Maximum entropy approach to H-theory: Statistical mechanics of hierarchical systems

Giovani L. Vasconcelos,[1,*] Domingos S. P. Salazar,[2] and A. M. S. Macêdo[1]

[1] *Laboratório de Física Teórica e Computacional, Departamento de Física, Universidade Federal de Pernambuco 50670-901 Recife, Pernambuco, Brazil*

[2] *Unidade de Educação a Distância e Tecnologia, Universidade Federal Rural de Pernambuco, 52171-900 Recife, PE, Brazil*

(Received 29 June 2017; published 5 February 2018)

A formalism, called H-theory, is applied to the problem of statistical equilibrium of a hierarchical complex system with multiple time and length scales. In this approach, the system is formally treated as being composed of a small subsystem—representing the region where the measurements are made—in contact with a set of "nested heat reservoirs" corresponding to the hierarchical structure of the system, where the temperatures of the reservoirs are allowed to fluctuate owing to the complex interactions between degrees of freedom at different scales. The probability distribution function (pdf) of the temperature of the reservoir at a given scale, conditioned on the temperature of the reservoir at the next largest scale in the hierarchy, is determined from a maximum entropy principle subject to appropriate constraints that describe the thermal equilibrium properties of the system. The marginal temperature distribution of the innermost reservoir is obtained by integrating over the conditional distributions of all larger scales, and the resulting pdf is written in analytical form in terms of certain special transcendental functions, known as the Fox H functions. The distribution of states of the small subsystem is then computed by averaging the quasiequilibrium Boltzmann distribution over the temperature of the innermost reservoir. This distribution can also be written in terms of H functions. The general family of distributions reported here recovers, as particular cases, the stationary distributions recently obtained by Macêdo *et al.* [Phys. Rev. E **95**, 032315 (2017)] from a stochastic dynamical approach to the problem.

DOI: 10.1103/PhysRevE.97.022104

I. INTRODUCTION

Complex systems with multiple time and length scales occur frequently in many areas of physics and interdisciplinary fields, such as turbulence [1], random-matrix theory [2], high-energy collision physics [3,4], and econophysics [5], to mention only a few. One common feature among many such systems is the appearance of probability distributions that deviate considerably from what one would expect (say, Gaussian or exponential behavior) on the basis of standard equilibrium statistical mechanics arguments. A great deal of effort has therefore been devoted to constructing physical models that generate such heavy-tailed distributions. One approach that has attracted considerable attention is the so-called nonextensive statistical mechanics formalism [6] whereby a power-law distribution, known as the Tsallis distribution, is obtained by maximizing a nonextensive entropy that generalizes the Boltzmann entropy formula. Heavy-tailed distributions can also be accounted for by a superposition of two statistics—a procedure known in mathematics as compounding [7] and in physics as superstatistics [8]. In particular, the Tsallis distribution can be readily obtained from the superstatistics approach by an appropriate choice of the weighting distribution [8]. Furthermore, this choice of weighting distribution can be justified from both a Bayesian analysis [9,10] and a maximum entropy principle based on the Boltzmann-Shannon entropy [11–15], thus circumventing the need to introduce a nonextensive entropy to justify the emergence of heavy-tailed distributions.

Recently, we introduced a general formalism [16–18] that extends the superstatistics approach to multiscale systems and gives rise to a large family of heavy-tailed distributions labeled by the number N of different scales present in the system. (Usual superstatistics corresponds to $N = 1$ [19].) In this hierarchical formalism, to which we refer as *H-theory*, it is assumed that at large scales the statistics of the system is described by a known distribution that contains a parameter (say, the temperature T_0) that characterizes the global equilibrium of the system. At short scales, however, the system deviates considerably from the large-scale distribution, owing to the complex multiscale dynamics (intermittency effects) of the system. The scale dependence of the relevant distributions can be effectively described by assuming that the environment (background) surrounding the small system under investigation changes slowly in time. The dynamics of the background is then formulated as a set of hierarchical stochastic differential equations whose form is derived from simple physical constraints, yielding only two "universality classes" for the stationary distributions of the background variables at each level of the hierarchy: (i) a gamma distribution and (ii) an inverse-gamma distribution. For both classes, analytical expressions are obtained for the marginal distribution of the background variable at the lowest level of the hierarchy in terms of Meijer G functions, from which the heavy-tailed distribution of the fluctuating signal is computed (and also written in terms of G functions). Here two classes of signal distributions are found [18] according to the behavior at the tails: (i) power-law decay and (ii) stretched-exponential

*giovani.vasconcelos@ufpe.br

tail. Applications of the H-theory to empirical data from several systems, such as turbulence [16,17], financial markets [18], and random fiber lasers [20] have yielded excellent results.

The dynamical formulation of the H-theory reviewed in the preceding paragraph represents a "microscopic" (i.e., small-scale) approach to the problem in that it tries to model the fluctuations in the environment under which the system evolves by a set of stochastic differential equations, which in principle provides a full description of the time-dependent stationary joint distribution function of the background variables. In this paper we take an alternative, thermodynamic-like approach in which the background distribution will be derived from a maximum entropy principle, thus bypassing the need to specify the underlying dynamics. We remark that this weakening of the basic dynamical hypothesis of H-theory leads to a considerable expansion of its domain of applicability, which may now include complex multiscale systems with non-Markovian stochastic dynamics.

The main purpose of the paper is to present a unified maximum-entropy principle suitable for hierarchical complex systems in statistical equilibrium. The main idea in our approach is to write the Boltzmann-Shannon entropy of the system in terms of the local equilibrium distribution of states of the small system under observation and the distributions of the background variables, representing the effective temperatures across the hierarchy of length scales. In other words, the system is treated as being effectively composed of a small system coupled to a set of "nested heat reservoirs" of increasingly larger size. Such hierarchy of reservoirs represents the distinct characteristic length scales of the system, where the temperature of each reservoir is allowed to fluctuate owing to the complex interactions between scales; see below.

In this multiscale picture, we seek to maximize the entropy with respect to the conditional temperature distributions at each level of the hierarchy, subject to certain physically motivated constraints. In doing so, we obtain a general family of distributions that includes two particular classes, namely, the *generalized gamma* and the *generalized inverse-gamma* distributions. The marginal distribution of temperature of the innermost reservoir (i.e., at the lowest level of the hierarchy) is obtained by integrating over the conditional distributions of all larger scales. Remarkably, the resulting distribution can be written explicitly in terms of a known special function, namely, the Fox H function. Averaging the quasiequilibrium Boltzmann distribution of the small system over the temperature of the innermost reservoir then yields the marginal distribution of states, which can also be written in terms of Fox H functions. Here again the distributions of states can be classified into two classes according to the tail behavior, namely, the power-law and stretched-exponential classes. For a particular choice of constraints our generalized distributions recover the distribution obtained in Ref. [18] in terms of Meijer G functions. The H-theory described here thus provides a rather general framework to describe the statistics of fluctuations in complex systems with multiple time and length scales.

II. MULTISCALE SYSTEMS

We consider a multiscale complex system that is characterized by N well-separated time scales τ_i, $i = 1,\ldots,N$, in addition to a large decorrelation time τ_0 above which fluctuations in the system are essentially uncorrelated. Let us order these time scales from smallest to largest: $\tau_N \ll \tau_{N-1} \ll \cdots \ll \tau_1 \ll \tau_0$. Thus, if one samples the system at time intervals larger than or comparable to τ_0, one will find the usual canonical distribution of states: $p(\boldsymbol{q}|\beta_0) = \exp[-\beta_0 E(\boldsymbol{q})]/Z(\beta_0)$, where $\boldsymbol{q}$ denotes the state variables, $\beta_0 = 1/k_B T_0$ with T_0 representing the "global" temperature of the system, $E(\boldsymbol{q})$ is the energy of the state labeled by $\boldsymbol{q}$, and $Z(\beta_0)$ is the large-scale partition function defined by $Z(\beta_0) = \int \exp[-\beta_0 E(\boldsymbol{q})]d\boldsymbol{q}$.

