Lecture 5: Entropy Theory in Metric Spaces

Overview: We discuss a dimension theoretic approach to entropy. Main Advantage: No need for smoothness!

Background on Metric Spaces

Metric Space (X,d): A set X with a non-negative function $d(x,y)$ s.t. $d(x,x)=0$, $d(x,y)=d(y,x)$, and $d(x,z)\le d(x,y)+d(y,z)$ for all $x,y,z\in X$.

- **ϵ-Balls**: $B(x,\epsilon):=\{y\in X: d(x,y)<\epsilon\}$
- **Convergence**: $x_n \xrightarrow[n\to\infty]{} y$ if $d(x_n,y)\longrightarrow 0$.
- **Compactness**: (X,d) is compact if every sequence $\{x_n\}_{n\ge1}$ has a convergent subsequence $x_{n_k}\xrightarrow[k\to\infty]{} y$

Fact: Suppose X is a compact metric space. Then for every $\epsilon>0$, X can be covered by finitely many ϵ-balls.

Proof. Fix $x_1\in X$. If $X\subseteq B(x_1,\epsilon)$, stop.
Otherwise $\exists x_2\in X\setminus B(x_1,\epsilon)$. If $X\subseteq \bigcup_{i=1}^{2} B(x_i,\epsilon)$, stop.
— " — $\exists x_3\in X\setminus \bigcup_{i=1}^{2} B(x_i,\epsilon)$. If $X\subseteq \bigcup_{i=1}^{3} B(x_i,\epsilon)$, stop.
Continue in this way.

At the moment the process stops, we obtain our finite cover. The process <u>must</u> stop, otherwise we obtain $x_1, x_2, x_3, \dots$ s.t.

$$i \neq j \Rightarrow d(x_i, x_j) \geq \epsilon \quad \left(\because x_i \notin \bigcup_{j<i} B(x_j, \epsilon)\right).$$

But such a sequence is paradoxical, because its convergent subsequence $x_{i_k} \to y$ satisfies

$$\epsilon \leq d(x_{i_k}, x_{i_{k+1}}) \leq d(x_{i_k}, y) + d(x_{i_{k+1}}, y) \to 0. \qquad \square$$

<u>How to Coarse-Grain a Compact Metric Space</u>

<u>Aim</u>: Replace (X,d) by a discrete object which represents it "at resolution ϵ."

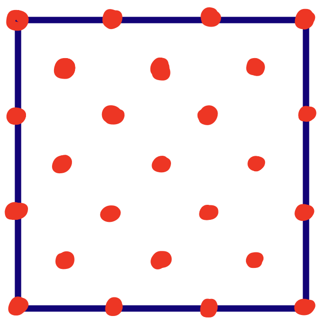
<u>Approach 1</u>: Replace $x \in X$ by the ϵ-ball $B(x,\epsilon)$. Replace X by the <u>smallest cover</u> of X by ϵ-balls.

$$s(X,\epsilon) := \left[\begin{array}{l}\text{cardinality of the smallest} \\ \text{cover of } X \text{ by } \epsilon\text{-balls}\end{array}\right]$$

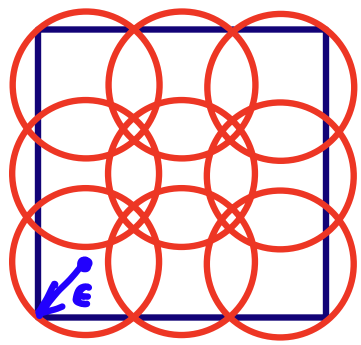
Approach 2: Replace X by a **maximal ϵ-separated** set F_ϵ.

- **ϵ-separated**: $\forall x, y \in F_\epsilon \ (x \neq y \Rightarrow d(x,y) \geq \epsilon)$
- **maximal**: if we add a single point to F_ϵ, it stops being maximal.

$$r(x,\epsilon) := \begin{bmatrix} \text{cardinality of largest} \\ \text{maximal } \epsilon\text{-separated set} \end{bmatrix}$$



ϵ-separated



ϵ-cover

Lemma: Suppose F_t is a maximal t-separated set. Then

$$|F_{2\epsilon}| \leq s(X,\epsilon) \leq |F_\epsilon|.$$

Corollary: $r(X,2\epsilon) \leq s(X,\epsilon) \leq r(X,\epsilon)$.

