On \(q\)-non-extensive statistics with non-Tsallisian entropy

Petr Jizba\(^{a,b}\), Jan Korbel\(^{a,c}\)

\(^{a}\)Faculty of Nuclear Sciences and Physical Engineering, Czech Technical University in Prague, Břehová 7, 11519, Prague, Czech Republic
\(^{b}\)Institute of Theoretical Physics, Freie Universität in Berlin, Arnimallee 14, 14195 Berlin, Germany
\(^{c}\)Max Planck Institute for the History of Science, Boltzmannstrasse 22, 14195 Berlin, Germany

Abstract

We combine an axiomatics of Rényi with the \(q\)-deformed version of Khinchin axioms to obtain a measure of information (i.e., entropy) which account both for systems with embedded self-similarity and non-extensivity. We show that the entropy thus obtained is uniquely solved in terms of a one-parameter family of information measures. The ensuing maximal-entropy distribution is expressible via a special function known under the name of the Lambert W–function. We analyze the corresponding “high” and “low-temperature” asymptotics and reveal a non-trivial structure of the parameter space.

© 2014 Published by Elsevier Ltd.

Keywords: Multifractals, Rényi’s information entropy, THC entropy, MaxEnt, heavy-tailed distributions

PACS: 65.40.Gr, 47.53.+n, 05.90.+m

1. Introduction

In his 1948 paper [1] Shannon formulated the theory of data compression. The paper established a fundamental limit to lossless data compression and showed that this limit coincides with the information measure presently known as Shannon’s entropy \(H\). In words, it is possible to compress the source, in a lossless manner, with compression rate close to \(H\), it is mathematically impossible to do better than \(H\). However, many modern communication processes, including signals, images and coding/decoding systems, often operate in complex environments dominated by conditions that do not match the basic tenets of Shannon’s communication theory. For instance, buffer memory (or storage capacity) of a transmitting channel is often finite, coding can have a non–trivial cost function, codes might have variable-length codes, sources and channels may exhibit memory or losses, etc. Information theory offers various generalized (non–Shannonian) measures of information to deal with such cases. Among the most frequently used one can mention Havrda–Charvát measure [3], Sharma–Mittal measure [3], Rényi’s measure [4] and Kapur’s measures [5]. Information entropies get even more complex by considering communication systems with quantum channels [6, 7]. There exists even attempts to generalize Shannon’s measure of information in the direction where no use of the concept of probability is needed hence demonstrating that information is more primary notion than probability [8].

In mid 1950 Jaynes [9] proposed the Maximum Entropy Principle (MaxEnt) as a general inference procedure that, among others, bears a direct relevance to statistical mechanics and thermodynamics. The conceptual frame of Jaynes’s...
MaxEnt is formed by Shannon’s communication theory with Shannon’s information measure as an inference functional. The central rôle of Shannon’s entropy as a tool for inductive inference (i.e., inference where new information is given in terms of expected values) was further demonstrated in works of Faddeyev [10], Shore and Johnson [11], Wallis [12], Topsøe [13] and others. In Jaynes’s procedure the laws of statistical mechanics can be viewed as inferences based entirely on prior information that is given in the form of expected values of energy, energy and number of particles, energy and volume, energy and angular momentum, etc., thus rederiving the familiar canonical ensemble, grand-canonical ensemble, pressure ensemble, rotational ensemble, etc., respectively [14]. Remarkable feature of this procedure is that often utilized additional hypotheses such as ergodicity or metric transitivity are not needed here. Following Jaynes, one should view the MaxEnt distribution (or maximizer) as a distribution that is maximally noncommittal with regard to missing information and that agrees with all what is known about prior information, but expresses maximum uncertainty with respect to all other matters [9]. By identifying the statistical sample space with the set of all (coarse-grained) microstates the corresponding maximizer yields the Shannon entropy that corresponds to the Gibbs entropy of statistical physics.

Surprisingly, despite the aforementioned linkage and related advancements in information theory, tendencies aiming at comparable extension of Gibbs’s entropy started to penetrate into statistical physics only in the last two decades. This happened when evidence accumulated showing that there are indeed many situations of practical interest requiring more “exotic” statistics which do not conform with Gibbsian exponential maximizers. Percolation, protein folding, critical phenomena, cosmic rays, turbulence, granular matter or stock market returns provide examples.

In attacking the problem of generalization of Gibbs’s entropy the information theoretic route to equilibrium statistical physics provides a very useful conceptual guide. The natural strategy that fits this framework would be then to revisit the axiomatic rules governing Shannon’s information measure and potential extensions translate into language of statistical physical. In fact, the usual axioms of Khinchin [15] is prone to several “plausible” generalizations. Among those, the additivity of independent mean information is a natural axiom to attack. Along those lines, two fundamentally distinct generalization schemes have been pursued in the literature; one redefining the statistical mean and another generalizing the additivity rule.

The first mentioned generalization was realized by Rényi by employing the most general means still compatible with Kolmogorov axioms of probability theory. These, so called, quasi-linear means were independently studied by Kolmogorov [16] and Nagumo [17]. It was shown that the generalization based on quasi-linear means unambiguously leads to information measures known as Rényi entropies [4, 18]. Although, the status of Rényi entropies (RE) in statistical physics is still debated they nevertheless provide an immensely important analyzing tool in classical statistical systems with a non-standard scaling behavior (e.g., fractals, multifractals, etc.) [19].

On the other hand, the second approach generalizes the additivity prescription but keeps the usual linear mean. Currently popular generalization is the \( q \)-additivity prescription and related \( q \)-calculus [20, 21]. The corresponding axiomatics [22] provides the entropy known as Tsallis–Havrda–Charvát’s (THC) entropy. As the classical additivity of independent information is destroyed in this case, a new more exotic physical mechanisms must be sought to comply with THC predictions. Recent theoretical advances in systems with long range interactions [23] and in correlated (or \( q \)-central limit theorems [29] indicate that the typical playground of THC entropy should be in cases where two statistically independent systems have non-vanishing long-range/time correlations or where the notion of statistical independence is an ill defined concept. Examples include, long-range Ising models, gravitational systems, statistical systems with quantum non-locality, etc. Along this line of reasoning there has been made a successful attempt to describe some general classes of entropies (including THC entropy) by their asymptotic scaling [24]. In passing we mention that there exists a number of works trying to compare both Rényi and THC entropies from both the theoretical and observational point of view (see, e.g., Refs. [27, 28]). Nevertheless, the intersection of both entropic paradigms has not been studied yet.

It is clear that an appropriate combination of the above generalizations could provide a new conceptual paradigm suitable for a statistical description of systems possessing both self-similarity and non-locality. Such systems are quite pertinent with examples spanning from the early universe cosmological phase transitions to currently much studied quantum phase transitions (frustrated spin systems, Fermi liquids, etc.). It is aim of this paper to pursue this line

---

1Other important approaches such as Kaniadakis’s [23] and Naudts’s [24] deformed Hartley’s logarithmic information also utilize linear means and generalized additivity rule (e.g., \( \alpha \)-additivity) but as yet they still lack the information-theoretic axiomatics that is crucial in our reasonings. For this reason we exclude these works from our consideration.
of reasonings and explore the resulting implications. In order to set a stage for our considerations we review in the following section some axiomatic essentials for both Shannon, Rényi and THC entropies that will be needed in the main body of the paper. In Section 2.2 we then formulate a new axiomatics which aims at bridging the Rényi and THC entropies. It is found that such axiomatics allows for only one one-parametric family of solutions. Basic properties of the new entropy that we denote as \( D_q \) are discussed in Section Appendix B. In particular, a simplification that \( D_q \) undergoes in multifractal systems is emphasized. The corresponding MaxEnt distributions are calculated in Section 4. We utilize both linear and non–linear moment constraints (applied to the energy) to achieve this goal. In both aforementioned cases the distributions are expressible through the Lambert W–function. We analyze the corresponding “high” and “low-temperature” asymptotics and discuss the ensuing non-trivial structure of the parameter space. Section 5 is devoted to conclusions. The paper is substituted with two appendices which clarify some finer mathematical points.

2. Brief review of entropy axiomatics

The information measure, or simply entropy, is supposed to represent the measure or degree of uncertainty or expectation in conveyed information which is going to be removed by the recipient. As a rule in information theory the exact value of entropy depends only on the information source — more specifically, on the statistical nature of the source. Generally speaking, the higher is the information measure the higher is the ignorance about the system (source) and thus more information will be uncovered after the message is received (or an actual measurement is performed). As often happens, this simple scenario is not frequently tenable as various restrictive factors are always present in realistic situations; finite buffer capacity, global patterns in messages, topologically non–trivial sample spaces, etc.. One may even entertain various information theoretic implications related with the quantum probability calculus or quantum communication channels. Thus, as we go to somewhat more elaborate and realistic models, the entropy prescriptions get more complicated and realistic!

