Stochastics and Dynamics, Vol. 1, No. 1 (2001) 1–17

PARTITION-BASED ENTROPIES OF DETERMINISTIC AND STOCHASTIC MAPS

W. EBELING* and R. STEUER†
Institute of Physics, Humboldt-University, 10115 Berlin, Germany

M. R. TITCHENER‡
Humboldt Fellow

Received 17 January 2001
Revised 5 February 2001

In this paper we explore the relationship between the Kolmogorov–Sinai entropy, the sum of positive Lyapunov exponents, denoted here as the *Pesin entropy* and a new measure, the *T-entropy*, for nonlinear maps. We demonstrate that threshold-crossing partitions are effective in deriving representative symbolic realisations for the real-valued time series. We describe the recently developed entropy measure for finite strings and compare with values derived from the application of Shannon's theory. These techniques are further applied to a simple stochastic system, appearing to further confirm recent theoretical results on Lyapunov exponents.

Keywords: Shannon entropy, complexity

1. Introduction

We consider several partition based measures of information and complexity, in particular Shannon's conditional entropy and Kolmogorov–Sinai entropy. We further consider a novel grammar-based complexity/information measure (*T-entropy*) defined for finite strings. Our aim is to compare these measures with dynamical measures based on Lyapunov exponents. We will demonstrate that there exist close relations between these measures, which seems to remain valid for stochastic systems. Our special interest in these relations is due to the fact that recently the concepts of Lyapunov exponents have been extended to stochastic systems [2], [3].

The paper is organized as follows. In the first section we review the concept of symbolic dynamics, a statistical description of the dynamics at the macroscopic

*E-mail: ebeling@physik.hu-berlin.de
†E-mail: steuer@physik.hu-berlin.de
‡E-mail: mark@tcode.tcs.auckland.ac.nz; on leave from Department of Computer Science, The University of Auckland, New Zealand.

level resulting from a coarse-graining of the state space. We then discuss n-block entropies, introduced by Shannon, as a well-known measure of information for symbolic sequences. This is supplemented by some properties of the Shannon entropy derived in earlier work. In the next section we introduce the grammar-based complexity measure referred to here as the *T-entropy*. The T-entropy is demonstrated to closely approximate the Kolmogorov–Sinai entropy for simple example systems and has the potential to be useful when investigating and characterizing empirical systems. In the last section we investigate the effects of noise on these measures giving support to the application on experimental and empirical time series and relating to recent results on Lyapunov exponents for stochastic systems.

2. Partition-Based Entropies

2.1. *Symbolic dynamics*

Symbolic dynamics is a coarse grained description of microscopic dynamics acting on a continuous state space Γ. This is obtained by introducing a partition $\mathcal{P} = \{P_1, \ldots, P_l\}$, which divides Γ into l non-overlapping sets, each labeled by symbols a_i drawn from an alphabet $\mathcal{A} = \{a_1, \ldots, a_l\}$ of cardinality $\#A = l$. Under the action of the dynamics the trajectory of the dynamical system $f : \Gamma \to \Gamma$ visits various elements of the partition $\mathcal{P}$. The time evolution is thus encoded as a sequence of symbols.

Given a careful choice of the partition, no relevant feature need to be lost in this encoding. In particular, the set $\Gamma_i^{(N)}$ that comprises all possible microscopic initial points which are compatible with a given finite symbol sequence $S_0 S_1 \cdots S_N$ is given by

$$\Gamma_i^{(N)} = P_{S_0} \cap f^{-1}(P_{S_1}) \cap \cdots \cap f^{-N}(P_{S_N}) \,, \tag{2.1}$$

where P_{S_k} denotes the set of all points assigned to the symbol S_k and $f^{-k}(P_{S_k})$ denotes the set of all points mapped by f^k to the partition cell P_{S_k}. For any N, the family $\{\Gamma_i^{(N)}\}$ again forms a partition of the state space Γ. Since these refined partitions are produced by the dynamics itself they are usually referred to as *dynamical refinement*. A partition is called *generating*, if the sets $\Gamma_i^{(N)}$ keep decreasing in size for increasing sequence length, i.e. each infinitely long symbol sequence corresponds to an individual initial point in state space [10].

Unfortunately, constructing a generating partition for arbitrary nonhyperbolic dynamical systems is not a trivial task. One construction rule we have is the one proposed by Grassberger and Kantz [15] for the dissipative Hénon map, and later applied to other systems [5]. An additional obstacle is that, in the presence of noise the concept of a generating partition is no longer well defined. This leads to the notion of an *optimal measurement* partition, which will be discussed in a subsequent section. For such nongenerating partitions some microscopic details of the dynamics might be lost on the symbolic level, but most temporal correlations are still embedded in the structure of the n-word distributions. Thus even without a generating partition, symbolic dynamics offers powerful methods for characterizing and investigating a given dynamical system.

2.2. *The shannon n-gram entropy*

Following Shannon's approach [27] the n-block entropies H_n of a symbolic sequence S are given by

$$H_n := -\sum p_i^{(n)} \log p_i^{(n)} , \tag{2.2}$$

where $p_i^{(n)}$ denotes the stationary probability of a substring i of length n occurring within the infinite sequence S. The summation is carried out over all substrings with $p_i^{(n)} > 0$. The n-block entropies H_n are a measure of *uncertainty* and give the *average* amount of information contained in a word of length n. Consequently, the *differential* or *conditional* entropies

$$h_n := H_{n+1} - H_n , \quad h_0 := H_1 \tag{2.3}$$

give the average amount of information required to predict the $(n+1)$th symbol, given the preceding n symbols. A quantity of particular interest is the *entropy of the source* or *limit* entropy h.

$$h := \lim_{n\to\infty} \frac{H_n}{n} = \lim_{n\to\infty} h_n . \tag{2.4}$$

Since the series of h_n is monotonically decreasing and bounded from below by zero, this limit does exist and can be interpreted as the average amount of information needed to predict the next symbol if the whole past is known. The convergence of the conditional entropies h_n to their limit h can be taken as a measure of correlations. If correlations do not exist beyond a finite range m (Markov chain property), the asymptotic value is reached for $n = m$.

$$h_n = h_m \quad \forall\, n \geq m . \tag{2.5}$$

However, for most systems the mth order Markov property applies to an approximate description rather than to the sequence itself. The generic case yields an exponential decay of the conditional entropy h_n to its limit [8], [9], [11]. Sequences showing a subexponential decay of the h_n are related to long-range correlations [16], [25].

