

Article

# Information Geometric Duality of $\phi$ -Deformed Exponential Families

Jan Korbel <sup>1,2</sup> , Rudolf Hanel <sup>1,2</sup> and Stefan Thurner <sup>1,2,3,4,\*</sup>

<sup>1</sup> Section for Science of Complex Systems, CeMSIIS, Medical University of Vienna, Spitalgasse 23, A-1090 Vienna, Austria; jan.korbel@meduniwien.ac.at (J.K.); rudolf.hanel@meduniwien.ac.at (R.H.)

<sup>2</sup> Complexity Science Hub Vienna, Josefstädter Strasse 39, A-1080 Vienna, Austria

<sup>3</sup> Santa Fe Institute, 1399 Hyde Park Road, Santa Fe, NM 87501, USA

<sup>4</sup> IIASA, Schlossplatz 1, A-2361 Laxenburg, Austria

\* Correspondence: stefan.thurner@meduniwien.ac.at

Received: 24 December 2018; Accepted: 16 January 2019; Published: 24 January 2019



**Abstract:** In the world of generalized entropies—which, for example, play a role in physical systems with sub- and super-exponential phase space growth per degree of freedom—there are two ways for implementing constraints in the maximum entropy principle: linear and escort constraints. Both appear naturally in different contexts. Linear constraints appear, e.g., in physical systems, when additional information about the system is available through higher moments. Escort distributions appear naturally in the context of multifractals and information geometry. It was shown recently that there exists a fundamental duality that relates both approaches on the basis of the corresponding deformed logarithms (deformed-log duality). Here, we show that there exists another duality that arises in the context of information geometry, relating the Fisher information of  $\phi$ -deformed exponential families that correspond to linear constraints (as studied by J. Naudts) to those that are based on escort constraints (as studied by S.-I. Amari). We explicitly demonstrate this information geometric duality for the case of  $(c, d)$ -entropy, which covers all situations that are compatible with the first three Shannon–Khinchin axioms and that include Shannon, Tsallis, Anteneodo–Plastino entropy, and many more as special cases. Finally, we discuss the relation between the deformed-log duality and the information geometric duality and mention that the escort distributions arising in these two dualities are generally different and only coincide for the case of the Tsallis deformation.

**Keywords:** generalized entropy;  $\phi$ -deformed family; Fisher information; information geometry; Cramér–Rao bound;  $(c, d)$ -entropy

## 1. Introduction

Entropy is one word for several distinct concepts [1]. It was originally introduced in thermodynamics, then in statistical physics, information theory, and last in the context of statistical inference. One important application of entropy in statistical physics, and in statistical inference in general, is the maximum entropy principle, which allows us to estimate probability distribution functions from limited information sources, i.e., from data [2,3]. The formal concept of entropy was generalized to also account for power laws that occur frequently in complex systems [4]. Literally dozens of generalized entropies have been proposed in various contexts, such as relativity [5], multifractals [6], or black holes [7]; see [8] for an overview. All these generalized entropies, whenever they fulfil the first three Shannon–Khinchin axioms (and violate the composition axiom) are special cases of the  $(c, d)$ -entropy asymptotically [9]. Generalized entropies play a role in non-multinomial, sub-additive systems (whose phase space volume grows sub-exponentially with the degrees of freedom) [10,11]

and in systems, whose phase space grows super-exponentially [12]. All generalized entropies, for sub-, and super-exponential systems, can be treated within a single, unifying framework [13].

With the advent of generalized entropies, depending on context, two types of constraint are used in the maximum entropy principle: traditional linear constraints (typically moments),  $\langle E \rangle = \sum_i p_i E_i$ , motivated by physical measurements, and the so-called escort constraints,  $\langle E \rangle_u = \sum_i u(p_i) E_i / \sum_i u(p_i)$ , where  $u$  is some nonlinear function. Originally, the latter were introduced with multifractals in mind [4]. Different types of constraint arise from different applications of relative entropy. While for physics-related contexts (such as thermodynamics), linear constraints are normally used, in other applications, such as non-linear dynamical systems or information geometry, it might be more natural to consider escort constraints. The question about their correct use and the appropriate form of constraints has caused a heated debate in the past decade [14–19]. To introduce escort distributions in the maximum entropy principle in a consistent way, two approaches have been discussed. The first [20] appears in the context of deformed entropies that are motivated by superstatistics [21]. It was later observed in [22] that this approach is linked to other deformed entropies with linear constraints through a fundamental duality (deformed-log duality), such that both entropies lead to the same functional form of MaxEnt distributions. The second way to obtain escort distributions was studied by Amari et al. and is motivated by information geometry and the theory of statistical estimation [23,24]. There, escort distributions represent natural coordinates on a statistical manifold [24,25].

In this paper, we show that there exists another duality relation between this information geometric approach with escort distributions and an approach that uses linear constraints. The relation can be given a precise information geometric meaning on the basis of the Fisher information. We show this within the framework of  $\phi$ -deformations [26–28]. We establish the duality relation for both cases in the relevant information geometric quantities. As an example, we explicitly show the duality relation for the class of  $(c, d)$ -exponentials, introduced in [9,10]. Finally, we discuss the relation between the deformed-log duality and the information geometric duality and show that these have fundamental differences. Each type of duality is suitable for different applications. We hope that this paper helps to avoid confusion in the use of escort distributions in the various contexts.

Let us start with reviewing central concepts of (non-deformed) information geometry, in particular relative entropy and its relation to the exponential family of distributions through the maximum entropy principle. Let us consider a probability simplex,  $S^n$ , with  $n$  independent probabilities,  $p_i$ , and probability  $p_0$ . Its value is not independent, but determined by the normalization condition,  $p_0 = 1 - \sum_i p_i$ . Further, consider a parametric family of distributions  $\mathbf{p}(\boldsymbol{\theta})$  with parameter vector  $\boldsymbol{\theta} \in \mathcal{M}$ , where  $\mathcal{M}$  is a parametric space. In this paper, we focus on probabilities over discrete sample spaces, only. For the continuous case see the mathematical formulation of Pistone and Sempi [29]. For the sake of simplicity, we sometimes do not explicitly write the parameter vector  $\boldsymbol{\theta}$  and consider  $p_i$  as the independent parameters of the distribution. It is easy to show that the choice of  $p_i$  determines the choice of the parameters  $\theta_i$ .

