Generalized entropies: what are they good for?

Jan Korbel
Faculty of Nuclear Sciences and Physical Engineering,
Czech Technical University in Prague, Břehová 7, 11519,
Prague, Czech Republic

26. January 2016

Motivation

A physicist, mathematician and programmer talk about entropy. Somebody asks:

- what is entropy?
  - measure of uncertainty, information, etc.
- how do you define entropy?
  - physicist: $S = \int \frac{\mathrm{d}Q}{T}$
  - mathematician: $S = -\sum_i p_i \log_2 p_i$
  - programmer: minimal length of message encoded in a binary code

different definitions and still one quantity

Introduction

- **entropy** is the central concept of thermodynamics, statistical mechanics and information theory
- there are close parallels between thermodynamical entropy and statistical entropy
- nevertheless, there are some differences and it is necessary to distinguish these concepts
- the aim of this talk is comparison of these concepts
- we discuss situations when a **generalized** form of entropy can be more convenient in description of a system

Historical overview

- **Thermodynamics**
  - 1803: Carnot - first formulation of 2nd law of TD
  - 1854: Clausius - equivalence-value (first formulation of TD entropy)
  - 1865: Clausius - first definition of entropy (he used term *entropy* for the first time)
- **Statistical mechanics**
  - 1877: Boltzman - $S = k_B \ln W$ (microcanonical)
  - 1880's: Gibbs - $S = -\sum_i p_i \ln p_i$
- **Information theory**
  - 1948: Shannon - mathematical theory of communication - Shannon entropy $S(P) = -\sum_i p_i \log_2 p_i$
  - 1957: Jaynes - MaxEnt principle (connection between information theory and statistical mechanics)
  - 1961: Rényi entropy
  - 1973: Havrda Charvát entropy, 1988 - Tsallis entropy

Thermodynamical entropy

- classical thermodynamics is based on **laws of thermodynamics**
- the first law of thermodynamics

$$\mathrm{d}U = \partial Q - \partial W$$

- differential form of heat is not closed, but there exists an integration factor, so

$$\partial Q = T\mathrm{d}S$$

- thermodynamical entropy is therefore defined as from the previous relation
- as a consequence, entropy is a state variable

Statistical mechanics entropy

- Classical thermodynamics works with macroscopic quantities
- Statistical mechanics works with microscopic quantities from which can be the macroscopic quantities derived
- The whole description is based on ensemble description and partition function
- Entropy can be determined as

$$
\begin{aligned}
S &= k_B \ln \Omega_{mic} \\
&= k_B(\ln Z_{can} + \beta < E >) \\
&= k_B(\ln \mathcal{Z}_{gc} + \beta(< E > - \mu < N >))
\end{aligned}
$$

Informational entropy

- information theory defines the entropy as a measure of information encoded in an arbitrary probability distribution
- entropy represents the amount of information necessary to fully determine a system
- exact form of entropy depends on the particular system
- the most common is the Shannon entropy (SE)

$$S(P) = \sum_i p_i \log_2 \left(\frac{1}{p_i}\right)$$

Important principles of entropy

To entropy are connected several important principles, among others:

- MaxEnt principle
- Legendre structure of TD
- Additivity
- Extensivity

we briefly discuss the importance of each principle

MaxEnt principle

- connection between thermodynamical entropy and informational entropy is given by Jaynes' **MaxEnt principle**
- thermodynamical entropy is obtained from the distribution which maximizes informational entropy under given constrains
- we define Lagrange function

$$L(P) = S(P) - \sum \lambda_j f_j(P) \tag{1}$$

  where S is the entropy and f_j are the constraints
- usually, we consider average energy constraint $< E >= \sum_i p_i E_i$ (canonical ensemble)
- corresponding Lagrange parameter represents the inverse temperature β
- resulting distribution maximizing entropy is the one which is realized in a thermodynamical system

