T'aurais pas une entropie?

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

Let f be a probability distribution 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

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

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.

2. Maximum (h, ϕ)-entropy distributions

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

Definition 1. Let $\phi : \mathcal{Y} \subset \mathbb{R}_+ \mapsto \mathbb{R}$ 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 $\mathcal{X}$ such that $f(\mathcal{X}) \subset \mathcal{Y}$,

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

is the ϕ-entropy of f. Since $\phi(x)$ 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. 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. With the same assumption in definition 1, the Bregman divergence associated with ϕ defined on a closed convex set $\mathcal{Y}$, is given by

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

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 $y_1 = y_2$.

Consider the problem of maximizing entropy (2) subject to constraints on some moments $\mathbb{E}\,[T_i(X)]$. 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),

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

Proposition 1. *The probability distribution f_X solution of the maximum entropy problem* (4) *satisfies the equation*

$$\phi'\big(f_X(x;t)\big) = \lambda_0 + \sum_{i=1}^{n} \lambda_i\, T_i(x), \tag{5}$$

where parameters λ_i are such that the constraints (normalization, moments) are satisfied.

Proof. The maximization problem being concave, the solution exists and is unique. Equation (5) results directly from the classical Lagrange multipliers technique [**?**]. 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 with given (fixed) moments $\mathbb{E}\,[T_i(X)]$. To this end, consider the functional Bregman divergence acting on functions defined on a common domain $\mathcal{X}$:

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

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

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

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

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

where we used the fact that f and f_X 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

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

for all distribution f with the same moments t than f_X, with equality if and only if $f = f_X$ almost everywhere. In other words, this shows that f_X, solution of (5), realizes the maximum of $H_\phi[f]$ 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. Therefore, 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.

Since we will look for the function ϕ for a given probability distribution $f_X(x)$ we also see that the corresponding λ 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 $f_X(\mathcal{X})$;
- From the strict convexity property of ϕ, necessarily ϕ' is increasing.

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

1. choose a set of functions $T_i(x)$, $i = 1, \ldots, n$,
2. find ϕ' satisfying $\lambda_0 + \sum_{i=1}^{n} \lambda_i T_i(x) = \phi'(f_X(x))$,
3. integrate the result 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.

Remind that ϕ' must be increasing, thus, necessarily, $\sum_{i=1}^{n} \lambda_i T_i(x)$ and $f_X(x)$ must have the same sense of variation. Moreover, at a first glance, eq. (5) requires these two quantities to share the same symmetries. Namely, if for two different values $x_1, x_2 \in \mathcal{X}$ the distribution satisfies $f_X(x_1) = f_X(x_2)$ 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). Thus, eq. (5) rewrites

$$\phi'(y) = \lambda_0 + \sum_{i=1}^{n} \lambda_i T_i\big(f_X^{-1}(y)\big)\,, \tag{6}$$

where f_X^{-1} can be multivalued; in such a situation, ϕ' remains well defined. Note again that for given T_i and f_X, the solution is not unique due to parameters λ_i, which can be chosen finely so as to simplify the expression of ϕ'. Eq. (6) can then be integrated, at least formally, to achieve H_ϕ (and thus any $H_{h,\phi}$ entropy with nondecreasing h).

For instance, for one moment constraint, if λ_1 is negative, then

- for $T_1(x) = x,\ f_X(x)$ must be decreasing,
- for $T_1(x) = x^2$ or $T_1(x) = |x|$, if $\mathcal{X} = \mathbb{R}$, $f_X(x)$ 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, 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 this limitation, 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. Let $\phi : \mathcal{X} \times \mathcal{Y} \mapsto \mathbb{R}$ such that for any x, $\phi(x, \cdot)$ is a strictly convex differentiable function 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 set $\mathcal{X}$ and such that $f(\mathcal{X}) \subset \mathcal{Y}$,

$$H_\phi[f] = -\int_{\mathcal{X}} \phi(x, f(x))\mathrm{d}\mu(x) \tag{7}$$

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

$$\phi(x, y) = \sum_{l=1}^{k} \phi_l(y) \mathbb{1}_{\mathcal{X}_l}(x) \tag{8}$$

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. 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

