On escort distributions, q-gaussians and Fisher information

J.-F. Bercher

Laboratoire d'Informatique Gaspard Monge, Université Paris-Est, ESIEE,
5 bd Descartes, 77454 Marne-la-Vallée Cedex 2, France

Abstract. Escort distributions are a simple one parameter deformation of an original distribution p. In Tsallis extended thermostatistics, the escort-averages, defined with respect to an escort distribution, have revealed useful in order to obtain analytical results and variational equations, with in particular the equilibrium distributions obtained as maxima of Rényi-Tsallis entropy subject to constraints in the form of a q-average. A central example is the q-gaussian, which is a generalization of the standard gaussian distribution.

In this contribution, we show that escort distributions emerge naturally as a maximum entropy trade-off between the distribution $p(x)$ and the uniform distribution. This setting may typically describe a phase transition between two states. But escort distributions also appear in the fields of multifractal analysis, quantization and coding with interesting consequences. For the problem of coding, we recall a source coding theorem by Campbell relating a generalized measure of length to the Rényi-Tsallis entropy and exhibit the links with escort distributions together with pratical implications.

That q-gaussians arise from the maximization of Rényi-Tsallis entropy subject to a q-variance constraint is a known fact. We show here that the (squared) q-gaussian also appear as a minimum of Fisher information subject to the same q-variance constraint.

INTRODUCTION

Escort distributions have been introduced in statistical physics for the characterization of chaos and multifractals. These distributions P are a simple one parameter transformation of an original distribution p according to

$$P(x) = \frac{p(x)^q}{\int p(x)^q dx}. \tag{1}$$

The parameter q behaves as a microscope for exploring different regions of the measure p: for $q > 1$, the more singular regions are amplified, while for $q < 1$ the less singular regions are accentuated. This behavior is illustrated in Figure 1.

In Tsallis extended thermostatistics, the escort-averages, defined with respect to an escort distribution, have revealed useful in order to obtain analytical results and variational equations, with in particular the equilibrium distributions obtained as maxima of Rényi-Tsallis entropy subject to constraints in the form of a q-average. In applied fields, Tsallis distributions (q-distributions) have encountered a large success because of their remarkable agreement with experimental data, see [1], [2], [3], and references therein.
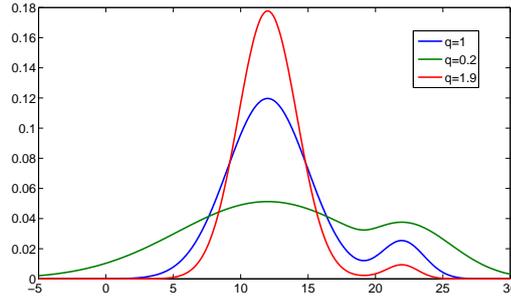
Figure 1. Behavior of the mapping $p(x) \to P(x) = p(x)^q / \int p(x)^q dx$. If U denotes the uniform distribution, it can be shown that $I(U||P) \geq I(U||p)$ if $q > 1$ with the reverse inequality if , where $I(p||q)$ denotes the Kullback-Leibler divergence from p to q.
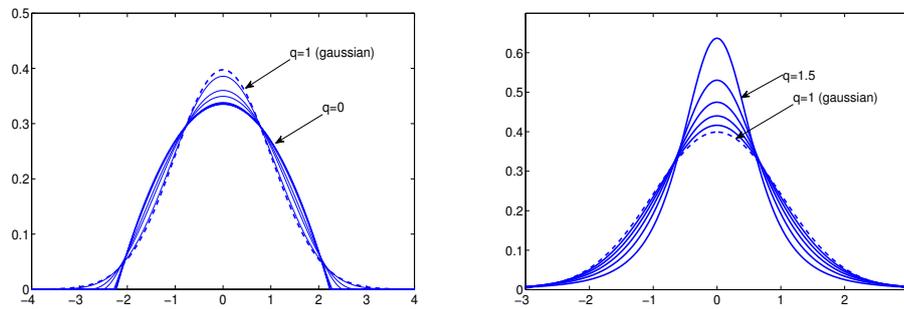

Figure 2. Examples of q-gaussian distributions with $\gamma = 1$ and $\sigma^2 = 1$, for several values of q. In the first case, $q \leq 1$, the distributions have compact support. In the second case where $q \geq 1$, the distributions have an infinite support and heavy tails. The limit case $q = 1$ is the standard gaussian distribution.

