Necessity of $q$-expectation value in nonextensive statistical mechanics

Sumiyoshi Abe$^{1,*}$ and G. B. Bagci$^2$

$^1$Institute of Physics, University of Tsukuba, Ibaraki 305-8571, Japan
$^2$Department of Physics, Middle East Technical University, Ankara 06531, Turkey

(Received 11 April 2004; published 28 January 2005)

In nonextensive statistical mechanics, two kinds of definitions have been considered for expectation value of a physical quantity: one is the ordinary definition and the other is the normalized $q$-expectation value employing the escort distribution. Since both of them lead to the maximum-Tsallis-entropy distributions of a similar type, it is of crucial importance to determine which the correct physical one is. A point is that the definition of expectation value is indivisibly connected to the form of generalized relative entropy. Studying the properties of the relative entropies associated with these two definitions, it is shown how the use of the escort distribution is essential. In particular, the Shore-Johnson theorem for consistent minimum cross-entropy (i.e., relative-entropy) principle is found to select the formalism with the normalized $q$-expectation value and to exclude the possibility of using the ordinary expectation value from nonextensive statistical mechanics.

DO: 10.1103/PhysRevE.71.016139 PACS number(s): 05.90.+m, 05.20.–y, 65.40.Gr

I. INTRODUCTION

Nonextensive statistical mechanics [1–4] pioneered by Tsallis [5] offers a consistent theoretical framework for the studies of complex systems in their nonequilibrium stationary states, systems with (multi)fractal and self-similar structures, long-range interacting systems, anomalous diffusion phenomena, and so on. The worked examples are dynamical systems at the edge of chaos [6–9], lattice Boltzmann models [10], magnetism of colossal magnetoresistance manganites [11], high-energy processes [12–16], cosmic rays [17], cellular aggregates [18], Lévy flights [19,20], semiclassical dynamics in optical lattices [21], astrophysics and self-gravitating systems [22,23], econophysical problems [24,25], kinetics of charged particles [26], Internet traffic [27], earthquakes [28,29], and complex networks [30–32].

In spite of these successes, still there remain some fundamental questions to be answered in the theory. One of them, which is of extreme importance, is concerned with the definition of expectation value. The frequently employed definition in nonextensive statistical mechanics is the normalized $q$-expectation value [33], which is given by

$$U^{(\text{norm})} = \langle H \rangle_q = \sum_i P_i \varepsilon_i,$$  

(1)

where $P_i$ is termed the escort distribution [34] associated with the basic distribution $p_i$, and $H$ denotes a physical random variable (e.g., the system energy) with its $i$th value $\varepsilon_i$. The index $q$ is taken to be positive. However, in the literature, there is an opinion that the ordinary expectation value

$$U^{(\text{ord})} = \langle H \rangle = \sum_i p_i \varepsilon_i,$$  

(3)

should be used also in nonextensive statistical mechanics. In Ref. [35], it has been shown that, for a class of power-law distributions, only the normalized $q$-expectation value is consistent with the method of steepest descents for (micro)canonical ensembles, but the situation remains unclear for the other class of distributions with compact supports.

It is discussed in Ref. [36] that there exists “duality” between the formalisms with the normalized $q$-expectation value and the ordinary expectation value. Notice that such a relationship can be realized only for the maximum entropy distributions.

Here, we address ourselves to the problem of choice of expectation value in nonextensive statistical mechanics. Our procedure is to examine the properties of the generalized relative entropies associated with the aforementioned two kinds of definitions. We shall see how the formalism with the normalized $q$-expectation value is superior to that with the ordinary expectation value.

The paper is organized as follows. In Sec. II, the ordinary and normalized $q$-expectation values are reexamined. An interesting geometric aspect of the maximum entropy principle is also pointed out, there. In Sec. III, two different kinds of the generalized relative entropies associated with these definitions of expectation value are considered and their properties are studied. In Sec. IV, an axiomatic approach to the issue is developed, and is shown to support the normalized $q$-expectation value, excluding the possibility of using the ordinary expectation value from nonextensive statistical mechanics. Section V is devoted to concluding remarks.

