An amended MaxEnt formulation for deriving Tsallis factors, and associated issues

Jean-François Bercher

Équipe Signal et Information,
Dept. Modélisation et Simulation
ESIEE, Noisy-le-Grand, France

Abstract.
An amended MaxEnt formulation for systems displaced from the conventional MaxEnt equilibrium is proposed. This formulation involves the minimization of the Kullback-Leibler divergence to a reference Q (or maximization of Shannon Q-entropy), subject to a constraint that implicates a second reference distribution P_1 and tunes the new equilibrium. In this setting, the equilibrium distribution is the generalized escort distribution associated to P_1 and Q. The account of an additional constraint, an observable given by a statistical mean, leads to the maximization of Rényi/Tsallis Q-entropy subject to that constraint. Two natural scenarii for this observation constraint are considered, and the classical and generalized constraint of nonextensive statistics are recovered. The solutions to the maximization of Rényi Q-entropy subject to the two types of constraints are derived. These optimum distributions, that are Levy-like distributions, are self-referential. We then propose two 'alternate' (but effectively computable) dual functions, whose maximizations enable to identify the optimum parameters. Finally, a duality between solutions and the underlying Legendre structure are presented.

Key Words: Rényi entropy, Levy distributions, optimization, nonextensive thermodynamics, duality

INTRODUCTION

The formalism of nonextensive statistical mechanics [1, 2] leads to a generalized Boltzmann factor in the form of a Tsallis distribution (or factor) that depends on an entropic index and recovers the classical Boltzmann factor as a special limit case [1]. This distribution is of high interest in many physical systems since it enables to model power-law phenomena. In a wide variety of fields, experiments, numerical results and analytical derivations fairly agree with the description by a Tsallis distribution.

Tsallis' distributions (sometimes called Levy distributions) are derived by maximization of Tsallis entropy [3], under suitable constraints. The present formulation is as follows: maximize Tsallis' entropy

$$T_\alpha(P) = \frac{1}{1-\alpha}\left[\int P(x)^\alpha dx - 1\right], \tag{1}$$

[0] This work was presented at MaxEnt2006, the 26^{th} International Workshop on Bayesian Inference and Maximum Entropy Methods in Science and Engineering CNRS, Paris, France, July 8-13, 2006.

subject to

$$m = \int x P^*(x) dx \text{ with } P^*(x) = \frac{P(x)^\alpha}{\int P(x)^\alpha dx}, \tag{2}$$

where the mean constraint is called a 'generalized' mean constraint in the nonextensive litterature, and $P^*(x)$ is called the 'escort' distribution. This formulation was preferred to the simple maximization with a classical mean constraint $m = \int x P(x) dx$ because of mathematical difficulties. The solution is given in the litterature as

$$P(x) = \frac{1}{Z} \left(1 - \frac{(1-\alpha)\beta}{Z^{1-\alpha}} (x - m) \right)^{\frac{1}{1-\alpha}}, \tag{3}$$

where Z is a partition function.

Of course, these distributions do not coincide with those derived by conventionnal MaxEnt and consequently will not be justified from a probabilistic point of view, because of the uniqueness of the rate function in the large deviations theory [4, 5]. Furthermore, the status and interest of generalized expectations and of escort distributions is unclear. Last, it is apparent that the expression of distribution (3) is implicit, so that both its manipulation and determination of its parameter β will be difficult.

However, in view of the success of nonextensive statistics, there should exist a probabilistic setting that provides a justification for the maximization of Tsallis entropy. There are now several indications that results of nonextensive statistics are physically relevant for partially equilibrated or nonequilibrated systems, with a stationary state characterized by fluctuations of an intensive parameter [6, 7]; for instance, the Tsallis factor is obtained from the Boltzmann-Gibbs' if the inverse of temperature fluctuates according to a gamma distribution.

