When Shannon and Khinchin meet Shore and Johnson: Equivalence of information theory and statistical inference axiomatics

Petr Jizba [1,*] and Jan Korbel [2,3,1,†]

[1] *Faculty of Nuclear Sciences and Physical Engineering, Czech Technical University in Prague, Břehová 7, 115 19, Prague, Czech Republic*

[2] *Section for Science of Complex Systems, Medical University of Vienna, Spitalgasse 23, 1090 Vienna, Austria*

[3] *Complexity Science Hub Vienna, Josefstädter Strasse 39, 1080 Vienna, Austria*

(Received 27 September 2019; revised manuscript received 8 March 2020; accepted 27 March 2020; published 22 April 2020)

We propose a unified framework for both Shannon-Khinchin and Shore-Johnson axiomatic systems. We do it by rephrasing Shannon-Khinchine axioms in terms of generalized arithmetics of Kolmogorov and Nagumo. We prove that the two axiomatic schemes yield identical classes of entropic functionals—the Uffink class of entropies. This allows to re-establish the entropic parallelism between information theory and statistical inference that has seemed to be "broken" by the use of non-Shannonian entropies.

DOI: 10.1103/PhysRevE.101.042126

I. INTRODUCTION

Entropy is undoubtedly one of the most important concepts in physics, information theory, and statistics [1]. The notion of entropy was originally developed by Claussius, Boltzmann, Gibbs, Carathéodory and others in the context of statistical thermodynamics. There it supplemented a new state function that was naturally extensive (due to its very formulation in terms of the heat one-form) and in any adiabatically isolated system it represented a nondecreasing function of its state variables (on account of the Clausius theorem). Roughly a half-century after these developments, the entropy paradigm was further conceptualized in the theory of information by Shannon [2]. In this later context the ensuing entropy (Shannon's entropy or measure of information) quantitatively represented the minimal number of binary (yes/no) questions which brings us from our present state of knowledge about the system in question to the one of certainty. The higher is the measure of information (more questions to be asked) the higher is the ignorance about the system and thus more information will be uncovered after an actual measurement. A proper axiomatization of Shannon's entropy is encapsulated in the so-called Shannon-Khinchin (SK) axioms [3]. Only one decade after Shannon's seminal works, Jaynes [4,5] promoted Shannon's information measure to the level of inference functional that was able to extract least biased probability distributions from measured data. This procedure is better known as the *Maximum entropy principle* (MEP). Since MEP is, in its essence, a statistical inference method, it needs a proper mathematical qualification to place Jaynes' heuristic arguments in a sound mathematical framework. The corresponding mathematical qualification was provided by Shore and Johnson (SJ) in the form of axioms that ensure that the MEP estimation procedure is consistent with desired properties of inference methods [6,7]. At this point, one should emphasize that in the statistical inference theory (SIT) entropy functionals serve only as convenient technical vehicles for unbiased assignment of distributions that are compatible with given constraints. In fact, one might say that it is the MEP distribution that is the primary object in SIT while the entropy itself is merely secondary (not having any operational role in the scheme). This is very different from the information theory or thermodynamics where entropies are primary objects with firm operational meanings (given, e.g., in terms of coding theorems or calorimetric measurements). In the original paper [6,7] Shore and Johnson concluded that their axioms yield only one "measure of bias", namely Shannon entropy. It might, however, seem a bit puzzling why "measure of bias" should have anything to do with additivity (i.e., one of the defining properties of Shannon's entropy). In the end, any monotonic function of such a measure should provide the same MEP distribution but might (and as a rule it does) yield nonadditive entropy. So, it is perhaps not so surprising that with the advent of generalized entropies [8–16], the past two decades have seen a renewed interest both in the SJ axiomatics and the associated classes of admissible entropies [17–22]. In particular, it has been shown in Ref. [21] that the SJ axiomatization of the inference rule does account for substantially wider class of entropic functionals than just SE—the so-called Uffink class [22], which include Shannon's entropy as a special case.

The main aim of this paper is to answer the following question: what generalization of the SK axioms would provide the Uffink class of entropic functional? This would not only allow to re-establish the "broken" entropic parallelism between information theory and statistical inference but it should also cast a new light on the Uffink class of entropies and its practical utility.

