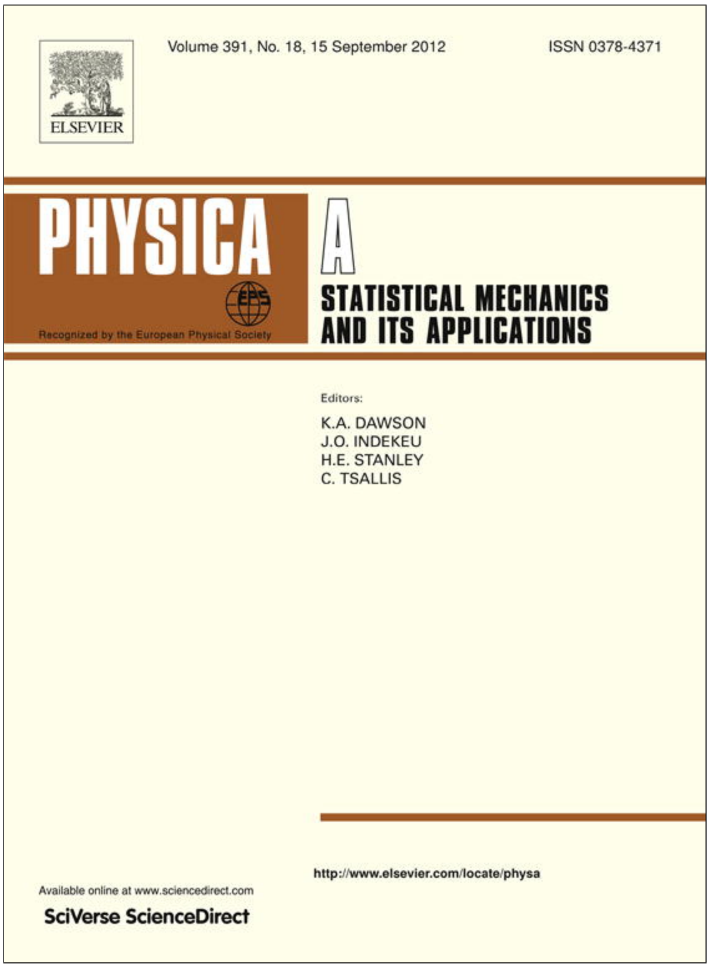
Volume 391, No. 18, 15 September 2012
ISSN 0378-4371
ELSEVIER
PHYSICA A
STATISTICAL MECHANICS AND ITS APPLICATIONS
Recognized by the European Physical Society
Editors:
K.A. DAWSON
J.O. INDEKEU
H.E. STANLEY
C. TSALLIS
http://www.elsevier.com/locate/physa
Available online at www.sciencedirect.com
SciVerse ScienceDirect

Physica A 391 (2012) 4308–4319

Contents lists available at SciVerse ScienceDirect

Physica A

journal homepage: www.elsevier.com/locate/physa

Geometry of deformed exponential families: Invariant, dually-flat and conformal geometries

Shun-ichi Amari [a], Atsumi Ohara [b,*], Hiroshi Matsuzoe [c]

[a] Riken Brain Science Institute, Wako-shi Hirosawa 2-1, Saitama 351-0198, Japan

[b] Department of Electrical and Electronics Engineering, University of Fukui, Bunkyo 3-9-1, Fukui 910-8507, Japan

[c] Department of Computer Science and Engineering, Graduate School of Engineering, Nagoya Institute of Technology, Gokiso-cho, Showa-ku, Nagoya 466-8555, Japan

ARTICLE INFO

Article history:
Received 18 January 2012
Received in revised form 10 April 2012
Available online 25 April 2012

Keywords:
Generalized entropies
Deformed exponential families
Information geometry
Invariance principle
Conformal transformation

ABSTRACT

An information-geometrical foundation is established for the deformed exponential families of probability distributions. Two different types of geometrical structures, an invariant geometry and a flat geometry, are given to a manifold of a deformed exponential family. The two different geometries provide respective quantities such as deformed free energies, entropies and divergences. The class belonging to both the invariant and flat geometries at the same time consists of exponential and mixture families. The q-families are characterized from the viewpoint of the invariant and flat geometries. The q-exponential family is a unique class that has the invariant and flat geometries in the extended class of positive measures. Furthermore, it is the only class of which the Riemannian metric is conformally connected with the invariant Fisher metric.

1. Introduction

Since the introduction of q-entropy by Tsallis [1] (see also an extensive monograph [2]), much attention has been paid to non-extensive statistical mechanics. It is related to various 'non-standard phenomena' subject to the power law not only in statistical physics but in economics and disaster statistics. Here, families of probability distributions of the q-exponential family and more general deformed exponential families play a major role. In the present paper, a geometrical foundation is given to these families of distributions from the point of view of information geometry [3].

The deformed exponential family was introduced and studied extensively by Naudts [4,5] (see also a monograph [6]). Kaniadakis et al. [7] studied the κ-exponential family which belongs to the deformed exponential family. Its mathematical structure was studied by Pistone [8] and Vigelis and Cavalcante [9]. See other examples with interesting discussions [10,11]. In statistics, a similar notion of a generalized exponential family [12] or the U-model [13,14] is discussed on the bases of respective motives.

Many useful concepts such as generalized entropy, divergence and escort probability distribution have been proposed. However, their relationships have not necessarily been well understood theoretically and are waiting for further geometrical and statistical elucidation. It is also useful to characterize the q-families in the class of general deformed exponential families.

In the present study, information geometry [3] is used to give a foundation to the deformed exponential families. Two types of geometry can be introduced in the manifold of a deformed exponential family: One is the *invariant geometry*, where the Fisher information is the unique Riemannian metric (Chentsov [15]; also see Ref. [3]) together with a dual pair of invariant affine connections (α-connections). The other is the *dually flat geometry* [3] (also see Ref. [16]), which is not necessarily

* Corresponding author.
E-mail addresses: amari@brain.riken.jp (S.-i. Amari), ohara@fuee.u-fukui.ac.jp (A. Ohara), matsuzoe@nitech.ac.jp (H. Matsuzoe).

0378-4371/$ – see front matter © 2012 Elsevier B.V. All rights reserved.
doi:10.1016/j.physa.2012.04.016

invariant but accompanies the Legendre structure. The escort probability distribution belongs to the latter geometry. The two geometries give different free-energies, entropies and divergences in general.

The exponential and mixture families are characterized by the property that they sit at the intersection of the classes of the invariant and flat geometries. The q-exponential family is then characterized from two viewpoints: One is invariance and flatness in the class of positive measures and the other is conformal geometry [17,18]. It is shown that the q-family is a unique class of flat geometry that is connected conformally to the invariant geometry.

