

ϕ -Informational Measures: Some Results and Interrelations

Steeve Zozor ^{1,*}  and Jean-François Bercher ²¹ GIPSA-Lab, CNRS, Grenoble INP, University Grenoble Alpes, 38000 Grenoble, France² CNRS, LIGM, University Gustave Eiffel, 77454 Marne-la-Vallée, France; jean-francois.bercher@esiee.fr

* Correspondence: steeve.zozor@cnrs.fr

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

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



Citation: Zozor, S.; Bercher, J.-F. ϕ -Informational Measures: Some Results and Interrelations. *Entropy* **2021**, *23*, 911. <https://doi.org/10.3390/e23070911>

Academic Editor: Philip Broadbridge

Received: 26 May 2021

Accepted: 13 July 2021

Published: 18 July 2021

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



Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

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

In the framework of these generalized entropies, the so-called maximum entropy principle takes a special place. This principle, advocated by Jaynes, states that the statistical distribution that describes a system in equilibrium maximizes the entropy while satisfying the system's physical constraints (e.g., the center of mass and energy) [31–35]. In other words, it is the less informative law given the constraints of the system. In the Bayesian approach, dealing with the stochastic modeling of a parameter, such a principle (or a minimum divergence principle) is often used to choose a prior distribution for the parameter [22,36–39]. It also finds its counterpart in communication, clustering, pattern recognition, problems, among many others [32,33,40–43]. In statistics, some goodness-of-fit tests are based on entropic criteria derived from the same idea of constrained maximal entropic law [44–49]. The principle behind such entropic tests lies in the Bregman divergence, measuring a kind of distance between probability distributions, i.e., the empirical distribu-

tion given by data and the distribution we assume for the data (reference). It appears that if the empirical distribution and the reference share the same moments, and if the latter is of maximum entropy with these moments as constraints, the divergence reduces to a difference of entropy. In a large number of works using the maximum entropy principle, the entropy used is the Shannon entropy. However, if for some reason, a generalized entropy is considered, the approach used in the Shannon case does not fundamentally change [50–53].

One can consider the inverse problem which consists in finding the moment constraints leading to the observed distribution as a maximal entropy distribution [50]. Kesavan and Kapur also envisaged a second inverse problem, where both the distribution and the moments are given. The question is thus to determine the entropy so that the distribution is its maximizer. As a matter of fact, dealing with the Shannon entropy, whatever the constraints considered, the maximum entropy distribution falls in the exponential family [33,34,52,54]. Remind that the exponential family is the set of parametric densities (with respect to a measure μ independent on the parameter) of the form $p(x) = C(\theta)h(x) \exp(R(\theta)^t S(x))$ where $S(x)$ is the sufficient statistics [39,55–60]. When $R(\theta) = \theta$, the family is said to be natural and $Z(\theta) = 1/C(\theta)$ is the partition function, the log-partition function $\varphi(\theta) = \log Z(\theta)$ being the cumulants generating function. Now, resolving the maximum entropy problem given later on by Equation (6) in the context of the Shannon entropy, it appears indeed that the maximum entropy distribution falls in the natural exponential family where the sufficient statistics is given by the moment constraints. Considering more general entropies allows to escape from this limitation. Moreover, if the Shannon entropy (or the Gibbs entropy in physics) is well adapted to the study of systems in the equilibrium (or in the thermodynamic limit), extended entropies allow a finer description of systems out of equilibrium [17,61–65], exhibiting their importance. While the problem was considered mainly in the discrete setting by Kesavan and Kapur in [50], we will recall it in the general framework of the ϕ -entropies probability densities with respect to any reference measure, and make a further step considering an extended class of these entropies. Resolving the inverse problem can find applications in goodness-of-fit tests for instance, allowing to design entropies adapted to such tests, in the same line as that of the approaches mentioned above [44–49].

While the entropy is a widely used tool for quantifying information (or uncertainty) attached to a random variable or to a probability distribution, other quantities are used as well, such as the moments of the variable (giving information, for instance, on center of mass, dispersion, skewness, or impulsive character), or the Fisher information. In particular, the Fisher information appears in the context of estimation [66,67], in Bayesian inference through the Jeffreys prior [39,68], but also for complex physical systems descriptions [67,69–73].

Although coming from different worlds (information theory and communication, estimation, statistics, and physics), these informational quantities are linked by well-known relations such as the Cramér–Rao inequality, the de Bruijn identity, and the Stam inequality [34,74–76]. These relationships have been proved very useful in various areas, for instance, in communications [34,74,75], in estimation [66], or in physics [77,78], among others. When generalized entropies are considered, it is natural to question the other informational measures' generalization and the associated identities or inequalities. This question gave birth to a large amount of work and is still an active field of research [28,79–90]. For instance, the Cramér–Rao inequality is very important as it gives the ultimate precision in terms of mean square error of an estimator of a parameter (i.e., the minimal error we can achieve). However, there is no reason for choosing a quadratic error in general. This choice is often made as it allows to simplify algebra or to derive estimators quite easily (e.g., of minimum mean square error). One may wish to choose other error criteria (mean of another norm of the error) and/or to stress parts of the distribution of the data in the mathematical average. It is thus of high interest to be able to derive Cramér–Rao inequalities in a context as broad as possible.

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

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

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

2. ϕ -Entropies—Direct and Inverse Maximum Entropy Problems

The direct problem, i.e., finding the probability distribution of maximum entropy given moments constraints, is a common problem and can find application, for instance, in the Bayesian framework, searching for prior probability distribution as less informative as possible, given some moments [22,36–39]. It also finds many other applications, as mentioned in the introduction.

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

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

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

is the ϕ -entropy of f .

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

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

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

Definition 2 (Bregman divergence and functional Bregman divergence [22,91]). *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 $\mathcal{Y} \times \mathcal{Y}$,*

$$B_\phi(y_1, y_2) = \phi(y_1) - \phi(y_2) - \phi'(y_2)(y_1 - y_2). \quad (2)$$

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

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

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

From its positivity and equality only when the distributions are (almost everywhere) equal, this divergence defines a kind of distance (it is not, being non-symmetrical) where f_2 serves as a reference.

More generally, the Bregman divergence is defined for multivariate convex functions, where the derivative is replaced by gradient operator [91]. Extensions for convex function of functions also exist, where the derivative is in the sense of Gâteaux [92]. Such general extensions are not useful for our purposes; thus, we restrict ourselves to the above definition where $\mathcal{Y} \subseteq \mathbb{R}_+$.

2.1. Maximum Entropy Principle: The Direct Problem

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

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

with

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

where $T_0(x) = 1$ and $t_0 = 1$ (normalization constraint), $T = (T_0, \dots, T_n)$, $t = (t_0, \dots, t_n)$. We are faced to a strictly concave optimization problem (the functional to maximize is concave w.r.t. f and the constraints are linear w.r.t. f , so that the functional restricted to a linear subspace is still concave). Therefore, the solution exists and is unique. A technique to solve the problem can be to use the classical Lagrange multipliers technique and to solve the Euler–Lagrange equation from the variational problem, but this approach requires mild conditions [50,51,53,93–95]. In the following proposition, we recall a sufficient condition relating f and ϕ so that f is the problem's solution. This result is proven without the use of the Lagrange technique.

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

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

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

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

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

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

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

for all distributions 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 Equation (6), then it is the desired solution. \square

Therefore, given an entropic functional ϕ and moments constraints T_i , Equation (6) leads to the maximum entropy distribution f^* . This distribution is parameterized by the λ_i s or, equivalently, by the moments t_i s.

Note that the reciprocal is not necessarily true, i.e., the maximum entropy distribution does not necessarily satisfy Equation (6) (i.e., Equation (6) has not necessarily a solution), as shown, for instance, in [53]. However, the reciprocal is true (i.e., Equation (6) has a solution) when \mathcal{X} is a compact [95] or for any \mathcal{X} provided that ϕ is locally bounded on \mathcal{X} [96].

2.2. Maximum Entropy Principle: The Inverse Problems

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

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

The second inverse problem consists in designing the entropy itself, given a target distribution f and given the T_i s. In other words, given a distribution f , Equation (6) may allow to determine the entropic functional ϕ so that f is its maximizer. As mentioned in the introduction, solving this inverse problem can find applications, for instance, in goodness-of-fit tests. In such tests, we would like to determine if data fit a given distribution, say f . A natural criterion of fit between an empirical distribution and distribution f can be a Bregman divergence, where distribution f serves as a reference. As shown in the proof of Proposition 1, when both distributions (empirical, reference) share the same moments and when reference f is of maximum entropy subject to these moments, the divergence turns to be a difference of entropy and approaches in the line of [44–49] can be applied. Distribution f and some moments being given/fixed, the problem is thus to determine the adequate entropy so that f is of maximum entropy. This is precisely the inverse problem we deal with now.

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

- the domain of definition of ϕ' must include $f(\mathcal{X})$; this will be satisfied by construction;

- from the strict convexity property of ϕ , ϕ' must be strictly increasing.

Therefore, because ϕ' must be strictly increasing, it is clear that solving Equation (6) requires the following two conditions:

- (C1) $f(x)$ and $\sum_{i=1}^n \lambda_i T_i(x)$ must have the same variations, i.e., $\sum_{i=0}^n \lambda_i T_i(x)$ is increasing (resp. decreasing, constant) where f is increasing (resp. decreasing, constant);
- (C2) $f(x)$ and $\sum_{i=1}^n \lambda_i T_i(x)$ must have the same level sets,
- $$f(x_1) = f(x_2) \Leftrightarrow \sum_{i=0}^n \lambda_i T_i(x_1) = \sum_{i=0}^n \lambda_i T_i(x_2).$$

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

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

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

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

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

Equation (7) provides thus an effective way to solve the inverse problem. However, there exist situations where there does not exist any set of λ_i s such that conditions (C1)–(C2) are satisfied (e.g., $T_1(x) = x^2$ with f not even). In such a case, we look for a solution for ϕ in a larger class, i.e., by extending the definition of the ϕ -entropy. This will be the purpose of Section 3. Before focusing on this, let us illustrate the previous result on some examples.

