

FURTHER DISCUSSION ON SECOND-ORDER EFFICIENCY FOR ESTIMATION

S. EGUCHI*

This paper is concerned with parameter estimation in a curved exponential family. A further discussion is given in the class of second-order-efficient estimators for the third-order optimality. An adjusted risk function equivalent to the information loss in the sense of second-order is proposed, and the optimality structure of third-order asymptotical behaviors in terms of this risk function is discussed.

Key words: Curved exponential family, information loss, maximum likelihood estimator, minimum contrast estimator, second-order efficiency.

1. INTRODUCTION

Let $\wp$ be a curved exponential family of dimension m , that is, $\wp$ has an embedded form $\{\theta(\mathbf{u}): \mathbf{u} \in U\}$ by m -parameter $\mathbf{u}$ in the n -coordinate (canonical parameters) θ which expresses an exponential family with the density as

$$f(\mathbf{x}, \theta) = \exp\{\mathbf{x}\theta - \Psi(\theta)\}.$$

*S. Eguchi. Department of Mathematics. Faculty of Sciences. Shimane University. Matsue 690 Japan.

—Article presentat al Seventh International Conference on Multivariate Analysis, setembre 1992.

—Acceptat el juny de 1993.

In a problem of estimation for the parameteres $\mathbf{u}$ Fisher (1934) is the first that challenged to proving the second-order efficiency of the maximum likelihood estimator (*MLE*) in the sense that the *MLE* attains asymptotically the lower bound of information loss in reducing from a sample to an estimator. This theorem has been proceeded to the deep discussion by Rao (1961, 1962), Ghosh and Subramanyam (1974), Efron (1975), Amari (1982) and Eguchi (1983). In the story Efron shed a geometric light on the proof with the fairly complicated calculation. Amari (1982) extended this methodology to a case of a multi-parameter family with a standard terminology in differential geometry. He also originates a new geometry beyond the Riemannian geometry in a wider framework with broad applications to mathematical sciences including system theory, coding theory and neural network theory. Barndorff-Nielsen (1984) presented the observed geometry with a unified look at the likelihood and conditional principles and has advocated the p^* formula.

In this paper we consider a further discussion on second-order efficiency. A construction of a wide class of second-order efficient estimators by minimum contrast method was given in Eguchi (1983, 1985). In particular a one-parameter family of second-order efficient estimators is proposed and compared with the maximum likelihood estimator in the real data related with genetics. This fact leads to an explicit motivation to investigation of third-order asymptotics in the class of second-order efficient estimators. Thus we wish to present further optimality in the class of second-order efficient estimators, proposing an adjusted risk function equivalent to the information loss in the sense of second-order. As a main result we show the structure of third-order asymptotical behaviors in terms of this risk function.

2. MAIN RESULT

Let $\mathbf{x}_1, \dots, \mathbf{x}_n$ be a sample from a distribution in a curved exponential family as given in Introduction. We restrict an estimator $\hat{\mathbf{u}}$ of the parameter $\mathbf{u}$ to a function of the canonical statistic $\bar{\mathbf{x}} = \frac{1}{n} \sum \mathbf{x}_i$, say, $\hat{\mathbf{u}}(\bar{\mathbf{x}})$, with Fisher-consistency, or $\hat{\mathbf{u}}(\eta(\mathbf{u})) = \mathbf{u} (\forall \mathbf{u} \in U)$, where $\eta(\mathbf{u}) = E[\bar{\mathbf{x}}]$. Then we define the ancillary subfamily associated with $\hat{\mathbf{u}}$ as

$$A(\mathbf{u}) = \{f(\cdot, \theta) : \theta = \partial\Psi(\eta), \hat{\mathbf{u}}(\eta) = \mathbf{u}\}.$$

Thus we can take a local coordinates $(\mathbf{u}, \mathbf{v})$ for the enveloping exponential family such that $\mathbf{u}$ designates the submodel $\wp$ when $\mathbf{v} = \mathbf{0}$ and that $\mathbf{v}$ does the ancillary subfamily $A(\mathbf{u})$ since $A(\mathbf{u})$ is transverse to $\wp$ at the true density $f(\cdot, \theta(\mathbf{u}))$.

Accordingly we can obtain the ancillary component $\hat{v}$ satisfying $\eta(\hat{v}, \hat{u}) = \hat{x}$, or equivalent to $\hat{x} \in A(\hat{u})$.

The information loss from the sample to the estimator $\hat{u}$ is defined as $\Delta(\hat{u}, \mathbf{u}) = nI_{\mathbf{u}} - \hat{I}_{\mathbf{u}}$ with the information matrices $nI_{\mathbf{u}}$ and $\hat{I}_{\mathbf{u}}$ with the sample and $\hat{u}$, respectively. Assume that $\hat{u}$ is first-order efficient, or equivalently $\lim_{n \rightarrow \infty} \Delta(\hat{u}, \mathbf{u}) = 0$. Let $\{e^a: 1 \leq a \leq m\}$ and $\{e^\lambda: m+1 \leq \lambda \leq n\}$ be the tangent bases of $\wp$ and $A(\mathbf{u})$ induced to the total space $\{e_i: 1 \leq i \leq n\}$ with $e_i = \bar{x}_i - \eta_i(\mathbf{u})$. Then $\{e^\lambda\}$ is also a basis of the normal space of $\wp$ because of the first-order efficiency, or equivalently the orthogonality of $A(\mathbf{u})$ with $\wp$ that

$$e_a = B_a^i(\mathbf{u})e_i \quad \text{and} \quad e_\lambda = B_\lambda^i(\mathbf{u})e_i,$$

where $B_a^i(\mathbf{u}) = \partial\theta^i(\mathbf{u})/\partial u^a$ and $B_\lambda^i(\mathbf{u})$ is an orthogonal matrix satisfying

