

Article

# Geometry Induced by a Generalization of Rényi Divergence

David C. de Souza <sup>1,†</sup>, Rui F. Vigelis <sup>2,\*,†</sup> and Charles C. Cavalcante <sup>3,†</sup>

<sup>1</sup> Instituto Federal do Ceará, Campus Maracanaú, Fortaleza 61939-140, Brazil; davidcs@ifce.edu.br

<sup>2</sup> Computer Engineering School, Campus Sobral, Federal University of Ceará, Sobral 62010-560, Brazil

<sup>3</sup> Department of Teleinformatics Engineering, Federal University of Ceará, Fortaleza 60455-900, Brazil; charles@ufc.br

\* Correspondence: rfvigelis@ufc.br

† These authors contributed equally to this work.

Academic Editors: Frédéric Barbaresco and Frank Nielsen

Received: 6 September 2016; Accepted: 11 November 2016; Published: 17 November 2016

**Abstract:** In this paper, we propose a generalization of Rényi divergence, and then we investigate its induced geometry. This generalization is given in terms of a  $\varphi$ -function, the same function that is used in the definition of non-parametric  $\varphi$ -families. The properties of  $\varphi$ -functions proved to be crucial in the generalization of Rényi divergence. Assuming appropriate conditions, we verify that the generalized Rényi divergence reduces, in a limiting case, to the  $\varphi$ -divergence. In generalized statistical manifold, the  $\varphi$ -divergence induces a pair of dual connections  $D^{(-1)}$  and  $D^{(1)}$ . We show that the family of connections  $D^{(\alpha)}$  induced by the generalization of Rényi divergence satisfies the relation  $D^{(\alpha)} = \frac{1-\alpha}{2}D^{(-1)} + \frac{1+\alpha}{2}D^{(1)}$ , with  $\alpha \in [-1, 1]$ .

**Keywords:** Rényi divergence;  $\varphi$ -function;  $\varphi$ -divergence;  $\varphi$ -family; statistical manifold; information geometry

## 1. Introduction

Information geometry, the study of statistical models equipped with a differentiable structure, was pioneered by the work of Rao [1], and gained maturity with the work of Amari and many others [2–4]. It has been successfully applied in many different areas, such as statistical inference, machine learning, signal processing or optimization [4,5]. In appropriate statistical models, the differentiable structure is induced by a (statistical) divergence. The Kullback–Leibler divergence induces a Riemannian metric, called the Fisher–Rao metric, and a pair of dual connections, the exponential and mixture connections. A statistical model endowed with the Fisher–Rao metric is called a (classical) statistical manifold. Amari also considered a family of  $\alpha$ -divergences that induce a family of  $\alpha$ -connections.

Much research in recent years has focused on the geometry of non-standard statistical models [6–8]. These models are defined in terms of a deformed exponential (also called  $\phi$ -exponential). In particular,  $\kappa$ -exponential models and  $q$ -exponential families are investigated in [9,10]. Non-parametric (or infinite-dimensional)  $\varphi$ -families were introduced by the authors in [11,12], which generalize exponential families in the non-parametric setting [13–16]. Based on the similarity between exponential and  $\varphi$ -families, we defined the so-called  $\varphi$ -divergence, with respect to which the Kullback–Leibler divergence is a particular case. Statistical models equipped with a geometric structure induced by  $\varphi$ -divergences, which are called generalized statistical manifolds, are investigated in [17,18]. With respect to these connections, parametric  $\varphi$ -families are dually flat.

The  $\varphi$ -divergence is intrinsically related to the  $(\rho, \tau)$ -model of Zhang, which was proposed in [19,20], extended to the infinite-dimension setting in [21], and explained in more details in [22,23].

For instance, the metric induced by  $\varphi$ -divergence and the  $(\rho, \tau)$ -generalization of the Fisher–Rao metric, for the choices  $\rho = \varphi^{-1}$  and  $f = \rho^{-1}$ , differ by a conformal factor.

Among many attempts to generalize Kullback–Leibler divergence, Rényi divergence [24] is one of the most successful, having found many applications [25]. In the present paper, we propose a generalization of Rényi divergence, which we use to define a family of  $\alpha$ -connections. This generalization is based on an interpretation of Rényi divergence as a kind of normalizing function. To generalize Rényi divergence, we considered functions satisfying some suitable conditions. To a function for which these conditions hold, we give the name of  $\varphi$ -function. In a limiting case, the generalized Rényi divergence reduces to the  $\varphi$ -divergence. In [17,18], the  $\varphi$ -divergence gives rise to a pair of dual connections  $D^{(-1)}$  and  $D^{(1)}$ . We show that the connection  $D^{(\alpha)}$  induced by the generalization of Rényi divergence satisfies the convex combination  $D^{(\alpha)} = \frac{1-\alpha}{2}D^{(-1)} + \frac{1+\alpha}{2}D^{(1)}$ .

Eguchi in [26] investigated a geometry based on a normalizing function similar to the one used in the generalization of Rényi divergence. In [26], results were derived supposing that this normalizing function exists; conditions for its existence were not given. In the present paper, the existence of the normalizing function is ensured by conditions involved in the definition of  $\varphi$ -functions.

The rest of the paper is organized as follows. In Section 2,  $\varphi$ -functions are introduced and some properties are discussed. The Rényi divergence is generalized in Section 3. We investigate in Section 4 the geometry induced by the generalization of Rényi divergence. Section 4.2 provides evidence of the role of the generalized Rényi divergence in  $\varphi$ -families.

## 2. $\varphi$ -Functions

Rényi divergence is defined in terms of the exponential function (to be more precise, the logarithm). A way of generalizing Rényi divergence is to replace the exponential function by another function, which satisfies some suitable conditions. To a function for which these conditions hold, we give the name  $\varphi$ -function. In this section, we define and investigate some properties of  $\varphi$ -functions.

Let  $(T, \Sigma, \mu)$  be a measure space. Although we do not restrict our analysis to a particular measure space, the reader can think of  $T$  as the set of real numbers  $\mathbb{R}$ ,  $\Sigma$  as the Borel  $\sigma$ -algebra on  $\mathbb{R}$ , and  $\mu$  as the Lebesgue measure. We can also consider  $T$  to be a discrete set, a case in which  $\mu$  is the counting measure.

We say that  $\varphi: \mathbb{R} \rightarrow (0, \infty)$  is a  $\varphi$ -function if the following conditions are satisfied:

- (a1)  $\varphi(\cdot)$  is convex;
- (a2)  $\lim_{u \rightarrow -\infty} \varphi(u) = 0$  and  $\lim_{u \rightarrow \infty} \varphi(u) = \infty$ ;
- (a3) there exists a measurable function  $u_0: T \rightarrow (0, \infty)$  such that

$$\int_T \varphi(c + \lambda u_0) d\mu < \infty, \quad \text{for all } \lambda > 0, \quad (1)$$

for each measurable function  $c: T \rightarrow \mathbb{R}$  satisfying  $\int_T \varphi(c) d\mu = 1$ .

Thanks to condition (a3), we can generalize Rényi divergence using  $\varphi$ -functions. These conditions appeared first at [12] where the authors constructed non-parametric  $\varphi$ -families of probability distributions. We remark that if  $T$  is finite, condition (a3) is always satisfied.

Examples of functions  $\varphi: \mathbb{R} \rightarrow (0, \infty)$  satisfying (a1)–(a3) abound. An example of great relevance is the exponential function  $\varphi(u) = \exp(u)$ , which satisfies conditions (a1)–(a3) with  $u_0 = \mathbf{1}_T$ . Another example of  $\varphi$ -function is the Kaniadakis'  $\kappa$ -exponential [12,27,28].

**Example 1.** The Kaniadakis'  $\kappa$ -exponential  $\exp_\kappa: \mathbb{R} \rightarrow (0, \infty)$  for  $\kappa \in [-1, 1]$  is defined as

$$\exp_\kappa(u) = \begin{cases} (\kappa u + \sqrt{1 + \kappa^2 u^2})^{1/\kappa}, & \text{if } \kappa \neq 0, \\ \exp(u), & \text{if } \kappa = 0, \end{cases}$$

whose inverse is the so called the Kaniadakis'  $\kappa$ -logarithm  $\log_\kappa: (0, \infty) \rightarrow \mathbb{R}$ , which is given by

$$\log_\kappa(u) = \begin{cases} \frac{u^\kappa - u^{-\kappa}}{2\kappa}, & \text{if } \kappa \neq 0, \\ \ln(u), & \text{if } \kappa = 0. \end{cases}$$

It is clear that  $\exp_\kappa(\cdot)$  satisfies (a1) and (a2). Let  $u_0: T \rightarrow (0, \infty)$  be any measurable function for which  $\int_T \exp_\kappa(u_0)d\mu < \infty$ . We will show that  $u_0$  satisfies expression (1). For any  $u \in \mathbb{R}$  and  $\alpha \geq 1$ , we can write

$$\begin{aligned} \exp_\kappa(\alpha u) &= \alpha^{1/|\kappa|} (|\kappa|u + \sqrt{1/\alpha^2 + |\kappa|^2 u^2})^{1/|\kappa|} \\ &\leq \alpha^{1/|\kappa|} (|\kappa|u + \sqrt{1 + |\kappa|^2 u^2})^{1/|\kappa|} \\ &= \alpha^{1/|\kappa|} \exp_\kappa(u), \end{aligned}$$

where we used that  $\exp_\kappa(\cdot) = \exp_{-\kappa}(\cdot)$ . Then, we conclude that  $\int_T \exp_\kappa(\alpha u_0)d\mu < \infty$  for all  $\alpha \geq 0$ . Fix any measurable function  $c: T \rightarrow \mathbb{R}$  such that  $\int_T \varphi(c)d\mu = 1$ . For each  $\lambda > 0$ , we have

$$\begin{aligned} \int_T \exp_\kappa(c + \lambda u_0)d\mu &\leq \frac{1}{2} \int_T \exp_\kappa(2c)d\mu + \frac{1}{2} \int_T \exp_\kappa(2\lambda u_0)d\mu \\ &\leq 2^{1/|\kappa|-1} \int_T \exp_\kappa(c)d\mu + 2^{1/|\kappa|-1} \int_T \exp_\kappa(\lambda u_0)d\mu \\ &< \infty, \end{aligned}$$

which shows that  $\exp_\kappa(\cdot)$  satisfies (a3). Therefore, the Kaniadakis'  $\kappa$ -exponential  $\exp_\kappa(\cdot)$  is an example of  $\varphi$ -function.