2.3. Second Inverse Maximum Entropy Problem: Some Examples

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

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

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

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

Example 2. Let us consider the q -Gaussian distribution, also known as Tsallis distribution or Student distribution [97,98], $f_X(x) = A_q \left(1 - (q-1) \frac{x^2}{\sigma^2}\right)_+^{\frac{1}{1-q}}$, where $q > 0$, $q \neq 1$, $x_+ =$

$\max(x, 0)$ and A_q is the normalization coefficient, defined over $\mathcal{X} = \mathbb{R}$ when $q < 1$ or over $\mathcal{X} = \left(-\frac{\sigma}{\sqrt{q-1}}; \frac{\sigma}{\sqrt{q-1}}\right)$ when $q > 1$, and let search for the ϕ -entropy so that the q -Gaussian is its maximizer with the constraint $T_1(x) = x^2$. Here, again, condition (C1) is satisfied if and only if $\lambda_1 < 0$, whereas condition (C2) is always satisfied. Rapid calculations detailed in Appendix A.2 lead to the entropic functional, after a reparameterization of the λ_i s, as,

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

where q is thus an additional parameter of the family. This entropy is nothing but the Havrda–Charvát or Daróczy or Tsallis entropy [12,14,17,97], up to the scaling factor α , and a shift (here also, to avoid the divergence of the entropy when \mathcal{X} is unbounded, one will take $\gamma = 0$). This entropy is also closely related to the Rényi entropy [10] via a one-to-one logarithmic mapping. One recovers the also well known fact that the q -Gaussian is the maximum Havrda–Charvát–Rényi–Tsallis entropy distribution with the second order moment constraint [97]. In the limit case $q \rightarrow 1$, the distribution f_X tends to the Gaussian, whereas the Havrda–Charvát–Rényi–Tsallis entropy tends to the Shannon entropy.

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

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

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

3. State-Dependent Entropic Functionals and Minimization Revisited

In order to follow asymmetries of the distribution f and address the limitation raised by conditions (C1) and (C2), we propose to allow the entropic functional to also depend on the state variable x . Indeed, imagine, for instance, that, for two values $x_1 \neq x_2$, the probability distribution is such that $f(x_1) = f(x_2)$, but, at the same time, $\sum_i \lambda_i T_i(x_1) \neq \sum_i \lambda_i T_i(x_2)$ (for any set of λ_i s). In such a situation, one cannot find a function ϕ so as to satisfy condition (C2). Choosing a functional ϕ depending both on $f(x)$ and x can allow to have $\phi(x_1, f(x_1)) = \phi(x_2, f(x_2))$ so that we expect it could compensate for the fact that, with a usual entropic functional, condition (C2) cannot be satisfied. In the same vein, imposing a particular form for $\phi(x, f(x))$, we also expect to be able to treat the case where condition (C1) cannot be satisfied with a usual entropic functional. Let us first define the hence extended state-dependent ϕ -entropy, before demonstrating that such an extension allows indeed to reach our goal.

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 $f(\mathcal{X}) \subseteq \mathcal{Y}$,

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

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

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

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

With these extended quantities, the direct problem becomes

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

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

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

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

for some $(\lambda_0, \dots, \lambda_n) \in \mathbb{R}^{n+1}$, then f is the unique solution of the extended maximum entropy problem (10).

If ϕ is chosen in the $(\mathcal{X}_1, \dots, \mathcal{X}_k)$ -multiform ϕ -entropy class, this sufficient condition writes

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

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

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

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

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

- design a partition $(\mathcal{X}_1, \dots, \mathcal{X}_k)$ 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) and
- determine ϕ_l as in Equation (7) in each \mathcal{X}_l , that is

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

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

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

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

$$\phi'_l(y) = \sum_{i=1}^{n_l} \lambda_{l,i} T_{l,i}(f_l^{-1}(y)) \tag{15}$$

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

Example 4. Let us now consider this arcsine distribution, constrained uniformly by $T_1(x) = x$. Clearly, neither condition (C1) nor condition (C2) can be satisfied. Note that the arcsine distribution is a one-to-one function on each set $\mathcal{X}_- = (-\frac{s}{\pi}; 0)$ and $\mathcal{X}_+ = [0; \frac{s}{\pi})$ that partitions \mathcal{X} . Therefore, considering multiform entropic functionals with this partition allows to overcome the issue on condition (C2), but that on condition (C1) remains. If we ignore this issue and apply Equation (14), after a reparameterization of the λ_i s, we obtain $\tilde{\phi}_{\pm}(y) = \tilde{\phi}_{\pm,u}(sy)$ with $\tilde{\phi}_{\pm,u}(u) = \pm\alpha \left(\sqrt{u^2 - 1} + \arctan\left(\frac{1}{\sqrt{u^2 - 1}}\right) \right) \mathbb{1}_{(1; +\infty)}(u) + \beta u + \gamma_{\pm}$ where s is thus an additional parameter of the family. It appears that whereas these functionals are defined for $u > 1$, one can extend them continuously and with a continuous derivative for any $u > 0$ imposing $\beta = 0$, which finally leads to the family

$$\begin{aligned} \tilde{\phi}_{\pm}(y) &= \tilde{\phi}_{\pm,u}(sy) \quad \text{with} \\ \tilde{\phi}_{\pm,u}(u) &= \pm\alpha \left(\sqrt{u^2 - 1} + \arctan\left(\frac{1}{\sqrt{u^2 - 1}}\right) \right) \mathbb{1}_{(1; +\infty)}(u) + \gamma_{\pm} \end{aligned}$$

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

Example 5. If now we impose the partial constraint $T_{\pm,1}(x) = x\mathbb{1}_{\mathcal{X}_{\pm}}(x)$, and search for the ϕ -entropy so that f_X is the maximizer subject to these constraints, condition (C1) can be now satisfied on each \mathcal{X}_{\pm} by imposing the $\pm\lambda_{\pm,1}$ given Equation (15) to be positive. We then obtain the associated multiform entropic functional, after a reparameterization of the λ_i s, as $\phi_{\pm}(y) = \phi_{\pm,u}(sy)$ with $\phi_{\pm,u}(u) = \alpha_{\pm} \left(\sqrt{u^2 - 1} + \arctan\left(\frac{1}{\sqrt{u^2 - 1}}\right) \right) \mathbb{1}_{(1; +\infty)}(u) + \beta u + \gamma_{\pm}$ with $\alpha_{\pm} > 0$ and where s is thus an additional parameter of the family. In this case, the entropic functionals can be considered for any $u > 0$ by imposing $\beta = 0$, and one can check that the obtained functions are of class C^1 . This finally leads to the family

$$\begin{aligned} \phi_{\pm}(y) &= \tilde{\phi}_{\pm,u}(sy) \quad \text{with} \\ \phi_{\pm,u}(u) &= \alpha_{\pm} \left(\sqrt{u^2 - 1} + \arctan\left(\frac{1}{\sqrt{u^2 - 1}}\right) \right) \mathbb{1}_{(1; +\infty)}(u) + \gamma_{\pm}, \quad \alpha_{\pm} > 0 \end{aligned}$$

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

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

In Sections 2 and 3, we established general entropies with a given maximizer. In what follows, we will complete the information theoretical setting by introducing generalized escort distributions, generalized moments, and generalized Fisher information associated to the same entropic functional. We will then explore some of their relationships. Indeed, as mentioned in the introduction, the Cramér–Rao inequality is very important as it gives the ultimate precision in terms of mean square error of an estimator of a parameter. As we would like to escape from the usual quadratic loss (that has often mathematical motivation but not physical one, and that even can not exist) and/or to stress parts of the distribution of the data so has to penalize for instance large errors depending of the tails of the distribution, it is thus of high interest to be able to derive Cramér–Rao inequalities in a broader framework, which can find natural applications in the estimation domain.

4. ϕ -Escort Distribution, (ϕ, α) -Moments, (ϕ, β) -Fisher Information, Generalized Cramér–Rao Inequalities

In this section, we begin by introducing the above-mentioned informational quantities. We will then show that generalizations of the celebrated Cramér–Rao inequalities hold and link the generalized moments and Fisher information. Furthermore, the lower bound of the inequalities are saturated precisely by maximal ϕ -entropy distributions. To derive such generalizations of this inequality, we thus need to precisely define the above mentioned generalization of the moments and of the Fisher information that will lower bound the moment (e.g., of any estimator of a parameter). The proposed generalizations are based on the notion of escort distribution we first need to introduce.

Escort distributions have been introduced as an operational tool in the context of multifractals [99,100], with interesting connections with the standard thermodynamics [101] and with source coding [26,27]. In our context, we also define (generalized) escort distributions associated with a particular convex function ϕ , and show how they pop up naturally. It is then possible to define generalized moments with respect to these escort distributions. Such distributions were previously introduced dealing with Rényi entropies and took the form f^q as we will see later on. When $q > 1$, the effect is to stress the head of the distribution, i.e., to penalize more the errors where the data fall in the head of the distribution. At the opposite, when $q < 1$, the tails of the distributions are stressed. As we will see later on in the proof of the generalized Cramér–Rao inequality, any form as an escort distribution can be chosen. However, as for the usual nonparametric Cramér–Rao inequality, one may wish the inequality to be saturated for the maximum entropy distribution, which fixes the form of the escort distribution as follows.

