Lecture 26: Nov. 14

SMB Theorem
Let $T$ be an ergodic MPT and $\alpha$ a (finite, measurable) partition. Then

$$\frac{1}{n} I_{\alpha_0^n} \to h(T, \alpha) = \int I_{\alpha|\alpha_1^\infty} = H(\alpha|\alpha_1^\infty) \text{ a.e. and in } L^1$$

Shannon-1948: irreducible stationary Markov chains; convergence in probability
McMillan-1953: ergodic stationary processes; convergence in $L^1$
Breiman-1957: ergodic stationary processes; convergence a.e.

Proof:
Recall:
— $\alpha(x)$ defined by $x \in A_{\alpha(x)}$.
— $I_{\alpha}(x) = -\log \mu(A_{\alpha(x)})$
— $I_{\alpha|\beta}(x) = -\log \mu(A_{\alpha(x)}|B_{\beta(x)})$

Lemma 1: $I_{\alpha \vee \beta} = I_{\beta} + I_{\alpha|\beta}$

Proof:

$I_{\alpha \vee \beta}(x) = -\log \mu(A_{\alpha(x)} \cap B_{\beta(x)}) = -\log \frac{\mu(A_{\alpha(x)} \cap B_{\beta(x)})}{\mu(B_{\beta(x)})} - \log(\mu(B_{\beta(x)}))$

$$= I_{\alpha|\beta}(x) + I_{\beta}(x)$$

Lemma 2: $I_{\alpha} \circ T = I_{T^{-1}(\alpha)}$

Proof:

$I_{\alpha} \circ T(x) = -\log \mu(A_{\alpha(T(x))}) = -\log \mu(A_{T^{-1}(\alpha)(x)}) = I_{T^{-1}(\alpha)}(x)$

Thus, $I_{\alpha \vee T^{-1}\alpha} = I_{T^{-1}(\alpha)} + I_{\alpha|T^{-1}\alpha} = I_{\alpha|T^{-1}\alpha} + I_{\alpha} \circ T$
By induction:

\[ I_{\alpha_0}^n = I_{\alpha|\alpha_1}^n + I_{\alpha|\alpha_1}^{n-1} \circ T + \ldots + I_{\alpha|\alpha_1} T^{n-1} + I_{\alpha} T^n = \sum_{k=0}^{n} I_{\alpha|\alpha_1}^{n-k} \circ T^k \]

\[ \frac{1}{n} I_{\alpha_0}^n = \frac{1}{n} \sum_{k=0}^{n} I_{\alpha|\alpha_1}^{n-k} \circ T^k \]

\[ = \frac{1}{n} \sum_{k=0}^{n} I_{\alpha|\alpha_1}^\infty \circ T^k + \frac{1}{n} \sum_{k=0}^{n} \left( I_{\alpha|\alpha_1}^{n-k} \circ T^k - I_{\alpha|\alpha_1}^\infty \circ T^k \right) \]

\[ = R_n + S_n \]

Since \( I_{\alpha|\alpha_1}^\infty \in L^1 \),

\[ \left( \int I_{\alpha|\alpha_1}^\infty d\mu = H(\alpha|\alpha_1^\infty) \right) \]

we can apply Ergodic Theorem to get

\[ R_n \to H(\alpha|\alpha_1^\infty) \text{ a.e. and in } L^1 \]

So, it suffices to show that \( S_n \to 0 \text{ a.e. and in } L^1 \).

(Note: In IID case, \( S_n = 0 \))

For \( L^1 \) convergence, observe:

\[ \int \left| S_n \right| d\mu \leq \frac{1}{n} \sum_{k=0}^{n} \int \left| I_{\alpha|\alpha_1}^{n-k} \circ T^k - I_{\alpha|\alpha_1}^\infty \circ T^k \right| \]

\[ = \frac{1}{n} \sum_{k=0}^{n} \int \left| I_{\alpha|\alpha_1}^{n-k} - I_{\alpha|\alpha_1}^\infty \right| \]

which converges to 0 by \( L^1 \) convergence in theorem on continuity of conditional information.
(in fact, we only need the Cesaro convergence in $L^1$ from that theorem).

Remains to prove:

$$S_n \to 0 \text{ a.e.}$$

(the interesting part).

Proof: Let $h_N = \sup_{n \geq N} |I_{\alpha|\alpha_1^n} - I_{\alpha|\alpha_1^\infty}|$.

By theorem on continuity of conditional entropy,

$$h_N \to 0 \text{ a.e.}$$

and

$$h_N \leq 2f^*$$

where $f^* = \sup_n I_{\alpha|\alpha_1^n} \in L^1$.

By DCT,

$$\int h_N d\mu \to 0$$

Fix $N$. Since for $k \leq n - N$, $|I_{\alpha|\alpha_1^{n-k}} - I_{\alpha|\alpha_1^\infty}| \leq h_N$

$$\limsup |S_n| \leq \limsup \left( \frac{1}{n} \sum_{k=0}^{n-N} h_N \circ T^k + \frac{1}{n} \sum_{k=n-N+1}^{n} h_1 \circ T^k \right) = \limsup (U_n + V_n)$$

Let $n \to \infty$.

By ergodic theorem applied to $h_N$, $U_n \to \int h_N d\mu$ a.e.

And $V_n$ consists of the last $N$ terms of $\frac{1}{n} \sum_{k=0}^{n} h_1 \circ T^k$, which converges by the ergodic theorem. So $V_n \to 0$ a.e. Thus,

$$\limsup |S_n| \leq \int h_N d\mu.$$

Let $N \to \infty$. Thus, $\limsup |S_n| = 0$ a.e. QED

Recall:

$$h(T, \alpha) = \lim_{n} (1/n) H(\vee_{i=0}^{n} \alpha_i)$$
\[ h(T) = \sup_\alpha h(T, \alpha) \]

Fact: \( h(T) \) is an isomorphism invariant, i.e. if \( T \) and \( S \) are isomorphic, then \( h(T) = h(S) \).

