Article

A Sequence of Escort Distributions and Generalizations of Expectations on q-Exponential Family

Hiroshi Matsuzoe

Department of Computer Science and Engineering, Nagoya Institute of Technology, Nagoya 466-8555, Japan; matsuzoe1@nitech.ac.jp; Tel.: +81-52-735-5143

Abstract: In the theory of complex systems, long tailed probability distributions are often discussed. For such a probability distribution, a deformed expectation with respect to an escort distribution is more useful than the standard expectation. In this paper, by generalizing such escort distributions, a sequence of escort distributions is introduced. As a consequence, it is shown that deformed expectations with respect to sequential escort distributions effectively work for anomalous statistics. In particular, it is shown that a Fisher metric on a q-exponential family can be obtained from the escort expectation with respect to the second escort distribution, and a cubic form (or an Amari–Chentsov tensor field, equivalently) is obtained from the escort expectation with respect to the third escort distribution.

1. Introduction

Long tailed probability distributions and their related probability distributions are important objects in anomalous statistical physics (cf. [1–3]). For such long tailed probability distributions, the standard expectation does not exist in general. Therefore, the notion of escort distribution has been introduced [4]. Since an escort distribution gives a suitable weight for tail probability, the escort expectation which is the expectation with respect to an escort distribution is more useful than the standard one.

In anomalous statistics, a deformed exponential function and a deformed logarithm function play essential roles. In fact, a deformed exponential family is an important statistical model in anomalous statistics. Such a statistical model is described by such a deformed exponential function. In particular, the set of all q-normal distributions (or Student's t-distributions, equivalently) is a q-exponential family, which is described by a q-deformed exponential function [5] (see also [6,7]).

On the other hand, a generalized score function is defined from a deformed logarithm function. In the previous works, the author showed that a deformed score function is unbiased with respect to the escort expectation [8,9]. This implies that a deformed score function is regarded as an estimating function on a deformed exponential family. In addition, in information geometry, it is known that a deformed exponential family has a statistical manifold structure. Then a deformed score function is regarded as a tangent vector on this statistical manifold [6,10]. Therefore, properties of escort expectations are closely related to geometric structures on a deformed exponential family.

In this paper, we introduce a sequence of escort distributions, then we consider a sequential structure of escort expectations. It is known that a deformed exponential family naturally has at

least three kind of different statistical manifold structures [6,11]. Then we show that such statistical manifold structures can be obtained from a sequential structure of escort expectations. In particular, we show that a Fisher metric on a q-exponential family can be obtained from the deformed expectations with respect to the second escort distribution, and a cubic form (or an Amari–Chentsov tensor field, equivalently) is obtained from the deformed expectations with respect to the third escort distribution.

This paper is written based on the proceeding paper [7]. However, this paper focuses on deformed expectations of a q-exponential family, whereas the previous paper focused on deformed independences. We remark that several authors have been studying deformed expectations recently. See [12,13], for example.

2. Deformed Exponential Families

In this paper, we assume that all objects are smooth for simplicity. Let us review preliminary facts about deformed exponential functions and deformed exponential families. For more details, see [2,6], for example. Historically, Tsallis [14] introduced the notion of q-exponential function and Naudts [5] introduced the notion of q-exponential family together with a further generalization. Such a historical note is provided in [2].

Let $\mathbf{R}_{++}$ be the set of all positive real numbers, $\mathbf{R}_{++} := \{x \in \mathbf{R} | x > 0\}$. Let χ be a strictly increasing function from $\mathbf{R}_{++}$ to $\mathbf{R}_{++}$. We define a χ-*logarithm function* or a *deformed logarithm function* by

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

The inverse of $\ln_\chi s$ is called a χ-*exponential function* or a *deformed exponential function*, which is defined by

$$\exp_\chi t := 1 + \int_0^t u(s) ds,$$

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

From now on, we suppose that χ is a power function, that is, $\chi(t) = t^q$. Then the deformed logarithm and the deformed exponential are defined by

$$\ln_q s := \frac{s^{1-q} - 1}{1 - q}, \qquad\qquad (s > 0),$$

$$\exp_q t := (1 + (1-q)t)^{\frac{1}{1-q}}, \qquad\qquad (1 + (1-q)t > 0).$$

We say that $\ln_q s$ is a q-*logarithm function* and $\exp_q t$ is a q-*exponential function*. In this case, the function $u(s)$ is given by

$$u(s) = (1 + (1-q)s)^{\frac{q}{1-q}} = \{\exp_q s\}^q.$$

By taking a limit $q \to 1$, these functions coincide with the standard logarithm $\ln s$ and the standard exponential $\exp t$, respectively.

A statistical model S_q is called a q-*exponential family* if

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

where $F_1(x), \ldots, F_n(x)$ are functions on a sample space Ω, $\theta = {}^t(\theta^1, \ldots, \theta^n)$ is a parameter, and $\psi(\theta)$ is the normalization with respect to the parameter θ. Under suitable conditions, S_q is regarded as a manifold with a local coordinate system $\{\theta^1, \ldots, \theta^n\}$. In this case, we call $\{\theta^i\}$ a *natural coordinate system*.

　　　In this paper, we focus on the q-exponential case. However, many results for the q-exponential family can be generalized for the χ-exponential family (cf. [6,8]). We remark that a q-exponential family and a χ-exponential family have further generalizations. See [15], for example.

