\mathbf{u}}^*(\bar{\mathbf{x}}))]$ and using (4.9), we can evaluate $C(\mathbf{u},\mathbf{t})$ in Lemma 1 as

$$
\begin{aligned}
C(\mathbf{u},\mathbf{t}) &= \frac{i}{\sqrt{n}}\frac{\mathrm{d}\phi}{\mathrm{d}\mathbf{u}'}i_{\mathbf{u}}^{-1}+\frac{i}{4n}(G_2-\bar{G}_2)E[e^{i\mathbf{t}'\bar{G}_1\mathbf{z}_n}(\mathbf{z}_n^{<2>}-\Psi_{20})(\mathbf{z}_n^{<2>}-\Psi_{20})']\bar{G}_2'+o\left(\tfrac{1}{n}\right) \\
&= \frac{i}{\sqrt{n}}\frac{\mathrm{d}\phi}{\mathrm{d}\mathbf{u}'}i_{\mathbf{u}}^{-1}+o\left(\tfrac{1}{n}\right)
\end{aligned}
$$

in view of (4.4a). We have from Lemma 2 and a consequence of the bias-correction

$$
\begin{aligned}
\phi =& e^{-\mathbf{t}'i_{\mathbf{u}}^{-1}\mathbf{t}/2}\left(E[\sqrt{n}(\mathbf{g}^*(\bar{\mathbf{x}})-\hat{\mathbf{u}}^*(\bar{\mathbf{x}}))]+\frac{i}{\sqrt{n}}\sum_{i=1}^{q}\int_0^{t_k}\frac{\mathrm{d}\phi(\mathbf{u},\mathbf{t}_k)}{\mathrm{d}\mathbf{u}'}e^{\mathbf{t}_k'i_{\mathbf{u}}^{-1}\mathbf{t}_k/2}dt_k i_{\mathbf{u}}^{-1}\right)+o\left(\tfrac{1}{n}\right) \\
=& \frac{ie^{-\mathbf{t}'i_{\mathbf{u}}^{-1}\mathbf{t}/2}}{\sqrt{n}}\sum_{i=1}^{q}\int_0^{t_k}\frac{\mathrm{d}\phi(\mathbf{u},\mathbf{t}_k)}{\mathrm{d}\mathbf{u}'}e^{\mathbf{t}_k'i_{\mathbf{u}}^{-1}\mathbf{t}_k/2}dt_k i_{\mathbf{u}}^{-1}+o\left(\tfrac{1}{n}\right).
\end{aligned}
$$

Substitution of $\phi=o\left(\frac{1}{\sqrt{n}}\right)$ shown in (4.10) into the integrand in the RHS above leads to $\phi=o\left(\frac{1}{n}\right)$. The proof of Theorem 3 is complete. *Q.E.D.*

Now we shall give a proof of Theorem 4. Consider Cases (III) and (IV) first. The bias-correction factors can be expanded as

$$\begin{aligned}\frac{1}{2\sqrt{n}}G_2(\bar{\mathbf{x}})\Psi_{20}(\theta(\mathbf{g}(\bar{\mathbf{x}}))) =& \frac{1}{2\sqrt{n}}G_2\Psi_{20} + \frac{1}{2n}G_2\Psi_{21}\Theta_1 G_1\mathbf{z}_n \\ &+ \frac{1}{2n}G_3(\Psi_{20}\otimes I_p)\mathbf{z}_n + o_p\left(\tfrac{1}{n}\right), \\ \frac{1}{2\sqrt{n}}G_2(\bar{\mathbf{x}})\Psi_{20}(\Psi_{10}^{-1}(\bar{\mathbf{x}})) =& \frac{1}{2\sqrt{n}}G_2\Psi_{20} + \frac{1}{2n}G_2\Psi_{21}\Psi_{11}^{-1}\mathbf{z}_n \\ &+ \frac{1}{2n}G_3(\Psi_{20}\otimes I_p)\mathbf{z}_n + o_p\left(\tfrac{1}{n}\right)\end{aligned}$$

