Measures of maximal entropy in symbolic dynamics

Adam Lott

8 May 2020

Outline

1. Intro to symbolic dynamics
2. Ergodic theory of symbolic systems
3. Measures of maximal entropy
4. Further topics

Setup

Setup

Let A be a finite set (the **alphabet**)

Setup

Let A be a finite set (the **alphabet**)

$A^{\mathbb{Z}}$ is the set of all bi-infinite sequences of letters from A

Setup

Let A be a finite set (the **alphabet**)

$A^{\mathbb{Z}}$ is the set of all bi-infinite sequences of letters from A

The **left shift** $\sigma : A^{\mathbb{Z}} \to A^{\mathbb{Z}}$ is defined by $\sigma(x)_n = x_{n+1} \quad (n \in \mathbb{Z})$

Setup

Let A be a finite set (the **alphabet**)

$A^{\mathbb{Z}}$ is the set of all bi-infinite sequences of letters from A

The **left shift** $\sigma : A^{\mathbb{Z}} \to A^{\mathbb{Z}}$ is defined by $\sigma(x)_n = x_{n+1} \quad (n \in \mathbb{Z})$

A **subshift** is a closed set $X \subseteq A^{\mathbb{Z}}$ with $\sigma(X) = X$ (shift-invariant)

Setup

Let A be a finite set (the **alphabet**)

$A^{\mathbb{Z}}$ is the set of all bi-infinite sequences of letters from A

The **left shift** $\sigma : A^{\mathbb{Z}} \to A^{\mathbb{Z}}$ is defined by $\sigma(x)_n = x_{n+1} \quad (n \in \mathbb{Z})$

A **subshift** is a closed set $X \subseteq A^{\mathbb{Z}}$ with $\sigma(X) = X$ (shift-invariant)

Examples:

- $X = A^{\mathbb{Z}}$ (the **full shift**)
- $X = \{x \in \{0,1\}^{\mathbb{Z}} : x$ does not contain the pattern $11\}$
- $X = \{x \in \{0,1,2\}^{\mathbb{Z}} : x_n - x_{n-1} = 0$ or $1 \mod 3\}$

Symbolic dynamics: studying properties of the dynamical system (X, σ)

Topology of symbolic space

Topology of symbolic space

Endow A with the discrete topology and $A^{\mathbb{Z}}$ with the product topology

Topology of symbolic space

Endow A with the discrete topology and $A^{\mathbb{Z}}$ with the product topology

The topology is generated by the family of **cylinder sets**

$$[a_0 a_1 \ \dots \ a_{n-1}] := \{x \in A^{\mathbb{Z}} : x_0 x_1 \dots x_{n-1} = a_0 a_1 \dots a_{n-1}\}$$

- Cylinder sets are clopen
- σ is continuous

Topology of symbolic space

Endow A with the discrete topology and $A^{\mathbb{Z}}$ with the product topology

The topology is generated by the family of **cylinder sets**

$$[a_0 a_1 \ \dots \ a_{n-1}] := \{x \in A^{\mathbb{Z}} : x_0 x_1 \dots x_{n-1} = a_0 a_1 \dots a_{n-1}\}$$

- Cylinder sets are clopen
- σ is continuous

It's also generated by a natural metric on $A^{\mathbb{Z}}$

$$d(x, y) := 2^{-\min\{|n| : x_n \neq y_n\}}$$

- This makes $A^{\mathbb{Z}}$ a compact metric space and open balls are cylinder sets

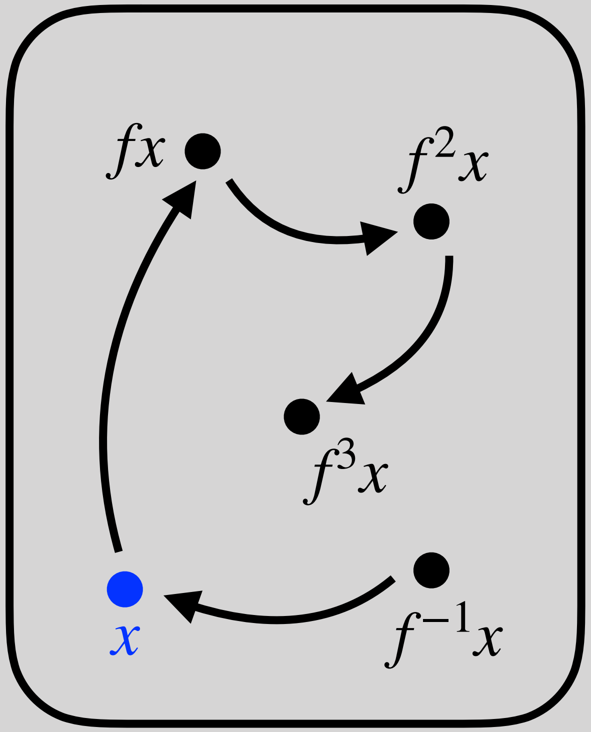
Motivation

Motivation

Many "real-world" dynamical systems of interest can be modeled by symbolic systems

