Homework 3 Solutions

Note: These solutions are not necessarily model answers. Rather, they are designed to be tutorial in nature, and sometimes contain a little more explanation than an ideal solution. Also, bear in mind that there may be more than one correct solution. The maximum total number of points available is 32.

1. Define the r.v. $X = \frac{X_1+X_2+\cdots+X_n}{n}$. Then by linearity of expectation we have $\mathrm{E}[X] = \frac{1}{n}\sum_i \mathrm{E}[X_i] = \mu$. *4pts*
Also, since the X_i's are independent, we have $\mathrm{Var}[X] = \frac{1}{n^2}\sum_{i=1}^n \mathrm{Var}[X_i] = \frac{\sigma^2}{n}$. Finally, by Chebyshev's inequality, for any $\varepsilon > 0$,

$$\Pr[|X-\mu| \geq \varepsilon] \leq \frac{\sigma^2}{n\varepsilon^2} \to 0$$

as $n \to \infty$. This completes the proof.

Some students forgot to show that $\mathrm{E}[X] = \mu$.

2. (a) For each $e \in E$, let X_e be the indicator random variable that assumes the value 1 if e is in the cut, and 0 otherwise. Then $X = \sum_{e\in E} X_e$. In addition, $\mathrm{E}[X_e] = \Pr[\text{endpoints of } e \text{ have different colors}] = \frac{1}{2}$. By linearity of expectation, we have $\mathrm{E}[X] = \frac{1}{2}|E| \geq \frac{\mathrm{OPT}}{2}$, since clearly $\mathrm{OPT} \leq |E|$. *2pts*

Note: Throughout this problem, it's important to remember the following points: (i) $\mathrm{OPT} \leq |E|$ *(because clearly no cut can contain more edges than the total number of edges in the graph!); and (ii) we do not know the value of* OPT*, so we cannot use it in our algorithm (though of course we do know the value of* $|E|$*). Several students got confused about this, and especially point (ii), by assuming that the algorithm knows the value of* OPT*. Note that it's actually NP-hard to compute* OPT*, so it's really not OK to assume this! Points were deducted for the first offense of this kind (but not for subsequent offenses).*

(b) Let $Y = |E| - X$, which is a non-negative random variable. Note that $\mathrm{E}[Y] = |E| - \mathrm{E}[X] = \frac{1}{2}|E|$. *3pts*
Now, $\Pr[X < 0.49|E|] = \Pr[Y > 0.51|E|]$, so we can apply Markov's inequality to Y to see that this probability is at most $\mathrm{E}[Y]/0.51|E| = \frac{1}{2}|E|/0.51|E| = \frac{50}{51}$. Hence, $\Pr[X \geq 0.49|E|] \geq \frac{1}{51}$. Again, since $\mathrm{OPT} \leq |E|$, we get that $p = \Pr[X \geq 0.49\mathrm{OPT}] \geq \frac{1}{51}$.

Some variations on the above argument are also valid. For example, we could instead use Markov's inequality to bound $\Pr[Y \geq |E| - 0.49\mathrm{OPT}]$, which is at most $\Pr[Y \geq 0.51\mathrm{OPT}]$. (However, note that we *cannot* use Markov to bound $\Pr[Y \geq 0.51\mathrm{OPT}]$ because $0.51\mathrm{OPT}$ may be smaller than $\mathrm{E}[Y] = |E|/2$). Another variation is to define $Z = \mathrm{OPT} - X$, and note that Z is a non-negative r.v. and $\mathrm{E}[Z] \leq 0.5\mathrm{OPT}$. We may then apply Markov's inequality to bound $\Pr[Z \geq 0.51\mathrm{OPT}]$

A rather different argument does not use Markov's inequality directly, but instead uses the same idea as in the *proof* of Markov's inequality. It goes as follows. Note that $\mathrm{E}[X] \leq p\cdot\mathrm{OPT} + (1-p)\cdot 0.49\mathrm{OPT}$. (This follows as in the proof of Markov's inequality; the first term bounds the contribution to $\mathrm{E}[X]$ from all values of X larger than $0.49\mathrm{OPT}$, and the second term bounds the contribution from the values less than or equal to $0.49\mathrm{OPT}$.) Since $\mathrm{E}[X] \geq \mathrm{OPT}/2$, we can cancel OPT through the inequality to get $p \geq 1/51$.