Proof of the Lemma: Let $\{B(x_1,\epsilon), \dots, B(x_N,\epsilon)\}$ be a cover of X with minimal cardinality.

- If $F_{2\epsilon} = \{y_1, \dots, y_M\}$ then each y_i is contained in some $B(x_j,\epsilon)$ and no two $y_i, y_{i'}$ are in the same $B(x_j,\epsilon)$ ($\because d(y_i,y_{i'}) > 2\epsilon$). We get a one-to-one map $y_i \mapsto x_j$. It follows that $|F_{2\epsilon}| \leq N = s(X,\epsilon)$.

• If $F_\epsilon = \{y_1, \dots, y_\ell\}$ then $\bigcup_{i=1}^{\ell} B(y_i, \epsilon) \supseteq X$, otherwise $\exists z \in X$ s.t. $d(z, y_i) \geq \epsilon$ for all i. But in this case $F_\epsilon \cup \{z\}$ is ϵ-separated, in contradiction to the maximality of F_ϵ.

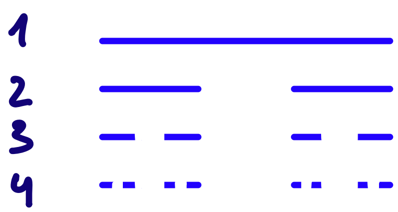
Thus $\{B(y_1,\epsilon), \dots, B(y_\ell, \epsilon)\}$ is a cover of X, and $|F_\epsilon| = \ell \geq s(X,\epsilon)$. $\square$

<u>Example 1</u>: $X = [0,1]^d$

Clearly, there's a maximal ϵ-separated set with $\sim (1/\epsilon)^d$ points. So

$$s(X,\epsilon) \sim (1/\epsilon)^d$$

<u>Example 2</u>: $X =$ Cantor set



Take $F_{1/3^n} := \left\{ \begin{array}{l} \text{endpoints of} \\ n^{\text{th}} \text{ level intervals} \end{array} \right\}$

This is a maximal $1/3^n$-separated set, and $|F_{1/3^n}| = 2^{n+1}$. Thus

$$s(X,\epsilon) \sim (1/\epsilon)^{\log 2 / \log 3}$$

<u>Def$^{\underline{n}}$</u>. The <u>upper box dimension</u>* of X is

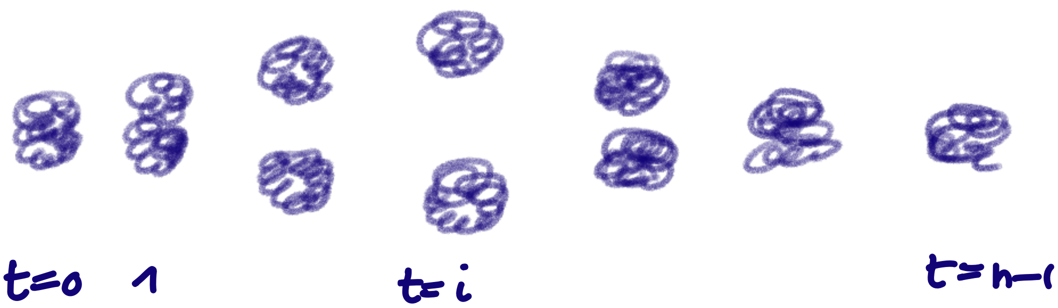
$$\overline{\dim}_B X = \overline{\lim_{\epsilon \to 0^+}} \frac{\log s(X,\epsilon)}{\log(1/\epsilon)}.$$

* aka "Kolmogorov Capacity", "Minkowski Dimension", "Entropy Dimension"

The Topological Entropy

Setup: Suppose T is a continuous map on a compact metric space (X, d).

