

Probability distribution and entropy as a measure of uncertainty

This article has been downloaded from IOPscience. Please scroll down to see the full text article.

2008 J. Phys. A: Math. Theor. 41 065004

(<http://iopscience.iop.org/1751-8121/41/6/065004>)

View [the table of contents for this issue](#), or go to the [journal homepage](#) for more

Download details:

IP Address: 171.66.16.152

The article was downloaded on 03/06/2010 at 07:24

Please note that [terms and conditions apply](#).

Probability distribution and entropy as a measure of uncertainty

Qiuping A Wang

Institut Supérieur des Matériaux et Mécaniques Avancés du Mans, 44 Av Bartholdi,
72000 Le Mans, France

Received 24 August 2007, in final form 2 December 2007

Published 29 January 2008

Online at stacks.iop.org/JPhysA/41/065004

Abstract

The relationship between three probability distributions and their maximizable entropy forms is discussed without postulating entropy property. For this purpose, the entropy I is defined as a measure of uncertainty of the probability distribution of a random variable x by a variational relationship $dI = d\bar{x} - \overline{dx}$, a definition underlying the maximization of entropy for corresponding distribution.

PACS numbers: 05.20.-y, 02.50.-r, 02.50.Tt

1. Introduction

It is well known that entropy and information can be considered as measures of uncertainty of probability distribution. However, the functional relationship between entropy and the associated probability distribution has since long been a question in statistical and informational science. There are many relationships established on the basis of the properties of entropy. In the conventional information theory and some of its extensions, these properties are postulated, such as the additivity and the extensivity in the Shannon information theory. The reader can refer to [1–8] to see several examples of entropies proposed on the basis of postulated entropy properties. Among all these entropies, the most famous one is the Shannon informational entropy ($S = -\sum_i p_i \ln p_i$) [2], which was almost the only one widely used in equilibrium thermodynamics (Boltzmann–Gibbs entropy) and in nonequilibrium dynamics (Kolmogorov–Sinai entropy for example). But the question remains open in the scientific community about whether or not Shannon entropy is the unique useful measure of statistical uncertainty or information [9].

The origin of this question can be traced back to the principle of maximum entropy (maxent) by Jaynes [10] who claimed the Shannon entropy was singled out to be the only consistent measure of uncertainty to be maximized in maxent. In view of the fact that this uniqueness was argued from the Shannon postulates of information property [2], one naturally asks what happens if some of these properties are changed. Some of the entropies in the list of

[1] are found by mathematical considerations which change the logic of Shannon. Recently, a nonextensive statistics (NES) [6–8] proposed to use some entropies for thermodynamics and stochastic dynamics of certain nonextensive systems. NES has given rise to a large number of papers with very different viewpoints dealing with equilibrium and nonequilibrium systems and incited considerable debate [11–13] within the statistical physics community. Some key questions in the debate are: whether or not it is necessary to replace Boltzmann–Gibbs–Shannon entropy with other ones in different physical situation? What are the possible entropy forms which can be maximized in order to derive probability distribution according to maxent?

One remembers that in the actual applications of maxent the used entropy forms are either directly postulated or derived from postulated properties of entropy [1–8]. The correctness of these entropies is then verified through the validity of derived probability distributions. In the present work, we will invert this reasoning in order to find maximizable entropy form directly from known probability distributions without postulating the properties of entropy. For this purpose, we need a generic entropy definition underlying in addition variational approach or maxent. Inspired by the first and second laws of thermodynamics for reversible process, we introduce a variational definition of entropy I such as $dI = d\bar{x} - \bar{d}x$ for the measure of probabilistic uncertainty of the simple situation with only one random variable x . We stress that the main objective of this work is to show the non-uniqueness of Shannon entropy as maximizable uncertainty measure. Other entropy forms must be introduced for different probability distributions. We would like to stress also that this is a conceptual work tackling the mathematical form of entropy without considering the detailed physics behind the distribution laws used in the calculations. In what follows, we first talk about three probability distributions and their invariant properties. The maximizable entropy form for each of them is then derived thanks to the definition $dI = d\bar{x} - \bar{d}x$.

2. Three probability laws and their invariance

In this section, by some trivial calculations one can find in textbooks, we want to underline the fact that a probability distribution may be derived uniquely from its invariance. By invariance of a function $f(x)$, we means that the dependence on x is invariant at transformation of x into x' , i.e., $f(x')\mu f(x)$. We consider three invariances corresponding to exponential, power law and q -exponential distributions, respectively.

