Theory: Complexity-Entropy analysis

June 3, 2021

1 Probability Distribution

Consider a weakly stationary time series of length L, denoted as $\{x_{t=1,...,L}\}$. The time series is then partitioned into chunks of size d, where d is the so-called embedding dimension. For each partition of the time series the size ordering of the d consecutive values is denoted as a permutation π. The permutation of all $L–d+1$ chunks of the time series are investigated, and the occurrence of each permutation is counted. The number of occurrences of each permutation is then divided by the number of partitions of the time series, $L-d+1$, so that the relative frequency of occurrence of each permutation is obtained. The total number of unique permutations for a given embedding dimension is $d!$. Denoting each permutation with a π_i, the relative frequency of each permutation, $(\pi_1, \pi_2, \ldots, \pi_{d!})$, is what is denoted as Bandt-Pompe probability distribution. This is summarized with the following formula [1]:

$$p(\pi) = \frac{\#[t|t \leq L-d+1, (x_{t+1}, ..., x_{t+d}) \text{ is of type } \pi]}{L-d+1} \tag{1}$$

Lets illustrate the algorithm with the following example.

Let the time series be the following sequence of numbers of length $L = 10$

$$x_t = [6, 2, 8, 10, 5, 4, 1, 3, 7, 9]$$

Also let the embedding dimension be $d = 3$.

Given the embedding dimension, the time series is then partitioned into chunks of size 3, also the total number of permutations the partitioned sequence can take is $d! = 3! = 6$.

Investigating the first partitioned chunk, $[6, 2, 8]$. The size ordering of this chunk is $x_1 < x_0 < x_2$, which means that the corresponding permutation is $\pi_3 = 102$. Next chunk is $[2, 8, 10]$. The entries in this sequence happens to be in increasing order, $x_0 < x_1 < x_2$, and the corresponding permutation $\pi_1 = 012$. The next chunk is $[8, 10, 5]$ which has the size ordering $x_2 < x_0 < x_1$ and thus corresponds to the permutation $\pi_5 = 201$. This is done to the remaining partitions of the time series. Counting the occurrences of all permutation and dividing by the total number of partitions, $L–d + 1 = 10–3 + 1 = 8$, gives the relative frequency of the permutations.

The number of occurrence and the relative frequency of all possible permutations are summarized in the following table:
To obtain the proper probability distribution for the permutations of the system the length of the time series must tend towards infinity [1]. However, it is assumed that the time series is long enough to get a decent estimation of the probability distribution. The embedding dimension cannot be chosen arbitrarily, and some thought must go into choosing the appropriate embedding dimension for the situation at hand [2].Also, there are practical limitations to the embedding dimension chosen. For the Bandt-Pompe algorithm to achieve accurate results the length of the time series must be significantly larger than the total number of permutations the system can access, that is $L \gg d!$ [1]. For embedding dimension $d = 6$ there are a total of $d! = 720$ possible permutations for amplitude orderings, while for embedding dimension $d = 10$ there are 3628800 possible permutations. The length of the time series is still beholden to the length requirement, $L \gg 3628800$. Meaning as

Table 1: List giving all the possible permutations, and the relative frequency of occurrence.

π	Number of occurrence	$p(\pi)$
$\pi_1 = 012$	3	3/8
$\pi_2 = 021$	0	0
$\pi_3 = 102$	1	1/8
$\pi_4 = 120$	1	1/8
$\pi_5 = 201$	1	1/8
$\pi_6 = 210$	2	2/8

the embedding dimension increases, the length requirement for the time series to obtain accurate results quickly becomes unreasonable. Consequently, Bandt and Pompe recommended that the embedding dimension to be limited to the range $d \in [3, 7]$ [1].

Papers give the common condition for the length of the time series in question to be $L > 5d!$ [3, 4]. As will be demonstrated in section **??**, a time series of length $L = 5d!$ should be viewed as a minimum requirement for the length. Using a white noise process it will be demonstrated that a time series of length $L < 5d!$ comes nowhere close to the theoretical point it should have in the Complexity-Entropy plane.

2 Permutation Entropy

Using the probability distribution obtained from the Bandt-Pompe algorithm, the entropy measure of the time series can be calculated. The entropy measure of choice is Shannon's logarithmic information entropy [1, 5, 2, 6, 7, 8]. Shannon's information entropy is given as:

$$S[P] = -\sum_{i=1}^{d!} p(\pi_i) \log_2 p(\pi_i) \tag{2}$$

This sum runs over all $d!$ permutations for the given embedding dimension d.

