Maximum Entropy Principle in Statistical Inference: Case for Non-Shannonian Entropies

Petr Jizba[1,*] and Jan Korbel[2,3,1,†]
[1]FNSPE, Czech Technical University in Prague, Břehová 7, 115 19, Prague, Czech Republic
[2]Section for Science of Complex Systems, CeMSIIS, Medical University of Vienna, Spitalgasse 23, A-1090 Vienna, Austria
[3]Complexity Science Hub Vienna, Josefstädter Strasse 39, 1080 Vienna, Austria

(Received 19 August 2018; revised manuscript received 12 January 2019; published 26 March 2019)

In this Letter, we show that the Shore-Johnson axioms for the maximum entropy principle in statistical estimation theory account for a considerably wider class of entropic functional than previously thought. Apart from a formal side of the proof where a one-parameter class of admissible entropies is identified, we substantiate our point by analyzing the effect of weak correlations and by discussing two pertinent examples: two-qubit quantum system and transverse-momentum behavior of hadrons in high-energy proton-proton collisions.

DOI: 10.1103/PhysRevLett.122.120601

The concept of entropy indisputably plays a pivotal role in modern physics [1,2], statistics [3–5], and information theory [6,7]. In each of these fields the entropy paradigm has been formulated independently and with different applications in mind. While, in physics, the entropy quantifies the number of distinct microstates compatible with a given macrostate, in statistics, it corresponds to the inference functional for an updating procedure, and in information theory, it determines a limit on the shortest attainable encoding scheme.

However, recent developments in high-energy physics [8,9] and complex dynamical systems in particular [10–13] have brought about the need for a further extension of the concept of entropy beyond the conventional Shannon-Gibbs paradigm. Consequently, numerous generalizations proliferate in the current literature ranging from the additive entropies of Rényi [14] and Burg [15] through the rich class of nonadditive entropies [16–21] to more exotic types of entropies [22]. There are also parallel efforts underway to classify all feasible entropic functionals according to their group properties [23], generalized additivity rules [24] or asymptotic scaling [13,25].

Regardless of a particular generalization, the key usage of entropy is in statistical estimation theory which, in turn, crucially hinges on the maximum entropy (ME) principle (MEP) and its various reincarnations (e.g., principle of maximum caliber, principle of minimum cross-entropy, or minimum Akaike information criterion). The MEP can be formulated as follows [7,26,27].

Theorem 1 (MEP).—Given the set of constraints $C = \{I_k\}_{k=1}^{\nu}$, the best estimate of the underlying (i.e., true) probability distribution $P = \{p_i\}_{i=1}^{n}$ is the one that maximizes the entropy functional $S(P)$ subject to the constraints; i.e., it maximizes the Lagrange functional

$$S(P) - \sum_{k=1}^{\nu} \lambda_k I_k. \tag{1}$$

In the case of inductive inference, the constraints, or prior information, are given in terms of linear expectation values; i.e., the constraints considered are of the form

$$I_k \equiv \langle \mathcal{I}_k \rangle = \sum_i \mathcal{I}_{k,i} p_i, \tag{2}$$

where $\{\mathcal{I}_{k,i}\}$ are possible realizations (alphabet) of the observable $\mathcal{I}_k$. Other types of constraints, such as escort means, quasilinear means, or noninductive prior information, such as the Lipshitz-Hölder exponent of probability distributions, are not considered at this stage.

MEP was pioneered by Jaynes who first employed Shannon's entropy (SE) functional in the framework of equilibrium statistical physics [26,27]. Since then, MEP has rapidly become a powerful tool, e.g., in nonequilibrium statistical physics, astronomy, geophysics, biology, medical diagnosis, or economics [11,28].

The rationale behind the MEP is typically twofold: first, maximizing entropy minimizes the amount of prior information built into the distribution (i.e., ME distribution is maximally noncommittal with regard to missing information); second, many physical systems tend to move towards (or concentrate extremely close to) ME configurations over time [1,2,11,26].

As successful as Shannon's information theory has been, it is clear by now that it is capable of dealing with only a limited class of systems. In fact, only recently, it has become apparent that there are many situations of practical interest requiring more "exotic" statistics which does not conform with the canonical prescription of the classical ME (known as Boltzmann-Gibbs statistics) [11]. However, it cannot be denied that the ME approach deals with statistical systems in a way that is methodically appealing, physically plausible, and intrinsically nonspeculative (ME invokes no hypotheses beyond the evidence that is in available data).

