

Article

A Non-Negative Measure of Information for Continuous Probability Distributions

François Xavier Machu ^{1,*}, Jeremy Cocks ¹, Ru Julie Wang ², Aziz El Kaabouchi ³, Yueqing Zhu ⁴, Maryam Lhernault ¹ and Qiuping Alexandre Wang ¹

¹ EsieaLab, Systèmes Complexes et Information Quantique, ESIEA, 9 Rue Vésale, 75005 Paris, France; jeremy.cocks@esiea.fr (J.C.); maryam.lhernault@esiea.fr (M.L.); alexandre.wang@esiea.fr (Q.A.W.)

² JOLIBRAIN, 77 Rue Pargaminières, 31000 Toulouse, France; wr0124@gmail.com

³ ESTACA, Parc Universitaire Laval-Changé, Rue Georg, 53000 Laval, France; aziz.elkaabouchi@estaca.fr

⁴ College of Mathematics and Computer Science, Wuhan Textile University, Wuhan 430079, China; zhuyy@mails.cnu.edu.cn

* Correspondence: francois.xavier.machu@esiea.fr

Abstract

In this work, we investigate the possibility of using varentropy, an information measure previously proposed for discrete probability distribution, as a measure of the probabilistic uncertainty of continuous probability distribution. We show that varentropy allows avoiding negative values and some undesirable features of informational entropy encountered while using the Boltzmann–Shannon formula (and others) for continuous probability distributions.

Keywords: information; entropy; differential entropy; measure of probabilistic uncertainty

MSC: 54C70; 82M30

1. Introduction

Entropy and information are among the most fundamental concepts in thermodynamics, statistical mechanics and information theory [1–7]. According to the common view, entropy and information (entropy for short in what follows) are two names for the same thing. They are a measure of disorder or statistical uncertainty associated with probability. The relationship between entropy and probability has been subject to rigorous mathematical study since the work of Shannon [2] and Khinchin [7]. Nevertheless, some confusion persists concerning the expression of entropy for continuous probability distributions [4,5,8].

The best known and most employed statistical expression of informational entropy is in the form of a logarithmic functional of probability distribution (Boltzmann–Shannon entropy), a form proposed for the first time by Boltzmann in his H-theorem [1] with continuous distribution of particles, then used by Gibbs in his work on statistical mechanics [3], and by Shannon in his information theory [2]. An axiomatic derivation (or proof of the uniqueness) of the formula was given in [2] by using discrete probability distributions. For a system having W discrete microstates, each having probability p_i ($i = 1, 2, \dots, W$), the Boltzmann–Shannon (BS) entropy S is given by

$$S = -\sum_{i=1}^W p_i \ln p_i \quad (1)$$



Academic Editors: Zdzisław E. Musielak and Rupam Das

Received: 27 April 2026

Revised: 5 June 2026

Accepted: 23 June 2026

Published: 30 June 2026

Copyright: © 2026 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.

(We suppose here that the Boltzmann constant $k_B = 1$). This information measure is always positive because $1 \geq p_i \geq 0$.

When the variable x of the states becomes continuous, entropy is sometimes called differential or continuous entropy given by [2–4,8]:

$$S = - \int \rho(x) \ln \rho(x) dx, \quad (2)$$

where $\rho(x)$ is the probability density distribution giving the probability $dp(x) = \rho(x)dx$ of finding the system in the state interval between x and $x + dx$, across the range of all possible states. The first use of this integral form dates back to Boltzmann [1]. Gibbs mentioned this formula in his book [3]. Shannon [2] as well intuitively took Equation (2) for granted as an analogue of Equation (1) without giving a mathematical derivation or proof of its uniqueness. As far as we know, an axiomatic derivation of Equation (2), as was done for Equation (1) by Shannon [2] and Khinchin [7], is still missing to date.

In this work, we focus on one specific mathematical property of Equation (2) concerning its sign. As mentioned above, in the case of Equation (1) with a discrete probability distribution, since p_i is positive and smaller than 1, $\ln p_i \leq 0$, which guarantees $S \geq 0$ [2]. However, in the case of a continuous probability distribution as in Equation (2), $\rho(x)$ can be larger than 1, leading to the possibility that $\ln \rho(x) > 0$ and, therefore, to negative informational entropy. The reader can see a list of continuous entropies calculated from Equation (2) for many probability density distributions, in which most entropies can be negative. The list includes some common distributions such as uniform, normal and exponential distributions [8]. This negative entropy problem can occur to many other entropy measures proposed in different contexts as generalization of BS formula (see, for example, [9]). As all these formulas contain adjustable parameters and recover the BS formula when the parameters take specific values, they risk negative values with continuous probability distribution in the same way as Equation (2). For example, the continuous Tsallis entropy $S_q = \frac{1 - \int \rho^q(x) dx}{q-1}$ [10] and the Renyi entropy $R_q = \frac{\ln \int \rho^q(x) dx}{1-q}$ [11] are both negative when $\int \rho^q(x) dx > 1$, which can happen for $\rho > 1$ with $q > 1$ and for $\rho < 1$ with $q < 1$. Most of the generalized continuous entropies containing the integral of $\int \rho^q(x) dx$ (counterpart of $\sum_i p_i^q$) [12] show negative values.