Proof: An isomorphism \( \phi : M \to N \) sets up a bijection between finite measurable partitions of \( M \) and those of \( N \):

\[ \alpha \mapsto \phi(\alpha), \quad h(T, \alpha) = h(S, \phi(\alpha)) \]

Prop 1: Let \( T \) be an MPT, \( \alpha \) a (finite, measurable) partition and \( \alpha_i = T^{-i}(\alpha) \). Then for all \( m \),

\[ h(T, \alpha) = h(T, \bigvee_{i=0}^m \alpha_i) \]

Proof:

\[ h(T, \bigvee_{i=0}^m \alpha_i) = \lim_n (1/n) H(\bigvee_{i=0}^{n+m} \alpha_i) = \lim_n (1/n) H(\bigvee_{i=0}^n \alpha_i) = h(T, \alpha) \]

Prop 2: Let \( T \) be an MPT, \( \alpha, \beta \) (finite, measurable) partitions. Then

\[ h(T, \beta) \leq h(T, \alpha) + H(\beta | \alpha) \]

Proof:

\[ H(\bigvee_{j=0}^n \beta_j) \leq H( (\bigvee_{i=0}^n \beta_i) \lor (\bigvee_{j=0}^n \alpha_i) ) = H(\bigvee_{i=0}^n \alpha_i) + H(\bigvee_{j=0}^n \beta_j | \bigvee_{i=0}^n \alpha_i) \]

\[ H(\bigvee_{j=0}^n \beta_j | \bigvee_{i=0}^n \alpha_i) \leq \sum_{j=0}^n H(\beta_j | \bigvee_{i=0}^n \alpha_i) \leq \sum_{j=0}^n H(\beta_j | \alpha_j) \]

Thus,

\[ H(\bigvee_{j=0}^n \beta_j) \leq H(\bigvee_{i=0}^n \alpha_i) + \sum_{j=0}^n H(\beta_j | \alpha_j) = H(\bigvee_{i=0}^n \alpha_i) + nH(\beta | \alpha) \]

Thus, \( (1/n)H(\bigvee_{i=0}^n \beta_i) \leq (1/n)H(\bigvee_{i=0}^n \alpha_i) + H(\beta | \alpha) \).

Let \( n \to \infty \).
Lecture 27: Nov. 16

Recall:
Prop 1: Let $T$ be an MPT, $\alpha$ a (finite, measurable) partition and $\alpha_i = T^{-i}(\alpha)$. Then for all $m$,

$$h(T, \alpha) = h(T, \vee_{i=0}^{m} \alpha_i)$$

Prop 2: Let $T$ be an MPT, $\alpha, \beta$ (finite, measurable) partitions. Then

$$h(T, \beta) \leq h(T, \alpha) + H(\beta | \alpha)$$

Notation: $\bigvee_{i=0}^{\infty} \alpha_i = \sigma(\bigcup_{i=0}^{\infty} \alpha_i)$

Defn: A 1-sided generator for a MPT $T$ on $(\mathcal{M}, \mathcal{A}, \mu)$ is a (finite, measurable) partition $\alpha$ such that

$$\bigvee_{i=0}^{\infty} \alpha_i = \mathcal{A}$$

Theorem (Kolmogorov-Sinai, 1958-1959). Let $\alpha$ be a 1-sided generator for an MPT $T$. Then

$$h(T) = h(T, \alpha)$$

(and so we have some hope of computing $h(T)$).

Kolmogorov - 1958: slightly different notion, for restricted class of processes; not quite right; but in 1957 lecture gave correct defn. for iid

Sinai - 1959: correct definition, completely general

Proof: Enough to show that for any (finite, measurable) partition $\beta$, we have $h(T, \beta) \leq h(T, \alpha)$.

By Prop 2 applied to $\bigvee_{i=0}^{n} \alpha_i$ and then Prop. 1, for all $n$,

$$h(T, \beta) \leq h(T, \bigvee_{i=0}^{n} \alpha_i) + H(\beta | \bigvee_{i=0}^{n} \alpha_i) = h(T, \alpha) + H(\beta | \bigvee_{i=0}^{n} \alpha_i)$$
But $\sigma(\bigvee_{i=0}^{n}\alpha_i) \uparrow \mathcal{A}$. By continuity of entropy

$$H(\beta | \bigvee_{i=0}^{n}\alpha_i) \to H(\beta | \mathcal{A})$$

But since $\beta \subset \mathcal{A}$, we have, by Property 3, $H(\beta | \mathcal{A}) = 0$. Thus, $h(T, \beta) \leq h(T, \alpha)$.

Example: If $T$ is the MPT corresponding to a one-sided stationary process $\overline{X}$, then

$$h(T) = h(\overline{X})$$

Proof: Recall that $h(T, \alpha) = h(\overline{X})$ where $\alpha$ is the partition $A_i = \{ x \in F^{\mathbb{Z}^+} : x_0 = i \}$.

Clearly $\alpha$ is a 1-sided generator.

So,

$$h(T_{iid^+ (p)}) = H(p)$$

$$h(T_{iid^+ (1/2,1/2)}) = \log 2, \quad h(T_{iid^+ (1/3,1/3,1/3)}) = \log 3$$

$$h(T_{Doubling}) = \log 2.$$ Show one-sided generator for doubling map.

For Markov $P$ with stationary vector $\pi$:

$$h(T_{P^+}) = \sum_{ij} \pi_i \log p_{ij}$$

Theorem: Let $T$ be an invertible MPT. If $T$ has a 2-sided generator $\alpha$, i.e.,

$$\bigvee_{i=-\infty}^{\infty} \alpha_i = \mathcal{A}$$

then $h(T) = h(T, \alpha)$.

Proof: In Prop. 1, replace $\bigvee_{i=0}^{m} \alpha_i$ with $\bigvee_{i=-m}^{m} \alpha_i$
Example: If $T$ is the MPT corresponding to a two-sided stationary process $X$, then
\[ h(T) = h(X) \]
So,
\[ h(T_{\text{iid}(p)}) = H(p) \]
\[ h(T_{\text{iid}(1/2,1/2)}) = \log 2, \quad h(T_{\text{iid}(1/3,1/3,1/3)}) = \log 3 \]
Thus, these MPT’s are not isomorphic.

