What do generalized entropies look like? An axiomatic approach for complex, non-ergodic systems

Stefan Thurner\textsuperscript{1,2} and Rudolf Hanel\textsuperscript{1}

\textsuperscript{1} Section for Science of Complex Systems, Medical University of Vienna, Spitalgasse 23, A-1090, Austria
\textsuperscript{2} Santa Fe Institute, 1399 Hyde Park Road, Santa Fe, NM 87501, USA

Shannon and Khinchin showed that assuming four information theoretic axioms the entropy must be of Boltzmann-Gibbs type, \( S = - \sum p_i \log p_i \). Here we note that in physical systems one of these axioms may be violated. For non-ergodic systems the so called separation axiom (Shannon-Khinchin axiom 4) will in general not be valid. We show that when this axiom is violated the entropy takes a more general form, \( S_{c,d} \propto \sum W \Gamma(d + 1, 1 - c \log p_i) \), where \( c \) and \( d \) are scaling exponents and \( \Gamma(a,b) \) is the incomplete gamma function. The exponents \( (c,d) \) define equivalence classes for all interacting and non interacting systems and unambiguously characterize any statistical system in its thermodynamic limit. The proof is possible because of two newly discovered scaling laws which any entropic form has to fulfill, if the first three Shannon-Khinchin axioms hold. \((c,d)\) can be used to define equivalence classes of statistical systems. A series of known entropies can be classified in terms of these equivalence classes. We show that the corresponding distribution functions are special forms of Lambert-\( W \) exponentials containing – as special cases – Boltzmann, stretched exponential and Tsallis distributions (power-laws). In the derivation we assume trace form entropies, \( S = \sum g(p_i) \), with \( g \) some function, however more general entropic forms can be classified along the same scaling analysis. In this contribution we largely follow the lines of thought presented in \cite{1}.

PACS numbers: 05.20.-y, 02.50.Cw, 05.90.+m

I. INTRODUCTION

Theorem number 2 in the seminal 1948 paper, The Mathematical Theory of Communication \cite{2}, by Claude Shannon, proves the existence of the one and only form of entropy, given that three fundamental requirements hold. A few years later A.I. Khinchin remarked in his Mathematical Foundations of Information Theory \cite{3}: “However, Shannon’s treatment is not always sufficiently complete and mathematically correct so that, besides having to free the theory from practical details, in many instances I have amplified and changed both the statement of definitions and the statement of proofs of theorems.” Khinchin adds a fourth axiom. The three fundamental requirements of Shannon, in the ‘amplified’ version of Khinchin, are known as the Shannon-Khinchin (SK) axioms. These axioms list the requirements needed for an entropy to be a reasonable measure of the ‘uncertainty’ about a finite probabilistic system. Khinchin further suggests to also use entropy as a measure of the information gained about a system when making an ‘experiment’, i.e. by observing a realization of the probabilistic system.

Khinchin’s first axiom states that for a system with \( W \) potential outcomes (states) each of which is given by a probability \( p_i \geq 0 \), with \( \sum p_i = 1 \), the entropy \( S(p_1, \ldots , p_W) \) as a measure of uncertainty about the system must take its maximum for the equi-distribution \( p_i = 1/W \), for all \( i \).

Khinchin’s second axiom (missing in \cite{2}) states that any entropy should remain invariant under adding zero-probability states to the system, i.e. \( S(p_1, \ldots , p_W) = S(p_1, \ldots , p_W, 0) \).

Khinchin’s third axiom (separability axiom) finally makes a statement of the composition of two finite probabilistic systems \( A \) and \( B \). If the systems are independent of each other, entropy should be additive, meaning that the entropy of the combined system \( A + B \) should be the sum of the individual systems, \( S(A + B) = S(A) + S(B) \). If the two systems are dependent on each other, the entropy of the combined system, i.e. the information given by the realization of the two finite schemes \( A \) and \( B \), \( S(A + B) \), is equal to the information gained by a realization of system \( A \), \( S(A) \), plus the mathematical expectation of information gained by a realization of system \( B \), after the realization of system \( A \), \( S(A + B) = S(A) + S|_A(B) \).

Khinchin’s fourth axiom is the requirement that entropy is a continuous function of all its arguments \( p_i \) and does not depend on anything else.

Given these axioms, the Uniqueness theorem \cite{3} states that the one and only possible entropy is \( S(p_1, \ldots , p_W) = -k \sum p_i \log p_i \), where \( k \) is an arbitrary positive constant. The result is of course the same as Shannon’s. We call the combination of 4 axioms the Shannon-Khinchin (SK) axioms.

