Journal of Statistical Physics, Vol. 70, Nos. 3/4, 1993

The Entropy Profile—A Function Describing Statistical Dependences

Christoph Bandt[1] **and Bernd Pompe**[2]

Received March 31, 1992; final July 8, 1992

In an attempt to find parameters of a time series which are absolutely robust with respect to nonlinear distortion, we introduce a function called the entropy profile which measures in some sense the distance between the given process and white noise. This concept combines a clear definition and a simple algorithm, which apply to arbitrary stationary time series, with an informative graphical representation similar to the Fourier spectrum. For sequences derived from one-dimensional maps, the entropy profile indicates periodic and almost periodic behavior and the presence of Markov partitions.

KEY WORDS: Symbolic dynamics; entropy; time series analysis; unimodal map.

1. INTRODUCTION

We define new entropy functions describing statistical dependences in a time series. Any number s defines a partition of the line into two sets: the numbers smaller than s and those larger than s. Thus the given data are transformed into a 0–1 sequence. The entropy profile of order n indicates the dependence of Shannon entropy, of words of length n of the symbol sequence, on the number s. The dependence of such entropies on the partition has been studied by several authors.[8, 15] Our new idea is a useful (quite simple and natural) standardization of thresholds s by so-called q-quantiles. The method has been applied successfully to a class of speech signals (ref. 2, cf. Fig. 1), and seems applicable to many experimental time series, in particular to those which combine nonlinear structure with random influences.

[1] Fachbereich Mathematik, Ernst-Moritz-Arndt-Universität, O-2200 Greifswald, Germany.
[2] Fachbereich Physik, Ernst-Moritz-Arndt-Universität, O-2200 Greifswald, Germany.

0022-4715/93/0200-0967$07.00/0

In this paper, we try to give a theoretical foundation for our method. We derive some general properties, and we take the well-known scenario of unimodal maps[4, 7, 19, 14] as a test example. Several entropy profiles of quadratic maps are obtained analytically, and new questions concerning one-dimensional dynamics are raised.

As a motivation, compare two main groups of methods in time series analysis. Classical methods, such as correlation functions, spectra, and ARMA models, work well for noisy data, but may turn out to be inadequate in the presence of nonlinear effects. On the other hand, there are parameters from nonlinear dynamical systems, such as fractal dimensions, Lyapunov exponents, metric entropies, and multifractal spectra.[1, 5–8, 10, 11, 15] They are more suitable to detect nonlinear structures and to characterize chaos. However, these parameters are limits which exist, by deep theorems, in the idealized context of deterministic dynamics only. In practice, where nonlinear structure is usually mixed with, or rather covered by, noise, such methods involve tremendous conceptual and numerical problems.

We have tried to find a method which works equally well for random and chaotic data. Entropy profiles combine a simple definition with a straightforward method of calculation. Being functions, they provide more information, and more intuitive information than single numbers. (While this has been stressed as a reason for the use of multifractal spectra instead of single dimensions, the multitude of shapes of entropy profiles is still more impressive.)

Our profiles are only loosely connected with the metric theory of maps, and we do not think they have similar theoretical importance. They rather have practical advantages: they apply to all kinds of time series.[2] The only assumptions we shall make is that the given process is stationary and that its one-dimensional distribution is continuous. The second assumption can be relaxed, as shown at the end of the next section, but the method is not applicable if only a few different values are attained.

2. DEFINITION OF ENTROPY PROFILE

Symbolic Dynamics. Shannon Entropy. Let $\{x_k\}$, $k = 1,..., N$, be the given time series. Using a partition of the range of x values, $\{x_k\}$ is transformed into a discrete sequence of symbols. To retain much information, it is common to use many symbols and partition sets. We take the simplest partitions with only two pieces. Given a threshold s, the time series $\{x_k\}$ is turned into a 0–1 sequence $\{i_k\}$, with $i_k = 0$ for $x_k < s$ and $i_k = 1$ otherwise. For unimodal maps (see Section 4), $\{i_k\}$ is called an itinerary when the critical point is chosen as s. The method of itineraries and kneading sequences is well established and was revived in recent work

of theoretical relevance.[9, 14] The idea here is to vary the threshold s in an appropriate way and to consider different symbol sequences at the same time.

Every symbol sequence is evaluated statistically by entropy. For each $n = 1, 2,...$ we determine the frequencies $p(w)$ of words $w = w_1 \cdots w_n$ in the 0–1 sequence, and the Shannon entropy as sum over all words of length n,

$$H_n = -\sum p(w_1 \cdots w_n) \operatorname{ld} p(w_1 \cdots w_n) \tag{1}$$

Quantiles. Our key point is to standardize thresholds appropriately so that quite different time series can be compared. For $0 < q < 1$, the q-quantile $s = s(q)$ of the one-dimensional distribution of $\{x_k\}$ is characterized by the property that at most q of the values x_k are smaller than s, and at most $1 - q$ larger.[18] In other words, the symbol sequence for $s(q)$ fulfils $p(0) = q$, $p(1) = 1 - q$. In our numerical studies, where time series had several thousand values, we estimated s from $q = \#\{x_k < s\}/N$. (When certain intervals $[s_1, s_2]$ do not contain any values x_k, the definition of quantile becomes ambiguous for some q. This does not matter, since in such cases all thresholds between s_1 and s_2 generate the same symbol sequence.)

