More higher-order efficiency: Concentration probability

Yutaka KANO

Osaka University

(Final version, May 1998)

Running Title: MORE HIGHER-ORDER EFFICIENCY.

Mailing Address: Faculty of Human Sciences, Osaka University, Suita, Osaka 565-0871, JAPAN. Email: kano@hus.osaka-u.ac.jp

Abstract

Based on concentration probability of estimators about a true parameter, third-order asymptotic efficiency of the first-order bias-adjusted MLE within the class of first-order bias-adjusted estimators has been well established in a variety of probability models. In this paper we consider the class of second-order bias-adjusted Fisher consistent estimators of a structural parameter vector on the basis of an i.i.d. sample drawn from a curved exponential-type distribution, and study the asymptotic concentration probability, about a true parameter vector, of these estimators up to the fifth-order. In particular, (i) we show that third-order efficient estimators are always fourth-order efficient; (ii) a necessary and sufficient condition for fifth-order efficiency is provided; and finally (iii) the MLE is shown to be fifth-order efficient.

Key Words and Phrases: Bias-adjustment, curved exponential distributions, Edgeworth expansion, maximum likelihood estimator, Fisher-consistency.

AMS 1991 subject classifications: 62F12

1 Introduction

In the recently published book by J. K. Ghosh (1994), he mentions, in Section 6.10 located after description of third-order (asymptotic) efficiency, that *Where do we go from here: Given that first-order efficiency implies second-order efficiency, it is natural to conjecture that third-order efficiency implies fourth-order efficiency. The proof of that must be very messy. One may also ask if anything like third-order efficiency holds for* $\widehat{\theta}$ *when we go to the fifth-order. Ghosh and Sinha (1982) show with a counterexample that this is not possible.*

With fairly simple proofs, this paper gives an affirmative answer to the conjecture above and establishes the fifth-order efficiency of the MLE, under a slightly different bias-adjustment.

Let $(\mathcal{X}, \mathcal{A}, \mathcal{P}_\theta)$ be a complete and separable probability space for each $\theta \in \Omega_\theta$, where Ω_θ, an open domain of ^{p}, is a parameter space of θ. Let $\mathbf{x}$ be a p-dimensional random vector ($\mathcal{X} \to {}^p$, $\mathcal{A}$-measurable) with a continuous exponential-type distribution:

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

where $\mu(\cdot)$ is a carrier measure on ^{p} (see, e.g., Amari, 1985, section 4.1) and the prime $'$ denotes the transpose of a vector or a matrix. The real-valued function, $\psi(\theta)$, on Ω_θ is defined as

$$\psi(\theta) = \log \int_p \exp(\theta' x)\mu(dx).$$

Assume that the parameter p-vector θ is a function of a more basic parameter q-vector $\mathbf{u}$ ($p > q$), i.e.,

$$\theta = \theta(\mathbf{u}) \tag{1.2}$$

is a function from Ω, an open domain of ^{q}, to $\Omega_\theta (\subset {}^p)$. The family of the distributions (1.1) with the structure $\theta = \theta(\mathbf{u})$ is said to be of curved exponential-type.

Let $\mathbf{x}_1, \ldots, \mathbf{x}_n$ be identically and independently distributed (i.i.d.) random p-vectors defined on the probability space $(\mathcal{X}^n, \mathcal{A}^n, \mathcal{P}_\theta^n)$, each of which is distributed according to (1.1) with the structure (1.2). In this paper, we shall study asymptotic properties of estimators for the unknown parameter vector $\mathbf{u}$ on the basis of the random sample $\mathbf{x}_1, \ldots, \mathbf{x}_n$.

The likelihood function $L(\mathbf{u})$, defined on Ω, is written as

$$L(\mathbf{u}) = \prod_{j=1}^{n} \exp(\theta(\mathbf{u})'\mathbf{x}_j - \psi(\theta(\mathbf{u}))) = \exp\{n(\theta(\mathbf{u})'\bar{\mathbf{x}} - \psi(\theta(\mathbf{u})))\},$$

where $\bar{\mathbf{x}} = \frac{1}{n}\sum_{j=1}^{n}\mathbf{x}_j$. Note that $\bar{\mathbf{x}}$ is a sufficient statistic for θ (and hence for $\mathbf{u}$). The maximum likelihood estimator (MLE), $\hat{\mathbf{u}} = \hat{\mathbf{u}}(\bar{\mathbf{x}})$, is defined as a solution to the maximization problem:

$$\max_{\mathbf{u}\,\in\,\Omega} L(\mathbf{u}). \tag{1.3}$$

Some assumptions on the MLE will be stated later.

Let $\mathbf{g}(\cdot)$ be a function from ^{p} to ^{q}, smooth in a neighborhood of an arbitrarily fixed $\eta(\mathbf{u})(= \mathrm{E}[\mathbf{x}])$ (see (3.3)). The function $\mathbf{g}(\bar{\mathbf{x}})$ is said to be a Fisher consistent estimator if and only if $\mathbf{g}(\eta(\mathbf{u})) = \mathbf{u}$ for all possible $\mathbf{u} \in \Omega$. The property excludes Hodges' pathological example (see Le Cam, 1953).

Let $\mathbf{g}_0(\bar{\mathbf{x}})$ and $\mathbf{g}(\bar{\mathbf{x}})$ be Fisher consistent estimators of $\mathbf{u}$, and let $\mathbf{g}_0^*(\bar{\mathbf{x}})$ and $\mathbf{g}^*(\bar{\mathbf{x}})$ be corresponding bias-adjusted estimators which have the same asymptotic bias up to a suitable order. Consider the following statement: For a positive integer k,

$$\mathrm{P}_{\theta(\mathbf{u})}\{\sqrt{n}(\mathbf{g}^*(\bar{\mathbf{x}}) - \mathbf{u}) \ \in \ C\} \le \mathrm{P}_{\theta(\mathbf{u})}\{\sqrt{n}(\mathbf{g}_0^*(\bar{\mathbf{x}}) - \mathbf{u}) \ \in \ C\} + o(n^{-(k-1)/2}), \tag{1.4}$$

where $\mathrm{P}_{\theta(\mathbf{u})}$ denotes the probability w.r.t. (1.1) with (1.2) and C is any Borel convex set of ^{q} symmetric about the origin. The statement (1.4) asserts that the estimator $\mathbf{g}_0^*(\bar{\mathbf{x}})$ is more concentrated about the true parameter $\mathbf{u}$ than $\mathbf{g}^*(\bar{\mathbf{x}})$ up to the order $o(n^{-(k-1)/2})$. If $\mathbf{g}_0^*(\bar{\mathbf{x}})$ satisfies the relation (1.4) for any Fisher consistent estimator $\mathbf{g}^*(\bar{\mathbf{x}})$ at $k = 1$, the estimator $\mathbf{g}_0(\bar{\mathbf{x}})$ is said to be first-order (asymptotically) efficient. A first-order efficient estimator $\mathbf{g}_0(\bar{\mathbf{x}})$ is said to be second-order efficient iff (1.4) holds true at $k = 2$ for any first-order efficient estimator $\mathbf{g}(\bar{\mathbf{x}})$; in general, $(K-1)$th-order efficient estimator $\mathbf{g}_0(\bar{\mathbf{x}})$ is said to be Kth-order efficient iff (1.4) holds true at $k = K$ for any $(K-1)$th-order efficient estimator $\mathbf{g}(\bar{\mathbf{x}})$ (see, e.g., Akahira and Takeuchi, 1981, page 8; Amari, 1985, page 130).

The purpose of this paper is to give an almost complete answer to some natural questions on the fourth- and fifth-order efficiency of the MLE. We show that under a certain second-order bias-adjustment, the fourth-order term of the concentration probability, which is of order $O\left(\frac{1}{n\sqrt{n}}\right)$, of third-order efficient estimators does not depend on the estimators, that is, third-order efficiency automatically implies fourth-order efficiency. The consequence corresponds to the impressive result, proved by Pfanzagl (1976), that first-order efficiency implies second-order efficiency. Next, a necessary and sufficient condition for fifth-order efficiency is given, and we apply the condition to establish the fifth-order efficiency of the MLE.