2.2.1. *Numerical estimation*

The application of such ideas to empirical sequences requires the precise estimation of higher order entropies. A naîve estimation of n-block probabilities by the observed frequencies of occurrence $k_i^{(n)}$ typically breaks down if words occur less than ten times. In this case the n-block entropies H_n get systematically underestimated [16].

$$H_n^{\text{observed}} = H_n - \frac{M_n - 1}{2N} + \mathcal{O}\left(\frac{1}{N^2}\right) . \tag{2.6}$$

Here $M_n \leq l^n$ denotes the number of different n-blocks within the sequence. A correction formula using higher moments has been given by Grassberger [14]. The estimation of higher order entropies was also improved considerably when applying reasonable assumptions about the n-word probabilities [25].

2.3. *Kolmogorov entropy and pesin identity*

Different partitions will yield a different symbolic representation of the system. Thus the resulting dynamical entropies $h_n(\mathcal{P})$ and their limit $h(\mathcal{P})$ are dependent on the chosen partition $\mathcal{P}$. To obtain a partition-independent value to characterize the underlying real-valued sequence, the *Kolmogorov–Sinai entropy* or *KS-entropy* is defined as the supremum over all possible partitions [21], [28].

$$h_{\mathrm{KS}} = \sup_{\mathcal{P}} h(\mathcal{P}) = \sup_{\mathcal{P}} \lim_{n\to\infty} h_n(\mathcal{P})\,. \tag{2.7}$$

An alternative definition could involve the size of the partition cells P_i tending to zero. In the case of a generating partition the supremum may be omitted [10], [26].

As another partition independent quantity, we consider the positive Lyapunov exponents λ^+, which characterize the exponential divergence of nearby trajectories. This divergence implies a loss of information about their future position. It is therefore plausible to argue that the (averaged) sum of positive Lyapunov exponents and the Kolmogorov–Sinai entropy are not independent [10], [24], [26]. For clarity we shall define the sum of positive Lyapunov exponents for a dynamical system to be the *Pesin entropy* h_p.

$$h_p := \sum_i \lambda_i^+ \text{ and } 0 \text{ if } \lambda_1 < 0\,. \tag{2.8}$$

The Pesin entropy h_p is not partition based but is identical to the Kolmogorov–Sinai entropy h_{KS} for many systems we are interested in (Pesin identity) [10]. Moreover, in most cases the Pesin entropy is the only way to obtain a reliable estimate of the Kolmogorov–Sinai entropy. However, in the present work we will use it as an independent way to check and compare the results obtained by partition-based entropies.

3. The T-entropy

3.1. *The lempel ziv string complexity*

Whereas the Kolmogorov–Sinai and Shannon entropies are well established in the literature, the new measure requires more explanation. The T-entropy is defined in terms of a grammar-based [8] string T-complexity measure [31]–[33] similar to the *production complexity* for finite strings proposed by Lempel and Ziv [22]. Thus it is useful to introduce the underlying concepts by describing the LZ algorithm.

We denote by $\mathcal{A}^*$ the free monoid generated by the alphabet $\mathcal{A}$ under catenation. $\mathcal{A}^*$ is thus the infinite set of all finite strings formed from the elements of $\mathcal{A}$ and $\mathcal{A}^+ := \mathcal{A}^* \backslash \Lambda$ excludes the *empty string*, denoted by Λ. $\mathcal{A}^n$ further denotes the set of all strings of length n.

Given a finite string $x(n) \in A^n$, Lempel and Ziv defined the *production complexity* $C_S(x)$ to be the least "effort" required to produce x from $\mathcal{A}$. They proposed a recursive linear-pattern copying (RLPC) algorithm for this and measure the effort as the total number of RLPC steps required to produce the string.

Parsing x left-to-right, and selecting substrings $p_i \in \mathcal{A}^+$, not previously seen, we may add each new pattern to a vocabulary set $\mathcal{V}$. With $\mathcal{V}$ initially empty we add the alphabet symbols as required and then new patterns formed from suffixing a single symbol from $\mathcal{A}$ to any existing pattern in $\mathcal{V}$. The very first pattern to appear in the vocabulary is necessarily an alphabet symbol.

To give an example, assume the alphabet $\mathcal{A} = \{0, 1\}$, and string $x =$ '1011010100010'. Here the first new pattern, left-to-right is a '1'. This is added to $\mathcal{V}$. The next new pattern is '0'. After the first two steps $\mathcal{V} = \{1, 0\}$. The third bit already appears in $\mathcal{V}$, so we append the next bit and add the resultant pattern '11' to $\mathcal{V}$. We now have $\mathcal{V} = \{1, 0, 11\}$. The vocabulary continues to evolve in this way until we finally have $\mathcal{V} = \{1, 0, 11, 01, 010, 00, 10\}$. Thus the vocabulary in effect records the history for the production of the string, and the size of the vocabulary, in this case $\#\mathcal{V} = 7$, is a measure of the production complexity of x.

There may exist for any given string, a multiplicity of choices in the evolution of a vocabulary. In practice, the minimum vocabulary size may be derived only by way of an exhaustive search process. Nevertheless, strings of a given length with repetitive structures will tend to result in smaller vocabularies than those which are more complex in structure.

