



Axioms and uniqueness theorem for Tsallis entropy

Sumiyoshi Abe*

College of Science and Technology, Nihon University, 7-24-1 Narashinodai, Funabashi, Chiba 274-8501, Japan

Received 19 January 2000; received in revised form 2 May 2000; accepted 2 May 2000

Communicated by C.R. Doering

Abstract

The Shannon-Khinchin axioms for the ordinary information entropy are generalized in a natural way to the nonextensive systems based on the concept of nonextensive conditional entropy and a complete proof of the uniqueness theorem for the Tsallis entropy is presented. This improves the discussion of dos Santos [J. Math Phys. 38 (1997) 4104]. © 2000 Elsevier Science B.V. All rights reserved.

PACS: 05.20. – y

1. Introduction

The concept of entropy is the fundamental ingredient in information theory and statistical mechanics. Given a probability distribution $\{p_i\}_{i=1,2,\dots,W}$ satisfying $0 \leq p_i$ ($i = 1, 2, \dots, W$) and $\sum_{i=1}^W p_i = 1$, the Boltzmann–Shannon entropy, i.e., the ordinary information entropy, is defined by

$$S(p_1, p_2, \dots, p_W) = -k \sum_{i=1}^W p_i \ln p_i, \quad (1)$$

where k is a positive constant, which is henceforth set equal to unity for simplicity. This quantity is a positive and concave function of $\{p_i\}_{i=1,2,\dots,W}$. It also fulfills the H-theorem. Its maximum is realized

by the equiprobability distribution, i.e., $p_i = 1/W$ ($i = 1, 2, \dots, W$) and the value itself is $S = \ln W$, which is the celebrated Boltzmann formula.

To characterize what the entropy is, Shannon [1] and Khinchin [2] investigated its axiomatic foundation. The Shannon–Khinchin axioms are given as follows:

[I]. $S(p_1, p_2, \dots, p_W)$ is continuous with respect to all its arguments and takes its maximum for the equiprobability distribution $p_i = 1/W$ ($i = 1, 2, \dots, W$),

[II]. $S[A, B] = S[A] + S[B|A]$,

[III]. $S(p_1, p_2, \dots, p_W, 0) = S(p_1, p_2, \dots, p_W)$.

In the second axiom, $S[A, B]$ and $S[A]$ are the entropies of the composite system (A, B) with the joint probability distribution $p_{ij}(A, B)$ ($i = 1, 2,$

* Fax: +81-47-464-9342.

E-mail address: suabe@sf6.so-net.ne.jp (S. Abe).

$\dots, W; j = 1, 2, \dots, U$) and the subsystem A with the marginal probability distribution $p_i(A) = \sum_{j=1}^U p_{ij}(A, B)$, respectively. $S[B|A]$ stands for the conditional entropy associated with the conditional probability distribution

$$p_{ij}(B|A) = p_{ij}(A, B) / p_i(A).$$

Writing the entropy of $p_{ij}(B|A)$ as $S[B|A_i]$, the conditional entropy is given by

$$S[B|A] = \langle S[B|A_i] \rangle^{(A)} = \sum_{i=1}^W p_i(A) S[B|A_i]. \quad (2)$$

In the special case when A and B are statistically independent, $S[B|A] = S[B]$, leading to the additivity:

$$S[A, B] = S[A] + S[B]. \quad (3)$$

The uniqueness theorem [1,2] states that a quantity satisfying the axioms [I]–[III] is uniquely equal to S in Eq. (1).

