Contemporary Physics
Vol. 50, No. 4, July–August 2009, 495–510

Generalised information and entropy measures in physics

Christian Beck*

School of Mathematical Sciences, Queen Mary, University of London, Mile End Road, London E1 4NS, UK

(*Received 21 December 2008; final version received 13 February 2009*)

The formalism of statistical mechanics can be generalised by starting from more general measures of information than the Shannon entropy and maximising those subject to suitable constraints. We discuss some of the most important examples of information measures that are useful for the description of complex systems. Examples treated are the Rényi entropy, Tsallis entropy, Abe entropy, Kaniadakis entropy, Sharma–Mittal entropies, and a few more. Important concepts such as the axiomatic foundations, composability and Lesche stability of information measures are briefly discussed. Potential applications in physics include complex systems with long-range interactions and metastable states, scattering processes in particle physics, hydrodynamic turbulence, defect turbulence, optical lattices, and quite generally driven nonequilibrium systems with fluctuations of temperature.

Keywords: measures of information; entropy; generalised statistical mechanics; complex systems

1. How to measure information

1.1. *Prologue*

How should one measure information? There is no unique answer to this. There are many different information measures, and what measure of information is the most suitable one will in general depend on the problem under consideration. Also, there are different types of information. For example, the information a reader gets from reading a book on quantum field theory is different from the one he gets from reading Shakespeare's *Romeo and Juliet*. In general one has to distinguish between elementary and advanced information concepts. The elementary information is just related to technical details such as, for example, the probability to observe certain letters in a long sequence of words (see Figure 1). The advanced information is related to the information the reader really gets out of reading and understanding a given text, i.e. this concept requires coupling to a very complex system such as the brain of a human being.

In physics, the missing information on the concrete state of a system is related to the *entropy* of the system. Entropy is an elementary information concept. Many different physical definitions of entropy can be given, and what makes up a 'physically relevant entropy' is often subject to 'heated' discussions. Misunderstandings with respect to the name 'entropy' seem to be the rule rather than the exception within the past 130 years. Generally one may use the name 'entropy' as a synonym for a possible quantity to measure missing information, keeping in mind that large classes of possible functions will potentially do the job, depending on application.

The entire formalism of statistical mechanics can be regarded as being based on maximising the entropy (= missing information) of the system under consideration subject to suitable constraints, and hence naturally the question arises how to measure this missing information in the first place [1]. While normally one chooses the Shannon information measure, in principle more general information measures (that contain the Shannon information as a special case) can be chosen as well. These then formally lead to generalised versions of statistical mechanics when they are maximised [2–7].

In this paper we describe some generalised information and entropy measures that are useful in this context. We discuss their most important properties, and point out potential physical applications. The physical examples we choose are the statistics of cosmic rays [8], defect turbulence [9], and optical lattices [10,11], but the general techniques developed have applications for a variety of other complex systems as well, such as driven nonequilibrium systems with large-scale fluctuations of temperature (so-called superstatistical systems [12,13]), hydrodynamic turbulence [14,15] scattering processes in particle physics [16,17], gravitationally interacting systems [18,19] and Hamiltonian systems with long-range interactions and metastable states [19,20]. There are applications outside physics as well, for example in mathematical finance [22], biology [23] and medicine [24].

*Email: c.beck@qmul.ac.uk

ISSN 0010-7514 print/ISSN 1366-5812 online

DOI: 10.1080/00107510902823517
http://www.informaworld.com

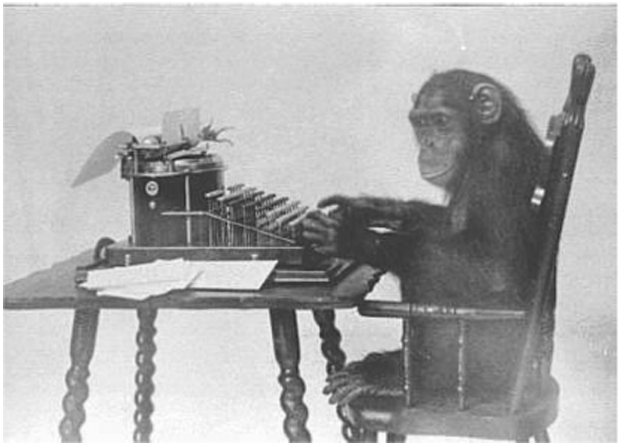

Figure 1. There is no obvious way to measure the information contents of given symbol sequences. While it is relatively easy to distinguish between a random sequence of symbols and Shakespeare's *Romeo and Juliet* in terms of suitable elementary information measures, it is less obvious how to distinguish the fact that the advanced information contents given by Shakespeare's *Romeo and Juliet* is different from the one given by a book on quantum field theory.

1.2. Basic concepts

One usually restricts the concept of an information measure to an information that is a function of a given probability distribution of events (and nothing else).[1] The basic idea is as follows. Consider a sample set of W possible events. In physics events are often identified as possible microstates of the system. Let the probability that event i occurs be denoted as p_i. One has from normalisation

$$\sum_{i=1}^{W} p_i = 1. \tag{1}$$

We do not know which event will occur. But suppose that one of these events, say j, finally takes place. Then we have clearly gained some information, because before the event occurred we did not know which event would occur.

Suppose that the probability p_j of that observed event j is close to 1. This means we gain very little information by the observed occurrence of event j, because this event was very likely anyway. On the other hand, if p_j is close to zero, then we gain a lot of information by the actual occurrence of event j, because we did not really expect this event to happen. The information gain due to the occurrence of a single event j can be measured by a function $h(p_j)$, which should be close to zero for p_j close to 1. For example, we could choose $h(p_j) = \log p_j$, the logarithm to some suitable basis a. If this choice of a is $a = 2$ then h is sometimes called a 'bit-number' [1]. But various other functions $h(p_j)$ are possible as well, depending on the application one has in mind. In other words, an information measure should better be regarded as a man-made construction useful for physicists who don't fully understand a complex system but try to do so with their limited tools and ability. Once again we emphasise that an information measure is not a universally fixed quantity. This fact has led to many misunderstandings in the community.

In a long sequence of independent trials, in order to determine an average information gain by the sequence of observed events i we have to weight the information gain associated with a single event with the probability p_i that event i actually occurs. That is to say, for a given function h the average information gained during a long sequence of trials is

$$I(\{p_i\}) = \sum_{i=1}^{W} p_i h(p_i). \tag{2}$$

Many information measures studied in the literature are indeed of this simple trace form. But other forms are possible as well. One then defines the entropy S as 'missing information', i.e.

$$S = -I. \tag{3}$$

This means the entropy is defined as our missing information on the actual occurrence of events, given that we only know the probability distribution of the events.

If the probability distribution is sharply peaked around one almost certain event j, we gain very little information from our long-term experiment of independent trials: the event j will occur almost all of the time, which we already knew before. However, if all events are uniformly distributed, i.e. $p_i = 1/W$ for all i, we get a large amount of information by doing this experiment, because before we did the experiment we had no idea which events would actually occur, since they were all equally likely. In this sense, it is reasonable to assume that an (elementary) information measure should take on an extremum (maximum or minimum, depending on sign) for the uniform distribution. Moreover, events i that cannot occur ($p_i = 0$) do not influence our gain of information in the experiment at all. In this way we arrive at the most basic principles an information measure should satisfy.

1.3. The Khinchin axioms

There is a more formal way to select suitable (elementary) information measures, by formulating a set of axioms and then searching for information measures that satisfy these axioms. *A priori*, there is an infinite set of possible information measures, not only of the simple form (2) but of more general forms as well, based on arbitrary functions of the entire set of p_i. How can we select the most suitable ones, given certain requirements we have in mind? Of course, what is 'most suitable' in this context will in general depend on the application we have in mind. The most appropriate way of dealing with this problem is to postulate some basic and essential properties the information measure one is interested in should have, and then to derive the functional form(s) that follows from these postulates.

Khinchin [25] has formulated four axioms that describe the properties a 'classical' information measure I should have (by 'classical' we mean an information measure yielding ordinary Boltzmann–Gibbs type of statistical mechanics):

Axiom 1

$$I = I(p_1, \ldots, p_W). \tag{4}$$

That is to say, the information measure I only depends on the probabilities p_i of the events and nothing else.

Axiom 2

$$I(W^{-1}, \ldots, W^{-1}) \leq I(p_1, \ldots, p_W). \tag{5}$$

This means the information measure I takes on an absolute minimum for the uniform distribution $(W^{-1}, \ldots, W^{-1})$, any other probability distribution has an information content that is larger or equal to that of the uniform distribution.