Example 1 (Student's t-distribution (cf. [2,6,7]))**.** *Fix a number q $(1 < q < 1 + 2/d,\ d \in \mathbf{N})$, and set $v = -d - 2/(1-q)$. We define a d-dimensional Student's t-distribution with degree of freedom v or a q-Gaussian distribution by*

$$p_q(x; \mu, \Sigma) := \frac{\Gamma\left(\frac{1}{q-1}\right)}{(\pi v)^{\frac{d}{2}} \Gamma\left(\frac{v}{2}\right) \sqrt{\det(\Sigma)}} \left[1 + \frac{1}{v}{}^t(x - \mu)\Sigma^{-1}(x - \mu)\right]^{\frac{1}{1-q}},$$

where $X = {}^t(X_1, \dots, X_d)$ is a random vector on $\mathbf{R}^d$, $\mu = {}^t(\mu^1, \dots, \mu^d)$ is a location vector on $\mathbf{R}^d$ and Σ is a scale matrix on $\mathrm{Sym}^+(d)$. For simplicity, we assume that Σ is invertible. Otherwise, we should choose a suitable basis $\{v^\alpha\}$ on $\mathrm{Sym}^+(d)$ such that $\Sigma = \sum_\alpha w_\alpha v^\alpha$. Then, the set of all Student's t-distributions is a q-exponential family. In fact, setting parameters by

$$z_q = \frac{(\pi v)^{\frac{d}{2}} \Gamma\left(\frac{v}{2}\right) \sqrt{\det(\Sigma)}}{\Gamma\left(\frac{1}{q-1}\right)}, \quad \tilde{R} = \frac{z_q^{q-1}}{(1-q)d + 2}\Sigma^{-1}, \quad and \quad \theta = 2\tilde{R}\mu, \tag{2}$$

we have

$$
\begin{aligned}
p_q(x; \mu, \Sigma) &= \frac{1}{z_q}\left[1 + \frac{1}{v}{}^t(x - \mu)\Sigma^{-1}(x - \mu)\right]^{\frac{1}{1-q}} \\
&= \left[\left(\frac{1}{z_q}\right)^{1-q} - \frac{1-q}{(1-q)d + 2}\left(\frac{1}{z_q}\right)^{1-q}{}^t(x - \mu)\Sigma^{-1}(x - \mu)\right]^{\frac{1}{1-q}} \\
&= \exp_q\left[-{}^t(x - \mu)\tilde{R}(x - \mu) + \ln_q\frac{1}{z_q}\right] \\
&= \exp_q\left[\sum_{i=1}^d \theta^i x_i - \sum_{i=1}^d \tilde{R}_{ii} x_i^2 - 2\sum_{i<j}\tilde{R}_{ij} x_i x_j - \frac{1}{4}{}^t\theta\tilde{R}^{-1}\theta + \ln_q\frac{1}{z_q}\right].
\end{aligned}
$$

Since $\theta \in \mathbf{R}^d$ and $\tilde{R} \in \mathrm{Sym}^+(d)$, the set of all Student's t-distributions is a $d(d+3)/2$-dimensional q-exponential family. The normalization $\psi(\theta)$ is given by

$$\psi(\theta) = \frac{1}{4}{}^t\theta\tilde{R}^{-1}\theta - \ln_q\frac{1}{z_q}.$$

　　　A univariate Student's t-distribution is a well-known probability distribution in elementary statistics. We denote it by

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

where $\mu \in \mathbf{R}$ is a location parameter, $\sigma \in \mathbf{R}_{++}$ is a scale parameter, and Z_q is the normalization defined by

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

In this case, the degree of freedom is $v = (3 - q)/(q - 1)$. Conversely, the parameter q is give by

$$q = \frac{v + 3}{v + 1}. \tag{4}$$

3. Escort Distributions and Generalizations of Expectations

In anomalous statistics, a generalized expectation, called an escort expectation, is often discussed since the standard expectation does not exist in general (cf. [2,5,6]). In this section, we recall generalizations of expectations and introduce a sequential structure of escort distributions.

Let S_q be a q-exponential family. For a given $p(x;\theta) \in S_q$ we define the *q-escort distribution* $P_q(x;\theta)$ of $p(x;\theta)$ and the *normalized q-escort distribution* $P_q^{esc}(x;\theta)$ by

$$
\begin{aligned}
P_q(x;\theta) &:= P_{q,(1)}(x;\theta) := \{p(x;\theta)\}^q, \\
P_q^{esc}(x;\theta) &:= \frac{1}{Z_q(p)}\{p(x;\theta)\}^q, \quad \text{where} \quad Z_q(p) = \int_\Omega \{p(x;\theta)\}^q dx,
\end{aligned}
$$

respectively. For a q-exponential family $S_q = \{p_q(x;\theta)\}$, the set of normalized escort distributions $S_{q'} = \{P_q^{esc}(x;\theta)\}$ is a q'-exponential family with $q' = (2q-1)/q$.

Example 2. *Let $t_\nu(x;\mu,\sigma)$ be a univariate Student's t-distribution with degree of freedom ν. Then its normalized escort distribution is also a univariate Student's t-distribution with degree of freedom $\nu + 2$. In fact, from Equation (4), a direct calculation shows that*

$$
q' = \frac{2q-1}{q} = \frac{\nu+5}{\nu+3}.
$$

This implies that the degree of freedom $\nu' = \nu + 2$. Therefore, we obtain a sequence of escort distributions from a given Student's t-distribution t_ν:

$$
t_\nu \to t_{\nu+2} \to t_{\nu+4} \to \cdots .
$$

This sequence is called a τ-sequence, and the procedure to obtain from a given t-distribution to another t-distribution through an escort distribution is called the τ-transformation [16].

For a given $p_q(x;\theta) \in S_q$, we can define the escort of an escort distribution

$$
\widetilde{P}_q(x;\theta) := P_{q,(2)}(x;\theta) := q\{P_q(x;\theta)\}^{q'} = q\{p_q(x;\theta)\}^{2q-1}.
$$

We call $\widetilde{P}_q(x;\theta)$ the *second escort distribution* of $p_q(x;\theta)$. The coefficient q before $\{p_q(x;\theta)\}^{2q-1}$ comes from considerations of U-information geometry [17]. We will discuss in the latter part of Section 5.

