Entropy & Complexity

José M. Amigó

Centro de Investigación Operativa

Universidad Miguel Hernández

03202 Elche (Alicante). Spain

jm.amigo@umh.es

OUTLINE

1. INTRODUCCION

SHORT CHRONOLOGY OF ENTROPY:

- In **physics** (as a measure of *disorder*):

  Clausius (1865), Boltzmann (1877), von Neumann (1927),...

- In **information theory** (as a measure of *uncertainty*):

  Shannon (1948),....

- In **dynamical systems** (as a measure of *randomness*):

  Kolmogorov (1958), Sinai (1959), Adler-Konheim-McAndrew (1965),...


S = k. log W
LVDWIG
BOLTZMANN
1844 – 1906
Dr PHIL PAULA
BOLTZMANN
ARTHUR
BOLTZMANN
LVDWIG
BOLTZMANN
HENRIETTE
BOLTZMANN

FIGURE 1

"*When Shannon had invented his quantity and consulted von Neumann on what to call it, von Neumann replied: 'Call it entropy. It is already in use under that name and besides, it will give you a great edge in debates because nobody knows what entropy is anyway'.*"

K. Denbigh (1990) How subjective is entropy. In *Maxwell's Demon, Entropy, Information, Computing* (ed. H.S. Leff and A.F. Rex), pp. 109-115, Princeton University Press.

COMPLEXITY:

- In **physics** (*complex phenomena*).
- In **statistical modelling** (*stochastic complexity*):

  Rissanen,...
- In **information theory** (*compressibility*):

  Lempel-Ziv,...
- In **dynamical systems** (*chaos*):

  Poincaré, Lyapunov, Yorke, Ruelle, Eckmann,...
- In **computer science** (*algorithmic complexity*):

  Kolmogorov, Chaitin,...

We will restrict to complexity in **time series analysis**.

2. COMPLEXITY AND STANDARD ENTROPY

$$x_1 x_2 \ldots x_n \ldots \equiv x_1^{\infty} \quad <====== \quad \boxed{\quad \mathbf{X} \quad}$$

where $x_n \in A$ (alphabet, state space, ...).

X (Data source)	**Realm**	**Entropy**
Stat. random process	Info. th./Dyn. syst.	Shannon, differential, permutation
Deterministic process	Dynamical systems	Metric, topological, permutation
Unknown / General	Time series analysis	Try permutation entropy

2.1 In information theory

Given the sequence

$$x_1^\infty = 010110100011011 1001...$$

its **LZ-76 complexity** is given by:

$$0|\,1|\,011|\,0100|\,011011|\,1001 \;\Rightarrow\; C(x_1^2) = 2, C(x_1^5) = 3, ..., C(x_1^{18}) = 6, ...$$

Theorem 1. [Lempel-Ziv] *If* $\mathbf{X}$ *is ergodic, then*

$$\limsup_{L\to\infty} \frac{C(x_1^L)}{L/\log L} = h(\mathbf{X}) \;(\textit{almost surely})$$

REMARK. Other definitions of complexity (e.g., **LZ-78**) are better suited for lossless data compression.

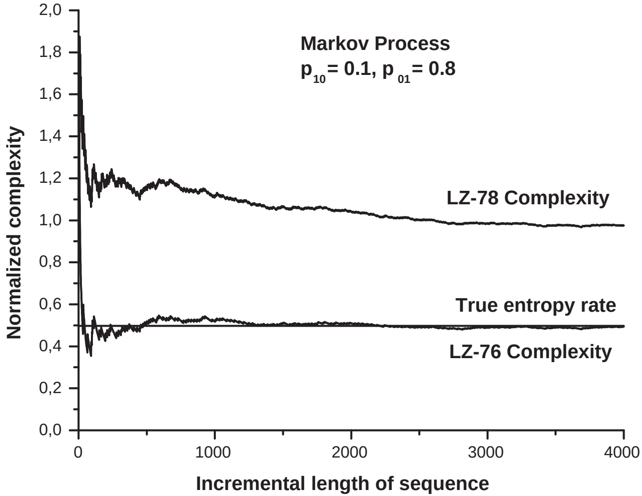

Markov Process
p10= 0.1, p 01= 0.8
LZ-78 Complexity
True entropy rate
LZ-76 Complexity
Normalized complexity
Incremental length of sequence
2,0
1,8
1,6
1,4
1,2
1,0
0,8
0,6
0,4
0,2
0,0
0
1000
2000
3000
4000