*p.jizba@fjfi.cvut.cz

†jan.korbel@meduniwien.ac.at

We first recall the original set of SK axioms [3]:

Let A and B be two discrete random variables with respective sets of possible values $\mathcal{A} = \{A_i\}_{i=1}^n$ and $\mathcal{B} = \{B_i\}_{i=1}^m$. With $\mathcal{A}$ one can associate a complete set of events $\{a_i\}_{i=1}^n$ so that a_i denotes the event that $A = A_i$ (similarly for $\mathcal{B}$). Elements a_i (and b_j) are known as elementary events. Let

$$\begin{aligned}
&P(A = A_i) = P(A_i) = p_i, \quad P(B_j) = q_j\,,\\
&P(A = A_i, B = B_j) = P(A_i, B_j) = r_{ij}\,,\\
&P(A = A_i|B = B_j) = P(A_i|B_j) = r_{i|j} = r_{ij}/q_j\,,\\
&1 \leqslant i \leqslant m; \quad 1 \leqslant j \leqslant n\,,
\end{aligned}$$

be corresponding elementary-event, joint, and conditional probabilities, respectively. For A and B we denote the ensuing probability distributions as $P_A = \{p_i\}_{i=1}^n$ and $P_B = \{q_j\}_{j=1}^m$. Likewise, we write $P_{A,B} = \{r_{ij}\}_{i,j=1}^{n,m}$, $P_{A|B} = \{r_{i|j}\}_{i,j=1}^{n,m}$, and $P_{A|B_j} = \{r_{i|j}\}_{i=1}^n$. The entropy of the probability distribution P_A (which may also be called the entropy of A) will be, with a slight abuse of the notation, denoted interchangeably as $\mathcal{H}(P_A)$ or $\mathcal{H}(A)$. Similar notation will be introduced for distributions P_B, $P_{A,B}$, $P_{A|B}$, and $P_{A|B_j}$.

SK1 *Continuity:* Entropy is a continuous function with respect to all its arguments, i.e., $\mathcal{H}(P) \in \mathcal{C}$.

SK2 *Maximality:* Entropy is maximal for uniform distribution, i.e., $\max_P \mathcal{H}(P) = \mathcal{H}(U_n)$, where $U_n = \{1/n, \ldots, 1/n\}$.

SK3 *Expandability:* Adding an elementary event with probability zero does not change the entropy, i.e.,

$$\mathcal{H}(p_1, \ldots, p_n, 0) = \mathcal{H}(p_1, \ldots, p_n).$$

SK4 ^{S} *Shannon additivity:*

$$\mathcal{H}(A, B) = \mathcal{H}(A|B) + \mathcal{H}(B) = \mathcal{H}(B|A) + \mathcal{H}(A),$$

where $\mathcal{H}(B|A) = \sum_i p_i \mathcal{H}(B|A = A_i)$.

We note that the conditional entropy $\mathcal{H}(B|A)$ can be calculated in two ways: *i)* from the entropy of the joint distribution of the pair (A, B) and marginal distribution of A, or *ii)* from the marginal distribution A and entropy of the conditional random variable $B|A = A_i$. This duality is crucial for the internal consistency the SK axiomatic scheme. The aforestated set of SK axioms has the unique solution—Shannon's entropy [1]

$$\mathcal{H}(P) = -\sum_i p_i \log p_i\,.$$

With the advent of generalized entropies [8–16] there arose two natural questions. First, is it possible to conceptualize such entropies in terms of information-theoretic axioms (*à la* SK axioms)? And second, can generalized entropies be used as consistent inference functionals with sound mathematical underpinning (*à la* SJ axioms)? As for the first question, it is well known that one can "judiciously" generalize the additivity axiom **SK4**^{S} to produce various generalized entropies. Typical examples are provided by Rényi and Tsallis-Havrda-Charvát (THC) entropies. For instance, for the Rényi entropy, one keeps axioms **SK1-3** and substitute **SK4**^{S} with [8]

SK4^R *Rényi additivity:* $\mathcal{R}_q(A, B) = \mathcal{R}_q(A|B) + \mathcal{R}_q(B) = \mathcal{R}_q(B|A) + \mathcal{R}_q(A)$, where $\mathcal{R}_q(B|A) = f^{-1}(\sum_i \rho_i^A(q) f(\mathcal{R}_q(B|A = A_i)))$.

Here, $\rho_i^A(q) = (p_i)^q / \sum_j (p_j)^q$ is the escort (or zooming) distribution [23,24] and f is an arbitrary invertible and positive function on $[0, \infty)$. Corresponding axiomatics is stringent enough to fix uniquely $f(x)$ to be either $f(x) = e^{(1-q)x}$ (for $q \neq 1$) or $f(x) = x$ (for $q = 1$), and yields the Rényi entropy

$$\mathcal{R}_q(P) = \frac{\log \sum_i p_i^q}{1 - q}\,,$$

as the unique solution.