$$D_\phi(x, y_1, y_2) = \phi(x, y_1) - \phi(x, y_2) - \phi'(x, y_2)\,(y_1 - y_2)\,. \tag{9}$$

where, by misuse of writing, ϕ' denotes the partial derivative versus the second argument of ϕ, i.e.,

$$\phi'(x,y) = \frac{\partial \phi}{\partial y}(x,y).$$

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 (7) subject to constraints on some moments $\mathbb{E}\left[T_i(X)\right]$,

$$\begin{cases} \max_f & -\displaystyle\int_{\mathcal{X}} \phi(x, f(x)) \mathrm{d}\mu(x) \\ \text{s.t.} & \displaystyle\int_{\mathcal{X}} f(x) \mathrm{d}\mu(x) = 1 \\ \text{s.t.} & \mathbb{E}\left[T_i(X)\right] = t_i, \quad i = 1, \ldots, n \end{cases} \tag{10}$$

Proposition 2. *The probability distribution f_X solution of the maximum entropy problem* (10) *satisfies the equation*

$$\phi'\big(x, f_X(x;t)\big) = \lambda_0 + \sum_{i=1}^{n} \lambda_i \, T_i(x), \tag{11}$$

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

$$\sum_{l=1}^{k} \phi_l'\big(f_X(x;t)\big) \, \mathbb{1}_{\mathcal{X}_l}(x) = \lambda_0 + \sum_{i=1}^{n} \lambda_i \, T_i(x), \tag{12}$$

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

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

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

Solving eq. (11) is difficult in general, but in specific context, its restriction given by eq. (12) 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. (8) (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_X solution of the maximum $(\mathcal{X}_1, \ldots, \mathcal{X}_k)$-multiform ϕ entropy problem (10) given by eq. (12) rewrites

$$\sum_{l=1}^{k} \left(\phi_l'\big(f_X(x)\big) - \lambda_0 - \sum_{i=1}^{n_l} \lambda_{l,i}\, T_{l,i}(x) \right) \mathbb{1}_{\mathcal{X}_l}(x) = 0 \tag{13}$$

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_X and $\lambda_0 + \sum_{i=1}^{n_l} \lambda_{l,i}\, T_{l,i}(x)$ share the same sense of variation.

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

1. define a partition $(\mathcal{X}_1, \ldots, \mathcal{X}_k)$ of $\mathcal{X}$ such that f_X is monotonous in each $\mathcal{X}_l$, denoting by $f_{X,l}$ the restriction of f_X to $\mathcal{X}_l$,
2. in each domain $\mathcal{X}_l$ choose n_l sets of functions $T_{l,i}(x)$ 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 sens of variation than f_X on $\mathcal{X}_l$
3. define
$$\phi_l'(y) = \lambda_0 + \sum_{i=1}^{n_l} \lambda_{l,i}\, T_{l,i}\Big(f_{X,l}^{-1}(y)\Big) \tag{14}$$
4. integrate the results to get $\phi_l(y) = \int \phi_l'(y)\mathrm{d}y$.

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_X. 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_X, adapting then the ϕ_l to each domain where f_X 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. (8), 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 notation $\widetilde{\phi}_l$ instead of 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 10 rewrites

$$\begin{cases} \operatorname{ext}_f & \left(-\sum_{l=1}^{k} \int_{\mathcal{X}_l} \widetilde{\phi}_l(f(x)) \mathrm{d}\mu(x) \right) \\ \text{s.t.} & \int_{\mathcal{X}} f(x) \mathrm{d}\mu(x) = 1 \\ \text{s.t.} & \mathbb{E}\left[T_i(X)\right] = t_i, \quad i = 1, \ldots, n \end{cases} \tag{15}$$

where the solution takes the same form than (13), i.e.,

$$\sum_{l=1}^{k} \left(\widetilde{\phi}_l'\big(f_X(x)\big) - \lambda_0 + \sum_{i=1}^{n} \lambda_i \, T_i(x) \right) \mathbb{1}_{\mathcal{X}_l}(x) = 0 \tag{16}$$

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 $f_X(x)$ is an associated extremal entropy distribution generalizes again following the steps

1. define a partition $(\mathcal{X}_1, \ldots, \mathcal{X}_k)$ of $\mathcal{X}$ such that f_X is monotonous in each $\mathcal{X}_l$,
2. choose n functions $T_i(x)$ defined over $\mathcal{X}$ and multipliers $\lambda_0 \lambda_i, i = 1, \ldots, n$.
3. define
$$\widetilde{\phi}_l'(y) = \lambda_0 + \sum_{i=1}^{n} \lambda_i \, T_i\Big(f_{X,l}^{-1}(y)\Big) \tag{17}$$
4. integrate the results to get $\widetilde{\phi}_l(y) = \int \widetilde{\phi}_l'(y) \mathrm{d}y$.

