

Foundations of Entropy IV

Axiomatic approaches

Lecture series at the
School on Information, Noise, and Physics of Life
Nis 19.-30. September 2022

by Jan Korbel

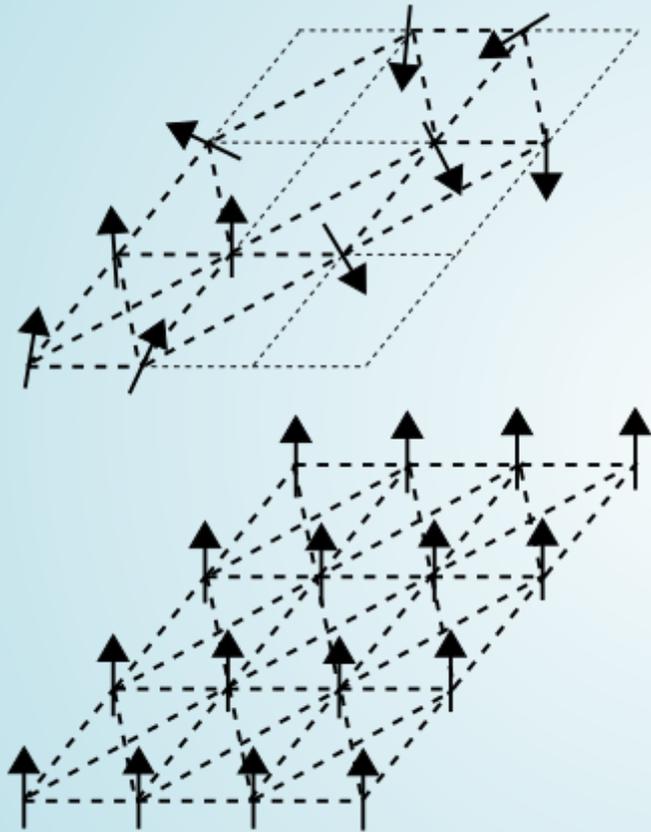
all slides can be found at: [**slides.com/jankorbel**](https://slides.com/jankorbel)

Activity IV

You have 3 minutes to write down on a piece of paper:

What is the most important
result/implication/phenomenon
that is related to **entropy**?

Spin glass



III. Niklas Elmehed © Nobel Prize
Outreach
Giorgio Parisi

Axiomatic approaches

Now we do the opposite approach compared to lecture **II**.
We postulate the properties we think entropy should have
and derive the corresponding entropic functional

These axiomatic approaches have different nature, we will
discuss their possible connection

Shannon-Khinchin axioms

you know them from the other lectures

Introduced independently by Shannon and Khinchin

Motivated by information theory

1. **Continuity.**—Entropy is a continuous function of the probability distribution only.
2. **Maximality.**— Entropy is maximal for the uniform distribution.
3. **Expandability.**— Adding an event with zero probability does not change the entropy.
4. **Additivity.**— $S(A \cup B) = S(A) + S(B|A)$ where
$$S(B|A) = \sum_i p_i^A S(B|A = a_i)$$

These four axioms uniquely determine Shannon entropy

$$S(P) = - \sum_i p_i \log p_i$$

SK axioms serve as a starting point for other axiomatic schemes

Non-additive SK axioms

Several axiomatic schemes generalize axiom **SK4**.

One possibility is to generalize additivity. The most prominent example is **q-additivity**

$$S(A \cup B) = S(A) \oplus_q S(B|A)$$

where $x \oplus_q y = x + y + (1 - q)xy$ is *q-addition*

$S(B|A) = \sum_i \rho_i(q)^A S(B|A = a_i)$ is *conditional entropy*

and $\rho_i = p_i^q / \sum_k p_k^q$ is *escort distribution*.

This uniquely determines Tsallis entropy

$$S_q(p) = \frac{1}{1 - q} \left(\sum_i p_i^q - 1 \right)$$

Kolmogorov-Nagumo average

Another possibility is to consider a different type of averaging

In the original SK axioms, the conditional entropy is defined as the arithmetic average of $S(B|A = a_i)$

We can use alternative averaging, as *Kolmogoro-Nagumo average*

$$\langle X \rangle_f = f^{-1} \left(\sum_i p_i f(x_i) \right)$$

By keeping additivity, but taking $S(B|A) = f^{-1}(\sum_i \rho_i(q)^A f(S(B|A = a_i)))$

for $f(x) = \frac{e^{(1-q)x} - 1}{1-q}$ we uniquely obtain Rényi entropy

$$R_q(p) = \frac{1}{1-q} \log \sum_i p_i^q$$

Generalized SK axioms and pseudo-additive entropies

Generalized Shannon–Khinchin axioms and uniqueness theorem for pseudo-additive entropies[☆]

