A simple probabilistic construction yielding generalized entropies and divergences, escort distributions and q-Gaussians

J.-F. Bercher

Université Paris-Est, LIGM, UMR CNRS 8049, ESIEE-Paris
5 bd Descartes, 77454 Marne la Vallée Cedex 2, France
tel: 33-1-45-92-65-15 fax: 33-1-45-92-66-99

Abstract

We give a simple probabilistic description of a transition between two states which leads to a generalized escort distribution. When the parameter of the distribution varies, it defines a parametric curve that we call an escort-path. The Rényi divergence appears as a natural by-product of the setting. We study the dynamics of the Fisher information on this path, and show in particular that the thermodynamic divergence is proportional to Jeffreys' divergence. Next, we consider the problem of inferring a distribution on the escort-path, subject to generalized moments constraints. We show that our setting naturally induces a rationale for the minimization of the Rényi information divergence. Then, we derive the optimum distribution as a generalized Gaussian distribution.

Keywords: Divergence measures, Generalized Rényi and Tsallis entropies, Escort distributions, q-gaussian distributions
PACS: 02.50.-r, 05.90.+m, 89.70.+c

1. Introduction

In this paper, we give a simple probabilistic description of a transition between two states, which leads to a parametric curve in the form of a generalized escort distribution. We call escort-path this parametric curve. In this setting, we show that the Rényi information divergence emerges naturally as a characterization of the transition. Along this escort-path, we study the Fisher information. In particular, we show that the thermodynamic divergence on the escort-path is proportional to Jeffreys' divergence. Finally, we consider the inference of a distribution subject to moments computed with respect to the escort distribution. First, we show that our setting leads to a rationale for the minimization of the Rényi information divergence. Then, we derive the optimum distribution as a generalized Gaussian distribution.

Before going into the details of the results, we shall present the context and introduce the main definitions on our main ingredients, that is the escort distributions,

---

Email address: `jf.bercher@esiee.fr` (J.-F. Bercher)

information divergences, and Fisher information.

Throughout the paper, we will work with univariate probability densities defined with respect to a general measure $\mu(x)$ on a set X. For instance, the Shannon-Boltzmann entropy will be expressed as

$$H[f] = -\int f(x) \log f(x) \mathrm{d}\mu(x). \tag{1}$$

As particular cases, we have that if X is the real line and μ the Lebesgue measure, then the expression above corresponds to the differential entropy. When the set X is $\mathbb{N}$ or a subset of $\mathbb{N}$ and μ the counting measure, then the expression reduces to the standard discrete entropy. When μ is a probability measure, then the expression (1) can also be seen as the relative entropy from the measure with density f to the measure μ.

Let us now turn to the notion of escort distribution. If $f(x)$ is an univariate probability density with respect to $\mu(x)$, then we define its escort distribution of order q, $q \geq 0$, by

$$f_q(x) = \frac{f(x)^q}{\int f(x)^q \mathrm{d}\mu(x)}, \tag{2}$$

provided that $M_q[f] = \int f(x)^q \mathrm{d}\mu(x)$ is finite. These escort distributions have been introduced as an operational tool in the context of multifractals [1], [2], with interesting connections with the standard thermodynamics. Discussion of their geometric properties can be found in [3, 4]. Escort distributions also prove to be useful in source coding where they enable to derive optimum codewords with a length bounded by the Rényi entropy [5].

The results presented in this paper are connected to the nonextensive statistical physics introduced by Tsallis, see e.g. [6]. Indeed, the nonextensive statistical physics uses a generalized entropy, makes use of escort distributions and exhibit generalized Gaussians. All these elements will pop up in our construction, which, therefore could lead to new viewpoints or interpretations in this context. It is particularly remarkable that the derivation of the maximum Tsallis entropy distributions in nonextensive thermostatistics requires a constraint in the form of an "escort mean value", that is computed with respect to an escort distribution like (2) [7, 8].

One can immediately extend the notion of escort distribution to deal with two probability densities $f(x)$ and $g(x)$ as follows.

Definition 1. Let f and g be two densities with respect to a common measure μ, with g dominated by f. For $q \geq 0$ such that $M_q[f,g] = \int f(x)^q g(x)^{1-q} d\mu(x) < \infty$, we call generalized escort distribution the function

$$f_q(x) = \frac{f(x)^q g(x)^{1-q}}{\int f(x)^q g(x)^{1-q} \mathrm{d}\mu(x)}. \tag{3}$$

We will also denote, when non ambigous, by $E_q[.]$ the statistical expectation with respect to the generalized escort distribution with index q.

This generalized escort distribution is simply a weighted geometric mean of $f(x)$ and $g(x)$, and reduces to $f_q(x) = f(x)$ for $q = 1$ and to $f_q(x) = g(x)$ for $q = 0$. Obviously, if $g(x)$ is a uniform density whose support includes the support of $f(x)$, then the generalized escort distribution gives back the standard one (2). Actually, the generalized escort (3) appeared in Chernoff analysis of the efficiency of hypothesis tests [9], and enables to define the best achievable exponent in the bayesian probability of error [10, Chapter 11]. As q varies, the generalized escort distribution defines a curve that connects $f(x)$ to $g(x)$ and further. In the general framework of information geometry [11], the generalized escort distribution (3) coincides with the geodesic joining f and g in the case of an exponential connection. Such interpretation also appeared in a work by Campbell [12].

Throughout the paper, we will focus on the generalized escort distribution and the path it defines, that we will call the escort-path.

Distances between probability distributions will be measured by means of information divergences. We will use the Kullback-Leibler directed information divergence which is defined as follows.

