The Annals of Statistics
1983, Vol. 11, No. 3, 793–803

SECOND ORDER EFFICIENCY OF MINIMUM CONTRAST ESTIMATORS IN A CURVED EXPONENTIAL FAMILY

BY SHINTO EGUCHI

Osaka University

This paper presents a sufficient condition for second order efficiency of an estimator. The condition is easily checked in the case of minimum contrast estimators. The α^*-minimum contrast estimator is defined and proved to be second order efficient for every α, $0 < \alpha < 1$. The Fisher scoring method is also considered in the light of second order efficiency. It is shown that a contrast function is associated with the second order tensor and the affine connection. This fact leads us to prove the above assertions in the differential geometric framework due to Amari.

1. Introduction. We consider an n-dimensional exponential family of densities

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

with respect to a dominating measure ω on the sample space $\mathbb{R}^n$, where $\langle \cdot\,, \cdot \rangle$ denotes the inner product in $\mathbb{R}^n$ and

$$\Theta = \{\theta \in \mathbb{R}^n;\ \int e^{\langle x,\theta\rangle} d\omega(x) < \infty\}.$$

A subfamily $\tilde{\mathscr{F}}^m$ of $\mathscr{F}^n$ $(m < n)$ is called an m-dimensional curved exponential family if there exists a nonlinear mapping $\theta(\cdot)$ of U into Θ with the Jacobian matrix of full rank over U such that $\tilde{\mathscr{F}}^m$ is locally expressed as

$$\{f(\cdot \mid \theta(u));\ u \in U\},$$

where U is an open subset of $\mathbb{R}^m$ (c.f. Efron [3]).

Let $(x_1, \cdots, x_N)$ be an i.i.d. sample with a density $f_u(\cdot) = f(\cdot \mid \theta(u))$. It follows from the non-linearity of $\theta(\cdot)$ that each of the statistics

$$\bar{x} = (x_1 + \cdots + x_N)/N$$

and $\bar{\theta} = (\nabla\psi)^{-1}(\bar{x})$ is minimal sufficient, where $\nabla = (\partial/\partial\theta^1, \cdots, \partial/\partial\theta^n)$. Therefore we may estimate the true value of u through $\bar{x}$ or $\bar{\theta}$. An estimator $\hat{u} = \hat{u}(\bar{\theta})$ is said to be Fisher-consistent if

$$\hat{u}(\theta(u)) = u$$

for all u in U. The information loss in reducing from the sample to the estimator $\hat{u}$ is defined as

$$\Delta^{(N)}(\hat{u}, u) = N\tilde{g}(u) - \hat{g}^{(N)}(u),$$

where $N\tilde{g}(u)$ and $\hat{g}^{(N)}(u)$ denote information matrices of the sample and the estimator $\hat{u}$, respectively. A Fisher-consistent estimator $\hat{u} = \hat{u}(\bar{\theta})$ is said to be first order efficient if

$$\lim_{N\to\infty} N^{-1}\Delta^{(N)}(\hat{u}, u) = 0.$$

Received September 1981; revised February 1983.

AMS 1980 *subject classifications.* 62F10, 62F12.

Key words and phrases. Affine connection, ancillary subspace of estimator, curvature, curved exponential family, Fisher consistency, Fisher information, Fisher scoring method, information loss, maximum likelihood estimator, minimum contrast estimator, Γ-transversality, searching curve of estimator, second order efficiency.

Further the second order efficiency of a first order efficient estimator $\hat{u}$ is defined by the property that

$$\lim_{N\to\infty}[\Delta^{(N)}(\tilde{u}, u) - \Delta^{(N)}(\hat{u}, u)] \geq 0$$

for any first order efficient estimator $\tilde{u}$, where "$A \geq 0$" denotes nonnegative definiteness of A.

Let us consider now the Fisher scoring method. The 2-step maximum likelihood estimator $\hat{u}_2 = \hat{u}_2(\bar{\theta})$ from an initial estimator $\hat{u}_0 = \hat{u}_0(\bar{\theta})$ is defined as

$$\hat{u}_2(\bar{\theta}) = S \circ S(\hat{u}_0(\bar{\theta})),$$

where $S(u)) = u + \tilde{g}^{-1}(u)\partial\ell(\bar{x} \mid \theta(u))$ with

$$\partial\ell(x \mid \theta(u)) = \left\{\frac{\partial}{\partial u^a} \log f(x \mid \theta(u))\right\}_{a=1,2,\cdots,m}.$$

The following theorem will be proved in Section 3.

THEOREM 1. *The 2-step maximum likelihood estimator* $\hat{u}_2 = \hat{u}_2(\bar{\theta})$ *from an initial estimator* $\hat{u}_0 = \hat{u}_0(\bar{\theta})$ *is second order efficient if the estimator* $\hat{u}_0$ *is Fisher-consistent.*

We next introduce a contrast function ρ over $\mathscr{F}^n \times \mathscr{F}^n$, which is defined by the conditions that

$$\rho(\theta_1, \theta_2) \geq 0$$

for all θ_1 and θ_2 in Θ and that $\rho(\theta_1, \theta_2) = 0$ is equivalent to $\theta_1 = \theta_2$ (see e.g. Pfanzagl [7]). We call $\hat{u}_\rho = \hat{u}_\rho(\bar{\theta})$ the minimum contrast estimator based on ρ if

$$\rho(\bar{\theta}, \theta(\hat{u}_\rho)) = \min_{u\in U}\rho(\bar{\theta}, \theta(u)).$$

By definition the estimator $\hat{u}_\rho$ is Fisher-consistent. A convex function $W: (0, \infty) \to \mathbb{R}$ with $W(1) = 0$ generates a function

$$\rho_W(\theta_1, \theta_2) = E_{\theta_1} W\left(\frac{f(X \mid \theta_2)}{f(X \mid \theta_1)}\right)$$

for all θ_1 and θ_2 in Θ, which becomes a contrast function by Jensen's inequality. We need the following assumption $(A_{p,q})$: $\rho_W(\theta_1, \theta_2)$ is p-times and q-times differentiable in θ_1 and θ_2, respectively, under the integral sign with respect to the dominating measure ω.