Velimir M. Ilić^{a,*}, Miomir S. Stanković^b

[GSK4] Let $P = (p_1, \dots, p_n) \in \Delta_n$, $PQ = (r_{11}, r_{12}, \dots, r_{nm}) \in \Delta_{nm}$, $n, m \in \mathbb{N}$, $n, m > 1$ such that $p_i = \sum_{j=1}^m r_{ij}$, and $Q_{jk} = (q_{1jk}, \dots, q_{mjk}) \in \Delta_m$, where $q_{ijk} = r_{ik}/p_k$ and $\alpha \in \mathbb{R}_0^+$ is some fixed parameter. Then,

$$\mathcal{H}_{nm}(PQ) = \mathcal{H}_n(P) \oplus \mathcal{H}_m(Q|P), \quad \text{where } \mathcal{H}_m(Q|P) = g^{-1} \left(\sum_k p_k^{(\alpha)} g(\mathcal{H}_m(Q_{jk})) \right), \quad (28)$$

where g is the invertible continuous function.

In the following theorem we establish the class of entropies characterized by the generalized Shannon–Khinchin axioms.

Theorem 5.1. A function $\mathcal{H}_n : \Delta_n \rightarrow \mathbb{R}_0^+$, satisfies [SK1]–[SK3] and [GSK4] iff it belongs to the following family parameterized by $\tau, \lambda, \alpha \in \mathbb{R}$:

$$\mathcal{H}_n(P) = h(\mathcal{N}_n(P)) \begin{cases} h \left(\tau \sum_{k=1}^n p_k \log_2 p_k \right), & \tau < 0, \text{ for } \lambda = 0, \\ h \left(\frac{1}{\lambda} \log_2 \left(\sum_{k=1}^n p_k^\alpha \right) \right), & \alpha > 0, \lambda(1 - \alpha) > 0, \text{ for } \lambda \neq 0. \end{cases} \quad (29)$$

Entropy composability and group entropies

Beyond the Shannon–Khinchin formulation: The composability axiom and the universal-group entropy

Piergiulio Tempesta *

Definition 1. An entropy S is strongly (or strictly) *composable* if there exists a smooth function of two real variables $\Phi(x, y)$ such that

(C1)

$$S(A \cup B) = \Phi(S(A), S(B); \{\eta\}) \quad (1)$$

where A and B are two statistically independent systems, each defined in terms of an arbitrary probability distribution $\{p_i\}_{i=1}^W$, $\{\eta\}$ is a possible set of real parameters, with the properties:

(C2) Symmetry:

$$\Phi(x, y) = \Phi(y, x); \quad (2)$$

(C3) Associativity:

$$\Phi(x, \Phi(y, z)) = \Phi(\Phi(x, y), z); \quad (3)$$

(C4) Null-composability:

$$\Phi(x, 0) = x. \quad (4)$$

Entropy composability and group entropies

Definition 3. Let $\{p_i\}_{i=1,\dots,W}$, $W \geq 1$, with $\sum_{i=1}^W p_i = 1$, be a discrete probability distribution. Let

$$G(t) = \sum_{k=0}^{\infty} a_k \frac{t^{k+1}}{k+1} \quad (9)$$

be a real analytic function, where $\{a_k\}_{k \in \mathbb{N}}$ is a sequence of parameters, with $a_0 \neq 0$, such that the function $S_U : \mathcal{P}_W \rightarrow \mathbb{R}^+ \cup \{0\}$, defined by

$$S_U(p_1, \dots, p_W) := k_B \sum_{i=1}^W p_i G\left(\ln \frac{1}{p_i}\right), \quad (10)$$

is a strictly concave one. This function will be called the *universal-group entropy*.

Entropy and scaling

We have been mentioning the issue of **extensivity** before

Let us see how the multiplicity and entropy scales with size N

This allows us to introduce a classification of entropies

Multiplicity scaling

How the sample space changes when we rescale its size $N \mapsto \lambda N$?

The ratio behaves like $\frac{W(\lambda N)}{W(N)} \sim \lambda^{c_0}$ for $N \rightarrow \infty$

the **exponent** can be extracted by $\frac{d}{d\lambda} \Big|_{\lambda=1}: c_0 = \lim_{N \rightarrow \infty} \frac{NW'(N)}{W(N)}$

For the **leading term** we have $W(N) \sim N^{c_0}$.