Some students incorrectly used OPT *in place of* $|E|$ *in Markov's inequality.*

(c) We expand the square and use linearity of expectation to write $\mathrm{E}[X^2] = \sum_e \mathrm{E}[X^2] + \sum_{e \neq e'} \mathrm{E}[X_e X_{e'}]$. *3pts* For the diagonal terms, we have $\mathrm{E}[X_e^2] = \mathrm{E}[X_e] = \frac{1}{2}$. To handle the cross terms $\mathrm{E}[X_e X_{e'}]$ for distinct edges e, e', observe that $\mathrm{E}[X_e X_{e'}] = \Pr[X_e = X_{e'} = 1] = \frac{1}{4}$. This is true both for the case where e, e' share an endpoint (fix the common endpoint, then the other two endpoints must both have the opposite color, which occurs with probability $\frac{1}{4}$), and for the case where e, e' do not share an endpoint (here, $X_e, X_{e'}$ are independent). It follows that $\mathrm{E}[X^2] = \sum_e \mathrm{E}[X_e] + \sum_{e \neq e'} \mathrm{E}[X_e X_{e'}] = \frac{1}{2}|E| + \frac{1}{4}|E|(|E|-1) = \frac{1}{4}|E|^2 + \frac{1}{4}|E|$. Hence, using part (a), $\mathrm{Var}[X] = \mathrm{E}[X^2] - \mathrm{E}[X]^2 = \frac{1}{4}|E|$.

Equivalently, we can generalize the formula $\mathrm{Var}[X+Y] = \mathrm{Var}[X] + \mathrm{Var}[Y] + \mathrm{Cov}(X,Y)$ from class to write

$$\mathrm{Var}[X] = \mathrm{Var}[\sum_e X_e] = \sum_e \mathrm{Var}[X_e] + \sum_{e \neq e'} \mathrm{Cov}(X_e, X_{e'}).$$

Now observe that every pair of r.v.'s $X_e, X_{e'}$ are independent as argued above, so all the covariances $\mathrm{Cov}(X_e, X_{e'})$ are zero. And since X_e is a 0-1 r.v. with expectation $\frac{1}{2}$, $\mathrm{Var}[X_e] = \frac{1}{4}$. Thus we get $\mathrm{Var}[X] = \sum_e \mathrm{Var}[X_e] = \frac{1}{4}|E|$, as before.

Note: Some people failed to explain why $X_e, X_{e'}$ are independent. Also, note that although each **pair** *of r.v.'s $X_e, X_{e'}$ is independent, it is* **not** *true that the X_e are mutually independent. To see this, consider three edges e, e', e'' that form a triangle. If $X_e = X_{e'} = 1$ then it must be the case that $X_{e''} = 0$. We say that the X_e are* **pairwise independent** *but not mutually independent. Since variance is only a second-order quantity, pairwise independence is enough to conclude that variances sum. We'll talk more about pairwise independence later in the class.*

(d) By Chebyshev's inequality, *2pts*

$$\begin{aligned} \Pr[X < 0.49|E|] &\le \Pr\left[|X - \tfrac{1}{2}|E|| > 0.01|E|\right] \\ &\le \tfrac{1}{4}|E|/(0.01|E|)^2 = O(1/|E|) \end{aligned}$$

Hence, $p \ge 1 - O(1/|E|)$ as required.

Again, some students incorrectly used OPT *in place of $|E|$ in Chebyshev's inequality.*

(e) We keep running the above algorithm until the size of the cut is at least $0.49|E|$. By the bound in *2pts* part (b) and the expectation of a geometric r.v., the expected number of repetitions we need is at most 51, so the expected running time is still linear. Correctness follows from the fact that $0.49|E| \ge 0.49\mathrm{OPT}$. We stress that the termination condition must compare the size of the cut with $0.49|E|$ and **not** with $0.49\mathrm{OPT}$ because we do not know the value of OPT!

The bound in part (d) yields a better upper bound on the expected running time; namely, it tells us that the expected number of repetitions is in fact only $1/(1 - O(|E|^{-1})) = 1 + O(|E|^{-1})$, which approaches 1 as $|E| \to \infty$ (i.e., for large graphs).

Some students just said that expected running time is $O(1)$, or just 1, ignoring the dependency on $|E|$. While it is indeed $O(1)$ (which just means it's bounded above by some constant), you can state the stronger fact that it's $1 + O(1/|E|)$, which tends to 1 as $|E| \to \infty$. And it's **not** *true that it's 1, since it's in fact slightly larger than 1 (by an amount $O(1/|E|)$).*

3. (a) As suggested in the Hint, write $X = \sum_{ij} X_{ij}$, where for each pair (i,j) with $i < j$ (which henceforth *2pts* we write ij for simplicity), X_{ij} is the indicator r.v. of the event that ij is an inversion. Now clearly

$\mathrm{E}[X_{ij}] = \frac{1}{2}$ for all ij by symmetry (since the number of permutations π with $\pi(i) < \pi(j)$ is equal to the number with $\pi(i) > \pi(j)$). Hence by linearity of expectation we have

$$\mathrm{E}[X] = \sum_{ij} \mathrm{E}[X_{ij}] = \binom{n}{2}\frac{1}{2} = \frac{n(n-1)}{4}.$$

