Ann. Inst. Statist. Math.
35 (1983), Part A, 1-24

DIFFERENTIAL GEOMETRY OF EDGEWORTH EXPANSIONS IN CURVED EXPONENTIAL FAMILY

SHUN-ICHI AMARI AND MASAYUKI KUMON

(Received Oct. 5, 1981; revised Aug. 12, 1982)

Summary

In order to construct a higher-order asymptotic theory of statistical inference, it is useful to know the Edgeworth expansions of the distributions of related statistics. Based on the differential-geometrical method, the Edgeworth expansions are performed up to the third-order terms for the joint distribution of any efficient estimators and complementary (approximate) ancillary statistics in the case of curved exponential family. The marginal and conditional distributions are also obtained. The roles and meanings of geometrical quantities are elucidated by the geometrical interpretation of the Edgeworth expansions. The results of the present paper provide an indispensable tool for constructing the differential-geometrical theory of statistics.

1. Introduction

Since Efron [10] introduced the concept of statistical curvature, it has gradually been recognized that the curvature of a statistical model plays an important role (Efron [11], Efron and Hinkley [12], Madsen [16]). Amari [3] has given a general differential-geometrical framework for analyzing spaces of statistical distributions, and proved the validity of the geometrical approach to treat the asymptotic properties of statistical inference, such as the second-order efficiency, second-order ancillarity and conditional inference (Amari [3]-[5]). The separate endeavors of Rao [18], Amari [2], Chentsov [7], Efron [10], etc. are thus unified to result in a new powerful differential-geometrical method of statistics.

Differential-geometry plays indeed an essential role in the higher-order asymptotic theory of statistical inference. On the other hand, the Edgeworth type expansions are necessary to know the asymptotic properties of inference. Hence, it is necessary to give a geometrical representation of Edgeworth expansions. The present paper performs this task in a curved exponential family. To this end, we introduce a new coordinate system consisting of an estimator (or a test statistic in the case of testing hypothesis) and complementary approximate ancillary statistics. We obtain the Edgeworth expansion up to the third order of the joint distribution of these statistics in terms of geometrical quantities. This provides us with an indispensable tool for analyzing higher-order asymptotic properties related to problems of statistical inference.

We shall show briefly that the present theory gives direct solutions to the problems of higher-order information loss and higher-order efficiency of an estimator (Efron [10], Amari [3], [4], see also Pfanzagl [17], Akahira and Takeuchi [1]), of higher-order ancillarity (see, e.g., Cox [8], Amari [5]), of conditional inference and of testing statistical hypothesis and interval estimation (Kumon and Amari [15]).

Refer to Amari [3], [4] for the differential-geometrical framework for statistics and to Schouten [19] for the index notations in differential geometry.

2. Geometry of curved exponential family

2.1. *Differential geometry of exponential family*

Let us consider a full, regular, minimally represented exponential family S of distributions. A member of S can be represented by the density function of the form

$$p(x, \theta)=\exp\{\theta^i x_i-\psi(\theta)\} \tag{2.1}$$

on some carrier measure $P(x)$ on the sample space X of a vector random variable $x=(x_i)=(x_1,\cdots, x_n)$, where $\theta=(\theta^i)=(\theta^1, \theta^2,\cdots, \theta^n)$ is a vector parameter specifying the distributions in S. Einstein's summation convention is assumed throughout the present paper, so that the summation is automatically taken over those indices (such as i in the above expression) that appear twice in a term once as a superscript and once as a subscript. Hence, $\theta^i x_i$ implies $\sum_{i=1}^{n} \theta^i x_i$ automatically without the summation symbol $\sum$. We treat the case where random variables x and parameters θ take on continuous values. The above parameter θ is called the natural parameter of the exponential family S. The family S forms a manifold, where θ plays the role of a local coordinate system.

We can introduce a Riemannian metric in S (Rao [18], Chentsov [7], Amari [3]).

DEFINITION 1. The metric tensor $g_{ij}(\theta)$ of S at θ is defined by the Fisher information matrix

$$g_{ij}(\theta)=\mathrm{E}\,[\partial_i l(x, \theta)\partial_j l(x, \theta)]\,, \tag{2.2}$$

where E denotes the expectation, ∂_i denotes the differentiation $\partial/\partial\theta^i$, and

$$l(x, \theta)=\log p(x, \theta) .$$

In the case of exponential family, we have

$$g_{ij}(\theta)=\partial_i\partial_j\psi(\theta) \tag{2.3}$$

from the definition. The inner product of two vectors $X=(X^i)$ and $Y=(Y^i)$ in the tangent space T_θ at θ is given by

$$(X, Y)=g_{ij}X^iY^j .$$

When this vanishes, X and Y are said to be orthogonal. The length $\|X\|$ of a vector X is also given by

$$\|X\|^2=g_{ij}X^iX^j .$$

A one-parameter family of affine connections is introduced in the following manner (Amari [3], cf. Chentsov [7], Dawid [9]). Let $\overset{\alpha}{\Gamma}_{ijk}$ be the covariant components of an affine connection parametrized by α, which we call the α-connection.

DEFINITION 2. The α-connection is defined by

$$\overset{\alpha}{\Gamma}_{ijk}(\theta)=\mathrm{E}\,[\partial_i\partial_j l(x, \theta)\partial_k l(x, \theta)]+\frac{1-\alpha}{2}T_{ijk}(\theta) , \tag{2.4}$$

where

$$T_{ijk}=\mathrm{E}\,[\partial_i l(x, \theta)\partial_j l(x, \theta)\partial_k l(x, \theta)] \tag{2.5}$$

is a symmetric tensor.

We can easily calculate the following relations in the space of exponential family,

$$\overset{\alpha}{\Gamma}_{ijk}=\frac{1-\alpha}{2}T_{ijk} , \tag{2.6}$$

$$T_{ijk}=\partial_i\partial_j\partial_k\psi=\partial_i g_{jk} . \tag{2.7}$$

The connection with $\alpha=1$ is called the exponential connection and that with $\alpha=-1$ is called the mixture connection. Refer to Amari [3], [4] for the detailed geometrical structures of the space of an exponential family.

It is convenient to use another coordinate system (parametrization) in S. The expectation of x with respect to the distribution $p(x, \theta)$, $\eta(\theta)=\mathrm{E}\,[x]$ is calculated in the component form as

(2.8) $$\eta_i(\theta)=\mathrm{E}\,[x_i]=\partial_i\psi(\theta)\,.$$

It is known (Chentsov [7], Barndorff-Nielsen [6]) that the transformation between θ and η is one-to-one, and we can use η as a parameter or a coordinate system specifying points in S. We call η the expectation parameter.

Since the Jacobian matrices of the coordinate transformations between θ and η are given by

(2.9) $$\partial\eta_i/\partial\theta^j=g_{ij}\,,\qquad \partial\theta^i/\partial\eta_j=g^{ij}\,,$$

where g^{ij} is the inverse matrix of g_{ij}, we can obtain the metric and the α-connections in terms of the expectation parameter η. Since a point $x=(x^i)$ in the sample space X can naturally be mapped to a point η in S having the same expectation coordinates $\eta_i=x_i$, we can identify the sample space X with the space S of distributions by the use of the expectation parameter. The expectation parameter η is convenient in this respect. We hereafter use mainly the expectation parameter. However, the same results can be obtained by using the natural parameter θ, because our theory is independent of the specific choice of coordinate systems.

2.2. *Curved exponential family and ancillary subspaces*

Let M be a statistical model consisting of the density functions $f(x, u)$, parametrized by an m-dimensional parameter $u=(u^a)=(u^1, u^2, \cdots, u^m)$. When M is smoothly imbedded in S by

$$f(x, u)=p[x, \theta(u)]\,,$$

M is called an (n, m)-curved exponential family. It forms an m-dimensional manifold with a local coordinate system u. It can also be considered as an m-dimensional submanifold of S, defined by the equation $\theta=\theta(u)$ or $\eta=\eta(u)$ in the respective coordinate systems of S.

Let us attach an $(n-m)$-dimensional smooth submanifold $A(u)$ to every point $\theta(u)$ or $\eta(u)$ of M imbedded in S, such that $A(u)$ transverses M at $\eta(u)$ and $A=\{A(u)\,|\,\eta(u)\in M\}$ forms a smooth family of submanifolds. We call the family $A=\{A(u)\}$ an ancillary family rigged to a curved exponential family M, and call each $A(u)$ the ancillary subspace at u of A. Let us introduce a coordinate system $v=(v^\kappa)$, $\kappa=m+1, m+2,\cdots, n$ in each $A(u)$ such that the origin $v=0$ is put at point $\eta(u)$ on M. Then, given an ancillary family A, a point η in S can uniquely be determined by the pair (u, v) (at least in some neighborhood of M), such that the point η is on the ancillary subspace $A(u)$ and the coordinates of the η on $A(u)$ are v. We can represent the point η by a smooth function of u and v as $\eta=\eta(u, v)$ or in the component form as

$$\eta_i=\eta_i(u^a, v^\kappa)\ , \qquad i=1,\cdots,n;\ a=1,\cdots,m;\ \kappa=m+1,\cdots,n\ ,$$

where we use indices i, j, k, etc. to represent quantities related to the coordinate system θ or η in S, indices a, b, c, etc. to represent those in the coordinate system u of M, and indices κ, λ, μ, etc. to represent those in the coordinate system v of the ancillary family A. Obviously, i, j, k, etc. run from 1 to n, a, b, c, etc. run from 1 to m, and κ, λ, μ, etc. run from $m+1$ to n.

