Math 254A Lecture 27 Notes

Daniel Raban

May 28, 2021

1 Equality of Entropy Rate and the Exponent Function

1.1 Proving that the entropy rate equals the exponent function for lattice models

In our current setting, we have a shift-invariant measure $\mu \in P^T(A^{\mathbb{Z}^d})$, and

$$s(\mu) = \inf_{W, U \ni \mu_W} \lim_{B \uparrow \mathbb{Z}^d} \frac{1}{|B|} \log |\Omega_B(U)|,$$

where $\Omega_B(U) = \{x \in A^B : P_x^W \in U\}$.

The Shannon entropy is

$$H(\mu_F) = -\sum_{y \in A^F} \mu_F(y) \log \mu_F(y),$$

and the entropy rate is

$$h(\mu) = \lim_B \frac{1}{|B|} H(\mu_B) = \inf_W \frac{1}{|W|} H(\mu_W)$$

Theorem 1.1.

$$s(\mu) = h(\mu).$$

To prove this, we will use two tools from last lecture:

Lemma 1.1. *If $A = B \sqcup C$, $p \in P(A)$, and $p(C) \leq \varepsilon \leq 1/2$, then*

$$H(p) \leq H(\varepsilon, 1-\varepsilon) + (1-\varepsilon) \log |B| + \varepsilon \log |A|.$$

Last time, we assumed $|B| \leq |C|$ in the above and got $\log |C|$ instead of $\log |A|$; this version is more useful. We also have the following corollary of Shearer's inequality:

Lemma 1.2. *If $W, B \subseteq \mathbb{Z}^d$ are finite and $\mu \in P(A^B)$, then*

$$H(\mu) \le \frac{1}{|W|} \sum_{v : v+W \subseteq B} H(\mu_{v+W}) + O\left(\frac{\log|A| \cdot |B| \cdot \operatorname{diam}(W)}{\text{min-side-length}(B)}\right).$$

Now let's prove the theorem:

Proof. We will prove the inequalities $\ge$ and $\le$ separately.

($\ge$): Denote $h = h(\mu)$. We want to show for any W, μ_W, we have

$$\frac{1}{|B|} \log |\{x \in A^B : P_x^W \in U\}| \ge h - o(1)$$

as $B \uparrow \mathbb{Z}^d$. Suppose we knew that

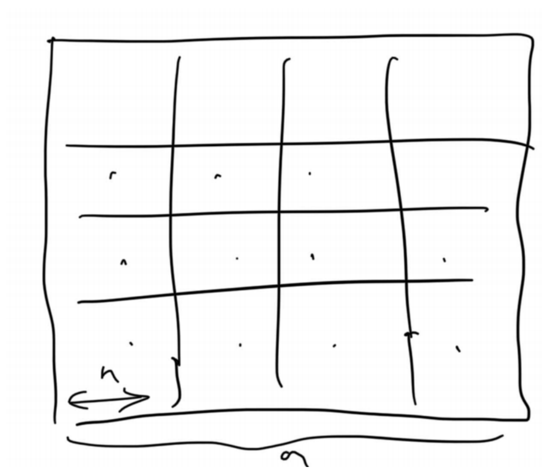
$$\mu_B(\{x \in A^B : P_x^W \in U\}) = 1 - o(1)$$

as $B \uparrow \mathbb{Z}^d$. Then, by the first lemma, we get

$$\begin{aligned}
\frac{1}{|B|} H(\mu_B) &\le \frac{1}{|B|} H(\varepsilon(B), 1-\varepsilon(B)) + \frac{1-\varepsilon(B)}{|B|} \log|\Omega_B(W,U)| + \frac{\varepsilon(B)}{|B|} \log|A^B| \\
&\le \frac{\log 2}{|B|} + \frac{1}{|B|} \log|\Omega_B(W,U)| + \varepsilon(B) \log|A| \\
&= \frac{1}{|B|} \log|\Omega_B(W,U)| + o(1)
\end{aligned}$$

as $B \uparrow \mathbb{Z}^d$. So $s(\mu) \ge h(\mu)$ if we have this property.

In general, this property does not hold, so we need a replacement for it. To do this, we may restrict attention to $B_n = \{0, \dots, n^2 - 1\}^d$. Let $\mathcal{Q}_n$ be the natural partition of B_n into boxed of side length n.



Let $\nu_n = \bigtimes_{Q \in \mathcal{Q}_n} \mu_Q$. Observe that $H(\nu_n) = \sum_W H(\mu_Q)$, so

$$\frac{1}{n^{2d}} H(\nu_n) = \frac{1}{n^d} H(\mu_{\{0,\dots,n-1\}^d}) \to h(\mu)$$

as $n \to \infty$. Also, if $x \in A^{B_n}$,

$$\begin{aligned} P_x^W &= \frac{1}{|\{v : v + W \subseteq B_n\}|} \sum_{v+W \subseteq B_n} \delta_{x_{v+W}} \\ &= \frac{1}{|\{v : v + W \subseteq B_n\}|} \sum_{Q \in \mathcal{Q}_n} \sum_{v+W \subseteq Q} \delta_{x_{v+W}} + \text{boundary terms} \end{aligned}$$

If we do the same analysis as we did before with this type of partition, we get

$$= \frac{1}{n^d} \sum_{Q \in \mathcal{Q}_n} P_{x_Q}^W + o(1).$$

The $P_{x_Q}^W$ are independent if $x \sim \nu_n$. The average of $P_{x_Q}^W(a)$ (with $a \in A^W$) over $x_Q \sim \mu_W$ is $\mu_W(a)$. So by the weak law of large numbers, $P_x^W \in U$ with high probability if $x \sim \nu_n$ and n is large enough. So the property we assumed works if we replace μ_B by ν_n. Now complete the argument as before.

