Lecture 4: "The Deletion Method"
(and more examples from previous lecture)

Sotiris Nikoletseas
Professor

Computer Engineering and Informatics Department
2017 - 2018

Summary of previous lecture

i. Proofs of existence using the *Linearity of Expectation method*

- Methodology
  (1) $X = \sum X_i$
  (2) Calculate $E[X]$
  (3) $\exists x \geq E[X]$ and $\exists x \leq E[X]$
- Examples
  (1) Tournament with at least $(n-1)!2^{-(n-1)}$Hamiltonian cycles.
  (2) Every graph $G = (V, E)$ has a bipartite subgraph with at least $\frac{|E|}{2}$ edges.

ii. Non-existence proofs using the *Markov Inequality*

- Methodology
  (1) $\Pr[X \geq t] \leq \frac{E[X]}{t}$ $(t > 0)$
  (2) $t = 1 \Rightarrow \Pr[X \geq 1] \leq E[X]$
  (3) if $E[X] \to 0 \Rightarrow \Pr[X = 0] \to 1$
- Example: $\nexists$ d.s. of size $< \ln n$ in $G_{n,\frac{1}{2}}$ w.h.p.

Summary of this lecture

i. Non-existence proofs
   - Example: Satisfiability of boolean formulas (SAT).

ii. Linearity of Expectation
   - Example: Dominating sets in arbitrary graphs.

iii. A variation of Linearity of Expectation method: The Deletion Method
   - Method's basic idea.
   - Examples
     (1) An improved lower bound for Ramsey numbers.
     (2) Independent sets (Turán's Theorem, 1941).

Satisfiability Problem (SAT)
Importance

- many practical applications
  - program and machine testing
  - artificial intelligence
  - VLSI design and testing (0,1 voltage values of logic gates)
- is a fundamental problem for computational complexity.
  - is the first known NP-complete problem, as proved by Stephen Cook in 1971 and independently by Leonid Levin in 1973.

- Literal = a variable or its negation
- Clause = a disjunction of literals
- Conjunctive Normal Form (CNF) = a formula F that is a conjunction of clauses

$$F = \left(\bigvee_i l_i\right) \wedge \left(\bigvee_j l_j\right) \wedge \cdots \wedge \left(\bigvee_k l_k\right)$$

- Truth assignment: a set of boolean $(0, 1)$ values to the n variables. $\Rightarrow \exists\ 2^n$ truth assignments.
- Satisfying truth assignment: one that makes F true (i.e., it satisfies all its clauses).

Definition 1 (SAT)

Determine if there exist a truth assignment of n variables such that $F = 1$ (i.e., is there a satisfying truth assignment?).

Definition 2 (k-SAT)

All clauses contain exactly k literals.

- $k = 3$: minimum value for which k-SAT is NP-hard.
- $k = 2$: $\exists$ polynomial algorithms.

Random boolean formulas of 3-SAT:

1. For each clause, choose independently and equiprobably the 3 variables that are contained in it.
2. For each variable x in a clause, choose eqiuprobably if it will be x or $\bar{x}$.

Let F be a random boolean formula with n variables and $m = cn$ clauses.

Intuition:

- $c \uparrow \Rightarrow Pr[\text{satisfiability}] \downarrow$
- $c \downarrow \Rightarrow Pr[\text{satisfiability}] \uparrow$

Experimental results: $\exists$ threshold c^* such that:

- $c > c^* \Rightarrow$ non-satisfiable w.h.p.
- $c < c^* \Rightarrow$ satisfiable w.h.p.

(I) Example of non-existence proofs

Theorem 1

For any $c \geq 5.191$, in a 3-SAT problem with n variables and $m = cn$ clauses, $Pr[satisfiability] \to 0$ as $n \to \infty$

Proof:

1. We construct a random sample space by determining the value of each variable, equiprobably for the two options (0,1) and independently for each variable. Clearly, the sample points of this space are random truth assignments. Let A be any fixed assignment.
2. Define r.v. X that corresponds to the number of satisfying truth assignments.
3. a. $X = \sum_A X_A$ where X_A are indicator variables

$$X_A = \begin{cases} 1 & \text{A is a satisfying t.a.} \\ 0 & \text{otherwise} \end{cases}$$

Proof of theorem 1

b. Calculate $E[X]$: Using linearity of expectation:

$$E[X] = E\left[\sum_A X_A\right] = \sum_A E[X_A]$$

- Calculate expectation of indicator variable X_A:

$$E[X_A] = \Pr\{\text{A is a satisfying t.a.}\} = \left(\frac{7}{8}\right)^m$$

- So,

$$E[X] = \sum_A E[X_A] = 2^n \left(\frac{7}{8}\right)^{cn}$$

c. if $2^n \left(\frac{7}{8}\right)^{cn} \to 0 \Rightarrow E[X] \to 0$

$\Leftrightarrow$ if $c \geq log_{\frac{8}{7}} 2 = 5.191$ then $E[X] \to 0$

4. $\Pr[X = 0] \to 1 \Rightarrow$ w.h.p. $\nexists$ satisfying truth assignment.

