Entropy and Equilibrium via Games of Complexity

Flemming Topsøe *
Department of Mathematics
University of Copenhagen
topsoe@math.ku.dk

Abstract

It is suggested that *thermodynamical equilibrium equals game theoretical equilibrium*. Aspects of this thesis are discussed. The philosophy is consistent with maximum entropy thinking of Jaynes, but goes one step deeper by deriving the maximum entropy principle from an underlying game theoretical principle. The games introduced are based on measures of complexity. Entropy is viewed as minimal complexity. It is demonstrated that Tsallis entropy (q-entropy) and Kaniadakis entropy (κ-entropy) can be obtained in this way, based on suitable complexity measures. A certain unifying effect is obtained by embedding these measures in a two-parameter family of entropy functions.

Keywords. Measure of complexity, maximum entropy, game theoretical equilibrium, Nash equilibrium, Tsallis entropy, Kaniadakis entropy.

1 Introduction, background

The *Maximum Entropy Principle* (MaxEnt) has been studied extensively by Jaynes from 1957 onwards, cf. [1], [2], [3] and [4]. MaxEnt dictates that the least biased probability distribution among the set of consistent distributions in some context, and hence the one best suited for predictions, is the one with maximal entropy. Following Jaynes, one should not think of

*Research supported by INTAS, project 00-738, and by the Danish Natural Science Research Council.

the ensuing distribution, the MaxEnt-*distribution*, as the "true" distribution, but rather as the distribution which best models our *knowledge* about the system under study. This distribution is the one best suited for predictions. If we could suggest better predictions than what can be obtained based on the MaxEnt-distribution that would reflect that *we actually knew something more*.

In [5] and [6] the author pointed out that a principle of *Game Theoretical Equilibrium* leads to MaxEnt and to a related principle of updating (based on a given prior distribution). In [7], [8], [9] and [10] the reader finds more recent publications on these issues.

The research alluded to above concerns the entropy measure associated with the names of Boltzmann, Gibbs and Shannon. When, below, we have this measure of entropy in mind, we refer to "classical entropy".

Based on the apparent relevance of information theory as e.g. in the MaxEnt-principle, the importance of this field as a basis for statistical physics is growing and there are tendencies that this is becoming more widely recognized. At least, this is the authors impression from participation in the NEXT2003 conference. The present contribution is in line with these tendencies. It emphasizes the role of measures of complexity and associated information theoretical games. Instead of focusing on the finer mathematical details (as in [8]) we concentrate on the basic ideas. The approach is axiomatic and allows for non-classical entropy measures such as the presently very popular *Tsallis entropies* (or *q-entropies*), cf. [11], and the more recently proposed *Kaniadakis entropies* (or κ*-entropies*), cf. [12] and [13].

2 Games of complexity

In this section we focus on concepts related to the information theoretical games we shall study.

Let $\mathbb{A}$, neutrally referred to as the *alphabet*, be a finite or countably infinite set. Introduce two players, *Player I* ("Nature", "the system" or $\cdots$) and *Player II* which we, less abstractly, identify with *the physicist*. Let the *strategies* available to Player I be given by a set $\mathcal{S}_I$ of distributions (always probability distributions) over $\mathbb{A}$. Often, we refer to $\mathcal{S}_I$ as *the preparation* [1].

As strategy set $\mathcal{S}_{II}$ for Player II we take the set of *all* distributions over $\mathbb{A}$. If $Q \in \mathcal{S}_{II}$, we may think of Q as a means to *describe* or to *code* outcomes

[1] 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. As a natural preparation, consider all distributions over $\mathbb{A}$ with a prescribed mean energy.

in $\mathbb{A}$ resulting from observations of the physical system under study. A few comments on codes related to classical entropy are in place. Technically, the concept can be avoided, but it *is* useful and supports the understanding. A *code*, more precisely an *idealized code*, is a map $\kappa : \mathbb{A} \to [0, \infty]$ such that *Kraft's equality*

$$\sum_{i \in \mathbb{A}} e^{-\kappa(i)} = 1 \tag{1}$$

holds. Intuitively, $\kappa(i)$ should be thought of as the length of the "codeword" which we imagine κ assigns to $i \in \mathbb{A}$. The good sense in this interpretation is well known in information theory. For a quick introduction, see [9]. The code *adapted* to $Q \in \mathcal{S}_{II}$ is denoted κ_Q and given by

$$\kappa_Q(i) = -\ln Q(i) \text{ for } i \in \mathbb{A}\,. \tag{2}$$

We now assume that there is given a *measure of complexity* which, to any pair (P, Q) of distributions, assigns a number $\Phi(P\|Q) \in [0, \infty]$, thought of as the *complexity* (or difficulty) for Player II involved in observations of the physical system when he uses the strategy Q and when Player I has chosen the strategy $P \in \mathcal{S}_I$. We sometimes refer to Q as the *reference* and may then say that $\Phi(P\|Q)$ is the *complexity of P with reference Q*. We have used the seperator $\|$ in the notation for values of Φ as the two distributions that appear in $\Phi(P\|Q)$ actually have different roles. Also, Φ is normally *not* symmetric.

By S_Φ we denote the Φ-*entropy* which is introduced for every distribution P by

$$\mathrm{S}_\Phi(P) = \inf_{Q \in \mathcal{S}_{II}} \Phi(P\|Q)\,. \tag{3}$$

Thus, by definition, *entropy equals minimal complexity.*

The key elements in the model we shall study are $(\mathbb{A}, \mathcal{S}_I, \Phi)$, the alphabet, the preparation and the complexity measure. We always take the set of all distributions over $\mathbb{A}$ as the strategy set for Player II.

We assume that the following axioms are fulfilled:

$$\mathrm{S}_\Phi(P) < \infty \text{ for } P \in \mathcal{S}_I\,, \tag{4}$$

$$\mathrm{S}_\Phi(P) = \Phi(P\|P) \text{ for } P \in \mathcal{S}_I\,, \tag{5}$$

$$\Phi(P\|Q) > \mathrm{S}_\Phi(P) \text{ for } P \in \mathcal{S}_I\,, Q \in \mathcal{S}_{II}\,, Q \neq P. \tag{6}$$

The basic axioms (5) and (6) are quite natural. Indeed, they express that the complexity of P is the smallest when P itself is taken as reference. And regarding (4), this is a truly innocent axiom. Therefore, it is understood in the sequel without further mentioning that this condition is fulfilled for the preparations we shall deal with.

To any pair $(P, Q) \in \mathcal{S}_I \times \mathcal{S}_{II}$ we can associate the Φ-*redundancy* defined by

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

i.e. as *actual complexity minus minimal complexity.* This identity is best written as

$$\Phi(P\|Q) = \mathrm{S}_\Phi(P) + \mathrm{D}_\Phi(P\|Q) \tag{7}$$

and is referred to as the *linking identity.*

Clearly, $\mathrm{D}_\Phi(P\|Q) \geq 0$ and equality holds if and only if $P = Q$. Instead of redundancy, we often use the terminology *divergence* for D_Φ. It can be seen as a kind of, typically, non-symmetric distance between distributions, but the interpretation indicated above as a redundancy is actually more appropriate for our purposes.

By the *classical complexity measure* we understand Φ_{clas} defined as *average code length*, i.e. as $\langle \kappa_Q, P \rangle$ or, directly in terms of Q,

$$\Phi_{\mathrm{clas}}(P\|Q) = \sum_{i\in\mathbb{A}} p_i \ln \frac{1}{q_i}\,. \tag{8}$$

(Here, the p_i and q_i denote the point probabilities of P, respectively Q).