2. Deformed exponential family

We follow Naudts [6,19] for the formulation of the deformed exponential family. Given a positive increasing function $\chi(s)$ on $(0, \infty) \in \boldsymbol{R}$, a deformed logarithm, called the χ-logarithm, is defined by

$$\ln_\chi(s) = \int_1^s \frac{1}{\chi(t)} \mathrm{d}t. \tag{1}$$

This is a concave monotonically increasing function. When χ is a power function,

$$\chi(s) = s^q, \quad q > 0, \tag{2}$$

(1) gives the q-logarithm

$$\ln_q(s) = \frac{1}{1-q}\left(s^{1-q} - 1\right). \tag{3}$$

The ordinary logarithm is obtained by taking the limit of $q = 1$.

The inverse of the χ-logarithm is the χ-exponential, given by

$$\exp_\chi(t) = 1 + \int_0^t \lambda(s)\mathrm{d}s, \tag{4}$$

where $\lambda(s)$ is defined by the relation

$$\lambda\left\{\ln_\chi(s)\right\} = \chi(s). \tag{5}$$

The q-exponential is given by

$$\exp_q(t) = \{1 + (1-q)t\}^{\frac{1}{1-q}}, \tag{6}$$

where the limit $q = 1$ gives the ordinary exponential.

A family $S = \{p(\boldsymbol{x}, \boldsymbol{\theta})\}$ parameterized by $\boldsymbol{\theta} = \left(\theta^1, \ldots, \theta^n\right)$ of probability distributions of a vector random variable $\boldsymbol{x} = (x_1, \ldots, x_n)$ is called a χ-exponential family, when its density function is given by

$$p(\boldsymbol{x}, \boldsymbol{\theta}) = \exp_\chi\left\{\sum_i \theta^i x_i - \psi(\boldsymbol{\theta})\right\} \tag{7}$$

with respect to a dominating measure $\mu(\boldsymbol{x})$. Here, $\psi(\boldsymbol{\theta})$ is determined from the normalization condition

$$\int p(\boldsymbol{x}, \boldsymbol{\theta}) d\mu(\boldsymbol{x}) = 1 \tag{8}$$

and is called the χ-free energy. Family S is regarded as an n-dimensional manifold, where $\boldsymbol{\theta}$ plays the role of a coordinate system. We call $\boldsymbol{\theta}$ the χ-coordinate system of the χ-exponential family. Any linear subspace in the $\boldsymbol{\theta}$-coordinates is also a χ-exponential family. The q-exponential family is χ-exponential family given by

$$p(\boldsymbol{x}, \boldsymbol{\theta}) = \exp_q\left\{\sum \theta^i x_i - \psi(\boldsymbol{\theta})\right\}. \tag{9}$$

Let us consider discrete random variable x, taking values on $X = \{0, 1, \ldots, n\}$. Let $S_n = \{p(x)\}$ be the family consisting of all such probability distributions. It is called the probability n-simplex. By introducing a new vector random variable $\boldsymbol{x} = \{\delta_i(x)\}$,

$$\delta_i(x) = \begin{cases} 1, & x = i \\ 0, & \text{otherwise}, \end{cases} \tag{10}$$

we have

$$p(x) = \sum_{i=0}^n p_i \delta_i(x), \tag{11}$$

where $p_i = \text{Prob}\{x = i\}$ with constraint $\sum p_i = 1$. Hence, S_n is an n-dimensional manifold, where $\boldsymbol{p} = (p_1, \ldots, p_n)$ plays the role of a coordinate system.

Theorem 1. *S_n is a χ-exponential family for any χ.*

Proof. Since $\delta_i(x)$ takes values 0 and 1, we easily have

$$\ln_\chi \left\{ \sum_{i=0}^{n} p_i \delta_i(x) \right\} = \sum_{i=1}^{n} \left\{ \ln_\chi p_i - \ln_\chi p_0 \right\} \delta_i(x) + \ln_\chi p_0. \tag{12}$$

By putting

$$\theta^i = \ln_\chi p_i - \ln_\chi p_0, \tag{13}$$

$$x_i = \delta_i(x), \tag{14}$$

$$\psi(\boldsymbol{\theta}) = -\ln_\chi \left\{ 1 - \sum_{i=1}^{n} p_i \right\} \tag{15}$$

we have

$$p(x) = \exp_\chi \left\{ \sum \theta^i x_i - \psi(\boldsymbol{\theta}) \right\}, \tag{16}$$

where ψ is a function of $\boldsymbol{\theta}$ determined from $\sum p(x) = 1$. The $\boldsymbol{\theta}$ determined by (13) is the χ-coordinate system of S_n. □

In the case of a continuous random variable x, a mathematically exact formulation is given by Pistone and Sempi [20], Cena and Pistone [21], Vigelis and Cavalcante [9] and Grasselli [22]. Here, we give intuitive but non-rigorous observation. We put the family of all the density functions, absolutely continuous with respect to the Lebesgue measure, as

$$F = \left\{ p(x) \;\middle|\; p(x) > 0,\ \int p(x)\mathrm{d}x = 1 \right\}. \tag{17}$$

Instead of p_i and $\delta_i(x)$ in the discrete case, by using the delta function $\delta(t)$, we have

$$p(x) = \int p(t)\delta(t-x)\mathrm{d}t \tag{18}$$

and

$$x(t) = \delta(t-x). \tag{19}$$

Then, by putting

$$\theta(t) = \ln_\chi p(t), \tag{20}$$

we have

$$\ln_\chi p(x) = \int \theta(t)\delta(t-x)\mathrm{d}t - \psi, \tag{21}$$

where the term ψ is added as a normalizing factor. It thus follows that

$$p(x) = \exp_\chi \left\{ \int \theta(t)\delta(t-x)\mathrm{d}t - \psi \right\} \tag{22}$$

is represented in the form of a χ-exponential family, where ψ is a functional of $\theta(t)$ satisfying

$$\int \exp_\chi \left\{ \theta(x) - \psi \right\} \mathrm{d}x = 1. \tag{23}$$

We have shown that both S_n and F are regarded as a χ-exponential family (χ-family for short) for any χ. When two functions $\chi(t)$ and $\tilde{\chi}(t)$ are linearly connected as

$$\chi(t) = c_1 \tilde{\chi}\,(c_2 t) \tag{24}$$

for constants c_1 and c_2, we say that χ and $\tilde{\chi}$ are equivalent. It is easy to see that, when χ and $\tilde{\chi}$ are equivalent, a χ-family is also a $\tilde{\chi}$-family. We have the following result which is a generalization of Theorem 4 of [5].

