Rank ordered beta distributions of nonlinear map symbolic dynamics families with a first-order transition between dynamical regimes

Roberto Alvarez-Martinez,[1,a)] Germinal Cocho,[2] and Gustavo Martinez-Mekler[3,4,5,b)]
[1]*Unidad de Microbiología Básica y Aplicada, Facultad de Ciencias Naturales,Universidad Autónoma de Querétaro, Carretera a Chichimequillas S/N, Ejido Bolaos, Qro. Codigo Postal 76140, Santiago de Quertaro, Mexico*
[2]*Instituto de Física, Universidad Nacional Autónoma de México, Apartado Postal 20-364, 01000 México, Distrito Federal, Mexico*
[3]*Instituto de Ciencias Físicas,Universidad Nacional Autónoma de México, Apartado Postal 48-3, 62251 Cuernavaca, Morelos, Mexico*
[4]*Centro Internacional de Ciencias, A.C. Avenida Universidad S/N, Apartado Postal 6-101, Cuernavaca 62131, Morelos, Mexico*
[5]*Centro de Ciencias de la Complejidad,Ciudad Universitaria, Universidad Nacional Autónoma de México, Circuito Mario de la Cueva 20, Insurgentes Cuicuilco, Apdo. Postal 20-364, 01000 México, Distrito Federal, Mexico*

(Received 06 March 2018; accepted 16 May 2018; published online 23 July 2018)

Rank-ordered distributions have been a matter of intense study. Often Zipf type invariant scaling is invoked; however, in the last decade the ubiquity of a Discrete Generalized Beta Distribution, DGBD, with two scaling exponents has been established. This distribution incorporates deviations from the power law at the extremes. A proper understanding of the meaning of these exponents is still lacking. Here, using two families of unimodal maps on the $[0, 1]$ interval, we construct binary sequences via standard symbolic dynamics. In both cases, the tent map, which is at the convex-concave border of the mapping families, separates intermittent regimes from chaotic dynamics. We show that the frequencies of n-tuples of the generated symbolic sequences are remarkably well fitted by the DGBD. We argue that in the underlying dynamics an order-disorder competition takes place and that one of the exponents is related to multiple range correlations, while the other is sensitive to disorder. In our study, we implement thermodynamic formalisms with which we can readily calculate n-tuple frequencies, in some particular cases, analytically. We show that for the convex mappings there is a first-order thermodynamic phase transition, while concave mappings have smooth free energy densities. Within our DGBD study, the transition between these two regimes coincides with a zero value for both exponents; in this sense, they may even be considered as indicators of the transition. An analysis of the difference between the exponents reinforces the interpretation we have assigned to them. Furthermore, the two regimes can be identified by the sign of such a difference. We also show that divergences in the invariant densities are responsible for the first order phase transitions observed in a range of the rank-frequency distributions. Our findings give further support to previous studies based on expansion-modification algorithms, birth-death processes, and random variable subtraction dynamics. *Published by AIP Publishing.* https://doi.org/10.1063/1.5027784

We study the probability distribution of symbolic strings of different lengths generated by the symbolic coding of unimodal maps taken from two families which include both the tent and logistic map as particular cases. When arranged in a decreasing order, these distributions, known as rank-frequency distributions, are determined by two exponents: exponent a, which we show can be associated to long-range correlations, and exponent b, related to disordered behavior. By examining the variation of these exponents with respect to the parameters of the studied unimodal families, we establish a correspondence between the difference $\Delta = (a - b) > 0$ and the prevalence of intermittency over chaos and the link between the prevalence of chaos, corresponding to $\Delta < 0$, and the existence of a first order phase transition in a thermodynamic model whose interaction potential is directly derived from the probability distribution of symbolic strings in the corresponding unimodal mapping.

a)Electronic mail: roberto.alvarez@uaq.mx
b)Electronic mail: mekler@icf.unam.mx

I. INTRODUCTION

Power law distributions are abundant in both natural and man-made phenomena and play an increasingly important role in the analysis of complex systems. In recent years, interest has escalated in them with developments in the study of networks.[7,19,26]

A particular instance which has attracted attention is rank-ordered distributions, where items (e.g., words, cities, wage-earners, electrical grids, forests, social interactions) are ordered decreasingly according to a numerical property (frequency, population, income, connectivity, area, etc.). Though these distributions are repeatedly reported to follow a power law, in many cases the power law is disputable and quite often it breaks down at the extremes.[11,25,33,38,62] Therefore, various power-law corrections have been suggested such as finite size scaling,[6,9,10] growth constraints,[4] and different kinds of cut-offs, e.g., stretched exponential,[27] Gaussian, and gamma.[4,23,56]

Recently, we proposed a functional form for rank-ordered distributions[35] that incorporates the deviations from power law behavior at the initial and end regions, which we named Discrete Generalized Beta Distribution (DGBD). Consider a set of items in descending order according to a numeric property denoted by f, such as frequency, size, connectivity, etc., and r the rank induced by this operation $r = 1, 2, 3 \ldots, N$, the DGBD is then defined by

$$f(r) = \mathcal{A}\frac{(N+1-r)^b}{r^a}, \tag{1}$$

where $\mathcal{A}$ is a scale constant (e.g., in some cases a normalization constant), a, b are fitting parameters, and N is the maximum number of elements.

In the DGBD, the scaling exponent b can be seen to control the curvature in a semi-log plot of $f(r)$ for large r values ($r \approx N$), while a determines the curvature for small values for $r \approx 1$. Notice that besides generalizing the Pareto power law distribution, when $a \neq 0$ and $b = 0$, the DGBD is also related to the Yule-Simon distribution[51,52,64] and recovers the Lavalette rank function when $a = b$,[28] which was recently related to the log-normal probability distribution.[24] A peculiar case is $a = b = 0$ which is the uniform distribution.[65]

The DGBD has been successfully applied to a wide range of phenomena, from motifs in arts,[35,48] impact factors,[34,50] rank-citation profiles of scientists,[41,42] sizes of communities in social networks,[58] human on-line activities,[63] energy flux and biomass in food networks,[22] baby name popularity,[31] distribution of h-index,[46,57] nuclear energy levels,[60] k-mer nucleotide distribution in the human genome,[32] popularity of programming languages,[59] Brexit results, and wealth distributions,[14–17] to mention a few.

Given the ubiquity of this distribution, the study of controllable processes that produce data which comply with it may contribute to the understanding and meaning of its parameters as well as the identification of relevant underlying characteristics of their genesis, all of which will contribute to the comprehension of phenomena leading to it. Previous work on these lines has been carried out, amongst others, in the context of expansion-modification models[3] and birth-death stochastic processes.[2,18] Here, we engage in this program by analyzing the symbolic dynamics of families of nonlinear unimodal mappings which include the tent and logistic maps as particular realizations. First, we show that frequency ranked n-tuples ensuing from symbolic sequences follow closely the DGBD. The possibility of studying the variation of the fitting exponents a, b, as we modify the family defining parameter, has enabled us to identify some features of the mappings relevant to the DGBD: (i) The concurrence of order-disorder conflicting dynamics, in a fluid dynamics context this can be paraphrased in terms of laminar flow—turbulent behavior dominance, (ii) The association of a with less disperse, correlated data and of b with more disperse, disordered data, (iii) The role of the difference $\Delta = a - b$ of the DGBD exponents in the determination of the transition between dynamical scenarios. There is value of the family parameter at which $\Delta = 0$ which signals a transition in the dynamics between intermittent and chaotic regimes, which coincides with a convexity inversion of the mappings. Furthermore, by means of a thermodynamical formalism, we show that the passage from negative to positive Δ is related to the onset of parameter range with first order phase transitions. This analysis relies on the absence or presence of singularities in the invariant measure of the mappings.