For brevity, we denote by $\mathrm{S}_{\mathrm{clas}}$ and Φ_{clas} the entropy- and redundancy mesures associated with Φ_{clas}. Entropy then is the classical *Boltzmann-Gibbs-Shannon entropy* and redundancy the well known *relative entropy* or *Kullback-Leibler divergence*, cf. [14], [15]:

$$\mathrm{S}_{\mathrm{clas}}(P) = \sum_{i\in\mathbb{A}} p_i \ln \frac{1}{p_i}\,, \tag{9}$$

$$\mathrm{D}_{\mathrm{clas}}(P\|Q) = \sum_{i\in\mathbb{A}} p_i \ln \frac{p_i}{q_i}\,. \tag{10}$$

The fact needed to prove (9) and (10) is the inequality $\mathrm{D}_{\mathrm{clas}}(P\|Q) \geq 0$ with equality only if $P = Q$ (use (10) with $-p_i \ln \frac{q_i}{p_i}$ in place of $p_i \ln \frac{p_i}{q_i}$ and apply Jensen's inequality). This also shows that our axioms are indeed fulfilled.

Now return to the case of a general measure of complexity, Φ, and introduce the two-person zero-sum game $\gamma_\Phi = \gamma_\Phi(\mathcal{S}_I)$ with Player I and Player II as the two players and with Φ as objective function, viewed as a cost to Player II. Thus, Player II, the physicist, aims at achieving a low complexity whereas Player I has the opposite aim.

Applying the 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 (3), 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* which is defined by

$$\mathrm{S}_\Phi^{\max} = \sup_{P \in \mathcal{S}_I} \mathrm{S}_\Phi(P) . \tag{11}$$

Similarly, Player II will associate a certain *risk* to any strategy $Q \in \mathcal{S}_{II}$. This risk is given by

$$\mathrm{R}_\Phi(Q) = \sup_{P \in \mathcal{S}_I} \Phi(P\|Q) . \tag{12}$$

Player II will, therefore, consider $Q^* \in \mathcal{S}_{II}$ to be an *optimal strategy* if $\mathrm{R}_\Phi(Q^*) = \mathrm{R}_\Phi^{\min}$, the *minimum risk value* which is defined by

$$\mathrm{R}_\Phi^{\min} = \inf_{Q \in \mathcal{S}_{II}} \mathrm{R}_\Phi(Q) . \tag{13}$$

By the general maximin/minimax inequality,

$$\mathrm{S}_\Phi^{\max} \le \mathrm{R}_\Phi^{\min} . \tag{14}$$

If equality holds in (14), we say that the game is in *equilibrium* and the common value is the *value* of the game. In such cases, *maximum entropy equals minimum risk* or, in other terms, *maximum entropy equals minimum complexity* (more precisely, minimum *guaranteed* complexity).

The *principle of game theoretical equilibrium* (GTE) simply 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. But the corresponding principle for Player II, call it the *minimum complexity principle*, ought to be just as interesting. It reflects what best the physicist can do.[2] In the classical case this second principle is, in a sense, not needed as usually the two optimal objects which you are led to consider agree and lead to the same value, $\mathrm{S}_\Phi^{\max} = \mathrm{R}_\Phi^{\min}$ (see Theorems 1 and 2 below and apply them to standard examples as in [8]). But when you turn to non-extensive statistical physics the situation may well be different and one should be aware of the possible significance of the point of view offered by the game theoretical approach.

If there is a discrepancy between $\mathrm{S}_\Phi^{\max}$ and $\mathrm{R}_\Phi^{\min}$ this is interpreted in the classical case that the excess value in complexity can be used to perform some work, hence the physical system is not in equilibrium. This interpretation

[2] In information theory this kind of thinking is usually referred to as the principle of *minimum description length*, a principle promoted by Rissanen, cf. the survey article [16].

will need a more thorough analysis but we shall not go into that here [3]. When these considerations are sound, they point to the thesis that *thermodynamical equilibrium is the same as game theoretical equilibrium.*

Often, 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 optimal strategies for Player II do not seem to have an analogy in established principles of statistical physics, at least not directly. In the classical case, the optimal strategies for Player II almost always exist – and are actually best understood in terms of coding – whereas the MaxEnt-distributions may fail to exist, cf. Ingarden and Urbanik [17] and the already cited papers [5] and [8].

3 Criteria for equilibrium

We need a general notion of equilibrium from game theory: A pair $(P^*, Q^*) \in \mathcal{S}_I \times \mathcal{S}_{II}$ is a *Nash equilibrium pair* if the *saddle value inequalities*

$$\Phi(P\|Q^*) \leq \Phi(P^*\|Q^*) \leq \Phi(P^*\|Q) \tag{15}$$

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 final concept is useful, at least for the classical case: A strategy Q for Player II is said to be *robust* if the complexity that may appear when using this strategy is finite and independent of Player I's choice of strategy, i.e. if, for some finite constant h, $\Phi(P\|Q) = h$ for all $P \in \mathcal{S}_I$. The set of all robust strategies forms the *exponential family* associated with the game γ_Φ[4].

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* $P \in \mathcal{S}_I$*, the trivial inequality* $\mathrm{S}_\Phi(P) \leq \mathrm{S}_\Phi^{\max}$ *can be sharpened to*

$$\mathrm{S}_\Phi(P) + \mathrm{D}_\Phi(P\|P^*) \leq \mathrm{S}_\Phi^{\max}\,, \tag{16}$$

[3]This depends, as indicated to me by Peter Harremoës, on the realization that, at least for the classical case, information in terms of bits, typically via redundancy, can be related to free energy.

[4]The family will be recognized as the usual family that emerges when studying standard cases related to the canonical or grand canonical ensemble of statistical thermodynamics.

and, for any $Q \in \mathcal{S}_{II}$, the trivial inequality $\mathrm{R}_\Phi^{\min} \le \mathrm{R}_\Phi(Q)$ can be sharpened to

$$\mathrm{R}_\Phi^{\min} + \mathrm{D}_\Phi(P^*\|Q) \le \mathrm{R}_\Phi(Q)\,. \tag{17}$$

(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). *Let P^* be a distribution. If $P^* \in \mathcal{S}_I$ and*

$$\Phi(P\|P^*) \le \mathrm{S}_\Phi(P^*) \text{ for all } P \in \mathcal{S}_I\,, \tag{18}$$

then γ_Φ is in equilibrium and P^* is the bi-optimal distribution.

(iv) (robustness). *If the distribution P^* is consistent ($P^* \in \mathcal{S}_I$) and robust, then (P^*, P^*) is a Nash equilibrium pair, hence γ_Φ is in equilibrium and has a bi-optimal strategy, viz P^*.*

Proof. (i): Assume that (P^*, Q^*) is a Nash equilibrium pair. By choosing P^* for Q in (15), we realize that $\Phi(P^*\|Q^*) \le \mathrm{S}_\Phi(P^*)$. By (6) this implies that $P^* = Q^*$, hence $\Phi(P^*\|Q^*) = \mathrm{S}_\Phi(P^*)$. Then the first inequality of (15) shows that $\mathrm{R}_\Phi(P^*) \le \mathrm{S}_\Phi(P^*)$. Therefore,

$$\mathrm{R}_\Phi^{\min} \le \mathrm{R}_\Phi(P^*) \le \mathrm{S}_\Phi(P^*) \le \mathrm{S}_\Phi^{\max}$$

and by (14) we conclude that here, equality must hold throughout. Thus, γ_Φ is in equilibrium and P^* is a bi-optimal strategy.

By (7), the left-hand-side in (16) equals $\Phi(P\|P^*)$ which is bounded above by $\mathrm{S}_\Phi(P^*) = \mathrm{S}_\Phi^{\max}$, hence (16) holds.

