Variational Principle

Measure-theoretic and topological entropy are related via the Variational Principle:

Theorem: Let (X, d) be a compact metric space and $T : X \to X$ a continuous map. Then

$$h_{top}(T) = \sup\{h_\mu(T) : \ \mu \text{ is a } T\text{-invariant probability measure}\}.$$

Any measure μ such that $h_{top}(T) = h_\mu(T)$ is called a measure of maximal entropy.

If there is a unique measure of maximal entropy $\mu_{\max}$, then (X, T) is called intrinsically ergodic.

Variational Principle

Remark: A measure of maximal entropy is automatically ergodic. Indeed, If $\mu_{\max}$ is not ergodic, say $\mu_{\max} = \alpha\mu_1 + (1-\alpha)\mu_2$, then

$$h_{\mu_{\max}}(T) = \alpha h_{\mu_1}(T) + (1-\alpha) h_{\mu_2}(T),$$

because measure-theoretic entropy is linear in the measure (check the definitions). But this means that at least one of μ_i, $i = 1, 2$ has $h_{\mu_i} \geq h_{\mu_{\max}}(T)$.

Remark: A measure of maximal entropy need not exist if T is discontinuous. For example, the Gauss map $G(x) = \frac{1}{x} - \lfloor \frac{1}{x} \rfloor$ has no measure of maximal entropy.

Exercise: Show that the Gauss map has infinite topological entropy.

Measures of Maximal Entropy

Most dynamical systems we see are intrinsically ergodic, but finding this measure of maximal entropy is not always simple.

- Any uniquely ergodic system is intrinsicially ergodic.
- For the full shift on N symbols, the $(\frac{1}{N}, \dots, \frac{1}{N})$-Bernoulli measure is the unique measure of maximal entropy.
- For transitive maps $T : [0,1] \to [0,1]$ of constant slope $\pm s$, $|s| > 1$, the measure that is absolutely continuous w.r.t Lebesgue is the unique measure of maximal entropy.
- Lebesgue measure is the unique measure of maximal entropy of hyperbolic toral automorphism.

The next main result (a theorem due to Parry) is about finding the maximal measure for subshifts of finite type.

Subshifts of Finite Type

Let $A = (A_{ij})_{i,j=1}^N$ be a non-negative $N \times N$ integer matrix.

- We A it transition matrix because A_{ij} usually indicates whether (or in how many ways) you can go from state i to state j in a Markov partition.
- A is irreducible if for every i, j there is k such that the i, j-entry of A^k is positive.
- Let $p(i) = \min\{k \geq 1 : \text{ the } i, i\text{-entry of } A^k \text{ is positive}\}$. A is aperiodic if $gcd\{p(i) : p(i) \text{ exists}\} = 1$.
- A is primitive if A is both irreducible and aperiodic. Alternatively, there is k such that A^k is a strictly positive matrix.

Subshifts of Finite Type

The set of (bi)infinite strings

$$\Sigma_A = \{(x_i)_{i\in\mathbb{Z}} : x_i \in \{1,\dots,N\}, A_{x_i,x_{i+1}} > 0 \text{ for all } i \in \mathbb{Z}\}$$

is shift-invariant and closed in the standard product topology of $\{1,\dots,N\}^{\mathbb{Z}}$. Hence it is a *subshift*.

It is called subshift of finite type (SFT) because of the finite collection of forbidden words (namely the pairs i, j such that $A_{i,j} = 0$) that fully determines Σ_A.

The word-complexity

$$p_n(\Sigma_A) := \#\{x_0 \dots x_{n-1} \text{ subword appearing in } \Sigma_A\}$$

Because the n-cylinders form an open 2^{-n}-cover of Σ_A:

$$h_{top}(\sigma|_{\Sigma_A}) = \lim_{n\to\infty} \frac{1}{n} \log p_n(\Sigma_A) = \log \lambda,$$

where λ is the leading eigenvalue of the transition matrix A.

