

On generalized Cramér-Rao inequalities, and an extension of the Shannon-Fisher-Gauss setting

Jean-François Bercher

► To cite this version:

Jean-François Bercher. On generalized Cramér-Rao inequalities, and an extension of the Shannon-Fisher-Gauss setting. Valérie Girardin, Christos H. Skiadas, James R. Bozeman. New Perspectives on Stochastic Modeling and Data Analysis, ISAST, pp.19-35, 2014, 6188069874. <http://www.amazon.com/Perspectives-Stochastic-Modeling-Data-Analysis/dp/6188069874/ref=sr_1_13?s=books

ie=UTF8

qid=1413297143

sr=1-13>. <hal-01087466>

HAL Id: hal-01087466

<https://hal.archives-ouvertes.fr/hal-01087466>

Submitted on 26 Nov 2014

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

On generalized Cramér-Rao inequalities, and an extension of the Shannon-Fisher-Gauss setting

Jean-François Bercher¹

Laboratoire d'informatique Gaspard Monge, UMR 8049 ESIEE-Paris, Université Paris-Est, France
(E-mail: jf.bercher@esiee.fr)

Abstract We begin with two possible extensions of Stam's inequality and of de Bruijn's identity. In both cases, a generalized q -Gaussian plays the same role as the standard Gaussian in the classical case. These generalized q -Gaussians are important in several areas of physics and mathematics. A generalized Fisher information also pops up, playing the same role as the classical Fisher information, but for the extended identity and inequality.

In the estimation theory context, we give several extensions of the Cramér-Rao inequality in the multivariate case, with matrix versions as well as versions for general norms. We define new forms of Fisher information, that reduce to the classical one in special cases.

In the case of a translation parameter, the general Cramér-Rao inequalities lead to an inequality for distributions, which involves the same generalized Fisher information as in the generalized de Bruijn's identity and Stam's inequality. This Cramér-Rao inequality is saturated by generalized q -Gaussian distributions. This shows that the generalized q -Gaussians also minimize the generalized Fisher information among distributions with a fixed moment. Similarly, the generalized q -Gaussians also minimize the generalized Fisher information among distributions with a given q -entropy.

Keywords: Cramér-Rao inequality, Fisher information, generalized q -entropy, generalized Gaussians, de Bruijn identity.

1 Introduction

Classical information theoretic inequalities, as described for instance in [12,14], interrelate information measures. These inequalities have proved to be useful for communication theoretic problems and engineering applications. They are also connected to uncertainty relations in physics, and to functional inequalities in mathematics. Thus new, or refinements of existing information theoretic inequalities, could be fruitful in several fields. Even before recalling the exact expressions of all the involved information measures, we find useful to first briefly describe their main interrelations.

The Shannon-Fisher-Gauss setting – The Gaussian distribution is well known to play a central role with respect to classical information measures and inequalities. More precisely, in a particular setting, that we will call the Shannon-Fisher-Gauss setting, these three quantities are interrelated in many ways. For instance, the Gaussian distribution maximizes the entropy over all distributions with the same variance; see [14, Lemma 5]. This can be stated as a

moment-entropy inequality [22]:

$$\frac{m_2[f]^{\frac{1}{2}}}{N[f]} \geq \frac{m_2[G]^{\frac{1}{2}}}{N[G]}, \quad (1)$$

where $m_2[f]$ is the second order moment, $N[f]$ is the entropy power, defined by $N[f] := \frac{1}{2\pi e} \exp\left(\frac{2}{n} H_1[f]\right)$, where $H_1[f]$ is the Shannon entropy of f , and G denotes the standard Gaussian. Similarly, the Cramér-Rao inequality, see e.g. [14, Theorem 20], shows that the minimum of the Fisher information among all distributions with a given variance is attained for the Gaussian distribution:

$$I_{2,1}[f] m_2[f] \geq I_{2,1}[G] m_2[G] = n. \quad (2)$$

Stam's inequality [26] shows that the minimum of the Fisher information over all distributions with a given entropy also occurs for the Gaussian distribution:

$$I_{2,1}[f] N[f] \geq I_{2,1}[G] N[G] \quad (3)$$

while de Bruijn's identity states that Fisher information is related to the derivative of the entropy, with respect to the variance of an additive Gaussian perturbation: if $X_t = X + \sqrt{2t}N$, and $N \sim G$ then

$$\frac{d}{dt} H[f_t] = I_{2,1}[f_t]. \quad (4)$$

In these expressions, the involved Fisher information is the Fisher information associated to a location parameter model. We will give the general definition and further results in what follows.

Clearly, it is both important and attracting to ask whether some similar inequalities, involving generalized information measures exist. A natural option is to consider the Rényi entropy, or the Tsallis entropy, which depend on a real parameter q , instead of the Shannon entropy. Then the answer to the question is positive: a series of information theoretic inequalities does exist which leads to a q -entropy- q -Fisher- q -Gaussian setting. This setting involves not only generalized q -entropies, but also a generalized Gaussian distribution and a suitable extension of Fisher information. It is one of the objectives of this paper to give extensions of inequalities (1)-(4), thus defining the q -entropy- q -Fisher- q -Gaussian setting as an extension of the Shannon-Fisher-Gauss triplet. Let us first recall some definitions, notations, and introduce the generalized q -Gaussian distribution.

Regarding q -entropies and generalized q -Gaussian distribution – Let $f(x)$ be a probability distribution defined on $X \subseteq \mathbb{R}^n$. For $q \geq 0$,

$$M_q[f] = \int_X f(x)^q dx \quad (5)$$

is the information generating function. With this notation,

$$S_q[f] = \frac{1}{1-q} (M_q[f] - 1) \quad (6)$$

is the Tsallis entropy, and

$$H_q[f] = \frac{1}{1-q} \log M_q[f] \quad (7)$$

the Rényi entropy. Both entropies reduce to the standard Shannon entropy $H_1[f] = -\int_X f(x) \log f(x) dx$ for $q = 1$. We will also note by $N_q[f]$ the q -entropy power. It is defined as an exponential of the Rényi entropy $H_q[f]$ as

$$N_q[f] = M_q[f]^{\frac{2}{n} \frac{1}{1-q}} = \exp\left(\frac{2}{n} H_q[f]\right) = \left(\int_{\Omega} f(x)^q dx\right)^{\frac{2}{n} \frac{1}{1-q}}, \quad (8)$$

for $q \neq 1$. For $q = 1$, we set $N[f] = N_1[f] = \exp\left(\frac{2}{n} H_1[f]\right)$.