Insight: Observing $T^i(x)$ $(i = 0, 1, \dots, n-1)$, we can distinguish ϵ-close initial conditions with resolution ϵ, whenever $d(T^i(x), T^i(y)) \geq \epsilon$ for some $0 \leq i \leq n-1$.



Definitions: **Bowen's Metrics**: $d_n(x,y) = \max\limits_{0 \leq i \leq n-1} d(T^i(x), T^i(y))$

- **(n,ϵ)-Bowen Balls**:

$$B(x, n, \epsilon) = \{ y \in X : d(T^i(x), T^i(y)) < \epsilon \ \ (i = 0, 1, \dots, n-1) \}$$

$$s(T, n, \epsilon) := \left[\begin{array}{l} \text{minimal cardinality of a cover of } X \\ \text{by } (n,\epsilon)\text{-balls} \end{array} \right]$$

- **Maximal (n,ϵ)-Separated Sets**: $F = \{ y_1, \dots, y_n \}$ s.t.
  - $\forall i \neq j$ $\ d(T^k(y_i), T^k(y_j)) \geq \epsilon$ for some $0 \leq k \leq n-1$
  - cannot add any point to F without destroying this

$$r(T, n, \epsilon) = \left[\begin{array}{l} \text{maximal cardinality of a maximal} \\ (n,\epsilon)\text{-separated set} \end{array} \right]$$

Def$^{\underline{n}}$. The <u>topological entropy</u> of $T: X \to X$ is

$$h_{top}(T) = \lim_{\epsilon \to 0} \left[\overline{\lim_{n\to\infty}} \frac{1}{n} \log s(T,n,\epsilon) \right]$$

$$\equiv \lim_{\epsilon \to 0} \left[\overline{\lim_{n\to\infty}} \frac{1}{n} \log r(T,n,\epsilon) \right]$$

$r(T,n,2\epsilon) \le s(T,n,\epsilon) \le r(T,n,\epsilon)$
by the lemma applied to Bowen's metric $d_n(\cdot,\cdot)$.

<u>Remark</u>: Why is $h_{top}(f)$ called "topological" and $h_\mu(T)$ is called "metric"? Because mathematicians are bad at naming things.

<u>The Variational Principle</u>: Suppose $T: X \to X$ is a continuous map on a compact metric space X. Then:

$$\boxed{h_{top}(T) = \sup \left\{ h_\mu(T) : \begin{array}{l} \mu \ T\text{-invariant} \\ \text{prob. measure} \end{array} \right\}}$$

<u>Corollary 1</u>: $h_{top}(T) > 0 \implies$ $\exists$ invariant measure with positive entropy.

Thus "positive topological entropy" is a criterion for "deterministic chaos." $\leftarrow$ simulating Bernoulli

Corollary 2: For every invariant prob. measure μ,

$$\boxed{h_\mu(T) \leq h_{top}(T)}$$

and this general bound is optimal

Def^{n}. A measure s.t. $h_\mu(T) = h_{top}(T)$ is called a <u>measure of maximal entropy</u>. Such measures exist sometimes, but not always.

(<u>Newhouse Thm</u>: Every <u>infinitely differentiable</u> map T on a compact smooth manifold has a measure of max entropy.)

Katok Entropy Formula

For general invariant measures, it could happen that $h_\mu(T) < h_{top}(T)$ "because μ occupies a lower-dimensional part of X." Given $0 < \delta < 1$, let

$$s_\mu(T, n, \epsilon, \delta) := \left[\begin{array}{l}\text{cardinality of smallest cover of a} \\ \text{set of measure} > 1-\delta \text{ by } (n,\epsilon)\text{-Bowen balls}\end{array}\right]$$

<u>Katok Entropy Formula</u>: For every $0 < \delta < 1$,

$$\boxed{h_\mu(T) = \lim_{\epsilon \to 0^+} \left[\overline{\lim_{n \to \infty}} \frac{1}{n} \log s_\mu(T, n, \epsilon, \delta) \right]}$$

Example:

- $X = \{ (x_0, x_1, x_2, \dots) : x_i = 0 \text{ or } 1 \}$
- $T : X \to X$ left shift

$$T(x_0, x_1, x_2, \dots) = (x_1, x_2, x_3, \dots)$$

- metric $d(\underline{x}, \underline{y}) = \exp[-\min\{i : x_i \neq y_i\}]$

In this metric $d(\underline{x}, \underline{y}) \leq e^{-k} \iff x_i = y_i$ for $i = 0, \dots, k-1$.
In addition, for $e^{-k-1} \leq \varepsilon < e^{-k}$, the (n, ϵ)-Bowen ball

is $B(\underline{x}, n, \epsilon) = \{ \underline{y} : y_i = x_i \text{ for } i = 0, \dots, n+k-1 \}$.

$$= [x_0, \dots, x_{n+k-1}].$$

Thus $s(X, \epsilon) = \#\{(n+k-1)\text{-cylinders}\} = 2^{n+k-1}$

$$h_{top}(T) = \lim_{k \to \infty} \left(\overline{\lim_{n \to \infty}} \frac{1}{n} \log 2^{n+k-1} \right)$$

$$= \lim_{k \to \infty} [\log 2] = \log 2.$$

(Note that the ϵ doesn't matter).

But if $\mu = \delta_{(1,1,\dots)}$ then it takes just one cylinder to cover most (even all) of the mass of μ. So $h_\mu(T) = 0$.

<u>A less trivial example:</u> Let μ be the Bernoulli measure $B(\frac{1}{3}, \frac{2}{3})$.

Fix $\delta, \kappa > 0$ very small, and let

$$\Omega_\kappa(n) = \left\{ \underline{x} \in X : \begin{array}{l} \#\{0 \le i \le n-1 : x_i = 0\} \le (\frac{1}{3}+\kappa)n \\ \#\{0 \le i \le n-1 : x_i = 1\} \ge (\frac{2}{3}-\kappa)n \end{array} \right\}$$

By the weak law of large numbers (or the ergodic thm), $\exists N_{\delta,\kappa}$ s.t. $\mu(\Omega_\kappa(n)) \ge 1-\delta$ for all $n > N_{\delta,\kappa}$.

To cover $\Omega_\kappa(n)$ by $n+k-1$ cylinders we just need

$$\sum_{\ell=(\frac{1}{3}-\kappa)n}^{(\frac{1}{3}+\kappa)n} \binom{n+k-1}{\ell} \sim \exp\Big[n\big(\underbrace{H(\tfrac{1}{3}, \tfrac{2}{3})}_{-\frac{1}{3}\log\frac{1}{3} - \frac{2}{3}\log\frac{2}{3}} + \varepsilon(\kappa)\big)\Big]$$

cylinders, where $\varepsilon(\kappa) \xrightarrow[\kappa \to 0]{} 0$.

Thus

$$\lim_{\epsilon \to 0} \overline{\lim_{n \to \infty}} \frac{1}{n} \log s_\mu(T, n, \epsilon, \delta) = H(\tfrac{1}{3}, \tfrac{2}{3})$$

Subshifts of Finite Type (SFT)

Setup: Suppose G is a connected aperiodic* finite directed graph with set of vertices S, and transition matrix

$$A = (t_{ab})_{S\times S}, \quad t_{ab} = \begin{cases} 1 & a\to b \\ 0 & a \not\to b \end{cases}$$

- $X := \{(x_0, x_1, x_2, \dots) : x_i \in S,\ t_{x_i x_{i+1}} = 1 \text{ for all } i\}$
- $T : X \to X$ is the left shift
- $d(\underline{x}, \underline{y}) = \exp\left[-\min\{i : x_i \neq y_i\}\right]$

Again, for $e^{-k-1} \le \epsilon < e^{-k}$, every (n,ϵ)-Bowen ball is a $(n+k-1)$-cylinder

$$B(\underline{x}, n, \epsilon) = [x_0, x_1, \dots, x_{n+k-1}]$$

Thus $s(X, n, \epsilon) = \#\left\{\begin{matrix}\text{non-empty cylinders} \\ \text{of length } n+k-1\end{matrix}\right\}$

$$= \sum_{x_0, \dots, x_{n+k-1}} t_{x_0 x_1} t_{x_1 x_2} \cdot \ldots \cdot t_{x_{n+k-2} x_{n+k-1}} \qquad (*)$$

(because the summand is one when $x_0 \to x_1 \to \dots \to x_{n+k-1}$ is a legitimate path, and zero otherwise).