This paper is organized as follows: Section II introduces the two unimodal one-parameter families of maps we study, and derives some of their properties based on their symbolic dynamics. In Section III, we present a thermodynamic formalism for these maps and focus on invariant measures and the calculation of n-tuple rank distributions. In Section IV, thermodynamic phase transitions associated to the mappings and their relation to the rank-ordered behaviors are studied, with emphasis on divergences in invariant measures. Section V discusses the genericity regarding the association of invariant measures with first order phase transitions and GDBD exponent scenarios. A section on conclusions closes the paper.

II. SYMBOLIC DYNAMICS

A. Logistic map family

The first family we study of unimodal mappings on the unit interval is directly related to the logistic map.[37] These maps are characterized by two parameters: λ and ν. The first one controls the transition from an ordered dynamics to a chaotic regime via period-doubling bifurcations, whereas ν modulates the geometric properties of the maps, such as the maximum curvature, the Schwarzian derivative, and convexity.[20,21] These mappings are defined as follows:

$$f(x_n) = x_{n+1} = \lambda(1 - |1 - 2x_n|^\nu), \tag{2}$$

where $0 < \lambda \leq 1$, and in order to avoid the existence of two unstable fixed points $\frac{1}{2} \leq \nu < \infty$. For our analysis, we restrict the upper bound for ν to 2 and select $\lambda = 1$ to guarantee an ergodic regime over the full [0,1] interval of the maps defined by Eq. (2). See in Fig. 1 the bifurcation diagrams for several ν values.

It is straightforward that this family interpolates between some of the most iconic 1D maps by selecting special values of the parameter ν:

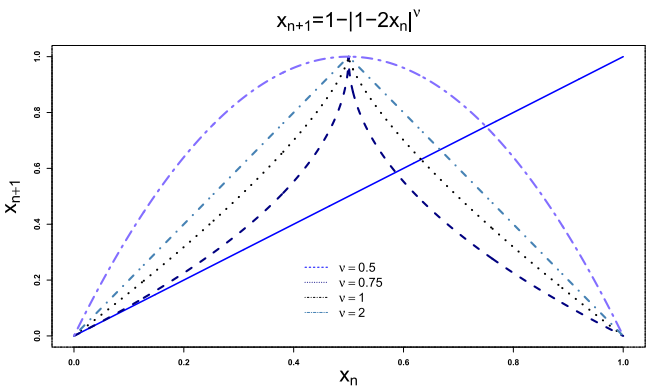

FIG. 1. Bifurcation diagram after 10 000 iterations of the logistic map family defined by Eq. (2) showing the effects of changes in the value of the parameter ν. Notice that in all the cases with ν above 1, the route to chaos via period doubling with the parameter λ encountered in Eq. (2) is preserved.

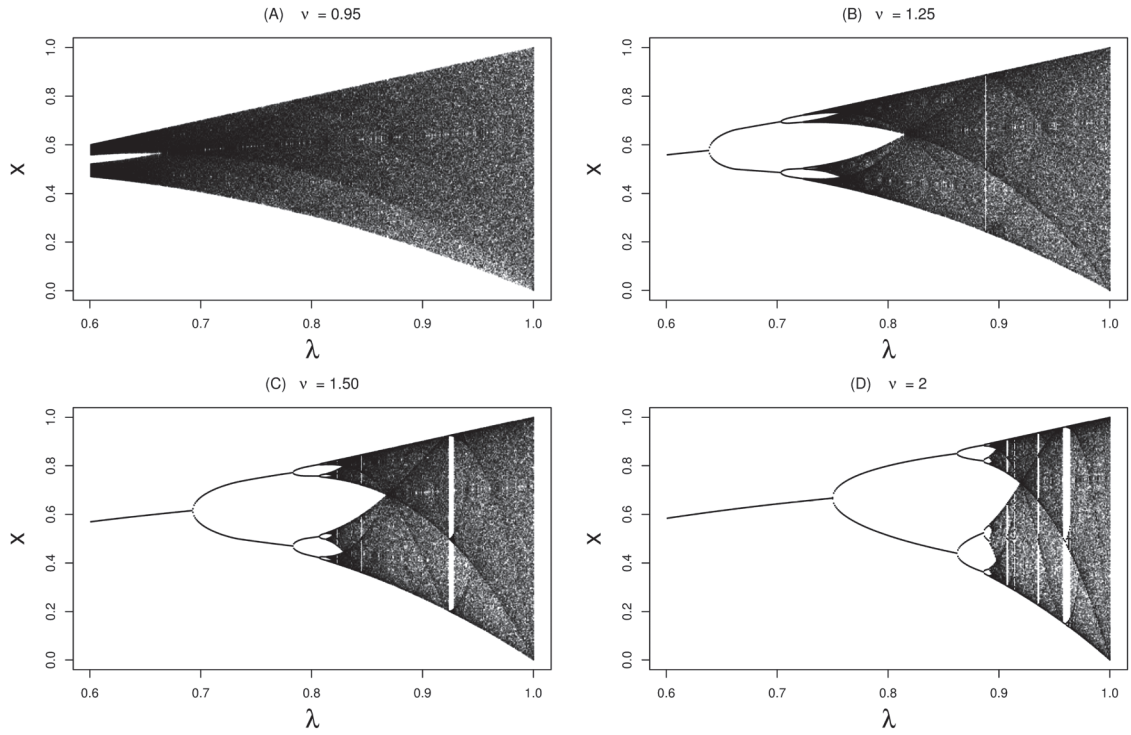

FIG. 2. Set of unimodal maps defined by Eq. (2). Note the convex-concave transition at the tent map with $\nu = 1$. Concavity produces intermittent behavior, $0.5 \leq \nu < 1$ is the normal form for type-I intermittency. Typical chaotic dynamics, as in the usual logistic equation, is generated when $\nu \geq 1$.

1. $\nu = \frac{1}{2}$, the map is tangent to the identity function at $x = 0$ (i.e., $f'(0) = 1$).
2. $\nu = 1$, $x_{n+1} = 1 - |1 - 2x_n|$, the tent map.
3. $\nu = 2$, $x_{n+1} = 1 - (1 - 2x_n)^2 = 4x_n(1 - x_n)$, the logistic map.

Figure 2 shows shape changes induced by changes in the parameter ν. The bottom, tangent map, is the normal form of intermittency-type I.[44] In this case, the "laminar" region prevails. In the top mapping with $\nu = 2$, corresponding to the logistic map, a chaotic regime prevails. Of special interest is the transition from concave to convex in the geometry of the mapping as $\nu = 1$ is crossed. Much of our work will be related to the consequences of this geometric transition on the dynamics. A thermodynamical formalism will be developed for this.

Our investigation focuses on the study of the left/right binary symbolic dynamics of these mappings, which we shall here after refer to simply as symbolic dynamics, and the rank-frequency distribution of n of its consecutive elements. Namely, we look into the frequency with which successive n-tuples (i.e., of non-overlapping groups of n consecutive elements) appear in decreasing order. The most frequent n-tuple is ranked in first place ($r = 1$), second place has ranking $r = 2$ and so on.

In a binary symbolic dynamics sequence, an n-tuple can have 2^n different configurations which we shall call n-mers. For example, in a given realization of a symbolic sequence $S = s_0 s_1 s_2 \ldots s_i s_{i+1} s_{i+2} \ldots$ where $s_i = 0$ or $1\ \forall\, i$, octamers are labeled by their decimal representation as shown below:

$$S = \underbrace{01000010}_{2+64=66}\ \underbrace{00111000}_{8+16+32=56}\ 0010\ldots$$

In Fig. 3, we display three rank-ordered distributions determined from 1×10^6 n-mers obtained from the standard symbolic dynamics of the unimodal maps defined by Eq. (2) for different values of ν.

This process is somewhat related to the expansion-modification systems proposed by Wentian Li,[30] where boolean sequences are generated via two stochastic processes that compete with each other: duplication that expands a sequence and tends to create long-range correlations, and modification that introduces changes that contribute to destroy them.

In order to fit the rank-frequency distribution with the DGBD, we estimate the parameters a, b via a multiple linear regression of the logarithm of Eq. (1), namely, $\log f(r) = \log A + b \log(N + 1 - r) - a \log r$. The coefficient A is an adjustable parameter that sets a scale, for normalized data it is a normalization constant. Fits are particularly good, for all the values of ν we considered, the square correlation coefficient R^2 was always ≥ 0.97.