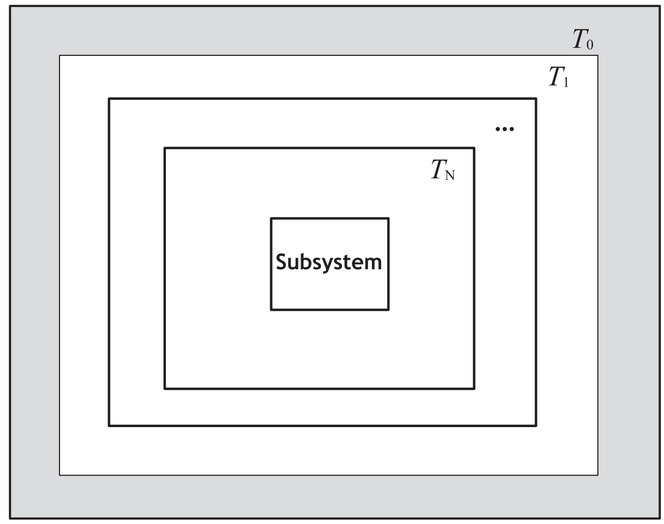

FIG. 1. Multiscale system at thermal equilibrium at temperature T_0. Each nested reservoir has an effective temperature T_j, $j = 1,\ldots,N$, which fluctuates around the same mean T_0.

At short time scales (say, smaller than the smallest characteristic time τ_N), the distribution of states $p(\boldsymbol{q})$ deviates considerably from the large-scale distribution $p(\boldsymbol{q}|\beta_0)$, owing to the complex multiscale dynamics of the system. In this scenario, it is convenient to consider the system as being composed of a small subsystem—corresponding to the effective region where the measurements are performed—and a large subsystem that has a slow internal dynamics characterized by several hierarchically arranged time scales. Thus, in contrast to the usual canonical formulation, the large subsystem can no longer be treated as a single heat reservoir with a fixed temperature. Instead, it must be viewed as a set of N nested reservoirs where each reservoir is described by a fluctuating temperature T_j, $j = 1,\ldots,N$; a cartoon view of the model is shown in Fig. 1. Physically, the fluctuations in these effective "temperatures" at different scales are caused by the complex interactions (exchange of energy) between scales in the hierarchy, in analogy with the phenomenon of intermittency in turbulence [1]. These interactions between scales can be roughly decomposed into two main components: (i) local energy exchange between adjacent reservoirs which tends to "equilibrate" their temperatures and (ii) energy transfer from larger scales onto smaller ones (not necessarily adjacent in the hierarchy) which may occur intermittently and represents a source of noise (i.e., fluctuations) for the temperature at the smaller scale. Notice, in particular, that if only local interactions were allowed the system as a whole would relax to a state described by only one (constant) temperature T_0, irrespective of the scale, and so one would be back to the usual canonical formalism. In other words, nonlocal interactions (in the above sense) are necessary to account for deviations from the usual Boltzmann-Gibbs distributions.

Invoking Bayes's theorem, the joint equilibrium distribution $p(\boldsymbol{q},\beta_1,\ldots,\beta_N)$, where $\beta_j = 1/k_B T_j$, can be factorized as

$$p(\boldsymbol{q},\boldsymbol{\beta}) = p(\boldsymbol{q}|\boldsymbol{\beta})p(\boldsymbol{\beta}), \tag{1}$$

where we introduced the notation $\boldsymbol{\beta} \equiv (\beta_1,\beta_2,\ldots,\beta_N)$. Because of the hierarchical nature of our system and separation of time scales, we assume that the stationary conditional distribution $p(\boldsymbol{q}|\boldsymbol{\beta})$ depends only on the inverse temperature β_N of the innermost reservoir, since this quantity characterizes the (slowly changing) state of the environment coupled to the small system of interest. We then write

$$p(\boldsymbol{q}|\boldsymbol{\beta}) = p(\boldsymbol{q}|\beta_N). \tag{2}$$

This means that the physical constraints imposed on the system at the large scale (and which fix the global temperature T_0) do not directly influence the small scales but rather are transferred down the hierarchy through the intervening scales. Under these assumptions, the marginal distribution $p(\boldsymbol{q})$ can be written as

$$P(\boldsymbol{q}) = \int_0^\infty P(\boldsymbol{q}|\beta_N)p(\beta_N)d\beta_N, \tag{3}$$

where the probability distribution $p(\beta_N)$ of the local inverse temperature β_N is given by

$$p(\beta_N) = \int_0^\infty \cdots \int_0^\infty p(\boldsymbol{\beta})d\beta_1 \cdots d\beta_{N-1}. \tag{4}$$

Owing to the separation of time scales, it is reasonable to assume that the small subsystem, which has a fast dynamics, is in local equilibrium with its immediate vicinity whose inverse temperature β_N changes much more slowly. In other words, over short time periods (during which β_N does not change appreciably) the conditional probability $p(\boldsymbol{q}|\beta_N)$ can be described by a Boltzmann distribution:

$$p(\boldsymbol{q}|\beta_N) = \frac{\exp[-\beta_N E(\boldsymbol{q})]}{Z(\beta_N)}. \tag{5}$$

The remaining task then is to find the distribution $p(\beta_N)$ of the local inverse temperature which encodes the complex dynamics of the multiscale background. This can be done by exploiting the hierarchical structure of the system, as argued below.

We assume that the inverse temperature of a subsystem (reservoir) at a given level j of the hierarchy depends conditionally only on the inverse temperature of the reservoir at the next level up the hierarchy (a Markov property), albeit its coupling to the remaining degrees of freedom in the upper levels are incorporated in the noise source (see below). We may thus write the joint distribution $p(\boldsymbol{\beta})$ as

$$p(\boldsymbol{\beta}) = \prod_{j=1}^{N} f(\beta_j|\beta_{j-1}), \tag{6}$$

where $f(\beta_j|\beta_{j-1})$ denotes the probability density of β_j conditioned on a fixed value of β_{j-1}. In view of Eqs. (4) and (6), the marginal distribution $p(\beta_N)$ can now be written as

$$p(\beta_N) = \int_0^\infty \cdots \int_0^\infty \prod_{j=1}^{N} f(\beta_j|\beta_{j-1})d\beta_1 \cdots d\beta_{N-1}, \tag{7}$$

In this way, our task has been reduced to computing the conditional distributions $f(\beta_j|\beta_{j-1})$, for $j = 1,\ldots,N$. In the next section we shall use a maximum entropy approach to solve this problem.

III. ENTROPY FORMULATION

A. Dynamical approach

As pointed out in Ref. [18], a simple way to accommodate the dynamical requirements of the hierarchical model is to introduce the following set of coupled stochastic differential equations for the time evolution of the inverse temperatures at each level of the hierarchy:

$$d\beta_j = -\gamma_j(\beta_j - \beta_{j-1})dt + \kappa_j \beta_j^s \beta_{j-1}^{1-s} dW_j; \quad j = 1,\ldots,N, \tag{8}$$

where $\gamma_j,\kappa_j > 0$ and $s \in \{1/2,1\}$ parametrizes the two universality classes. The specific form of the stochastic process described by Eq. (8) is fixed by the requirements of invariance under scale transformation ($\beta_j \to \lambda\beta_j$), positivity ($\beta_j(t) > 0$), and linear regression to the average global inverse temperature [$\langle\beta_j(t)\rangle \to \beta_0$ as $t \to \infty$]. Notice that the deterministic term in the dynamical model (8) represents the local interaction between scales, whereas the noise term accounts for nonlocal interactions leading to intermittency.

The corresponding Fokker-Planck equation is

$$\partial_t P = -\sum_{j=1}^{N} \partial_{\beta_j} J_j, \tag{9}$$

where

$$J_j = -\gamma_j(\beta_j - \beta_{j-1})P - \frac{\kappa_j^2}{2}\partial_{\beta_j}\left(\beta_j^{2s}\beta_{j-1}^{2-2s}P\right). \tag{10}$$

To be consistent with Eq. (6), we assume widely separated time scales $\gamma_N \gg \gamma_{N-1} \gg \cdots \gg \gamma_1$ and $\kappa_N \gg \kappa_{N-1} \gg \cdots \gg \kappa_1$, so that the condition of microscopic reversibility $J_j = 0$ applies for the equilibrium distribution. We thus have

$$\partial_{\beta_j}\left(\beta_j^{2s}\beta_{j-1}^{2-2s}P_{\text{eq}}\right) = -\alpha_j(\beta_j - \beta_{j-1})P_{\text{eq}}, \tag{11}$$

where $\alpha_j = 2\gamma_j/\kappa_j^2$ are free parameters. Equation (11) implies in turn that the stationary distribution $P_{\text{eq}}(\boldsymbol{\beta})$ can be written in the factorized form shown in Eq. (6).