Existence of two-sided generators (Krieger, 1970): if $T$ is an invertible ergodic MPT and $h(T)$ is finite, then it has a two-sided generator with at most $\lfloor e^{h(T)} \rfloor + 1$ sets.

Note: this bound is necessary:

Proof: If $\alpha$ is a two-sided generator, then
\[ h(T) = h(T, \alpha) = H(\alpha | \alpha_1^\infty) \leq H(\alpha) \leq \log \alpha \]
Thus
\[ e^{h(T)} \leq |\alpha| \]
Thus,
\[ \lfloor e^{h(T)} \rfloor \leq |\alpha| \]
LHS is $\lfloor e^{h(T)} \rfloor + 1$ unless $e^{h(T)}$ is an integer.

But we will see an example where $e^{h(T)}$ is an integer $k$ and there is no generator with $k$ sets; namely $h(T) = 0$ (rotation of the circle).

Prop: $h(T, \alpha) = 0$ iff $\alpha \subset \alpha_1^\infty$.
Proof: $h(T, \alpha) = H(\alpha | \alpha_1^\infty)$.
But Property 3 says that RHS is 0 iff $\alpha \subset \alpha_1^\infty$.

Interpretation in terms of predictability (determinism).
Lecture 28: Nov. 18

Recall:
Prop: \( h(T, \alpha) = 0 \) iff \( \alpha \subset \alpha_1^\infty \).

Interpretation in terms of predictability (determinism).

Cor: \( h(T) = 0 \) iff \( \alpha \subset \alpha_1^\infty \) for all \( \alpha \).

Prop: If \( T \) is invertible and has a 1-sided generator, then \( h(T) = 0 \).

Proof:
\[
\alpha \subset \mathcal{A} = T^{-1}(\mathcal{A}) = T^{-1}(\alpha_0^\infty) = \alpha_1^\infty
\]

So, we do not generally expect to have one-sided generators for invertible MPT’s.

Example 1:
For Baker; \( h(T) = \log 2 \), and since Baker is invertible, it has no 1-sided generator. Show 2-sided Generator

Example 2:
Irrational rotation (where \( T_r \) is rotation of circle by angle \( r \) set \( r/ (2\pi) \not\in Q \)).

Claim: \( h(T_r) = 0 \)

Let \( a = e^{r2\pi i} \).

\( T_r(z) = az \)

Partition, \( \alpha = \{A_{up}, A_{down}\} \): upper and lower semi-circles bounded by 1 and -1. Then \( T^{-n}(\alpha) \) are semi-circles bounded by \( a^{-n} \) and \( -a^{-n} \). But \( \{a^{-n}\}_{n \in \mathbb{Z}^+} \) is dense (note: \( -r \) is just as irrational as \( r \)).

Thus, all arcs belongs to \( \alpha_0^\infty \). Thus, \( \alpha_0^\infty = \mathcal{A} \). QED

Note: \( T_r \) has a two-sided generator with 2 sets, but not with 1 set. Thus, in this case, we need more than \( \lceil e^{h(T)} \rceil \) sets in a two-sided generator (compare with Krieger generator theorem).
Example: For a rational rotation $T_r$, i.e., $r/(2\pi) = p/q \in Q$, $h(T_r) = 0$.

Proof: $T^q = I$. For any partition $\alpha$, $\alpha_0^{nq-1} = \alpha_0^{q-1}$. Thus,

$$h(T) = \lim_{n \to \infty} \left( \frac{1}{nq} H(\alpha_0^{nq-1}) \right) = \lim_{n \to \infty} \left( \frac{1}{nq} H(\alpha_0^{q-1}) \right) \to 0$$

Note: there is no finite generator for $T_r$, $r$ rational (non-ergodic, compare with Krieger generator theorem)

Note: More generally, any rotation on a compact group has zero entropy.

Defn: Kolmogorov (K) MPT

Invertible and:

There exists a sub-sigma-algebra $\mathcal{B}$ s.t.

1. $\mathcal{B} \subset T(\mathcal{B})$
2. $\bigvee_{n=0}^{\infty} T^n(\mathcal{B}) = \mathcal{A}$
3. $\bigcap_{n} T^{-n}(\mathcal{B}) = \{\emptyset, M\}$ (mod 0)

Think of $\mathcal{B}$ as $\alpha_0^\infty$.

i.e., satisfies the Kolmogorov 0-1 law.

Deep Theorem (Kolmogorov, Rohlin): Let $T$ be invertible. Then $T$ is K iff for all non-trivial $\alpha$, $h(T, \alpha) > 0$.

K-MPT’s are extreme opposites of zero entropy MPT’s:

For all $\alpha$, $h(T, \alpha) > 0$ —vs— For all $\alpha$, $h(T, \alpha) = 0$

At one time, it was conjectured that every invertible MPT is a direct product of a K MPT and a zero entropy MPT. This is false, but there is a weaker notion of product for which this is true.

Deep Theorem (Ornstein 1969) $T_{iid(p)} \sim T_{iid(q)}$ iff $H(p) = H(q)$.

Defn: An MPT is B (Bernoulli) if it isomorphic to $T_{iid(p)}$ for some $p$. 


Note that $B \subseteq K$ by virtue of the Kolmogorov 0-1 Law. At one time, it was conjectured that $B = K$, but this is false.

Lecture 29: Nov. 21

– Posted Exercises 3 and 4 on website
– Student talks: reserve Math 126 for every weekday Dec. 1 - 15, 2-3:30
– Teaching evaluations, enrolled students.
– Problem of isomorphism of MPT’s is unsolved, except for special classes
– Lebesgue spaces are quite general

Defn: A topological dynamical system (TDS) is a continuous map on a compact metric space $T : M \to M$.

Defn: An invertible top dyn sys (ITDS) is a homeomorphism $T : M \to M$ (equivalently, a bijective continuous map, since $M$ is compact.

Orbit:
—- If non-invertible, $\{T^n(x)\}_{n \geq 0}$
—- If invertible, $\{T^n(x)\}_{n \in \mathbb{Z}}$

Iterate dynamics, just as in ergodic theory.

