Journal of Statistical Physics, Vol. 54, Nos. 3/4, 1989

Chaotic Dynamics, Markov Partitions, and Zipf's Law

G. Nicolis,[1] **C. Nicolis,**[2] **and John S. Nicolis**[3]

Received June 7, 1988; final August 19, 1988

A chaotic dynamics model creating Markovian strings of symbols as well as sequences of "words" is presented, and its possible relevance to Zipf's law is discussed.

KEY WORDS: Chaotic dynamics; Markov chains; information processing; linguistics; Zipf's law.

1. INTRODUCTION

Large classes of deterministic dynamical systems possessing few degrees of freedom and giving rise to chaotic attractors can generate complexity in the form of an interplay between randomness and order.[1]

Inasmuch as randomness ensures variety and information generation, while order ensures reliability, it is legitimate to expect that chaotic dynamics should be relevant in biological information processing.[2] In this communication we show how, starting from a deterministic dynamical system operating in the chaotic region, one may (a) generate strings of symbols obeying a well-defined Markov statistics, and (b) combine these symbols in words interrupted by the pause (blank space), whose abundance is described by an inverse power law similar to Zipf's law of experimental linguistics.[3]

Inverse power law distributions have been investigated in recent years in connection with self-similar (fractal) processes in physics, biology, cognitive psychology, and social sciences.[4–6] The best-known examples

[1] Faculté des Sciences de l'Université Libre de Bruxelles, 1050 Bruxelles, Belgium.
[2] Institut d'Aéronomie Spatiale de Belgique, 1180 Bruxelles, Belgium.
[3] Department of Electrical Engineering, University of Patras, Patras, Greece.

0022-4715/89/0200-0915$06.00/0

refer to the spectral domain ($1/f$ noise) or to highly multivariable processes. In this respect, therefore, Zipf's law

$$P(r) \sim \frac{A}{r^{\lambda}} \tag{1}$$

relating the probability $P(r)$ of appearance of words in a natural language with their rank r appears to be rather exceptional. Here we show that Eq. (1) can be recovered in some asymptotic sense (to be defined shortly) from quite simple dynamical systems involving few degrees of freedom, provided an appropriate mapping is constructed between the (chaotic) dynamics and Markovian strings of symbols, and an adequate definition of rank is adopted.

2. MAPPING CHAOTIC DYNAMICS INTO A MARKOV PROCESS

Consider a dissipative recurrent dynamical system

$$\mathbf{x}_{n+1} = \mathbf{F}(\mathbf{x}_n, \mu) \tag{2}$$

possessing a smooth invariant probability density $\rho(\mathbf{x})$. We partition the state space into N nonoverlapping cells C_i such that the boundaries between cells are preserved by the dynamics. Clearly, the shift process induced on such a *Markov partition* by Eq. (2) generates sequences of "symbols" whose number is equal to N. We have shown that, under certain conditions, the probability distribution of these strings obeys a Chapman–Kolmogorov equation giving rise to an irreversible approach to a stationary state and to an H-theorem[7]:

$$P_{n+1}(i) = \sum_{j=1}^{N} W_{ji} P_n(j), \qquad i = 1,..., N \tag{3}$$

A simple example of transition between (2) and (3), which will be used in the sequel, is provided by the logistic map in the fully chaotic region,

$$x_{n+1} = 4x_n(1 - x_n), \qquad 0 \leqslant x \leqslant 1 \tag{4}$$

For this dynamical system there exists a family of Markov partitions whose cells are separated by the points on the unstable periodic orbits. For instance, the points of the period-two orbit $\bar{x}_1 \simeq 0.345$, $\bar{x}_2 \simeq 0.905$ define a three-cell partition (Fig. 1). The resulting three "states" α, β, γ, which can also be viewed as "letters" of an alphabet, are then continuously transfor-

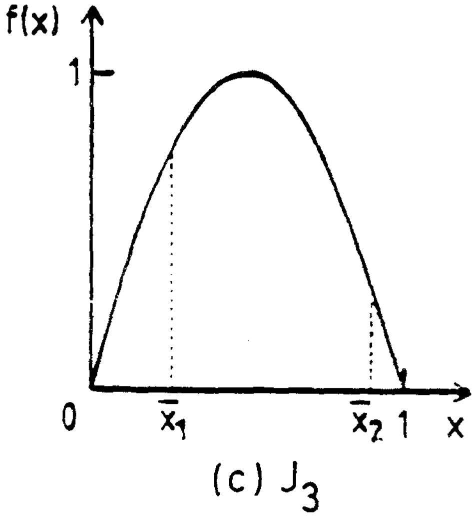

Fig. 1. A three-cell Markov partition of the logistic map.

