Complex Systems Methods — 4. Statistical complexity II

Eckehard Olbrich

e.olbrich@gmx.de
http://personal-homepages.mis.mpg.de/olbrich/complex_systems.html

Potsdam WS 2007/08

Overview

1. Summary: Excess entropy of temporal sequences
2. Examples
   - The logistic map
   - The henon map
   - Cellular Automata
   - Cellular Automata
3. Generalization of the excess entropy for systems without a linear ordering
   - Measuring statistical dependencies: Integration
   - Tononi, Sporns and Edelman's (TSE) neural complexity
4. How should statistical complexity scale with system size?
   - Adding one element
   - Adding a subsystem
   - Adding an independent subsystem
   - Adding a copy

Summary: Excess entropy of stationary temporal sequences

- Entropy of a subsequence $H_n = H(X_1^n) = H(X_1, \ldots, X_n)$ is splitted into a unpredictable, "random" part nh_∞ and a part representing the regularities in the sequence $H_n \approx E + n \cdot h_\infty$. This remaining **non-extensive** part was the excess entropy

$$E = \lim_{n\to\infty} E_n = \lim_{n\to\infty} (H_n - n \cdot h_n) \qquad h_n = H(X_0 | X_{-n}^{-1}) \ .$$

- The excess entropy is related to the rate of convergence of the conditional entropies

$$E = \sum_{k=1}^{\infty} k \delta h_k \quad \delta h_n := h_{n-1} - h_n = MI(X_0 : X_{-n} | X_{-n+1}^{-1}) \ .$$

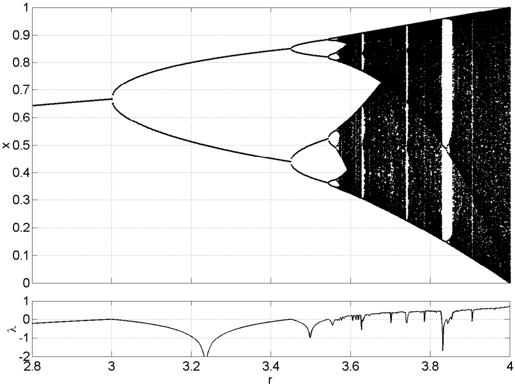
- If the excess entropy is finite it is equal to the "predicitive information"

$$I_{pred} = MI(X_{-\infty}^0 : X_1^\infty) = E \ .$$

- The excess entropy is a lower bound for the amount of information needed for an optimal prediction.

Examples: The logistic map

The logistic map: $x_{n+1} = rx_n(1 - x_n)$. Invariant measure:



Entropy rate equal to the Lyapunov exponent, $h_\infty = \lambda$, if $\lambda \geq 0$.

Excess entropy

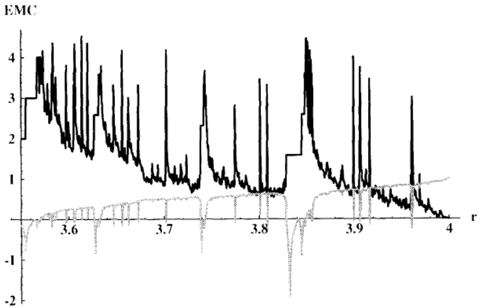


Fig. 1. Plot of EMC estimated to l = 16th order, versus r-values in [3.55, 4.00] of the logistic map. Also included in the plot we have the Lyapunov exponent λ as a function of r, calculated with logarithms of base two (gray line). Therefore one can identify the Feigenbaum-points, where the EMC diverges. Values of the EMC for high-entropy data sets (mostly $r > 3.85$) are calculated up to l = 12th order only, owing to limitations of computer memory.

R. Günther, B. Schapiro, P. Wagner, Complex Systems, Complexity Measures, Grammars and Model-Inferring. Chaos, Solitons & Fractals, 4(1994),p. 635-651.

Excess entropy and ϵ-machine complexity

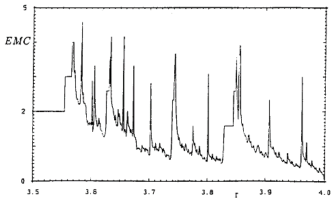


Fig. 10. Effective measure complexity calculated for P_{16}^{G*}. EMC vanishes for $r = 4$ and is given by $\log p$ for periodic behavior.

