Entropy **2015**, *17*, 5729-5751; doi:10.3390/e17085729

OPEN ACCESS
entropy
ISSN 1099-4300
www.mdpi.com/journal/entropy

Article

Deformed Algebras and Generalizations of Independence on Deformed Exponential Families

Hiroshi Matsuzoe [1,*] and Tatsuaki Wada [2]

[1] Department of Computer Science and Engineering, Nagoya Institute of Technology, Gokiso-cho, Showa-ku, Nagoya, 466-8555, Japan
[2] Department of Electrical and Electronic Engineering, Ibaraki University, 4-12-1 Nakanarusawa, Hitachi, Ibaraki, 316-8511, Japan; E-Mail: wada@mx.ibaraki.ac.jp

* Author to whom correspondence should be addressed; E-Mail: matsuzoe@nitech.ac.jp; Tel.: +81-52-735-5143.

External Editor: Giorgio Kaniadakis

Received: 1 February 2015 / Accepted: 4 August 2015 / Published: 10 August 2015

Abstract: A deformed exponential family is a generalization of exponential families. Since the useful classes of power law tailed distributions are described by the deformed exponential families, they are important objects in the theory of complex systems. Though the deformed exponential families are defined by deformed exponential functions, these functions do not satisfy the law of exponents in general. The deformed algebras have been introduced based on the deformed exponential functions. In this paper, after summarizing such deformed algebraic structures, it is clarified how deformed algebras work on deformed exponential families. In fact, deformed algebras cause generalization of expectations. The three kinds of expectations for random variables are introduced in this paper, and it is discussed why these generalized expectations are natural from the viewpoint of information geometry. In addition, deformed algebras cause generalization of independences. Whereas it is difficult to check the well-definedness of deformed independence in general, the κ-independence is always well-defined on κ-exponential families. This is one of advantages of κ-exponential families in complex systems. Consequently, we can well generalize the maximum likelihood method for the κ-exponential family from the viewpoint of information geometry.

Keywords: deformed algebra; deformed exponential family; expectation functional; information geometry; statistical manifold; generalized maximum likelihood method

MSC Classifications: 82B03; 94A17; 53A15; 62F99

1. Introduction

An exponential family is a set of probability distributions and an important statistical model in mathematical sciences. For example, the set of all Gaussian distributions is an exponential family. A deformed exponential family is one of generalizations of exponential families, and it has been studied in anomalous statistical physics (*cf.* [1]) and in machine learning theory (*cf.* [2,3]). A useful class of power law tailed distributions, such as the set of all Student's t-distributions, is a deformed exponential family.

In the study of deformed exponential families, a deformed exponential function and a deformed logarithm function play important roles. However, these functions do not satisfy the law of exponents in general. Hence, deformed algebraic structures and deformed differentials have been introduced in anomalous statistical physics (*cf.* [4–7]). In addition, a random variable that follows a power law tailed distribution may not have its mean and variance. To overcome this problem, a deformed probability distribution called an escort distribution (*cf.* [1,8]) has been introduced. Then, an expectation with respect to the escort distribution has been discussed.

In this paper, after summarizing such deformed algebraic structures, we clarify how a deformed algebra works on deformed exponential families. In particular, we elucidate that a deformed sum works on the sample space and a deformed product works on the target functional space. This difference makes clear how to use deformed algebras.

Since the deformed sum works on the sample space (*i.e.*, the domain of random variables), the sample space can be regarded as some algebraic space, not the standard Euclidean space. This deformation causes generalizations of expectations of random variables. In this paper, we consider three kinds of expectations, which include the expectation with respect to the escort distribution mentioned above. Then, we elucidate why these expectations are natural from the viewpoint of information geometry. Here, information geometry is one of the differential geometric methods for statistical estimation (*cf.* [9]). As a consequence, generalized expectations give local coordinate systems of deformed exponential families, and such coordinate systems have close relations to a dually-flat structure and to a projective structure of deformed exponential families. (see also [10–13], *etc.*)

The deformed product works on the target space of probability distributions. This deformation causes generalizations of independences. Though it is difficult to check the well-definedness of deformed independence, the κ-independence for the κ-exponential family is always well defined. This is an advantage of κ-exponential families among the class of deformed exponential families. Hence, we consider κ-generalization of the maximum likelihood method. In information geometry, it is known that the maximum likelihood estimator for a curved exponential family is characterized by the Kullback–Leibler divergence projection from the observed data. Based on this fact, we give a κ-generalization of the divergence projection-type theorem for the κ-maximum likelihood estimator.

In this paper, new contributions are stated as theorems (*i.e.*, Theorems 1, 3, 4 and 7), whereas known results are stated as propositions.

2. Deformed Exponential Families

In this section, we give definitions of deformed exponential functions and deformed exponential families. For more details, see [1,10,11,14], for example. We assume that all functions are real functions and that variables are defined in a real number field, since we will consider probability distributions in a real number field.

Let χ be a strictly increasing function from $\mathbf{R}_{>0}$ to $\mathbf{R}_{>0}$. We define a χ-logarithm function (or a deformed exponential function) by:

$$\ln_\chi s \;\; := \;\; \int_1^s \frac{1}{\chi(t)}\, dt.$$

The inverse of the χ-logarithm function is called a χ-exponential function (or a deformed exponential function), and it is given by:

$$\exp_\chi t \;\; := \;\; 1 + \int_0^t \lambda(s) dx,$$

where the function $\lambda(s)$ is given by $\lambda(\ln_\chi s) = \chi(s)$.

We remark that the χ-logarithm function $\ln_\chi$ and the χ-exponential function $\exp_\chi$ are usually called ϕ-logarithm and ϕ-exponential, respectively (*cf.* [1,15]). However, the symbol ϕ is used as the dual Hessian potential function in information geometry, so we use χ as the deformation function in this paper.

Example 1. *Suppose that a deformation function $\chi(s)$ is given by:*

$$\chi(s) = \frac{2s}{s^\kappa + s^{-\kappa}}, \qquad (-1 < \kappa < 1,\ \kappa \neq 0).$$

Then, the deformed exponential and the deformed logarithm are given by:

$$\begin{aligned} \ln_\kappa s \;\; &:= \;\; \frac{s^\kappa - s^{-\kappa}}{2\kappa}, \qquad (s > 0), \\ \exp_\kappa t \;\; &:= \;\; (\kappa t + \sqrt{1 + \kappa^2 t^2})^{\frac{1}{\kappa}}, \end{aligned}$$

respectively. The function $\ln_\kappa s$ is called a κ-logarithm function and $\exp_\kappa t$ a κ-exponential function (cf. *[6]). By taking a limit $\kappa \to 0$, these functions coincide with the standard logarithm and the standard exponential, respectively.*

While $s > 0$ is needed for defining the κ-logarithm function $\ln_\kappa s$, the κ-exponential function $\exp_\kappa t$ is defined entirely on $\mathbf{R}$, since $\kappa t + \sqrt{1 + \kappa^2 t^2}$ is always positive.

Example 2. *Suppose that $\chi(s)$ is given by a power function $\chi(s) = s^q$, $(q > 0, q \neq 1)$, Then, the deformed exponential and the deformed logarithm are given by:*

$$\begin{aligned} &\ln_q s := \frac{s^{1-q} - 1}{1 - q}, && (s > 0), \\ &\exp_q t := (1 + (1 - q)t)^{\frac{1}{1-q}}, && (1 + (1 - q)t > 0). \end{aligned}$$

The function $\ln_q s$ *is called a* q*-logarithm, and* $\exp_q t$ *a* q*-exponential* (cf. *[1,8]). Taking a limit* $q \to 1$*, these functions coincide with the standard logarithm and the standard exponential, respectively.*

The condition $s > 0$ *is needed for defining* $\ln_q s$*. In the* q*-exponential case, the condition:*

$$1 + (1 - q)t > 0 \tag{1}$$

is also necessary, since the base of the exponential function must be positive. Condition (1) is called the anti-exponential condition for the q*-exponential function.*

Let Ω be a total sample space. We say that a statistical model S_χ on Ω is a χ-exponential family or a deformed exponential family if S_χ is a set of probability density functions, such that:

$$S_\chi := \left\{ p(x;\theta) \,\middle|\, p(x;\theta) = \exp_\chi \left[\sum_{i=1}^n \theta^i F_i(x) - \psi(\theta) \right],\ \theta \in \Theta \subset \mathbf{R}^n \right\},$$

where $F_1(x), \ldots, F_n(x)$ are functions on Ω, $\theta = \{\theta^1, \ldots, \theta^n\}$ is a parameter and $\psi(\theta)$ is the normalization with respect to the parameter θ. We assume that S_χ is a statistical model in the sense of information geometry. That is, a probability density $p(x;\theta) \in S_\chi$ has support entirely on Ω. See Chapter 2 in [9] for more details. The normalization function ψ is convex, but it may not be strictly convex in general. We assume that ψ is strictly convex in this paper, and then, we can induce a Riemannian metric from this normalization function ψ (see Section 7). In addition, functions $F_1(x), \ldots, F_n(x), \psi(\theta)$ and a parameter θ must satisfy the anti-exponential condition. For example, in the q-exponential case,

$$\sum_{i=1}^n \theta^i F_i(x) - \psi(\theta) < -\frac{1}{1-q}.$$

We remark that it is a bit of a difficult problem how the anti-exponential condition imposes the domain of $\{\theta^i\}$ and the range of $\{F_i(x)\}$. We will give a further discussion at the end of this section.