med into each other by the dynamics according to a first-order Markov chain whose conditional probability matrix turns out to be[7]

$$W = \begin{pmatrix} 1/2 & 1/2 & 0 \\ 0 & 1/2 & 1/2 \\ 1 & 0 & 0 \end{pmatrix} \tag{5}$$

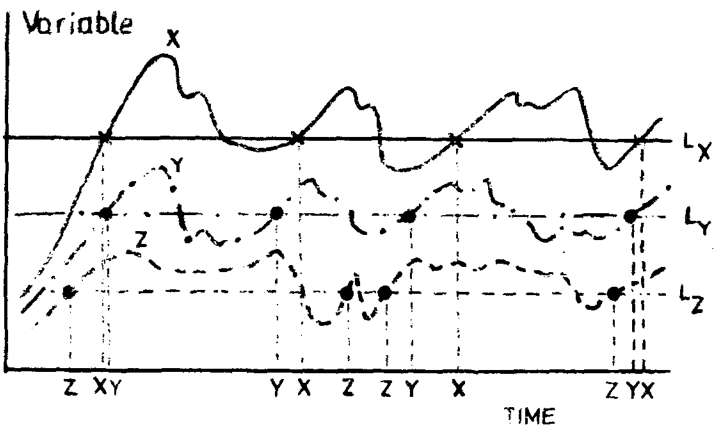

Fig. 2. Asymmetric sequence involving the three symbols X, Y, Z generated as the three variables x, y, z of the model of Eq. (6) cross the threshold values L_x, L_y, L_z with a positive slope.

One can show that the lumping of certain cells of the above-defined partitions generates Markov chains of higher order characterized by the existence of correlations between successive symbols.[7]

A second mechanism of producing information-rich sequences of symbols and hypersymbols (groups of symbols) has been also suggested recently.[8] One considers a dissipative flow whose state variables x, y, z perform sustained (aperiodic) oscillations and assumes that when a variable crosses a certain predetermined level with, say, a positive slope a symbol "forms" and is subsequently "typed" (Fig. 2). One can envisage in this way a sequence of level crossing variables—symbols—standing as a one-dimensional trace of the underlying multidimensional flow. The sequence is by necessity asymmetric (a syndrome that all languages share) as a result of the dissipative character of the flow in phase space. In ref. 8 numerical examples are given for the Rössler attractor,

$$\begin{aligned} \dot{x} &= -y - z \\ \dot{y} &= x + ay \\ \dot{z} &= bx - cz + xz \end{aligned} \tag{6}$$

with $a = 0.38$, $b = 0.3$, $c = 4.5$, and thresholds $L_x = L_y = L_z = 3$. A typical sequence generated by this mechanism is

$$zyx\ zxyx\ zxyx\ zyx\ zxyx\ zyx\ zyx\ zx\ zyx\ zyx\ zxyx\ zyx... \tag{7a}$$

Remarkably, one can verify that the above sequence can be formulated more succintly by introducing the hypersymbols

$$\alpha = zyx, \qquad \beta = zxyx, \qquad \gamma = zx \tag{7b}$$

giving rise to

$$\alpha\beta\beta\alpha\beta\alpha\alpha\gamma\alpha\alpha\beta\alpha... \tag{7c}$$

A statistical analysis reveals strong correlations in the sequence (7a) which to a very good approximation can be fitted by a fifth-order Markov process. On the other hand, the hypersymbol sequence is definitely more random first order chain, indicating that the "compression" achieved by the hypersymbols has indeed removed much of the structure of the original sequence.[8]

3. WORD GENERATION AND ZIPF'S LAW

Having now at our disposal the above two algorithms for generating information-rich strings of symbols from a deterministic dynamics, we come to the main object of our study. Specifically, we show that starting from the "alphabet" induced in the shift space, one can generate sequences of "words" having some well-defined statistical properties.