To see why a new generalization of the entropy is desirable let us briefly dwell into 3 most common entropy protagonist, namely Shannon’s, Rényi’s and THC entropy.

2.1. Shannon’s entropy — Khinchin axioms

The best known and widely used information measure is Shannon’s entropy. For the completeness sake we now briefly recapitulate the Khinchin axiomatics as this will prove important in what follows. The latter consist of four axioms [15]:

1. For a given integer \( n \) and given \( \mathcal{P} = \{p_1, p_2, \ldots, p_n\} (p_k \geq 0, \sum_{k=1}^n p_k = 1) \), \( S(\mathcal{P}) \) is a continuous with respect to all its arguments.
2. For a given integer \( n \), \( H(p_1, p_2, \ldots, p_n) \) takes its largest value for \( p_k = 1/n (k = 1, 2, \ldots, n) \) with the normalization \( I \left( \frac{1}{n}, \frac{1}{n} \right) = 1 \).
3. For a given \( q \in \mathbb{R} \); \( H(A \cap B) = H(A) + H(B|A) \) with \( H(B|A) = \sum_k p_k H(B|A = A_k) \), and distribution \( \mathcal{P} \) corresponds to the experiment \( A \).
4. \( H(p_1, p_2, \ldots, p_n, 0) = H(p_1, p_2, \ldots, p_n) \), i.e., adding an event of probability zero (impossible event) we do not gain any new information.

The corresponding information measure, Shannon’s entropy, then reads

\[
S(\mathcal{P}) = -\sum_{k=1}^n p_k \ln p_k. \tag{1}
\]

In passing we should stress two important points. Firstly, axiom 3 indicates that Shannon’s entropy of two independent experiments (sources) is additive. Secondly, by identifying thermodynamic microstates with all possible outcomes of some experiment/measurement then Shannon’s entropy will coincide with Gibbs entropy.
2.2. Rényi’s entropy: entropy of multifractal systems

As already mentioned, RE represents a step further towards more realistic situations encountered in information theory. Among a myriad of information measures RE distinguish themselves by firm operational characterizations. These were established by Arikan [30] for the theory of guessing, by Jelinek [31] for the buffer overflow problem in lossless source coding, by Cambell [32] for the lossless variable–length coding problem with an exponential cost constraint, etc. Recently an interesting operational characterization of RE was provided by Csiszár [33] in terms of block coding and hypotheses testing. In the latter case the Rényi parameter \( q \) was directly related to so–called \( \beta \)-cutoff rates [33].

Apart from information theory RE have proved to be indispensable tool also in numerous branches of physics. Typical examples are provided by chaotic dynamical systems and multifractal statistical systems (see e.g., [34] and citations therein). Fully develop turbulence, earthquake analysis and generalized dimensions of strange attractors provide examples.

RE of order \( q (q > 0) \) of a discrete distribution \( \mathcal{P} = \{p_1, \ldots, p_n\} \) are defined as

\[
I_q(\mathcal{P}) = \frac{1}{1-q} \log_2 \left( \sum_{k=1}^{n} (p_k)^q \right).
\]

In his original work Rényi [4, 18] introduced a one-parameter family of information measures \( I \) which he based on axiomatic considerations. In the course of time these axioms have been sharpened by Darótzky [35] and others [36]. Most recently we have proved in Ref. [34] that RE can be uniquely derived from the following set of axioms:

1. For a given integer \( n \) and given \( \mathcal{P} = \{p_1,\ldots,p_n\} \) \( (p_k \geq 0, \sum_{k} p_k = 1) \), \( I(\mathcal{P}) \) is a continuous with respect to all its arguments.
2. For a given integer \( n \), \( I(p_1, p_2, \ldots, p_n) \) takes its largest value for \( p_k = 1/n \) \( (k = 1, 2, \ldots, n) \) with the normalization \( I(\frac{1}{n}, \ldots, \frac{1}{n}) = 1 \).
3. For a given \( q \in \mathbb{R} \); \( I(A \cap B) = I(A) + I(B|A) \) with \( I(B|A) = g^{-1}(\sum_k \varphi_k(q) g(I(B|A = A_k))) \), and \( \varphi_k(q) = p_k^q / \sum_k p_k^q \) (distribution \( \mathcal{P} \) corresponds to the experiment \( A \)).
4. \( g \) is invertible and positive in \([0, \infty)\).
5. \( I(p_1, p_2, \ldots, p_n, 0) = I(p_1, p_2, \ldots, p_n) \).

Former axioms markedly differ from those utilized in [4, 18, 35, 36]. Particularly distinct is the appearance of the escort (or zooming) distribution \( \varphi(q) \) in the 3rd axiom. Distribution \( \varphi(q) \) was originally introduced by Beck and Schrögl [37] in the context of non-linear dynamics.

We briefly remind some elementary properties of \( I_q \); it is symmetric in all arguments, for \( q \leq 1 \) is \( I_q \) a concave function and \( \mathcal{H}(\mathcal{P}) \leq I_q(\mathcal{P}) \), while for \( q \geq 1 \) it is neither concave nor convex and \( I_q(\mathcal{P}) \leq \mathcal{H}(\mathcal{P}) \). Some further properties can be found, e.g., in Refs. [4, 18, 35].

Note particularly that RE of two independent experiments (sources) is additive. In fact, it was noticed in Ref. [4] that RE is the most general information measure compatible with additivity of independent information and Kolmogorov axioms of probability theory.

2.3. THC entropy: entropy of long distance correlated systems

THC entropy was originally introduced in 1967 by Havrda and Charvát in the context of information theory of computerized systems [2] and together with the \( \alpha \)-norm entropy measure \( H_\alpha \) it belongs to class of pseudo-additive entropies. In contrast with Rényi’s or Shannon’s entropy THC entropy does not have (as yet) an operational characterization. Havrda–Charvát structural entropy, though quite well known among information theorists, had remained largely unknown in physics community. It took more than two decades till Tsallis in his pioneering work [39] on generalized (or non-extensive) statistics rediscovered this entropy. Since then the THC entropy has been employed in many physical systems. In this connection one may particularly mention, Hamiltonian systems with long-range interactions, granular systems, complex networks, stock market returns, etc. For recent review see, e.g., Ref. [40].

In the case of a discrete distribution \( \mathcal{P} = \{p_1, \ldots, p_n\} \) the THC entropy takes the form:
\[ S_q(\mathcal{P}) = \frac{1}{1-q} \left[ \sum_{k=1}^{n} (p_k)^q - 1 \right], \quad q > 0. \]  

Axiomatic treatment was recently proposed in Ref. [22] and it consists of four axioms:

1. For a given integer \( n \) and given \( \mathcal{P} = \{p_1, p_2, \ldots, p_n\} (p_k \geq 0, \sum p_k = 1) \), \( S(\mathcal{P}) \) is a continuous with respect to all its arguments.
2. For a given integer \( n \), \( S(p_1, p_2, \ldots, p_n) \) takes its largest value for \( p_k = 1/n \) (\( k = 1, 2, \ldots, n \)).
3. For a given \( q \in \mathbb{R} \), \( S(A \cap B) = S(A) + S(B|A) + (1 - q)S(A)S(B|A) \) with 
   \[ S(B|A) = \sum_k q_k(q) S(B|A = A_k), \]
   and \( q_k(q) = p_k^q / \sum_k p_k^q \) (distribution \( \mathcal{P} \) corresponds to the experiment \( A \)).
4. \( S(p_1, p_2, \ldots, p_n, 0) = S(p_1, p_2, \ldots, p_n) \).