and thus for a 3OE estimator $\mathbf{g}^*(\bar{\mathbf{x}})$ (i.e., $G_1 = \bar{G}_1$ and $G_2 = \bar{G}_2$) we have

$$\sqrt{n}(\mathbf{g}^*(\bar{\mathbf{x}}) - \hat{\mathbf{u}}^*(\bar{\mathbf{x}})) = \frac{1}{6n}(G_3 - \bar{G}_3)\{\mathbf{z}_n^{<3>} - 3(\Psi_{20}\otimes\mathbf{z}_n)\} + o_p\left(\tfrac{1}{n}\right) = O_p\left(\tfrac{1}{n}\right)$$

for the both cases. For all the cases, we shall write

$$\begin{aligned}\sqrt{n}(\mathbf{g}^*(\bar{\mathbf{x}}) - \hat{\mathbf{u}}^*(\bar{\mathbf{x}})) =& \frac{1}{6n}(G_3 - \bar{G}_3)\{\mathbf{z}_n^{<3>} - 3(\Psi_{20}\otimes\mathbf{z}_n)\} + \frac{1}{2n}B(\mathbf{u})\mathbf{z}_n + o_p\left(\tfrac{1}{n}\right) \\ =& O_p\left(\tfrac{1}{n}\right),\end{aligned} \tag{4.14}$$

where $B(\mathbf{u})$ is a constant matrix of order $q\times p$ given as

$$B(\mathbf{u}) = \begin{cases} (G_3 - \bar{G}_3)(\Psi_{20}\otimes I_p) - \frac{\mathrm{d}}{\mathrm{d}\bar{\mathbf{x}}'}[\widehat{G_2\Psi_{20}} - \widehat{\bar{G}_2\Psi_{20}}]|_{\bar{\mathbf{x}}=\eta(\mathbf{u})} & \text{for Case (I)} \\ (G_3 - \bar{G}_3)(\Psi_{20}\otimes I_p) - (G_3 - \bar{G}_3)(\Psi_{21}\Theta_1\bar{G}_1\otimes\Psi_{20}) & \text{for Case (II)} \\ 0 & \text{for Cases (III), (IV).}\end{cases}$$

Since $\sqrt{n}(\mathbf{g}^*(\bar{\mathbf{x}}) - \hat{\mathbf{u}}^*(\bar{\mathbf{x}})) = O_p\left(\frac{1}{n}\right)$, the same derivation as in (4.7) can apply to get

$$\begin{aligned}E[e^{i\mathbf{t}'\sqrt{n}(\mathbf{g}^*(\bar{\mathbf{x}})-\mathbf{u})}] - &E[e^{i\mathbf{t}'\sqrt{n}(\hat{\mathbf{u}}^*(\bar{\mathbf{x}})-\mathbf{u})}] \\ &= (i\mathbf{t})'E[e^{i\mathbf{t}'\sqrt{n}(\hat{\mathbf{u}}^*(\bar{\mathbf{x}})-\mathbf{u})}\sqrt{n}(\mathbf{g}^*(\bar{\mathbf{x}}) - \hat{\mathbf{u}}^*(\bar{\mathbf{x}}))] + o\left(\frac{1}{n\sqrt{n}}\right).\end{aligned}$$

Put

$$\phi = \phi(\mathbf{u},\mathbf{t}) = E[e^{i\mathbf{t}'\sqrt{n}(\hat{\mathbf{u}}^*(\bar{\mathbf{x}})-\mathbf{u})}\sqrt{n}(\mathbf{g}^*(\bar{\mathbf{x}}) - \hat{\mathbf{u}}^*(\bar{\mathbf{x}}))] \quad (= O\left(\tfrac{1}{n}\right)),$$