Definition. We determine entropies with $s = s(q)$ for each q and obtain H_n as a function of q. The *entropy profile* of order $n \geqslant 2$ of the time series is defined as

$$r_n(q) = 1 - \frac{H_n(q) - H_{n-1}(q)}{H_1(q)} \tag{2}$$

We shall also consider the average profile of order $\leqslant n$,

$$R_n(q) = \frac{nH_1(q) - H_n(q)}{(n-1)\,H_1(q)} = 1 - \frac{H_n(q) - H_1(q)}{(n-1)\,H_1(q)} = \frac{1}{n-1} \sum_{k=2}^{n} r_k(q) \tag{3}$$

Remarks. Both r_n and R_n measure the distance of the time series from a sequence of statistically independent symbols. r_n uses the conditional entropy $H_n - H_{n-1}$ of a letter provided its $n-1$ predecessors are known, and R_n is based on the conditional entropy $H_n - H_1$ of a word $w_1 \cdots w_n$ when its first letter is known.

Entropy profiles for $n = 2,...,10$ of a "chaotic" experimental time series are shown in Fig. 1. The vowel "a" spoken by a male speaker in a modal register was sampled at a frequency of 8 kHz. The picture is typical for a large class of speech signals. The r_n and R_n do not differ much, except perhaps for $n \leqslant 5$, and they have some characteristic peaks in common. In this situation, we recommend R_5 as a rather stable overall parameter.

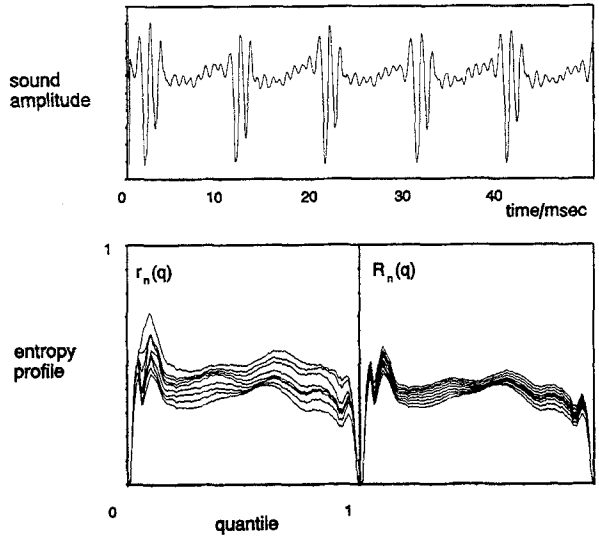

Fig. 1. An experimental time series with its entropy profiles.

It is clear that R_n as an average quantity behaves more regularly than r_n. On the other hand, r_n is quite sensitive with respect to periodic patterns with periods smaller than n. In the framework of one-dimensional dynamics, r_n is the appropriate function showing some remarkable features (Section 4). In view of the special nature of the partitions, it is not clear whether entropy profiles will be equally successful in the study of higher-dimensional attractors.

The Case of Discrete Distribution. Let us assume now that the one-dimensional distribution of the x_k is not continuous, so that some value x is attained with positive probability $p = P\{x_k = x\}$. Let $p' = P\{x_k < x\}$. Then for $p' < q < p' + p$, the quantile $s(q)$ equals x. Intuitively, $s(q)$ cuts the mass p at point x, leaving $q - p'$ on the left and $p' + p - q$ on the right.

This idea applies to the sampling of word frequencies $p(w)$. If $s(q) = x$, define the symbol sequence $i_k = 0$ for $x_k < x$, $i_k = 1$ for $x_k > x$, and $i_k = *$ for $x_k = x$. Determine $p(w)$ for words from 0, 1, and $*$. Afterwards eliminate the words with $*$, one by one, as follows. If w contains one $*$, let w^0, w^1 be the words obtained when $*$ is replaced by 0 and 1, respectively. Add $p(w)(q - p')/p$ to $p(w^0)$ and $p(w)(p' + p - q)/p$ to $p(w^1)$, and cancel w. If w contains two symbols $*$, let w^{00}, w^{01}, w^{10}, and w^{11} denote the words obtained by replacing these symbols by 0 or 1. Add $p(w)(q - p')^2/p^2$ to $p(w^{00})$, add $p(w)(q - p')(p' + p - q)/p^2$ to $p(w^{01})$ and $p(w^{10})$, and so on.

In other words, the point x is replaced by an interval I of length p, and each occurrence $x_k = x$ in our time series by a uniformly distributed random number from I. The definition of entropy profiles now applies to the modified word frequencies. This works perfectly well if w only occasionally contains more than one $*$. In the exceptional case of a periodic sequence with small period m, however, the method inserts randomness into a deterministic pattern. One could argue that $r_n(q) \equiv 1$ for $n > m$ is more reasonable (cf. below). For that reason, we shall avoid entropy profiles of periodic sequences with order larger than the period.

3. PROPERTIES

Preliminaries. We list some well-known facts concerning arbitrary stationary symbol sequences $\{i_k\}$. (Everything could be done for more than two symbols and more than one quantile, but we found it difficult to visualize and interpret functions r_n of more than one variable.) Let us start with

$$nH_1 \geqslant H_n \geqslant H_{n-1} \geqslant H_1 \tag{4}$$

Here $nH_1 = H_n$ if and only if any n consecutive values i_k are statistically independent, and $H_n = H_{n-1}$ if the symbol sequence is deterministic: $i_{k+n} = f(i_{k+1}, \ldots, i_{k+n-1})$ for each k. Now,

$$H_1 \geqslant H_n - H_{n-1} \geqslant H_{n+1} - H_n \tag{5}$$

says that the uncertainty of a letter in the sequence becomes smaller if we know more of its predecessors. Here equality holds if and only if the conditional probabilities

$$p(w_n \mid w_0, \ldots, w_{n-1}) = \frac{p(w_0 \cdots w_n)}{p(w_0 \cdots w_{n-1})}$$

and

$$p(w_n \mid w_1, \ldots, w_{n-1}) = \frac{p(w_1 \cdots w_n)}{p(w_1 \cdots w_{n-1})}$$

coincide.

