Tsallis maximum entropy principle and the law of large numbers

Brian R. La Cour and William C. Schieve

Ilya Prigogine Center for Studies in Statistical Mechanics and Complex Systems, Department of Physics,
The University of Texas at Austin, Austin, Texas 78712

Tsallis has suggested a nonextensive generalization of the Boltzmann-Gibbs entropy, the maximization of which gives a generalized canonical distribution under special constraints. In this Brief Report, we show that the generalized canonical distribution so obtained may differ from that predicted by the law of large numbers when empirical samples are held to the same constraint. This conclusion is based on a result regarding the large deviation property of conditional measures and is confirmed by numerical evidence.

PACS number(s): 02.50.Cw, 05.20.Gg, 05.30.Ch

I. INTRODUCTION

From considerations of multifractals, Tsallis [1] was led to conjecture a generalization of the Boltzmann-Gibbs entropy given by

$$S_q(p) = \frac{1}{q-1}\left[1 - \sum_{i=1}^{m} p_i^q\right], \qquad (1)$$

where $p = (p_1, \ldots, p_m)$ is a probability distribution for a discrete random variable with values $\epsilon_1, \ldots, \epsilon_m$ and q is any real number different from one. S_1 is defined to be the usual Boltzmann-Gibbs entropy, in agreement with the limit $q \rightarrow 1$. (Boltzmann's constant is set to one.) Non-Gibbsian distributions are obtained by extremizing the Tsallis entropy under special constraints, described below, while using q as an adjustable parameter. The parameter q typically has no direct physical interpretation, but when it is used as an adjustable parameter the resulting distributions can give surprisingly good agreement with experimental data in a wide variety of fields [2]. In a few cases, q is uniquely determined by the constraints of the problem and may thereby bear some physical interpretation [3,4].

Although the Tsallis entropy preserves all of the familiar thermodynamic formalism, Curado [5] has noted that this is true of a much broader class of entropies. Given the myriad of possible entropy functions, one is led to ask why the Tsallis entropy is special, and a natural place to look for answers is in the theory of large deviations [6], which gives a probabilistic justification for the maximum entropy principle in terms of a unique entropy function. In this Brief Report, we compare the probabilities obtained by Tsallis's maximum entropy principle with the asymptotic frequencies predicted by large deviation theory (i.e., the law of large numbers) under similar constraints. We find that the two do not in general agree.

II. TSALLIS MAXIMUM ENTROPY PRINCIPLE

If no constraints are imposed upon p (other than that it be nonnegative and normalized), S_q is readily seen to be extremized by $p_i = 1/m \equiv \mu_i$. (The case $q = 0$ is special, as S_0 is a constant function.) This conclusion, independent of q, agrees with the usual Boltzmann-Gibbs result and corre-

sponds to a microcanonical ensemble. If we view μ as a sampling distribution, then the empirical distribution of frequencies obtained from a random sample $x_1, \ldots, x_n$ converges to μ almost surely as n grows large. This well-known result, originally due to Boltzmann [7], may be viewed as an example of the (strong) law of large numbers. Since S_q has a global extremum at μ, the distribution predicted by extremizing S_q agrees with the actual asymptotic empirical distribution.

Placing additional constraints when extremizing S_q may result in a distribution dependent upon q, i.e., one at variance with that predicted from the Boltzmann-Gibbs case $q = 1$. As a generalization of the internal energy constraint, Tsallis [8] has suggested the following constraint be used when extremizing S_q:

$$\sum_{i=1}^{m} (\epsilon_i - u)p_i^q = 0, \qquad (2)$$

where u is a given fixed constant. For $q = 1$ this of course reduces to the usual expectation value constraint. By extremizing Eq. (1) subject to Eq. (2), one obtains a solution in general different from the Boltzmann distribution. This solution is given explicitly by

$$p_i \propto [1 - (1-q)\alpha(\epsilon_i - u)]^{1/(1-q)}, \qquad (3)$$

where α is chosen such that Eq. (2) is satisfied. It has been noted that this explicit form of the distribution appears to be more numerically robust than the more common implicit form, for which $\alpha = \beta/\sum_{j=1}^{m} p_j^q$ [9].

