MDPI

Article

ϕ-informational measures: some results and interrelations

Steeve Zozor [1,*] and Jean-François Bercher [2]

[1] Univ. Grenoble Alpes, CNRS, Grenoble INP*, GIPSA-Lab, 38000 Grenoble, France
*Institute of Engineering Univ. Grenoble Alpes
steeve.zozor@cnrs.fr

[2] LIGM, Univ Gustave Eiffel, CNRS, F-77454 Marne-la-Vallée, France
jean-francois.bercher@esiee.fr

* Correspondence: steeve.zozor@cnrs.fr

Abstract: In this paper we focus on extended informational measures based on a convex function ϕ : entropies, extended Fisher information, generalized moments. Both the generalization of the Fisher information and the moments rely on the definition of an escort distribution linked to the (entropic) functional ϕ. We revisit the usual maximum entropy principle –more precisely its inverse problem, starting from the distribution and constraints–, which leads to the introduction of state-dependent ϕ-entropies. Then, we examine interrelations between the extended informational measures and generalize relationships such that Cramér-Rao inequality and the de Bruijn identity in this broader context. In this particular framework, the maximum entropy distributions play a central role. Of course, all the results derived in the paper include the usual ones as special cases.

Keywords: ϕ-entropy; state-dependent ϕ-entropy; (inverse) maximum ϕ-entropy problem; ϕ-escort distributions; ϕ-Fisher information; ϕ-moments; generalized Cramér-Rao inequality; ϕ-heat equation; generalized de Bruijn identity.

1. Introduction

Since the pioneer works of von Neumann [1], Shannon [2], Boltzmann, Maxwell, Planck and Gibbs [3–9], many investigations were devoted to the generalization of the so-called Shannon entropy and its associated measures [10–22]. If the Shannon measures are compelling, especially in the communication domain, for compression purposes, many generalizations proposed later on has also showed promising interpretations and applications (Panter-Dite formula in quantification where the Rényi or Havdra-Charvát entropy emerges [23–25], codification penalizing long codewords where the Rényi entropy appears [26,27] for instance). The great majority of the extended entropies found in the literature belongs to a very general class of entropic measures called (h, ϕ)-entropies [13,19,20,28–30]. Such a general class (or more precisely the subclass of ϕ-entropies) can be traced back to the work of Burbea & Rao [28]. They offer not only a general framework to study general properties shared by special entropies, but they also offer many potential applications as described for instance in [30]. Note that if a large amount of work deals with divergences, entropies occur as special cases when one takes a uniform reference measure.

In the framework of these generalized entropies, the so-called maximum entropy principle takes a special place. This principle, advocated by Jaynes, states that the statistical distribution that describes a system in equilibrium maximizes the entropy while satisfying the system's physical constraints (e.g., the center of mass, energy) [31–35]. In other words, it is the less informative law given the constraints of the system. In the Bayesian approach, dealing with the stochastic modeling of a parameter, such a principle (or a minimum divergence principle) is often used to choose a prior distribution for the parameter [22,36–39]. It also finds its counterpart in communication, clustering, pattern

Citation: Zozor, S.; Bercher, J.-F. ϕ-informational measures: some results and interrelations. *Entropy* **2021**, *1*, 0. https://doi.org/

Received:
Accepted:
Published:

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.

recognition, problems, among many others [32,33,40–43]. In statistics, some goodness-of-fit tests are based on entropic criteria derived from the same idea of constrained maximal entropic law [44–49]. In a large number of works using the maximum entropy principle, the entropy used is the Shannon entropy. However, if for some reason a generalized entropy is considered, the approach used in the Shannon case does not fundamentally change [50–53].

One can consider the inverse problem which consists in finding the moment constraints leading to the observed distribution as a maximal entropy distribution [50]. Kesavan & Kapur also envisaged a second inverse problem, where both the distribution and the moments are given. The question is thus to determine the entropy so that the distribution is its maximizer. As a matter of fact, dealing with the Shannon entropy, whatever the constraints considered, the maximum entropy distribution falls in the exponential family [33,34,52,54]. Considering more general entropies allows to escape from this limitation. Moreover, if the Shannon entropy (or the Gibbs entropy in physics) is well adapted to the study of systems in the equilibrium (or in the thermodynamic limit), extended entropies allow a finer description of systems out of equilibrium [17,55–59], exhibiting their importance. While the problem was considered mainly in the discrete setting by Kesavan & Kapur in [50], we will recall it in the general framework of the ϕ-entropies probability densities with respect to any reference measure, and make a further step considering an extended class of these entropies.

While the entropy is a widely used tool for quantifying information (or uncertainty) attached to a random variable or to a probability distribution, other quantities are used as well, such as the moments of the variable (giving information for instance on center of mass, dispersion, skewness, impulsive character), or the Fisher information. In particular, the Fisher information appears in the context of estimation [60,61], in Bayesian inference through the Jeffreys prior [39,62], but also for complex physical systems descriptions [61,63–67].

Although coming from different worlds (information theory and communication, estimation, statistics, physics), these informational quantities are linked by well-known relations such as the Cramér-Rao inequality, the de Bruijn identity, the Stam inequality [34,68–70]. These relationships have been proved very useful in various areas, for instance in communications [34,68,69], in estimation [60] or in physics [71,72], among others. When generalized entropies are considered, it is natural to question the other informational measures' generalization and the associated identities or inequalities. This question gave birth to a large amount of work and is still an active field of research [28,73–84].

In this paper, we show that it is possible to build a whole framework, which associates a target maximum entropy distribution to generalized entropies, generalized moments and generalized Fisher information. In this setting, we derive generalized inequalities and identities relating these quantities, which are all linked in some sense to the maximum entropy distribution.

The paper is organized as follows. In section 2 we recall the definition of the generalized ϕ-entropy. Thus, we come back to the maximum entropy problem in this general settings. Following the sketch of [50], we present a sufficient condition linking the entropic functional and the maximizing distribution, allowing to both solve the direct and the inverse problems. When the sufficient conditions linking the entropic function and the distribution cannot be satisfied, the problem can be solved by introducing state-dependent generalized entropies, which is the purpose of section 3. In section 4, we introduce informational quantities associated to the generalized entropies of the previous sections, such that a generalized escort distribution, generalized moments and generalized Fisher informations. These generalized informational quantities allow to extend the usual informational relations such that the Cramér-Rao inequality, relations precisely saturated (or valid) for the generalized maximum entropy distribution. Finally, in section 5, we show that the extended quantities allows to obtain an extended de Bruijn identity, provided the distribution follows a non-linear heat equation. Some examples of ϕ-entropies solving the inverse maximum entropy problem are provided in a short series of appendices, showing in particular that the usual quantities are recovered as particular cases (Gaussian distribution, Shannon entropy, Fisher information, variance).

In what follows, we will define a series of generalized information quantities relative to a probability density defined with respect to a given reference measure μ (e.g., the Lebesgue measure when dealing with continuous random variables, discrete measure for discrete-state random variables,...). Therefore, rigorously, all these quantities depend on the particular choice of this reference measure. However, for simplicity, we will omit to mention this dependence in the notations along the paper.

2. ϕ-entropies – direct and inverse maximum entropy problems

Let us first recall the definition of the generalized ϕ-entropies introduced by Csiszàr in terms of divergence, and by Burbea and Rao in terms of entropy:

Definition 1 (ϕ-entropy [28]). *Let $\phi : \mathcal{Y} \subseteq \mathbb{R}_+ \mapsto \mathbb{R}$ be a convex function defined on a convex set $\mathcal{Y}$. Then, if f is a probability distribution defined with respect to a general measure μ on a set $\mathcal{X} \subseteq \mathbb{R}^d$ such that $f(\mathcal{X}) \subseteq \mathcal{Y}$, when this quantity exists,*

$$H_\phi[f] = -\int_{\mathcal{X}} \phi(f(x))\,\mathrm{d}\mu(x) \tag{1}$$

is the ϕ-entropy of f.

The (h,ϕ)-entropy is defined by $H_{(h,\phi)}[f] = h\big(H_\phi[f]\big)$ where h is a nondecreasing function. The definition is extended by allowing ϕ to be concave, together with h nonincreasing [13,19,20,29,30]. If additionally h is concave, then the entropy functional $H_{(h,\phi)}[f]$ is concave.

Since we are interested in the maximum entropy problem and because h is monotone, we can restrict our study to the ϕ-entropies. Additionally, we will assume that ϕ is *strictly* convex and *differentiable*.

A related quantity is the Bregman divergence associated with convex function ϕ:

Definition 2 (Bregman divergence and functional Bregman divergence [22,85]). *With the same assumptions as in def. 1, the Bregman divergence associated with ϕ defined on a convex set $\mathcal{Y}$ is given by the function defined on $\mathcal{Y} \times \mathcal{Y}$,*

$$B_\phi(y_1, y_2) = \phi(y_1) - \phi(y_2) - \phi'(y_2)(y_1 - y_2). \tag{2}$$

Applied to two functions $f_i : \mathcal{X} \mapsto \mathcal{Y}$, $i = 1, 2$, the functional Bregman divergence writes

$$\mathcal{B}_\phi(f_1, f_2) = \int_{\mathcal{X}} \phi(f_1(x))\,\mathrm{d}\mu(x) - \int_{\mathcal{X}} \phi(f_2(x))\,\mathrm{d}\mu(x) - \int_{\mathcal{X}} \phi'(f_2(x))(f_1(x) - f_2(x))\,\mathrm{d}\mu(x). \tag{3}$$

A direct consequence of the strict convexity of ϕ is the nonnegativity of the (functional) Bregman divergence: $B_\phi(y_1, y_2) \geq 0$ and $\mathcal{B}_\phi(f_1, f_2) \geq 0$, with equality if and only if $y_1 = y_2$ and $f_1 = f_2$ almost everywhere respectively.

More generally, the Bregman divergence is defined for multivariate convex functions, where the derivative is replaced by gradient operator [85]. Extensions for convex function of functions also exist, where the derivative is in the sense of Gâteau [86]. Such general extensions are not useful for our purposes; thus, we restrict ourselves to the above definition where $\mathcal{Y} \subseteq \mathbb{R}_+$.

2.1. Maximum entropy principle: the direct problem

Let us first recall the maximum entropy problem that consists in searching for the distribution maximizing the ϕ-entropy (1) subject to constraints on some moments $\mathbb{E}[T_i(X)]$ with $T_i : \mathbb{R}^d \mapsto \mathbb{R}$, $i = 1, \ldots, n$. This direct problem writes

$$f^\star = \underset{f \in \mathcal{D}_{T,t}}{\operatorname{argmax}} \left(- \int_{\mathcal{X}} \phi(f(x)) \, \mathrm{d}\mu(x) \right) \tag{4}$$

with

$$\mathcal{D}_{T,t} = \{ f \geq 0 : \mathbb{E}[T_i(X)] = t_i, i = 0, \ldots, n \}, \tag{5}$$

where $T_0(x) = 1$ and $t_0 = 1$ (normalization constraint), $T = (T_0, \ldots, T_n)$, $t = (t_0, \ldots, t_n)$. The maximization problem being strictly concave, the solution exists and is unique. A technique to solve the problem can be to use the classical Lagrange multipliers technique and to solve the Euler-Lagrange equation from the variational problem, but this approach requires mild conditions [50,51,53,87–89]. In the following proposition, we recall a sufficient condition relating f and ϕ so that f is the problem's solution. This result is proven without the use of the Lagrange technique.

Proposition 1 (Maximal ϕ-entropy solution [50])**.** *Suppose that there exists a probability distribution $f \in \mathcal{D}_{T,t}$ satisfying*

$$\phi'(f(x)) = \sum_{i=0}^{n} \lambda_i \, T_i(x), \tag{6}$$

for some $(\lambda_0, \ldots, \lambda_n) \in \mathbb{R}^{n+1}$. Then, f is the unique solution of the maximal entropy problem (4).

Proof. Suppose that distribution f satisfies eq. (6) and consider any distribution $g \in \mathcal{D}_{T,t}$. The functional Bregman divergence between f and g writes

$$\begin{aligned}
\mathcal{B}_\phi(g, f) &= \int_{\mathcal{X}} \phi(g(x)) \, \mathrm{d}\mu(x) - \int_{\mathcal{X}} \phi(f(x)) \, \mathrm{d}\mu(x) - \int_{\mathcal{X}} \phi'(f(x))(g(x) - f(x)) \, \mathrm{d}\mu(x) \\
&= -H_\phi[g] + H_\phi[f] - \sum_{i=0}^{n} \lambda_i \int_{\mathcal{X}} T_i(x)(g(x) - f(x)) \, \mathrm{d}\mu(x) \\
&= H_\phi[f] - H_\phi[g]
\end{aligned}$$

where we used the fact that g and f are both probability distributions with the same moments $\mathbb{E}[T_i(X)] = t_i$. By nonnegativity of the Bregman functional divergence, we finally get that

$$H_\phi[f] \geq H_\phi[g]$$

for all distribution g with the same moments as f, with equality if and only if $g = f$ almost everywhere. In other words, this shows that if f satisfies eq. (6), then it is the desired solution. □

Hence, given an entropic functional ϕ and moments constraints T_i, eq. (6) leads the the maximum entropy distribution $f^\star$. This distribution is parameterized by the λ_is or, equivalently, by the moments t_is.

Note that the reciprocal is not necessarily true, as shown for instance in [53]. However, the reciprocal is true when $\mathcal{X}$ is a compact [89] or for any $\mathcal{X}$ provided that ϕ is locally bounded on $\mathcal{X}$ [90].

2.2. Maximum entropy principle: the inverse problems

As stated in the introduction, two inverse problems can be considered starting from a given distribution f. These problems were considered by Kesavan & Kapur in [50] in the discrete framework.

The first inverse problem consists in searching for the adequate moments so that a desired distribution f is the maximum entropy distribution of a given ϕ-entropy. This amounts to find functions T_i and coefficients λ_i satisfying eq. (6). This is not always an easy task, and even not always possible. For instance, it is well known that the maximum Shannon entropy distribution given moment constraints falls in the exponential family [33,34,52,54]. Therefore, if f does not belong to this family, the problem has no solution.

The second inverse problem consists in designing the entropy itself, given a target distribution f and given the T_is. In other words, given a distribution f, eq. (6) may allow to determine the entropic functional ϕ so that f is its maximizer.

As for the direct problem, in the second inverse problem, the solution is parameterized by the λ_is. Here, required properties on ϕ will shape the domain the λ_is live in. In particular, ϕ must satisfy:

- the domain of definition of ϕ' must include $f(\mathcal{X})$; this will be satisfied by construction;
- from the strict convexity property of ϕ, ϕ' must be strictly increasing.