The restriction that  $\int_T \varphi(c)d\mu = 1$  can be weakened, as asserted in the next result.

**Lemma 1.** Let  $\tilde{c}: T \rightarrow \mathbb{R}$  be any measurable function such that  $\int_T \varphi(\tilde{c})d\mu < \infty$ . Then,  $\int_T \varphi(\tilde{c} + \lambda u_0)d\mu < \infty$  for all  $\lambda > 0$ .

**Proof.** Notice that if  $\int_T \varphi(\tilde{c})d\mu \geq 1$ , then  $\int_T \varphi(\tilde{c} - \alpha u_0)d\mu = 1$  for some  $\alpha > 1$ . From the definition of  $u_0$ , it follows that  $\int_T \varphi(\tilde{c} + \lambda u_0)d\mu = \int_T \varphi(c + (\alpha + \lambda)u_0)d\mu < \infty$ , where  $c = \tilde{c} - \alpha u_0$ . Now assume that  $\int_T \varphi(\tilde{c})d\mu < 1$ . Consider any measurable set  $A \subseteq T$  with measure  $0 < \mu(A) < \mu(T)$ . Let  $u: T \rightarrow [0, \infty)$  be a measurable function supported on  $A$  satisfying  $\varphi(\tilde{c} + u)\mathbf{1}_A = [\varphi(\tilde{c}) + \alpha]\mathbf{1}_A$ , where  $\alpha = (1 - \int_T \varphi(\tilde{c})d\mu) / \mu(A)$ . Defining  $c = (\tilde{c} + u)\mathbf{1}_A + \tilde{c}\mathbf{1}_{T \setminus A}$ , we see that  $\int_T \varphi(c)d\mu = 1$ . By the definition of  $u_0$ , we can write

$$\int_T \varphi(\tilde{c} + \lambda u_0)d\mu \leq \int_T \varphi(c + \lambda u_0)d\mu < \infty, \quad \text{for any } \lambda > 0,$$

which is the desired result.  $\square$

As a consequence of Lemma 1, condition (a3) can be replaced by the following one:

(a3') There exists a measurable function  $u_0: T \rightarrow (0, \infty)$  such that

$$\int_T \varphi(c + \lambda u_0)d\mu < \infty, \quad \text{for all } \lambda > 0, \tag{2}$$

for each measurable function  $c: T \rightarrow \mathbb{R}$  for which  $\int_T \varphi(c)d\mu < \infty$ .

Without the equivalence between conditions (a3) and (a3'), we could not generalize Rényi divergence in the manner we propose. In fact,  $\varphi$ -functions could be defined directly in terms of (a3'), without mentioning (a3). We chose to begin with (a3) because this condition appeared initially in [12].

Not all functions  $\varphi: \mathbb{R} \rightarrow (0, \infty)$ , for which conditions (a1) and (a2) hold, satisfy condition (a3). Such a function is given below.

**Example 2.** Assume that the underlying measure  $\mu$  is  $\sigma$ -finite and non-atomic. This is the case of the Lebesgue measure. Let us consider the function

$$\varphi(u) = \begin{cases} e^{(u+1)^2/2}, & u \geq 0, \\ e^{(u+1/2)}, & u \leq 0, \end{cases} \tag{3}$$

which clearly is convex, and satisfies the limits  $\lim_{u \rightarrow -\infty} \varphi(u) = 0$  and  $\lim_{u \rightarrow \infty} \varphi(u) = \infty$ . Given any measurable function  $u_0: T \rightarrow (0, \infty)$ , we will find a measurable function  $c: T \rightarrow \mathbb{R}$  with  $\int_T \varphi(c) d\mu < \infty$ , for which expression (2) is not satisfied.

For each  $m \geq 1$ , we define

$$v_m(t) := \left( m \frac{\log(2)}{u_0(t)} - \frac{u_0(t)}{2} - 1 \right) \mathbf{1}_{E_m}(t),$$

where  $E_m = \{t \in T : m \frac{\log(2)}{u_0(t)} - \frac{u_0(t)}{2} - 1 > 0\}$ . Because  $v_m \uparrow \infty$ , we can find a sub-sequence  $\{v_{m_n}\}$  such that

$$\int_{E_{m_n}} e^{(v_{m_n}+u_0+1)^2/2} d\mu \geq 2^n.$$

According to (Lemma 8.3 in [29]), there exists a sub-sequence  $w_k = v_{m_{n_k}}$  and pairwise disjoint sets  $A_k \subseteq E_{m_{n_k}}$  for which

$$\int_{A_k} e^{(w_k+u_0+1)^2/2} d\mu = 1.$$

Let us define  $c = \bar{c} \mathbf{1}_{T \setminus A} + \sum_{k=1}^{\infty} w_k \mathbf{1}_{A_k}$ , where  $A = \bigcup_{k=1}^{\infty} A_k$  and  $\bar{c}$  is any measurable function such that  $\varphi(\bar{c}(t)) > 0$  for  $t \in T \setminus A$  and  $\int_{T \setminus A} \varphi(\bar{c}) d\mu < \infty$ . Observing that

$$e^{(w_k(t)+u_0(t)+1)^2/2} = 2^{m_{n_k}} e^{(w_k(t)+1)^2/2}, \quad \text{for } t \in A_k,$$

we get

$$\int_{A_k} e^{(w_k+1)^2/2} d\mu = \frac{1}{2^{m_{n_k}}}, \quad \text{for every } m \geq 1.$$

Then, we can write

$$\begin{aligned} \int_T \varphi(c) d\mu &= \int_{T \setminus A} \varphi(\bar{c}) d\mu + \sum_{k=1}^{\infty} \int_{A_k} e^{(w_k+1)^2/2} d\mu \\ &= \int_{T \setminus A} \varphi(\bar{c}) d\mu + \sum_{k=1}^{\infty} \frac{1}{2^{m_{n_k}}} < \infty. \end{aligned}$$

On the other hand,

$$\begin{aligned} \int_T \varphi(c + u_0) d\mu &= \int_{T \setminus A} \varphi(\bar{c}) d\mu + \sum_{k=1}^{\infty} \int_{A_k} e^{(u_0+w_k+1)^2/2} d\mu \\ &= \int_{T \setminus A} \varphi(\bar{c}) d\mu + \sum_{k=1}^{\infty} 1 = \infty, \end{aligned}$$

which shows that (2) is not satisfied.

### 3. Generalization of Rényi Divergence

In this section, we provide a generalization of Rényi divergence, which is given in terms of a  $\varphi$ -function. This generalization also depends on a parameter  $\alpha \in [-1, 1]$ ; for  $\alpha = \pm 1$ , it is defined as a limit. Supposing that the underlying  $\varphi$ -function is continuously differentiable, we show that this limit exists and results in the  $\varphi$ -divergence [12]. In what follows, all probability distributions are assumed to have positive density. In other words, they belong to the collection

$$\mathcal{P}_\mu = \left\{ p \in L^0 : \int_T p d\mu = 1 \text{ and } p > 0 \right\},$$

where  $L^0$  is the space of all real-valued, measurable functions on  $T$ , with equality  $\mu$ -a.e. ( $\mu$ -almost everywhere).

The Rényi divergence of order  $\alpha \in (-1, 1)$  between two probability distributions  $p$  and  $q$  in  $\mathcal{P}_\mu$  is defined as

$$\mathcal{D}_R^{(\alpha)}(p \parallel q) = \frac{4}{\alpha^2 - 1} \log \left( \int_T p^{\frac{1-\alpha}{2}} q^{\frac{1+\alpha}{2}} d\mu \right). \quad (4)$$

For  $\alpha = \pm 1$ , the Rényi divergence is defined by taking a limit:

$$\mathcal{D}_R^{(-1)}(p \parallel q) = \lim_{\alpha \downarrow -1} \mathcal{D}_R^{(\alpha)}(p \parallel q), \quad (5)$$

$$\mathcal{D}_R^{(1)}(p \parallel q) = \lim_{\alpha \uparrow 1} \mathcal{D}_R^{(\alpha)}(p \parallel q). \quad (6)$$

Under some conditions, the limits in (5) and (6) are finite-valued, and converge to the Kullback–Leibler divergence. In other words,

$$\mathcal{D}_R^{(-1)}(p \parallel q) = \mathcal{D}_R^{(1)}(q \parallel p) = \mathcal{D}_{\text{KL}}(p \parallel q) < \infty,$$

where  $\mathcal{D}_{\text{KL}}(p \parallel q)$  denotes the Kullback–Leibler divergence between  $p$  and  $q$ , which is given by

$$\mathcal{D}_{\text{KL}}(p \parallel q) = \int_T p \log \left( \frac{p}{q} \right) d\mu.$$

These conditions are stated in Proposition 1, given in the end of this section, for the case involving the generalized Rényi divergence.