PROPOSITION 1. *Under* $(A_{1,1})$, *the minimum contrast estimator* $\hat{u}_{\rho_W}$ *based on* ρ_W *is first order efficient.*

THEOREM 2. *Under* $(A_{2,1})$, *the minimum contrast estimator* $\hat{u}_{\rho_W}$ *based on* ρ_W *is second order efficient if*

$$W'''(1) + 2\,W''(1) = 0, \tag{1.1}$$

where $W''(\cdot)$ *and* $W'''(\cdot)$ *denote the second and third order derivatives, respectively.*

Proofs of Proposition 1 and Theorem 2 will be given in Section 3.

Let us mention some examples of ρ_W.

(1) Kullback-Leibler:

$$\rho_{\mathrm{KL}}(\theta_1, \theta_2) = E_{\theta_1}\left\{-\log\frac{f(X \mid \theta_2)}{f(X \mid \theta_1)}\right\} = \langle \theta_1 - \theta_2, \nabla\psi(\theta_1)\rangle - \psi(\theta_1) + \psi(\theta_2).$$

(2) Jeffreys:

$$\rho_J(\theta_1, \theta_2) = \{\rho_{\mathrm{KL}}(\theta_1, \theta_2) + \rho_{\mathrm{KL}}(\theta_2, \theta_1)\}/2 = \tfrac{1}{2}\langle \theta_1 - \theta_2, \nabla\psi(\theta_1) - \nabla\psi(\theta_2)\rangle.$$

(3) Hellinger:

$$\rho_H(\theta_1, \theta_2) = 4E_{\theta_1}\left\{1 - \left[\frac{f(X\mid\theta_2)}{f(X\mid\theta_1)}\right]^{1/2}\right\} = 4\left[1 - \exp\left\{\psi\left(\frac{\theta_1+\theta_2}{2}\right) - \frac{\psi(\theta_1)+\psi(\theta_2)}{2}\right\}\right].$$

(4) α-Chernoff ($-1 < \alpha < 1$):

$$\rho_\alpha(\theta_1, \theta_2) = \frac{4}{1-\alpha^2} E_{\theta_1}\left\{1 - \left[\frac{f(X\mid\theta_2)}{f(X\mid\theta_1)}\right]^{(1+\alpha)/2}\right\}$$
$$= \frac{4}{1-\alpha^2}\left[1 - \exp\left\{\psi\left(\frac{1-\alpha}{2}\theta_1 + \frac{1+\alpha}{2}\theta_2\right) - \frac{1-\alpha}{2}\psi(\theta_1) - \frac{1+\alpha}{2}\psi(\theta_2)\right\}\right].$$

(5) α^*-contrast ($0 < \alpha < 1$):

$$\rho^*_\alpha(\theta_1, \theta_2) = \frac{1}{\alpha^2}\left\{\frac{1-\alpha}{2}\rho_\alpha(\theta_1, \theta_2) + (\alpha^2 - 1)\rho_H(\theta_1, \theta_2) + \frac{1+\alpha}{2}\rho_\alpha(\theta_2, \theta_1)\right\}.$$

The minimum contrast estimator based on ρ_{KL} is nothing but the maximum likelihood estimator. Estimators based on ρ_α and ρ^*_α will be called the α-minimum and the α^*-minimum contrast estimators, respectively. The α^*-minimum contrast estimator is first proposed here and satisfies the following corollary, which will be proved in Section 3.

COROLLARY 1. *The α^*-minimum contrast estimator is second order efficient for every α, $0 < \alpha < 1$.*

2. Differential geometric framework. Amari [1] considered a parametric family of distributions as a Riemannian manifold with the metric g whose components form the Fisher information matrix. The differential structure is associated with all re-parameterizations which are diffeomorphic to the original parameters. We adopt the framework due to Amari [1].

The metric g, the third order tensor T and the α-connections Γ^α for $\alpha \in [-1, 1]$ over $\mathscr{F}^n$ have the following components:

$$g_{ij}(\theta) = E_\theta\left[\frac{\partial \ell}{\partial\theta^i}\frac{\partial \ell}{\partial\theta^j}\right]\left(= \frac{\partial^2}{\partial\theta^i\partial\theta^j}\psi(\theta)\right),$$

$$T_{ijk}(\theta) = E_\theta\left[\frac{\partial \ell}{\partial\theta^i}\frac{\partial \ell}{\partial\theta^j}\frac{\partial \ell}{\partial\theta^k}\right]\left(= \frac{\partial^3}{\partial\theta^i\partial\theta^j\partial\theta^k}\psi(\theta)\right), \tag{2.1}$$

$$\overset{\alpha}{\Gamma}{}^i_{jk}(\theta) = g^{il}(\theta)\left\{\frac{1+\alpha}{2}T_{ljk}(\theta) + \frac{1-\alpha}{2}E_\theta\left[\frac{\partial \ell}{\partial\theta^l}\frac{\partial^2 \ell}{\partial\theta^j\partial\theta^k}\right]\right\}\left(= \frac{1-\alpha}{2}T^i_{jk}(\theta)\right),$$

respectively, for $i, j, k = 1, 2, \cdots, n$ with respect to the natural coordinate system (θ^i) of $\mathscr{F}^n$, where $\ell = \log f(x\mid\theta)$ and $g^{il}(\theta)$ is the inverse element of $g_{li}(\theta)$. The summation convention is used hereafter as in (2.1). The parameter $\eta = (\eta_i)$ of $\mathscr{F}^n$ defined by

$$\eta_i(\theta) = E_\theta x_i\left(= \frac{\partial}{\partial\theta^i}\psi(\theta)\right)$$

is called the dual coordinate. It is noted that the affine connections $\overset{1}{\Gamma}$ and $\overset{-1}{\Gamma}$ have vanishing components with respect to (θ^i) and (η_i), respectively. In Amari [1], the connections $\overset{1}{\Gamma}$ and $\overset{-1}{\Gamma}$ are referred to as the Efron and the mixture connections and denoted by $\overset{e}{\Gamma}$ and $\overset{m}{\Gamma}$, respectively (cf. Dawid [2]). We shall also use this notation in the following.