Subshifts of Finite Type

Perron-Frobenius Theorem: Let A be a primitive nonnegative $N \times N$-matrix. Then A has a unique (up to scaling) eigenvector with all entries > 0. The corresponding eigenvalue λ is positive, has multiplicity one, and is larger than the absolute value of every other eigenvalue of A.

- λ is called the leading or Perron-Frobenius eigenvalue.
- If A is not irreducible, then λ can have higher multiplicity. For example $A = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$.
- If A is not aperiodic, then there can be other eigenvalues of the same absolute value as λ. For example $A = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$.
- The Perron-Frobenius Theorem holds both for left and right eigenvalues.

Parry Measure

Bill Parry constructed the measure of maximal entropy, which is now called after him. Let (Σ_A, σ) be a subshift of finite type on alphabet $\{1, \dots, N\}$ with transition matrix $A = (A_{i,j})_{i,j=1}^N$, $A_{ij} \in \{0, 1\}$, so $x = (x_n) \in \Sigma_A$ if and only if $A_{x_n, x_{n+1}} = 1$ for all n.

We assume that A is aperiodic and irreducible. Then by the Perron-Frobenius Theorem, the leading eigenvalue λ has multiplicity one, is larger in absolute value than every other eigenvalue, and $h_{top}(\sigma) = \log \lambda$.

The left and right eigenvectors

$$u = (u_1, \dots, u_N) \text{ and } v = (v_1, \dots, v_N)^T$$

associated to λ are unique up to a multiplicative factor. We will scale them such that they are positive and

$$\sum_{i=1}^N u_i v_i = 1.$$

Parry Measure

Define the Parry measure by

$$\begin{aligned} p_i &:= u_i v_i = \mu([i]), \\ p_{i,j} &:= \frac{A_{i,j} v_j}{\lambda v_i} = \mu([ij] \mid [i]), \end{aligned}$$

so $p_{i,j}$ indicates the conditional probability that $x_{n+1} = j$ knowing that $x_n = i$. Therefore $\mu([ij]) = \mu([i])\mu([ij] \mid [i]) = p_i p_{i,j}$. It is stationary (i.e., shift-invariant) but not quite a product measure:

$$\mu([i_m \dots i_n]) = p_{i_m} \cdot p_{i_m, i_{m+1}} \cdots p_{i_{n-1}, i_n}.$$

Theorem: The Parry measure μ is the unique measure of maximal entropy for a subshift of finite type with aperiodic irreducible transition matrix.

Parry Measure

Proof: In this proof, we will only show that $h_\mu(\sigma) = h_{top}(\sigma) = \log \lambda$, and skip the (more complicated) uniqueness part.

The definitions of the masses of 1-cylinders and 2-cylinders are compatible, because (since v is a right eigenvector)

$$\sum_{j=1}^{N} \mu([ij]) = \sum_{j=1}^{N} p_i p_{i,j} = p_i \sum_{j=1}^{N} \frac{A_{i,j} v_j}{\lambda v_i} = p_i \frac{\lambda v_i}{\lambda v_i} = p_i = \mu([i]).$$

Summing over i, we get $\sum_{i=1}^{N} \mu([i]) = \sum_{i=1}^{N} p_i = \sum_{i=1}^{N} u_i v_i = 1$, due to our scaling.

Parry Measure

To show that μ is shift-invariant, we take any cylinder set $Z = [i_m \dots i_n]$ and compute

$$\begin{aligned}
\mu(\sigma^{-1} Z) &= \sum_{i=1}^{N} \mu([i i_m \dots i_n]) = \sum_{i=1}^{N} \frac{p_i p_{i,i_m}}{p_{i_m}} \mu([i_m \dots i_n]) \\
&= \mu([i_m \dots i_n]) \sum_{i=1}^{N} \frac{u_i v_i \ A_{i,i_m} v_{i_m}}{\lambda v_i \ u_{i_m} v_{i_m}} \\
&= \mu(Z) \sum_{i=1}^{N} \frac{u_i A_{i,i_m}}{\lambda u_{i_m}} = \mu(Z) \frac{\lambda u_{i_m}}{\lambda u_{i_m}} = \mu(Z).
\end{aligned}$$

This invariance carries over to all sets in the σ-algebra $\mathcal{B}$ generated by the cylinder sets.

