Activity 7

Entropy of a dynamical system

Goals

Let N copies of a dynamical systems be initially in close states (initial conditions close in the phase space). If the system presents instabilities in the neighbourhood of these initial conditions, the orbits of the N copies will diverge each other after the Lyapunov horizon. The disorder in the set of the N copies increases then by the flow. We want to measure the mean increase of disorder generated by the flow by unit of time or at each iteration.
This activity does not concern the L-systems.

Theory

Discrete and continuous time conservative dynamical systems

Definition 36 (Partition). *Let (Γ, φ, μ) be a dynamical system with Γ compact of finite measure. A partition of the phase space is a set $\mathfrak{X} = \{\sigma\} \subset \mathscr{T}$ $(\sigma \neq \varnothing)$ such that*

$$\forall \sigma, \tau \in \mathfrak{X}, \sigma \neq \tau \quad \mu(\sigma \cap \tau) = 0 \qquad \textit{and} \qquad \bigcup_{\sigma \in \mathfrak{X}} \sigma \simeq \Gamma$$

$A \simeq B$ meaning $A = B \cup Z$ or $B = A \cup Z$ with $\mu(Z) = 0$ (A and B are μ-almost equal). $\sigma \in \mathfrak{X}$ is called classical micro-state.

Property 2. *Let $\mathfrak{X}$ be a partition of Γ (compact, $\mu(\Gamma) < +\infty$). If φ preserves μ and is injective, then $\varphi(\mathfrak{X}) = \{\varphi(\sigma)\}_{\sigma \in \mathfrak{X}}$ is another partition of Γ.*

Proof : $\forall \sigma, \tau \in \mathfrak{X}$, $\sigma \neq \tau$, $\varphi(\sigma \cap \tau) = \varphi(\sigma) \cap \varphi(\tau)$ ($\varphi(\sigma \cap \tau) \subset \varphi(\sigma) \cap \varphi(\tau)$ obviously; $x \in \varphi(\sigma) \cap \varphi(\tau) \Rightarrow \exists y \in \sigma, z \in \tau$ with $\varphi(y) = \varphi(z) = x$, but $\varphi(y) = \varphi(z) \Rightarrow y = z$ because φ is injective, then $y \in \sigma \cap \tau$ and $x = \varphi(y) \in \varphi(\sigma \cap \tau)$, and then $\varphi(\sigma) \cap \varphi(\tau) \subset \varphi(\sigma \cap \tau)$). Then $\mu(\varphi(\sigma) \cap \varphi(\tau)) = \mu(\varphi(\sigma \cap \tau)) = \mu(\sigma \cap \tau) = 0$. $\mu\left(\bigcup_{\sigma \in \mathfrak{X}} \varphi(\sigma)\right) = \bigcup_{\sigma \in \mathfrak{X}} \mu(\varphi(\sigma))$ since $\varphi(\sigma)$ are μ-almost two by two disjoint. Then $\mu\left(\bigcup_{\sigma \in \mathfrak{X}} \varphi(\sigma)\right) = \bigcup_{\sigma \in \mathfrak{X}} \mu(\sigma) = \mu(\Gamma)$. Then $\bigcup_{\sigma \in \mathfrak{X}} \varphi(\sigma) = \Gamma \setminus Z$ with $\mu(Z) = 0$. ■

Let ν be a probability measure defined on the same σ-algebra $\mathscr{T}$ than μ. The statistical entropy $S_{\nu,\mathfrak{X}}$ relative to $\mathfrak{X}$ and ν, is the measure of the lack of information on the system described by the set of micro-states $\mathfrak{X}$ with a lack of knowledge modelled by the probability distribution ν. The following properties must be satisfies:

- a sure micro-state $\nu(\sigma) = 1$ does not contribute to the lack of knowledge (we know that this micro-state is occupied).

- an inaccessible micro-state $\nu(\sigma) = 0$ does not contribute to the lack of knowledge (we know that this micro-state is not occupied).
- if the system is composed by two subsystems of phase spaces with partitions $\mathfrak{X}$ and $\mathfrak{Y}$, we must have $S_{\nu,\mathfrak{X}\times\mathfrak{Y}} = S_{\nu,\mathfrak{X}} + S_{\nu,\mathfrak{Y}}$ (the lack of knowledge concerning the composite system is the sum of the lacks of knowledge concerning its subsystems).