The Rényi divergence in its standard form is given by

$$\mathcal{D}^{(\alpha)}(p \parallel q) = \frac{1}{1-\alpha} \log \left( \int_T p^\alpha q^{1-\alpha} d\mu \right), \quad \text{for } \alpha \in (0, 1). \quad (7)$$

Expression (4) is related to this form by

$$\mathcal{D}_R^{(\alpha)}(p \parallel q) = \frac{2}{1-\alpha} \mathcal{D}^{((1-\alpha)/2)}(p \parallel q).$$

Beyond the change of variables, which results in  $\alpha$  ranging in  $[-1, 1]$ , expressions (4) and (7) differ by the factor  $2/(1-\alpha)$ . We opted to insert the term  $2/(1-\alpha)$  so that some kind of symmetry could be maintained when the limits  $\alpha \downarrow -1$  and  $\alpha \uparrow 1$  are considered. In addition, the geometry induced by the version (4) conforms with Amari's notation [5].

The Rényi divergence  $\mathcal{D}_R^{(\alpha)}(\cdot \parallel \cdot)$  can be defined for every  $\alpha \in \mathbb{R}$ . However, for  $\alpha \notin (-1, 1)$ , the expression (4) may not be finite-valued for every  $p$  and  $q$  in  $\mathcal{P}_\mu$ . To avoid some technicalities, we just consider  $\alpha \in [-1, 1]$ .

Given  $p$  and  $q$  in  $\mathcal{P}_\mu$ , let us define

$$\kappa(\alpha) = -\log\left(\int_T p^{\frac{1-\alpha}{2}} q^{\frac{1+\alpha}{2}} d\mu\right), \quad \text{for } \alpha \in [-1, 1],$$

which can be used to express the Rényi divergence as

$$\mathcal{D}_R^{(\alpha)}(p \parallel q) = \frac{4}{1-\alpha^2} \kappa(\alpha), \quad \text{for } \alpha \in (-1, 1).$$

The function  $\kappa(\alpha)$ , which depends on  $p$  and  $q$ , can be defined as the unique non-negative real number for which

$$\int_T \exp\left(\frac{1-\alpha}{2} \ln(p) + \frac{1+\alpha}{2} \ln(q) + \kappa(\alpha)\right) d\mu = 1. \tag{8}$$

The function  $\kappa(\alpha)$  makes the role of a normalizing term. The generalization of Rényi divergence, which we propose, is based on the interpretation of  $\kappa(\alpha)$  given in (8). Instead of the exponential function, we consider a  $\varphi$ -function in (8).

Fix any  $\varphi$ -function  $\varphi: \mathbb{R} \rightarrow (0, \infty)$ . Given any  $p$  and  $q$  in  $\mathcal{P}_\mu$ , we take  $\kappa(\alpha) = \kappa(\alpha; p, q) \geq 0$  so that

$$\int_T \varphi\left(\frac{1-\alpha}{2} \varphi^{-1}(p) + \frac{1+\alpha}{2} \varphi^{-1}(q) + \kappa(\alpha) u_0\right) d\mu = 1, \tag{9}$$

or, in other words, the term inside the integral is a probability distribution in  $\mathcal{P}_\mu$ . The existence and uniqueness of  $\kappa(\alpha)$  as defined in (9) is guaranteed by condition (a3').

We define a generalization of the Rényi divergence of order  $\alpha \in (-1, 1)$  as

$$\mathcal{D}_\varphi^{(\alpha)}(p \parallel q) = \frac{4}{1-\alpha^2} \kappa(\alpha). \tag{10}$$

For  $\alpha = \pm 1$ , this generalization is defined as a limit:

$$\mathcal{D}_\varphi^{(-1)}(p \parallel q) = \lim_{\alpha \downarrow -1} \mathcal{D}_\varphi^{(\alpha)}(p \parallel q), \tag{11}$$

$$\mathcal{D}_\varphi^{(1)}(p \parallel q) = \lim_{\alpha \uparrow 1} \mathcal{D}_\varphi^{(\alpha)}(p \parallel q). \tag{12}$$

The cases  $\alpha = \pm 1$  are related to a generalization of the Kullback–Leibler divergence, the so-called  $\varphi$ -divergence, which was introduced by the authors in [12]. The  $\varphi$ -divergence is given by (It was pointed out to us by an anonymous referee that this form of divergence is a special case of the  $(\rho, \tau)$ -divergence for  $\rho = \varphi^{-1}$  and  $f = \rho^{-1}$  (see Section 3.5 in [19]) apart from a conformal factor, which is the denominator of (13)):

$$\mathcal{D}_\varphi(p \parallel q) = \frac{\int_T \frac{\varphi^{-1}(p) - \varphi^{-1}(q)}{(\varphi^{-1})'(p)} d\mu}{\int_T \frac{u_0}{(\varphi^{-1})'(p)} d\mu}. \tag{13}$$

Under some conditions, the limit in (11) or (12) is finite-valued and converges to the  $\varphi$ -divergence:

$$\mathcal{D}_\varphi^{(-1)}(p \parallel q) = \mathcal{D}_\varphi^{(1)}(q \parallel p) = \mathcal{D}_\varphi(p \parallel q) < \infty. \tag{14}$$

To show (14), we make use of the following result.

**Lemma 2.** Assume that  $\varphi(\cdot)$  is continuously differentiable. If for  $\alpha_0, \alpha_1 \in \mathbb{R}$ , the expression

$$\int_T \varphi\left(\frac{1-\alpha}{2} \varphi^{-1}(p) + \frac{1+\alpha}{2} \varphi^{-1}(q)\right) d\mu < \infty \tag{15}$$

is satisfied for all  $\alpha \in [\alpha_0, \alpha_1]$ , then the derivative of  $\kappa(\alpha)$  exists at any  $\alpha \in (\alpha_0, \alpha_1)$ , and is given by

$$\frac{\partial \kappa}{\partial \alpha}(\alpha) = -\frac{1}{2} \frac{\int_T [\varphi^{-1}(q) - \varphi^{-1}(p)] \varphi'(c_\alpha) d\mu}{\int_T \varphi'(c_\alpha) u_0 d\mu}, \tag{16}$$

where  $c_\alpha = \frac{1-\alpha}{2} \varphi^{-1}(p) + \frac{1+\alpha}{2} \varphi^{-1}(q) + \kappa(\alpha) u_0$ .

**Proof.** For  $\alpha \in (\alpha_0, \alpha_1)$  and  $\kappa > 0$ , define

$$g(\alpha, \kappa) = \int_T \varphi\left(\frac{1-\alpha}{2} \varphi^{-1}(p) + \frac{1+\alpha}{2} \varphi^{-1}(q) + \kappa u_0\right) d\mu.$$

The function  $\kappa(\alpha)$  is defined implicitly by  $g(\alpha, \kappa(\alpha)) = 1$ . If we show that

- (i) the function  $g(\alpha, \kappa)$  is continuous in a neighborhood of  $(\alpha, \kappa(\alpha))$ ,
- (ii) the partial derivatives  $\frac{\partial g}{\partial \alpha}$  and  $\frac{\partial g}{\partial \kappa}$  exist and are continuous at  $(\alpha, \kappa(\alpha))$ ,
- (iii) and  $\frac{\partial g}{\partial \kappa}(\alpha, \kappa(\alpha)) > 0$ ,

then by the Implicit Function Theorem  $\kappa(\alpha)$  is differentiable at  $\alpha \in (\alpha_0, \alpha_1)$ , and

$$\frac{\partial \kappa}{\partial \alpha}(\alpha) = -\frac{(\partial g / \partial \alpha)(\alpha, \kappa(\alpha))}{(\partial g / \partial \kappa)(\alpha, \kappa(\alpha))}. \tag{17}$$

We begin by verifying that  $g(\alpha, \kappa)$  is continuous. For fixed  $\alpha \in (\alpha_0, \alpha_1)$  and  $\kappa > 0$ , set  $\kappa_0 = 2\kappa$ . Denoting  $A = \{t \in T : \varphi^{-1}(q(t)) > \varphi^{-1}(p(t))\}$ , we can write

$$\begin{aligned} \varphi\left(\frac{1-\beta}{2} \varphi^{-1}(p) + \frac{1+\beta}{2} \varphi^{-1}(q) + \lambda u_0\right) &\leq \varphi\left(\varphi^{-1}(p) + \frac{1+\beta}{2} [\varphi^{-1}(q) - \varphi^{-1}(p)] + \kappa_0 u_0\right) \\ &\leq \varphi\left(\varphi^{-1}(p) + \frac{1+\alpha_1}{2} [\varphi^{-1}(q) - \varphi^{-1}(p)] + \kappa_0 u_0\right) \mathbf{1}_A \\ &\quad + \varphi\left(\varphi^{-1}(p) + \frac{1+\alpha_0}{2} [\varphi^{-1}(q) - \varphi^{-1}(p)] + \kappa_0 u_0\right) \mathbf{1}_{T \setminus A}, \end{aligned} \tag{18}$$

for every  $\beta \in (\alpha_0, \alpha_1)$  and  $\lambda \in (0, \kappa_0)$ . Because the function on the right-hand side of (18) is integrable, we can apply the Dominated Convergence Theorem to conclude that

$$\lim_{(\beta, \lambda) \rightarrow (\alpha, \kappa)} g(\beta, \lambda) = g(\alpha, \kappa).$$

