Information geometry of Bayesian statistics

Hiroshi Matsuzoe

Department of Computer Science and Engineering, Graduate School of Engineering, Nagoya Institute of Technology, Nagoya 466-8555, Japan

Abstract. A survey of geometry of Bayesian statistics is given. From the viewpoint of differential geometry, a prior distribution in Bayesian statistics is regarded as a volume element on a statistical model. In this paper, properties of Bayesian estimators are studied by applying equiaffine structures of statistical manifolds. In addition, geometry of anomalous statistics is also studied. Deformed expectations and deformed independeces are important in anomalous statistics. After summarizing geometry of such deformed structures, a generalization of maximum likelihood method is given. A suitable weight on a parameter space is important in Bayesian statistics, whereas a suitable weight on a sample space is important in anomalous statistics.

Keywords: information geometry, equiaffine structure, statistical manifold, anomalous statistics, Tsallis statistics

PACS: 02.40.Ky, 02.50.Tt, 02.50.Cw

INTRODUCTION

Information geometry is a differential geometric approach of statistical inferences. After the formulation of dual affine connections on statistical models by Amari and Nagaoka (cf. [1]), information geometry has been applied various fields of mathematical sciences (e.g. [3] and [10]). In particular, a Riemannian manifold with mutually dually flat affine connection is called a dually flat space. This geometric structure has close relations among geometry of statistical inferences and relative entropies.

Information geometry of Bayesian statistics has been studied by Komaki [4], Takeuchi and Amari [14], and Takeuchi, Amari and Matsuzoe [6], etc. In Bayesian statistics, a prior distribution on a parameter space plays an important role. From a viewpoint of differential geometry, a prior distribution is regarded as a volume element on a manifold of statistical model since a parameter space is a local coordinate system of a statistical model. For the arguments of projected Bayesian estimators and bias correction of estimators, differential geometric gradient vector fields of volume elements, called the Tchebychev vector fields, are important.

On the other hand, a method of statistical inference based on non-additive entropy has been studied (cf. [2], [9], and [11]), which is called Tsallis statistics or anomalous statistics. In this method, deformed probability distributions, called escort distributions, play important roles. (See Definition 9 and Proposition 10.) Deformations of probability measures on sample spaces are important in anomalous statistics while deformations of measures on parameter spaces are important in Bayesian statistics. A dually flat structure on a deformed exponential family is introduced from the escort distribution.

In this paper, we give a survey about information geometry of Bayesian statistics. Then we consider relations between Bayesian statistics and anomalous statistics.

GEOMETRY OF STATISTICAL MODELS

We assume that all the objects are smooth throughout this paper. We also assume that differentials and integrals are interchangeable, and a given manifold is an open domain.

Let Ω be a sample space, and let Ξ be an open subset on \mathbb{R}^n . We say that S is a *statistical model* (or a *parametric model*) on Ω if S is a set of probability densities with parameter $\xi \in \Xi$, that is,

$$S = \left\{ p(x;\xi) \left| \int_{\Omega} p(x;\xi) dx = 1, p(x;\xi) > 0, \xi \in \Xi \subset \mathbb{R}^n \right\}.$$

Under suitable conditions, *S* can be regarded as a manifold with local coordinate system $\{\Xi, \xi^1, \ldots, \xi^n\}$. (cf. [1])

We define a symmetric (0,2)-tensor field g^F and a totally symmetric (0,3)-tensor field T^F on S by

$$\begin{split} g_{ij}^{F}(\xi) &= \int_{\Omega} \left(\frac{\partial}{\partial \xi^{i}} \log p(x;\xi) \right) \left(\frac{\partial}{\partial \xi^{j}} \log p(x;\xi) \right) p(x;\xi) dx \\ &= E_{p}[\partial_{i}l_{\xi}\partial_{j}l_{\xi}], \\ T_{ijk}^{F}(\xi) &= \int_{\Omega} \left(\frac{\partial}{\partial \xi^{i}} \log p(x;\xi) \right) \left(\frac{\partial}{\partial \xi^{j}} \log p(x;\xi) \right) \left(\frac{\partial}{\partial \xi^{k}} \log p(x;\xi) \right) p(x;\xi) dx \\ &= E_{p}[\partial_{i}l_{\xi}\partial_{j}l_{\xi}\partial_{k}l_{\xi}], \end{split}$$

where $\partial_i = \partial/\partial \xi^i$, $l_{\xi} = \log p(x; \xi)$ and $E_p[*]$ is the standard expectation with respect to $p(x; \xi)$. Under suitable conditions, g^F is a Riemannian metric on *S*. We call g^F the *Fisher metric* on *S*, and T^F the *cubic form* (or the *skewness tensor field*) on *S*.