Theorem 2. *When χ and $\tilde{\chi}$ are not equivalent, χ-family is not a $\tilde{\chi}$-family, except for S_n and F which are χ-families for any χ.*

Proof. We prove the theorem in the discrete case. Any family M of discrete distributions is a submanifold of S_n. Let us assume that M belongs to χ- and $\tilde{\chi}$-family for non-equivalent χ and $\tilde{\chi}$. Let $\boldsymbol{\theta}_\chi$ and $\boldsymbol{\theta}_{\tilde{\chi}}$ be the χ- and $\tilde{\chi}$-coordinates of S_n, where

$$\theta_\chi^i = \exp_\chi \left\{ \ln_{\tilde{\chi}} \left(\theta_{\tilde{\chi}}^i \right) \right\}, \quad i = 1, \ldots, n \tag{25}$$

holds. M is given by a linear subspace of $\boldsymbol{\theta}_\chi$ and also of $\tilde{\boldsymbol{\theta}}_\chi$. However, any linear constraints on $\boldsymbol{\theta}_\chi$ are not linear in $\tilde{\boldsymbol{\theta}}_{\tilde{\chi}}$, provided χ and $\tilde{\chi}$ are not equivalent. □

3. Invariant geometry of χ-family

3.1. α-geometry

A χ-family (7) is considered as a manifold with a (local) coordinate system $\boldsymbol{\theta}$. We use the invariance principle to introduce a unique geometrical structure (see Refs. [3,15]).

Invariance principle: The geometry is invariant under the transformation of random variable $\boldsymbol{x}$ *to* $\boldsymbol{y}$, *provided* $\boldsymbol{y}$ *is a sufficient statistics.*

The invariant geometry is characterized by the two tensors,

$$g_{ij}(\boldsymbol{\theta}) = E\left[\partial_i \log p(\boldsymbol{x}, \boldsymbol{\theta}) \partial_j \log p(\boldsymbol{x}, \boldsymbol{\theta})\right], \tag{26}$$

$$T_{ijk}(\boldsymbol{\theta}) = E\left[\partial_i \log p(\boldsymbol{x}, \boldsymbol{\theta}) \partial_j \log p(\boldsymbol{x}, \boldsymbol{\theta}) \partial_k \log p(\boldsymbol{x}, \boldsymbol{\theta})\right] \tag{27}$$

where $\partial_i = \partial/\partial\theta^i$ and E denotes expectation. The first tensor, g_{ij}, is the Fisher information matrix, playing the role of a Riemannian metric. T_{ijk} is a third-order symmetric tensor that defines a pair of dual affine connections, $\Gamma^{(\alpha)}_{ijk}$ and $\Gamma^{(-\alpha)}_{ijk}$, for any real number α,

$$\Gamma^{(\alpha)}_{ijk} = [i, j; k] - \frac{\alpha}{2} T_{ijk}, \tag{28}$$

called the α-connection. Here, $[i, j; k]$ is the Christoffel symbol defining the Riemannian (Levi–Civita) connection,

$$\Gamma^{(0)}_{ijk} = [i, j; k] = \frac{1}{2}\left[\partial_i g_{jk} + \partial_j g_{ik} - \partial_k g_{ij}\right]. \tag{29}$$

The ($\alpha = 1$)-connection is called the e-connection (exponential connection) and $\alpha = -1$ is called the m-connection (mixture connection).

3.2. Invariant divergence

Let us consider a divergence function $D[p(x) : q(x)]$ between two probability distributions. Here, D should satisfy

$$D[p(x) : q(x)] \geq 0, \quad \text{the equality holds} \quad \text{iff} \quad p(x) = q(x). \tag{30}$$

A divergence is said to be decomposable when it is written as

$$D[\boldsymbol{p} : \boldsymbol{q}] = \sum d(p_i, q_i) \tag{31}$$

for function d. The invariance can be stated in terms of the divergence. Let us divide the set $X = \{0, 1, \dots, n\}$ into m subclasses $(m \leq n)$, $G = \{G_1, \dots, G_m\}$,

$$\cup \; G_i = X, \qquad G_i \cap G_j = \phi. \tag{32}$$

This is coarse graining of X. It induces a coarse-grained probability distribution over G. Given $\boldsymbol{p}$, the derived probability of $\bar{p}_i = \text{Prob}\{x \in G_i\}$ is

$$\bar{p}_i = \sum_{j \in G_i} p_j. \tag{33}$$

A divergence is said to be information-monotonic when it does not increase by coarse graining,

$$D[\boldsymbol{p} : \boldsymbol{q}] \geq D[\bar{\boldsymbol{p}} : \bar{\boldsymbol{q}}], \tag{34}$$

and the equality holds when and only when, for each class G_k,

$$\frac{q_i}{p_i} = \text{const} \quad i \in G_k. \tag{35}$$

The following theorem is known [23,24].

Theorem 3. *A decomposable invariant (information-monotonic) divergence is given by the following* f*-divergence* [23,24]

$$D[\boldsymbol{p} : \boldsymbol{q}] = \sum p_i f\left(\frac{q_i}{p_i}\right) \tag{36}$$

where f is a convex function satisfying $f(1) = 0$.

Invariant divergence $D[\boldsymbol{p} : \boldsymbol{q}]$ and its dual, $D[\boldsymbol{q} : \boldsymbol{p}]$, give a Riemannian metric together with dual affine connections. The following theorem is known [3].

Theorem 4. *The geometry derived from an invariant divergence consists of the Fisher information metric together with a pair of* $\pm\alpha$*-connections, where* α *is given by*

$$\alpha = 2f'''(1) + 3, \tag{37}$$

when f *is normalized to satisfy* $f''(1) = 1$.

3.3. Dually flat manifold

A Riemannian manifold with dual affine connections is dually flat when the curvature vanishes with respect to the dual affine connections. Such a manifold plays a fundamental role in information geometry. The following theorem is established in information geometry [3].

Theorem 5. *When a manifold M has dually flat structure, the following propositions hold:*
(1) *There are two affine coordinate systems,* $\boldsymbol{\theta}$ *and* $\boldsymbol{\eta}$*, with which the coefficients (Christoffel symbols)* Γ_{ijk} *and* Γ^*_{ijk} *of affine and dual affine connections vanish, respectively.*
(2) *There are two convex functions,* $\psi(\boldsymbol{\theta})$ *and* $\varphi(\boldsymbol{\eta})$*, with which the affine coordinates* $\boldsymbol{\theta}$ *and* $\boldsymbol{\eta}$ *are mutually connected by the Legendre transformation,*

$$\eta_i = \frac{\partial}{\partial\theta^i}\psi(\theta), \quad \theta^i = \frac{\partial}{\partial\eta_i}\varphi(\boldsymbol{\eta}), \tag{38}$$

$$\psi(\boldsymbol{\theta}) + \varphi(\boldsymbol{\eta}) - \boldsymbol{\theta}\cdot\boldsymbol{\eta} = 0. \tag{39}$$

(3) *Given two points P and Q in M, a canonical divergence exists and is given by*

$$D[P : Q] = \psi(\boldsymbol{\theta}_P) + \varphi\left(\boldsymbol{\eta}_Q\right) - \boldsymbol{\theta}_P\cdot\boldsymbol{\eta}_Q, \tag{40}$$

where $\boldsymbol{\theta}_P$ *and* $\boldsymbol{\theta}_Q$ *are the affine coordinates, and* $\boldsymbol{\eta}_P$ *and* $\boldsymbol{\eta}_Q$ *are the dual affine coordinates, of P and Q, respectively.*

We call $\psi(\boldsymbol{\theta})$ the free energy and $\varphi(\boldsymbol{\eta})$ the negative entropy for a dually flat manifold. We study the invariant geometry given to a χ-family. When a χ-family is dually flat, it naturally has a χ-free energy, χ-entropy and χ-divergence based on the invariant geometry.