Hence, because ϕ' must be strictly increasing, its clear that solving eq. (6) requires the following two conditions:

(C1) $f(x)$ and $\sum_{i=1}^{n} \lambda_i \, T_i(x)$ must have the same variations, i.e., $\sum_{i=0}^{n} \lambda_i \, T_i(x)$ is increasing (resp. decreasing, constant) where f is increasing (resp. decreasing, constant);

(C2) $f(x)$ and $\sum_{i=1}^{n} \lambda_i \, T_i(x)$ must have the same level sets,

$$f(x_1) = f(x_2) \;\Leftrightarrow\; \sum_{i=0}^{n} \lambda_i \, T_i(x_1) \;=\; \sum_{i=0}^{n} \lambda_i \, T_i(x_2).$$

For instance, in the univariate case, for one moment constraint,

- for $\mathcal{X} = \mathbb{R}_+$, $T_1(x) = x$, λ_1 must be negative and $f(x)$ must be decreasing,
- for $\mathcal{X} = \mathbb{R}$, $T_1(x) = x^2$ or $T_1(x) = |x|$, λ_1 must be negative and $f(x)$ must be even and unimodal.

Under conditions (C1) and (C2), the solutions of eq. (6) are given by

$$\phi'(y) = \sum_{i=0}^{n} \lambda_i \, T_i\Big(f^{-1}(y)\Big) \tag{7}$$

where f^{-1} can be multivalued. However, even if f^{-1} is multivalued, because of condition (C2), ϕ' is defined univocally.

Eq. (7) provides thus an effective way to solve the inverse problem. However, there exist situations where there do not exist any set of λ_is such that conditions (C1)-(C2) are satisfied (e.g., $T_1(x) = x^2$ with f not even). In such a case, a way to go is to extend the definition of the ϕ-entropy. This will be the purpose of section 3. Before, let us illustrate the previous result on some examples.

2.3. Second inverse maximum entropy problem: some examples

To illustrate the previous subsection, let us analyze briefly three examples: the famous Gaussian distribution (example 1), the q-Gaussian distribution also intensively studied (example 2) and the arcsine distribution (example 3), both three with a second-order moment constraint. The Gaussian, q-Gaussian, and arcsine distributions will serve as a guideline all along the paper. The details of the calculus, together with a deeper study related to the sequel of the paper, are rejected in the appendix. Other examples are also given in this appendix. In both three examples, except in the next section, we consider the second-order moment constraint $T_1(x) = x^2$.

Example 1. *Let us consider the well-known Gaussian distribution* $f_X(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{x^2}{2\sigma^2}\right)$, *defined over* $\mathcal{X} = \mathbb{R}$, *and let us search for the* ϕ*-entropy so that the Gaussian is its maximizer subject to the constraint* $T_1(x) = x^2$. *To satisfy condition (C1) we must have* $\lambda_1 < 0$, *whereas condition (C2) is always satisfied. Rapid calculations, detailed in appendix A.1, and a reparameterization of the* λ_i*s, give the entropic functional*

$$\phi(y) = \alpha\, y \log(y) + \beta\, y + \gamma \quad \textit{with} \quad \alpha > 0.$$

This is nothing but the Shannon entropy, up to the scaling factor α, *and a shift (to avoid the divergence of the entropy when* $\mathcal{X}$ *is unbounded, one will take* $\gamma = 0$*). One thus recovers the long outstanding fact that the Gaussian is the maximum Shannon entropy distribution with the second order moment constraint.*

Example 2. *Let us consider the* q*-Gaussian distribution, also known as Tsallis distribution or Student distribution [91,92],* $f_X(x) = A_q \left(1 - (q-1)\frac{x^2}{\sigma^2}\right)_+^{\frac{1}{1-q}}$, *where* $q > 0$, $q \neq 1$, $x_+ = \max(x, 0)$ *and* A_q *is the normalization coefficient, defined over* $\mathcal{X} = \mathbb{R}$ *when* $q < 1$ *or over* $\mathcal{X} = \left(-\frac{\sigma}{\sqrt{q-1}}\,;\, \frac{\sigma}{\sqrt{q-1}}\right)$ *when* $q > 1$, *and let search for the* ϕ*-entropy so that the* q*-Gaussian is its maximizer with the constraint* $T_1(x) = x^2$. *Here again, condition (C1) is satisfied if and only if* $\lambda_1 < 0$ *whereas condition (C2) is always satisfied. Rapid calculations detailed in appendix A.2 lead to the entropic functional, after a reparameterization of the* λ_i*s, as,*

$$\phi(y) = \alpha\, \frac{y^q - y}{q - 1} + \beta\, y + \gamma \quad \textit{with} \quad \alpha > 0,$$

where q *is thus an additional parameter of the family. This entropy is nothing but the Havrda-Charvát or Daróczy or Tsallis entropy [12,14,17,91], up to the scaling factor* α, *and a shift (here also, to avoid the divergence of the entropy when* $\mathcal{X}$ *is unbounded, one will take* $\gamma = 0$*). This entropy is also closely related to the Rényi entropy [10] via a one-to-one logarithmic mapping. One recovers the also well known fact that the* q*-Gaussian is the maximum Havrda-Charvát-Rényi-Tsallis entropy distribution with the second order moment constraint [91]. In the limit case* $q \to 1$, *the distribution* f_X *tends to the Gaussian, whereas the Havrda-Charvát-Rényi-Tsallis entropy tends to the Shannon entropy.*

Example 3. *Consider the arcsine distribution,* $f_X(x) = \frac{1}{\sqrt{s^2 - \pi^2 x^2}}$ *where* $s > 0$, *defined over* $\mathcal{X} = \left(-\frac{s}{\pi}\,;\, \frac{s}{\pi}\right)$ *and let us determine the entropic functionals* ϕ *so that* f_X *is the maximum* ϕ*-entropy distribution subject to the constraint* $T_1(x) = x^2$. *Condition (C2) is always satisfied and now, to fulfill condition (C1) we must impose* $\lambda_1 > 0$. *Some algebra detailed in appendix A.4.1 leads to the entropic functional, after a reparameterization of the* λ_i*s,*

$$\phi(y) = \frac{\alpha}{y} + \beta\, y + \gamma \quad \textit{with} \quad \alpha > 0$$

(again, to avoid the divergence of the entropy one can adjust parameter γ*). This entropy is unusual and, due to its form, is potentially finite only for densities defined over a bounded support and that are divergent in its boundary (integrable divergence).*

3. State-dependent entropic functionals and mimization revisited

In order to follow asymmetries of the distribution f and address the limitation raised by conditions (C1) and (C2), an idea is to allow the entropic functional to also depend on the state variable x:

Definition 3 (State-dependent ϕ-entropy)**.** *Let $\phi : \mathcal{X} \times \mathcal{Y} \mapsto \mathbb{R}$ such that for any $x \in \mathcal{X} \subseteq \mathbb{R}^d$, function $\phi(x, \cdot)$ is a convex function on the closed convex set $\mathcal{Y} \subseteq \mathbb{R}_+$. Then, if f is a probability distribution defined with respect to a general measure μ on set $\mathcal{X}$ and such that $f(\mathcal{X}) \subseteq \mathcal{Y}$,*

$$H_\phi[f] = -\int_{\mathcal{X}} \phi(x, f(x))\, \mathrm{d}\mu(x) \tag{8}$$

will be called state-dependent ϕ-entropy of f. Since $\phi(x, \cdot)$ is convex, then the entropy functional $H_\phi[f]$ is concave. A particular case arises when, for a given partition $(\mathcal{X}_1, \ldots, \mathcal{X}_k)$ of $\mathcal{X}$, functional ϕ writes

$$\phi(x, y) = \sum_{l=1}^{k} \phi_l(y) \mathbb{1}_{\mathcal{X}_l}(x) \tag{9}$$

where $\mathbb{1}_A$ denotes the indicator of set A. This functional can be viewed as a "$(\mathcal{X}_1, \ldots, \mathcal{X}_k)$-extension" over $\mathcal{X} \times \mathcal{Y}$ of a multiform function defined on $\mathcal{Y}$, with k branches ϕ_l and the associated ϕ-entropy will be called $(\mathcal{X}_1, \ldots, \mathcal{X}_k)$-multiform ϕ-entropy.

As in the previous section, we restrict our study to functionals $\phi(x, y)$ *strictly convex and differentiable* with respect to y.

Following the lines of section 2, a generalized Bregman divergence can be associated to ϕ under the form $B_\phi(x, y_1, y_2) = \phi(x, y_1) - \phi(x, y_2) - \frac{\partial \phi}{\partial y}(x, y_2)(y_1 - y_2)$, and a generalized functional Bregman divergence $\mathcal{B}_\phi(f_1, f_2) = \int_{\mathcal{X}} B_\phi(x, f_1(x), f_2(x))\, \mathrm{d}\mu(x)$.

With these extended quantities, the direct problem becomes

$$f^\star = \underset{f \in \mathcal{D}_{T,t}}{\operatorname{argmax}} \left(-\int_{\mathcal{X}} \phi(x, f(x))\, \mathrm{d}\mu(x) \right) \tag{10}$$

Although the entropic functional is now state dependent, the approach adopted before can be applied here, leading to

Proposition 2 (Maximum state-dependent ϕ-entropy solution)**.** *Suppose that there exists a probability distribution f satisfying*

$$\frac{\partial \phi}{\partial y}(x, f(x)) = \sum_{i=0}^{n} \lambda_i\, T_i(x), \tag{11}$$

for some $(\lambda_0, \ldots, \lambda_n) \in \mathbb{R}^{n+1}$, then f is the unique solution of the extended maximum entropy problem (10)*.*
If ϕ is chosen in the $(\mathcal{X}_1, \ldots, \mathcal{X}_k)$-multiform ϕ-entropy class, this sufficient condition writes

$$\sum_{l=1}^{k} \phi_l'(f(x))\, \mathbb{1}_{\mathcal{X}_l}(x) = \sum_{i=0}^{n} \lambda_i\, T_i(x), \tag{12}$$

Proof. The proof follows the steps of Proposition 1, using the generalized functional Bregman divergence instead of the usual one. □

Resolving eq. (11) is not possible in all generality. However the sufficient condition (12) can be rewritten as

$$\sum_{l=1}^{k}\left(\phi_l'(f(x)) - \sum_{i=0}^{n}\lambda_i\,T_i(x)\right)\mathbb{1}_{\mathcal{X}_l}(x) = 0. \tag{13}$$

Therefore, if there exists (at least) a set of λ_is such that condition (C1) is satisfied (but not necessarily (C2)), one can always

- design a partition $(\mathcal{X}_1,\ldots,\mathcal{X}_k)$ so that (C2) is satisfied *in each* $\mathcal{X}_l$ (at least, such that f is either strictly monotonic, or constant, on $\mathcal{X}_l$)
- determine ϕ_l as in eq. (7) in each $\mathcal{X}_l$, that is

$$\phi_l'(y) = \sum_{i=0}^{n}\lambda_i\,T_i\left(f_l^{-1}(y)\right) \tag{14}$$

  where f_l^{-1} is the (possibly multivalued) inverse of f on $\mathcal{X}_l$. By the way, when $\mathcal{X}_l$ is such that f_X is monotonic on it ensures that f_l^{-1} is univalued.

In short, in the case where only condition (C1) is satisfied, one can obtain an extended entropic functional of $(\mathcal{X}_1,\ldots,\mathcal{X}_k)$-multiform class so that eq. (13) provides an effective way to solve the inverse problem in the state-dependent entropic functional context. This is given by eq. (14).

Note however that it still may happen that there is no set of λ_is allowing to satisfy (C1). In this harder context, the problem remains solvable when the moments are defined as partial moments like $\mathbb{E}\left[T_{l,i}(X)\mathbb{1}_{\mathcal{X}_l}(X)\right] = t_{l,i}$, $l = 1,\ldots,k$ and $i = 1,\ldots,n_l$ and when there exists on $\mathcal{X}_l$ a set of $\lambda_{l,i}$s such that (C1) and (C2) hold. The solution still writes as in eq. (14), but where now n, the λ_is and the T_is are replaced by n_l, the $\lambda_{l,i}$s and $T_{l,i}$s respectively,

$$\phi_l'(y) = \sum_{i=0}^{n_l}\lambda_{l,i}\,T_{l,i}\left(f_l^{-1}(y)\right) \tag{15}$$

Let us now come back to the arcsine example $f_X(x) = \frac{1}{s^2-\pi^2x^2}$, defined over $\mathcal{X} = \left(-\frac{s}{\pi}\,;\,\frac{s}{\pi}\right)$ (example 3) of the previous section, when now we constraint the first order moment or partial first order moments.

Example 3-1. *Let us now consider this arcsine distribution, constrained uniformly by* $T_1(x) = x$. *Clearly, neither condition (C1) nor condition (C2) can be satisfied. Note that the arcsine distribution is a one-to-one function on each set* $\mathcal{X}_- = \left(-\frac{s}{\pi}\,;0\right)$ *and* $\mathcal{X}_+ = \left[0\,;\frac{s}{\pi}\right)$ *that partitions* $\mathcal{X}$. *Therefore, considering multiform entropic functionals with this partition allows to overcome the issue on condition (C2), but that on condition (C1) remains. If we ignore this issue and apply eq. (14), after a reparameterization of the* λ_i*s, we obtain* $\widetilde{\phi}_\pm(y) = \widetilde{\phi}_{\pm,\mathrm{u}}(sy)$ *with* $\widetilde{\phi}_{\pm,\mathrm{u}}(y) = \pm\alpha\left(\sqrt{u^2-1}+\arctan\left(\frac{1}{\sqrt{u^2-1}}\right)\right)\mathbb{1}_{(1\,;+\infty)}(u)+\beta\,u+\gamma_\pm$ *where s is thus an additional parameter of the family. It appears that whereas these functionals are defined for* $u > 1$, *one can extend them continuously and with a continuous derivative for any* $u > 0$ *imposing* $\beta = 0$, *which finally leads to the family*

$$\widetilde{\phi}_\pm(y) = \widetilde{\phi}_{\pm,\mathrm{u}}(sy) \qquad \textit{with}$$

$$\widetilde{\phi}_{\pm,\mathrm{u}}(y) = \pm\alpha\left(\sqrt{u^2-1}+\arctan\left(\frac{1}{\sqrt{u^2-1}}\right)\right)\mathbb{1}_{(1\,;+\infty)}(u)\;+\;\gamma_\pm$$

However, the functional are no more convex. (see appendix A.4.2 for more details).

