Math 254B Lecture 17 Notes

Daniel Raban

May 7, 2019

1 The Variational Principle

1.1 Relationship between KS and topological entropy

Let $X \subseteq \{0, \dots, k-1\}^{\mathbb{N}}$ be a subshift, and let $\pi_n : X \to \{0, \dots, k-1\}^n$ be the projection onto the first n coordinates. Then $\pi_n(X)$ is the permitted words of length n. We defined

$$h_{\text{top}}(\sigma|_X) = \lim_n \frac{1}{n} \log |\pi_n(X)|.$$

Observe that

$$0 \leq h_{\text{top}} \leq \log(k).$$

Suppose $\mu \in P^\sigma(X)$. Let $\alpha_0 : X \to \{0, \dots, k-1\}$ be the time-zero observable. Then $\pi_n = (\alpha_0, \dots, \alpha_{n-1})$. Now define $\mu_n = \pi_{n*}\mu$, the joint distribution of $\alpha_0, \dots, \alpha_{n-1}$. Then the entropy rate is

$$h_\mu(\sigma) = \lim_n \frac{1}{n} H(\mu_n) \leq \lim_n \frac{1}{n} \log |\pi_n(x)| \leq h_{\text{top}}(X).$$

It turns out that we can approximate h_{top} by h_μ:

Theorem 1.1 (Variational principle)**.** *For all (X, T),*

$$h_{\text{top}}(X, T) = \sup_{\mu \in P^T(X)} h_\mu(T).$$

1.2 Shearer's inequality

We need to prove a fact from information theory.

Lemma 1.1 (Shearer's inequality[1])**.** *Let $\alpha_1, \dots, \alpha_n$ be finite-valued random variables. Assume that $\mathcal{S} \subseteq \mathscr{P}(\{1, \dots, n\})$ such that every $i \in \{1, \dots, n\}$ is contained in $\geq k$ members of $\mathcal{S}$ (a "k cover of $\{1, \dots, n\}$"). Then*

$$H(\alpha_1, \dots, \alpha_n) \leq \frac{1}{k} \sum_{S \in \mathcal{S}} H(\alpha_S), \qquad \alpha_S = (\alpha_i : i \in S).$$

[1] This gets used in graph theory, but it's not very widely known nowadays. Most proofs of the variational principle that you see will not use this.

Proof. Let $\mathcal{S} = \{S_1, \ldots, S_M\}$, and enumerate $S_m = \{i_{m,1} < \cdots < i_{m,r_m}\}$. Now write

$$
\begin{aligned}
A &= \sum_{m=1}^{M} H(\alpha_{S_m}) \\
&= H(\alpha_{i_1,1}, \ldots, \alpha_{i_1,r_1}) + H(\alpha_{i_2,1}, \ldots, \alpha_{i_2,r_2}) + \cdots + H(\alpha_{i_M,1}, \ldots, \alpha_{i_M,r_M}) \\
&= H(\alpha_{i_1,1}) + H(\alpha_{i_1,2} \mid \alpha_{i_1,1}) + \cdots + H(\alpha_{i_1,r_1} \mid \alpha_{i_1,1}, \ldots, \alpha_{i_1,r_1-1}) \\
&\qquad + \cdots \\
&\qquad + H(\alpha_{i_M,1}) + H(\alpha_{i_M,2} \mid \alpha_{i_M,1}) + \cdots + H(\alpha_{i_M,r_M} \mid \alpha_{i_M,1}, \ldots, \alpha_{i_M,r_M-1}) \\
&\geq H(\alpha_{i_1,1} \mid \alpha_{<i_1,1}) + H(\alpha_{i_1,2} \mid \alpha_{<i_1,2}) + \cdots + H(\alpha_{i_1,r_1} \mid \alpha_{<i_1,r_1}) \\
&\qquad + \cdots \\
&\qquad + H(\alpha_{i_M,1} \mid \alpha_{<i_M,1}) + H(\alpha_{i_M,2} \mid \alpha_{<i_M,2}) + \cdots + H(\alpha_{i_M,r_M} \mid \alpha_{<i_M,r_M}) \\
&\geq k \sum_{i=1}^{n} H(\alpha_i \mid \alpha_{<i}) \\
&= kH(\alpha_1, \ldots, \alpha_n).
\end{aligned}
$$

□

1.3 Variational principle for subshifts

Here is the construction:

Proof. Let $S_n \subseteq X$ be a selection of one preimage in X for each element of π_n. Let

$$\nu^{(n)} = \frac{1}{|S_n|} \sum_{x \in S_n} \delta_x.$$

Bogoliubov-Krylov gives us that if

$$\mu^{(n)} = \frac{1}{n} \sum_{i=0}^{n-1} \sigma_*^i \nu^{(n)},$$

then there is a subsequence $\mu^{(n_i)} \xrightarrow{\text{weak}^*} \mu$. We will show that $h_\mu(\sigma) \geq h_{\text{top}}(\sigma|_X) =: h$. That is, we must show that

$$H_\mu(\underbrace{\alpha_0, \ldots, \alpha_{m-1}}_{=:\alpha_{[0;m-1}}) \geq km \qquad \forall m.$$