We first search for a special class of χ-families of which invariant geometries are dually flat. The discrete case is studied for simplicity's sake. There are two well-known families, namely the exponential family written as

$$p(\boldsymbol{x}, \boldsymbol{\theta}) = \exp\{\boldsymbol{\theta}\cdot\boldsymbol{x} - \psi(\boldsymbol{\theta})\} \tag{41}$$

and the mixture family represented by

$$p(\boldsymbol{x}, \boldsymbol{\theta}) = \boldsymbol{\theta}\cdot\boldsymbol{k}(x), \qquad \sum\theta^i = 1, \tag{42}$$

where $k_i(x) > 0$ is a probability distribution on $x \in X$. They are special χ-families defined by

$$\exp_\chi(t) = \exp(t), \tag{43}$$

$$\exp_\chi(t) = t, \tag{44}$$

respectively.

Theorem 6. *The exponential and mixture families are the only two families that have an invariant dually flat geometry, provided the dimension number m of M is larger than 1. They are dually flat with respect to the e-connection* $(\alpha = 1)$ *and m-connection* $(\alpha = -1)$*. The invariant canonical divergence is the Kullback–Leibler divergence*

$$D[P : Q] = \sum p(x, \boldsymbol{\theta}_P)\log\frac{p(x, \boldsymbol{\theta}_P)}{p\left(x, \boldsymbol{\theta}_Q\right)}. \tag{45}$$

Remark 1. The case with $m = 1$ is trivial, since M is a curve and hence its Riemann–Christoffel curvature always vanishes.

The geometrical entropy is

$$-\varphi(\boldsymbol{\eta}) = -\sum p(x, \boldsymbol{\eta})\log p(x, \boldsymbol{\eta}) \tag{46}$$

in both cases, coinciding with the Shannon entropy. The geometrical free energy $\psi(\boldsymbol{\theta})$ is a cumulant generating function.

3.4. Characterization of q-family from the invariance viewpoint

In order to characterize the q-families, we extend the probability simplex, S_n, to a set of positive measures over X,

$$\boldsymbol{R}^{n+1}_+ = \left\{\boldsymbol{\xi}\,\middle|\,\xi^i > 0\right\}, \tag{47}$$

where ξ^i is the measure of $x = i$, and the normalization constraint $\sum \xi^i = 1$ is removed. It is known that $\boldsymbol{R}_+^{n+1}$ has an invariant and dually flat structure with respect to the α-connection for any α [24], while S_n is flat only with respect to $\alpha = \pm 1$ connections.

We have used parameter α from the information geometry point of view, where α and $-\alpha$ are dually coupled. In physics, it is traditional to use parameter q which is related to α by

$$q = \frac{1+\alpha}{2}, \qquad 1-q = \frac{1-\alpha}{2}. \tag{48}$$

Therefore, we hereafter use q instead of α frequently, such that q-connection implies $\alpha = 2q - 1$ connection.

Since $\boldsymbol{R}_+^{n+1}$ is flat with respect to the q-connection, the corresponding affine coordinate system, the q-coordinate system, is given by

$$\theta^i = \xi_i^{\frac{1-q}{2}} \tag{49}$$

and its dual, the $(1-q)$-coordinates, is given by

$$\eta_i = \xi_i^q. \tag{50}$$

Rescaling and adjusting constants, we have the q- and $(1-q)$-expression of affine coordinates

$$\theta^i = \ln_q(\xi_i) = \frac{1}{1-q}\left\{\xi_i^{1-q} - 1\right\}, \tag{51}$$

$$\eta_i = \ln_{1-q}(\xi_i) = \frac{1}{q}\left(\xi_i^q - 1\right). \tag{52}$$

The q-potential (q-free energy) is given by

$$\psi_q\left(\boldsymbol{\theta}_q\right) = (1-q)\sum\left(\theta_q^i\right)^{1-q} \tag{53}$$

and its dual is the $(1-q)$-potential (negative of q-entropy) given by

$$\varphi_{1-q}\left(\boldsymbol{\eta}_{1-q}\right) = q\left\{\sum\left(\eta_i^q\right)\right\}^{\frac{1}{q}}. \tag{54}$$

Hence, the Riemannian metric is

$$g_{ij} = \frac{q}{1-q}\left(\theta_q^i\right)^{\frac{2q-1}{1-q}}\delta_{ij}. \tag{55}$$

The q-entropy is also obtained from the invariant divergence. A dually-flat manifold has a canonical divergence. The canonical divergence with respect to the q-connection is

$$D_q\left[\tilde{\boldsymbol{p}} : \tilde{\boldsymbol{q}}\right] = (1-q)\sum \tilde{p}_i + q\sum \tilde{q}_i - \sum \tilde{p}_i^{(1-q)}\tilde{q}_i^q, \tag{56}$$

where $\tilde{\boldsymbol{p}}, \tilde{\boldsymbol{q}} \in \boldsymbol{R}_+^{n+1}$. Hence, the q-divergence in its subspace S_n is naturally given by

$$D_q[\boldsymbol{p} : \boldsymbol{q}] = 1 - \sum p_i^{(1-q)} q_i^q, \tag{57}$$

where $\sum p_i = \sum q_i = 1$ holds. The q-entropy of $\boldsymbol{p}$ is defined from the q-divergence by

$$H_q(\boldsymbol{p}) = -D_q[\mathbf{1} : \boldsymbol{p}], \tag{58}$$

where $\mathbf{1}$ is the uniform distribution. It reduces to the Tsallis entropy, except for a scale and constant,

$$H_q(\boldsymbol{p}) = \frac{1}{1-q}\left(\sum p_i^q - 1\right). \tag{59}$$

We define a q-family of probability distribution by

$$p(\boldsymbol{x}, \boldsymbol{\theta})^{1-q} = \boldsymbol{\theta} \cdot \boldsymbol{x} - \psi(\boldsymbol{\theta}), \tag{60}$$

of which $q = 0$ is the mixture family. The limiting case of $q = 1$ is the exponential family. In general, a q-family is

$$p(x, \boldsymbol{\theta})^{1-q} = \sum \theta_i c_i(x) - \psi(\boldsymbol{\theta}) \tag{61}$$

where $c_i(x)$ are positive measures.

