

Some results on a χ -divergence, an extended Fisher information and generalized Cramér-Rao inequalities ^{*}

Jean-François Bercher

Laboratoire d'Informatique Gaspard Monge, UMR 8049
Université Paris-Est, ESIEE, Cité Descartes 93162 Noisy-le-Grand Cedex
jf.bercher@esiee.fr

Abstract. We propose a modified χ^β -divergence, give some of its properties, and show that this leads to the definition of a generalized Fisher information. We give generalized Cramér-Rao inequalities, involving this Fisher information, an extension of the Fisher information matrix, and arbitrary norms and power of the estimation error. In the case of a location parameter, we obtain new characterizations of the generalized q -Gaussians, for instance as the distribution with a given moment that minimizes the generalized Fisher information. Finally we indicate how the generalized Fisher information can lead to new uncertainty relations.

1 Introduction

In this communication, we begin with a variation on a χ^β -divergence, which introduces an averaging with respect to an arbitrary distribution. We give some properties of this divergence, including monotonicity and a data processing inequality. Then, in the context of parameter estimation, we show that this induces both an extension of the standard Fisher information, that reduces to the standard Fisher information in a particular case, and leads to an extension of the Cramér-Rao inequality for the estimation of a parameter. We show how these results can be expressed in the multidimensional case, with general norms. We also show how the classical notion of Fisher information matrix can be extended in this context, but unfortunately in a non-explicit form.

In the case of a translation parameter and using the concept of escort distributions, the general Cramér-Rao inequality leads to an inequality for distributions which is saturated by generalized q -Gaussian distributions. These generalized q -Gaussians are important in several areas of physics and mathematics. They are known to maximize the q -entropies subject to a moment constraint. The Cramér-Rao inequality shows that the generalized q -Gaussians also minimize the generalized Fisher information among distributions with a fixed moment. In information theory, the de Bruijn identity links the Fisher information and the

^{*} To be presented during the conference “Geometric Sciences of Information”, Paris, August 28-30, 2013 (and a version published in the proceedings)

derivative of the entropy. We show that this identity can be extended to generalized versions of entropy and Fisher information. More precisely, the generalized Fisher information naturally pops up in the expression of the derivative of the entropy. Finally, we give an extended version of the Weyl-Heisenberg uncertainty relation as a consequence of the generalized multidimensional Cramér-Rao inequality. Due to the lack of space, we will omit or only sketch the proofs, but will reference to the literature whenever possible.

2 Context

Let $f_1(x; \theta)$, $f_2(x; \theta)$ and $g(x; \theta)$ be three probability distributions, with $x \in X \subseteq \mathbb{R}^k$ and θ a parameter of these densities, $\theta \in \mathbb{R}^n$. We will deal with a measure of a divergence between two probability distributions, say $f_1(x)$ and $f_2(x)$, and we will also be interested in the estimation of a vector $h(\theta) \in \mathbb{R}^m$, with $T(x)$ the corresponding estimator.

If $\|\cdot\|$ is an arbitrary norm on \mathbb{R}^m , its dual norm $\|\cdot\|_*$ is defined by

$$\|Y\|_* = \sup_{\|X\| \leq 1} X.Y,$$

where $X.Y$ is the standard inner product. In particular, if $\|\cdot\|$ is an L_p norm, then $\|\cdot\|_*$ is an L_q norm, with $p^{-1} + q^{-1} = 1$. A very basic idea in this work is that it may be useful to vary the probability density function with respect to which are computed the expectations. Typically, in the context of estimation, if the error is $T(x) - h(\theta)$, then the bias can be evaluated as $B_f(h(\theta)) = \int_X (T(x) - h(\theta)) f(x; \theta) dx = E_f [T(x) - h(\theta)]$ while a general moment of an arbitrary norm of the error can be computed with respect to another probability distribution, say $g(x; \theta)$, as in $E_g [\|T(x) - h(\theta)\|^\beta] = \int_X \|T(x) - h(\theta)\|^\beta g(x; \theta) dx$.