Similarly, for the case of nonadditive THC entropy [9,10] one can augment axioms **SK1-3** with [25,26]

SK4^T *Tsallis additivity:* $\mathcal{S}_q(A, B) = \mathcal{S}_q(B|A) + \mathcal{S}_q(A) + (1 - q)\mathcal{S}_q(B|A)\mathcal{S}_q(A)$ where $\mathcal{S}_q(B|A) = \sum_i \rho_i^A(q) \mathcal{S}_q(B|A = A_i)$,

where $\rho_i^A(q)$ is again the escort distribution. The unique solution of this axiomatic system gives the THC entropy

$$\mathcal{S}_q(P) = \frac{\sum_i p_i^q - 1}{1 - q}\,.$$

In parallel with this there has been several successful attempts to classify entropic functionals according to various desirable information-theoretic properties. Here, we should mention, e.g., the class of strongly pseudoadditive entropies (SPA) based on generalization of Rényi entropy axioms for nonadditive entropies [27], Z-entropies based on group properties of the entropic functionals [28], or classification according to the asymptotic scaling leading to (c, d)-entropies [13] and ensuing generalizations [29].

As for the second question, there has been notable progress in recent years in the classification of entropic functionals satisfying SJ axioms [21,22,30]. Our aim here is to employ generic arithmetical principles to generalize, in a logically sound way, the SK axiomatic scheme. To this end we will use the framework of Kolmogorov-Nagumo (KN) arithmetics [31,32], KN quasiarithmetic means [33–35], and escort distributions [23,24]. Ensuing class of admissible entropies will be compared with the class of entropies solving SJ axioms—Uffink class. We will see that both classes not only coincide, and hence bolster the entropic parallelism between information theory and statistical inference, but there also is a close parallelism between the two axiomatic schemes.

The rest of the paper is organized as follows. In Sec. II, we briefly summarize the concept of generalized arithmetics and outline the key role that Kolmogorov-Nagumo functions play in this context. In Sec. III, we introduce the class of Shannon-Khinchin axioms based on the Kolmogorov-Nagumo generalized arithmetics and derive the generic class of entropic functionals satisfying these axioms. In Sec. IV, we show the equivalence of the aforementioned class and the Uffink's entropic class. This will, in turn, cast new light on the relationship between SK and SJ axiomatic schemes. This will in turn re-establish the entropic parallelism between information theory and statistical inference. The last section is devoted to some further observations, remarks, and conclusions.

[1]Here and throughout we use the base of natural logarithms. Entropy thus defined is then measured in natural units — *nats*, rather than *bits*. To convert, note that 1 bit = 0.693 nats.

II. GENERALIZED ARITHMETICS AND KOLMOGOROV AND NAGUMO FUNCTIONS

Let us now introduce the concept of generalized arithmetics. From abstract algebra it is known that arithmetic operations can be defined in various ways, even if one assumes commutativity and associativity of addition and multiplication, and distributivity of multiplication with respect to addition [31,32]. In consequence, whenever one encounters "plus" or "times" one has certain flexibility in interpreting these operations. A change of realization of arithmetic, without altering the remaining structures of equations involved, plays an analogous role as a symmetry transformation in physics.

Let us considering a bijection $f^{-1} : M \mapsto N \subset \mathbb{R}$, where M is some set. The map f allows us to define addition, subtraction, multiplication, and division in M, as follows:

$$\begin{aligned} x \oplus y &= f(f^{-1}(x) + f^{-1}(y)), \\ x \ominus y &= f(f^{-1}(x) - f^{-1}(y)), \\ x \otimes y &= f(f^{-1}(x) f^{-1}(y)), \\ x \oslash y &= f(f^{-1}(x)/f^{-1}(y)). \end{aligned} \tag{1}$$

One can readily verify the following standard properties: (1) associativity $(x \oplus y) \oplus z = x \oplus (y \oplus z)$, $(x \otimes y) \otimes z = x \otimes (y \otimes z)$, (2) commutativity $x \oplus y = y \oplus x$, $x \otimes y = y \otimes x$, (3) distributivity $(x \oplus y) \otimes z = (x \otimes z) \oplus (y \otimes z)$. For a future convenience we will explicitly affiliate with the arithmetic operations $\oplus$, $\ominus$, $\otimes$, and $\oslash$ the symbol of the function f, so for instance, we will write $\oplus_f$ instead of $\oplus$, etc.

This generalized arithmetical structure motivated Kolmogorov and Nagumo [33,34] to formulate the most general class of means, so-called quasilinear means, that are fully compatible with the usual Kolmogorov postulates of probability theory [36], with interesting applications in thermostatistics [35].

