CS5314
Randomized Algorithms

Lecture 17: Probabilistic Method
(Introduction)

Objectives

- Introduce Probabilistic Method
  - a powerful way of proving existence of certain objects
- Idea: If a certain object can be selected with positive probability (in some sample space), then this object must exist
- Introduce two techniques :

  Basic Counting and Expectation

Basic Counting Argument

Question: Is it possible to color edges of a complete graph in red or green, so that there is no large monochromatic clique ?

- monochromatic = same color

- Let n = # vertices in the graph
- Let K_k = A clique of k vertices

Then, we have the following theorem:

Basic Counting Argument (2)

Theorem:

If $2C(n,k) / 2^{C(k,2)} < 1$, we can color edges of K_n using red or green such that there is no red K_k and no green K_k subgraphs

Proof:

Define a sample space

S = all possible colorings of edges of K_n

Proof (cont)

- Consider choosing a coloring from S, uniformly at random

  Let G = the chosen colored graph

- One way to choose:

  Start with an empty graph ; Color each edge in a particular color with prob = 1/2

Thus, for a particular k-vertex clique, it is monochromatic with probability $2/2^{C(k,2)}$

Proof (cont)

Let x = #distinct k-vertex clique in K_n

Question: What is the value of x?

Let A be the event such that

A := there is a k-vertex clique in K_n

and $A_1, A_2, \ldots, A_x$ be the events such that

A_j := j^{th} k-vertex clique is monochromatic

➔ $\Pr(A) \leq \Pr(A_1) + \ldots + \Pr(A_x)$

$\leq x \; 2/2^{C(k,2)} = 2C(n,k) / 2^{C(k,2)} < 1$

Proof (cont)

- In other words, the event A' (the complement of A), which is the event that there is no k-vertex clique in K_n happens with probability:

$$\Pr(A') = 1 - \Pr(A) > 1 - 1 = 0$$

➔ there must exist a coloring with no monochromatic K_k subgraph

Example

Question: In a group of 1000 people, is it possible that for any 20 people selected, some pair of people know each other, while some pair of people don't know each other?

Answer: Yes (why??)

Example (2)

- Mapping each person to a vertex, and each relationship (friend/non-friend) to one of the colors, we have

$$n = 1000 \text{ and } k = 20 \quad (\text{note: } n \le 2^{k/2})$$

So, $2C(n,k) / 2^{C(k,2)}$

$$\le 2(n^k / k!)/2^{C(k,2)} \le 2(2^{k^2/2} / k!)/2^{C(k,2)}$$

$$= 2(2^{k/2}/k!) = 2048/(20!) < 1$$

➔ Thus, the answer of the question is YES

A Side Note

- An interesting branch in Mathematics, called Ramsey Theory, studies the minimum n to guarantee the two-coloring of K_n must either contain a red r-clique or a green s-clique

- Such an n is denoted by $R(r,s)$
- For example,

$$R(2,s) = s, \quad R(3,3) = 6$$

A Side Note

A related quote by Joel Spencer:

" Erdos asks us to imagine an alien force, vastly more powerful than us, landing on Earth and demanding the value of R(5,5) or they will destroy our planet.

In that case, he (Erdos) claims, we should marshal all our computers and all our mathematicians and attempts to find the value.

But suppose, instead, that they ask for R(6,6). In that case, he believes, we should attempt to destroy the aliens."

Expectation Argument

- Another method to prove the existence of an object is by averaging argument

- Based on the fact that :

  For a random variable X,

  $$Pr(X \geq E[X]) > 0$$

  $$Pr(X \leq E[X]) > 0$$

- The proof is given in the next slides

Expectation Argument (2)

Lemma: Let S be a probability space and a random variable X (defined on S)

Suppose $E[X] = \mu$. Then,

$$\Pr(X \geq \mu) > 0$$

Proof: Suppose on contrary $\Pr(X \geq \mu) = 0$

Then, $\mu = \sum x \Pr(X=x) = \sum_{x < \mu} x \Pr(X=x)$

$$< \sum_{x < \mu} \mu \Pr(X=x) = \mu$$

Expectation Argument (3)

