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

The investigation on

ϕ-entropies c'est important; On trouve deja des premieres briques dans Burbea Rao et dpuis ces balbutiements (cf si pas plus anciennes), de tres nombreux travaux se sont empare de celles-ci. Outre le fait qu'elles offrent un cadre tres general d'etude de nombreuses entropies qu'elles contiennent (Shannon, HCT...), elles offrent egalement champs a de nombreuses applications allant de... a...[Pardo]. Souvent phi-divergence [Csizar, Pardo...] mais entropie pour ref uniforme.

Dans le champs des applications utilisant phi-ent, concept de max ent tient une place particuliere. Principe physique [Jaynes]. Pb direct est que etant donnes contraintes physiques (centre de gravite, puissance) loi la moins "informative", max ent. Meme idee peut etre utilise contexte Bayesien pour le choix de loi a priori par exemple. Base sur entropie naturelle, de Shannon (Boltzman, von Neuman selon le contexte), mais memes ides pour des entropies plus generales sans que la methodologie change fondamentalement.

On peut cependant se poser un probleme inverse, ou une loi est observee/donnee. Following Kesavan s'interesse a deux problemes inverses. Si on suppose principe physique de max ent, on s'interesse donc aux contraintes imposee par le systeme, la loi devant (physiquement) etre max ent. Second probleme inverse consiste a se dire que, la loi etant donnee et a priori le systeme etant egalement contraint, quelle serait l'entropie adequate decrivant le systeme, i.e., telle que f soit max ent pour celle-ci. Pb a du sens ou Shannon/Boltzman decrit systeme en en regime en equilibre, ce qui n'est pas necessairement cas de tout systeme (cf Tsallisseries). Ici, on generalise approche complete de Kesavan concernant cette problematique maxent.

Adosse a Shannon, diverses mesures alternative moment, Fisher, le tout relie pr inegalites ou identites (CR, Stam, de Bruijn...). Ce caneva complet a ete generalise pour des classes particulieres, essentiellement Renyi-Tsallis [Liese, Vajda, Lutwak, JFB et cf ancienne intro].

The maximal entropy principle (in short MaxEnt principle), which consists in searching maximum entropy law under moments constrainst, is very usual in physics. The underlying idea is to determine the less informative law which governs a system (in equilibrium) subject to physical constraints [1–4]. Such an approach finds also its counterpart in estimation problems (in communication, clustering, pattern recongnition, in others) following the same principle of modeling laws given some constraints and with minimal information [3, 5, 6]. In the statistics field, some goodness-of-fit tests are based on entropic criteria that sometimes exploit the same idea of constrained maximal entropic law [7–12]. The principle of such entropy-based tests consist in comparing the entropy of the tested law, where the theoretical moments are replaced by their estimates from the data, to a direct estimation of the entropy using the data. As we will see later on, when the entropy used is such that the tested law is a maximal entropic law subject to the considered moment constraints, the difference of the entropy is nothing more than a Bregman divergence, that can be viewed as the distance of a law to a reference law. In the same vein, in estimation domain, when seeking the (likelihood or Bayesian) estimation of some parameters underlying a parametrized probability distribution, if too complicated, an alternative approach is to search the parameters that maximize the entropy of the law [13, 14] or that minimize a (Bregman, Csiszàr) divergence between laws when some *a priori* are given [15–17].

To briefly come back to the maximum entropy principle, let f be a probability distribution (of a random variable X) defined with respect to a general measure μ on a set $\mathcal{X}$ and $H[f] = -\displaystyle\int_{\mathcal{X}} f(x) \log f(x) \mathrm{d}\mu(x)$ be the Shannon entropy of f. Subject to n moment constraints such as $\mathbb{E}\left[T_i(X)\right] = t_i, i = 1, \ldots, n$ and to normalization, it is well known that the maximum entropy distribution lies within the exponential family, of the form [2–4, 18]

$$f(x) = \exp\left(\sum_{i=1}^{n} \lambda_i T_i(x) + \lambda_0\right).$$

Thus, given constraints, the MaxEnt law is know, provided one can determine the Lagrangian parameters λ_i insuring the normalization and moment constraints.