Relative entropy, or Kullback–Leibler divergence, is defined as:

$$D_{KL}(\mathbf{p} \parallel \mathbf{q}) = \sum_i p_i \log(p_i/q_i) \quad . \quad (1)$$

For the uniform distribution  $\mathbf{q} = \mathbf{u}_n$ , i.e.,  $q_i = 1/n$ , we have:

$$D_{KL}(\mathbf{p} \parallel \mathbf{u}_n) = \log n - S(\mathbf{p}) \quad , \quad (2)$$

where  $S(\mathbf{p})$  is the Shannon entropy,  $S(\mathbf{p}) = -\sum_i p_i \log p_i$ . Let us consider a set of linear constraints,  $\sum_j p_j E_{ij} = \langle E_i \rangle$ , and denote the configuration vector as  $E_i$ . Shannon entropy is maximized under this

set of linear constraints and the normalization condition, by functions belonging to the exponential family of probability distributions, which can be written as:

$$p_i(\boldsymbol{\theta}) \equiv p(\mathbf{E}_i; \boldsymbol{\theta}) = \exp(\Psi(\boldsymbol{\theta}) + \boldsymbol{\theta} \cdot \mathbf{E}_i) = \exp\left(\Psi(\boldsymbol{\theta}) + \sum_j \theta_j E_{ij}\right) . \quad (3)$$

$\Psi(\boldsymbol{\theta})$  guarantees normalization. Fisher information defines the metric on the parametric manifold  $\mathcal{M}$  by taking two infinitesimally-separated points,  $\boldsymbol{\theta}_0$  and  $\boldsymbol{\theta} = \boldsymbol{\theta}_0 + \delta\boldsymbol{\theta}$ , and by expanding  $D_{KL}(\mathbf{p}(\boldsymbol{\theta}_0) \parallel \mathbf{p}(\boldsymbol{\theta}))$ ,

$$g_{ij}^F(\boldsymbol{\theta}) = \left. \frac{\partial^2 D_{KL}(\mathbf{p}(\boldsymbol{\theta}_0) \parallel \mathbf{p}(\boldsymbol{\theta}))}{\partial \theta_i \partial \theta_j} \right|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0} . \quad (4)$$

For the exponential family of distributions, it is a well-known fact that Fisher information is equal to the inverse of the probability in Equation (3):

$$g_{ij}^F(\boldsymbol{\theta}) \equiv g_{ij}^F(\mathbf{p}(\boldsymbol{\theta})) = \frac{1}{p_i} \delta_{ij} + \frac{1}{p_0} . \quad (5)$$

## 2. Deformed Exponential Family

We briefly recall the definition of  $\phi$ -deformed logarithms and exponentials as introduced by Naudts [26]. The deformed logarithm is defined as:

$$\log_{\phi}(x) = \int_1^x dy \frac{1}{\phi(y)} , \quad (6)$$

for some positive, strictly-increasing function,  $\phi(x)$ , defined on  $(0, +\infty)$ . Then,  $\log_{\phi}$  is an increasing, concave function with  $\log_{\phi}(1) = 0$ .  $\log_{\phi}(x)$  is negative on  $(0, 1)$  and positive on  $(1, +\infty)$ . Naturally, the derivative of  $\log_{\phi}(x)$  is  $1/\phi(x)$ . The inverse function of  $\log_{\phi}(x)$  exists; we denote it by  $\exp_{\phi}(x)$ . Finally, the  $\phi$ -exponential family of probability distributions is defined as a generalization of Equation (3):

$$p_i(\boldsymbol{\theta}) = \exp_{\phi}\left(\Psi(\boldsymbol{\theta}) + \sum_j \theta_j E_{ij}\right) . \quad (7)$$

We can express  $\Psi(\boldsymbol{\theta})$  in the form:

$$\Psi(\boldsymbol{\theta}) = \log_{\phi}(p_i(\boldsymbol{\theta})) - \boldsymbol{\theta} \cdot \mathbf{E}_i , \quad (8)$$

which allows us to introduce dual coordinates to  $\boldsymbol{\theta}$ . This is nothing but the Legendre transform of  $\Psi(\boldsymbol{\theta})$ , which is defined as:

$$\varphi(\boldsymbol{\eta}) = \boldsymbol{\eta} \cdot \boldsymbol{\theta} - \Psi(\boldsymbol{\theta}) , \quad (9)$$

where:

$$\boldsymbol{\eta} = \nabla \Psi(\boldsymbol{\theta}) . \quad (10)$$

Because:

$$\partial_{\theta_j} p_i(\boldsymbol{\theta}) = \exp'_{\phi}(\Psi(\boldsymbol{\theta}) + \boldsymbol{\theta} \cdot \mathbf{E}_i) (\partial_{\theta_j} \Psi(\boldsymbol{\theta}) + E_{ij}) \quad (11)$$

holds, and using  $\sum_i \partial_{\theta_j} p_i(\boldsymbol{\theta}) = 0$ , we obtain that:

$$\eta_j = \frac{\sum_i \exp'_{\phi}(\Psi(\boldsymbol{\theta}) + \boldsymbol{\theta} \cdot \mathbf{E}_i) E_{ij}}{\sum_i \exp'_{\phi}(\Psi(\boldsymbol{\theta}) + \boldsymbol{\theta} \cdot \mathbf{E}_i)} = \mathbf{E}_j \cdot \mathbf{P}^{\phi} , \quad (12)$$

where  $P^\phi$  is the so-called escort distribution. With  $\exp'_\phi(\log_\phi(x)) = \phi(x)$ , the elements of  $P^\phi$  are given by:

$$p_j^\phi = \frac{\phi(p_j)}{\sum_i \phi(p_i)} = \frac{\phi(p_j)}{h_\phi(\mathbf{p})} \quad , \quad (13)$$

where we define  $h_\phi(\mathbf{p}) \equiv \sum_i \phi(p_i)$ . The Legendre transform provides a connection between the exponential family and the escort family of probability distributions, where the coordinates are obtained in the form of escort distributions. This generalizes the results for the ordinary exponential family of distributions, where the dual coordinates form a mixture family, which can be obtained as the superposition of the original distribution. The importance of dual coordinates in information geometry comes from the existence of a dually-flat geometry for the pair of coordinates. This means that there exist two affine connections with vanishing coefficients (Christoffel symbols). For the exponential family of distributions, the connection determined by the exponential distribution is called  $e$ -connection, and the dual connection leading to a mixture family is called  $m$ -connection [25]. For more details, see, e.g., [24]. We next look at generalizations of the Kullback–Leibler divergence and the Fisher information for the case of  $\phi$ -deformations.

### 3. Deformed Divergences, Entropies, and Metrics

For the  $\phi$ -deformed exponential family of distributions, we have to define the proper generalizations of the relevant quantities, such as the entropy, divergence, and metric. A natural approach is to start with the deformed Kullback–Leibler divergence, denoted by  $D_\phi(\mathbf{p}||\mathbf{q})$ .  $\phi$ -entropy can then be defined as:

$$S_\phi(\mathbf{p}) \sim -D_\phi(\mathbf{p}||\mathbf{u}_n) \quad , \quad (14)$$

where  $\sim$  means that the relation holds up to a multiplicative constant depending only on  $n$ . Similarly, the  $\phi$ -deformed Fisher information is:

$$g_{\phi,ij}(\boldsymbol{\theta}) = \left. \frac{\partial^2 D_\phi(p(\boldsymbol{\theta}_0)||p(\boldsymbol{\theta}))}{\partial \theta_i \partial \theta_j} \right|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0} \quad . \quad (15)$$

