Symbolic Sequences and Tsallis Entropy

H. V. Ribeiro,* E. K. Lenzi, and R. S. Mendes

Departamento de Física and National Institute of Science and Technology for Complex Systems,
Universidade Estadual de Maringá
Av. Colombo, 570 - 87020-900 Maringá, PR, Brazil

G. A. Mendes and L. R. da Silva

Departamento de Física and National Institute of Science and Technology for Complex Systems,
Universidade Federal do Rio Grande do Norte
59072-970 Natal, RN, Brazil

We address this work to investigate symbolic sequences with long-range correlations by using computational simulation. We analyze sequences with two, three and four symbols that could be repeated l times, with the probability distribution $p(l) \propto 1/l^{\mu}$. For these sequences, we verified that the usual entropy increases more slowly when the symbols are correlated and the Tsallis entropy exhibits, for a suitable choice of q, a linear behavior. We also study the chain as a random walk-like process and observe a nonusual diffusive behavior depending on the values of the parameter μ.

Keywords: Symbolic sequences, Long-range correlations, Tsallis entropy, Non-usual diffusion.

1. INTRODUCTION

Two basic assumptions of the statistical mechanics are: the "equal a priori probabilities" and ergodicity. When these assumptions do not hold, we need other suitable tools to study systems which exhibit a nonusual behavior. A typical situation can be found by analyzing systems which have an intermediate regime between periodic and chaotic [1]. This kind of system commonly shows a power law spectra and appears in several fields of science. Aspects of nonusual behavior have been explored, for instance, in biology[2], nuclear physics [3], financial market [4], music [5] and linguistics [6]. In this context, there are also works that search for correlations in DNA sequences [7–11] by using entropic indexes [12–16]. To provide a possible description for these systems which are not conveniently explained by the usual formalism, Tsallis[17] proposes an extension of the Boltzmann-Gibbs entropy. Many systems have been investigated by using this approach, e.g., long-range Hamiltonian systems like the HMF model [18], the generalized Lennard-Jonnes gas [19], self-gravitating systems [20] and anomalous diffusion [21].

In this direction, to try to clarify in a more direct way basic aspects related to Tsallis entropy, it may be convenient to consider specifc models with a kind of long-range behavior. Considering this, the aim of this work is to explore the nonusual behavior of a symbolic model with an adjustable long-range behavior. More precisely, we investigate one dimensional symbolic sequences with long-range correlations which are generated by using the numerical experiment presented in Ref. [22]. The procedure uses two random numbers to obtain a lattice with N sites which represent the symbolic sequence. One of them, x, has a uniform distribution in the interval [0,1] and the other emerges from the expression

$$y = A \left[\frac{1}{(1-x)^{1/(\mu-1)}} - 1 \right] , \qquad (1)$$

where A and μ are real parameters. We go through the symbolic sequence drawing x and filling $N_y = [y] + 1$ sites with the same value z, where $[y]$ denotes the integer part of y and z is a signal generator that can have one of four distinct values (0, 1, 2, 3) with the same statistical weight. A typical example obtained within this procedure is

$$Q = \left\{ \underbrace{0,0,0,}_{N_y=3} \underbrace{1,1,1,1,}_{N_y=5} \underbrace{0,0,}_{N_y=2} \underbrace{1,1,1,1,}_{N_y=4} \underbrace{0}_{N_y=1} \right\}$$

for a sequence with two symbols.

For the sequences generated with the procedure described above, we may obtain the probability distribution function of the variable y, $p(y)$, and show that, depending on the values of the μ, it can be asymptotically related to a Lévy distribution (for $y \geq 0$). In fact, after some calculations, one can show that $p(y)$ is given by

