Deriving generalized entropies from specific probability distributions

by jfb & co

Abstract

Where we show that it is possible to derive new entropies yielding a particular specified maximum entropy distribution. There are (probably) many errors –I hope not fundamental but is is possible; (certainly many) approximations, typos, maths and language mistakes.Suggestions and improvements will be much appreciated.

1. Maximum entropy distributions

classique, interesse par maximum d'entropies etant donnees un certain nombre de contraintes ohysiques. Si mesures donnent une distribution, on peut se poser la question de la voir comme max ent. Shannon, max ent sous contraintes... En notant cela, deux constatations, lois exponentielles, sinon out. Si tel est le cas, ok, mais moment peuvent être complexes. Par exemple, gamma avec moments... Les deux limitations peuvent etre levees en

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 $S[f] = -\int_{\mathcal{X}} f(x)\log f(x) \mathrm{d}\mu(x)$ 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

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

In order to recover known probability distributions (that must belong to the exponential family), it is then sufficient to specify a set of functions T_i. This has been used by many authors [**? ? ?**]. For instance, the gamma distribution can be viewed as a maximum entropy distribution if one knows the moments $\mathbb{E}\left[X\right]$ and $\mathbb{E}\left[\log(X)\right]$ [**? ? ? ? ?**]. In order to find maximum entropy distributions with simpler constraints or distributions outside of the exponential family, it is possible to consider other entropies, which is discussed below. This problem find interests in goodness-and-fit tests based on maximum entropy principle [**? ? ?**] **some words on the principle to motivate this stuff.**.

2. Maximum (h,ϕ)-entropy distributions

2.1. Definition and maximum (h,ϕ)-entropy solution

Definition 1 (ϕ-entropy [**? ? ?**]). Let $\phi : \mathcal{Y} \subseteq \mathbb{R}_+ \mapsto \mathbb{R}$ be a strictly convex differentiable function defined on the closed 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 $f(\mathcal{X}) \subseteq \mathcal{Y}$,

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

is the ϕ-entropy of f. Since ϕ is convex, then the entropy functional $H_\phi[f]$ is concave. Also note that the composition of a concave function with a nondecreasing concave function preserves concavity, and that composition of a convex function with a nonincreasing convex function yields a concave functional.

Definition 2 ((h,ϕ)-entropy [**? ? ?**]). With the same assumption as in definition 1,

$$H_{h,\phi}[f] = h\left(-\int_{\mathcal{X}} \phi(f(x))\mathrm{d}\mu(x)\right) \tag{2}$$

is called (h,ϕ)-entropy of f, where

- either ϕ is convex and h concave nondecreasing,
- or ϕ is concave and h convex nonincreasing

These (h,ϕ)-entropies have been studied in [**? ? ?**] for instance. In these works neither concavity (resp. convexity) of h, nor the differentiability of ϕ are imposed.

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

Definition 3 (Bregman divergence [? ?]). With the same assumption than in definition 1, the Bregman divergence associated with ϕ defined on a closed convex set $\mathcal{Y}$, is given by the function defined on $\mathcal{Y} \times \mathcal{Y}$,

$$D_\phi(y_1, y_2) = \phi(y_1) - \phi(y_2) - \phi'(y_2)\,(y_1 - y_2)\,. \tag{3}$$

A direct consequence of the strict convexity of ϕ is the nonnegativity of the Bregman divergence: $D_\phi(y_1, y_2) \geq 0$ with equality if and only if $y_1 = y_2$.

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 $\mathcal{Y} \subseteq \mathbb{R}_+$.

Consider the problem of maximizing entropy (2) subject to constraints on some moments $\mathbb{E}\left[T_i(X)\right]$ with $T_i : \mathbb{R}^d \mapsto \mathbb{R}$ (e.g., a location constraints consists of d constraints of this kind). Without loss of generality, we consider in the sequel that ϕ is convex. Since h is nondecreasing, it is enough to look for the maximum of the ϕ-entropy (1),

$$\begin{cases} \max_f & -\displaystyle\int_{\mathcal{X}} \phi(f(x))\mathrm{d}\mu(x) \\ \text{s.t.} & \displaystyle\int_{\mathcal{X}} f(x)\mathrm{d}\mu(x) = 1 \\ \text{s.t.} & \mathbb{E}\left[T_i(X)\right] = t_i, \quad i = 1, \ldots, n \end{cases} \tag{4}$$

Proposition 1 (Maximal ϕ-entropy solution)**.** *The probability distribution f solution of the maximum entropy problem* (4) *satisfies the equation*

$$\phi'\big(f(x)\big) = \lambda_0 + \sum_{i=1}^{n} \lambda_i\, T_i(x), \tag{5}$$

where parameters λ_i are such that the constraints are satisfied.

