ϕ-informational measures: some results in a generalized settings

by jfb & sz

Abstract

To be modified in the light of the new direction of the paper This paper focus on maximum entropy problems under moment constraints. Contrary to the usual problem of finding the maximizer of a given entropy, or of selecting constraints such that a given distribution is a maximizer of the considered entropy, we consider here the problem of the determination of an entropy such that a given distribution is its maximizer. Our goal is to adapt the entropy to its maximizer, with potential application in entropy-based goodness-of-fit tests for instance. It allows us to consider distributions outside the exponential family – to which the maximizers of the Shannon entropy belong -, and also to consider simple moment constraints, estimated from the observed sample. Out approach also yields entropic functionals that are function of both probability density and state, allowing us to include skew-symmetric or multimodal distributions in the setting. Finally, extended informational quantities are introduced, such that generalized moments and generalized Fisher informations. With thes extended quantities, we propose extended version of the Cramér-Rao inequality and of the de Bruijn identity, valid or saturated for the maximal entropy distribution corresponding to the generalized entropy previsouly studied.

1. Introduction

Since the pionner works of von Neuman [1], Shannon [2], Boltzmann, Maxwell, Planck and Gibbs [3–7] on the entropy as a tool for uncertainty or information measure, many investigations were devoted to the generalization of the so-called Shannon entropy and its associated measures [8–20]. 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 [21–23], codification penalizing long codewords where the Rényi entropy appears [24, 25] 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 [11, 17, 18, 26–28]. Such a general class (or more precisely the subclass of ϕ-entropies) traced back to the work of Burbea & Rao [26]. 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 [28]. Note that if a large amount of work deals with divergences, entropies occur as special cases when one takes a unform reference measure.

In the settings 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) [29–32]. In other words, it is the less informative law given the constraints of the system. In the Bayesian approach, dealing with the stochastic modelisation of a parameter, such a principle (or a minimum divergence principle) is often used to choose a prior distribution for the parameter [20, 33–36]. It also finds its counterpart in communication, clustering, pattern recognition, problems, among many others [30, 31, 37–39]. In statistics, some goodness-of-fit tests are based on entropic criteria derived from the same idea of constrained maximal entropic law [40–45]. In a large number of works using the maximum entropy principle, the entropy used is the Shannon entropy, although several extensions exist in the literature. However, if for some reason a generalized entropy is considered, the approach used in the Shannon case does not fundamentally change [46–49].

One can consider the inverse problem which consists in finding the moment constraints leading to the observed distribution as a maximal entropy distribution [46]. 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. jfbThis problem has also a physical interpretation, in the sense that the Shannon entropy described system in the thermodynamic limit, i.e., in equilibrium, whereas other entropies are often evoked to describe systems out of equilibrium [15, 50–54]. While the problem was considered mainly in the discrete Shannon settings by Kesavan & Kapur in [46], we will recall it in the general framework of the (h, ϕ)-entropies, 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 [55, 56], in Bayesian inference through the Jeffrey's prior [36, 57], but also for complex physical systems descriptions [56, 58–62].

Although coming from different worlds (information theory and communication, estimation, statistics, physics), these informational quantities are linked by well-known relations such the Cramér-Rao's inequality, the de Bruijn's identity, the Stam's inequality [32, 63–65]. These relationships have been proved very useful in various areas, for instance in communications [32, 63, 64], in estimation [55] or in physics [66, 67], 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 [26, 68–79].

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 [46], 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 saturated (or valid) dealing precisely for the generalized maximum entropy distribution. Finally, in section 5, we show that the extended quantities allows to obtain an extended de Bruijn identity, profided the distribution follows a non-linear heat equation. Some exemple of determination of ϕ-entropies solving the inverse maximum entropy problem are provided in a short series of appendix, showing in other that the usual quantities are recovered in the well known cases (Gausian distribution, Shannon entropy, Fisher information, variance).

In what follows we will define a series of generalized informational quantities of a probability density that is 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 variabes,...). Therefore, rigorously, all this quantities depend on the particular choice of this reference measure. However, for sake of simplicity we will omit to mention this dependence in the notation 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 divergences, and by Burbea and Rao in terms of entropies:

Definition 1 (ϕ-entropy [26]). 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 $H_{(h,\phi)}[f] = h\left(H_\phi[f]\right)$ where h is a nondecreasing function. The definition is extended by allowing ϕ to be concave, together with h nonincreasing [11, 17, 18, 27, 28]. If additionnaly h is concave, then the entropy functional $H_{(h,\phi)}[f]$ is concave.

In the following, since we are interested by the maximum entropy problem and because h is monotone, we can restrict our study to the ϕ-entropies. Additionnaly, we will assume that ϕ is *strictly* convex and *differentiable*.

A useful related quantity to these entropies is the Bregman divergence associated with convex function ϕ:

Definition 2 (Bregman divergence and functional Bregman divergence [20, 80]). With the same assumptions as in definition 1, the Bregman divergence associated with ϕ defined on a convex set $\mathcal{Y}$, is given by the function defined on $\mathcal{Y} \times \mathcal{Y}$,

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

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

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

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

Note that, more generally, the Bregman divergence is defined for multivariate convex functions, where the derivative is replaced by gradient operator [80]. Extensions for convex function of functions also exist, where the derivative is in the sense of Gâteau [81]. Such general extensions are not useful for our purposes, thus, we restrict to the above definition where $\mathcal{Y} \subseteq \mathbb{R}_+$.

2.1. Maximum entropy principle: the direct problem

Let us here 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

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

with

$$C_t = \{ f \geq 0 : \; \mathbb{E}[T_i(X)] = t_i, i = 0, \ldots, n \}, \tag{5}$$

where $T_0(x) = 1$ and $t_0 = 1$ (normalization constraint). 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, but this approach requires mild conditions [46, 47, 49, 82–84]. A sufficient condition relating f and ϕ so that f is the desired solution of the problem is then obtained, as recalled in the following proposition. Below, we prove the result without the use of the Lagrange technique.

Proposition 1 (Maximal ϕ-entropy solution [46])**.** *Suppose that there exists a probability distribution $f \in C_t$ satisfying*

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

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 (6) and consider any distribution $g \in C_t$. The functional Bregman divergence between f and g writes

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

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 (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 parametrized by the λ_i or, equivalently, by the moments t_i.

Note that the reciprocal is not necesarilly true, as shown for instance in [49]. However, the reciprocal is true when $\mathcal{X}$ is a compact [84] or for any $\mathcal{X}$ provided that ϕ is locally bounded on $\mathcal{X}$ [85].

2.2. Maximum entropy principle: the inverse problems

As stated in the introdution, two inverse problems can be considered started from a given distribution f. These problems were considered by Kesavan & Kapur in [46] 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. A solution can thus consist in identifying functions T_i and coefficients λ_i in order to satisfy eq. (6). Obviously, 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 fall in the exponential family [31, 32, 48]. 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_i. 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 parametrized by the λ_i. Here, required properties on ϕ will shape the domain the λ_i live in. In particular ϕ, must satisfy the following properties:

- 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 stricly increasing, its clear that solving eq. (6) requires the following two conditions:

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

(C2) $f(x)$ and $\displaystyle\sum_{i=1}^{n} \lambda_i T_i(x)$ must have the same level sets, $f(x_1) = f(x_2) \Leftrightarrow \displaystyle\sum_{i=0}^{n} \lambda_i T_i(x_1) = \sum_{i=0}^{n} \lambda_i T_i(x_2)$

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 3 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.

Eq. (6) provides an effective way to solve the inverse problem. However, there exist situations where there do not exist any set of λ_i such that conditions 3-(C2) are satisfied (e.g., $T_1(x) = x^2$ with f not even). In such a case, a way to go is to extent the definition of the ϕ-entropy, purpose of section 3.