Fix a real number $\alpha \in R$. We define the α -connection $\nabla^{(\alpha)}$ on *S* by

$$g^F(\nabla_X^{(\alpha)}Y,Z) = g^F(\nabla_X^{(0)}Y,Z) - \frac{\alpha}{2}T^F(X,Y,Z),$$

where $\nabla^{(0)}$ is the Levi-Civita connection on *S* with respect to g^F . An α -connection $\nabla^{(\alpha)}$ is torsion-free and satisfies $(\nabla^{(\alpha)}_X g^F)(Y,Z) = (\nabla^{(\alpha)}_Y g^F)(X,Z)$. Two affine connections $\nabla^{(\alpha)}$ and $\nabla^{(-\alpha)}$ satisfy

$$Xg^{F}(Y,Z) = g^{F}(\nabla_{X}^{(\alpha)}Y,Z) + g^{F}(Y,\nabla_{X}^{(-\alpha)}Z).$$

We say that $\nabla^{(\alpha)}$ and $\nabla^{(-\alpha)}$ are mutually *dual* with respect to g^F .

A statistical model S is said to be an *exponential family* if

$$S = \left\{ p(x;\theta) \left| p(x;\theta) = \exp\left[Z(x) + \sum_{i=i}^{n} \theta^{i} F_{i}(x) - \psi(\theta) \right], \theta \in \Theta \subset \mathbb{R}^{n} \right\},\$$

where F_1, \ldots, F_n, Z are functions on Ω , Θ is an open subset in \mathbb{R}^n , and ψ is a function on Θ . Without loss of generality, we can choose the dominating measure on Ω such that Z(x) = 0. In this paper, we assume that Z(x) = 0.

Suppose that *M* is a submanifold of *S*. We call *M* a *curved exponential family* of *S*.

BAYESIAN STATISTICS OF CURVED EXPONENTIAL FAMILIES

In this section, we give a review of Bayesian statistics of curved exponential families. For more details, see [4], [6] and [14]. Our notation follow those adopted in their papers. From now on, we assume that subscripts and superscripts vary over the noted range values otherwise stated.

$$a, b, \ldots = 1, 2, \ldots, m, \qquad i, j, \ldots = 1, 2, \ldots, n$$

Let *S* be an exponential family $(\dim S = n)$ and let *M* be a curved exponential family embedded in *S* $(\dim M = m)$. We assume that $p(x; u) = p(x; \theta(u)) \in M$ is an underlying distribution on *M*. Let $\rho(u)du$ be a prior distribution on *M*. The *posterior distribution* for given *x* with respect to $\rho(u)du$ is defined by

$$\rho'(u|x)du = \frac{p(x;u)\rho(u)}{\int_U p(x;u)\rho(u)du}du$$

We recall that we regarded a statistical model as a manifold, and the integration of the posterior distribution is carried out on a parameter space. Hence a prior distribution is regarded as a volume element (*n*-th differential form) on M. (cf. [14] and [6])

Let $x^N = (x_1, ..., x_N)$ is *N*-observations generated from $p(x; u) \in M$. We define the *Bayesian mixture distribution* by

$$f_{\rho}[x^N](x) = \int_U p(x;u)\rho'(u|x^N)du.$$

The Bayesian mixture distribution $f_{\rho}[x^N](x)$ is not contained in *M* generally. We consider the projection of $f_{\rho}[x^N](x)$ to *M* with respect to the Kullback-Leibler divergence.

For two points $p(x; \theta)$ and $p(x; \theta')$ in *S*, we define the *Kullback-Leibler divergence* (or the *relative entropy*) by

$$D(p(x;\theta)||p(x;\theta')) = \int_{\Omega} p(x;\theta) \log \frac{p(x;\theta)}{p(x;\theta')} dx.$$

Nevertheless *D* is not a distance, it measures a dissimilarity of two probability distributions. Hence the *projected Bayes estimator* $u(\tilde{f}_{\rho}[x^N])$ with respect to a prior distribution $\rho(u)du$ is defined by

$$u(\tilde{f}_{\rho}[x^N]) = \operatorname*{argmin}_{u \in U} D(f_{\rho}[x^N] || p(x^N; u)).$$

In this case, the estimated distribution is given by $p(x; u(\tilde{f}_{\rho}[x^N]))$.