and then the inversion of the Fourier transform is made, so that

$$\begin{aligned}&\mathrm{P}_{\theta(\mathbf{u})}\{\sqrt{n}(\mathbf{g}^*(\bar{\mathbf{x}}) - \mathbf{u}) \in C\} \\ &\quad = \mathrm{P}_{\theta(\mathbf{u})}\{\sqrt{n}(\hat{\mathbf{u}}^*(\bar{\mathbf{x}}) - \mathbf{u}) \in C\} \\ &\qquad + \int_C \frac{1}{(2\pi)^{q/2}}\int_{\mathbf{R}^q} e^{-i\mathbf{x}'\mathbf{t}}(i\mathbf{t})'\phi(\mathbf{u},\mathbf{t})d\mathbf{t}d\mathbf{x} + o\left(\frac{1}{n\sqrt{n}}\right)\end{aligned} \tag{4.15}$$

for every Borel convex set $C \subset \mathbf{R}^q$. Let us apply Lemmas 1 and 2 to evaluate ϕ. We have

$$
\begin{aligned}
&iE[e^{i\mathbf{t}'\sqrt{n}(\hat{\mathbf{u}}^*(\bar{\mathbf{x}})-\mathbf{u})}\sqrt{n}(\mathbf{g}^*(\bar{\mathbf{x}})-\hat{\mathbf{u}}^*(\bar{\mathbf{x}}))\{\sqrt{n}(\hat{\mathbf{u}}^*(\bar{\mathbf{x}})-\mathbf{u})-\bar{G}_1\mathbf{z}_n\}'] \\
&\quad=\frac{i}{n\sqrt{n}}E[e^{i\mathbf{t}'\bar{G}_1\mathbf{z}_n}\{\tfrac{1}{6}(G_3-\bar{G}_3)(\mathbf{z}_n^{<3>}-3(\Psi_{20}\otimes\mathbf{z}_n))+\tfrac{1}{2}B(\mathbf{u})\mathbf{z}_n\} \\
&\qquad\qquad\cdot\{\tfrac{1}{2}\bar{G}_2(\mathbf{z}_n^{<2>}-\Psi_{20})\}']+o\left(\tfrac{1}{n\sqrt{n}}\right) \qquad (\text{by } (4.14)) \\
&\quad=\frac{i}{4n\sqrt{n}}B(\mathbf{u})E[e^{i\mathbf{t}'\bar{G}_1\mathbf{z}_n}\mathbf{z}_n(\mathbf{z}_n^{<2>}-\Psi_{20})']\bar{G}_2'+o\left(\tfrac{1}{n\sqrt{n}}\right) \qquad (\text{by } (4.5)) \\
&\quad=-\frac{e^{-\mathbf{t}'i_{\mathbf{u}}^{-1}\mathbf{t}/2}}{4n\sqrt{n}}B(\mathbf{u})\left[(2\Psi_{11}-\Psi_{11}\bar{G}_1'\mathbf{t}\mathbf{t}'\bar{G}_1\Psi_{11})\otimes(\mathbf{t}'\bar{G}_1\Psi_{11})\right]\bar{G}_2' \\
&\qquad+o\left(\tfrac{1}{n\sqrt{n}}\right) \qquad (\text{by } (4.3\text{a})).
\end{aligned}
$$

As a result, the $C(\mathbf{u},\mathbf{t})$ in Lemma 2 can be written as

$$
C(\mathbf{u},\mathbf{t})=\frac{i}{\sqrt{n}}\frac{\mathrm{d}\phi}{\mathrm{d}\mathbf{u}'}i_{\mathbf{u}}^{-1}+\frac{e^{-\mathbf{t}'i_{\mathbf{u}}^{-1}\mathbf{t}/2}}{n\sqrt{n}}B(\mathbf{u})D(\mathbf{u},\mathbf{t})+o\left(\tfrac{1}{n\sqrt{n}}\right), \tag{4.16}
$$

where

$$
D(\mathbf{u},\mathbf{t})=-\frac{1}{4}\left[(2\Psi_{11}-\Psi_{11}\bar{G}_1'\mathbf{t}\mathbf{t}'\bar{G}_1\Psi_{11})\otimes(\mathbf{t}'\bar{G}_1\Psi_{11})\right]\bar{G}_2'.
$$

