TECHNICAL REPORTS

Differential Geometry of Smooth Families of Probability Distributions

Hiroshi NAGAOKA and Shun-ichi AMARI

METR 82-7 October 1982

MATHEMATICAL ENGINEERING SECTION

DEPARTMENT OF MATHEMATICAL ENGINEERING AND INSTRUMENTATION PHYSICS
FACULTY OF ENGINEERING, UNIVERSITY OF TOKYO
BUNKYO-KU, TOKYO, JAPAN

Differential Geometry of Smooth Families of Probability Distributions

(Abbreviation: Geometry of Probability Distributions)

Hiroshi Nagaoka* and Shun-ichi Amari

University of Tokyo

"Summary" A smoothly parametrized family of probability distributions forms a manifold. Its differential-geometrical structures are elucidated by introducing a Riemannian metric and one-parameter families of affine connections (α-connections). There exists duality between α- and $-\alpha$-connections, so that an α-flat manifold is automatically $-\alpha$-flat. In an α-flat manifold, a natural quasi-distance, called the α-divergence can naturally be introduced from the intrinsic dualistic structure. When $\alpha = -1$, this reduces to the Kullback divergence, and when $\alpha = 0$ it is the Hellinger distance (which in this case is related to the Riemannian distance). The geometry of α-divergence is connected with the α- and $-\alpha$-geodesics due to the α-and $-\alpha$-connections. It is important in many statistical problems to approximate a distribution by one belonging to a prescribed family of distributions that is closest to the distribution in the sense of the α-divergence. This problem of α-approximation is solved with the help of the α-geodesic and $-\alpha$-geodesic. The geometrical structures of the function space of distributions are also touched upon.

AMS 1980 subject classifications. Primary 60-E05; Secondary 53-B-05, 49-F22, 62-B10.

Key words and phrases. Geometry of probability distributions, α-connection, α-divergence, α-geodesic, α-family, exponential family, Fisher information, differential geometry of statistics.

1. Introduction

It is important to study the properties which a family of probability distributions possess as a whole, since a parametrized statistical model constitutes a family of distributions. Such a family forms a geometrical manifold in many cases, and geometric structures, such as Riemannian metric, affine connections, divergences of two points, can be shown to be introduced naturally. These intrinsic geometrical structures represent some important properties of the family, which are not mere aggregates of the properties of the distributions in the family but are given rise to by the mutual relations of the distributions. The present paper aims at elucidating the intrinsic geometrical structures of parametric families of probability distributions from the differential-geometrical point of view.

Rao [20] seems to be the first who explicitly introduced a Riemannian metric in a model of probability distributions with the help of Fisher information matrix. Čencov [6] used the differential-geometrical method to study the properties of families of probability distributions by introducing affine connections, while Csiszár [7], [8] used the f-divergence including Kullback divergence [16] as a special case as a quasi squared distances to study the geometry of a family of distributions. (See also Ingarden [13].) However, statistical roles of these geometrical structures have not yet been sufficiently elucidated and their important works have remained not widely known.

It was Efron [11] who studied the role of the curvature of a statistical model in the framework of theory of estimation. Dawid [9], [10] pointed out the geometrical fertility of Efron's rather intuitive approach. Amari [1], [2], [3] defined a one-parameter family of affine connections (α-connections), which turned out to be tightly connected with Čencov's, to give a differential-geometrical foundation of the asymptotic theory of statistical inference. (See also Madsen [19]). The work has further been developed in related papers (Amari and Kumon [4], Kumon and Amari [17]), where the curvatures due to these connections play a fundamental role. Amari [3] has also defined α-quasi-distances or α-divergences between two probability distributions, and has shown that the α-geodesic is essentially related to the criterion of minimizing the α-divergence. Thus, the α-geodesic is shown to play the role of the straight line in the geometry of the α-divergence.

All of the above works suggest that there exist beautiful geometrical structures in a family of probability distributions. We elucidate in the present paper such structures by introducing a Riemannian metric and α-connections (α being a real prameter) in the manifolds of parametrized families of not only probability distributions but also of finite measures. There exists a dualistic structure between α- and $-\alpha$-connections, so that we can prove that an α-flat manifold is $-\alpha$-flat. The geometry of an α-flat manifold is studied, in which it is proved that α-affine and $-\alpha$-affine coordinate systems are connected by the Legendre transformation. Moreover, α-divergence is naturally defined in an α-flat space, which reduces to the Kullback divergence for $\alpha = -1$, to Hellinger distance for $\alpha = 0$, and equivalent to the Chernoff distance in general. It is proved that the α-geodesic (complemented by $-\alpha$-geodesic) plays a role of the straight line in the geometry of the α-distance. Thus, along the line shown in the appendix of Amari [3], we generalize and unify Csiszar's geometry, Cencov's geometry, and Amari's arguments in the sense that the intrinsic relation of the α-divergence and α-connections are elucidated. The properties of the α-projection or α-approximation, by which a distribution is mapped to one belonging to a family of distributions which is closest to the former in the sense of the α-divergence, are fully studied.

It seems important to extend the theory to an infinite-dimensional function space of distributions. However, there are some difficulties in constructing the differential-geometry of such function spaces (see Čencov [6], Lang [18]). We touch upon the geometrical structures of such spaces. We refer to [14], [15], [23] for differential geometry.

2. Differential Geometry of Parametrized Measures

2.1. Manifolds of measures

Let μ be a measure defined on a σ-field of a sample space X, and let M be a set of finite measures which are absolutely continuous with respect to μ and are parametrized by $\theta \in \Theta$, where Θ is an open set in R^n. The members of M can be represented by the density functions $m(x,\theta)$ with respect to μ, $x \in X$, satisfying $\int m(x, \theta) d\mu(x) < \infty$. It is assumed that the natural mapping $\tau : \Theta \to M$, $\tau\theta = m(x, \theta)$ is a bijection between Θ and M. Furthermore, the following regularity conditions are assumed:

i) $m(x,\theta)$ has a common support X_0 in X, i.e., $m(x,\theta) > 0$, $x \in X_0$ and $m(x,\theta) = 0$, $x \notin X_0$, for all $\theta \in \Theta$.

ii) $m(x,\theta)$ is sufficiently smooth in θ.

We can introduce M the structure of an n-dimensional differentiable manifold, by considering $\tau^{-1} : M \to \Theta$ as a coordinate function defining a coordinate system $\theta = (\theta^i)$, $i = 1, \ldots, n$, in M (cf. [14]). Moreover, it is assumed that Θ itself is diffeomorphic to R^n, so that θ gives a coordinate system for the entire M, since we study only local structures of M in the present paper.

Let $T_\theta(M)$ be the tangent space of M at θ. It is spanned by the natural basis $\{\partial_i\}$, where ∂_i is the abbreviation for $\partial/\partial\theta^i$. We use

$$\ell(x, \theta) = \log m(x, \theta)$$

to define a linear space T_θ^1 spanned by the n functions $\partial_i \ell(\cdot, \theta)$ in x, i = 1, ..., n. It is further assumed that

iii) n functions $\partial_i \ell(\cdot, \theta)$ are linearly independent, for any θ, in the linear space of functions in x.

We then have a natural isomorphism between T_θ and T_θ^1 by which $\partial_i \in T_\theta$ and $\partial_i \ell(x, \theta) \in T_\theta^1$ correspond to each other. We call T_θ^1 the 1-representation of T_θ. A vector $A = A^i \partial_i \in T_\theta$ corresponds to

$$A\ell(x, \theta) = A^i \partial_i \ell(x, \theta) \in T_\theta^1 ,$$

where, and throughout the paper, the Einstein summation convention is used so that the summation is automatically taken for those indices, such as i in the above, that appear twice in one term once as upper and once as lower indices.

Let $\mathcal{T}(M)$ be the set of smooth vector fields on M, and let A, B be its elements. Then, $AB\ell(x,\theta)$ at θ is a function in x but it does not in general belong to $T_\theta{}^1$. In the component form with $A = A^i(\theta)\partial_i$, $B = B^i(\theta)\partial_i$,

$$AB\ell = A^i(\partial_i B^j)\partial_j \ell(x, \theta) + A^i B^j \partial_i \partial_j \ell(x, \theta) .$$

2.2. Metric and α-connections

A metric is introduced in T_θ by defining the inner product A, B of two vectors A and B by

$$\langle A, B \rangle = E_\theta[A\ell B\ell] , \tag{2.1}$$

where $E_\theta[f(x)]$, which sometimes is abbreviated as $E[f(x)]$, is the integration with respect to $m(x,\theta)$,

$$E_\theta[f(x)] = \int f(x)m(x,\theta)d\mu .$$

The existence of integrations is assumed, and the commutativity of the integration with respect to μ and the derivation with respect to θ is also assumed. The metric tensor $g_{ij}(\theta)$ in the coordinate system θ is given by

$$g_{ij} = \langle \partial_i, \partial_j \rangle = E[\partial_i \ell \partial_j \ell] , \tag{2.2}$$

which is a positive-definite matrix because of iii). This reduces to the Fisher information matrix, when M is a space of probability distributions. (Rao [20], Čencov [6], Amari [1], [3]).