Proof. The maximization problem being concave, the solution exists and is unique. Equation (5) results directly from the classical Lagrange multipliers technique (see Girardin's theorem [?]). An alternative derivation of the result consists in checking that the distribution (5) is effectively a maximum entropy distribution, by showing that $H_\phi[f] > H_\phi[g]$ for all probability distributions g with given (fixed) moments $\mathbb{E}\left[T_i(X)\right] = t_i$. To this end, consider the functional Bregman divergence [?], acting on functions defined on a common domain $\mathcal{X}$:

$$\mathcal{D}_\phi(f_1, f_2) = \int_{\mathcal{X}} \phi(f_1(x))\mathrm{d}\mu(x) - \int_{\mathcal{X}} \phi(f_2(x))\mathrm{d}\mu(x) - \int_{\mathcal{X}} \phi'(f_2(x))\,(f_1(x) - f_2(x))\,\mathrm{d}\mu(x). \tag{6}$$

From the nonnegativity of the Bregman divergence this functional divergence is nonnegative as well, and zero if and only if $f_1 = f_2$ almost everywhere. Define by

$$C_t = \left\{ g : \mathcal{X} \mapsto \mathbb{R}_+ : \ \int_{\mathcal{X}} g(x)\mathrm{d}\mu(x) = 1,\ \mathbb{E}\left[T_i(X)\right] = t_i,\ i = 1, \ldots, n \right\}$$

the set of all probability distributions defined on $\mathcal{X}$ with given moments $t = (t_1, \ldots, t_n)$. Consider now $f \in C_t$ such that $\phi'(f(x)) = \lambda_0 + \displaystyle\sum_{i=1}^{n} \lambda_i\, T_i(x)$ and any given function $g \in C_t$. Then

$$\begin{aligned} \mathcal{D}_\phi(g, f) &= \int_{\mathcal{X}} \phi(g(x))\mathrm{d}\mu(x) - \int_{\mathcal{X}} \phi(f(x))\mathrm{d}\mu(x) - \int_{\mathcal{X}} \phi'(f(x))\,(g(x) - f(x))\,\mathrm{d}\mu(x) \\ &= -H_\phi[g] + H_\phi[f] - \int_{\mathcal{X}} \left(\lambda_0 + \sum_{i=1}^{n} \lambda_i\, T_i(x)\right)(g(x) - f(x))\,\mathrm{d}\mu(x) \\ &= H_\phi[f] - H_\phi[g] \end{aligned}$$

where we used the fact that g and f have 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 t than f, with equality if and only if $g = f$ almost everywhere. In other words, this shows that f, solution of (5), realizes the maximum of $H_\phi[g]$ over C_t. □

2.2. Defining new entropy functionals

Given an entropy functional, we thus obtain a maximum entropy distribution. There exists numerous (h,ϕ)-entropies in the literature. However a few of them lead to explicit forms for the maximum entropy distribution. Conversely, it is of high interest to look for the entropies that lead to a specified distribution as a maximum entropy solution. As pointed out previously, this find interests in goodness-and-fit tests based in entropies: it seems convenient to realize such tests using the entropy such that the distribution tested corresponds to its maximum entropy [**? ? ?**] **principle to roughly recall**.

Since we will look for a functional such that a given probability distribution f is its ϕ-entropy maximizer under moment constraints, we also see that the corresponding λ_i parameters can be included in the definition of the function.

Let us recall some implicit properties of ϕ.

- Its derivative ϕ' is defined on a domain that includes $f(\mathcal{X})$;
- From the strict convexity property of ϕ, necessarily ϕ' is strictly increasing.

