Chapter 8

Information and Entropy

8.1 Definitions

A sequence of N coin tosses (of an assumed unbiased coin) has 2^N possible outcomes: there is an uncertainty in the outcome that we measure by the "entropy" which is the log of the number of possible equally likely outcomes:

$$S = \log 2^N = N \log 2. \tag{8.1}$$

We could use the sequences e.g. heads,heads,...tail to send a message, and we could send 2^N different messages. The "information capacity" of this scheme is again measured by the log of the number of possible messages

$$I = \log 2^N = N \log 2. \tag{8.2}$$

In this context we would often use base 2 for the log and say there are N bits of information. An alternative point of view is that the measurement of a particular result i.e. sequence of heads and tails has told us something about the system and we have learned $N \log 2$ bits of information The ideas of uncertainty of outcome (entropy) and what has been learned from a measurement (information) are complementary.

Generalizing these ideas to a system of N possible results with independent probabilities p_i (that may be different) gives the expression for the entropy of the system S or the information learned by finding a particular result

$$I = S = -\sum_{i=1}^{N} p_i \log p_i. \tag{8.3}$$

This result is familiar from thermodynamics (see below for a discussion of this analogy) and is due to Shannon [1] in the context of information.

8.2 Dynamical Systems

In a dynamical system the entropy is defined through a partition of the phase space with the p_i given by the integral of the invariant measure $\rho(\vec{x})$ (probability density) over the ith element..

Define the partition $\beta = \{B_i\}$, $i = 1 \ldots N$ with B_i non-empty, non-intersecting sets that cover the attractor (e.g. n-boxes for an n dimensional phase space). Then the probability of finding a point in the box B_i is $\int_{B_i} dV_i\, \rho(\vec{x})$. The entropy of the partition of the dynamical system is defined as

$$S = -\sum_{i=1}^{N} p_i \log p_i . \tag{8.4}$$

This tells us about the uncertainty coming from the "random" aspect of the dynamics. This quantity is not immediately useful, since it depends on the scheme of partitioning (e.g. the box size) as well as intrinsic properties of the attractor. Two related quantities have been defined to give intrinsic properties. One is the scaling of the entropy as the box size of the partition is reduced: this defines the "information density" of the attractor and is discussed in chapter 9. The information density is a static property of the attractor. A second quantity tells us how the uncertainty or information of the system evolves in time, or under iteration for a map. This is known as the Kolmogorov, Kolmogorov-Sinai, or metric entropy. A useful reference is by Farmer [2].

8.2.1 Kolmogorov Entropy

The idea is usefully introduced using the shift map (figure 8.1a).

Suppose we are limited to measuring the variable x to a precision of $\frac{1}{2}$, i.e. we can only measure whether x is in the range or $\frac{1}{2} < x < 1$. We can ask how our knowledge of $x = x_0$ increases if we iterate the map, measuring x after each iteration with this same precision. Suppose after one iteration we find x_1 to be in the range $0 < x < \frac{1}{2}$. Then we know that x_0 must lie in ranges of the interval that are preimages of this range: the intersection of this result with the direct measurement of x_0 has now localized x_0 to one quarter of the unit interval. As the

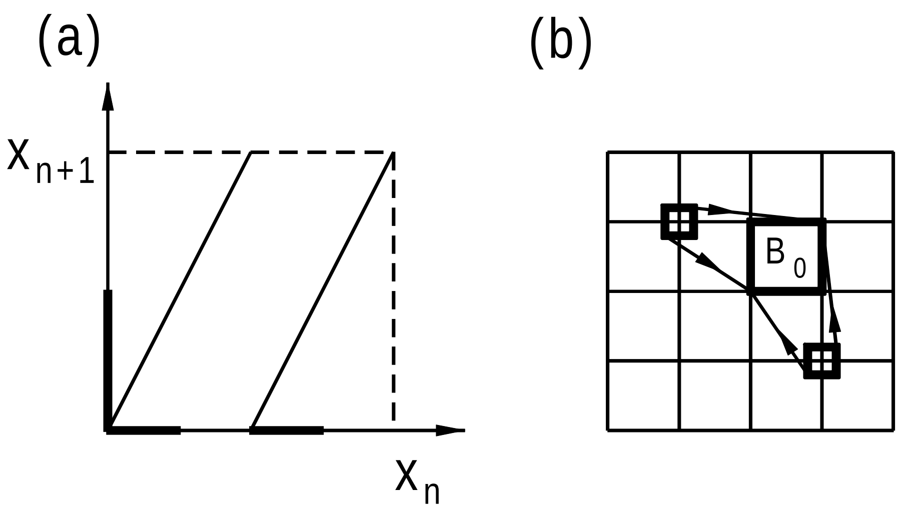

Figure 8.1: Kolmogorov Entropy. (a) Entropy production in the shift map. The heavy lines on the abscissa denote the preimages of $0 < x < \frac{1}{2}$ on the ordinate, and serve to refine the knowledge of x. (b) Refining the partition: the preimages of the box B_0 refine the partition.

iteration continues we learn 1 bit of information at each iteration, and this is the Kolmogorov entropy of the dynamical system.