It is well-known that the maximum of the Rényi-Tsallis entropy, among all distributions with a fixed moment $m_\alpha = E_g[\|X\|^\alpha]$, is obtained for a generalized q -Gaussian distribution with parameter γ

$$G_\gamma(x) \propto \begin{cases} (1 - (q-1)\gamma\|x\|^\alpha)_+^{\frac{1}{q-1}} & \text{for } q \neq 1 \\ \exp(-\gamma\|x\|^\alpha) & \text{for } q = 1. \end{cases} \quad (9)$$

which is a generalized (or stretched) Gaussian for $q > 1$, and a standard Gaussian for $q = 1$, $\alpha = 2$. This extends the classical result that the standard Gaussian maximizes Shannon entropy subject to a second order moment constraint. Note that $\gamma^{-\frac{1}{\alpha}}$ is simply a scale parameter which allows in particular to fix the moment of order α . For instance, in the case of a standard Euclidean norm, $\gamma = ((1 + \alpha/n)q - 1) m_\alpha$. For $q > 1$, the generalized q -Gaussian distributions exhibit heavy tails, while they have compact support when $q \leq 1$. The generalized q -Gaussians are used in nonextensive statistical physics, where they present significant agreement with experimental data. They also appear in problems involving non-linear diffusion equations (as we will see in the sequel) and as extremal functions of Sobolev, log-Sobolev or Gagliardo–Nirenberg inequalities. The fact that these distributions maximize the q -entropies can be seen as a consequence of a moment-entropy inequality

$$\frac{m_\alpha[f]^{\frac{1}{2}}}{N_q[f]} \geq \frac{m_\alpha[G]^{\frac{1}{2}}}{N_q[G]} \quad (10)$$

similar to (1), and where the lower bound is attained if f is any generalized q -Gaussian G_γ as in (9). This general inequality is due to Lutwak, Yang and Zhang [22]. In (10), we dropped the index γ of the generalized q -Gaussian G_γ to reflect the fact that the lower bound is scale invariant, i.e. does not depend on γ .

On Fisher information – The Fisher information $I_{2,1}[f]$ involved in the Shannon-Fisher-Gauss setting (1)-(4) is a measure of information attached to the distribution f . This Fisher information is used in inference and understanding in physics [16]. It is also used as a tool for characterizing complex signals or systems, with applications, e.g. [30], [24], [13]. Actually, the Fisher information

has been originally introduced in the wider context of estimation theory. It measures the information about a parameter θ in a distribution, corresponds to the Hessian of the log-likelihood and defines, by the Cramér-Rao inequality, a fundamental lower bound on the variance of any estimator. Let $f(x; \theta)$ be a probability density function defined over a subset X of \mathbb{R}^n , $\theta \in \Theta \subset \mathbb{R}$, $h(\theta)$ a scalar valued function and $\eta(\theta) = E[T(X)]$. Then, under some regularity conditions:

$$E \left[|T(X) - h(\theta)|^2 \right] I_{2,1}[f; \theta] \geq \left| \frac{\partial}{\partial \theta} \eta(\theta) \right|^2, \quad (11)$$

with

$$I_{2,1}[f, \theta] = \int_X \left(\frac{\partial \ln f(x; \theta)}{\partial \theta} \right)^2 f(x; \theta) dx, \quad (12)$$

and with equality iff $\frac{\partial}{\partial \theta} \ln f(x; \theta) = k(\theta) (T(x) - h(\theta))$. The indexes in the notation $I_{2,1}[f, \theta]$ are for possible extensions of the definition. The first index corresponds to the quadratic case. Indeed, the definition, here given as the second order moment of the derivative of the log-likelihood – the score function, can be extended to other moments, leading to a generalized version of the Cramér-Rao inequality. This extension, which does not seem to be well known, can be traced back to Barankin [3, Corollary 5.1] and Vajda [27]. The second index will correspond to an extension associated with q -entropies. In the case of a translation family, i.e. $f(x; \theta) = f(x - \theta)$, that is if the parameter θ is a location parameter, the Fisher information, expressed at $\theta = 0$, becomes a characteristic of information in the distribution

$$I_{2,1}[f, \theta = 0] = \int_X \left(\frac{d \ln f(x)}{dx} \right)^2 f(x) dx.$$

In this case, the Cramér-Rao inequality leads to inequality (2) for the moments of the distribution, where the equality is achieved by the Gaussian distribution.

A part of this paper will be devoted to extensions of the Cramér-Rao inequality for the estimation of a multivariate parameter, in particular with variations on the way of averaging the estimation error, and on ways for handling the multivariate case. These new Cramér-Rao inequalities may prove useful in their own right. A second objective of this paper is to show that the classical Shannon-Fisher-Gauss setting can be extended to generalized q -entropies, to the generalized q -gaussian distribution, and a suitable generalized Fisher information, that has an estimation theoretic foundation. These ideas have grown in a series of papers [4], [6], [5], [7]. Of course, some of these results will be reviewed here. However, many proofs are revisited and the paper also provides several new results, in particular the matrix versions of the multivariate Cramér-Rao inequality and further results on the generalized Fisher information.

The rest of this paper is organized as follows. In section 2, we give two possible extensions of Stam's inequality (3) and of de Bruijn's identity (4) to the q -setting. In both cases, the generalized q -Gaussian (9) plays a key role; namely as a limit case or as the distribution which saturates the inequality. In these two relations, a generalized Fisher information also pops up, playing the same role

as the classical Fisher information, but for the extended identity and inequality. In section 3, we turn to the extension of the Cramér-Rao inequality. We first describe the setting where we consider a moment of any order of the estimation error, computed with respect to an arbitrary distribution. In section 3.1, we obtain a matrix version of the Cramér-Rao inequality, in the multivariate case, where the Fisher information matrix is defined implicitly as a function of a generalized score function. Special cases are examined. In section 3.2 we also give other extensions of the Cramér-Rao inequality for general norms of the estimation error. In the special case of a translation parameter and using the notion of escort distributions, these generalized Cramér-Rao inequalities lead to an extension of the Cramér-Rao inequality (2), saturated by generalized q -Gaussian distributions. Finally, in section 4, we discuss the main contributions and describe several directions for future work.

2 Extensions of de Bruijn's identity and of Stam's inequality

As mentioned above, a fundamental connection between the Shannon entropy, Fisher information, and the Gaussian distribution is given by the de Bruijn identity [26]. We show here that this important connection can be extended to the q -entropies, a suitable generalized Fisher information and the generalized q -Gaussian distributions.