Lempel and Ziv show that for binary strings the production complexity $C_S(x(n))$ is bounded from above by [22]:

$$C_S(x) < \frac{n}{(1-\varepsilon_n)\log_2 n} \quad \text{where} \quad \varepsilon_n = 2\frac{1+\log_2\log_2(2n)}{\log_2 n}\,.$$

Further, for an ergodic source $C_S(x(n))$ is in the limit $n \to \infty$ bounded by a function of Shannon's source entropy. This connection has motivated the investigation of the symbolic dynamics of nonlinear dynamical systems using the LZ complexity [20]. However an explicit relationship between the LZ complexity and the Kolmogorov–Sinai entropy of the system is not known.

3.2. *The T-complexity of a string*

The *T-complexity* [31], [32] of a string $x \in \mathcal{A}^n$ is defined along similar lines to the LZ production complexity but a *recursive hierarchical pattern copying* (RHPC) algorithm [33] is used instead.

Our measure computes the *effective number* of *T-augmentation* steps [33] required to generate x from $\mathcal{A}$. By using the RHPC algorithm no search is necessary to determine an optimal value. The T-complexity may be thus computed efficiently and directly from a any string[a] and the resultant value is unique [23].

The string $x(n)$ is parsed to derive constituent patterns $p_i \in \mathcal{A}^+$, and associated *copy-exponents* $k_i \in \mathbb{N}^+$, $i = 1, 2, \ldots, q$, where $q \in \mathbb{N}^+$ satisfying:

$$x = p_q^{k_q} p_{q-1}^{k_{q-1}} \cdots p_i^{k_i} \cdots p_1^{k_1} a_0 \,, \quad a_0 \in \mathcal{A} \,. \tag{3.9}$$

Each pattern p_i is further constrained to satisfy:

$$p_i = p_{i-1}^{m_{i,i-1}} p_{i-2}^{m_{i,i-2}} \cdots p_j^{m_{i,j}} \cdots p_1^{m_{i,1}} a_i \,, \quad a_i \in \mathcal{A} \text{ and } 0 \le m_{i,j} \le k_j \,. \tag{3.10}$$

The *T-complexity* $C_T(x(n))$ is defined[b] in terms of the copy-exponents k_i:

$$C_T(x(n)) = \sum_i^q \ln(k_i + 1) \,. \tag{3.11}$$

One may verify that $C_T(x(n))$ is minimal for a string comprising a single repeating character. From Eq. (3.11) we have:

$$\ln n \le C_T(x(n)) \,. \tag{3.12}$$

The upper bound is more difficult to derive. However, for $n > n_0$

$$C_T(x(n)) \le \text{li}(\ln 2 \ln(\#\mathcal{A}^n)) \,, \tag{3.13}$$

where $\text{li}(z) := \int_0^z du/\ln u$ is the logarithmic integral function [1]. For a binary alphabet $n_0 \approx 15$, i.e. small enough to be of no consequence as we are typically concerned with strings in the range of $n = 10^4$–10^6 bits.

We now illustrate the RHPC parsing algorithm using the example, introduced already in connection with the LZ production complexity, to derive the hierarchy of patterns $p_i \in \mathcal{A}^*$ and corresponding copy-exponents $k_i \in \mathbb{N}^+$. In practice we parse the string repeatedly from *left-to-right* but select the patterns from *right-to-left*.

With $\mathcal{A} = \{0, 1\}$, and $x =$ '1011010100010', we see by inspection that the symbol $a_0 =$ '0'. It is clear from Eq. (3.9) that p_1 immediately precedes a_0, so evidently $p_1 =$ '1' (the *penultimate* symbol, assuming a left-to-right scanning '10110101000$\underline{1}$0'). Clearly, $p_1 =$ '1' $= a_1 \in \mathcal{A}$ satisfies Eq. (3.10).

As p_1 does not repeat immediately to the left, we conclude $k_1 = 1$. We next group the symbols in x *left-to-right*, prefixing the pattern p_1 each time it appears with its immediate successor symbol. Emphasizing the formation of groups by way of underlining, we have $x =$ '$\underline{10}$ $\underline{11}$ $\underline{0}$ $\underline{10}$ $\underline{10}$ $\underline{0}$ $\underline{0}$ $\underline{10}$'.

[a] A general purpose UNIX compatible software application is freely available for this upon request.
[b] In contrast to earlier descriptions [31], [32] we have here used the natural logarithm. This allows us to more properly state the upper bound for the measure.

The *penultimate* (second-to-last) group is $p_2 =$ '0'. However, the pattern repeats once more to the left, i.e. $k_2 = 2$ (10 11 0 10 10 0 0 10). Once more we reform the existing groups *left-to-right* but this time each appearance of $p_2 =$ '0', or consecutive run of p_2 up to a maximum of $k_2 = 2$ times, is the cue for forming a new single group by appending the immediate successor pattern. We now have $x =$ '10 11 0 10 10 0 0 10'.

This process of, (i) identifying the penultimate pattern (left-to-right), (ii) identifying the corresponding repetition factor, and (iii) reforming the subgroups, is repeated until only one single group remains, encompassing the whole of x.

For completeness we trace the remaining steps to exhaustion. The penultimate group now is $p_3 =$ '10', with $k_3 = 1$. Reforming the existing groups results in 10 11 0 10 10 0 0 10. We see then that $p_4 =$ '010' and $k_4 = 1$ and we have, 10 11 0 10 10 0 0 10. Finally, $p_5 =$ '1011' and $k_5 = 1$ and 10 11 0 10 10 0 0 10.