There is now more than one way to generalize the ordinary Kullback–Leibler divergence. The first is Csiszár’s divergence: [30]

$$I_f(\mathbf{p}||\mathbf{q}) = \sum_i q_i f(p_i/q_i) \quad , \quad (16)$$

where  $f$  is a convex function. For  $f(x) = x \ln x$ , we obtain the Kullback–Leibler divergence. Note, however, that the related information geometry based on the generalized Fisher information is trivial, because we have:

$$g_{ij}^f(\mathbf{p}) = f''(1)g_{ij}^F(\mathbf{p}) \quad , \quad (17)$$

i.e., the rescaled Fisher information metric; see [27]. The second possibility is to use the divergence of Bregman type, usually defined as:

$$D_f(\mathbf{p}||\mathbf{q}) = f(\mathbf{p}) - f(\mathbf{q}) - \nabla f(\mathbf{q}) \cdot (\mathbf{p} - \mathbf{q}) \quad , \quad (18)$$

where the symbol “ $\cdot$ ” denotes the scalar product. This type of divergence can be understood as the first-order Taylor expansion of  $f$  around  $\mathbf{q}$ , evaluated at  $\mathbf{p}$ . Let us next discuss the two possible types of the Bregman divergence, which naturally correspond to the  $\phi$ -deformed family of distributions. For both, the  $\phi$ -exponential family of distributions is obtained from the maximum entropy principle of the corresponding  $\phi$ -entropy, however, under different constraints. Note that the maximum entropy principle is just a special version of the more general minimal relative entropy principle, which minimizes the divergence functional  $D(\mathbf{p}||\mathbf{q})$  w.r.t.  $\mathbf{p}$ , for some given prior distribution  $\mathbf{q}$ .

### 3.1. Linear Constraints: Divergence a là Naudts

One generalization of the Kullback–Leibler divergence was introduced by Naudts [26] by considering  $f(\mathbf{p}) = \sum_i \left( \int_1^{p_i} \log_\phi(x) dx + (1 - p_i) \right)$ , which leads to:

$$D_\phi^N(\mathbf{p} \parallel \mathbf{q}) = \sum_j \int_{q_j}^{p_j} dx \left( \log_\phi(x) - \log_\phi(q_j) \right) \quad . \quad (19)$$

The corresponding entropy can be expressed as:

$$S_\phi^N(\mathbf{p}) = - \sum_j \int_0^{p_j} dx \ln_\phi(x) \quad . \quad (20)$$

$S_\phi^N(\mathbf{p})$  is maximized by the  $\phi$ -exponential family of distributions under linear constraints. The Lagrange functional is:

$$\mathcal{L}_\phi(\mathbf{p}) = S_\phi^N(\mathbf{p}) - \Psi \sum_i p_i - \sum_j \theta_j \sum_i p_i E_{ij} \quad , \quad (21)$$

which leads to:

$$- \log_\phi(p_i) - \Psi - \sum_j \theta_j E_{ij} = 0 \quad , \quad (22)$$

and we get:

$$\Psi(\boldsymbol{\theta}) = - \sum_i p_i \log_\phi(p_i) - \sum_j \theta_j \langle E_j \rangle = - \langle \log_\phi(\mathbf{p}) \rangle - \sum_j \theta_j \langle E_j \rangle \quad , \quad (23)$$

which is just Equation (8), averaged over the distribution  $p_i$ . Note that Equation (23) provides the connection to thermodynamics, because  $\Psi(\boldsymbol{\theta})$  is a so-called Massieu function. For a canonical ensemble, i.e., one constraint on the average energy  $\langle E \rangle$ , parameter  $\theta$  plays the role of an inverse temperature, and  $\Psi$  can be related to the free energy,  $F(\theta) = \theta \Psi(\theta)$ . Thus, the term  $\langle \log_\phi(\mathbf{p}) \rangle$  can be interpreted as the thermodynamic entropy, which is determined from Equation (23). This is a consequence of the Legendre structure of thermodynamics.

The corresponding MaxEnt distribution can be written as:

$$p_i(\boldsymbol{\theta}) = \exp_\phi \left( - \langle \log_\phi(\mathbf{p}) \rangle - \sum_j \theta_j (\langle E_j \rangle - E_{ij}) \right) = \exp_\phi (\Psi(\boldsymbol{\theta}) + \boldsymbol{\theta} \cdot \mathbf{E}_i) \quad . \quad (24)$$

Finally, the Fisher information metric can be obtained in the following form:

$$g_{\phi,ij}^N(\mathbf{p}) = \log'_\phi(p_i) \delta_{ij} + \log'_\phi(p_0) = \frac{1}{\phi(p_i)} \delta_{ij} + \frac{1}{\phi(p_0)} \quad . \quad (25)$$

### 3.2. Escort Constraints: Divergence a là Amari

Amari et al. [23,24] used a different divergence introduced in [31], which is based on the choice of  $f(\mathbf{p}) = \sum_i P_i^\theta \log_\phi(p_i)$ . This choice is motivated by the fact that the corresponding entropy is just the dual function of  $\Psi(\boldsymbol{\theta})$ , i.e.,  $\varphi(\boldsymbol{\eta})$ . This is easy to show, because:

$$\varphi(\boldsymbol{\eta}) = \boldsymbol{\eta} \cdot \boldsymbol{\theta} - \Psi(\boldsymbol{\theta}) = \sum_j P_j^\theta (\theta_j E_{ij} - \Psi(\boldsymbol{\theta})) = \sum_j P_j^\theta \log_\phi(p_j) \quad . \quad (26)$$

Thus, the divergence becomes:

$$D_\phi^A(\mathbf{p} \parallel \mathbf{q}) = \frac{1}{h_\phi(\mathbf{p})} \sum_j \phi(p_j) (\log_\phi(p_j) - \log_\phi(q_j)) \quad , \quad (27)$$

and the corresponding entropy can be expressed from Equation (26) as:

$$S_\phi^A(\mathbf{p}) = -\frac{1}{h_\phi(\mathbf{p})} \sum_j \phi(p_j) \log_\phi(p_j) \quad , \quad (28)$$

so it is a dual function of  $\Psi(\boldsymbol{\theta})$ . For this reason, the entropy is called “canonical”, because it is obtained by the Legendre transform from the Massieu function  $\Psi$ . Interestingly, the entropy is maximized by the  $\phi$ -exponential family of distributions under escort constraints. The Lagrange function is:

$$\mathcal{L}_\phi(\mathbf{p}) = S_\phi^A(\mathbf{p}) - \Psi \sum_i p_i - \sum_j \theta_j \sum_i P_i^\phi E_{ij} \quad . \quad (29)$$

After a straightforward calculation, we get:

$$\Psi(\boldsymbol{\theta}) = -\sum_i \phi(p_i(\boldsymbol{\theta})) \quad , \quad (30)$$

and the corresponding MaxEnt distribution can be expressed as:

$$p_i(\boldsymbol{\theta}) = \exp_\phi \left( -\langle \log_\phi(\mathbf{p}) \rangle_\phi - \sum_j \theta_j (\langle E_j \rangle_\phi - E_{ij}) \right) = \exp_\phi (\Psi(\boldsymbol{\theta}) + \boldsymbol{\theta} \cdot \mathbf{E}_i) \quad , \quad (31)$$

where:

$$\Psi(\boldsymbol{\theta}) = -\langle \log_\phi(\mathbf{p}) \rangle_\phi - \sum_j \theta_j \langle E_j \rangle_\phi \quad . \quad (32)$$

