Groups, Information Theory and Einstein's Likelihood Principle

Gabriele Sicuro*
Centro Brasileiro de Pesquisas Físicas, Rua Dr. Xavier Sigaud, 150, 22290-180, Rio de Janeiro, Brazil

Piergiulio Tempesta†
Departamento de Física Teórica II (Métodos Matemáticos de la física), Facultad de Físicas, Universidad Complutense de Madrid, 28040 – Madrid, Spain, and Instituto de Ciencias Matemáticas, C/ Nicolás Cabrera, No 13–15, 28049 Madrid, Spain

(Dated: December 2, 2015)

We propose a unifying picture where the notion of generalized entropy is related to information theory by means of a group-theoretical approach. The group structure comes from the requirement that an entropy be well defined with respect to the composition of independent systems, in the context of a recently proposed generalization of the Shannon–Khinchin axioms. We associate a general information functional to each member of a large class of non-extensive entropies, satisfying the additivity property on a set of independent systems on the basis of the underlying group law. At the same time, we also show that the Einstein likelihood function naturally emerges as a byproduct of our informational interpretation of nonadditive entropies. These results confirm the adequacy of composable entropies both in physical and social science contexts.

The study of the relations among Statistical Mechanics, Information Theory and the notion of entropy is at the heart of the science of complexity, and in the last decades has been widely explored. After the seminal works of Shannon [1] and Khinchin [2] on the foundations of Information Theory, Jaynes [3] re-formulated Boltzmann–Gibbs (BG) equilibrium statistical mechanics as a statistical inference theory, where all fundamental equations are consequences of the maximum entropy principle applied to the *BG entropy* S^{BG}. Subsequently, Rényi [4, 5] introduced a generalized measure of information, now called *Rényi entropy* S^{R}_{α}, depending on a real parameter α having the BG entropy as particular case in the $\alpha \to 1$ limit.

The recent explosion of interest in non–equilibrium statistical physics motivated, among many research lines, the search for new entropic functionals. The aim was to extend the classical BG formulation to new contexts, especially complex systems and non-extensive systems in strongly correlated dynamical regimes. The S^{T}_{q} entropy, introduced by Havrda and Charvát [6] and Tsallis [7], has been the prototype of the nonadditive entropies studied in the last decades [8–13]. These functionals are generalizations of the BG entropy and they depend on one or more parameters, in such a way that the BG entropy is recovered as a particular limit.

This research has led to a new flow of ideas regarding the old problem of the probabilistic versus the dynamical foundations of the notion of entropy. It is well known [14] that Einstein's approach was very different with respect to the probabilistic methodology of Boltzmann (which eventually emerged as the predominant one). Indeed, Einstein argued that the probabilities of occupation of the various regions of the phase space associated with a physical systems cannot be postulated *a priori*. Instead, only a knowledge of dynamics, obtained by solving the equations of motion, could provide this information. For this reason, Einstein [15] introduced the likelihood function $\mathcal{W} \propto \exp\left(S^{\mathrm{BG}}\right)$ as fundamental quantity (for the sake of semplicity, here and in the following we put $k_{\mathrm{B}} \equiv 1$, k_{B} being the Boltzmann constant). He observed that, by composing two independent systems $\mathcal{A}$ and $\mathcal{B}$, the fundamental relation

$$\mathcal{W}(\mathcal{A} \cup \mathcal{B}) = \mathcal{W}(\mathcal{A})\mathcal{W}(\mathcal{B}) \quad (1)$$

holds. Eq. (1) expresses the fact that the physical description of the system $\mathcal{A}$ does not depend on the physical description of the system $\mathcal{B}$, and *vice versa*. Needless to say, the factorization in Eq. (1) is epistemologically crucial. Moreover, it anticipated the additivity requirement of the information content of independent systems, as it will be explained below.