As is well known, thermodynamic entropy cannot be negative due to the third law of thermodynamics [6]. It may contain a constant in addition to the function term of probability, as proposed by Boltzmann, Gibbs and Shannon [1–3]. It is this function, depending only on probability, that assumes the role of entropy as a measure of disorder or probabilistic uncertainty. Logically, this function is expected to be zero in a non-probabilistic situation, as is the case of all the different forms of entropy, generalized or not, for discrete probability distribution; when only one state has $p_i = 1$, all other states have $p_{j \neq i} = 0$. Hence, the functions of probability of both thermodynamic entropy and informational entropy are expected to have positive lower bounds. But it is not the case with those entropy functions using continuous distribution as mentioned above. Many of them enter into contradiction with the definition of entropy as an uncertainty measure and even lose the finite lower bound. For example, the continuous entropy of the uniform distribution $\rho(x) = \frac{1}{b-a}$ in the interval $a \leq x \leq b$ is $S = \log(b-a)$, as given by Equation (2) [8,13]. We straightforwardly have $S = 0$ if $b-a = 1$, while the probabilistic uncertainty of such a distribution should not be zero. On the other hand, when $b \rightarrow a$, $\rho(x) \rightarrow \infty$, which can be considered equivalent to the Dirac delta function (see Section 3.1 below) with the normalization $\int_{-\infty}^{\infty} \rho(x) dx = 1$. The Dirac delta function is a continuous counterpart of the discrete probability distribution: $p_a = 1$ and all other $p_{x \neq a} = 0$, a situation where the probabilistic uncertainty should be zero from Equation (1). However, Equation (2) yields

$S = \log(b - a) \rightarrow -\infty$. This loss of lower bound implies that it is impossible to add some constant to S to shift it to a positive domain as proposed by Jaynes (see Section 4 below) [4]. A similar paradox happens with other continuous entropies, for instance $S = 1 - \log \lambda$ [8] of the exponential distribution $\rho(x) \sim e^{-\lambda x}$ ($0 \leq x \leq \infty$).

These undesirable features of continuous entropy raise the question whether Equation (2) is still suitable for measuring probabilistic uncertainty and whether it deserves the name entropy. Despite this uncertain situation, continuous entropies have been considered acceptable [8,13] and widely used in many applications (see [11,14,15], for example) sometimes with their negative values [14]. There are also other solutions to this negativity issue in using relative entropy or Kulback–Leibler divergence instead of Equation (2) [13,14], although this is not a solution to the issue of Equation (2) itself.

As is well known, the use of Equation (2) can be traced back to Boltzmann [1]. It has been proposed rather by intuition than by mathematical proof as has been done for Equation (1) by Shannon [2]. From a mathematical point of view, Equation (1) cannot be simply replaced by Equation (2) because, when p_i is replaced by its continuous counterpart $dp(x) = \rho dx$, a divergent term $-\ln dx \propto \ln W$ appears [4,5], which implies that the replacement of p_i by ρdx in Equation (1) is questionable. Jaynes has tried to avoid negative entropy using a continuous version of BS entropy $S = -\int_1^\infty \rho(x) \ln \frac{\rho(x)}{m(x)} dx + \ln W$ where $m(x)$ is called the invariant measure of the density of discrete values of x [4,5]. The term $\ln W$, divergent when $W \rightarrow \infty$, was simply removed, giving a continuous entropy $S^c = S + \int_1^\infty \rho(x) \ln m(x) dx$. According to Jaynes [4], it is possible to choose the invariant measure $m(x)$ in an appropriate (albeit ad hoc) way for S^c to be positive.

In what follows, we present an alternative informational entropy called varentropy, which possibly helps to avoid negative value and other undesirable properties of continuous entropy. Varentropy is an extension of the fundamental equation of equilibrium thermodynamics $\delta U = T\delta S + \delta W$, where δU is a variation of internal energy U in a reversible process, δS a variation of thermodynamic entropy, T the temperature, and δW the work done during the quasi-equilibrium equilibrium process [6]. We show that varentropy, when extended to continuous probability distributions, allows us to avoid negative informational entropy in many cases where BS formula yields negative values.

2. Definition of Varentropy

Varentropy was defined on the basis of the variational form of the entropy of the second law of thermodynamics [16,17]. This is a definition from scratch without any prerequisite or postulate about the property of entropy. The motivation was to look for a probabilistic uncertainty measure that has a sound physical background and is maximizable to generate probability distribution. This objective implies that this measure cannot be defined with a given formula because the maximization of a given functional can only yield one or two distributions. One just cannot maximize a given functional for any distribution. For example, the Shannon entropy in Equation (1), which has been widely used as a universal uncertainty measure for any probability distribution, can only be maximized for exponential or uniform distribution, depending on the constraints [4,5] and the Tsallis or Renyi entropy can only be maximized for q-exponential distributions [10,11]. Hence, this measure we looked for should be something more general and fundamental than the formulas of the different entropies [9,12]; more fundamental in that it should have some physical background relative to thermodynamic entropy; more general in that it can reproduce the different entropies for different distributions, more or less in similar ways as a differential equation of physical law generates different trajectories with different interactions, as has been shown in [16–18].

