Probabilistic Method and Random Graphs

Lecture 8. The Method of Counting&Expectation

Xingwu Liu

Institute of Computing Technology, Chinese Academy of Sciences, Beijing, China

[1]The slides are mainly based on Chapter 6 of Probability and Computing.

Comments, questions, or suggestions?

Recap of Lecture 6

- Random graph models
  - $\mathcal{G}_{n,p}$, $\mathcal{G}_{n,m}$
    - Decoupling: $\mathcal{G}_{n,m} \sim (\mathcal{G}_{n,p} | \text{there are } m \text{ edges})$

$$
\begin{tikzcd}
\mathcal{G}_{n,m} & & \mathcal{G}_{n,p} \arrow[ll, "\text{approximation}"'] \arrow[dl, "\text{$\approx$, proved by conditioning on a proper range of edge numbers}"] \\
& \mathcal{G}_{n,p} | \text{there are } m \text{ edges} \arrow[ul, "\text{$=$, in the sense of distribution}"] &
\end{tikzcd}
$$

  - Other random graphs: with specified degree distribution, scale-free, small world

Recap of Lecture 6

- The existence of (sharp) threshold functions
  - Mostly on $\mathcal{G}_{n,p}$
  - Sharp threshold functions of connectivity: $\frac{\ln n}{n}$
  - That of major component existence: $\frac{1}{n}$
- A Hamiltonian cycle can be found efficiently w.h.p. in $\mathcal{G}_{n,p}$ with $p \geq 40\frac{\ln n}{n}$

Probabilistic Method
-Elegance from graph theory

- A warm-up example:
  – n players against each other
  – "Top-k" players get prize
- But, are you sure no controversy exists?
  – Controversy: prize-winners all defeated by a loser
- Unfortunately, when players are too many, controversy does exist w.h.p.

Proof (non-constructive)

- S: a k-subset of players
- A_S: no player beats all member of S
- Consider a random tournament

- $\Pr(A_S) = (1 - 2^{-k})^{n-k}$
- $\Pr(\text{no controversy}) = \Pr(\bigcup A_S)$

$$\leq \sum \Pr(A_S) = \binom{n}{k}(1 - 2^{-k})^{n-k}$$

$$= o\left(\frac{1}{n}\right)$$

- Find a controversial one? Just sampling

Cool?

A piece of cake in probabilistic method!

What is the Probabilistic Method?

- **Proving the existence** of an object satisfying a certain property **without constructing** it
- Underlying principle

Probability space → Nonzero probability of satisfying the property → Existence

- Pioneered by Erdős in 1940's
- Construct by flipping coins
- Naturally lead to (randomized) algorithms

Main Probabilistic Methods

- Counting argument
- First-moment method
- Second-moment method
  - Higher-moment methd
- Lovasz local lemma

Algorithms

Counting Argument

- Construct a probability space and calculate the probability

- Algorithm design: sampling

- Application
  - Tournament
  - Ramsey number: an observation by S. Szalai, 1950's

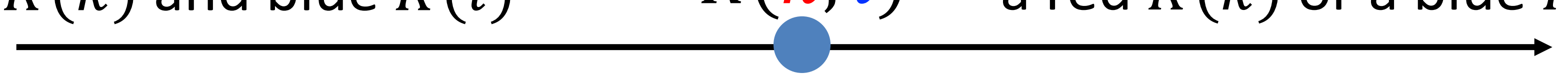
Ramsey Number

- Given a complete graph $K(n)$, 2-color its edges
- Ramsey number $R(k, l)$
  - the smallest number n such that for every complete graph G with at least n vertices, any 2-coloring of G would either has a red $K(k)$ or a blue $K(l)$

There is a coloring without red $K(k)$ and blue $K(l)$

$R(k, l)$

Every coloring has either a red $K(k)$ or a blue $K(l)$



Number of vertices of the complete graph

Ramsey Number is Well defined

- Ramsey Theorem: $R(k,l)$ is finite for any k,l
  - Proved by F. Ramsey in 1930 and determined $R(3{,}3)$
  - Origin of Ramsey theory
    - The existence of rather large good substructure in a big structure
  - How much is $R(k,l)$?
- Upper bound: $R(k,l) \leq R(k-1,l) + R(k,l-1)$
  - Proved by P. Erdős and G. Szekeres in 1935
  - The 2^{nd} cornerstone of Ramsey theory
  - By $R(k,2) = R(2,k) = k$, $R(k,l) \leq \binom{k+l-2}{k-1}$
  - It implies $R(k,k) \leq 4^k$
    - Best: $k^{-c\frac{\ln k}{\ln \ln k}} 4^k$ by Conlon in 2009