We next introduce in M a one-parameter family of affine connections as follows (Amari [1], [3]). Let A, B, C be vector fields. The α-covariant derivative $\nabla^{\alpha}_{A}B$ of B in the direction of A, where α is a real parameter, is defined by

$$\langle \Delta^{\alpha}_{A}B,\ C \rangle = E[AB\ell C\ell] + \frac{1-\alpha}{2} E[A\ell B\ell C\ell] \ . \tag{2.3}$$

The components of the related α-connection Γ^{α}_{ijk} is written, in the coordinate system θ, as

$$\Gamma^{\alpha}_{ijk}(\theta) = \langle \nabla_{\partial i}\partial_j,\ \partial_k \rangle$$

$$= E[\partial_i \partial_j \ell(x,\ \theta)\partial_k \ell(x,\ \theta)] + \frac{1-\alpha}{2} E[\partial_i \ell \partial_j \ell \partial_k \ell] \ . \tag{2.4}$$

The covariant derivative $\nabla^{\alpha}_{A}B$ can be regarded as the orthogonal projection of the function $AB\ell + (1/2)(1-\alpha)A\ell B\ell$ in x to ${T_\theta}^1$ in the linear function space $\mathcal{F}$ in x, where the inner product of two functions f(x) and g(x) is given by $E_\theta[fg]$.

The torsion T is a mapping from $\mathcal{T} \times \mathcal{T}$ to $\mathcal{T}$ defined by

$$T(A,\ B) = \nabla^{\alpha}_{A}B - \nabla^{\alpha}_{B}A - (AB - BA) \ . \tag{2.5}$$

By using the ${T_\theta}^1$-representation, we have

$$\langle T(A,B)\ell,\ C\ell \rangle = 0$$

for arbitrary vector C. This implies that T(A,B) vanishes identically, or the α-connection is torsion-free. Hence, the α-connection is symmetric,

$$\Gamma^{\alpha}_{ijk} = \Gamma^{\alpha}_{jik} \,. \tag{2.6}$$

We give two typical examples.

Example 1. Exponential family. Let M consist of the following density functions

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

where $x = (x_i)$ is a point in $X \subset R^n$, $\theta = (\theta^1, \ldots, \theta^n) \in \Theta$, $\psi(\theta)$ is a smooth convex function satisfying

$$\exp\{\psi(\theta)\} = \int \exp\theta^i x_i d\mu(x) \,, \quad \int m(x,\ \theta) d\mu(x) = 1 \,.$$

This M is composed of a family of probability distributions and is well known as the exponential family in statistics. From the relations

$$\partial_i \ell(x,\ \theta) = x_i - \partial_i \psi(\theta) \,, \quad \partial_i \partial_j \ell(x,\ \theta) = -\partial_i \partial_j \psi(\theta) \,,$$

we have

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

$$\Gamma^{\alpha}_{ijk}(\theta) = \frac{1-\alpha}{2} \partial_i \partial_j \partial_k \psi(\theta) ,$$

after a little complicated calculations (see Amari [3]).

<u>Example</u> 2. Let M_n be the set of measures on a finite number of atoms, x_1, x_2, ..., x_n. Let θ^i be the measure of an atom x_i. We then have $m(x, \theta) = \theta^i \delta_i(x)$ where $\delta_i(x)$ is equal to 1 when $x = x_i$ and is otherwise equal to 0. The M_n forms an n-dimensional manifold with coordinates $\theta = (\theta^i)$, $\theta^i > 0$. From the relations

$$\partial_i \ell(x, \theta) = (\theta^i)^{-1} \delta_i(x) ,$$

$$\partial_i \partial_j \ell(x, \theta) = -(\theta^i)^{-2} \delta_i(x) \delta_j(x) ,$$

where the summation is not taken for i and j, we have

$$g_{ij}(\theta) = (\theta^i)^{-1} \delta_{ij} ,$$

$$\Gamma^{\alpha}_{ijk}(\theta) = -\frac{1+\alpha}{2} (\theta^i)^{-2} \delta_{ik} \delta_{jk} ,$$

δ_{ij} being equal to 1 for $i = j$ and is otherwise equal to 0.

2.3. α-representation

Let $F_\alpha(m)$ be a function whose derivative $F'_\alpha(m)$ is homogeneous of degree $-(1+\alpha)/2$. By the use of this function, we can define

$$(2.7)\qquad \phi^\alpha(x,\theta) = F_\alpha[m(x,\theta)] = \begin{cases} \frac{2}{1-\alpha} m(x,\theta)^{(1-\alpha)/2}, & \alpha \neq 1, \\ \ell(x,\theta), & \alpha = 1. \end{cases}$$

Since n functions

$$(2.8)\qquad \partial_i \phi^\alpha(x,\theta) = m^{(1-\alpha)/2} \partial_i \ell$$

are linearly independent, the linear space T_θ^α spanned by $\partial_i \phi^\alpha(\cdot,\theta)$ is isomorphic to T_θ. The space T_θ^α is called the α-representation of T_θ and vector $A = A^i \partial_i \in T_\theta$ is represented by $A\phi^\alpha = A^i \partial_i \phi(x,\theta) \in T_\theta^\alpha$.

The inner product of A and B is written as

$$(2.9)\qquad \langle A, B \rangle = E_\alpha[A\phi^\alpha B\phi^\alpha] = \int A\phi^\alpha B\phi^{-\alpha} d\mu$$

where the α-integration E_α is defined by

$$E_\alpha[f(x)] = \int [m(x,\theta)]^\alpha f(x) d\mu(x) .$$

The metric tensor is written as

$$(2.10)\qquad g_{ij}(\theta) = E_\alpha[\partial_i \phi^\alpha \partial_j \phi^\alpha] .$$

For vector fields A, B, C, the α-covariant derivative can be written as

$$(2.11) \qquad \langle \nabla^{\alpha}_{A}B,\ C\rangle = E_{\alpha}[AB\phi^{\alpha}C\phi^{\alpha}] = \int AB\phi^{\alpha}C^{\ -\alpha}d\mu\ ,$$

$$(2.12) \qquad \Gamma^{\alpha}_{ijk}(\theta) = E_{\alpha}[\partial_i\partial_j\phi^{\alpha}\partial_k\phi^{\alpha}] = \int\partial_i\partial_j\phi^{\alpha}\partial_k\phi^{-\alpha}d\mu\ .$$

Hence, the α-representation is convenient for studying the properties of the α-connection, because $\nabla^{\alpha}_{A}B \in T_{\theta}$ is represented by the projection of the function $AB\phi^{\alpha}$ to the space $T_{\theta}{}^{\alpha}$. In other words, the vector $\nabla^{\alpha}_{A}B$ is represented in $T_{\theta}{}^{\alpha}$ by $AB\phi^{\alpha} + f(x,\ \theta)$, where $f(x,\ \theta)$ is a function perpendicular to $T_{\theta}{}^{\alpha}$, i.e., $E_{\alpha}[f(x,\ \theta)\ \partial_i\,\phi^{\alpha}(x,\ \theta)] = 0$. It should be remarked that the expressions (2.9) and (2.11) are invariant, because F'_{α} is homogeneous.

2.4. Dual connections and parallel displacements

Let $c : I \to M$ be a smooth curve in M, $c : t \to \theta(t)$ with the tangent vector $\dot{\theta}(t) = \dot{\theta}^{i}\partial_i$ where I is a real interval and . implies the derivative with respect to t. Let A(t) be a vector field along the curve. When it satisfies

$$(2.13) \qquad \nabla_{\dot{\theta}}A(t) = 0$$

for an affine connection with covariant derivative ∇, or

$$\dot{A}^{i} + \Gamma_{jk}{}^{i}A^{k}\dot{\theta}^{j} = 0\ ,$$

in the component form, $A(t_1)$ is said to be the parallel displacement of $A(t_0)$ from $\theta(t_0)$ to $\theta(t_1)$ along the curve c. We denote the parallel displacement along curve c by $\Pi_c : T_\theta \to T_\theta$,

$$\Pi_c A(t_0) = A(t_1)$$

Let ∇ and ∇^* be the covariant derivatives of two affine connections in a manifold with a Riemannian metric. When

$$(2.14) \qquad A\langle B, C\rangle = \langle \nabla_A B, C\rangle + \langle B, \nabla_A^* C\rangle$$

holds for any vector fields A, B, C, the two connections are said to be dual. Obviously, the duality satisfies $\nabla^{**} = \nabla$. By putting $A = \partial_i$, $B = \partial_j$, $C = \partial_k$, the component form of the duality is obtained as

$$(2.15) \qquad \Gamma_{ijk} + \Gamma^*_{ikj} = \partial_i g_{jk} .$$

This proves the existence and uniqueness of the dual of a connection. A connection is self-dual, i.e., $\nabla = \nabla^*$, when and only when it is metric.

Let c be a smooth curve connecting two points θ and θ', and let Π_c and $\Pi_c{}^*$ be, respectively, the parallel displacements with respect to the connections ∇ and ∇^* from θ to θ' along the curve c.

Lemma 1. For any $A, B \in T_\theta$,

$$(2.16) \qquad \langle A, B\rangle_\theta = \langle \Pi_c A, \Pi^*_c B\rangle_{\theta'},$$

where $\langle A, B \rangle_\theta$ is the inner product at the tangent space T_θ.

Proof. Let A(t) and B(t) be the parallel displacements of $A, B \in T_\theta$ along the curve c with respect to the connections ∇ and ∇^*, respectively. From $\nabla_{\dot\theta}A(t) = 0$, $\nabla^*_{\dot\theta}B(t) = 0$, we obtain the relation

$$\frac{t}{dt}\langle A(t), B(t)\rangle = \dot\theta \langle A(t), B(t)\rangle$$

$$= \langle \nabla_{\dot\theta}A, B\rangle + \langle A, \nabla^*_{\dot\theta}B\rangle = 0 ,$$

which proves the lemma.