Let us now consider the dynamical entropy defined as

$$S(t) = -\int P(\boldsymbol{\beta},t)\ln P(\boldsymbol{\beta},t)d\boldsymbol{\beta}, \tag{12}$$

where we use the short-hand notation $d\boldsymbol{\beta} = \prod_{j=1}^{N} d\beta_j$. We may use the existence of $P_{\text{eq}}(\boldsymbol{\beta})$ to write Eq. (12) as $S(t) = S_i(t) + S_e(t)$, where

$$S_i(t) = -\int d\boldsymbol{\beta}\, P(\boldsymbol{\beta},t)\ln[P(\boldsymbol{\beta},t)/P_{\text{eq}}(\boldsymbol{\beta})], \tag{13}$$

and

$$S_e(t) = -\int d\boldsymbol{\beta}\, P(\boldsymbol{\beta},t) \ln P_{\rm eq}(\boldsymbol{\beta}). \quad (14)$$

Here the two terms of entropy have their changes, $dS_i(t)$ and $dS_e(t)$, related to irreversible entropy production and reversible entropy flow, respectively [21,22]. Taking the derivative of $S(t)$ with respect to time we get

$$\dot{S}_i(t) = -\sum_{j=1}^{N} \int d\boldsymbol{\beta}\, J_j \partial_{\beta_j} \ln[P(\boldsymbol{\beta},t)/P_{\rm eq}(\boldsymbol{\beta})], \quad (15)$$

and

$$\dot{S}_e(t) = -\frac{d}{dt}\langle \ln P_{\rm eq}(\boldsymbol{\beta})\rangle_t. \quad (16)$$

The sign of $\dot{S}_e(t)$ depends on the initial condition $P_0(\boldsymbol{\beta}) \equiv P(\boldsymbol{\beta},0)$. In contrast, on using Eqs. (10) and (11) we see that

$$\dot{S}_i(t) = \sum_{j=1}^{N} \int d\boldsymbol{\beta} \frac{2J_j^2}{\kappa_j^2 \beta_j^{2s} \beta_{j-1}^{2-2s} P(\boldsymbol{\beta},t)} \geqslant 0. \quad (17)$$

We thus conclude that the stationary solution $P_{\rm eq}(\boldsymbol{\beta})$ is a dynamical maximum of $S_i(t)$. This result is consistent with a more general analysis of the entropy evolution of stochastic dynamical systems [23]. A Fokker-Planck approach to systems governed by a hierarchical dynamics has also been considered in the literature [24]. To proceed further we use Eq. (6) to write

$$\ln P_{\rm eq}(\boldsymbol{\beta}) = \sum_j \ln f(\beta_j|\beta_{j-1}) \quad (18)$$

and thus

$$S_i(t) = -\int d\boldsymbol{\beta}\, P(\boldsymbol{\beta},t)\left[\ln P(\boldsymbol{\beta},t) - \sum_j \ln f(\beta_j|\beta_{j-1})\right]. \quad (19)$$

We may now extend S_i to a Lagrange functional $S_i[P]$ and surmise that it reaches a conditional maximum at the stationary solution $P = P_{\rm eq}$, with the averages $\langle \ln f(\beta_j|\beta_{j-1})\rangle$ being constraints. This is the basic principle of the maximum entropy method which we describe in the next section.

B. Multiscale entropy

We start by defining the information entropy of the joint distribution $p(\boldsymbol{q},\boldsymbol{\beta})$ by

$$S[p(\boldsymbol{q},\boldsymbol{\beta})] = -\int\int p(\boldsymbol{q},\boldsymbol{\beta}) \ln p(\boldsymbol{q},\boldsymbol{\beta})\, d\boldsymbol{q}\, d\boldsymbol{\beta}. \quad (20)$$

In view of Eqs. (1), (2), and (6), the entropy (20) can be rewritten as

$$S[p(\boldsymbol{q},\boldsymbol{\beta})] = \int p(\boldsymbol{\beta}) s(\beta_N) d\boldsymbol{\beta} - \sum_{k=1}^{N} \int p(\boldsymbol{\beta}) \ln f(\beta_k|\beta_{k-1}) d\boldsymbol{\beta}, \quad (21)$$

where $s(\beta_N)$ is the thermodynamic entropy of the small subsystem

$$s(\beta_N) = -\int p(\boldsymbol{q}|\beta_N) \ln p(\boldsymbol{q}|\beta_N)\, d\boldsymbol{q}, \quad (22)$$

which is a multiscale generalization of the entropy described in superstatistics [11,15] for the case $N = 1$. Let us also define the entropy at level j, for $j = 0,\ldots,N-1$, as the average of $s(\beta_N)$ over all scales below this level, that is,

$$s(\beta_j) = \int s(\beta_N) p(\boldsymbol{\beta}) d\beta_{j+1}\cdots d\beta_N. \quad (23)$$

We now seek to maximize Eq. (21) with respect to the distributions $f(\beta_j|\beta_{j-1})$. To this end, let us first discuss the constraints under which we shall carry out this maximization procedure.

C. Constraints

The first set of constraints is given by the usual normalization condition

$$\int f(\beta_j|\beta_{j-1}) d\beta_j = 1, \quad j = 1,\ldots,N. \quad (24)$$

The second set of constraints entails the choice of a moment to be kept fixed in the maximization procedure. Usually, the first moment (mean) is the preferred choice [13,15]. Here, however, we shall adopt a more general approach and fix the rth moment of the distributions $f(\beta_j|\beta_{j-1})$. More specifically, we require that

$$\int \beta_j^r f(\beta_j|\beta_{j-1}) d\beta_j = \beta_{j-1}^r, \quad j = 1,\ldots,N, \quad (25)$$

for some arbitrary real $r \neq 0$ (not necessarily an integer). Notice that Eq. (25) implies that

$$\langle \beta_j^r \rangle \equiv \int \beta_j^r p(\boldsymbol{\beta}_j) d\boldsymbol{\beta}_j = \beta_0^r, \quad j = 1,\ldots,N, \quad (26)$$

where we introduced the notation

$$\boldsymbol{\beta}_j \equiv (\beta_1,\ldots,\beta_j).$$

Equation (26) can be seen as a generalized equilibrium condition in the sense that the average value of β_j^r is the same at all levels of the hierarchy. We anticipate that this generalized constraint allows us to obtain a larger class of distributions than that generated by the dynamical [18] approach described in Sec. III A, as will be discussed in Sec. IV.

As an additional constraint we use the average entropy

$$\langle s(\beta_N)\rangle \equiv \int s(\beta_N) p(\boldsymbol{\beta}) d\boldsymbol{\beta} = s(\beta_0), \quad (27)$$

where $s(\beta_0)$ is fixed. It then follows from definition (23) that the average entropy is the same across all scales:

$$\langle s(\beta_j)\rangle \equiv \int s(\beta_j) p(\boldsymbol{\beta}_j) d\boldsymbol{\beta}_j = s(\beta_0), \quad j = 1,\ldots,N, \quad (28)$$

which is a reasonable equilibrium condition. Furthermore, we shall assume that the thermodynamic entropy defined in Eq. (22) satisfies the following relation:

$$s(\beta_N) \sim s_0 \ln \beta_N, \quad (29)$$

where s_0 is a constant and the notation $\sim$ indicates equality except for an additive constant. [In other words, $f(x) \sim g(x)$ means here that $f(x) = g(x) + C$, where C is a constant.] We recall that relation (29) is valid for a large class of systems, such as those that obey the equipartition theorem, for which the internal energy is proportional to the temperature [13,15].