$$B_\lambda^i(\mathbf{u})g_{ij}(\theta(\mathbf{u}))B_a^i(\mathbf{u}) = 0 \quad (a = 1, \dots, m)$$

with respect to the information metric $G_{\theta(\mathbf{u})} = [g_{ij}(\theta(\mathbf{u}))]$ with $g_{ij}(\theta(\mathbf{u})) = \partial^2\Psi(\theta(\mathbf{u}))/\partial\theta^i\partial\theta^j$. The exponential connection $\mathbf{I}^{(e)}$ and mixture connection $\mathbf{I}^{(m)}$ have the natural parameter θ and the expectation parameter η , respectively, as affine parameters in the exponential family with relation to the transformation $\eta = \partial\Psi(\theta)/\partial\theta$. Hence the embedding curvatures $\tilde{H}^{(m)}$ and $\tilde{H}^{(e)}$ to the model $\wp$ with respect to $\mathbf{I}^{(e)}$ and $\mathbf{I}^{(m)}$ are written as componentwise

$$\tilde{H}_{ab\lambda}^{(m)} = B_\lambda^i(\mathbf{u})\partial_b B_{ai}(\mathbf{u}) \quad \text{and} \quad \tilde{H}_{ab\lambda}^{(e)} = B_{\lambda i}(\mathbf{u})\partial_b B_a^i(\mathbf{u})$$

where $\partial_a = \partial/\partial u^a$ and $B_{ai}(\mathbf{u}) = \partial_a \eta_i(\mathbf{u})$. Similarly the embedding curvatures $\hat{H}^{(m)}$ and $\hat{H}^{(e)}$ to $A(\mathbf{u})$ are

$$\hat{H}_{\lambda\mu b}^{(m)} = B_a^i(\mathbf{u})\partial_\mu B_{\lambda i}(\mathbf{u}) \quad \text{and} \quad \hat{H}_{\lambda\mu a}^{(e)} = B_{ai}(\mathbf{u})\partial_\mu B_\lambda^i(\mathbf{u})$$

where $\partial_\lambda = \partial/\partial v^\lambda$ and $B_{\lambda i}(\mathbf{u}) = B_\lambda^j(\mathbf{u})g_{ij}(\theta(\mathbf{u}))$. A characterization of second-order efficiency on the basis of the proposed risk function,

$$R(\hat{u}, \mathbf{u}) = E \{ {}^t[\hat{x} - \eta(\hat{u})]G_{\theta(\hat{u})}^{-1}[\hat{x} - \eta(\hat{u})] \}$$

is given in Eguchi (1984). The minimization of $R(\hat{u}, \mathbf{u})$ in an estimator $\hat{u}$ is shown to be equivalent to second-order efficiency. In a standard theory of linear regression $R(\hat{u}, \mathbf{u})$ is referred to as the residual sum of squares. Thus the risk measures the fitness of $\hat{u}$ into $\hat{x}$. Note that the risk function $R(\hat{u}, \mathbf{u})$ is decomposed into

$$R_{\text{tan}}(\hat{u}, \mathbf{u}) = E \{ {}^t[\hat{x} - \eta(\hat{u})]P(\mathbf{u})[\hat{x} - \eta(\hat{u})] \}$$

and

$$R_{\text{nor}}(\hat{\mathbf{u}}, \mathbf{u}) = \text{E} \{ {}^t[\hat{\mathbf{x}} - \eta(\hat{\mathbf{u}})] \mathcal{Q}(\mathbf{u}) [\hat{\mathbf{x}} - \eta(\hat{\mathbf{u}})] \},$$

where $P(\mathbf{u})$ and $\mathcal{Q}(\mathbf{u})$ denote the projection matrices onto the tangent space and the normal space of the subspace $\mathfrak{P}$ at $\theta(\mathbf{u})$, respectively. The characterization result follows that a first-order efficient estimator $\hat{\mathbf{u}}$ has the residual vectors expanded as

$$r_a = B_a^i(\eta(\hat{\mathbf{u}}) - \eta(\mathbf{u}))_i = e_a + \tilde{H}_{ab\lambda}^{(e)} e^\lambda e^b - \frac{1}{2} \hat{H}_{a\lambda\mu}^{(m)} e^\mu e^\lambda$$

and

$$r_\lambda = B_\lambda^i(\eta(\hat{\mathbf{u}}) - \eta(\mathbf{u}))_i = e_\lambda - \frac{1}{2} \tilde{H}_{ab\lambda}^{(m)} e^b e^\lambda.$$

This expression leads to

$$R_{\text{tan}}(\hat{\mathbf{u}}, \mathbf{u}) = \frac{1}{N^2} \left\| \tilde{H}^{(e)} \right\|^2 + \frac{1}{4N^2} \left\| \hat{H}^{(m)} \right\|^2$$

and

$$R_{\text{nor}}(\hat{\mathbf{u}}, \mathbf{u}) = \frac{1}{N}(n - m) + \frac{1}{N^2}(\tilde{H}^{(m)}, T) + \frac{1}{4N^2} \left\| \tilde{H}^{(m)} \right\|^2.$$

Hence the dependency of $\hat{\mathbf{u}}$ in the risk $R(\hat{\mathbf{u}}, \mathbf{u})$ is confined only to $\|\hat{H}^{(m)}\|^2/(4N^2)$ since the terms expressed by $\tilde{H}^{(e)}$ and $\tilde{H}^{(m)}$ depend only on the model $\mathfrak{P}$. Thus the minimization of $R(\hat{\mathbf{u}}, \mathbf{u})$ among first-order estimators is equivalent that the embedding mixture tensor $\tilde{H}^{(m)}$ vanishes over $\mathfrak{P}$, which implies the equivalence between the risk and the second-order efficiency. In the risk function the normal part $R_{\text{nor}}(\hat{\mathbf{u}}, \mathbf{u})$ measures the first-order efficiency with the lower band given as the codimension of $\mathfrak{P}$ with the order N^{-1} , while the tangent $R_{\text{tan}}(\hat{\mathbf{u}}, \mathbf{u})$ does second-order efficiency with the lower band $\tilde{H}^{(e)}$ with the order N^{-2} .