The de Bruijn identity states that if $Y_t = X + \sqrt{2t}Z$ where Z is a standard Gaussian vector and X a random vector of \mathbb{R}^n , independent of Z , then

$$\frac{d}{dt}H[f_{Y_t}] = I_{2,1}[f_{Y_t}] = \phi_{2,1}[f_{Y_t}], \quad (13)$$

where f_{Y_t} denotes the density of $Y_t = X + \sqrt{2t}Z$, and $I_{2,1}[f_{Y_t}]$, $\phi_{2,1}[f_{Y_t}]$ both denote the classical Fisher information (the meaning of which will be made clear in the following). Although the de Bruijn identity holds in a wider context, its classical proof uses the fact that if Z is a standard Gaussian vector, then Y_t satisfies the well-known heat equation $\frac{\partial f}{\partial t} = \Delta f$, where Δ denotes the Laplace operator.

Nonlinear versions of the heat equation are of interest in a large number of physical situations, including fluid mechanics, nonlinear heat transfer or diffusion. Other applications have been reported in mathematical biology, lubrication, boundary layer theory, etc; see the series of applications presented in [28, chapters 2 and 21] and references therein. The porous medium equation and the fast diffusion equation correspond to the differential equation $\frac{\partial f}{\partial t} = \Delta f^m$, with $m > 1$ for the porous medium equation and < 1 for the fast diffusion. These two equations have been exhaustively studied and characterized by J. L. Vazquez, e.g. in [28,29].

These equations are included as particular cases into the *doubly nonlinear equation*, which involves a p -Laplacian operator $\Delta_p f := \operatorname{div}(|\nabla f|^{p-2} \nabla f)$, and the power m of the porous medium or fast diffusion equation. This doubly

nonlinear equation takes the form

$$\frac{\partial}{\partial t} f = \Delta_{\beta} f^m = \operatorname{div} (|\nabla f^m|^{\beta-2} \nabla f^m), \quad (14)$$

where we use $p = \beta$ for convenience and coherence with notation in the present paper. The β -Laplacian typically appears in the minimization of a Dirichlet energy such as $\int |\nabla f|^{\beta} dx$ which leads to the Euler-Lagrange equation. As we can see, the doubly nonlinear equation includes the standard heat equation ($\beta = 2, m = 1$), the β -Laplace equation ($\beta \neq 2, m = 1$), the porous medium equation ($\beta = 2, m > 1$) and the fast diffusion equation ($\beta = 2, m < 1$). It can be shown, see [29, page 192], that for $m(\beta - 1) + (\beta/n) - 1 > 0$, (14) has a unique self-similar solution, called a Barenblatt profile $B()$, whose initial value is the Dirac mass at the origin. This fundamental solution is usually given as a function of m . Here, if we set $q = m + 1 - \alpha/\beta$, the solution can be written as a q -Gaussian distribution:

$$f(x, t) = \frac{1}{t^{\frac{n}{\delta}}} B\left(\frac{x}{t^{\frac{1}{\delta}}}\right), \quad \text{with } B(x) = \begin{cases} (C - k|x|^{\alpha})_+^{\frac{1}{q-1}} & \text{for } q \neq 1 \\ \frac{1}{\sigma} \exp\left(-\frac{|\beta-1|}{\beta\alpha}|x|^{\alpha}\right) & \text{for } q = 1 \end{cases} \quad (15)$$

with $\delta = n(\beta - 1)m + \beta - n > 0$, $k = \frac{m(\beta-1)-1}{\beta} \left(\frac{1}{\delta}\right)^{\frac{1}{\beta-1}}$ and $\alpha = \frac{\beta}{\beta-1}$. The constants C and σ are uniquely determined by mass conservation, e.g. $\int f(x, t) dx = 1$. Of course, we observe that the function $B(x)$ above is analogous to the generalized q -Gaussian (9).

As mentioned above, the doubly nonlinear diffusion equation allows to derive a nice extension of the de Bruijn identity (13), and leads to a possible definition of a generalized Fisher information. This is stated in the next Proposition. The case $\beta = 2$ has been proved in [17].

Proposition 1. *[Extended de Bruijn identity, [7]] Let $f(x, t)$ be a probability distributions defined on a subset X of \mathbb{R}^n and satisfying the doubly nonlinear equation (14). Assume that X is independent of t , that $f(x, t)$ is differentiable with respect to t , continuously differentiable over X , and that $\frac{\partial}{\partial t} f(x, t)^q$ is absolutely integrable and locally integrable with respect to t . Then, for $\beta > 1$, α and β Hölder conjugates of each other, $q = m + 1 - \frac{\alpha}{\beta}$, $M_q[f] = \int f^q$ and $S_q[f] = \frac{1}{1-q} (M_q[f] - 1)$ the Tsallis entropy, we have*

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

$$\text{with } \begin{cases} \phi_{\beta,q}[f] = \int_X f(x)^{\beta(q-1)+1} \left(\frac{|\nabla f(x)|}{f(x)}\right)^{\beta} dx \\ \text{and } I_{\beta,q}[f] = \frac{1}{M_q[f]^{\beta}} \phi_{\beta,q}[f]. \end{cases} \quad (17)$$

In (17), $\phi_{\beta,q}[f]$ and $I_{\beta,q}[f]$ are two possible generalization of Fisher information. Of course, the standard Fisher information is recovered in the particular case $\alpha = \beta = 2$, and $q = m = 1$, and so de Bruijn's identity (13). The proof of this result relies on integration by part (actually using the Green identity)

along the solutions of the nonlinear heat equation (14). See [7] for a proof not repeated here. A variant of the result for $\beta = 2$, which considers a free-energy instead of the entropy above, is well-known in certain commities, see e.g. [15,10]. Even more, by carefully using calculations in [15], one can check that $\frac{d}{dt}\phi_{2,q}[f] \leq 0$ for $q > 1 - \frac{1}{n}$, which means the Tsallis entropy is a monotone increasing concave function along the solutions of (14). In their recent work [25], Savaré and Toscani have shown that in the case $\beta = 2$, $m = q$, the entropy power, up to a certain exponent, is a concave function of t , thus generalizing the well-known concavity of the (Shannon) entropy power to the case of q -entropies. This allows one to obtain as a by-product a generalized version of the Stam inequality, valid for the solutions of (14).

Actually, this extension of Stam's inequality, which links the generalized Fisher information to the q -entropy power (8), holds in a broader context.

Proposition 2. *[Generalized Stam inequality, cf [7]] Let $n \geq 1$, β and α be Hölder conjugates of each other, $\alpha > 1$, and $q > \max\{(n-1)/n, n/(n+\alpha)\}$. Then for any continuously differentiable probability density on \mathbb{R}^n , the following generalized Stam inequality holds*

$$I_{\beta,q}[f]^{\frac{1}{\beta}} N_q[f]^{\frac{1}{2}} \geq I_{\beta,q}[G]^{\frac{1}{\beta}} N_q[G]^{\frac{1}{2}}. \quad (18)$$

with equality if and only if f is a generalized q -Gaussian (9).

