Lecture 8: Oct 4

Lecturer: Mohammad R. Salavatipour *Scribe: Joshua Gane*

8.1 The Probabilistic Method

In the next few lectures we will be talking about the Probabilistic Method. This method which was invented by Erdös more than 50 years ago is a powerful technique in combinatorics which has found numerous applications in theoretical computer science, specially in the design of randomized algorithms. The basic probabilistic method can be described as follows: in order to prove the existence of a combinatorial structure or object with certain properties, we construct an appropriate probability space and show that a randomly chosen element in this space has the desired properties with positive probability. In most applications, this probability is not only positive, but is actually high and frequently tends to 1 as the parameters of the problem tend to infinity. This yields easy randomized algorithms for constructing an object with the desired properties. Sometimes this positive probability is extremely small (like in applications of the Lovász Local Lemma). In those situations more complicated techniques are required to turn the proof to a constructive one.

8.1.1 Example 1: MAX_CUT

The input to the problem is similar to the min-cut problem: a graph $G(V, E)$ find. Our goal is to find a cut with maximum size. Unlike min-cut problem, Max-Cut is NP-hard. Here we show, using the probabilistic method, that *any* graph G has a cut of size at least $\frac{|E|}{2}$. Since the maximum cut size is at most $|E|$, if we can find a cut of size at least $\frac{|E|}{2}$ it yields a $\frac{1}{2}$-approximation. Consider the following simple randomized algorithm. We are creating the two parts $S, \overline{S}$ as follows: for every $v \in V$ we place it in S or in $\bar{S}$ with probability $\frac{1}{2}$ u.r. and independtly.

Claim: E[size of cut] $= \frac{|\mathrm{E}|}{2}$.

Consider every edge $e = uv$. The probability that u falls in a different part as v does is $\frac{1}{2}$. In other e contributes to the size of the cut with probability $\frac{1}{2}$. Thus the expected size of the cut is $\frac{|E|}{2}$. Note that this does not show that the randomized algorithm finds a cut of this size. It only gives a bound for the expected size.

8.1.2 Example 2: MAX_SAT

Consider the problem of Max-SAT. We have a CNF boolean formula Φ with over the boolean variables $x_1 ... x_n$ and which has clauses $C_1 ... C_m$. e.g. $(x_1 \vee \bar{x_2}) \wedge (x_3 \vee \bar{x_4} \vee \bar{x_1}) \wedge x_2$. The Max-SAT problem asks to find a truth assignment to x_i's to maximize the number of satisfied clauses.

It is easily seen that we can solve the SAT problem if we can solve the Max-SAT (simply see if the size of Max-SAT solution is m or not. Thus:

Theorem 8.1 *MAX_SAT is NP-hard.*

Lemma 8.2 *There is a truth assignment with size (# of clauses) at least* $\frac{m}{2}$.

Proof: Assign x_i's True or Fals uniformly randomly with probablity $\frac{1}{2}$ each. Let $Z_i = 1$ iff C_i is satisfied and define $Z = \sum_{i=1}^{m} Z_i$. Thus Z is the size of the solution and $\mathrm{E}[Z] = \sum_{i=1}^{m} \mathrm{E}[Z_i]$. If C_i has k variables then the probability that C_i is not satisfied is at most $1/2^k$. Thus $Z_i = 1$ with probability at least $1 - 2^{-k} \geq \frac{1}{2}$. Thus $\mathrm{E}[Z_i]\frac{1}{2}$ and $\mathrm{E}[Z] \geq \frac{m}{2}$. ∎

8.1.3 Example 3: Ramsey numbers

Consider the complete graph on 6 vertices K_6 and color every edge of it Red or Blue. It is an easy exercise to prove that no matter how we do the coloring, there is always a red triangle or a blue triangle. It is also not difficult to show that for K_5 (complete graph on 5 vertices) there is a 2-coloring of the edges with no monochromatic triangle. So 6 is the smallest integer n such that any 2-coloring of the edges of K_n either has a red triangle (i.e. complete subgraph on 3 vertices) or a blue triangle. This problem is a special case of a class of problems called after mathematician Ramsey.