In summary we have $p_5 =$ '1011', $p_4 =$ '010', $p_3 =$ '10', $p_2 =$ '0', $p_1 =$ '1', and $k_5 = 1$, $k_4 = 1$, $k_3 = 1$, $k_2 = 2$, $k_1 = 1$. The *T-complexity* by Eq. (3.11), is thus $\sum_{i=1}^{5} \ln(k_i + 1) = \ln(\prod_{i=1}^{5}(k_i + 1)) = \ln(48) \approx 3.87$.

3.3. *The T-information and average T-entropy of a string*

Motivation for the following definitions is broadly outlined in [33]. The *T-information* $I_T(x(n))$ of the string $x(n)$ is defined to be the *inverse logarithmic integral* of the T-complexity divided by a scaling constant, $\ln 2$:

$$I_T(x(n)) = \mathrm{li}^{-1}\left(\frac{C_T(x(n))}{\ln 2}\right). \tag{3.14}$$

In the limit of $n \to \infty$ we have $I_T(x(n)) \le \ln(\#\mathcal{A}^n)$. The form of the right-hand side may be recognizable as the maximum possible n-block entropy of Shannon's definition. The natural logarithm implicitly gives the T-information the units of *nats*[c]. $I_T(x(n))$ is the total *T-information* for $x(n)$.

The *average T-information rate* per symbol, referred to here as the *average T-entropy* of $x(n)$ and denote by $\bar{h}_T(x(n))$ is defined simply:

$$\bar{h}_T(x(n)) = \frac{I_T(x(n))}{n}(\text{nats/symbol}). \tag{3.15}$$

Clearly we note that in the limit of $n \to \infty$, $\bar{h}_T(x(n)) \le \ln\#\mathcal{A}$. The correspondence that is superficially evident between the T-information and T-entropy on the one hand and Shannon's entropy definitions on the other hand, is reinforced in subsequent investigations.

Finishing off with the above example, we can compute the T-information and average T-entropy for the string from its T-complexity. Thus $I_T(1011010100010) \approx$

[c]The T-information may be converted to *bits* by dividing by $\ln 2$.

8.7 nats ($\approx$ 12.6 bits), and since $n = 13$, $\bar{h}_T(1011010100010) \approx 8.7/13 \approx 0.669$ (nats/bit), or 0.965 bits/bit.

The algorithm perhaps looks complicated, but is in fact recursive. Software implementations thus benefit from using standard recursion techniques making the algorithm easy to implement, and debug. This recursive approach to identifying pattern structures along the whole of a string, means that long range dependancies and structures in patterning are effectively detected. The T-entropy calculation is broadly sensitive both to the local and global pattern structures and yields entropy values that are indicative of the Shannon n-block entropy for large n, as will be demonstrated below.

4. The Logistic Map as a Testing Ground

To illustrate the given concepts we use the logistic map

$$x_{n+1} = f(x_n) = rx_n(1 - x_n)\,, \quad r \in [0, 4]\,. \tag{4.16}$$

The logostic map has been studied in detail by several authors [9], [19], which gives the possibility to compare the numerical estimations with known analytical results. In particular, expressions for the conditional entropy and KS-entropy are available for the period accumulation point $r_\infty = 3.5699\ldots$ and for fully developed chaos $r = 4.0$ [9], [11], [19]. As in all one-dimensional maps, a generating partition is defined by the coordinates of the critical points [6]. Here

$$S_n = \begin{cases} 0 & \text{if } x_n \in [0, c] \quad \text{and} \\ 1 & \text{if } x_n \in (c, 1] \end{cases} \tag{4.17}$$

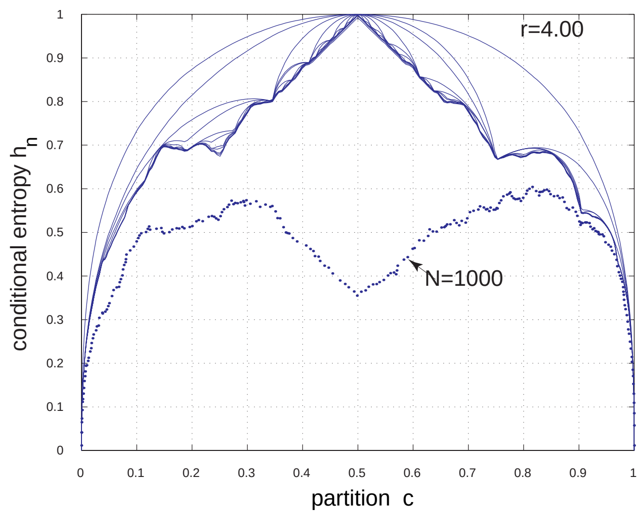
with $c = 0.5$. The convergence of the conditional entropies h_n to the positive Lyapunov exponent, hence to the Pesin entropy h_p, has been verified [7]. Of certain importance is that the higher order conditional entropies attain their maximal value for a partition threshold $c = 0.5$. While for the logistic map this coincides with the known generating partition and is thus expected, it could be used to establish a heuristic criterion for choosing a particular partition c. With respect to Eq. (2.7) we will call the partition threshold c^* for which the higher order conditional entropies $h_n(c)$ attain their maximal value an *optimal binary measurement partition*[d] [7], [29].

Figure 1 shows the conditional entropies h_n and the T-entropy of the logistic map as a function of the measurement partition c.