The pair (u, v) can be regarded as a local coordinate system of S in some neighborhood of M. It is convenient to introduce a new single vector variable w by $w=(u, v)$ or in the component form

$$w^\alpha=(u^a, v^\kappa)\ , \qquad \alpha=1,\cdots,n;\ a=1,\cdots,m;\ \kappa=m+1,\cdots,n$$

i.e., $w^1=u^1,\cdots,\ w^m=u^m,\ w^{m+1}=v^{m+1},\cdots,\ w^n=v^n$. We use indices α, β, γ, etc. to denote quantities in the coordinate system w. The transformation from w to η can be written as

$$\eta=\eta(w)\ .$$

Let $x_1, x_2,\cdots, x_N$ be N independent observations from a distribution $f(x, u_0)$ in M. Then, $\bar{x}=\frac{1}{N}\sum_{i=1}^{n} x_i$ is a sufficient (vector) statistic. Since a point $\bar{x}$ in the sample space X can be identified with a point $\bar{\eta}$ in S having the same components in the η-coordinates $\bar{\eta}=\bar{x}$, we can represent this point $\bar{x}$ in terms of the new w-coordinates related to A as

$$\bar{x}=\eta(\hat{u}, \hat{v})\ . \tag{2.10}$$

This defines new statistics $\hat{u}$ and $\hat{v}$, which together form a sufficient statistic. Obviously $\hat{u}$ and $\hat{v}$ are determined, depending on the ancillary family A rigged to M.

Reduction of information is carried into effect by a statistical inference procedure from $\bar{x}$ to some statistic $\hat{u}$, e.g., $\hat{u}$ is an estimator of u_0 in the case of estimation, $\hat{u}$ is a test statistic in the case of testing, etc. If we choose an ancillary family A such that the statistic under consideration can be derived by the M-part $\hat{u}$ of $\hat{w}=(\hat{u}, \hat{v})$, which is another expression of $\bar{x}$, this coordinate system w associated with A is convenient to analyze the performance of the statistical procedure. The M-part $\hat{u}$ represents the summarized information, and the A-part $\hat{v}$ represents the abandoned information. This is the reason why we introduce an ancillary family A.

The metric tensor is written as

$$g_{\alpha\beta}(u)=g^{ij}B_{\alpha i}B_{\beta j} \tag{2.11}$$

in the new coordinate system w, where

$$B_{ai}(u)=\partial_a\eta_i(u, 0)\ , \qquad \partial_\alpha=\partial/\partial w^\alpha \tag{2.12}$$

is the Jacobian matrix of the coordinate transformation from η to w. Here, all the quantities are evaluated on M, i.e., at $v=0$, and hence are functions of u only. The m vectors $B_1,\cdots, B_m$,

$$B_a(u)=(B_{ai})\ , \qquad a=1,\cdots, m$$

form a set of vectors spanning the tangent space at u of M. Similarly, $n-m$ vectors $B_{m+1},\cdots, B_n$

$$B_\kappa(u)=(B_{\kappa i})\ , \qquad \kappa=m+1,\cdots, n$$

form a set of vectors spanning the tangent space of $A(u)$ at $v=0$. Their contravariant expressions are

$$B_a^i=g^{ij}B_{aj}\ , \qquad B_\kappa^i=g^{ij}B_{\kappa j}\ .$$

The M-part of $g_{\alpha\beta}$,

$$g_{ab}=g_{ij}B_a^iB_b^j=g^{ij}B_{ai}B_{bj} \tag{2.13}$$

is the Fisher information matrix of M, and hence is the metric tensor of M. The A-part $g_{\kappa\lambda}$ of $g_{\alpha\beta}$,

$$g_{\kappa\lambda}(u)=g_{ij}B_\kappa^iB_\lambda^j=g^{ij}B_{\kappa i}B_{\lambda j} \tag{2.14}$$

is the metric of $A(u)$ at $v=0$. The mixed part

$$g_{a\kappa}(u)=B_a^iB_\kappa^jg_{ij} \tag{2.15}$$

is the inner product of the tangent vectors of M and $A(u)$, and it vanishes when M and $A(u)$ are orthogonal at u. The ancillary family A is said to be orthogonal at u_0, when

$$g_{a\kappa}(u_0)=0\ , \tag{2.16}$$

and is said to be orthogonal when $g_{a\kappa}(u)=0$ holds for all u. When a family A is orthogonal at u_0, we can define

$$Q_{ab\kappa}=\partial_ag_{b\kappa}(u_0)\ , \tag{2.17}$$

which is a tensor because of (2.16). This tensor can be used to evaluate $g_{a\kappa}(u)$, when u is close to u_0. We have indeed

$$g_{a\kappa}(u)=Q_{ba\kappa}(u^b-u_0^b)+O(|u-u_0|^2)\ . \tag{2.18}$$

The α-connection can be represented in the w-coordinates as

$$\overset{\alpha}{\Gamma}_{\beta\gamma\delta}=\overset{\alpha}{\Gamma}_{ijk}B_\beta^iB_\gamma^jB_\delta^k+B_\delta^j(\partial_\beta B_\gamma^i)g_{ij}\ , \tag{2.19}$$

because the Jacobian matrix of the transformation from θ to w is given by B^i_α. Taking account of

$$(\partial_\beta B^i_\gamma)g_{ij}=\partial_\beta B_{\gamma j}-B^i_\gamma\partial_\beta g_{ij}=\partial_\beta B_{\gamma j}-B^i_\beta B^k_\gamma T_{ijk} ,$$

we have

$$\overset{\alpha}{\Gamma}_{\beta\gamma\delta}=-\frac{1+\alpha}{2}T_{\beta\gamma\delta}+C_{\beta\gamma\delta} ,$$

where

$$C_{\beta\gamma\delta}=(\partial_\beta B_{\gamma j})B^j_\delta , \qquad T_{\beta\gamma\delta}=T_{ijk}B^i_\beta B^j_\gamma B^k_\delta .$$

By putting $\alpha=-1$ and $\alpha=1$, we have

(2.20) $$\overset{\mathrm{m}}{\Gamma}_{\beta\gamma\delta}=C_{\beta\gamma\delta} ,$$

(2.21) $$\overset{\mathrm{e}}{\Gamma}_{\beta\gamma\delta}=C_{\beta\gamma\delta}-T_{\beta\gamma\delta} ,$$

where "m" and "e" represent the mixture and exponential connections, respectively. The M-part of the α-connection

(2.22) $$\overset{\alpha}{\Gamma}_{abc}=-\frac{1+\alpha}{2}T_{abc}+C_{abc}$$

gives the α-connection in M. The A-part

(2.23) $$\overset{\alpha}{\Gamma}_{\kappa\lambda\mu}=-\frac{1+\alpha}{2}T_{\kappa\lambda\mu}+C_{\kappa\lambda\mu}$$

gives the α-connection in $A(u)$.

We next consider curvatures of a manifold. The α-curvature of a manifold is defined by the rate of the "α-intrinsic" change in the tangent directions as positions change in the manifold. The "α-intrinsic" change is measured by the covariant derivative of the tangent vectors with respect to the α-connection. The α-curvature of M is thus defined by a tensor

(2.24) $$\overset{\alpha}{H}{}^i_{ab}(u)=B^j_a\overset{\alpha}{\nabla}_j B^i_b ,$$

where $\overset{\alpha}{\nabla}_j$ is the covariant derivative with respect to the α-connection. Similarly, the α-curvature of an ancillary subspace $A(u)$ is defined by

(2.25) $$\overset{\alpha}{H}{}^i_{\kappa\lambda}(u)=B^j_\kappa\overset{\alpha}{\nabla}_j B^i_\lambda .$$

We consider the exponential (i.e., $\alpha=1$) curvature of the model M and the mixture (i.e., $\alpha=-1$) curvature of the ancillary subspace of an orthogonal ancillary family A, because they play fundamental roles in the problems of statistical inference. The exponential curvature of

M is given in this case by the tensor

(2.26) $$\overset{\mathrm{e}}{H}_{ab\kappa}(u)=\overset{\mathrm{e}}{H}{}^{i}_{ab}B_{\kappa i}=(\partial_a B^i_b)B_{\kappa i}\ ,$$

where "e" denotes the underlying exponential connection because of $\overset{\mathrm{e}}{H}{}^{i}_{ab}B_{ci}=0$. The mixture curvature of $A(u)$ is given by the tensor

(2.27) $$\overset{\mathrm{m}}{H}_{\kappa\lambda a}(u)=\overset{\mathrm{m}}{H}{}^{i}_{\kappa\lambda}B_{ai}=(\partial_\kappa B_{\lambda i})B^i_a\ ,$$

where "m" denotes the underlying mixture ($\alpha=-1$) connection.