2.3. Second inverse maximum entropy problem: some examples

To illustrate the previous subsection, let us analyze very briefly three examples: the very 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 constrainst. The Gaussian, q-Gaussian, and arcsine distributions will serve as a 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 The Gaussian distribution $f_X(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{x^2}{2\,\sigma^2}\right)$ defined over $\mathcal{X} = \mathbb{R}$, with the constraint $T_1(x) = x^2$ viewed as a maximum ϕ-entropy imposes that $\lambda_1 < 0$. Rapid calculus leads to the entropic functional, after a reparametrization of the λ_i's, of the form,

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

that is nothing more than the Shannon entropy, up to the scaling factor α, and a shift (except for $\gamma = -\beta$). One thus recovers the long outstanding fact that the Gaussian is the maximum Shannon entropy distribution with the second order moment constraint.

Example 2 The q-Gaussian distribution, also known as Tsallis distribution, $f_X(x) = C_q \left(1 - (q-1)\frac{x^2}{\sigma^2}\right)_+^{\frac{1}{1-q}}$, where $q > 0$, $q \neq 1$, $x_+ = \max(x, 0)$ and C_q is the normalization coefficient, defined over $\mathcal{X} = \mathbb{R}$, with the constraint $T_1(x) = x^2$ viewed as a maximum ϕ-entropy imposes also that $\lambda_1 < 0$. Rapid calculus leads to the entropic functional, after a reparametrization of the λ_i's, as,

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

that is nothing more than the Havrdat-Charvát-Tsallis entropy [10, 12, 15, 86], up to the scaling factor α, and a shift (except for $\gamma = -\beta$). One recover the also well known fact that the q-Gaussian is the maximum Shannon entropy distribution with the second order moment constraint [86]. In the limit case $q \to 1$, the distribution f_X tends to the Gaussian, whereas the Havrdat-Charvát-Tsallis entropy tends to the Shannon entropy.

Example 3 The arcsince distribution, $f_X(x) = \frac{1}{\pi\sqrt{2\,\sigma^2 - x^2}}$, defined over $\mathcal{X} = \left(-\sigma\sqrt{2}\,;\,\sigma\sqrt{2}\right)$, with the constraint $T_1(x) = x^2$ viewed as a maximum ϕ-entropy imposes now that $\lambda_1 > 0$. Short algebra leads to the entropic functional, after a reparametrization of the λ_i's,

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

This entropy is non usual 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 adress the limiation raised above, an idea is to allow the entropic functional to be depend also 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* versus y.

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

With these extended quantities, the direct problem becomes

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

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

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

Proof. The proof is the very same as that of Proposition 1, using the generalized functional Bregman divergence instead of the usual one. □

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

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

Thus, if there exists (at least) a set of λ_i such that condition 3 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

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

  where f_l^{-1} is the (possibly multivalued) inverse of f on $\mathcal{X}_l$.

In a conclusion, in the case where only condition 3 is satisfied, one can obtain an extended entropic functionnal 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.

Note however that it still may happen that there is no set of λ_i allowing to satisfy 3. In such an harder context, the problem remains solvable when then 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 exist on $\mathcal{X}_l$ a set of $\lambda_{l,i}$ such that 3 and (C2) holds. The solution still writes as in eq. (14), but where now n, the λ_i and the T_i are replaced by n_l, $\lambda_{l,i}$ and $T_{l,i}$ respectively.

Let us now come back to the arcsine example (example 3) of the previous section, when now we constraint the first order moment or partial first order moments.

Example 3-1 The arcsince distribution, $f_X(x) = \frac{1}{\pi \sqrt{2\,\sigma^2 - x^2}}$, defined over $\mathcal{X} = \left(-\sigma\sqrt{2}\,;\,\sigma\sqrt{2}\right)$, is bijective on each set $\mathcal{X}_- = (-\sigma\sqrt{2}\,;\,0)$ and $\mathcal{X}_+ = [0\,;\,\sigma\sqrt{2})$ that partitions $\mathcal{X}$. With the partial constraint $T_{\pm,1}(x) = x \mathbb{1}_{\mathcal{X}_\pm}(x)$, this distribution viewed as a maximum ϕ-entropy imposes that $\lambda_{\pm,1} > 0$ and the associated multiform entropic functional, after a reparametrization of the λ_i's, writes

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

(see appendix for more details).

Example 3-2 This arcsince distribution, now constraint uniformely by $T_1(x) = x$, viewed as an extremal ϕ-entropy imposes again that $\lambda_1 > 0$ and the associated multiform entropic functional, after a reparametrization of the λ_i's, is given by

$$\widetilde{\phi}_\pm(y) = \pm\alpha \left(\sqrt{2\pi^2\sigma^2y^2 - 1} + \arctan\left(\frac{1}{\sqrt{2\pi^2\sigma^2y^2 - 1}} \right) \right) \mathbb{1}_{(1\,;\,+\infty)}(\sqrt{2}\,\pi\sigma\,y) + \beta\,y + \gamma \quad \text{with} \quad \alpha > 0$$

(see appendix for more details). The entropic functional is no more convex.

At a first glance, the two solutions seem to be identical. In fact, they drastically differ. Indeed, let us insist on the fact that in the first case, the problem has two constraints whereas in the second case there is only one. The consequence is that the first solution is parametrized by 5 parameters $\beta, \gamma_\pm$ and, especially, $\alpha_\pm$, while only 3 parametrize the second solution β, γ and α. This difference is not insignificant: the second case cannot be viewed as a special case of the first one because $\alpha_\pm$ must be positive, which cannot be possible with only one parameter because $\pm\alpha$ rule the $\widetilde{\phi}_\pm$. The consequcence is that for the second example, the solution does not lead to a convex function, otherwise, it would have contradict the required condition on the parts $\mathcal{X}_\pm$.

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

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

In this section, after introducing the above mentioned informational quantities, we will show that generalizations of the celebrated Cramér-Rao inequalities hold. 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 [87, 88], with interesting connections with the standard thermodynamics [89] and with source coding [24, 25]. In our context, we also define (generalized) escort distributions associated with a particular ϕ-entropy, and show how they pop up naturally. It is then possible to define generalized moments with respect 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 $f(\mathcal{X}) \subseteq \mathcal{Y}$, such that

$$C_\phi[f] = \int_{\mathcal{X}} \frac{\mathrm{d}\mu(x)}{\frac{\partial^2 \phi}{\partial y^2}(x, f(x))} < +\infty \tag{15}$$

we define by

$$E_{\phi,f}(x) = \frac{1}{C_\phi[f] \, \dfrac{\partial^2 \phi}{\partial y^2}(x, f(x))} \tag{16}$$

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 [89].

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-distribution where the "scaling" would not be related to the exponent q. It is thus not suprising to recover an escort distribution associated to this family.

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

$$M_{\alpha,\phi}[f; X] = \int_{\mathcal{X}} \|x\|_{\mathcal{X}}^{\alpha} \, E_{\phi,f}(x) \, \mathrm{d}\mu(x) \tag{17}$$

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

Note that:

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 Rény-Tsallis context the generalized moments introduced in [50, 90] are recovered.

Example 3 For $\phi(x, y) = \phi(y) = \frac{1}{y}$ one also naturally find the generalized moments introduced in [50, 90] (see the items related to the escort distributions).

The importance of the Fisher information 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 [32, 55]. The Fisher information is also used as a method of inference and understanding in statistical physics and biology, as promoted by Frieden [56] and has been generalized in the Rényi-Tsallis context in a series of papers [70, 73, 75–78, 91, 92]. In what follows, we generalize these definitions a step further in our ϕ-entropy context by using the above defined ϕ-escort distribution.