Figure 2

2.2 In dynamical systems

Dynamical complexity = 'chaos'

Theorem 2. [Ledrappier-Young] *Let (i) T be a C^2-diffeomorphism on a compact manifold, (ii) μ a T-invariant SRB measure (i.e., absolutely continuous along the unstable directions) and (iii) T ergodic w.r.t μ, then*

$$h_\mu(T) = \sum_j \lambda_j^+,$$

where λ_j are the (eventually repeated) Lyapunov exponents of (Ω, μ, T).

The *topological entropy* is related to the volumen growth, growth of periodic points and horseshoes, etc.

2.3 In time series analysis

The **algorithmic complexity** $K(x_0^{L-1})$ *of the finite sequence* x_0^{L-1} *is the length of the shortest binary program that, run on a universal Turing machine, outputs* x_0^{L-1}.

Let $\mathbf{X}^\gamma = (X_n^\gamma)$ be the associated symbolic dynamics of (Ω, μ, T) wrt the coarse-graining γ and

$$x_n = X_n^\gamma(\omega) \quad (\text{or } \mathbf{x} = \mathbf{X}^\gamma(\omega)), \;\; \omega \in \Omega.$$

Theorem 3. [Brudno 1983] *If* (Ω, μ, T) *is ergodic, then*

$$h_\mu(T) = \sup\nolimits_\gamma \lim\nolimits_{L\to\infty} \sup \tfrac{1}{L} K(x_0^{L-1}) \quad \mu\text{-}a.e.$$

REMARKS

(1) *The concept of* **linear complexity** *[Massey-Belerkamp] of a binary sequence is similarly defined using linear feedback shift registers.*

(2) *Contrarily to linear complexity, algorithmic complexity is* **not** *a* **computable** *quantity.*

(3) *Pseudo-random sequences can have a very low algorithmic complexity.*

3. ENTROPY AND ORDINAL PATTERNS

- Suppose that the alphabet A of $\mathbf{X}$ has a linear ordering $<$.
- Let $\pi \equiv \langle \pi_0, \pi_1, ..., \pi_{L-1} \rangle$ be the permutation of $\{0, 1, ..., L-1\}$:

$$\text{For ex.:} \quad L = 3 \;\; \text{and } \pi = \langle \pi_0, \pi_1, \pi_2 \rangle = \langle 2, 0, 1 \rangle \, .$$

The word $x_0, x_1, ..., x_{L-1}$ $(x_j \in A)$ *defines the* **ordinal L-pattern** π *if*

$$x_{\pi_0} \prec x_{\pi_1} \prec ... \prec x_{\pi_{L-1}},$$

where (for definiteness)

$$x_i \prec x_j \;\Leftrightarrow\; \begin{cases} x_i < x_j & \text{in } A \\ i < j & \text{if } x_i = x_j \end{cases} .$$

EXAMPLE. *If, e.g.,*

$$x_0^\infty = 4, 2, 10, 6, 10, 3, ...$$

Then:

$L = 2$	$4, 2 \sim \langle 1, 0 \rangle$;	$2, 10 \sim \langle 0, 1 \rangle$;	$10, 6 \sim \langle 1, 0 \rangle$;
$L = 3$	$4, 2, 10 \sim \langle 1, 0, 2 \rangle$;	$2, 10, 6 \sim \langle 0, 2, 1 \rangle$;	$10, 6, 10, ... \sim \langle 1, 0, 2 \rangle$.

What can ordinal patterns do for you?

3.1 In information theory

- **Metric permutation entropy** *of* $\mathbf{X}$,

$$h_m^*(\mathbf{X}) = -\lim_{L\to\infty} \frac{1}{L-1} \sum_{\pi} p(\pi) \log p(\pi),$$

where $p(\pi)$ = *probability of the ordinal L-pattern* π.