With this properties we have:

Definition 37 (Shannon entropy). *The statistical entropy relative to a partition $\mathfrak{X}$ and to a probability measure ν is*

$$S_{\nu,\mathfrak{X}} = -\sum_{\sigma\in\mathfrak{X}} \nu(\sigma)\ln\nu(\sigma)$$

The third property is well satisfied:

$$\begin{aligned} S_{\nu,\mathfrak{X}\times\mathfrak{Y}} &= -\sum_{\sigma\in\mathfrak{X}}\sum_{\tau\in\mathfrak{Y}} \nu(\sigma\times\tau)\ln\nu(\sigma\times\tau) \\ &= -\sum_{\sigma\in\mathfrak{X}}\sum_{\tau\in\mathfrak{Y}} \nu(\sigma)\nu(\tau)\ln\left(\nu(\sigma)\nu(\tau)\right) \\ &= -\underbrace{\sum_{\tau\in\mathfrak{Y}}\nu(\tau)}_{=1}\sum_{\sigma\in\mathfrak{X}}\nu(\sigma)\ln\nu(\sigma) - \underbrace{\sum_{\sigma\in\mathfrak{X}}\nu(\sigma)}_{=1}\sum_{\tau\in\mathfrak{Y}}\nu(\tau)\ln\nu(\tau) \\ &= S_{\nu,\mathfrak{X}} + S_{\nu,\mathfrak{Y}} \end{aligned}$$

Definition 38 (Joint partition). *Let $\mathfrak{X}$ and $\mathfrak{Y}$ two partitions of Γ. we denote by $\mathfrak{X}\vee\mathfrak{Y} = \{\sigma\cap\tau, \sigma\in\mathfrak{X}, \tau\in\mathfrak{Y}\}$ the joint partition.*

Property 3. $S_{\nu,\mathfrak{X}\vee\mathfrak{Y}} \leq S_{\nu,\mathfrak{X}} + S_{\nu,\mathfrak{Y}}$

Proof : This results from $\nu(\sigma\cup\tau) = \nu(\sigma) + \nu(\tau) - \nu(\sigma\cap\tau)$. ■

Let N copies of a conservative dynamical system (Γ, φ, μ) (with N large), $\mathfrak{X}$ be a partition of Γ (of finite measure), and $\{X_0^i\}_{i=1,\dots,N}$ be the N initial conditions. The initial dispersion is

$$d_0 = \max_{i,j} \|X_0^i - X_0^j\|$$

The measure of the disorder in the set of copies of the system, which is equivalent to the measure of the lack of information concerning a system randomly chosen into the set of N copies, is given by the Shannon entropy:

$$S_{\mathfrak{X},N}(t) = -\sum_{\sigma\in\mathfrak{X}} p_\sigma(t)\ln p_\sigma(t)$$

where $p_\sigma(t)$ is the rate of copies being in the micro-state σ at the date/iteration t:

$$p_\sigma(t) = \frac{\mathrm{card}\{i \text{ such that } \varphi^t(X_0^i)\in\sigma\}}{N}$$

$\lim_{N\to+\infty} p_\sigma(t) = \nu_t(\sigma)$ with ν_t a probability measure, and $\lim_{N\to+\infty} S_{\mathfrak{X},N}(t) = S_{\mathfrak{X},\nu_t}$.

To simplify, suppose that the length $d_\square$ of the micro-states is the same for all ($d_\square = \sup_{X,Y\in\sigma}\|X - Y\|$). If $d_0 < d_\square$, all copies can be initially in a same single micro-state σ_0: $p_{\sigma_0}(0) = 1$, $p_{\sigma\neq\sigma_0}(0) =$

0, and $S_{\mathfrak{X},N}(0) = 0$. During the evolution, the dispersion can be evaluated by the asymptotic linear approximation:

$$\begin{aligned} d_t &= \max_{i,j} \|\varphi^t(X_0^i) - \varphi^t(X_0^j)\| \\ &\leq e^{\underline{\lambda} t} d_0 \end{aligned}$$

where $\underline{\lambda}$ is the main asymptotic Lyapunov exponent of the orbits starting in σ_0 (we suppose the partition sufficiently thin in order to the Lyapunov exponents be approximately constant inside the micro-states). We can define the Lyapunov horizon of predictability for the partition as

$$t_H = \frac{\ln d_\square - \ln d_0}{\max(\underline{\lambda}, 0)}$$

After t_H the distribution of the copies into Γ implies necessarily several micro-states. $S_{\mathfrak{X},N}(t)$ is then an increasing function of t from t_H.