We say that a deformed exponential family is a κ-exponential family if its deformed exponential function is a κ-exponential function $\exp_\kappa$ and a q-exponential family if its deformed exponential function is a q-exponential function $\exp_q$. These deformed exponential families are denoted by S_κ and S_q, respectively.

Suppose that M_χ is a submanifold of S_χ, that is,

$$M_\chi := \left\{ p(x;\theta(u)) \,\middle|\, p(x;\theta(u)) = \exp_\chi \left[\sum_{i=1}^n \theta^i(u) F_i(x) - \psi(\theta(u)) \right],\ u \in U \subset \mathbf{R}^m \subset \mathbf{R}^n \right\}.$$

The submanifold M_χ is called a curved χ-exponential family of S_χ. From similar arguments, we can define a curved q-exponential family M_q in S_q and a curved κ-exponential family M_κ in S_κ.

Example 3 (Discrete distributions (*cf.* [10]))**.** *Suppose that* $\Omega = \{x_0, x_1, \ldots, x_n\}$ *is a finite sample space. Denote by* S_n *the set of all probability distributions on* Ω*:*

$$S_n = \left\{ p(x;\eta) \,\middle|\, \eta_i > 0,\ \sum_{i=0}^n \eta_i = 1,\ p(x;\eta) = \sum_{i=0}^n \eta_i \delta_i(x) \right\}.$$

The natural parameters and the normalization are given by:

$$\begin{aligned}
\theta^i &= \ln_\chi p(x_i) - \ln_\chi p(x_0) = \ln_\chi \eta_i - \ln_\chi \left(1 - \sum_{i=1}^{n} \eta_i\right), \\
\psi(\theta) &= -\ln_\chi \eta_0 = -\ln_\chi \left(1 - \sum_{i=1}^{n} \eta_i\right).
\end{aligned}$$

Then, we obtain:

$$\begin{aligned}
\ln_\chi p(x;\theta) &= \ln_\chi \left(\sum_{i=0}^{n} \eta_i \delta_i(x)\right) \\
&= \sum_{i=1}^{n} (\ln_\chi \eta_i - \ln_\chi \eta_0)\, \delta_i(x) + \ln_\chi \eta_0 \\
&= \sum_{i=1}^{n} \theta^i \delta_i(x) - \psi(\theta).
\end{aligned}$$

This implies that S_n is a χ-exponential family for any χ.

Example 4 (Student t-distributions)**.** *Fix a parameter q $(1 < q < 3)$. A probability density functions $p(x;\mu,\sigma)$ on $\Omega = \mathbf{R}$ is said to be a Student t-distribution or a q-normal distribution if:*

$$p(x;\mu,\sigma) := \frac{1}{Z_q(\sigma)} \left[1 - \frac{1-q}{3-q}\frac{(x-\mu)^2}{\sigma^2}\right]^{\frac{1}{1-q}},$$

where (μ,σ) are parameters, such that $-\infty < \mu < \infty$ and $0 < \sigma < \infty$, and $Z_q(\sigma)$ is the normalization of probability density defined by:

$$Z_q(\sigma) := \frac{\sqrt{3-q}}{\sqrt{q-1}} \mathrm{Beta}\left(\frac{3-q}{2(q-1)}, \frac{1}{2}\right)\sigma.$$

By taking a limit $q \to 1$, a Student t-distribution converges to a normal distribution.

The set of all Student t-distributions S_q is a q-exponential family. In fact, natural parameters are given by:

$$\theta^1 := \frac{2}{3-q}\{Z_q(\sigma)\}^{q-1}\frac{\mu}{\sigma^2}, \quad \theta^2 := -\frac{1}{3-q}\{Z_q(\sigma)\}^{q-1}\frac{1}{\sigma^2},$$

respectively. Then, we obtain:

$$\begin{aligned}
\ln_q p(x) &= \frac{1}{1-q}(\{p(x)\}^{1-q} - 1) \\
&= \frac{1}{1-q}\left\{\frac{1}{\{Z_q(\sigma)\}^{1-q}}\left(1 - \frac{1-q}{3-q}\frac{(x-\mu)^2}{\sigma^2}\right) - 1\right\} \\
&= \frac{2\mu\{Z_q(\sigma)\}^{q-1}}{(3-q)\sigma^2}x - \frac{\{Z_q(\sigma)\}^{q-1}}{(3-q)\sigma^2}x^2 - \frac{\{Z_q(\sigma)\}^{q-1}}{3-q}\frac{\mu^2}{\sigma^2} + \frac{\{Z_q(\sigma)\}^{q-1} - 1}{1-q} \\
&= \theta^1 x + \theta^2 x^2 - \psi(\theta),
\end{aligned}$$

where ψ is the normalization defined by:

$$\psi(\theta) := -\frac{(\theta^1)^2}{4\theta^2} - \frac{\{Z_q(\sigma)\}^{q-1} - 1}{1-q}.$$

Hence, the set of all Student t-distributions S_q is a q-exponential family.

Let us give further considerations about deformed exponential families. In the case $0 < q < 1$, a q-normal distribution has the following form:

$$p(x;\mu,\sigma) \;:=\; \frac{1}{Z_q(\sigma)}\left[1-\frac{1-q}{3-q}\frac{(x-\mu)^2}{\sigma^2}\right]^{\frac{1}{1-q}} \;=\; \frac{1}{Z_q(\sigma)}\exp_q\left[-\frac{(x-\mu)^2}{(3-q)\sigma^2}\right],$$

where the normalization $Z_q(\sigma)$ is given by:

$$Z_q(\sigma) \;:=\; \frac{\sqrt{3-q}}{\sqrt{1-q}}\mathrm{Beta}\left(\frac{2-q}{1-q},\frac{1}{2}\right)\sigma.$$

The anti-exponential condition for this q-normal distribution is:

$$1-\frac{1-q}{3-q}\frac{(x-\mu)^2}{\sigma^2} > 0,$$

hence the domain of random variable x is given by:

$$\mu-\frac{\sqrt{3-q}}{\sqrt{1-q}}\sigma \;<\; x \;<\; \mu+\frac{\sqrt{3-q}}{\sqrt{1-q}}\sigma. \tag{2}$$

In this case, the set of q-normal distributions $S_q = \{p(x;\mu,\sigma)\}$ is not a statistical model in the sense of information geometry [9], since the support of $p(x;\mu,\sigma)$ depends on its parameter (μ,σ).

On the other hand, for a q-normal distribution, fix parameters q,μ,σ. By introducing a new parameter α $(0 < \alpha < q/(1-q))$, we set:

$$q_\alpha = \frac{q-\alpha(1-q)}{1-\alpha(1-q)}, \quad \sigma_\alpha^2 = \frac{3-q}{3-q-2\alpha(1-q)}\sigma^2. \tag{3}$$

The transformation $(q,\sigma) \mapsto (q_\alpha,\sigma_\alpha)$ defined by (3) is called a τ-transformation [16]. From straightforward calculations, we have:

$$\frac{3-q_\alpha}{1-q_\alpha}\sigma_\alpha^2 \;=\; \frac{3-q}{1-q}\sigma^2.$$

This equation implies that, from Equation (2), the domain of random variable x is invariant under τ-transformations. Hence, a one-dimensional statistical model is defined by:

$$S_{q_\alpha} = \left\{p(x;\alpha)\;\middle|\;p(x;\alpha) := \frac{1}{Z_{q_\alpha}(\sigma_\alpha)}\exp_{q_\alpha}\left[-\frac{(x-\mu)^2}{(3-q_\alpha)\sigma_\alpha^2}\right],\; 0<\alpha<\frac{q}{1-q}\right\}.$$

However, S_{q_α} is not a deformed exponential family in our setting, since the exponent q_α of the deformed exponential function depends on the parameter α.

3. Non-Additive Differentials

In this section, we consider deformed algebras and deformed differential equations to characterize deformed exponential functions.

3.1. κ-Deformed Algebras and κ-Exponential Functions

We begin with the κ-exponential case. For more details about κ-deformed algebras, see [6].

Let $\exp_\kappa$ be a κ-exponential function and $\ln_\kappa$ a κ-logarithm function. Since $\exp_\kappa$ and $\ln_\kappa$ do not satisfy the law of exponents, we introduce the κ-sum $\tilde{\oplus}^\kappa$ and the κ-product $\otimes_\kappa$ as follows.

$$\begin{aligned} x_1 \tilde{\oplus}^\kappa x_2 &:= \ln_\kappa \left[\exp_\kappa x_1 \cdot \exp_\kappa x_2\right] \\ &= x_1\sqrt{1+\kappa^2 x_2^2} + x_2\sqrt{1+\kappa^2 x_1^2}, \\ y_1 \otimes_\kappa y_2 &:= \exp_\kappa \left[\ln_\kappa y_1 + \ln_\kappa y_2\right], \quad (y_1 > 0 \text{ and } y_2 > 0). \end{aligned}$$

The conditions $y_1 > 0$ and $y_2 > 0$ are necessary for defining the κ-logarithm function. On the other hand, such conditions are not necessary for defining the κ-exponential function.