(II)Example: Dominating sets in arbitrary graphs

Theorem 2

Any graph $G = (V, E)$ on n vertices with minimum degree $\delta > 1$ has a dominating set of size at most $n\frac{1+ln(\delta+1)}{\delta+1}$

Proof:

- We construct a random sample space by choosing for each vertex if it is contained or not in a set X with probability $p = \frac{ln(\delta+1)}{\delta+1}$ (X is a d.s. under construction)
- Define Y_X the set that contains vertices which
  - are not contained in X and
  - are not neighbours of any vertex in X

  Clearly, $X \cup Y_X$ is a dominating set.
- By Linearity of Expectation:
  $E[|X \cup Y_X|] = E[|X|] + E[|Y_X|]$ (X, Y disjoint)
- $|X| \sim B(n, p) \Rightarrow E[|X|] = np = n\frac{ln(\delta+1)}{\delta+1}$

Proof of theorem 2

- $\Pr[u \in Y_X] = \Pr[u \notin X \cap \forall v \in X : (u, v) \text{ not an edge}]$

$$\Rightarrow \Pr[u \in Y_X] = (1-p) \cdot (1-p)^{d(u)} = (1-p)^{1+d(u)}$$

$$\leq (1-p)^{1+\delta} \leq e^{-p(1+\delta)} = e^{-\frac{ln(\delta+1)}{\delta+1}(1+\delta)} = e^{-ln(\delta+1)} = \frac{1}{\delta+1}$$

- Thus, $E[Y_X] \leq n\frac{1}{\delta+1}$
- $E[|X \cup Y_X|] = E[|X|] + E[|Y_X|] \leq n\frac{ln(\delta+1)}{\delta+1} + n\frac{1}{\delta+1}$
- Thus, there must exist a d.s. of size $\leq n\frac{1+ln(\delta+1)}{\delta+1}$

(1) Prove the existence of a structure that:
- doesn't have the desired property but
- it is “very similar” to have the desired property.

(2) Modify the random structure (e.g. delete problematic parts of it) in order to have a structure with the desired property.

Definition 3

The diagonal ramsey number $R(k,k)$ is the smallest integer n such that in any two-coloring of the edges of the complete graph on n vertices K_n there is a monochromatic K_k.

Theorem 3

For any integer n: $R(k,k) > n - \binom{n}{k} 2^{1-\binom{k}{2}}$

Proof:

- Construct a probability sample space of edge coloring by two-coloring at random, equiprobably (for the two colors) and independently the edges of K_n.
- Let S be any fixed set of k vertices.
- We define the r.v. $X = \sum_{S,|S|=k} X_S$ that corresponds to the number of monochromatic sets and X_S:

$$X_S = \begin{cases} 1 & \text{S is monochromatic} \\ 0 & \text{otherwise} \end{cases}$$

- $E[X_S] = \Pr[\text{S is monochromatic}] = 2\left(\frac{1}{2}\right)^{\binom{k}{2}} = 2^{1-\binom{k}{2}}$

Proof of theorem 3

- By linearity of expectation:
  $E[X] = E\left[\sum_{S,|S|=k} X_S\right] = \sum_S E[X_S] = \binom{n}{k} 2^{1-\binom{k}{2}}$
- $\exists$ point: $X \leq E[X] \Rightarrow \exists$ edge two-coloring of K_n of at most $\binom{n}{k} 2^{1-\binom{k}{2}}$ monochromatic sets of size k.
- If we remove at most 1 vertex from any set S we have a graph of at least $n - \binom{n}{k} 2^{1-\binom{k}{2}}$ vertices with no monochromatic subsets of size k.
- Thus, $R(k,k) > n - \binom{n}{k} 2^{1-\binom{k}{2}}$

Example 2 - Independent set in arbitrary graphs

Definition 4 (Independent Set)

An independent set is a set of vertices such that for every two vertices in it, there is no edge connecting them.

Theorem 4 (Turán, 1941)

Consider a graph $G = (V, E)$ of n vertices and $\frac{nd}{2}$ edges. Then, G contains an independent set of size at least $\frac{n}{2d}$.

Proof of theorem 4

- We construct a random sample space by choosing for each vertex if it is contained or not in a set X (with probability p it is contained), independently for each vertex. X is an i.s. under construction.
- Let Y be the number of edges between vertices of X.
- We can denote $Y = \sum_e Y_e$ where

$$Y_e = \begin{cases} 1 & \text{e is an edge between vertices of X} \\ 0 & \text{otherwise} \end{cases}$$

- $E[Y_e] = p^2$
- Thus, $E[Y] = \sum_e E[Y_e] = |E|p^2 = \frac{nd}{2}p^2$
- $|X| \sim B(n,p) \Rightarrow E[|X|] = np$
- By Linearity of Expectation $E\left[|X| - |Y|\right] = E[|X|] - E[|Y|] = np - \frac{nd}{2}p^2$

Proof of theorem 4

- An independent set has the maximum value of $|X| - |Y|$.
- The value $p = \frac{1}{d}$ maximizes the expectation of the difference $|X| - |Y|$ which becomes:

$$E\left[|X| - |Y|\right] = \frac{n}{2d}$$

- $\exists$ point such that $|X| - |Y| \geq E\left[|X| - |Y|\right] = \frac{n}{2d}$
- $\exists$ set of vertices such that $|X| \geq |Y| + \frac{n}{2d}$
- $\exists$ set of vertices such that # vertices $\geq$ #edges $+\frac{n}{2d}$
- We can modify X by removing one vertex from each edge inside X. We obtain a set with at least $\frac{n}{2d}$ vertices and no edges, which is clearly independent.
- $\exists$ independent set of size at least, $\frac{n}{2d}$.