$$p(y) = (\mu - 1) \frac{A^{\mu-1}}{(A+y)^{\mu}} , \qquad (2)$$

and the first moment of this distribution is $\langle y \rangle = A/(\mu - 2)$. By comparing the asymptotic limit of Eq.(2), $p(y) \sim 1/y^{\mu}$, with the asymptotic limit of the Lévy distributions, $p(y) \sim 1/y^{1+\eta}$, the relation between μ and η is $\mu = 1 + \eta$. Note also that $\langle y \rangle$ diverges for $\mu \to 2$. This fact indicates that, when μ is close to two, N_y may assume large values and fill a large part of the symbolic sequence with the same symbol. On the other hand, when μ is far from two ($\mu \gg 2$), large values of N_y become very rare and consequently the sequence has more alternated symbols.
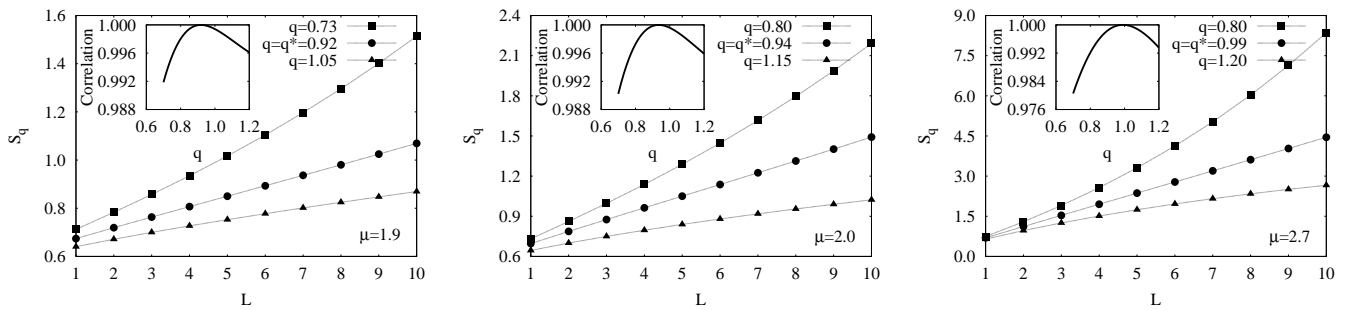
*Electronic address: hvr@dfi.uem.br

FIG. 1: S_q versus L for some values of μ (indicated in the figure) for a two-symbol sequence. We use $A = 2$ and $N = 10^8$ in all the three figures.

2. ENTROPY AND SEQUENCE

The Tsallis entropy is defined, for a system with W microstates and occupation probabilities p_i, as follows:

$$S_q = \frac{1 - \sum_{i=1}^{W} p_i^q}{q - 1}, \qquad (3)$$

where q is a real parameter. In the limit $q \to 1$ we have the standard Boltzmann-Gibbs entropy. S_q is extensive for a composite system consisting of independent subsystems for $q = 1$ and nonextensive for $q \neq 1$; for this reason, S_q is sometimes referred to as nonextensive entropy. However, when we have long-range interactions or long-range correlations, the subsystems cannot be independent. In this case we will see that S_q can be extensive for a particular value of $q \neq 1$.
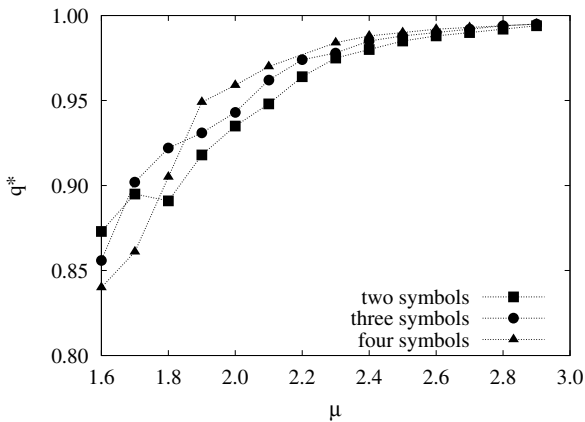

FIG. 2: The entropic index $q*$ versus μ for two, three and four-symbol sequences, with $A = 2$ and $N = 10^8$.

In order to evaluate the Tsallis entropy, for the symbolic sequence generated with the previous procedure, we fix windows of length L which are moved along the sequence. Then, we count how many times a given configuration (string) occurs, determining the probability p_i of a specific configuration i. To illustrate this procedure, suppose that we have the following sequence:

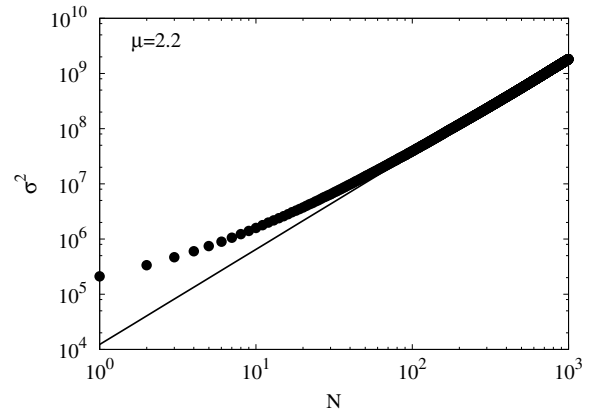$$Q = \{0, 0, 1, 0, 0, 1, 1, 1, 0, 0, 0, 1, 0, 0, 0, 0, 1, 0, 1, 1\},$$