This shows that Π_c and $\Pi_{-c}{}^*$ are adjoint operators, where $\Pi_{-c}{}^*$ is the parallel displacement from θ' to θ along c with respect to ∇^*, in the sense

$$\langle \Pi_c A, B\rangle_{\theta'} = \langle A, \Pi^*_{-c}B\rangle_\theta \quad .$$

Theorem 1. The α-connection and $-\alpha$-connection are mutually dual. The 0-connection is self-dual, and is the Riemannian connection derived from the metric.

Proof. For vector fields A, B, C, we have

$$A\langle B, C\rangle = A\int B\phi^{\alpha}C\phi^{-\alpha}d\mu = \langle \nabla^{\alpha}_{A}B, C\rangle + \langle B, \nabla^{-\alpha}_{A}C\rangle \ .$$

Hence, ∇^{α} and $\nabla^{-\alpha}$ are dual. For $\alpha = 0$, ∇^{0} is self-dual so that it is metric. Since ∇^{0} is torsion-free the 0-connection is derived from the Riemannian metric by the Levi-Civita parallelism.

2.5. Submanifolds

Let M' be a smooth submanifold in M. The tangent space $T_{\theta}(M')$ of M' at $\theta \in M'$ is a linear subspace of the tangent space $T_{\theta}(M)$ of M at θ. The geometrical structures, the Riemannian metric and the α-connections, can be defined in M' in the same manner as in M. Let A, B be tangent vectors at θ of M'. They can also be regarded as tangent vectors of M. As can easily be shown from the definition, the inner product $\langle A, B\rangle'$ in $T_{\theta}(M')$ coincides with that in $T_{\theta}(M)$,

$$\langle A, B\rangle' = \langle A, B\rangle = E_{\alpha}[A\phi^{\alpha}B\phi^{\alpha}] .$$

Hence, the metric of M' coincides with that induced from the metric of M.

Let A, B, C be vector fields in M'. From the definitions of the α-covariant derivatives in M' and in M, the relation

$$\langle \nabla'^{\alpha}_{A}B, C\rangle' = E_{\alpha}[AB\phi^{\alpha}C\phi^{\alpha}] = \langle \nabla^{\alpha}_{A}B, C\rangle ,$$

where ∇'^{α} is the covariant derivative of the α-connection in M'. However, this does not imply that $\nabla'^{\alpha}_{A}B$ is equal to $\nabla^{\alpha}_{A}B$. The vector $\nabla^{\alpha}_{A}B$ does not necessarily belong to $T_{\theta}(M')$. The orthogonal projection (with respect to the Riemannian metric) of $\nabla^{\alpha}_{A}B$ to $T_{\theta}(M')$ coincides with $\nabla'^{\alpha}_{A}B$. The α-connection of M' is induced from that of M in this sense.

A submanifold M' of M is said to be α-autoparallel, when, for any vector fields A, B belonging to $\mathcal{J}(M')$, the covariant derivative of B along A in M belongs to $\mathcal{J}(M')$, $\nabla^{\alpha}_{A}B \in T_{\theta}(M')$, i.e., $\nabla^{\alpha}_{A}B = \nabla'^{\alpha}_{A}B$ holds. A one-dimensional α-autoparallel submanifold is said to be an α-geodesic curve. Let c : θ(t) be an α-geodesic curve. Then, it satisfies $\nabla^{\alpha}_{\dot\theta}\dot\theta = 0$ or

$$\ddot\theta^{i} + \Gamma^{\alpha\ i}_{jk}\dot\theta^{j}\dot\theta^{k} = 0$$

by choosing an appropriate parameter t. This parameter is called the α-affine parameter of the geodesic.

A submanifold M' of M is said to be totally geodesic, when any geodesic curve θ(t) in M remains in M' if $\theta = \theta(t_0) \in M'$ and $\dot\theta(t_0) \in T_{\theta}(M')$ at a point θ. Since the present manifolds of measures are torsion-free, an autoparallel submanifold is totally geodesic and vice versa (cf. [15]). Let T'_{θ} be a subvector space of the tangent space T_{θ}. We can then construct an α-autoparallel submanifold M' which includes θ and whose tangent space at θ coincides with T'_{θ}. This M' is composed of all the α-geodesics which pass through θ and which have tangents at θ included in T'_{θ}. We call M' the α-autoparallel submanifold generated by T'_{θ}.

2.6. Manifolds of probability distributions

A manifold P of measures p(x, θ) is called a manifold of probability distributions, when $\int p(x, \theta)d\mu = 1$ holds for all θ. An n-dimensional manifold P of probability distributions can be extended to an (n+1)-dimensional manifold M of finite measures consisting of the density functions

(2.17) $$m(x, \theta, t) = tp(x,\theta) ,$$

where t is positive and (t,θ) forms an (n+1)-dimensional coordinate system of M. This M is called the manifold of measures extended from P, and denoted by M(P). The original P is a submanifold of M(P) specified by $t = 1$.

Let $\partial_0 = \partial/\partial t$ and $\partial_i = \partial/\partial\theta^i$ be the fields of the natural basis vectors associated with the coordinates (t,θ) on the whole M(P). When a vector field $A(\theta) = A^i(\theta)\partial_i$ is defined on P, it can naturally be extended in M(P) by $A(t,\theta) = A^i(\theta)\partial_i$. From

$$\partial_0 \log m(x, \theta, t) = \frac{1}{t}, \quad \langle \partial_0, \partial_0 \rangle = \frac{1}{t^2},$$

$$\langle \partial_0, \partial_i \rangle = \frac{1}{t} E[\partial_i \log p(x, \theta)] = 0 ,$$

it is proved that ∂_0 is the unit normal vector field on the submanifold P in M(P).

Let ∇^α and $\tilde{\nabla}^\alpha$ be, respectively, the α-covariant derivatives in P and in M(P). The following lemma shows their relations.

Lemma 2. For vector fields A, B $\mathcal{T}(P)$, their extensions in $\mathcal{T}[M(P)]$ satisfy

(2.18) $$\tilde{\nabla}^\alpha_A B = \nabla^\alpha_A B - \frac{1+\alpha}{2} \langle A, B \rangle \partial_0$$

(2.19) $$\tilde{\nabla}^\alpha_\partial A = \nabla^\alpha_A \partial_0 = \frac{1-\alpha}{2t} A ,$$

(2.20) $$\tilde{\nabla}^{\alpha}_{\partial} \partial_0 = -\frac{1+\alpha}{2t} \partial_0 .$$

Proof. From the definition, we have

$$t\left\langle \tilde{\nabla}^{\alpha}_{A} B, \partial_0 \right\rangle = E[AB\ell] + \frac{1-\alpha}{2} E[A\ell B\ell]$$

because of $\partial_0 \ell = 1/t$. By using the relation

$$0 = AB/pd\mu = E[AB\ell + A\ell B\ell] ,$$

$$t\left\langle \tilde{\nabla}^{\alpha}_{A} B, \partial_0 \right\rangle = -\frac{1+\alpha}{2} \left\langle A, B \right\rangle$$

is derived. Hence, (2.18) is proved. We can prove (2.19) and (2.20) similarly.

Corollary 1. The space P is autoparallel in the extended M(P) in the sense of the -1-connection ($\alpha = -1$).

Corollary 2. For any fixed θ and for any α the curve $c : t \to \{tp(x, \theta)\}$ is an α-geodesic intersecting P orthogonally at θ.

Theorem 2. Let P' be a submanifold of P and let M' = M(P') and M = M(P) be the extended manifolds of measures from P' and P, respectively. Then, P' is α-autoparallel in P, if and only if M' is α-autoparallel in M. Especially, a curve c in P is an α-geodesic, when and only when its extension M(c) is α-autoparallel in M(P).

Proof. Let A', B' be vector fields of P', and let N be a vector field of P orthogonal to the tangent spaces of P' at P'. Then, P' is α-autoparallel in P, when and only when

$$\langle \nabla^{\alpha}_{A'}, B', N \rangle = 0 \tag{2.21}$$

holds for any A', B', N. The tangent space T(M') of M' is the direct sum of the tangent space T(P') of P' and the vector ∂_0. Hence, M' is α-autoparallel in M, when and only when

$$\begin{aligned} &\langle \tilde{\nabla}^{\alpha}_{A'} B', N \rangle = 0 , \quad \langle \tilde{\nabla}^{\alpha}_{\partial} \partial_0, N \rangle = 0 , \\ &\langle \tilde{\nabla}^{\alpha}_{\partial} A', N \rangle = 0 , \quad \langle \tilde{\nabla}^{\alpha}_{A'} \partial_0, N \rangle = 0 \end{aligned} \tag{2.22}$$

holds for any A', B', N, where A' is extended in M' when necessary. However, the latter three holds automatically because of Lemma 2. Moreover, (2.18) guarantees that (2.21) and the first of (2.22) are equivalent. This proves the theorem.

Corollary of Theorem 2. Let P' be the α-autoparallel submanifold in P generated by subspace $T'_\theta \subset T_\theta(P)$ and let M' = M(P') be its extension. Then, M' is the α-autoparallel submanifold in M(P) generated by the subspace spanned by T'_θ and ∂_0, and $P' = P \cap M'$ holds.