Definition 6 (Nonparametric (β, ϕ)-Fisher information)**.** With the same assumption as in definition 5, denoting by $\|\cdot\|_{\mathcal{X}*}$ the dual norm, for any differentiable density f, we define the quantity

$$I_{\beta,\phi}[f] = \int_{\mathcal{X}} \left\| \frac{\nabla_x f(x)}{E_{\phi,f}(x)} \right\|_{\mathcal{X}*}^{\beta} E_{\phi,f}(x) \, \mathrm{d}\mu(x) \tag{18}$$

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.

Definition 7 (Parametric (β, ϕ)-Fisher information)**.** Let consider the same assumption as in definition 5, such that density f is parametrized by a parameter $\theta \in \Theta \subseteq \mathbb{R}^m$. The set Θ is equipped with a norm $\|\cdot\|_{\Theta}$ and the corresponding dual norm is denoted $\|\cdot\|_{\Theta*}$. Assume that f is differentiable with respect to θ. We define by

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

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

Note that, as for the usual Fisher information, when the norm on $\mathcal{X}$ and on Θ are the same, the nonparametric and parametric information coincide when θ is a location parameter. Note also that:

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 [76–78] are recovered.

Example 3 For $\phi(x, y) = \phi(y) = \frac{1}{y}$ one also naturally find the generalized moments introduced in [50, 90] (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 $\alpha \ge 1$ and α^* Holder-conjugated $\frac{1}{\alpha} + \frac{1}{\alpha^*} = 1$, and that $x f(x)$ vanishes at the boundary of $\mathcal{X}$. Thus, density f satisfies the (α,ϕ) extended Cramér-Rao inequality*

$$M_{\alpha,\phi}[f;X]^{\frac{1}{\alpha}}\, I_{\alpha^*,\phi}[f]^{\frac{1}{\alpha^*}} \;\ge\; d \tag{20}$$

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 [78], starting from the differentiable probability density f (derivative denoted $\nabla_x f$), since $x f(x)$ vanishes in the boundaries of X from the divergence theorem one has

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

Now, for the first term, we use the fact that $\nabla_x x = d$ 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 [78, Lemma 2]

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

The proof ends by choosing $g = E_{\phi,f}$ the ϕ-escort density associated to density f. Note now that, again from [78, 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 probabiliy 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 spirit is thus to extend such an inequality to bound any α order mean error thanks to 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 define over a set $\mathcal{X}$, where f is parametrized by a parameter $\theta \in \Theta \subseteq \mathbb{R}^m$ but μ does not depend on θ and satisfying the conditions of definition 7. Assume that domain $\mathcal{X}$ does not depend on θ neither, that f is a jointly measurable function of x and θ, is integrable with respect to x, is absolutely continuous with respect to θ and that the derivatives 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*

$$M_{\alpha,\phi}\Big[f; \widehat{\theta}(X) - \theta\Big]^{\frac{1}{\alpha}} \;\; I_{\alpha^*,\phi}[f;\theta]^{\frac{1}{\alpha^*}} \;\ge\; \left| \, m \,+\, \nabla_\theta^t\, b(\theta) \, \right| \tag{21}$$

where

$$b(\theta) = \mathbb{E}\left[\widehat{\theta}(X) - \theta \right] \tag{22}$$

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 $T(x) = \|\Theta(x) - \theta\|_{\Theta}^{\alpha}$.

Proof. The proof follows again that of [78], and start first 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, thus

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

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

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

for any density g non-zero on $\mathcal{X}$. The proof ends with the very same steps that in proposition 4 using [78, Lemma 2]. □

Note that:

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 [76–78] 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 $\phi(x,y) = \phi(y) = \frac{1}{y}$ one also naturally find the generalized moments introduced in [50, 90] (see the items related to the escort distributions).

Beyond the mathematical aspect of these relations, they may have great interest to asses for instance an estimator when the usual variance/mean square error does not exist. Moreover, the escort distribution is also a way to emphasis some part of a distribution. For instance, in the Rényi-Tsallis context, one can see that in f^q either the tails of the head of the distribution is emphasis. Playing with q is thus a way to penalize more either the tails, or the head of the distribution in the estimation (estimation using the escort and adequate Cramér-Rao bound to assess the estimator).

5. ϕ-heat equation and extended de Bruijn identity

An important relation connecting the Shannon entropy H, comming 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 independant 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 in the heart of the proof of the entropy power inequality, and then to a proof of the Stam's inequality [32]. The key point of the proof of this 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 [93].

Inspired by the work [79], we consider in the following, probability distributions f parametrized by a parameter $\theta \in \Theta \subseteq \mathbb{R}^m$, satisfying what we will call *generalized ϕ-heat equation*,

$$\nabla_\theta f = K \operatorname{div}\left(\|\nabla_x \phi'(f)\|_{\mathcal{X}^*}^{\beta-2} \nabla_x f \right) \tag{23}$$

for some $K \in \mathbb{R}^m$ (possibly dependent on θ) and where ϕ is a convex twice differentiable function defined over a set $\mathcal{X} \in \mathbb{R}_+$.

Proposition 5 (Extended de Bruijn identity)**.** *Let f be a probability distribution, parametrized by a parameter $\theta \in \Theta \subseteq \mathbb{R}^m$, defined over a set $\mathcal{X} \subset \mathbb{R}^d$ that do not depend on θ, and satisfying the nonlinear ϕ-heat equation eq.* (23) *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,*

$$\nabla_\theta H_\phi[f] = K C_\phi^{1-\beta}\, I_{\beta,\phi}[f] \tag{24}$$

with C_ϕ is the normalisation constant given eq. (15).

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

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

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. (15), (16) 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 [79] and the results of the paper are naturally recovered. In particular, the generalized ϕ-heat equation appears in anomalous diffusion in porous medium [79, 94, 95].

Example 3 For $\phi(x, y) = \phi(y) = \frac{1}{y}$ one also naturally find the generalized moments introduced in [50, 90] (see the items related to the escort distributions).

Note that various physical non linear diffusions equation are encompassed in the generalized ϕ-heat equation [95, 96].
Anomalous diffusion, NL FK: [94?]

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 main 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 done, it is extended here in a more general context, where densities are considered with respect to a measure not necessarily discrete or of Lebesgue, and especially when the distribution and the constraints do not share the same symmetries. To overcome the last drawback, we introduced a state-dependent entropic functional. 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 link with the associated ϕ-entropy resides in the fact that the inequality is saturated precisely for the maximum ϕ-entropy distribution constraints to the moment of the Cramér-Rao inequality. Finally, our third result is the generalized de Bruijn identity, linking the ϕ-entropy rate and the ϕ-Fisher information of a distribution that follows a extended heat equation, called ϕ-heat equation. Such nonlinear differential equations appear in various physical contexts. Moreover, dealing with usual distributions (Gaussian, q-Gaussian, exponential) and usual moments (mean, second order), the classical results are recovered as limit cases. As a matter of fact, dealing with the Shannon entropy, whatever the constraints considered, the maximum entropy distribution falls in the exponential family [31, 32, 48]. 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 thermodynamic limit), extended entropies allow a finer description of systems out of equilibrium [15, 52], exhibiting their importance.

In this panel, two important inequalities still miss. The first one is 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 of the sum of the individual entropy powers [1]. The second one is the Stam's 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 [97–99]. In the continuous framework, the EPI was also extended to the class of the Rényi entropies (log of a ϕ-entropy with $\phi(u) = u^\alpha$) [100]. Important properties that play a key role in the inequality is that the Rényi's entropies are invariant to an afine 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. About 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's entropies, base on the Gagliardo-Nirenberg inequality [73, 75, 76, 101]. 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 that based on optimal transport [102]. This approach precisely drops off any use of Young or Sobolev-like 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 Stam inequalities are left as a perspective.