We want estimate the speed of the entropy increase. Suppose that at a time/step t we know that the system is in the micro-state $\sigma \in \mathfrak{X}$. At the previous step the system is in $\varphi^{-\Delta t}(\sigma)$ (where $\Delta t = 1$ for discrete time systems and is a discretisation time step for continuous time systems). But in general $\varphi^{-\Delta t}(\sigma) \notin \mathfrak{X}$ (it straddles several micro-states). To describe the situation with two steps, it needs then to refine the partition by $\mathfrak{X}_1 = \varphi^{-\Delta t}(\mathfrak{X}) \vee \mathfrak{X}$. The entropy variation induces by one step is then

$$\Delta S_{\mathfrak{X},1} = -\sum_{\sigma \in \mathfrak{X}_1} \mu(\sigma) \ln \mu(\sigma) + \sum_{\sigma \in \mathfrak{X}} \mu(\sigma) \ln \mu(\sigma)$$

By considering two steps, it needs anew to refine the partition by $\mathfrak{X}_2 = \varphi^{-\Delta t}(\mathfrak{X}_1) \vee \mathfrak{X}_1$, with an average increase of entropy:

$$\Delta S_{\mathfrak{X},2} = \frac{-\sum_{\sigma \in \mathfrak{X}_2} \mu(\sigma) \ln \mu(\sigma) + \sum_{\sigma \in \mathfrak{X}} \mu(\sigma) \ln \mu(\sigma)}{2}$$

and so with n steps we have

$$\Delta S_{\mathfrak{X},n} = \frac{-\sum_{\sigma \in \mathfrak{X}_n} \mu(\sigma) \ln \mu(\sigma) + \sum_{\sigma \in \mathfrak{X}} \mu(\sigma) \ln \mu(\sigma)}{n}$$

with $\mathfrak{X}_n = \varphi^{-1}(\mathfrak{X}_{n-1}) \vee \mathfrak{X}_{n-1}$. So the mean entropy increase by one step is

$$h_{\mathfrak{X},\varphi,\mu} = \lim_{n \to +\infty} \Delta S_{\mathfrak{X},n} = -\lim_{n \to +\infty} \frac{1}{n} \sum_{\sigma \in \mathfrak{X}_n} \mu(\sigma) \ln \mu(\sigma)$$

And to have a measure independent of the choice of the partition, we set the following definition:

Definition 39 (Kolmogorov-Sinaï entropy)**.** *We call metric entropy (so-called Kolmogorov-Sinaï entropy) of a conservative dynamical system (Γ, φ, μ) the quantity:*

$$h_{\varphi,\mu} = \sup_{\mathfrak{X}} \lim_{n \to +\infty} \frac{-1}{n} \sum_{\sigma \in \mathfrak{X}_n} \mu(\sigma) \ln \mu(\sigma)$$

with $\mathfrak{X}_n = \bigvee_{i=0}^{n-1} \varphi^{-i}(\mathfrak{X})$.

A very important result is the following:

Theorem 9 (Pesin theorem)**.** *If μ is the microcanonical measure (i.e. the measure of uniform probability on Γ) and φ is a conservative flow for μ we have*

$$h_{\varphi,\mu} = \int_\Gamma \sum_i m_i(X) \max(\underline{\lambda}_i(X), 0) d\mu(X)$$

where $(\underline{\lambda}_i(X))_i$ are the Lyapunov exponents of φ at X and $m_i(X)$ are the algebraic multiplicities (degree of degenerescence) of $\underline{\lambda}_i(X)$.

Dissipative systems

The definition of the entropy of dissipative system is not simple because by definition the measure is not preserved. A first possibility could consist to ignore the transient regime to focus on the permanent regime onto an attractor A. But if A is a stable fixed point or a stable invariant cycle or torus, the question has no interest because by definition the dynamics restricted to A is stable, and no disorder appears on A. The only interesting case would be a strange attractor A. Unfortunately, it is impossible to define an inner invariant measure μ_A onto a strange attractor (we will study the topology of the strange attractors in a next activity). It is then impossible to use the theory presented in the previous section.

An efficient approach consists to consider a logistic flow (I, f, dx) associated with the dynamical system. Let x_c be the critical point of the logistic flow and $\mathrm{Orb}(x_c) = \{x_0, x_1, ..., x_n\}$ be its orbit (which is n-cyclic because of the Fatou theorem). This orbit defines a partition of I in n intervals. Let $\sigma \in S_n$ be the permutation of $\{0, ..., n\}$ such that $x_{\sigma(0)} < x_{\sigma(1)} < ... < x_{\sigma(n)}$ (σ sort $\mathrm{Orb}(x_c)$ by increasing values). The partition of I is $\mathfrak{X} = \{I_i\}_{i=1,...,n}$ with $I_i = [x_{\sigma(i-1)}, x_{\sigma(i)}]$. Let $T^f \in \mathfrak{M}_{n\times n}(\{0,1\})$ be the transition matrix of f:

$$T^f_{ij} = \begin{cases} 1 & \text{if } I_j \subset f(I_i) \\ 0 & \text{otherwise} \end{cases}$$