A central example is the q-gaussian distribution, which has the following probability density function

$$f(x) = \frac{1}{Z_q(\beta)} \left(1 - (1-q)\beta x^2\right)_+^{\frac{\gamma}{1-q}}, \quad q \neq 1 \tag{2}$$

where γ and β are positive parameters, and where we use the notation $(x)_+ = \max\{x, 0\}$. The partition function converges for $q < 2\gamma + 1$, and the variance exists for $q < 2\gamma/3 + 1$. Examples of q-gaussian distributions for different values of q are given on Figure 2, with $\gamma = 1$ and with variance $\sigma^2 = 1$.

In this contribution, we show that escort distributions (1) emerge naturally as a maximum entropy trade-off between the distribution $p(x)$ and the uniform distribution. This setting may typically describe a phase transition between two states. Then, looking for the distribution p with maximum entropy subject to a constraint on the mean energy computed with respect to the escort distribution, we arrive at the q-gaussian (2).

But escort distributions also appear in other fields with interesting consequences. In the problem of coding, we recall a source coding theorem by Campbell relating a generalized measure of length to the Rényi-Tsallis entropy. We show that the associated optimal codes can easily be obtained using considerations on escort-distributions and

that this provide an easy implementation procedure. We also show that these generalized lengths are bounded below by the Rényi entropy.

That q-gaussians arise from the maximization of Rényi-Tsallis entropy subject to a q-variance constraint is a known fact. We show here that the q-gaussian (with $\gamma = 2$) also appears as a minimum of Fisher information subject to the same q-variance constraint, and of course, that we recover the standard results in the limit $q = 1$.

ESCORT DISTRIBUTIONS

Escort as a model of displaced equilibrium

It has been noted that Tsallis' extended thermodynamic seems appropriate in the case of modified, perturbed, 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 we discussed in a previous MaxEnt workshop [4]. 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 f, and B with distribution g, and the whole system is viewed with distribution p^* that is some intermediate between f and g. This can also model a phase transition: a system leaves a state g toward f and presents an intermediate distribution p^*. This intermediate distribution shall minimize its divergence to the initial distribution $g(x)$, but also be 'not too far' from the attractor distribution $f(x)$. This can be stated as $I(p||g) - I(p||f) \leq \theta$, or equivalently as $I(p||f) \leq \theta'$, where $I(p||q)$ denotes the Kullback-Leibler divergence from a distribution p to another distribution q. Remark that the first constraint can be interpreted as a constraint on the mean log-likelihood. The problem can be written as follows:

$$
\begin{cases}
\min_p I(p||g) \\
\text{s.t. } I(p||g) - I(p||f) \leq \theta \\
\quad \text{or } I(p||f) < \theta',
\end{cases}
\tag{3}
$$

where "s.t." stands for "subject to". The optimum distribution solution of these two equivalent problems is

$$
f^*(x) = \frac{f(x)^q g(x)^{1-q}}{\int f(x)^q g(x)^{1-q} dx},
\tag{4}
$$

The solution of the minimization problem satisfies a Pythagorean equality: $I(p||g) = I(p||f^*) + I(f^*||g)$ for any distribution p such that $\theta = I(p||g) - I(p||f)$. It is interesting

to note that the solution (4) is nothing but a generalized version of the *escort* or *zooming* distribution of nonextensive thermostatistics. Obviously, one recover (1) when $g(x)$ is the (improper) uniform measure. With the expression of the solution $p^*(x)$, we obtain that

$$I(f^*||g) = q\theta - \log(\int f(x)^q g(x)^{1-q} dx) = q\theta - (q-1)I_q(f||g), \qquad (5)$$

where we have recognized the Rényi divergence with index q, $I_q(f||g)$ from f to g.