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

4. ϕ-escort, ϕ-Fisher information and generalized Cramér-Rao inequality

In this section, we associate a specific generalized Fisher information with the ϕ-entropies. More than that, we show that a generalization of the celebrated Cramér-Rao inequality holds, and that the maximum ϕ-entropy distribution precisely saturates the inequality. Finally, we define a notion of ϕ-escort distribution which enable to derive an estimation theoretic version of the Cramér-Rao inequality.

Define $\diamondsuit(f)$ as the function such that

$$\dot{\diamondsuit}(f) = f \dot{\phi}(f),$$

or

$$\diamondsuit(f) = \int_0^f t \ddot{\phi}(t) \mathrm{d}t.$$

It is understood that in the case where ϕ is multiform, then so is $\diamondsuit(f)$. We assume that $f\diamondsuit(f)$ decreases to zero on the boundaries $\mathcal{B}$ of the domain $\mathcal{D}$ of f. In such case, by integration by part, we have

$$\begin{aligned}\int_{\mathcal{D}} \diamondsuit(f)\mathrm{d}x &= [x\diamondsuit(f)]_{\mathcal{B}} - \int_{\mathcal{D}} x\dot{\diamondsuit}(f)\mathrm{d}x,\\ &= -\int_{\mathcal{D}} x\dot{\diamondsuit}(f)\mathrm{d}x = -\int_{\mathcal{D}} x\frac{\dot{\diamondsuit}(f)}{f(x)}f(x)\,\mathrm{d}x.\end{aligned}$$

By the Hölder inequality applied to the last integral, we obtain

$$\int_{\mathcal{D}} \diamondsuit(f)\mathrm{d}x \leq \left(\int_{\mathcal{D}} |x|^{\alpha}\, f(x)\,\mathrm{d}x\right)^{\frac{1}{\alpha}} \left(\int_{\mathcal{D}} \left|\frac{\dot{\diamondsuit}(f)}{f(x)}\right|^{\beta} f(x)\,\mathrm{d}x\right)^{\frac{1}{\beta}}$$

where $1/\alpha + 1/\beta = 1$. The inequality can also be rewritten as

$$\left(\int_{\mathcal{D}} |x|^{\alpha}\, f(x)\,\mathrm{d}x\right)^{\frac{1}{\alpha}} \frac{\left(\int_{\mathcal{D}} \left|\frac{\dot{f}(x)}{f(x)}\dot{\diamondsuit}(f)\right|^{\beta} f(x)\,\mathrm{d}x\right)^{\frac{1}{\beta}}}{\int_{\mathcal{D}} \diamondsuit(f)\mathrm{d}x} \geq 1.$$

This has the form of a generalized Cramér-Rao inequality, where the first term is the moment of order α and the second one is a generalized ϕ-Fisher information of order β.

A Cramér-Rao inequality for the estimation of a parameter

The problem of estimation is to determine a function $\hat{\theta}(x)$ in order to estimate an unknown parameter θ. Let $f(x;\theta)$ and $g(x;\theta)$ be two probability density functions, with $x \in X \subseteq \mathbb{R}^k$ and θ a parameter of these densities, $\theta \in \mathbb{R}^n$. An underlying idea in the statement of the new Cramér-Rao inequality is that it is possible to evaluate the moments of the error with respect to different probability distributions. For instance, in the estimation setting the estimation error is $\hat{\theta}(x) - \theta$. The bias can be evaluated with respect to f according to

$$B_f(\theta) = \int_X \Big(\hat{\theta}(x) - \theta\Big)\; f(x;\theta)\,\mathrm{d}x = \mathrm{E}_f\left[\hat{\theta}(x) - \theta\right] \tag{18}$$