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

if f is a probability distribution defined with respect to a general measure μ on a set \mathcal{X} such that $f(\mathcal{X}) \subseteq \mathcal{Y}$, and such that

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

we define by

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

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

Note that from the strict convexity of ϕ with respect to its second argument, this probability density is well defined and is strictly positive. We can note that, with the above definition, the ϕ -escort distribution will tend to stress the parts of the distribution where $\phi(x, f(x))$ has a small "curvature." Moreover, coming back to the previous examples, one can see the following.

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

Example 2 (cont.). In the Rényi–Tsallis context, entropy for which the q -Gaussian is the maximal entropy law for the second-order moment constraint $\phi(x, y) = \phi(y) = \frac{y^q - y}{q-1}$, and $E_{\phi, f} \propto f^{2-q}$ which recovers the escort distributions used in the Rényi–Tsallis context up to a duality transformation [101].

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

Definition 5 ((α, ϕ) -moments). Under the assumptions of Definition 4, with \mathcal{X} equipped with a norm $\|\cdot\|_{\mathcal{X}}$, we define the (α, ϕ) -moment of a random variable X associated to distribution f by

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

if this quantity exists.

This definition goes further than the usual definition of variance as a measure of dispersion, both by generalizing the exponent, the norm, and by taking the mean with respect to an escort distribution. Thanks to the escort distribution, one can stress special parts of the distribution (heads, tails, parts where ϕ has a small curvature that is with a small informational content in a sense). Here, again, any escort distribution could have been chosen, but, as pointed out previously, that of the definition allows to saturate the Cramér–Rao inequality we will derive in a while for the maximum entropy distribution. Note that, in the particular case of the Euclidean norm and $\alpha = 2$, the second-order moment statistics are indeed contained in the second-order moments matrix given by the mathematical mean of XX^t . In such a context, the definition above coincides with the trace of this second order moment matrix and represents the total power of X .

This said, for our three examples, we have the following.

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

Example 2 (cont.). In the Rényi–Tsallis context the generalized moments introduced in [61,102] are recovered.

Example 3 (cont.). For $\phi(x, y) = \phi(y) = \frac{1}{y}$, one also naturally finds generalized moments with the same form as those introduced in [61,102] (see the items related to the escort distributions).

The Fisher information’s importance is well known in estimation theory: the estimation error of a parameter is bounded by the inverse of the Fisher information associated with this distribution [34,66]. The Fisher information is also used as a method of inference and understanding in statistical physics and biology, as promoted by Frieden [67] and has been generalized in the Rényi–Tsallis context in a series of papers [81,84,86–89,103,104]. In the following, we generalize these definitions a step further in our ϕ -entropy context.

Definition 6 (Nonparametric (β, ϕ) -Fisher information). With the same assumption as in Definition 4, denoting by $\|\cdot\|_{\chi^*}$ the dual norm (the norm induced in the dual space that gives here $\|z\|_{\chi^*} = \sup_{\|x\|_{\chi}=1} z^t x$ [105,106]), for any differentiable density f , we define the quantity

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

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

Note that the Fisher information can be viewed as local, as it is sensitive to the variation of a distribution, rather than to the distribution itself. As for the generalized moments, through the power β other moments for the gradient of f than the second one can be considered, so that more or less weight can be put in the variations of the distribution. Moreover, as for the case of generalized moments, any escort distribution could have been chosen, but, again this choice is dictated by our wish to saturate the Cramér–Rao inequality for the maximum entropy distribution.

Note also 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 has $I_{\beta, \phi}[g] = I_{\beta, \phi}[f]$. This property is unfortunately lost in the state-dependent context. Furthermore, whereas the Fisher information have scaling properties $I[a^{-d} f(\cdot/a)] = I[f]/a^2$, this is lost for $I_{\beta, \phi}$, except when ϕ'' is a power (which corresponds either to the Shannon or Rényi–Tsallis entropy).

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

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

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

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

Note that in the classical setting, the information on X in the sense of Fisher is given by the so-called Fisher information matrix, which is the mathematical mean of $\nabla f \nabla^t f$. Taking the trace of the Fisher information matrix, one obtains what is often called Fisher information (without the term “matrix”), which is nothing but the expectation of $\|\nabla f\|^2$ [58,67,107]. This is in the line of the above definitions. Extending these definitions to obtain a matrix would have been possible by averaging over the ϕ -escort distribution the element-wise power $\beta/2$ of matrix $(\nabla f \nabla^t f) / E_{\phi, f}^2$, but the trace of this matrix does not coincide anymore with the above definition. Moreover, it is not obvious that it will allow a generalization of the matrix form of the Cramér–Rao inequality we will see in the following. Such a matrix extended Fisher information is left as a perspective.

For our three examples, we have the following.

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

Example 2 (cont.). Similarly, in the Rényi–Tsallis context, the generalizations proposed in [87–89] are recovered.

Example 3 (cont.). For $\phi(x, y) = \phi(y) = \frac{1}{y}$, one also naturally finds, the generalizations proposed in [87–89] (see the items related to the escort distributions).

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

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

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

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

Proof. The approach follows [89], starting from the differentiable probability density f (derivative denoted $\nabla_x f$), as $xf(x)$ vanishes in the boundaries of X from the divergence theorem one has

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

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

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

for any function g non-zero on \mathcal{X} . Now, noting that $d > 0$, we obtain from the work in [89] (Lemma 2)

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

The proof ends by choosing $g = E_{\phi, f}$ the ϕ -escort density associated to density f . Note now that, again from [89] (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 Equation (6) of Proposition 1. \square

Analyzing minutely the proof, it is clear that both in the generalized moments and the generalized Fisher information, any escort distribution g can be chosen (being identical for both quantities), including the probability distribution itself. The saturation will be achieved for the distribution f satisfying $\nabla_x f(x) g(x) = \lambda_1 \nabla_x \|x\|_{\mathcal{X}}^\alpha$, but the ϕ -escort distribution Definition 4 is the only choice which allows to recover maximal ϕ -entropy as the saturating distribution; of course with the same ϕ as in the escort distribution, and with the moment constraint similar to that of the inequality but averaged over the distribution itself.

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

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

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

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

where

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

is the bias of the estimator and α and α^* are Hölder 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 $T(x) = \|\hat{\theta}(x) - \theta\|_{\Theta}^\alpha$.

Proof. The proof follows again that of [89], and starts by evaluating the divergence of the bias. The regularity conditions in the statement of the theorem enable to interchange integration with respect to x and differentiation with respect to θ , so that

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

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

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

for any density g non-zero on \mathcal{X} . The proof ends with the very same steps that in Proposition 4 using [89] (Lemma 2). \square

In the classical setting, in the multivariate context ($m > 1$), the Cramér–Rao inequality takes a matrix form, stating that the difference of the second order moment matrix of the estimation error of an estimator with the inverse Fisher information matrix is positive definite [34,58,66,67,108,109]. Several scalar forms can be derived, for instance by taking the determinant, the trace, and/or by mean of trace [58,66,67,108] or determinant/trace inequalities [110]. Typically, by mean of the trace, the scalar equivalent of the above results are recovered. Conversely, extending our result in a matrix context is not immediate and left as a perspective.

For our three examples, Propositions 3 and 4 lead to what follows.

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

Example 2 (cont.). *In the Rényi–Tsallis context, the generalizations proposed in [87–89] are recovered and, again, when $\alpha = 2$, the bound is saturated in the nonparametric context for the q -Gaussian, maximal entropy law under the second order moment constraint.*

Example 3 (cont.). *For $\phi(x, y) = \phi(y) = \frac{1}{y}$, again, one finds inequalities with the same form as those of the generalizations proposed in [87–89] (see the items related to the escort distributions).*

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

5. ϕ -Heat Equation and Extended de Bruijn Identity

An important relation connecting the Shannon entropy H , coming from the “information world”, with the Fisher information I , living in the “estimation world”, is given by the de Bruijn identity and it is closely linked to the Gaussian distribution. Considering a noisy random variable $Y_{\theta} = X + \sqrt{\theta}N$ where N is a zero-mean d -dimensional standard Gaussian random vector and X a d -dimensional random vector independent of N , and of support independent on parameter θ , then

$$\frac{d}{d\theta} H[f_{Y_{\theta}}] = \frac{1}{2} I[f_{Y_{\theta}}]$$

where $f_{Y_{\theta}}$ stands for the probability distribution of Y_{θ} . This identity is a critical ingredient in proving the entropy power and Stam inequalities [34]. The de Bruijn identity has applications in communication by characterizing a channel face to noise [34,76,111,112] or in mismatch estimation [113]. It is involved in the Entropy Power Inequality, which itself is involved in an informational proof of the central limit theorem [114–116]. Extending the de Bruijn identity is thus of great interest as, for instance, it may allow to characterize more general communication channels in the same line than that in [117] or for non-additive noise or to give rise to generalized central limit theorem [115,116].

The starting point to establish the de Bruijn identity is the heat equation satisfied by the probability distribution f_{Y_θ} , $\frac{\partial f}{\partial \theta} = \frac{1}{2} \Delta f$, where Δ stands for the Laplacian operator [118].

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

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

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

When θ is scalar, this equation is an instance of what are known as quasilinear parabolic equations [119] (§ 8.8) and arises in various physical problems.

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

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

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

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

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

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 hand side reduces to the integral of $\|\nabla_x \phi'(f)\|_{\chi^*}^{\beta-2} \nabla_x \phi(f)$ on the boundary of \mathcal{X} , that vanishes from the assumption of the proposition, while the second term of the right hand side gives the right hand side of (25) from C_ϕ and the (β, ϕ) -Fisher information given by Equations (16) and (17) and Definition 6. \square

As for the Cramér–Rao inequality, in the classical settings there exist matrix variants of the de Bruijn identity, the scalar form being a special one [115,117].

