Entropy measures of physics via complexity

Giorgio Kaniadakis and Flemming Topsøe
Politecnico of Torino, Department of Physics
and
University of Copenhagen, Department of Mathematics

1 Introduction, Background

A large number of papers have been devoted to entropy (or uncertainty) measures going beyond Boltzmann-Shannon entropy, aiming at covering the needs of non-extensive statistical physics. In particular, *q-entropy* or *Tsallis entropy*, is now popular among physicists (the database maintained by Tsallis, cf. [6], contains more than 1000 entries!). In fact, the q-entropy functions were first considered in mathematics by Havrda and Charvát, [1]. They are monotone functions of Rényi entropies and one might dismiss them on this ground. However, the point is that they are of significance for studies of statistical physics as witnessed in many papers. More recently, motivated by the needs of relativistic statistical physics, Kaniadakis introduced a family of entropy measures which he called *κ-entropies*, cf. [3],

We will point to the significance of q-entropy and κ-entropy by showing that they fit into a general theory based on measures of complexity.

In Section 2 the notion of complexity together with a related game is discussed. This follows ideas which can be traced from Topsøe [5]. Basically, the philosophy is consistent with maximum entropy thinking of Jaynes, cf. [2], but goes one step deeper by deriving the maximum entropy principle from an underlying game theoretical principle. In Section 3, q- and κ-entropies are derived from suitable measures of complexity.

2 Games of complexity

Let $\mathbb{A}$ be finite or countably infinite. Introduce *Player I*, "Nature", and *Player II*, "the physicist". Let the *strategies* available to Player I be given by a set $\mathcal{S}_I$ of distributions over $\mathbb{A}$, *the preparation.* As an indicative example, think of an Ising spin system provided with an energy function. Then $\mathbb{A}$ consists of all sequences $(i_1, \cdots, i_n)$ of 0's and 1's with n the number of particles and as the preparation you could consider all distributions over $\mathbb{A}$ with a prescribed mean energy. As strategy set $\mathcal{S}_{II}$ for Player II we take the set of *all* distributions over $\mathbb{A}$.

We consider a *measure of relative complexity* (below just *complexity*) which, to any pair (P, Q) of distributions, assigns a number $\Phi(P\|Q) \in [0, \infty]$, thought of as the complexity for Player II in observing the physical system based on the strategy Q when Player I has chosen the strategy P. We say that $\Phi(P\|Q)$ is the *complexity of* P *with reference* Q.

By definition, Φ-*entropy* equals minimal complexity, i.e. we put

$$\mathrm{S}_\Phi(P) = \inf_{Q \in \mathcal{S}_{II}} \Phi(P\|Q)\,. \tag{1}$$

We assume that the following axioms are fulfilled:

$$\mathrm{S}_\Phi(P) < \infty \text{ for } P \in \mathcal{S}_I\,, \tag{2}$$

$$\mathrm{S}_\Phi(P) = \Phi(P\|P) \text{ for } P \in \mathcal{S}_I\,, \tag{3}$$

$$\Phi(P\|Q) > \mathrm{S}_\Phi(P) \text{ for } P \in \mathcal{S}_I\,, Q \in \mathcal{S}_{II}\,, Q \neq P. \tag{4}$$

The basic axioms (3) and (4) are quite natural. Indeed, they express that the complexity of P is the smallest when P itself is taken as reference.

We define Φ-*redundancy* (or *divergence*) by $\mathrm{D}_\Phi(P\|Q) = \Phi(P\|Q) - \mathrm{S}_\Phi(P)$, i.e. as *actual complexity minus minimal complexity.* Often, this is written

$$\Phi(P\|Q) = \mathrm{S}_\Phi(P) + \mathrm{D}_\Phi(P\|Q)\,, \tag{5}$$

the *linking identity*. Clearly, $\mathrm{D}_\Phi(P\|Q) \geq 0$ with equality if and only if $P = Q$. In certain models of statistical physics, $\mathrm{D}(P\|Q) > 0$ can be related to *free energy* and if $\mathrm{D}(P\|Q) > 0$ this gives a possibility to perform some work.