The identification of a function ϕ such that a given distribution f is the associated maximum entropy distribution amounts to solve (5), that is:

1. choose a set of functions T_i, $i = 1, \ldots, n$,
2. find ϕ' satisfying $\lambda_0 + \sum_{i=1}^{n} \lambda_i T_i(x) = \phi'(f(x))$,
3. integrate the result ϕ' to get $\phi(y) = \int \phi'(y) \mathrm{d}y$
4. Parameters λ_i may be chosen case by case in order to simplify the expression of ϕ.

Note that to be solvable, eq. (5) requires that $f(x)$ and $\sum_{i=1}^{n} \lambda_i T_i(x)$ share the same isoprobability subsets, namely, if for two different values $x_1, x_2 \in \mathcal{X}$ the distribution satisfies $f(x_1) = f(x_2)$ then $\sum_{i=1}^{n} \lambda_i T_i(x_1) = \sum_{i=1}^{n} \lambda_i T_i(x_2)$ (this does not mean that $T_i(x_1)$ and $T_i(x_2)$ must be equal). Moreover, ϕ' must be increasing, thus, necessarily, $\sum_{i=1}^{n} \lambda_i T_i(x)$ and $f(x)$ must have the same sense of variation. If both previous conditions are satisfied, eq. (5) rewrites

$$\phi'(y) = \lambda_0 + \sum_{i=1}^{n} \lambda_i T_i\big(f^{-1}(y)\big)\,, \tag{7}$$

where f^{-1} can be multivalued; in such a situation, ϕ' remains well defined. Note again that for given T_i and f, the solution is not unique due to parameters λ_i, which can be chosen finely so as to simplify the expression of ϕ'. Eq. (7) can then be integrated, at least formally, to achieve H_ϕ (and thus any $H_{h,\phi}$ entropy with nondecreasing h).

For instance, in the case $d = 1$, for one moment constraint, if λ_1 is negative, then

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

3. State-dependent entropy functionals

Of course, the preceding derivations require that (5) is effectively solvable. In addition, one has also to choose or design specific $T_i(X)$ statistics, as well as the parameters λ_i so as to respect the symmetries of the problem. In the examples above, for $d = 1$, we used $T_1(x) = x$, and thus f_X must be monotone. Similarly the choice $T_1(x) = x^2$ or $|x|$ obviously lead to symmetrical densities as already mentioned.

For nonsymmetrical densities for instance, the situation can be more involved. For instance, if we take $T_1(x) = x$, then, on $\mathcal{X} = \mathbb{R}$, eq. (5) has no solution.

To intent to overcome such limitations, a natural way a making should be to extend the ϕ-entropy class by letting function ϕ to be a function of both the state and of the probability distribution:

Definition 4 (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 strictly convex differentiable 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.

Similarly to the classical case, a generalized Bregman divergence can be associated to ϕ under the form:

Definition 5 (Generalized Bregman divergence)**.** With the same assumption in definition 4, the generalized Bregman divergence associated with ϕ defined on $\mathcal{X}\times\mathcal{Y}$ where $\mathcal{Y}$ is closed convex is given by

$$D_\phi(x,y_1,y_2)=\phi(x,y_1)-\phi(x,y_2)-\frac{\partial\phi}{\partial y}(x,y_2)\,(y_1-y_2)\,. \tag{10}$$

A direct consequence of the strict convexity of $\phi(x,\cdot)$ is the nonnegativity of the Bregman divergence: $D_\phi(x,y_1,y_2)\geq 0$ with equality if and only if $y_1=y_2$ (whatever $x\in\mathcal{X}$).

Consider the modified problem of maximizing entropy (8) subject to constraints on some moments $\mathbb{E}\left[T_i(X)\right]$,

$$\begin{cases}\max_f & -\displaystyle\int_{\mathcal{X}}\phi(x,f(x))\mathrm{d}\mu(x)\\ \text{s.t.} & \displaystyle\int_{\mathcal{X}}f(x)\mathrm{d}\mu(x)=1\\ \text{s.t.} & \mathbb{E}\left[T_i(X)\right]=t_i,\quad i=1,\ldots,n\end{cases} \tag{11}$$

