Entropies and Predictability of Nonlinear Processes and Time Series

Werner Ebeling

Saratov State University, Faculty of Physics, Saratov, Russia, werner_ebeling@web.de, home page: www.ebelinge.de

Abstract. We analyze complex model processes and time series with respect to their predictability. The basic idea is that the detection of local order and of intermediate or long-range correlations is the main chance to make predictions about complex processes. The main methods used here are discretization, Zipf analysis and Shannon's conditional entropies. The higher order conditional Shannon entropies and local conditional entropies are calculated for model processes (Fibonacci, Feigenbaum) and for time series (Dow Jones). The results are used for the identification of local maxima of predictability.

1 Introduction

Our everyday experience with the prediction of complex processes is showing us that predictions may be done only with certain probability. Based on our knowledge on the present state and on certain history of the process we make predictions, sometimes we succeed and in other cases the predictions are wrong [1]. Considering a mechanical process, we need only some knowledge about the initial state. The character of the dynamics, regular or chaotic, and the precision of the measurement of the initial states decide about the horizon of predictability. For most complex systems, say e.g. meteorological or financial processes, we have at best a few general ideas about their predictability.

The problem we would like to discuss here is, in which cases our chances to predict future states are good and in which cases they are rather bad. Our basic tool to analyze these questions are the conditional entropies introduced by Shannon and used by many workers [2-6]. By using the methods of symbolic dynamics any trajectory of a dynamic system is first mapped to a string of letters on certain alphabet [2, 4, 5]. This string of letters is analyzed then by Shannon's information-theoretical methods.

2 Conditional Entropies

This section is devoted to the introduction of several basic terms stemming from information theory which were mostly used already by Shannon. Let us assume that the processes to be studied are mapped to trajectories on discrete state spaces (sequences of letters) with the total length L. Let λ be the length of the alphabet. Further let $A_1A_2...A_n$ be the letters of a given subtrajectory of length $n \leq L$. Let further $p^{(n)}(A_1...A_n)$ be the probability to find in the total trajectory a block (subtrajectory) with the letters $A_1...A_n$. Then according to Shannon the entropy per block of length n is:

$$H_n = -\sum p^{(n)}(A_1 \dots A_n) \log p^{(n)}(A_1 \dots A_n)$$
(1)

From this we derive conditional entropies as $h_n = (H_{n+1} - H_n) \leq \log(\lambda)$ The limit of the dynamic n-gram entropies for large n is the entropy of the source h (called also dynamic entropy or Kolmogorov - Sinai entropy). Further we define

$$r_n = \log(\lambda) - h_n \tag{2}$$

as the average predictability of the state following after a measured n-trajectory. We remember that $log(\lambda)$ is the maximum of the uncertainty, so the predictability is defined as the difference between the maximal and the actual uncertainty. In other words, predictability is the information we get by exploration of the next state in the future in comparison to the available knowledge. In the following we shall use λ as the unit of the logarithms.

The predictability of processes is closely connected with the dynamic entropies [7]. Let us consider now certain section of length n of the trajectory, a time series, or another sequence of symbols $A_1 \ldots A_n$, which often is denoted as a subcylinder. We are interested in the uncertainty of the predictions of the state following after this particular subtrajectory of length n. Following again the concepts of Shannon we define the expression

$$h_n^{(1)}(A_1...A_n) = -\sum p(A_{n+1}|A_1...A_n) \log p(A_{n+1}|A_1...A_n)$$
(3)

as the conditional uncertainty of the next state (1 step into the future) following behind the measured trajectory $A_1
dots A_n$. Further we define

$$r_n^{(1)}(A_1 \dots A_n) = 1 - h_n^{(1)}(A_1 \dots A_n)$$
(4)

as the predictability of the next state following after a measured subtrajectory, which is a quantity between zero and one. We note that the average of the local uncertainty leads us back to Shannon's conditional entropy h_n . The predictability may be improved by taking into account longer blocks. In other words, one can gain advantage for predictions by basing the predictions not only on actual states but on whole trajectory blocks which represent the actual state and its history.

