PACS Part 2, Lecture 3

Probabilistic Method

- show existence of objects satisfying property E by showing $\mathbb{P}(E) > 0$

First Moment

- show satisfiability of $X \geq \mu$ by showing $\mathbb{E}X \geq \mu$
- uses the inequality $\mathbb{P}(X \geq \mathbb{E}X) > 0$
- Example: large cuts

  Let $G = (V, E)$ be an undirected graph with $m = |E|$. Then there is a cut of G with value at least $m/2$.

  Proof: We divide the vertices of V into two disjoint sets A and B, assigning each vertex to A with probability $1/2$, independently of the other choices. Then, defining $X_e = 1$ iff edge $e \in E$ is in the cut defined by A and B, we have

  $$\mathbb{E}C(A, B) = \mathbb{E}\sum_{e\in E} X_e = \frac{m}{2}$$

  for the expected value of the cut.

- Transforming this into an algorithm for finding a large cut, define $p = \mathbb{P}(C(A, B) \geq m/2)$ and note $C(A, B) \leq m$ to get

  $$\frac{m}{2} = \mathbb{E}\sum_{e\in E} X_e \leq (1-p)\left(\frac{m}{2} - 1\right) + pm$$

  which implies $p \geq 1/(1 + m/2)$. This gives a Las Vegas algorithm with an expected number of iterations of $O(m)$.

- We can *derandomize* this algorithm by fixing any enumeration $v_1, \ldots, v_n$ of the vertices. Letting x_i denote the choice of set A or B for vertex v_i, we show that it is possible to achieve

  $$\frac{m}{2} \leq \mathbb{E}[C(A, B) \mid x_1, \ldots, x_k] \leq \mathbb{E}[C(A, B) \mid x_1, \ldots, x_{k+1}]$$

  which follows from the law of total expectation conditioning on the value of x_{k+1}. The base case is $\mathbb{E}[C(A, B) \mid x_1] = \mathbb{E}C(A, B) = m/2$. To make the choice x_{k+1} that maximizes the conditional expectation, we note that it is equal to the number of edges in the cut between vertices among $v_1, \ldots, v_{k+1}$ plus half the remaining edges. This can be computed in linear time.

Sample and Modify

- sometimes it is not sufficient to make all choices randomly, we are thus led to modifying the random sample to satisfy the specification
- Example: independent sets

  If $G = (V, E)$ is a connected undirected graph with n vertices and $m \geq n/2$ edges, then G has an independent set of size $\geq n^2/4m$.

  Proof: Let $d = 2m/n$ be the average degree of vertices. First, select each vertex with probability $1/d$. Then remove one vertex for each induced edge.

  Let X be the number of selected vertices and Y the number of induced edges. Then:

$$\mathbb{E}(X - Y) = \mathbb{E}X - \mathbb{E}Y = \frac{n}{d} - \frac{n}{2d} = \frac{n}{2d} = \frac{n^2}{4m}$$

Second Moment

- using Chebyshev's inequality, we will show that $p = n^{-2/3}$ is a threshold function for the occurrence of cliques of size 4 in Erdős–Rényi graphs
- Let $C_1, \ldots, C_{\binom{n}{4}}$ be an enumeration of all possible 4-cliques and define the indicator variable $X_i = 1$ iff C_i is a clique.
- Set $X = \sum_{i=1}^{\binom{n}{4}} X_i$.
- First, let $p = o(n^{-2/3})$. Then

$$\mathbb{P}(X \geq 1) \leq \mathbb{E}X = \binom{n}{4} p^6 = O(n^4 p^6) = o(1)$$

  as $n \to \infty$.

- Now let $p = \omega(n^{-2/3})$. We have

$$\mathrm{Var}(X) = \sum_i \mathrm{Var}(X_i) + \sum_{i \neq j} \mathrm{Cov}(X_i, X_j) \leq \mathbb{E}X + \mathrm{Cov}(X_i, X_j)$$

  since the X_i are indicator variables and thus $\mathrm{Var}(X_i) \leq \mathbb{E}X_i^2 = \mathbb{E}X_i$. Depending on the number $|C_i \cap C_j|$ of vertices in the intersection of the potential cliques, we either have $\mathrm{Cov}(X_i, X_j) = 0$ (if it is 0 or 1) or a positive term (if it is 2 or 3). Collecting the terms, we have:

$$\mathrm{Var}(X) \leq \binom{n}{4} p^6 + \binom{n}{6} \binom{6}{2; 2; 2} p^{11} + \binom{n}{5} \binom{5}{3; 1; 1} p^9$$

  Compared to $(\mathbb{E}X)^2 = O(n^8 p^{12})$, we have:

$$\mathrm{Var}(X) = o(n^8 p^{12}) = o((\mathbb{E}X)^2)$$

But this implies

$$\mathbb{P}(X=0) \le \mathbb{P}(|X-\mathbb{E}X| \ge \mathbb{E}X) \le \frac{\mathrm{Var}(X)}{(\mathbb{E}X)^2} = o(1)$$

as $n \to \infty$.