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 known from the \( q \)-calculus [41]; there one defines the Jackson basic number \([X]_q \) of quantity \( X \) as

\[ [X]_q = (q^X - 1)/(q - 1) \Rightarrow [X + Y]_q = [X]_q + [Y]_q + (q - 1)[X]_q[Y]_q. \]  

The connection with axiom 3 is then established when \( q \to (2 - q) \). Nice feature of the \( q \)-calculus is that it formalizes many mathematical manipulations. For instance, using the \( q \)-logarithm

\[ \ln_{[2-q]} x = -\ln_{[2-q]} \left( \frac{1}{x} \right) = \frac{1}{1-q} (x^{1-q} - 1), \]  

THC entropy can be concisely written as the \( q \)-deformed Shannon’s entropy, i.e.,

\[ S_q(\mathcal{P}) = -\sum_{k=1}^{n} p_k \ln_{[2-q]} p_k = -\sum_{k=1}^{n} p_k^q \ln_{[q]} p_k = \sum_{k=1}^{n} p_k \ln_{[q]} \left( \frac{1}{p_k} \right). \]

Some elementary properties of \( S_q \) are positivity, concavity for all values of \( q \) and indeed nonextensivity. There hold also inequalities between all three entropies, namely:

\[ \mathcal{H}(\mathcal{P}) \leq I_q(\mathcal{P}) \leq S_q(\mathcal{P}) \]  

for \( 0 < q \leq 1 \), and

\[ S_q(\mathcal{P}) \leq I_q(\mathcal{P}) \leq \mathcal{H}(\mathcal{P}) \]  

for \( q \geq 1 \)

for monographs that cover this subject in more depth the reader is referred to Ref. [42].

### 3. J-A axioms and solutions

In this section we propose a unifying axiomatic framework in which both Rényi’s and THC entropies are evenly represented. To do so we synthesize the previous two axiomatics in the following natural way:

1. For a given integer \( n \) and given \( \mathcal{P} = \{p_1, p_2, \ldots, p_n\} (p_k \geq 0, \sum p_k = 1) \), \( D(\mathcal{P}) \) is a continuous with respect to all its arguments.
2. For a given integer \( n \), \( D(p_1, p_2, \ldots, p_n) \) takes its largest value for \( p_k = 1/n \) (\( k = 1, 2, \ldots, n \)).
3. For a given \( q \in \mathbb{R} \), \( D(A \cap B) = D(A) + D(B|A) + (1 - q)D(A)D(B|A) \) with

\[ D(B|A) = f^{-1} \left( \sum_k q_k(q) D(B|A = A_k) \right), \]

and \( q_k(q) = p_k^q / \sum_k p_k^q \) (distribution \( \mathcal{P} \) corresponds to the experiment \( A \)).
4. $f$ is invertible and positive in $[0, \infty)$.
5. $D(p_1, p_2, \ldots, p_n, 0) = D(p_1, p_2, \ldots, p_n)$.

Note particularly that due to the non-linear nature of the non-additivity condition there is no need to select a normalization condition for $D_q$. In Ref. [30] it is shown that above axioms allow for only one class of solutions, leading to an entirely new family of physically conceivable entropy functions. For reader’s convenience are the basic steps of the proof sketched in Appendix A. This results into the following form of entropy:

$$D_q(A) = \frac{1}{1-q} \left( e^{-(1-q)dI_q/dq} \sum_{k=1}^{n} (p_k)^q - 1 \right) = \ln(q) e^{-(\ln P(a))}.$$  \hspace{1cm} (9)

Let us further remark that axiom 5 restricts the possible values of $q$ to $q \geq 0$. This is because $D_q$ would otherwise tend to infinity if some of $p_k$ would tend to zero. The latter would be counter-intuitive, because without changing the probability distribution we would gain an infinite information. Value $q = 0$ must be also ruled out on the basis of axiom 2, because $D_0$ would yield an expression not dependent on the probability distribution $P$ but only on the number of outcomes (or events) — i.e., $D_0$ would be a system (source) insensitive. The basic properties of hybrid entropy are presented in Appendix B.

Before studying the implications of the formula (9), there are two immediate consequences which warrant special mention. The first is that, from the condition $dI_q/dq \leq 0$ (see Section 2.2) we have

$$D_q(A) = \begin{cases} \geq S_q(A) & \text{if } q \leq 1 \\ \leq S_q(A) & \text{if } q \geq 1 \end{cases},$$  \hspace{1cm} (10)

with equality holding, if and only if, $q = 1$ or $dI_q/dq = 0$. These mean that either $D_q(A)$ and $S_q(A)$ jointly coincide with Shannon’s entropy or that $P$ is uniform or $\{1, 0, \ldots, 0\}$. Hence, combining with inequalities between THC, Rényi and Shannon entropy, we obtain

$$0 \leq H(P) \leq I_q(P) \leq S_q(P) \leq D_q(P) \leq H'(P) \leq \ln n \quad \text{for } 0 < q \leq 1 ,$$

$$0 \leq D_q(P) \leq S_q(P) \leq I_q(P) \leq H(P) \leq \ln n \quad \text{for } q \geq 1 .$$  \hspace{1cm} (11)

The result (11) implies that by investigating the information measure $D_q$ with $q < 1$ we receive more information than restricting our investigation just to entropies $I_q$ or $S_q$. On the other hand, when $q > 1$ then both $I_q$ and $S_q$ are more informative than $D_q$. In practical cases one usually requires more than one $q$ to gain more complete information about the system. In fact, when entropies $I_q$ or $S_q$ are used, it is necessary to know them for all $q$ in order to obtain a full information on a given statistical system [34]. For applications in strange attractors the reader may consult Ref. [44], for reconstruction theorems see, e.g., Refs. [44] [44].

The second comment to be made concerns the fact that when the statistical system in question is a multifractal then relations [33-36] assert that

$$(1 - q) \frac{dI_q}{dq} = (a - f(a)) \ln e = \ln \left( \sum_{k} p_k(e) \right).$$  \hspace{1cm} (12)

where summation runs only over support boxes of size $e$ with the scaling exponent $a$. Alternatively, we could have started with the first relation in Eq. (A19) and used the multifractal canonical relations (see Ref. [44]) in which case the result would have been again (12). So for the coarse-grained multifractal with the mesh size $e$ the corresponding entropy $D_q$ reads

$$D_q(A) = \frac{1}{(1-q)} \left( \frac{\sum_{k=1}^{n} (p_k(e))^q}{\sum_{k}^{N_0} p_k(e)} - 1 \right) .$$  \hspace{1cm} (13)

\footnote{The necessary essentials on multifractals are presented in Appendix C.}
Now, the passage from multifractals to single-dimensional statistical systems is done by assuming that the \( a \)-interval is infinitesimally narrow and that PDF is smooth \([14, 15]\). In such a case Cvitanovic’s condition \([45]\) holds, namely both \( a \) and \( f(a) \) collapse to \( a = f(a) = D \) and \( q = f'(a) = 1 \). So, for example, for a statistical system with a smooth PDF and the support space \( \mathbb{R}^d \) the entropy \( D_q \) coincides with Shannon’s \( \mathcal{H} \). As a byproduct of Eq. \((11)\) we may notice that for single-dimensional systems with smooth PDF’s \( S_q \) and \( I_q \) must approach Shannon’s entropy \([3] \). We remark that this may help to understand why Shannon’s entropy plays such a predominant rôle in physics of single-dimensional sets.

In what follows, we examine the class of distributions that represent maximizers for \( D_q(A) \) subject to certain physical constraints.

4. MaxEnt distribution

According to information theory, the maximal entropy principle yields distributions which reflect least bias and maximum uncertainty about information not provided to a recipient (i.e., observer). Important feature of the usual Gibbsian MaxEnt formalism is that the maximal value of entropy is a concave function of the values of the prescribed constraints (moments), and maximizing probabilities are all greater than zero \([46]\). The first is important for thermodynamical stability and the second for mathematical consistency. In this section we will see that both mentioned features hold true also in the case of the \( D_q \) entropy.

Let us first address the issue of maximizers for \( D_q \). To this end we shall seek the conditional extremum of \( D_q \) subject to the constraints imposed by the averaged value of energy \( E \) (or generally any random quantity representing the constant of the motion) in the form

\[
\langle E \rangle_r = \sum_k q_k(r) E_k.
\]  

(14)

For the future convenience we initially keep \( r \) not necessary coincident with \( q \). Taking into account the normalization condition for \( p_i \), we ought to extremize the functional

\[
L_{q,r}(\mathcal{P}) = D_q(\mathcal{P}) - \Omega \sum_k (p_k \ln q_k E_k) - \Phi \sum_k p_k,
\]  

(15)

with \( \Omega \) and \( \Phi \) being the Lagrange multipliers. Setting the derivatives of \( L_{q,r}(\mathcal{P}) \) with respect to \( p_1, p_2, \ldots, \) etc., to zero, we obtain

\[
\frac{\partial L_{q,r}(\mathcal{P})}{\partial p_i} = e^{(q-1)\sum_k \ln p_k} \left( q \langle \ln \mathcal{P} \rangle_q - \ln p_i \right) \frac{(p_i)^{q-1}}{\sum_k (p_k)^q} - r \Omega \langle E_i \rangle - \langle E \rangle_q - \Phi = 0, \quad i = 1, 2, \ldots, n.
\]  

(16)

Note that when both \( q \) and \( r \) approach 1, \((16)\) reduces to the usual condition for Shannon’s maximizer. This, in turn, ensures that in the \((q, r) \rightarrow (1, 1)\) limit the maximizer of \((15)\) is Gibbs’s canonical distribution.