2.3. *Stochastic expansion*

Let us define new random variables

$$\tilde{x}=\sqrt{N}\,[\bar{x}-\eta(u,0)]\ ,\quad \tilde{u}=\sqrt{N}\,(\hat{u}-u)\ ,\quad \tilde{v}=\sqrt{N}\,\hat{v}\ ,\quad \tilde{w}=(\tilde{u},\tilde{v})\ ,$$

where u is the true parameter of distribution. Since the expectation of $\bar{x}$ is $\eta(u,0)$, $\tilde{x}$ is asymptotically normally distributed with zero mean and covariance matrix g_{ij}. In order to obtain the joint distribution of $\tilde{u}$ and $\tilde{v}$, we expand (2.10) at $(u,0)$. The stochastic expansion yields

(2.28) $$\tilde{x}_i=B_{\alpha i}\tilde{w}^\alpha+\frac{1}{2\sqrt{N}}C_{\alpha\beta i}\tilde{w}^\alpha\tilde{w}^\beta+\frac{1}{6N}D_{\alpha\beta\gamma i}\tilde{w}^\alpha\tilde{w}^\beta\tilde{w}^\gamma+O_p(N^{-3/2})\ ,$$

where

$$B_{\alpha i}=\partial_\alpha\eta_i\ ,\qquad C_{\alpha\beta i}=\partial_\alpha\partial_\beta\eta_i\ ,\qquad D_{\alpha\beta\gamma i}=\partial_\alpha\partial_\beta\partial_\gamma\eta_i$$

are evaluated at $(u,0)$. By multiplying the inverse of $B_{\alpha i}$ or $g^{\alpha\beta}B_{\beta j}g^{ji}$, we have

(2.29) $$\tilde{w}^\alpha=g^{\alpha\beta}B^i_\beta\tilde{x}_i-\frac{1}{2\sqrt{N}}C^\alpha_{\beta\gamma}\tilde{w}^\beta\tilde{w}^\gamma-\frac{1}{6N}D^\alpha_{\beta\gamma\delta}\tilde{w}^\beta\tilde{w}^\gamma\tilde{w}^\delta+O_p(N^{-3/2})\ ,$$

where $g^{\alpha\beta}$ is the inverse of $g_{\alpha\beta}$,

$$C^\alpha_{\beta\gamma}=C_{\beta\gamma i}B^i_\delta g^{\delta\alpha}\ ,\qquad D^\alpha_{\beta\gamma\delta}=D_{\beta\gamma\delta i}B^i_\varepsilon g^{\varepsilon\alpha}\ ,$$

and indices α, β, γ, etc. are raised or lowered by the use of $g_{\alpha\beta}$ or $g^{\alpha\beta}$.

Since the quantity $C_{\alpha\beta\gamma}$, which gives the mixture connection in the w-coordinate system, plays an important role, we clarify the geometric meanings of its M-, A-, and mixed parts.

THEOREM 1. *The M-, A-, and mixed parts of $C_{\alpha\beta\gamma}$ represent the following geometric quantities.*

(2.30) $$C_{abc}=\overset{\mathrm{m}}{\Gamma}_{abc}\ ,$$

(2.31) $$C_{\kappa\lambda\mu}=\overset{\mathrm{m}}{\Gamma}_{\kappa\lambda\mu}\ ,$$

(2.32) $$C_{ab\kappa}=\overset{\mathrm{m}}{H}_{ab\kappa} ,$$

(2.33) $$C_{\kappa\lambda a}=\overset{\mathrm{m}}{H}_{\kappa\lambda a} ,$$

(2.34) $$C_{a\kappa b}=Q_{ab\kappa}-\overset{\mathrm{e}}{H}_{ab\kappa} ,$$

especially

(2.35) $$C_{a\kappa b}=-\overset{\mathrm{e}}{H}_{ab\kappa} ,$$

for an orthogonal family A, because $g_{b\kappa}(u)=0$ implies $Q_{ab\kappa}=0$. It is possible to choose the coordinate system v^κ for each $A(u)$, such that

(2.36) $$C_{a\kappa\lambda}=C_{a\lambda\kappa}=\frac{1}{2}T_{a\kappa\lambda}$$

holds.

PROOF. The equations (2.30)–(2.33) are the direct consequences of (2.22), (2.23), (2.24) and (2.27), respectively. We have (2.34) from

$$C_{a\kappa b}=C_{\kappa ab}=(\partial_a B_{\kappa i})B^i_b=\partial_a(B_{\kappa i}B^i_b)-B_{\kappa i}\partial_a B^i_b=\partial_a g_{\kappa b}-\overset{\mathrm{e}}{H}_{ab\kappa}=Q_{ab\kappa}-\overset{\mathrm{e}}{H}_{ab\kappa} .$$

It is always possible to choose the coordinate system v^κ in each $A(u)$ such that $g_{\kappa\lambda}(u)$ is kept constant for any u. In this case, from $0=\partial_a g_{\kappa\lambda}=\partial_a(B^i_\kappa B^j_\lambda g_{ij})$, we have

$$C_{a(\kappa\lambda)}=\frac{1}{2}T_{a\kappa\lambda} ,$$

where the bracket ( ) implies the symmetrization of indices,

$$C_{a(\kappa\lambda)}=\frac{1}{2}(C_{a\kappa\lambda}+C_{a\lambda\kappa}) .$$

Moreover, by applying an adequate orthogonal transformation to each $A(u)$, the coordinate v^κ in each $A(u)$ can be made to satisfy

$$B^i_\lambda\partial_a B_{\kappa i}-B^i_\kappa\partial_a B_{\lambda i}=0 .$$

We have in this case $C_{a\kappa\lambda}=C_{a\lambda\kappa}$, proving the theorem.

By taking the expectation of (2.29), we have

(2.37) $$\mathrm{E}\,[\tilde{w}^\alpha]=-\frac{1}{2\sqrt{N}}C^\alpha+O(N^{-1}) ,$$

where

$$C^\alpha=C^\alpha_{\beta\gamma}g^{\beta\gamma}$$

because of

$$\mathrm{E}\,[\tilde{w}^{\alpha}\tilde{w}^{\beta}]=g^{\alpha\beta}+O(N^{-1})\,.$$

This shows that the expectation of the statistic $\hat{w}=(\hat{u},\hat{v})$ is biased from $w_0=(u_0, 0)$ by $-C^{\alpha}(u_0)/2N+O(N^{-3/2})$. We can modify the statistic $\hat{w}$ to yield the unbiased version

$$\hat{w}^{*}=\hat{w}+C^{\alpha}(\hat{w})/2N\,, \tag{2.38}$$

and thus

$$\mathrm{E}\,[w^{*}]=w_0+O(N^{-3/2})\,.$$

In terms of the deviation part

$$\tilde{w}^{*}=\sqrt{N}\,(\hat{w}^{*}-w_0)\,,$$

we have

$$\tilde{w}^{*}=\tilde{w}-\mathrm{E}_{\hat{w}}\,[\tilde{w}]\,, \tag{2.39}$$

where $\mathrm{E}_{\hat{w}}$ denotes the expectation with respect to the distribution in S specified by $\hat{w}$. This can be written as

$$\tilde{w}^{*\alpha}=\tilde{w}^{\alpha}+\frac{1}{2\sqrt{N}}C^{\alpha}(\hat{w})=\tilde{w}^{\alpha}+\frac{1}{2\sqrt{N}}C^{\alpha}(u_0)+\frac{1}{2N}\partial_{\beta}C^{\alpha}\tilde{w}^{\beta}+O(N^{-3/2})\,. \tag{2.40}$$

Since the statistic $\tilde{w}^{*}$ is unbiasedly modified up to order N^{-1}, it is convenient to obtain the Edgeworth expansion of the distribution of $\tilde{w}^{*}$ first and then to obtain that of $\tilde{w}$ therefrom. Moreover, the unbiasedly modified statistic $\tilde{u}^{*}$ itself plays a very important role in the theory of estimation, as will be seen in the next section.

One may consider that the statistic

$$\tilde{w}^{**}=\tilde{w}-\mathrm{E}_{\hat{u}}\,[\tilde{w}]\,, \tag{2.41}$$

where $\mathrm{E}_{\hat{u}}$ denotes the expectation with respect to the distribution specified by $(\hat{u}, 0)$ or $f(x, \hat{u})$, is more natural than $\tilde{w}^{*}$. The points η satisfying

$$\hat{u}^{**}(\eta)=\text{const.}$$

indeed lie in one $A(u)$, while those satisfying

$$\hat{u}^{*}(\eta)=\text{const.}$$

do not. However, we can prove that the distributions of $\tilde{u}^{*}$ and $\tilde{u}^{**}$ are the same except for the term of $O(N^{-3/2})$. Hence, we can regard them equivalent so long as the third-order asymptotic theory is concerned.

3. Geometric aspects of Edgeworth expansions

3.1. *Joint distribution of $\tilde{u}$ and $\tilde{v}$*

Vector random variable $\tilde{w}$ or $\tilde{w}^*$ is asymptotically normally distributed, with asymptotic covariance matrix $g^{\alpha\beta}$. The probability density function $p(\tilde{w}^*; u_0)$ of $\tilde{w}^*$, where u_0 is the true parameter of the distribution, can be expanded in the following Gram-Charlier series,

$$(3.1) \qquad p(\tilde{w}^*; u_0)=\varphi[\tilde{w}^*; g_{\alpha\beta}(u_0)]\left\{1+\sum_{k=1}^{\infty} c_{\alpha_1\cdots\alpha_k}(u_0)h^{\alpha_1\cdots\alpha_k}(\tilde{w}^*)\right\},$$

where

$$(3.2) \qquad \varphi(\tilde{w}^*; g_{\alpha\beta})=(\det|g_{\alpha\beta}|)^{1/2}(2\pi)^{-n/2}\exp\left\{-\frac{1}{2}g_{\alpha\beta}\tilde{w}^{*\alpha}\tilde{w}^{*\beta}\right\}$$

and $h^{\alpha_1\cdots\alpha_k}(\tilde{w}^*)$ is the tensorial Hermite polynomial in $\tilde{w}^*$ of degree k with respect to the normal distribution $\varphi(\tilde{w}^*; g_{\alpha\beta})$. See Appendix 1 for the Hermite polynomials.