Figure 3(a) shows that there are two clearly different regimes: the first one with $a > b$ and $\frac{1}{2} \leq \nu < 1$ (intermittent regime), and the second one, in which $b > a$ with $1 \leq \nu < 2$ (chaotic regime). The parameter ν controls the preponderance between two competing processes one favors permanence, the other change. This behavior is similar to the one previously reported in expansion-modification schemes.[3]

Notice that $a = b = 0$ when $\nu = 1, 2$; in Section III, we prove analytically that this coincidence is a consequence of the topological conjugacy between these cases.[1] In these cases, the DGBD distribution is uniform, for any given n.

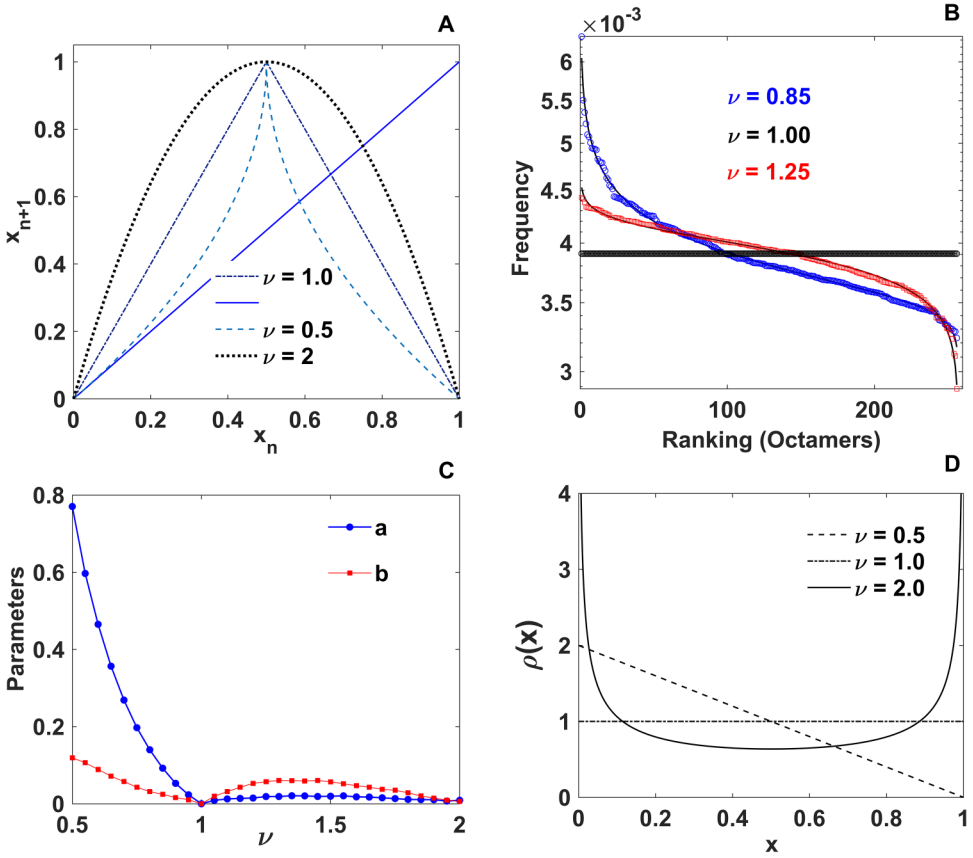

FIG. 3. (a) Unimodal maps defined by Eq. (2). This subplot shows the variation of the mappings with the parameter ν. The "laminar" phase is obtained with $0.5 \leq \nu < 1$, and the chaotic phase is generated with $\nu \geq 1$. (b) This panel shows the excellent agreement between the frequency-ordered 8-tuples generated via symbolic dynamics and the DGBD ($R^2 \geq 0.96$, for all cases). The cases plotted are $\nu = 0.85, 1, 1.25$. We have performed this operation also for $n = 6, 10, 12, 16$ with similar results. (c) This subplot shows the variation of the parameters (a, b) with parameter ν. (d) Histograms of empirical invariant densities for $\nu = 0.5, 1, 2$.

B. ϵ family

Here, we study another family of mappings determined by a linear combination of tent and logistic maps. The family is defined as follows:

$$x_{n+1} = 1 - \epsilon(1 - 2x_n)^2 + (\epsilon - 1)|1 - 2x_n|. \tag{3}$$

In this case, the parameter ϵ $(-\frac{1}{2} \leq \epsilon \leq 1)$ controls the relative weight of "laminar" and "turbulent" regimes.

- $\epsilon = -\frac{1}{2}$, intermittent map.
- $\epsilon = 0$, tent map.
- $\epsilon = 1$, logistic map.

An analysis for this family similar to the one performed for the logistic family [Eq. (2)] leads to results that resemble qualitatively those encountered previously. This suggests a certain degree of generality. From Fig. 4, we can see that ϵ controls the prevalence of either the intermittent or the chaotic regime.

It is known that intermittent maps,[5,43] as opposed to the chaotic ones, are particularly appropriate for the generation of long range correlation features. Figures 3 and 4 suggest that high values of parameter a may be an indication of the presence of long range correlations. On the other hand, values of b higher than those of a coincide with chaotic regimes.

III. THERMODYNAMIC FORMALISM

Though the following analysis can be applied to any mapping, for our purposes, from now on our study only refers to the logistic family of maps. For a characterization of the transition numerically observed between the intermittent and chaotic regimes, which as mentioned above coincides with a concavity reversal, we resort to the thermodynamic formalism for dynamical systems.[8] This approach also enables us to circumvent problems of the lack of convergence criteria in the simulations of rank-frequency distributions.[49] In particular, we can obtain some analytic results regarding differences in the behavior of the n-mers frequencies obtained from symbolic dynamics. We use this mathematical formalism to study the symbolic patterns generated by unimodal maps. Amongst others, with this method we can associate n-mers to "microstates" and their frequencies to the "energy" of the system. With this treatment, we can relate well-known thermodynamic properties (e.g., free energy and phase transitions) to dynamic features such as determination of invariant ergodic densities, n-mer distributions, and bifurcations.

The n-mers can be interpreted as n-spin configurations in a lattice magnetic system. Under this analogy, a Hamiltonian H of a n-mer $S = s_0 s_1 \ldots s_{n-1}$ can be defined[53] as follows:

$$H(s_0, s_1, \ldots, s_{n-1}) = -\ln p(s_0 s_1 \ldots s_{n-1}), \tag{4}$$

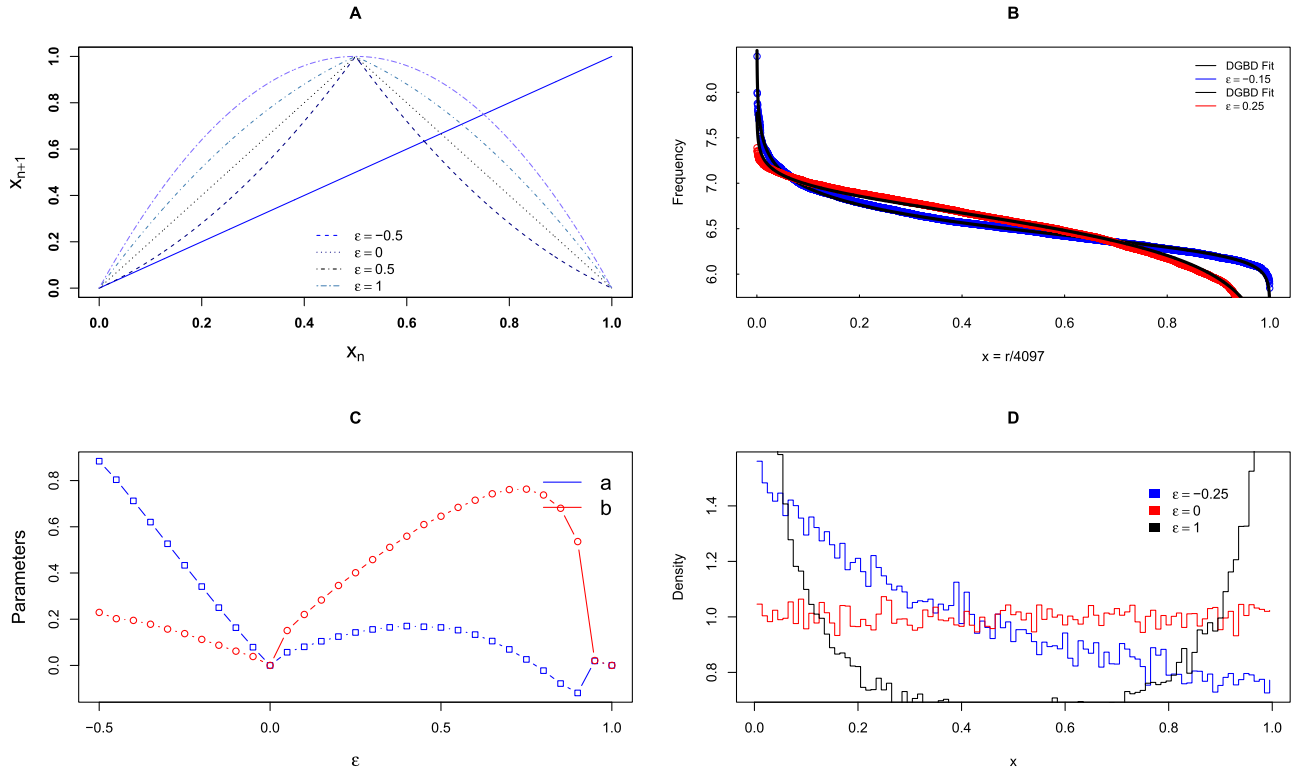

FIG. 4. (a) Unimodal maps defined by Eq. (3). This panel shows generic behaviors with changes in the parameter ϵ, i.e., the "laminar" phase is obtained with $0.5 \leq \epsilon < 0$, and the more chaotic phase is generated with $\epsilon \geq 0$. (b) This panel shows the excellent agreement of the frequency-ordered 12-tuples generated via symbolic dynamic with the DGBD ($R^2 \geq 0.97$, for all cases). We have performed this operation with $n = 6, 8, 10, 16$ obtaining similar results. (c) This panel shows the variation of the parameters (a, b) with parameter ϵ. (d) Histograms of empirical invariant density for $\epsilon = -0.25, 0, 1$ Again here the association of divergences to the transition from $a > b$ to $a < b$ is suggestive.

where $s_0, s_1, \ldots s_{n-1}$ is a symbolic sequence with $s_i = 0, 1$ $\forall\, i = 0, 1, \ldots, n-1$, and $p(s_0 s_1 \ldots s_{n-1})$ is the probability of the occurrence of a state the S n-mer.

A non-uniform occurrence of certain symbolic blocks in configuration space can be ascribed to interactions. The Hamiltonian of less (more) frequent symbolic sequences corresponds to less (more) frequent micro states [i.e., the value of hamiltonian is higher (lower)].

If we define escort distributions P_i by[8]

$$P_i = \frac{p_i^{\mu}}{\sum p_i^{\mu}}, \tag{5}$$

where p_i is the probability of encountering the ith-state and the parameter μ is an arbitrary real number. From Eqs. (4) and (5), we obtain

$$P_i = \frac{\exp^{-\mu H(s_0 s_1 \ldots s_{n-1})}}{\sum_{\{s_0 s_1 \ldots s_{n-1}\}} \exp^{-\mu H(s_0 s_1 \ldots s_{n-1})}}, \tag{6}$$

where the parameter μ plays the role of a Boltzmann's factor $\mu = \frac{1}{\kappa T}$. Note that the denominator of Eq. (6) has the functional form and role of a partition function:

$$Z = \sum_{\{s_0 s_1 \ldots s_{n-1}\}} \exp^{-\mu H(s_0 s_1 \ldots s_{n-1})}, \tag{7}$$

where the sum is over all the possible symbolic sequences of length n, hence we may write

$$P_i = \frac{e^{-\mu H(s_0 s_1 \ldots s_{n-1})}}{Z}, \tag{8}$$

On the other hand, for a given finite symbol sequence, $s_0, s_1, \ldots, s_{n-1}$—obtained from the symbolic dynamics of a map—the set denoted by $I[s_0 s_1 \ldots s_{n-1}]$, of initial conditions x_0 that generate this sequence is called an n-cylinder. Let $\rho(x)$ be the invariant density of a mapping the interval $[0, 1]$, the invariant measure of a n-cylinder, measures the probability $p(s_0, s_1, \ldots, s_{n-1})$ of the symbolic sequence.

$$p(s_0, s_1, \ldots, s_{n-1}) = \int_{I[s_0, s_1, \ldots, s_{n-1}]} \rho(x) dx, \tag{9}$$

or, in terms of Hamiltonian defined in Eq. (4)

$$H(s_0 s_1 \ldots s_{n-1}) = -\ln \int_{I[s_0 s_1 \ldots s_{n-1}]} \rho(x) dx. \tag{10}$$

A. Invariant density *ansatz*

In the above formalism, the knowledge of the invariant density of the maps is central. In general, smooth interval maps do not have nice invariant measures.[12] Given a map, it is difficult to obtain an invariant closed form analytic density. If a relevant invariant density exists,[13,29,55] in principle, it could be determined as a fixed point problem of the Frobenius-Perron operator. In practice, an approximate numerical approximation for the operator, based on an *ansatz* by Ulam, now known as the Ulam-Garlerkin method, is a good strategy worthwhile exploring. However, in this work, we rather propose and explore the following *ansatz* for the invariant densities of the family of maps defined by Eq. (2), which is a straightforward generalization of the logistic map

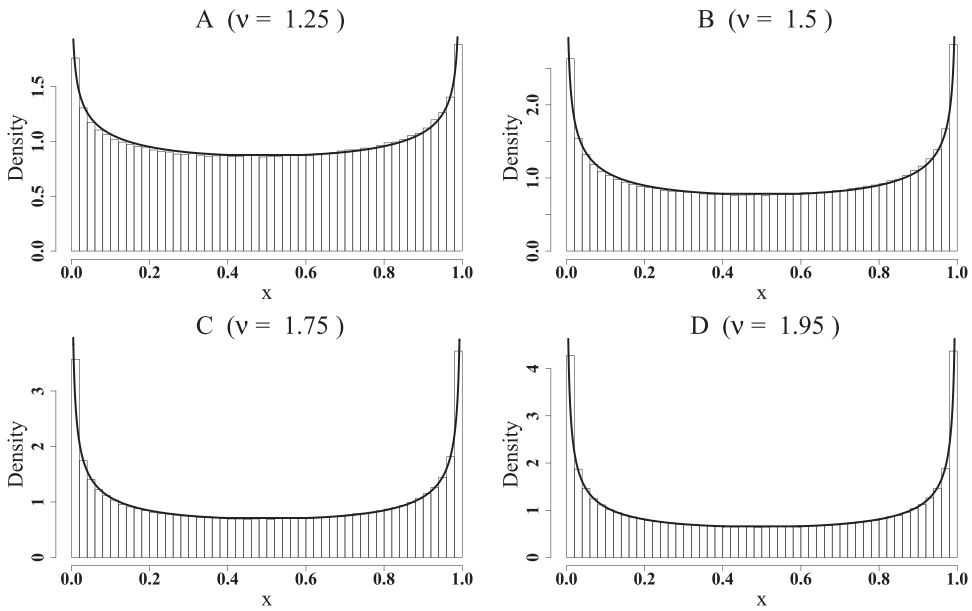

FIG. 5. Comparison of empirical histograms and *ansatz* results. Bars are the empirical histogram and the *ansatz* is represented by the solid black line. The cases $\nu = 1, 2$ are exact (not plotted here).

invariant density, $\rho_2(x) = \frac{1}{\pi\sqrt{x(1-x)}}$, that also recovers the tent map invariant density $\rho_1(x) = 1$, namely:

$$\rho_\nu(x) \equiv B(x; \frac{1}{\nu}, \frac{1}{\nu}) = A_\nu x^{1/\nu-1}(1-x)^{1/\nu-1}, \qquad (11)$$

where $B(x; \frac{1}{\nu}, \frac{1}{\nu})$ is the beta probability distribution function, $B(\nu, \nu)$ is the beta function, and $A_\nu = \frac{\Gamma(\frac{2}{\nu})}{\Gamma(\frac{1}{\nu})^2}$ the normalization constant. We assume that this density is valid for $1 \leq \nu \leq 2$. Figure 5 shows a comparison of the numerically determined histograms, from simulations after 100,000 iterations, with the *ansatz*, for several values of ν.