II. ORDINARY AND NORMALIZED $q$-EXPECTATION VALUES

Before discussing the problem of expectation value, first we wish to point out a geometric aspect of the maximum entropy principle, which will be used later. The idea was

*Corresponding author. Email address: suabe@s6.so-net.ne.jp
inspired by the work in Ref. [37]. Consider a functional $\Phi$ defined in the space $\Sigma$ of probability distributions. Two operations on $\Sigma$ are of interest: one is translation and the other is dilatation. The corresponding generators are given by

$$T_i = \frac{\delta}{\delta p_i},$$

$$D = \sum_i p_i \frac{\delta}{\delta p_i},$$

respectively, where $p_i \in \Sigma$. Clearly, they satisfy the following closed algebra: $[T_i, T_j] = 0$, $[T_i, D] = T_i$, $[D, D] = 0$. Invariance of the functional $\Phi$ under these operations implies

$$T_i \Phi = 0,$$

$$D \Phi = 0.$$  (6)

If $\Phi$ is an entropic functional, the solution to these equations yields the maximum entropy distribution. A point to be noticed is that the dilatation operation is constrained by the normalization condition

$$\sum_i p_i - 1 = 0.$$  (8)

This, in turn, determines the value of the associated Lagrange multiplier.

Now, let us apply this method to the Tsallis entropy indexed by $q$ [5],

$$S_q[p] = \frac{1}{1-q} \left[ \sum_i (p_i)^q - 1 \right].$$  (9)

Here and hereafter, the Boltzmann constant is set equal to unity for the sake of simplicity.

If the constraint is imposed on the ordinary expectation value, then the functional to be maximized is

$$\Phi^{(ord)}[p; \alpha, \beta] = S_q[p] - \alpha \left( \sum_i p_i - 1 \right) - \beta \left( \sum_i p_i \epsilon_i - U^{(ord)} \right),$$

where $\alpha$ and $\beta$ are the Lagrange multipliers. Equations (6)–(8) are found to give

$$\frac{q}{1-q} (p_i^{(ord)})^{q-1} - \alpha - \beta \epsilon_i = 0,$$

$$\alpha = \frac{q}{1-q} \left[ 1 + (1-q) S_q^{(ord)} \right] - \beta U^{(ord)},$$

where $S_q^{(ord)}$ and $U^{(ord)}$ are the values of $S_q$ and $U_q$ calculated in terms of the maximum entropy distribution $p_i^{(ord)}$, respectively. From these equations, it follows that

$$\tilde{p}_i^{(ord)} = \left[ 1 + (1-q) S_q^{(ord)} \right]^{1/(q-1)}$$

$$\times \left[ 1 - \frac{q-1}{q} \beta' \epsilon_i - \tilde{U}^{(ord)} \right]^{1/(q-1)},$$

where

$$\beta' = \frac{\beta}{\sum_i (\tilde{p}_i^{(ord)})^q},$$

and $[\alpha_+^+_a] = \max\{0, \alpha\}$. On the other hand, if the normalized $q$-expectation value is employed, the corresponding functional reads

$$\Phi^{(nor)}[p; \alpha, \beta] = S_q[p] - \alpha \left( \sum_i p_i - 1 \right) - \beta \left[ \sum_i (p_i)^q \epsilon_i - q \tilde{U}^{(nor)} \right].$$

Here, we are using the same notation for the Lagrange multipliers as in Eq. (10), but it will not cause any confusion. The operator $D$ acting on the third term on the right-hand side trivially vanishes since this term is manifestly invariant under the dilatation. Equations (6)–(8) give rise to

$$\frac{q}{1-q} (p_i^{(nor)})^{q-1} - \alpha - q \beta' (\epsilon_i - \tilde{U}^{(nor)})(p_i^{(nor)})^{q-1} = 0,$$

$$\alpha = \frac{q}{1-q} \left[ 1 + (1-q) \tilde{S}_q^{(nor)} \right],$$

where

$$\beta' = \frac{\beta}{\sum_i (\tilde{p}_i^{(nor)})^q}.$$  (18)