Analogue of ergodicity:

Defn: Topologically Transitive: for all nonempty open $A, B$, there exists $n > 0$ s.t. $T^{-n}(A) \cap B \neq \emptyset$.

Fact: Top. transitive iff there exists a point with a forward dense orbit, i.e., $\{T^n(x)\}_{n \geq 0}$ is dense in $M$; sometimes we say that $x$ is a dense orbit (equivalently, a dense $G_\delta$ of dense orbits).
Proof:
Only if: Let \( \{A_m\} \) be basis for topology. Then
\[
\bigcap_{m=1}^{\infty} (\bigcup_{n=1}^{\infty} T^{-n}(A_m))
\]
is a countable intersection of dense open sets.

If: Let \( x \in V \) have forward dense orbit. Then \( x \in T^{-n}(A) \cap B \)
for some \( n \geq 0 \).

Analogue of mixing:

Defn: Topologically mixing: for all nonempty open \( A, B \), there exists \( N > 0 \) s.t. for all \( n \geq N \), \( T^{-n}(A) \cap B \neq \emptyset \).

Note: suffices to check Top. mixing and Top. transitive on a basis of open sets.

Note that if there is a \( T \)-invariant Borel probability measure \( \mu \) which is strictly positive on all nonempty open sets, then:

— Ergodicity (wrt \( \mu \)) implies Top. Transitive.

— Mixing (wrt \( \mu \)) implies Top. Mixing.

Defn: Periodic point: \( T^p(x) = x \) (\( x \) is periodic with period \( p \); sometimes we say that the orbit is periodic).

e.g., \( p = 1 \) means fixed point.

(no real analogue in ergodic theory)

Examples:

1. circle rotation: \( z \mapsto az, \ a = e^{ir}, \ r \in [0, 2\pi) \)
   — \( r/(2\pi) \in Q \): all orbits are periodic; not top. trans.; therefore, not top, mixing
   — \( r/(2\pi) \notin Q \): all orbits are dense; no periodic orbits; top. trans., since ergodic w.r.t Lebesgue; not top. mixing
2. doubling map on circle: $z \mapsto z^2 \ (e^{i\theta} \mapsto e^{2i\theta})$.
   this is similar to the doubling map on $[0, 1]: x \mapsto 2x \mod 1$, but rescaled.
   top. trans. and top. mixing, since the doublign map on $[0, 1]$ is mixing w.r.t. Lebesgue
   will see that it has a dense set of periodic points.
   (so, a rich mixture of periodic and "ergodic" behaviour).

3. Smale horseshoe with caps (a continuous analogue of Baker’s transformation).

   $T : M \to M$, where $M$ is union of square $S$ and left and right caps.

   $T$ is not onto, but can be extended to a self-homeomorphism of $\mathbb{R}^2$.

   Dynamics:

   Let $\Omega = \cap_{n \in \mathbb{Z}} T^n(\text{ red } \cup \text{ blue })$.

   $T$ has a fixed point $x_0$ in Left Cap.

   Claim: If $x \notin \Omega$, then $\lim_{n \to \infty} T^n(x) = x_0$.

   Proof: black $\to$ right cap $\to$ left cap $\to x_0$.

   If $x \notin \Omega$, then for some $m \in \mathbb{Z}$, then
   $T^m(x) \in \text{ left cap } \cup \text{ black } \cup \text{ right cap}$, and so $T^{n+m}(x) \to x_0$.

   QED

   Interesting dynamics is on $\Omega$.

   Looks like Baker, except it is continuous.

   Will see $T|_\Omega$ is top. mixing, therefore also top. trans., and also show that it has dense periodic points.
Lecture 30: Nov. 23

Example: Full two-sided Shift
Let $F$ be a finite set. Let $M = F^\mathbb{Z}$. Let $T = \sigma$, the left shift.

Metric: $d(x, y) = 2^{-k}$ where $k$ is maximal such that $x[−k,k] = y[−k,k]$.

if $x = y$, set $d(x, y) = 0$; if $x_0 \neq y_0$, set $d(x, y) = 2$; this is a metric.

meaning: $x$ and $y$ are close if they agree in a large central block.

the topology generated is the product topology, which is the same as the topology generated by cylinder sets.

note that agreement in a large central block is equivalent, in the sense of generating the topology, to agreement in a large block (central or otherwise).

Top, mixing because for $A = A_{a_0...a_{Σ−1}}$ and $B = A_{b_0...b_{u−1}}$

$\sigma^{-n}(A) \cap B \neq \emptyset$ for $n \geq u$.

Thus, also top. transitive.

Periodic points: periodic sequences.

Periodic points are dense: Given $x \in M$, define $y = (x[−n,n])^\infty$. Then $y$ is periodic and $d(x, y) \leq 2^{-n}$,

Note: Same holds for one-sided shift.

Defn: Topological conjugacy: $T \sim S$ if there exists homeomorphism $\phi : M \to N$ s.t. $\phi \circ T = S \circ \phi$.

(in this case, homeo is equivalent to bijective and continuous).

Invariants of top. conjugacy:
a. Top. Trans.
b. Top, Mixing

c. Density of Periodic points

Claim: Full two sided shift on $F = \{R, B\}$ is conjugate to horseshoe, restricted to $\Omega$,

Define $\phi : \Omega \rightarrow F^\mathbb{Z}$. If $x \in \Omega$, define $\phi(x) = z \in F^\mathbb{Z}$, where

$z_n = R$ if $T^n(x) \in \text{red}$

$z_n = B$ if $T^n(x) \in \text{blue}$

Proof of: $\phi \circ T = \sigma \circ \phi$:

$\phi(Tx)_n = R$ if $T^{n+1}(x) \in \text{red}$

$\phi(Tx)_n = B$ if $T^{n+1}(x) \in \text{blue}$

Thus, $\phi(Tx)_n = \phi(x)_{n+1} = (\sigma(\phi(x)))_n$. QED

Lemma: For all $N$ and all sequences $z_{-N} \ldots z_0 \ldots z_N \in \{B, R\}^{2N+1}$, the set $S(z_{-N} \ldots z_0 \ldots z_N) = \cap_{n=-N}^N T^{-n}(z_n)$

- is compact
- nonempty
- has diameter $\approx 2^{-N}$

Draw pictures of $S(3\text{-blocks})$

Proof of 1-1: Let $x \in \Omega$ and $z = \phi(x)$.