Conversely to the search for MaxEnt laws, one may whish to view a given probability distribution as a maximal entropy distribution subject to appropriate constraints. This approach has interest for instance in goodness-of-fit tests, where the law to be tested has to be the MaxEnt distribution. Provided that the considered distribution belong to the exponential family, it is then sufficient to determine the set of functions T_i adequately [2, 4, 19]. For instance, the gamma distribution can be viewed as a maximum entropy distribution if one knows the moments $\mathbb{E}\left[X\right]$ and $\mathbb{E}\left[\log(X)\right]$ [2, 4, 8, 19, 20]. A problem that may arise for instance in goodness-of-fit tests is the difficulty to estimate too complicate moments and/or to deal with laws outside the exponential family. Thus, in order to extend the idea of using the maximum entropy distributions principle, a natural idea is to consider generalized entropies [16, 21–32]. As we will see later on, in such a generalized context, the difference of the generalized entropy of the MaxEnt law subject to given moment constraints, and the entropy of a law sharing the same moments, remains a Bregman divergence, which can justify "informationally" that such an extension can potentially be used in goodness-to-fit tests. Moreover, to impose simple constraints, it is very natural to whish to adapt the entropy to both the law and the constraints, i.e., to adapt the entropic functional to a fixed law, given simple moment constraints. This objective is the central part of this paper.

If the entropy is a widely tool used for quantifying an information (or uncertainty) attached to a random variable or to a probability distribution, other quantities are very often used such as the moments of the variable, or the Fisher information, for instance in estimation contexts [33, 34] or for complex pysical systems descriptions [34–39]. Although coming form different worlds (information theory and communication, estimation, statistics, physics), both these informational quantities are linked by well known relations such that the famous Cramér-Rao's inequality, the de Bruijn's identity, the Stam's inequality [4, 40–42]. These relationships have been proved very usefull in various areas, for instance in communications [4, 40, 41], in estimation [33] or in physics [43, 44], in others. When generalizing entropy are considered, one naturally expect to generalized the other informational measure and the associated identities or inegalities. Such generalization gave birth to a huge amount of work and is still an active field of research [45–57]. It is natural to expect to put together generalized entropies, generalized moments and generalized Fisher information starting from the same entropic functional, and to derive generalized inequalities and identities from these quantities (both linked in some sense to the MaxEnt distribution). This is the second goal of this work.

The paper is organized as follows. Section 2 deals with the generalized MaxEnt distributions subject to a given set of constraints. In a first step, given an entropic functional, we derive a sufficient condition satisfied by this functional, the probability distribution and the set of constraints, to solve the problem. In passing, we show that the Bregman divergence between the MaxEnt and any distribution sharing the same moments (the MaxEnt being the reference one) reduces to a diference of entropies, which potential motivate their use in goodness-of-fit tests. In this same section, we then reverse the MaxEnt problem: fixing the probability distribution and a set of moments, we show the conditions that they must satisfy to be able to determine the distribution-adapted entropic functional and we formally determine this last one. When these necessary conditions are not 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 and the de Bruijn identity, relations saturated (or valid) dealing precisely for the generalized MaxEnt distribution. Finally, in section 6, we expose some example, from the well known cases (Gausian distribution, Shannon entropy, Fisher information, variance) to more exotic examples. The paper ends then with some discussions and further directions of investigations and applications.

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

Let us first recall the definition of the generalized (h, ϕ)-entropies, also known as entropies of the Salicrú's class or of the Csiszàr's class:

Definition 1 (ϕ-entropy [24, 30, 31])**.** 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 [24, 30, 31, 58]. 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 since h is monotone, we can restrict our study to the ϕ-entropies. Additionnaly, we will assume that ϕ is *strictly* convex and *differentiable*. A voir car ensuite les identites

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

Definition 2 (Bregman and functional divergence [16, 59])**.** 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. Such a general extension is 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

Refaire au vu des modif: Our purpose now is to define a ϕ-entropy for a given maximizing distribution f. This is the inverse problem of the very well known "MaxEnt problem" which consists in searching for the distribution maximizing the ϕ-entropy (1) subject to constraints on some moments $\mathbb{E}\left[T_i(X)\right]$ with $T_i:\mathbb{R}^d\mapsto\mathbb{R}$, $i=1,\ldots,n$. Here, instead of determining f given ϕ, we look for ϕ given a MaxEnt solution f. Let us recall that the MaxEnt 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

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

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 [60–64]. We recall herefater a sufficient condition relating f and ϕ so that f is the (resp. ϕ) is the desired solution of the problem.

Proposition 1 (Maximal ϕ-entropy solution [65–67])**.** *Suppose that there exists a probability distribution f 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 MaxEnt 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 $\mathbb{E}\left[T_i(X)\right]=t_i$. 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. $\square$

Hence, given an entropic functional ϕ, eq. (6) leads the the maximum entropy distribution $f^\star$.