Denote by \hat{u} the maximum likelihood estimator from x^N . Then two estimators $u(\tilde{f}_{\rho}[x^N])$ and \hat{u} are related as follows:

$$u^{c}(\tilde{f}_{\rho}[x^{N}]) = \hat{u}^{c} + \frac{1}{N} \sum_{a=1}^{n} \left(\frac{\partial}{\partial u^{a}} \log \rho(\hat{u}) - \sum_{b=1}^{m} \hat{\Gamma}_{ab}^{(1)b} \right) \hat{g}^{ac} + o\left(\frac{1}{N}\right), \tag{1}$$

where \hat{g}^{ab} is the inverse matrix of Fisher metric at the maximum likelihood estimation \hat{u} , and $\hat{\Gamma}_{ab}^{(1)c}$ is the 1-connection coefficient at \hat{u} . (cf. [4] and [14])

STATISTICAL MANIFOLDS AND EQUIAFFINE STRUCTURES

For Bayesian statistics, volume elements on statistical models have important roles from the viewpoint of differential geometry. In this section, we consider geometry of volume elements.

Definition 1 Let (M,g) be a Riemannian manifold, and let T be a totally symmetric (0,3)-tensor field on M. The triplet (M,g,T) is called a *statistical manifold*, and the tensor field T is called a *cubic form* (or a *skewness tensor field*).

In this case, for a fixed $\alpha \in R$, we can define a torsion-free affine connection by

$$g(\nabla_X^{(\alpha)}Y,Z) = g(\nabla_X^{(0)}Y,Z) - \frac{\alpha}{2}T(X,Y,Z),$$
(2)

where $\nabla^{(0)}$ is the Levi-Civita connection with respect to g. The connection $\nabla^{(\alpha)}$ is called an α -connection. Two affine connections $\nabla^{(\alpha)}$ and $\nabla^{(-\alpha)}$ is mutually *dual* with respect to g. The difference of two affine connections $\nabla^{(\alpha)}$ and $\nabla^{(\beta)}$ is given by

$$g(\nabla_X^{(\alpha)}Y - \nabla_X^{(\beta)}Y, Z) = \frac{\beta - \alpha}{2}T(X, Y, Z).$$

Next, let us define an equiaffine structure on a manifold. Let M be a manifold $(\dim M = m)$, and ∇ a torsion-free affine connection on M. Let ω be a volume element on M. That is, ω is an *m*-th differential form which does not vanish everywhere on M.

Definition 2 A pair $\{\nabla, \omega\}$ is said to be a *(locally) equiaffine structure* on *M* if ω is parallel with respect to ∇ , that is, $\nabla \omega = 0$. In this case, ∇ is called a *(locally) equiaffine connection* and ω is called a *parallel volume element* on *M*. (cf. [12] and [13])

Example 3 Let (M,g) be a Riemannian manifold. Denote by $\nabla^{(0)}$ the Levi-Civita connection on M, and set an *m*-form $\omega^{(0)}$ by

$$\omega^{(0)} = (\det |g_{ab}|)^{1/2} du$$

= $(\det |g_{ab}|)^{1/2} du^1 \wedge \dots \wedge du^m$

Then $\{\nabla^{(0)}, \omega^{(0)}\}$ is an equiaffine structure on *M*. In particular, the normalized volume element

$$\omega^{(0)} = \frac{(\det|g_{ab}|)^{1/2}}{\int_U (\det|g_{ab}|)^{1/2} du} du$$
(3)

is called the Jeffreys prior distribution in Bayesian statistics.

For an equiaffine structure $\{\nabla, \omega\}$, the following holds. (See Proposition 1 in [14].)

Lemma 4 Let ∇ be a torsion-free affine connection on M. Suppose that a volume element ω on M is given by $\omega (\partial/\partial u^1, \ldots, \partial/\partial u^m) = \rho(u) du^1 \wedge \cdots \wedge du^m$. Then $\{\nabla, \omega\}$ is equiaffine if and only if

$$\frac{\partial}{\partial u^a}\log\rho(u) = \sum_{b=1}^m \Gamma^b_{ab},$$

where Γ_{ab}^{c} is a connection coefficient of ∇ .

