



# Journal of Applied Sciences

ISSN 1812-5654

**science**  
alert

**ANSI***net*  
an open access publisher  
<http://ansinet.com>

## Concerning Tsallis's Condition of Pseudo-additivity as a Definition of Non-extensivity and the Use of the Maximum Entropy Formalism

<sup>1</sup>B.H. Lavenda and <sup>2</sup>J. Dunning-Davies

<sup>1</sup>Università degli Studi Camerino 62032 (MC) Italy

<sup>2</sup>Department of Physics, University of Hull, Hull HU6 7RX, England

---

**Abstract:** The pseudo-additive relation satisfied by the Tsallis entropy is not linked with the super- or sub-additive properties of the entropy. These latter properties, like concavity and convexity, are fundamentally geometric inequalities and cannot be reduced to equalities. The pseudo-additivity relation is, rather, a functional equation that determines the functional forms of the random entropies. A similar pseudo-additive relation is satisfied by the Arimoto entropy which is a first-order homogeneous form. No conclusions, based on the pseudo-additive functional equation, may be drawn about the extensive nature of systems from either the Tsallis or Arimoto entropy. Further, it is shown that Tsallis' statistical thermodynamic formulation of the non-additive entropy of degree- $\alpha$  is neither correct nor self-consistent.

**Key words:** Pseudo-additive entropies, maximum entropy formalism, non-extensivity, homogeneity, Tsallis and Arimoto entropies, Gauss's principle

---

### INTRODUCTION

According to Tsallis<sup>[1]</sup>, the entropy of a composite system composed of two independent systems A and B is given by the functional equation:

$$S_{\alpha}(A+B) = S_{\alpha}(A) + S_{\alpha}(B) + (1-\alpha)S_{\alpha}(A)S_{\alpha}(B), \quad (1)$$

where:

$$S_{\alpha}(P) = \frac{1 - \sum_{i=1}^m p_i^{\alpha}}{\alpha - 1} \quad (2)$$

is the Tsallis entropy for all  $\alpha > 0$  and energy units for which  $k = 1$  have been used. Also, the probability distribution  $P = (p_1, \dots, p_m)$  is complete. However, this pseudo-additive relation has nothing to do with the super- and sub-additivity properties of the entropy. These latter properties, like concavity and convexity, are couched in geometric inequalities and simply cannot be reduced to equalities. Rather, (1) is a functional equation that determines the functional forms of the random entropies. A similar pseudo-additive relation is satisfied by the Arimoto entropy which, unlike (2), is a first-order homogeneous form. Hence, no conclusions concerning the extensive nature of systems may be drawn from either the Tsallis or Arimoto entropies based on the functional equation of the form (1).

Further, it is well-known that the maximum entropy formalism<sup>[2]</sup>, the minimum discrimination information<sup>[3]</sup> and

Gauss' principle<sup>[4,5]</sup>, all lead to the same results when a certain condition on the prior probability distribution is imposed<sup>[6]</sup>. All these methods lead to the same form of the posterior probability distribution; namely, the exponential family of distributions. Tsallis *et al.*<sup>[7]</sup> have attempted to adapt the maximum entropy formalism that uses the

Shannon entropy,  $S_1 = -\sum_{i=1}^m p_i \log p_i$ , to the pseudo-

additive entropy (2). In order to come out with analytic expressions for the probabilities that maximise the nonadditive entropy, they found it necessary to use escort probabilities<sup>[8]</sup> of the same power as the nonadditive entropy. If the procedure they use is correct then Gauss' principle should yield the same optimum probabilities. However, it will be shown that the Tsallis result requires the prior probability distribution be given by the same unphysical condition as the maximum entropy formalism and that the potential of the error law be required to vanish. The potential of the error law is what information theory refers to as the error<sup>[9]</sup>; that is, the difference between the inaccuracy and the entropy. Unless the 'true' probability distribution,  $P = \{p(x_1), p(x_2), \dots, p(x_m)\}$  coincides with the estimated probability distribution,  $Q = \{q(x_1), q(x_2), \dots, q(x_m)\}$ , the error does not vanish. Moreover, the two averaging procedures, one using the escort probabilities explicitly, will be shown to

be inequivalent and the relation between the potential of the error law and the pseudo-additive entropy to require the latter to vanish when the former vanishes.

**Condition of pseudo-additivity as a definition of non-extensivity:** Supposedly<sup>[1]</sup>, A and B are two independent systems in the sense that the probability of (A + B) factorises into those of A and B (that is,  $p_{ij}(A + B) = p_i(A)p_j(B)$ ). It follows immediately that, since  $S_\alpha \geq 0$  (non-negativity property) in all cases,  $\alpha < 1$ ,  $\alpha = 1$ ,  $\alpha > 1$ , respectively correspond to superadditivity (superextensivity), additivity (extensivity) and subadditivity (subextensivity). It does seem odd that criteria of super- and sub-additivity may be obtained via a functional equation rather than as a geometric inequality as are the criteria for convexity and concavity. The subadditive property is the triangle inequality<sup>[10]</sup>, whereas the geometrical interpretation of a concave function is one that never rises above its tangent plane at any point. There are classes of functions which are defined by inequalities that are weaker than convexity (concavity) and stronger than superadditivity (subadditivity)<sup>[11]</sup>.

