Math 254A Lecture 25 Notes

Daniel Raban

May 24, 2021

1 The Entropy Rate of Shift-Invariant Measures

1.1 Recap

Our alphabet is $A^{\mathbb{Z}^d}$ as before, and we have been moving around finite windows $W \subseteq \mathbb{Z}^d$ and looking at what patterns appear. The empirical distribution of x in W is

$$P_x^W = \frac{1}{|\{v : v + W \subseteq B\}|} \sum_{v+W \subseteq B} \delta_{v+W} \qquad (x \in A^B).$$

Last time, we saw that if U is an open, convex subset of $P(A^W)$ (or $\mathbb{R}^{A^W}$), then

$$\underbrace{|\{x \in A^B : P_x^W \in U\}|}_{=:\Omega_B(W,U)} = e^{|B| \cdot s(U) + o(|B|)},$$

if $U \cap \{s > -\infty\} \neq \varnothing$ or $\overline{U} \cap \overline{\{s > -\infty\}} = \varnothing$. Here, $s(U) = \sup\{s(x) : x \in U\}$. We have not yet verified that if $U \subseteq U_1 \cup \cdots \cup U_k$, then $s(U) \leq \max_i s(U_i)$, but this is a quick check.

1.2 Counting microscopic configurations by their empirical measures — consistency of the entropy rate

If $W \subseteq W'$, B is large, and $\pi : A^{W'} \to A^W$ is the projection, then

$$\pi_* P_x^{W'} = P_x^W + O\left(\frac{|W|}{\text{min-side-length}(B)}\right)$$

As a result, inside $P(A^{\mathbb{Z}^d})$, consider weak* open sets of the form $\widehat{U} := \{\mu \in P(A^{\mathbb{Z}^d}) : \mu_W \in U\}$ for some finite $W \subseteq \mathbb{Z}^d$ and open convex $U \subseteq P(A^W)$, where $\mu \mapsto \mu_W$ is the projection of μ to A^W. These sets form a base $\mathcal{U}$ for the weak* topology on $P(A^{\mathbb{Z}^d})$.

We would like to try to define

$$s(\widehat{U}) := s(U),$$

where the right hand side is defined using the particular window W. We must show that this is consistent with respect to the choice of W: We want $s^{(W)}(U) = s^{(W')}(U')$ whenever $U \subseteq P(A^W)$ is open and convex and $U' = \{\nu \in P(A^{W'}) : \nu_W \in U\}$. This holds because of the result proven last time:

If U and U' are as above, assume $U \cap \{s^{(W)} > -\infty\} \cap \varnothing$ or $\overline{U} \cap \overline{\{s^{(W)} > -\infty\}} = \varnothing$. This condition implies that

$$\inf_{\delta > 0} s^{(W)}(B_\delta(U)) = s^{(W)}(U) = \sup_{\delta > 0} s^{(W)}(U_\delta).$$

Now observe from the aforementioned result that for any $\delta > 0$, if B is large enough,

$$P_x^W \in U \implies (P_x^{W'})_W = P_x^W + O\left(\frac{|W|}{\text{min-side-length}(B)}\right).$$

Hence,

$$|\Omega_B(W, U)| \leq |\Omega_B(W', U')|,$$

and similarly,

$$|\Omega_B(W, U)| \geq |\Omega_B(W', U')|.$$

Now let $B \uparrow \mathbb{Z}^d$ and then $\delta \downarrow 0$. Then set $s^{(W')}(U') = s^{(W)}(U)$. We then obtain

$$|\Omega_B(\widehat{U})| = \exp\left(|B| \cdot \sup_{\mu \in \widehat{U}} s(\mu) + o(|B|)\right),$$

as $B \uparrow \mathbb{Z}^d$. Interpret $\Omega_B(\widehat{U})$ as $\Omega_B(W, U)$ for any suitable W and U. Note that the left hand side is not precisely well-defined, but it is asymptotically well-defined by these considerations, so this statement still makes sense. This exponent function s is a concave, upper semicontinuous function on $M(A^{\mathbb{Z}^d})$.