Parry Measure

Based on the interpretation of conditional probabilities, the identities

$$\sum_{\substack{i_{m+1},\ldots,i_n=1 \\ A_{i_k,i_{k+1}}=1}}^{N} p_{i_m} p_{i_m,i_{m+1}} \quad \cdots \quad p_{i_{n-1},i_n} = p_{i_m}$$
$$\text{and} \tag{1}$$
$$\sum_{\substack{i_m,\ldots,i_{n-1}=1 \\ A_{i_k,i_{k+1}}=1}}^{N} p_{i_m} p_{i_m,i_{m+1}} \quad \cdots \quad p_{i_{n-1},i_n} = p_{i_n}$$

follows because the left hand side indicates the total probability of starting in state i_m and reaching some state after $n-m$ steps, respectively starting at some state and reaching state n after $n-m$ steps.

Parry Measure

To compute $h_\mu(\sigma)$, we will take the partition $\mathcal{P}$ of 1-cylinder sets; this partition is generating, so this restriction is justified by the Kolmogorov-Sinaĭ Theorem (on generating partitions).

$$\begin{aligned}
H_\mu\Big(\bigvee_{k=0}^{n-1} \sigma^{-k}\mathcal{P}\Big) &= -\sum_{\substack{i_0,\dots,i_{n-1}=1\\ A_{i_k,i_{k+1}}=1}}^{N} \mu([i_0\dots i_{n-1}])\log\mu([i_0\dots i_{n-1}]) \\
&= -\sum_{\substack{i_0,\dots,i_{n-1}=1\\ A_{i_k,i_{k+1}}=1}}^{N} p_{i_0}p_{i_0,i_1}\cdots p_{i_{n-1},i_n}\big(\log p_{i_0} \\
&\qquad\qquad + \log p_{i_0,i_1} + \dots + \log p_{i_{n-2},i_{n-1}}\big) \\
&= -\sum_{i_0=1}^{N} p_{i_0}\log p_{i_0} - (n-1)\sum_{i,j=1}^{N} p_i p_{i,j}\log p_{i,j},
\end{aligned}$$

by (1) used repeatedly.

Parry Measure

Hence

$$\begin{aligned} h_\mu(\sigma) &= \lim_{n\to\infty} \frac{1}{n} H_\mu\Big(\bigvee_{k=0}^{n-1} \sigma^{-k}\mathcal{P}\Big) \\ &= -\sum_{i,j=1}^{N} p_i p_{i,j} \log p_{i,j} \\ &= -\sum_{i,j=1}^{N} \frac{u_i A_{i,j} v_j}{\lambda} \left(\log A_{i,j} + \log v_j - \log v_i - \log \lambda\right). \end{aligned}$$

The first term in the brackets is zero because $A_{i,j} \in \{0,1\}$.

Parry Measure

The second term $-\sum_{i,j=1}^{N} \frac{u_i A_{i,j} v_j}{\lambda} \log v_j$ (summing first over i) simplifies to

$$-\sum_{j=1}^{N} \frac{\lambda u_j v_j}{\lambda} \log v_j = -\sum_{j=1}^{N} u_j v_j \log v_j,$$

The third term $\sum_{i,j=1}^{N} \frac{u_i A_{i,j} v_j}{\lambda} \log v_i$ (summing first over j) simplifies to

$$\sum_{i=1}^{N} \frac{u_i \lambda v_i}{\lambda} \log v_i = \sum_{i=1}^{N} u_i v_i \log v_i.$$

Hence these two terms cancel each other.

Parry Measure

The remaining term is

$$\sum_{i,j=1}^{N} \frac{u_i A_{i,j} v_j}{\lambda} \log \lambda = \sum_{i=1}^{N} \frac{u_i \lambda v_i}{\lambda} \log \lambda = \sum_{i=1}^{N} u_i v_i \log \lambda = \log \lambda.$$

This finishes the proof.

Remark: To deal with entries $A_{ij} \in \{2, 3, 4, \dots\}$, we can split states and regain a $0, 1$-matrix.