Ergodic Theory and Measured Group Theory
Lecture 20

Example. A Γ-invariant Borel $X \subseteq \mathbb{N}^{\Gamma}$ admits a generating partition, namely: $P_n := \{x \in X : x(1_\Gamma) = n\}$. Indeed, the itinerary map (under the shift action) is the identity. Up to Borel isomorphism, these are all the examples (by the def. of a generating partition).

Theorem (Weiss for $\mathbb{Z}$, Kechris for Γ). Any aperiodic Borel action $\Gamma \curvearrowright X$ of a ctbl group Γ admits a ctbl generating partition (typically infinite).

Thus a more instructive notion is that of a finite generating partition, which may not exist in general. If it does exist for say a $\mathbb{Z}$-action, then whatever the static entropy of that partition is, it is the most clever partition for the dynamical 20-questions game.

Def. Let $\mathbb{N} \curvearrowright (X, \mu)$ be a pmp action by a transform. T. Let $\mathcal{P}$ be a finite partition of X into measurable pieces. The *dynamic entropy* of $\mathcal{P}$ wrt T is:

$$h(\mathcal{P},T) := \overset{\overset{\inf_n}{\shortparallel}}{\lim_{n\to\infty}} \frac{1}{n}\, h(\mathcal{P} \overset{\swarrow \text{common refinement}}{\vee} T^{-1}\mathcal{P} \vee T^{-2}\mathcal{P} \vee \dots \vee T^{-(n-1)}\mathcal{P})$$

This measures the expected info per day on average (over days) gained by Player 2 if they play $\mathcal{P}$. Thus, the closer $h(\mathcal{P},T)$ is to $\log|\mathcal{P}|$, the smarter/more informative is the partition $\mathcal{P}$. The entropy of T is:

$$h(T) := \sup_{\mathcal{P}} h(\mathcal{P},T),$$

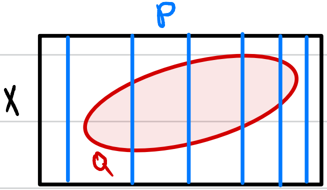
where $\mathcal{P}$ ranges over all finite partitions.
This sup may not be achieved, in fact, it may be ∞.

Let's understand why the quantity $\frac{1}{n} h(\bigvee_{i<n} T^{-i}\mathcal{P})$ is indeed the average expected info gained in a day, and why the limit exists.

Lemma (Fekete). If (a_n) is a subadditive sequence of nonnegative reals, i.e. $a_{n+m} \leq a_n + a_m$, then the limit $\lim_n \frac{a_n}{n}$ exist and is equal to $\inf_n$.

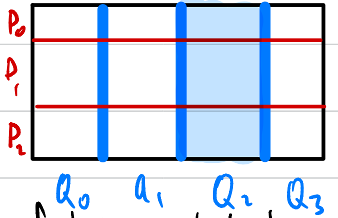
Lemma. If $\mathcal{P}$ & $\mathcal{Q}$ are two finite partitions of X, then $h(\mathcal{P} \vee \mathcal{Q}) \leq h(\mathcal{P}) + h(\mathcal{Q})$.

For the proof of this, also for its own sake, we introduce the notions of conditional information and entropy. For a set $Q \subseteq X$, define $\mu_Q := \frac{1}{\mu(Q)} \cdot \mu|_Q$, which makes (X, μ_Q) into a prob. space. For any finite partition $\mathcal{P}$ of X, the entropy of $\mathcal{P}$ wrt μ_Q is $h_{\mu_Q}(\mathcal{P}) = -\sum_{P \in \mathcal{P}} \mu_Q(P) \cdot \log \mu_Q(P)$. We think of this as the expected info given by learning which piece of $\mathcal{P}$ a random $x \in X$ is, if we already know that $x \in Q$. Now for a partition



$\mathcal{Q}$ of X, the *expected info of $\mathcal{P}$ conditioned on $\mathcal{Q}$* is $i_{\mathcal{P}|\mathcal{Q}} : X \to \mathbb{R}^+$ given by $i_{\mathcal{P}|\mathcal{Q}}(x) := h_{\mu_Q}(\mathcal{P})$ for $x \in Q \in \mathcal{Q}$. The *entropy of $\mathcal{P}$ conditioned on $\mathcal{Q}$* is then the expectation of $i_{\mathcal{P}|\mathcal{Q}}$: $h(\mathcal{P}|\mathcal{Q}) := \mathbb{E}(i_{\mathcal{P}|\mathcal{Q}}) = \int_X i_{\mathcal{P}|\mathcal{Q}} \, d\mu = \sum_{Q \in \mathcal{Q}} h_{\mu_Q}(\mathcal{P}) \cdot \mu(Q)$

$$= -\sum_{Q \in \mathcal{Q}} \sum_{P \in \mathcal{P}} \mu_Q(P) \cdot \log \mu_Q(P) \cdot \mu(Q) = -\sum_{Q \in \mathcal{Q}} \sum_{P \in \mathcal{P}} \mu(P \cap Q) \cdot \log \frac{\mu(P \cap Q)}{\mu(Q)}.$$



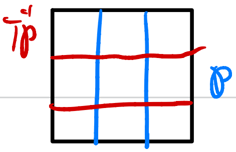
Lemma: (a) $h(\mathcal{P} \vee \mathcal{Q}) = h(\mathcal{P}|\mathcal{Q}) + h(\mathcal{Q})$.

(b) $h(\mathcal{P}|\mathcal{Q}) \leq h(\mathcal{P})$.