Suppose now that we get an additional constraint, in the form of an observable, given as some expectation under p^*, the equilibrium distribution between f and g. Then, a new question arises, that is the determination of the more general distribution f, compatible with this constraint.

We may use a MaxEnt point of view that consists in further minimizing the Kullback-Leibler information divergence $I(f^*||g)$, with respect to f. This finally amounts to the minimization of the Rényi divergence in (5), $I_q(f||g)$ subject to the mean constraint. In the case of a uniform distribution for g and a variance constraint we obtain the so-called q-variance constraint

$$\sigma_q^2 = \frac{\int x^2 f(x)^q dx}{\int f(x)^q dx}. \qquad (6)$$

Obsiously, the problem reduces to the maximization of Rényi-Tsallis entropy subject to this q-variance constraint, whose solution is known to be the standard q-gaussian distribution (2) with $\gamma = 1$.

Another possibility is to look for the less informative distribution in the Fisher sense, subject to the same constraint. We will see that this also leads to a q-gaussian distribution.

Rényi entropy, escort distributions and multifractals

Tsallis was originally inspired by the multifractals for the proposal of a generalized entropy. In this setting, one describes the properties of a singular measure in terms of singularity spectrum and generalized dimensions. A standard construction is to imagines that the support of the measure is covered with boxes of size l. Then if p_i is the probability in the i-th box, one puts $p_i(l) \sim l^{-\alpha_i}$, where α_i is the singularity exponent. Similarly, the number of boxes $N(\alpha)$ with an exponent α defines the singularity spectrum $f(\alpha)$ with $N(\alpha) \sim l^{-f(\alpha)}$. An alternative description is provided by the q-th 'moments' of the measure $N_q = \sum_i p_i(l)^q \sim l^{(q-1)D_q} \sim l^{\tau(q)}$, where D_q is the 'Rényi dimension' and $\tau(q)$ is the correlation exponent. The important connection between the two descriptions is that the pairs $(\alpha, f(\alpha))$ and $(q, \tau(q))$ are related by a Legendre transform [5], i.e. $f(\alpha) = q\alpha - \tau(q)$ with $\alpha = d\tau(q)/dq$.

It is well known that MaxEnt problems, as well as the standard thermodynamics are characterized by a beautiful Legendre structure. Thus, we readily see that the previous Legendre transform can be given a maximum entropy interpretation, where $f(\alpha)$, the singularity spectrum, is an entropy and $\tau(q)$ is the free energy. Here, the free energy $\tau(q)$ has the form $\tau(q) = \log \sum_i p_i(l)^q = \log \sum_i \exp(q \log p_i(l))$, and we may identify

$\tau(q)$ as $\tau(q) = \log \sum_i \exp(\beta E_i)$ with $\beta = q$ the inverse temperature and $E_i = \log p_i(l)$ the energy (also called here the 'bit-number'). This leads us to an easy thermodynamic interpretation: $\exp(\beta E_i)$ is the Boltzmann factor which results from the maximization of the entropy $-\sum_i P_i \log P_i$ subject to the mean constraint $U = \alpha = E_P[E_i] = E_P[\log p_i(l)]$. The standard resolution of this problem [6, 7] leads us to the solution $P_i = \frac{p_i^q}{\sum_i p_i^q}$, which is nothing else but the escort distribution of order q of $\{p_i\}$. Obviously, we then obtain that the singularity spectrum is the entropy of the escort, and the mean exponent is the mean of the bit-number:

$$f(\alpha) = \sum_i P_i \log P_i \text{ and } \alpha = E_P[\log p_i(l)].$$

These relationships are the rationale basis for the method introduced in [8] for the direct estimation of the singularity spectrum. In the latter paper, the escort distributions were introduced as a simple operational tool. They were later named "escort distributions" in the book by Beck [9].

Source coding

We also have an interesting connection with source coding, which was first reported in [10]. The mains formulas and comparisons are given on Table 1.