In this paper, I present a framework for the maximization of Rényi/Tsallis $Q-$entropy, that leads to the so-called Levy distribution (or Tsallis factor). The Rényi information divergence, the opposite of Rényi Q-entropy, is given by

$$D_\alpha(P||Q) = \frac{1}{\alpha - 1} \log \int P(x)^\alpha Q(x)^{1-\alpha} dx, \tag{4}$$

where α is a real parameter called the entropic index. Using L'Hospital's rule, the Kullback-Leibler divergence is recovered for $\alpha \to 1$

$$D(P||Q) = \int P(x) \log \frac{P(x)}{Q(x)} dx. \tag{5}$$

Its opposite is the Shannon $Q-$entropy, the correct, coordinate invariant, extension of the classical Shannon entropy to the continuous case [8]. This divergence can be interpreted as a "distance" between two distributions. Rényi and Tsallis Q-entropies are related by a simple monotonic function. Therefore, their maximization under the same constraint lead to the same distribution.

In the following, I propose an amended MaxEnt formulation for systems with a displaced equilibrium, find that the relevant entropy in this setting is the Rényi entropy, interpret the mean constraints, derive the correct form of solutions, propose numerical procedures for estimating the parameters of the Tsallis factor and characterize the associated entropies. I will also indicate a duality between the solutions associated with classical and generalized mean constraint. Finally I will discuss the underlying Legendre structure of generalized thermodynamics associated to this setting.

THE AMENDED MAXENT FORMULATION

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 for instance a new constraint that displaces 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^* simultaneously close to two distinct references: such a distribution will be localized somewhere 'between' the two references P_1 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 $P_1,$ and B with distribution $Q,$ and the whole system is viewed with distribution P^* that is some intermediate between P_1 and Q. This can also model a phase transition: a system leaves a state Q toward P_1 and presents an intermediate distribution P^*.

This can be stated as: find P^* such that the Kullback-Leibler divergence to Q, $D(P||Q)$ is minimum (or equivalently the Shannon Q-entropy is maximum), but under the constraint that $D(P||Q) = D(P||P_1) + \theta,$ where θ can be expressed as a log-likelihood. The problem simply writes

$$\begin{cases} \min_P \; D(P||Q) = \min_P \; \int P(x) \log \frac{P(x)}{Q(x)} dx \\ s.t \;\; \theta = D(P||Q) - D(P||P_1) = \int P(x) \log \frac{P_1(x)}{Q(x)} dx \end{cases} \tag{6}$$

and its solution was given by Kullback [9, page 39] as an illustration of his general theorem on constrained minimization of $D(P||Q)$:

$$P^*(x) = \frac{P_1(x)^\alpha Q(x)^{1-\alpha}}{\int P_1(x)^\alpha Q(x)^{1-\alpha} dx}, \tag{7}$$

which is nothing else but the escort distribution (2) of nonextensive statistics [10] (although it is generalized here with reference Q). The parameter α is simply the Lagrange parameter associated to the constraint, and it can be shown that necessarily $\alpha \leq 1$. Clearly, distribution P^* which is the geometric mean between $P_1$and Q realizes a trade-off, governed by $\alpha,$ between the two references. By dual attainment, we have

$$\left\{\begin{array}{c}\min_P\ D(P||Q)\\ s.t\ \ \theta=D(P||Q)-D(P||P_1)\end{array}\right. =\sup_\alpha\left(\alpha\theta-\log\left(\int P_1(x)^\alpha Q(x)^{1-\alpha}dx\right)\right). \quad (8)$$

In this last relation, the term $\log\left(\int P_1(x)^\alpha Q(x)^{1-\alpha}dx\right)$ is directly proportional to the Rényi divergence (4).

Observable mean values

Observable values are as usual the statistical mean under some distributions. Depending on the viewpoint, the observable may be a mean under distribution P_1, the distribution of an isolated subsystem, or under P^*, the equilibrium distribution between P and Q. Hence, the problem will be completed by an additionnal constraint, and a possible approach would be to select distribution P_1 by further minimizing the Kullback-Leibler information divergence $D(P||Q)$, but over $P_1(x)$ and subject to the mean constraint. So, the whole problem writes

$$K=\left\{\begin{array}{c}\min_{P_1}\left\{\begin{array}{c}\min_P\ D(P||Q)=\min_P\ \int P(x)\log\frac{P(x)}{Q(x)}dx\\ \text{subject to: }\ \theta=\int P(x)\log\frac{P_1(x)}{Q(x)}dx\end{array}\right.\\ \text{subject to: } m=E_{P_1}[X] \text{ or } m=E_{P^*}[X]\end{array}\right., \quad (9)$$