Definition 2. Let f and g be two univariate densities with respect to a common measure μ, with f absolutely continuous with respect to g. The Kullback-Leibler directed information divergence is given by

$$D(f||g) = \int f(x) \log \frac{f(x)}{g(x)} \mathrm{d}\mu(x). \tag{4}$$

It is understood, as usual, that $0 \log 0 = 0 \log 0/a = 0 \log 0/0 = 0$. Note that if we take $g(x) = 1$ in the expression above, then we obtain minus the Shannon entropy $H[f]$. Let us also recall that the minimization of the Kullback-Leibler divergence is a well established inference method, analog to Jaynes' maximum entropy approach and which is supported in particular by large deviation results [13]. We will also make use of the Rényi information divergence introduced in [14].

Definition 3. Let f and g be two probability densities with respect to a measure μ. If f is absolutely continuous with respect to g, then, for $q \geq 0$ such that $M_q[f, g] = \int f(x)^q g(x)^{1-q} d\mu(x) < \infty$, the Rényi divergence is defined by

$$D_q(f||g) = \frac{1}{q-1} \log \int f(x)^q g(x)^{1-q} \mathrm{d}\mu(x). \tag{5}$$

Let us recall that the divergence is always non negative $D_q(f||g) \geq 0$ with the equality sign iff $f = g$. By L'Hôpital's rule, the Kullback divergence is recovered in the limit $q \to 1$. Taking $g(x) = 1$ in the expression of the Rényi divergence yields the negative of the Rényi entropy, noted $H_q[f]$.

We will study Fisher information along the escort-path. Indeed, it is well known that the Fisher information metric is a Riemannian metric that can be defined on a smooth statistical manifold [15, 16]. Furthermore, the Fisher information serves as

a measure of the information about a parameter in a distribution. It has intricate relationships with maximum likelihood and has many implications in estimation theory, as exemplified by the Cramér-Rao bound which provides a fundamental lower bound on the variance of an estimator [17]. It is also used as a method of inference and understanding in statistical physics and biology, as promoted by Frieden [18, 19].

Definition 4. Let $f(x; \theta)$ denote a probability density with respect to a measure μ, where θ is a real parameter, and suppose that $f(x; \theta)$ is differentiable with respect to θ. Then, the Fisher information in the density f about the parameter θ is defined as

$$I[f, \theta] = \int \left(\frac{\partial \ln f(x; \theta)}{\partial \theta} \right)^2 f(x; \theta) \mathrm{d}\mu(x). \tag{6}$$

The remaining of the paper is structured as follows. In section 2 we show that the generalized escort presented above arises naturally in a simple probabilistic description of a transition between two states. Interestingly, the Rényi information divergence, and in a particular case the Rényi entropy, emerges as a characterization of the transition.

In section 3, we study the Fisher information, with respect to q, along the escort-path. We show in particular that the integral of the Fisher information along the path, the thermodynamic divergence, is proportional to Jeffreys' divergence.

In section 4, we consider the problem of inferring the distribution $f(x)$ in (2) or (3) on the escort-path when the only available information is given as a mean value. This mean value is the statistical expectation taken with respect to an escort distribution: this is the escort mean value used in nonextensive statistics. Different possible approaches, such as minimizing the directed divergence, or Jeffreys divergence or the thermodynamic divergence, reduce to the minimization of the Rényi information divergence. In this case, the probability distribution that emerges is a generalized Gaussian distribution, which is particularly important in applications.

2. The escort-path

It has been observed that Tsallis' extended thermodynamics seems particularly appropriate in the case of deviations from the classical Boltzmann-Gibbs equilibrium. This suggests that the original MaxEnt formulation "find the closest distribution to a reference under a mean constraint" may be amended by introducing a new constraint that displaces the equilibrium. The partial or displaced equilibrium can be imagined as an equilibrium characterized by two distributions, say $p_0(x)$ and $p_1(x)$. Instead of selecting the nearest distribution to a reference under a mean constraint, we may look for a distribution $p_q(x)$ simultaneously close, in some sense, to two distinct references: such a distribution will be localized somewhere 'between' $p_0(x)$ and $p_1(x)$.

2.1. Displaced equilibrium

We consider two equilibrium states with respective probability densities $p_0(x)$ and $p_1(x)$ with respect to a common measure μ, at some point x in the phase space, and

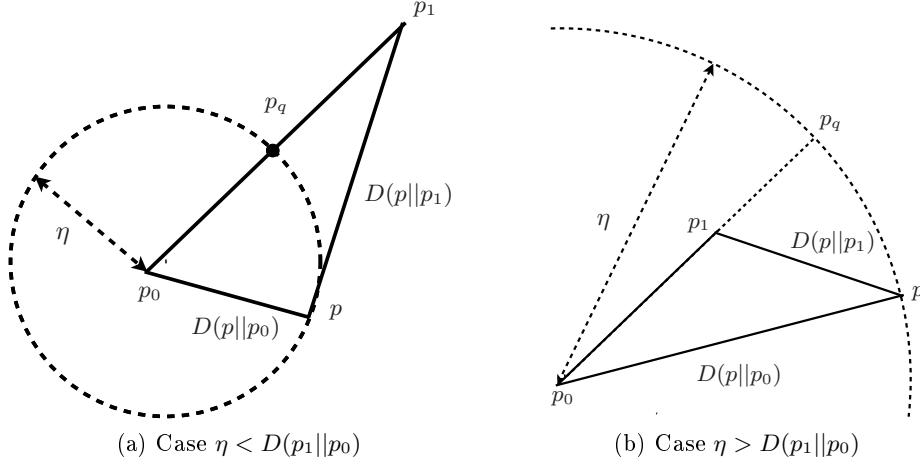(a) Case $\eta < D(p_1||p_0)$ (b) Case $\eta > D(p_1||p_0)$

Figure 1: Constrained equilibrium between states p_0 and p_1: the equilibrium distribution is sought in the set of all distributions such that $D(p||p_0) = \eta$, and with minimum Kullback distance to p_1. The equilibrium distribution p_q , the generalized escort distribution, is "aligned" with p_0 and p_1 and intersects the set $D(p||p_0) = \eta$.