Here,  $\langle \cdot \rangle_\phi$  denotes the average under the escort probability measure,  $\mathbf{P}^\phi$ . Interestingly, in the escort constraints scenario, the “MaxEnt” entropy is the same as the “thermodynamic” entropy in the case of linear constraints. We call this entropy,  $S_\phi^A(\mathbf{p})$ , the dual entropy. Finally, one obtains the corresponding metric:

$$g_{\phi,ij}^A(\mathbf{p}) = -\frac{1}{h_\phi(\mathbf{p})} \left( \frac{\log_\phi''(p_i)}{\log_\phi'(p_i)} \delta_{ij} + \frac{\log_\phi''(p_0)}{\log_\phi'(p_0)} \right) = \frac{1}{h_\phi(\mathbf{p})} \left( \frac{\phi'(p_j)}{\phi(p_j)} \delta_{ij} + \frac{\phi'(p_0)}{\phi(p_0)} \right) \quad . \quad (33)$$

Note that the metric can be obtained from  $\Psi(\boldsymbol{\theta})$  as:  $g_{\phi,ij}^A(\boldsymbol{\theta}) = \frac{\partial^2 \Psi(\boldsymbol{\theta})}{\partial \theta_i \partial \theta_j}$ , which is the consequence of the Legendre structure of escort coordinates [24]. For a summary of the  $\phi$ -deformed divergence, entropy, and metric, see Table 1.

**Table 1.**  $\phi$ -deformation of divergence, entropy, and Fisher information corresponding to the  $\phi$ -exponential family under linear and escort constraints. For the ordinary logarithm,  $\phi(x) = x$ , the two entropies become the Shannon entropy, and the divergence is Kullback–Leibler.

|            | $\phi$ -Deformation              | Linear Constraints   | Escort Constraints  |
|------------|----------------------------------|--|---|
| divergence | $D_\phi(\mathbf{p}  \mathbf{q})$ | $\sum_j \int_{q_j}^{p_j} dx (\log_\phi(x) - \log_\phi(q_j))$ | $\frac{\sum_i \phi(p_i)(\log_\phi(p_i) - \log_\phi(q_i))}{\sum_k \phi(p_k)}$  |
| entropy    | $S_\phi(\mathbf{p})$             | $-\sum_i \int_0^{p_i} \log_\phi(x) dx$                       | $-\sum_i \phi(p_i) \log_\phi(p_i) / \sum_k \phi(p_k)$   |
| metric     | $g_{ij}^\phi(\mathbf{p})$        | $\frac{1}{\phi(p_i)} \delta_{ij} + \frac{1}{\phi(p_0)}$      | $\frac{1}{\sum_k \phi(p_k)} \left( \frac{\phi'(p_i)}{\phi(p_i)} \delta_{ij} + \frac{\phi'(p_0)}{\phi(p_0)} \right)$ |

### 3.3. Cramér–Rao Bound of Naudts Type

One of the important applications of the Fisher metric is the so-called Cramér–Rao bound, which is the lower bound for the variance of an unbiased estimator. The generalization of the Cramér–Rao bound for two families of distributions was given in [26,27]. Assume these two families of distributions to be denoted by  $\mathbf{p}(\boldsymbol{\theta})$  and  $\mathbf{P}(\boldsymbol{\theta})$ , with their corresponding expectation values,  $\langle \cdot \rangle_{\mathbf{p}(\boldsymbol{\theta})}$  and  $\langle \cdot \rangle_{\mathbf{P}(\boldsymbol{\theta})}$ . Let  $c_k$

denote the estimator of the family  $\mathbf{p}(\boldsymbol{\theta})$  that fulfills  $\langle c_k \rangle_{\mathbf{p}(\boldsymbol{\theta})} = \frac{\partial}{\partial \theta_k} f(\boldsymbol{\theta})$ , for some function  $f$ , and let us consider a mild regularity condition  $\left\langle \frac{1}{P(\boldsymbol{\theta})} \frac{\partial}{\partial \theta_k} \mathbf{p}(\boldsymbol{\theta}) \right\rangle_{P(\boldsymbol{\theta})} = 0$ . Then:

$$\frac{\langle c_k c_l \rangle_{P(\boldsymbol{\theta})} - \langle c_k \rangle_{P(\boldsymbol{\theta})} \langle c_l \rangle_{P(\boldsymbol{\theta})}}{\left( \frac{\partial^2 f(\boldsymbol{\theta})}{\partial \theta_k \partial \theta_l} \right)^2} \geq \frac{1}{I_{kl}(\boldsymbol{\theta})} \quad , \tag{34}$$

where:

$$I_{kl}(\boldsymbol{\theta}) = \sum_i \frac{1}{P_i(\boldsymbol{\theta})} \frac{\partial p_i(\boldsymbol{\theta})}{\partial \theta_k} \frac{\partial p_i(\boldsymbol{\theta})}{\partial \theta_l} \quad . \tag{35}$$

If  $\mathbf{p}(\boldsymbol{\theta}) = \mathbf{p}_\phi(\boldsymbol{\theta})$  is the  $\phi$ -exponential family of distributions, in Equation (34), equality holds for the escort distribution  $P(\boldsymbol{\theta}) = P^\phi(\boldsymbol{\theta})$ , [28]. It is easy to see that for this case, i.e., for the  $\phi$ -exponential family and the corresponding escort distribution, the following is true:

$$I_{\phi;kl}^N(\mathbf{p}) = h_\phi(\mathbf{p}) g_{\phi;kl}^N(\mathbf{p}) \quad . \tag{36}$$

This provides a connection between the Cramér–Rao bound and the  $\phi$ -deformed Fisher metric. In the next section, we show that the Cramér–Rao bound can be also estimated for the case of the Fisher metric of the “Amari type”.

#### 4. The Information Geometric “Amari–Naudts” Duality

In the previous section, we have seen that there are at least two natural ways to generalize divergence, such that the  $\phi$ -exponential family of distributions maximizes the associated entropy functional, however under different constraint types. These two ways result in two different geometries on the parameter manifold. The relation between the metric  $g_{\phi,ij}^A$  and  $g_{\phi,ij}^N$  can be expressed by the operator,  $T$ :

$$g_{ij}^A(p) = T(g_{ij}^N(p)) \quad , \tag{37}$$

where

$$T(g(x)) = -N_g (\log g(x))' \quad , \tag{38}$$

with the normalization factor,  $N_g = \sum_i 1/g(p_i)$ . Note that the operator acts locally on the elements of the metric. In order to establish the connection to the Cramér–Rao bound, let us focus on the transformation of  $g^A$ .