Example 3-2. *If now we impose the partial constraint $T_{\pm,1}(x) = x \mathbb{1}_{\mathcal{X}_\pm}(x)$, and search for the ϕ-entropy so that f_X is the maximizer subject to these constraints, condition (C1) can be now satisfied on each $\mathcal{X}_\pm$ by imposing the $\pm\lambda_{\pm,1}$ given eq. (15) to be positive. We then obtain the associated multiform entropic functional, after a reparameterization of the λ_is, as $\phi_\pm(y) = \phi_{\pm,\mathrm{u}}(sy)$ with $\phi_{\pm,\mathrm{u}}(u) = \alpha_\pm\left(\sqrt{u^2-1} + \arctan\left(\frac{1}{\sqrt{u^2-1}}\right)\right) \mathbb{1}_{(1\,;+\infty)}(u) + \beta\, u + \gamma_\pm$ with $\alpha_\pm > 0$ and where s is thus an additional parameter of the family. In this case, the entropic functionals can be considered for any $u > 0$ by imposing $\beta = 0$ and one can check that the obtained functions are of class C^1. This finally leads to the family*

$$\phi_\pm(y) = \widetilde{\phi}_{\pm,\mathrm{u}}(sy) \qquad \textit{with}$$

$$\phi_{\pm,\mathrm{u}}(y) = \alpha_\pm\left(\sqrt{u^2-1} + \arctan\left(\frac{1}{\sqrt{u^2-1}}\right)\right) \mathbb{1}_{(1\,;+\infty)}(u) \,+\, \gamma_\pm, \quad \alpha_\pm > 0$$

In addition, remarkably, the entropic functional can be made univalued by choosing $\alpha_+ = \alpha_-$ and $\gamma_+ = \gamma_-$. In fact, such a choice is equivalent to considering the constraint $T_1(x) = |x|$ which respects the symmetries of the distribution, and allows to recover a classical ϕ-entropy. (see appendix A.4.2 for more details).

At a first glance, the solutions of example 3-1 and example 3-2 seem to be identical. In fact, they drastically differ. Indeed, let us emphasize that the problem has one constraint in the first case, but two in the second case. The consequence is that 4 parameters parameterize the first solution $\beta, \gamma_\pm$ and α, while 5 parameters $\beta, \gamma_\pm$ and $\alpha_\pm$ parameterize the second solution. This difference is not insignificant: the first case cannot be viewed as a special case of the second one, because $\alpha_\pm$ must be positive, which cannot be possible with only parameter α since $\pm\alpha$ rule the $\widetilde{\phi}_\pm$. For the first example, the solution does not lead to a convex function, because this would contradict the required condition (C1) on the parts $\mathcal{X}_\pm$. Coming back to the direct problem, the "ϕ-like-entropy" defined with $\widetilde{\phi}$ is no more concave (indeed, it is no more an entropy in the sense of definition 1), so that the maximum ϕ-entropy problem is no more concave: one cannot guarantee the uniqueness and even the existence of a maximum so that there is no guarantee that the arcsine distribution would be a maximizer. Indeed, equation (6) coming from the Euler-Lagrange equation (see paragraph previous to prop. 1), one can just conclude that the arcsine is a critical point (either extremal, or inflection point) of the identified ϕ-like-entropy.

In section 2 and 3 we established general entropies with a given maximizer. In what follows, we will complete the information theoretical setting by introducing generalized escort distributions, generalized moments, and generalized Fisher information associated to the same entropic functional. We will then explore some of their relationships.

4. ϕ-escort distribution, (ϕ, α)-moments, (ϕ, β)-Fisher informations, generalized Cramér-Rao inequalities

In this section, we begin by introducing the above-mentioned informational quantities. We will then show that generalizations of the celebrated Cramér-Rao inequalities hold and link the generalized moments and Fisher information. Furthermore, the lower bound of the inequalities are saturated precisely by maximal ϕ-entropy distributions.

Escort distributions have been introduced as an operational tool in the context of multifractals [93,94], with interesting connections with the standard thermodynamics [95] and with source coding [26,27]. In our context, we also define (generalized) escort distributions associated with a particular convex function ϕ, and show how they pop up naturally. It is then possible to define generalized moments with respect to these escort distributions.

Definition 4 (ϕ-escort)**.** *Let $\phi : \mathcal{X} \times \mathcal{Y} \mapsto \mathbb{R}$ such that for any $x \in \mathcal{X} \subseteq \mathbb{R}^d$ function $\phi(x, \cdot)$ is a strictly convex twice differentiable function defined on the closed convex set $\mathcal{Y} \subseteq \mathbb{R}_+$. Then, if f is a probability distribution defined with respect to a general measure μ on a set $\mathcal{X}$ such that $f(\mathcal{X}) \subseteq \mathcal{Y}$, and such that*

$$C_\phi[f] = \int_{\mathcal{X}} \frac{\mathrm{d}\mu(x)}{\frac{\partial^2 \phi}{\partial y^2}(x, f(x))} < +\infty \tag{16}$$

we define by

$$E_{\phi,f}(x) = \frac{1}{C_\phi[f] \, \dfrac{\partial^2 \phi}{\partial y^2}(x, f(x))} \tag{17}$$

the ϕ-escort density with respect to measure μ, associated to density f.

Note that from the strict convexity of ϕ with respect to its second argument, this probability density is well defined and is strictly positive. Moreover, coming back to the previous examples, one can see that:

Example 1. *In the context of the Shannon entropy, entropy for which the Gaussian is the maximal entropy law for the second order moment constraint, $\phi(x, y) = \phi(y) = y \log y$, the ϕ-escort density associated to f restricts to density f itself.*

Example 2. *In the Rényi-Tsallis context, entropy for which the q-Gaussian is the maximal entropy law for the second order moment constraint $\phi(x, y) = \phi(y) = \frac{y^q - y}{q-1}$, and $E_{\phi,f} \propto f^{2-q}$ which recovers the escort distributions used in the Rényi-Tsallis context up to a duality transformation [95].*

Example 3. *For the entropy that is maximal for the arcsine distribution under the second order moment constraint, $\phi(x, y) = \phi(y) = \frac{1}{y}$, and $E_{\phi,f} \propto f^3$ which is nothing more than an escort distributions used in the Rényi-Tsallis context. Indeed, although the arcsine distribution does not fall in the q-Gaussian family, its form is very similar to a q-Gaussian distribution (with $q = -1$) where the "scaling" parameter would not be related to the exponent q. It is thus not surprising to recover an escort distribution associated to this family.*

Definition 5 ((α, ϕ)-moments)**.** *Under the assumptions of definition 4, with $\mathcal{X}$ equipped with a norm $\| \cdot \|_{\mathcal{X}}$, we define by*

$$M_{\alpha,\phi}[f; X] = \int_{\mathcal{X}} \|x\|_{\mathcal{X}}^{\alpha} \, E_{\phi,f}(x) \, \mathrm{d}\mu(x) \tag{18}$$

if this quantity exists, as the (α, ϕ)-moment of a random variable X associated to distribution f.

For our three examples, we have:

Example 1. *In the context of the Shannon entropy, the (α, ϕ)-moments are the usual moments of $\|X\|_{\mathcal{X}}^{\alpha}$.*

Example 2. *In the Rényi-Tsallis context the generalized moments introduced in [55,96] are recovered.*

Example 3. *For $\phi(x, y) = \phi(y) = \frac{1}{y}$ one also naturally finds generalized moments with the same form as those introduced in [55,96] (see the items related to the escort distributions).*

The Fisher information's importance is well known in estimation theory: the estimation error of a parameter is bounded by the inverse of the Fisher information associated with this distribution [34,60]. The Fisher information is also used as a method of inference and understanding in statistical physics and biology, as promoted by Frieden [61] and has been generalized in the Rényi-Tsallis context in a series of papers [75,78,80–83,97,98]. In what follows, we generalize these definitions a step further in our ϕ-entropy context.

Definition 6 (Nonparametric (β,ϕ)-Fisher information)**.** *With the same assumption as in definition 4, denoting by $\|\cdot\|_{\mathcal{X}*}$ the dual norm, for any differentiable density f, we define the quantity*

$$I_{\beta,\phi}[f] = \int_{\mathcal{X}} \left\| \frac{\nabla_x f(x)}{E_{\phi,f}(x)} \right\|_{\mathcal{X}*}^{\beta} E_{\phi,f}(x)\,\mathrm{d}\mu(x) \tag{19}$$

if this quantity exists, as the nonparametric (β,ϕ)-Fisher information of f.

Note that when ϕ is state-independent, $\phi(x,y) = \phi(y)$, as for the usual Fisher information, this quantity is shift-invariant, i.e., for $g(x) = f(x - x_0)$ one have $I_{\beta,\phi}[g] = I_{\beta,\phi}[f]$. This property is unfortunately lost in the state-dependent context. Furthermore, whereas the Fisher information have scaling properties $I[a^{-d} f(\cdot/a)] = I[f]/a^2$, this is lost for $I_{\beta,\phi}$, except when ϕ'' is a power (which corresponds either to the Shannon or Rényi-Tsallis entropy).

Definition 7 (Parametric (β,ϕ)-Fisher information)**.** *Let us consider the same assumptions as in definition 4, and a density f parameterized by $\theta \in \Theta \subseteq \mathbb{R}^m$ where set Θ is equipped with a norm $\|\cdot\|_\Theta$ and with the corresponding dual norm denoted $\|\cdot\|_{\Theta*}$. Assume that f is differentiable with respect to θ. We define by*

$$I_{\beta,\phi}[f;\theta] = \int_{\mathcal{X}} \left\| \frac{\nabla_\theta f(x)}{E_{\phi,f}(x)} \right\|_{\Theta*}^{\beta} E_{\phi,f}(x)\,\mathrm{d}\mu(x) \tag{20}$$

as the parametric (β,ϕ)-Fisher information of f.

Note that, as for the usual Fisher information, when the norms on $\mathcal{X}$ and on Θ are the same, the nonparametric and parametric information coincide when θ is a location parameter. For our three examples, we have:

Example 1. *In the Shannon entropy context, when the norm is the Euclidean norm and $\beta = 2$, the nonparametric and parametric informations (β,ϕ)-Fisher give the usual nonparametric and parametric Fisher informations respectively.*

Example 2. *Similarly, in the Rényi-Tsallis context, the generalizations proposed in [81–83] are recovered.*

Example 3. *For $\phi(x,y) = \phi(y) = \frac{1}{y}$ one also naturally find the generalizations proposed in [81–83] (see the items related to the escort distributions).*

We have now the quantities that allow to generalize the Cramér-Rao inequalities as follows.

Proposition 3 (Nonparametric (α,ϕ)-Cramér-Rao inequality)**.** *Assume that a differentiable probability density function with respect to a measure μ, defined on a domain $\mathcal{X}$, admits an (α,ϕ)-moment and a (α^*,ϕ)-Fisher information with $\alpha \geq 1$ and α^* Holder-conjugated $\frac{1}{\alpha} + \frac{1}{\alpha^*} = 1$,* *and that $xf(x)$ vanishes at the boundary of $\mathcal{X}$. Thus, density f satisfies the (α,ϕ) extended Cramér-Rao inequality*

$$M_{\alpha,\phi}[f;X]^{\frac{1}{\alpha}}\, I_{\alpha^*,\phi}[f]^{\frac{1}{\alpha^*}} \geq d \tag{21}$$

When ϕ is state independent, $\phi(x,y)=\phi(y)$, the equality occurs when f is the maximal ϕ entropy distribution subject to the moment constraint $T(x)=\|x\|_{\mathcal{X}}^{\alpha}$.

Proof. The approach follows [83], starting from the differentiable probability density f (derivative denoted $\nabla_x f$), since $xf(x)$ vanishes in the boundaries of X from the divergence theorem one has

$$0=\int_{\mathcal{X}}\nabla_x^t(xf(x))\,\mathrm{d}\mu(x)=\int_{\mathcal{X}}(\nabla_x^t\,x)f(x)\,\mathrm{d}\mu(x)+\int_{\mathcal{X}}x^t(\nabla_x f(x))\,\mathrm{d}\mu(x)$$

Now, for the first term, we use the facts that $\nabla_x^t x = d$ and that f is a density to achieve

$$d=-\int_{\mathcal{X}}x^t\,\frac{\nabla_x f(x)}{g(x)}\,g(x)\,\mathrm{d}\mu(x)$$

for any function g non-zero on $\mathcal{X}$. Now, noting that $d>0$, we obtain from [83, Lemma 2]

$$\begin{aligned} d &= \left|\int_{\mathcal{X}}x^t\left(\frac{\nabla_x f(x)}{g(x)}\right)g(x)\,\mathrm{d}\mu(x)\right| \\ &\leq \left(\int_{\mathcal{X}}\|x\|_{\mathcal{X}}^{\alpha}\,g(x)\,\mathrm{d}\mu(x)\right)^{\frac{1}{\alpha}}\left(\int_{\mathcal{X}}\left\|\frac{\nabla_x f(x)}{g(x)}\right\|_{\mathcal{X}^*}^{\alpha^*}g(x)\,\mathrm{d}\mu(x)\right)^{\frac{1}{\alpha^*}} \end{aligned}$$

The proof ends by choosing $g=E_{\phi,f}$ the ϕ-escort density associated to density f. Note now that, again from [83, Lemma 2] the equality is obtained when

$$\nabla_x f(x)\,\frac{\partial^2\phi}{\partial y^2}(x,f(x))=\lambda_1\nabla_x\|x\|_{\mathcal{X}}^{\alpha}$$

where λ_1 is a negative constant. Consider now the case where $\phi(x,y)=\phi(y)$ is state-independent. Thus, $\nabla_x f(x)\,\frac{\partial^2\phi}{\partial y^2}(x,f(x))=\nabla_x\phi'(f(x))$, that gives

$$\phi'(f(x))=\lambda_0+\lambda_1\|x\|_{\mathcal{X}}^{\alpha}$$

This last equation has precisely the form eq. (6) of proposition 1. □

An obvious consequence of the proposition is that the probability density that minimizes the (α^*,ϕ)-Fisher information subject to the moment constraint $T(x)=\|x\|_{\mathcal{X}}^{\alpha}$ coincides with the maximal ϕ-entropy distribution subject to the same moment constraint.

In the problem of estimation, the purpose is to determine a function $\hat{\theta}(x)$ in order to estimate an unknown parameter θ. In such a context, the Cramér-Rao inequality allows to lowerbound the variance of the estimator thanks to the parametric Fisher information. The idea is thus to extend this to bound any α order mean error using our generalized Fisher information.

Proposition 4 (Parametric (α,ϕ)-Cramér-Rao inequality)**.** *Let f be a probability density function with respect to a general measure μ defined over a set $\mathcal{X}$, where f is parameterized by a parameter $\theta\in\Theta\subseteq\mathbb{R}^m$, and satisfies the conditions of definition 7. Assume that both μ and $\mathcal{X}$ do not depend on θ, that f is a jointly measurable function of x and θ which is integrable with respect to x and absolutely continuous with respect to θ and that the derivatives of f with respect* *to each component of θ are locally integrable. Thus, for any estimator $\widehat{\theta}(X)$ of θ that does not depend on θ, we have*

$$M_{\alpha,\phi}\Big[f;\widehat{\theta}(X)-\theta\Big]^{\frac{1}{\alpha}}\, I_{\alpha^*,\phi}[f;\theta]^{\frac{1}{\alpha^*}} \;\geq\; \big|\, m \,+\, \nabla_\theta^t\, b(\theta)\,\big| \tag{22}$$

where

$$b(\theta) = \mathbb{E}\Big[\widehat{\theta}(X)-\theta\Big] \tag{23}$$

is the bias of the estimator and α and α^* are Holder conjugated. When ϕ is state independent, $\phi(x,y)=\phi(y)$, the equality occurs when f is the maximal ϕ entropy distribution subject to the moment constraint $T(x)=\Big\|\widehat{\theta}(x)-\theta\Big\|_\Theta^\alpha$.

