A non-negative informational entropy for continuous probability distribution

Ru Wang, François-Xavier Machu, Aziz El Kaabouchi, Qiuping A. Wang

A non-negative informational entropy for continuous probability distribution

R. Wang[1], F.X. Machu[1], A. El Kaabouchi[3] and Q.A. Wang[1,2]

[1]Laboratoire SCIQ, ESIEA, 9 Rue Vésale, Paris, France

[2]IMMM, CNRS UMR 6283, Le Mans Université, Ave. O. Messiaen, Le Mans, France

[3]ESTACA, Parc universitaire Laval-Changé, Rue Georges Charpak, Laval, France

Abstract

In this work, we propose to use varentropy, an information measure defined from a generalization of thermodynamic entropy, for the calculation of informational entropy in order to avoid negative entropy in case of continuous probability distribution.

Keywords: Information; Entropy; Differential entropy; Measure of probabilistic uncertainty

1) Introduction

Thermodynamic entropy and information are among the most fundamental concepts in thermodynamics, statistical mechanics and information theory [1][2][3][4][5][6][7]. According to a common view, thermodynamic entropy and information (entropy for short in what follows) are two names of a same thing, they are a measure of disorder, or statistical uncertainty in the presence of 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 distribution [4][5][8].

The best known and most employed statistical expression of informational entropy is in a form of logarithmic functional of probability distribution, a form proposed for the first time by Boltzmann in the study of H-theorem [1], subsequently by Gibbs in his work on statistical mechanics [3], and by Shannon in his information theory [2]. For a system having W discrete microstates, each having the probability p_i ($i = 1,2 \ldots W$), the Boltzmann-Gibbs (BG) entropy S_{BG} is given by

$$S_{BG} = -\sum_{i=1}^{W} p_i \ln p_i \tag{1}$$

(Suppose Boltzmann constant $k_B = 1$). This information measure is always positive because $1 \geq p_i \geq 0$.

When the states become continuous with a variable x characterizing the states in the phase (state) space, entropy is sometime called differential or continuous entropy and given by [2][3][4][8] :

$$S_{BG} = -\int \rho(x) \ln \rho(x) dx \tag{2}$$

where $\rho(x)$ is the probability density distribution giving the probability $dp(x) = \rho(x)dx$ of finding the system in the state interval $x \to x + dx$, the integral being carried out over the interval including all the possible states. The first use of this integral form dates back to Boltzmann [1], Gibbs [3], as well as Shannon [2] who intuitively took Eq.(2) for granted as a analogue of Eq.(1) without giving mathematical proof . As far as we know, the derivation of Eq.(2) from first principles, just as has been done for Eq.(1) by Shannon [2] and Khinchin [7], is still missing to date.

In this work, we focus on a specific mathematical problem of Eq.(2) concerning the sign of the continuous entropy. As mentioned above, in the case of discrete probability with Eq.(1), as p_i is positive and smaller than unity, so $\ln p_i \leq 0$ to guarantee $S_{BG} \geq 0$ [2]. However, in the case of continuous probability distribution with Eq.(2), $\rho(x)$ can be larger than unity, leading to $\ln\rho(x) > 0$, and to negative informational entropy. The reader can see a list of continuous entropies calculated from Eq.(2) for many probability density distributions, in which most entropies can be negative, including those for some common and ordinary distributions such as uniform, normal and exponential distributions [8].

As well known, the thermodynamic entropy cannot be negative due to the third law of thermodynamics [6]. In statistical mechanics and information theory, informational entropy is regarded as a measure of disorder or uncertainty, negative information or entropy is senseless. As a matter of fact, Eq.(1) cannot be simply replaced by Eq.(2) because when p_i is replaced by the continuous counterpart $dp(x) = \rho dx$, a divergent term $-\ln dx \propto \ln W$ takes place [4][5]. This term implies that the shift from p_i to ρdx in Eq.(1) is questionable. Jaynes has tried to avoid negative entropy using a continuous version of Boltzmann-Shannon entropy $S_{BG}^C = -\int_1^\infty \rho(x)\ln\frac{\rho(x)}{m(x)}dx + \ln W$ where $m(x)$ is called invariant measure of the density of discrete values of x [7]. The term $\ln W$, divergent when $W \to \infty$, is simply removed, giving a continuous entropy $S_{BG}^C = S_{BG} + \int_1^\infty \rho(x)\ln m(x)dx$. According to Jaynes [4], it is possible to choose the invariant measure $m(x)$ in an appropriate (albeit ad hoc) way for S_{BG}^C to be positive.

