

# Projections Maximizing Tsallis Entropy Projections Maximizing Tsallis Entropy

Peter Harremoës    Peter Harremoës

**Abstract.** We consider a small closed physical systems. It has fixed energy and if the energy is a power of its generale coordinates then the distribution of any single coordinate will follow a distribution that maximizes Tsallis entropy. For many aspects of entropy can be discussed for such small systems with out going to the thermodynamical limit. By letting the number of degrees of freedom tend to infinity the ordinary entropy formalism is recovered.

**Keywords:** <Missing keywords>

**PACS:** <Missing classification>

Centrum voor Wiskunde en Informatica, Kruislaan 413, 1090-GB Amsterdam, The Netherlands

## INTRODUCTION

The second law of thermodynamics is by many considered as one of the most mysterious law of physics. Actually to such an extend that people do not agree about what the problem really is. Even people that do not agree that entropy is mysterious would admit that nobody has really succeeded in combining the different definitions of entropy in thermodynamics, statistical mechanics and information theory into a simple consistent theory. In order to make a consistent theory one often appeal to the thermodynamic limit where one let the number of particles tend to infinity. Many physicists therefor claim that concepts like entropy and temperature do not exist for systems of finite size which unfortunately includes all real world systems. For small systems with only few degrees of freedom one may therefore think that concepts like entropy are not applicable at all. Here we shall see that entropy plays a role even for some systems of finite size and that many discussions related to maximum entropy are equally relevant for such small systems.

## RÉNYI AND TSALLIS ENTROPY

For a discrete distribution with point probabilities  $p_1, p_2, \dots$  the Rényi entropy is defined as

$$H_q = \frac{\log \sum_i p_i^q}{1 - q}.$$

Although this quantity is less important than the Shannon entropy it has found a number of application in information theory and other areas. For a distribution in  $\mathbb{R}^k$  with density

$f$  the differential Rényi entropy of order  $q$  is defined by

$$h_q = \frac{\log \int_{\mathbb{R}^k} f^q}{1 - q}. \quad (1)$$

This quantity shares many properties with the discrete Rényi entropy, but it is far from clear what interpretation of it should be. Like the differential entropy it may even be negative or undefined.

An alternative to the Rényi entropy was introduced first in [1] and later, independently, in [2]. In the discrete case it is given by

$$S_q = \frac{\sum_i p_i^q - 1}{q - 1}$$

and in the continuous case by

$$s_q = \frac{\int_{\mathbb{R}^k} f^q - 1}{q - 1}.$$

There is a simple functional relationship between  $S_q$  and  $H_q$  so that one is a monotone function of the other. In particular Rényi entropy and Tsallis entropy have maximum at exactly the same distributions. Maximizers of Rényi entropy under moment constraints have been studied in [3], [4] and [5]. In this paper we emphasize the physical relevance of such results.

## TSALLIS ENTROPY MAXIMIZERS AS PROJECTIONS

Consider a small gas consisting of  $l$  identical molecules of mass  $m$ . The molecules are confined in a cylinder with elastic walls and all collisions are considered to be elastic. Each velocity vector has 3 coordinates so that the velocities can be described by a vector in  $\vec{v} \in \mathbb{R}^{3l}$ . Then the energy is given by  $\frac{1}{2}m\vec{v}^2$ . If all the degrees of freedom interact (by collisions of the molecules) and the system is isolated then it is natural to assume that the distribution on the sphere  $\{\vec{v} \in \mathbb{R}^{3l} \mid \frac{1}{2}m\vec{v}^2 = lE\}$  is uniform where  $E$  denotes the mean energy per molecule. We may now focus our attention to a single molecule and ask what the distribution of its velocity vector  $\vec{u} \in \mathbb{R}^3$  is, i.e. the marginal distribution of the uniform distribution on a sphere on a three dimensional subspace. It turns out that the distribution of the velocity vector of a single molecule maximizes the Tsallis entropy of order

$$q = \frac{3l - 3}{3l - 5}, \quad l \geq 2 \quad (2)$$

under the constraint that the mean value of  $\frac{1}{2}m\vec{u}^2$  is  $E$ . For  $l$  tending to infinity  $q$  will tend to 1 and the distribution of  $\vec{u}$  will tend to the distribution that maximizes the Shannon entropy.

