<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), 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 \( \tilde{\mathcal{F}} \) 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 \( \tilde{\mathcal{F}} \).
A necessary and sufficient condition for second order
efficiency of an estimator \( \hat{u} \) is given by

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

for any first order efficient estimator \( \tilde{u} \). 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-\( \tilde{\mathcal{F}} \)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

$$\{ f(x \mid \theta) = e^{\langle x, \theta \rangle - \psi(\theta)} : \theta \in \Theta \}$$

by the natural co-ordinate system $\theta \equiv (\theta_1, \ldots, \theta^n)$ with the usual inner product $\langle \cdot, \cdot \rangle$ 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 $\eta$:

$$\eta[\theta] = E_{\theta} x.$$

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

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

or $\hat{\theta} = \theta[\hat{x}]$, 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

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

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 \( \mathcal{F} \) is expressed as

\[
\rho_{KL}(\theta_1, \theta_2) = \eta(\theta_1), \theta_1 - \theta_2 > - \psi(\theta_1) + \psi(\theta_2)
\]

with respect to \( \theta \), where \( f_p = f(\cdot|\theta_p) \) 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) \( \lim_{N \to \infty} \mathbb{E} \left[ \rho_{KL}(\hat{\theta}, \theta(\hat{u})) - \rho_{KL}(\hat{\theta}, (\hat{u})) \right] \geq 0 \)

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

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

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 \( \mathcal{F} \) (see Figure).

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 \( \mathcal{F} \), i.e., the case including the sufficient statistic \( \hat{\theta} \). However we, here, consider this as a criterion between estimators \( \hat{u} \) and \( \hat{u}' \).

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

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

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

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

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

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

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

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

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

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

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:

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

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

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

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

for any \( \hat{v} \) 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 $\mathcal{F}$:

1. $\rho(f_1, f_2) \geq 0$
2. $\rho(f_1, f_2) = 0 \iff f_1 = f_2$ a.e. $\omega$.

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.,

$$T_f(\mathcal{F}) = T_f^t(\mathcal{F}) + T_f^n(\mathcal{F}).$$

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

satisfy

$$B^\lambda(u) g_{ij}(\theta(u)) B^\lambda(u) = 0$$

(2.1)
for $a = 1, 2, \ldots, m$, where $\mathcal{E}_a(u) = \partial \theta^i(u) / \partial u^a$. In the sequel we use the summation convention as in (2.1). With respect to co-ordinates $\theta$ and $u$, the bases of $T_f(\mathcal{X})$, $T_f(\mathcal{Y})$ and $T_f(\mathcal{Z})$ at $f = f(\cdot | \theta(u))$ are represented as

\[
\{ e_i(u) = \hat{x}_i - \eta_i(u) \}_{i=1,2,\ldots,n'} \\
\{ e_a(u) = B_a^i(u) \ e_i(u) \}_{a=1,2,\ldots,m}
\]

and

\[
\{ e_\lambda(u) = B_\lambda^i(u) \ e_i(u) \}_{\lambda=m+1,\ldots,n'}
\]

respectively, where $\eta(u) \equiv \eta[\theta(u)]$. The induced components of $g$ to $T_f(\mathcal{Y})$ and $T_f(\mathcal{Z})$ at $f = f(\cdot | \theta(u))$ are expressed as

\[
\mathcal{C}_{ab}(u) = B_a^i(u) \mathcal{C}_{ij}(\theta(u)) B_b^j(u)
\]

and

\[
\mathcal{C}_{\lambda\mu}(u) = B_\lambda^i(u) \mathcal{C}_{ij}(\theta(u)) B_\mu^j(u),
\]

respectively.

The second fundamental tensors of $\mathcal{Y}$ with respect to $e$ and $\mathcal{E}$ are denoted by $H$ and $\mathcal{H}$, respectively. The components of $\mathcal{H}$ and $\mathcal{H}$ are expressed as

\[
H_{ab\lambda}(u) = B_{\lambda}^i(u) \partial_a [B_b^j(u) \mathcal{C}_{ij}(\theta(u))] \\
\]

and

\[
H_{ab\lambda}(u) = B_{\lambda}^i(u) \mathcal{C}_{ij}(\theta(u)) \partial_a B_b^j(u)
\]

with respect to $u$ with $\partial_a \equiv \partial / \partial u^a$. 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.,

\[
H_{b\lambda} = H_{cb\lambda}(u) \mathcal{C}_{a}(u) \\
\]

and

\[
e^\lambda = \mathcal{C}^{\lambda\mu}(u) e_\mu(u),
\]
where \( g_{\alpha}^{\mu}(u) \) and \( g_{\mu}^{\lambda}(u) \) are the inverse elements of 
\( \{ g_{ac}^{(u)} \}_{a,c=1,2,\ldots,m} \) and \( \{ g_{\mu}^{\lambda}(u) \}_{\lambda, \mu=m+1, \ldots, n} \) 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 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

\begin{equation}
\hat{u}^a - u^a = e^a - \frac{1}{2} \Gamma_{bce} e^b e^c + \frac{1}{2} e^b e^\lambda e^\mu - \frac{1}{2} \hat{H}_{b\lambda} e^\lambda e^\mu + O(||e||^3).
\end{equation}

Furthermore the estimator \( \hat{u} \) is second order efficient if and only if the tensor \( \hat{H} \) vanishes on \( \partial \).