2.2. Maximum entropy principle: the inverse problems

These problems were also been considered by Kesavan & in [?], although essentially derived in the Shannon entropy 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 [4?].

As also investigated by Kesavan [?], the second inverse problem consists in design the entropy itself, given a target distribution f and given the T_i. In other words, given a distribution f, eq. (6) should 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. In the direct problem, they are determined by the moments t_i. In the inverse problem, 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 sense of variations**, $\displaystyle\sum_{i=0}^{n} \lambda_i\, T_i(x_)$ 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, negative, then

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

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 (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 extent the definition of the ϕ-entropy, purpose of the next section.

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

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

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*

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

for some $(\lambda_0,\ldots,\lambda_n)\in\mathbb{R}^{n+1}$, then f is the unique solution of the extended MaxEnt 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 (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
$$\phi_l'(y) = \sum_{i=0}^{n}\lambda_i\,T_i\big(f_l^{-1}(y)\big) \tag{14}$$
where f_l^{-1} is the (possibly multivalued) inverse of f on $\mathcal{X}_l$.

In a conclusion, in the weak case where only condition (C1) 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 (C1). 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 (C1) 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.

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 [68, 69], with interesting connections with the standard thermodynamics [70] and with source coding [71, 72]. 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 $\mathcal{Y} \subset \mathbb{R}_+$. Then, if f is a probability distribution defined with respect to a general measure μ on a set $\mathcal{X}$ such that $f(\mathcal{X}) \subset \mathcal{Y}$, 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, in the context of the Shannon entropy, $\phi(x, y) = \phi(y) = y \log y$, the ϕ-escort density associated to f restricts to density f itself. In the Rényi-Tsallis context $\phi(x, y) = \phi(y) = y^q$, and $E_{\phi,f} \propto f^{2-q}$ which recovers the escort distributions used in the Rény-Tsallis context up to a duality transformation [70].

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

In the context of the Shannon entropy, the (α, ϕ)-moments are the usual moments of $\|X\|_{\mathcal{X}}^{\alpha}$, while in the Rény-Tsallis context the generalized moments introduced in [73, 74] are recovered.

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 [4, 33]. The Fisher information is also used as a method of inference and understanding in statistical physics and biology, as promoted by Frieden [34] and has been generalized in the Rényi-Tsallis context in a series of papers [47, 51, 53–56, 75**?**]. In what follows, we generalize these definition 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 $\|.\|_{\mathcal{X}*}$ 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 $\|.\|_\Theta$ and the corresponding dual norm is denoted $\|.\|_{\Theta*}$. 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 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. Similarly, in the Rényi-Tsallis context, the generalizations proposed in [54–56] are recovered.

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

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 $x f(x)$ vanishes in the boundary of $\mathcal{X}$. Thus, density f satisfies the (α, ϕ) extended Cramér-Rao inequality*

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

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 $T(x) = \|x\|_{\mathcal{X}}^{\alpha}$.

Note that the usual nonparametric Cramér-Rao inequality is recovered in the usual Shannon context $\phi(x, y) = y \log y$, using the euclidean norm and $\alpha = 2$.

Proof. The approach follows [56], 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 [56, Lemma 2]

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

The proof ends by choosing $g = E_{\phi,f}$ the ϕ-escort density associated to density f. Note now that, again from [56, 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 pdf that minimize 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 $\mathcal{X}$, parametrized by a parameter $\theta \in \Theta \subseteq \mathbb{R}^m$ and satisfying the conditions of definition 7. Assume that domain $\mathcal{X}$ does not depend on θ, 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*

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

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

Here again the usual parametric Cramér-Rao inequality and its Rény-Tsallis extentions are recovered as special cases.

Proof. The proof follows again that of [56], 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 [56, Lemma2]. □

micro conclusion et transition

5. ϕ-heat equation and extended de Bruijn identity

An important relation connecting the Shannon entropy H, comming from the "information world", to the Fisher information I, living in the "estimation world", is given by the de Bruijn identity and 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 [4]. 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 [76].