Another perspective lies in the estimation of the generalized moments from data (or from estimates). Such a possibility would have the merit to give an operational role of the Cramér-Rao inequality, i.e., by estimating generalized moments and compare them to the bound. A difficulty resides in the presence of the ϕ-escort distribution. This forbids empirical or Monte-Carlo approaches, i.e., the escort distribution need clearly 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 [103–105], nearest neighbor approaches [105, 106], or minimal spanning tree approaches [38]. This perspective goes far beyond the scope of this paper.

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

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

[1] In fact, there exists other equivalent versions which can be found e.g., in [32, 64].

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 $y = f_X(x)$ where $x \in \mathbb{R}_+$ for instance, which gives

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

The judicious choice

$$\lambda_0 = 1 - \log(\sqrt{2\pi}\sigma) \qquad \text{and} \qquad \lambda_1 = -\frac{1}{2\,\sigma^2}$$

leads to function

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

that gives nothing more than the Shannon entropy as expected.

Now, $\phi''(y) \propto \frac{1}{y}$ leading to the escort distribution $E_{\phi,f} = f$ so that, as expected, the (α, ϕ) moments are the usual moments of order α. When $\beta = 2$ and the usual euclidean norm is considered, the (β, ϕ)-Fisher informations are the usual Fisher information and the usual Cramér-Rao inequalities are recovered for $\alpha = 2$. Finally, for $\beta = 2$, the usual euclidean norm, the ϕ-heat equation turns to be the heat equation, satisfied by the gaussian, so that the usual de Bruijn identity is naturally recovered.

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

For q-normal distribution, also known as Tsallis distributions, Student-t and -r, and a second order moment constraint,

$$f_X(x) = C_q\Big(1 - (q-1)\beta x^2\Big)_+^{\frac{1}{(q-1)}} \qquad \text{and} \qquad T_1(x) = x^2 \qquad \text{on} \qquad \mathcal{X} = \mathbb{R},$$

where $q > 0$, $x_+ = \max(x, 0)$ and C_q is a normalization coefficient, we get

$$\phi'(y) = \left(\lambda_0 + \frac{\lambda_1}{(q-1)\beta}\right) - \frac{\lambda_1\, y^{q-1}}{C_q^{q-1}(q-1)\beta}.$$

In this case, a judicious choice of parameters is

$$\lambda_0 = \frac{q\,C_q^{q-1} - 1}{q-1} \qquad \text{and} \qquad \lambda_1 = -q\,C_q^{q-1}\beta$$

that yields to

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

and an associated entropy is then

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

It is nothing but Havrdat-Charvát-Tsallis entropy [10, 12, 15, 86].

Then, $\phi''(y) = q y^{q-2}$: $M_{\phi,\alpha}[f]$ and $I_{\phi,\alpha}[f]$ are respectively the q-moment of order α and the (q, β)-Fisher information defined previously in [73–78] (with the symmetric q index given here by $2 - q$). The extended Cramér-Rao inequality proved in [73, 77, 78] is then recovered.

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

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

It suffices to follow the very same steps as above, leading again to the Havrdat-Charvát-Tsallis entropy, the q-moments of order α and the (q, β)-Fisher information defined previously in [73–78] (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 [77, 78] as well.

Now when $q \to 1$: f_X tends to the exponential distribution, known to be of maximum Shannon's entropy on $\mathbb{R}_+$ under the first order moment constraint. Again H_ϕ tends to the Shannon's entropy, $I_{\phi,2}$ to the usual Fisher's 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 logistic distribution

In this case,

$$f_X(x) = \frac{1 - \tanh^2\left(\frac{x}{2s}\right)}{4s} \qquad \text{and} \qquad T_1(x) = x^2 \qquad \text{on} \qquad \mathcal{X} = \mathbb{R}.$$

This distribution, which resembles the normal distribution but has heavier tails, has been used in many applications. One can then check that over each interval

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

the inverse distribution writes

$$f_{X,\pm}^{-1}(y) = \pm 2s \operatorname{argtanh} \sqrt{1 - 4sy}, \qquad y \in \left[0\,;\, \frac{1}{4s}\right]$$

We concentrate now on a second order constraint, that respect the symmetry of the distribution, and on first order constrain(s) that does not respect the symmetry.

Appendix A.4.1. Second order moment constraint

In this case, immediately

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

for $y \in \left[0\,;\, \frac{1}{4s}\right]$ and where the positive factors $\frac{1}{4s}$ and s are absorbed in λ_0 and λ_1 respectively. To impose the convexity of ϕ, one must impose

$$\lambda_1 < 0$$

that gives the family of entropy functionals $\phi(y) = \phi_{\mathrm{u}}(4sy)$ with

$$\phi_{\mathrm{u}}(u) = c \,+\, \lambda_0\, u \,+\, \lambda_1 \left[u\left(\operatorname{argtanh}\sqrt{1-u}\right)^2 - 2\sqrt{1-u}\, \operatorname{argtanh}\sqrt{1-u} - \log u\right] \mathbb{1}_{[0\,;\,1]}(u).$$

where c is an integration constant. Figure A.1(a) depicts function ϕ_{u} for the special choice $\lambda_0 = 0, \lambda_1 = -1$ and $\mathcal{X}$ being unbounded, c is chosen to be zero.

Appendix A.4.2. (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 $\mathcal{X}_\pm = \mathbb{R}_\pm$ allows such an interpretation, using then multiform entropies, while the alternative is to relax the concavity property of the entropy. To be more precise, one chooses either functions $T_{-,1}(x)$ and $T_{+,1}$, or function T_1 under the form

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

Over each set $\mathcal{X}_\pm$ we immediately get

$$\phi_\pm'(y) = 4s\left(\lambda_0 + \lambda_{\pm,1} \operatorname{argtanh}\sqrt{1-4sy}\right) \qquad \text{or} \qquad \widetilde{\phi}_\pm'(y) = 4s\left(\lambda_0 \pm \lambda_1 \operatorname{argtanh}\sqrt{1-4sy}\right)$$

where the sign and the factors are absorbed on λ_0 and $\lambda_{\pm,1}$. A judicious choice is then to impose

$$\lambda_{-,1} = \lambda_{+,1} = \overline{\lambda}_1 < 0 \qquad (\lambda_1 < 0)$$

and the same integration constant c for each branch leading either to the family of (convex) uniform functions $\phi(y) = \phi_{\mathrm{u}}(4sy)$ with,

$$\phi_{\mathrm{u}}(u) = c \,+\, \lambda_0\, u \,+\, \overline{\lambda}_1 \left(u \operatorname{argtanh}\sqrt{1-u} \,-\, \sqrt{1-u}\right) \mathbb{1}_{[0\,;\,1]}(u)$$

or to the family of multiform function $\widetilde{\phi}$, with branches $\widetilde{\phi}_{\pm,\mathrm{u}}(4sy)$,

$$\widetilde{\phi}_{\pm,\mathrm{u}}(u) = c \,+\, \lambda_0\, u \,\pm\, \lambda_1 \,\left(u \operatorname{argtanh}\sqrt{1-u} \,-\, \sqrt{1-u}\right) \mathbb{1}_{[0\,;\,1]}(u)$$

Function ϕ_{u} is represented figure A.1(b) for the special choice $c = \lambda_0 = 0, \overline{\lambda}_1 = -1$ (here, for $c = \lambda_0 = 0, \lambda_1 = -1$, $\widetilde{\phi}_\pm = \pm\phi$). The choice of equal $\lambda_{\pm,1}$ is equivalent than considering the constraint $T_1(x) = |x|$, and thus allows to respect the symmetries of the distribution, allowing thus to recover a classical ϕ-entropy.