Technically, the following two things are new: (i) It is seen that an alternative bias-adjustment factor must be employed, otherwise Ghosh and Sinha's (1982) counterexample to the fifth-order efficiency of the MLE is alive; (ii) partial differential equations (PDE) are derived to evaluate characteristic functions. The method of evaluation through the PDE is new and successfully reduces the computation, and even though the method is applied to prove the third-order efficiency, it is new. Rather than emphasizing the novelty of these techniques, we hope that one will develop these technical tools to establish so-called *absolute asymptotic efficiency*, i.e., the MLE is asymptotically efficient at any order, a conjecture by Rao, Sinha and Subramanyam (1982, Remark 3.1).

An overview of the paper is as follows: Section 2 briefly reviews the asymptotic efficiency, and Section 3 gives notation used throughout the paper. Main results are stated in Section 4. An outline of the proofs of the main results is described in Section 5; some technical details are presented in Appendix.

2 Brief historical review

The notion of (asymptotic) efficiency in estimation problem was first discussed by Fisher (1925). The first-order efficiency of the MLE in the sense of (1.4) was proved by Rao(1962), Bahadur (1964) and Kaufman (1966) for the case where a general density function w.r.t. the Lebesgue measure is permitted. Several refinements of the first-order efficiency have been made by Hajek (1970), Inagaki (1970), and Pfanzagl (1973a) among others.

The second-order efficiency of the MLE was proved by Pfanzagl (1973b); and then, as noted in Introduction, he established that first-order efficiency implies second-order efficiency under bias-adjustments (see Pfanzagl, 1976). Extensions to the multi-parameter case or to not i.i.d. cases were made by Akahira (1975), Akahira and Takeuchi (1976), Hosoya (1979) and Taniguchi (1983) among others. Yoshida (1992, 93) recently made an attempt to establish second-order efficiency in diffusion processes.

The third-order efficiency of the MLE was shown by Takeuchi and Akahira (1978), Pfanzagl and Wefelmeyer (1978) and Ghosh, Sinha and Wieand (1980). Extensions to more complicated models were made by Takeuchi and Morimune (1985) and Taniguchi (1986,87). Thorough theory of asymptotic efficiency up to the third-order in estimation problems is described by the well-written text books by Akahira and Takeuchi (1981), Pfanzagl (1985), Taniguchi (1991), and Ghosh (1994). In the discussion of second- or

third-order optimality in estimation, the main ideas are bias-adjustments and Edgeworth expansions.

There are more tractable but somewhat indirect criteria of asymptotic efficiency such as the quadratic loss and loss of information. Plenty of research on second- or third-order efficiency based on these criteria have been done by many authors including Rao (1961,63), Ghosh and Subramanyam (1974), Efron (1975), Eguchi (1983), and Amari (1985). Efron (1975) and Hosoya (1990) have pointed out a certain equivalence between these criteria.

There is little work on studying properties of n^{-3} terms of loss functions, see Efron (1975), Rao, Sinha and Subramanyam (1982), and Ghosh and Sinha (1982). Note that the evaluation of n^{-3} terms of loss functions corresponds to that of n^{-2} terms of concentration probability. Kano (1997) recently proved that the MLE minimizes the n^{-3} terms of an expansion of the quadratic loss of Fisher consistent estimators. No work on over third-order efficiency on the basis of concentration probability has been yet made; the present work is the first attempt in this direction.

3 Notation

Let $\mathbf{x} = [x_1, \ldots, x_p]'$, and let $A(\mathbf{x}) = [a_{ij}(\mathbf{x})]$ 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 $\frac{\mathrm{d}A(\mathbf{x})}{\mathrm{d}\mathbf{x}'} = \frac{\mathrm{d}}{\mathrm{d}\mathbf{x}'}A(\mathbf{x}) = \left[\frac{\partial A(\mathbf{x})}{\partial x_1}, \cdots, \frac{\partial A(\mathbf{x})}{\partial x_p}\right]$, 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 e.g., Bentler and Lee (1978), Magnus and Neudecker (1985,88), and Kano (1993,97). 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. Define $A^{<0>} = 1$. 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. Sternberg (1964, Section 1.3) for the symmetric tensor. Let $\mathbf{g}(\mathbf{x})$ be an ^{q}-valued smooth function defined on a neighborhood of $\mathbf{x}_0 \in {}^p$. The Taylor series of $\mathbf{g}(\mathbf{x})$ about $\mathbf{x}_0$ is expressible in the form:

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

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

$$G_k N_{p^{<k>}} = G_k. \tag{3.1}$$

Recall that $\mathbf{x}_1, \ldots, \mathbf{x}_n$ are an i.i.d. random sample distributed according to the distribution (1.1) with (1.2). Define

$$\Psi_{k\ell} = \Psi_{k\ell}(\theta(\mathbf{u})) = \left(\frac{\mathrm{d}}{\mathrm{d}\theta}\right)^{<k>} \left(\frac{\mathrm{d}}{\mathrm{d}\theta'}\right)^{<\ell>} \psi(\theta)\Big|_{\theta=\theta(\mathbf{u})}. \tag{3.2}$$

Since the cumulant generating function of the distribution is written as $\log \mathrm{E}_{\theta(\mathbf{u})}[e^{i\mathbf{s}'\mathbf{x}_j}] = \psi(\theta(\mathbf{u}) + i\mathbf{s}) - \psi(\theta(\mathbf{u}))$, we have

$$\mathrm{E}_{\theta(\mathbf{u})}[\mathbf{x}_j] = \Psi_{10}(\theta(\mathbf{u}) \,(= \eta(\mathbf{u}), \quad \text{say}), \qquad j = 1, \ldots, n \tag{3.3}$$

where $\mathrm{E}_{\theta(\mathbf{u})}[\cdot]$ denotes the expectation w.r.t. the probability measure in (1.1) with the structure (1.2). Note that

$$\Psi_{k\ell} = \mathrm{E}_{\theta(\mathbf{u})}[(\mathbf{x}_j - \eta(\mathbf{u}))^{<k>}(\mathbf{x}_j - \eta(\mathbf{u}))^{<\ell>\prime}]$$

for $k + \ell = 2, 3$. Put $\mathbf{z}_n = \sqrt{n}(\bar{\mathbf{x}} - \eta(\mathbf{u}))$ and $\Psi_{k\ell}(n) = \mathbf{z}_n^{<k>}\mathbf{z}_n^{<\ell>\prime}$. Then

$$\mathrm{E}_{\theta(\mathbf{u})}[\Psi_{k\ell}(n)] = \Psi_{k\ell} \quad \text{for } k + \ell = 2, \qquad \mathrm{E}_{\theta(\mathbf{u})}[\Psi_{k\ell}(n)] = \frac{1}{\sqrt{n}}\Psi_{k\ell} \quad \text{for } k + \ell = 3;$$
$$\mathrm{E}_{\theta(\mathbf{u})}[\Psi_{40}(n)] = 3N_{p^{<4>}}\Psi_{20}^{<2>} + O\left(\tfrac{1}{n}\right).$$

See Holmquist (1988) and Kano (1993, 97) for more general results.

4 Main results

First of all, we state some assumptions. Let $\mathbf{u}_0$ be an arbitrarily fixed (inner) point of Ω, and let $U_0 (\subset \Omega)$ be a neighborhood of $\mathbf{u}_0$. We denote by $_k$ the k-dimensional Borel σ-field.

(A1) $\theta(\mathbf{u})$ is $(_q \cap \Omega, _p \cap \Omega_\theta)$-measurable and smooth in $\mathbf{u} \in U_0$.

