16 December 1996

PHYSICS LETTERS A

ELSEVIER

Physics Letters A 223 (1996) 348-354

Some features of the López-Ruiz–Mancini–Calbet (LMC) statistical measure of complexity

C. Anteneodo [a,b,1], A.R. Plastino [b,2]

[a] *Instituto de Biofísica CCF, Universidade Federal do Rio de Janeiro, CCS, G026, Cidade Universitária, Ilha do Fundão, 21949-900 Rio de Janeiro, Brazil*

[b] *Centro Brasileiro de Pesquisas Físicas/CNPq, Rua Dr. Xavier Sigaud 150, Urca, 22290-180 Rio de Janeiro, Brazil* [3]

Received 8 May 1996; revised manuscript received 1 October 1996; accepted for publication 2 October 1996
Communicated by A.P. Fordy

Abstract

We discuss some aspects of a statistical measure of complexity recently introduced by López-Ruiz, Mancini and Calbet (LMC) [Phys. Lett. A 209 (1995) 321]. By means of a numerical procedure and analytical considerations, we exhibit the qualitative features of discrete probability distributions that maximize the LMC measure. We also study the behavior of the LMC measure under replications of the system. Moreover, we show some peculiarities concerning continuous probability distributions.

PACS: 05.20.−y; 02.50.+s; 02.70.+d

1. Introduction

In the last years many efforts have been devoted towards a rigorous definition of complexity but there is not yet a consensus on a precise definition. Concepts such as *information* and *entropy* are frequently present in the proposals for quantifying the complexity of a system or process [1]. Some recent attempts, coming from the field of physics, involve the notion of *edge of chaos* [2] or that of *self-organized criticality* [3]. Other concepts such as *computational complexity*, *algorithmic complexity* and *logical depth* [2,4] arose in the field of computational sciences. Complexity is encountered in many other contexts (biology, economy, linguistics, sociology, etc.) and the main barrier in arriving to a consensus appears to come from the lack of a common language in the different fields of science.

Intuitively, one of the conditions that a measure of complexity must fulfill is to vanish both for completely ordered and for completely random systems. Recently, López-Ruiz, Mancini and Calbet [5] introduced a measure of complexity fulfilling the above criterium. The LMC measure is a function of the probabilities p_i $(i = 1, \dots, N)$ associated to the N possible accessible states of a given physical system. Within this context, one can identify a wholly ordered system with one having a unique nonvanishing probability state, whereas equiprobability describes fully random systems.

It is the aim of the present work to explore some

[1] E-mail: celia@cat.cbpf.br.

[2] Permanent address: Facultad de Ciencias Astronómicas y Geofísicas, Casilla de Correo 727, Universidad Nacional de La Plata, 1900 La Plata, Argentina.

[3] Mailing address.

PII S0375-9601(96)00756-6

features of the LMC measure. For the discrete case, by Metropolis Monte Carlo simulations we find the probability distributions that maximize the LMC measure. Numerical results are supported by analytical considerations. We also study the behavior of the LMC measure under "replication" of the system of interest. For the continuous case, we construct a family of relatively simple distributions that lead to arbitrarily large values of the LMC measure and we analyze the behavior of the measure under rescaling transformations.

2. Optimal discrete distributions

The LMC statistical measure of complexity [5] is given by

$$C = H \cdot D, \tag{1}$$

where H stands for the logarithmic entropy

$$H = -\sum_{i=1}^{N} p_i \ln p_i, \tag{2}$$

and D is the quadratic distance of the actual probability distribution $\{p_i\}$ to equiprobability,

$$D = \sum_{i=1}^{N} (p_i - 1/N)^2. \tag{3}$$

The main motivation behind the LMC proposal is the construction of a complexity measure that vanishes for both fully ordered systems and completely random ones. In fact, we have $H = 0$ in the former case, while D vanishes in the latter one. Hence, the measure C, as defined by Eq. (1), is expected to achieve maximum values somewhere between order and randomness. This quantity, which looks quite correct as a measure of complexity in those extreme situations, is somewhat arbitrary at intermediate levels of order/disorder. Indeed, all definitions of complexity are, to some extent, arbitrary in the above sense. After all, they are just human theoretical constructs. Only by a thorough analysis of its properties it is possible to assess how useful, in the description of nature, a given measure of complexity is. With regards to the LMC proposal, it is possible to define many alternative measures fulfilling, equally well, the requirements at order and randomness. For instance, the simple product of two functions G_1 and G_2 of the probability distribution $\{p_i\}$ with suitable convexity and such that G_1 is maximal at equiprobability and vanishes at certainty while G_2 is maximal at certainty and vanishes at equiprobability. Clearly, the degree of order/disorder for which a measure is maximal will depend on the particular measure. In the present work, we will circumscribe our scope to the analysis of the LMC measure as a simple paradigmatic member of that family of measures.

