Some Possible rationales for Rényi-Tsallis entropy maximization

Jean-François Bercher

Some possible rationales for Rényi-Tsallis entropy maximization

Jean-François Bercher

LabInfo-IGM,
Université Paris-Est/ESIEE,
77454 Marne la Vallée Cedex 2, France
(e-mail: jf.bercher@esiee.fr)

Abstract. Distributions derived from the maximization of Rényi-Tsallis entropy are often called Tsallis' distributions. We first indicate that these distributions can arise as mixtures, and can be interpreted as the solution of a standard maximum entropy problem with fluctuating constraints. Considering that Tsallis' distributions appear for systems with displaced or fluctuating equilibriums, we show that they can be derived in a standard maximum entropy setting, taking into account a constraint that displace the standard equilibrium and introduce a balance between several distributions. In this setting, the Rényi entropy arises as the underlying entropy.

Another interest of Tsallis distributions, in many physical systems, is that they can exhibit heavy-tails and model power-law phenomena. We note that Tsallis' distributions are similar to Generalized Pareto distributions, which are widely used for modeling the tail of distributions, and appear as the limit distribution of excesses over a threshold. This suggests that they can arise in many contexts if the system at hand or the measurement device introduces some threshold. We draw a possible asymptotic connection with the solution of maximum entropy. This view gives a possible interpretation for the ubiquity of Tsallis' (GPD) distributions in applications and an argument in support to the use of Rényi-Tsallis entropies.

Keywords. Rényi entropy, Tsallis entropy, Displaced equilibriums, Generalized Pareto Distributions.

1 Introduction

The Tsallis (1988) entropy was introduced in statistical physics, originally for the description of multifractals systems. It is defined by

$$S_\alpha(f_X) = \frac{1}{1-\alpha}\left(\int f_X\left(x\right)^\alpha dx - 1\right) \tag{1}$$

where α is a real parameter and f_X an univariate distribution. Tsallis entropy is simply related to the Rényi (1961) entropy introduced on firm axiomatics grounds. In the two cases, the Shannon entropy $S_1(f_X) = -\int f_X(x) \log f_X(x) dx$ is obtained in the limit case $\alpha = 1$.

It is also convenient here to introduce the Kullback-Leibler and Rényi divergences from a density p to q:

$$I(p||q) = \int p(x) \log \frac{p(x)}{q(x)} dx \text{ and } I_\alpha(p||q) = \frac{1}{\alpha - 1} \int p(x)^\alpha q(x)^{1-\alpha} dx. \quad (2)$$

The definition of $I(p||q)$ requires that p is absolutely continuous with respect to q. It is understood, as usual, that $0 \log 0 = 0 \log 0/0 = 0$ and that $a \log(a/0) = +\infty$ if $a > 0$.

Since the introduction of Tsallis entropy in statistical physics, a considerable amount of work has been devoted to the study of these alternative entropies and to the associated thermostatistics, called nonextensive thermostatistics. Indeed, it is known that for some complex systems, the celebrated Gibbs-Boltzmann formalism seems insufficient for a good description. Applications include fully developed turbulence, Levy anomalous diffusion, statistics of cosmic rays, econometry, and many others.

For this kind of problems, the formalism of nonextensive statistical mechanics (which recovers the classical one as a special limit case) leads to Tsallis' distributions. In a wide variety of fields, experiments, numerical results and analytical derivations fairly agree with the formalism and the description by a Tsallis distribution; see for instance Tsallis (2002) and references therein.

These distributions are of very high interest in many physical systems, since they can exhibit heavy-tails, and model power-law phenomena (i.e. with density $f_X(x) \propto x^{-a}, a > 0$). Indeed, power-laws are especially interesting since they appear widely in physics, biology, economy, and many other fields, see the review in Newman (2005).

In physical systems, the equilibrium Boltzmann-Gibbs distribution can be derived, à la Jaynes (1957), as a maximum entropy distribution, or as a minimum Kullback-Leibler divergence distribution, for a system submitted to a mean constraint (the value of the internal energy for instance). This derives from Sanov theorem whose essence is that, for a parent distribution q, then in the set of empirical probability distributions compatible with a given constraint, the one that becomes overwhelmingly preponderant, i.e. that has the greatest probability, is the nearest to q in the Kullback-Leibler sense.

