



**HAL**  
open science

# Probability distribution and entropy as a measure of uncertainty

Qiuping A. Wang

► **To cite this version:**

Qiuping A. Wang. Probability distribution and entropy as a measure of uncertainty. 2007. hal-00117452v3

**HAL Id: hal-00117452**

**<https://hal.science/hal-00117452v3>**

Preprint submitted on 24 Aug 2007 (v3), last revised 9 Jan 2008 (v4)

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Probability distribution and entropy as a measure of uncertainty

Qiuping A. Wang

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

## Abstract

The relationship between three probability distributions and their optimal entropy forms is discussed without postulating entropy property as usual. For this purpose, the entropy  $I$  is defined as a measure of uncertainty of the probability distribution  $p(x)$  of a random variable  $x$  by a variational relationship  $dI = \overline{dx} - \overline{dx}$ , a definition assuring the optimization 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 probability distribution and the associated entropy 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 the references [1] to [8] to see several examples of entropies proposed on the basis of different 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 measure of statistical uncertainty or information [9].

Recently, a nonextensive statistics [6][7][8] (NES) proposed to use other entropies for thermodynamics and stochastic dynamics of certain nonextensive systems. NES has given rise to a large number of papers in the last decade with very different viewpoints dealing with equilibrium and nonequilibrium systems, which have incited more and more debates [10][12] within the statistical physics community. Some of the key questions in the debates are: whether or not it is necessary to replace Boltzmann-Gibbs-Shannon entropy with other ones in different physical situation? what is the useful forms of entropy? It should be remembered that all the known entropy forms have been either postulated directly or derived from some postulated properties of entropy [1-8]. The correctness of these entropies is often verified through their application in the inference and the probability assignment.

The present work inverts the reasoning and investigates entropy form under a different angle without postulating the properties of entropy. Inspired by a thermodynamic relationship between entropy and energy, we introduce a variational definition of entropy as the measure of probabilistic uncertainty. One of the objectives of this work is to show that, Shannon entropy is not unique as uncertainty measure. Other forms are possible if we change the rules of reasoning and introduce new criteria, as we have done in this work in introducing the variational definition which underlies a criterion for the entropy: the uncertainty measure is optimal for the corresponding stable probability distribution. We indicate that this work is a

conceptual one 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 invariant probability distributions and their invariant properties. Their optimal entropy forms are then derived from the variational entropy definition.

## 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.,  $f(x') \propto f(x)$ . We consider three invariances corresponding to exponential, power law and q-exponential distributions, respectively, which are the most interesting for our work.

### a) Translation invariance and exponential law

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

$$f(x+b) = g(b)f(x) \quad (1)$$

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

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

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.,

$$p(x') = \frac{1}{Z}f(x') = \frac{1}{Z}f(x) = p(x) \text{ since } Z' = \sum_{x'} f(x') = \sum_x f(x+b) = g(b) \sum_x f(x) = Zg(b).$$

### b) Scale invariance and power law

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

$$f(bx) = g(b)f(x) \quad (3)$$

where  $b$  is the scale factor of the transformation. We make following calculation

$$\frac{df(bx)}{db} = \frac{df(bx)}{d(bx)} x = g'(b)f(x) \text{ to get } \frac{df(x)}{dx} x = g'(1)f(x), \text{ which means}$$

$$f(x) = c x^{g'(1)}. \quad (4)$$

This kind of laws is widely observed in nature for different dynamical systems such as language systems[13] and scale free networks[14] among many others[15]. The well known Levy flight for large  $x$  is a good example of power law with  $g'(1) = -1 - \alpha$  where  $0 < \alpha < 2$ .

### c) 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:

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

where  $a$  and  $\beta$  are some constants. The Zipf-Mandelbrot law  $f(x) = c [1 + x]^{-\alpha}$  observed in textual systems and other evolutionary systems[16] 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[17], which implies

$$f(x) = c [1 + a\beta x]^{\frac{1}{a}} \xrightarrow{a \rightarrow 0} c e^{\beta x}.$$

This distribution is not a power law in the sense of Eq.(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  $x +_a b = x + b + a\beta x b$  [18], 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  $g(b) = [1 + a\beta b]^{\frac{1}{a}}$ .