Proof. Rearrange (1) to get

$$H_n - H_{n-1} = -\sum p(w_1 \cdots w_n) \operatorname{ld} p(w_n \mid w_1, \ldots, w_{n-1})$$

and

$$H_{n+1} - H_n = -\sum p(w_0 \cdots w_n) \operatorname{ld} p(w_n \mid w_0, \ldots, w_{n-1})$$

Their difference is

$$H_n - H_{n-1} - (H_{n+1} - H_n) = \sum p(w_0 \cdots w_n) \operatorname{ld} \frac{p(w_n | w_0, \ldots, w_{n-1})}{p(w_n | w_1, \ldots, w_{n-1})}$$

$$= \sum p(w_0 \cdots w_n) \operatorname{ld} \frac{p(w_0 \cdots w_n)}{\tilde{p}(w_0 \cdots w_n)}$$

with

$$\tilde{p}(w_0 \cdots w_n) = p(w_0 \cdots w_{n-1})\, p(w_n | w_1, \ldots, w_{n-1})$$

This can be interpreted as information gain[18] or Kullback–Leibler entropy,[11] which is nonnegative, and equals zero if and only if the distributions $\tilde{p}$ and p coincide.

Thus (5) is true. Equality holds in particular when $\{i_k\}$ is an $(n-1)$th-order Markov chain. This means that the probability for a symbol w_n to occur is determined by its $n-1$ predecessors $w_1, \ldots, w_{n-1}$, or $p(w_n | w_1, \ldots, w_{n-1}) = p(w_n | w_k, \ldots, w_{n-1})$ for all $k \leqslant 0$ and all $w_k, \ldots, w_n$. By the above proof, this is equivalent to $H_{m+1} - H_m = H_n - H_{n-1}$ for all $m \geqslant n$.

Since $H_n - H_{n-1}$ is a decreasing bounded sequence, the limit $h = \lim_{n \to \infty} H_n - H_{n-1}$ exists. It is called the entropy of the source of the symbol sequence. Clearly, $h \leqslant H_1$, with equality for independent symbols.

These facts will now be applied to our entropy profiles.

Bounds for Profiles. Randomness and Determinism. From (4) and (5) we have

$$0 \leqslant r_n(q) \leqslant 1 \quad \text{and} \quad 0 \leqslant R_n(q) \leqslant 1 \quad \text{for} \quad 0 < q < 1, \quad n \geqslant 2 \tag{6}$$

We add some comments on equality which show that the line $r = 0$ represents "randomness" and the line $r = 1$ "determinism." $R_n(q) = r_n(q) = 0$ holds for $n = 2, \ldots, m$ if any m successive values of the symbol sequence are independent. This holds independently of q if the x_k are independent (white noise). On the other hand, $r_n(q) = 1$ means that each symbol is determined by its $n-1$ predecessors: $i_{k+n} = f(i_{k+1}, \ldots, i_{k+n-1})$. This will happen when the sequence x_k with period n alternates between values $> s(q)$ and $< s(q)$ (e.g., n-band attractor for a chaotic map).

$R_n(q) = 1$ for some n implies $r_2(q) = 1$ by (3) and (6). Hence $H_2 = H_1$ and $i_{k+1} = f(i_k)$. Since $q \neq 0, 1$, it is not possible that $f(0) = 0$ or $f(1) = 1$. This means that $\{i_k\} = 010101\ldots$ or $101010\ldots$. Thus $q = 1/2$, and x_k alternates between values less than $s(q)$ and values larger than $s(q)$ ("two-band attractor").

Increase with n. Markov Partitions. From (5) and (3) it is immediate that

$$r_n(q) \leqslant r_{n+1}(q) \qquad \text{and} \qquad R_n(q) \leqslant R_{n+1}(q) \qquad \text{for all } q \text{ and } n \quad (7)$$

As proved above, $r_m(q) = r_{m+1}(q)$ for some m holds exactly if $p(w_m | w_1, \ldots, w_{m-1}) = p(w_m | w_0, \ldots, w_{m-1})$ for all w_i. Equality is true for all $n \geqslant m$ if and only if the symbol sequence forms a Markov chain of order $m-1$. This happens only for special values of q: it implies that the threshold $s(q)$ defines a Markov partition for the process which generates our time series (see below). A simple calculation shows that $R_m(q) = R_{m+1}(q)$ implies $r_2(q) = r_3(q) = \cdots = r_{m+1}(q)$.

Continuity and Limit. If the one-dimensional distribution of the x_k is continuous, $p(w)$ is a continuous function of q for each word $w = w_1 \cdots w_n$. (If $\{i_k\}$ corresponds to q and $\{\tilde{i}_k\}$ to $\tilde{q}$, then the probability that $i_k \neq \tilde{i}_k$ is $|q - \tilde{q}|$ and the probability for $i_{k+1} \cdots i_{k+n} \neq \tilde{i}_{k+1} \cdots \tilde{i}_{k+n}$ is not larger than $n \cdot |q - \tilde{q}|$.) As a consequence, $H_n(q)$, $r_n(q)$, and $R_n(q)$ are continuous functions, which can also be proved for our definition in the discrete case.

Next, let us mention that

$$r_\infty(q) := \lim_{n \to \infty} r_n(q) = \lim_{n \to \infty} R_n(q) = 1 - \frac{h(q)}{H_1(q)} \quad (8)$$

which measures the redundancy per symbol of the sequence $\{i_k\}$. Only in the Markov case [which includes the deterministic case $r_n(q) = 1$] will this limit be reached for finite n.