- **Topological permutation entropy** *of* $\mathbf{X}$,

$$h_{top}^*(\mathbf{X}) = -\lim_{L\to\infty} \frac{1}{L-1} \log N(\mathbf{X}, L),$$

where $N(\mathbf{X}, L)$ *is the number of* allowable *ordinal L-patterns.*

Theorem 4. *If* $\mathbf{X}$ *is an* ergodic *source with finite alphabet, then*

$$h(\mathbf{X}) = h_m^*(\mathbf{X}).$$

3.2 In dynamical systems

Let $I \subset \mathbb{R}^d$ compact interval, (I, μ, f) dynamical system.

Kolmogorov's strategy: $\mathbf{X}^\gamma = (X_n^\gamma)$ symbolic dynamics wrt a product partition γ.

Define
- **Metric permutation entropy** *of* f: $h_\mu^*(f) := \lim_{\|\gamma\| \to 0} h_m^*(\mathbf{X}^\gamma)$.
- **Topological permutation entropy** *of* f: $h_{top}^*(f) := \lim_{\|\gamma\| \to 0} h_{top}^*(\mathbf{X}^\gamma)$.

Theorem 5.

- *If f is ergodic, then $h_\mu(f) = h_\mu^*(f)$.*
- *If f is expansive, then $h_{top}^*(f) = h_{top}(f)$.*

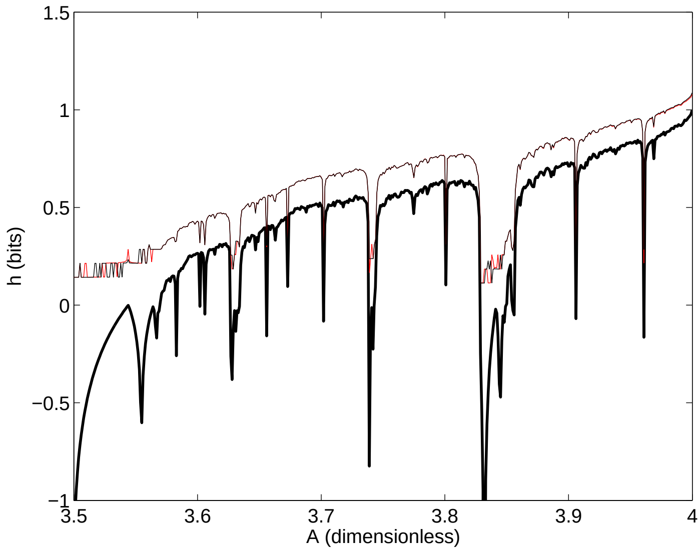
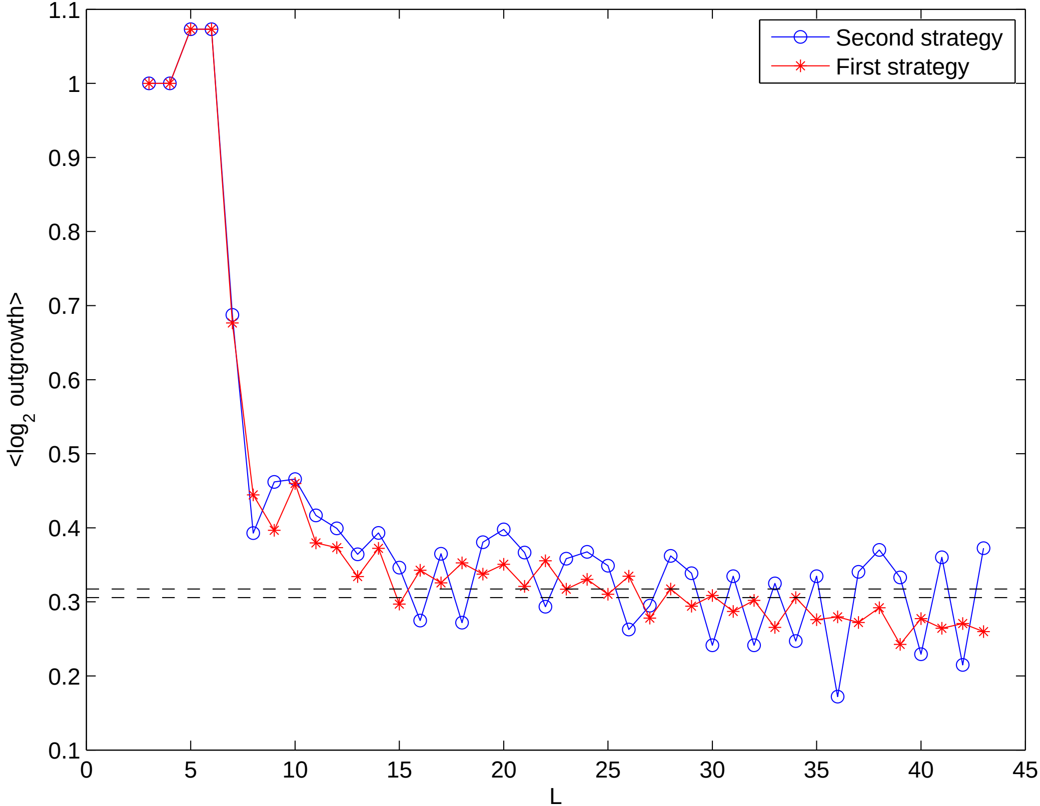