In virtue of the orthogonal relations of the Hermite polynomials, we have

$$(3.3) \qquad c^{\alpha_1\cdots\alpha_k}(u_0)=\frac{1}{k!}\int p(\tilde{w}^*; u_0)h^{\alpha_1\cdots\alpha_k}(\tilde{w}^*)d\tilde{w}^*=\frac{1}{k!}\mathrm{E}\,[h^{\alpha_1\cdots\alpha_k}(\tilde{w}^*)],$$

where $c^{\alpha_1\cdots\alpha_k}$ is the contravariant (upper index) version of $c_{\alpha_1\cdots\alpha_k}$, indices being raised or lowered by the use of $g^{\alpha\beta}$ or $g_{\alpha\beta}$. Hence, the coefficients can be calculated from the moments

$$(3.4) \qquad \mu^{\alpha_1\cdots\alpha_k}=\mathrm{E}\,[\tilde{w}^{*\alpha_1}\cdots\tilde{w}^{*\alpha_k}].$$

(Appendix 2). It is more tractable to use the cumulants $\kappa^{\alpha_1\cdots\alpha_k}$ instead of the moments, since the cumulant $\kappa^{\alpha_1\cdots\alpha_k}$ of $\tilde{w}^*$ is of order $N^{-(k-2)/2}$ $(k\geqq 2)$. We put

$$(3.5) \qquad K^{\alpha\beta\gamma}=\sqrt{N}\,\kappa^{\alpha\beta\gamma},$$

$$(3.6) \qquad K^{\alpha\beta\gamma\delta}=N\kappa^{\alpha\beta\gamma\delta}.$$

The higher-order cumulants are $o(N^{-1})$ and hence are not necessary to evaluate the distribution function up to the term of $O(N^{-1})$.

In order to calculate the moments $\mu^{\alpha_1\cdots\alpha_k}$ or cumulants $\kappa^{\alpha_1\cdots\alpha_k}$ from (2.29) we need to know the moments or cumulants of $\tilde{x}_i$. Since $\psi(\theta)$ is the cumulant generating function of x, we have

$$\mathrm{E}\,[\tilde{x}_i]=0, \qquad \mathrm{E}\,[\tilde{x}_i\tilde{x}_j]=g_{ij}, \qquad \mathrm{E}\,[\tilde{x}_i\tilde{x}_j\tilde{x}_k]=T_{ijk}/\sqrt{N},$$

$$\mathrm{E}\,[\tilde{x}_i\tilde{x}_j\tilde{x}_k\tilde{x}_m]=3g_{(ij}g_{km)}+S_{ijkm}/N,$$

where T_{ijk} is the third cumulant, and

$$S_{ijkm}=\partial_i\partial_j\partial_k\partial_m\psi(\theta)$$

is the fourth cumulant of x.

By calculating the moments or cumulants of $\tilde{w}^*$ up to the necessary order (Appendix 2), we have the following geometric expression of the Edgeworth expansion of the joint distribution of $\tilde{u}^*$ and $\tilde{v}^*$.

THEOREM 2. *The distribution of $\tilde{w}^*$ is expanded as*

$$p(\tilde{w}^*;u_0)=\varphi(\tilde{w}^*;g_{\alpha\beta})\left\{1+\frac{1}{6\sqrt{N}}K_{\alpha\beta\gamma}h^{\alpha\beta\gamma}(\tilde{w}^*)+\frac{1}{4N}C^2_{\alpha\beta}h^{\alpha\beta}(\tilde{w}^*)\right.$$
$$\left.+\frac{1}{24N}K_{\alpha\beta\gamma\delta}h^{\alpha\beta\gamma\delta}(\tilde{w}^*)+\frac{1}{72N}K_{\alpha\beta\gamma}K_{\delta\varepsilon\zeta}h^{\alpha\beta\gamma\delta\varepsilon\zeta}(\tilde{w}^*)\right\}$$
$$+O(N^{-3/2}), \tag{3.7}$$

where

$$K_{\beta\gamma\delta}=T_{\beta\gamma\delta}-3C_{(\beta\gamma\delta)}=-3\overset{\alpha}{\Gamma}_{\beta\gamma\delta}\qquad(\alpha=-1/3) \tag{3.8}$$

$$C^2_{\alpha\beta}=\overset{m}{\Gamma}_{\gamma\delta\alpha}\overset{m}{\Gamma}_{\varepsilon\zeta\beta}g^{\gamma\varepsilon}g^{\delta\zeta} \tag{3.9}$$

$$K_{\alpha\beta\gamma\delta}=S_{\alpha\beta\gamma\delta}-4D_{(\alpha\beta\gamma\delta)}+12(\overset{m}{\Gamma}_{\varepsilon\alpha\beta}+\overset{e}{\Gamma}_{\alpha\beta\varepsilon})\overset{m}{\Gamma}_{\zeta\gamma\delta}g^{\varepsilon\zeta}\,. \tag{3.10}$$

The distribution of unmodified $\tilde{w}$ or partly modified $\tilde{w}^{**}$ can be derived as follows.

THEOREM 3. *The distribution of $\tilde{w}$ is derived by replacing $\tilde{w}^*$ in* (3.7) *by $\tilde{w}'$,*

$$\tilde{w}'^{\alpha}=\tilde{w}^{\alpha}+\frac{1}{2\sqrt{N}}C^{\alpha} \tag{3.11}$$

and by adding the term

$$-\frac{1}{2N}(\partial_{\alpha}C^{\gamma})g_{\gamma\beta}h^{\alpha\beta}$$

in the last bracket. The distribution of $\tilde{w}^{**}$ is obtained from (3.7) *by replacing $\tilde{w}^*$ by $\tilde{w}^{**}$ and by adding the term*

$$-\frac{1}{2N}(\partial_{\kappa}C^{\gamma})g_{\gamma\beta}h^{\kappa\beta}$$

in the last bracket.

3.2. *Efficiency of estimators and distribution of $\tilde{u}^*$ or $\tilde{u}$*

Let $\hat{u}$ be an estimator of u, $\hat{u}=f(\bar{x})$ which is a smooth function of the sufficient statistic $\bar{x}$ independently of N. Then, the set of the points $\bar{\eta}$ in S which are mapped to u by f, forms an $(n-m)$-dimensional submanifold $A(u)$,

$$A(u)=\{\bar{\eta}\in S\,|\,f(\bar{\eta})=u\}\ .$$

This implies that, when $\bar{x}\in A(u)$, the value of the estimator is u. When the estimator is consistent, we can show that $\eta(u)\in A(u)$. Thus, a consistent estimator $\hat{u}$ in general defines an ancillary family A consisting of the above $A(u)$'s. This A is said to be associated with the estimator. The coordinates $\hat{u}$ of the point $\bar{x}$ in the associated A, $\bar{x}=\eta(\hat{u}, \hat{v})$ give the value of the estimator. The unbiasedly modified estimator is given by $\hat{u}^*$. The random variables $\tilde{u}$ and $\tilde{u}^*$ are $\sqrt{N}$ times of the estimation error. Hence, by knowing the distribution of $\tilde{u}$ or $\tilde{u}^*$, we can evaluate the estimator. We can also calculate the amount of information loss caused by summarizing the data $\bar{x}$ into an estimator $\hat{u}$.

The distribution of $\tilde{u}^*$ is obtained by integrating (3.7) with respect to $\tilde{v}^*$. Let us put

$$\bar{g}_{ab}=g_{ab}-g_{a\kappa}g_{b\lambda}g^{\kappa\lambda}\ , \tag{3.12}$$

which is the inverse of the A-part g^{ab} of $g^{\alpha\beta}$. The inequality

$$\bar{g}_{ab}\leqq g_{ab} \tag{3.13}$$

holds, implying that $g_{ab}-\bar{g}_{ab}$ is positive semi-definite. The equality holds, when and only when $g_{a\kappa}(u_0)=0$, i.e., A is orthogonal at u_0.

THEOREM 4. *The distribution of* $\tilde{u}^*$ *or* $\tilde{u}$ *is given by*

$$p(\tilde{u}^*;u_0)=\varphi[\tilde{u}^*;\bar{g}_{ab}(u_0)]+O(N^{-1/2})\ . \tag{3.14}$$

According to (3.13), the theorem shows that the estimator $\hat{u}^*$ (or $\hat{u}$) is mostly concentrated at around the true value u_0, when the associated ancillary family A is orthogonal. We can prove from (3.13) that the Fisher information of $\hat{u}$ (or $\hat{u}^*$) is $N\bar{g}_{ab}(u_0)+O(1)$, and that the mean square error (or the covariance of the estimator $\hat{u}$) is $(N\bar{g}_{ab})^{-1}+O(N^{-2})$. Hence, an estimator is (first-order) efficient, when the associated A is orthogonal.

We next evaluate the higher-order terms of the distribution of $\tilde{u}^*$ for an orthogonal A. The same result has already been obtained in Akahira and Takeuchi [1] without geometrical interpretations.

THEOREM 5. *When* A *is an orthogonal family, the distribution of* $\tilde{u}^*$ *is given by*

$$(3.15)\quad p(\tilde{u}^*;u_0)=\varphi(\tilde{u}^*;g_{ab})\Big\{1+\frac{1}{6\sqrt{N}}K_{abc}h^{abc}(\tilde{u}^*)+\frac{1}{4N}C^2_{ab}h^{ab}(\tilde{u}^*)+\frac{1}{24N}K_{abcd}h^{abcd}(\tilde{u}^*)+\frac{1}{72N}K_{abc}K_{def}h^{abcdef}(\tilde{u}^*)\Big\}+O(N^{-3/2})\,,$$

where the third and fourth cumulant terms

$$(3.16)\quad K_{abc}=T_{abc}-3\overset{\mathrm{m}}{\Gamma}_{abc}=-3\overset{\alpha}{\Gamma}_{abc}\qquad (\alpha=-1/3)\,,$$

$$(3.17)\quad K_{abcd}=S_{abcd}-4D_{abcd}+12(\overset{\mathrm{m}}{\Gamma}_{eab}+\overset{\mathrm{e}}{\Gamma}_{abe})\overset{\mathrm{m}}{\Gamma}_{fcd}g^{ef}$$

are common to all orthogonal ancillary families independently of a specific A, and only the term

$$(3.18)\quad C^2_{ab}=C_{\alpha\beta a}C_{\gamma\delta b}g^{\alpha\gamma}g^{\beta\delta}=\overset{\mathrm{m}}{\Gamma}_{cda}\overset{\mathrm{m}}{\Gamma}_{efb}g^{ce}g^{df}+2\overset{\mathrm{e}}{H}_{ac\kappa}\overset{\mathrm{e}}{H}{}^{\kappa}_{bd}g^{cd}+\overset{\mathrm{m}}{H}_{\kappa\lambda a}\overset{\mathrm{m}}{H}{}^{\kappa\lambda}_{b}$$

depends on A through the mixture curvature $\overset{\mathrm{m}}{H}_{\kappa\lambda a}$.