Proof. The proof follows again that of [83], and starts by evaluating the divergence of the bias. The regularity conditions in the statement of the theorem enable to interchange integration with respect to x and differentiation with respect to θ, so that

$$\nabla_\theta^t\, b(\theta) \;=\; \int_{\mathcal{X}} \Big(\nabla_\theta^t\, \widehat{\theta}(x) \;-\; \nabla_\theta^t\, \theta\Big) f(x)\, \mathrm{d}\mu(x) \;+\; \int_{\mathcal{X}} \Big(\widehat{\theta}(x)-\theta\Big)^t \nabla_\theta f(x)\, \mathrm{d}\mu(x)$$

Note then that $\nabla_\theta^t\,\theta \;=\; m$ and that $\widehat{\theta}$ being independent on θ, one has $\nabla_\theta^t\,\widehat{\theta}(x) \;=\; 0$. Thus, f being a probability density, the equality becomes

$$m+\nabla_\theta^t\, b(\theta) = \int_{\mathcal{X}} \Big(\widehat{\theta}(x)-\theta\Big)^t \frac{\nabla_\theta f(x)}{g(x)}\, g(x)\, \mathrm{d}\mu(x)$$

for any density g non-zero on $\mathcal{X}$. The proof ends with the very same steps that in proposition 4 using [83, Lemma 2]. □

For our three examples, this leads to what follows.

Example 1. *The usual parametric and nonparametric Cramér-Rao inequality are recovered in the usual Shannon context $\phi(x,y)=y\log y$, using the Euclidean norm and $\alpha=2$. The bound in the nonparametric context is saturated for the maximal entropy law, namely the Gaussian.*

Example 2. *In the Rényi-Tsallis context, the generalizations proposed in [81–83] are recovered and, again, when $\alpha=2$, the bound is saturated in the nonparametric context for the q-Gaussian, maximal entropy law under the second order moment constraint.*

Example 3. *For $\phi(x,y)=\phi(y)=\frac{1}{y}$, again, one finds inequalities with the same form as those of the generalizations proposed in [81–83] (see the items related to the escort distributions).*

Beyond the mathematical aspect of these relations, they may have great interest to assess an estimator when the usual variance/mean square error does not exist. Moreover, the escort distribution is also a way to emphasize some part of a distribution. For instance, in the Rényi-Tsallis context, one can see that in f^q either the tails or the head of the distribution is emphasized. Playing with q is a way to penalize either the tails, or the head of the distribution in the estimation process.

5. ϕ-heat equation and extended de Bruijn identity

An important relation connecting the Shannon entropy H, coming from the "information world", with the Fisher information I, living in the "estimation world", is given by the de Bruijn identity and is closely linked to the Gaussian distribution. Considering a noisy random variable $Y_t = X+\sqrt{t}N$ where N is a zero-mean d-dimensional standard

Gaussian random vector and X a d-dimensional random vector independent of N, and of support independent on parameter t, then

$$\frac{d}{dt} H[f_{Y_t}] = \frac{1}{2} I[f_{Y_t}]$$

where f_{Y_t} stands for the probability distribution of Y_t. This identity is a critical ingredient in proving the entropy power and Stam inequalities [34]. The starting point to establish the de Bruijn identity is the heat equation satisfied by the probability distribution f_{Y_t}, $\frac{\partial f}{\partial t} = \frac{1}{2}\Delta f$, where Δ stands for the Laplacian operator [99].

Let us consider probability distributions f parameterized by a parameter $\theta \in \Theta \subseteq \mathbb{R}^m$, satisfying what we will call *generalized ϕ-heat equation*,

$$\nabla_\theta f = K \operatorname{div}\Big(\|\nabla_x \phi'(f)\|_{\mathcal{X}^*}^{\beta-2} \nabla_x f \Big) \tag{24}$$

for some $K \in \mathbb{R}^m$, possibly dependent on θ but not on x, and where ϕ is a convex twice differentiable function defined over a set $\mathcal{X} \in \mathbb{R}_+$.

When θ is scalar, this equation is an instance of what is known as quasilinear parabolic equations [100, § 8.8] and arises in various physical problems.

Proposition 5 (Extended de Bruijn identity)**.** *Let f be a probability distribution with respect to a measure μ. Suppose that f is parameterized by a parameter $\theta \in \Theta \subseteq \mathbb{R}^m$, and is defined over a set $\mathcal{X} \subset \mathbb{R}^d$. Assume that both $\mathcal{X}$ and μ do not depend on θ, and that f satisfies the nonlinear ϕ-heat equation eq. (24) for a twice differentiable convex function ϕ. Assume that $\nabla_\theta \phi(f)$ is absolutely integrable and locally integrable with respect to θ, and that the function $\|\nabla_x \phi'(f)\|_{\mathcal{X}^*}^{\beta-2} \nabla_x \phi(f)$ vanishes at the boundary of $\mathcal{X}$. Thus, distribution f satisfies the extended de Bruijn identity, relating the ϕ-entropy of f and its nonparametric (β, ϕ)-Fisher information as follows,*

$$\nabla_\theta H_\phi[f] = K\, C_\phi^{1-\beta}\, I_{\beta,\phi}[f] \tag{25}$$

with C_ϕ is the normalization constant given eq. (16).

Proof. From the definition of the ϕ-entropy, the smoothness of the assumption enabling to use the Leibnitz' rule and differentiate under the integral,

$$\begin{aligned}
\nabla_\theta H_\phi[f] &= -\int_{\mathcal{X}} \phi'(f(x))\, \nabla_\theta f(x)\, \mathrm{d}\mu(x)\\
&= -K \int_{\mathcal{X}} \phi'(f(x))\, \operatorname{div}\Big(\|\nabla_x \phi'(f(x))\|_{\mathcal{X}^*}^{\beta-2} \nabla_x f(x) \Big)\, \mathrm{d}\mu(x)\\
&= -K \int_{\mathcal{X}} \operatorname{div}\Big(\phi'(f(x)) \|\nabla_x \phi'(f(x))\|_{\mathcal{X}^*}^{\beta-2} \nabla_x f(x) \Big)\, \mathrm{d}\mu(x)\\
&\quad + K \int_{\mathcal{X}} \nabla_x^t \phi'(f(x)) \|\nabla_x \phi'(f(x))\|_{\mathcal{X}^*}^{\beta-2} \nabla_x f(x)\, \mathrm{d}\mu(x)\\
&= -K \int_{\mathcal{X}} \operatorname{div}\Big(\|\nabla_x \phi'(f(x))\|_{\mathcal{X}^*}^{\beta-2} \nabla_x \phi(f(x)) \Big)\, \mathrm{d}\mu(x)\\
&\quad + K \int_{\mathcal{X}} \left(\phi''(f(x))\right)^{\beta-1} \|\nabla_x f(x)\|_{\mathcal{X}^*}^{\beta}\, \mathrm{d}\mu(x)
\end{aligned}$$

where the second line comes from the ϕ-heat equation and where the third line comes from the product derivation rule.

Now, from the divergence theorem, the first term of the right handside reduces to the integral of $\|\nabla_x \phi'(f)\|_{\mathcal{X}^*}^{\beta-2} \nabla_x \phi(f)$ on the boundary of $\mathcal{X}$, that vanishes from the assumption of the proposition, while the second term of the right handside related to C_ϕ and the (β, ϕ)-Fisher information from eqs. (16), (17) and definition 6. □

Coming back to the special examples we presented all along the paper:

Example 1. *In the Shannon entropy context, for $K = \frac{1}{2}$ and $\beta = 2$, the standard heat-equation is recovered and the usual de Bruijn identity is recovered.*

Example 2. *The case where $\phi(y) = y^q$ was intensively studied in [84] and the results of the paper are naturally recovered. In particular, the generalized ϕ-heat equation appears in anomalous diffusion in porous medium [84,100–103].*

Example 3. *For $\phi(x, y) = \phi(y) = \frac{1}{y}$, once again one find the same form for the generalized heat equation than in [84,101,102], and therefore the same form of the generalized de Bruijn identity of [84] (see the items related to the escort distributions).*

6. Concluding remarks

In this paper, we extended as far as possible the identities and inequalities which link the classical informational quantities – Shannon entropy, Fisher information, moments, ..., in the framework of the ϕ-entropies. Our first result concerns the inverse maximum entropy problem, starting with a probability distribution and constraints and searching for which entropy the distribution is the maximizer. If such a study was already tackled, it is extended here in a much more general context. We used general reference measures — not necessarily discrete or of Lebesgue. We also considered the case where the distribution and constraints do not share the same symmetries, which leads to state-dependent entropic functionals. Our second result is the generalization of the Cramér-Rao inequality in the same setting: to this end, a generalized Fisher information and generalized moments are introduced, both based on a convex function ϕ (and a so-called ϕ-escort distribution). The Cramér-Rao inequality is saturated precisely for the maximum ϕ-entropy distribution with the same moment constraints, linking all information quantities together. Finally, our third result is the statement of a generalized de Bruijn identity, linking the ϕ-entropy rate and the ϕ-Fisher information of a distribution satisfying an extended heat equation, called ϕ-heat equation.

Two important inequalities still miss. The first one is the entropy power inequality (EPI), which states that the entropy power (exponential of twice the entropy) of the sum of two continuous independent random variables is higher than the sum of the individual entropy powers [1]. The second one is the Stam inequality which lowerbounds the product of the entropy power and the Fisher information. For the former, despite many efforts, the literature on extended version only treat special cases. For instance, some extensions in the classical settings exist for discrete variables but are somewhat limited [104–106]. In the continuous framework, the EPI was also extended to the class of the Rényi entropy (log of a ϕ-entropy with $\phi(u) = u^\alpha$) [107]. Important properties that play a key role in the inequality is that the Rényi entropy is invariant to an affine transform of unit determinant and monotonic under convolution, properties that seem lost in the very general setting considered here. This fact leaves little room to extend the EPI in our general settings. Concerning the Stam inequality, at a first glance, the fact that the proof is based on the EPI seems to close any hope to extend it to the ϕ-entropy framework. However, it was remarkably extended to the Rényi entropy, based on the Gagliardo-Nirenberg inequality [78,80,81,108]. Nevertheless, a key property is that both the entropy power and the extended Fisher information have scaling properties, that are lost in the general setting of the ϕ-entropies. A possible way to overcome the (apparent) limits just evoked could be to mimic alternative proofs such those based on optimal transport [109]. This approach precisely drops off any use of Young or Sobolev-like

[1] In fact, there exist other equivalent versions which can be found e.g., in [34,69].

inequalities. As far as we feel, there is thus a little room for extensions in the settings of the paper. Both the extension of the EPI and the Stam inequality are left as perspectives.

Another perspective lies in the estimation of the generalized moments from data (or from estimates). Such a possibility would confer an operational role to our Cramér-Rao inequality, i.e., by computing the estimator's generalized moments and comparing them to the bound. A difficulty resides in the presence of the ϕ-escort distribution which forbids empirical or Monte-Carlo approaches. The escort distribution needs to be estimated. This problem seems not far from the estimation of entropies from data and plug-in approaches used in such problems can thus be considered, like kernel approaches [110–112], nearest neighbor approaches [112,113], or minimal spanning tree approaches [42]. Of course this perspective goes far beyond the scope of this paper.

Author Contributions: The authors contributed equally to this work. All authors have read and agreed to the published version of the manuscript.

Funding: This research was partially funded by a grant from the LabEx PERSYVAL-Lab (ANR-11-LABX-0025-01).

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

Appendix A. Inverse maximum entropy problem and associated inequalities: some examples

In this appendix, we will now derive in details several examples of the maximum entropy inverse problem. In each case, we provide the quantities and inequalities associated with the entropic functional ϕ, as derived in the text. In the sequel, for sake of simplicity, we restrict our examples to the univariate context $d = 1$.

Appendix A.1. Normal distribution and second-order moment

For a normal distribution and second order moment constraint

$$f_X(x) = \frac{1}{\sqrt{2\pi}\,\sigma} \exp\left(-\frac{x^2}{2\,\sigma^2}\right) \qquad \text{and} \qquad T_1(x) = x^2 \qquad \text{on} \qquad \mathcal{X} = \mathbb{R},$$

we begin by computing the inverse of $y = f_X(x)$, which yields $T_1(x) = x^2 = -\sigma^2 \ln\left(2\pi\sigma^2 y^2\right)$. Note that f_X^{-1} is multivalued, but $T_1\left(f_X^{-1}(\cdot)\right)$ is univalued. Injecting the expression of $T_1\left(f_X^{-1}(y)\right)$ into eq. (7) we obtain

$$\phi'(y) = \left(\lambda_0 - \sigma^2 \log(2\pi\sigma^2)\,\lambda_1\right) - 2\sigma^2\lambda_1 \log y \qquad \text{with} \quad \lambda_1 < 0$$

where the requirement $\lambda_1 < 0$ is necessary to satisfy condition (C1), condition (C2) being already satisfied because f_X and T_1 share the same symmetries. This gives, after a reparameterization of the λ_is,

$$\phi(y) = \alpha\, y \log(y) + \beta\, y + \gamma \qquad \text{with} \quad \alpha > 0.$$

The judicious choice $\alpha = 1$, $\beta = \gamma = 0$ leads to function

$$\phi(y) = y \log y$$

which gives nothing but the Shannon entropy, as expected,

$$H_\phi[f] = -\int_{\mathcal{X}} f(x) \ln f(x)\, \mathrm{d}\mu(x)$$

where $\mathcal{X}$ is now the support of f (overall, the obtained family of entropy is the Shannon one up to a scaling and a shift).

Now, $\phi''(y) \propto \frac{1}{y}$ leads to the escort distribution def. 4 as $E_{\phi,f} = f$ so that, as expected, the (α, ϕ) moments def. 5 are the usual moments of order α. When $\beta = 2$ and the usual Euclidean norm is considered, the (β, ϕ)-Fisher informations def. 6 & 7 are the usual Fisher informations and the usual Cramér-Rao inequalities prop. 3 & 4 are recovered for $\alpha = 2$. Finally, for $\beta = 2$, the usual Euclidean norm, the ϕ-heat equation eq. (24) turns to be the heat equation, satisfied by the Gaussian, so that the usual de Bruijn identity is naturally recovered from prop. 5.

Appendix A.2. q-Gaussian distribution and second-order moment

For q-Gaussian distribution, also known as Tsallis distributions, Student-t and Student-r [91,92], and a second order moment constraint, we have

$$f_X(x) = A_q\left(1 - (q-1)\frac{x^2}{\sigma^2}\right)_+^{\frac{1}{(q-1)}} \qquad \text{and} \qquad T_1(x) = x^2,$$

where $q > 0$, $q \neq 1$, $x_+ = \max(x, 0)$ and A_q is a normalization coefficient. The support of f_X is $\mathcal{X} = \mathbb{R}$ when $q < 1$ and $\mathcal{X} = \left(-\frac{\sigma}{\sqrt{q-1}}\,;\, \frac{\sigma}{\sqrt{q-1}}\right)$ when $q > 1$.