From the definitions of κ-deformed algebras, we have the following deformed law of exponents.

$$\begin{aligned} &\exp_\kappa(x_1 \tilde{\oplus}^\kappa x_2) = \exp_\kappa x_1 \cdot \exp_\kappa x_2, && \ln_\kappa(y_1 \cdot y_2) = \ln_\kappa y_1 \tilde{\oplus}^\kappa \ln_\kappa y_2, \\ &\exp_\kappa(x_1 + x_2) = \exp_\kappa x_1 \otimes_\kappa \exp_\kappa x_2, && \ln_\kappa(y_1 \otimes_\kappa y_2) = \ln_\kappa y_1 + \ln_\kappa y_2. \end{aligned} \tag{4}$$

Since the inverse element of x with respect to the κ-sum is $-x$, we define the κ-difference $\tilde{\ominus}^\kappa$ by:

$$\begin{aligned} x_1 \tilde{\ominus}^\kappa x_2 &:= x_1 \tilde{\oplus}^\kappa (-x_2) \\ &= x_1\sqrt{1+\kappa^2 x_2^2} - x_2\sqrt{1+\kappa^2 x_1^2}. \end{aligned} \tag{5}$$

By taking a limit with respect to the κ-difference, we define a (non-additive) κ-differential as follows.

$$\frac{d_\kappa}{d_\kappa x} f(x) := \lim_{x' \to x} \frac{f(x') - f(x)}{x' \tilde{\ominus}^\kappa x}. \tag{6}$$

We remark that a non-additive κ-differential $d_\kappa / d_\kappa x$ characterizes the κ-exponential function. Consider the following deformed differential equations:

$$\frac{d_\kappa}{d_\kappa x} f(x) = f(x), \tag{7}$$

$$\frac{d}{dx} f(x) = \frac{1}{\sqrt{1+\kappa^2 x^2}} f(x). \tag{8}$$

Then, the eigenfunction $f(x)$ of both equations is the κ-exponential function. That is,

$$f(x) = \exp_\kappa x = (\kappa x + \sqrt{1+\kappa^2 x^2})^{\frac{1}{\kappa}}.$$

In fact, from the definition of the κ-difference (5), we have:

$$\frac{d_\kappa}{d_\kappa x} = \sqrt{1+\kappa^2 x^2}\frac{d}{dx},$$

hence two deformed differential equations, (7) and (8), are essentially equivalent. We call a non-additive differential equation (7) a non-additive representation and a deformed differential equation (8) an escort representation.

Remark 1. *A κ-sum works on the domain of a κ-exponential function (*i.e.*, the sample space Ω), and a κ-product works on the target space. This implies that the sample space can be regarded as some deformed algebraic space, not the standard Euclidean space. In fact, the sample space and the target space are regarded as commutative fields (equivalently, Abelian fields in the usage of [7]). The κ-sum is an additive group structure of a commutative field structure on the sample space, and the κ-product is a multiplicative group structure on the target space (see also Remark 2 and [7]). We consider that this fact is very important in the theory of non-extensive statistical physics.*

Recall the definition of Napier's constant. The standard exponential function has the following infinite product expression:

$$\exp x = \lim_{n\to\infty}\left(1+\frac{x}{n}\right)^n .$$

In the κ-exponential case, we have the following.

Theorem 1. *Fix a real number $x \in \mathbf{R}$. Suppose that $n > |x|$ and $n \in \mathbf{N}$. Then, we have:*

$$\exp_\kappa x \;=\; \lim_{n\to\infty}\left(1+\frac{x}{n}\right)^{\otimes_\kappa n},$$

where:

$$\left(1+\frac{x}{n}\right)^{\otimes_\kappa n} := \underbrace{\left(1+\frac{x}{n}\right)\otimes_\kappa\cdots\otimes_\kappa\left(1+\frac{x}{n}\right)}_{n\ \textit{times}}.$$

Proof. From the assumption, the inequality $1 + x/n > 0$ always holds. Hence, we have:

$$\begin{aligned}\ln_\kappa\left(1+\frac{x}{n}\right)^{\otimes_\kappa n} &= n\,\ln_\kappa\left(1+\frac{x}{n}\right)\\ &= n\,\frac{\left(1+\frac{x}{n}\right)^\kappa-\left(1+\frac{x}{n}\right)^\kappa}{2\kappa}.\end{aligned} \tag{9}$$

(the assumption $n > |x|$ is a condition that the κ-logarithm in Equation (9) defines). From the definition of the κ-product (4), using asymptotic expansions, we have:

$$\begin{aligned}\left(1+\frac{x}{n}\right)^\kappa &= 1\,+\,\kappa\frac{x}{n}\,+\,O\left(\left(\frac{x}{n}\right)^2\right),\\ \left(1+\frac{x}{n}\right)^{-\kappa} &= 1\,-\,\kappa\frac{x}{n}\,+\,O\left(\left(\frac{x}{n}\right)^2\right).\end{aligned}$$

Substituting asymptotic expansions into (9), we have:

$$\ln_\kappa\left(1+\frac{x}{n}\right)^{\otimes_\kappa n} \;=\; x\,+\,\frac{n}{2\kappa}\cdot O\left(\left(\frac{x}{n}\right)^2\right).$$

Hence, we have:

$$\left(1+\frac{x}{n}\right)^{\otimes_\kappa n} \;=\; \exp_\kappa\left[x+\frac{n}{2\kappa}\cdot O\left(\left(\frac{x}{n}\right)^2\right)\right].$$

By taking a limit $n \to \infty$, we obtain the result. $\square$

3.2. q-Deformed Algebras and q-Exponential Functions

Let us consider the q-exponential case (*cf.* [4]). Let $\exp_q$ be a q-exponential function, and let $\ln_q$ be a q-logarithm function. The q-deformed algebras, *i.e.*, the q-sum $\tilde{\oplus}^q$ and the q-product $\otimes_q$, are defined as follows.

$$\begin{aligned} x_1 \tilde{\oplus}^q x_2 &:= \ln_q \left[\exp_q x_1 \cdot \exp_q x_2\right] \\ &= x_1 + x_2 + (1-q)x_1 x_2, \\ y_1 \otimes_q y_2 &:= \exp_q \left[\ln_q y_1 + \ln_q y_2\right] \\ &= \left[y_1^{1-q} + y_2^{1-q} - 1\right]^{\frac{1}{1-q}}, \end{aligned}$$

where conditions $1 + (1-q)x_1 > 0,\ 1 + (1-q)x_2 > 0,\ y_1^{1-q} + y_2^{1-q} - 1 > 0$ are needed for defining q-exponential functions and $y_1 > 0,\ y_2 > 0$ are for q-logarithm. Under the q-deformed algebras, the q-deformed law of exponents holds:

$$\begin{aligned} &\exp_q(x_1 \tilde{\oplus}^q x_2) = \exp_q x_1 \cdot \exp_q x_2, \qquad && \ln_q(y_1 \cdot y_2) = \ln_q y_1 \tilde{\oplus}^q \ln_q y_2, \\ &\exp_q(x_1 + x_2) = \exp_q x_1 \otimes_q \exp_q x_2, && \ln_q(y_1 \otimes_q y_2) = \ln_q y_1 + \ln_q y_2. \end{aligned} \tag{10}$$

The inverse element of x with respect to the q-sum is given by:

$$[-x]_q := \ln_q\left(\frac{1}{\exp_q x}\right) = \frac{-x}{1+(1-q)x}.$$

Hence, the q-difference should be defined by:

$$\begin{aligned} x_1 \tilde{\ominus}^q x_2 &:= x_1 \tilde{\oplus}^q [-x_2]_q \\ &= x_1 - \frac{1+(1-q)x_1}{1+(1-q)x_2} x_2. \end{aligned} \tag{11}$$

By taking a limit with respect to the q-difference, we define a (non-additive) q-differential as follows.

$$\frac{d_q}{d_q x} f(x) := \lim_{x' \to x} \frac{f(x') - f(x)}{x' \tilde{\ominus}^q x}.$$

Let us consider the following deformed differential equations:

$$\frac{d_q}{d_q x} f(x) = f(x), \tag{12}$$

$$\frac{d}{dx} f(x) = \{f(x)\}^q. \tag{13}$$

Then, the eigenfunction $f(x)$ of both equations is the q-exponential function. That is,

$$f(x) = \exp_q x = (1 + (1-q)x)^{\frac{1}{1-q}}.$$

In the same way as the κ-exponential, we say that the non-additive differential equation (12) is the non-additive representation and the deformed differential equation (13) is the escort representation.

