Generalized entropy optimized by a given arbitrary distribution

2003 J. Phys. A: Math. Gen. 36 8733

(http://iopscience.iop.org/0305-4470/36/33/301)

INSTITUTE OF PHYSICS PUBLISHING JOURNAL OF PHYSICS A: MATHEMATICAL AND GENERAL

J. Phys. A: Math. Gen. **36** (2003) 8733–8738 PII: S0305-4470(03)61363-6

Generalized entropy optimized by a given arbitrary distribution

Sumiyoshi Abe

Institute of Physics, University of Tsukuba, Ibaraki 305-8571, Japan

Received 25 March 2003, in final form 04 July 2003
Published 5 August 2003
Online at stacks.iop.org/JPhysA/36/8733

Abstract

An ultimate generalization of the maximum entropy principle is presented. An entropic measure, which is optimized by a given arbitrary distribution with the finite linear expectation value of a physical random quantity of interest, is constructed. It is concave irrespective of the properties of the distribution and satisfies the H-theorem for the master equation combined with the principle of microscopic reversibility. This offers a unified basis for a great variety of distributions observed in nature. As examples, the entropies associated with the stretched exponential distribution postulated by Anteneodo and Plastino (1999 *J. Phys. A: Math. Gen.* **32** 1089) and the κ-deformed exponential distribution by Kaniadaki (2002 *Phys. Rev.* E **66** 056125) and Naudts (2002 *Physica* A **316** 323) are derived. To include distributions with divergent moments (e.g., the Lévy stable distributions), it is necessary to modify the definition of the expectation value.

PACS number: 05.90.+m

Statistical distributions observed in nature have great diversity. In particular, a number of distributions, which are anomalous in view of ordinary statistical mechanics, are found in a variety of complex systems in their metaequilibrium states, including granular materials, glassy systems, self-gravitating systems and biological systems. What is remarkable there is that such metaequilibrium states often survive for periods much longer than typical time scales of underlying microscopic dynamics. To understand better the properties of such states of complex systems it is desirable to characterize these distributions within a unified framework of the statistical principles. The maximum entropy principle can be thought of as one [1]. Quite often in physical experiments, what is measured is the distribution of a physical random quantity (e.g., the energy) and not directly the entropy itself. Accordingly, it is of importance to find the corresponding entropic measure optimized by the observed distribution under appropriate constraints.

In this paper, we present the entropy-generating algorithm and construct an entropic measure that is optimized by a given arbitrary distribution with the finite *linear expectation value* of a physical random quantity. A question that arises then is how such a measure can possess the properties to be satisfied by the *physical entropy*. Here, we show that the proposed measure is manifestly concave irrespective of the behaviour of the distribution and fulfils the H-theorem for the master equation combined with the principle of microscopic reversibility. We also explicitly *derive* the entropies associated with the stretched exponential distribution and the κ-deformed exponential distribution, as examples.

Let us start our discussion by considering a continuous function f of $s \in D \subseteq \mathbf{R}$, whose range is [0, 1] and is assumed to be integrable over D

$$F = \int_D \mathrm{d}s\, f(s) < \infty. \tag{1}$$

f need not be a monotonic function, in general.

In terms of a probability distribution $\{p_i\}_{i=1,2,\ldots,W}$ with W being the number of accessible states, we define the following quantity:

$$A[p; s) = \sum_{i=1}^{W} (p_i - f(s))_+ \tag{2}$$

which can be thought of as a generalization of the quantities considered in [2, 3]. In this equation, the symbol $(x)_+$ stands for

$$(x)_+ = \max\{0, x\} \tag{3}$$

which satisfies

$$(\lambda x + (1 - \lambda)y)_+ \leqslant \lambda(x)_+ + (1 - \lambda)(y)_+ \qquad (\forall\lambda \in (0, 1)) \tag{4}$$

$$(x)_+ = x\theta(x) \tag{5}$$

where $\theta(x)$ is the Heaviside unit step function defined by $\theta(x) = 0$ for $x < 0$ and $\theta(x) = 1$ for $x > 0$. Though we are considering a discrete distribution, generalizations of the subsequent results to the continuous case are straightforward. In the case when the distribution is continuous, the range of f should be extended from [0, 1] to $[0, \infty)$, in general.