Is it only possible scaling? We have $\frac{W(\lambda N)}{W(N)} \frac{N^{c_0}}{(\lambda N)^{c_0}} \sim 1$

Let us use the other rescaling $N \mapsto N^\lambda$

The we get that $\frac{W(N^\lambda)}{W(N)} \frac{N^{c_0}}{N^{\lambda c_0}} \sim \lambda^{c_1}$

First correction is $W(N) \sim N^{c_0} (\log N)^{c_1}$

It is the same scaling like for **(c, d)-entropy**

Can we go further?

Multiplicity scaling

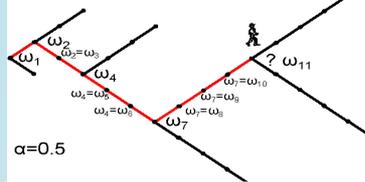
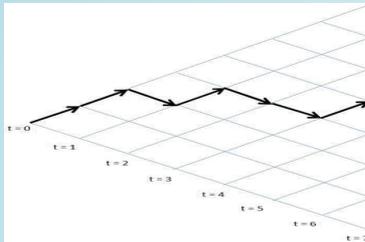
- We define the set of rescalings $r_\lambda^{(n)}(x) := \exp^{(n)}(\lambda \log^{(n)}(x))$
 - $f^{(n)}(x) = \underbrace{f(f(\dots(f(x))\dots))}_{n \text{ times}}$
 - $r_\lambda^{(0)}(x) = \lambda x$, $r_\lambda^{(1)}(x) = x^\lambda$, $r_\lambda^{(2)}(x) = e^{\log(x)^\lambda}$, ...
 - They form a **group**: $r_\lambda^{(n)}(r_{\lambda'}^{(n)}) = r_{\lambda\lambda'}^{(n)}$, $(r_\lambda^{(n)})^{-1} = r_{1/\lambda}^{(n)}$
 $r_1^{(n)}(x) = x$
- We **repeat** the procedure: $\frac{W(N^\lambda)}{W(N)} \frac{N^{c_0} (\log N)^{c_1}}{N^{\lambda c_0} (\log N^\lambda)^{c_1}} \sim 1$,
 - We take $N \mapsto r_\lambda^{(2)}(N)$
 - $\frac{W(r_\lambda^{(2)}(N))}{W(N)} \frac{N^{c_0} (\log N)^{c_1}}{r_\lambda^{(2)}(N)^{c_0} (\log r_\lambda^{(2)}(N))^{c_1}} \sim \lambda^{c_2}$,
 - **Second correction** is $W(N) \sim N^{c_0} (\log N)^{c_1} (\log \log N)^{c_2}$

Multiplicity scaling

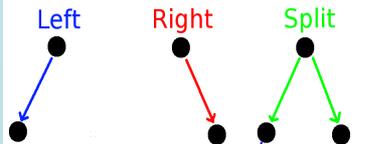
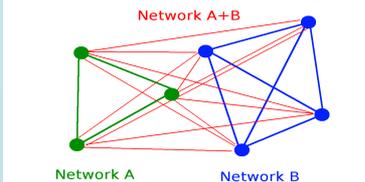
- **General correction** $\frac{W(r_\lambda^{(k)}(N))}{W(N)} \prod_{j=0}^{k-1} \left(\frac{\log^{(j)} N}{\log^{(j)}(r_\lambda^{(k)}(N))} \right)^{c_j} \sim \lambda^{c_k}$
- **Possible issue:** what if $c_0 = +\infty$? $W(N)$ grows faster than any N^α
- We replace $W(N) \mapsto \log W(N)$
- The leading order scaling is $\frac{\log W(\lambda N)}{\log W(N)} \sim \lambda^{c_0}$ for $N \rightarrow \infty$
- So we have $W(N) \sim \exp(N^{c_0})$
- If this is not enough, we replace $W(N) \mapsto \log^{(l)} W(N)$
so that we get **finite** c_0
- General expansion of $W(N)$ is
$$W(N) \sim \exp^{(l)}(N^{c_0} (\log N)^{c_1} (\log \log N)^{c_2} \dots)$$

