<table>
<thead>
<tr>
<th>Title</th>
<th>A CHARACTERIZATION OF SECOND ORDER EFFICIENCY FOR ESTIMATORS IN A CURVED EXPONENTIAL FAMILY</th>
</tr>
</thead>
<tbody>
<tr>
<td>Author(s)</td>
<td>Eguchi, Shinto</td>
</tr>
<tr>
<td>Citation</td>
<td>数理解析研究所講究録 1983年12月号 507: 1-15</td>
</tr>
<tr>
<td>Issue Date</td>
<td>1983-12</td>
</tr>
<tr>
<td>URL</td>
<td><a href="http://hdl.handle.net/2433/103763">http://hdl.handle.net/2433/103763</a></td>
</tr>
<tr>
<td>Type</td>
<td>Departmental Bulletin Paper</td>
</tr>
<tr>
<td>Textversion</td>
<td>publisher</td>
</tr>
</tbody>
</table>

Kyoto University
A CHARACTERIZATION OF SECOND ORDER EFFICIENCY
FOR ESTIMATORS IN A CURVED EXPONENTIAL FAMILY

By Shinto Eguchi (江口 真透)

Osaka University

Asymptotic properties of estimators are considered in an m-dimensional curved exponential family \( \mathcal{F} \) which is embedded in an exponential family of dimension n. It is shown that any first order efficient estimator is induced to a unique right triangle with sides \( \sqrt{m/2} \), \( \sqrt{(n-m)/2} \) and \( \sqrt{n/2} \). Let \( L(u) \) be the likelihood function of a sample of size N with respect to an m-component parameter \( u \) describing \( \mathcal{F} \). A necessary and sufficient condition for second order efficiency of an estimator \( \hat{u} \) is given by

\[
\lim_{N \to \infty} E \left( \frac{L(\hat{u})}{L(u)} \right)^N \geq 1
\]

for any first order efficient estimator \( \hat{u} \). The condition implies second order efficiency of the maximum likelihood estimator which is famous as Fisher-Rao's theorem.

AMS subject 80 60F, 62F.

Key words and phrases. almost-Metric structure, contrast function, curved exponential family, exponential family, Kullback-Leibler divergence, maximum likelihood estimator, second order efficiency, second fundamental tensor.
1. Introduction and main results. Let $\mathcal{F}$ be an $n$-dimensional exponential family of densities on the data-space $\mathbb{R}^n$ with respect to a carrier measure $\omega$. The family $\mathcal{F}$ is expressed as

$$
\{ f(x \mid \theta) \equiv e^{<x, \theta>-\psi(\theta)} : \theta \in \Theta \}
$$

by the natural co-ordinate system $\theta \equiv (\theta_1, \ldots, \theta^n)$ with the usual inner product $<\cdot, \cdot>$ of $\mathbb{R}^n$. The dual co-ordinates $\eta \equiv (\eta_1, \eta_2, \ldots, \eta_n)$ of $\mathcal{F}$ is defined by the transformation of $\theta$ into $\eta$:

$$
\eta(\theta) \equiv E^\theta x.
$$

Then the maximum likelihood estimator of $\eta$ or $\theta$ based on a sample $(x_1, x_2, \ldots, x_N)$ is given by

$$
\hat{x} \equiv \frac{1}{N} (x_1 + x_2 + \cdots + x_N)
$$

or $\hat{\theta} \equiv \theta[\hat{x}]$, respectively, where $\theta[\cdot]$ denotes the inverse transformation of $\eta[\cdot]$.

An $m$-dimensional curved exponential family is denoted by $\tilde{\mathcal{F}}$ $(m < n)$, i.e.,

$$
\tilde{\mathcal{F}} \equiv \{ f(x \mid \theta(u)) : u \in U \},
$$

where $U$ is an open set in $\mathbb{R}^m$ and the map $\theta(\cdot)$ from $U$ to $\Theta$ is nonlinear with the Jacobian matrix of rank $m$ on $U$. Let $(x_1, x_2, \ldots, x_n)$ be an i.i.d. sample from a density $f(\cdot \mid \theta(u))$. We may confine estimators of $u$ to the form of mappings of $\hat{x}$ or $\hat{\theta}$ since each of statistics $\hat{x}$ and $\hat{\theta}$ is minimal sufficient owing to the nonlinearity of $\theta(\cdot)$. Fisher-consistency of an estimator $\hat{u} = \hat{u}(\hat{\theta})$ is defined by
\( \hat{U}(\theta(u)) = u \)