### 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 this probability distribution of  $x$  can be measured by many quantities. For example, the standard deviation  $\sigma$  or the variance  $\sigma^2 = \overline{x^2} - \bar{x}^2$  can surely be used if they exist. A disadvantage of  $\sigma^2$  is that it may not exist for many probability distributions. Here we propose another measure which seems much general. This is a variational definition of entropy as a measure of uncertainty given by following relationship

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

This choice of uncertainty measure has been in a way inspired by the first and second laws of thermodynamics in equilibrium statistical thermodynamics. Considering the definition of internal energy  $\bar{E} = \sum_i p_i E_i$  where  $E_i$  is the energy of the state  $i$  with probability  $p_i$ , we can

write  $\delta\bar{E} = \sum_i \delta p_i E_i + \sum_i p_i \delta E_i = \sum_i \delta p_i E_i + \overline{\delta E_i}$ . It can be proved that

$\overline{\delta E_i} = \sum_i p_i \delta E_i = \sum_j \left( \sum_i p_i \frac{\partial E_i}{\partial q_j} \right) \delta q_j$  is the work done to the system by external forces

$F_j = \left( \sum_i p_i \frac{\partial E_i}{\partial q_j} \right)$  where  $q_j$  is extensive variables such as volume, distance or electrical

polarization. According to the first law of thermodynamics, the quantity  $\sum_i \delta p_i E_i = \delta\bar{E} - \overline{\delta E_i}$

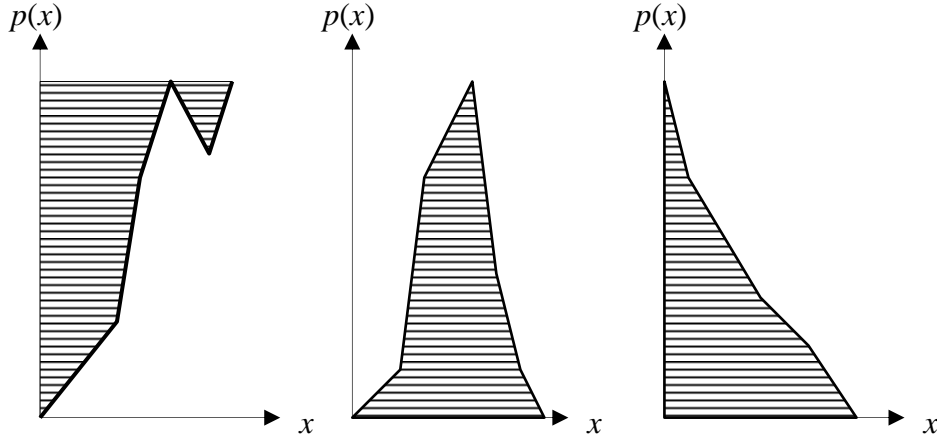
must be the heat change in the system, that is,  $\sum_i \delta p_i E_i = \delta Q = T \delta S$  for a reversible process,

where  $S$  is the thermodynamic entropy and  $T$  the absolute temperature. Hence the thermodynamic entropy must satisfy the following variational relation

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

This relationship is extended in Eq.(6) to arbitrary random variables  $x$ . By this definition, it is obvious that if the distribution is not exponential, the entropy functional may not be logarithmic.

The geometrical aspect of the uncertainty measure defined by Eq.(6) can be seen 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.



**Figure 1**, Three geometrical representations of the variation  $dI$  defined in Eq.(6) for some distributions. The hatched areas represent the value of the entropy variation  $dI = d\bar{x} - \bar{dx} = \sum_i x_i d p_i$  for each case. It is easily seen that  $dI=0$  if the width (uncertainty) of the distribution is zero or if the distribution is uniform.

#### 4) Three probability distribution laws and their entropy

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

##### a) Translation invariant probability and Shannon entropy

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

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

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.

*b) Scale invariant probability and entropy functional*

We have in this case power law probability distribution  $p_i$  :

$$p_i = \frac{1}{Z} x_i^{-a} . \quad (8)$$

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

$$dI = \sum_i (p_i Z)^{-1/a} d p_i = Z^{-1/a} \frac{1}{1-1/a} d \sum_i p_i^{1-1/a} = -Z^{-1/a} d[\sum_i p_i^{1-1/a} / (1-1/a) + c] \quad (9)$$

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

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

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

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

i.e.,

$$c = -\frac{1}{1-1/a} \quad (12)$$

We finally get

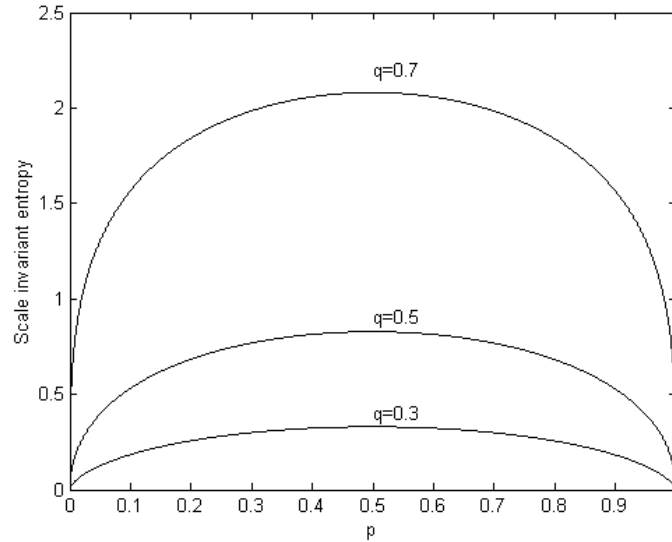
$$I = \frac{\sum_i p_i^{1-1/a} - 1}{1-1/a} \quad (13)$$

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

$$I = \frac{1 - \sum_i p_i^{1-q}}{1-q} = -\sum_i \frac{p_i - p_i^{1-q}}{1-q} \quad (14)$$



Notice 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  $\frac{1}{3} < q < 1$ .