Instead of molecules in a cylinder one may think of particles in a quadratic potential. In this case the energy is also proportional to  $\vec{v}^2$  where  $\vec{v}$  is the vector in phase space an appropriate scaling of the different axes. We note that any for almost any system of finitely many classical particles the energy is approximately quadratic in the phase

space vector so the above results should approximately hold for any such system. As discussed in [5] the results can even be extended to  $p$ -spheres defined by a condition on the  $p$ -norm. As long as the energy is partly kinetic the expression for the energy will contain terms of order 2 and therefore potentials very different from the elastic potential discussed here will lead to maximizers of Rényi entropy.

Results as 2 are actually easy to derive. We introduce the ball as the set

$$B_n(0, r) = \left\{ (x_1, x_2, \dots, x_n) \in \mathbb{R}^n \mid \sum_{i=1}^n x_i^2 \leq r^2 \right\}.$$

First we note that for any  $k \in \mathbb{N}$  the maximizer of the differential Tsallis entropy  $s_q$  under the variance constraint

$$E \left( \frac{1}{k} \sum_{i=1}^k X_i^2 \right) = \mu$$

is

$$f(x) \sim \left( 1 - \frac{x_1^2 + x_2^2 + \dots + x_k^2}{k\mu} \right)_+^{\frac{1}{q-1}}$$

(see [3] and [? ]). The case  $q = \infty$  is of special interest since we get the uniform distribution on the ball. The uniform distribution on  $B_n(0, (n\mu)^{1/2})$  is projected onto a distribution with support  $B_k(0, (n\mu)^{1/2})$ . The set of points in  $B_n(0, (n\mu)^{1/2})$  with projection  $(x_1, x_2, \dots, x_k) \in B_k(0, (n\mu)^{1/2})$  is the set

$$\left\{ (x_1, x_2, \dots, x_n) \in B_n(0, (n\mu)^{1/2}) \mid \sum_{i=k+1}^n x_i^2 = n\mu - \sum_{i=1}^k x_i^2 \right\}.$$

This is a ball of radius  $(n\mu - \sum_{i=1}^k x_i^2)^{1/2}$  so it has volume proportional to

$$\left( \left( n\mu - \sum_{i=1}^k x_i^2 \right)^{1/2} \right)^{n-k} = \left( n\mu - \sum_{i=1}^k x_i^2 \right)^{\frac{n-k}{2}} \quad (3)$$

$$= (n\mu)^{\frac{n-k}{2}} \left( 1 - \frac{\sum_{i=1}^k x_i^2}{n\mu} \right)^{\frac{n-k}{2}}. \quad (4)$$

We see that this distribution maximizes the differential Tsallis entropy of order  $q$  if

$$\frac{1}{q-1} = \frac{n-k}{2}.$$

This is equivalent to

$$q = \frac{n-k+2}{n-k}. \quad (5)$$

Next we shall consider the projection of a thin area lying between  $S_n(0, n\mu)$  and  $S_n(0, n(\mu + \varepsilon))$  each equipped with uniform distribution. The projection of the (un-normalized) uniform distribution on  $S_{n,2}(0, n\mu)$  is

$$\left( n\mu - \sum_{i=1}^k x_i^2 \right)_+^{\frac{n-k}{2}}$$

and the derivative with respect to  $\mu$  is

$$n \frac{n-k}{2} \left( n\mu - \sum_{i=1}^k x_i^2 \right)_+^{\frac{n-k}{2}-1}.$$

Thus the projection of the sphere with the disintegration measure has density proportional to

$$\left( 1 - \frac{\sum_{i=1}^k x_i^2}{n\mu} \right)_+^{\frac{n-k}{2}-1}$$

and this distribution maximizes the differential entropy of order  $q$  given by the equation

$$\frac{1}{q-1} = \frac{n-k}{2} - 1$$

that has the solution

$$q = \frac{n-k}{n-k-2}. \quad (6)$$

Equation 2 can easily be derived from this more general result.

