*Sankhyā: The Indian Journal of Statistics*
1997, Volume 59, Series A, Pt. 2, pp.179-197

# BEYOND THIRD-ORDER EFFICIENCY

*By* YUTAKA KANO
*University of Tsukuba*

*SUMMARY.* Fifth-order (asymptotic) efficiency of the second-order bias-corrected MLE, minimizing the $n^{-3}$ term of an expansion of the quadratic risk of Fisher-consistent estimators bias-corrected similarly, is established in a general curved exponential family with a structural parameter vector. A characterization theorem of the MLE in terms of its higher-order derivatives is provided, and an alternative bias correction factor is proposed. Both the characterization and the new bias correction play an important role in proving the fifth-order efficiency. The matrix form symmetric tensor and higher-order derivatives are utilized, rather than usual elementwise tensors with Einstein's convention, to derive all the results of this article.

## 1. Introduction

The Fisher-Rao theorem [Fisher (1925) and Rao (1961, 1962)] states that the information loss of the MLE is minimal among an appropriate class of best asymptotic normal estimators in a curved multinomial model with a single parameter, and they say that the MLE is second-order efficient. A new terminology recently more often used for this efficiency is *third*-order efficiency, see e.g., Pfanzagl and Wefelmeyer [(1978), Note 1, page 2] and Ghosh (1994). We shall follow this terminology here. Rao (1963) points out the equivalence between two measures of third-order efficiency, namely information loss and quadratic risk. Since then, a lot of literature generalizes the basic Fisher-Rao theorem to many cases such as more general loss functions, general multi-parameter curved exponential family, regression models, and some time series models [see e.g., Ghosh and Subramanyam (1974), Efron (1975), Eguchi (1983), Amari (1985), Hosoya (1990), and Taniguchi and Watanabe (1994)]. In particular, we appreciate Efron's introduction of statistical curvature, Amari's geometrical methods in asymptotic theory, and developments on concentration probability by Pfanzagl and Wefelmeyer (1978), Ghosh, Sinha and Subramanyam (1979), Ghosh, Sinha and Wieand (1980), Akahira and Takeuchi (1981), Pfanzagl (1982), Takeuchi and Morimune (1985), and Taniguchi (1986, 1987, 1991). Note that Ghosh,

Paper received. July 1995; revised August 1996.
*AMS* (1991) *subject classifications.* 62F12.

*Key words and phrases.* Curved exponential family, Fisher-consistency, higher-order matrix derivatives, maximum likelihood estimator, quadratic risk, symmetric tensor.

Sinha and Subramanyam (1979) gave a proof of the third-order efficiency of the MLE without Cramer's condition on population distributions. Ghosh (1994) gives an excellent review on higher-order asymptotics.

Except for a few works, all the research on higher-order efficiency has aimed at studying to what extent - in what kind of models or under what kind of distribution assumption - the third-order efficiency holds, but has not proceed to *more* higher-order efficiency such as fifth-order. Efron [(1975), page 1210] mentions that the fifth-order term of Pitman's estimator may be smaller than that of the MLE, so that the MLE may not be fifth-order efficient. Rao, Sinha and Subramanyam (1982) show, by very complicated computation, that the fifth-order efficiency of the MLE holds among a relatively small class of estimators in the multinomial model with a single parameter. On the contrary, Ghosh and Sinha (1982) present a counterexample that the MLE is not fifth-order efficient. Akahira (1995), recently, shows the fourth-order efficiency of the third-order efficient estimators, based on concentration probability.

Suppose that $\mathbf{x}_1, \ldots, \mathbf{x}_n$ are independent and identically distributed $p$-dimensional random sample drawn from a population with a curved exponential family of distributions defined as

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

where $\theta$ is a $p$-vector of natural parameters with structure $\theta(u)$, $u$ is a $q$-vector of structural parameters $(q < p)$, $\psi(\theta)$ is a normalizing factor, and $\mu(dx)$ is a carrier measure [see e.g., Amari (1985), Section 4.1]. Assume that $\theta$ and $u$ are in certain open subsets of $\mathbf{R}^p$ and $\mathbf{R}^q$ and that $\psi(\theta)$ and $\theta(u)$ are continuously differentiable up to suitable order. Let $\overline{\mathbf{x}} = \frac{1}{n}\sum_{j=1}^{n}\mathbf{x}_j$, and let $g(\overline{\mathbf{x}})$ be an estimator for $u$, where $g(\cdot)$ is continuously differentiable. Let $E[\overline{\mathbf{x}}] = \eta(u)$. Fisher-consistency requires

$$g(\eta(u)) = u \qquad \ldots(1.1)$$

for all possible $u \in \mathbf{R}^q$. Let $g^*(\overline{\mathbf{x}})$ be a suitably bias-corrected version of $g(\overline{\mathbf{x}})$, and let $\mathcal{F}$ be a class of such estimators $g^*(\overline{\mathbf{x}})$. It is known that the quadratic risk of $g^*(\bar{\mathbf{x}})$ can be expanded (formally) as

$$E[(g^*(\overline{\mathbf{x}}) - u)(g^*(\overline{\mathbf{x}}) - u)'] = \frac{1}{n}V_1(u) + \frac{1}{n^2}V_2(u) + \frac{1}{n^3}V_3(u) + \ldots$$

An estimator $g^*(\overline{\mathbf{x}})$ is said to be first-order efficient in a class $\mathcal{F}$ if and only if $V_1(u)$ is minimal at all possible $u$ among all estimators in $\mathcal{F}$. A first-order efficient estimator is said to be third-order efficient in a class $\mathcal{F}$ if and only if $V_2(u)$ is minimal at all possible $u$ among all first-order efficient estimators in $\mathcal{F}$. In the same manner, we define fifth-order efficiency [cf. Amari (1985), Section 5.1].

The paper makes two main contributions to the estimation problem in the curved exponential family. One is to give a characterization theorem to the matrices $\overline{G}_k$ of $k$-th order derivatives of the MLE $\widehat{u}(\overline{\mathbf{x}})$ (Theorem 3.1), which states that for each $k, \overline{G}_k$ is a solution to a certain minimization problem under some linear constraints (see Section 2.2 for the precise definition of $\overline{G}_k$). The second is to establish the fifth-order efficiency of the bias-corrected MLE with a certain bias correction factor among the class of all third-order efficient Fisher-consistent estimators bias-corrected in a similar way, employing quadratic risk as a measure of efficiency (Theorem 4.2).

We shall end this section by stating the idea of this article. Let $Cov(\overline{\mathbf{x}}, \overline{\mathbf{x}}) = \frac{1}{n}\Psi_{11}$, and let $G_k$ be a $q \times p^k$ matrix of the $k$-th order derivatives of $g(\overline{\mathbf{x}})$. The delta theorem shows

$$V_1(u) = G_1 \Psi_{11} G_1'.$$

It is well-known that the (bias-corrected) MLE minimizes $V_1(u)$ under the condition (1.1), implying the first-order efficiency. We will extend the derivation to higher-order efficiency. That is, we will show that the bias-corrected MLE minimizes

$$V_2(u) = \frac{1}{2!} G_2 (\Psi_{11} \otimes \Psi_{11}) G_2'$$

under the condition that $g^*(\overline{\mathbf{x}})$ be first-order efficient and bias-corrected. As a result, the bias-corrected MLE is third-order efficient. Looking at these facts, one will expect that the bias-corrected MLE again minimizes

$$V_3(u) = \frac{1}{3!} G_3 (\Psi_{11} \otimes \Psi_{11} \otimes \Psi_{11}) G_3' \ (+ \text{ terms free from estimators})$$

under the condition that $g^*(\overline{\mathbf{x}})$ be third-order efficient and bias-corrected suitably. This establishes the fifth-order efficiency of the MLE. One of the results of this article is to prove the above expectation if one makes a particular bias correction.

The derivation described above relies upon the two facts : for any $(k-1)$-th order efficient estimator, i) $V_k(u)$ is essentially representable as $G_k(\Psi_{11} \otimes \ldots \otimes \Psi_{11}) G_k'$, ii) the MLE minimizes the quadratic form $G_k(\Psi_{11} \otimes \ldots \otimes \Psi_{11}) G_k'$. In Section 3 we prove the statement ii) for any $k$, which is a sort of characterization of the MLE. The statement i) is shown for $k = 1,2,3$ in Section 4. Section 2 describes some preliminary results on special matrices, higher-order matrix derivatives, moments, and the MLE.

## 2. **Preliminaries**

In this section, we summarize many useful formulas that help reduce computation very much. In the later sections, the formulas will often be used without specific citations. Elementary proofs of the formulas are given in Kano (1993).