We define a symmetric tensor $g^{(\rho)}$ associated with a contrast function ρ by the components

$$g^{(\rho)}_{ij}(\theta) = -\frac{\partial}{\partial\theta^i_1}\frac{\partial}{\partial\theta^j_2}\rho(\theta_1, \theta_2)\big|_{\theta_1=\theta_2=\theta}$$

with respect to (θ^i). It approximately holds that

$$\rho(\theta_1, \theta_2) = [\theta_1^i - \theta_2^i] g_{ij}^{(\rho)}(\theta)[\theta_1^j - \theta_2^j]/2$$

for θ_1 and θ_2 in a small neighbourhood of θ in Θ. The tensor $g^{(\rho)}$ is said to be equivalent to the metric g over $\tilde{\mathscr{F}}^m$ if there exists a positive scalar function $\varepsilon(\theta)$ such that

$$g_{ij}^{(\rho)}(\theta(u)) = \varepsilon(\theta(u)) g_{ij}(\theta(u))$$

for all $u \in U$. In this case we normalize the contrast function ρ by

$$\tilde{\rho}(\theta_1, \theta_2) = \frac{1}{\varepsilon(\theta_1)} \rho(\theta_1, \theta_2)$$

to let $g^{(\tilde{\rho})}$ and g be identical over $\tilde{\mathscr{F}}^m$. By definition it holds that $\hat{u}_\rho(\theta) = \hat{u}_{\tilde{\rho}}(\theta)$ for any θ in Θ. The examples (1)–(5) in Section 1 are already normalized.

For a contrast function ρ with the tensor $g^{(\rho)}$ equivalent to g, we define an affine connection $\Gamma^{(\rho)}$ associated with ρ. The components of $\Gamma^{(\rho)}$ with respect to (θ^i) are

$$\Gamma_{jk}^{(\rho)i}(\theta) = g^{il}(\theta)\left[-\frac{\partial^2}{\partial\theta_1^k \partial\theta_1^j} \frac{\partial}{\partial\theta_2^l} \rho(\theta_1, \theta_2)|_{\theta_1=\theta_2=\theta} \right]. \tag{2.2}$$

We arbitrarily fix a coordinate $\tau = (\tau^{i'})$ of $\mathscr{F}^n$ with the coordinate transformation $\phi: \tau \to \theta$. Let $(B_{i'}^i(\tau))$ be the Jacobian matrix of the inverse transformation ϕ at τ. It follows from the identity of $g^{(\rho)}$ with g that the components of $\Gamma^{(\rho)}$ with respect to $(\tau^{i'})$ are

$$\begin{aligned} \Gamma_{j'k'}^{(\rho)i'}(\tau) &= g^{i'l'}(\tau)\left[-\frac{\partial^2}{\partial\tau_1^{k'} \partial\tau_1^{j'}} \frac{\partial}{\partial\tau_2^{l'}} \rho(\phi(\tau_1), \phi(\tau_2))|_{\tau_1=\tau_2=\tau} \right] \\ &= B_i^{i'}(\phi(\tau)) \left\{ \frac{\partial}{\partial\tau^{j'}} B_k^i(\phi(\tau)) + \Gamma_{jk}^{(\rho)i}(\phi(\tau)) B_{j'}^j(\phi(\tau)) B_{k'}^k(\phi(\tau)) \right\}, \end{aligned} \tag{2.3}$$

where $\{B_i^{i'}(\phi(\tau))\}$ and $g^{i'j'}(\tau)$ are the inverses of $\{B_{i'}^i(\tau)\}$ and

$$\{g_{i'j'}(\tau) = B_{i'}^i(\tau) g_{ij}(\phi(\tau)) B_{j'}^j(\tau)\},$$

respectively, with $\tau_p = \phi^{-1}(\theta_p)$ for $p = 1, 2$.

Therefore $\Gamma^{(\rho)}$ satisfies the transformation rule of affine connections (c.f. Kobayashi and Nomizu [6]).

The above geometric quantities g, T, $\overset{\alpha}{\Gamma}$, $g^{(\rho)}$ and $\Gamma^{(\rho)}$ on $\mathscr{F}^n$ can be induced to $\tilde{\mathscr{F}}^m$. The tangent space T_{f_0} of $\mathscr{F}^n$ at f_0 in $\mathscr{F}^n$ is decomposed into the direct sum

$$T_f = \tilde{T}_f + \tilde{T}_f^\perp$$

at every f in $\tilde{\mathscr{F}}^m$, where $\tilde{T}_f$ and $\tilde{T}_f^\perp$ are the tangent and the normal spaces of $\tilde{\mathscr{F}}^m$, respectively. The connecting tensor B: $T_f \to \tilde{T}_f$ at $f = f_u$ has the components

$$B_a^i(u) = \partial_a \theta^i(u), \quad a = 1, \cdots, m$$

with respect to (θ^i) and (u^a) where $\partial_a = \partial/\partial u^a$. We appropriately choose components $B_\lambda^i(u)$, $\lambda = m+1, \cdots, n$, of the connecting tensor $B^\perp: T_f \to \tilde{T}_f^\perp$, i.e.,

$$B_\lambda^i(u) g_{ij}(\theta(u)) B_a^j(u) = 0 \tag{2.4}$$

for $a = 1, \cdots, m$. For example, the metric g has induced components

$$\tilde{g}_{ab}(u) = B_a^i(u) g_{ij}(\theta(u)) B_b^j(u), \tag{2.5}$$

and

$$\tilde{g}_{\lambda\mu}(u) = B_\lambda^i(u) g_{ij}(\theta(u)) B_\mu^j(u) \tag{2.6}$$

on $\tilde{T}_f \times \tilde{T}_f$ and $\tilde{T}_f^\perp \times \tilde{T}_f^\perp$ with respect to the local coordinate (u^a), respectively, where $f = f_u$.