In source coding, one considers a source that produces a set of symbols $\mathcal{X} = \{x_1, x_2, \ldots x_N\}$ with probabilities p_i. The aim of source coding is to map each symbol x_i to a codeword c_i of length l_i expressed using the D letters of the alphabet. The Shannon source coding theorem (noiseless coding theorem) indicates that the expected length $\bar{L}$ of any uniquely decodable code is bounded below by the entropy of the source, $H_1(p)$.

It is well-known that Huffman coding yields a prefix code which minimizes the expected length and approaches the optimum limit $l_i = -\log_D p_i$ for the individual symbols. Other forms of lengths have also been considered, the first and fundamental contribution being Campbell's one [11]. In Shannon's result, low probabilities yield very long words. In order to penalize the apparition of such long codes, Campbell proposed a different length measure, featuring an exponential account of the elementary lengths of the codewords.

The remarkable result [11] is that just as Shannon entropy is the lower bound on the average codeword length of an uniquely decodable code, the Rényi entropy of order q, with $q = 1/(\beta + 1)$, is the lower bound on the exponentially weighted codeword length, see line 2 of Table 1.

In order to increase the impact of the longer lengths with low probabilities, the Campbell's length uses an exponential of the length. A different approach to the problem can be to modify the weigths in the linear combination, so as to raise the importance of the terms with low probabilities. A simple way to achieve this is to deform, flatten, the original probability distribution so as to scale down its features with highest probabilities and raise the areas of lower probability. Of course, a very good candidate is the escort distribution, which leads us to the 'average length measure' $M_q = \sum_{i=1}^N P_i l_i$. For the virtual source with distribution P, the standard expected length is M_q, and the classical

Table 1. Source coding with generalized lengths, with $P_i = p_i^q / \sum_i p_i^q$ and where $\log_D$ denotes the base D logarithm

	length	bound	opt. length
Shannon	$\bar{L} = \sum_i p_i l_i$	$H_1(p)$	$l_i = -\log_D(p_i)$
Campbell	$C_\beta = \frac{1}{\beta} \log_D \sum_{i=1}^N p_i D^{\beta l_i}$	$H_{1/(\beta+1)}(p)$	$l_i = -\log_D(P_i)$
Gen. mean	$M_q = \sum_{i=1}^N P_i l_i$	$H_1(P)$	$l_i = -\log_D(P_i)$
Mixed	$L_q = \frac{1}{q-1} \log_D \left[\sum_{i=1}^N \frac{p_i^q}{\sum_j p_j^q} D^{(q-1)l_i} \right]$	$H_q(p)$	$l_i = -\log_D(p_i)$

Shannon noiseless source coding theorem immediately applies, leading to $M_q \geq H_1(P)$ with equality for $l_i = -\log_D P_i$, which is nothing else but Campbell's optimal lengths. The simple relation $l_i = -\log_D P_i$ for the minimization of M_q has an immediate application for the computation of Campbell's lengths: it suffices to feed a standard coding algorithm, namely a Huffman coder, with the escort distribution P instead of the natural distribution p.

Another measure of length, Table 1, line 4, can mix both a an exponential weight of individual lengths l_i and an escort distribution. This length of order q is lower bounded by the Rényi entropy of the same order. Interestingly, the optimum legths are nothing but the optimal lengths in the Shannon coding theorem.

FISHER, ESCORT AND q-GAUSSIANS

Let f denote the probability density of a random variable X. Then the Fisher information with respect to a translation parameter is defined as

$$I[f] = \int \left(\frac{d \ln f(x)}{dx} \right)^2 f(x) dx = \int \left(\frac{d f(x)}{dx} \right)^2 \frac{1}{f(x)} dx, \qquad (7)$$

where $f(x)$ is supposed differentiable and both $f(x)$ and its derivative $f'(x)$ are square integrable on $\mathbb{R}$.