2.1 *Commutation matrix and symmetric tensor.* The commutation matrix $K_{pq}$ is defined by the relation :

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

for all $p$-vector $\mathbf{a}$ and $q$-vector $\mathbf{b}$, where the symbol $\otimes$ denotes the (right) Kronecker product. For properties of the Kronecker product, see Rao [(1973), Section 1b.8]. Henderson and Searle (1981) present a good review of $K_{pq}$. The following basic results hold. See Magnus and Neudecker [(1979); (1988), page 47] for proofs.

PROPOSITION 2.1. *Let $A, B$ and $\mathbf{b}$ be of order $p_1 \times p_2, q_1 \times q_2$ and $q_1 \times 1$.*

i) $K_{pq}^{-1} = K_{pq}' = K_{qp}$, $K_{p1} = K_{1p} = I_p$, *and* $K_{pp}$ *is symmetric;*

ii) $K_{q_1p_1}(A \otimes B)K_{p_2q_2} = B \otimes A$, *and* $K_{q_1p_1}(A \otimes \mathbf{b}) = \mathbf{b} \otimes A$;

iii) $K_{p,qr}(I_q \otimes K_{rp}) = K_{pq} \otimes I_r$.

Magnus and Neudecker [(1988), page 48] also define $N_{p<2>} = (I_{p^2} + K_{pp})/2$. It has the property:

$$N_{p<2>}(\mathbf{a}_1 \otimes \mathbf{a}_2) = (\mathbf{a}_1 \otimes \mathbf{a}_2 + \mathbf{a}_2 \otimes \mathbf{a}_1)/2$$

for $p$-vectors $\mathbf{a}_1, \mathbf{a}_2$. The matrix $N_{p<2>}$ is equivalent to $M_p$ defined by Browne (1974), who uses $M_p$ to express the covariance matrix of a vectorized sample covariance matrix. We generalize $N_{p<2>}$ to the case of $k$-fold Kronecker product. Define the $p^k \times p^k$ matrix $N_{p<k>}$ by

$$N_{p<k>}(\mathbf{a}_1 \otimes \ldots \otimes \mathbf{a}_k) = \sum_{\binom{1\ldots k}{\sigma(1)\ldots\sigma(k)} \in S_k} \left(\mathbf{a}_{\sigma(1)} \otimes \ldots \otimes \mathbf{a}_{\sigma(k)}\right)/k!,$$

where $\mathbf{a}_j$'s are any $p$-vectors and $S_k$ denotes the $k$-th symmetric group, that is, the summation is taken over all permutations of $(1, \ldots, k)$. Note that $N_{p<1>} = I_p$ and $N_{p<k>} = 1$ for $p = 1$. The matrix is known to be symmetrizer or symmetric tensor in tensor analysis. In statistics, Holmquist (1988) uses $N_{p<k>}$ to express the higher-order moments of $N_p(0, \Sigma)$ in matrix form. We simply write $k$-fold Kronecker product of the same matrix $A$ as

$$A^{<k>} = A \otimes \ldots \otimes A,$$

and it is convenient to define $A^{<0>} = 1$. For proofs of the next proposition, see Sternberg [(1964), Section 1.3] and Satake [(1974), Section 5.3].

PROPOSITION 2.2.

i) $N_{p<k>}$ *is symmetric, idempotent and rank* $(N_{p<k>}) = \binom{p+k-1}{k}$:

ii) $N_{p^{<k>}}\mathbf{a}^{<k>} = \mathbf{a}^{<k>}$ *for* $p$*-vector* $\mathbf{a}$;

iii) $N_{p^{<k>}}A^{<k>} = A^{<k>}N_{q^{<k>}} = N_{p^{<k>}}A^{<k>}N_{q^{<k>}}$ *for* $A$ *of* $p \times q$;

iv) $(I_{p^{k_1}} \otimes K_{p^{k_2},p^{k_3}} \otimes I_{p^{k_4}})N_{p^{<k>}}$ *for any* $k_j \geq 0$ *with* $\sum_j^4 k_j = k$;
$(I_{p^{k_1}} \otimes N_{p^{<k_2>}} \otimes I_{p^{k_3}})N_{p^{<k>}} = N_{p^{<k>}}$ *for any* $k_j \geq 0$ *with* $\sum_j^3 k_j = k$.

2.2 *Higher-order matrix derivatives.* Let $\mathbf{x} = [x_1, \ldots, x_p]'$ and $A(\mathbf{x}) = (a_{ij}(\mathbf{x}))$ be an $a_1 \times a_2$ matrix-valued function. Put $\frac{\partial A(\mathbf{X})}{\partial x_k} = \left(\frac{\partial a_{ij}(\mathbf{X})}{\partial x_k}\right)$, an $a_1 \times a_2$ matrix. Then we define

$$\frac{dA(\mathbf{x})}{d\mathbf{x}'} = \frac{d}{d\mathbf{x}'}A(\mathbf{x}) = \left[\frac{\partial A(\mathbf{x})}{\partial x_1}, \ldots, \frac{\partial A(\mathbf{x})}{\partial x_p}\right],$$

which is of order $a_1 \times pa_2$ and

$$\frac{dA(\mathbf{x})}{d\mathbf{x}} = \frac{d}{d\mathbf{x}}A(\mathbf{x}) = \begin{bmatrix} \frac{\partial A(\mathbf{x})}{\partial x_1} \\ \vdots \\ \frac{\partial A(\mathbf{x})}{\partial x_p} \end{bmatrix},$$

which is of order $pa_1 \times a_2$. The definition coincides with that of Magnus and Neudecker [(1988, page 114] when the variate is a vector.

The higher-order derivatives are defined inductively by $\left(\frac{d}{d\mathbf{X}'}\right)^{<k>} = \frac{d}{d\mathbf{X}'}\left(\frac{d}{d\mathbf{X}'}\right)^{<k-1>}$. Let $g(\mathbf{x})$ be a $q$-vector valued analytic function. The Taylor series of $g(\mathbf{x})$ about $\mathbf{x}_0$ is expressible in the form :

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

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

$$G_k N_{p^{<k>}} = G_k, \qquad \ldots(2.1)$$

The following are basic but useful formulas on matrix differentiations [cf. Bentler and Lee (1978) and Magnus and Neudecker (1985)].

PROPOSITION 2.3. *Let* $\mathbf{x}$ *be a* $p$*-vector and let* $A(\mathbf{x})$ *and* $B(\mathbf{x})$ *be of order* $a_1 \times a_2$ *and* $b_1 \times b_2$. *Suppose that the following matrix products in the LHS are permissible.*

i) $\frac{d}{d\mathbf{X}'}[A(\mathbf{x})B(\mathbf{x})] = \frac{dA(\mathbf{X})}{d\mathbf{X}'}(I_p \otimes B(\mathbf{x})) + A(\mathbf{x})\frac{dB(\mathbf{X})}{d\mathbf{X}'}$;

ii) $\frac{d}{d\mathbf{X}'}[A(\mathbf{x}) \otimes B(\mathbf{x})] = \left(\frac{dA(\mathbf{X})}{d\mathbf{X}'} \otimes B(\mathbf{x})\right) + \left(A(\mathbf{x}) \otimes \frac{dB(\mathbf{X})}{d\mathbf{X}'}\right)(K_{a_2 p} \otimes I_{b_2})$;

iii) $\frac{d}{d\mathbf{t}'}[A(\mathbf{x}(\mathbf{t}))B] = \frac{dA(\mathbf{x})}{d\mathbf{x}'}\left(\frac{d\mathbf{x}(\mathbf{t})}{d\mathbf{t}'} \otimes B\right)$, *where* $\mathbf{x}(\mathbf{t})$ *is a* $p$*-vector valued function of* $\mathbf{t}$ *and* $B$ *is a constant matrix.*

The corresponding formulas on $\left(\frac{d}{d\mathbf{x}}\right)^{<k>}$ are easily obtained in view of the fact that $\left(\left(\frac{d}{d\mathbf{x}}\right)^{<k>}A(\mathbf{x})\right)' = \left(\frac{d}{d\mathbf{x}'}\right)^{<k>}A(\mathbf{x})'$.

