



# Journal of Applied Sciences

ISSN 1812-5654

**science**  
alert

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

## Are Superstatistics and Nonextensive Statistical Mechanics Viable?

<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, UK

---

**Abstract:** Like Jaynes' maximum entropy formulation, the Tsallis entropy is to be maximised 'with respect to all that is known about the system'. However, instead of customary averages, the averaging is performed with 'escort' probabilities resulting in Student's t-distribution, as opposed to exponential distributions in Jaynes' formalism. The same class of non-exponential distributions, possessing 'fat' tails, is said to be derived from so-called 'superstatistics'. Here it is shown that there is neither motivation nor need to introduce a 'super' statistics.

**Key words:** Superstatistics, non-extensive statistical mechanics, maximum entropy formalism, escort averaging, power law versus exponential distributions

---

### INTRODUCTION

In 1988, a study appeared<sup>[1]</sup> which supposedly introduced, with no recognisable preamble, a new entropy expression into the field of statistical thermodynamics. As has been shown on several occasions<sup>[2-4]</sup>, the said expression had been known to information scientists since the late 1960's. However, its introduction to statistical thermodynamics has spawned a plethora of articles which are aimed primarily at determining the exponent in the Tsallis entropy for individual systems. That would mean that each such system would have its own thermodynamics. As this is quite impossible, the entropy cannot depend upon a parameter that is different for different systems. While the apparent claim of novelty in the 1988 study is definitely questionable, some of the subsequent studies which make use of it do little other than provoke qualms over the validity of their content. Here attention will be restricted to just three such articles which, at first sight, might seem aimed at extending the field of statistical thermodynamics.

### BECK AND COHEN'S 'SUPERSTATISTICS'

Beck and Cohen<sup>[5]</sup> attempted to generalise the Boltzmann factor in order to obtain a more general statistics, which they termed 'superstatistics'. They did this by performing a Laplace transform on the probability density function (pdf) of an intensive variable  $f(\beta)$ , where  $\beta$  is the inverse temperature. The criteria used for choosing  $f(\beta)$  were:

(i) it must be normalised:

$$\int_0^{\infty} f(\beta) d\beta = 1. \quad (1)$$

(ii) it must be such that its Laplace transform:

$$B(E) = \int_0^{\infty} e^{-\beta E} f(\beta) d\beta$$

is normalised, or at least is normalised with respect to a density of states  $\rho(E)$ :

$$\int_0^{\infty} B(E) \rho(E) dE = \int_0^{\infty} d\beta f(\beta) \int_0^{\infty} dE e^{-\beta E} \rho(E) = 1.$$

(iii) the superstatistics should reduce to BG-statistics when there are no fluctuations in  $\beta$ .

In thermodynamics, there are two types of variable: extensive and intensive. In mathematical statistics they correspond to observable and estimable variables, respectively<sup>[6,7]</sup>. Estimable variables relate the pdf of the extensive variables to the properties of the physical system. They are referred to as the 'state of nature' in mathematical statistics. Estimable variables, like the inverse temperature, change due to the nature of the physical interaction and therefore, cannot be considered random variables in the frequency sense. Estimable variables cannot be treated in the limit-of-frequency sense; rather they must be interpreted in the subjective sense that some values of  $\beta$  are more 'probable' than others.

If the parameter's value is completely unknown insofar as it can take on any conceivable value from

0 to  $\infty$ , Jeffreys' rule for choosing the prior is to take its logarithm as uniform so that<sup>[8]</sup>:

$$f(\beta)d\beta \propto \frac{d\beta}{\beta} \text{ for } 0 < \beta < \infty. \quad (2)$$

Since the integral of the prior is infinite, it is improper. Jeffreys uses  $\infty$  to represent the probability of a certain event rather than 1. The indeterminacy for predicting whether the value will fall within the interval (0, b),

$\int_0^b d\beta/\beta = \infty$ , or  $(b, \infty)$ ,  $\int_b^\infty d\beta/\beta = \infty$ , is merely a formal representation of ignorance.

However, Beck and Cohen<sup>[5]</sup> do not imply that their  $f(\beta)$  is a prior pdf. Transforming an improper pdf into a proper pdf may be done by using Bayes' theorem<sup>[9]</sup>. The method uses a likelihood function which essentially inverts the roles of the observable and estimable variables. Due to the additivity of the observable variables, the likelihood function will be a product of the individual probabilities. The likelihood function takes us from an improper, uninformative, pdf to a proper, informative one. This is the meaning of  $f(\beta)$ .

Bayes' theorem may be written as<sup>[7]</sup>:

$$f(\beta; \bar{E}) = \frac{e^{L(\beta, \bar{E})}}{B(\bar{E})} f(\beta) \quad (3)$$

where,  $L(\beta, \bar{E})$  is the log-likelihood function which transforms the prior pdf  $f(\beta)$  into the posterior pdf  $f(\beta, \bar{E})$ , where,  $\bar{E}$  is the average energy of the sample. Since the posterior pdf is normalised, we obtain the explicit expression for  $B(\bar{E})$  as:

$$B(\bar{E}) = \int_0^\infty e^{-\beta \bar{E} - L(\beta)} f(\beta) d\beta = \int_0^\infty e^{-\beta \bar{E} - L(\beta)} \frac{d\beta}{\beta},$$

where,  $L(\beta)$  is the Legendre transform of the entropy with respect to the energy and Jeffreys' improper prior pdf has been introduced in the second expression.