Theorem 1. *Let $f(x;\theta)$ be a multivariate probability density function defined over a subset $X \subseteq \mathbb{R}^n$, and $\theta \in \Theta \subseteq \mathbb{R}^k$ a parameter of the density. The set Θ is equipped with a norm $\|.\|$, and the corresponding dual norm is denoted $\|.\|_*$. Let $g(x;\theta)$ denote another probability density function also defined on $(X;\Theta)$. Assume that $f(x;\theta)$ is a jointly measurable function of x and θ, is integrable with respect to x, is absolutely continuous with respect to θ, and that the derivatives with respect to each component of θ are locally integrable. For any estimator $\hat{\theta}(x)$ of θ, we have*

$$E\left[\left\|\hat{\theta}(x) - \theta\right\|^{\alpha}\right]^{\frac{1}{\alpha}} I_{\beta}[f|g;\theta]^{\frac{1}{\beta}} \geq |n + \nabla_{\theta}.\, B_f(\theta)| \tag{19}$$

with α and β Hölder conjugates of each other, i.e. $\alpha^{-1} + \beta^{-1} = 1$, $\alpha \geq 1$, and where the (β, g)-Fisher information

$$I_{\beta}[f|g;\theta] = \int_X \left\|\frac{\nabla_{\theta} f(x;\theta)}{g(x;\theta)}\right\|_*^{\beta} g(x;\theta)\; dx \tag{20}$$

is the generalized Fisher information of order β on the parameter θ contained in the distribution f and taken with respect to g. The equality case is obtained if

$$\frac{\nabla_{\theta} f(x;\theta)}{g(x;\theta)} = K\left\|\hat{\theta}(x) - \theta\right\|^{\alpha-1} \nabla_{\hat{\theta}(x)-\theta}\|\hat{\theta}(x) - \theta\|, \tag{21}$$

with $K > 0$.

Proof. The bias in (18) is a n-dimensional vector. Let us consider its divergence with respect to variations of θ:

$$\operatorname{div} B_f(\theta) = \nabla_\theta . B_f(\theta). \tag{22}$$

The regularity conditions in the statement of the theorem enable to interchange integration with respect to x and differentiation with respect to θ, and

$$\nabla_\theta . B_f(\theta) = \int_X \nabla_\theta . \left(\hat{\theta}(x) - \theta\right) \, f(x;\theta) \, \mathrm{d}x + \int_X \nabla_\theta f(x;\theta) . \left(\hat{\theta}(x) - \theta\right) \, \mathrm{d}x. \tag{23}$$

In the first term on the right, we have $\nabla_\theta . \theta = n$, and the integral reduces to $-n \int_X f(x;\theta) \, \mathrm{d}x = -n$, since $f(x;\theta)$ is a probability density on X. The second term can be rearranged so as to obtain an integration with respect to the density $g(x;\theta)$, assuming that the derivatives with respect to each component of θ are absolutely continuous with respect to $g(x;\theta)$, i.e. $g(x;\theta) \gg \nabla_\theta f(x;\theta)$. This gives

$$n + \nabla_\theta . B_f(\theta) = \int_X \frac{\nabla_\theta f(x;\theta)}{g(x;\theta)} . \left(\hat{\theta}(x) - \theta\right) \, g(x;\theta) \, \mathrm{d}x. \tag{24}$$

Now, it only remains to apply the generalized Hölder-type inequality (**??**) in Lemma **??** to the integral on the right side, with $X(x) = \hat{\theta}(x) - \theta$, $Y(x) = \frac{\nabla_\theta f(x;\theta)}{g(x;\theta)}$, and $w(x) = g(x;\theta)$. This yields in all generality

$$\left(\int_X \left\| \hat{\theta}(x) - \theta \right\|^\alpha g(x;\theta) \, \mathrm{d}x \right)^{\frac{1}{\alpha}} \left(\int_X \left\| \frac{\nabla_\theta f(x;\theta)}{g(x;\theta)} \right\|_*^\beta g(x;\theta) \, \mathrm{d}x \right)^{\frac{1}{\beta}} \geq |n + \nabla_\theta . B_f(\theta)| \tag{25}$$

which is (19). By Lemma **??** again, we know that the case of equality occurs if $Y(t) = K \|X(t)\|^{\alpha-1} \nabla_{X(t)} \|X(t)\|$, $K > 0$, which gives (21). □

5. Some examples

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 $q > 0$ and $x_+ = \max(x, 0)$, 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 can be

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

It is nothing but Tsallis entropy.

