CS 174: Combinatorics and Discrete Probability Fall 2012

Homework 6 Solutions

Problem 1. *(Exercise 5.16 from MU – 6 points)* Let G be a random graph generated using the $G_{n,p}$ model. Write the answers for these questions in the limit as $n \to \infty$, so you should ignore constant and lower order terms.

(a) A *clique* of k vertices in a graph is a subset of k vertices such that all $\binom{k}{2}$ edges between these vertices lie in the graph. For what value of p, as a function of n, is the expected number of cliques of five vertices in G equal to 1?

Solution: Fix any subset S of 5 vertices. The probability S is a clique in G is p^{10} since S contains $\binom{5}{2} = 10$ edges. There are $\binom{n}{5}$ possible choices for S, so by linearity of expectation, the expected number of 5-cliques in G is $\binom{n}{5}p^{10}$. Solving for $\binom{n}{5}p^{10} = 1$ yields $p^{10} = \Theta(1/n^5)$, so $p = \Theta(1/\sqrt{n})$.

(b) A $K_{3,3}$ is a complete bipartite graph with three vertices on each side. In other words, it is a graph with six vertices and nine edges; six distinct vertices are arranged in two groups of three, and the nine edges connect each of the nine pairs over vertices with one vertex in each group. For what value of p, as a function of n, is the expected number of $K_{3,3}$ subgraphs of G equal to 1?

Solution: Fix any unordered pair of disjoint subsets of 3 vertices; there are $\frac{1}{2}\binom{n}{3}\binom{n-3}{3}$ choices. The probability that this induces a $K_{3,3}$ is $p^9(1-p)^6$. Hence, the expected number of $K_{3,3}$ is $\frac{1}{2}\binom{n}{3}\binom{n-3}{3}p^9(1-p)^6$. Setting this quantity to 1 yields $p^9(1-p)^6 = \Theta(1/n^6)$, which solves to $p = \Theta(1/n^{2/3})$.

This question is a bit ambiguous – if by a $K_{3,3}$ subgraph, you mean an induced subgraph on 6 vertices, then the above calculations are correct. Otherwise, you simply want the 9 edes to be present. Thus, the expected number is $\frac{1}{2}\binom{n}{3}\binom{n-3}{3}p^9$. This also gives, $p = \Theta(1/n^{2/3})$.

(c) For what value of p, as a function of n, is the expected number of Hamiltonian cycles in the graph equal to 1?

Solution: Fix an ordered sequence of n vertices; there are $\frac{1}{2}(n-1)!$ choices (because we may always assume that the first vertex is fixed, and each order is equivalent to its reversal). The probability this induces a Hamiltonian cycle is p^n. Hence, the expected number of Hamiltonian cycles is $\frac{1}{2}(n-1)! \cdot p^n$. Setting this quantity to 1 yields $p = \Theta(1/n)$, using the bounds $(n/e)^n \leq n! \leq n^n$.

Problem 2. *(Exercise 5.21 from MU – 8 points)* In hashing with *open addressing*, the hash table is implemented as an array and there are no linked lists or chaining. Each entry in the array either contains one hashed item or is empty. The hash function defines, for each key k, a *probe sequence* $h(k,0), h(k,1), \ldots$ of table locations. To insert the key k, we first examine the sequence of table locations in the order defined by the key's probe sequence until we find an empty location; then we insert the item at that position. When searching for an item in the hash table, we examine the sequence of table locations in the order defined by the key's probe sequence until either the item is found or we have found an empty location in the sequence. If an empty location is found, this means the item is not present in the table.

An open-address hash table with $2n$ entries is used to store n items. Assume that the table location $h(k, j)$ is uniform over the $2n$ possible table locations and that all $h(k, j)$ are independent.

(a) Show that, under these conditions, the probability of an insertion requiring more than m probes is at most 2^{-m}.

Solution: Consider the i^{th} iteration. Since $i - 1$ entries are already filled, the success probability is $\frac{2n-(i-1)}{2n} > \frac{1}{2}$. Thus, at each step with probability at least $\frac{1}{2}$, we will find an empty entry.

Since $h(k, j)$ are iid, then:

$$\Pr(\text{Requires more than } m \text{ probes}) \leq \prod_{i=1}^{m} \frac{1}{2} = 2^{-m}$$

(b) Show that, for $i = 1, 2, \ldots, n$, the probability that the i^{th} insertion requires more than $2\log(n)$ probes is at most $1/n^2$.

Solution: Using part (a), we get that $\Pr(X_i > m) \leq 2^{-m}$, which implies that

$$\Pr(X_i > 2\log n) \leq 2^{-2\log n} = n^{-2}$$

(c) Now, let X_i denote the number of probes required by the i^{th} insertion. You showed above that $\Pr(X_i \geq 2\log(n)) \leq 1/n^2$. Let the random variable $X = \max_{1\leq i\leq n} X_i$ denote the maximum number of probes required by any of the n insertions. Show that $\Pr(X > 2\log(n)) \leq 1/n$.

Solution: Using the union bound, we can show that $\Pr(X > 2\log n) \leq \frac{1}{n}$.

$$\begin{aligned}
\Pr(X > 2\log n) &= \Pr\left(\bigcup_{i=1}^{n}(X_i > 2\log n)\right) \\
&\leq \sum_{i=1}^{n} \Pr(X_i > 2\log n) \\
&\leq \sum_{i=1}^{n} \frac{1}{n^2} = \frac{1}{n}
\end{aligned}$$

(d) Use the above to conclude that the expected length of the longest probe sequence, $\mathbb{E}[X] = O(\log(n))$.