FIG. 3: The standard-deviation versus N for $\mu = 2.2$ and $A = 0.1$.

then we fix a window of length 2 and move it along the sequence, i.e., we have

$$Q = \left\{ \underbrace{0, 0,}_{1} \underbrace{1, 0,}_{2} \underbrace{0, 1,}_{3} \underbrace{1, 1,}_{4} \underbrace{0, 0,}_{5} \underbrace{0, 1,}_{6} \underbrace{0, 0,}_{7} \underbrace{0, 0,}_{8} \underbrace{1, 0,}_{9} \underbrace{1, 1,}_{10} \right\},$$

where the index below the keys indicates time steps of the window's motion. The next step is to count how many times a given configuration occurs, for example, the configuration $\{0,0\}$ occurred 4 times (in the instants of "time" 1, 5, 7 and 8), leading to the probability $4/10$. Similarly, we calculate the probability of other configurations and for other window lengths as well.

Figure (1) shows S_q as a function of L for some values of μ. Note that for each value of μ there is only one value of $q = q*$ that makes the relation S_q versus L linear. This feature becomes evident when we look at the linear correlations (see the insets in Fig. (1)). We can observe from the above results that when μ decreases $q*$ also decreases.

Motivated by the previous results, we investigate the relation $q*$ versus μ for two, three and four-symbol sequences. The results are shown in Fig. (2). Note that, when μ increases, q^* tends to unity, and that the more symbols the sequence has, the faster it reaches towards one. This feature shows that large values of μ generate small values of N_y and consequently the

terms of the symbolic sequence becomes noncorrelated leading to the usual description based on the Boltzmann-Gibbs entropy. However, when μ decreases, N_y is generally very large (remember that, when $\mu < 2$, all the moments of $p(y)$ diverge) and introduces correlation among the terms of the symbolic sequence which are not properly described by the usual formalism. The decreasing values of $q*$ reflects this nonusual behavior. We emphasize that in this case the Tsallis entropy is extensive and Boltzmann-Gibbs entropy is not, indicating the applicability and robustness of the generalized entropy.

3. DIFFUSION AND SEQUENCE

In order to explore further aspects of a symbolic sequence, let us consider it as an erratic trajectory and establish a cor-
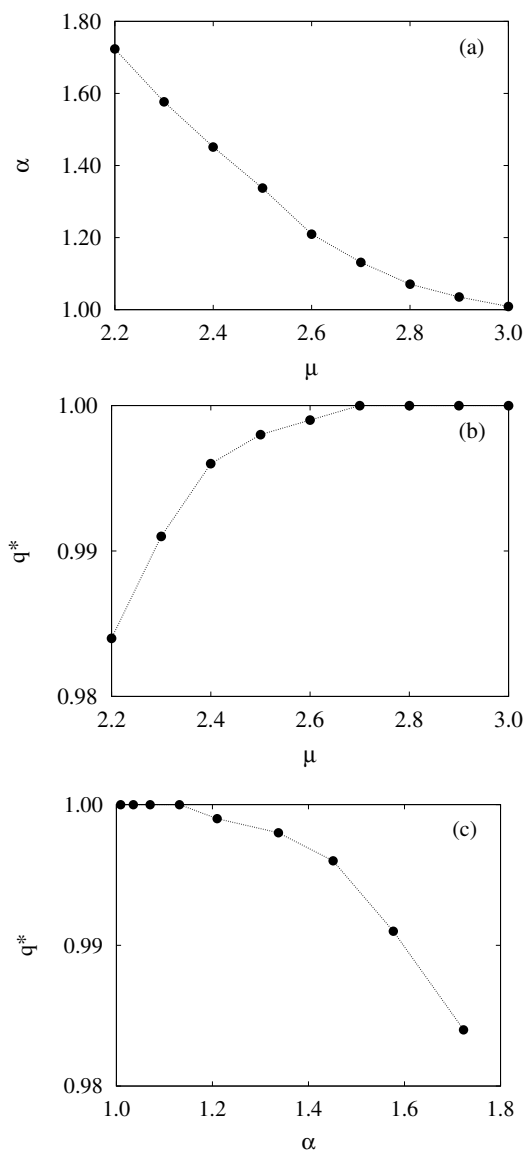
respondence with a diffusive process. For the case of two-symbol sequences, we associate the symbol "0" with a jump of unit length to the right and the symbol "1" with a jump of unit length to the left. That is, a random walk-like process.