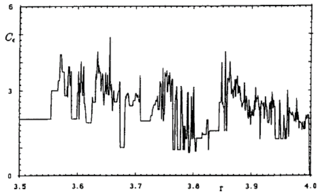


Fig. 12. ϵ-complexity calculated for P_8^{G*} and for the parameter combination $l_2 = 1, 2, \cdots, 8$; $l_1 = 2l_2$; $\epsilon = 0.02, 0.04, \cdots, 0.2$. The small complexity values in the parameter range [3.80, 3.83] are caused by the relative small lengths (l_1, l_2) of trees and subtrees.

R. Wackerbauer, A. Witt, H. Atmanspacher, J. Kurths, H. Scheingraber, A Comparative Classification of Complexity Measures.

Chaos, Solitons & Fractals, 4(1994),p. 133-173.

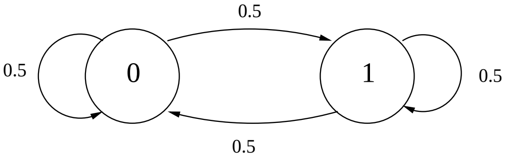
ϵ-machine reconstruction - Fully developed chaos

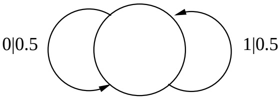
r=4: Fully developed chaos.

- Generating partition:

$$s_n = \begin{cases} 1 \text{ for } x_n \geq 0.5 \\ 0 \text{ for } x_n < 0.5 \end{cases}$$

- Symbolic dynamics: Bernoulli shift — random sequence of 0's and 1's with $p(0) = p(1) = p(1|0) = p(0|1) = 1/2$.





ϵ-machine reconstruction

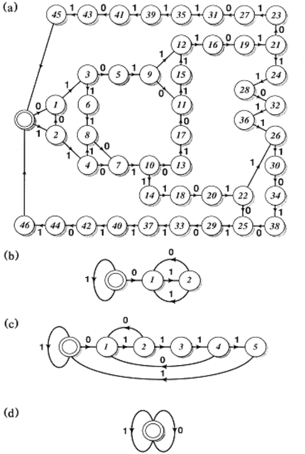


FIG. 1. Topological ε-machine l-digraphs for the logistic map at (a) the first period-doubling accumulation $r_c = 3.569\,945\,671\ldots$, (b) the band merging $r_{2B\to 1B} = 3.678\,59\ldots$, (c) the "typical" chaotic value $r=37$, and (d) the most chaotic value $r=4$. (a) and (c) show approximations of infinite ε-machines. The start vertex is indicated by a double circle; all states are accepting; otherwise, see Ref. 11.

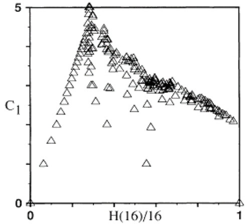


FIG. 2. Graph complexity C_1 vs specific entropy $H_1(16)/16$, using the binary, generating partition {[0,0.5),[0.5,1]}, for the logistic map at 193 parameter values $r \in [3,4]$ associated with various period-doubling cascades. For most, the underlying tree was constructed from 32-cylinders and machines from 16-cylinders. From high-entropy data sets smaller cylinders were used as determined by storage. Note the phase transition (divergence) at $H^* \approx 0.28$. Below H^* behavior is periodic and $C_a = H_a = \log(\text{period})$. Above H^*, the data are chaotic. The lower bound $C_a = \log(B)$ is attained at $B \to B/2$ band mergings.

J.P.Crutchfield, K. Young, Inferring Statistical Complexity,

PRL 63(1989), 105.

ϵ-machine at the period-doubling onset of chaos

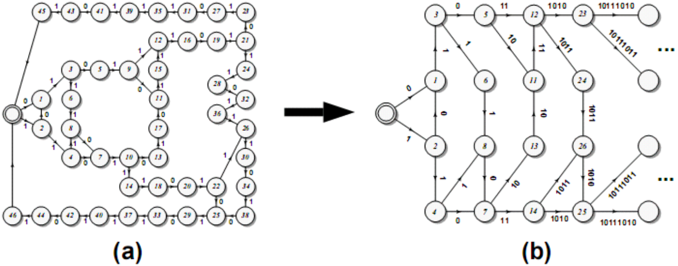


Figure 7 (a) Approximation of the critical ϵ-machine at the period-doubling onset of chaos. (After [24].) (b) The dedecorated version of the machine in (a). Here the deterministic state chains have been replaced by their equivalent strings. (After [56].)

J.P.Crutchfield, The Calculi of Emergence: Computation, Dynamics and Induction, Physica D 75(1994), 11-54.