2.1. Translation invariance and exponential law

Suppose that $f(x)$ is invariant by a translation of $x \rightarrow x + b$, i.e.,

$$f(x + b) = g(b)f(x), \tag{1}$$

where $g(b)$ depends on the form of $f(x)$. We have $\frac{df(x+b)}{db} = \frac{df(x+b)}{d(x+b)} = g'(b)f(x)$ and $\frac{df(x)}{dx} = g'(0)f(x)$ or $\frac{df(x)}{f} = g'(0) dx$ ($b = 0$). This means

$$\ln f(x) = g'(0)x + c \quad \text{or} \quad f(x) = ce^{g'(0)x}, \tag{2}$$

where c is some constant. If $f(x)$ gives a probability such as $p(x) = \frac{1}{Z}f(x)$ where $Z = \sum_x f(x)$, the normalization condition $\sum_x p(x) = 1$ will make $p(x)$ strictly invariant versus the transformation $x \rightarrow x' = x + b$, i.e., $p(x') = \frac{1}{Z'}f(x') = \frac{1}{Z}f(x) = p(x)$ since $Z' = \sum_{x'} f(x') = \sum_x f(x + b) = g(b)\sum_x f(x) = Zg(b)$.

2.2. Scale invariance and power law

Now suppose that $f(x)$ is scale invariant, we should have

$$f(bx) = g(b)f(x), \tag{3}$$

where b is the scale factor of the transformation. We make following calculation $\frac{df(bx)}{db} = \frac{df(bx)}{d(bx)}x = g'(b)f(x)$ to get $\frac{df(x)}{dx}x = g'(1)f(x)$, which means

$$f(x) = cx^{g'(1)}. \tag{4}$$

This kind of laws is widely observed in nature for different dynamical systems such as language systems [14] and scale free networks [15] among many others [16]. The well-known Levy flight for large x is a good example of power law with $g'(1) = -1 - \alpha$ where $0 < \alpha < 2$.

2.3. The q -exponential and its invariance properties

Here we would like to mention a probability which has attracted a lot of attention in the last years,

$$f(x) = c[1 + a\beta x]^{\frac{1}{a}}, \tag{5}$$

where a and β are some constants. The Zipf–Mandelbrot law $f(x) = c[1 + x]^{-\alpha}$ observed in textual systems and other evolutionary systems [17] can be considered as a kind of q -exponential law. Another example of this law is the equilibrium thermodynamic distribution for finite systems in equilibrium with a finite heat bath, where a can be related to the number of elements N of the heat bath and tends to zero if N is very large [18], which implies $f(x) = c[1 + a\beta x]^{\frac{1}{a}} \xrightarrow{a \rightarrow 0} ce^{\beta x}$.

This distribution is not a power law in the sense of equation (4). It has neither the scale invariance nor the translation invariance mentioned above. The operator on x that keeps $f(x)$ invariant is a generalized addition $x + {}_a b = x + b + a\beta x b$ [19], i.e. $f(x + {}_a b) = c[1 + a\beta(x + b + a\beta x b)]^{\frac{1}{a}} = [1 + a\beta b]^{\frac{1}{a}}c[1 + a\beta x]^{\frac{1}{a}} = g(b)f(x)$, where $g(b) = [1 + a\beta b]^{\frac{1}{a}}$.

The derivation of equation (5) from $f(x + {}_a b) = g(b)f(x)$ is given as follows. First, we make a derivative such as $\frac{df(x+{}_a b)}{db} = \frac{df(x+{}_a b)}{d(x+{}_a b)} \frac{d(x+{}_a b)}{db} = \frac{df(x+{}_a b)}{d(x+{}_a b)}(1 + a\beta x) = g'(b)f(x)$. Then let $b = 0$, we get $\frac{df(x)}{dx}(1 + a\beta x) = g'(0)f(x)$ which means $\frac{df(x)}{f(x)} = g'(0) \frac{dx}{(1+a\beta x)}$ or $\ln f = \ln(1 + a\beta x)^{1/a} + c$ or equation (5) with $\beta = g'(0)$.

