Math 254B Lecture 4 Notes

Daniel Raban

April 8, 2019

1 Fano's Inequality, Conditional Entropy, and Introduction to Ergodic Theory

1.1 Fano's inequality

Lemma 1.1 (Fano's inequality)**.** *Let $p = \mathbb{P}(\alpha \neq \beta)$ be the "probability of error." Then*

$$H(\alpha) - H(\beta) \leq H(\alpha \mid \beta) \leq H(p, 1-p) + p \log(|\mathcal{X}| - 1),$$

where $p, 1-p$ is a distribution on $\{0,1\}$ with probability $p, 1-p$.

Proof. Introduce ξ valued in $\{0,1\}$ by

$$\xi = \begin{cases} 1 & \alpha = \beta \\ 0 & \alpha - \beta. \end{cases}$$

Then $\mathbb{P}(\xi = 1) = 1 - p$. Then

$$H(\alpha \mid \beta) \leq H(\alpha, \xi \mid \beta) = \underbrace{H(\xi \mid \beta)}_{\leq H(\xi) = H(p, 1-p)} + H(\alpha \mid \xi, \beta)$$

The right hand side is equal to

$$\cancel{\mathbb{P}(\xi = 1) H_{\mathbb{P}(\cdot \mid \xi = 1)}(\alpha \mid \beta)}^{0} + \underbrace{\mathbb{P}(\xi = 0)}_{p} \underbrace{H_{\mathbb{P}(\cdot \mid \xi = 0)}(\alpha \mid \beta)}_{\leq \log(|\mathcal{X}| - 1)}.$$

□

1.2 Entropy conditioned on a σ-algebra

Definition 1.1. Let α take values in a finite set, and let $\mathcal{G} \subseteq \mathcal{F}$ be a σ-algebra. The **conditional entropy** $H(\alpha \mid \mathcal{G})$ is

$$H(\alpha \mid \mathcal{G}) := \inf\{H(\alpha \mid \beta) : \beta \text{ is } \mathcal{G}\text{-measurable}\}.$$

How can we actually determine this value?

Proposition 1.1. *Suppose $\mathcal{G} = \sigma(\beta_1, \beta_2, \dots)$. Then for all α,*

$$H(\alpha \mid \mathcal{G}) = \lim_n H(\alpha \mid \beta_1, \dots, \beta_m).$$

Proof. The limit exists, and $\leq$ follows. So we need to prove $\geq$. Fix a $\mathcal{G}$-measurable β taking values in $Y = \{y_1, \dots, y_k\}$. Claim: for any $\varepsilon > 0$, there exist m and $\beta' : \Omega \to Y$ such that β' is determined by $(\beta_1, \dots, \beta_m)$ and $\mathbb{P}(\beta \neq \beta') < \varepsilon$. First, if $E \in \mathcal{G}$¡ then for all $\varepsilon > 0$¡ there exists m and $E' \in \sigma(\beta_1, \dots, \beta_m)$ such that $\mathbb{P}(E \Delta E') < \varepsilon$ (as such Es form a σ-algebra). Choose m and $F_1, \dots, F_{k-1} \in \sigma(\beta_1, \dots, \beta_m)$ such that $\mathbb{P}(\{\beta = y_i\} \Delta F_i) < \varepsilon/(k-1)$. Let $E_1' : F_1$, $E_2' := F_2 \setminus F_1, \dots E_{k-1}' := F_{k-1} \setminus (F_1 \cup \dots \cup F_{k-1})$, $E_k = \Omega \setminus (F_1 \cup \dots \cup F_{k-1})$. Define β' by $\beta'(\omega) = y_i$ fi $\omega \in E_i$. We have

$$\mathbb{P}(\{\beta \neq \beta'\}) \leq \sum_{i=1}^{k-1} \mathbb{P}(\{\beta = y_i\} \Delta F_i) < \varepsilon.$$

Let $\varepsilon > 0$, and choose m and $\sigma(\beta_1, \dots, \beta_m)$-measurable β' as above. Now

$$\begin{aligned} H(\alpha \mid \beta_1, \dots, \beta_m) &= H(\alpha \mid \beta_1, \dots, \beta_m, \beta') \\ &\leq H(\alpha \mid \beta') \\ &\leq H(\alpha, \beta \mid \beta') \\ &= H(\beta \mid \beta') + H(\alpha \mid \beta', \beta) \\ &\leq H(\varepsilon, 1 - \varepsilon) + \varepsilon \log(|Y| - 1) + H(\alpha \mid \beta). \end{aligned}$$

□

Example 1.1. The Borel σ-algebra on $[0, 1]$ is generated by (B_n), where B_n is uniform on $\{0, 1\}$ and represents the n-th binary digit.

1.3 Introduction to Ergodic theory

In dynamical systems, we have a state space X and a transformation $T : X \to X$ which tells you what happens after a unit of time. Different kinds of dynamics arise when you talk about preserving different kinds of structure on X.

Example 1.2. Let X be a smooth manifold and T be a smooth map. This is called smooth dynamics.

Example 1.3. Let X be a topological space and T be continuous. This is called topological dynamics.

Example 1.4. Let $X \subseteq \mathbb{C}$ be open and T be holomorphic. This is called complex dynamics.

Ergodic theory studies long-run statistical properties in dynamics. The key object is a T-invariant probability measure on X.

Definition 1.2. A **measure-preserving system (MPS)** is a triple (X, μ, T), where X is a measurable space, $\mu \in P(X)$, and $T : X \to X$ is measurable such that $\mu(T^{-1}[A]) = \mu(A)$ for all measurable $A \subseteq X$ (i.e. $T_*\mu = \mu$). We say T is **invertible** if there exists a measurable $S : X \to X$ with $S_*\mu = \mu$ such that $S \circ T = T \circ S = \mathrm{id}_X$ a.e.

Example 1.5. Let $\mathbb{T} = \mathbb{R}/\mathbb{Z}$. Fix $\alpha \in \mathbb{T}$, and define R_α sending $x \mapsto x + \alpha$. Similarly, we can look at $\mathbb{T}^d$. Lebesgue measure on the unit-interval is T-invariant.

Example 1.6. Let $k \geq 2$. Then let $T_{\times k} : \mathbb{T} \to \mathbb{T}$ send $x \mapsto kx$. Lebesgue measure is $T_{\times k}$-invariant.

These two are very different systems in terms of the dynamics. Here is how this relates to probability theory:

Let $\mathcal{X}$ be a finite set, and let $(\xi_n)_{n\in\mathbb{N}}$ be an $\mathcal{X}$-valued process.

Definition 1.3. $(\xi_n)_n$ is **stationary** if for all $k \geq 1$, the distribution of $(\xi_n, \ldots, \xi_{n+k-1}) \in P(\mathcal{X}^k)$ is independent of n.

Lemma 1.2. *Define $T : \mathcal{X}^{\mathbb{N}} \to \mathcal{X}^{\mathbb{N}}$ sending $(x_n)_n \mapsto (x_{n+1})_n$. Then $(\xi_n)_n$ is stationary if and only if its joint distribution μ is T-invariant.*

We will call this a **shift system**.