Axiom 3

$$I(p_1, \ldots, p_W) = I(p_1, \ldots, p_W, 0). \tag{6}$$

This means the information measure I should not change if the sample set of events is enlarged by another event that has probability zero.

Axiom 4

$$I\left(\left\{p_{ij}^{\mathrm{I,II}}\right\}\right) = I(\{p_i^{\mathrm{I}}\}) + \sum_i p_i^{\mathrm{I}} I(\{p^{\mathrm{II}}(j|i)\}). \tag{7}$$

This axiom is slightly more complicated and requires a longer explanation. The axiom deals with the composition of two systems I and II (not necessarily independent). The probabilities of the first system are p_i^{I} , those of the second system are p_j^{II}. The joint system I,II is described by the joint probabilities $p_{ij}^{\mathrm{I,II}} = p_i^{\mathrm{I}} p^{\mathrm{II}}(j|i)$, where $p^{\mathrm{II}}(j|i)$ is the conditional probability of event j in system II under the condition that event i has already occurred in system I. $I(\{p^{\mathrm{II}}(j|i)\})$ is the conditional information of system II formed with the conditional probabilities $p^{\mathrm{II}}(j|i)$, i.e. under the condition that system I is in state i.

The meaning of the above axiom is that it postulates that the information measure should be independent of the way the information is collected. We can first collect the information in the subsystem II, assuming a given event i in system I, and then sum the result over all possible events i in system I, weighting with the probabilities p_i^{I}.

For the special case that system I and II are inde pendent the probability of the joint system factorises as

$$p_{ij}^{\mathrm{I,II}} = p_i^{\mathrm{I}} p_j^{\mathrm{II}}, \tag{8}$$

and only in this case, Axiom 4 reduces to the rule of additivity of information for independent subsystems:

$$I(\{p_{ij}^{\mathrm{I,II}}\}) = I(\{p_i^{\mathrm{I}}\}) + I(\{p_j^{\mathrm{II}}\}). \tag{9}$$

Whereas there is no doubt about Axioms 1–3, the reader immediately notices that Axiom 4 requires a much longer explanation. From a physical point of view, Axiom 4 is a much less obvious property. Why should information be independent from the way we collect it?

To illustrate this point, we may consider a simple example of an information-collecting system, a first-year undergraduate student trying to understand physics. This student will learn much more if he first attends a course on classical mechanics, collecting all available information there, and then attends a course on quantum mechanics. If he does it the other way round, he will probably hardly understand anything in the course on quantum mechanics, since he does not have the necessary prerequisites. So attending the quantum mechanics course first leads to zero information gain. Apparently, the order in which the information of the two courses (the two subsystems) is collected is very important and leads to different results in the achieved knowledge.

In general complex systems, the order in which information is collected can be very relevant. This is a kind of information hysteresis phenomenon. In these cases we have situations where the replacement of Axiom 4 by something more general makes physical sense. We will come back to this in Section 3.

1.4. The Shannon entropy

It is easy to verify that the celebrated Shannon entropy, defined by

$$S = -k \sum_{i=1}^{W} p_i \ln p_i, \tag{10}$$

satisfies all four of the Khinchin axioms. Indeed, up to an arbitrary multiplicative constant, one can easily show (see, e.g. [1]) that this is the only entropic form that satisfies all four Khinchin axions, and that it follows uniquely (up to a multiplicative constant) from these postulates. k denotes the Boltzmann constant, which in the remaining sections will be set equal to 1. For the uniform distribution, $p_i = 1/W$, the Shannon entropy takes on its maximum value

$$S = k \ln W, \tag{11}$$

which is Boltzmann's famous formula, carved on his grave in Vienna (Figure 2). Maximising the Shannon entropy subject to suitable constraints leads to ordinary statistical mechanics (see Section 4.3). In thermodynamic equilibrium, the Shannon entropy can be identified as the 'physical' entropy of the system, with the usual thermodynamic relations. Generally, the Shannon entropy has an enormous range of applications not only in equilibrium statistical mechanics but also in coding theory, computer science, etc.

It is easy to verify that S is a concave function of the probabilities p_i, which is an important property to formulate statistical mechanics. Remember that concavity of a differentiable function $f(x)$ means $f''(x) \leq 0$ for all x. For the Shannon entropy one has

$$\frac{\partial}{\partial p_i} S = -\ln p_i - 1, \tag{12}$$

$$\frac{\partial^2}{\partial p_i \partial p_j} S = -\frac{1}{p_i}\delta_{ij} \leq 0, \tag{13}$$

and hence, as a sum of concave functions of the p_i, it is concave.

In classical mechanics, one often has a continuous variable u with some probability density $p(u)$, rather than discrete microstates i with probabilities p_i. In this case the normalisation condition reads $\int_{-\infty}^{\infty} p(u)\,\mathrm{d}u = 1$, and the Shannon entropy associated with this probability density is defined as

$$S = -\int_{-\infty}^{\infty} \mathrm{d}u\, p(u) \ln(\sigma p(u)), \tag{14}$$

where σ is a scale parameter that has the same dimension as the variable u. For example, if u is a velocity (measured in units of m s^{-1}), then $p(u)$, as a probability density of velocities, has the dimension s m^{-1}, since $p(u)\mathrm{d}u$ is a dimensionless quantity. As a consequence, one needs to introduce the scale parameter σ in Equation (14) to make the argument of the logarithm dimensionless.

Figure 2. The grave of Boltzmann in Vienna. On top of the gravestone the formula $S = k \log W$ is engraved. Boltzmann laid the foundations for statistical mechanics, but his ideas were not widely accepted during his time. He committed suicide in 1906.

Besides the Shannon information, there are lots of other information measures. We will discuss some of the most important examples in the next section. Some information measures are more suitable than others for the description of various types of complex systems. We will discuss the axiomatic foundations that lead to certain classes of information measures. Important properties to check for a given information measure are convexity, additivity, composability, and stability. These properties can help to select the most suitable generalised information measure to describe a given class of complex systems.

2. More general information measures

2.1. *The Rényi entropies*

We may replace Axiom 4 by the less stringent condition (9), which just states that the entropy of independent systems should be additive. In this case one ends up with other information measures which are called the *Rényi entropies* [26]. These are defined for an arbitrary real parameter q as

$$S_q^{(\mathrm{R})} = \frac{1}{q-1} \ln \sum_i p_i^q. \tag{15}$$

The summation is over all events i with $p_i \neq 0$. The Rényi entropies satisfy the Khinchin Axioms 1–3 and the additivity condition (9). Indeed, they follow uniquely from these conditions, up to a multiplicative constant. For $q \to 1$ they reduce to the Shannon entropy:

$$\lim_{q\to 1} S_q^{(\mathrm{R})} = S, \tag{16}$$

as can be easily derived by setting $q = 1 + \varepsilon$ and doing a perturbative expansion in the small parameter ε in Equation (15).

The Rényi information measures are important for the characterisation of multifractal sets (i.e. fractals with a probability measure on their support [1]), as well as for certain types of applications in computer science. But do they provide a good information measure to develop a generalised statistical mechanics for complex systems?

At first sight it looks nice that the Rényi entropies are additive for independent subsystems for general q, just as the Shannon entropy is for $q = 1$. But for non-independent subsystems I and II this simplicity vanishes immediately: there is no simple formula of expressing the total Rényi entropy of a joint system as a simple function of the Rényi entropies of the interacting subsystems.

Does it still make sense to generalise statistical mechanics using the Rényi entropies? Another problem arises if one checks whether the Rényi entropies are a concave function of the probabilities. The Rényi entropies do not possess a definite concavity – the second derivative with respect to the p_i can be positive or negative. For formulating a generalised statistical mechanics, this poses a serious problem. Other generalised information measures are better candidates – we will describe some of those in the following.

2.2. *The Tsallis entropies*

The *Tsallis entropies* (also called q-entropies) are given by the following expression [2]:

$$S_q^{(\mathrm{T})} = \frac{1}{q-1}\left(1 - \sum_{i=1}^{W} p_i^q\right). \tag{17}$$

One finds definitions similar to Equation (17) already in earlier papers such as e.g. [27], but it was Tsallis in his seminal paper [2] who for the first time suggested to generalise statistical mechanics using these entropic forms. Again $q \in \mathcal{R}$ is a real parameter, the entropic index. As the reader immediately sees, the Tsallis entropies are different from the Rényi entropies: there is no logarithm anymore. A relation between Rényi and Tsallis entropies is easily derived by writing

$$\sum_i p_i^q = 1 - (q-1) S_q^{(\mathrm{T})} = \exp\left[(q-1) S_q^{(\mathrm{R})}\right], \tag{18}$$

which implies

$$S_q^{(\mathrm{T})} = \frac{1}{q-1}\left(1 - \exp\left[(q-1) S_q^{(\mathrm{R})}\right]\right). \tag{19}$$

Apparently the Tsallis entropy is a monotonous function of the Rényi entropy, so any maximum of the Tsallis entropy will also be a maximum of the Rényi entropy and vice versa. But still, Tsallis entropies have many distinguished properties that make them a better candidate for generalising statistical mechanics than, say, the Rényi entropies.