Thus, one might be tempted to extend MEP also to other entropy functionals, particularly when the ensuing ME distributions differ from Boltzmann-Gibbs ones in some desirable way (e.g., they have specific heavy tails). Entropy functionals in question should not be arbitrary, they ought to satisfy some "reasonable" properties. For instance, in information theory, such properties are typified by coding theorems [29,30] or axiomatic rules (à la Shannon-Kchinchine type of axioms [14,31]). Recently, however, doubts have been raised about feasibility of this program. Arguments involved rest primarily on Shore-Johnson (SJ) axioms of statistical estimation theory [32,33].

Shore-Johnson axioms.—From the point of statistics, MEP is an estimation method, approximating probability distribution from the limited prior information. As such, it should obey some consistency rules. SJ introduced a set of axioms, which ensure that the MEP estimation procedure is consistent with desired properties of inference methods. These axioms are [4]

Axiom 1: Uniqueness.—The result should be unique.

Axiom 2: Permutation invariance.—The permutation of states should not matter.

Axiom 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.

Axiom 4: 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 constraints and joint distribution.

Axiom 5: Maximality.—In absence of any prior information, the uniform distribution should be the solution.

For the sake of simplicity, we focus on a discrete version of SJ axioms only. The corresponding generalization of SJ axioms to continuous distributions was done, e.g., in Refs. [4,34], and it is easy to verify that results obtained here are (with minor adjustments) valid also in continuous-state spaces.

In recent years, there has been much discussion of the consistency of MEP for generalized, i.e., non-Shannonian entropies. A typical claim has been that the SJ axioms preclude the use of MEP for generalized entropies, since these introduce an extra bias in the estimation of the ensuing ME distributions [32,33,35]. Here, we show that the SJ axioms, as they stand, certainly allow for a wider class of entropic functional than just SE. Central to this is the following theorem due to Uffink [34].

Theorem 2 (Uffink theorem).—MEP satisfies Shore-Johnson consistency axioms if and only if the following prescription holds: Maximize $\mathcal{U}_q(P)$ under the set of constraints $C = \{I_k\}$, where $\mathcal{U}_q(P) = (\sum_{i=1}^n p_i^q)^{1/(1-q)}$ for any $q > 0$, modulo equivalency condition.

The equivalency condition means that all functionals $f[\mathcal{U}_q(P)]$ for strictly increasing functions f are equivalent ($\sim$) in the sense that they provide the same ME distribution [34]. A simple variant of the proof together with related discussion is provided in the Supplemental Material [36]. Here, we list some pertinent results. (a) From Axioms 1–3 alone follows that the entropy is equivalent to the sum-form functional

$$\mathcal{U}(P) = \sum_{i=1}^n g(p_i) \sim f\left(\sum_{i=1}^n g(p_i)\right). \tag{3}$$

Thus, Axioms 1–3 alone exclude a wide class of existent entropies. Examples include the following. (a,λ)-escort entropies [44]; $S_{a,\lambda}(P) = 1/(\lambda - a)[(\sum_i p_i^a)^\lambda (\sum_i p_i^\lambda)^{-a} - 1]$ or Jizba-Arimitsu hybrid entropy [20,21]; $\mathcal{D}_q(P) = \ln_q \exp[-\sum_i P_i(q) \ln p_i]$, where $\ln_q x = (x^{1-q} - 1)/(1-q)$ is the q logarithm and $P_i(q) = p_i^q / \sum_j p_j^q$ is the escort distribution [2]. (b) Axiom 4 ensures that any entropy functional consistent with SJ axioms should be equivalent to $\sum_i p_i^q$. There are a number of entropic functionals that do not conform to this form, examples include: (c,d)-entropy [10,25]; $S_{c,d}(P) = \sum_i \Gamma(1+d, 1 - c \log p_i)$ or the Kaniadakis entropy [45]; $S_\kappa(P) = (1/2\kappa)\sum_i (p_i^{1+\kappa} + p_i^{1-\kappa})$. (c) Axiom 5 implies that the inference functional should be of the form

$$\mathcal{U}_q(P) \sim \left(\sum_i p_i^q\right)^{1/(1-q)} \quad \text{for } q > 0. \tag{4}$$

Only for $q > 0$, it is guaranteed that $\mathcal{U}_q(P)$ is Schur-concave which is a sufficient property for the maximality axiom [36]. For example, the Burg entropy [46] $\mathcal{K}(P) = K \sum_i \ln p_i$ provides an example of an entropy functional belonging to the class of $\mathcal{U}_q(P)$, but not fulfilling the maximality axiom. (d) SE is a unique candidate for MEP only when an extra desideratum is added to SJ axioms, namely, strong system independence (SSI): Whenever two subsystems of a system are disjoint, we can treat the subsystems in terms of independent distributions.