Now, we will show that the derivative of  $g(\alpha, \kappa)$  with respect to  $\alpha$  exists and is continuous. Consider the difference

$$\frac{g(\gamma, \lambda) - g(\beta, \lambda)}{\gamma - \beta} = \int_T \frac{1}{\gamma - \beta} \left[ \varphi\left(c_\beta + \frac{\gamma - \beta}{2} [\varphi^{-1}(q) - \varphi^{-1}(p)] + \lambda u_0\right) - \varphi(c_\beta + \lambda u_0) \right] d\mu, \tag{19}$$

where  $c_\beta = \frac{1-\beta}{2} \varphi^{-1}(p) + \frac{1+\beta}{2} \varphi^{-1}(q)$ . Represent by  $f_{\beta, \gamma, \lambda}$  the function inside the integral sign in (19). For fixed  $\alpha \in (\alpha_0, \alpha_1)$  and  $\kappa > 0$ , denote  $\bar{\alpha}_0 = (\alpha_0 + \alpha)/2$ ,  $\bar{\alpha}_1 = (\alpha + \alpha_1)/2$ , and  $\kappa_0 = 2\kappa$ . Because  $\varphi(\cdot)$  is convex and increasing, it follows that

$$|f_{\beta, \gamma, \lambda}| \leq f_{\bar{\alpha}_1, \alpha_1, \kappa_0} \mathbf{1}_A - f_{\bar{\alpha}_0, \alpha_0, \kappa_0} \mathbf{1}_{T \setminus A} =: f, \quad \text{for all } \beta, \gamma \in (\bar{\alpha}_0, \bar{\alpha}_1) \text{ and } \lambda \in (0, \kappa_0),$$

where  $A = \{t \in T : \varphi^{-1}(q(t)) > \varphi^{-1}(p(t))\}$ . Observing that  $f$  is integrable, we can use the Dominated Convergence Theorem to get

$$\lim_{\gamma \rightarrow \beta} \int_T f_{\beta, \gamma, \lambda} d\mu = \int_T (\lim_{\gamma \rightarrow \beta} f_{\beta, \gamma, \lambda}) d\mu,$$

and then

$$\frac{\partial g}{\partial \alpha}(\beta, \lambda) = \frac{1}{2} \int_T [\varphi^{-1}(q) - \varphi^{-1}(p)] \varphi'(c_\beta + \lambda u_0) d\mu. \tag{20}$$

For  $\beta \in (\bar{\alpha}_0, \bar{\alpha}_1)$  and  $\lambda \in (0, \kappa_0)$ , the function inside the integral sign in (20) is dominated by  $f$ . As a result, a second use of the Dominated Convergence Theorem shows that  $\frac{\partial g}{\partial \alpha}$  is continuous at  $(\alpha, \kappa)$ :

$$\lim_{(\beta, \lambda) \rightarrow (\alpha, \kappa)} \frac{\partial g}{\partial \alpha}(\beta, \lambda) = \frac{\partial g}{\partial \alpha}(\alpha, \kappa).$$

Using similar arguments, one can show that  $\frac{\partial g}{\partial \kappa}(\alpha, \kappa)$  exists and is continuous at any  $\alpha \in (\alpha_0, \alpha_1)$  and  $\kappa > 0$ , and is given by

$$\frac{\partial g}{\partial \kappa}(\alpha, \kappa) = \int_T u_0 \varphi'(c_\alpha + \kappa u_0) d\mu. \tag{21}$$

Clearly, expression (21) implies that  $\frac{\partial g}{\partial \kappa}(\alpha, \kappa) > 0$  for all  $\alpha \in (0, \alpha_0)$  and  $\kappa > 0$ .

We proved that items (i)–(iii) are satisfied. As consequence, the derivative of  $\kappa(\alpha)$  exists at any  $\alpha \in (\alpha_0, \alpha_1)$ . Expression (16) for the derivative of  $\kappa(\alpha)$  follows from (17), (20) and (21).  $\square$

As an immediate consequence of Lemma 2, we get the proposition below.

**Proposition 1.** Assume that  $\varphi(\cdot)$  is continuously differentiable.

(a) If, for some  $\alpha_0 < -1$ , expression (15) is satisfied for all  $\alpha \in [\alpha_0, -1)$ , then

$$\mathcal{D}_\varphi^{(-1)}(p \parallel q) = \lim_{\alpha \downarrow -1} \mathcal{D}_\varphi^{(\alpha)}(p \parallel q) = 2 \frac{\partial \kappa}{\partial \alpha}(-1) = \mathcal{D}_\varphi(p \parallel q) < \infty.$$

(b) If, for some  $\alpha_1 > 1$ , expression (15) is satisfied for all  $\alpha \in (1, \alpha_1]$ , then

$$\mathcal{D}_\varphi^{(1)}(p \parallel q) = \lim_{\alpha \uparrow 1} \mathcal{D}_\varphi^{(\alpha)}(p \parallel q) = -2 \frac{\partial \kappa}{\partial \alpha}(1) = \mathcal{D}_\varphi(q \parallel p) < \infty.$$

#### 4. Generalized Statistical Manifolds

Statistical manifolds consist of a collection of probability distributions endowed with a metric and  $\alpha$ -connections, which are defined in terms of the derivative of  $l(t; \theta) = \log p(t; \theta)$ . In a generalized statistical manifold, the metric and connection are defined in terms of  $f(t; \theta) = \varphi^{-1}(p(t; \theta))$ . Instead of the logarithm, we consider the inverse  $\varphi^{-1}(\cdot)$  of a  $\varphi$ -function. Generalized statistical manifolds were introduced by the authors in [17,18]. Among examples of the generalized statistical manifold, (parametric)  $\varphi$ -families of probability distributions are of greatest importance. The non-parametric counterpart was investigated in [11,12]. The metric in  $\varphi$ -families can be defined as the Hessian of a function; i.e.,  $\varphi$ -families are Hessian manifolds [30]. In [17,18], the  $\varphi$ -divergence gives rise to a pair of dual connections  $D^{(-1)}$  and  $D^{(1)}$ ; and then for  $\alpha \in (-1, 1)$  the  $\alpha$ -connection  $D^{(\alpha)}$  is defined as the convex combination  $D^{(\alpha)} = \frac{1-\alpha}{2} D^{(-1)} + \frac{1+\alpha}{2} D^{(1)}$ . In the present paper, we show that the connection induced by  $\mathcal{D}_\varphi^{(\alpha)}(\cdot \parallel \cdot)$ , the generalization of Rényi divergence, corresponds to  $D^{(\alpha)}$ .

##### 4.1. Definitions

Let  $\varphi: \mathbb{R} \rightarrow (0, \infty)$  be a  $\varphi$ -function. A generalized statistical manifold  $\mathcal{P} = \{p(t; \theta) : \theta \in \Theta\}$  is a collection of probability distributions  $p_\theta(t) := p(t; \theta)$ , indexed by parameters  $\theta = (\theta^1, \dots, \theta^n) \in \Theta$  in a one-to-one relation, such that

- (m1)  $\Theta$  is a domain (open and connected set) in  $\mathbb{R}^n$ ;
- (m2)  $p(t; \theta)$  is differentiable with respect to  $\theta$ ;



(m3) the matrix  $g = (g_{ij})$  defined by

$$g_{ij} = -E_{\theta}' \left[ \frac{\partial^2 \varphi^{-1}(p_{\theta})}{\partial \theta^i \partial \theta^j} \right], \tag{22}$$

is positive definite at each  $\theta \in \Theta$ , where

$$E_{\theta}'[\cdot] = \frac{\int_T (\cdot) \varphi'(\varphi^{-1}(p_{\theta})) d\mu}{\int_T u_0 \varphi'(\varphi^{-1}(p_{\theta})) d\mu}, \tag{23}$$

(m4) the operations of integration with respect to  $\mu$  and differentiation with respect to  $\theta^i$  commute in all calculations found below, which are related to the metric and connections.

The matrix  $g = (g_{ij})$  equips  $\mathcal{P}$  with a metric. By the chain rule, the tensor related to  $g = (g_{ij})$  is invariant under change of coordinates. The (classical) statistical manifold is a particular case in which  $\varphi(u) = \exp(u)$  and  $u_0 = \mathbf{1}_T$ .

We introduce a notation similar to Equation (23) that involves higher order derivatives of  $\varphi(\cdot)$ . For each  $n \geq 1$ , we define

$$E_{\theta}^{(n)}[\cdot] = \frac{\int_T (\cdot) \varphi^{(n)}(\varphi^{-1}(p_{\theta})) d\mu}{\int_T u_0 \varphi^{(n)}(\varphi^{-1}(p_{\theta})) d\mu}. \tag{24}$$

We also use  $E_{\theta}'[\cdot]$ ,  $E_{\theta}''[\cdot]$  and  $E_{\theta}'''[\cdot]$  to denote  $E_{\theta}^{(n)}[\cdot]$  for  $n = 1, 2, 3$ , respectively. The notation (24) appears in expressions related to the metric and connections.

Using property (m4), we can find an alternate expression for  $g_{ij}$  as well as an identification involving tangent spaces. The matrix  $g = (g_{ij})$  can be equivalently defined by

$$g_{ij} = E_{\theta}'' \left[ \frac{\partial \varphi^{-1}(p_{\theta})}{\partial \theta^i} \frac{\partial \varphi^{-1}(p_{\theta})}{\partial \theta^j} \right]. \tag{25}$$

As a consequence of this equivalence, the tangent space  $T_{p_{\theta}}\mathcal{P}$  can be identified with  $\tilde{T}_{p_{\theta}}\mathcal{P}$ , the vector space spanned by  $\frac{\partial \varphi^{-1}(p_{\theta})}{\partial \theta^i}$ , and endowed with the inner product  $\langle \tilde{X}, \tilde{Y} \rangle_{\theta} := E_{\theta}''[\tilde{X}\tilde{Y}]$ . The mapping

$$\sum_i a_i \frac{\partial}{\partial \theta^i} \mapsto \sum_i a_i \frac{\partial \varphi^{-1}(p_{\theta})}{\partial \theta^i}$$

defines an isometry between  $T_{p_{\theta}}\mathcal{P}$  and  $\tilde{T}_{p_{\theta}}\mathcal{P}$ .