3. A definition of entropy as a measure of dynamical uncertainty

Suppose we have a random (discrete) variable x_i with a probability distribution $p_i = \frac{1}{Z} f(x_i)$, where i is the state index. The average of x_i is given by $\bar{x} = \sum_i x_i p_i$ and the normalization is $\sum_i p_i = 1$. The uncertainty in a probability distribution of x can be measured by many quantities. The standard deviation σ or the variance $\sigma^2 = \overline{x^2} - \bar{x}^2$, for example, can surely be used if they exist. Shannon entropy, as some of other known entropy forms, can also be used as a measure of uncertainty of any p_i . But certainly any given entropy form, including Shannon one, cannot be maximized for any distribution p_i according to maxent rules, which can be seen below. The main task of this work, let us recall it, is to search for a general definition of uncertainty measure underlying maxent, in such a way that each derived entropy can be maximized to give the corresponding distribution. Here, we propose a quite general

measure I by a variational definition as follows:

$$dI = d\bar{x} - \overline{dx} = \sum_i x_i dp_i, \quad (6)$$

where $\overline{dx} = \sum_i p_i dx_i$. This definition has been in a way inspired by the first and second laws of thermodynamics in equilibrium thermodynamics. Considering the definition of internal energy $\overline{E} = \sum_i p_i E_i$, where E_i is the energy of the microstate i with probability p_i , we can write $\delta\overline{E} = \sum_i \delta p_i E_i + \sum_i p_i \delta E_i = \sum_i \delta p_i E_i + \delta\overline{E}_i$ in which $\delta\overline{E}_i = \sum_i p_i \delta E_i = \sum_j (\sum_i p_i \frac{\partial E_i}{\partial q_j}) \delta q_j$ is the work done to the system by external forces $F_j = \sum_i p_i \frac{\partial E_i}{\partial q_j}$, where q_j is the extensive variables such as volume, distance or electrical polarization. According to the first law of thermodynamics, the quantity $\sum_i \delta p_i E_i = \delta\overline{E} - \delta\overline{E}_i$ is the heat change in the system, that is, $\sum_i \delta p_i E_i = \delta Q = T\delta S$ for a reversible process, where S is the thermodynamic entropy and T is the absolute temperature. S has the following variational relation:

$$\delta S = \frac{1}{T} (\delta\overline{E} - \delta\overline{E}_i). \quad (7)$$

As is well known, S measures the dynamical disorder of the system or the uncertainty of the probability distribution of energy in the dynamics. In contrast with another measure $\sigma^2 = \overline{E^2} - \overline{E}^2$, S can be maximized in maxent for equilibrium system to derive probability distribution. At this point, we must mention a work by Plastino and Curado [20] on the equivalence between the particular thermodynamic relation $\delta S = \beta\delta\overline{E}$ and maxent for probability assignment. In order to obtain this heat–energy relationship from the first and second laws of thermodynamics, they considered the particular reversible process affecting only the microstate population, i.e., the distribution p_i has a variation δp_i due to a heat transfer $\delta Q = T\delta S$. Although their variational approach from a particular process still needs further justification, their conclusion is important and consequential. Not only have they shown the equivalence between $\delta S = \beta\delta\overline{E}$ and maxent, they also reached a quite generic variational principle by virtue of the fundamental laws of thermodynamics. To our opinion, as a variational method, $\delta S = \beta\delta\overline{E}$ is much more general and powerful than maxent with originally a unique maximizable measure, since it enables one to maximize other form of entropy if any for equilibrium system, which opens the door for other development of statistical mechanics on the basis of maxent for equilibrium system.

Coming back to equation (6) which is just an extension of equation (7) to arbitrary random variables x and to arbitrary system (even out of equilibrium) and a generic definition of maximizable uncertainty measure for any random variable. The maximizability of this measure consists to put $\overline{dx} = 0$ as discussed below.

The geometrical aspect of the uncertainty measure defined by equation (6) can be illustrated in the examples of figure 1 which shows that dI and I are related to the width of the distributions on the one hand, and to the form of the distribution on the other. dI is not an increasing function of the distribution width. For example, $dI = 0$ for uniform distribution whatever the width of $p(x) = \text{constant}$. This means that I is a constant.

4. Three probability distribution laws and their entropy

In this section, on the basis of the uncertainty measure defined in equation (6), we derive the entropy functionals for the three probability laws discussed in section 2.