Motivation

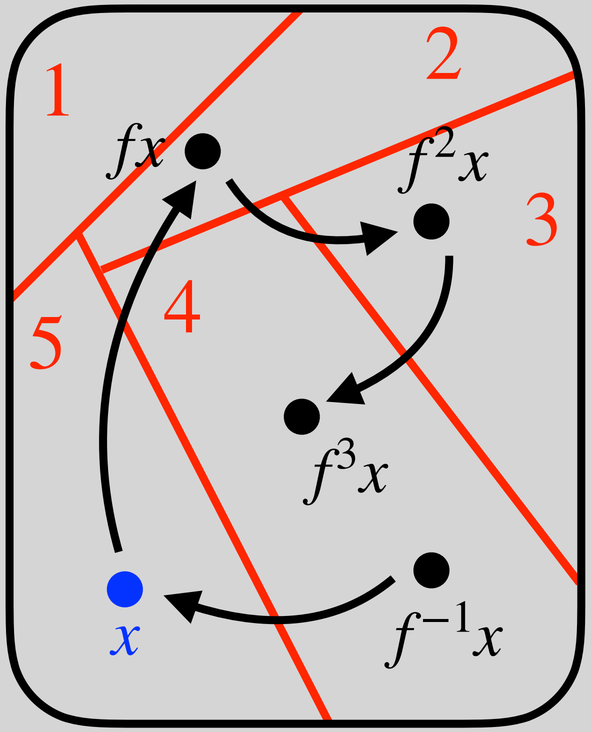
Many "real-world" dynamical systems of interest can be modeled by symbolic systems



(M, f)

Motivation

Many "real-world" dynamical systems of interest can be modeled by symbolic systems

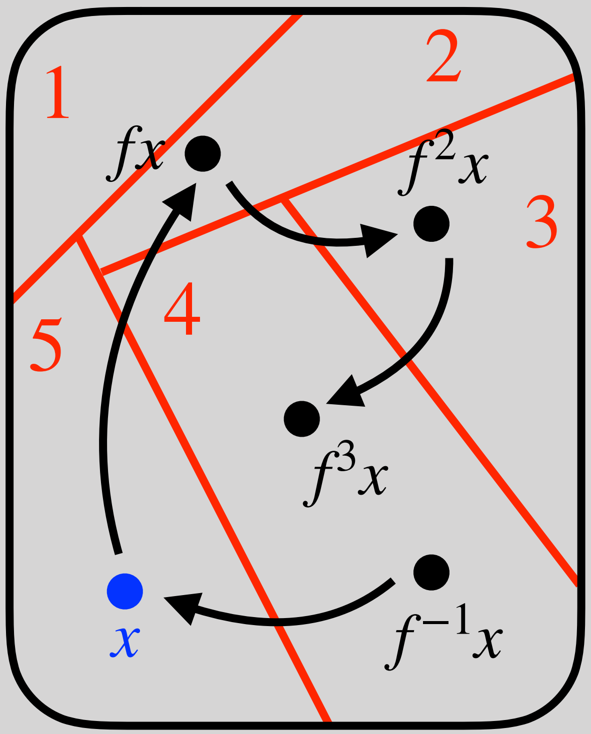


$$(M, f)$$

P = partition of M

Motivation

Many "real-world" dynamical systems of interest can be modeled by symbolic systems



(M,f)

$$x \longleftrightarrow (\ldots,4,5,2,3,4,\ldots)$$

P = partition of M

Topological entropy

Topological entropy

Given a subshift X, let

$$\mathscr{L}_n := \{a_1 \dots a_n \in A^n : x_i \dots x_{i+n-1} = a_1 \dots a_n \text{ for some } x \in X, i \in \mathbb{Z}\}$$

be the family of all **words** of length n that appear in X

Topological entropy

Given a subshift X, let

$\mathscr{L}_n := \{a_1 \ldots a_n \in A^n : x_i \ldots x_{i+n-1} = a_1 \ldots a_n$ for some $x \in X, i \in \mathbb{Z}\}$

be the family of all **words** of length n that appear in X

The **topological entropy** of X is

$$h_{\text{top}}(X) := \lim_{n\to\infty} \frac{1}{n} \log |\mathscr{L}_n| = \inf_n \frac{1}{n} \log |\mathscr{L}_n|$$

- Measures the "size" or "complexity" of X
- $|\mathscr{L}_n| = \exp(h_{\text{top}} n) \cdot$ (sub-exponential factor)

Examples

$X = \{0,1\}^{\mathbb{Z}}$

- $|\mathscr{L}_n| = 2^n \implies h_{\text{top}}(X) = \log 2$

$X = \{x \in \{0,1\}^{\mathbb{Z}} : x$ does not contain the patttern $11\}$

- $\mathscr{L}_n =$ binary strings of length n avoiding 11
- $|\mathscr{L}_n| = F_n =$ n[th] Fibonacci number $\sim \varphi^n$
- $h_{\text{top}}(X) \approx \log(1.618)$