3. Geometry of α-flat Manifold

3.1. α-flat manifold and α-affine coordinates

A torsion-free manifold M is said to be α-flat, when the α-curvature (the curvature due to the α-connection) vanishes. The curvature R associated with covariant derivative ∇ is a mapping from $\mathcal{T}(M) \times \mathcal{T}(M) \times \mathcal{T}(M)$ to $\mathcal{T}(M)$ defined by

$$R(A, B, C) = [\nabla_A, \nabla_B]C - \nabla_{[A,B]}C \tag{3.1}$$

where A, B, C are vector fields and [,] implies the alternation, for example

$$[\nabla_A, \nabla_B] = \nabla_A\nabla_B - \nabla_B\nabla_A , \quad [A, B] = AB - BA .$$

When M is α-flat, there exists a (local) coordinate system θ for which the followings hold,

$$\overset{\alpha}{\nabla}_{\partial_i} \partial_j = 0 , \quad \overset{\alpha}{\Gamma}_{ijk}(\theta) = 0 .$$

On the contrary, when the above holds for some coordinate system θ, M is α-flat. This θ is called an α-affine coordinate system of the α-flat manifold M. All the α-affine coordinate systems are related to each other by affine transformations.

When M is α-flat, the parallel displacement of a vector A from θ_0 to θ_1 does not depend on the curve connecting them, and vice versa. Hence, the parallel displacement of a vector A along a loop is invariant, provided M is simply connected.

Theorem 3. An α-flat manifold M is also $-\alpha$-flat.

Proof. Let c be a loop in M, and let Π_c^α be the operator of the parallel displacement encircling the loop. For any vectors A, B,

$$\langle \Pi_c^\alpha A, \Pi_c^{-\alpha} B \rangle = \langle A, B \rangle$$

holds because of (2.16) and Theorem 1. Since M is α-flat, $\Pi_c^\alpha A = A$, so that $\Pi_c^{-\alpha} B = B$ holds, implying M is $-\alpha$-flat.

3.2. Dual coordinate systems and Legendre transformations

Let $\theta = (\theta^i)$ and $\eta = (\eta_i)$ be two coordinate systems in M, where the lower index is used to describe the components of η. The natural bases of the tangent space $T_\theta(M)$ are given by $\{\partial_i\}$ and $\{\partial^i\}$, respectively, where $\partial_i = \partial/\partial\theta^i$, $\partial^i = \partial/\partial\eta_i$. When they form biorthogonal bases, i.e., when the inner product of ∂_i and ∂^j satisfies, for any θ,

$$\langle \partial_i, \partial^j \rangle = \delta_i^j , \tag{3.2}$$

where δ_i^j is the Kronecker delta, being equal to 1 for $i = j$ and otherwise equal to 0, the coordinate systems θ and η are said to be mutually dual. In this case, from (2.3) and $\partial^j \ell = (\partial\theta^k/\partial\eta_j)\partial_k \ell$, we have the following relations

$$\frac{\partial\theta^i}{\partial\eta_j} = g^{ij} , \quad \frac{\partial\eta_k}{\partial\theta^j} = g_{jk} , \tag{3.3}$$

where the metric tensor g^{ij} defined by $g^{ij}(\eta) = E[\partial^i \ell \partial^j \ell]$ is the inverse of the metric tensor g_{ji}.

Theorem 4. When M has a set of dual coordinate systems (θ, η), there exist convex functions $F(\theta)$ and $G(\eta)$ such that θ and η are obtained from one another by the Legendre transformations

$$\theta^i = \partial^i G(\eta) , \quad \eta_i = \partial_i F(\theta) \tag{3.4}$$

and that the two functions are related to each other by

$$F(\theta) + G(\eta) - \theta^i \eta_i = 0 . \tag{3.5}$$

The metric tensors are derived by

$$g_{ij} = \partial_i \partial_j F, \quad g^{ij} = \partial^i \partial^j G . \tag{3.6}$$

On the contrary, when there exists a coordinate system θ (or η) and a function F (or G) for which (3.6) holds, there exists the dual coordinate systems (θ, η) in M.

Proof. When (θ, η) are dual coordinates, $\partial_i g_{jk} - \partial_j g_{ik} = 0$ holds because of (3.3). Since g_{ij} is symmetric, the Poincare lemma guarantees the existence of the potential function $F(\theta)$, provided M is simply connected. Similarly, the existence of $G(\eta)$ is derived. The relations (3.4) and (3.6) hold for them. The functions F and G are determined to within linear and constant terms, and we can choose them such that (3.5) holds. On the contrary, if there exists a function F for which (3.6) holds, the dual of θ is obtained by $\eta_i = \partial_i F$.

Theorem 5. Let M be a torsion-free flat manifold with respect to an affine connection ∇ whose dual ∇^* is also torsion free, and let θ be a ∇-affine coordinate system. Then, there exists a coordinate system η such that (θ, η) are dual coordinate systems. Moreover, M is flat with respect to the dual connection and η is a ∇^*-affine coordinate system.

Proof. Since θ is ∇-affine, $\Gamma_{ijk} = 0$ and since ∇^* is torsion free or $\Gamma^*_{ijk} = \Gamma^*_{jik}$, we have from (2.15) $\partial_i g_{jk} - \partial_j g_{ik} = 0$. This proves the existence of the potential $F(\theta)$. The related dual coordinate system η is derived from $F(\theta)$. Applying (2.14) again to the natural basis vectors ∂_i of θ and ∂^k of η, we have

$$0 = \partial_i \langle \partial_j, \partial^k \rangle = \langle \nabla_{\partial_i} \partial_j, \partial_k \rangle + \langle \partial_j, \overset{*}{\nabla}_{\partial_i} \partial^k \rangle$$

$$= \langle \partial_j, \overset{*}{\nabla}_{\partial_i} \partial_k \rangle,$$

where $\nabla_{\partial i} \partial_j = 0$ is used because θ is ∇-affine. Hence, we have $\overset{*}{\nabla}_{\partial i} \partial^k = 0$. Since $\nabla^*_A B$ is linear in A, we obtain $\overset{*}{\nabla}_{\partial_i} \partial^k = 0$, which implies η is a ∇^*-affine coordinate system.

Corollary of Theorem 5. When M is α-flat with an α-affine coordinate system θ, there exists a dual coordinate system η which is a −α-affine coordinate system.

3.3. α-divergence

Let M be a manifold admitting dual coordinate systems (θ, η) with potential functions F and G. A function $D : M \times M \to R$ is defined by

$$D(\theta_1, \theta_2) = F(\theta_1) + G(\eta_2) - \theta_1^i \eta_{2i} \tag{3.7}$$

in terms of the coordinates θ_1 and θ_2 of two points in M, where θ_1^i and η_{2i} are the components of θ_1 and $\eta_2 = \eta(\theta_2)$ which is the dual coordinates of θ_2. This function satisfies the following relations.

Lemma 3.

1) $D(\theta, \xi) \geqq 0$. The equality holds when, and only when, $\theta = \xi$.
2) $\partial/\partial\theta^i D(\theta, \xi) = \partial/\partial\xi^i D(\theta, \xi) = 0$, at $\theta = \xi$.
3) $(\partial^2/\partial\theta^i\partial\theta^j)D(\theta, \xi) = g_{ij}(\theta)$.

Proof. By differenting the definition (3.7), we have

$$(\partial/\partial\theta^i)D(\theta, \xi) = \eta_i(\theta) - \eta_i(\xi) \ ,$$

$$(\partial/\partial\xi^i)D(\theta, \xi) = g_{ij}(\xi)(\xi^j - \theta^j) \ ,$$

$$(\partial^2/\partial\theta^i\partial\theta^j)D(\theta, \xi) = g_{ij}(\theta) \ ,$$

which prove 2) and 3). Since g_{ij} is positive-definite, $D(\theta, \xi)$ is a strictly convex function with respect to θ. Hence, $D(\theta, \xi) > 0$ when $\theta \neq \xi$ is proved from $D(\theta, \theta) = 0$ and 2).

When $d\theta$ is "infinitesimally small", the relation

$$D(\theta, \theta+d\theta) = D(\theta+d\theta, \theta) = \frac{1}{2} g_{ij} d\theta^i d\theta^j$$

holds. Hence $2D(\theta, \xi)$ can be regarded as a generalization of the square of the Riemannian distance on M, although it is not symmetric $D(\theta, \xi) \neq D(\xi, \theta)$ in general. It does not necessarily satisfy the triangular inequality. We call $D(\theta_1, \theta_2)$ the divergence from θ_1 to θ_2 in M. Although the potential functions F, G are determined only to within linear terms, the divergence is uniquely determined.

Let M be an α-flat manifold. Then, there exists the dual coordinate systems (θ, η) with functions F and G. Hence, the divergence is introduced in M, which we call the α-divergence and denote it by $D_\alpha(\theta_1, \theta_2)$. Since M is also $-\alpha$-flat, it has the $-\alpha$-divergence $D_{-\alpha}(\theta_1, \theta_2)$, too. Since η is the $-\alpha$-affine coordinate system with potential $G(\eta)$, the $-\alpha$-divergence is written in terms of η as

$$D_{-\alpha}(\eta_1, \eta_2) = G(\eta_1) + F(\theta_2) - \eta_{1i}\theta_2^i ,$$

where $\theta_2 = \theta(\eta_2)$. The following theorem is easy to prove.

Theorem 6. The α-divergence from θ_1 to θ_2 is equal to the $-\alpha$-divergence from θ_2 to θ_1, (3.8) $$D_\alpha(\theta_1, \theta_2) = D_{-\alpha}(\theta_2, \theta_1) .$$

Theorem 7. Let c_1 be the α-geodesic connecting two points θ_2 and θ_1, let c_2 be the $-\alpha$-geodesic connecting two points θ_2 and θ_3 in an α-flat manifold M. Let γ be the angle of the two geodesics c_1 and c_2, i.e., the angle between the tangent vectors of c_1 and c_2 at θ_2. Then

(3.9) $$D_\alpha(\theta_1, \theta_3) \gtreqless D_\alpha(\theta_1, \theta_2) + D_\alpha(\theta_2, \theta_3)$$

according as $\cos\gamma < 0$, $\cos\gamma = 0$ or $\cos\gamma > 0$.

