CS395T Pseudorandomness and Combinatorial Constructions
Lecture 4, January 25, 2001
Lecturer: Prof. David Zuckerman, Scribe: Ramgopal Mettu

In this lecture, we start by examining some bounds on the deviation of random variables, namely Markov's inequality and Chernoff bounds. We then look at how Chernoff bounds can be used in the probabilistic method to establish some interesting facts in combinatorics. We then discuss the application of the probabilistic method in establishing lower bounds.

1 Useful Theorems From Probability

1.1 Markov's Inequality

Proposition 1 (Markov's Inequality) *Let X be a nonnegative random variable. Then for all $k > 0$, $\Pr[X \geq k] \leq \mathrm{E}[X]/k$.*

Proof: Suppose that $\Pr[X \geq k] > \mathrm{E}[X]/k$. Then, $\mathrm{E}[X] \geq \sum_{i \geq k} i \cdot \Pr[X = i] \geq k \cdot \Pr[X \geq k] > \mathrm{E}[X]$, which is a contradiction. □

1.2 Chernoff Bounds

We will now use Markov's inequality to bound the deviation of the sum of independent random variables from its expected value. We first establish a useful lemma.

Lemma 1 *Let $Z \in [-1, 1]$ be a random variable with $\mathrm{E}[Z] = 0$. Then, for all t, $\mathrm{E}\left[e^{t \cdot Z}\right] \leq e^{t^2/2}$.*

Proof: First, we can observe that $\mathrm{E}\left[e^{t \cdot Z}\right] \leq \frac{1}{2}(e^{-t} + e^{t})$. Intuitively, the way to maximize the expectation for all t is to set $\Pr[Z = 1] = \Pr[Z = -1] = 1/2$. Informally speaking, we can check that given any other distribution of mass, pushing some of the mass out symmetrically causes the expectation of $e^{t \cdot Z}$ to increase.

Then, we can verify that $\frac{1}{2}(e^{-t} + e^{t}) \leq e^{t^2/2}$ by examining the Taylor series of each side of the inequality and using induction to show that the inequality holds for corresponding terms in the Taylor expansions. □

Theorem 1 (Chernoff Bound) *Let $X = \sum_{i=1}^{n} X_i$, where the X_i's are independent and $|X_i - \mathrm{E}[X_i]| \leq c$. Then, $\Pr[X - \mathrm{E}[X] \geq c\lambda\sqrt{n}] \leq e^{-\lambda^2/2}$.*

Proof: First, we will assume without loss of generality that $c = 1$, since we can always normalize the X_i's. The main idea of the proof will be to use Markov's inequality on the random variable $e^{t(X - \mathrm{E}[X])}$, where t is a parameter to be chosen later. First, note that

$$\begin{aligned}
\mathrm{E}\left[e^{t(X-\mathrm{E}[X])}\right] &= \mathrm{E}\left[\prod_{i=1}^{n} e^{t(X_i - \mathrm{E}[X_i])}\right] \\
&= \prod_{i=1}^{n} \mathrm{E}\left[e^{t(X_i - \mathrm{E}[X_i])}\right] \\
&\leq \prod_{i=1}^{n} e^{t^2/2} \\
&= e^{nt^2/2},
\end{aligned}$$

where the second step follows since the X_i's are independent and the third step follows from Lemma 1. Now, we can observe that

$$\begin{aligned}
\Pr\left[X - \mathrm{E}[X] \geq \lambda\sqrt{n}\right] &\leq \Pr\left[e^{t(X-\mathrm{E}[X])} \geq e^{t\lambda\sqrt{n}}\right] \\
&\leq e^{-t\lambda\sqrt{n}} \cdot \mathrm{E}\left[e^{t(X-\mathrm{E}[X])}\right] \\
&\leq e^{\frac{nt^2}{2} - t\lambda\sqrt{n}},
\end{aligned}$$

where the second step follows from Markov's inequality.

By setting $t = \lambda/\sqrt{n}$, we can minimize $e^{\frac{nt^2}{2} - t\lambda\sqrt{n}}$ and obtain the bound in the theorem statement. □

Using arguments similar to those in the proof above, one can also obtain the following bound, which is useful when the expected value of X is $o(n)$.

Theorem 2 *Let $X = \sum_{i=1}^{n} X_i$, where the X_i's are independent and $X_i \in [0,1]$. Let $\mu = \mathrm{E}[X]$. Then, for all $\delta \geq 0$,*

$$\Pr[X > (1+\delta)\mu] < \left(\frac{e^{\delta}}{(1+\delta)^{(1+\delta)}}\right)^{\mu}$$

and

$$\Pr[X < (1+\delta)\mu] < e^{-\mu\delta^2/2}.$$

2 Applications of the Probabilistic Method

We will see that straightforward (in retrospect) probabilistic arguments yield interesting and strong bounds for some combinatorial questions.

2.1 Coloring Hypergraphs

Definition 1 *A* hypergraph *is a pair (V, E) where V is a set of vertices and E is a set of subsets of V.*

Definition 2 *A hypergraph is n-*uniform *if for all $e \in E$, $|e| = n$.*

Remark: A graph is a 2-uniform hypergraph.