FIGURE 3



Second strategy
First strategy
<log$_2$ outgrowth>
L
1.1
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
5
10
15
20
25
30
35
40
45

3.3 In discrete mathematics

If your state space $\mathcal{S}$ is finite and $F : \mathcal{S} \to \mathcal{S}$ one-to-one, you can associate to the *discrete dynamical system* $(\mathcal{S},F)$ a **discrete entropy**, which is a kind of discretization of the metric permutation entropy.

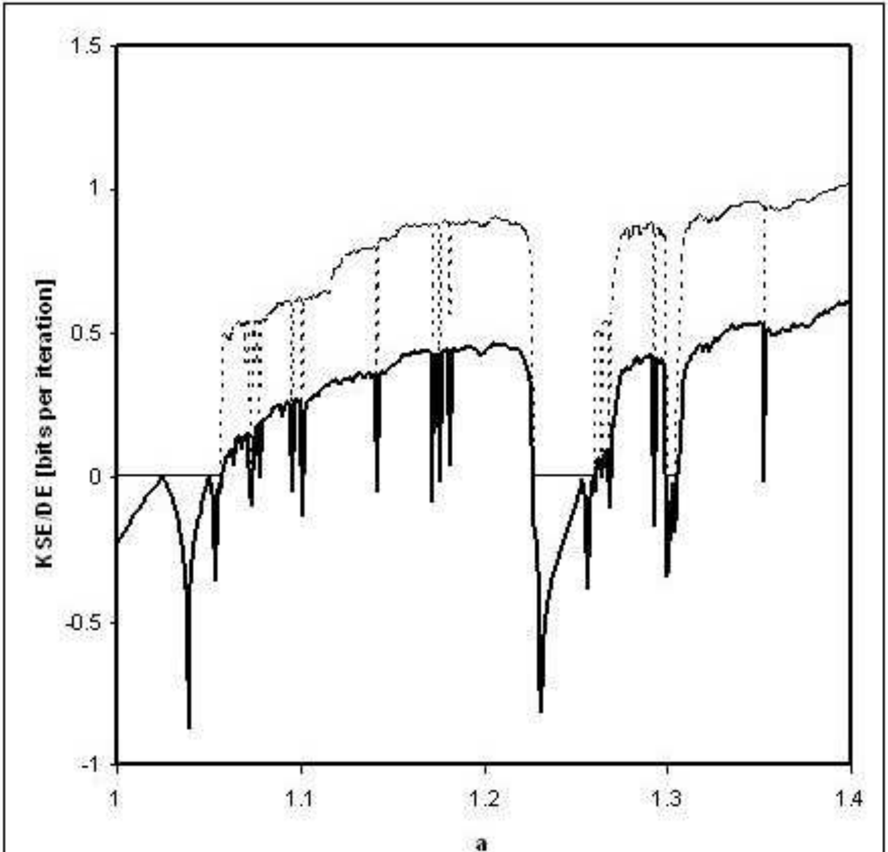
Discrete entropy can be used to estimate the metric entropy of d-dimensional interval maps.

Example. Two-dimensional quadratic map:

$$x_{n+1} = 1 - ax_n^2 + 0.3y_n$$

$$y_{n+1} = x_n$$

with $0 \leq x, y \leq 1$ and $1 \leq a \leq 1.4$.