Outline

1. Intro to symbolic dynamics
2. Ergodic theory of symbolic systems
3. Measures of maximal entropy
4. Further topics

Ergodic theory

Ergodic theory

Let μ be a **shift-invariant** Borel probability measure on X
($\mu(\sigma^{-1}A) = \mu(A)$ for all measurable $A \subseteq X$)

Ergodic theory

Let μ be a **shift-invariant** Borel probability measure on X ($\mu(\sigma^{-1}A) = \mu(A)$ for all measurable $A \subseteq X$)

Two interpretations:

- μ as an invariant measure on the system (X, σ)
- μ as the joint distribution of a stationary A-valued stochastic process $(\ldots, \alpha_{-1}, \alpha_0, \alpha_1, \ldots)$

Ergodic theory

Let μ be a **shift-invariant** Borel probability measure on X
($\mu(\sigma^{-1}A) = \mu(A)$ for all measurable $A \subseteq X$)

Two interpretations:

- μ as an invariant measure on the system (X, σ)
- μ as the joint distribution of a stationary A-valued stochastic process $(\ldots, \alpha_{-1}, \alpha_0, \alpha_1, \ldots)$

Measure entropy

Measure entropy

Recall the **Shannon entropy** of a probability measure ν on a finite set F is

$$H(\nu) = \sum_{x \in F} -\nu(x) \log \nu(x)$$

- A measure of how uniform ν is
- Key features: $H(\nu) \leq \log |\operatorname{supp}(\nu)|$, with equality iff $\nu = \mathrm{Unif}$

Measure entropy

Recall the **Shannon entropy** of a probability measure ν on a finite set F is

$$H(\nu) = \sum_{x \in F} -\nu(x) \log \nu(x)$$

- A measure of how uniform ν is
- Key features: $H(\nu) \leq \log |\mathrm{supp}(\nu)|$, with equality iff $\nu = \mathrm{Unif}$

The **entropy rate** of a shift-invariant measure μ on X is

$$h(\mu) := \lim_{n \to \infty} \frac{1}{n} H(\mu_{[1,n]}) = \inf_n \frac{1}{n} H(\mu_{[1,n]})$$

where $\mu_{[1,n]}$ is the marginal of μ on coordinates $[1,n]$ (or the joint law of $\alpha_1, \ldots, \alpha_n$)

Measure entropy

Recall the **Shannon entropy** of a probability measure ν on a finite set F is

$$H(\nu) = \sum_{x \in F} -\nu(x) \log \nu(x)$$

- A measure of how uniform ν is
- Key features: $H(\nu) \leq \log|\mathrm{supp}(\nu)|$, with equality iff $\nu = \mathrm{Unif}$

The **entropy rate** of a shift-invariant measure μ on X is

$$h(\mu) := \lim_{n \to \infty} \frac{1}{n} H(\mu_{[1,n]}) = \inf_n \frac{1}{n} H(\mu_{[1,n]})$$

where $\mu_{[1,n]}$ is the marginal of μ on coordinates $[1,n]$ (or the joint law of $\alpha_1, \ldots, \alpha_n$)

- $h(\mu)$ is high if μ gives roughly equal probabilities to all long words of the same length
- $h(\mu)$ is low if a few long words are much more likely than all the rest
- "h_{top} counts words, $h(\mu)$ counts words weighted according to μ"

The variational principle

The variational principle

Theorem: For any subshift X, one has $h_{\mathrm{top}}(X) = \sup_\mu h(\mu)$ where the sup is taken over all shift-invariant measures, and there is always at least one μ that achieves the sup.

The variational principle

Theorem: For any subshift X, one has $h_{\text{top}}(X) = \sup_\mu h(\mu)$ where the sup is taken over all shift-invariant measures, and there is always at least one μ that achieves the sup.

Proof (sketch):

The variational principle

Theorem: For any subshift X, one has $h_{\text{top}}(X) = \sup_\mu h(\mu)$ where the sup is taken over all shift-invariant measures, and there is always at least one μ that achieves the sup.

Proof (sketch):

- For any μ, note $\mu_{[1,n]}$ is supported on $\mathscr{L}_n$ (permitted words of length n), so $H(\mu_{[1,n]}) \leq \log|\mathscr{L}_n| \Longrightarrow h(\mu) \leq h_{\text{top}}(X)$

The variational principle

Theorem: For any subshift X, one has $h_{\text{top}}(X) = \sup_\mu h(\mu)$ where the sup is taken over all shift-invariant measures, and there is always at least one μ that achieves the sup.

Proof (sketch):

- For any μ, note $\mu_{[1,n]}$ is supported on $\mathscr{L}_n$ (permitted words of length n), so $H(\mu_{[1,n]}) \leq \log |\mathscr{L}_n| \implies h(\mu) \leq h_{\text{top}}(X)$
- For the other direction, construct an optimal μ directly using the principle that uniform measures maximize entropy

The variational principle, cont.

The variational principle, cont.

- Let ν_n be any measure that gives mass $|\mathscr{L}_n|^{-1}$ to each cylinder set $[w], w \in \mathscr{L}_n$ (ν_n is not shift-invariant)

The variational principle, cont.