Therefore, distributions derived using a maximum Tsallis entropy will not coincide with those derived for the classical maximum entropy approach and the maximum Tsallis entropy distribution, in the words of Grendar and Grendar (2004), *will be asymptotically improbable*. In view of the success of nonextensive statistics and of the successful identification of Tsallis' distributions in physical problems, there should exist some probabilistic setting that provides a justification for the maximization of Tsallis entropy. Some such possible rationales are described in this communication.

The Tsallis' distributions are obtained as the result of the maximization of Tsallis entropy (1) subject to a mean constraint and normalization. Two kind of constraints have been considered: the standard mean $\mathrm{E}[X] = m$, and a generalized α-expectation $\mathrm{E}_\alpha[X] = m$, which is taken with respect to the 'zooming distribution' $p(x)^\alpha / \int p(x)^\alpha dx$. These two constraints lead to a solution with the general form

$$f_X(x) = \frac{1}{\sigma}\left(1 + \frac{\gamma}{\sigma}x\right)^{-\frac{1}{\gamma}-1} \text{ for } x_F > x \geq 0, \tag{3}$$

where σ and γ are respectively the scale and shape parameters; for $\gamma < 0$, the density has a finite support, with $x_F = -\sigma/\gamma$ if $\gamma < 0$ and an infinite support otherwise. In the first case, the exponent is $1/(\alpha - 1)$ and $\sigma = 1/\beta$, and in the second case, the exponent is $1/(1-\alpha)$ and $\sigma = 1/\beta(2-\alpha)$. For $\gamma = 0$ ($\alpha = 1$), the solution reduces to $f_X(x) = 1/\sigma \exp(-x/\sigma)$ for $\gamma = 0$.

2 Tsallis distributions as a mixture

Tsallis' distributions, have been found physically relevant for partially equilibrated systems, characterized by fluctuations of an intensive parameter, namely temperature fluctuations. They are obtained if the inverse of temperature, proportional to the parameter β of a Boltzmann distribution $\beta \exp(-\beta x)$, fluctuates according to a gamma distribution. This technique was later generalized as 'Superstatistics' by Beck and Cohen (2003) for more general distributions $f_\beta(\beta)$. In fact, this approach is the well-known technique of mixing, which is a standard technique for the derivation of distributions.

It is also possible (see Bercher (2008)), to relate this result to a (standard) maximum entropy approach, but with fluctuating constraints. Indeed, the canonical distribution maximizes the Shannon-Boltzmann entropy $S = -\int f_E(x) \log f_E(x) d\Gamma$ subject to normalization and to the observation constraint $\bar{E} = \int x f_E(x) d\Gamma$. The solution is $f_E(x|\bar{E}) = e^{-\beta x - \log Z(\beta)}$, where $Z(\beta) = \int e^{-\beta E} d\Gamma$ is the partition function. The entropy $S(\bar{E})$ and the potential $\Phi(\beta) = \log Z(\beta)$ are conjugated functionals while $\bar{E}$ and β are conjugated variables. This gives the standard relation $\bar{E} = -\frac{d \log Z(\beta)}{d\beta}$. Hence, it appears that variations of the thermodynamic β implies variations of its dual variable $\bar{E}$ and reciprocally. Thus, if the mean energy varies according to a distribution $f_{\bar{E}}(\bar{E})$, then this also gives a distribution $f_\beta(\beta)$ for β, and the global distribution becomes

$$P(E) = \int_0^{+\infty} f_E(E|\bar{E}) f_{\bar{E}}(\bar{E}) d\bar{E} = P(E) = \int_0^{+\infty} f_E(E|\beta) f_\beta(\beta) d\beta. \tag{4}$$

This is similar to the 'Superstatistics', but with the recognition of the link between the intensive parameter β and the observable $\bar{E}$, so that a model of fluctuations can be naturally introduced for $\bar{E}$. For instance an inverse-gamma model for $\bar{E}$ leads to Tsallis distributions[1]. Then, this distribution, which arises as the solution of a standard maximum entropy with

[1] A gamma model leads to the so-called K-distributions that also have interest in physical applications.