Using the Kolmogorov-Smirnov statistic test[36,54] and Monte Carlo simulations as proposed by Clauset for power laws,[11] we checked the goodness of the probability density distributions given by the *ansatz* and the empirical probability density distributions and found that the p-values are statistically significant for $1 \leq \nu \leq 2$. This demonstrates that the proposed invariant densities are very good approximations, even though they are not exact solutions. This is, to our knowledge, the first statistical analysis of the goodness of this *ansatz* for this family of unimodal maps.

B. Tent and logistic maps: Analytical results

The DGBD can be obtained analytically for the n-mers of the symbolic dynamics of both the tent ($\nu = 1$) and logistic ($\nu = 2$) maps. Since the invariant density of the tent map is $\rho_1(x) = 1$, the associated n-mer cylinders are $I(i) = [\frac{i}{2^n}, \frac{i+1}{2^n}]$, where $i = 0, 1, \ldots, 2^n - 1$ and the probability of sequence $s_0 s_1 \ldots s_{2^n-1}$ [Eq. (9)] is

$$P(s_0 s_1 \ldots s_{2^n-1}) = \int_{\frac{i}{2^n}}^{\frac{i+1}{2^n}} dx = \frac{1}{2^n}\int dx = \frac{1}{2^n}. \qquad (12)$$

Given the topological conjugacy between the logistic and tent map, by means of the homeomorphism $h(x) = \sin^2(\frac{\pi}{2}x)$, the cylinders for the logistic map are given by $L_i = [h(\frac{i}{2^n}), h(\frac{i+1}{2^n})] = [\sin^2(\frac{\pi j}{2^{n+1}}), \sin^2(\frac{\pi(i+1)}{2^{n+1}})]$, where $i = 0, 1, \ldots, 2^n - 1$, hence,

$$P(s_0 s_1, \ldots s_{2^n-1}) = \frac{1}{\pi}\int_{x_1=\sin^2(\frac{\pi i}{2^{n+1}})}^{x_2=\sin^2(\frac{\pi(i+1)}{2^{n+1}})} \frac{1}{\sqrt{x(1-x)}}dx$$

$$B_z(x_2, x_1, 1/2, 1/2) = \frac{1}{2^n}, \qquad (13)$$

where B_z is the incomplete beta function.[40]

We have thus shown analytically that for any n-mer of the symbolic dynamics of the tent and logistic maps, the exponents a and b defined by Eq. (1) are 0, the distribution is uniform with a constant value of $1/2^n$, this result is confirmed numerically.

IV. PHASE TRANSITIONS

The thermodynamics of dynamical systems[8] provides a framework to study the presence of phase transitions. If we choose a regular partition of the phase space of size $\delta = \frac{1}{2^n}$ and use for invariant measure the *ansatz* given by (11), then the probability p_i of encountering an n-mer of the logistic map family [Eq. (2)] is given by

$$p_i = A_\nu \int_{(i-1)\delta}^{i\delta} x^{\frac{1}{\nu}-1}(1-x)^{\frac{1}{\nu}-1}dx, \qquad (14)$$

where $i = 1, 2, \ldots, \frac{1}{\delta}$ and A_ν is a normalization constant and ν is restricted to the closed interval [1,2]. This probability can be expressed in terms of the incomplete beta function $B_z(\alpha, \beta)$[61]

$$B_z(\alpha, \beta) \equiv \frac{B(z; \alpha, \beta)}{B(\alpha, \beta)} \equiv A_\nu \int_0^z t^{\alpha-1}(1-t)^{\beta-1}dt,$$

as

$$p_i = A_\nu\left(B_{i\delta}\left(\frac{1}{\nu}, \frac{1}{\nu}\right) - B_{(i-1)\delta}\left(\frac{1}{\nu}, \frac{1}{\nu}\right)\right), \qquad (15)$$

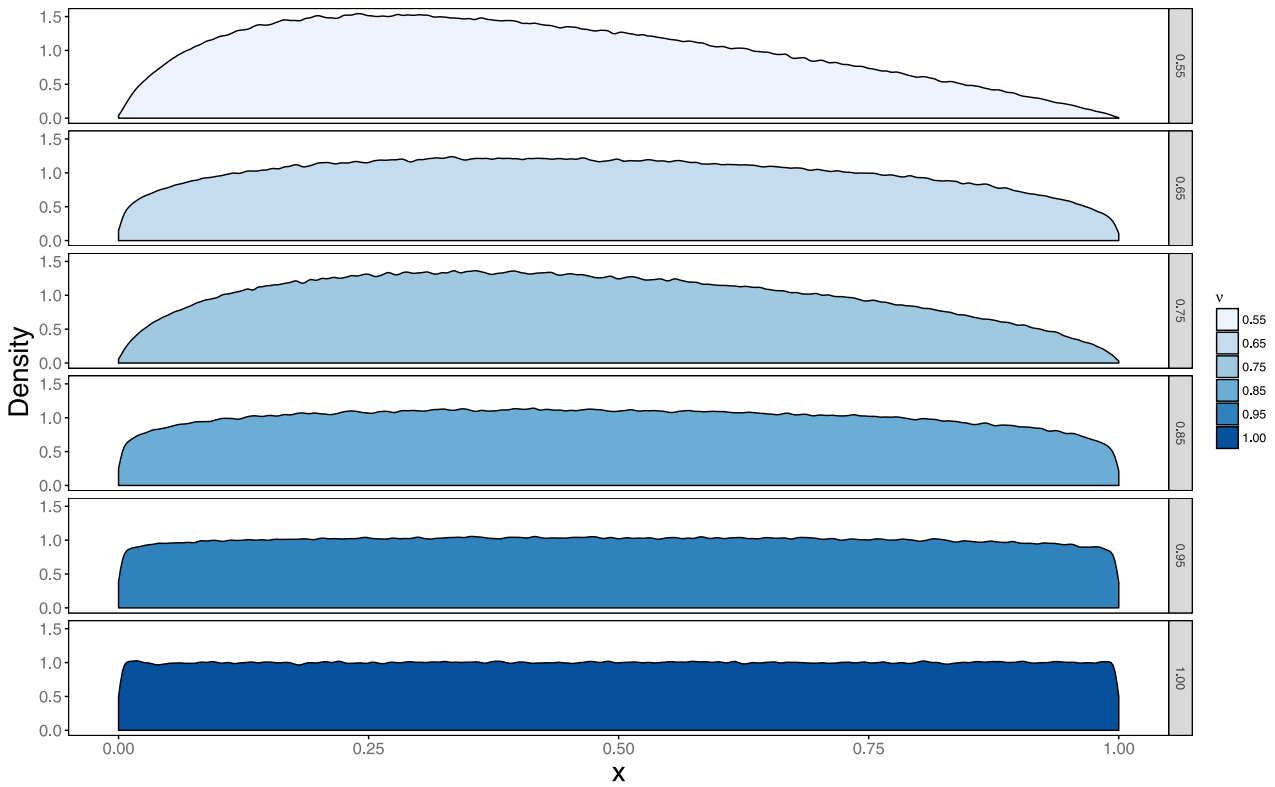

FIG. 6. Empirical densities (histograms) from maps defined by Eq. (2) for values in $0.5 < \nu \leq 1$. It is worth noting that for $\nu = 0.5$ and $\nu = 1$ analytical closed forms are available for invariant densities $\rho(x) = 2(1-x)$ and $\rho(x) = 1$. This figure shows the absence of divergences of invariant densities in contrast to $\nu > 1$. These histograms were performed over 100 different initial conditions and 1000 iterations each one the bin width is equal to 0.01.

