Random discrete structures (MPI, 2014) **Lecturers:** Kunal Dutta and Arijit Ghosh
Topic: Final exam **Date:** 04/08/2014
Scribe: Kunal Dutta and Arijit Ghosh **Due:** 08/08/2014

By *random variables* or *discrete random variables* we mean random variables taking either finitely many values or countably infinite values.

Short problems. Each of the following short problems comprise of five credits.

1. Let X be a random variable.
   (a) For all even integer $k \geq 2$, show that $\mathbf{E}\left[X^k\right] \geq \mathbf{E}[X]^k$.
   (b) If X is a non-negative integer valued random variable then $\mathbf{Pr}[X > 0] \leq \mathbf{E}[X]$.
   (c) Show that $\mathbf{Pr}[X = 0] \leq \frac{\mathbb{V}ar[X]}{\mathbf{E}[X]^2}$.
2. Let a coin be such that heads comes up independently with probability p on each flip. What is the expected number of flips until you get k heads?
3. Suppose cards are being drawn at random with replacement from a deck of n cards.
   (a) What is the expected number of cards we must draw until we have seen all n cards?
   (b) Suppose we have drawn only $2n$ times. What is the expected number of different cards chosen from the deck?
4. Consider the random walk on the integer numbers starting at 0. Let $F(n)$ be the expected number of steps for this walk to reach either $-n$ or $+n$. Give exact expression for $F(n)$.
5. For $n \geq 4$ and let $H = (V, E)$ be an n-uniform hypergraph.
   (a) If $|E| \leq 4^{n-1}$, then there exists a coloring of vertices of H such that very edge in H contains vertices with at least two different colors.
   (b) If $|E| \leq \frac{4^{n-1}}{3^n}$, then show that there exists a coloring of vertices of H such that very edge in H contains vertices of all four colors.
6. We saw in class that $(X, d) \xhookrightarrow{O(\log n)} l_1^{O(\log^2 n)}$ where n is the number of points in the metric space (X, d), i.e., $\#X = n$. Using this result and Holder's inequality show that $(X, d) \xhookrightarrow{O(\log n)} l_p^{O(\log^2 n)}$.

   Hölder's inequality: $\|x\|_p \cdot \|y\|_q \geq \langle x, y\rangle$ where $\frac{1}{p} + \frac{1}{q} = 1$.
7. Let $\mathcal{F} = \{S_1, \ldots, S_m\}$ be a collection of m sets with $S_i \subset [n] = \{1, \ldots, n\}$. Given an assignment $\chi : [n] \to \{1, -1\}$, the *discrepancy* of a set $S \subseteq [n]$ is defined as

   $$\mathrm{disc}_\chi(S) = \mid \sum_{x \in S} \chi(x) \mid .$$

   Discrepancy of $\mathcal{F}$ wrt χ is defined as

   $$\mathrm{disc}_\chi(\mathcal{F}) = \max_{S_i \in \mathcal{F}} \mathrm{disc}_\chi(S_i).$$

   Using a random assignment $\chi : [n] \to \{1, -1\}$ and Chernoff bound show existence of a coloring such that the discrepancy of $\mathcal{F}$ wrt to that coloring is $O(\sqrt{n \log m})$.

8. Let C_n denotes a cycle (graph) with n-vertices, and $d(,)$ denotes the shortest distance metric between the vertices of C_n (note that we assume edges in C_n and unit length). Show that C_n can be embedded in a familty of trees $\mathcal{F}_n$ on n-vertices such that

$$1 \leq \frac{E_{T \leftarrow \mathcal{F}}[d_T(x, y)]}{d(x, y)} < 2,$$

where $d_T(x, y)$ denotes the shorest distance metric in the tree T.

9. Let p be a prime number, and m, n are positive integers. Given a set of m linear equalities over n variables mod$p$ show that there exists an assignment to the variables that will satisfy at least $\frac{1}{p}$ of linear equalities. What happens if p is not a prime?

10. Consider the general setting of the Lovász local lemma, i.e. probability space Ω, and "bad" events $A_1, \ldots, A_m$ which we want to forbid. Now let B be another event in the same probability space. Suppose that B depends on the set of events $\Gamma(B) \subseteq \{A_1, \ldots, A_m\}$. Then prove that,

$$\mathbf{Pr}\left[B \mid \prod_{i=1}^{m} \bar{A}_i\right] \leq \frac{\mathbf{Pr}[B]}{\prod_{i \in \Gamma(B)}(1 - x_i)},$$

where x_i's are the reals corresponding to the events A_i in the usual statement of the Lovász Local Lemma.

Long problems. Each of the following long problems comprise of 10 credits.

1. Consider a fair die, i.e., when we roll the die the probability we get i, for $i \in \{1, 2, 3, 4, 5, 6\}$, is equal to 1/6. What is the expected number of rolls until the first pair of consecutive sixes appears?