The idea of [16] was to start from the first law of equilibrium thermodynamics given by $\delta U = T\delta S + \delta W$. As is well known, in classical statistical mechanics, the internal energy is the average $\bar{E}$ of the energy E_i of all microstates i , i.e., $U = \bar{E} = \sum_{i=1}^W p_i E_i$, and the work done in an infinitesimal reversible process is given by the average of the energy change δE_i of each state: $\delta W = \overline{\delta E} = \sum_{i=1}^W p_i \delta E_i$, where $p_i = f(E_i)$ is the probability of finding the system at the state $i = 1, 2, \dots, W$ having energy E_i . The statistical expression of the fundamental equation $\delta U = T\delta S + \delta W$ becomes then $\delta \sum_{i=1}^W p_i E_i = T\delta S + \sum_{i=1}^W p_i \delta E_i$, giving a statistical expression of the variation of entropy during the process $\delta S = \frac{1}{T} \left(\delta \sum_{i=1}^W p_i E_i - \sum_{i=1}^W p_i \delta E_i \right) = \frac{1}{T} (\delta \bar{E} - \overline{\delta E})$, giving $\delta S = \frac{1}{T} \sum_{i=1}^W \delta p_i E_i$. From the above calculation, we see that this variational expression of thermodynamic entropy as a function of δp_i multiplied by the random variable energy E_i is a statistical form of the first law. For us, this expression reveals a general kinship between a measure of probabilistic uncertainty (S) and the related random variable (E_i) with its probability distribution p_i . This kinship was deeply hidden in the first and second law of thermodynamics and can only be seen when these laws are expressed in statistical form as shown above. In a previous work [10], we have extended this relationship to any single random variable x_i with its probability distribution $p_i = f(x_i)$ and defined varentropy S_V in the variational form: $\delta S_V = A \sum_{i=1}^W x_i \delta p_i$. This extension is purely mathematical; x_i can be any random variable (frequency of words, price, population, position, etc.). S_V measures the uncertainty in x_i , and its functional form and property (extensivity, additivity, concavity, etc.) are entirely determined by x_i and p_i . S_V is the thermodynamic entropy only when x_i is energy E_i .

It turns out that this variational definition of varentropy could really generate many known entropies with discrete distributions. For example, S_V has the form of the Boltzmann–Shannon entropy Equation (1) when p_i is exponential distribution, it is the nonadditive Tsallis or additive Renyi entropy if p_i is q -exponential distribution, and can take different forms for other distributions (Power law, stretched exponential, Cauchy, Gauss, etc.), and $S_V = -\frac{1 - \sum_i p_i^{1-\frac{1}{\alpha}}}{1-\frac{1}{\alpha}}$ (nonadditive) for power law for example [16–19]. A remarkable property of varentropy is that each functional form of S_V can be maximized using the calculus of variation $\delta(S_V - A\bar{x}) = 0$ to generate its original distribution. For instance, maximizing Equation (1) generates exponential distribution, maximizing Tsallis entropy generates q -exponential, and maximizing the power law varentropy $S_V = -\frac{1 - \sum_i p_i^{1-\frac{1}{\alpha}}}{1-\frac{1}{\alpha}}$ generates the power law $p_i \propto x_i^{-\alpha}$ (α positive) [16]. This property of varentropy is the meaning of the statement “maximizable measure” and is an intrinsic nature of varentropy, because its origin is in the variational definition $\delta S = \frac{1}{T} \sum_{i=1}^W \delta p_i E_i$, which is equivalent to writing $\delta S_V = A(\delta \bar{x} - \overline{\delta x})$ or $\delta(S_V - A\bar{x}) = -A\overline{\delta x}$. This means that the maximization of S_V subject to the constraint of the constant $\bar{x}$, i.e., $\delta(S_V - A\bar{x}) = 0$, implies $\overline{\delta x} = 0$. This expression can be better understood if we consider the thermodynamic case where x_i is energy E_i and $\overline{\delta x} = \overline{\delta E} = \sum_{i=1}^W p_i \delta E_i = \delta W$, which is the work in the variational process considered as the virtual work. In this way, $\delta W = 0$ can be understood as prescribed by the principle of virtual work and yields the maximum entropy calculus (Maxent principle) $\delta(S_V - A\bar{x}) = 0$. This kinship between two fundamental principles was studied in [20] with the conclusion that, in the case of thermodynamic entropy, maxent with the calculus $\delta(S - A\bar{E}) = 0$ can be considered as a law of physics derived from the fundamental principle of virtual work.

In what follows, we extend this varentropy to continuous probability replacing $p_i = f(x_i)$ by $dp(x) = \rho(x)dx$. Considering $\bar{x} = \int x\rho(x)dx$ and $\overline{\delta x} = \int \delta x\rho(x)dx$, we arrive at

$$\delta S_V = A(\delta \bar{x} - \overline{\delta x}) = A \int x \delta \rho(x) dx, \quad (3)$$

where $A \in \mathbb{R}$ is a constant to be chosen according to the nature of S_V . For example, in a reversible thermodynamic process where $x = E$, S_V is the thermodynamic entropy with $A = \frac{1}{T}$, as shown above; for the exponential decay distribution, we can choose $A = 1$ or $A = -1$, depending on the domain of x in the considered distribution (see below). It should be noticed in the above discussion of maximization of varentropy that A also plays the role of Lagrange multiplier in the maximum varentropy to generate probability distribution. It is then natural to see different A for different varentropies and distributions due to the different physics or processes. It is worth stressing that an important role of A is to guarantee S_V be positive, as required both for entropy and for any measure of statistical uncertainty, which is the main aim of this work.