- Let ν_n be any measure that gives mass $|\mathscr{L}_n|^{-1}$ to each cylinder set $[w], w \in \mathscr{L}_n$ (ν_n is not shift-invariant)
- Let $\nu_n' = \frac{1}{n}\sum_{i=0}^{n-1} \sigma_*^i \nu_n$ and let μ be any weak-* limit point of the ν_n' (now μ is shift-invariant)

The variational principle, cont.

- Let ν_n be any measure that gives mass $|\mathscr{L}_n|^{-1}$ to each cylinder set $[w], w \in \mathscr{L}_n$ (ν_n is not shift-invariant)
- Let $\nu_n' = \frac{1}{n}\sum_{i=0}^{n-1} \sigma_*^i \nu_n$ and let μ be any weak-* limit point of the ν_n' (now μ is shift-invariant)
- With some entropy calculations one can show that $h(\mu) = h_{\text{top}}(X)$

The variational principle, cont.

- Let ν_n be any measure that gives mass $|\mathscr{L}_n|^{-1}$ to each cylinder set $[w], w \in \mathscr{L}_n$ (ν_n is not shift-invariant)
- Let $\nu_n' = \frac{1}{n}\sum_{i=0}^{n-1} \sigma_*^i \nu_n$ and let μ be any weak-* limit point of the ν_n' (now μ is shift-invariant)
- With some entropy calculations one can show that $h(\mu) = h_{\text{top}}(X)$

Measures achieving the sup in the variational principle are called **measures of maximal entropy (MMEs)**

Outline

1. Intro to symbolic dynamics
2. Ergodic theory of symbolic systems
3. Measures of maximal entropy
4. Further topics

Question: what conditions on a subshift X guarantee that it has a **unique** MME?

(assuming $h_{\text{top}}(X) > 0$)

Easy example

$X = \{0,1\}^{\mathbb{Z}}$

- $\sup_{\mu} h(\mu) = \sup_{p \in \mathrm{Prob}\{0,1\}} \sup_{\mu : \mu_1 = p} h(\mu)$
- Given the constraint $\mu_1 = p$, $h(\mu)$ is uniquely maximized by the product measure $p^{\times \mathbb{Z}}$, and $h(\mu) = H(p)$
- $H(p)$ is uniquely maximized by $p = (1/2, 1/2)$
- So $(1/2, 1/2)^{\times \mathbb{Z}}$ is the unique MME

Markov shifts

Markov shifts

We will focus on a certain class of subshift:

Markov shifts

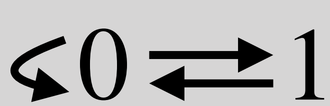
We will focus on a certain class of subshift:

Given an alphabet A, let G be a directed graph with vertex set A. Define a subshift $X_G \subseteq A^{\mathbb{Z}}$ by the condition $x \in X$ iff $x_n \to x_{n+1}$ is an edge in G for all $n \in \mathbb{Z}$

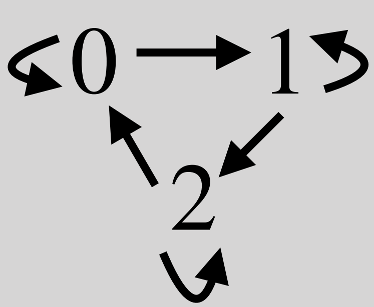
Markov shifts

We will focus on a certain class of subshift:

Given an alphabet A, let G be a directed graph with vertex set A. Define a subshift $X_G \subseteq A^{\mathbb{Z}}$ by the condition $x \in X$ iff $x_n \to x_{n+1}$ is an edge in G for all $n \in \mathbb{Z}$



$\{x \in \{0,1\}^{\mathbb{Z}} : x$ does not contain the pattern $11\}$

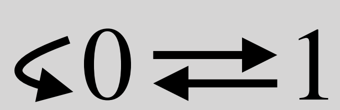


$\{x \in \{0,1,2\}^{\mathbb{Z}} : x_n - x_{n-1} = 0$ or $1 \bmod 3\}$

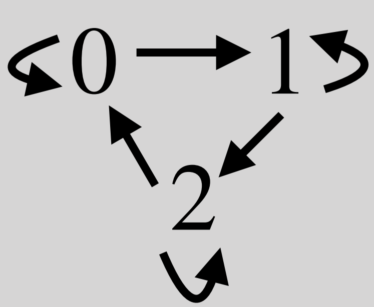
Markov shifts

We will focus on a certain class of subshift:

Given an alphabet A, let G be a directed graph with vertex set A. Define a subshift $X_G \subseteq A^{\mathbb{Z}}$ by the condition $x \in X$ iff $x_n \to x_{n+1}$ is an edge in G for all $n \in \mathbb{Z}$



$\{x \in \{0,1\}^{\mathbb{Z}} : x$ does not contain the pattern $11\}$



$\{x \in \{0,1,2\}^{\mathbb{Z}} : x_n - x_{n-1} = 0$ or $1 \mod 3\}$

Such subshifts are called **Markov shifts** or **subshifts of finite type**

Many "nice" dynamical systems can be modeled by Markov shifts (Y. Sinai, R. Bowen, D. Ruelle, 1960s-80s, "Markov partitions")

A classical result

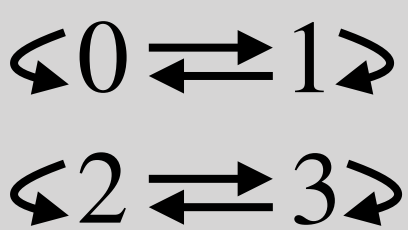
A classical result

When will a Markov shift have a unique MME?