The aforementioned generalized arithmetics can be extended quite naturally to real multivariate functions. For instance, for a function of two variables $\mathcal{G}(x, y)$ it can be defined as

$$\mathcal{G}_f(x, y) \equiv f(\mathcal{G}(f^{-1}(x), f^{-1}(y)).$$

Let us state in this connection a couple of important consequences that can be easily verified:

(i) when $z = x \otimes_f y$, then $g(z) = g(x) \otimes_{g \cdot f} g(y)$,

(ii) $x \oplus_f y = x \otimes_{f \cdot \log} y$. Here, by $f \cdot g$ we implicitly mean the composition of two functions. Particularly important for our purposes will be the so-called q-deformed algebra where

$$f(x) \equiv f_q(x) = \log_q(x) = \frac{(x^{1-q} - 1)}{(1-q)}.$$

Ensuing operation $\otimes_{f_q}$ is traditionally denoted as q-addition and the notation $\oplus_q$ is often used instead.

(iii) For the generalized product $\otimes_f$ the function f is not determined uniquely. In fact, there exists a two-parametric class of functions $f_{a,b}$, so that $f(x) \mapsto f_{a,b}(x) = f(ax^b)$, which yield the same product. Indeed,

$$\begin{aligned} x \otimes_{f_{a,b}} y &= f(a[(f^{-1}(x)/a)^{1/b}(f^{-1}(y)/a)^{1/b}]^b) \\ &= x \otimes_f y. \end{aligned} \tag{2}$$

This result will be particularly important in Sec. III.

III. KOLMOGOROV-NAGUMO GENERALIZATION OF SHANNON-KHINCHIN AXIOMS

Let us now generalize the Shannon-Khinchin (SK) entropic axioms in terms of the Kolmogorov-Nagumo arithmetics in the following way:

SK1 *Continuity:* Entropy is a continuous function with respect to all its arguments, i.e., $S(P) \in \mathcal{C}$.

SK2 *Maximality:* Entropy is maximal for uniform distribution, i.e., $\max_P S(P) = S(U_n)$, where $U_n = \{1/n, \ldots, 1/n\}$.

SK3 *Expandability:* Adding an elementary event with probability zero does not change the entropy, i.e.,

$$S(p_1, \ldots, p_n, 0) = S(p_1, \ldots, p_n).$$

SK4 *Composability:* Joint entropy for a pair (A, B) of random variables can be expressed as

$$S(A, B) = S(A|B) \otimes_f S(B),$$

where $S(A|B)$ is conditional entropy satisfying consistency requirements **I), II)** (see below).

In passing, we can observe from the two illustrative axiomatic schemes $\mathbf{SK4}^R$ and $\mathbf{SK4}^T$ that viable entropic functionals should obey two natural conditions:

I) For two independent random variables A and B the joint entropy $S(A, B)$ should be *composable* from entropies $S(A)$ and $S(B)$, i.e., $S(A, B) = F(S(A), S(B))$.

II) Conditional entropy should be *decomposable* into entropies of conditional distributions, i.e., $S(B|A) = G(P_A, \{S(B|A = A_i)\}_{i=1}^n)$.

Here, F and G are functionals to be determined shortly. The motivation for these two conditions is taken from the original SK axioms for Shannon, Rényi, and Tsallis entropy. They all are *composable* from marginal entropies if the subsystems are independent and they all are *decomposable* into conditional entropies and (escort) marginal distributions.

Let us also note that the conditional entropy $S(A|B)$ automatically fulfills several important properties:

(a) *Entropic Bayes' rule:* $S(A|B) = S(B|A) \oslash_f S(B) \otimes_f S(A)$,

(b) *Generalized Gibbs inequality:* $S(A|B) \leqslant S(A)$.

The Bayes rule is easy to show from the interchangeability of $S(A, B) = S(B, A)$ and by using the definition of conditional entropy. The second law of thermodynamics is easy to show because $S(A, B) \oslash_f S(B) \leqslant S(A)$.

Moreover, we can define the mutual information as

$$I(A, B) = S(A, B) \oslash_f \big(S(B) \otimes_f S(A)\big).$$

The composition requirement **I)** is equivalent to $I(A, B) = f(1)$ for independent random variables. We might note that the requirement **I)** is equivalent to *strict composability axiom* introduced in Ref. [28].

Let us now prove the following theorem:

Theorem 1. The most general class of entropic functionals S satisfying the aforestated axioms **SK1-4** can be expressed as

$$S_q^f(P) = f\left[\left(\sum_i p_i^q\right)^{1/(1-q)}\right], \tag{3}$$

where $f(x)$ is a generic strictly increasing function defined on $x \in [0, \infty)$.