The henon map

Extension of the logistic map to make it invertible:

$$\begin{aligned} x_{n+1} &= 1 - a x_n^2 + b y_n \\ y_{n+1} &= x_n \end{aligned}$$

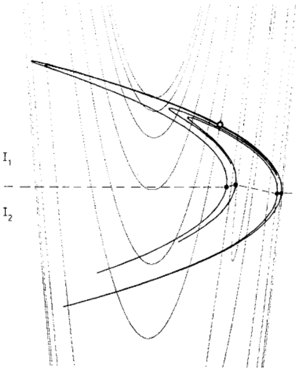


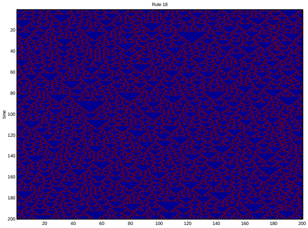
Fig. 2. Hénon map for $a = 1.4$, $b = 0.3$. The heavy lines are the unstable manifold and coincide with the attractor. The open dot is an unstable fixed point, and the light lines are part of its stable manifold. The heavy dots are the "principal" tangent points between stable and unstable manifolds. The dashed line connecting them forms the division line between the two sets of the partition.

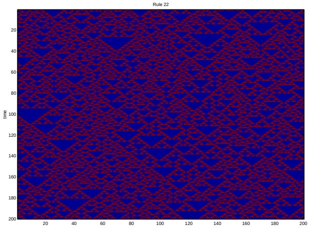
- Kantz/Grassberger (1985): $h_n - h_\infty \propto e^{-\gamma n}$ with $\gamma \approx 0.19$ for the standard parameter $a = 1.4$ and $b = 0.3$.

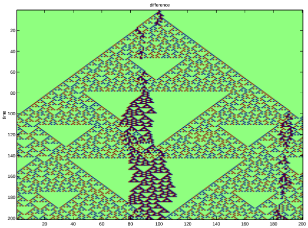
Cellular Automata — Rules 18 and 22

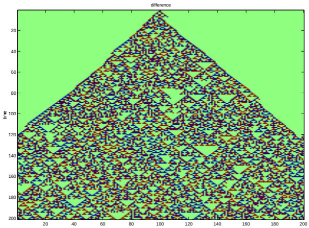
111	110	101	100	011	010	001	000
0	0	0	1	0	0	1	0

111	110	101	100	011	010	001
0	0	0	1	0	1	1

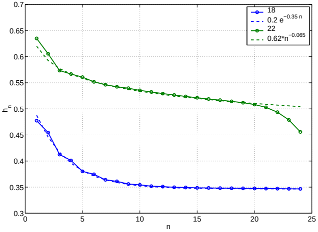


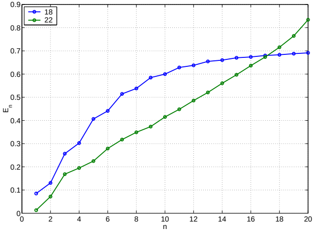






Cellular Automata — Rules 18 and 22





Rule 18: conditional entropies h_n decay exponentially, $h_\infty > 0$
$\Rightarrow$ "Chaotic"

Rule 22: h_n decay with a power law — Grassberger (1986): temporal decay $\propto n^{-0.06}$, spatial decay $\propto n^{-0.18}$, $h_\infty \to 0$
$\Rightarrow$ "Complex"

Generalization of the excess entropy — the setting

In the case of temporal sequences we considered unpredictability from the past as “randomness”. For a general joint probability distribution of n random variables with no particular ordering, there is a priori no restriction, which variables should be used to “predict” the others. Thus we consider in the following as randomness the sum of the remaining uncertainties of each random variable assuming the knowledge of all others.

- “World”: a set V of $1 \leq N < \infty$ elements (agents, nodes) with state sets $\mathcal{X}_v$, $v \in V$.
- Given a finite subset $A \subseteq V$ we write $\mathcal{X}_A$ instead of $\times_{v \in A} \mathcal{X}_v$
- Given a probability vector p on $\mathcal{X}_V$ we get random variables X_A.