we look at intermediate states defined by the following scenario. The system with initial state p_0, subject to a generalized force, is moved at a distance $\eta = D(p||p_0)$ from p_0, where $D(p||p_0)$ is the Kullback-Leibler divergence (or relative entropy) from p to p_0. Then, the system is attracted toward the final state p_1. Therefore, the new intermediate equilibrium state, say p_q, is chosen as the one which minimizes its divergence to the attractor p_1 while being hold on at the distance η from p_0. As illustrated in Figure 1, the intermediate probability density is located on the "straight line" $p_0 - p_1$ and intersects the circle with radius η centered at p_0. More precisely, the problem can be written as follows:

$$\begin{cases} \min_p \ D(p||p_1) \\ \text{s.t.} \ D(p||p_0) = \eta \\ \text{and} \int p(x)\mathrm{d}\mu(x) = 1 \end{cases} \qquad (7)$$

where "s.t." stands for "subject to", and where the Kullback-Leibler divergence $D(f||g)$ is defined by (4). The solution is given by the following Theorem.

Theorem 5. *Let p_1 a probability density function with respect to μ, and p_0 a non negative function. Assume that p_1 is absolutely continuous with respect to p_0. Let p_q denote the generalized escort distribution with index $q \geq 0$*

$$p_q(x) = \frac{p_1(x)^q p_0(x)^{1-q}}{\int p_1(x)^q p_0(x)^{1-q}\mathrm{d}\mu(x)}, \qquad (8)$$

with $M_q(p_1, p_0) = \int p_1(x)^q p_0(x)^{1-q}\mathrm{d}\mu(x) < \infty$. If $E_q\left[\log \frac{p_1}{p_0}\right]$ is finite, where $E_q[.]$ denote the statistical expectation with respect to p_q, and if q is chosen such that

5

$D(p_q||p_0) = \eta$, then the generalized escort distribution (8) is the unique solution of problem (7).

Proof. Let us evaluate the divergence $D(p||p_q)$. For all densities p satisfying $D(p||p_0) = \eta$, we have

$$D(p||p_q) = \int p(x) \log \frac{p(x)}{p_q(x)} \mathrm{d}\mu(x) = \int p(x) \log \frac{p(x)^q p(x)^{1-q}}{p_1(x)^q p_0(x)^{1-q}} \mathrm{d}\mu(x) + \log M_q(p_1, p_0) \quad (9)$$

$$= q \int p(x) \log \frac{p(x)}{p_1(x)} \mathrm{d}\mu(x) + (1-q) \int p(x) \log \frac{p(x)}{p_0(x)} \mathrm{d}\mu(x) + \log M_q(p_1, p_0) \quad (10)$$

$$= q\, D(p||p_1) + (1-q)\eta + \log M_q(p_1, p_0) \quad (11)$$

Observe that $D(p_q||p_0) = qE_q\left[\log \frac{p_1}{p_0}\right] - \log M_q$ and that $D(p_q||p_1) = (1-q)E_q\left[\log \frac{p_1}{p_0}\right] - \log M_q$ so that both divergences exist. Therefore, taking $p = p_q$, the last equality gives

$$D(p_q||p_q) = q\, D(p_q||p_1) + (1-q)\eta + \log M_q(p_1, p_0). \quad (12)$$

Finally, subtracting (11) and (12) yields

$$D(p||p_q) - D(p_q||p_q) = q\, (D(p||p_1) - D(p_q||p_1)).$$

Since $q \geq 0$ and since $D(p||p_q) \geq 0$ with equality iff $p = p_q$, we obtain that $D(p||p_1) \geq D(p_q||p_1)$ which proves the Theorem. $\qquad\square$

It is interesting to note that (8) is nothing else but a generalized version of the *escort* or *zooming* distribution of nonextensive thermostatistics, and that the corresponding statistical expectations are the so-called escort-means or generalized averages. Obviously, one recovers a standard escort distribution like (1) when $p_0(x)$ is uniform with respect to μ. This is immediate if μ has a compact support. However, if one wants to use a uniform measure on the whole real axis, with μ the Lebesgue measure, then such a measure is no more a probability density since it integrates to infinity. In such case, it still possible to modify the formulation to include this case as well. Indeed, with $p_0(x) = 1$, the expression of the Kullback-Leibler divergence $D(p||p_0)$ becomes nothing but minus the standard entropy

$$H[p] = - \int p(x) \log p(x) \mathrm{d}\mu(x).$$

Therefore, the problem turns into the research of a distribution with a given entropy, which minimizes the divergence to p_1:

$$\begin{cases} \min_p \ D(p||p_1) \\ \text{s.t. } H[p] = -\eta \\ \text{and } \int p(x)\mathrm{d}\mu(x) = 1. \end{cases} \quad (13)$$

This setting can be illustrated as was done in Figure 1, excepted that the circle now corresponds to the set of distributions with a given level of entropy. Observe that neither the Theorem 5 nor its proof require that p_0 is a probability density. Therefore we can take $p_0(x) = 1$ and obtain the solution of (13) as a simple corollary.

Corollary 6. *Let p_q denote the escort distribution with index q, associated with p_1, defined by*

$$p_q(x) = \frac{p_1(x)^q}{\int p_1(x)^q \mathrm{d}\mu(x)}, \tag{14}$$

provided that $M_q(p_1) = \int p_1(x)^q \mathrm{d}\mu(x) < \infty$. If $E_q\left[\log p_1\right]$ is finite, where $E_q\left[.\right]$ denote the statistical expectation with respect to p_q, and if q is chosen such that $H[p_q] = -\eta$, then the escort distribution (14) is the unique solution of problem (13).