and the partition function is

$$Z = \sum_{i=1}^{\frac{1}{\delta}} p_i^{\mu} = (A_\nu)^{\mu} \sum_{i=1}^{\frac{1}{\delta}} \left[B_{i\delta}\left(\frac{1}{\nu}, \frac{1}{\nu}\right) - B_{(i-1)\delta}\left(\frac{1}{\nu}, \frac{1}{\nu}\right) \right]^{\mu}. \tag{16}$$

For $\nu \in (1, 2]$ due to the divergence of the invariant density function at its extremes, it follows for small δ that

$$p_1 = B_\delta\left(\frac{1}{\nu}, \frac{1}{\nu}\right) \approx \delta^{\frac{1}{\nu}}, \tag{17}$$

$$p_{\frac{1}{\delta}} = 1 - B_{1-\frac{1}{\delta}}\left(\frac{1}{\nu}, \frac{1}{\nu}\right) \approx (1-\delta)^{\frac{1}{\nu}}. \tag{18}$$

This result agrees with the divergence exponents calculated for the invariant measures of the mappings based on an analysis of their behavior in neighborhood around the maximum when $1 < \nu \leq 2$.[39] Substituting Eqs. (17) and (18) in (16) and taking into account the fact that for almost all i at the central support of the invariant density, $p_i \approx \delta$, we obtain

$$\begin{aligned} Z &\approx (A_\nu)^{\mu} \left\{ 2\delta^{\frac{\mu}{\nu}} + \sum_{i=2}^{\frac{1}{\delta}-1} \left[B_{i\delta}\left(\frac{1}{\nu}, \frac{1}{\nu}\right) - B_{(i-1)\delta}\left(\frac{1}{\nu}, \frac{1}{\nu}\right) \right]^{\mu} \right\}, \\ &\approx (A_\nu)^{\mu} \left[2\delta^{\frac{\mu}{\nu}} + \left(\frac{1}{\delta} - 2\right)\delta^{\mu} \right] \approx \left(\delta^{\frac{\mu}{\nu}} + \delta^{\mu-1}\right), \end{aligned} \tag{19}$$

where we recall that μ is an arbitrary parameter which plays the role of an inverse physical temperature and $2^n \gg 1$. From (19), it follows that the "static"[8] Helmholtz free energy density $\varphi(\mu, \nu)$ in the $\delta \to 0$ thermodynamic limit is given by

$$\varphi(\mu, \nu) = \begin{cases} 1 - \frac{1}{\mu}, & \mu < \mu_c, \\ \frac{1}{\nu}, & \mu \geq \mu_c, \end{cases} \tag{20}$$

where $\mu_c = \frac{\nu}{\nu - 1}$.

For $\nu > 1$, $\varphi(\mu, \nu)$ has a discontinuity at μ_c that produces a first order phase transition. It is a piece-wise linear function, with derivatives from left and right that lead to a step-function in $\partial\phi(\mu, \nu)/\partial\mu$. As $\nu \to 1$, $\mu_c \to \infty$, hence only one of the functional parts of φ is accessible and consequently the discontinuity disappears. Notice that μ_c going to infinity corresponds to a critical temperature T_c going to zero. As a function of temperature, the free energy density is linear with respect to T with a negative slope and a finite constant value at the $\nu \to \infty$ limit. This latter result is consistent with the $\nu = 1$ calculation based on the exact finding of Eq. (12) that φ is constant for all T.

For $\nu \in (0.5, 1)$, the absence of divergences in the numerical calculations of invariant measures shown in Fig. 6, together with the exact result $\rho = 2(1-x)$ for $\nu = 0.5$ (obtainable from the Perron-Frobenius operator), produces a smooth behavior of the energy density ϕ with no discontinuities, divergences, and hence no phase transitions. We hence have found that at tent map value ($\nu = 1$), a change in curvature of the family of logistic maps, signaling a transition between intermittency and chaos, produces a transition from a range of maps exhibiting thermodynamic first order phase transitions to maps with smooth thermodynamic behavior. For our rank-ordered investigation, it is interesting to remark that

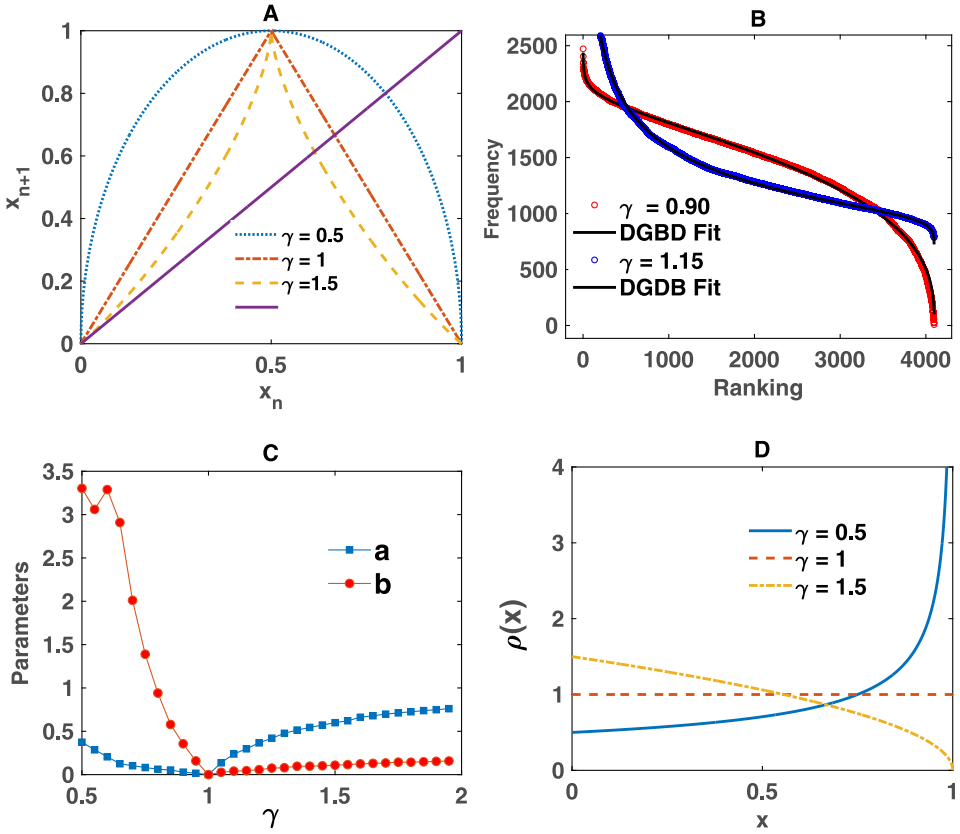

FIG. 7. (a) Family of unimodal maps defined by Eq. (23). This subplot shows the mapping changes for several values of the parameter. (b) DGBD fitting from symbolic dynamics with $n = 12$, 12-mers, this figure shows two values of $\gamma = 0.90$ and $\gamma = 1.15$ one before the tent map and the other after the tent map, illustrating the two different behaviors: $a < b$ and $a > b$, respectively. (c) This subplot shows the variation of the parameters (a, b) with parameter γ. Notice the resemblance with the behavior shown in Figs. 3 and 4. (d) This panel exhibits the invariant measure behavior from the smooth $\gamma > 1$ cases to the divergent case of $\gamma \leq 1$ with singularity at $x = 1$.

the above takes place at the exponent values $a = b = 0$. Furthermore, since, as shown in Fig. 2, at this point the difference $\Delta = a - b$ changes sign, the order-disorder interpretation of a and b is reinforced.