4.1. The \( r = q \) case

In case when we chose \( r = q \) (i.e., when the non-linear moment constraints are implemented via zooming distribution) it follows from \((16)\) that

\[
\Phi(p_i)^{1-q} \sum_k (p_k)^q = e^{(q-1)\langle \ln \mathcal{P} \rangle_q} \left( q \left( \langle \ln \mathcal{P} \rangle_q - \ln p_i \right) - 1 \right) - q \Omega \langle E_i \rangle - \langle E \rangle_q.
\]  

(17)

Multiplying both sides of \((17)\) by \( q_k(q) \), summing over \( i \) and taking the normalization condition \( \sum_k p_k = 1 \) we obtain

\[
\Phi = -e^{(q-1)\langle \ln \mathcal{P} \rangle_q} \Rightarrow \frac{\ln(-\Phi)}{q - 1} = \langle \ln \mathcal{P} \rangle_q \Rightarrow D_q(\mathcal{P}) \big|_{\max} = \frac{1}{q - 1} (\Phi + 1).
\]

(18)

Footnote: For Rényi’s entropy this fact was already observed in \([13]\).
Plugging result (15) back into (17) we obtain after some algebra

\[ \sum_k (p_k)^q = (p_\ast)^{q-1} \left[ q \ln p_i + \left( 1 - \frac{q \ln(-\Phi)}{q - 1} - \frac{q \Omega}{\Phi} (E_i - \langle E \rangle_q) \right) \right], \]

which must be true for any index \( i \). On the substitution

\[ \mathcal{E}_i = 1 - \frac{q \ln(-\Phi)}{q - 1} - \frac{q \Omega}{\Phi} (E_i - \langle E \rangle_q) , \]

this leads to the equation

\[ \kappa(p_\ast)^{1-q} = q \ln p_i + \mathcal{E}_i . \]

Here we have denoted \( \sum_k (p_k)^q \equiv \kappa \). Equation (21) has the solution

\[ p_i = \frac{q}{\kappa(q-1)} W \left( \frac{\kappa(q-1)}{q} e^{(q-1)\mathcal{E}_i/q} \right)^{1/(1-q)} = e^{\frac{\kappa(q-1)}{q} (q-1)^{-1} \mathcal{E}_i/q - \mathcal{E}_i/q} , \]

with \( W(x) \) being the Lambert–W function [47].

A couple of comments are now in order. First, \( p_i \)'s as prescribed by (22) are positive for any value of \( q > 0 \). This is a straightforward consequence of the following two identities [47]:

\[ W(x) = \sum_{n=1}^{\infty} \frac{x^n}{(n-1)!} \ln n^{-1} n^{n-2} , \]

\[ W(x) = x e^{-W(x)} . \]

Indeed, Eq. (22) ensures that for \( x < 0 \) also \( W(x) < 0 \) and hence \( W(x)/x > 0 \). Thus for \( 0 < q < 1 \) the positivity of \( p_i \)'s is proven. Positivity for \( q \geq 1 \) follows directly from the relation (24). Second, as \( q \to 1 \) the entropy \( D_q \to \mathcal{H} \) and we expect that \( p_i \)'s defined by (22) should approach the Gibbs canonical distribution in the \( q \to 1 \) limit. To see that this is indeed the case, let us note that

\[ \Phi|_{q=1} = -1, \quad \mathcal{E}|_{q=1} = 1 + \mathcal{H} + \Omega(\mathcal{E}_i - \langle E \rangle) , \quad \text{and} \quad \kappa|_{q=1} = 1 . \]

Then

\[ p_i|_{q=1} = e^{1-(1+\mathcal{H}+\Omega(\mathcal{E}_i - \langle E \rangle))} = e^\Omega(\mathcal{E}_i - \langle E \rangle) = e^\Omega / |Z| , \]

which after identification \( \Omega|_{q=1} = \beta \) leads to the desired result. Note also that (22) is invariant under uniform translation of the energy spectrum, i.e., the corresponding \( p_i \) is independent on the choice of the energy origin.

Third, there are situations, when Eq. (21) has no solution, or it gives solution for \( p_i \not\in [0, 1] \). To see this note, that when \( q > 1 \), the left-hand side of (21) is greater than \( \kappa \), from which follows that \( \mathcal{E}_i \geq \kappa \) for all \( i \)'s. For \( q < 1 \) the left-hand side of (21) acquires values from \([0, \kappa]\) which leads again to the condition \( \mathcal{E}_i \geq \kappa \). In both cases are therefore \( \mathcal{E}_i \) positive. Thus, for energies, for which \( \Delta_i E_i = E_i - \langle E \rangle_q \) is too negative, Eq. (21) has no solution, and the corresponding occupation probability is zero. Contrary to MaxEnt distributions of other commonly used entropies, there exist energy levels here, for which are MaxEnt distributions of \( D_q \) has zero occupation probabilities. This might provide a natural conceptual playground for statistical systems with energy gaps (e.g., disordered systems, carbon nanotubes) or for system with various super-selection rules (e.g., first-quantized relativistic systems).

Finally, there is no direct way to any simple method for a unique determination of \( \Phi \) and \( \Omega \) from the constraint conditions. In fact, only asymptotic situations for large and vanishingly small \( \Omega \) can be successfully resolved (this will be relegated to Sections 4.1.2 and 4.1.3). But very fortunately in the case of multifractals we can give to our relations (22) a very satisfactory physical interpretation, without resolving \( \langle E \rangle_q \) in terms of \( \Phi \) and \( \Omega \); cf. the discussion in the following subsection.
4.1.1. Multifractal case

Assuming the multifractality of the system (described in Appendix C), we can rewrite Eq. (21) into the form

$$e^{(q-1)/\Phi} \sim 1 + q \left[ a_i - \langle a \rangle_q(\epsilon) \right] \left[ 1 - \frac{\Omega}{\Phi} \right] \ln \epsilon. \quad (27)$$

where the exponent $\langle a \rangle_q(\epsilon)$ is the scaling exponent of $q$-mean of log-PDF:

$$\langle \ln P \rangle_q = \frac{\sum_i p_i(q) \ln p_i}{\sum_i p_i(q)} = \frac{\sum_i p_i(q)a_i \ln \epsilon}{\sum_i p_i(q)} = \langle a \rangle_q(\epsilon) \ln \epsilon. \quad (28)$$

We remind that in thermodynamic limit, one needs to interpret $E_i$ as $E_i \propto \log p_i$.

Equation (27) has a few implications. Firstly, from the polynomial coefficients, we get a version of a curdling theorem, where in the limit $\epsilon \to 0$, i.e., thermodynamic limit, is the only scaling exponent dominant, whereas the other are suppressed. In this case we have that

$$a_i = \frac{\tau(q)}{q-1} = D_q, \quad (29)$$

which is property inherited from Rényi's entropy. Indeed, in thermodynamic limit is possible to observe only a part of the whole multifractal system the scales with the scaling exponent $\langle a \rangle_q$ and the choice of $q$ influences on which part are focused.