The induced connection $\overset{\tilde{\alpha}}{\Gamma}$ of $\overset{\alpha}{\Gamma}$ to $\tilde{\mathscr{F}}^m$ has components

$$\overset{\tilde{\alpha}}{\Gamma}{}^c_{ab}(u) = B^c_i(u)\{\partial_b B^i_a(u) + \Gamma^i_{jk}(\theta(u))B^k_a(u)\}, \tag{2.7}$$

where

$$B^c_j(u) = \tilde{g}^{cd}(u)B^i_d(u)g_{ij}(\theta(u)).$$

The second fundamental form $\overset{\alpha}{H}$ of $\tilde{\mathscr{F}}^m$ on $\tilde{T}_f \times \tilde{T}_f \times \tilde{T}_f^{\perp}$ with respect to $\overset{\alpha}{\Gamma}$ has components

$$\begin{aligned}\overset{\alpha}{H}_{ab\lambda}(u) = \partial_a B^i_b(u)B^j_\lambda(u)g_{ij}(\theta(u)) + \overset{\alpha}{\Gamma}{}^l_{ij}(\theta(u))g_{lk}(\theta(u)) \\ \times B^i_a(u)B^j_b(u)B^k_\lambda(u)\end{aligned} \tag{2.8}$$

with respect to (u^a). In Amari [1], $\overset{1}{H}$ is referred to as the Efron curvature tensor, which will be denoted by $\underset{e}{H}$.

For an estimator $\hat{u} = \hat{u}(\bar{\theta})$ the set

$$A = A(\hat{u}, u) = \{f(\cdot \mid \theta);\ \hat{u}(\theta) = u\}$$

is called the ancillary subspace of $\hat{u}$ at f_u. Henceforth we assume that the Jacobian matrix of $\hat{u}$ at θ is of full rank for each θ in Θ. Then $A(\hat{u}, u)$ is a submanifold of codimension m and transverse to $\tilde{\mathscr{F}}^m$ at $f = f_u$ (c.f. Hattori [5]). In other words it holds for every $f = f_u$ that

$$T_f = \tilde{T}_f + T_f(A),$$

where $T_f(A)$ denotes the tangent space of $A = A(\hat{u}, u)$ at f. This property of $\hat{u}$ is the Fisher consistency of $\hat{u}$. For the estimator $\hat{u} = \hat{u}(\bar{\theta})$, a C^∞-curve C: $(-\varepsilon, \varepsilon) \to A(\hat{u}, u)$ passing through f_u at $t = 0$ is called a searching curve of $\hat{u}$ (passing through f_u). Amari [1] proved in Theorem 6 that the first order efficiency of $\hat{u}$ means the orthogonality of $A(\hat{u}, u)$ to $\tilde{\mathscr{F}}^m$ at $f = f_u$, i.e.,

$$T_f(A) = \tilde{T}_f^{\perp}. \tag{2.9}$$

Let $(u^a, v^\lambda)_{a=1,\cdots m, \lambda=m+1,\cdots,n}$ be a local coordinate system of $\mathscr{F}^n$ around f_{u_0} such that the coordinates (u_0, v) and (u, v_0) represent $A(\hat{u}, u_0)$ and $\tilde{\mathscr{F}}^m$ for fixed u_0 and v_0, respectively. Existence of such a coordinate is guaranteed by the transversality of $A(\hat{u}, u)$ to $\tilde{\mathscr{F}}^m$. In the case of (2.9), the second fundamental form of A at $f = f_u$ on $T_f(A) \times T_f(A) \times T_f^{\perp}(A)$, i.e., $\tilde{T}_f^{\perp} \times \tilde{T}_f^{\perp} \times \tilde{T}_f$ with respect to $\overset{m}{\Gamma}$ is defined as

$$\overset{m}{H}_{\kappa\lambda a}(u) = B^i_a(u)g_{ij}(\theta(u))\ \{\partial_\lambda \hat{B}^j_\kappa(u, v_0) + \overset{m}{\Gamma}{}^j_{lk}(\theta(u))B^k(u)B^l_\lambda(u)\}, \tag{2.10}$$

where

$$\partial_\lambda \hat{B}^i_\kappa(u, v) = \frac{\partial^2}{\partial v^\lambda \partial v^\kappa}\theta^i(u, v)$$

with the coordinate transformation $\theta(u, v)$ of (u, v) into θ.

3. Theorems and proofs. We investigate asymptotic properties of the minimum contrast estimator based on ρ in terms of the geometry associated with ρ.

PROPOSITION 2. *A minimum contrast estimator $\hat{u}_\rho = \hat{u}_\rho(\bar{\theta})$ based on ρ is first order efficient if the tensor $g^{(\rho)}$ is equivalent to the metric g over $\tilde{\mathscr{F}}^m$.*

PROOF. Suppose that $g^{(\rho)}$ is equivalent to g over $\tilde{\mathscr{F}}^m$. Since $\theta(u)$ gives a local minimum of the contrast function ρ from $\theta[t]$ to the model $\tilde{\mathscr{F}}^m$, every searching curve C of $\hat{u}_\rho$ satisfies the system of equations

$$\frac{\partial}{\partial u^a}\rho(\theta[t], \theta(u)) = 0 \tag{3.1}$$

for $a = 1, 2, \cdots, m$, where C is expressed as the mapping $t \to \theta[t]$ with $\theta[0] = \theta(u)$. Differentiating (3.1) with respect to t, we have

$$\dot{\theta}^i[t]C_{ij}(\theta[t], \theta(u))B^j_a(u) = 0, \tag{3.2}$$

where $\dot{\theta}^i[t] = (d/dt)\theta^i[t]$ and

$$C_{ij}(\theta_1, \theta_2) = \frac{\partial}{\partial \theta_1^i}\frac{\partial}{\partial \theta_2^j}\rho(\theta_1, \theta_2). \tag{3.3}$$

It follows from the equivalence of $g^{(\rho)}$ to g over $\tilde{\mathscr{F}}^m$ that

$$\dot{\theta}^i[0]g_{ij}(\theta(u))B^j_a(u) = 0 \tag{3.4}$$

by substituting $t = 0$ in (3.2). The relation (3.4) for every searching curve means the orthogonality of $A(\hat{u}_\rho, u)$ to $\tilde{\mathscr{F}}^m$ at $f = f_u$, i.e., the first order efficiency of the estimator $\hat{u}_\rho$ from Theorem 6 of Amari [1]. The proof is completed.