Lovász Local Lemma

- if we can bound the probabilities of bad events $E_1, \ldots, E_n$ individually, then we can use the union bound to bound the probability of none of them occurring
- however, if $\sum_{i=1}^n \mathbb{P}(E_i) \ge 1$, then this doesn't give a meaningful bound
- if the E_i are mutually independent, then it suffices to have the very weak bound $\mathbb{P}(E_i) < 1$ to conclude

  $$\mathbb{P}\left(\bigcap_{i=1}^n \bar{E}_i\right) = \prod_{i=1}^n (1-\mathbb{P}(E_i)) > 0$$

- this can be generalized to the case of limited dependence
- Definition: E_i is mutually independent of the set $\{E_j \mid j \in J\}$ of events if:

  $$\forall I \subseteq J: \quad \mathbb{P}\left(E_i \mid \bigcap_{j \in I} E_j\right) = \mathbb{P}(E_i)$$

- Definition: Let $E_1, \ldots, E_n$ be events. A dependency graph of the events is a graph $G = ([n], E)$ such that E_i is mutually independent of the set $\{E_j \mid (i,j) \notin E\}$.
- Lovász Local Lemma:

  Let $E_1, \ldots, E_n$ be events and a dependency graph G such that:

  1. $\mathbb{P}(E_i) \le p$ for all i
  2. the maximum vertex degree of G is at most d
  3. $4dp \le 1$

  Then $\mathbb{P}(\bigcap_{i=1}^n \bar{E}_i) > 0$.
- Example: k-SAT

  If all variables appear in at most $T = 2^k/4k$ clauses, then the formula is satisfiable.

  Proof: Assign truth values uniformly i.i.d. Let E_i be the event that clause i is not satisfied. Then $\mathbb{P}(E_i) = 2^{-k}$. By the pigeonhole principle, the maximum degree of G is at most $d \leq kT \leq 2^k/4$. We verify:

$$4dp \leq 4\frac{2^k}{4}2^{-k} = 1$$

An application of the local lemma thus concludes the proof.

- Proof of the local lemma: We prove by induction on $0 \leq s \leq n-1$ that

$$\mathbb{P}\left(E_k \mid \bigcap_{j \in S} \bar{E}_j\right) = \mathbb{P}(E_k \mid F_S) \leq 2p$$

  for all $S \subseteq [n]$ and $k \in [n] \setminus S$. The lemma then follows: Setting $S_i = [i-1]$, we get

$$\mathbb{P}\left(\bigcap_{i=1}^{n} \bar{E}_i\right) = \prod_{i=1}^{n} \mathbb{P}(\bar{E}_i \mid F_{S_i}) \geq \prod_{i=1}^{n}(1-2p) > 0$$

  The base case $s = 0$ of the induction is just assumption 1.

  For the induction step, define $S_1 = \{j \in S \mid (k,j) \in E\}$ and $S_2 = S \setminus S_1$. If $S_1 = \emptyset$, then the events are mutually exclusive and the inequality follows from assumption 1. If $S_1 \neq \emptyset$, then

$$\mathbb{P}(E_k \mid F_S) = \frac{\mathbb{P}(E_k \cap F_S)}{\mathbb{P}(F_S)} = \frac{\mathbb{P}(E_k \cap F_{S_1} \mid F_{S_2})}{\mathbb{P}(F_{S_1} \mid F_{S_2})}$$

  since $F_S = F_{S_1} \cap F_{S_2}$. By the definition of S_2 and assumption 1, we can bound the numerator by $\mathbb{P}(E_k) \leq p$. By applying the induction hypothesis to $\mathbb{P}(E_i \mid F_{S_2})$, we can bound the denominator by

$$\mathbb{P}(F_{S_1} \mid F_{S_2}) \geq 1 - \sum_{i \in S_1} \mathbb{P}(E_i \mid F_{S_2}) \geq 1 - 2pd \geq \frac{1}{2}$$

  using assumptions 2 and 3. This concludes the induction step and the proof of the lemma.