(b) Generalizing the formula $\mathrm{Var}[X+Y] = \mathrm{Var}[X] + \mathrm{Var}[Y] + 2\mathrm{Cov}(X,Y)$ from class to a sum over any finite number of r.v.'s, we can express $\mathrm{Var}[X]$ as *4pts*

$$\mathrm{Var}[X] = \sum_{ij} \mathrm{Var}[X_{ij}] + 2\sum_{ij \prec kl} \mathrm{Cov}(X_{ij}, X_{kl}). \tag{1}$$

Note that in the second sum here we need to enumerate each pair of distinct pairs ij, kl exactly once. To ensure this, we define a natural order $\prec$ on such pairs by $ij \prec kl$ iff either (i) $i < k$, or (ii) $i = k$ and $j < l$.

Since each X_{ij} is a 0-1 r.v. with parameter $p = \frac{1}{2}$, we have that $\mathrm{Var}[X_{ij}] = p(1-p) = \frac{1}{4}$, so the first sum in (1) is $\binom{n}{2}\frac{1}{4} = \frac{n(n-1)}{8}$.

Now we turn to the covariances in (1). Observe that, for pairs ij and kl such that i,j,k,l are all distinct, the r.v.'s X_{ij} and X_{kl} are independent (since the relative order of i,j doesn't affect the relative order of k,l). Hence most of the covariances in (1) are in fact zero. We only need to handle the cases where ij and kl overlap. Recall that we are assuming $ij \prec kl$.

Case 1: $i = k$. In this case we must have $j < l$. There are $\binom{n}{3}$ ways to choose i,j,k,l with $i = k$ and $j < l$: namely, pick three distinct values in $\{1,\ldots,n\}$, set $i = k$ to be the smallest one, l the largest, and j the one in the middle. For each such pair of r.v.'s X_{ij}, X_{kl}, we have

$$\mathrm{Cov}(X_{ij}, X_{kl}) = \mathrm{E}[X_{ij}X_{kl}] - \mathrm{E}[X_{ij}]\mathrm{E}[X_{kl}]. \tag{2}$$

The first term in (2) is just $\Pr[(\pi(i) > \pi(j)) \cap (\pi(i) > \pi(l))]$: since exactly two of the six possible relative orderings of $\pi(i), \pi(j), \pi(k)$ satisfy this condition, we see that this probability is $\frac{1}{3}$. And the second term in (2) is just $\frac{1}{2} \times \frac{1}{2} = \frac{1}{4}$. Hence $\mathrm{Cov}(X_{ij}, X_{kl}) = \frac{1}{3} - \frac{1}{4} = \frac{1}{12}$, and the total contribution of this case is $\frac{1}{12}\binom{n}{3}$.

Case 2: $j = k$. Again there are $\binom{n}{3}$ ways to choose i,j,k,l in this case: pick three distinct values, set $j = k$ to be the middle one, i the smallest, and l the largest. Writing the covariance again as in (2), the first term is $\Pr[(\pi(i) > \pi(j)) > \pi(l))]$, which is $\frac{1}{6}$ since only one of the six possible orderings works. Thus in this case $\mathrm{Cov}(X_{ij}, X_{kl}) = \frac{1}{6} - \frac{1}{4} = -\frac{1}{12}$, and the total contribution here is $-\frac{1}{12}\binom{n}{3}$. [Note that the covariances are negative in this case!]

Case 3: $j = l$. Again there are $\binom{n}{3}$ ways to choose i,j,k,l: pick three values, set $j = l$ to be the largest, i the smallest, and k the middle one. This time the first term in (2) is $\Pr[(\pi(i) > \pi(j)) \cap (\pi(k) > \pi(j))]$. This is symmetrical with Case 1, so we get an overall contribution of $\frac{1}{12}\binom{n}{3}$ to the sum of covariances.

Plugging all of this into (1), we get

$$\mathrm{Var}[X] = \frac{n(n-1)}{8} + 2\frac{1}{12}\binom{n}{3} - 2\frac{1}{12}\binom{n}{3} + 2\frac{1}{12}\binom{n}{3} = \frac{n(n-1)}{8} + \frac{1}{6}\binom{n}{3}.$$

Finally, we can simplify this to $\mathrm{Var}[X] = \frac{n^3}{36} + O(n^2)$, as required.