To verify (25), we differentiate  $\int_T p_{\theta} d\mu = 1$ , with respect to  $\theta^i$ , to get

$$0 = \frac{\partial}{\partial \theta^i} \int_T p_{\theta} d\mu = \int_T \frac{\partial}{\partial \theta^i} \varphi(\varphi^{-1}(p_{\theta})) d\mu = \int_T \frac{\partial \varphi^{-1}(p_{\theta})}{\partial \theta^i} \varphi'(\varphi^{-1}(p_{\theta})) d\mu. \tag{26}$$

Now, differentiating with respect to  $\theta^j$ , we obtain

$$0 = \int_T \frac{\partial^2 \varphi^{-1}(p_{\theta})}{\partial \theta^i \partial \theta^j} \varphi'(\varphi^{-1}(p_{\theta})) d\mu + \int_T \frac{\partial \varphi^{-1}(p_{\theta})}{\partial \theta^i} \frac{\partial \varphi^{-1}(p_{\theta})}{\partial \theta^j} \varphi''(\varphi^{-1}(p_{\theta})) d\mu,$$

and then (25) follows. In view of (26), we notice that every vector  $\tilde{X}$  belonging to  $\tilde{T}_{p_{\theta}}\mathcal{P}$  satisfies  $E_{\theta}'[\tilde{X}] = 0$ .

The metric  $g = (g_{ij})$  gives rise to a Levi-Civita connection  $\nabla$  (i.e., a torsion-free, metric connection), whose corresponding Christoffel symbols  $\Gamma_{ijk}$  are given by

$$\Gamma_{ijk} := \frac{1}{2} \left( \frac{\partial g_{ki}}{\partial \theta^j} + \frac{\partial g_{kj}}{\partial \theta^i} - \frac{\partial g_{ij}}{\partial \theta^k} \right). \tag{27}$$

Using expression (25) to calculate the derivatives in (27), we can express

$$\begin{aligned} \Gamma_{ijk} = & E''_{\theta} \left[ \frac{\partial^2 \varphi^{-1}(p_{\theta})}{\partial \theta^i \partial \theta^j} \frac{\partial \varphi^{-1}(p_{\theta})}{\partial \theta^k} \right] + \frac{1}{2} E'''_{\theta} \left[ \frac{\partial \varphi^{-1}(p_{\theta})}{\partial \theta^i} \frac{\partial \varphi^{-1}(p_{\theta})}{\partial \theta^j} \frac{\partial \varphi^{-1}(p_{\theta})}{\partial \theta^k} \right] \\ & - \frac{1}{2} E''_{\theta} \left[ \frac{\partial \varphi^{-1}(p_{\theta})}{\partial \theta^i} \frac{\partial \varphi^{-1}(p_{\theta})}{\partial \theta^k} \right] E''_{\theta} \left[ u_0 \frac{\partial \varphi^{-1}(p_{\theta})}{\partial \theta^j} \right] - \frac{1}{2} E''_{\theta} \left[ \frac{\partial \varphi^{-1}(p_{\theta})}{\partial \theta^j} \frac{\partial \varphi^{-1}(p_{\theta})}{\partial \theta^k} \right] E''_{\theta} \left[ u_0 \frac{\partial \varphi^{-1}(p_{\theta})}{\partial \theta^i} \right] \\ & + \frac{1}{2} E''_{\theta} \left[ \frac{\partial \varphi^{-1}(p_{\theta})}{\partial \theta^i} \frac{\partial \varphi^{-1}(p_{\theta})}{\partial \theta^j} \right] E''_{\theta} \left[ u_0 \frac{\partial \varphi^{-1}(p_{\theta})}{\partial \theta^k} \right]. \end{aligned}$$

As we will show later, the Levi-Civita connection  $\nabla$  corresponds to the connection derived from the divergence  $\mathcal{D}_{\varphi}^{(\alpha)}(\cdot \parallel \cdot)$  with  $\alpha = 0$ .

#### 4.2. $\varphi$ -Families

Let  $c: T \rightarrow \mathbb{R}$  be a measurable function for which  $p = \varphi(c)$  is a probability density in  $\mathcal{P}_{\mu}$ . Fix measurable functions  $u_1, \dots, u_n: T \rightarrow \mathbb{R}$ . A (parametric)  $\varphi$ -family  $\mathcal{F}_p = \{p_{\theta} : \theta \in \Theta\}$ , centered at  $p = \varphi(c)$ , is a set of probability distributions in  $\mathcal{P}_{\mu}$ , whose members can be written in the form

$$p_{\theta} := \varphi \left( c + \sum_{i=1}^n \theta^i u_i - \psi(\theta) u_0 \right), \quad \text{for each } \theta = (\theta^i) \in \Theta, \tag{28}$$

where  $\psi: \Theta \rightarrow [0, \infty)$  is a normalizing function, which is introduced so that expression (28) defines a probability distribution belonging to  $\mathcal{P}_{\mu}$ .

The functions  $u_1, \dots, u_n$  are not arbitrary. They are chosen to satisfy the following assumptions:

- (i)  $u_0, u_1, \dots, u_n$  are linearly independent,
- (ii)  $\int_T u_i \varphi'(c) d\mu = 0$ , and
- (iii) there exists  $\varepsilon > 0$  such that  $\int_T \varphi(c + \lambda u_i) d\mu < \infty$ , for all  $\lambda \in (-\varepsilon, \varepsilon)$ .

Moreover, the domain  $\Theta \subseteq \mathbb{R}^n$  is defined as the set of all vectors  $\theta = (\theta^i)$  for which

$$\int_T \varphi \left( c + \lambda \sum_{i=1}^n \theta^i u_i \right) d\mu < \infty, \quad \text{for some } \lambda > 1.$$

Condition (i) implies that the mapping defined by (28) is one-to-one. Assumption (ii) makes of  $\psi$  a non-negative function. Indeed, by the convexity of  $\varphi(\cdot)$ , along with (ii), we can write

$$\int_T \varphi(c) d\mu = \int_T \left[ \varphi(c) + \left( \sum_{i=1}^n \theta^i u_i \right) \varphi'(c) \right] d\mu \leq \int_T \varphi \left( c + \sum_{i=1}^n \theta^i u_i \right) d\mu,$$

which implies  $\psi(\theta) \geq 0$ . By condition (iii), the domain  $\Theta$  is an open neighborhood of the origin. If the set  $T$  is finite, condition (iii) is always satisfied. One can show that the domain  $\Theta$  is open and convex. Moreover, the normalizing function  $\psi$  is also convex (or strictly convex if  $\varphi(\cdot)$  is strictly convex). Conditions (ii) and (iii) also appears in the definition of non-parametric  $\varphi$ -families. For further details, we refer to [11,12].

In a  $\varphi$ -family  $\mathcal{F}_p$ , the matrix  $(g_{ij})$  given by (22) or (25) can be expressed as the Hessian of  $\psi$ . If  $\varphi(\cdot)$  is strictly convex, then  $(g_{ij})$  is positive definite. From

$$\frac{\partial \varphi^{-1}(p_{\theta})}{\partial \theta^i} = u_i - \frac{\partial \psi}{\partial \theta^i}, \quad -\frac{\partial^2 \varphi^{-1}(p_{\theta})}{\partial \theta^i \partial \theta^j} = -\frac{\partial^2 \psi}{\partial \theta^i \partial \theta^j},$$

it follows that  $g_{ij} = \frac{\partial^2 \psi}{\partial \theta^i \partial \theta^j}$ .

The next two results show how the generalization of Rényi divergence and the  $\varphi$ -divergence are related to the normalizing function in  $\varphi$ -families.

**Proposition 2.** In a  $\varphi$ -family  $\mathcal{F}_p$ , the generalization of Rényi divergence for  $\alpha \in (-1, 1)$  can be expressed in terms of the normalizing function  $\psi$  as follows:

$$\mathcal{D}_\varphi^{(\alpha)}(p_\theta \parallel p_\vartheta) = \frac{2}{1+\alpha}\psi(\theta) + \frac{2}{1-\alpha}\psi(\vartheta) - \frac{4}{1-\alpha^2}\psi\left(\frac{1-\alpha}{2}\theta + \frac{1+\alpha}{2}\vartheta\right), \tag{29}$$

for all  $\theta, \vartheta \in \Theta$ .

**Proof.** Recall the definition of  $\kappa(\alpha)$  as the real number for which

$$\int_T \varphi\left(\frac{1-\alpha}{2}\varphi^{-1}(p_\theta) + \frac{1+\alpha}{2}\varphi^{-1}(p_\vartheta) + \kappa(\alpha)u_0\right) d\mu = 1.$$

Using expression (28) for probability distributions in  $\mathcal{F}_p$ , we can write

$$\begin{aligned} \frac{1-\alpha}{2}\varphi^{-1}(p_\theta) + \frac{1+\alpha}{2}\varphi^{-1}(p_\vartheta) + \kappa(\alpha)u_0 &= c + \sum_{i=1}^n \left(\frac{1-\alpha}{2}\theta^i + \frac{1+\alpha}{2}\vartheta^i\right)u_i - \left(\frac{1-\alpha}{2}\psi(\theta) + \frac{1+\alpha}{2}\psi(\vartheta) - \kappa(\alpha)\right)u_0 \\ &= c + \sum_{i=1}^n \left(\frac{1-\alpha}{2}\theta^i + \frac{1+\alpha}{2}\vartheta^i\right)u_i - \psi\left(\frac{1-\alpha}{2}\theta + \frac{1+\alpha}{2}\vartheta\right)u_0. \end{aligned}$$

The last equality is a consequence of the domain  $\Theta$  being convex. Thus, it follows that

$$\kappa(\alpha) = \frac{1-\alpha}{2}\psi(\theta) + \frac{1+\alpha}{2}\psi(\vartheta) - \psi\left(\frac{1-\alpha}{2}\theta + \frac{1+\alpha}{2}\vartheta\right).$$

By the definition of  $\mathcal{D}_\varphi^{(\alpha)}(\cdot \parallel \cdot)$ , we get (29).  $\square$

**Proposition 3.** In a  $\varphi$ -family  $\mathcal{F}_p$ , the  $\varphi$ -divergence is related to the normalizing function  $\psi$  by the equality

$$\mathcal{D}_\varphi(p_\theta \parallel p_\vartheta) = \psi(\vartheta) - \psi(\theta) - \nabla\psi(\theta) \cdot (\vartheta - \theta), \tag{30}$$

for all  $\theta, \vartheta \in \Theta$ .