The generalized Stam's inequality implies that the generalized q -Gaussians minimize the generalized Fisher information within the set of probability distributions with a fixed q -entropy power; or alternatively minimize the entropy power among the distributions with a given (β, q) -Fisher information. A related, but different, Generalized Stam inequality has been given by Lutwak *et al* [22]. Their proof involves optimal transportation of probability measures. The proof below is quite different.

Proof. We sketch the proof in the case $q < 1$. The inequality follows from a sharp Gagliardo-Nirenberg inequality valid for $n > 1$, due to Cordero *et al* [11]:

$$\|\nabla u\|_{\beta} \|u\|_{a(\beta-1)+1}^{\frac{1}{\theta}-1} \geq K \|u\|_{a\beta}^{\frac{1}{\theta}}$$

with $a > 1$ and $\|\nabla u\|_{\beta} = \left(\int \|\nabla u\|_{*}^{\beta} dx\right)^{\frac{1}{\beta}}$, and where K is a sharp constant attained if and only if u is a generalized Gaussian with exponent $1/(1-a)$, and $\theta = n(a-1)/a(n\beta - (a\beta + 1 - a)(n - \beta))$. The idea is to take $u = g^t$, for g a probability density function, with $a\beta t = 1$, and to set $q = [a(\beta - 1) + 1]t$. With these notations, we get that $\beta t = \beta(q - 1) + 1$, and $a = 1/\beta t > 1$ implies that $q < 1$. Rearranging the exponents, (18) follows. Similarly, the case $q > 1$ follows from the sharp Gagliardo-Nirenberg inequality with $a < 1$. In the case $n = 1$, the inequality follows similarly from a sharp Gagliardo-Nirenberg inequality on the real line [1] [23]. \square

3 Extended Cramér-Rao inequalities

Let $f(x; \theta)$ be a probability distribution, with $x \in X \subseteq \mathbb{R}^n$ and $\theta \in \mathbb{R}^k$. We will first deal here with the estimation of a scalar function $h(\theta)$ of θ , with $T(x)$ the corresponding estimator. We extend here the classical Cramér-Rao inequality in two directions: firstly, we give results for a general moment of the estimation error instead of the second order moment, and secondly we introduce the possibility of computing the moment of this error with respect to a distribution $g(x; \theta)$ instead of $f(x; \theta)$. In estimation theory, the error is $T(X) - h(\theta)$, and the bias can be evaluated as

$$\int_X (T(x) - h(\theta)) f(x; \theta) dx = E_f [T(X) - h(\theta)] = \eta(\theta) - h(\theta),$$

while a general moment of the error can be computed with respect to another probability distribution $g(x; \theta)$, 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 arbitrarily. 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 $g(x; \theta) = [f(x; \theta)]$, where $[.]$ denotes the integer part. Another important case appears when $g(x; \theta)$ is defined as the escort distribution of order q of $f(x; \theta)$:

$$f(x; \theta) = \frac{g(x; \theta)^q}{\int g(x; \theta)^q dx} \quad \text{and} \quad g(x; \theta) = \frac{f(x; \theta)^{\bar{q}}}{\int f(x; \theta)^{\bar{q}} dx}, \quad (19)$$

where q is a positive parameter, $\bar{q} = 1/q$, provided of course that all involved integrals are finite. These escort distributions are an essential ingredient in the nonextensive thermostatics context. When $f(x; \theta)$ and $g(x; \theta)$ are a pair of escort distributions we will recover the generalized Fisher information (17) obtained in the extended de Bruijn identity.

3.1 Multivariate generalized Cramér-Rao inequality – matrix version

Previous results on generalized Fisher information can be found in [6,5] in the case of the direct estimation of the parameter θ . We propose here a new derivation, introducing in particular a notion of generalized Fisher information matrix, in the case of the estimation of a function of the parameters. Let us first state the result.

Proposition 3. *Let $f(x; \theta)$ be a multivariate probability density function defined for $x \in X \subseteq \mathbb{R}^n$, with parameter $\theta \in \Theta \subseteq \mathbb{R}^k$. Let $g(x; \theta)$ denote another probability density function also defined on $(X; \Theta)$. Assume that $f(x; \theta)$ is a jointly measurable function of x and θ , is integrable with respect to x , absolutely*

continuous with respect to θ . Suppose that the derivatives with respect to each component of θ are locally integrable. Let $T(x)$ be an estimator of a scalar valued function $h(\theta) : \Theta \rightarrow \mathbb{R}$ and set $\eta(\theta) = E_f[T(X)]$. Then, for any estimator $T(x)$ of $h(\theta)$, we have

$$E_g [|T(X) - h(\theta)|^\alpha]^\frac{1}{\alpha} \geq \sup_{A>0} \frac{\dot{\eta}(\theta)^T A \dot{\eta}(\theta)}{E_g [|\dot{\eta}(\theta)^T A \psi_g(X; \theta)|^\beta]^\frac{1}{\beta}}. \quad (20)$$

Equality in (20) holds if and only if

$$T(x) - h(\theta) = c(\theta) \text{sign}(\dot{\eta}(\theta)^T A \psi_g(x; \theta)) |\dot{\eta}(\theta)^T A \psi_g(x; \theta)|^{\beta-1} \quad (21)$$

with $c(\theta) > 0$ and where $\alpha^{-1} + \beta^{-1} = 1$, $\alpha > 1$, A is a positive definite matrix and $\psi_g(x; \theta)$ a score function given with respect to $g(x; \theta)$:

$$\psi_g(x; \theta) := \frac{\nabla_\theta f(x; \theta)}{g(x; \theta)}. \quad (22)$$

Proof. Let $\eta(\theta) = E_f[T(X)]$. Let us first observe that

$$E_g[\psi_g(x; \theta)] = \frac{d}{d\theta} \int_X f(x; \theta) dx = 0.$$

Differentiating $\eta(\theta) = E_f[T(X)]$ with respect to each θ_i we get

$$\begin{aligned} \dot{\eta}(\theta) &= \nabla_\theta \eta(\theta) = \nabla_\theta \int_X T(x) f(x; \theta) dx \\ &= \int_X T(x) \frac{\nabla_\theta f(x; \theta)}{g(x; \theta)} g(x; \theta) dx \\ &= \int_X (T(x) - h(\theta)) \psi_g(x; \theta) g(x; \theta) dx. \end{aligned}$$

