Math 254A Lecture 2 Notes

Daniel Raban

March 30, 2021

1 Counting Empirical Distributions Close to a Given Distribution

1.1 Easier upper bound for the size of a type class

Recall our setting: A is a finite alphabet, and for $x \in A^n$, $p_x(a) = \frac{|\{i \leq n : x_i = a\}|}{n}$ is the empirical distribution. The type class is

$$T_n(p) = \{x \in A^n : p_x = p\}.$$

Last time, we used Stirling's approximation to show that $|T_n(p)| = e^{H(p)n + o(n)}$, where $H(p) = -\sum_a p(a) \log p(a)$.

Today we will focus on a variant of the question: counting how many empirical distributions are close to p. We will prove an alternative proof that $|T_n(p)| \leq e^{H(p)n}$, the arguments for which will help us in the later analytic case when there is no exact answer.

Proposition 1.1. $|T_n(p)| \leq e^{H(p)n}$.

Proof. Choose $X \in A^n$ at random with iid p coordinates, i.e. the law of x is $p^{\times n}$. Given $x \in T_n(p)$, then

$$\begin{aligned}
\mathbb{P}(X = x) &= \prod_{i=1}^n p(x_i) \\
&= \exp\left(\sum_{i=1}^n \log p(x_i)\right) \\
&= \exp\left(\sum_a p_x(a) \cdot n \cdot \log p(a)\right) \\
&= \exp\left(n \sum_a p(a) \log p(a)\right)
\end{aligned}$$

$$= e^{-H(p)n}.$$

So

$$1 \geq \mathbb{P}(x \in T_n(p)) = \sum_{x \in T_n(p)} \mathbb{P}(X = x) = |T_n(p)| e^{-H(p)n}. \qquad \square$$

Remark 1.1. It's also true that $|T_n(p)| \geq e^{H(p)n - o(n)}$ if $p(a) \in \mathbb{N}/n$ for all a.

1.2 Asymptotic analysis of number of empirical distributions close to p

Next, we estimate the size of

$$T_{n,\delta}(p) = \{x \in A^n : \|p_x - p\| < \delta\}.$$

Proposition 1.2. *For any $\varepsilon > 0$ and $p \in P(A)$, there is a $\delta_0 > 0$ such that for all $\delta \in (0, \delta_0)$, we have*

$$e^{H(p)n - \varepsilon n - o(n)} \leq |T_{n,\delta}(p)| \leq e^{H(p)n + \varepsilon n + o(n)}.$$

Proof. (Upper bound):

$$T_{n,\delta}(p) = \bigcup_{\substack{\|q-p\|<\delta \\ nq(a) \in \mathbb{N}\ \forall a}} T_n(q),$$

so

$$|T_{n,\delta}(p)| \leq \sum_q |T_n(q)| \leq \sum_q e^{H(q)n}.$$

H is continuous on $\mathbb{P}(A)$, so there exists a δ_0 such that $\|q - p\| < \delta_0 \implies H(q) < H(p) + \varepsilon$, and then

$$\begin{aligned}
|T_{n,\delta}(p)| &\leq e^{H(p)n + \varepsilon n} |\{q \in P(A) : \|q - p\| < \delta, nq(a) \in \mathbb{N}\ \forall a\}| \\
&\leq (n+1)^{|A|} e^{H(p)n + \varepsilon n} \\
&= e^{H(p)n + \varepsilon n + o(n)}.
\end{aligned}$$

(Lower bound): If $X \sim p^{\times n}$, so

$$\begin{aligned}
\mathbb{P}(X \in T_{n,\delta}(p)) &= \mathbb{P}\left(\sum_a |p_X(a) - p(a)| < \delta\right) \\
&= \mathbb{P}\left(\sum_a \left|\frac{|\{i : X_i = a\}|}{n} - p(a)\right| < \delta\right)
\end{aligned}$$

$$= \mathbb{P}\left(\sum_a \left| \frac{\sum_{i=1}^n \mathbb{1}_{\{X_i = a\}}}{n} - p(a) \right| < \delta \right).$$

The $\mathbb{1}_{\{X_i=a\}}$ are iid Bernoulli random variables with mean $p(a)$, so by the Weak Law of Large Numbers, this stays $< \delta/|A|$ with high probability as $n \to \infty$. So this probability equals $1 - o(1)$. So we must have

$$\sum_{x \in T_{n,\delta}(p)} \underbrace{\mathbb{P}(X = x)}_{= e^{-n \sum_a p_x(a) \log p(a)}} = 1 - o(1).$$