Extensive entropy

- We can do the same procedure with **entropy** $S(W)$
- **Leading order** scaling: $\frac{S(\lambda W)}{S(W)} \sim \lambda^{d_0}$
- **First correction** $\frac{S(W^\lambda)}{S(W)} \frac{W^{d_0}}{W^{\lambda d_0}} \sim \lambda^{d_1}$
 - First two scalings correspond to **(c, d)-entropy**
for $c = 1 - d_0$ and $d = d_1$
- **Scaling expansion** of entropy
$$S(W) \sim W^{d_0} (\log W)^{d_1} (\log \log W)^{d_2} \dots$$
- Requirement of **extensivity** $S(W(N)) \sim N$ determines the relation between c and d :
 - $d_l = 1/c_0,$ $d_{l+k} = -c_k/c_0$ for $k = 1, 2, \dots$



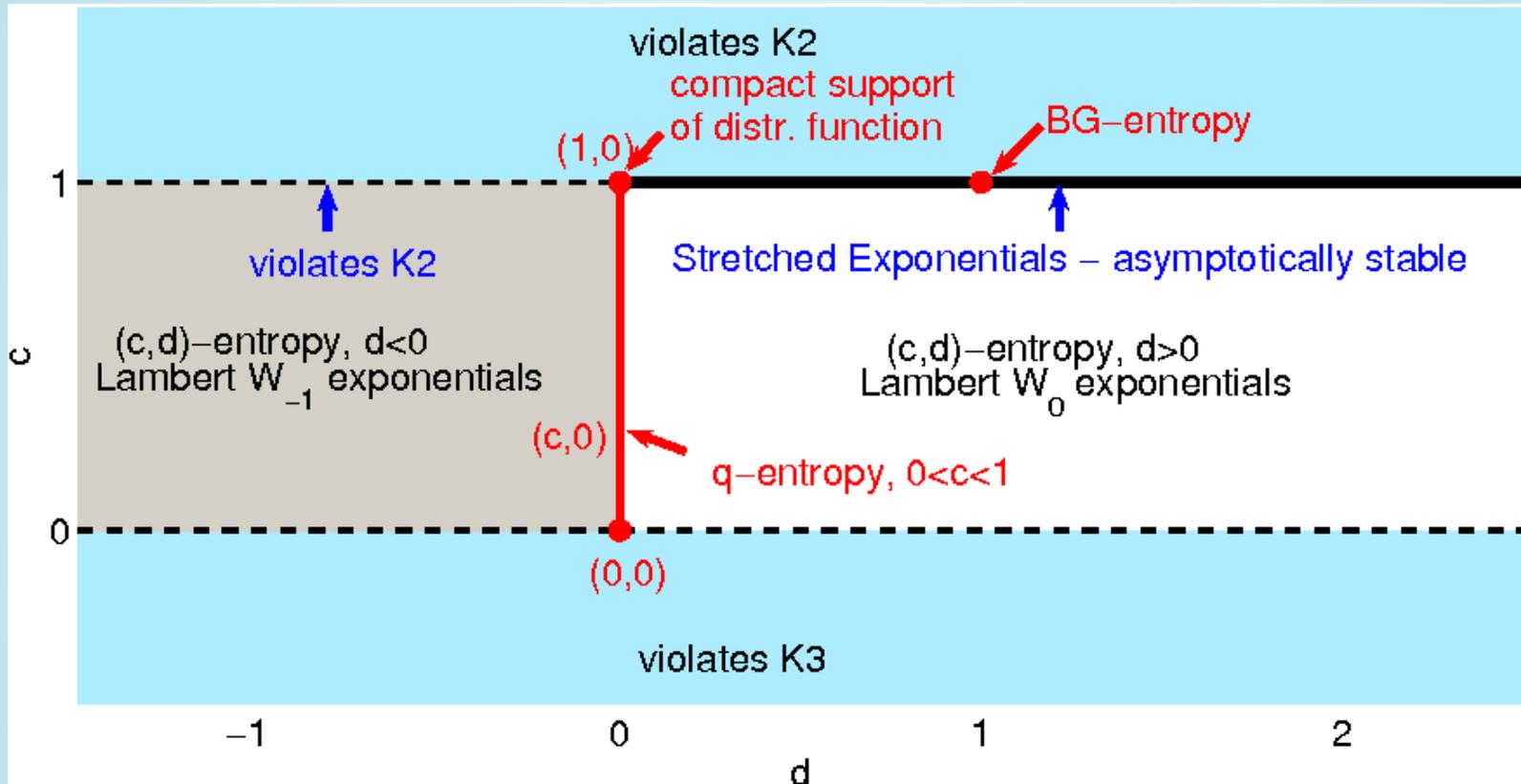
N	W(N)	Configurations
1	2	
2	5	
3	14	
4	43	



