Probability and Algorithms | **Caltech CS150, Fall 2018**
Leonard J. Schulman, schulman@caltech.edu | **TA: Jenish Mehta, jenishc@gmail.com**
Problem set 4 | **Out F 16/Nov Due W 28/Nov in Jenish's mailbox**

1. The Chernoff bounds I showed in class were each derived for particular distributions, so we could write down an mgf explicitly. But sometimes we don't have that luxury, and need to use a Chernoff bound that relies only on limited information. Here is a case in point.

   Let $X_1, \dots, X_n$ be independent (but not necessarily identically distributed) random variables, each taking values in the interval $[0, 1]$. Let $\overline{X} = \sum X_i/n$ and let $\theta = E(\overline{X})$. Let $\varepsilon > 0$.

   (a) Show that $\Pr(\overline{X} > (1+\varepsilon)\theta) < \exp(-\theta n((1+\varepsilon)\log(1+\varepsilon) - \varepsilon))$.

   (b) Show that $\Pr(\overline{X} < (1-\varepsilon)\theta) < \exp(-\theta n\varepsilon^2/2)$.

   (We used (b) as a lemma in the proof of Bourgain's L_1-embedding theorem. We needed there only the special case that the X_i are Bernoulli rvs.)

   Hint: Upper bound the mgf of X_i by that of a Bernoulli distribution.

2. In this problem I'm going to ask you to prove a generalization of the threshold we studied for appearance of K_4 in $G(n, p)$. Given a graph H, let its "peak density" be

   $$\rho(H) = \max_{H' \subseteq H} |E(H')|/|V(H')|$$

   where $H' \subseteq H$ denotes that H' is a subgraph of H, and where V and E are the vertex and edge sets respectively.

   You are to show that $p = n^{-1/\rho}$ is the threshold for appearance of H in $G(n, p)$. To simplify matters, it is enough for this problem if you show that for $\varepsilon > 0$, the probability tends to 0 for $p = n^{-1/\rho-\varepsilon}$ and to 1 for $p = n^{-1/\rho+\varepsilon}$.

   Let $h = |V|$ and $e = |E|$. It is certainly possible for ρ to be larger than e/h. For example in the "spatula graph" (Fig. 1) with $h = 5, e = 6$, $\rho = 5/4$ and the threshold is $n^{-4/5}$ (rather than $n^{-5/6}$ as you might have been tempted to think).

   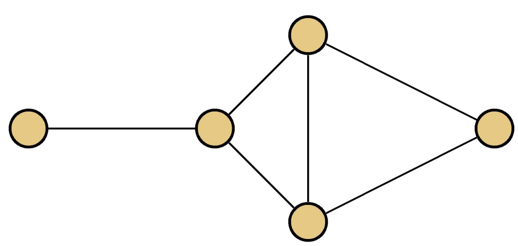

   Figure 1: An abstract spatula

   Hint: One of the superficial complications in this problem is that the automorphism group of H may be anything from trivial up through all permutations of the vertices (which is what we encountered for K_4). For instance the automorphism group of the spatula graph has size 2. This range of possibilities complicates counting how many ways H might turn up in the graph G that we sample from $G(n, p)$. The trick is that, since H is of fixed size, you can afford to count wastefully.

   I want you to notice something carefully: there is in this problem a "fake threshold" $n^{-h/e}$. Once p rises above $n^{-h/e}$, the expected number of copies of X in G is $\omega(1)$. Despite this, with probability $1 - o(1)$ there are no copies of X in G, until p reaches $n^{-1/\rho}$.

3. Show that the value of the nine-free problem, $f(n)$ (see previous problem set), is $O(n^{5/3})$.

   (In actual fact there is a matching $\Omega(n^{5/3})$ bound, but I am not assigning that.)

4. Consider a collection of vectors $v_i = (x_i, y_i)$ (for $i = 1, \ldots, n$), such that for all i, x_i and y_i are integer, and $|x_i|, |y_i| < 2^{n/2}/(c\sqrt{n})$. For a subset $I \subseteq [n]$ let $v_I = \sum_{i \in I} v_i$.

   Show that there is a $c < \infty$ such that (for any n and any set of vectors as above) there are disjoint subsets I, J s.t. $v_I = v_J$.

   Hint: Sample a uniform random subset I, and (writing $v_I = (x_I, y_I)$), consider the variance of x_I.

5. There are two disjoint sets A, B with $A \cup B = [n]$ and $|A| \geq cn$, for some constant $c > 0$. Consider sampling a set $R \subseteq [n]$ as follows (with R_i denoting the event that $i \in R$): the R_i are pairwise independent, and $\Pr(R_i) = p$ for all i.

   Show that there is a $d > 0$ depending only on c such that for all n there is a p such that $\Pr((A \cap R \neq \emptyset) \wedge (B \cap R = \emptyset)) \geq d$.

   How can we use this in regard to Bourgain's L_1 embedding?