Invariance under Distortion. An entropy profile does not change when the time series undergoes a distortion $y_k = \psi(x_k)$, where ψ is an arbitrary (in general nonlinear) strictly increasing function. Namely, for any q, the threshold goes with the transformation: $s_y(q) = \psi(s_x(q))$. (Note that ψ may even be constant on intervals which do not contain any x_k. For a decreasing transformation ψ, the quantiles q and $1-q$ will interchange, and the profile will be reflected at the line $q = 1/2$.) In this sense, entropy profiles are independent of the one-dimensional distribution of the time series. Actually, our construction of profiles can be interpreted as a distortion which transforms $\{x_k\}$ into a time series which is uniformly distributed in $[0, 1]$.

Example. Let g be an arbitrary strictly increasing function on $[0, 1]$, ε a small, positive, irrational number, and $x_k = g(k\varepsilon - [k\varepsilon])$, where $[x]$ denotes the largest integer contained in x. The map $\psi = g^{-1}$ trans-

forms x_k into the sawtooth sequence $y_k = k\varepsilon - [k\varepsilon]$ for which $s(q) = q$. We shall prove that the graph of $r_n(q)$ forms a bell with a broad middle part near to 1 and with steep descent to 0 for $q \to 0$ and $q \to 1$. For sine functions, we get a similar result. More generally, this effect can be observed whenever we sample values x_k from a sufficiently smooth function, as in Fig. 1, with a very small sampling period.

Let $q > n\varepsilon$ and $1 - q > n\varepsilon$. Then the possible 0–1 words are $0^k 1^{n-k}$ and $1^k 0^{n-k}$ with $0 \leqslant k \leqslant n$. For instance, $y_k \cdots y_{k+n-1} = 01^{n-1}$ if and only if y_k lies in $]q - \varepsilon, q]$. Hence, taking Lebesgue measure as distribution of the y_k, we obtain $p(0^n) = q - (n-1)\varepsilon$, $p(1^n) = 1 - q - (n-1)\varepsilon$, and $p(w) = \varepsilon$ for the other $2n-2$ words of length n. We write $\alpha = q - (n-1)\varepsilon$, $\beta = 1 - q - (n-1)\varepsilon$, and $\varphi(p) = -p \operatorname{ld} p$, so that

$$H_n - H_{n-1} = \varphi(\alpha) - \varphi(\alpha + \varepsilon) + \varphi(\beta) - \varphi(\beta + \varepsilon) + 2\varphi(\varepsilon)$$

which goes to zero with ε, by continuity of φ at q and $1 - q$. Thus $r_n(q) \approx 1$ if ε is small enough, and q, $1 - q$ are much larger than ε. In a similar way one shows that for fixed ε and $q \to 0$ or $q \to 1$ we still have $r_n(q) \to 0$.

4. ONE-DIMENSIONAL DYNAMICS—ANALYTICAL RESULTS

Topological Invariance. For sequences of the form $x_{k+1} = f(x_k)$, where f is a continuous mapping of an interval I, entropy profiles show remarkable features. If two mappings $f, \tilde{f}: I \to I$ are topologically conjugate (that is, $f \circ \psi = \psi \circ \tilde{f}$ for some homeomorphism $\psi: I \to I$), then the corresponding profiles either coincide or $\tilde{r}_n(q) = r_n(1-q)$. The latter case corresponds to decreasing ψ. Thus our profiles describe general, topological features of the dynamics. We cannot expect information on properties which depend on differentiability conditions on f.

Thus, when we consider unimodal maps,[4, 19] any suitable one-parameter family will yield the same entropy profiles. In other words, the $r_n(q)$ indicate at which position in the bifurcation diagram ("Feigenbaum scenario"[4, 7, 14]) we are. Our numerical studies will be made with the quadratic family (or logistic map) $f(x) = ax(1-x)$, $0 < a \leqslant 4$, while piecewise linear maps, in particular those with an invariant ergodic measure, are more appropriate for exact calculations.

For our first remarks we do not even need an ergodic measure. Nevertheless, the following illustration might be helpful. The lines $x = s(q)$ and $y = s(q)$ divide the square $I \times I$ into two squares and two rectangles. The numbers $p(00)$, $p(01)$, $p(10)$, and $p(11)$ denote the relative frequency of points (x_k, x_{k+1}) in these four pieces, for a "typical orbit" $x_k = f^k(x_0)$. A similar illustration, with 2^n pieces, holds for n-words of the symbolic sequence.

A Standard Form for r_2. Let q be given. Clearly, $q = p(0) = p(00) + p(01) = p(00) + p(10)$. Thus, if we know $p(00)$, we can determine $p(01) = p(10)$ and $p(11)$. In other words, for fixed q we can consider $H_2(q)$ and $r_2(q)$ as functions of one parameter $p(00)$. (For the order $n = 3$ we get three free parameters.)

If $p(00) = 0$, which can happen only for $q \leqslant 1/2$, we get $H_2(q) = 2\varphi(q) + \varphi(1 - 2q)$ with $\varphi(p) = -p \operatorname{ld} p$, $\varphi(0) = \varphi(1) = 1$. Similarly, $H_2(q) = 2\varphi(1 - q) + \varphi(2q - 1)$ for $p(11) = 0$, which is possible for $q \geqslant 1/2$. These two expressions for H_2 together define a function $r_2^*(q)$, $0 < q < 1$, which is common to *all* time series of unimodal maps in the period-two window—from the first bifurcation up to the band-merging point (cf. Figs. 2a–2c and 5a). Namely, if the values of the time series alternate between a "lower band" and an "upper band," then for $q < 1/2$ our threshold will be in the lower band, and 00 cannot appear in the symbol sequence. For $q > 1/2$ we are in the upper band, and 11 is impossible. In Section 3 we proved already that $r_2(1/2) = 1$.