**Proof.** To show (30), we use

$$\frac{\partial\psi}{\partial\theta^i}(\theta) = \frac{\int_T u_i \varphi'(\varphi^{-1}(p_\theta)) d\mu}{\int_T u_0 \varphi'(\varphi^{-1}(p_\theta)) d\mu'}$$

which is a consequence of (Lemma 10 in [12]). In view of  $(\varphi^{-1})'(u) = 1/\varphi'(\varphi^{-1}(u))$ , expression (13) with  $p = p_\theta$  and  $q = p_\vartheta$  results in

$$\mathcal{D}_\varphi(p_\theta \parallel p_\vartheta) = \frac{\int_T [\varphi^{-1}(p_\theta) - \varphi^{-1}(p_\vartheta)] \varphi'(\varphi^{-1}(p_\theta)) d\mu}{\int_T u_0 \varphi'(\varphi^{-1}(p_\theta)) d\mu}. \tag{31}$$

Inserting into (31) the difference

$$\begin{aligned} \varphi^{-1}(p_\theta) - \varphi^{-1}(p_\vartheta) &= \left(c + \sum_{i=1}^n \theta^i u_i - \psi(\theta)u_0\right) - \left(c + \sum_{i=1}^n \vartheta^i u_i - \psi(\vartheta)u_0\right) \\ &= \psi(\vartheta)u_0 - \psi(\theta)u_0 - \sum_{i=1}^n (\vartheta^i - \theta^i)u_i, \end{aligned}$$

we get expression (30).  $\square$

In Proposition 2, the expression on the right-hand side of Equation (29) defines a divergence on its own, which was investigated by Jun Zhang in [19]. Proposition 3 asserts that the  $\varphi$ -divergence  $\mathcal{D}_\varphi(p_\vartheta \parallel p_\theta)$  coincides with the Bregman divergence [31,32] associated with the normalizing function  $\psi$  for points  $\vartheta$  and  $\theta$  in  $\Theta$ . Because  $\psi$  is convex and attains a minimum at  $\theta = 0$ , it follows that  $\frac{\partial \psi}{\partial \theta^i}(\theta) = 0$  at  $\theta = 0$ . As a result, equality (30) reduces to  $\mathcal{D}_\varphi(p \parallel p_\theta) = \psi(\theta)$ .

### 4.3. Geometry Induced by $\mathcal{D}_\varphi^{(\alpha)}(\cdot \parallel \cdot)$

In this section, we assume that  $\varphi(\cdot)$  is continuously differentiable and strictly convex. The latter assumption guarantees that

$$\mathcal{D}_\varphi^{(\alpha)}(p \parallel q) = 0 \quad \text{if and only if} \quad p = q. \tag{32}$$

The generalized Rényi divergence induces a metric  $g = (g_{ij})$  in generalized statistical manifolds  $\mathcal{P}$ . This metric is given by

$$g_{ij} = - \left[ \left( \frac{\partial}{\partial \theta^i} \right)_p \left( \frac{\partial}{\partial \theta^j} \right)_q \mathcal{D}_\varphi^{(\alpha)}(p \parallel q) \right]_{q=p}. \tag{33}$$

To show that this expression defines a metric, we have to verify that  $g_{ij}$  is invariant under change of coordinates, and  $(g_{ij})$  is positive definite. The first claim follows from the chain rule. The positive definiteness of  $(g_{ij})$  is a consequence of Proposition 4, which is given below.

**Proposition 4.** *The metric induced by  $\mathcal{D}_\varphi^{(\alpha)}(\cdot \parallel \cdot)$  coincides with the metric given by (22) or (25).*

**Proof.** Fix any  $\alpha \in (-1, 1)$ . Applying the operator  $\left( \frac{\partial}{\partial \theta^i} \right)_{p_\vartheta}$  to

$$\int_T \varphi(c_\alpha) d\mu = 1,$$

where  $c_\alpha = \frac{1-\alpha}{2} \varphi^{-1}(p_\vartheta) + \frac{1+\alpha}{2} \varphi^{-1}(p_\theta) + \kappa(\alpha)u_0$ , we obtain

$$\int_T \left( \frac{1+\alpha}{2} \frac{\partial \varphi^{-1}(p_\vartheta)}{\partial \theta^j} + \left( \frac{\partial}{\partial \theta^j} \right)_{p_\vartheta} \kappa(\alpha)u_0 \right) \varphi'(c_\alpha) d\mu = 0,$$

which results in

$$\left( \frac{\partial}{\partial \theta^j} \right)_{p_\vartheta} \kappa(\alpha) = - \frac{1+\alpha}{2} \frac{\int_T \frac{\partial \varphi^{-1}(p_\vartheta)}{\partial \theta^j} \varphi'(c_\alpha) d\mu}{\int_T u_0 \varphi'(c_\alpha) d\mu}.$$

By the standard differentiation rules, we can write

$$\begin{aligned} \left( \frac{\partial}{\partial \theta^i} \right)_{p_\vartheta} \left( \frac{\partial}{\partial \theta^j} \right)_{p_\vartheta} \kappa(\alpha) &= - \frac{1+\alpha}{2} \frac{\int_T \left[ \frac{1-\alpha}{2} \frac{\partial^2 \varphi^{-1}(p_\vartheta)}{\partial \theta^i \partial \theta^j} + \left( \frac{\partial}{\partial \theta^i} \right)_{p_\vartheta} \kappa(\alpha)u_0 \right] \frac{\partial \varphi^{-1}(p_\vartheta)}{\partial \theta^j} \varphi''(c_\alpha) d\mu}{\int_T u_0 \varphi'(c_\alpha) d\mu} \\ &\quad + \frac{1+\alpha}{2} \frac{\int_T \frac{\partial \varphi^{-1}(p_\vartheta)}{\partial \theta^j} \varphi'(c_\alpha) d\mu}{\int_T u_0 \varphi'(c_\alpha) d\mu} \frac{\int_T u_0 \left[ \frac{1-\alpha}{2} \frac{\partial^2 \varphi^{-1}(p_\vartheta)}{\partial \theta^i \partial \theta^j} + \left( \frac{\partial}{\partial \theta^i} \right)_{p_\vartheta} \kappa(\alpha)u_0 \right] \varphi''(c_\alpha) d\mu}{\int_T u_0 \varphi'(c_\alpha) d\mu}. \end{aligned} \tag{34}$$

Noticing that  $\int_T \frac{\partial \varphi^{-1}(p_\vartheta)}{\partial \theta^j} \varphi'(c_\alpha) d\mu = 0$  for  $p_\vartheta = p_\theta$ , the second term on the right-hand side of Equation (34) vanishes, and then

$$\left[ \left( \frac{\partial}{\partial \theta^i} \right)_{p_\vartheta} \left( \frac{\partial}{\partial \theta^j} \right)_{p_\vartheta} \kappa(\alpha) \right]_{p_\vartheta=p_\theta} = - \frac{1-\alpha^2}{4} \frac{\int_T \frac{\partial \varphi^{-1}(p_\vartheta)}{\partial \theta^i} \frac{\partial \varphi^{-1}(p_\vartheta)}{\partial \theta^j} \varphi''(\varphi^{-1}(p_\theta)) d\mu}{\int_T u_0 \varphi'(\varphi^{-1}(p_\theta)) d\mu}.$$

If we use the notation introduced in (24), we can write

$$g_{ij} = - \left[ \left( \frac{\partial}{\partial \theta^i} \right)_{p_\theta} \left( \frac{\partial}{\partial \theta^j} \right)_{p_\theta} \mathcal{D}_\varphi^{(\alpha)}(p_\theta \| p_\theta) \right]_{p_\theta=p_\theta} = E_\theta'' \left[ \frac{\partial \varphi^{-1}(p_\theta)}{\partial \theta^i} \frac{\partial \varphi^{-1}(p_\theta)}{\partial \theta^j} \right].$$

It remains to show the case  $\alpha = \pm 1$ . Comparing (13) and (23), we can write

$$\mathcal{D}_\varphi(p_\theta \| p_\theta) = E_\theta'[\varphi^{-1}(p_\theta) - \varphi^{-1}(p_\theta)]. \tag{35}$$

We use the equivalent expressions

$$g_{ij} = \left[ \left( \frac{\partial^2}{\partial \theta^i \partial \theta^j} \right)_p \mathcal{D}_\varphi^\alpha(p \| q) \right]_{q=p} = \left[ \left( \frac{\partial^2}{\partial \theta^i \partial \theta^j} \right)_q \mathcal{D}_\varphi^\alpha(p \| q) \right]_{q=p},$$

which follows from condition (32), to infer that

$$g_{ij} = \left[ \left( \frac{\partial^2}{\partial \theta^i \partial \theta^j} \right)_{p_\theta} \mathcal{D}_\varphi(p_\theta \| p_\theta) \right]_{p_\theta=p_\theta} = -E_\theta' \left[ \frac{\partial^2 \varphi^{-1}(p_\theta)}{\partial \theta^i \partial \theta^j} \right]. \tag{36}$$