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

\begin{equation}
\rho_{KL}(\hat{x}, \eta(\hat{u})) = \frac{1}{2} e_i(\hat{u}) e_j(\hat{u}) g_{ij}^{\hat{u}} + O(||e||^3),
\end{equation}

where \( g_{ij}^{\hat{u}} \) is the inverse element of \( \{ g_{ij}^{\hat{u}} \} \). It follows from THEOREM A that

\[ e_i(\hat{u}) = B_{i\lambda} e^\lambda + O(||e||^2). \]

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

\[
\lim_{N \to \infty} E \left[ N \left( e_i(\hat{u})e_j(\hat{u})g^{ij} - (n-m) \right) \right] = -\frac{1}{4} ||\Gamma||^2 + \frac{e}{4} ||\hat{H}||^2 + \frac{1}{4} ||\hat{H}||^2 - 2(H,\hat{H}) - (H,T),
\]

where

\[
||\Gamma||^2 = \sum_{bc} \Gamma_{bc} \Gamma_{ef} \sum_{ad} g^{ae} g^{bd},
\]

\[
||\hat{H}||^2 = \sum_{a} \hat{H}_{a\mu} \hat{H}_{a\mu} \sum_{\nu\xi} g^{\nu\xi} g^{\mu\nu},
\]

\[
(H,T) = \sum_{a} \hat{H}_{a\mu} T_{d\mu} \sum_{\nu\xi} g^{\nu\xi} g^{\mu\nu},
\]

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

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

\[
e_i(\hat{u}) = B_{i\lambda} e^{\lambda} + B_{i\lambda} \Delta^\lambda - \frac{1}{2} \partial_{\lambda} B_{i\lambda} e^{\lambda} - \partial_{\lambda} B_{i\lambda} e^{\lambda} \Delta^\lambda - \frac{1}{6} \partial_{\lambda} \partial_{\mu} B_{i\lambda} e^{\lambda} e^{\mu} + O(||e||^4),
\]

by Taylor's theorem, where \( \Delta^\lambda = e^{\lambda} - \hat{u}^{\lambda} \). It follows from (2.2) that

\[
\Delta^\lambda = \frac{1}{2} \Gamma_{bc} e^{bc} - \hat{H}_{a\lambda} e^{\lambda} + \frac{1}{2} \hat{H}_{a\lambda} e^{\lambda} + O(||e||^3).
\]

Substituting (2.6) into (2.5), we have

\[
e_j(\hat{u})e_j(\hat{u})g^{ij} = \sum_{a} B_{a\lambda} e^{\lambda} e^{\mu} + \sum_{a} \partial_{\lambda} B_{a\lambda} e^{\lambda} \Delta^\lambda + \frac{1}{4} \partial_{\lambda} \partial_{\mu} B_{a\lambda} e^{\lambda} e^{\mu} e^{\nu} + 2H_{a\mu} e^{\lambda} e^{\mu} + O(||e||^5)
\]

\[
+ \sum_{a} \hat{H}_{a\lambda} e^{\lambda} e^{\mu} + \frac{1}{2} \sum_{a} \partial_{\lambda} \partial_{\mu} \hat{H}_{a\lambda} e^{\lambda} e^{\mu} e^{\nu} + \hat{H}_{a\mu} e^{\lambda} e^{\mu} + O(||e||^5)
\]

\[
+ \sum_{a} \hat{H}_{a\lambda} e^{\lambda} e^{\mu} + \frac{1}{4} \sum_{a} \hat{H}_{a\lambda} e^{\lambda} e^{\mu} e^{\nu} e^{\xi} + \hat{H}_{a\mu} e^{\lambda} e^{\mu} e^{\nu} + \hat{H}_{a\mu} e^{\lambda} e^{\mu}
\]