For any positive definite matrix A , multiplying on the left by $\dot{\eta}(\theta)^T A$ gives

$$\dot{\eta}(\theta)^T A \dot{\eta}(\theta) = \int_X (T(x) - h(\theta)) \dot{\eta}(\theta)^T A \psi_g(x; \theta) g(x; \theta) dx,$$

and by the Hölder inequality

$$E_g [|T(x) - h(\theta)|^\alpha]^\frac{1}{\alpha} E_g [|\dot{\eta}(\theta)^T A \psi_g(x; \theta)|^\beta]^\frac{1}{\beta} \geq \dot{\eta}(\theta)^T A \dot{\eta}(\theta),$$

with equality if and only if $(T(x) - h(\theta)) \dot{\eta}(\theta)^T A \psi_g(x; \theta) > 0$ and $|T(x) - h(\theta)|^\alpha = k(\theta) |\dot{\eta}(\theta)^T A \psi_g(x; \theta)|^\beta$, $k(\theta) > 0$. The latter inequality, in turn, provides us with the lower bound (20) for the moment of order α of the estimation error, computed with respect to g . \square

The inverse of the matrix A which maximizes the right hand side of (20) is the Fisher information matrix of order β . Unfortunately, we do not have a closed-form expression for this matrix in the general case. Nevertheless, two particular cases are of interest.

Corollary 4. [Scalar extended Cramér-Rao inequality] In the scalar case (or the case of a single component of θ), the following inequality holds

$$E_g [|T(X) - h(\theta)|^\alpha]^{\frac{1}{\alpha}} \geq \frac{|\dot{\eta}(\theta)|}{E_g [|\psi_g(X; \theta)|^\beta]^{\frac{1}{\beta}}}, \quad (23)$$

with equality if and only if $T(x) - h(\theta) = c(\theta)\text{sign}(\psi_g(x; \theta)) |\psi_g(x; \theta)|^{\beta-1}$.

In the simple scalar case, we see that $A > 0$ can be simplified in (20) and thus that (23) follows. Note that for $\alpha = 2$, the equality case implies that $E_g[\psi_g] = 0 = E_g[T(X) - h(\theta)]$, which means that $E_g[T(X)] = h(\theta)$, i.e. the estimator is unbiased (with respect to g) – but this also mean that it will be generally biased with respect to f . Actually, the inequality (23) recovers at once the generalized Cramér-Rao inequality we presented in the univariate case [5]. The denominator plays the role of the Fisher information in the classical case, which corresponds to the case $g(x; \theta) = f(x; \theta)$, $\beta = 2$.

A second interesting case is the multivariate case with $\alpha = \beta = 2$. Indeed, in that case, we get an explicit form for the generalized Fisher information matrix and an inequality which looks like the classical one.

Corollary 5. [Multivariate Cramér-Rao inequality with $\alpha = \beta = 2$] For $\alpha = \beta = 2$, we have

$$E_g [|T(X) - h(\theta)|^2] \geq \dot{\eta}(\theta)^T J_g(\theta)^{-1} \dot{\eta}(\theta) \quad (24)$$

with $J_g(\theta) = E_g [\psi_g(X; \theta)\psi_g(X; \theta)^T]$, and with equality if and only if

$$|T(X) - h(\theta)| = k(\theta) |\dot{\eta}(\theta)^T J_g(\theta)^{-1} \psi_g(X; \theta)|. \quad (25)$$

Proof. The denominator of (20) is a quadratic form and we have

$$\begin{aligned} E_g [|T(X) - h(\theta)|^2] &\geq \sup_{A>0} \frac{(\dot{\eta}(\theta)^T A \dot{\eta}(\theta))^2}{E_g [|\dot{\eta}(\theta)^T A \psi_g(X; \theta)|^2]} \\ &\geq \sup_{A>0} \frac{(\dot{\eta}(\theta)^T A \dot{\eta}(\theta))^2}{\dot{\eta}(\theta)^T A E_g [\psi_g(X; \theta)\psi_g(X; \theta)^T] A^T \dot{\eta}(\theta)}. \end{aligned} \quad (26)$$

Let $J_g(\theta) = E_g [\psi_g(X; \theta)\psi_g(X; \theta)^T]$ and set $z(\theta) = A^{\frac{1}{2}}\dot{\eta}(\theta)$. Then, using the inequality $(z^T z)^2 \leq (z^T B z)(z^T B^{-1} z)$ which holds for any $B > 0$, we obtain that

$$\dot{\eta}(\theta)^T J_g(\theta)^{-1} \dot{\eta}(\theta) \geq \sup_{A>0} \frac{(z(\theta)^T z(\theta))^2}{z(\theta)^T A^{\frac{1}{2}} J_g(\theta) \left(A^{\frac{1}{2}}\right)^T z(\theta)},$$

with $B = A^{\frac{1}{2}} J_g(\theta) \left(A^{\frac{1}{2}}\right)^T$.

Since the upper bound is it readily checked to be attained for $A = J_g(\theta)^{-1}$, we finally get (24). Of course, for $g = f$, the inequality (24) reduces to a classical multivariate Cramér-Rao inequality. \square

In the quadratic case, it is also possible to derive 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). Obviously, the inequality involves the generalized Fisher information matrix $J_g(\theta)$. The proof follows the lines in [20, pp. 296-297] and we get the following result.

Proposition 6. *[Multivariate Cramér-Rao inequality – covariance version] Consider the quadratic case, with $\beta = 2$, and the assumptions of Proposition 3, but with $T(x)$ vector valued being an estimator of a vector valued function $h(\theta)$. Set $\eta(\theta) = E_f[T(X)]$, denote $\psi_g(x; \theta) := \frac{\nabla_{\theta} f(x; \theta)}{g(x; \theta)}$ the generalized score function, and define by $J_g(\theta) = E_g[\psi_g(X; \theta)\psi_g(X; \theta)^T]$ the Fisher Fisher information matrix $J_g(\theta)$. Assume that $J_g(\theta)$ is invertible. Then*

$$E_g \left[(T(X) - h(\theta)) (T(X) - h(\theta))^T \right] \succeq \dot{\eta}(\theta)^T J_g(\theta)^{-1} \dot{\eta}(\theta) \quad (27)$$

with $\dot{\eta}(\theta) = \nabla_{\theta} E_f[T(X)^T]$, and equality if and only if

$$(T(X) - h(\theta)) = \dot{\eta}(\theta) J_g(\theta)^{-1} \psi_g(X). \quad (28)$$

Of course, (24) appears as a special case of (27).