On the other hand, Rao (1963) discussed the usual mean squared error $\text{E} \|\hat{\mathbf{u}}, \mathbf{u}\|^2$ as optimality characterization of second-order efficiency. By bias correction of $\hat{\mathbf{u}}$ the second-order efficiency is interpreted as the minimization of the mean squared error among first-efficient estimators. The relation of $R(\hat{\mathbf{u}}, \mathbf{u})$ with $\text{E} \|\hat{\mathbf{u}}, \mathbf{u}\|^2$ is given as

$$\text{E} \|\hat{\mathbf{u}}^* - \mathbf{u}\|^2 = \frac{m}{N} + \frac{1}{2N^2} \left\| \tilde{I}^{(m)} \right\|^2 + R_{\text{tan}}(\hat{\mathbf{u}}, \mathbf{u}),$$

where $\hat{\mathbf{u}}^*$ is a bias-corrected version of $\hat{\mathbf{u}}$. Consequently, we obtain the equivalence of $R(\hat{\mathbf{u}}, \mathbf{u})$ with the mean squared error after bias-reduction since $R_{\text{nor}}(\hat{\mathbf{u}}, \mathbf{u})$ is independent of $\hat{\mathbf{u}}$.

We here return to an elementary case where the model $\wp$ is itself of exponential-type, that is, the order of a minimal sufficient statistic is the same with the dimension of the parameter in $\wp$, so that $n - m = 0$. Then we meet some discordance of the risk function with the information loss $\Delta(\hat{\mathbf{u}}, \mathbf{u})$ as follows:

$$\Delta(\hat{\mathbf{u}}, \mathbf{u}) = 0 \quad \text{but} \quad R(\hat{\mathbf{u}}, \mathbf{u}) = \frac{1}{N^2}(\tilde{H}^{(m)}, T) + \frac{1}{4N^2} \left\| \tilde{H}^{(m)} \right\|^2$$

since the exponential flatness of $\wp$ does not imply necessarily the mixture flatness.

For cancellation of this discordance we propose a new measures for risk

$$R^*(\hat{\mathbf{u}}, \mathbf{u}) = R_{\tan}(\hat{\mathbf{u}}, \mathbf{u}) + R_{\text{nor}}^*(\hat{\mathbf{u}}, \mathbf{u})$$

by modifying $R_{\text{nor}}(\hat{\mathbf{u}}, \mathbf{u})$ into

$$R_{\text{nor}}^*(\hat{\mathbf{u}}, \mathbf{u}) = E^t[\hat{\mathbf{x}} - \eta(\hat{\mathbf{u}})]^* Q(\mathbf{u})[\hat{\mathbf{x}} - \eta(\hat{\mathbf{u}})]^*,$$

where

$$[\hat{\mathbf{x}} - \eta(\hat{\mathbf{u}})]^* = \hat{\mathbf{x}} - \eta(\mathbf{u}) - G_{\theta(\mathbf{u})}^{-1}[\theta(\hat{\mathbf{u}}) - \theta(\mathbf{u})].$$

Note that if the transformation of θ into η is affine, *e.g.* in the case of normal regression cases, then $R^*(\hat{\mathbf{u}}, \mathbf{u})$ is reduced to as the original version $R(\hat{\mathbf{u}}, \mathbf{u})$. Thus we have directly the following relation

$$R_{\text{nor}}^*(\hat{\mathbf{u}}, \mathbf{u}) = \frac{1}{N}(n - m) + \frac{1}{N^2}(\tilde{H}^{(e)}, T) + \frac{1}{4N^2} \left\| \tilde{H}^{(e)} \right\|^2.$$

In accordance $R^*(\hat{\mathbf{u}}, \mathbf{u}) = 0$ if and only if $\Delta(\hat{\mathbf{u}}, \mathbf{u}) = 0$. We can rewrite the result in Eguchi (1984) on the modified risk:

Theorem 1

A Fisher-consistent estimator $\hat{\mathbf{u}}$ is second-order efficient if the risk function $R^*(\hat{\mathbf{u}}, \mathbf{u})$ attains the lower bound

$$\frac{1}{N}(n - m) + \frac{1}{N^2}(\tilde{H}^{(e)}, T) + \frac{5}{4N^2} \left\| \tilde{H}^{(e)} \right\|^2$$

to the second-order of N^{-1} .

We will investigate the higher-order expansion of estimator $\hat{\mathbf{u}}$ via the canonical statistic $\hat{\mathbf{x}}$ in the following section, where a main theorem will be shown.

Theorem 2

Let $\hat{\mathbf{u}}$ and $\mathbf{u}^\dagger$ be second-order efficient estimators. Then in the expansion to the third-order of $\hat{\mathbf{x}}, \hat{\mathbf{u}} - \mathbf{u}^\dagger$ vanishes except for the diagonal part of $e^\perp = (e^\lambda)$, that is,

$$(\hat{\mathbf{u}}, \mathbf{u}^\dagger)^a = \left(\hat{\boldsymbol{\Xi}} - \boldsymbol{\Xi}^\dagger \right)_{\lambda\mu\nu}^a e^\nu e^\mu e^\lambda.$$

Further,

$$R^*(\hat{\mathbf{u}}, \mathbf{u}) - R^*(\mathbf{u}^\dagger, \mathbf{u}) = \frac{1}{N^3} \left\| \hat{\boldsymbol{\Xi}} - \boldsymbol{\Xi}^\dagger \right\|^2 + \frac{2}{N^3} (\mathbf{h} - \hat{\xi}, \hat{\xi} - \xi^\dagger),$$

where

$$\mathbf{h} = (h^{a\lambda}) = (-\tilde{H}_{(b}^{(e)\lambda a} \tilde{T}_{cd}^b \tilde{g}^{dc}) \quad \text{and} \quad \xi = (\xi^{a\lambda}) = \left(3\boldsymbol{\Xi}_{\mu\nu\delta}^a \tilde{g}^{(\delta\nu} \tilde{g}^{\lambda m)} \right).$$

and the bracket () means symmetrization with respect to the enclosed indices.