By Lemma 2,

$$
\phi(\mathbf{u},\mathbf{t})=e^{-\mathbf{t}'i_{\mathbf{u}}^{-1}\mathbf{t}/2}\left(E[\sqrt{n}(\mathbf{g}^*(\bar{\mathbf{x}})-\hat{\mathbf{u}}^*(\bar{\mathbf{x}}))]+\sum_{k=1}^{q}\int_0^{t_k}\mathbf{c}_k(\mathbf{u},\mathbf{t}_k)e^{\mathbf{t}_k'i_{\mathbf{u}}^{-1}\mathbf{t}_k/2}dt_k\right). \tag{4.17}
$$

We know that $\phi=O\left(\frac{1}{n}\right)$, so that $C(\mathbf{u},\mathbf{t})=O\left(\frac{1}{n\sqrt{n}}\right)$ from (4.16), implying $\phi=O\left(\frac{1}{n\sqrt{n}}\right)$ in view of (3.3) and (4.17). Thus, since $C(\mathbf{u},\mathbf{t})=\frac{e^{-\mathbf{t}'i_{\mathbf{u}}^{-1}\mathbf{t}/2}}{n\sqrt{n}}B(\mathbf{u})D(\mathbf{u},\mathbf{t})+o\left(\frac{1}{n\sqrt{n}}\right)$, we have

$$
\begin{aligned}
\phi(\mathbf{u},\mathbf{t})=&e^{-\mathbf{t}'i_{\mathbf{u}}^{-1}\mathbf{t}/2}\left\{E[\sqrt{n}(\mathbf{g}^*(\bar{\mathbf{x}})-\hat{\mathbf{u}}^*(\bar{\mathbf{x}}))]+\frac{1}{n\sqrt{n}}B(\mathbf{u})\sum_{i=1}^{q}\int_0^{t_k}\mathbf{d}_k(\mathbf{u},\mathbf{t}_k)dt_k\right\} \\
&+o\left(\tfrac{1}{n\sqrt{n}}\right),
\end{aligned} \tag{4.18}
$$

where $D(\mathbf{u},\mathbf{t})=[\mathbf{d}_1(\mathbf{u},\mathbf{t}),\dots,\mathbf{d}_q(\mathbf{u},\mathbf{t})]$.

When $E[\sqrt{n}(\mathbf{g}^*(\bar{\mathbf{x}})-\hat{\mathbf{u}}^*(\bar{\mathbf{x}}))]=o\left(\frac{1}{n\sqrt{n}}\right)$ and $B(\mathbf{u})=0$, which is met by Cases (III) and (IV), we have from (4.18) that $\phi=o\left(\frac{1}{n\sqrt{n}}\right)$. This, along with (4.15), proves the second statement in Theorem 4. We shall prove Theorem 4-(i). First the reader should note that $\phi(\mathbf{u},\mathbf{t})$ is even in argument $\mathbf{t}$ since $D(\mathbf{u},\mathbf{t})$ is odd in $\mathbf{t}$. Thus, $(i\mathbf{t})'\phi(\mathbf{u},\mathbf{t})$ is odd in $\mathbf{t}$ and of order $O\left(\frac{1}{n\sqrt{n}}\right)$. The second term in the RHS in (4.15) vanishes if C is symmetric about the origin. The proof is complete. *Q.E.D.*

Kano [19] proved 4OE of the MLE for Case (III) as a corollary of his main result on 5OE. The proof given here is direct and much simpler.

Appendix

We shall give proofs of Lemmas 1 and 2.

There is a useful formula in evaluating expectations involving the score function, see e.g., Akahira and Takeuchi [3](Lemma 5.1.1), Amari [5](page 124) or Kano [17](Eq.(4.5)):

$$E[\mathbf{z}_n'\Theta_1 \otimes A(\mathbf{z}_n, \mathbf{u})] = \frac{1}{\sqrt{n}}\left\{\frac{\mathrm{d}}{\mathrm{d}\mathbf{u}'}E[A(\mathbf{z}_n, \mathbf{u})] - E\left[\frac{\mathrm{d}}{\mathrm{d}\mathbf{u}'}A(\mathbf{z}_n, \mathbf{u})\right]\right\}, \tag{A.1}$$

where $A(\mathbf{z}, \mathbf{u})$ is a matrix-valued function measurable in $\mathbf{z}$ in R^p and continuously differentiable in $\mathbf{u}$.