The two distributions $f(x; \theta)$ and $g(x; \theta)$ can be chosen very arbitrary. However, one can also build $g(x; \theta)$ as a transformation of $f(x; \theta)$ that highlights, or on the contrary scores out, some characteristics of $f(x; \theta)$. For instance, $g(x; \theta)$ can be a weighted version of $f(x; \theta)$, i.e. $g(x; \theta) = h(x; \theta)f(x; \theta)$, or a quantized version as $g(x; \theta) = [f(x; \theta)]$, where $[.]$ denotes the integer part. Another important special case is when $g(x; \theta)$ is defined as the escort distribution of order q of $f(x; \theta)$, where q plays the role of a tuning parameter: $f(x; \theta)$ and $g(x; \theta)$ are a pair of escort distributions, which are defined as follows:

$$f(x; \theta) = \frac{g(x; \theta)^q}{\int g(x; \theta)^q dx} \quad \text{and} \quad g(x; \theta) = \frac{f(x; \theta)^{\bar{q}}}{\int f(x; \theta)^{\bar{q}} dx}, \quad (1)$$

where q is a positive parameter, $\bar{q} = 1/q$, and provided of course that involved integrals are finite. These escort distributions are an essential ingredient in the nonextensive thermostatics context. Actually, the escort distributions have been introduced as an operational tool in the context of multifractals, c.f. [8], [4], with interesting connections with the standard thermodynamics. Discussion of their geometric properties can be found in [1,14]. Escort distributions also prove useful in source coding, as noticed by [5].

Finally, we will see that in our results, the generalized q -Gaussians play an important role. These generalized q -Gaussians appear in statistical physics, where they are the maximum entropy distributions of the nonextensive thermostatics [16]. The generalized q -Gaussian distributions, which reduce to the standard Gaussian in a particular case, define a versatile family that can describe problems with compact support as well as problems with heavy tailed distributions. They are also analytical solutions of actual physical problems, see e.g. [13], [15] and are sometimes known as Barenblatt-Pattle functions, following their identification by Barenblatt and Pattle. We shall also mention that the Generalized q -Gaussian distributions appear in other fields, namely as the solution of non-linear diffusion equations, or as the distributions that saturate some sharp inequalities in functional analysis [10], [9].

3 The modified χ^β -divergence

The results presented in this section build on a beautiful, but unfortunately overlooked, work by Vajda [17]. In this work, Vajda presented and characterized an extension of the Fisher information popping up as a limit case of a χ^α divergence (for consistency with our notations in previous papers, we will use here the superscript β instead of α as in Vajda's paper). Here, we simply make a step beyond on this route.

The χ^β -divergence between two probability distributions f_1 and f_2 is defined by

$$\chi^\beta(f_1, f_2) = E_{f_2} \left[\left| 1 - \frac{f_1}{f_2} \right|^\beta \right], \quad (2)$$

with $\beta > 1$. In the case $\beta = 2$, and a parametric density $f_\theta(x) = f(x; \theta)$, it is known that the Fisher information of f_θ is nothing but

$$I_{2,1}[f_\theta; \theta] = \lim_{|t| \rightarrow 0} \chi^2(f_{\theta+t}, f_\theta) / t^2.$$

In [17], Vajda extended this to any $\beta > 1$, and defined a generalized Fisher information as $I_\beta[f_\theta; \theta] = \lim_{|t| \rightarrow 0} \chi^\beta(f_{\theta+t}, f_\theta) / |t|^\beta = E_{f_\theta} \left[\left| \frac{\dot{f}_\theta}{f_\theta} \right|^\beta \right]$, assuming that f_θ is differentiable wrt to θ . Let us consider again a χ^β -divergence as in (2), but modified in order to involve a third distribution $g(x; \theta)$:

$$\chi_g^\beta(f_1, f_2) = E_g \left[\left| \frac{f_2 - f_1}{g} \right|^\beta \right].$$