The starting point is to choose one of the symbols of our alphabet to be the pause (blank space). As the dynamics unfolds in the shift space, the remaining $N-1$ symbols are then organized in words C_L, of varying lengths L, interrupted by the pause. We want to find the probability $P(C_L)$ of formation of such words. Notice that the sequence C_L is in general a *non-Markovian* process.[9]

We first carry out the analysis on the simple example of the three-cell partition of the logistic map. Choosing β to be the pause limits the "language" to words involving a single nontrivial letter α, since the role of γ is trivial ($W_{\gamma\alpha}=1$). We obtain, using the explicit form of the transition probability matrix (5),

$$P(C_L)=W_{\alpha\alpha}^{L-1}=(1/2)^{L-1}, \qquad L\geqslant 2 \tag{8a}$$

If, on the other hand, γ is used as pause, a richer language involving two nontrivial symbols α and β is created. Arguing as above, we obtain

$$\begin{aligned} P(C_L) &= \sum_{m=1}^{L-1} W_{\alpha\alpha}^{m} W_{\alpha\beta} W_{\beta\beta}^{L-1-m} \\ &= (1/2)^{L}(L-1), \qquad L\geqslant 2 \end{aligned} \tag{8b}$$

Although (8a) and (8b) differ significantly for small integer values of C_L, they tend to the same asymptotic form for long words, $L\to\infty$. Stated differently, in this limit the word processor is universally penalized in an exponential fashion with the length of the word generated. The argument can clearly be extended to partitions involving more cells, the difference being merely the occurrence of higher powers of L multiplying an exponential of the form a^L, a being a suitable combination of elements of the conditional probability matrix.

Figure 3 depicts the dependence of the logarithm of $P(C_L)$ versus L obtained by iterating the dynamical system of Eq. (4) and subsequently monitoring the frequency of appearance of strings of symbols of given length between two pauses. The agreement with Eqs. (8a)–(8b) (solid lines in Fig. 3) is very satisfactory.

Following Mandelbrot's discussion[6] of the concept of "lexicographical tree," we now define the rank r_L of a word of length L as the sum of all words of length equal to or less than L,

$$r_L=1+\sum_{i=1}^{L} K^i \tag{9}$$

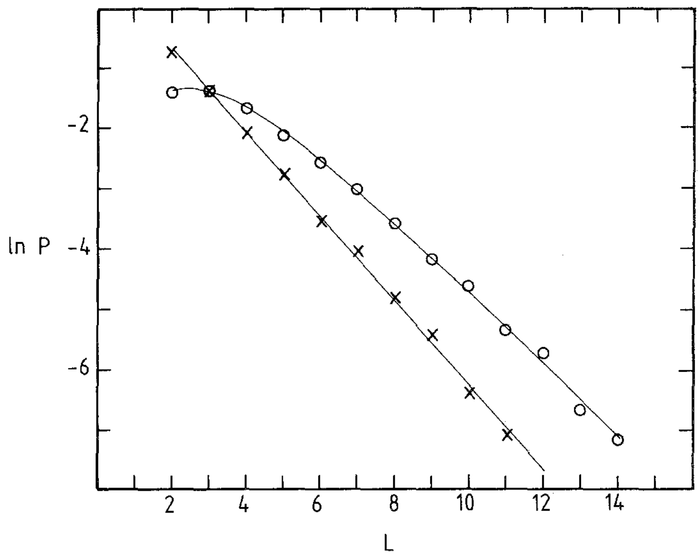

Fig. 3. Probability of words of length L plotted against L. The words are generated by the shift process induced by the dynamics [Eq. (4)] on the three-cell Markov partition of Fig. 1. The crosses and circles stand for the numerically computed probability using, respectively, letter β and letter γ as the pause. Solid lines represent the analytically deduced laws (8a) and (8b).

K being the number of symbols in the alphabet other than the pause. For long words, such that $r_L \gg 1/(K-1)$, one can easily deduce from (9) that

$$r_L \sim K^{L-L_0} \tag{10a}$$

where L_0 is defined through

$$K/(K-1) = K^{-L_0} \tag{10b}$$

Let us apply this to our previous examples. We have $K=2$, i.e., $L_0 = -1$ and $r_L \sim 2^{L+2}$. Consequently, from (8a)–(8b) we obtain for large values of L,

$$P(r_L) \sim 1/r_L \tag{11}$$

which is precisely Zipf's law [Eq. (1)] with $\lambda = 1$.