for all \( u \) in \( U \). For an estimator \( \hat{U} \), \( \Delta_N(\hat{U}, u) \) denotes the difference between the information matrix of the sample and that of the estimator, which is called the information loss incurred by \( \hat{U} \). A Fisher-consistent estimator \( \hat{U} \) is said to be first order efficient if

\[
\lim_{N \to \infty} \frac{1}{N} \Delta_N(\hat{U}, u) = 0.
\]

Furthermore, a first order efficient estimator \( \hat{U} \) is said to be second order efficient if

\[
\lim_{N \to \infty} \left[ \Delta_N(\hat{U}, u) - \Delta_N(\hat{U}, u) \right] \geq 0
\]

for all first order efficient estimator \( \hat{U} \), where \( M \geq 0 \) denotes the non-negative definiteness of \( M \). The Kullback-Leibler divergence \( \rho_{KL}(f_1, f_2) \) between \( f_1 \) and \( f_2 \) in \( \mathcal{Y} \) is expressed as

\[
\rho_{KL}(\theta_1, \theta_2) = \log \left( \frac{\hat{f}_1(\theta_1)}{\hat{f}_2(\theta_2)} \right) - \psi(\theta_1) + \psi(\theta_2)
\]

with respect to \( \theta \), where \( \hat{f}_p = f(\cdot | \theta_p) \) with \( p = 1, 2 \).

The following theorems 1, 2 and 3 will be proved in Section 2.

**Theorem 1.** First order efficiency of a Fisher-consistent estimator \( \hat{U} \) is equivalent to each of the following conditions (i), (ii) and (iii);

(i) \( \lim_{N \to \infty} \mathbb{E} \left[ \rho_{KL}(\hat{U}, \theta(u)) - \rho_{KL}(\hat{U}, \theta(u)) \right] \geq 0 \)

for any Fisher-consistent estimator \( \hat{U} \).

(ii) \( \lim_{N \to \infty} \mathbb{E} \rho_{KL}(\theta(\hat{U}), \theta(u)) = m/2. \)
(iii) \( \lim_{N \to \infty} N \mathbb{E} \rho_{KL}(\hat{\theta}, \theta(u)) = (n-m)/2 \).

THEOREM 2 enables us to associate the common property of all first order efficient estimators with a right triangle, since the Kullback-Leibler divergence is the same order as the squared distance of \( F \) (see Figure).

\\

\text{\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\\n

The measure

\[
N \mathbb{E} [\rho_{KL}(\hat{\theta}, \theta(\hat{u})) - \rho_{KL}(\hat{\theta}, \theta(\hat{u}))]
\]

is closely related to the discrimination rate of \( \rho_{KL} \), introduced by Kuboki [5], in the model \( \mathbb{F} \), i.e., the case including the sufficient statistic \( \hat{\theta} \). However we, here, consider this as a criterion between estimators \( \hat{u} \) and \( \hat{v} \).

Let \( L(u) \) be the likelihood function based on the sample \( (x_1, \ldots, x_N) \). Since we have the relation

(1.1) \( \log L(u_1) - \log L(u_2) = N(\rho_{KL}(\hat{\theta}, \theta(u_2)) - \rho_{KL}(\hat{\theta}, \theta(u_1))) \)

for all \( u_1 \) and \( u_2 \) in \( U \), THEOREM 1 can be rewritten as

COROLLARY 1. A Fisher-consistent estimator \( \hat{u} \) is first order efficient if and only if

\[
\lim_{N \to \infty} \mathbb{E} [\frac{L(\hat{u})}{L(\hat{v})}] \geq 1
\]

for all Fisher-consistent estimators \( \hat{v} \).
Moreover we shall show

**THEOREM 2.** A first-order efficient estimator \( \hat{u} \) is second order efficient if and only if

\[
(1.2) \quad \lim_{N \to \infty} E N^2 \left[ \rho_{KL}(\hat{\theta}(\hat{u}), \theta(\hat{u})) - \rho_{KL}(\hat{\theta}, \theta(\hat{u})) \right] \geq 0
\]

for all first order efficient estimators \( \hat{u} \).