PROOF. The expansion (3.15) is obtained by integrating (3.7) with respect to $\tilde{v}^*$ and by using the relation (A.1.4) in Appendix 1. When A is orthogonal, the density function $p(\tilde{u}^*;u_0)$ depends on A only through the mixture curvature $\overset{\mathrm{m}}{H}_{\kappa\lambda a}$ of $A(u_0)$ up to the terms of order N^{-1}. Hence, the higher-order efficiency is evaluated by the mixture curvature of the associated A.

By the use of the above expansion, we can easily calculate the covariance of an (unbiasedly modified) first-order efficient estimator $\hat{u}^*$ by

$$(3.19)\quad \mathrm{E}\,[\tilde{u}^{*a}\tilde{u}^{*b}]=g^{ab}+\frac{1}{N}C^{2ab}+o(N^{-1})=g^{ab}+\frac{1}{N}\{(\overset{\mathrm{m}}{\Gamma}{}^2)^{ab}+2(\overset{\mathrm{e}}{H}{}^2)^{ab}+(\overset{\mathrm{m}}{H}{}^2)^{ab}\}+o(N^{-1})\,,$$

where

$$(3.20)\quad (\overset{\mathrm{m}}{\Gamma}{}^2)^{ab}=\overset{\mathrm{m}}{\Gamma}{}^{a}_{cd}\overset{\mathrm{m}}{\Gamma}{}^{b}_{ef}g^{ce}g^{df}\,,$$

$$(3.21)\quad (\overset{\mathrm{e}}{H}{}^2)^{ab}=\overset{\mathrm{e}}{H}{}^{a}_{c\kappa}\overset{\mathrm{e}}{H}{}^{b}_{d\lambda}g^{cd}g^{\kappa\lambda}\,,$$

$$(3.22)\quad (\overset{\mathrm{m}}{H}{}^2)^{ab}=\overset{\mathrm{m}}{H}{}^{a}_{\kappa\lambda}\overset{\mathrm{m}}{H}{}^{b}_{\nu\mu}g^{\kappa\nu}g^{\lambda\mu}\,,$$

are positive semi-definite matrices. The first term of the right-hand side in (3.19) shows that the estimator is (first-order) efficient. The remainings are the second-order terms. (These terms are called the third-order terms in some literatures, where terms of order $1/\sqrt{N}$ are said to be second-order terms.) We have from this the following theorem, which summarizes the higher-order efficiency of an first-order efficient estimator, by taking account that the mixture curvature $\overset{\mathrm{m}}{H}_{\kappa\lambda a}$ of the associated ancillary subspaces vanishes for the maximum likelihood estimator (Amari [3]).

THEOREM 6. *The second-order mean square error of the unbiased version of a first-order efficient estimator is decomposed into three non-negative terms. One is the square of the mixture connection of M, which is the Bhattacharyya bound, sometimes called the naming curvature. There is a parametrization of M such that this term vanishes identically, when and only when the mixture Riemann-Christoffel curvature tensor of M vanishes. Another term is the square of the exponential curvature of M. This is a tensor, and the above two are common to all the first-order efficient estimators. The third is the square of the mixture curvature of the ancillary subspace* $A(u_0)$, *which depends on the estimator or the associated A. This term vanishes for the maximum likelihood estimator, showing that it is second-order most efficient among all the (unbiasedly modified) first-order efficient estimators.*

We can prove the stronger assertion that the unbiasedly modified maximum likelihood estimator minimizes the expected loss $\mathrm{E}[f(\tilde{u}^*)]$ for any convex loss function. Moreover, it can be easily shown that the amount of information loss Δg_{ab} by taking a first-order efficient estimator is given by

$$\Delta g_{ab} = (\overset{\mathrm{e}}{H}{}^2)_{ab} + \frac{1}{2}(\overset{\mathrm{m}}{H}{}^2)_{ab}\,, \tag{3.23}$$

and is minimized for the maximum likelihood estimator (Amari [3]). However, we can prove that there exist no estimators which uniformly minimize the information loss of order N^{-1}, by the use of (3.7).

The distribution of the unmodified $\tilde{u}$ is given by replacing $\tilde{u}^*$ in (3.15) by

$$\tilde{u}'^a = \tilde{u}^a + \frac{1}{2\sqrt{N}} C^a$$

and adding the term $-(\partial_a C^d) g_{bd} h^{ab}/2N$. The distribution of $\tilde{u}^{**}$ is the same as that of $\tilde{u}^*$ up to the terms of order N^{-1}.

3.3. *Edgeworth expansion in a locally orthogonal ancillary family*

We can construct a higher-order asymptotic theory of testing statistical hypothesis in a similar manner. Let us consider a test for the null hypothesis $H_0: u = u_0$ against $H_1: u \neq u_0$. We can associate an ancillary family A with a test T such that the critical region R of the test T is bounded by some of $A(u)$'s in A, where the sample space X is identified with S by $\eta = x$ as before.

We can easily prove that a test T is first-order most powerful when the associated A is orthogonal at u_0. However, an orthogonal A does not in general gives the most powerful test for higher-order, in contrast with the problem of estimation. In order to analyze the second- and third-order powers of a test T at

$$u_t = u_0 + te/\sqrt{N}\ , \tag{3.24}$$

where e is a vector satisfying

$$g_{ab}(u_0)e^a e^b = 1\ ,$$

we need to know the Edgeworth expansion of the distribution of $\tilde{u} = \sqrt{N}(\tilde{u} - u_t)$ or its unbiasedly modified version $\tilde{u}^*$, when the true parameter of the distribution is u_t. The point is that the ancillary subspace is orthogonal at u_0, but not u_t.

THEOREM 7. *The Edgeworth expansion of $\tilde{u}^*$ at u_t, when A is locally orthogonal at u_0, is given by*

$$\begin{aligned} p(\tilde{u}^*;\, u_t) = \varphi[\tilde{u}^*;\, g_{ab}(u_t)]\Big\{1 &+ \frac{1}{6\sqrt{N}} K_{abc}(u_t) h^{abc}(\tilde{u}^*) \\ &+ \frac{t}{2N}(2Q_{ab\kappa} - \overset{e}{H}_{ab\kappa})Q_{cd\lambda} g^{\kappa\lambda} e^d h^{abc} \\ &+ \frac{1}{4N}[\overset{\circ}{C}{}^2_{ab} + 2Q_{ca\kappa}(Q_{db\lambda} g^{cd} - 2\overset{e}{H}_{db\lambda} g^{cd} \\ &+ t^2 Q_{db\lambda} e^c e^d) g^{\kappa\lambda} + \overset{m}{H}{}^2_{ab}] h^{ab} \\ &+ \frac{1}{24N}[\overset{\circ}{K}_{abcd} + 12 Q_{ab\kappa}(Q_{cd\lambda} - \overset{e}{H}_{cd\lambda}) g^{\kappa\lambda}] h^{abcd} \\ &+ \frac{1}{72N} K_{abc} K_{def} h^{abcdef}\Big\} + O(N^{-3/2})\ , \end{aligned} \tag{3.25}$$

where

$$\overset{\circ}{C}{}^2_{ab} = \overset{m}{\Gamma}_{cda} \overset{m}{\Gamma}_{efb} g^{ce} g^{df} + 2\overset{e}{H}_{ac\kappa} \overset{e}{H}_{bd\lambda} g^{cd} g^{\kappa\lambda}\ ,$$

$$\overset{\circ}{K}_{abcd} = S_{abcd} - 4D_{abcd} + 12(\overset{m}{\Gamma}_{eab} + \overset{e}{\Gamma}_{abe}) \overset{m}{\Gamma}_{fcd} g^{ef}\ .$$

The proof can be accomplished from (3.7) by careful calculations, taking the relation

$$g_{a\kappa}(u_t) = \frac{t}{\sqrt{N}} Q_{ba\kappa} e^b$$

into account. See Appendix 3. The distribution depends on A only through two quantities $\overset{m}{H}_{\kappa\lambda a}$ and $Q_{ab\kappa}$ up to the order of N^{-1}. We will show a higher-order asymptotic theory of statistical test and a theory of interval estimation by applying the above result (Kumon and Amari [15]).

3.4. *Distribution of $\tilde{v}^*$ and higher-order ancillarity*

We can similarly obtain the distribution of $\tilde{v}^*$ by integrating (3.7) with respect to $\tilde{u}^*$.

