The Probabilistic Method

Instructor: Noga Alon
Notes by Nilava Metya

Spring 2024

Contents

Lecture 1

01/30/2024

1.1 Philosophy

Main philosophy of the probabilistic method: To prove existence of a structure (or a substructure of a given one), define a probability space of structures, and show that a random point in it satisfies the required properties with positive (often high) probability.

We will look at two examples today.

1.2 Example: Graph Theory

Definition 1 (Ramsey numbers)
For $k, \ell \geq 1$, let $r = r(k, \ell)$ be the smallest integer, if there exists any, satisfying the following property: for every coloring of edges of $G = K_r$ (the complete graph on r nodes) by red and blue, either $\exists$ a blue $K_k \subseteq G$ or a red $K_\ell \subseteq G$.

Example 1. $r(3, 3) = 6$.

A special case of Ramsey's theorem says that $\exists r(k, l) < \infty \forall k, l$. The proof, by induction (using Erdös-Szekeres theorem), gives $r(k, \ell) \leq \begin{pmatrix} k + \ell - 2 \\ k - 1 \end{pmatrix}$. In particular, $r(k, k) \leq \binom{2k-2}{k-1} < 4^k$.

Remark 1
The following are easy to observe: $r(k, \ell) = r(l, k), r(1, \ell) = 1, r(2, \ell) = \ell$.

All the exactly known Ramsey numbers for $\ell \geq k \geq 3$ are $r(3, 3) = 6, r(3, 4) = 9, r(3, 5) = 14, r(3, 6) = 18, r(3, 7) = 23, r(3, 8) = 28, r(3, 9) = 36, r(4, 4) = 18, r(4, 5) = 25$. It is only known that $41 \leq r(3, 10) \leq 42, 36 \leq r(4, 6) \leq 40, 43 \leq r(5, 5) \leq 48$, and some similar bounds for other Ramsey numbers.

Theorem 2 (Erdos '47)
If $\binom{n}{k} 2^{1-\binom{k}{2}} < 1$ then $r(k,k) > n$. Therefore $r(k,k) \geq [1-o(1)] \frac{k}{e} 2^{\frac{k-1}{2}}$.

Proof. Take the complete graph on n labelled vertices $[n] = \{1, \cdots, n\}$. Color each edge $\{i, j\}$ (for $1 \leq i < j \leq n$) randomly uniformly and independently either red or blue. For fixed $K \subseteq [n]$ with $k = |K|$, the probability that the graph induced by K is monochromatic is $2^{-\binom{k}{2}} + 2^{-\binom{k}{2}} = 2^{1-\binom{k}{2}}$. So

$$\begin{aligned} \mathbb{P}[\exists \text{ such monochromatic } K] &\leq \sum_{\substack{K \subseteq [n] \\ |K|=k}} \mathbb{P}[K \text{ induces a monochromatic graph}] \\ &= \binom{n}{k} 2^{1-\binom{k}{2}} \overset{\text{given}}{<} 1. \end{aligned}$$

Therefore, $\mathbb{P}\left[\nexists \text{ such monochromatic } K\right] > 0$. This means $r(k,k) > n$, which proves the first part.

Now,

$$\binom{n}{k} 2^{1-\binom{k}{2}} \leq 2 \left(\frac{en}{k}\right)^k \cdot 2^{-\binom{k}{2}} = 2 \left(\frac{en}{2^{\frac{k-1}{2}} \cdot k}\right)^k$$

where the first inequality is due to $\binom{a}{b} \leq \left(\frac{ea}{b}\right)^b$. If $\frac{en}{2^{\frac{k-1}{2}} \cdot k} < 1 - \varepsilon$ then for $k > k_0(\varepsilon)$ for some $k_0(\varepsilon)$, the RHS is < 1. This implies that $r(k,k) \geq [1-o(1)] \frac{k}{e} 2^{\frac{k-1}{2}}$.[1] ■

Remark 2
The lower bound was improved only by a factor of two since 1947.
The upper bound was improved several times, last time in 2023 by Campos, Griffiths, Morris, Sahasrabudhe to $(4-\varepsilon)^k$, for an absolute constant $\varepsilon > 0$.
Open: Does $\lim r(k,k)^{1/k}$ exist (for USD 100)? If exists, find it (for USD 250).

