18.317 Combinatorics, Probability, and Computations on Groups 24 September 2001

Lecture 7

Lecturer: Igor Pak *Scribe: Etienne Rassart*

The Erdös-Rényi "Machine"

Let $g_1, g_2, \ldots, g_k \in G$ and consider $h = {g_1}^{\varepsilon_1} \cdots {g_k}^{\varepsilon_k}$, where the $\varepsilon_i \in \{0,1\}$ are i.i.d. random variables. These h are called *random subproducts.*

A theorem of Erdös and Rényi shows that when k is large, the distributions of the h becomes "close" to the uniform distribution on G.

Definition 1 *Pick $\bar{g} = (g_1, \ldots, g_k)$ uniformly in G^k, and fix it. We can then define the probability distribution*

$$Q_{\bar{g}}(h) = \mathrm{Pr}_{\bar{\varepsilon}}[{g_1}^{\varepsilon_1} \cdots {g_k}^{\varepsilon_k} = h]\,,$$

where the $\varepsilon_i \in \{0,1\}$ are i.i.d. random variables and $\bar{\varepsilon} = (\varepsilon_1, \ldots, \varepsilon_k)$.

Theorem 2 (Erdős-Rényi, 1965)

For all $\epsilon, \delta > 0$, we have

$$\mathrm{Pr}_{\bar{g}}\left[\max_{h\in G}\left|Q_{\bar{g}}(h) - \frac{1}{|G|}\right| < \frac{\epsilon}{|G|}\right] > 1 - \delta$$

for $k > 2\log_2|G| + 2\log_2\frac{1}{\epsilon} + \log_2\frac{1}{\delta} + 1$.

We first need the following lemma.

Lemma 3

$$\mathbf{E}_{\bar{g}}\left[\sum_{h\in G}\left(Q_{\bar{g}}(h) - \frac{1}{|G|}\right)^2\right] = \frac{1}{2^k}\left(1 - \frac{1}{|G|}\right).$$

Proof: (Lemma)

From the usual formula for the variance $\mathbf{E}\left[(X - \mathbf{E}[X])^2\right] = \mathbf{E}[X^2] - \mathbf{E}[X]^2$, we first get that

$$\mathbf{E}_{\bar{g}}\left[\sum_{h\in G}\left(Q_{\bar{g}}(h) - \frac{1}{|G|}\right)^2\right] = \mathbf{E}_{\bar{g}}\left[\sum_{h\in G} Q_{\bar{g}}(h)^2\right] - \frac{1}{|G|}\,,$$

since

$$\mathbf{E}_{\bar{g}}[Q_{\bar{g}}(h)] = \frac{1}{|G|} \qquad \forall\, h \in G\,.$$

The next step is to observe that for $\bar{g}$ fixed in G^k,

$$Q_{\bar{g}}(h) = \frac{1}{2^k}\sum_{\bar{\varepsilon}:\bar{g}^{\bar{\varepsilon}}=h} 1\,,$$

and thus that

$$\sum_{h \in G} Q_{\bar{g}}(h)^2 = \frac{1}{2^{2k}} \sum_{\bar{\varepsilon}, \bar{\varepsilon'} : \bar{g}^{\bar{\varepsilon}} = \bar{g}^{\bar{\varepsilon'}}} 1 \,,$$

So when we let $\bar{g}$ be variable again and take the expectation over G^k, we get

$$\mathbf{E}_{\bar{g}} \left[\sum_{h \in G} Q_{\bar{g}}(h)^2 \right] = \frac{1}{2^{2k}} \sum_{\bar{\varepsilon}, \bar{\varepsilon'} : \{0,1\}^k} \mathrm{Pr}_{\bar{g}}[{g_1}^{\varepsilon_1} \cdots {g_k}^{\varepsilon_k} = {g_1}^{\varepsilon'_1} \cdots {g_k}^{\varepsilon'_k}] \,.$$