Known bounding ranges

r \ s	1	2	3	4	5	6	7	8	9	10
1	**1**	1	1	1	1	1	1	1	1	1
2		**2**	3	4	5	6	7	8	9	10
3			**6**	9	14	18	23	28	36	40–42
4				**18**	25[5]	36–41	49–61	59[10]–84	73–115	92–149
5					43-48	58–87	80–143	101–216	133–316	149[10]–442
6						**102–165**	115[10]–298	134[10]–495	183–780	204–1171
7							**205–540**	217–1031	252–1713	292–2826
8								**282–1870**	329–3583	343–6090
9									**565–6588**	581–12677
10										**798–23556**

Vigleik Angeltveit; Brendan McKay (2017). "R(5,5)≤48". arXiv

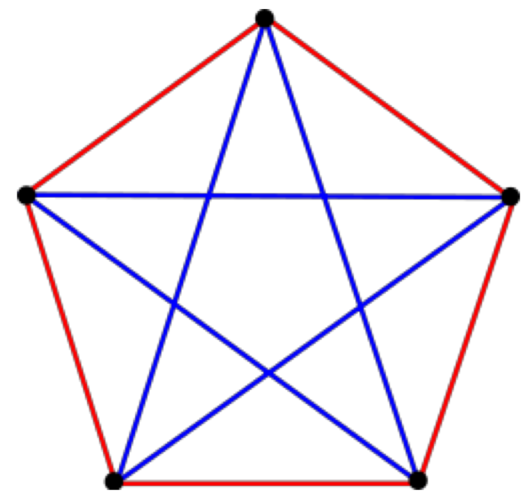
Proof of the upper bound

- 2-color the complete graph on $R(k-1,l) + R(k,l-1)$ vertices
- Pick a vertex u
- Define subgraphs G_r and G_b: $\forall\, v$,
  - $v \in \begin{cases} G_r & \text{if } (u, v) \text{ is red} \\ G_b & \text{if } (u, v) \text{ is blue} \end{cases}$
- Either $|G_r| \geq R(k-1,l)$ or $|G_b| \geq R(k,l-1)$
  - Do case-by-case analysis

Example: Ramsey Number $R(3,3)$

$R(3,3) \leq 6$

Actually, $R(3,3) > 5$



What is $R(k,k)$?

- $R(k,k) > 2^{k/2} = \sqrt{2}^k$ (Erdős, 1947)
  - Best: $[1+o(1)]\frac{k}{e}\sqrt{2}^{1+k}$ by Spencer in 1975
- For any complete graph with at most $2^{k/2}$ vertices, there is a 2-coloring without red $K(k)$ and blue $K(k)$

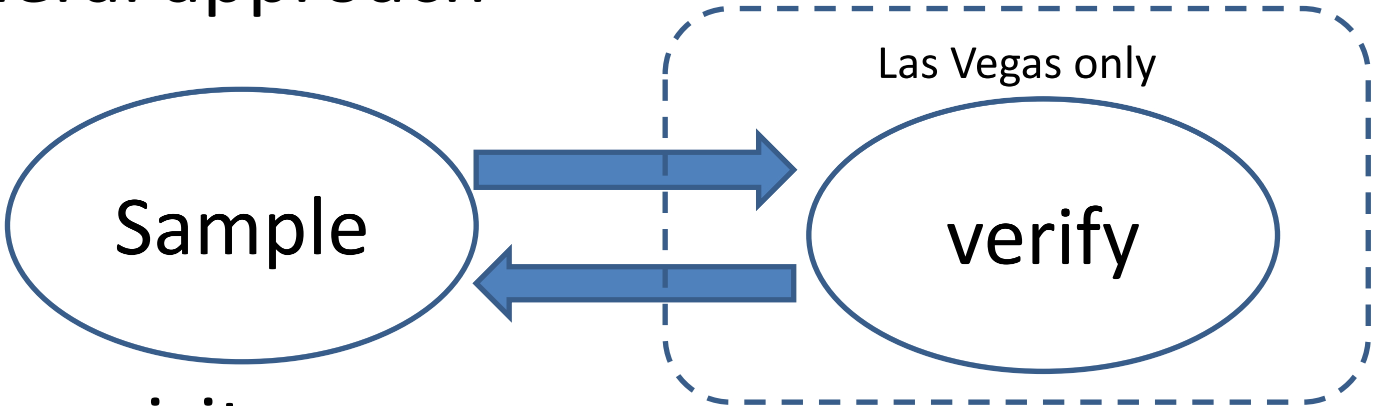
- Prove by the probabilistic method

Proof

- Consider a graph G on n vertices, and 2-color each edge randomly
  - Uniform distribution on all 2-coloring
- A_S: the subgraph on S is monochromatic
  - S is a k-subset of the vertices
- $\Pr(A_S) = 2^{1-\binom{k}{2}}$
- $\Pr(\bigcup_S A_S) \leq \binom{n}{k} 2^{1-\binom{k}{2}} < \frac{2^{1+\frac{k}{2}}}{k!} \frac{n^k}{2^{\frac{k^2}{2}}}$<1   (If $n = \lfloor 2^{k/2} \rfloor$)
- There is a 2-coloring of the edges such that $K(\lfloor 2^{k/2} \rfloor)$ has no monochromatic k subgraph
  - Together with the definition of $R(k,k)$, $R(k,k) > 2^{k/2}$