This result leads to the proof of Proposition 1 in Section 1. Henceforth we write $C: \tau = \tau[t]$ if a curve C of $\mathscr{F}^n$ is expressed as the mapping $t \to \tau[t]$ with respect to the coordinate system (τ^i) of $\mathscr{F}^n$.

PROOF OF PROPOSITION 1. It follows from the assumption $(A_{1,1})$ that

$$g^{(\rho w)}_{ij}(\theta) = W''(1)g_{ij}(\theta)$$

with respect to (θ^i). This relation means the equivalence of $g^{(\rho w)}$ to the metric g, which completes the proof from Proposition 2.

Let Γ be an affine connection on $\mathscr{F}^n$. A first order efficient estimator $\hat{u} = \hat{u}(\bar{\theta})$ is said to be Γ-transversal to the model $\tilde{\mathscr{F}}^m$ if for every searching curve $C: \theta = \theta[t]$ of $\hat{u}$,

$$B^i_a(u)g_{ij}(\theta(u))\{\ddot{\theta}^j[0] + \Gamma^j_{lk}(\theta(u))\dot{\theta}^k[0]\dot{\theta}^l[0]\} = 0 \tag{3.5}$$

for $a = 1, 2, \cdots, m$, where $\theta[0] = \theta(u)$ and $\{\Gamma^j_{lk}(\theta)\}$ denote the components of Γ with respect to (θ^i). Let $\tau = (\tau^{i'})$ be local coordinates of $\mathscr{F}^n$, obtained from θ through the transformation ϕ^{-1}. Then the relation (3.5) can be expressed as

$$B^{i'}_a(u)g_{i'j'}(\tau(u))\{\ddot{\tau}^{j'}[0] + \Gamma^{j'}_{k'l'}(\tau(u))\dot{\tau}^{k'}[0]\dot{\tau}^{l'}[0]\} = 0, \tag{3.6}$$

with respect to $(\tau^{i'})$, where $\{B^{i'}_a(u)\}$, $\{g_{i'j'}(\tau(u))\}$, and $\{\Gamma^{j'}_{k'l'}(\tau(u))\}$ are components of B, g and Γ, respectively, with respect to $(\tau^{i'})$. In particular we have for $\Gamma = \overset{m}{\Gamma}$ over $\tilde{\mathscr{F}}^m$ that

$$B_{ai}(u)g^{ij}(\theta(u))\ddot{\eta}_j[0] = 0 \tag{3.7}$$

with respect to the dual coordinate (η_i) on account of the vanishing of $\overset{m}{\Gamma}$, where $\{B_{ai}(u)\}$ are the components of B with respect to (η_i).

PROPOSITION 3. *A minimum contrast estimator $\hat{u}_\rho$ based on ρ is $\Gamma^{(\rho)}$-transversal to the model $\tilde{\mathscr{F}}^m$ if the tensor $g^{(\rho)}$ is equivalent to the metric g over $\tilde{\mathscr{F}}^m$.*

PROOF. By a similar argument as in the proof of Proposition 2, it holds for every searching curve $C: \theta = \theta[t]$ of $\hat{u}_{\rho_W}$ with $\theta[0] = \theta(u)$ that

$$\frac{\partial}{\partial u^a}\rho_W(\theta[t], \theta(u)) = 0 \tag{3.8}$$

for $a = 1, 2, \cdots, m$. Twice differentiating (3.8) in t, we have

$$B^i_a(u)\{C_{ji}(\theta[t], \theta(u))\ddot{\theta}^j[t] + D_{kji}(\theta[t], \theta(u))\dot{\theta}^j[t]\dot{\theta}^k[k]\} = 0, \tag{3.9}$$

where we put

$$D_{kji}(\theta_1, \theta_2) = \frac{\partial^2}{\partial \theta_1^k \partial \theta_1^j}\frac{\partial}{\partial \theta_2^i}\rho(\theta_1, \theta_2),$$

whereas $\{C_{ji}(\theta_1, \theta_2)\}$ are defined in (3.3). Then the system of equations (3.9) reduces to the relations

$$B_a^i(u)g_{ji}(\theta(u))\{\ddot{\theta}^j[t] + \Gamma_{kl}^{(\rho)j}(\theta(u))\dot{\theta}^k[0]\dot{\theta}^l[0]\} = 0$$

at $t = 0$ from the equivalence of $g^{(\rho)}$ to g, where $\{\Gamma_{kl}^{(\rho)j}(\theta)\}$ are defined in (2.2). Hence the proof is completed.

THEOREM 3. *A first order efficient estimator $\hat{u} = \hat{u}(\bar{\theta})$ is second order efficient if the estimator $\hat{u}$ is $\overset{m}{\Gamma}$-transversal to the model $\tilde{\mathscr{F}}^m$.*

PROOF. Suppose that the estimator $\hat{u}$ is Γ-transversal to $\tilde{\mathscr{F}}^m$. It holds for each searching curve $C: \eta = \eta[t]$ with $\eta[0] = \eta(u)$ that

$$(3.10)\quad \begin{aligned} B_{ai}(u)g^{ij}(\theta(u))\{\eta_j[t] - \eta_j(u)\} &= B_{ai}(u)g^{ij}(\theta(u))\{\dot{\eta}_j[0]t + \tfrac{1}{2}\ddot{\eta}_j[0]t^2\} + O(t^3) \\ &= -\tfrac{1}{2}t^2 B_{ai}(u)g^{ij}(\theta(u))\Gamma_j^{kl}(\eta(u))\dot{\eta}_k[0]\dot{\eta}_l[0] + O(t^3) \end{aligned}$$

because of the relation (3.6) and the orthogonality of $A(\hat{u}, u)$ to $\tilde{\mathscr{F}}^m$ at f_u, where $\{\Gamma_j^{kl}(\eta)\}$ are components of Γ with respect to (η_i).