Introduce the two-person zero-sum game γ_Φ with Player I and Player II as players and with Φ as objective function, viewed as a cost to Player II. Applying usual "minimax/maximin thinking" of game theory, Player I will, when contemplating whether or not to use the strategy $P \in \mathcal{S}_I$, pay attention to the best counter strategy by Player II, i.e. the strategy which minimizes complexity. By (1), this leads to the value $\mathrm{S}_\Phi(P)$. Therefore, Player I considers $P^* \in \mathcal{S}_I$ to be an *optimal strategy* if $\mathrm{S}_\Phi(P^*) = \mathrm{S}_\Phi^{\max}$, the *maximum entropy value*, given by $\mathrm{S}_\Phi^{\max} = \sup_{P \in \mathcal{S}_I} \mathrm{S}_\Phi(P)$.

Player II will pay attention to the *risk function* $\mathrm{R}_\Phi(Q) = \sup_{P \in \mathcal{S}_I} \Phi(P\|Q)$, and consider Q^* to be an *optimal strategy* if $\mathrm{R}_\Phi(Q^*) = \mathrm{R}_\Phi^{\min}$, the *minimum risk value*, defined by $\mathrm{R}_\Phi^{\min} = \inf_{Q \in \mathcal{S}_{II}} \mathrm{R}_\Phi(Q)$.

Clearly, $\mathrm{S}_\Phi^{\max} \le \mathrm{R}_\Phi^{\min}$. If equality holds, the game is in *equilibrium*. A pair $(P^*, Q^*) \in \mathcal{S}_I \times \mathcal{S}_{II}$ is a *Nash equilibrium pair* if the *saddle value inequalities*

$$\Phi(P\|Q^*) \le \Phi(P^*\|Q^*) \le \Phi(P^*\|Q) \tag{6}$$

hold for any $P \in \mathcal{S}_I$ and any $Q \in \mathcal{S}_{II}$. If the players choose strategies prescribed by such a pair, none of the players will benefit from changing strategy – assuming that the other player does not do so either.

A strategy Q for Player II is *robust* if, for some $h < \infty$, $\Phi(P\|Q) = h$ for all $P \in \mathcal{S}_I$. The set of robust strategies forms the Φ-*exponential family*.

The *principle of game theoretical equilibrium* (GTE) dictates that you search for optimal strategies for the players and investigate if the game is in equilibrium. By the identification of the optimal strategies for Player I, we already see that this leads to the MaxEnt-principle. If there is a discrepancy between $\mathrm{S}_\Phi^{\max}$ and $\mathrm{R}_\Phi^{\min}$, the excess value in complexity can, via the relation between redundancy and free energy hinted at above, be used to perform some work, hence the physical system is not in equilibrium. A more thorough analysis on this point is needed. Anyhow, we maintain the thesis that *thermodynamical equilibrium is the same as game theoretical equilibrium*.

The optimal strategies for Player II do not seem to have an analogy in established principles of statistical physics. They almost always exist - in contrast to the MaxEnt-distributions. Often it turns out that one and the same distribution is optimal for Player I as well as for Player II. Such a distribution is said to be a *bi-optimal strategy*. The main facts about γ_Φ are as follows:

Theorem 1. (GTE-fundamentals for γ_Φ). *Assume that* $\mathrm{S}_\Phi^{\max} < \infty$.

(i) (Nash equilibrium properties). *If* (P^*, Q^*) *is a Nash equilibrium pair, then* $Q^* = P^*$, γ_Φ *is in equilibrium,* P^* *is the unique optimal strategy for Player I, hence the unique* MaxEnt-*distribution, and* Q^* *is the unique optimal strategy for Player II. In particular,* P^* *is the unique bi-optimal distribution. Furthermore, for any* $P \in \mathcal{S}_I$, *and* $Q \in \mathcal{S}_{II}$,

$$\mathrm{S}_\Phi(P) + \mathrm{D}_\Phi(P\|P^*) \le \mathrm{S}_\Phi^{\max} = \mathrm{R}_\Phi^{\min} \le \mathrm{R}_\Phi(Q) - \mathrm{D}_\Phi(P^*\|Q)\,. \tag{7}$$

(ii) (necessity of Nash equilibrium). *If* γ_Φ *is in equilibrium and if both players have optimal strategies, then there exists a Nash equilibrium pair for the game, hence also a bi-optimal strategy.*

(iii) (identification). *A distribution,* P^*, *is the bi-optimal distribution if and only if* $P^* \in \mathcal{S}_I$ *and*

$$\Phi(P\|P^*) \le \mathrm{S}_\Phi(P^*) \text{ for all } P \in \mathcal{S}_I\,. \tag{8}$$

(iv) (robustness). *If the distribution* P^* *is consistent* ($P^* \in \mathcal{S}_I$) *and robust, then* (P^*, P^*) *is a Nash equilibrium pair.*

The proof is simple, but will not be given here. For standard Shannon entropy, the proof can be deduced from [5] and references there.