$\cap_{n=-\infty}^{\infty} T^{-n}(z_n) = \{x\}$

So, $x$ can be uniquely recovered from $T(x)$.

Proof of onto: Let $z \in \{B, R\}^\mathbb{Z}$.

Then $\cap_{n=-N}^N T^{-n}(z_n)$ is a nested decreasing sequence of compact sets and thus its infinite intersection is nonempty. In fact, since
diameters decrease to zero, its intersection is a single point \( x \in \Omega \),
and \( z = \phi(x) \).

Proof that \( \phi^{-1} \) is continuous:
If \( d(z, z') \) is small, then for a large \( N \),
\[ z_N \ldots z_0 \ldots z_N = z'_N \ldots z'_0 \ldots z'_N. \]
Now, \( \phi^{-1}(z) \in S(z_N \ldots z_0 \ldots z_n) \) and \( \phi^{-1}(z') \in S(z'_N \ldots z'_0 \ldots z'_n) \),
and so
\[ \rho(\phi^{-1}(z), \phi^{-1}(z')) < 2^{-N}. \]

Note: \( \Omega \) is a Cantor set.

Defn: Topological factor: \( T \downarrow S \) if there exists onto, continuous
\( \phi : M \to N \) s.t. \( \phi \circ T = S \circ \phi \).

If \( T \) has one of property a, b or c, and \( S \) is a topological factor of
\( T \) (e.g., conjugacy), then so does \( S \).

Example:
Claim: Full one sided shift on \( F = \{0, 1\} \) factors onto doubling
map on circle.
\[ \phi(x_0x_1 \ldots) = e^{2\pi i (x_0x_1 \ldots)} \]
\[ \phi \circ \sigma = T \circ \phi: \]
\[ \phi \circ \sigma(x_0x_1 \ldots) = \phi(x_1x_2 \ldots) = e^{2\pi i (x_1x_2 \ldots)} = e^{2\pi i 2(x_0x_1 \ldots)} = (\phi(x_0x_1 \ldots))^2 \]
\( \phi \) is continuous: if \( x_{[0,n]} = y_{[0,n]} \), then \( .x_0x_1 \ldots \) and \( .y_0y_1 \ldots \)
are close.

onto: every angle is of the form \( 2\pi (.x_0x_1 \ldots) \),

Note: full one sided shift and doubling map are not top. conjugate
because \( F^\mathbb{Z} \) and circle are not homoemorphic (e.g., disconnected vs. connected).
Thus, both doubling map on circle and horseshoe map on $\Omega$ are top. mixing, therefore top. transitive, and have dense periodic points.

Lecture 31: Nov. 25

Another Example:

Defn: Let $C$ be a square 0-1 matrix (say $m \times m$).

Let $F = \{0, \ldots, m - 1\}$.

Let $M_C = \{x \in F^\mathbb{Z} : C_{x_i, x_{i+1}} = 1 \text{ for all } i \in \mathbb{Z}\}$

Consider graph of $C$

$\sigma|_{M_C}$ is called a Topological Markov Chain (TMC).

Notation: $M_C$ refers to the set and to $\sigma|_{C_A}$.

Examples:

1.

$$C = \begin{bmatrix} 1 & 1 \\ 1 & 0 \end{bmatrix}$$

Draw graph

Then $M_C$ is the set of all bi-infinite binary sequences such that 11 is forbidden.

2.

$$C = \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}$$

Then $M_C$ is a full shift.

WMA: $C$ has no all zero rows and no all zero columns.

Recall that we defined irreducibility and primitivity only for stochastic matrices, but the same concepts apply to nonnegative matrices.
Prop:
1. $\sigma$ on $M_C$ is top. transitive iff $C$ is irreducible.
2. $\sigma$ on $M_C$ is top. mixing iff $C$ is primitive.

Proof is nearly identical to proofs of ergodicity and mixing for irreducible and primitive Markov chains.

Reasons for our interest:

a. TMC’s model many interesting smooth dynamical systems, e.g. partial horseshoe.

b. Prop: TFAE
   1. $M$ is the support of a stationary Markov chain (i.e., the smallest closed subset of $F^\mathbb{Z}$ that has measure one).
   2. $M = M_C$ is a TMC which is the disjoint union of finitely many irreducible TMC’s,
   3. $M = M_C$ has dense periodic points

Example:

\[
P = \begin{bmatrix} 1/3 & 2/3 \\ 1 & 0 \end{bmatrix}
\]

$\pi = [3/5, 2/5]$.

TOPOLOGICAL ENTROPY:

For simplicity, assume Invertible!

Define: for open covers $\alpha, \beta$ (possibly infinite!)

$\alpha \lor \beta = \{A_i \cap B_j : \text{nonempty}\}$

$T^{-1}(\alpha) = \{T^{-1}(A_i)\}$

$\alpha_m^n = T^{-m}(\alpha) \lor \cdots \lor T^{-n}(\alpha)$.

$\alpha \preceq \beta$: every element of $\beta$ is a subset of an element of $\alpha$
Recall: every open cover of a compact space has a finite subcover.

Defn: Let $M$ be compact metric space. Let $\alpha$ be an open cover of $M$. Let $N(\alpha)$ be size of smallest subcover. Let $H(\alpha) = \log N(\alpha)$.

Easy properties:
- $H(\alpha \vee \beta) \leq H(\alpha) + H(\beta)$
- If $\alpha \preceq \beta$, then $H(\alpha) \leq H(\beta)$.
- $H(\alpha) = H(T^{-1}(\alpha))$.

Defn:
$$h(T, \alpha) = \lim_{n \to \infty} (1/n)H(\alpha_0^{n-1})$$

Prop: Limit exists. Proof:
Let $a_n = H(\alpha_0^{n-1})$. Then $a_{n+m} \leq a_n + a_m$ (is subadditive).