Essential in the above reasoning was the fact that our dynamical system releases the successive symbols at regular time intervals. In a dissipative flow such as Rössler's system [Eq. (6)] this will not be the case. Figure 4 depicts the numerically computed ln $P(C_L)$ versus L for the hyper-

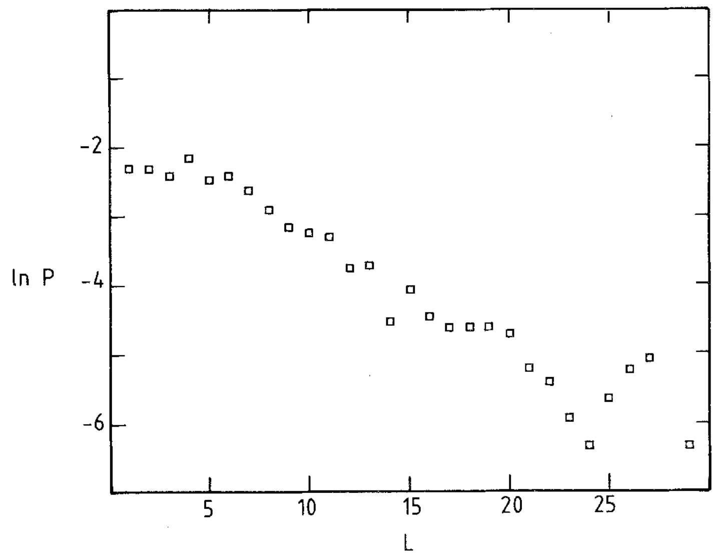

Fig. 4. Probability of words of length L plotted against L. The words are generated by Rössler's model [Eq. (6)] in the hypersymbol space, γ being used as the pause.

symbol sequence of this model using γ as the pause. The dependence is now more complicated than in Fig. 3, despite certain similarities in the general trend.

One can verify that the mean time of formation of a word increases with its length. On the other hand, there exists a large dispersion around the mean, leading to crossovers in the times of formation observed in a given realization of particular words of different lengths. We conjecture that this phenomenon might be at the origin of the observed deviation from Zipf's law.

An interesting question is whether one can identify a function measuring the "quality" of the three processors corresponding to Eq. (8a), Eq. (8b), and the Rössler model. In the theory of dynamical systems it is customary to characterize the qualitative properties of the trajectories by the topological entropy.[10] However, being an invariant, this quantity cannot differentiate between strings of symbols involving alphabets with different numbers of letters. For instance, the processors corresponding to Eq. (8a) and (8b) both have the same topological entropy $h = \ln 2$, which is nothing but the topological entropy of the logistic map. A similar remark holds for the Rössler system. It has been shown[11] that in the region in which the chaotic attractor considered in Section 2 exists, the successive iterates of variable x arising from the intersection of the flow by the plane $(y=0,\ x<0,\ z<1)$ give rise to a cubiclike map involving three monotonous segments. This gives a topological entropy $h = \ln 3$ which is again an invariant.

It is well known that topological entropy is an upper bound of the measure-theoretic entropy, the latter being the supremum of the rate of change of the entropy of an initial partition of state space in the limit of inifite refinement.[10] In the present paper we are not concerned with such a limit, since we are able to obtain a well-defined stochastic process involving a finite number of states, through coarse graining. We therefore introduce the information entropy of the stochastic process $\{C_L\}$,

$$S_I = -\sum_L P(C_L) \ln P(C_L) \tag{12}$$

Contrary to h, this quantity is not an invariant, but depends on the partition, that is, on the algorithm generating the stochastic process $\{C_L\}$ from the original dynamical system. Notice, however, that S_I should not be identified with the entropy of the partition: the latter involves the probability $P(j)$ of being in cell j [Eq. (3)] rather than $P(C_L)$.

Computing S_I using the analytical expressions (8a)–(8b) and the numerical values corresponding to Fig. 4 yields $S_I = 2 \ln 2 \simeq 1.39$ for Eq. (8a), $S_I \simeq 1.88$ for Eq. (8b), and $S_I \simeq 2.87$ for Rössler's model. This trend, which must be contrasted with the invariance of h, reflects, in a sense, the increasing richness of the "repertoires" of the corresponding languages.

