Review of Lectures 5-8: Entropy in Dynamical Systems

Maximilian Puelma Touzel

October 30, 2012

1. Entropy and Information in a Dynamical System

How can dynamics affect the information about the state of a dynamical system? In particular, we want to know how the map of a dynamical system influences the amount of information we obtain about the state of the system.

1.1. What is state information?

Definition 1.1. With a partition, $\alpha = \{\alpha_i\}$, and a measure, μ, on a space, X, the entropy of the partition is

$$H(\alpha) = \sum \mu(\alpha_i) \log(\mu(\alpha_i)). \tag{1.1}$$

More generally, a cover can be used instead of a partition. One can also forgo defining a measure by just using the uniform measure on α, or, given a probability density function, ρ, use $\mu(\alpha_i) := \int_{\alpha_i} \rho dx$. In the latter case, the measure is of probability and one obtains the Shannon entropy of the partition, interpreted as a measure of the average uncertainty associated with observing the state in a particular partition element.

If the partition is a realization of the finite measurement precision with which we can specify the state of the system, the information about the system acquired through such a measurement is the average reduction in our prior uncertainty about in which partition element the system resides, as captured by the entropy,

$$\begin{aligned} I(\alpha) := &\ H_{before} - H_{after} && (1.2)\\ = &\ H(\alpha) - 0 && (1.3)\\ I(\alpha) = &\ H(\alpha) && (1.4) \end{aligned}$$

1.2. What is the question? Now, given a map, $T : X \to X$, we can ask what additional information we can obtain about the state. In particular, with the knowledge that $x \in \alpha_i$, *what more does* Tx,T^2x,..., *i.e. future iterates of* x, *tell us?*

While this is perhaps the most natural question given the context, and so the one that we consider in what follows, there are important other questions that we encourage the reader to consider. For example, what additional information is associated to how the map evolves distributions on X (see section on information loss), or what information do past iterates of x contain about future iterates of x.

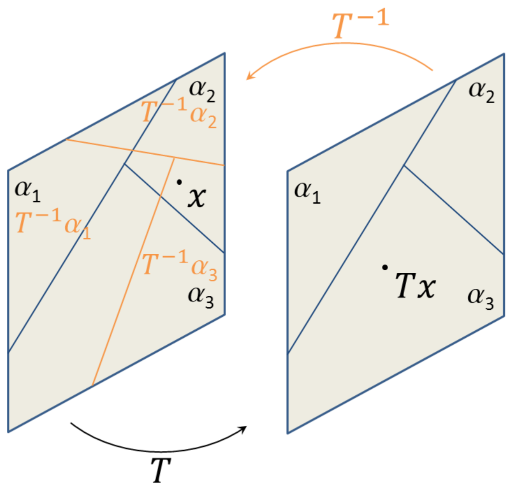

FIGURE 1. Refining our knowledge of x with back iterated partition elements. See text for more details.

1.3. The conceptual answer through an example. For purposes of example, consider a uniform measure on the partition given in Fig.1. The information associated with a measurement that determines in which partition element the system resides is

$$I = \log|\alpha| \tag{1.5}$$

$$= \log 3. \tag{1.6}$$

Now, with access to the evolution of the system as defined by the map T, and knowing $x \in \alpha_2$, we specify x in a refined partition $\alpha \vee T^{-1}\alpha$:

(1) x is mapped forward using T and we observe $Tx \in \alpha_3$.
(2) The pre-image set of α_3 must then contain x.
(3) x is thus contained in the intersection of α_2 and $T^{-1}\alpha_3$,

So now that x can be specified in an element of the refined partition $\alpha \vee T^{-1}\alpha$, we receive more information upon measurement of x,

$$I = \log\left|\alpha \vee T^{-1}\alpha\right| \tag{1.7}$$

$$= \log 7. \tag{1.8}$$

We can apply this procedure using higher powers of T so that the sequence of partition elements occupied through successive applications of the transformation, $Tx, T^2x, ...$, generates an ever refined partition, $\alpha \vee T^{-1}\alpha \vee T^{-2}\alpha \vee ...$, within which the state, x, can be uniquely specified with ever higher precision. Thus,

$$\begin{array}{ccc} occupancy & \approx & refinement \\ \text{in time} & & \text{in space} \end{array}$$

This is a core idea in symbolic dynamics where symbolic trajectories, i.e. sequences of partition elements (akin to Shannon-style symbolic strings, or 'words'), can be mapped to initial conditions.

2. Entropies and their Relations

The program we have followed is to study the controlled asymptotic behaviour of something like $\eta_N = H(\alpha \vee ... \vee T^{-N+1}\alpha)$ in spaces with different structures.