Randomized Algorithms

- But how to find a good coloring? By sampling!
- General approach



- Prerequisites
  - Efficient sampling
  - Small probability of failure
  - Efficient verification (Las Vegas only)

First-Moment method

- Use the expectation in probabilistic reasoning
- Two types of first-moment method
  - Expectation argument

    $$\Pr(X \geq \mathbb{E}[X]) > 0, \Pr(X \leq \mathbb{E}[X]) > 0$$

  - Markov's inequality for non-negative X
    - $\Pr(X \geq a) \leq \frac{\mathbb{E}[X]}{a}$
    - When X is integer-valued,

      $$\Pr(X \neq 0) = \Pr(X > 0) = \Pr(X \geq 1) \leq \mathbb{E}[X]$$

First-Moment argument

- 3-CNF Boolean formula
  - $(x_1 \vee x_2 \vee x_3) \wedge \ldots \wedge (\overline{x_1} \vee \overline{x_3} \vee x_4)$
- For such a formula, at most how many clauses can be satisfied simultaneously?
  - **MAX-3SAT** is NP-hard
- Theorem: there is a truth assignment which satisfies $\geq \frac{7}{8}$-fraction of the clauses

Proof

- Randomly assign truth values to each variable
- Define r.v. X_i indicating whether clause i is true
- $\mathbb{E}[X_i] = \frac{7}{8} \Rightarrow \mathbb{E}[\sum_{i=1}^{n} X_i] = \frac{7}{8}n$

- Remark: probability of sampling a good truth assignment $\geq \frac{1}{n+1}$, leading to an efficient alg.
  – Optimum, since impossible to get a $\left(\frac{7}{8} + \varepsilon\right)$-approx.
    - J. Hastad. Some optimal inapproximability results. STOC 1997

Proof of $\Pr\left(\sum X_i \geq \frac{7}{8}n\right) \geq \frac{1}{n+1}$

- Let $X = \sum X_i$ and $p = \Pr\left(X \geq \frac{7}{8}n\right)$
- $\frac{7}{8}n = \mathbb{E}[X]$

$$= \sum_{i<\frac{7}{8}n} i * \Pr(X = i) + \sum_{i\geq\frac{7}{8}n} i * \Pr(X = i)$$

$$\leq \left(\frac{7}{8}n - \frac{1}{8}\right)(1 - p) + np$$

$$= \frac{7}{8}n - \frac{1}{8} + \frac{n+1}{8}p$$

Expectation argument

- Turán Theorem
  - Any graph G=(V,E) contains an independent set of size at least $\frac{|V|}{D+1}$, where $D = \frac{2|E|}{|V|}$
- Proof: Consider the following random process for constructing an independent set S:
  - Initialize S to be the empty set
  - For each vertex u in V in random order,

    if no neighbors of u are in S, add u to S
  - Return S

Proof (Continued)

- S is an independent set
- Vertex u is selected with probability $\geq \frac{1}{d(u)+1}$
  - See the next slide
- So, $\mathbb{E}[|S|] \geq \sum \frac{1}{d(u)+1} \geq \frac{|V|}{D+1}$ due to convexity

- **Remark:** probability of sampling a good independent set is $\geq \frac{1}{2D|V|^2}$

Proof: Pr(u is selected)$\geq \frac{1}{d(u)+1}$

- u is selected if and only if A occurs
  - A: when sampling first occurs in the neighborhood of u, u rather than its neighbors is sampled
    - Neighborhood: u and its then-valid neighbors
  - Denote the neighborhood by N, and the number of then-valid neighbors by x. Note that $x \leq d(u)$
- $\Pr(A) = \Pr(u \text{ is chosen} | \text{sampling occurs in } N)$

$$= \frac{1}{x+1} \geq \frac{1}{d(u)+1}$$

References

- http://www.cse.buffalo.edu/~hungngo/classes/2011/Spring-694/lectures/sm.pdf
- http://www.cse.cuhk.edu.hk/~chi/csc5160-2007/notes/Probabilistic.ppt
- Erdős. Graph theory and probability I. 1959
- Erdős. Graph theory and probability II. 1961
- Alon&Krivelevich. Extremal and Probabilistic Combinatorics. 2006