Integration or Multi-information

The *integration* is a generalization of the mutual information for more than two random variables. The integration of the system X_V with respect to its nodes is defined as

$$I(X_V) := \sum_{v \in V} H(X_{\{v\}}) - H(X_V) = D\left(p(x_V) || \prod_{v \in V} p_v(x_{\{v\}})\right)$$

It is the difference between the sum of the variety of the elements and the variety of the system as a whole.
It becomes zero if and only if the probability distribution p has the product structure

$$p(x_V) = \prod_{v \in V} p_v(x_{\{v\}}) \,,$$

where each $p_{\{v\}}$ denotes the marginal distribution of the $X_{\{v\}}$.

Does the integration measure complexity?

$$I(X_V) := \sum_{v \in V} H(X_{\{v\}}) - H(X_V)$$

Pro:

- The integration vanishes in the case of complete randomness, because of the indepedence of the elements:

$$H(X_V) = \sum_{v \in V} H(X_{\{v\}})$$

- $I(X_V)$ vanishes in the case of complete determinism, given by a distribution that is concentrated in one configuration:

$$H(X_V) = H(X_{\{v\}}) = 0$$

Does the integration measure complexity?

$$I(X_V) := \sum_{v \in V} H(X_{\{v\}}) - H(X_V)$$

Contra:

- The integration maximal in the case of "synchronization"

$$H(X_V) = H(X_{\{v\}}) \qquad \Rightarrow \qquad I(X_V) = (N-1)H(X_{\{v\}}) ,$$

i.e. in an "ordered" state.

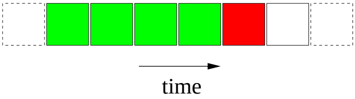
Does the integration measure complexity?

$$I(X_V) := \sum_{v \in V} H(X_{\{v\}}) - H(X_V)$$

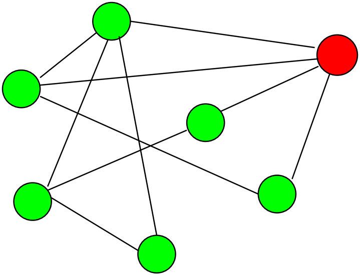
Conclusion:

- Multi-information measures all statistical dependencies in a system, also redundant ones.
- No general measure of statistical complexity, but might be used in special cases.

Excess entropy for a system without a linear order



time series



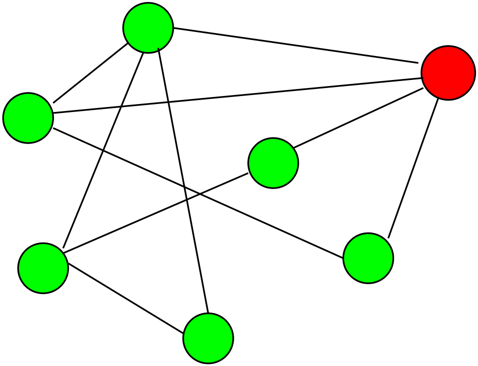
finite system

- $E'_N = \sum_{i=1}^{N}(h_i - h_{N-1})$ depends on the order of the elements, because of h_{N-1} defines the "randomness".
- Order independent possibility

$$E(X_V) := H(X_V) - \sum_{v \in V} H(X_{\{v\}} | X_{V \setminus \{v\}}) \, .$$

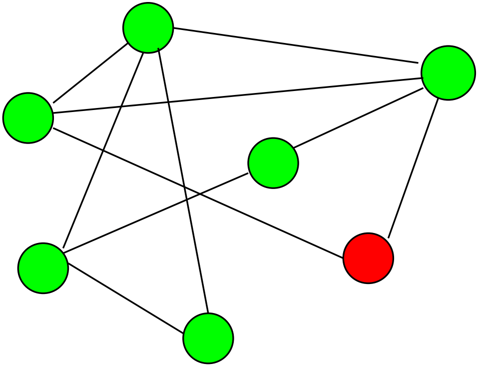
Excess entropy - Randomness

- Entropy rate replaced by $H(X_{\{v\}}|X_{V\setminus\{v\}})$
- $H(X_{\{v\}}|X_{V\setminus\{v\}})$ quantifies the amount to which the state of a single element cannot be explained by dependencies in the system and is therefore considered as random.



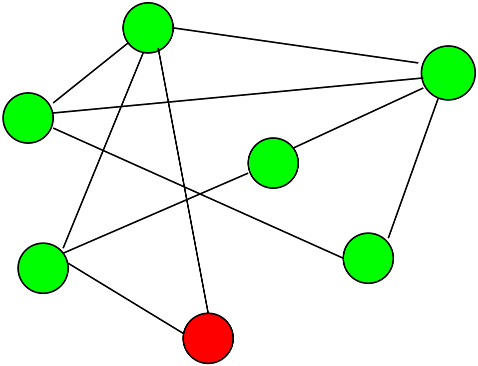
Excess entropy - Randomness

- Entropy rate replaced by $H(X_{\{v\}}|X_{V\setminus\{v\}})$
- $H(X_{\{v\}}|X_{V\setminus\{v\}})$ quantifies the amount to which the state of a single element cannot be explained by dependencies in the system and is therefore considered as random.