In terms of Theorem 2 the following classification for third-order optimality of estimation in the sense of minimization for the risk R^* holds:

Theorem 3

- (i) In the case of codimension $n - m = 1$ a second-order efficient estimator $\mathbf{u}^\dagger$ is best in the class of second-order efficient estimators if and only if $\mathbf{u}^\dagger$ has the third-order term of $e^\perp$ with coefficients satisfying $\boldsymbol{\Xi}^\dagger = \mathbf{h}$.
- (ii) In the case of codimension $n - m \geq 2$ there exists generally no best estimator.

We can easily check that the third-order diagonal term of $e^\perp$ for the *MLE* vanishes, which directly yields the following theorem.

Theorem 4

The *MLE* is the best estimator if the skewness tensor $\tilde{T}$ induced to the model $\wp$ is free.

Proof follows from $\mathbf{h} = \xi = \mathbf{0}$ by applying Theorem 2.

The condition $\tilde{T} = \mathbf{0}$ for $\wp$ may be fairly limited but it holds for any nonlinear regression model with normal errors.

3. THE N^{-3} -ORDER INVESTIGATIONS

First we give the following expression to a formal expansion of $\hat{\mathbf{u}}$ via $\hat{\mathbf{x}}$:

$$(3.1) \quad (\hat{\mathbf{u}}, \mathbf{u})^a = e^a + \Delta^a + \hat{\Xi}^a + \hat{\Theta}^a$$

to the fourth-order. In particular writing

$$\hat{\Xi}^a = \Xi_{bcd}^a e^d e^c e^b + \Xi_{bc\lambda}^a e^\lambda e^c e^b + \Xi_{b\lambda\mu}^a e^\mu e^\lambda e^b + \Xi_{\lambda\mu\nu}^a e^\nu e^\mu e^\lambda,$$

we get

Lemma 1

In the third-order term Ξ^a ,

$$\begin{aligned} \Xi_{bcd}^a &= -\frac{1}{2} \tilde{\mathbf{I}}_{bf}^{(m) a} \tilde{\mathbf{I}}_{cd}^{(m) f} + \frac{1}{2} \tilde{H}_{b\lambda}^{(e) a} \tilde{H}_{cd}^{(m) \lambda} - \frac{1}{6} B^{ai} \partial_b \partial_c B_{di}, \\ \Xi_{bc\lambda}^a &= -\frac{3}{2} \tilde{\mathbf{I}}_{bf}^{(m) a} \tilde{H}_{cd}^{(e) f} + \tilde{H}_{b\mu}^{(e) a} \tilde{H}_{c\lambda}^{(m) \mu} - \frac{1}{2} B^{ai} \partial_b \partial_c B_{\lambda i}, \\ \Xi_{b\lambda\mu}^a &= -\tilde{H}_{c\mu}^{(e) a} \tilde{H}_{b\lambda}^{(e) c} \quad \text{and} \\ \Xi_{\lambda\mu\nu}^a &= -\frac{1}{6} B^{ai} \partial_\lambda \partial_\mu B_{\nu i}. \end{aligned}$$

Further, only the coefficients $(\Xi_{\lambda\mu\nu}^a)$ depend on the estimator $\hat{\mathbf{u}}$.

Proof

A Taylor expansion with respect to $(\mathbf{u}, \mathbf{v})$ gives

$$\begin{aligned} (\hat{\mathbf{x}} - \eta(\mathbf{u}))_i &= [\eta(\hat{\mathbf{u}}, \hat{\mathbf{v}}) - \eta(\mathbf{u})]_i = B_{ai} \bar{\mathbf{u}}^a + B_{\lambda i} \bar{\mathbf{v}}^\lambda + \\ &+ \frac{1}{2} \{ \partial_a B_{bi} \bar{\mathbf{u}}^b \bar{\mathbf{u}}^a + 2 \partial_\lambda B_{ai} \bar{\mathbf{u}}^a \bar{\mathbf{v}}^\lambda + \partial_\mu B_{\lambda i} \bar{\mathbf{v}}^\lambda \bar{\mathbf{v}}^\mu \} + \\ &+ \frac{1}{6} \partial_b \partial_c B_{di} \bar{\mathbf{u}}^d \bar{\mathbf{u}}^c \bar{\mathbf{u}}^b + \frac{1}{2} \partial_b \partial_c B_{\lambda i} \bar{\mathbf{v}}^\lambda \bar{\mathbf{u}}^c \bar{\mathbf{u}}^b + \\ &+ \frac{1}{2} \partial_b \partial_\mu B_{\lambda i} \bar{\mathbf{v}}^\lambda \bar{\mathbf{v}}^\mu \bar{\mathbf{u}}^b + \frac{1}{6} \partial_\nu \partial_\mu B_{\lambda i} \bar{\mathbf{v}}^\lambda \bar{\mathbf{v}}^\mu \bar{\mathbf{v}}^\nu. \end{aligned}$$