We can take a local coordinate (u^a, v^λ) $a = 1, \cdots, m$, $\lambda = m + 1, \cdots, n$ of $\mathscr{F}^n$ around $f = f_u$ which specifies $\tilde{\mathscr{F}}^m$ and $A(\hat{u}, u)$ by fixing (v_0^λ) and (u_0^a), respectively. Let $\eta(u, v)$ be the transformation of (u^a, v^λ) into η. It follows from the orthogonality of $A(\hat{u}, u)$ to $\tilde{\mathscr{F}}^m$ at f_u that

$$(3.11)\quad \frac{\partial \eta_i}{\partial v^\lambda}(u, v_0) = B_\lambda^j(u)g_{ji}(\theta(u))$$

for $\lambda = m + 1, \cdots, n$. Then the curve C is expressed as

$$\eta_i[t] = \eta_i(u, v[t])$$

by the coordinate (u^a, v^λ). We have from (3.10) that

$$(3.12)\quad \dot{\eta}_i[0]t = B_{i\lambda}(u)v^\lambda,$$

neglecting the second order terms or more, where $B_{i\lambda}(u) = g_{ij}(\theta(u))B_\lambda^j(u)$ and $v^\lambda = v^\lambda[t]$. Substitution of (3.12) into (3.10) yields that

$$B_a^i(u)\{\eta_i(u, v) - \eta_i(u)\} = -\tfrac{1}{2}\overset{m}{H}_{\kappa\lambda a}(u)(v^\kappa - v_0^\kappa)(v^\lambda - v_0^\lambda) + O(|v - v_0|^3),$$

where

$$\overset{m}{H}_{\kappa\lambda a}(u) = B_a^i(u)\Gamma_i^{kl}(\theta(u))B_{k\lambda}(u)B_{l\kappa}(u).$$

The statistic $\bar{x}$ can be expressed as $(\hat{u}, \hat{v})$ in the coordinate (u^a, v^λ) for a large sample size N because of the almost-sure convergence of $\bar{x}$ to $\eta(u)$. Then the score function

$$\bar{S}_a = \frac{\partial}{\partial u^a}\log f(\bar{x} \mid \theta(u))$$

is represented as

$$\begin{aligned} B_a^i(u)\{\eta_i(u, v) - \eta_i(u)\} &= \tilde{g}_{ab}(u)\bar{u}^b + \tfrac{1}{2}\overset{m}{\Gamma}_{abc}(u)\bar{u}^b\bar{u}^c \\ &\quad - \overset{e}{H}_{ab\kappa}(u)\bar{u}^b\bar{v}^\kappa - \overset{m}{H}_{a\lambda\kappa}(u)\bar{v}^\kappa\bar{v}^\lambda + O(|(\bar{u}, \bar{v})|^3), \end{aligned}$$

where $\bar{u} = \hat{u} - u$, $\bar{v} = \hat{v} - v_0$ and quantities $\{\overset{m}{\Gamma}_{abc}(u)\}$ and $\{\overset{e}{H}_{ab\kappa}(u)\}$ are defined in (2.7) and (2.8), respectively. The limiting distribution of $(\bar{u}, \bar{v})$ follows the n-variate Gaussian law with mean 0 and covariance matrix

$$\begin{pmatrix} \tilde{g}^{ab}(u) & 0 \\ 0 & \tilde{g}^{\lambda\kappa}(u) \end{pmatrix}_{\substack{a,b=1,2,\cdots m, \\ \kappa,\lambda=m+1,\cdots,n,}}$$

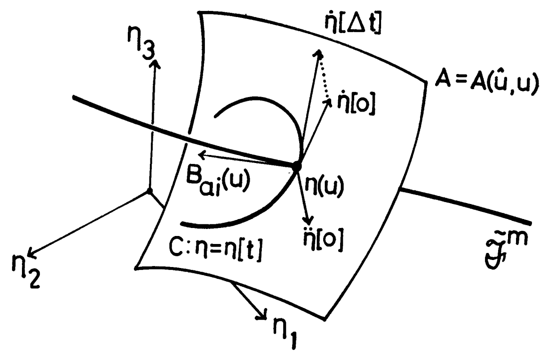

FIG. 1. *We consider the case of* $(n, m) = (3, 1)$. *In the dual coordinate system* (η_1, η_2, η_3), *both the velocity vector* $(\dot\eta_i[0])$ *and the acceleration vector* $(\ddot\eta_i[0])$ *of every searching curve* $C: \eta_i = \eta_i[t]$ *are orthogonal to the model* $\tilde{\mathscr{F}}^m$.

where $\{\tilde g^{ab}(u)\}$ and $\{\tilde g^{\kappa\lambda}(u)\}$ are the inverses of $\{\tilde g_{ba}(u)\}$ and $\{\tilde g_{\lambda\kappa}(u)\}$, respectively. Set

$$\hat S_a = \tfrac{1}{2}\overset{m}{H}_{a\lambda\kappa}(u)\bar v^\kappa \bar v^\lambda + H^l_{ab\kappa(u)}\bar u^b \bar v^\kappa.$$

Then it follows that

$$\lim_{n\to\infty}\Delta^{(n)}_{ab}(\hat u, u) = \lim_{n\to\infty} E \operatorname{cov}[\hat S_a, \hat S_b \mid \hat u = u]$$

by replacing $\bar S_a$ with $\hat S_a$. Hence the limiting information loss by $\hat u$ is decomposed into the sum of non-negative definite terms

$$\overset{e}{H}_{ac\kappa}(u)\overset{e}{H}_{bd\lambda}(u)\tilde g^{\kappa\lambda}(u)\tilde g^{cd}(u) + \overset{m}{H}_{\lambda\mu a}(u)\overset{m}{H}_{\kappa\nu b}(u)\tilde g^{\lambda\kappa}(u)\tilde g^{\mu\nu}(u),$$

which depend only on $\tilde{\mathscr{F}}^m$ and $A(\hat u, u)$, respectively. If the connection Γ coincides with $\overset{m}{\Gamma}$, the terms $\{\overset{m}{H}_{\kappa\lambda a}(u)\}$ vanishes over U. Therefore the $\overset{m}{\Gamma}$-transversality of $\hat u$ to $\tilde{\mathscr{F}}^m$ implies the second order efficiency of $\hat u$, which completes the proof.