Since $\boldsymbol{R}_+^{n+1}$ is dually flat with respect to the q-connection, we have naturally the Max–Ent theorem, which characterizes the q-family.

Theorem 7. *Given k constraints for constants a_i,*

$$\sum_{x=0}^{n} p^{\frac{1-\alpha}{2}}(x)c_i(x) = a_i, \quad i = 1, \ldots, k, \tag{62}$$

the family of positive measures that maximizes the q-entropy is the q-family

$$\{p(x; \boldsymbol{\theta})\}^q = \sum_{i=1}^{k} \theta_i c_i(x). \tag{63}$$

In regard to information geometry, the α-geometry has been well studied [3,25]. Its relation with the Tsallis entropy and physical applications are studied in Refs. [26,27].

4. Flat geometry of χ-family

4.1. Flat χ-geometry

Apart from the invariancy principle, we give a new dually flat structure to the χ-family, different from the invariant geometry. This is called the χ-geometry. When $\chi(s) = s^q$, $q > 0$, we particularly call it the q-geometry. The χ-free-energy $\psi(\boldsymbol{\theta})$ of a χ-family is known to be a convex function [6]. We give its simple proof for later use. We have, by putting $u(x) = \exp_\chi(x)$ for simplicity,

$$\partial_i p(\boldsymbol{x}, \boldsymbol{\theta}) = u'(\boldsymbol{\theta} \cdot \boldsymbol{x} - \psi)(x_i - \partial_i \psi), \tag{64}$$

$$\partial_i \partial_j p(\boldsymbol{x}, \boldsymbol{\theta}) = u''(\boldsymbol{\theta} \cdot \boldsymbol{x} - \psi)\left\{(x_i - \partial_i \psi)\left(x_j - \partial_j \psi\right)\right\} - u'(\boldsymbol{\theta} \cdot \boldsymbol{x} - \psi)\partial_i \partial_j \psi. \tag{65}$$

Since

$$\int \partial_i p(\boldsymbol{x}, \boldsymbol{\theta}) \mathrm{d}\boldsymbol{x} = \int \partial_i \partial_j p(\boldsymbol{x}, \boldsymbol{\theta}) \mathrm{d}\boldsymbol{x} = 0, \tag{66}$$

we have

$$\partial_i \partial_j \psi(\boldsymbol{\theta}) = \frac{\int u''(\boldsymbol{\theta} \cdot \boldsymbol{x} - \psi)(x_i - \partial_i \psi)\left(x_j - \partial_j \psi\right) \mathrm{d}\boldsymbol{x}}{h_\chi(\boldsymbol{\theta})}, \tag{67}$$

where

$$h_\chi(\boldsymbol{\theta}) = \int u'(\boldsymbol{\theta} \cdot \boldsymbol{x} - \psi) \mathrm{d}\boldsymbol{x} = \int \chi\{p(\boldsymbol{x}, \boldsymbol{\theta})\} \mathrm{d}\boldsymbol{x}. \tag{68}$$

Since

$$h_\chi(\boldsymbol{\theta}) > 0, \qquad u''(\boldsymbol{\theta} \cdot \boldsymbol{x} - \psi) > 0, \tag{69}$$

it is clear from (67) that

$$g_{ij}^\chi = \partial_i \partial_j \psi(\boldsymbol{\theta}) \tag{70}$$

is positive-definite and defines a Riemannian metric.

The χ-free-energy ψ is a convex function of $\boldsymbol{\theta}$ so that it gives a dually-flat structure, which is not invariant except for the cases of the exponential and mixture families. The χ-Fisher metric is given by (70), and is different from the invariant Fisher information matrix in general. The third-order tensor is given by

$$T_{ijk}^\chi = \partial_i \partial_j \partial_k \psi, \tag{71}$$

from which a pair of affine connections

$$\Gamma_{ijk}^\chi(\boldsymbol{\theta}) = [i, j, k]^\chi - \frac{1}{2} T_{ijk} \tag{72}$$

$$\Gamma_{ijk}^{*\chi}(\boldsymbol{\theta}) = [i, j; k]^\chi + \frac{1}{2} T_{ijk} \tag{73}$$

are derived, where $[i, j; k]^\chi$ is the coefficient of the Levi–Civita connection with respect to the Riemannian metric g_{ij}^χ. The connections are flat and dually coupled. (We consider only $\alpha = \pm 1$ connections here.) The affine coordinate system of the χ-geometry is $\boldsymbol{\theta}$ itself, so

$$\Gamma_{ijk}^\chi(\boldsymbol{\theta}) = 0. \tag{74}$$

Hence, the geodesic connecting two distributions $p(\boldsymbol{x}, \boldsymbol{\theta}_1)$ and $p(\boldsymbol{x}, \boldsymbol{\theta}_2)$ in a χ-family is written as

$$p(\boldsymbol{x}, t) = u\{t\boldsymbol{\theta}_1 \cdot \boldsymbol{x} + (1-t)\boldsymbol{\theta}_2 \cdot \boldsymbol{x} - \psi(t)\}. \tag{75}$$

4.2. Escort distribution as dual χ-affine system

Since the χ-geometry is dually flat, we have a dual affine coordinate system $\boldsymbol{\eta}$ in which $\Gamma^{*\chi}_{ijk}(\boldsymbol{\eta}) = 0$ holds. This is given by the Legendre transformation of $\psi(\boldsymbol{\theta})$,

$$\boldsymbol{\eta} = \nabla\psi(\boldsymbol{\theta}). \tag{76}$$

From (64), we have

$$\eta_i = \frac{\int u'(\boldsymbol{\theta}\cdot\boldsymbol{x} - \psi)x_i \mathrm{d}\boldsymbol{x}}{h_\chi(\boldsymbol{\theta})}, \tag{77}$$

where we put

$$h_\chi(\boldsymbol{\theta}) = \int u'(\boldsymbol{\theta}\cdot\boldsymbol{x} - \psi)\mathrm{d}\boldsymbol{x}. \tag{78}$$