From information theory now to physics, where systems may exist that violate the separability axiom. This might especially be the case for non-ergodic, complex systems exhibiting long-range and strong interactions. Such complex systems may show extremely rich behavior in contrast to simple ones, such as gases. There exists some hope that it should be possible to understand such systems also on a thermodynamical basis, meaning that a few measurable quantities would be sufficient to understand their macroscopic phenomena. If this would be possible, through an equivalent to the second law of thermodynamics, some appropriate entropy would enter as a fundamental concept relating the number of microstates.
in the system to its macroscopic properties. Guided by this hope, a series of so called generalized entropies have been suggested over the past decades, see [4-8] and Table 1. These entropies have been designed for different purposes and have not been related to a fundamental origin. Here we ask how generalized entropies can look like if they fulfill some of the Shannon-Khinchin axioms, but explicitly violate the separability axiom. We do this axiomatically as first presented in [9]. By doing so we can relate a large class of generalized entropies to a single root.

The reason why this axiom is violated in some physical, biological or social systems is broken ergodicity, i.e. that not all regions in phase space are visited and many micro states are effectively ‘forbidden’. Entropy relates the number of micro states of a system to an extensive property that plays the fundamental role in the systems thermodynamical description. Extensive means that if two initially isolated, i.e. sufficiently separated systems, A and B, with WA and WB the respective numbers of states, are brought together, the entropy of the combined system A + B is S(WA+B) = S(WA) + S(WB). WA+B is the number of states in the combined system A + B.

This is not to be confused with the form of the Shannon-Khinchin axioms in the light of the combined system A + B is S(WA+B) = S(WA) + S(WB). Both, extensive and additivity coincide if number of states in the combined system is W_{A+B} = W_AW_B. Clearly, for a non-interacting system Boltzmann-Gibbs-Shannon entropy, S_{BG}[p] = -\sum_g p_i \ln p_i, is extensive and additive. By ‘non-interacting’ (short-range, ergodic, sufficiently mixing, Markovian, ...) systems we mean W_{A+B} = W_AW_B. For interacting statistical systems the latter is in general not true: phase space is only partly visited and W_{A+B} < W_AW_B. In this case, an additive entropy such as Boltzmann-Gibbs can no longer be extensive and vice versa. To ensure extensivity of entropy, an entropic form should be found for the particular interacting statistical systems at hand. These entropic forms are called generalized entropies and usually assume trace form 33:

\[ S_g[p] = \sum_{i=1}^{W} g(p_i) \]  (1)

W being the number of states. Obviously not all generalized entropic forms are of this type. Rényi entropy e.g. is of the form G(\sum g(p_i)), with G a monotonic function. We use trace forms Eq. (1) for simplicity. Rényi forms can be studied in exactly the same way as will be shown, however at more technical cost.

Let us revisit the Shannon-Khinchin axioms in the light of generalized entropies of trace form Eq. (1). Specifically axioms SK1-SK3 (now re-ordered) have implications on the functional form of g

- SK1: The requirement that S depends continuously on p implies that g is a continuous function.
- SK2: The requirement that the entropy is maximal for the equi-distribution p_i = 1/W (for all i) implies that g is a concave function.
- SK3: The requirement that adding a zero-probability state to a system, W+1 with p_{W+1} = 0, does not change the entropy, implies that g(0) = 0.
- SK4 (separability axiom): The entropy of a system – composed of sub-systems A and B – equals the entropy of A plus the expectation value of the entropy of B, conditional on A. Note that this also corresponds exactly to Markovian processes.

As mentioned, if SK1 to SK4 hold, the only possible entropy is the Boltzmann-Gibbs-Shannon entropy. We are now going to derive the extensive entropy when the separability axiom SK4 is violated. Obviously this entropy will be more general and should contain BG entropy as a special case.

We now assume that axioms SK1, SK2, SK3 hold, i.e. we restrict ourselves to trace form entropies with g continuous, concave and g(0) = 0. These systems we call admissible systems.

This generalized entropy for (large) admissible statistical systems (SK1-SK3 hold) is derived from two hitherto unexplored fundamental scaling laws of extensive entropies. Both scaling laws are characterized by exponents c and d, respectively, which allow to uniquely define equivalence classes of entropies, meaning that two entropies are equivalent in the thermodynamic limit if their exponents (c, d) coincide. Each admissible system belongs to one of these equivalence classes (c, d). [10]

In terms of the exponents (c, d) we show in the following that all generalized entropies have the form

\[ S_{c,d}[W] \propto \sum_i^{W} \Gamma(d+1,1-c \log p_i) \quad \text{with} \quad \Gamma(a,b) = \int_0^\infty dt t^{a-1} \exp(-t) \quad \text{the incomplete Gamma-function}. \]

Admissible systems when combined with a maximum entropy principle show remarkably simple mathematical properties as recently demonstrated in [10].