Legendre structure of thermodynamics

- one important consequence of MaxEnt principle is the Legendre structure of thermodynamics
- for canonical ensemble represents Lagrange function the *free energy* of the system
- therefore we get the well-known relation

$$F(T,V) = U(T,V) - TS(T,V) \tag{2}$$

  not only for Shannon entropy, but also for other entropy functionals

Additivity

- additivity and extensivity are important properties of statistical systems
- nevertheless, they have slightly different meaning
- additivity: informational property - entropy of a compound system can be expressed as

$$S(A \cup B) = S(A) + S(B|A) \tag{3}$$

- it is possible to introduce a concept of *generalized additivity* so for independent A, B

$$S(A \cup B) = \Phi(S(A), S(B)) \tag{4}$$

EXTENSIVITY

- extensivity is, on the other hand, a thermodynamical property depending on a particular system
- entropy is extensive if for state space $W(N)$ is

$$\lim_{N \to \infty} \frac{S(W(N))}{N} = \omega \tag{5}$$

  the entropy scales as the number of particles so

$$S(2N) \simeq 2S(N)$$

- Shannon entropy is extensive for $W(N) \propto 2^N$ (no correlations)
- other entropies are extensive systems growing polynomially N^α (Tsallis) or subexponentially 2^{N^γ}

Axiomatic definition of entropy

- Entropy can be straightforwardly defined as a functional on a probability space
- Alternatively, more rigorous definition was done by A. Kolmogorov
- The axiomatic definition of the entropy enables us to understand main properties of entropy
- It is also a good starting point for possible generalizations

Axiomatic definition of Shannon entropy

Shannon entropy is defined by the following four axioms:

1. *continuity*: $\mathcal{H}(\mathcal{P})$ is continuous in every argument,
2. *maximality*: $\mathcal{H}(\mathcal{P})$ takes the maximal value for the uniform distribution,
3. *expansibility*: $\mathcal{H}(p_1, \dots, p_n, 0) = \mathcal{H}(p_1, \dots, p_n)$.
4. *additivity*: $\mathcal{H}(A \cup B) = \mathcal{H}(A) + \mathcal{H}(B|A)$, where $\mathcal{H}(B|A) = \sum_k p_k \mathcal{H}(B|A = a_k)$.

From these four axioms can be determined the form of Shannon entropy

$$\mathcal{H}(\mathcal{P}) = -\sum_k p_k \ln p_k \, .$$

Generalized entropies

- functional form of Shannon entropy corresponds to **Boltzman-Gibbs** entropy from statistical mechanics
- information theory provides many more information measures
- one way how to define these entropies is to generalize Shannon axiomatic
- these entropies could be understand as statistical mechanical entropies resulting into a different thermodynamics
- question at stake is if this approach is legitimate and if it brings any advantages
- "everybody can have his/her own entropy"

Rényi entropy

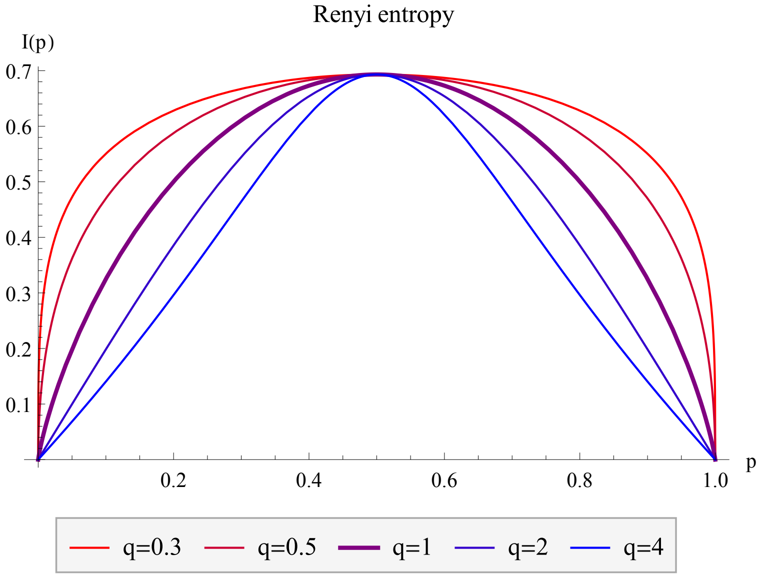
Entropy of multifractal systems

- Shannon entropy is not the only solution for additivity axiom $\mathcal{S}(A \cup B) = \mathcal{S}(A) + \mathcal{S}(B|A)$
- The defining axioms remain the same but the fourth is slightly changed:
  4. *Rényi additivity*: $\mathcal{I}_q(A \cup B) = \mathcal{I}_q(A) + \mathcal{I}_q(B|A)$, where $\mathcal{I}_q(B|A) = g^{-1}\left[\sum_k \rho_k(q) g[\mathcal{I}_q(B|A = a_k)]\right]$, $\rho_k(q) = p_k^q / \sum_j p_j^q$ and g is positive and invertible function on $[0, \infty)$.
- Similarly to Shannon entropy, it is possible to find an *operational* definition of RE (minimal cost)