When q varies, the function $\eta(q) = D(p_q||p_0)$ is monotonically increasing, and particular intermediate values satisfy the implicit relationship $D(p_q||p_0) = \eta$. This property will be proved in section 3, corollary 10, as a simple consequence of a result on Fisher information. For $q = 0$ we have $\eta = 0$ and for $q = 1$, we have $\eta = D(p_1||p_0)$. Accordingly, as q varies, p_q traces out a curve, the *escort-path,* that connects p_0 ($q = 0$) and p_1 ($q = 1$). In the case $q > 1$, we have $\eta > D(p_1||p_0)$ as shown in Figure 1b.

Interestingly enough, recent results have shown that the average dissipated work during a transition can be expressed as a relative entropy [20, 21]. Along these lines, with an Hamiltonian even in the momenta, the minimization of $D(p||p_1)$ may be understood as a minimization of the average dissipated work for a transition from p to p_1.

2.2. Rényi and Jeffreys' divergences as by-products

It is interesting to outline that the Rényi divergence and entropy arise as a by-product of our construction. Indeed, the minimum of the Kullback-Leibler divergence can be expressed as follows.

Corollary 7. *The minimum divergence is given by*

$$D(p_q||p_1) = \left(1 - \frac{1}{q}\right)(\eta - D_q(p_1||p_0)) \tag{15}$$

where $D_q(p_1||p_0)$ is the Rényi information divergence with index q, from p_1 to p_0.

Proof. By direct calculation from the expression of the solution $p_q(x)$, or by a direct consequence of relation (12). $\square$

If p_0 is a uniform distribution, then $-D_q(p_1||p_0) = H_q(p_1)$, the Rényi entropy, p_q is the standard escort distribution and (15) becomes

$$D(p_q||p_1) = \left(1 - \frac{1}{q}\right)(\eta + H_q(p_1)).$$

Although it is convenient to think of the Kullback-Leibler divergence $D(f||g)$ (4) as a distance between f and g, it is not symmetric and does not satisfy the triangle inequality. Kullback and Leibler themselves introduced a symmetrised version, which was also considered before by Jeffreys. This Jeffreys' divergence appears here to be a simple affine function of Rényi information divergence $D_q(p_1||p_0)$.

Corollary 8. *The Jeffreys divergence between p_1 and the generalized escort distribution p_q is given by*

$$J(p_1, p_q) = D(p_1 || p_q) + D(p_q || p_1) = \frac{(q-1)^2}{q} \left(D_q(p_1 || p_0) - \eta \right). \qquad (16)$$

Proof. This is a simple consequence of (11), which gives $D(p_1 || p_q) = (1-q)\eta + \log M_q(p_1, p_0)$ if $p = p_1$, and of (12) that gives $D(p_q || p_1) = (1 - \frac{1}{q})\eta - \frac{1}{q} \log M_q(p_1, p_0)$. $\square$

As an interesting consequence, we see that if one wants to minimize the symmetric divergence between p_1 and p_q, subject to additional constraints, then this simply amounts to the minimization of the Rényi information divergence with the same constraints. When p_0 is uniform, this becomes the maximization of the Rényi entropy, or equivalently of the Tsallis entropy. It is thus interesting that our setting induces both an escort distribution and a Rényi divergence (or entropy), and besides with a common index q. Actually, although these two quantities are essential ingredients in nonextensive statistical mechanics, their relationships are discussed, e.g. [22].

3. Fisher information along the escort-path

Suppose now that $p_0(x)$ and $p_1(x)$ depend on a parameter θ. The Fisher information metric is based on the Fisher information matrix on a vector parameter θ attached to a density $p(x; \theta)$. This Fisher information matrix has entries

$$[I(\theta)]_{i,j} = \int p(x; \theta) \left(\frac{\partial}{\partial \theta_i} \log p(x; \theta) \right) \left(\frac{\partial}{\partial \theta_j} \log p(x; \theta) \right) d\mu(x).$$

The derivative of the logarithm of the density with respect to the parameter is called the score function. The mean of the score function is zero, so that the Fisher information matrix is the covariance of the score function.

The length of a curve parametrized by t, from 0 to T, is given by

$$\mathcal{L} = \sum_i \sum_j \int_0^T \sqrt{\frac{d\theta_i}{dt} [I(\theta)]_{i,j} \frac{d\theta_j}{dt}} \, dt.$$

In the context of thermodynamics, this quantity is called the thermodynamic length [23, 24, 25]. A related quantity is the thermodynamic divergence, or energy of the curve, given by

$$\mathcal{J} = \sum_i \sum_j \int_0^T \frac{d\theta_i}{dt} [I(\theta)]_{i,j} \frac{d\theta_j}{dt} \, dt.$$

By Jensen's inequality, we have immediately that $\mathcal{J} \geq \mathcal{L}^2$. An interesting point, that outlines the importance of these quantities, is the fact that the thermodynamic divergence asymptotically bounds the dissipation induced by a finite time transformation of a thermodynamic system [26, 24]. Hence, it is interesting here to study

some characteristics of the Fisher information along the escort-path. The general study of the Fisher information on the escort-path with respect to a general parameter θ is interesting in its own right. However, in order to save space, we will focus here on a special case. Let us still simply mention that when p_0 is uniform, the related Fisher information is the *escort-Fisher information* which has been considered in [27, 28, 29].

As we have seen, the generalized escort distribution describes a geometric path, the escort-path, connecting distributions p_0 and p_1 for the values $q = 0$ and $q = 1$. Clearly, the densities on the escort-path are characterized by the index q. Hence it is quite natural to evaluate the distance between two densities on the path, as well as the Fisher information with respect to q. Let us begin by a general expression of the Fisher information on the path. Then, we will be able to link this Fisher information to information divergences on the path.

