22C:296 Seminar on Randomization

Scribe: Saurav Pandit

September 17, 2003

From the result we proved last class, we can prove the following result. Let $f(k)$ denote the size of the smallest tournament with property S_k then $f(k) \leq k^2 \cdot 2^k \cdot \ln 2 \cdot (1 + o(1))$.

First Moment Method

For any random variable X, the quantity $E[X^k]$ is called the kth moment of X. So, in the first moment method, we will be talking about $E[X]$.

Here are the results we will use:

1. If $E[X] \leq t$, then $\text{Prob}[X \leq t] > 0$ (First Moment Principle).

2. If X is a non-negative random variable, then

$$\text{Prob}[X \geq t] \leq \frac{E[X]}{t} \qquad \text{(Markov's Inequality)}$$

3. Let $X_1, X_2, \ldots, X_t$ be random variables. Let, $c_1, c_2, \ldots, c_t$ be scalar constants, then

$$E[\sum_{i=1}^{t} c_i X_i] = \sum_{i=1}^{t} c_i E[X_i] \qquad \text{(Linearity of Expectation)}$$

These results can be easily proved. Consider for example, the following simple proof of the Markov's inequality.

$$\begin{aligned} E[X] &= \sum_i i \cdot \text{Prob}[X = i] \\ &\geq \sum_{i \geq t} i \cdot \text{Prob}[X = i] \\ &\geq \sum_{i \geq t} t \cdot \text{Prob}[X = i] \\ &= t \cdot \text{Prob}[X \geq t] \end{aligned}$$

This implies that $\text{Prob}[X \geq t] \leq E[X]/t$.

In the subsequent classes we will study three applications of the first moment method.

1. k-SAT.

2. Exisence of a high-girth, large-chromatic-number graphs.

3. Turan's theorem on dominating sets in graphs.

Here is the definition of the k-SAT problem.
`INPUT`: A boolean formula in CNF such that each clause has exactly k literals.
`QUESTION`: Is there a truth-assignment to the variables that satisfies all the clauses?
For example, an input to 3-SAT could be

$$(x_1 \vee \overline{x}_2 \vee x_4) \wedge (\overline{x}_1 \vee \overline{x}_2 \vee \overline{x}_3).$$

This is a *decision problem*. The optimization version of this problem is *MAX-k-SAT*. This problem seeks a truth-assignment that maximises the number of satisfied clauses.

Theorem 1 *Any instance of k-SAT with less than 2^k clauses is satisfiable.*

Proof: Construct a random truth assignment by setting each variable (independently) to TRUE or FALSE with equal probability. Let X_i be an indicator random variable defined as: $X_i = 1$ if clause i is not satisfied and $X_i = 0$ if clause i is satisfied. Let $X = \sum_i X_i$. Then X is the number of unsatisfied causes and $E[X] = E[\sum_i X_i] = \sum_i E[X_i]$ (by linearity of expectation). Since $\text{Prob}[X_i = 1] = 1/2^k$, we have $E[X_i] = 1/2^k$ for each clause. Hence,

$$E[X] = \frac{1}{2^k} \cdot (\text{number of clauses}).$$

If the number of clauses is less than 2^k, $E[X] < 1$. Therefore, $\text{Prob}[X < 1] > 0$ (by the first moment principle) and this implies that $\text{Prob}[X = 0] > 0$. Hence, there exists a truth assignment in which 0 clauses are unsatisfied. □

Here are some closely related results. The proof of (1) is very similar, the proof of (2) is given below.

1. Consider the CNF formula $F = C_1 \wedge C_2 \wedge \cdots \wedge C_m$. If $\sum_{i=1}^{m} 2^{-|C_i|} < 1$, then F is satisfiable.

2. For any $\varepsilon > 0$, there is a simply poly-time algorithm that solves SAT on any CNF formula on n variables such that each clause has size $\geq \varepsilon \cdot n$.

Proof of (2): If the instance has $< 2^{\varepsilon \cdot n}$ clauses, then, by (1), it is satisfiable. If the instance has $\geq 2^{\varepsilon \cdot n}$ clauses, then simply check all the 2^n truth-assignments since this is polynomial in $2^{\varepsilon \cdot n}$.

Our second example is on the existence of high-girth, large-chromatic number graphs. First some notation. For any graph G, we denote by $\omega(G)$ the size of a largest clique in G and by $\chi(G)$ the chromatic number of G. Clearly, $\chi(G) \geq \omega(G)$. Odd cycles are examples of graphs with $\chi(G) > \omega(G)$. Is it possible to construct graphs with small $\omega(G)$ but large $\chi(G)$? We will show the following theorem.

Theorem 2 *For any k there is a graph G_k such that $\omega(G_k) = 2$ and $\chi(G_k) = k$.*