In the case of S_n,

$$\boldsymbol{\theta}\cdot\boldsymbol{x} - \psi = \sum \theta^i\delta_i(x) - \psi(\boldsymbol{\theta}) \tag{79}$$

so

$$\eta_i = \frac{u'\left\{\theta^i - \psi(\boldsymbol{\theta})\right\}}{h_\chi(\boldsymbol{\theta})}, \tag{80}$$

$$h_\chi(\boldsymbol{\theta}) = \sum_{i=0}^{n} u'\left\{v\left(p_i\right)\right\}, \tag{81}$$

where

$$v(t) = \ln_\chi(t). \tag{82}$$

From the dual coordinates (80), a new family of probability distributions of $\boldsymbol{x}$ specified by $\boldsymbol{\theta}$ is defined as

$$\hat{p}(\boldsymbol{x},\boldsymbol{\theta}) = \frac{1}{h_\chi(\boldsymbol{\theta})}u'\left\{\boldsymbol{\theta}\cdot\boldsymbol{x} - \psi\right\}, \tag{83}$$

which is called a χ-escort distribution. Since the following equation holds:

$$\frac{\mathrm{d}}{\mathrm{d}t}\exp_\chi(t) = \chi\left\{\exp_\chi(t)\right\}, \tag{84}$$

we have

$$h_\chi(\boldsymbol{\theta}) = \int u'\left(\boldsymbol{\theta}\cdot\boldsymbol{x} - \psi\right)\mathrm{d}\boldsymbol{x} = \int \chi\left\{p(\boldsymbol{x})\right\}\mathrm{d}\boldsymbol{x}. \tag{85}$$

Hence, the escort distribution is written as

$$\hat{p}(\boldsymbol{x},\boldsymbol{\theta}) = \frac{\chi\left\{p(\boldsymbol{x},\boldsymbol{\theta})\right\}}{h_\chi(\boldsymbol{\theta})}. \tag{86}$$

The dual coordinates $\boldsymbol{\eta}$ are hence the expectation of $\boldsymbol{x}$ with respect to $\hat{p}(\boldsymbol{x},\boldsymbol{\theta})$,

$$\boldsymbol{\eta} = E_{\hat{p}}[\boldsymbol{x}], \tag{87}$$

which is the $\boldsymbol{\eta}$-coordinates of $\hat{p}(\boldsymbol{x},\boldsymbol{\theta})$. In particular, for a q-family, the following simple relations are obtained,

$$h_q(\boldsymbol{\theta}) = \int p(\boldsymbol{x},\boldsymbol{\theta})^q\mathrm{d}\boldsymbol{x}, \tag{88}$$

$$\hat{p}(\boldsymbol{x},\boldsymbol{\theta}) = \frac{1}{h_q(\boldsymbol{\theta})}p(\boldsymbol{x},\boldsymbol{\theta})^q. \tag{89}$$

In S_n, the following is easily obtained

$$\eta_i = \hat{p}(i,\boldsymbol{\theta}). \tag{90}$$

Hence, the escort distribution gives the dual coordinate system.

A dual geodesic is linear in $\boldsymbol{\eta}$ so that the dual geodesic connecting two distributions is given in terms of the escort distributions by

$$\hat{p}(\boldsymbol{x},t) = \frac{t}{h\left(\boldsymbol{\theta}_1\right)}\chi\left\{p\left(\boldsymbol{x},\boldsymbol{\theta}_1\right)\right\} + \frac{(1-t)}{h\left(\boldsymbol{\theta}_2\right)}\chi\left\{p\left(\boldsymbol{x},\boldsymbol{\theta}_2\right)\right\}. \tag{91}$$

4.3. χ-entropy

The dual of χ-free-energy, which we call negative χ-entropy, is

$$\varphi(\boldsymbol{\eta}) = \boldsymbol{\theta} \cdot \boldsymbol{\eta} - \psi(\boldsymbol{\theta}) \tag{92}$$

where $\boldsymbol{\theta}$ is considered as a function of $\boldsymbol{\eta}$. Obviously,

$$\nabla\varphi(\boldsymbol{\eta}) = \boldsymbol{\theta}. \tag{93}$$

Theorem 8. *The χ-entropy is given by*

$$\varphi(\boldsymbol{p}) = \frac{1}{h_\chi} \sum_{i=0}^{n} \frac{v(p_i)}{v'(p_i)}. \tag{94}$$

Proof. From

$$\varphi(\boldsymbol{\eta}) = \boldsymbol{\theta} \cdot \boldsymbol{\eta} - \psi(\boldsymbol{\theta}) \tag{95}$$

$$= E_{\hat{p}}\left[\boldsymbol{\theta} \cdot \boldsymbol{x} - \psi(\boldsymbol{\theta})\right] \tag{96}$$

$$= E_{\hat{p}}\left[v\left\{p(x)\right\}\right], \tag{97}$$

holds

$$\varphi(\boldsymbol{\eta}) = \frac{1}{h_\chi} \sum u'\left\{v(p_i)\right\} v(p_i). \quad \square \tag{98}$$

In the case of the q-family in S_n,

$$\varphi_q(\boldsymbol{\eta}) = \frac{1}{1-q}\left\{\frac{1}{h_q} - 1\right\}. \tag{99}$$

This shows that the q-entropy in the flat geometry is defined, except for a scale and a constant, by

$$H_q = \frac{1}{h_q} \tag{100}$$

instead of the Tsallis entropy,

$$H_{\text{Tsallis}} = -h_q, \tag{101}$$

which is the entropy due to the invariant geometry.

4.4. χ-divergence

The canonical divergence of the χ-family is the χ-divergence. It is obtained by Vigelis and Cavalcante [9] in the function space of probability distributions. It is, for $\boldsymbol{p} = p\left(\boldsymbol{x}, \boldsymbol{\theta}_p\right)$ and $\boldsymbol{r} = p\left(\boldsymbol{x}, \boldsymbol{\theta}_r\right)$, given as

$$D_\chi[\boldsymbol{p} : \boldsymbol{r}] = \psi\left(\boldsymbol{\theta}_p\right) + \varphi\left(\boldsymbol{\eta}_r\right) - \boldsymbol{\theta}_p \cdot \boldsymbol{\eta}_r. \tag{102}$$

Theorem 9. *The χ-divergence is given by*

$$D_\chi\left[\boldsymbol{r} : \boldsymbol{p}\right] = E_{\hat{p}}\left[\ln_\chi\left\{p\left(\boldsymbol{x}, \boldsymbol{\theta}_p\right)\right\} - \ln_\chi\left\{p\left(\boldsymbol{x}, \boldsymbol{\theta}_r\right)\right\}\right]. \tag{103}$$

Proof. The right-hand side of (103) is calculated as

$$\frac{1}{h_\chi(\boldsymbol{p})} \sum_{i=0}^{n} \left[u'\left\{v(p_i)\right\} v(p_i) - u'\left\{v(p_i)\right\} v(r_i)\right] \tag{104}$$

$$= \varphi\left(\boldsymbol{\eta}_p\right) - \frac{1}{h_\chi(p)} \sum_{i=1}^{n} u'\left\{v(p_i)\right\}\left\{\theta_r^i + v(r_0)\right\} \tag{105}$$

$$= \varphi\left(\boldsymbol{\eta}_p\right) - v(r_0) - \sum \boldsymbol{\theta}_r \cdot \boldsymbol{\eta}_p \tag{106}$$

$$= \psi\left(\boldsymbol{\theta}_r\right) + \varphi\left(\boldsymbol{\eta}_p\right) - \boldsymbol{\theta}_r \cdot \boldsymbol{\eta}_p \tag{107}$$

$$= D_\chi\left[\boldsymbol{r} : \boldsymbol{p}\right]. \quad \square \tag{108}$$

Corollary 10. *The χ-entropy is related to the χ-divergence by*

$$H_\chi(\boldsymbol{p}) = -D[\mathbf{1} : \boldsymbol{p}] = -\varphi_\chi\left(\boldsymbol{\eta}_p\right) + \psi(\mathbf{1}). \tag{109}$$

The Pythagorean theorem holds with respect to the χ-divergence.