Lets investigate the conditions required to maximize and minimize the entropy measure to gain some intuition. The entropy measure is minimized for a monotone time series. For a monotonically increasing or decreasing time series only one amplitude ordering permutation is accessed by the system with probability $p(\pi) = 1$. This makes the Shannon entropy:

$$S[P] = -\sum_{i=1}^{d!} p(\pi_i) \log_2 p(\pi_i) = -1 \cdot \log_2 1 = 0$$

The maximal value of the entropy measure is for a time series where all possible permutations occur with equal probability. This happens for a completely random time series from an independent, identically distributed sequence [1]. The probability that the system access one state is $p(\pi) = p = \frac{1}{d!}$.

The Shannon entropy is then:

$$S[P] = -\sum_{i=1}^{d!} p(\pi_i) \log_2 p(\pi_i) = -d!(p \log_2 p) = -d! \frac{1}{d!} \log_2 \frac{1}{d!} = -\log_2 \frac{1}{d!} = \log_2 d!$$

The permutation entropy is then defined as the Shannon information entropy normalized to its maximum value:

$$H[P] = \frac{S[P]}{\log_2 d!} \tag{3}$$

The discussion above makes it clear that $0 \leq H[P] \leq 1$.

From this, the entropy measure can be considered as a measure of the randomness of the time series [1, 7]. Ordered time series will have low entropy. As the randomness of the signal increases, so does the entropy until it is maximized for a completely random, uncorrelated time series. But the entropy measure only captures the randomness of the physical system the time series is generated from and this measure cannot determine if there are correlation structures present in the physical system [7]. To aid in capturing these structures a statistical complexity measure is used. The measure considered here is the Jensen-Shannon complexity measure.

3 Jensen-Shannon Complexity

The statistical complexity measure should capture the relation between the components of the physical system which influence the probability distribution from the system [7]. The desired behavior for the chosen statistical complexity measure should be small for high degree of order, and for high degree of randomness. For the extreme cases, being a perfectly ordered system and completely random system, the complexity measure should vanish as these types of systems do not possess any structure to speak of [7]. In between these extremes the complexity measure should reflect the possible degrees of physical structure from the underlying probability distribution [7].

In the paper by Martin et al. [7] a family of statistical complexity measures was presented, one of them was the Jensen-Shannon complexity measure. These statistical complexity measures were defined to, in a way, represent a "distance" between the probability distribution obtained from the system to the uniform probability distribution. In general, these complexity measures have the functional form [7]:

$$C[P] = Q[P]H[P] \tag{4}$$

Here $H[P]$ is a measure of entropy and $Q[P]$ represents the "distance" to, or "disequilibrium" from the uniform distribution. $Q[P]$ is given as:

$$Q[P] = Q_0 D[P, P_e] \tag{5}$$

Q_0 is a normalization factor such that $0 \leq Q[P] \leq 1$, and $D[P, P_e]$ is the "distance" from the probability distribution P to the uniform distribution P_e [7].

The disequilibrium $Q[P]$ will be different from zero only if there exist "privileged states", states which the system is more likely to be in. If all states are accessed by the system with equal probability, then there are no privileged states, and the system probability distribution is uniform. Therefore, in this case the disequilibrium measure is zero. Which in turns mean that the complexity measure is zero. For the opposite case where the system only access one state, the disequilibrium measure from the uniform distribution is maximized, meaning that $Q[P] = 1$, but the complexity measure should still be zero in this case. Notice that the complexity measure is the product of the disequilibrium measure $Q[P]$ and an entropy measure $H[P]$ (equation (4)). In the case where the system only access one state, the system is completely ordered meaning that the entropy measure is zero which results in the complexity measure also being zero. Thus, the complexity measure is compatible with its desired behavior described earlier and vanishes in the trivial cases.

For the following Complexity-Entropy analysis the Jensen-Shannon complexity measure is chosen. As mentioned in the paper by Martin et al. [7] and Lamberti et al. [8], the Jensen-Shannon complexity measure is an intensive quantity. Because of its intensive property, the Jensen-Shannon complexity measure is the best choice as a statistical complexity measure for systems of physical origins and hence is the reason the Jensen-Shannon complexity measure was chosen.