3 The conditonal entropy for model processes and time series

The first mathematical model of a nonlinear process was formulated in 1202 by the Italian mathematician Leonardo da Pisa, better known as Fibonacci, in his book Liber Abaci. Fibonacci considered the problem how many rabbit pairs are generated after n breeding sessions assuming the following simple rules:

- the game starts with an immature pair,
- rabbits mature in one season after birth,
- mature rabbit pairs produce one new pair every breeding session,

- rabbits never die. This game generates the famous sequence of Fibonacci numbers 1, 1, 2, 3, 5, 8, 13, 21, 34, 55, The Fibonacci model may be encoded as a sequence of zeroths and ones by using the rules $0 \rightarrow 1$ denoting "young rabbits grow old" and $1 \rightarrow 10$ standing for "old rabbits stay old and beget young ones". Beginning with a single 0, continued iteration gives 1, 10, 101, 10110, etc., resulting finally in the infinite selfsimilar Fibonacci sequence 1011010110110 ... Alternatively we may formulate the rules by a grammar:

$$S_0 = 0 \tag{5}$$

$$S_1 = 1 \tag{6}$$

$$S_{n+1} = S_n S_{n-1} \qquad (n = 1, 2, 3, \ldots)$$
(7)

The conditional entropy of the Fibonacci sequence is exactly known [8]. These entropies behave in the limit of large n as

$$h_n = \frac{C}{n} \tag{8}$$

Another well-studied simple model of a nonlinear process is the logistic map:

$$x(n+1) = rx(n)(1 - x(n))$$
(9)

In order to generate a discrete string from this map we use the bipartition ($\lambda = 2$)

$$C_1 := [0, \frac{1}{2}) \longrightarrow 0 \tag{10}$$

$$C_2 := [\frac{1}{2}, 1] \longrightarrow 1 \tag{11}$$

This way the states are mapped on the symbols 0 and 1 and the process is mapped on binary strings. We denote these sequences as Feigenbaum sequences. The rank-ordered word distributions for Feigenbaum strings were discussed by several authors [2,9,11]. For r = 4 all the words of fixed length are equally distributed and the entropy is maximal $h_n = 1$. For the Feigenbaum accumulation point $r_a = 3.5699456...$ we get also a simple word distributions consisting only of one or two steps in dependence on the word length [2,9,11]. The construction rules for the generation of these sequences generate selfsimilar structures. Accordingly the n-gram block entropies satisfy the relations

$$H_{n_{k+1}} = H_{n_k} + 1 = H_{n_1} + k \tag{12}$$

By specialization of eqs.(12) we get for $n = 2^k$ a result first obtained in 1986 by Grassberger [7]:

$$H_n = \log_2(3n/2) \tag{13}$$

In a similar way we obtain the H_n for all the other sequences [2,9]. For the conditional entropies we get for the Grassberger numbers n = 2, 4, 8, 16, ... the conditional entropies

$$h_n = \frac{4}{3n} \tag{14}$$

In between two Grassberger numbers namely at n = 3, 6, 12, 24, ... the dynamic entropies jump to the value according to the next Grassberger number. In this way a simple step function is obtained. We see that the dynamic entropy itself (which is the limit of infinite n) is zero. For infinite histories the predictability is 1 i.e. 100%. This corresponds to a zero Lyapunov exponent $\lambda = 0$ [7,11]. In the region r > 3.5699... the Lyapunov exponent is in general larger zero, corresponding to chaotic states. Then the Pesin theorem $h = \lambda$ may be used in order to obtain a lower border for the conditonal entropies [5]. The convergence to the limit is rather fast. This may be exploited also for the investigation of optimal partitions [5]. Further we may use the knowledge of the lower order entropies and of the limit for the construction of Pade approximations.