II. THE SCALING LAWS OF ENTROPIES

We discuss two – a primary and a secondary – scaling properties of generalized entropies of trace form and assume the validity of the first 3 KS axioms. For equi-distribution p_i = 1/W (for all i) obviously, \[ S_g[W] = W g(1/W). \]

The primary scaling law is found from the relation

\[ S_g(\lambda W) = \lambda^{c} S_g(W) \cdot g(1/W). \]  (2)

We define a scaling function

\[ f(z) \equiv \lim_{x \to 0} \frac{g(zx)}{g(x)} \quad (0 < z < 1) \]  (3)

This function f for systems satisfying SK1-SK3, but violating SK4, can only be a power f(z) = z^c, with
tropies for stretched exponentials to \((1, d > 0)\). We expect values in the constraints \([9, 15]\).

With the observation that scaling law for entropies, particularly simple choice is

\[ \lambda \text{Eq. (4).} \]

Note that if \(x\) independent of property, see Theorem 3 in the appendix.

Remarkably, \(h_{c,d}(a)\) does not explicitly depend on \(c\). \(h_{c,d}(a)\) is an asymptotic property which is independent of property Eq. (4). Note that if \(c = 1\), concavity of \(g\) implies \(d \geq 0\). In principle this scheme can be iterated to higher orders, see appendix.

III. DERIVATION OF ENTROPY

We now ask which families of entropies, i.e. functions \(g_{c,d}\) fulfill the primary and the secondary scaling law. A particularly simple choice is

\[ g_{c,d,r}(x) = r A^{-d} e^{A (1 + d, A - c \ln x) - r c x} \]  

where \(A = \frac{cd}{1 - (1 - c)r}\) and \(r\) is an arbitrary constant \(r > 0\) (see below). For all choices of \(r\) the function \(g_{c,d,r}\) is a representative of the class \((c, d)\). This allows to choose \(r\) as a suitable function of \(c\) and \(d\). For example choose \(r = (1 - c + cd)^{-1}\), so that \(A = 1\), and

\[ S_{c,d}[p] = \frac{e \sum_i W^c}{1 - c + cd} \]  

It can be easily verified that \(S_{c,d}\) has the correct asymptotic properties, see Theorem 3 in the appendix.

A. Special cases of entropic equivalence classes

Let us look at some specific equivalence classes \((c, d)\)

- Boltzmann-Gibbs entropy belongs to the \((c, d) = (1, 1)\) class. One gets from Eq. (3)

\[ S_{1,1}[p] = \sum_i g_{1,1}(p_i) = -\sum_i p_i \ln p_i + 1 \]  

- Tsallis entropy belongs to the \((c, d) = (c, 0)\) class. From Eq. (5) and the choice \(r = 1/(1 - c)\) (see below) we get

\[ S_{c,0}[p] = \sum_i g_{c,0}(p_i) = \frac{1 - \sum_i p_i^c}{c - 1} + 1 \]  

Note, that although the pointwise limit \(c \to 1\) of Tsallis entropy yields BG entropy, the asymptotic properties \((c, 0)\) do not change continuously to \((1,1)\) in this limit! In other words the thermodynamic limit and the limit \(c \to 1\) do not commute.

- The entropy related to stretched exponentials [3] belongs to the \((c, d) = (1, d)\) classes, see Table 1. As a specific example we compute the \((c, d) = (1, 2)\) case,

\[ S_{1,2}[p] = 2 \left( 1 - \sum_i p_i \ln p_i \right) + \frac{1}{2} \sum_i p_i (\ln p_i)^2 \]

leading to a superposition of two entropy terms, the asymptotic behavior being dominated by the second.

- All entropies associated to distributions with compact support belong to \((c, d) = (1, 0)\). Clearly, all of these have the same (trivial) asymptotic behavior.

Other entropies which are special cases of our scheme are found in Table 1.