To prove (17) note that

$$\begin{aligned}\mathrm{R}_\Phi^{\min} + \mathrm{D}_\Phi(P^*\|Q) &\le \mathrm{R}_\Phi(P^*) + \mathrm{D}_\Phi(P^*\|Q)\\ &= \mathrm{S}_\Phi(P^*) + \mathrm{D}_\Phi(P^*\|Q)\\ &= \Phi(P^*\|Q)\\ &\le \mathrm{R}_\Phi(Q)\,.\end{aligned}$$

By (16) and (17), the uniqueness assertions regarding optimality of P^* follow. All parts of (i) are now verified.

To prove (ii), assume that $P^* \in \mathcal{S}_I\,, Q^* \in \mathcal{S}_{II}$ and that $\mathrm{S}_\Phi(P^*) = \mathrm{S}_\Phi^{\max} = \mathrm{R}_\Phi^{\min} = \mathrm{R}_\Phi(Q^*)$. Then, for $P \in \mathcal{S}_I$ and $Q \in \mathcal{S}_{II}$ we have

$$\begin{aligned}\Phi(P\|Q^*) &\le \mathrm{R}_\Phi(Q^*) = \mathrm{S}_\Phi(P^*) \le \Phi(P^*\|Q^*)\\ &\le \mathrm{R}_\Phi(Q^*) = \mathrm{S}_\Phi(P^*) \le \Phi(P^*\|Q)\,,\end{aligned}$$

and we conclude that (P^*, Q^*) is a Nash equilibrium pair. This proves (ii).

The property (iii) follows from the analysis above regarding property (i).

To prove (iv) note that if $\Phi(P\|P^*) = h$ for all $P \in \mathcal{S}_I$ and $P^* \in \mathcal{S}_I$, then $h = \Phi(P^*\|P^*) = \mathrm{S}_\Phi(P^*)$, hence (iii) applies and the desired result follows. $\square$

NOTE 1. No convexity assumptions regarding the preparation $\mathcal{S}_I$ or concavity assumptions regarding the Φ-entropy function are required in the result. But, of course, such conditions will be fulfilled for many applications of interest.

NOTE 2. The inequality (16) is of the "*Pythagorean type*" and goes back to Čencov, [18], and to Csiszár, [19]. The "dual" inequality (17) appeared in [5].

NOTE 3. The proof of (ii) is a general argument that applies to any two-person zero-sum game. Thus, if such a game is in equilibrium and allow optimal strategies for both players, then the two optimal strategies must form a Nash equilibrium pair.

NOTE 4. If γ_Φ is in equilibrium and admits a bi-optimal distribution, we see from (iii) that the strategy set $\mathcal{S}_I$ can be extended without destroying the equilibrium properties. In fact, the maximal extension consists of all distributions P which satisfy the inequality in (18).

NOTE 5. Regarding the soundness of the GTE principle it is appropriate, not so much to think of the game as such but more so to think of the actual Nash equilibrium pair as an expression of physical equilibrium.

NOTE 6. The criterion (iv) – which in practise does not even require the introduction of Lagrange multipliers in order to identify the MaxEnt-distribution – is very powerful when it applies, cf. [8]. However, one cannot expect that it works when dealing with non-extensive entropy functions. The usefulness is limited to the classical case and to preparations defined via moment constraints as in the discussion of the canonical and the grand canonical ensemble.

It is extremely easy to suggest concrete measures of complexity which satisfy our axioms. We can even achieve that any given non-negative function of P occurs as the Φ-entropy function. In fact, just add to the given function a suitable function $\mathrm{D}(P\|Q)$ such as a true metric or a *Csiszár f-divergence* (for these, see the next section). Of course, not all complexity measures arising in this way are of interest. The fact is that though Theorem 1 is important, there are other aspects which should be considered when suggesting reasonable entropy functions.

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 $\gamma_\Phi(\mathcal{S}_I)$ to be in equilibrium for any convex preparation $\mathcal{S}_I$ or at least that the MaxEnt-distribution exists for these preparations. This points to concavity of the Φ-entropy function as a natural requirement. Other consequences, e.g. related to a study of certain derived games with a given prior, point to the good sense in requiring that Φ-redundancy is convex (in the first variable or even jointly).

In the next section we suggest a widely applicable method to generate suitable measures of complexity. Here we point to a general property which guarantees existence of equilibrium for a large class of preparations.

Theorem 2. (Equilibrium for convex preparations). *Assume that, besides axioms (5) and (6), $\Phi(P\|Q)$ is concave in P for any distribution Q. Then S_Φ is concave. In fact, the following strong form of the concavity inequality holds: For any mixture $\overline{P} = \sum \alpha_\nu P_\nu$ (α_ν's non-negative with sum 1),*

$$\mathrm{S}_\Phi\left(\sum \alpha_\nu P_\nu\right) \geq \sum \alpha_\nu \, \mathrm{S}_\Phi(P_\nu) + \sum \alpha_\nu \, \mathrm{D}_\Phi(P_\nu\|\overline{P})\,. \tag{19}$$

Furthermore, under mild and natural continuity conditions (see the proof below), the game $\gamma_\Phi(\mathcal{S}_I)$ is in equilibrium and the bi-optimal distribution exists if only $\mathcal{S}_I$ is convex.

Proof. By the linking identity, the right hand side of (19) can be written as $\sum \alpha_\nu \Phi(P_\nu\|\overline{P})$ which is dominated by $\Phi(\sum \alpha_\nu P_\nu\|\overline{P}) = \Phi(\overline{P}\|\overline{P}) = \mathrm{S}_\Phi(\overline{P})$, the left hand side of (19).

Now assume that $\mathcal{S}_I$ is convex and that $\mathcal{S}_\Phi(P^*) = \mathrm{S}_\Phi^{\max}$ with $P^* \in \mathcal{S}_I$ (if, e.g. $\mathcal{S}_I$ is compact and convex and S_Φ continuous, the distribution P^* with the stated properties exists). Let $P \in \mathcal{S}_I$. Then, for every $\varepsilon \in [0,1]$, $P_\varepsilon \in \mathcal{S}_I$ with $P_\varepsilon = (1-\varepsilon)P^* + \varepsilon P$. Therefore, $\mathrm{S}_\Phi(P_\varepsilon) \leq \mathrm{S}_\Phi^{\max} = \mathrm{S}_\Phi(P^*)$. By (19),

$$(1-\varepsilon)\,\mathrm{S}_\Phi(P^*) + \varepsilon\,\mathrm{S}_\Phi(P) + \varepsilon\,\mathrm{D}_\Phi(P\|P_\varepsilon) \leq \mathrm{S}_\Phi(P^*)\,.$$

After rearrangement, division by ε and appeal to the linking identity we get $\Phi(P\|P_\varepsilon) \leq \mathrm{S}_\Phi(P^*)$. Assuming that Φ is lower semi-continuous in the second variable, we conclude that $\Phi(P\|P^*) \leq \mathrm{S}_\Phi(P^*)$ by letting ε tend to 0. As this holds for every $P \in \mathcal{S}_I$, the desired result follows by (iii) of Theorem 1. $\square$

Note. Following ideas in [5] or [8], it is possible to refine the argument and establish equilibrium in cases when the MaxEnt-distribution does not exist.

4 Generation of complexity triples

We now propose to look closer into a simple and quite general class of *complexity triples* by which we mean triples $(\Phi, \mathrm{S}, \mathrm{D})$ consisting of a complexity measure and its associated entropy and redundancy functions. Rather than redundancy we shall in the sequel talk mainly about divergence since the approach we shall adopt depends on properties of *Csiszár f-divergences*. We only need a few basic properties, some of which are derived below for the readers convenience. For more information, see Liese and Vajda, [20] or, more easily accessibly, Österreicher and Vajda, [21], and references therein.