1.5
1
0.5
0
-0.5
-1
1
1.1
1.2
1.3
1.4
KSE/DE [bits per iteration]
a

4. ORDINAL PATTERNS AND COMPLEXITY

Given $f : \Omega \to \Omega$, $x \in \Omega$ defines the **ordinal** L**-pattern** $\pi = \langle \pi_0, \pi_1, ..., \pi_{L-1} \rangle$ if

$$f^{\pi_0}(x) < f^{\pi_1}(x) < ... < f^{\pi_{L-1}}(x) \quad (0 \le \pi_i \le L-1).$$

π is a **forbidden (ordinal) pattern**, if no $x \in \Omega$ defines π.

Theorem 6. *If I is a compact, d-dimensional interval and $f : I \to I$ expansive with $h^*_{top}(f) < \infty$, then f has forbidden patterns, and*

$$\#\{\textit{allowable ordinal } L\textit{-patterns}\} = e^{L h^*_{top}(f)}.$$

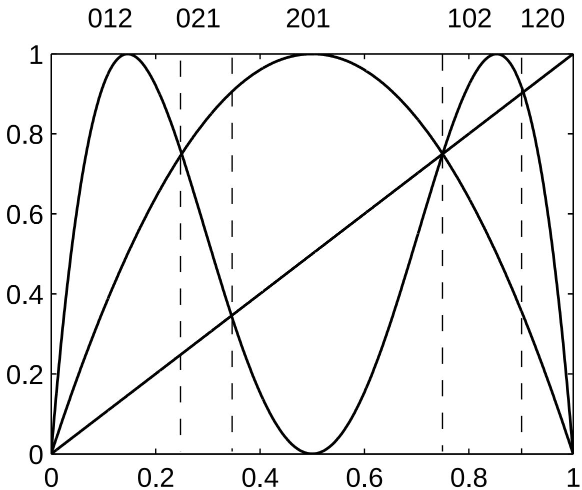


Figure 5

Properties of the forbidden patterns:

1. Forbidden patterns are invariant only under **order-ismorphisms**.

logistic map	*not order-isomorphic to*	shift map: $x \mapsto 2x \bmod 1$
one-sided $\left(\frac{1}{2}, \frac{1}{2}\right)$-Bernoulli shift	*order-isomorphic to*	shift map: $x \mapsto 2x \bmod 1$
two-sided $\left(\frac{1}{2}, \frac{1}{2}\right)$-Bernoulli shift	*order-isomorphic to*	the baker map

REMARK. A one- ot two-sided shift on N symbols have forbidden patterns for every $L \geq 2N$.

2. A given forbidden pattern generates a superexponentially growing trail of longer and longer forbidden patterns called **outgrowth forbidden patterns**. For ex., for the logistic map

$$\langle \pi_0, \pi_1, ..., \pi_{L-1} \rangle \text{ forb.} \Rightarrow \langle *, \pi_0 + n, *, \pi_1 + n, *, ..., *, \pi_{L-1} + n, * \rangle \quad \text{forb.}$$

EXAMPLE. The logistic map has 12 forbidden patterns of $L = 4$: seven are **outgrowth** of $\langle 2, 1, 0 \rangle$

$$(n = 0) \quad \langle 3, 2, 1, 0 \rangle , \langle 2, 3, 1, 0 \rangle , \langle 2, 1, 3, 0 \rangle , \langle 2, 1, 0, 3 \rangle$$
$$(n = 1) \quad \langle 0, 3, 2, 1 \rangle , \langle 3, 0, 2, 1 \rangle , \langle 3, 2, 0, 1 \rangle$$

and five are new (‘**roots**’):

$$\langle 0, 2, 1, 3 \rangle , \langle 1, 0, 2, 3 \rangle , \langle 1, 0, 3, 2 \rangle , \langle 1, 3, 0, 2 \rangle , \langle 3, 1, 2, 0 \rangle .$$

3. Forbidden patterns are **robust** against noise because they involve inequalities.

EXAMPLE. Let $f(x) = 4x(1-x)$, $0 \leq x \leq 1$, and consider the noisy time series

$$z_k = f^k(x_0) + \xi_k, \quad \xi_k \text{ white noise uniformly distributed in } [-\tfrac{1}{2}, \tfrac{1}{2}]$$