We note that there is a correspondence relation between the Bayes multiplication law and the axiom [II]:

$$p_{ij}(A, B) = p_i(A) p_{ij}(B|A) \leftrightarrow S[A, B] = S[A] + S[B|A]. \quad (4)$$

Now, in the field of statistical mechanics, there is a growing interest in nonextensive generalization of Boltzmann–Gibbs theory. In particular, the one initiated by Tsallis [3,4]¹ is receiving much attention. This formalism is based on the following single-parameter generalization of the Boltzmann–Shannon entropy:

$$S_q(p_1, p_2, \dots, p_W) = \frac{1}{1-q} \left[\sum_{i=1}^W (p_i)^q - 1 \right] \quad (q > 0). \quad (5)$$

In the limit $q \rightarrow 1$, S_q converges to S in Eq. (1) (with $k \equiv 1$). This quantity is also positive and concave, and satisfies the H-theorem. In addition, as the

Boltzmann–Shannon entropy, it takes its maximum for the equiprobability distribution $p_i = 1/W$ ($i = 1, 2, \dots, W$): Its value is $S_q = (1-q)^{-1} (W^{1-q} - 1)$, which is a monotonically increasing function of W . However, the additivity is violated. For statistically independent systems, A and B , S_q yields

$$S_q[A, B] = S_q[A] + S_q[B] + (1-q) S_q[A] S_q[B], \quad (6)$$

which is referred to as the pseudoadditivity. Clearly, the additivity holds only in the limit $q \rightarrow 1$. Eq. (6) has been carefully discussed in Ref. [5].

In a recent paper [6], dos Santos has made an interesting discussion about uniqueness of the Tsallis entropy in Eq. (5). He has shown that that if a quantity is: [i] continuous with respect to the probability distribution $\{p_i\}_{i=1,2,\dots,W}$; [ii] a monotonically increasing function of W in the case of the equiprobability distribution $p_i = 1/W$ ($i = 1, 2, \dots, W$); [iii] satisfies the pseudoadditivity in Eq. (6) for the statistically independent systems, A and B ; and [iv] fulfills the relation

$$\begin{aligned} S_q(p_1, p_2, \dots, p_W) &= S_q(p_L, p_M) + (p_L)^q S_q \\ &\times (p_1/p_L, p_2/p_L, \dots, p_W/p_L) \\ &+ (p_M)^q S_q(p_1/p_M, p_2/p_M, \dots, p_W/p_M), \end{aligned}$$

where $p_L = \sum_{i=1}^{W_L} p_i$ and $p_M = \sum_{i=W_L+1}^W p_i$, then it is identical to the Tsallis entropy.

Here, comparing the set of the Shannon–Khinchin axioms [I]–[III] with that of [i]–[iv], we find that the complete parallelism is missing between the two. The main reason for this is due to the absence of the concept of ‘nonextensive conditional entropy’.

In this Letter, we present the axioms for the Tsallis entropy by introducing the nonextensive conditional entropy. Then we prove the uniqueness theorem for the Tsallis entropy. Our proof essentially follows a natural generalization of the line presented in Ref. [2]. This improves the discussion given in Ref. [6] and establishes the complete parallelism with the Shannon–Khinchin axioms.

¹ A comprehensive list of references can currently be obtained from <http://tsallis.cat.cbpf.br/biblio.htm>

2. Nonextensive conditional entropy

In nonextensive statistical mechanics, it is known [4] that the average of a physical quantity $Q = \{Q_i\}_{i=1, 2, \dots, W}$ is given in terms of the normalized q -expectation value:

$$\langle Q \rangle_q = \sum_{i=1}^W Q_i P_i \equiv \frac{\sum_{i=1}^W Q_i (p_i)^q}{\sum_{i=1}^W (p_i)^q}, \quad (7)$$

where $P_i \equiv (p_i)^q / \sum_{i=1}^W (p_i)^q$ is the escort distribution associated with p_i [7]. To be consistent with the nonextensive formalism, we use this concept to generalize the definition in Eq. (2). For this purpose, we calculate the Tsallis entropy of the conditional probability distribution

$$S_q[B|A_i] = \frac{1}{1-q} \left\{ \sum_{j=1}^U [p_{ij}(B|A)]^q - 1 \right\}. \quad (8)$$

From this quantity, we define the nonextensive conditional entropy as follows:

$$\begin{aligned} S_q[B|A] &= \langle S_q[B|A_i] \rangle_q^{(A)} \\ &= \frac{\sum_{i=1}^W [p_i(A)]^q S_q[B|A_i]}{\sum_{i=1}^W [p_i(A)]^q}. \end{aligned} \quad (9)$$