Proof:
$$a_0^{n+m-1} = \alpha_0^{n-1} \vee T^{-n}(\alpha_0^{m-1}).$$

Thus,
$$a_{n+m} = H(\alpha_0^{n+m-1}) \leq H(\alpha_0^{n-1}) + H(\alpha_0^{m-1}) = a_n + a_m$$

QED

Fekete’s Lemma: For a subadditive sequence $a_n$,
$$\lim_{n} (1/n)a_n = \inf_{n} (1/n)a_n$$

Rough idea of proof: $a_{kn} \leq ka_n$ and so $\frac{a_{kn}}{kn} \leq \frac{a_n}{n}$.

So, if limit exists, then it must be the inf.

Defn:
$$h(T) = \sup_{\alpha} h(T, \alpha)$$
Notation: If $T$ is a ITDS and and also has an invariant Borel probability measure $\mu$, then we use $h_\mu(T)$ to denote the measure theoretic entropy and $h(T)$ to denote the topological entropy.

Lecture 32: Nov. 28

Recall defn of $h(T, \alpha)$ and $h(T)$

For simplicity, assume $T$ is ITDS.

Prop: Topological entropy is an invariant of topological conjugacy, i.e., if $S \approx T$, then $h(S) = h(T)$.

Want to compute $h(T)$

Easy facts:

- $h(T, \alpha) = h(T, \alpha_m^n)$.
- If $\alpha \preceq \beta$, then $h(T, \alpha) \leq h(T, \beta)$.

Defn: Diameter of an open cover is the sup of diameters of its elements.

Prop: If $\alpha_m$ is a sequence of open covers and diameter($\alpha_m$) approaches 0, then $h(T, \alpha_m) \to h(T, \alpha)$.

Proof: Assume $h(T) < \infty$. Let $\epsilon > 0$. Let $\beta$ be an open cover such that $h(T, \beta) > h(T) - \epsilon$. Choose $M$ such that for $m \geq M$, the diameter of $\alpha_m$ is less than the Lebesgue number of $\beta$, and so $\beta \preceq \alpha_m$. Thus,

\[ h(T) \geq h(T, \alpha_m) \geq h(T, \beta) > h(T) - \epsilon. \]

If $h(T) = \infty$. Let $N > 0$ and $\beta$ be an open cover such that $h(T, \beta) > N$. Continue as above. QED

Example: Let $T$ be rotation of circle. Let $\alpha_m$ be cover consisting of all open intervals of diameter $< 1/m$. 

Then $T^{-1}(\alpha_m) = \alpha_m$, and so $\alpha_m = (\alpha_m)^n$ and so \( h(T, \alpha_m) = 0 \).

Thus, \( h(T) = 0 \). QED

**Defn:** A *topological generator* for an ITDS is a finite open cover \( \alpha = \{A_1, \ldots, A_m\} \) such that every sequence \( \cap_{k=-\infty}^{\infty} A_{i_k} \) contains at most one point.

**Theorem:** If \( \alpha \) is a top. generator, then \( h(T) = h(T, \alpha) \).

**Claim:** \( \text{diam} (\alpha^n_{-n}) \to 0 \).

**Proof of Claim:** Use compactness:

If not, then there exists \( \delta > 0 \) and \( x_n, y_n \) in same element of \( (\alpha^n_{-n}) \). and \( d(x_n, y_n) > \delta \). Let the \( x_n, y_n \) accumulate on \( x, y \).

Then \( x \neq y \). By diagonalization can find an infinite sequence \( i_k \) s.t. \( x, y \in \cap_{k=-\infty}^{\infty} A_{i_k} \).

**Proof of Prop:** By Previous Prop, \( h(T, \alpha^n_{-n}) \to h(T) \).

By easy Prop., \( h(T, \alpha^n_{-n}) = h(T, \alpha) \). QED

For the full shift, the standard partition \( \alpha \) given by \( A_i = \{x \in F^\Z : x_0 = i\} \) is a top. generator (in particular it is an open cover).

Thus, \( h(\sigma_{F^\Z}) = h(\sigma_{F^\Z}, \alpha) \).

Observe \( N(\alpha_0^{n-1}) = |F|^n \).

Thus, \( h(\sigma_{F^\Z}) = \log |F| \)

**Corollary:** entropy of horseshoe is \( \log(2) \) (the complement of the invariant set has zero entropy).

The standard partition for a TMC \( M_C \) is also a top. generator (for the same reason).

**Notation:** \( \sigma_C \) denotes the shift on \( M_C \).
Proposition: For a TMC $M_C$, 

$$h(\sigma_C) = \log \lambda_C$$

where $\lambda_C$ is the spectral radius of $C$, i.e.,

$$\lambda_C = \max\{|\lambda| : \lambda \text{ eigenvalues of } C\}$$

Will use:

Perron-Frobenius Theorem:

Let $C$ be an irreducible matrix with spectral radius $\lambda_C$. ($C$ is a nonnegative square matrix s.t. for all $i, j$ there exists $n$ s.t. $(C^n)_{ij} > 0$.) Then

1. $\lambda_C > 0$ and is an eigenvalue of $A$.
2. $\lambda_C$ has a (strictly) positive eigenvector and is the only eigenvalue with a (strictly) positive eigenvector ((right Perron eigenvector is denoted $v$ and left Perron eigenvector is denoted $w$).)
3. $\lambda_C$ is a simple eigenvalue.
4. If $C$ is primitive (i.e., for some $n > 0$, $C^n > 0$), then

$$\frac{C^n}{\lambda_C^n} \to (v_i w_j)$$

(where $w \cdot v = 1$).
5. If $C$ is primitive, then $\lambda$ is the only eigenvalue with modulus $= \lambda$.

Note the consistency of the Proposition above with computation above for the Full shift:

$C$ is the all 1’s matrix of size $|F|$. So,

Claim: $\lambda_C = |F|$
Proof of Claim: \(|F|\) is an eigenvalue of \(C\) with eigenvector \(\mathbf{1}\)

Apply parts 1 and 2 of P-F Theorem.