Similarly, we can define the *n-th escort distribution* $P_{q,(n)}(x;\theta)$ from the sequence of escort distributions:

$$
P_{q,(n)}(x;\theta) := \{q(2q-1)\cdots((n-1)q-(n-2))\}\{p_q(x;\theta)\}^{nq-(n-1)}. \tag{5}
$$

Let $f(x)$ be a function on Ω. The *q-expectation* $E_{q,p}[f(x)]$ and the *normalized q-expectation* $E_{q,p}^{esc}[f(x)]$ with respect to $p(x;\theta) \in S_q$ are defined by

$$
\begin{aligned}
E_{q,p}[f(x)] &:= \int_\Omega f(x)P_q(x;\theta)dx, \\
E_{q,p}^{esc}[f(x)] &:= \int_\Omega f(x)P_q^{esc}(x;\theta)dx,
\end{aligned}
$$

respectively. We denote by $\widetilde{E_{q,p}}[f(x)]$ the expectation with respect to the second escort distribution $\widetilde{P}_q(x;\theta)$, that is,

$$
\widetilde{E_{q,p}}[f(x)] := \int_\Omega f(x)\widetilde{P}_q(x;\theta)dx = q\int_\Omega f(x)\{p_q(x;\theta)\}^{2q-1}dx.
$$

Since a differential of a power function is also a power function, we can give a characterization for escort distributions.

Proposition 1. *Suppose that S_q is a q-exponential family defined by (1). Then the n-th escort distribution is given by the n-th differential of q-exponential function. That is, by setting $u(t) = (\exp_q t)'$, we have the following formula:*

$$p_q(x;\theta) = \exp_q\left(\sum_{i=1}^{n}\theta^i F_i(x) - \psi(\theta)\right),$$

$$P_q(x;\theta) = P_{q,(1)}(x;\theta) = u\left(\sum_{i=1}^{n}\theta^i F_i(x) - \psi(\theta)\right),$$

$$\widetilde{P}_q(x;\theta) = P_{q,(2)}(x;\theta) = u'\left(\sum_{i=1}^{n}\theta^i F_i(x) - \psi(\theta)\right),$$

$$\vdots \qquad\qquad \vdots$$

$$P_{q,(n)}(x;\theta) = u^{(n-1)}\left(\sum_{i=1}^{n}\theta^i F_i(x) - \psi(\theta)\right),$$

$$\vdots \qquad\qquad \vdots$$

Proof. Since a q-exponential function is $\exp_q(x) = (1 + (1-q))^{1/(1-q)}$, its differential is given by

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

Therefore, we obtain $P_q(x;\theta) = u\left(\sum_{i=1}^{n}\theta^i F_i(x) - \psi(\theta)\right)$.

By induction, the n-th differential of $u(x)$ coincides with the n-th escort distribution $P_{q,(n)}$, which is given by Equation (5). □

4. Statistical Manifolds and Their Generalized Conformal Structures

In this section, we us review the geometry of statistical manifolds. For more details about the geometry of statistical manifolds, see [18,19].

Let (S, g) be a Riemannian manifold and ∇ be a torsion-free affine connection on S. We say that the triplet (S, ∇, g) is a *statistical manifold* if ∇g is totally symmetric. In this case, we can define a totally symmetric $(0,3)$-tensor field by

$$C(X,Y,Z) := (\nabla_X g)(Y,Z) = Xg(Y,Z) - g(\nabla_X Y, Z) - g(Y, \nabla_X Z),$$

where X, Y and Z are arbitrary vector fields on S. The tensor field C is called a *cubic form* or an *Amari–Chentsov tensor field*.

The notion of statistical manifold was introduced by Lauritzen [20]. He called the triplet (S, g, C) a statistical manifold. In this paper, the definition is followed to Kurose [18]. Though these two definitions are different, the other statistical manifold structure can be obtained from a given one, However, the motivation for the notion of conformal equivalence using (S, g, C) is different from that one using (S, ∇, g), which we will discuss in the latter part of this section.

For a given statistical manifold (S, ∇, g), we can define another torsion-free affine connection ∇^* on S by

$$Xg(Y,Z) = g(\nabla_X Y, Z) + g(Y, \nabla^*_X Z).$$

The connection ∇^* is called the *dual connection* of ∇ with respect to g. The triplet (S, ∇^*, g) is also a statistical manifold, which is called the *dual statistical manifold* of (S, ∇, g). The cubic form is given by the difference of two affine connections ∇^* and ∇:

$$C(X, Y, Z) = g(\nabla^*_X Y - \nabla_X Y, Z).$$

We define generalized conformal structures for statistical manifolds followed to Kurose [18]. Two statistical manifolds (S, ∇, g) and $(S, \bar{\nabla}, \bar{g})$ are said to be 1-*conformally equivalent* if there exists a function $\lambda : S \to \mathbf{R}_{++}$ such that

$$\begin{align}
\bar{g}(X, Y) &= \lambda g(X, Y), \tag{6} \\
\bar{\nabla}_X Y &= \nabla_X Y - g(X, Y) \operatorname{grad}_g(\ln \lambda), \tag{7}
\end{align}$$

where $\operatorname{grad}_g(\ln \lambda)$ is the gradient vector field of $\ln \lambda$ with respect to g, that is, $g(X, \ln \lambda) = X(\ln \lambda)$. We say that (S, ∇, g) is 1-*conformally flat* if (S, ∇, g) is locally 1-conformally equivalent to a flat statistical manifold.