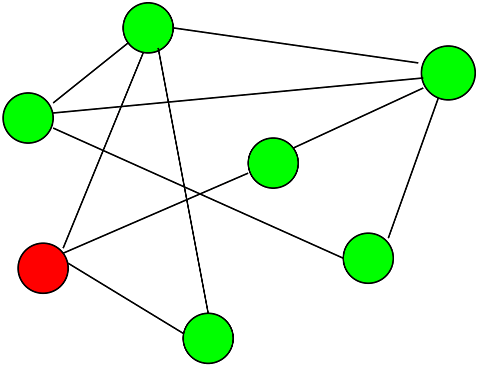


Excess entropy - Randomness

- Entropy rate replaced by $H(X_{\{v\}}|X_{V\setminus\{v\}})$
- $H(X_{\{v\}}|X_{V\setminus\{v\}})$ quantifies the amount to which the state of a single element cannot be explained by dependencies in the system and is therefore considered as random.

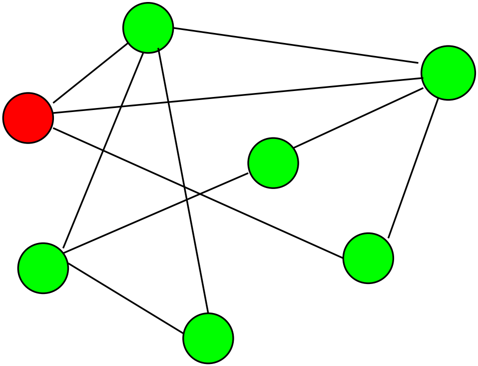


- Entropy rate replaced by $H(X_{\{v\}}|X_{V\setminus\{v\}})$
- $H(X_{\{v\}}|X_{V\setminus\{v\}})$ quantifies the amount to which the state of a single element cannot be explained by dependencies in the system and is therefore considered as random.



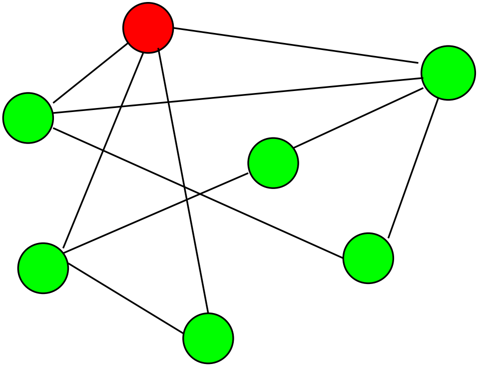
Excess entropy - Randomness

- Entropy rate replaced by $H(X_{\{v\}}|X_{V\setminus\{v\}})$
- $H(X_{\{v\}}|X_{V\setminus\{v\}})$ quantifies the amount to which the state of a single element cannot be explained by dependencies in the system and is therefore considered as random.



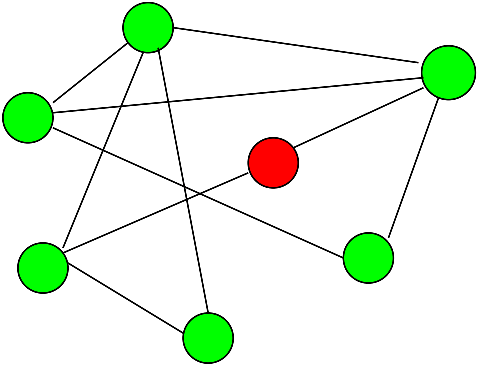
Excess entropy - Randomness

- Entropy rate replaced by $H(X_{\{v\}}|X_{V\setminus\{v\}})$
- $H(X_{\{v\}}|X_{V\setminus\{v\}})$ quantifies the amount to which the state of a single element cannot be explained by dependencies in the system and is therefore considered as random.



Excess entropy - Randomness

- Entropy rate replaced by $H(X_{\{v\}}|X_{V\setminus\{v\}})$
- $H(X_{\{v\}}|X_{V\setminus\{v\}})$ quantifies the amount to which the state of a single element cannot be explained by dependencies in the system and is therefore considered as random.