Secondly, the speed of convergence is given by the fact that $\Delta a(q, \epsilon) = a_i(\epsilon) - \langle a \rangle_q(\epsilon)$ has to converge at least as $\frac{1}{\ln \epsilon}$. Finally, because of the dimensional analysis, $\Omega$ has to scale similarly to $\Phi$, so

$$\Omega_{eq} \sim e^{(q-1)/\langle a \rangle_q(\epsilon)}. \quad (30)$$

4.1.2. “High-temperature” expansion

Let us now make an important remark concerning the asymptotic behavior of $p_i$ in regard to $\Omega$. If we assume that the Lagrange multiplier $\Omega \ll 1$ then from (24) the following expansion holds

$$W\left(-\frac{\kappa(q-1)}{\Phi q} \exp\left(\frac{q-1}{q}\right) \exp\left(-\left(q - 1\right) \frac{\Omega}{\Phi} \Delta_q E_i\right)\right) \approx W(x) \left[ 1 - (1 - q) \Omega^* \Delta_q E_i \right], \quad (31)$$

with

$$\Delta_q E_i = E_i - \langle E_i \rangle_q, \quad \Omega^* = -\frac{\Omega}{\Phi(W(x) + 1)}, \quad x = -\frac{\kappa(q-1)}{\Phi q} \exp\left(\frac{q-1}{q}\right). \quad (32)$$

Hence, if we use the relation (22) we can write

$$p_i = \frac{\left[ 1 - (1 - q) \Omega^* \Delta_q E_i \right]^{1/(1-q)}}{\sum_k \left[ 1 - (1 - q) \Omega^* \Delta_q E_k \right]^{1/(1-q-1)}} \approx Z^{-1} \left[ 1 - (1 - q) \Omega^* \Delta_q E_i \right]^{1/(1-q)}, \quad (33)$$

with

$$Z = \sum_k \left[ 1 - (1 - q) \Omega^* \Delta_q E_k \right]^{1/(1-q)} = \left[ \frac{q}{\kappa(q-1)} W(x) \right]^{1/(1-q-1)}. \quad (34)$$

The distribution (33) agrees with the so-called 3rd version of thermostatistics introduced by Tsallis et al. [48]. It might by also formally identified with the maximizer for Rényi’s entropy [54]. Clearly, $\Omega^*$ is not a Lagrange multiplier, but $\Omega^*$ passes to $\beta$ at $q \to 1$ (in fact, $\Phi \to -1$, $\Omega \to \beta$ and $W(x) \to 0$ at $q \to 1$). Note also that when $\Omega = 0$ (i.e., no energy constraint) then $p_i = 1/n$ which reconfirms the fact that $D_q$ attains its largest value for the uniform distribution.
Figure 1. Two real branches of the Lambert–W function. Solid line: \( W_0(x) \equiv W(x) \) defined for \(-1/e \leq x < +\infty\) (so far considered). Dashed line: \( W_{-1}(x) \) defined for \(-1/e \leq x < 0\). The two branches meet at point \((-1/e, -1)\).

4.1.3. “Low-temperature” expansion

From the physical standpoint it is the asymptotic behavior at \( \Omega \gg 1 \) (or more precisely at \( \Omega |(q - 1)/\Phi| \gg 1 \)), i.e., “low-temperature” expansion, that is most intriguing. This is because the branching properties of the Lambert–W function at negative argument values make the structure of \( P \) rather non-trivial. We thus split our task into four distinct cases:

\[
\begin{align*}
&\text{a}_1 \quad (q - 1) > 0 \quad \text{and} \quad \Delta_q E < 0, \\
&\text{a}_2 \quad (q - 1) > 0 \quad \text{and} \quad \Delta_q E \geq 0, \\
&\text{b}_1 \quad (q - 1) < 0 \quad \text{and} \quad \Delta_q E < 0, \\
&\text{b}_2 \quad (q - 1) < 0 \quad \text{and} \quad \Delta_q E \geq 0.
\end{align*}
\]

Cases \( \text{a}_1 \) and \( \text{a}_2 \) are much simpler to start with as the argument of \( W \) is positive. \( W \) is then a real and single valued function which belongs to the principal branch of \( W_0 \), see Fig 1.

When \( \Delta_q E < 0 \) then \( \text{a}_1 \) implies the asymptotic expansion:

\[
W(z) \approx z \quad \Rightarrow \quad p_i = \left( \frac{1}{|\Phi|} \right)^{1/(1-q)} e^{\Delta_q E} \lesssim Z_1 e^{\frac{\Delta_q E}{|\Phi|^q}} \equiv Z_1^{-1} e^{\frac{\Omega |(q - 1)/\Phi|}{q \exp(q - 1)}}, \tag{35}
\]

with

\[
Z_1 = \left( \frac{1}{|\Phi|} \right)^{1/(1-q)} e^{1/q}.
\]

Note that in this case \( p_i \) is of a Boltzmann type (\( \langle E \rangle_q \) can be canceled against the same term in \( Z_1 \)).

On the other hand, \( \text{a}_2 \) situation implies the asymptotic expansion:

\[
W(z) \approx \log(z) - \log(\log(z)) \quad \Rightarrow \quad p_i = Z_2^{-1} [1 - (1 - q) \Omega^* \Delta_q E]^{1/(1-q)} \tag{36},
\]

with

\[
Z_2 = \left( \frac{q}{k(q - 1)} \right) \log \left( \frac{\kappa(q - 1)}{|\Phi|^q} \exp \left( \frac{q - 1}{q} \right) \right)^{1/(1-q)} ; \quad \Omega^* = \frac{\Omega}{|\Phi| \log(2q - 1) \exp(q - 1)}.
\]

Although the distribution \( \text{a}_2 \) formally agrees with Tsallis et al. distribution it cannot be identified with it as \( \Omega^* \) does not tend to \( \beta \) in \( q \to 1 \) limit. In fact, the limit \( q \to 1 \) is prohibited in this case as it violates the “low-temperature” condition \( \Omega |(q - 1)/\Phi| \gg 1 \). Note particularly that our MaxEnt distribution represents in the “low-temperature” regime
a heavy tailed distribution with Boltzmannian outset. When Ω and q > 1 are fixed one may find κ and Φ from the normalization condition

\[ Z_1^{-1} \sum_{k : \Delta_q E_k < 0} \exp\left(-\frac{\Omega}{|\Phi|} \Delta_q E_k\right) + Z_2^{-1} \sum_{k : \Delta_q E_k \geq 0} \left[1 - (1 - q)\Omega' \Delta_q E_k\right]^{1/(1-q)} = 1, \]  

and sewing condition at Δ_qE = 0. However, because the “low-temperature” approximation does not allow to probe regions with small Δ_qE one must numerically optimize the sewing by interpolating the forbidden parts of Δ_qE axis. Example of such a numerical optimization is presented in Fig. 2.

Cases (b1) and (b2) are a little bit more technically difficult, because (q − 1) < 0 causes the negativity of the argument of W(·). In case (b1) we obtain that for low temperatures \( \frac{\Delta_q E_k}{q} \to -\infty \). Nevertheless, the complex Lambert–W function has a branch cut in the interval \([-\infty, -\frac{1}{e}]\), so the real-valued Lambert–W function is defined only for \( x > -\frac{1}{e} \) and Eq. (24) has no real solution. This situation corresponds previous discussions about existence of solution of Eq. (21).

In case (b2) there exist two solutions of Eq. (24), i.e. \( W_0(z) \) and \( W_{-1}(z) \) (see Fig. 1). In case of principal branch \( W_0(z) \), the Lambert W-function can be approximated as \( W(z) \approx z \), and the solution corresponds to the case (a1). In case of principal branch \( W_{-1}(z) \), the asymptotic expansion for \( z \to 0^{-} \) is

\[ W_{-1}(z) \approx \log(-z) - \log(-\log(-z)) \]  

(38)

so the resulting probability is similar to the case (a2), only with \( \Omega^* = \frac{\Omega}{|\Phi| \log\left(\frac{\theta}{e^{\Phi}}\right)} \).

We should stress that for all cases it is necessary to check the validity of the asymptotic expansion and its applicability to the MaxEnt distribution. In some cases can the expansion violate the condition \( p_i \leq 1 \) and then it is not possible to use such approximations.

### 4.2. The r = 1 case

When \( r = 1 \) is chosen (i.e., when the constraints are implemented via the usual linear averaging) then Eq. (16) implies

\[ \Phi = e^{(q-1)\ln P_0} \frac{q\left(\ln P_0 - \ln p_i\right) - 1}{\sum_k (p_k)_{q-1}} - \Omega(E_i - \langle E \rangle). \]  

(39)

Multiplying by \( p_i \) and summing over \( i \) we obtain the constraint

\[ \Phi = -e^{(q-1)\ln P_0} \frac{\ln(-\Phi)}{-1} = \ln\langle P \rangle. \]  

(40)
Upon insertion of (40) into (39) we get a transcendental equation for $p_i$, which reads

$$\kappa p_i^{1-q} = \frac{\Phi}{\Phi + \Omega(E_i - \langle E \rangle)} \left[ q \ln p_i - \frac{q \ln(-\Phi)}{q - 1} + 1 \right]. \quad (41)$$

The solution can be again written in terms of the Lambert W–function, namely

$$p_i = \left[ \frac{q \Phi}{(q - 1) \kappa (\Phi + \Omega \Delta E_i)} W\left( -\frac{\kappa (q - 1)}{\Phi q} \exp\left( \frac{q - 1}{q} \left( 1 + \frac{\Omega}{\Phi} \Delta E_i \right) \right) \right) \right]^{1/(1-q)}$$

$$= \frac{1}{(\Phi)^{1/(1-q)}} \exp \left\{ \frac{1}{(q - 1)} \left[ \frac{\kappa (q - 1)}{\Phi q} \exp\left( \frac{q - 1}{q} \left( 1 + \frac{\Omega}{\Phi} \Delta E_i \right) \right) \right] \right\}.$$ \quad (42)

Relations (23) and (24) again ensure that all $p_i$’s are positive. In addition, it is easy to check that in the limit case $q \to 1$, the formula (42) approaches the classical Gibbsian maximizer. Indeed, if we utilize the identities:

$$\kappa_{q=1} = 1, \; \Phi_{q=1} = -1, \; \left[-\Phi\right]_{q=1}^{1/(1-q)} = e^{-\beta}, \; \text{and} \; \Omega_{q=1} = \beta,$$

then

$$p_{i|q=1} = e^{-\beta \Omega(\beta - E_i)} = e^{\beta E_i / Z}.$$ \quad (44)

Similarly as Eq. (26) also the relation (23) represents an important consistency check of our procedure.