For $q = 1$ the constraint on the expectation may be interpretation as a constraint on the sample mean, the two being equivalent for large samples. Thus, if we consider random samples $x_1, \ldots, x_n$ from μ that satisfy

$$\frac{1}{n}\sum_{k=1}^{n} x_k = u, \qquad (4)$$

then the empirical distributions of such samples will approach the Boltzmann distribution $p_i \propto e^{-\alpha\epsilon_i}$ as n grows large.

The question arises of whether a similar interpretation may be made of the constraint in Eq. (2) for $q \neq 1$ and, more

importantly, whether the resulting empirical distribution converges to that given by Eq. (3). As our observable is discrete, let $f_{n,i}(x_1, \ldots, x_n)$ denote the observed frequency of ϵ_i in the sample $x_1, \ldots, x_n$. (There is no obvious interpretation for continuous values.) We may interpret Eq. (2) to mean

$$\sum_{i=1}^{m} (\epsilon_i - u) f_{n,i}(x_1, \ldots, x_n)^q = 0. \qquad (5)$$

We will show that random samples drawn from μ that satisfy Eq. (5) do not in general give rise to empirical distributions that converge to the Tsallis prediction of Eq. (3).

III. CONDITIONAL CONVERGENCE OF THE EMPIRICAL DISTRIBUTION

The general problem we are considering is the convergence in probability of the empirical frequencies $f_n = (f_{n,1}, \ldots, f_{n,m})$, where f_n is a random vector with domain $\{\epsilon_1, \ldots, \epsilon_m\}^n$ taking values in the convex set $\mathcal{P} = \{p \in \mathbb{R}^m : p_i \geq 0, \Sigma_{i=1}^m p_i = 1\}$. Unconstrained, an infinite random sample $x_1, x_2, \ldots$, from μ gives rise to a sequence of empirical frequencies that converge in probability to μ. Sanov's theorem [7] gives the large deviation rate function for this convergence to be just the negative of the Boltzmann-Gibbs entropy:

$$I_\mu(p) = -S_1(p) - \ln m. \qquad (6)$$

Loosely speaking, Sanov's theorem states that for $A \subseteq \mathcal{P}$, $\mu^n[f_n \in A] \sim \exp[-n \inf_{p \in A} I_\mu(p)]$ for large n (cf. the Boltzmann-Einstein formula $W = e^S$). The asymptotic measure μ is the unique minimum of the rate function I_μ, which is continuous and strictly convex.

When we impose additional constraints on f_n, the asymptotic value changes from μ to a new distribution that minimizes I_μ under the added restrictions [7,10]. If we condition on the sample mean, for example,

$$\sum_{i=1}^{m} \epsilon_i f_{n,i}(x_1, \ldots, x_n) = u, \qquad (7)$$

the resulting asymptotic distribution is no longer μ but the canonical distribution $P_i \propto e^{-\beta \epsilon_i}$, where β satisfies

$$\sum_{i=1}^{m} \epsilon_i P_i = u. \qquad (8)$$

It is in this sense that finding the asymptotic empirical distribution under Eq. (7) is equivalent to maximizing S_1 under Eq. (8).

More generally, imposing condition (5) results in an asymptotic distribution that minimizes I_μ (maximizes S_1) subject to Eq. (2). This distribution is given implicitly by

$$P_i \propto \exp[-\beta(\epsilon_i - u) P_i^{q-1}], \qquad (9)$$

where β is such that Eq. (2) is satisfied with p replaced by P. Comparison with Eq. (3) shows that both p and P will agree when $q \rightarrow 1$.

IV. COMPARISON OF THE TWO DISTRIBUTIONS

For $q = 0$, Eq. (3) gives $p_i = [1 - \alpha(\epsilon_i - u)]/m$, with α unrestricted, while Eq. (9) implies $P_i = 1/m$. Clearly, both agree if α is arbitrarily chosen to be zero. However, as we have noted, S_0 is a constant function, so the entropy extremization procedure may be expected to break down in this case.