We remark again that a q-sum (a deformed sum) works on the domain of a q-exponential function and that a q-product (a deformed product) works on the target space. Hence, the sample space Ω may not be the standard Euclidean space.

An infinite product expression of the q-exponential function is given as follows.

Proposition 2 (*cf.* [17])**.** *For all integers $n \in \mathbf{N}$, suppose that:*

$$n\left(1+\frac{x}{n}\right)^{1-q}-(n-1)>0. \tag{14}$$

Then, we have:

$$\exp_q x \;=\; \lim_{n\to\infty}\left(1+\frac{x}{n}\right)^{\otimes_q n},$$

where:

$$\left(1+\frac{x}{n}\right)^{\otimes_1 n} := \underbrace{\left(1+\frac{x}{n}\right)\otimes_q\cdots\otimes_q\left(1+\frac{x}{n}\right)}_{n \text{ times}}.$$

Proof. From the definition of q-product (10) and the anti-exponential condition (14), we have:

$$\begin{aligned}\ln_q\left(1+\frac{x}{n}\right)^{\otimes_q n} &= n\,\ln_q\left(1+\frac{x}{n}\right)\\ &= n\,\frac{\left(1+\frac{x}{n}\right)^{1-q}-1}{1-q}.\end{aligned}$$

Using an asymptotic expansion:

$$\left(1+\frac{x}{n}\right)^{1-q} \;=\; 1\;+\;(1-q)\frac{x}{n}\;+\;O\left(\left(\frac{x}{n}\right)^2\right),$$

we have:

$$\ln_q\left(1+\frac{x}{n}\right)^{\otimes_q n} \;=\; x\;+\;\frac{n}{1-q}\cdot O\left(\left(\frac{x}{n}\right)^2\right).$$

Hence, we have:

$$\left(1+\frac{x}{n}\right)^{\otimes_q n} \;=\; \exp_q\left[x+\frac{n}{1-q}\cdot O\left(\left(\frac{x}{n}\right)^2\right)\right].$$

By taking a limit $n\to\infty$, we obtain the result. $\square$

Remark 2. *If a deformed sum and a deformed product are well-defined, then we can give similar arguments for any χ-exponential functions. However, it is difficult to describe the anti-exponential conditions in general. If we can admit a complex number field for the domain and the target of statistical model* (cf. *[18]), then the deformed algebras are well defined [7]. In fact, we can define the following commutative field structures if all of the objects are well defined:*

$$\begin{aligned} x_1\tilde{\oplus}^\chi x_2 &:= \ln_\chi\left[\exp_\chi x_1\cdot\exp_\chi x_2\right], & x_1\tilde{\otimes}^\chi x_2 &:= \ln_\chi\left[\exp[\ln(\exp_\chi x_1)\cdot\ln(\exp_\chi x_2)]\right],\\ y_1\oplus_\chi y_2 &:= \exp_\chi\left[\ln[\exp(\ln_\chi y_1)+\exp(\ln_\chi y_2)]\right], & y_1\otimes_\chi y_2 &:= \exp_\chi\left[\ln_\chi y_1+\ln_\chi y_2\right].\end{aligned}$$

Usages of the multiplicative group structure $\tilde{\otimes}^\chi$ on the sample space and the additive group structure $\oplus_\chi$ on the target space are not clear. We may need algebraic probability theory to clarify these group structures (for usages of algebraic structures for statistics, see [19,20], for example).

4. Expectation Functionals

As we have seen in the previous section, the sample space Ω may not be the standard Euclidean space. Let us consider suitable expectations for deformed exponential families.

For a χ-exponential probability $p(x;\theta) \in S_\chi$, we define the escort distribution $P_\chi(x;\theta)$ and the normalized escort distribution $P_\chi^{esc}(x;\theta)$ of $p(x;\theta)$ by:

$$\begin{aligned} P_\chi(x;\theta) &:= \chi\{p(x;\theta)\}, \\ P_\chi^{esc}(x;\theta) &:= \frac{1}{Z_\chi(\theta)}\chi\{p(x;\theta)\}, \qquad Z_\chi(\theta) := \int_\Omega \chi\{p(x;\theta)\}dx, \end{aligned}$$

respectively. The χ-canonical expectation $E_{\chi,p}[*]$ and the normalized χ-escort expectation $E_{\chi,p}^{esc}[*]$ are defined by:

$$\begin{aligned} E_{\chi,p}[f(x)] &:= \int_\Omega f(x)P_\chi(x;\theta)\,dx = \int_\Omega f(x)\chi\{p(x;\theta)\}dx, \\ E_{\chi,p}^{esc}[f(x)] &:= \int_\Omega f(x)P_\chi^{esc}(x;\theta)\,dx = \frac{1}{Z_\chi(\theta)}\int_\Omega f(x)\chi\{p(x;\theta)\}dx. \end{aligned}$$

Even though the integration of χ-canonical expectation is carried out with respect to a positive density, as we will see in later sections, this expectation is natural from the viewpoint of differential geometry. On the other hand, we call the standard expectation with respect to $p(x;\theta)$ a simple expectation and denote it by:

$$E_p[f(x)] := \int_\Omega f(x)p(x;\theta)dx.$$

A χ-canonical expectation and a normalized χ-escort expectation with respect to a κ-exponential probability $p(x;\theta)$ are called the κ-canonical expectation and the normalized κ-escort expectation and are denoted by $E_{\kappa,p}[*]$ and $E_{\kappa,p}^{esc}[*]$, respectively. In the q-exponential case, they are called the q-canonical expectation and the normalized q-escort expectation and denoted by $E_{q,p}[*]$ and $E_{q,p}^{esc}[*]$, respectively.

For a Student t-distribution $p(x;\mu,\theta) \in S_q$, the normalized q-escort mean μ_q and the normalized q-escort variance σ_q^2 are given by:

$$\begin{aligned} \mu_q &:= E_{q,p}^{esc}[x] = \mu, \\ \sigma_q^2 &:= E_{q,p}^{esc}\left[(x-\mu)^2\right] = \sigma^2, \end{aligned}$$

respectively. Hence, the normalized q-escort expectation $E_{q,p}^{esc}[*]$ is a natural generalization of the simple expectation $E_p[*]$.

Next, we consider non-additive integrals to elucidate the relations between the deformed algebras and the escort expectations. In particular, we discuss the κ-exponential case.

Let $f(x)$ be a function on the sample space Ω. Then, we define a (non-additive) κ-integral (*cf.* [5]) by the following formula:

$$\int_\Omega f(x)d_\kappa x := \int_\Omega \frac{f(x)}{\sqrt{1+\kappa^2x^2}}dx = \int_\Omega f(x)w_\kappa(x)dx, \tag{15}$$

where $w(x)$ is a weight function defined by:

$$w_\kappa(x) = \frac{1}{\sqrt{1+\kappa^2x^2}}.$$

Obviously, this is the inverse operation of the non-additive κ-differential (6).

When Ω is a discrete set, $\Omega = \{x_0.x_1, \ldots, x_n\}$, then we define a (non-additive) κ-summation by:

$$\sum_{i=0}^{n}{}_{\circledast} f(x_i) \ := \ \sum_{i=0}^{n} \frac{f(x_i)}{\sqrt{1+\kappa^2 x_i^2}} \ = \ \sum_{i=0}^{N} f(x_i) w(x_i).$$

From the definition of the κ-exponential function, we have the following.

Theorem 3. *Suppose that $\chi(s) = 2s/(s^{\kappa} + s^{-\kappa})$ is the deformation function with respect to the κ-logarithm function. Then, $\chi(\exp_\kappa x)$ coincides with the weight function $w(x)$ with respect to the non-additive κ-integral. That is, the following formula holds:*

$$\chi(\exp_\kappa x) = \frac{1}{\sqrt{1+\kappa^2 x^2}} = w_\kappa(x).$$

We think that the canonical expectation $E_{\kappa,p}[*]$ gives a suitable weight for the sample space Ω from the above theorem. We may consider a non-additive χ-integral as a general discussion (in the q-exponential case, the corresponding q-integral is introduced in [4]). However, we have to check carefully the well-definedness of the χ-integral since the anti-exponential condition must be satisfied.

5. Geometry of χ-Exponential Families with Simple Expectations

In this section, we consider the geometry of χ-exponential families by generalizing the e-representation and the m-representation of probability densities. For more details, see [11].