The importance of the Fisher information as a measure of the information about a parameter in a distribution is well known. It has intricate relationships with maximum likelihood and has many implications in estimation theory, as exemplified by the Cramér-Rao bound which provides a fundamental lower bound on the variance of an estimator. It is used as a method of inference and understanding in statistical physics and biology, as promoted by Frieden [12]. Information theoretic inequalities involving Fisher information have attracted lot of attention for characterizing statistical systems through their localization in information planes, e.g. the Fisher-Shannon information plane [13, 14] or the Cramér-Rao information plane.

If we return to the problem of selecting the distribution $f(x)$ with a given q-variance, as indicated in subsection , we see that such distribution can be sought in a minimum Fisher sense. Actually, there are at least two possibilities that makes sense: (i) find $f(x)$ with a given variance such that its escort has minimum Fisher information, or (ii) find a

minimum Fisher information $f(x)$ such that its escort has a given variance, i.e. $f(x)$ has a given q-variance. Due to lack of space, we will only consider here the second variational problem:

$$\begin{cases} \inf_f I[f], \\ s.t. \quad \sigma_q^2 = \frac{\int x^2 f(x)^q dx}{\int f(x)^q dx}, \\ s.t. \quad f(x) \geq 0, \int f(x)dx = 1. \end{cases} \tag{8}$$

The solution to (i) can be simply deduced from the solution to (ii).

The main difficulty is that even though the Fisher information $I[f]$ is a strictly convex function, the set defined by the q-variance constraint is not convex and uniqueness of a solution is not guaranteed. It is convenient to introduce a new parameter and reformulate the awkward q-variance constraint into two simpler elementary constraints, that is $N_q = \int f(x)^q dx$ and $V_q = \int x^2 f(x)^q dx$, with, of course, $\sigma_q^2 = V_q/N_q$. Then, we can note that the initial problem is equivalent to the following two steps procedure:

$$\begin{cases} \inf_{N_q} \begin{cases} \inf_f I[f], \\ s.t. \quad N_q = \int f(x)^q dx \\ \text{and} \quad V_q = \sigma_q^2 N_q = \int x^2 f(x)^q dx \end{cases} \\ s.t. \quad \int f_{N_q}(x)dx = 1 \end{cases} \tag{9}$$

provided that the second problem is feasible with the parametric solution obtained in the first step. For now, we only consider the first step. Again and as in (8), although the objective functional $I[f]$ is strictly convex, since the constraints sets are not convex then the uniqueness of a minimum cannot be guaranteed. However, it is still possible to relax the equality constraints into inequalities, e.g. $\text{Sign}(q-1)\left(\int f(x)^q dx - N_q\right) \leq 0$ so that the problem becomes a convex optimization problem, whose solution is known to be unique. If the constraints are active at the solution, that is with equality, then we obtain the solution of the initial problem using the (relaxed) convex formulation.

Let a and b denote the Lagrange multipliers associated with the inequality constraints. These multipliers are necessarily nonnegative. Now, can check that with the following candidate

$$f(x) = u(x)^2 = \alpha^{\frac{2}{1-q}} \left(1 - (1-q)\beta x^2\right)^{\frac{2}{1-q}} \tag{10}$$

satisfies the Euler-Lagrange equation associated with the relaxed problem, with $a = 4\text{Sign}(q-1)\frac{\alpha^2 \beta}{q}$ and $b = -4\text{Sign}(q-1)\frac{\alpha^2 \beta^2(1-q)}{q}$. With $\beta > 0$, we check that these Lagrange multipliers are simultaneously positive for $q > 1$, but do not have the same sign for $q < 1$. This means that for $q > 1$, the q-gaussian (10) is a possible solution of the relaxed problem.