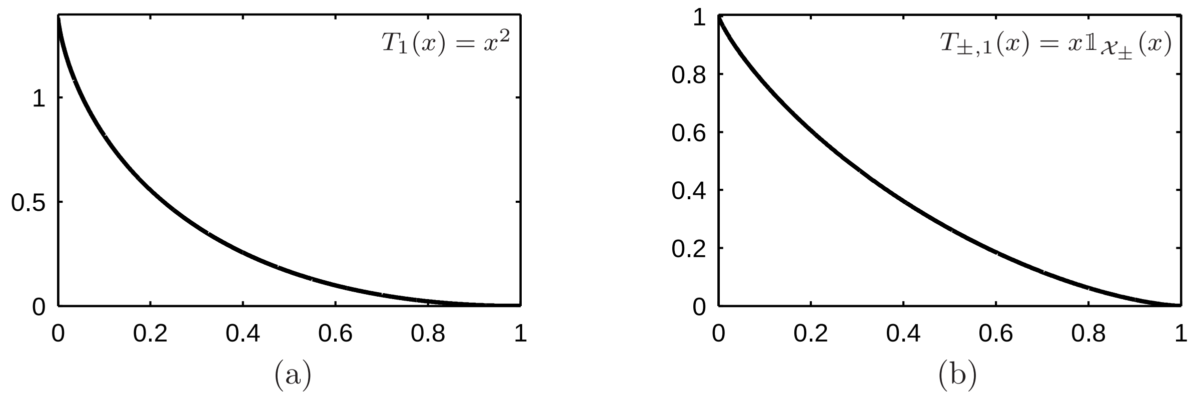

Figure A.1: Entropy functional ϕ_{u} derived from the logistic distribution: (a) with $T_1(x) = x^2$ and (b) with $T_{\pm,1}(x) = x \mathbb{1}_{\mathcal{X}_\pm}(x)$.

Appendix A.5. The arcsine distribution

The arcsine distribution is a special case of the beta distribution with $\alpha = \beta = \frac{1}{2}$. We consider here the centered and scaled version of this distribution which writes

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

The inverse distributions $f_{X,\pm}^{-1}$ on $\mathcal{X}_- = \left(-\sigma\sqrt{2}\,;\,0\right)$ and $X_+ = \left[0\,;\,\sigma\sqrt{2}\right)$ write then

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

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

Appendix A.5.1. Second order moment

When the second order moment $T_1(x) = x^2$ is constrained, one immediately obtains

$$\phi'(y) = \lambda_0 + \lambda_1\left(2\sigma^2 - \frac{1}{\pi^2y^2}\right)$$

The family of entropy functional is then

$$\phi(y) = c + \left(\lambda_0 + 2\sigma^2\lambda_1\right)y + \frac{\lambda_1}{\pi^2 y}$$

which drastically simplifies with the special choice

$$c = 0, \qquad \lambda_0 = -\frac{\alpha^2}{\pi^2} \qquad \text{and} \qquad \lambda_1 = \pi^2 \qquad \text{to} \qquad \phi(y) = \frac{1}{y}$$

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

Since the distribution does not share the sense of variation of $T_1(x) = x$, either we turn out to consider it as an extremal distribution of an entropy that is not concave, or as a maximum entropy when constraints are of the type

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

now

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

where the different factors and the sign are absorbed in the factors $\lambda_0, \lambda_{\pm,1}$. A judicious choice can be to impose

$$\lambda_{-,1} = \lambda_{+,1} = \overline{\lambda}_1 > 0$$

and the same integration constant c for each branch, leading then either to the family of (convex) uniform of functions $\phi(y) = \phi_{\mathrm{u}}(\sqrt{2}\pi\sigma y)$ with

$$\phi_{\mathrm{u}}(u) = c + \lambda_0\,u + \overline{\lambda}_1\left(\sqrt{u^2-1} + \arctan\left(\frac{1}{\sqrt{u^2-1}}\right)\right)\mathbb{1}_{(1\,;\,+\infty)}(u)$$

or, in the non-convex case, to the family of functions with branches $\widetilde{\phi}_\pm(y) = \widetilde{\phi}_{\pm,\mathrm{u}}(\sqrt{2}\pi\sigma y)$,

$$\widetilde{\phi}_{\pm,\mathrm{u}}(u) = c + \lambda_0\, u \pm \lambda \left(\sqrt{u^2-1} + \arctan\left(\frac{1}{\sqrt{u^2-1}}\right)\right) \mathbb{1}_{(1\,;\,+\infty)}(u)$$

The uniform function ϕ_u is represented figure A.2 for the special choice $c = \lambda_0 = 0, \overline{\lambda}_1 = 1$ (here again, for $c = \lambda_0 = 0, \lambda_1 = 1$, $\widetilde{\phi}_\pm = \pm\phi$). In this case again, the symmetrical choice for $\lambda_{\pm,1}$ allows to recover the symmetries of the probability density, and thus to a uniform convex entropy functional in the first context.

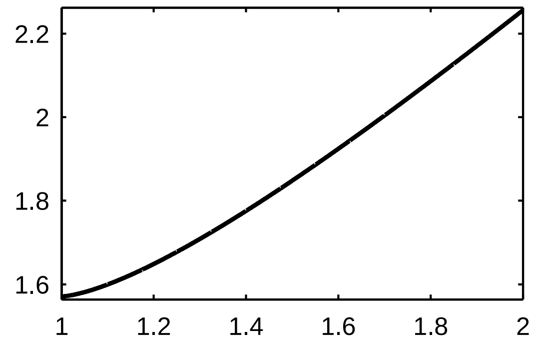

Figure A.2: Entropy functional ϕ_u derived from the arcsine distribution with partial constraints $T_{\pm,1}(x) = x \mathbb{1}_{\mathcal{X}_\pm}(x)$.

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

As a very special case, consider here this distribution, expressed as

$$f_X(x) = \frac{\beta^\alpha x^{\alpha-1} \exp(-\beta x)}{\Gamma(\alpha)} \qquad \text{on} \qquad \mathcal{X} = \mathbb{R}_+.$$

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

$$x_\mathrm{m} = \frac{\alpha-1}{\beta} \qquad \text{and} \qquad \tau = \frac{\Gamma(\alpha)}{\beta\,(\alpha-1)^{\alpha-1}}$$

Thus, here again it cannot be viewed as a maximum entropy constraint neither by any p-order moment. Here, we can again interpret it as a maximum entropy constrained by partial moments

$$T_{k,1}(x) = x^p, \quad k \in \{0, -1\} \qquad \text{over} \qquad \mathcal{X}_0 = [0\,;\, x_\mathrm{m}) \qquad \text{and} \qquad \mathcal{X}_{-1} = [x_\mathrm{m}\,;\, +\infty).$$

or as an extremal entropy constrained by the moment

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

where $p > 0$. Inverting $y = f_X(x)$ leads to the equation

$$-\frac{x}{x_\mathrm{m}} \exp\left(-\frac{x}{x_\mathrm{m}}\right) = -(\tau\, y)^{\frac{1}{\alpha-1}}$$

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 $u \rightsquigarrow u\exp(u)$[107, § 1], leading to the inverse functions

$$f_{X,k}^{-1}(y) = -x_\mathrm{m}\, \mathrm{W}_k\left(-(\tau y)^{\frac{1}{\alpha-1}}\right), \qquad y \in \left[0\,;\, \frac{1}{\tau e^{\alpha-1}}\right],$$

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.