Obviously, this formula includes the standard divergence. For $\theta \in \mathbb{R}^n$, let us denote by $\partial_i f_\theta$ the partial derivative with respect to the i^{th} component, and let t_i be a vector that increments this component. Then, we have $\lim_{|t_i| \rightarrow 0} \chi^\beta(f_{\theta+t_i}, f_\theta) / |t_i|^\beta = E_{g_\theta} \left[\left| \frac{\partial_i f_\theta}{g_\theta} \right|^\beta \right]$. Doing this for all components and summing the resulting vector,

we finally arrive at the following definition

$$I_\beta[f_\theta|g_\theta;\theta] = \sum_i E_{g_\theta} \left[\left| \frac{\partial_i f_\theta}{g_\theta} \right|^\beta \right] = E_{g_\theta} \left[\left\| \frac{\nabla f_\theta}{g_\theta} \right\|_\beta^\beta \right], \quad (3)$$

where ∇f_θ denotes the gradient of f_θ and $\|\cdot\|_\beta$ is the β -norm. A version involving a general norm instead of the β -norm is given in [7]. Vajda's generalized Fisher information [17] corresponds to the scalar case and $g_\theta = f_\theta$. We will see that this generalized Fisher information, which includes previous definitions as particular cases, is involved in a generalized Cramér-Rao inequality for parameter estimation.

The modified χ^β -divergence enjoys some important properties:

Property 1. The modified χ^β -divergence has information monotonicity. This means that coarse-graining the data leads to a loss of information. If \tilde{f}_1, \tilde{f}_2 and \tilde{g} denote the probability densities after coarse-graining, then we have $\chi_g^\beta(f_1, f_2) \geq \chi_{\tilde{g}}^\beta(\tilde{f}_1, \tilde{f}_2)$. A proof of this result can be obtained following the lines in [2]. A consequence of this result is a data processing inequality: if $Y = \phi(X)$, and if f_1^ϕ, f_2^ϕ and g^ϕ denotes the densities after this transformation, then

$$\chi_g^\beta(f_1, f_2) \geq \chi_{g^\phi}^\beta(f_1^\phi, f_2^\phi),$$

with equality if the transformation $Y = \phi(X)$ is invertible. It must be mentioned here that this also yields an important data processing inequality for the generalized Fisher information: $I_\beta[f_\theta|g_\theta;\theta] \geq I_\beta[f_\theta^\phi|g_\theta^\phi;\theta]$.

Property 2. Matrix Fisher data processing inequality. Consider the quadratic case, i.e. $\beta = 2$, $\theta \in \mathbb{R}^n$ and t an increment on θ . Assuming that the partial derivatives of f wrt the components of θ exist and are absolutely integrable, a Taylor expansion about θ gives $f_{\theta+t} = f_\theta + \sum_i t_i \partial_i f_\theta + \frac{1}{2} \sum_i \sum_j t_i t_j \partial_i \partial_j f_\theta + \dots$. Hence, to within the second order terms,

$$\chi_g^2(f_\theta, f_{\theta+t}) = E_g \left[\left(\sum_i t_i \frac{\partial_i f_\theta}{g} \right)^2 \right] = E_g \left[\sum_i \sum_j t_i t_j \frac{\partial_i f_\theta \partial_j f_\theta}{g^2} \right] = t^T I_{2,g}[\theta] t,$$

where $I_{2,g}[\theta] = E_g [\psi_g \psi_g^T] = E_g \left[\frac{\nabla f_\theta}{g_\theta} \frac{\nabla f_\theta^T}{g_\theta} \right]$ is a Fisher information matrix computed wrt to g_θ and ψ_g is a generalized score function. By information monotonicity, $\chi_g^2(f_\theta, f_{\theta+t}) \geq \chi_{g^\phi}^2(f_\theta^\phi, f_{\theta+t}^\phi)$, and therefore we get that

$$I_{2,g}[\theta] \geq I_{2,g^\phi}[\theta]$$

(the difference of the two matrices is positive semi definite) which is a data processing inequality for Fisher information matrices.