To prove that (2) is always subadditive, consider the sum  $\sum_{i=1}^m p_i^\alpha$ . Companions to Minkowski's inequalities are<sup>[12]</sup>

$$\sum_{i=1}^m (p_i + q_i)^\alpha > \sum_{i=1}^m p_i^\alpha + \sum_{i=1}^m q_i^\alpha \tag{3}$$

for  $\alpha > 1$  and

$$\sum_{i=1}^m (p_i + q_i)^\alpha < \sum_{i=1}^m p_i^\alpha + \sum_{i=1}^m q_i^\alpha \tag{4}$$

for  $0 < \alpha < 1$ , where,  $Q = (q_1, \dots, q_m)$  is another complete distribution. Inequality (3) is the condition for superadditivity, the negative of which is subadditive. Hence, for  $\alpha > 1$  the Tsallis entropy is subadditive. Inequality (4) is the criterion for subadditivity and hence, the Tsallis entropy is subadditive for  $0 < \alpha < 1$ . This contradicts the conclusion, stated above, that the entropy (2) is superadditive for  $0 < \alpha < 1$  on the basis of (1).

Actually, the functional equation (1) is incorrect and should be replaced by<sup>[12]</sup>:

$$S_\alpha(PQ) = S_\alpha(P) + S_\alpha(Q) + (1 - \alpha)S_\alpha(P)S_\alpha(Q) \tag{5}$$

given in terms of two sets of complete probability distributions P and Q. Expression (5) has nothing to say about the extensivity of the entropy or the lack thereof; rather, it is a functional equation which determines the

form of the entropy. The pseudo-additive entropies are averages of the generators of power means. Power means are the only quasi-arithmetic means with a certain homogeneity property that will now be used to derive these random entropies.

Power means are first order homogeneous<sup>[12]</sup>

$$M_\phi(\lambda x) = \lambda M_\phi(x), \tag{6}$$

where, x stands for a set of discrete variables,  $x_1, \dots, x_m$ . Since  $\phi$  is defined up to a constant, set

$$\phi(1) = 0. \tag{7}$$

Now (6) stands for

$$M_\phi(x) = \lambda^{-1} M_\phi(\lambda x) = \lambda^{-1} \phi^{-1} \left( \sum_{i=1}^m p_i \phi(\lambda x_i) \right) = M_\psi(x). \tag{8}$$

On account of the translational invariance of the weighted mean

$$\psi(x) = a(\lambda)\phi(x) + b(\lambda), \tag{9}$$

where,  $a \neq 0$  and b are functions of the parameter  $\lambda$ . Since it is apparent from (8) that  $\psi(x) = \phi(\lambda x)$ , then

$$\phi(\lambda x) = a(\lambda)\phi(x) + b(\lambda). \tag{10}$$

However, if (7) is to be satisfied,  $b(\lambda) = \phi(\lambda)$ . Then, letting  $\lambda$  become another positive variable, y, (9) becomes the functional equation

$$\phi(xy) = a(y)\phi(x) + \phi(y). \tag{11}$$

Also,

$$\phi(xy) = a(x)\phi(y) + \phi(x). \tag{12}$$

Subtracting (12) from (11) leads to

$$\frac{a(x) - 1}{\phi(x)} = \frac{a(y) - 1}{\phi(y)} = c \tag{13}$$

where, c is a constant which both expressions must equal since each depends on a different independent variable. Solving (13) for a(y) and substituting back into (11) gives the functional equation

$$\phi(xy) = \phi(x) + \phi(y) + c\phi(x)\phi(y), \tag{14}$$

which is known in information science as the additivity property of degree- $\alpha$ .

If  $c = 0$ , this reduces to the classical equation

$$\phi(xy) = \phi(x) + \phi(y)$$

whose most general solution is  $\phi = A \log x$ .

If  $x$  represents the probability  $p_i$ , then the best code length for an input symbol of probability  $p_i$  is

$$\phi(p_i) = -\log p_i = n_i, \tag{15}$$

where,  $A = -1$  and  $n_i$  is the length of sequence  $i$ . Then, the

average length  $\sum_{i=1}^m p_i n_i$  corresponds to the average entropy<sup>[13]</sup>

$$S_1 = -\sum_{i=1}^m p_i \log p_i, \tag{16}$$

which is an additive entropy of degree-1, namely, the Shannon entropy.