The quantity in equation (2) has some remarkable properties. Among others, what we note here are the following two:

$$0 < A[p; s) < 1 \tag{6}$$

$$A[\lambda p + (1 - \lambda)p'; s) \leqslant \lambda A[p; s) + (1 - \lambda)A[p'; s) \qquad (\forall\, \lambda \in [0, 1]). \tag{7}$$

Equation (7) directly follows from equation (4).

Next, we consider the integral

$$I[p] = \int_D \mathrm{d}s (1 - A[p; s)). \tag{8}$$

Clearly, this is a positive functional due to equation (6). Using the expression in equation (5), we find this integral to be rewritten as follows:

$$I[p] = \sum_{i=1}^{W} \int_D \mathrm{d}s (p_i - f(s))[1 - \theta(p_i - f(s))] + WF \tag{9}$$

where F is given in equation (1).

The entropy-generating algorithm presented here consists of identifying the generalized entropy, S, with the positive functional I:

$$S[p] = kI[p] + a \tag{10}$$

where k and a are constants, and in particular k is positive. Since the entropy should vanish for the completely ordered state, $p_i = p_i^{(0)} = \delta_{in}$ (with n a natural number between 1 and W), the constant a satisfies the condition: $kI[p^{(0)}] + a = 0$. Consequently, the generalized entropy is given by

$$S[p] = k(I[p] - I[p^{(0)}]). \tag{11}$$

By virtue of equation (7), S is a manifestly concave functional:

$$S[\lambda p + (1-\lambda)p'] \geqslant \lambda S[p] + (1-\lambda)S[p'] \qquad (\forall\, \lambda \in [0, 1]). \tag{12}$$

We wish to emphasize that this concavity property is established irrespective of the properties of f.

Let us first discuss a relatively simpler case when f is a monotonically decreasing function defined in $D = [0, \infty)$. In this case, I in equation (9) can be further rewritten as

$$I[p] = \sum_{i=1}^{W} \left[p_i f^{-1}(p_i) - \int_0^{f^{-1}(p_i)} \mathrm{d}s\, f(s) \right] + WF \tag{13}$$

where f^{-1} is the inverse function of f. Therefore, the generalized entropy reads

$$S[p] = k \left\{ \sum_{i=1}^{W} \left[p_i f^{-1}(p_i) - \int_0^{f^{-1}(p_i)} \mathrm{d}s\, f(s) \right] + c \right\} \tag{14}$$

where c is a constant given by

$$c = W \int_0^{f^{-1}(0)} \mathrm{d}s\, f(s) + \int_{f^{-1}(0)}^{f^{-1}(1)} \mathrm{d}s\, f(s) - f^{-1}(1). \tag{15}$$

Let us employ S in equation (14) as the entropy for the maximum entropy principle. Consider the functional

$$\Phi[p : \alpha, \beta] = S[p] - \alpha \left(\sum_{i=1}^{W} p_i - 1 \right) - \beta \left(\sum_{i=1}^{W} p_i Q_i - \langle Q \rangle \right) \tag{16}$$

where Q_i is the ith value of the basic random variable, Q, and α and β are the Lagrange multipliers associated with the constraints on the normalization condition and on the linear expectation value of Q denoted by $\langle Q \rangle$, respectively. (Here, clearly the ordinary expectation value is assumed to be well defined. However, there exist distributions with no finite moments. Celebrated examples are the Lévy stable distributions. To treat such distributions, it is necessary to modify the definition of the expectation value. See the later remarks.) Under the assumption of differentiability of f^{-1}, variation of Φ with respect to $\{p_i\}_{i=1,2,\ldots,W}$ leads to the following stationary distribution:

$$p_i = f(\alpha + \beta Q_i) \tag{17}$$

provided that the positive constant, k, has been eliminated by rescaling the Lagrange multipliers. α can be determined by the normalization condition: $\sum_{i=1}^{W} f(\alpha + \beta Q_i) = 1$. Therefore, an arbitrary monotonically decreasing distribution with the finite linear expectation value of Q could be derived from the maximum entropy principle.

We emphasize that the variational principle is not sufficient for identifying the generalized entropy associated with a given distribution, $p_i = f(\alpha + \beta Q_i)$ [4]. An obvious example revealing this point may be illustrated by taking the functional: $\Sigma[p] \sim \sum_{i=1}^{W} \int^{p_i} \mathrm{d}s\, f^{-1}(s)$, which also gives rise to the distribution of the same form as $p_i = f(\alpha + \beta Q_i)$. However, this Σ cannot be identified as a kind of entropy, since its concavity property is not consistent with arbitrary f, in general. This is in marked contrast to the present construction, in which S in equation (11) is always concave, as already stressed.