Theorem 9. *Let p_q be the generalized escort distribution as in (8). Then, the Fisher information with respect to q of the generalized escort distribution is given by*

$$I(q) = \int \frac{1}{p_q(x)} \left(\frac{dp_q(x)}{dq} \right)^2 \mathrm{d}\mu(x) = \int \frac{dp_q(x)}{dq} \log \frac{p_1(x)}{p_0(x)} \, \mathrm{d}\mu(x) \qquad (17)$$

provided that $E_r \left[\left(\log \frac{p_1}{p_0} \right)^2 \right]$ is finite for r in a compact neighborhood of q. The Fisher information with respect to q can also be written as the variance of the log-likelihood ratio:

$$I(q) = E_q \left[\left(\log \frac{p_1(x)}{p_0(x)} - E_q \left[\log \frac{p_1(x)}{p_0(x)} \right] \right)^2 \right]. \qquad (18)$$

Proof. The second order moment condition on the log-likelihood ratio implies, by Jensen inequality, that both $E_q \left[\left| \log \frac{p_1}{p_0} \right| \right]$ and $E_q \left[\log \frac{p_1}{p_0} \right]$ are finite. Let us first consider $M_q(p_1, p_0) = \int p_1(x)^q p_0(x)^{1-q} \mathrm{d}\mu(x)$. The integrand is clearly differentiable with respect to q, and this derivative, which is equal to $p_q \log \frac{p_1}{p_0}$ is continuous and is absolutely integrable since $E_q \left[\left| \log \frac{p_1}{p_0} \right| \right]$ is finite. Furthermore, by the second order moment hypothesis, the last expression is also locally integrable with respect to q. This enables to use Leibniz' rule and differentiate under the integral sign, which gives

$$\frac{\mathrm{d} \log M_q}{\mathrm{d}q} = \int p_q(x) \log \frac{p_1(x)}{p_0(x)} \mathrm{d}\mu(x) = E_q \left[\log \frac{p_1(x)}{p_0(x)} \right]. \qquad (19)$$

Then, by direct calculation, we also have

$$\frac{\mathrm{d}p_q(x)}{\mathrm{d}q} = p_q(x) \left(\log \frac{p_1(x)}{p_0(x)} - E_q \left[\log \frac{p_1(x)}{p_0(x)} \right] \right), \qquad (20)$$

which, inserted in the definition of the Fisher information in (17) gives (18).

By (20), we have that

$$\int \left| \frac{\mathrm{d}p_q(x)}{\mathrm{d}q} \right| \mathrm{d}\mu(x) \le E_q \left[\left| \log \frac{p_1}{p_0} \right| \right] + \left| E_q \left[\log \frac{p_1}{p_0} \right] \right| < \infty. \tag{21}$$

Moreover, by the second order moment hypothesis, (21) is also locally integrable with respect to q. Since $\int p_q(x)\mathrm{d}\mu(x) = 1$, then by Leibniz' rule we get that $\frac{\mathrm{d}}{\mathrm{d}q} \int p_q(x)\mathrm{d}\mu(x) = \int \frac{\mathrm{d}p_q(x)}{\mathrm{d}q}\mathrm{d}\mu(x) = 0$. Finally, the right hand side of (17) is obtained by using (20) and the fact that

$$\int \frac{\mathrm{d}p_q(x)}{\mathrm{d}q} \frac{\mathrm{d}\log M_q}{\mathrm{d}q}\mathrm{d}\mu(x) = \frac{\mathrm{d}\log M_q}{\mathrm{d}q} \int \frac{\mathrm{d}p(x)}{\mathrm{d}q}\mathrm{d}\mu(x) = 0.$$

$\square$

As a simple consequence, we can now check that $\eta = D(p_q||p_0)$ is indeed a monotone increasing function of q, as announced in section 2.

Corollary 10. *Let p_q be a generalized escort distribution, with $q > 0$, and assume that $E_r \left[\left(\log \frac{p_1}{p_0} \right)^2 \right] < \infty$ for r in a compact neighborhood of q. Then $\eta(q) = D(p_q||p_0)$ is a strictly monotone increasing function of q, with*

$$\frac{\partial}{\partial q}\eta(q) = q\,I(q) > 0 \tag{22}$$

Proof. Note that $\eta(q) = \int p_q(x) \log \frac{p_q(x)}{p_0(x)}\mathrm{d}\mu(x) = q \int p_q(x) \log \frac{p_1(x)}{p_0(x)}\mathrm{d}\mu(x) - \log M_q$. Under the second order moment condition, one can differentiate under the integral sign, take into account (19) and it remains

$$\frac{\partial}{\partial q}\eta(q) = q\frac{\partial}{\partial q} \int p_q(x) \log \frac{p_1(x)}{p_0(x)}\mathrm{d}\mu(x) = q \int \frac{dp_q(x)}{dq} \log \frac{p_1(x)}{p_0(x)}\mathrm{d}\mu(x),$$

where we recognize the Fisher information in (17). Therefore, taking into account the fact that both q and the Fisher information are positive, we get (22). $\square$

Finally, an important result is that the integral of the Fisher information, the "energy" of the curve, is nothing but Jeffreys' divergence. This result is mentioned in [30]. Alternatively, this can also be obtained as a consequence of the general integral representation of the Kullback-Leibler divergence [11, eq. 3.71]. We propose here a direct proof of the result.