No convexity assumptions regarding $\mathcal{S}_I$ or concavity assumptions regarding the Φ-entropy function are required. But, of course, such conditions will be fulfilled for many applications of interest.

In (7) we recognize "*Pythagorean type*" inequalities.

3 Entropy measures via complexity

It is extremely easy to suggest concrete measures of complexity which satisfy our axioms. In a sense, Theorem 1 is "too general". It only deals with situations when equilibrium holds, and says nothing about how often this happens. For instance, one might want γ_Φ to be in equilibrium for any convex preparation $\mathcal{S}_I$. This points to concavity of the Φ-entropy function as a natural requirement. We may also require that Φ-redundancy is convex (in the first variable or even jointly).

We shall now demonstrate one way of obtaining non-Boltzmann-Shannon entropies by "digging one level deeper" via suitably chosen measures of complexity. For any pair of real numbers, α, β, we define the (α, β)-*logarithmic function* on $]0, \infty[$ by

$$\ln_{\alpha,\beta} x = \frac{x^\beta - x^\alpha}{\beta - \alpha} \text{ if } \beta \neq \alpha \,, \tag{9}$$

and $\ln_{\alpha,\beta} x = x^\alpha \ln x$ if $\beta = \alpha$. We assume below that $\alpha \neq \beta$ as the case of equality only leads to standard complexity (average code length), standard entropy (Boltzmann-Shannon entropy) and standard divergence (relative entropy or Kullback-Leibler divergence).

We point out that $\ln_{\alpha,\beta} 1 = 0$ (*normalization*), $\ln_{\alpha,\beta} = \ln_{\beta,\alpha}$ (*symmetry*), $\ln_{\alpha+c,\beta+c} x = x^c \ln_{\alpha,\beta} x$ (*translation property*) and that the *functional equation*

$$\ln_{\alpha,\beta}(xy) = y^\alpha \ln_{\alpha,\beta} x + x^\beta \ln_{\alpha,\beta} y \tag{10}$$

holds.

First use the functional equation with $x = p_i$ and $y = \frac{q_i}{p_i}$ together with natural generalizations of standard quantities in order to ensure that the linking identity (5) holds. This leads to the following objects:

$$\Phi_{\alpha,\beta}(P\|Q) = \sum_{i\in\mathbb{A}} \phi(p_i) p_i^{1+\alpha-\beta} q_i^{-\alpha} (-\ln_{\alpha,\beta} q_i)\,, \tag{11}$$

$$\mathrm{S}_{\alpha,\beta}(P) = \sum_{i\in\mathbb{A}} \phi(p_i) p_i^{1-\beta} (-\ln_{\alpha,\beta} p_i)\,, \tag{12}$$

$$\mathrm{D}_{\alpha,\beta}(P\|Q) = \sum_{i\in\mathbb{A}} \phi(p_i) p_i^{1+\alpha} q_i^{-\alpha} (-\ln_{\alpha,\beta} \frac{q_i}{p_i})\,, \tag{13}$$

where ϕ is some non-negative function on $[0,1]$.

With these definitions, $\Phi_{\alpha,\beta}$ and $\mathrm{S}_{\alpha,\beta}$ are non-negative and the linking identity holds. It remains to investigate if ϕ can be chosen such that $\mathrm{D}_{\alpha,\beta}$ is non-negative and only vanishes on the diagonal. It lies nearby to achieve this by ensuring that $\mathrm{D}_{\alpha,\beta}$ is a Csiszár f-divergence, i.e. of the form $\sum_{i\in\mathbb{A}} p_i f(\frac{q_i}{p_i})$ for a strictly convex function f with $f(1)=0$. This forces ϕ to be constant and we may then take $\phi \equiv 1$. A simple computation then leads to the following result, which may be taken as a supportive argument in favour of Tsallis entropy.