Proof. The proof follows the main lines of [20, pp. 296-297] with some small adaptations. With $\eta(\theta) = E_f[T(X)]$, the transpose of the Jacobian matrix, here denoted by $\dot{\eta}(\theta)$, is given by $\dot{\eta}(\theta) = \nabla_{\theta} E_f[T(X)^T]$, which can be rewritten as $\dot{\eta}(\theta) = E_g[\psi_g(X) T(X)^T]$. Recall that $E_g[\psi_g(X)] = 0$. Therefore, for every vectors u and v of ad hoc dimensions, we have

$$E_g \left[u^T (T(X) - h(\theta)) \psi_g(X)^T v \right] = u^T \dot{\eta}(\theta)^T v.$$

By the Schwarz inequality

$$(u^T \dot{\eta}(\theta)^T v)^2 \leq (u^T C_g u) (v^T J_g v) \quad (29)$$

with $C_g := E_g \left[(T(X) - h(\theta)) (T(X) - h(\theta))^T \right]$. If $v = J_g^{-1} \dot{\eta}(\theta) u$, then (29) reduces to $(u^T \dot{\eta}(\theta)^T J_g^{-1} \dot{\eta}(\theta) u) \leq (u^T C_g u)$, which implies (27). For the case of equality, simply observe that

$$E_g \left[(T(X) - h(\theta)) (\dot{\eta}(\theta)^T J_g \psi_g(X))^T \right] = \dot{\eta}(\theta)^T J_g^{-1} \dot{\eta}(\theta),$$

given that $E_g[\psi_g(X)] = 0$. Thus

$$\begin{aligned} & E_g \left[(T(X) - h(\theta) - \dot{\eta}(\theta)^T J_g^{-1} \psi_g(X)) (T(X) - h(\theta) - \dot{\eta}(\theta)^T J_g^{-1} \psi_g(X))^T \right] \\ &= E_g \left[(T(X) - h(\theta)) (T(X) - h(\theta))^T \right] - \dot{\eta}(\theta)^T J_g^{-1} \dot{\eta}(\theta) \end{aligned}$$

which is equal to zero, meaning that the bound is attained in (27) if and only if $\dot{\eta}(\theta) J_g(\theta)^{-1} \psi_g(X) = (T(X) - h(\theta))$. \square

An important consequence of the results above is obtained for a translation parameter, where the generalized Cramér-Rao inequality induces a new class of inequalities. Let $\theta \in \mathbb{R}$ be a scalar location parameter, $x \in X \subseteq \mathbb{R}^n$, and define by $f(x; \theta)$ the family of density $f(x; \theta) = f(x - \theta \mathbf{1})$, where $\mathbf{1}$ is a vector of ones. In this case, we have $\nabla_{\theta} f(x; \theta) = -\mathbf{1}^T \nabla_x f(x - \theta \mathbf{1})$ (provided that f is differentiable at $x - \theta \mathbf{1}$) and the Fisher information becomes a characteristic of the information in the distribution. If X is a bounded subset, we will assume that $f(x)$ vanishes and is differentiable on the boundary ∂X . Without loss of generality, we can assume that the mean of $f(x)$ is zero. Set $h(\theta) = \theta$ and take $T(X) = \mathbf{1}^T X/n$, with of course $\eta(\theta) = E[T(X)] = \theta$ and $\dot{\eta}(\theta) = 1$. Finally, let us choose the particular value $\theta = 0$. In these conditions, the generalized Cramér-Rao inequality (23) becomes

$$E_g \left[|\mathbf{1}^T X|^{\alpha} \right]^{\frac{1}{\alpha}} E_g \left[\left| \mathbf{1}^T \frac{\nabla_x f(X)}{g(X)} \right|^{\beta} \right]^{\frac{1}{\beta}} \geq n, \quad (30)$$

with equality if and only if $\mathbf{1}^T \frac{\nabla_x f(x)}{g(x)} = c(\theta) \text{sign}(\mathbf{1}^T X) |\mathbf{1}^T X|^{\alpha-1}$.

3.2 Multivariate generalized Cramér-Rao inequality – general norms

In the multivariate case, we have another result, which involves an arbitrary norm of the estimation error. Recall that if $\|\cdot\|$ is an arbitrary norm, then its dual norm $\|\cdot\|_*$ is defined by

$$\|Y\|_* = \sup_{\|X\| \leq 1} X.Y, \quad (31)$$

where $X.Y$ is the standard scalar product. For instance, if $\|\cdot\|$ is a L_p -norm, then $\|\cdot\|_*$ is the L_q -norm. The multivariate generalized Cramér-Rao inequality relies on an Hölder-type inequality for vector-valued functions, with arbitrary norm. This inequality was proved in [6]. Let $E = (\mathbb{R}^n, \|\cdot\|)$ be a n -dimensional normed space and denote $E^* = (\mathbb{R}^n, \|\cdot\|_*)$ its dual space. If $X(t)$ and $Y(t)$ are two functions taking values respectively in E and E^* , and if $w(t)$ is a weight function, then

$$\left(\int \|X(t)\|^{\alpha} w(t) dt \right)^{\frac{1}{\alpha}} \left(\int \|Y(t)\|_*^{\beta} w(t) dt \right)^{\frac{1}{\beta}} \geq \left| \int X(t).Y(t) w(t) dt \right| \quad (32)$$

with $\alpha^{-1} + \beta^{-1} = 1$, $\alpha \geq 1$. The equality is obtained if

$$Y(t) = K \|X(t)\|^{\alpha-1} \nabla_{X(t)} \|X(t)\|, \quad \text{with } K > 0. \quad (33)$$

The condition is also necessary if the dual norm is strictly convex. We can now state the following result.

Proposition 7. *[Generalized Cramér-Rao inequality for arbitrary norms, c.f. [6]] Let $f(x; \theta)$ be a multivariate probability density function defined for $x \in$*

$X \subseteq \mathbb{R}^n$, and $\theta \in \Theta \subseteq \mathbb{R}^k$. Let $g(x; \theta)$ denote another probability density function also defined on $(X; \Theta)$. Assume that $f(x; \theta)$ is a jointly measurable function of x and θ , is integrable with respect to x , is absolutely continuous with respect to θ , and that its derivatives with respect to each component of θ are locally integrable. Then, for any estimator $\hat{\theta}(X)$ of $\theta \in \mathbb{R}^n$, and arbitrary norms,

$$E_g \left[\left\| \hat{\theta}(X) - \theta \right\|^\alpha \right]^{\frac{1}{\alpha}} E_g \left[\left\| \frac{\nabla_\theta f(X; \theta)}{g(X; \theta)} \right\|_*^\beta \right]^{\frac{1}{\beta}} \geq \left| \nabla_\theta \cdot E_f[\hat{\theta}(X)] \right| \quad (34)$$

with $\alpha^{-1} + \beta^{-1} = 1$, $\alpha \geq 1$, and where

$$I_\beta[f|g; \theta] = \int_X \left\| \frac{\nabla_\theta f(x; \theta)}{g(x; \theta)} \right\|_*^\beta g(x; \theta) dx \quad (35)$$

is generalized Fisher information of order β on θ taken with respect to g . Equality occurs in (34) if (and only if the dual norm is strictly convex)

$$\frac{\nabla_\theta f(x; \theta)}{g(x; \theta)} = K \left\| \hat{\theta}(x) - \theta \right\|^{\alpha-1} \nabla_{\hat{\theta}(x) - \theta} \left\| \hat{\theta}(x) - \theta \right\|, \quad \text{with } K > 0. \quad (36)$$