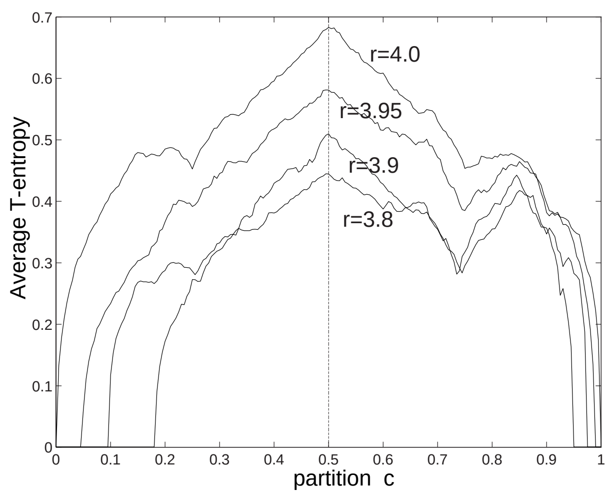
Fixing the partition to $c = 0.5$ we are now able to compare the estimated entropies to the previously defined Pesin entropy h_p. While for the conditional entropy h_n the convergence is somewhat expected (Fig. 2), also the T-entropy

[d] Of course this heuristic criterion does not meet Kolmogorov's criterion for the supremum over all partition. Still it provides a simple rule how the choice of a partition should be performed.

is successful in estimating the Pesin entropy, hence positive Lyapunov Exponent (Fig. 3). In particular, the deviation between the T-entropy values and corresponding positive Lyapunov exponent values over the corresponding parameter values $3.55 \leq r \leq 4$ value, is RMS 1.04% of full scale (ln 2). The T-entropy values are found to be dependent to a small degree on the length of the string. For a given source the



(a)



(b)

Fig. 1. (a) The conditional entropies h_1 to h_{10} (from above) for the logistic map at $r = 4.0$ as a function of the measurement partition c. The dotted line indicates the finite size effects for h_{10} and $N = 1000$. (b) The T-entropy as a function of the measurement partition c for different values of r. Both complexity estimates attain their maximal value at the critical point $c^* = 0.5$.

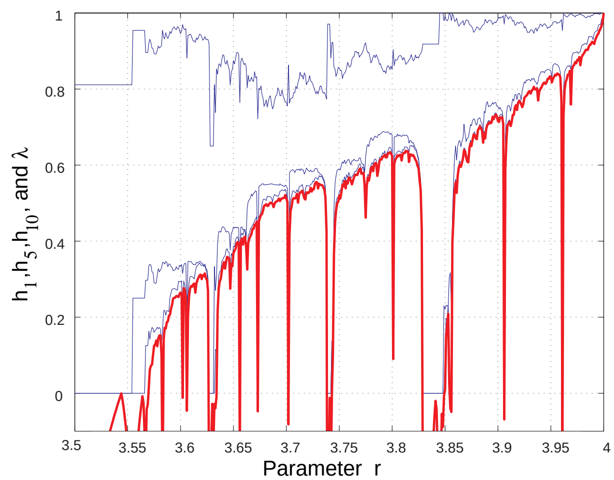


Fig. 2. The convergence of the conditional entropies h_n to the positive Lyapunov exponent: The figure shows the conditional entropies h_1, h_5 and h_{10} (from above) as a function of map parameter $r \in [3.5, 4]$. The measurement partition was $c^* = 0.5$ throughout. The thick line represent the Lyapunov exponent, hence the positive part is the Pesin entropy h_p.

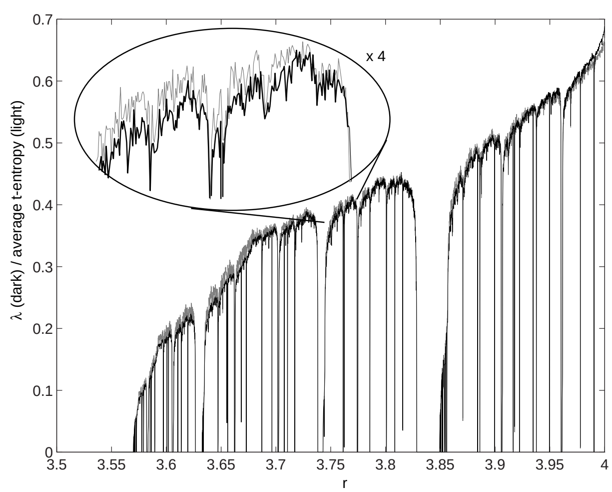


Fig. 3. The average T-entropy (the scale factor is $\ln 2$) and the Pesin entropy (dark line) for the logistic map as a function of $r \in [3.5, 4]$. The average T-entropy was computed from strings of $500,000$ bits long and sampled using the generating bi-partition.

computed T-entropy values approach the Lyapunov exponent asymptotically from above as a function of increasing string length N. This gives rise to a relationship $\bar{h}_T = kh_P$ where k is a scaling constant that needs to be determined for each string length. Once the sample string size has been selected, this scaling constant may determined rather easily using for example the known results of the logisitic map

$(r = 4)\lambda = \log(2)$ as a calibration point. This scaling factor then applies across the whole range of entropy values.

A similar comparison involving the Lyapunov exponent and the algorithmic redundancy was given by Wackerbauer *et al.* [34].

5. The Influence of Noise

5.1. *How noise affects the dynamics*

In the previous sections we have outlined the concept of symbolic dynamics and we have given criteria to perform the coarse-graining of a given dynamical system. However, any real physical system will always be subject to fluctuations. Thus it is essential to investigate this influence on quantities we wish to examine. As in the previous section we will only consider time-discrete maps as simple exemplary systems and will briefly review results for the Shannon entropies given by other authors. Before we proceed we distinguish between several, though not independent ways noise can affect the dynamics.