The next observation is that

$$\mathrm{Pr}_{\bar{g}}[\bar{g}^{\bar{\varepsilon}} = \bar{g}^{\bar{\varepsilon'}}] = \begin{cases} 1 & \text{if } \bar{\varepsilon} = \bar{\varepsilon'} \\ \dfrac{1}{|G|} & \text{otherwise.} \end{cases}$$

Hence

$$\begin{aligned} \mathbf{E}_{\bar{g}} \left[\sum_{h \in G} \left(Q_{\bar{g}}(h) - \frac{1}{|G|} \right)^2 \right] &= \mathbf{E}_{\bar{g}} \left[\sum_{h \in G} Q_{\bar{g}}(h)^2 \right] - \frac{1}{|G|} \\ &= \frac{1}{2^{2k}} \sum_{\bar{\varepsilon}, \bar{\varepsilon'} \in \{0,1\}^k} \mathrm{Pr}_{\bar{g}}[\bar{g}^{\bar{\varepsilon}} = \bar{g}^{\bar{\varepsilon'}}] - \frac{1}{|G|} \\ &= \frac{1}{2^{2k}} \left(\sum_{\bar{\varepsilon} = \bar{\varepsilon'}} 1 + \sum_{\bar{\varepsilon} \neq \bar{\varepsilon'}} \frac{1}{|G|} \right) - \frac{1}{|G|} \\ &= \frac{1}{2^{2k}} \left(2^k + (2^{2k} - 2^k) \frac{1}{|G|} \right) - \frac{1}{|G|} \\ &= \frac{1}{2^k} \left(1 - \frac{1}{|G|} \right) \,. \end{aligned}$$

■

Proof: (Theorem)

First observe that

$$\max_{h \in G} \left(Q_{\bar{g}}(h) - \frac{1}{|G|} \right)^2 \leq \sum_{h \in G} \left(Q_{\bar{g}}(h) - \frac{1}{|G|} \right)^2 \,.$$

Therefore

$$\mathrm{Pr}_{\bar{g}} \left[\max_{h \in G} \left(Q_{\bar{g}}(h) - \frac{1}{|G|} \right) > \frac{\epsilon}{|G|} \right] \leq \mathrm{Pr}_{\bar{g}} \left[\sum_{h \in G} \left(Q_{\bar{g}}(h) - \frac{1}{|G|} \right)^2 > \frac{\epsilon^2}{|G|^2} \right] \,.$$

If we let $X = \sum_{h \in G} \left(Q_{\bar{g}}(h) - \frac{1}{|G|} \right)^2$, then $\mathbf{E}_{\bar{g}}[X] = \frac{1}{2^k} \left(1 - \frac{1}{|G|} \right)$ by the previous lemma.

We can then use Markov's inequality $\mathrm{Pr}[X > \lambda \mathbf{E}[X]] < \frac{1}{\lambda}$ with X as above and $\lambda = \frac{\epsilon^2}{|G|^2 \mathbf{E}[X]}$ to get

$$\mathrm{Pr}_{\bar{g}} \left[\max_{h \in G} \left(Q_{\bar{g}}(h) - \frac{1}{|G|} \right) > \frac{\epsilon}{|G|} \right] \leq \mathrm{Pr}_{\bar{g}} \left[\sum_{h \in G} \left(Q_{\bar{g}}(h) - \frac{1}{|G|} \right)^2 > \frac{\epsilon^2}{|G|^2} \right] < \frac{|G|^2}{2^k \left(1 - \frac{1}{|G|} \right) \epsilon^2} < \frac{|G|^2}{2^{k-1} \epsilon^2} \,.$$

In particular, this will be less than δ if

$$2^{k-1} > \frac{|G|^2}{\delta \epsilon^2} \,,$$

or

$$k > 2 \log_2 |G| - 2 \log_2 \epsilon - \log_2 \delta + 1 \,.$$

■