With  $c \neq 1$ , insert  $c\phi(x) + 1 = g(x)$  into (14) to give

$$g(xy) = g(x)g(y). \tag{17}$$

Since the general solution of the functional Eq. 17 is  $g(x) = x^r$ , the random entropy function is

$$\phi(p_i) = \frac{p_i^r - 1}{c}. \tag{18}$$

Since (18) has to reduce to (15) in the limit as  $-c \rightarrow +0$ , so that  $r = -c$ . Hence (18) becomes

$$S_\alpha = \sum_{i=1}^m p_i \phi(p_i) = -\frac{\sum p_i^\alpha - 1}{\alpha - 1}, \tag{19}$$

where,  $c = 1 - \alpha$ . In the limit  $\alpha \rightarrow 1$ , it reduces to (16). The average entropy (19) is the pseudo-additive entropy of degree- $\alpha$ <sup>[14-21]</sup> (2), which reduces to the additive entropy (16) in the limit as  $\alpha \rightarrow 1$ .

There is nothing unique about (19). If our only concern is that it reduces to (15) in the limit  $\alpha \rightarrow 1$ , then an equally likely candidate is:

$$\phi(p_i) = \frac{1 - p_i^{1-\alpha}}{1 - \alpha}. \tag{20}$$

If (20) is averaged with respect to  $Q$ , then  $P$  must be chosen in terms of  $Q$  such that the new  $P$  are normalised. In terms of a variational problem, it is necessary that

$$\sum_{i=1}^m q_i \frac{1 - p_i^{1-\alpha}}{1 - \alpha} + \mu \sum_{i=1}^m p_i = \min.,$$

where, the constraint has been introduced using the Lagrange multiplier  $\mu$ . Performing the variation leads to  $p_i = (q_i / \mu)^{1/\alpha}$  and so that  $P$  be normalised

$$p_i = \frac{q_i^{\frac{1}{\alpha}}}{\sum_{i=1}^m q_i^{\frac{1}{\alpha}}}.$$

Introducing these so-called ‘escort’ probabilities<sup>[8]</sup> into (20) and averaging gives

$$S_\beta(Q) = \frac{\beta}{\beta - 1} \left\{ 1 - \left( \sum_{i=1}^m q_i^\beta \right)^{\frac{1}{\beta}} \right\}, \tag{21}$$

where,  $\alpha = 1/\beta$ . The first order homogeneous entropy (21) appears to have been introduced first by Arimoto<sup>[22]</sup> and subsequently given a complete characterisation by Boekee and van der Lubbe<sup>[23]</sup>.

Just as the Tsallis entropy (2) obeys the pseudo-additivity relation (5), the Arimoto entropy (21) satisfies

$$S_\beta(PQ) = S_\beta(P) + S_\beta(Q) - \left( \frac{\beta - 1}{\beta} \right) S_\beta(P) S_\beta(Q). \tag{22}$$

Adopting Tsallis’ line of reasoning, it would follow from (22) that (21) is subadditive for  $\beta > 1$  and superadditive for  $\beta < 1$ . However, according to Minkowski’s inequalities<sup>[12]</sup>

$$\left( \sum_{i=1}^m (p_i + q_i)^\beta \right)^{\frac{1}{\beta}} < \left( \sum_{i=1}^m p_i^\beta \right)^{\frac{1}{\beta}} + \left( \sum_{i=1}^m q_i^\beta \right)^{\frac{1}{\beta}} \tag{23}$$

for  $\beta > 1$  and

$$\left( \sum_{i=1}^m (p_i + q_i)^\beta \right)^{\frac{1}{\beta}} > \left( \sum_{i=1}^m p_i^\beta \right)^{\frac{1}{\beta}} + \left( \sum_{i=1}^m q_i^\beta \right)^{\frac{1}{\beta}} \tag{24}$$

for  $\beta < 1$ . Thus,  $\left( \sum_{i=1}^m p_i^\beta \right)^{\frac{1}{\beta}}$  is subadditive for  $\beta > 1$  and superadditive for  $\beta < 1$ . Since the negative of the former and positive of the latter are used to construct the Arimoto entropy, it may be concluded that it is always superadditive.

Moreover, concavity follows by setting  $p_i = \lambda a_i$  and  $q_i = (1 - \lambda) b_i$ , where the sets of numbers  $a_i$  and  $b_i$  are non-negative and  $0 \leq \lambda \leq 1$ . Then inequality (23) becomes

the criterion for convexity of  $\left( \sum_{i=1}^m p_i^\beta \right)^{\frac{1}{\beta}}$ ,

$$\left( \sum_{i=1}^m (\lambda a_i + (1 - \lambda) b_i)^\beta \right)^{\frac{1}{\beta}} < \lambda \left( \sum_{i=1}^m a_i^\beta \right)^{\frac{1}{\beta}} + (1 - \lambda) \left( \sum_{i=1}^m b_i^\beta \right)^{\frac{1}{\beta}}$$

for  $\beta > 1$  and

$$\left( \sum_{i=1}^m (\lambda a_i + (1-\lambda)b_i)^\beta \right)^{\frac{1}{\beta}} > \lambda \left( \sum_{i=1}^m a_i^\beta \right)^{\frac{1}{\beta}} + (1-\lambda) \left( \sum_{i=1}^m b_i^\beta \right)^{\frac{1}{\beta}}$$

for  $\beta < 1$  is the condition for its concavity. Since the Arimoto entropy takes the negative of the former and positive of the latter, it is concave everywhere<sup>[23]</sup>.