- *Observational noise*: In the simplest case we have a deterministic system, whose measured data is polluted by observational or measurement noise. That is, we get a time series $\tilde{x}_n = x_n + \varepsilon\xi_n$ instead of the underlying deterministic values x_n. At this stage we are not specific about the nature of the noise ξ_n (e.g. white, $1/f$-noise, Gaussian, ...).
- *Dynamical noise*: Another possibility would be, that the noise affects the dynamics of the system. This could be modeled by adding a noise term to the map (*additive dynamical noise*).

$$x_{n+1} = f^{\varepsilon}(x_n) = f(x_n) + \varepsilon\xi_n \,. \tag{5.18}$$

  Again we are not specific about the nature of the noise. Here special precaution must be paid to the fact that the noise can push the trajectory out of the basin of attraction. Still, additive dynamical noise is perhaps the most widely studied case.
- *Parametric noise*: A different approach is a fluctuating parameter in the map. In the case of the logistic map this could be $r^{\varepsilon} = r + \varepsilon\xi_n$. However, it is not obvious how this changes the invariant measure, given it exists at all. Analytical result for situation are given by Hubermann *et al.* [17].
- *Fluctuating partition borders*: In all previous cases the noise affects the dynamics prior to the procedure of coarse-graining. But one could also think of a fluctuating partition threshold $\tilde{c} = c + \varepsilon\xi_n$, when converting the trajectory into a sequence of symbols. Obviously this case closely resembles observational noise.
- *Noise-induced cell flips*: Already investigated by Shannon was the transmission of a sequence through a noisy channel, i.e. independent flips of symbols. In this case analytical results for this situation are available [12], [27].

Whichever of these different mechanisms of introducing noise is an appropriate model depends on the situation at hand. In empirically investigated physical systems a mixture of different noise effects is present and it is not possible to distinguish between them from the data only. Other challenges arise from the nature of the noise itself. In considering noise effects here we think of thermal-like fluctuations drawn from a stationary ensemble at each time step. However, this is subject of a recent debate [4], [13] and will not be addressed in this paper.

5.2. *Numerical results*

Before we discuss the numerical results of our investigation, we outline the effects of noise on our estimated quantities. If we consider a source of equidistributed uncorrelated random numbers (white noise) the joint probabilities in Eq. (2.2) factorize and we get $H_n = nH_1$, thus $h_n = H_1$. Here H_1 is determined by the partition, i.e. $H_1 = \log \#\mathcal{A}$ if all symbols occur with equal probability. Thus the KS-entropy is not bounded and will go to infinity as the alphabet size increases. The same is also true for the T-entropy.

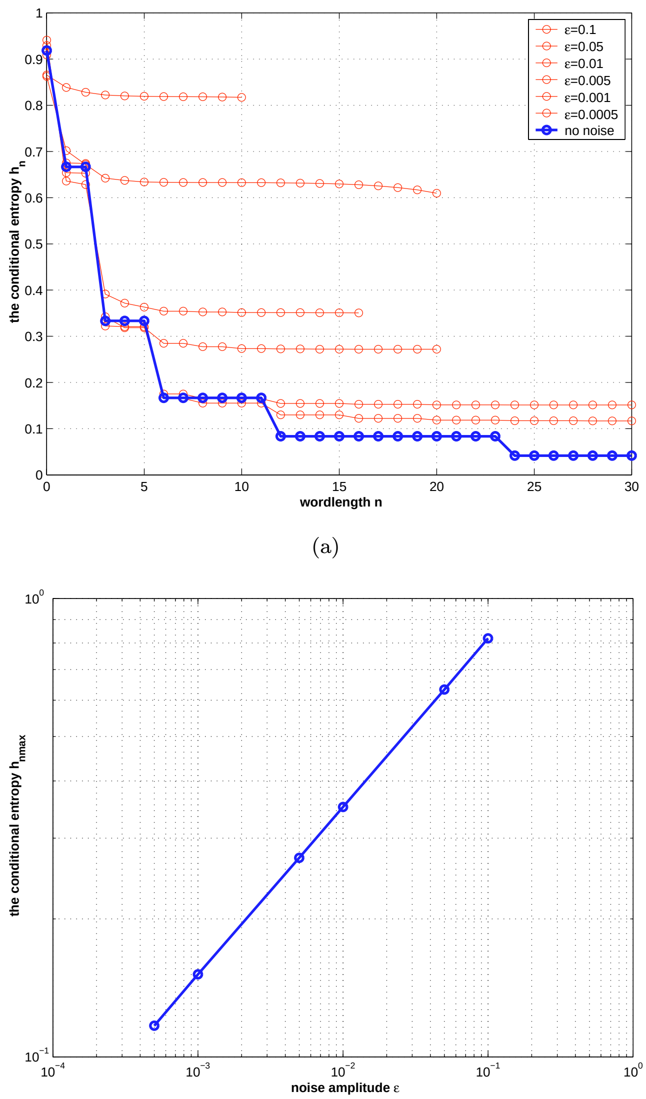
One approach to obtain meaningful results is to introduce a spatial resolution ε, corresponding to the size of the partition cells, and generalize the entropies to a function of both ε and n. This generalization of the KS-entropy to the (ε, τ)-entropy from deterministic chaotic processes to stochastic systems was introduced by Gaspard and Wang [13] and recently investigated by other authors [4], [18].

Here we follow a somewhat simpler approach. The number of partition cells, and hence the size of the alphabet, stays fixed and the partition is chosen to be an optimal measurement partition. To confirm earlier results [7], [12] we first look at the conditional entropies h_n^ε of the logistic map with additive dynamical noise for various noise amplitudes ε.

$$x_{n+1} = r_\infty x_n(1 - x_n) + \varepsilon\xi_n \quad \xi_n \in [-1, 1]\,. \tag{5.19}$$

Here ξ_n are equidistributed delta-correlated random numbers (white noise) and $r_\infty \approx 3.5699$ is the Feigenbaum accumulation point. The first thing to observe is that the conditional entropies h_n^ε remain constant beyond a certain value $n_{\max}$, which depends on the noise amplitude ε (*noise floor*). In other words, by observing additional symbols no additional information about the initial condition is gained, the self-refinement is destroyed on length scales smaller than of order ε. This corresponds to an $n_{\max}$th order Markov chain property (see Fig. 4). The limit entropies h_∞^ε, approximated by h_n^ε with $n > n_{\max}$ display a power law behaviour $h_\infty^\varepsilon \sim \varepsilon^\alpha$ [7], [12]. Similar results are found for the T-entropy, as could be observed in Fig. 5.