One has thus to solve the equation

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

where the positive factor are absorbed in the λ_0, $\lambda_{k,1}$ and where to insure 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 $\lambda_{k,1}$. Integrating the previous expression is not an easy task. Relation $u\,(1+\mathrm{W}_k(u))\,\mathrm{W}_k'(x) = \mathrm{W}_k(u)$ [107, Eq. 3.2] suggests that a way to make the integration is to search for $\phi_k(y) = \phi_{k,\mathrm{u}}(\tau y)$ where $\phi_{k,\mathrm{u}}(u)$ is searched as the product of $u\left[-\mathrm{W}_k\left(-u^{\frac{1}{\alpha-1}}\right)\right]^p$ and a series of $\left[-\mathrm{W}_k\left(-u^{\frac{1}{\alpha-1}}\right)\right]$ and then to recognize the coefficients of the series. Such an approach leads to the family of entropic functional $\phi_k(y) = \phi_{k,\mathrm{u}}(\tau y)$ with

$$\begin{aligned}\phi_{k,\mathrm{u}}(u) &= c_k + \lambda_0\, u\\ &+ \lambda_{k,1}\, u\left[-\mathrm{W}_k\left(-u^{\frac{1}{\alpha-1}}\right)\right]^p \left[1 - \tfrac{p}{p+\alpha-1}\ {}_1F_1\left(1\,;\,p+\alpha\,;\,(1-\alpha)\,\mathrm{W}_k\left(-u^{\frac{1}{\alpha-1}}\right)\right)\right] \mathbb{1}_{(0\,;\,e^{1-\alpha})}(u)\end{aligned}$$

where ${}_1F_1$ is the confluent hypergeometric (or Kummer's) function [108, § 13] and c_k are integration constants. One can verify a posteriori that these functions are the ones we search for. The integration constant can be chosen such that ϕ_k coincide in 0 for instance, that gives

$$c_{-1} - c_0 = \frac{p\,\Gamma(p+\alpha-1)}{(\alpha-1)^{p+\alpha-1}}\,\lambda_{-1,1}$$

using successively [107, Eq. 3.1] and [108, Eq. 13.1.2] for W_0, and successively [108, Eq. 13.1.4] (W_{-1} tending to $-\infty$ in 0^-), $\mathrm{W}_{-1}(u)\exp(\mathrm{W}_{-1}(u)) = u$, and [107, Eq. 4.6 & lines that follow] for W_{-1}. The same algebra leads to the same expression for the $\widetilde{\phi}_k$, except that $\lambda_{k,1}$ are replaced by a unique λ_1.

The multivalued function ϕ_u in the concave context is represented figure A.3 for $p = 2, \alpha = 2$ and $\alpha = 5$, and with the choices $c_{-1} = \lambda_0 = 0, \lambda_{0,1} = 1, \lambda_{-1,1} = -0.1$.

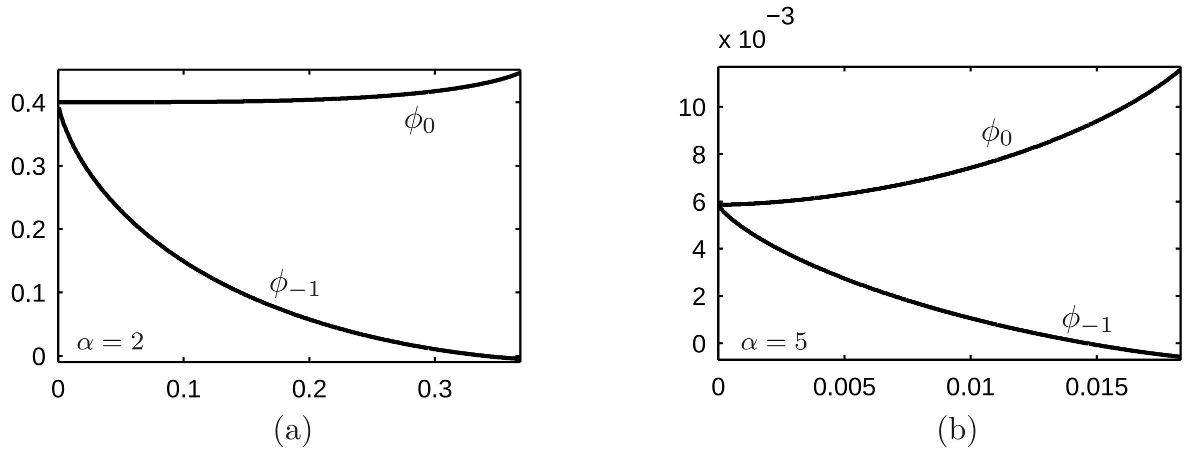

Figure A.3: Multiform entropy functional ϕ_u derived from the gamma distribution with the partial moment constraints $T_{k,1}(x) = x^2 \mathbb{1}_{\mathcal{X}_k}(x)$, $k \in \{0, -1\}$. (a): $\alpha = 2$; (b): $\alpha = 5$.

References

[1] J. von Neumann. Thermodynamik quantenmechanischer gesamtheiten. *Nachrichten von der Gesellschaft der Wissenschaften zu Göttingen*, 1:273–291, 1927.

[2] C. E. Shannon. A mathematical theory of communication. *The Bell System Technical Journal*, 27(4):623–656, October 1948.

[3] L. Boltzmann (translated by Stephen G. Brush). *Lectures on Gas Theory*. Dover, Leipzig, Germany, 1964.

[4] M. Planck. *Eight Lectures on Theoretical Physics*. Columbia University Press, New-York, 2015.

[5] F. R. S. W. D. Nieven, M. A. *The scientific papers of James Clerk Maxwell*, volume 2. Dover, New-York, 1952.

[6] E. T. Jaynes. Gibbs vs Boltzmann entropies. *American Journal of Physics*, 33(5):391–398, May 1965.

[7] I. Müller and W. H. Müller. *Fundamentals of Thermodynamics and Applications. With Historical Annotations and Many Citations from Avogadro to Zermelo*. Springer, Berlin, 2009.

[8] A. Rényi. On measures of entropy and information. *in Proceeding of the 4th Berkeley Symposium on Mathematical Statistics and Probability*, 1:547–561, 1961.

[9] R. S. Varma. Generalization of Rényi's entropy of order α. *Journal of Mathematical Sciences*, 1:34–48, 1966.

[10] J. Havrda and F. Charvát. Quantification method of classification processes: Concept of structural α-entropy. *Kybernetika*, 3(1):30–35, 1967.

[11] I. Csiszàr. Information-type measures of difference of probability distributions and indirect observations. *Studia Scientiarum Mathematicarum Hungarica*, 2:299–318, 1967.

[12] Z. Daróczy. Generalized information functions. *Information and Control*, 16(1):36–51, March 1970.

[13] J Aczél and Z. Daróczy. *On Measures of Information and Their Characterizations*. Academic Press, New-York, 1975.

[14] Z. Daróczy and A. Járai. On the measurable solution of a functional equation arising in information theory. *Acta Mathematica Academiae Scientiarum Hungaricae*, 34(1-2):105–116, March 1979.

[15] C. Tsallis. Possible generalization of Boltzmann-Gibbs statistics. *Journal of Statistical Physics*, 52(1-2):479–487, July 1988.

[16] M. Salicrú. Funciones de entropía asociada a medidas de Csiszár. *Qüestiió*, 11(3):3–12, 1987.

[17] M. Salicrú, M. L. Menéndez, D. Morales, and L. Pardo. Asymptotic distribution of (h, ϕ)-entropies. *Communications in Statistics – Theory and Methods*, 22(7):2015–2031, 1993.

[18] M. Salicrú. Measures of information associated with Csiszár's divergences. *Kybernetica*, 30(5):563–573, 1994.

[19] F. Liese and I. Vajda. On divergence and informations in statistics and information theory. *IEEE Transactions on Information Theory*, 52(10):4394–4412, October 2006.

[20] M. Basseville. Divergence measures for statistical data processing – an annotated bibliography. *Signal Processing*, 93(4):621–633, April 2013.