---

* aperiodic : $\gcd\left\{n : \begin{matrix}a \text{ connects to} \\ \text{itself in } n \text{ steps}\end{matrix}\right\} = 1$

To continue with the calculation we note that the powers of A are given by

$$A^n = \left(t^{(n)}_{ab}\right)_{S\times S}, \quad t^{(n)}_{ab} = \sum_{\xi_1 \ldots \xi_{n-1} \in S} t_{a\xi_1} t_{\xi_1\xi_2} \cdots t_{\xi_{n-1}b}$$

Thus

$$\boxed{S(X,n,\epsilon) = \text{Sum of the } |S|\times|S| \text{ entries of } A^{n+k-1}}$$

A is a positive matrix. By the Perron-Frobenius theorem (and the assumption that G is connected),

$$\forall a,b \quad \overline{\lim_{n\to\infty}} \frac{1}{n} \log (A^n)_{ab} = \log \lambda$$

where λ = maximal positive e.v. of A. Thus,

$$\boxed{h_{top}(\tau) = \log \lambda}$$

Approximation By Periodic Points on SFT:

Let $\mathrm{Fix}(T^n) := \{ \underline{x} \in X : T^n(\underline{x}) = \underline{x} \}$.

Each $\underline{x} \in \mathrm{Fix}(T^n)$ has the form

$$(\underbrace{a, x_1, x_2, \ldots, x_{n-1}, a}_{\text{determines } \underline{x}}, x_1, x_2, \ldots, x_{n-1}, a, \ldots)$$

Thus the same calculation as before gives

$$Z_n := |\mathrm{Fix}(T^n)| = \sum_{a \in S} (A^n)_{aa} = \mathrm{tr}(A^n) \sim \lambda^n .$$

<u>Corollary</u>: $\mathrm{Fix}(T^n)$ is a (non-maximal) (n,ϵ) – separated set of cardinality $\sim \exp[n\, h_{top}(T)]$

<u>Thm (Bowen)</u>: Let $\mu_n = \frac{1}{Z_n} \sum_{\underline{x} \in \mathrm{Fix}(T^n)} \delta_{\underline{x}}$. This is a T-invariant atomic prob measure, and

$$\mu_n \xrightarrow[n \to \infty]{} \mu = \text{measure of maximal entropy}$$

More generally, suppose $U : X \to \mathbb{R}$ is a function s.t.

$$|U(\underline{x}) - U(\underline{y})| < \text{const.} \exp\left[-\gamma \cdot \min\{i : x_i \neq y_i\}\right]$$

Let $U_n(\underline{x}) = U(\underline{x}) + U(T(\underline{x})) + \cdots + U(T^{n-1}(\underline{x}))$.

Fix an "inverse temperature" $\beta > 0$ and set

$$\mu_\varphi^{(n)} = \frac{1}{Z_n(\varphi)} \sum_{\underline{x} \in \mathrm{Fix}(T^n)} e^{-\beta U_n(\underline{x})} \delta_{\underline{x}}, \quad Z_n(\varphi) = \sum_{\underline{x} \in \mathrm{Fix}(T^n)} e^{-\beta U_n(\underline{x})}$$

Thm (Bowen, Ruelle): $\mu_\varphi^{(n)} \xrightarrow[n \to \infty]{} \mu_\varphi$ where μ_φ is the unique measure which minimizes the "free energy"

$$\underbrace{\int U \, d\mu - \frac{1}{\beta} h_\mu(T)}_{\text{"energy} - \text{temperature} \times \text{entropy"}}.$$

The value of the minimized free energy is

$$-\frac{1}{\beta} \times \underbrace{\lim_{n \to \infty} \frac{1}{n} \log Z_n(\varphi)}_{\text{"pressure of } \varphi\text{"}}$$

There are additional "thermodynamic" results like this, including linear response formulas etc.