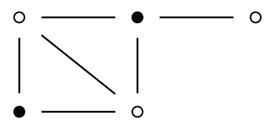
Remark 3
Open problem: Find an explicit coloring showing $r(k,k) > 1.0001^k$.

Remark 4
This proof provides a randomized algorithm for finding a coloring that shows $r(k,k) > \left\lfloor \sqrt{2^k} \right\rfloor$.
But given such a coloring, we don't know how to efficiently check that $\nexists$ a monochromatic K_k.

[1] Explanation for the last 'implies': We note that for every n satisfying the given condition, we have $r(k,k) > n$. Now for any $n < [1-\varepsilon] \frac{k}{e} 2^{\frac{k-1}{2}}$, the condition is satisfied. Thus, $r(k,k)$ is more than all such n's, which is written as $[1-o(1)] \frac{k}{e} 2^{\frac{k-1}{2}}$.

1.3 Example: Dominating Sets

Definition 3
If $G = (V, E)$ is a graph, we say $S \subseteq V$ is dominating if $\forall v \in V \setminus S \exists u \in S$ such that $\{u, v\} \in E$.

Example 2. The set of bold vertices in  form a dominating set.

Theorem 4
If $G = (V, E)$ is a graph with $|V| = n$ and minimum degree δ, then it has a dominating set of size at most $n \cdot \dfrac{1 + \ln(1+\delta)}{1+\delta}$.

Proof. Let $p = \frac{\ln(1+\delta)}{1+\delta}$. Clearly $p \in [0, 1]$. Let $X \subseteq V$ be a random subset of V obtained by choosing each $v \in V$ to randomly and independently lie in X with probability p. Since X is not necessarily a dominating set, we can *alter* it by

$$Y_X := \left\{ v \in V \setminus X \;\middle|\; \nexists u \in X \text{ with } \{u, v\} \in E \right\}.$$

By construction, $X \sqcup Y_X$ is a dominating set (note that they are disjoint).

Let's estimate the expected size of $X \cup Y_X$. First observe that $\mathbb{E}[|X \cup Y_X|] = \mathbb{E}[|X| + |Y_X|]$ due to disjointness, and this is further equal to $\mathbb{E}[|X|] + \mathbb{E}[|Y_X|]$ by linearity of expectation. $|X|$ is a sum of independent indicators, one for each vertex which takes 1 with probability p and 0 with probability $1 - p$. So $\mathbb{E}[|X|] = np$.
Note that $\mathbb{P}[v \in Y_X] = \mathbb{P}[v \notin X] \cdot \mathbb{P}[\text{no neighbor of } v \text{ is in } X] = (1-p)^{d_v} \leq (1-p)^{1+\delta} = \frac{1}{1+\delta}$ where d_v is the degree of v in G. Again $|Y_X| = \sum_{v \in V} \mathbf{1}_{v \in Y_X}$ whence $\mathbb{E}[|Y_X|] \leq \frac{n}{1+\delta}$.
This means $\mathbb{E}[|X \cup Y_X|] \leq n \left[\frac{1+\ln(1+\delta)}{1+\delta}\right]$. Since the 'average size' of a dominating set is less than or equal to the given quantity, $\exists$ a choice of X such that $X \cup Y_X$ is a dominating set of size at most $n \cdot \dfrac{1+\ln(1+\delta)}{1+\delta}$. ∎

Remark 5
We used *linearity of expectation*: $\mathbb{E}[X + Y] = \mathbb{E}[X] + \mathbb{E}[Y]$. We also used *alteration*: making a change after initial random choice, in this case we added Y_X to X. (To be discussed more)

Remark 6
Here $\exists$ an efficient algorith to find such a dominating set. Start with $\varnothing$ and keep adding vertices that dominate maximum of yet non-dominated vertices.

Remark 7
Estimate is essentially that for $n \gg \delta \gg 1$.