Two statistical manifolds (S, ∇, g) and $(S, \bar{\nabla}, \bar{g})$ are said to be (-1)-*conformally equivalent* if there exists a function $\lambda : S \to \mathbf{R}_{++}$ such that

$$\begin{align}
\bar{g}(X, Y) &= \lambda g(X, Y), \\
\bar{\nabla}_X Y &= \nabla_X Y + d(\ln \lambda)(Y) X + d(\ln \lambda)(X) Y, \tag{8}
\end{align}$$

where $d(\ln \lambda)(X) = X(\ln \lambda)$. If two statistical manifolds (S, ∇, g) and $(S, \bar{\nabla}, \bar{g})$ are 1-conformally equivalent, then their dual statistical manifolds (S, ∇^*, g) and $(S, \bar{\nabla}^*, \bar{g})$ are (-1)-conformally equivalent.

Proposition 2. *If two statistical manifolds (S, ∇, g) and $(S, \bar{\nabla}, \bar{g})$ are 1-conformally equivalent, then their cubic forms have the following relation:*

$$\frac{1}{\lambda} \bar{C}(X, Y, Z) = C(X, Y, Z) + g(Y, Z) d(\ln \lambda)(X) + g(Z, X) d(\ln \lambda)(Y) + g(X, Y) d(\ln \lambda)(Z).$$

Proof. From Equations (7) and (8), we obtain

$$\bar{\nabla}_X Y = \nabla_X Y + d(\ln \lambda)(Y) X + d(\ln \lambda)(X) Y + g(X, Y) \operatorname{grad}_g(\ln \lambda).$$

By taking an inner product with respect to g, we obtain the result. □

5. Statistical Manifold Structures on q-Exponential Families

In this section, we consider statistical manifold structures on a q-exponential family. It is known that a q-exponential family naturally has at least three kinds of statistical manifold structures (cf. [6,8]). We reformulate these structures from the viewpoint of the sequence of escort distributions. In this paper, we omit the details about information geometry. See [21,22] for further details.

Firstly, we review basic facts about q-exponential family. Let S_q be a q-exponential family. The normalization $\psi(\theta)$ on S_q is convex, but may not be strictly convex. In fact, we obtain the following proposition.

Proposition 3. *Let $S_q = \{p(x;\theta)\}$ be a q-exponential family. Then the normalization function $\psi(\theta)$ is convex.*

Proof. Set $u(x) = (\exp_q x)'$ and $\partial_i = \partial/\partial\theta^i$. Then we have

$$
\begin{aligned}
\partial_i p(x;\theta) &= u\left(\sum \theta^k F_k(x) - \psi(\theta)\right)(F_i(x) - \partial_i\psi(\theta)), \\
\partial_i\partial_j p(x;\theta) &= u'\left(\sum \theta^k F_k(x) - \psi(\theta)\right)(F_i(x) - \partial_i\psi(\theta))(F_j(x) - \partial_j\psi(\theta)) \\
&\quad - u\left(\sum \theta^k F_k(x) - \psi(\theta)\right)\partial_i\partial_j\psi(\theta).
\end{aligned}
\tag{9}
$$

Since $\partial_i \int_\Omega p(x;\theta)dx = \int_\Omega \partial_i p(x;\theta)dx = 0$ and $\int_\Omega \partial_i\partial_j p(x;\theta)dx = 0$, we have

$$
\begin{aligned}
Z_q(p) &= \int_\Omega \{(p(x;\theta)\}^q dx = \int_\Omega u\left(\sum \theta^k F_k(x) - \psi(\theta)\right)dx, \\
\partial_i\partial_j\psi(\theta) &= \frac{1}{Z_q(p)}\int_\Omega u'\left(\sum \theta^k F_k(x) - \psi(\theta)\right)(F_i(x) - \partial_i\psi(\theta))(F_j(x) - \partial_j\psi(\theta))dx.
\end{aligned}
\tag{10}
$$

For an arbitrary vector $c = {}^t(c^1, c^2, \ldots, c^n) \in \mathbf{R}^n$, since $Z_q(p) > 0$ and $u''(x) > 0$, we have

$$
\sum_{i,j=1}^n c^i c^j (\partial_i\partial_j\psi(\theta)) = \frac{1}{Z_q(p)}\int_\Omega u''\left(\sum_{k=1}^n \theta^k F_k(x) - \psi(\theta)\right)\left\{\sum_{i=1}^n c^i(F_i(x) - \partial_i\psi(\theta))\right\}^2 dx \geq 0.
$$

This implies that the Hessian matrix $(\partial_i\partial_j\psi(\theta))$ is semi-positive definite. □

We assume that ψ is strictly convex in this paper. Under this assumption, we can induce many geometric structures for a q-exponential family.

Since ψ is strictly convex, we can define a Riemannian metric and a cubic form by

$$
\begin{aligned}
g_{ij}^q(\theta) &:= \partial_i\partial_j\psi(\theta), \\
C_{ijk}^q(\theta) &:= \partial_i\partial_j\partial_k\psi(\theta).
\end{aligned}
$$

We call g^q and C^q a *q-Fisher metric* and a *q-cubic form*, respectively [23,24]. Since g^q is a Hessian of a function ψ, g^q is a *Hessian metric*, and ψ is the *potential* of g^q with respect to the natural coordinate $\{\theta^i\}$ [25].

For a fixed real number α, set

$$
g^q\left(\nabla_X^{q(\alpha)}Y, Z\right) := g^q\left(\nabla_X^{q(0)}Y, Z\right) - \frac{\alpha}{2}C^q(X, Y, Z),
\tag{11}
$$

where $\nabla^{q(0)}$ is the Levi-Civita connection with respect to g^q. Since g^q is a Hessian metric, from standard arguments in Hessian geometry [25], $\nabla^{q(e)} := \nabla^{q(1)}$ and $\nabla^{q(m)} := \nabla^{q(-1)}$ are flat affine connections and mutually dual with respect to g^q. Therefore, the triplets $(S_q, \nabla^{q(e)}, g^q)$ and $(S_q, \nabla^{q(m)}, g^q)$ are flat statistical manifolds, and the quadruplet $(S_q, g^q, \nabla^{q(e)}, \nabla^{q(m)})$ is a dually flat space.