In what follows, we present an alternative solution to this mathematical problem. The starting point is a definition of informational entropy as an extension of the fundamental equation of 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 process [6]. We show that varentropy allows avoiding negative informational entropy. This is the main objective of this work.

2) Definition of varentropy

Varentropy is defined by mimicking the variational form of the entropy of the second law of thermodynamics [9][10]. This is a definition from scratch without any prerequisite or postulate on the property of entropy. The motivations was to look for an uncertainty measure that is optimal or maximized for any probability distribution. Although S_{BG} has been widely used as a universal uncertainty measure for any probability distribution, it is only maximized for exponential and uniform distribution. The question is whether it is possible to define a more general measure of probabilistic uncertainty that is optimal (maximized) for other distributions as well, and recovers S_{BG} for exponential distribution.

The origin of the approach is the thermodynamic equation $\delta U = T\delta S_{BG} + \delta W$. As 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 in the infinitesimal reversible process is given by the average of the energy change δE_i of each state: $\delta W = \overline{\delta E} = \sum_{i=1}^{W} p_i \delta E_i$. The statistical expression of the fundamental equation $\delta U = T\delta S_{BG} + \delta W$ becomes then $\delta \sum_{i=1}^{W} p_i E_i = T\delta S_{BG} + \sum_{i=1}^{W} p_i \delta E_i$, giving a statistical expression of the variation of entropy δS_{BG} during the process with $\delta S_{BG} = \frac{1}{T}(\delta \sum_{i=1}^{W} p_i E_i - \sum_{i=1}^{W} p_i \delta E_i) = \frac{1}{T}\left(\delta\bar{E} - \overline{\delta E}\right)$. This straightforwardly leading to $\delta S_{BG} = \frac{1}{T}\sum_{i=1}^{W} \delta p_i E_i$. For continuous distribution when p_i becomes $dp(x) = \rho dx$, $\bar{E} = \int E\,\rho(E)dE$, $\overline{\delta E} = \int \delta E\,\rho(E)dE$, and we have $\delta S_{BG} = \int E\,\delta\rho(E)dE$.

In the above statistical expressions of thermodynamic entropy, the random variable is the energy E of the microstates. In a previous work [9], we have extended this expression of thermodynamic entropy in order to define a measure of statistical uncertainty for any single random variable, say, x. This measure has been called varentropy S_V since it is defined in a variational form:

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

where $A \in \mathbb{R}$ is a constant to choose 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 of the considered distribution (see below). It is worth stressing that an important role of A is to guarantee S_V positive, as expected for entropy or any measure of statistical uncertainty, which is the main aim of this work.

Varentropy turns out to be equivalent to a generalized entropy defined in [11] by mathematical consideration from the principle of maximum entropy. From Eq.(3), we can write $\delta(S_V - A\bar{x}) = -A\overline{\delta x}$. The maximization of S_V subject to the constraint of the constant $\bar{x}$, i.e., $\delta(S_V - A\bar{x}) = 0$, implies that $\overline{\delta x} = 0$, which has been interpreted as a probabilistic extension of the principle of virtual work in the case where the random variable x is the energy of the system under consideration [12]. It is worth mentioning once again that varentropy was motivated by the search for an uncertainty measure that is maximized for any probability distribution [9][11], implying that S_V should be equal to S_{BG} for exponential distribution and larger than S_{BG} for other distributions. Although a general proof of this property of S_V is still missing, the reader can find some examples of this advantage of S_V compared with S_{BG} in [13] and [14].

3) Examples of varentropy

a) Exponential distributions

The continuous entropy of exponential distribution $\rho(x) = \frac{1}{Z} e^{-\alpha x}$ for positive $0 < x < \infty$ and $\alpha > 0$, Z being the normalization constant. Its entropy has been calculated and reads $S_{BG} = 1 - ln\alpha$ [6], which is inevitably *negative* when α is sufficiently large.