This leads to the definition of RE in the form:

$$\mathcal{I}_q(\mathcal{P}) = \frac{1}{1-q} \ln \left(\sum_k p_k^q \right)$$

RÉNYI ENTROPY



Application of RE

Multifractal thermodynamics

- Rényi entropy has tight relation to theory of **multifractals**
- fractal systems: characteristic scaling relation and fractal dimension (Koch snowflake,etc.)
- multifractal systems: multiple local scaling exponents, probability distribution $p_i(s) \sim s^{\alpha_i}$ (s - scale)
- distribution of scaling exponents $p(\alpha, s) \sim s^{-f(\alpha)} \mathrm{d}\alpha$ - $f(\alpha)$: multifractal spectrum, fractal dimension of α-subset
- partition function $Z(q, s) = \sum_i p_i(s)^q \propto s^{\tau(q)} \sim s^{\alpha q - f(\alpha)}$
- the whole system can be described alternatively by multifractal spectrum $f(\alpha)$ or scaling function $\tau(q)$
- Rényi entropy $\mathcal{I}_q = \frac{1}{1-q} \ln \sum_i p_i^q \sim \frac{\tau(q)}{q-1} = D(q)$

Application of RE

Multifractal thermodynamics

- The connection to thermodynamics can be done via the partition function $\sum_i p_i^q = \sum_i \exp(-\beta E_i)$
- As a consequence $E_i = -\ln p_i$ and $\beta = q$
- Temperature is equal to parameter q which is called zooming parameter because for p_i^q accentuates different parts of PDF
- Relation between multifractal entropy $I_q = \frac{1}{1-q} \ln Z(q,s)$ and thermodynamical entropy $S_q = (\ln Z(q,s) + \beta < E >)$ is more complicated

Tsallis entropy

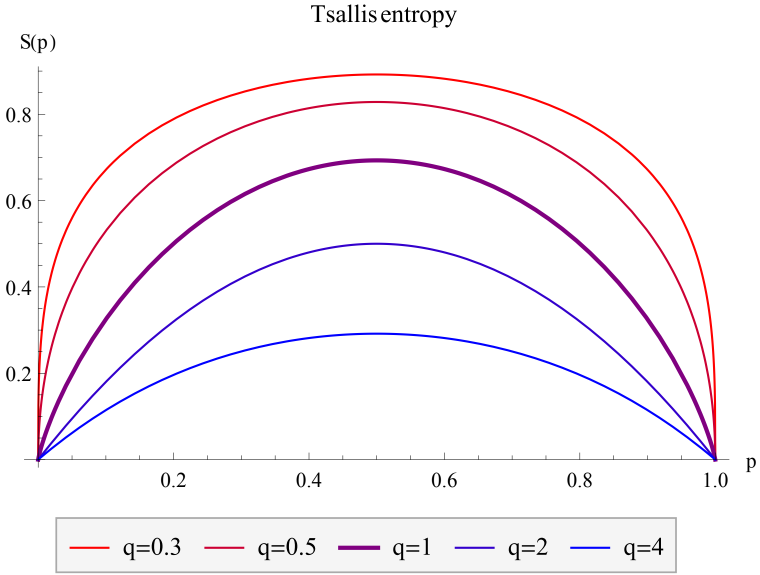
Entropy of non-extensive systems

- Tsallis entropy generalizes the concept of entropy to non-extensive systems
- it was firstly discovered in information theory by Czech mathematicians J. Havrda and F. Charvát
- later it was introduced to physics by C. Tsallis
- The additivity axiom is now changed as follows:
  4. *Tsallis additivity*:
     $\mathcal{S}_q(A \cup B) = \mathcal{S}_q(A) + \mathcal{S}_q(B|A) + (1-q)S_q(A)\mathcal{S}_q(B|A)$,
     where $\mathcal{S}_q(B|A) = \sum_k \rho_k(q)\mathcal{S}_q(B|A = a_k)$ and
     $\rho_k(q) = p_k^q / \sum_j p_j^q$

