CSE599s: Extremal Combinatorics September 28, 2011

Lecture 1 Review of Some Basics

Lecturer: Anup Rao

1 Refresher of Basic Facts

The number of subsets of $[n] = 1, 2, \ldots, n$ of size k is $\binom{n}{k} = \frac{n!}{k! \cdot (n-k)!}$.

Theorem 1 (Binomial Theorem)**.** $(x+y)^n = \sum_{i=0}^{n} \binom{n}{i} x^i y^{n-i}$.

Proof There are exactly $\binom{n}{i}$ ways to get the monomial $x^i y^{n-i}$. ■

Proposition 2. *If* $0 < k < n$, $\binom{n}{k} = \binom{n-1}{k-1} + \binom{n-1}{k}$.

Proof Idea The set of size k either contains n or not. ■

Some estimates:

Proposition 3. *For* $n \geq k > 0$, $\left(\frac{en}{k}\right)^k \geq \binom{n}{k} \geq \left(\frac{n}{k}\right)^k$.

Proof $\binom{n}{k} = \frac{n}{k} \cdot \frac{n-1}{k-1} \ldots \frac{n-k+1}{1} \geq (n/k)^k$, since each of the k ratios in the product is at at least n/k. For the upper bound, observe that Taylor expansion gives $e^t = 1 + t + t^2/2! + \ldots \geq 1 + t$. Thus $(e^{k/n})^n \geq (1 + k/n)^n = \sum_{i=0}^{n} \binom{n}{i} (k/n)^i$ by the binomial theorem. Considering just the k'th term, we get $e^k \geq \binom{n}{k} (k/n)^k$. ■

An entropy based bound:

Proposition 4. $\binom{n}{\alpha \cdot n} = \frac{(1+o(1)) 2^{H(\alpha) n}}{\sqrt{2\pi\alpha(1-\alpha)n}}$, *where* $H(\alpha) = \alpha \log(1/\alpha) + (1-\alpha)\log(1/(1-\alpha))$ *is the binary entropy function.*

The proof is not pretty, but the intuition is that picking a random set of size k is similar to picking a random set where each element is included independently with probability k/n.

Selection with repetitions:

Proposition 5. *The number of non-negative integer solutions to* $x_1 + x_2 + \ldots + x_n = r$ *is* $\binom{n+r-1}{n-1}$.

Proof For every choice of $n-1$ elements S of $[n+r-1]$, we obtain such a solution. x_1 is the number of elements less than the first element of S, x_2 is the number of elements between the first two elements of S and so on. ■

$2^{[n]}$ represents the power set of $[n]$, and $\binom{[n]}{k}$ represents the set of sets of size k. A graph $G = (V = [n], E)$ is specified by a family of sets of size 2, $E \subseteq \binom{[n]}{2}$.

Proposition 6. $\binom{n}{2} = \binom{k}{2} + k(n-k) + \binom{n-k}{2}$.

Proof Idea Count the edges in the complete graph by counting the number of edges inside a set S of k vertices, the number outside, and the number that cross. ■

Proposition 7 (Little Fermat)**.** *For any prime p, $a^p = a \mod p$.*

Proof First note that $(a+1)^p = \sum_{i=0}^{p} \binom{p}{i} a^i$ by the binomial theorem. However, $\binom{p}{i} = \frac{p(p-1)\dots}{i(i-1)\dots}$ is divisible by p when $0 < i < p$. Therefore $(a+1)^p = a^p + 1 \mod p$. The proof is then completed by induction on a. ■

Given a family of sets $\mathcal{F} \subseteq 2^{[n]}$, let $d(x)$ denote the number of sets containing x.

Proposition 8 (Double counting)**.** $\sum_{x \in [n]} d(x) = \sum_{A \in \mathcal{F}} |A|$.

Proof Consider the bipartite graph where the left vertices are $[n]$ and the right vertices are $\mathcal{F}$, with an edge (x, S) exactly when $x \in S$. The left hand side is the number of edges in the graph counted from the left. The right hand side is the number of edges counted from the right. ■

2 The Chernoff-Hoeffding Bound

Here we give a proof of the Chernoff bound from [1]. The proof is simple, and applies in a variety of settings where true independence is not available.

Let $X_1, \dots, X_n$ be independent binary random variables such that

$$X_i = \begin{cases} 1 & \text{with probability } p \\ 0 & \text{with probability } 1-p \end{cases}$$

Then we shall prove:

Theorem 9. $\Pr[\sum_i X_i \geq pn(1+\epsilon)] < 2^{-\epsilon^2 pn/4}$.

Proof Consider the following mental experiment. We sample the X_i's, and if at least $k = pn(1+\epsilon)$ of the variables turn out to be 1, we pick a uniformly random subset $S \subset [n]$ of $t = \epsilon pn/2$ of the coordinates that are 1, and blame S.

The probability that any fixed set $T \subset [n]$ is blamed is at most

$$\Pr[X_i = 1, \forall i \in T] \cdot \Pr[T \text{ is blamed} | X_i = 1, \forall i \in T] \leq \frac{p^t}{\binom{k}{t}}.$$

Note that in general there can be more than k coordinates that are 1, in which case the odds that T will be picked can only be smaller than $1/\binom{k}{t}$.

$\Pr[\sum_i X_i \geq pn(1+\epsilon)]$ is the same as the probability that *any* set is blamed, which by the union bound is at most

$$\begin{aligned}
\frac{p^t \cdot \binom{n}{t}}{\binom{k}{t}} &= \frac{p^t \cdot n! \cdot (k-t)! \cdot t!}{(n-t)! \cdot k! \cdot t!} \\
&< \frac{p^t \cdot n^t}{(k-t)^t} \\
&= \left(\frac{pn}{k-t}\right)^t \\
&= \left(\frac{1}{1+\epsilon/2}\right)^{\epsilon pn/2}.
\end{aligned}$$

Using the fact that $2^x \leq 1+x$ for $x \in [0,1]$, we get that this probability is at most $2^{-\epsilon^2 pn/4}$.
■

Remark For the above proof to work, it is sufficient to have that the probability of seeing only ones in a fixed set T of t coordinates is exponentially small in t. This is a condition that can often be satisfied even if the variables are not truly independent.

References

[1] Jeanette P. Schmidt, Alan Siegel, and Aravind Srinivasan. Chernoff–Hoeffding bounds for applications with limited independence. *SIAM Journal on Discrete Mathematics*, 8(2):223–250, May 1995.