4. VARIATIONAL FORMULATION OF ZIPF'S LAW

In this section we show that one can deduce Zipf's law by employing a maximum entropy formalism[4] as follows: we search for the probability density function $P(x)$ characterizing a word of rank order x, which maximizes the *a priori* uncertainty

$$S(x) = -\int P(x) \ln P(x)\, dx, \qquad \int P(x)\, dx = 1 \tag{13}$$

or the maximum, per word average, information conveyed by a language using such a "syntax"—under a given constraint. Such a constraint is associated with a certain average "cost" and is of the general form

$$\int \xi(x)\, P(x)\, dx = \text{const} \tag{14a}$$

If we choose

$$\xi(x) \sim \ln x \tag{14b}$$

(something which has been dismissed as "unphysical" by Montroll and Shlesinger,[4] thereby prompting them to derive inverse-power law distributions as limiting cases of log normal ones), we end up with the relationship between length L and rank r_L derived earlier [Eq. (10a)]. In other words, we express the natural idea that the "cost" of a word is on average proportional to its length.

The maximum entropy subject to (14a) may now be reduced to the search of extrema of the functional

$$H(P) = -\int P(x) \ln P(x)\, dx - \lambda_1 \int P(x)\, dx - \lambda_2 \int \ln x P(x)\, dx$$

Setting $\partial H/\partial P = 0$, one finds

$$P(x) \sim \exp\{(-\lambda_1 - \lambda_2 \ln x)\} = A/x^{\lambda_2} \tag{15}$$

where A is the normalization constant over the interval of x. It is obvious that $P(x)$ obeys the scaling relation

$$P(ax)\, d(ax) = a^{1-\lambda_2} P(x)\, dx \tag{16}$$

implying a lack of fundamental scale in the process underlying $P(x)$. Alternatively, if the rank order x with a distribution $P(x)$ is known in a given interval, that interval can be extended: the scaling implies that the fluctuations of the random variable x are generated at each scale in a statistically identical (self-similar) fashion.

The above arguments provide an additional qualitative explanation of the deviations from Zipf's law found in the preceding section for continuous time flows. Indeed, as pointed out earlier, there exists in this case a large dispersion around the average time needed to form a word of a given length. As a result, a "cost" function proportional to the length of a word does not describe adequately the conditions that must be met for its formation.

5. DISCUSSION

An experimental apparatus—or a cognitive processor for that matter—recognizes the external world in a "coarse-grained" fashion. In this paper we have considered two particular forms of a coarse-graining operation: partitioning of state space into cells satisfying certain conditions; and monitoring the crossings of appropriately defined threshold values by the state variables. We have shown that deterministic chaos induces in the discrete state space constructed in this manner stochastic processes of varying complexity. In particular, we have identified some mechanisms leading to a statistical distribution of "words" of a given rank, described by Zipf's law. We have also indicated some reasons limiting the full applicability of this law in continuous time dynamical systems.

We believe that our results provide further support for the interest in chaotic dynamics in information and cognitive sciences. In future investigations it is planned to extend the analysis to alternative mechanisms of producing information-rich strings of symbols, such as spatially distributed networks or transitions between coexisting attractors.

ACKNOWLEDGMENT

The work of C.N. is supported in part by the European Economic Community under contract ST2J-0079-1-B(EDB).

REFERENCES

1. H. G. Schuster, *Deterministic Chaos* (Physik Verlag, Weinheim, 1984).
2. J. S. Nicolis, *Dynamics of Hierarchical systems* (Springer, Berlin, 1986).
3. G. K. Zipf, *Human Behavior and the Principle of Least Effort* (Addison-Wesley, Cambridge, Massachusetts, 1949).
4. E. Montroll and M. F. Shlesinger, *J. Stat. Phys.* **32**:209 (1983).
5. B. West, in *Lecture Notes in Biomathematics*, No. 62 (Springer, Berlin, 1985).
6. B. Mandelbrot, *Fractals: Form, Chance, Dimension* (Freeman, San Francisco, 1977).
7. G. Nicolis and C. Nicolis, *Phys. Rev.* **38A**:427 (1988).
8. G. Nicolis, S. Rao, G. Rao, and C. Nicolis, in *Coherence and Chaos in Dynamical Systems* (Manchester University Press, 1988).
9. W. Feller, *An Introduction to Probability Theory and Its Applications*, 3rd ed. (Wiley, New York, 1968), Vol. I, p. 426.
10. P. Collet and J. P. Eckmann, *Iterative maps of the interval as dynamical systems* (Birkhäuser, Basel, 1980).
11. P. Gaspard and G. Nicolis, *J. Stat. Phys.* **31**:499 (1983).
12. J. S. Nicolis and I. Tsuda, *Progr. Theor. Phys.*, submitted (1988).