The inverse of $y = f_X(x)$ gives $T_1(x) = x^2 = \frac{\sigma^2}{q-1}\left(1 - \left(\frac{y}{A_q}\right)^{q-1}\right)$. Note that, again, f_X^{-1} is multivalued, but $T_1\left(f_X^{-1}(\cdot)\right)$ is univalued. Injecting the expression of $T_1\left(f_X^{-1}(y)\right)$ into eq. (7) we get

$$\phi'(y) = \left(\lambda_0 + \frac{\lambda_1\sigma^2}{q-1}\right) - \frac{\lambda_1\sigma^2}{(q-1)\,A_q^{q-1}}\,y^{q-1} \qquad \text{with} \quad \lambda_1 < 0$$

where the requirement $\lambda_1 < 0$ is necessary to satisfy condition (C1), condition (C2) being satisfied since f_X and T_1 share the same symmetries. This gives, after a reparameterization of the λ_is,

$$\phi(y) = \alpha\,\frac{y^q - y}{q-1} + \beta y + \gamma \qquad \text{with} \quad \alpha > 0.$$

Note that the inverse of f_X is defined over $\left(0\,;\, A_q\right)$ but, without contradiction, the domain of definition of the entropic functional can be extended to $\mathbb{R}_+$.

Then, a judicious choice of parameters is $\alpha = 1, \beta = \gamma = 0$ that yields

$$\phi(y) = \frac{y^q - y}{q-1}.$$

and an associated entropy is then

$$H_\phi[f] = \frac{1}{1-q}\left(\int_{\mathcal{X}} f(x)^q \,\mathrm{d}\mu(x) - 1\right) :$$

where $\mathcal{X}$ is now the support of f. This entropy is nothing but the Havrda-Charvát-Tsallis entropy [12,14,17,91] (overall, we obtain this entropy up to a scaling and a shift).

Then, $\phi''(y) = q y^{q-2}$, so that, from def. 4, and then from def. 5, def. 6 & 7 respectively, we obtain $M_{\phi,\alpha}[f]$ and $I_{\phi,\alpha}[f]$ as respectively the q-moment of order α and the (q, β)-Fisher information defined previously in [78–83] (with the symmetric q index given here by $2 - q$). The extended Cramér-Rao inequality proved in [78,82,83] is then recovered from prop. 3 & 4, and the generalized de Bruijn identity of [84] is also recovered from eq. (24) & prop. 5.

Note that when $q \to 1$: f_X tends to the Gaussian distribution. It appears that H_ϕ tends to the Shannon entropy, $I_{\phi,2}$ to the usual Fisher information and $M_{\phi,\alpha}$ to the usual moments (both considering the Euclidean norm): all the settings related to the Gaussian distribution are naturally recovered.

Appendix A.3. q-exponential distribution and first-order moment

The same entropy functional can readily be obtained for the so-called q-exponential and a first order moment constraint:

$$f_X(x) = B_q(1-(q-1)\beta x)_+^{\frac{1}{(q-1)}} \qquad \text{and} \qquad T_1(x) = x \qquad \text{on} \qquad \mathcal{X} = \mathbb{R}_+,$$

where B_q is a normalization coefficient. It suffices to follow the very same steps as above, leading again to the Havrda-Charvát-Tsallis entropy, the q-moments of order α and the (q,β)-Fisher information defined previously in [78–83] (with the symmetric q index given here by $2-q$) as for the q-Gaussian distribution and to the extended Cramér-Rao inequality proved in [82,83] as well.

Now when $q \to 1$: f_X tends to the exponential distribution, known to be of maximum Shannon entropy on $\mathbb{R}_+$ under the first order moment constraint [34]. Again H_ϕ tends to the Shannon entropy, $I_{\phi,2}$ to the usual Fisher information and $M_{\phi,\alpha}$ to the usual moments (both considering the Euclidean norm): all the settings related to the exponential distribution is naturally recovered.

Appendix A.4. The arcsine distribution

The arcsine distribution is a special case of the beta distribution with shaping parameter $\alpha = \beta = \frac{1}{2}$ and appears in various application, see e.g. [92]. We consider here the centered and scaled version of this distribution which writes

$$f_X(x) = \frac{1}{\sqrt{s^2 - \pi^2 x^2}} \qquad \text{on} \qquad \mathcal{X} = \left(-\frac{s}{\pi}\,;\,\frac{s}{\pi}\right)$$

where $s > 0$. The inverse distributions $f_{X,\pm}^{-1}$ on $\mathcal{X}_- = \left(-\frac{s}{\pi}\,;0\right)$ and $\mathcal{X}_+ = \left(0\,;\frac{s}{\pi}\right)$ are

$$f_{X,\pm}^{-1}(y) = \pm\frac{\sqrt{s^2y^2-1}}{\pi y}, \qquad y \ge \frac{1}{s}.$$

Let us now consider again either a second order moment as the constraint, or (partial) first order moment(s).

Appendix A.4.1. Second order moment

When the second order moment $T_1(x) = x^2$ is constrained, conditions(C2) is satisfied, so that, injecting the expression of $T_1\left(f_X^{-1}(y)\right)$ into eq. (7) one immediately obtains

$$\phi'(y) = \lambda_0 + \lambda_1\left(\frac{s^2}{\pi^2} - \frac{1}{\pi^2 y^2}\right) \qquad \text{with} \quad \lambda_1 > 0$$

where the requirement $\lambda_1 < 0$ is necessary to satisfy condition (C1). After a reparameterization of the λ_is, the family of entropy functionals is then

$$\phi(y) = \frac{\alpha}{y} + \beta y + \gamma \quad \text{with} \quad \alpha > 0$$

Although the inverse of the arcsine distribution exists for $y \le \frac{1}{s}$, the entropy functional can be defined over $\mathbb{R}_+^*$.

Note that this entropy can be viewed as Havrda-Charvát-Tsallis entropy for $q = -1$, so that all the generalizations (escort, moments, Cramér-Rao inequality, de Bruijn identity) set out appendix A.2 are recovered in the limit $q \to -1$.

Appendix A.4.2. (Partial) first-order moment(s)

Since the distribution has not the same variation as $T_1(x) = x$, condition (C1) cannot be satisfied. Therefore, either we turn out to consider the arcsine distribution as a critical point (extremal, inflection point) of a non concave "entropy", or as a maximum entropy when constraints are of the type

$$T_{\pm,1}(x) = x\,\mathbb{1}_{\mathcal{X}_\pm}(x).$$

Now, dealing respectively with the partial-moment constraints $T_{\pm,1}$ and with the uniform constraint T_1, we obtain from eq. (15) and eq. (14) respectively,

$$\phi'_\pm(y) = \lambda_0 + \lambda_{\pm,1} \frac{\sqrt{s^2 y^2 - 1}}{\pi y} \qquad \text{and} \qquad \widetilde{\phi}'_\pm(y) = \lambda_0 \pm \lambda_1 \frac{\sqrt{s^2 y^2 - 1}}{\pi y}$$

where the sign is absorbed in the factors $\lambda_{\pm,1}$ in the first case. Dealing with the partial moments, one must impose

$$\lambda_{\pm,1} > 0$$

to satisfy condition (C1). At the opposite, condition (C1) cannot be satisfied for the second case (one would have to impose $\pm\lambda_1 > 0$ on $\mathcal{X}_\pm$). After a reparameterization of the λ_is, one obtains the branches of the entropic functional under the form $\phi_\pm(y) = \phi_{\pm,\mathrm{u}}(sy)$ with $\phi_{\pm,\mathrm{u}}(u) = \alpha_\pm\left(\sqrt{u^2-1} + \arctan\left(\frac{1}{\sqrt{u^2-1}}\right)\right)\mathbb{1}_{(1\,;+\infty)}(u) + \beta\, u + \gamma_\pm$ and with $\alpha_\pm > 0$, and the branches for the non-convex case $\widetilde{\phi}_\pm(y) = \widetilde{\phi}_{\pm,\mathrm{u}}(sy)$ with $\widetilde{\phi}_{\pm,\mathrm{u}}(u) = \pm\alpha\left(\sqrt{u^2-1} + \arctan\left(\frac{1}{\sqrt{u^2-1}}\right)\right)\mathbb{1}_{(1\,;+\infty)}(u) + \beta\, u + \gamma_\pm$.

In this case, s appears as an additional parameter of this family of the ϕ-entropy.

In both cases, the entropic functionals are defined for $u > 1$ because of the domain where f_X is invertible. However, in the first case, one can extend the domain to $\mathbb{R}_+$ ensuring both the continuity of the entropic functional and its derivative at $u = 1$ (and thus everywhere), by vanishing the derivative of the entropic functional at $u = 1$, which imposes $\beta = 0$. This is also possible for the functionals $\widetilde{\phi}_{\pm,\mathrm{u}}$ setting condition $\beta = 0$. This leads respectively to

$$\phi_\pm(y) = \phi_{\pm,\mathrm{u}}(sy) \qquad \text{with}$$

$$\phi_{\pm,\mathrm{u}}(u) = \alpha_\pm\left(\sqrt{u^2-1} + \arctan\left(\frac{1}{\sqrt{u^2-1}}\right)\right)\mathbb{1}_{(1\,;+\infty)}(u) + \gamma_\pm, \qquad \alpha_\pm > 0$$

and the branches for the non-convex case

$$\widetilde{\phi}_\pm(y) = \widetilde{\phi}_{\pm,\mathrm{u}}(sy) \qquad \text{with}$$

$$\widetilde{\phi}_{\pm,\mathrm{u}}(u) = \pm\alpha\left(\sqrt{u^2-1} + \arctan\left(\frac{1}{\sqrt{u^2-1}}\right)\right)\mathbb{1}_{(1\,;+\infty)}(u) + \gamma_\pm.$$

Remarkably, in the first case, an univalued entropic functional can be obtain imposing both $\alpha_+ = \alpha_-$, $\gamma_+ = \gamma_-$. Looking more attentively to this choice, one observe that it corresponds to the one obtained for the moment constraint $T_1(x) = |x|$, which have the same symmetries as f_X.

The uniform function ϕ_u is represented figure A1 for $\alpha_\pm = 1$, $\gamma_\pm = 0$.

Appendix A.5. The logistic distribution

In this case, one can write the distribution under the form

$$f_X(x) = \frac{1 - \tanh^2\left(\frac{2x}{s}\right)}{s} \qquad \text{and} \qquad T_1(x) = x^2 \qquad \text{on} \qquad \mathcal{X} = \mathbb{R}.$$

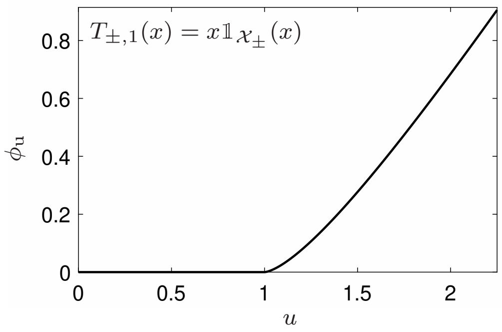


Figure A1. Univalued entropic functional ϕ_{u} derived from the arcsine distribution with partial constraints $T_{\pm,1}(x) = x\mathbb{1}_{\mathcal{X}_\pm}(x)$.

This distribution, which resembles the normal distribution but has heavier tails, has been used in various applications, see e.g., [92]. One can then check that over each interval

$$\mathcal{X}_\pm = \mathbb{R}_\pm$$

the inverse distribution writes

$$f_{X,\pm}^{-1}(y) = \pm\frac{s}{2}\operatorname{argtanh}\sqrt{1-sy}, \qquad y \in \left(0\,;\,\frac{1}{s}\right].$$

Let us now concentrate ourselves on a second order constraint, that respects the symmetry of the distribution, and on first order constraint(s) that does not respect the symmetry.

Appendix A.5.1. Second order moment constraint

In this case, injecting the expression of $T_1\left(f_X^{-1}(y)\right)$ into eq. (7), we immediately obtain

$$\phi'(y) = \lambda_0 + \frac{\lambda_1 s^2}{4}\left(\operatorname{argtanh}\sqrt{1-sy}\right)^2 \qquad \text{with} \qquad \lambda_1 < 0$$

where $\lambda_1 < 0$ is required to satisfy condition (C1). After a reparameterization, we thus obtain the family of entropy functionals $\phi(y) = \phi_{\mathrm{u}}(sy)$ with $\phi_{\mathrm{u}}(u) = -\alpha\left[u\left(\operatorname{argtanh}\sqrt{1-u}\right)^2 - 2\sqrt{1-u}\,\operatorname{argtanh}\sqrt{1-u} - \log u\right]\mathbb{1}_{(0\,;1]}(u) + \beta u + \gamma$ with $\alpha > 0$.

Here again, s is an additional parameter for this family of ϕ-entropies.

The entropy functional is defined for $u \le 1$ due to the domain f_X is invertible. To evaluate the ϕ-entropy for a given distribution f, one can play with parameter s so as to restrict, if possible, sf to be on $[0\,;1]$. But one can also extend the functional to $\mathbb{R}_+$ while remaining of class C^1 by vanishing the derivative at $u = 1$: this imposes $\beta = 0$ and leads to the entropy functional

$$\phi(y) = \phi_{\mathrm{u}}(sy) \qquad \text{with}$$

$$\phi_{\mathrm{u}}(u) = \gamma - \alpha\left[u\left(\operatorname{argtanh}\sqrt{1-u}\right)^2 - 2\sqrt{1-u}\,\operatorname{argtanh}\sqrt{1-u} - \log u\right]\mathbb{1}_{(0\,;1]}(u), \;\; \alpha > 0$$

depicted figure A2(a) for $\alpha = 1$, $\gamma = 0$.