2.3 *Moments.* Let $\mathbf{x}_1, \ldots, \mathbf{x}_n$ be independent and identically distributed random $p$-vectors with $E[\mathbf{x}_1] = \eta$, and let

$$\Psi_{kl} = E[(\mathbf{x}_1 - \eta)^{<k>}(\mathbf{x}_1 - \eta)^{<l>'}].$$

Put $\mathbf{z}_n = \sqrt{n}(\overline{\mathbf{x}} - \eta)$, where $\overline{\mathbf{x}}$ is the sample mean vector. Using the symmetrizer $N_{p^{<k>}}$ in Section 2.1, we can express some expectations of the sample moments in a fairly simple form :

$$\begin{aligned}
E[\mathbf{z}_n\mathbf{z}_n'] &= \Psi_{11}, \quad E[\mathbf{z}_n^{<2>}] = \Psi_{20}, \quad E[\mathbf{z}_n^{<3>}] = \frac{1}{\sqrt{n}}\Psi_{30}; \\
E[\mathbf{z}_n^{<2>}\mathbf{z}_n^{<2>\prime}] &= 2N_{p^{<2>}}\Psi_{11}^{<2>}N_{p^{<2>}} + \Psi_{20}\Psi_{02} + O\left(\frac{1}{n}\right); \\
E[\mathbf{z}_n^{<3>}\mathbf{z}_n'] &= 3N_{p^{<3>}}(\Psi_{20}\otimes\Psi_{11}) + O\left(\frac{1}{n}\right) = 3N_{p^{<3>}}(\Psi_{11}\otimes\Psi_{20}) + O\left(\frac{1}{n}\right); \\
E[\mathbf{z}_n^{<4>}] &= 3N_{p^{<4>}}\Psi_{20}^{<2>} + O\left(\frac{1}{n}\right); \\
E[\mathbf{z}_n^{<3>}\mathbf{z}_n^{<3>\prime}] &= 6N_{p^{<3>}}\Psi_{11}^{<3>}N_{p^{<3>}} \\
&\quad + 9N_{p^{<3>}}(\Psi_{20}\otimes\Psi_{11}\otimes\Psi_{02})N_{p^{<3>}} + O\left(\frac{1}{n}\right).
\end{aligned}$$

Notice that

$$N_{p^{<3>}}(\Psi_{02}\otimes\Psi_{11}\otimes\Psi_{20})N_{p^{<3>}} = N_{p^{<3>}}(\Psi_{20}\otimes\Psi_{11}\otimes\Psi_{02})N_{p^{<3>}} \qquad \ldots(2.2)$$

by virtue of $N_{p^{<3>}}$. Moreover $N_{p^{<3>}}(\Psi_{20}\otimes\Psi_{11}\otimes\Psi_{02})N_{p^{<3>}}$ remains unchanged under any permutations of $\Psi_{20}, \Psi_{11}, \Psi_{02}$.

Some immediate consequences involving $N_{p^{<k>}}$ and moment matrices are ready for later use. Let $G_3 = \left(\frac{d}{d\mathbf{x}'}\right)^{<3>} g(\mathbf{x})|_{\mathbf{x}=\eta}$ as in Section 2.2. It follows that

$$\begin{aligned}
&G_3E[(\mathbf{z}_n^{<3>} - 3(\mathbf{z}_n\otimes\Psi_{20}))\mathbf{z}_n'] \\
&= G_3\left[3N_{p^{<3>}}(\Psi_{11}\otimes\Psi_{20}) + O\left(\frac{1}{n}\right) - 3(\Psi_{11}\otimes\Psi_{20})\right] = O\left(\frac{1}{n}\right) \qquad \ldots(2.3a)
\end{aligned}$$

in view of (2.1). Noting (2.2), a similar argument leads to

$$G_3E[(\mathbf{z}_n^{<3>} - 3(\mathbf{z}_n\otimes\Psi_{20}))\mathbf{z}_n^{<3>\prime}]G_3' = 6G_3\Psi_{11}^{<3>}G_3' + O\left(\frac{1}{n}\right), \qquad \ldots(2.3b)$$

$$G_3 E[(\mathbf{z}_n^{<3>} - 3(\mathbf{z}_n \otimes \Psi_{20}))(\mathbf{z}_n^{<3>} - 3(\mathbf{z}_n \otimes \Psi_{20}))']G_3' = 6G_3\Psi_{11}^{<3>}G_3' + O\left(\frac{1}{n}\right). \quad \ldots(2.3c)$$

Let

$$\Psi_{kl}(\theta) = \left(\frac{d}{d\theta}\right)^{<k>}\left(\frac{d}{d\theta'}\right)^{<l>}\psi(\theta).$$

In particular, $E[\mathbf{x}_1] = \Psi_{10}(\theta(u)) = \eta(u)$, being the expectation parameter. It is also known that

$$E[(\mathbf{x}_1 - \eta(u))^{<k>}(\mathbf{x}_1 - \eta(u))^{<l>'}] = \Psi_{kl}(\theta(u)) \text{ for } k + l = 2, 3.$$

Accordingly, the same symbol $\Psi_{kl}$ is used as the moment matrices.

2.4 *MLE* Let

$$\Theta_1(u) = \frac{d}{du'}\theta(u).$$

The likelihood function $L(u|\overline{\mathbf{x}})$ of $u$ based on the sample $\mathbf{x}_1, \ldots, \mathbf{x}_n$ is of the form : $L(u|\overline{\mathbf{x}}) = \exp\{n(\theta(u)'\overline{\mathbf{x}} - \psi(\theta(u)))\}$, and the derivatives of the log-likelihood is written as

$$\frac{d}{du}logL(u|\overline{\mathbf{x}}) = n\Theta_1(u)'\{\overline{\mathbf{x}} - \eta(u)\} = \sqrt{n}\Theta_1(u)'\mathbf{z}_n \quad \ldots(2.4)$$

with $\mathbf{z}_n = \sqrt{n}\{\overline{\mathbf{x}} - \eta(u)\}$. The MLE $\hat{u} = \hat{u}(\overline{\mathbf{x}})$ is assumed to be a unique solution to the equation

$$\Theta_1(u)'\{\overline{\mathbf{x}} - \eta(u)\} = 0. \quad \ldots(2.5)$$

It is also known that the Fisher information matrix of $u$ is given as

$$i_u = \Theta_1(u)'\Psi_{11}(\theta(u))\Theta_1(u), \quad \ldots(2.6)$$

and that

$$\left.\frac{d\hat{u}(\overline{\mathbf{x}})}{d\overline{\mathbf{x}}'}\right|_{\overline{\mathbf{x}}=\eta(u)} = i_u^{-1}\Theta_1(u)'. \quad \ldots(2.7)$$

Basically, the following sections take over these definitions and notations. For simplicity, sometimes, let us write $\eta, \Theta_l$ and $\Psi_{kl}$ for $\eta(u), \Theta_l(u)$ and $\Psi_{kl}(\theta(u))$ respectively if there is no confusion.

## 3. Characterization of the MLE

When $\psi(\theta)$ and $\theta(u)$ are $k+1$ times continuously differentiable and the Fisher information matrix $i_u$ in (2.6) is positive definite, then the implicit function theorem guarantees that the likelihood equation (2.5) uniquely determines the

MLE $\hat{u}(\overline{\mathbf{x}})$ in a neighborhood of $u$ for all $\overline{\mathbf{x}}$ in a neighborhood of $\eta(u)$ and that $\hat{u}(\overline{\mathbf{x}})$ is $k$ times continuously differentiable. We write

$$\overline{G}_j = \overline{G}_j(\eta(u)) = \left(\frac{d}{d\overline{\mathbf{x}}'}\right)^{<j>} \hat{u}(\overline{\mathbf{x}})\Big|_{\overline{\mathbf{x}}=\eta(u)} \quad (j = 1, \ldots, k). \qquad \ldots(3.1)$$

Let us denote by $g(\overline{\mathbf{x}})$ an arbitrary Fisher-consistent estimator as in Section 1. When it is differentiable, we write

$$G_j = G_j(\eta(u)) = \left(\frac{d}{d\overline{\mathbf{x}}'}\right)^{<j>} g(\overline{\mathbf{x}})\Big|_{\overline{\mathbf{x}}=\eta(u)} \quad (j = 1, \ldots, k). \qquad \ldots(3.2)$$

It is convenient to write $\overline{G}_0 = \hat{u}(\eta(u)) = u$ and $G_0 = g(\eta(u)) = u$.

The following theorem asserts that the derivative matrix $\overline{G}_k$ is obtained as a solution to an optimization problem subject to some linear constraints. Let $\mathbf{N}$ be the set of natural numbers.

THEOREM 3.1 *Let $k \in \mathbf{N}$ be arbitrarily fixed. Assume that $\psi(\theta)$ and $\theta(u)$ are $k+1$ times continuously differentiable and that $i_u$ and $\Psi_{11}$ are positive definite. Let $\overline{G}_k$ be a matrix of the $k$-th order derivatives of the MLE $\hat{u}(\overline{\mathbf{x}})$, defined by (3.1). Then, $\overline{G}_k$ is a unique solution to the minimization problem :*

$$\min_{X:q\times p^k} X\Psi_{11}^{<k>}X'$$

*subject to*

$$X(\Psi_{11}\Theta_1 \otimes I_{p^{k-1}}) = \frac{d}{du'}\overline{G}_{k-1}(\eta(u)), \qquad \ldots(3.3a)$$

$$XN_{p^{<k>}} = X. \qquad \ldots(3.3b)$$

The theorem leads to the following important relation between the derivatives of the two estimators $\hat{u}(\overline{\mathbf{x}})$ and $g(\overline{\mathbf{x}})$.