Lecture 33: Nov. 30

Re-state P-F Theorem

Brief discussion of proof of P-F:

Let \(S\) be the unit simplex in \(\mathbb{R}^n\). Define

\[
f : S \rightarrow S, \quad x \mapsto \frac{xS}{xS\mathbf{1}}
\]

In primitive case, \(C^n > 0\) for some \(n\), and so \(f\) contracts into a single point, giving a unique eigenvalue \(\lambda\) with positive eigenvector. And \(\lambda\) is the only eigenvalue with largest modulus; for otherwise we would have leakage out of \(S\).

Modify to take care of the general irreducible case.

Compare with stochastic case.

Proposition: For a TMC \(M_C\),

\[
h(\sigma_C) = \log \lambda_C
\]

where \(\lambda_C\) is the spectral radius of \(C\), i.e.,

\[
\lambda_C = \max\{ |\lambda| : \lambda\text{ eigenvalues of } C \}
\]

Proof of Proposition:

Will prove Prop in special case that \(C\) is irreducible (in general, reduce to disjoint union of irreducible components plus transient connections).

Proof: Let \(\alpha\) be standard partition, a top. generator.

Then \(N(\alpha_0^{n-1})\) is the number of vertex sequences that occur in the graph \(G(C)\).
Thus, \( N(\alpha_0^{n-1}) = \log 1C^{m-1}1 \).

So,

\[
h(\sigma_C) = \lim_{n \to \infty} (1/n) \log 1C^n1
\]

P-F Theorem guarantees a positive (right) eigenvector \( v \) corresponding to \( \lambda = \lambda_C \).

Claim: There exist constants \( K_1, K_2 > 0 \) s.t.

\[
K_1\lambda^n \leq 1C^n1 \leq K_2\lambda^n
\]

Proof of Claim:

\[
1C^n1 \leq 1C^n\left(\frac{v}{\min(v)}\right) = (1 \cdot v)\frac{1}{\min(v)}\lambda^n
\]

\[
1C^n1 \geq 1C^n\left(\frac{v}{\max(v)}\right) = (1 \cdot v)\frac{1}{\max(v)}\lambda^n
\]

Take log, divide by \( n \), take \( \lim_{n} \), to get:

\[
h(\sigma_C) = \log \lambda
\]

QED

Example: Golden Mean.

\[
C = \begin{bmatrix} 1 & 1 \\ 1 & 0 \end{bmatrix}
\]

\[
\lambda_C = (1 + \sqrt{5})/2.
\]

For an ITDS \( T \), let \( \mathcal{M} = \mathcal{M}_T \) denote the set of all \( T \)-invariant Borel probability measures \( \mu \), i.e., \( T \) is an MPT with respect to \( (M, \text{Borel}, \mu) \).
Theorem (Variational Principle for irreducible TMC) Let $\mathcal{M} = \mathcal{M}_{\sigma C}$.

$$h(\sigma C) = \log \lambda_C = \sup_{\mu \in \mathcal{M}} h(\mu(\sigma C))$$

and the sup is achieved uniquely by an explicitly describable Markov chain:

$$P_{ij} = C_{ij} \frac{v_j}{\lambda_C v_i}$$

where $Cv = \lambda_C v$.

Lecture 34: Dec. 2

Proof of Variational Principle:

Terminology: $\lambda = \lambda_G$ is called the Perron eigenvalue of $C$ and it corresponding eigenvectors are called Perron eigenvectors.

Step 1: For all $\mu \in \mathcal{M}$, $h(\mu(\sigma C)) \leq \log \lambda = h(T)$.

Proof: Let $\alpha$ be the standard partition. Then

$$h(\mu(\sigma C)) = \lim_{n \to \infty} (1/n) H(\alpha^{n-1}) \leq \lim_{n \to \infty} (1/n) \log 1C^{m-1}1 = \log \lambda = h(\sigma C).$$

Step 2: Let $\mathcal{M}$ denote the set of all first-order stationary Markov $\nu \in \mathcal{M}$. Then

$$\sup_{\mu \in \mathcal{M}} h(\mu(\sigma C)) = \sup_{\nu \in \mathcal{M}} h(\nu(\sigma C))$$

Proof: Let $\mu \in \mathcal{M}$. Let

$$S = \{i : \mu(A_i) > 0\}$$

(recall $A_i = \{x \in M_C : x_0 = i\}$).

Let $\nu$ denote the first-order stationary Markov chain defined by the $|S| \times |S|$ matrix $P$:

$$P_{ij} = \frac{\mu(A_{ij})}{\mu(A_i)}$$
where
\[ A_{ij} = \{ x \in M_C : x_0 = i, \ x_1 = j \} \]
Then \( \nu \in \mathcal{M} \) and is called the Markovization of \( \mu \). Then
\[
h_{\mu}(\sigma) \leq H_{\mu}(\alpha|\sigma^{-1}(\alpha)) = -\sum_{ij} \mu(A_{ij}) \log P_{ij} = h_{\nu}(\sigma_C)
\]

Step 3: Find optimal \( \nu \in \overline{\mathcal{M}} \)

What is a good guess for the joint (NOT conditional) distribution \( Q_{ij} = \nu(A_{ij}) \)?

Make distribution on long words roughly uniform.

Assuming primitivity of \( C \), we get from P-F (part 4):
\[
\frac{C^m}{\lambda^n} \to (v_i w_j)
\]
(where \( C v = \lambda v, w C = \lambda w \), and \( w \cdot v = 1 \)).

So,
\[
(C^m)_{ki} \sim v_k w_i \lambda^n
\]
and
\[
(C^m)_{j\ell} \sim v_j w_\ell \lambda^n
\]

So, we guess:
\[
Q_{ij} \sim (1 C^m e_i) C_{ij} (e_j C^m 1) \sim C_{ij} w_i v_j \lambda^{2n} \sim C_{ij} w_i v_j
\]
Then
\[
\pi_i \sim \sum_j C'_{ij} w_i v_j = \lambda w_i v_i
\]
Thus,
\[
P_{ij} = \frac{Q_{ij}}{\pi_i} = C_{ij} \frac{w_i v_j}{\lambda w_i v_i} = C_{ij} \frac{v_j}{\lambda v_i}
\]
is a good guess. Can verify directly that this Markov chain achieves the sup, namely $\log \lambda$:

$$h_\nu = -\sum_{ij} Q_{ij} \log C_{ij} \frac{v_j}{\lambda v_i} = \log \lambda + \sum_{ij} Q_{ij} C_{ij} (\log v_j - \log v_i)$$

$$= \log \lambda + \sum_{ij} Q_{ij} (\log v_j - \log v_i) = \log \lambda$$

since

$$\sum_{ij} Q_{ij} (\log v_j - \log v_i) = \sum_j \pi_j \log v_j - \sum_i \pi_i \log v_i = 0.$$  
(Note: this is valid assuming irreducibility).

