Available online at www.sciencedirect.com
SCIENCE DIRECT®

Physica A 340 (2004) 110–116

PHYSICA A

www.elsevier.com/locate/physa

Generalized statistics: yet another generalization

Petr Jizba*, Toshihico Arimitsu

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

Abstract

We provide a unifying axiomatics for Rényi's entropy and non-extensive entropy of Tsallis. It is shown that the resulting entropy coincides with Csiszár's measure of directed divergence known from communication theory.

PACS: 05.90.+m; 65.40.Gr; 02.90.+p

Keywords: Rényi's information entropy; Tsallis entropy; Non-extensive entropy

1. Introduction

It has been known already since Shannon's seminal paper [1] that Shannon's information measure (or entropy) represents mere idealized information appearing only in situations when the buffer memory (or storage capacity) of a transmitting channel is infinite. As the latter is not satisfied in many practical situations, information theorists have invented various remedies to deal with such cases. This usually consists of substituting Shannon's information measure with information measures of other types. Particularly distinct among them is a one-parametric class of information measures discovered by A. Rényi. It was later on realized by Linnik that these, so-called, Rényi entropies (REs) are associated to the decoding limit if the source is compressed to $\mathscr{I}_q$ and the parameter q essentially tells how much the tail of a probability distribution should count in the calculation of the Rényi entropy. Recently, an operational characterization of RE in terms of β-cutoff rates was provided by Csiszár [2].

On the other hand, pioneering works of E. Jaynes [3] in mid 1950s revealed that the Gibbs entropy of statistical physics represents the Shannon entropy whenever the sample space of Shannon's entropy is identified with the set of all (coarse-grained)

* Corresponding author. Tel.: +81-298-53-4322; fax: +81-298-53-4492.
E-mail addresses: petr@cm.ph.tsukuba.ac.jp (P. Jizba), arimitsu@cm.ph.tsukuba.ac.jp (T. Arimitsu).

0378-4371/$ - see front matter
doi:10.1016/j.physa.2004.03.085

microstates. However, contrary to information theory, tendencies trying to extend the concept of Gibbs's entropy have started to penetrate into statistical physics just recently. This happened by observing that there are indeed many situations of practical interest requiring more "exotic" statistics which do not conform with the classical Gibbsian MaxEnt. Percolation, polymers, protein folding, critical phenomena, cosmic rays, turbulence or stock market returns provide examples.

One obvious way of generalizing Gibbs's entropy would be to look on the axiomatic rules determining Shannon's information measure. In fact, the usual axiomatics of Khinchin [4] offers various "plausible" generalizations. The additivity of independent mean information is then natural axiom to attack. Along these lines only two distinct generalization schemes have been explored in the literature so far. First consists of a redefinition of the statistical mean and second generalizes additivity rule. Respective entropies are then REs [5] and various deformed entropies see e.g. Ref. [6]. While REs are natural tools in statistical systems with a non-standard scaling behavior, deformed entropies seem to be relevant to systems with embedded non-locality. A suitable merger of the above generalizations could provide a new conceptual playground suitable for a statistical description of systems possessing both self-similarity and non-locality. Examples being the early universe cosmological phase transitions or currently much studied quantum phase transitions. In this paper, we attempt to merge REs with a particular class of deformed entropies — Tsallis entropies.

2. Rényi's entropy: entropy of self-similar systems

As already mentioned, RE represents a step towards more realistic situations encountered in communication theory. Among a myriad of information measures REs discern themselves by a firm operational characterization given in terms of block coding and hypotheses testing. Rényi parameter q then represents the so-called β-cutoff rates [2]. RE of order q ($q > 0$) of a discrete distribution $\mathscr{P} = \{p_1, \ldots, p_n\}$ reads

$$\mathscr{I}_q(\mathscr{P}) = \frac{1}{(1-q)} \ln\left(\sum_{k=1}^{n} (p_k)^q \right) . \tag{1}$$

Apart from coding theory REs have proved to be indispensable tools in various branches of physics. Examples being chaotic dynamical systems or multifractals. In his original work Rényi [5] introduced a one-parameter family of information measures (=RE) which he based on axiomatic considerations. In the course of time his axioms have been sharpened by Daróczy [7] and others [8]. Most recently, it was proved in Ref. [9] that RE can be conveniently characterized by the following set of axioms:

(1) For a given integer n and given $\mathscr{P} = \{p_1, p_2, \ldots, p_n\}$ ($p_k \geqslant 0, \sum_k^n p_k = 1$), $\mathscr{I}(\mathscr{P})$ is a continuous with respect to all its arguments.
(2) For a given integer n, $\mathscr{I}(p_1, p_2, \ldots, p_n)$ takes its largest value for $p_k = 1/n$ ($k = 1, 2, \ldots, n$) with the normalization $\mathscr{I}\left(\frac{1}{2}, \frac{1}{2}\right) = \ln 2$.
(3) For a given $q \in \mathbb{R}$ define $\varrho_k(q) = (p_k)^q / \sum_k (p_k)^q$ ($\mathscr{P}$ is affiliated to A) then $\mathscr{I}(A \cap B) = \mathscr{I}(A) + \mathscr{I}(B|A)$ where $\mathscr{I}(B|A) = g^{-1}(\sum_k \varrho_k(q) g(\mathscr{I}(B|A = A_k)))$.

(4) g is invertible and positive in $[0,\infty)$.
(5) $\mathscr{I}(p_1, p_2, \ldots, p_n, 0) = \mathscr{I}(p_1, p_2, \ldots, p_n)$.

Former axioms markedly differ from those utilized in Refs. [5,7,8]. One particularly distinct point is the appearance of the escort distribution $\varrho(q)$ in axiom 3. Note also that RE of two independent experiments is additive. In fact, it was shown in Ref. [5] that RE is the most general information measure compatible with additivity of independent information and Kolmogorov axioms of probability theory.

3. Tsallis entropy: entropy of long-distance correlated systems

Among variety of deformed entropies the currently popular one is the q-additivity prescription and related Tsallis entropy (TE). As the classical additivity of independent information is destroyed in this case, a new more exotic physical mechanisms must be sought to comply with TE predictions. One may guess that the typical playground for TE should be cases when two statistically independent systems have non-vanishing long-range/time correlations: e.g., statistical systems with quantum non-locality. In the case of discrete distributions $\mathscr{P} = \{p_1, \ldots, p_n\}$ TE takes the form

$$\mathscr{S}_q(\mathscr{P}) = \frac{1}{(1-q)}\left[\sum_{k=1}^{n} (p_k)^q - 1\right] . \tag{2}$$

Axiomatic treatment was recently proposed in Ref. [10] and it consists of four axioms:

(1) For a given integer n and given $\mathscr{P} = \{p_1, p_2, \ldots, p_n\}$ $(p_k \geqslant 0, \sum_k^n p_k = 1)$, $\mathscr{S}(\mathscr{P})$ is a continuous with respect to all its arguments.
(2) For a given integer n, $\mathscr{S}(\mathscr{P})$ takes its largest value for $p_k = 1/n$ $(k = 1, 2, \ldots, n)$.
(3) For a given $q \in \mathbb{R}$; $\mathscr{S}(A \cap B) = \mathscr{S}(A) + \mathscr{S}(B|A) + (1-q)\mathscr{S}(A)\mathscr{S}(B|A)$ with $\mathscr{S}(B|A) = \sum_k \varrho_k(q)\mathscr{S}(B|A = A_k)$.
(5) $\mathscr{S}(p_1, p_2, \ldots, p_n, 0) = \mathscr{S}(p_1, p_2, \ldots, p_n)$.

As said before, one keeps here the linear mean but generalizes the additivity law. In fact, the additivity law in axiom 3 is nothing but the Jackson sum of the q-calculus.

4. J-A axioms and solutions

Let us combine the previous two axiomatics in the following natural way:

(1) For a given integer n and given $\mathscr{P} = \{p_1, p_2, \ldots, p_n\}$ $(p_k \geqslant 0, \sum_k^n p_k = 1)$, $\mathscr{D}(\mathscr{P})$ is a continuous with respect to all its arguments.
(2) For a given integer n, $\mathscr{D}(\mathscr{P})$ takes its largest value for $p_k = 1/n$ $(k = 1, 2, \ldots, n)$.
(3) For a given $q \in \mathbb{R}$; $\mathscr{D}(A \cap B) = \mathscr{D}(A) + \mathscr{D}(B|A) + (1-q)\mathscr{D}(A)\mathscr{D}(B|A)$ with $\mathscr{D}(B|A) = f^{-1}(\sum_k \varrho_k(q) f(\mathscr{D}(B|A = A_k)))$.
(4) f is invertible and positive in $[0,\infty)$.
(5) $\mathscr{D}(p_1, p_2, \ldots, p_n, 0) = \mathscr{D}(p_1, p_2, \ldots, p_n)$.