$\tilde{S}_q^{(nor)}$ and $\tilde{U}^{(nor)}$ are the values of $S_q$ and $U_q$ calculated in terms of the maximum entropy distribution $\tilde{p}_i^{(nor)}$. Equations (16) and (17) lead to

$$\tilde{p}_i^{(nor)} = \left( \frac{1}{Z_q} \right)^{1/(1-q)} \left[ 1 + (1-q) \tilde{S}_q^{(nor)} \right]^{1/(1-q)},$$

$$\tilde{Z}_q^{(nor)} = \left[ 1 + (1-q) \tilde{S}_q^{(nor)} \right]^{1/(1-q)}.$$  (20)

Equations (13) and (19) are quite similar to each other, but the signs of the exponents are opposite.

An important point is that with both of the definitions of expectation value the following thermodynamic relations hold:

$$\frac{\delta S_q^{(ord)}}{\delta U^{(ord)}} = \beta,$$

$$\frac{\delta S_q^{(nor)}}{\delta U^{(nor)}} = \beta.$$  (22)

which may indicate that the thermodynamic Legendre-transform structure is established in both cases. However, it
is still an open problem in nonextensive statistical mechanics if $\beta$ is the physical inverse temperature [38].

It is clear that, in the limit $q \to 1$, the Tsallis entropy in Eq. (9) tends to the Boltzmann-Gibbs-Shannon entropy $S[p] = -\sum p_i \ln p_i$, and accordingly both of the distributions in Eqs. (13) and (19) converge to the familiar Boltzmann-Gibbs distribution $\bar{p}_i \sim \exp(-\beta e_i)$.

III. GENERALIZED RELATIVE ENTROPIES

Relative entropy plays a fundamental role for comparing two distributions. There exist two different kinds of the generalized relative entropies in nonextensive statistical mechanics. One is given by [39]

$$I_q[p || r] = \frac{1}{q-1} \sum_i p_i [(p_i)^q - (r_i)^q] - \sum_i (p_i - r_i)(r_i)^{q-1},$$

(23)

which is of the Bregman type [40], and the other is [41-43]

$$K_q[p || r] = \frac{1}{1-q} \left[ 1 - \sum_i (p_i)^q (r_i)^{1-q} \right],$$

(24)

which is of the Csiszár type [44], where $r_i$ is a reference distribution (i.e., prior). In what follows, we shall look at the properties of these two quantities in detail. In particular, in Sec. III B, we shall see that $I_q[p || r]$ and $K_q[p || r]$ are the relative entropies associated with the ordinary expectation value and the normalized $q$-expectation value, respectively.

A. Correspondence relation and nonnegativity

First of all, we notice that, in the limit $q \to 1$, both $I_q[p || r]$ and $K_q[p || r]$ tend to the Kullback-Leibler relative entropy

$$H[p || r] = \sum_i p_i \ln \frac{p_i}{r_i}.$$  

(25)

The following expression is often used for this quantity:

$$H[p || r] = \frac{d}{dx} \left[ \sum_i (p_i)^q (r_i)^{1-q} \right]_{x=1}.$$  

(26)

$K_q[p || r]$ in Eq. (24) is obtained by replacing the differential operator by the Jackson $q$-differential operator [45],

$$K_q[p || r] = D_q \left[ \sum_i (p_i)^q (r_i)^{1-q} \right]_{x=1},$$

(27)

where $D_q f(x) = [f(qx) - f(x)]/[x(q-1)]$, which satisfies the $q$-deformed Leibniz rule $D_q [f(x)g(x)] = D_q f(x)g(x) + f(x)D_q [g(x)] + x(q-1)D_q f(x)D_q g(x)$ and converges to the ordinary differential in the limit $q \to 1$.