Proof. (Sketch of proof, see [6]). (a) evaluate the divergence of $\eta(\theta) = E_f[\hat{\theta}(X)]$, i.e. $\nabla_\theta \cdot \eta(\theta)$; (b) average with respect to g and use the fact that the expectation of the score is zero $\nabla_\theta \cdot \eta(\theta) = \int_X \frac{\nabla_\theta f(x; \theta)}{g(x; \theta)} \cdot \left(\hat{\theta}(x) - \theta \right) g(x; \theta) dx$, (c) apply the Hölder inequality (32), with $X(x) = \hat{\theta}(x) - \theta$, $Y(x) = \frac{\nabla_\theta f(x; \theta)}{g(x; \theta)}$, and $w(x) = g(x; \theta)$. \square

Let us consider again the case of a location parameter, but in the multivariate case $\theta \in \mathbb{R}^n$. The translation family is $f(x; \theta) = f(x - \theta)$. Then $\nabla_\theta f(x; \theta) = -\nabla_x f(x - \theta)$. 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 $\theta = 0$, (34) leads to

Proposition 8. [Functional Cramér-Rao inequality] For any pair of probability density functions, and under some technical conditions,

$$\left(\int_X \|x\|^\alpha g(x) dx \right)^{\frac{1}{\alpha}} \left(\int_X \left\| \frac{\nabla_x f(x)}{g(x)} \right\|_*^\beta g(x) dx \right)^{\frac{1}{\beta}} \geq n, \quad (37)$$

with equality if (and only if when the dual norm is strictly convex) $\nabla_x f(x) = -K g(x) \|x\|^{\alpha-1} \nabla_x \|x\|$.

Finally, let $f(x)$ and $g(x)$ be a pair of escort distributions such as in (19). Then, the generalized Fisher information (17) is obtained in the following result, which also yields a new characterization of q -Gaussian distributions.

Corollary 9. [*q-Cramér-Rao inequality*] Assume that $g(x)$ is a measurable differentiable function of x , which vanishes and is differentiable on the boundary ∂X . Suppose that the involved integrals exist and are finite. Then, for the pair of escort distributions (19), the following *q-Cramér-Rao inequality* holds

$$E_g [\|X\|^\alpha]^{\frac{1}{\alpha}} I_{\beta,q} [g]^{\frac{1}{\beta}} \geq n, \quad (38)$$

$$\text{with } I_{\beta,q} [g] = (q/M_q [g])^\beta E \left[g(x)^{\beta(q-1)} \left\| \frac{\nabla_x g(x)}{g(x)} \right\|_*^\beta \right],$$

with equality if and only if $g(x)$ is a generalized *q-Gaussian*

$$g(x) \propto (1 - \gamma(q-1)\|x\|^\alpha)^{\frac{1}{q-1}} \quad (39)$$

Proof. The result follows from (37), or (30) with $n = 1$, and the fact that for escort distributions,

$$\frac{\nabla_x f(X)}{g(X)} = \frac{q}{M_q [g]} g(X)^{q-1} \frac{\nabla_x g(X)}{g(X)}.$$

The case of equality is obtained by solving the general equality condition $\nabla_x f(x) = -K g(x) \|x\|^{\alpha-1} \nabla_x \|x\|$. \square

As a direct consequence of the *q-Cramér-Rao inequality* (38), the minimum of the generalized Fisher information among all distributions with a given moment of order α , say $m_\alpha = E_g [\|X\|^\alpha]$, is obtained for g a generalized *q-Gaussian* distribution, with parameter γ such that the distribution has the prescribed moment. This extends the well-known result that the minimum of Fisher information with a fixed variance is attained for the Gaussian distribution. This parallels, and complements the known fact that the *q-Gaussians* maximize the *q-entropies* subject to a moment constraint, and yields new variational characterizations of generalized *q-Gaussian* distributions.

Let us also mention that the inequality (38) is similar, but different, to an inequality given by Lutwak et al [21] which is also saturated by the generalized Gaussians (39).

Finally, for a location parameter, the matrix inequality (27) reduces to

$$\text{Cov}_g [X] \geq (J_g)^{-1} = E_g [\psi_g \psi_g^t]^{-1}. \quad (40)$$

When (f, g) is a pair of escort distributions, the equality condition in Proposition 6 shows that equality occurs if and only if g is a generalized *q-Gaussian* with covariance matrix $(J_g)^{-1}$.

4 Conclusions

To sum up and emphasize the main results, let us point out that we have exhibited a generalized Fisher information, both as a by-product of a generalization of de Bruijn identity and Stam inequality and as a fundamental

measure of information in estimation theory. We have drawn a nice interplay between q -entropies, generalized q -Gaussians and the generalized Fisher information. These interrelations yield the generalized q -Gaussians as minimizers of the (β, q) -Fisher information under adequate constraints, or as minimizers of functionals involving q -entropies, (β, q) -Fisher information and/or moments. This is shown through inequalities and identities involving all quantities and generalizing classical information relations (Cramér-Rao's inequality, Stam's inequality, De Bruijn's identity).

In a recent work [8], we introduced a variant of a χ^α -divergence, which enables us to interpret the generalized Fisher information as a limit case of this divergence. Furthermore, this leads to a new derivation of the generalized Cramér-Rao inequality. More interestingly, this also enables us to derive a generalization of the Fisher Information Inequality (FII) – an inequality that characterizes the behavior of the Fisher information under convolution. The equality case, as usual, occurs for generalized q -Gaussian distributions. Details will come in a future paper.

The entropy power inequality (EPI) is a remaining important inequality of information theory, whose proof is difficult. An extension of the EPI to the q -entropy power

$$N_q(X + Y) \geq c_q (N_q(X) + N_q(Y)) \quad (q \geq 1)$$

has been proved very recently in the case $q \geq 1$ [9], and the subject is very hot [31]. In the usual case, the EPI can be derived using de Bruijn's identity and the FII. Thus it would be very interesting to try to derive it from the results presented in this paper and the aforementioned FII.