One such property is concavity. One easily verifies that

$$\frac{\partial}{\partial p_i} S_q^{(\mathrm{T})} = -\frac{q}{q-1} p_i^{q-1}, \tag{20}$$

$$\frac{\partial^2}{\partial p_i \partial p_j} S_q^{(\mathrm{T})} = -q p_i^{q-2} \delta_{ij}. \tag{21}$$

This means $S_q^{(\mathrm{T})}$ is concave for all $q > 0$ (convex for all $q < 0$). This property is missing for the Rényi entropies. Another such property is the so-called Lesche-stability, which is satisfied for the Tsallis entropies but not satisfied by the Rényi entropies (see Section 3.3 for more details).

The Tsallis entropies also contain the Shannon entropy

$$S = -\sum_{i=1}^{W} p_i \ln p_i \tag{22}$$

as a special case. Letting $q \to 1$ we have

$$S_1^{(\mathrm{T})} = \lim_{q\to 1} S_q^{(\mathrm{T})} = S. \tag{23}$$

As expected from a good information measure, the Tsallis entropies take on their extremum for the uniform distribution $p_i = 1/W \, \forall i$. This extremum is given by

$$S_q^{(\mathrm{T})} = \frac{W^{1-q} - 1}{1-q}, \tag{24}$$

which, in the limit $q \to 1$, reproduces Boltzmann's celebrated formula $S = \ln W$.

It is also useful to write down the definition of Tsallis entropies for a continuous probability density

$p(u)$ with $\int_{-\infty}^{\infty} p(u)\,\mathrm{d}u = 1$, rather than a discrete set of probabilities p_i with $\sum_i p_i = 1$. In this case one defines

$$S_q^{(\mathrm{T})} = \frac{1}{q-1}\left(1 - \int_{-\infty}^{+\infty} \frac{\mathrm{d}u}{\sigma}(\sigma p(u))^q\right), \tag{25}$$

where again σ is a scale parameter that has the same dimension as the variable u. It is introduced for a similar reason as before, namely to make the integral in Equation (25) dimensionless so that it can be subtracted from 1. For $q \to 1$ Equation (25) reduces to the Shannon entropy

$$S_1^{(\mathrm{T})} = S = -\int_{-\infty}^{\infty} \mathrm{d}u p(u) \ln(\sigma p(u)). \tag{26}$$

A fundamental property of the Tsallis entropies is the fact that they are not additive for independent subsystems. In fact, they have no chance to do so, since they are different from the Rényi entropies, the only solution to Equation (9).

To investigate this in more detail, let us consider two independent subsystems I and II with probabilities p_i^{I} and p_j^{II}, respectively. The probabilities of joint events i, j for the combined system I,II are $p_{ij} = p_i^{\mathrm{I}} p_j^{\mathrm{II}}$. We may then consider the Tsallis entropy for the first system, denoted as S_q^{I}, that of the second system, denoted as S_q^{II}, and that of the joint system, denoted as $S_q^{\mathrm{I,II}}$. One has

$$S_q^{\mathrm{I,II}} = S_q^{\mathrm{I}} + S_q^{\mathrm{II}} - (q-1) S_q^{I} S_q^{\mathrm{II}}. \tag{27}$$

Proof of Equation (27): We may write

$$\sum_i (p_i^{\mathrm{I}})^q = 1 - (q-1) S_q^{\mathrm{I}}, \tag{28}$$

$$\sum_j (p_j^{\mathrm{II}})^q = 1 - (q-1) S_q^{\mathrm{II}}, \tag{29}$$

$$\sum_{i,j} p_{ij}^q = \sum_i (p_i^{\mathrm{I}})^q \sum_j (p_j^{\mathrm{II}})^q = 1 - (q-1) S_q^{\mathrm{I,II}}. \tag{30}$$

From Equations (28) and (29) it also follows that

$$\sum_i (p_i^{\mathrm{I}})^q \sum_j (p_j^{\mathrm{II}})^q = 1 - (q-1) S_q^{\mathrm{I}} - (q-1) S_q^{\mathrm{II}} + (q-1)^2 S_q^{\mathrm{I}} S_q^{\mathrm{II}}. \tag{31}$$

Combining Equations (30) and (31) one ends up with Equation (27). □

Apparently, if we put together two independent subsystems then the Tsallis entropy is not additive but there is a correction term proportional to $q-1$, which vanishes for $q = 1$ only, i.e. for the case where the Tsallis entropy reduces to the Shannon entropy. Equation (27) is sometimes called the 'pseudo-additivity' property.

Equation (27) has given rise to the name *non-extensive statistical mechanics*. If we formulate a generalised statistical mechanics based on maximising Tsallis entropies, then the (Tsallis) entropy of *independent* systems is not additive (Figure 3). However, it turns out that for special types of *correlated* subsystems, the Tsallis entropies do become additive if the subsystems are put together [28]. This means, for these types of correlated complex systems a description in terms of Tsallis entropies in fact can make things simpler as compared to using the Shannon entropy, which is non-additive for correlated subsystems.

2.3. *Landsberg–Vedral entropy*

Let us continue with a few other examples of generalised information measures. Consider

$$S_q^{(\mathrm{L})} = \frac{1}{q-1}\left(\frac{1}{\sum_{i=1}^W p_i^q} - 1\right). \tag{32}$$

This measure was studied by Landsberg and Vedral [29]. One immediately sees that the Landsberg–Vedral entropy is related to the Tsallis entropy $S_q^{(\mathrm{T})}$ by

$$S_q^{(\mathrm{L})} = \frac{S_q^{(\mathrm{T})}}{\sum_{i=1}^W p_i^q}, \tag{33}$$

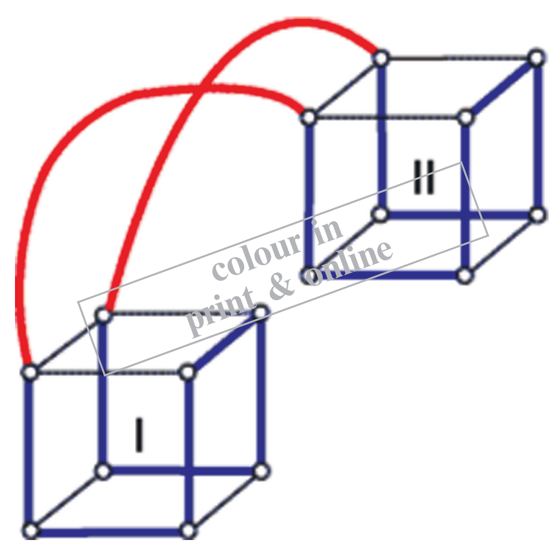

Figure 3. If the nonadditive entropies S_q are used to measure information, then the information contents of two systems I, II (blue) that are put together is not equal to the sum of the information contents of the isolated single systems. In other words, there is always an interaction between the subsystems (red).

and hence $S_q^{(\mathrm{L})}$ is sometimes also called *normalised Tsallis entropy*. $S_q^{(\mathrm{L})}$ also contains the Shannon entropy as a special case

$$\lim_{q\to 1} S_q^{(\mathrm{L})} = S \tag{34}$$

and one readily verifies that it also satisfies a pseudo-additivity condition for independent systems, namely

$$S_q^{(\mathrm{L})\mathrm{I,II}} = S_q^{(\mathrm{L})\mathrm{I}} + S_q^{(\mathrm{L})\mathrm{II}} + (q-1)S_q^{(\mathrm{L})\mathrm{I}} S_q^{(\mathrm{L})\mathrm{II}}. \tag{35}$$

This means that in the pseudo-additivity relation (27) the role of $(q-1)$ and $-(q-1)$ is exchanged.