Proof. We have

$$D_\alpha(\theta_1, \theta_2) + D_\alpha(\theta_2, \theta_3) = F(\theta_1) + G(\eta_3) - \theta_1\cdot\eta_3$$

$$+ (\theta_1 - \theta_2)\cdot(\eta_3 - \eta_2) = D_\alpha(\theta_1, \theta_3) + (\theta_1 - \theta_2)\cdot(\eta_3 - \eta_2) ,$$

where $\theta_1\cdot\eta_2$, for example, implies $\theta_1^i\eta_{2i}$. Since θ is α-affine coordinates, the α-geodesic curve c_1 can be written in the form $c_1 : \theta(t) = \theta_2 + t(\theta_1 - \theta_2)$. Hence its tangent vector at θ_2 is given by $\theta_1 - \theta_2 = (\theta_1^i - \theta_2^i)\partial_i$. Similarly, the $-\alpha$-geodesic curve c_2 can be written in the η-coordinates as $c_2 : \eta(t) = \eta_2 + t(\eta_3 - \eta_2)$, and the tangent vector at η_2 is $\eta_3 - \eta_2 = (\eta_{3i} - \eta_{2i})\partial^i$. Hence, $(\theta_1 - \theta_2)\cdot(\eta_3 - \eta_2)$ has the same sign as $\cos\gamma$. This proves the theorem.

The theorem shows that D_α satisfies the Pythagorean relation as the squared distance does in the Euclidean geometry, and $\pm\alpha$-geodesics play the role of straight lines in an α-flat M. This can be shown in the following discussions more clearly.

3.4. α-projection

Let M' be an m-dimensional smooth submanifold of an n-dimensional α-flat manifold M ($m < n$). We consider the problem of obtaining the point $\theta' \in M'$ which is closest to a given point $\theta \in M$ in the sense of the α-divergence $D_\alpha(\theta, \theta')$. This is the problem of approximating a measure $\theta \in M$ by a measure included in M'. Such a problem has been studied by Kullback [16], Csiszár [7], [8] and Čencov [6], where the Kullback divergence is used. Amari [3] used the α-connection to solve the problem where the α-distance is defined as the criterion. He proved that the α-projection to M' gives the best approximation of a distribution by one belonging to M'. Here, the α-projection of θ to M' is the point $\theta' \in M'$ such that the α-geodesic c connecting θ and θ' are orthogonal to M' at θ'. Thus, the role of the α-geodesic or the α-connection is elucidated in the geometry of the α-divergence. We generalize his results by giving a more general and rigorous framework.

A point $\theta' \in M'$ is called an α-extreme point for $\theta \in M$, when it satisfies, for any $A' \in T_{\theta'}(M')$

$$A' D_\alpha(\theta, \theta') = 0 ,$$

where the vector A' operates on the second variable θ',

$$A'D_\alpha(\theta, \theta') = A'^i(\partial/\partial\theta'^i)D_\alpha(\theta, \theta') .$$

The α-divergence $D_\alpha(\theta, \xi')$, $\xi' \in M'$ attains the extreme value at an α-extreme point $\xi' = \theta'$ of θ. Let $O_\alpha(\theta', M')$ be the inverse image of the α-projection. Then, it consists of all the points on the α-geodesics which intersect M' perpendicularly at θ'. We call $O_\alpha(\theta', M')$ the orthogonal α-submanifold at θ' of M', because it is the α-autoparallel submanifold generated by the subspace of $T_{\theta'}$ which are the orthogonal complement of $T_{\theta'}(M')$. When θ belongs to $O_\alpha(\theta', M')$, there is an α-geodesic c connecting θ' and θ, $c : \theta(t) = \theta' + t(\theta - \theta')$ in the α-affine coordinate system θ and the tangent vector $\theta - \theta'$ of c at θ' is perpendicular to any $A' \in T_{\theta'}(M')$, $\langle A', \theta - \theta' \rangle = 0$.

One more definition is necessary before stating the relation between the α-divergence and α-connection. A subset S of M is said to be α-convex, when there exists inside S a unique α-geodesic of M connecting any two points $\theta_1, \theta_2 \in S$.

Theorem 8. For an α-flat α-convex M, the α-projection of $\theta \in M$ to M' gives the α-extreme point, and conversely, if $\theta' \in M'$ is the extreme point for θ, θ' is the α-projection of θ.

Proof. Let θ be a point for which $\theta' \in M'$ is the α-extreme point. Then, for any $A' \in T_{\theta'}(M')$,

$$0 = A'D_\alpha(\theta, \theta') = A'_i\partial^i[G(\eta') - \theta^i\eta'_i] = -\langle A', \theta - \theta' \rangle$$

holds, where (θ, η) are the dual affine coordinate systems. Since M is α-convex, there exists the α-geodesic $\theta(t) = \theta' + t(\theta - \theta')$ connecting θ and θ' and its tangent vector at θ' is $\theta - \theta'$. Hence, θ' is the α-projection of θ. On the contrary, when the α-projection of θ is θ', $A'D_\alpha(\theta, \theta') = 0$ holds, so that θ' is the α-extreme point of θ.

The theorem shows that the α-projection gives the best approximation (the α-approximation) of θ by the distributions in M' in the sense of the α-divergence provided the minimum of $D_\alpha(\theta, \theta')$ is given by the α-extreme point. However, the α-projection might not be unique. We give conditions which guarantee the uniqueness of the α-projection.

Theorem 9. 1) The α-projection to M' is unique, when M' is a $-\alpha$-convex submanifold. 2) The α-projection from $M-\bar{V}$ to the boundary ∂V of V is unique, when V is a closed $-\alpha$-convex set having a smooth boundary.

Proof. It should be noted that a $-\alpha$-convex submanifold is $-\alpha$-autoparallel. We prove only case 2), because the proof of case 1) is almost similar. Let $M' = \partial V$. We assume the contrary that there exists a point ξ of which both θ_1 and θ_2 $(\theta_1 \neq \theta_2)$ are the α-projections to M'. We connect ξ and θ_1 by the α-geodesics c_1, and connect θ_1 and θ_2 by the $-\alpha$-geodesic c_0, constructing a triangle $(\xi, \theta_1, \theta_2)$. Since c_1 is orthogonal to $M' = \partial V$ and lies outside V while c is included in V, the angles τ_1 between c_1 and c_0 at θ_1 satisfies $\cos\tau_1 \leq 0$. Hence, Theorem 7 yields

$$D_\alpha(\xi, \theta_2) \geq D_\alpha(\xi, \theta_1) + D_\alpha(\theta_1, \theta_2) ,$$

$$D_\alpha(\xi, \theta_1) \geq D_\alpha(\xi, \theta_2) + D_\alpha(\theta_2, \theta_1) .$$

It then follows $D_\alpha(\theta_1, \theta_2) + D_\alpha(\theta_2, \theta_1) \leq 0$, which is a contradiction, proving the theorem.

We next study the problem of the α-projection in a manifold P of probability distributions whose associated manifold M(P) is α-flat. The α-divergence can be introduced in P by restricting M(P) to P. The α-projection in P also gives the α-approximation, even though P itself is not α-flat.

Theorem 10. Let P' be a smooth submanifold of a manifold P of probability distributions whose extension M(P) is α-flat. Then, the α-projection of θ to P' is θ' when, and only when $\theta' \in P'$ is the α-extreme point.

Proof. Let $\tilde{O}_\alpha(\theta', P')$ and $O_\alpha(\theta', P')$ be the orthogonal α-submanifolds at θ' of P' in M(P) and in P, respectively. Since $O_\alpha(\theta', P') = \tilde{O}_\alpha(\theta', P') \cap P$ holds (Corollary of Theorem 2), the α-projection of $\theta \in P$ to P' is θ', when and only when $\theta \in \tilde{O}_\alpha(\theta', P)$, and this holds when and only when θ' is the α-extreme point of θ.

Theorem 11. When M(P) is α-flat, 1) the α-projection from P to a smooth -α-convex submanifold P' is unique, and 2) the α-projection from P-V to the boundary $P' = \partial V$ of a closed -α-convex set V is unique.

Proof. We prove only case 2). Let $M(V)$ be the set of measures extended from the probability distributions of V. Then, $M(V)$ is also $-\alpha$-convex. Since $\tilde{O}_\alpha(\theta', P')$ is α-autoparallel and since it includes the one-dimensional submanifold $M(\theta')$ extended from $\theta' \in P'$, the α-projection of $\theta \in \tilde{O}_\alpha(\theta', P')$ to $M(V)$ is included in $M(\theta')$. Let θ'_1, θ'_2 be two distinct points in P'. Then, $\tilde{O}_\alpha(\theta'_1, P')$ and $\tilde{O}_\alpha(\theta'_2, P')$ are disjoint in $M(P) - M(V)$, because, if they have a common point ξ, its α-projection to $M(V)$ is on $M(\theta'_1)$ and on $M(\theta'_2)$ at the same time, contradicting the uniqueness of the α-projection in M. Hence, $O_\alpha(\theta'_1, P')$ and $O_\alpha(\theta'_2, P')$ are also disjoint in P, proving the theorem.

Let us define the α-sphere $S_\alpha(\theta_0, r)$ with center θ_0 and radius r by

$$S_\alpha(\theta_0, r) = \{\theta \mid D_\alpha(\theta_0, \theta) = r^2\} .$$

Theorem 12. The α-geodesic connecting θ_0 and a point θ $S_\alpha(\theta_0, r)$ in the sphere is orthogonal to the sphere in an α-flat M or in a P whose $M(P)$ is α-flat.