5.3. q-exponential distribution and first-order moment

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

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

It suffices to follow the very same steps as above, leading again to the Tsallis entropy.

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

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

for $y \in \left[0\,;\,\frac{1}{4s}\right]$. The simple choice

$$\lambda_0 = 0 \qquad \text{and} \qquad \lambda_1 = -\frac{1}{s}$$

gives then

$$\phi(y) = \left(-4sy\left(\operatorname{argtanh}\sqrt{1-4sy}\right)^2 + 2\sqrt{1-4sy}\,\operatorname{argtanh}\sqrt{1-4sy} + \log(4sy)\right)\mathbb{1}_{\left[0\,;\,\frac{1}{4s}\right]}(y).$$

Figure 1 depicts this function ϕ for $s = 1$.

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

$$\phi'_\pm(y) = \lambda_0 + 2s\,\lambda_{\pm,1}\operatorname{argtanh}\sqrt{1-4sy} \qquad \text{or} \qquad \widetilde{\phi}'_\pm(y) = \lambda_0 \pm 2s\,\lambda_1\operatorname{argtanh}\sqrt{1-4sy}$$

where the sign is absorbed on $\lambda_{\pm,1}$. A judicious choice is then

$$\lambda_0 = 0 \qquad \text{and} \qquad \lambda_{\pm,1} = -1 \qquad \text{or} \qquad \lambda_1 = -1$$

leading either to the (convex) uniform function ϕ,

$$\phi(y) = \left(\frac{1}{2}\sqrt{1-4sy} - 2sy\operatorname{argtanh}\sqrt{1-4sy}\right)\mathbb{1}_{\left[0\,;\,\frac{1}{4s}\right]}(y)$$

or to the multiform function ϕ with branches $\widetilde{\phi}_\pm$,

$$\widetilde{\phi}_\pm(y) = \pm\left(\frac{1}{2}\sqrt{1-4sy} - 2sy\operatorname{argtanh}\sqrt{1-4sy}\right)\mathbb{1}_{\left[0\,;\,\frac{1}{4s}\right]}(y)$$

Function ϕ is represented figure 2 for $s = 1$ (here, $\widetilde{\phi}_\pm(y) = \pm\phi(y)$). The choice of $\lambda_{\pm,1}$ allows in a sense to respect the symmetries of the distribution, allowing thus to recover a classical ϕ-entropy.

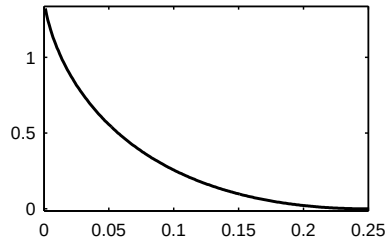

Figure 1: Entropy functional ϕ derived from the logistic distribution with $T_1(x) = x^2$.

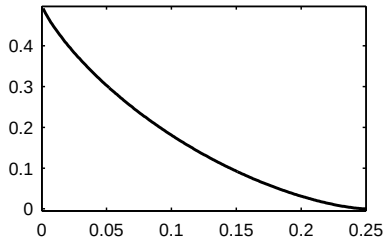

Figure 2: Entropy functional ϕ ($\widetilde{\phi}_\pm = \pm\phi$) derived from the logistic distribution with either partial moments $T_{\pm,1}(x) = x\mathbb{1}_{\mathcal{X}_\pm}(x)$, or global moment $T_1(x) = x$.

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 $f^{-1}_{X,\pm}$ on $\mathcal{X}_- = (-\sigma\sqrt{2}\,;\,0)$ and $X_+ = [0\,;\,\sigma\sqrt{2})$ write then

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

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

5.5.1. Second order moment

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

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

With the special choice

$$\lambda_0 = 0 \qquad \text{and} \qquad \lambda_1 = 1$$

the entropy functional is then

$$\phi(y) = \left(2\sigma^2y + \frac{1}{\pi^2y}\right)\mathbb{1}_{\left[\frac{1}{\pi\sigma\sqrt{2}}\,;\,+\infty\right)}(y)$$

which is represented figure 3 for $\sigma = 1$.