Theorem 3 gives a sufficient condition for second order efficiency of estimators, which is an adaptation of Theorem 7 in Amari [1] to Γ-transversal estimators. Theorem 3 enables us to calculate limiting information losses of various estimators. $\overset{m}{\Gamma}$-transversality of estimators leads us to perceive the following dynamical interpretation (see Figure 1).

If the conditions

$$B_{ai}(u)g^{ij}(\theta(u))\dot\eta_j[0] = 0, \tag{3.13}$$

and

$$B_{ai}(u)g^{ij}(\theta(u))\ddot\eta_j[0] = 0 \tag{3.14}$$

hold for every searching curve $C: \eta = \eta[t]$ of a Fisher consistent estimator $\hat u$ with $\eta[0] = \eta(u)$, then the estimator $\hat u$ is second order efficient.

We now prove the statements in Section 1 by using Theorem 3. First the following lemma is well-known but necessary to prove Theorem 1. We denote by $\hat u\{\bar x\}$ an estimator expressed in terms of $\bar x$.

LEMMA 1. *The* 1-*step maximum likelihood estimator* $\hat u_1 = S(\hat u_0)$ *from any Fisher consistent estimator* $\hat u_0 = \hat u_0\{\bar x\}$ *is first order efficient.*

PROOF. By the definition of $\hat u_1$ it holds for each searching curve $C: \eta = \eta[t]$ of $\hat u$ with $\eta[0] = \eta(u)$ that

$$S^a(\eta[t], \hat u_0\{\eta[t]\}) = u^a \tag{3.15}$$

for any t, $-\varepsilon < t < \varepsilon$, with a small $\varepsilon > 0$, where

$$S^a(\eta, u) = u^a + B^{ai}(u)[\eta_i - \eta_i(u)]$$

with $B^{ai}(u) = \tilde{g}^{ab}(u)B^i_b(u)$. Differentiating (3.15) in t, we have

$$\dot{\eta}_j[t]D_0^{jb}(\eta[t])\partial_b B^{ai}(\hat{u}_0[t])(\eta_i[t] - \eta_i(u)) + B^{ai}(\hat{u}_0[t])\dot{\eta}_i[t] = 0 \tag{3.16}$$

because of the identity

$$B^{ai}(u)B_{bi}(u) = \delta^a_b \text{ (Kronecker delta)},$$

where we put

$$D_0^{ib}(\eta) = \frac{\partial}{\partial \eta_i}\hat{u}_0^b\{\eta\}$$

and $\hat{u}_0[t] = \hat{u}_0\{\eta[t]\}$. It follows from the Fisher-consistency of $\hat{u}_0$ that

$$B^{ai}(u)\dot{\eta}_i[0] = 0 \tag{3.17}$$

for $a = 1, \cdots, m$ by substituting $t = 0$ in (3.15). The relation (3.17) implies (3.13), which completes the proof through a similar argument as in the proof of Proposition 2.

From Lemma 1, the Jacobian matrix of $\hat{u}_1\{\eta\}$ satisfies

$$D_1^{ia}(\eta(u)) = B^{ai}(u) \tag{3.18}$$

for any $u \in U$.

PROOF OF THEOREM 1. Every searching curve $C: \eta = \eta[t]$ of $\hat{u}_2$ with $\eta[0] = \eta(u)$ satisfies

$$S^a(\eta[t], \hat{u}_1\{\eta[t]\}) = u^a \tag{3.19}$$

for $a = 1, \cdots m$ and any t in $(-\varepsilon, \varepsilon)$. Twice differentiating (3.19) in t, we have

$$\begin{aligned}&\left[\frac{d}{dt}\{\dot{\eta}_j[t]D_1^{jb}(\eta[t])\partial_b B^{ai}(\hat{u}_1[t])\}\right][\eta_i[t] - \eta_i(u)]\\ &\quad + 2\dot{\eta}_i[t]\dot{\eta}_j[t]D_1^{jb}(\eta[t])\partial_b B^{ai}(\hat{u}_1[t]) + B^{ai}(\hat{u}_1[t])\ddot{\eta}_i[t] = 0\end{aligned} \tag{3.20}$$

for $a = 1, \cdots, m$. The equations (3.20) lead to the relation (3.14) at $t = 0$ by reason of (3.18). This shows the $\overset{m}{\Gamma}$-transversality of $\hat{u}_2$, which completes the proof by Theorem 3.

PROOF OF THEOREM 2. Under the assumption $(A_{2,1})$ the affine connection $\Gamma^{(\rho w)}$ associated with ρ has the components

$$\Gamma^{(\rho w)i}_{jk}(\theta) = -\frac{W'''(1) + W''(1)}{W''(1)}T^i_{jk}(\theta)$$

with respect to the θ^i-coordinate, where $\{T^i_{jk}(\theta)\}$ are defined in (2.1). By the transformation rule (2.3) of affine connections, the components of $\Gamma^{(\rho w)}$ are calculated as

$$\Gamma^{(\rho w)jk}_{i}(\eta) = -\frac{W'''(1) + 2W''(1)}{W''(1)}T_i^{jk}(\theta\{\eta\})$$

with respect to the η_i-coordinate, where

$$T_i^{jk}(\theta) = g^{jj'}(\theta)g^{kk'}(\theta)T^{i'}_{j'k'}(\theta)g_{i'i}(\theta)$$

with the inverse elements $\{g^{ij}(\theta)\}$ of $g_{ji}(\theta)$. Therefore the condition (1.1) implies the coincidence of $\Gamma^{(\rho_W)}$ with $\overset{m}{\Gamma}$. Then by Proposition 3, the estimator $\hat{u}_{\rho_W}$ is $\overset{m}{\Gamma}$-transversal to the model $\tilde{\mathscr{F}}^m$. This completes the proof by Theorem 3.