So far, the additivity property of the entropy functional was not our concern. Note, however, that $\mathcal{U}_q(P)$—known also as Rényi entropy powers [47], obey the multiplicative composition rule $\mathcal{U}_q(A \cup B) = \mathcal{U}_q(A)\mathcal{U}_q(B)$ for independent events. With appropriate f one can construct entropies with various types of composition rules. For instance, for $f(x) = \ln x$, we get a class of additive Rényi entropies [14], if $f(x) = \ln_{\mathcal{Q}} x$ we obtain $\mathcal{Q}$-additive Sharma-Mittal entropies [19]. For $\mathcal{Q} = q$, we end up with the class of Tsallis entropies [1]. Consequently, the MEP procedure implied by SJ axioms does not preclude any additivity rule as long as the entropy is $\sim \mathcal{U}_q(P)$.

Despite this, it is asserted in a number of recent works, cf. e.g., [32,33,35], that the only inference functional consistent with the SJ desiderata is SE, i.e., the $q = 1$ case. This was also the original result of SJ. The point of disagreement with these works can be retraced back to the axiom of system independence and its implementation in the original SJ proof [4,5]. Notably, SJ assumed that, because the prior distributions are independent and because the data-driven constraints I_1 and I_2 are independent (i.e., they give no information about any interaction between the two systems), the posterior distribution P must be written as a product of marginal distributions U and V. However, this goes well beyond the original SJ Axiom 4 in that the presumed independency of constraints invokes (unwarranted though often correct) a unique factorization rule for P. Clearly, having no information about interaction encoded in constraints (i.e., having independent constraints) is not the same as having no correlations among systems. Now, we show that there is an implicit assumption about the state-space structure in the SJ proof implied by the factorization rule $p_{ij} = u_i v_j$.

Factorization rule revisited.—Let us concentrate on the composition rule of ME distributions for two systems described by marginal distributions $U = \{u_i\}_{i=1}^n$ and $V = \{v_j\}_{j=1}^m$ and related constraints $\sum_{i=1}^n \mathcal{I}_i u_i = I$ and $\sum_{j=1}^m \mathcal{J}_j v_j = J$. The ME distributions U and V are obtained by maximizing $\mathcal{U}_q(U)$ and $\mathcal{U}_q(V)$, respectively, giving

$$\frac{q}{1-q}[\mathcal{U}_q(U)]^q u_i^{q-1} - \alpha_{\mathcal{I}} - \beta_{\mathcal{I}}\mathcal{I}_i = 0, \tag{5}$$

and likewise for V. The solution is

$$u_i = [\mathcal{U}_q(U)]^{-1}\left(1-(q-1)\frac{\beta_{\mathcal{I}}\Delta\mathcal{I}_i}{q\mathcal{U}_q(U)}\right)^{1/(q-1)}, \tag{6}$$

and analogously for u_j. Here, $\Delta\mathcal{I}_i = \mathcal{I}_i - I$ (similarly for $\Delta\mathcal{J}_j$). Lagrange multiplier $\alpha_{\mathcal{I}}$ (and similarly $\alpha_{\mathcal{J}}$) was eliminated via the normalization condition. The ME distribution of the joint system $P = \{p_{ij}\}$ includes both constraints, and hence,

$$\frac{q}{1-q}[\mathcal{U}_q(P)]^q p_{ij}^{q-1} - \alpha_{\mathcal{I}\mathcal{J}} - \beta_{\mathcal{I}\mathcal{J}}(\mathcal{I}_i + \mathcal{J}_j) = 0. \tag{7}$$

By inserting (5) into (7), we obtain [36]

$$[p_{ij}\mathcal{U}_q(P)]^{q-1} - 1 = \{[u_i\mathcal{U}_q(U)]^{q-1} - 1\} + \{[v_j\mathcal{U}_q(V)]^{q-1} - 1\}, \tag{8}$$

which can be rewritten in terms of the q product $x \otimes_q y = [x^{1-q} + y^{1-q} - 1]_+^{1/(1-q)}$ (with $x, y > 0$) as

$$\frac{1}{p_{ij}\mathcal{U}_q(P)} = \frac{1}{u_i\mathcal{U}_q(U)} \otimes_q \frac{1}{v_j\mathcal{U}_q(V)}. \tag{9}$$