Now to consider the examples treated by Beck and Cohen. According to them, the most relevant example of superstatistics is the gamma pdf:

$$f(\beta) = \frac{1}{b\Gamma(c)} \left(\frac{\beta}{b}\right)^{c-1} e^{-\beta/b}. \quad (4)$$

This Gamma pdf for the inverse temperature has been derived from Bayes' theorem (3) already by Lavenda<sup>[7]</sup>, Eq. (4.97), where,  $2c = 3N$  is the effective number of degrees of freedom and  $b = 1/E_0$  is a fixed parameter. Setting their  $f(\beta) = e^{L(\beta)}/\beta$ , it is seen that:

$$L(\beta) = \beta/b - c \log(\beta/b) + \log \Gamma(c).$$

Since  $L(\beta)$  is the Legendre transform of the entropy, it is also the logarithm of the partition function, so that:

$$\frac{\partial L}{\partial \beta} = \frac{1}{b} - \frac{c}{\beta} = -\bar{E}$$

which leads to the thermal equation of state:

$$\beta = \frac{\beta_0}{1 + b\bar{E}}, \quad (5)$$

where,  $\beta_0 = bc = 3N/2E_0$  in the notation adopted here. The entropy is the Legendre transform of  $L$  so that, if  $c$  is large enough for Stirling's approximation to hold:

$$S(\bar{E}) = c \log(1 + b\bar{E}). \quad (6)$$

This means that:

$$S(\bar{E}) = \beta \bar{E} + L(\beta) \quad (7)$$

where,  $\beta$  is given by (5) and

$$B(\bar{E}) = e^{-S(\bar{E})}.$$

It is imperative to emphasise that absolutely no reference to superstatistics has been made here! Fluctuations in intensive variables occur in BG-statistics. In fact, for  $b\bar{E} \gg 1$  the thermal equation of state (5) reduces to the thermal equation of state of an ideal gas with  $2c$  degrees of freedom,  $\beta = c/\bar{E}$  and (6) gives the energy dependency of the entropy of an ideal gas,  $S = c \log \bar{E}$  and Superstatistics supposedly enters when Beck and Cohen identify  $B$  with the Tsallis distribution where  $c = 1/(q-1)$  and  $q$  is the exponent in the Tsallis entropy. They write  $B = \exp\{-c \log(1 + b\bar{E})\}$  and expand it for small  $b\bar{E}$ . Yet their resulting expansion is not in powers of  $b\bar{E}$ ! The parameter  $b = \beta_0/c = \beta_0/(q-1)$ , which in the present notation is  $b = 1/E_0$ , the inverse of the total energy. The expansion is:

$$\begin{aligned} (1 + b\bar{E})^{-c} &= \sum_{k=c}^{\infty} \binom{k-1}{c-1} (-b\bar{E})^{k-c} \\ &= 1 - \beta_0 \bar{E} + \frac{(\beta_0 \bar{E})^2}{2!} - \frac{(\beta_0 \bar{E})^3}{3!} + \dots \\ &\dots + (q-1) \frac{(\beta_0 \bar{E})^2}{2!} - (2q+1)(q-1) \frac{(\beta_0 \bar{E})^3}{3!} + \dots \end{aligned}$$

and not their Eq. 14 in which  $e^{-b\bar{E}}$  may be factored out of the sum. This has fatal consequences for their conclusion, in their Eq. 25, that the Tsallis entropy exponent  $q = \beta^2/\beta_0$ , which already looked suspicious since it excluded values of  $q < 1$ .

The log-normal pdf for the inverse temperature was derived<sup>[7]</sup>, where there is a  $1/T_i$  missing. This term is due to the Jacobian of the transformation on going from 4.114 to 4.115. It was derived from an error law for the temperature for which the geometric mean is the most probable value of the temperature measured. The geometric mean value of the temperature is the lowest temperature attainable when two bodies at different initial temperatures are placed in thermal contact and where the processes of heat withdrawal and injection and conversion into work are all carried out reversibly. Again there is no mention, or need, of superstatistics.

Beck and Cohen<sup>[5]</sup> first example has been left until last because it illustrates two general principles. Firstly, Beck and Cohen<sup>[5]</sup> choose a uniform prior pdf in an interval  $a$  to  $a + b$ . According to Jeffreys' first rule, a uniform prior is appropriate when the parameter involved may conceivably assume all values from  $-\infty$  to  $+\infty$ . Since the temperature is non-negative,  $\beta$  may assume values from  $0$  to  $\infty$  and for such a parameter, Jeffrey<sup>[8]</sup> suggests taking its logarithm uniform, (2).

Secondly, the Legendre transform, as already mentioned, is equivalent to the Laplace integral here but this is not justified in this case because it is not a thermodynamic system with a large number of degrees of freedom. Consider the following Laplace-type integral:

$$B(E) = \int_a^{a+b} e^{-\beta E} \beta^m d\beta. \tag{8}$$

Under change of variable  $\beta = m\gamma$ , the integral becomes:

$$B(E) = m^{m+1} \int_{a'}^{a'+b'} e^{-m\phi(\gamma)} d\gamma,$$

where,  $a' = a/m$ ,  $b' = b/m$  and  $\phi(\gamma) = \gamma E - \log \gamma$ .

The function  $\phi$  has a minimum at  $\hat{\gamma} = E^{-1}$  or equivalently  $\hat{\beta} = m/E$ , which is the thermal equation of state for a system of  $2m$  degrees of freedom. Expanding  $\phi(\gamma)$  up to second order in the small difference  $(\gamma - \hat{\gamma})$  gives:

$$B(E) \approx m^{m+1} e^{-m-(m+1)\log E} \int_{-\sqrt{(m/2)(1-a/\hat{\beta})}}^{\sqrt{(m/2)((a+b)/\hat{\beta}-1)}} \frac{2}{m} e^{-\tau^2} d\tau,$$

where, a further change of variable,  $\tau = \sqrt{(m/2)}\gamma^{-1}(\gamma - \hat{\gamma})$ , has been introduced.

It might be noted that, no matter how small  $b$  is,  $m$  may be taken so large that the value of the integral is altered only slightly when the limits of integration are replaced by  $\pm\infty$ <sup>[10]</sup>. Thus, finally:

$$B(E) \approx \frac{m^m e^{-m} \sqrt{2\pi m}}{E^{m+1}}.$$

For large  $m$ , the numerator is Stirling's approximation to  $m!$  up to order  $1/m$ . Hence, the final expression for the entropy is:

$$S(E) = \log \left( \frac{E^m}{m!} \right).$$

Therefore, the Legendre transform may be used instead of evaluating the Laplace transform for values of  $m$  for which Stirling's approximation holds. Only in this case may the principles of statistical thermodynamics be used<sup>[7]</sup>. This does not apply to Beck and Cohen's uniform distribution. Finally, it might be noted that (8) is essentially the Gamma pdf (4) in the case of large  $b$  and so the normalisation condition (1) is irrelevant.