2.4. Abe entropy

Abe [30] introduced a kind of symmetric modification of the Tsallis entropy, which is invariant under the exchange $q \longleftrightarrow q^{-1}$. This is given by

$$S_q^{\mathrm{Abe}} = -\sum_i \frac{p_i^q - p_i^{q^{-1}}}{q - q^{-1}}. \tag{36}$$

This symmetric choice in q and q^{-1} is inspired by the theory of quantum groups which often exhibits invariance under the 'duality transformation' $q \to q^{-1}$. Like Tsallis entropy, the Abe entropy is also concave. In fact, it is related to the Tsallis entropy S_q^{T} by

$$S_q^{\mathrm{Abe}} = \frac{(q-1)S_q^{\mathrm{T}} - (q^{-1}-1)S_{q^{-1}}^{\mathrm{T}}}{q - q^{-1}}. \tag{37}$$

Clearly the relevant range of q is now just the unit interval (0,1], due to the symmetry $q \to q^{-1}$: Replacing q by q^{-1} in Equation (36) does not change anything.

2.5. Kaniadakis entropy

The *Kaniadakis entropy* (also called κ-entropy) is defined by the following expression [4]

$$S_\kappa = -\sum_i \frac{p_i^{1+\kappa} - p_i^{1-\kappa}}{2\kappa}. \tag{38}$$

Again this is a kind of deformed Shannon entropy, which reduces to the original Shannon entropy for $\kappa = 0$. We also note that for small κ, and by writing $q = 1+\kappa$, $q^{-1} \approx 1-\kappa$, the Kaniadakis entropy approaches the Abe entropy. Kaniadakis was motivated to introduce this entropic form by special relativity: the relativistic sum of two velocities of particles of mass m in special relativity satisfies a similar relation as the Kaniadakis entropy does, identifying $\kappa = 1/mc$. Kaniadakis entropies are also concave and Lesche stable (see Section 3.3).

2.6. Sharma–Mittal entropies

These are two-parameter families of entropic forms [31]. They can be written in the form

$$S_{\kappa,r} = -\sum_i p_i^r \left(\frac{p_i^\kappa - p_i^{-\kappa}}{2\kappa} \right). \tag{39}$$

Interestingly, they contain many of the entropies mentioned so far as special cases. The Tsallis entropy is obtained for $r = \kappa$ and $q = 1$–2κ. The Kaniadakis entropy is obtained for $r = 0$. The Abe entropy is obtained for $\kappa = \frac{1}{2}(q - q^{-1})$ and $r = \frac{1}{2}(q + q^{-1}) - 1$. The Sharma–Mittal entropies are concave and Lesche stable.

3. Selecting a suitable information measure

3.1. Axiomatic foundations

The Khinchin axioms apparently are the right axioms to obtain the Shannon entropy in a unique way, but this concept may be too narrow-minded if one wants to describe general complex systems. In physics, for example, one may be interested in nonequilibrium systems with a stationary state, glassy systems, long transient behaviour in systems with long-range interactions, systems with multifractal phase space structure etc. In all these cases one should be open-minded to allow for generalisations of Axiom 4, since it is this axiom that is least obvious in the given circumstances.

Abe [32] has shown that the Tsallis entropy follows uniquely (up to an arbitrary multiplicative constant) from the following generalised version of the Khinchin axioms. Axioms 1–3 are kept, and Axiom 4 is replaced by the following more general version:

New Axiom 4

$$S_q^{\mathrm{I,II}} = S_q^{\mathrm{I}} + S_q^{\mathrm{II|I}} - (q-1)S_q^{\mathrm{I}} S_q^{\mathrm{II|I}}. \tag{40}$$

Here $S_q^{\mathrm{II|I}}$ is the conditional entropy formed with the conditional probabilities $p(j|i)$ and averaged over all states i using the so-called escort distributions P_i:

$$S_q^{\mathrm{II|I}} = \sum_i P_i S_q(\{p(j|i)\}). \tag{41}$$

Escort distributions P_i were introduced quite generally in [1] and are defined for any given probability distribution p_i by

$$P_i = \frac{p_i^q}{\sum_i p_i^q}. \tag{42}$$

For $q = 1$, the new axiom 4 reduces to the old Khinchin Axiom 4, i.e. $S_q^{\mathrm{I,II}} = S_q^{\mathrm{I}} + S_q^{\mathrm{II|I}}$. For independent systems I and II, the new Axiom 4 reduces to the pseudo-additivity property (27).

The meaning of the new Axiom 4 is quite clear. It is a kind of minimal extension of the old Axiom 4: if we collect information from two subsystems, the total information should be the sum of the information collected from system I and the conditional information from system II, plus a correction term. This correction term can *a priori* be anything, but we want to restrict ourselves to information measures where

$$S^{\mathrm{I,II}} = S^{\mathrm{I}} + S^{\mathrm{II|I}} + g(S^{\mathrm{I}}, S^{\mathrm{II|I}}), \tag{43}$$

where $g(x, y)$ is some function. The property that the entropy of the composed system can be expressed as a function of the entropies of the single systems is sometimes referred to as the *composability* property. Clearly, the function g must depend on the entropies of *both* subsystems, for symmetry reasons. The simplest form one can imagine is that it is given by

$$g(x, y) = \mathrm{const} \cdot xy, \tag{44}$$

i.e. it is proportional to both the entropy of the first system and that of the second system. Calling the proportionality constant $(q - 1)$, we end up with the new Axiom 4.

It should, however, be noted that we may well formulate other axioms, which then lead to other types of information measures. The above generalisation is perhaps the one that requires least modifications as compared to the Shannon entropy case. But clearly, depending on the class of complex systems considered, and depending on what properties we want to describe, other axioms may turn out to be more useful. For example, Wada and Suyari [33] have suggested a set of axioms that uniquely lead to the Sharma–Mittal entropy.

3.2. *Composability*

Suppose we have a given complex system which consists of subsystems that interact in a complicated way. Let us first analyse two subsystems I and II in an isolated way and then put these two dependent systems I and II together. Can we then express the generalised information we have on the total system as a simple function of the information we have on the single systems? This question is sometimes referred to as the composability problem.

The Tsallis entropies are composable in a very simple way. Suppose the two systems I and II are not independent. In this case one can still write the joint probability p_{ij} as a product of the single probability p_i and conditional probability $p(j|i)$, i.e. the probability of event j under the condition that event i has already occurred is

$$p_{ij} = p(i|j)p_j. \tag{45}$$

The conditional Tsallis entropy associated with system II (under the condition that system I is in state i) is given by

$$S_q^{\mathrm{II}|i} = \frac{1}{q-1}\left(1 - \sum_j p(j|i)^q\right). \tag{46}$$

One readily verifies the relation

$$S_q^{\mathrm{I}} + \sum_i p_i^q S_q^{\mathrm{II}|i} = S_q^{\mathrm{I,II}}. \tag{47}$$

This equation is very similar to that satisfied by the Shannon entropy in Axiom 4. In fact, the only difference is that there is now an exponent q that wasn't there before. It means our collection of information is biased: instead of weighting the events in system I with p_i we weight them with p_i^q. For $q = 1$ the above equation of course reduces to the fourth of the Khinchin axioms, but only in this case. Hence, for general $q \neq 1$, the Tsallis information is *not* independent of the way it is collected for the various subsystems.

To appreciate the simple composability property of the Tsallis entropy, let us compare with other entropy-like functions, for example the Rényi entropy. For the Rényi entropy there is no simple composability property similar to Equation (47). Only the exponential of the Renyi entropy satisfies a relatively simple equation, namely

$$\exp\left((q-1)S_q^{(\mathrm{R})\mathrm{I,II}}\right) = \sum_i p_i^q \exp\left((q-1)S_q^{(\mathrm{R})\mathrm{II}|i}\right). \tag{48}$$

However, by taking the exponential one clearly removes the logarithm in the definition of the Rényi entropies in Equation (15). This means one is effectively back to the Tsallis entropies.

3.3. *Lesche stability*

Physical systems contain noise. A necessary requirement for a generalised entropic form $S[p]$ to make physical sense is that it must be stable under small perturbations. This means a small perturbation of the set of probabilities $p := \{p_{\mathrm{i}}\}$ to a new set $p' = \{p_{i'}\}$ should have only a small effect on the value $S_{\max}$ of $S_{\mathrm{q}}[p]$ in the thermodynamic state that maximises the entropy. This should in particular be true in the limit

$W \to \infty$ (recall that W denotes the number of microstates). The stability condition can be mathematically expressed as follows [34].

Stability condition

For every $\varepsilon > 0$ there is a $\delta > 0$ such that

$$||p - p'||_1 \leq \delta \Rightarrow \left| \frac{S[p] - S[p']}{S_{\max}} \right| < \varepsilon \tag{49}$$

for arbitrarily large W. Here $||A||_1 = \sum_{i=1}^{W} |A_i|$ denotes the L_1 norm of an observable A.