Denote $\Theta_1 = \Theta_1(\mathbf{u}) = \frac{\mathrm{d}}{\mathrm{d}\mathbf{u}'}\theta(\mathbf{u})$ for $\mathbf{u} \in U_0$. Recall the definitions of $\Psi_{k\ell}(\theta)$ in (3.2) and $\eta(\mathbf{u})$ in (3.3).

(A2) $\mathrm{rank}(\Theta_1(\mathbf{u})) = q$ and $\mathrm{rank}(\Psi_{11}(\theta(\mathbf{u}))) = p$ at $\mathbf{u} = \mathbf{u}_0$.

Remember that $\mathbf{g}(\mathbf{x})$ is an ^{q}-valued function from ^{p}.

(A3) $\mathbf{g}(\mathbf{x})$ is $({}_p, {}_q)$-measurable and smooth in a neighborhood of $\eta(\mathbf{u}_0)$, and $\mathbf{g}(\bar{\mathbf{x}})$ is Fisher consistent, i.e., $\mathbf{g}(\eta(\mathbf{u})) = \mathbf{u}$ for any $\mathbf{u} \in U_0$.

Assumption (A3) is also said to be *locally stable* (see Pfanzagl and Wefelmeyer, 1978, page 7).

(A4) The optimization problem (1.3) has a unique solution, $\hat{\mathbf{u}}(\bar{\mathbf{x}})$, say, for any $\bar{\mathbf{x}}$ in a neighborhood of $\eta(\mathbf{u})$. The MLE $\hat{\mathbf{u}}(\bar{\mathbf{x}})$ is $({}_p, {}_q)$-measurable and continuous at $\bar{\mathbf{x}} = \eta(\mathbf{u})$ $(\mathbf{u} \in U_0)$.

The continuity assumption ensures (strong) consistency of $\hat{\mathbf{u}}(\bar{\mathbf{x}})$. Assumptions on Ω and $\mathbf{u}_0$ are needed to guarantee the continuity; we do not discuss it here. See Rao (1973) and Kano (1986) for details.

It follows from (A4) that for any $\bar{\mathbf{x}}$ in a neighborhood of $\eta(\mathbf{u}_0)$, $\hat{\mathbf{u}}(\bar{\mathbf{x}})$ is a solution to the equation:

$$\frac{\mathrm{d}}{\mathrm{d}\mathbf{u}} L(\mathbf{u}|\bar{\mathbf{x}}) = n\Theta_1(\mathbf{u})'\{\bar{\mathbf{x}} - \eta(\mathbf{u})\} = 0. \tag{4.1}$$

The function $\frac{\mathrm{d}}{\mathrm{d}\mathbf{u}} L(\mathbf{u}|\bar{\mathbf{x}})$ is smooth in $(\bar{\mathbf{x}}, \mathbf{u})$ in a neighborhood of $(\eta(\mathbf{u}_0), \mathbf{u}_0)$, and

$$\frac{\mathrm{d}}{\mathrm{d}\mathbf{u}'} \frac{\mathrm{d}}{\mathrm{d}\mathbf{u}} L(\mathbf{u}|\bar{\mathbf{x}})\bigg|_{(\bar{\mathbf{x}},u)=(\eta(\mathbf{u}_0),\mathbf{u}_0)} = n\Theta_1(\mathbf{u}_0)\Psi_{11}(\theta(\mathbf{u}_0))\Theta_1(\mathbf{u}_0) \ \ (= ni_{\mathbf{u}_0}, \quad \text{say}) \tag{4.2}$$

is positive definite in view of (A2). The continuity in (A4) ensures that $(\bar{\mathbf{x}}, \hat{\mathbf{u}}(\bar{\mathbf{x}}))$ stays in a neighborhood of $(\eta(\mathbf{u}_0), \mathbf{u}_0)$ for any $\bar{\mathbf{x}}$ close to $\eta(\mathbf{u}_0)$. Thus, application of the implicit function theorem to (4.1) shows that $\hat{\mathbf{u}}(\bar{\mathbf{x}})$ is smooth in a neighborhood of $\eta(\mathbf{u}_0)$. Note that $i_{\mathbf{u}_0}$ in (4.2) is the Fisher information matrix.

Bias adjustments play an important role in studying higher-order asymptotic optimality of estimators. Recall $\mathbf{z}_n = \sqrt{n}(\bar{\mathbf{x}} - \eta(\mathbf{u}))$ and write

$$G_k = G_k(\eta(\mathbf{u})) = \left(\frac{\mathrm{d}}{\mathrm{d}\bar{\mathbf{x}}'}\right)^{<k>} \mathbf{g}(\bar{\mathbf{x}})\bigg|_{\bar{\mathbf{x}}=\eta(\mathbf{u})}.$$

The function $G_k(\mathbf{x})$ is defined merely on a neighborhood of $\eta(\mathbf{u}_0)$, but we extend $G_k(\mathbf{x})$ to a function defined on ^{p} which is measurable in $\mathbf{x}$ and smooth in a neighborhood of $\eta(\mathbf{u}_0)$.

The estimator $\mathbf{g}(\bar{\mathbf{x}})$ is then expanded formally as

$$\mathbf{g}(\bar{\mathbf{x}}) = \mathbf{u} + \frac{1}{\sqrt{n}} G_1 \mathbf{z}_n + \frac{1}{2n} G_2 \mathbf{z}_n^{<2>} + \frac{1}{6n\sqrt{n}} G_3 \mathbf{z}_n^{<3>} + \cdots ,$$

and hence the asymptotic bias is expressible in the form:

$$\mathrm{E}_{\theta(\mathbf{u})}[\mathbf{g}(\bar{\mathbf{x}}) - \mathbf{u}] = \frac{1}{2n} G_2 \Psi_{20} + \frac{1}{6n^2} G_3 \Psi_{30} + \frac{1}{8n^2} G_4 \Psi_{20}^{<2>} + O\left(\tfrac{1}{n^3}\right).$$

We shall make a first-order bias adjustment (or correction) via

$$\frac{1}{n} b_1(\bar{\mathbf{x}}) = \frac{1}{2n} G_2(\bar{\mathbf{x}}) \Psi_{20}(\theta(g(\bar{\mathbf{x}}))). \tag{4.3}$$

It should be noted that the adjustment factor is slightly but *essentially* different from the one employed in the literature: $\frac{1}{2n} G_2 \hat{\Psi}_{20} = \frac{1}{2n} G_2(\eta(g(\bar{\mathbf{x}}))) \Psi_{20}(\theta(g(\bar{\mathbf{x}})))$. Either of the factors can be employed to establish the second- and third-order optimality of the MLE, in other words, the optimality does not depend on the distribution of $b_1(\bar{\mathbf{x}})$ but only on its convergent point. In contrast, the distribution does influence upon the terms which contribute to more higher-order optimality. Thus, the choice of the bias-adjustment factor is essential.

