

Homework 6 Solutions

Note: These solutions are not necessarily model answers. Bear in mind that there may be more than one correct solution. The maximum total number of points available is 40.

1. (a) Clearly, we may write the r.v. X for the number of balls in bin 1 as the sum of n independent 0-1 r.v.'s (one for each ball). Then, $E[X] = n \cdot \frac{1}{n} = 1$. Applying the Chernoff bound with $\delta = k - 1$ and $\mu = 1$, we get $p \leq \exp(-\frac{(k-1)^2}{k+1}) = 2^{-\Omega(k)}$. 2pts
- (b) Fix a subset of k balls. The probability that all k balls fall into bin 1 is $(\frac{1}{n})^k$. Taking a union bound over all subsets of k balls, we may deduce that the probability that there exists a subset of k balls all of which fall into bin 1 is at most $\binom{n}{k}(\frac{1}{n})^k$. Hence, $p \leq \binom{n}{k}(\frac{1}{n})^k \leq \frac{n^k}{k!}(\frac{1}{n})^k = \frac{1}{k!} \leq (e/k)^k$. 2pts
- (c) Under the Poisson approximation, the probability that bin 1 gets exactly j balls is $\Pr[Z = j]$ where Z is a Poisson r.v. with parameter 1. Hence, p may be approximated by $\sum_{j=k}^{\infty} \Pr[Z = j] = \frac{1}{e} \sum_{j=k}^{\infty} \frac{1}{j!}$. The event “first bin gets at least k balls” is monotonically increasing, so applying the inequality from Q2 of HW5 yields $p \leq 2 \cdot \frac{1}{e} \sum_{j=k}^{\infty} \frac{1}{j!}$. 2pts
- (d) The values are:
- For $p = n^{-\log \log n} = 2^{-\log n \log \log n}$: (a) yields $k = \Omega(\log n \log \log n)$ and (b) yields $k = \Omega(\log n)$. 2pts
 - For $p = \frac{1}{n^{100}}$: (a) yields $k = \Omega(\log n)$ and (b) yields $k = \Omega(\frac{\log n}{\log \log n})$. 2pts
 - For $p = 2^{-0.001n}$: (a) yields $k = \Omega(n)$ and (b) yields $k = \Omega(\frac{n}{\log n})$. 2pts
- Note that the values for (b) are systematically smaller than for (a).
2. (a) Fix any subset S of 5 vertices. The probability S is a clique in G is p^{10} since S contains 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 $p = \Theta(1/\sqrt{n})$. 4pts
- (b) 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 $p = \Theta(1/n^{2/3})$. 4pts
- (c) 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 $(n/e)^n \leq n! \leq n^n$. 4pts
3. (a) The expected number of edges is $\frac{1}{2}pn(n-1) = 8(n-1)$. 1pt
- (b) Clearly, we may write the r.v. X for the number of edges in G as the sum of $\frac{1}{2}n(n-1)$ independent 0-1 r.v.'s (one for each edge). Applying the Chernoff bound with $\mu = 8(n-1)$ and $\delta = \frac{1}{4}$, we obtain $\Pr[X > 10(n-1)] < e^{-(1/4)^2 8(n-1)/3} = 2^{-\Omega(n)}$. Hence, $\Pr[X \leq 10(n-1)] \geq 1 - 2^{-\Omega(n)}$ and the required inequality follows. 2pts
- (c) Let r be the number of red vertices in the coloring. Then the number of pairs of vertices of the same color is $\binom{r}{2} + \binom{n-r}{2}$. Let Y be the r.v. for the number of violated edges in G . Then, $E[Y] = p(\binom{r}{2} + \binom{n-r}{2}) \geq 2p\binom{n/2}{2} = 4(n-2)$ (assuming that n is even to avoid messy rounding details). Again, Y can be written as a sum of independent 0-1 r.v.'s. Applying the Chernoff bound with $\mu = 4(n-2)$ and $\delta = \frac{3}{4}$, we obtain $\Pr[Y \leq n-2] \leq e^{-(3/4)^2 \cdot 4(n-2) \cdot 1/2} = e^{-9(n-2)/8}$. The required result now follows. 3pts

(d) There are 2^n different ways to color the graph. By part (c) and a union bound over colorings, we have 2pts

$$\Pr_G[\exists \text{ an assignment of colors to } G \text{ with } \leq n - 2 \text{ violated edges}] \leq 2^n \cdot e^{-\frac{9}{8}(n-2)} \leq \frac{1}{4}.$$

When an assignment has $n - 1$ violated edges, then removing any $n - 2$ edges still leaves at least one violated edge. Since with probability $\geq \frac{3}{4}$ every assignment has $\geq n - 1$ violated edges, we see that if we remove any $n - 2$ edges of G there is still no valid 2-coloring.

(e) Fix a sequence of k distinct vertices v_1, \dots, v_k in G . The probability (over G) that (v_1, \dots, v_k) is a cycle in G is p^k . There are at most n^k sequences of k distinct vertices, and for each such sequence, we may assign an indicator variable that tells us whether it corresponds to a k -cycle in G . By linearity of expectation, the expected number of k -cycles in G is at most $n^k \cdot p^k = 16^k$. 3pts

The expected number of cycles of length at most $\frac{1}{8} \log n$ is bounded by $\sum_{k=3}^{\frac{1}{8} \log n} 16^k \leq 16^{\frac{1}{8} \log n + 1} = 16\sqrt{n}$.

(f) By Markov's inequality, with probability at least $3/4$ (over G), G has at most $64\sqrt{n}$ cycles of length at most $\frac{1}{8} \log n$. We may delete an edge from each of these cycles, thereby removing at most $64\sqrt{n}$ edges. 2pts

(g) Taking a union bound, we know that for all sufficiently large n , with probability at least $1/2 - 2^{-\Omega(n)} > 0$ over G from $\mathcal{G}_{n,16/n}$, G satisfies the properties in parts (b), (d) and (f) simultaneously. This means that by deleting at most $64\sqrt{n}$ edges from G , we obtain a graph G' with the following properties: 3pts

- the induced subgraph on any subset of $\frac{1}{8} \log n$ vertices of G' is 2-colorable (by part (f));
- G' has at most $10(n - 1)$ edges (by part (b));
- G' is not 2-colorable even if we delete any $n - 2 - 64\sqrt{n}$ edges (by part (d); note that we may have already deleted $64\sqrt{n}$ edges from G itself in part (f)).

Finally, note that for sufficiently large n we have (very crudely) $n - 2 - 64\sqrt{n} \geq \frac{1}{2}(n - 1)$, so the number of edges we can delete is at least a 0.05 fraction of the edges of G (and thus also of G'). Thus G' is the required graph G_n , and we have verified the existence of G' by the probabilistic method.