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

$$T_{\pm,1}(x) = x \quad \text{over} \quad \mathcal{X}_- = \left(-\sigma\sqrt{2}\,;\,0\right), \quad \text{and} \quad \mathcal{X}_+ = \left[0\,;\,\sigma\sqrt{2}\right).$$

now

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

where the sign is absorbed in the factors $\lambda_{\pm,1}$. A judicious choice can be

$$\lambda_0 = 0 \qquad \text{and} \qquad \lambda_{\pm,1} = 1 \qquad \text{or} \qquad \lambda_1 = 1$$

leading then either to the (convex) uniform function

$$\phi(y) = \left(\frac{1}{\pi}\sqrt{2\pi^2\sigma^2y^2-1} + \frac{1}{\pi}\arctan\left(\frac{1}{\sqrt{2\pi^2\sigma^2y^2-1}}\right)\right)\mathbb{1}_{\left[\frac{1}{\pi\sigma\sqrt{2}}\,;\,+\infty\right)}(y)$$

or to the multiform function with branches $\widetilde{\phi}_{\pm}$,

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

The uniform function is represented figure 4 for $\sigma = 1$ (here again $\widetilde{\phi}_{\pm}(y) = \pm\phi(y)$).

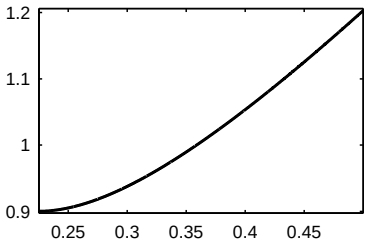

Figure 3: Entropy functional ϕ derived from the centered and scaled arcsine distribution with constraint $T_1 = x^2$.

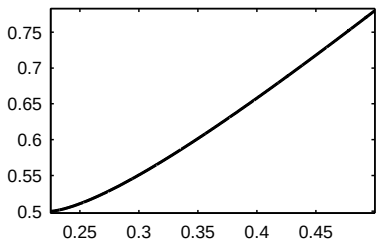

Figure 4: Entropy functional ϕ ($\widetilde{\phi}_{\pm} = \pm\phi$) derived from the arcsine distribution either with partial constraints $T_{\pm,1} = x\mathbb{1}_{\mathcal{X}_{\pm}}(x)$, or with global constraint $T_1(x) = x$.

5.6. The gamma distribution and (partial) first-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 $x = \frac{\alpha-1}{\beta}$ and $f_X(\mathbb{R}_+) = \left[0\,;\,\frac{\beta}{\Gamma(\alpha)}\left(\frac{\alpha-1}{e}\right)^{\alpha-1}\right]$. Thus, here again it cannot be viewed as a maximum entropy constraint neither by the first order moment, nor by the second order moment. Here, we can again interpret it as a maximum entropy constrained by partial moments

$$T_{0,1}(x) = x^i \quad \text{over} \quad \mathcal{X}_0 = \left[0\,;\,\frac{\alpha-1}{\beta}\right), \quad \text{and} \quad T_{-1,1}(x) = x^i \quad \text{over} \quad \mathcal{X}_{-1} = \left[\frac{\alpha-1}{\beta}\,;\,+\infty\right).$$

or as an extremal entropy constrained by the moment

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

where $i = 1$ or $i = 2$. Inverting $y = f_X(x)$ leads to the equation

$$\frac{\beta x}{1-\alpha}\exp\left(\frac{\beta x}{1-\alpha}\right) = -\frac{1}{\alpha-1}\left(\frac{\Gamma(\alpha)y}{\beta}\right)^{\frac{1}{\alpha-1}}$$

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

$$f^{-1}_{X,k}(y) = \frac{\alpha-1}{\beta}\,\mathrm{W}_k\left(-\frac{1}{\alpha-1}\left(\frac{\Gamma(\alpha)y}{\beta}\right)^{\frac{1}{\alpha-1}}\right), \qquad y \in \left[0\,;\,\frac{\beta}{\Gamma(\alpha)}\left(\frac{\alpha-1}{e}\right)^{\alpha-1}\right],$$

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.