By means of Monte Carlo simulations using the Metropolis algorithm, the distributions of probabilities $\{p_i\}$ that maximize the measure C can be obtained. At each Montecarlo step, we randomly choose two values of i, (i_1 and i_2) and a real number Δ uniformly distributed on the interval $[0, p_{i_1}]$. Probabilities are actualized as $p_{i_1}^{\rm new} = p_{i_1}^{\rm old} - \Delta$ and $p_{i_2}^{\rm new} = p_{i_2}^{\rm old} + \Delta$ ($p_i^{\rm new} = p_i^{\rm old}$, for $i \neq i_1, i_2$). The new distribution is accepted or rejected according to the standard Metropolis prescription where the optimized quantity is the measure C and the "temperature" is appropriately chosen.

For different values of N, we obtained that the distributions optimizing C have the same qualitative features: one of the states has probability p, while all the $(N-1)$ remaining states have the same probability $(1-p)/(N-1)$. We verified that different initial distributions (uniform, random, etc.) led to the same final one.

Our numerical results are also supported by analytical considerations. In fact, it can be shown that, at maximal C, the probabilities p_i ($i = 1, \dots, N$) admit at most two values. (In the following analysis, we exclude the trivial cases when $C = 0$, which clearly correspond to minimal C.) By optimizing C with the normalization constraint, i.e.,

$$\delta\left(C - \lambda\left(\sum_{i=1}^{N} p_i - 1\right)\right) = 0, \tag{4}$$

straightforwardly, we obtain

$$\ln p_i - a_1 p_i + a_2 = 0, \quad i = 1, \dots, N, \tag{5}$$

where

$$a_1 = \frac{2H}{D}, \tag{6}$$

$$a_2 = 1 + \frac{\lambda}{D} + \frac{2H}{ND}. \tag{7}$$

Now we want to know the maximum number of different values of p_i that Eq. (5) admits. In order to perform this analysis, a_1 and a_2 must be regarded as constants, since they are the same for all p_i with $i = 1, \ldots, N$. With this in mind, the first derivative of the l.h.s. of Eq. (5) with respect to p_i has only one real positive root. Therefore, by recourse to the mean value theorem, we infer that Eq. (5) can not have more than two real positive roots. The positivity requirement ($p_i \geqslant 0$) determines a subset $\mathcal{A}$ of the $(N-1)$-dimensional hyperplane defined by the normalization constraint. The Lagrange multipliers method supplies the extreme values in the interior of $\mathcal{A}$ but since $\mathcal{A}$ is a bounded domain, extrema could also occur at the boundary. It can be shown (see Appendix) that maximum values only occur strictly in the interior of $\mathcal{A}$. Consequently, we restricted our analysis to the behavior of measure C when the distribution of probabilities has the form

$$p_i = p \quad \text{for } 1 \leqslant i \leqslant n, \qquad = \frac{1-pn}{N-n} \quad \text{for } n < i \leqslant N, \tag{8}$$

with $p \in \mathbb{R}^+$ and $n \in \mathbb{N}$. The optimal values of parameters p and n satisfy

$$(2/n - 3p + 1/N)\ \ln\left(\frac{1-np}{N-n}\right) + (3p - 1/N)\ \ln p = 0, \tag{9}$$

$$(1/n - 2p + np/N)\ \ln\left(\frac{1-np}{N-n}\right) + p\,(2 - n/N)\ \ln p - (p - 1/N) = 0, \tag{10}$$

derived from the conditions $\partial C/\partial p = 0$ and $\partial C/\partial n = 0$, respectively. Numerical resolution of this set of equations for integer values of n leads to maximal C for $n = 1$ (and also $n = N - 1$). In Fig. 1a we show, for given values of n, maximal values of measure C ($C^{\max}$) versus $\ln N$ obtained by numerically solving Eq. (9) for p. The values obtained in the case $n = 1$ are coincident with those resulting from Monte Carlo simulations. Indeed, our analytical arguments and Monte Carlo results are consistent.