Observe that for any $\varepsilon > 0$, there exists a δ such that $\|p_x - p\| < \delta \implies \sum_a p_x(a) \log p(a) \leq \sum_a p(a) \log p(a) + \varepsilon$. So for this δ, we get

$$|T_{n,\delta}(p)| e^{-H(p)n + \varepsilon n} \geq \mathbb{P}(X \in T_{n,\delta}(p)) = 1 - o(1),$$

and so $|T_{n,\delta}(p)| \geq e^{H(p)n - \varepsilon n - o(n)}$. $\square$

1.3 Superadditivity and convexity arguments for counting type classes of sets

What we've done is specify a ball in the space of empirical distributions and calculated how many distributions end up in the ball. Here is an approach that does not rely on an exact answer. Given $U \subseteq P(A)$, let $T_n(U) = \{x \in A^n : p_x \in U\}$ and $S_n(U) := \log |T_n(U)|$. Here is a key fact.

Proposition 1.3. *If U is convex, then $S_{n+m}(U) \geq S_n(U) + S_m(U)$ for all n, m; i.e. $S_\cdot(U)$ is **superadditive**.*

Proof. Suppose $x \in T_n(U)$ and $y \in T_m(U)$. Then

$$p_{(x,y)}(a) = \frac{n}{n+m} p_x(a) + \frac{m}{n+m} p_y(a),$$

so $p_{(x,y)} \in U$ by convexity of U. So $T_n(U) \times T_m(U) \subseteq T_{n+m}(U)$. This gives $|T_n(U)| \cdot |T_m(U)| \leq |T_{n+m}(U)|$. Now take log. $\square$

Lemma 1.1 (Fekete)**.** *Suppose $a_n \in \mathbb{R}$ for all n is superadditive: $a_{n+m} \geq a_n + a_m$. Then*

$$\lim_n \frac{a_n}{n} = \sup_n \frac{a_n}{n} \in (-\infty, \infty].$$

Proof. By iterating this condition, $a_n \geq n a_1$ for all n. Rearrange this to $a_n/n \geq a_1$ for all n. Now suppose that $c < \sup_n a_n/n$. We will show that $a_n/n > c$ for all sufficiently large

n. Choose m such that $a_m/m > c$. Now consider $n \gg m$ such that $n = km + p$, where $k \geq 1$ and $0 \leq p < m$. Then $a_n \geq ka_m + a_p$, so

$$\frac{a_n}{n} \geq \frac{k}{km+p} a_m + \frac{p}{km+p} a_1 = \underbrace{\frac{km}{km+p}}_{\xrightarrow{n\to\infty} 1} \underbrace{\frac{a_m}{m}}_{>c} + \underbrace{\frac{p}{km+p}}_{\xrightarrow{n\to\infty} 0} a_1. \qquad \square$$

Corollary 1.1. *If $U \subseteq P(A)$ is convex, then $S(U) := \lim_n \frac{1}{n} S_n(U)$ exists; i.e. $|T_n(U)| = e^{S(U)n + o(n)}$.*

Next, we will derive properties of S.

Lemma 1.2. *If $U \subseteq V$, then $S(U) \leq S(V)$.*

Here is somewhat of an improvement:

Lemma 1.3. *If $U \subseteq U_1 \cup \cdots \cup U_k$, then $S(U) \leq \max_i S(U_i)$.*

Proof.

$$|T_n(U)| \leq \sum_i |T_n(U_i)| \leq k \cdot \max_i |T_n(U_i)|,$$

so

$$\frac{1}{n} S_n(U) \leq \frac{\log k}{n} + \max_i \frac{1}{n} S_n(U_i).$$

Now let $n \to \infty$. $\square$

How can a function of convex sets U be like this?

Example 1.1. Let $\widetilde{S} : P(A) \to \mathbb{R}$ be continuous, and let $S(U) = \sup\{\widetilde{S}(p) : p \in U\}$. This example will have the property in the above lemma.

Next time, we will give conditions on S for it to have this form. When we come to the analytic case, we will be able to repeat this analysis without needing to know the exact value of S.