We find it convenient to introduce the notion of a *generator* as a real-valued function f defined on $[0, \infty[$ such that

$$f(0) = f(1) = 0\,, \tag{20}$$

$$f'(1) = 1\,, \tag{21}$$

$$f \text{ is strictly convex} \tag{22}$$

and, furthermore, we assume that f is "smooth" , say continuous on $[0, \infty[$ and twice continuously differentiable on $]0, \infty[$ (possibly with less strict conditions, e.g. allowing for exceptional points regarding the requirement of differentiability).

Equation (21) is just a normalization condition. By (20) and (22), f is negative on $]0, 1[$, positive on $]1, \infty[$. Putting $f(\infty) = \infty$, f maps $[0, \infty]$ continuously onto the interval $[a, \infty]$ where $a > -\infty$ is the minimal value of f. The *limit slopes*

$$f'(0) = \lim_{x \to 0^+} \frac{f(x)}{x}\ ,\ f'(\infty) = \lim_{x \to \infty} \frac{f(x)}{x} \tag{23}$$

play a certain role. Often, $f'(0) = -\infty$, $f'(\infty) = \infty$.

As a typical example, consider the *classical generator*, by which we mean the all-important function of information theory, $f(x) = x \ln x$. Other concrete examples will be discussed in the next section.

For any generator f, we define the *associated triple* consisting of *f-complexity*, *f-entropy* and *f-divergence* by

$$\Phi_f(P\|Q) = \sum_{i \in \mathbb{A}} \big(q_i f(\frac{p_i}{q_i}) - f(p_i) \big)\,, \tag{24}$$

$$\mathrm{S}_f(P) = -\sum_{i \in \mathbb{A}} f(p_i) \tag{25}$$

$$\mathrm{D}_f(P\|Q) = \sum_{i \in \mathbb{A}} q_i f(\frac{p_i}{q_i})\,. \tag{26}$$

We use the standard convention $0f(\frac{0}{0}) = 0$, thus all sums above could be replaced by sums only over $i \in \mathbb{A}$ with $p_i > 0$. If $p > 0$, we define $0f(\frac{p}{0})$ by continuity, i.e. by

$$0f(\frac{p}{0}) = \lim_{q \to 0} qf(\frac{p}{q}) = pf'(\infty)\,. \tag{27}$$

Clearly, S_f is well-defined and $0 \leq \mathrm{S}_f \leq \infty$. Also D_f is well-defined. In fact, an application of Jensens inequality shows that the sum of negative terms in (26) is bounded below by $Q^- f(P^-/Q^-)$ where P^-, respectively Q^- is the sum of p_i, respectively q_i with $p_i < q_i$. A further application of Jensens inequality shows that $\mathrm{D}_f(P\|Q) \geq 0$ with equality if and only if $P = Q$. Then, for any preparation $\mathcal{S}_I$ (which satisfies (4)), axioms (5) and (6) are satisfied. We may, therefore, following the approach of Section 2, consider the associated game, denoted $\gamma_f(\mathcal{S}_I)$.

Regarding technical issues, we note that $\Phi_f\,, \mathrm{S}_f$ and D_f are lower semi-continuous (using the topology of pointwise convergence in the set of distributions). This property is important for certain finer investigations (not pursued here – for an indication, see [8]).

As a final comment to the definitions, note that, clearly, the linking identity holds:

$$\Phi_f(P\|Q) = \mathrm{S}_f(P) + \mathrm{D}_f(P\|Q)\,. \tag{28}$$

It is convenient, given the generator f, to consider also the *Csiszár-dual* [5] which is defined as the function $\widetilde{f}$ given by

$$\widetilde{f}(x) = xf\big(\frac{1}{x}\big)\,,\ \ 0 \leq x \leq \infty\,. \tag{29}$$

Usual conventions are observed, e.g. $0 \cdot \infty = 0$. For the classical generator, we see that $\widetilde{f}(x) = \ln \frac{1}{x}$, the (second-) most important function of information theory. Expressed in terms of the Csiszár-dual, the definitions (24)-(26) take the form:

$$\Phi_f(P\|Q) = \sum_{i \in \mathbb{A}} p_i\Big(\widetilde{f}(\frac{q_i}{p_i}) - \widetilde{f}(\frac{1}{p_i})\Big)\,, \tag{30}$$

$$\mathrm{S}_f(P) = -\sum_{i \in \mathbb{A}} p_i \widetilde{f}(\frac{1}{p_i})\,, \tag{31}$$

$$\mathrm{D}_f(P\|Q) = \sum_{i \in \mathbb{A}} p_i \widetilde{f}(\frac{q_i}{p_i})\,. \tag{32}$$

Some general remarks on Csiszár-duals are in place. Clearly, $\underset{\sim}{f}^{\sim\sim} = f$. The duality operation is easy to illustrate geometrically. Indeed, $\widetilde{f}(x)$ is the

[5]not to be confused with the *Fenchel conjugate* of convexity theory .

slope of the chord OA corresponding to the arguments 0 and x^{-1}, cf. Figure 1. Also, $\widetilde{f}'(x)$ has a simple interpretation as the ordinate of the point B on Figure 1 (this point is obtained by intersecting the tangent to f in A with the second axes). These interpretations show that $\widetilde{f}$ is decreasing and $\widetilde{f}'$ increasing, in particular, $\widetilde{f}$ is convex, in fact strictly convex. This also follows analytically since

$$\widetilde{f}''(x) = x^{-3} f''(x^{-1}) \,. \tag{33}$$

Adding a few more simple observations we can summarize the discussion as follows:

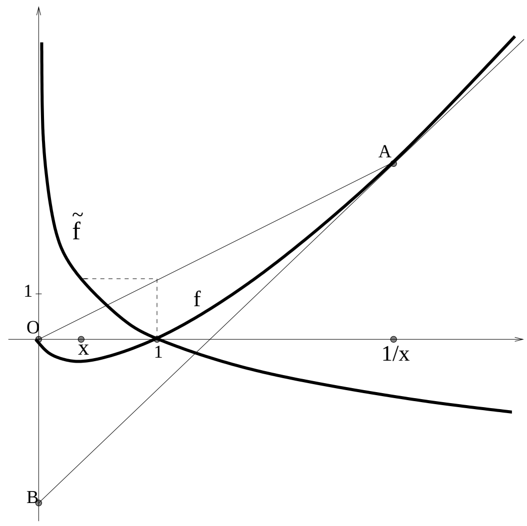

Figure 1. The functions f and $\widetilde{f}$.

Lemma 1. *The Csiszár-dual $\widetilde{f}$ of the generator f is strictly decreasing and strictly convex on* $[0, \infty]$, $\widetilde{f}(0) = f'(\infty)\,, \widetilde{f}(1) = 0\,, \widetilde{f}(\infty) = f'(0)$ *and* $\widetilde{f}'(1) = -1$.

Inspired by the form of (30), we define *local complexity*, ϕ_f, and *contributed complexity*, ψ_f, by

$$\phi_f(p\|q) = \widetilde{f}\big(\frac{q}{p}\big) - \widetilde{f}\big(\frac{1}{p}\big)\,, \tag{34}$$

$$\psi_f(p\|q) = p\phi_f(p\|q)\,. \tag{35}$$

These definitions are used for $(p, q) \in]0, 1] \times [0, 1]$ (and (35) also for $p = 0$).

We see that complexity ("total" complexity) is the average of local complexity which equals the sum of contributed complexity.

Let us summarize some facts which are relevant for the study of the games γ_f:

Theorem 3. *Let f be a generator. Then f-entropy $\mathrm{S}_f(P)$ is strictly concave in P, f-divergence $\mathrm{D}_f(P\|Q)$ is strictly convex in both variables and f-complexity $\Phi_f(P\|Q)$ is strictly convex in Q. For any preparation $\mathcal{S}_I$, the risk*

$$\mathrm{R}_f(Q) = \sup_{P\in\mathcal{S}_I} \Phi_f(P\|Q)$$

is convex in Q.