Using the definition of the Tsallis entropy, we find that Eq. (9) can be expressed as

$$S_q[B|A] = \frac{S_q[A, B] - S_q[A]}{1 + (1-q) S_q[A]}. \quad (10)$$

An important point here is that, with this definition, a natural nonextensive generalization of the correspondence relation in Eq. (4) is established in conformity with the pseudoadditivity in Eq. (6):

$$\begin{aligned} p_{ij}(A, B) &= p_i(A) p_{ij}(B|A) \leftrightarrow \\ &\times S_q[A, B] \\ &= S_q[A] + S_q[B|A] \\ &+ (1-q) S_q[A] S_q[B|A]. \end{aligned} \quad (11)$$

In fact, the pseudoadditivity in Eq. (6) is recovered in the special case when A and B are statistically independent each other.

3. Axioms and uniqueness theorem for Tsallis entropy

Now, the set of the axioms we present for the Tsallis entropy is the following:

[I]*. $S_q(p_1, p_2, \dots, p_W)$ is continuous with respect to all its arguments and takes its maximum for the equiprobability distribution $p_i = 1/W$ ($i = 1, 2, \dots, W$),

[II]*. $S_q[A, B] = S_q[A] + S_q[B|A] + (1-q) S_q[A] S_q[B|A]$,

[III]*. $S_q(p_1, p_2, \dots, p_W, 0) = S_q(p_1, p_2, \dots, p_W)$.

Theorem. A quantity satisfying [I]*–[III]* is uniquely equal to the Tsallis entropy.

Proof. First, let us consider the equiprobability distribution $p_i = 1/W$ ($i = 1, 2, \dots, W$) and put

$$S_q\left(\frac{1}{W}, \frac{1}{W}, \dots, \frac{1}{W}\right) = L_q(W). \quad (12)$$

From [III]*, it follows that

$$\begin{aligned} L_q(W) &= S_q\left(\frac{1}{W}, \frac{1}{W}, \dots, \frac{1}{W}, 0\right) \\ &\leq S_q\left(\frac{1}{W+1}, \frac{1}{W+1}, \dots, \frac{1}{W+1}, \frac{1}{W+1}\right) = L_q(W+1), \end{aligned} \quad (13)$$

which means that $L_q(W)$ is a nondecreasing function of W .

Consider m statistically independent systems, A_1, A_2, \dots, A_m , each of which contains $r (\geq 2)$ equally likely events. Then, we have

$$\begin{aligned} S_q[A_k] &= S_q\left(\frac{1}{r}, \frac{1}{r}, \dots, \frac{1}{r}\right) \\ &= L_q(r) \quad (1 \leq k \leq m). \end{aligned} \quad (14)$$

Using [III]* for these independent systems, we find

$$\begin{aligned} S_q[A_1, A_2, \dots, A_m] &= \sum_{k=1}^m \binom{m}{k} (1-q)^{k-1} [L_q(r)]^k \\ &= \frac{1}{1-q} \left\{ [1 + (1-q) L_q(r)]^m - 1 \right\}. \end{aligned} \quad (15)$$

Since $S_q[A_1, A_2, \dots, A_m] = L_q(r^m)$, we have

$$L_q(r^m) = \frac{1}{1-q} \left\{ [1 + (1-q) L_q(r)]^m - 1 \right\}. \quad (16)$$

Similarly, for any other positive integers n and s equal to or larger than 2, we have

$$L_q(s^n) = \frac{1}{1-q} \left\{ [1 + (1-q) L_q(s)]^n - 1 \right\}. \quad (17)$$

