Levy distributions and the maximization of Rényi-Tsallis entropy

J.-F. BERCHER
Laboratoire Signaux et Télécoms ESIEE
and Laboratoire Systèmes de Communications,
Université de Marne la Vallée
93 162 Noisy-le-Grand, FRANCE
bercherj@esiee.fr

C. VIGNAT
Équipe Signal, Institut Gaspard Monge
Université de Marne la Vallée
93 166 Noisy-le-Grand, FRANCE
vignat@univ-mlv.fr

The Levy distribution is extensively used in the field of *non-extensive statistical mechanics* where it succeeds in modelling power-law phenomena, long-range interactions, long-range memory or space-time multifractal structure. Applications include fully developed turbulence, Levy anomalous diffusion, statistics of cosmic rays... There are several indications that non-extensive statistical mechanics results are physically relevant for partially equilibrated or nonequilibrated systems, with a stationary state that is characterized by fluctuations of an intensive parameter.

We show that the Levy distribution can be derived as the minimizer of the Kullback-Leibler information divergence with respect to a reference distribution, where the minimization is carried under a mean log-likelihood constraint and a (mean) observation constraint. Here, the mean log-likelihood constraint characterizes the 'displacement' from the conventional equilibrium.

We examine two scenario for the mean observation constraint, that lead to two Levy distributions with opposite exponents and involve naturally the so-called 'escort' or 'zooming' distributions. We derive two unimodal modified dual functions, whose maximization enable to find the index of the Levy distribution.

The minimum of the Kullback-Leibler information divergence under our constraints is a level-one entropy. In some special cases for the reference distribution, we derive the expression of the partition function, and recover some well-known entropy functionals in the classical Gibbs-Boltzmann-Shannon limit.

An amended MaxEnt formulation for displaced equilibriums

The original maxent formulation "find the closest distribution to a reference under a mean constraint" may be amended by introducing for instance a new constraint that will displace the equilibrium. The partial or displaced equilibrium may be imagined as an equilibrium characterized by two references, say P_1 and Q. Instead of selecting the nearest distribution to a reference under a mean constraint, we may look for a distribution P^* close simultaneously to two distinct references: such a distribution will be localized somewhere 'between' the two references P_1 and Q. For instance, we may consider two systems in contact characterized by two prior reference distributions. The global equilibrium is attained for some intermediate distribution

Find P^* such that the

$$\left\{ \begin{array}{c} \min_P \; D(P||Q) = \min_P \; \int P(x) \log \frac{P(x)}{Q(x)} dx \\ s.t \;\; \theta = D(P||Q) - D(P||P_1) = \int P(x) \log \frac{P_1(x)}{Q(x)} dx \end{array} \right. \tag{1}$$

where θ is some constant that tune the equilibrium between P_1 and Q.

Solution is the *escort* or *zooming* distribution

$$P^*(x) = \frac{P_1(x)^\alpha Q(x)^{1-\alpha}}{\int P_1(x)^\alpha Q(x)^{1-\alpha} dx}, \tag{2}$$

under observable mean values (constraints)

In order to take into account an observable as a mean value under some distribution, we may adjust one of the reference distribution. Suppose that the subsystem with distribution P_1 is isolated, and that one can have access to the mean under P_1. Then what shall be the distribution P_1 with given mean such that the equilibrium distribution P^* remains simultaneously close to P_1 and Q ? But the observable may also be measured as the mean under the equilibrium distribution P^*, that triggers the new question: what shall be the distribution P_1 such that P^* possess a known mean and remains simultaneously close to P_1 and Q ?

$$\mathcal{F}(m) = \left\{ \begin{array}{l} \min_{P_1} \left\{ \begin{array}{c} \min_P \; D(P||Q) = \min_P \; \int P(x) \log \frac{P(x)}{Q(x)} dx \\ \text{subject to: } \theta = \int P(x) \log \frac{P_1(x)}{Q(x)} dx \end{array} \right. \\ \quad \text{subject to: } m = E_{P_1}[X] \text{ or } m = E_{P^*}[X] \end{array} \right. = \sup_\alpha \left[\alpha\theta - \left\{ \begin{array}{c} \max_{P_1} (\alpha - 1) D_\alpha(P_1||Q) \\ \text{subject to: } m = E_{P_1}[X] \text{ or } m = E_{P^*}[X] \end{array} \right. \right]$$

where $E_P[X]$ represents the statistical mean under distribution $P: E_P[X] = \int xP(x)dx$. $E_{P^*}[X]$ is nothing else but the Tsallis-Mendes-Plastino (TMP) generalized mean constraint, the 'α-expectation' of non-extensive thermodynamics.