We also make the assumption that the distribution $f_k(\beta_k|\beta_{k-1})$ is invariant under a rescaling of the variables $\beta \to \lambda\beta$:

$$f_k(\beta_k|\beta_{k-1})d\beta_k = f_k(\lambda\beta_k|\lambda\beta_{k-1})d(\lambda\beta_k). \tag{30}$$

Physically, this means that the temperature distributions should remain of the same form regardless of the temperature scale one chooses. Now, if we make $\lambda = 1/\beta_{k-1}$ in Eq. (30) we get

$$f_k(\beta_k|\beta_{k-1})d\beta_k = g_k\left(\frac{\beta_k}{\beta_{k-1}}\right)\frac{d\beta_k}{\beta_{k-1}} = g_k(u)du, \tag{31}$$

for some function $g_k(u)$, where $u = \beta_k/\beta_{k-1}$. Relation (31) leads to the following two useful relations, which are proven in Appendix A:

$$\int p(\boldsymbol{\beta}_k)\ln\beta_k d\boldsymbol{\beta}_k \sim \int p(\boldsymbol{\beta}_j)\ln\beta_j d\boldsymbol{\beta}_j, \quad \text{for } j \leqslant k, \tag{32}$$

and

$$\int p(\boldsymbol{\beta}_k)\ln f(\beta_k|\beta_{k-1})d\boldsymbol{\beta}_k \sim -\int p(\boldsymbol{\beta}_j)\ln\beta_j d\boldsymbol{\beta}_j, \quad \text{for } j < k. \tag{33}$$

Now, inserting Eq. (29) into Eq. (23) and using Eq. (32), one finds that

$$s(\beta_j) = s_0\ln\beta_j + s_j, \tag{34}$$

where s_j is a constant that does not depend on β_j. In view of this relation, the constraint (28) can be written as

$$\int (\ln\beta_j)p(\boldsymbol{\beta}_j)d\boldsymbol{\beta}_j = c_j, \tag{35}$$

where c_j is a constant.

D. Entropy maximization

In order to maximize Eq. (21) with respect to $f(\beta_j|\beta_{j-1})$, for any given j, it is necessary to make explicit the dependence of $S[p(\boldsymbol{q},\boldsymbol{\beta})]$ on $f(\beta_j|\beta_{j-1})$. To this end, we first note that on use of Eqs. (6) and (24) we can rewrite Eq. (21) as

$$\begin{aligned} S[p(\boldsymbol{q},\boldsymbol{\beta})] = &\int s(\beta_j)p(\boldsymbol{\beta}_j)d\boldsymbol{\beta}_j \\ &- \sum_{k=1}^{j-1}\int p(\boldsymbol{\beta}_k)\ln f(\beta_k|\beta_{k-1})d\boldsymbol{\beta}_k \\ &- \int p(\boldsymbol{\beta}_j)\ln f(\beta_j|\beta_{j-1})d\boldsymbol{\beta}_j \\ &- \sum_{k=j+1}^{N}\int p(\boldsymbol{\beta}_k)\ln f(\beta_k|\beta_{k-1})d\boldsymbol{\beta}_k. \end{aligned} \tag{36}$$

Now using Eqs. (33) and (34) in Eq. (36), one finds that

$$\begin{aligned} S[p(\boldsymbol{q},\boldsymbol{\beta})] \sim &\, c_j\int p(\boldsymbol{\beta}_j)\ln\beta_j d\boldsymbol{\beta}_j \\ &- \int p(\boldsymbol{\beta}_j)\ln f(\beta_j|\beta_{j-1})d\boldsymbol{\beta}_j \\ &- \sum_{k=1}^{j-1}\int p(\boldsymbol{\beta}_k)\ln f(\beta_k|\beta_{k-1})d\boldsymbol{\beta}_k, \end{aligned} \tag{37}$$

where $c_j = N - j - s_0$. Note that the entropy $S[p(\boldsymbol{q},\boldsymbol{\beta})]$ depends on $f(\beta_j|\beta_{j-1})$ only through the first two terms in the right-hand side of Eq. (37).

Maximizing Eq. (37) with respect to $f(\beta_j|\beta_{j-1})$, subject to the constraints in Eqs. (24), (25), and (35), yields

$$\int\left[\ln f(\beta_j|\beta_{j-1}) + A_j + B_j\beta_j^r + C_j\ln\beta_j\right]\delta_j p(\boldsymbol{\beta}_j)d\boldsymbol{\beta}_j = 0, \tag{38}$$

where A_j, B_j, and C_j are Lagrange multipliers and $\delta_j p(\boldsymbol{\beta}_j) \equiv p(\boldsymbol{\beta}_{j-1})\delta f(\beta_j|\beta_{j-1})$. The solution to Eq. (38) takes the form

$$f(\beta_j|\beta_{j-1}) = e^{-A_j}\beta_j^{-C_j}\exp\left(-B_j\beta_j^r\right). \tag{39}$$

To enforce the constraint (25) we choose $B_j = \alpha_j/\beta_{j-1}^r$ and set $C_j = -r\alpha_j + 1$, where $\alpha_j > 0$. Using these parameters in Eq. (39) one obtains the following general distribution:

$$f_j(\beta_j|\beta_{j-1}) = \frac{|r|\alpha_j^{\alpha_j}}{\beta_j\Gamma(\alpha_j)}\left(\frac{\beta_j}{\beta_{j-1}}\right)^{r\alpha_j}\exp\left[-\alpha_j\left(\frac{\beta_j}{\beta_{j-1}}\right)^r\right]. \tag{40}$$

For $r > 0$ this distribution corresponds to the *generalized gamma distribution*, whereas for $r < 0$ it gives the *generalized inverse gamma distribution*.

We note furthermore that for the particular case $r = 1$ the distribution (40) yields the usual gamma distribution,

$$f_j(\beta_j|\beta_{j-1}) = \frac{(\alpha_j/\beta_{j-1})^{\alpha_j}}{\Gamma(\alpha_j)}\beta_j^{\alpha_j-1}\exp\left(-\frac{\alpha_j\beta_j}{\beta_{j-1}}\right), \tag{41}$$

whereas for $r = -1$ it gives the standard inverse gamma distribution:

$$f_j(\beta_j|\beta_{j-1}) = \frac{(\alpha_j\beta_{j-1})^{\alpha_j}}{\Gamma(\alpha_j)}\beta_j^{-\alpha_j-1}\exp\left(-\frac{\alpha_j\beta_{j-1}}{\beta_j}\right). \tag{42}$$

It is interesting to note that the generalized inverse gamma distribution has been used to model the statistics of certain complex systems, such as the wealth distribution in ancient Egypt [25]. The Weibull and the Frechet distributions, which are particular cases of the generalized gamma and generalized inverse-gamma distributions, respectively, have also found important applications in extreme value statistics [26] and sum of correlated random variables [27]. Here, however, our interest is to use Eq. (40) not so much as a stand-alone distribution but rather as a means to obtain the distribution $p(\beta_N)$ of inverse temperatures at the innermost reservoir, from which the distribution of states $p(\boldsymbol{q})$ can be found. This is done next.

IV. EQUILIBRIUM DISTRIBUTIONS

As discussed in Sec. II, the complex dynamics of the large system (background) is felt by the small subsystem only through the fluctuations of the inverse temperature β_N of the innermost reservoir. Thus, in order to determine the marginal distribution of states $p(\boldsymbol{q})$ of the small subsystem, it is necessary first to compute the distribution $p(\beta_N)$; see Eq. (3). It is remarkable that both these distributions can be obtained in analytical form in terms of some special transcendental functions known as the Fox H functions [28], as shown below.

A. Background distribution

The marginal distribution $p(\beta_N)$ at the lowest level of the hierarchy is given by Eq. (7), where each of the distributions $f(\beta_j|\beta_{j-1})$ appearing in this expression is as shown in Eq. (40). In computing the multiple integrals in Eq. (7) the cases $r > 0$ and $r < 0$ need to be treated separately, but for both cases these integrals can be calculated explicitly in terms of the Fox H functions.