Property 3. If $T(X)$ is a statistic, then with $\alpha^{-1} + \beta^{-1} = 1$, $\alpha \geq 1$, we have

$$|E_{f_2}[T] - E_{f_1}[T]| \leq E_g[|T|^\alpha]^{\frac{1}{\alpha}} \chi_g^\beta(f_1, f_2)^{\frac{1}{\beta}}. \quad (4)$$

It suffices here to consider $\left| E_g \left[T \left(\frac{f_2 - f_1}{g} \right) \right] \right| = |E_{f_2}[T] - E_{f_1}[T]|$ and then apply the Hölder inequality to the left hand side.

4 The generalized Cramér-Rao inequality

The Property 3 above can be used to derive a generalized Cramér-Rao inequality involving the generalized Fisher information (3). Let us consider the scalar case. Set $f_2 = f_{\theta+t}$, $f_1 = f_\theta$, and denote $\eta = E_f[T(X) - h(\theta)]$. Then divide both side of (4) by t , substitute $T(X)$ by $T(X) - h(\theta)$, and take the limit $t \rightarrow 0$. Assuming that we can exchange the order of integrations and derivations, and using the definition(3) in the scalar case, we obtain

$$E_g[|T(X) - h(\theta)|^\alpha]^{\frac{1}{\alpha}} I_\beta[f_\theta|g_\theta; \theta]^{\frac{1}{\beta}} \geq \left| \frac{d}{d\theta} \eta \right| \quad (5)$$

which reduces to the standard Cramér-Rao inequality in the case $\alpha = \beta = 2$ and $g = f$. A multidimensional version involving arbitrary norms can be obtained by the following steps: (a) evaluate the divergence of the bias, $\nabla_\theta \cdot B_f(\theta)$, (b) introduce an averaging with respect to g in the resulting integral (c) apply a version of the Hölder inequality for arbitrary norms. The proof can be found in [7] for the direct estimation of the parameter θ . For the estimation of any function $h(\theta)$, we have a generalized Cramér-Rao inequality that enables to lower bound a moment of an arbitrary norm of the estimation error, this moment being computed wrt any distribution g .

Proposition 1. *[Generalized Cramér-Rao inequality, partially in [7]] Under some regularity conditions, then for any estimator $T(X)$ of $h(\theta)$,*

$$E_g[||T(X) - h(\theta)||^\alpha]^{\frac{1}{\alpha}} E_g \left[\left\| H \frac{\nabla_\theta f(X; \theta)}{g(X; \theta)} \right\|_*^\beta \right]^{\frac{1}{\beta}} \geq |m + \nabla_{h(\theta)} \cdot B_f(h(\theta))| \quad (6)$$

where α and β are Hölder conjugate of each other, i.e. $\alpha^{-1} + \beta^{-1} = 1$, $\alpha \geq 1$, with $H_{ij} = \partial \theta_j / \partial h(\theta)_i$, and where the second factor in the left side is actually $I_\beta[f_\theta|g_\theta; h(\theta)]^{\frac{1}{\beta}}$.

Many consequences can be obtained from this general result. For instance, if one chooses $h(\theta) = \theta$ and an unbiased estimator $T(X) = \hat{\theta}(X)$, then the inequality above reduces to

$$E_g \left[\left\| \hat{\theta}(X) - \theta \right\|^\alpha \right]^{\frac{1}{\alpha}} E_g \left[\left\| \frac{\nabla_\theta f(X; \theta)}{g(X; \theta)} \right\|_*^\beta \right]^{\frac{1}{\beta}} \geq n. \quad (7)$$

Taking $f = g$, we obtain an extension of the standard Cramér-Rao inequality, featuring a general norm and an arbitrary power; in the scalar case, we obtain the Barankin-Vajda result [3,17], see also [19].