Inversely, for any given entropy we are now in the remarkable position to characterize all large SK1-SK3 systems by a pair of two exponents \((c,d)\), i.e. their scaling functions \(f\) and \(h_c\). See Fig. [11] For example, for \(g_{BG}(x) = -x \ln(x)\) we have \(f(z) = z\), i.e. \(c = 1\), and \(h_c(a) = 1 + a\), i.e. \(d = 1\). \(S_{BG}\) therefore belongs to the universality class \((c, d) = (1, 1)\). For
one finds f and Tsallis entropy, S, They must not be confused with power-laws.

c, d, and β, q, are the universality classes (c, d) = (c, d). The universality classes (c, d) are associated with stretched exponential distributions. All classes associated with (c, d) = (1, d), for d > 0 are associated with stretched exponential distributions. Expanding the k = 0 branch of the Lambert-W function W_0(x) = x - x^2 + ... for 1 > |x|, the limit c → 1 is seen to be a stretched exponential

\[ \lim_{c \to 1} E_{c,d,r}(x) = e^{-dr\left[\frac{1}{1-c}\right]^{\frac{1}{d} - 1}}. \]

Clearly, r does not effect its asymptotic properties (tail of the distribution), but can be used to incorporate finite size properties of the distribution function for small x. Examples of all of the above distribution functions are shown in Fig. 2.

A. Special cases of distribution functions

It is easy to verify that the class (c, d) = (1, 1) leads to Boltzmann distributions, and the class (c, d) = (c, 0) exhibits power-laws, or more precisely, Tsallis distributions i.e. q-exponentials. All classes associated with (c, d) = (1, d), for d > 0 are associated with stretched exponential distributions. Expanding the k = 0 branch of the Lambert-W function W_0(x) = x - x^2 + ... for 1 > |x|, the limit c → 1 is seen to be a stretched exponential

\[ \lim_{c \to 1} E_{c,d,r}(x) = e^{-dr\left[\frac{1}{1-c}\right]^{\frac{1}{d} - 1}}. \]

B. Finite size effects and the parameter r

In Eq. 9 we chose r = (1 - c + cd)^{-1}. This is not the most general case and the only limitations on r – if one requires the generalized logarithms to have the usual properties Λ(1) = 0 and Λ'(1) = 1 to hold – are

\[ d > 0 : \quad r < \frac{1}{1-c}, \]
\[ d = 0 : \quad r = \frac{1}{1-c}, \]
\[ d < 0 : \quad r > \frac{1}{1-c}. \]

Every choice of r gives a representative of the equivalence class (c, d), i.e. r has no effect on the thermodynamic limit and therefore can be used to encode potential finite-size characteristics of a system at hand. In case of no such effects, practical choices are r = (1 - c + cd)^{-1} for d > 0, and r = exp(-d)/(1 - c) for d < 0.

IV. DISTRIBUTION FUNCTIONS

Distribution functions associated with our Γ-entropy, Eq. 8, can be derived from so-called generalized logarithms of the entropy. Under the maximum entropy principle (given ordinary constraints) the inverse functions of these logarithms, E = Λ^{-1}, are the distribution functions, p(e) = E_{c,d,r}(-c). Following 8, 15 the gen-

![Graph](image-url)
V. A PHYSICAL SYSTEM AS AN EXAMPLE

From axiomatic derivations no physical relevance can be inferred. To demonstrate the applicability of the proposed classification scheme to concrete physical systems, consider over-damped interacting particles moving in a narrow channel,

$$\mu \vec{v}_i = \sum_{j \neq i} \vec{J}(\vec{r}_i - \vec{r}_j) + \vec{F}(\vec{r}_i) + \eta(\vec{r}_i, t) \ , \ (17)$$

where $v_i$ is velocity of $i$th particle, $\mu$ viscosity, $F$ an external force, $\vec{J}(\vec{r})$ a linear repulsive particle-particle interaction, $\eta$ uncorrelated thermal noise with $\langle \eta \rangle = 0$ and $\langle \eta^2 \rangle = \kappa T$, with $\kappa$ a characteristic length of the pair interaction. For details see [22], where the non-linear Fokker-Planck equation for the spatial distribution $\rho(x)$ of the particles was solved. It was shown that the corresponding entropy is a superposition of BG and Tsallis entropy with $q = 2$. Note the similarity to our example in Eq. (12) where a similar superposition emerged naturally. At high temperatures the BG contribution dominates i.e. particles diffuse normally, while at zero temperature the system is governed by Tsallis entropy with a non-Gaussian parabolic diffusion profile of compact support. The concrete stationary distribution functions given in [22] can be used to demonstrate that the system is found either in the asymptotic equivalence class $(c, d) = (1, 1)$ (BG entropy) or in $(c, d) = (1, 0)$ (compact support entropies) depending on the temperature of the heat bath. Expressing the integral by a discrete sum the entropy given in [22] can be used to demonstrate the existence of the two sets of classes. This shows that the classification is applicable for concrete physical systems and suggests further that the equivalence class $(c, d)$ may even depend on macro variables of the system such as the temperature of the heat bath.