The first-order bias-adjusted estimator has the following second-order asymptotic bias:

$$\begin{aligned}
\mathrm{E}_{\theta(\mathbf{u})}\left[\mathbf{g}(\bar{\mathbf{x}}) - \mathbf{u} - \frac{1}{n}\mathbf{b}_1(\bar{\mathbf{x}})\right] =& -\frac{1}{2n^2}\left\{ \left(\frac{\mathrm{d}}{\mathrm{d}\bar{\mathbf{x}}'}\right)^{<2>} \mathbf{b}_1(\bar{\mathbf{x}}) \Big|_{\bar{\mathbf{x}}=\eta(\mathbf{u})} \right\} \Psi_{20} \\
& + \frac{1}{6n^2} G_3 \Psi_{30} + \frac{1}{8n^2} G_4 \Psi_{20}^{<2>} + O\left(\tfrac{1}{n^3}\right) \\
=& \frac{1}{n^2}\mathbf{b}_2(\mathbf{u}) + O\left(\tfrac{1}{n^3}\right), \qquad \text{say.}
\end{aligned}$$

We make a second-order bias-adjustment by $\mathbf{b}_2(\mathbf{g}(\bar{\mathbf{x}}))$ $(= \mathbf{b}_2(\bar{\mathbf{x}})$, simply). The estimator after the second-order bias-adjustment is denoted by

$$\mathbf{g}^*(\bar{\mathbf{x}}) = \mathbf{g}(\bar{\mathbf{x}}) - \frac{1}{n}\mathbf{b}_1(\bar{\mathbf{x}}) - \frac{1}{n^2}\mathbf{b}_2(\bar{\mathbf{x}}). \tag{4.4a}$$

In the same manner, the MLE $\hat{\mathbf{u}}(\bar{\mathbf{x}})$ is bias-adjusted up to the second-order: Let

$$\bar{G}_k = \bar{G}_k(\eta(\mathbf{u})) = \left(\frac{\mathrm{d}}{\mathrm{d}\bar{\mathbf{x}}'}\right)^{<k>} \hat{\mathbf{u}}(\bar{\mathbf{x}}) \Big|_{\bar{\mathbf{x}}=\eta(\mathbf{u})} .$$

Then,

$$\hat{\mathbf{u}}^*(\bar{\mathbf{x}}) = \hat{\mathbf{u}}(\bar{\mathbf{x}}) - \frac{1}{n}\bar{\mathbf{b}}_1(\bar{\mathbf{x}}) - \frac{1}{n^2}\bar{\mathbf{b}}_2(\bar{\mathbf{x}}), \tag{4.4b}$$

where $\bar{\mathbf{b}}_1(\bar{\mathbf{x}})$ and $\bar{\mathbf{b}}_2(\bar{\mathbf{x}})$ are are defined as $\mathbf{b}_1(\bar{\mathbf{x}})$ and $\mathbf{b}_2(\bar{\mathbf{x}})$ with $\bar{G}_k$ for G_k. As a result, it holds formally

$$\mathrm{E}_{\theta(\mathbf{u})}[\mathbf{g}^*(\bar{\mathbf{x}}) - \mathbf{u}] = O\left(\tfrac{1}{n^3}\right), \quad \mathrm{E}_{\theta(\mathbf{u})}[\hat{\mathbf{u}}^*(\bar{\mathbf{x}}) - \mathbf{u}] = O\left(\tfrac{1}{n^3}\right). \tag{4.5}$$

It is known that $\mathbf{g}(\bar{\mathbf{x}})$ is first-order efficient iff $G_1 = \bar{G}_1$ $(= i_{\mathbf{u}}^{-1}\Theta_1')$, see e.g., Takeuchi and Akahira (1978), formula (2.3); under the first-order bias adjustment, the first-order efficient estimator is always second-order efficient (Pfanzagl, 1976), and the estimator is third-order efficient iff $G_2 = \bar{G}_2$. Although there would not be suitable references which explicitly state the last statement, it could be easily deduced from the literature (see e.g. Akahira and Takeuchi, 1981, pages 155-158). As a result, the bias-adjusted MLE is first-, second- and third-order efficient within the class of bias-adjusted Fisher consistent estimators.

The next theorem is the main result of this paper.

Theorem 1. Suppose that $\mathbf{x}_1 \ldots, \mathbf{x}_n$ are i.i.d. ^{p}-valued random vectors defined on $(\mathcal{X}^n, \mathcal{A}^n, \mathcal{P}_\theta^n)$, where $\mathcal{P}_\theta$ is of continuous curved exponential-type (1.1) with the structure $\theta = \theta(\mathbf{u})$ $(\mathbf{u} \in U_0)$. Write $\bar{\mathbf{x}} = \frac{1}{n}\sum_{j=1}^n \mathbf{x}_j$. Assume (A1)–(A4). Let the MLE $\hat{\mathbf{u}}(\bar{\mathbf{x}})$ be defined by (1.3) and let $\mathbf{g}(\bar{\mathbf{x}})$ be third-order efficient; let $\hat{\mathbf{u}}^*(\bar{\mathbf{x}})$ and $\mathbf{g}^*(\bar{\mathbf{x}})$ be second-order bias-adjusted in (4.4). Let $C \in {}_q$ be an arbitrary convex set symmetric about the origin. Then there exists a nonnegative quantity $\Delta^2_{\hat{u},g,C}$ such that for each $\mathbf{u} \in U_0$,

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

uniformly in C and that

(ii) $\Delta^2_{\hat{u},g,C} = 0$ if and only if $G_3 = \bar{G}_3$, provided that C is of positive Lebesgue measure.

The result (i) implies that the term of order $n^{-3/2}$ of the LHS coincides with that of the RHS, that is, under the second-order bias-adjustment, the terms of order $n^{-3/2}$ of expansions of the concentration probability of third-order efficient estimators are independent of the estimators; the terms of all such estimators are the same as the MLE's. As a result, we have the following.

Corollary 1. Under the second-order bias-adjustment in (4.4), third-order efficiency implies fourth-order efficiency.

It is seen that one can not distinguish third-order efficient estimators via Edgeworth expansions up to the order $n^{-3/2}$. The statement (ii) of Theorem 1 gives a necessary and sufficient condition for fifth-order efficiency. We apply the condition to get

Corollary 2. Under the second-order bias-adjustment in (4.4), the MLE is fifth-order efficient.

We shall end by stating a consequence on asymptotic completeness. Let

$$\hat{\mathbf{u}}^{**}(\bar{\mathbf{x}}) = \hat{\mathbf{u}}(\bar{\mathbf{x}}) - \left(\frac{1}{n}\bar{\mathbf{b}}_1(\bar{\mathbf{x}}) + \frac{1}{n^2}\bar{\mathbf{b}}_2(\bar{\mathbf{x}})\right) + \left(\frac{1}{n}\mathbf{b}_1(\bar{\mathbf{x}}) + \frac{1}{n^2}\mathbf{b}_2(\bar{\mathbf{x}})\right), \tag{4.6}$$

where $\mathbf{b}_i(\bar{\mathbf{x}})$ and $\bar{\mathbf{b}}_i(\bar{\mathbf{x}})$ are the bias-adjustment factors for $\mathbf{g}(\bar{\mathbf{x}})$ and $\hat{\mathbf{u}}(\bar{\mathbf{x}})$, respectively.

Theorem 2. Suppose that $\mathbf{x}_1 \ldots, \mathbf{x}_n$ are i.i.d. ^{p}-valued random vectors defined on $(\mathcal{X}^n, \mathcal{A}^n, \mathcal{P}_\theta^n)$, where $\mathcal{P}_\theta$ is of continuous curved exponential-type (1.1) with the structure $\theta = \theta(\mathbf{u})$ $(\mathbf{u} \in U_0)$. Write $\bar{\mathbf{x}} = \frac{1}{n}\sum_{j=1}^n \mathbf{x}_j$. Assume (A1)–(A4). Let the MLE $\hat{\mathbf{u}}(\bar{\mathbf{x}})$ be defined by (1.3), and let $C \in {}_q$ be an arbitrary convex set symmetric about the origin. For any Fisher consistent third-order efficient estimator $\mathbf{g}(\bar{\mathbf{x}})$, we define $\hat{\mathbf{u}}^{**}(\bar{\mathbf{x}})$ by (4.6). Then,

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

uniformly in C, where $\Delta^2_{\hat{u},g,C}$ is defined in Theorem 1.

Theorem 2 states that for any third-order efficient estimator $\mathbf{g}(\bar{\mathbf{x}})$, there exists a modified MLE $\hat{\mathbf{u}}^{**}(\bar{\mathbf{x}})$ which is not inferior to $\mathbf{g}(\bar{\mathbf{x}})$ up to the order n^{-2} and that $\hat{\mathbf{u}}^{**}(\bar{\mathbf{x}})$ does have a higher concentration probability than $\mathbf{g}(\bar{\mathbf{x}})$ when $G_3 \neq \bar{G}_3$ and C is of positive Lebesgue measure. The result shows a kind of asymptotic completeness to the fourth- and fifth-order, cf. Pfanzagl and Wefelmeyer (1978) and Akahira, Hirakawa and Takeuchi (1988).