Using the previous prescription, we calculate the standard-deviation for $i = 1$ to N over 10^5 events as we can see in Fig. (3). We know that the slope α of this curve is one for a usual diffusion, but in the case of Fig.(3) α is greater than one. We also observed that α depends on μ. This behavior is shown in Fig.(4a). Note that for small values of μ ($\mu < 3$) the diffusion is anomalous, i.e., we have a superdiffusion, and for large values ($\mu \geq 3$) the diffusion regimes tend to a usual diffusion. This behavior can be explained if we remember that for small values of μ, N_y can be very large and consequently the walker can make large steps without changing the direction. When μ is large, this event becomes very rare because N_y is in general small, making the walker change directions, producing a usual erratic trajectory. We may also connect α with $q*$ through the values of μ. In order to do this we evaluate the relation $q*$ versus μ as shown in Fig. (4b) and exhibit $q*$ versus α in Fig. (4c).

FIG. 4: (a) The α slope of the curve in Fig (3) versus μ (b) the entropic index $q*$ versus μ and (c) $q*$ versus α. In the three figures we use $A = 0.1$.

4. DISCUSSION AND CONCLUSION

We verified that by varying the value of μ we can produce long-range correlations in symbolic sequences. This is evidenced by the nonlinear growing of the Boltzmann-Gibbs entropy. This feature led us to use the Tsallis entropy with suitable values of q to obtain a satisfactory description of these sequences. Specifically, we observed that the Tsallis entropy preserves the extensivity even when the terms of the symbolic sequence are correlated. We also considered the symbolic sequence as a random walk-like process and evaluated the standard deviation. The result showed that the diffusive process presents a superdiffusive regime which emerges for small values of μ ($\mu < 3$). The usual diffusion is recovered when $\mu \geq 3$.

[1] P. Gaspard and X. J. Wang, Proc. Nat. Acad. Sci. USA **85**, 4591 (1988).

[2] R. S. Mendes, L. R. Envagelista, S. M. Thomaz, A. A. Agostinho, and L. C. Gomes, Ecography **31**, 450 (2008).

[3] A. Drago, A. Lavagno, and P. Quarati, Physica A **344**, 472 (2004).

[4] S. M. D. Queirós, L. G. Moyano, J. de Souza, and C. Tsallis, Eur. Phys. J. B **55**, 161 (2007).

[5] E. P. Borges, Eur. Phys. J. B **30**, 593 (2002).

[6] M. A. Montemurro, Physica A **300**, 567 (2001).

[7] K. Krishnamachari and A. Krishnamachari, Neural Information Processing **3316**, 1354 (2004).

[8] C. K. Peng, S. V. Buldyrev, A. L Goldberger, S. Havlin, F. Sciortino, M. Simons, and H. E. Stanley, Nature **356**, 168 (1992).

[9] H. Herzel and I. Grosse, Physica A **216**, 518 (1995).

[10] W. T. Li, T. G. Marr, and K. Kaneko, Physica D **75**, 392 (1994).

[11] P. Allegrini, P. Grigolini, and B. J. West, Phys. Lett. A **211**, 217 (1996).

[12] W. Ebeling, R. Feistel, and H. Herzel, Phys. Scripta **35**, 761 (1987).

[13] W. Li and K. Kaneko, Europhys. Lett. **17**, 655 (1992).

[14] H. Herzel, W. Ebeling, and A. O. Schmitt, Phys. Rev. E **50**, 5061 (1994).

[15] A. O. Schmitt and H. Herzel, J. Theor. Biol. **188**, 369 (1997).

[16] P. Liò, A. Politi, M. Buiatti, and S. Ruffo, J. Theor. Biol. **180**, 151 (1996).

[17] C. Tsallis, J. Stat. Phys. **52**, 479 (1988).

[18] A. Pluchino, A. Rapisarda, and C. Tsallis, Europhys. Lett. **80**, 26002 (2007).

[19] E. P. Borges and C. Tsallis, Physica A **305**, 148 (2002).

[20] J. L. Du, Astrophysics and space science **305**, 247 (2006).

[21] P. C. Assis, L. R. da Silva, E. K. Lenzi, L. C. Malacarne, and R. S. Mendes, J. Math. Phys. **46**, 123303 (2005).

[22] M. Buiatti, P. Grigolini, and L. Palatella, Physica A **268**, 214 (1999).