When we apply to (9) the q logarithm, we obtain

$$I_q(P)\ominus_q S_q(P) = [I_q(U)\ominus_q S_q(U)] + U \leftrightarrow V. \tag{10}$$

Here, $I_q(r_k) = \ln_q(1/r_k)$ is the Tsallis-type Hartley information, $S_q(R) = \ln_q \mathcal{U}_q(R)$ is the Tsallis entropy and $x\ominus_q y = (x-y)/[1+(1-q)y]$ is the q difference. Note that (10) represents a q deformed version of the additive entropic rule. For $q \to 1$, both (9) and (10) reduce to $p_{ij} = u_i v_j$, which is tantamount to SSI. The reverse is true as well. By reexpressing (9) in terms of escort distributions $P_{ij}(q)$, $U_i(q)$, and $V_j(q)$ 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). The latter have (save for uniform and deterministic distributions) a unique solution [2] $q = 1$.

Systems where SSI fails are, e.g., systems where the number of accessible states $W(N)$ of a state set A_N does not grow exponentially with the number of distinguishable subsystems N, i.e., $W(N) \neq \mu^N, \mu > 1$ for $N \gg 1$ [hence, $W(N+M) \neq W(N)W(M)$] [48]. By the asymptotic equipartition property [49], the "typical" size $\tilde{W}(N)$ of A_N is $e^{S(A_N)}$, S is SE. The ensuing typical region contains almost all the probability of A_N as N increases [50]. Of all probability assignments compatible with constraints, MEP defines the largest typical region of A_N, [7]. If A and B are independent, then by SJ MEP, $S_{\max}(A_N \cup B_M) = S_{\max}(A_N) + S_{\max}(B_M)$, and hence, $\tilde{W}_{\max}(N+M) = \tilde{W}_{\max}(N)\tilde{W}_{\max}(M)$. By enforcing SSI, this extends to $W(N+M) = W(N)W(M)$ which, however, conflicts assumed $W(N+M) \neq W(N)W(M)$. So, states in $A_{N+M}\backslash\tilde{A}_{N+M}$ carry correlations not reflected in "independent" constraints. In fact, sampling procedures implying independent constraints and, indeed, the entire notion of independent sets largely sample only states in typical sets which might misrepresent existent correlations. Note that, despite the minuscule probability carried by $A_{N+M}\backslash\tilde{A}_{N+M}$, the cardinality of this set is $\propto W(N+M)$ for $N, M \gg 1$, i.e., truly huge [50]. With SSI, one implicitly assumes that $W(N) \propto \mu^N$. If the scaling is unknown, one should use MEP with $\mathcal{U}_q(P)$ as this is noncommittal about the form of $W(N)$. Sub- (or super-) exponential scaling is often found, e.g., in strongly correlated systems in quantum mechanics [51,52] or astrophysics [53,54].

Issue of correlations.—Now, we will see that, for $q \neq 1$, intrinsic system correlations are present even when constraints are SJ independent. Let us investigate the regime where q is close to 1 and expand a generic escort distribution $R_k(q)$ in the vicinity of $q = 1$ ($q \equiv 1 + \Delta q$) as

$$R_k(q) = r_k - r_k\Delta q[I(r_k) - \Gamma_1^R] + r_k\frac{(\Delta q)^2}{2} \times \{[I(r_k) - \Gamma_1^R]^2 - \Gamma_2^R\} + \mathcal{O}[(\Delta q)^3], \tag{12}$$

where $I(r_k) \equiv I_1(r_k) = \ln(1/r_k)$ is the Hartley information of the kth event and Γ_n^R are the bit-cumulants [2]. Notably, $\Gamma_1^R = S = -\sum_k r_k \ln r_k$ is the SE and $\Gamma_2^R = \sum_k r_k \ln^2 r_k - (\sum_k r_k \ln r_k)^2$ is the varentropy [55]. By inserting (12) into (11), we obtain

$$\begin{aligned} I(p_{ij}) - I(u_i) - I(v_j) - (\Gamma_1^P - \Gamma_1^U - \Gamma_1^V) \\ = \frac{\Delta q}{2}\{[I(p_{ij}) - \Gamma_1^P]^2 - [I(u_i) - \Gamma_1^U]^2 \\ -[I(v_j) - \Gamma_1^V]^2 - (\Gamma_2^P - \Gamma_2^U - \Gamma_2^V)\}. \end{aligned} \tag{13}$$