\[
\hat{H}_{a\mu} e^{\lambda} e^{\mu} + \hat{H}_{a\mu} e^{\lambda} e^{\mu} e^{\nu} e^{\xi} + \hat{H}_{a\mu} e^{\lambda} e^{\mu} e^{\nu} e^{\xi} + \hat{H}_{a\mu} e^{\lambda} e^{\mu} e^{\nu} e^{\xi} + \hat{H}_{a\mu} e^{\lambda} e^{\mu} e^{\nu} e^{\xi}.
\]
\[ -\frac{m}{2} H_{a\mu} H_{\lambda\xi} e^{\lambda} e^{\lambda} e^{\mu} e^{\xi} + \frac{m}{2} H_{a\lambda} H_{\mu\xi} e^{\lambda} e^{\mu} e^{\xi} + \frac{m}{3} B^i_{\lambda\alpha} B^i_{\beta} a_{\alpha\beta} e^{\lambda} e^{\xi} e^{\xi} - \frac{m}{2} \Gamma_{\alpha\beta\gamma} \Gamma_{\delta\epsilon\phi} e^{\alpha} e^{\beta} e^{\gamma} + \Gamma_{\alpha\beta\gamma} H_{d\lambda} e^{\alpha} e^{\beta} e^{\gamma} e^{\lambda} + O(||e||^5) \]

since

\[ g^{ij}_{\lambda} = B_i^a g_{ab} B_j^b + B_j^a g^{\lambda}_{\mu} B_\mu^j. \]

Hence from (2.7) and Appendix II,

\[ E[e_i(\hat{u}) e_j(\hat{u}) g^{ij}_{\lambda}] = \frac{n-m}{N^2} + \frac{1}{n} \left( \frac{m}{|H|^2} \right)^2 + \frac{e}{N^2} \left( \frac{m}{|H|^2} \right)^2 - 2(H_{ij} - (H_{ij} T)) + O(N^{-2}). \]

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

\[ (2.8) \quad \rho_{KL}(x, n(\hat{u})) = \frac{1}{2} \hat{e}_i \hat{e}_j g^{ij}_{\lambda} - \frac{1}{2} \hat{T}^{ijk}_{\lambda} \hat{e}_i \hat{e}_j \hat{e}_k + \frac{1}{3} \hat{T}^{ijk}_{\lambda} \hat{e}_i \hat{e}_j \hat{e}_k + \frac{3}{4} \hat{s}^{ijkl} \hat{e}_i \hat{e}_j \hat{e}_k \hat{e}_l + \frac{1}{8} \hat{s}^{ijkl} \hat{e}_i \hat{e}_j \hat{e}_k \hat{e}_l + O(||e||^5), \]

where \( \hat{s}^{ijkl} \equiv \frac{3}{\eta_i} T^{ijkl} \) and \( \hat{e}_i \equiv e_i(\hat{u}) \). By a similar argument as in the proof of LEMMA 1, we have from (2.8) that

\[ (2.9) \quad \rho_{KL}(x, n(\hat{u})) = \frac{1}{2} \hat{e}_i \hat{e}_j g^{ij}_{\lambda} - \frac{1}{2} T_{\lambda \mu \nu} e^{\lambda} e^{\nu} e^{\mu} + T_{\lambda \mu \nu} e^{\lambda} e^{\mu} e^{\nu} + T_{\lambda \mu \nu} e^{\lambda} e^{\mu} e^{\nu} e^{\xi} + O(||e||^5) \]

Let \( \hat{H} \) be the second fundamental tensor of the ancillary subspace of \( \hat{u} \). It follows from (2.9) and LEMMA 1 that
(2.10) \( E \rho_{KL}(\hat{\theta}, \theta(\hat{u})) = \frac{n-m}{N} + \frac{1}{\sqrt{2}} \left( \frac{3}{5} \Vert \hat{H} \Vert^2 + M \right) + O\left( \frac{1}{\sqrt{N}} \right) \),

where

\[
M = -\frac{1}{6} \Vert \Gamma \Vert^2 + \frac{1}{2} \Vert \mathbf{e} \Vert^2 - \frac{1}{2} \langle H, T \rangle - \langle H, H \rangle - \frac{m}{2} \mathbf{T}^{\lambda \mu} T_{\lambda \mu} - \frac{1}{2} \mathbf{T}^{\lambda \mu} \mathbf{T}_{\lambda \mu} + \frac{15}{6} S_{\lambda \mu \nu} \mathbf{g}^{\lambda \mu} \mathbf{g}^{\nu \xi} + \frac{3}{4} S_{\lambda \mu \nu} \tilde{g}^{ab} \mathbf{g}^{\lambda \mu}.
\]

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

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

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 \( \Vert \hat{H} \Vert = 0 \) 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) \[ \frac{1}{2} \mathbf{g}(\rho) \mathbf{g}(\rho) + \frac{1}{6} \mathbf{T}(\rho) \mathbf{T}(\rho) + \frac{1}{5} \mathbf{\Gamma}(\rho) \mathbf{\Gamma}(\rho) + \mathbf{D}(\rho) + O(\mathbf{e}^5), \]

where \( A(\rho) \equiv (g(\rho), \Gamma(\rho), \Gamma(\rho)) \) and

\[ D(\rho) = \frac{\partial}{\partial \eta_1 \partial \eta_j \partial \eta_k \partial \eta_l} \rho(\eta, \eta)|_{\eta=\eta(u)}. \]

If \( A(\rho) = A \) on \( \mathcal{Y} \), 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 \( g^{(\rho)} = g \) on \( \mathcal{U} \) in place of \( \rho_{KL} \).