When $h(\theta)$ is scalar valued, we have the following variation on the theme (proof omitted), which involves a Fisher information matrix, but unfortunately in a non-explicit form: for any matrix A , we have

$$E_g [|T(X) - h(\theta)|^\alpha]^\frac{1}{\alpha} \geq \frac{|\nabla_\theta h(\theta)^t A \nabla_\theta h(\theta)|}{E_g [|\nabla h(\theta)^t A \psi_g|^\beta]^\frac{1}{\beta}}, \quad (8)$$

where $\psi_g = \frac{\nabla_\theta f(x;\theta)}{g(x;\theta)}$ is a score function. We define as Fisher information matrix the matrix A which maximizes the right hand side. In the quadratic case $\alpha = \beta = 2$, and using the inequality $(x^t x)^2 \leq (x^t B x) (x^t B^{-1} x)$ valid for any positive definite matrix B , one can check that the maximum is precisely attained for $A^{-1} = I_{2,g}[\theta] = E_g[\psi_g \psi_g^t]$, that is the Fisher information matrix we obtained above in Property 2, in the quadratic case. The inequality reduces to the inequality $E_g [|T(X) - h(\theta)|^2] \geq \nabla_\theta h(\theta)^t A \nabla_\theta h(\theta)$ which is known in the standard case.

In the quadratic case, it is also possible to obtain an analog of the well known result that the covariance of the estimation error is greater than the inverse of the Fisher information matrix (in the Löwner sense). The proof follows the lines in [11, pp. 296-297] and we get that

$$E_g \left[(T(X) - h(\theta)) (T(X) - h(\theta))^t \right] \geq \dot{\eta}^t I_{2,g}[\theta]^{-1} \dot{\eta} \quad (9)$$

with $\dot{\eta}$ the matrix defined by $\dot{\eta} = \nabla_\theta E_f[\hat{\theta}^t]$, and with equality iff $\dot{\eta} I_{2,g}[\theta]^{-1} \psi_g(X) = \lambda (T(x) - h(\theta))$, with $\lambda > 0$.

Let us now consider the case of a location parameter $\theta \in \mathbb{R}^n$ for a translation family $f(x; \theta) = f(x - \theta)$. In such a case, we have $\nabla_\theta f(x; \theta) = -\nabla_x f(x - \theta)$. Let us also assume, without loss of generality, that $f(x)$ has zero mean. In these conditions, the estimator $T(X) = \hat{\theta}(X) = X$ is unbiased. Finally, taking $\theta = 0$, the relation (7) leads to

Proposition 2. *[Functional Cramér-Rao inequality] For any pair of probability density functions, and under some technical conditions,*

$$\left(\int_X \|x\|^\alpha g(x) dx \right)^\frac{1}{\alpha} \left(\int_X \left\| \frac{\nabla_x f(x)}{g(x)} \right\|_*^\beta g(x) dx \right)^\frac{1}{\beta} \geq n, \quad (10)$$

with equality if $\nabla_x f(x) = -K g(x) \|x\|^{\alpha-1} \nabla_x \|x\|$.

At this point, we can obtain an interesting new characterization of the generalized q -Gaussian distributions, which involves the generalized Fisher information. Indeed, if we take $f(x)$ and $g(x)$ as a pair of escort distributions, with $\bar{q} = 1/q$,

$$f(x) = \frac{g(x)^q}{\int g(x)^q dx} \quad \text{and} \quad g(x) = \frac{f(x)^{\bar{q}}}{\int f(x)^{\bar{q}} dx}, \quad (11)$$