VI. RÉNYI-TYPE ENTropies

Rényi entropy is obtained by relaxing SK4 to the pure (unconditional) additivity condition. Following the same scaling idea for Rényi-type entropies, $S = G(\sum_{i=1}^{W} g(p_i))$, with $G$ and $g$ some functions, one gets

$$\lim_{W \to \infty} \frac{S(\lambda W)}{S(W)} = \lim_{s \to +\infty} \frac{G(\lambda f_\gamma(s))}{G(s)} \ , \ (18)$$

where $f_\gamma(z) = \lim_{x \to z} g(x)/g(x)$. The expression $f_\gamma(s) \equiv \lim_x G(sy)/G(s)$, provides the starting point for deeper analysis which now gets more involved. In particular, for Rényi entropy with $G(x) \equiv \ln(x)/(1 - \alpha)$ and $g(x) \equiv x^\alpha$, the asymptotic properties yield the class $(c, d) = (1, 1)$ (BG entropy) meaning that Rényi entropy is additive. However, in contrast to the trace form entropies used above, Rényi entropy can be shown to be not Lesche stable, as was observed before [16, 20]. All of the $S = \sum_{i=1}^{W} g(p_i)$ entropies can be shown to be Lesche stable, see Theorem 4 in the appendix.

VII. DISCUSSION

We studied the scaling laws of trace form entropies which are constrained by the first three Shannon-Khinchin axioms in the thermodynamic limit. In analogy to critical exponents these laws are characterized by two scaling exponents $(c, d)$, which define equivalence relations on entropic forms. We showed that a single entropic form – parameterized by the two exponents – covers all admissible systems (Shannon-Khinchin axioms 1-3 hold, 4 is violated). In other words every statistical system has its pair of unique exponents in the large size limit, its entropy is then given by $S_{c,d} \sim \sum_{i=1}^{W} \Gamma(1 + d, 1 - c \ln p_i)$. The exponents for BG systems are $(c, d) = (1, 1)$, systems characterized by stretched exponentials belong to the class $(c, d) = (1, d)$, and Tsallis systems have

| entropy | $c$ | $d$ | reference |
|---------|-----|-----|-----------|
| $S_{BG}$ | 1 | 1 | [3] |
| $S_{q<1}(p) = \frac{1}{1-q}\sum_{j=1}^{p_i}$ | $q < 1$ | $c = q < 1$ | [4] |
| $S_{q>1}(p) = \frac{1}{1-q}\sum_{j=1}^{p_i}$ | $q > 1$ | 1 | [4] |
| $S_{\gamma} = \frac{1}{\gamma} \sum_{i=1}^{p_i}$ | $0 < \kappa < 1$ | 1 | [6] |
| $S_{\beta} = \beta \sum_{p_i}$ | $b > 0$ | 1 | [7] |
| $S_{\beta} = \beta \sum_{p_i}$ | $c = \beta$ | 1 | [12] |
(c, d) = (q, 0). In the context of a maximum entropy principle, the associated distribution functions of all systems (c, d) are shown to belong to a class of exponentials involving Lambert-W functions, given in Eq. (4). There are no other options for tails in distribution functions other than these.

The equivalence classes characterized by the exponents (c, d) form basins of asymptotic equivalence. In general these basins characterize interacting statistical (non-additive) systems. There exists an analogy between these basins of asymptotic equivalence and the basin of attraction of weakly interacting, uncorrelated systems subject to the law of large numbers, i.e. the central limit theorem. Any system within a given equivalence class may show individual characteristics as long as it is small. Systems belonging to the same class will start behaving similarly as they become larger and in the thermodynamic limit they become identical. Distribution functions converge to those uniquely determined by (c, d).

Our framework shows that for non-interacting systems c = 1. Setting \( \lambda = W_B \) in Eq. (2) and Eq. (3), immediately implies \( S(W_A W_B) / S(W_A) \sim W_B^{-c} \). This means that if for such a system it would be true that c ≠ 1, then adding only a few independent states to a system would explosively change its entropy and extensivity would be strongly violated. A further interesting feature of admissible systems is that they all are what has been called Lesche stable. As a practical note Lesche stability corresponds one-to-one to the continuity of the scaling function f and can now be checked by a trivial verification of this property (Eq. (3)). The proof is given in Theorem 4 in the appendix.

Finally we remark that the classification scheme for generalized entropic forms of type \( S = \sum_i g(p_i) \) can be extended to entropies of e.g. Rényi type, i.e. \( S = G(\sum_i g(p_i)) \). We demonstrated that generalized entropies can be applied to actual physical systems. We hypothesize that many complex statistical systems are indeed admissible systems of equivalence classes (c, d), with 0 < c < 1.