Inspired by the work [57], 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*,

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

for some $K \in \mathbb{R}^m$ (possibly dependent on θ) and where ϕ is a convex twice differentiable function defined over a set $\mathcal{X} \in \mathbb{R}_+$. Clearly, when $\phi(y) = y \log y$, for $K = \frac{1}{2}$ and $\beta = 2$, the standard heat-equation is recovered. The case where $\phi(y) = y^q$ was intensively studied in [57]. **finir avec discussion eq physiques**

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 $\|\nabla_x \phi'(f)\|^{\beta-2} \nabla_x \phi(f)$ vanishes in 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,

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

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 $\|\nabla_x \phi'(f)\|^{\beta-2} \nabla_x \phi(f)$ 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. $\square$

Micro conclusion/bilan

6. Some examples

Dans une telle construction, lorsque loi cible tends vers la Gaussienne/puissance par exemple, la phi va tendre vers la Shannon.

In the sequel, for sake of simplicity, we restricts our example to the univariate context $d = 1$.

6.1. 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

$$\phi'(y) = \big(\lambda_0 - \sigma^2 \log(2\pi\sigma^2)\,\lambda_1\big) - 2\sigma^2 \lambda_1 \log y.$$

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.

6.2. 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 [23, 25, 28, 77].

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 [54–56] (with the symmetric q index given here by $2-q$). The extended Cramér-Rao inequality proved in [55, 56] is then recovered.

6.3. q-exponential distribution and first-order moment

The same entropy functional can readily be obtained for the so-called q-exponential

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

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.

6.4. 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.

6.4.1. 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 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.

6.4.2. (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 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.

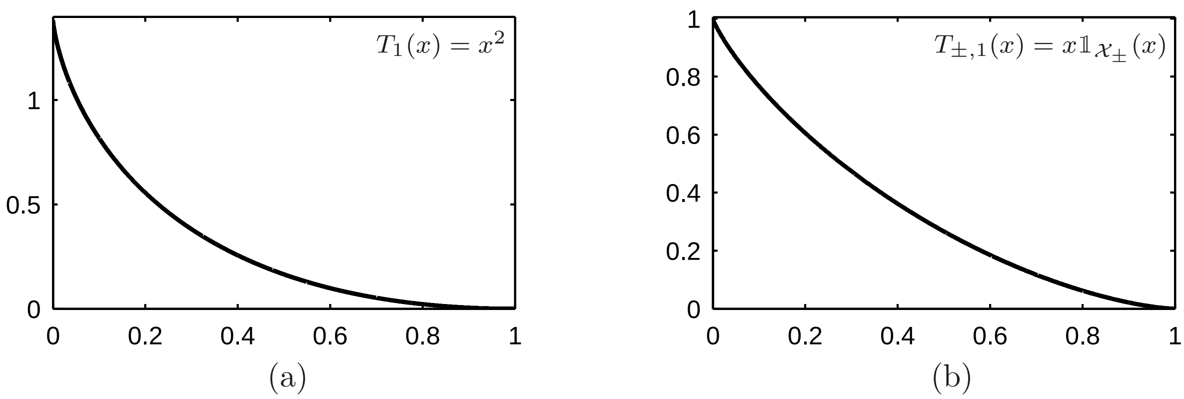

Figure 1: Entropy functional ϕ_{u} derived from the logistic distribution: (a) with $T_1(x) = x^2$ and (b) with $T_{\pm,1}(x) = x\mathbb{1}_{\mathcal{X}_\pm}(x)$.

6.5. 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

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

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

6.5.1. 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

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

which drastically simplifies with the special choice

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

6.5.2. (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

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

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

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

The uniform function ϕ_{u} is represented figure 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.

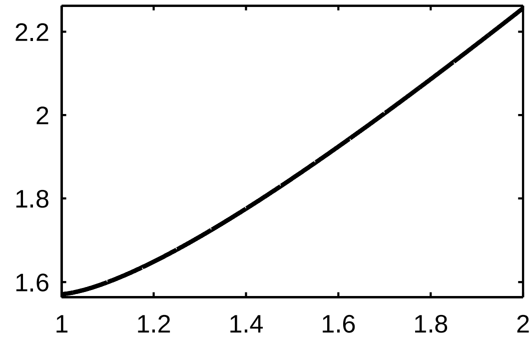

Figure 2: Entropy functional ϕ_{u} derived from the arcsine distribution with partial constraints $T_{\pm,1}(x) = x \mathbb{1}_{\mathcal{X}_\pm}(x)$.