Taking u to be the equilibrium value $u_* = \Sigma_{i=1}^m \epsilon_i/m$ also results in general agreement between p and P for all $q \neq 0$. Indeed, by choosing $\alpha = \beta = 0$ we see that $p_i = 1/m$ is the unique solution for both Eq. (3) and Eq. (9). This agreement simply reflects the fact that both S_1 and S_q have the same global extremum.

When $m = 2$ the two constraints are sufficient to uniquely determine the distribution, and for this reason general agreement is also expected. In particular, we find

$$p = P \propto ((\epsilon_2 - u)^{1/q}, (u - \epsilon_1)^{1/q}), \qquad (10)$$

assuming $\epsilon_1 < \epsilon_2$ and $q \neq 0$. It is readily verified that Eq. (2) is satisfied. By solving for α and β, Eqs. (3) and (9), respectively, may be satisfied as well.

Disagreement between p and P is therefore expected when $m \geq 3$. To show this explicitly, we may compute p from Eq. (3) for an arbitary u and then search for a value of β such that Eq. (9) is satisfied when p is substituted for P. The claim is that a single β cannot always be found that satisfies this equation for all values of i when $m \geq 3$.

The case $q = 1/2$ is particularly amenable to analytic study [11] and appears in an early application of the Tsallis entropy to turbulence in a two-dimensional electron plasma [12]. For this case, Eq. (3) may be solved explicitly in terms of u to obtain

$$p_i \propto \left[\frac{1}{m} \sum_{j=1}^{m} (\epsilon_j - u)^2 - (\epsilon_i - u)(u_* - u) \right]^2. \qquad (11)$$

Using a given value of u and the corresponding p given above, we then consider zeros of the functions d_i, where

$$d_i(\beta) = \frac{\exp[-\beta(\epsilon_i - u) p_i^{-1/2}]}{\sum_{j=1}^{m} \exp[-\beta(\epsilon_j - u) p_j^{-1/2}]} - p_i, \qquad (12)$$

for $i = 1, \ldots, m$. A plot of these functions is shown in Fig. 1 for selected parameter values. The failure of all three graphs to have a zero at the same value of β indicates that p and P are in this case distinct.

From this example, one can derive a general necessary condition for agreement with P. Suppose that for given q, ϵ, and u there exists a simultaneous solution to both Eqs. (3) and (9). [More generally, p may be any probability distribution satisfying Eq. (2).] Substituting the former into the latter we find

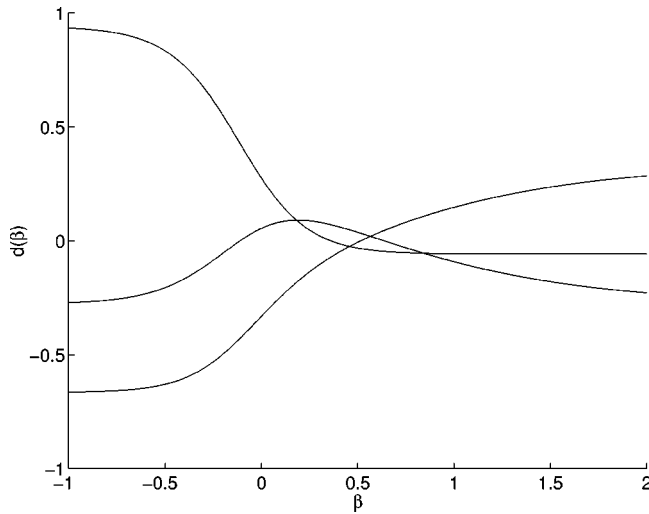$$p_i = \exp[-\beta(\epsilon_i - u) p_i^{q-1}]/Z(\beta), \qquad (13)$$

where

FIG. 1. Plot of $d_i(\beta) = \rho_i(\beta) - p_i$ for $\epsilon = (0,1,2)$, $q = 1/2$, and $u = 7/11$, for which $p = (289,121,25)/435$. The positive roots are found numerically to be $0.514\,509$, $0.637\,715$, $0.360\,903$ for $i = 1,2,3$, respectively.