**THEOREM 2** does not hold if the relation (1.2) is replaced by

\[
\lim_{N \to \infty} E N^2 \left[ \rho_{KL}(\theta(\hat{u}), \theta(u)) - \rho_{KL}(\theta(\hat{u}), \theta(u)) \right] \geq 0.
\]

This phenomenon comes from the naming term by the parametrization, which may be similar to the discussion of the mean squared errors for estimators (c.f. Rao [8], Efron [3] and Amari [1]).

The relation (1.1) leads us directly to

**COROLLARY 2.** Second order efficiency of a first order efficient \( \hat{u} \) is equivalent to the condition:

\[
\lim_{N \to \infty} E \left[ \frac{L(\hat{u})}{L(\hat{u})} \right]^N \geq 1
\]

for all first order efficient estimators \( \hat{u} \).

By definition, the maximum likelihood estimator \( \hat{u}_{ML} \) satisfies

\[
\frac{L(\hat{u}_{ML})}{L(\hat{u})} \geq 1
\]

for any \( \hat{u} \) in \( U \) and any sample size \( N \). So **COROLLARY 2** implies promptly in second order efficiency of the maximum like likelihood estimator, which is famous for Fisher-Rao's theorem.
A contrast function \( \rho \) on \( \mathcal{F} \) is defined by satisfying the following conditions for any \( f_1 \) and \( f_2 \) in \( \mathcal{F} \):

\begin{align*}
(1) & \quad \rho(f_1, f_2) \geq 0 \\
(2) & \quad \rho(f_1, f_2) = 0 \iff f_1 = f_2 \text{ a.e. } \omega.
\end{align*}

Dawid-Amari's almost-metric structure is denoted by \( A \), and the almost-metric structure associated with \( \rho \) is denoted by \( A(\rho) \) (see Appendix I).

There may remain a question of whether this criterion

\[ E \rho_{KL}(\hat{\theta}, \theta(\cdot)) \]

is favorable only to the maximum likelihood estimator \( \hat{\theta}_{ML} \) since the minimum Kullback-Leibler divergence estimator is nothing but \( \hat{\theta}_{ML} \). However this question will vanish by

**Theorem 3.** Let \( \rho \) be a contrast function with the almost-metric structure \( A(\rho) \). The condition (1.2) for \( \rho \) in place of \( \rho_{KL} \) holds if \( A(\rho) = A \) on \( \mathcal{F} \).

2. Proofs of the results. We adopt the differential geometric formulation, due to Amari [1], including the almost-metric structure \( A = (g, \Gamma, \Gamma) \) over \( \mathcal{F} \).

For any \( f \) in \( \mathcal{F} \), let \( T_f(\mathcal{F}) \) be the tangent space of \( \mathcal{F} \) at \( f \). \( T_f(\mathcal{F}) \) is decomposed into the tangent and normal spaces of \( \mathcal{F} \) at \( f \) with respect to the information metric \( g \), i.e.,

\[ T_f(\mathcal{F}) = T_f(\mathcal{F}) + T_f^\perp(\mathcal{F}). \]

An \( n \times (n-m) \) matrix \( [B_{\lambda}^i(u)]_{i=1,2,\ldots,n} \) can be chosen to satisfy

\[ B_{\lambda}^i(u)g_{ij}(\theta(u))B_{\lambda}^j(u) = 0 \]

for \( \lambda = m+1, \ldots, n \).
for \( a = 1, 2, \ldots, m \), where \( B^i_a(u) = \vartheta^i(u)/\vartheta^a \). In the sequel we use the summation convention as in (2.1). With respect to co-ordinates \( \theta \) and \( u \), the bases of \( T_f(\mathcal{G}) \), \( T^*_f(\mathcal{G}) \), and \( T^*_f(\mathcal{G}) \) at \( f = f(\cdot | \theta(u)) \) are represented as

\[
\{ e^i_1(u) = \hat{x}^i_1 - \eta^i_1(u) \}_{i=1,2,\ldots,n} \\
\{ e^a_1(u) = B^i_a(u) e^i_1(u) \}_{a=1,2,\ldots,m}
\]