Thus from the above relation to the second-order we have

$$\bar{\mathbf{u}}^a = e^a + \Delta^a = e^a - \frac{1}{2} \tilde{\mathbf{I}}_{bc}^{(m) a} e^c e^b + \tilde{H}_{b\lambda}^{(e) a} e^\lambda e^b$$

and

$$\bar{v}^\lambda = e^\lambda + \Delta^\lambda = e^\lambda - \frac{1}{2} \tilde{H}_{cd}^{(m)\lambda} e^d e^c + \hat{H}_{c\lambda}^{(e)\lambda} e^\lambda e^c - \frac{1}{2} \tilde{F}_{\mu\gamma}^{(m)\lambda} e^\gamma e^\mu.$$

Next we proceed to the third-order as follows:

$$\begin{aligned}
(3.2) \ e^a &= \bar{u}^a + \frac{1}{2} \tilde{F}_{bc}^{(m)a} \bar{u}^c \bar{u}^b - \tilde{H}_{b\lambda}^{(e)a} \bar{u}^b \bar{v}^\lambda + \\
&+ \frac{1}{6} B^{ai} \partial_b \partial_c B_{di} \bar{u}^d \bar{u}^c \bar{u}^b + \frac{1}{2} B^{ai} \partial_b \partial_c B_{\lambda i} \bar{v}^\lambda \bar{u}^c \bar{u}^b + \\
&+ \frac{1}{2} B^{ai} \partial_b \partial_\mu B_{\lambda i} \bar{v}^\lambda \bar{v}^\mu \bar{u}^b + \frac{1}{6} B^{ai} \partial_v \partial_\mu B_{\lambda i} \bar{v}^\lambda \bar{v}^\mu \bar{v}^v = \\
&= \bar{u}^a + \frac{1}{2} \tilde{F}_{bc}^{(m)a} \left(e^c e^b + 2e^c \Delta^b \right) - \tilde{H}_{b\lambda}^{(m)a} \left(e^b e^\lambda + e^b \Delta^\lambda + \Delta^b e^\lambda \right) + \\
&+ \frac{1}{6} B^{ai} \partial_b \partial_c B_{di} e^c e^b + \frac{1}{2} B^{ai} \partial_b \partial_c B_{\lambda i} e^\lambda e^c e^b + \\
&+ \frac{1}{2} B^{ai} \partial_b \partial_\mu B_{\lambda i} e^\lambda e^\mu e^b + \frac{1}{6} B^{ai} \partial_v \partial_\mu B_{\lambda i} e^\lambda e^\mu e^v = \\
&= \bar{u}^a + \frac{1}{2} \tilde{F}_{bc}^{(m)a} e^c e^b - \tilde{H}_{b\lambda}^{(e)a} e^b e^\lambda - \\
&- \frac{1}{2} \tilde{F}_{bc}^{(m)a} \tilde{F}_{gh}^{(m)c} e^h e^g e^b + \tilde{F}_{bc}^{(m)a} \tilde{H}_{gm}^{(e)c} e^\mu e^g e^b + \\
&+ \frac{1}{2} \tilde{H}_{b\lambda}^{(e)a} \tilde{F}_{df}^{(m)b} e^f e^d e^\lambda - \tilde{H}_{b\lambda}^{(e)a} \tilde{H}_{d\gamma}^{(e)b} e^\gamma e^d e^\lambda + \\
&+ \frac{1}{2} \tilde{H}_{b\lambda}^{(e)a} \tilde{H}_{cd}^{(m)\lambda} e^d e^c e^b - \tilde{H}_{b\lambda}^{(e)a} \hat{H}_{c\mu}^{(e)\lambda} e^\mu e^c e^b + \frac{1}{2} \tilde{H}_{b\lambda}^{(e)a} \tilde{F}_{\mu\gamma}^{(m)\lambda} e^\gamma e^\mu e^b + \\
&+ \frac{1}{6} B^{ai} \partial_b \partial_c B_{di} e^d e^c e^b + \frac{1}{2} B^{ai} \partial_b \partial_c B_{\lambda i} e^\lambda e^c e^b + \\
&+ \frac{1}{2} B^{ai} \partial_b \partial_\mu B_{\lambda i} e^\lambda e^\mu e^b + \frac{1}{6} B^{ai} \partial_v \partial_\mu B_{\lambda i} e^\lambda e^\mu e^v = \\
&= \bar{u}^a + \frac{1}{2} \tilde{F}_{bc}^{(m)a} e^c e^b - \tilde{H}_{b\lambda}^{(e)a} e^b e^\lambda + \\
&+ \left(-\frac{1}{2} \tilde{F}_{bc}^{(m)a} \tilde{F}_{gh}^{(m)c} + \frac{1}{2} \tilde{H}_{b\lambda}^{(e)a} \tilde{H}_{cd}^{(m)\lambda} + \frac{1}{6} B^{ai} \partial_b \partial_c B_{di} \right) e^d e^c e^b + \\
&+ \left(\frac{3}{2} \tilde{F}_{bf}^{(m)a} \tilde{H}_{c\mu}^{(e)f} - \tilde{H}_{b\mu}^{(e)a} \hat{H}_{c\lambda}^{(e)\mu} + \frac{1}{2} B^{ai} \partial_b \partial_c B_{\lambda i} \right) e^\lambda e^c e^b + \\
&+ \left(-\tilde{H}_{c\lambda}^{(e)a} \tilde{H}_{b\mu}^{(e)c} + \frac{1}{2} \tilde{H}_{bv}^{(e)a} \tilde{F}_{\mu\lambda}^{(m)v} + \frac{1}{2} B^{ai} \partial_b \partial_\mu B_{\lambda i} \right) e^\lambda e^\mu e^b + \\
&+ \frac{1}{6} B^{ai} \partial_v \partial_\mu B_{\lambda i} e^\lambda e^\mu e^v.
\end{aligned}$$