In passing it is useful to note that Eq. (3) can be equivalently expressed as

$$S_q^f(P) = f\left[\exp_q\left(\sum_i p_i \log_q\left(\frac{1}{p_i}\right)\right)\right], \tag{4}$$

where $\exp_q(x) = [1 + (1-q)x]^{1/(1-q)}$. Equation (4) is a simple consequence of the fact that

$$\begin{aligned}\exp_q \sum_i p_i \log_q\left(\frac{1}{p_i}\right) &= \left(1 + (1-q)\sum_i p_i \frac{p_i^{1-q} - 1}{1-q}\right)^{1/(1-q)} \\ &= \left(\sum_i p_i^q\right)^{1/(1-q)}.\end{aligned}$$

Proof of Theorem 1: First, we see that the functional has to be symmetric in all components of $P = \{p_i\}$. This is because by relabeling points in a set of elementary events should not change the information about the underlying stochastic process. Consequently, S must be symmetric. Second, the entropy of the uninform distribution $S(n) \equiv S(1/n, \ldots, 1/n)$ can be obtained from composability axiom. To this end we denote the random variable with uniform distribution as $U_{nm} = U_n U_m$. We abbreviate $S(U_n)$ as $S(n)$. Then [see Eq. (1)]

$$S(nm) = S(n) \otimes_f S(m) \Rightarrow S(n) = f(n).$$

Here, we have explored the freedom in the definition of the function f [see Eq. (2)] and scaled back the generic solution $S(n) = f(n^x)$, $x \in \mathbb{R}$ to the solution $S(n) = f(n)$.

Third, let us take two random variables A and B with distributions $P_A = \{p_i\}_{i=1}^n$ and $P_B = \{q_j = 1/m\}_{j=1}^m$. Let us also introduce the so-called Daróczy mapping [27,37], i.e., $S \mapsto f^{-1}S$. After this mapping we get multiplicative entropy. From the definition of $S(A|B)$ we then obtain that

$$m f^{-1}S(p_1/m, \ldots, p_n/m) = f^{-1}S(p_1, \ldots, p_n), \tag{5}$$

since the conditional entropy is for each random variable just the usual unconditional one. Therefore, entropy must be a first order homogenous, symmetric function. According to [38] the solution of homogeneous equation (5) can be (under mildly restrictive assumptions) expressed as

$$f^{-1}S(x_1, \ldots, x_n) = b\prod_{i=1}^n x_i^{a_i} \quad \text{where} \quad \sum_i a_i = 1. \tag{6}$$

Here, a_i and b are constants to be specified later. However, this solution is not symmetric in its variables. This can be achieved by symmetrization of Eq. (6) that can be then rewritten in the following form:

$$f^{-1}S(p_1, \ldots, p_n) = b \sum_{\{j_1,\ldots,j_n\}\in\sigma(n)} \prod_{i=1}^n p_i^{a_{j_k}}.$$

This expression can be equivalently recast to

$$f^{-1}S(p_1, \ldots, p_n) = b\prod_{i=1}^n \left(\sum_{k_i} p_{k_i}^{a_i}\right),$$

that can further be rewritten as

$$f^{-1}S(p_1, \ldots, p_n) = b\prod_{i=1}^n \left(\sum_{k_i} p_{k_i}^{a_i}\right)^{c/(1-a_i)}, \tag{7}$$

which still keeps the entropy to be a homogeneous function of the first order. The parameter c is a free parameter that will be determined later. Note that this representation is also mentioned in [39].

Let us now show that in order to fulfill the decomposability axiom **II**), only one a_j must be nonzero. To this end, we explicitly express $f^{-1}S(A|B)$ as

$$\begin{aligned}f^{-1}S(A|B) = &\frac{\left(\sum_{k_1,l_1}(r_{k_1|l_1}q_{l_1})^{a_1}\right)^{c/(1-a_1)}}{\left(\sum_{l_1} q_{l_1}^{a_1}\right)^{c/(1-a_1)}} \\ &\times \cdots \times \frac{\left(\sum_{k_n,l_n} r_{k_n|l_n}q_{l_n}^{a_n}\right)^{c/(1-a_n)}}{\left(\sum_{l_n} q_{l_n}^{a_n}\right)^{c/(1-a_n)}}.\end{aligned}$$