Appendix A.5.2. (Partial) first-order moment(s) constraint(s)

Since f_X and $T(x) = x$ do no share the same symmetries, one cannot interpret the logistic distribution as a maximum entropy constraint by the first order moment. However, constraining the partial means over $\mathcal{X}_\pm = \mathbb{R}_\pm$ and using multiform entropies allows such an interpretation, while the alternative is to relax the concavity property of the entropy – but again, one would only be able to ensure that the distribution from which it comes is a critical point. To be more precise, one chooses

$$T_{\pm,1}(x) = x\,\mathbb{1}_{\mathcal{X}_\pm}(x) \qquad \text{or} \qquad T_1(x) = x$$

We thus obtain from eq. (15) and eq. (14) respectively, over each set $\mathcal{X}_\pm$, the branches

$$\phi_\pm'(y) = \lambda_0 + \frac{\lambda_{\pm,1} s}{2} \operatorname{argtanh} \sqrt{1-sy} \qquad \& \qquad \widetilde{\phi}_\pm'(y) = \lambda_0 \pm \frac{\lambda_1 s}{2} \operatorname{argtanh} \sqrt{1-sy}$$

where the sign is absorbed on $\lambda_\pm$ for the first case. Dealing with the partial moments, to satisfy condition (C1) one must impose

$$\lambda_\pm < 0.$$

At the opposite, condition (C1) cannot be satisfied for the second case (one would have to impose $\pm\lambda_1 < 0$ on $\mathcal{X}_\pm$). After a reparameterization of the λ_is, one obtains the branches of the entropic functional under the form $\phi_\pm(y) = \phi_{\pm,\mathrm{u}}(sy)$ with $\phi_{\pm,\mathrm{u}}(u) = -\alpha_\pm\big(u \operatorname{argtanh}\sqrt{1-u} - \sqrt{1-u}\big)\mathbb{1}_{(0\,;1]}(u) + \beta\, u + \gamma_\pm$ where $\alpha_\pm > 0$ and the branches for the non-convex case $\widetilde{\phi}_\pm(y) = \widetilde{\phi}_{\pm,\mathrm{u}}(sy)$ with $\widetilde{\phi}_{\pm,u}(u) = \pm\alpha\big(u \operatorname{argtanh}\sqrt{1-u} - \sqrt{1-u}\big)\mathbb{1}_{(0\,;1]}(u) + \beta\, u + \gamma_\pm$.

Once again, appears an additional parameter, s, for these families of entropies.

In both cases, even if the inverse of f_X restricts u to be lower than 1, one can either play with parameter s to allow to compute the ϕ-entropy of any distribution f, or to extend the entropic functionals to $\mathbb{R}_+$ by vanishing the derivative at $u = 1$. This impose $\beta = 0$ and thus the entropic functional,

$$\begin{aligned} &\phi_\pm(y) = \phi_{\pm,\mathrm{u}}(sy) \qquad \text{with} \\ &\phi_{\pm,u}(u) = \gamma_\pm - \alpha_\pm\Big(u \operatorname{argtanh}\sqrt{1-u} - \sqrt{1-u}\Big)\mathbb{1}_{(0\,;1]}(u), \qquad \alpha_\pm > 0 \end{aligned}$$

and the branches for the non-convex case

$$\begin{aligned} &\widetilde{\phi}_\pm(y) = \widetilde{\phi}_{\pm,\mathrm{u}}(sy) \qquad \text{with} \\ &\widetilde{\phi}_{\pm,u}(u) = \gamma_\pm \pm \alpha\Big(u \operatorname{argtanh}\sqrt{1-u} - \sqrt{1-u}\Big)\mathbb{1}_{(0\,;1]}(u) \end{aligned}$$

Remarkably, in the first case, an univalued entropic functional can be obtain imposing both $\alpha_+ = \alpha_-$, $\gamma_+ = \gamma_-$. Here also, such a choice is equivalent to consider the constraint $T_1(x) = |x|$, which allows to respect the symmetries of the distribution and to recover a classical ϕ-entropy.

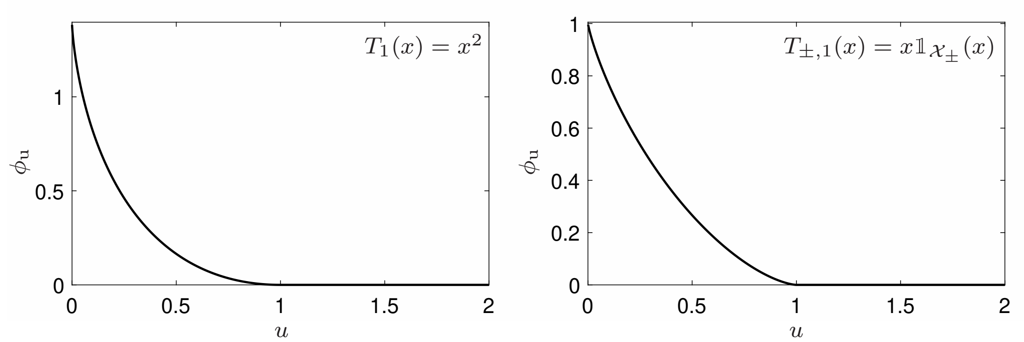
The uniform function ϕ_u is represented figure A2(b) for $\alpha_\pm = 1$, $\gamma_\pm = 0$.

Appendix A.6. The gamma distribution and (partial) p-order moment(s)

As a very special case, consider here the gamma distribution expressed as

$$f_X(x) = \frac{(\Gamma(q)x)^{q-1} \exp\left(-\frac{\Gamma(q)}{r}\, x\right)}{r^q} \qquad \text{on} \qquad \mathcal{X} = \mathbb{R}_+.$$

Parameter $q > 0$ is known as shape parameter of the law, while $\sigma = \frac{r}{\Gamma(q)} > 0$ is a scaling parameter. This distribution also appears in various applications, as described for instance in [114].



(a) (b)

Figure A2. Entropy functional ϕ_u derived from the logistic distribution: (a) with $T_1(x) = x^2$ and (b) with $T_{\pm,1}(x) = x\mathbb{1}_{\mathcal{X}_\pm}(x)$.

Let us concentrate on the case $q > 1$ for which the distribution is non-monotonous, unimodal, where the mode is located at $x = \frac{r\,(q-1)}{\Gamma(q)}$, and where $f_X(\mathbb{R}_+) = \left[0\,;\, \frac{(q-1)^{q-1}\,e^{1-q}}{r}\right]$.

Here again it cannot be a maximizer of a ϕ-entropy constraint subject to a moment of order $p > 0$ because x^p and f_X do not share the same symmetries. Therefore, we can again consider partial moments as constraints,

$$T_{k,1}(x) = x^p\,\mathbb{1}_{\mathcal{X}_k}(x), \qquad k \in \{0, -1\} \qquad \text{where}$$

$$\mathcal{X}_0 = \left[0\,;\, \frac{r\,(q-1)}{\Gamma(q)}\right) \qquad \text{and} \qquad \mathcal{X}_{-1} = \left[\frac{r\,(q-1)}{\Gamma(q)}\,;\, +\infty\right),$$

or interpret f_X as a critical point of a ϕ-like entropy by constraining the moment

$$T_1(x) = x^p \quad \text{over} \quad \mathcal{X} = \mathbb{R}_+$$

Inverting $y = f_X(x)$ leads to the equation

$$-\frac{\Gamma(q)\,x}{r\,(q-1)} \exp\left(-\frac{\Gamma(q)\,x}{r\,(q-1)}\right) = -\frac{(ry)^{\frac{1}{q-1}}}{q-1}$$

to be solved. As expected, this equation has two solutions. These solutions can be expressed via the multivalued Lambert-W function W defined by $z = \mathrm{W}(z)\exp(\mathrm{W}(z))$, i.e., W is the inverse function of $u \mapsto u\exp(u)$ [115, § 1], leading to the inverse functions

$$f_{X,k}^{-1}(y) = -\frac{r\,(q-1)}{\Gamma(q)}\,\mathrm{W}_k\left(-\frac{(ry)^{\frac{1}{q-1}}}{q-1}\right), \qquad ry \in \left[0\,;\, \left(\frac{q-1}{e}\right)^{q-1}\right],$$

where k denotes the branch of the Lambert-W function. $k = 0$ gives the principal branch and here it is related to the entropy part on $\mathcal{X}_0$, while $k = -1$ gives the secondary branch, related to $\mathcal{X}_{-1}$ here.

Applying (15) to obtain the branches of the functionals of the multiform entropy, one has thus to integrate the functions

$$\phi_k'(y) = \lambda_0 + \lambda_{k,1}\left[-\frac{r\,(q-1)}{\Gamma(q)}\,\mathrm{W}_k\left(-\frac{(ry)^{\frac{1}{q-1}}}{q-1}\right)\right]^p$$

where, to ensure the convexity of the ϕ_k,

$$(-1)^k \lambda_{k,1} > 0$$

The same approach allows to design $\widetilde{\phi}_k$, with a unique λ_1 instead of the $\lambda_{k,1}$s and without restriction on λ_1.

First, let us reparameterize the λ_is so as to absorb the factor $r/\Gamma(q)$ inside $\lambda_{k,1}$ so that one can write formally

$$\phi_k(y) = \phi_{k,\mathrm{u}}(ry) \qquad \text{with}$$

$$\phi_{k,\mathrm{u}}(u) = \gamma_k + \beta\, u + (-1)^k \alpha_k \int \left[(1-q)\, \mathrm{W}_k\left(-\frac{u^{\frac{1}{q-1}}}{q-1} \right) \right]^p du, \qquad \alpha_k \geq 0$$

Obtaining a closed-form expression for the integral term is not an easy task. But relation $z\,(1+\mathrm{W}_k(z))\ \mathrm{W}_k'(z) = \mathrm{W}_k(z)$ [115, Eq. 3.2] suggests that a way to make the integration is to search for it under the form of a series

$$\int \left[(1-q)\, \mathrm{W}_k\left(-\frac{u^{\frac{1}{q-1}}}{q-1} \right) \right]^p du \;=\; u \sum_{l \geq 0} a_l \left[(1-q)\, \mathrm{W}_k\left(-\frac{u^{\frac{1}{q-1}}}{q-1} \right) \right]^{l+p}$$

Hence, (i) we differentiate both side, (ii) we use the relation $z\,\mathrm{W}_k'(z) = \frac{\mathrm{W}_k(z)}{1+\mathrm{W}_k(z)}$ given above applied to $z = -\frac{u^{\frac{1}{q-1}}}{q-1}$, (iii) we thus multiply both side of the obtained equality by $1+\mathrm{W}_k\left(-\frac{u^{\frac{1}{q-1}}}{q-1}\right)$ and (iv) we equal the coefficients of the terms in $\left[(1-q)\,\mathrm{W}_k\left(-\frac{u^{\frac{1}{q-1}}}{q-1}\right)\right]^{p+l}$ to obtain a recursion on the a_l. The a_l can thus be evaluated explicitly and we recognize in the series the confluent hypergeometric (or Kummer's) function ${}_1F_1(1; p+q; \cdot)$ [116, Eq. 13.1.2] or [117, Eq. 9.210-1] (up to a factor and an additive constant), so that, we achieve to

$$\phi_{k,\mathrm{u}}(u) \;=\; \gamma_k + \beta\, u + (-1)^k\, \alpha_k\, u \left[(1-q)\ \mathrm{W}_k\left(-\frac{u^{\frac{1}{q-1}}}{q-1} \right) \right]^p \times$$

$$\left[1 - \frac{p}{p+q-1}\, {}_1F_1\left(1\,;\, p+q\,;\, (1-q)\, \mathrm{W}_k\left(-\frac{u^{\frac{1}{q-1}}}{q-1} \right) \right) \right] \mathbb{1}_{\left(0\,;\,\left(\frac{q-1}{e}\right)^{q-1}\right)}(u), \qquad \alpha_k > 0$$

One can check that these functions are indeed the ones we search for. To this end, (i) one derives the previous expression, (ii) one notes that from $z\,\mathrm{W}_k'(z) = \frac{\mathrm{W}_k(z)}{1+\mathrm{W}_k(z)}$ [115, Eq. 3.2] we have $u\left[(1-q)\,\mathrm{W}_k\left(-\frac{u^{\frac{1}{q-1}}}{q-1}\right)\right]' = -\frac{\mathrm{W}_k\left(-\frac{u^{\frac{1}{q-1}}}{q-1}\right)}{1+\mathrm{W}_k\left(-\frac{u^{\frac{1}{q-1}}}{q-1}\right)}$, (iii) ones finally uses the relation $(p+q-1-z)\,{}_1F_1(1\,;\,p+q\,;\,z) + z\,{}_1F_1'(1\,;\,p+q\,;\,z) = (p+q-1)\,{}_1F_1(0\,;\,p+q\,;\,z)$ [116, 13.4.11] together with ${}_1F_1(0\,;\,b\,;\,z) = 1$ [116, 13.1.2].

Again, p, q, r are additional parameters for this family of entropies.

Then, from the domain of definition of the inverse of f_X, u is restricted to $\left(0\,;\,\left(\frac{q-1}{e}\right)^{q-1}\right)$, which can be compensated for by playing with parameter r. At the opposite, noting that $\mathrm{W}_k(-e^{-1}) = -1$, to extend the entropic functionals to C^1 functions on $\mathbb{R}_+$, one would have to impose $\beta + (-1)^k \alpha_k = 0$ to vanish the derivatives at $u = e^{1-a}$. This is impossible because from $\alpha_k > 0$ one cannot impose $\beta = \alpha_{-1} = -\alpha_0$. Moreover, even a convex extension relaxing the C^1 condition is impossible since we would have to impose $\beta + (-1)\alpha_k \le \beta$ to insure the increase of both the ϕ_ks on $\mathbb{R}_+$.

We can however choose the coefficients so as to impose special conditions at the boundary(ies) of the domain of definition. As an example, we can which to vanish the ϕ_k at $u = 0$ (e.g., to ensure the convergence of the integral of $\phi_{-1}(f)$, being $\mathcal{X}_{-1}$ unbounded). To this end, one can evaluate the values of the ϕ_k in the boundaries of the domain.

From [115, Eq. 3.1] we have $\mathrm{W}_0(0) = 0$ and from [116, Eq. 13.1.2] ${}_1F_1(1; p+q; 0) = 1$, so that

$$\phi_{0,\mathrm{u}}(0) = \gamma_0 \qquad \text{and} \qquad \phi_{0,\mathrm{u}}'(0) = \beta$$

Then, $\lim_{x \to 0^-} \mathrm{W}_{-1}(x) = -\infty$ (see [115, Fig. 1 or Eq. 4.18]) so that, (i) from the asymptotics [116, Eq. 13.1.4] of the confluent hypergeometric function for a large argument, (ii) using $\mathrm{W}(z) e^{\mathrm{W}(z)} = z$ for $z = -\frac{u^{\frac{1}{q-1}}}{q-1}$, we obtain

$$\phi_{-1,\mathrm{u}}(0) = \gamma_{-1} + p\,\Gamma(p+q-1)\,\alpha_{-1} \qquad \text{and} \qquad \lim_{u \to 0^-} \phi_{-1,\mathrm{u}}'(u) = -\infty$$

Finally, from $\mathrm{W}_k(-e^{-1}) = -1$ we immediately have

$$\phi_{k,\mathrm{u}}\left(\left(\frac{q-1}{e}\right)^{q-1}\right) = \gamma_k + \left(\frac{q-1}{e}\right)^{q-1}\left(\beta + (-1)^k \alpha_k\,(q-1)^p \left[1 - \frac{p}{p+q-1}\,{}_1F_1(1; p+q; q-1)\right]\right)$$

and

$$\phi_{k,\mathrm{u}}'\left(\left(\frac{q-1}{e}\right)^{q-1}\right) = \beta + (-1)^k\,\alpha_k\,(q-1)^p$$

Interestingly, at the $q \to 1^+$, the Gamma distribution reduces to the exponential law. It is known that subject to the first order moment constraint, it is a maximum Shannon entropy distribution [34]. From the results above, one can notice that when $q \to 1^+$ one has

$$\lim_{q \to 1^+} \mathcal{X}_0 = \emptyset, \qquad \lim_{q \to 1^+} \mathcal{X}_{-1} = \mathbb{R}_+ = \mathcal{X}$$

Hence, in accordance

- The constraints degenerate to a single uniform constraint $T_1(x) = x^p$;
- In this limit, conditions **??** and (C2) are both satisfied.
- The entropic functional become state-independent (uniform), where only the branch ϕ_{-1} remains.