We mention that similar long range correlations are also generated by intermittent processes [11]. A special group of discrete intermittent maps of such type was investigated by Szepfaluzy and coworkers [10]. The following scaling for the approach to the limit was found

$$h_n - h = \frac{1}{n^{\alpha}} . \tag{15}$$

The processes considered so far, correspond to the limiting case of processes which are predictable on the basis of a long observation. This property is lost, if noise is added which leads always to an upper limit of the predictability $r_{max} < 1$ [11,9].

There exist several types of noise, the simplest is the perturbation by a noisy channel. Assuming that the percentage ϵ of the symbols is subject to random flips, the source entropy as a function of ϵ may be estimated by

$$h(\epsilon) = h(0) - \epsilon \log_{\lambda}(\epsilon) - (1 - \epsilon) \log_{\lambda}(1 - \epsilon)$$
(16)

Real data behave similar to maps with noise. By combination of a solvable map with local structures on any scale, say the Feigenbaum map, with channel noise (measurement noise) and by adaptation of the parameter ϵ a wide variety of shapes of the entropy function h_n may be represented.

Let us present now an application of these concepts to the analysis of real strong noisy data [4]. Prediction of strong noisy data using classical linear methods usually fails to give accurate predictions and a reliable confidence level. The concept of entropy and local predictability in combination with classical methods is a good candidate to give reliable results. Applications of these concepts to meteorological strings were given in [12, 13] and to nerve signals in [5].

In the following our concept will be demonstrated on daily stock index data S_t : Dow Jones 1900-1999 (27044 trading days). Since the stock index itself has an exponentially growing trend we will use daily logarithmic index changes

$$x_t = \ln(S_t) - \ln(S_{t-1}) \quad . \tag{17}$$

An application of the entropy concept requires a partitioning of the real value data x_t into symbols A_t of an alphabet. To find an optimal partition and alphabet is a process of maximizing the Kolmogorov-Sinai entropy [5]. However for strong noisy signals with short memory an equal frequency of the letters is near to optimal. To be concrete we used $\lambda = 3$ and $A_t = 0$; $x_t < -0.0025$ (strong decrease in the stock value), $A_t = 2$; $x_t > 0.0034$ (strong increase), $A_t = 1$ (intermediate) were chosen [4]. In Fig. 1 the result of calculations of the conditional entropy is presented [4].



Fig. 1. Conditional entropy $h_n = H_{n+1} - H_n$ as a function of word length n; the strong decrease for $n \ge 5$ is an articlated due to length effects.

We see that the average predictability is rather small. For $n \ge 4-5$ the error is growing due to length effects [14]. The further decay seems to be an artefact, the true entropy probably remains constant for $n \ge 5$. Therefore the average uncertainty of the daily stock index is very high und the average predictability is less than 5%.

4 Predictions Based on a Local Analysis

Sometimes the analysis of the average entropies fails to detect existing correlations. On the other hand the average uncertainty of predictions is in many cases (e.g. for the stock market as shown above) higher than 0.9 (i.e. higher than 1.8 bits). Therefore the average predictability is rather low. For practical applications, one is not so much interested in an average value but even more in a concrete prediction based on the the observation of a concrete string of finite length n. In other words one is more interested in concrete predictions than in "average predictabilities".

Therefore we have studied also the predictabilities of the states following right after the particular strings $A_1 \ldots A_n$ which we denoted by $r_n^{(1)}(A_1 \ldots A_n)$

This is a quantity depends on the local "history" $A_1 \ldots A_n$ and fluctuates therfore while going along the string. Another closely related fluctuation quantity is the transinformation, which is connected with the local predictability.

For the Fibonacci sequences as well as for the Feigenbaum sequences the local regularities follow from the grammar rules. Sometimes the next letter is nearly predetermined. Let us give just one example. In the Fibonacci sequence as well as in the Feigenbaum sequence the subsequence 00 is forbidden by the grammatical rules. Therefore in the state 0 the predictability of the next state is 1, after the symbol 0 comes the symbol 1 with 100% certainty. The rule that 00 is forbidden, creates a special local order.