A classical result

When will a Markov shift have a unique MME?

There is one obvious obstruction:

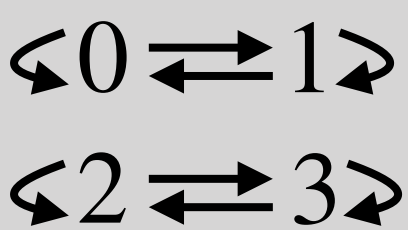


$\mu_1 = (1/2,1/2,0,0)^{\times\mathbb{Z}}$ and $\mu_2 = (0,0,1/2,1/2)^{\times\mathbb{Z}}$ are both MMEs

A classical result

When will a Markov shift have a unique MME?

There is one obvious obstruction:



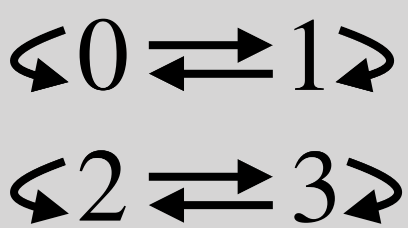
$\mu_1 = (1/2, 1/2, 0, 0)^{\times \mathbb{Z}}$ and $\mu_2 = (0, 0, 1/2, 1/2)^{\times \mathbb{Z}}$ are both MMEs

Say a Markov shift X is **irreducible** if the graph generating it is (strongly) connected

A classical result

When will a Markov shift have a unique MME?

There is one obvious obstruction:



$\mu_1 = (1/2,1/2,0,0)^{\times\mathbb{Z}}$ and $\mu_2 = (0,0,1/2,1/2)^{\times\mathbb{Z}}$ are both MMEs

Say a Markov shift X is **irreducible** if the graph generating it is (strongly) connected

Theorem (W. Parry, 1960s): Any irreducible Markov shift X has a unique MME

Proof, step 1

Proof, step 1

By irreducibility, there is some **gap bound** $g \in \mathbb{N}$ such that any two letters in A can be joined by a path of length $\leq g$

- Equivalently, if w and w' are any two permitted words, they can be joined into a permitted word wuw' with $|u| \leq g$

Proof, step 1

By irreducibility, there is some **gap bound** $g \in \mathbb{N}$ such that any two letters in A can be joined by a path of length $\leq g$

- Equivalently, if w and w' are any two permitted words, they can be joined into a permitted word wuw' with $|u| \leq g$

This gives good control on the number of allowed words:

Proof, step 1

By irreducibility, there is some **gap bound** $g \in \mathbb{N}$ such that any two letters in A can be joined by a path of length $\leq g$

- Equivalently, if w and w' are any two permitted words, they can be joined into a permitted word wuw' with $|u| \leq g$

This gives good control on the number of allowed words:

- $w^1, w^2, \ldots, w^r \in \mathscr{L}_n \mapsto w^1 u^1 w^2 \ldots u^{r-1} w^r s \in \mathscr{L}_{nr+g(r-1)}$

Proof, step 1

By irreducibility, there is some **gap bound** $g \in \mathbb{N}$ such that any two letters in A can be joined by a path of length $\leq g$

- Equivalently, if w and w' are any two permitted words, they can be joined into a permitted word wuw' with $|u| \leq g$

This gives good control on the number of allowed words:

- $w^1, w^2, \ldots, w^r \in \mathcal{L}_n \mapsto w^1 u^1 w^2 \ldots u^{r-1} w^r s \in \mathcal{L}_{nr+g(r-1)}$
- So $|\mathcal{L}_n|^r \leq |\mathcal{L}_{r(n+g)-g}| \implies \log|\mathcal{L}_n| \leq \frac{1}{r} \log|\mathcal{L}_{r(n+g)-g}|$

Proof, step 1

By irreducibility, there is some **gap bound** $g \in \mathbb{N}$ such that any two letters in A can be joined by a path of length $\leq g$

- Equivalently, if w and w' are any two permitted words, they can be joined into a permitted word wuw' with $|u| \leq g$

This gives good control on the number of allowed words:

- $w^1, w^2, \ldots, w^r \in \mathcal{L}_n \mapsto w^1 u^1 w^2 \ldots u^{r-1} w^r s \in \mathcal{L}_{nr+g(r-1)}$
- So $|\mathcal{L}_n|^r \leq |\mathcal{L}_{r(n+g)-g}| \implies \log|\mathcal{L}_n| \leq \frac{1}{r}\log|\mathcal{L}_{r(n+g)-g}|$
- Take $r \to \infty$: $|\mathcal{L}_n| \leq \exp((n+g)h_{top}) = C\exp(nh_{top})$