Abe [35] has proved that the Tsallis entropies are Lesche stable, i.e. they satisfy Equation (49) for all q, whereas the Rényi entropies and the Landsberg entropies are not stable for any $q \neq 1$ (for a discrete set of probabilities p_i with $W \to \infty$). This is an important criterion to single out generalised entropies that may have potential physical meaning. According to the stability criterion, the Tsallis entropies are stable and thus may be associated with physical states, whereas the other two examples of entropic forms have a stability problem. Kaniadakis entropies and Sharma–Mittal entropies are also Lesche stable. Only entropies that are Lesche stable are good candidates for physically relevant information measures. For a recent re-investigation of these types of stability problems in a physical setting, see [36].

4. Maximising generalised entropies

4.1. A rigorous derivation of statistical mechanics?

The rigorous foundations of statistical mechanics are a kind of miracle. There is little progress in rigorously deriving statistical mechanics from the microscopic classical Hamiltonian equations of motion, neither there is a rigorous derivation starting from quantum mechanics or quantum field theory. It is almost surprising how well statistical mechanics works in daily life, given its lack of rigorous derivation from other microscopic theories.

The problem is that for a rigorous derivation of statistical mechanics from dynamical systems theory one needs the underlying dynamical system to be *ergodic*, and even that is not enough: it should have the stronger property of *mixing*. Ergodicity essentially means that typical trajectories fill out the entire phase space (which implies that for typical trajectories the time average is equal to the ensemble average) and mixing means asymptotic independence, i.e. the correlation function of distant events decays to zero if the time difference between the events goes to infinity. For strongly chaotic dynamical systems (i.e. those exhibiting exponential sensitive dependence on the initial conditions) one normally expects the mixing property to hold (though there are some mathematical subtleties here). From a mathematical point of view, the mixing property is the theoretical ingredient that is needed to guarantee the approach to an equilibrium state in statistical mechanics.

Unfortunately, ergodicity and mixing can only be rigorously proved for simple toy examples of dynamical systems, for example the discrete-time map $x_{n+1} = 1 - 2x_n^2$ with initial values in the interval $[-1,1]$ or other very simple toy models (see, e.g. [1]). For realistic systems of physical relevance, such as the Hamiltonian equations of a large number of weakly or strongly interacting particles, a rigorous mathematical proof of the mixing property does not exist, and the deeper reason why statistical mechanics works so well in typical situations remains a miracle.

4.2. Jaynes' information theory

In view of the fact that there are no rigorous foundations of statistical mechanics, one usually sticks to some simple principle such as the maximum entropy principle in order to 'derive' it. Jaynes has given a simple and plausible interpretation of the maximum entropy principle [37]. His interpretation is purely based on concepts from information theory, and applicable to many problems, not only to equilibrium statistical mechanics.

In simple words, the idea is as follows. Assume we have only limited information on a system containing many particles or constituents. We may know the mean values of some observables M^σ, $\sigma = 1, \ldots, s$ but nothing else. For example, we may just know one such quantity, the mean energy of all particles and nothing else ($s = 1$). What probability distributions p_i should we now assume, given that we have such limited information on the system?

Suppose we measure information with some information measure $I(\{p_i\}) =: I[p]$. Among all distributions possible that lead to the above known mean values M^σ we should select those that do not contain any unjustified prejudices. In other words, our information measure for the relevant probabilities should take on a minimum, or the entropy (= negative information) should take a maximum, given the constraints. For, if the information associated with the selected probability distribution does not take on a minimum, we have more information than the minimum information, but this means we are pre-occupied by a certain belief or additional information, which we should have entered as a condition of constraint in the first place.

Of course, if we have no knowledge on the system at all ($s = 0$), the principle yields the uniform distribution $p_i = 1/W$, $i = 1, \ldots, W$ of events. For this to happen, the information measure $I[p]$ must only satisfy the second Khinchin axiom, nothing else. In statistical mechanics, the corresponding ensemble is the microcanonical ensemble.

If some constraints are given, we have to minimise the information (= maximise the entropy) subject to the given constraints. A constraint means that we know that some observable $\tilde{M}$ of the system, which takes on the values M_i in the microstates i, takes on the fixed mean value M. In total, there can be s such constraints, corresponding to s different observables $\tilde{M}^\sigma$:

$$\sum_i p_i M_i^\sigma = M^\sigma \quad (\sigma = 1, \ldots, s). \tag{50}$$

For example, for the canonical ensemble of equilibrium statistical mechanics one has the constraint that the mean value U of the energies E_i in the various microstates is fixed:

$$\sum_i p_i E_i = U. \tag{51}$$

We may also regard the fact that the probabilities p_i are always normalised as a constraint obtained for the special choice $\tilde{M} = 1$:

$$\sum_i p_i = 1. \tag{52}$$

To find the distributions that maximise the entropy under the given constraints one can use the method of Lagrange multipliers. One simply defines a function $\Psi[p]$ which is the information measure under consideration plus the condition of constraints multiplied by some constants β_σ (the Lagrange multipliers):

$$\Psi[p] = I[p] + \sum_\sigma \beta_\sigma \left(\sum_i p_i M_i^\sigma \right). \tag{53}$$

One then looks for the minimum of this function in the space of all possible probabilities p_i. In practice, these distributions p_i are easily obtained by evaluating the condition

$$\frac{\partial}{\partial p_i} \Psi[p] = 0 \quad (i = 1, \ldots, W), \tag{54}$$

which means that Ψ has a local extremum. We obtain

$$\frac{\partial}{\partial p_i} I[p] + \sum_\sigma \beta_\sigma M_i^\sigma = 0. \tag{55}$$

All this is true for *any* information measure $I[p]$, it need not be the Shannon information. At this point we see why it is important that the information measure $I[p]$ is convex: we need a well-defined inverse function of $(\partial/\partial p_i)I[p]$, in order to uniquely solve Equation (55) for the p_i. This means $(\partial/\partial p_i)I[p]$ should be a monotonous function, which means that $I[p]$ must be convex.

Note that Jaynes' principle is (in principle) applicable to all kinds of complex systems, many different types of observables, and various types of information measures. There is no reason to restrict it to equilibrium statistical mechanics only. It is generally applicable to all kinds of problems where one has missing information on the actual microscopic state of the system and wants to make a good (unbiased) guess of what is happening and what should be done. The concept of avoiding unjustified prejudices applies in quite a general way. An important question is which information measure is relevant for which system. Clearly, the Shannon entropy is the right information measure to analyse standard types of systems in equilibrium statistical mechanics. But other systems of more complex nature can potentially be described more effectively if one uses different information measures, for examples those introduced in the previous section.

4.3. *Ordinary statistical mechanics*

For ordinary statistical mechanics, one has $I[p] = \sum_i p_i \ln p_i$ and $(\partial/\partial p_i)I[p] = 1 + \ln p_i$. For the example of a canonical ensemble Equation (53) reads

$$\Psi[p] = \sum_i p_i \ln p_i + \alpha \sum_i p_i + \beta \sum_i p_i E_i \tag{56}$$

and Equation (55) leads to

$$\ln p_i + 1 + \alpha + \beta E_i = 0. \tag{57}$$

Hence, the maximum entropy principle leads to the canonical distributions

$$p_i = \frac{1}{Z} \exp(-\beta E_i). \tag{58}$$

The partition function Z is related to the Lagrange multiplier α by

$$Z := \sum_i \exp(-\beta E_i) = \exp(1 + \alpha). \tag{59}$$

4.4. *Generalised statistical mechanics*

More generally we may start from a generalised information measure of the trace form

$$I[p] = -S[p] = \sum_i p_i h(p_i), \tag{60}$$

where h is some suitable function, as introduced before. Tsallis entropy, Abe entropy, Kaniadakis entropy, Sharma–Mittal entropy and Shannon entropy are examples that can all be cast into this general form, with different functions h of course. Again let us consider the canonical ensemble (the extension to further constraints/other ensembles is straightforward). The functional to be maximised is then

$$\Psi[p] = \sum_i p_i h(p_i) + \alpha \sum_i p_i + \beta \sum_i p_i E_i. \tag{61}$$

By evaluating the condition

$$\frac{\partial}{\partial p_i} \Psi[p] = 0 \tag{62}$$

we obtain

$$h(p_i) + p_i h'(p_i) + \alpha + \beta E_i = 0. \tag{63}$$

Defining a function g by

$$g(p_i) := h(p_i) + p_i h'(p_i), \tag{64}$$

we end up with

$$g(p_i) = -\alpha - \beta E_i. \tag{65}$$

Hence, if a unique inverse function g^{-1} exists, we have

$$p_i = g^{-1}(-\alpha - \beta E_i) \tag{65}$$

and this is the generalised canonical distribution of the generalised statistical mechanics.