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

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

Example 2 (cont.). The case where $\phi(y) = y^q$ was intensively studied in [90] and the results of the paper are naturally recovered. In particular, the generalized ϕ -heat equation appears in anomalous diffusion in porous medium [90,119–122].

Example 3 (cont.). For $\phi(x, y) = \phi(y) = \frac{1}{y}$, once again one finds the same form for the generalized heat equation than in [90,120,121], and therefore the same form of the generalized de Bruijn identity of [90] (see the items related to the escort distributions).

6. Concluding Remarks

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

As a direct perspective, the extensions of the generalized moments and Fisher information in terms of matrix, and matrix form of the generalized Cramér–Rao inequalities and de Bruijn identities are still open problems. Several ways to define matrix moments and Fisher information may be considered, such as in a term-wise manner as evoked in this paper. However, deriving matrix forms of the inequalities and identities does not seem trivial, and neither does obtaining the scalar form, for instance, through trace operator. Moreover, as the de Bruijn identity can be closely related to the generalized Price’s theorem [123–125], studying the connections between the extended de Bruijn and this theorem, or generalizing following the work of [125] is also of great interest.

Furthermore, two important inequalities are still lacking: The first one is the entropy power inequality (EPI), which states that the entropy power (exponential of twice the entropy) of the sum of two continuous independent random variables is higher than the sum of the individual entropy powers (In fact, there exist other equivalent versions which can be found, e.g., in [34,75,107,126–128].). The second one is the Stam inequality which lower bounds the product of the entropy power and the Fisher information. For the former, despite many efforts, the literature on extended version only considers special cases. For instance, some extensions in the classical settings exist for discrete variables but are somewhat limited [129–131]. In the continuous framework, the EPI was also extended to the class of the Rényi entropy (log of a ϕ -entropy with $\phi(u) = u^\alpha$) [132]. Note that variants of the EPI also exist in the context where one of the variables is Gaussian. This is equivalent to the convexity property of $\theta \mapsto N(X + \sqrt{\theta}Y)$ with N the entropy power and Y a Gaussian noise independent on X [133–137]; property also extended in the context of the Rényi entropy [132,138–140]. An important property that plays a key role in the inequality is the fact that the Rényi entropy is invariant to an affine transform of unit determinant and monotonic under convolution, a property which seems lost in the very general setting considered here. This fact leaves little room to extend the EPI in our general

settings. Concerning the Stam inequality, at a first glance, the fact that the proof is based on the EPI seems to close any hope to extend it to the ϕ -entropy framework. However, it was remarkably extended to the Rényi entropy, based on the Gagliardo–Nirenberg inequality [84,86,87,141]. Nevertheless, a key property is that both the entropy power and the extended Fisher information have scaling properties that are lost in the general setting of the ϕ -entropies. A possible way to overcome the (apparent) limits just evoked could be to mimic alternative proofs such as those based on optimal transport [142]. This approach precisely drops off any use of Young or Sobolev-like inequalities. As far as we see, there is thus a little room for extensions in the settings of the paper. Both the extension of the EPI and the Stam inequality are left as perspectives.

Another perspective lies in the estimation of the generalized moments from data (or from estimates). Such a possibility would confer an operational role to our Cramér–Rao inequality, i.e., by computing the estimator’s generalized moments and comparing them to the bound. A difficulty resides in the presence of the ϕ -escort distribution which forbids empirical or Monte Carlo approaches. The escort distribution needs to be estimated. This problem seems not far from the estimation of entropies from data and plug-in approaches used in such problems can thus be considered, like kernel approaches [143–145], nearest neighbor approaches [145,146], or minimal spanning tree approaches [42]. Of course, this perspective goes far beyond the scope of this paper.

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

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

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Acknowledgments: The authors wish to warmly thank the three reviewers who gave a careful reading of this manuscript. Their very valuable remarks and suggestions led to the improvement of the manuscript and opened various perspectives.

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

Appendix A. Inverse Maximum Entropy Problem and Associated Inequalities: Some Examples

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

Appendix A.1. Normal Distribution and Second-Order Moment

For a normal distribution and second-order moment constraint

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

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

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

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

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

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

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

which gives nothing but the Shannon entropy, as expected,

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

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

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

Appendix A.2. q -Gaussian Distribution and Second-Order Moment

For q -Gaussian distribution, also known as Tsallis distribution, Student-t, and Student-r [97,98], and a second-order moment constraint, we have

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

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

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

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

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

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

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

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

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

and an associated entropy is then

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

where \mathcal{X} is now the support of f . This entropy is nothing but the Havrda–Charvát–Tsallis entropy [12,14,17,97] (overall, we obtain this entropy up to a scaling and a shift).

Then, $\phi''(y) = qy^{q-2}$, so that, from Definition 4, and then from Definitions 5–7, respectively, we obtain $M_{\phi,\alpha}[f]$ and $I_{\phi,\alpha}[f]$ as, respectively, the q -moment of order α and the (q, β) -Fisher information defined previously in [84–89] (with the symmetric q index given here by $2 - q$). The extended Cramér–Rao inequality proved in [84,88,89] is then recovered from Propositions 3 and 4, and the generalized de Bruijn identity of [90] is also recovered from Equation (24) and Proposition 5.

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

Appendix A.3. q -Exponential Distribution and First-Order Moment

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

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

where B_q is a normalization coefficient. It suffices to follow the very same steps as above, leading again to the Havrda–Charvát–Tsallis entropy, the q -moments of order α and the (q, β) -Fisher information defined previously in [84–89] (with the symmetric q index given here by $2 - q$) as for the q -Gaussian distribution and to the extended Cramér–Rao inequality proved in [88,89] as well.

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

Appendix A.4. The Arcsine Distribution

The arcsine distribution is a special case of the beta distribution with shaping parameter $\alpha = \beta = \frac{1}{2}$ and appears in various application, e.g., see in [98]. We consider here the centered and scaled version of this distribution which writes

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

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

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

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

Appendix A.4.1. Second-Order Moment

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

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

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

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

Although the inverse of the arcsine distribution does not exist for $y \leq \frac{1}{s}$, the entropy functional can be defined over \mathbb{R}_+^* .

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

Appendix A.4.2. (Partial) First-Order Moment(s)

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

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

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

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

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

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

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

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

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

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

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

and the branches for the non-convex case

$$\tilde{\phi}_{\pm}(y) = \tilde{\phi}_{\pm,u}(sy) \quad \text{with}$$

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

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

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

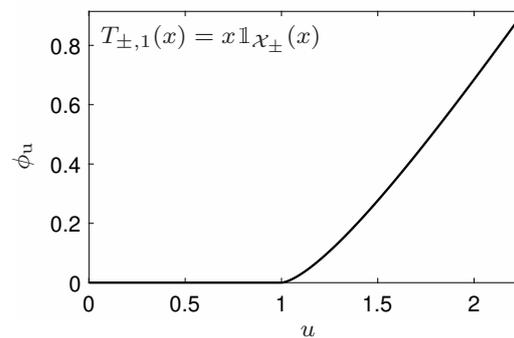


Figure A1. Univalued entropic functional ϕ_u derived from the arcsine distribution with partial constraints $T_{\pm,1}(x) = x \mathbb{1}_{\mathcal{X}_{\pm}}(x)$.

Appendix A.5. The Logistic Distribution

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

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

This distribution, which resembles the normal distribution but has heavier tails, has been used in various applications, e.g., see in [98]. One can then check that over each interval

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

the inverse distribution writes

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

Let us now focus on a second-order constraint, that respects the symmetry of the distribution, and on first-order constraint(s) that do(es) not respect the symmetry.

Appendix A.5.1. Second Order Moment Constraint

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

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

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

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

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

$$\phi(y) = \phi_u(sy) \quad \text{with}$$

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

depicted Figure A2a for $\alpha = 1, \gamma = 0$.

Appendix A.5.2. (Partial) First-Order Moment(s) Constraint(s)

As f_X and $T(x) = x$ do not share the same symmetries, one cannot interpret the logistic distribution as a maximum entropy constraint by the first order moment. However, constraining the partial means over $\mathcal{X}_\pm = \mathbb{R}_\pm$ and using multiform entropies allow such an interpretation, while the alternative is to relax the concavity property of the entropy—but again, one would only be able to ensure that the distribution from which it comes is a critical point. To be more precise, one chooses

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

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

$$\phi'_\pm(y) = \lambda_0 + \frac{\lambda_{\pm,1}s}{2} \operatorname{arctanh} \sqrt{1-sy} \quad \& \quad \tilde{\phi}'_\pm(y) = \lambda_0 \pm \frac{\lambda_1 s}{2} \operatorname{arctanh} \sqrt{1-sy}$$

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

$$\lambda_\pm < 0.$$

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

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

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

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

$$\phi_{\pm,u}(u) = \gamma_\pm - \alpha_\pm \left(u \operatorname{arctanh} \sqrt{1-u} - \sqrt{1-u} \right) \mathbb{1}_{(0;1]}(u), \quad \alpha_\pm > 0$$

and the branches for the non-convex case

$$\tilde{\phi}_\pm(y) = \tilde{\phi}_{\pm,u}(sy) \quad \text{with}$$

$$\tilde{\phi}_{\pm,u}(u) = \gamma_\pm \pm \alpha \left(u \operatorname{arctanh} \sqrt{1-u} - \sqrt{1-u} \right) \mathbb{1}_{(0;1]}(u)$$

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

The uniform function ϕ_u is represented Figure A2b for $\alpha_{\pm} = 1$, $\gamma_{\pm} = 0$.