It is easy to show that, for independent systems, one has $\Gamma_k^U + \Gamma_k^V = \Gamma_k^{\mathrm{UV}}$, where $UV \equiv U \times V$ is the ensuing joint distribution. Thus, the differences $(\Gamma_k^P - \Gamma_k^U - \Gamma_k^V)$ quantify the correlations in the system. This can be seen by considering $p_{ij} = (1 + \epsilon_{ij}) u_i v_j$, where $\max_{ij}|\epsilon_{ij}| \ll 1$. In this case, we have

$$\Gamma_1^P = \Gamma_1^U + \Gamma_1^V - \frac{1}{2}\langle \epsilon^2 \rangle_0 + \mathcal{O}(\epsilon^3), \tag{14}$$

$$\Gamma_2^P = \Gamma_2^U + \Gamma_2^V + \langle \epsilon \ln^2(UV) \rangle_0 + \mathcal{O}(\epsilon^2), \tag{15}$$

where (see also Supplemental Material [36])

$$\langle \epsilon^2 \rangle_0 = \sum_{ij} \epsilon_{ij}^2 u_i v_j, \tag{16}$$

$$\langle \epsilon \ln^2(UV) \rangle_0 = \sum_{ij} \epsilon_{ij} \ln^2(u_i v_j) u_i v_j. \tag{17}$$

The term $\langle \epsilon^2 \rangle_0$ represents the strength of the correlations, and is always non-negative. The case $\langle \epsilon^2 \rangle_0 = 0$ happens only for independent distributions corresponding to $q = 1$. Γ_2^P represents a specific heat of the system (e.g., C_p in thermal systems) [2,56]. Thus, expression (17) represents the difference in specific heats ΔC with and without correlations ϵ_{ij}. A connection of the q parameter with ϵ_{ij} can be established by inserting (16)–(17) into (13), multiplying the whole equation by $u_i v_j$, and summing over i and j. At the leading order in ϵ, we get

$$q = 1 - 2\frac{\langle \epsilon^2 \rangle_0}{\langle \epsilon \ln^2(UV) \rangle_0} = 1 + 4\frac{\Delta S}{\Delta C}. \tag{18}$$

Now, let us provide two apt examples of MEP with $q \neq 1$.

Examples.—First, we consider a generic two-qubit quantum system (e.g., a bipartite spin-$\frac{1}{2}$ system). Starting from unentangled states $|11\rangle, |10\rangle, |01\rangle, |00\rangle$, we pass to the Bell basis of maximally entangled orthonormal states $|\Psi^\pm\rangle = (1/\sqrt{2})(|00\rangle \pm |11\rangle)$ and $|\Phi^\pm\rangle = (1/\sqrt{2})(|01\rangle \pm |10\rangle)$. Let us examine the situation where the only available constraint is given by a Bell-CHSH observable [51,57] $B = |\Phi^+\rangle\langle\Phi^+| - |\Psi^-\rangle\langle\Psi^-|$ mean value of which yields the (scaled) CHSH Bell inequality [57,58]. According to MEP, we should maximize $S(\rho) = [\mathrm{Tr}(\rho^q)]^{1/(1-q)}$ $(q > 0)$ subject to constraints $\mathrm{Tr}(\rho) = 1$ and $\mathrm{Tr}(\rho B) = b$ with $|b| \leq 1$. The corresponding MEP state, is given by [36]

$$\begin{aligned} \rho_{\mathrm{MEP}} = Z^{-1}(x, q)[(|\Phi^-\rangle\langle\Phi^-| + |\Psi^+\rangle\langle\Psi^+|) \\ + (1 + x)^{1/(q-1)}|\Phi^+\rangle\langle\Phi^+| \\ + (1 - x)^{1/(q-1)}|\Psi^-\rangle\langle\Psi^-|], \end{aligned} \tag{19}$$

where $x = \beta/\alpha$ is the ratio of Lagrange multipliers and $Z(x, q) = 2 + (1 + x)^{1/(q-1)} + (1 - x)^{1/(q-1)}$. We see that ρ_{MEP} is diagonal in the Bell basis. This Bell-diagonal state is not entangled if and only if [57] all its eigenvalues are less than or equal to $\frac{1}{2}$. From concavity of $(1 \pm x)^{1/(q-1)}$ for $q \geq 2$ and ensuing Jensen's inequality, it is easy to conclude [36] that all eigenvalues of ρ_{MEP} are $\leq 1/2$. Consequently, for $q \geq 2$ we obtain that ρ_{MEP} is not entangled (i.e., is separable). The situation for $q < 2$ is not conclusive, though inseparability can be deduced numerically. Fortunately, the case $q = 1$ (i.e., the SE case) is accessible analytically [36]. In this case, the eigenvalues of ρ_{MEP} are $p_{\Phi^-} = p_{\Psi^+} = \frac{1}{4}(1 - b^2)$, $p_{\Phi^+} = \frac{1}{4}(1 - b)^2$, $p_{\Psi^-} = \frac{1}{4}(1 + b)^2$. So, particularly for $b \in (\sqrt{2} - 1, 1]$, Shannonian MEP clearly predicts entanglement. However, one can find a non-MEP state [57], namely,