THEOREM 8. *The distribution of $\tilde{v}^*$ can be expanded for an orthogonal A as*

$$p(\tilde{v}^*; u_0)=\varphi[\tilde{v}^*; g_{\kappa\lambda}(u_0)]\left\{1+\frac{1}{6\sqrt{N}}K_{\kappa\lambda\mu}h^{\kappa\lambda\mu}(\tilde{v}^*)+\frac{1}{4N}C^2_{\kappa\lambda}h^{\kappa\lambda}(\tilde{v}^*)+\frac{1}{24N}K_{\kappa\lambda\mu\nu}h^{\kappa\lambda\mu\nu}(\tilde{v}^*)+\frac{1}{72N}K_{\kappa\lambda\mu}K_{\nu\tau\sigma}h^{\kappa\lambda\mu\nu\tau\sigma}(\tilde{v}^*)\right\}+O(N^{-3/2}), \tag{3.26}$$

where

$$C^2_{\kappa\lambda}=\overset{\mathrm{m}}{\Gamma}_{\nu\mu\kappa}\overset{\mathrm{m}}{\Gamma}{}^{\nu\mu}_{\lambda}+\overset{\mathrm{m}}{H}_{ab\kappa}\overset{\mathrm{m}}{H}{}^{ab}_{\lambda}+\frac{1}{2}T_{a\nu\kappa}T^{a\nu}_{\lambda}. \tag{3.27}$$

The Fisher information which the statistic $\tilde{v}^*$ carries is measured by the matrix

$$I_{ab}=\mathrm{E}\,[\partial_a l(\tilde{v}^*; u_0)\partial_b l(\tilde{v}^*; u_0)], \tag{3.28}$$

where

$$l(\tilde{v}^*; u_0)=\log p(\tilde{v}^*; u_0).$$

As can easily be shown from (3.26), $I_{ab}=O(1)$. The statistic $\tilde{v}^*$ which summarizes N independent observations $x_1, \cdots, x_N$ carries the Fisher information of only order 1. (The statistic $\tilde{u}^*$ obtained by summarizing N observations carries the Fisher information of order N.) Hence, $\tilde{v}^*$ can be said to be an approximately ancillary statistic. We call a statistic T an ancillary statistic of order p, when it carries the Fisher information of order N^{-p}. The above $\tilde{v}^*$ is an ancillary statistic of order 0 in this sense.

Given an ancillary family A, we can take any coordinate system v^κ in each $A(u)$, especially such coordinates v^κ in each $A(u)$ that

$$g_{\kappa\lambda}(u)=\text{const.} \tag{3.29}$$

holds for all u at $v=0$. The statistic $\tilde{v}^*$ in this coordinate system is an ancillary statistic of order 1, because the Fisher information of this $\tilde{v}^*$ is of $O(N^{-1})$. We can further modify the coordinate system in each $A(u)$ such that

(3.30) $$K_{\kappa\lambda\mu}(u)=-3\overset{\alpha}{\Gamma}_{\kappa\lambda\mu}(u)=0 \qquad (\alpha=-1/3),$$

holds, by taking the normal coordinate system of the $\alpha=-1/3$ connection. We have in this case the second-order ancillary statistic.

We cannot proceed further to get a higher-order ancillary statistic. It is indeed possible to modify the coordinate system such that $K_{\kappa\lambda\mu\nu}(u)=0$ holds at $v=0$ for all u, besides (3.29) and (3.30). However, the term $C^2_{\kappa\lambda}(u)$ still remains, and it is in general impossible to have a set of coordinate systems in all the $A(u)$ such that $g_{\kappa\lambda}(u)$ and $C^2_{\kappa\lambda}(u)$ do not depend on u at the same time. This yields the following generalization of the result of Amari [5].

THEOREM 9. *A maximal set of ancillary statistics of the* 0*th, first, and second order can always be given by the above procedures. However, third-order ancillary statistics do not in general exist.*

3.5. *Conditional distribution and mutual information*

From (3.7), (3.15) and (3.26), the following relation easily follows for an orthogonal family A,

(3.31) $$\frac{p(\tilde{u}^*, \tilde{v}^*; u_0)}{p(\tilde{u}^*; u_0)p(\tilde{v}^*; u_0)}=1+\frac{1}{2\sqrt{N}}(K_{ab\kappa}h^{ab\kappa}+K_{a\kappa\lambda}h^{a\kappa\lambda})+\frac{1}{N}R +O(N^{-3/2}),$$

where the term R is given by

(3.32) $$\begin{aligned} R=&\frac{1}{4}[C^2_{\alpha\beta}h^{\alpha\beta}-C^2_{ab}h^{ab}-C^2_{\kappa\lambda}h^{\kappa\lambda}] \\ &+\frac{1}{24}[K_{\alpha\beta\gamma\delta}h^{\alpha\beta\gamma\delta}-K_{abcd}h^{abcd}-K_{\kappa\lambda\mu\nu}h^{\kappa\lambda\mu\nu}] \\ &+\frac{1}{72}[K_{\alpha\beta\gamma}K_{\delta\epsilon\zeta}h^{\alpha\beta\gamma\delta\epsilon\zeta}-K_{abc}K_{def}h^{abcdef}-K_{\kappa\lambda\mu}K_{\nu\tau\sigma}h^{\kappa\lambda\mu\nu\tau\sigma}] \\ &-\frac{1}{12}(K_{ab\kappa}h^{ab\kappa}+K_{a\kappa\lambda}h^{a\kappa\lambda})(K_{abc}h^{abc}+K_{\kappa\lambda\mu}h^{\kappa\lambda\mu}) \\ &-\frac{1}{36}K_{abc}K_{\kappa\lambda\mu}h^{abc}h^{\kappa\lambda\mu}. \end{aligned}$$

Noting that $K_{ab\kappa}=\overset{e}{H}_{ab\kappa}$, $K_{a\kappa\lambda}=-\overset{m}{H}_{\kappa\lambda a}$, we have the following theorem.

THEOREM 10. *The conditional probability of $\tilde{u}^*$ given the ancillary statistic $\tilde{v}^*$ is expanded as*

(3.33) $$p(\tilde{u}^*; u_0, \tilde{v}^*)=p(\tilde{u}^*; u_0)\left\{1+\frac{1}{2\sqrt{N}}(\overset{e}{H}_{ab\kappa}h^{ab}h^{\kappa}-\overset{m}{H}_{\kappa\lambda a}h^{a}h^{\kappa\lambda})+\frac{1}{N}R\right\} +O(N^{-3/2}).$$

This relation is obtained in Amari [5] up to the term of order $N^{-1/2}$. The above theorem gives the term of order N^{-1}. The above relation is useful for studying problems of conditional inference.

We finally give the Shannon mutual information $I(\hat{U}:\hat{V})$ between a first-order efficient estimator $\hat{u}$ and the related ancillary statistic $\hat{v}$. This is the same as the mutual information between $\tilde{u}^*$ and $\tilde{v}^*$ except for the term of $o(N^{-1})$. The Shannon mutual information $I(\hat{U}:\hat{V})$ is given by

$$I(\hat{U}:\hat{V})=\mathrm{E}\left[\log\frac{p(\tilde{u}^*,\tilde{v}^*;u_0)}{p(\tilde{u}^*;u_0)p(\tilde{v}^*;u_0)}\right] \tag{3.34}$$

By expanding (3.34) by the use of (3.31) and taking account of (3.32), we have

THEOREM 11.

$$I(\hat{U}:\hat{V})=\frac{1}{4N}(\overset{\mathrm{e}}{H}_{ab\kappa}\overset{\mathrm{e}}{H}_{cd\lambda}g^{ac}g^{bd}g^{\kappa\lambda}+\overset{\mathrm{m}}{H}_{\kappa\lambda a}\overset{\mathrm{m}}{H}_{\nu\mu b}g^{\kappa\nu}g^{\lambda\mu}g^{ab})\ . \tag{3.35}$$

APPENDICES

1. *Tensorial Hermite polynomials with metric tensor* $g_{\alpha\beta}$

The tensorial polynomials are shown in Grad [13] in the orthogonal coordinate system, where the metric $g_{\alpha\beta}$ reduces to the unit matrix. Let us consider a general coordinate system in which the metric is given by $g_{\alpha\beta}$, and define the tensorial Hermite polynomials in this coordinate system.

By operating the derivative $D^\alpha=g^{\alpha\beta}(\partial/\partial w^\beta)$ successively to the normal distribution with the covariance $g^{\alpha\beta}$

$$\varphi(w)=(\det|g_{\alpha\beta}|)^{1/2}(2\pi)^{-n/2}\exp\left(-\frac{1}{2}g_{\alpha\beta}w^\alpha w^\beta\right)\ ,$$

we have

$$D^\alpha\varphi(w)=-w^\alpha\varphi(w)\ ,$$

$$D^{\alpha\beta}\varphi(w)=(w^\alpha w^\beta-g^{\alpha\beta})\varphi(w)\ ,$$

$$D^{\alpha\beta\gamma}\varphi(w)=(3g^{(\alpha\beta}w^{\gamma)}-w^\alpha w^\beta w^\gamma)\varphi(w)\ ,$$

and so on, where $D^{\alpha_1\cdots\alpha_k}=D^{\alpha_1}\cdots D^{\alpha_k}$. The results will be, in general, polynomials in w multiplied by $\varphi(w)$. We define the tensorial Hermite polynomial $h^{\alpha_1\cdots\alpha_k}(w)$ by the identity

$$(-1)^k D^{\alpha_1\cdots\alpha_k}\varphi(w)=h^{\alpha_1\cdots\alpha_k}(w)\varphi(w)\ . \tag{A.1.1}$$

Evidently $h^{\alpha_1\cdots\alpha_k}(w)$ is of degree k in w and the coefficient of the highest degree term $w^{\alpha_1}\cdots w^{\alpha_k}$ is unity. By convention $h^0(w)=1$.