With a, b positive the Lagrangian is a convex function of f and of f' ($I[f]$ is a strictly convex function of f and that this is also true pointwise). In that case, any singular point corresponds to the global minimum [15, theorem 2.1]. Therefore, we obtain that the q-gaussian solves the relaxed minimization problem. Finally, it always possible to choose α so that the solution is normalized, then select β and adjust the parameter N_q so that the inequality constraints are active at the solution, *i.e.* are verified with equality. As a result, we finally obtain the global minimum f^* of the initial problem is obtained

with the q-gaussian (2) with $\gamma = 2$. This approach does not cover the case $q < 1$. In this case, we also obtain that a q-gaussian, with compact support, is the minimizer of the Fisher information, but the proof of the general case is more involved and omitted here. A direct consequence of this result is a generalization of Cramér-Rao inequality which readily follows from the inequality $I[f] \geq I[f^*]$ for all f with a given q-variance.

CONCLUSIONS AND FURTHER WORK

In this contribution, we have underlined the apparition of escort distributions in several domains. We have proposed a view of escort distributions as a trade-off between two distributions, in the context of displaced equilibriums or phase transition. When q varies, the escort distribution defines a path between two distributions, that shall be characterized; e.g. the thermodynamic length which is defined as the flux of Fisher information along the path. Associated with escort distributions, we recall that the q-gaussians are Rény-Tsallis Maxent distributions. This characterization is completed by the fact that q-gaussians are also minimum Fisher information distributions. Further work includes the study of the interaction between (generalized) Fisher information and Rényi-Tsallis entropy, as well as the study of the Fisher path, defined as the curve of minimum Fisher distributions when q varies.

REFERENCES

1. C. Tsallis, *Introduction to Nonextensive Statistical Mechanics*, 1st ed. Springer, Apr. 2009.
2. H. J. Hilhorst and G. Schehr, "A note on q-Gaussians and non-Gaussians in statistical mechanics," *Journal of Statistical Mechanics: Theory and Experiment*, vol. 2007, no. 06, pp. P06 003–P06 003, 2007.
3. C. Vignat and A. Plastino, "Why is the detection of q-Gaussian behavior such a common occurrence?" *Physica A*, vol. 388, no. 5, pp. 601–608, Mar. 2009.
4. J. F. Bercher, "An amended MaxEnt formulation for deriving Tsallis factors, and associated issues," *AIP Conference Proceedings*, vol. 872, no. 1, pp. 441–448, Nov. 2006. [Online]. Available: http://arxiv.org/abs/math-ph/0609077
5. R. H. Riedi, "Multifractal processes," *Theory and applications of long range dependence, eds. Doukhan, Oppenheim and Taqqu*, pp. 625–716, 2003.
6. D. Harte, *Multifractals: Theory and Applications*, 1st ed. Chapman & Hall/CRC, Jun. 2001.
7. J. F. Bercher, "Tsallis distribution as a standard maximum entropy solution with 'tail' constraint," *Physics Letters A*, vol. 372, no. 35, pp. 5657–5659, 2008.
8. A. Chhabra and R. V. Jensen, "Direct determination of the $f(\alpha)$ singularity spectrum," *Physical Review Letters*, vol. 62, no. 12, p. 1327, Mar. 1989.
9. C. Beck and F. Schloegl, *Thermodynamics of Chaotic Systems*. Cambridge University Press, 1993.
10. J. F. Bercher, "Source coding with escort distributions and Rényi entropy bounds," *Physics Letters A*, vol. 373, no. 36, pp. 3235–3238, Aug. 2009.
11. L. L. Campbell, "A coding theorem and Rényi's entropy," *Information and Control*, vol. 8, no. 4, pp. 423–429, 1965.
12. B. R. Frieden, *Science from Fisher Information: A Unification*. Cambridge University Press, 2004.
13. C. Vignat and J. F. Bercher, "Analysis of signals in the Fisher-Shannon information plane," *Physics Letters A*, vol. 312, no. 1-2, pp. 27–33, Jun. 2003.
14. E. Romera and J. S. Dehesa, "The Fisher-Shannon information plane, an electron correlation tool," *The Journal of Chemical Physics*, vol. 120, no. 19, pp. 8906–12, May 2004.
15. G. M. Ewing, "Sufficient conditions for global minima of suitably convex functionals from variational and control theory," *SIAM Review*, vol. 19, no. 2, pp. 202–220, Apr. 1977.