Excess entropy - Randomness

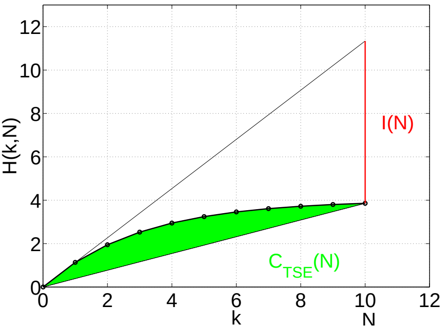
- Entropy rate replaced by $H(X_{\{v\}}|X_{V\setminus\{v\}})$
- $H(X_{\{v\}}|X_{V\setminus\{v\}})$ quantifies the amount to which the state of a single element cannot be explained by dependencies in the system and is therefore considered as random.
- The excess entropy is then the difference between the uncertainty of the state of the whole system and the sum of the unreducible uncertainties of the state of the elements using **all** information available in the system

$$E(X_V) := H(X_V) - \sum_{v \in V} H(X_{\{v\}}|X_{V\setminus\{v\}}) \,.$$

- It quantifies the "explainable" part of the variety of the system.

Tononi, Sporns and Edelmans (TSE) neural complexity

- Introduced as “neural complexity” by Tononi, Sporns and Edelman (1994)
- Should optimize between high entropy and high integration
- Integration formally: $I(X_V) := \sum_{v \in V} H(X_v) - H(X_V)$

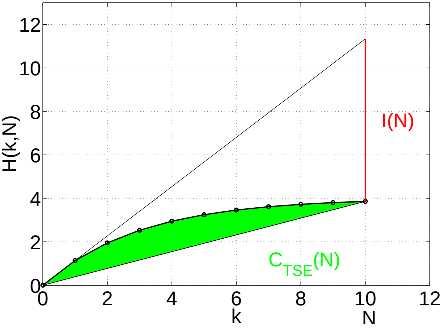


$$C_{TSE}(X_V): = \sum_{k=1}^{N} \left(H(k,N) - \frac{k}{N} H(N) \right)$$

$$H(k,N) = \binom{N}{k}^{-1} \sum_{\substack{Y \subseteq V \\ |Y|=k}} H(X_Y)$$

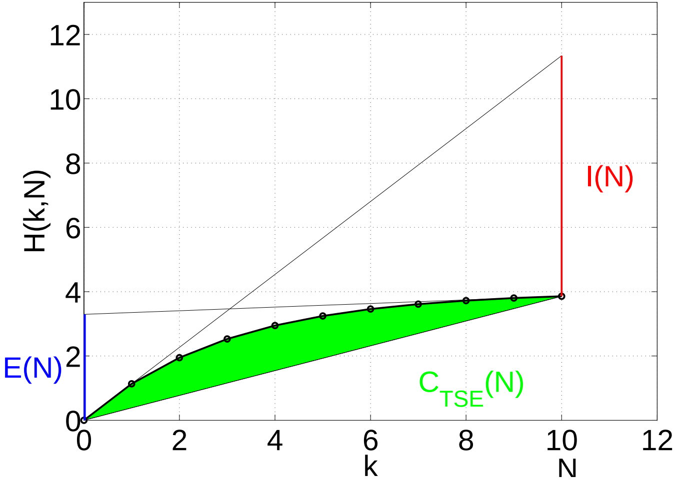
Tononi, Sporns and Edelmans (TSE) neural complexity

- Introduced as "neural complexity" by Tononi, Sporns and Edelman (1994)
- Should optimize between high entropy and high integration
- Integration formally: $I(X_V) := \sum_{v \in V} H(X_v) - H(X_V)$



- Attempt to measure the potential ability of a neural system to produce consciousness based on the intuition that a neural state corresponding conscious experience should contain a large amount of information, i.e. high entropy, which, however, is strongly integrated, i.e. also a large multi-information on the system level.

Excess entropy and TSE-complexity



Excess entropy and TSE-complexity

- Excess entropy relates the system level to the level of elements, the TSE-complexity considers also the level in between. Excess entropy depends on the system level entropy and its growth at the last step from $N-1$ to N, the TSE is governed by the growing behaviour at all levels.
- Even if the Excess entropy remains constant for growing k, the TSE-complexity will grow extensively.
- It can be shown that (exercise)

$$C_{TSE}(X_V) = \frac{1}{2}\sum_{k=1}^{N} E(k,N) = \frac{1}{2}\sum_{Y\subseteq V}\frac{1}{\binom{N}{|Y|}} E(X_Y)\ ,$$

i.e. the TSE-complexity is proportional to the sum over the mean excess entropies averaged over all subsets of the same size.