The above discussion can immediately be applied to some important cases. Here, we present two examples: (A) the stretched exponential distribution and (B) the κ-deformed exponential distribution.

(A) The generalized entropy associated with the stretched exponential distribution: in this case, f is taken to be

$$f(s) = \exp(-s^{\gamma}) \qquad (s \in [0, \infty), \gamma \in (0, 1)). \tag{18}$$

Then, the associated generalized entropy is found to be given by

$$S_{\gamma}[p] = \sum_{i=1}^{W} \Gamma(1 + 1/\gamma, -\ln p_i) - \frac{1}{\gamma}\Gamma(1/\gamma) \tag{19}$$

where $\Gamma(u, x)$ is the incomplete gamma function of the second kind defined by $\Gamma(u, x) = \int_x^{\infty} \mathrm{d}t\, t^{u-1}\, \mathrm{e}^{-t}$, and $\Gamma(u) = \Gamma(u, 0)$ is the gamma function. The generalized entropy in equation (19) is identical to the one recently discussed in [4], in which it seems to be postulated. We mention that, taking the limit $\gamma \to 1-0$ in the above discussion, f becomes the exponential function of s and equation (19) converges to the ordinary Boltzmann–Gibbs–Shannon entropy, $S_{\mathrm{BGS}}[p] = -\sum_{i=1}^{W} p_i \ln p_i$, as it should do [2].

(B) The generalized entropy associated with the κ-deformed exponential distribution [5, 6]: in this case, f is taken to be

$$f(s) = \exp_{\{\kappa\}}(-s) \equiv \left(\sqrt{1 + \kappa^2 s^2} - \kappa s\right)^{1/\kappa} \qquad (s \in [0, \infty), \kappa \in (0, 1)). \tag{20}$$

Though the range of κ can be extended to $(-1, 1)$ [5, 6], we here consider only $\kappa \in (0, 1)$ for the sake of simplicity. For this function, equation (11) is calculated to be

$$S_{\kappa}[p] = \sum_{i=1}^{W} \left[c_{-\kappa}\left(p_i^{1-\kappa} - p_i\right) + c_{\kappa}\left(p_i^{1+\kappa} - p_i\right)\right] \tag{21}$$

where

$$c_{\kappa} = -\frac{1}{2}\left(\frac{1}{\kappa} + \frac{1}{1 + \kappa}\right). \tag{22}$$

This is precisely the κ-deformed entropy given in [5]. Similar to the previous example (A), f in equation (20) and S_{κ} in equation (21) respectively converge to the exponential function e^{-s} and the Boltzmann–Gibbs–Shannon entropy in the limit $\kappa \to +0$.

Now, we briefly look at the H-theorem for the entropy (11) with equation (9). For this, consider the master equation

$$\frac{\mathrm{d}p_i}{\mathrm{d}t} = \sum_{j=1}^{W} (A_{ij} p_j - A_{ji} p_i) \tag{23}$$

where A_{ij} is the transition probability per unit time from the state j to the state i. Taking the time derivative of equation (11), we have

$$\begin{aligned} \frac{\mathrm{d}S[p]}{\mathrm{d}t} = {} & k \sum_{i=1}^{W} \int_D \mathrm{d}s\, \frac{\mathrm{d}p_i}{\mathrm{d}t}[1 - \theta(p_i - f(s))] \\ & - k \sum_{i=1}^{W} \int_D \mathrm{d}s (p_i - f(s))\delta(p_i - f(s))\, \frac{\mathrm{d}p_i}{\mathrm{d}t}. \end{aligned} \tag{24}$$

Noting that the second term on the right-hand side vanishes and using equation (23), we find

$$\frac{\mathrm{d}S[p]}{\mathrm{d}t} = k\sum_{i=1}^{W}\int_D \mathrm{d}s \sum_{j=1}^{W}(A_{ij}p_j - A_{ji}p_i)\,[1-\theta\,(p_i - f(s))]\,. \tag{25}$$

Assuming the principle of microscopic reversibility [4, 7, 8]

$$A_{ij} = A_{ji} \tag{26}$$

we obtain

$$\frac{\mathrm{d}S[p]}{\mathrm{d}t} = \frac{k}{2}\sum_{i,j=1}^{W}\int_D \mathrm{d}s A_{ij}(p_i - p_j)[\theta(p_i - f(s)) - \theta(p_j - f(s))] \geqslant 0. \tag{27}$$