---

**VIII. APPENDIX**

**Theorem 1:** Let \( g \) be a continuous, concave function on \([0, 1]\) with \( g(0) = 0 \) and let \( f(z) = \lim_{x \to 0^+} g(zx) / g(x) \) be continuous, then f is of the form \( f(z) = z^c \) with \( c \in (0, 1) \).

**Proof.** Note that \( f(ab) = \lim_{x \to 0} g(abx) / g(x) = \lim_{x \to 0} g(abx) / (g(bx))((g(bx))/g(x)) = f(a)f(b) \). All pathological solutions are excluded by the requirement that \( f \) is continuous. So \( f(ab) = f(a)f(b) \) implies that \( f(z) = z^c \) is the only possible solution of this equation. Further, since \( g(0) = 0 \), also \( \lim_{x \to 0} g(0x)/g(x) = 0 \), and it follows that \( f(0) = 0 \). This necessarily implies that \( c > 0 \). \( f(z) = z^c \) also has to be concave since \( g(zx)/g(x) \) is concave in \( z \) for arbitrarily small, fixed \( x > 0 \). Therefore \( c \leq 1 \).

**Theorem 2:** Let \( g \) be like in Theorem 1 and let \( f(z) = z^c \) then \( h_c \) in Eq. (4) is a constant of the form \( h_c(a) = (1 + a)^d \) for some constant \( d \).

**Proof.** We can determine \( h_c(a) \) again by a similar trick as we have used for \( f \).

\[
h_c(a) = \lim_{x \to 0} g(y^{x^{-d}}) / g^{(x^{-d} + 1)}
= g^{(x^{-d} + 1)} / g^{(1)}
= h_c \left( \frac{a + 1}{b} - 1 \right) h_c(b - 1),
\]

for some constant \( b \). By a simple transformation of variables, \( a = bb' - 1 \), one gets \( h_c(bb' - 1) = h_c(b - 1)h_c(b' - 1) \). Setting \( H(x) = h_c(x - 1) \) one again gets \( H(bb') = H(b)H(b') \). So \( H(x) = x^d \) for some constant \( d \) and consequently \( h_c(a) \) is of the form \((1 + a)^d\).
**Theorem 3:** The entropy based on $g_{c,d,r}$, Eq. (8), has the desired asymptotic properties.

**Proof.** Let $g$ be like in Theorem 1, i.e. let $f(z) = z^c$ with $0 < c \leq 1$, then

$$\lim_{x \to 0^+} \frac{g'(x)}{g(x)} = c. \quad (19)$$

Consider

$$\lim_{x \to 0^+} \frac{g(z)x}{g(x)} = \frac{1}{1-z} \left( \frac{g(x)}{g(z)} - g(x) \right) = z^{c-1} - 1. \quad (20)$$

Taking the limit $z \to 1$ on both sides completes the first part of the proof. Further, two functions $g_A$ and $g_B$ generate equivalent entropic forms if

$$\lim_{x \to 0^+} \frac{g_A(x)}{g_B(x)} = \lim_{x \to 0^+} \frac{g_A'(x)}{g_B'(x)},$$

which is just the rule of L'Hospital shown to hold for the considered families of functions $g$. This is true since, either $\lim_{x \to 0^+} g_A(x)/g_B(x) = \phi$ with $0 < \phi < \infty$ and $c_A = c_B$, i.e. $g_A$ and $g_B$ are equivalent, or $g_A$ and $g_B$ are inequivalent, i.e. $c_A \neq c_B$ but $\phi = 0$ or $\phi \to \infty$.

So if one can find a function $g_{test}$, having the desired asymptotic exponents $c$ and $d$, it suffices to show that $0 < -\lim_{x \to 0^+} \Lambda_{c,d,r}(x)/g_{test}(x) < \infty$, where $\Lambda_{c,d,r}$ is the generalized logarithm Eq. (13) associated with the generalized entropy Eq. (6). The test function $g_{test}(x) = x^c \log(1/x)^d$ is of class $(c,d)$, as can be verified easily. Unfortunately $g_{test}$ can not be used to define the generalized entropy due to several technicalities. In particular $g_{test}$ lacks concavity around $x \sim 1$ for a considerable range of $(c,d)$ values, which then makes it impossible to define proper generalized logarithms and generalized exponential functions on the entire interval $x \in [0,1]$. However, we only need the asymptotic properties of $g_{test}$ and for $x \sim 0$ the function $g_{test}$ does not violate concavity or any other required condition. The first derivative is $g_{test}'(x) = x^{c-d-1} \log(1/x)^d (c \log(1/x) - d)$. With this we finally get

$$\lim_{x \to 0^+} \frac{\Lambda_{c,d,r}(x)}{g_{test}'(x)} = \frac{r - D^{-1}(\frac{z}{c})^{c-1} (\log(\frac{z}{c}))^d \frac{x}{x^{c-1} \log(\frac{z}{c})^d (c \log(\frac{1}{x}) - d)}}{\frac{2^{1-c}}{kld^{d}}}. \quad (21)$$

Since $0 < \frac{2^{1-c}}{kld^{d}} < \infty$ this proves that the Gamma-entropy $g_{c,d,r}$, Eq. (10), represents the equivalence classes $(c,d)$.