Similarly, the concavity of the Tsallis entropy (2) may be established from the observations that  $\sum_{i=1}^m p_i^\alpha$  is convex for  $\alpha > 1$  and concave for  $0 < \alpha < 1$ . Since the Tsallis entropy is formed from the negative of the former and positive of the latter, it is concave for all  $\alpha > 0$ .

Furthermore, since the entropy is defined to within a constant classically, the Arimoto entropy (21) is a first order homogeneous function and yet it obeys the pseudo-additive relation (22), just as the Tsallis entropy (2) obeys the pseudo-additive relation (5). Hence, it may be concluded that this relation has nothing whatsoever to do with the extensivity properties, or lack thereof, of the pseudo-additive entropies.

**Use of the maximum entropy formalism:** Now let  $X$  be a random variable whose values  $x_1, x_2, \dots, x_m$  are obtained at  $m$  independent trials. Prior to the observations, the distribution is  $Q$  and after the observations, the unknown probability distribution is  $P$ . The observer has at his disposal the statistic

$$\hat{a} = \frac{1}{m} \sum_{i=1}^m x_i$$

to help him formulate a guess as to the form of  $Q$ . Gauss' principle assumes that the probability distribution  $P$  depends on a parameter  $a$

$$a = E(X) = \sum_{i=1}^m x_i p(x_i), \tag{25}$$

such that the arithmetic mean,  $\hat{a}$ , is the maximum likelihood estimate of  $a$ . Furthermore,  $P$  will depend on the parameter  $a$  so that there is value  $a^0$  for which  $p(x_i; a^0) = q(x_i)$ , the prior distribution.

The maximum likelihood estimate,

$$\frac{\partial}{\partial a} \log L(a) = 0,$$

will lead to the exponential family of distributions when the log-likelihood function

$$\log L(a) = \sum_{i=1}^m \log p(x_i; a).$$

The likelihood equation

$$\frac{\partial}{\partial a} \psi(x_i; a) = 0,$$

where  $\psi(x_i; a) = \log p(x_i; a)$  is the same as requiring

$$\sum_{i=1}^m (x_i - a) = 0,$$

and any deviations in one will immediately lead to deviations in the other. Hence, they must be proportional to one another. Choosing the coefficient of proportionality as the second derivative of some appropriate scalar function  $V$  gives<sup>[6]</sup>

$$\frac{\partial}{\partial a} \psi(x_i; a) = V''(a)(x_i - a), \tag{26}$$

where, the prime denotes differentiation with respect to the argument. The scalar potential,  $V(a)$ , must be independent of the  $x_i$  because the left-hand side is a function of  $x_i$  alone and a similar equation for  $x_j$  would lead to a contradiction. The potential may be assumed such that  $V(a^0) = 0$  and consequently, (26) may be written as:

$$\frac{\partial}{\partial a} \psi(x_i; a) = \frac{\partial}{\partial a} \{V'(a)(x_i - a) + V(a)\}.$$

Integrating from  $a^0$  to  $a$  gives

$$\psi(x_i; a) = \psi(x_i; a^0) + \lambda(a)(x_i - a) + V(a),$$

where,  $\lambda(a) = V'(a)$ .

Usually, the log-likelihood function is logarithmic and the exponential family of distributions

$$\log p(x_i; a) = \log q(x_i) + \lambda(a)(x_i - a) + V(a) \tag{27}$$

results. Averaging both sides with respect to the probability distribution  $P$  gives

$$\sum_{i=1}^m \{p(x_i; a) \log p(x_i; a) - p(x_i; a) \log q(x_i)\} = V(a).$$

In information theory, the first term is the negative of the Shannon entropy, the second term is the inaccuracy and the right-hand side is the error<sup>[9]</sup>. On the strength of Shannon's inequality

$$\sum_{i=1}^m p(x_i; a) \log \left( \frac{p(x_i; a)}{q(x_i)} \right) = V(a) \geq 0, \quad (28)$$

the inaccuracy cannot be smaller than the Shannon entropy. Shannon's inequality follows from the arithmetic-geometric mean inequality,

$$\prod_{i=1}^m x_i^{p(x_i; a)} \leq \sum_{i=1}^m x_i p(x_i; a)$$

with  $x_i = q(x_i)/p(x_i; a)$

When Q is the uniform distribution, that is,  $q(x_i) = 1/m$  for all  $i$ , Shannon's inequality, (28), becomes

$$S_0(1/m) - S_1(P) = V(a)$$

which has been referred to as the entropy reduction caused by the application of a constraint that produces a finite value of  $a$ <sup>[24]</sup>.  $S_0(1/m) = \log m$  is the maximum entropy and is known as Hartley entropy in information theory. Classically, the entropy is defined to within a constant; only entropy differences are measurable.