Suppose f is bijective in $[a, b]$, so we can write $x = f^{-1}(\rho)$, Equation (3) becomes $\delta S_V = A \int_a^b f^{-1}[\rho(x)] \delta \rho(x) dx$. Then we define the function of the upper bound as follows: $\forall x \in [a, b]$, $F[\rho(x)] = \int_0^{\rho(x)} f^{-1}(t) dt$. According to the second fundamental theorem of calculus, F is differentiable with respect $\rho(x) \forall x \in [a, b] : \frac{d}{d\rho(x)} F[\rho(x)] = f^{-1}[\rho(x)]$, so that we can write $\delta S_V = A \int_a^b \left[\frac{d}{d\rho(x)} \int_0^{\rho(x)} f^{-1}(t) dt \right] \rho(x) dx = A \int_a^b \left[\delta \int_0^{\rho(x)} f^{-1}(t) dt \right] dx = \delta \left\{ A \int_a^b \left[\int_0^{\rho(x)} f^{-1}(t) dt \right] dx + C \right\}$, which implies

$$S_V = A \int_a^b \left[\int_0^{\rho(x)} f^{-1}(t) dt \right] dx + C, \quad (4)$$

where C is a constant.

If f is not bijective in $[a, b]$, we can cut the domain $[a, b]$ into w sub-domains $[a_k, a_{k+1}]$ with $k = 0, 1, 2, \dots, w-1$ ($a_0 = a$, $a_w = b$) in such a way that, in each sub-domain k , $f(x)$ is bijective and a local varentropy $S_V^{(k)}$ can be calculated using $S_V^{(k)} = A_k \int_{a_k}^{a_{k+1}} \left[\int_0^{\rho(x)} f^{-1}(t) dt \right] dx$. The total varentropy S_V is then the sum $S_V = \sum_k S_V^{(k)}$ [19] or:

$$S_V = \sum_{k=0}^{w-1} A_k \int_{a_k}^{a_{k+1}} \left[\int_0^{\rho(x)} f^{-1}(t) dt \right] dx + C. \quad (5)$$

If $f(x)$ is not bijective in $[a, b]$ but is differentiable, one can use the relationships $\delta \rho(x) = f'(x) \delta x$ and $xf'(x) \delta x = \delta \int_0^x tf'(t) dt$ to change the definition of varentropy in the following way: $\delta S_V = A \int x \delta \rho(x) dx = A \int_a^b xf'(x) \delta x dx = \int_a^b \left[\delta \int_0^x tf'(t) dt \right] dx$, which implies

$$S_V = A \int_a^b \left[\int_0^x tf'(t) dt \right] dx + C, \quad (6)$$

which can be used for any differentiable distribution.

3. Examples of Varentropy

3.1. Uniform Distribution

The uniform distribution is given by $\rho(x) = f(x) = \begin{cases} \frac{1}{b-a}, & a \leq x \leq b \\ 0, & x < a \text{ or } x > b \end{cases}$, which has the BS entropy given by $S = \ln(b-a)$. This entropy is zero when $b-a=1$, becomes negative in the interval $b-a < 1$ and tends to negative infinity when $b \rightarrow a$ instead of zero as expected for a deterministic case [13]. $S = \ln(b-a)$ has another issue related to the scale of x . Suppose x is a coordinate distance (idem in following discussion), logarithm of a dimensioned quantity is improper, leading to different values of entropy over the same

distance but measured in different units (for example, $S = 0$ over 1 m, but $S = \ln 100 \approx 4.61$ over 100 cm and $S < 0$ over 0.001 km). The origin of this paradox is the improper use of logarithm $\ln \rho(x)$ in Equation (2) since $\rho(x)$ has a dimension m^{-1} .

For calculating varentropy of this distribution, let's consider a linear distribution between a and b : $\rho(x) = f(x) = \begin{cases} \frac{1}{Z} \left(s(x-a) + \frac{1}{b-a} \right), & a \leq x \leq b \\ 0, & x < a \text{ or } x > b \end{cases}$ with a positive slope s , where normalization constant $Z = \int_a^b \left(s(x-a) + \frac{1}{b-a} \right) dx = \left[\frac{s}{2} (x-a)^2 + \frac{x}{b-a} \right]_a^b = \frac{s}{2} (b-a)^2 + 1$. For simplicity, we use Equation (6) where $f'(x) = \frac{s}{Z}$, leading to

$$S_V = A \int_a^b \left[\int_0^x t f'(t) dt \right] dx + C = A \int_a^b \left[\frac{s}{Z} \int_0^x t dt \right] dx = \frac{sA}{6Z} (b^3 - a^3) + C.$$

This expression of S_V does not have the dimensional problem since slope s has the dimension m^{-2} and A has the dimension m^{-1} by definition in Equation (3), yielding dimensionless S_V . As A is an arbitrary constant, let $A = 1/s$ and $C = 0$, and we obtain $S_V = \frac{1}{6Z} (b^3 - a^3)$, which becomes the varentropy of the uniform distribution when $s \rightarrow 0$ ($Z \rightarrow 1$):

$$S_V = \frac{1}{6} (b^3 - a^3),$$

which is always positive. $S_V \rightarrow 0$ when $b \rightarrow a$, which is natural for this non-probabilistic case (no uncertainty) where $\rho(x)$ is a Dirac delta function at $x = a$, i.e., $\rho(x-a) = \begin{cases} \infty & \text{if } x = a \\ 0 & \text{if } x \neq a \end{cases}$ with the normalization $\int_a^b \rho(x-a) dx = 1$. This expression of S_V can also be derived, with a bit longer computation, from Equation (4), which we will use for the following bijective distribution functions.