Proof, step 1

By irreducibility, there is some **gap bound** $g \in \mathbb{N}$ such that any two letters in A can be joined by a path of length $\leq g$

- Equivalently, if w and w' are any two permitted words, they can be joined into a permitted word wuw' with $|u| \leq g$

This gives good control on the number of allowed words:

- $w^1, w^2, \ldots, w^r \in \mathscr{L}_n \mapsto w^1u^1w^2\ldots u^{r-1}w^r s \in \mathscr{L}_{nr+g(r-1)}$
- So $|\mathscr{L}_n|^r \leq |\mathscr{L}_{r(n+g)-g}| \implies \log|\mathscr{L}_n| \leq \frac{1}{r}\log|\mathscr{L}_{r(n+g)-g}|$
- Take $r \to \infty$: $|\mathscr{L}_n| \leq \exp((n+g)h_{top}) = C\exp(nh_{top})$
- From definition of h_{top}: $|\mathscr{L}_n| \geq \exp(nh_{top})$

Proof, step 1

By irreducibility, there is some **gap bound** $g \in \mathbb{N}$ such that any two letters in A can be joined by a path of length $\leq g$

- Equivalently, if w and w' are any two permitted words, they can be joined into a permitted word wuw' with $|u| \leq g$

This gives good control on the number of allowed words:

- $w^1, w^2, \ldots, w^r \in \mathscr{L}_n \mapsto w^1 u^1 w^2 \ldots u^{r-1} w^r s \in \mathscr{L}_{nr+g(r-1)}$
- So $|\mathscr{L}_n|^r \leq |\mathscr{L}_{r(n+g)-g}| \implies \log|\mathscr{L}_n| \leq \frac{1}{r} \log|\mathscr{L}_{r(n+g)-g}|$
- Take $r \to \infty$: $|\mathscr{L}_n| \leq \exp((n+g)h_{top}) = C\exp(nh_{top})$
- From definition of h_{top}: $|\mathscr{L}_n| \geq \exp(nh_{top})$

To summarize: $|\mathscr{L}_n| \asymp \exp(nh_{top})$ for all n

Proof, step 2

Proof, step 2

Recall the general procedure for constructing MMEs:

- $\nu_n =$ uniform over $\mathscr{L}_n$, $\mu =$ limit of running averages of the ν_n

Proof, step 2

Recall the general procedure for constructing MMEs:

- $\nu_n =$ uniform over $\mathscr{L}_n$, $\mu =$ limit of running averages of the ν_n

So we can estimate the μ-mass of any fixed cylinder $[w]$, $w \in \mathscr{L}_k$:

$$\mu[w] \approx \frac{1}{n}\left(\nu_n[w] + \nu_n\sigma^{-1}[w] + \ldots + \nu_n\sigma^{n-1}[w]\right) \quad \text{(consider } n \gg k\text{)}$$

Proof, step 2

Recall the general procedure for constructing MMEs:

- $\nu_n =$ uniform over $\mathscr{L}_n$, $\mu =$ limit of running averages of the ν_n

So we can estimate the μ-mass of any fixed cylinder $[w]$, $w \in \mathscr{L}_k$:

$$\mu[w] \approx \frac{1}{n}\left(\nu_n[w] + \nu_n\sigma^{-1}[w] + \ldots + \nu_n\sigma^{n-1}[w]\right) \quad \text{(consider } n \gg k\text{)}$$

By definition

$$\nu_n\sigma^{-i}[w] = |\mathscr{L}_n|^{-1} \cdot \#\left\{u \in \mathscr{L}_n : u_{[i,i+k)} = w_{[i,i+k)}\right\}$$

Proof, step 2

Recall the general procedure for constructing MMEs:

- $\nu_n =$ uniform over $\mathcal{L}_n$, $\mu =$ limit of running averages of the ν_n

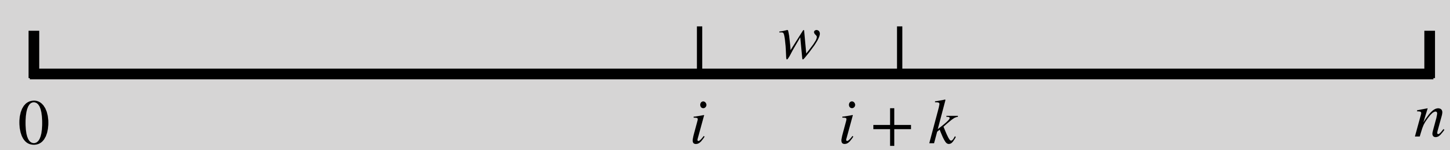
So we can estimate the μ-mass of any fixed cylinder $[w]$, $w \in \mathcal{L}_k$:

$$\mu[w] \approx \frac{1}{n}\left(\nu_n[w] + \nu_n\sigma^{-1}[w] + \ldots + \nu_n\sigma^{n-1}[w]\right) \quad \text{(consider } n \gg k\text{)}$$

By definition

$$\nu_n\sigma^{-i}[w] = |\mathcal{L}_n|^{-1} \cdot \#\left\{u \in \mathcal{L}_n : u_{[i,i+k)} = w_{[i,i+k)}\right\}$$