From the second order efficiency of $\hat{\mathbf{u}}$, or $\hat{H}^{(m)} = 0$ it follows that

$$0 = \partial_b \hat{H}_{a\lambda\mu}^{(m)} = \partial_b B_a^i \partial_\lambda B_{\mu i} + B_a^i \partial_b \partial_\lambda B_{\mu i} = \tilde{H}_{ab\gamma}^{(e)} \hat{\Gamma}_{\lambda\mu}^{(m)\gamma} + B_a^i \partial_b \partial_\lambda B_{\mu i}$$

or

$$B_a^i \partial_b \partial_\lambda B_{\mu i} = -\tilde{H}_{ab\gamma}^{(e)} \hat{\Gamma}_{\lambda\mu}^{(m)\gamma}$$

or which substitution into the coefficients of $e^\mu e^\lambda e^b$ in (3.2) yields

$$\Xi_{b\lambda\mu}^a = -\tilde{H}_{c\mu}^{(e)a} \tilde{H}_{b\lambda}^{(e)c}.$$

We show the last statement of Lemma 1. It is clear that the coefficients (Ξ_{bcd}^a) and $(\Xi_{b\lambda\mu}^a)$ are independent of the choice of $\hat{\mathbf{u}}$, because they are expressed by the geometric quantities only on the model $\wp$. From two expressions

$$(3.3) \quad \partial_a \hat{H}_{bc\lambda}^{(e)} = -\partial_a (B_b^i \partial_c B_{\lambda i}) = \tilde{\Gamma}_{abd}^{(e)} \tilde{H}_{c\lambda}^{(e)d} + \tilde{H}_{ab\mu}^{(e)} \hat{H}_{c\lambda}^{(e)\mu} + B_b^i \partial_a \partial_c B_{\lambda i}$$

and

$$(3.4) \quad \partial_a \hat{H}_{bc\lambda}^{(e)} = \partial_a (B_{\lambda i} \partial_c B_d^i) = -\tilde{\Gamma}_{acb}^{(e)} \tilde{H}_{a\lambda}^{(e)d} - \tilde{H}_{cb\mu}^{(e)} \hat{H}_{a\lambda}^{(e)\mu} + B_{\lambda i} \partial_a \partial_c B_d^i$$

we have

$$\begin{aligned} \tilde{H}_{cb\mu}^{(e)} \hat{H}_{a\lambda}^{(e)\mu} &= -\tilde{\Gamma}_{acb}^{(e)} \tilde{H}_{a\lambda}^{(e)d} - \partial_a \hat{H}_{bc\lambda}^{(e)} + B_{\lambda i} \partial_a \partial_c B_d^i \\ B_b^i \partial_a \partial_c B_{\lambda i} &= -2\partial_a \hat{H}_{bc\lambda}^{(e)} + B_{\lambda i} \partial_a \partial_c B_b^i. \end{aligned}$$

This concludes that the coefficients $(\Xi_{bc\lambda}^a)$ are written only in terms of $\wp$.

We now prove Theorem 2 by the use of Lemma 1. It is a direct result that

$$(\hat{\mathbf{u}} - \mathbf{u})^a = (\hat{\Xi} - \Xi^\dagger)_{\lambda\mu\nu}^a e^\nu e^\mu e^\lambda.$$

By definition

$$\begin{aligned} R^*(\hat{\mathbf{u}}, \mathbf{u}) - R^*(\mathbf{u}^\dagger, \mathbf{u}) &= E^t[r(\hat{\mathbf{u}}) - r(\mathbf{u}^\dagger)]P(\mathbf{u})[r(\hat{\mathbf{u}}) - r(\mathbf{u}^\dagger)] \\ E^t[r^*(\hat{\mathbf{u}}) - r^*(\mathbf{u}^\dagger)]Q^*(\mathbf{u})[r^*(\hat{\mathbf{u}}) - r^*(\mathbf{u}^\dagger)] &+ 2E^t[e - r(\hat{\mathbf{u}})]P(\mathbf{u})[r(\hat{\mathbf{u}}) - r(\mathbf{u}^\dagger)] + \\ + 2E^t[e^* - r^*(\hat{\mathbf{u}})]Q^*(\mathbf{u})[r^*(\hat{\mathbf{u}}) - r^*(\mathbf{u}^\dagger)]. \end{aligned}$$

We expand $P(u)r^*(\hat{u})$ and $Q(u)r^*(\hat{u})$ by the formal expansion (3.1):

$$\begin{aligned}
\hat{r}^a &= B^{ai}(\eta(\hat{u}) - \eta(u))_i = \bar{u}^a + \frac{1}{2}\tilde{I}_{bc}^{(m)a} \bar{u}^c \bar{u}^b + \frac{1}{6}B^{ai}\partial_b\partial_c B_{di}\bar{u}^d \bar{u}^c \bar{u}^b + \\
&+ \frac{1}{24}B^{ai}\partial_b\partial_c\partial_d B_{fi}\bar{u}^f \bar{u}^d \bar{u}^c \bar{u}^b = \\
&= (e + \Delta + \hat{\Xi} + \hat{\Theta})^a + \frac{1}{2}\tilde{I}_{bc}^{(m)a} \left(e^c e^b + \Delta^c \Delta^b + 2e^c \Delta^b + e^c \hat{\Xi}^b \right) + \\
&+ \frac{1}{6}B^{ai}\partial_b\partial_c B_{di} \left(e^d e^c e^b + 3e^d e^c \Delta^b \right) + \frac{1}{24}B^{ai}\partial_b\partial_c\partial_d B_{fi} e^f e^d e^c e^b
\end{aligned}$$