In case f-complexity $\Phi_f(P\|Q)$ is concave (not necessarily strictly) in P for each distribution Q, in particular if, for $0 < p \leq 1$ and $0 < q \leq 1$, the inequality

$$\frac{\partial^2}{\partial p^2}\psi_f(p\|q) \leq 0 \tag{36}$$

holds, then the game $\gamma_f(\mathcal{S}_I)$ is in equilibrium and the bi-optimal distribution exists for every convex and compact preparation $\mathcal{S}_I$.

Proof. The defining relation (25) shows that f-entropy is strictly concave.

Consider a mixture $\overline{P} = \sum \alpha_\nu P_\nu$ and any Q. Then, by (26),

$$\begin{aligned}
&\sum_\nu \alpha_\nu \,\mathrm{D}_f(P_\nu\|Q) = \sum_i q_i \sum_\nu \alpha_\nu f\big(\frac{p_{\nu,i}}{q_i}\big)\\
&\geq \sum_i q_i f\Big(\frac{\overline{p}_i}{q_i}\Big) = \mathrm{D}_f(\overline{P}\|Q)\,,
\end{aligned}$$

hence $\mathrm{D}_f(\cdot\|Q)$ is convex. Closer inspection of the argument tells us that the convexity is strict.

The strict convexity for every P of $\mathrm{D}_f(P\|\cdot)$, and then also of $\Phi_f(P\|\cdot)$, follows in a similar way as above but based on (32) and on Lemma 1. Then, the function $Q \curvearrowright \mathrm{R}_f(Q)$ can be viewed as a supremum of convex functions, hence this function is convex, too.

Finally, the statement about equilibrium when $\Phi_f(\cdot\|Q)$ is concave follows from Theorem 2. □

To check (36), we may use the formulas

$$\begin{aligned}
\frac{\partial^2}{\partial p^2}\psi_f(p\|q) &= \frac{1}{q} f''\big(\frac{p}{q}\big) - f''(p)\,, &(37)\\
&= p^{-3}\Big(q^2 \widetilde{f}''\big(\frac{q}{p}\big) - \widetilde{f}''\big(\frac{1}{p}\big)\Big)\,. &(38)
\end{aligned}$$

For the classical generator, (36) holds with equality, reflecting the fact that then $\Phi_f(P\|Q)$ is even affine in P for every distribution Q.

5 Possible interpretations

In this section we consider the measure of complexity Φ_f associated with a generator f. The section is partly speculative. It is intended as a first attempt to reach a better understanding of the significance of the notion of complexity as used in this paper. This is highly desirable as an overall aim of the present paper is to promote complexity as a key concept, even more fundamental than entropy. An interpretation of complexity in practical terms will support this view.

There is a completely satisfactory interpretation of Φ_{clas} via coding as hinted at in Section 2. For more general complexity functions, we do not know of a fully convincing interpretation. However, the formula (30) is a step in the right direction. Firstly, complexity is written as a kind of average with respect to the distribution P. The average is "entangled" as it involves the p_i's as well as the q_i's. Secondly, the quantity being avaraged, i.e. the local complexity, can with some good will be seen as a sensible one for which the $\widetilde{f}$-function represents difficulty or complexity associated to various situations, measured relatively so that only differences occur.

We suggest to interpret $\phi_f(p\|q)$ as the complexity of an event, say B, with probability p in case our reference probability attached to the event is q. Consider first the extremal case when B is deterministic, i.e. $p = 1$. This does not – so we imagine – give rise to any interaction (disturbance) and the complexity is given by $\widetilde{f}(q)$. Then assume that p is (a bit) smaller than 1. This corresponds – again according to our imagination – to a *surprise factor* $\rho = \frac{1}{p} > 1$. The reference value to be used should really be q but the surprise factor results in a disturbance to the level ρq which results in a smaller complexity, $\widetilde{f}(\rho q)$, than that associated with q. However, we have to accept a penalty associated with the level of the surprise. Thus we have to add $-\widetilde{f}(\frac{1}{p})$.

It looks as if interference is due to events that have not occured as there is no disturbance associated with deterministic events.

As a special case, consider the possibility that the local complexity is independent of the surprise factor. Then $\phi_f(p\|q) = \phi_f(1\|q)$ or

$$\widetilde{f}(\frac{q}{p}) - \widetilde{f}(\frac{1}{p}) = \widetilde{f}(q)\,. \tag{39}$$

If this is to hold generally we see, recalling also the normalization requirement (21), that there is only one such possible generator, viz the classical one, $\widetilde{f}(x) = \ln \frac{1}{x}$.

Thus, even though our considerations are preliminary and highly speculative they do indicate the significance of the classical quantities. They also point to what is missing in non-classical (non-extensive) modelling, viz a substitute for the classical coding process. A deeper analysis of the measuring process may be needed. Possibly – another speculation – this may work better if one extend the quantities found to the quantum setting.

6 Non-extensive entropy via complexity

In statistical physics, the search for non-extensive entropy measures which have reasonable structural properties in relation to mixtures (concavity) and in relation to superposition, and otherwise satisfy the needs of statistical thermodynamics led Tsallis to suggest his by now well known *q-entropy functions*, cf. [11] and the many papers which followed. These further papers can be traced from the bibliographical file maintained by Tsallis, cf. [11]. Here, we only point to Tsallis [22], Plastino and Plastino [23], Kaniadakis [12] and Souza and Tsallis [24].

The q-entropy functions of Tsallis were first considered in mathematics as we shall comment on in the final section, but it was not until Tsallis and followers argued for the q-entropies as useful in the study of certain phenomena of statistical physics that these entropy measures gained in popularity, now predominantly within the physical community. More recently, Kaniadakis introduced another family of non-extensive entropies, called κ*-entropies*, cf. [12] and [13]. The motivation was to serve the needs of relativistic non-extensive statistical physics.

We shall introduce a two-parameter family $(\Phi_{\alpha,\beta}, \mathrm{S}_{\alpha,\beta}, \mathrm{D}_{\alpha,\beta})$ of concrete complexity triples which will contain both the entropy functions of Tsallis and those of Kaniadakis. The family is defined via a family $(f_{\alpha,\beta})$ of generators and these, in turn, are defined via *deformed logarithms* For any pair (α, β) of real numbers we define the (α, β)-*deformed logarithm* as the following function on $]0, \infty[$:

$$\ln_{\alpha,\beta} x = \frac{x^{\beta} - x^{\alpha}}{\beta - \alpha} \text{ if } \beta \neq \alpha\,, \tag{40}$$

$$\ln_{\alpha,\beta} x = x^{\alpha} \ln x \text{ if } \beta = \alpha\,. \tag{41}$$

When needed, these definitions are extended in the natural way, allowing also for $x = 0$ or $x = \infty$.