Proposition 2 (Maximum state-dependent ϕ-entropy solution)**.** *The probability distribution f solution of the maximum entropy problem* (11) *satisfies the equation*

$$\frac{\partial\phi}{\partial y}\big(x,f(x)\big)=\lambda_0+\sum_{i=1}^{n}\lambda_i\,T_i(x), \tag{12}$$

where parameters λ_i are such that the constraints (normalization, moments) are satisfied.
If ϕ is chosen in the $(\mathcal{X}_1,\ldots,\mathcal{X}_k)$-multiform ϕ-entropy class, this equation writes

$$\sum_{l=1}^{k}\phi_l'\big(f(x)\big)\,\mathbb{1}_{\mathcal{X}_l}(x)=\lambda_0+\sum_{i=1}^{n}\lambda_i\,T_i(x), \tag{13}$$

Proof. The proof is similar to that of Proposition 1, for instance using the generalized functional Bregman divergence

$$\mathcal{D}_\phi(f_1,f_2)=\int_{\mathcal{X}}\phi(x,f_1(x))\mathrm{d}\mu(x)-\int_{\mathcal{X}}\phi(x,f_2(x))\mathrm{d}\mu(x)-\int_{\mathcal{X}}\frac{\partial\phi}{\partial y}(x,f_2(x))\,(f_1(x)-f_2(x))\,\mathrm{d}\mu(x).$$

which is nonnegativity and zero if and only if $f_1=f_2$ almost everywhere. □

Solving eq. (12) is difficult in general, but in specific context, its restriction given by eq. (13) is less complicated to be solved.

3.1. Concave entropy with partial moments constraints

Let us consider

- a partition $(\mathcal{X}_1,\ldots,\mathcal{X}_k)$ of $\mathcal{X}$ and
- an entropic functional ϕ under the form eq. (9) (i.e., k functions ϕ_l)
- partial moment constraints defined over each set $\mathcal{X}_l$, $\mathbb{E}\left[T_{l,i}(X)\mathbb{1}_{\mathcal{X}_l}(X)\right]=t_{l,i}$, $l=1,\ldots,k$ and $i=1,\ldots,n_l$ (by convention, this set is empty if $n_l=0$).

Thus, the probability distribution f solution of the maximum $(\mathcal{X}_1,\ldots,\mathcal{X}_k)$-multiform ϕ entropy problem (11) given by eq. (13) rewrites

$$\sum_{l=1}^{k}\left(\phi_l'\big(f(x)\big)-\lambda_0-\sum_{i=1}^{n_l}\lambda_{l,i}\,T_{l,i}(x)\right)\mathbb{1}_{\mathcal{X}_l}(x)=0 \tag{14}$$

where $\sum_{i=1}^{0}$ is empty (or zero) by convention. This equation is solvable with convex function ϕ_l if and only if over $\mathcal{X}_l$ distribution f and $\lambda_0 + \sum_{i=1}^{n_l} \lambda_{l,i}\, T_{l,i}(x)$ share the same isoprobability subsets and sense of variation.

The identification of a multiform extension of function ϕ such that a given $f(x)$ is the associated maximum multiform entropy distribution generalizes following the steps

1. define a partition $(\mathcal{X}_1, \ldots, \mathcal{X}_k)$ and denote by f_l the restriction of f to $\mathcal{X}_l$,
2. in each domain $\mathcal{X}_l$ choose n_l sets of functions $T_{l,i}$ and choose parameters λ_0 and $\lambda_{l,i}$ such that $\sum_{i=1}^{n_l} \lambda_{l,i} T_{l,i}(x)$, has the same isoprobability subsets and sense of variation than f_l on $\mathcal{X}_l$
3. define
$$\phi_l'(y) = \lambda_0 + \sum_{i=1}^{n_l} \lambda_{l,i}\, T_{l,i}\big(f_l^{-1}(y)\big) \tag{15}$$
4. integrate the results to get $\phi_l(y) = \int \phi_l'(y) \mathrm{d}y$.

For instance, in the univariate context $d = 1$, the partition can be chosen such that the f_l are monotonous and the partial constraints adequately to be able to determine the ϕ_l.

An advantage of this approach is that it allows to view any distribution as a maximum entropy distribution subjected to simple constraints since, although defined via a multiform entropic functional ϕ, the concavity of the ϕ-entropy is preserved. Moreover, the classical case is naturally included in this extension.