Let us consider a few examples of interesting functions h. For the Shannon entropy one has of course

$$h(p_i) = \ln p_i. \tag{67}$$

For the Tsallis entropy,

$$h(p_i) = \frac{p_i^{q-1} - 1}{q-1} =: \log_{2-q}(p_i). \tag{68}$$

This is like a deformed logarithm that approaches the ordinary logarithm for $q \to 1$. In fact, a useful definition commonly used in the field is the so-called q-logarithm defined by

$$\log_q(x) := \frac{x^{1-q} - 1}{1-q}. \tag{69}$$

Its inverse function is the q-exponential

$$\exp_q(x) := (1 + (1-q)x)^{1/(1-q)}. \tag{70}$$

For the Kaniadakis entropy one has

$$h(p_i) = \frac{p_i^{\kappa} - p_i^{-\kappa}}{2\kappa} =: \ln_\kappa(x), \tag{71}$$

where the κ-logarithm is defined as

$$\ln_\kappa(x) = \frac{x^\kappa - x^{-\kappa}}{2\kappa}. \tag{72}$$

Its inverse is the κ-exponential

$$\exp_\kappa(x) = \left((1 + \kappa^2 x^2)^{1/2} + \kappa x\right)^{1/\kappa}. \tag{73}$$

Essentially, the generalised canonical distributions obtained by maximising Tsallis entropies are given by q-exponentials of the energy E_{i} and those by maximising Kaniadakis entropies are κ-exponentials. Both decay with a power law for large values of the energy E_{i}.

4.5. *Nonextensive statistical mechanics*

Let us consider in somewhat more detail a generalised statistical mechanics based on Tsallis entropies. If we start from the Tsallis entropies $S_q^{(\mathrm{T})}$ and maximise those subject to suitable constraint, the corresponding formalism is called *nonextensive statistical mechanics*. We have

$$I_q^{(\mathrm{T})}[p] = -S_q^{(\mathrm{T})}[p] = \frac{-1}{q-1}\left(1 - \sum_i p_i^q\right), \tag{74}$$

thus

$$\frac{\partial}{\partial p_i} I_q^{(\mathrm{T})}[p] = \frac{q}{q-1} p_i^{q-1}. \tag{75}$$

For a canonical ensemble Equation (55) leads to

$$\frac{q}{q-1} p_i^{q-1} + \alpha + \beta E_i = 0. \tag{76}$$

Thus, the maximum entropy principle leads to generalised canonical distributions of the form

$$p_i = \frac{1}{Z_q}(1 - \beta(q-1)E_i)^{1/(q-1)}, \tag{77}$$

where Z_q is a normalisation constant. This is the original formula Tsallis introduced in his paper [2]. These days, however, the convention has become to replace the parameter q by $q' = 2 - q$ and then rename $q' \to q$. That is to say, the generalised canonical distributions in nonextensive statistical mechanics are given by the following q-exponentials:

$$p_i = \frac{1}{Z_q}(1 + \beta(q-1)E_i)^{-1/(q-1)}. \tag{78}$$

They live on a bounded support for $q < 1$ and exhibit power-law decays for $q > 1$.

Starting from such a q-generalised approach, one can easily derive formal q-generalised thermodynamic relations. The details depend a bit on how the constraints on energy are taken into account [3]. All relevant thermodynamic quantities now get an index q. Typical examples of such formulas are

$$1/T = \beta = \partial S_q^{(\mathrm{T})} / \partial U_q, \quad \forall q \tag{79}$$

with

$$\sum_{i=1}^{W} (p_i)^q = (\bar{Z}_q)^{1-q}, \tag{80}$$

$$F_q \equiv U_q - TS_q = -\frac{1}{\beta}\frac{(Z_q)^{1-q} - 1}{1-q} \tag{81}$$

and

$$U_q = -\frac{\partial}{\partial\beta}\frac{(Z_q)^{1-q} - 1}{1-q}, \tag{82}$$

where

$$\frac{(Z_q)^{1-q} - 1}{1-q} = \frac{(\bar{Z}_q)^{1-q} - 1}{1-q} - \beta U_q \tag{83}$$

and

$$\sum_{i=1}^{W} P_i E_i = \frac{\sum_{i=1}^{W} p_i^q E_i}{\sum_{i=1}^{W} p_i^q} = U_q. \tag{84}$$

There are some ambiguities on how to take into account the constraints, using for example the original p_i or the escort distributions P_i, but we will not comment on these technicalities here.

5. Some physical examples

5.1. *Making contact with experimental data*

It should be clear that direct physical measurements of generalised entropy measures are impossible since these are basically man-made information-theoretic tools. However, what can be measured is the stationary probability distribution of certain observables of a given complex system, as well as possibly some correlations between subsystems. As we have illustrated before, measured probability densities in some complex system that deviate from the usual Boltzmann factor $\exp(-\beta E)$ can then be formally interpreted as being due to the maximisation of a more general information measure that is suitable as an effective description for the system under consideration.

In this approach one regards the complex system as a kind of 'black box'. Indeed many phenomena in physics, biology, economics, social sciences, etc. are so complicated that there is not a simple equation describing them, or at least we do not know this equation. *A priori* we do not know what is the most suitable way to measure information for any output that we get from our black box. But if a distribution p_i of some observable output is experimentally measured, we can indirectly construct a generalised entropic form that takes a maximum for this particular observed distribution. This allows us to make contact with experimental measurements, make some predictions, e.g. on correlations of subsystems and translate the rather abstract information theoretical concepts into physical reality.

5.2. *Statistics of cosmic rays*

Our first example of making contact to concrete measurements is cosmic ray statistics [8]. The Earth is constantly bombarded with highly energetic particles, cosmic rays. Experimental data of the measured cosmic ray energy spectrum are shown in Figure 4. It has been known for a long time that the observed distribution of cosmic rays with a given energy E

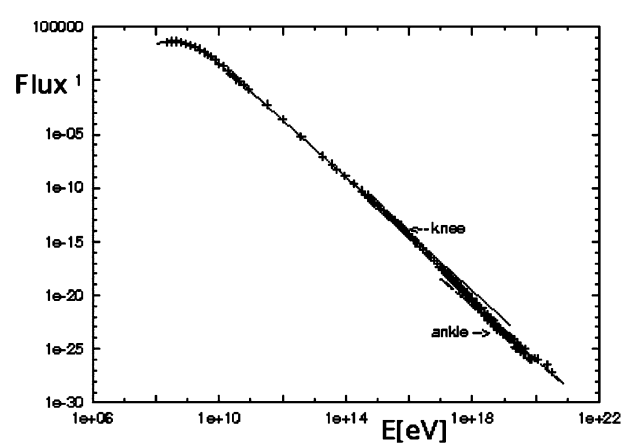

Figure 4. Measured energy spectrum of cosmic rays and a fit by Equation (85) with $q = 1.215$. The 'knee' and 'ankle' are structures that go beyond the simple model considered here.

exhibits strongly pronounced power laws rather than exponential decay. It turns out that the observed distribution is very well fitted over a very large range of energies by the formula

$$p(E) = C \cdot \frac{E^2}{(1 + b(q-1)E)^{1/(q-1)}}. \qquad (85)$$

Here E is the energy of the cosmic ray particles,

$$E = \left(c^2 p_x^2 + c^2 p_y^2 + c^2 p_z^2 + m^2 c^4 \right)^{1/2}, \qquad (86)$$

$b = (k\tilde{T})^{-1}$ is an effective inverse temperature variable, and C is a constant representing the total flux rate. For highly relativistic particles the rest mass m can be neglected and one has $E \approx c|\boldsymbol{p}|$. The reader immediately recognises the distribution (85) as a q-generalised relativistic Maxwell–Boltzmann distribution, which maximises the Tsallis entropy. The factor E^2 takes into account the available phase space volume. As seen in Figure 4, the cosmic ray spectrum is very well fitted by the distribution (85) if the entropic index q is chosen as $q = 1.215$ and if the effective temperature parameter is given by $k\tilde{T} = b^{-1} = 107$ MeV. Hence, the measured cosmic ray spectrum effectively maximises the Tsallis entropy.