4. α-Families of Probability Distributions

4.1. Completely α-flat manifold

A manifold M is said to be completely α-flat or shortly c. α-flat, when, for any vector fields A, $B \in \mathcal{T}(M)$, $AB\phi^\alpha$ belongs to T^α_θ. In this case, the α-representation of the covariant derivative is written as

(4.1) $$(\nabla^{\alpha}_{A}B)\phi^{\alpha} = AB\phi^{\alpha}$$

because of (2.11), implying that the function $\partial_i\partial_j\phi^{\alpha}(x, \theta)$ is included in T^{α}_{θ}.

Theorem 13. A completely α-flat manifold is α-flat.

Proof. From (4.1), we have

$$[\nabla_A, \nabla_B]C\phi = (ABC - BAC)\phi = \nabla_{[AB]}C\phi ,$$

so that $R(A,B,C)\phi = 0$, proving the theorem.

Let θ be the α-affine coordinate system of c. α-flat M. Then, $\nabla^{\alpha}_{\partial i}\partial_j = 0$ or $\partial_i\partial_j\phi^{\alpha}(x, \theta) = 0$ holds, so that we have

(4.3) $$\phi^{\alpha}(x, \theta) = \theta^i\phi_i(x) + \phi_0(x) .$$

Hence, T^{α}_{θ} is spanned by n functions $\phi_i(x)$. The metric tensor $g_{ij}(\theta)$ is given by

(4.4) $$g_{ij}(\theta) = E_{\alpha}[\phi_i(x)\phi_j(x)] .$$

Since the density function $m(x, \theta)$ is connected with $\phi^{\alpha}(x, \theta)$ by (2.7), we have

$$m(x, \theta) = \begin{cases} [\frac{1-\alpha}{2}\{\theta^i \phi_i(x) + \phi_0(x)\}]^{(1-\alpha)/2}, & \alpha \neq 1, \\ \exp\{\theta^i \phi_i(x) + \phi_0(x)\}, & \alpha = 1. \end{cases}$$

If we further assume that, for real $t > 0$, $tm(x,\theta)$ is also included in M, we can choose an α-affine θ for which $\phi_0(x) = 0$ ($\alpha \neq 1$). In the case of $\alpha = 1$, by changing the dominating measure from $d\mu$ to $e^{-\phi(x)}d\mu$, we can eliminate the term $\phi_0(x)$: Hence, we have

$$(4.5) \qquad m(x, \theta) = \begin{cases} \{\frac{1-\alpha}{2}\theta^i \phi_i(x)\}^{(1-\alpha)/2}, & \alpha \neq 1 \\ \exp\{\theta^i \phi_i(x)\}, & \alpha = 1 \end{cases}$$

In order to obtain the potential functions F_α and G_α explicitly, let us define two functions of θ. One is the total measure $T(\theta)$, and the other is the entropy-like function $H(\theta)$,

$$(4.6) \qquad T(\theta) = \int m(x, \theta)d\mu ,$$

$$(4.7) \qquad H(\theta) = - \int m(x, \theta)\log m(x, \theta)d\mu .$$

Theorem 14. The potential functions of c. α-flat manifold M are given by

$$F_\alpha(\theta) = \begin{cases} \frac{2}{1+\alpha} T, & \alpha \neq -1, \\ -H - T, & \alpha = -1 \end{cases} \tag{4.8}$$

$$G_\alpha(\eta) = F_{-\alpha}(\theta) . \tag{4.9}$$

Proof. By differentiating (4.8) twice with respect to θ, it can be shown that $\partial_i \partial_j F_\alpha = g_{ij}(\theta)$ holds for all α. The potential $G_\alpha(\eta)$ can be calculated from (3.5), giving (4.9). It should be noted that c. α-flat manifold M is $-\alpha$-flat but is not necessarily c. $-\alpha$-flat.

In order to obtain the α-divergence in a compact form, let us define a one-parameter family of convex functions,

$$f_\alpha(u) = \begin{cases} \frac{4}{1-\alpha^2} \left[\frac{1-\alpha}{2} + \frac{1+\alpha}{2} u - u^{(1+\alpha)/2}\right], & \alpha \neq \pm 1, \\ u\log u - u + 1, & \alpha = 1, \\ -\log u + u - 1, & \alpha = -1. \end{cases} \tag{4.10}$$

Theorem 15. The α-divergence $D_\alpha(\theta_1, \theta_2)$ is written as

$$D_\alpha(\theta_1, \theta_2) = E_1\left[f_\alpha\left(\frac{m(x, \theta_2)}{m(x, \theta_1)}\right)\right] \tag{4.11}$$

in c. α-flat manifold M, where E_1 implies the integration with respect to the measure $m(x, \theta_1)$.

4.2. α-family

A manifold P of probability distributions is called an α-family, when its extended manifold M(P) of measures is completely α-flat.

Theorem 16. A manifold P is α-family, when and only when

$$\partial_i \partial_j \phi^\alpha(x, \xi) = A_{ij}{}^k(\xi) \partial_k \phi^\alpha(x, \xi) - \frac{1+\alpha}{2} g_{ij} p(x, \xi)^{(1-\alpha)/2} \tag{4.12}$$

holds for some functions $A_{ij}{}^k(\xi)$ in any coordinates $\xi = (\xi^i)$.

Proof. When $P = \{p(x, \xi)\}$ is an α-family, the extended $M(P) = \{tp(x, \xi)\}$ is c. α-flat. In the coordinates $\theta = (t, \xi)$ in M(P), we have

$$\partial_i \partial_j \phi^\alpha(x, \xi) = \Gamma^{\alpha}_{ij}{}^{k}(\xi) \partial_k \phi^\alpha + \Gamma^{\alpha}_{ij}{}^{0} \partial_0 \phi^\alpha$$

evaluated on the submanifold P in M(P), i.e., t = 1, where $\theta^0 = t$ and i, j, k are indices for the coordinates ξ in P. Since $\partial_i \partial_j \phi^\alpha$ is a linear combination of $\partial_k \phi^\alpha$ and $\partial_0 \phi^\alpha$ from (2.18) we have

$$\Gamma^{\alpha}_{ij}{}^{0} = -\frac{1+\alpha}{2} g_{ij}$$

so that (4.12) holds because of $\partial_0 \phi^\alpha(x, t, \xi) = p(x, \xi)^{(1-\alpha)/2}$. On the contrary, when (4.12) holds, $\partial_i \partial_j \phi^\alpha$ belongs to $\tilde{T}^\alpha_\theta$ spanned by $\partial_k \phi^\alpha$ and $\partial_0 \phi^\alpha$. It is easy to show that $\partial_0 \partial_i \phi^\alpha$ and $\partial_0 \partial_0 \phi^\alpha$ also belong to $\tilde{T}^\alpha_\theta$. Hence M(P) is c. α-flat and P is an α-family.

By choosing an (n+1)-dimensional α-affine coordinates θ in M(P) for an n-dimensional α-family P, it can be represented in the form

$$(4.13)\qquad p(x,\theta) = \begin{cases} \{\frac{1-\alpha}{2}\theta^i\phi_i(x)\}^{2/(1-\alpha)}, & \alpha \neq 1 \\ \exp\{\theta^i\phi_i(x)\}, & \alpha = 1 \end{cases}$$

where n among n+1 components θ^i of θ are independent because of the constraint

$$(4.14)\qquad T(\theta) = \int p(x,\theta)d\mu = 1 .$$

It should be remarked that the coordinates (t,ξ) is not in general α-affine. The n+1 affine coordinates θ can be regarded as homogeneous coordinates of P ($\alpha \neq 1$), since T(θ) is a homogeneous function of degree $2/(1-\alpha)$. The following theorem is clear from Theorem 2.

Theorem 17. A submanifold P' of an α-family P is α-family, when and only when P' is α-autoparallel.

The α-families include some famous families well known in statistics. For example, when α = 1, we can choose first n components $\bar{\theta}^i = \theta^i$ $(i = 1, \ldots, n)$ of θ as a coordinate system in P. By taking an adequate dominating measure dμ, the 1-family can be written as

$$p(x, \bar{\theta}) = \exp \{\bar{\theta}^i \phi_i(x) - \psi(\bar{\theta})\}$$

where $\exp[\psi(\bar{\theta})] = \int \exp[\bar{\theta}^i \phi_i(x)] d\mu$. This family is well known as the exponential family and plays a fundamental role in statistics. The parameter $\bar{\theta}$ is known as the natural parameter. We have already seen in Example 1 that an exponential family is 1-flat (and is hence -1-flat) with the affine coordinate system $\bar{\theta}$. Since an α-family is not in general α-flat for $\alpha \neq \pm 1$, this is a special property for $\alpha = 1$.

When $\alpha = -1$, we have $p(x, \theta) = \theta^i \phi_i(x)$. We can choose the scale of θ^i such that $\int \phi_i(x) d\mu = 1$ always holds. Then, the constraint (4.14) becomes $\Sigma\theta^i = 1$. This family in well known in statistics as the mixture family. We can use the first n components of θ as the coordinates $\bar{\theta}$ of P, with

$$p(x, \bar{\theta}) = \bar{\theta}_i \bar{\phi}_i(x) + \phi_{n+1}(x) \ ,$$

$$\bar{\phi}_i(x) = \phi_i(x) - \phi_{n+1}(x) \ .$$

This shows that a -1-family is c. -1-flat, and hence is 1-flat (but not c. -1-flat). This is a special property for $\alpha = -1$. The special property for $\alpha = \pm 1$ is given rise to by the fact that P is -1-autoparallel in M (Corollary 1).