3.2. Exponential Distributions

Exponential distribution is given by $\rho(x) = \frac{1}{Z} e^{-\alpha x}$ for $0 < x < \infty$ and $\alpha > 0$, Z being the normalization constant. Its BS entropy has been calculated and reads $S = 1 - \ln \alpha$ [6], which is inevitably negative when α is sufficiently large.

Now we calculate varentropy with Equation (4) using $f^{-1}(t) = \frac{\ln(Zt)}{-\alpha}$. It is straightforward to calculate $\int_0^{\rho(x)} \frac{\ln(Zt)}{-\alpha} dt = \frac{\rho(x) \ln(Z\rho(x)) - \rho(x)}{-\alpha}$, so $S_V = -\frac{A}{\alpha} \int_0^\infty [\rho(x) \ln(Z\rho(x)) - \rho(x)] dx + C$. Let $C = -\frac{A}{\alpha} \int_0^\infty \rho(x) dx = -\frac{A}{\alpha}$ and $A = 1$, and we get

$$S_V = -\frac{1}{\alpha} \int_0^\infty \rho \ln Z \rho dx,$$

which is different from Equation (2). There is no problem of improper use of logarithm because $Z\rho$ is dimensionless. The positive value of this expression can be seen by substituting $\rho(x) = \frac{1}{Z} e^{-\alpha x}$ into the integral to obtain $S_V = -\frac{1}{\alpha} \int_0^\infty \frac{1}{Z} e^{-\alpha x} (-\alpha x) dx = \frac{1}{\alpha} > 0$.

In order to show the role of the constant A to guarantee positive S_V , let us suppose an increasing exponential distribution $\rho(x) = \frac{1}{Z} e^{-\alpha x}$ for $-\infty < x \leq 0$ with $\alpha < 0$. After the same calculation as above, we reach $S_V = \frac{A}{\alpha}$. As $\alpha < 0$, we can choose $A = -1$ to have a positive entropy $S_V = -\frac{1}{\alpha}$.

3.3. Stretched Exponential Distribution

We consider the continuous stretched exponential distribution $\rho(x) = \frac{1}{Z}e^{-x^\beta}$ for positive $0 < x < \infty$ and $\beta > 0$, Z being the normalization constant $Z = \int_0^\infty e^{-x^\beta} dx$. Let us first calculate its entropy S_{BS} :

$$S = - \int_0^\infty \frac{1}{Z} e^{-x^\beta} \ln \left(\frac{1}{Z} e^{-x^\beta} \right) dx = \frac{\ln Z}{Z} \int_0^\infty e^{-x^\beta} dx + \frac{1}{Z} \int_0^\infty x^\beta e^{-x^\beta} dx.$$

Considering the change in variable $t = x^\beta$ and the definition of Gamma function, we obtain $S = \frac{\ln Z}{Z\beta} \Gamma\left(\frac{1}{\beta}\right) + \frac{1}{Z\beta} \Gamma\left(\frac{1}{\beta} + 1\right)$, which becomes, with the equality $\Gamma(x+1) = x\Gamma(x)$, $S = \frac{1}{Z\beta} \Gamma\left(\frac{1}{\beta}\right) [\ln Z + \frac{1}{\beta}]$. Considering the normalization constant $Z = \frac{1}{\beta} \Gamma\left(\frac{1}{\beta}\right)$, we have:

$$S = \frac{1}{\beta} - \ln \beta + \ln \left(\Gamma\left(\frac{1}{\beta}\right) \right),$$

which is *negative* whenever $\ln \beta > \frac{1}{\beta} + \ln \left(\Gamma\left(\frac{1}{\beta}\right) \right)$.

Now let us see the varentropy of the continuous stretched exponential distribution $\delta S_V = A \int_0^\infty x \delta \rho(x) dx$. Introducing the change in variable $x = t^{\frac{1}{\beta}}$, we have $\delta \rho(x) = \delta \left(\frac{1}{Z} e^{-t} \right) = -\frac{1}{Z} e^{-t} \delta t$ and $\delta S_V = -\frac{A}{Z} \int_0^\infty t^{\frac{1}{\beta}} e^{-t} \delta t dx = -\frac{A}{Z} \int_0^\infty \delta \left[\int_0^t \tau^{\frac{1}{\beta}} e^{-\tau} d\tau + C \right] dx = -\frac{A}{Z} \delta \left[\int_0^\infty \gamma\left(\frac{1}{\beta} + 1, t\right) dx + C \right]$, where C is a constant of integration and $\gamma\left(\frac{1}{\beta} + 1, t\right) = \int_0^t \tau^{\frac{1}{\beta}} e^{-\tau} d\tau$ is the lower incomplete gamma function. As $\rho(x)$ is bijective, $t = x^\beta = \ln \frac{1}{Z\rho(x)}$, the varentropy reads

$$S_V = -\frac{A}{Z} \left[\int_0^\infty \gamma\left(\frac{1}{\beta} + 1, \ln \frac{1}{Z\rho(x)}\right) dx + C \right].$$

Let us choose $A = -1$ and $C = 0$, the stretched exponential varentropy reads

$$S_V = \frac{\beta}{\Gamma\left(\frac{1}{\beta}\right)} \int_0^\infty \gamma\left(\frac{1}{\beta} + 1, x^\beta\right) dx,$$

where we used $Z = \frac{1}{\beta} \Gamma\left(\frac{1}{\beta}\right)$, and $S_V \geq 0$ because $\beta > 0$, $\Gamma\left(\frac{1}{\beta}\right) > 0$, and $\gamma\left(\frac{1}{\beta} + 1, x^\beta\right) > 0$.