Under q-expectations, we have the following proposition (cf. [10]).

Proposition 4. *For S_q a q-exponential family,* (1) *Set $\eta_i = E_{q,p}^{esc}[F_i(x)]$. Then $\{\eta_i\}$ is a $\nabla^{q(m)}$-affine coordinate system such that*

$$
g^q\left(\frac{\partial}{\partial\theta^i}, \frac{\partial}{\partial\eta_j}\right) = \delta_i^j.
$$

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

Next, let us consider the standard Fisher metric and the standard cubic form. Suppose that $S := \{p(x;\theta)\}$ is a statistical model. Set $p_\theta := p(x;\theta)$, for simplicity. We define the *(standard) Fisher metric* g^F on S_q by

$$g_{ij}^F(\theta) := \int_\Omega (\partial_i \ln p_\theta)(\partial_j \ln p_\theta) p_\theta dx,$$

and the *(standard) cubic form* or the *Amari–Chentsov vector field* C^F by

$$C_{ijk}^F(\theta) := \int_\Omega (\partial_i \ln p_\theta)(\partial_j \ln p_\theta)(\partial_j \ln p_\theta) p_\theta dx.$$

From similar arguments of (11), we can define an *α-connection* $\nabla^{(\alpha)}$ on S_q, and we can obtain a statistical manifold structure $(S_q, \nabla^{(\alpha)}, g^F)$. In this case, $(S_q, \nabla^{(\alpha)}, g^F)$ is called an *invariant statistical manifold* [21,22].

A Fisher metric and a cubic form have the following representation using a sequence of escort distributions,

Theorem 1. *Let S_q be a q-exponential family. For $p(x;\theta) \in S_q$, suppose that $P_{q,(2)}(x;\theta)$ and $P_{q,(3)}(x;\theta)$ are the second and the third escort distribution of $p(x;\theta)$, respectively. Then the Fisher metric g^F and the cubic form C^F are given as follows:*

$$g_{ij}^F(\theta) = \frac{1}{q}\int_\Omega (\partial_i \ln_q p_\theta)(\partial_j \ln_q p_\theta) P_{q,(2)}(x;\theta) dx, \tag{12}$$

$$C_{ijk}^F(\theta) = \frac{1}{q(2q-1)}\int_\Omega (\partial_i \ln_q p_\theta)(\partial_j \ln_q p_\theta)(\partial_k \ln_q p_\theta) P_{q,(3)}(x;\theta) dx. \tag{13}$$

Proof. Differentiating the q-logarithm, we have

$$\partial_i \ln_q p_\theta = \partial_i \left(\frac{p_\theta^{1-q}-1}{1-q}\right) = p_\theta^{-q}\partial_i p(\theta) = p_\theta^{1-q}\partial_i \ln p(\theta).$$

Therefore, we obtain

$$\begin{aligned}
\frac{1}{q}\int_\Omega (\partial_i \ln_q p_\theta)(\partial_j \ln_q p_\theta) P_{q,(2)}(x;\theta) dx &= \int_\Omega p_\theta^{1-q}(\partial_i \ln p_\theta) p_\theta^{1-q}(\partial_j \ln p_\theta) p_\theta^{2q-1}(x;\theta) dx \\
&= \int_\Omega (\partial_i \ln p_\theta)(\partial_j \ln p_\theta) p_\theta(x;\theta) dx \\
&= g_{ij}^F(\theta).
\end{aligned}$$

By a similar argument, we obtain the representation for C^F. $\square$

We define an *α-divergence* $D^{(\alpha)}$ with $\alpha = 1 - 2q$ and a *q-relative entropy* (or a *normalized Tsallis relative entropy*) D_q^T by

$$D^{(1-2q)}(p(x),r(x)) = \frac{1}{q}E_{q,p}[\ln_q p(x) - \ln_q r(x)] = \frac{1 - \int_\Omega p(x)^q r(x)^{1-q} dx}{q(1-q)}, \tag{14}$$

$$D_q^T(p(x),r(x)) = E_{q,p}^{esc}[\ln_q p(x) - \ln_q r(x)] = \frac{1 - \int_\Omega p(x)^q r(x)^{1-q} dx}{(1-q)Z_q(p)}, \tag{15}$$

respectively. It is known that the α-divergence $D^{(1-2q)}(r,p)$ induces a statistical manifold structure $(S_q, g^F, \nabla^{(2q-1)})$, where g^F is the Fisher metric on S_q and $\nabla^{(2q-1)}$ is the α-connection with $\alpha = 2q - 1$, and the q-relative entropy $D_q^T(r,p)$ induces $(S_q, g, \nabla^{q(e)})$.

Theorem 2 (cf. [10,24]). *For a q-exponential family S_q, two statistical manifolds $(S_q, g^F, \nabla^{(2q-1)})$ and $(S_q, g, \nabla^{q(e)})$ are 1-conformally equivalent. In particular, an invariant statistical manifold $(S_q, g^F, \nabla^{(2q-1)})$ is 1-conformally flat. Riemannian metrics and cubic forms have the following relations:*

$$g_{ij}^q(\theta) = \frac{q}{Z_q(p)} g_{ij}^F(\theta), \tag{16}$$

$$\begin{aligned} C_{ijk}^q(\theta) = {} & \frac{q}{Z_q(p)}(2q-1)C_{ijk}^F(\theta) \\ & - \frac{q}{Z_q(p)}\left\{ g_{ij}^F \partial_k \ln Z_q(p) + g_{jk}^F(\theta)\partial_i \ln Z_q(p) + g_{ki}^F(\theta)\partial_j \ln Z_q(p) \right\}. \end{aligned} \tag{17}$$

Proof. The results were essentially obtained in [10]. However, we give a simpler proof for Equations (16) and (17). The key idea is a sequence of escort distributions and the escort representations of g^F and C^F in Theorem 1.