The behavior of $C^{\max}/\ln N$ versus $\ln N$, for $n = 1$ is depicted in Fig. 1b. It follows from Eq. (9) that, for $n = 1$, in the limit $N \to \infty$, $p \to 2/3$. Then, in that limit, $C^{\max}/\ln N$ approaches a finite value $(4/27)$, indicating that, for large values of N, $C^{\max}$ grows logarithmically with N.

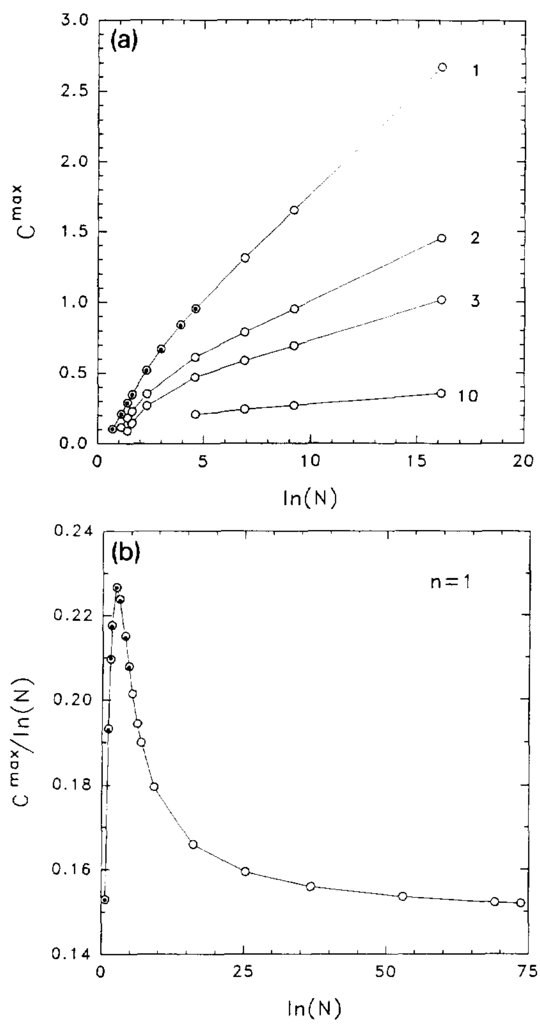

Fig. 1. (a) Maximal values of measure C ($C^{\max}$) versus $\ln N$ for different values of n. Open circles correspond to values obtained from Eq. (9) for fixed values of n indicated in the figure. (b) $C^{\max}/\ln N$ versus $\ln N$ for $n = 1$. As $N \to \infty$, $C^{\max}/\ln N$ approaches 4/27. In both cases, full circles correspond to Monte Carlo results.

We analyzed the logistic map $x_{i+1} = \alpha\, x_i\,(1 - x_i)$, $\alpha \in [0,4]$, following Ref. [5]. The dynamics was reduced to a binary sequence (0 if $x \leqslant \frac{1}{2}$, 1 if $x > \frac{1}{2}$) and binary sequences of length 12 were considered as states of the system. The concomitant probabilities are assigned according to the frequency of occurrence