Now with varentropy Eq.(3), it is straightforward to get $\delta S_V = A \int_0^\infty \frac{\ln(Z\rho)}{-\alpha} \delta\rho dx = -\frac{A}{\alpha}\int_0^\infty \delta[\rho \ln(Z\rho)]\, dx = \delta\left[-\frac{A}{\alpha}\int_0^\infty \rho \ln(Z\rho)\, dx\right]$ which implies

$$S_V = -\frac{A}{\alpha}\int_0^\infty \rho \ln Z\rho \, dx + C \tag{4}$$

where C is an arbitrary constant, we can choose C=0, leading to the expression of conventional Boltzmann-Gibbs entropy. The positivity of this expression can be seen by substituting $\rho(x) = \frac{1}{Z} e^{-\alpha x}$ into Eq.(4), which yields $S_V = -\frac{A}{\alpha}\int_0^\infty \frac{1}{Z} \mathrm{e}^{-\alpha x}(-\alpha x) dx = \frac{A}{Z}\int_0^\infty x e^{-\alpha x} dx = \frac{A}{\alpha}$. Let $A = 1$, $S_V = \frac{1}{\alpha}$ which is positive.

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 $S_V = -\frac{1}{\alpha}$.

c) 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 BG entropy:

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

Considering the change of variable $t = x^\beta$, and the definition of Gamma function, we obtain:

$$S_{BG} = \frac{\ln Z}{Z\beta} \Gamma\left(\frac{1}{\beta}\right) + \frac{1}{Z\beta} \Gamma\left(\frac{1}{\beta} + 1\right).$$

which becomes, with the equality $\Gamma(x+1) = x\Gamma(x)$:

$$S_{BG} = \frac{1}{Z\beta} (\Gamma\left(\frac{1}{\beta}\right) [\ln Z + \frac{1}{\beta}].$$

Considering the normalization constant $Z = \frac{1}{\beta} \Gamma\left(\frac{1}{\beta}\right)$, we have:

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

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

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 of 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 $t = x^\beta = \ln\frac{1}{Z\rho(x)}$, 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

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

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 $\gamma\left(\frac{1}{\beta}+1, \ln\frac{1}{Z\rho(x)}\right) = \gamma\left(\frac{1}{\beta}+1, x^{\beta}\right)$>0.

d) Continuous normal distribution

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

Before calculating its varentropy, let us make a change of variable $t = \left(\frac{x-\mu}{\sqrt{2}\sigma}\right)^2$, with $\rho(t) = \frac{1}{Z}e^{-t}$. Its varentropy is $\delta S_V = A\int_0^\infty x\delta\rho(x)dx = A\int_0^\infty(\sqrt{2}\sigma t^{1/2}+m)\delta(\frac{1}{Z}e^{-t})dx$. The result is $S_V = -\frac{A}{Z}\int_0^\infty \sqrt{2}\sigma\gamma\left(\frac{3}{2}, \ln\frac{1}{Z\rho(x)}\right)dx - A\mu + C$. Let $A = -1$ and $C = 0$, we get

$$S_V = \frac{\gamma}{\sqrt{\pi}}\int_0^\infty\left(\frac{3}{2}, \ln\frac{1}{\sigma\sqrt{2\pi}\rho(x)}\right)dx + \mu$$

which should always be positive.

e) Power law distribution

We consider here 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] :

$$S_{BG} = -\int_1^\infty \rho(x)\ln\rho(x)\,dx = -\ln\beta + 1 + \frac{1}{\beta} \tag{5}$$

The trouble takes place for the interval $\ln\beta > 1 + \frac{1}{\beta}$ or $\beta > 3.59$ where S_{BG} becomes negative.

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

$$S_V = \int\rho\frac{(Z\rho)^{-b} - m}{1-b}dx \tag{6}$$

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 $x_{max} = \infty$, Eq.(7) gives

$$S_V = \frac{\int_1^{\infty} \beta^{-(\beta+1)} \rho^{\beta/(\beta+1)} dx - C}{\beta/(\beta+1)} = \frac{\beta+1}{\beta}\left(\frac{\beta}{\beta-1} - C\right), \tag{7}$$

Let $C = 1$, we obtain a positive varentropy

$$S_V = \frac{\beta + 1}{\beta(\beta - 1)} \tag{8}$$

which is always positive for $1 < \beta < \infty$ as plotted in **Figure 4** as a function of β.