No such simple correspondence is known to exist between $H[p || r]$ and $I_q[p || r]$. One could, however, still dare to write

$$H[p || r] = \frac{dG(x)}{dx} \bigg|_{x=1},$$

(28)

$$G(x) = \frac{1}{x} \sum_i [(p_i)^q - (r_i)^q] - \sum_i (p_i - r_i)(r_i)^{q-1},$$

(29)

and accordingly

$$I_q[p || r] = qD_q G(x)|_{x=1}.$$  

(30)

Like $H[p || r]$, $I_q[p || r]$ and $K_q[p || r]$ are nonnegative and vanish if and only if $p_i = r_i$ ($\forall i$). This can be seen as follows. Regarding $I_q[p || r]$, it is convenient to employ the integral representation [39]

$$I_q[p || r] = \frac{1}{q-1} \sum_i p_i [(x)^q - (r_i)^q],$$

(31)

from which nonnegativity follows immediately. On the other hand, $K_q[p || r]$ can be rewritten as

$$K_q[p || r] = \frac{1}{1-q} \sum_i p_i [(x)^q - (r_i)^q].$$

(32)

Noticing $(1 - x^{-q})/(1 - q) \geq 1 - x$ for $x > 0$ and $q > 0$ with the equality for $x = 1$, $K_q[p || r]$ is also seen to be nonnegative.

B. Free energy difference

Let us discuss the physical meanings of $I_q[p || r]$ and $K_q[p || r]$, in particular, their relevance to the definitions of expectation value. For this purpose, we take the maximum entropy distributions as the reference distributions. The exponents of the maximum entropy distributions in Eqs. (13) and (19) together with the dependencies of $I_q[p || r]$ and $K_q[p || r]$ on $r_i$ should be noticed.

Putting $r_i = \bar{p}_i^{(\text{ord})}$ in Eq. (23), we have

$$I_q[p || \bar{p}^{(\text{ord})}] = \frac{1}{q-1} \sum_i \bar{p}_i^{(\text{ord})} [(x)^q - (r_i)^q].$$

(33)

where

$$F_q^{(\text{ord})} = U^{(\text{ord})} - \frac{1}{\beta} S_q, \quad F_q^{(\text{ord})} = \tilde{U}^{(\text{ord})} - \frac{1}{\beta} \tilde{S}_q.$$  

(34)

On the other hand, substituting $r_i = \tilde{p}_i^{(\text{non})}$ into Eq. (24), we obtain

$$K_q[p || \tilde{p}^{(\text{non})}] = \frac{1}{q-1} \sum_i \tilde{p}_i^{(\text{non})} [(x)^q - (r_i)^q],$$

(35)

where

$$\tilde{F}_q^{(\text{non})} = U^{(\text{non})} - \frac{1}{\hat{\beta}} S_q, \quad \tilde{F}_q^{(\text{non})} = \tilde{U}^{(\text{non})} - \frac{1}{\hat{\beta}} \tilde{S}_q.$$  

(36)

Equations (33) and (35) show that $I_q[p || r]$ and $K_q[p || r]$ are essentially the free energy differences and therefore are recognized as the generalized relative entropies associated with the ordinary expectation value and the normalized $q$-expectation value, respectively. We also mention that the
quantum mechanical counterpart of $K_d[p\|r]$ has recently been employed to prove the second law of thermodynamics [46].

**C. Convexity**

Convexity is one of the most important properties to be fulfilled by relative entropy. Taking the second-order derivatives of $I_d[p\|r]$ with respect to the arguments, one finds that it is convex in $p_i$, but not in $r_i$. On the other hand, like the Kullback-Leibler relative entropy, $K_d[p\|r]$ is seen to be jointly convex (see Ref. [47] in the quantum mechanical case):

$$K_d\left[\sum_a \lambda_a p(a) \| \sum_a \lambda_a r(a)\right] \leq \sum_a \lambda_a K_d[p(a) \| r(a)],$$  \hspace{1cm} (38)

where $\lambda_a > 0$ and $\sum_a \lambda_a = 1$. This property is stronger than individual convexity in $p_i$ and $r_i$.

**D. Composability**

Finally, we notice that, like the Kullback-Leibler relative entropy, $K_d[p\|r]$ is “composable” [48], but $I_d[p\|r]$ is not. In fact, for factorized joint distributions of a composite system $(A,B)$, $p_i(A,B)=p_i(1)(A)p_i(2)(B)$ and $r_i(A,B)=r_i(1)(A)r_i(2)(B)$, $K_d[p_i(p_A)p_i(r_A)]$ yields

$$K_d[p_i(p_A)p_i(r_A)] = K_d[p_i(p_A)p_i(r_A)] + K_d[p_i(r_A)p_i(p_B)]$$

where no such closed relation exists for $I_d[p_i(p_A)p_i(r_A)]$. Equation (39) has its origin in the $q$-deformed Leibniz rule satisfied by the Jackson $q$-differential operator.