Recall that $\bar{G}_1 = i_{\mathbf{u}}^{-1}\Theta_1'$. Differentiation of ϕ w.r.t. $\frac{\mathrm{d}}{\mathrm{d}\mathbf{t}'}$ gives

$$\begin{aligned}\frac{\mathrm{d}\phi}{\mathrm{d}\mathbf{t}'} &= iE[e^{i\mathbf{t}'\sqrt{n}(T_n^{(1)}-\mathbf{u})}T_n^{(2)}\sqrt{n}(T_n^{(1)} - \mathbf{u})'] \\ &= iE[e^{i\mathbf{t}'\sqrt{n}(T_n^{(1)}-\mathbf{u})}T_n^{(2)}\{\sqrt{n}(T_n^{(1)} - \mathbf{u}) - \bar{G}_1\mathbf{z}_n\}'] \\ &\qquad + iE[(\mathbf{z}_n'\Theta_1) \otimes (e^{i\mathbf{t}'\sqrt{n}(T_n^{(1)}-\mathbf{u})}T_n^{(2)})]i_{\mathbf{u}}^{-1}. \end{aligned} \tag{A.2}$$

Exchangeability of the differentiation and expectation is permitted for exponential families (see e.g., Lehmann [20] page 59). Application of the formula (A.1) to the second term in (A.2) leads to

$$\begin{aligned}&\frac{i}{\sqrt{n}}\left(\frac{\mathrm{d}}{\mathrm{d}\mathbf{u}'}E[e^{i\mathbf{t}'\sqrt{n}(T_n^{(1)}-\mathbf{u})}T_n^{(2)}] - E[\frac{\mathrm{d}}{\mathrm{d}\mathbf{u}'}e^{i\mathbf{t}'\sqrt{n}(T_n^{(1)}-\mathbf{u})}T_n^{(2)}]\right)i_{\mathbf{u}}^{-1} \\ &\qquad = \frac{i}{\sqrt{n}}\frac{\mathrm{d}\phi}{\mathrm{d}\mathbf{u}'}i_{\mathbf{u}}^{-1} - (\mathbf{t}'i_{\mathbf{u}}^{-1} \otimes \phi).\end{aligned}$$

Substitution of this into (A.2) gives (4.11) *Q.E.D.*

Let us prove Lemma 2. It suffices to show that the following partial differential equations (A.3) has the solution (A.6):

$$\frac{\mathrm{d}\phi(\mathbf{t})}{\mathrm{d}\mathbf{t}'} = -(\mathbf{t}'i_{\mathbf{u}}^{-1} \otimes \phi(\mathbf{t})) + C(\mathbf{t}). \tag{A.3}$$

Put $\phi(\mathbf{t}) = \widetilde{\phi}(\mathbf{t}) \cdot e^{-\mathbf{t}'i_{\mathbf{u}}^{-1}\mathbf{t}/2}$, and then we have

$$\frac{\mathrm{d}\widetilde{\phi}(\mathbf{t})}{\mathrm{d}\mathbf{t}'} = e^{\mathbf{t}'i_{\mathbf{u}}^{-1}\mathbf{t}/2}C(\mathbf{t}) \left(= e^{\mathbf{t}'i_{\mathbf{u}}^{-1}\mathbf{t}/2}[\mathbf{c}_1(\mathbf{t}), \dots, \mathbf{c}_q(\mathbf{t})],\ \text{say}\right).$$

That is,

$$\frac{\partial\widetilde{\phi}(\mathbf{t})}{\partial t_k} = e^{\mathbf{t}'i_{\mathbf{u}}^{-1}\mathbf{t}/2}\mathbf{c}_k(\mathbf{t}) \quad (k = 1, \dots, q). \tag{A.4}$$