Process	d_0	d_1	d_2	$S(W)$
Random walk $W(N) = 2^N$	0	1	0	$\log W$
Aging random walk $W(N) \approx 2^{\sqrt{N}/2} \sim 2^{N^{1/2}}$	0	2	0	$(\log W)^2$
Magnetic coins * $W(N) \approx N^{N/2} e^{2\sqrt{N}} \sim e^{N \log N}$	0	1	-1	$\log W / \log \log W$
Random network $W(N) = 2^{\binom{N}{2}} \sim 2^{N^2}$	0	1/2	0	$(\log W)^{1/2}$
Random walk cascade $W(N) = 2^{2^N} - 1 \sim 2^{2^N}$	0	0	1	$\log \log W$

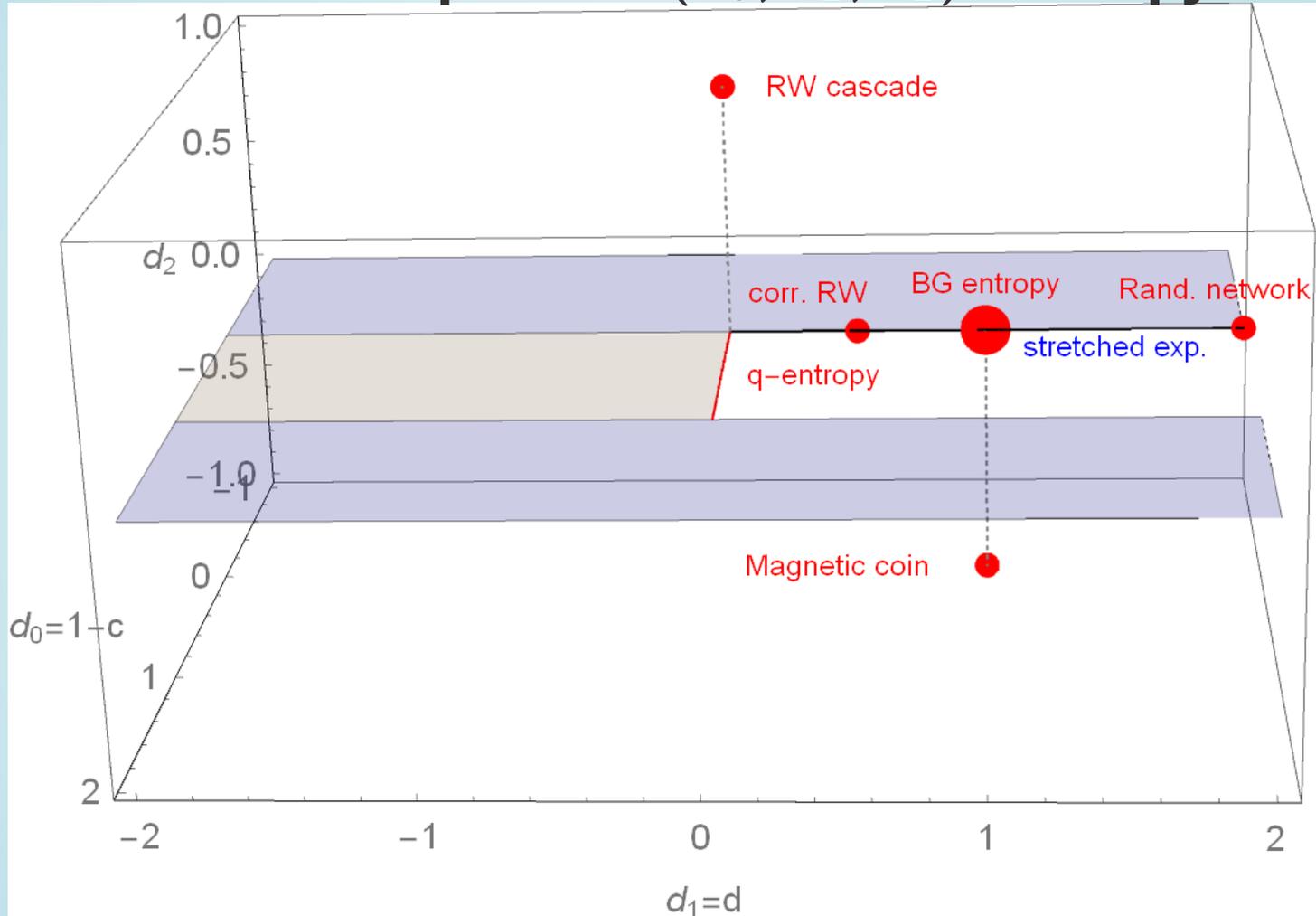
* H. Jensen et al. *J. Phys. A: Math. Theor.* **51** 375002

Parameter space of (c, d) entropy



How does it change for one more scaling exponent?