**IV. SHORE-JOHNSON THEOREM SUPPORTS NORMALIZED $q$-EXPECTATION VALUE AND RULES OUT ORDINARY EXPECTATION VALUE**

In the preceding section, we have seen how $K_d[p\|r]$ associated with the normalized $q$-expectation value has the properties, which are more favorable than those of $I_d[p\|r]$ corresponding to the ordinary expectation value. In this section, we discuss that the Shore-Johnson theorem for consistent minimum cross-entropy (relative-entropy) principle supports $K_d[p\|r]$ and rules out $I_d[p\|r]$, leading to necessity of using the normalized $q$-expectation value in nonextensive statistical mechanics.

About a quarter a century ago, Shore and Johnson [49] have proposed the axioms for minimum cross-entropy (i.e., relative-entropy) principle. These authors have made an attempt to answer to the question why the correct rule of inference is to minimize relative entropy, in conformity with a vindication of Jaynes’ claim that every other rule will lead to contradiction [50]. The axioms are composed of the following five statements (presented in a nonabstract manner):

(i) **Axiom I** (uniqueness): If the same problem is solved twice, then the same answer is expected to result both times.

(ii) **Axiom II** (invariance): The same answer is expected when the same problem is solved in two different coordinate systems, in which the posteriors in the two systems should be related by the coordinate transformation.

(iii) **Axiom III** (system independence): It should not matter whether one accounts for independent information about independent systems separately in terms of their marginal distributions or in terms of the joint distribution.

(iv) **Axiom IV** (subset independence): It should not matter whether one treats independent subsets of the states of the systems in terms of their separate conditional distributions or in terms of the joint distribution.

(v) **Axiom V** (expansibility): In the absence of new information, the prior (i.e., the reference distribution) should not be changed.

These axioms are natural in the sense that all of them are fulfilled by the ordinary Kullback-Leibler relative entropy in Eq. (25), which gives rise to the free energy difference in Boltzmann-Gibbs statistical mechanics.

For the Tsallis entropy in Eq. (9), the axioms and uniqueness theorem are known in the literature [51]. In contrast to this fact, the above set of axioms is quite general and not very restrictive, and therefore does not uniquely determine the definition of the relative entropy. However, the Shore-Johnson theorem [49] states that the relative entropy $J_d[p\|r]$ with the prior $r_i$ and the posterior $p_i$ satisfying the axioms’ I–V has the following form:

$$J[p\|r] = \sum_i p_i h(p_i/r_i),$$  \hspace{1cm} (40)

where $h(x)$ is some function.

At this juncture, it is crucial to recognize that the function $h(x)$ surely exists for $K_d[p\|r]$:

$$h(x) = \frac{1}{1-q} (1-x^{q-1}).$$  \hspace{1cm} (41)

On the other hand, $I_d[p\|r]$ cannot be recast to the form in Eq. (40). This is because, as can be seen, $I_d[p\|r]$ may violate Axiom III.

Therefore we conclude that the Shore-Johnson theorem supports the normalized $q$-expectation value and excludes the possibility of using the ordinary expectation value from nonextensive statistical mechanics.

**V. CONCLUDING REMARKS**

We have discussed two kinds of definitions of expectation value in nonextensive statistical mechanics, that is, the ordinary expectation value and the normalized $q$-expectation value. To determine which is the correct definition, we have studied the corresponding generalized relative entropies. It was found that the generalized relative entropy associated with the normalized $q$-expectation value has the properties,
which are superior to those associated with the ordinary expectation value. More decisively, the Shore-Johnson theorem is shown to select the formalism with the normalized $q$-expectation value and to exclude the possibility of using the ordinary expectation value from nonextensive statistical mechanics.


S.A. thanks Andrei Bashkirov and Jan Naudts for discussions. He also acknowledges the Grant-in-Aid for Scientific Research of Japan Society for the Promotion of Science for financial support.