(inverse-gamma) fluctuating constraints, can in turn be viewed as the maximum Rényi-Tsallis entropy solution.

3 Amended MaxEnt

A key for the apparition of Levy distributions and a probabilistic justification might be that it seems to appear in the case of modified, perturbated, or displaced classical Boltzmann-Gibbs equilibrium. This means that the original MaxEnt formulation "find the closest distribution to a reference under a mean constraint" may be amended by introducing a new constraint that displaces the equilibrium, as discussed in Bercher (2006). The partial or displaced equilibrium may be imagined as an equilibrium characterized by two references, say r and q. Instead of selecting the nearest distribution to a reference under a mean constraint, we may look for a distribution p^* simultaneously close to two distinct references: such a distribution will be localized somewhere 'between' the two references r and q. For instance, we may consider a global system composed of two subsystems characterized by two prior reference distributions. The global equilibrium is attained for some intermediate distribution, and the observable may be, depending on the viewpoint or on the experiment, either the mean under the distribution of the global system or under the distribution of one subsystem. This can model a fragmentation process: a system $\Sigma(A, B)$ fragments into A, with distribution r, and B with distribution q, and the whole system is viewed with distribution p^* that is some intermediate between r and q. This can also model a phase transition: a system leaves a state q toward r and presents an intermediate distribution p^*. This intermediate distribution shall minimize its divergence to its "parent" distribution $q(x)$, but also be 'not too far' from its attractor $r(x)$. This can be stated as $I(p||q) - I(p||r) \leq \theta$, or equivalently as $I(p||r) \leq \theta'$. Remark that the first constraint can be interpreted as a constraint on the mean log-likelihood. The problem can be written as follows:

$$\begin{cases} \min_p \; I(p||q) \\ \text{s.t. } I(p||q) - I(p||r) \leq \theta \\ \text{or } I(p||r) \leq \theta' \end{cases} \tag{5}$$

The optimum distribution solution of these two equivalent problems is

$$p^*(x) = \frac{r(x)^\alpha q(x)^{1-\alpha}}{\int r(x)^\alpha q(x)^{1-\alpha} dx}, \tag{6}$$

The solution of the minimization problem satisfies a Pythagorean equality: $I(p||q) = I(p||p^*) + I(p^*||q)$ for any distribution p such that $\theta = I(p||q) - I(p||r)$. It is interesting to note that the solution (6) is nothing but the *escort* or *zooming* distribution of nonextensive thermostatistics. With the expression of the solution $p^*(x)$, we obtain that

$$I(p^*||q) = \alpha\theta - \log(\int R(x)^\alpha q(x)^{1-\alpha} dx) = \alpha\theta - (\alpha - 1) I_\alpha(r||q), \tag{7}$$

where we have recognized the Rényi divergence.

Suppose now that the original problem is completed by an additional constraint. For instance, one classically has to account for an observable. The observable values are as usual the statistical mean under some distributions. Depending on the viewpoint, the observable may be a mean under distribution r, the distribution of an isolated subsystem, or under p^*, the equilibrium distribution between r and q. Hence, the problem will be completed by an additional constraint, and we need to adjust the distribution r by further minimizing the Kullback-Leibler information divergence $I(p^*||q)$, but with respect to r and subject to the mean constraint. This finally amounts to the minimization of the Rényi divergence in (7), $I_\alpha(r||q)$ subject to the chosen mean constraint. Of course, the problem is similar to the maximization of Rényi-Tsallis entropy and leads to the Tsallis distribution (3), but with respect to the measure $q(x)$.