It is always possible to take m , r , n , and s which satisfy

$$r^m \leq s^n \leq r^{m+1}. \quad (18)$$

Since L_q is a nondecreasing function, we have the inequalities

$$L_q(r^m) \leq L_q(s^n) \leq L_q(r^{m+1}), \quad (19)$$

which lead to

$$\begin{aligned} &\frac{1}{1-q} \left\{ [1 + (1-q) L_q(r)]^m - 1 \right\} \\ &\leq \frac{1}{1-q} \left\{ [1 + (1-q) L_q(s)]^n - 1 \right\} \\ &\leq \frac{1}{1-q} \left\{ [1 + (1-q) L_q(r)]^{m+1} - 1 \right\}. \end{aligned} \quad (20)$$

Here, it is necessary to examine two cases, $0 < q < 1$ and $q > 1$, separately. After simple algebra, we find that in both cases the following inequalities hold:

$$\frac{m}{n} \leq \frac{\ln [1 + (1-q) L_q(s)]}{\ln [1 + (1-q) L_q(r)]} \leq \frac{m}{n} + \frac{1}{n}. \quad (21)$$

We note that, in deriving these inequalities, the following Ansatz has to be made:

$$0 < 1 + (1-q) L_q(r), \quad 1 + (1-q) L_q(s) < 1$$

for $q > 1$. (22)

Later, we shall see that this is in fact justified. Also, from Eq. (18), it is evident that

$$\frac{m}{n} \leq \frac{\ln s}{\ln r} \leq \frac{m}{n} + \frac{1}{n}. \quad (23)$$

Combining this with Eq. (21), we have

$$\left| \frac{\ln [1 + (1-q) L_q(s)]}{\ln [1 + (1-q) L_q(r)]} - \frac{\ln s}{\ln r} \right| \leq \frac{1}{n}. \quad (24)$$

Since n can be arbitrarily large, we obtain

$$\begin{aligned} &\frac{\ln [1 + (1-q) L_q(r)]}{\ln r} \\ &= \frac{\ln [1 + (1-q) L_q(s)]}{\ln s} := \lambda(q), \end{aligned} \quad (25)$$

where $\lambda(q)$ is a separation constant dependent on q . Therefore, we find

$$L_q(r) = \frac{1}{1-q} [r^{\lambda(q)} - 1]. \quad (26)$$

Clearly, $\lambda(1) = 0$. Next, let us consider any rational numbers

$$p_i = \frac{g_i}{g} \quad (i = 1, 2, \dots, W), \quad (27)$$

where g_i ($i = 1, 2, \dots, W$) are any positive integers and $g = \sum_{i=1}^W g_i$. The system A is assumed to be described by the probability distribution $\{p_i = g_i/g\}_{i=1, 2, \dots, W}$. We construct the system B dependent on A as follows. B contains g events, which are partitioned into W groups: B_1, B_2, \dots, B_W . B_j ($1 \leq j \leq W$) has g_j events. Once the i th event A_i of the system A was found, i.e., $A = A_i$, then, in the system B , g_i events of the group $B_{j=i}$ have the same conditional probability $1/g_i$ and all the events of the other groups $B_{j \neq i}$ have the vanish-

ing probability. S_q of B thus constructed is calculated to be

$$\begin{aligned} S_q[B|A_i] &= S_q\left(\frac{1}{g_i}, \frac{1}{g_i}, \dots, \frac{1}{g_i}\right) = L_q(g_i) \\ &= \frac{1}{1-q} \left[(g_i)^{\lambda(q)} - 1 \right]. \end{aligned} \quad (28)$$

Therefore, the nonextensive conditional entropy is given by

$$\begin{aligned} S_q[B|A] &= \langle S_q[B|A_i] \rangle_q^{(A)} \\ &= \frac{\sum_{i=1}^W (p_i)^q S_q[B|A_i]}{\sum_{i=1}^W (p_i)^q} \\ &= \frac{1}{1-q} \left[\frac{\sum_{i=1}^W (p_i)^q (g_i)^{\lambda(q)}}{\sum_{i=1}^W (p_i)^q} - 1 \right]. \end{aligned} \quad (29)$$