V. ON THE OCCURRENCE OF THE LOGISTIC FAMILY *a, b* PARAMETER SPACE RELATION

Here, we investigate how recurrent is the a, b pattern we have encountered in Figs. 3(c) and 4(c). Our current results rely heavily on the divergence of the mapping invariant measures. In this section, we show that the presence of these divergences is a sufficient condition for such a behavior to arise.

Consider a simple family of invariant density measures with only one divergence at one of its extremes:

$$\rho(x) = \gamma(1 - x)^{\gamma - 1}, \tag{21}$$

if $\gamma < 1$, the existence of a divergence is guaranteed, when $\gamma > 1$ there are no singularities and the case $\gamma = 1$ is the tent map. We explore $0.5 \leq \gamma$ to study the transition between the absence and presence of divergences.

In order to determine a family of mappings that produce these measures, we implement the inverse Perron-Frobenius procedure. For $\rho(x)$, the inverse Perron-Frobenius equation is given by[43]

$$\frac{dy}{dx} = \frac{x^{\gamma - 1} + (1 - x)^{\gamma - 1}}{(1 - y)^{\gamma - 1}}, \tag{22}$$

and the corresponding family of unimodal maps are

$$y = f(x_n) = x_{n+1} = 1 - |x_n^{\gamma} - (1 - x_n)^{\gamma}|^{\frac{1}{\gamma}}. \tag{23}$$

In Fig. 7, we show that indeed, the behavior of this family of mappings is very similar to the one we encountered for the logistic and ϵ families. Notice, however, as shown in subplot A of Fig. 7, that the maps generated by the prescribed invariant densities Eq. (21) show a reverse dependence on γ with respect to the one displayed on ν in Figs. 3 and 4.

The fitting value diagrams illustrates that the mappings (23) have the same type of transition from a regime dominated by the intermittency to the one where a chaotic regime prevails. This change now occurs when the parameter γ is equal to one. There are two regimes, and changes in the exponent difference Δ appear to have the same predictive value as before. Furthermore, a thermodynamic study gives similar results as with the logistic family, giving support to the proposal that divergences in invariant densities are a sufficient condition for first order phase transitions.

VI. CONCLUSIONS

The explanation and understanding of why there is such an abundance of phenomena, where the DGBD gives excellent fits to rank-ordered data, is challenging. Acquiring knowledge of the meaning of the distribution exponents (a, b) is one of the main issues. Here, we have approached this task by looking into distributions of n-mers in symbolic dynamic sequences of families of unimodal maps. As a starting point, we have shown numerically and in some cases analytically that DGBD fits are outstanding. For three different families, we obtain a recurrent pattern in the variation of the values of the exponents as the mapping defining parameter is changed, the difference $\Delta = a - b$ gradually diminishes goes to 0 and increases (Figs. 3, 4, and 7). By construction, in all

cases the tent map marks a convex to concave divide in the mappings. In the logistic family, this has been typified as a transition from an intermittent dynamics to a chaotic dynamics, in fluid representations the transition is between a laminar dominated flow and a turbulent flow. The underlying feature is an order-disorder transition. When a conflicting dynamics can be defined, the exponent a is an ordering factor related to less dispersion, higher correlated fluctuations, while b is more linked to disorder, higher dispersion, and large fluctuation components.

It is interesting to note that the order-disorder phenomenon appears at two distinct levels. At the individual mapping level, this can appear as a competition between sensitive dependent expanding dynamics and unpredictable reentrant dynamics essential for chaos, or it may surface as a jammed dynamics due to tangency which induces laminar evolution and escape episodes introducing disorder situation typical of intermittent behavior. At the family level, the transition is between intermittent-chaotic transitions mentioned above.

For our study, we have taken full advantage of the thermodynamic approach to dynamical systems implemented via the symbolic dynamics. In this formalism, n-mers can be identified with thermodynamic states and the probability of their occurrence is associated with the weight of their cylinders. With this approach, we have been able to find exact n-mer distributions for particular cases. Moreover, with the thermodynamic formalism, together with an invariant density *ansatz*, we have been able to relate the intermittency-chaos transition to the vanishing of the occurrence of first order thermodynamic phase transitions within an interval of the mapping family parameters. This takes place at the tent map after which an interval of smooth static free energies settles in. From the DGBD outlook, this happens when $\Delta = 0$, it and might be associated with a balance between disordering and ordering factors. Though the exponent interpretation ensuing from this study reinforces previous studies,[2,3,47] we clarify that we are not proposing that this is a necessary condition for a good DGBD fit.

ACKNOWLEDGMENTS

R.A.M. would like to thank Consejo Nacional de Ciencia y Tecnologia (CONACYT) for postdoctoral grant 105368 and research project (UAQ) FNB-2016-12. G.M.M. and G.C. also acknowledge support from DGAPA-UNAM-IN115908 at initial stages of this investigation. We would also like to express our gratitude to a referee for his/her comments and suggestions which improved our work considerably and widened our horizons.

[1]K. T. Alligood, T. D. Sauer, and J. A. Yorke, *One-Dimensional Maps* (Springer, 1997).
[2]R. Alvarez-Martinez, G. Cocho, R. Rodríguez, and G. Martínez-Mekler, "Birth and death master equation for the evolution of complex networks," Physica A **402**, 198–208 (2014).
[3]R. Alvarez-Martinez, G. Martinez-Mekler, and G. Cocho, "Order–disorder transition in conflicting dynamics leading to rank–frequency generalized beta distributions," Physica A **390**, 120–130 (2011).
[4]L. Amaral, A. Scala, M. Barthélémy, and H. Stanley, "Classes of small-world networks," Proc. Natl. Acad. Sci. **97**, 11149–11152 (2000).
[5]R. Artuso and A. Prampolini, "Correlation decay for an intermittent area-preserving map," Phys. Lett. A **246**, 407–411 (1998).
[6]M. Ausloos and R. Cerqueti, "A universal rank-size law," PLoS ONE **11**, e0166011 (2016).
[7]A.-L. Barabási and R. Albert, "Emergence of scaling in random networks," Science **286**, 509–512 (1999).
[8]C. Beck and F. Schögl, *Thermodynamics of Chaotic Systems: An Introduction* (Cambridge University Press, 1995), Vol. **4**.
[9]J. L. Cardy, "Conformal invariance and universality in finite-size scaling," J. Phys. A **17**, L385 (1984).
[10]J. Chayes, L. Chayes, D. S. Fisher, and T. Spencer, "Finite-size scaling and correlation lengths for disordered systems," Phys. Rev. Lett. **57**, 2999 (1986).
[11]A. Clauset, C. Shalizi, and M. Newman, "Power-law distributions in empirical data," SIAM Rev. **51**, 661–703 (2009).
[12]W. De Melo and S. Van Strien, *One-Dimensional Dynamics* (Springer Science & Business Media, 2012), Vol. **25**.
[13]J. Ding, "Absolutely continuous invariant measure of a piecewise concave mapping of [0, 1]," Nonlin. Anal. Theory Methods Appl. **28**, 1133–1140 (1997).
[14]I. Eliazar, "Growth and inequality," Physica A **437**, 457–470 (2015).
[15]I. Eliazar, "Black swans and dragon kings: A unified model," Europhys. Lett. **119**, 60007 (2017a).
[16]I. Eliazar, "Lindy's law," Physica A **486**, 797–805 (2017b).
[17]I. Eliazar and M. H. Cohen, "Hierarchical socioeconomic fractality: The rich, the poor, and the middle-class," Physica A **402**, 30–40 (2014).
[18]I. I. Eliazar and M. H. Cohen, "A Langevin approach to the log–gauss–pareto composite statistical structure," Physica A **391**, 5598–5610 (2012).
[19]M. Faloutsos, P. Faloutsos, and C. Faloutsos, "On power-law relationships of the internet topology," in *ACM SIGCOMM Computer Communication Review* (ACM, 1999), Vol. 29, pp. 251–262.
[20]M. J. Feigenbaum, "The metric universal properties of period doubling bifurcations and the spectrum for a route to turbulence," Ann. N. Y. Acad. Sci. **357**, 330–336 (1980).
[21]M. J. Feigenbaum, "Universal behavior in nonlinear systems," Physica D **7**, 16–39 (1983).
[22]Y. Feng and J. Zhang, "Universal biomass and energy flow distribution in weighted food webs," Procedia Environ. Sci. **13**, 818–828 (2012).
[23]T. Fenner, M. Levene, and G. Loizou, "A model for collaboration networks giving rise to a power-law distribution with an exponential cutoff," Soc. Netw. **29**, 70–80 (2007).
[24]O. Fontanelli, P. Miramontes, Y. Yang, G. Cocho, and W. Li, "Beyond Zipf's law: The lavalette rank function and its properties," PLoS ONE **11**, e0163241 (2016).
[25]M. L. Goldstein, S. A. Morris, and G. G. Yen, "Problems with fitting to the power-law distribution," Eur. Phys. J. B **41**, 255–258 (2004).
[26]H. Jeong, B. Tombor, R. Albert, Z. N. Oltvai, and A.-L. Barabási, "The large-scale organization of metabolic networks," Nature **407**, 651–654 (2000).
[27]J. Laherrere and D. Sornette, "Stretched exponential distributions in nature and economy: Fat tails with characteristic scales," Eur. Phys. J. B **2**, 525–539 (1998).
[28]D. Lavalette, "Facteur dimpact: Impartialité ou impuissance," Orsay (France): Institut CurieRecherche, Bât 112 (1996).
[29]T.-Y. Li, "Finite approximation for the Frobenius-Perron operator. A solution to Ulam's conjecture," J. Approx. Theory **17**, 177–186 (1976).
[30]W. Li, "Expansion-modification systems: A model for spatial 1/f spectra," Phys. Rev. A **43**, 5240 (1991).
[31]W. Li, "Analyses of baby name popularity distribution in us for the last 131 years," Complexity **18**, 44–50 (2012).
[32]W. Li, J. Freudenberg, and P. Miramontes, "Diminishing return for increased mappability with longer sequencing reads: Implications of the k-mer distributions in the human genome," BMC Bioinformatics **15**, 2 (2014).
[33]G. Lima-Mendez and J. van Helden, "The powerful law of the power law and other myths in network biology," Mol. Biosyst. **5**, 1482–1493 (2009).
[34]R. Mansilla, E. Köppen, G. Cocho, and P. Miramontes, "On the behavior of journal impact factor rank-order distribution," J. Informetr. **1**, 155–160 (2007).
[35]G. Martínez-Mekler, R. Alvarez-Martínez, M. Beltran-del-Río, R. Mansilla, P. Miramontes, and G. Cocho, "Universality of rank-ordering distributions in the arts and sciences," PLoS ONE **4**, e4791 (2009).