Theorem 11. *Let p_r and p_s be two generalized escort distributions. Assume that $E_q \left[\left(\log \frac{p_1}{p_0} \right)^2 \right] < \infty$ for all $q \in [r, s]$. Then, the integral of the Fisher information along the escort-path, from $q = r$ to $q = s$ is proportional to Jeffreys' divergence between p_r and p_s:*

$$(s - r) \int_r^s I(q)\mathrm{d}q = J(p_s, p_r) = D(p_s||p_r) + D(p_r||p_s). \tag{23}$$

With $r = 0$ and $s = 1$, we get the integral along the whole path connecting p_0 and p_1, that is

$$\int_0^1 I(q)\mathrm{d}q = J(p_1, p_0) = D(p_1||p_0) + D(p_0||p_1). \tag{24}$$

Proof. The Fisher information is finite on the escort path; therefore its integral over a compact interval is also finite. Let us integrate the right equality in (17):

$$\int_r^s I(q)\mathrm{d}q = \int_r^s \int \frac{\mathrm{d}p_q(x)}{\mathrm{d}q} \log \frac{p_1(x)}{p_0(x)} \,\mathrm{d}\mu(x)\,\mathrm{d}q. \tag{25}$$

Since $I(q)$ is positive and $\int_r^s I(q)\mathrm{d}q$ finite, it is possible to apply Fubini's theorem to the right hand side of (25), and exchange the order of integrations. Thus, integrating with respect to q yields

$$\int_r^s I(q)\mathrm{d}q = \int (p_s(x) - p_r(x)) \log \frac{p_1(x)}{p_0(x)} \,\mathrm{d}\mu(x). \tag{26}$$

On the other hand, the divergence $D(p_s||p_r)$ writes

$$D(p_s||p_r) = (s - r) \int p_s(x) \log \frac{p_1(x)}{p_0(x)} \,\mathrm{d}\mu(x) - \log M_s + \log M_r,$$

and similarly for $D(p_r||p_s)$. Adding the two divergences and taking into account (26) give the result (23). $\qquad\square$

Finally, let θ_i, $i = 1..M$ denote a set of intensive variables, which are some functions of the index q. Then, we have that $\frac{d\log p}{dq} = \sum_{i=1}^M \frac{\partial \log p}{\partial \theta_i} \frac{d\theta_i}{dq}$ and the Fisher information with respect to q can be expressed as

$$I(q) = \int p(x) \left(\frac{\mathrm{d}\log p(x)}{\mathrm{d}q} \right)^2 \mathrm{d}\mu(x) = \sum_{i=1}^M \sum_{j=1}^M \frac{\mathrm{d}\theta_i}{\mathrm{d}q} \left[I(\theta) \right]_{i,j} \frac{\mathrm{d}\theta_j}{\mathrm{d}q},$$

where $I(\theta)$ is the Fisher information matrix with respect to θ. Therefore, for the escort-path we introduced, we obtain that the thermodynamic divergence is nothing but the Jeffreys divergence:

$$\mathcal{J} = \int_0^1 I(q)\mathrm{d}q = \sum_{i=1}^M \sum_{j=1}^M \int_0^1 \frac{\mathrm{d}\theta_i}{\mathrm{d}q} \left[I(\theta) \right]_{i,j} \frac{\mathrm{d}\theta}{\mathrm{d}q} \,\mathrm{d}q = D(p_1||p_0) + D(p_0||p_1). \tag{27}$$

4. Inference of a distribution subject to q-moments constraints

In the last section of this paper, we investigate some relationships between escort-distributions, Fisher information and generalized Gaussians. Let us return to the model of states transition as presented in section 2 that led us to the generalized escort distribution (8) as the optimum intermediate between p_0 and p_1.

11

Assume that the distribution p_1 is not exactly known but that the available information is given as an expectation under the escort p_q. This expectation is the so-called generalized expectation, or q-average which is largely used in nonextensive statistics, although it is generalized here with the presence of p_0. In our context, it has the clear meaning of an expectation with respect to the intermediate distribution p_q at a given distance of a reference p_0, c.f. Theorem 5, or with a given entropy, c.f. Corollary 6. Let the observable be given as the absolute moment of order α:

$$m_{\alpha,q}[p_1] = E_q\left[|x|^\alpha\right] = \frac{\int |x|^\alpha p_1(x)^q p_0(x)^{1-q} \mathrm{d}\mu(x)}{\int p_1(x)^q p_0(x)^{1-q} \mathrm{d}\mu(x)}. \tag{28}$$

Typically, the observable could be a mean energy, where the statistical mean is taken with respect to the escort distribution. Then, the question that arises is the determination of a general distribution p_1 compatible with this constraint.

One may keep the idea of minimizing the divergence to p_1, as in the original problem (7) which led us to the generalized escort distribution. Since the Kullback divergence is a directed divergence, we shall keep the notion of direction by minimizing $D(p_q||p_1)$ for $q < 1$ and $D(p_1||p_q)$ for $q > 1$. In both cases, the divergence is an affine function of the Rényi divergence $D_q(p_1||p_0)$, c.f. (15). Therefore, these minimizations are finally equivalent to the minimization of the Rényi divergence under the generalized mean constraint.

In the same vein, we may consider the minimization of the symmetric Jeffreys' divergence between p_q and p_1. We have noticed (16) that this divergence is also an affine function of the Rényi divergence $D_q(p_1||p_0)$. Therefore, its minimization is also equivalent to the minimization of the Rényi divergence under the generalized mean constraint.

Finally, a natural idea is to select the distribution p, thus its escort p_q, so as to minimize the thermodynamic divergence $\int_q^1 I(t)\mathrm{d}t$ or $\int_1^q I(t)\mathrm{d}t$ from p_q to p, while satisfying the constraint (28). We have seen that Jeffreys' divergence $J(p_1, p_q)$ is proportional to the thermodynamic divergence, as indicated in (23). As a consequence, the minimization of the thermodynamic divergence between p_q and p_1 is also equivalent to the minimization of the Rényi information divergence $D_q(p_1||p_0)$.

It is known [6] that the maximization of Rényi entropy subject to generalized q-moments constraints, or equivalently of Tsallis entropy under the same constraints, leads to generalized Gaussian distributions. As far as the minimization of the Rényi information divergence is concerned, a direct proof based on a simple inequality can be derived along the lines in [31, Appendix 1] or in [32, Proposition 4]. Therefore, we have the following result.