4 Tsallis distribution as the distribution of excesses

A remarkable feature of the maximum Tsallis entropy construction is its ability to exhibit heavy tailed distributions. Tsallis distributions have the form (3) of a Generalized Pareto Distribution (GPD). The interesting point is that the GPD is employed outside the statistical physics field for modeling heavy tailed distributions. Examples of applications are numerous, ranging from reliability theory, traffic in networks, hydrology, climatology, geophysics, materials science, radar imaging or actuarial sciences.

These uses are related to the POT (Peaks over Threshold) method; see Leadbetter (1991). The underlying rationale is the Balkema-de Haan-Pickands theorem (see Pickands (1975)), which asserts that the distribution of excesses over a (high) threshold often follows approximately a GPD.

Indeed, there are many situations in which the system at hand, the measurement process or device only give access to values of a variate X greater than a parameter u. Hence, the recorded values are the excesses of X over the threshold u, that is values of the *conditional* random variable X_u

$$X_u = (X - u)1_{X>u}.$$

The Pickands' theorem indicates, roughly speaking, that if F is in the domain of attraction of an extreme distribution, for instance the Fréchet distribution or Gumbel distribution, the distribution function of excesses F_{X_u} converges (in the infinite norm sense) to a limit distribution which is nothing else but a GPD.

To the accumulation of random variables in the classical Central Limit Theorem corresponds here an increase in the threshold u. Furthermore, the GPD verifies a remarkable *stability* property with respect to thresholding, in the sense that the distribution of excesses of a GPD remains a GPD, with the same shape (exponent) parameter γ but a different scale parameter.

It is possible to draw another connection between the distribution of excesses and the maximum of Tsallis entropy. In Bercher and Vignat (2008), we indicate that the 1-norm, first moment and Rényi-Tsallis entropy of a suitably normalized version of the excess variable (in the Fréchet and Gumbel domain) converge asymptotically to constant values. Then, with an appropriate choice of α, this shows that the distribution of excesses is necessarily, asymptotically, the GPD solution of Tsallis' α-entropy maximization).

We have presented here several possible rationales for the apparition of Tsallis distributions and of the underlying Rényi-Tsallis entropy. A common feature is that the distribution or entropy appears in what sounds as exceptional situations: modified, perturbated or fluctuating equilibriums, or excess over a threshold. However, we believe that there is not probably a single reason that explains the success of Rényi-Tsallis entropy and distribution, and that the multiplicity of possible rationales is itself an explanation of this success.

References

Beck, C. and Cohen, E. (2003): Superstatistics. *Physica A*, *322*, 267–275.

Bercher, J.F. (2006): An amended maxent formulation for deriving Tsallis factors, and associated issues. *AIP Conference Proceedings*, *872*, 441–448.

Bercher, J.F. and Vignat, C. (2008): An entropic view of Pickands' theorem. In: *International Symposium on Information Theory ISIT2008*, Toronto.

Bercher, J.F. (2008): Maximum entropy with fluctuating constraints - The example of K-distributions. *Physics Letters A*, in press.

Grendar, J. and Grendar, M. (2004): Maximum entropy method with non-linear moment constraints: challenges. In: G.J. Erickson and Y. Zhai, eds., *Bayesian Inference and Maximum Entropy Methods in Science and Engineering*, vol. 707, 97–109.

Jaynes, E.T. (1957): Information theory and statistical mechanics. *Phys. Rev.*, *106*, 620–630

Leadbetter, R. (1991): On a basis for 'Peaks over Threshold' modeling. *Statistics and Probability Letters*, *12*, 357–362.

Newman, M.E.J. (2005): Power laws, Pareto distributions and Zipf's law. *Contemporary Physics*, *46*, 323–351.

Pickands, J. (1975): Statistical inference using extreme order statistics. *The Annals of Statistics*, *3*, 119–131.

Rényi, A. (1961): On measures of entropy and information. Univ. California Press, Berkeley, Calif., pp. 547–561.

Tsallis, C. (1988): Possible generalization of Boltzmann-Gibbs statistics. *Journal of Statistical Physics*, *52*, 479–487.

Tsallis, C. (2002): Entropic nonextensivity: a possible measure of complexity. *Chaos, Solitons,& Fractals*, *13*, 371–391.