CS174 Midterm 2 Solutions

1. Consider a random graph G in the $G_{n,p}$ model. Suppose that we choose one of two colors for each vertex of G, uniformly and independently. (In this question, the expectations are over both the choice of G and the colors.)

   (a) What is the expected number of monochromatic cycles of length k in G, as a function of k, n, and p?

   The number of potential cycles of length k is

   $$\binom{n}{k}\frac{(k-1)!}{2}.$$

   For each of these, the probability that the cycle is included in G is p^k. And the probability that the cycle is monochromatic given that it is included in G is $2^{-(k-1)}$. Thus, the expected number of monochromatic cycles of length k in G is

   $$\binom{n}{k}(k-1)!\left(\frac{p}{2}\right)^k.$$

   (b) We say that a vertex v is color-isolated if all of the neighbors of v are colored differently from v. What is the expected number of color-isolated vertices in G, as a function of n and p?

   Consider a vertex v. We'll determine the probability that v is color-isolated. For any other vertex u, the probability that u is either not connected or has a different color from v is $(1-p)+p/2$. Thus, the probability that v is color-isolated is

   $$\left(1-\frac{p}{2}\right)^{n-1}.$$

   Hence, the expected number of color-isolated vertices is

   $$n\left(1-\frac{p}{2}\right)^{n-1}.$$

2. Consider the following scheme for broadcasting data to a set of n hosts on an unreliable network: The server sends a packet to the first host. If it does not receive an acknowledgement in one second, it sends another packet, and so on. Once it receives an acknowledgement from the first host, it goes on to the second host. Assume that, for each packet sent to host i, the probability that an acknowledgement is received by the server in one second is p, and that all of these 'acknowledgement-received' events are independent. Let X_i be the number of packets sent to host i. Notice that X_i has a geometric distribution with parameter p. Define X as the total number of packets sent to n hosts, $X = \sum_{i=1}^{n} X_i$.

   (a) What is $\mathbb{E}[X]$?

$$\mathbb{E}[X] = \sum_{i=1}^{n} \mathbb{E}[X_i] = \frac{n}{p}.$$

   (b) Give an upper bound on the probability that X exceeds twice its expectation. Your upper bound should decrease exponentially in n.
   (Hint: relate this probability to a probability involving the number of successes in a sequence of independent Poisson trials.)

   If $X \geq 2\mathbb{E}[X]$, then the number of successes in $2\mathbb{E}[X]$ trials with probability of success p is no more than n. Thus, if Z is a Bernoulli$(2\mathbb{E}[X], p)$ random variable, we have $\mathbb{E}[Z] = 2p\mathbb{E}[X] = 2n$, and so

$$\begin{aligned}
\Pr(X \geq 2\mathbb{E}[X]) &\leq \Pr\left(Z \leq \frac{\mathbb{E}[Z]}{2}\right) \\
&\leq \left(\frac{e^{-1/2}}{(1/2)^{1/2}}\right)^{2n} \\
&= \left(\frac{2}{e}\right)^{n}.
\end{aligned}$$

3. We throw m balls in n bins, uniformly and independently. Let X_i be the indicator for bin i containing no more than $m/(2n)$ balls, and let X be the number of bins containing no more than $m/(2n)$ balls. Show that

$$\Pr(X_1 = 1) \leq 2\left(\frac{2}{e}\right)^{m/(2n)},$$

and hence

$$\mathbb{E}[X] \leq 2n\left(\frac{2}{e}\right)^{m/(2n)}.$$

(You may assume that $\Pr(X_1 = 1)$ decreases monotonically in m.)

Under the Poisson approximation, with $\mu = m/n$, we have

$$\begin{aligned}\Pr(X_1 = 1) &\leq \frac{e^{-\mu}(e\mu)^{\mu/2}}{(\mu/2)^{\mu/2}} \\ &= e^{-\mu}(2e)^{\mu/2} \\ &= \left(\frac{2}{e}\right)^{\mu/2}.\end{aligned}$$

Thus, under the true distribution,

$$\Pr(X_1 = 1) \leq 2\left(\frac{2}{e}\right)^{m/(2n)}.$$

The second claim follows from linearity of expectation.

4. A database is stored in a set of files. There is some redundancy: if S is the set of files, the database can be reconstructed completely from a subset $S' \subseteq S$, provided that S' intersects each of a set of *essential* subsets of S. More precisely, there is a set $\mathbb{B}$ of subsets of S, such that if, for all $B \in \mathbb{B}$, $S' \cap B \neq \emptyset$, then S can be reconstructed completely from S'.

   We wish to partition S into two disjoint sets S_1, S_2 so that $S_1 \cup S_2 = S$. These two sets of files will be stored separately. The partition should be chosen so that, if one of the sets is destroyed, the full set S can be reconstructed completely.

   Suppose that there are constants a, b such that each B in $\mathbb{B}$ satisfies the following conditions:

   - $|B| = a$, and
   - $|\{A \in \mathbb{B} : B \cap A \neq \emptyset\}| \leq b$.

   Show that, provided a is sufficiently large compared to b, a suitable partition exists. (Try to give the weakest condition that you can on the relationship between a and b.)

   Use the Lovasz local lemma. Assign elements to S_1 or S_2 uniformly at random. The bad events are

$$E_A = \{S_1 \cap A = \emptyset \text{ or } S_2 \cap A = \emptyset\}.$$

   First, $\Pr(E_A) = 2^{-|A|+1} = 2^{-a+1}$. Second, the dependency graph has degree no more than b. By the Lovasz local lemma, provided $4b2^{-a+1} \leq 1$, which is equivalent to

$$a \geq 3 + \log_2(b),$$

   the probability of a suitable assignment is positive.