COROLLARY 3.2 *In addition to the assumptions in Theorem 3.1, suppose that a Fisher-consistent estimator $g(\overline{\mathbf{x}})$ is $k$ times continuously differentiable, and let $G_k$ be defined in (3.2). If $G_{k-1} = \overline{G}_{k-1}$, then $G_k\Psi_{11}^{<k>}G_k' \geq \overline{G}_k\Psi_{11}^{<k>}\overline{G}_k'$. The equality holds if and only if $G_k = \overline{G}_k$.*

REMARK 3.1 For any Fisher-consistent estimator, we have $G_0(\eta(u)) = u = \overline{G}_0(\eta(u))$, which implies that the assumption in Corollary 3.2 is met at $k = 1$. Thus, $G_1\Psi_{11}G_1' \geq \overline{G}_1\Psi_{11}\overline{G}_1'$. This shows that the MLE is first-order efficient. In addition, $g(\overline{\mathbf{x}})$ is first-order efficient if and only if $G_1 = \overline{G}_1$. Such logic will be used to prove third- and fifth-order efficiency in the next section.

REMARK 3.2 From the proof of the corollary stated later, it will be seen that the constraint (3.3a) comes from $G_{k-1} = \overline{G}_{k-1}$. What does (3.3b) imply? An implication thereof can be described as follows : Some columns of the derivatives $\overline{G}_k$ are functionally dependent, for example, $\frac{\partial^2\hat{u}(\overline{\mathbf{x}})}{\partial\bar{x}_1\partial\bar{x}_2}$ and $\frac{\partial^2\bar{u}(\overline{\mathbf{x}})}{\partial\bar{x}_2\partial\bar{x}_1}$ (at

$\bar{\mathbf{x}} = \eta$) are identical but located in different columns in $\bar{G}_2$. The constraint $XN_{p^{<k>}} = X$ reflects such restrictions. Note that the number of functionally independent columns of $\bar{G}_k$ is $\binom{p+k-1}{k}$, which equals the rank of $N_{p^{<k>}}$ as shown in Proposition 2.2 i).

REMARK 3.3 Eguchi (1985) uses differential-geometric terminology to present some result on $\bar{G}_k$. His result may be related to the following lemma.

We state a lemma to prove Theorem 3.1. A proof thereof will be given in the appendix.

LEMMA 3.3 *For any $k \in \mathbf{N}$, there exists a $q \times qp^{k-1}$ matrix $B_k$ such that*

$$\bar{G}_k = B_k(\Theta_1' \otimes I_{p^{k-1}})N_{p^{<k>}}. \qquad \ldots(3.5)$$

PROOF OF THEOREM 3.1. It follows that

$$X\Psi_{11}^{<k>}X' = (X - \bar{G}_k)\Psi_{11}^{<k>}(X - \bar{G}_k)' + \bar{G}_k\Psi_{11}^{<k>}\bar{G}_k'$$

because

$$\begin{aligned}(X - \bar{G}_k)\Psi_{11}^{<k>}\bar{G}_k' &= (X - \bar{G}_k)N_{p^{<k>}}\Psi_{11}^{<k>}(\Theta_1 \otimes I_{p^{k-1}})B_k' \\ &= (X - \bar{G}_k)(\Psi_{11}\Theta_1 \otimes I_{p^{k-1}})(I_q \otimes \Psi_{11}^{<k-1>})B_k' \\ &= 0.\end{aligned}$$

The first equality holds from Lemma 3.3 and Proposition 2.2 iii); the second and last equalities are due to the constrains in (3.3b) and (3.3a). The uniqueness is obvious.

*Q.E.D.*

*Proof of Corollary 3.2* Remember that $G_{k-1} = G_{k-1}(\eta(u))$ with $\eta(u) = \Psi_{10}(\theta(u))$. Thus, differentiation of $G_{k-1} = \bar{G}_{k-1}$ leads to $G_k(\Psi_{11}\Theta_1 \otimes I_{p^{k-1}}) = \frac{d}{du'}\bar{G}_{k-1}(\eta(u))$ in view of Proposition 2.3 iii). Note that $G_kN_{p^{<k>}} = G_k$, as described in (2.1). Thus, the matrix $G_k$ satisfies the constraints on $X$ in (3.3). As a result, $G_k\Psi_{11}^{<k>}G_k' \geq \bar{G}_k\Psi_{11}^{<k>}\bar{G}_k'$. The uniqueness of $\bar{G}_k$ in the minimization problem proves the last statement.

*Q.E.D.*

## 4. **Asymptotic efficiency**

In this section, we develop the theory of higher-order asymptotic efficiency. Recall that $\hat{u}(\bar{\mathbf{x}})$ and $g(\bar{\mathbf{x}})$ are the MLE and an arbitrary Fisher-consistent, differentiable estimator. We expand

$$\hat{u}(\bar{\mathbf{x}}) = u + \frac{1}{\sqrt{n}}\bar{G}_1\mathbf{z}_n + \frac{1}{2n}\bar{G}_2\mathbf{z}_n^{<2>} + \frac{1}{6n\sqrt{n}}\bar{G}_3\mathbf{z}_n^{<3>} + o_p\left(\frac{1}{n\sqrt{n}}\right)$$

and

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

with $\mathbf{z}_n = \sqrt{n}(\bar{\mathbf{x}} - \eta(u))$.

The bias of $g(\bar{\mathbf{x}})$ has the form :

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

The first-order bias correction is made with

$$\frac{1}{2n} \widehat{G_2 \Psi}_{20}, \qquad \ldots(4.1)$$

so that

$$\begin{aligned} & E[g(\bar{\mathbf{x}}) - \frac{1}{2n} \widehat{G_2 \Psi}_{20} - u] \\ = \; & \frac{1}{n^2} \left[ - \left( \frac{1}{4} \frac{d}{d\bar{\mathbf{x}}'}^{<2>} \widehat{G_2 \Psi}_{20} \Big|_{\bar{\mathbf{x}} = \eta(u)} \right) \Psi_{20} + \frac{1}{6} G_3 \Psi_{30} + \frac{1}{8} G_4 \Psi_{20}^{<2>} \right] + o\left(\frac{1}{n^2}\right) \\ = \; & \frac{1}{n^2} b_2 + o\left(\frac{1}{n^2}\right) \quad \text{say.} \end{aligned}$$

Here $b_2$ is said to be the second-order bias.

Let $\hat{b}_2$ be an estimator of $b_2$. We now consider a family $\mathcal{F}$ of second order bias corrected Fisher-consistent estimators :

$$\begin{aligned} \mathcal{F} = \Bigg\{ & g^*(\bar{\mathbf{x}}) = g(\bar{\mathbf{x}}) - \frac{1}{2n} \widehat{G_2 \Psi}_{20} - \frac{1}{n^2} \hat{b}_2 \Bigg| \\ & E[g^*(\bar{\mathbf{x}}) = u] = o\left(\frac{1}{n^2}\right), g(\eta(u)) = u \;\; \text{for all possible } u \in \mathbf{R}^q \Bigg\} \qquad \ldots(4.2) \end{aligned}$$

The class $\mathcal{F}$ depends on choice of bias correction factors. Here we take

$$\widehat{G_2 \Psi}_{20} = G_2(\bar{\mathbf{x}}) \Psi_{20}(\theta(g(\bar{\mathbf{x}}))). \qquad \ldots(4.3)$$

Let us refer to the bias-correction factor in some details. The only requirement for $\widehat{G_2 \Psi}_{20}$ in (4.1) is $E[\widehat{G_2 \Psi}_{20}] = G_2 \Psi_{20} + o\left(\frac{1}{n}\right)$ in order to prove third-order efficiency. The estimation method does not matter. On the other hand, the requirement is not enough to study fifth-order optimality because the distribution of $\widehat{G_2 \Psi}_{20}$ does influence upon the $n^{-3}$ terms. The counterexample in Ghosh and Sinha (1982) shows that the usual bias correction factor does not work, which is given by $\widehat{G_2 \Psi}_{20} = G_2(\eta(g(\bar{\mathbf{x}}))) \Psi_{20}(\theta(g(\bar{\mathbf{x}})))$. The choice for an estimator $\widehat{G_2 \Psi}_{20}$ in (4.3) enables us to establish the fifth-order efficiency of the MLE.