$$\rho = b|\Phi^+\rangle\langle\Phi^+| + \frac{1}{2}(1 - b)(|\Psi^+\rangle\langle\Psi^+| + |\Phi^-\rangle\langle\Phi^-|), \tag{20}$$

which satisfies the MEP constraint and is separable for $b \leq \frac{1}{2}$. Hence, Shannonian MEP predicts entanglement even if, for $b \in (\sqrt{2} - 1, \frac{1}{2}]$, there is a separable state that is fully compatible with the constraining data.

Clearly, the correct inference scheme (such as a presumed Shannonian MEP) should not yield an inseparable state if there exists (albeit only theoretically) a separable state compatible with the constraining data, or else one may get erroneous results (e.g., in quantum communication) by trying to use the entanglement inferred by MEP, while in reality, there is no entanglement present [57]. Note that, when MEP with $\mathcal{U}_q$, $q \geq 2$ is chosen, one can avoid the fake entanglement for any $b \leq 1$. The reason why Shannonian MEP implies spurious (quantum) correlations is that, in that analyzed quantum system, it does not comply with SSI due to use of the nonlocal Bell-CHSH observable. We note that problems with Shannonian MEP should be generically expected in entangled systems as entanglement does not conform to SSI because measurement results on (possibly distant) noninteracting subsystems (giving independent constraints) are still correlated. The situation should be particularly pressing in strongly entangled N-partite systems because there $W(N) \propto N^\rho, \rho > 0$, cf. [35,59].

As a second example, we consider the transverse momentum (p_T) distributions of hadrons produced in pp collisions at very high energies (center-of-mass energies $\sim 10^2 - 10^3$ GeV) as measured in RHIC and LHC experiments. The term transverse relates to the direction of colliding protons. From particle phenomenology, it is known that, in these cases, the effective number of distinguishable states with energy E shows a subexponential growth [60,61], i.e., $W(E) \propto \exp(\langle N \rangle^\gamma)$ with $0 < \gamma < 1$ and $\langle \cdots \rangle$ taken with respect to an appropriate multiplicity distribution. Thus, SSI (and hence, Shannon's MEP) is not warranted. In fact, the single-particle p_T distributions are best fitted by the q-exponential distributions (resulting from MEP based on $\mathcal{U}_q$) with $q \in [1.05, 1.10]$ depending on the type of the collision [8,62–65]. In these cases, the constraint (2) is represented by the mean of the transverse energy $E_T = \sqrt{p_T 2 + m^2}$ (m is hadron's rest mass). The typical picture is that, out of many hadrons produced in a given event, only one is selected (system A). The remaining $(N-1)$ particles (N is event dependent) act as a kind of a heat bath (HB) (system B) described by some apparent temperature. In this HB, the single-hadron p_T is effectively distributed according to the Maxwell-Jüttner distribution. The final distribution $u(p_T)$ is obtained by averaging over many events with distinct apparent temperatures. Systems A and B are clearly disjoint, but due to event-to-event temperature fluctuations the joint distribution $p(p_T, p_B) \neq u(p_T)v(p_B)$, so SSI is, indeed, violated. Now, since q is close to 1, we can consider only the leading order of ϵ_{ij} in $(q-1)$, i.e., $\epsilon_{ij} = (1-q)\beta^2 \Delta E_i^u \Delta E_j^v$. From (18), then, follows that [36]

$$q = 1 + \frac{\langle N \rangle - 1}{\beta^2 \langle (\Delta E^v)^2 \rangle_0} = 1 + \frac{\langle N \rangle - 1}{C_V^v}. \tag{21}$$

where $\langle (\Delta E^v)^2 \rangle_0 = \partial^2 \log Z^v / \partial \beta^2 = C_V^v / \beta^2$ (Z^v and C_V^v represent partition function [i.e., $\mathcal{U}_q(V)$] and heat capacity of the HB) and $\langle N - 1 \rangle = \beta \langle E^v \rangle_0$ is the virial relation where $1/\beta$ is the kinetic temperature of the hadronic HB. Note that system A factored out. Relations of the type (21) frequently appear in phenomenological studies on high-energy pp collisions [8,66,67].