The term "standard form of r_2" is justified by the property that for all maps of the quadratic family with $3.5 \leqslant a < 4$, there is $q_1 > 0$ and $q_2 < 1$ such that $r_2(q) = r_2^*(q)$ for $0 < q < q_1$ and $q_2 < q < 1$ (cf. Figs. 2–5). To prove the second assertion (which is true also for $a = 4$, with $q_2 = 2/3$), we note that the nonzero fixed point x^* of the mapping f must represent a

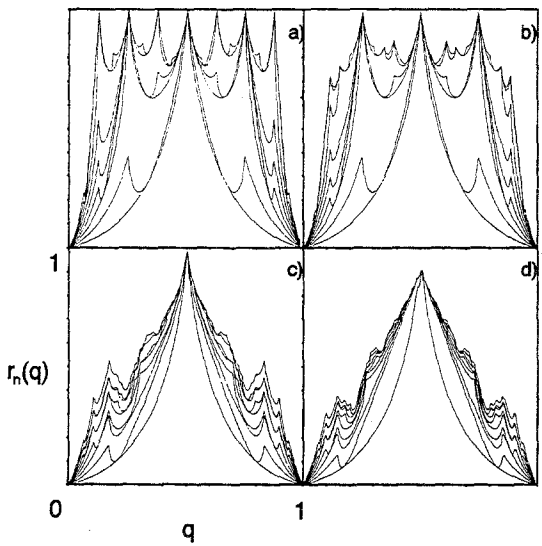

Fig. 2. Entropy profiles r_2 to r_{10} of the quadratic map for a band-merging scenario. (a) $a = 3.569945$ (near the Feigenbaum point), (b) $a = 3.579$ (chaotic four-band attractor), (c) $a = 3.675$ (chaotic two-band attractor), (d) $a = 3.68$ (chaotic one-band attractor).

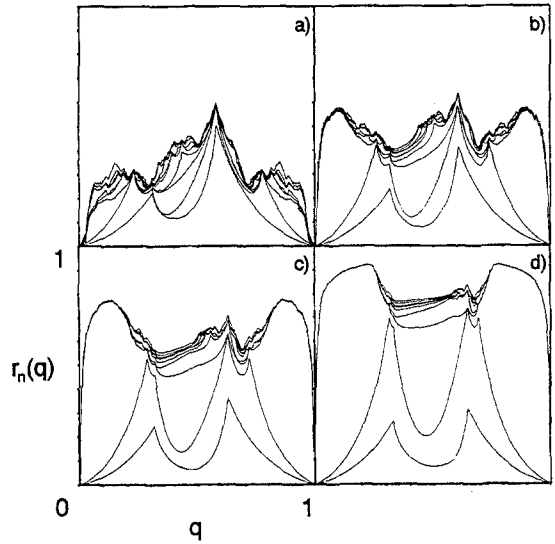

Fig. 3. Entropy profiles r_2 to r_{10} of the quadratic map near the period-three-window. (a) $a=3.82$, (b) $a=3.8274$, (c) $a=3.8282$, (d) $a=3.8284$.

quantile $q_2<1$ since for points x between 0 and x^* we find a $k\geqslant 1$ with $f^k(x)>x^*$. It is then obvious that $p(11)=0$ for $q>q_2$. For the first assertion, let x_- denote the (essential) infimum of a typical orbit of f. If $a<4$, then $x_-\neq 0$, so $f(x_-)>x_-$ and the quantile q_1 corresponding to $f(x_-)$ is larger than zero. For $q<q_1$ we have $p(00)=0$.

The universality of r_2^* also holds for multimodal maps f which do not "start or end with a fixed point." (For $n>2$ we find a universal r_n^* if we exclude maps which start or end with a periodic point of period smaller

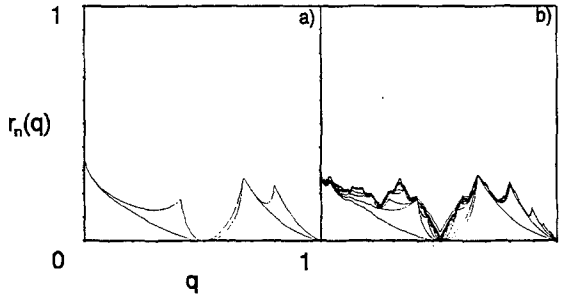

Fig. 4. Entropy profiles for the quadratic map at $a=4$ (fully developed chaos). (a) Analytical result for r_2 and r_3, (b) numerical result for r_2 to r_{10}.

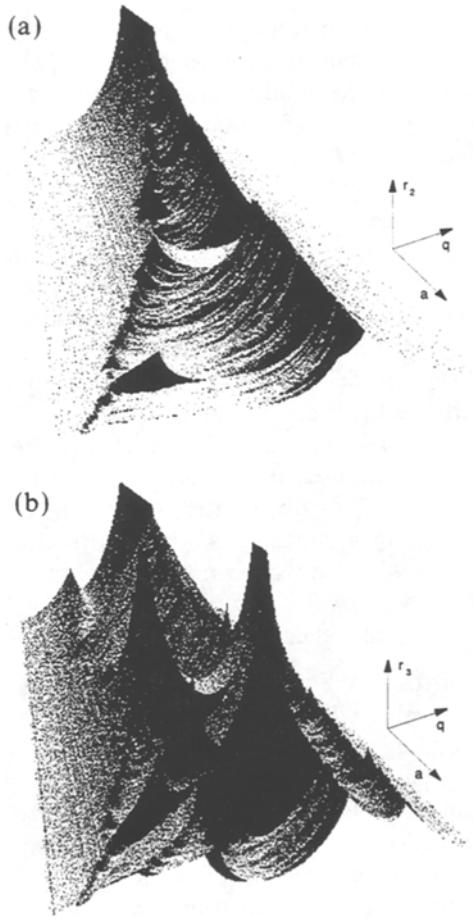

Fig. 5. Entropy profiles of 491 quadratic maps $f(x) = ax(1-x)$, $3.5 \leqslant a \leqslant 3.99$. (a) r_2, (b) r_3.

than n.) For continuous time series, however, as shown in Fig. 1, these border effects are negligible since the minimum and maximum are almost fixed points.