Definition 8.3 *Ramsey nubmer $R(k,l)$ is the smallest integer n such that any 2-coloring of edges of K_n either has a red K_k or a blue K_l.*

From our definition and the discussion before we have $R(3,3) = 6$. Finding the exact value of $R(k,l)$ is a very difficult problem. Even finding reasonable upper and lower bounds for them is a challenging problem.

The first application of the probabilistic method was the following clever result of Erdös in 1947.

Theorem 8.4 $R(k,k) > \frac{k2^{\frac{k}{2}}}{e\sqrt{2}}$

Proof: First we prove the following easier result.

Lemma 8.5 *if* $\binom{n}{k}2^{1-\binom{k}{2}} < 1$ *then* $R(k,k) > n$

Proof: Take K_n, color the edges red and blue randomly with probability $\frac{1}{2}$ each. If S is a set with $|S| = s$ let $\mathcal{E}_s$ be the event that all the edges of S have the same color (i.e. the subgraph induced by S is monochromatic). Then $\Pr[\mathcal{E}_s] = 2^{1-\binom{s}{2}}$. There are $\binom{n}{k}$ sets S with size k. Therefore: $\Pr[\exists$ a monochromatic set of size $k] \leq \binom{n}{k}2^{1-\binom{k}{2}} < 1$. That is With positive probability, no event $\mathcal{E}_s$ happens (for any set of size k) i.e. no monochromatic subgraph on k vertices. ∎ Now using Stirling approximation for $n!$ we have: $n! \approx \frac{n}{e}^n\sqrt{2\pi n}$. With $n = \frac{k2^{\frac{k}{2}}}{e\sqrt{2}}$ we have:

$$\begin{aligned}
\binom{n}{k}2^{1-\binom{k}{2}} &< \frac{n^k}{k!} \cdot 2^{1-\frac{k^2}{2}+\frac{k}{2}} \\
&= \frac{2^{\frac{k^2}{2}}\left(\frac{k}{e\sqrt{2}}\right)^k}{k!} 2^{-\frac{k^2}{2}}2^{\frac{k}{2}+1} \\
&< 1
\end{aligned}$$

So it's enough to use the previous lemma. ∎

8.2 Derandomization using the method of conditional probability

We can turn each of the above two existential results into deterministic algorithms whose solution is as good as the bound for the expected size of randomized algorithm. This method of derandomization is due to Erdös and Spencer and is called the method of conditional probability. We explain the method through the Max-SAT example.

Lemma 8.6 *Suppose that we have assigned values $x_1 = a_1, \ldots, x_i = a_i$. Then we can compute the expected value of the solution given that the values for $x_1, \ldots, x_i$ are $a_1, \ldots, a_i$, respectively and the other variables are assigned u.r.*

Proof: Let Φ' be the reduced formula on $x_{i+1}...x_n$ obtained from Φ by substituting $x_1 = a_1, \ldots, x_i = a_i$, deleting the satisfied clauses, and also removing every occurence of $x_1, \ldots, x_i$ from the remaining clauses. Clearly this can be easily done in polynomial time. Also we can compute the expected value of the solution of Φ' assuming that the variables of Φ' (which are $x_{i+1}, \ldots, x_n$) are assigned T/F u.r. ■

The derandomized algorithm for Max-SAT will be as follows. Consider an ordering of the variables, say $x_1, \ldots, x_n$. We go through the variables one by one in this order and assign a value to each of them. Consider x_1: x_1 can be T or F. We compute the expected value for each case using the above lemma. Let Z be the number of satisfied clauses. If $\mathrm{E}[Z|x_1 = T] > \mathrm{E}[Z|x_1 = F]$ then we set x_1 = T. Otherwise set x_1 = F. Since $\mathrm{E}[Z] = \mathrm{E}[Z|x_1 = T] \cdot \Pr[x_1 = T] + \mathrm{E}[Z|x_1 = F] \cdot \Pr[x_1 = F] = \frac{1}{2}(E[Z|x_1 = T] + E[Z|x_1 = f])$, if we set $x_1 = v$ as above $E[Z|x_1 = v] \geq E[Z]$. Now a simple induction on i shows that after we have assigned values to $x_1, \ldots, x_i$ there is an extension of the current solution which has size at least $\frac{m}{2}$.