5.6.1. (Partial) first-order moment(s)

In the context of the first order moment, $i = 1$, one gets

$$\phi'_k(y) = \lambda_0 + \frac{\alpha-1}{\beta}\lambda_{k,1}\,\mathrm{W}_k\left(-\frac{1}{\alpha-1}\left(\frac{\Gamma(\alpha)y}{\beta}\right)^{\frac{1}{\alpha-1}}\right)$$

and similarly for $\widetilde{\phi}_k$ (with a unique λ_1 instead of the $\lambda_{k,1}$). To design a concave entropy, to respect the sense of variation imposed on $\phi'_k(y)$, one can choose

$$\lambda_0 = 0 \qquad \text{and } \lambda_{k,1} \text{ such that} \qquad (-1)^{k+1}\lambda_{k,1} > 0$$

Indeed, there is no judicious choice allowing to compensate for the asymmetry of f_X and then allowing the definition of an uniform function ϕ. In general, there is no closed form for $\phi_k(y)$. However, when $\alpha = 2, \ldots$ is integer[1], it can be checked that

$$\phi_k(y) = \frac{(-1)^{k+1}\,\overline{\lambda}_{k,1}\,y\displaystyle\sum_{m=0}^{\alpha}\mu_m\left[\mathrm{W}_k\left(-\frac{1}{\alpha-1}\left(\frac{\Gamma(\alpha)y}{\beta}\right)^{\frac{1}{\alpha-1}}\right)\right]^m}{\left[\mathrm{W}_k\left(-\frac{1}{\alpha-1}\left(\frac{\Gamma(\alpha)y}{\beta}\right)^{\frac{1}{\alpha-1}}\right)\right]^{\alpha-1}} + c_k$$

where

$$\mu_m = \frac{(-1)^{\alpha-m}\,\Gamma(\alpha)}{\Gamma(m+1)\,(\alpha-1)^{\alpha-m}}, \quad m = 0, \ldots, \alpha-1, \quad \text{and} \quad \mu_\alpha = 1$$

c_k is an integration constant and the multiplicative factor is absorbed in the strictly positive $\overline{\lambda}_{k,1}$. $\mathcal{X}_{-1}$ being unbounded, c_{-1} is chosen to be zero and c_0 can be chosen such that $\phi_k\left(\frac{\beta}{\Gamma(\alpha)}\left(\frac{\alpha-1}{e}\right)^{\alpha-1}\right)$ coincide for instance. The same algebra leads to the same expression for the $\widetilde{\phi}_k$, except that $(-1)^{k+1}\lambda_{k,1}$ are replaced by a unique λ_1.

The multivalued function ϕ in the concave context is represented figure 5 for $\beta = 3$, $\alpha = 2$ and $\alpha = 5$, and with the choices $\overline{\lambda}_{k,1} = 1$ (for the other context, with $\overline{\lambda}_1 = -1$, $\widetilde{\phi}_k = (-1)^k\phi_k$.)

[1]Note that in this case, for $\beta = \frac{1}{2}$, f_X is a chi-squared distribution with 2α degrees of freedom

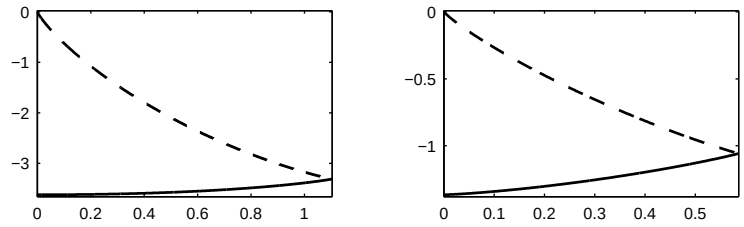

Figure 5: Multiform entropy functional ϕ derived from the gamma distribution with $\beta = 3$, $T_{k,1}(x) = x\mathbb{1}_{\mathcal{X}_k}(x)$, $k \in \{0, -1\}$ (solid line ϕ_0 and dashed line ϕ_{-1}). Left: $\alpha = 2$; Right: $\alpha = 5$.