One can determine the limit entropic functional using [118, Th. 3.2] that states for any $t > 0$,

$$\left|\mathrm{W}_{-1}\left(-e^{-(t+1)}\right) + \log(t+1) + (t+1)\right| \le 1 - \log(e-1) = a$$

We apply this theorem to the positive real t given by

$$e^{-(t+1)} = \frac{u^{\frac{1}{q-1}}}{q-1} \qquad \text{i.e.,} \qquad t = -\frac{1}{q-1}\log u + \log(q-1) - 1$$

(see domain where u lives), which thus gives, from $q > 1$

$$\left| (1-q)\,\mathrm{W}_{-1}\left(-\frac{u^{\frac{1}{q-1}}}{q-1} \right) + \log u - (q-1)\log\Big((q-1)\log(q-1) - \log u \Big) \right| \le (q-1)a$$

As a consequence, the left handside tends uniformly to 0 when $q \to 1^+$ and one can see that $(q-1)\log\Big((q-1)\log(q-1) - \log u\Big)$ goes also uniformly to 0 as $q \to 1^+$, which allows to obtain

$$\lim_{q\to 1^+} (1-q)\,\mathrm{W}_k\left(-\frac{u^{\frac{1}{q-1}}}{q-1} \right) = -\log u$$

As a conclusion, from the continuity of ${}_1F_1$ both w.r.t., its parameters and its variable, we have

$$\lim_{q\to 1^+} \phi_{1,\mathrm{u}}(u) = \gamma_{-1} + \beta u - \alpha_{-1}\, u\, (-\log u)^p \Big(1 - {}_1F_1(1; p+1; -\log u) \Big)$$

$u \in (0\,;\,1)$ but the domain can be expanded to $\mathbb{R}_+$.

Finally, for $p = 1$, due to [116, 13.6.14] stating that ${}_1F_1(1;2;x) = \frac{e^x-1}{x}$, we obtain after simple algebra

$$\lim_{q\to 1^+, p=1} \phi_{-1,\mathrm{u}} = \alpha_{-1}\, u\, \log u \,+\, (\beta - \alpha_{-1})\, u \,+\, \gamma_{-1} + \alpha_{-1}$$

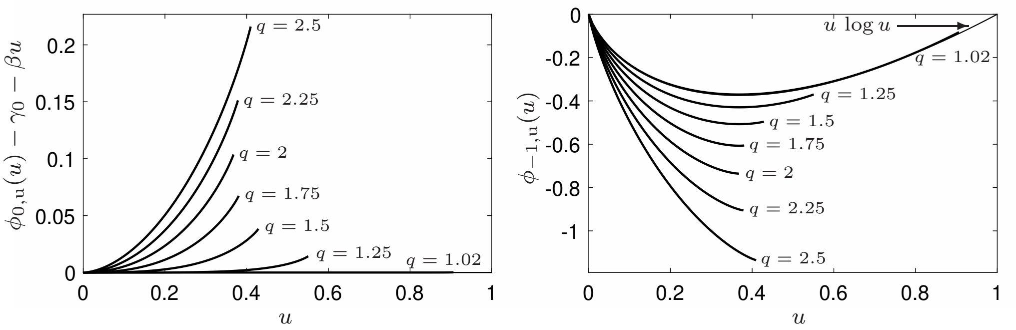
which is nothing but than the Shannon entropic functional, as expected

In passing, because W_0 is bounded on the considered domain, one has immediately

$$\lim_{q\to 1^+} \phi_{0,\mathrm{u}}(u) = \gamma_0 + \beta u$$

but remember that at the limit, this entropic branch disappear from the multiform entropy (i.e., the entropy becomes uniform).

The behavior of the multivalued function ϕ_{u} is represented figures A3 for $p = 1, q = 1.02, 1.25, 1.5, 1.75, 2, 2.25, 2.5$ respectively, and with the choices $\alpha_0 = \alpha_{-1} = \beta = 1$, $\gamma_0 = 0$, $\gamma_{-1} = -\Gamma(q)$. In (a) is represented $\phi_{0,\mathrm{u}} - \gamma_0 - \beta u$ so as to emphasis the behavior of the non lineal term; In (b) is depicted $\phi_{-1,\mathrm{u}}$ which, with the chosen parameters, tends to $u\,\log u$ (Shannon entropic functional) when $q \to 1^+$, together with this limit.



(a) (b)

Figure A3. Multiform entropy functional ϕ_{u} derived from the gamma distribution with the partial moment constraints $T_{k,1}(x) = x\mathbb{1}_{\mathcal{X}_k}(x)$ $(p = 1)$, $k \in \{0, -1\}$ for $q = 1.02, 1.25, 1.5, 1.75, 2, 2.25, 2.5$. (a): $\phi_{0,\mathrm{u}} - \gamma_0 - \beta u$ $(\alpha_0 = 1)$; (b): $\phi_{-1,\mathrm{u}}$ with $\alpha_{-1} = \beta = 1$, $\gamma_{-1} = -\Gamma(q)$, and Shannon entropic functional $u\,\log u$ (thin line).

References

1. von Neumann, J. Thermodynamik quantenmechanischer Gesamtheiten. *Nachrichten von der Gesellschaft der Wissenschaften zu Göttingen* **1927**, *1*, 273–291.
2. Shannon, C.E. A Mathematical Theory of Communication. *The Bell System Technical Journal* **1948**, *27*, 623–656. doi:10.1002/j.1538-7305.1948.tb00917.x.
3. Boltzmann, L. *Lectures on Gas Theory (translated by S. G. Brush)*; Dover: Leipzig, Germany, 1964.
4. Boltzmann, L. *vorlesungen über Gastheorie - I*; Verlag von Johann Ambrosius Barth: Leipzig, Germany, 1896.
5. Boltzmann, L. *vorlesungen über Gastheorie - II*; Verlag von Johann Ambrosius Barth: Leipzig, Germany, 1898.
6. Planck, M. *Eight Lectures on Theoretical Physics*; Columbia University Press: New-York, 2015.
7. W. D. Nieven, M. A., F.R.S. *The scientific papers of James Clerk Maxwell*; Vol. 2, Dover: New-York, 1952.
8. Jaynes, E.T. Gibbs vs Boltzmann Entropies. *American Journal of Physics* **1965**, *33*, 391–398. doi:10.1119/1.1971557.
9. Müller, I.; Müller, W.H. *Fundamentals of Thermodynamics and Applications. With Historical Annotations and Many Citations from Avogadro to Zermelo*; Springer: Berlin, 2009. doi:10.1007/978-3-540-74648-5.
10. Rényi, A. On measures of entropy and information. *in Proceeding of the 4th Berkeley Symposium on Mathematical Statistics and Probability* **1961**, *1*, 547–561.
11. Varma, R.S. Generalization of Rényi's Entropy of Order α. *Journal of Mathematical Sciences* **1966**, *1*, 34–48.
12. Havrda, J.; Charvát, F. Quantification Method of Classification Processes: Concept of Structural α-Entropy. *Kybernetika* **1967**, *3*, 30–35.
13. Csiszàr, I. Information-Type Measures of Difference of Probability Distributions and Indirect Observations. *Studia Scientiarum Mathematicarum Hungarica* **1967**, *2*, 299–318.
14. Daróczy, Z. Generalized Information Functions. *Information and Control* **1970**, *16*, 36–51.
15. Aczél, J.; Daróczy, Z. *On Measures of Information and Their Characterizations*; Academic Press: New-York, 1975.
16. Daróczy, Z.; Járai, A. On the measurable solution of a functional equation arising in information theory. *Acta Mathematica Academiae Scientiarum Hungaricae* **1979**, *34*, 105–116. doi:10.1007/bf01902599.
17. Tsallis, C. Possible Generalization of Boltzmann-Gibbs Statistics. *Journal of Statistical Physics* **1988**, *52*, 479–487. doi:10.1007/BF01016429.
18. Salicrú, M. Funciones de entropía asociada a medidas de Csiszár. *Qüestiió* **1987**, *11*, 3–12.
19. Salicrú, M.; Menéndez, M.L.; Morales, D.; Pardo, L. Asymptotic distribution of (h,ϕ)-entropies. *Communications in Statistics – Theory and Methods* **1993**, *22*, 2015–2031. doi:10.1080/03610929308831131.
20. Salicrú, M. Measures of information associated with Csiszár's divergences. *Kybernetica* **1994**, *30*, 563–573.
21. Liese, F.; Vajda, I. On Divergence and Informations in Statistics and Information Theory. *IEEE Transactions on Information Theory* **2006**, *52*, 4394–4412. doi:10.1109/TIT.2006.8811731.
22. Basseville, M. Divergence measures for statistical data processing – An annotated bibliography. *Signal Processing* **2013**, *93*, 621–633. doi:10.1016/j.sigpro.2012.09.003.
23. Panter, P.F.; Dite, W. Quantization distortion in pulse-count modulation with nonuniform spacing of levels. *Proceedings of the IRE* **1951**, *39*, 44–48. doi:10.1109/JRPROC.1951.230419.
24. Lloyd, S.P. Least Squares Quantization in PCM. *IEEE Transactions on Information Theory* **1982**, *28*, 129–137. doi:10.1109/TIT.1982.1056489.
25. Gersho, A.; Gray, R.M. *Vector quantization and signal compression*; Kluwer: Boston, 1992. doi:10.1007/978-1-4615-3626-0.
26. Campbell, L.L. A coding theorem and Rényi's entropy. *Information and Control* **1965**, *8*, 423–429. doi:10.1016/S0019-9958(65)90332-3.
27. Bercher, J.F. Source coding with escort distributions and Rényi entropy bounds. *Physics Letters A* **2009**, *373*, 3235–3238. doi:10.1016/j.physleta.2009.07.015.
28. Burbea, J.; Rao, C.R. On the Convexity of Some Divergence Measures Based on Entropy Functions. *IEEE Transactions on Information Theory* **1982**, *28*, 489–495. doi:10.1109/TIT.1982.1056497.
29. Menéndez, M.L.; Morales, D.; Pardo, L.; Salicrú, M. (h,Φ)-entropy differential metric. *Applications of Mathematics* **1997**, *42*, 81–98. doi:10.1023/A:1022214326758.
30. Pardo, L. *Statistical Inference Based on Divergence Measures*; Chapman & Hall: Boca Raton, FL, USA, 2006.
31. Jaynes, E.T. Information Theory and Statistical Mechanics. *Physical Review* **1957**, *106*, 620–630. doi:10.1103/PhysRev.106.620.
32. Kapur, J.N. *Maximum Entropy Model in Sciences and Engineering*; Wiley Eastern Limited: New-Dehli, 1989.
33. Arndt, C. *Information Measures: Information and Its Description in Sciences and Engeniering*; Springer Verlag: Berlin, 2001. doi:10.1007/978-3-642-56669-1.
34. Cover, T.M.; Thomas, J.A. *Elements of Information Theory*, 2nd ed.; John Wiley & Sons: Hoboken, New Jersey, 2006.
35. Gokhale, D.V. Maximum entropy characterizations of some distributions. In *A Modern Course on Statistical Distributions in Scientific Work*; G. P. Patil, S.K.; Ord, J.K., Eds.; Reidel: Dordrecht, Holland, 1975; Vol. III, pp. 299–304. doi:10.1007/978-94-010-1848-7.
36. Jaynes, E.T. Prior probabilities. *IEEE transactions on systems science and cybernetics* **1968**, *4*, 227–241. doi:10.1109/TSSC.1968.300117.
37. Csiszàr, I. Why Least Squares and Maximum Entropy? An Axiomatic Approach to Inference for Linear Inverse Problems. *The Annals of Statistics* **1991**, *19*, 2031–2066.