Using (A.4), we have

$$\begin{aligned}\widetilde{\phi}(\mathbf{t}) &= \widetilde{\phi}(0) + \sum_{k=1}^{q}\left\{\widetilde{\phi}(\mathbf{t}_k) - \widetilde{\phi}(\mathbf{t}_{k-1})\right\} \\ &= \widetilde{\phi}(0) + \sum_{k=1}^{q}\int_0^{t_k}\frac{\partial\widetilde{\phi}(\mathbf{t}_k)}{\partial t_k}dt_k \\ &= \widetilde{\phi}(0) + \sum_{k=1}^{q}\int_0^{t_k}\mathbf{c}_k(\mathbf{t}_k)e^{\mathbf{t}_k' i_{\mathbf{u}}^{-1}\mathbf{t}_k/2}dt_k. \qquad \text{(A.5)}\end{aligned}$$

Conversely, the function (A.5) is actually a solution to (A.4), provided that (A.4) has a solution. Thus, we obtain

$$\phi(\mathbf{t}) = e^{-\mathbf{t}' i_{\mathbf{u}}^{-1}\mathbf{t}/2}\left(\phi(0) + \sum_{k=1}^{q}\int_0^{t_k}\mathbf{c}_k(\mathbf{t}_k)e^{\mathbf{t}_k' i_{\mathbf{u}}^{-1}\mathbf{t}_k/2}dt_k\right). \qquad \text{(A.6)}$$

ACKNOWLEDGMENT

The author would like to express his thanks to Professor M. Akahira of University of Tsukuba and Dr. M. Taniguchi of Osaka University for invaluable discussion on higher-order efficiency. He also thanks Professor S. Amari of The University of Tokyo for showing his bias-correction factor (Case (IV) in Theorem 4).

REFERENCES

[1] Akahira, M. (1975). A note on the second order asymptotic efficiency of estimators in an autoregressive process. *Rep. Univ. Electro-Comm.*, **26**, 143-149.

[2] Akahira, M. and Takeuchi, K. (1976). On the second order asymptotic efficiency of estimators in multiparameter cases. *Rep. Univ. Electro-Comm.*, **26**, 261-269.

[3] Akahira, M. and Takeuchi, K. (1981). *Asymptotic Efficiency of Statistical Estimators: Concepts and Higher Order Asymptotic Efficiency,* Lecture Notes in Statistics, Vol.7. Springer-Verlag, New York.

[4] Amari, S. (1982). Differential geometry of curved exponential families: Curvatures and information loss. *Ann. Statist.*, **10**, 357-387.

[5] Amari, S. (1985). *Differential-Geometrical Methods in Statistics,* Lecture Notes in Statistics, Vol.28. Springer-Verlag, Berlin.

[6] Amari, S. (1995). Personal communications.

[7] Barndorff-Nielsen, O. E. and Cox, D. R. (1982). *Asymptotic Techniques for Use in Statistics.* Chapman and Hall, London.

[8] Bentler, P. M. and Lee, Sik-Yum (1978). Matrix derivatives with chain rule and rules for simple, Hadamard, and Kronecker products. *J. Math. Psychol.*, **17**, 255-262.

[9] Bhattacharya, R. and Denker, M. (1990). *Asymptotic Statistics.* Birkhäuser, Basel.

[10] Efron, B. (1975). Defining the curvature of a statistical problem (with applications to second order efficiency). *Ann. Statist.*, **3**, 1189-1242.

[11] Eguchi, S. (1983). Second order efficiency of minimum contrast estimators in a curved exponential family. *Ann. Statist.*, **11**, 793-803.

[12] Ghosh, J. K. (1994). *Higher Order Asymptotics.* NSF-CBMS Regional Conference Series in Probability and Statistics, Vol.4. IMS, Hayward, CA.

[13] Ghosh, J. K. and Subramanyam, K. (1974). Second order efficiency of maximum likelihood estimates. *Sankhyā, Ser. A*, **36**, 325-358.