This can be more explicitly rewritten as

$$\begin{aligned}f^{-1}S(A|B) = &\left(\sum_{l_1} \rho_{l_1}^B(a_1) \sum_{k_1}(r_{k_1|l_1})^{a_1}\right)^{c/(1-a_1)} \\ &\times \cdots \times \left(\sum_{l_n} \rho_{l_n}^B(a_n) \sum_{k_n}(r_{k_n|l_n})^{a_n}\right)^{c/(1-a_n)},\end{aligned}$$

where $\rho_l^B(a) = q_l^a / \sum_l q_l^a$ is the escort distribution [23,24]. This expression is an unconditional entropy of the conditional distribution only if one of a_j is nonzero and the remaining ones are zero. With this we get that

$$\begin{aligned}f^{-1}S(A|B) &= \left(\sum_l \rho_l^B(a) \sum_k (r_{k|l})^a\right)^{1/(1-a)} \\ &= \left\{\sum_l \rho_l^B(a)[S(A|B = b_l)]^{1-a}\right\}^{1/(1-a)},\end{aligned}$$

which directly implies the decomposability function G. With this result Eq. (7) boils down to

$$\begin{aligned}f^{-1}S(p_1, \ldots, p_n) &= b\left(\sum_k p_k^a\right)^{c/(1-a)} \\ &= b\left[\exp_a\left(\sum_k p_k \log_a \frac{1}{p_k}\right)\right]^c,\end{aligned}$$

which by Eq. (2) is equivalent to Eq. (3) provided we identify a with q.

The function f must be must be strictly monotonic because in the proof we needed inverse of f, and must be strictly increasing because S has by **SK2** the maximum for uniform distribution [and not, for instance, for $P = (1, 0, 0, , 0)$]. This concludes the proof. ■

Let us finally note that the original axiom **SK4**^{S} is recovered from **SK4** by taking $f(x) = \ln x$ and the decomposability function $G(x_i, y_i) = \sum_i x_i y_i$.

IV. EQUIVALENCE WITH SHORE-JOHNSON AXIOMS

Let us now turn our attention to MEP and corresponding consistency requirements. The MEP can be formulated in the following way [4,5]:

Proposition (Maximum entropy principle). Given the set of linear constraints $\sum_i p_i E_i^{(k)} = \langle E^{(k)} \rangle$, the least biased estimate of the underlying distribution $P = \{p_i\}$ is obtained from maximization of the entropic functional $S(P)$ under normalization constraint and set of constraints $\langle E^{(k)} \rangle$, i.e., by maximizing the Lagrange functional

$$S(P) - \alpha \sum_{i=1}^{N} p_i - \sum_{k=1}^{\nu} \beta^{(k)} \sum_{i=1}^{N} p_i E_i^{(k)} . \tag{8}$$

In Eq. (8) the index "i" runs over all possible *states*, i.e., over all elements from the set of possible outcomes associated with a given random system.

Shore and Johnson formulated the set of consistency requirements that the MEP should satisfy [6,7]:

SJ1 *Uniqueness*: the result should be unique.

SJ2 *Permutation invariance*: the permutation of states should not matter.

SJ3 *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.

SJ4 *System independence*: It should not matter whether one accounts for independent constraints related to independent systems separately in terms of marginal distributions or in terms of full system.

SJ5 *Maximality*: In absence of any prior information, the uniform distribution should be the solution.

Let us now state without proof the theorem that provides the most general class of admissible entropic functionals consistent with the aforestated **SJ** axioms:

Theorem 2 (Uffink theorem). The class of entropic functionals S satisfying the axioms **SJ1-5** can be expressed as

$$S_q^f(P) = f\left[\left(\sum_i p_i^q\right)^{1/(1-q)}\right], \tag{9}$$

for any $q > 0$ and for any strictly increasing function f.

In particular, the Uffink theorem shows that members of this entropic class admit representation in the form given by Eq. (3), and hence the SK and SJ axiomatic systems are equivalent. A detailed proof of Theorem 2 can be found in Ref. [21].

Let us now discuss some salient results of the proof. First two axioms assert that the entropic functional must be a symmetric functional in the probability components. The third axiom determines the function in the sum form, i.e., in the form $S(P) = f(\sum_k g(p_k))$, with g being an arbitrary increasing concave function. The fourth axiom gives us the final form of the entropic functional (without specifying the range for q's), and finally the fifth axiom guaranties that $q > 0$. Note that the class obtained from Theorem 1 and epitomized by Eq. (3) is the same as the class given by Eq. (9) from the Uffink theorem. Therefore, we immediately see that in axiom **SK4** the requirement **II)** (decomposability) corresponds to axiom **SJ3**, while requirement **I)** (composability) corresponds to axiom **SJ4**. Moreover, the interpretation of f and q is now clear. The function f determines the scaling of the entropy for uniform distribution (as it is independent of q), see also [29], while the parameter q determines the correlations in the system through MaxEnt distribution, which can be expressed as (see Ref. [21])