We can deduce that Tsallis entropy has form

$$\mathcal{S}_q(\mathcal{P}) = \frac{1}{1-q}\left(\sum_k p_k^q - 1\right)$$

TSALLIS ENTROPY



Application of Tsallis entropy

Thermodynamics with finite bath

- let us consider a system S coupled with a heat bath B
- in microcanonical ensemble is the partition function given by number of states $\Omega_{tot}(E) = \sum_s \Omega_S(E_s)\Omega_B(E - E_s)$
- for finite bath is $\Omega_B(E - E_s) \propto (E - E_s)^{n_s - 1}$
- probability distribution is therefore $p(E_s) = \frac{\Omega_B(E-E_s)}{\sum_k \Omega_B(E-E_k)} = \frac{1}{Z_q^{-1}}[1 - \beta(q-1)E_s]^{1/(q-1)}$
- equivalently, Tsallis q-gaussian distribution can be derived as a MaxEnt distribution of Tsallis entropy under constraints $\sum_i p_i = 1 \sum_j p_j E_j =< E >$
- There are problems with temperature definition (self-referential temperature - depending on distribution)

Hybrid entropy

Overlap between self–similarity and non–extensivity

- recently, there have been many attempts on further generalizations - one of them is called hybrid entropy
- hybrid entropy is based on both properties of Rényi entropy and Tsallis entropy
- resulting entropy can describe self–similar and non–extensive systems
- additivity axiom is now changed as
  4. *J.-A. additivity*:
     $\mathcal{D}_q(A \cup B) = \mathcal{D}_q(A) + \mathcal{D}_q(B|A) + (1-q)D_q(A)\mathcal{D}_q(B|A)$,
     where $\mathcal{D}_q(B|A) = f^{-1}\left[\sum_k \rho_k(q) f[\mathcal{D}_q(B|A = a_k)]\right]$,
     $\rho_k(q) = p_k^q / \sum_j p_j^q$ is escort distribution and f is a positive and invertible function on $[0, \infty)$.

These axioms define the hybrid entropy in form:

$$\mathcal{D}_q(\mathcal{P}) = \frac{1}{1-q}\left(e^{-(1-q)\sum_k \rho_k(q) \ln p_k} - 1\right) = \frac{1}{1-q}\left(e^{-(1-q)\langle \ln \mathcal{P} \rangle_q} - 1\right)$$

- we demand normalization condition $\sum_k p_k = 1$ and given expectation value of energy $\langle E\rangle_r = \sum_k \rho_k(r) E_k$.
- two most common choices are linear averaging ($r = 1$, $\langle E\rangle = \sum_k p_k E_k$) and q-averaging ($r = q$)
- for both cases we find the MaxEnt distributions and discuss their properties
- the maximization under constraints is done via the Lagrange multiplier method:

$$\mathcal{L}_{q,r}(\mathcal{P}) = \mathcal{D}_q(\mathcal{P}) - \Omega\left(\sum_k \rho_k(r) E_k - \langle E\rangle_r\right) - \Phi\left(\sum_k p_k - 1\right)$$

Hybrid entropy

MaxEnt for $r = q$

- condition $\frac{\partial \mathcal{L}_{q,q}(\mathcal{P})}{\partial p_i} = 0$ leads to equation

$$\kappa p_i^{1-q} = q \ln p_i + \mathcal{E}_i$$

where $\kappa = \sum_j p_j^q$ and $\mathcal{E}_i = 1 + \frac{q \ln(-\Phi)}{1-q} + \frac{q\Omega}{-\Phi}(E_i - \langle E \rangle_q)$

- previous equation can be solved in terms of Lambert W-function defined as $W(x)e^{W(x)} = x$
- we obtain $p_i = \left[\frac{q}{\kappa(q-1)} W\left(\frac{\kappa(q-1)}{q} e^{(q-1)\mathcal{E}_i/q}\right)\right]^{1/(1-q)}$
- Lambert W-function is defined only on interval $[-1/e, \infty)$, so there are energy regions with zero probability - energy gaps

Hybrid entropy

MaxEnt for $r = q$

- Similarly to previous case, from condition $\frac{\partial \mathcal{L}_{q,1}(\mathcal{P})}{\partial p_i} = 0$ we obtain