where $E_P[X]$ represents the statistical mean under distribution $P: E_P[X]=\int xP(x)dx$. This may be tackled in two steps: first minimize with respect to P taking into account the mean log-likelihood constraint, and obtain (7), and second, minimize with respect to P_1. Taking into account (8), problem (9) becomes

$$K=\sup_\alpha\left[\alpha\theta-\left\{\begin{array}{c}\max_{P_1}(\alpha-1)D_\alpha(P_1||Q)\\ \text{subject to: } m=E_{P_1}[X] \text{ or } m=E_{P^*}[X]\end{array}\right.\right] \quad (10)$$

and *amounts to the extremization of Rényi information divergence under a mean constraint*. Therefore, we find that the amended MaxEnt formulation leads to the maximization of Rényi (or equivalently Tsallis) entropy subject to a statistical mean constraint. We can note that the second constraint, $m=E_{P^*}[X]$ is nothing else but the 'generalized expectation' of nonextensive statistics that has here a clear interpretation.

It is important to note that the minimization of Kullback-Leibler divergence with respect to P and P_1, subject to the two constraints, may not always reduce to the two-steps procedure above.

SOLUTIONS TO THE MAXIMIZATION OF RÉNYI Q-ENTROPY

We now consider the maximization of Rényi Q-entropy subject to the classical mean constraint (C) $m=E_{P_1}[X]$ and the generalized mean constraint (G) $m=E_{P^*}[X]$ as we obtained in (10). We first begin by some results on a general 'Tsallis' distribution, that simplify the derivation of exact solutions (proofs are omitted to save space).

Preliminary results

Definition 1 *Distribution $P_\nu^\#(x)$ is defined by:*

$$P_\nu^\#(x) = [\gamma(x-\overline{x})+1]^\nu Q(x) e^{D_\alpha(P_\nu^\# || Q)}, \tag{11}$$

on domain $\mathcal{D} = \mathcal{D}_Q \cap \mathcal{D}_\gamma$, where $\mathcal{D}_Q = \{x : Q(x) \geq 0\}$ and $\mathcal{D}_\gamma = \{x : \gamma(x-\overline{x})+1 \geq 0\}$. In this expression, $\overline{x}$ is either (a) a fixed parameter, say m, and $P_\nu^\#(x)$ is a two parameters distribution, (b) or some statistical mean with respect to $P_\nu^\#(x)$, e.g. its "classical" or "generalized" mean, and as such a function of γ. Observe that distribution $P_\nu^\#(x)$ is not necessarily normalized to one. Associated with $P_\nu^\#(x)$, we also define a partition function

$$Z_\nu(\gamma, \overline{x}) = \int_{\mathcal{D}} [\gamma(x-\overline{x})+1]^\nu Q(x) dx. \tag{12}$$

Notation 2 *We will denote by $E_\nu[X]$ the statistical mean with respect to the probability distribution associated with $P_\nu^\#(x)$, and by $E_\nu^{(\alpha)}[X]$ the generalized $\alpha-$mean. One can observe that in the case of the Levy distribution (11), we have $E_\nu^{(\alpha)}[X] = E_{\alpha\nu}[X]$. In the special case $\nu = \pm\xi$, we obtain $E_{\pm\xi}^{(\alpha)}[X] = E_{\pm(\xi+1)}[X]$, because $\xi\alpha = (\xi+1) = \frac{\alpha}{\alpha-1}$.*

Theorem 3 *The Levy distribution $P_\xi^\#(x)$ with exponent $\nu = \xi$, is normalized to one if and only if $\overline{x} = E_\xi[x]$, the statistical mean of the distribution, and $D_\alpha(P_\xi^\# || Q) = -\log Z_{\xi+1}(\gamma, \overline{x}) = -\log Z_\xi(\gamma, \overline{x})$.*

In the same way, the Levy distribution $P_{-\xi}^\#(x)$ with exponent $\nu = -\xi$, is normalized to one if and only if $\overline{x} = E_{-\xi-1}[x] = E_{-\xi}^{(\alpha)}[x]$, the generalized $\alpha-$expectation of the distribution, and $D_\alpha(P_{-\xi}^\# || Q) = -\log Z_{-(\xi+1)}(\gamma, \overline{x}) = -\log Z_{-\xi}(\gamma, \overline{x})$, with $\alpha\xi = (\xi+1)$.