However, none of what has been said so far applies to equilibrium thermodynamics<sup>[5]</sup>. If (27) is averaged with respect to the Q distribution instead of the P distribution and Shannon's inequality,

$$\sum_{i=1}^m q(x_i) \log [q(x_i)/p(x_i; a)] \geq 0,$$

is used, a problem is encountered immediately because  $V(a)$  must be negative now. In statistical mechanics,  $q(x_i)$  represents the surface of constant energy of a hypersphere of high dimensionality<sup>[25]</sup>. Because of its high dimensionality, the volume of the hypersphere lies very close to its surface so that  $q(x_i)$  may be thought of as the volume of phase space occupied by the system. Averages are performed with respect to this non-normalisable prior probability distribution<sup>[25]</sup>. In order to keep the error  $V(a)$ , which will be identified as the thermodynamic entropy, positive, it is necessary to introduce a sign change in (27). This sign change may be rationalised as follows: the exponential factor,  $e^{\lambda(a)x_i}$ , will not overpower the rapidly increasing factor of the density of states,  $q(x_i)$ . However, the density of states cannot increase faster than a certain power of the radius,  $x_i$ <sup>[25]</sup>, of the phase space volume, which is proportional to  $x_i^m$  in a hypersphere of  $m$ -dimensions. What is required is an even more rapidly decreasing exponential factor  $e^{-\lambda x_i}$ .

According to the Boltzmann-Planck interpretation,  $q(x_i)$  is not a normalised probability but, rather, a

'thermodynamic' probability, being proportional to the volume of phase space occupied by the system. The (random) entropy  $S(x_i)$  is defined as the logarithm of the thermodynamic probability

$$S(x_i) = \log q(x_i).$$

The phase average is given by:

$$E[X] = \sum_{i=1}^m x_i q(x_i) / \sum_{i=1}^m q(x_i).$$

The thermodynamic entropy is the phase space average of  $\log [q(x_i)/p(x_i; a)]$ , that is:

$$S(a) = \frac{\sum_{i=1}^m q(x_i) \log [q(x_i)/p(x_i; a)]}{\sum_{i=1}^m q(x_i)}$$

and its Legendre transform

$$S(a) - \lambda(a)a = \log Z(\lambda),$$

defines the logarithm of the generating function,  $Z(\lambda)$ . The inaccuracy now appears as the difference between the thermodynamic entropy and the average of the random entropies

$$-\sum_{i=1}^m q(x_i) \log p(x_i; a) = S(a) - \sum_{i=1}^m q(x_i) S(x_i) / \sum_{i=1}^m q(x_i) \geq 0. \quad (29)$$

The inequality follows from the fact that  $S$  increases in the wide sense and is concave. The expectation  $a$  may be taken with respect to either P or Q. The two averages must coincide, otherwise there would not be a single general thermodynamics, but, rather a "microcanonical thermodynamics" and a separate "canonical thermodynamics"<sup>[26]</sup>. Taken with respect to Q, (29) is Jensen's inequality for a concave function, where the Q has positive components which are arbitrary otherwise. Taken with respect to the normalised P, (29) is the

Jensen-Petrovic inequality<sup>[27]</sup>, where  $\sum_{i=1}^m p_i(x_i) \geq x_j$  for each  $j = 1, \dots, m$ .

The average of  $m$  variables is likely to be considerably greater than any of its components. Then, if  $S$  is increasing,

$$S \left( \sum_{i=1}^m x_i p(x_i) \right) \geq S(x_j)$$

for  $j = 1, \dots, m$ . Multiplying by  $q(x_i)$  and summing give back (29). This does not mean that  $S(x_i)/x_i$  should not decrease: A sufficient condition for

$$S\left(\sum_{i=1}^m x_i\right) \leq \sum_{i=1}^m S(x_i) \text{ is that } S(x_i)/x_i \text{ should decrease.}$$