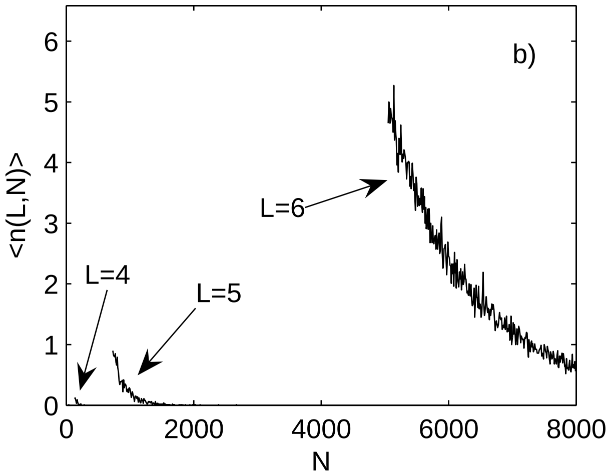


FIGURE 6

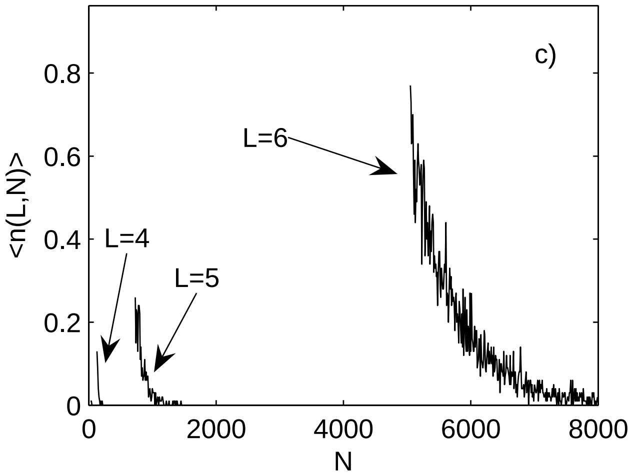


Figure 7

Complexity in time series analysis can mean (at least!) two things:

A. Complexity as **incompressibility** (in the sense of Kolmogorov-Chaitin): *A binary sequence is 'incompressible' if* $\exists c$ *constant such that* $K(x_0^{L-1}) \geq L - c$ *for* $\forall L$.

B. Complexity as **typicality** (in the sense of Martin-Löf): *There is a universal (or maximal) sequential test that, if passed, defines a sequence as being random or 'typical'.*

Theorem 7. [Levin-Schnorr-Chaitin] *A binary sequence is incompressible' iff it is typical.*

Consider now *real-valued* sequences.

- **deterministic sequences** have forbidden ordinal patterns
- **random sequences** have no forbidden ordinal patterns with probability 1.

QUESTION: Can forbidden patterns (or topological permutation entropy) be used to define the concept of complexity for real-valued sequences?

5. **CONCLUSION**

- Entropy is related from to such interesting concepts as complexity, randomness and typicality.
- Ordinal patterns are good for you: they provide a *unified* and conceptually *simple* approach to entropy in different frameworks.
- Ordinal patterns and/or topological permutation entropy can distinguish between random and deterministic sequences (at least, in the limit $L \to \infty$).

REFERENCES

(1) J.M. Amigó, M.B. Kennel and L. Kocarev, "The permutation entropy rate equals the metric entropy rate for ergodic information sources and ergodic dynamical systems". *Physica D* **210** (2005) 77-95.

(2) J.M. Amigó, L. Kocarev and J. Szczepanski, "Order patterns and chaos". *Physics Letters A* **355** (2006) 27-31.

(3) J.M. Amigó, L. Kocarev and I. Tomovski, "Discrete entropy". *Physica* D **228** (2007) 77-85.

(4) J.M. Amigó and M.B. Kennel, "Topological permutation entropy". *Physica* D (in press).

(5) J.M. Amigó, S. Elizalde, "Forbidden patterns and shift systems" (submitted).

(6) J.M. Amigó, S. Zambrano and M.A.F. Sanjuán, "True and false forbidden patterns in deterministic and random dynamics" (submitted).