and

$$\begin{aligned}
\hat{r}_\lambda^* &= B_{\lambda i}(\theta(\hat{u}) - \theta(u))^i = \frac{1}{2}\tilde{H}_{\lambda ab}^{(e)} \bar{u}^b \bar{u}^a + \frac{1}{6}B_{\lambda i}\partial_a\partial_b B_c^i \bar{u}^c \bar{u}^b \bar{u}^a \bar{u}^a + \\
&+ \frac{1}{24}B_{\lambda i}\partial_a\partial_b\partial_c B_d^i \bar{u}^d \bar{u}^c \bar{u}^b \bar{u}^a + \frac{1}{120}B_{\lambda i}\partial_a\partial_b\partial_c\partial_d B_f^i \bar{u}^f \bar{u}^d \bar{u}^c \bar{u}^b \bar{u}^a = \\
&= \frac{1}{2}\tilde{H}_{\lambda ab}^{(e)} \left(e^b e^a + \Delta^b + \Delta^a + 2e^b \Delta^a + 2e^b \Xi^a + 2e^b \Theta^a + 2\Delta^b \Xi^a \right) + \\
&+ \frac{1}{6}B_{\lambda i}\partial_a\partial_b B_c^i \left(e^c e^b e^a + 3e^c e^b \Delta^a + 3e^c e^b \Xi^a + 3e^c \Delta^b \Delta^a \right) + \\
&+ \frac{1}{24}B_{\lambda i}\partial_a\partial_b\partial_c B_d^i \left(e^d e^c e^b e^a + 4e^d e^c e^b \Delta^a \right) + \\
&+ \frac{1}{120}B_{\lambda i}\partial_a\partial_b\partial_c\partial_d B_f^i e^f e^d e^c e^b e^a.
\end{aligned}$$

Hence

$$\begin{aligned}
(\hat{r} - r)^a &= (\hat{\Xi} - \Xi^\dagger)^a + (\hat{\Theta} - \Theta^\dagger)^a + \tilde{I}_{bc}^{(m)a} e^c (\hat{\Xi} - \Xi^\dagger)^b, \\
(\hat{r}^* - r^*)_\lambda &= \tilde{H}_{\lambda ab}^{(e)} \left(e^b (\hat{\Xi} - \Xi^\dagger)^a + e^b (\hat{\Theta} - \Theta^\dagger)^a + \Delta^b (\hat{\Xi} - \Xi^\dagger)^a \right) + \\
&+ \frac{1}{2}B_{\lambda i}\partial_c\partial_b B_a^i e^a e^b (\hat{\Xi} - \Xi^\dagger)^c, \\
(e - \hat{r})^a &= \tilde{H}_{\lambda b}^{(e)a} e^b e^\lambda - \hat{\Xi}^a - \tilde{I}_{bc}^{(m)a} e^c e^b - \frac{1}{6}B_a^i \partial_b\partial_c B_{di} e^c e^b e^a
\end{aligned}$$

and

$$(e - \hat{r})^\lambda = e^\lambda - \tilde{H}_{cd}^{(m)\lambda} e^d e^c.$$

By this relation we obtain

$$\begin{aligned}
& \mathbb{E}^t[\mathbf{r}(\hat{\mathbf{u}}) - \mathbf{r}(\mathbf{u})^\dagger]P(\mathbf{u})[\mathbf{r}(\hat{\mathbf{u}}) - \mathbf{r}(\mathbf{u})^\dagger] = \\
& = \frac{15}{N^3}(\hat{\Xi} - \Xi^\dagger)_{\lambda\mu\nu}^a(\hat{\Xi} - \Xi^\dagger)_{\varphi\gamma\delta}^b g_{ab} g^{(\nu\mu} g^{\lambda\delta} g^{\gamma\varphi)}, \\
& \mathbb{E}^t[\mathbf{r}^*(\hat{\mathbf{u}}) - \mathbf{r}^*(\mathbf{u})]Q^*(\mathbf{u})[\mathbf{r}^*(\hat{\mathbf{u}}) - \mathbf{r}^*(\mathbf{u})] = O(N^{-4}), \\
& \mathbb{E}^t[\mathbf{e} - \mathbf{r}(\hat{\mathbf{u}})]P(\mathbf{u})[\mathbf{r}(\hat{\mathbf{u}}) - \mathbf{r}(\mathbf{u})^\dagger] = \mathbb{E}(\mathbf{e} - \mathbf{r})^a(\hat{\mathbf{r}}^* - \mathbf{r}^{\dagger*})_a = \\
& = \mathbb{E}\left((\hat{\Xi} - \Xi^\dagger)^a + (\hat{\Theta} - \Theta^\dagger)^a + \tilde{\mathbf{I}}_{bc}^{(m)a} e^c(\hat{\Xi} - \Xi^\dagger)^b\right) \cdot \\
& \cdot \left(-\tilde{H}_{\lambda b}^{(e)a} e^b e^\lambda - \hat{\Xi}^a - \tilde{\mathbf{I}}_{bc}^{(m)a} e^c e^b - \frac{1}{6} B_a^i \partial_b \partial B_c B_{di} e^c e^b e^a\right) = \\
& = \mathbb{E}\left\{\tilde{H}_{\lambda b}^{(e)a} e^b e^\lambda ((\hat{\Xi} - \Xi^\dagger)^a + (\hat{\Theta} - \Theta^\dagger)^a - \tilde{H}_{\lambda ba}^{(e)} \tilde{\mathbf{I}}_{dc}^{(m)a} e^c e^b e^\lambda - \right. \\
& \left. - \hat{\Xi}^a (\hat{\Xi} - \Xi^\dagger)_a - \tilde{\mathbf{I}}_{abc}^{(m)} e^c \Delta^b (\hat{\Xi} - \Xi^\dagger)^a - \frac{1}{6} B_a^i \partial_b \partial_c B_{di} e^d e^c e^b (\hat{\Xi} - \Xi^\dagger)^a\right\}
\end{aligned}$$