By following the analysis in [16], we shall call Eq. (1) the Einstein's likelihood principle. The likelihood function has a clear physical meaning. It is indeed the number of accessible configurations in the entire space of possible configurations. In many circumstances, this number is exponential in the size N of the system and we can write $\mathcal{W} \propto \exp(N\Sigma)$, where Σ is an (adimensional) entropy density. Mézard and Parisi [17] introduced this quantity in the study of disorderd systems, calling Σ complexity, or configurational entropy, and an equivalent quantity is used in the study of random optimization problems by means of statistical physics techniques [18].

The aim of this paper is to provide a novel approach that relates, in a unique framework, classical information theory with both the notion of generalized entropy and Einstein's likelihood principle. Precisely, we shall show that an intrinsic *group–theoretical structure* is at the heart of the multiple connections among these foundational perspectives.

The physical root of this group structure relies in a generalization of the classical axiomatic formulation, originally proposed by Shannon and by Khinchin to characterize the BG entropy. The first three postulates, nowadays called the Shannon–Khinchin (SK) axioms [1, 2], define some crucial properties that any entropic functional S should satisfy. Let us consider the set $\mathscr{P}$ of finite discrete probability distributions $P \in \mathscr{P}$, $P \coloneqq \{p_1, \ldots, p_W\}$, $W \in \mathbb{N}$, $p_i > 0$, $\sum_{i=1}^{W} p_i = 1$. We can state the postulates as follows.

Continuity: The function $S(p_1, \ldots, p_W)$ is continuous respect to all its arguments.

Maximum principle: The function $S(p_1, \ldots, p_W)$ is maximized by $p_i = W^{-1}\ \forall i$.

Expansibility: $S(p_1, \ldots, p_W, 0) \equiv S(p_1, \ldots, p_W)$ (adding an event of zero probability does not change the value of the entropy).

Also, a fourth axiom, i.e. *additivity* with respect to the composition of two systems, was required. Under these assumptions, Khinchin proved that the only admissible entropy turns out to be the BG entropy $S^{\mathrm{BG}}[P] \coloneqq -\sum_{i=1}^{W} p_i \ln p_i$.

The additivity property was thought to be a sufficient condition for the *extensivity* of BG entropy, which in the formulation of Clausius is an essential requisite for thermodynamics. However, in the last decades it became evident that the two properties, i.e. additivity and extensivity, are completely unrelated [19]. Indeed, denoting by $W(N)$ the number of accessible states of a system with N particles, the BG entropy is not extensive, *on the uniform distribution*, if $W(N) \sim N^{\alpha}$ for a certain $\alpha \in \mathbb{R}^{+}$. This scaling is not atypical and it appears often in the framework of complex systems. Recently, the non-extensivity of Boltzmann's entropy over a large class of probability distributions was proved. Surprisingly, Rényi entropy can be extensive in the same contexts [20]. Additivity is therefore not an intrinsic physical requirement. At the same time, by renouncing to the additivity postulate, new possibilities arise.

In the context of Information Theory, non–additive entropies provided useful tools, for example in the study of quantum entanglement [9, 21]. However, the lack of additivity appears to be an important flaw if we want to interpret generalized entropic functionals as classical information functionals [5, 22]. Indeed, it is expected that, when composing two statistically independent systems, the total amount of information is nothing but the sum of the information content of the two systems. Moreover, any measurement or change of information content in one of the two systems does not affect, nor is influenced by, any other property of the other system, being the systems uncorrelated.

This property is certainly satisfied by Boltzmann's and Rényi's entropies, but using non–additive entropic functionals *tout–court* it is not possible to fulfill the requirement above. Therefore, to preserve the meaning of entropy as a measure of the information content of a given system $\mathcal{A}$, we firstly postulate the general relation

$$\mathcal{I}(\mathcal{A}) = f(S(\mathcal{A})), \tag{2}$$

where the information is assumed to be a function f of the entropy of the system only.