For a given statistical manifold (M, g, T), we define the *Tchebychev form* by

$$\tau(X) = \operatorname{trace}_g\{(Y,Z) \mapsto T(X,Y,Z)\},\$$

and its metric dual vector field is the *Tchebychev vector field*, that is, ${}^{\#}\tau$ is defined by

$$g(^{\#}\tau, X) = \tau(X).$$

Denote by $\operatorname{Ric}^{(\alpha)}$ the Ricci tensor field of $\nabla^{(\alpha)}$. That is, $\operatorname{Ric}^{(\alpha)}(X,Y) = \operatorname{trace}\{Z \mapsto R^{(\alpha)}(Z,X)Y\}$, where $R^{(\alpha)}$ is the curvature tensor field of $\nabla^{(\alpha)}$.

Proposition 5 Let (M, g, T) be a statistical manifold. Then the following conditions are equivalent:

- 1. $\nabla^{(\alpha)}$ is equiaffine.
- 2. The Ricci tensor field is symmetric, i.e., $\operatorname{Ric}^{(\alpha)}(X,Y) = \operatorname{Ric}^{(\alpha)}(Y,X)$.
- 3. The Tchebychev form T is closed, i.e. $d\tau = 0$.

In this case, there exists a function ϕ on M such that $\tau = d\phi$.

From Equation (3), a Levi-Civita connection is always equiaffine. On the other hand, an α -connection $\nabla^{(\alpha)}$ is equiaffine if and only if its Ricci tensor is symmetric.

Proposition 6 Let (M, g, T) be a statistical manifold. Suppose that $\nabla^{(\alpha)}$ and $\nabla^{(-\alpha)}$ are mutually dual affine connections determined by g and T, and ϕ is a function on M determined by the Tchebychev form τ , i.e., $\tau = d\phi$. Then $\{\nabla^{(\alpha)}, \omega^{(\alpha)}\}$ is equiaffine if and only if $\{\nabla^{(-\alpha)}, \omega^{(-\alpha)}\} = \{\nabla^{(-\alpha)}, e^{-\alpha\phi}\omega^{(\alpha)}\}$ is equiaffine.

We remark that the proposition above implies that the Tchebychev vector field is the gradient vector field of logarithmic ratio of volume elements.

BAYESIAN STATISTICS OF α **-PARALLEL PRIORS**

As we have seen in the previous section, the Jeffreys prior is a parallel volume element of the Levi-Civita connection with respect to the Fisher metric. From the viewpoint of this geometric property, we generalize the Jeffreys prior to another equiaffine connection.

Suppose that *M* is a curved exponential family which is embedded into *S*, and $\nabla^{(\alpha)}$ is an α -connection on *M* whose Ricci tensor field is symmetric. Then we say that a prior distribution $\omega^{(\alpha)}$ is an α -parallel prior if $\nabla^{(\alpha)}\omega^{(\alpha)} = 0$.

For example, suppose that $\{u^1, \ldots, u^m\}$ is an affine coordinate of $\nabla^{(\alpha)}$. That is, the connection coefficients $\Gamma_{ab}^{(\alpha)c}$ vanish for all a, b, c. Then a uniform prior $\omega^{(\alpha)} = du^1 \wedge \cdots \wedge du^m$ is an α -parallel prior on M.

From Equations 1, 2, and Lemma 4, the following Proposition holds:

Proposition 7 Let \hat{u} be a maximum likelihood estimator, and let $u(\tilde{f}_{\rho}[x^N])$ by a projected Bayes estimator with respect to an α -parallel prior distribution. Denote by $^{\#}\hat{\tau}$ the Tchebychev vector at \hat{u} . Then we obtain

$$u^{c}(\tilde{f}_{\rho}[x^{N}]) = \hat{u}^{c} + \frac{1-\alpha}{N}^{\#}\hat{\tau}^{c} + o\left(\frac{1}{N}\right).$$

GEOMETRY OF *q***-EXPONENTIAL FAMILIES**

In this section, we review geometry of *q*-exponential families. For more details of deformed exponential families, see [2], [7], [8] and [9]. The framework based on deformed exponential families is called the *anomalous statistics*.

To begin with, we generalize the exponential and the logarithm functions.

Definition 8 Fix a positive real number q. The q-exponential function is defined by

$$\exp_q x = (1 + (1 - q)x)^{\frac{1}{1 - q}}, \quad (q \neq 1, \text{ and } 1 + (1 - q)x > 0),$$

and the *q*-logarithm function by

$$\log_q x = \frac{x^{1-q}-1}{1-q}, \quad (q \neq 1, \text{ and } x > 0).$$

By taking the limit $q \rightarrow 1$, the standard exponential and the standard logarithm are recovered, respectively. For further generalizations of exponential functions, see [11].