Proposition 12. *Among all distributions with a given q-moment of order α as in (28), the distribution p with minimum thermodynamic divergence, or equivalently which minimizes Jeffreys' or Rényi divergence to its escort, is a generalized Gaussian*

distribution given by

$$p(x) = \begin{cases} \frac{1}{Z_q(\gamma)} \left(1 - (1-q)\gamma |x|^\alpha\right)_+^{\frac{1}{1-q}} p_0(x) & \text{for } q \neq 1 \\ \frac{1}{Z_1(\gamma)} \exp\left(-\gamma |x|^\alpha\right) p_0(x) & \text{for } q = 1, \end{cases} \tag{29}$$

where we use the notation $(x)_+ = \max\{x, 0\}$, *and where* $Z_q(\gamma)$ *is the normalization factor.*

When p_0 is uniform the distribution becomes the standard Gaussian distribution, for $\alpha = 2$, in the limit case $q = 1$, by l'Hôpital's rule. This gives the rationale for the denomination of "generalized Gaussians". For $q < 1$, the probability density has a compact support, while for $q > 1$, the probability density has heavy tails with a power-law behavior and is analog to a Student distribution. These generalized Gaussians appear in statistical physics, where they are the maximum entropy distributions of the nonextensive thermostatistics [6]. In this context, these distributions have been observed to present a significant agreement with experimental data, and also to be the analytical solution of actual physical problems [33, 34], [35]. In an other field, the generalized Gaussians are the one dimensional instances of explicit extremal functions of Sobolev, log-Sobolev or Gagliardo–Nirenberg inequalities on $\mathbb{R}^n$, with $n \geq 2$ [36].

Finally, let us close this paper with the example of a q-variance constraint, i.e. $m_{2,q}[p] = \sigma_q^2$, with $p_0(x) = 1$ and μ the Lebesgue measure. We have seen that among all distributions with a given differential entropy, the standard escort distribution p_q minimizes the Kullback-Leibler divergence to p, for some value of the index q (Proposition 6). If p is free but its escort is known to have a given variance, then the distribution p which minimizes the thermodynamic divergence or Jeffreys' divergence (Proposition 12), or equivalently that maximizes the Rényi entropy, is the generalized Gaussian (29) with $\alpha = 2$. In this setting, p_q is located at the intersection of the set of distributions with a given variance and of the set of distributions with a given Shannon differential entropy. When q varies, the optimum distributions follow a path indexed by q which is nothing but the path followed by the generalized Gaussians, with compact support for $q < 1$ and infinite support for $q > 1$. In the limit case $q = 1$, we obtain a standard Gaussian distribution, which is its own escort distribution, and that has the maximum entropy among all escort distributions with the same variance. These situations are illustrated in Figure 2.

5. Conclusions

In this paper, we have presented a simple probabilistic model of transition between two states, which leads naturally to a generalized escort distribution. This generalized escort distribution enables to describe a path, the escort-path, that connects the two states. Then, we have connected several information measures, and studied their evolution along the escort-path. In particular, we have obtained that
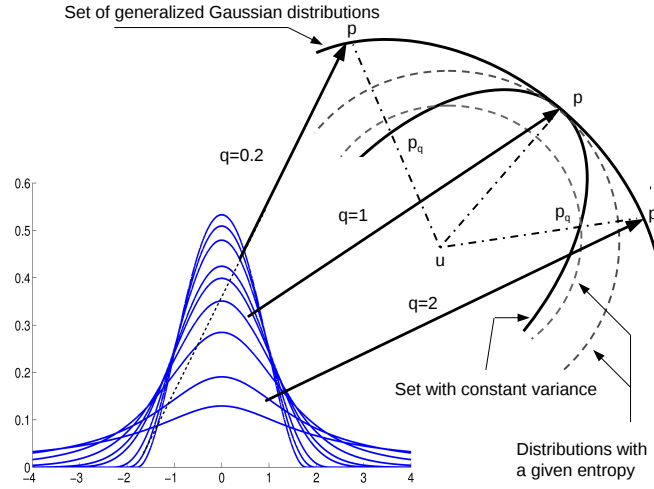
Figure 2: Path of distributions with maximum Rényi entropy and fixed q-variance. For each value of q, the optimum distribution p whose escort p_q has a given variance is a generalized Gaussian. Thus, when q varies, the path followed by p is the manifold of generalized Gaussians with index q.

the Rényi information divergence appears naturally as a characterization of the transition, and that the notion of escort mean values, as used in nonextensive thermostatistics, receives a clear interpretation. We have studied the properties and the evolution of Fisher information along the escort-path. In particular, we have shown that the thermodynamic divergence on the escort-path is a simple function of Jeffreys divergence. We have also considered the problem of inferring a distribution on the escort-path, subject to a moment constraint on its escort. Looking for the distribution as the minimizer of the thermodynamic divergence, we have shown that this procedure is equivalent to the minimization of Rényi divergence subject to a q-moment constraint, which gives a rationale for this approach. Finally, we have recalled that generalized Gaussian distributions arise as solutions of the previous problem.

Beyond the intrinsic interest of our geometric contruction, which enables to connect several quantities of information theory, we have also pointed out possible connections with finite thermostatistics. Furthermore, we have indicated that our findings interrelates several ingredients of the nonextensive statistics. Let us also add that the literature usually points out that the standard entropy (or divergence) is a particular case of generalized Rényi or Tsallis entropies. Our setting suggests a possible additional layer where the generalized quantities are derived from a construction involving the classical information measures. Therefore, we believe that the presented construction, and the series of observations we made can be useful to workers in this field. Future work should consider the extension of this setting in the multivariate case. In future work, we plan to look for possible connections

with finite time thermodynamics. We also intend to study the information theoretic relationships between generalized moments, Fisher information and generalized Gaussians.