Because  $\mathcal{D}_\varphi^{(-1)}(p \| q) = \mathcal{D}_\varphi^{(1)}(q \| p) = \mathcal{D}_\varphi(p \| q)$ , we conclude that the metric defined by (22) coincides with the metric induced by  $\mathcal{D}_\varphi^{(-1)}(\cdot \| \cdot)$  and  $\mathcal{D}_\varphi^{(1)}(\cdot \| \cdot)$ .  $\square$

In generalized statistical manifolds, the generalized Rényi divergence  $\mathcal{D}_\varphi^{(\alpha)}(\cdot \| \cdot)$  induces a connection  $D^{(\alpha)}$ , whose Christoffel symbols  $\Gamma_{ijk}^{(\alpha)}$  are given by

$$\Gamma_{ijk}^{(\alpha)} = - \left[ \left( \frac{\partial^2}{\partial \theta^i \partial \theta^j} \right)_p \left( \frac{\partial}{\partial \theta^k} \right)_q \mathcal{D}_\varphi^{(\alpha)}(p \| q) \right]_{q=p}.$$

Because  $\mathcal{D}_\varphi^{(\alpha)}(p \| q) = \mathcal{D}_\varphi^{(-\alpha)}(q \| p)$ , it follows that  $D^{(\alpha)}$  and  $D^{(-\alpha)}$  are mutually dual for any  $\alpha \in [-1, 1]$ . In other words,  $\Gamma_{ijk}^{(\alpha)}$  and  $\Gamma_{ijk}^{(-\alpha)}$  satisfy the relation  $\frac{\partial g_{jk}}{\partial \theta^i} = \Gamma_{ijk}^{(\alpha)} + \Gamma_{ikj}^{(-\alpha)}$ . A development involving expression (35) results in

$$\Gamma_{ijk}^{(1)} = E_\theta'' \left[ \frac{\partial^2 \varphi^{-1}(p_\theta)}{\partial \theta^i \partial \theta^j} \frac{\partial \varphi^{-1}(p_\theta)}{\partial \theta^k} \right] - E_\theta' \left[ \frac{\partial^2 \varphi^{-1}(p_\theta)}{\partial \theta^i \partial \theta^j} \right] E_\theta'' \left[ u_0 \frac{\partial \varphi^{-1}(p_\theta)}{\partial \theta^k} \right], \tag{37}$$

and

$$\begin{aligned} \Gamma_{ijk}^{(-1)} &= E_\theta'' \left[ \frac{\partial^2 \varphi^{-1}(p_\theta)}{\partial \theta^i \partial \theta^j} \frac{\partial \varphi^{-1}(p_\theta)}{\partial \theta^k} \right] + E_\theta''' \left[ \frac{\partial \varphi^{-1}(p_\theta)}{\partial \theta^i} \frac{\partial \varphi^{-1}(p_\theta)}{\partial \theta^j} \frac{\partial \varphi^{-1}(p_\theta)}{\partial \theta^k} \right] \\ &\quad - E_\theta'' \left[ \frac{\partial \varphi^{-1}(p_\theta)}{\partial \theta^j} \frac{\partial \varphi^{-1}(p_\theta)}{\partial \theta^k} \right] E_\theta'' \left[ u_0 \frac{\partial \varphi^{-1}(p_\theta)}{\partial \theta^i} \right] \\ &\quad - E_\theta'' \left[ \frac{\partial \varphi^{-1}(p_\theta)}{\partial \theta^i} \frac{\partial \varphi^{-1}(p_\theta)}{\partial \theta^k} \right] E_\theta'' \left[ u_0 \frac{\partial \varphi^{-1}(p_\theta)}{\partial \theta^j} \right]. \end{aligned} \tag{38}$$

For  $\alpha \in (-1, 1)$ , the Christoffel symbols  $\Gamma_{ijk}^{(\alpha)}$  can be written as a convex combination of  $\Gamma_{ijk}^{(-1)}$  and  $\Gamma_{ijk}^{(1)}$ , as asserted in the next result.

**Proposition 5.** *The Christoffel symbols  $\Gamma_{ijk}^{(\alpha)}$  induced by the divergence  $\mathcal{D}_\varphi^{(\alpha)}(\cdot \| \cdot)$  satisfy the relation*

$$\Gamma_{ijk}^{(\alpha)} = \frac{1 - \alpha}{2} \Gamma_{ijk}^{(-1)} + \frac{1 + \alpha}{2} \Gamma_{ijk}^{(1)}, \quad \text{for } \alpha \in [-1, 1]. \tag{39}$$

**Proof.** For  $\alpha = \pm 1$ , equality (39) follows trivially. Thus, we assume  $\alpha \in (-1, 1)$ . By (34), we can write

$$\begin{aligned} \left(\frac{\partial}{\partial \theta^i}\right)_{p_\theta} \left(\frac{\partial}{\partial \theta^k}\right)_{p_\theta} \kappa(\alpha) &= -\frac{1 + \alpha \int_T \left[\frac{1-\alpha}{2} \frac{\partial \varphi^{-1}(p_\theta)}{\partial \theta^i} + \left(\frac{\partial}{\partial \theta^i}\right)_{p_\theta} \kappa(\alpha) u_0\right] \frac{\partial \varphi^{-1}(p_\theta)}{\partial \theta^k} \varphi''(c_\alpha) d\mu}{2 \int_T u_0 \varphi'(c_\alpha) d\mu} \\ &+ \frac{1 + \alpha \int_T \frac{\partial \varphi^{-1}(p_\theta)}{\partial \theta^k} \varphi'(c_\alpha) d\mu}{2 \int_T u_0 \varphi'(c_\alpha) d\mu} \frac{\int_T u_0 \left[\frac{1-\alpha}{2} \frac{\partial \varphi^{-1}(p_\theta)}{\partial \theta^i} + \left(\frac{\partial}{\partial \theta^i}\right)_{p_\theta} \kappa(\alpha) u_0\right] \varphi''(c_\alpha) d\mu}{\int_T u_0 \varphi'(c_\alpha) d\mu}. \end{aligned} \tag{40}$$

Applying  $\left(\frac{\partial}{\partial \theta^j}\right)_{p_\theta}$  to the first term on the right-hand side of (40), and then equating  $p_\theta = p_\theta$ , we obtain

$$\begin{aligned} -\frac{1 - \alpha^2}{4} E_\theta'' \left[ \frac{\partial^2 \varphi^{-1}(p_\theta)}{\partial \theta^i \partial \theta^j} \frac{\partial \varphi^{-1}(p_\theta)}{\partial \theta^k} \right] - \frac{1 + \alpha}{2} \left(\frac{\partial^2}{\partial \theta^i \partial \theta^j}\right)_{p_\theta} \kappa(\alpha) E_\theta'' \left[ u_0 \frac{\partial \varphi^{-1}(p_\theta)}{\partial \theta^k} \right] \\ - \frac{1 - \alpha^2}{4} \frac{1 - \alpha}{2} E_\theta''' \left[ \frac{\partial \varphi^{-1}(p_\theta)}{\partial \theta^i} \frac{\partial \varphi^{-1}(p_\theta)}{\partial \theta^j} \frac{\partial \varphi^{-1}(p_\theta)}{\partial \theta^k} \right] \\ + \frac{1 - \alpha^2}{4} \frac{1 - \alpha}{2} E_\theta'' \left[ \frac{\partial \varphi^{-1}(p_\theta)}{\partial \theta^i} \frac{\partial \varphi^{-1}(p_\theta)}{\partial \theta^k} \right] E_\theta'' \left[ u_0 \frac{\partial \varphi^{-1}(p_\theta)}{\partial \theta^j} \right]. \end{aligned} \tag{41}$$

Similarly, if we apply  $\left(\frac{\partial}{\partial \theta^j}\right)_{p_\theta}$  to the second term on the right-hand side of (40), and make  $p_\theta = p_\theta$ , we get

$$\frac{1 - \alpha^2}{4} \frac{1 - \alpha}{2} E_\theta'' \left[ \frac{\partial \varphi^{-1}(p_\theta)}{\partial \theta^j} \frac{\partial \varphi^{-1}(p_\theta)}{\partial \theta^k} \right] E_\theta'' \left[ u_0 \frac{\partial \varphi^{-1}(p_\theta)}{\partial \theta^i} \right]. \tag{42}$$

