CSE599s: Extremal Combinatorics October 3, 2011

Lecture 2 Jensen's Inequality and Probabilistic Magic

Lecturer: Anup Rao

1 Using Random Choices

The probabilistic method allows us to show that objects with nice combinatorial properties exist (even if it may be hard to find such objects). We shall illustrate this with some examples. We shall need the following fact, which you can prove yourself:

Fact 1 (Linearity of Expectation)**.** *If X, Y are real valued random variables,*

$$\mathbb{E}[X+Y] = \mathbb{E}[X] + \mathbb{E}[Y].$$

1.1 MAX Cut

A *cut* in a graph $G = (V, E)$ is a subset $S \subset V$. The size of the cut is the number of edges that cross the cut, i.e. $|\{u, v\} \in E | u \in S, v \notin S\}|$.

In the MAX-Cut problem, we are given a graph and wish to find a cut of largest size. The decision version (i.e. given a graph and k, decide whether or not there is a cut of size k) is NP-hard. However, an easy algorithm can give pretty good cuts:

Proposition 2. *Every graph has a cut of size at least $|E|/2$.*

Proof Let S be a uniformly random subset of the vertices. For each edge e, let X_e be the indicator random variable for whether or not e is cut. Then X_e is just a random bit, so $\mathbb{E}[X_e] = 1/2$. The size of the cut is $\sum_{e \in E} X_e$, and we have that $\mathbb{E}\left[\sum_{e \in E} X_e\right] = |E|/2$ by linearity of expectation.

Thus, there must exist one set S which cuts at least $|E|/2$ edges. ■

The above argument can be easily derandomized to give a deterministic algorithm that always finds a cut of size at least $|E|/2$, and so is within a factor of 2 from best possible.

1.2 Independent Sets

A subset S of the vertices of the graph is said to be *independent* if $\binom{S}{2}$ contains no edge of the graph. Given a graph, how can one find an independent set of largest size? Again, the decision version of this problem is NP-hard. We shall give a simple probabilistic algorithm for finding a large independent set when the number of edges in the graph is small. This will illustrate the *deletion method*.

Suppose the graph has n vertices and m edges. For a parameter p, let us pick a random subset S by including each vertex independently with probability p. Let $X = |S|$. Then $\mathbb{E}[X] = pn$. Let Y be the number of edges included in S. Then $\mathbb{E}[Y] = p^2 m$. We can obtain an independent set of size $X - Y$ by simply deleting one vertex from the set for each of the Y edges. In this way, we obtain an independent set of size $X - Y$. $\mathbb{E}[X - Y] = \mathbb{E}[X] - \mathbb{E}[Y] = pn - p^2 m$.

This quantity is maximized when $p = n/2m$, which gives an expected size of $n^2/2m - n^2/4m = n^2/4m$. We have proved:

Proposition 3. *There is always an independent set of size $n^2/4m$.*

2 Jensen's Inequality

Definition 4. *We say that a function $f : (a, b) \to \mathbb{R}$ is* convex *if for every $x, y \in (a, b)$ and every $\lambda \in (0, 1)$,*

$$f(\lambda x + (1-\lambda)y) \leq \lambda f(x) + (1-\lambda)f(y).$$

Examples of convex functions include x, e^x, x^2 and $\log(1/x)$. If $-f$ is convex, we shall say that f is *concave.* Note that if f'' is strictly positive, then f is convex.

The following is a useful inequality for dealing with the entropy function and its derivatives:

Lemma 5 (Jensen's Inequality)**.** *If f is a convex function on (a, b) and X is a random variable taking values in (a, b), then*

$$f(\mathbb{E}[X]) \leq \mathbb{E}[f(X)]$$

Proof We prove the case when X takes on finitely many values. The general case follows by continuity arguments.

We prove the statement by induction on the number of elements in the support of X. If X is supported on 2 elements, the lemma immediately follows from the definition of convexity. In the general case, let us assume X is supported on $x_1, \ldots, x_n$. Then,

$$\begin{aligned}
\mathbb{E}[f(X)] &= p(x_1)f(x_1) + \sum_{i=2}^{n} p(x_i)f(x_i) \\
&= p(x_1)f(x_1) + (1-p(x_1)\sum_{i=2}^{n} p(x_i)f(x_i)/(1-p(x_1)) \\
&\geq p(x_1)f(x_1) + (1-p(x_1)f\left(\sum_{i=2}^{n} p(x_i)x_i/(1-p(x_1))\right) \\
&\geq f\left(p(x_1)x_1 + (1-p(x_1)\left(\sum_{i=2}^{n} p(x_i)x_i/(1-p(x_1))\right)\right) \\
&= f(\mathbb{E}[X]),
\end{aligned}$$

where the first inequality follows by applying the lemma for the case when there are $n-1$ elements in the support, and the second inequality is a direct consequence of the definition of convexity. ■