Exercise: Derandomize the randomized algorithm for Max-Cut.

For Example 3, we don't know how to make the proof into a constructive one. In other words, we don't have any explicit construction of lower bounds for $R(k, k)$ that are as good as the one guaranteed in the proof of Example 3.

8.3 Random Graph Models

We can define random graphs many different ways. Two very common models for random graphs are $G_{n,p}$ and $G_{n,M}$.

In model $G_{n,p}$, we have a graph on n vertices. Every edge is present with probability p u.r. and independetly. In model $G_{n,M}$ we have a graph on n vertices and M edges, where among all $\binom{\binom{n}{2}}{M}$ possible graphs on n vertices and M edges each graph has equal probabability of being chosen. Often the results proved for one model hold for the other model too.

To generate a graph from $G_{n,p}$: pick n vertices, place every edge with probability p. To generate a graph from $G_{n,M}$: pick n vertices, then place M edges u.r. without replacement.

Definition 8.7 *A vertex k-coloring of a graph G is a function $f : V \longrightarrow \{1, \ldots, k\}$ such that for every edge $e = uv$: $f(u) \neq f(v)$. The minimum value of k for which G has a k-coloring is the chromatic number of G and is denoted by $\chi(G)$.*

For example, it is easy to see that a graph is bipartite if and only if $\chi(G) = 2$.

Theorem 8.8 *For any $k > 2$ there is a triangle-free graph G with $\chi(G) \geq k$*

Proof: Consider $G_{n,p}$ with $p = n^{-\frac{2}{3}}$. Note that in any feasible coloring, the color classes are all independent sets. Therefore, if we show that the largest independent set of G, denoted by $\alpha(G)$, is smaller than $\frac{n}{k}$, then $\chi(G) \geq k$.

We prove that with positive probability $G_{n,p}$ does not have an independent set of size $\frac{n}{2k}$. This implies that there is some graph G on n vertices with $\alpha(G) > \frac{n}{2k}$.

Let X be the number of independent sets of size $a = \frac{n}{2k}$. Using the fact that $(1-x) \leq e^{-x}$:

$$\begin{aligned}
\mathrm{E}[X] &\leq \binom{n}{a}(1-p)^{\binom{a}{2}} \\
&\leq \left(\frac{en}{a}\right)^a \cdot e^{-p[\frac{a^2}{2}-\frac{a}{2}]} \\
&= (2ke)^{\frac{n}{2k}} \cdot e^{-n^{-\frac{2}{3}}[\frac{n^2}{4k^2}-\frac{n}{2k}]} \\
&= (2ek)^{\frac{n}{2k}} \cdot e^{[\frac{-n^{\frac{4}{3}}}{4k^2}+\frac{n^{\frac{1}{3}}}{2k}]} \\
&\leq e^{[\frac{n(\ln k+1)}{2k}-\frac{n^{\frac{4}{3}}}{4k^2}+\frac{n^{\frac{1}{3}}}{2k}]},
\end{aligned}$$

which is $<< \frac{1}{2}$ if n is large enough with respect to k (since $-\frac{n^{4/3}}{4k^2}$ will be the dominant term in the exponent).

Now let's compute the expected number of triangles in G.

$$\begin{aligned}
\mathrm{E}[\#\text{ of triangles}] &\leq \binom{n}{3}p^3 \\
&\leq \frac{n^3}{6} \cdot (n^{-\frac{2}{3}})^3 \\
&\leq \frac{n}{6} \\
&< \frac{n}{2}.
\end{aligned}$$

Thus, more than half of the graphs in $G_{n,p}$ have less than $\frac{n}{2}$ triangles each and for more than half of the graphs in $G_{n,p}$: $\alpha(G) < \frac{n}{2k}$. This implies that there is one graph with $\alpha(G) < \frac{n}{2k}$ which has less than $\frac{n}{2}$ triangles. Take such a graph and delete one vertex from each triangle. We get a triangle-free graph with chromatic number at least $\frac{n/2}{n/2k} = k$. ■