Profiles of Periodic Windows. Generalizing our argument on r_2^*, we shall now see that within each window of period p of a family of one-dimensional maps, all entropy profiles of order $n \leqslant p$ remain unchanged! This is illustrated in Fig. 5.

A window of period p in the quadratic family is the set of all mappings between a tangent bifurcation which gives birth to an attractive orbit of period p, up to the corresponding band-merging point where p bands unify.

For our purpose, we can define a period-p window in an arbitrary parametric family of one-dimensional maps by a weaker property. We require that the functions f of the window are characterized by a cyclic permutation $(m_0 = 1, m_1, ..., m_{p-1})$ as follows. There are p pairwise disjoint subintervals (or "bands") $I_1, ..., I_p$ of I, written in natural order with I_1 on the left, such that $f^k(x) \in I_{m_k}$ for $x \in I_1$ and $0 \leqslant k < p$, and $f^p(x) \in I_1$ again. Now if $(j-1)/p < q < j/p$, the threshold belongs to band I_j, and we define the word $u = i_0 i_1 \cdots i_{p-1} i_0 \cdots i_{n-2}$ with $i_k = 0$ if $m_k < j$, with $i_k = 1$ if $m_k > j$, and $i_k = *$ for $m_k = j$. It is not difficult to see that the possible words w of length n for the symbol sequence are just the subwords of u of length n. If w does not contain a $*$, then $p(w) = 1/p$, since a whole band is concerned. If w contains $*$, then we consider the words w^0 and w^1 defined at the end of Section 2, and distribute the mass $1/p$ to $p(w^0)$ and $p(w^1)$ according to the position of q between $(j-1)/p$ and j/p. In this way the profiles $r_n(q)$ with $n \leqslant p$ are correctly determined, using only the information contained in the cyclic permutation. It does not matter whether f has an attractive orbit of period p, $2p$, $4p$,..., or a continuous p-band attractor.

For the case $q = k/p$ where we have no $*$, we obtain $r_p(q) = 1$ [see the discussion of (6)]. Moreover, $r_n(q) = 1$ for $n < p$ if the sequence $i_0 i_1 \cdots i_{p-1} i_0 \cdots i_{p-1} i_0 \cdots$ is periodic with period n.

Feigenbaum Points. A Feigenbaum point in a parametric family of maps is a parameter value which belongs to a decreasing sequence of windows. The corresponding sequence of periods tends to ∞. So, by the results of the preceding paragraph, there is a dense set of rationals q for which $r_n(q) = 1$ for sufficiently large n, and in the limit $r_\infty(q) \equiv 1$ for each q. This agrees with the well-known fact that the action of f on Feigenbaum attractors has topological entropy $h = 0$.[19] For the "classical" Feigenbaum point of type 2^∞, the situation is illustrated in Fig. 2a.

Maps with Ergodic Invariant Measure. Let us now assume that there is a probability measure μ on I which is invariant with respect to f [i.e., $\mu(B) = \mu(f^{-1}(B))$ for all Borel sets $B \subset I$] and ergodic [a set A with $A = f^{-1}(A)$ has μ-measure 0 or 1]. The ergodic theorem says that for μ-almost all initial points x_0, the measures μ_N which assign mass $1/(N-1)$ to each of the points $f^k(x_0)$, $k = 0, 1, ..., N-1$, converge weakly to μ for $N \to \infty$. Thus in this case, one can use μ to determine H_n and r_n directly, instead of evaluating trajectories.

To determine quantiles, one uses the distribution function $F(t) = \mu(\{x \mid x < t\})$ of the probability measure. F maps I onto the unit interval $[0, 1]$, and there is a mapping $\tilde{f}$ from $[0, 1]$ into itself with $F \cdot f = \tilde{f} \cdot F$ which has the uniform distribution, the usual Lebesgue measure λ on $[0, 1]$, as an ergodic invariant probability measure. For any q we can now

determine $p(00)=\lambda(\{x \mid x<q, \tilde{f}(x)<q\})$, and similarly for each word of length n.

The Tent Map. We demonstrate the method for the tent map, which has λ as ergodic invariant measure. The tent map $g(x)=2x$ for $x<1/2$, $g(x)=2(1-x)$ for $x \geqslant 1/2$, is conjugate to $f(x)=4x(1-x)$.[4, 7] (Actually, the above method applied to f yields $\tilde{f}=g$.) It is interesting to note that from the viewpoint of *linear* time series analysis, these functions provide a model of white noise: the time series generated by both functions are known to be δ-correlated for almost all starting points in [0, 1]. In fact the values of r_n were rather small in this case (Fig. 4). Nevertheless, nonlinear dependences in the data were sufficient to keep the entropy profiles well away from zero, especially for small q.

r_2 and r_3 for the Tent Map. We determine $p(00)$ as the sum of the lengths of intervals on which $x<q$, $g(x)<q$. Now $g(x)>q$ if and only if $q/2<x<1-q/2$. For $q<1-q/2$ this implies $p(00)=p(01)=p(10)=q/2$. Consequently, $H_2=3\varphi(q/2)+\varphi(1-3q/2)$ for $q<2/3$. For $q>2/3$ we obtain $p(11)=0$, $p(01)=1-q=p(10)$, which leads to $H_2(q)=2\varphi(1-q)+\varphi(2q-1)$. Note the coincidence with the result for H_2 above.