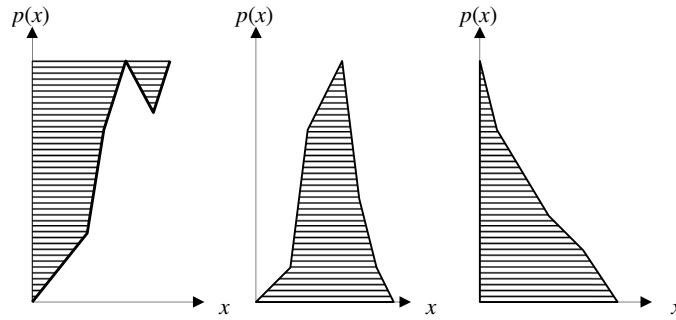


Figure 1. Three geometrical representations of the variation dI defined in equation (6) for some distributions. The hatched areas represent the value of the entropy variation $dI = d\bar{x} - \overline{dx} = \sum_i x_i dp_i$ for each case. In general, dI for any increasing or decreasing part of a distribution curve is the area between the $p(x)$ -axis and the considered part. So in this figure, dI for any part of the distribution curve is the area from the $p(x)$ -axis to the left of the part. The sign of dI depends on the way it is calculated. If it is calculated with increasing x , dI is positive for the increasing part of $p(x)$ since $dp_i > 0$ (main parts of left panel) and negative for the decreasing part of $p(x)$ since $dp_i < 0$ (right panel). In the middle panel, the hatched area is the difference of $dI_{inc} > 0$ for the increasing part and the $dI_{dec} < 0$ for the decreasing part. Since the region of dI_{inc} is smaller than that of dI_{dec} , the difference is negative. It is easily seen that $dI = 0$ if the width (uncertainty) of the distribution is zero or if the distribution is uniform.

4.1. Translation invariant probability and Shannon entropy

The following calculation is trivial. From equation (6), for exponential distribution $p_i = \frac{1}{Z} e^{-x_i}$, we have

$$dI = -\sum_i \ln(p_i Z) dp_i = -\sum_i \ln p_i dp_i - \ln Z \sum_i dp_i = -\sum_i \ln p_i dp_i = -d\sum_i p_i \ln p_i$$

and

$$I = -\sum_i p_i \ln p_i + c.$$

This is Shannon information if the constant c is neglected. Within the conventional statistical mechanics, this is the Gibbs formula for Clausius entropy. Remember that the maximization of this entropy using Lagrange multiplier associated with expectation of x yields exponential distribution law.

4.2. Scale invariant probability and entropy functional

We have in this case power law probability distribution p_i ,

$$p_i = \frac{1}{Z} x_i^{-a}, \tag{8}$$

where $Z = \sum_i x_i^{-a}$. Put it into equation (6) to get

$$\begin{aligned} dI &= \sum_i (p_i Z)^{-1/a} dp_i = Z^{-1/a} \frac{1}{1-1/a} d\sum_i p_i^{1-1/a} \\ &= -Z^{-1/a} d\left[\sum_i p_i^{1-1/a} / (1-1/a) + c \right], \end{aligned} \tag{9}$$

where c is an arbitrary constant. Since we are addressing a given system to find its entropy form, Z can be considered as a constant for the variation in x (the reader will find below that this constant can be given by the Lagrange multiplier in the maximum entropy formalism). Hence we can write

$$I \propto \sum_i p_i^{1-1/a} / (1 - 1/a) + c. \tag{10}$$

In order to determine c , we imagine a system with two states $i = 1$ and 2 with $p_1 = 0$ and $p_2 = 1$. In this case, $I = 0$ so that

$$\left(\frac{0+1}{1-1/a} \right) + c = 0, \tag{11}$$

i.e.,

$$c = -\frac{1}{1-1/a}. \tag{12}$$

We finally get

$$I = \frac{\sum_i p_i^{1-1/a} - 1}{1-1/a}. \tag{13}$$

Let $q = \frac{1}{a}$, we can write

$$I = -\frac{1 - \sum_i p_i^{1-q}}{1-q} = -\sum_i \frac{p_i - p_i^{1-q}}{1-q}. \tag{14}$$

Note that this functional does not yield Shannon entropy for $q \rightarrow 1$. As a matter of fact, q must be positive and smaller than unity. I is negative if q is greater than unity or smaller than zero, which does not make sense. For large x Lévy flight, for example, $1 < a < 3$, so $\frac{1}{3} < q < 1$.