[14] Henderson H. V. and Searle, S. R. (1981). The vec-permutation matrix, the vec operator and Kronecker products: A review. *Linear and Multilinear Algebra*, **9**, 271-288.

[15] Hosoya, Y. (1979). Higher-order efficiency in the estimation of linear processes. *Ann. Statist.*, **7**, 516-530.

[16] Kano, Y. (1986). Conditions on consistency of estimators in covariance structure model. *J. Japan Statist. Soc.*, **16**, 75-80.

[17] Kano, Y. (1993). Beyond second-order efficiency. (submitted)

[18] Kano, Y. (1993). Beyond second-order efficiency: Supplement. *Technical Report*, University of Osaka Prefecture, Osaka, Japan.

[19] Kano, Y. (1994). *More* higher-order efficiency: Concentration probability. *Technical Reports of Statistical Group:* DATA95-04, Tokyo, Japan.

[20] Lehmann, E. L. (1986). *Testing Statistical Hypotheses* (2nd Ed.). John Wiley, New York.

[21] Magnus, J. R. and Neudecker, H. (1988). *Matrix Differential Calculus with Applications in Statistics and Econometrics.* John Wiley, Chichester.

[22] Pfanzagl, J. (1973). Asymptotic expansions related to minimum contrast estimators. *Ann. Statist.*, **1**, 993-1026.

[23] Pfanzagl, J. (1976). First order efficiency implies second order efficiency. In *Contribution to Statistics (Hajek Memorial Volume)*, (J. Jureckova, Ed.), pp. 167-196,

Academia, Prague.

[24] Pfanzagl, J. (1985). *Asymptotic Expansions for General Statistical Models.* Lecture Notes in Statistics, Vol.31. Springer-Verlag, Berlin.

[25] Pfanzagl, J. and Wefelmeyer, W. (1978). A third order optimum property of the maximum likelihood estimator. *J. Multivariate Anal.*, **8**, 1-29.

[26] Rao, C. R. (1961). Asymptotic efficiency and limiting information. *Proc. Fourth Berkeley Symp.*, **1**, 531-545.

[27] Rao, C. R. (1962). Efficient estimates and optimum inference procedures in large samples. *J. Royal Statist. Soc., Ser. B*, **24**, 46-72.

[28] Rao, C. R. (1963). Criteria of estimation in large samples. *Sankhyā, Ser. A*, **25**, 189-206.

[29] Rao, C. R., Sinha, B. K. and Subramanyam, K. (1982). Third order efficiency of the maximum likelihood estimator in the multinomial distribution. *Statist. Decisions*, **1**, 1-16.

[30] Satake, I. (1974). *Senkei Daisu Gaku (Linear Algebra)* (2nd Ed.). Shokabo, Tokyo (in Japanese).

[31] Sternberg, S. (1964). *Lectures on Differential Geometry.* Prentice-Hall Inc., Englewood Cliffs, NJ.

[32] Takemura, A. (1991). *Tahenryo Suisoku Tokei no Kiso (The foundation of Multivariate Statistical Inference).* Kyoritsu Shuppan, Tokyo (in Japanese).

[33] Takeuchi, K. (1974). *Tokeiteki Suisoku no Zenkin Riron (Asymptotic Theory of Statistical Estimation).* Kyoiku Shuppan, Tokyo (in Japanese).

[34] Taniguchi, M. (1983). On the second order asymptotic efficiency of estimators of Gaussian ARMA Processes. *Ann. Statist.*, **11**, 157-169.

[35] Taniguchi, M. (1991). *Higher Order Asymptotic Theory for Time Series Analysis*, Lecture Notes in Statistics, Vol.68. Springer-Verlag, Berlin.

[36] Yoshida, N. (1992). Asymptotic expansions of maximum likelihood estimators for small diffusions via the theory of Malliavin–Watanabe. *Probab. Theory Relat. Fields*, **92**, 275-311.

[37] Yoshida, N. (1993). Asymptotic expansion of Bayes estimators for small diffusions. *Probab. Theory Relat. Fields*, **95**, 429-450.