(Note: we get the same result if $Q$ is replaced by any other stationary $Q'$)

Thus, sup is achieved and equals $\log \lambda = h(\sigma_C)$.

Step 4: Uniqueness

a. if $h_\mu(\sigma_C) = h(\sigma_C)$, then $h_\mu(\sigma_C) = h_\nu(\sigma_C)$, where $\nu$ is the Markovization of $\mu$. By exercise 6 on Exercise Set 1, $\mu$ is Markov and $\mu = \nu$.

b. Let $P$ and $Q$ be the conditional and joint probability matrices for $\nu$ as above (i.e.

$$P_{ij} = C_{ij} \frac{v_j}{\lambda v_i}$$

Let $P'$ and $Q'$ be the conditional and joint probability matrices for an arbitrary $\nu' \in \mathcal{M}$.

If $h_{\nu'} = h_\nu$, then

$$-E_{Q'} \log P' = -E_Q \log P = -E_{Q'} \log P$$

and so

$$E_{Q'} \log \frac{P}{P'} = 0$$
And
\[ \log E_{Q'} \frac{P}{P'} = \log \sum_{ij} Q'_{ij} \frac{P_{ij}}{P'_{ij}} = \log \sum_{ij} \pi'_{ij} P_{ij} = 0. \]
(here, \( \pi' = \pi' P \)). Thus,
\[ E_{Q'} \log \frac{P}{P'} = 0 = \log E_{Q'} \frac{P}{P'} \]
By Jensen, equality happens only if \( P' = P \) on the support of \( Q' \).
But if the support of \( Q' \) is strictly smaller than the support of \( Q \),
then \( P' \) cannot be stochastic.

Example: Golden Mean
\[ v = w = \frac{[\lambda, 1]}{\lambda^2 + 1} \]
So,
\[ P = \begin{bmatrix} 1/\lambda & 1/\lambda^2 \\ 1 & 0 \end{bmatrix} \]
where \( \lambda \) is the golden mean.

Defn: Let \( C \) be an irreducible \( m \times m \) 0-1 matrix. Let \( f_1, \ldots, f_m \in \mathbb{R} \). Let \( C_f \) be the matrix defined by:
\[ (C_f)_{ij} = C_{ij} e^{f_j} \]
The pressure is defined:
\[ P_{\sigma_C}(f) = \lim_{n \to \infty} \left( \frac{1}{n} \right) \log 1(C_f)^{n-1} 1 \]
Note: If \( f \equiv 0 \), then \( P_{\sigma_C}(f) = h(\sigma_C) \).

Theorem (Variational Principle for pressure of irreducible TMC's)
Let \( \mathcal{M} = \mathcal{M}_{\sigma_C} \).
\[ P_{\sigma_C}(f) = \log \lambda_{C_f} = \sup_{\mu \in \mathcal{M}} h_{\mu}(\sigma_C) + \int_{\mathcal{M}_C} f d\mu \]
where \( f(x) = f_{x_0} \) and the sup is uniquely achieved by the Markov chain \( \nu \):

\[
P_{ij} = C_f \frac{v_j}{\lambda v_i}
\]

where \( v \) is the right Perron eigenvector of \( C_f \).

Compute:

Let \( Q_{ij} = \nu(A_{ij}) \).

\[
h\nu = -\sum_{ij} Q_{ij} \log C_{ij} e^{f_j} \frac{v_j}{\lambda v_i} = \log \lambda - \sum_{ij} Q_{ij} f_j + \sum_{ij} Q_{ij} C_{ij} (\log v_j - \log v_i)\]

\[
= \log \lambda - \int_{M_C} f \, d\nu + 0
\]

as before. Thus,

\[
h\nu + \int_{M_C} f \, d\nu = \log \lambda
\]

The main point:

Solution to a variational problem (equilibrium state) has an explicit exponential form (Gibbs state).

This generalizes the result discussed in lectures 1 and 2.

Further generalization:

Defn: Let \( T \) be an ITDS. Let \( f : M \to \mathbb{R} \) be continuous.

The pressure is defined as follows.

Define \( S_n f(x) = \sum_{i=0}^{n-1} f \circ T^i(x) \).

An \((n, \epsilon)-spanning \) set is a set \( F \) such that for all \( x \in M \), there exists \( y \in F \) such that \( d(T^i x, T^i y) < \epsilon \), for \( i = 0, \ldots, n - 1 \). Let

\[
P_T(f,\epsilon,n) = \inf_{(n,\epsilon)-spanning \, \text{sets} \, F} \sum_{x \in F} e^{S_n f(x)}
\]
Let
\[ P_T(f, \epsilon) = \limsup_{n \to \infty} \frac{1}{n} \log P_T(f, \epsilon, n) \]

Define the pressure as
\[ P_T(f) = \lim_{\epsilon \to 0} P_T(f, \epsilon) \]

Note: when \( f = 0 \), \( P_T(f) = h(T) \); to see this, use the spanning set definition of topological entropy given in Walters or Keller.

Theorem (generalized variational principle)
\[ P_T(f) = \sup_{\mu \in \mathcal{M}} h_\mu(T) + \int f \, d\mu \]

Note: sup is not always achieved. When it is achieved, it need not be unique.

Note: But there are many important classes where the sup is achieved uniquely.

Note: The RHS above can be expressed as a constrained maximization of \( h_\mu(T) \), subject to constraint on \( \int f \, d\mu \).

Note: And cases where it is achieved non-uniquely correspond to phase transitions.

THE END