The criteria for inequality attenuation<sup>[28]</sup> are that  $S(x_i)$  be an increasing function and  $S(x_i)/x_i$  decrease, that is, it is anti-star shaped. Fluctuations give rise to inaccuracy (29) and in their absence, a function of the average equals an average of the function. Therefore, if the exponential probability distribution, (27), is to coincide with Gauss' error law, written in terms of the concavity of the entropy,

$$\log p(x_i; a) = S(x_i) - S'(a)(x_i - a) - S(a) = \frac{1}{2} S''(\tilde{a})(x_i - a)^2, \tag{30}$$

where,  $\tilde{a}$  lies between  $x_i$  and  $a$ , then sign changes are required. When this is done, (27) becomes

$$\log p(x_i; a) = \log q(x_i) - \lambda(a)(x_i - a) - V(a). \tag{31}$$

A comparison of these latter two equations shows that the entropy,  $S(a)$ , is the potential,  $V(a)$ , that determines the error law<sup>[5]</sup>. The concavity of the entropy ensures that the exponent will be negative and hence  $p(x_i; a)$  will be less than unity. The parameter,  $\lambda(a)$ , is still the derivative of the scalar potential,  $V(a)$ , but since this potential now coincides with the thermodynamic entropy,  $S(a)$ , the Lagrange multiplier  $\lambda(a)$  is identified as the intensive variable divided by the absolute temperature in the entropy representation.

Information theoretic entropies and the entropy reduction of the thermodynamics of extremes<sup>[24]</sup> are not amenable to the previous thermodynamic interpretation, where the entropy is defined as the logarithm of the volume of phase space occupied by the system. Since all the volume lies very near the surface in a thermodynamic system of high dimensionality, the volume of phase space will coincide with the surface area, which is referred to as the structure function<sup>[25]</sup>.

In what was eventually a futile attempt to justify Tsallis' formalism, Plastino and Plastino<sup>[29]</sup> considered a structure function of the form  $E_i^{m-1}$  for the energy. Assuming a bounded phase space-for no stated reason-whose total energy is  $E_0$ , the Tsallis exponent was identified as  $\alpha = (m-2)/(m-1)$  and the inverse temperature was defined as  $\beta = (m-1)/E_0$ . What was not realised is that, in order to define a temperature,  $m$  must be much greater than 1 so that  $\alpha \approx 1$ . More precisely,  $m$  must be large enough to allow the use of Stirling's formula<sup>[5]</sup>. Considering the claimed conditions under which Tsallis'

statistical mechanics is supposed to apply, it follows that it cannot be applied to thermodynamic systems for such systems would be too small to be capable of defining intensive quantities, such as temperature and pressure.

Consider the P and Q as two sets of complete probability distributions. For a given probability distribution, Q, seek the set of probability P which resembles Q most closely. This is the minimum discrimination statistic of Kullback<sup>[3]</sup>.

In order to derive the pseudo-additive entropies, the logarithm is replaced by the limit

$$\log p(x_i; a) \rightarrow \frac{p^{\alpha-1}(x_i; a) - 1}{\alpha - 1} = \psi(x_i; a),$$

and a similar relation for  $\log q(x_i)$ , in the exponential law to give

$$\frac{p^{\alpha-1}(x_i; a) - q^{\alpha-1}(x_i)}{\alpha - 1} = \lambda(a)(x_i - a) + V(a). \tag{32}$$

Multiplying (32) by  $p(x_i; a)$  and summing leads to<sup>[30]</sup>

$$I_\alpha(Q) - S_\alpha(P) = \frac{\sum_{i=1}^m p(x_i; a) (p^{\alpha-1}(x_i; a) - q^{\alpha-1}(x_i))}{\alpha - 1} = V(a) \geq 0 \tag{33}$$

where:

$$I_\alpha(Q) = \frac{1 - \sum_{i=1}^m p(x_i; a) q^{\alpha-1}(x_i)}{\alpha - 1}$$

has been referred to as the inaccuracy<sup>[11]</sup>.

From now on the dependence of the probability distribution P on the average (25) will be suppressed because Tsallis statistical thermodynamics makes no pretext at statistical inference and the inaccuracy is a convex function of Q for given P provided  $\alpha \leq 2$ . The inaccuracy is defined as the sum of the entropy (2) and the error  $V(a)$ . The inaccuracy has the property that

$$\lim_{q \rightarrow 1/m} I_\alpha(Q) = S_\alpha(1/m).$$

the negative of the error is what has been termed the entropy reduction<sup>[24]</sup>.

The inequality in (33) follows from Hölder's inequality. Consider the case when all the P are rational; then they may be expressed in the form

$$p(x_i) = x_i / \sum_{i=1}^m x_i \text{ and } Q \text{ is the uniform distribution,}$$

$q(x_i) = 1/m$  for all  $i$ . (32) then becomes

$$\frac{1}{(\alpha - 1)} \left( \frac{\sum_{i=1}^m x_i^\alpha}{\left(\sum_{i=1}^m x_i\right)^\alpha} - m^{1-\alpha} \right) = V(a)$$

because of Hölder's inequalities<sup>[12]</sup>

$$m^{-1} \sum_{i=1}^m x_i > \left( \sum_{i=1}^m x_i^\alpha \right)^{1/\alpha} m^{-1/\alpha} \text{ for } \alpha < 1$$

with the inequality sign being reversed for  $\alpha > 1$  so that inequality (33) always holds.