and

\[
\{ e_\lambda^i(u) = B^i_\lambda(u) e^i_1(u) \}_{\lambda=m+1,\ldots,n},
\]

respectively, where \( \eta(u) = \eta(\theta(u)) \). The induced components of \( g \) to \( T_f(\mathcal{G}) \) and \( T^*_f(\mathcal{G}) \) at \( f = f(\cdot | \theta(u)) \) are expressed as

\[
\gamma^a_{ab}(u) = B^i_a(u) \gamma^{ij}(\theta(u)) B^j_b(u)
\]

and

\[
\gamma^\lambda_\mu(u) = B^i_\lambda(u) \gamma^{ij}(\theta(u)) B^j_\mu(u),
\]

respectively.

The second fundamental tensors of \( \mathcal{G} \) with respect to \( \Gamma \) and \( \Gamma \) are denoted by \( H \) and \( H_e \), respectively. The components of \( H \) and \( H_e \) are expressed as

\[
m \begin{align*}
H^a_{ab\lambda}(u) &= B^i_\lambda(u) \vartheta^a[B^j_b(u) \gamma^{ij}(\theta(u))] \\
H^e_{ab\lambda}(u) &= B^i_\lambda(u) \gamma^{ij}(\theta(u)) \vartheta^a B^j_b(u)
\end{align*}
\]

with respect to \( u \) with \( \vartheta^a = \vartheta/\vartheta^a \). Henceforth we omit the arguments of the above geometric quantities at the true value \( u \) and freely raise or lower indices of them, e.g.,

\[
m \begin{align*}
H^a_{b\lambda} &= H^m_{cb\lambda}(u) \gamma^{ca}(u) \\
e^\lambda &= \gamma^\lambda_\mu(u) e^\mu(u),
\end{align*}
\]
where \( g^\alpha_\mu(u) \) and \( g^\lambda_\mu(u) \) are the inverse elements of \( \{ g^a_\alpha(u) \}_{a,c=1,2,\ldots,m} \) and \( \{ g^\lambda_\mu(u) \}_{\lambda,\mu=m+1,\ldots,n} \) respectively. For a first order efficient estimator \( \hat{u}(\theta) \), the set
\[
\{ f(\cdot | \theta); \hat{u}(\theta) = u \}
\]
is called the ancillary subspace of \( \hat{u} \) of which the second \( m \) fundamental tensor at \( f = f(\cdot | \theta(u)) \) with respect to \( \Gamma \) is denoted by \( \hat{H} \). Then we can rewrite THEOREM 7 in Amari [1] in the following convenient form:

**THEOREM A.** Let \( \hat{u} \) be a first order efficient estimator of \( u \). Then
\[
(2.2) \quad \hat{u}^a - u^a = e^a - \frac{1}{2} \hat{H}^a_{b\lambda} e^b e^\lambda - \frac{1}{2} \hat{H}^a_{\lambda\mu} e^\lambda e^\mu + O(||e||^3). 
\]
Furthermore the estimator \( \hat{u} \) is second order efficient if and only if the tensor \( \hat{H} \) vanishes on \( \hat{g} \).

In practice THEOREM A is also equivalent to THEOREM 1 (ii) by Ghosh and Subramanyan [7] in the case of one parameter. We set about, on the basis of THEOREM A and Appendix II,

**PROOF of THEOREM 1.** The Kullback-Leibler divergence \( \rho_{KL} \) can be also expressed as
\[
\rho_{KL}(\eta_1, \eta_2) = \langle \eta_1, \theta[\eta_2] - \theta[\eta_2] \rangle - \psi(\theta[\eta_1]) + \psi(\theta[\eta_2])
\]
with respect to \( \eta \). Let \( \hat{u} \) be a first order efficient estimator. Then the statistic \( \rho_{KL}(\hat{x}, \eta(\hat{u})) \) can be expanded as
\[
(2.3) \quad \rho_{KL}(\hat{x}, \eta(\hat{u})) = \frac{1}{2} e_1(\hat{u}) e_j(\hat{u}) g^{1j} + O(||e||^3),
\]
where \( g^{1j} \) is the inverse element of \( \{ g_{ij} \} \). It follows from THEOREM A that
\[
e_1(\hat{u}) = B_1 e^\lambda + O(||e||^2).
\]
Hence we have from Appendix II that
\[
\lim_{N \to \infty} N E \rho_{KL}(\hat{x}, \eta(\hat{u})) = g^\lambda_\mu g^\lambda_\mu = (n-m)/2
\]
The rest of the assertions are similar to the above and so
we complete the proof.