For $H_3(q)$ we get $4\varphi(q/4)+2\varphi(q/2)+\varphi(1-2q)$ if $q \leqslant 2/5$, $4\varphi(q/4)+3\varphi(1/2-3q/4)+\varphi(5q/4-1/2)$ for $2/5 \leqslant q \leqslant 2/3$, $2\varphi(q/4)+\varphi(1-5q/4)+\varphi(1-q)+\varphi(7q/4-1)$ for $2/3 \leqslant q \leqslant 4/5$, and $3\varphi(1-q)+\varphi(3q-2)$ for $q \geqslant 4/5$.

The corresponding profiles r_2, r_3 are shown in Fig. 4a. Note that $\lim_{q \to 0} r_2(q)=\lim_{q \to 0} r_3(q)=1/2$—the only case of a unimodal map for which this limit is nonzero. Let us now discuss the peaks of r_2 and r_3.

Markov Partitions. For $q=1/2$, the sequence $\{i_k\}$ consists of statistically independent symbols if the starting point x_1 is chosen with respect to Lebesgue measure (i_n depends on the nth digit in the binary representation of x_1). Thus, $r_n(1/2)=0$ for $n \geqslant 2$.

Next, consider $q=2/3$. With $I_0=[0, 2/3]$ and $I_1=[2/3, 1]$, we have $g(I_1)=I_0$. In terms of conditional probabilities, $p(w_n=0 \mid w_k, \ldots, w_{n-1})=1$ whenever $w_{n-1}=1$, for all $k<n$ and all possible values of $w_k, \ldots, w_{n-2}$. Moreover, g maps the left half of I_0 onto I_0 and the right half twice onto I_1. Consequently, $p(w_n=0 \mid w_{n-1}=0)=1/2$. We show that $p(w_n=0 \mid w_k, \ldots, w_{n-2}, w_{n-1}=0)=1/2$ for any $k<n$ and any values of $w_k, \ldots, w_{n-2}$. The set of all x_k which generate the symbols $w_k, \ldots, w_{n-2}, 0$ consists of several intervals which are mapped by g^{n-1-k} onto I_0. We can choose the intervals so that g^{n-1-k} is a linear function on each interval.

Then g^{n-k} maps exactly one half of each of these intervals onto I_0. This proves that we have an ordinary Markov chain. Thus

$$r_n(\tfrac{2}{3}) = r_2(\tfrac{2}{3}) = (\mathrm{ld}\, 3 - \tfrac{4}{3})/(\mathrm{ld}\, 3 - \tfrac{2}{3}) \approx 0.274 \qquad \text{for} \quad n \geqslant 3$$

as seen in Fig. 4b.

For $q = 2/5$, the partition $I_{00} = [0, 1/5]$, $I_{01} = [1/5, 2/5]$, $I_{11} = [2/5, 4/5]$, and $I_{10} = [4/5, 1]$ induces a second-order Markov chain with $p(0\,|\,01) = 0$, $p(0\,|\,00) = p(0\,|\,10) = p(0\,|\,11) = 1/2$. The argument is the same as above: each of the intervals, except I_{01}, divides into two halves when we consider the next letter. We have

$$r_n(\tfrac{2}{5}) = r_3(\tfrac{2}{5}) \approx 0.176 \qquad \text{for} \quad n > 3$$

For $q = 4/5$ the dynamics seems even more deterministic: $I_1 = I_{10} = I_{100} = [4/5, 1]$ and $I_{01} = [2/5, 3/5]$, and the remaining two intervals constitute I_{00}. This is not a Markov partition, however, since the image of I_{00} does not evenly cover the partition sets. In fact, $r_3(4/5) \approx 0.237$, $r_4(4/5) \approx 0.246$, and $r_5(4/5) \approx 0.249$.

Problems. At this point, we have to mention the intricacies connected with the fact that a mapping, even a quadratic one, need not generate a stationary process. It was Milnor who asked whether for quadratic maps f without attracting periodic orbits, the time average of the orbit $\{x_k\}$ converges weakly to a fixed continuous f-invariant distribution for Lebesgue-almost-all choices of the initial point x_0 (ref. 16; cf. ref. 12). The physicist's obvious answer is "yes." A proof was given when f either has a Feigenbaum attractor[4, 19] or an absolutely continuous invariant measure,[3] which covers a lot of parameter values a, including $a = 4$. Recently, however, Hofbauer and Keller[13] have found other values of a where either weak convergence does not take place, or the limit is a discrete measure. Strictly speaking, this implies that the famous bifurcation diagram is not properly defined!

This raises the question whether the entropy profiles, estimated by taking limits from an orbit $f^k(x_0)$, are the same for Lebesgue-almost-all choices of x_0. The answer is positive when an attracting measure in the sense of Milnor exists, but also in the presence of attractive periodic orbits. Note that not all x_0 can be taken.

We think that the connection between statistical and topological features in dynamics provides new and challenging problems. We confine ourselves to the topic of this paper: which functions $r_2(q)$ of quadratic maps are identical, outside periodic windows? Which absolutely continuous invariant measures of quadratic maps have the property that the

critical point corresponds to $q = 1/2$? (We know only $a = 4$.) In periodic windows, the q of the critical point may be arbitrarily near to 1, but is there a lower bound?