For $\alpha = q - 1\,, \beta = 0$ we obtain the *q-logarithms* which are related to the entropy functions introduced by Tsallis:

$$\ln_q x = \frac{1 - x^{q-1}}{1 - q}\,, \tag{42}$$

and for $\alpha = -\kappa\,, \beta = \kappa$ we recognize the *κ-logarithms* which are related to the entropy functions introduced by Kaniadakis:

$$\ln_{\{\kappa\}} x = \frac{x^{\kappa} - x^{-\kappa}}{2\kappa}\,. \tag{43}$$

Geometrically, $\ln_{\alpha,\beta} x$ is the slope of the chord on the function $y \curvearrowright x^y$ corresponding to the two values α and β for the independent variable y, with the understanding that the chorde is replaced by the tangent in case $\beta = \alpha$. 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*),

$$\ln_{\alpha,\beta} \frac{1}{x} = -\ln_{-\beta,-\alpha} x \tag{44}$$

(*inversion*) and, finally, that the following *functional equation* holds:

$$\ln_{\alpha,\beta}(xy) = y^{\beta} \ln_{\alpha,\beta} x + x^{\alpha} \ln_{\alpha,\beta} y\,. \tag{45}$$

For suitable values of the parameters α and β, the generators we shall consider are given by

$$f_{\alpha,\beta}(x) = x \ln_{\alpha,\beta}(x)\,. \tag{46}$$

According to standard conventions, it is understood that $f_{\alpha,\beta}(0) = 0$. The Csiszár-duals are

$$\widetilde{f}_{\alpha,\beta}(x) = \ln_{\alpha,\beta} \frac{1}{x}\,.$$

Let us clarify when the $f_{\alpha,\beta}$'s are genuine generators:

Lemma 2. *Let Ω denote the set of pairs of real numbers (α, β) for which $f_{\alpha,\beta}$ is a generator according to the definition in Section 4. Then Ω consists of all points in $]-1, 0] \times [0, \infty[$ and in the symmetric set obtained by interchanging α and β.*

Proof. For $\alpha = \beta$, only $f_{0,0}$ is a generator. Assume that $\alpha < \beta$ and that $f_{\alpha,\beta}$ is a generator. As

$$f''_{\alpha,\beta}(x) = \frac{x^{\alpha-1}}{\beta - \alpha}\Big((1+\beta)\beta x^{\beta-\alpha} - (1+\alpha)\alpha\Big) \geq 0\,, \tag{47}$$

we must have $-1 \leq \alpha \leq 0 \leq \beta$. As $f_{\alpha,\beta}$ is strictly convex, we must exclude the case $\alpha = -1\,, \beta = 0$. But also other cases with $\alpha = -1$ are excluded as we require that f is continuous at 0. [6]

[6] The cases excluded in this way could be allowed, though. Then the entropy function will not be continuous, even for a finite alphabet $\mathbb{A}$. For $\alpha = -1\,, \beta = 1$, for example, one finds the entropy function $\mathrm{S}(P) = \frac{1}{2}(n_P - \sum p_i^2)$ where n_P is the number of non-zero p_i's and divergence is closely related to the well known *Pearson χ^2-divergence* (cf. [25]). In fact, for equivalent distributions, $\mathrm{D}(P\|Q) = \frac{1}{2}\sum(p_i - q_i)^2/q_i$, half the χ^2-divergence.

The cases which we are left with, all define legitimate generators. We have, therefore, identified all generators with $\alpha \leq \beta$. The case $\alpha \geq \beta$ is handled by symmetry. $\square$

The set Ω is referred to as *the natural domain*. As pointed out in the footnote, one could add certain parameter values ($(-1, \beta)$ with $\beta > 0$ and $(\alpha, -1)$ with $\alpha > 0$) and obtain a larger domain, the *extended domain*, $\overline{\Omega}$, cf. Figure 2.

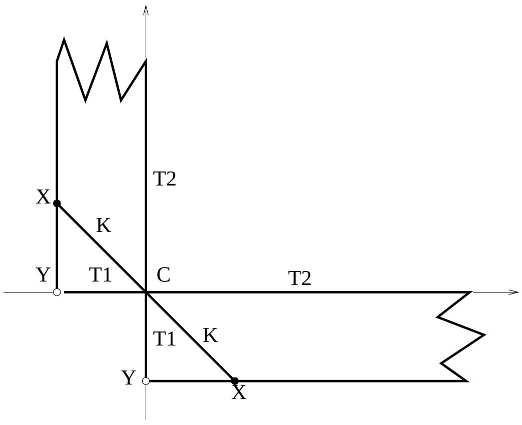

Figure 2. The extended domain $\overline{\Omega}$ in the (α, β)-plane.

We now define (α, β)-*complexity*, (α, β)-*entropy* and (α, β)-*redundancy* (or -*divergence*) for $(\alpha, \beta) \in \Omega$ in the obvious way by substituting $f_{\alpha,\beta}$ for f in the formulas (24)-(26). In terms of deformed logarithms we find the expressions

$$\Phi_{\alpha,\beta}(P\|Q) = \sum_{i \in \mathbb{A}} p_i \Big(\ln_{\alpha,\beta} \frac{p_i}{q_i} - \ln_{\alpha,\beta} p_i \Big) \,, \tag{48}$$

$$\mathrm{S}_{\alpha,\beta}(P) = -\sum_{i \in \mathbb{A}} p_i \ln_{\alpha,\beta} p_i \,, \tag{49}$$

$$\mathrm{D}_{\alpha,\beta}(P\|Q) = \sum_{i \in \mathbb{A}} p_i \ln_{\alpha,\beta} \frac{p_i}{q_i} \,. \tag{50}$$

The local and the contributed complexity are denoted in a similar way ($\phi_{\alpha,\beta}$ and $\psi_{\alpha,\beta}$) and so are the associated games ($\gamma_{\alpha,\beta}$).

Let us discuss two special subfamilies which are known to be of significance for statistical physics.

First consider the cases obtained by choosing $(\alpha, \beta) = (q-1, 0)$. In order to assure that $(\alpha, \beta) \in \Omega$, we assume that q, called the *Tsallis index*, is positive. This choice leads to the *Tsallis family*. For this family, we use the suffix "q" instead of "$q-1,0$" for all associated concepts (Φ, S, D, ϕ, ψ and γ). In particular, for $q \neq 1$, we find the expressions:

$$\Phi_q(P\|Q) = \frac{1}{1-q}\sum_{i\in\mathbb{A}} p_i^q\big(1-q_i^{1-q}\big)\,, \tag{51}$$

$$\mathrm{S}_q(P) = \frac{1}{1-q}\sum_{i\in\mathbb{A}} p_i\big(p_i^{q-1}-1\big) = \frac{1}{1-q}\Big(\sum_{i\in\mathbb{A}} p_i^q - 1\Big)\,, \tag{52}$$

$$\mathrm{D}_q(P\|Q) = \frac{1}{1-q}\sum_{i\in\mathbb{A}} p_i\Big(1-\big(\frac{p_i}{q_i}\big)^{q-1}\Big) = \frac{1}{1-q}\Big(1-\sum_{i\in\mathbb{A}} p_i\big(\frac{p_i}{q_i}\big)^{q-1}\Big)\,. \tag{53}$$

For $q=1$ we are back to the classical functions Φ_{clas}, $\mathrm{S}_{\mathrm{clas}}$ and $\mathrm{D}_{\mathrm{clas}}$. In (52) we recognize the celebrated Tsallis entropy functions. For $q=\frac{1}{2}$, (53) gives us the popular *Hellinger divergence* which is the only symmetric divergence in the Tsallis family (it is even a squared metric, cf. LeCam [25]).

Another important family arises by choosing $(\alpha,\beta)=(-\kappa,\kappa)$ with $0 \leq \kappa < 1$ (the value $\kappa=1$ could also be allowed, though, see the previous footnote). This gives the *Kaniadakis family*. Here, we use the suffix "$\{\kappa\}$" instead of "$-\kappa,\kappa$" for all associated concepts and quantities. In particular, for $\kappa \neq 0$, we find the expressions:

$$\Phi_{\{\kappa\}}(P\|Q) = \frac{1}{2\kappa}\sum_{i\in\mathbb{A}} p_i(1-q_i^\kappa)\Big(p_i^{-\kappa}+\big(\frac{p_i}{q_i}\big)^\kappa\Big)\,, \tag{54}$$

$$\mathrm{S}_{\{\kappa\}}(P) = \frac{1}{2\kappa}\sum_{i\in\mathbb{A}} p_i\big(p_i^{-\kappa}-p_i^\kappa\big)\,, \tag{55}$$

$$\mathrm{D}_{\{\kappa\}}(P\|Q) = \frac{1}{2\kappa}\sum_{i\in\mathbb{A}} p_i\big((\frac{p_i}{q_i})^\kappa-(\frac{q_i}{p_i})^\kappa\big)\,. \tag{56}$$

For $\kappa=0$ one obtains the classical quantities Φ_{clas}, $\mathrm{S}_{\mathrm{clas}}$ and $\mathrm{D}_{\mathrm{clas}}$. In (55) we recognize the family of entropy functions introduced by Kaniadakis.