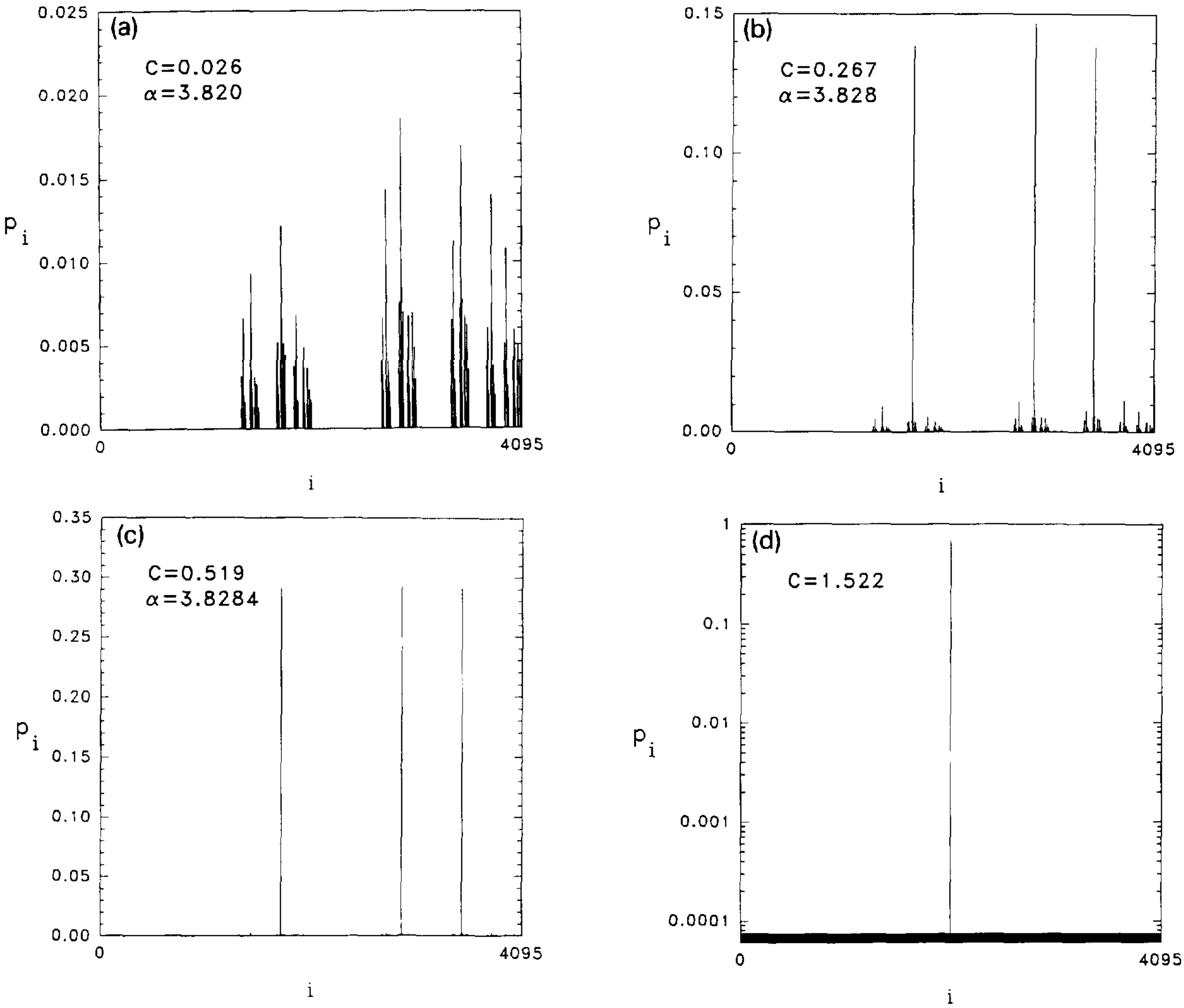

Fig. 2. (a)-(c) Probability distributions associated with the logistic map for different values of parameter α close to the transition point from chaos to a period three orbit. Probabilities were computed running over 2^{26} iterations. (d) Optimal probability distribution (one of the N possibilities). In all cases $N = 2^{12}$. The respective values of C are indicated in the figure.

after running over at least 2^{22} iterations. By means of the above procedure, we obtained the probability distributions for a set of values of parameter α approaching the transition point from chaos to a period three orbit ($\alpha \simeq 3.8284$). These distributions are shown in Fig. 2 together with the one corresponding to the optimal case. As we approach the critical point from the left, the associated distribution becomes less uniform (hence, C increases) until a few peaks predominate (Figs. 2a, 2b). Very close and at the left of the critical point (at which C presents a local maximum), the distribution tends to present three equal peaks (Fig. 2c). For values of α just beyond the critical point, there is only one peak with unit probability (result not shown) and, therefore, C vanishes. There is only one peak because, in our description of the dynamics, the length of the (non-overlapping) state-sequences is an exact multiple of the orbit period. We also considered the first critical point of the logistic map (accumulation point of the subharmonic cascade) and the transition point from regular to chaotic behaviors in the Lorenz map. Around those critical points, we obtained probability distributions with similar features as those described above, although the detailed dynamics involved is very different. The optimal probability distribution (Fig. 2d) is very simple and does not seem to represent a physical situation exhibiting higher complexity than the logistic map close to the transition point.