From an empirical standpoint one might also be interested in the predictability of a system, given a finite prehistory n. While the conditional entropy generally increases under the influence of noise, we find that for some systems there is a slight



(a)

(b)

Fig. 4. The logistic map with additive dynamical noise. (a) The conditional entropies h_n^ε as a function of wordlength n for various noise amplitudes ε. (b) The limit entropies h_∞ as a function of noise amplitude. h_∞ was approximated by h_n having reached a reliable plateau.

decrease for small noise amplitudes (see Fig. 6). This means the predictability could be enhanced when including a small amount of noise. Similar results for the Hénon map were reported by Tang and co-workers [30].

However in all previous examples, the partition was fixed to be the generating partition for the noise-free case $c = 0.5$. The decrease in entropy happens only

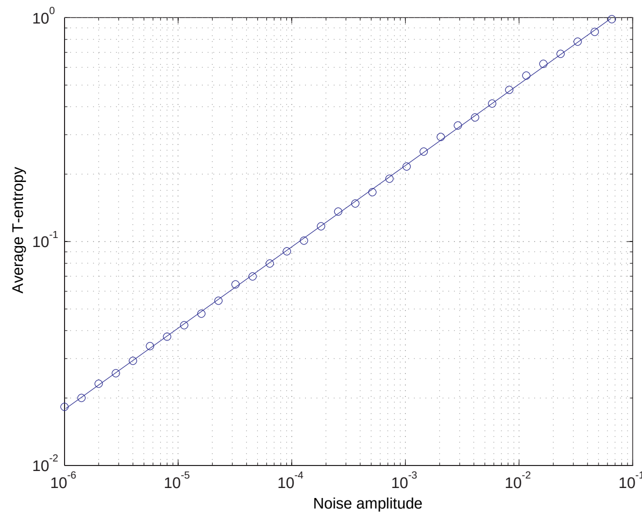
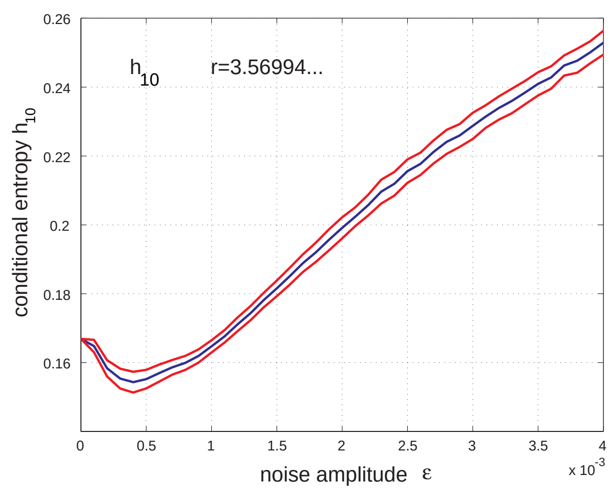


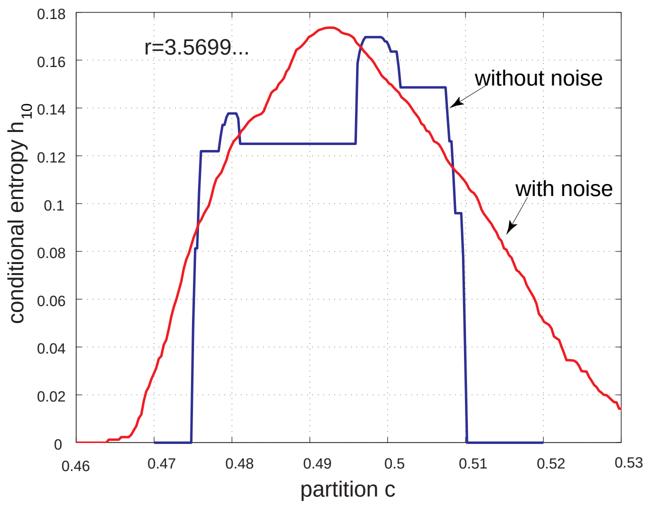
Fig. 5. The average T-entropy versus noise amplitude confirming a power law behavior over a wide dynamical range.



(a)

Fig. 6. The logistic map with additive dynamical noise. (a) The conditional entropy h_{10} as a function of noise amplitude ε for the fixed binary partition $c = 0.5$ (the upper and lower line indicating the standard deviation). (b) The conditional entropy h_{10} as a function of measurement partition in the vicinity of $c \approx 0.5$ in the no-noise case and for a fixed noise amplitude.

for certain partition thresholds, while for others the conditional entropy does increase with the influence of noise. This is illustrated in Fig. 6(b), which shows the conditional entropy h_{10} for the noise amplitude $\varepsilon \approx 0.0004$ in the vicinity of $c = 0.5$. Thus a statement about the influence of noise on the investigated measures may only



(b)

Fig. 6 (*Continued*)

be made with respect to the partition threshold. The more general conjecture that one could always find a *binary* measurement partition c, for which the conditional entropy does increase is not supported by the data.

6. Conclusions

This work was focused on a comparison of partition-based entropies such as Shannon, Kolmogorov–Sinai, and T-entropy and trajectory-based measures such as Lyapunov exponents and Pesin entropy. Partition-based measures may be calculated independently of the deterministic or stochastic nature of the system. Therefore the full exploration of the relation between partition-based and trajectory-based measures seems to offer an approach for extending the concepts developed for deterministic systems to stochastic systems. Here we used the logistic map across the full spectrum of parameter values, including regular and chaotic motions, and under the influence of noise as a setting for comparing the entropy measures.