The major drawback of this approach is that in general constraints must be specified on subsets $\mathcal{X}_l$ and not on the whole domain of definition $\mathcal{X}$ of f. This is somewhat unnatural, even if practically such partial moments can be estimated by thresholding properly the data.

3.2. Extremum entropy with uniform moments constraints

An alternative to the previous approach should be to preserve the definition of constraints over the whole domain of definition $\mathcal{X}$ of f, adapting then the ϕ_l to each domain where f is monotone. The consequence of this way of making is that the concavity of the ϕ-entropy we will derive is lost.

To be clearer, let us again consider a multiform ϕ-entropy as in eq. (9), but relaxing the concavity of the ϕ_l. The ϕ-entropy is then not necessarily concave. To distinguish this case to the previous one, we will use the notations $\widetilde{\phi}$ and $\widetilde{\phi}_l$ instead of ϕ and ϕ_l. Thus, it is no more possible to interpret a distribution as a maximal entropy when this last one turns to be not concave. However, by the Lagrange technique, we can achieve an *extremal entropy* that can be either a maximum, or a minimum, or a saddle-point. Let us denote by ext such an extremal distribution, thus, problem 11 rewrites

$$\begin{cases} \mathrm{ext}_f & \left(- \sum_{l=1}^{k} \int_{\mathcal{X}_l} \widetilde{\phi}_l(f(x)) \mathrm{d}\mu(x) \right) \\ \text{s.t.} & \int_{\mathcal{X}} f(x) \mathrm{d}\mu(x) = 1 \\ \text{s.t.} & \mathbb{E}\left[T_i(X)\right] = t_i, \quad i = 1, \ldots, n \end{cases} \tag{16}$$

where the solution takes the same form than (14), i.e.,

$$\sum_{l=1}^{k} \left(\widetilde{\phi}_l'\big(f(x)\big) - \lambda_0 + \sum_{i=1}^{n} \lambda_i\, T_i(x) \right) \mathbb{1}_{\mathcal{X}_l}(x) = 0 \tag{17}$$

where both functions T_i and the Lagrange multiplier λ_i are the same for any domain $\mathcal{X}_l$ of the partition.

The identification of a multivalued function $\widetilde{\phi}$ such that a given f is an associated extremal entropy distribution generalizes again following the steps

1. define a partition $(\mathcal{X}_1, \ldots, \mathcal{X}_k)$ of $\mathcal{X}$ and denote by f_l the restriction of f on $\mathcal{X}_l$,
2. choose n functions T_i defined over $\mathcal{X}$ such that $T_i(f_l^{-1}(y))$ is defined unambiguously, and multipliers $\lambda_i, i = 0, \ldots, n$.
3. define
$$\widetilde{\phi}_l'(y) = \lambda_0 + \sum_{i=1}^{n} \lambda_i\, T_i\big(f_l^{-1}(y)\big) \tag{18}$$

4. integrate the results to get $\widetilde{\phi}_l(y) = \displaystyle\int \widetilde{\phi}_l'(y)\mathrm{d}y$.

Note that here the same λ_i are linked to any $\widetilde{\phi}_l$. Now, the $\widetilde{\phi}_l'$ are not imposed to be increasing, thus

- if in $\mathcal{X}_l$, f_X and $\sum_i \lambda_i T_i$ share the same isoprobability subsets and sense of variation, $\widetilde{\phi}_l$ is convex,
- if in $\mathcal{X}_l$, f_X and $\sum_i \lambda_i T_i$ share the same isoprobability subsets but have opposite sense of variation, $\widetilde{\phi}_l$ is concave,
- otherwise $\widetilde{\phi}_l$ is neither convex, nor concave.

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

In this section, we associate a specific generalized escort distribution, moments and Fisher informations associate to ϕ-entropies. More than that, by the mean for these extended quantities, we show that generalizations of the celebrated Cramér-Rao inequalities hold. In some particular context, the lower bound of the inequalities are saturated precisely by maximal ϕ-entropy distributions.

Definition 6 (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

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

we define by

$$E_{\phi,f}(x) = \frac{1}{C_\phi[f] \, \dfrac{\partial^2 \phi}{\partial y^2}(x, f(x))} \tag{19}$$

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.