5 Proofs: Outline

In this section we shall describe an outline of the proofs of the theorems given in Section 4.

The formal Edgeworth expansion of the distribution of $\sqrt{n}(\mathbf{g}^*(\bar{\mathbf{x}}) - \mathbf{u})$ is obtained as follows: For any fixed $(\mathbf{u}, \mathbf{t}) \in U_0 \times {}^q$, expand formally the characteristic function of $\sqrt{n}(\widetilde{\mathbf{g}}^*(\bar{\mathbf{x}}) - \mathbf{u})$ with $\widetilde{\mathbf{g}}^*(\bar{\mathbf{x}})$ an approximate statistic to $\mathbf{g}^*(\bar{\mathbf{x}})$ up to the order n^{-2} and make the inverse Fourier transform of the expanded characteristic function, to get an expansion of the density function of $\sqrt{n}(\mathbf{g}^*(\bar{\mathbf{x}}) - \mathbf{u})$. We fix $(\mathbf{u}, \mathbf{t}) \in U_0 \times {}^q$ throughout this section.

Notice that

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}\bar{\mathbf{x}}'}\mathbf{b}_1(\bar{\mathbf{x}})\Big|_{\bar{\mathbf{x}}=\eta(\mathbf{u})} &= \frac{\mathrm{d}}{\mathrm{d}\bar{\mathbf{x}}'}\frac{1}{2}G_2(\bar{\mathbf{x}})\Psi_{20}(\theta(\mathbf{g}(\bar{\mathbf{x}})))\Big|_{\bar{\mathbf{x}}=\eta(\mathbf{u})} \\ &= \frac{1}{2}\{G_3(I_p \otimes \Psi_{20}) + G_2\Psi_{21}\Theta_1 G_1\},\end{aligned}$$

and then the approximate statistic $\widetilde{\mathbf{g}}^*(\bar{\mathbf{x}})$ to $\mathbf{g}^*(\bar{\mathbf{x}})$ obtained by expanding $\mathbf{g}^*(\bar{\mathbf{x}})$ up to the order $n^{-5/2}$ can be expressed in the form:

$$\begin{aligned}\widetilde{\mathbf{g}}^*(\bar{\mathbf{x}}) = \mathbf{u} + \frac{1}{\sqrt{n}}G_1\mathbf{z}_n + \frac{1}{2n}G_2(\mathbf{z}_n^{<2>} - \Psi_{20}) - \frac{1}{2n\sqrt{n}}G_2\Psi_{21}\Theta_1 G_1\mathbf{z}_n \\ + \frac{1}{6n\sqrt{n}}G_3\{\mathbf{z}_n^{<3>} - 3(I_p \otimes \Psi_{20})\mathbf{z}_n\} + \frac{1}{n^2}P_n(\mathbf{z}_n),\end{aligned} \tag{5.1}$$

where $P_n(\mathbf{z}_n) = P^{(1)}(\mathbf{z}_n) + \frac{1}{\sqrt{n}}P^{(2)}(\mathbf{z}_n) = O_p(1)$ is a finite polynomial of $\mathbf{z}_n$.

Since $\mathbf{x}_j$ has a continuous distribution, the distribution meets Cramer's condition: $\limsup_{||\mathbf{s}||\to\infty} \mathrm{E}_{\theta(\mathbf{u})}[e^{i\mathbf{s}'(\mathbf{x}_j - \eta(\mathbf{u}))}] < 1$ by the virtue of Bhattacharya and Denker (1990, Lemma 1.5). Therefore, the so-called *formal* Edgeworth expansion of the distribution of $\sqrt{n}(\mathbf{g}^*(\bar{\mathbf{x}}) - \mathbf{u})$ is *valid* up to the order n^{-2} in view of Theorem 2.1 of Bhattacharya and Denker (1990) (see also Bhattacharya and Ghosh, 1978). Accordingly, we evaluate asymptotic concentration probability of the approximate estimator $\widetilde{\mathbf{g}}^*(\bar{\mathbf{x}})$ about the true parameter $\mathbf{u}$ up to the order $n^{-5/2}$ or of the standardized form $\sqrt{n}(\widetilde{\mathbf{g}}^*(\bar{\mathbf{x}}) - \mathbf{u})$ about the zero vector up to the order n^{-2}, rather than that of $\mathbf{g}^*(\bar{\mathbf{x}})$ itself.

Let $\widetilde{\mathbf{u}}^*(\bar{\mathbf{x}})$ be the approximate estimator constituted from the MLE $\hat{\mathbf{u}}^*(\bar{\mathbf{x}})$ in the same manner as $\widetilde{\mathbf{g}}^*(\bar{\mathbf{x}})$. By (5.1) we can express the standardized forms of $\widetilde{\mathbf{g}}^*(\bar{\mathbf{x}})$ and $\widetilde{\mathbf{u}}^*(\bar{\mathbf{x}})$ as

$$\begin{aligned}\sqrt{n}(\widetilde{\mathbf{g}}^*(\bar{\mathbf{x}}) - \mathbf{u}) = G_1\mathbf{z}_n + \frac{1}{2\sqrt{n}}G_2(\mathbf{z}_n^{<2>} - \Psi_{20}) - \frac{1}{2n}G_2\Psi_{21}\Theta_1 G_1\mathbf{z}_n \\ + \frac{1}{6n}G_3\{\mathbf{z}_n^{<3>} - 3(I_p \otimes \Psi_{20})\mathbf{z}_n\} + \frac{1}{n\sqrt{n}}P_n(\mathbf{z}_n)\end{aligned}$$

and

$$\sqrt{n}(\widetilde{\mathbf{u}}^*(\bar{\mathbf{x}}) - \mathbf{u}) = \bar{G}_1\mathbf{z}_n + \frac{1}{2\sqrt{n}}\bar{G}_2(\mathbf{z}_n^{<2>} - \Psi_{20}) - \frac{1}{2n}\bar{G}_2\Psi_{21}\Theta_1\bar{G}_1\mathbf{z}_n + \frac{1}{6n}\bar{G}_3\{\mathbf{z}_n^{<3>} - 3(I_p \otimes \Psi_{20})\mathbf{z}_n\} + \frac{1}{n\sqrt{n}}\bar{P}_n(\mathbf{z}_n), \tag{5.2}$$

where $\bar{P}_n(\mathbf{z}_n)$ is defined as $P_n(\mathbf{z}_n)$ with $\bar{G}_k$ for G_k. We have noticed that $G_1 = \bar{G}_1$ and $G_2 = \bar{G}_2$ for $\mathbf{g}(\bar{\mathbf{x}})$ to be third-order efficient. In this case, we have

$$\begin{aligned}\sqrt{n}(\widetilde{\mathbf{g}}^*(\bar{\mathbf{x}}) - \widetilde{\mathbf{u}}^*(\bar{\mathbf{x}})) &= \sqrt{n}(\widetilde{\mathbf{g}}^*(\bar{\mathbf{x}}) - \mathbf{u}) - \sqrt{n}(\widetilde{\mathbf{u}}^*(\bar{\mathbf{x}}) - \mathbf{u}) \\ &= \frac{1}{6n}(G_3 - \bar{G}_3)\{\mathbf{z}_n^{<3>} - 3(I_p \otimes \Psi_{20})\mathbf{z}_n\} + \frac{1}{n\sqrt{n}}Q_n(\mathbf{z}_n),\end{aligned} \tag{5.3}$$

where $Q_n(\mathbf{z}_n) = P_n(\mathbf{z}_n) - \bar{P}_n(\mathbf{z}_n)$ is bounded in probability and a finite polynomial of $\mathbf{z}_n$ (with coefficients smooth in $\mathbf{u} \in U_0$) because so are $P_n(\mathbf{z}_n)$ and $\bar{P}_n(\mathbf{z}_n)$. As deduced from (4.5), the second-order bias-adjustment leads to

$$\mathrm{E}_{\theta(\mathbf{u})}[\sqrt{n}(\widetilde{\mathbf{g}}^*(\bar{\mathbf{x}}) - \widetilde{\mathbf{u}}^*(\bar{\mathbf{x}}))] = O\left(\tfrac{1}{n^2\sqrt{n}}\right), \tag{5.4}$$

which is not merely formal but valid.