Collecting (41) and (42), we can write

$$\begin{aligned} \Gamma_{ijk}^{(\alpha)} &= -\frac{4}{1 - \alpha^2} \left[ \left(\frac{\partial^2}{\partial \theta^i \partial \theta^j}\right)_{p_\theta} \left(\frac{\partial}{\partial \theta^k}\right)_{p_\theta} \kappa(\alpha) \right]_{p_\theta = p_\theta} \\ &= E_\theta'' \left[ \frac{\partial^2 \varphi^{-1}(p_\theta)}{\partial \theta^i \partial \theta^j} \frac{\partial \varphi^{-1}(p_\theta)}{\partial \theta^k} \right] + \frac{1 - \alpha}{2} E_\theta''' \left[ \frac{\partial \varphi^{-1}(p_\theta)}{\partial \theta^i} \frac{\partial \varphi^{-1}(p_\theta)}{\partial \theta^j} \frac{\partial \varphi^{-1}(p_\theta)}{\partial \theta^k} \right] \\ &- \frac{1 - \alpha}{2} E_\theta'' \left[ \frac{\partial \varphi^{-1}(p_\theta)}{\partial \theta^j} \frac{\partial \varphi^{-1}(p_\theta)}{\partial \theta^k} \right] E_\theta'' \left[ u_0 \frac{\partial \varphi^{-1}(p_\theta)}{\partial \theta^i} \right] \\ &- \frac{1 - \alpha}{2} E_\theta'' \left[ \frac{\partial \varphi^{-1}(p_\theta)}{\partial \theta^i} \frac{\partial \varphi^{-1}(p_\theta)}{\partial \theta^k} \right] E_\theta'' \left[ u_0 \frac{\partial \varphi^{-1}(p_\theta)}{\partial \theta^j} \right] \\ &- \frac{1 + \alpha}{2} E_\theta' \left[ \frac{\partial^2 \varphi^{-1}(p_\theta)}{\partial \theta^i \partial \theta^j} \right] E_\theta'' \left[ u_0 \frac{\partial \varphi^{-1}(p_\theta)}{\partial \theta^k} \right], \end{aligned} \tag{43}$$

where we used

$$\begin{aligned} \left(\frac{\partial^2}{\partial \theta^i \partial \theta^j}\right)_{p_\theta} \kappa(\alpha) &= \frac{1 - \alpha^2}{4} \left[ \left(\frac{\partial^2}{\partial \theta^i \partial \theta^j}\right)_{p_\theta} \mathcal{D}_\varphi^{(\alpha)}(p_\theta \parallel p_\theta) \right]_{p_\theta = p_\theta} \\ &= \frac{1 - \alpha^2}{4} g_{ij} = -\frac{1 - \alpha^2}{4} E_\theta' \left[ \frac{\partial^2 \varphi^{-1}(p_\theta)}{\partial \theta^i \partial \theta^j} \right]. \end{aligned}$$

Expression (39) follows from (37), (38) and (43).  $\square$

### 5. Conclusions

In [17,18], the authors introduced a pair of dual connections  $D^{(-1)}$  and  $D^{(1)}$  induced by  $\varphi$ -divergence. The main motivation of the present work was to find a (non-trivial) family of  $\alpha$ -divergences, whose induced  $\alpha$ -connections are convex combinations of  $D^{(-1)}$  and  $D^{(1)}$ . As a result of our efforts, we proposed a generalization of Rényi divergence. The connection  $D^{(\alpha)}$  induced by the generalization of Rényi divergence satisfies the relation  $D^{(\alpha)} = \frac{1-\alpha}{2} D^{(-1)} + \frac{1+\alpha}{2} D^{(1)}$ . To generalize Rényi divergence, we made use of properties of  $\varphi$ -functions. This makes evident the importance of  $\varphi$ -functions in the geometry of non-standard models. In standard statistical manifolds, even though Amari's  $\alpha$ -divergence and Rényi divergence (with  $\alpha \in [-1, 1]$ ) do not coincide, they induce the same family of  $\alpha$ -connections. This striking result requires further investigation. Future work should focus

on how the generalization of Rényi divergence is related to Zhang's  $(\rho, \tau)$ -divergence, and also how the present proposal is related to the model presented in [33].

**Acknowledgments:** The authors are indebted to the anonymous reviewers for their valuable comments and corrections, which led to a great improvement of this paper. Charles C. Cavalcante also thanks the CNPq (Proc. 309055/2014-8) for partial funding.

**Author Contributions:** All authors contributed equally to the design of the research. The research was carried out by all authors. Rui F. Vigelis and Charles C. Cavalcante gave the central idea of the paper and managed the organization of it. Rui F. Vigelis wrote the paper. All the authors read and approved the final manuscript.

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

1. Rao, C.R. Information and the accuracy attainable in the estimation of statistical parameters. *Bull. Calcutta Math. Soc.* **1945**, *37*, 81–91.
2. Amari, S.-I. Differential geometry of curved exponential families—Curvatures and information loss. *Ann. Stat.* **1982**, *10*, 357–385.
3. Amari, S.-I. *Differential-Geometrical Methods in Statistics*; Springer: Berlin/Heidelberg, Germany, 1985; Volume 28.
4. Amari, S.-I.; Nagaoka, H. *Methods of Information Geometry (Translations of Mathematical Monographs)*; American Mathematical Society: Providence, RI, USA, 2000; Volume 191.
5. Amari, S.-I. *Information Geometry and Its Applications*; Applied Mathematical Sciences Series; Springer: Berlin/Heidelberg, Germany, 2016; Volume 194.
6. Amari, S.-I.; Ohara, A.; Matsuzoe, H. Geometry of deformed exponential families: Invariant, dually-flat and conformal geometries. *Physica A* **2012**, *391*, 4308–4319.
7. Matsuzoe, H. Hessian structures on deformed exponential families and their conformal structures. *Differ. Geom. Appl.* **2014**, *35* (Suppl.), 323–333.
8. Naudts, J. Estimators, escort probabilities, and  $\phi$ -exponential families in statistical physics. *J. Inequal. Pure Appl. Math.* **2004**, *5*, 102.
9. Pistone, G.  $\kappa$ -exponential models from the geometrical viewpoint. *Eur. Phys. J. B* **2009**, *70*, 29–37.
10. Amari, S.-I.; Ohara, A. Geometry of  $q$ -exponential family of probability distributions. *Entropy* **2011**, *13*, 1170–1185.
11. Vigelis, R.F.; Cavalcante, C.C. The  $\Delta_2$ -Condition and  $\varphi$ -Families of Probability Distributions. In *Geometric Science of Information*; Springer: Berlin/Heidelberg, Germany, 2013; Volume 8085, pp. 729–736.
12. Vigelis, R.F.; Cavalcante, C.C. On  $\varphi$ -families of probability distributions. *J. Theor. Probab.* **2013**, *26*, 870–884.
13. Cena, A.; Pistone, G. Exponential statistical manifold. *Ann. Inst. Stat. Math.* **2007**, *59*, 27–56.
14. Grasselli, M.R. Dual connections in nonparametric classical information geometry. *Ann. Inst. Stat. Math.* **2010**, *62*, 873–896.
15. Pistone, G.; Sempi, C. An infinite-dimensional geometric structure on the space of all the probability measures equivalent to a given one. *Ann. Stat.* **1995**, *23*, 1543–1561.
16. Santacroce, M.; Siri, P.; Trivellato, B. New results on mixture and exponential models by Orlicz spaces. *Bernoulli* **2016**, *22*, 1431–1447.
17. Vigelis, R.F.; Cavalcante, C.C. Information Geometry: An Introduction to New Models for Signal Processing. In *Signals and Images*; CRC Press: Boca Raton, FL, USA, 2015; pp. 455–491.
18. Vigelis, R.F.; de Souza, D.C.; Cavalcante, C.C. New Metric and Connections in Statistical Manifolds. In *Geometric Science of Information*; Springer: Berlin/Heidelberg, Germany, 2015; Volume 9389, pp. 222–229.
19. Zhang, J. Divergence function, duality, and convex analysis. *Neural Comput.* **2004**, *16*, 159–195.
20. Zhang, J. Referential Duality and Representational Duality on Statistical Manifolds. In Proceedings of the 2nd International Symposium on Information Geometry and Its Applications, Pescara, Italy, 12–16 December 2005; pp. 58–67.
21. Zhang, J. Nonparametric information geometry: From divergence function to referential-representational biduality on statistical manifolds. *Entropy* **2013**, *15*, 5384–5418.
22. Zhang, J. Divergence Functions and Geometric Structures They Induce on a Manifold. In *Geometric Theory of Information*; Springer: Berlin/Heidelberg, Germany, 2014; pp. 1–30.

23. Zhang, J. On monotone embedding in information geometry. *Entropy* **2015**, *17*, 4485–4489.
24. Rényi, A. On measures of entropy and information. In *Proceedings of 4th Berkeley Symposium on Mathematical Statistics and Probability*; University California Press: Berkeley, CA, USA, 1961; Volume I, pp. 547–561.
25. Van Erven, T.; Harremoës, P. Rényi divergence and Kullback-Leibler divergence. *IEEE Trans. Inform. Theory* **2014**, *60*, 3797–3820.
26. Eguchi, S.; Komori, O. Path Connectedness on a Space of Probability Density Functions. In *Geometric Science of Information*; Springer: Berlin/Heidelberg, Germany, 2015; Volume 9389, pp. 615–624.
27. Kaniadakis, G.; Lissia, M.; Scarfone, A.M. Deformed logarithms and entropies. *Physica A* **2004**, *340*, 41–49.
28. Kaniadakis, G. Theoretical foundations and mathematical formalism of the power-law tailed statistical distributions. *Entropy* **2013**, *15*, 3983–4010.
29. Musielak, J. *Orlicz Spaces and Modular Spaces*; Springer: Berlin/Heidelberg, Germany, 1983; Volume 1034.
30. Shima, H. *The Geometry of Hessian Structures*; World Scientific: Singapore, 2007.
31. Banerjee, A.; Merugu, S.; Dhillon, I.S.; Ghosh, J. Clustering with Bregman Divergences. *J. Mach. Learn. Res.* **2005**, *6*, 1705–1749.
32. Bregman, L.M. The relaxation method of finding the common point of convex sets and its application to the solution of problems in convex programming. *USSR Comput. Math. Math. Phys.* **1967**, *7*, 200–217.
33. Zhanga, J.; Hästö, P. Statistical manifold as an affine space: A functional equation approach. *J. Math. Psychol.* **2006**, *50*, 60–65.



© 2016 by the authors; licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC-BY) license (<http://creativecommons.org/licenses/by/4.0/>).