To prove THEOREM 2, we prepare

LEMMA 1. Let \( \hat{u} \) be a first order efficient estimator

with the fundamental tensor \( \hat{H} \). It holds that

\[
(2.4) \quad \lim_{N \to \infty} E \left[ N \left( e_1(\hat{u}) e_2(\hat{u}) g_{ij} - (n-m) \right) \right] = - \frac{1}{4} \left| \Gamma \right|^2 + \frac{1}{4} \left| \hat{H} \right|^2 + \frac{1}{4} \left| \hat{H} \right|^2 - 2(H,\hat{H}) - (H,T),
\]

where

\[
\left| \Gamma \right|^2 = \sum_{bc} \Gamma_{bc} \Gamma_{ef} \Gamma_{ad} \Gamma_{gf} \Gamma_{gh},
\]

\[
\left| \hat{H} \right|^2 = \sum_{b\lambda} \frac{1}{2} b_{b\lambda} \frac{1}{2} \lambda_{b\lambda} \frac{1}{4} \left( \frac{1}{2} \lambda_{b\lambda} \lambda_{b\lambda} - \frac{1}{4} \lambda_{b\lambda} \right).\]

\[
(H,T) = \sum_{\lambda \mu} \frac{1}{2} H_{\lambda \mu} \frac{1}{2} T_{\lambda \mu} \frac{1}{4} \left( \frac{1}{2} \mu_{\lambda \mu} \mu_{\lambda \mu} - \frac{1}{4} \mu_{\lambda \mu} \right),
\]

with the tensor \( T = \Gamma - \Gamma \).

PROOF. The statistic \( e_1(\hat{u}) \) is expanded as

\[
(2.5) \quad e_1(\hat{u}) = B_{\lambda 1} e^\lambda + B_{\lambda 1} \Delta^a - \frac{1}{2} a B_{\lambda 1} e^{a b} - \frac{1}{2} a B_{\lambda 1} e^a \Delta^b
\]

\[
- \frac{1}{6} \partial^a \partial^b B_{\lambda 1} e^{a b c} + O(\left| e \right|^4)
\]

by Taylor's theorem, where \( \Delta^a = e^a - \bar{u}^a \). It follows from

\[
(2.2)
\]

\[
\Delta^a = \frac{1}{2} \sum_{bc} B_{bc} - \left( b_{b\lambda} e^b \lambda - \frac{1}{2} \lambda_{b\lambda} e^b \lambda \right) + O(\left| e \right|^3).
\]

Substituting (2.6) into (2.5), we have

\[
(2.7) \quad e_j(u) e_j(u) g^{ij} = \sum_{\lambda \mu} e_{\lambda \mu} e^\lambda e^\mu + \sum_{ab} \Delta^a \Delta^b + \frac{1}{4} a B_{\lambda 1} g^{ij}
\]

\[
- \frac{1}{3} B \sum_{bc} \Delta^a \partial^b e^{a c e f} + H_{\lambda \mu} e^{a b} \lambda_{b\lambda} + O(\left| e \right|^5)
\]

\[
= \sum_{\lambda \mu} e_{\lambda \mu} e^\lambda e^{\mu} + \frac{1}{2} \sum_{b\lambda} \lambda_{b\lambda} e^b \lambda_{b\lambda} e^\lambda + H_{\lambda \mu} e^{a b} \lambda_{b\lambda} + O(\left| e \right|^5)
\]

\[
\times \partial^a \partial^b B_{1 \lambda} e^a e^b \lambda_{b\lambda} e^\lambda + \frac{1}{2} \lambda_{b\lambda} e^{a b} \lambda_{b\lambda} e^\lambda + H_{\lambda \mu} e^b \lambda_{b\lambda} e^\lambda + O(\left| e \right|^5)
\]