1.3 The entropy rate of shift-invariant measures

Proposition 1.1. *Consider the collection of measures*

$$\{\mu \in M(A^{\mathbb{Z}^d}) : s(\mu) > -\infty\} = \{\mu : \forall B_n \uparrow \mathbb{Z}^d,\ \exists x_n \in A^{B_n} \text{ s.t. } P_{x_n}^W \to \mu_W\ \forall W\}.$$

This is contained in

$$P^T(A^{\mathbb{Z}^d}) = \{\mu \in P(A^{\mathbb{Z}^d}) : \text{shift-invariant, i.e. } T_*^v \mu = \mu\ \forall v \in \mathbb{Z}^d\},$$

where $T^v : A^{\mathbb{Z}^d} \to A^{\mathbb{Z}^d}$ *sends* $\langle a_n \rangle_n \mapsto \langle a_{n-v} \rangle_n$ *and* $(T_*^v \mu)(B) = \mu(T^{-v}(B))$ *for all Borel* $B \subseteq A^{\mathbb{Z}^d}$.

Proof. Here is the proof of shift invariance: Suppose $B_n \uparrow \mathbb{Z}^d$ and $x_n \in A^{B_n}$ are such that $P_{x_n}^W \to \mu_W$ for all finite $W \subseteq \mathbb{Z}^d$. Pick a window V and $a \in A^V$. We will show that $\mu_V(A) = \mu_{V-u}(a)$ for all $u \in \mathbb{Z}^d$.

Pick $W \supseteq V \cup (V-u)$, and let $\psi_1, \psi_2 : A^W \to \{0,1\}$ be defined by

$$\psi_1(b) = \mathbb{1}_{\{b_V = a\}}, \qquad \psi_2(b) = \mathbb{1}_{\{b_{V-u}=a\}}.$$

We know $\mu_W = \lim_n P_{x_n}^W$, and so

$$\mu_V(a) = (\mu_W)_V(a) = \lim_n (P_{x_n}^W)_V(a)$$

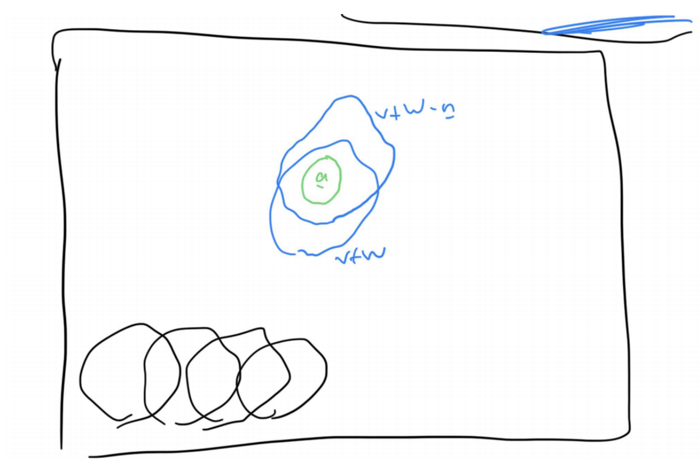
and

$$\mu_{V-u}(a) = \lim_n (P_{x_n}^W)_{V-u}(a).$$

These respectively equal:

$$= \frac{1}{|\{v : v+W \subseteq B_n\}|} |\{v : v+W \subseteq B_n, (x_n)_{v+V} = a\}|,$$

$$= \frac{1}{|\{v : v+W \subseteq B_n\}|} |\{v : v+W \subseteq B_n, (x_n)_{v+V-u} = a\}|,$$



These will agree except for points on the boundary. So the difference is

$$\mu_V(a) - \mu_{V-u}(a) = O\left(\frac{(|v|+|u|)|\text{boundary of } B_n|}{|B_n|}\right) \xrightarrow{n\to\infty} 0.$$

So $T_*^V \mu = \mu$. $\square$

So $\{s > -\infty\} \subseteq P^T(A^{\mathbb{Z}^d})$. We want to generalize the formula "$s(p) = H(p)$ for $p \in P(A)$" from the non-interacting case. To do this we need a digression into the properties of Shannon entropy.

From before, we had that if $p \in P(A)$, then

$$H(p) = -\sum_{a \in A} p(a) \log p(a).$$

Here is some notation: If α is an A-valued random variable and if the distribution of α is p: $\mathbb{P}(\alpha = a) = p(a)$, then $H(\alpha) = H(p)$. We interpret this as a measure of the "uncertainty" in α.

Recall that $0 \leq H(\alpha) \leq \log |A|$, where equality is achieved on the left iff α is deterministic (i.e. $p = \delta_a$ for some letter a) and equality on the right is achieved iff $\alpha \sim \mathrm{Unif}(A)$. Next time, we will discuss some more properties of Shannon entropy and return to $s(\mu)$ for $\mu \in P(A^{\mathbb{Z}^d})$.