Let \( \rho \) be a contrast function on \( \mathcal{U} \) 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
\[
\lim_{N \to \infty} N^2 \mathbb{E}[\rho(\hat{\theta}, \hat{u}) - \min_{u \in U} \rho(\hat{\theta}, \theta(u)) \mid H(\rho) - \hat{\mathcal{H}} |^2 \]
for any first order efficient estimator \( \hat{u} \), where \( \mathcal{H} \) and \( H(\rho) \) denote the second fundamental tensors of the ancillary subspace of \( \hat{u} \) and the subspace
\[
\{ f(\cdot \mid \theta) \mid \rho(\theta, \theta(u)) = \min_{u \in U} \rho(\theta, \theta(u')) \}
\]
at \( f = f(\cdot \mid \theta(u)) \), respectively, with respect to \( \Gamma \). Finally
\[
\lim_{N \to \infty} N^2 \mathbb{E}\left[(\hat{u}^a - u^a)_{\rho} \mathcal{G}_{ab}(\hat{u}^b - u^b)_{\rho}\right] = \lim_{N \to \infty} N^2 \mathbb{E}[\rho(\hat{\theta}, \theta(\hat{u})) - \min_{u \in U} \rho(\hat{\theta}, \theta(u))],
\]
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 \( \mathcal{U} \) with the transformation \( \theta(\cdot) \) of \( \tau \) into \( \theta \). The log-likelihood \( \log f(x \mid \theta(\tau)) \) 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

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

\[
\Gamma_{ij}^{kl}(\tau) = \mathbb{E}\left[\frac{\partial^2 \ell}{\partial \tau^i \partial \tau^j} \frac{\partial \ell}{\partial \tau^k}\right] + \mathbb{E}\left[\frac{\partial \ell}{\partial \tau^i} \frac{\partial \ell}{\partial \tau^j} \frac{\partial \ell}{\partial \tau^k}\right]
\]
and

\[ e^{\Gamma_{ij}|k}(\tau) = \mathbb{E}[\frac{\partial^2 \lambda}{\partial \tau^i \partial \tau^j} \frac{\partial \lambda}{\partial \tau^k}], \]

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 \( \mathcal{F} \) is defined as the following components

\[ g^{(\rho)}_{ij}(\tau) = -\frac{\partial^2}{\partial \tau^i \partial \tau^j} \rho(\tau_1, \tau_2)|_{\tau_1=\tau_2=\tau}, \]

\[ \Gamma^{(\rho)}_{ij}|k(\tau) = -\frac{\partial^3}{\partial \tau^i \partial \tau^j \partial \tau^k} \rho(\tau_1, \tau_2)|_{\tau_1=\tau_2=\tau} \]

and

\[ \star \Gamma^{(\rho)}_{ij}|k(\tau) = -\frac{\partial}{\partial \tau^i \partial \tau^j \partial \tau^k} \rho(\tau_2, \tau_1)|_{\tau_2=\tau_1=\tau}, \]

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

\[ \mathbb{E} e^{a} e^{b} = \frac{1}{N} g^{ab}, \]

\[ \mathbb{E} e^{\lambda} e^{\mu} = \frac{1}{N} g^{\lambda \mu}, \]

\[ \mathbb{E} e^{a} e^{b} e^{\lambda} = \frac{1}{N^2} T^{ab\lambda}, \]

\[ \mathbb{E} e^{a} e^{b} e^{\lambda} e^{\mu} = \frac{1}{N^3} g^{ab} g^{\lambda \mu} + \frac{1}{N^3} S^{ab\lambda \mu} \]

and

\[ \mathbb{E} e^{a} e^{b} e^{c} e^{\lambda} = \frac{1}{N^3} S^{abc \lambda}, \]

where

\[ S^{ab\lambda \mu} \equiv B^{ai} B^{bj} B^{\lambda k} B^{\mu l} \partial_i \partial_j \partial_k \partial_l \psi \]

with \( \partial_i = \partial / \partial \theta^i \).
References


Fig. 1. The right triangle. Let \( \hat{u} \) be first order efficient estimator of \( u \). The triangle with sides \( \sqrt{N E \rho(\hat{0}, \theta(u))} \), \( \sqrt{N E \rho(\hat{0}, \theta(u))} \) and \( \sqrt{N E \rho(\theta(u), \theta(u))} \) converges to the right triangle with \( \sqrt{n/2} \), \( \sqrt{(n-m)/2} \) and \( \sqrt{m/n} \).