Finally we would like to make a few remarks on the computational aspects. Entropy calculations from finite strings are possible with rather modest computational effort, the problems are due to finite-size effects. For the calculation of the Shannon entropy analytical estimates are available, as discussed in Sec. 2.2.1. For the T-entropy strict analytical estimates of finite-size errors are not yet available. Our computational experience led us to the conclusion that the T-entropy is less sensitive to finite-size effects than the Shannon entropy, this is also known for the LZ complexity. These problems need further clarification.

References

1. Eds. M. Abramowitz and I. A. Stegun, **Handbook of Mathematical Functions** (Dover, 1970).
2. L. Arnold, **Random Dynamical Systems** (Springer, 1998).
3. L. Arnold and P. Imkeller, *The Kramers oscillator revisited*, in: **Stochastic Processes in Physics, Chemistry and Biology** eds. J. Freund and T. Poeschel (Springer, 2000).
4. M. Cencini, M. Falconi, E. Olbrich, H. Kantz and A. Vulpiani, *Chaos or noise: Difficulties of a distinction*, *Phys. Rev.* **E62** (2000).
5. F. Christiansen and A. Politi, *A generating partition for the standard map*, *Phys. Rev.* **E51** (1995).
6. Collet and J. P. Eckmann, **Iterated Maps on the Interval as Dynamical Systems** (Birkäuser, 1980).
7. J. P. Crutchfield and N. H. Packard, *Symbolic dynamics of noisy chaos*. *Physica* **D7** (1983).
8. W. Ebeling and M. A. Jiménez-Montaño, *On grammars, complexity and information measures of biological macromolecules*, *Math. Biosci.* **52** (1980).
9. W. Ebeling and K. Rateitschak, Symbolic dynamics, entropy and complexity of the Feigenbaum map at the accumulation point, *Discrete Dynamics in Nature Soc.* **2** (1998) 187-194.
10. J. P. Eckmann and D. Ruelle, *Ergodic theory of chaos and strange attractors*, *Rev. Mod. Phys.* **57** (1985).
11. J. Freund, W. Ebeling and K. Rateitschak, *Self similar sequences and universal scaling of dynamical entropies*, *Phys. Rev.* **E54** (1996).
12. J. Freund and K. Rateitschak, *Entropy analysis of noise contaminated sequences*, *Internat. J. Bifurcation Chaos* **8** (1998) 933–940.
13. P. Gaspard and X.-J. Wang, *Noise, chaos, and* (ε, τ)*-entropy per unit time*, *Phys. Rep.* **235** (1993).
14. P. Grassberger, *Finite sample corrections to entropy and dimension estimates*, *Phys. Lett.* **A128** (1988).
15. P. Grassberger and H. Kantz, *Generating partitions for the dissipative Hénon map*, *Phys. Lett.* **A113** (1985).
16. H. Herzel, A. O. Schmitt, and W. Ebeling, *Finite sample effects in sequence analysis*, *Chaos, Solitons Fractals* **4** (1994).
17. B. A. Hubermann and J. Rudnick, *Scaling behaviour of chaotic flows*, *Phys. Rev. Lett.* **45** (1980).
18. H. Kantz and E. Olbrich, *Coarse grained dynamical entropies - investigations of high-entropic dynamical systems*, Elsevier preprint, (1999).
19. K. Karamanos and G. Nicolis, *Symbolic dynamics and entropy analysis of Feigenbaum limit sets*, *Chaos, Solitons & Fractals* **10** (1999) 1135–1150.
20. F. Kaspar and H. G. Schuster, *Easily calculable measure for the complexity of spatiotemporal patterns*, *Phys. Rev.* **A36** (1987).
21. A. N. Kolmogorov, *A new metric invariant of transitive dynamical systems and automorphisms in Lebesgue space*, *Dokl. Acad. Nauk. SSSR* **119** (1958).
22. A. Lempel and J. Ziv, *On the complexity of finite sequences*, *IEEE Trans. Inform. Theory* **22** (1976) 75–81.
23. R. Nicolescu and M. R. Titchener, *Uniqueness theorems for t-codes*, *Romainian J. Inform. Sci. Tech.* **1** (1998).
24. J. B. Pesin, *Characteristic Lyapunov exponents and smooth ergodic theory*, *Russ. Math. Surveys* **32** (1977) 355.

25. T. Pöschel, W. Ebeling and H. Rosé, *Guessing probability distributions from small samples*, *J. Statist. Phys.* **80** (1995).
26. H. G. Schuster, *Deterministic Chaos* (Weinheim VCH, 1989).
27. C. E. Shannon, *A mathematical theory of communication*, *The Bell System Tech. J.* **27** (1948).
28. W.-H. Steeb, *A Handbook of Terms Used in Chaos and Quantum Chaos* (BI Wissenschaftsverlag, 1991).
29. R. Steuer, L. Molgedey, W. Ebeling and M. A. Jiménez-Montaño, *Entropy and optimal partition for data analysis*, *EPJ* (submitted).
30. X. Z. Tang, E. R. Tracy, A. D. Boozer, A. deBrauw and R. Brown, *Symbol sequence statistics in noisy signal reconstruction*, *Phys. Rev.* **E51** (1995).
31. M. R. Titchener, *Deterministic computation of string complexity, information and entropy*, in **Internat. Symp. on Information Theory**, August 16–21, 1998, MIT, Boston.
32. M. R. Titchener, *A deterministic theory of complexity, information and entropy*, in **IEEE Information Theory Workshop**, February 1998, San Diego.
33. M. R. Titchener, *A measure of information*, in **IEEE Data Compression Conference**, March 2000, Snowbird.
34. R. Wackerbauer, A. Witt, H. Atmanspacher, J. Kurths and H. Scheingraber, *Quantification of structural and dynamical complexity*, *Chaos, Solitons & Fractals* **4** (1994) 133.