\[
\times \sum_{cd} c_{cd} e^{a b} e^a e^b + \frac{1}{2} \sum_{b\lambda} \lambda_{b\lambda} e^{a b} \lambda_{b\lambda} e^a e^b + H_{\lambda \mu} e^b \lambda_{b\lambda} e^\lambda + O(\left| e \right|^5).
\]
- \frac{m}{2} H_{\alpha \beta} e^\alpha e^\beta + \frac{m}{2} H_{\alpha \beta} H_{\mu \xi} e^\alpha e^\mu e^\xi + \frac{1}{3} B^i_{\lambda \gamma} a^i B_{\alpha \beta} e^\alpha e^\beta e^\lambda - \frac{m}{2} G_{\alpha \beta \gamma} e^\alpha e^\beta e^\gamma - \frac{m}{2} G_{\alpha \beta \gamma} e^\alpha e^\beta e^\gamma + \frac{m}{2} G_{\gamma \delta \lambda} e^\gamma e^\delta e^\lambda + O(||e||^5)

since

\epsilon_{1J}^i = B^i_{\alpha \beta} - \frac{1}{2} B^i_{\alpha \beta} + B^i_{\alpha \beta} X_{\lambda \mu} B^i_{\mu \nu}.

Hence from (2.7) and Appendix II,

\begin{align*}
E[\epsilon_{1J}(\hat{u}) \epsilon_{1J}(\hat{u})] &= \frac{n-m}{N} + \frac{1}{N^2} \left\{ - \frac{1}{4} ||\Gamma||^2 + ||H||^2 + \frac{1}{4} ||\hat{H}||^2 \right. \\
& \quad - 2(\hat{H},H) - (H,T) \left. \right\} + O(N^{-3}).
\end{align*}

This completes the proof.

Now we set about

**PROOF of THEOREM 2.** Let \hat{u} be a first order efficient estimator. Then the statistic \rho_{KL}(x, n(\hat{u})) is expanded as

(2.8) \quad \rho_{KL}(x, n(\hat{u})) = \frac{1}{2} \hat{\epsilon}_{1i}^i \hat{\epsilon}_{1j}^j - \frac{1}{2} \hat{\epsilon}_{1i}^i \hat{\epsilon}_{1j}^j e^i + \frac{1}{3} \hat{\epsilon}_{1i}^i \hat{\epsilon}_{1j}^j \hat{\epsilon}_{1k}^k + \frac{3}{4} S_{ijkl} \hat{\epsilon}_{1i}^i \hat{\epsilon}_{1j}^j \hat{\epsilon}_{1k}^k + \frac{1}{8} S_{ijkl} \hat{\epsilon}_{1i}^i \hat{\epsilon}_{1j}^j \hat{\epsilon}_{1k}^k \hat{\epsilon}_{1l}^l + O(||e||^5),

where S_{ijkl} \equiv \frac{3}{2} T_{ijkl} and \hat{\epsilon}_{1i} = \epsilon_{1i}(\hat{u}). By a similar argument as in the proof of LEMMA 1, we have from (2.8) that

(2.9) \quad \rho_{KL}(x, n(\hat{u})) = \frac{1}{2} \hat{\epsilon}_{1i}^i \hat{\epsilon}_{1j}^j - \frac{1}{2} T_{\lambda \mu \nu} e^\lambda e^\mu e^\nu \left. \right\} \\
- T_{\alpha \beta \gamma} e^\alpha e^\beta e^\gamma + \frac{1}{3} T_{\lambda \mu \nu} e^\lambda e^\mu e^\nu + T_{\lambda \mu} e^\lambda e^\mu \Delta\alpha \left. \right\} \\
- \frac{3}{4} S_{\lambda \mu \nu} e^\lambda e^\mu e^\nu + \frac{1}{8} S_{\lambda \mu \nu} e^\lambda e^\mu e^\nu e^\xi \left. \right\} + O(||e||^5)