We will now show that the above axioms allow for only one class of solutions which will be closely related to the cross-entropy measures of Havrda and Charvat [11].

5. Basic steps in the proof

Let us first denote $\mathscr{D}(1/n, 1/n, \ldots, 1/n) = \mathscr{L}(n)$. Axioms 2 and 5 then imply that $\mathscr{L}(n) = \mathscr{D}(1/n, \ldots, 1/n, 0) \leqslant \mathscr{D}(1/n+1, \ldots, 1/n+1) = \mathscr{L}(n+1)$. Consequently $\mathscr{L}$ is a non-decreasing function. To determine the form of $\mathscr{L}(n)$ we will assume that $\mathscr{A}^{(1)}, \ldots, \mathscr{A}^{(m)}$ are independent experiments each with r equally probable outcomes

$$\mathscr{D}(\mathscr{A}^{(k)}) = \mathscr{D}(1/r, \ldots, 1/r) = \mathscr{L}(r) \quad (1 \leqslant k \leqslant m)\,. \tag{3}$$

Repeated application of axiom 3 then leads to

$$\begin{aligned}\mathscr{D}(\mathscr{A}^{(1)} \cap \mathscr{A}^{(2)} \cap \cdots \cap \mathscr{A}^{(m)}) = \mathscr{L}(r^m) &= \sum_{i=1}^{m} \binom{m}{i} (1-q)^{i-1} \mathscr{D}^i(\mathscr{A}^{(i)}) \\ &= \frac{1}{(1-q)} \left[(1 + (1-q)\mathscr{L}(r))^m - 1\right].\end{aligned} \tag{4}$$

Taking partial derivative of both sides of (4) with respect to m and putting $m = 1$ afterwards we get the differential equation

$$\frac{(1-q)\,\mathrm{d}\mathscr{L}}{(1 + (1-q)\mathscr{L})[\ln(1 + (1-q)\mathscr{L})]} = \frac{\mathrm{d}r}{r \ln r}\,. \tag{5}$$

It is easy to verify that the general solution of (5) has the form

$$\mathscr{L}(r) \equiv \mathscr{L}_q(r) = \frac{1}{1-q}\,(r^{c(q)} - 1)\,. \tag{6}$$

Function $c(q)$ will be determined later on. Right now we just note that because at $q = 1$ Eq. (4) boils down to $\mathscr{L}(r^m) = m\mathscr{L}(r)$ we have $c(1) = 0$. We proceed by considering the experiment with outcomes $\mathscr{A} = (\mathscr{A}_1, \mathscr{A}_2, \ldots, \mathscr{A}_n)$ and the distribution $\mathscr{P} = (p_1, p_2, \ldots, p_n)$. Assume moreover that p_k $(1 \leqslant k \leqslant n)$ are rational numbers, i.e., $p_k = g_k/g$, $\sum_{k=1}^{n} g_k = g$ with $g_k \in \mathbb{N}$. Let us have, furthermore, an experiment $\mathscr{B} = (\mathscr{B}_1, \mathscr{B}_2, \ldots, \mathscr{B}_g)$ with distribution $\mathscr{Q} = \{q_1, q_2, \ldots, q_g\}$. We split $(\mathscr{B}_1, \mathscr{B}_2, \ldots, \mathscr{B}_g)$ into n groups containing $g_1, g_2, \ldots, g_n$ outcomes, respectively. Consider now a particular situation in which whenever event $\mathscr{A}_k$ happens then in $\mathscr{B}$ all g_k events of kth group occur with the equal probability $1/g_k$ and all the other events in $\mathscr{B}$ have probability zero. Hence $\mathscr{D}(\mathscr{B}|\mathscr{A} = \mathscr{A}_k) = \mathscr{D}(1/g_k, \ldots, 1/g_k) = \mathscr{L}_q(g_k)$ and so by axiom 3 we have

$$\mathscr{D}(\mathscr{B}|\mathscr{A}) = f^{-1}\left(\sum_{k=1}^{n} \varrho_k(q) f(\mathscr{L}_q(g_k))\right)\,. \tag{7}$$