##### 4.1. Cramér–Rao Bound of the Amari Type

The metric of the “Amari case” can be seen as a conformal transformation [32] of the metric that is obtained in the “Naudts case”, for a different deformation of the logarithm. Two metric tensors are connected by a conformal transformation if they have the same form, except for the global conformal factor,  $\Omega(p)$ , which depends only on the point  $p$ . Our aim is to connect the Amari metric with the Cramér–Rao bound and obtain another type of bound for the estimates that are based on escort distributions. To this end, let us consider a general metric of the Naudts type, corresponding to  $\chi$ -deformation, and a metric of the Amari type, corresponding to  $\zeta$ -deformation. They are connected through the conformal transformation, which acts globally on the whole metric. The relation can be expressed as:

$$g_{\chi,ij}^N(\mathbf{p}) = \Omega(\mathbf{p}) g_{\zeta,ij}^A(\mathbf{p}) \quad . \tag{39}$$

By using previous results in this relation, we obtain:

$$\frac{1}{\chi(p_i)} = \frac{\Omega(\mathbf{p}) \zeta'(p_i)}{h_\zeta(\mathbf{p}) \zeta(p_i)} \quad , \tag{40}$$



from which we see that  $\Omega(\mathbf{p}) = h_{\zeta}(\mathbf{p})$  and  $\log_{\chi}(x) = \log(\zeta(x))$ , i.e.,

$$\zeta(x) = \exp(\log_{\chi}(x)) \Rightarrow \log_{\zeta}(x) = \int_1^x \exp(-\log_{\chi}(y)) dy \quad (41)$$

Note that  $\log_{\chi}$  might not be concave because:

$$\frac{d^2}{dx^2} \log_{\chi}(x) = \frac{\zeta(x)\zeta''(x) - \zeta'(x)^2}{\zeta(x)^2} \quad (42)$$

Concavity exists, if  $\zeta''(x) \leq \frac{\zeta'(x)^2}{\zeta(x)}$ . To now make the connection with the Cramér–Rao bound, let us take  $\chi(x) = \phi(x)$ , so  $\zeta(x) = \exp \log_{\phi}(x)$ , and:

$$I_{\phi;kl}^A(\mathbf{p}) = h_{\exp(\log_{\phi})}(\mathbf{p}) g_{\exp(\log_{\phi});kl}^A(\mathbf{p}) \quad (43)$$

As a consequence, there exist two types of Cramér–Rao bound for a given escort distribution, which might be used to estimate the lower bound of the variance of an unbiased estimator, obtained from two types of Fisher information.

#### 4.2. Example: Duality of (c, d)-Entropy

We demonstrate the “Amari–Naudts” duality on the general class of (c, d)-entropies [9,10], which include all deformations associated with statistical systems that fulfil the first three Shannon–Khinchin axioms. These include most of the popular deformations, including Tsallis q-exponentials [4] and stretched exponentials studied in connection with the entropies by Anteneodo and Plastino [33]. The generalized (c, d)-logarithm is defined as:

$$\log_{(c,d)}(x) = r - rx^{c-1} \left( 1 - \frac{(1 - (1 - c)r}{dr} \log x \right)^d, \quad (44)$$

where c and d are the scaling exponents [8,9] and r is a free scale parameter (that does not influence the asymptotic behavior). The associated  $\phi$ -deformation is:

$$\phi_{(c,d)}(x) = \frac{x}{r - \log_{c,d}(x)} \left( \frac{(-cr + r - 1) \log(x) + dr}{(c - 1)((c - 1)r + 1) \log(x) + d} \right) \quad (45)$$

The inverse function of  $\log_{(c,d)}$ , the deformed (c, d)-exponential, can be expressed in terms of the Lambert W-function, which is the solution of equation,  $W(z)e^{W(z)} = z$ . The deformed (c, d)-exponential is:

$$\exp_{(c,d)}(x) = \exp \left( -\frac{d}{1 - c} \left[ W \left( B(1 - x/r)^{1/d} \right) - W(B) \right] \right), \quad (46)$$

where  $B = \frac{(1-c)r}{1-(1-c)r} \exp \left( \frac{(1-c)r}{1-(1-c)r} \right)$ . The corresponding entropy that is maximized by (c, d)-exponentials (see [8] for their properties) is the (c, d)-entropy:

$$S_{(c,d)}(\mathbf{p}) = rA^{-d}e^A \sum_i \Gamma(1 + d, A - c \ln p_i) - rc \quad (47)$$



where  $A = \frac{cdr}{1-(1-c)r}$ . This is an entropy of “Naudts type”, since it is maximized with  $(c, d)$ -exponentials under linear constraints. We can immediately write the metric as:

$$g_{(c,d),ij}^N(\mathbf{p}) = \left( \frac{r - \log_{c,d}(p_i)}{p_i} \left( \frac{(c-1)((c-1)r+1)\log(p_i)+d}{(-cr+r-1)\log(p_i)+dr} \right) \right) \delta_{ij} + \frac{r - \log_{c,d}(p_0)}{p_0} \left( \frac{(c-1)((c-1)r+1)\log(p_i)+d}{(-cr+r-1)\log(p_0)+dr} \right) . \tag{48}$$

The corresponding entropy of “Amari type”, i.e., maximized with  $(c, d)$ -exponentials under escort constraints:

$$\frac{\sum_i \phi_{(c,d)}(p_i) E_i}{\sum_j \phi_{(c,d)}(p_j)} = \langle E \rangle_{(c,d)} , \tag{49}$$

is:

$$S_{(c,d)}^A(\mathbf{p}) = -\frac{1}{h_{(c,d)}(\mathbf{P})} \sum_i \frac{p_i \log_{c,d} p_i}{r - \log_{c,d}(p_i)} \left( \frac{(-cr+r-1)\log(p_i)+dr}{(c-1)((c-1)r+1)\log(p_i)+d} \right) , \tag{50}$$

and its metric is:

$$g_{(c,d),ij}^A(\mathbf{p}) = \frac{1}{p_i} \left( 2-c - \frac{(d-1)((c-1)r+1)}{((c-1)r+1)\log(p_i)-dr} - \frac{(c-1)^2r+c-1}{(c-1)dr-cdr+(c-1)((c-1)r+1)\log(p_i)+d+dr} \right) \delta_{ij} + \frac{1}{p_0} \left( 2-c - \frac{(d-1)((c-1)r+1)}{((c-1)r+1)\log(p_0)-dr} - \frac{(c-1)^2r+c-1}{(c-1)dr-cdr+(c-1)((c-1)r+1)\log(p_0)+d+dr} \right) . \tag{51}$$

The metric of the Amari type of the  $(c, d)$ -entropy was already discussed in [34] based on  $(c, d)$ -logarithms. However, as demonstrated above, the metric can be found without using the inverse  $\phi$ -deformed logarithms, which in the case of  $(c, d)$ -logarithms lead to Lambert  $W$ -functions. The Fisher metric of Naudts and Amari type and the corresponding Cramér–Rao bound are shown in Figure 1. The scaling parameter is set (following [9]) to  $r = 1/(1-c+cd)$ , for  $d \leq 0$ , and  $r = \exp(-d)/(1-c)$ , for  $d < 0$ . The Fisher metric of both types is displayed in Figure 2 as a function of the parameters  $c$  and  $d$  for a given point,  $P = (1/3, 2/3)$ . We see that both types of metric have a singularity for  $(c, d) = (1, 0)$ . This point corresponds to distributions with compact support. For one-dimensional distributions, the singularity corresponds to the transition between distributions with support on the real line and distributions with support on a finite interval.