We have obtained general Cramér-Rao inequalities, in the multivariate case, for a pair of arbitrary densities (f, g) . This led us to the definition of a Fisher information $I_\beta[f|g; \theta]$ about a parameter θ in f , taken with respect to g . In the case of a pair of escort distributions, this Fisher information reduces to the (β, q) -Fisher information $I_{\beta, q}[f; \theta]$, which is associated to the q -entropies and generalized q -Gaussians. Hence, a natural question would be to search for possible generalized entropies associated with the generalized Fisher information $I_\beta[f|g; \theta]$, and perhaps with other families of maximum entropy distributions.

The extensions presented here rely on the idea of modifying the averaging distribution in the computation of the moment of the estimation error. A further step would be to consider general loss functions instead of a simple power of the estimation error. [18], [19] and a work in progress bring interesting results in this direction. However, a remaining issue is again to associate some entropy with the new Cramér-Rao inequalities.

Finally, it would certainly be of interest to go further into estimation and look for estimators efficient with respect to the new Cramér-Rao bounds.

References

1. M. Agueh, Sharp Gagliardo-Nirenberg inequalities via p-Laplacian type equations, *Nonlinear Differential Equations and Applications NoDEA*, 15, 457–472, 2008.

2. S.I. Amari, α -divergence is unique, belonging to both f -divergence and Bregman divergence classes. *IEEE Transactions on Information Theory*, 55, 11, 4925–4931, 2009.
3. E. W. Barankin, Locally best unbiased estimates, *The Annals of Mathematical Statistics*, 20, 477–501, 1949.
4. J.-F. Bercher, On a (β, q) -generalized Fisher information and inequalities involving q -Gaussian distributions, *Journal of Mathematical Physics*, 53, p. 063303, 2012.
5. J.-F. Bercher, On generalized Cramér-Rao inequalities, generalized Fisher informations and characterizations of generalized q -Gaussian distributions, *Journal of Physics A: Mathematical and Theoretical*, 45, p. 255303, 2012.
6. J.-F. Bercher, On multidimensional generalized Cramér-Rao inequalities, uncertainty relations and characterizations of generalized q -Gaussian distributions, *Journal of Physics A: Mathematical and Theoretical*, 46, p. 095303, 2013.
7. J.-F. Bercher, Some properties of generalized Fisher information in the context of nonextensive thermostatics, *Physica A: Statistical Mechanics and its Applications*, 392, 3140–3154, 2013.
8. J.-F. Bercher, Some results on a χ -divergence, an extended Fisher information and generalized Cramér-Rao inequalities, in *Geometric Science of Information*, F. Nielsen and F. Barbaresco, eds., no. 8085 in Lecture Notes in Computer Science, Springer Berlin Heidelberg, 487–494, 2013.
9. S. G. Bobkov and G. P. Chistyakov, Entropy power inequality for the Rényi entropy, eprint: <http://www.math.umn.edu/~bobkov/preprints/>, May 2013.
10. J. A. Carrillo and G. Toscani, Asymptotic L1-decay of solutions of the porous medium equation to self-similarity, *Indiana University Mathematics Journal*, 49, 113–142, 2000.
11. D. Cordero-Erausquin, B. Nazaret, and C. Villani, A mass-transportation approach to sharp Sobolev and Gagliardo-Nirenberg inequalities, *Advances in Mathematics*, 182, 307–332, 2004.
12. T. M. Cover and J. A. Thomas, *Elements of Information Theory*, 2nd Edition, Wiley-Interscience, 2 ed., July 2006.
13. J. Dehesa, P. Sánchez-Moreno, and R. Y. áñez, Cramér-Rao information plane of orthogonal hypergeometric polynomials, *Journal of Computational and Applied Mathematics*, 186, 523–541, 2006.
14. A. Dembo, T. Cover, and J. Thomas, Information theoretic inequalities, *IEEE Transactions on Information Theory*, 37, 1501–1518, 1991.
15. J. Dolbeault and G. Toscani, *Improved interpolation inequalities, relative entropy and fast diffusion equations*, *Annales de l'Institut Henri Poincaré (C) Non Linear Analysis*, 30, 917–934, 2013.
16. B. R. Frieden, *Science from Fisher Information: A Unification*, Cambridge University Press, 2004.
17. O. Johnson and C. Vignat, Some results concerning maximum rényi entropy distributions, *Annales de l'Institut Henri Poincaré (B) Probability and Statistics*, 43, 339–351, 2007.
18. A. Kozek, Efficiency and Cramér-Rao type inequalities for convex loss functions, *Journal of Multivariate Analysis*, 7, 89–106, 1977.
19. F. Liese, A Cramer-Rao type inequality for a convex loss function, in Transactions of the Tenth Prague Conference on Information Theory, Statistical Decision Functions, Random Processes, no. 10A-B, Springer Netherlands, 121–128, Jan. 1987.

20. F. Liese and K.-J. Miescke, *Statistical Decision Theory: Estimation, Testing, and Selection*, Springer, 2008 ed., June 2008.
21. E. Lutwak, S. Lv, D. Yang, and G. Zhang, Extensions of Fisher information and Stam's inequality, *IEEE Transactions on Information Theory*, 58, 1319–1327, 2012.
22. E. Lutwak, D. Yang, and G. Zhang, Cramér-Rao and moment-entropy inequalities for Rényi entropy and generalized Fisher information, *IEEE Transactions on Information Theory*, 51, 473–478, 2005.
23. B. S. Nagy, Über integralungleichungen zwischen einer funktion und ihrer ableitung, *Acta Scientiarum Mathematicarum (Szeged)*, 10, 64–74, 1941.
24. E. Romera and J. S. Dehesa, The Fisher-Shannon information plane, an electron correlation tool, *The Journal of Chemical Physics*, 120, 8906–12, 2004.
25. G. Savaré and G. Toscani, The concavity of Rényi entropy power, arXiv:1208.1035, 2012.
26. A. Stam, Some inequalities satisfied by the quantities of information of Fisher and Shannon, *Information and Control*, 2, 101–112, 1959.
27. I. Vajda, χ^α -divergence and generalized Fisher information, in Transactions of the Sixth Prague Conference on Information Theory, Statistical Decision Functions and Random Processes, 1973, 223–234, 1973.
28. J. L. Vázquez, *The Porous Medium Equation: Mathematical Theory*, Oxford University Press, USA, 1 ed., Dec. 2006.
29. J. L. Vázquez, *Smoothing and Decay Estimates for Nonlinear Diffusion Equations: Equations of Porous Medium Type*, Oxford University Press, USA, Oct. 2006.
30. C. Vignat and J. F. Bercher, Analysis of signals in the Fisher-Shannon information plane, *Physics Letters A*, 312, 27–33, 2003.
31. L. Wang and M. Madiman, Beyond the entropy power inequality, via rearrangements, arXiv e-print 1307.6018, July 2013.