3.4. Continuous Normal Distribution

We consider the continuous normal distribution $\rho(x) = \frac{1}{Z} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$ for any $x \in \mathbb{R}$, where $Z = \sigma\sqrt{2\pi}$ is the normalization constant, and μ is the mean. Its BS entropy has been calculated and reads $S = \ln(\sigma\sqrt{2\pi e})$ [6], which is negative when $\sigma\sqrt{2\pi e} < 1$.

Before calculating its varentropy, let us remind that the normal distribution is not a bijective function. According to Equation (5), we must separate the domain of x into two parts, $-\infty \leq x - \mu \leq 0$ and $0 \leq x - \mu \leq \infty$, in each of which the normal distribution is bijective. Let $y = x - \mu$ and $\rho(y) = \frac{1}{Z} e^{-\frac{y^2}{2\sigma^2}}$, we have $\delta S_V = A \int_{-\infty}^\infty (y + \mu) \delta \rho(y) dy = A \int_{-\infty}^0 y \delta \rho(y) dy + A \int_0^\infty y \delta \rho(y) dy + A \int_{-\infty}^\infty \mu \delta \rho(y) dy$. The third term should be zero since $\int_{-\infty}^\infty \mu \delta \rho(y) dy = \mu \delta(1) = 0$ due to the normalization $\int_{-\infty}^\infty \rho(y) dy = 1$.

Now let us make a change in variable $t = \left(\frac{y}{\sqrt{2}\sigma}\right)^2 \geq 0$ or $y = \pm \sqrt{2}\sigma t^{1/2}$ and write $\rho(t) = \frac{1}{Z} e^{-t}$. The first term of varentropy is $A \int_{-\infty}^0 y \delta \rho(y) dy = -A \int_0^\infty \sqrt{2}\sigma t^{1/2} \delta \left(\frac{1}{Z} e^{-t} \right) dy = -\frac{A}{Z} \delta \int_0^\infty \sqrt{2}\sigma \gamma\left(\frac{3}{2}, \ln \frac{1}{Z\rho(y)}\right) dx$ where we used $t = \ln \frac{1}{Z\rho(y)}$. The second term gives the

same result as the first one. So $\delta S_V = -\frac{2A}{\sqrt{\pi}} \delta \int_0^\infty \gamma\left(\frac{3}{2}, \ln \frac{1}{\sigma\sqrt{2\pi\rho(x)}}\right) dx$, which implies $S_V = -\frac{2A}{\sqrt{\pi}} \int_0^\infty \gamma\left(\frac{3}{2}, \ln \frac{1}{\sigma\sqrt{2\pi\rho(y)}}\right) dy + C$. Let $A = -1$ and $C = 0$, we get

$$S_V = \frac{2}{\sqrt{\pi}} \int_0^\infty \gamma\left(\frac{3}{2}, \frac{y^2}{2\sigma^2}\right) dy,$$

which is always positive and independent from μ .

3.5. Power Law Distribution

A typical example of continuous power law is the Pareto law $\rho(x) = \frac{\beta}{x^{\beta+1}}$, for $1 < \beta < \infty$. The continuous Boltzmann–Gibbs entropy S_{BG} is calculated as follows [6]:

$$S = -\int_1^\infty \rho(x) \ln \rho(x) dx = -\ln \beta + 1 + \frac{1}{\beta}.$$

The negative value of BS entropy takes place in the interval $\ln \beta > 1 + \frac{1}{\beta}$ or $\beta > 3.59$.

Now let us calculate the varentropy for the power law distribution $\rho(x) = \frac{1}{Z} x^{-\frac{1}{b}}$. From the definition of varentropy, we write $\delta S_V = A \int (Z\rho)^{-b} \delta \rho dx = A \int \delta \left(\frac{Z^{-b}}{1-b} \rho^{1-b} \right) dx = \delta \left\{ \frac{A}{1-b} \int (Z^{-b} \rho^{1-b} - m) dx \right\}$. Let $A = 1$, the continuous varentropy reads

$$S_V = \int \rho \frac{(Z\rho)^{-b} - m}{1-b} dx,$$

where the function m is such that $\int \rho m(x) dx = C$ is a constant of the variation, i.e., $\delta C = 0$.

For Pareto PDF $\rho(x) = \frac{\beta}{x^{\beta+1}}$ with $\beta = \frac{1}{b} - 1$, $x_{\min} = 1$ and $x_{\max} = \infty$, we obtain

$$S_V = \frac{\int_1^\infty \beta^{-(\beta+1)} \rho^{\beta/(\beta+1)} dx - C}{\beta/(\beta+1)} = \frac{\beta+1}{\beta} \left(\frac{\beta}{\beta-1} - C \right).$$

Let $C = 1$, we obtain

$$S_V = \frac{\beta+1}{\beta(\beta-1)},$$

which is always positive for $1 < \beta < \infty$ as plotted in Figure 1 as a function of β .