**Lesche stability of trace form entropies**

The Lesche stability criterion is a uniform-equicontinuity property of functionals $S[p]$ on families of probability functions $\{p^{(W)}\}_{W=1}^{\infty}$ where $p^{(W)} = (p^{(W)}_i)_{i=1}^{W}$. The criterion is phrased as follows:

Let $p^{(W)}$ and $q^{(W)}$ be probabilities on $W$ states. An entropic form $S$ is Lesche stable if for all $\epsilon > 0$ and all $W$ there is a $\delta > 0$ such that

$$||p^{(W)} - q^{(W)}||_1 < \delta \Rightarrow |S[p^{(W)}] - S[q^{(W)}]| < \epsilon S(W), \quad (22)$$

where $S(W)$ is again the maximal possible entropy for $W$ states.

We characterize Lesche stability on the class of our generalized entropic forms in terms of continuity of $f$ in

**Theorem 4:** Let $p_i \geq 0$ be a probability and $W$ the number of states $i$. Let $g$ be a concave, continuous function on $[0,1]$, continuously differentiable on the semi-open interval $(0,1]$ and $g(0) = 0$. The entropic form $S_g[p] = \sum_{i=1}^{W} g(p_i)$ is Lesche stable iff the function $f(z) = \lim_{x \to 0^+} g(xz)/g(x)$ is continuous on $z \in [0,1]$.

**Proof.** SK2 states that maximal entropy is given by $\hat{S}_g(W) = Wg(1/W)$. We identify the worst case scenario for $|S_g[p] - S_g[q]|$, where $p$ and $q$ are probabilities on the $W$ states. For this maximize $G[p,q] = |S_g[p] - S_g[q]| - \alpha(\sum_i p_i - 1) - \beta(\sum_i q_i - 1) - \gamma(\sum_i |p_i - q_i| - \delta)$, where $\alpha$, $\beta$ and $\gamma$ are Lagrange multipliers. Without loss of generality assume that $S_g[p] > S_g[q]$. Thus condition $\partial G/\partial p_i = 0$ gives $g'(p_i) + \gamma \text{sign}(p_i - q_i) - \alpha = 0$, where $g'$ is the derivative of $g$ and sign is the sign function. Similarly, $\partial G/\partial q_i = 0$ leads to $g'(q_i) + \gamma \text{sign}(p_i - q_i) + \beta = 0$. From this we see that both $p$ and $q$ can only possess two values $p_+$, $p_-$ and $q_+$, $q_-$, where one can assume (without loss of generality) that $p_+ > q_+$ and $q_+ > p_-$. We can now assume that for $w$ indices $i$, $p_+ > q_i$, $q_i > p_-$ and for $W - w$ indices $j$, $p_- < q_j < q_+$. where $w$ may range from $1$ to $W - 1$. This leads to seven equations

$$wp_+ + (W-w)p_- = 1, \quad g'(p_+) + \gamma - \alpha = 0 \quad (23)$$

which allow to express $p_-$, $q_-$, and $q_+$ in terms of $p_+$

$$p_- = \frac{1 - wp_+}{W - w}, \quad q_- = \frac{p_+ - \delta}{2W}, \quad q_+ = \frac{1 - wp_+}{W - w} + \frac{\delta}{2(W-w)}. \quad (24)$$
Further we get the equation
\[
g'(p_+) - g'(p_-) + g'(q_+) - g'(q_-) = 0.
\] (24)