Theorem 11. *For three points $\boldsymbol{p}, \boldsymbol{q}, \boldsymbol{r}$ in the χ-family,*

$$D_\chi[\boldsymbol{p} : \boldsymbol{q}] + D_\chi[\boldsymbol{q} : \boldsymbol{r}] = D_\chi[\boldsymbol{p} : \boldsymbol{r}] \tag{110}$$

holds, when the dual χ-geodesic connecting $\boldsymbol{p}$ and $\boldsymbol{q}$ is orthogonal to the χ-geodesic connecting $\boldsymbol{q}$ and $\boldsymbol{r}$ with respect to the χ-metric.

4.5. χ-version of Max–Ent theorem

The Max–Ent theorem is a direct consequence of the generalized Pythagorean theorem in a dually flat manifold. Let us consider the following constraints imposed on the probability distributions

$$E_{\hat{p}}[c_i(\boldsymbol{x})] = a_i, \quad i = 1, \dots, k. \tag{111}$$

The constraints are linear in the escort distributions, that is, linear with respect to the dual affine coordinates $\boldsymbol{\eta}$. Hence, submanifold $M(\boldsymbol{a})$ with $\boldsymbol{a} = (a_1, \dots, a_k)$, consisting of distributions satisfying the constraints (111), are dually flat.

Given $M(\boldsymbol{a})$, we search for the distribution $q(\boldsymbol{x}, \boldsymbol{a})$ that maximizes the χ-entropy H_χ for given $\boldsymbol{a} = (a_1, \dots, a_k)$:

$$q(\boldsymbol{x}, \boldsymbol{a}) = \arg\max\left\{H_\chi(q) \mid q \in M(\boldsymbol{a})\right\}. \tag{112}$$

The χ-entropy is the negative of the χ-divergence from q to the uniform distribution $\mathbf{1}$,

$$H_\chi(q) = c - D_\chi[\mathbf{1} : q(\boldsymbol{x})]. \tag{113}$$

Hence, the maximizer is given by the geodesic projection of $\mathbf{1}$ to $M(\boldsymbol{a})$. Let us denote it by $p(x)$. Indeed, the Pythagorean theorem

$$D_\chi[\mathbf{1} : q(x)] = D_\chi[\mathbf{1} : p(x)] + D_\chi[p(x) : q(x)] \tag{114}$$

shows that $D_\chi[\mathbf{1} : \boldsymbol{p}]$ is minimized when $q(x) = p(x)$.

Such $p(x)$'s form a geodesically flat submanifold intersecting $M(\boldsymbol{a})$'s orthogonally. Hence, they form a χ-family,

$$p(\boldsymbol{x}, \boldsymbol{\theta}) = u\left\{\sum_{i=1}^{k} \theta_i c_i(\boldsymbol{x}) - \psi(\boldsymbol{\theta})\right\}. \tag{115}$$

5. Conformal transformation connecting invariant and flat geometries

Two geometries, invariant and flat, induced in the χ-structure of S_n were studied in the previous sections. It is interesting to know how they are related. In order to answer this question, we calculate the χ-Fisher metric by using the χ-divergence.

Theorem 12. *The χ-Fisher metric of S_n is given by*

$$g_{ij}^\chi = -\frac{1}{2h_\chi}\left\{v''(p_i)\, u'(v(p_i))\, \delta_{ij} + v''(p_0)\, u'(v(p_0))\right\}, \quad i, j = 1, \dots, n. \tag{116}$$

Proof. The Taylor expansion of the χ-divergence is calculated as

$$D_\chi[\boldsymbol{p} : \boldsymbol{p} + \mathrm{d}\boldsymbol{p}] = \frac{1}{h_\chi}\sum_{i=0}^{n} u'\{v(p_i)\}\{v(p_i) - v(p_i + \mathrm{d}p_i)\} \tag{117}$$

$$= -\frac{1}{2h_\chi}\sum u'\{v(p_i)\}\, v''(p_i)\, \mathrm{d}p_i^2. \tag{118}$$

Hence, the χ-Fisher metric on $\boldsymbol{R}_+^{n+1}$ is diagonal,

$$g_{ij}^\chi = \gamma(p_i)\,\delta_{ij}, \quad i, j = 0, \dots, n, \tag{119}$$

having diagonal terms

$$\gamma\left(p_i\right) = -\frac{1}{2h_\chi} u'\left\{v\left(p_i\right)\right\} v''\left(p_i\right). \tag{120}$$

Here, the homogeneous coordinates $(p_0, p_1, \ldots, p_n)$ are used for S_n so that $\mathrm{d}\boldsymbol{p}$ satisfies

$$\mathrm{d}p_0 = -\sum_{i=1}^{n} \mathrm{d}p_i. \tag{121}$$

The χ-Fisher information in terms of $p_1, \ldots, p_n$ is calculated from this by eliminating $\mathrm{d}p_0$. Then we have (116). □

The invariant Fisher metric g is given by

$$\sum_{i,j=1}^{n} g_{ij} \mathrm{d}p_i \mathrm{d}p_j = \sum_{i,j=0}^{n} \frac{1}{p_i} \delta_{ij} \mathrm{d}p_i \mathrm{d}p_j \tag{122}$$

with equality constraint (121).

A transformation of the metric g in a Riemannian manifold is said to be conformal, when the new metric $\tilde{g}$ is given by

$$\tilde{g}_{ij}(\boldsymbol{p}) = \sigma(\boldsymbol{p}) g_{ij}(\boldsymbol{p}) \tag{123}$$

for a positive scalar function $\sigma(\boldsymbol{p})$. We search for the condition that the χ-geometry is a conformal transform of the invariant geometry.