As shown in Appendix B, for the case $r > 0$ one finds

$$p(\beta_N) = \omega_\rho \Omega H^{N,0}_{0,N}\left(\begin{matrix} - \\ (\boldsymbol{\alpha} - \rho\mathbf{1}, \rho\mathbf{1}) \end{matrix} \middle| \frac{\omega_\rho \beta_N}{\beta_0} \right), \tag{43}$$

while for $r < 0$ the result is

$$p(\beta_N) = \frac{\Omega}{\omega_\rho} H^{0,N}_{N,0}\left(\begin{matrix} ((1-\rho)\mathbf{1} - \boldsymbol{\alpha}, \rho\mathbf{1}) \\ - \end{matrix} \middle| \frac{\beta_N}{\omega_\rho \beta_0} \right), \tag{44}$$

where $\rho = 1/|r|$, $\omega_\rho = \prod_{j=1}^N \alpha_j^\rho$, and $\Omega = 1/[(\beta_0 \Gamma(\boldsymbol{\alpha})]$. Here we have introduced the vector notation $\boldsymbol{\alpha} \equiv (\alpha_1, \ldots, \alpha_N)$ and $\Gamma(\mathbf{a}) \equiv \prod_{j=1}^N \Gamma(a_j)$. We have also used a dash in the top row of the H function in Eq. (43) and in the low row of the H function in Eq. (44) to indicate that the respective parameters are not present.

We note in passing that after setting $|r| = 1$ in expressions (43) and (44) we recover the two classes of universality for the background distributions obtained in Ref. [18] from a stochastic dynamical model. To see this, we note that for $\rho = 1$ the set of parameters $\rho\mathbf{1} \equiv (\rho, \ldots, \rho)$ appearing in each of the H functions above becomes simply the identity vector, in which case the H function reduces to a simpler function, namely, the Meijer G function [28]. Setting $\rho = 1$ in Eq. (43) then yields

$$p(\beta_N) = \omega\Omega G^{N,0}_{0,N}\left(\begin{matrix} - \\ \boldsymbol{\alpha} - \mathbf{1} \end{matrix} \middle| \frac{\omega\beta_N}{\beta_0} \right), \tag{45}$$

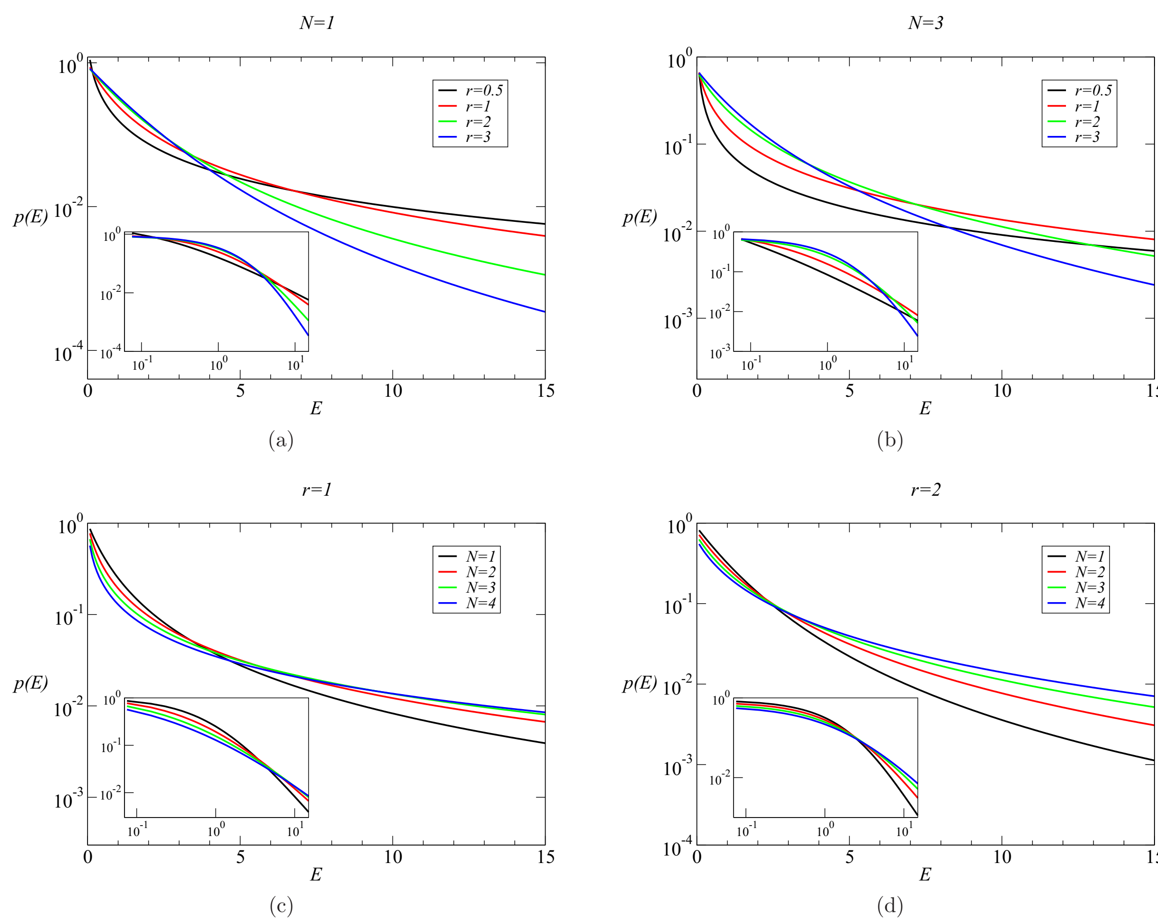

FIG. 2. (a) Distribution of states $p(E)$ for the power-law class for the following values of parameters: (a) $N = 1$, $r = 0.5,1,2,3$; (b) $N = 3$, $r = 0.5,1,2,3$; (c) $r = 1$, $N = 1,2,3,4$; and (d) $r = 2$, $N = 1,2,3,4$. In all cases shown here we have used $\gamma = 1$, $\beta_0 = 1$, $Z(\beta_0) = 1$, and $\alpha_j = \alpha = 1.0$, for $j = 1, \ldots, N$.