$$Z(\beta) = \sum_{j=1}^{m} \exp[-\beta(\epsilon_j - u)p_i^{q-1}]. \tag{14}$$

The value of each p_i is fixed in terms of the given parameters, so a single value of β must simultaneously satisfy Eq. (13) for $i = 1, \ldots, m$. If any $p_i = 0$, then Eq. (13) cannot possibly be satisfied, so suppose all p_i are nonzero. For any given $j \neq i$,

$$\beta = -[\ln p_j + \ln Z(\beta)]/[(\epsilon_j - u)p_j^{q-1}]. \tag{15}$$

Substituting this expression back into Eq. (13) gives

$$\ln Z(\beta) = \frac{(\epsilon_i - u)p_i^{q-1} \ln p_j - (\epsilon_j - u)p_j^{q-1} \ln p_i}{(\epsilon_j - u)p_j^{q-1} - (\epsilon_i - u)p_i^{q-1}}. \tag{16}$$

The RHS of Eq. (16) is invariant under the interchange of i and j, so it has at most $m(m-1)/2$ distinct values. The

LHS, of course, is the same for all choices of i and j. Now, the RHS will be independent of the choice of i and j if either (1) $q = 1$, (2) $m = 2$, or (3) $p_i = p_j$ for all i and j, the latter being equivalent to $u = u_*$, which is equivalent to $\alpha = 0$. Assuming none of these three conditions hold, the RHS must be the same for all choices of i and j if indeed $p = P$. This gives a necessary condition for agreement.

V. DISCUSSION

We have compared the probability distribution over m states predicted from Tsallis's maximum entropy principle, which constrains the normalized q expectation to a value u, to the asymptotic frequencies when the empirical q expectation is similarly constrained. The two will always agree if either (1) $q = 1$, (2) $m = 2$, or (3) $u = u_*$. A specific example for which $q = 1/2$ and $m = 3$ was used to demonstrate numerically that the two distributions may be different. For the case in which none of these three conditions hold, we derived a necessary condition to be satisfied by any candidate distribution in order that it be identical to true asymptotic distribution.

From the point of view of large deviation theory, the maximum entropy principle specifies the overwhelmingly most probable distribution to be realized by a large-sample empirical distribution under given constraints. The uniqueness of the rate function in large deviation theory implies that the Boltzmann-Gibbs entropy plays a special role in determining this most likely distribution. For this reason, entropy functions such as that proposed by Tsallis may give results that are at variance with actual sample frequencies except, as observed, in some special cases.

[1] C. Tsallis, J. Stat. Phys. **52**, 479 (1988).
[2] C. Tsallis, Braz. J. Phys. **29**, 1 (1999).
[3] C. Beck, Physica A **277**, 115 (2000).
[4] M.L. Lyra and C. Tsallis, Phys. Rev. Lett. **80**, 53 (1998).
[5] E.M.F. Curado, Braz. J. Phys. **29**, 36 (1999).
[6] R.S. Ellis, *Entropy, Large Deviations, and Statistical Mechanics* (Springer-Verlag, New York, 1985).
[7] R.S. Ellis, Physica D **133**, 106 (1999).
[8] C. Tsallis, R. Mendes, and A.R. Plastino, Physica A **261**, 534 (1998).
[9] A.R. Lima and T.J.P. Penna, Phys. Lett. A **256**, 221 (1999).
[10] B.R. La Cour and W.C. Schieve, J. Stat. Phys. **99**, 1225 (2000).
[11] L. Rebolla-Neira, A. Plastino, and J. Fernandez-Rubio, Physica A **258**, 458 (1998).
[12] B.M.R. Boghosian, Phys. Rev. E **53**, 4754 (1996).