Definition 7 ((α,ϕ)-moments)**.** Under the assumptions of definition 6, with $\mathcal{X}$ equiped with a norm $\|.\|_{\mathcal{X}}$, we define by

$$M_{\alpha,\phi}[f;X] = \int_{\mathcal{X}} \|x\|_{\mathcal{X}}^{\,\alpha} \; E_{\phi,f}(x)\,\mathrm{d}\mu(x) \tag{20}$$

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

In the context of the Shannon entropy, $\phi(x,y) = \phi(y) = y\log y$, the ϕ-escort density associated to f being density f, the (α,ϕ)-moments are the usual moments of $\|X\|_{\mathcal{X}}^{\alpha}$.

Definition 8 (Nonparametric (β,ϕ)-Fisher information)**.** With the same assumption as in definition 7, denoting by $\|.\|_{\mathcal{X}*}$ the dual norm, for any differentiable density f, we define the quantity

$$I_{\beta,\phi}[f] = \int_{\mathcal{X}} \left\| \frac{\nabla_x f(x)}{E_{\phi,f}(x)} \right\|_{\mathcal{X}*}^{\beta} E_{\phi,f}(x)\,\mathrm{d}\mu(x) \tag{21}$$

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 9 (Parametric (β,ϕ)-Fisher information)**.** Let consider the same assumption as in definition 7, 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

$$I_{\beta,\phi}[f;\theta] = \int_{\mathcal{X}} \left\| \frac{\nabla_\theta f(x)}{E_{\phi,f}(x)} \right\|_{\Theta*}^{\beta} E_{\phi,f}(x)\,\mathrm{d}\mu(x) \tag{22}$$

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, $\phi(x,y) = y \log y$, 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.

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*

$$M_{\alpha,\phi}[f;X]^{\frac{1}{\alpha}} \, I_{\alpha^*,\phi}[f]^{\frac{1}{\alpha^*}} \;\geq\; d \tag{23}$$

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 $\phi(x,y) = y \log y$, dealing with euclidean norms and $\alpha = 2$.

Proof. The approach follows [**?**], 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 [**?** , 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 [**?** , Lemma 2] the equality in is obtain 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. (5) of proposition prop. 1. ☐

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 9. 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*

$$M_{\alpha,\phi}\Big[f; \widehat{\theta}(X) - \theta\Big]^{\frac{1}{\alpha}} \;\; I_{\alpha^*,\phi}[f;\theta]^{\frac{1}{\alpha^*}} \;\geq\; \big| \, m \, + \, \nabla_\theta^t \, b(\theta) \, \big| \tag{24}$$

where

$$b(\theta) = \mathbb{E}\left[\widehat{\theta}(X) - \theta \right] \tag{25}$$

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 is recovered in the usual Shannon context $\phi(x,y) = y \log y$, dealing with euclidean norms and $\alpha = 2$.

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

5. Some examples

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

5.1. 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

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

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.

5.2. 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 [? ? ? ?].

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

5.3. 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.

5.4. 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

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

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.

5.4.1. Second order moment constraint

In this case, immediately

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

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

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

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.

5.4.2. (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

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

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,

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

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

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

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.