A statistical model S_q is said to be a *q*-exponential family if

$$S_q = \left\{ p(x;\theta) \left| p(x;\theta) = \exp_q \left[\sum_{i=i}^n \theta^i F_i(x) - \psi(\theta) \right], \theta \in \Theta \subset \mathbb{R}^n \right\} \right\}$$

where F_1, \ldots, F_n are functions on Ω , Θ is an open subset in \mathbb{R}^n , and ψ is a function on Θ . In the same manner as an exponential family, S_q is regarded as a manifold with local coordinate system $\{\Theta; \theta^1, \ldots, \theta^n\}$. The normalization function ψ is convex, but it may not be strictly convex. Hence we assume that ψ is strictly convex from now on.

For a q-exponential family S_q , we define the q-Fisher metric and the q-cubic form by

$$g_{ij}^{q}(\theta) = \partial_i \partial_j \psi(\theta), \qquad T_{ijk}^{q}(\theta) = \partial_i \partial_j \partial_k \psi(\theta),$$

respectively, where $\partial_i = \partial/\partial \theta^i$. Since ψ is strictly convex, g^q is a Riemannian metric on S_q . For a fixed real number α , set

$$g^{q}\left(\nabla_{X}^{q(\alpha)}Y,Z\right) = g^{q}\left(\nabla_{X}^{q(0)}Y,Z\right) - \frac{\alpha}{2}T^{q}\left(X,Y,Z\right),$$

where $\nabla^{q(0)}$ is the Levi-Civita connection with respect to g^q . The affine connection $\nabla^{q(\alpha)}$ is torsion-free. In particular, $\nabla^{q(e)} := \nabla^{q(1)}$ and $\nabla^{q(m)} := \nabla^{q(-1)}$ are flat affine connections and mutually dual with respect to g^q . Hence the quadruplet $(S_q, g^q, \nabla^{q(e)}, \nabla^{q(m)})$ is a dually flat space.

Next, we consider deformed expectations for q-exponential families.

Definition 9 We say that $P_q(x)$ is the *escort distribution* of $p(x) \in S_q$ if

$$P_q(x) = \frac{1}{Z_q(p)} p(x)^q$$
, where $Z_q(p) = \int_{\Omega} p(x)^q dx$.

The *q*-expectation of a function f(x) is defined by

$$E_{q,p}[f(x)] = \int_{\Omega} f(x)P_q(x)dx = \frac{1}{Z_q(p)}\int_{\Omega} f(x)p(x)^q dx.$$

Under q-expectations, we have the following proposition. (cf. [7])

Proposition 10 Let S_q be a q-exponential family.

Set $\eta_i = E_{q,p}[F_i(x)]$. Then $\{\eta_i\}$ is a $\nabla^{q(m)}$ -affine coordinate system such that

$$g^q\left(rac{\partial}{\partial heta^i},rac{\partial}{\partial \eta_j}
ight)=\delta^j_i.$$

Set $\phi(\eta) = E_{q,p}[\log_q p(x; \theta)]$. Then $\phi(\eta)$ is the potential of g^q with respect to $\{\eta_i\}$.

We define the q-relative entropy (or the normalized Tsallis relative entropy) by

$$D_q^T(p(x), r(x)) = E_{q,p}[\log_q p(x) - \log_q r(x)] = \frac{1 - \int_{\Omega} p(x)^q r(x)^{1-q} dx}{(1-q)Z_q(p)}.$$

By taking the limit $q \to 1$, the Kullback-Leibler divergence is recovered. It is known that a *q*-relative entropy on a *q*-exponential family coincides with the canonical divergence on a dually flat space $(S_q, g^q, \nabla^{q(m)}, \nabla^{q(e)})$.

Let us generalize the maximum likelihood method for deformed exponential families. Suppose that two random variables X and Y follow probability distributions $p_1(x)$ and $p_2(y)$, respectively. We say that X and Y are *independent* if the joint probability distribution p(x,y) is decomposed by $p(x,y) = p_1(x)p_2(y)$. This formula is written by

$$p(x,y) = \exp[\log p_1(x) + \log p_2(y)].$$

Hence we can regard that the independence of random variables is attributed to the duality of the exponential and the logarithm. By virtue of this duality, we can generalize the notion of independence.