It can be calculated that

$$I = -\sum_i \frac{p_i - p_i^{1-q}}{1-q} = -\sum_i p_i \frac{1 - p_i^{-q}}{1-q} = -\sum_i p_i \frac{1 - \frac{1}{Z^{-q}}x}{1-q} = -\frac{1 - Z^q \bar{x}}{1-q}. \tag{15}$$

Its behavior with probability is shown in figure 2. The maximization of I conditioned with a Lagrange multiplier β such as $\delta(I - \beta \bar{x}) = 0$ directly yields the power law of equation (8) with $\beta = Z^q = Z^{1/a}$.

4.3. The entropy for q -exponential probability

We have seen above that the probability $p_i = c[1 - a\beta x_i]^{-\frac{1}{a}}$ had a special invariant property. Let us express x as a function of p_i and put it into equation (6) to get

$$dI = \sum_i \frac{1 - (p_i/c)^a}{a\beta} dp_i = -\frac{1}{a\beta c^a} \sum_i p_i^a dp_i = -\frac{1}{a\beta c^a(1+a)} d\left(\sum_i p_i^{1+a} + c\right). \tag{16}$$

By the same tricks for determining c in the above section, we get $c = -1$. So we can write

$$I = -\frac{\sum_i p_i^{1+a} - 1}{a} = -\sum_i \frac{p_i - p_i^q}{1-q}. \tag{17}$$

Where $q = 1 + a$ and we have used the normalization $\sum_i p_i = 1$. This is the Tsallis entropy which tends to the Shannon entropy for $q \rightarrow 1$ or $a \rightarrow 0$. In this case, $p_i = c[1 - a\beta x_i]^{-\frac{1}{a}}$ tends to an exponential distribution.

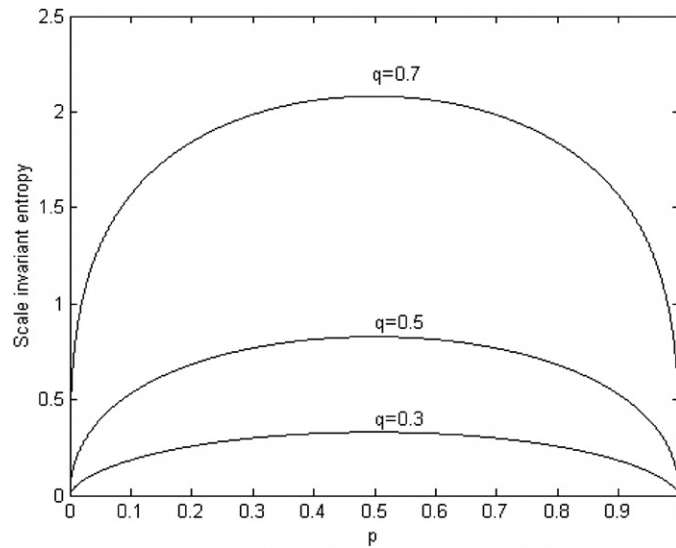


Figure 2. The variation of the scale invariant entropy $I = -\sum_{i=1}^2 \frac{p_i - p_i^{1-q}}{1-q}$ with $p_1 = p$ and $p_2 = 1-p$ for different q values. It can be shown that if $q \rightarrow 0$, $S = -\frac{q}{1-q} \sum p \ln p \rightarrow 0$, and If $q \rightarrow 1$, $S = \frac{\sum 1-p}{1-q} + \sum \ln p \rightarrow \infty$.

5. Concluding remarks

We have derived the entropy functionals for three probability distributions. This was done on the basis of a variational definition of uncertainty measure, or entropy without postulating entropy property (such as additivity) as in the usual information theory. The variational definition $dI = d\bar{x} - \overline{dx}$ is valid for any probability distributions of x as long as it has finite expectation value. According to the results, the exponential probability has the Shannon entropy, the power-law distribution has an entropy $I = -\frac{1 - \sum_i p_i^{1-q}}{1-q}$ where $0 < q < 1$, and the q -exponential distribution has Tsallis entropy $I = -\sum_i \frac{p_i - p_i^q}{1-q}$ where q is positive.