Lemma: Let S be a probability space and a random variable X (defined on S)

Suppose $E[X] = \mu$. Then,

$$\Pr(X \leq \mu) > 0$$

Proof: Suppose on contrary $\Pr(X \leq \mu) = 0$

Then, $\mu = \sum x \Pr(X=x) = \sum_{x > \mu} x \Pr(X=x)$

$> \sum_{x > \mu} \mu \Pr(X=x) = \mu$

Example 1: Large Cut

Definition: A cut is a partition of the set of vertices V of a graph into two disjoint sets V_1 and V_2

Definition: The size of a cut (V_1,V_2) is #edges with one endpoint from V_1 and one endpoint from V_2

Fact: Finding a cut in a graph G with maximum size is NP-hard

Example 1: Large Cut (2)

New Target:

Can we get a sub-optimal cut, but whose size is at least half of the optimal?

We begin with some observations:

Let m = #edges in G

Trivial Fact: Size of any cut in $G \leq m$

Question: How about a lower bound?

Lower Bound of Maximum Cut

Theorem: There is some cut in G whose size is at least $m/2$

Proof: Let us construct a cut (V_1, V_2) by assigning each vertex of G randomly and independently into one of the two sets. That is, $\Pr(v \text{ in } V_1) = \Pr(v \text{ in } V_2) = 1/2$

Let X = size of this cut

Question: What is $E[X]$?

Lower Bound of Maximum Cut (2)

Let $X_1, X_2, \ldots, X_m$ be indicators such that

$X_j = 1$ if j^{th} edge has one endpoint from V_1 and one endpoint from V_2

$X_j = 0$ otherwise

Then, $E[X_j] = 1/2$ … (why??)

Also, $X = X_1 + X_2 + \ldots + X_m$

→ $E[X] = E[X_1 + X_2 + \ldots + X_m] = m\,E[X_1] = m/2$

→ Theorem thus follows

Example 1: Large Cut (3)

- Now, we know that by random assignment of vertices, it is possible to get a cut whose size is at least m/2

  → at least half of optimal

- Let us see the probability that we can generate such a cut (of size at least m/2) in one random assignment

Example 1: Large Cut (4)

Let X = size of the random cut

$p = Pr(\text{success}) = Pr(X \geq m/2)$

Then, we have

$$m/2 = E[X]$$

$$= \Sigma_{x < m/2}\, x\, Pr(X=x) + \Sigma_{x \geq m/2}\, x\, Pr(X=x)$$

$$\leq (1-p)(m/2 - 1) + pm \quad \text{... (why??)}$$

➔ $Pr(\text{success}) = p \geq 1 / (m/2 + 1)$

➔ can get a sub-optimal cut by repeating random assignment

Example 2: MAXSAT

Definition: A literal is a Boolean variable or a negated Boolean variable. E.g., x, $\neg y$

Definition: A clause is several literals connected with $\vee$'s. E.g., $(x \vee y \vee \neg z)$

Definition: A SAT formula is an expression made of clauses connected with $\wedge$'s. E.g., $(x \vee y \vee \neg z) \wedge (\neg y \vee z) \wedge (\neg x)$

Definition: A formula is satisfiable if there is an assignment of variables to T/F such that the value of formula is TRUE

Example 2: MAXSAT (2)

Fact: Determining a SAT formula is satisfiable is NP-complete

- Related problem: Find an assignment of variables that maximize #true clauses

Fact: Finding the above assignment is an NP-hard problem

➔ Finding optimal assignment with max clauses satisfied may be time-consuming

Example 2: MAXSAT (3)

New Target:

Can we get a sub-optimal assignment?

We begin with some observations:

Let m = # clauses in the formula

Trivial Fact: #clauses satisfied in any assignment is at most m

Question: How about a lower bound?

Lower Bound of MAXSAT

Theorem:

Let k = #literals in the smallest clause

Then, there is an assignment that satisfies at least

$m(1 - 1/2^k)$ clauses

How to prove? (Left as an exercise)