Proof. (a) is by def and (b) is by the concavity of $t \mapsto -t \log t$. □

(a) + (b) $\Rightarrow h(\mathcal{P} \vee \mathcal{Q}) \leq h(\mathcal{P}) + h(\mathcal{Q})$, which implies that the sequence $a_n := h(\bigvee_{i<n} T^{-i}\mathcal{P})$ is subadditive, so by Fekete's lemma, $\lim_n \frac{a_n}{n} = \inf_n \frac{a_n}{n}$ exists. Moreover, note that the new info gained on day $k \in \mathbb{N}$ is $h(T^{-k}\mathcal{P} | \bigvee_{i<k} T^{-i}\mathcal{P})$, so $\frac{1}{n} h(\bigvee_{i<n} T^{-i}\mathcal{P}) = \frac{1}{n} \sum_{k<n} h(T^{-k}\mathcal{P} | \bigvee_{i<k} T^{-i}\mathcal{P})$, which justifies thinking of this as the average info gained over n days.

Obs. $h(T^{-n}\mathcal{P}) = h(\mathcal{P})$ (because T is pmp).



Example. For a Bernoulli shift $(k^{\mathbb{Z}}, \nu^{\mathbb{Z}})$, ν a prob. measure on k, taking as P the base partition

$$P_i := \{x \in k^{\mathbb{Z}} : x(0) = i\}, \quad i \in k,$$

we see that

$$h(T^{-1}P \mid P) \underset{\text{by independence}}{=} h(T^{-1}\mathcal{P}) \underset{\text{pmp}}{=} h(\mathcal{P}).$$

Hence, $h\left(\bigvee_{i<n} T^{-i}\mathcal{P}\right) = n \cdot h(\mathcal{P})$, so

$$h(\mathcal{P},T) = h(\mathcal{P}) = h(\nu) := -\sum_{i<k} \nu(i) \log \nu(i).$$

In particular, if ν is the uniform $\frac{1}{k}$ measure, then $h(\mathcal{P},T) = h(\mathcal{P}) = \log k$.

How does one find $h(T)$? Also, if $\mathcal{P}$ is a finite generating partition, what's the relationship between $h(T)$ and $h(\mathcal{P},T)$?

Theorem (Kolmogorov – Sinai, 1958-59). If $\mathcal{P}$ is a finite generating partition, then $h(T) = h(\mathcal{P},T)$.

This follows from the more general fact:

Theorem (Abramov). If $(\mathcal{P}_n)_{n\in\mathbb{N}}$ is a sequence of finite partitions s.t.

$\mathcal{P}_{n+1}$ refines $\mathcal{P}_n$ and $\bigcup_{n\in\mathbb{N}} \mathcal{P}_n$ generates $\mathcal{B}(X)$, then

$$h(T) = \lim_n h(\mathcal{P}_n, T).$$

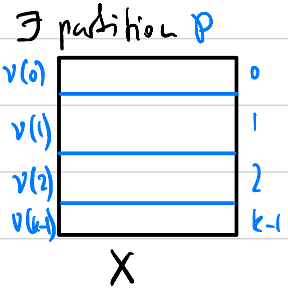
Indeed, if $\mathcal{P}$ is a finite generating partition then taking $\mathcal{P}_n := \bigvee_{i<n} T^{-i}\mathcal{P}$ works. Recalling that there always exists an infinite ctbl generating partition $\mathcal{Q} := (Q_i)_{i\in\mathbb{N}}$, we can use Abramov's theorem to compute $h(T)$. Indeed, take $\mathcal{P}_n := \bigvee_{i<n} T^{-i}\tilde{\mathcal{Q}}_n$, where $\tilde{\mathcal{Q}}_n := \{Q_0, Q_1, \dots, Q_{n-1}, \bigcup_{i=n}^{\infty} Q_i\}$.

Obs. If a pmp action $\mathbb{Z} \overset{T}{\curvearrowright} (X,\mu)$ factors onto another action $\mathbb{Z} \overset{S}{\curvearrowright} (Y,\lambda)$, then $h(T) \geq h(S)$.

Proof. Any partition of Y pulls back to a partition of X of the same entropy. □

When (Y,λ) is a finite Bernoulli shift, the converse is true!

Sinai's theorem (1964). For a pmp action $\mathbb{Z} \overset{T}{\curvearrowright} (X,\mu)$, if $h(T) \geq h(\nu)$ for some measure on a finite set k, then the action factors onto the corresponding Bernoulli shift $\mathbb{Z} \curvearrowright (k^{\mathbb{Z}}, \nu^{\mathbb{Z}})$. In other words, there is a partition $\mathcal{P} = (P_i)_{i<k}$ of X with $\mu(P_i) = \nu(i)$ and the itinerary map $\pi_{\mathcal{P}} : X \to k^{\mathbb{Z}}$ for $\mathcal{P}$ maps μ to $\nu^{\mathbb{Z}}$.



This theorem shows that the entropy is large because the action encodes (weakly contains) a Bernoulli action.

The partitions constructed in Sinai's theorem need not be generating. This makes sense since a finite generating partition $\mathcal{P}$ bounds $h(T)$ from above (as opposed to below) by the Kolmogorov–Sinai theorem:

$$h(T) = h(\mathcal{P}, T) \leq \log |\mathcal{P}|.$$

Krieger showed that the converse of this is true!

Krieger's theorem (1970). If $h(T) < \log n$, then $\exists$ a generating partition of size n.

This shows that entropy is small (not infinite) only because there is a finite generating partition (i.e. Player 2 has a winning strategy in the dynamical 20-questions game).