The Jensen-Shannon complexity measure is defined as

$$C_{JS}[P] = Q_J[P, P_e]H[P] \tag{6}$$

Here, P denotes the probability distribution obtained from the system in question using the Bandt-Pompe algorithm. P_e is the uniform distribution. $H[P]$ is the permutation entropy defined by equation (3). $Q_J[P, P_e]$ is the normalized Jensen-Shannon divergence, or disequilibrium measure, defined as

$$Q_J[P, P_e] = Q_0 \left[S\left[\frac{P + P_e}{2}\right] - \frac{S[P]}{2} - \frac{S[P_e]}{2} \right] \tag{7}$$

$P + P_e$ denotes the addition between the probability distribution obtained from the system and the uniform distribution. $S[\cdot]$ is the Shannon entropy measure defined by equation (2). The normalization constant Q_0 is given as:

$$Q_0 = -\frac{2}{\frac{d!+1}{d!}\log_2(d!+1) - 2\log_2(2d!) + \log_2(d!)} \tag{8}$$

The complexity measure is not a trivial function of entropy [7]. For a value of entropy in the range $H \in (0, 1)$ the complexity measure can take a range of values between a minimum and a maximum value. Martin et al. [7] proved with the use of Lagrangian multipliers that one can define probability distributions that maximizes and minimizes the complexity measure in the range $H \in (0, 1)$, essentially defining two curves that gives the upper and lower bound for the complexity measure. These probability distributions were nicely summarized in two tables in the paper by Zhu et al. [2], and are presented in table 2 nad 3 below. Notice that the probability distributions which defines the maximum and minimum complexity lines only depend on the embedding dimension d.

Table 2: The probability distributions that minimizes the complexity measure, taken from table 1 in [2].

Number of states p_i	p_i	range p_i
1	p_{min}	$\left[\frac{1}{d!}, 1\right]$
$d! - 1$	$\frac{1-p_{\text{min}}}{d!-1}$	$\left[0, \frac{1}{d!}\right]$

The probability distribution that defines the minimum complexity line starts of as a uniform distribution where all entries have the same value $p_{min} = \frac{1}{d!}$. Then the probability that the system will access one state increases until the state will be accessed with probability $p_{min} = 1$. The probability for the rest of the $d! - 1$ states decreases accordingly such that $\sum p_i = 1$. For small enough step-size between possible p_{min} values, a smooth curve for minimum complexity is achieved.

Table 3: The probability distributions that maximizes the complexity measure, taken from table 2 in [2]. The range for the number n is $0 \leq n \leq (d! - 1)$

Number of states p_i	p_i	range p_i
n	0	0
1	p_{max}	$\left[0, \frac{1}{d!-n}\right]$
$d! - n - 1$	$\frac{1-p_{\text{max}}}{d!-n-1}$	$\left[\frac{1}{d!-n-1}\right]$

The maximum complexity line is defined by $d! - 1$ different probability distributions. n denotes the number of states in a probability distribution that are zero and increases from 0 states to $d! - 1$ states. For each case of n, one state varies in probability of occurring from 0 to $\frac{1}{d!-n}$. The rest of the states change accordingly such that $\sum p_i = 1$. As is visible in figure 1, the maximum complexity line is not a smooth function [2].

Figure 1 shows how the maximum and minimum complexity line changes with the embedding dimension d. For lower values of d numerical effects are visible.