From (5.3) and (5.4), one can show

$$\mathrm{E}_{\theta(\mathbf{u})}[e^{i\mathbf{t}'\sqrt{n}(\widetilde{\mathbf{u}}^*(\bar{\mathbf{x}})-\mathbf{u})}\sqrt{n}(\widetilde{\mathbf{g}}^*(\bar{\mathbf{x}}) - \widetilde{\mathbf{u}}^*(\bar{\mathbf{x}}))] = O\left(\tfrac{1}{n^2\sqrt{n}}\right) \tag{5.5}$$

and

$$\begin{aligned}&\mathrm{E}_{\theta(\mathbf{u})}[e^{i\mathbf{t}'\sqrt{n}(\widetilde{\mathbf{u}}^*(\bar{\mathbf{x}})-\mathbf{u})}\sqrt{n}(\widetilde{\mathbf{g}}^*(\bar{\mathbf{x}}) - \widetilde{\mathbf{u}}^*(\bar{\mathbf{x}}))\sqrt{n}(\widetilde{\mathbf{g}}^*(\bar{\mathbf{x}}) - \widetilde{\mathbf{u}}^*(\bar{\mathbf{x}}))'] \\ &\qquad = \frac{1}{6n^2}(G_3 - \bar{G}_3)\Psi_{11}^{<3>}(G_3 - \bar{G}_3)' \cdot e^{-\mathbf{t}'i_{\mathbf{u}}^{-1}\mathbf{t}/2} + O\left(\tfrac{1}{n^2\sqrt{n}}\right).\end{aligned} \tag{5.6}$$

The Appendix gives the flow of the proof of (5.5). The full proofs of these consequences and (5.9) are given in Kano (1995).

It follows that

$$\begin{aligned}&\mathrm{E}_{\theta(\mathbf{u})}[e^{i\mathbf{t}'\sqrt{n}(\widetilde{\mathbf{g}}^*(\bar{\mathbf{x}})-\mathbf{u})}] \\ &\quad = \mathrm{E}_{\theta(\mathbf{u})}[e^{i\mathbf{t}'\sqrt{n}(\widetilde{\mathbf{u}}^*(\bar{\mathbf{x}})-\mathbf{u})} \cdot e^{i\mathbf{t}'\sqrt{n}(\widetilde{\mathbf{g}}^*(\bar{\mathbf{x}})-\widetilde{\mathbf{u}}^*(\bar{\mathbf{x}}))}] \\ &\quad = \mathrm{E}_{\theta(\mathbf{u})}[e^{i\mathbf{t}'\sqrt{n}(\widetilde{\mathbf{u}}^*(\bar{\mathbf{x}})-\mathbf{u})}\{1 + \tfrac{1}{2!}(i\mathbf{t}'\sqrt{n}(\widetilde{\mathbf{g}}^*(\bar{\mathbf{x}}) - \widetilde{\mathbf{u}}^*(\bar{\mathbf{x}})))^2 + R_n\}] \\ &\qquad\qquad + O\left(\tfrac{1}{n^2\sqrt{n}}\right) \qquad \text{(by (5.5))} \\ &\quad = \mathrm{E}_{\theta(\mathbf{u})}[e^{i\mathbf{t}'\sqrt{n}(\widetilde{\mathbf{u}}^*(\bar{\mathbf{x}})-\mathbf{u})}] \\ &\qquad + \frac{1}{12n^2}(i\mathbf{t})'(G_3 - \bar{G}_3)\Psi_{11}^{<3>}(G_3 - \bar{G}_3)'(i\mathbf{t}) \cdot e^{-\mathbf{t}'i_{\mathbf{u}}^{-1}\mathbf{t}/2} \\ &\qquad\qquad + O\left(\tfrac{1}{n^2\sqrt{n}}\right) \qquad \text{(by (5.6))},\end{aligned} \tag{5.7}$$

where R_n is the reminder term of the expansion above, and we have used

$$\begin{aligned} |\mathrm{E}_{\theta(\mathbf{u})}[e^{i\mathbf{t}'\sqrt{n}(\widetilde{\mathbf{u}}^*(\bar{\mathbf{x}})-\mathbf{u})}R_n]| &\leq \mathrm{E}_{\theta(\mathbf{u})}[|R_n|] \\ &\leq \mathrm{E}_{\theta(\mathbf{u})}[|\mathbf{t}'\sqrt{n}(\widetilde{\mathbf{g}}^*(\bar{\mathbf{x}})-\widetilde{\mathbf{u}}^*(\bar{\mathbf{x}}))|^3/3!] \\ &= O\left(\tfrac{1}{n^3}\right) \end{aligned} \tag{5.8}$$

in view of (5.3). As a result of (5.7), the difference in the concentration probability between $\sqrt{n}(\widetilde{\mathbf{g}}^*(\bar{\mathbf{x}})-\mathbf{u})$ and $\sqrt{n}(\widetilde{\mathbf{u}}^*(\bar{\mathbf{x}})-\mathbf{u})$ based on the Edgeworth expansions appears in the n^{-2} terms, which result from the second term of the RHS of (5.7). Thus, for each $\mathbf{u}\in U_0$, the probability that $\sqrt{n}(\widetilde{\mathbf{g}}^*(\bar{\mathbf{x}})-\mathbf{u})$ be in the convex region C can be expressed in the form:

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

uniformly in C, where

$$\begin{aligned} \Delta^2_{\hat{u},g,C} = \frac{1}{12}&\mathrm{vec}(G_3-\bar{G}_3)' \\ &\cdot\left(\Psi_{11}^{<3>}\otimes\int_C\{i_{\mathbf{u}}-i_{\mathbf{u}}\mathbf{x}\mathbf{x}'i_{\mathbf{u}}\}N_q(\mathbf{x}|0,i_{\mathbf{u}}^{-1})d\mathbf{x}\right)\mathrm{vec}(G_3-\bar{G}_3). \end{aligned} \tag{5.9}$$

The vec-operator $\mathrm{vec}(A)$ for an $a_1\times a_2$ matrix A denotes an a_1a_2-vector obtained by stacking a_2 column vectors of A in order, see Henderson and Searle (1981) and Magnus and Neudecker (1988) for its properties. The expression (5.9) is obtained by making the inverse Fourier transform of the second term in (5.7).

It is known that the matrix of the integrals

$$\int_C\{i_{\mathbf{u}}-i_{\mathbf{u}}\mathbf{x}\mathbf{x}'i_{\mathbf{u}}\}N_q(\mathbf{x}|0,i_{\mathbf{u}}^{-1})d\mathbf{x}$$

is positive definite, provided that C is of positive Lebesgue measure and convex, symmetric about the origin. See Pfanzagl (1985, Lemma 13.2.4) for a proof. Thus, $\Delta^2_{\hat{u},g,C}$ is nonnegative, and $\Delta^2_{\hat{u},g,C}=0$ if and only if $G_3=\bar{G}_3$. The proof of Theorem 1 is complete.