and

$$\begin{aligned}
& \mathbb{E}^t[\mathbf{e}^* - \mathbf{r}^*(\hat{\mathbf{u}})]Q^*(\mathbf{u})[\mathbf{r}^*(\hat{\mathbf{u}}) - \mathbf{r}^*(\mathbf{u})] = \mathbb{E}(\mathbf{e} - \hat{\mathbf{r}})^\lambda(\hat{\mathbf{r}}^* - \mathbf{r}^*)_\lambda = \\
& = \mathbb{E}\left\{\tilde{H}_{\lambda ab}^{(e)}\left(e^b(\hat{\Xi} - \Xi^\dagger)^a + e^b(\hat{\Theta} - \Theta^\dagger)^a\right) + \Delta^b(\hat{\Xi} - \Xi^\dagger)^a e^\lambda + \right. \\
& \left. + \frac{1}{2} B_{\lambda i} e^a e^b (\hat{\Xi} - \Xi^\dagger)^c e^\lambda - \frac{1}{2} \tilde{H}_{cd}^{(e)\lambda} \tilde{H}_{\lambda ab}^{(e)} e^b e^a (\hat{\Xi} - \Xi^\dagger)^c\right\}.
\end{aligned}$$

Hence the sum of the cross terms is given as

$$\begin{aligned}
& 2\mathbb{E}^t[\mathbf{e} - \mathbf{r}(\hat{\mathbf{u}})]P(\mathbf{u})[\mathbf{r}(\hat{\mathbf{u}}) - \mathbf{r}(\mathbf{u})] + 2\mathbb{E}^t[\mathbf{e}^* - \mathbf{r}^*(\hat{\mathbf{u}})]Q^*(\mathbf{u})[\mathbf{r}^*(\hat{\mathbf{u}}) - \mathbf{r}^*(\mathbf{u})^\dagger] \\
& = 2\mathbb{E}\left\{\left(\frac{1}{2} B_{\lambda i} \partial_c \partial_b B_a^i + \frac{5}{2} \tilde{H}_{ad\lambda}^{(e)} \tilde{\mathbf{I}}_{bc}^{(m)a}\right) e^a e^b (\hat{\Xi} - \Xi^\dagger)^c e^\lambda - \hat{\Xi}^a (\hat{\Xi} - \Xi^\dagger)^a\right\} \\
& = -6\tilde{H}_{ad\lambda}^{(e)} \tilde{T}_{bc}^d (\hat{\Xi} - \Xi^\dagger)_{\mu\nu\delta}^c g^{ab} g^{\delta\nu} g^{\lambda\mu} - 30\hat{\Xi}_{a\lambda\varphi\gamma} (\hat{\Xi} - \Xi^\dagger)_{\mu\nu\delta}^a g^{\delta\nu} g^{\mu\gamma} g^{\varphi\lambda}
\end{aligned}$$

since it follows from Lemma 1, (3.3) and (3.4) that

$$\begin{aligned}
& \frac{1}{2} B_{\lambda i} \partial_c \partial_b B_a^i + \frac{5}{2} \tilde{H}_{ad\lambda}^{(e)} \tilde{\mathbf{I}}_{bc}^{(m)a} - \hat{\Xi}_{a,bc\lambda} \\
& = \frac{1}{2} b_{\lambda i} \partial_c \partial_b B_a^i - \tilde{H}_{ad\lambda}^{(e)} \tilde{\mathbf{I}}_{bc}^{(m)a} - \tilde{H}_{b\mu}^{(e)a} \tilde{H}_{c\lambda}^{(m)\mu} + \frac{1}{2} B^{ai} \partial_b \partial_c B_{\lambda i} = \\
& = -\tilde{H}_{ad\lambda}^{(e)} \tilde{T}_{bc}^d,
\end{aligned}$$

which concludes the second result of Theorem 2.

REFERENCES

- [1] **Amari, S-I.** (1982). "Differential geometry of curved exponential families —curvatures and information loss". *Ann. Stat.*, **10**, 357–387.
- [2] **Amari, S-I.** (1982). "Differential geometrical Methods in statistics". *Lec. Notes in Statist.*, **28**. Springer-Verlag. Heidelberg.
- [3] **Barndorff-Nielsen, O.E.** (1986). "Likelihood and observed geometries". *Ann. Statist.*, **14**, 856–873.
- [4] **Efron, B.** (1975). "Defining the curvature of a statistical problem (with application to second order efficiency) (with discussion)". *Ann. Statist.*, **3**, 1189–1242.
- [5] **Eguchi, S.** (1983). "Second order efficiency of minimum contrast in a curved exponential family". *Ann. Statist.*, **11**, 793–803.
- [6] **Eguchi, S.** (1984). "A characterization of second order efficiency in a curved exponential family". *Ann. Inst. Statist. Math.*, **36 A**, 199–206.
- [7] **Eguchi, S.** (1985). "A differential geometric approach to statistical inference on the basis of contrast functionals". *Hiroshima Math. J.*, **15**, 341–391.
- [8] **Fisher, R.A.** (1925). "Theory of statistical estimation". *Proc. Cambridge Philos. Soc.*, **122**, 700–725.
- [9] **Ghosh, J.K. and Subramanyam, K.** (1974). "Second order efficiency of the maximum likelihood estimator". *Sankhyā*, **A 36**, 325–358.
- [10] **Rao, C.R.** (1961). "Asymptotic efficiency and limiting information". *Proc. Fifth. Berkely Symposium*, vol. **1**, 531–546.
- [11] **Rao, C.R.** (1962). "Efficient estimates and optimum inference procedures in large samples (with discussion)". *J. Roy. Statist. Soc.*, **B 24**, 46–72.