How should statistical complexity scale with system size?

Three special cases:

A) Adding an independent element: The element has no structure itself, so it has no own complexity. Because it is independent on the rest of the system the complexity should **not change**.

B) Adding an independent subsystem: Because there are no dependencies between the two systems the complexity of the union should be simply the **sum of the complexities** of the subsystems.

C) Adding an identical copy: Because there is no need for additional information to describe the second system one could argue that the complexity should be **equal to the complexity of one system**. One has, however, to include the fact in the description that the second system is a copy of the first one. At least this part should be not extensive with respect to the system size.

Excess entropy: Adding one element v'

$$E(X_{V\cup\{v'\}}) = E(X_V) + \sum_{v\in V} MI(X_{\{v\}} : X_{\{v'\}} | X_{V\setminus\{v\}})$$

- excess entropy increases monotonically with system size
- if there are **no statistical dependencies** between X_V and $X_{\{v'\}}$ then

$$\begin{aligned} MI(X_V : X_{\{v'\}}) &= MI(X_{\{v'\}} : X_{V\setminus\{v\}}) + MI(X_{\{v\}} : X_{\{v'\}} | X_{V\setminus\{v\}}) \\ &= 0 \\ \Rightarrow \quad & MI(X_{\{v\}} : X_{\{v'\}} | X_{V\setminus\{v\}}) = 0 \end{aligned}$$

- Thus $E(X_{V\cup\{v'\}}) = E(X_V)$ for an independent v'.

Adding a subsystem

1. The excess entropy of a system consisting of two subsystems A and B is always larger than the mutual information between these two subsystems:
$$E(X_{A\cup B}) \geq I(X_A : X_B)\ .$$
2. The excess entropy of the union of two subsystems is always larger than the excess entropy of one of the subsystems.
$$E(X_{A\cup B}) \geq E(X_A) \qquad E(X_{A\cup B}) \geq E(X_B)$$
3. In general the sum of the excess entropies of the subsystems is neither a lower nor an upper bound for the excess entropy of the whole system.
$$\begin{aligned} E(X_{A\cup B}) \quad &= \quad E(X_A) + E(X_B) + \sum_{v\in A} I(X_{\{v\}} : X_B | X_{A\setminus\{v\}}) + \\ &\quad + \sum_{v\in X_B} I(X_{\{v\}} : X_A | X_{B\setminus\{v\}}) - I(X_A : X_B)\ . \end{aligned}$$

Excess entropy: Adding an independent subsystem

- Two independent subsystems A and B, thus

$$MI(X_A : X_B) = 0 \quad \Rightarrow H(X_{A\cup B}) = H(X_A) + H(X_B)$$

- Moreover

$$\begin{aligned} MI(X_{\{v^A\}} : X_B | X_{A\setminus\{v^A\}}) &= H(X_{\{v^A\}} | X_{A\setminus\{v^A\}}) - H(X_{\{v^A\}} | X_B, X_{A\setminus\{v^A\}}) \\ &= 0 \quad \forall v^A \in A \qquad \text{(same for } v^B \in B) \end{aligned}$$

- Thus

$$E(X_{A\cup B}) = E(X_A) + E(X_B) \, .$$

Excess entropy: Adding a copy

- Two copies $X_A = X_B$. Thus

$$H(X_{A\cup B}) = H(X_A) \qquad H(X_{v^A}|X_{(A\setminus v^A)\cup B}) = 0$$

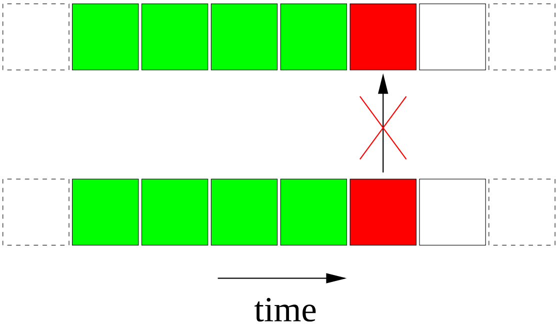
  because $\exists v^B \in B$ with $X_{v^A} = X_{v^B}$.