**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  $q \rightarrow 1$ ,  $S = \frac{\sum 1-1}{1-q} + \sum \ln p \rightarrow \infty$ .

It can be calculated that

$$I = -\sum_i \frac{p_i - p_i^{1-q}}{1-q} = -\sum_i p_i \frac{1 - p_i^{-q}}{1-q} = -\sum_i p_i \frac{1 - \frac{1}{Z^{-q}x}}{1-q} = -\frac{1 - Z^{-q}x}{1-q} \quad (15)$$

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

c) *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 Eq.(6) to get

$$dI = \sum_i \frac{1 - (p_i/c)^a}{a\beta} d p_i = -\frac{1}{a\beta c^a} \sum_i p_i^a d p_i = -\frac{1}{a\beta c^a (1+a)} d(\sum_i p_i^{1+a} + c) \quad (16)$$

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

$$I = -\frac{\sum_i p_i^{1+a} - 1}{a} = -\sum_i \frac{p_i - p_i^q}{1-q} \quad (17)$$

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.

## 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 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. We straightforwardly get  $dI-\beta d\overline{x}=0$  or  $d(I+\alpha\sum_i p_i+\beta\overline{x})=0$  if we add the normalization condition. This is the usual maximum entropy principle using Lagrange multipliers  $\alpha$  and  $\beta$ . The entropy  $I$  defined by  $dI = d\overline{x} - \overline{dx}$  goes naturally to the conditioned maximum for its corresponding probability distribution.

## Acknowledgements

We thank Professor F. Tsohnang and Professor S. Abe for very useful discussion and comments.

## References

- [1] M.D. Esteban, *Kybernetika*, **31**(1995)337
- [2] C.E. Shannon, *The Bell System Technical Journal*, **27**(1948)379
- [3] A. Rényi, *Calcul de probabilit e*,(Paris, Dunod, 1966)P522  
A. Wehrl, *Rev. Mod. Phys.*, **50**(1978)221
- [4] J. Harvda and F. Charvat, *Kybernetika*, **3**(1967)30
- [5] Z. Daroczy, *Information and control*, **16**(1970)36
- [6] C. Tsallis, *J. Statis. Phys.*, **52**(1988)479
- [7] G. Kaniadakis, *Phys. Rev. E*, **296** (2000)405
- [8] Q.A. Wang, *Chaos, Solitons & Fractals*, **12**(2001)1431;  
Q.A. Wang, *Entropy*, **5**(2) (2003)
- [9] Jos Uffink, Can the maximum entropy principle be explained as a consistency requirement, *Studies in History and Philosophy of Modern Physics*, **26B** (1995): 223-261
- [10] R. Balian and M. Nauenberg, *Europhysics news*, **37/2** (2006)9;  
R. Luzzy, A.R. Vasconcellos and J.G. Ramos, *Europhysics news*, **37/2** (2006)11
- [11] Nauenberg M., *Phys. Rev. E*, **67** (2003) 036114; *Phys. Rev. E*, **69** (2004) 038102; cond-mat/0305365v1.
- [12] See also the special issue of *Comptes Rendus Physique*, (2006) and the special issue of *Europhysics news*, **36/6** (2005)
- [13] W. Li, Random texts exhibit Zipf's-law-like word frequency distribution, *IEEE Transactions on Information Theory*, **38**(6)(1992)1842-1845
- [14] M. E. J. Newman, The structure and function of complex networks, *SIAM Review*,

**45**, (2003) 167-256

- [15] M. E. J. Newman, Power laws, Pareto distributions and Zipf's law, Contemporary Physics **46**(2005) 323-351
- [16] Z. K. Silagadze, Citations and the Zipf-Mandelbrot's law, physics/9901035
- [17] Y.P. Terletsii, Statistical physics, North-Holland Publishing Company, Amsterdam, 1971
- [18] L. Nivanen, A. Le Méhauté, and Q.A. Wang, Generalized algebra within a nonextensive statistics, Reports on Mathematical Physics, **52**(2003)437;  
E.P. Borges, A possible deformed algebra and calculus inspired in nonextensive thermostatics, Physica A, **340**(2004)95