### THE NONEXTENSIVE STATISTICAL MECHANICS OF BECK

There is always a certain delight in seeing old ideas from a new perspective, but there is no delight in seeing professed 'new' ideas that are not new. The notion of 'superstatistics', as discussed above<sup>[5]</sup> grew out of an idea that if one began with a stationary Maxwell distribution and introduced a randomised operational temperature, then a new process could be derived which had something to do with the so-called Tsallis distribution<sup>[11]</sup>. This distribution had been obtained as a stationary condition for maximising a nonadditive entropy of degree  $q$  with respect to the energy constraint<sup>[12]</sup>. The generalisation to 'superstatistics' then consisted in replacing the exponential in the Maxwell distribution by  $e^{-\beta H}$ , where,  $H$  is any generic Hamiltonian and using an arbitrary probability distribution for the conjugate  $\beta$  one could obtain a new probability distribution for  $H$  by integrating the product of the two probability distributions over all values of  $\beta$ . The physical justification of this procedure has been questioned above already.

The actual idea of randomising temperature is not new as Beck would have us believe; it was discussed by Lavenda<sup>[13]</sup>. It is based on subordination, which is discussed by Feller<sup>[14]</sup>. If  $X(t)$  represents a brownian motion then, by randomising the operational time  $t$ , a variety of new processes may be derived. That is, if the displacements of brownian motion are governed by the pdf:

$$f_{t_0}(x) = \frac{1}{\sqrt{2\pi t_0}} e^{-x^2/2t_0},$$

where,  $x$  is the random variable and  $t_0$  the parameter, then exchanging their roles leads to a Lévy pdf:

$$f_{x_0}(t) = \frac{X_0}{\sqrt{2\pi t^3}} e^{-x_0^2/2t}, \tag{9}$$

for the randomised time,  $t$ , where,  $x_0$  is the parameter. The two pdf are related by the transform  $x^2/t_0 = x_0^2/t$ . The Cauchy process is then said to be subordinated to brownian motion<sup>[14]</sup>:

$$f_{x_0}(x) = \int_{-\infty}^{+\infty} f_t(x) f_{x_0}(t) dt = \frac{X_0}{2\pi} \int_{-\infty}^{\infty} \frac{1}{t^2} e^{-(x^2+x_0^2)/2t} dt = \frac{1}{\pi} \frac{X_0}{x_0^2 + x^2},$$

where, the two processes are evaluated at the same time.

The Maxwell speed,  $u(T)$ , in a single dimension and the temperature,  $T$ , stand in the same relation as the displacement of a brownian motion process,  $X(t)$  and the time,  $t$ <sup>[13]</sup>. The Cauchy distribution for the speed is, therefore, subordinated to the Maxwell distribution in one dimension<sup>[13]</sup>. Therefore, Beck<sup>[11]</sup> does not have to suppose that the fluctuations in the inverse temperature,  $\beta$ , follow a  $\chi^2$  distribution given by his Eq. 2:

$$f_{\beta_0}(\beta) = \frac{1}{\Gamma(n/2)} \left(\frac{n}{2\beta_0}\right)^{n/2} \beta^{n/2-1} e^{-n\beta/2\beta_0}, \tag{10}$$

which, on the strength of equipartition,  $u_0^2\beta_0 = n$ , could be written:

$$f_{u_0}(\beta) = \frac{1}{\Gamma(n/2)} \left(\frac{u_0^2}{2}\right)^{n/2} \beta^{n/2-1} e^{-\beta u_0^2/2}. \tag{11}$$

By what is shown to be an incorrect association of Tsallis' distribution with the subordinated process, Beck finds  $n = 2/(q-1)$ , where,  $q$  is the degree of the nonadditive entropy or what is commonly known as the Tsallis exponent. A single mode of (10) exists at  $\beta = (1-2/n)\beta_0$ . The single mode vanishes, therefore, for all  $q \geq 2$ . Hence, a qualitative change in the behaviour of the process might be expected for  $q < 2$  and  $q \geq 2$ . That no behavioural change has been predicted casts doubt on the association of half the number of degrees of freedom with the inverse of  $(q-1)$ .

It follows from the Lévy transform:

$$\beta u_0^2 = \beta_0 u^2 \tag{12}$$

that the  $n$ -dimensional Maxwell speed distribution,

$$f_{\beta_0}(u) = \frac{2}{\Gamma(n/2)} \left(\frac{\beta_0}{2}\right)^{n/2} u^{n-1} e^{-\beta_0 u^2/2}, \tag{13}$$

corresponds to the  $\chi^2$  distribution for inverse temperature

(11) and not Beck's one dimensional Maxwell pdf (Eq. 21). What has happened is that, in a space of  $n$  dimensions, the starting point was a Markov process,  $u(\beta)$ , whose stationary transition probability is given by (13). A host of new processes may be derived by randomising the temperature or its inverse,  $\beta > 0$ , corresponding to a new random variable  $B(\beta)$  whose pdf is (11). The process  $b(\beta)$  is called the 'directing' process<sup>[14]</sup>. The pdf of the new process  $u(b(\beta))$  will be given by:

$$f_{u_0}(u) = \int f_{\beta}(u) f_{u_0}(\beta) d\beta = \frac{2}{B(n/2, n/2)} \frac{(u/u_0)^{n-1}}{(1+(u/u_0)^2)^n} \frac{1}{u_0}, \tag{14}$$

where,  $\beta$  has been set equal to  $\beta_0$  to ensure that the two processes are in thermal equilibrium and  $B(\dots)$  is the beta function. Contrary to what has been claimed<sup>[9]</sup>, (14) is not Tsallis' canonical probability distribution Eq. 9<sup>[11]</sup>:

$$f_{\beta_0}(u) = \frac{\Gamma\left(\frac{n+1}{2}\right)}{\sqrt{n/2}\Gamma(n/2)} \sqrt{\frac{\beta_0}{2\pi}} \left(1 + \frac{\beta_0 u^2}{n}\right)^{-(n+1)/2}. \tag{15}$$

From symmetry consideration, all odd order moments of (15) are zero. In particular, the second moment is:

$$\overline{u^2} = \frac{n\beta_0^{-1}}{n-2}, \tag{16}$$

which does not reflect equipartition that follows from (13) (Eq. 25).