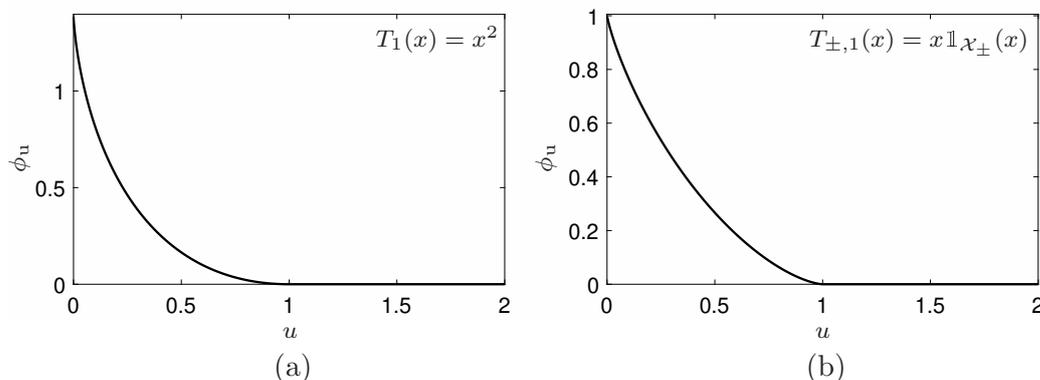


Figure A2. Entropy functional ϕ_u derived from the logistic distribution: (a) with $T_1(x) = x^2$ and (b) with $T_{\pm,1}(x) = x \mathbb{1}_{\mathcal{X}_{\pm}}(x)$.

Appendix A.6. The Gamma Distribution and (Partial) P-Order Moment(s)

As a very special case, let us finally consider the gamma distribution expressed as

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

Parameter $q > 0$ is known as the shape parameter of the law, while $\sigma = \frac{r}{\Gamma(q)} > 0$ is a scaling parameter. This distribution appears in various applications, as described, for instance, in [147].

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

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

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

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

or interpret f_X as a critical point of a ϕ -like entropy with a constraint on the moment

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

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

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

As expected, this equation has two solutions. These solutions can be expressed thanks to the multivalued Lambert-W function W defined by $z = W(z) \exp(W(z))$, i.e., W is the inverse function of $u \mapsto u \exp(u)$ [148] (§ 1), leading to the inverse functions

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

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

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

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

where, to ensure the convexity of the ϕ_k ,

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

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

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

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

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

Obtaining a closed-form expression for the integral term is not an easy task. However, relation $z(1+W_k(z))W'_k(z) = W_k(z)$ [148] (Equation (3.2)) suggests that a way to make the integration is to search for it under the form of a series

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

Therefore, to obtain a recursion on the a_l , we proceed as follows: (i) we differentiate both side, (ii) we use the relation $z W'_k(z) = \frac{W_k(z)}{1+W_k(z)}$ given above applied to $z = -\frac{u^{\frac{1}{q-1}}}{q-1}$, (iii)

we thus multiply both side of the obtained equality by $1 + W_k\left(-\frac{u^{\frac{1}{q-1}}}{q-1}\right)$, and finally (iv)

we equal the coefficients of the terms in $\left[(1-q) W_k\left(-\frac{u^{\frac{1}{q-1}}}{q-1}\right) \right]^{p+l}$. The a_l can thus be evaluated explicitly, and we recognize in the series the confluent hypergeometric (or

Kummer’s function ${}_1F_1(1; p + q; \cdot)$ [149] (Equation (13.1.2)) or [150] (Equation (9.210-1)) (up to a factor and an additive constant), so that

$$\phi_{k,u}(u) = \gamma_k + \beta u + (-1)^k \alpha_k u \left[(1 - q) W_k \left(-\frac{u^{\frac{1}{q-1}}}{q-1} \right) \right]^p \times \left[1 - \frac{p}{p+q-1} {}_1F_1 \left(1; p+q; (1-q) W_k \left(-\frac{u^{\frac{1}{q-1}}}{q-1} \right) \right) \right] \mathbb{1}_{\left(0; \left(\frac{q-1}{e}\right)^{q-1}\right)}(u), \quad \alpha_k > 0$$

One can check that these functions are indeed the ones we search for. To this end, (i) one derives the previous expression, (ii) one notes that from $z W'_k(z) = \frac{W_k(z)}{1+W_k(z)}$ [148]

(Equation (3.2)) we have $u \left[(1 - q) W_k \left(-\frac{u^{\frac{1}{q-1}}}{q-1} \right) \right]' = -\frac{W_k \left(-\frac{u^{\frac{1}{q-1}}}{q-1} \right)}{1+W_k \left(-\frac{u^{\frac{1}{q-1}}}{q-1} \right)}$, (iii) one finally

uses the relation $(p + q - 1 - z) {}_1F_1(1; p + q; z) + z {}_1F_1'(1; p + q; z) = (p + q - 1) {}_1F_1(0; p + q; z)$ [149] (13.4.11) together with ${}_1F_1(0; b; z) = 1$ [149] (13.1.2).

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

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

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

From [148] (Equation (3.1)), we have $W_0(0) = 0$ and from [149] (Equation (13.1.2)) ${}_1F_1(1; p + q; 0) = 1$, so that

$$\phi_{0,u}(0) = \gamma_0 \quad \text{and} \quad \phi'_{0,u}(0) = \beta.$$

Then, $\lim_{x \rightarrow 0^-} W_{-1}(x) = -\infty$ (see [148] (Figure 1 or Equation (4.18))) so that, (i) from the asymptotic [149] (Equation (13.1.4)) of the confluent hypergeometric function for a large argument, (ii) using $W(z)e^{W(z)} = z$ for $z = -\frac{u^{\frac{1}{q-1}}}{q-1}$, we obtain

$$\phi_{-1,u}(0) = \gamma_{-1} + p \Gamma(p + q - 1) \alpha_{-1} \quad \text{and} \quad \lim_{u \rightarrow 0^-} \phi'_{-1,u}(u) = -\infty.$$

Finally, from $W_k(-e^{-1}) = -1$ we immediately have

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

and

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

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

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

Therefore, in accordance

- The constraints degenerate to a single uniform constraint $T_1(x) = x^p$;
- In this limit, conditions (C1) and (C2) are both satisfied.
- The entropic functional becomes state-independent (uniform), where only the branch ϕ_{-1} remains.

One can determine the limit entropic functional using [151] (Th. 3.2) that states for any $t > 0$,

$$\left| W_{-1} \left(-e^{-(t+1)} \right) + \log(t+1) + (t+1) \right| \leq 1 - \log(e-1) = a$$

We apply this theorem to the positive real t given by

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

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

$$\left| (1-q) W_{-1} \left(-\frac{u^{\frac{1}{q-1}}}{q-1} \right) + \log u - (q-1) \log \left((q-1) \log(q-1) - \log u \right) \right| \leq (q-1)a.$$

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

$$\lim_{q \rightarrow 1^+} (1-q) W_k \left(-\frac{u^{\frac{1}{q-1}}}{q-1} \right) = -\log u.$$

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

$$\lim_{q \rightarrow 1^+} \phi_{1,u}(u) = \gamma_{-1} + \beta u - \alpha_{-1} u (-\log u)^p \left(1 - {}_1F_1(1; p+1; -\log u) \right)$$

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

Finally, for $p = 1$, using [149] (13.6.14) which states that ${}_1F_1(1; 2; x) = \frac{e^x - 1}{x}$, we obtain after simple algebra

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

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

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

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

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

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

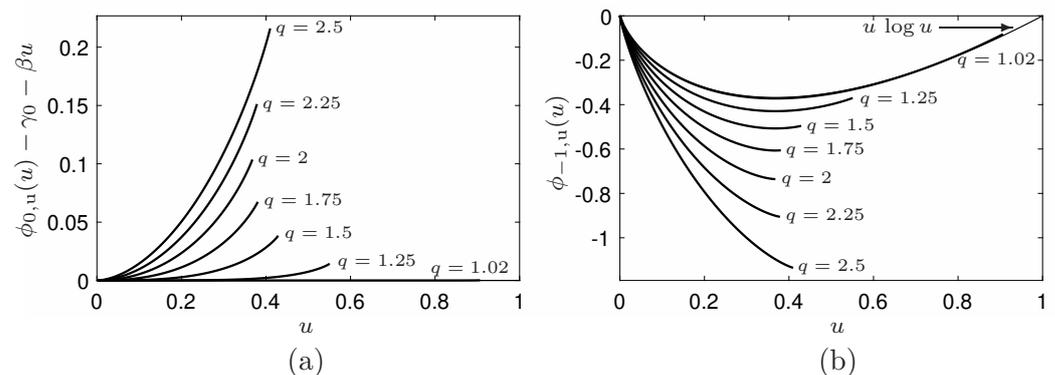


Figure A3. Multiform entropy functional ϕ_u derived from the gamma distribution with the partial moment constraints $T_{k,1}(x) = x \mathbb{1}_{\mathcal{X}_k}(x)$ ($p = 1$), $k \in \{0, -1\}$ for $q = 1.02, 1.25, 1.5, 1.75, 2, 2.25, 2.5$. (a): $\phi_{0,u} - \gamma_0 - \beta u$ ($\alpha_0 = 1$); (b):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