Let S_χ be a χ-exponential family. We define a χ-score function $s^{\chi}(x;\theta) : S_\chi \to \mathbf{R}^n$, $s^{\chi}(x;\theta) = {}^t\left((s^{\chi})^1(x;\theta),\ \ldots,\ (s^{\chi})^n(x;\theta)\right)$ by:

$$(s^{\chi})^i(x;\theta) \ := \ \frac{\partial}{\partial\theta^i} \ln_\chi p(x;\theta), \quad (i = 1, \ldots, n). \tag{16}$$

Under suitable conditions, we can define Riemannian metrics on S_χ by:

$$g_{ij}^E(\theta) \ := \ \int_\Omega \partial_i \ln_\chi p(x;\theta) \partial_j \ln_\chi p(x;\theta) \chi\{p(x;\theta)\}\, dx \tag{17}$$
$$= \ E_{\chi,p}[(s^{\chi})^i(x;\theta)(s^{\chi})^j(x;\theta)],$$
$$g_{ij}^M(\theta) \ := \ \int_\Omega \partial_i p(x;\theta) \partial_j \ln_\chi p(x;\theta)\, dx, \tag{18}$$
$$g_{ij}^N(\theta) \ := \ \int_\Omega \frac{1}{\chi\{p(x;\theta)\}} \partial_i p(x;\theta) \partial_j p(x;\theta)\, dx. \tag{19}$$

In the same manner as an invariant statistical manifold, a differential $\partial_i p(x;\theta)$ and a χ-score function $\partial_i \ln p(x;\theta)$ are regarded as tangent vectors for a χ-exponential family S_χ. Hence, the χ-score function is a generalization of the e-representation of $p(x;\theta)$.

Theorem 4. *Riemannian metrics g^E, g^M and g^N on S_χ coincide. That is,*

$$g^E(\theta) = g^M(\theta) = g^N(\theta).$$

Proof. For a χ-exponential distribution $p(x;\theta)$, its differential is given as follows:

$$\begin{aligned}\frac{\partial}{\partial\theta^i}p(x;\theta) &= \chi(p(x;\theta))\left(F_i(x) - \frac{\partial}{\partial\theta^i}\psi(\theta)\right),\\ \frac{\partial}{\partial\theta^i}\ln_\chi p(x;\theta) &= F_i(x) - \frac{\partial}{\partial\theta^i}\psi(\theta).\end{aligned}$$

By substituting the above formulas into (17)–(19), we obtain the results. □

We remark that integrations are carried out with respect to un-normalized χ-escort distributions. If we define Riemannian metrics by normalized χ-escort expectations, they do not coincide in general. Their Riemannian metrics are conformally equivalent (*cf.* [11]).

By differentiating Equation (18), we can define dual affine connections $\nabla^{M(e)}$ and $\nabla^{M(m)}$ on S_χ by:

$$\begin{aligned}\Gamma^{M(e)}_{ij,k}(\theta) &:= \int_\Omega \partial_k p(x;\theta)\partial_i\partial_j \ln_\chi p(x;\theta)dx,\\ \Gamma^{M(m)}_{ij,k}(\theta) &:= \int_\Omega \partial_i\partial_j p(x;\theta)\partial_k \ln_\chi p(x;\theta)dx.\end{aligned}$$

From the definitions of the χ-exponential family and the χ-logarithm function, we obtain $\Gamma^{M(e)}_{ij,k}(\theta) \equiv 0$. Hence, a parameter $\theta = \{\theta^i\}$ is a $\nabla^{M(e)}$-affine coordinate system, and the connection $\nabla^{M(e)}$ is flat. These imply that the triplet $(S_\chi, \nabla^{M(e)}, g^M)$ is a Hessian manifold. The cubic form C^M_{ijk} of $(S_\chi, \nabla^{M(e)}, g^M)$ is:

$$C^M_{ijk} = \Gamma^{M(m)}_{ij,k} - \Gamma^{M(e)}_{ij,k} = \Gamma^{M(m)}_{ij,k}.$$

To give Hessian potential functions of $(S_\chi, \nabla^{M(e)}, g^M)$, we define functions I_χ and Φ by:

$$\begin{aligned}I_\chi(p_\theta) &:= -\int_\Omega \{V_\chi(p(x;\theta)) + (p(x;\theta)-1)V_\chi(0)\}\,dx,\\ \Psi(\theta) &:= \int_\Omega p(x;\theta)\ln_\chi p(x;\theta)dx + I_\chi(p_\theta) + \psi(\theta),\end{aligned}$$

where the function $V_\chi(t)$ is given by:

$$V_\chi(t) := \int_1^t \ln_\chi(s)\,ds.$$

We call I_χ a generalized entropy functional and Ψ a generalized Massieu potential.

Proposition 5 (*cf.* [21,22])**.** *For a χ-exponential family S_χ, (1) the generalized Massieu potential $\Psi(\theta)$ is the potentials of g^M and C^M with respect to $\{\theta^i\}$:*

$$\begin{aligned}g^M_{ij}(\theta) &= \partial_i\partial_j\Psi(\theta),\\ C^M_{ijk}(\theta) &= \partial_i\partial_j\partial_k\Psi(\theta).\end{aligned}$$

(2) Let η_i be the simple expectation of $F_i(x)$, i.e., *$\eta_i := E_p[F_i(x)]$. Then, $\{\eta_i\}$ is the dual affine coordinate system of $\{\theta^i\}$ with respect to g^M, and each η_i is given by:*

$$\eta_i = \partial_i\Psi(\theta).$$

(3) Let $\Phi(\eta)$ be the negative generalized entropy functional, i.e., *$\Phi(\eta) := -I_\chi(p_\theta)$. Then, $\Phi(\eta)$ is the potential of g^M with respect to $\{\eta_i\}$.*

Let us consider a divergence function on χ-exponential family. The canonical divergence D on $(S_\chi, \nabla^{M(e)}, g^M)$ is defined by:

$$D(p,r) = \Psi(\theta(p)) + \Phi(\eta(r)) - \sum_{i=1}^{n} \theta^i(p)\eta_i(r).$$

On the other hand, the χ-divergence (or U-divergence) on S_χ is defined by:

$$D_\chi(p,r) = \int_\Omega \{U_\chi(\ln_\chi r(x)) - U_\chi(\ln_\chi p(x)) - p(x)(\ln_\chi r(x) - \ln_\chi p(x))\}\, dx,$$

where the function $U_\chi(t)$ is given by:

$$U_\chi(s) := \int_0^s \exp_\chi(t)\, dt.$$

Then, the χ-divergence D_χ coincides with the canonical divergence D on $(S_\chi, \nabla^{M(m)}, g^M)$. We remark that the χ-divergence is naturally constructed from a bias corrected χ-score function. See [11,23]. for more details.

In the q-exponential case, the χ-divergence is given by:

$$D_{1-q}(p,r) = \frac{1}{(1-q)(2-q)} \int_\Omega p(x)^{2-q} dx - \frac{1}{1-q} \int_\Omega p(x) r(x)^{1-q} dx + \frac{1}{2-q} \int_\Omega r(x)^{2-q} dx.$$

The divergence $D_{1-q}(p,r)$ is called a β-divergence ($\beta = 1-q$) or a density power divergence in statistics [24]. This divergence is useful in robust statistics.

We remark that the generalization of e- and m-representations through an arbitrary monotone embedding function was first studied in [25]. For further generalizations through monotone embedding functions, see [26,27]. These generalizations of e- and m-representations are also related to the U-geometry in information geometry (*cf.* [21,22]). When the embedding function $\chi(t)$ is identity ($q = 1$ in the q-exponential case and $\kappa = 0$ in the κ-exponential case), the results in this section reduce to the standard results in exponential families [11].

6. Geometry of Deformed Exponential Families with χ-Escort Expectation

Since a χ-exponential distribution has a normalization term $\psi(\theta)$, we induce geometric structures directly from the potential function ψ. For more details, see [10,11]. When the embedding function $\chi(t)$ is identity, the results in this section also reduce to the standard results in exponential families [11].

We define a χ-Fisher metric g^χ and a χ-cubic form C^χ by:

$$\begin{aligned} g^\chi_{ij}(\theta) &:= \partial_i\partial_j\psi(\theta), \\ C^\chi_{ijk}(\theta) &:= \partial_i\partial_j\partial_k\psi(\theta), \end{aligned}$$

respectively. Denote by $\Gamma^{\chi(0)}_{ij,k}$ the Christoffel symbol of the Levi–Civita connection with respect to the χ-Fisher metric g^χ. From standard arguments in Hessian geometry [28], we can define mutually dual flat connections by:

$$\begin{aligned} \Gamma^{\chi(e)}_{ij,k}(\theta) &:= \Gamma^{\chi(0)}_{ij,k}(\theta) - \frac{1}{2}C^\chi_{ijk}(\theta) &&\equiv 0, \\ \Gamma^{\chi(m)}_{ij,k}(\theta) &:= \Gamma^{\chi(0)}_{ij,k}(\theta) + \frac{1}{2}C^\chi_{ijk}(\theta) &&= C^\chi_{ijk}(\theta), \end{aligned}$$

respectively. We call $\nabla^{\chi(e)}$ a χ-exponential connection and $\nabla^{\chi(m)}$ a χ-mixture connection. In this case, $\{\theta^i\}$ is a $\nabla^{\chi(e)}$-affine coordinate system, and triplets $(S_\chi, \nabla^{\chi(e)}, g^\chi)$ and $(S_\chi, \nabla^{\chi(m)}, g^\chi)$ are mutually dual Hessian manifolds.

Proposition 6 (*cf.* [10,11])**.** *For a χ-exponential family S_χ,*

(1) $\psi(\theta)$ is the potential of g^χ and C^χ with respect to $\{\theta^i\}$.