PROOF OF COROLLARY 1. By definition the α^*-minimum contrast estimator is generated by the function

$$W^*_\alpha(t) = \frac{1}{\alpha^2}\left\{\frac{2}{1+\alpha}(1 - t^{(1+\alpha)/2}) + 4(\alpha^2 - 1)(1 - t^{1/2}) + \frac{2}{1-\alpha}(1 - t^{(1-\alpha)/2})\right\},$$

which satisfies the condition (1.1) for every α, $0 < \alpha < 1$. The contrast function generated by W^*_α is easily seen to satisfy $(A_{2,1})$ for every α, $0 < \alpha < 1$. The proof is completed by Theorem 2.

By l'Hospital's theorem we have that

$$\lim_{\alpha \searrow 0} W^*_\alpha(t) = ½\, t^{1/2}(\log t - 2)^2 - 8t^{1/2} + 6,$$

which also generates a second order efficient estimator.

Let ρ_W be a non-symmetric contrast function. For any β, $0 < \beta < 1$, a new contrast function is defined by

$$\rho^{[\beta]}_W(\theta_1, \theta_2) = (1 - \beta)\rho_W(\theta_1, \theta_2) + \beta\rho_W(\theta_2, \theta_1).$$

Then we obtain the following corollary of Theorem 3.

COROLLARY 2. *The minimum contrast estimator based on* $\rho^{[\beta_0]}_W$ *is second order efficient for*

$$\beta_0 = \frac{2W''(1) + W'''(1)}{3W''(1) + 2W'''(1)}, \tag{3.21}$$

if $0 < \beta_0 < 1$.

PROOF. Let $\{\Gamma^{[\beta]i}_{Wjk}(\theta)\}$ be components of $\Gamma^{[\beta]}_W$ associated with $\rho^{[\beta]}_W$ with respect to θ^i-coordinate. It follows from a straightforward calculation that

$$\Gamma^{[\beta]i}_{Wjk}(\theta) = \frac{(3\beta\text{-}1)\,W''(1) + (2\beta - 1)\,W'''(1)}{W''(1)}\,T^i_{jk}(\theta)$$

where $\{T^i_{jk}(\theta)\}$ are defined in (2.1). Therefore $\Gamma^{[\beta_0]}_W$ for the case (3.21) is equal to $\overset{m}{\Gamma}$. This completes the proof by Theorem 3.

We note that $\Gamma^{[1/2]}_W$ is the same as the metric connection $\overset{\alpha}{\Gamma}$ for $\alpha = 0$ for any ρ_W (e.g. the Jeffreys contrast function in Section 1).

EXAMPLE. We mention a 1-parameter curved exponential family of multinomial distributions with 4 cells, which have probabilities

$$\frac{2+u}{4}, \frac{1-u}{4}, \frac{1-u}{4}, \frac{u}{4}$$

for u, $0 < u < 1$ (cf. Chapter IV in Fisher [4]). The model is curved (non-flat) in the natural coordinate. We adopt the observed frequencies 125, 18, 20, 34 shown in Chapter 5 of Rao [8]. We note that the α-Chernoff contrast function is well defined for all $\alpha \in R$ if the common support of $\mathscr{F}^n$ is finite. Then some estimators in Section 1 are computed as in Table 1, which shows the slight differences between the first order and the second order efficient estimators.

TABLE 1

method	α	estimated value of u
maximum likelihood		.6268215
	3.0	.6268217
	.8	.6268215
α^*-minimum contrast	.6	.6268215
	.4	.6268214
	.2	.6268212
	3.0	.6264057
	.8	.6266366
	.6	.6266574
α-minimum contrast	.4	.6266781
	.2	.6266988
	.0	.6267193
	−3.0	.6264057

Acknowledgments. I would like to thank the referees and the Associate Editor for valuable comments on the original version of this paper. I express my thanks to Professor S. Amari of University of Tokyo for kind suggestions and providing me with his unpublished lecture notes. I am grateful to Professors M. Okamoto and N. Inagaki and Dr. Y. Toyooka of Osaka University for helpful comments. Professor N. Inagaki suggested Theorem 1 to me from a practical viewpoint.

REFERENCES

[1] AMARI, S. (1982). Differential geometry of curved exponential families—curvature and information loss. *Ann. Statist.* **10** 357–385.
[2] EFRON, B. (1975). Defining the curvature of a statistical problem (with applications to second order efficiency). *Ann. Statist.* **3** 1189–1217.
[3] DAWID A. P. (1975). Comments on the paper by Efron. *Ann. Statist.* **3** 1231–1234.
[4] FISHER, R. A. (1970). *Statistical Methods for Research Workers.* Oliver and Boyd, Edinburgh.
[5] HATTORI, A. (1980). *Manifolds (in Japanese).* Iwanami Shoten, Tokyo.
[6] KOBAYASHI, S. and NOMIZU, K. (1963). *Foundations of Differential Geometry, vol I.* Interscience Publishers, New York.
[7] PFANZAGL, J. (1973). Asymptotic expansions related to minimum contrast estimators. *Ann. Statist.* **1** 993–1026.
[8] RAO, C. R. (1973). *Linear Statistical Inference and its Applications.* Wiley, New York.

DEPARTMENT OF APPLIED MATHEMATICS
FACULTY OF ENGINEERING SCIENCE
OSAKA UNIVERSITY
TOYONAKA, OSAKA, JAPAN.