Use the gap bound property again:



Proof, step 2

Recall the general procedure for constructing MMEs:

- ν_n = uniform over $\mathscr{L}_n$, μ = limit of running averages of the ν_n

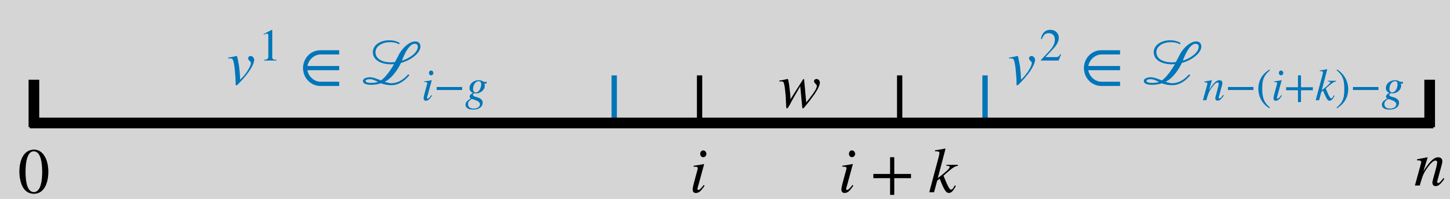
So we can estimate the μ-mass of any fixed cylinder $[w]$, $w \in \mathscr{L}_k$:

$$\mu[w] \approx \frac{1}{n}\left(\nu_n[w] + \nu_n\sigma^{-1}[w] + \ldots + \nu_n\sigma^{n-1}[w]\right) \quad \text{(consider } n \gg k\text{)}$$

By definition

$$\nu_n\sigma^{-i}[w] = |\mathscr{L}_n|^{-1} \cdot \#\left\{u \in \mathscr{L}_n : u_{[i,i+k)} = w_{[i,i+k)}\right\}$$

Use the gap bound property again:



Proof, step 2

Recall the general procedure for constructing MMEs:

- ν_n = uniform over $\mathscr{L}_n$, μ = limit of running averages of the ν_n

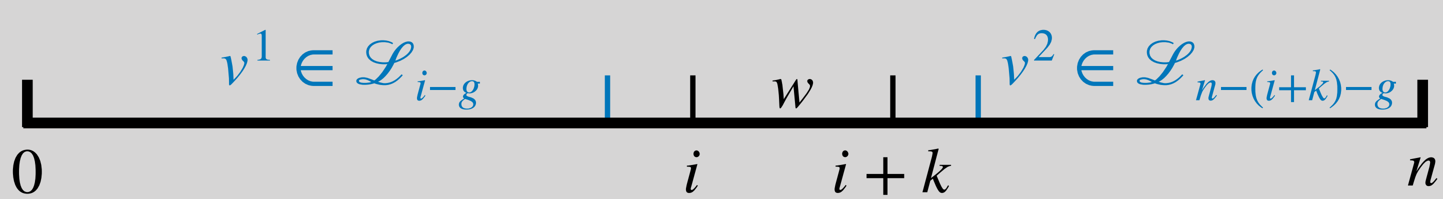
So we can estimate the μ-mass of any fixed cylinder $[w]$, $w \in \mathscr{L}_k$:

$$\mu[w] \approx \frac{1}{n}\left(\nu_n[w] + \nu_n\sigma^{-1}[w] + \ldots + \nu_n\sigma^{n-1}[w]\right) \quad \text{(consider } n \gg k\text{)}$$

By definition

$$\nu_n\sigma^{-i}[w] = |\mathscr{L}_n|^{-1} \cdot \#\left\{u \in \mathscr{L}_n : u_{[i,i+k)} = w_{[i,i+k)}\right\}$$

Use the gap bound property again:



So this number is $\asymp \exp(h_{top}(n-k))$

Proof, step 2

Recall the general procedure for constructing MMEs:

- $\nu_n =$ uniform over $\mathscr{L}_n$, $\mu =$ limit of running averages of the ν_n

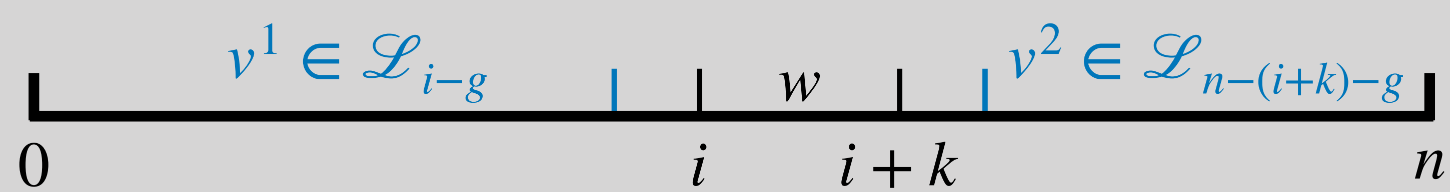
So we can estimate the μ-mass of any fixed cylinder $[w]$, $w \in \mathscr{L}_k$:

$$\mu[w] \approx \frac{1}{n}\left(\nu_n[w] + \nu_n\sigma^{-1}[w] + \ldots + \nu_n\sigma^{n-1}[w]\right) \quad \text{(consider } n \gg k\text{)}$$