Secondly, along the lines of [23, 24], we discuss an axiomatic formulation of entropy, generalizing the fourth SK axiom and requiring, instead of additivity, the *composability* property [25, 26]. In this settings, we simply postulate that the entropy of a composite system be a function of the entropy of the two components only. This simple requirement is however far rich of consequences. In particular, a specific algebraic group structure appears, in which the composition operation plays the role of group operation. In the following, we will show that, for a class of entropies, an additive information functional having the form of Eq. (2) is automatically inferred by the algebraic structure itself. The composition process of two or more systems is formalized as follows.

Composability axiom: Given two statistically independent systems $\mathcal{A}$ and $\mathcal{B}$, each defined over a given probability distribution in $\mathscr{P}$, there exists a symmetric function $\Phi(x, y)$ such that

$$S(\mathcal{A} \cup \mathcal{B}) = \Phi\left(S(\mathcal{A}), S(\mathcal{B})\right). \tag{3}$$

Moreover, we require the associativity property

$$\Phi\left(\Phi(x, y), z\right) = \Phi\left(x, \Phi(y, z)\right) \tag{4}$$

and the relation $\Phi(x, 0) = x$.

When the systems $\mathcal{A}$ and $\mathcal{B}$ are not independent, we postulate the relation

$$S(\mathcal{A} \cup \mathcal{B}) = \Phi\left(S(\mathcal{A}), S(\mathcal{B}|\mathcal{A})\right), \tag{5}$$

where $S(\mathcal{B}|\mathcal{A})$ is the entropy evaluated on the conditional distribution $p_{ij}^{\mathcal{B}|\mathcal{A}} \coloneqq \frac{p_{ij}^{\mathcal{A}\cup\mathcal{B}}}{p_i^{\mathcal{A}}}$.

We shall define *admissible entropies* those satisfying the first three SK axioms and the *composability axiom*. The axiom contains crucial requirements, that allow the existence of a zeroth law of thermodynamics. Also, when composing a system with another in a certainty state (zero entropy), the entropy of the composed system will remain unchanged. This natural generalization of the SK axiom allows an infinite number of admissible entropic forms [23, 24].

The classification and the study of the properties of all functions $\Phi(x, y)$ satisfying the composability axiom was performed in the context of formal group theory [27].

This branch of algebraic topology was established in the second half of the XX century, starting from the work of Bochner [28]. In particular, there exists a universal group law, the *Lazard formal group*, whose general expression is given by

$$\Phi(x,y) = G\left(G^{-1}(x) + G^{-1}(y)\right). \tag{6}$$

Here $G(t)$ is a general strictly monotonically increasing function whose power series expansion is of the form $G(t) = \sum_{k=0}^{\infty} a_k \frac{t^{k+1}}{k+1}$, where $\{a_k\}_{k\in\mathbb{N}}$ a real sequence, with $a_0 \neq 0$. This series, by means of the Lagrange inversion theorem, is invertible with respect to the composition (i.e., there exists a series G^{-1} such that $G^{-1}(G(t)) = t$). One can prove that, given a group law, there exists a specific function G such that the law takes the form in Eq. (6) [29].

Consequently, if a functional S satisfies the generalized SK axioms, then the associated composition law (3) can be realized in terms of the universal law (6) by means of a specific function G. The rich algebraic structure underlying the simple composability axiom naturally allows to classify all entropies possessing a given composition law. Also, we can generate new examples, according to the composition law adopted. For the trace-form family (i.e., the family of entropies having the structure $S[P] = \sum_i p_i f(p_i)$ for a certain concave function f), we have the form [24]

$$S_G[P] = \sum_{i=1}^{W} p_i G\left(\ln \frac{1}{p_i}\right). \tag{7}$$

For example, the celebrated Boltzmann entropy corresponds to the *additive group*:

$$\Phi(x,y) = x + y \Rightarrow G(t) = t. \tag{8}$$

The *multiplicative group*

$$\Phi(x,y) = x + y + (1-q)xy \Rightarrow G(t) = \frac{e^{(1-q)t} - 1}{1-q} \tag{9}$$

leads to the Tsallis entropy, defined for $q \in \mathbb{R}$ as [30]

$$S_q^{\mathrm{T}}[P] := \sum_{i=1}^{W} p_i \log_q \frac{1}{p_i} \xrightarrow{q\to 1} S^{\mathrm{BG}}[P], \tag{10}$$

where $\log_q(x) := \frac{x^{1-q}-1}{1-q} \xrightarrow{q\to 1} \ln x$, $x \in \mathbb{R}^+$. Remarkably, the BG entropy and the Tsallis entropy are the *only* trace form entropies that are admissible. All other entropies in the form (7) satisfy the composability property only on the uniform distribution, i.e. in the microcanonical ensemble. Dropping out the trace-form hypothesis, new entropic forms are allowed [24]. For example, the Rényi entropy

$$S_\alpha^{\mathrm{R}}[P] := \frac{\ln \sum_{i=1}^{W} p_i^\alpha}{1-\alpha} \xrightarrow{\alpha\to 1} S^{\mathrm{BG}}[P] \tag{11}$$

belongs to the additive group of composable entropies.

Motivated by the previous discussion, we can propose now a notion of information functional that comes directly from the group-theoretical structure. A priori, apart from the entropies belonging to the additive group, all the other ones *do not* satisfy the additivity property for the composition of probabilistic independent systems. However, we can overcome this difficulty by associating to each of them an information functional that is indeed additive on statistically independent systems. This functional is determined by the composition class itself, i.e., by the function G appearing in Eq. (6). Consequently, we propose the following main definition of a *group-theoretical information functional*.

Given a composable entropy S_G, with a group law of the form (6), the information functional of S for any system $\mathcal{A}$ is defined to be

$$\mathcal{I}_G(\mathcal{A}) = G^{-1}(S_G(\mathcal{A})). \tag{12}$$

In the specific case of an entropy of trace-form class (7), we recover for our functional the expression of the *Kolmogorov-Nagumo mean* [31, 32]:

$$\mathcal{I}_G(\mathcal{A}) = G^{-1}\left(\sum_{i=1}^{W} p_i G\left(\ln \frac{1}{p_i}\right)\right). \tag{13}$$

The information functional (12), however, is defined in the generic case of entropies that are composable and not only on trace-form entropies. We are ready now to present one of the main results of this Letter.

Theorem. *Let S be a composable entropy, with a group law defined by (6) for a certain function G. Then for two statistically independent systems $\mathcal{A}$ and $\mathcal{B}$ we have*

$$\mathcal{I}_G(\mathcal{A}\cup\mathcal{B}) = \mathcal{I}_G(\mathcal{A}) + \mathcal{I}_G(\mathcal{B}). \tag{14}$$

Moreover, the information functional $\mathcal{I}_G$ satisfies the following further properties.

Continuity: *$\mathcal{I}_G$ is continuous respect to its arguments;*

Maximum principle: *$\mathcal{I}_G$ is maximized on the uniform distribution;*

Expansibility: *The addition of a zero–probability event do not change the value of $\mathcal{I}_G$.*

Proof. Observe indeed that

$$\begin{aligned}\mathcal{I}_G(\mathcal{A}\cup\mathcal{B}) &= G^{-1}(S_G(\mathcal{A}\cup\mathcal{B})) = G^{-1}\left(\Phi(S_G(\mathcal{A}), S_G(\mathcal{B}))\right) \\ &= G^{-1}\left\{G\left[G^{-1}(S_G(\mathcal{A})) + G^{-1}(S_G(\mathcal{B}))\right]\right\} \\ &= \mathcal{I}_G(\mathcal{A}) + \mathcal{I}_G(\mathcal{B}). \end{aligned} \tag{15}$$