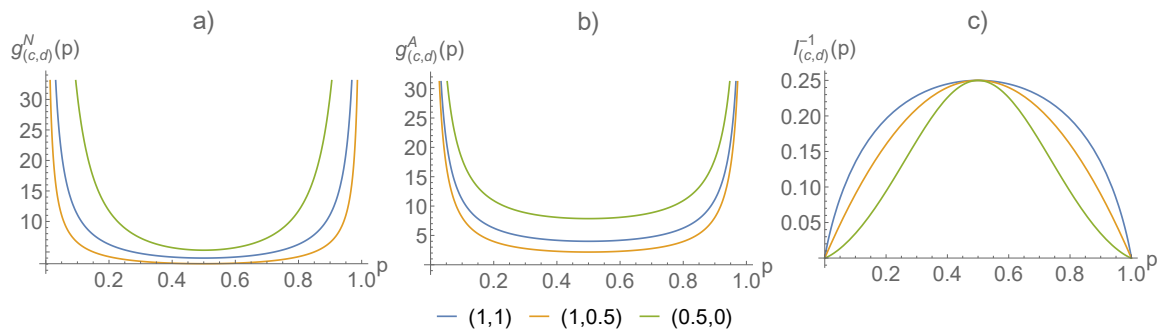
Interestingly, for  $(c, d) = (q, 0)$ , the metric simplifies to:

$$g_{(q,0),ij}^A(\mathbf{p}) = \frac{2-q}{p_i} \delta_{ij} + \frac{2-q}{p_0} , \tag{52}$$

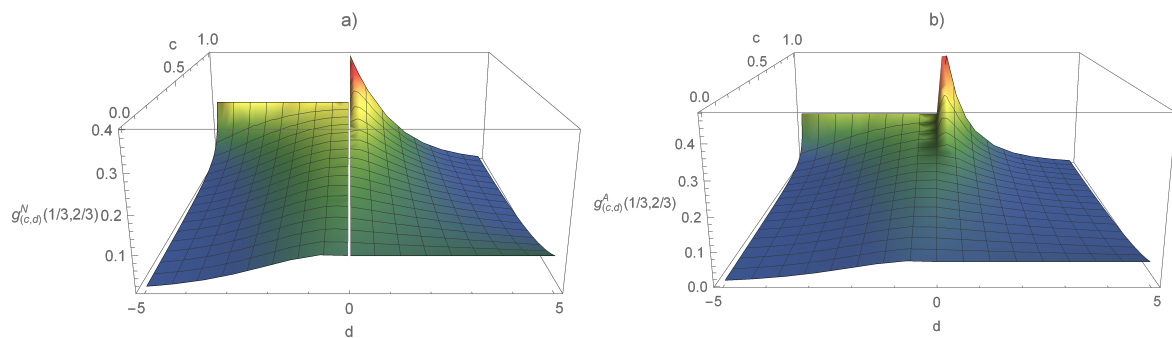
which corresponds to the Tsallis  $q$ -exponential family of distributions. Therefore,  $g_{(q,0),ij}^A(\mathbf{p})$  is just a conformal transformation of the Fisher information metric for the exponential family of distributions, as shown in [24]. Note, that only for Tsallis  $q$ -exponentials, the relation between  $S_q^N(\mathbf{p})$  and  $S_q^A(\mathbf{p})$  can be expressed as (see also Table 2):

$$S_q^A(\mathbf{p}) = f(S_{q'}^N(\mathbf{p})) , \tag{53}$$

where  $f(x) = (2-q)/x$  and  $q' = 2-q$ . This is nothing but the well-known additive duality  $q \leftrightarrow 2-q$  of Tsallis entropy [11]. Interestingly,  $q$ -escort distributions form a group with  $\phi_q(\phi_{q'}(x)) = (\phi_{q \cdot q'}(x))$  and  $\phi_q^{-1}(x) = \phi_{1/q}(x)$ , where  $q \leftrightarrow 1/q$  is the multiplicative duality [35].



**Figure 1.** Fisher metric for  $p = (p, 1 - p)$  corresponding to various  $(c, d)$ -deformations  $((c, d) = (1, 1), (1, 1/2), (1/2, 0))$  for (a) the Naudts type, (b) the Amari type, and (c) the Cramér–Rao bound corresponding to the metric.



**Figure 2.** Fisher metric for  $(c, d)$ -deformations as a function of  $c$  and  $d$  of Naudts type (a) and Amari type (b). The metric is evaluated at a point  $p = (1/3, 2/3)$ .

This is not the case for more general deformations, because typically, the inverse does not belong to the class of escort distributions. Popular deformations belonging to the  $(c, d)$ -family, as the Tsallis  $q$ -exponential family or the stretched exponential family, are summarized in Table 2.

**Table 2.** Two important special cases of  $(c, d)$ -deformations and related quantities: Power laws (Tsallis) and stretched exponentials.

|                        | Tsallis $q$ -Exponential [4]                                    | Stretched $\eta$ -Exponential [33]                             |
|------------------------|---|--|
| $\phi(x)$              | $x^q$   | $x\eta \log(x)^{1-1/\eta}$                                     |
| $\log_\phi(x)$         | $\frac{x^{1-q}-1}{1-q}$   | $\log(x)^{1/\eta}$   |
| $\exp_\phi(x)$         | $(1 + (1 - q)x)^{1/(1-q)}$                                      | $\exp(x^\eta)$   |
| $\chi_\phi(x)$         | $\frac{x}{q}$   | $\frac{x\eta \log(x)}{(\eta-1)+\eta \log(x)}$                  |
| $S_\phi^N(\mathbf{p})$ | $\frac{1}{q-1} \left( \sum_i \frac{p_i^{2-q}}{2-q} - 1 \right)$ | $\sum_i \Gamma \left( \frac{\eta+1}{\eta}, -\log p_i \right)$  |
| $S_\phi^A(\mathbf{p})$ | $\frac{1}{q-1} \left( \frac{1}{\sum_i p_i^q} - 1 \right)$       | $\frac{\sum_i p_i \log p_i}{\sum_i p_i (\log p_i)^{1-1/\eta}}$ |

### 5. Connection to the Deformed-Log Duality

A different duality of entropies and their associated logarithms under linear and escort averages has been discussed in [22]. There, two approaches were discussed. The first uses the generalized entropy of trace form under linear constraints. It was denoted by:

$$S^{HT}(\mathbf{p}) = \sum_i s^{HT}(p_i) = \sum_i \int_0^{p_i} \log^{HT}(x) dx \quad . \tag{54}$$

It corresponds to the Naudts case,  $\log^{HT}(x) = \log_{\phi^{HT}}^N(x)$ . The second approach, originally introduced by Tsallis and Souza [20], uses the trace form entropy:

$$S^{TS}(\mathbf{p}) = \sum_i s^{TS}(p_i) = \sum_i \int_0^{p_i} \log^{TS}(x) dx \quad . \tag{55}$$