For a generic value of  $n$ , (14) is a special case of the inverse beta pdf:

$$f(x) = \frac{1}{B(n/2, n/2)} \frac{x^{n/2-1}}{(1+x)^n}, \tag{17}$$

where,  $x = (u/u_0)^2$ . In economics, (17) is known as the Pareto pdf since 'it was thought (rather naively from a modern statistical standpoint) that income distributions should have a tail with density  $\sim Ax^{-\alpha}$  as  $x \rightarrow \infty$  and (17) fulfils this requirement<sup>[14]</sup>.

In a single dimension, (14) is the half-Cauchy pdf for a positive variate. Just as the Cauchy process for the displacement is subordinated to brownian motion in one dimension when time is randomised, the Cauchy process for the kinetic energy is subordinated to the Maxwellian in a single dimension when the temperature, or its inverse, is randomised<sup>[13]</sup>.

In  $n = 3$  dimensions, (14) is<sup>[13]</sup>:

$$f_{n=3}(v) = \frac{8}{\pi} \frac{v^2}{(1+v^2)^3}. \tag{18}$$

This is the pdf of the length of a random velocity

vector  $v = \sqrt{v_x^2 + v_y^2 + v_z^2}$  in  $\mathfrak{R}^3$  dimensions. It is related to the density:

$$f_{n-1}(v) = \frac{1}{B(\frac{1}{2}, m - \frac{1}{2})} \frac{1}{(1 + v^2)^m}, \quad (19)$$

in a fixed direction by<sup>[14]</sup>:

$$f_{n-3}(v) = -v f'_{n-1}(v), \quad (20)$$

where, the prime denotes differentiation with respect to the argument. The pdf (19) has  $(2m-1)$  degrees of freedom. For  $m = 1$ , (19) reduces to the Cauchy pdf, while for  $m = 2$ , implying that the distribution has three degrees of freedom, (18) is obtained from (20). Hence, (18) is the density of the length  $v$  of a random velocity vector in  $n = 3$  dimensions<sup>[13]</sup>.

Although (10) is valid for any  $n > 0$ , Beck's 'conditional' pdf for the speed is Eq. 13, corresponding to his Eq. 6, with  $n = 1$ . For any  $n \neq 1$ , there is an incompatibility in the dimensions of the two distributions. This is responsible for the seemingly close appearance of the pdf of the subordinated process with the Tsallis distribution. In other words, if Beck's Eq. 6 is meant to be the speed distribution, it does not reflect the dimensionality of the  $\chi^2$  pdf for  $\beta$ , (10). Rather Beck finds that the subordinated process is governed by Student's t-distribution, (15). The first factor tends to unity as  $n \rightarrow \infty$  and for every fixed  $u$ <sup>[15]</sup>:

$$-\frac{n+1}{2} \log \left( 1 + \frac{\beta_0 u^2}{n} \right) \rightarrow -\frac{1}{2} \beta_0 u^2$$

so that in the limit as  $n \rightarrow \infty$

$$f_{\beta_0}(u) = \sqrt{\frac{\beta_0}{2\pi}} e^{-\frac{1}{2}\beta_0 u^2}. \quad (21)$$

This is the  $n = 1$  Maxwellian of (13) which is the conditional pdf with which Beck started!

What has happened is that the derivation of the Student t-pdf (15) formally resembles subordination but, in reality, has nothing to do with it. Suppose<sup>[16]</sup> that  $(n + 1)$  random variables  $u$  and  $u_1, u_2, \dots, u_n$  are independent and identically distributed according to the normal distribution with zero mean and standard deviation  $1/\sqrt{\beta_0}$ . The distribution of  $u$  is given by (21), while the distribution of the square root of the average of the sum

of squares,  $v = \sqrt{\frac{1}{2} \sum_{i=1}^n u_i^2}$  is

$$f_{\beta_0}(v) = \frac{2}{\Gamma(n/2)} \left( \frac{n\beta_0}{2} \right)^{n/2} v^{n-1} e^{-\frac{1}{2}n\beta_0 v^2}. \quad (22)$$

On account of the independence of the random variables  $u$  and  $v$ , their joint pdf will be given by:

$$f_{\beta_0}(u) f_{\beta_0}(v) = \frac{2}{\Gamma(n/2)} \sqrt{\frac{2}{\pi}} \left( \frac{n}{2} \right)^{n/2} \beta_0^{(n+1)/2} v^{n-1} e^{-\frac{1}{2}\beta_0(u^2 + n v^2)}. \quad (23)$$

The probability that  $u/v \leq t$  is the integral of (23) over the region  $v > 0$  and  $u \leq tv$ . Introducing the transformation  $u = xy$  and  $v = y$ , whose Jacobian is  $y$ , gives:

$$F(t) = \frac{2}{\Gamma(n/2)} \sqrt{\frac{2}{\pi}} \left( \frac{n}{2} \right)^{n/2} \beta_0^{(n+1)/2} \int_0^\infty dx \int_0^\infty y^n e^{-\frac{1}{2}\beta_0 n(1+x^2/n)y^2} dy.$$

Its derivative yields Student's t-pdf:

$$f(t) = \frac{1}{\sqrt{n\pi}} \frac{\Gamma(\frac{n+1}{2})}{\Gamma(\frac{n}{2})} \left( 1 + \frac{t^2}{n} \right)^{-(n+1)/2}. \quad (24)$$

It is simply a coincidence that the  $\chi^2$  pdf which Beck chose to represent fluctuations in the inverse temperature, (10), has the same form as Maxwell's density for the speed

$\sqrt{\frac{1}{n} \sum_{i=1}^n u_i^2}$ . Surprisingly, the Student t-distribution (24) is independent of the variance,  $1/\beta_0$ . Although this is important insofar as testing hypotheses of the mean of a population is concerned, because it is independent of the variance, it nevertheless implies that information has been lost (Eq. 28 and 30).