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

- The solution of previous equation is

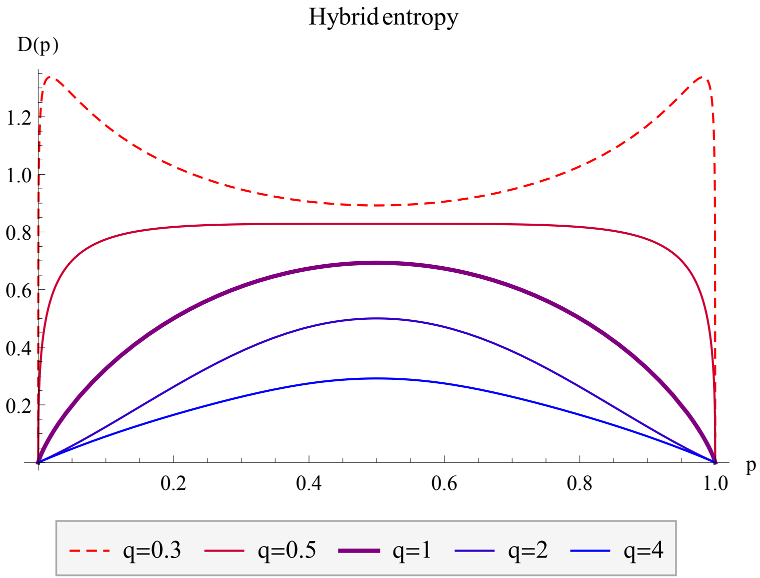
$$p_i = \left[\frac{q\Phi}{(q-1)\kappa(\Phi + \Omega \Delta E_i)} W\left(-\frac{\kappa(q-1)}{\Phi q} \exp\left(\frac{q-1}{q}\right) \left(1 + \frac{\Omega}{\Phi} \Delta E_i\right) \right) \right]^{1/(1-q)}$$

- Also in this case we observe energy gaps and forbidden energies

Maximality axiom and and (Schur-)concavity

- In the derivation of the hybrid entropy was not used the maximality axiom, it is necessary to check its validity
- We focus on the term $\langle \ln \mathcal{P} \rangle_q = \sum_k \rho_k(q) \ln p_k$
- For $q < 1/2$, $\langle \ln \mathcal{P} \rangle_q$ does not become maximal value for equal probabilities
- Thus, the hybrid entropy is properly defined only for $q \geq 1/2$
- hybrid entropy is for concave for $q \geq 1$ and Schur-concave for $q \geq 1/2$ (weaker version of concavity)

Hybrid entropy



Discussion

- the main question is if generalized entropies are a fundamental concept, similarly to thermodynamical entropy
- one could think about systems with q-gaussian distributions
- MaxEnt distribution does not determine entropy: q-gaussian distribution can be obtained from
  - **Tsallis entropy** under constraints $\sum_i p_i = 1$, $\sum_i p_i E_i = < E >$
  - **Shannon entropy** under constraints $\sum_i p_i = 1, \sum_i p_i^q = N_q$, $\sum_i E_i \rho_i(q) = < E >_q$
- same distributions - different entropies (correlated walk vs accelerated walk)
- problems with operational definitions (Tsallis entropy cannot be measured) and self-referential entropy

CONCLUSIONS

- entropy is the concept used in many different scientific fields
- the talk compared its definition and connection is thermodynamics, statistical mechanics and information theory
- several generalized entropies were presented - Rényi entropy, Tsallis entropy
- we have combined axiomatic of Rényi entropy and Tsallis entropy and obtained a new class of hybrid entropies
- generalized entropies can be useful tools for description of various systems
- usually, they are not fundamental concepts (contrary to thermodynamical entropy)

References

- C.E. Shannon. A mathematical theory of communication. Bell System Technical Journal, 27(4):623656, 1948.
- E. T. Jaynes. Information theory and statistical mechanics. Physical Review, 106:620630, 1957.
- A. Rényi. Selected Papers of Alfréd Rényi. Akadmiai Kiad, 1976.
- J. Havrda and F. Charvát. Quantification method of classification processes. concept of structural α-entropy. Kybernetika, 03(1):3035, 1967.
- C. Tsallis. Possible generalization of boltzmann-gibbs statistics. Journal of Statistical Physics, 52(1-2):479487, 1988.
- P. Jizba, J. Korbel. On q-non-extensive statistics with non-Tsallisian entropy. Physica A 444, 808-827, 2016.

Thank you for your attention.