The right hand side is obtained by dual attainment, provided the solution exists, by computing the RHS in two steps: first minimize with respect to P taking into account the mean log-likelihood constraint, and obtain (2), and secondly, minimize with respect to P_1. Hence the whole problem *amounts to the extremization of Rényi information divergence under a mean constraint.*

The 'contracted' Rényi information divergence or of Kullback-Leibler information divergence for a given mean m and a log-likelihood constraint θ, functionals $F_\alpha^{(1)}(m)$ and $F_\alpha^{(\alpha)}(m)$ are level-one entropy, positive and strictly convex for $\alpha \in [0,1]$:

$$F_\alpha^{(1)}(m) \ \left(\text{resp. } F_\alpha^{(\alpha)}(m)\right) = \begin{cases} \min_{P_1} (1-\alpha) D_\alpha(P_1||Q) \\ \text{subject to: } m = E_{P_1}[X] \ \ (\text{ resp. } m = E_{P^*}[X]) \end{cases}, \tag{3}$$

In the two cases, the general solution is given (implicitely) by

$$\boxed{P_1(x) = [\gamma(x - x_{\gamma,\alpha}) + 1]_+^\nu \, Q(x) e^{D_\alpha(P_1||Q)}.} \tag{4}$$

where

(i) $\nu = \xi = \frac{1}{\alpha} - 1$ and $x_{\lambda,\alpha}$ is the statistical mean under P_1, for the classical mean constraint,

(ii) $\nu = -\xi = \frac{1}{1-\alpha}$ and $x_{\lambda,\alpha}$ is the generalized α-expectation under P_1 for the generalized mean constraint.

Parameter γ is the Lagrange parameter such that the constraint is verified. This distribution is self-referential because it is implicitely defined by its mean and Rényi divergence to Q.

Dual functions

The corresponding dual functions are given by

$$D(\gamma) = (1-\alpha) D_\alpha(P||Q) + \lambda(m - x_{\gamma,\alpha})$$

with

(i) $D_\alpha(P_1||Q) = -\log\left(Z_{\xi+1}(\gamma, x_{\gamma,\alpha})\right)$ and $x_{\gamma,\alpha}$ is the statistical mean under P_1, for the classical mean constraint,

(ii) $D_\alpha(P_1||Q) = -\log\left(Z_{-\xi}(\gamma, x_{\gamma,\alpha})\right)$ and $x_{\gamma,\alpha}$ is the generalized α-expectation under P_1 for the generalized mean constraint,

and where we used the *partition function*

$$Z_\nu(\gamma, x_{\gamma,\alpha}) = \int [\gamma(x - x_{\gamma,\alpha}) + 1]_+^\nu \, Q(x) dx$$

But again such an expression is not exploitable in practice, because the distribution, and its partition function, are implicitely defined. Thus we have to look after some 'alternate' dual function, that is computable and provides the same optimum solution as the original dual function.

Two alternate 'computable' dual functions, with the same maximum as the original one can be found:

$$D^\#(\lambda) = \left(\frac{1}{\xi}\right) \log Z_{\xi+1}(\lambda, m), \quad \text{or} \quad D^\#(\lambda) = \left(\frac{1}{\xi}\right) \log Z_{-\xi}(\lambda, m),$$

for the classical mean and for the generalized mean respectively.

A generalized thermodynamics

Define an entropy

$$S = -\frac{1}{\nu+1} \log Z_{\nu+1}(\gamma, U),$$

and a functional $\Phi(\gamma, U) = S - \gamma U$ (or the free energy $F = U - \frac{1}{\gamma} S$). From general properties of Levy distribution,

$$E_\nu[x - m] = \frac{1}{\nu+1} \frac{\frac{dZ_{\nu+1}(\gamma, U)}{d\gamma}}{Z_\nu(\gamma, U)}$$

and since $Z_{\nu+1}(\gamma^*, U) = Z_\nu(\gamma^*, U)$ for the optimum γ^*, we have

$$\frac{dS}{d\gamma} = 0 \quad \frac{dS}{dU} = -\gamma \quad \frac{d\Phi((\gamma, U))}{d\gamma} = -U \quad \frac{d\Phi((\gamma, U))}{dU} = 0$$

that is the whole Legendre structure of thermodynamics.

Some special cases of entropy functionals

Partition functions are computable for particular cases of references measures. Then one can look for $\gamma^*(m)$ such that $dD^{\#}(\gamma^*) = 0$ and then express $F_\alpha^{(1)}(m)$ resp. $F_\alpha^{(\alpha)}(m)$ since they are equal to $D^{\#}(\gamma^*)$ at the optimum.

For $\alpha \to 1$, we obtain

1. $F_\xi(x) = x \log x + (1-x)\log(1-x) - \log(2)$ for Bernoulli reference (Fermi-Dirac entropy),
2. $F_\xi(x) = \frac{1}{\xi}\left(-x\beta - \ln\left(-x\beta\right) - \ln\left(-\beta\right) - 1\right)$ for the exponential reference (Itakura-Saito or Burg entropy)
3. $F_\xi(x) = \frac{1}{\xi}\left(x \ln \frac{x}{\mu} + (\mu - x)\right)$ for Poisson reference (Shannon entropy).

[Jay82, Ell85, GG04, BS02, Kul59]

References

[BS02] A.G. Bashkirov and A.D. Sukhanov. The distribution function for a subsystem experiencing temperature fluctuations. *Journal of Experimental and Theoretical Physics*, 95(3):440 – 446, September 2002.

[Ell85] Richard S. Ellis. *Entropy, Large Deviations, and Statistical Mechanics*, volume 271 of *Grundlehren der mathematischen Wissenschaften*. Springer-Verlag, 1985.

[GG04] Jr. Grendar, M. and M. Grendar. Maximum entropy method with non-linear moment constraints: challenges. In G. Erickson (ed.), editor, *Bayesian inference and maximum entropy methods in science and engineering*. AIP, 2004.

[Jay82] E.T. Jaynes. On the rationale of maximum entropy methods. *Proc. IEEE*, 70:939–952, 1982.

[Kul59] S. Kullback. *Information Theory and Statistics*. Wiley, New York, 1959.