Here, let us succinctly discuss the case when f in equation (2) is not monotonic. In such a case, the domain interval D of f should be divided into the subintervals, $\{D_a\}$, in which f is piecewise monotonic. D is now the disjoint union of D_a's. Accordingly, equations (8) and (9) are written as

$$I[p] = \sum_a \int_{D_a} \mathrm{d}s(1 - A[p;s]) \tag{28}$$

and

$$I[p] = \sum_a \sum_{i=1}^{W}\int_{D_a} \mathrm{d}s(p_i - f(s))[1-\theta(p_i - f(s))] + WF \tag{29}$$

respectively. Let $D_a^* = D_a^*(f^{-1}(p_i))$ be the subinterval of D_a, in which $p_i < f(s)$. Then, equation (29) can further be rewritten as follows:

$$\begin{aligned} I[p] &= \sum_a \sum_{i=1}^{W}\int_{D_a^*} \mathrm{d}s(p_i - f(s)) + WF \\ &= \sum_a \sum_{i=1}^{W}[p_i l(D_a^*(f^{-1}(p_i))) - F_a^*(f^{-1}(p_i))] + WF \end{aligned} \tag{30}$$

where $l(D_a^*)$ is the length of the interval D_a^* and $F_a^* = \int_{D_a^*}\mathrm{d}s\, f(s)$. Substitution of this quantity into equation (11) leads to the generalized entropy optimized by a non-monotonic distribution.

Finally, we mention that the Legendre transform structure highlighted by the relation

$$\frac{\partial S}{\partial \langle Q\rangle} = \beta \tag{31}$$

exists in the present generalized theory. This is due to the fact [9, 10] that equation (31) holds for an arbitrary form of the entropy and an arbitrary definition of the expectation value.

In conclusion, we have constructed the generalized entropy, which is optimized by any given statistical distribution. We have shown that it is concave irrespective of the properties of the distribution and satisfies the H-theorem for the master equation combined with the principle of microscopic reversibility. This can be regarded as an ultimate generalization of the maximum entropy principle. The present approach assumes that the distribution has a finite linear expectation value of a basic random quantity of interest. If the lowest moments are divergent, then the definition of the expectation value should be modified. For example, in Tsallis statistics [11–13], the optimal distribution is the so-called q-exponential distribution, which can asymptotically be a power-law distribution. In this case, the expectation value should be defined in terms of the escort distribution [12, 14, 15]. Regarding the necessity of modifying the definition of the expectation value, approaches from the different perspectives are needed [16].

Acknowledgments

The author would like to thank Drs Celia Anteneodo, Ramandeep S Johal and Tatsuaki Wada for useful discussions. This work was supported in part by the Grant-in-Aid for Scientific Research of Japan Society for the Promotion of Science.

References

[1] Rosenkrantz R D (ed) 1989 *E. T. Jaynes: Papers on Probability, Statistics and Statistical Physics* (Dordrecht: Kluwer)
[2] Lesche B 1982 *J. Stat. Phys.* **27** 419
[3] Abe S 2002 *Phys. Rev.* E **66** 046134
[4] Anteneodo C and Plastino A R 1999 *J. Phys. A: Math. Gen.* **32** 1089
[5] Kaniadakis G 2002 *Phys. Rev.* E **66** 056125
[6] Naudts J 2002 *Physica* A **316** 323
[7] Mariz M A 1992 *Phys. Lett.* A **165** 409
[8] Landsberg P T 1961 *Thermodynamics* (New York: Interscience)
[9] Plastino A and Plastino A R 1997 *Phys. Lett.* A **226** 257
[10] Yamano T 2000 *Eur. Phys. J.* B **18** 103
[11] Tsallis C 1988 *J. Stat. Phys.* **52** 479
[12] Tsallis C, Mendes R S and Plastino A R 1998 *Physica* A **261** 534
[13] Abe S and Okamoto Y (ed) 2001 *Nonextensive Statistical Mechanics and Its Applications* (Heidelberg: Springer)
[14] Beck C and Schlögl F 1993 *Thermodynamics of Chaotic Systems: An Introduction* (Cambridge: Cambridge University Press)
[15] Abe S 2000 *Phys. Lett* A **275** 250
[16] Abe S and Rajagopal A K 2000 *J. Phys. A: Math. Gen.* **33** 8733