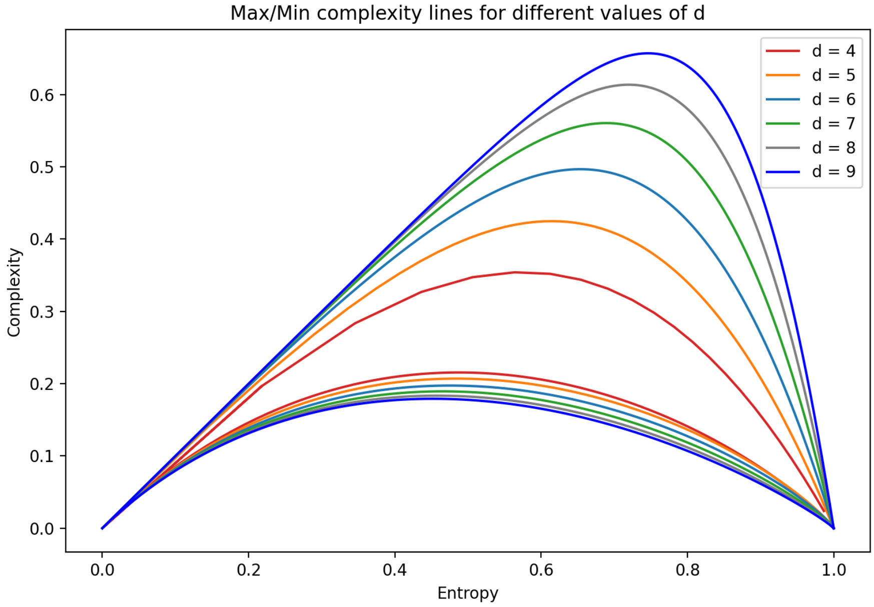

Figure 1: The maximum and minimum complexity lines for different embedding dimensions as function of entropy.

4 Complexity-Entropy Location of Selected Time Series

4.1 Linear mode and white noise

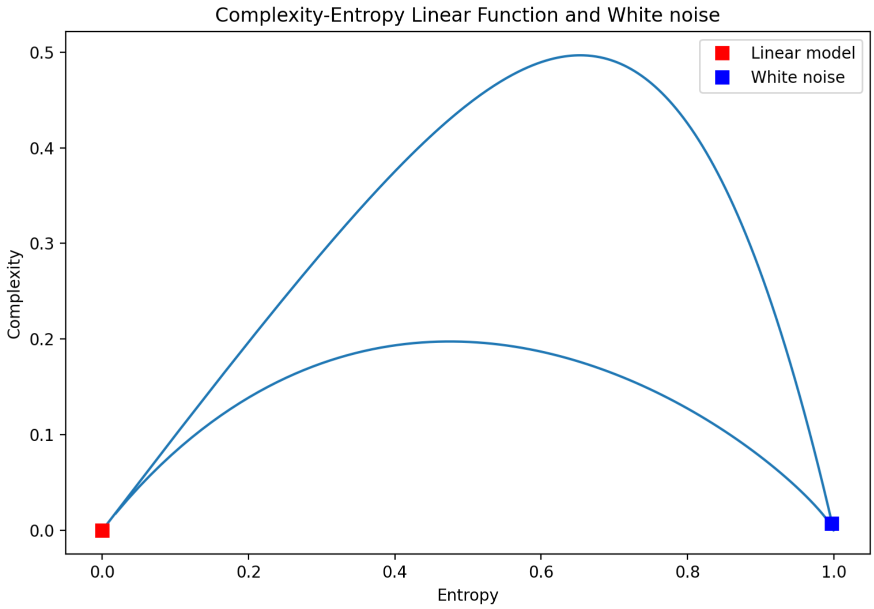

Figure 2: Complexity-Entropy plot of the linear model and the white noise process. Both models appear in their theoretical locations.

4.2 Sine-function

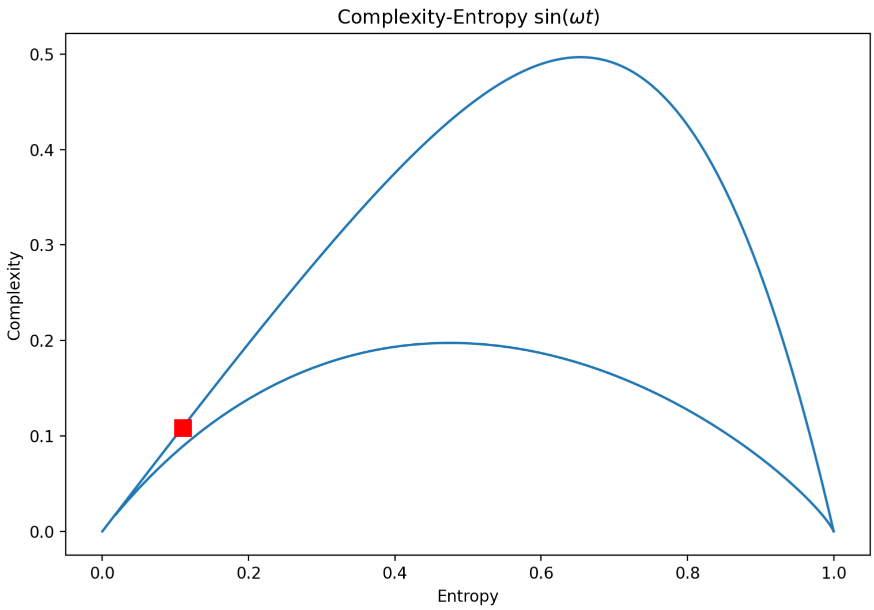

Figure 3: Complexity-Entropy plot of the sine function.