Conclusions.—In summary, we have shown that the SJ axiomatization of the inference rule does account for a substantially wider class of entropic functionals than just SE. The root cause could be retraced to unreasonably strong assumptions employed by SJ in their proof—assumptions that go beyond the original SJ axioms. In particular, we have shown that Shannonian MEP is singled out as a unique method of statistical inference only insofar as an extra axiom of strong system independence is added to the SJ desiderata. While, for systems where state space scales exponentially with its size [as, e.g., in (quasi-) ergodic systems [50]] SE is the only entropy compatible with SJ axioms, for systems with sub- (super-) exponential growth, the assumption of SSI is not justified and the original proof of SJ needs revision. In our revised version of the proof, we identified a one-parameter class of admissible entropies whose utility was illustrated with two phenomenologically relevant examples.

P. J. and J. K. were supported by the Czech Science Foundation (GAČR), Grant No. 17-33812L. J. K. was also supported by the Austrian Science Foundation (FWF) under Project No. I3073.

*p.jizba@fjfi.cvut.cz
†jan.korbel@meduniwien.ac.at

[1] See, e.g., C. Tsallis, *Introduction to Nonextensive Statistical Mechanics: Approaching a Complex World* (Springer, New York, 2009).
[2] C. Beck and F. Schögl, *Thermodynamics of Chaotic Systems: An Introduction* (Cambridge University Press, Cambridge, 1993).
[3] See, e.g., R. B. Millar, *Maximum Likelihood Estimation and Inference* (John Wiley & Sons Ltd., Chichester, 2011).
[4] J. E. Shore and R. W. Johnson, IEEE Trans. Inf. Theory **26**, 26 (1980).
[5] J. E. Shore and R. W. Johnson, IEEE Trans. Inf. Theory **27**, 472 (1981).
[6] See, e.g., D. Petz, *Quantum Information Theory and Quantum Statistics* (Springer, Berlin, 2008).
[7] E. T. Jaynes, *Papers on Probability, Statistics, and Statistical Physics* (D. Reidel Publishing Co., Dordrecht, Holland, 1983).
[8] G. Bíró, G. G. Barnaföldi, T. S. Biró, K. Ürmössy, and Á. Takács, Entropy **19**, 88 (2017).
[9] T. S. Biró and P. Ván, Phys. Rev. E **83**, 061147 (2011).
[10] S. Thurner and R. Hanel, AIP Conf. Proc. **965**, 68 (2007).
[11] S. Thurner, P. Klimek, and R. Hanel, *Introduction to the Theory of Complex Systems* (Oxford University Press, Oxford, 2018).
[12] R. Hanel and S. Thurner, Europhys. Lett. **96**, 50003 (2011).
[13] R. Hanel, S. Thurner, and M. Gell-Mann, Proc. Natl. Acad. Sci. U.S.A. **111**, 6905 (2014).
[14] P. Jizba and T. Arimitsu, Ann. Phys. (Amsterdam) **312**, 17 (2004).
[15] J. P. Burg, Geophysics **37**, 375 (1972).
[16] C. Tsallis, J. Stat. Phys. **52**, 479 (1988).
[17] J. Havrda and F. Charvát, Kybernetika **3**, 30 (1967).
[18] T. Frank and A. Daffertshofer, Physica (Amsterdam) **285A**, 351 (2000).
[19] B. D. Sharma, J. Mitter, and M. Mohan, Inf. Control **39**, 323 (1978).
[20] P. Jizba and J. Korbel, Physica (Amsterdam) **444A**, 808 (2016).
[21] P. Jizba and T. Arimitsu, Physica (Amsterdam) **340A**, 110 (2004).
[22] G. Vos, Nucl. Phys. **B899**, 91 (2015).
[23] P. Tempesta, Phys. Rev. E **84**, 021121 (2011).
[24] V. M. Ilić and M. S. Stanković, Physica (Amsterdam) **411A**, 138 (2014).