Theorem 2. (Tsallis entropy via complexity). *If (11)-(13) are to define a genuine measure of complexity together with its associated entropy and redundancy functions, and if we require that* $\mathrm{D}_{\alpha,\beta}$ *be a Csiszár* f*-divergence then, necessarily,* $\beta-\alpha\leq 1$ *and* ϕ *must be a positive constant function. Assume now that* $\beta-\alpha\leq 1$*. Then, with* $\phi\equiv 1$*, all stated properties hold. As* $\Phi_{\alpha,\beta}=\Phi_{\alpha+c,\beta+c}$ *for any constant* c*, this defines a one-parameter family of functions. If we put* $q=1-\beta+\alpha$*, then* $q\geq 0$ *and* $\Phi_{\alpha,\beta}=\Phi_{q,1}\,,\mathrm{S}_{\alpha,\beta}=\mathrm{S}_{q,1}$ *and* $\mathrm{D}_{\alpha,\beta}=\mathrm{D}_{q,1}$ *and these functions are given by*

$$\Phi_{q,1}(P\|Q) = \frac{1}{1-q}\sum_{i\in\mathbb{A}} p_i^q\big(1-q_i^{1-q}\big)\,, \tag{14}$$

$$\mathrm{S}_{q,1}(P) = \frac{1}{1-q}\sum_{i\in\mathbb{A}} p_i\big(p_i^{q-1}-1\big) = \frac{1}{1-q}\Big(\sum_{i\in\mathbb{A}} p_i^q - 1\Big)\,, \tag{15}$$

$$\mathrm{D}_{q,1}(P\|Q) = \frac{1}{1-q}\sum_{i\in\mathbb{A}} p_i\Big(1-(\frac{p_i}{q_i})^{q-1}\Big) = \frac{1}{1-q}\Big(1-\sum_{i\in\mathbb{A}} p_i(\frac{p_i}{q_i})^{q-1}\Big)\,. \tag{16}$$

In (15) we recognize the q-entropy of Tsallis. Note also that for $q=\frac{1}{2}$ (16) gives us the popular *Hellinger divergence.*

The discussion shows that, if the approach as explained based on (10) is taken as starting point, Tsallis q-entropy emerges as the only possible family of decently behaved entropy measures. But, of course, one could suggest different approaches. One other possibility is to base the definitions on the *κ-logarithms* introduced in Kaniadakis [3] by $\ln_{\{\kappa\}} x = \ln_{-\kappa,\kappa} x$. It is convenient to introduce, as in Kaniadakis [4], the *κ-deformed product* by

$$\ln_{\{\kappa\}}(x \underset{\kappa}{\otimes} y) = \ln_{\{\kappa\}} x + \ln_{\{\kappa\}} y\,. \tag{17}$$

Theorem 3. (κ-entropy via complexity). *In case $-1 \le \kappa \le 1$, the formulas*

$$\Phi_{\{\kappa\}}(P\|Q) = \sum_{i\in\mathbb{A}} p_i \ln_{\{\kappa\}} \big(\frac{p_i}{q_i} \underset{\kappa}{\otimes} \frac{1}{p_i}\big)\,, \tag{18}$$

$$\mathrm{S}_{\{\kappa\}}(P) = \sum_{i\in\mathbb{A}} p_i \ln_{\{\kappa\}} \frac{1}{p_i}\,, \tag{19}$$

$$\mathrm{D}_{\{\kappa\}}(P\|Q) = \sum_{i\in\mathbb{A}} p_i \ln_{\{\kappa\}} \frac{p_i}{q_i}\,, \tag{20}$$

define a genuine complexity measure with its associated entropy- and divergence-functions and $\mathrm{S}_{\{\kappa\}}$ *is strictly concave and* $\mathrm{D}_{\{\kappa\}}$ *is a Csiszár f-divergence. For no other values of κ does this hold.*

In (19) we recognize the κ-entropy function of Kaniadakis. You may note that for $\kappa = 1$, $\mathrm{S}_{\{\kappa\}}$ is proportional to χ^2-divergence.

References

[1] J. Havrda and F. Charvát. Quantification method of classification processes. Concept of structural a-entropy. *Kybernetika*, 3:30–35, 1967.

[2] E. T. Jaynes. *Probability Theory - The Logic of Science.* Cambridge University Press, Cambridge, 2003.

[3] G. Kaniadakis. Non-linear kinetics underlying generalized statistics. *Physica A*, 296:405–425, 2001.

[4] G. Kaniadakis. Statistical mechanics in the context of special relativity. *Physical Review E*, 66:056125,1–17, 2002.

[5] F. Topsøe. Maximum entropy versus minimum risk and applications to some classical discrete distributions. *IEEE Trans. Inform. Theory*, 48:2368–2376, 2002.

[6] C. Tsallis. Possible generalization of Boltzmann-Gibbs statistics. *J. Stat. Physics*, 52:479, 1988. See http://tsallis.cat.cbpf.br/biblio.htm for a complete and updated bibliography.