Tsallis *et al.*<sup>[7]</sup> found that the maximisation procedure for the pseudo-additive entropy (32) with respect to the constraint

$$a_\alpha = E_\alpha(X) = \sum_{i=1}^m x_i p^\alpha(x_i) / \sum_{i=1}^m p^\alpha(x_i), \quad (34)$$

using the escort probabilities<sup>[8]</sup>, yields the stationary condition<sup>[7]</sup>

$$p(x_i) = \frac{[1 - (1 - \alpha)\lambda_\alpha(x_i - a_\alpha)]^{1/(1-\alpha)}}{Z(\lambda_\alpha)} \quad (35)$$

where:

$$\lambda_\alpha = \lambda / \sum_{i=1}^m p^\alpha(x_i) \quad (36)$$

and  $\lambda$  is the Lagrange multiplier for constraint (34). The normalisation condition for the  $p(x_i)$  gives the partition function as

$$Z(\lambda_\alpha) = \sum_{i=1}^m [1 - (1 - \alpha)\lambda_\alpha(x_i - a_\alpha)]^{1/(1-\alpha)}. \quad (37)$$

At best, (35) may be considered an implicit relation for the probabilities since (36) contains the probabilities explicitly period. In order to reduce Gauss' principle (32) to something looking even vaguely like the 'optimal' probabilities (35) that maximise the Tsallis entropy in, (33), it is necessary to

- (i) assume that P is an incomplete distribution,
- (ii) set  $q(x_i) = 1$  for all i
- and (iii) set  $V(a) = 0$ .

Then

$$\frac{p(x_i)}{\sum_{i=1}^m p(x_i)} = \frac{[1 - (1 - \alpha)\lambda_\alpha(x_i - a_\alpha)]^{1/(1-\alpha)}}{Z(\lambda_\alpha)},$$

where, the partition function is given by (37) and the escort probabilities (34) were used to define the parameter a, rather than the weighted average (25).

However, if (31) is taken and the approximation

$$\frac{p^{1-\alpha}(x_i) - 1}{1 - \alpha} \rightarrow \log p(x_i)$$

introduced, with a similar expression for  $q(x_i)$ ,

$$\frac{p^{1-\alpha}(x_i) - q^{1-\alpha}(x_i)}{1 - \alpha} = -\lambda_\alpha(x_i - a_\alpha) - V(a) \quad (38)$$

results. Setting  $q(x_i) = 1$  and  $V(a) = 0$  and requiring the probability distribution P be normalised result in (35) or, equivalently

$$p^{1-\alpha}(x_i) = \frac{1 - (1 - \alpha)\lambda_\alpha(x_i - a_\alpha)}{Z^{1-\alpha}(\lambda_\alpha)}.$$

Multiplying both sides by  $p^\alpha(x_i)$  and summing lead to

$$Z^{1-\alpha}(\lambda_\alpha) = \sum_{i=1}^m p^\alpha(x_i),$$

provided  $\lambda_\alpha$  is given by (36). On the other hand, if both sides of (35) are raised to the power  $\alpha$ , summed from 1 to m and rearranged

$$Z^\alpha(\lambda_\alpha) \sum_{i=1}^m p^\alpha(x_i) = \sum_{i=1}^m [1 - (1 - \alpha)\lambda_\alpha(x_i - a_\alpha)]^{\alpha/(1-\alpha)} \neq Z(\lambda_\alpha). \quad (39)$$

The only difference between the two forms of averaging is that, in the first case, use was made of the escort probability average, (34). Since the two results do not coincide, it must be concluded that there is something amiss with the escort probability average (34) since this is the one totally new aspect introduced into the considerations.

Moreover, if the  $\alpha \rightarrow 1$  limit is taken in (39)

$$Z(\lambda_1) = \sum_{i=1}^m e^{-\lambda_1(x_i - a)}$$

results and this is not the correct expression for the partition function even in the unphysical case of a density of states equal to unity.

Finally, multiplying both sides of (38) by  $p^\alpha(x_i)$  and summing gives:

$$\frac{1 - \sum_{i=1}^m q(x_i) [p(x_i)/q(x_i)]^\alpha}{1 - \alpha} = - \sum_{i=1}^m p^\alpha(x_i) V(a).$$

If  $q(x_i) = 1$  for all i then

$$S_\alpha(P) = \sum_{i=1}^m p^\alpha(x_i) V(a). \quad (40)$$

This illustrates the correspondence between the Shannon entropy and the potential  $V(a)$  in the  $\alpha \rightarrow 1$  limit that was alluded to above in the thermodynamic formulation which takes into account a non-normalised



prior probability distribution. However, the prior probability distribution has been set equal to unity, as in the maximum entropy method. Also, in order to derive the probability distribution (35) from Gauss' law,  $V$  had to be assumed identically equal to zero. Consequently, (40) would require the vanishing of the pseudo-additive entropy (2).