The reason why Beck has  $\beta_0$  as a parameter in his pdf (15) is that he obtained:

$$\begin{aligned} & \frac{1}{\sqrt{2\pi}} \frac{1}{\Gamma(n/2)} \left( \frac{n}{2\beta_0} \right)^{\frac{n}{2}} \int_0^\infty \beta_0^{(n-1)/2} e^{-\frac{1}{2}\beta_0(u^2 + n/\beta_0)} d\beta \\ &= \frac{1}{B(\frac{1}{2}, n/2)} \sqrt{\frac{\beta_0}{n}} \left( 1 + \frac{\beta_0 u^2}{n} \right)^{-(n+1)/2}, \end{aligned}$$

using (10) instead of (11). Had he used (11), he would have obtained:

$$f_{u_0}(u) = \frac{1}{B(\frac{1}{2}, n/2)} u_0^{-1} \left( 1 + \frac{u^2}{u_0^2} \right)^{-(n+1)/2}.$$

The two are the same if equipartition holds:

$$u_0 = \sqrt{\frac{n}{\sum_{i=1}^n u_i^2}} = \sqrt{\frac{n}{\beta_0}}. \quad (25)$$

In general, the estimator of the variance,  $\frac{1}{n} \sum_{i=1}^n u_i^2$ , will be different from the variance  $\beta_0^{-1}$ . The estimator is

$\beta^{-1}$  and usually this temperature will be different from the temperature,  $\beta_0^{-1}$ , of the reservoir. In other words, the

statistic for the Student distribution,  $t = u / \sqrt{\sum_{i=1}^n u_i^2}$ , is not distributed as a normal random variable as it would be

had the statistic been given by  $t = u / \sqrt{n/\beta_0}$ . The actual variance may be completely unknown and it suffices to know only the sample variance in order to make statistical predictions. Nevertheless, as the number of degrees of freedom increases, it is expected that the distribution of  $t$  will be very similar to that of a standard normal variable; that is, it is a ‘consistent’ estimator. Accordingly, the  $\chi_c^2$  in Beck’s Eq. 3, which are distributed as  $\chi_c^2$ , may be identified with the kinetic energies of the individual particles and the left-hand side should be  $n/\beta$  and not  $\beta$  itself.

The whole idea of ‘superstatistics’ is that the exponential pdf<sup>[17]</sup>:

$$f_{\beta_0}(E) = \frac{e^{-\beta_0 E} \rho(E)}{Z(\beta_0)} \quad (26)$$

is actually a conditional pdf and specifying a normalised pdf for what was previously a mere parameter by the  $\chi_c^2$  pdf:

$$f_{E_0}(\beta) = \frac{1}{\Gamma(n/2)} E_0^{\frac{n}{2}} \beta^{\frac{n}{2}-1} e^{-\beta E_0} \quad (27)$$

leads to a new distribution of  $E$  which is precisely the Tsallis pdf. However, (27) is no arbitrarily chosen pdf for the inverse temperature, but rather the directing process which results when the parameter of the original distribution is randomised.

Consider the usual case of a power law for the structure function<sup>[7]</sup>:

$$\rho(E) = \frac{E^{m-1}}{\Gamma(m)},$$

which leads to equipartition because the partition function is  $Z(\beta) = \beta^{-m}$ . Then the process which is subordinated to the exponential pdf (26) has a pdf:

$$f_{E_0}(E) = \int_0^{\infty} f_{\beta}(E) f_{E_0}(\beta) d\beta = \frac{1}{B(m, n/2)} \frac{(E/E_0)^{m-1}}{(1 + E/E_0)^{m+n/2}} \frac{1}{E_0} \quad (28)$$

which is still the inverted beta pdf Eq. 14. This is because the Lévy transform:

$$\beta_0 E = \beta E_0 \quad (29)$$

in (26) produces a randomised inverse temperature which is distributed according to (27) with  $n = 2m$  degrees of freedom. Consequently, the inverse beta pdf (28) is the Fisher-Snedecor pdf with equal degrees of freedom,  $m = n/2$ , or what is more commonly known as the Pareto distribution. The transformation to a new process by randomising the operational temperature cannot change the number of degrees of freedom of the system.

‘F-superstatistics’<sup>[5,17]</sup> is not a directing process but, rather, a subordinated process that is obtained from (11) and (13) by averaging the product of these densities over a common value of the kinetic energy, viz.<sup>[11]</sup>:

$$f_{\beta_0}(\beta) = \int f_{\beta_0}(u) f_u(\beta) du = \frac{1}{B(n/2, n/2)} \frac{(\beta/\beta_0)^{n/2-1}}{(1 + \beta/\beta_0)^n} \frac{1}{\beta_0} \quad (30)$$

Since it is necessary that both the original and directing processes have the same number of degrees of freedom, (30) is a symmetrical inverse beta distribution, which necessarily has the same number of degrees of freedom. Comparing (30) with (28) it is seen easily that they are interchangeable on the strength of the Lévy transformation (29). Reference to a Tsallis distribution in  $\beta$  space is completely inappropriate.

The symmetry of the inverse beta pdf (30) means that we are dealing with a composite system comprised of two subsystems with the same number of degrees of freedom. The total energy  $E_t = E_0 + E$ , formed from two subsystems with energies  $E_0$  and  $E$ , is fixed. Thus, the inverse beta pdf (28) is transformed into the beta pdf which may be written as the composition law<sup>[13]</sup>:

$$\rho(E_t) = \int \rho(E_t - E) \rho(E) dE = \frac{E_t^{2m-1}}{\Gamma(2m)}$$

for the structure function.