### 4.2.1. Multifractal case

Similarly to the case $r = q$, we plug multifractal relations to Eq. (40) and obtain

$$e^\tau \sim -q e^{\alpha_{q-1} - 1} \ln e (a_i - \langle a \rangle_\alpha) + e^{\alpha_{q-1}} + \frac{\Omega e^\tau}{\Phi} (a_i - \langle a \rangle_\tau) \ln e$$

$$e^\tau - e^{\alpha_{q-1}} - \left[ \Delta \alpha_i(q, \epsilon) q e^{\alpha_{q-1}} + \frac{\Omega e^\tau}{\Phi} \Delta \alpha_i(1, \epsilon) \right] \ln e.$$ \quad (46)

Again, the analog of the curdling theorem can be extracted from the left-hand side. The right-hand side of the relation gives again the convergence to the prevailing exponent. We get two interesting results. Firstly $\Delta a_i(q, \epsilon)$ and $\Delta a_i(1, \epsilon)$ have to tend to zero at least as $e^{-\epsilon q}/\ln e$, and

$$\Omega_{i=1} \sim q \Phi \frac{\Delta \alpha_i(q, \epsilon)}{\Delta \alpha_i(1, \epsilon)} = q \Phi = q \Omega_{i=q}.$$ \quad (47)

### 4.2.2. “High-temperature” expansion

Similarly as in the $r = q$ case we can find the “high-temperature” expansion by assuming that $\Omega \ll 1$. In such a case we have

$$\frac{\Phi}{(\Phi + \Omega \Delta E_i)} W\left( -\frac{\kappa (q - 1)}{\Phi q} \exp\left( \frac{q - 1}{q} \left( 1 + \frac{\Omega}{\Phi} \Delta E_i \right) \right) \right) \approx W(x) \left[ 1 - (1 - q) \Omega^* \Delta E_i \right]. \quad (48)$$

Here

$$\Omega^* = -\frac{q}{q - 1} \frac{\Omega}{\Phi} W(x), \; \quad x = -\frac{q (q - 1)}{\Phi q} \exp\left( \frac{q - 1}{q} \right).$$ \quad (49)

Through (42) this implies that

$$p_i = Z^{-1} \left[ 1 - (1 - q) \Omega^* \Delta E_i \right]^{1/(1-q)},$$ \quad (50)

with

$$Z = \sum_k \left[ 1 - (1 - q) \Omega^* \Delta E_k \right]^{1/(1-q)} = \left[ \frac{q}{k (q - 1)} W(x)\right]^{1/(q - 1)}.$$ \quad (51)

Note that for $q \to 1$ the factor $\Omega^*$ approaches the inverse temperature $\beta$. Central to this analysis is the observation that $\lim_{q \to 1} W(x)/(q - 1) = 1$. 

---
4.2.3. “Low-temperature” expansion

We now wish to consider the “low-temperature” expansion — i.e., $\Omega \gg 1$. Similarly as in the $r = q$ case, we divide the situation into four sub-cases:

\[ a_1) \ (q - 1) > 0 \ \text{and} \ \Delta E < 0, \]
\[ a_2) \ (q - 1) > 0 \ \text{and} \ \Delta E \geq 0, \]
\[ b_1) \ (q - 1) < 0 \ \text{and} \ \Delta E < 0, \]
\[ b_2) \ (q - 1) < 0 \ \text{and} \ \Delta E \geq 0. \]

Unlike case $r = q$, the sub-cases group into two qualitatively distinct classes:

1. cases $a_2)$ and $b_1)$ lead to the asymptotic expansion $W(z) \propto \log(z) - \log(\log(z))$, because $-\frac{\log(1)\Delta E}{\phi_q} > 0$.

2. cases $a_1)$ and $b_2)$ lead to the situation, when the Lambert $W$-function is not defined, which corresponds to the fact that Eq. (11) has no solution.

So in particular, we see that in cases when $q < 1$ our hybrid entropy cannot be consistently used over the whole temperature range. It can be at best used as an effective entropy in higher-temperature regimes. This might be particularly pertinent in the high-energy particle phenomenology where the host of phase transitions is happening under conditions that are far from thermal equilibrium (e.g., chiral phase transition in QCD and ensuing quark-gluon plasma formation). In the case when $q > 1$, the asymptotic expansion can be written in the form

\[ p_i = \left\{ \frac{e^{(q-1)}_q}{q} (1 + \Omega \Delta E_i) \right\} \frac{1}{\ln\left[ \frac{e^{(q-1)}_q \exp\left( \frac{q-1}{q} \right)}{\ln[1 + \Omega \Delta E_i]} \right]}^{q-1}. \]  \hspace{1cm} (52)

Contrary to $r = q$, the resulting distribution has functionally different form from both the Boltzmann distribution and Tsallis distribution, even in the generalized form, i.e. with the self-referential temperature. For large temperatures, the log-term is negligible and the distribution becomes similar power-like behavior. We shall again note that it is necessary to check consistency of asymptotic expansions.

4.3. Concavity issue

We now turn to the question of whether or not $D_{q_{\lambda_{\max}}}$ is a concave function of $\langle E \rangle_r$. Note that in contrast to the information-theoretic entropy $D_q$, $D_{q_{\lambda_{\max}}}$ is the system entropy, i.e., it depends on the actual system state variables. We wish to show now that $D_{q_{\lambda_{\max}}}$ is a concave function of the constraints as in the case of Gibbsian MaxEnt. In other words we wish to show that

\[ D_q(\lambda(\langle E^{(1)-gallery} \rangle_r + (1 - \lambda)(\langle E^{(2)-gallery} \rangle_r))_{\max} \geq \lambda D_q(\langle E^{(1)-gallery} \rangle_{\max}) + (1 - \lambda) D_q(\langle E^{(2)-gallery} \rangle_{\max})_{\max}. \]  \hspace{1cm} (53)

for any $\lambda \in [0, 1]$ and $r = 1$ or $r = q$.

For this purpose we first show that for $q \leq 1$ the entropy $D_q(P)$ is a concave function of the probability distribution. Using the $q$-logarithm we can write

\[ D_q(P) = \ln_{(q)} e^{-\langle \ln P \rangle_q}. \]  \hspace{1cm} (54)

Because the $q$-logarithm is a concave function for any $q \geq 0$ so is $\ln_{(q)} e^x$ for $x \in \mathbb{R}^+$. In addition,

\[ -x^q \ln x \]  \hspace{1cm} (55)

is a convex function whenever $q$ and hence

\[ -\langle \ln P \rangle_q = -\sum_k q_k \ln p_k \quad q \leq 1, \]  \hspace{1cm} (56)

is a concave function of $P$.
5. Conclusions and outlooks

We have presented a plausible generalization of the information entropy concept. Our approach is based on an axiomatic merger of two currently widely used information measures: Rényi’s and Tsallis–Havrda–Charvát’s. Such a merger is natural from the mathematical point of view as both above measures have an axiomatic underpinning with a very similar axiomatics. From the physics viewpoint the above merger is interesting because it combines two entropies with analogous MaxEnt distributions but with very different scope of applicability in physics.

We have shown that the maximizers for \( D_q \) subject to constant averaged energy are represented in terms of the Lambert W–function. The Lambert W–function is a special function that appears in numerous exactly solvable statistical systems. Tonks gas [49], Richards growth model and Lotka–Volterra models [50] may serve as examples. The Lambert W–function was recently also used in quantum statistics [51] and statistics of weak long-range repulsive potentials [49]. This usage nicely bolsters our suggestion that a typical playground for statistical systems, Tonks gas [49], Richards growth model and Lotka–Volterra models [50] may serve as examples.