In the following the existence of local regularities will be demonstrated on the daily stock index data S_t discussed above.

The result of the calculation of the local uncertainty $h_n(A_1, \ldots, A_n)$ for the next trading day following behind an observation of n trading days A_1, \ldots, A_n for n = 5 is plotted in Fig. 2. The local uncertainty is almost near one, i.e. the average predictability is very small. However behind certain patterns of stock movements A_1, \ldots, A_n the local predictability reaches 8% – a notable value for the stock market, which in average is near to random. The mean predictability over the full data set is less then 2% (see Fig.1).

The question of the significance of the prediction is treated by calculating a distribution of local uncertainty $h_n^S(A_1, \ldots, A_n)$ by help of surrogates. We constructed surrogate sequences having the same two point probabilities as the original sequence [4]. The level of significance K was calculated as

$$K_n(A_1,\ldots,A_n) = \frac{h_n(A_1,\ldots,A_n) - \langle h_n^S(A_1,\ldots,A_n) \rangle}{\sigma} \quad , \tag{18}$$

where $\langle h_n^S(A_1, \ldots, A_n) \rangle$ is the mean and σ is the standard deviation of the local uncertainty distribution for the word A_1, \ldots, A_n .

Assuming Gaussian statistics $|K| \leq 2$ represents confidence greater then 95%. However since the local uncertainty distribution is more exponential like larger K-values are required to guarantee significance. For the analyzed data set a word length up to 6 seems to give still reliable results. In Fig. 2 we represented the uncertainty of the state subsequent to six observed states as a function of time, the interval corresponds to the last months of 1987 [4]. The greyvalue codes the level of significance calculated from a first order Markov surrogate. Dark represents a large deviation from the noise level (good significance).

It is remarkable that higher local predictabilities coincide with larger levels of significance. This can be seen also from Table 1.

Since we used a timeseries over a very long period we have to address the problem of non-stationary by dividing the original timeseries into smaller pieces.



Fig. 2. Local uncertainty of the the daily Dow Jones index (in symbolic representation) which follows after an observation of 5 subsequent days. We represented an interval corresponding to the second half of 1987.

Furthermore instead of producing surrogates on the level of Symbols one can discuss surrogates obtained by modells of a stockmarkets like ARCH/GARCH-modells. This has been done in [4].

Analyzing the data in Table 1 we see, that in spite of the fact that the average predictability is very low (about 2%) there are special days, where the predictability is up to 8%, i.e. up to 4 times higher than in avaerage. We remember that in our way of coding 0 stands for a day with a strong downswing of the index, 2 stands for a strong upswing and 1 stands for a day where the index remains nearly constant. Remarkable is, that the highest predictability correponds to the days following the October-Crash in 1987.

As a result of these investigations we may state that in spite of the fact that the stock market index index is in average very uncertain, some local order might be detected which is helpful for predictions. Similar results were obtained for meteorological data and for nerve signals [5, 12, 13].

5 Conclusions

Our results show that the dynamic entropies are an appropriate measure for studying the predictability of complex processes. Of particular interest are local studies of the predictabilities after certain local histories. Local minima of the uncertainty may be found in many processes including even the index of the stock market. The basic problem for improving predictions is the detection of middle range and of long range correlations. These correlations are of specific interest since they improve the predictability. If long range correlations exist, one can improve the resultes by basing the predictions at longer observations.

word	predic.	K	word	predic.	K	word	predic.	Κ
020	0.029	-27.9	1112	0.046	-14.3	11110	0.081	-9.5
112	0.029	-30.5	0000	0.043	-12.0	11120	0.074	-9.1
110	0.023	-23.4	1110	0.042	-15.0	20000	0.074	-6.9
120	0.019	-29.4	0110	0.040	-14.9	11112	0.069	-8.7
000	0.018	-18.5	0020	0.039	-12.5	10120	0.067	-5.2
212	0.017	-19.5	1102	0.038	-12.4	22202	0.067	-9.4
202	0.016	-22.9	2020	0.034	-10.6	00000	0.066	-6.4
111	0.015	-20.1	0200	0.032	-10.6	11011	0.063	-9.7
121	0.015	-12.4	0202	0.031	-14.6	02000	0.061	-5.7
012	0.013	-14.2	0120	0.029	-12.3	02020	0.059	-5.1
102	0.012	-10.5	2112	0.029	-9.0	00020	0.057	-6.4

Table 1. Sequences of 3-5 daily indices of the Dow Jones with the highest predictabilityof the following (nextday) index.