4.3 Logistic map

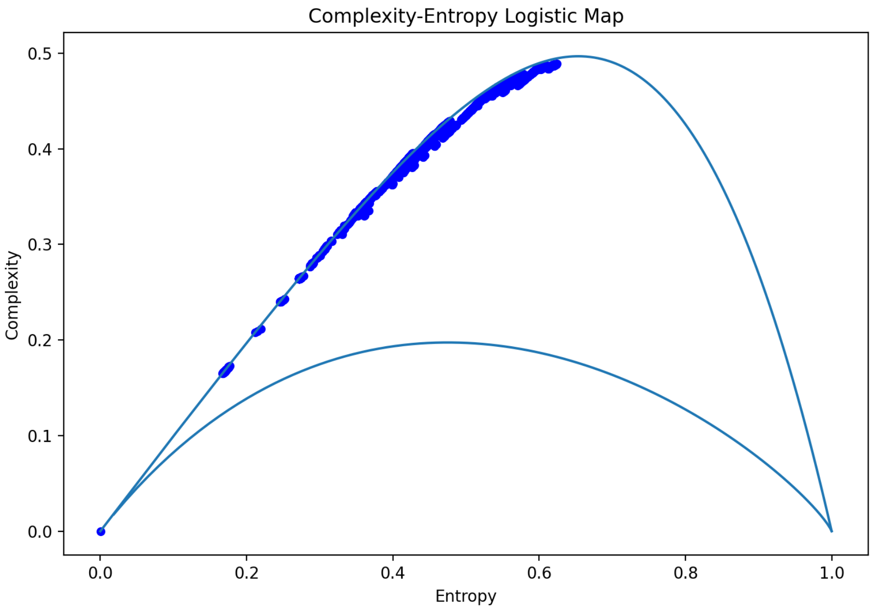

Figure 4: Complexity-Entropy plot for Logistic map realizations with growth parameter in the range $3.5 \leq r \leq 4$

References

[1] Christoph Bandt and Bernd Pompe. Permutation entropy: A natural complexity measure for time series. *Phys. Rev. Lett.*, 88:174102, Apr 2002. doi:10.1103/PhysRevLett.88.174102.

[2] Z. Zhu, A. E. White, T. A. Carter, S. G. Baek, and J. L. Terry. Chaotic edge density fluctuations in the alcator c-mod tokamak. *Physics of Plasmas*, 24(4):042301, 2017. doi:10.1063/1.4978784.

[3] P. J. Weck, D. A. Schaffner, M. R. Brown, and R. T. Wicks. Permutation entropy and statistical complexity analysis of turbulence in laboratory plasmas and the solar wind. *Phys. Rev. E*, 91:023101, Feb 2015. doi:10.1103/PhysRevE.91.023101.

[4] M. Riedl, A. Müller, and N. Wessel. Practical considerations of permutation entropy. *The European Physical Journal Special Topics*, 222(2):249–262, Jun 2013. doi:10.1140/epjst/e2013-01862-7.

[5] O. A. Rosso, H. A. Larrondo, M. T. Martin, A. Plastino, and M. A. Fuentes. Distinguishing noise from chaos. *Phys. Rev. Lett.*, 99:154102, Oct 2007. doi:10.1103/PhysRevLett.99.154102.

[6] J E Maggs and G J Morales. Permutation entropy analysis of temperature fluctuations from a basic electron heat transport experiment. *Plasma Physics and Controlled Fusion*, 55(8):085015, jun 2013. doi:10.1088/0741-3335/55/8/085015.

[7] M.T. Martin, A. Plastino, and O.A. Rosso. Generalized statistical complexity measures: Geometrical and analytical properties. *Physica A: Statistical Mechanics and its Applications*, 369(2):439 – 462, 2006. doi:https://doi.org/10.1016/j.physa.2005.11.053.

[8] P.W Lamberti, M.T Martin, A Plastino, and O.A Rosso. Intensive entropic non-triviality measure. *Physica A: Statistical Mechanics and its Applications*, 334(1):119–131, 2004. doi:https://doi.org/10.1016/j.physa.2003.11.005.