The deeper reason why this is so could be temperature fluctuations during the production process of the primary cosmic ray particles [8]. Consider quite generally a superposition of ordinary Maxwell–Boltzmann distributions with different inverse temperatures β:

$$p(E) \sim \int f(\beta) E^2 \exp(-\beta E)\, d\beta. \qquad (87)$$

Here $f(\beta)$ is the probability density to observe a given inverse temperature β. If $f(\beta)$ is a Gamma distribution, then the integration in Equation (87) can be performed and one ends up with Equation (85) (see [8] for more details). This is the basic idea underlying so-called superstatistical models [12]: one does a kind of generalised statistical mechanics where the inverse temperature β is a random variable as well.

The effective temperature parameter $\tilde{T}$ (a kind of average temperature in the above superstatistical model) is of the same order of magnitude as the so-called Hagedorn temperature T_H [38], an effective temperature well known from collider experiments. The fact that we get from the fits something of the order of the Hagedorn temperature is encouraging. The Hagedorn temperature is much smaller than the centre-of-mass energy E_{CMS} of a typical collision process and represents a kind of 'boiling temperature' of nuclear matter at the confinement phase transition. It is a kind of maximum temperature that can be reached in a collision experiment. Even the largest E_{CMS} cannot produce a larger average temperature than T_H due to the fact that the number of possible particle states grows exponentially.

Similar predictions derived from nonextensive statistical mechanics also fit measured differential cross-sections in e^+e^-annihilation processes and other scattering data very well (see e.g. [16,17] for more details). The hadronic cascade process underlying these scattering data is not well understood, though it can be simulated by Monte Carlo simulations. If we don't have any better theory, then the simplest model to reproduce the measured cross-sections is indeed a generalised Hagedorn theory where the Shannon entropy is replaced by Tsallis entropy [17].

5.3. *Defect turbulence*

Our next example is the so-called 'defect turbulence'. Defect turbulence shares with ordinary turbulence only the name as otherwise it is very different. It is a phenomenon related to convection and has nothing to do with fully developed hydrodynamic turbulence. Consider a Rayleigh–Bénard convection experiment: a liquid is heated from below and cooled from above. For large enough temperature differences, interesting convection patterns start to evolve. An inclined layer convection experiment [9] is a kind of Rayleigh–Bénard experiment where the apparatus is tilted by an angle (say 30°), moreover, the liquid is confined between two very narrow plates. For large temperature differences, the convection rolls evolve chaotically. Of particular interest are the defects in this pattern, i.e. points where two convection rolls merge into one (see Figure 5). These defects behave very much like particles. They have a well-defined position and velocity, they are created and annihilated in pairs, and one can even formally attribute a 'charge' to them: there are positive and negative defects, as indicated by the black and white boxes in Figure 5. But the theory underlying these highly nonlinear excitations is pretty unclear, they are like a 'black box' complex system whose measured output is velocity.

The probability density of defect velocities has been experimentally measured with high statistics [9]. As shown in Figure 6, the measured distribution is well fitted by a q-Gaussian with $q \approx 1.45$. The defects are also observed to exhibit anomalous diffusion. Their position $X(t)$ roughly obeys an anomalous diffusion law of the type

$$\langle X^2(t) \rangle \sim t^\alpha, \qquad (88)$$

with $\alpha \approx 1.3$. The relation $\alpha \approx 2/(3-q)$ can be theoretically derived [9].

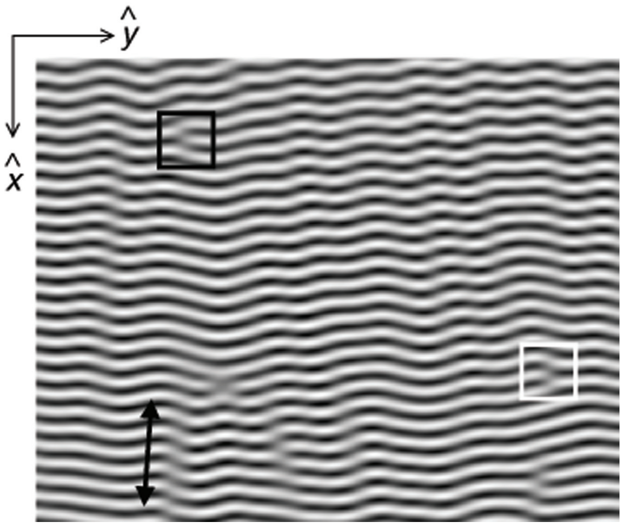

Figure 5. Convection rolls and defects (black and white boxes) as observed in the experiment of Daniels et al. [9]

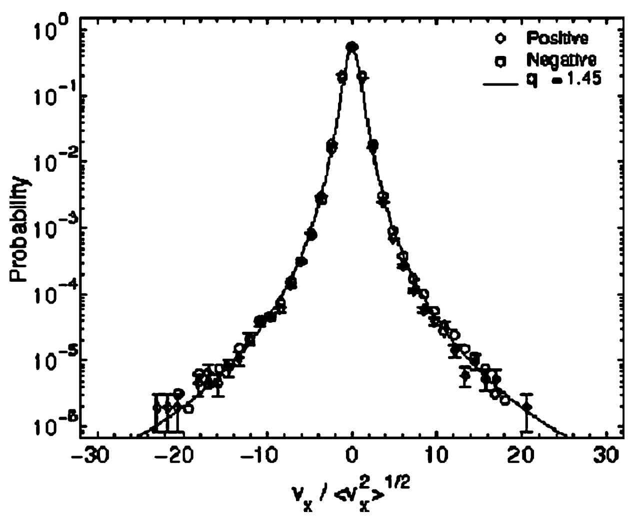

Figure 6. Measured probability density of defect velocities and fit with a q-Gaussian with $q = 1.45$.

Apparently defects are a very complicated nonlinear system with complicated interactions in a nonequilibrium environment. Their dynamics is not fully understood so far. But we see that effectively they seem to behave like a gas of nonextensive statistical mechanics that leads to q-exponential Boltzmann factors rather than ordinary Boltzmann factors.

5.4. *Optical lattices*

Optical lattices are standing-wave potentials obtained by superpositions of counter-propagating laser beams. One obtains easily tunable periodic potentials in which atoms can perform normal and anomalous quantum transport processes. If the potential is very deep, there is diffusive motion. If it is very shallow, there is ballistic motion. In between, there is a regime with anomalous diffusion that is of interest here.

Optical lattices can be theoretically described by a nonlinear Fokker–Planck equation for the Wigner function $W(p,t)$ (the Wigner function is an important statistical tool for the quantum mechanical description in the phase space). The above Fokker–Planck equation admits Tsallis statistics as a stationary solution. This was pointed out by Lutz [10]. The equation is given by

$$\frac{\partial W}{\partial t} = -\frac{\partial}{\partial p}[K(p)W] + \frac{\partial}{\partial p}\left[D(p)\frac{\partial W}{\partial p}\right], \tag{89}$$

where

$$K(p) = -\frac{\alpha p}{1+(p/p_c)^2} \tag{90}$$

is a momentum-dependent drift force and

$$D(p) = D_0 + \frac{D_1}{1+(p/p_c)^2} \tag{91}$$

is a momentum-dependent diffusion constant. The stationary solution is

$$W(p) = C\frac{1}{(1+\beta(q-1)E)^{1/(q-1)}}, \tag{92}$$

where

$$E = \frac{1}{2}p^2, \tag{93}$$

$$\beta = \frac{\alpha}{D_0 + D_1}, \tag{94}$$

$$q = 1 + \frac{2D_0}{\alpha p_c^2}. \tag{95}$$

So the optical lattice effectively maximises Tsallis entropy in its nonequilibrim stationary state. Another way to express the entropic index in terms of physical parameters is the formula

$$q = 1 + \frac{44E_{\mathrm{R}}}{U_0}, \tag{96}$$

where E_{R} is the so-called recoil energy and U_0 is the potential depth. These types of q-exponential predictions have been experimentally confirmed [11].

Lutz' microscopic theory thus yields a theory of the relevant entropic index q in terms of system parameters.