All other properties of the functional $\mathcal{I}_G$ derive from the properties of S_G imposed by the generalized SK axioms and from the strict monotonicity of G. □

Therefore all composable entropies possess an associated information functional which is additive and can be constructed directly through the function G. Observe that the strict monotonicity of G (and therefore of G^{-1}) implies that $S(\mathcal{A}) < S(\mathcal{B}) \Rightarrow \mathcal{I}_G(\mathcal{A}) < \mathcal{I}_G(\mathcal{B})$, coherently with the fact that the entropic forms allowed by the axioms above do possess indeed an information content (see for example [9, 33] for the study of the Tsallis entropic form as information functional and its application).

Observe also that Rényi entropy fits naturally in our scheme. Indeed, for S^{R}_{α}, $G(t) = t$, then the associated information functional $\mathcal{I}^{\mathrm{R}}_{\alpha}$ coincides with the entropic functional, i.e.

$$S^{\mathrm{R}}_{\alpha}(\mathcal{A}) \mapsto \mathcal{I}^{\mathrm{R}}_{\alpha}(\mathcal{A}) \equiv S^{\mathrm{R}}_{\alpha}(\mathcal{A}). \tag{16}$$

The previous identity holds also in the $\alpha \to 1$ limit, i.e., in the BG case. In other words, the BG entropy and the Rényi entropy are stable with respect to definition (12): the associated group-theoretical information functional coincides with the corresponding entropy. If we consider instead the Tsallis entropy, we see that

$$S^{\mathrm{T}}_{q}(\mathcal{A}) \mapsto \mathcal{I}^{\mathrm{T}}_{q}(\mathcal{A}) \equiv S^{\mathrm{R}}_{q}(\mathcal{A}), \tag{17}$$

i.e., the information functional associated to S^T_q is the Rényi entropic functional with parameter $1-q$. Again, this result is not surprising, being the Rényi entropy the only information functional that satisfy the additivity requirement and has the structure of a Kolmogorov–Nagumo mean [22]. However, as explained above, our formalism goes beyond the requirement of a Kolmogorov–Nagumo structure and it holds, indeed, for all non-trace form admissible entropies, having a more general form for the corresponding information functional. For a large classe of non-trace form entropies (for example for the entropic functionals discussed in [24]), the analitic expression of the information functional is of the type

$$S_G(\mathcal{A}) \mapsto \mathcal{I}_G(\mathcal{A}) = \sum_i c_i S^{\mathrm{R}}_{\alpha_i}(\mathcal{A}), \tag{18}$$

where the parameters c_i and α_i depend on the parameters appearing in the entropy S_G.

Finally, the presented group-theoretical approach allows to generalize easily the Einstein principle, and to connect it with information theory in a natural way. Given an entropy S_G, whose associated group law is (6), we introduce the likelihood function

$$\mathcal{W}_G(\mathcal{A}) =: e^{\mathcal{I}_G(\mathcal{A})} = e^{G^{-1}(S_G(\mathcal{A}))}. \tag{19}$$

The motivation for this definition is twofold. First, it relates the Einstein likelihood function directly with the group-theoretical information functional. Second, it generalizes Einstein's relations [15, 16] in the case of independent systems for all composable entropies. Indeed, let S be a composable entropy, whose group law is given by (6). Then Einsten's likelihood principle (1) follows immediately from the additivity property of the information functional (12). In the case of the multiplicative group (9), we recover the likelihood function recently introduced in [16]. Once again, for generalized entropies composable only over the uniform distribution, just a weak formulation of the principle holds. Needless to say, this situation is not very satisfactory,

In the light of the whole analysis of this paper, we conclude that the composability axiom allows a potentially fruitful interpretation of generalized entropies in information theory. Indeed, composable entropies both possess an information theoretical content and satisfy the Einstein principle, which is a crucial statement for any physical application of the notion of entropy. As a byproduct of our analysis, it emerges that also non-trace form but composable entropies can play an important role in statistical mechanics.