Proposition 3. [*q-Cramér-Rao inequality [7]*] For any probability density g , if $m_\alpha[g] = E[\|x\|^\alpha]$ is the moment of order α of the norm of x , and if

$$I_{\beta,q}[g] = (q/M_q[g])^\beta E_g \left[g(x)^{\beta(q-1)} \|\nabla_x \ln g(x)\|_*^\beta \right] \quad (12)$$

is the Fisher information of order (β, q) , with $M_q[g] = \int g(x)^q dx$, then

$$m_\alpha[g]^\frac{1}{\alpha} I_{\beta,q}[g]^\frac{1}{\beta} \geq n \quad (13)$$

with equality if and only if $g(x)$ is a generalized Gaussian of the form

$$g(x) \propto (1 - \gamma(q-1)\|x\|_+^\alpha)^\frac{1}{q-1} \quad (14)$$

Let us simply note here that $g(x)$ is also called stretched q -Gaussian, become a stretched Gaussian for $q = 1$, and a standard Gaussian when in addition $\alpha = 2$. The inequality (13) shows that the generalized q -Gaussians minimize the generalized Fisher information among all distributions with a given moment. Let us also note and mention that the inequality (13) is similar, but different, to an inequality given by Lutwak et al [12] which is also saturated by the generalized Gaussians (14). Finally, for a location parameter, the matrix inequality (9) reduces to $E_g[XX^t] \geq (I_{2,g})^{-1} = E_g[\psi_g \psi_g^t]^{-1}$, with equality iff f is a generalized q -Gaussian with covariance matrix $(I_{2,g})^{-1}$.

The generalized Fisher information (12) also pops up in an extension of the de Bruijn identity. This identity is usually shown for the solutions of a heat equation. An extension is obtained by considering the solutions of a doubly-nonlinear equation [18]

$$\frac{\partial}{\partial t} f = \Delta_\beta f^m = \operatorname{div}(|\nabla f^m|^{\beta-2} \nabla f^m). \quad (15)$$

Proposition 4. [*Extended de Bruijn identity [6]*] For $q = m + 1 - \frac{\alpha}{\beta}$, $M_q[f] = \int f^q$ and $S_q[f] = \frac{1}{1-q}(M_q[f] - 1)$ the Tsallis entropy, we have

$$\frac{d}{dt} S_q[f] = \left(\frac{m}{q}\right)^{\beta-1} M_q[f]^\beta I_{\beta,q}[f]. \quad (16)$$

Of course, the standard de Bruijn identity is recovered in the particular case $\alpha = \beta = 2$, and $q = m + 1$.

We close this paper by indicating that it is possible to exhibit new uncertainty relations, beginning with the generalized Cramér-Rao inequality (13). These inequalities involve moments computed with respect to escort distributions like (11). We denote by $E_q[\cdot]$ an expectation computed wrt an escort of order q . If ψ is a wave function, x and ξ two Fourier conjugated variables, then

Proposition 5. [Uncertainty relations] For $k = \beta / (\beta(q - 1) + 1)$, $\lambda = n(q - 1) + 1$, and $\gamma \geq 2$, $\theta \geq 2$,

$$\frac{M_{\frac{k}{2}}[|\psi|^2]^{\frac{1}{2}}}{M_{\frac{kq}{2}}[|\psi|^2]} E_{\frac{k}{2}}[\|x\|_2^\gamma]^{\frac{1}{\gamma}} E\left[\|\xi\|_2^\theta\right]^{\frac{1}{\theta}} \geq \frac{n}{2\pi kq}. \quad (17)$$

For $\gamma = \theta = 2$, the lower bound is attained if and only if $|\psi|$ is a generalized Gaussian. For $\gamma = \theta = 2$, $q = 1$, this inequality yields a multidimensional version of the Weyl-Heisenberg uncertainty principle. For $\frac{3}{2} - \frac{1}{\beta} > q$, we also get the inequality

$$\left(E_{\frac{k}{2}}[\|x\|_2^\gamma]\right)^{\frac{1}{\gamma}} \left(E\left[\|\xi\|_2^\theta\right]\right)^{\frac{1}{\theta\lambda}} > \frac{1}{M_{\frac{k}{2}}[|\psi|^2]^{\frac{1}{k\lambda}}} \left(E_{\frac{k}{2}}[\|x\|_2^\gamma]\right)^{\frac{1}{\gamma}} \left(E\left[\|\xi\|_2^\theta\right]\right)^{\frac{1}{\theta\lambda}} \geq K.$$