Let \hat{\mathcal{H}} be the second fundamental tensor of the ancillary subspace of \hat{u}. It follows from (2.9) and LEMMA 1 that
(2.10) \[ E \rho_{KL}(\hat{\theta}, \theta(\hat{u})) = \frac{n-m}{N} + \frac{1}{N^2} \left( \frac{3}{8} ||\hat{H}||^2 + M \right) + O\left(\frac{1}{N^3}\right), \]

where

\[ M = -\frac{1}{8} ||\Gamma||^2 + \frac{1}{2} ||H||^2 - \frac{1}{2}(H,T) - (H,H) \]

\[ - \frac{1}{6} T_{\lambda \mu \nu} T^{\lambda \mu \nu} - \frac{1}{2} T_{\lambda \mu \nu} T^{\lambda \mu \nu} \]

\[ + \frac{15}{8} S_{\lambda \mu \nu} \xi \xi \mu \nu \xi \mu \nu + \frac{3}{4} S_{\lambda \mu \nu} \xi \xi \mu \nu \xi \mu \nu. \]

Clearly the term \( M \) in the RHS of (2.10) is independent of \( \hat{u} \) and dependent only on the model \( \hat{g} \). Hence it holds for a second order efficient estimator \( \hat{u} \) that

(2.11) \[ E[\rho_{KL}(\hat{\theta}, \theta(\hat{u}) - \rho(\hat{\theta}, \theta(\hat{u}))]) = \frac{1}{N^2} \left( \frac{3}{8} ||\hat{H}||^2 + O\left(\frac{1}{N^3}\right) \right) \]

since the estimator \( \hat{u} \) has the vanishing second fundamental tensor on account of THEOREM A. Therefore the second order efficiency of \( \hat{u} \) implies the inequality (1.2). The inverse assertion is clear since \( ||\hat{H}|| = 0 \) implies \( \hat{H} = 0 \). This completes the proof.

Similarly we have

PROOF of THEOREM 3. The statistic \( \rho(\hat{\theta}, \theta(\hat{u})) \) is expanded as

(2.12) \[ \frac{1}{2} g^{(\rho)}_{ij}(\hat{\theta}) e^i e^j + \frac{1}{6} \varepsilon^{ij}_{\kappa} g^{(\rho)}_{ij} k_{\kappa} + \varepsilon^{ij}_{\kappa} g^{(\rho)}_{ij} k_{\kappa} + \varepsilon^{ij}_{\kappa} g^{(\rho)}_{ij} k_{\kappa} \]

\[ + D^{(\rho)}_{ijkl} e^i e^j e^k e^l + O(||e||^5), \]

where \( A(\rho) \equiv (g^{(\rho)}, \varepsilon^{ij}_{\kappa} g^{(\rho)}_{ij} k_{\kappa}) \) and

\[ D^{(\rho)}_{ijkl} = \frac{\partial}{\partial \eta_i \partial \eta_j \partial \eta_k \partial \eta_l} \rho(\theta[\eta], \theta(\eta)) \mid_{\eta=\eta(u)}. \]

If \( A(\rho) = A \) on \( \hat{g} \), the expansion (2.12) is equal to (2.8) on \( \rho_{KL} \) except the last term because of \( A(\rho_{KL}) = A \). This implies the condition (2.11) with \( \rho_{KL} \) replaced by \( \rho \), which completes the proof.
REMARK 1. From the first term of (2.12), it follows that Theorem 1 holds for \( \rho \) with \( \rho^{(p)} = g \) on \( \mathcal{Y} \) in place of \( \rho_{KL} \).

Let \( \rho \) be a contrast function on \( \mathcal{Y} \) with the almost-metric structure \( A(\rho) \). By a similar argument as in the proof of THEOREMS 2 and 3, we may conclude the relation

\[
\lim_{N \to \infty} N^2 E[\rho(\hat{\theta}, \hat{u}) - \min_{u \in U} \rho(\hat{\theta}, \theta(u))] = \frac{3}{2} || H(\rho) - \hat{H} ||^2
\]

for any first order efficient estimator \( \hat{u} \), where \( \hat{H} \) and \( H(\rho) \) denote the second fundamental tensors of the ancillary subspace of \( \hat{u} \) and the subspace

\[
\{ f(\cdot | \theta); \rho(\theta, \theta(u)) = \min_{u \in U} \rho(\theta, \theta(u')) \}
\]

at \( f = f(\cdot | \theta(u)) \), respectively, with respect to \( \Gamma \). Finally

\[
\lim_{N \to \infty} N^2 E[\hat{u}^a - u^a(\hat{u}^b - u^b)]
= \lim_{N \to \infty} N^2 E[\rho(\hat{\theta}, \theta(\hat{u})) - \min_{u \in U} \rho(\hat{\theta}, \theta(u))],
\]

where \( \hat{u}_\rho \) denotes the minimum contrast estimator based on \( \rho \).