The α-geodesic of an α-family can easily be obtained in terms of the homogeneous coordinates.

Theorem 18. The α-geodesic connecting two points θ_1 and θ_2 is given by

(4.15) $$\theta(t) = c(t)\{\theta_1 + t(\theta_2 - \theta_1)\}\ , \quad \alpha \neq 1\ ,$$

(4.16) $$\bar{\theta}(t) = \bar{\theta}_1 + t(\bar{\theta}_2 - \bar{\theta}_1)\ , \quad \alpha = 1\ .$$

where $c(t)$ is determined from the normalization condition $T[\theta(t)] = 1$.

Proof. The α-geodesic connecting θ_1 and θ_2 is given by $\theta(t) = \theta_1 + t(\theta_2 - \theta_1)$ in M(P). Since the two-dimensional submanifold given by $\theta(t_1,t_2) = t_1\theta_1 + t_2\theta_2$ is α-autoparallel in M(P), the α-geodesic in P is given by the intersection $\theta(t_1, t_2) \cap P$. Hence, we get (4.15) or (4.16).

Example 3. The functions $m(x, \xi)$ of the manifold M_n of the measures on a finite number of atoms are given by $m(x,\xi) = \xi^i\delta_i(x)$ (see Example 2), where ξ^i is the measure of atom x_i, ξ forming a coordinate system of M_n. By defining the new coordinate system θ_α depending on α,

$$\theta_\alpha^i = \begin{cases} \frac{2}{1-\alpha}\,(\xi^i)^{(1-\alpha)/2}\ , & \alpha \neq 1 \\ \log\xi^i\ , & \alpha = 1 \end{cases}$$

we have $\phi^\alpha(x,\theta_\alpha) = \theta_\alpha^i\delta_i(x)$ so that $\partial_i\partial_j\phi^\alpha(x, \theta_\alpha) = 0$. Hence, M_n is c. α-flat for any α and the θ_α is the α-affine coordinates.

The probability distribution P_{n-1} on n atoms $x_1, \ldots, x_n$ is an (n-1)-dimensional α-family for any α, because its extended manifold is

M_n. Hence, we can define the α-divergence, the α-projection, etc. in P_{n-1}. It can also be shown in M_n that $\theta_{-\alpha}$ plays the role of the dual of θ_α. This defines the dual (homogeneous) coordinates η_α in P_{n-1}.

4.3. Duality in α-families

The dual homogeneous coordinates η are obtained in an α-family ($\alpha \neq \pm 1$) by

$$\eta_j = \frac{2}{1+\alpha} \partial_j T , \qquad \theta^i = \frac{2}{1-\alpha} \partial^i T .$$

Obviously, η should satisfy the constraint $T(\eta) = 1$. It is interesting that the normalization constraint given by T plays the central role in both the Legendre and its inverse transformations. The normal vectors of the constraint surface T = const. give the dual coordinates in the both transformations.

For the exponential family P where $\alpha = 1$, the dual coordinates $\bar{\eta}$ of $\bar{\theta}$ is given by $\bar{\eta}_i = \partial_i \psi(\bar{\theta})$ since P itself is 1-flat. The inverse transformation is given by $\bar{\theta}^i = \partial^i \phi(\bar{\eta})$, where $\phi(\bar{\eta}) = -H(\bar{\eta}) - 1$, (ψ, ϕ) being the potential functions. The dual coordinate system $\bar{\eta}$ is known as the expectation parameter.

For the mixture family where $\alpha = -1$, the dual coordinates $\bar{\eta}$ are given by

$$\bar{\eta}_i = \int \bar{\phi}_i(x)\{1 + \log p(x, \bar{\theta})\} d\mu .$$

The dual coordinates $(\bar{\theta}, \bar{\eta})$ of the mixture family P have the potentials $\psi(\bar{\theta})$ and $\phi(\bar{\eta})$,

$$\psi(\bar{\theta}) = -H(\bar{\theta}) , \quad \phi(\bar{\theta}) = -\int \phi_{n+1} \log p \, d\mu$$

and are related by $\bar{\eta}_i = \partial_i \psi(\bar{\theta})$, $\bar{\theta}^i = \partial^i \phi(\bar{\eta})$.

4.4. α-divergence and α-entropy in α-family

Because of the constraint (4.14), the α-divergence has a simpler and more familiar form in the α-family. For the functions

$$(4.17) \qquad h_\alpha(u) = \begin{cases} \frac{4}{1-\alpha^2}\{1 - u^{(1+\alpha)/2}\}, & \alpha \neq \pm 1 \\ u \log u , & \alpha = 1 \\ -\log u , & \alpha = -1 \end{cases}$$

the α-divergence is expressed as

$$(4.18) \qquad D_\alpha(\theta_1, \theta_2) = E_1[h_\alpha\{\frac{p(x, \theta_2)}{p(x, \theta_1)}\}]$$

in the α-family.

The -1-divergence is well known as the Kullback-Leibler divergence, and its geometry has extensively been studied so far (Csiszár [8], Čencov [6], Efron [12], Ingarden [13], Amari [3]). The 1-divergence is the dual of the -1-divergence, i.e., $D_{-1}(\theta_1, \theta_2) = D_1(\theta_2, \theta_1)$. The α-divergence

($\alpha \neq \pm 1$) coincides with the so-called Chernoff distance (Chernoff [5]). Especially, the 0-divergence is well known as the Hellinger distance. The 0-divergence is symmetric and it satisfies the triangular inequality. As will be seen later, the Hellinger distance is related to the Riemannian distance.

It should be noted that the α-divergence is closely related with Rényi's α-entropy or α-information (Rényi [21], [22]). Since the α-information is defined for $0 < \alpha \leq 1$, we put $\alpha' = (1 - \alpha)/2$. The α'-information $I_{\alpha'}(p,q)$ of p relative to distribution q is defined by

$$(4.19) \qquad I_{\alpha'}(p, q) = \frac{1}{\alpha'-1} \log\{\int p(x)^{\alpha'} q(x)^{1-\alpha'} d\mu\} \ .$$

When $\alpha' = 1$, the limit $\alpha' \to 1$ is used. When $\alpha' = 1$ and $qd\mu$ is the uniform measure, it reduces to the ordinary Shannon entropy. The α'-information and the α-divergence is connected by

$$(4.20) \qquad I_{\alpha'}(p, q) = \frac{2}{1+\alpha}\{\frac{1-\alpha^2}{4} D_\alpha(p, q) - 1\} \ ,$$

where the limit is taken for $\alpha = -1$ ($\alpha' = 1$).

5. Remarks on the Geometry of Function Spaces of Distributions

We have so far treated only parametrized families of distributions, because there are some difficulties in constructing a differential-geometrical theory of infinite-dimensional function spaces of distributions, (cf. Čencov, [6]). Here, we suggest the structure of the geometry of the space of density function, without rigorous arguments. The function space is expected to be completely α-flat for any α in analogy with the case of Example 3 where the dominating measure is concentrated on a finite number of atoms.

Let $M = \{m(x)\}$ be the set of density functions on the real Euclidean space X with respect to the Lebesgue measure dx. It is assumed that $m(x) > 0$ and that $m(x)$ has the moments of any order. M is a subset of $L^1(X)$. Let P be its subset, defined by

$$T(m) = \int m(x)dx = 1 . \tag{5.1}$$

The α-representation of $m(x)$ is

$$\phi^\alpha(x) = \begin{cases} \frac{2}{1-\alpha} m(x)^{(1-\alpha)/2} , & \alpha \neq 1 \\ \log m(x) , & \alpha = 1 \end{cases} \tag{5.2}$$

The set $M_\alpha = \{\phi^\alpha(x)\}$ is obviously a subset of $L^\beta(X)$, $\beta = 2/(1 - \alpha)$, with the convention that $L^\infty(X)$ $(\alpha = 1)$ implies the set of functions $\ell(x)$ for which $\exp\{\ell(x)\}$ is integrable. The following lemma is obtained immediately from the Minkowski and Hölder inequalities.

Lemma 4. For $\alpha \neq 1$, M_α is a cone in the sense that, for $\phi_1, \phi_2 \in M_\alpha$ and $c, d \geqq 0$, $c\phi_1 + d\phi_2 \in M_\alpha$. For $\alpha = 1$, M_1 is a convex set in the sense that $c\phi_1 + d\phi_2 \in M_1$ for $\phi_1, \phi_2 \in M_1$, $c + d = 1$, $c, d \geqq 0$.

Let $m(x, t)$, $0 \leq t$, be a smooth curve in M in a suitable sense (cf. Čencov [6]). We assume that its derivatives with respect to t belong to $L^1(X)$. We call the one-sided first derivative $A(x) = \dot{m}(x, 0)$ at $t = 0$, where $\dot{}$ denotes d/dt, the tangent of the curve at $m(x, 0)$. The α-representation of the tangent is given by

$$A_\alpha(x) = \dot{\phi}^\alpha(x, 0) = m^{-(1+\alpha)/2}\,\dot{m}\,, \tag{5.3}$$

where $\phi^\alpha(x, t)$ is the α-representation of $m(x, t)$. When $m_1(x, t)$ and $m_2(x, t)$ are two curves starting at $m(x)$ with $t = 0$,

$$m(x, t) = \frac{1}{2}m_1(x, 2ct) + \frac{1}{2}m_2(x, 2dt)$$

is another curve starting at $m(x)$ for $c, d \geq 0$. Its tangent is given by $\dot{m}(x, 0) = c\dot{m}_1(x, 0) + d\dot{m}_2(x, 0)$. Hence, a non-negative linear combination of two tangents is also a tangent. Hence, the set of all the tangents at $m(x)$ forms a cone. The set $T_\alpha(m)$ of the α-representations of the tangents at m forms a cone. Unfortunately, the set $T_\alpha(m)$ is not necessarily a linear space, because $m(x, t)$ can not necessarily be extended for negative t. Hence, the space M of distributions does not form a manifold. The inner product of two tangents $A(x)$ and $B(x)$ is defined by

$$\langle A, B \rangle = E_\alpha[A_\alpha(x)B_\alpha(x)] = \int A_\alpha(x)B_{-\alpha}(x)dx \tag{5.4}$$

where A_α is the α-representation of A.