Now we have observed a maximizer of Tsallis entropy may appear as a projection of uniform distribution on a sphere. Our first observation is that this approach works equally well with Tsallis and Rényi entropy. Therefore it tells absolutely nothing about the interpretation of any of these quantities. Another observation is that derivation has no obvious relation to information theory. So although Rényi entropy has applications in information theory it may be the case that Tsallis entropy may be useful in physics for reasons that have very little to do with information theory. The kind of arguments we have used are mostly related to scaling, so if a link to information theory can be provided then it should be via the scaling properties of Rényi entropy.

For a finite system the Tsallis entropy of a marginal distribution may exhibit the following behavior.

1. A system may have large Tsallis entropy of all marginal distributions although the joint distribution is not uniform.
2. If the dynamics of the system is completely deterministic, the Tsallis entropy a marginal distribution may vary periodically up and down.
3. Even if the dynamics of the system is mixing, the mixing time may be so large that the deviation from a maximum entropy distribution is considerable. Computer experiments confirm this.

Thus we see that several of the properties that are often discussed for ordinary entropy for large systems also appear in small systems with Tsallis entropy in place of ordinary entropy.

## RANDOM MATRICES

What is special about the sphere is its high degree of symmetry, i.e. its invariance under the action of the rotation group. The uniform distribution on the sphere plays a central role for the derivation, so there is a natural relation to the study of processes like random group actions that gives convergence to the uniform distribution. Information theoretic tools have proved useful in the study of such processes [6] and provide via rate distortion theory a natural framework to relate discrete and continuous versions of the different notions of entropy. Here we shall just mention some results and conjectures of physical relevance.

The set of special orthogonal matrices in  $n$  dimensions is denoted  $SO(n)$ . The group is compact and therefore it has a Haar measure, i.e. a measure that is invariant under left and right translations. This measure is finite and therefore it can be normalized to a probability measure  $U$  that we shall call the uniform distribution on the group. The Haar measure can be constructed as follows. Choose  $n$  unit vectors  $\vec{u}_1, \vec{u}_2, \dots, \vec{u}_n$  at random, i.e. they are chosen independently according to a uniform distribution on the unit sphere. We then use the Gram-Smidt orthogonalization method to find an orthonormal basis  $\vec{v}_1, \vec{v}_2, \dots, \vec{v}_n$ . Then the matrix with  $\vec{v}_i$  as its  $i$ 'th column is an orthogonal matrix. We note that  $\vec{v}_1 = \vec{u}_1$  and that the last vector  $\vec{v}_n$  can be chosen so that the matrix becomes a special orthogonal matrix. The distribution of the matrix constructed in this way is obviously invariant under left and right multiplication by a unitary matrix, and therefore the random matrix has the uniform distribution. From the construction we see that for any subset of the entries of the first column of the matrix is a vector that has a distribution that maximizes Tsallis entropy of some order. The columns can be interchanged by multiplication by unitary matrices so any subset of any column forms a vector that maximizes Tsallis entropy of some order. The same kind of argument can be used to show that a vector formed by any subset of the entries in a row has a distribution that maximizes Tsallis entropy.

Next we shall consider an  $n$ -dimensional ball where  $SO(n)$  acts on the system. The dynamics of the system is given actions of elements of  $SO(n)$  so that at each time step an element of  $SO(n)$  acts. We shall assume that there is a little uncertainty about which element of  $SO(n)$  is acting on the system that that this uncertainty is described by the probability distribution  $P$  on  $SO(n)$ . In [6] the following theorems were proved.