By definition

$$\nu_n\sigma^{-i}[w] = |\mathscr{L}_n|^{-1} \cdot \#\left\{u \in \mathscr{L}_n : u_{[i,i+k)} = w_{[i,i+k)}\right\}$$

Use the gap bound property again:



So this number is $\asymp \exp(h_{top}(n-k))$

Combining this and the estimate on $|\mathscr{L}_n|$ yields

$\mu[w] \asymp \exp(-kh_{top})$ for all $w \in \mathscr{L}_k$, any fixed k **("Gibbs property for μ")**

Proof, step 3

Proof, step 3

Suppose ν is a different MME. After some reductions we can assume μ and ν are mutually singular.

Proof, step 3

Suppose ν is a different MME. After some reductions we can assume μ and ν are mutually singular.

Let n be large, and partition X into two disjoint families of n-cylinders $M_\mu^{(n)}$ and $M_\nu^{(n)}$ such that $\mu(M_\nu^{(n)}) \approx \nu(M_\mu^{(n)}) \approx 0$

Proof, step 3

Suppose ν is a different MME. After some reductions we can assume μ and ν are mutually singular.

Let n be large, and partition X into two disjoint families of n-cylinders $M_\mu^{(n)}$ and $M_\nu^{(n)}$ such that $\mu(M_\nu^{(n)}) \approx \nu(M_\mu^{(n)}) \approx 0$

Now estimate

$$\begin{aligned} nh_{top} &= nh(\nu) \leq H(\nu_{[1,n]}) \lessapprox \log \# M_\nu^{(n)} \\ &\leq \log(Ce^{nh_{top}}\mu(M_\nu^{(n)})) \qquad \text{(using the Gibbs property)} \\ &= O(1) + nh_{top} + \log(\mu(M_\nu^{(n)})) \end{aligned}$$

Contradiction.

Outline

1. Intro to symbolic dynamics

2. Ergodic theory of symbolic systems

3. Measures of maximal entropy

4. Further topics

Specification

Specification

The key to the proof was the "gap bound" property

Specification

The key to the proof was the "gap bound" property

Definition: A subshift X has the **specification property** if there exists $g \in \mathbb{N}$ such that whenever w and w' are permitted words, there is a joining word u such that wuw' is permitted and $|u| \leq g$

Specification

The key to the proof was the "gap bound" property

Definition: A subshift X has the **specification property** if there exists $g \in \mathbb{N}$ such that whenever w and w' are permitted words, there is a joining word u such that wuw' is permitted and $|u| \leq g$

- Much larger class of systems than the irreducible Markov shifts
- Example: $X = \{x \in \{0,1,2\}^{\mathbb{Z}} : \frac{1}{n}(x_i + x_{i+1} + \ldots + x_{i+n-1}) \leq 1$ for all $i, n\}$

Specification

The key to the proof was the "gap bound" property

Definition: A subshift X has the **specification property** if there exists $g \in \mathbb{N}$ such that whenever w and w' are permitted words, there is a joining word u such that wuw' is permitted and $|u| \leq g$

- Much larger class of systems than the irreducible Markov shifts
- Example: $X = \{x \in \{0,1,2\}^{\mathbb{Z}} : \frac{1}{n}(x_i + x_{i+1} + \ldots + x_{i+n-1}) \leq 1$ for all $i, n\}$

Theorem (R. Bowen, 1970s): Any subshift with the specification property has a unique MME

Specification

The key to the proof was the "gap bound" property

Definition: A subshift X has the **specification property** if there exists $g \in \mathbb{N}$ such that whenever w and w' are permitted words, there is a joining word u such that wuw' is permitted and $|u| \leq g$

- Much larger class of systems than the irreducible Markov shifts
- Example: $X = \{x \in \{0,1,2\}^{\mathbb{Z}} : \frac{1}{n}(x_i + x_{i+1} + \ldots + x_{i+n-1}) \leq 1$ for all $i, n\}$

Theorem (R. Bowen, 1970s): Any subshift with the specification property has a unique MME

Theorems (R. Pavlov, V. Climenhaga, D. Thompson, 2010s): different "weak specification" properties also imply uniqueness of MME

Stronger statistical properties

Stronger statistical properties

Once uniqueness of the MME is established, what other properties does it have?

Stronger statistical properties

Once uniqueness of the MME is established, what other properties does it have?

- Mixing?
- Bernoulli?
- Exponential decay of correlations?
- Central limit theorem?

Stronger statistical properties

Once uniqueness of the MME is established, what other properties does it have?

- Mixing?
- Bernoulli?
- Exponential decay of correlations?
- Central limit theorem?

Theorem (R. Bowen, 1970s): If X is an irreducible Markov shift, then its unique MME has all of the above

Theorem (V. Climenhaga, 2018): If X is a subshift with the specification property, then its unique MME has all of the above.