Proof of Theorem 2. Theorem 2 can be proved in the same manner. The reader should note that

$$\sqrt{n}(\mathbf{g}(\bar{\mathbf{x}})-\hat{\mathbf{u}}^{**}(\bar{\mathbf{x}})) = \sqrt{n}(\mathbf{g}^*(\bar{\mathbf{x}})-\hat{\mathbf{u}}^*(\bar{\mathbf{x}}))$$

(so that the corresponding approximate statistics are the same) and that the same derivation in (5.7) holds even though $\widetilde{\mathbf{u}}^*(\bar{\mathbf{x}})$ is replaced with $\widetilde{\mathbf{u}}^{**}(\bar{\mathbf{x}})$ in $e^{i\mathbf{t}'\sqrt{n}(\widetilde{\mathbf{u}}^*(\bar{\mathbf{x}})-\mathbf{u})}$. *Q.E.D.*

Appendix

Here we give a condensed proof of (5.5). See Kano (1995) for the full detailed proof.

Kano (1997, Lemma 3.3) showed that for any $k \in$, there exists a $q \times qp^{k-1}$ matrix-valued smooth function $B_k = B_k(\mathbf{u})$ defined on U_0 such that $\bar{G}_k = B_k(\Theta_1' \otimes I_{p^{k-1}})N_{p<k>}$ for any $\mathbf{u} \in U_0$. The third-order efficiency, i.e., $G_2 = \bar{G}_2$, imposes a restriction on G_3:

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

In particular, $(G_3 - \bar{G}_3)\Psi_{11}^{<3>}\bar{G}_3' = 0$.

Let

$$\phi_1 = \phi_1(\mathbf{u}, \mathbf{t}) = \mathrm{E}_{\theta(\mathbf{u})}[\sqrt{n}(\widetilde{\mathbf{g}}^*(\bar{\mathbf{x}}) - \widetilde{\mathbf{u}}^*(\bar{\mathbf{x}}))e^{i\mathbf{t}'\sqrt{n}(\widetilde{\mathbf{u}}^*(\bar{\mathbf{x}})-\mathbf{u})}]$$

and

$$\phi_2 = \phi_2(\mathbf{u}, \mathbf{t}) = \mathrm{E}_{\theta(\mathbf{u})}[\sqrt{n}(\widetilde{\mathbf{g}}^*(\bar{\mathbf{x}}) - \widetilde{\mathbf{u}}^*(\bar{\mathbf{x}}))e^{i\mathbf{t}'\sqrt{n}(\widetilde{\mathbf{u}}^*(\bar{\mathbf{x}})-\mathbf{u})}\mathbf{z}_n'],$$

where $(\mathbf{u}, \mathbf{t}) \in U_0 \times {}^q$. Differentiating ϕ_i's by $\mathbf{t}'$, we have the following partial differential equations (PDE):

$$\frac{\mathrm{d}\phi_i}{\mathrm{d}\mathbf{t}'} = -\left\{(\mathbf{t}' i_{\mathbf{u}}^{-1}) \otimes \phi_i\right\} + C_i(\mathbf{u}, \mathbf{t}) \qquad (i = 1, 2), \tag{A.1}$$

where

$$\begin{aligned}C_1(\mathbf{u}, \mathbf{t}) = &\frac{i}{\sqrt{n}}\frac{\mathrm{d}\phi_1}{\mathrm{d}\mathbf{u}'}i_{\mathbf{u}}^{-1} + \frac{i}{2\sqrt{n}}\phi_1(B_2\mathrm{vec}(\Psi_{11}\Theta_1) - \bar{G}_2\Psi_{20})' \\ &- \frac{1}{2\sqrt{n}}(\mathbf{t}' \otimes \phi_2)B_2' + \frac{i}{2n}\frac{\mathrm{d}\phi_2}{\mathrm{d}\mathbf{u}'}B_2' + O\left(\tfrac{1}{n^2\sqrt{n}}\right)\end{aligned} \tag{A.2a}$$

and

$$C_2(\mathbf{u}, \mathbf{t}) = \frac{i}{\sqrt{n}}\frac{\mathrm{d}\phi_2}{\mathrm{d}\mathbf{u}'}(i_{\mathbf{u}}^{-1} \otimes I_p) + i\phi_1\mathrm{vec}(\Psi_{11}\bar{G}_1)' + O\left(\tfrac{1}{n^2}\right). \tag{A.2b}$$

If $C_i(\mathbf{u}, \mathbf{t})$'s were regarded as unrelated to ϕ_i's, the PDE (A.1) could be solved in a closed form: Let $\mathbf{t}_k = [t_1, \ldots, t_k, 0, \ldots, 0]'$, $C_1(\mathbf{u}, \mathbf{t}) = [\mathbf{c}_1^{(1)}(\mathbf{u}, \mathbf{t}), \ldots, \mathbf{c}_q^{(1)}(\mathbf{u}, \mathbf{t})]$ and

$C_2(\mathbf{u},\mathbf{t}) = [C_1^{(2)}(\mathbf{u},\mathbf{t}),\ldots,C_q^{(2)}(\mathbf{u},\mathbf{t})]$. Then, using $\mathrm{E}_{\theta(\mathbf{u})}[\sqrt{n}(\widetilde{\mathbf{g}}^*(\bar{\mathbf{x}}) - \widetilde{\mathbf{u}}^*(\bar{\mathbf{x}}))] = O\left(\frac{1}{n^2\sqrt{n}}\right)$ and $\mathrm{E}_{\theta(\mathbf{u})}[\sqrt{n}(\widetilde{\mathbf{g}}^*(\bar{\mathbf{x}}) - \widetilde{\mathbf{u}}^*(\bar{\mathbf{x}}))\mathbf{z}_n'] = O\left(\frac{1}{n^2}\right)$, we have

$$\begin{aligned}\phi_1 =& e^{-\mathbf{t}' i_{\mathbf{u}}^{-1}\mathbf{t}/2}\left\{\mathrm{E}_{\theta(\mathbf{u})}[\sqrt{n}(\widetilde{\mathbf{g}}^*(\bar{\mathbf{x}}) - \widetilde{\mathbf{u}}^*(\bar{\mathbf{x}}))] + \sum_{k=1}^{q}\int_0^{t_k}\mathbf{c}_k^{(1)}(\mathbf{u},\mathbf{t}_k)e^{\mathbf{t}_k' i_{\mathbf{u}}^{-1}\mathbf{t}_k/2}dt_k\right\}\\ =& e^{-\mathbf{t}' i_{\mathbf{u}}^{-1}\mathbf{t}/2}\sum_{k=1}^{q}\int_0^{t_k}\mathbf{c}_k^{(1)}(\mathbf{u},\mathbf{t}_k)e^{\mathbf{t}_k' i_{\mathbf{u}}^{-1}\mathbf{t}_k/2}dt_k + O\left(\frac{1}{n^2\sqrt{n}}\right) \qquad \text{(A.3a)}\end{aligned}$$

and

$$\begin{aligned}\phi_2 =& e^{-\mathbf{t}' i_{\mathbf{u}}^{-1}\mathbf{t}/2}\left\{\mathrm{E}_{\theta(\mathbf{u})}[\sqrt{n}(\widetilde{\mathbf{g}}^*(\bar{\mathbf{x}}) - \widetilde{\mathbf{u}}^*(\bar{\mathbf{x}}))\mathbf{z}_n'] + \sum_{k=1}^{q}\int_0^{t_k}C_k^{(2)}(\mathbf{u},\mathbf{t}_k)e^{\mathbf{t}_k' i_{\mathbf{u}}^{-1}\mathbf{t}_k/2}dt_k\right\}\\ =& e^{-\mathbf{t}' i_{\mathbf{u}}^{-1}\mathbf{t}/2}\sum_{k=1}^{q}\int_0^{t_k}C_k^{(2)}(\mathbf{u},\mathbf{t}_k)e^{\mathbf{t}_k' i_{\mathbf{u}}^{-1}\mathbf{t}_k/2}dt_k + O\left(\frac{1}{n^2}\right). \qquad \text{(A.3b)}\end{aligned}$$