THEOREM 4.1. *Assume that $\psi(\theta)$ and $\theta(u)$ are four times continuously differentiable and $i_u$ in (2.6) is positive definite. Let $\hat{u}(\bar{\mathbf{x}})$ and $g(\bar{\mathbf{x}})$ be, respectively, the MLE and an arbitrary Fisher-consistent, three times continuously differentiable estimator, with $\bar{G}_k$ in (3.1) and $G_k$ in (3.2) being matrices of the k-th order derivatives of $\hat{u}(\bar{\mathbf{x}})$ and $g(\bar{\mathbf{x}})$. Let $\mathcal{F}$ be defined in (4.2) with (4.3). Let $\hat{u}^*(\bar{\mathbf{x}})$ and $g^*(\bar{\mathbf{x}})$, in the $\mathcal{F}$, be second-order bias corrected versions of $\hat{u}(\bar{\mathbf{x}})$ and $g(\bar{\mathbf{x}})$. Then*

*i) for any $g^*(\bar{\mathbf{x}}) \in \mathcal{F}$,*

$$E[g^*(\bar{\mathbf{x}}) - u)(g^*(\bar{\mathbf{x}}) - u)'] = \frac{1}{n}G_1\Psi_{11}G_1' - \frac{1}{n}\bar{G}_1\Psi_{11}\bar{G}_1' + E[(\hat{u}^*(\bar{\mathbf{x}}) - u)(\hat{u}(\bar{\mathbf{x}})^* - u)'] + o\left(\frac{1}{n}\right); \quad \ldots(4.4a)$$

*ii) for any $g^*(\bar{\mathbf{x}}) \in \mathcal{F}$ with $G_1 = \bar{G}_1$,*

$$E[g^*(\bar{\mathbf{x}}) - u)(g^*(\bar{\mathbf{x}}) - u)'] = \frac{1}{2n^2}G_2\Psi_{11}^{<2>}G_2' - \frac{1}{2n^2}\bar{G}_2\Psi_{11}^{<2>}\bar{G}_2' + E[(\hat{u}^*(\bar{\mathbf{x}}) - u)(\hat{u}(\bar{\mathbf{x}}) - u)'] + o\left(\frac{1}{n^2}\right); \quad \ldots(4.4b)$$

*iii) for any $g^*(\bar{\mathbf{x}}) \in \mathcal{F}$ with $G_1 = \bar{G}_1$ and $G_2 = \bar{G}_2$,*

$$E[g^*(\bar{\mathbf{x}}) - u)(g^*(\bar{\mathbf{x}}) - u)'] = \frac{1}{3!n^3}G_3\Psi_{11}^{<3>}G_3' - \frac{1}{3!n^3}\bar{G}_3\Psi_{11}^{<3>}\bar{G}_3' + E[(\hat{u}^*(\bar{\mathbf{x}}) - u)(\hat{u}(\bar{\mathbf{x}}) - u)'] + o\left(\frac{1}{n^3}\right). \quad \ldots(4.4c)$$

REMARK 4.1 (4.4a) holds for estimators without bias correction; (4.4b) holds for those without second-order bias correction.

REMARK 4.2 The terms depending on the estimator $g(\bar{\mathbf{x}})$ are $G_k\Psi_{11}^{<k>}G_k'$ with $k = 1, 2, 3$; the other terms relate only to the MLE.

REMARK 4.3 The second and third terms in the RHS in (4.4a) are canceled to be $o\left(\frac{1}{n}\right)$; those in (4.4b) reduce to $\frac{1}{n}\bar{G}_1\Psi_{11}\bar{G}_1'$ [see Kano (1993) for a proof].

We are now in a position to study higher-order asymptotic efficiency with the help of Corollary 3.1 and Theorem 4.1. We shall make a similar discussion to that in Remark 3.1. Assume that $\Psi_{11}$ is positive definite. Recall that $\hat{u}^*(\bar{\mathbf{x}})$ and $g^*(\bar{\mathbf{x}})$ belong to $\mathcal{F}$. The quadratic risk of $g^*(\bar{\mathbf{x}})$ can be expressed as (4.4a), and Corollary 3.1 asserts that $g^*(\bar{\mathbf{x}})$ minimizes the quadratic form $G_1\Psi_{11}G_1'$ in $\mathcal{F}$ if and only if $G_1 = \bar{G}_1$. Thus, the estimators $g^*(\bar{\mathbf{x}})$ meeting $G_1 = \bar{G}_1$ are first-order efficient in $\mathcal{F}$, and particularly the (bias-corrected) MLE $\hat{u}^*(\bar{\mathbf{x}})$ is first-order efficient in $\mathcal{F}$. The quadratic risk of a first-order efficient estimator $g^*(\bar{\mathbf{x}})$ can be expressed as (4.4b), and Corollary 3.1 asserts that $g^*(\bar{\mathbf{x}})$ minimizes the quadratic form $G_2\Psi_{11}^{<2>}G_2'$ in the class of first-order efficient estimators in $\mathcal{F}$ if and only if $G_2 = \bar{G}_2$. Thus, the estimators $g^*(\bar{\mathbf{x}})$ meeting $G_2 = \bar{G}_2$ are third-

order efficient in $\mathcal{F}$, and particularly the bias-corrected MLE $\hat{u}^*(\bar{\mathbf{x}})$ is third-order efficient in $\mathcal{F}$.

The same argument holds for fifth-order efficiency. Indeed, let $g^*(\bar{\mathbf{x}})$ be third-order efficient, that is, $G_1 = \bar{G}_1$ and $G_2 = \bar{G}_2$, and hence the quadratic risk of $g^*(\bar{\mathbf{x}})$ can be expressed as (4.4c). Corollary 3.2 shows $g^*(\bar{\mathbf{x}})$ minimizes the $G_3 \Psi_{1\bar{1}}^{<3>} G_3'$ in the class of third-order efficient estimators in $\mathcal{F}$ if and only if $G_3 = \bar{G}_3$. Thus, the second-order bias-corrected MLE $\hat{u}^*(\bar{\mathbf{x}})$ is fifth-order efficient in $\mathcal{F}$.

We summarize the above results as the following two statements.

THEOREM 4.2 *In addition to the assumptions in Theorem 4.1, assume that* $\Psi_{11}$ *is positive definite. Then, an estimator* $g^*(\bar{\mathbf{x}}) \in \mathcal{F}$ *is first-order efficient in* $\mathcal{F}$ *if and only if* $G_1 = \bar{G}_1$*; a first-order efficient estimator* $g^*(\bar{\mathbf{x}})$ *is third-order efficient in* $\mathcal{F}$ *if and only if* $G_2 = \bar{G}_2$*; finally a third-order efficient estimator* $g^*(\bar{\mathbf{x}})$ *is fifth-order efficient in* $\mathcal{F}$ *if and only if* $G_3 = \bar{G}_3$.

COROLLARY 4.3 *Assume the same assumptions as in Theorem 4.2. The second-order bias-corrected MLE* $\hat{u}^*(\bar{\mathbf{x}}) \in \mathcal{F}$ *be first-, third- and fifth-order efficient in* $\mathcal{F}$.

REMARK 4.4 It follows from Remark 4.1 that the first-order efficiency (the third-order efficiency) in Theorem 4.2 and Corollary 4.3 holds for a class of estimators without bias correction (without second-order bias correction).

Before proving Theorem 4.1, we state a useful formula in evaluating expectations involving the score function, cf. Akahira and Takeuchi [(1981), Lemma 5.1.1], Amari [(1985), page 124] and Taniguchi and Watanabe [(1994), formula (5.10)]. A proof of the lemma will be given in the appendix.

LEMMA 4.4 *Let* $A(\mathbf{z}_n|u)$ *be of* $a_1 \times a_2$. *Then*

$$E[\mathbf{z}_n' \Theta_1 \otimes A(\mathbf{z}_n|u)] = \frac{1}{\sqrt{n}} \left\{ \frac{d}{du'} E[A(\mathbf{z}_n|u)] - E\left[ \frac{d}{du'} A(\mathbf{z}_n|u) \right] \right\}. \qquad \ldots(4.5)$$

Note two things : if $A(\mathbf{z}_n|u)$ is independent of $u$, the second term of the RHS in (4.5) disappears; for a vector-valued function $\mathbf{a}(\mathbf{z}_n|u), \mathbf{z}_n' \Theta_1 \otimes \mathbf{a}(\mathbf{z}_n|u) = \mathbf{a}(\mathbf{z}_n|u)\mathbf{z}'\Theta_1$.