**Theorem 1** *Let  $P$  be a probability measure on  $SO(n)$  and assume that the support of  $P$  is not contained in any nontrivial coset of a subgroup of  $G$ . Then  $P^{*k}$  converges to the uniform distribution in the weak topology for  $k$  tending to infinity.*

**Theorem 2** *Let  $P$  be a probability measure on  $SO(n)$  and assume that the support of  $P$  is not contained in any nontrivial coset of a subgroup of  $G$ . If  $P$  has finite entropy then  $D(P^{*k} || U) \rightarrow 0$  for  $k \rightarrow \infty$  where  $D$  denotes information divergence.*

Let  $P^{*k}(\vec{v})$  denote the random vector obtained by applying a random matrix distributed according to  $P^{*k}$ . Combined with our previous discussion we get the following result.

**Theorem 3** *Let  $P$  be a probability measure on  $SO(n)$  and assume that the support of  $P$  is not contained in any nontrivial coset of a subgroup of  $G$ . Let  $v$  denote any vector in  $\mathbb{R}^n$ . Then the  $k$ -dimensional marginal distribution of  $P^{*k}(\vec{v})$  tends a maximizer of Tsallis entropy under a variance constraint in the weak topology for  $n$  tending to infinity. If  $P$  has finite entropy then  $P^{*k}(\vec{v})$  tends the maximizer of Tsallis entropy under a variance constraint in the total variation for  $n$  tending to infinity.*

**Example 1** *Let  $A_0$  and  $A_1$  denote two special orthogonal matrices such that any matrix that commutes with both  $A_0$  and  $A_1$  is a scalar times the identity. Let  $X_1, X_2, \dots$  denote a sequence of independent Bernoulli random variables each with success probability  $1/2$ . Then the distribution of*

$$A_{X_1} A_{X_2} \cdots A_{X_k} \vec{v}$$

*tends to a maximizer of Tsallis entropy in the weak topology for  $k$  tending to infinity.*

One should also note that Wigner's semi-circular law and Grikov's law of the eigenvalue distributions of random matrices are maximizers of Tsallis entropy of order 3 and  $\infty$ .

## DISCUSSION

Many applications of Tsallis entropy can be explained from these results. One may complain that the results do not tell what Tsallis entropy measures, and one may note that Rényi entropy is just a function of Tsallis entropy, so it has the same maximizers. But exactly the same can be said for similar applications of Shannon entropy.

In this talk we shall also discuss the validity of the physical assumptions leading to the maximizer of the Tsallis entropy. Some of these physical assumptions are easier to discuss than normal because the system under consideration is essentially finite so we do not have to go the "thermodynamical limit" or anything like that.

## ACKNOWLEDGMENTS

The author want to thank Christophe Vignat for useful discussions.

## REFERENCES

1. J. Havrda, and F. Charvát, *Kybernetika* **3**, 30–35 (1967), Review by I. Csizár in MR, vol. 34, no.8875.
2. C. Tsallis, *J. Stat. Physics* **52**, 479 (1988), See <http://tsallis.cat.cbpf.br/biblio.htm> for a complete and updated bibliography.
3. J. Costa, A. Hero, and C. Vignat, *On solutions to multivariate maximum  $\alpha$ -entropy problems*, Springer, Berlin, 2003, vol. 2688 of *Lecture Notes in Computer Science*, pp. 211–228.
4. O. Johnson, and C. Vignat, *Annales de l'Institut Henri Poincaré (B) Probability and Statistics* **43**, 339–351 (2007).

5. P. Harremoës, and C. Vignat, Rényi entropies of marginal distributions (2007), URL [homepages.cwi.nl/~ph/Else.pdf](http://homepages.cwi.nl/~ph/Else.pdf), submitted to Journal of Multivariate Analysis.
6. P. Harremoës, “Maximum Entropy on Compact Groups,” in *Proceedings of International Symposium on Information Theory*, IEEE, Seattle, USA, 2006, pp. 108–112, URL <http://homepages.cwi.nl/~ph/compgro.pdf>.