Due to complicated analytical structure of the Lambert W–function we have resorted in our discussion to the “low” and “high temperature” asymptotic regimes. We have shown that under certain parameter conditions these have the heavy tailed behavior that is identical with Tsallissian maximizers. The fact that this is true only asymptotically might be at first sight a bit surprising, as there exists perception that both THC and Rényi’s entropies have the same maximizer and hence the merger entropy should again possess the same MaxEnt distribution. This anticipation is clearly erroneous. Indeed, both Rényi entropy and THC maximizers have the same functional form but their respective “temperature” parameters \( \beta^* \) are entirely different functions of \( q \), and in the case of THC entropy \( \beta^* \) is even self-referential (i.e., it depends on the distribution itself) [54].

In summary, we have proved that there exists a well defined sense in which one can combine Rényi and THC entropic paradigms. We have found the associated entropy measure, namely (9) and the ensuing MaxEnt distribution (22). It can be rightly objected that apart from the axiomatic side more is needed to consider a merger is natural from the mathematical point of view as both above measures have an axiomatic underpinning with a very similar axiomatics. From the physics viewpoint the above merger is interesting because it combines two entropies with analogous MaxEnt distributions but with very different scope of applicability in physics.

6. Acknowledgements

We acknowledge very helpful discussions with P.T. Landsberg, P. Haramoës and H. Lavička which have helped us to understand better the ideas discussed in this paper. This work was supported by the Grant Agency of the Czech Republic, Grant No. GCP402/12/J077.

Appendix A. Derivation of \( D_q \) from J-A axioms

In this appendix we show the basic steps in the derivation of functional form of hybrid entropy \( D_q \).

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

\[
D(A^{(k)}) = D(1/r, \ldots, 1/r) = L(r), \quad (1 \leq k \leq m). \tag{A.1}
\]

Repeated application of axiom 3 then leads to

\[
D(A^{(1)} \cap A^{(2)} \cap \ldots \cap A^{(m)}) = L(r^m) = \sum_{k=1}^{m} m C_k (1-q)^{k-1} D(A^{(k)}) \]

\[
= \frac{1}{(1-q)} [(1 + (1-q)L(r))^m - 1]. \tag{A.2}
\]
\(^{m} C\_k\) is the binomial coefficient. Taking partial derivative of both sides of Eq. (A.2) with respect to \(m\) and setting \(m = 1\) afterwards we obtain the differential equation
\[
\frac{(1-q) d L}{(1+(1-q) L) [\ln(1+(1-q) L)]} = \frac{dr}{r \ln r}.
\]
(A.3)

The general solution of Eq. (A.3) has the form
\[
L(r) \equiv L_q(r) = \frac{1}{1-q}(e^q - 1).
\]
(A.4)

Function \(c(q)\) will be determined shortly. Right now we just note that because at \(q = 1\) Eq. (A.2) boils down to \(L(r^m) = mL(r)\) we must have \(c(1) = 0\). In addition, the monotonicity of \(L(r)\) ensures that \(c(q)/(1-q) \geq 0\). To proceed further let us consider the experiment with outcomes \(A = (A_1, A_2, \ldots, A_n)\) and the distribution \(P = \{p_1, p_2, \ldots, p_n\}\). Assume moreover that \(p_k\) (\(1 \leq k \leq n\)) are rational numbers, i.e.,
\[
p_k = \frac{g_k}{g}, \quad \sum_{k=1}^{n} g_k = g, \quad g_k \in \mathbb{Z}^+.
\]
(A.5)

Let have, in addition, an experiment \(B = (B_1, B_2, \ldots, B_k)\) with associated distribution \(Q = \{q_1, q_2, \ldots, q_k\}\). We split \((B_1, B_2, \ldots, B_k)\) into \(n\) groups containing \(g_1, g_2, \ldots, g_n\) outcomes respectively. Consider now a particular situation in which whenever event \(A_i\) in \(A\) happens then in \(B\) all \(g_k\) events of \(k\)-th group occur with the equal probability \(1/g_k\) and all the other events in \(B\) have probability zero. Hence
\[
\mathcal{D}(B|A = A_k) = \mathcal{D}(1/g_k, \ldots, 1/g_k) = L_q(g_k).
\]
(A.6)

and so axiom 3 implies that
\[
\mathcal{D}(B|A) = f^{-1} \left( \sum_{k=1}^{n} q_k(g) f(L_q(g_k)) \right).
\]
(A.7)

On the other hand, in the stated system the entropy \(\mathcal{D}(A \cap B)\) can be easily evaluated. Realizing that the joint probability distribution corresponding to \(A \cap B\) is
\[
\mathcal{R} = \{r_{kl} = p_k q_{|l}\} = \begin{pmatrix} \frac{P_1}{g_1} & \cdots & \frac{P_1}{g_1} \\ \frac{P_2}{g_2} & \cdots & \frac{P_2}{g_2} \\ \vdots & \ddots & \vdots \\ \frac{P_n}{g_n} & \cdots & \frac{P_n}{g_n} \end{pmatrix} = \{1/g, \ldots, 1/g\}.
\]
(A.8)

we obtain that \(\mathcal{D}(A \cap B) = L_q(g)\). Applying axiom 3 together with Eq. (A.2) we get
\[
\mathcal{D}(A) \left( 1+ (1-q) f^{-1} \left( \sum_k q_k(g) f(L_q(p_k)[1+(1-q)L_q(g)] + L_q(g)) \right) \right)
= L_q(g) - f^{-1} \left( \sum_k q_k(g) f(L_q(p_k)[1+(1-q)L_q(g)] + L_q(g)) \right).
\]
(A.9)

Define \(f\_a(x) = f(-ax + y) \Rightarrow f^{-1}(x) - y = -af\_a^{-1}(x)\) then
\[
\mathcal{D}(A) = \frac{f\_a^{-1}(L_q(g)) \left( \sum_k q_k(g) f(a L_q(p_k)) \right)}{1 - (1-q)f\_a^{-1}(L_q(g)) \left( \sum_k q_k(g) f(a L_q(p_k)) \right)}.
\]
(A.10)

with \(a = [1 + (1-q)L_q(g)]\).
To proceed further, let us formally put \( p_k = 1/r \). Eq. (A.4) then indicates that it is \( L_q(1/p_k) \) and not \(-L_q(p_k)\) which represents the elementary information of order \( q \) affiliated with \( p_k \) (cf. with Eq. (5)). It is thus convenient to reformulate (A.10) directly in terms of \( L_q(1/p_k) \). This can be done via relation

\[
L_q(p_k) = \frac{L_q(1/p_k)}{1 + (1 - q)L_q(1/p_k)},
\]

(A.11)

If we now write

\[
g(x) = f_{x,L_q}(\frac{x}{1 + (1 - q)x}),
\]

(A.12)

we easily obtain from (A.10) that

\[
\mathcal{D}(A) = g^{-1} \left( \sum_k q_k(g(L_q(1/p_k))) \right).
\]

(A.13)

Moreover, if we set in the second part of axiom 3, \( B = A \) then \( \mathcal{D}(A) \) is given as

\[
\mathcal{D}(A) = f^{-1} \left( \sum_k q_k(f(L_q(1/p_k))) \right).
\]

(A.14)

Using the fact that two quasi-linear means with the same \( \mathcal{P} \) are identical iff their respective Kolmogorov–Nagumo functions are linearly related [4], we may write

\[
g(x) = f \left( \frac{-x + y}{1 + (1 - q)x} \right) = \theta_q(y)f(x) + \theta_q(y).
\]

(A.15)

Here \( y = L_q(g) \). In order to solve (A.15) we define \( \varphi(x) = f(x) - f(0) \). With this notation Eq. (A.15) turns into

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

(A.16)

By setting \( x = y \) we obtain that \( \theta_q(y) = -1 \), and hence

\[
\varphi(x + y + (1 - q)xy) = \varphi(x) + \varphi(y).
\]

(A.17)

According to axiom 1 we may now extend (A.17) to real valued \( x \) and \( y \). Eq. (A.17) is Pixder’s functional equation which can be solved by the standard method of iterations [43]. In [64] is show that (A.17) has only one non-trivial class of solutions, namely

\[
\varphi(x) = \frac{1}{\alpha} \ln \left[ 1 + (1 - q)x \right].
\]

(A.18)

\( \alpha \) is here a parameter. Inserting this solution back to (A.14) gives

\[
\mathcal{D}_q(A) = \frac{1}{1 - q} \left( e^{-c(q)} \sum_k \varphi(k) \ln p_k - 1 \right) = \frac{1}{1 - q} \left( \prod_k (p_k)^{c(q)\varphi(k)} - 1 \right).
\]

(A.19)

Note that the constant \( \alpha \) got canceled. We have also denoted the explicit order of the entropy \( \mathcal{D} \) with the subscript \( q \). It remains to determine \( c(q) \). Utilizing the conditional entropy constructed from (A.19) and using axiom 3, we obtain \( c(q) = 1 - q \). In result we can recast (A.19) into more expedient form, by application of relation

\[
\langle \ln \mathcal{P} \rangle_q = (1 - q) \left( \frac{dI_q(\mathcal{P})}{dq} - I_q(\mathcal{P}) \right)
\]

(A.20)

to the following form:

\[
\mathcal{D}_q(A) = \frac{1}{1 - q} \left( e^{-(1-q)^2I_q/A} \sum_{k=1}^n (p_k)^q - 1 \right).
\]

(A.21)
Appendix B. Basic properties of $D_q$ entropy

In this appendix, we present some basic properties of the hybrid entropy.