In Figure 2 we have indicated by "T" respectively "K", those parts of the extended domain which define the Tsallis, respectively the Kaniadakis family ("T1" corresponds to $q \leq 1$, "T2" to $q \geq 1$). The classical case is obtained when we consider the point "C". The points "X" give divergence measures related to χ^2-divergence, cf. footnote 7. These points were added to the natural domain when we formed the extended domain. Note that the points "Y" on Figure 2 are truly forbidden[7].

We shall now investigate which of the games $\gamma_{\alpha,\beta}$ with $(\alpha,\beta)\in\Omega$ give rise to equilibrium for convex preparations. It lies nearby to apply the criterion of Theorem 3. For that purpose we note that

[7]in physical terms, they correspond to situations of chaos and high temperatures.

$$\frac{\partial^2}{\partial p^2}\psi_{\alpha,\beta}(p\|q) = \frac{p^{\alpha-1}}{\beta-\alpha}\Big((1+\beta)\beta p^{\beta-\alpha}\big(q^{-\beta}-1\big)-(1+\alpha)\alpha\big(q^{-\alpha}-1\big)\Big)\,. \quad (57)$$

The "concave/convex" behaviour (first concave, then convex) of the functions $\psi_{\alpha,\beta}(\cdot\|q)$, considered on all of $[0,\infty[$, is of significance (cf. indications in the final section). Here we focus on the special cases when the functions (keeping α and β fixed, but varying q) are either concave or convex. These cases are closely related to the Tsallis family. Indeed, the functions in question are concave on all of $[0,\infty[$ if and only if we are in the "T1"-case (see Figure 2) and they are convex in $[0,\infty]$ if and only if we are in the "T2"-case[8]. We can now prove the following main result about the Tsallis family:

Theorem 4. (Equilibrium under Tsallis entropy). *If $0 < q \le 1$, then $\gamma_q = \gamma_q(\mathcal{S}_I)$ is in equilibrium and the bi-optimal distribution exists for any compact and convex preparation. This does not hold for any other value of q.*

Proof. The first part follows by the discussion above.

To prove the second half, assume that $q > 1$. Consider the case of a finite alphabet $\mathbb{A}$ with n elements, say, and the preparation $\mathcal{S}_I$ of *all* distributions on $\mathbb{A}$. As S_q is strictly concave (Theorem 3) a symmetry argument shows that the uniform distribution, call it U, is the unique MaxEnt-distribution, hence

$$\mathrm{S}_q^{\max} = \mathrm{S}_q(U) = \frac{1}{q-1}(1-n^{-q+1})\,.$$

In order to calculate $\mathrm{R}_q^{\min}$, first note that by symmetry and convexity of the risk function (cf. Theorem 3), it follows that $\mathrm{R}_q^{\min} = R_q(U)$. The analysis preceeding the proof shows that Φ_q is convex in the first variable and it follows that $\Phi_q(P\|U)$ achieves its maximal value for an extreme distribution P, i.e. for a deterministic distribution. This shows that

$$\mathrm{R}_q(U) = \frac{1}{1-q}(1-q_{min}^{1-q})$$

with $q_{\min}$ the smallest point probability in Q. It follows that

$$\mathrm{R}_q^{\min} = \frac{1}{q-1}(n^{q-1}-1)\,.$$

Thus

$$\mathrm{R}_q^{\min} - \mathrm{S}_q^{\max} = \frac{1}{q-1}(n^{q-1}+n^{-q+1}-2)$$

[8]if we consider the extended domain, convexity also holds for the added parameter values.

which is positive (unless $n = 1$), thus violating the condition of equilibrium. $\square$

Without going into details we remark that arguing a bit like in the above proof, one finds that for $(\alpha, \beta) \in \Omega$ with $\alpha + \beta > 0$, the game $\gamma_{\alpha,\beta}$ need not be in equilibrium for simple convex preparations.

7 Discussion, conclusions

The paper is conceptual in nature. Complexity is promoted as the key concept. Based on this notion, one can define entropy as minimal complexity and redundancy (divergence) as actual minus minimal complexity. Games based on complexity allow a discussion of equilibrium in models used in statistical thermodynamics. It is shown that the entropy measures of Tsallis and Kaniadakis can be discussed from this point of view by embedding them in one and the same two-parameter family of entropy functions (and associated complexity measures).

A preliminary interpretation of complexity functions via the introduced local complexity is provided.

Apparently, Tsallis entropy (and associated objects) with parameter $q \leq 1$ is better behaved from the game theoretical point of view than Tsallis entropy with parameter > 1 or than Kaniadakis entropy. However, this conclusion is not all that obvious. Fact is that the game theoretical notion of equilibrium which we have applied works with "worst possible scenarios". This points to the conclusion that *when* the principle of game theoretical equilibrium (GTE) applies, it is appropriate to infer that also equilibrium in a physical sense is ensured. However, if GTE *does not* apply in the form used here, this may be because weaker concepts of game theoretical equilibrium are more appropriate. As one such form we propose to look closer into the condition that the bi-optimal strategy exists. Appropriate interpretations of the possible discrepancy between $\mathrm{R}^{\min}$ and $\mathrm{S}^{\max}$, may then point to situations where physical equilibrium is not excluded. Let us illustrate this point of view without going into too many details. As example we take the Kaniadakis family.

EXAMPLE. Let $\mathbb{A}$ be finite with $n \geq 2$ elements, let $0 < \kappa < 1$ and consider the preparation of all distributions over $\mathbb{A}$. Denote by U_ν a typical uniform distribution over ν elements in $\mathbb{A}$. By previous results, U_n is the bi-optimal strategy and $\mathrm{S}^{\max}_{\{\kappa\}} = \ln_{\{\kappa\}} n$. Further,

$$\mathrm{R}^{\min}_{\{\kappa\}} = \mathrm{R}_{\{\kappa\}}(U_n) = \frac{1 - n^{-\kappa}}{2\kappa} \sup_P \sum_{i \in \mathbb{A}} \left(p_i^{1-\kappa} + p_i^{1+\kappa} n^\kappa \right).$$

The function appearing in the summation is of the concave/convex type with inflection point ξ given by

$$\xi = \frac{1}{\sqrt{n}}\Big(\frac{1-\kappa}{1+\kappa}\Big)^{\frac{1}{2\kappa}}\,.$$

We can then apply a technical lemma, the “lemma of replacement” , cf. Theorem III.1 of [26], and conclude that the suppremum above is attained for a mixture $(1-x)U_1 + xU_n$ of a deterministic distribution and U_n. It is then easy to see that

$$\mathrm{R}^{\min}_{\{\kappa\}} > R_{\{\kappa\}}(U_1) = R_{\{\kappa\}}(U_n) = \mathrm{S}^{\max}_{\{\kappa\}}\ .$$

The system is not in equilibrium in our strict sense. However, the players can agree on one and the same choice of distribution, the uniform distribution. And even if Player II knew beforehand that Player I had chosen this distribution, he would not feel inclined to change his strategy. Not so for Player I. Had Player I known that Player II had chosen U_n as his strategy, Player I would have done well in changing the strategy by considering an appropriate mixture of U_n and a deterministic distribution.