An alternative point of view is to suppose we know the initial value x_0 to a certain precision (e.g. 8 bits). The Kolmogorov entropy tells us how the precision of our *prediction* for the nth iterate x_n decreases with n, due to the "sensitive dependence on initial conditions".

More generally we again define the partition $\beta_0 = \{B_i\}$ dividing the non-empty region of phase space into non-overlapping boxes B_i and we suppose that at each iteration (or after each fixed time step) all we know is in which box $\vec{x}$ lies. We can define the preimage $M^{-1}(B_i)$ of each B_i which is all points that will be mapped into B_i after a single iteration or time step (figure 8.1b). Note that this operation does not require that the inverse mapping (time reversed dynamics) exists. The measurement that $\vec{x}_0$ lies in box i_0 say, and $\vec{x}_1$ lies in box i_1 then tells us that $\vec{x}_0$ in fact lies in the region $B_{i_0} \cap M^{-1}(B_{i_1})$ i.e. the intersection of B_{i_0} with the preimage of B_{i_1}. The intersection of these two sets gives a finer partition β_1 of the attractor

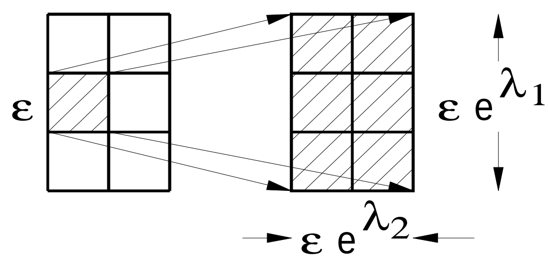

Figure 8.2: Growth of uncertainty due to expansion at a rate given by the positive Lyapunov exponents.

$\beta_1 = \left\{ B_i \cap M^{-1}(B_j) \right\}$ and the entropy of this partition is

$$S(\beta_1) = -\sum_{ij} p_{ij} \log p_{ij} \tag{8.5}$$

where p_{ij} is the integral of the measure over the box $B_i \cap M^{-1}(B_j)$. The Kolmogorov entropy is defined as the rate of change of entropy due to the finer partitioning given by each iteration or time step

$$K = \lim_{m\to\infty} \frac{1}{m} S(\beta_m) \tag{8.6}$$

(actually the sup of this over all choices of initial partition β_0) with

$$\beta_m = \left\{ B_{i_0} \cap M^{-1}(B_{i_1}) \cap M^{-2}(B_{i_2}) \ldots \cap M^{-m}(B_{i_m}) \right\}. \tag{8.7}$$

A positive value of K may be used to define the existence of chaos.

8.2.2 Relationship with Lyapunov Exponents

Since the growth of uncertainty, or refinement of the partition, is due to the divergence of nearby trajectories, it might be expected that there is a relationship between the Kolmogorov entropy K and the Lyapunov exponents. In fact Ruelle has shown that K is bounded by the sum of positive exponents

$$K \le \sum_{\lambda^{(i)}>0} \lambda^{(i)} \tag{8.8}$$

and the equality has been proven for the special class of "Axiom A" attractors. This idea is easily motivated, since a box of side ε is expanded in the growing directions to sides $\varepsilon e^{n\lambda^{(1)}}$, $\varepsilon e^{n\lambda^{(2)}}$ ... etc, so that after n iterations an initial condition in one of the boxes will lie somewhere in one of $e^{n(\lambda^{(1)}+\lambda^{(2)}+\cdots)}$ boxes (figure 8.2).

8.3 Comparison with Thermodynamics

The entropy in thermodynamic systems has dynamical *predictive* power expressed in the statement that "the entropy of an isolated system tends to increase with time". To put meaning into this statement we define $S(W_j)$ the entropy as a function of certain macroscopic (or "thermodynamic") variables W_j, where S is given by (8.3) with the sum over the microstates i consistent with the macrostate values W_j. The content of the statement is then that the W_j will evolve in the direction of increasing S. This profound result depends crucially on the Hamiltonian nature of the underlying dynamics and the resulting Liouville's theorem (the statement that the probability measure in phase space evolves as an incompressible fluid). This leads to the result that the probabilities p_i are known *independently of a detailed solution to the dynamics*—in an isolated system each "state" or volume of phase space is equally likely. In dissipative dynamical systems, where there is no Liouville's theorem and the p_i are not *a priori* known, the concept of entropy becomes descriptive (diagnostic) rather than predictive.

January 16, 2000

Bibliography

[1] C.E. Shannon and W. Weaver, *The Mathematical Theory of Communication*, University of Illinois Press (Urbana 1963)

[2] J.D. Farmer, Z. Naturforsch. **37a**, 1304 (1982)