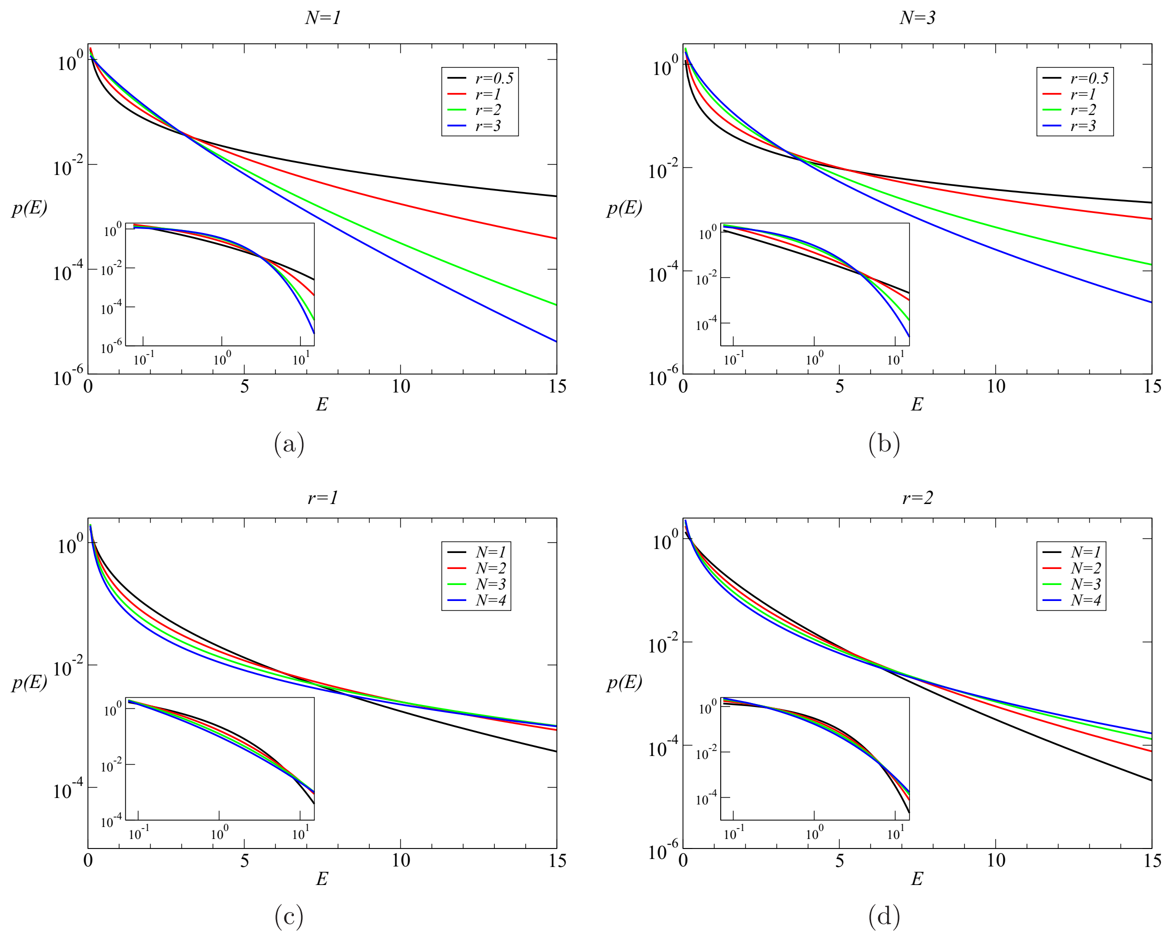

FIG. 3. Distribution of states $p(E)$ for the stretched exponential class with the same choice of parameters as in Fig. 2.

while from Eq. (44) one has

$$p(\beta_N) = \frac{\Omega}{\omega} G^{0,N}_{N,0}\left(\begin{matrix} -\boldsymbol{\alpha} \\ - \end{matrix} \middle| \frac{\beta_N}{\beta_0 \omega} \right), \tag{46}$$

where $\omega = \prod_{j=1}^{N} \alpha_j$. In comparing the distributions (45) and (46) with the corresponding expressions given in Ref. [18] one has to bear in mind that there the distributions are written in terms of a variable ε_N which corresponds in the notation of the present paper to $1/\beta_N$.

B. Distribution of states

In view of Eqs. (3) and (5), the marginal distribution of states $p(\boldsymbol{q})$ of the small subsystem can be written as

$$p(\boldsymbol{q}) = \int_0^\infty \frac{\exp[-\beta_N E(\boldsymbol{q})]}{Z(\beta_N)} p(\beta_N) d\beta_N, \tag{47}$$

where $p(\beta_N)$ is given by either Eq. (43) or (44). In order to carry out this integral one needs to know the dependence of the partition function $Z(\beta_N)$ on β_N. Recalling that $S = \partial(k_B T \ln Z)/\partial T$, one then sees that the assumption (29) is compatible with the behavior $Z(\beta_N) \sim \beta_N^{-\gamma}$, for some exponent $\gamma > 0$, and so we write

$$Z(\beta_N) = Z(\beta_0)\left(\frac{\beta_N}{\beta_0}\right)^{-\gamma}. \tag{48}$$

Inserting Eq. (48) into (47) yields

$$p(\boldsymbol{q}) = \frac{1}{Z(\beta_0)} \int_0^\infty \left(\frac{\beta_N}{\beta_0}\right)^{\gamma} \exp[-\beta_N E(\boldsymbol{q})] p(\beta_N) d\beta_N. \tag{49}$$

It is also remarkable that this integral can be carried out explicitly in terms of Fox H functions for both classes of background distributions, with the resulting distributions being classified into two classes according to the behavior at the tails, as follows:

(i) *Power-law class*. This is the case when $r > 0$. Upon inserting Eq. (43) into (49) and using a convolution property of the H function [28], the resulting integral can be performed explicitly (see Appendix C), yielding

$$p(E) = \frac{1}{Z(\beta_0)\omega_\rho^{\gamma}\Gamma(\boldsymbol{\alpha})} H^{1,N}_{N,1}\left(\begin{matrix} ((1-\gamma\rho)\mathbf{1} - \boldsymbol{\alpha}, \rho\mathbf{1}) \\ (0,1) \end{matrix} \middle| \frac{\beta_0 E}{\omega_\rho} \right). \tag{50}$$

Here we have omitted the state variable $\boldsymbol{q}$ for simplicity of notation, with the understanding that $p(E)$ denotes the probability of a state $\boldsymbol{q}$ with energy $E(\boldsymbol{q})$. From the asymptotic expansion of the H function for large arguments one finds [28] that the $p(E)$ decays as a power-law for large values of E:

$$p(E) \sim \sum_{j=1}^{N} \frac{c_j}{E^{\gamma+|r|\alpha_j}}, \quad \text{for} \quad E \to \infty, \tag{51}$$

where the c_i's are constants. To illustrate the power-law class of distributions we show in Fig. 2 some plots of the function $p(E)$ given in Eq. (50) for cases where $\gamma = 1$, $\beta_0 = 1$, $Z(\beta_0) = 1$, and $\alpha_j = \alpha = 1.0$. The values of the parameters N and r for each plot are indicated in the figure caption. The main plots in Fig. 2 are in semilogarithmic scale, while the insets show the same data in log-log scale. One clearly sees from Figs. 2(a) and 2(b) that the smaller the value of the parameter r, for N fixed, the heavier the tail of the distribution. This is in agreement with the asymptotic behavior given in Eq. (51) which shows that the exponent of the power law decreases as r decreases. Similarly, from Figs. 2(c) and 2(d) one sees that the larger the number N of scales, for r fixed, the heavier the tails. Note, however, that the exponent of the power law does not depend on N; see Eq. (51). It is instead the prefactor that increases with N, since we are taking $\alpha_j = \alpha$, for $j = 1, \ldots, N$, thus causing a slower decay of the tail.

(ii) *Stretched-exponential class*. This corresponds to the case $r < 0$. Here the integral (49), with $p(\beta_N)$ as given in Eq. (44), can be written as

$$p(E) = \frac{\omega_\rho^\gamma}{Z(\beta_0)\Gamma(\boldsymbol{\alpha})} H_{0,N+1}^{N+1,0}\left(\overline{(\boldsymbol{\alpha} - \gamma\rho\mathbf{1}, \rho\mathbf{1}), (0,1)} \middle| \omega_\rho \beta_0 E \right), \tag{52}$$

as also shown in Appendix C. The asymptotic behavior in this case is given by a modified stretched exponential:

$$p(E) \sim E^\theta \exp[-A(\omega_\rho \beta_0 E)^{1/(\rho N+1)}], \quad \text{for} \quad E \to \infty, \tag{53}$$

where $\theta = N(\bar{\alpha} - \gamma\rho - 1/2)/(\rho N + 1)$, $\bar{\alpha} = (1/N)\sum_{i=1}^{N} \alpha_i$ and $A = (\rho N + 1)\rho^{-\rho N/(\rho N+1)}$. Some illustrative plots of the function $p(E)$ given in Eq. (52) are shown in Fig. 3 for the same choice of parameters as in Fig. 2. The same qualitative dependence of the tails on the parameters N and r are observed here: the larger the value of N or the smaller the choice of r, the heavier the tails. This behavior is in agreement with Eq. (53) which shows that the exponent of the stretched exponential decreases with both the increase of N and the decrease of r.

We note in passing that the particular cases $r = \pm 1$ yield results consistent with those obtained in Ref. [18], in that the corresponding distributions can also be written in terms of G functions. For $\rho = 1$ the expression (50) simplifies to

$$p(E) = \frac{1}{Z(\beta_0)\omega^\gamma \Gamma(\boldsymbol{\alpha})} G_{N,1}^{1,N}\left(\begin{matrix} (1-\gamma)\mathbf{1} - \boldsymbol{\alpha} \\ 0 \end{matrix} \middle| \frac{\beta_0 E}{\omega} \right), \tag{54}$$

whereas the distribution (52) reads

$$p(E) = \frac{\omega^\gamma}{Z(\beta_0)\Gamma(\boldsymbol{\alpha})} G_{0,N+1}^{N+1,0}\left(\overline{\boldsymbol{\alpha} - \gamma\mathbf{1}, 0} \middle| \omega\beta_0 E \right). \tag{55}$$

V. CONCLUSIONS

In this paper, we have used a maximum entropy principle to derive a generalized version of the multicanonical formalism (H-theory) introduced in Refs. [17,18]. In our approach the system is considered to be effectively composed of a small subsystem in thermal equilibrium with a hierarchical set of heat reservoirs, whose local temperatures fluctuate owing to weak interactions between different scales. We characterized the joint equilibrium distribution of the state variables and the local inverse temperatures by means of its Shannon information entropy. This entropy was maximized with respect to the conditional temperature distributions at each level of the hierarchy, subject to certain physically motivated constraints. The large family of distributions that were found by this procedure can be grouped into two classes: the *generalized gamma* and the *generalized inverse-gamma* distributions. The knowledge of these conditional distributions of inverse temperatures allowed us to obtain the marginal distribution $p(\beta_N)$ of the inverse temperature at the lowest level of the hierarchy, which was explicitly written for both classes in terms of the Fox H functions.