(2) Let η_i be the normalized χ-escort expectation of $F_i(x)$, i.e., *$\eta_i := E^{esc}_{\chi,p}[F_i(x)]$. Then, $\{\eta_i\}$ is the dual affine coordinate system of $\{\theta^i\}$ with respect to g^χ, and each η_i is given by:*

$$\eta_i = \partial_i \psi(\theta).$$

(3) Let $\phi(\eta)$ by the negative χ-deformed entropy, i.e., *$\phi(\eta) := E^{esc}_{\chi,p}[\ln_\chi p(x;\theta)]$.*

Then, $\phi(\eta)$ is the potential of g^χ with respect to $\{\eta_i\}$.

Let us consider divergence functions. The canonical divergence of $(S_\chi, \nabla^{\chi(e)}, g^\chi)$ is given by:

$$D(p,r) = \psi(\theta(p)) + \phi(\eta(r)) - \sum_{i=1}^n \theta^i(p)\eta_i(r).$$

On the other hand, a χ-relative entropy (or a generalized relative entropy) $D^\chi(p,r)$ on S_χ is defined by:

$$D^\chi(p,r) \;:=\; E^{esc}_{\chi,p}[\ln_\chi p(x) - \ln_\chi r(x)]. \tag{20}$$

If the deformation function χ is an identity function $\chi(s) = s$, then the χ-relative entropy coincides with the Kullback–Leibler divergence. In addition, the χ-relative entropy D^χ coincides with the canonical divergence on $(S_\chi, \nabla^{\chi(m)}, g^\chi)$. In fact, in the same way as a standard exponential family, we have:

$$\begin{aligned}
D^\chi(p(\theta), p(\theta')) &= E^{esc}_{\chi,p}\left[\left(\sum_{i=1}^n \theta^i F_i(x) - \psi(\theta)\right) - \left(\sum_{i=1}^n (\theta')^i F_i(x) - \psi(\theta')\right)\right] \\
&= \left(\sum_{i=1}^n \theta^i \eta_i(x) - \psi(\theta)\right) - \left(\sum_{i=1}^n (\theta')^i \eta_i(x) - \psi(\theta')\right) \\
&= \psi(\theta') + \phi(\theta) - \sum_{i=1}^n (\theta')^i \eta_i \\
&= D(p(\theta'), p(\theta)).
\end{aligned}$$

In the κ-exponential case, we call a χ-relative entropy (20) a κ-relative entropy and denote it by D^κ.

On the other hand, in the q-exponential case, a χ-relative entropy for q-exponential family is called a normalized Tsallis relative entropy, which is given by:

$$\begin{aligned}
D^T_q(p,r) &:= E^{esc}_{q,p}[\ln_q p(x) - \ln_q r(x)] \\
&= \int_\Omega P^{esc}_q(x)\,(\ln_q p(x) - \ln_q r(x))\,dx \\
&= \frac{1}{(1-q)Z_q(p)}\left\{1 - \int_\Omega p(x)^q r(x)^{1-q}\right\} dx,
\end{aligned}$$

where $Z_q(p)$ is the normalization of the escort distribution $P_q^{esc}(x)$ of $p(x)$. Denote by $(S_q, \nabla^{q(e)}, g^q)$ and $(S_q, \nabla^{q(m)}, g^q)$ the induced Hessian manifolds from the normalization $\psi(\theta)$. Then, the normalized Tsallis relative entropy coincides with the canonical divergence for a Hessian manifold $(S_q, \nabla^{q(m)}, g^q)$.

For a q-exponential family, we can also define an α-divergence $(\alpha = 1 - 2q)$ by:

$$\begin{aligned} D^{(1-2q)}(p, r) &:= \frac{1}{q} E_{q,p}[\ln_q p(x) - \ln_q r(x)] \\ &= \frac{1}{q} \int_\Omega P_q(x) \left(\ln_q p(x) - \ln_q r(x)\right) dx \\ &= \frac{1}{q(1-q)} \left\{ 1 - \int_\Omega p(x)^q r(x)^{1-q} \right\} dx. \end{aligned}$$

It is known that the α-divergence $(\alpha = 1 - 2q)$ induces an invariant statistical manifold $(S_q, \nabla^{(1-2q)}, g)$.

Remark 3. *For a q-exponential family S_q, a normalized Tsallis entropy induces a Hessian manifold (*i.e.*, a flat statistical manifold) $(S_q, \nabla^{q(m)}, g^q)$, whereas an α-divergence induces an invariant statistical manifold $(S_q, \nabla^{(1-2q)}, g)$. Since a constant multiplication is not essential in differential geometry, the difference is caused by the normalization of the escort distribution:*

$$D_q^T(p, r) = \frac{q}{Z_q(p)} D^{(1-2q)}(p, r).$$

In this case, the two statistical manifolds $(S_q, \nabla^{q(m)}, g^q)$ and $(S_q, \nabla^{(1-2q)}, g)$ are (-1)-conformally equivalent (cf. *[29,30]). This implies that the normalization of a probability density is not a trivial problem. The normalization does affect the induced geometric structures and, consequently, the estimating methods for statistical inference.*

7. Discussion about Expectations

We give further discussions about expectation functionals. Since a deformed exponential family S_χ is regarded as a manifold, we can choose an arbitrary local coordinate system for S_χ. From this point of view, simple expectations $\{E_p[F_i(x)]\}$ and normalized χ-escort expectations $\{E_{\chi,p}^{esc}[F_i(x)]\}$ are nothing but local coordinates of the statistical model. However, in differential geometry, we often use appropriate coordinates depending on the background geometry, e.g., Darboux coordinates in symplectic geometry and isothermal coordinates in geometry of minimal surfaces. From Propositions 5 and 6, the simple expectations $\{E_p[F_i(x)]\}$ and the normalized χ-escort expectations $\{E_{\chi,p}^{esc}[F_i(x)]\}$ give appropriate coordinates for $(S_\chi, g^M, \nabla^{M(e)}, \nabla^{M(m)})$ and $(S_\chi, g^\chi, \nabla^{\chi(e)}, \nabla^{\chi(m)})$, respectively, since they are the dual affine coordinates of the natural parameters $\{\theta^i\}$.

From the assumptions of deformed exponential families, there always exists a dually flat structure $(S_\chi, g^\chi, \nabla^{\chi(e)}, \nabla^{\chi(m)})$, but there does not exist $(S_\chi, g^M, \nabla^{M(e)}, \nabla^{M(m)})$ in general (see [31] for more details). In addition, from Theorem 3, the deformed algebra on sample space Ω is reflected in the canonical expectation $E_{\chi,p}[*]$. Hence, we think that the canonical expectation $E_{\chi,p}[*]$ and the normalized χ-escort expectation $E_{\chi,p}^{esc}[*]$ are more natural than the simple expectation $E_p[*]$.

8. Maximum κ-Likelihood Estimators

In Section 3, we discussed deformed algebras for deformed exponential functions. As a consequence, it is natural to regard that a sample space is not the standard Euclidean space. In this section, we construct a maximum likelihood method that is in accordance with the deformed algebras.

Suppose that X is a random variable that follows a probability $p_1(x)$, and Y follows $p_2(y)$. We say that two random variables X and Y are independent if the joint probability $p(x,y)$ coincides with the product of marginal distributions $p_1(x)$ and $p_2(y)$:

$$p(x,y) = p_1(x)p_2(y).$$

Suppose that $p_1(x)$ and $p_2(y)$ have support entirely on Ω, that is $p_1(x) > 0$ and $p_2(y) > 0$ hold for all $x \in \Omega$. The independence is given by a duality of an exponential function and a logarithm function:

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

We generalize the notion of independence using the χ-exponential and χ-logarithm.

Suppose that X_i is a random variable on Ω_i, which follows $p_i(x)$ $(i = 1, 2, \ldots, N)$. Random variables $X_1, X_2, \ldots, X_N$ may not be independent on the standard algebra. Let $p(x_1, x_2, \ldots, x_N)$ be the joint probability density of $X_1, X_2, \ldots, X_N$.

We say that $X_1, X_2, \ldots, X_N$ are χ-independent with m-normalization if:

$$p(x_1, x_2, \ldots, x_N) = \frac{p_1(x_1) \otimes_\chi p_2(x_2) \otimes_\chi \cdots \otimes_\chi p_N(x_N)}{Z_{p_1, p_2, \cdots, p_N}},$$

where $Z_{p_1,p_2,\cdots,p_N}$ is the normalization of $p_1(x_1) \otimes_\chi p_2(x_2) \otimes_\chi \cdots \otimes_\chi p_N(x_N)$ defined by:

$$Z_{p_1,p_2,\cdots,p_N} := \int \cdots \int_{Supp\{p(x_1,x_2,\ldots,x_N)\} \subset \Omega_1 \cdots \Omega_N} p_1(x_1) \otimes_\chi p_2(x_2) \otimes_\chi \cdots \otimes_\chi p_N(x_N) dx_1 \cdots dx_N.$$

We remark that the domain of integration may not be entirely $\Omega_1 \times \cdots \times \Omega_N$ because of the anti-exponential conditions. In addition, N is not an arbitrary integer. The maximum number of N depends on the deformation function χ.