3. Behavior of the LMC measure under replications

The behavior under replications of the system is relevant in connection with one of the criteria for complexity measures advanced by Lloyd and Pagels [4]. According to these authors, complexity must not be an additive property of the parts that constitute an object and must remain essentially unchanged under a simple "duplication".

Let us consider m independent systems A_j ($j = 1,\dots,m$), characterized by probability distributions $\{p_i^{A_j}\}$ ($i = 1,\dots,N_j$).

If we regard the set of those m systems as a unique system Σ with probability distribution $\{p^{\Sigma}_{i_1 i_2 \dots i_m}\}$ ($i_j = 1,\dots,N_j; j = 1,\dots,m$), given by

$$p^{\Sigma}_{i_1 i_2 \dots i_m} \equiv \prod_{j=1}^{m} p_{i_j}^{A_j} \qquad (i_j = 1,\dots,N_j; j = 1,\dots,m), \tag{11}$$

it is easy to show, after some algebra, that the LMC complexity measure associated to the combined system is

$$C^{\Sigma} = \left\{\sum_{j=1}^{m} H^{A_j}\right\}\left\{\prod_{j=1}^{m}\left(D^{A_j} + \frac{1}{N_j}\right) - \prod_{j=1}^{m}\frac{1}{N_j}\right\}, \tag{12}$$

in self-explanatory notation. In the particular situation where all the m systems are identical, the combined system Σ consists just of m copies of an original single one. In this case, the complexity C_m of the m replicas is related to the complexity C_1 of the original system by

$$C_m = \frac{m}{D}\left[\left(D + \frac{1}{N}\right)^m - \frac{1}{N^m}\right]C_1, \tag{13}$$

where D and N refer to the original single system. From the above relation we can see that, depending on the value of D, the LMC measure of m identical copies of a system may be greater, equal or even smaller than the LMC measure of the original system.

4. Some properties of the LMC measure for continuous distributions

For a continuous probability distribution $f(x)$, the LMC measure is defined by [5]

$$C = -\left(\int_{-\infty}^{\infty} f(x)\ln f(x)\mathrm{d}x\right)\cdot\left(\int_{-\infty}^{\infty}[f(x)]^2\mathrm{d}x\right). \tag{14}$$

Let us consider the following probability function,

$$F(x) = \frac{1}{n\epsilon} \quad \text{for } 0 \leqslant x \leqslant \epsilon, \qquad = \frac{(n-1)\epsilon}{n} \quad \text{for } \epsilon < x \leqslant \epsilon + \frac{1}{\epsilon}, \tag{15}$$

where n and ϵ are positive real numbers such that $n \geqslant 1$. The associated complexity reads

$$C = \frac{1}{n^3}\left(\frac{1}{\epsilon} + (n-1)^2\epsilon\right)[n\ln n - (n-1)\ln(n-1) + (2-n)\ln\epsilon]. \tag{16}$$

It is plain from the above expression that C diverges when ϵ vanishes: for $1 \leqslant n < 2$, we get $C \to -\infty$, while, if $n \geqslant 2$, we have $C \to +\infty$. This means that, given any probability distribution, no matter how intricate, we can find (for appropriate values of n and ϵ) a trivial function F with a higher "complexity" measure C. This behavior is coherent with that found in the discrete case.

Let us consider the behavior of measure C under the rescaling of an arbitrary distribution function $f(x)$. The measure C_γ of the rescaled function $f_\gamma(x) = \gamma f(x\gamma)$ reads

$$C_\gamma = \gamma(C_1 - D_1 \ln\gamma). \tag{17}$$

Thus, the LMC measure is not invariant under a rescaling transformation. As $\gamma \to 0$, i.e., as the distribution broadens, $C_\gamma \to 0$. As $\gamma \to \infty$, i.e., as the distribution becomes more peaked, $C_\gamma \to -\infty$. We can identify the first case with randomness and the second one with order. While in the former instance C vanishes, in the latter, C diverges to $-\infty$ instead of vanishing. Specially, it is worth noting that the measure C varies drastically under a simple scaling transformation which does not change the shape of the distribution.

5. Discussion

The goal of the present work is to study some aspects of the LMC measure in order to contribute to the elucidation of the criteria for defining a useful measure of complexity.