From all that has gone before, it must be concluded that Tsallis' statistical thermodynamic formulation of his entropy (2) is neither self consistent nor correct.

### REFERENCES

1. Tsallis, C., 1999. Nonextensive statistics: Theoretical, experimental and computational evidences and connections. *Brazilian J. Phys.*, 29: 1.
2. Jaynes, E.T., 1957. Information theory and statistical mechanics. *Phys. Rev.*, 106: 620; *Information Theory and Statistical Mechanics II*, *Phys. Rev.*, 108: 171.
3. Kullback, S., 1959. *Information Theory and Statistics*, Wiley, New York.
4. Keynes, J.M., 1921. *Treatise on Probability*. St. Martin's Press, New York.
5. Lavenda, B.H., 1991. *Statistical Physics: A Probabilistic Approach*. Wiley-Interscience, New York.
6. Campbell, L.L., 1970. Equivalence of Gauss' principle and minimum discrimination information estimation of probabilities. *Ann. Math. Stat.*, 41: 1011.
7. Tsallis, C., R.S. Mendes and A.R. Plastino, 1998. The role of constraints within generalized nonextensive statistics. *Physica, A* 261: 534.
8. Beck, C. and F. Schlögl, 1993. *Thermodynamics of Chaotic Systems*. Cambridge U. P., Cambridge.
9. Kerridge, D.F., 1961. Inaccuracy and inference. *J. Roy. Stat. Soc. Ser B.*, 23: 184.
10. Beckbach, E.F. and R. Bellman, 1961. *Inequalities*. Springer, Berlin.
11. Bruckner, A.M. and E. Ostrow, 1962. Some function classes related to the class of convex functions. *Pacific J. Math.*, 12: 1203.
12. Hardy, G., J.E. Littlewood and G. Pólya, 1952. *Inequalities*. 2nd Edn. Cambridge U. P., Cambridge, pp: 31, 68.
13. Campbell, L.L., 1966. Definition of entropy by means of a coding problem. *Z. Wahrscheinlichkeitstheorie verw. Geb.*, 6: 113.
14. Havrda, J. and F. Charvát, 1967. Quantification method of classification processes. Concept of structural  $\alpha$ -entropy. *Kybernetika*, 3: 30
15. Daróczy, Z., 1970. Generalized information function. *Information and Control*, 16: 36.
16. Vajda, I., 1968. Axiomy  $\alpha$ -entropie zobecneneho pravdivopodnostniho schématu, *Kybernetika*, 4: 105.
17. Aggarwal, N.L., Y. Cesari and C.F. Picard, 1972. Propriétés de branchement liées aux questionnaires de Campbell et à l'information de Rényi. *C.R. Acad. Sci. Paris Série A*, 275: 437.
18. Forte, B. and C.T. Ng, 1973. On characterisation of the entropies of degree  $\beta$ . *Utilitas Mathematica*, 4: 193.
19. Aczél, J. and Z. Daróczy, 1975. *On Measures of Information and Their Characterization*. Academic Press, New York. Formula (6.3.3).
20. Mathai, A.M. and P.N. Rathie, 1975. *Basic Concepts in Information Theory and Statistics*. Wiley, New York. Formula (1.2.2).
21. Tsallis, C., 1988. Possible generalization of Boltzmann-Gibbs statistics. *J. Stat. Phys.*, 52: 479.
22. Arimoto, S., 1971. Information theoretical considerations on estimation problems. *Inform. Contr.*, 19: 181.
23. Boekee, D.E. and J.C.A. Van der Lubbe, 1980. The R-norm information measure. *Inform. Contr.*, 45: 136
24. Lavenda, B.H., 1995. *Thermodynamics of Extremes*. Horwood, Chichester.
25. Khinchin, A.I., 1949. *Mathematical Foundations of Statistical Mechanics*. Dover, New York.
26. Greene, R.F. and H.B. Callen, 1951. On the formalism of thermodynamic fluctuation theory. *Phys. Rev.*, 83: 1231.
27. Pecaric, J.E., F. Proschan and Y.L. Tong, 1992. *Convex Functions, Partial Orderings and Statistical Applications*. Academic Press, San Diego.
28. Arnold, B.C., 1987. *Majorization and the Lorenz Order*, Springer, Berlin.
29. Plastino, A. and A.R. Plastino, 1999. Tsallis entropy and Jaynes' information theory formalism. *Brazilian J. Phys.*, 29: 50.
30. Lavenda, B.H., 1998. Measures of information and error laws. *Intl. J. Theor. Phys.*, 37: 3119.