When $\overline{x}$ is a fixed parameter m, this will be only true for a special value γ^* of γ such that $E_\xi[x] = m$ or $E_{-\xi}^{(\alpha)}[x] = m$, respectively in the first and second case.

Remark 4 *Here takes place an important remark on* the mapping $\overline{x} \leftrightarrow \gamma$. *Consider the normalized distribution $P_\xi^\#(x)$ with $\overline{x} = E_\xi[x]$. This distribution depends on the sole parameter γ, and $\overline{x}$ is a function of γ. But contrary to the intuition,* the mapping $\overline{x} \leftrightarrow \gamma$ is not necessarily one to one. *This means that a specified value of the mean $\overline{x} = m$ may correspond to several values of γ, and conversely a specified value of γ may give several different means $\overline{x}$. This can be illustrated through numerical examples.*

Lemma 5 *Partition functions $Z_{\xi+1}(\gamma, m)$ and $Z_{-\xi}(\gamma, m)$ are convex functions of γ.*

Solutions

The solutions to the maximization of Rényi Q-entropy subject to the classical mean constraint (C) $m = E_{P_1}[X]$ and the generalized mean constraint (G) $m = E_{P^*}[X]$ are found using standard Lagrangian techniques The optimum solution, see for instance [11], is a saddle point of the Lagrangian and we may proceed in two steps: first minimize the Lagrangian in $P(x)$, and thus obtain a solution in terms of the Lagrange parameters, and then maximize the resulting Lagrangian, the dual function, in order to exhibit the optimum Lagrange parameters. Taking into account the normalization conditions described above, these solutions are easily derived and simplified:

$$(C) \quad P_C(x) = \frac{[\gamma(x-\overline{x})+1]^{\xi}}{Z_{\xi}(\gamma,\overline{x})} Q(x), \text{ with } \overline{x} = E_{P_C}[X] = E_{\xi}[X] \tag{13}$$

$$(G) \quad P_G(x) = \frac{(1+\gamma(x-\overline{x}))^{-\xi}}{Z_{-\xi}(\gamma,\overline{x})} Q(x) \text{ with } \overline{x} = E_{P_G}[X] = E_{-(\xi+1)}[X] \tag{14}$$

where $\xi = \frac{1}{\alpha-1}$, and $Z_{\nu}(\gamma,\overline{x})$ is the partition function. It is important to emphasize that $\overline{x}$ in (13) is the statistical mean with respect to $P_C(x)$, $\overline{x}$ in (14) is the generalized α-mean with respect to $P_G(x)$, and as such a function of γ. It is a common mistake in the large majority of reported results and calculations to improperly take for $\overline{x}$ the fixed value m of the constaint, which is only correct for the optimum value of the Lagrange parameter.

These optimum distributions appear to be self-referential, since their expressions involve their statistical mean. Therefore, the direct determination of their parameters is difficult, if not intractable.

Alternate dual functions

From the Lagrangian theory, one should maximize the dual function in order to obtain the remaining Lagrange parameter. But in the present cases, the dual functions are implicitely defined. Thus, in order to identify the value of the natural parameter associated to the mean constraints, I propose two 'alternate' (but effectively computable) dual functions, whose numerical maximizations enable to exhibit the optimum parameters.

For the classical mean, I just sketch the procedure. At the optimum, we have $D(\gamma^*) = \sup_{\gamma}\sup_{\mu}\inf_P L(P,\gamma,\mu)$. For any value $\widetilde{\mu}$ of μ, letting $\widetilde{D}(\gamma) = L(P^*_{\gamma,\widetilde{\mu}},\gamma,\widetilde{\mu})$, we have $D(\gamma^*) \geq \widetilde{D}(\gamma)$. Thus, if $\widetilde{D}(\gamma^*) = D(\gamma^*)$ for the optimum γ^*, then $\widetilde{D}(\gamma^*)$ will be a maximum of $\widetilde{D}(\gamma)$ and the maximization of the dual function can be carried equivalently via the maximization of $\widetilde{D}(\gamma)$. Condition $\widetilde{D}(\gamma^*) = D(\gamma^*)$ is achieved with $\widetilde{\mu}(\gamma) = -(\xi+1)(1-\gamma m)$. Then, after some algebra, we obtain the very simple form