It is worth mentioning again that the present definition of entropy as a measure of uncertainty offers the possibility of introducing the maximum entropy principle in a natural way with Lagrange multipliers associated with expectation of the random variables. It is easy to verify, with the above three entropies, that the maximum entropy calculus yields the original probability distributions. This is not an ordinary and fortuitous mutual invertibility, since the probability and the entropy are not reciprocal functions and the maximum entropy calculus is not a usual mathematical operation. As a matter of fact, this invertibility between entropy and probability resides in the variational definition $dI = d\bar{x} - \overline{dx}$. As discussed in the section 3, \overline{dx} can be considered as an extended work whatever the nature of x . So to get the ‘equilibrium state’ or stable probability distribution, we can put $\overline{dx} = 0$ just as in the mechanical equilibrium condition where the vector sum of all forces acting on an object should be 0, an underlying idea of the principle of virtual work in mechanics. We straightforwardly get $dI - \beta d\bar{x} = 0$ or $d(I + \alpha \cdot \sum_i p_i + \beta \overline{dx}) = 0$ if we add the normalization condition. This reasoning leads in a more natural way to the maxent idea $\delta S - \beta \delta E = 0$ of Plastino and Curado [20] for thermodynamic equilibrium.

Acknowledgments

We thank Professor F Tsohnang and Professor S Abe for useful discussion and comments. We would like to express here our thanks to the referees for their pertinent criticism and very useful comments and suggestions.

References

- [1] Esteban M D 1995 *Kybernetika* **31** 337
- [2] Shannon C E 1948 *Bell Syst. Tech. J.* **27** 379
- [3] Rényi A 1966 *Calcul de probabilité* (Dunod: Paris) p 522
Wehrl A 1978 *Rev. Mod. Phys.* **50** 221
- [4] Harvda J and Charvat F 1967 *Kybernetika* **3** 30
- [5] Daroczy Z 1970 *Inform. Control* **16** 36
- [6] Tsallis C 1988 *J. Stat. Phys.* **52** 479
- [7] Kaniadakis G 2000 *Phys. Rev. E* **296** 405
- [8] Wang Q A and Chaos 2001 *Solitons Fractals* **12** 1431
Wang Q A 2003 *Entropy* **5**
- [9] Uffink J 1995 Can the maximum entropy principle be explained as a consistency requirement *Studies Hist. Phil. Mod. Phys.* **26B** 223–61
- [10] Jaynes E T 1984 The evolution of Carnot's principle *The opening talk at the EMBO Workshop on Maximum Entropy Methods in x-ray Crystallographic and Biological Macromolecule Structure Determination, (Orsay, France, April) pp 24–8*
Gibbs vs Boltzmann entropies 1965 *Am. J. Phys.* **33** 391
Ray Smith C, Grandy W T Jr and Reidel D (ed) 1985 Where do we go from here? *Maximum Entropy and Bayesian Methods in Inverse Problems* (Publishing Company) pp 21–58
- [11] Balian R and Nauenberg M 2006 *Europhys. News* **37** 9
Luzzy R, Vasconcellos A R and Ramos J G 2006 *Europhys. News* **37** 11
- [12] Nauenberg M 2003 *Phys. Rev. E* **67** 036114
Nauenberg M 2004 *Phys. Rev. E* **69** 038102 *Preprint* cond-mat/0305365v1
- [13] See also the special issue of *Comptes Rendus Physique*, **7**, (2006) and the special issue of *Europhysics news*, **36/6** (2005)
- [14] Li W 1992 Random texts exhibit Zipf's-law-like word frequency distribution *IEEE Trans. Inform. Theory* **38** 1842–5
- [15] Newman M E J 2003 The structure and function of complex networks *SIAM Rev.* **45** 167–256
- [16] Newman M E J 2005 Power laws, Pareto distributions and Zipf's law *Contemp. Phys.* **46** 323–51
- [17] Silagadze Z K 1999 Citations and the Zipf–Mandelbrot's law, physics/9901035
- [18] Terletskii Y P 1971 *Statistical Physics* (Amsterdam: North-Holland)
- [19] Nivonen L, Le Méhauté A and Wang Q A 2003 Generalized algebra within a nonextensive statistics *Rep. Math. Phys.* **52** 437
Borges E P 2004 A possible deformed algebra and calculus inspired in nonextensive thermostatics *Physica A* **340** 95
- [20] Plastino A and Curado E F M 2005 *Phys. Rev. E* **72** 047103
Curado E F M and Plastino A 2006 *Preprint* cond-mat/0601076
Curado E F M and Plastino A 2005 *Preprint* cond-mat/0509070