Thus, it may be concluded that the statement ‘‘Of course, other distribution functions  $f(\beta)$  can also be considered which may lead to other generalised statistics’’<sup>[18]</sup> is devoid of meaning when the original process is distributed according to (26). More importantly, as has been concluded elsewhere, ‘‘subordination can be considered as the probabilistic origin of power laws in physics’’<sup>[13]</sup>.

### TOUCHETTE AND BECK’S ‘ASYMPTOTIC SUPERSTATISTICS’

The basic idea contained in Touchette and Beck’s<sup>[19]</sup> study is to obtain ‘fat’ tail power distributions by a ‘mixing’ of Boltzmann factors. Precisely why such a mixing will produce fat tail power laws is left to the reader’s imagination.

A ‘mixing’ distribution,  $f(\beta)$ , is used to obtain a new ‘distribution’,  $B(E)$ , via a Laplace transform:

$$B(E) = \int_0^{\infty} f(\beta) e^{-\beta E} d\beta \tag{31}$$

where,  $E$  is the energy and  $\beta$  its conjugate intensive parameter, as in (ii) of section 2. Both  $f(\beta)$  and  $B(E)$  are claimed unimodal but this is both unnecessary and misleading since it is required only that they be monotone<sup>[13]</sup>. In the article under discussion here, the Laplace integral was evaluated in the high energy,  $E \rightarrow \infty$ ,  $\beta \rightarrow 0$ , limit using Laplace’s method. However, when the inverse transform is considered:

$$f(\beta) = \frac{1}{2\pi i} \int_{b-i\infty}^{b+i\infty} B(E) e^{\beta E} dE, \tag{32}$$

along the vertical line  $\Re\beta = b$ , in the region of existence of  $B(E)$ , it is surprising to learn that the same method may be applied only in the opposite limit as  $E \rightarrow 0$ ,  $\beta \rightarrow \infty$ . This appears all the more disconcerting since no information is lost in a Laplace transformation: A distribution is determined uniquely by its Laplace transform and vice-versa. Both should yield the same expressions in the same asymptotic limit.

Any relation giving the asymptotic behaviour of  $B(E)$  in terms of  $f(\beta)$  is known as a Tauberian theorem, whereas the inverse relation is termed Abelian<sup>[14]</sup>. The behaviour of  $B(E)$  as  $E \rightarrow \infty$  completely determines the behaviour of  $f(\beta)$  near the origin and vice-versa. Without loss of generality, power laws may be considered for both  $f(\beta)$  and  $B(E)$ , one of which must be monotonically increasing and the other monotonically decreasing. For the Laplace transform (31) to exist, it is necessary that the exponent in:

$$f(\beta) = \frac{\beta^{m-1}}{\Gamma(m)}, \tag{33}$$

be confined to the semi-open interval  $[0, \infty)$ . This gives:

$$B(E) = E^m, \tag{34}$$

which is the tail distribution that Touchette and Beck derive by Laplace’s method in the limit as  $E \rightarrow \infty$ . They are quick to point out that (34) has the same asymptotic form as Tsallis’  $q$ -exponential distributions<sup>[13]</sup>, where,  $m = (q-1)^{-1}$ .

However, the Laplace transform (31) defines a generating function,  $B(E)$  and not a normalisable probability density. The conjugate distribution is the  $\chi^2$  distribution<sup>[16]</sup>:

$$p(\beta|E_0) = E_0 \frac{(\beta E_0)^{m-1}}{\Gamma(m)} e^{-\beta E_0}, \tag{35}$$

characterised by the ‘parameter’,  $E_0$  and not the probability distribution<sup>[19]</sup>:

$$p(E) = B(E)/Z,$$

where:

$$Z = \int B(E) dE,$$

since the integral diverges. Hence, a simple Laplace transform on the ‘density of states’,  $f(\beta)$ , will not lead to a power-tail distribution.

However, the  $\chi^2$ -distribution (35) for the inverse temperature may be derived from the  $\chi^2$ -distribution for the energy by Bayes’ theorem of inverse probability<sup>[5]</sup>. Moreover, it is the distribution of the directing process which transforms the  $\chi^2$ -distribution:

$$p(E|\beta_0) = \beta_0 \frac{(\beta_0 E)^{m-1}}{\Gamma(m)} e^{-\beta_0 E}, \tag{36}$$

for the energy into a beta distribution of the second kind for the energy. This is accomplished again through sudordination<sup>[14]</sup>, in which the parameter  $\beta_0$  is replaced by the lower limit on the energy distribution<sup>[13]</sup>.

Once again, the Lévy transformation (Eq. 12):

$$\beta_0 E = \beta E_0,$$

where, the subscripted quantities refer to the heat reservoir, enters the deliberations, since it provides the relation between the two  $\chi^2$ -distributions, (35) and (36). By analogy with Brownian motion, where a distribution of particles may be considered, or, equivalently, a single particle whose motion in time is followed, the conjugate variables of position and time become those of energy and temperature in thermodynamics<sup>[13]</sup>. The ‘pure’ spatial process, where the initial position replaces the time parameter, is the Cauchy distribution and the Cauchy process is subordinated to the Brownian motion<sup>[13]</sup>. Analogously, in thermodynamics, the beta density of the second kind is subordinate to the  $\chi^2$ -distribution (36), since<sup>[13]</sup>:

$$p(E|E_0) = \int_0^{\infty} p(E|\beta) p(\beta|E_0) d\beta = B^{-1}(m,m) \frac{E_0 (E_0 E)^{m-1}}{(E_0 + E)^{2m}}, \tag{37}$$

where, as before,  $B(.,.)$  is the beta function. For  $E_0 \ll E$ , the beta density (37) transforms into the Pareto density:

$$p(E|E_0) \approx \frac{E_0^m}{E^{m+1}}.$$



The energy  $E_0$  is then interpreted as the smallest energy for which the distribution is valid, corresponding to the minimum income in Pareto's distribution of income above a given value.