5. SIMULATION RESULTS FOR UNIMODAL MAPS

The "Symmetry-Breaking Scenario" of Fig. 5. All entropy profiles of quadratic mappings before the band-merging point are symmetric, with a peak at $q = 1/2$ reaching to 1 (Figs. 2a–2c). Beyond that point, symmetry is broken, but only in the central part, which increases with a. Near 0 and 1, symmetry is preserved for $a < 4$, as we proved (Figs. 2d, 3, and 4). The main peak is clearly seen, except for parameters near a periodic window. It wanders from 1/2 in Fig. 2 to 2/3 in Fig. 4.

Beyond the band-merging point (Fig. 3d), the function r_2 decreases with increasing a, but only in the central part, where it becomes asymmetric. Note that the decrease of $r_2(q, a)$ for fixed q is interrupted by periodic windows. The period-three window, where r_2 remains unchanged, is clearly seen in the middle of Fig. 5a, and the little shadow in the foreground indicates the period-four window at $a \approx 3.96$. It is curious to contrast the successive symmetry breaking in r_2 with the position of the critical point. r_2 is symmetric when the critical point is far from the median (1/2-quantile) of the distribution of the x_k, and r_2 becomes completely asymmetric at $a = 4$, where the critical point coincides with the median.

Intermittency. An interesting scenario, typical for intermittency, is shown in Fig. 3. Just below $a_c = 1 + \sqrt{8}$ where a tangent bifurcation gives rise to the period-three window, the dynamics of f can be characterized as a long-lasting rather periodic change between three "bands" [located at the "channels" between the graph of f^3 and $g(x) = x$], interrupted by short chaotic bursts.[17] In the "quasiperiodic" parts, f^3 maps each "band" into itself in a monotonic manner. When a comes near to a_c, the chaotic parts become negligible for the estimation of $p(w)$, and the action of f^3 is the same as in the example of Section 3, with a small ε. Hence $r_n(q)$ tends to 1 when a approaches a_c, for every $n > 3$ and $0 < q < 1$. For $q < 1/3$ and $q > 2/3$ convergence is three times faster as in Section 3 (indicated by the coincidence of r_n), while the middle part carries the effects of chaotic bursts.

Numerical Methods and Accuracy. All simulations were made on a personal computer. Applying a quicksort algorithm to $\{x_k\}$, we determine the $s(q)$. The $p(w)$ can be calculated by integer arithmetic. As a rule, one needs at least ten times more data than words w for which we estimate frequencies. For quadratic maps we took $N = 30{,}000$. When the order n of profiles does not exceed five, there will be quite accurate results with 1000

values. Note that the sampling error is maximal for values of q near to zero or one, due to the bad statistics when one symbol is rare. Compare Figs. 4a and 4b at $q = 0$ (in this particular case, another initial value yields better coincidence). On the other hand, the method is quite stable with respect to outliers (even wrong values) and noise.

6. CONCLUSION

We have introduced new functions describing statistical dependences in a time series. Entropy profiles are defined by a certain level-crossing statistics combined with an appropriate standardization of thresholds. Their properties were discussed both in general and through the example of quadratic (logistic) maps. Entropy profiles are invariant with respect to arbitrary monotonic distortion of the domain, which makes them applicable to time series of ordinal data. The method applies well to a large class of speech signals[2] and some experiments to study attractors have been very promising.

ACKNOWLEDGMENTS

We are grateful to Manfred Streich for cooperation in the early stages of this work. He determined the first interesting entropy profiles. We also thank Matthias Heilfort for advice concerning the computer programming.

REFERENCES

1. R. Badii and A. Politi, *J. Stat. Phys.* **40**:725–750 (1985).
2. C. Bandt and B. Pompe, Entropy profiles of speech signals, preprint, Greifswald (1992).
3. A. M. Blokh and M. Yu. Lyubich, *Funct. Anal. Appl.* **21**:148–150 (1987).
4. P. Collet and J.-P. Eckmann, *Iterated Maps on the Interval as Dynamical Systems* (Birkhäuser, Basel, 1980).
5. P. Collet, J. L. Lebowitz, and A. Porzio, *J. Stat. Phys.* **47**:609–643 (1987).
6. C. D. Cutler, *J. Stat. Phys.* **62**:651–708 (1991).
7. J.-P. Eckmann and D. Ruelle, *Rev. Mod. Phys.* **57**:617–656 (1985).
8. A. M. Fraser, *IEEE Trans. Information Theory* **35**:245–262 (1989); A. M. Fraser and H. L. Swinney, *Phys. Rev. A* **33**:1134–1140 (1986).
9. P. Grassberger, *Z. Naturforsch.* **43a**:671–680 (1988).
10. P. Grassberger and I. Procaccia, *Phys. Rev. A* **28**:2591–2593 (1983).
11. P. Grassberger, T. Schreiber, and C. Schaffrath, *Bifurcation Chaos* **1**:521–548 (1991).
12. J. Guckenheimer and S. Johnson, *Ann. Math.* **132**:71–130 (1990).
13. F. Hofbauer and G. Keller, *Commun. Math. Phys.* **127**:319–337 (1990).
14. S. Isola and A. Politi, *J. Stat. Phys.* **61**:263–291 (1990).

15. P. Martien, S. C. Pope, P. L. Scott, and R. S. Shaw, *Phys. Lett.* **110A**:399–404 (1985); R. Shaw, *The Dripping Faucet as a Model Chaotic System* (Aerial Press, Santa Cruz, California, 1984).
16. J. Milnor, *Commun. Math. Phys.* **99**:177–195 (1985); **102**:517–519 (1985).
17. Y. Pomeau and P. Manneville, *Commun. Math. Phys.* **74**:189–197 (1980).
18. A. Renyi, *Probability Theory* (North-Holland, Amsterdam, 1970).
19. S. van Strien, Smooth dynamics on the interval, in *New Directions in Dynamical Systems*, T. Bedford and J. Swift, eds. (Cambridge University Press, 1988), pp. 57–119.