5.5. *Epilogue*

There are many other examples of physical systems where generalised entropies yield a useful tool to effectively describe the complex system under consideration. Important examples include Hamiltonian systems with long-range interactions that exhibit metastable states [19,20] as well as driven nonequilibrium systems with large-scale fluctuations of temperature or energy dissipation, i.e. superstatistical systems [12,39,40]. The best way to define generalised entropies for superstatistical systems is still the subject of current research [6,41,42]. Superstatistical turbulence models yield excellent agreement with experimental data [13–15]. Generalised statistical mechanics methods have also applications outside physics, for example in mathematical finance [22,43], for traffic delay statistics [44], for biological systems [7], or in the medical sciences [24]. It is often in these types of complex systems that one does not have a concrete equation of motion and hence is forced to do certain 'unbiased guesses' on the behaviour of the system – which for sufficiently complex systems may lead to other entropic forms than the usual Shannon entropy that are effectively maximised. The beauty of the formalism is that it can be applied to a large variety of complex systems from different subject areas, without knowing the details of the dynamics.

Note

1. An exception to this rule is the Fisher information, which depends on gradients of the probability density but will not be discussed here.

Notes on contributor

Christian Beck obtained his Ph.D. in Physics at the Technical University of Aachen in 1988. After postdoctoral positions at the Universities of Warwick, Copenhagen, Budapest and Maryland, he became a Lecturer in Applied Mathematics at Queen Mary, University of London, in 1994. He was subsequently promoted to Reader and Professor and is currently Director of Applied Mathematics at Queen Mary. In 1999 he obtained a Royal Society Senior Research Fellowship and in 2005 an EPSRC Springboard Fellowship. He is an Advisory Editor of *Physica A* and author of nearly 100 publications, including two books. His research interests cover a broad range of different subject areas, in particular the mathematical modelling of complex systems using novel techniques from statistical mechanics and dynamical systems theory.

References

[1] C. Beck and F. Schlögl, *Thermodynamics of Chaotic Systems*, Camdridge University Press, Cambridge, UK, 1993.
[2] C. Tsallis, *Possible generalization of Boltzmann-Gibbs statistics*, J. Stat. Phys. 52 (1988), pp. 479–487.
[3] C. Tsallis, R.S. Mendes, and A.R. Plastino, *The role of constraints within generalized nonextensive statistics*, Physica A 261 (1998), pp. 534–554.
[4] G. Kaniadakis, *Statistical mechanics in the context of special relativity*, Phys. Rev. E 66 (2002), 056125.
[5] J. Naudts, *Generalized thermostatistics based on deformed exponential and logarthemic functions*, Physica A 340 (2004), pp. 32–40.
[6] S. Abe, C. Beck, and E.G.D. Cohen, *Superstatistics, thermodynamics, and fluctuations*, Phys. Rev. E 76 (2007), 031102.
[7] P.-H. Chavanis, *Nonlinear mean field Fokker-Planck equations. Application to the chemotaxis of biological populations*, Eur. Phys. J. B 62 (2008), pp. 179–208.
[8] C. Beck, *Generalized statistical machanics of cosmic rays*, Physica A 331 (2004), pp. 173–181.
[9] K.E. Daniels, C. Beck, and E. Bodenschatz, *Defect turbulence and generalized statistical machanics*, Physica D 193 (2004), pp. 208–217.
[10] E. Lutz, *Anomalous diffusion and Tsallis statistics in an optical lattice*, Phys. Rev. A 67 (2003), 051402.
[11] P. Douglas, S. Bergamini, and F. Renzoni, *Tunable Tsallis distributions in dissipative optical lattices*, Phys. Rev. Lett. 96 (2006), 110601.
[12] C. Beck and E.G.D. Cohen, *Superstatistics*, Physica A 322 (2003), pp. 267–275.
[13] C. Beck, E.G.D. Cohen, and H.L. Swinney, *From time series to superstatistics*, Phys. Rev. E 72 (2005), 056133.
[14] A. Reynolds, *Superstatistical machanics of tracer-particle motions in turbulence*, Phys. Rev. Lett. 91 (2003), 084503.
[15] C. Beck, *Statistics of three-dimentional Lagrangian turbulence*, Phys. Rev. Lett. 98 (2007), 064502.
[16] I. Bediaga, E.M.F. Curado, and J.M. de Miranda, *A nonextensive thermodynamic equilibrium approach in $e^+e^- \rightarrow hadrons$*, Physica A 286 (2000), pp. 156–163.
[17] C. Beck, *Non-extensive statistical mechanics and particle spectra in elementary interactions*, Physica A 286 (2000), pp. 164–180.
[18] A.R. Plastino and A. Plastino, *Stellar polytropes and Tsallis' entropy*, Phys. Lett. A 174 (1993), pp. 384–386.
[19] P.-H. Chavanis, *Generalized thermodynamics and Fokker-Planck equations: Applications to stellar dynamics and two-dimensional turbulence*, Phys. Rev. E 68 (2003), 036108.
[20] A. Pluchino, V. Latora, and A. Rapisarda, *Glassy dynamics in the HMF model*, Physica A 340 (2004), pp. 187–195.
[21] A. Pluchino, A. Rapisarda, and C. Tsallis, *Nonergodicity and central-limit behavior for long-range Hamiltonians*, Europhys. Lett. 80 (2007), 26002.
[22] L. Borland, *Option pricing formulas based on a non-Gaussian stock price model*, Phys. Rev. Lett. 89 (2002), 098701.
[23] A. Upadhyaya, J.-P. Rieu, J.A. Glazier, and Y. Sawada, *Anomalous diffusion and non-Gaussian velocity distribution of Hydra cells in cellular aggregates*, Physica A 293 (2001), pp. 549–558.

[24] L.L. Chen and C. Beck, *A superstatistical model of metastasis and cancer survival*, Physica A 387 (2008), pp. 3162–3172.
[25] A.I. Khinchin, *Mathematical Foundations of Information Theory*, Dover, New York, 1957.
[26] A. Rényi, *Probability Theory*, North Holland, Amsterdam, 1970.
[27] J.H. Havrada and F. Charvat, *Quantification methods of classification processes: Concepts of structural α entropy*, Kybernetica 3 (1967), pp. 30–35.
[28] C. Tsallis, M. Gell-Mann, and Y. Sato, *Asymptotically scale-invariant occupancy of phase space makes the entropy S_q extensive*, Proc. Nat. Acad. Sci. 102 (2005), pp. 15377–15382.
[29] P.T. Landsberg and V. Vedral, *Distributions and channel capacities in generalized statistical mechanics*, Phys. Lett. A 247 (1998), pp. 211–217.
[30] S. Abe, *A note on the q-deformation-theoretic aspect of the generalized entropies in nonextensive physics*, Phys. Lett. A 224 (1997), pp. 326–330.
[31] B.D. Sharma and D.P. Mittal, J. Math. Sci. 10 (1975), pp. 28–40.
[32] S. Abe, *Axioms and uniqueness theorem for Tsallis entropy*, Phys. Lett. A 271 (2000), pp. 74–79.
[33] T. Wada and H. Suyari, *A two-parameter generalization of Shannon-Khinchin axioms and the uniqueness theorem,* cond-mat/0608139.
[34] B. Lesche, *Instabilities of Rényi entropies*, J. Stat. Phys. 27 (1982), pp. 419–422.
[35] S. Abe, *Stability of Tsallis entropy and instabilities of Rényi and normalized Tsallis entropies: A basis for q-exponential distributions*, Phys. Rev. E 66 (2002), 046134.
[36] R. Hanel, S. Thurner, and C. Tsallis, *On the robustness of q-expectation values and Rényi entropy*, Europhys. Lett. 85 (2009), 20005.
[37] E.T. Jaynes, *Information theory and statistical mechanics*, Phys. Rev. 106 (1957), pp. 620–630.
[38] R. Hagedorn, *Statistical thermodynamics of strong interactions at high energies*, Nuovo Cim. Suppl. 3 (1965), pp. 147–186.
[39] G. Wilk and Z. Wlodarczyk, *Interpretation of the nonextensive parameter q in some applications of Tsallis statistics and Lévy distributions*, Phys. Rev. Lett. 84 (2000), pp. 2770–2773.
[40] C. Beck, *Dynamical foundations of nonextensive statistical mechanics*, Phys. Rev. Lett. 87 (2001), 180601.
[41] E. Van der Straeten and C. Beck, *Superstatistical distributions from a maximum entropy principle*, Phys. Rev. E 78 (2008), 051101.
[42] C. Tsallis and A.M.C. Souza, *Constructing a statistical mechanics for Beck-Cohen superstatistics*, Phys. Rev. E 67 (2003), 026106.
[43] J.-P. Bouchard and M. Potters, *Theory of Financial Risk and Derivative Pricing*, Cambridge University Press, Cambridge, UK, 2003.
[44] K. Briggs and C. Beck, *Modelling train delays with q-exponential functions*, Physica A 378 (2007), pp. 498–504.