$$\widetilde{D}_C(\gamma) = -\log Z_{\xi+1}(\gamma,m) \tag{15}$$

that is simply the expression of the divergence from $P_{\xi}^{\#}$ to Q, $D_{\alpha}(P_{\xi}^{\#}||Q)$. We know that $Z_{\xi+1}(\gamma,m)$ is a convex function. Thus, if $Z_{\xi+1}(\gamma,m)$ is defined on a continuous domain, $\widetilde{D}_C(\gamma)$ has an only maximum for $\gamma=\gamma^*$. If $Z_{\xi+1}(\gamma,m)$ is defined (and convex) on several intervals, $\widetilde{D}_C(\gamma)$ may have a maximum on each of these intervals, and one has to select the minimum of these maxima (that is the maximum associated with the minimum divergence). Hence, the identification of the optimum parameter γ^* simply amounts to the unconstrained maximization of an unimodal functional, possibly in several intervals.

For the generalized mean, the rationale for an alternate dual function is as follows. We know that $D_\alpha(P^{\#}_{-\xi}||Q) = -\log Z_{-\xi}(\gamma, m)$ when the generalized mean constraint is satisfied. Since $\frac{d\log Z_{-\xi}(\gamma,m)}{d\gamma} = -\xi\,(\overline{x}-m)\,\frac{Z_{-\xi-1}(\gamma,m)}{Z_{-\xi}(\gamma,m)},\ -\log Z_{-\xi}(\gamma,m)$ is maximum when the constraint $\overline{x} = m$ is satisfied. Hence, the search of the optimum Lagrange parameter can be carried using the very simple alternate dual function

$$\widetilde{D}_G(\gamma) = -\log Z_{-\xi}(\gamma, m). \tag{16}$$

The partition function $Z_{-\xi}(\gamma, m)$ is a convex function for $\alpha \leq 1$. If it is defined on a continuous domain, $\widetilde{D}_G(\gamma)$ has an only maximum that is simply reached for γ^* such that $m = E_{-\xi-1}[x]$, the generalized α-mean. If the domain is given by several intervals, then $\widetilde{D}_G(\gamma)$ may present several maxima, and the minimum of these maxima, associated with the minimum divergence $D_\alpha(P^{\#}_{-\xi}||Q)$, has to be selected. We thus obtain two practical numerical schemes for the identification of the distributions parameters, and it is also possible to study the behaviour of entropies associated with some particular references Q. We come to a close to this presentation by considering the relationship between the two minimization problems and an underlying Legendre structure.

DUALITY AND LEGENDRE STRUCTURE

The $\alpha \leftrightarrow 1/\alpha$ duality

The dual functions associated to the two problems are $-\log Z_{\xi_1+1}(\gamma, m)$ and $-\log Z_{-\xi_2}(\gamma, m)$. Thus, we will have pointwise equality of dual functions, and of course of the optima, if $\xi_1 + 1 = -\xi_2$, that is if indexes α_1 and α_2 satisfy $\alpha_1 = 1/\alpha_2$. We can also remark that with $-\xi_2 = \xi_1 + 1 = \alpha_1\xi_1$, we have the following relations between the two optimum probability density functions:

$$P_G = \frac{P_C^{\alpha_1} Q^{1-\alpha_1}}{Z_{\xi_1}^{1-\alpha_1}} \ \text{ and } \ P_C = \frac{P_G^{\alpha_2} Q^{1-\alpha_2}}{Z_{\xi_1}^{1-\alpha_2}}, \ \text{ with } \ \alpha_2 = 1/\alpha_1, \tag{17}$$

and using the fact that $Z_{\xi_1+1}(\gamma, m) = Z_{\xi_1}(\gamma, m)$ for the optimum value of γ. It means that P_G is the escort distribution of P_C with index α_1 and that P_C is the escort distribution associated with P_G and index α_2. It can be checked in the general case that *always* have the equality $D_{\frac{1}{\alpha}}(P^*||Q) = D_\alpha(P_1||Q)$ between the $1/\alpha$ Rényi divergence of the escort distribution to Q and the standard α divergence Hence, the minimization of the α Rényi divergence subject to the generalized mean constraint is exactly equivalent to the minimization of the $1/\alpha$ Rényi divergence subject to the classical mean constraint so that generalized and classical mean constraints can always be swapped, provided the index α is changed into $1/\alpha$, as was argued in [12, 13].