[36]F. J. Massey, Jr., "The Kolmogorov-Smirnov test for goodness of fit," J. Am. Stat. Assoc. **46**, 68–78 (1951).
[37]H. Nagashima and Y. Baba, *Introduction to Chaos: Physics and Mathematics of Chaotic Phenomena* (CRC Press, 1998).
[38]M. Newman, "Power laws, pareto distributions and Zipf's law," Contemp. Phys. **46**, 323–351 (2005).
[39]E. Ott, *Chaos in Dynamical Systems* (Cambridge University Press, 2002).
[40]K. Pearson and G. Britain, *Tables of the Incomplete Beta-function* (University Press Cambridge, 1968).
[41]A. M. Petersen, H. E. Stanley, and S. Succi, "Statistical regularities in the rank-citation profile of scientists," Sci. Rep. **1**, 181 (2011).
[42]A. M. Petersen and S. Succi, "The *Z*-index: A geometric representation of productivity and impact which accounts for information in the entire rank-citation profile," J. Informetr. **7**, 823–832 (2013).
[43]D. Pingel, P. Schmelcher, and F. Diakonos, "Theory and examples of the inverse Frobenius–Perron problem for complete chaotic maps," Chaos **9**, 357–366 (1999).
[44]P. Manneville and Y. Pomeau, "Intermittent transition to turbulence in dissipative dynamical systems," Commun. Math. Phys. **74**, 189–197 (1980).
[45]I.-I. Popescu, M. Ganciu, M.-C. Penache, and D. Penache, "On the Lavalette ranking law," Rom. Rep. Phys. **49**, 003–028 (1997).
[46]M. R. Quigley, E. B. Holliday, C. D. Fuller, M. Choi, and C. R. Thomas, Jr., "Distribution of the h-index in radiation oncology conforms to a variation of power law: Implications for assessing academic productivity," J. Cancer Educ. **27**, 463–466 (2012).
[47]M. Beltrán-del Río, G. Cocho, and R. Mansilla, "General model of subtraction of stochastic variables. Attractor and stability analysis," Physica A **390**, 154–160 (2011).
[48]M. Beltrán-del Río, G. Cocho, and G. Naumis, "Universality in the tail of musical note rank distribution," Physica A **387**, 5552–5560 (2008).
[49]R. Salgado-García and E. Ugalde, "Exact scaling in the expansion-modification system," J. Stat. Phys. **153**, 842–863 (2013).
[50]J. M. Sarabia, F. Prieto, and C. Trueba, "Modeling the probabilistic distribution of the impact factor," J. Informetr. **6**, 66–79 (2012).
[51]H. Simon, "On a class of skew distribution functions," Biometrika 425–440 (1955).
[52]H. A. Simon, "Models of man; social and rational" (Wiley, 1957).
[53]Y. G. Sinai, "Gibbs measures in ergodic theory," Russ. Math. Surv. **27**, 21 (1972).
[54]N. Smirnov, "Table for estimating the goodness of fit of empirical distributions," Ann. Math. Stat. **19**, 279–281 (1948).
[55]A. C. Smith, "Using Ulam's method to test for mixing," Ph.D. thesis (University of Florida, 2010).
[56]D. Sornette, L. Knopoff, Y. Kagan, and C. Vanneste, "Rank-ordering statistics of extreme events: Application to the distribution of large earthquakes," J. Geophys. Res. **101**, 13883–13893 (1996).
[57]C. M. Spearman, M. J. Quigley, M. R. Quigley, and J. E. Wilberger, "Survey of the h index for all of academic neurosurgery: Another power-law phenomenon? Clinical article," J. Neurosurg. **113**, 929–933 (2010).
[58]M. Šuvakov, M. Mitrović, V. Gligorijević, and B. Tadić, "How the online social networks are used: Dialogues-based structure of myspace," J. R. Soc. Interface **10**, 20120819 (2013).
[59]S. Valverde and R. V. Solé, "A cultural diffusion model for the rise and fall of programming languages," Hum. Biol. **87**, 224–234 (2015).
[60]V. Velázquez, E. Landa, C. Vargas, R. Fossion, J. López-Vieyra, I. Morales, and A. Frank, "Self similitude in the power spectra of nuclear energy levels," *J. Phys. Conf. Ser.* (IOP Publishing, 2013), Vol. 475, p. 012014.
[61]E. W. Weisstein, "Incomplete beta function. from mathworld—A wolfram web resource," see http://mathworld.wolfram.com/IncompleteBetaFunction.html; accessed 18 March 2016.
[62]E. P. White, B. J. Enquist, and J. L. Green, "On estimating the exponent of power-law frequency distributions," Ecology **89**, 905–912 (2008).
[63]L. Wu and J. Zhang, "Accelerating growth and size-dependent distribution of human online activities," Phys. Rev. E **84**, 026113 (2011).
[64]G. Yule, "A mathematical theory of evolution, based on the conclusions of Dr. J. C. Willis, F. R. S," Philos. Trans. R. Soc. Lond. Ser. B **213**, 21–87 (1925).
[65]D. Zwillinger, *CRC Standard Mathematical Tables and Formulae* (CRC Press, 2002).