$$\text{(1)} \quad \begin{array}{c}\text{Topological}\\ \text{Space}\\ (X,\tau)\end{array} : h_{\text{top}}(T) := \sup_{\substack{\mathbb{A} \\ \text{covers of X}}} h(\mathbb{A},T), \;\; h(\mathbb{A},T) := \lim_{N\to\infty} \frac{\eta_N}{N}$$

$$\text{(2)} \quad \begin{array}{c}\text{Probability}\\ \text{Space}\\ (X,\mathbb{B},\mu)\end{array} : h_\mu(T) := \sup_{\substack{\alpha \\ \text{partitions of X}}} h_\mu(\alpha,T), \quad \begin{array}{rl} h_\mu(\alpha,T) & := \lim_{N\to\infty} \frac{\eta_N}{N} \\ & \equiv \lim_{N\to\infty} \eta_N - \eta_{N-1}\end{array}$$

$$\text{(3)} \quad \begin{array}{c}\text{Differentiable}\\ \text{Manifold}\\ (X,\mathbb{B},\mu)\\ F: M \to M,\\ \text{diffeomorphism},\\ \dot{x} = F(x)\end{array} : h_\lambda(F) := \textstyle\sum_{\lambda_i>0} \lambda_i, \;\; \lambda_i = \displaystyle\lim_{N\to\infty} \frac{1}{N} \ln \frac{|\xi(N)|}{|\xi(0)|}, \xi = \delta x,$$

with the important identities

(2)→(1): $h_{\text{top}}(T) \equiv \sup_\mu h_\mu(T)$ and
(3)→(2): $h_\mu(F) \equiv h_\lambda(F)$ under certain restrictions. (Pesin Identity)

While both $h_{\text{top}}(T)$ and $h_\mu(T)$ are accessible quantities for only the most simple of maps, $h_\lambda(F)$ offers a straightforward calculation for any differentiable map, and so the Pesin Identity is a crucial result. However, the conditions for it to apply involve subtle notions relating measures and maps.

3. Measures and Maps: Two Approaches

3.1. Given μ, study T's that preserve μ, i.e. $\mu(T^{-1}\alpha_i) = \mu(\alpha_i)$. The key property used for the Pesin Identity is that of ergodicity.

Definition 3.1. A transformation, T, is ergodic with respect to μ if T-invariant subsets of X have full or null measure.

The general importance of ergodicity is seen in the following theorem.

Theorem 3.2 (Birkhoff). *If T is ergodic, for any $f \in \mathbb{L}(\mu)$ and for μ-almost all x:*

$$\lim_{N\to\infty} \frac{1}{N} \sum_{i=0}^{N-1} f(T^N x) = \int f d\mu \tag{3.1}$$

In other words, *a single orbit samples the whole (μ-relevant) space.*

The Lyapunov exponents, as quantities calculated as averages over trajectories, are subject to this theorem, which ensures their uniqueness regardless of starting point, x, and their relevance as a legitimate property of the attractor. This theorem thus provides the logic behind the Pesin Identity. However, the key assumption on T, that of ergodicity, is difficult to prove in general. For that reason, a second approach is often taken (see next section).

The practical limitation provided by the supremum over partitions is addressed in the following definition and theorem.

Definition 3.3. A partition, α, is generating if, for the finite subalgebra of $\mathcal{B}$, $\mathcal{A}$, that it generates,

$$\bigvee_{N=-\infty}^{\infty} T^N \mathcal{A} = \mathcal{B}. \tag{3.2}$$

THEOREM 3.4 (Kolmolgorov-Sinai). *For an invertible, measure-preserving T of $(X, \mathcal{B}, \mu)$, if α is a generating partition*

$$h_\mu = h_\mu(\alpha, T). \tag{3.3}$$

In other words, a generating partition achieves the supremum in the definition of $h_\mu(T)$. In practice, the heuristic is to take a partition of hypercubes of size $\epsilon \ll 1$. See pg. 198-202 of Vulpiani.

3.2. Given T, study μ's that are invariant under T. The existence of such an invariant measure is guaranteed by a fixed-point theorem applied the measure-evolving Perron-Frobenius operator.

When T is ergodic with respect to this μ we can apply Birkhoff's Theorem, on which the following important theorem by Ruelle relies.

THEOREM 3.5 (Ruelle). *For μ an F-invariant probability measure on M*

$$h_\mu(F) \leq \int \sum_{i:\lambda_i > 0} \lambda_i(x) d\lambda(x) \tag{3.4}$$

If, in addition, F is $\mathcal{C}^2$ and ergodic (with μ equivalent to the Reimannian metric), then

$$h_\mu(F) \equiv \sum_{i:\lambda_i > 0} \lambda_i(x) \quad \text{(Pesin Identity)}. \tag{3.5}$$

However, the equation in Birkhoff's theorem holds for μ-almost all x and μ may only have support on an unphysical subset of X (where physical means typical, e.g. stable in the presence of noise, in the physical system that the model purports to represent). Thus, since we are interested in physical systems, the following is a useful property of a measure.

DEFINITION 3.6. A T-invariant measure, μ, is SRB (*Sinai-Ruelle-Bowen*) if Birkhoff's equation holds for all $f \in \mathcal{C}^2$ and for all $x \in \mathcal{B} \subseteq M$, where $\mathcal{B}$ has finite Lebesgue measure.

In other words, SRB measures allow us to start the trajectory from any initial condition in a Lebesgue measurable, i.e. physical, region and still end up with Birkhoff's '$\langle\rangle_{time} = \langle\rangle_{space}$' relation. SRB thus strips ergodicity down to what we need it for in a realistic situation, articulated in the following theorem.

THEOREM 3.7. *For $F \in \mathcal{C}^2$, ergodicity implies SRB if and only if the Pesin Identity holds.*

That is, when SRB can be substituted for ergodicity, we can obtain $h_\mu(F)$ from $h_\lambda(F)$. As for when the substitution is valid, see the section on Hyperbolicity in the Jost text as a start. As far as I can tell, this question is generally difficult to answer and at the forefront of current research.