References

1. von Neumann, J. Thermodynamik quantenmechanischer Gesamtheiten. *Nachr. Ges. Wiss. Gött.* **1927**, *1*, 273–291.
2. Shannon, C.E. A Mathematical Theory of Communication. *Bell Syst. Tech. J.* **1948**, *27*, 623–656. [[CrossRef](#)]
3. Boltzmann, L. *Lectures on Gas Theory (Translated by S. G. Brush)*; Dover: Leipzig, Germany, 1964.
4. Boltzmann, L. *Vorlesungen Über Gastheorie—I*; Verlag von Johann Ambrosius Barth: Leipzig, Germany, 1896.
5. Boltzmann, L. *Vorlesungen Über Gastheorie—II*; Verlag von Johann Ambrosius Barth: Leipzig, Germany, 1898.
6. Planck, M. *Eight Lectures on Theoretical Physics*; Columbia University Press: New York, NY, USA, 2015.
7. Maxwell, J.C. *The Scientific Papers of James Clerk Maxwell*; Dover: New York, NY, USA, 1952; Volume 2.
8. Jaynes, E.T. Gibbs vs Boltzmann Entropies. *Am. J. Phys.* **1965**, *33*, 391–398. [[CrossRef](#)]
9. Müller, I.; Müller, W.H. *Fundamentals of Thermodynamics and Applications. With Historical Annotations and Many Citations from Avogadro to Zermelo*; Springer: Berlin/Heidelberg, Germany, 2009; [[CrossRef](#)]
10. Rényi, A. On measures of entropy and information. *Proc. Berkeley Symp. Math. Stat. Probab.* **1961**, *1*, 547–561.
11. Varma, R.S. Generalization of Rényi's Entropy of Order k . *J. Math. Sci.* **1966**, *1*, 34–48.
12. Havrda, J.; Charvát, F. Quantification Method of Classification Processes: Concept of Structural α -Entropy. *Kybernetika* **1967**, *3*, 30–35.
13. Csiszár, I. Information-Type Measures of Difference of Probability Distributions and Indirect Observations. *Stud. Sci. Math. Hung.* **1967**, *2*, 299–318.
14. Daróczy, Z. Generalized Information Functions. *Inf. Control* **1970**, *16*, 36–51. [[CrossRef](#)]
15. Aczél, J.; Daróczy, Z. *On Measures of Information and Their Characterizations*; Academic Press: New York, NY, USA, 1975.
16. Daróczy, Z.; Járjai, A. On the measurable solution of a functional equation arising in information theory. *Acta Math. Acad. Sci. Hung.* **1979**, *34*, 105–116. [[CrossRef](#)]
17. Tsallis, C. Possible Generalization of Boltzmann-Gibbs Statistics. *J. Stat. Phys.* **1988**, *52*, 479–487. [[CrossRef](#)]
18. Salicrú, M. Funciones de entropía asociada a medidas de Csiszár. *Qüestió* **1987**, *11*, 3–12.
19. Salicrú, M.; Menéndez, M.L.; Morales, D.; Pardo, L. Asymptotic distribution of (h, ϕ) -entropies. *Commun. Stat. Theory Methods* **1993**, *22*, 2015–2031. [[CrossRef](#)]
20. Salicrú, M. Measures of information associated with Csiszár's divergences. *Kybernetika* **1994**, *30*, 563–573.
21. Liese, F.; Vajda, I. On Divergence and Informations in Statistics and Information Theory. *IEEE Trans. Inf. Theory* **2006**, *52*, 4394–4412. [[CrossRef](#)]
22. Basseville, M. Divergence measures for statistical data processing—An annotated bibliography. *Signal Process.* **2013**, *93*, 621–633. [[CrossRef](#)]

23. Panter, P.F.; Dite, W. Quantization distortion in pulse-count modulation with nonuniform spacing of levels. *Proc. IRE* **1951**, *39*, 44–48. [[CrossRef](#)]
24. Lloyd, S.P. Least Squares Quantization in PCM. *IEEE Trans. Inf. Theory* **1982**, *28*, 129–137. [[CrossRef](#)]
25. Gersho, A.; Gray, R.M. *Vector Quantization and Signal Compression*; Kluwer: Boston, MA, USA, 1992. [[CrossRef](#)]
26. Campbell, L.L. A coding theorem and Rényi's entropy. *Inf. Control* **1965**, *8*, 423–429. [[CrossRef](#)]
27. Bercher, J.F. Source coding with escort distributions and Rényi entropy bounds. *Phys. Lett. A* **2009**, *373*, 3235–3238. [[CrossRef](#)]
28. Burbea, J.; Rao, C.R. On the Convexity of Some Divergence Measures Based on Entropy Functions. *IEEE Trans. Inf. Theory* **1982**, *28*, 489–495. [[CrossRef](#)]
29. Menéndez, M.L.; Morales, D.; Pardo, L.; Salicrú, M. (h, Φ) -entropy differential metric. *Appl. Math.* **1997**, *42*, 81–98. [[CrossRef](#)]
30. Pardo, L. *Statistical Inference Based on Divergence Measures*; Chapman & Hall: Boca Raton, FL, USA, 2006.
31. Jaynes, E.T. Information Theory and Statistical Mechanics. *Phys. Rev.* **1957**, *106*, 620–630. [[CrossRef](#)]
32. Kapur, J.N. *Maximum Entropy Model in Sciences and Engineering*; Wiley Eastern Limited: New Dehli, India, 1989.
33. Arndt, C. *Information Measures: Information and Its Description in Sciences and Engineering*; Springer: Berlin/Heidelberg, Germany, 2001. [[CrossRef](#)]
34. Cover, T.M.; Thomas, J.A. *Elements of Information Theory*, 2nd ed.; John Wiley & Sons: Hoboken, NJ, USA, 2006.
35. Gokhale, D.V. Maximum entropy characterizations of some distributions. In *A Modern Course on Statistical Distributions in Scientific Work*; Patil, S.K., Ord, J.K., Eds.; Reidel: Dordrecht, The Netherlands, 1975; Volume III, pp. 299–304. [[CrossRef](#)]
36. Jaynes, E.T. Prior probabilities. *IEEE Trans. Syst. Sci. Cybern.* **1968**, *4*, 227–241. [[CrossRef](#)]
37. Csiszár, I. Why Least Squares and Maximum Entropy? An Axiomatic Approach to Inference for Linear Inverse Problems. *Ann. Stat.* **1991**, *19*, 2031–2066. [[CrossRef](#)]
38. Frigyi, B.A.; Srivastava, S.; Gupta, M.R. Functional Bregman Divergence and Bayesian Estimation of Distributions. *IEEE Trans. Infom. Theory* **2008**, *54*, 5130–5139. [[CrossRef](#)]
39. Robert, C.P. *The Bayesian Choice. From Decision-Theoretic Foundations to Computational Implementation*, 2nd ed.; Springer: New York, NY, USA, 2007.
40. Jaynes, E.T. On the rational of maximum-entropy methods. *Proc. IEEE* **1982**, *70*, 939–952. [[CrossRef](#)]
41. Jones, L.K.; Byrne, C.L. General Entropy Criteria for Inverse Problems, with Applications to Data Compression, Pattern Classification, and Cluster Analysis. *IEEE Trans. Inf. Theory* **1990**, *36*, 23–30. [[CrossRef](#)]
42. Hero III, A.O.; Ma, B.; Michel, O.J.J.; Gorman, J. Application of Entropic Spanning Graphs. *IEEE Signal Process. Mag.* **2002**, *19*, 85–95. [[CrossRef](#)]
43. Park, S.Y.; Bera, A.K. Maximum entropy autoregressive conditional heteroskedasticity model. *J. Econom.* **2009**, *150*, 219–230. [[CrossRef](#)]
44. Vasicek, O. A Test for Normality Based on Sample Entropy. *J. R. Stat. Soc. B* **1976**, *38*, 54–59. [[CrossRef](#)]
45. Gokhale, D. On entropy-based goodness-of-fit tests. *Comput. Stat. Data Anal.* **1983**, *1*, 157–165. [[CrossRef](#)]
46. Song, K.S. Goodness-of-fit tests based on Kullback-Leibler discrimination information. *IEEE Trans. Inf. Theory* **2002**, *48*, 1103–1117. [[CrossRef](#)]
47. Lequesne, J. A goodness-of-fit test of Student distributions based on Rényi entropy. In *AIP Conference Proceedings of the 34th International Workshop on Bayesian Inference and Maximum Entropy Methods (MaxEnt'14)*; Djafari, A., Barbaresco, F., Eds.; American Institute of Physics: College Park, MD, USA, 2014; Volume 1641, pp. 487–494. [[CrossRef](#)]
48. Lequesne, J. Tests Statistiques Basés sur la Théorie de L'information, Applications en Biologie et en Démographie. Ph.D. Thesis, Université de Caen Basse-Normandie, Caen, France, 2015.
49. Girardin, V.; Regnault, P. Escort distributions minimizing the Kullback-Leibler divergence for a large deviations principle and tests of entropy level. *Ann. Inst. Stat. Math.* **2015**, *68*, 439–468. [[CrossRef](#)]
50. Kesavan, H.K.; Kapur, J.N. The Generalized Maximum Entropy Principle. *IEEE Trans. Syst. Man Cybern.* **1989**, *19*, 1042–1052. [[CrossRef](#)]
51. Borwein, J.M.; Lewis, A.S. Duality Relationships for Entropy-Like Minimization Problems. *SIAM J. Control Optim.* **1991**, *29*, 325–338. [[CrossRef](#)]
52. Borwein, J.M.; Lewis, A.S. Convergence of best entropy estimates. *SIAM J. Optim.* **1991**, *1*, 191–205. [[CrossRef](#)]
53. Borwein, J.M.; Lewis, A.S. Partially-finite programming in L_1 and the existence of maximum entropy estimates. *SIAM J. Optim.* **1993**, *3*, 248–267. [[CrossRef](#)]
54. Mézard, M.; Montanari, A. *Information, Physics, and Computation*; Oxford University Press: New York, NY, USA, 2009.
55. Darmois, G. Sur les lois de probabilités à estimation exhaustive. *C. R. l'Académie Sci.* **1935**, *200*, 1265–1966.
56. Koopman, B.O. On distributions admitting a sufficient statistic. *Trans. Am. Math. Soc.* **1936**, *39*, 399–399. [[CrossRef](#)]
57. Pitman, E.J.G. Sufficient statistics and intrinsic accuracy. *Math. Proc. Camb. Philos. Soc.* **1936**, *32*, 567–579. [[CrossRef](#)]
58. Lehmann, E.L.; Casella, G. *Theory of Point Estimation*, 2nd ed.; Springer: New York, NY, USA, 1998.
59. Mukhopadhyay, N. *Probability and Statistical Inference*, 5th ed.; Statistics: Textbooks and Monographs; Marcel Dekker: New York, NY, USA, 2000; Volume 162.
60. Rao, C.R. *Linear Statistical Inference and Its Applications*; John Wiley & Sons: New York, NY, USA, 2001.
61. Tsallis, C.; Mendes, R.M.; Plastino, A.R. The role of constraints within generalized nonextensive statistics. *Physica A* **1998**, *261*, 534–554. [[CrossRef](#)]