We can calculate $\varphi(w-t)$ in the following two ways,

$$\varphi(w-t)=(\det|g_{\alpha\beta}|)^{1/2}(2\pi)^{-n/2}\exp\left(-\frac{1}{2}g_{\alpha\beta}w^\alpha w^\beta+g_{\alpha\beta}t^\alpha w^\beta-\frac{1}{2}g_{\alpha\beta}t^\alpha t^\beta\right)$$
$$=\varphi(w)\exp\left(g_{\alpha\beta}t^\alpha w^\beta-\frac{1}{2}g_{\alpha\beta}t^\alpha t^\beta\right)$$

and by Taylor's theorem

$$\varphi(w-t)=\sum_{k=0}^{\infty}\frac{(-1)^k}{k!}t_{\alpha_1\cdots\alpha_k}D^{\alpha_1\cdots\alpha_k}\varphi(w)=\sum_{k=0}^{\infty}\frac{t_{\alpha_1\cdots\alpha_k}}{k!}h^{\alpha_1\cdots\alpha_k}(w)\varphi(w)\ ,$$

where $t_{\alpha_1\cdots\alpha_k}$ is the covariant version of $t^{\alpha_1\cdots\alpha_k}=t^{\alpha_1}t^{\alpha_2}\cdots t^{\alpha_k}$, indices being raised or lowered by the use of $g^{\alpha\beta}$ or $g_{\alpha\beta}$. Consequently, the Hermite polynomial $h^{\alpha_1\cdots\alpha_k}(w)$ is given by the coefficient of $t_{\alpha_1\cdots\alpha_k}/k!$ in $\exp(g_{\alpha\beta}\cdot t^\alpha w^\beta-(1/2)g_{\alpha\beta}t^\alpha t^\beta)$.

The first six polynomials are

$$h^0=1\ ,\qquad h^\alpha=w^\alpha\ ,\qquad h^{\alpha\beta}=w^\alpha w^\beta-g^{\alpha\beta}\ ,\qquad h^{\alpha\beta\gamma}=w^\alpha w^\beta w^\gamma-3g^{(\alpha\beta}w^{\gamma)}\ ,$$
$$h^{\alpha\beta\gamma\delta}=w^\alpha w^\beta w^\gamma w^\delta-6g^{(\alpha\beta}w^\gamma w^{\delta)}+3g^{(\alpha\beta}g^{\gamma\delta)}\ ,$$
$$h^{\alpha\beta\gamma\delta\epsilon}=w^\alpha w^\beta w^\gamma w^\delta w^\epsilon-10g^{(\alpha\beta}w^\gamma w^\delta w^{\epsilon)}+15g^{(\alpha\beta}g^{\gamma\delta}w^{\epsilon)}\ ,$$
$$h^{\alpha\beta\gamma\delta\epsilon\zeta}=w^\alpha w^\beta w^\gamma w^\delta w^\epsilon w^\zeta-15g^{(\alpha\beta}w^\gamma w^\delta w^\epsilon w^{\zeta)}+45g^{(\alpha\beta}g^{\gamma\delta}w^\epsilon w^{\zeta)}-15g^{(\alpha\beta}g^{\gamma\delta}g^{\epsilon\zeta)}\ .$$

Differentiating the identity

$$\exp\left(g_{\alpha\beta}t^\alpha w^\beta-\frac{1}{2}g_{\alpha\beta}t^\alpha t^\beta\right)=\sum_{k=0}^{\infty}\frac{t_{\alpha_1\cdots\alpha_k}}{k!}h^{\alpha_1\cdots\alpha_k}(w)$$

with respect to w and identifying the coefficients in $t_{\alpha_1\cdots\alpha_k}$, we have

(A.1.2) $$D^\beta h^{\alpha_1\cdots\alpha_k}(w)=kg^{(\beta\alpha_1}h^{\alpha_2\cdots\alpha_k)}(w)\ .$$

From this, we can derive the following important orthogonal property

(A.1.3) $$\int_{-\infty}^{\infty}h^{\alpha_1\cdots\alpha_k}(w)h^{\beta_1\cdots\beta_l}(w)\varphi(w)dw=\begin{cases}k!g^{(\alpha_1|\beta_1|}\cdots g^{\alpha_k)\beta_k}\ , & \text{if } k=l\ ,\\ 0\ , & \text{otherwise}\ ,\end{cases}$$

where the symmetrization ( ) is taken for the indices $\alpha_1,\cdots,\alpha_k$ only. In fact, integrating by parts, we have

$$\int_{-\infty}^{\infty}h^{\alpha_1\cdots\alpha_k}h^{\beta_1\cdots\beta_l}\varphi dw=(-1)^l\int_{-\infty}^{\infty}h^{\alpha_1\cdots\alpha_k}D^{\beta_1\cdots\beta_l}\varphi dw$$
$$=(-1)^{l-1}\int_{-\infty}^{\infty}D^{\beta_1}h^{\alpha_1\cdots\alpha_k}D^{\beta_2\cdots\beta_l}\varphi dw$$

$$=k(-1)^{l-1}\int_{-\infty}^{\infty} g^{(\beta_1\alpha_1}h^{\alpha_2\cdots\alpha_k)}D^{\beta_2\cdots\beta_l}\varphi dw\ .$$

Continuing the integration by parts, we find that the above integration is zero when k is not equal to l and is $k!g^{(\alpha_1|\beta_1|}\cdots g^{\alpha_k)\beta_k}$ when $k=l$.

When w^α consists of u^a- and v^κ- parts, $a=1,\cdots,m$, $\kappa=m+1,\cdots,n$, and when the orthogonality $g_{a\kappa}=0$ holds between them, we have

$$\varphi(w)=(\det|g_{ab}|)^{1/2}(\det|g_{\kappa\lambda}|)^{1/2}(2\pi)^{-n/2}\exp\left(-\frac{1}{2}g_{ab}u^a u^b-\frac{1}{2}g_{\kappa\lambda}v^\kappa v^\lambda\right)$$
$$=\varphi(u)\varphi(v)\ .$$

Hence, from

$$D^{a_1\cdots a_r\kappa_{r+1}\cdots\kappa_k}\varphi(w)=D^{a_1\cdots a_r}\varphi(u)D^{\kappa_{r+1}\cdots\kappa_k}\varphi(v)\ ,$$

we have

$$h^{a_1\cdots a_r\kappa_{r+1}\cdots\kappa_k}(w)=h^{a_1\cdots a_r}(u)h^{\kappa_{r+1}\cdots\kappa_k}(v)\ .$$

Furthermore, the expression $c_{\alpha_1\cdots\alpha_k}h^{\alpha_1\cdots\alpha_k}(w)$ can be decomposed into

$$\sum_{r=0}^{k}\binom{k}{r}c_{a_1\cdots a_r\kappa_{r+1}\cdots\kappa_k}h^{a_1\cdots a_r}(u)h^{\kappa_{r+1}\cdots\kappa_k}(v)\ .$$

Hence, using (A.1.3), we have a useful relation

$$\text{(A.1.4)}\qquad \int_{-\infty}^{\infty} c_{\alpha_1\cdots\alpha_k}h^{\alpha_1\cdots\alpha_k}(w)\varphi(w)dv=c_{a_1\cdots a_k}h^{a_1\cdots a_k}(u)\varphi(u)\ ,$$

which is used to derive (3.15).

2. *Calculations of moments or cumulants of* $\tilde{w}^*$

The coefficients $c^{\alpha_1\cdots\alpha_k}$ in (3.1) are expressed as

$$c^\alpha=\mu^\alpha\ ,\qquad 2c^{\alpha\beta}=\mu^{\alpha\beta}-g^{\alpha\beta}\ ,\qquad 6c^{\alpha\beta\gamma}=\mu^{\alpha\beta\gamma}-3\mu^{(\alpha}g^{\beta\gamma)}\ ,$$
$$24c^{\alpha\beta\gamma\delta}=\mu^{\alpha\beta\gamma\delta}-6g^{(\alpha\beta}\mu^{\gamma\delta)}+3g^{(\alpha\beta}g^{\gamma\delta)}\ ,$$
$$120c^{\alpha\beta\gamma\delta\epsilon}=\mu^{\alpha\beta\gamma\delta\epsilon}-10g^{(\alpha\beta}\mu^{\gamma\delta\epsilon)}+15\mu^{(\alpha}g^{\beta\gamma}g^{\delta\epsilon)}\ ,$$
$$720c^{\alpha\beta\gamma\delta\epsilon\zeta}=\mu^{\alpha\beta\gamma\delta\epsilon\zeta}-15g^{(\alpha\beta}\mu^{\gamma\delta\epsilon\zeta)}+45g^{(\alpha\beta}g^{\gamma\delta}\mu^{\epsilon\zeta)}-15g^{(\alpha\beta}g^{\gamma\delta}g^{\epsilon\zeta)}\ .$$

Since we know that $\mu^\alpha=o(N^{-1})$, $c^{\alpha\beta}=O(N^{-1})$ and the cumulant $\kappa^{\alpha_1\cdots\alpha_k}$ of $\tilde{w}^*$ is of order $N^{-(k-2)/2}$ $(k\geqq 2)$, it is useful to express the $c^{\alpha_1\cdots\alpha_k}$ in terms of the cumulants. Taking the above evaluations of the orders into account, we have the following simple expressions by neglecting the terms of order higher than N^{-1}.

$$2c^{\alpha\beta}=\kappa^{\alpha\beta}-g^{\alpha\beta}\ ,\qquad 3c^{\alpha\beta\gamma}=\kappa^{\alpha\beta\gamma}\ ,\qquad 24c^{\alpha\beta\gamma\delta}=\kappa^{\alpha\beta\gamma\delta}\ ,$$

$$120c^{\alpha\beta\gamma\delta\epsilon}=\kappa^{\alpha\beta\gamma\delta\epsilon}\,, \qquad 720c^{\alpha\beta\gamma\delta\epsilon\zeta}=10\kappa^{(\alpha\beta\gamma}\kappa^{\delta\epsilon\zeta)}\,.$$

For the expressions in terms of the cumulants, see Kendall [14] in the one-dimensional case.