The tangent set can similarly be defined for the set P of all the probability distributions imbedded in M. The tangent of a curve $p(x, t)$, $t \geq 0$, in P is given by $A(x) = \dot{p}(x, 0)$. For the curve $m(x, t) = (1 + t)p(x)$ in M starting at $p(x) \in P$, the tangent $\partial_0(x)$ is given by $\partial_0 = p(x)$ or by its α representation $\phi^\alpha(x)$. (When $\alpha = 1$, $\partial_0(x) = 1$.) A tangent $A(x) = \dot{p}(x, 0)$ of a curve $p(x, t)$ in P always satisfies

$$E_\alpha[\partial_0(x) A_\alpha(x)] = 0$$

because of $E_\alpha[p^{(1-\alpha)/2} \phi^\alpha] = 0$. Hence, $\partial_0(x)$ gives a normal field orthogonal to P.

The α-geodesic connecting two points $m_1(x), m_2(x) \in M$ is defined by the curve

$$\phi^\alpha(x, t) = \phi_1^\alpha(x) + t\{\phi_2^\alpha(x) - \phi_1^\alpha(x)\} \tag{5.5}$$

in M_α, where ϕ_i^α is the α-representation of m_i, $i = 1, 2$. This suggests that M is completely α-flat. The α-geodesic connecting $p_1(x)$ and $p_2(x)$ in P is given by

(5.6)

$$\phi^\alpha(x, t) = c(t)[\phi_1^\alpha(x) + t\{\phi_2^\alpha(x) - \phi_1^\alpha(x)\}] , \quad \alpha \neq 1$$

$$\ell(x, t) = c(t) + \ell_1(x) + t\{\ell_2(x) - \ell_1(x)\} , \quad \alpha = 1$$

where $c(t)$ is determined from the normalization condition $T[m(x, t)] = 1$.

Let

$$(5.7) \qquad F_\alpha(\phi^\alpha) = \begin{cases} \frac{2}{1+\alpha} T, & \alpha \neq 1 \\ - H - T, & \alpha = 1 \end{cases}$$

be one-parameter functionals on M, where H is the entropy-like function as in (4.7). The Fréchet derivative of F_α is a linear mapping $F'_\alpha : T_\alpha \to R$, and the second derivative F''_α is a bilinear symmetric mapping $F''_\alpha : T_\alpha \times T_\alpha \to R$. It is easy to show that, for $A_\alpha, B_\alpha \in T_\alpha$,

$$(5.8) \qquad F''_\alpha(A_\alpha, B_\alpha) = E_\alpha[A_\alpha B_\alpha] = \langle A, B \rangle .$$

Hence, F''_α is equal to the α-expectation operator E_α which defines the inner product in T_α. From Theorem 4, we see M_α is an α-affine representation of M, which is completely α-flat for all α.

A linear mapping $K : L^\beta(X) \to R$ can be represented by a function $f_K \in L^{\beta'}(X)$, $(1/\beta) + (1/\beta') = 1$, such that, for $A \in L^\beta(X)$,

$$KA = \int A(x) f_k(x) dx .$$

Since T_α is a subset of $L^\beta(X)$, $\beta = 2/(1 - \alpha)$, the linear operator F'_α is represented by a function in $L^{\beta'}(X)$, $\beta' = 2/(1 + \alpha)$. We have by calculations

$$(5.9) \qquad F'_\alpha(\phi^\alpha)A = \int \phi^{-\alpha}(x) A(x) dx$$

for $A \in T_\alpha$. This implies that the $-\alpha$-representation $\phi^{-\alpha}$ is the dual of ϕ^α

(5.10) $$\phi^{-\alpha}(x) = F_\alpha'(\phi^\alpha) \ .$$

This is the Legendre transformation, and the inverse transformation is given by

(5.11) $$\phi^\alpha(x) = G_\alpha'(\phi^{-\alpha})$$

with the functional G_α satisfying $G_\alpha(\cdot) = F_{-\alpha}(\cdot)$ and

$$F_\alpha(\phi^\alpha) + G_\alpha(\phi^{-\alpha}) - \int \phi^\alpha \phi^{-\alpha} dx = 0 \ .$$

The α-divergence from m_1 to m_2 is then defined by

$$D_\alpha(\phi_1^\alpha, \phi_2^\alpha) = F_\alpha(\phi_1^\alpha) + G_\alpha(\phi_2^\alpha) - \int \phi_1^\alpha \phi_2{}^{\alpha} dx \ ,$$

which can be written in the form

$$D\alpha(m_1, m_2) = \int f_\alpha\left(\frac{m_2(x)}{m_1(x)}\right) m_1(x) dx \ .$$

The important Theorem 7 is valid in this case, too.

The α-projection can also be defined in M and in P. By defining a region V having a smooth boundaries, the relation

$$A'D_\alpha(m, m') = -\left\langle A', \phi^\alpha - \phi'^\alpha \right\rangle$$

holds for a tangent A' of the boundary, where A' is a Fréchet derivative operator operating to the second variable in D_α. Therefore, Theorems 9∿12 hold in this case, too. Hence, the inverse image of the α-approximation is given by the orthogonal α-totally geodesic set. It should be remarked that the sets M and P are α-convex for any α. A statistical model S is a parametrized family $S = \{p(x, \theta)\}$ of probability distributions. We often treat a smooth model imbedded in P. The α-approximation of a distribution $p(x)$ by a member in S is an important problem in applications, and is solved in this way (see Amari, [3]).

6. Conclusion

We have treated the geometrical structures of a smooth family of probability distributions by introducing a Riemannian metric and α-connections. The α - $(-\alpha)$ dualistic structures are elucidated. It has been shown that the α-divergence is naturally introduced in an α-flat family. It is closely related to the α-geodesics, and they together constitute the α-geometry. The α-family of probability distributions are also defined. These geometrical structures play important roles in studying the properties of various statistical models. We have also discussed the geometrical properties of the function space of distributions. However, there remain many problems to be studied further, because we need to modify or generalize the concept of manifold to treat the geometry of function spaces.

References

[1] Amari, S. (1980). Theory of information spaces: A differential -geometrical foundation of statistics. POST RAAG Rep., No.106.

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

[3] Amari, S. (1982b). Differential geometry of curved exponential families --- curvatures and information loss. Ann. Statist., 10, 357-387.

[4] Amari, S. and Kumon, M. (1983). Differential geometry of Edgeworth expansions in curved exponential family. Ann. Statist. Math., to appear.

[5] Chernoff, H. (1952). A measure of asymptotic efficiency for test of a hypothesis bases on a sum of observations. Ann. Math. Statist., 23, 493-507.

[6] Čencov, N.N. (1972). Statistical Decision Rules and Optimal Conclusions. Nauka, Moscow (in Russian), translated into English (1982), AMS, Rhode Island.

[7] Csiszár, I. (1967). Information-type measures of difference of probability distributions and indirect observations. Studia Sci Math. Hungar., 2, 299-318.

[8] Csiszár, I. (1975). I-divergence geometry of probability distributions and minimization problems. Ann. Prob., 3, 146-158.

[9] Dawid, A.P. (1977). Discussions to Efron's paper. Ann. Statist., 3, 1231-1234.

[10] Dawid, A.P. (1977). Further comments on a paper by Bradley Efron. Ann. Statist., 5, 1249.

[11] Efron, B. (1975). Defining the curvature of a statistical problem (with application to second order efficiency) (with Discussion). Ann. Statist., 3, 1189-1242.

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

[13] Ingarden, R.S. (1981). Information geometry in function spaces of classical and quantum finite statistical systems. Internat. J. Engrg. Sci., 19, 1609-1633.

[14] Kobayashi, S. and Nomizu, K. (1963). Foundations of Differential Geometry, I. John Wiley, New Yourk & London.

[15] Kobayashi, S. and Nomizu, K. (1969). Foundations of Differential Geometry, II. John Wiley, New York & London.

[16] Kullback, S.L. (1959). Information Theory and Statistics. John Wiley, New York.

[17] Kumon, M. and Amari, S. (1983). Geometrical theory of higher-order asymptotics of test, interval estimator and conditional inference. Biometrika, To appear.

[18] Lang, S. (1972). Differential Manifolds. Addison-Wesley, Reading, Massachusetts.

[19] Madsen, L.T. (1979). The geometry of statistical model --- a generalization of curvature. Research Report., 79-1, Statist. Res. Unit., Danish Medical Res. Council.

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

[21] Rényi, A. (1961). On measures of entropy and information. Proc. 4th Berkeley Symp. Math. Statist. Prob., University of California Press, 347-561.

[22] Rényi, A. (1970). Probability Theory. North-Holland, Amsterdam.

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

Dept. of Mechanical Eng.
Osaka University, Toyonaka,
560 Japan

Dept. of Mathematical Engineering
and Instrumentation Physics,
University of Tokyo
Tokyo, 113 Japan