Observe that since $\alpha_0, \ldots, \alpha_{m-1}$ is continuous, the left hand side equals

$$\lim_i H_{\mu^{(n_i)}}(\alpha_{[0;m)})$$

Since Shannon entropy is a concave function,

$$
\begin{aligned}
H_{\mu^{(n_i)}}(\alpha_{[0;m)}) &\geq \frac{1}{n_i}\sum_{t=0}^{n_i-1} H_{\sigma_*^t \nu^{(n_i)}}(\alpha_{[0;m)}) \\
&= \frac{1}{n_i}\sum_{t=0}^{n_i-1} H_{\nu^{(n_i)}}(\alpha_{[t;t+m)}) \\
&\geq \frac{1}{n_i}\sum_{t=m}^{n_i-1} H_{\nu^{(n_i)}}(\alpha_{[t,t+m)\cap[0;n_i-1)})
\end{aligned}
$$

These are an m-cover of $[m, n_i - 1)$. By Shearer's inequality,

$$
\begin{aligned}
&\geq \frac{m}{n_i} H_{\nu^{(n_i)}}(\alpha_{[m;n_i-1)}) \\
&\geq \frac{m}{n_i}\left[H_{\nu^{(n_i)}}(\alpha_{[0,n_i-1)}) - m\log(k)\right] \\
&= \frac{m}{n_i}\left[\log|\pi_{n_i}(X)| - m\log(l)\right] \\
&= m h_{\text{top}}(\sigma|_X) + o(1)
\end{aligned}
$$

as $i \to \infty$. $\square$

1.4 Variational principle (general case)

For the general case, we want to show that

$$h_{\text{top}}(X, T) \leq \sup_{\mu} h_\mu(T).$$

Let $\delta > 0$ and choose maximal sets $S_n \subseteq X$ which are $|delta$-separated according to S_n (so $|S_n|$ is the δ-packing number). Let

$$\nu^{(n)} = \frac{1}{|S_n|}\sum_{x \in S_n} \delta_x$$

$$\mu^{(n)} = \frac{1}{n}\sum_{t=0}^{n-1} T_*^i \delta_x,$$

so $\mu^{(n_i)} \xrightarrow{\text{weak}^*} \mu$. We want to pick $\alpha_0 : X \to \{0, \ldots, k-1\}$ and estimate $h_\mu(T, \alpha_0)$. There are 2 complications:

1. $(\alpha_0, \ldots, \alpha_{n-1})$ must be able to distinguish points in S_n: This follows if $\operatorname{diam}(\{\alpha = i\}) < \delta$ for all $i = 0, \ldots, k-1$.

2. α_0 might not be continuous: But it is enough if $\mu(\partial\{\alpha_0 = i\}) = 0$ for each $i = 0, \ldots, k-1$ (by the portmanteau theorem).

Lemma 1.2. *For all μ and $\delta > 0$ there is an α_0 such that each $\{\alpha_0 = i\}$ has diameter $< \delta$ and $\mu(\partial\{\alpha_0 = i\}) = 0$.*

Proof. Here is the idea. Let $x \in X$. $\partial B_r(x) \subseteq \{y : \rho(x, y) = r\}$. If you fix x and vary r, there are uncountably many of them, they are disjoint, and they are closed. So for co-countably many r, $\mu(\partial B_r(x)) = 0$. Choose a finite cover, and chop into a partition. $\square$

The other direction is to show that for all μ,

$$h_{\text{top}} \geq h_\mu(T).$$

Pick $\alpha_0 : X \to \{0, \ldots, k-1\}$ such that

$$H_\mu(\alpha_0, \ldots, \alpha_{n-1}) \geq n h_\mu(T\alpha_0) \qquad \forall n.$$

Find a compact subset G_i in each of the level sets of α_0 such that $\mu(\bigcup_i g_i) > 1 - \varepsilon$. Then let $\delta = \min_{i \neq j} \operatorname{dist}(G_i, G_j)$. To finish, we must used that most $x \in X$ spend most of their time in $\bigcup_i G_i$. If $x, y \in X$ and $\alpha_{[0,n-1)}(x), \alpha_{[0,n-1)}(X)$ differ in some coordinate t at which $T^t x, T^t y \in G$, then $\rho_n(x, y) \geq \delta$.

Sometimes, there exists a μ such that $h_\mu(T) = h_{\text{top}}(\mu, T)$. This is true for subshifts but not always. Such a μ is called a **measure of maximal entropy** or an **equilibrium measure**.

Proposition 1.1. *If (X, T) is expansive, then there exists a measure of maximal entropy.*

Proof. $h_{(\cdot)}(T)$ is upper semicontinuous. $\square$