Suppose that x > 0, y > 0 and $x^{1-q} + y^{1-q} - 1 > 0$ (q > 0). The *q*-product of x and y is defined by

$$x \otimes_q y := [x^{1-q} + y^{1-q} - 1]^{\frac{1}{1-q}} = \exp_q [\log_q x + \log_q y].$$

Let $S_q = \{p(x; \theta) | \theta \in \Theta\}$ be a *q*-exponential family. Suppose that $\{x_1, \ldots, x_N\}$ are *N*-observations from $p(x; \theta) \in S_q$. We define a *q*-likelihood function $L_q(\theta)$ by

$$L_q(\theta) = p(x_1; \theta) \otimes_q p(x_2; \theta) \otimes_q \cdots \otimes_q p(x_N; \theta).$$

The parameter which maximizes the q-likelihood function is called the q-maximum likelihood estimator, that is,

$$\theta = \underset{\theta \in \Theta}{\operatorname{argmax}} L(\theta).$$

Theorem 11 (cf. [9]) Let $S_q = \{p(x; \theta) | \theta \in \Theta\}$ be a q-exponential family. Suppose that $\{x_1, \ldots, x_N\}$ are N-observations from $p(x; \theta) \in S_q$. Then the q-likelihood attains the maximum if and only if the normalized Tsallis relative entropy attains the minimum. In particular, the maximum q-likelihood estimator for η -coordinates is given by

$$\hat{\eta}_i = \frac{1}{N} \sum_{i=1}^N F_i(x_j).$$

CONCLUSIONS

In this paper, we summarized geometry of Bayesian statistics. The prior distribution is regarded as a volume element on a statistical manifold, and the Tchebychev vector field plays an important role. We also studied geometry of deformed exponential families. A suitable weight on a parameter space is considered in Bayesian statistics, whereas a suitable weight on a sample space is considered in anomalous statistics.

The author would like to express his gratitude to the reviewers for the comments for improvements of this paper. This research is partially supported by MEXT KAKENHI no. 23740047 and no. 26108003.

REFERENCES

- 1. S. Amari and H. Nagaoka, *Method of Information Geometry*, Amer. Math. Soc., Oxford University Press, 2000
- 2. S. Amari, A. Ohara and H. Matsuzoe, Geometry of deformed exponential families: invariant, duallyflat and conformal geometry, Physica A., **391**(2012), 4308–4319
- 3. S. Ikeda, T. Tanaka and S. Amari, *Information geometry of turbo and low-density parity-check codes*, IEEE Trans. Inform. Theory, **50**(2004), 1097–1114.
- 4. F. Komaki, On asymptotic properties of predictive distributions, Biometrica, 83(1996), 299–313.
- S. L. Lauritzen, *Statistical manifolds*, Differential Geometry in Statistical Inferences, in IMS Lecture Notes Monograph Series, vol. 10, Hayward California, 1987, pp. 96–163.
- 6. H. Matsuzoe, J. Takeuchi and S. Amari, *Equiaffine structures on statistical manifolds and Bayesian statistics*, Diff. Geom. Appl., **24**(2006), 567–578.
- 7. H. Matsuzoe, *Statistical manifolds and geometry of estimating functions*, Recent Progress in Differential Geometry and Its Related Fields, World Scientific, 2013, pp. 187–202.
- 8. H. Matsuzoe, *Hessian structures on deformed exponential families and their conformal structures*, Diff. Geom. Appl., **35**, Supplement, (2014), 323–333.
- H. Matsuzoe and M. Henmi, *Hessian structures and divergence functions on deformed exponential families*, Geometric Theory of Information, Signals and Communication Technology, Springer, 2014, pp. 57–80.
- 10. N. Murata, T. Takenouchi, T. Kanamori and S. Eguchi, *Information geometry of U-Boost and Bregman divergence*, Neural Computation, **16**, 7(2004), 1437-1481.
- 11. J. Naudts, Generalised Thermostatistics, Springer, 2011.
- 12. K. Nomizu. and T. Sasaki, Affine differential geometry Geometry of Affine Immersions –, Cambridge University Press, 1994.
- 13. U. Simon, A. Schwenk-Schellschmidt and H. Viesel, *Introduction to the affine differential geometry* of hypersurfaces, Lecture notes of the Science University of Tokyo, 1991.
- 14. J. Takeuchi and S. Amari, *α-parallel prior and its properties*, IEEE Trans. Inform. Theory, **51**(2005), 1011–1023.