We studied numerically the discrete distributions yielding maximal complexity according to the LMC measure. Our results are sustained by analytical arguments. In the continuous case we studied a family of distributions, depending on two parameters n and ϵ, that yield arbitrarily large values of LMC measure C. Results found in the discrete and the continuous cases are consistent. In both instances, extreme values of C are observed for distributions characterized by a sharp peak superimposed to a uniform sea. Clearly, this kind of distribution constitutes an intermediate case between full order (certainty) and complete randomness (equiprobability).

We considered the behavior of the LMC measure under replications of the system. An expression was derived showing that, depending on the value of the distance D from equiprobability, the complexity of m identical replicas of a given system may be larger, equal or lower than the complexity of the original system.

Another remarkable feature of the LMC measure is the fact that it depends only on the values of probabilities p_i and not on the particular states i they are assigned to. This property is valid both for discrete and continuous distributions. For instance, in the continuous case, let us consider a step distribution with just two probability values p_1 and p_2 (like the one given by Eq. (15)). It is clear that C will not change if we arbitrarily redistribute the regions where each value of p is adopted (without changing their total lengths). A possible way to take into account this aspect of a distribution is by recourse to a complexity measure that involves, simultaneously, the properties of the system at different scales of coarse-graining as recently proposed by Zhang [6].

For the continuous case, we also studied the behavior of the LMC measure under a rescaling transformation. We showed that the value of C changes drastically under this simple kind of transformation.

The complexity measure advanced by LMC has the merits of being simple and verifying the criterium of vanishing both at total order and complete disorder. It is a functional of the probability distribution characterizing the system of interest at a given (and unique) resolution scale. Further theoretical research may help to find out to what extent the LMC measure constitutes a useful quantitative measure embodying our intuitive notions about complexity. We think that the following points are particularly problematic. First of all, it is not clear for us in what sense the probability distributions that optimize C can be regarded as complex. The behavior of the LMC measure under a rescaling transformation also seems to be in conflict with what common knowledge tells us is complex, since we feel that complexity must not change dramatically under simple transformations. Of course, these objections arise exclusively from our intuition about complexity. Another counterintuitive result is that, following the LMC description, the complexity of m replicas of a given system could be smaller than the complexity of the original.

Perhaps, a configurational approach as the LMC measure is not able to deal with the above discussed difficulties. According to Lloyd and Pagels [4] complexity must be a function of the process that brought the object into existence and not a property of the object actual physical configuration. Exploring the connection between the "historical" and the configurational approaches could throw new light on the nature of complexity. Aside from its possible lasting value as a complexity measure, the LMC proposal proved to be worth studying as it raised many interesting issues about the meaning of complexity.

Acknowledgement

We acknowledge Brazilian agency CNPq for financial support.

Appendix A

We will prove that, given any probability distribution $\{p_i\}$ $(i = 1, \ldots, N)$ that lies on the boundary of domain $\mathcal{A}$, we can always find another distribution $\{q_i\}$ with measure $C(\{q_i\}) > C(\{p_i\})$. This means that the measure C does not have maxima at the boundaries of $\mathcal{A}$.

Since $\{p_i\}$ lies on the boundary, at least one of the p_i's vanishes. Thus, without loss of generality we assume $p_N = 0$. Now, let $\{q_i\}$ be defined as

$$\begin{aligned} q_i &= \epsilon \cdot p_i \quad \text{for } 1 \leqslant i < N, \\ &= 1 - \epsilon \quad \text{for } i = N, \end{aligned}$$

where $0 < \epsilon < (1 - \sum p_i^2)/(1 + \sum p_i^2)$. Straightforwardly, it follows that $C(\{q_i\}) > C(\{p_i\})$ holds. We have excluded the case of certainty, i.e., only one of the p_i's is non-zero (hence, $\sum p_i^2 = 1$). In this case H and, consequently, C vanish, therefore, it is obvious that measure C will be greater for any other distribution $\{q_i\}$ with at least two non-vanishing probabilities.

References

[1] G. Nicolis and I. Prigogine, Exploring complexity: an introduction (Freeman, New York, 1989).
[2] W.H. Zurek, ed., Complexity, entropy and the physics of information (Adisson-Wesley, Redwood City, 1990).
[3] P. Bak and M. Paczuski, Physics World (December 1993) 39.
[4] S. Lloyd and H. Pagels, Ann. Phys. 188 (1988) 186.
[5] R. López-Ruiz, H.L. Mancini and X. Calbet, Phys. Lett. A 209 (1995) 321.
[6] Y.-C. Zhang, J. Phys. I (Paris) 1 (1991) 971.