The dynamical system is then reduced to a directed network of n nodes $\{I_i\}$. If f sends points of I_i into I_j the nodes I_i and I_j are linked (by a link form I_i to I_j) and then $T^f_{ij} = 1$ ($T^f_{ij} = 0$ means that no link goes from I_i to I_j).
Consider N copies of the logistic flow with initial conditions in a node I_i. The dynamics is equivalent that at each time, each copy moves to a node by randomly chosen an outgoing link of its current node. The disorder increases if the copies initially in the same nodes spread in several nodes. So the disorder must be increase with respect to the "mean" number of nodes reached by outgoing links. A quantity which can be estimated by the spectral radius of T^f : $r(T^f) = \max|\mathrm{Sp}(T^f)|$. Indeed, if T^f presents only one 1 by row (a single outgoing link by node), T^f is equivalent to the identity matrix by row permutation and obviously $r(T^f) = 1$. If we have $T^f_{ij} = 1\ \forall i, j$ (each node is linked by outgoing links to the others nodes). It is not difficult to show that $\mathrm{Sp}(T^f) = \{n, -1\}$ with

$$\begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix} \text{ eigenvector associated with } n$$

$$\begin{pmatrix} -1 \\ 0 \\ \vdots \\ 0 \\ 1 \end{pmatrix}, \begin{pmatrix} -1 \\ 0 \\ \vdots \\ 1 \\ 0 \end{pmatrix}, \begin{pmatrix} -1 \\ 1 \\ \vdots \\ 0 \\ 0 \end{pmatrix} \text{ eigenvectors associated with } -1$$

We have then $r(T^f) = n$.
The measure of disorder must be zero if the copies remain in the same node at each iteration (i.e. if $r(T^f) = 1$) and must be an increasing function of $r(T^f)$. The disorder is equivalent to the lack of information concerning the node where is a randomly chosen a copy after an iteration. The lack of information is measured by the Shannon entropy:

$$S = -\sum_{i=1}^{m} p_i \ln p_i$$

where p_i is the probability to find the copy at the node I_i. But after large number of iterations, we can think that the average number of nodes where the copy can be is equal to $r(T^f)$. And so there are $r(T_f)$

nodes $\{i_1, ..., i_{r(T^f)}\}$ occupied with a probability $\frac{1}{r(T^f)}$, the other nodes being not occupied (probability equal to zero). Then

$$S \simeq -\sum_{j=1}^{r(T^f)} p_{i_j} \ln p_{i_j} = -\sum_{j=1}^{r(T^f)} \frac{1}{r(T^f)} \ln \frac{1}{r(T^F)} = \ln r(T^f)$$

Finally we have

Definition 40 (Topological entropy)**.** *The topological entropy of a logistic flow f is the quantity*

$$h_f^{top} = \log_2 r(T^f)$$

where $r(T^f) = \max |\mathrm{Sp}(T^f)|$ is the spectral radius of the transition matrix of the flow.

Since T^f is binary valued, it is chosen to normalise the topological entropy by $\ln 2$ ($\log_2 r = \frac{\ln r}{\ln 2}$).

Example : Let $\mathrm{Orb}(x_c) = \{x_0, ..., x_4\}$ be the orbit of a critical point with $x_2 < x_0 < x_3 < x_4 < x_1$ ($\sigma : \{0,1,2,3,4\} \mapsto \{2,0,3,4,1\}$). We have then

$$\begin{aligned}
I_1 = [x_2, x_0] &\xrightarrow{f} [x_3, x_1] = I_3 \cup I_4 \\
I_2 = [x_0, x_3] &\xrightarrow{f} [x_1, x_4] = I_4 \\
I_3 = [x_3, x_4] &\xrightarrow{f} [x_4, x_0] = I_2 \cup I_3 \\
I_4 = [x_4, x_1] &\xrightarrow{f} [x_0, x_2] = I_1
\end{aligned}$$

$f(I_1)$ contains I_3 and I_4 then the first row of T^f is $(0,0,1,1)$, $f(I_2)$ contain only I_4, the second row is then $(0,0,0,1)$, etc.

$$T^f = \begin{pmatrix} 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 1 \\ 0 & 1 & 1 & 0 \\ 1 & 0 & 0 & 0 \end{pmatrix}$$

And after diagonalisation of T^f we have $h_f^{top} \simeq 0.597$.