($\leq$): We want to show that if $\varepsilon > 0$, W is large enough, and $U \ni \mu_W$ is small enough, then

$$\frac{1}{|B|} \log |\Omega_B(W, U)| \leq h + \varepsilon + o(1)$$

as $B \uparrow \mathbb{Z}^d$. To estimate the left hand side, let ν_B be the uniform probability distribution on $\Omega_B(W, U)$, so the left hand side equals $\frac{1}{|B|}$. So the second lemma (the corollary of Shearer's inequality) tells us that

$$\frac{1}{|B|} H(\nu_B) \leq \frac{1}{|B|} \sum_{v+v+W \subseteq B} H(\nu_{v+W}) + \underbrace{O\left(\frac{\log |A| \cdot \operatorname{diam}(W)}{\text{min-side-length}(B)} \right)}_{= o_B(1)}.$$

What can we say about the family ν_{v+W}, where $v + W \subseteq B$? Observe that

$$\begin{aligned} \frac{1}{|\{v : v + W \subseteq B\}|} \sum_{v+W \subseteq B} \nu_{v+W} &= \frac{1}{|\{v : v + W \subseteq B\}|} \sum_{v+W \subseteq B} \int \delta_{x_{v+W}} \, d\nu_B(x) \\ &= \int \frac{1}{|\{v : v + W \subseteq B\}|} \sum_{v+W \subseteq B} \delta_{x_{v+W}} \, d\nu_B(x) \\ &= \int P_x^W \, d\nu_B(x) \\ &=: \widehat{\mu} \in U. \end{aligned}$$

Since Shannon entropy is concave and continuous, we get

$$\frac{1}{|\{v : v + W \subseteq B\}|} \sum_{v+W\subseteq B} H(\nu_{v+W}) \le H(\widehat{\mu}) \le H(\mu_W) + \varepsilon$$

if we choose U small enough.

If we put it all together, we get

$$\frac{1}{|B|} \log |\Omega_B(W,U)| = \frac{1}{|B|} H(\nu_B) \le (1 + o_B(1)) \frac{1}{|W|} (H(\mu_W) + \varepsilon) + o_B(1).$$

So for every ε and W, there is a U such that

$$\lim_B \frac{1}{|B|} \log |\Omega_B(W,U)| \le \frac{1}{|W|} H(\mu_W) + \varepsilon.$$

So, if we choose W large enough depending on ε, we get that the left hand side is $\le h + 2\varepsilon$, Since ε is arbitrary, we get $s(\mu) \le h(\mu)$. $\square$

1.2 A digression concerning ergodic measures

If $\mu \in P^T(A^{\mathbb{Z}^d})$, B is a large box, and $x \in A^{\mathbb{Z}^d}$, then

$$\begin{aligned} P_x^W &= \frac{1 + o(1)}{|B|} \sum_{v+W\subseteq B} \delta_{x_{v+W}} \\ &= (1 + o(1)) \cdot \frac{1}{|B|} \sum_{v\in B} \delta_{x_{v+W}}. \end{aligned}$$

When do we have $P_{x_B}^W \to \mu_W$ in weak* as $B \uparrow \mathbb{Z}^d$ when $x \sim \mu$? Equivalently, we test against $\psi : A^{\mathbb{Z}^d} \to \mathbb{R}$ dependent only on coordinates in W: When do we have

$$\mu\left(\left\{x \in A^{\mathbb{Z}^d} : \left|\frac{1}{|B|}\sum_{v\in B} \psi(x_{v+W}) - \int \psi \, d\mu\right| < \varepsilon\right\}\right) \to 1$$

as $B \uparrow \mathbb{Z}^d$? Write $\frac{1}{|B|}\sum_{v\in B} \psi(x_{v+W}) = \frac{1}{|B|}\sum_{v\in B} \psi(T^v x)$. Then we really want

$$\frac{1}{|B|}\sum_{v\in B} \psi \circ T^v \to \int \psi \, d\mu$$

in probability for all ψ.

Theorem 1.2 (Mean Ergodic Theorem)**.** *Let (X, μ) be a probability space (e.g. above $X = A^{\mathbb{Z}^d}$). Let $(T^n)_{n\in\mathbb{Z}^d}$ be an action on X that preserves μ (e.g. above this equals translation). The following are equivalent:*

1. *For all* $\psi \in L^1(\mu)$*, we have*

$$\frac{1}{|B|}\sum_{v\in B}\psi\circ T^v \to \int \psi\, d\mu$$

   in L^1 *as* $B \uparrow \mathbb{Z}^d$*.*

2. *The system* (X,μ,T) *is* ***ergodic****: there is no measurable partition* $X = Y \sqcup Z$ *such that* $T^v(Y) = Y$ *and* $T^v(Z) = Z$ *for all* v *and* $\mu(Y), \mu(Z) > 0$*.*