62. Tsallis, C. Nonextensive Statistics: theoretical, Experimental and Computational Evidences and Connections. *Braz. J. Phys.* **1999**, *29*, 1–35. [[CrossRef](#)]
63. Tsallis, C. *Introduction to Nonextensive Statistical Mechanics—Approaching a Complex World*; Springer: New York, NY, USA, 2009; [[CrossRef](#)]
64. Essex, C.; Schulzsky, C.; Franz, A.; Hoffmann, K.H. Tsallis and Rényi entropies in fractional diffusion and entropy production. *Physica A* **2000**, *284*, 299–308. [[CrossRef](#)]
65. Parvan, A.S.; Biró, T.S. Extensive Rényi statistics from non-extensive entropy. *Phys. Lett. A* **2005**, *340*, 375–387. [[CrossRef](#)]
66. Kay, S.M. *Fundamentals for Statistical Signal Processing: Estimation Theory*; Prentice Hall: Upper Saddle River, NJ, USA, 1993; Volume 1.
67. Frieden, B.R. *Science from Fisher Information: A Unification*; Cambridge University Press: Cambridge, UK, 2004.
68. Jeffrey, H. An Invariant Form for the Prior Probability in Estimation Problems. *Proc. R. Soc. A* **1946**, *186*, 453–461. [[CrossRef](#)]
69. Vignat, C.; Bercher, J.F. Analysis of signals in the Fisher-Shannon information plane. *Phys. Lett. A* **2003**, *312*, 27–33. [[CrossRef](#)]
70. Romera, E.; Angulo, J.C.; Dehesa, J.S. Fisher entropy and uncertainty like relationships in many-body systems. *Phys. Rev. A* **1999**, *59*, 4064–4067. [[CrossRef](#)]
71. Romera, E.; Sánchez-Moreno, P.; Dehesa, J.S. Uncertainty relation for Fisher information of D -dimensional single-particle systems with central potentials. *J. Math. Phys.* **2006**, *47*, 103504. [[CrossRef](#)]
72. Sánchez-Moreno, P.; González-Férez, R.; Dehesa, J.S. Improvement of the Heisenberg and Fisher-information-based uncertainty relations for D -dimensional potentials. *New J. Phys.* **2006**, *8*, 330. [[CrossRef](#)]
73. Toranzo, I.V.; Lopez-Rosa, S.; Esquivel, R.; Dehesa, J.S. Heisenberg-like and Fisher-information uncertainties relations for N -fermion d -dimensional systems. *Phys. Rev. A* **2015**, in press. [[CrossRef](#)]
74. Stam, A.J. Some Inequalities Satisfied by the Quantities of Information of Fisher and Shannon. *Inf. Control* **1959**, *2*, 101–112. [[CrossRef](#)]
75. Dembo, A.; Cover, T.M.; Thomas, J.A. Information Theoretic Inequalities. *IEEE Trans. Inf. Theory* **1991**, *37*, 1501–1518. [[CrossRef](#)]
76. Guo, D.; Shamai, S.; Verdú, S. Mutual Information and Minimum Mean-Square Error in Gaussian Channels. *IEEE Trans. Inf. Theory* **2005**, *51*, 1261–1282. [[CrossRef](#)]
77. Folland, G.B.; Sitaram, A. The uncertainty principle: A mathematical survey. *J. Fourier Anal. Appl.* **1997**, *3*, 207–233. [[CrossRef](#)]
78. Sen, K.D. *Statistical Complexity. Application in Electronic Structure*; Springer: New York, NY, USA, 2011. [[CrossRef](#)]
79. Vajda, I. χ^{α} -divergence and generalized Fisher's information. In *Transactions of the 6th Prague Conference on Information Theory, Statistics, Decision Functions and Random Processes*; Academia: Prague, Czech Republic, 1973; pp. 873–886.
80. Boekee, D.E. An extension of the Fisher information measure. Topics in information theory. In *Proceedings 2nd Colloquium on Information Theory*; Csiszár, I., Elias, P., Eds.; Series: Colloquia Mathematica Societatis János Bolyai; North Holland: Keszthely, Hungary, 1977; Volume 16, pp. 113–123.
81. Hammad, P. Mesure d'ordre α de l'information au sens de Fisher. *Rev. Stat. Appl.* **1978**, *26*, 73–84.
82. Boekee, D.E.; Van der Lubbe, J.C.A. The R -Norm Information Measure. *Inf. Control* **1980**, *45*, 136–155. [[CrossRef](#)]
83. Lutwak, E.; Yang, D.; Zhang, G. Moment-Entropy Inequalities. *Ann. Probab.* **2004**, *32*, 757–774. [[CrossRef](#)]
84. Lutwak, E.; Yang, D.; Zhang, G. Cramér-Rao and Moment-Entropy Inequalities for Rényi Entropy and Generalized Fisher Information. *IEEE Trans. Inf. Theory* **2005**, *51*, 473–478. [[CrossRef](#)]
85. Lutwak, E.; Yang, D.; Zhang, G. Moment-Entropy Inequalities for a Random Vector. *IEEE Trans. Inf. Theory* **2007**, *53*, 1603–1607. [[CrossRef](#)]
86. Lutwak, E.; Lv, S.; Yang, D.; Zhang, G. Extension of Fisher Information and Stam's Inequality. *IEEE Trans. Inf. Theory* **2012**, *58*, 1319–1327. [[CrossRef](#)]
87. Bercher, J.F. On a (β, q) -generalized Fisher information and inequalities involving q -Gaussian distributions. *J. Math. Phys.* **2012**, *53*, 063303. [[CrossRef](#)]
88. Bercher, J.F. On generalized Cramér-Rao inequalities, generalized Fisher information and characterizations of generalized q -Gaussian distributions. *J. Phys. A* **2012**, *45*, 255303. [[CrossRef](#)]
89. Bercher, J.F. On multidimensional generalized Cramér-Rao inequalities, uncertainty relations and characterizations of generalized q -Gaussian distributions. *J. Phys. A* **2013**, *46*, 095303. [[CrossRef](#)]
90. Bercher, J.F. Some properties of generalized Fisher information in the context of nonextensive thermostatics. *Physica A* **2013**, *392*, 3140–3154. [[CrossRef](#)]
91. Bregman, L.M. The relaxation method of finding the common point of convex sets and its application to the solution of problem in convex programming. *USSR Comput. Math. Math. Phys.* **1967**, *7*, 200–217. [[CrossRef](#)]
92. Nielsen, F.; Nock, R. Generalizing Skew Jensen Divergences and Bregman Divergences With Comparative Convexity. *IEEE Signal Process. Lett.* **2017**, *24*, 1123–1127. [[CrossRef](#)]
93. Ben-Tal, A.; Bornwein, J.M.; Teboulle, M. Spectral Estimation via Convex Programming. In *Systems and Management Science by Extremal Methods*; Phillips, F.Y., Rousseau, J.J., Eds.; Springer US: New York, NY, USA, 1992; Chapter 18, pp. 275–290. [[CrossRef](#)]
94. Teboulle, M.; Vajda, I. Convergence of Best ϕ -Entropy Estimates. *IEEE Trans. Inf. Theory* **1993**, *39*, 297–301. [[CrossRef](#)]
95. Girardin, V. Méthodes de réalisation de produit scalaire et de problème de moments avec maximisation d'entropie. *Stud. Math.* **1997**, *124*, 199–213. [[CrossRef](#)]