$$p_i = \frac{1}{Z_q} \exp_q(-\beta \Delta E_i),$$
$$Z_q = \sum_i [\exp_q(-\beta \Delta E_i)],$$
$$\beta = \frac{\beta}{q f'(Z_q) Z_q},$$

where $\Delta E_i = E_i - \langle E \rangle$. The connection of the q parameter with correlations can be understood from the MaxEnt distribution of a joint system composed from two disjoint subsystems. Let us denote the MaxEnt distribution of the joint system as p_{ij} and the MaxEnt distribution of the subsystems as u_i and v_j. In [21] it was shown that the MaxEnt distributions involved fulfill the composition rule that can be formulated as

$$\frac{1}{p_{ij}\,\mathcal{U}_q(P)} = \frac{1}{u_i\,\mathcal{U}_q(U)} \otimes_q \frac{1}{v_j\,\mathcal{U}_q(V)}, \tag{10}$$

where $x \otimes_q y = [x^{1-q} + y^{1-q} - 1]_+^{1/(1-q)}$ (with $x, y > 0$) is the so-called q-product [40], and $\mathcal{U}_q(P) = (\sum p_{ij}^q)^{1/(1-q)}$, and similarly for U and V. For $q \to 1$, Eq. (10) reduces to $p_{ij} = u_i v_j$. The reverse is true as well. By re-expressing Eq. (10) in terms of escort distributions $P_{ij}(q)$, $U_i(q)$, and $V_j(q)$ (associated with p_{ij}, u_i, and v_j, respectively) as

$$\frac{P_{ij}(q)}{p_{ij}} = \frac{U_i(q)}{u_i} + \frac{V_j(q)}{v_j} - 1, \tag{11}$$

and using $p_{ij} = u_i v_j$, we obtain $U_i(q) = u_i$, $V_k(q) = v_k$ (for all i, k). Latter have a unique solution [23] $q = 1$. This implies that q parametrizes correlations between system's subsystems since only for $q = 1$ the Pearson correlation coefficient is zero.

As discussed, e.g., in [30], a monotonic function of an entropic functional gives the same MEP distribution and redefines only the Lagrange multipliers but does not change the actual form of the distribution. This can be interpreted as a sort of *gauge invariance* $S(P) \mapsto f(S(P))$. Finally, let us mention that the $q = 1$ case corresponds to uncorrelated MEP distributions for disjoint systems, for which we get a stronger version of system independence axiom [21]:

SJ4^{SSI} *Strong system independence:* Whenever two subsystems of a system are disjoint, we can treat the subsystems in terms of independent distributions.

The solution is then

$$S_1^f(P) = f\left(\exp\left[\sum_i p_i \log(1/p_i)\right]\right),$$

which is equivalent (through Daróczy mapping) to Shannon entropy—as expected. In this case, the composition rules in Eqs. (10) and (11) reduce to the composition rule of independent systems, i.e.,

$$p_{ij} = u_i v_j .$$

On the other hand, if we require that the entropy must be in the trace form [13,28], i.e., $S(P) = \sum_i g(p_i)$, then we get that $f(x) = \log_q(x)$ and we end up with the class of THC entropies

$$S_q^{\log_q}(P) = \sum_i p_i \log_q(1/p_i) .$$

V. CONCLUSIONS

Here we have reformulated Shannon-Khinchin axioms of information theory in terms of generalized arithmetics of Kolmogorov and Nagumo. Apart from the axiomatic structure itself, the novelty of this work is in showing that the general class of entropic functional satisfying such SK axioms is identical with the Uffink class of entropies. Since the Uffink class is known to represent the general solution of Shore-Johnson axioms of statistical-inference theory, both axiomatic systems have to be equivalent. We have shown that Uffink functionals S_q^f are characterized by the Kolmogorov-Nagumo function f and a positive parameter q, where f determines a scaling behavior of entropy for uniform distributions and q quantifies correlations of MEP distributions for disjoint subsystems. In passing, we can note that the form (4) of the class S_q^f can also be found in the literature under the name *strongly pseudoadditive* (SPA) entropies [41] or Z-entropies [28].

The outlined entropic parallelism between information theory and statistical inference can serve as a good starting point for further research. In this context it would be particularly interesting to investigate how robust the aforementioned equivalence between the two axiomatic systems is and assess the extent and consequences resulting from a prospective breakdown. One might instigate such a breakdown by working, e.g., with more general constraints (nonlinear constraints or scalings as in noninductive inference) or by relaxing some of the presented axioms. In fact, it is well known that many complex systems do not satisfy SK axioms, not even in our generalized sense [13,29,42]. By relaxing some of these axioms, one might gain further maneuvering space allowing to accommodate entropies of such systems as path-dependent and super-exponential systems or complex systems with nontrivial constraints.