Proposition 6. *For a fixed non-negative integer a, $\binom{x}{a} = \frac{x\cdot(x-1)\cdot\ldots(x-a+1)}{a\cdot(a-1)\cdot\ldots 1}$ is a convex function on $[a, \infty]$.*

Proof Every derivative of this function is non-negative. ■

A consequence is the arithmetic geometric mean inequality:

Proposition 7. *For positive $x_1, \ldots, x_n$, $\frac{x_1+x_2+\ldots+x_n}{n} \geq \sqrt[n]{x_1 \cdot x_2 \cdot \ldots x_n}$.*

Proof Let Y be a random variable taking the value $\log x_i$ with probability $1/n$. Then the left hand side is $\mathbb{E}\left[2^Y\right]$ and the right hand side is $2^{\mathbb{E}[Y]}$. The inequality follows from the convexity of exponentiation. ■

3 Zarankiewicz's Problem

This is the first true example of an extremal problem that we have encountered. We shall focus on bipartite graphs G with n vertices on the left and n vertices on the right. Recall that a $a \times b$ is a set of a vertices on the left and a set of b vertices on the right such that every pair of two vertices is an edge of the graph.

Question: How many edges can G have without containing an $a \times a$ clique?

Obviously, the complete graph has such a clique, and the empty graph doesn't. So there must be some critical number $k_a(n)$ such that any graph with $k_a(n)$ edges has an $a \times a$ clique, but there is a graph with one fewer edges that does not have such a clique.

Theorem 8. *For n large enough in terms of a,*

$$2(a-1)^{1/a} n^{2-1/a} \geq k_a(n) \geq n^{2-2/a}/2.$$

3.1 Lower Bound

We use the deletion method. For a parameter p that we shall pick later, include each edge (u, v) in the graph with probability p. Let X denote the number of edges in the graph, and let Y denote the number of $a \times a$ cliques. We can find a graph with $X - Y$ edges that does not contain any cliques by deleting one edge from each of the Y cliques.

$\mathbb{E}[X] = pn^2$, and $\mathbb{E}[Y] = \binom{n}{a}^2 p^{a^2}$, so by linearity of expectation, $\mathbb{E}[X - Y] = pn^2 - \binom{n}{a}^2 p^{a^2}$, so there is a graph that many edges that has no clique. This quantity is maximized when

$$n^2 = a^2 \binom{n}{a}^2 p^{a^2-1} \Rightarrow p = \left(\frac{n^2}{a^2 \binom{n}{a}^2} \right)^{1/(a^2-1)},$$

but to make our life easy let us calculate what we get when $p = n^{-2/a}$, which is close enough. This gives

$$\begin{aligned} k_a(n) &\geq pn^2 - \binom{n}{a}^2 p^{a^2} \\ &= n^{-2/a} n^2 - \binom{n}{a}^2 n^{-2a} \\ &\geq n^{2-2/a} - 1 \end{aligned}$$

3.2 Upper Bound

Now we wish to show that if the graph has enough edges, there must be a clique. We shall use a little bit of double counting, and a little bit of Jensen.

Suppose the graph has k edges and does not have any $a \times a$ cliques. Given a set of A of a vertices on the left, say that the tuple (A, x) forms a *star* if every (u, x) for $u \in A$ is an edge of the graph.

The point of counting stars is that unlike cliques, stars are much easier to count in terms of the number edges in the graph, and the number of stars is closely related to the number of cliques. Since the graph has no $a \times a$ cliques, each such set A can have at most $a-1$ stars. Therefore, the total number of stars is at most $\binom{n}{a}(a-1)$. On the other hand, each x contributes $\binom{d(x)}{a}$ stars. The average value of $d(x)$ is k/n. If every vertex on the right had degree exactly k/n, then we would find $n\binom{k/n}{a}$ stars, establishing that

$$\begin{aligned}
\binom{n}{a}(a-1) &\geq n\binom{k/n}{a} \\
\Rightarrow (a-1)/n &\geq \frac{\binom{k/n}{a}}{\binom{n}{a}} \\
&\geq (k/n - a + 1)^a / n^a \\
\Rightarrow (a-1)^{1/a} n^{2-1/a} + (a-1)n \geq k
\end{aligned}$$

Of course, in reality, we don't know that each vertex has exactly k/n degree. But Jensen will save us, showing us that the case of equal degrees is actually the worst case. The actual number of stars in the graph is

$$\sum_x \binom{d(x)}{a} = n\left(1/n \sum_x \binom{d(x)}{a}\right) \leq n\binom{1/n \sum_x d(x)}{a} = n\binom{k/n}{a},$$

where the inequality is by the convexity of the function $f(y) = \binom{y}{a}$.