(c) By Chebyshev's inequality, and using the values of $\mathrm{E}[X]$ and $\mathrm{Var}[X]$ from parts (a) and (b), we have *2pts*

$$\Pr[X \geq (\frac{1}{4}+\varepsilon)n^2] \leq \Pr[|X-\mathrm{E}[X]| \geq \varepsilon n^2] \leq \frac{\mathrm{Var}[X]}{\varepsilon^2 n^4} = \frac{n^3+O(n^2)}{36\varepsilon^2 n^4},$$

which tends to 0 as $n \to \infty$ for any fixed $\varepsilon > 0$.

4. (a) We focus on the event $\mathcal{E}_d$; the event $\mathcal{E}_u$ follows by a symmetric argument. Writing X for the r.v. that denotes the number of elements of R that are less than or equal to the median m, we have that $\Pr[\mathcal{E}_d] = \Pr[|\{r \in R \mid r \leq m\}| < \frac{1}{2}n^\alpha - n^\beta]$. But the r.v. X is has distribution $\mathrm{Bin}(n^\alpha, 1/2)$ (actually the 1/2 may be slightly larger due to rounding, which only makes our bounds better), so $\mathrm{E}[X] = \frac{1}{2}n^\alpha$ and $\mathrm{Var}[X] = \frac{1}{4}n^\alpha$. Thus by Chebyshev we get *3pts*

$$\Pr[\mathcal{E}_d] \leq \Pr[|X-\mathrm{E}[X]| \geq n^\beta] \leq \frac{\mathrm{Var}[X]}{n^{2\beta}} = \frac{n^\alpha}{4n^{2\beta}} = O(n^{-2\beta+\alpha}),$$

as required.

(b) As in class, we introduce two new events: *3pts*

$$\begin{aligned} \mathcal{E}'_C &: \text{ at least } 2n^{1-\alpha+\beta} \text{ elements of } C \text{ are } < m; \\ \mathcal{E}''_C &: \text{ at least } 2n^{1-\alpha+\beta} \text{ elements of } C \text{ are } > m. \end{aligned}$$

Clearly if $\mathcal{E}_C$ happens then at least one of $\mathcal{E}'_C, \mathcal{E}''_C$ must happen. By symmetry it suffices to bound $\Pr[\mathcal{E}'_C]$, as the same bound will hold for $\Pr[\mathcal{E}''_C]$.

Now note that, by analogy with what we did in class, in order for $\mathcal{E}'_C$ to happen, the rank of d in (the sorted order of) S must be at most $\frac{1}{2}n - 2n^{1-\alpha+\beta}$, which implies that our random sample R must contain at least $\frac{1}{2}n^\alpha - n^\beta$ elements within the smallest $\frac{1}{2}n - 2n^{1-\alpha+\beta}$ elements of S. Let the r.v. Y denote the number of elements of R that fall within this portion of S. Then Y is distributed as $\mathrm{Bin}(n^\alpha, \frac{n/2-2n^{1-\alpha+\beta}}{n}) = \mathrm{Bin}(n^\alpha, \frac{1}{2} - \frac{2}{n^{\alpha-\beta}})$, and hence $\mathrm{E}[Y] = \frac{1}{2}n^\alpha - 2n^\beta$ and $\mathrm{Var}[Y] = n^\alpha(\frac{1}{2} - \frac{2}{n^{\alpha-\beta}})(\frac{1}{2} + \frac{2}{n^{\alpha-\beta}}) = n^\alpha(\frac{1}{4} - 4n^{2(\alpha-\beta)}) < \frac{1}{4}n^\alpha$. Finally, we use Chebyshev's inequality to deduce that

$$\Pr[\mathcal{E}'_C] \leq \Pr[|Y-\mathrm{E}[Y]| > n^\beta] \leq \frac{\mathrm{Var}[Y]}{n^{2\beta}} \leq \frac{n^\alpha}{4n^{2\beta}} = O(n^{-2\beta+\alpha}).$$

(c) From parts (a) and (b), we deduce by a union bound that the probability that any of the four "bad" events above occurs is at most $\Pr[\mathcal{E}_d] + \Pr[\mathcal{E}_u] + \Pr[\mathcal{E}'_C] + \Pr[\mathcal{E}''_C] = O(n^{-2\beta+\alpha})$, which tends to zero as $n \to \infty$ provided $\alpha < 2\beta$. As in class, the absence of any of these events ensures that the algorithm will not fail, and will output the median m of S. Moreover, the running time of the algorithm will be bounded by $O(n)$ plus the time to sort the sets R and C. By construction, $|R| = n^\beta = o(n)$ since $\beta < 1$, so R can be sorted in sublinear (in n) time. And because $\mathcal{E}_C$ doesn't hold, we also know that $|C| \leq 4n^{1-\alpha+\beta}$, which is $o(n)$ provided $1-\alpha+\beta < 1$, i.e., $\beta < \alpha$. Thus the algorithm works as claimed provided $\beta < \alpha < 2\beta$. *2pts*