Let us start with features that are common to both Rényi’s and THC entropies we find

(a) $D_q(P = \{1, 0, \ldots, 0\}) = 0$

(b) $D_q(P) \geq 0$

(c) $D_1 = I_1 = S_1 = \mathcal{H}$

(d) $D_q$ involves a single free parameter - $q$

(e) $D_q$ is symmetric, i.e., $D_q(p_1, \ldots, p_n) = D_q(p_{k(1)}, \ldots, p_{k(n)})$

(f) $D_q$ is bounded

On the other hand, among features inherited from Rényi’s entropy we can find that

(g) $D_q(A) = f^{-1} \left( \sum_k \varrho_k(q) f(D_q(A_k)) \right)$

(h) For single-dimensional statistical systems with continuous PDF $D_q(A)$ reduces to $\mathcal{H}$

(i) $D_q$ is a strictly decreasing function of $q$, i.e., $dD_q/dq \leq 0$, for any $q > 0$

Result (i) follows from the fact that $D_q$ is a monotonically decreasing function of $A_q \equiv \sum_k \varrho_k(q) \ln p_k$ (see Eq. (A.19)) and that $A_q$ is a monotonically increasing function of $q$, indeed

$$\frac{dA_q}{dq} = \langle (\ln(P))^2 \rangle_q - \langle \ln(P) \rangle_q^2 \geq 0.$$ (B.1)

Here $\langle \ldots \rangle_q$ is defined with respect to the distribution $\varrho_k(q)$. The last relation in (B.1) is Jensen’s inequality. Note that $dD_q/dq = 0$ happens only for the degenerate case $P = \{1, \ldots, 0\}$.

Finally, properties imprinted from THC entropy include

(j) $\max_P D_q(P) = D_q(P = \{1/n, \ldots, 1/n\}) = \ln n$

(k) $D_q$ is $q$ non–extensive, i.e., $\mathcal{D}(A \cap B) = \mathcal{D}(A) + \mathcal{D}(B|A) + (1 - q)\mathcal{D}(A)\mathcal{D}(B|A)$

Appendix C. Brief revision of multifractal formalism

We present here some essentials of the fractal and multifractal analysis that are needed in the main body of the text.

Fractals are sets with a generally non–integer dimension exhibiting property of self–similarity. The key characteristic of fractals is fractal dimension which is defined as follows: Consider a set $M$ embedded in a $d$–dimensional space. Let us cover the set with a mesh of $d$–dimensional cubes of size $\varepsilon^d$ and let $N_\varepsilon(M)$ is a number of the cubes needed for the covering. The fractal dimension of $M$ is then defined as

$$D = -\lim_{\varepsilon \to 0} \frac{\ln N_\varepsilon(M)}{\ln \varepsilon}.$$ (C.1)

In most cases of interest the fractal dimension (C.1) coincides with the Hausdorff–Besicovich dimension used by Mandelbrot [53].
Multifractals, on the other hand, are related to the study of a distribution of physical or other quantities on a generic support (be it or not fractal) and thus provide a move from the geometry of sets as such to geometric properties of distributions. Let us suppose that over some support (usually a subset of a metric space) is distributed a probability of certain phenomenon. If we pave the support with a grid of spacing $\varepsilon$ and denote the integrated probability in the $i$th box as $p_i$, then the scaling exponent $a_i$ is defined [55, 56]

$$p_i(\varepsilon) \sim \varepsilon^{a_i}. \quad (C.2)$$

The exponent $a_i$ is called singularity or Lipshitz–Hölder exponent. Counting boxes $N(a)$ where $p_i$ has $a_i \in (a, a + da)$, the singularity spectrum $f(a)$ is defined as [55, 56]

$$N(a) \sim \varepsilon^{-f(a)}. \quad (C.3)$$

Thus a multifractal is the ensemble of intertwined (uni)fractals each with its own fractal dimension $f(a_i)$. It is further convenient to define a “partition function” [55]

$$Z(q) = \sum_i (p_i(\varepsilon))^q = \int da' \rho(a') e^{-f(a') \varepsilon^q}, \quad (C.4)$$

$(\rho(a))$ is a proportionality function having its origin in relations $C.2$ and $C.3$). In the small $\varepsilon$ limit the method of steepest descent yields the scaling

$$Z(q) \sim \varepsilon^\tau(q), \quad (C.5)$$

with

$$\tau(q) = \min_a (qa - f(a)), \quad f'(a) = q \quad \text{and} \quad \tau'(a) = a(q). \quad (C.6)$$

These are precisely Legendre transform relations. Scaling exponent $\tau$ is often called the correlation exponent. Legendre transform $C.6$ ensures that pairs $(f(a), a)$ and $(\tau(q), q)$, are conjugates comprising the same mathematical content.

Multifractal formalism has direct applications in the turbulent flow of fluids [57], percolation [58], diffusion–limited aggregation (DLA) systems [59], DNA sequences [60], finance [61], and string theory [62]. In chaotic dynamical systems all $I_q$ are necessary to describe uniquely, e.g., strange attractors [44]. More generally, one may argue [34] that when the outcome space is discrete then all $I_q$ (or $S_q$) with $q \in [1, \infty]$ are needed to reconstruct the underlying distribution, while when the outcome space is $d$–dimensional subset of $\mathbb{R}^d$ then all $I_q$ (or $S_q$), $q \in (0, \infty)$, are required to pinpoint uniquely the underlying PDF. The latter is nothing but the information–theoretic variants of Hausdorff’s moment problem of mathematical statistics. The connection of Rényi entropies with multifractals is established via relation $C.4$. Note particularly that when $\varepsilon$ is finite then $I_2$ plays the rôle of the Helmholtz free energy. *** Closer analysis of the related implications can be found in Refs. [34, 63].

References

[1] C.E. Shannon, *A mathematical theory of communication*, Bell Syst. Tech. J. 27, (1948), 379; 623.
[2] J.H. Havrda and F. Charvat, Kybernetika 3 (1967) 30.
[3] B.D. Sharma and D.P. Mittal, J. Math. Sci. 10 (1975) 28.
[4] A. Rényi, *Selected Papers of Alfred Rényi*, Vol.2 (Akadémia Kiado, Budapest, 1976).
[5] I.N.Kapur, The Mathematics Seminar 4 (1967) 78; Ind. Jour. Pure and App. Maths. 17 (1986) 429.
[6] M. Müller-Lennert, F. Dupuis, O. Szehr, S. Fehr and M. Tomamichel, J. Math. Phys. 54 (2013) 122203.
[7] P.J. Coles and M. Piani, Phys. Rev. A 89 (2014) 022112.
[8] K. Urbanik, Rep. Math. Phys. 4 (1973) 289.
[9] E.T. Jaynes, Phys. Rev. 106 (1957) 171; 108 (1957) 620.
[10] D.K.Faddeyev, Uspekhi Mat. Nauk, 11 (1956) 227.
[11] J.E. Shore and R.W. Johnson, IEEE Trans. Inform. Theory 26 (1980) 26.
[12] in, E.T. Jaynes, *Probability Theory, The Logic of Science* (Cambridge Un. Press, Cambridge, 2003).
[13] F. Topsoe, Kybernetica 15 (1979) 8; IEEE Trans. Inform. Theory 48 (2002) 2368.
[14] E.T. Jaynes, Revs. Modem. Phys. 27 (1955) 189.
[15] A.I. Khinchin, *Mathematical Foundations of Information Theory* (Dover Publications, Inc., New York, 1957).
[16] A. Kolmogorov, Atti della R. Accademia Nazionale dei Lincei 12 (1930) 388.