Since \( g \) is concave \( g' \) is monotonically decreasing and
\( g'(p_+) - g'(q_-) > 0 \) and \( g'(q_+) - g'(p_-) > 0 \). Thus Eq. [24] has no solution, meaning that there is no extremum with \( p_+ \) and \( q_+ \) in \((0,1)\), and extrema are at the boundaries. The possibilities are \( p_+ = 1 \) or \( p_- = 0 \), then \( q_+ = 1 \) and \( q_- = 0 \). Only \( p_+ = 1 \) or \( p_- = 0 \) are compatible with the assumption that \( S[p] > S[q] \); \( p_+ = 1 \) is only a special case of \( p_- = 0 \) with \( n = 1 \). Since \( g(0) = 0 \) this immediately leads to the inequality
\[
\frac{|S_g[p] - S_g[q]|}{S_{\max}} \leq (1 - \phi) \left\{ \frac{\delta}{2(1 - \phi)} \right\} \left( \frac{1}{\phi} \right)^c - \phi \left( \frac{1 - \delta}{\phi} \right)^c \]
\[+ \phi \left| g\left( \frac{1}{\phi} \right) - g\left( \frac{1 - \delta}{\phi} \right) \right|,
\]
(25)

where \( \phi = w/W \) is chosen such that the right hand side of the equation is maximal. Obviously, for any finite \( W \) the right hand side can always be made as small as needed by choosing \( \delta > 0 \) small enough. Now take the limit \( W \to \infty \). If \( f \) is continuous
\[
\frac{|S_g[p] - S_g[q]|}{S_{\max}} \leq \delta^c + \left| 1 - \left( \frac{1 - \delta}{2} \right) \right| \leq \delta^c + \delta.
\]
(26)

Lesche-stability of \( S_g \) follows since the right hand side of Eq. [26] can be made smaller than any given \( \epsilon > 0 \) by choosing \( \delta > 0 \) small enough. This completes the first direction of the proof.

If, on the other hand, \( S_g \) is not Lesche-stable then there exists an \( \epsilon > 0 \), such that \( |S_g[p] - S_g[q]|/S_{\max} \geq \epsilon \), \( \forall N \), implying
\[
(1 - \phi) f\left( \frac{\delta}{2(1 - \phi)} \right) + \phi \left\{ f\left( \frac{1}{\phi} \right) - f\left( \frac{1 - \delta}{\phi} \right) \right\} \geq \epsilon.
\]
(27)
\[\forall \delta > 0. \]
This again means that either \( f(z) \) is discontinuous at \( z = 1/\phi \), or \( \lim_{z \to 0} f(z) > 0 \). Since \( g(0) = 0 \) implies that \( f(0) = 0 \), \( f(z) \) has to be discontinuous at \( z = 0 \).

Remark on higher order scaling exponents:

In principle the scheme of finding scaling exponents can be iterated to higher orders, i.e. to find a sequence of exponents \( (c, d, d_2, d_3, \ldots) \). To see this define
\[
\log^{(n+1)}(x) = \log(\log^{(n)}(x)), \quad \log^{(1)}(x) = \log(x) \text{ and } \exp^{(n+1)}(x) = \exp(\exp^{(n)}(x)), \quad \text{with } \exp^{(1)}(x) = \exp(x),
\]
then up to a precision \( M \)
\[
g(x) \sim x^c \prod_{m=1}^{M} \log^{(n)}\left( \frac{1}{x} \right),
\]
(28)
for small \( x \). The \( n \)th scaling exponent \( d_n \) can be determined by
\[
\lim_{x \to 0} \frac{g\left( \lambda_n(x)x \right)}{\lambda_n(x)x^{d_m}} \prod_{m=1}^{n-1} \left( \frac{\log^{(n)}\left( \frac{1}{\lambda_n(x)} \right)}{\log^{(m)}\left( \frac{1}{x} \right)} \right) = (1 + a_n)^{d_m},
\]
(29)
where the generalized scaling factor reads
\[
\lambda_n(x)^{-1} = x \exp^{(n)}\left( (1 + a_n) \log^{(n)}\left( \frac{1}{x} \right) \right).
\]
(30)
Note that this is a nested scheme of sequentially incorporating more information with higher levels, i.e. if \( g \in (c, d, d_2, \ldots, d_{n+1}) \), then also \( g \in (c, d, d_2, \ldots, y_n) \).

Using higher order asymptotic exponents makes it difficult to find a concave representative \( g_{c,d,d_2,\ldots,d_{\Omega}}(x) \) and \( g \) quickly becomes experimentally inaccessible. For macroscopic systems there may exist an upper limit \( M \) such that measuring \( d_m \) for \( m > M \) becomes senseless. For example, consider a system with \( N \sim 10^{23} \) particles and \( \Omega \propto e^N \) states. Then \( x \propto 1/\Omega \) and it does not make any sense to measure \( d_0 \) or higher because \( \log^{(6)}(x) \sim -1.13 \) is already negative, and \( \log^{(7)}(x) \) is no longer defined on the principal branch of the logarithm.
(a) $c=0.99999$, $r=0.0001$

(b) $d=0.025$, $r=0.9/(1-c)$

(c) $r=\exp(-d/2)/(1-c)$