PROOF OF THEOREM 4.1. The result (4.4a) is obvious. As a consequence of the bias correction, we have

$$\frac{1}{n} E[g^*(\bar{\mathbf{x}}) - \hat{u}^*(\bar{\mathbf{x}}))] = o\left(\frac{1}{n^2}\right) \text{ or } o\left(\frac{1}{n^3}\right) \qquad \ldots(4.6)$$

and

$$E\left[ (g^*(\bar{\mathbf{x}}) - \hat{u}^*(\bar{\mathbf{x}})) \left( \frac{1}{\sqrt{n}} \bar{G}_1 \mathbf{z}_n \right)' \right] = o\left(\frac{1}{n^2}\right) \text{ or } o\left(\frac{1}{n^3}\right), \qquad \ldots(4.7)$$

where the RHS's are $o(\frac{1}{n^2})$ if the first-order bias correction, with (4.1), is made; they are $o(\frac{1}{n^3})$ if the second-order bias correction is made. The evaluation (4.7)

is obtained because the LHS is equal to

$$\frac{1}{\sqrt{n}}E[(g^*(\bar{\mathbf{x}})-\hat{u}^*(\bar{\mathbf{x}}))\mathbf{z}_n'\Theta_1]i_u^{-1}=\frac{1}{n}\frac{d}{du'}E[g^*(\bar{\mathbf{x}})-\hat{u}^*(\bar{\mathbf{x}})]i_u^{-1}$$

in view of (2.7) and Lemma 4.4, and then (4.6) applies.

Let us show (4.4b). We have

$$\hat{u}^*(\bar{\mathbf{x}})-u=\frac{1}{\sqrt{n}}\bar{G}_1\mathbf{z}_n+\frac{1}{2n}\bar{G}_2(\mathbf{z}_n^{<2>}-\Psi_{20})+o_p\left(\frac{1}{n}\right), \qquad \ldots(4.8)$$

and $g^*(\bar{\mathbf{x}})$ permits the same expansion with $G_2$ for $\bar{G}_2$. Therefore,

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

which implies

$$E[(g^*(\bar{\mathbf{x}})-\hat{u}^*(\bar{\mathbf{x}}))(g^*(\bar{\mathbf{x}})-\hat{u}^*(\bar{\mathbf{x}}))']=\frac{1}{2n^2}(G_2-\bar{G}_2)\Psi_{11}^{<2>}(G_2-\bar{G}_2)'+o\left(\frac{1}{n^2}\right) \qquad \ldots(4.9)$$

and

$$E\left[(g^*(\bar{\mathbf{x}})-\hat{u}^*(\bar{\mathbf{x}}))\left(\frac{1}{2n}\bar{G}_2(\mathbf{z}_n^{<2>}-\Psi_{20})\right)'\right]=\frac{1}{2n^2}(G_2-\bar{G}_2)\Psi_{11}^{<2>}\bar{G}_2'+o\left(\frac{1}{n^2}\right). \qquad \ldots(4.10)$$

Use of (4.7), (4.8) and (4.10) leads to

$$E[(g^*(\bar{\mathbf{x}})-\hat{u}^*(\bar{\mathbf{x}}))(\hat{u}^*(\bar{\mathbf{x}})-u)']=\frac{1}{2n^2}(G_2-\bar{G}_2)\Psi_{11}^{<2>}\bar{G}_2'+o\left(\frac{1}{n^2}\right). \qquad \ldots(4.11)$$

The result (4.4b) follows from (4.9), (4.11) and the idea that $g^*(\bar{\mathbf{x}})-u=(g^*(\bar{\mathbf{x}})-\hat{u}^*(\bar{\mathbf{x}}))+(\hat{u}^*(\bar{\mathbf{x}})-u)$.

The result (4.4c) can be established on the same lines. We first note that the factor (4.3) permits an expansion as

$$\frac{1}{2n}G_2\widehat{\Psi}_{20}=\frac{1}{2n}\bar{G}_2\Psi_{20}+\frac{1}{2n\sqrt{n}}\{G_3(I_p\otimes\Psi_{20})+\bar{G}_2\Psi_{21}\Theta_1\bar{G}_1\}\mathbf{z}_n+o_p\left(\frac{1}{n\sqrt{n}}\right),$$

and the corresponding factor for $\hat{u}(\bar{\mathbf{x}})$ has the same expansion with $\bar{G}_3$ for $G_3$. Thus we have

$$\begin{aligned}\hat{u}^*(\bar{\mathbf{x}})-u &= \frac{1}{\sqrt{n}}(I_q-\frac{1}{2n}\bar{G}_2\Psi_{21}\Theta_1)\bar{G}_1\mathbf{z}_n \\ &\quad+\frac{1}{2n}\bar{G}_2(\mathbf{z}_n^{<2>}-\Psi_{20})+\frac{1}{6n\sqrt{n}}\bar{G}_3(\mathbf{z}_n^{<3>}-3(\mathbf{z}_n\otimes\Psi_{20}))+o_p\left(\frac{1}{n\sqrt{n}}\right)\end{aligned} \qquad \ldots(4.12)$$

and

$$g^*(\bar{\mathbf{x}}) - \hat{u}^*(\bar{\mathbf{x}}) = \frac{1}{6n\sqrt{n}}(G_3 - \bar{G}_3)(\mathbf{z}_n^{<3>} - 3(\mathbf{z}_n \otimes \Psi_{20})) + o_p\left(\frac{1}{n\sqrt{n}}\right). \quad \ldots(4.13)$$

Therefore, using (2.3c), we have

$$E[g^*(\bar{\mathbf{x}}) - \hat{u}^*(\bar{\mathbf{x}}))(g^*(\bar{\mathbf{x}}) - \hat{u}^*(\bar{\mathbf{x}}))'] = \frac{1}{6n^3}(G_3 - \bar{G}_3)\Psi_{11}^{<3>}(G_3 - \bar{G}_3)' + o\left(\frac{1}{n^3}\right). \quad \ldots(4.14)$$

We can evaluate

$$\begin{aligned} & E\left[(g^*(\bar{\mathbf{x}}) - \hat{u}^*(\bar{\mathbf{x}}))\left(\frac{1}{2n}\bar{G}_2(\mathbf{z}_n^{<2>} - \Psi_{20})\right)'\right] \\ & \quad = \frac{1}{2n}E[(g^*(\bar{\mathbf{x}}) - \hat{u}^*(\bar{\mathbf{x}}))(\mathbf{z}_n'\Theta_1 \otimes \mathbf{z}_n')]B_2' + o\left(\frac{1}{n^3}\right) \\ & \qquad\qquad \text{(by (4.6), Lem. 3.3 and Prop. 2.2 ii))} \\ = \; & \frac{1}{2n}E[\mathbf{z}_n'\Theta_1 \otimes \{(g^*(\bar{\mathbf{x}}) - \hat{u}^*(\bar{\mathbf{x}}))\mathbf{z}_n'\}]B_2' + o\left(\frac{1}{n^3}\right) \\ = \; & \frac{1}{2n\sqrt{n}}\left\{\frac{d}{du'}E[g^*(\bar{\mathbf{x}}) - \hat{u}^*(\bar{\mathbf{x}})\mathbf{z}_n'] - E\left[\frac{d}{du'}(g^*(\bar{\mathbf{x}}) - \hat{u}^*(\bar{\mathbf{x}})\mathbf{z}_n'\right]\right\}B_2' + o\left(\frac{1}{n^3}\right) \\ & \qquad\qquad \text{(by Lem. 4.4)} \\ = \; & \frac{1}{12n^3}\frac{d}{du'}\{(G_3 - \bar{G}_3)E[(\mathbf{z}_n^{<3>} - 3(\mathbf{z}_n \otimes \Psi_{20}))\mathbf{z}_n']\}B_2' \\ & \quad + \frac{1}{2n}E[g^*(\bar{\mathbf{x}}) - \hat{u}^*(\bar{\mathbf{x}})]\frac{d\eta'}{du'}(I_q \otimes B_2') + o\left(\frac{1}{n^3}\right) \quad \text{(by (4.13))} \quad \ldots(4.15) \\ = \; & o\left(\frac{1}{n^3}\right) \quad \text{(by (2.3a) and (4.6))}. \end{aligned}$$

Finally, we calculate

$$\begin{aligned} & E\left[(g^*(\bar{\mathbf{x}}) - \hat{u}^*(\bar{\mathbf{x}}))\left(\frac{1}{6n\sqrt{n}}\bar{G}_3\mathbf{z}_n^{<3>}\right)'\right] \\ = \; & \frac{1}{36n^3}(G_3 - \bar{G}_3)E[(\mathbf{z}_n^{<3>} - 3(\mathbf{z}_n \otimes \Psi_{20}))\mathbf{z}_n^{<3>'}]\bar{G}_3 + o\left(\frac{1}{n^3}\right) \quad \text{(by (4.13))} \\ = \; & \frac{1}{6n^3}(G_3 - \bar{G}_3)\Psi_{11}^{<3>}\bar{G}_3' + o\left(\frac{1}{n^3}\right) \quad \text{(by (2.3b))}. \end{aligned} \quad \ldots(4.16)$$

Thus, use of (4.7), (4.12), (4.15) and (4.16) results in

$$E[(g^*(\bar{\mathbf{x}}) - \hat{u}^*(\bar{\mathbf{x}}))(\hat{u}^*(\bar{\mathbf{x}}) - u)'] = \frac{1}{6n^3}(G_3 - \bar{G}_3)\Psi_{11}^{<3>}\bar{G}_3' + o\left(\frac{1}{n^3}\right)$$

which, along with (4.14), establishes (4.4c). *Q.E.D.*

## 5. Final remarks

One of the feature of our paper is that the explicit form of the derivatives $\bar{G}_k$ is not used to evaluate the quadratic risk, unlike in Rao, Sinha and Subramanyam (1982). Of course the explicit form, using higher-order derivatives of $\psi(\theta)$ and $\theta(u)$, can be obtained but is very complicated especially for $k \geq 3$. Instead, we develop an important expression $\bar{G}_k = B_k(\Theta_1' \otimes I_{p^{k-1}})N_{p<k>}$ in Lemma 3.3, which is the only property of $\bar{G}_k$ that we need in order to establish the higher-order efficiency. The idea successfully reduces the computation.