Parameter space of (d_0, d_1, d_2) -entropy



To fulfill SK axiom 2 (maximality): $d_i > 0$, to fulfill SK axiom 3 (expandability): $d_0 < 1$

Shore-Johnson axioms

- Axiomatization from the **Maximum entropy principle** point of view
- Principle of maximum entropy is an **inference method** and it should obey some statistical **consistency requirements**.
- **Shore and Johnson** set the consistency requirements:
 1. **Uniqueness**.—The result should be unique.
 2. **Permutation invariance**.—The permutation of states should not matter.
 3. **Subset independence**.—It should not matter whether one treats disjoint subsets of system states in terms of separate conditional distributions or in terms of the full distribution.
 4. **System independence**.—It should not matter whether one accounts for independent constraints related to disjoint subsystems separately in terms of marginal distributions or in terms of full-system constraints and joint distribution.
 5. **Maximality**.—In absence of any prior information the uniform distribution should be the solution.

History & Controversy of Shore-Johnson axioms

J. E. Shore, R. W. Johnson. *Axiomatic derivation of the principle of maximum entropy and the principle of minimum cross-entropy*. IEEE Trans. Inf. Theor. 26(1) (1980), 26. - **only Shannon**

J. Uffink, *Can the Maximum Entropy Principle be Explained as a Consistency Requirement?* Stud. Hist. Phil. Mod. Phys. 26(3), (1995), 223. - **larger class of entropies including Tsallis, Rényi, ..**

S. Pressé, K. Ghosh, J. Lee, K.A. Dill, *Nonadditive Entropies Yield Probability Distributions with Biases not Warranted by the Data*. Phys. Rev. Lett., 111 (2013), 180604. - **only Shannon - not Tsallis**

C. Tsallis, *Conceptual Inadequacy of the Shore and Johnson Axioms for Wide Classes of Complex Systems*. Entropy 17(5), (2015), 2853. - **S.-J. axioms are not adequate**

S. Pressé K. Ghosh, J. Lee, K.A. Dill, *Reply to C. Tsallis' Conceptual Inadequacy of the Shore and Johnson Axioms for Wide Classes of Complex Systems*. Entropy 17(7), (2015), 5043. - **S.-J. axioms are adequate**

B. Bagci, T. Oikonomou, Rényi entropy yields artificial biases not in the data and incorrect updating due to the finite-size data Phys. Rev. E 99 (2019) 032134 - **only Shannon - not Rényi**

P. Jizba, J.K. *Phys. Rev. Lett.* **122** (2019), 120601 - **Uffink is correct!**

(and the show goes on)

Shannon & Khinchin meet Shore & Johnson

Are the axioms set by theory of information and statistical inference different or can we find some overlap?

Let us consider the 4th SK axiom
in the form equivalent to composability axiom by P. Tempesta:

$$4. S(A \cup B) = f[f^{-1}(S(A)) \cdot f^{-1}(S(B|A))]$$
$$S(B|A) = S(B) \text{ if } B \text{ is independent of } A.$$

Entropies fulfilling SK and Sj:

$$S_q^f(P) = f \left[\left(\sum_i p_i^q \right)^{1/(1-q)} \right] = f \left[\exp_q \left(\sum_i p_i \log_q(1/p_i) \right) \right]$$

Non-equilibrium thermodynamics axioms

In ST lecture, you saw that Shannon entropy fulfills the second law of thermodynamics for linear Markov dynamics with detailed balance.

But is it the only possible entropy?

Our axioms are:

1. Linear Markov evolution - $\dot{p}_m = \sum_n (w_{mn}p_n - w_{nm}p_m)$

2. Detailed balance - $w_{mn}p_n^{st} = w_{nm}p_m^{st}$

3. Second law of thermodynamics: $\dot{S} = \dot{S}_i + \dot{S}_e$

where $\dot{S}_e = \beta Q$, and $\dot{S}_i \geq 0$ where $\dot{S}_i = 0 \Leftrightarrow p = p^{st}$

Then $S = - \sum_m p_m \log p_m$

This is a special case of more general result connecting non-linear master equations and generalized entropies

Summary