Solution:

$$\begin{aligned}
\mathbb{E}[X] &= \mathbb{E}[X|X \leq 2\log n]\Pr(X \leq 2\log n) + \mathbb{E}[X|X > 2\log n]\Pr(X > 2\log n) \\
&\leq (2\log n)\Pr(X \leq 2\log n) + \mathbb{E}[X|X > 2\log n]\frac{1}{n} \\
&\leq (2\log n) + \mathbb{E}[X|X > 2\log n]\frac{1}{n} \quad \text{by part (c)}
\end{aligned}$$

But

$$\begin{aligned}
\mathbb{E}[X|X > 2\log n] &\leq \mathbb{E}\left[\sum_{i=1}^{n} X_i \middle| X > 2\log n\right] \\
&\leq \sum_{i=1}^{n} \mathbb{E}[X_i|X_i > 2\log n] \\
&\leq n(2\log n + 2) \quad \text{by the memoryless property of the geometric distribution}
\end{aligned}$$

So

$$\mathbb{E}[X] \leq \Theta(\log n)$$

Problem 3 *(Exercise 6.2 from MU – 8 points)*

(a) Prove that, for every integer n, there exists a colouring of the edges of the complete graph, K_n, using two colours – say red and black – so that the total number of monochromatic K_4 is at most $\binom{n}{4}/2^5$.

Solution: Construct a uniform probabiliy space over all possible 2-colorings by independently assigning each edge to color 1 w.p. 1/2 and color 2 w.p. 1/2. Let X be the number of monochromatic K_4s for this coloring.

A particular K_4 gets the same color w.p. $2 \cdot 2^{-\binom{4}{2}} = 2^{-5}$, and linearity of expectation gives $\mathbb{E}[X] = \binom{n}{4}2^{-5}$. So by the expectation argument, there exists a 2-coloring with at most $\binom{n}{4}2^{-5}$ monochromatic K_4s.

(b) Give a Las Vegas algorithm for finding such a colouring (one with at most $\binom{n}{4}/2^5$ monochromatic K_4) that runs in expected polynomial time in n. Recall that a Las Vegas algorithm always returns a correct output, but its worst case running time may be unbounded.

Solution: Each trial of the randomized algorithm independently picks a random 2-coloring as in part (a) and checks if the coloring satisfies the desired property. The trials are repeated until one succeeds. (Note, that counting the number of monochromatic K_4s can be done in $O(n^4)$ time.)

Each trial fails w.p. $\Pr[X \geq \binom{n}{4}2^{-5} + 1] \leq \frac{\binom{n}{4}2^{-5}}{\binom{n}{4}2^{-5}+1}$ using Markov' inequality. The number of trials thus follows a geometric distribution with parameter $\frac{1}{\binom{n}{4}2^{-5}+1}$, and has expectation $\binom{n}{4}2^{-5} + 1 = O(n^4)$.

(c) Show how to construct such a colouring deterministically in polynomial time using the method of conditional expectations.

Solution: We use the same notation as the initial note. To answer this part, it suffices to specify the z_is and f, whether to use max or min operator, and how to compute the conditional expectations. For some numbering of the edges, let z_i be the color (color1 or color2) for the i^{th} edge (so $k = \binom{n}{2}$), and $f(z)$ be the number of monochromatic K_4s for this coloring. We use the min operator in each step to find a coloring a with $f(a) \leq \binom{n}{4}2^{-5}$.

Suppose only the first $i-1$ edges have been colored, then a particular K_4 is monochromatic w.p. (a) 2^{-5} if it has no colored edges, (b) 2^{-6+x} if it has $x > 0$ edges of the same color and $(6-x)$ uncolored edges, and (c) 0 if it has two or more edges of different color. We can use this face and the linearity of expectation to find the deesired conditional expectations. It is easy to verify that each of these $k = O(n^2)$ steps take polynomial time.

Problem 4 *(Exercise 6.3 – 8 points)* Given an n-vertex undirected graph $G = (V, E)$, consider the following method of generating an independent set. Given a permutation σ of the vertices, define a subset $S(\sigma)$ of the vertices as follows: for each vertex i, $i \in S(\sigma)$ if and only if no neighbour j of i precedes i in the permutation σ.

(a) Show that each $S(\sigma)$ is an independent set in G.

Solution: Suppose not, then there exist nodes $i, j \in S(\sigma)$ that are connected by an edge. However, if $\sigma(i) < \sigma(j)$, then $j \notin S(\sigma)$; or vice versa. This is a contradiction. Hence $S(\sigma)$ must be an independent set.

(b) Suggest a natural randomized algorithm to produce σ for which you can show that the expected cardinality of $S(\sigma)$ is

$$\sum_{i=1}^{n} \frac{1}{d_i + 1},$$

where d_i is the degree of vertex i.

Solution: Choose the permutation σ uniformly at random. $i \in S(\sigma)$ iff node i comes before all its neighbors in the permutation. This happens with probability $\frac{1}{d_i+1}$. The cardinality of $S(\sigma)$ is the sum of the indicator variables for each node being in $S(\sigma)$. Thus

$$\mathbb{E}[|S(\sigma)|] = \sum_{i=1}^{n} \frac{1}{d_i + 1}$$

(c) Prove that G has an independent set of size at least $\sum_{i=1}^{n} 1/(d_i + 1)$.

Solution: This follows by the expectation argument of the probabilistic method.