Expand ϕ_1 and ϕ_2 as $\phi_1 = \frac{1}{n}A_1 + \frac{1}{n\sqrt{n}}A_2 + \frac{1}{n^2}A_3 + O\left(\frac{1}{n^2\sqrt{n}}\right)$ and $\phi_2 = \frac{1}{n}B_1 + \frac{1}{n\sqrt{n}}B_2 + O\left(\frac{1}{n^2}\right)$. It suffices to show $A_1 = A_2 = A_3 = 0$ for each $(\mathbf{u},\mathbf{t})$. We first note that

$$\int_0^{t_k}\phi_1 dt_k = O\left(\tfrac{1}{n}\right), \qquad \int_0^{t_k}\frac{\mathrm{d}\phi_1}{\mathrm{d}\mathbf{u}'}dt_k = O\left(\tfrac{1}{n}\right); \qquad \text{(A.4a)}$$

$$\int_0^{t_k}(\mathbf{t}'\otimes\phi_2)dt_k = O\left(\tfrac{1}{n}\right), \qquad \int_0^{t_k}\frac{\mathrm{d}\phi_2}{\mathrm{d}\mathbf{u}'}dt_k = O\left(\tfrac{1}{n}\right) \qquad \text{(A.4b)}$$

$(k = 1,\ldots,q)$ uniformly in $(\mathbf{u},\mathbf{t})$ in any compact set of $U_0 \times {}^q$.

It follows from (A.2a) and (A.4) that $\int_0^{t_k} C_1(\mathbf{u},\mathbf{t})dt_k = O\left(\frac{1}{n\sqrt{n}}\right)$ $(k = 1,\ldots,q)$ uniformly in $(\mathbf{u},\mathbf{t})$ in any compact set of $U_0 \times {}^q$ containing $(\mathbf{u}_0, 0)$, so that $\int_0^{t_k}\mathbf{c}_k^{(1)}(\mathbf{u},\mathbf{t}_k)dt_k = O\left(\frac{1}{n\sqrt{n}}\right)$, which leads to $\phi_1 = O\left(\frac{1}{n\sqrt{n}}\right)$ in view of (A.3a). As a result, A_1 equals 0 and hence,

$$\int_0^{t_k}\phi_1 dt_k = O\left(\frac{1}{n\sqrt{n}}\right), \qquad \int_0^{t_k}\frac{\mathrm{d}\phi_1}{\mathrm{d}\mathbf{u}'}dt_k = O\left(\frac{1}{n\sqrt{n}}\right) \qquad \text{(A.5)}$$

uniformly in $(\mathbf{u},\mathbf{t})$ in any compact set of $U_0 \times {}^q$. Substitution of (A.4b) and (A.5) into (A.2b) shows $\int_0^{t_k} C_2(\mathbf{u},\mathbf{t})dt_k = O\left(\frac{1}{n\sqrt{n}}\right)$ $(k = 1,\ldots,q)$ uniformly in $(\mathbf{u},\mathbf{t})$ in any compact set of $U_0 \times {}^q$ containing $(\mathbf{u}_0, 0)$, so that $\int_0^{t_k} C_k^{(2)}(\mathbf{u},\mathbf{t}_k)dt_k = O\left(\frac{1}{n\sqrt{n}}\right)$, which leads to $\phi_2 = O\left(\frac{1}{n\sqrt{n}}\right)$ in view of (A.3b). As a result, B_1 equals 0.

Thus, it is seen that starting $\phi_i = O\left(\frac{1}{n}\right)$, we get $A_1 = B_1 = 0$ and $\phi_i = O\left(\frac{1}{n\sqrt{n}}\right)$. Repeating this process, we have $A_2 = B_2 = 0$ and $\phi_i = O\left(\frac{1}{n^2}\right)$. Then, we get $A_3 = 0$ and $\phi_1 = O\left(\frac{1}{n^2\sqrt{n}}\right)$. The proof is complete.

Acknowlegments. 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.

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., Hirakawa, F. and Takeuchi, K. (1988). Second and third order asymptotic completeness of the class of estimators. In *Probability Theory and Mathematical Statistics,* Lecture Notes in Mathematics, Vol.1299 (S. Watanabe and Yu. V. Prokhorov, Eds). pp. 11-27. Springer-Verlag, Berlin.

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

[4] 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.

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

[6] Bahadur, R. R. (1964). On Fisher's bound for asymptotic variance. *Ann. Math. Statist.*, **35**, 1545-1552.

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

[8] Bhattacharya, R. and Ghosh, J. K. (1978). On the validity of the formal Edgeworth expansion. *Ann. Statist.*, **6**, 434-451.

[9] 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.

[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] Fisher, R. A. (1925). Theory of statistical estimation. *Proc. Cambridge Philos. Soc.*, **122**, 700-725.

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

[14] Ghosh, J. K. and Sinha, B. K. (1982). Third order efficiency of the MLE — A counterexample. *Calcutta Statist. Assoc. Bull.*, **31**, 151-158.

[15] Ghosh, J. K., Sinha, B. K. and Wieand, H. S. (1980). Second order efficiency of the MLE with respect to any bounded bowl-shaped loss function. *Ann. Statist.*, **8**, 506-521.

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

[17] Hajek, J. (1970). A characterization of limiting distributions of regular estimators. *Z. Wahr. verw. Geb.*, **14**, 323-330.

[18] 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.

[19] Holmquist, B. (1988). Moments and cumulants of the multivariate normal distribution. *Stoch. Anal. App.*, **6**, 273-278.

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

[21] Hosoya, Y. (1990). Information amount and higher-order efficiency in estimation. *Ann. Inst. Statist. Math.*, **42**, 37-49.

[22] Inagaki, N. (1970). On the limiting distribution of a sequence of estimators with uniformity property. *Ann. Inst. Statist. Math.*, **22**, 1-13.

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

[24] Kano, Y. (1997). Beyond third-order efficiency. *Sankhyā, Ser. B*, **59**, 179-197.

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

[26] Kano, Y. (1995). More higher-order efficiency: Concentration Probability. *Technical Report of Statistical Group:* DATA95-04, Tokyo, Japan.

[27] Kaufman, S. (1966). Asymptotic efficiency of the maximum likelihood estimator. *Ann. Inst. Statist. Math.*, **18**, 155-178.

[28] Le Cam, L. (1953). On some asymptotic properties of maximum likelihood estimates and related Bayes' estimates. *Univ. California Publ. Statist.*, **1**, 277-330.

[29] Magnus, J. R. and Neudecker, H. (1985). Matrix differential calculus with applications to simple, Hadamard, and Kronecker products. *J. Math. Psychol.*, **29**, 474-492.

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

[31] Pfanzagl, J. (1973a). The Accuracy of the normal approximation for estimates of vector parameters. *Z. Wahr. verw. Geb.*, **25**, 171-198.

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

[33] 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.

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

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

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

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

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

[39] Rao, C. R. (1973). *Linear Statistical Inference and its Applications,* 2nd ed. Wiley, New York.

[40] 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.

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

[42] Takeuchi, K. and Akahira, M. (1978). Third order asymptotic efficiency of maximum likelihood estimator for multiparameter exponential case. *Rep. Univ. Electro-Comm.*, **28**, 271-293.

[43] Takeuchi, K. and Morimune, K. (1985). Third-order efficiency of the extended maximum likelihood estimators in a simultaneous equation system. *Econometrica*, **53**, 177-200.

[44] Taniguchi, M. (1983). On the second order asymptotic efficiency of estimators of

Gaussian ARMA Processes. *Ann. Statist.*, **11**, 157-169.

[45] Taniguchi, M. (1986). Third order asymptotic properties of maximum likelihood estimators for Gaussian ARMA processes. *J. Multivariate Anal.*, **18**, 1-31.

[46] Taniguchi, M. (1987). Third order asymptotic properties of BLUE and LSE for a regression model with ARMA residual. *J. Time Ser. Anal.*, **8**, 111-114.

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

[48] 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.

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