Touchette and Beck claim that Laplace's method will only work on the inverse transform (32) in the limit as  $\beta \rightarrow \infty$  and  $E \rightarrow 0$ , i.e., the low temperature limit. The distribution they find by using Laplace's method is:

$$f(\beta) \approx \frac{e^{\beta E_0 + \ln B(E_0)}}{\sqrt{-(\ln B(E_0))'}}$$

where, the prime denotes differentiation with respect to the argument, which is the implicit solution to:

$$(\ln B(E))' = -\beta. \tag{38}$$

However,  $\beta E + \ln B(E)$  has a minimum at  $E_\beta = m/\beta$  and not a maximum as contended by the authors. Hence, Laplace's method will not work regardless of which asymptotic limit is being considered. It is surprising also that the authors find it necessary to work in the opposite asymptotic limit when discussing the inverse transform. As noted earlier, no information is lost upon taking the Laplace transform, or its inverse and so the two should correspond to the same asymptotic limit.

As is well-known, the inverse Laplace transform (32) may be performed by contour integration. There is a pole of order  $m$  at the origin and  $B(E) \rightarrow 0$  as  $E \rightarrow \infty$ . If the pole at the origin is enclosed by a contour  $C_m$ , Cauchy's theorem gives:

$$f(\beta) = \int_{C_m} E^{-m} e^{\beta E} dE = \frac{1}{\Gamma(m)} \left[ \frac{d^{m-1}}{dE^{m-1}} e^{\beta E} \right]_{\beta=0} = \frac{\beta^{m-1}}{\Gamma(m)},$$

which is the density (33) with which we started.

It seems sensible now to enquire into the physical meaning of  $f(\beta)$  and  $B(E)$ . According to the process of subordination, the primary process is the  $\chi^2$ -distribution (36) for the energy. The directing process is governed by the inverse temperature distribution (35). This is analogous to randomising time in Brownian motion where the Brownian motion process for the increments of a Brownian particle is transformed into a Lévy distribution for the random increments in time.

Equilibrium statistical mechanics teaches that the fundamental quantity is the improper probability distribution, or 'structure' function<sup>[20]</sup>:

$$\Omega(E) = \frac{E^{m-1}}{\Gamma(m)}, \tag{39}$$

representing the density of states in phase space. The conjugate probability distribution is the exponential law

$$p(E|\beta) = \frac{e^{-\beta E}}{Z(\beta)} \Omega(E), \tag{40}$$

where, the generating function is:

$$Z(\beta) = \int_0^\infty \Omega(E) e^{-\beta E} dE = \beta^{-m}. \tag{41}$$

The functions  $f(\beta)$ , (33) and  $B(E)$ , (34), refer to the inverse of the generating function, (41) and the inverse of the structure function, (39), respectively. Hence, they are devoid of any physical content.

Khinchin<sup>[20]</sup> contends that the conjugate distribution (40) may be approximated by the normal distribution on the strength of the central limit theorem. Setting:

$$\frac{e^{-\beta E}}{Z(\beta)} \Omega(E) = \frac{\exp\left\{-\left(E + (\ln Z)'\right)^2\right\}}{\sqrt{2\pi(\ln Z)''}}, \tag{42}$$

and evaluating it at the mean energy,  $\bar{E} = m/\beta = -(\ln Z)'$ , in the high temperature limit gives the important result:

$$\Omega(\bar{E}) = \frac{Z(\beta) e^{\beta \bar{E}}}{\sqrt{2\pi(\ln Z)''}}. \tag{43}$$

Now, introducing this explicitly into (43) gives<sup>[7]</sup>:

$$\Omega(\bar{E}) = \frac{\bar{E}^{-m-1} m}{\sqrt{2\pi m^{m+\frac{1}{2}} e^{-m}}} = \frac{\bar{E}^{-m-1} m}{m!} = \frac{\bar{E}^{-m-1}}{\Gamma(m)},$$

which is precisely Stirling's expression for  $m!$

The reason for this is that the normal law (43) applies in the limit where the number of degrees of freedom of the system has been allowed to increase without limit. Alternatively, if a finite number of degrees of freedom is maintained and the conjugate distribution is identified with the  $\chi^2$ -distribution (36) for the energy, then the expression for the density of states is found to be:

$$\Omega(E) = Z(\beta) e^{\beta E} \frac{\beta(\beta E)^{m-1}}{\Gamma(m)} e^{-\beta E} = \frac{E^{m-1}}{\Gamma(m)}, \tag{44}$$

without any approximation. Obviously, if (44) was evaluated at  $E = \bar{E} = m/\beta$ , Stirling's approximation,  $m! = \sqrt{2\pi m} m^{m+\frac{1}{2}} e^{-m}$ , would be obtained once again.

At this point it is appropriate to enquire after the distinction between the  $\chi^2$ -and normal distributions as far as the processes are involved. This is a matter for order statistics. The probability that one of the  $(n-1)$  molecules has energy falling in the interval from

$E_1$  to  $E_1 + dE_1$  is  $(n-1)dE/E$ . There are then  $(n-2)$  molecules remaining,  $(m-1)$  of which have energies  $\leq dE_1$ , while  $(n-m-1)$  have energies  $\geq dE_1$ . The probability that  $(m-1)$  have energies  $\leq E_1$  is  $(E_1/E)^{m-1}$ , while those having energies  $> E_1$  is  $(1-E_1/E)^{n-m-1}$ . The product of the three factors is proportional to the  $\beta$  density of the first kind:

$$p(E_1) = \frac{1}{B(m, n-m)E} \left(\frac{E_1}{E}\right)^{m-1} \left(1 - \frac{E_1}{E}\right)^{n-m-1}$$

The mean energy of the molecules with energies  $\leq E_1$  is simply the fraction  $(m/n)$  of the total energy. If  $n$  is so large that the equipartition result for the total energy,  $E = n/\beta$ , may be used and  $n \rightarrow \infty$ , the  $\beta$  density of the first kind will transform into the  $\chi^2$ -distribution (36). The mean value  $\bar{E}_1 = m/\beta$  is again given by the equipartition result, which is entirely independent of the asymptotically infinite total number of particles  $n$ .