38. Frigyik, B.A.; Srivastava, S.; Gupta, M.R. Functional Bregman Divergence and Bayesian Estimation of Distributions. *IEEE Transactions on Infomation Theory* **2008**, *54*, 5130–5139. doi:10.1109/TIT.2008.929943.
39. Robert, C.P. *The Bayesian Choice. From Decision-Theoretic Foundations to Computational Implementation*, 2nd ed.; Springer: New-York, 2007.
40. Jaynes, E.T. On the rational of maximum-entropy methods. *Proceedings of the IEEE* **1982**, *70*, 939–952. doi:10.1109/PROC.1982.12425.
41. Jones, L.K.; Byrne, C.L. General Entropy Criteria for Inverse Problems, with Applications to Data Compression, Pattern Classification, and Cluster Analysis. *IEEE transactions on Information Theory* **1990**, *36*, 23–30. doi:10.1109/18.50370.
42. Hero III, A.O.; Ma, B.; Michel, O.J.J.; Gorman, J. Application of Entropic Spanning Graphs. *IEEE Signal Processing Magazine* **2002**, *19*, 85–95. doi:10.1109/MSP.2002.1028355.
43. Park, S.Y.; Bera, A.K. Maximum entropy autoregressive conditional heteroskedasticity model. *Journal of Econometrics* **2009**, *150*, 219–230. doi:10.1016/j.jeconom.2008.12.014.
44. Vasicek, O. A Test for Normality Based on Sample Entropy. *Journal of the Royal Statistical Society B* **1976**, *38*, 54–59. doi:10.1111/j.2517-6161.1976.tb01566.x.
45. Gokhale, D. On entropy-based goodness-of-fit tests. *Computational Statistics and Data Analysis* **1983**, *1*, 157–165. doi:10.1016/0167-9473(83)90087-7.
46. Song, K.S. Goodness-of-fit tests based on Kullback-Leibler discrimination information. *IEEE Transactions on Information Theory* **2002**, *48*, 1103–1117. doi:10.1109/18.995548.
47. Lequesne, J. A goodness-of-fit test of Student distributions based on Rényi entropy. AIP conference proceedings of the 34th international workshop on Bayesian Inference and Maximum Entropy Methods (MaxEnt'14); Djafari, A.; Barbaresco, F., Eds.; , 2014; Vol. 1641, pp. 487–494. doi:10.1063/1.4906014.
48. Lequesne, J. Tests statistiques basés sur la théorie de l'information, applications en biologie et en démographie. PhD thesis, Université de Caen Basse-Normandie, Caen, France, 2015.
49. Girardin, V.; Regnault, P. Escort distributions minimizing the Kullback-Leibler divergence for a large deviations principle and tests of entropy level. *Annals of the Institute of Statistical Mathematics* **2015**, *68*, 439–468. doi:10.1007/s10463-014-0501-x.
50. Kesavan, H.K.; Kapur, J.N. The Generalized Maximum Entropy Principle. *IEEE Transactions on Systems Man and Cybernetics* **1989**, *19*, 1042–1052.
51. Borwein, J.M.; Lewis, A.S. Duality Relationships for Entropy-Like Minimization Problems. *SIAM Journal on Control and Optimization* **1991**, *29*, 325–338. doi:10.1137/0329017.
52. Borwein, J.M.; Lewis, A.S. Convergence of best entropy estimates. *SIAM Journal of Optimization* **1991**, *1*, 191–205. doi:10.1137/0801014.
53. Borwein, J.M.; Lewis, A.S. Partially-finite programming in L_1 and the existence of maximum entropy estimates. *SIAM Journal of Optimization* **1993**, *3*, 248–267. doi:10.1137/0803012.
54. Mézard, M.; Montanari, A. *Information, Physics, and Computation*; Oxford University Press: New-York, 2009.
55. Tsallis, C.; Mendes, R.M.; Plastino, A.R. The role of constraints within generalized nonextensive statistics. *Physica A* **1998**, *261*, 534–554. doi:10.1016/S0378-4371(98)00437-3.
56. Tsallis, C. Nonextensive Statistics: theoretical, Experimental and Computational Evidences and Connections. *Brazilian Journal of Physics* **1999**, *29*, 1–35.
57. Tsallis, C. *Introduction to Nonextensive Statistical Mechanics – Approaching a Complex World*; Springer Verlag: New-York, 2009. doi:10.1007/978-0-387-85359-8.
58. Essex, C.; Schulzsky, C.; Franz, A.; Hoffmann, K.H. Tsallis and Rényi entropies in fractional diffusion and entropy production. *Physica A* **2000**, *284*, 299–308. doi:10.1016/S0378-4371(00)00174-6.
59. Parvan, A.S.; Biró, T.S. Extensive Rényi statistics from non-extensive entropy. *Physics Letters A* **2005**, *340*, 375–387. doi:10.1016/j.physleta.2005.04.036.
60. Kay, S.M. *Fundamentals for Statistical Signal Processing: Estimation Theory*; vol. 1, Prentice Hall: Upper Saddle River, NJ, 1993.
61. Frieden, B.R. *Science from Fisher Information: A Unification*; Cambridge University Press: Cambridge, UK, 2004.
62. Jeffrey. An Invariant Form for the Prior Probability in Estimation Problems. *Proceedings of the Royal Society A* **1946**, *186*, 453–461. doi:10.1098/rspa.1946.0056.
63. Vignat, C.; Bercher, J.F. Analysis of signals in the Fisher-Shannon information plane. *Physics Letters A* **2003**, *312*, 27–33. doi:10.1016/S0375-9601(03)00570-X.
64. Romera, E.; Angulo, J.C.; Dehesa, J.S. Fisher entropy and uncertainty like relationships in many-body systems. *Physical Review A* **1999**, *59*, 4064–4067. doi:10.1103/PhysRevA.59.4064.
65. Romera, E.; Sánchez-Moreno, P.; Dehesa, J.S. Uncertainty relation for Fisher information of D-dimensional single-particle systems with central potentials. *Journal of Mathematical Physics* **2006**, *47*, 103504. doi:10.1063/1.2357998.
66. Sánchez-Moreno, P.; González-Férez, R.; Dehesa, J.S. Improvement of the Heisenberg and Fisher-information-based uncertainty relations for D-dimensional potentials. *New Journal of Physics* **2006**, *8*, 330. doi:10.1088/1367-2630/8/12/330.
67. Toranzo, I.V.; Lopez-Rosa, S.; Esquivel, R.; Dehesa, J.S. Heisenberg-like and Fisher-information uncertainties relations for N-fermion d-dimensional systems. *Physical Review A* **2015**, p. on press.

68. Stam, A.J. Some Inequalities Satisfied by the Quantities of Information of Fisher and Shannon. *Information and Control* **1959**, *2*, 101–112. doi:10.1016/S0019-9958(59)90348-1.
69. Dembo, A.; Cover, T.M.; Thomas, J.A. Information Theoretic Inequalities. *IEEE Transactions on Information Theory* **1991**, *37*, 1501–1518. doi:10.1109/18.104312.
70. Guo, D.; Shamai, S.; Verdú, S. Mutual Information and Minimum Mean-Square Error in Gaussian Channels. *IEEE Transactions on Information Theory* **2005**, *51*, 1261–1282. doi:10.1109/TIT.2005.844072.
71. Folland, G.B.; Sitaram, A. The uncertainty principle: A mathematical survey. *Journal of Fourier Analysis and Applications* **1997**, *3*, 207–233. doi:10.1007/BF02649110.
72. Sen, K.D. *Statistical Complexity. Application in Electronic Structure*; Springer Verlag: New-York, 2011. doi:10.1007/978-90-481-3890-6.
73. Vajda, I. χ^α-divergence and generalized Fisher's information. Transactions of the 6th Prague Conference on Information Theory, Statistics, Decision Functions and Random Processes, 1973, pp. 873–886.
74. Boekee, D.E. An extension of the Fisher information measure. Topics in information theory. Proc. 2nd Colloquium on Information Theory; Csiszàr, I.; Elias., P., Eds.; Colloquia Mathematica Societatis János Bolyai and North Holland, Amsterdam: Keszthely, Hungary, 1977; Vol. 16, pp. 113–123.
75. Hammad, P. Mesure d'ordre α de l'information au sens de Fisher. *Revue de Statistique Appliquée* **1978**, *26*, 73–84.
76. Boekee, D.E.; Van der Lubbe, J.C.A. The *R*-Norm Information Measure. *Information and Control* **1980**, *45*, 136:155. doi:10.1016/S0019-9958(80)90292-2.
77. Lutwak, E.; Yang, D.; Zhang, G. Moment-Entropy Inequalities. *The Annals of Probability* **2004**, *32*, 757–774. doi:10.1214/aop/1079021463.
78. Lutwak, E.; Yang, D.; Zhang, G. Cramér-Rao and Moment-Entropy Inequalities for Rényi Entropy and Generalized Fisher Information. *IEEE Transactions on Information Theory* **2005**, *51*, 473–478. doi:10.1109/TIT.2004.840871.
79. Lutwak, E.; Yang, D.; Zhang, G. Moment-Entropy Inequalities for a Random Vector. *IEEE Transactions on Information Theory* **2007**, *53*, 1603–1607. doi:10.1109/TIT.2007.892780.
80. Lutwak, E.; Lv, S.; Yang, D.; Zhang, G. Extension of Fisher Information and Stam's Inequality. *IEEE Transactions on Information Theory* **2012**, *58*, 1319–1327. doi:10.1109/TIT.2011.2177563.
81. Bercher, J.F. On a (β, q)-generalized Fisher information and inequalities invoving q-Gaussian distributions. *Journal of Mathematical Physics* **2012**, *53*, 063303. doi:10.1063/1.4726197.
82. Bercher, J.F. On generalized Cramér-Rao inequalities, generalized Fisher information and characterizations of generalized q-Gaussian distributions. *Journal of Physics A* **2012**, *45*, 255303. doi:10.1088/1751-8113/45/25/255303.
83. Bercher, J.F. On multidimensional generalized Cramér-Rao inequalities, uncertainty relations and characterizations of generalized q-Gaussian distributions. *Journal of Physics A* **2013**, *46*, 095303. doi:10.1088/1751-8113/46/9/095303.
84. Bercher, J.F. Some properties of generalized Fisher information in the context of nonextensive thermostatistics. *Physica A* **2013**, *392*, 3140–3154. doi:10.1016/j.physa.2013.03.062.
85. Bregman, L.M. The relaxation method of finding the common point of convex sets and its application to the solution of problem in convex programming. *USSR Computational Mathematics and Mathematical Physics* **1967**, *7*, 200–217. doi:10.1016/0041-5553(67)90040-7.
86. Nielsen, F.; Nock, R. Generalizing Skew Jensen Divergences and Bregman Divergences With Comparative Convexity. *IEEE Signal Processing Letters* **2017**, *24*, 1123–1127. doi:10.1109/lsp.2017.2712195.
87. Ben-Tal, A.; Bornwein, J.M.; Teboulle, M. Spectral Estimation via Convex Programming. In *Systems and Management Science by Extremal Methods*; Phillips, F.Y.; Rousseau, J.J., Eds.; Springer, 1992; chapter 18, pp. 275–290. doi:10.1007/978-1-4615-3600-0_18.
88. Teboulle, M.; Vajda, I. Convergence of Best ϕ-Entropy Estimates. *IEEE Transactions on Infomation Theory* **1993**, *39*, 297–301. doi:10.1109/18.179378.
89. Girardin, V. Méthodes de réalisation de produit scalaire et de problème de moments avec maximisation d'entropie. *Studia Mathematica* **1997**, *124*, 199–213. doi:10.4064/sm-124-3-199-213.
90. Girardin, V. Relative Entropy and Spectral Constraints : Some Invariance Properties of the ARMA Class. *Journal of Time Series Analysis* **2007**, *28*, 844–866. doi:0.1111/j.1467-9892.2007.00535.x.
91. Costa, J.A.; Hero III, A.O.; Vignat, C. On Solutions to Multivariate Maximum α-Entropy Problems. 4th International Workshop on Energy Minimization Methods in Computer Vision and Pattern Recognition (EMMCVPR); Rangarajan, A.; Figueiredo, M.A.T.; Zerubia, J., Eds.; Springer Verlag: Lisbon, Portugal, 2003; Vol. 2683, *Lecture Notes in Computer Sciences*, pp. 211–226.
92. Johnson, N.L.; Kotz, S.; Balakrishnan, N. *Continuous Univariate Distributions*, 2nd ed.; Vol. 2, John Wiley & Sons: New-York, 1995.
93. Chhabra, A.; Jensen, R.V. Direct determination of the $f(\alpha)$ singularity spectrum. *Physical Review Letters* **1989**, *62*, 1327. doi:10.1103/PhysRevLett.62.1327.
94. Beck, C.; Schögl, F. *Thermodynamics of chaotic systems: an introduction*; Cambridge University Press: Cambridge, 1993. doi:10.1017/CBO9780511524585.
95. Naudts, J. *Generalized Thermostatistics*; Springer: London, 2011. doi:10.1007/978-0-85729-355-8.
96. Martínez, S.; Nicolás, F.; Pennini, F.; Plastino, A. Tsallis' entropy maximization procedure revisited. *Physica A* **2000**, *286*, 489–502. doi:10.1016/S0378-4371(00)00359-9.
97. Chimento, L.P.; Pennini, F.; Plastino, A. Naudts-like duality and the extreme Fisher information principle. *Physical Review E* **2000**, *62*, 7462–7465. doi:10.1103/PhysRevE.62.7462.

98. Casas, M.; Chimento, L.; Pennini, F.; Plastino, A.; Plastino, A.R. Fisher information in a Tsallis non-extensive environment. *Chaos, Solitons and Fractals* **2002**, *13*, 451–459. doi:10.1016/S0960-0779(01)00027-3.
99. Widder, D.V. *The Heat Equation*; Academic Press: New-York, 1975.
100. Roubíček, T. *Nonlinear partial differential equations with applications*; Birkhaäuser: Basel, Switzerland, 2005.
101. Tsallis, C.; Lenzi, E.K. Anomalous diffusion: nonlinear fractional Fokker-Planck equation. *Chemical Physics* **2002**, *284*, 341–347. doi:10.1016/S0301-0104(02)00557-8.
102. Vázquez, J.L. *Smoothing and Decay Estimates for Nonlinear Diffusion Equations – Equation of Porous Medium Type*; Oxford University Press: New-York, USA, 2006.
103. Gilding, B.H.; Kersner, R. *Travelling Waves in Nonlinear Diffusion-Convection Reaction*; Springer: Basel, Switzerland, 2004. doi:10.1007/978-3-0348-7964-4.
104. Harremoës, P.; Vignat, C. An Entropy Power Inequality for the Binomial Family. *Journal of Inequalities in Pure and Applied Mathematics* **2003**, *4*, 93.
105. Johnson, O.; Yu, Y. Monotonicity, Thinning, and Discrete Versions of the Entropy Power Inequality. *IEEE Transactions on Information Theory* **2010**, *56*, 5387–5395. doi:10.1109/tit.2010.2070570.
106. Haghighatshoar, S.; Abbe, E.; Telatar, I.E. A New Entropy Power Inequality for Integer-Valued Random Variables. *IEEE Transactions on Information Theory* **2014**, *60*, 3787–3796. doi:10.1109/tit.2014.2317181.
107. Bobkov, S.G.; Chistyakov, G.P. Entropy Power Inequality for the Rényi Entropy. *IEEE Transactions on Information Theory* **2015**, *61*, 708–714. doi:10.1109/TIT.2014.2383379.
108. Zozor, S.; Puertas-Centeno, D.; Dehesa, J.S. On Generalized Stam Inequalities and Fisher–Rényi Complexity Measures. *Entropy* **2017**, *19*, 493. doi:10.3390/e19090493.
109. Rioul, O. Yet Another Proof of the Entropy Power Inequality. *IEEE Transactions on Information Theory* **2017**, *63*, 3595–3599. doi:10.1109/tit.2017.2676093.
110. Rosenblatt, M. Remarks on Some Nonparametric Estimates of a Density Function. *The Annals of Mathematical Statistics* **1956**, *27*, 832–837.
111. Parzen, E. On Estimation of a Probability Density Function and Mode. *The Annals of Mathematical Statistics* **1962**, *33*, 1065–1076.
112. Beirlant, J.; Dudewicz, E.J.; Györfi, L.; van der Meulen, E.C. Nonparametric Entropy Estimation: An Overview. *International Journal of Mathematical and Statistical Sciences* **1997**, *6*, 17–39.
113. Leonenko, N.; Pronzato, L.; Savani, V. A Class of Rényi Information Estimators for Multidimensional Densities. *Annals of Statistics* **2008**, *36*, 2153–2182. doi:10.1214/07-AOS539.
114. Johnson, N.L.; Kotz, S.; Balakrishnan, N. *Continuous Univariate Distributions*, 2nd ed.; Vol. 1, John Wiley & Sons: New-York, 1995.
115. Corless, R.M.; Gonnet, G.H.; Hare, D.E.G.; Jeffrey, D.J.; Knuth, D.E. On the Lambert W Function. *Advances in Computational Mathematics* **1996**, *5*, 329–359. doi:10.1007/BF02124750.
116. Abramowitz, M.; Stegun, I.A. *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables*; 9th printing, Dover: New-York, 1970.
117. Gradshteyn, I.S.; Ryzhik, I.M. *Table of Integrals, Series, and Products*, 8th ed.; Academic Press: San Diego, 2015.
118. Alzahrani, F.; Salem, A. Sharp bounds for the Lambert W function. *Integral Transforms and Special Functions* **2018**, *29*, 971–978. doi:10.1080/10652469.2018.1528247.