Symbolic dynamical systems

Let $(\mathscr{A}^{\mathbb{Z}}, \varphi, \sharp)$ be a 1D cellular automaton (φ being the global transition rule). Let $R(s,n)$ be the number of distinct rectangles of n rows obtains with s cells (as columns) during the iteration. For example with $s = 4$

c_0	■□□■
c_1	□■■□
c_2	■□□■
c_3	□■■□
c_4	■□□■

with $c_{n+1} = \varphi(c_n)$ (here φ changes the color of each cell at each iteration). In this example we have $R(4,1) = 2$ (■□□■ and □■■□), $R(4,2) = 2$ (■□□■
□■■□ and □■■□
■□□■), etc.

Clearly, the disorder created by the cellular automaton is an increasing function of $R(s,n)$. The disorder can be assimilated to a lack of information. If we choose randomly a rectangle of s columns and n rows, there is then $R(s,n)$ possibilities for this one. So the probability to obtain a particular possible pattern can be estimated as $\frac{1}{R(s,n)}$. The Shannon entropy measuring the lack of information is then

$$S = -\sum_{\text{pattern}} p_{\text{pattern}} \ln p_{\text{pattern}} = -\sum_{\text{pattern}} \frac{1}{R(s,n)} \ln \frac{1}{R(s,n)} = \ln R(s,n)$$

Since we have considered n rows, this quantity is the entropy generated by n iterations. The average entropy generated by one iteration is then

Definition 41 (Topological entropy)**.** *The topological entropy of a 1D cellular automata is*

$$h_\varphi^{top} = \lim_{s\to+\infty} \lim_{n\to+\infty} \frac{\ln R(s,n)}{n}$$

where $R(s,n)$ is the number of distinct rectangles of n rows with s cells generated by the global transition map φ.

h_φ^{top} is not dependent of the initial row when $s \to +\infty$ and $n \to +\infty$ (except for a small number of special configurations). But in a numerical estimation where s and n are large but finite, it is preferable to make an average onto several random initial rows.

Work to be done

- If your system is conservative (Hamiltonian):
  1. – Define a partition $\mathfrak{X}$ of the phase space and choose a micro-state σ_0 for which the main asymptotic Lyapunov exponent is positive. Compute the associated Lyapunov horizon of predictability t_H.
     – Choose randomly N initial conditions in σ_0 with N large and compute the orbits of these N points. At each time step, compute $p_\sigma(t)$ the rate of points occupying each micro-state σ, and compute the associated Shannon entropy $S_{\mathfrak{X},N}(t)$.
     – Draw the evolution of the Shannon entropy. This one starts at 0 and if $\mathfrak{X}$ is sufficiently thin and properly, chosen it increases almost linearly from t_H until to reach a maximal value. If $S_{\mathfrak{X},N}$ does not seem to increase almost linearly, change $\mathfrak{X}$. Compute the slope of the increase (which is an estimation of the entropy of the dynamical system in the neighbourhood of σ_0.
     – If your system presents a large variation of positive main asymptotic Lyapunov exponents, remake this analysis for the different initial micro-states of different Lyapunov exponents and compute the average entropy of your dynamical system.
  2. If the microcanonical measure is preserved by the flow, compute the Kolmogorov-Sinaï entropy of your dynamical system by using the Pesin theorem. Compare with the previous result.
- If your system is dissipative and can be reduced to a logistic flow:
  1. – Define the partition associated with the orbit of the critical point.
     – Choose randomly N initial conditions in an interval of the partition (with N large) and compute the orbits of these N point. At each iteration n, compute $p_{i,n}$ the rate of points occupying each interval I_i, and compute the associated Shannon entropy S_n. Plot the sequence (S_n) which increases almost linearly before to reach a maximal value. Compute the slope of the increase.
     – Repeat this for all initial interval and compute the average value of the entropy slopes. This quantity is an estimation of the entropy of the dynamical system.

2. Compute the transition matrix of the flow, its spectral radius and the topological entropy of your dynamical system. Compare with the previous result.

- If your system is a cellular automaton:
  1. For a 1D representation of your cellular automaton, code a function permitting to count the number of distinct $n \times s$ rectangles induced by the iterations.
  2. Choose a large value of the number of cells s, compute the evolution of an initial random raw and compute $R(s, n)$ for some increasing values of n.
  3. Plot the sequence $(\ln R(s, n)/n)_n$. If s has been chosen sufficiently large, this one increases almost linearly (if it is not the case, increase the value of s). Compute the slope of the increase, this one is an estimation of the topological entropy.
  4. Change the initial random raw to verify the stability of the estimation.