From Equation (10), we directly obtain the conformal equivalence relation (16) using the escort representation of g^F in (12).

By differentiating (9) and taking an integration, we obtain

$$\begin{aligned} 0 = {} & \int_\Omega u''\left(\sum \theta^l F_l(x) - \psi(\theta)\right)(F_i(x) - \partial_i\psi(\theta))(F_j(x) - \partial_j\psi(\theta))(F_k(x) - \partial_k\psi(\theta))dx \\ & - \int_\Omega u'\left(\sum \theta^l F_l(x) - \psi(\theta)\right)(F_k(x) - \partial_k\psi(\theta))\partial_i\partial_j\psi(\theta)dx \\ & - \int_\Omega u'\left(\sum \theta^l F_l(x) - \psi(\theta)\right)(F_i(x) - \partial_i\psi(\theta))\partial_j\partial_k\psi(\theta)dx \\ & - \int_\Omega u'\left(\sum \theta^l F_l(x) - \psi(\theta)\right)(F_j(x) - \partial_j\psi(\theta))\partial_k\partial_i\psi(\theta)dx \\ & - \int_\Omega u\left(\sum \theta^l F_l(x) - \psi(\theta)\right)\partial_i\partial_j\partial_k\psi(\theta)dx. \end{aligned}$$

Since $Z_q(p) = \int_\Omega P_q(x;\theta)dx$, we have

$$\partial_i Z_q(p) = \partial_i \int_\Omega P_q(x;\theta)dx = \int_\Omega \partial_i P_q(x;\theta)dx = \int_\Omega \tilde{P}_q(x;\theta)(F_i(x) - \partial_i\psi(\theta))dx.$$

From the escort representation of C^F in (13), and Proposition 1, we obtain Equation (17) since $g_{ij}^q(\theta) = \partial_i\partial_j\psi(\theta)$ and $C_{ijk}^q(\theta) = \partial_i\partial_j\partial_k\psi(\theta)$. □

We remark that the cubic form of $(S_q, g^F, \nabla^{(2q-1)})$ is not C^F but $(2q-1)C^F$.

The difference of a α-divergence and a q-relative entropy is only the normalization $q/Z_q(p)$. This implies that a normalization for probability density imposes a generalized conformal change for a statistical model.

In the next part of this section, let us consider another statistical manifold on S_q (cf. [6,17,26]).

Recall that a Fisher metric g^F has the following representation:

$$g_{ij}^F(\theta) = \int_\Omega (\partial_i \ln p_\theta)(\partial_j p_\theta)dx.$$

In information geometry, $\partial_i \ln p_\theta$ is called an *e-representation (exponential representation)* of p_θ, and $\partial_j p_\theta$ is called a *m-representation (mixture representation)*. Intuitively, $\partial_i \ln p_\theta$ and $\partial_j p_\theta$ are regarded as tangent vectors on a statistical model. Hence a Fisher metric is regarded as a L^2-inner product of *e*- and *m*-representations.

Let us generalize *e*- and *m*-representations for a q-exponential family. For $p_\theta \in S_q$, we call $\partial_i \ln_q p_\theta$ a *q-score function*. Then we define a Riemannian metric g^M by

$$g_{ij}^M(\theta) = \int_\Omega (\partial_i \ln_q p_\theta)(\partial_j p_\theta)dx. \tag{18}$$

By differentiating the above equation, we can define mutually dual torsion-free affine connections $\nabla^{M(e)}$ and $\nabla^{M(m)}$:

$$\Gamma_{ij,k}^{M(e)}(\theta) \quad := \quad \int_\Omega (\partial_i \partial_j \ln_q p_\theta)(\partial_k p_\theta)dx,$$

$$\Gamma_{ij,k}^{M(m)}(\theta) \quad := \quad \int_\Omega (\partial_k \ln_q p_\theta)(\partial_i \partial_j p_\theta)dx,$$

where $\Gamma_{ij,k}^{M(e)}$ and $\Gamma_{ij,k}^{M(m)}$ are the Christoffel symbols of $\nabla^{M(e)}$ and $\nabla^{M(m)}$ of the first kind, respectively. It is known that g^M is a Hessian metric, and the quadruplet $(S_q, g^M, \nabla^{M(e)}, \nabla^{M(m)})$ is a dually flat space. In addition, a natural parameter $\{\theta^i\}$ is a $\nabla^{M(e)}$-affine coordinate sysem. Therefore, the cubic form for $(S_q, \nabla^{M(e)}, g^M)$ is

$$C_{ijk}^M(\theta) = \Gamma_{ij,k}^{M(m)}(\theta). \tag{19}$$

We remark that the statistical manifold structure $(S_q, \nabla^{M(e)}, g^M)$ is induced from a *β-divergence* [17,26] (or a *density power divergence* [27]):

$$D_{1-q}(p,r) := \int_\Omega \left\{ p(x)\frac{p(x)^{1-q} - r(x)^{1-q}}{1-q} - \frac{p(x)^{2-q} - r(x)^{2-q}}{2-q} \right\} dx. \tag{20}$$

Theorem 3. *For the statistical manifold structure $(S_q, \nabla^{M(e)}, g^M)$, the escort representations of the Riemannian metric g^M and the cubic form C^M are given as follows:*

$$g_{ij}^M(\theta) \quad = \quad \int_\Omega (\partial_i \ln_q p_\theta)(\partial_j \ln_q p_\theta)P_q(x;\theta)dx, \tag{21}$$

$$C_{ijk}^M(\theta) \quad = \quad \int_\Omega (\partial_i \ln_q p_\theta)(\partial_j \ln_q p_\theta)(\partial_k \ln_q p_\theta)\tilde{P}_q(x;\theta)dx. \tag{22}$$

Proof. For the Riemannian metric g^M, since $\partial_i p_\theta = (\partial_i \ln_q p_\theta)P_q(x;\theta)$, we immediately obtain Equation (21) from the definition of g^M.

