CSE599s: Extremal Combinatorics October 5, 2011

Lecture 3 Sets with Small Pairwise Intersection, Inclusion Exclusion

Lecturer: Anup Rao

1 Sets with Small Pairwise Intersection

Suppose we want a family of sets $\mathcal{F} \subseteq 2^{[n]}$ such that every set in the family is of size exactly r, and every pair of sets has an intersection of size k. How many such sets can we have?

This question has found a very beautiful application in computer science. It was used in the construction of the Nisan Wigderson pseudorandom generator, and is a crucial component used in the proof of the following theorem:

Theorem 1. *One of the following statements must hold:*

- *Every language that can be computed in time $2^{O(r)}$ (here r is the input length) can be computed by a boolean circuit family with circuits of size $2^{o(r)}$.*
- *Every randomized polynomial time algorithm has a deterministic polynomial time simulation (i.e. $P = BPP$).*

Just to give you an idea of how the proof will go, suppose f is a boolean function that violates the first condition, and let $\mathcal{F} = \{S_1, \ldots, S_t\}$. Then given n random bits $x_1, \ldots, x_n$, we will generate t "pseudorandom bits" $f(x_{S_1}), f(x_{S_2}), \ldots, f(x_{S_t})$ by evaluating f on each of the projections of the bits to the r coordinates of S_i. The fact that the pairwise intersections of the sets is small will be used to show that no circuit can distinguish the future bits from the past bits (this step is not at all trivial).

These results are based on showing that for every constant c, there is such a family $\mathcal{F}$ of size $|\mathcal{F}| = 2^{n^{1/2c}}$, with sets of size $r = \sqrt{n}$ whose pairwise intersection is at most $k = n^{1/2c}$.

We shall discuss this proof in detail later in the course. For now, we shall prove a simple upper bound on the size of the family.

Lemma 2 (Corradi)**.** *If $r^2 > kn$, $|\mathcal{F}| \leq \frac{rn-kn}{r^2-kn}$.*

Proof For every set S in the family, we have

$$\sum_{x \in S} d(x) = \sum_{T \in \mathcal{F}} |S \cap T| \leq r + (|\mathcal{F}| - 1)k.$$

So this bound holds even for the average set S, i.e.:

$$r + (|\mathcal{F}| - 1)k \geq (1/|\mathcal{F}|) \sum_{S \in \mathcal{F}, x \in S} d(x).$$

Suppose the family was nice in that every $d(x)$ took the average value, i.e. $d(x) = |\mathcal{F}|r/n$. Then we would get

$$\begin{aligned} r + (|\mathcal{F}| - 1)k &\geq (1/|\mathcal{F}|) \sum_{S \in \mathcal{F}, x \in S} |\mathcal{F}|r/n = |\mathcal{F}|r^2/n \\ \Rightarrow |\mathcal{F}|(r^2 - kn) &\leq rn - kn \\ \Rightarrow |\mathcal{F}| &\leq \frac{rn - kn}{r^2 - kn}, \end{aligned}$$

where we used the fact that $r^2 > kn$ in the last step.

To complete the proof, we show that the equal degree case is the worst case.

$$\sum_{S \in \mathcal{F}, x \in S} d(x) = \sum_{x \in [n]} d(x)^2 = n \sum_{x \in [n]} d(x)^2/n \leq n \left(\sum_{x \in [n]} d(x)/n \right)^2 = n|\mathcal{F}|^2 r^2/n^2 = |\mathcal{F}|^2 r^2/n.$$

■

2 The Inclusion Exclusion Principle

Given two sets A, B, we have that $|A \cup B| = |A| + |B| - |A \cap B|$. Suppose we are given a family of sets $\{A_1, \ldots, A_n\}$. Can we estimate the size of their union in terms of their intersection sizes?

For every non-empty subset $I \subseteq [n]$, define $A_I = \bigcap_{i \in I} A_i$.

Then we can estimate $|\bigcup_{i=1}^n A_i|$ by $\sum_{i=1}^n |A_i|$ but we would have over counted elements that appear in multiple sets. We can subtract $\sum_{I \subseteq \binom{[n]}{2}} |A_I|$ but then we would have undercounted elements that appear in 3 sets.

Proposition 3 (Inclusion Exclusion)**.** $|\bigcup_{i=1}^n A_i| = \sum_{\emptyset \neq I \subseteq [n]} (-1)^{|I|+1} |A_I|$

Proof Consider each element x which is in the union of all the sets. Let $J = \{i : x \in A_i\}$. Then x contributes to a term in above sum exactly when $x \in A_I$, which means that $I \subseteq J$. Thus, the total contribution of x is

$$\sum_{\emptyset \neq I \subseteq J} (-1)^{|I|+1} = 1 + \sum_{I \subseteq J} (-1)^{|I|} = 1 - (1-1)^{|J|} = 1.$$

■

If we adopt the convention that $A_\emptyset$ is the set of all elements in the universe, then the above argument shows:

Proposition 4. *The number of elements not in* any *of the sets is* $\sum_{I \subseteq [n]} (-1)^{|I|} |A_I|$.

Proof If the universe is of size t, then

$$\sum_{I \subseteq [n]} (-1)^{|I|} |A_I| = t - \sum_{\emptyset \neq I \subseteq [n]} (-1)^{|I|+1} |A_I| = t - |\bigcup_{i=1}^n A_i|.$$

■

Next, let us see some applications of inclusion-exclusion.

2.1 Counting derangements

Question: How many permutations π of $[n]$ are there such that for each $j \in [n]$, $\pi(j) \neq j$?

Define A_j to be the set of permutations where j is mapped to j. So $|A_I| = (n - |I|)!$. Then the set of permutations that *do not* fix a point is just

$$\sum_{I \subseteq [n]} (-1)^{|I|} |A_I| = \sum_{i=0}^{n} (-1)^i \binom{n}{i} (n-i)! = n! \sum_{i=0}^{n} \frac{(-1)^i}{i!}$$

2.2 Euler's Totient function

Question : Given a positive integer N, with prime factorization $p_1^{e_1} p_2^{e_2} \cdots p_t^{e_t}$, how many numbers from 1 to N are relatively prime to N?

Let A_i denote the set of numbers that are divisible by p_i. Then $|A_I| = \frac{N}{\prod_{j \in I} p_j}$. So the number of relatively prime numbers is

$$\sum_{I \subseteq [t]} (-1)^{|I|} |A_I| = \sum_{I \subseteq [t]} (-1)^{|I|} \frac{N}{\prod_{j \in I} p_j} = N \prod_{i=1}^{t} (1 - 1/p_i)$$