[21] P. F. Panter and W. Dite. Quantization distortion in pulse-count modulation with nonuniform spacing of levels. *Proceedings of the IRE*, 39(1):44–48, January 1951.

[22] S. P. Lloyd. Least squares quantization in PCM. *IEEE Transactions on Information Theory*, 28(2):129–137, March 1982.

[23] A. Gersho and R. M. Gray. *Vector quantization and signal compression*. Kluwer, Boston, 1992.

[24] L. L. Campbell. A coding theorem and Rényi's entropy. *Information and Control*, 8(4):423–429, August 1965.

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

[26] J. Burbea and C. R. Rao. On the convexity of some divergence measures based on entropy functions. *IEEE Transactions on Information Theory*, 28(3):489–495, May 1982.

[27] M. L. Menéndez, D. Morales, L. Pardo, and M. Salicrú. (h, ϕ)-entropy differential metric. *Applications of Mathematics*, 42(1-2):81–98, 1997.

[28] L. Pardo. *Statistical Inference Based on Divergence Measures*. Chapman & Hall, Boca Raton, FL, USA, 2006.

[29] E. .T Jaynes. Information theory and statistical mechanics. *Physical Review*, 106(4):620–630, may 1957.

[30] J. N. Kapur. *Maximum Entropy Model in Sciences and Engineering*. Wiley Eastern Limited, New-Dehli, 1989.

[31] C. Arndt. *Information Measures: Information and Its Description in Sciences and Engeniering*. Springer Verlag, Berlin, 2001.

[32] T. M. Cover and J. A. Thomas. *Elements of Information Theory*. John Wiley & Sons, Hoboken, New Jersey, 2nd edition, 2006.

[33] E. T. Jaynes. Prior probabilities. *IEEE transactions on systems science and cybernetics*, 4(3):227–241, September 1968.

[34] I. Csiszàr. Why least squares and maximum entropy? an axiomatic approach to inference for linear inverse problems. *The Annals of Statistics*, 19(4):2031–2066, December 1991.

[35] B. A. Frigyik, S. Srivastava, and M. R. Gupta. Functional Bregman divergence and Bayesian estimation of distributions. *IEEE Transactions on Infomation Theory*, 54(11):5130–5139, November 2008.

[36] C. P. Robert. *The Bayesian Choice. From Decision-Theoretic Foundations to Computational Implementation*. Springer, New-York, 2nd edition, 2007.

[37] L. K. Jones and C. L. Byrne. General entropy criteria for inverse problems, with applications to data compression, pattern classification, and cluster analysis. *IEEE transactions on Information Theory*, 36(1):23–30, January 1990.

[38] A. O. Hero III, B. Ma, O. J. J. Michel, and J. Gorman. Application of entropic spanning graphs. *IEEE Signal Processing Magazine*, 19(5):85–95, September 2002.

[39] S. Y. Park and A. K. Bera. Maximum entropy autoregressive conditional heteroskedasticity model. *Journal of Econometrics*, 150(2):219–230, June 2009.

[40] O. Vasicek. A test for normality based on sample entropy. *Journal of the Royal Statistical Society B*, 38(1):54–59, 1976.

[41] D.V. Gokhale. On entropy-based goodness-of-fit tests. *Computational Statistics and Data Analysis*, 1:157–165, march 1983.

[42] K.-S. Song. Goodness-of-fit tests based on Kullback-Leibler discrimination information. *IEEE Transactions on Information Theory*, 48(5):1103–1117, May 2002.

[43] J. Lequesne. A goodness-of-fit test of Student distributions based on Rényi entropy. In A. Djafari and F. Barbaresco, editors, *AIP conference proceedings of the 34th international workshop on Bayesian Inference and Maximum Entropy Methods (MaxEnt'14)*, volume 1641, pages 487–494, Amboise, France, 21-25 september 2014.

[44] J. Lequesne. *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.

[45] V. Girardin and P. Regnault. Escort distributions minimizing the Kullback-Leibler divergence for a large deviations principle and tests of entropy level. *Annals of the Institute of Statistical Mathematics*, 68(2):439–468, April 2015.

[46] H. K. Kesavan and J. N. Kapur. The generalized maximum entropy principle. *IEEE Transactions on Systems Man and Cybernetics*, 19:1042–1052, 1989.

[47] J. M. Borwein and A. S. Lewis. Duality relationships for entropy-like minimization problems. *SIAM Journal on Control and Optimization*, 29(2):325–338, March 1991.

[48] J. M. Borwein and A. S. Lewis. Convergence of best entropy estimates. *SIAM Journal of Optimization*, 1(2):191–205, May 1991.

[49] J. M. Borwein and A. S. Lewis. Partially-finite programming in L_1 and the existence of maximum entropy estimates. *SIAM Journal of Optimization*, 3(2):248–267, May 1993.

[50] C. Tsallis, R. M. Mendes, and A. R. Plastino. The role of constraints within generalized nonextensive statistics. *Physica A*, 261(3-4):534–554, December 1998.

[51] C. Tsallis. Nonextensive statistics: theoretical, experimental and computational evidences and connections. *Brazilian Journal of Physics*, 29(1):1–35, March 1999.

[52] C. Tsallis. *Introduction to Nonextensive Statistical Mechanics – Approaching a Complex World*. Springer Verlag, New-York, 2009.

[53] C. Essex, C. Schulzsky, A. Franz, and K. H. Hoffmann. Tsallis and Rényi entropies in fractional diffusion and entropy production. *Physica A*, 284(1-4):299–308, September 2000.

[54] A. S. Parvan and T. S. Biró. Extensive Rényi statistics from non-extensive entropy. *Physics Letters A*, 340(5-6):375–387, June 2005.

[55] S. M. Kay. *Fundamentals for Statistical Signal Processing: Estimation Theory*. vol. 1. Prentice Hall, Upper Saddle River, NJ, 1993.

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

[57] Jeffrey. An invariant form for the prior probability in estimation problems. *Proceedings of the Royal Society A*, 186(1007):453–461, September 1946.

[58] C. Vignat and J.-F. Bercher. Analysis of signals in the Fisher-Shannon information plane. *Physics Letters A*, 312(1-2):27–33, June 2003.

[59] E. Romera, J. C. Angulo, and J. S. Dehesa. Fisher entropy and uncertainty like relationships in many-body systems. *Physical Review A*, 59(5):4064–4067, May 1999.

[60] E. Romera, P. Sánchez-Moreno, and J. S. Dehesa. Uncertainty relation for Fisher information of D-dimensional single-particle systems with central potentials. *Journal of Mathematical Physics*, 47(10):103504, October 2006.

[61] P. Sánchez-Moreno, R. González-Férez, and J. S. Dehesa. Improvement of the Heisenberg and Fisher-information-based uncertainty relations for D-dimensional potentials. *New Journal of Physics*, 8:330, December 2006.

[62] I. V. Toranzo, S. Lopez-Rosa, R.O. Esquivel, and J. S. Dehesa. Heisenberg-like and Fisher-information uncertainties relations for N-fermion d-dimensional systems. *Physical Review A*, page on press, 2015.

[63] A. J. Stam. Some inequalities satisfied by the quantities of information of Fisher and Shannon. *Information and Control*, 2(2):101–112, June 1959.

[64] A. Dembo, T. M. Cover, and J. A. Thomas. Information theoretic inequalities. *IEEE Transactions on Information Theory*, 37(6):1501–1518, November 1991.

[65] D. Guo, S. Shamai, and S. Verdú. Mutual information and minimum mean-square error in Gaussian channels. *IEEE Transactions on Information Theory*, 51(4):1261–1282, April 2005.