Example 5 (Bivariate Student t-distributions (*cf.* [32]))**.** *Suppose that X and Y are random variables that follow Student t-distributions $p_q(x; \mu_x, \sigma_x)$ and $p_q(y; \mu_y, \sigma_y)$, respectively. Even if X and Y are independent, the joint distribution $p(x,y) = p_q(x)p_q(y)$ is not a bivariate Student t-distribution. On the other hand, if X and Y are q-independent with m-normalization, then the joint distribution:*

$$p_q(x,y) = \frac{p_q(x) \otimes_q p_q(y)}{Z_{p_q(x), p_q(y)}}$$

is a bivariate Student t-distribution. Note that neither $p_q(x)$ nor $p_q(y)$ is the marginal distribution, because:

$$\int_{\Omega_Y} p_q(x,y)dy \neq p_q(x).$$

However, in this paper, we say that $p_q(x)$ and $p_q(y)$ are the q-marginal distributions of the joint distribution $p_q(x,y)$.

Recall that we cannot consider infinitely many q-products to define a joint distribution. In the case of Student t-distributions, the number of q-marginal distributions must satisfy $N < 2(q-1)$. Otherwise, the normalization Z diverges.

Let us consider the κ-exponential case. We say that random variables $X_1, X_2, \ldots, X_N$ are κ-independent with m-normalization if:

$$p(x_1, x_2, \ldots, x_N) = \frac{p_1(x_1) \otimes_\kappa p_2(x_2) \otimes_\kappa \cdots \otimes_\kappa p_N(x_N)}{Z_{p_1, p_2, \cdots, p_N}},$$

where $Z_{p_1, p_2, \cdots, p_N}$ is the normalization of $p_1(x_1) \otimes_\kappa p_2(x_2) \otimes_\kappa \cdots \otimes_\kappa p_N(x_N)$ defined by:

$$Z_{p_1, p_2, \cdots, p_N} := \int \cdots \int_{\Omega_1 \cdots \Omega_N} p_1(x_1) \otimes_\kappa p_2(x_2) \otimes_\kappa \cdots \otimes_\kappa p_N(x_N) dx_1 \cdots dx_N.$$

In the κ-exponential case, the domain of integration is entirely $\Omega_1 \times \cdots \times \Omega_N$, since the κ-exponential function is defined entirely on $\mathbf{R}$.

Similarly, we say that $X_1, X_2, \ldots, X_N$ are κ-independent with e-normalization (or exponential normalization) if:

$$p(x_1, x_2, \ldots, x_N) = p_1(x_1) \otimes_\kappa p_2(x_2) \otimes_\kappa \cdots \otimes_\kappa p_N(x_N) \otimes_\kappa (-c),$$

where c is the normalization of $p_1(x_1) \otimes_\kappa p_2(x_2) \otimes_\kappa \cdots \otimes_\kappa p_N(x_N)$ defined by:

$$\int \cdots \int_{\Omega_1 \cdots \Omega_N} p_1(x_1) \otimes_\kappa p_2(x_2) \otimes_\kappa \cdots \otimes_\kappa p_N(x_N) \otimes_\kappa (-c) dx_1 \cdots dx_N = 1.$$

We remark that the e-normalization is different form the m-normalization in general. See [33] for further discussion.

A normalization of joint distribution is not required in several problems. In these cases, we define a joint positive distribution (not a probability distribution) by the κ-marginal probability distributions,

$$f(x_1, x_2, \ldots, x_N) := p_1(x_1) \otimes_\kappa p_2(x_2) \otimes_\kappa \cdots \otimes_\kappa p_N(x_N), \tag{21}$$

and we say that $X_1, X_2, \ldots, X_N$ are simply κ-independent.

Remark 4. *As we mentioned in Remark 2, it is difficult to describe explicitly the anti-exponential conditions for the χ-exponential case. Though several authors have introduced χ-independence (which is called U-independence in [2,3] and F-independence in [34]), they did not mention the anti-exponential conditions. Hence, the χ-independence was not well defined in their papers.*

On the other hand, the anti-exponential condition of the κ-deformed algebra (4) is always satisfied, since $p(x; \theta) \in S_\kappa$ can be defined entirely on $\mathbf{R}$. *Therefore, the κ-independence is well defined for a κ-exponential family. This is an advantage of the κ-exponential families.*

Before we discuss a generalization of maximum likelihood methods, we recall the difference between Gauss' law of error and the maximum likelihood method.

In the case of Gauss' law of error, we consider the following likelihood function:

$$L(\theta) := p(x_1 - \theta) p(x_2 - \theta) \cdots p(x_N - \theta).$$

Suppose that N-observations $\{x_1, \ldots, x_N\}$ are obtained. If the likelihood function $L(\theta)$ attains the maximum at the sample mean $\theta = \bar{x}_N = (x_1, \ldots, x_N)/N$, then the probability density function p must be a Gaussian distribution. Hence, we specify a probability distribution from a given likelihood function and observed data. Generalizations of Gauss's law of error in non-extensive statistical physics have been obtained in [35,36], *etc*.

On the other hand, in the case of the maximum likelihood method, we suppose a statistical model $S = \{p(x;\theta)\}$ and define a likelihood function $L(\theta)$ by:

$$L(\theta) := p(x_1;\theta)p(x_2;\theta)\cdots p(x_N;\theta).$$

Suppose that N-observations $\{x_1, \ldots, x_N\}$ are obtained. If the likelihood function attains the maximum at $\hat{\theta}$, then the probability distribution $p(x;\hat{\theta})$ is expected to be closest to the true distribution in the given statistical model. Hence, we specify a parameter on a given statistical model from a likelihood function and observed data.

Later in this section, we consider a κ-generalization of the maximum likelihood method and give a characterization of the maximum κ-likelihood estimator from the viewpoint of information geometry.

Let $S_\kappa = \{p(x;\theta) | \theta \in \Theta\}$ be a κ-exponential family, and let $\{x_1, \ldots, x_N\}$ be N-observations from $p(x;\theta) \in S_\kappa$. We define a κ-likelihood function $L_\kappa(\theta)$ and a κ-logarithm κ-likelihood function $l_\kappa(\theta)$ by:

$$\begin{aligned} L_\kappa(\theta) &:= p(x_1;\theta) \otimes_\kappa p(x_2;\theta) \otimes_\kappa \cdots \otimes_\kappa p(x_N;\theta), \qquad (22) \\ l_\kappa(\theta) &:= \ln_\kappa L_\kappa(\theta) = \sum_{i=1}^{N} \ln_\kappa p(x_i;\theta), \end{aligned}$$

respectively. By taking a limit $\kappa \to 0$, L_κ is the standard likelihood function on θ.

The maximum κ-likelihood estimator $\hat{\theta}$ is the maximizer of κ-likelihood function. We assume the existence of $\hat{\theta}$ in this paper. Since the parameter space Θ is assumed to be an open subset, $\hat{\theta}$ should be an interior point in Θ. From the monotonicity of the κ-logarithm $\ln_\kappa$, $\hat{\theta}$ is also the maximizer of κ-logarithm κ-likelihood function:

$$\hat{\theta} := \operatorname*{argmax}_{\theta\in\Theta} L_\kappa(\theta) \;=\; \operatorname*{argmax}_{\theta\in\Theta} \ln_\kappa L_\kappa(\theta).$$

Theorem 7. *Let $S_\kappa = \{p(x;\theta) | \theta \in \Theta\}$ be a κ-exponential family. Suppose that $M_\kappa = \{p(x;\theta(u)) | u \in U\}$ is a curved κ-exponential family of S_κ and $\{x_1, \ldots, x_N\}$ are N-observations from $p(x;\theta(u)) \in M_\kappa$. Then,*

(1) the maximum κ-likelihood estimator for S_κ in η-coordinates is given by:

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

(2) The κ-likelihood attains the maximum if and only if the κ-relative entropy attains the minimum.

Proof. (1) The κ-logarithm κ-likelihood function is given by:

$$l_\kappa(\theta) \;=\; \sum_{j=1}^{N} \ln_\kappa p(x_j;\theta) \;=\; \sum_{j=1}^{N}\left\{\sum_{i=1}^{n}\theta^i F_i(x_j) - \psi(\theta)\right\} \;=\; \sum_{i=1}^{n}\theta^i \sum_{j=1}^{N} F_i(x_j) - N\psi(\theta).$$

Hence, we obtain the κ-logarithm κ-likelihood equation:

$$\partial_i l_\kappa(\theta) = \sum_{j=1}^{N} F_i(x_j) - N\partial_i\psi(\theta) = 0.$$

From Proposition 6, the maximum κ-likelihood estimator for S_κ is given by:

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

(2) Denote $p(\hat{\eta}) = p(x;\hat{\eta}) \in S_\kappa$ by the probability distribution whose parameter is determined by the maximum likelihood $\hat{\eta}$. Since a κ-relative entropy coincides with a canonical divergence, we obtain:

$$\begin{aligned} D^\kappa(p(\hat{\eta}), p(\theta(u))) &= D(p(\theta(u)), p(\hat{\eta})) = \psi(\theta(u)) + \phi(\hat{\eta}) - \sum_{i=1}^{n} \theta^i(u)\hat{\eta}_i \\ &= \phi(\hat{\eta}) - \frac{1}{N}\ln_\kappa L_\kappa(\theta(u)). \end{aligned}$$

This implies that the κ-likelihood attains the maximum if and only if the κ-relative entropy attains the minimum. □

Since the κ-relative entropy attains the minimum at the κ-maximum likelihood estimator, we say that Theorem 7 is a divergence projection theorem for the κ-exponential family. We remark again that similar arguments hold for any χ-exponential families if the χ-independence is well defined.