[25] R. Hanel and S. Thurner, Europhys. Lett. **93**, 20006 (2011).
[26] E. T. Jaynes, Phys. Rev. **106**, 620 (1957).
[27] E. T. Jaynes, Phys. Rev. **108**, 171 (1957).
[28] See, e.g., *Entropy Measures, Maximum Entropy Principle and Emerging Applications*, edited by Karmeshu (Springer-Verlag, Berlin, 2003).
[29] L. L. Campbell, Inf. Control **8**, 423 (1965).
[30] P. Nath, Inf. Control **29**, 234 (1975).
[31] S. Abe, Phys. Lett. A **224**, 326 (1997).
[32] S. Pressé, K. Ghosh, J. Lee, and K. A. Dill, Phys. Rev. Lett. **111**, 180604 (2013).
[33] S. Pressé, K. Ghosh, J. Lee, and K. A. Dill, Entropy **17**, 5043 (2015).
[34] J. Uffink, Stud. Hist. Phil. Mod. Phys. **26**, 223 (1995).
[35] C. Tsallis, Entropy **17**, 2853 (2015).
[36] See Supplemental Material at http://link.aps.org/supplemental/10.1103/PhysRevLett.122.120601 for finer technical details, which include Refs. [9,37–43].
[37] J. Aczel and J. Dhombres, *Functional Equations in Several Variables* (Cambridge University Press, Cambridge, 1989).
[38] A. W. Marshall and I. Olkin, *Inequalities: Theory of Majorization and Its Application*, Mathematics in Science and Engineering Vol. 143 (Academic Press Inc., London, 1979).
[39] A. W. Roberts and D. E. Varberg, Pure Appl. Math. **57**, 12 (1973).
[40] P. Jizba, J. Korbel, and V. Zatloukal, Phys. Rev. E **95**, 022103 (2017).
[41] R. Korus, St. Mrowczýnski, M. Rybczýnski, and Z. Wlodarczyk, Phys. Rev. C **64**, 054908 (2001).
[42] M. Stephanov, K. Rajagopal, and E. V. Shuryak, Phys. Rev. D **60**, 114028 (1999).
[43] A. Mekjian, Phys. Rev. C **73**, 014901 (2006).
[44] J.-F. Bercher, Phys. Lett. A **375**, 2969 (2011).
[45] G. Kaniadakis, Physica (Amsterdam) **365A**, 17 (2006).
[46] A. N. Gorban and I. V. Karlin, Phys. Rev. E **67**, 016104 (2003).
[47] P. Jizba, Y. Ma, A. Hayes, and J. A. Dunningham, Phys. Rev. E **93**, 060104(R) (2016).
[48] J. Korbel, R. Hanel, and S. Thurner, New J. Phys. **20**, 093007 (2018).
[49] S. B. Nicholson, M. Alaghemandi, and J. R. Green, J. Chem. Phys. **145**, 084112 (2016).
[50] P. Billingsley, *Ergodic Theory and Information* (Wiley, New York, 1965).
[51] N. Canosa and R. Rossignoli, Phys. Rev. Lett. **88**, 170401 (2002).
[52] A. K. Rajagopal, R. S. Mendes, and E. K. Lenzi, Phys. Rev. Lett. **80**, 3907 (1998).
[53] A. Taruya and M. A. Sakagami, Phys. Rev. Lett. **90**, 181101 (2003).
[54] A. R. Plastino and A. Plastino, Phys. Lett. A **174**, 384 (1993).
[55] F. Schlögl, Z. Phys. B **52**, 51 (1983).
[56] L. D. Landau and I. M. Lifschitz, *Course of Theoretical Physics: Statistical Physics* (Pergmon Press, New York, 1958).
[57] R. Horodecki, M. Horodecki, and P. Horodecki, Phys. Rev. A **59**, 1799 (1999).
[58] J. F. Clauser, M. A. Horne, A. Shimony, and R. A. Holt, Phys. Rev. Lett. **23**, 880 (1969).
[59] F. Caruso and C. Tsallis, Phys. Rev. E **78**, 021102 (2008).
[60] T. S. Biró and G. Purcsel, Phys. Rev. Lett. **95**, 162302 (2005).
[61] T. S. Biró and Z. Néda, Physica (Amsterdam) **499A**, 335 (2018).
[62] PHENIX Collaboration, Phys. Rev. Lett. **91**, 241803 (2003).
[63] STAR Collaboration, Phys. Rev. Lett. **108**, 072302 (2012).
[64] ALICE Collaboration, Eur. Phys. J. C **74**, 3108 (2014).
[65] ALICE Collaboration, Eur. Phys. J. C **75**, 226 (2015).
[66] G. Wilk and Z. Wlodarczyk, Phys. Rev. Lett. **84**, 2770 (2000).
[67] G. Wilk and Z. Wlodarczyk, Phys. Rev. C **79**, 054903 (2009).