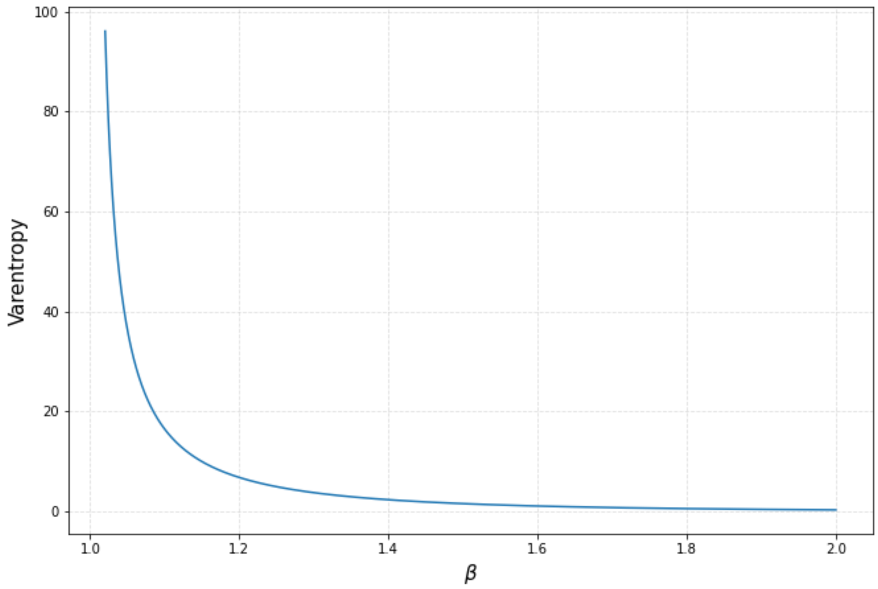

Figure 4. Evolution of $S_V = \frac{\beta+1}{\beta} \beta^{-\frac{\beta^2+\beta+1}{\beta+1}} \cdot \frac{1}{\beta-1}$ which decreases from infinity to zero with increasing β in the intervals $1 < \beta < \infty$.

4) Conclusion

We have proposed a solution of the problem of negative value of informational entropy for continuous probability distribution with the help of varentropy which has been previously defined as a general maximizable measure of probabilistic uncertainty. We have shown that varentropy can be used to avoid negative informational entropy of continuous probability distribution. Several examples of varentropy for some well-known continuous probability has been calculated.

References

[1] L. Boltzmann, Weitere Studien über das Wärmegleichgewicht unter Gasmolekülen. Sitzungsberichte Akademie der Wissenschaften 66 (1872) 275-370. English translation: L. Boltzmann, Further Studies on the Thermal Equilibrium of Gas Molecules. The Kinetic Theory of Gases. History of Modern Physical Sciences. 1. pp. 262–349.

[2] C.E. Shannon, A Mathematical Theory of Communication. Bell System Technical Journal, **27** (1948) 379.

[3] J.W. Gibbs, Elementary Principles in Statistical Mechanics. New York: Charles Scribner's Sons (1902)

[4] E.T. Jaynes, Gibbs vs Boltzmann Entropies. American Journal of Physics. **33** (1965) 391.

[5] E.T. Jaynes, Information theory and statistical mechanics, *Brandeis University Summer Institute Lectures in Theoretical Physics.* ***3*** *(1963)181-218*

[6] I. Müller, W.H. Müller, Fundamentals of thermodynamics and applications, Springer-Verlag, Berlin, 2009

[7] A. I. Khinchin, Mathematical Foundations of Information Theory, Dover, New York, 1957.

[8] See *https://en.wikipedia.org/wiki/Differential_entropy* and references therein.

[9] Q.A. Wang, J. Phys. A: Math. Theor. **41**(2008)065004

[10] Ou Congjie, El Kaabouchi, Nivanen Laurent Maximizable Informational Entropy as a Measure of Probabilistic Uncertainty, International Journal of Modern Physics B, **24**(2010)3461-3468 arXiv:0803.3110

[11] S. Abe, Generalized entropy optimized by a given arbitrary distribution, J. Phys. A: Math. Gen. **36**(2003)8733

[12] Q.A. Wang, From the principle of virtual work to the principle of maximum entropy, arXiv:0704.1076 and arXiv:0712.2583.

[13] J. Jiang, F. Metz, C. Beck, S. Lefevre, J.C. Chen, M. Pezeril, Q.A. Wang, Double power law degree distribution and informational entropy in urban road networks, International Modern Physics C, **22**(2011)33

[14] A. El Kaabouchi, F.X. Machu, R. Wang, Y.Y. Zhu, Q.A. Wang, 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, Advances in Complex Systems **24** (2021) 2150013