The marginal distribution of states $p(\boldsymbol{q})$ was then obtained by averaging the conditional distribution of states $p(\boldsymbol{q}|\beta_N)$ over the local inverse-temperature β_N and the resulting distribution was also written in terms of Fox H functions. These distributions exhibit heavy tails that can be classified into two classes, namely, the power-law and stretched-exponential classes. The distributions derived in Ref. [18] from a stochastic dynamical approach, which were written in terms of Meijer G functions, were shown to be particular cases of the Fox H functions obtained from the maximum entropy approach. The H-theory presented here thus provides a rather general framework to describe the statistics of fluctuations in complex systems with multiple time and space scales, quite irrespective of the detailed underlying dynamics. Applications of H-theory in the context of Eulerian and Lagrangian turbulence, mathematical finance, and random lasers have had great success. Further applications of the generalized formalism presented here, including cases where the conditional distribution $p(\boldsymbol{q}|\beta_N)$ is not the Boltzmann-Gibbs distribution, are under current investigation.

ACKNOWLEDGMENTS

We thank W. Sosa for generating the data used in the figures. This work was supported in part by the Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), under Grants No. 308290/2014-3 and No. 311497/2015-2, and by Fundação de Amparo à Ciência e Tecnologia de Pernambuco (FACEPE), under Grant No. APQ-0073-1.05/15.

APPENDIX A: DERIVATION OF EQS. (32) AND (33)

First consider a term of the form

$$\int p(\boldsymbol{\beta}_k) \ln \beta_k \, d\boldsymbol{\beta}_k. \tag{A1}$$

This can be rewritten as

$$\int p(\boldsymbol{\beta}_k) \ln \beta_k \, d\boldsymbol{\beta}_k = \int p(\boldsymbol{\beta}_k)\left[\ln\left(\frac{\beta_k}{\beta_{k-1}}\right) + \ln \beta_{k-1} \right] d\boldsymbol{\beta}_k. \tag{A2}$$

Upon using property (31) we then obtain

$$\int p(\boldsymbol{\beta}_k)\ln\beta_k d\boldsymbol{\beta}_k = \left(\int g_k(u)\ln u\, du\right)\int p(\boldsymbol{\beta}_{k-1})d\boldsymbol{\beta}_{k-1} + \int p(\boldsymbol{\beta}_{k-1})\ln\beta_{k-1}\, d\boldsymbol{\beta}_{k-1} = A_k + \int p(\boldsymbol{\beta}_{k-1})\ln\beta_{k-1}\, d\boldsymbol{\beta}_{k-1}, \quad \text{(A3)}$$

where $A_k = \int_0^\infty g_k(u)\ln u\, du$ is a constant. This implies that

$$\int p(\boldsymbol{\beta}_k)\ln\beta_k d\boldsymbol{\beta}_k \sim \int p(\boldsymbol{\beta}_{k-1})\ln\beta_{k-1}d\boldsymbol{\beta}_{k-1}, \quad \text{(A4)}$$

where we recall that the notation $\sim$ implies equality, except for an irrelevant additive constant. If we repeat this procedure recursively we get Eq. (32).

Next consider terms of the form

$$\int p(\boldsymbol{\beta}_k)\ln f(\beta_k|\beta_{k-1})d\boldsymbol{\beta}_k. \quad \text{(A5)}$$

Using Eq. (31), we have

$$\int p(\boldsymbol{\beta}_k)\ln f(\beta_k|\beta_{k-1})d\boldsymbol{\beta}_k = \int p(\boldsymbol{\beta}_k)\ln\left[\frac{1}{\beta_{k-1}}g_k\left(\frac{\beta_k}{\beta_{k-1}}\right)\right]d\boldsymbol{\beta}_k = \int g(u)\ln g_k(u)du - \int p(\boldsymbol{\beta}_{k-1})\ln\beta_{k-1}d\boldsymbol{\beta}_{k-1} = B_k - \int p(\boldsymbol{\beta}_{k-1})\ln\beta_{k-1}d\boldsymbol{\beta}_{k-1}, \quad \text{(A6)}$$

where $B_k = \int_0^\infty g_k(u)\ln g_k(u)\, du$. Neglecting this additive constant we can then write

$$\int p(\boldsymbol{\beta}_k)\ln f(\beta_k|\beta_{k-1})d\boldsymbol{\beta}_k \sim -\int p(\boldsymbol{\beta}_{k-1})\ln\beta_{k-1}d\boldsymbol{\beta}_{k-1}, \quad \text{(A7)}$$

which in view of Eq. (32) yields Eq. (33), as desired.

APPENDIX B: DERIVATION OF EQS. (43) AND (44)

Here we calculate $p(\beta_N)$ explicitly in terms of Fox H functions. We begin by introducing the variable

$$y = \frac{\beta_N}{\beta_0} = \prod_{j=1}^N \xi_j, \quad \text{(B1)}$$

where $\xi_j = \beta_j/\beta_{j-1}$, so that $p(\beta_N) = g(y)/\beta_0$ and

$$g(y) = \int_0^\infty \cdots \int_0^\infty \prod_{j=1}^N g_j(\xi_j)d\xi_j\delta(y - \xi_1\xi_2\cdots\xi_N). \quad \text{(B2)}$$

For $r > 0$ we obtain from Eq. (40) that

$$g_j(\xi_j) = \frac{r\alpha_j^{\alpha_j}}{\Gamma(\alpha_j)}\xi_j^{r\alpha_j-1}e^{-\alpha_j\xi_j^r}, \quad \text{(B3)}$$

while for $r < 0$ we find

$$g_j(\xi_j) = \frac{r'\alpha_j^{\alpha_j}}{\Gamma(\alpha_j)}\xi_j^{-r'\alpha_j-1}e^{-\alpha_j\xi_j^{-r'}}, \quad \text{(B4)}$$

where we defined $r' = -r > 0$.

Now applying the Mellin transform, defined as

$$\mathcal{M}[g;s] \equiv \int_0^\infty dy y^{s-1}g(y), \quad \text{(B5)}$$

to both sides of Eq. (B2), we find

$$\mathcal{M}[g;s] = \prod_{j=1}^N \mathcal{M}[g_j;s], \quad \text{(B6)}$$

where

$$\mathcal{M}[g_j;s] = \frac{\Gamma(\alpha_j + (s-1)/r)}{\alpha_j^{(s-1)/r}\Gamma(\alpha_j)} \quad \text{(B7)}$$

is the Mellin transform of Eq. (B3), and

$$\mathcal{M}[g_j;s] = \alpha_j^{(s-1)/r'}\frac{\Gamma(\alpha_j + (1-s)/r')}{\Gamma(\alpha_j)} \quad \text{(B8)}$$

is the Mellin transform of Eq. (B4). Next, we use the following property of the Fox H function [28]. If the Mellin transform of $g(y)$ is

$$\mathcal{M}[g;s] = \frac{\lambda^{-s}\prod_{j=1}^m\Gamma(b_j + B_js)\prod_{j=1}^n\Gamma(1-a_j-A_js)}{\prod_{j=m+1}^q\Gamma(1-b_j-B_js)\prod_{j=n+1}^p\Gamma(a_j+A_js)}$$

then

$$g(y) = H_{p,q}^{m,n}\left(\begin{matrix}(\boldsymbol{a},\boldsymbol{A})\\(\boldsymbol{b},\boldsymbol{B})\end{matrix}\middle|\lambda y\right), \quad \text{(B9)}$$

where we introduced the notation $(\boldsymbol{x},\boldsymbol{X}) \equiv \{(x_1,X_1),\ldots,(x_d,X_d)\}$, with $d\in\{p,q\}$. Using Eqs. (B6), (B7), and (B9) we obtain Eq. (43); while using Eqs. (B6), (B8), and (B9) we get Eq. (44), as desired.