Further we have found that there are specific local substrings, where the uncertainty is much smaller than the average, i.e. the predictability is better than in average. In other words, even for the case of noisy data, there are specific situations where local predictions are possible, since the local predictability is much better than the average predictability. It may be of practical importance to find out all substrings which belong to this particular class. Our results clearly demonstrate that the best chance for predictions is based on the observation of ordered local structures. The entropy-like measures studied here operate on the sentence and the word level. In some sense entropies are the most complete quantitative measures of correlation relations. This is due to the fact that the entropies include also many point-correlations. On the other hand the calculation of the higher order entropies is extremely difficult and at the present moment there is no hope to extend the entropy analysis to the level of hundreds of letters. In conclusion we may say that a more careful study of the correlations in time series sequences of mediate and long range may contribute to better predictions of complex processes.

The author thanks J. Freund, L. Molgedey, T. Pöschel, K. Rateitschak, and R. Steuer for many fruitful discussions and a collaboration on special topics of the problems discussed here.

References

- 1. Feistel, R., Ebeling, W.: Evolution of Complex Systems, Kluwer Academic Publ., Dordrecht 1989.
- Ebeling, W., Nicolis, W.: Word frequency and entropy of symbolic sequences: a dynamical perspective, Solitons & Fractals 2 (1992) 635-640.
- 3. Ebeling, W.: Prediction and entropy of sequences with LRO. Physica D 109 (1997) 42-50.

- 4. L. Molgedey, W. Ebeling: Local order, entropy and predictability of financial time series, Eur. Phys. J B 15 (2000) 733-737; Physica A 287 (2000) 420-427.
- 5. R. Steuer. L. Molgedey, W. Ebeling, M.A. Jimenez-Montano: Entropy and optimal partition for data analysis, Eur. Phys. J. B **19** (2001) 265-269.
- Ebeling, W., Steuer, R., Titchener, M.R.: Partition-based entropies of deterministic and stochastic maps. Stochastics and Dynamics 1 (2001) 45-61.
- 7. Grassberger, P.: Entropy and complexity. Int. J. Theor. Phys. 25 (1986) 907-915.
- 8. T. Gramss, T.: Entropy of Fibonacci sequences. Phys. Rev. E 50 (1994) 2616-2620.
- Ebeling, W., Rateitschak, K.: Symbolic dynamics, entropy and complexity of the Feigenbaum map at the accumulation point. Discrete Dyn. in Nat. & Soc.2 (1998) 187-194.
- Szepfaluzy, P., Györgyi, G.: Entropy of nonlinear maps. Phys. Rev. A 33 (1986) 2852-2860.
- Freund, J., Ebeling, W., Rateitschak, K.: Self similar sequences and universal scaling of dynamical entropies, Phys. Rev. E 54 (1996); Int. J. Bifurc. & Chaos 6 (1996) 611-620.
- Nicolis, C., Ebeling, W., Baraldi, C.: Markov processes, dynamical entropies and the statistical prediction of mesoscale wheather regimes. Tellus 49 A (1997) 108-118.
- Werner, P.C., Gerstengarbe, F.-W., Ebeling, W.: Changes in the probability of sequences, exit time distribution and dynamical entropy in the Potsdam temperature record, Theor. Appl. Climatol. 62 (1999) 125-132.
- Pöschel, T., Ebeling, W., Rosé, H.: Guessing probability distributions from small samples, J. Stat. Phys 80 (1995) 1443-1452.