The above points to an essential asymmetry among the players. Really, this is only natural. We expect the laws of nature to be fixed once and for all, hence Player I really has no intention or possibility to change his strategy, whereas Player II, the physicist, should use all means available to improve the performance, i.e. to decrease the complexity.

The example also points to other sensible variations of the theme of equilibrium. Fact is that a mixture of a deterministic distribution and U_n turned up. But *which* deterministic distribution? As symmetry appears to place all n deterministic distributions on an equal footing, a sensible principle of symmetry would lead to U_n as the natural choice of Player I after all, in spite of a potential for Player I to improve the performance.

Speculative comments as in the example above or as in Section 5 – a mathematicians attempt to think in physical terms – may guide further research. But also, several purely mathematical questions are open to further investigation. This concerns questions about strict convexity of the risk function, about extensions of the Pythagorean inequalities (16) and (17) when only weaker forms of equilibrium hold, and so on. It would also be appropriate to illuminate the choice of model in Section 4. It looks so natural, but with a given entropy function, is it really the best way to associate a measure of complexity as in (24)? A number of topological questions could also be looked into, e.g. variation of the method of observation, decribed by a point $(\alpha, \beta) \in \overline{\Omega}$, e.g. a move towards "chaos" , the forbiddeen point "Y" on Figure 2.

Lastly, some comments on the complexity-, entropy-, and divergence-measures here considered. As will be seen, the notion of Csiszár f-divergences plays an important role. These divergence measures were introduced, independently, by Csiszár, [27], and by Ali and Silvey, [28]. The literature on specific divergence measures is rich and we do not claim originality regarding these as such. The novelty here lies in the focus on complexity and the related games and in the possibility to choose suitable families of relevance to statistical physics.

The more concrete two-parameter family (49) has been considered in the mathematical literature by Mittal [29] and by Sharma and Taneja [30] from the point of view of functional equations of relevance to information theory. More relevant here is the work by Borges and Roditi [31] which looks at these entropies from a physical point of view and also notes the key problem of proper interpretations – but with only very sporadic comments on this important issue. These authors, in contrast to the previous ones, also make use of the deformed logarithms[9].

Regarding Tsallis entropy, they were introduced to the mathematical readership by Havrda and Charvát, [32], twenty years before Tsallis, but were not met with great enthusiasm[10]. However, a number of papers did appear, in particular dealing with axiomatics, cf. Forte and Ng [33] (other references can be traced via the rewiew of [32] on MathSciNet). Though the "pre-Tsallis" literature on these issues did not really catch on, one should also note the paper by van der Lubbe [34] as this author addressed the problem of coding, though in a rather artificial way. Anyhow, it is an important open problem as we have put much emphasis on.

Acknowledgements

The author thanks Giorgio Kaniadakis for useful comments, in particular for providing important input which helped in choosing what appears as the most sensible choice of the families based on the deformed logarithms. Further, the author acknowledges useful discussions with Peter Harremoës of all aspects of the paper.

[9]apparently, the device to base definitions of non-extensive entropy measures on such functions was first suggested by Tsallis, [22].

[10]In the review, cf. [32], Csiszár writes (with reference to Rényi entropies) "*In the opinion of the reviewer, however, the difference is not essential since "a-entropy" and "entropy of order α" are in an obvious one-to-one functional relationship (with $\alpha = a$).*"

References

[1] E. T. Jaynes. Information theory and statistical mechanics, I and II. *Physical Reviews*, 106 and 108:620–630 and 171–190, 1957.

[2] E. T. Jaynes. Clearing up mysteries – the original goal. In J. Skilling, editor, *Maximum Entropy and Bayesian Methods.* Kluwer, Dordrecht, 1989.

[3] E. T. Jaynes. Webpage maintained by L. Bretthorst, dedicated to Jaynes work, available ONLINE from http://bayes.wustl.edu.

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

[5] F. Topsøe. Information Theoretical Optimization Techniques. *Kybernetika*, 15:8–27, 1979.

[6] F. Topsøe. Game theoretical equilibrium, maximum entropy and minimum information discrimination. In A. Mohammad-Djafari and G. Demoments, editors, *Maximum Entropy and Bayesian Methods*, pages 15–23. Kluwer Academic Publishers, Dordrecht, Boston, London, 1993.

[7] P. Harremoës. Binomial and Poisson distributions as maximum entropy distributions. *IEEE Trans. Inform. Theory*, 47(5):2039–2041, July 2001.

[8] P. Harremoës and F. Topsøe. Maximum entropy fundamentals. *Entropy*, 3:191–226, Sept. 2001. http://www.unibas.ch/mdpi/entropy/ [ONLINE].

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

[10] F. Topsøe. Information Theory at the Service of Science. 2004. To appear in a special volume of the János Bolyai Mathematical Society.

[11] 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.

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

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

[14] S. Kullback and R. Leibler. On information and sufficiency. *Ann. Math. Statist.*, 22:79–86, 1951.

[15] S. Kullback. *Informaton Theory and Statistics*. Wiley, New York, 1959.

[16] J. Rissanen A. Barron and B. Yu. The minimum description length principle in coding and modeling. *IEEE Trans. Inform. Theory*, 44:2743–2760, 1998.

[17] R. S. Ingarden and K. Urbanik. Quantum informational thermodynamics. *Acta Physica Polonica*, 21:281–304, 1962.

[18] N. N. Čencov. A nonsymmetric distance between probability distributions, entropy and the Pythagorean theorem. *Math. Zametki*, 4:323–332, 1968. (in Russian).

[19] I. Csiszár. I-divergence geometry of probability distributions and minimization problems. *Ann. Probab.*, 3:146–158, 1975.

[20] F. Liese and I. Vajda. *Convex Statistical Distances*. Teubner, Leipzig, 1987.

[21] F. Österreicher and I. Vajda. A new class of metric divergences on probability spaces and and its statistical applications. *Ann. Inst. Statist. Math.*, 55:639–653, 2003.

[22] C. Tsallis. What are the numbers that experiments provide? *Quimica Nova*, 17:468, 1994.

[23] A. Plastino and A. R. Plastino. Tsallis Entropy and Jaynes' Information Theory Formalism. *Brazilian Journal of Physics*, 29:50–60, 1999.

[24] A. M. C Souza and C. Tsallis. Stability of the entropy for superstatistics. *Physica A*, 2004. under publication.

[25] L. Le Cam. *Asymptotic Methods in Statistical Theory*. Springer-Verlag, New York, 1986.

[26] P. Harremoës and F. Topsøe. Inequalities between entropy and index of coincidence derived from information diagrams. *IEEE Trans. Inform. Theory*, 47:2944–2960, 2001.

[27] I. Csiszár. Information-type measures of difference of probability distributions and indirect observations. *Studia Sci. Math. Hungar.*, 2:299–318, 1967.

[28] S. M. Ali and S. D. Silvey. A general class of coefficients of divergence of one distribution from another. *J. Roy. Statist. Soc. Ser B*, 28:131–142, 1966.

[29] D. P. Mittal. On some functional equations concerning entropy, directed divergence and inaccuracy. *Metrika*, 22:35–45, 1975.

[30] B. D. Sharma and I. J. Taneja. Entropy of type (α, β) and other generalized measures in information theory. *Metrika*, 22:205–215, 1975.

[31] E. P. Borges and I. Roditi. A family of nonextensive entropies. *Physics Letters A*, 246:399–402, 1998.

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

[33] B. Forte and C. T. Ng. Derivation of a class of entropies including those of degree β. *Information and Control*, 28:335–351, 1975.

[34] J. C. A. van der Lubbe. On certain coding theorems for the information of order α and of type β. In *Trans. Eighth Prague Conf. on Inform. Theory, Statist. Decision Functions, Random Processes*, Prague, 1978. Czech. Acad. Science, Academia Publ. Prague, 1979.