Note that in [20], the notion of the deformed logarithm is not used (as in Equation (55)). However, it is again an entropy of the Naudts type with the deformed logarithm  $\log^{TS}(x) = \log_{\phi^{TS}}^N(x)$ . Equation (55) is maximized under the escort constraints:

$$\langle E \rangle_{TS} = \frac{\sum_j u(p_j) E_j}{\sum_j u(p_j)} \quad , \tag{56}$$

where  $u(p_j) = p_j + \nu s^{TS}(p_j)$ . The linear constraints are recovered for  $\nu = 0$ . This form is dictated by the Shannon–Khinchin axioms, as discussed in the next section. Let us assume that the maximization of both approaches—Equation (54) under linear, and Equation (55) under escort constraints—leads to the same MaxEnt distribution. One can then show that there exists the following duality (deformed-log duality) between  $\log^{HT}(x)$  and  $\log^{TS}(x)$ :

$$\log^{TS}(x) = \frac{1}{\frac{1}{\log^{HT}(x)} + \nu} \quad . \tag{57}$$

Let us focus on specific  $\phi$ -deformations, so that  $\log^{HT}(x) = \log_{\phi}^{HT}(x) \equiv \log_{\phi}^N(x)$ . Then,  $\log^{TS}(x)$  is also a  $\phi$ -deformation, with:

$$\phi^{TS}(x) = \phi^{HT}(x) \left( 1 + \nu \log_{\phi}^{HT}(x) \right)^2 = \phi^{HT}(x) \left( 1 + \nu \int_1^x \frac{1}{\phi^{HT}(y)} dy \right)^2 \quad . \tag{58}$$

It is straightforward to calculate the metric corresponding to the entropy  $S^{TS}(\mathbf{p})$ :

$$g_{\phi,ij}^{TS}(\mathbf{p}) = \frac{1}{\phi^{TS}(p_i)} \delta_{ij} + \frac{1}{\phi^{TS}(p_0)} = \mathcal{T}_{\nu}(g_{\phi,ij}^{HT}(\mathbf{p})) \quad ,$$

where:

$$\mathcal{T}_{\nu}(g(x)) = g(x) \left( 1 + \nu \int_1^x g(y) dy \right)^2 \quad . \tag{59}$$

Thus, the Tsallis–Souza approach results in yet another information matrix. We may also start from the other direction and look at the situation when the escort distribution for the information geometric approach is the same as the escort distribution for the Tsallis–Souza approach. In this case, we get:

$$\phi(x) = x + \nu s^{TS}(x) = x + \int_0^x \mathcal{T}_{\nu}(\log_{\phi}(y)) dy \quad . \tag{60}$$

We find that the entropy must be expressed as:

$$S_{\phi}(\mathbf{p}) = \sum_i \frac{\phi(p_i) - p_i}{\nu} \quad . \tag{61}$$

Note that for  $\phi(x) = x^q$  and  $\nu = 1 - q$ , we obtain the Tsallis entropy:

$$S_{\phi}(\mathbf{p}) = \sum_i \frac{p_i^q - p_i}{1 - q} \quad , \tag{62}$$

which corresponds to  $S^{TS}(\mathbf{p})$  for  $q' = 2 - q$ , which is nothing but the mentioned Tsallis additive duality. It turns out that Tsallis entropy is the only case where the deformed-log duality and the information geometric duality result in the same class of functionals. In general, the two dualities have different escort distributions.

## 6. Discussion

We discussed the information geometric duality of entropies that are maximized by  $\phi$ -exponential distributions under two types of constraint: linear constraints, which are known from contexts such as thermodynamics, and escort constraints, which appear naturally in the theory of statistical estimation and information geometry. This duality implies two different entropy functionals:  $S^N(\mathbf{p}) = -\sum_i \int_0^{p_i} dx \log_\phi(x)$  and  $S^A(\mathbf{p}) = -\langle \log_\phi(P) \rangle_\phi$ . For  $\phi(x) = x$ , they both boil down to Shannon entropy. The connection between the entropy of Naudts type and the one of Amari type can be established through the corresponding Fisher information through the Cramér–Rao bound. Contrary to the deformed-log duality introduced in [22], the information theoretic duality introduced here cannot be established within the framework of  $\phi$ -deformations, since  $S^A(\mathbf{p})$  is not a trace form entropy. We demonstrated the duality between the Naudts approach with linear constraints and the Amari approach with escort constraints, within the example of  $(c, d)$ -entropies, which include a wide class of popular deformations, including Tsallis and Anteneodo–Plastino entropy as special cases. Finally, we compared the information geometric duality to the deformed-log duality and showed that they are fundamentally different and result in other types of Fisher information.

Let us now discuss the role of information geometric duality and possible applications in information theory and thermodynamics. Recall that the Shannon entropic functional is determined by the four Shannon–Khinchin (SK) axioms. In many contexts, at least three of the axioms should hold:

- (SK1) Entropy is a continuous function of the probabilities  $p_i$  only and should not explicitly depend on any other parameters.
- (SK2) Entropy is maximal for the equi-distribution  $p_i = 1/W$ .
- (SK3) Adding a state  $W + 1$  to a system with  $p_{W+1} = 0$  does not change the entropy of the system.

Originally, the Shannon–Khinchin axioms contain four axioms. The fourth describes the “composition rule” for entropy of a joint system  $S(A + B) = S(A) + S(B|A)$ . The only entropy satisfying all four SK axioms is Shannon entropy. However, Shannon entropy is not sufficient to describe the statistics of complex systems [10] and can lead to paradoxes in applications in thermodynamics [12]. Therefore, instead of imposing the fourth axiom in situations where it does not apply, it is convenient to consider a weaker requirement, such as generic scaling relations of entropy in the thermodynamic limit [9,13]. It is possible to show that the only type of duality satisfying the first three Shannon–Khinchin axioms is the deformed-log duality of [22]. Moreover, entropies that are neither trace-form, nor sum-form (*sum-form* entropies are in the form  $S(\mathbf{p}) = f(\sum_i g(p_i))$ ) might be problematic from the view of information theory and coding. For example, it is then not possible to introduce a conditional entropy consistently [36] because the corresponding conditional entropy cannot be properly defined. This is related to the fact that the Kolmogorov definition of conditional probability is not generally valid for escort distributions [37]. Additional issues arise from the theory of statistical estimation, since only entropies of the form  $S(\mathbf{p}) = f(\sum_i g(p_i))$ , i.e., sum-form entropies, can fulfil the consistency axioms [38]. From this point of view, the deformed-log duality using the class of Tsallis–Souza escort distributions can play a role in thermodynamical applications [39], because the corresponding entropy fulfills the SK axioms. On the other hand, the importance of escort distributions considered by Amari and others is in the realm of information geometry (e.g., dually-flat geometry or generalized Cramér–Rao bound), and their applications in thermodynamics might be limited. Finally, for the case of the Tsallis  $q$ -deformation, both dualities, the information geometric and the deformed-log duality, reduce to the well-known additive duality  $q \leftrightarrow 2 - q$ .

**Author Contributions:** All authors contributed to the conceptualization of the work, the discussion of the results, and their interpretation. J.K. took the lead in technical computations. J.K. and S.T. wrote the manuscript.