Definition 3 *A* 2-coloring *is a coloring of the vertices with red and blue.*

Definition 4 *A* valid 2*-coloring of the hypergraph is a 2-coloring in which no edge is monochromatic.*

Theorem 3 (Erdös) *If $H = (V, E)$ is an n-uniform hypergraph, and $|E| < 2^{n-1}$, then there exists a valid 2-coloring.*

Proof: We will use the probabilistic method to establish the claim. That is, we will randomly 2-color a given hypergraph and show that the probability that the 2-coloring is invalid is less than 1. Thus, one of the random colorings we could pick is valid.

Fix a graph H and color each vertex uniformly and at random. The probability that a coloring is invalid is exactly the probability that there is a monochromatic edge. By the union bound, this probability is at most $|E|$ times the probability that a fixed edge is monochromatic. This quantity is at most $|E| \cdot 2/2^n$. Since $|E| < 2^{n-1}$, we have that the probability that a coloring is invalid is less than 1, establishing the claim. $\square$

2.2 Cliques and Independent Sets in Ramsey Graphs

Definition 5 *A* clique *in $G = (V, E)$ is a subset C of V, such that for all u, v in C, $\{u, v\} \in E$.*

Definition 6 *An* independent set *in $G = (V, E)$ is a subset I of V, such that for all $\{u, v\} \subseteq I$, $\{u, v\} \notin E$.*

Theorem 4 (Erdös) *There exist graphs on n vertices with no clique or independent set of size $k \geq 2\log_2 n$, assuming $k \geq 3$.*

Proof: We will use the probabilistic method to establish this claim. That is, we will construct a graph randomly and show that the probability that it has a clique or independent set of the desired size k is less than 1. Thus, one of the possible graphs we could construct must have a clique or independent set of size at least k.

First, construct a graph G on n vertices by choosing each edge uniformly at random. Then, the probability that there is a clique on k vertices is at most $\binom{n}{k}$ times the probability that a given set of size k is a clique. This probability is $2^{-\binom{k}{2}}$, since we must include exactly $\binom{k}{2}$ edges to construct a clique. Thus, the probability that there is a clique with k vertices is at most $C(n, k)2^{-\binom{k}{2}}$. To establish a bound on the probability that there is an independent set of size k, we use a similar argument, except we leave out edges instead of including them, yielding the same bound. Then, the probability that there is either a clique or an independent set of size k is $2 \cdot C(n, k)2^{-\binom{k}{2}}$. Now, observe that

$$
\begin{aligned}
2 \cdot C(n,k)2^{-\binom{k}{2}} &< 2 \cdot \frac{n^k}{k!} \cdot 2^{1-(k^2-k)/2} \\
&= \frac{2^{1+k/2}}{k!} \cdot \frac{n^k}{2^{k^2/2}} \\
&< 1.
\end{aligned}
$$

The last step follows since $k \geq 3$, i.e., $2^{1+k/2} < k!$, and since $n \leq 2^{k/2}$, i.e., $n^k \leq 2^{k^2/2}$ $\square$

It is worth noting that this non-constructive argument yields a much stronger upper bound than the best known explicit construction. The best known upper bound for the size of a clique and independent set in an explicitly constructed graph is $e^{\Theta(\sqrt{(\log n)\log\log n})}$. In the case of bipartite graphs, (where a clique is a complete bipartite subgraph and an independent set is still a subgraph with no edges), the gap between non-constructive and explicit results is even larger. A probabilistic argument yields an upper bound of $O(\log n)$, while the best known constructive upper bound is $O(\sqrt{n})$.

2.3 Lower Bounds

In the last lecture, we saw that the number of functions that have circuits of size at most ℓ is at most $(3\ell^2)^\ell$. If we think of ℓ as being polynomial in n, then there are very few "concise" functions: there are no more than an exponential (in n) number of functions that can be computed by polynomially sized circuits, whereas the total number of functions is doubly exponential in n. Indeed, the parameter ℓ needs to be exponential in n in order yield a reasonable number of functions. Thus, "most" functions require circuits of exponential size, which leads us to suggest the following conjecture and open question.

Conjecture: **NP** requires super-polynomial functions.

Open Question: Can we find a function in **NP** requiring super-linear circuits?

Evidence of the difficulty of constructing pseudo-random generators is that pseudo-random generators imply circuit lower bounds. To see this, suppose that we have access to a pseudorandom generator $f : \{0,1\}^s \to \{0,1\}^r$ which "fools" circuits of size at most ℓ. Then, we can construct an indicator function $g : \{0,1\}^r \to \{0,1\}$ for the range of the pseudorandom generator f, i.e., $g(x) = 1$ if $x \in f(\{0,1\}^s)$ and $g(x) = 0$ otherwise. Then, g cannot be computed by any circuit of size ℓ, otherwise there would be a circuit which could not be "fooled" by f.

Interestingly, circuit lower bounds also imply a derandomization result. Impagliazzo and Wigderson showed the following:

Theorem 5 (Impagliazzo & Wigderson) *If there is a language in $\mathcal{DTIME}(2^{O(n)})$ without circuits of size $2^{\alpha n}$, where $\alpha > 0$, for infinitely many values of n, then* $\mathbf{BPP} = \mathbf{P}$.