Let us consider the expression for cubic form (22). The q-score function $\partial_i \ln_q p_\theta$ is unbiased under the q-expectation. In fact,

$$E_{q,p}[\partial_i \ln_q p_\theta] = \int_\Omega (\partial_j \ln_q p_\theta)P_q(x;\theta)dx = \int_\Omega \partial_j p_\theta dx = 0.$$

From Equation (19), we obtain

$$
\begin{aligned}
C_{ijk}^M(\theta) \quad = \quad & \Gamma_{ij,k}^{M(m)}(x;\theta) \\
= \quad & \int_\Omega (\partial_k \ln_q p_\theta)(\partial_i \partial_j p_\theta)dx \\
= \quad & \int_\Omega (\partial_k \ln_q p_\theta)\partial_i \left\{ (\partial_j \ln_q p_\theta) P_q(x;\theta) \right\} dx \\
= \quad & -\partial_{ij}\psi(\theta) \int_\Omega (\partial_k \ln_q p_\theta)P_q(x;\theta)dx + \int_\Omega (\partial_k \ln_q p_\theta)(\partial_j \ln_q p_\theta)\{\partial_i P_q(x;\theta)\}dx \\
= \quad & \int_\Omega (\partial_k \ln_q p_\theta)(\partial_j \ln_q p_\theta)(\partial_i \ln_q p_\theta)\tilde{P}_q(x;\theta)dx.
\end{aligned}
$$

□

We remark that Naudts [5] gave another generalization of Fisher metric g^N, which is defined by

$$g_{ij}^N(\theta) := \int_\Omega \frac{1}{P_q^{esc}(x;\theta)}(\partial_i p_\theta)(\partial_j p_\theta)dx,$$

The metric g^N is conformally equivalent to g^M with conformal factor $Z_q(p_\theta) = \int_\Omega \{p(x;\theta)\}^q dx$. That is, $g^N(\theta) = Z_q(p_\theta) g^M(\theta)$. (See also [6]). Naudts gave a further generalization of Fisher metric and he showed a Cramér–Rao type bound theorem [5].

6. Concluding Remarks

In this paper, we introduced a sequence of escort distributions. Then we gave representations of Riemannian metrics and cubic forms from a viewpoint of the sequence of escort distributions.

In particular, we can define the following $(0,2)$-tensor fields on a q-exponential family. For $p_\theta \in S_q$, set $\eta_i = \partial_i \psi(\theta)$.

(1) From the standard expectation, we obtain

$$
\begin{aligned}
g_{ij}^{(0)}(\theta) \ := \ G_{ij}(\theta) \ := \ & \int_\Omega (\partial_i \ln_q p_\theta)(\partial_j \ln_q p_\theta) p_\theta dx \\
= \ & E_p[(F_i(x) - \eta_i)(F_j(x) - \eta_j)].
\end{aligned}
$$

The tensor G is a covariance matrix. However, G may not be important in anomalous statistics.

(2) From the q-expectation, we obtain

$$
\begin{aligned}
g_{ij}^{(1)}(\theta) \ := \ g_{ij}^M(\theta) \ = \ & \int_\Omega (\partial_i \ln_q p_\theta)(\partial_j \ln_q p_\theta)\{p_\theta\}^q dx \\
= \ & \int_\Omega (\partial_i \ln_q p_\theta)(\partial_j \ln_q p_\theta) P_q(x;\theta) dx \\
= \ & E_{q,p}[(F_i(x) - \eta_i)(F_j(x) - \eta_j)].
\end{aligned}
$$

The Riemannian metric g^M is a Hessian metric, and it is induced from the β-divergence (20).

(3) From the expectation with respect to the second escort distribution, we obtain

$$
\begin{aligned}
g_{ij}^{(2)}(\theta) \ := \ g_{ij}^F(\theta) \ = \ & \int_\Omega (\partial_i \ln_q p_\theta)(\partial_j \ln_q p_\theta)\{p_\theta\}^{2q-1} dx \\
= \ & \frac{1}{q} \int_\Omega (\partial_i \ln_q p_\theta)(\partial_j \ln_q p_\theta) \tilde{P}_q(x;\theta) dx \\
= \ & \frac{1}{q} E_{q,(2),p}[(F_i(x) - \eta_i)(F_j(x) - \eta_j)]. \\
g_{ij}^q(\theta) \ = \ & \frac{Z_q(p)}{q} g_{ij}^F.
\end{aligned}
$$

The Riemannian metric g^F is a Fisher metric. Hence g^F is invariant to the choice of reference measure on Ω, but it is not a Hessian metric. In addition, g^F is induced from the α-divergence (14). The conformal Riemannian metric g^q is a q-Fisher metric. It is a Hessian metric, and it is induced from a normalized Tsallis relative entropy (15).