Theorem 13. *The q-geometry is unique in the class of χ-geometries that is derived by a conformal transformation from the invariant geometry.*

Proof. Since

$$-2h_\chi(\boldsymbol{p})\gamma\left(p_i\right) = \frac{v''\left(p_i\right)}{v'\left(p_i\right)} = \frac{\mathrm{d}}{\mathrm{d}p_i} \log v'\left(p_i\right), \tag{124}$$

a conformal change is derived when

$$\frac{\mathrm{d}}{\mathrm{d}p_i} \log v'\left(p_i\right) = \frac{c}{p_i}, \tag{125}$$

where c is a constant not depending on i. This gives

$$\log v'\left(p_i\right) = c \log p_i + \mathrm{d} \tag{126}$$

and hence

$$v'\left(p_i\right) = \mathrm{d}' p_i^c, \tag{127}$$

for constants d and d′. This holds only for the q-logarithm with $q = -c$. □

The conformal geometry of a dual Riemannian manifold is given by

$$\tilde{g}_{ij} = \sigma g_{ij} \tag{128}$$

$$\tilde{T}_{ijk} = \sigma T_{ijk} + \left(\partial_i \sigma\right) g_{jk} + \left(\partial_j \sigma\right) g_{ik} + \left(\partial_k \sigma\right) g_{ij}. \tag{129}$$

(See Refs. [28–32]). The above transformations are derived from the conformal change of divergence

$$\tilde{D}[\boldsymbol{p} : \boldsymbol{q}] = \sigma(\boldsymbol{p}) D[\boldsymbol{p} : \boldsymbol{q}]. \tag{130}$$

The q-geometry is the case of

$$\sigma(\boldsymbol{p}) = \frac{1}{h_q(\boldsymbol{p})}. \tag{131}$$

It should be remarked that

$$-\log \sigma(\boldsymbol{p}) = H_{\mathrm{Renyi}}(\boldsymbol{p}) \tag{132}$$

gives the Renyi entropy except for a scale factor and constant. In other words, the Renyi entropy is the negative of the power exponent of the conformal transformation.

6. Conclusions

The geometrical structures of the manifold of a general deformed exponential family were studied. The two geometrical structures were introduced, one from the viewpoint of invariance and the other from the viewpoint of flatness. They give different definitions of generalized free energy, entropy and divergence. The exponential and mixture families are characterized uniquely by the fact that the two geometries coincide. The q-exponential families are characterized by the fact that the two geometries coincide in the extended manifold of positive measures. It is also the unique family for which the Riemannian metrics of the two geometries are connected by the conformal transformation, where the Renyi entropy plays the role of expansion exponent.

References

[1] C. Tsallis, Possible generalization of Boltzmann–Gibbs statistics, J. Stat. Phys. 52 (1988) 479–487.
[2] C. Tsallis, Introduction to Nonextensive Statistical Mechanics, Springer, New York NY, USA, 2009.
[3] S. Amari, H. Nagaoka, Methods of Information Geometry (Translations of Mathematical Monographs), Oxford University Press, Oxford, UK, 2000.
[4] J. Naudts, Estimators, escort probabilities, and phi-exponential families in statistical physics, J. Ineq. Pure Appl. Math. 5 (2004) 102.
[5] J. Naudts, Generalized exponential families and associated entropy functions, Entropy 10 (2008) 131–149.
[6] J. Naudts, Generalized Thermostatistics, Springer, London, 2011.
[7] G. Kaniadakis, M. Lissia, A.M. Scarfone, Deformed logarithms and entropies, Physica A 340 (2004) 41–49.
[8] G. Pistone, κ-exponential models from the geometrical viewpoint, Eur. Phys. J. B 70 (2009) 29–37.
[9] R. Vigelis, C. Cavalcante, On the φ-exponential family of probability distributions, J. Theor. Probab. 21 (2011) 1–15.
[10] F. Topsœ, Entropy and equilibrium via games of complexity, Physica A 340 (2004) 11–31.
[11] C. Beck, Generalized information and entropy measures in physics, Contemp. Phys. 50 (2009) 495–510.
[12] P.D. Grunwald, A.P. David, Game theory, maximum entropy, minimum discrepancy and robust Bayesian decision theory, Ann. Stat. 32 (2004) 1367–1433.
[13] S. Eguchi, Information geometry and statistical pattern recognition, Sugaku Expositions 19 (2006) 197–216 (in Japanese) (originally Sūgaku 56 (2004) 380–399.
[14] Y. Fujimoto, N. Murata, A generalization of independence in naive Bayes model, Lect. Notes Comput. Sci. 6283 (2010) 153–161.
[15] N.N. Chentsov, Statistical Decision Rules and Optimal Inference, Amer. Math. Soc., Rhode Island, 1982 (originally published in Russian Nauka, Moscow, 1972).
[16] H. Shima, The Geometry of Hessian Structure, World Scientific, Singapore, 2007.
[17] A. Ohara, H. Matsuzoe, S.-I. Amari, A dually flat structure on the space of escort distributions, J. Phys.: Conf. Ser. 201 (2010) 012012 (electronic).
[18] S. Amari, A. Ohara, Geometry of q-exponential family of probability distributions, Entropy 13 (2011) 1170–1185.
[19] J. Naudts, Deformed exponentials and logarithms in generalized thermostatistics, Physica A 316 (2002) 323–334.
[20] G. Pistone, C. Sempi, An infinite-dimensional geometric structure on the space of all the probability measures equivalent to a given one, Ann. Stat. 23 (1995) 1543–1561.
[21] A. Cena, G. Pistone, Exponential statistical manifold, Ann. Inst. Statist. Math. 59 (1) (2007) 27–56.
[22] M. Grasselli, Dual connections in nonparametric classical information geometry, Annals Inst. Statist. Math. 62 (2010) 873–896.
[23] I. Csiszár, Why least squares and maximum entropy? an axiomatic approach to inference for linear inverse problems, Ann. Statist. 19 (1991) 2032–2066.
[24] S. Amari, α-divergence is unique, belonging to both f-divergence and Bregman divergence classes, IEEE Trans. Inform. Theory 55 (2009) 4925–4931.
[25] S. Amari, Differential-Geometrical Methods in Statistics, in: Lecture Notes in Statistics, vol. 28, Springer, New York, 1985.
[26] A. Ohara, Geometry of distributions associated with Tsallis statistics and properties of relative entropy minimization, Phys. Lett. A 370 (2007) 184–193.
[27] A. Ohara, Geometric study for the Legendre duality of generalized entropies and its application to the porous medium equation, Eur. Phys. J. B 70 (2009) 15–28.
[28] I. Okamoto, S. Amari, K. Takeuchi, Asymptotic theory of sequential estimation: differential geometrical approach, Ann. Statist. 19 (1991) 961–981.
[29] T. Kurose, On the divergence of 1-conformally flat statistical manifolds, Tôhoku Math. J. 46 (1994) 427–433.
[30] H. Matsuzoe, On realization of conformally-projectively flat statistical manifold and the divergences, Hokkaido Math. J. 27 (1998) 409–421.
[31] H. Matsuzoe, Geometry of contrast functions and conformal geometry, Hiroshima Math. J. 29 (1999) 175–191.
[32] T. Kurose, Conformal-projective geometry of statistical manifolds, Interdiscip. Inform. Sci. 8 (2002) 89–100.