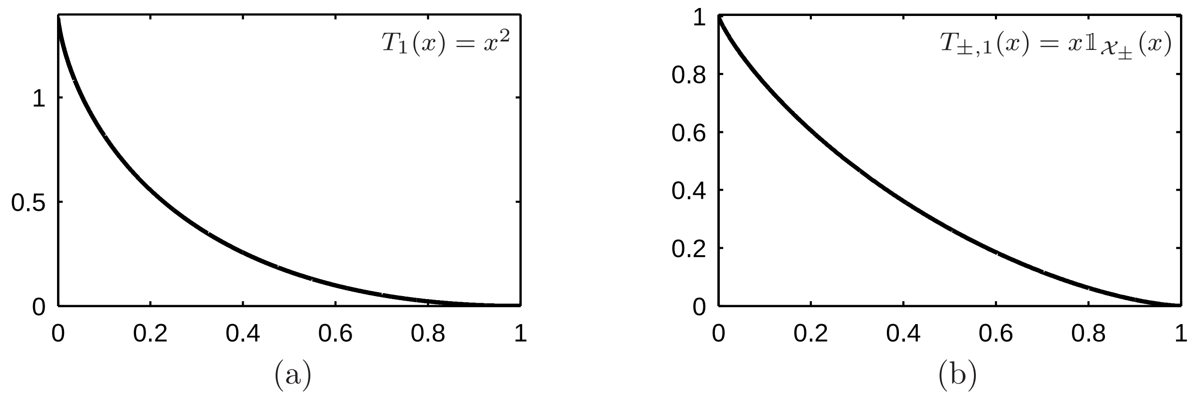

Figure 1: Entropy functional ϕ_u derived from the logistic distribution: (a) with $T_1(x) = x^2$ and (b) with $T_{\pm,1}(x) = x \mathbb{1}_{\mathcal{X}_\pm}(x)$.

5.5. 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

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

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

5.5.1. Second order moment

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

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

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

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

5.5.2. (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

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

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.

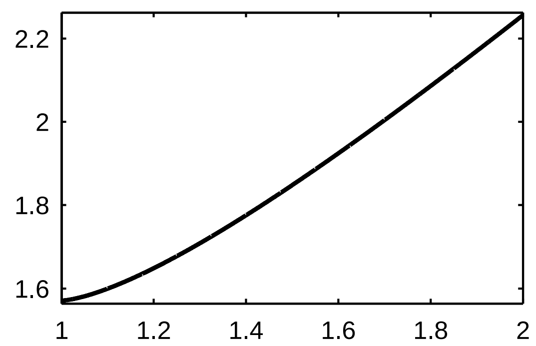

Figure 2: Entropy functional ϕ_u derived from the arcsine distribution with partial constraints $T_{\pm,1}(x) = x \mathbb{1}_{\mathcal{X}_\pm}(x)$.

5.6. 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

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

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 [**?** , eq. 13.1.4] and [**?** , 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 $c_{-1} = \lambda_0 = 0, \lambda_{0,1} = 1, \lambda_{-1,1} = -0.1$.

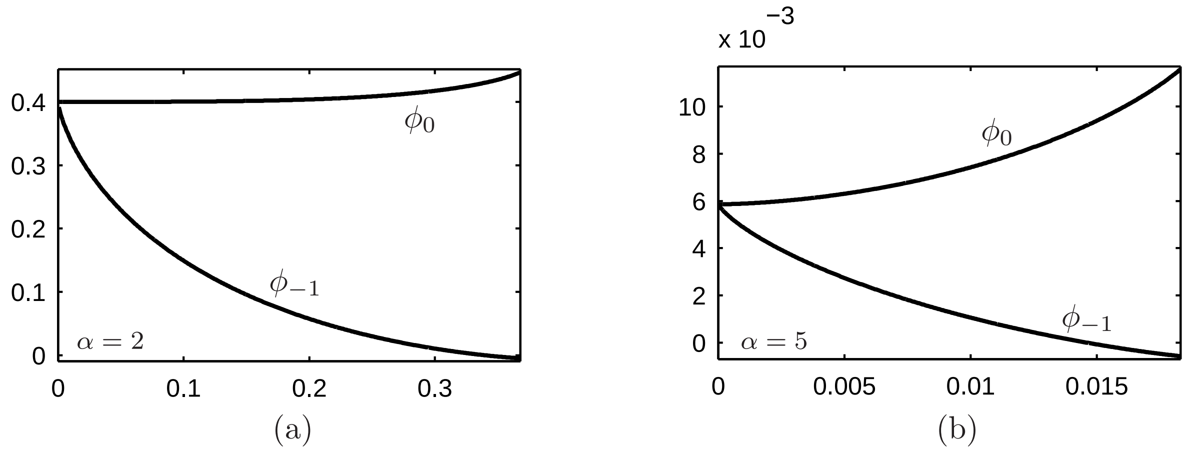

Figure 3: Multiform entropy functional ϕ_u derived from the gamma distribution with the partial moment constraints $T_{k,1}(x) = x^2 \mathbb{1}_{\mathcal{X}_k}(x)$, $k \in \{0, -1\}$. (a): $\alpha = 2$; (b): $\alpha = 5$.