The Legendre structure

In the study of alternative entropies, considerable efforts have been directed to the analysis of associated thermodynamics. The concave entropies corresponding to our two problems are $S_C = \log Z_{\xi+1}(-\frac{\lambda}{(\xi+1)}, \overline{x})$, and $S_G = \log Z_{-\xi}(\lambda/\xi, \overline{x})$. Let us consider the general form $S = \log Z_{\nu+1}(\gamma, \overline{x})$.

In terms of the Lagrange multiplier λ, it can be shown that

$$\frac{dS}{d\lambda} = \frac{dS}{d\gamma}\frac{d\gamma}{d\lambda} = -\gamma\,(\nu+1)\,\frac{d\overline{x}}{d\lambda}. \tag{18}$$

Specializing the result to the two entropies, we obtain in both cases the Euler formula:

$$\frac{dS}{d\lambda} = \lambda\frac{d\overline{x}}{d\lambda}. \tag{19}$$

Next, the derivative of the entropy with respect to the mean is simply

$$\frac{dS}{d\overline{x}} = \frac{dS}{d\lambda}\frac{d\lambda}{d\overline{x}} = \lambda\frac{d\overline{x}}{d\lambda}\frac{d\lambda}{d\overline{x}} = \lambda. \tag{20}$$

Let us now introduce the Massieu potential $\phi(\lambda) = S - \lambda\overline{x}$ (or equivalently the free energy). Derivations with respect to the Lagrange parameter and to the mean give

$$\frac{d\phi}{d\lambda} = -\overline{x}, \text{ and } \frac{d\phi}{d\overline{x}} = -\overline{x}\frac{d\lambda}{d\overline{x}}. \tag{21}$$

These four relations show that S and ϕ are conjugated with variables $\overline{x}$ and $\lambda : S\,[\overline{x}] \rightleftharpoons \phi\,[\lambda]$, so that the basic Legendre structure of thermodynamics is preserved (but care must be taken for interpretations, for instance a valid definition of temperature requires that λ always remains positive).

REFERENCES

1. C. Tsallis, "Nonextensive statistics: Theoretical, experimental and computational evidences and connections," *Brazilian Journal of Physics*, vol. 29, pp. 1–35, March 1999.
2. C. Tsallis, "Entropic nonextensivity: a possible measure of complexity," *Chaos, Solitons,& Fractals*, vol. 13, pp. 371–391, 2002.
3. C. Tsallis, "Possible generalization of Boltzmann-Gibbs statistics," *Journal of Statistical Physics*, vol. 52, no. 1-2, pp. 479–487, 1988.
4. B. R. La Cour and W. C. Schieve, "Tsallis maximum entropy principle and the law of large numbers," *Phys. Rev. E*, vol. 62, pp. 7494 – 7496, November 2000.
5. J. Grendar, M. and M. Grendar, "Maximum entropy method with non-linear moment constraints: challenges," in *Bayesian inference and maximum entropy methods in science and engineering* (G. Erickson, ed.), AIP, 2004.
6. C. Beck, "Generalized statistical mechanics of cosmic rays," *Physica A*, vol. 331, pp. 173–181, january 2004.
7. G. Wilk and Z. Wodarczyk, "Interpretation of the nonextensitivity parameter q in some applications of Tsallis statistics and Lévy distributions," *Physical Review Letters*, vol. 84, pp. 2770–2773, March 2000.

8. E. T. Jaynes, *Statistical Physics*, ch. Information Theory and Statistical Mechanics, pp. 181–218. Benjamin, New York, 1963.
9. S. Kullback, *Information Theory and Statistics*. Wiley, New York, 1959.
10. C. Beck and F. Schloegl, *Thermodynamics of Chaotic Systems*. Cambridge University Press, 1993.
11. S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge University Press, 1st ed., March 2004. ISBN: 0521833787.
12. G. A. Raggio, “On equivalence of thermostatistical formalisms.” http://arxiv.org/abs/cond-mat/9909161, 1999.
13. J. Naudts, “Dual description of nonextensive ensembles,” *Chaos, Solitons, and Fractals*, vol. 13, no. 3, pp. 445–450, 2002.