We may define a Riemannian metric and a cubic form from higher order escort expectations:

$$
\begin{aligned}
g_{ij}^{(n)}(\theta) \ := \ & \int_\Omega (\partial_i \ln_q p_\theta)(\partial_j \ln_q p_\theta) P_{q,(n)}(x;\theta) dx, \\
C_{ij}^{(n)}(\theta) \ := \ & \int_\Omega (\partial_i \ln_q p_\theta)(\partial_j \ln_q p_\theta)(\partial_k \ln_q p_\theta) P_{q,(n+1)}(x;\theta) dx.
\end{aligned}
$$

Then we obtain a sequence of statistical manifold structures.

$$
(S_q, g^{(1)}, C^{(1)}) \ \rightarrow \ (S_q, g^{(2)}, C^{(2)}) \ \rightarrow \ \cdots \ \rightarrow \ (S_q, g^{(n)}, C^{(n)}) \ \rightarrow \ \cdots
$$

However, the geometric meaning of this sequence is not clear at this moment. Elucidating geometric properties of this sequence is a future problem.

Conflicts of Interest: The author declares no conflict of interest.

References

1. Tsallis, C. *Introduction to Nonextensive Statistical Mechanics: Approaching a Complex World*; Springer: New York, NY, USA, 2009.
2. Naudts, J. *Generalised Thermostatistics*; Springer: London, UK, 2011.
3. Kaniadakis, G. Theoretical foundations and mathematical formalism of the power-law tailed statistical distributions. *Entropy* **2013**, *15*, 3983–4010.
4. Beck, C.; Schlögl, F. *Thermodynamics of Chaotic Systems: An Introduction*; Cambridge University Press: Cambridge, UK, 1993.
5. Naudts, J. Estimators, escort probabilities, and ϕ-exponential families in statistical physics. *J. Inequal. Pure Appl. Math.* **2004**, *5*, 102.
6. Matsuzoe, H.; Henmi, M. Hessian structures and divergence functions on deformed exponential families. In *Geometric Theory of Information, Signals and Communication Technology*; Nielsen, F., Ed.; Springer: Basel, Switzerland, 2014; pp. 57–80.
7. Sakamoto, M.; Matsuzoe, H. A generalization of independence and multivariate Student's *t*-distributions. In *Geometric Science of Information*, Proceedings of Second International Conference on Geometric Science of Information (GSI 2015), Palaiseau, France, 28–30 October 2015; Volume 9389, pp. 740–749.
8. Matsuzoe, H.; Wada, T. Deformed algebras and generalizations of independence on deformed exponential families. *Entropy* **2015**, *17*, 5729–5751.
9. Wada, T.; Matsuzoe, H.; Scarfone, A.M. Dualistic hessian structures among the thermodynamic potentials in the κ-Thermostatistics. *Entropy* **2015**, *17*, 7213–7229.
10. 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 Geomerty and Its Related Fields, Veliko Tarnovo, Bulgaria, 3–7 September 2012*; Adachi, T., Hashimoto, H., Hristov, M.J., Eds.; World Scientific: Hackensack, NJ, USA, 2013; pp. 187–202.
11. 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*; Springer: Berlin/Heidelberg, Germany, 2015; Volume 8085, pp. 275–282.
12. Eguchi, S.; Komori, O. Path connectedness on a space of probability density functions. In *Geometric Science of Information, Proceedings of Second International Conference on Geometric Science of Information (GSI 2015), Palaiseau, France, 28–30 October 2015*; Springer: Berlin/Heidelberg, Germany, 2015; Volume 9389, pp. 615–624.
13. Scarfone, A.M.; Matsuzoe, H.; Wada, T. Consistency of the structure of Legendre transform in thermodynamics with the Kolmogorov-Nagumo average. *Phys. Lett. A* **2016**, *380*, 3022–3028.
14. Tsallis, C. What are the numbers that experiments provide? *Quim. Nova* **1994**, *17*, 468–471.
15. Zhang, J. On monotone embedding in information geometry. *Entropy* **2015**, *17*, 4485–4499.
16. Tanaka M. Meaning of an escort distribution and τ-transformation. *J. Phys. Conf. Ser.* **2010**, *201*, 012007.
17. Murata, N.; Takenouchi, T.; Kanamori, T.; Eguchi, S. Information geometry of U-boost and Bregman divergence. *Neural Comput.* **2004**, *16*, 1437–1481.
18. Kurose, T. On the divergences of 1-conformally flat statistical manifolds. *Tôhoku Math. J.* **1994**, *46*, 427–433.
19. Matsuzoe, H. Statistical manifolds and affine differential geometry. *Adv. Stud. Pure Math.* **2010**, *57*, 303–321.
20. Lauritzen, S.L. Statistical manifolds. In *Differential Geometry in Statistical Inferences*; Gupta, S.S., Ed.; IMS Lecture Notes Monograph Series 10; Institute of Mathematical Statistics: Hayward, CA, USA, 1987; pp. 96–163.
21. Amari, S.; Nagaoka, H. *Method of Information Geometry*; Translations of Mathematical Monographs; American Mathematical Society: Providence, RI, USA; Oxford University Press: Oxford, UK, 2000.
22. Amari, S. *Information Geometry and Its Applications*; Springer: Tokyo, Japan, 2016.
23. Amari, S.; Ohara, A.; Matsuzoe, H. Geometry of deformed exponential families: Invariant, dually-flat and conformal geometry. *Physica A* **2012**, *391*, 4308–4319.

24. 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 Geometry and Its Related Fields, Veliko Tarnovo, Bulgaria, 6–10 September 2010*; Adachi, T., Hashimoto, H., Hristov, M.J., Eds.; World Scientific: Hackensack, NJ, USA, 2011; pp. 55–71.
25. Shima, H. *The Geometry of Hessian Structures*; World Scientific: Hackensack, NJ, USA, 2007.
26. Ohara, A.; Wada, T. Information geometry of *q*-Gaussian densities and behaviors of solutions to related diffusion equations. *J. Phys. A* **2010**, *43*, 035002.
27. 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.