6.6. 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 $z = \mathrm{W}(z) \exp(\mathrm{W}(z))$, 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 $\lambda_0, \lambda_{k,1}$ 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. Noting that $\mathrm{W}_k'(x) = \frac{\mathrm{W}_k(x)}{x(1+\mathrm{W}_k(x))}$, a way to make the integration is to search $\phi_k(y) = \phi_{k,\mathrm{u}}(\tau y)$ 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 and c_k are integration constants. The integration constant can be chosen such that ϕ_k coincide in 0 for instance, that gives

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

(see [78, eq. 13.1.4] and [79, eq.]). 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 3 for $p = 2, \alpha = 2$ and $\alpha = 5$, and with the choices $c_{-1} = \lambda_0 = 0, \lambda_{0,1} = 1, \lambda_{-1,1} = -0.1$.

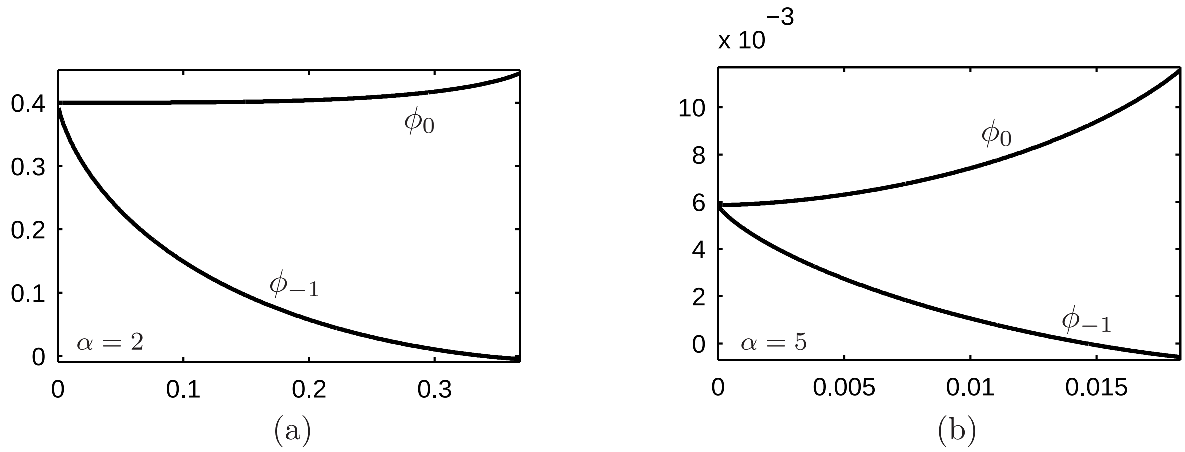

Figure 3: Multiform entropy functional ϕ_{u} derived from the gamma distribution with the partial moment constraints $T_{k,1}(x) = x^2 \mathbb{1}_{\mathcal{X}_k}(x)$, $k \in \{0, -1\}$. (a): $\alpha = 2$; (b): $\alpha = 5$.

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

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

[3] Arndt. *Information Measures: Information and Its Description in Sciences and Ingeniering*. Springer Verlag, Berlin, 2001.

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

[5] 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.

[6] 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.

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

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

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

[10] 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.

[11] 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.

[12] 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.

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

[14] E. T. Jaynes. On the rational of maximum-entropy methods. *Proceedings of the IEEE*, 70(9):939–952, September 1982.

[15] 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.

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

[17] 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.

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

[19] D. V. Gokhale. Maximum entropy characterizations of some distributions. In S. Kotz G. P. Patil and J. K. Ord, editors, *A Modern Course on Statistical Distributions in Scientific Work*, volume III, pages 299–304. Reidel, Dordrecht, Holland, 1975.

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

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

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

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

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

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

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

[27] 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.

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

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

[30] 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.

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

[32] 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.

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

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

[35] 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.

[36] 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.

[37] 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.

[38] 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.

[39] 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.

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

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

[42] 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.

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

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

[45] 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.

[46] 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.

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

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

[49] 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.

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

[51] 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.

[52] 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.

[53] 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.

[54] 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.

[55] 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.

[56] 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.

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

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

[59] 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.

[60] 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.

[61] 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.

[62] 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.

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

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

[65] C. Vignat, A. O. Hero III, and J. A. Costa. About closedness by convolution of the Tsallis maximizers. *Physica A*, 340(1-3):147–152, September 2004.

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

[67] J.-F. Bercher, V. Girardin, J. Lequesne, P. Regnault, and S. Zozor. Sur la recherche de ϕ-entropie à maximisante donnée. In *Colloque GRETSI*, Lyon, France, 8-11 September 2015.

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

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

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

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

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

[73] 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.

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

[75] 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.

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

[77] 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.

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

[79] 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.