2. Given a graph $G = (V; E)$ with $V = n$, a dominating set for G is a subset $D \subseteq V$ such that each vertex $v \in V$ is either in D or has a neighbor in D.
   (a) Show that any graph with minimum degree δ has a dominating set of size at most $\frac{n \log n}{\delta+1}$.
   (b) Improve the bound for dominating set to $\frac{n(1+\log(\delta+1))}{\delta+1}$

3. See the definition of discrepancy given above in problem-(7). In that problem we were interested in upper bounding discrepancy of a family of sets $\mathcal{F}$. In this problem we want to lower bound discrepancy, i.e., we want to show there exists a family of sets $\mathcal{F}$ with n subsets of $[n]$ such that for every coloring $\chi : [n] \to \{1, -1\}$, $\text{disc}_\chi(\mathcal{F}) > \Omega(\sqrt{n})$. Complete the proof by proving the following subproblems:
   (a) For a fixed $\chi : [n] \to \{1, -1\}$, pick a subset $S \subseteq [n]$ by including each element in S with probability $\frac{1}{2}$. Show that there exists a constant $c > 0$ such that

$$\mathbf{Pr}\left[\text{disc}_{chi}(S) > \sqrt{n}/c\right] > 1/2.$$

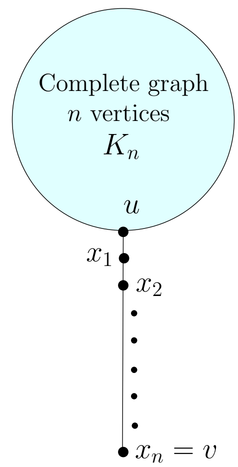
   (b) Let $\mathcal{F}$ consists of n sets picked independently as above. Show that for any fixed assignment $\chi : [n] \to \{1, -1\}$, $\text{disc}_\chi(\mathcal{F}) > \sqrt{n}/c$ with probability $> 1 - \frac{1}{2^n}$.
   (c) Using union bound over all 2^n assignments to show existence of a family $\mathcal{F}$ with n sets for which all assignments $\chi : [n] \to \{1, -1\}$ have discrepancy $> \sqrt{n}/c$.

4. Let F be an infinite family of graphs. Let $G \in \mathcal{G}(n, p)$, where $p = p(n)$ and $\mathcal{G}(n, p)$ denotes the Erdős-Reńyi random graph model. Suppose F, p are such that the expected number of subgraphs on k vertices, belonging to the family F, in G is $t > 1$, t fixed. Prove that there exists a graph on n vertices which does not contain any subgraph of size k from the family F.

5. Let $G = (V, E)$ be an undirected graph and suppose each $v \in V$ is associated with a set $S(v)$ of $8r$ colors, where $r \geq 1$. Suppose, in addition that for each $v \in V$ and $c \in S(v)$ there are at most r neighbours u of v such that c lies in $S(u)$. Prove that there is a proper coloring of G assigning to each vertex v a color from its class $S(v)$ such that, for any edge $(u, v) \in E$, the colors assigned to u and v are different.

   [Hint: Consider the family of events $A_{u,v,c}$, such that u and v are both colored with color c.]

6. Consider the Lollipop graph given below:



   The top part is a complete graph K_n on n vertices, and the lower tail is a path of length n. Compute tight bounds for the following quantities:

   (a) The expected time of the random walk starting at u to arrive at v.

   (b) The expected time of the random walk starting at v to arrive at u.

   (c) The expected time of the random walk starting at v to visit all the vertices in the graph.

   (d) The expected time of the random walk starting at u to visit all the vertices in the graph.

7. Show that for a n point set $P \subset \mathbb{R}^D$ and $d = O\left(\frac{\log n}{\varepsilon^2}\right)$ there exists a linear map $f : \mathbb{R}^D \to \mathbb{R}^d$ such that for all p_i, p_j, $p_k \in P$,

$$| (p_j - p_i)^{\mathrm{T}}(p_k - p_i) - (f(p_j) - f(p_i))^{\mathrm{T}}(f(p_k) - f(p_i)) | \leq \varepsilon \, \|p_j - p_i\| \|p_k - p_i\|.$$

   [Hint: See the proof of Johnson-Lindenstrauss lemma covered in class.]

8. Suppose we are given a discrete probability distribution X on $[n]$, i.e., we are given $p_i \geq 0$ such that $\mathbf{Pr}[X = i] = p_i$ and $\sum_{i=1}^{n} p_i = 1$. We want to preprocess this distribution in $O(n)$ time such that given an oracle that uniformly generates random number from the interval $[0, 1]$, we should be able to sample the distribution X in $O(1)$ time for each query. Assume that we are working in the real RAM model of computation.

9. We are given n points distributed uniformly at random within the unit square in a plane. Each point connects to the k-closest points. Let us denote the resulting graph as $\Gamma(n, k)$. Show that there exists α such that if $k \geq \alpha \log n$, then $\Gamma(n, k)$ is connected with probability at least $1 - 1/n$.

10. A random variable X taking on positive integer values is said to be a *Poisson* random variable with parameter $\lambda > 0$ if

$$\mathbf{Pr}[X = i] = \exp(-\lambda)\frac{\lambda^i}{i!}$$

for $i = 0, 1, 2, \ldots$

(a) Show that $\mathbf{E}[X] = \lambda$.

(b) Let $X_1, \ldots, X_n$ be identically distributed Poisson random variable with parameter λ, and let $X = \sum_{i=1}^{n} X_i$. Show that

i. Show that X is a Poisson random variable with parameter $n\lambda$.

ii. $\mathbf{Pr}[X > (1+\epsilon)n\lambda] \leq \exp(-\epsilon^2 n\lambda)$.