References

1. Abe, S.: Geometry of escort distributions. *Physical Review E* 68(3), 031101 (2003)
2. Amari, S.I.: α -divergence is unique, belonging to both f -divergence and Bregman divergence classes. *IEEE Transactions on Information Theory* 55(11), 4925–4931 (2009)
3. Barankin, E.W.: Locally best unbiased estimates. *The Annals of Mathematical Statistics* 20(4), 477–501 (Dec 1949)
4. Beck, C., Schloegl, F.: *Thermodynamics of Chaotic Systems*. Cambridge University Press (1993)
5. Bercher, J.F.: Source coding with escort distributions and Rényi entropy bounds. *Physics Letters A* 373(36), 3235–3238 (Aug 2009)
6. Bercher, J.F.: Some properties of generalized Fisher information in the context of nonextensive thermostatics (2012), <http://hal.archives-ouvertes.fr/hal-00766699>
7. Bercher, J.F.: On multidimensional generalized Cramér-Rao inequalities, uncertainty relations and characterizations of generalized q -Gaussian distributions. *Journal of Physics A: Mathematical and Theoretical* 46(9), 095303 (Feb 2013), <http://hal-upec-upem.archives-ouvertes.fr/hal-00766695>
8. Chhabra, A., Jensen, R.V.: Direct determination of the $f(\alpha)$ singularity spectrum. *Physical Review Letters* 62(12), 1327 (Mar 1989)
9. Cordero-Erausquin, D., Nazaret, B., Villani, C.: A mass-transportation approach to sharp Sobolev and Gagliardo-Nirenberg inequalities. *Advances in Mathematics* 182(2), 307–332 (Mar 2004)
10. Del Pino, M., Dolbeault, J.: Best constants for Gagliardo-Nirenberg inequalities and applications to nonlinear diffusions. *Journal de Mathématiques Pures et Appliquées* 81(9), 847–875 (Sep 2002)
11. Liese, F., Miescke, K.J.: *Statistical Decision Theory: Estimation, Testing, and Selection*. Springer, 2008 edn. (Jun 2008)
12. Lutwak, E., Lv, S., Yang, D., Zhang, G.: Extensions of Fisher information and Stam’s inequality. *IEEE Transactions on Information Theory* 58(3), 1319–1327 (Mar 2012)
13. Lutz, E.: Anomalous diffusion and Tsallis statistics in an optical lattice. *Physical Review A* 67(5), 051402 (2003)

14. Ohara, A., Matsuzoe, H., Amari, S.I.: A dually flat structure on the space of escort distributions. *Journal of Physics: Conference Series* 201, 012012 (2010)
15. Schwämmle, V., Nobre, F.D., Tsallis, C.: q -Gaussians in the porous-medium equation: stability and time evolution. *The European Physical Journal B-Condensed Matter and Complex Systems* 66(4), 537–546 (2008)
16. Tsallis, C.: *Introduction to Nonextensive Statistical Mechanics*. Springer (Apr 2009)
17. Vajda, I.: χ^α -divergence and generalized Fisher information. In: *Transactions of the Sixth Prague Conference on Information Theory, Statistical Decision Functions and Random Processes*. pp. 873–886 (1973)
18. Vázquez, J.L.: *Smoothing and Decay Estimates for Nonlinear Diffusion Equations: Equations of Porous Medium Type*. Oxford University Press, USA (Oct 2006)
19. Weinstein, E., Weiss, A.: A general class of lower bounds in parameter estimation. *IEEE Transactions on Information Theory* 34(2), 338–342 (1988)