[1] A. Chhabra, R. V. Jensen, Direct determination of the $f(\alpha)$ singularity spectrum, Physical Review Letters 62 (12) (1989) 1327.

[2] C. Beck, F. Schloegl, Thermodynamics of Chaotic Systems, Cambridge University Press, 1993.

[3] S. Abe, Geometry of escort distributions, Physical Review E 68 (3) (2003) 031101.

[4] A. Ohara, H. Matsuzoe, S. Amari, A dually flat structure on the space of escort distributions, Journal of Physics: Conference Series 201 (2010) 012012.

[5] J.-F. Bercher, Source coding with escort distributions and Rényi entropy bounds, Physics Letters A 373 (36) (2009) 3235–3238.

[6] C. Tsallis, Introduction to Nonextensive Statistical Mechanics, 1st Edition, Springer, 2009.

[7] S. Abe, G. B. Bagci, Necessity of q -expectation value in nonextensive statistical mechanics, Physical Review E 71 (1) (2005) 016139.

[8] C. Tsallis, A. R. Plastino, R. F. Alvarez-Estrada, Escort mean values and the characterization of power-law-decaying probability densities, Journal of Mathematical Physics 50 (2009) 043303.

[9] H. Chernoff, A measure of asymptotic efficiency for tests of a hypothesis based on the sum of observations, The Annals of Mathematical Statistics 23 (4) (1952) 493–507.

[10] T. M. Cover, J. A. Thomas, Elements of Information Theory, 2nd Edition, Wiley-Interscience, 2006.

[11] S. Amari, H. Nagaoka, Methods of Information Geometry, AMS & Oxford University Press, 2000.

[12] L. L. Campbell, The relation between information theory and the differential geometry approach to statistics, Information Sciences 35 (3) (1985) 199–210.

[13] R. S. Ellis, The theory of large deviations: from Boltzmann's 1877 calculation to equilibrium macrostates in 2D turbulence, Physica D 133 (1-4) (1999) 106–136.

[14] A. Rényi, On measures of entropy and information, in: Proc. 4th Berkeley Sympos. Math. Statist. and Prob., Vol. I, Univ. California Press, Berkeley, Calif., 1961, pp. 547–561.

[15] C. R. Rao, Information and accuracy attainable in the estimation of statistical parameters., Bull. Calcutta Math. Soc 37 (1945) 81—91.

[16] J. Burbea, C. R. Rao, Differential metrics in probability spaces, Probability and Mathematical Statistics 3 (2) (1985) 241–258.

[17] E. W. Barankin, Locally best unbiased estimates, The Annals of Mathematical Statistics 20 (4) (1949) 477–501.

[18] B. R. Frieden, P. M. Binder, Physics from Fisher information: A unification, American Journal of Physics 68 (11) (2000) 1064.

[19] B. R. Frieden, Science from Fisher Information: A Unification, Cambridge University Press, 2004.

[20] R. Kawai, J. M. R. Parrondo, C. V. den Broeck, Dissipation: The Phase-Space perspective, Physical Review Letters 98 (8) (2007) 080602.

[21] J. M. R. Parrondo, C. V. den Broeck, R. Kawai, Entropy production and the arrow of time, New Journal of Physics 11 (7) (2009) 073008.

[22] F. Pennini, A. Plastino, G. Ferri, Semiclassical information from deformed and escort information measures, Physica A 383 (2) (2007) 782–796.

[23] F. Weinhold, Metric geometry of equilibrium thermodynamics, The Journal of Chemical Physics 63 (6) (1975) 2479–2483.

[24] G. E. Crooks, Measuring thermodynamic length, Physical Review Letters 99 (10) (2007) 100602–4.

[25] D. Shenfeld, H. Xu, M. Eastwood, R. Dror, D. Shaw, Minimizing thermodynamic length to select intermediate states for free-energy calculations and replica-exchange simulations, Physical Review E 80 (4).

[26] J. Nulton, P. Salamon, B. Andresen, Q. Anmin, Quasistatic processes as step equilibrations, The Journal of Chemical Physics 83 (1) (1985) 334–338.

[27] P. Hammad, Mesure d'ordre α de l'information au sens de Fisher, Revue de Statistique Appliquée 26 (1) (1978) 73–84.

[28] F. Pennini, A. R. Plastino, A. Plastino, Rènyi entropies and Fisher informations as measures of nonextensivity in a Tsallis setting, Physica A: Statistical and Theoretical Physics 258 (3-4) (1998) 446–457.

[29] F. Pennini, A. Plastino, Escort husimi distributions, Fisher information and nonextensivity, Physics Letters A 326 (1-2) (2004) 20–26.

[30] A. G. Dabak, D. H. Johnson, Relations between Kullback-Leibler distance and Fisher information (2002).
URL http://www.ece.rice.edu/ dhj/distance.pdf

[31] J.-F. Bercher, On some entropy functionals derived from Rényi information divergence, Information Sciences 178 (12) (2008) 2489–2506.

[32] J.-F. Bercher, Escort entropies and divergences and related canonical distribution, Physics Letters A 375 (33) (2011) 2969–2973.

[33] E. Lutz, Anomalous diffusion and Tsallis statistics in an optical lattice, Physical Review A 67 (5) (2003) 051402.

[34] V. Schwämmle, F. D. Nobre, C. Tsallis, q-Gaussians in the porous-medium equation: stability and time evolution, The European Physical Journal B-Condensed Matter and Complex Systems 66 (4) (2008) 537–546.

[35] A. Ohara, T. Wada, Information geometry of q-Gaussian densities and behaviors of solutions to related diffusion equations, Journal of Physics A: Mathematical and Theoretical 43 (3) (2010) 035002.

[36] M. D. Pino, J. Dolbeault, Best constants for Gagliardo-Nirenberg inequalities and applications to nonlinear diffusions, Journal de Mathématiques Pures et Appliquées 81 (9) (2002) 847–875.