**CONCLUSIONS**

Here attention has been drawn to discrepancies appearing in discussions of so-called ‘superstatistics’ which were introduced into the statistical mechanical literature following Tsallis’s study<sup>[1]</sup>. The Tsallis entropy is a pseudo-additive entropy of degree- $q$ <sup>[2]</sup> which has been known and used by information scientists for many years. In fact, the basic relation,

$$S(xy) = S(x) + S(y) + (1 - q)S(x)S(y), \tag{45}$$

which Tsallis<sup>[21]</sup> used as a definition of a nonextensive thermodynamic system, is not the definition of nonextensivity at all! For  $q = 1$ , the solution of the functional equation is  $S(x) = \log x$ , while for  $q \neq 1$ , the solution is  $S(x) = (x^q - 1)/(1 - q)$ . The  $x$  and  $y$  in (45) are sets of probabilities and not extensive thermodynamic variables. The major hurdle in characterising information theoretic entropies is that the set of random quantities is not independent of the set of their probabilities. Equality (45) has no meaning as a definition of a nonextensive entropy unless a prescription is given that relates the set of probabilities to the set of random quantities. Furthermore there is no reason to think that this prescription would be unique or that it will preserve the functional form of (45).

Equation 45 has been shown many years ago by Hardy, Littlewood and Pólya<sup>[22]</sup> to characterise the generators of homogeneous means which constitute an equivalent class. Two generators are said to be equivalent if there exists a linear relation between them:

$$\tilde{S}(x) = \alpha S(x) + \beta, \tag{46}$$

where,  $\alpha$  and  $\beta$  are constants. If  $\tilde{S}(x) = S(\lambda x)$ , where  $\lambda > 0$ , (32) becomes:

$$S(\lambda x) = \alpha(\lambda)S(x) + \beta(\lambda)$$

and if  $S(1) = 0$ , then  $\beta(\lambda) = S(\lambda)$ . Hence, if  $\lambda = y$ , (46) becomes:

$$S(xy) = \alpha(y)S(x) + S(y).$$

Interchanging  $x$  and  $y$  and eliminating  $S(xy) = S(yx)$  between the two expressions gives:

$$\frac{\alpha(x) - 1}{S(x)} = \frac{\alpha(y) - 1}{S(y)} = 1 - q = \text{const.}$$

Introducing this into (34) gives the pseudo-additive relation for the generators of the power means. For reasons given by Rényi<sup>[23]</sup>, it is the negative logarithm of these means which should be identified with information theoretic entropies-and not the generators of these means which, at most, may be considered as pseudo-additive entropies<sup>[3]</sup>.

**REFERENCES**

1. Tsallis, C., 1988. Possible generalisation of Boltzmann-Gibbs statistics. *J. Stat. Phys.*, 52: 470.
2. Lavenda, B.H. and J. Dunning-Davies, 2005. Pseudo-additive entropies of degree- $q$  and the Tsallis entropy. *J. Applied Sci.*, 5: 315.
3. Lavenda, B.H., 2004. Information and coding discrimination of pseudo-additive entropies. *Open Sys. Inform. Dyn.*, 11: 257.
4. Lavenda, B.H., 2004. Entropies as means. *Open Sys. Inform. Dyn.* (In Press).
5. Beck, C. and E.G.D. Cohen, 2003. Superstatistics. *Physica, A* 322: 267.
6. Mandelbrot, B., 1956. An outline of a purely phenomenological theory of statistical thermodynamics. *IRE Trans. Inform. Theory*, 2: 190.
7. Lavenda, B.H., 1991. *Statistical Physics: A Probabilistic Approach*, Wiley-Interscience, New York.
8. Jeffrey, H., 1961. *Theory of Probability*. 3rd. Edn. Clarendon, Oxford.
9. Zellner, A., 1971. *An Introduction to Bayesian Inference in Economics*, Wiley, New York.
10. Bleistein, N. and R.A. Handelsman, 1986. *Asymptotic Expansions of Integrals*. Dover, New York.
11. Beck, C. *Dynamical foundation of nonextensive statistical mechanics*. <http://arxiv.org/conclmat/010537+>

12. de Souza, A.M.C. and C. Tsallis, 1997. Student's t- and r-distributions. A unified derivation from an entropic variational principle. *Physica, A* 236: 52.
13. Lavenda, B.H., 1995. *Thermodynamics of Extremes*. Horwood, Chichester.
14. Feller, W., 1971. *An Introduction to Probability Theory and its Applications*, Vol. II, 2nd. Edn. Wiley, New York.
15. Cramér, H., 1946. *Mathematical Methods of Statistics*. Princeton U.P., Princeton NJ.
16. Lavenda, B.H., 2004. On the definition of fluctuating temperature. *Open Sys. Inform. Dyn.*, 11: 139.
17. Beck, C., 2004. Superstatistics, escort distributions and applications. *Physica, A* 342: 139.
18. Beck, C., 2002. Nonextensive methods in turbulence and particle physics. *Physica, A* 305: 209.
19. Touchette, H. and C. Beck, 2004. *Asymptotic Superstatistics*.
20. Khinchin, A.I., 1949. *Mathematical Foundations of Statistical Mechanics*, Dover, New York.
21. Tsallis, C., 2005. Is the entropy  $S_q$  extensive or nonextensive? cond-mat/0409631. In: *Complexity, Metastability and Nonextensivity*, Beck, C., A. Rapisarda, C. Tsallis, Eds., World Scientific, Singapore.
22. Hardy, G., J.E. Littlewood and G. Pólya, 1934. *Inequalities*. Cambridge U. P., Cambridge.
23. Rényi, A., 1961. On measures of entropy and information. In: *Proc. 4th. Berkeley Symp. Math. Statist. Probability 1960*, (Univ. Calif. Press, Berkeley) 547; 1965. On the foundations of information theory. *Rev. Intl. Statist. Inst.*, 33: 1.