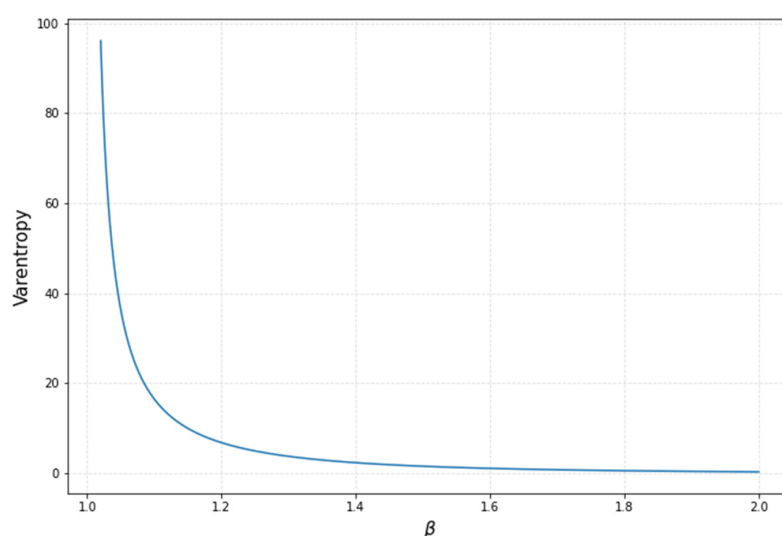


Figure 1. Evolution of $S_V = \frac{\beta+1}{\beta(\beta-1)}$, which decreases from infinity to zero with increasing β in the intervals $1 < \beta < \infty$.

4. Discussion

To summarize, despite some undesirable features, the continuous entropies, as a heritage of the long history of statistical mechanics and information theory, continue to be considered as a measure of the uncertainty in continuous random variables, sometimes with the help of relative entropies to avoid negative values [11,13–15]. We have proposed here an alternative measure called varentropy to avoid several undesirable features of continuous entropies. Examples of varentropy for several well-known continuous probabilities have been calculated from its variational definition, showing the following features with respect to continuous entropies.

1. Varentropy S_V is positive for the distributions studied in this work. The continuous entropies of these distributions can be negative.
2. S_V is zero for deterministic case as expected for a measure of probabilistic uncertainty, while continuous entropy goes to minus infinity.
3. S_V can avoid the improper use of logarithm $\ln \rho$ because the probability density distribution is a dimensioned quantity. Other generalized continuous entropies [10,12] containing the term $\sum_i \rho_i^q$ have the same undesirable feature: loss of scale invariance for example.
4. S_V has a sound physical background because it is defined with a variational equation, which is just the statistical form of the first law of thermodynamics. Different from other entropies defined with given functional formulas, it has great flexibility to generate different functionals for different distributions.
5. As each given formula of varentropy is maximized for its distribution, it is the optimal measure of the uncertainty of that distribution, meaning that its value is always the largest one among all the possible measures (Shannon, Tsallis, etc.) for the same distribution. A case study of this feature was presented in [19,21].

Further investigation is necessary to confirm these features of varentropy for other continuous distributions.

One should have noticed that, in this work, for varentropy to be positive, a constant A must be chosen according to the kind of distribution function. The existence of such a constant is quite natural in such a general definition of a quantity whose nature differs in different situations. Such a constant is also in the Boltzmann–Shannon formula, written as $S = -A \sum_{i=1}^W p_i \ln p_i$. For a binary system ($W = 2$), the choice of $A = 1$ defines the unit of the information measure in bit. When the formula is used for ideal gas to measure the probabilistic uncertainty of the distribution of internal energy, we must write $A = k_B$, the Boltzmann constant, for S to be the thermodynamic entropy of the gas. But it is impossible to choose a single constant A to make continuous entropies always positive because, as discussed above in Section 3, they can be positive and negative for a given distribution, and sometimes go to minus infinity.

Finally, we would like to mention that we need information measures for continuous probability distributions in both classical and quantum physics. An example is the calculation of the path entropy [22,23] of random dynamics using path integral method [24] and the classical path probability [23] or the quantum propagator [24], which are both continuous exponential functions of the classical action [25,26]; hence, the use of the BS formula (or other generalized ones) risks giving negative values. The path entropy is an increasing function of time of random motion [23,25,26] and an important ingredient in the study of the irreversible processes in both classical and quantum world [22].

Author Contributions: Conceptualization, Q.A.W.; Validation, F.X.M., J.C., R.J.W., A.E.K., Y.Z., M.L. and Q.A.W.; Formal analysis, R.J.W., A.E.K., Y.Z. and Q.A.W.; Investigation, F.X.M., R.J.W., A.E.K., Y.Z., M.L. and Q.A.W.; Resources, R.J.W. and Y.Z.; Writing—original draft, F.X.M., J.C., R.J.W. and

Q.A.W.; Writing—review and editing, F.X.M. and Q.A.W.; Supervision, F.X.M. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: The original contributions presented in this study are included in the article. Further inquiries can be directed to the corresponding author.

Conflicts of Interest: Author Ru Julie Wang was employed by the company JOLIBRAIN, 77 rue Pargaminières, 31000, Toulouse, France. The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential of conflict of interest.