Acknowledgements. I would express my hearty thanks to Professor M. Okamoto and Dr. Y. Toyooka of Osaka University.

Appendix I

Let \( \tau \) be a co-ordinate system (parameter vector) of \( \mathcal{Y} \) with the transformation \( \theta(\cdot) \) of \( \tau \) into \( \theta \). The log-likelihood \( \log f(x | \theta(\tau)) \) with respect to \( \tau \) is denoted by \( \ell(\tau) \). Then the almost-metric structure \( A = (g, \Gamma, \Gamma) \), introduced by Amari [1], is defined as the following components

\[
g_{ij}(\tau) = E[\frac{\partial \ell}{\partial \tau^i} \frac{\partial \ell}{\partial \tau^j}],
\]

\[
\Gamma_{ij | k}(\tau) = E[\frac{\partial^2 \ell}{\partial \tau^i \partial \tau^j} \frac{\partial \ell}{\partial \tau^k}] + E[\frac{\partial \ell}{\partial \tau^i} \frac{\partial \ell}{\partial \tau^j} \frac{\partial \ell}{\partial \tau^k}]
\]
and
\[ e^T_{ij}(\tau) = E[\frac{\partial^2 \lambda}{\partial \tau_i \partial \tau_j} \frac{\partial \lambda}{\partial \tau_k}], \]
with respect to \( \tau \) with \( \lambda = \lambda(\tau) \) (see also Dawid [2]).

On the other hand, the almost-metric structure associated with a contrast function \( \rho \) on \( \mathcal{S} \) is defined as the following components
\[ g_{ij}(\tau) = -\frac{\partial^2}{\partial \tau_i \partial \tau_j} \rho(\tau_1, \tau_2) |_{\tau_1=\tau_2=\tau}, \]
\[ \Gamma_{ij}(\rho)(\tau) = -\frac{\partial^3}{\partial \tau_i \partial \tau_j \partial \tau_k} \rho(\tau_1, \tau_2) |_{\tau_1=\tau_2=\tau} \]
and
\[ \Gamma_{ij}^{(\rho)}(\tau) = -\frac{\partial}{\partial \tau_i \partial \tau_j \partial \tau_k} \rho(\tau_1, \tau_2) |_{\tau_1=\tau_2=\tau}, \]
with \( \rho(\tau_1, \tau_2) = \rho(f(\cdot | \theta(\tau_1)), f(\cdot | \theta(\tau_2))) \) (c.f. Eguchi [4]).

Appendix II
It holds for any sample size \( N \) that
\[ E e^{a b} = \frac{1}{N} g^{a b}, \]
\[ E e^{\lambda \mu} = \frac{1}{N} g^{\lambda \mu}, \]
\[ E e^{a b} e^{\lambda} = \frac{1}{N^2} T^{a b \lambda}, \]
\[ E e^{a b} e^{\lambda} e^{\mu} = \frac{1}{N^3} g^{a b} g^{\lambda \mu} + \frac{1}{N^3} S^{a b \lambda \mu} \]
and
\[ E e^{a b} e^{c e} = \frac{1}{N^3} S^{a b c \lambda}, \]
where
\[ S^{a b \lambda \mu} = B^{a i b j} B^{k l} B^{m u} \partial_i \partial_j \partial_k \partial_l \psi \]
with \( \partial_i = \partial / \partial \theta^i \).
References


    (Forthcoming)


    Sankhyā A, 36, 325-358.

    Sankhyā A, 25, 189-206.
Fig. 1. The right triangle. Let \( \hat{u} \) be first order efficient estimator of \( u \). The triangle with sides \( \sqrt{N E \rho(\hat{\theta}, \theta(u))} \), \( \sqrt{N E \rho(\hat{\theta}, \hat{\theta}(\hat{u}))} \) and \( \sqrt{N E \rho(\hat{\theta}(\hat{u}), \theta(u))} \) converges to the right triangle with \( \sqrt{n/2} \), \( \sqrt{(n-m)/2} \) and \( \sqrt{m/n} \).