[66] G. B. Folland and A. Sitaram. The uncertainty principle: A mathematical survey. *Journal of Fourier Analysis and Applications*, 3(3):207–233, May 1997.

[67] K. D. Sen. *Statistical Complexity. Application in Electronic Structure*. Springer Verlag, New-York, 2011.

[68] I. Vajda. χ^α-divergence and generalized Fisher's information. In *Transactions of the 6th Prague Conference on Information Theory, Statistics, Decision Functions and Random Processes*, pages 873–886, 1973.

[69] D. E. Boekee. An extension of the Fisher information measure. In I. Csiszàr and P. Elias., editors, *Topics in information theory. Proc. 2nd Colloquium on Information Theory*, volume 16, pages 113–123, Keszthely, Hungary, 25-29 August 1975 1977. Colloquia Mathematica Societatis János Bolyai and North Holland, Amsterdam.

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

[71] D. E. Boekee and J. C. A. Van der Lubbe. The r-norm information measure. *Information and Control*, 45(2):136:155, May 1980.

[72] E. Lutwak, D. Yang, and G. Zhang. Moment-entropy inequalities. *The Annals of Probability*, 32(1B):757–774, January 2004.

[73] E. Lutwak, D. Yang, and G. Zhang. Cramér-Rao and moment-entropy inequalities for Rényi entropy and generalized Fisher information. *IEEE Transactions on Information Theory*, 51(2):473–478, February 2005.

[74] E. Lutwak, D. Yang, and G. Zhang. Moment-entropy inequalities for a random vector. *IEEE Transactions on Information Theory*, 53(4):1603–1607, April 2007.

[75] E. Lutwak, S. Lv, D. Yang, and G. Zhang. Extension of Fisher information and Stam's inequality. *IEEE Transactions on Information Theory*, 58(3):1319–1327, March 2012.

[76] J.-F. Bercher. On a (β, q)-generalized Fisher information and inequalities invoving q-Gaussian distributions. *Journal of Mathematical Physics*, 53(6):063303, june 2012.

[77] J.-F. Bercher. On generalized Cramér-Rao inequalities, generalized Fisher information and characterizations of generalized q-Gaussian distributions. *Journal of Physics A*, 45(25):255303, June 2012.

[78] J.-F. Bercher. On multidimensional generalized Cramér-Rao inequalities, uncertainty relations and characterizations of generalized q-Gaussian distributions. *Journal of Physics A*, 46(9):095303, March 2013.

[79] J.-F. Bercher. Some properties of generalized Fisher information in the context of nonextensive thermostatistics. *Physica A*, 392(15):3140–3154, August 2013.

[80] L. M. Bregman. 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*, 7(3):200–217, 1967.

[81] F. Nielsen and R. Nock. Generalizing skew Jensen divergences and Bregman divergences with comparative convexity. *IEEE Signal Processing Letters*, 24(8):1123–1127, August 2017.

[82] A. Ben-Tal, J. M. Bornwein, and M. Teboulle. Spectral estimation via convex programming. In Fred Young Phillips and John James Rousseau, editors, *Systems and Management Science by Extremal Methods*, chapter 18, pages 275–290. Springer, 1992.

[83] M. Teboulle and I. Vajda. Convergence of best ϕ-entropy estimates. *IEEE Transactions on Infomation Theory*, 39(1):297–301, January 1993.

[84] V. Girardin. Méthodes de réalisation de produit scalaire et de problème de moments avec maximisation d'entropie. *Studia Mathematica*, 124(3):199–213, 1997.

[85] V. Girardin. Relative entropy and spectral constraints : Some invariance properties of the ARMA class. *Journal of Time Series Analysis*, 28(6):844–866, November 2007.

[86] J. A. Costa, A. O. Hero III, and C. Vignat. On solutions to multivariate maximum α-entropy problems. In Anand Rangarajan, Mário A. T. Figueiredo, and Josiane Zerubia, editors, *4th International Workshop on Energy Minimization Methods in Computer Vision and Pattern Recognition (EMMCVPR)*, volume 2683 of *Lecture Notes in Computer Sciences*, pages 211–226, Lisbon, Portugal, July 7-9 2003. Springer Verlag.

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

[88] C. Beck and F. Schögl. *Thermodynamics of chaotic systems: an introduction*. Cambridge University Press, Cambridge, 1993.

[89] J. Naudts. *Generalized Thermostatistics*. Spri, London, 2011.

[90] S. Martínez, F. Nicolás, F. Pennini, and A. Plastino. Tsallis' entropy maximization procedure revisited. *Physica A*, 286(3-4):489–502, November 2000.

[91] L. P. Chimento, F. Pennini, and A. Plastino. Naudts-like duality and the extreme Fisher information principle. *Physical Review E*, 62(5):7462–7465, November 2000.

[92] M. Casas, L. Chimento, F. Pennini, A. Plastino, and A. R. Plastino. Fisher information in a Tsallis non-extensive environment. *Chaos, Solitons and Fractals*, 13(3):451–459, March 2002.

[93] D. V. Widder. *The Heat Equation*. Academic Press, New-York, 1975.

[94] C. Tsallis and E. K. Lenzi. Anomalous diffusion: nonlinear fractional Fokker-Planck equation. *Chemical Physics*, 284(1-2):341–347, November 2002.

[95] J. L. Vázquez. *Smoothing and Decay Estimates for Nonlinear Diffusion Equations – Equation of Porous Medium Type*. Oxford University Press, New-York, USA, 2006.

[96] B. H. Gilding and R. Kersner. *Travelling Waves in Nonlinear Diffusion-Convection Reaction*. Springer, Basel, Switzerland, 2004.

[97] P. Harremoës and C. Vignat. An entropy power inequality for the binomial family. *Journal of Inequalities in Pure and Applied Mathematics*, 4(5):93, 2003.

[98] O. Johnson and Y. Yu. Monotonicity, thinning, and discrete versions of the entropy power inequality. *IEEE Transactions on Information Theory*, 56(11):5387–5395, nov 2010.

[99] S. Haghighatshoar, E. Abbe, and I. E. Telatar. A new entropy power inequality for integer-valued random variables. *IEEE Transactions on Information Theory*, 60(7):3787–3796, jul 2014.

[100] S. G. Bobkov and G. P. Chistyakov. Entropy power inequality for the Rényi entropy. *IEEE Transactions on Information Theory*, 61(2):708–714, February 2015.

[101] S. Zozor, D. Puertas-Centeno, and J. S. Dehesa. On generalized Stam inequalities and Fisher–Rényi complexity measures. *Entropy*, 19(9):493, September 2017.

[102] O. Rioul. Yet another proof of the entropy power inequality. *IEEE Transactions on Information Theory*, 63(6):3595–3599, June 2017.

[103] M. Rosenblatt. Remarks on some nonparametric estimates of a density function. *The Annals of Mathematical Statistics*, 27(3):832–837, September 1956.

[104] E. Parzen. On estimation of a probability density function and mode. *The Annals of Mathematical Statistics*, 33(3):1065–1076, September 1962.

[105] J. Beirlant, E. J. Dudewicz, L. Györfi, and E. C. van der Meulen. Nonparametric entropy estimation: An overview. *International Journal of Mathematical and Statistical Sciences*, 6(1):17–39, June 1997.

[106] N. Leonenko, L. Pronzato, and V. Savani. A class of Rényi information estimators for multidimensional densities. *Annals of Statistics*, 36(5):2153–2182, October 2008.

[107] R. M. Corless, G. H. Gonnet, D. E. G. Hare, D. J. Jeffrey, and D. E. Knuth. On the Lambert W function. *Advances in Computational Mathematics*, 5(1):329–359, December 1996.

[108] M. Abramowitz and I. A. Stegun. *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables*. 9th printing. Dover, New-York, 1970.