In this paper we have not discussed validity of the asymptotic expansions. All the results here are based on "approximate estimators", obtained by Taylor expansions of estimators in $\mathcal{F}$. To confirm the results will necessitate appropriate conditions as in Fuller [(1976), Theorem 5.4.3], or replacement with a truncated loss function in Ghosh and Subramanyam (1974).

There are many third-order efficient estimators. For instance, Rao, Sinha and Subramanyam (1982) and Eguchi (1983) have formed some estimators asymptotically equivalent to the MLE up to order $n^{-2}$; estimators obtained by iterating the scoring method twice are third-order efficient as well [see e.g., Eguchi (1983)]. Thus, the results developed here are useful in distinguishing them. Moreover, we emphasize that this is the first step to establish a beautiful property that the MLE would enjoy, namely *absolute asymptotic efficiency*, i.e., the MLE is asymptotically efficient at any order, see Rao, Sinha and Subramanyam [(1982), Remark 3.1]. We hope many researchers will jump in the area to establish the absolute asymptotic efficiency.

## Appendix

In this section, we give proofs of Lemma 3.3 and Lemma 4.4

Proof of Lemma 3.3. Let

$$\Theta_{11}(u) = \frac{d^2}{dudu'}\theta(u)$$

Remember that

$$\bar{G}_k(\mathbf{x}) = \left(\frac{d}{d\bar{\mathbf{x}}'}\right)^{<k>} \hat{u}(\bar{\mathbf{x}}) \quad \text{and } \bar{G}_k(u) = \bar{G}_k(\Psi_{10}(\theta(u)))$$

with $\Psi_{10}(\theta(u)) = \eta(u)$ and that the MLE satisfies the equation

$$\Theta_1(\hat{u}(\bar{\mathbf{x}}))'\{\bar{\mathbf{x}} - \Psi_{10}(\theta(\hat{u}(\bar{\mathbf{x}}))\} = 0 \qquad \ldots(A.1)$$

(see (2.5)). For simplicity we write

$$\Theta_1,\ \Psi_{kl}, \bar{G}_k \quad \text{for } \Theta_1(u),\ \Psi_{kl}(\theta(u)),\ \bar{G}_k(u)$$

and

$$\Theta_1(\bar{\mathbf{x}}), \Theta_{11}(\bar{\mathbf{x}}), \Psi_{kl}(\bar{\mathbf{x}}) \quad \text{for } \Theta_1(\hat{u}(\bar{\mathbf{x}})),\ \Theta_{11}(\hat{u}(\bar{\mathbf{x}})),\ \Psi_{kl}(\theta(\hat{u}(\bar{\mathbf{x}})).$$

Further, we denote $i_u(\bar{\mathbf{x}}) = \Theta_1(\bar{\mathbf{x}})'\Psi_{11}(\bar{\mathbf{x}})\Theta_1(\bar{\mathbf{x}})$. Then (A.1) becomes

$$\Theta_1(\bar{\mathbf{x}})'(\bar{\mathbf{x}} - \Psi_{10}(\bar{\mathbf{x}})) = 0. \qquad \ldots(A.2)$$

Differentiation of (A.2) with respect to $\bar{\mathbf{x}}$ and application of Proposition 2.3 i), iii) give

$$\Theta_{11}(\bar{\mathbf{x}})'\{\bar{G}_1(\bar{\mathbf{x}}) \otimes (\bar{\mathbf{x}} - \Psi_{10}(\bar{\mathbf{x}}))\} + \Theta_1(\bar{\mathbf{x}})'\{I_p - \Psi_{11}(\bar{\mathbf{x}})\Theta_1(\bar{\mathbf{x}})\bar{G}_1(\bar{\mathbf{x}})\} = 0,$$

or

$$i_u(\bar{\mathbf{x}})\bar{G}_1(\bar{\mathbf{x}}) = A_0(\bar{\mathbf{x}})\bar{G}_1(\bar{\mathbf{x}}) + \Theta_1(\bar{\mathbf{x}})' \qquad \ldots(A.3)$$

with $A_0(\bar{\mathbf{x}}) = \Theta_{11}(\bar{\mathbf{x}})'\{I_q \otimes (\bar{\mathbf{x}} - \Psi_{10}(\bar{\mathbf{x}}))\}$. Note that

$$A_0(\Psi_{10}) = 0. \qquad \ldots(A.4)$$

Using Proposition 2.1 ii), we calculate

$$\begin{aligned}
\tfrac{d}{d\bar{\mathbf{x}}'}[i_u(\bar{\mathbf{x}})\bar{G}_1(\bar{\mathbf{x}})] &= \frac{di_u(\bar{\mathbf{x}})}{d\bar{\mathbf{x}}'}K_{pq}(\bar{G}_1(\bar{\mathbf{x}}) \otimes I_p)K_{pp} + i_u(\bar{\mathbf{x}})\bar{G}_2(\bar{\mathbf{x}}) \\
\tfrac{d}{d\bar{\mathbf{x}}'}[A_0(\bar{\mathbf{x}})\bar{G}_1(\bar{\mathbf{x}})] &= \frac{dA_0(\bar{\mathbf{x}})}{d\bar{\mathbf{x}}'}K_{pq}(\bar{G}_1(\bar{\mathbf{x}}) \otimes I_p)K_{pp} + A_0(\bar{\mathbf{x}})\bar{G}_2(\bar{\mathbf{x}}) \\
\tfrac{d}{d\bar{\mathbf{x}}'}\Theta_1(\bar{\mathbf{x}})' &= \Theta_{11}(\bar{\mathbf{x}})'(\bar{G}_1(\bar{\mathbf{x}}) \otimes I_p)
\end{aligned}$$

and hence, differentiation of both sides in (A.3), postmultiplication by $N_{p<2>}$, and use of (2.1) and Proposition 2.2 iv) results in

$$i_u(\bar{\mathbf{x}})\bar{G}_2(\bar{\mathbf{x}}) = A_0(\bar{\mathbf{x}})\bar{G}_2(\bar{\mathbf{x}}) + A_{21}(\bar{\mathbf{x}})(\bar{G}_1(\bar{\mathbf{x}}) \otimes I_p)N_{p<2>}$$

with $A_{21}(\bar{\mathbf{x}}) = \frac{dA_0(\bar{\mathbf{x}})}{d\bar{\mathbf{x}}'}K_{pq} + \Theta_{11}(\bar{\mathbf{x}})' - \frac{di_u(\bar{\mathbf{x}})}{d\bar{\mathbf{x}}'}K_{pq}$, being of order $q \times pq$.