9. Conclusion

In this paper, we discussed deformed algebras and generalizations of expectations for χ-exponential families. In particular, we clarified how to use deformed algebraic structures for deformed exponential families. We introduced the canonical expectation for χ-exponential families, whereas the normalized χ-escort expectation has been known in anomalous statistical physics. We then considered information geometric properties of deformed exponential families. Though the canonical expectation is not an expectation with respect to a probability density, it naturally characterizes a generalized Fisher metric and the α-divergence.

In addition, we studied the generalization of independence and introduced a generalized maximum likelihood method for the κ-exponential family. In particular, a divergence projection-type theorem was obtained in the case of the κ-maximum likelihood method. A deformed independence is not defined explicity in general, since it is difficult to describe anti-exponential conditions for χ-exponential functions. On the other hand, the κ-independence for the κ-exponential family is always well defined. This is an advantage of the κ-exponential family in the class of χ-exponential families.

Acknowledgments

The authors would like to express their sincere gratitude to the anonymous reviewers for the constructive comments that improved this paper. Hiroshi Matsuzoe is partially supported by The Ministry of Education, Culture, Sports, Science and Technology (MEXT) Grants-in-Aid for Scientific

Research (KAKENHI) Grant Numbers 23740047, 26108003 and 15K04842. Tatsuaki Wada is partially supported by Japan Society for the Promotion of Science (JSPS) KAKENHI Grant Number 25400188.

Author Contributions

This work has been conceived of and prepared by both authors. Both authors have read and approved the final manuscript.

Conflicts of Interest

The authors declare no conflict of interest.

References

1. Naudts, J. *Generalised Thermostatistics*; Springer: London, UK, 2011.
2. Fujimoto, Y.; Murata, N. A generalization of independence in naive Bayes model. In *Intelligent Data Engineering and Automated Learning—IDEAL 2010*, Proceedings of 11th International Conference on Intelligent Data Engineering and Automated Learning, Paisley, UK, 1–3 September 2010; Lecture Notes in Computer Science, Volume 6283; pp. 153–161.
3. Fujimoto, Y.; Murata, N. A generalisation of independence in statistical models for categorical distribution. *Int. J. Data Min. Model. Manag.* **2012**, *4*, 172–187.
4. Borgesa, E.P. A possible deformed algebra and calculus inspired in nonextensive thermostatistics. *Physica A* **2004**, *340*, 95–101.
5. Kaniadakis, G. Non-linear kinetics underlying generalized statistics. *Physica A* **2001**, *296*, 405–425.
6. Kaniadakis, G. Theoretical foundations and mathematical formalism of the power-law tailed statistical distributions. *Entropy* **2013**, *15*, 3983–4010.
7. Scarfone, A.M. Entropic forms and related algebras. *Entropy* **2013**, *15*, 624–649.
8. Tsallis, C. *Introduction to Nonextensive Statistical Mechanics: Approaching a Complex World*; Springer: New York, NY, USA, 2009.
9. Amari, S.; Nagaoka, H. *Method of Information Geometry*; Translations of Mathematical Monographs; American Mathematical Society: Providence, RI, USA and Oxford University Press: Oxford, UK, 2000.
10. Amari, S.; Ohara, A.; Matsuzoe, H. Geometry of deformed exponential families: Invariant, dually-flat and conformal geometry. *Physica A* **2012**, *391*, 4308–4319.
11. Matsuzoe, H.; Henmi, M. Hessian structures and divergence functions on deformed exponential families. In *Geometric Theory of Information, Signals and Communication Technology*; Nielsen, F., Eds.; Springer: Basel, Switzerland, 2014; pp. 57–80.
12. Ohara, A. Geometry of distributions associated with Tsallis statistics and properties of relative entropy minimization. *Phys. Lett. A*, **2007**, *370*, 184–193.
13. Ohara, A. Geometric study for the Legendre duality of generalized entropies and its application to the porous medium equation. *Eur. Phys. J. B.*, **2009**, *70*, 15–28.

14. Matsuzoe, H.; Henmi, M. Hessian structures on deformed exponential families. In *Geometric Science of Information*, Proceedings of First International Conference on Geometric Science of Information (GSI 2013), Paris, France, 28–30 August 2013; Lecture Notes in Computer Science, Volume 8085; pp. 275–282.
15. Naudts, J. Estimators, escort probabilities, and ϕ-exponential families in statistical physics, *J. Inequal. Pure Appl. Math.* **2004**, *5*, 102.
16. Tanaka, M. Meaning of an escort distribution and τ-transformation. *J. Phys. Conf. Ser.* **2010**, *201*, 012007.
17. Suyari, H. *Fundamental Mathematics for Complex Systems*; Makinoshoten: Tokyo, Japan, 2010. (in Japanese)
18. Wilk G.; Wlodarczyk Z. Tsallis distribution with complex nonextensivity parameter q. *Physica A* **2014**, *413*, 53–58.
19. Lauritzen, S.L. *Extremal Families and Systems of Sufficient Statistics*; Lecture Notes in Statistics, Volume 49; Springer: New York, NY, USA, 1988.
20. Pistone, G. κ-exponential models from the geometrical viewpoint. *Eur. Phys. J. B*, **2009**, *70*, 29–37.
21. Murata, N.; Takenouchi, T.; Kanamori, T.; Eguchi, S. Information geometry of U-boost and Bregman divergence. *Neural Comput.* **2004**, *16*, 1437–1481.
22. Ohara, A.; Wada, T. Information geometry of q-Gaussian densities and behaviors of solutions to related diffusion equations. *J. Phys. A* **2010**, *43*, 035002.
23. Matsuzoe, H. Statistical manifolds and geometry of estimating functions. In *Prospects of Differential Geometry and Its Related Fields*, Proceedings of the 3rd International Colloquium on Differential Geomentry and Its Related Fields; Adachi, T., Hashimoto, H., Hristov, M.J., Eds.; World Scientific: Hackensack, NJ, USA, 2013; pp. 187–202.
24. Basu, A.; Harris, I.R.; Hjort, N.L.; Jones, M.C. Robust and efficient estimation by minimising a density power divergence. *Biometrika* **1998**, *85*, 549–559.
25. Zhang, J. Divergence function, duality, and convex analysis. *Neural Comput.* **2004**, *16*, 159–195.
26. Harsha, K.V.; Subrahamanian Moosath, K.S. Dually flat geometries of the deformed exponential family. *Physica A* **2015**, *433*, 136–147
27. Zhang, J. A note on monotone embedding in information geometry. *Entropy* **2015**, *17*, 4485–4499.
28. Shima, H. The Geometry of Hessian Structures, World Scientific: Hackensack, NJ, 2007.
29. Kurose, T. On the divergences of 1-conformally flat statistical manifolds. *Tôhoku Math. J.* **1994**, *46*, 427–433.
30. Matsuzoe, H.; Ohara, A. Geometry for q-exponential families. In *Recent progress in differential geometry and its related fields*; Proceedings of the 2nd International Colloquium on Differential Geomentry and Its Related Fields; Adachi, T., Hashimoto, H., Hristov, M.J., Eds.; World Scientific: Hackensack, NJ, USA, 2011; pp. 55–71.
31. Matsuzoe, H. Hessian structures on deformed exponential families and their conformal structures. *Differ. Geom. Appl.* **2014**, *35 Supplement*, 323–333.
32. Sakamoto, M.; Matsuzoe, H. A generalization of independence and multivariate Student's t-distributions. To appear in *Lecture Notes in Comput. Sci.*

33. Takatsu, A. Behaviors of φ-exponential distributions in Wasserstein geometry and an evolution equation, *SIAM J. Math. Anal.* **2013**, *45*, 2546–2556.
34. Harsha, K.V.; Subrahamanian Moosath, K.S. Geometry of F-likelihood estimators and F-Max-Ent theorem. *AIP Conf. Proc.* **2015**, *1641*, 263–270.
35. Suyari, H.; Tsukada, M. Law of error in Tsallis statistics, *IEEE Trans. Inf. Theory* **2005**, *51*, 753–757.
36. Wada T.; Suyari, H. κ-generalization of Gauss's law of error. *Phys. Lett. A* **2006**, *348*, 89–93.