On the other hand, the composite system (A, B) consists of the events $A_i B_{j=i}$ ($1 \leq i \leq W$). For a given i , the number of possible events $A_i B_{j=i}$ ($1 \leq i \leq W$) is g_i , and therefore the total number of events in the composite system is $\sum_{i=1}^W g_i = g$. The probability of finding the event $A_i B_{j=i}$ is $p_i \times (1/g_i)$, which is the equiprobability $1/g$. Therefore, S_q of the composite system is

$$S_q[A, B] = L_q(g) = \frac{1}{1-q} \left[g^{\lambda(q)} - 1 \right]. \quad (30)$$

Substituting Eqs. (29) and (30) into $[\text{II}]^*$ and using Eq. (27), we have

$$S_q[A] = \frac{1}{1-q} \left[\frac{\sum_{i=1}^W (p_i)^q}{\sum_{i=1}^W (p_i)^{q+\lambda(q)}} - 1 \right]. \quad (31)$$

This holds for any rational p_i ($i = 1, 2, \dots, W$), but actually for any probability distribution $\{p_i\}_{i=1, 2, \dots, w}$ due to the assumption of continuity in $[\text{I}]^*$.

A remaining task is to determine $\lambda(q)$. For this purpose, it is sufficient to calculate the nonextensive conditional entropy using the form in Eq. (31) and impose $[\text{II}]^*$ on it. Consequently, we find

$$\lambda(q) = 1 - q. \quad (32)$$

At this stage, we also see that the Ansatz in Eq. (22) is in fact justified. Thus, we see that S_q satisfying $[\text{I}]^* - [\text{III}]^*$ is uniquely equal to the Tsallis entropy in Eq. (5). (Q.E.D.)

4. Concluding remarks

We have constructed the nonextensive conditional entropy in conformity with the Bayes multiplication law and the pseudoadditivity of the Tsallis entropy. We have generalized the Shannon–Khinchin axioms for the Boltzmann–Gibbs entropy to the nonextensive systems. Based on the proposed set of axioms, we have proved the uniqueness theorem for the Tsallis entropy.

Recently, the nonextensive (nonadditive) conditional entropy has been discussed in the quantum context in Ref. [8]. There, it has been shown to give rise to the strongest criterion for separability of the density matrix of a bipartite spin-1/2 system for validity of local realism. We also mention that characterization of the Tsallis entropy has been considered in Ref. [9] from the viewpoint of the concept of ‘composability’, which means that the entropy of the total system composed of statistically independent subsystems is expressed as a certain function of the entropies of such subsystems. (The additivity of the Boltzmann–Shannon entropy and the pseudoadditivity of the Tsallis entropy are the actual examples.) The authors of Ref. [9] has shown that if a quantity satisfies the composability (and some other supplementary conditions), then it is given by the Tsallis entropy with $q > 1$.

Acknowledgements

The author would like thank Professor A.K. Rajagopal for discussions about diverse topics of nonextensive statistical mechanics. This work was

supported in part by the GAKUJUTSU-SHO Program of College of Science and Technology, Nihon University.

References

- [1] C.E. Shannon, W. Weaver, *The Mathematical Theory of Communication*, University of Illinois Press, Urbana, 1963.
- [2] A.I. Khinchin, *Mathematical Foundations of Information Theory*, Dover, New York, 1957.
- [3] C. Tsallis, *J. Stat. Phys.* 52 (1988) 479.
- [4] C. Tsallis, R.S. Mendes, A.R. Plastino, *Physica A* 261 (1998) 534.
- [5] R.P. Di Sisto, S. Martínez, R.B. Orellana, A.R. Plastino, A. Plastino, *Physica A* 265 (1999) 590.
- [6] R.J.V. dos Santos, *J. Math. Phys.* 38 (1997) 4104.
- [7] C. Beck, F. Schlögl, *Thermodynamics of Chaotic Systems: An Introduction*, Cambridge University Press, Cambridge, 1993.
- [8] S. Abe, A.K. Rajagopal, *Nonadditive Conditional Entropy and its Significance for Local Realism*, e-print quant-ph/0001085.
- [9] M. Hotta, I. Joichi, *Phys. Lett. A* 262 (1999) 302.