5.6.2. (Partial) second-order moment(s)

Now, we consider as constraints the (partial) second-order moment(s), $i = 2$. The same approach than in the previous case leads to

$$\phi_k'(y) = \lambda_0 + \left(\frac{\alpha-1}{\beta}\right)^2 \lambda_{k,1} \left[\mathrm{W}_k\left(-\frac{1}{\alpha-1}\left(\frac{\Gamma(\alpha)y}{\beta}\right)^{\frac{1}{\alpha-1}}\right)\right]^2$$

To respect the sense of variation imposed on $\phi_k'(y)$, one can choose again

$$\lambda_0 = 0 \qquad \text{and } \lambda_{k,1} \text{ such that} \qquad (-1)^{k+1}\lambda_{k,1} > 0$$

and when α is integer ($\alpha \geq 2$), it can be checked that

$$\phi_k(y) = \frac{(-1)^{k+1}\,\overline{\lambda}_{k,1}\, y \sum_{m=0}^{\alpha+1} \mu_m \left[\mathrm{W}_k\left(-\frac{1}{\alpha-1}\left(\frac{\Gamma(\alpha)y}{\beta}\right)^{\frac{1}{\alpha-1}}\right)\right]^m}{\left[\mathrm{W}_k\left(-\frac{1}{\alpha-1}\left(\frac{\Gamma(\alpha)y}{\beta}\right)^{\frac{1}{\alpha-1}}\right)\right]^{\alpha-1}} + c_k$$

where

$$\mu_m = \frac{2\,(-1)^{\alpha-m+1}\,\Gamma(\alpha+1)}{\Gamma(m+1)\,(\alpha-1)^{\alpha-m+1}}, \quad m = 0, \ldots, \alpha, \quad \text{and} \quad \mu_{\alpha+1} = 1$$

and c_k is an integration constant and the multiplicative factor is absorbed in the strictly positive $\overline{\lambda}_{k,1}$. Again c_{-1} is chosen to be zero and c_0 can be chosen such that $\phi_k\left(\frac{\beta}{\Gamma(\alpha)}\left(\frac{\alpha-1}{e}\right)^{\alpha-1}\right)$ coincide. This multivalued function is represented figure 6 for $\beta = 3$, $\alpha = 5$ and $\alpha = 10$, and with the choices $\overline{\lambda}_{k,1} = 1$

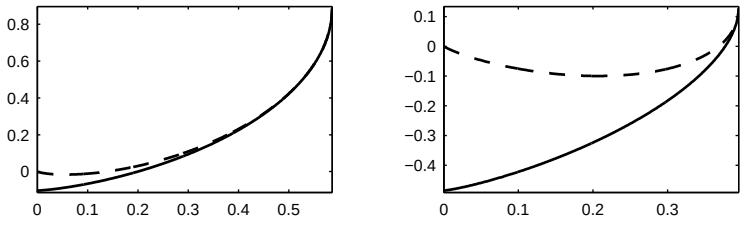

Figure 6: Multiform entropy functional ϕ derived from the gamma distribution with $\beta = 3$, $T_{k,1}(x) = x^2\mathbb{1}_{\mathcal{X}_k}(x)$, $k \in \{0, -1\}$ (solid line ϕ_0 and dashed line ϕ_{-1}). Left: $\alpha = 5$; Right: $\alpha = 10$.

Here again, the approach with a global constraint $T_1(x) = x^2$ over $\mathcal{X} = \mathbb{R}_+$, with the choice $\overline{\lambda}_1 = -1$, leads simply to $\widetilde{\phi}_k = (-1)^k \phi_k$.

Let us consider some specific cases.

1. Let $f_X(x)$ be the hyperbolic nt distribution, with density

$$f_X(x) = \frac{1}{2}\text{sech}(\frac{\pi}{2}x) = \frac{1}{2}\cosh^{-1}(\frac{\pi}{2}x).$$

Obviously, $\frac{\pi}{2}x = \cosh(2y) = \phi'(y)$ with $T(x) = x$, $\lambda = \frac{\pi}{2}$, and

$$\phi(y) = \sinh(2y).$$

So doing, we obtain an hyperbolic sine entropy with the hyperbolic secant distribution as the associated maximum entropy distribution.

document