APPENDIX C: DERIVATION OF EQS. (50) AND (52)

We start by considering the Laplace transform of the Fox H function [28]

$$\int_0^\infty dx\, x^\gamma e^{-sx}H_{p,q}^{m,n}\left(\begin{matrix}(\boldsymbol{a},\boldsymbol{A})\\(\boldsymbol{b},\boldsymbol{B})\end{matrix}\middle|\lambda x\right) = s^{-(\gamma+1)}H_{p+1,q}^{m,n+1}\left(\begin{matrix}(\boldsymbol{a},\boldsymbol{A}),(-\gamma,1)\\(\boldsymbol{b},\boldsymbol{B})\end{matrix}\middle|\lambda s^{-1}\right), \quad \text{(C1)}$$

where $(\boldsymbol{x},\boldsymbol{X}) \equiv \{(x_1,X_1),\ldots,(x_d,X_d)\}$, with $d\in\{p,q\}$. Using the identities

$$H_{p,q}^{m,n}\left(\begin{matrix}(\boldsymbol{a},\boldsymbol{A})\\(\boldsymbol{b},\boldsymbol{B})\end{matrix}\middle|z\right) = H_{q,p}^{n,m}\left(\begin{matrix}(\boldsymbol{1-b},\boldsymbol{B})\\(\boldsymbol{1-a},\boldsymbol{A})\end{matrix}\middle|\frac{1}{z}\right) \quad \text{(C2)}$$

and

$$z^\sigma H_{p,q}^{m,n}\left(\begin{matrix}(\boldsymbol{a},\boldsymbol{A})\\(\boldsymbol{b},\boldsymbol{B})\end{matrix}\middle|z\right) = H_{p,q}^{m,n}\left(\begin{matrix}(\boldsymbol{a}+\sigma\boldsymbol{A},\boldsymbol{A})\\(\boldsymbol{b}+\sigma\boldsymbol{B},\boldsymbol{B})\end{matrix}\middle|z\right) \quad \text{(C3)}$$

we may rewrite Eq. (C1) as

$$\int_0^\infty dx x^\gamma e^{-sx}H_{p,q}^{m,n}\left(\begin{matrix}(\boldsymbol{a},\boldsymbol{A})\\(\boldsymbol{b},\boldsymbol{B})\end{matrix}\middle|\lambda x\right) = \frac{1}{\lambda^{\gamma+1}}H_{q,p+1}^{n+1,m}\left(\begin{matrix}(\boldsymbol{1-b}-(\gamma+1)\boldsymbol{B},\boldsymbol{B})\\(\boldsymbol{1-a}-(\gamma+1)\boldsymbol{A},\boldsymbol{A}),(0,1)\end{matrix}\middle|\frac{s}{\lambda}\right). \quad \text{(C4)}$$

We are now in position to calculate the Laplace transform of $p(\beta_N)$. Using Eqs. (43) and (C4), we get for the case $r > 0$:

$$\int_0^\infty d\beta_N\, \beta_N^\gamma e^{-\beta_N E} p(\beta_N) = \frac{\beta_0^\gamma}{\omega_\rho^\gamma \Gamma(\boldsymbol{\alpha})} H_{N,1}^{1,N}\left(\begin{matrix} ((1-\gamma\rho)\mathbf{1}-\boldsymbol{\alpha},\rho\mathbf{1}) \\ (0,1) \end{matrix} \middle| \frac{\beta_0 E(\boldsymbol{q})}{\omega_\rho} \right). \tag{C5}$$

Similarly, in view of Eq. (44), the result for $r < 0$ is

$$\int_0^\infty d\beta_N\, \beta_N^\gamma e^{-\beta_N E} p(\beta_N) = \frac{(\beta_0\omega_\rho)^\gamma}{\Gamma(\boldsymbol{\alpha})} H_{0,N+1}^{N+1,0}\left(\begin{matrix} - \\ (\boldsymbol{\alpha}-\gamma\rho\mathbf{1},\rho\mathbf{1}),(0,1) \end{matrix} \middle| \omega_\rho \beta_0 E(\boldsymbol{q}) \right). \tag{C6}$$

Using Eqs. (C5) and (C6), we obtain Eqs. (50) and (52) respectively.

[1] U. Frisch, *Turbulence: The Legacy of A. N. Kolmogorov* (Cambridge University Press, Cambridge, England, 1995).
[2] M. L. Mehta, *Random Matrices* (Elsevier, Amsterdam, 2004).
[3] G. Wilk and Z. Wlodarczyk, Eur. Phys. J. A **48**, 161 (2012).
[4] C. Beck, Eur. Phys. J. A **40**, 267 (2009), and references therein.
[5] S. Ghashghaie, W. Breymann, J. Peinke, P. Talkner, and Y. Dodge, Nature **381**, 767 (1996).
[6] C. Tsallis, J. Stat. Phys. **52**, 479 (1988); *Introduction to Nonextensive Statistical Mechanics* (Springer, New York, 2009).
[7] O. Barndor-Nielsen, J. Kent, and M. Srensen, Int. Stat. Rev. **50**, 145 (1982); S. D. Dubey, Metrika **16**, 27 (1970).
[8] C. Beck, Phys. Rev. Lett. **87**, 180601 (2001); Europhys. Lett. **64**, 151 (2003); Phys. Rev. Lett. **98**, 064502 (2007).
[9] F. Sattin, Eur. Phys. J. B **49**, 219 (2006).
[10] G. L. Vasconcelos and D. S. P. Salazar, Rev. Cub. Fis. **29**, 1E51 (2012).
[11] S. Abe, C. Beck, and E. G. D. Cohen, Phys. Rev. E **76**, 031102 (2007).
[12] G. E. Crooks, Phys. Rev. E **75**, 041119 (2007).
[13] E. V. Vakarin and J. P. Badiali, Phys. Rev. E **74**, 036120 (2006).
[14] E. Van der Straeten and C. Beck, Phys. Rev. E **78**, 051101 (2008).
[15] P. Dixit, Phys. Chem. Chem. Phys. **17**, 13000 (2015).
[16] D. S. P. Salazar and G. L. Vasconcelos, Phys. Rev. E **82**, 047301 (2010).
[17] D. S. P. Salazar and G. L. Vasconcelos, Phys. Rev. E **86**, 050103(R) (2012).
[18] A. M. S. Macêdo, I. R. R. González, D. S. P. Salazar, and G. L. Vasconcelos, Phys. Rev. E **95**, 032315 (2017).
[19] An extension of superstatistics for $N = 2$ was also considered by D. N. Sob'yanin, Phys. Rev. E **84**, 051128 (2011).
[20] I. R. Roa-González, B. C. Lima, P. I. R. Pincheira, A. A. Brum, A. M. S. Macêdo, G. L. Vasconcelos, L. S. Menezes, E. P. Raposo, A. S. L. Gomes, and R. Kashyap, Nat. Commun. **8**, 15731 (2017).
[21] C. Van den Broeck and M. Esposito, Phys. Rev. E **82**, 011144 (2010).
[22] T. Tomé and M. J. de Oliveira, Phys. Rev. E **82**, 021120 (2010).
[23] X. San Liang, Entropy **16**, 3605 (2014).
[24] S. Abe, Found. Phys. **44**, 175 (2014).
[25] A. Y. Abul-Magd, Phys. Rev. E **66**, 057104 (2002).
[26] S. Kotz and S. Nadarajah, *Extreme Value Distributions: Theory and Applications* (Imperial College Press, London, 2000).
[27] E. Bertin and M. Clusel, J. Phys. A **39**, 7607 (2006).
[28] A. M. Mathai, R. K. Saxena, and H. J. Haubold, *The H-Function: Theory and Applications* (Springer, New York, 2010).