We can show by induction that for any $\alpha \geq 2$,

$$i_u(\bar{\mathbf{x}})\bar{G}_\alpha(\bar{\mathbf{x}}) = A_0(\bar{\mathbf{x}})\bar{G}_\alpha(\bar{\mathbf{x}}) + \sum_{\beta=1}^{\alpha-1} A_{\alpha\beta}(\bar{\mathbf{x}})(\bar{G}_\beta(\bar{\mathbf{x}}) \otimes I_{p^{\alpha-\beta}})N_{p<\alpha>} \qquad \ldots(A.5)$$

for some $A_{\alpha\beta}(\bar{\mathbf{x}})$ of order $q \times qp^{\alpha-\beta}$. The proof is obtained by differentiating both sides of (A.5) and postmultiplying them by $N_{p<\alpha+1>}$. Here we just notice that

$$\frac{d}{d\bar{\mathbf{x}}'}[A_{\alpha\beta}(\bar{\mathbf{x}})(\bar{G}_\beta(\bar{\mathbf{x}})\otimes I_{p^{\alpha-\beta}})N_{p^{<\alpha>}}]$$

$$= \left(\frac{dA_{\alpha\beta}(\bar{\mathbf{x}})}{d\bar{\mathbf{x}}'}(I_p\otimes\bar{G}_\beta(\bar{\mathbf{x}})\otimes I_{p^{\alpha-\beta}})+A_{\alpha\beta}(\bar{\mathbf{x}})(\bar{G}_{\beta+1}(\bar{\mathbf{x}})\otimes I_{p^{\alpha-\beta}})\right)(I_p\otimes N_{p^{<\alpha>}})$$

$$= \left(\frac{dA_{\alpha\beta}(\bar{\mathbf{x}})}{d\bar{\mathbf{x}}'}K_{p,qp^{\alpha-\beta}}(\bar{G}_\beta(\bar{\mathbf{x}})\otimes I_{p^{\alpha+1-\beta}})K_{p^\alpha,p}\right.$$
$$\left.+A_{\alpha\beta}(\bar{\mathbf{x}})(\bar{G}_{\beta+1}(\bar{\mathbf{x}})\otimes I_{p^{\alpha-\beta}})\right)(I_p\otimes N_{p^{<\alpha>}})$$

and $(I_p\otimes N_{p^{<\alpha>}})N_{p^{<\alpha+1>}}=N_{p^{<\alpha+1>}}$ and $K_{p^\alpha,p}N_{p^{<\alpha+1>}}=N_{p^{<\alpha+1>}}$ [see Prop. 2.2 iv)]. Evaluating (A.5) at $\bar{\mathbf{x}}=\Psi_{10}$ and using (A.4) gives

$$i_u\bar{G}_\alpha=\sum_{\beta=1}^{\alpha-1}A_{\alpha\beta}(\bar{G}_\beta\otimes I_{p^{\alpha-\beta}})N_{p^{<\alpha>}} \qquad \ldots(A.6)$$

for $\alpha\geq 2$, where $A_{\alpha\beta}=A_{\alpha\beta}(\Psi_{10})$. Recall that $\bar{G}_1=i_u^{-1}\Theta_1'$ in (2.7), and hence we have from (A.6) with $\alpha=2$

$$\bar{G}_2=B_2(\Theta_1'\otimes I_p)N_{p^{<2>}},$$

where $B_2=i_u^{-1}A_{21}(i_u^{-1}\otimes I_p)$. Thus, (3.5) is seen to be true for $k=1,2$. Suppose that (3.5) holds true for $k=1,2,\ldots,\alpha$. Then, substitution of (3.5) into (A.6) with $\alpha+1$ for $\alpha$ gives

$$\begin{aligned}\bar{G}_{\alpha+1} &= i_u^{-1}\sum_{\beta=1}^{\alpha}A_{\alpha+1,\beta}\{(B_\beta(\Theta_1'\otimes I_{p^{\beta-1}})N_{p^{<\beta>}})\otimes I_{p^{\alpha+1-\beta}}\}N_{p^{<\alpha+1>}}\\ &= i_u^{-1}\sum_{\beta=1}^{\alpha}A_{\alpha+1,\beta}(B_\beta\otimes I_{p^{\alpha+1-\beta}})(\Theta_1'\otimes I_{p^{\beta-1}}\otimes I_{p^{\alpha+1-\beta}})(N_{p^{<\beta>}}\otimes I_{p^{\alpha+1-\beta}})N_{p^{<\alpha+1>}}\\ &= B_{\alpha+1}(\Theta_1'\otimes I_{p^\alpha})N_{p^{<\alpha+1>}}\end{aligned}$$

with $B_{\alpha+1}=i_u^{-1}\sum_{\beta=1}^{\alpha}A_{\alpha+1,\beta}(B_\beta\otimes I_{p^{\alpha+1-\beta}})$. This proves Lemma 3.3 by induction
*Q.E.D.*

Proof of Lemma 4.4. Use of (2.4) gives

$$E[\mathbf{z}_n'\Theta_1\otimes A(\mathbf{z}_n|u)]=\frac{1}{\sqrt{n}}E\left[\frac{dlogL(u|\bar{\mathbf{x}})}{du'}\otimes A(\mathbf{z}_n|u)\right]=\frac{1}{\sqrt{n}}\int\frac{dL(u|\bar{\mathbf{x}})}{du'}\otimes A(\mathbf{z}_n|u)\mu(dX).$$

Proposition 2.3 ii) helps differentiate

$$\frac{d}{du'}[L(u|\bar{\mathbf{x}})\otimes A(\mathbf{z}_n|u)]=\frac{dL(u|\bar{\mathbf{x}})}{du'}\otimes A(\mathbf{z}_n|u)+L(u|\bar{\mathbf{x}}).\frac{dA(\mathbf{z}_n|u)}{du'}.$$

These equalities complete the proof, provided that suitable exchangeability in order between differentiation and integration is permitted. *Q.E.D.*

ACKNOWLEDGMENT. The author would like to express his thanks to Professor M. Akahira of University of Tsukuba, Dr. M. Taniguchi of Osaka University, Dr. S. Eguchi of the Institute of Statistical Mathematics and Dr. Y. Takagi of University of Osaka Prefecture for stimulating discussions on higher-order efficiency.

## References

AKAHIRA, M. (1995). Third order efficiency implies fourth order efficiency : A resolution of the conjecture of J.K. Ghosh. To appear in *Ann. Inst. Statist. Math.*

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.

AMARI, S. *Differential-Geometrical Methods in Statistics*, Lecture Notes in Statistics, Vol. 28. Springer-Verlag, Berlin.

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.

BROWNE, M.W. (1974). Generalized least squares estimators in the analysis of covariance structures. *South Afr. Statist. J.*, **8**, 1-24.

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

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

– – – – (1985). Model fidelity of the maximum likelihood estimator in a curved exponential family. In *Statistical Theory and Data Analysis* (K. Matusita, ed.) 207-223. North-Holland, Amsterdam.

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

FULLER, W.A. (1976). *Introduction to Statistical Time Series*. Wiley, New York.

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

GHOSH, J.K. and SINHA, B.K. (1982). Third order efficiency of the MLE – A counterexample. *Calcutta Statist. Assoc. Bull.*, **31**, 151-158.

GHOSH, J.K., SINHA, B.K. and SUBRAMANYAM, K. (1979). Edgeworth expansions for Fisher-consistent estimators and second order efficiency. *Calcutta Statist. Assoc. Bull.*, **28**, 1-18.

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.

GHOSH, J.K. and SUBRAMANYAM, K. (1974). Second order efficiency of maximum likelihood estimates. *Sankhyā*, Ser. A, **36**, 325-358.

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.

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

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

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

MAGNUS, J.R. and NEUDECKER, H. (1979). The commutation matrix : Some properties and applications. *Ann. Statist.* **7**, 381-394.

– – –– (1985). Matrix differential calculus with applications to simple, Hadamard, and Kronecker products. *J. Math. Psychol.* **29**, 474-492.

– – –– (1988). *Matrix Differential Calculus with Applications in Statistics and Econometrics*. John Wiley, Chichester.

PFANZAGL, J. (1982). *Contributions to General Asymptotic Statistical Theory*, Lecture Notes in Statistics, Vol. 13, Springer-Verlag, New York.

PFANZAGL, J. and WEFELMEYER, W. (1978). A third order optimum property of the maximum likelihood estimator. *J. Multivariate Anal.*, **8**, 1-29.

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

– – –– (1962). Efficient estimates and optimum inference procedures in large samples. *J. Royal Statist. Soc.*, Ser. B, **24**, 46-72.

– – –– (1963). Criteria of estimation in large samples. *Sankhyā*, Ser. A, **25**, 189-206.

– – –– (1973). *Linear Statistical Inference and its Applications*, 2nd ed. Wiley, New York.

RAO, C.R., SINHA, B.K. and SUBRAMANYAM, K. (1982). Third order efficiency of the maximum likelihood estimator in the multinomial distribution. *Statistics & Decisions*, **1**, 1-16.

SATAKE, I. (1974). *Linear Algebra*, 2nd ed. (in Japanese). Shokabo, Tokyo.

STERNBERG, S. (1964). *Lectures on Differential Geometry*. Prentice-Hall Inc., Englewood Cliffs, NJ.

TAKEUCHI, K. and MORIMUNE, M. (1985). Third-order efficiency of the extended maximum likelihood estimators in a simultaneous equation system. *Econometrica*, **53**, 177-200.

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

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

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

TANIGUCHI, M. and WATANABE, Y. (1994). Statistical analysis of curved probability densities. *J. Multivariate Anal.*, **48**, 228-248.

INSTITUTE OF MATHEMATICS
UNIVERSITY OF TSUKUBA
TSUKUBA, IBARAKI 305
JAPAN