- The excess entropy of the two copies is equal to the entropy of one subsystem.

$$E(X_{A\cup B}) = H(X_A) \geq E(X_A)\ .$$

- Problem: The "complexity" of two identical copies measured by the excess entropy is independent of the complexity of the single system which is clearly counterintuitive and shows a severe limitation of the excess entropy as a complexity measure for finite systems.
- Note: This problem does not occur for the excess entropy for time series.

Two identical time series



- Excess entropy for time series:

$$E = \lim_{N\to\infty} (H(X_1, \ldots, X_N) - Nh_\infty) \quad h_\infty = \lim_{N\to\infty} H(X_t | X_{t-1}, \ldots, X_{t-N})$$

- Only conditioning on the past allowed.
- Causal Explanation

TSE complexity: Adding one element v'

$$\begin{aligned} C_{TSE}(X_{V\cup\{v'\}}) &= (1+\frac{1}{N+1})C_{TSE}(X_V) + \\ & \sum_{k=1}^{N+1} \binom{N+1}{k}^{-1} \sum_{\substack{Y\subseteq V \\ |Y|=k-1}} MI(X_{\{v'\}} : X_{V\setminus Y} | X_Y) \end{aligned}$$

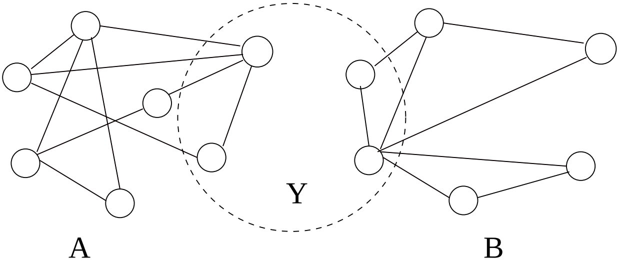
- C_{TSE} increases monotonically with system size
- Thus $C_{TSE}(X_{V\cup\{v'\}}) = (1+\frac{1}{N+1})C_{TSE}(X_V)$ for an independent v'.
- Suggests to use a normalized TSE-complexity

$$\tilde{C}_{TSE}(X_V) = \frac{1}{N+1} C_{TSE}(X_V)$$

TSE-complexity: Two independent subsystems

- Using $E(X_{A\cup B}) = E(X_A) + E(X_B)$ we have for an arbitrary subset $Y \subseteq A \cup B$

$$E(X_Y) = E(X_{Y\cap A}) + E(X_{Y\cap B}) \; .$$



TSE-complexity: Two independent subsystems

- Using $E(X_{A\cup B}) = E(X_A) + E(X_B)$ we have for an arbitrary subset $Y \subseteq A \cup B$
$$E(X_Y) = E(X_{Y\cap A}) + E(X_{Y\cap B})\ .$$
- By using this property and $C_{TSE}(X_V) = \frac{1}{2}\sum_{Y\subseteq V} \frac{1}{\binom{N}{|Y|}} E(X_Y)$ we could show that
$$C_{TSE}(X_{A\cup B}) = \frac{N_A + N_B + 1}{N_A + 1} C_{TSE}(X_A) + \frac{N_A + N_B + 1}{N_B + 1} C_{TSE}(X_B)$$
or
$$\tilde{C}_{TSE}(X_{A\cup B}) = \tilde{C}_{TSE}(X_A) + \tilde{C}_{TSE}(X_B)$$

TSE-Complexity: Two copies

- Using a reasoning similar to the case of the two independent subsystems one finally arrives at

$$C_{TSE}(X_{A\cup B}) = \frac{2N_A+1}{N_A+1} C_{TSE}(X_A) + \sum_{Y_A\subseteq A}\sum_{Y_B\subseteq A} \binom{N}{|Y_A|+|Y_B|}^{-1} H(X_{Y_B}|X_{Y_A})$$

- This leads to a very reasonable lower bound for the normalized TSE-complexity

$$\tilde{C}_{TSE}(X_{A\cup B}) \geq \tilde{C}_{TSE}(X_A)\ .$$

- But, a similar problem as in the case of the excess entropy can occur.

Summary

- The TSE-complexity is closely related to the excess entopy: it is proportional to the mean subset excess entropy averaged over all subsets.
- Both the excess entropy and the normalized TSE-complexity seems to be reasonable but not ideal complexity measures.
- The case of two copies indicated a certain limitation of the studied complexity measures for finite systems, which is not present in the time series case: One has to be more careful about what is explained using which information from the system.
- More specific complexity measures, e.g. restricted on causal explanations