96. Girardin, V. Relative Entropy and Spectral Constraints: Some Invariance Properties of the ARMA Class. *J. Time Ser. Anal.* **2007**, *28*, 844–866. [[CrossRef](#)]
97. Costa, J.A.; Hero, A.O., III; Vignat, C. On Solutions to Multivariate Maximum α -Entropy Problems. In *4th International Workshop on Energy Minimization Methods in Computer Vision and Pattern Recognition (EMMCVPR)*; Lecture Notes in Computer Sciences; Rangarajan, A., Figueiredo, M.A.T., Zerubia, J., Eds.; Springer: Lisbon, Portugal, 2003; Volume 2683, pp. 211–226.
98. Johnson, N.L.; Kotz, S.; Balakrishnan, N. *Continuous Univariate Distributions*, 2nd ed.; John Wiley & Sons: New York, NY, USA, 1995; Volume 2
99. Chhabra, A.; Jensen, R.V. Direct determination of the $f(\alpha)$ singularity spectrum. *Phys. Rev. Lett.* **1989**, *62*, 1327. [[CrossRef](#)]
100. Beck, C.; Schögl, F. *Thermodynamics of Chaotic Systems: An Introduction*; Cambridge University Press: Cambridge, UK, 1993. [[CrossRef](#)]
101. Naudts, J. *Generalized Thermostatistics*; Springer: London, UK, 2011. [[CrossRef](#)]
102. Martínez, S.; Nicolás, F.; Pennini, F.; Plastino, A. Tsallis' entropy maximization procedure revisited. *Physica A* **2000**, *286*, 489–502. [[CrossRef](#)]
103. Chimento, L.P.; Pennini, F.; Plastino, A. Naudts-like duality and the extreme Fisher information principle. *Phys. Rev. E* **2000**, *62*, 7462–7465. [[CrossRef](#)] [[PubMed](#)]
104. Casas, M.; Chimento, L.; Pennini, F.; Plastino, A.; Plastino, A.R. Fisher information in a Tsallis non-extensive environment. *Chaos Solitons Fractals* **2002**, *13*, 451–459. [[CrossRef](#)]
105. Rudin, W. *Functional Analysis*, 2nd ed.; McGraw-Hill: New York, NY, USA, 1991.
106. Morrison, T.J. *Functional Analysis. An Introduction to Banach Space Theory*; John Wiley & Sons: New York, NY, USA, 2000.
107. Rioul, O. Information Theoretic Proofs of Entropy Power Inequalities. *IEEE Trans. Inf. Theory* **2011**, *57*, 33–55. [[CrossRef](#)]
108. Rao, C.R.; Wishart, J. Minimum variance and the estimation of several parameters. *Math. Proc. Camb. Philos. Soc.* **1947**, *43*, 280–283. [[CrossRef](#)]
109. van den Bos, A. *Parameter Estimation for Scientists and Engineers*; John Wiley & Sons: Hoboken, NJ, USA, 2007.
110. Magnus, J.R.; Neudecker, H. *Matrix Differential Calculus with Applications in Statistics and Econometrics*, 3rd ed.; John Wiley & Sons: New York, NY, USA, 1999.
111. Guo, D.; Shamaï, S.; Verdú, S. Additive Non-Gaussian Noise Channels: Mutual Information and Conditional Mean Estimation. *IEEE Int. Symp. Inf. Theory* **2005**, 719–723. [[CrossRef](#)]
112. Palomar, D.P.; Verdú, S. Gradient of Mutual Information in Linear Vector Gaussian Channels. *IEEE Trans. Inf. Theory* **2006**, *52*, 141–154. [[CrossRef](#)]
113. Verdú, S. Mismatched Estimation and Relative Entropy. *IEEE Trans. Inf. Theory* **2010**, *56*, 3712–3720. [[CrossRef](#)]
114. Barron, A.R. Entropy and the Central Limit Theorem. *Ann. Probab.* **1986**, *14*, 336–342. [[CrossRef](#)]
115. Johnson, O. *Information Theory and the Central Limit Theorem*; Imperial College Press: London, UK, 2004.
116. Madiman, M.; Barron, A. Generalized Entropy Power Inequalities and Monotonicity Properties of Information. *IEEE Trans. Inf. Theory* **2007**, *53*, 2317–2329. [[CrossRef](#)]
117. Toranzo, I.V.; Zozor, S.; Brossier, J.M. Generalization of the de Bruijn identity to general ϕ -entropies and ϕ -Fisher informations. *IEEE Trans. Inf. Theory* **2018**, *64*, 6743–6758. [[CrossRef](#)]
118. Widder, D.V. *The Heat Equation*; Academic Press: New York, NY, USA, 1975.
119. Roubíček, T. *Nonlinear Partial Differential Equations with Applications*; Birkhäuser: Basel, Switzerland, 2005.
120. Tsallis, C.; Lenzi, E.K. Anomalous diffusion: Nonlinear fractional Fokker-Planck equation. *Chem. Phys.* **2002**, *284*, 341–347. [[CrossRef](#)]
121. Vázquez, J.L. *Smoothing and Decay Estimates for Nonlinear Diffusion Equations—Equation of Porous Medium Type*; Oxford University Press: New York, NY, USA, 2006.
122. Gilding, B.H.; Kersner, R. *Travelling Waves in Nonlinear Diffusion-Convection Reaction*; Springer: Basel, Switzerland, 2004. [[CrossRef](#)]
123. Price, R. A Useful Theorem for Nonlinear Devices Having Gaussian Inputs. *IEEE Trans. Inf. Theory* **1958**, *4*, 69–72. [[CrossRef](#)]
124. Pawula, R. A modified version of Price's theorem. *IEEE Trans. Inf. Theory* **1967**, *13*, 285–288. [[CrossRef](#)]
125. Riba, J.; de Cabrera, F. A Proof of de Bruijn Identity based on Generalized Price's Theorem. *IEEE Int. Symp. Inf. Theory* **2019**, 2509–2513. [[CrossRef](#)]
126. Lieb, E.H. Proof of an Entropy Conjecture of Wehrl. *Commun. Math. Phys.* **1978**, *62*, 35–41. [[CrossRef](#)]
127. Costa, M.; Cover, T. On the Similarity of the Entropy Power Inequality and the Brunn-Minkowski Inequality. *IEEE Trans. Inf. Theory* **1984**, *30*, 837–839. [[CrossRef](#)]
128. Carlen, E.A.; Soffer, A. Entropy Production by Block Variable Summation and Central Limit Theorems. *Commun. Math. Phys.* **1991**, *140*, 339–371. [[CrossRef](#)]
129. Harremoës, P.; Vignat, C. An Entropy Power Inequality for the Binomial Family. *J. Inequalities Pure Appl. Math.* **2003**, *4*, 93.
130. Johnson, O.; Yu, Y. Monotonicity, Thinning, and Discrete Versions of the Entropy Power Inequality. *IEEE Trans. Inf. Theory* **2010**, *56*, 5387–5395. [[CrossRef](#)]
131. Haghigatshoar, S.; Abbe, E.; Telatar, I.E. A New Entropy Power Inequality for Integer-Valued Random Variables. *IEEE Trans. Inf. Theory* **2014**, *60*, 3787–3796. [[CrossRef](#)]
132. Bobkov, S.G.; Chistyakov, G.P. Entropy Power Inequality for the Rényi Entropy. *IEEE Trans. Inf. Theory* **2015**, *61*, 708–714. [[CrossRef](#)]

133. Costa, M. A New Entropy Power Inequality. *IEEE Trans. Inf. Theory* **1985**, *31*, 751–760. [[CrossRef](#)]
134. Dembo, A. Simple Proof of the Concavity of the Entropy Power with Respect to Added Gaussian Noise. *IEEE Trans. Inf. Theory* **1989**, *35*, 887–888. [[CrossRef](#)]
135. Villani, C. A Short Proof of the “Concavity of Entropy Power”. *IEEE Trans. Inf. Theory* **2000**, *46*, 1695–1696. [[CrossRef](#)]
136. Toscani, G. Heat Equation and Convolution Inequalities. *Milan J. Math.* **2014**, *82*, 183–212. [[CrossRef](#)]
137. Toscani, G. A Strengthened Entropy Power Inequality for Log-Concave Densities. *IEEE Trans. Inf. Theory* **2015**, *61*, 6550–6559. [[CrossRef](#)]
138. Ram, E.; Sason, I. On Rényi Entropy Power Inequalities. *IEEE Trans. Inf. Theory* **2016**, *62*, 6800–6815. [[CrossRef](#)]
139. Bobkov, S.G.; Marsiglietti, A. Variants of the Entropy Power Inequality. *IEEE Trans. Inf. Theory* **2017**, *63*, 7747–7752. [[CrossRef](#)]
140. Savaré, G.; Toscani, G. The Concavity of Rényi Entropy Power. *IEEE Trans. Inf. Theory* **2014**, *60*, 2687–2693. [[CrossRef](#)]
141. Zozor, S.; Puertas-Centeno, D.; Dehesa, J.S. On Generalized Stam Inequalities and Fisher–Rényi Complexity Measures. *Entropy* **2017**, *19*, 493. [[CrossRef](#)]
142. Rioul, O. Yet Another Proof of the Entropy Power Inequality. *IEEE Trans. Inf. Theory* **2017**, *63*, 3595–3599. [[CrossRef](#)]
143. Rosenblatt, M. Remarks on Some Nonparametric Estimates of a Density Function. *Ann. Math. Stat.* **1956**, *27*, 832–837. [[CrossRef](#)]
144. Parzen, E. On Estimation of a Probability Density Function and Mode. *Ann. Math. Stat.* **1962**, *33*, 1065–1076. [[CrossRef](#)]
145. Beirlant, J.; Dudewicz, E.J.; Györfi, L.; van der Meulen, E.C. Nonparametric Entropy Estimation: An Overview. *Int. J. Math. Stat. Sci.* **1997**, *6*, 17–39.
146. Leonenko, N.; Pronzato, L.; Savani, V. A Class of Rényi Information Estimators for Multidimensional Densities. *Ann. Stat.* **2008**, *36*, 2153–2182. [[CrossRef](#)]
147. Johnson, N.L.; Kotz, S.; Balakrishnan, N. *Continuous Univariate Distributions*, 2nd ed.; John Wiley & Sons: New York, NY, USA, 1995; Volume 1.
148. Corless, R.M.; Gonnet, G.H.; Hare, D.E.G.; Jeffrey, D.J.; Knuth, D.E. On the Lambert W Function. *Adv. Comput. Math.* **1996**, *5*, 329–359. [[CrossRef](#)]
149. Abramowitz, M.; Stegun, I.A. *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables*, 9th ed.; Dover: New York, NY, USA, 1970.
150. Gradshteyn, I.S.; Ryzhik, I.M. *Table of Integrals, Series, and Products*, 8th ed.; Academic Press: San Diego, CA, USA, 2015.
151. Alzahrani, F.; Salem, A. Sharp bounds for the Lambert W function. *Integral Transform. Spec. Funct.* **2018**, *29*, 971–978. [[CrossRef](#)]