In order to calculate the moments or cumulants of $\tilde{w}^*$, it is better to express $\tilde{w}^*$ by the Hermite polynomials in $\tilde{x}$. The expression is derived by the use of (2.29) and (2.40) as

$$\text{(A.2.1)} \quad \tilde{w}^{*\alpha}=\left(U^{\alpha i}+\frac{1}{N}V^{\alpha i}\right)\tilde{x}_i+\frac{1}{\sqrt{N}}W^{\alpha ij}h_{ij}(\tilde{x})+\frac{1}{N}X^{\alpha ijk}h_{ijk}(\tilde{x})$$

where

$$U^{\alpha i}=g^{\alpha\beta}B^i_\beta\,, \qquad V^{\alpha i}=\frac{1}{2}C^\alpha_{\gamma\delta}T^{\beta\gamma\delta}B^i_\beta\,, \qquad W^{\alpha ij}=-\frac{1}{2}C^{\beta\gamma\alpha}B^i_\beta B^j_\gamma\,,$$

$$X^{\alpha ijk}=\frac{1}{2}\left(C^{\beta\alpha}_\epsilon C^{\gamma\delta\epsilon}-\frac{1}{3}D^{\beta\gamma\delta\alpha}\right)B^i_\beta B^j_\gamma B^k_\delta\,.$$

From (A.2.1), we can calculate the moments of $\tilde{w}^*$ necessary to obtain (3.7).

3. *Proof of Theorem* 7

From (3.7), we have

$$\text{(A.3.1)} \quad p(\tilde{w}^*;u_t)=\varphi[\tilde{w}^*;g_{\alpha\beta}(u_t)]\left\{1+\frac{1}{6\sqrt{N}}K_{\alpha\beta\gamma}(u_t)h^{\alpha\beta\gamma}(\tilde{w}^*)\right.$$
$$+\frac{1}{4N}C^2_{\alpha\beta}h^{\alpha\beta}+\frac{1}{24N}K_{\alpha\beta\gamma\delta}h^{\alpha\beta\gamma\delta}$$
$$\left.+\frac{1}{72N}K_{\alpha\beta\gamma}K_{\delta\epsilon\zeta}h^{\alpha\beta\gamma\delta\epsilon\zeta}\right\}+O(N^{-3/2})\,.$$

It is necessary to integrate (A.3.1) with respect to $\tilde{v}^*$. This is done by the following calculations, where we use $g_{a\kappa}(u_t)=tQ_{ba\kappa}e^b/\sqrt{N}$:

$$\text{(A.3.2)} \quad \int_{-\infty}^{\infty}\varphi[\tilde{w}^*;g_{\alpha\beta}(u_t)]d\tilde{v}^*$$
$$=\varphi[\tilde{u}^*;g_{ab}(u_t)]\left\{1+\frac{t^2}{2N}Q_{ca\kappa}Q_{db\lambda}g^{\kappa\lambda}e^ce^dh^{ab}\right\}+O(N^{-3/2})\,,$$

$$\text{(A.3.3)} \quad \int_{-\infty}^{\infty}\varphi[\tilde{w}^*;g_{\alpha\beta}(u_t)]K_{\alpha\beta\gamma}(u_t)h^{\alpha\beta\gamma}(\tilde{w}^*)d\tilde{v}^*$$
$$=\varphi[\tilde{u}^*;g_{ab}(u_t)]\left\{K_{abc}(u_t)h^{abc}(\tilde{u}^*)+\frac{t}{2\sqrt{N}}(2Q_{ab\kappa}\right.$$
$$\left.-\overset{e}{H}_{ab\kappa})Q_{cd\lambda}g^{\kappa\lambda}e^dh^{abc}\right\}+O(N^{-1})\,,$$

where the second term arises from the integral of $K_{ab\kappa}h^{ab\kappa}$, and those derived from $K_{a\kappa\lambda}h^{a\kappa\lambda}$ and $K_{\kappa\lambda\mu}h^{\kappa\lambda\mu}$ become of order N^{-1}.

(A.3.4)
$$\int_{-\infty}^{\infty} \varphi(\tilde{w}^*)C^2_{\alpha\beta}h^{\alpha\beta}d\tilde{v}^*$$
$$=\varphi(\tilde{u}^*)C^2_{ab}h^{ab}+O(N^{-1/2})$$
$$=\varphi(\tilde{u}^*)\{\mathring{C}^2_{ab}+2Q_{ca\kappa}(Q_{db\lambda}-2\overset{e}{H}_{db\lambda})g^{cd}g^{\kappa\lambda}+\overset{m}{H}{}^2_{ab}\}h^{ab}+O(N^{-1/2}) ,$$

because of (2.34), where

$$\mathring{C}^2_{ab}=\overset{m}{\Gamma}_{cda}\overset{m}{\Gamma}_{efb}g^{ce}g^{df}+2\overset{e}{H}_{ac\kappa}\overset{e}{H}_{bd\lambda}g^{cd}g^{\kappa\lambda} , \qquad \overset{m}{H}{}^2_{ab}=\overset{m}{H}_{\kappa\lambda a}\overset{m}{H}_{\mu\nu b}g^{\kappa\mu}g^{\lambda\nu} .$$

(A.3.5)
$$\int_{-\infty}^{\infty} \varphi(\tilde{w}^*)K_{\alpha\beta\gamma\delta}h^{\alpha\beta\gamma\delta}d\tilde{v}^*$$
$$=\varphi(\tilde{u}^*)K_{abcd}h^{abcd}+O(N^{-1/2})$$
$$=\varphi(\tilde{u}^*)\{\mathring{K}_{abcd}+12Q_{ab\kappa}(Q_{cd\lambda}-\overset{e}{H}_{cd\lambda})g^{\kappa\lambda}\}h^{abcd}+O(N^{-1/2}) ,$$

where

$$\mathring{K}_{abcd}=S_{abcd}-4D_{abcd}+12(\overset{m}{\Gamma}_{eab}+\overset{e}{\Gamma}_{abe})\overset{m}{\Gamma}_{fcd}g^{ef} .$$

(A.3.6)
$$\int_{-\infty}^{\infty} \varphi(\tilde{w}^*)K_{\alpha\beta\gamma}K_{\delta\epsilon\zeta}h^{\alpha\beta\gamma\delta\epsilon\zeta}d\tilde{v}^*=\varphi(\tilde{u}^*)K_{abc}K_{def}h^{abcdef}+O(N^{-1/2}) .$$

Summarizing (A.3.2)-(A.3.6), we have the Edgeworth expansion given by (3.25).

UNIVERSITY OF TOKYO

REFERENCES

[1] Akahira, M. and Takeuchi, K. (1981). Asymptotic efficiency of statistical estimators: Concepts and higher order asymptotic efficiency, *Lecture Notes in Statistics*, No. 7, Springer.

[2] Amari, S. (1968). Theory of information spaces—a geometrical foundation of the analysis of communication systems, *RAAG Memoirs* (ed. K. Kondo), Gakujutsu Bunken Fukyukai, Tokyo, **4**, F-X, 378-418.

[3] Amari, S. (1980). Theory of information space: A differential-geometrical foundation of statistics, *Post-RAAG Reports*, 106.

[4] Amari, S. (1982). Differential geometry of curved exponential families—curvatures and information loss, *Ann. Statist.*, **10**, 357-385.

[5] Amari, S. (1982). Geometrical theory of asymptotic ancillarity and conditional inference, *Biometrika*, **69**, 1-17.

[6] Barndorff-Nielsen, O. (1978). *Information and Exponential Families in Statistical Theory*, John Wiley, New York.

[7] Chentsov, N. N. (1972). *Statistical Decision Rules and Optimal Conclusions* (in Russian), Nauka, Moscow.

[8] Cox, D. R. (1980). Local ancillarity, *Biometrika*, **67**, 279-286.

[9] Dawid, A. P. (1975). Discussions to Efron's paper, *Ann. Statist.*, **3**, 1231-1234.

[10] Efron, B. (1975). Defining the curvature of statistical problem (with application to

second order efficiency), *Ann. Statist.*, **3**, 1189-1242.

[11] Efron, B. (1978). The geometry of exponential families, *Ann. Statist.*, **6**, 362-376.

[12] Efron, B. and Hinkley, D. V. (1978). Assessing the accuracy of the maximum likelihood estimator: Observed versus expected Fisher information, *Biometrika*, **65**, 457-487.

[13] Grad, H. (1949). Note on N-dimensional Hermite polynomials, *Commun. Pure Appl. Math.*, **3**, 325-330.

[14] Kendall, M. G. and Stuart, A. (1963). *The Advanced Theory of Statistics*, **1**, Charles Griffin & Co., Ltd., London.

[15] Kumon, M. and Amari, S. (1982). Geometrical theory of higher-order asymptotics of test, interval estimator and conditional inference, *Proc. Roy. Soc.*, **A**, in press.

[16] Madsen, L. T. (1979). The geometry of statistical models—a generalization of curvature, *Res. Rep.*, 79-1, *Statist. Res. Unit.*, Danish Medical Res. Council.

[17] Pfanzagl, J. (1980). Asymptotic expansions in parametric statistical theory, *Developments in Statistics*, Vol. 3 (ed. P. R. Krishnaiah), Academic Press, Chap. 1, 1-97.

[18] Rao, C. R. (1945). Information and accuracy attainable in the estimation of statistical parameters, *Bull. Calcutta Math. Soc.*, **37**, 81-91.

[19] Schouten, J. A. (1954). *Ricci-Calculus*, 2nd ed., Springer, Berlin.