**Funding:** This work was supported by the Austrian Science Fund FWFunder I 3073.

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

## References

1. Thurner, S.; Corominas-Murtra, B.; Hanel, R. Three faces of entropy for complex systems: Information, thermodynamics, and the maximum entropy principle. *Phys. Rev. E* **2017**, *96*, 032124. [[CrossRef](#)] [[PubMed](#)]
2. Jaynes, E.T. Information theory and statistical mechanics. *Phys. Rev.* **1957**, *106*, 620. [[CrossRef](#)]
3. Harremoës, P.; Topsøe, F. Maximum entropy fundamentals. *Entropy* **2001**, *3*, 191–226. [[CrossRef](#)]
4. Tsallis, C. Possible generalization of Boltzmann–Gibbs statistics. *J. Stat. Phys.* **1988**, *52*, 479–487. [[CrossRef](#)]
5. Kaniadakis, G. Statistical mechanics in the context of special relativity. *Phys. Rev. E* **2002**, *66*, 056125. [[CrossRef](#)] [[PubMed](#)]
6. Jizba, P.; Arimitsu, T. The world according to Rényi: Thermodynamics of multifractal systems. *Ann. Phys.* **2004**, *312*, 17–59. [[CrossRef](#)]
7. Tsallis, C.; Cirto, L.J. Black hole thermodynamical entropy. *Eur. Phys. J. C* **2013**, *73*, 2487. [[CrossRef](#)]
8. Thurner, S.; Hanel, R.; Klimek, P. *Introduction to the Theory of Complex Systems*; Oxford University Press: Oxford, UK, 2018.
9. Hanel, R.; Thurner, S. A comprehensive classification of complex statistical systems and an axiomatic derivation of their entropy and distribution functions. *Europhys. Lett.* **2011**, *93*, 20006. [[CrossRef](#)]
10. Hanel, R.; Thurner, S. When do generalized entropies apply? How phase space volume determines entropy. *Europhys. Lett.* **2011**, *96*, 50003. [[CrossRef](#)]
11. Tsallis, C.; Gell-Mann, M.; Sato, Y. Asymptotically scale-invariant occupancy of phase space makes the entropy  $S_q$  extensive. *Proc. Natl. Acad. Sci. USA* **2005**, *102*, 15377–15382. [[CrossRef](#)]
12. Jensen, H.J.; Pazuki, R.H.; Pruessner, G.; Tempesta, P. Statistical mechanics of exploding phase spaces: Ontic open systems. *J. Phys. A* **2018**, *51*, 375002. [[CrossRef](#)]
13. Korbel, J.; Hanel, R.; Thurner, S. Classification of complex systems by their sample-space scaling exponents. *New J. Phys.* **2018**, *20*, 093007. [[CrossRef](#)]
14. Tsallis, C.; Mendes, R.S.; Plastino, A.R. The role of constraints within generalized nonextensive statistics. *Physica A* **1998**, *261*, 534–554. [[CrossRef](#)]
15. Abe, S. Geometry of escort distributions. *Phys. Rev. E* **2003**, *68*, 031101. [[CrossRef](#)] [[PubMed](#)]
16. Ohara, A.; Matsuzoe, H.; Amari, S.I. A dually flat structure on the space of escort distributions. *J. Phys. Conf. Ser.* **2010**, *201*, 012012. [[CrossRef](#)]
17. Bercher, J.-F. A simple probabilistic construction yielding generalized entropies and divergences, escort distributions and q-Gaussians. *Physica A* **2012**, *391*, 4460–4469. [[CrossRef](#)]
18. Hanel, R.; Thurner, S.; Tsallis, C. On the robustness of q-expectation values and Renyi entropy. *Europhys. Lett.* **2009**, *85*, 20005. [[CrossRef](#)]
19. Hanel, R.; Thurner, S.; Tsallis, C. Limit distributions of scale-invariant probabilistic models of correlated random variables with the q-Gaussian as an explicit example. *Eur. Phys. J. B* **2009**, *72*, 263–268. [[CrossRef](#)]
20. Tsallis, C.; Souza, A.M.C. Constructing a statistical mechanics for Beck-Cohen superstatistics. *Phys. Rev. E* **2003**, *67*, 026106. [[CrossRef](#)]
21. Beck, C.; Cohen, E.D.G. Superstatistics. *Physica A* **2003**, *322*, 267–275. [[CrossRef](#)]
22. Hanel, R.; Thurner, S.; Gell-Mann, M. Generalized entropies and logarithms and their duality relations. *Proc. Natl. Acad. Sci. USA* **2012**, *109*, 19151–19154. [[CrossRef](#)]
23. Amari, S.I.; Cichocki, A. Information geometry of divergence functions. *Bull. Pol. Acad. Sci. Tech. Sci.* **2010**, *58*, 183–195. [[CrossRef](#)]
24. Amari, S.I.; Ohara, A.; Matsuzoe, H. Geometry of deformed exponential families: Invariant, dually-flat and conformal geometries. *Physica A* **2012**, *391*, 4308–4319. [[CrossRef](#)]
25. Ay, N.; Jost, J.; Le, H.V.; Schwachhöfer, L. *Information Geometry*; Springer: Berlin, Germany, 2017.
26. Naudts, J. Deformed exponentials and logarithms in generalized thermostatics. *Physica A* **2002**, *316*, 323–334. [[CrossRef](#)]

27. Naudts, J. Continuity of a class of entropies and relative entropies. *Rev. Math. Phys.* **2004**, *16*, 809–822. [[CrossRef](#)]
28. Naudts, J. *Generalised Thermostatistics*; Springer Science & Business Media: Berlin, Germany, 2011.
29. Pistone, G.; Sempì, 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. [[CrossRef](#)]
30. Csiszar, I. Why least squares and maximum entropy? An axiomatic approach to inference for linear inverse problems. *Ann. Stat.* **1991**, *19*, 2032–2066. [[CrossRef](#)]
31. Vigelis, R.F.; Cavalcante, C.C. On  $\phi$ -Families of probability distributions. *J. Theor. Probab.* **2013**, *26*, 870–884. [[CrossRef](#)]
32. Ohara, A. Conformal flattening for deformed information geometries on the probability simplex. *Entropy* **2018**, *20*, 186. [[CrossRef](#)]
33. Anteneodo, C.; Plastino, A.R.I Maximum entropy approach to stretched exponential probability distributions. *J. Phys. A* **1999**, *32*, 1089. [[CrossRef](#)]
34. Ghikas, D.P.K.; Oikonomou, F.D. Towards an information geometric characterization/classification of complex systems. I. Use of generalized entropies. *Physica A* **2018**, *496*, 384–398. [[CrossRef](#)]
35. Tsallis, C. Generalization of the possible algebraic basis of  $q$ -triplets. *Eur. Phys. J. Spec. Top.* **2017**, *226*, 455–466. [[CrossRef](#)]
36. Ilić, V.M.; Stanković, M.S. Generalized Shannon–Khinchin axioms and uniqueness theorem for pseudo-additive entropies. *Physica A* **2014**, *411*, 138–145. [[CrossRef](#)]
37. Jizba, P.; Korbel, J. On the uniqueness theorem for pseudo-additive entropies. *Entropy* **2017**, *19*, 605. [[CrossRef](#)]
38. Uffink, J. Can the maximum entropy principle be explained as a consistency requirement? *Stud. Hist. Philos. Sci. B* **1995**, *26*, 223–261. [[CrossRef](#)]
39. Hanel, R.; Thurner, S.; Gell-Mann, M. Generalized entropies and the transformation group of superstatistics. *Proc. Natl. Acad. Sci. USA* **2011**, *108*, 6390–6394. [[CrossRef](#)]



© 2019 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/>).