The authors are grateful to Constantino Tsallis for useful discussions. G.S. acknowledges the financial support of the John Templeton Foundation. The research of P.T. has been partly supported by the project FIS2011–22566, Ministerio de Ciencia e Innovación, Spain.

* sicuro@cbpf.br
† p.tempesta@fis.ucm.es

[1] C. Shannon, The Bell System Technical Journal **27**, 379 (1948).
[2] A. Khinchin, *Mathematical Foundations of Information Theory*, Dover Books on Mathematics (Dover Publications, 1957).
[3] E. T. Jaynes, Physical Review **106**, 620 (1957).
[4] A. Rényi, in *Proc. 4th Berkeley Symp. Math. Statist. Probability* (1961) pp. 547–561.
[5] A. Rényi, *Probability theory* (North–Holland, 1970).
[6] J. Havrda and F. Charvát, Kybernetika **3**, 30 (1967).
[7] C. Tsallis, Journal of statistical physics **52**, 479 (1988).
[8] E. P. Borges and I. Roditi, Physics Letters A **246**, 399 (1998).
[9] N. Canosa and R. Rossignoli, Physical Review Letters **88**, 170401 (2002).
[10] C. Beck and E. Cohen, Physica A: Statistical mechanics and its applications **322**, 267 (2003).
[11] R. Hanel, S. Thurner, and M. Gell-Mann, Proceedings of the National Academy of Sciences **108**, 6390 (2011).
[12] P. Tempesta, Physical Review E **84**, 021121 (2011).
[13] C. Tsallis and L. J. Cirto, The European Physical Journal C **73**, 1 (2013).
[14] E. Cohen, Physica A: Statistical Mechanics and its Applications **305**, 19 (2002).
[15] A. Einstein, Annalen der Physik **338**, 1275 (1910).
[16] C. Tsallis and H. J. Haubold, EPL (Europhysics Letters) **110**, 30005 (2015).
[17] M. Mézard and G. Parisi, Journal of Statistical Physics **111**, 1 (2003).
[18] V. Bapst, L. Foini, F. Krzakala, G. Semerjian, and F. Zamponi, Physics Reports **523**, 127 (2013).

[19] H. Touchette, Physica A: Statistical Mechanics and its Applications **305**, 84 (2002).
[20] H. Bergeron, E. Curado, J. Gazeau, and L. M. Rodrigues, Physica A **441**, 23 (2016).
[21] M. Berta, M. Christandl, R. Colbeck, J. M. Renes, and R. Renner, Nature Physics **6**, 659 (2010).
[22] P. Jizba and T. Arimitsu, Annals of Physics **312**, 17 (2004).
[23] P. Tempesta, Annals of Physics (2014), in press.
[24] P. Tempesta, arXiv preprint arXiv:1507.07436 (2015).
[25] H. Callen, *Thermodynamics and an Introduction to Thermostatistics* (Wiley, 1985).
[26] C. Tsallis, *Introduction to Nonextensive Statistical Mechanics: Approaching a Complex World* (Springer New York, 2009).
[27] M. Hazewinkel, *Formal groups and applications*, Vol. 78 (Elsevier, 1978).
[28] S. Bochner, Annals of Mathematics , 192 (1946).
[29] This set of properties in particular holds for group laws on zero–characteristic fields. For further details, see [24].
[30] C. Tsallis, Quimica Nova **17**, 468 (1994).
[31] A. Kolmogorov, in *Atti della R. Accademia Nazionale dei Lincei*, Vol. 12 (1930) pp. 388–391.
[32] M. Nagumo, in *Japanese journal of mathematics: transactions and abstracts*, Vol. 7 (The Mathematical Society of Japan, 1930) pp. 71–79.
[33] C. Tsallis, Physical Review E **58**, 1442 (1998).