On the other hand, for our system the entropy $\mathscr{D}(\mathscr{A} \cap \mathscr{B})$ can be easily evaluated. Realizing that the joint probability distribution corresponding to $\mathscr{A} \cap \mathscr{B}$ is

$$\mathscr{R} = \{r_{kl} = p_k q_{l|k}\} = \left\{\underbrace{\frac{p_1}{g_1}, \ldots, \frac{p_1}{g_1}}_{g_1\times}, \underbrace{\frac{p_2}{g_2}, \ldots, \frac{p_2}{g_2}}_{g_2\times}, \ldots, \underbrace{\frac{p_n}{g_n}, \ldots, \frac{p_n}{g_n}}_{g_n\times}\right\} = \{1/g, \ldots, 1/g\}\,,$$

we obtain that $\mathscr{D}(\mathscr{A} \cap \mathscr{B}) = \mathscr{L}_q(g)$. Utilizing the first part of axiom 3 and defining $f_{(a,y)}(x) = f(-ax + y)$ we can write

$$\mathscr{D}(\mathscr{A}) = \frac{f^{-1}_{(a,\mathscr{L}(g))}\left(\sum_k \varrho_k(q) f_{(a,\mathscr{L}(g))}(-\mathscr{L}_q(p_k))\right)}{1-(1-q)f^{-1}_{(a,\mathscr{L}(g))}\left(\sum_k \varrho_k(q) f_{(a,\mathscr{L}(g))}(-\mathscr{L}_q(p_k))\right)} ,$$

$$a = [1 + (1-q)\mathscr{L}_q(g)] .$$

As Eq. (6) indicates it is $\mathscr{L}_q(1/p_k)$ and not $-\mathscr{L}_q(p_k)$ which represents the elementary information of order q affiliated with p_k. Thus using the relation

$$\mathscr{L}_q(p_k) = -\frac{\mathscr{L}_q(1/p_k)}{1 + (1-q)\mathscr{L}_q(1/p_k)} , \tag{8}$$

together with transformation

$$g(x) = f_{(a,\mathscr{L}(g))}\left(\frac{x}{1+(1-q)x}\right) , \tag{9}$$

we easily obtain that

$$\mathscr{D}(\mathscr{A}) = g^{-1}\left(\sum_k \varrho_k(q) g(\mathscr{L}_q(1/p_k))\right) = f^{-1}\left(\sum_k \varrho_k(q) f(\mathscr{L}_q(1/p_k))\right) . \tag{10}$$

The last identity is due to second part of Axiom 3. It is well known from the theory of means [12] that (10) can be fulfilled if $g(x)$ is a linear function of $f(x)$, i.e.,

$$g(x) = f\left(\frac{-x+y}{1+(1-q)x}\right) = \theta_q(y) f(x) + \vartheta_q(y) . \tag{11}$$

In order to solve (11) we define $\varphi(x) = f(x) - f(0)$. With this notation (11) turns into

$$\varphi\left(\frac{-x+y}{1+(1-q)x}\right) = \theta_q(y)\varphi(x) + \varphi(y), \quad \varphi(0) = 0 . \tag{12}$$

By setting $x = y$ we see that $\theta_q(y) = -1$, hence one finds

$$\varphi(x + y + (1-q)xy) = \varphi(x) + \varphi(y) . \tag{13}$$

Eq. (13) is Pixeder's functional equation which can be solved by the standard method of iterations [13]. Eq. (13) has only one non-trivial class of solutions, namely

$$\varphi(x) = \frac{1}{1-\alpha} \ln(1 + (1-q)x) . \tag{14}$$

α is here a free parameter. Plugging this solution back to (10) we obtain

$$\mathscr{D}_q(\mathscr{A}) = \frac{1}{1-q}(\mathrm{e}^{-c(q)\sum_k \varrho_k(q) \ln p_k} - 1) = \frac{1}{1-q}\left(\prod_k (p_k)^{-c(q)\varrho_k(q)} - 1\right) . \tag{15}$$