References

1. Boltzmann, L. Weitere Studien über das Wärmegleichgewicht unter Gasmolekülen. *Sitzungsberichte Akad. Der Wiss.* **1872**, *66*, 275–370.
2. Shannon, C.E. A Mathematical Theory of Communication. *Bell Syst. Tech. J.* **1948**, *27*, 379. [CrossRef]
3. Gibbs, J.W. *Elementary Principles in Statistical Mechanics*; Charles Scribner's Sons: New York, NY, USA, 1902.
4. Jaynes, E.T. Gibbs vs Boltzmann Entropies. *Am. J. Phys.* **1965**, *33*, 391. [CrossRef]
5. Jaynes, E.T. Information theory and statistical mechanics. In *Statistical Physics*; Brandeis University Summer Institute Lectures in Theoretical Physics; W. A. Benjamin, Inc.: New York, NY, USA, 1963; Volume 3, pp. 181–218.
6. Müller, I.; Müller, W.H. *Fundamentals of Thermodynamics and Applications*; Springer: Berlin/Heidelberg, Germany, 2009.
7. Khinchin, A.I. *Mathematical Foundations of Information Theory*; Dover: New York, NY, USA, 1957.
8. Available online: https://en.wikipedia.org/wiki/Differential_entropy (accessed on 30 January 2025).
9. Amigó, J.M.; Balogh, S.G.; Hernández, S. A Brief Review of Generalized Entropies. *Entropy* **2018**, *20*, 813. [CrossRef] [PubMed]
10. Alomani, G.; Kayid, M. Further Properties of Tsallis Entropy and Its Application. *Entropy* **2023**, *25*, 199. [CrossRef] [PubMed]
11. Alawady, M.A.; Barakat, H.M.; Al Luhayb, A.S.M.; Mansour, G.M. Some properties on dynamic cumulative Tsallis residual entropy measures based on Sarmanov family with applications to motor data. *AIMS Math.* **2026**, *11*, 8271–8307. [CrossRef]
12. Esteban, M.D.; Morales, B. A Summary on Entropy Statistics. *Kybernetika* **1995**, *31*, 337.
13. Rioul, O. This is IT: A Primer on Shannon's Entropy and Information. In *L'Information, Séminaire Poincaré 2018*; Birkhäuser: Basel, Switzerland, 2021; Volume XXIII, pp. 43–77.
14. Grassucci, E.; Communiello, D.; Uncini, A. An Information-Theoretic Perspective on Proper Quaternion Variational Autoencoders. *Entropy* **2021**, *23*, 856. [CrossRef] [PubMed]
15. Via, J.; Ramirez, D.; Santamaria, I. Properness and Widely Linear Processing of Quaternion Random Vectors. *IEEE Trans. Inf. Theory* **2010**, *56*, 3502–3515. [CrossRef]
16. Wang, Q.A. Probability distribution and entropy as a measure of uncertainty. *J. Phys. A Math. Theor.* **2008**, *41*, 065004. [CrossRef]
17. Ou, C.; El Kaabouchi, A.; Nivanen, L.; Chen, J.; Tsobnang, F.; Le Méhauté, A.; Wang, Q.A. Maximizable Informational Entropy as a Measure of Probabilistic Uncertainty. *Int. J. Mod. Phys. B* **2010**, *24*, 3461–3468. [CrossRef]
18. Abe, S. Generalized entropy optimized by a given arbitrary distribution. *J. Phys. A Math. Gen.* **2003**, *36*, 8733. [CrossRef]
19. Jiang, J.; Metz, F.; Beck, C.; Lefevre, S.; Chen, J.C.; Pezeril, M.; Wang, Q.A. Double power law degree distribution and informational entropy in urban road networks. *Int. Mod. Phys. C* **2011**, *22*, 33. [CrossRef]
20. Wang, Q.A.; Ye, Q. Derivation of the Maximum Entropy Principle from the Virtual Work Principle. *Eur. Phys. J. Plus* **2025**, *140*, 1234. [CrossRef]
21. El Kaabouchi, A.; Machu, F.X.; Cocks, J.; Wang, R.; Zhu, Y.Y.; Wang, Q.A. Study of a measure of efficiency as a tool for applying the principle of least effort to the derivation of the Zipf and Pareto laws. *Adv. Complex Syst.* **2021**, *24*, 2150013. [CrossRef]
22. Davis, S.; Gonzales, S. Hamiltonian formalism and path entropy maximization. *J. Phys. A Math. Theor.* **2015**, *48*, 425003. [CrossRef]
23. Lin, T.; Wang, R.; Bi, W.P.; El Kaabouchi, A.; Pujos, C.; Calvayrac, F.; Wang, Q.A. Path probability distribution of stochastic motion of non dissipative systems: A classical analog of Feynman factor of path integral. *Chaos Solitons Fractals* **2013**, *57*, 129. [CrossRef]
24. Feynman, R.P.; Hibbs, A.R. *Quantum Mechanics and Path Integrals*; McGraw-Hill Publishing Company: New York, NY, USA, 1965.
25. Wang, Q.A.; El Kaabouchi, A. From Random Motion of Hamiltonian Systems to Boltzmann H Theorem and Second Law of Thermodynamics—A Pathway by Path Probability. *Entropy* **2014**, *16*, 885. [CrossRef]
26. General, I. Principle of maximum caliber and quantum physics. *Phys. Rev. E* **2018**, *98*, 012110. [CrossRef] [PubMed]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.