ACKNOWLEDGMENTS

P.J. and J.K. were supported by the Czech Science Foundation (GAČR), Grant No. 19-16066S. J.K. was also supported by the Austrian Science Foundation (FWF) under Project no. I3073.

[1] S. Thurner, B. Corominas-Murtra, and R. Hanel, Phys. Rev. E **96**, 032124 (2017).
[2] C. Shannon, Bell Syst. Tech. J. **27**, 379 (1948); **27**, 623 (1948).
[3] A. I. Khinchin, *Mathematical Foundations of Information Theory* (Dover Publications, New York, 1957).
[4] E. T. Jaynes, Phys. Rev. **106**, 620 (1957).
[5] E. T. Jaynes, Phys. Rev. **108**, 171 (1957).
[6] J. E. Shore and R. W. Johnson, IEEE Trans. Inf. Theor. **26**, 26 (1980).
[7] J. E. Shore and R. W. Johnson, IEEE Trans. Inf. Theor. **27**, 472 (1981).
[8] P. Jizba and T. Arimitsu, Ann. Phys. (NY) **312**, 17 (2004).
[9] C. Tsallis, J. Stat. Phys. **52**, 479 (1988).
[10] J. Havrda and F. Charvát, Kybernetika **3**, 30 (1967).
[11] G. Kaniadakis, Physica A **365**, 17 (2006).
[12] B. D. Sharma, J. Mitter, and M. Mohan, Inf. Control **39**, 323 (1978).
[13] R. Hanel and S. Thurner, Europhys. Lett. **93**, 20006 (2011).
[14] P. Jizba and J. Korbel, Physica A **444**, 808 (2016).
[15] T. Wada and H. Suyari, Phys. Lett. A **368**, 199 (2007).
[16] P. Jizba and T. Arimitsu, Physica A **340**, 110 (2004).
[17] S. Pressé, K. Ghosh, J. Lee, and K. A. Dill, Phys. Rev. Lett. **111**, 180604 (2013).
[18] C. Tsallis, Entropy **17**, 2853 (2015).
[19] S. Pressé, K. Ghosh, J. Lee, and K. A. Dill, Entropy **17**, 5043 (2015).
[20] T. Oikonomou and G. B. Bagci, Phys. Rev. E. **99**, 032134 (2019).
[21] P. Jizba and J. Korbel, Phys. Rev. Lett. **122**, 120601 (2019).
[22] J. Uffink, Stud. Hist. Phil. Mod. Phys. **26**, 223 (1995).
[23] C. Beck and F. Schlögl, *Thermodynamics of Chaotic Systems: An Introduction* (Cambridge University Press, Cambridge, 2008).
[24] C. Tsallis, R. S. Mendes, and A. R. Plastino, Physica A **261**, 534 (1998).
[25] S. Abe, Phys. Lett. A **271**, 74 (2000).
[26] H. Suyari, IEEE Trans. Inf. Theor. **50**, 1783 (2004).
[27] V. M. Ilić and M. S. Stanković, Physica A **411**, 138 (2014).
[28] P. Tempesta, Proc. Royal Soc. A **472**, 20160143 (2016).
[29] J. Korbel, R. Hanel, and S. Thurner, New J. Phys. **20**, 093007 (2018).
[30] P. Jizba and J. Korbel, Phys. Rev. E. **100**, 026101 (2019).
[31] M. Czachor, Quantum Stud.: Math. Found. **3**, 123 (2016).
[32] J. Naudts, *Generalised Thermostatistics* (Springer, London, 2011).
[33] A. Kolmogorov, Atti della R. Accademia Nazionale dei Lincei **12**, 388 (1930).
[34] M. Nagumo, Japan. J. Math. **7**, 71 (1930).

[35] M. Czachor and J. Naudts, Phys. Lett. A **298**, 369 (2002).
[36] A. N. Kolmogorov, *Foundations of Probability* (Chelsea Publishing Company, New York, 1950).
[37] P. Jizba and J. Korbel, Entropy **19**, 605 (2017).
[38] J. Aczél and J. Dhombres, *Functional Equations in Several Variables* (Cambridge University Press, Cambridge, 1989).
[39] M. A. Rodríguez, A. Romaniega, and P. Tempesta, Proc. R. Soc. A **475**, 20180633 (2019).
[40] E. P. Borges, Physica A **340**, 95 (2004).
[41] V. Ilić, A. M. Scafone, and T. Wada, Phys. Rev. E **100**, 062135 (2019).
[42] B. Corominas-Murtra, R. Hanel, and S. Thurner, Proc. Nat. Akad. Sci. USA **112**, 5348 (2015).