Note that the constant α got cancelled. It remains to determine $c(q)$. Utilizing the conditional entropy constructed from (15) and using Axiom 3, we obtain $c(q)=1-q$. Inasmuch we can recast (15) into more expedient form

$$\mathscr{D}_q(\mathscr{A}) = \frac{1}{1-q}\left(\mathrm{e}^{-(1-q)^2 \mathrm{d}\mathscr{I}_q/\mathrm{d}q} \sum_{k=1}^{n} (p_k)^q - 1 \right) . \tag{16}$$

Eqs. (15) and (16) are the sought results. In passing we note that $\mathscr{D}_q \geqslant 0$ for $\forall q \in \mathbb{R}$ and $\lim_{q\to 1} \mathscr{D}_q = \mathscr{I}_1 = \mathscr{S}_1$.

6. Conclusions and outlooks

Presented axiomatics might provide a novel playground for a q non-extensive systems with embedded self–similarity. Indeed, one could expect that the obtained measure of information could play a relevant rôle in q non-extensive statistical systems near critical points. Research in this direction is currently in progress.

A curious result arises when one restricts values of $\mathscr{P}$ by the constraint $\mathrm{d}\mathscr{I}_q(\mathscr{P})/\mathrm{d}q = \max_{p_i} \mathscr{I}_q(\mathscr{P})/(1-q)$. Eq. (16) then boils down to

$$\mathscr{D}_q(\mathscr{A}) = \frac{1}{1-q}\left(n^{q-1} \sum_{k=1}^{n} (p_k)^q - 1 \right) \equiv -\mathscr{C}_q(\mathscr{A}) . \tag{17}$$

The reader may recognize in $\mathscr{C}_q$ the generalized measure of cross entropy of Havrda and Charvat [11] (also known as Csiszár's measure of directed divergence [14]) used in communication theory. For $q=2$ we recover the χ^2 measure. This suggests that the non-extensivity together with self-similarity may be important concepts also in information theory. In this connection such issues as the channel capacitance and cutoff rates would deserve a separate discussion.

The generalized entropy $\mathscr{D}_q$ has many desirable features: like Tsallis entropy it satisfies the non-extensive q-additivity, involves a single parameter q, and goes over into the standard Shannon entropy in the limit $q \to 1$. On that basis it would appear that both $\mathscr{S}_q$ and $\mathscr{D}_q$ have an equal right to furnish a generalization of statistical mechanics.

References

[1] C.E. Shannon, Bell Syst. Tech. J. 27 (1948) 379, 623.
[2] I. Csiszár, IEEE Trans. Inf. Theory 41 (1995) 26.
[3] E.T. Jaynes, Phys. Rev. 106 (1957) 171;
E.T. Jaynes, Phys. Rev. 108 (1957) 620.
[4] A.I. Khinchin, Mathematical Foundations of Information Theory, Dover Publications, Inc., New York, 1957.
[5] A. Rényi, Probability Theory, North-Holland, Amsterdam, 1970;
A. Rényi, Selected Papers of Alfred Rényi, Vol. 2, Akadémia Kiado, Budapest, 1976.
[6] C. Tsallis, J. Stat. Phys. 52 (1988) 479;
C. Tsallis, Braz. J. Phys. 29 (1999) 1;
J. Naudts, Physica A 316 (2002) 323;
G. Kaniadakis, Phys. Rev. E 66 (2002) 056125.

[7] Z. Daróczy, Acta Math. Acad. Sci. Hung. 15 (1964) 203.
[8] S. Guias, Information Theory with Applications, McGraw-Hill, New York, 1977.
[9] P. Jizba, T. Arimitsu, Ann. Phys. (N.Y.) (to be published), cond-mat/0207707.
[10] S. Abe, Phys. Lett. A 271 (2000) 74.
[11] J.H. Havrda, F. Charvat, Kybernatica 3 (1967) 30.
[12] G.H. Hardy, J.E. Littlewood, G. Pólya, Inequalities, Cambridge University Press, Cambridge, 1952.
[13] J. Aczél, Lectures on Functional Equations and their Applications, Academic Press, New York, 1966.
[14] I. Csiszár, Period Math. Hung. 2 (1972) 191.