10/10

Peilun Dai

1. (Random hats)

(a) Solution: Let X be the random variable that denote the number of pair of changes, and X_{ij} where $i < j$ be the indicator random variable that people i and j exchanged their hats. Then from linearity of expectation,

4/4

$$\mathbf{E}(X) = \mathbf{E}\Big(\sum_{i=1}^{n-1} \sum_{j=i+1}^{n} X_{ij}\Big) \tag{1}$$

$$= \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} \mathbf{E}(X_{ij}) \tag{2}$$

$$= \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} \frac{1}{n} \frac{1}{n-1} \tag{3}$$

$$= \frac{n(n-1)}{2} \frac{1}{n(n-1)} = \frac{1}{2} \tag{4}$$

(b) Solution: Since X_{ij} is a binary indicator Bernoulli random variable taking values from $\{0, 1\}$, then the random variable X_{ij}^2 is also a Bernoulli random variable taking values from $\{0, 1\}$, and $\mathbf{E}(X_{ij}^2) = Pr[X_{ij}^2 = 1] = Pr[X_{ij} = 1] = \frac{1}{n(n-1)}$.

6/6

$$Var[X^2] = \mathbf{E}[X^2] - \mathbf{E}[X]^2 \tag{5}$$

$$= \mathbf{E}\Big[\Big(\sum_{i<j} X_{ij}\Big)^2\Big] - (1/2)^2 \tag{6}$$

$$= \mathbf{E}\Big(\sum_{i<j} X_{ij}^2 + \sum_{i \neq m \text{ or } j \neq n} X_{ij} X_{mn}\Big) - 1/4 \tag{7}$$

$$= \sum_{i<j} \mathbf{E}(X_{ij}^2) + \sum_{i \neq m \text{ or } j \neq n} \mathbf{E}(X_{ij} X_{mn}) - 1/4 \tag{8}$$

$$= 1/2 + \sum_{i \neq j \neq m \neq n, i<j, m<n} \mathbf{E}(X_{ij} X_{mn}) - 1/4 \tag{9}$$

$$= 1/2 + \binom{n}{2}\binom{n-2}{2} Pr[X_{ij} = 1 \,\&\&\, X_{mn} = 1] - 1/4 \tag{10}$$

$$= 1/2 + \frac{n(n-1)}{2!} \frac{(n-2)(n-3)}{2!} \frac{1}{n(n-1)} \frac{1}{(n-2)(n-3)} - 1/4 \tag{11}$$

$$= 1/2 + 1/4 - 1/4 \tag{12}$$

$$= 1/2 \tag{13}$$

From (7) to (8) we have used linearity of expectation, and from (8) to (9) we have used the fact that if i, j and m, n are not four unique numbers, then $X_{ij} X_{mn}$ will be zero. Thus we have only kept the terms that are non-zero.

 Peilun Dai

$\frac{10}{10}$

2. (MaxCut)

(a) Let X_{ij} denote the Bernoulli random variable whether the edge $\{i,j\}$ is in the cut set S, where $\{i,j\} \in E$, $i < j$ and $m = |E|$ is the total number of edges in the graph. Then $X = \sum_{\{i,j\}\in E} X_{ij}$. For R.V. $\{i,j\}$, since it is a Bernoulli random variable, $\mathbf{E}(X_{ij}) = Pr[X_{ij} = 1] = 1/2$. Thus,

$$\begin{aligned}
\mathbf{E}[X] &= \mathbf{E}\Big[\sum_{\{i,j\}\in E} X_{ij}\Big] && (1)\\
&= \sum_{\{i,j\}\in E} \mathbf{E}[X_{ij}] && (2)\\
&= \sum_{\{i,j\}\in E} Pr[X_{ij} = 1] && (3)\\
&= \sum_{\{i,j\}\in E} \frac{1}{2} && (4)\\
&= \frac{1}{2}m && (5)
\end{aligned}$$

$\frac{1}{1}$

It is easy to show that $OPT \leq m$ since the size of max cut cannot exceed the total number of edges in the graph, thus $\mathbf{E}(X) = \frac{1}{2}m \geq \frac{OPT}{2}$.

(b)

Proof. Let R.V. Y denote the number of edges not in the cut set S, then $Y = m - X$.

$$\begin{aligned}
Pr[X \geq 0.49OPT] &= Pr[Y \leq m - 0.49OPT] && (6)\\
&= 1 - Pr[Y \geq m - 0.49OPT] && (7)\\
&\geq 1 - \frac{\mathbf{E}(Y)}{m - 0.49OPT} && (8)\\
&= 1 - \frac{m}{2(m - 0.49OPT)} && (9)\\
&\geq 1 - \frac{m}{2(0.51m)} && (10)\\
&= 1 - \frac{1}{2 * 0.51} = \frac{1}{51} && (11)
\end{aligned}$$

$\frac{3}{3}$

From (7) to (8), we have used Markov Inequality. From (8) to (9), we have used the fact that $\mathbf{E}(Y) = m - \mathbf{E}(\cancel{m}) = 0.5m$. And from (9) to (10), we used the obvious fact that $OPT \leq m$.

(c) Similar to (a), we write $X = \sum_{\{i,j\}\in E} X_{ij}$.

$$\begin{aligned}
Var[X] &= \mathbf{E}[X^2] - \mathbf{E}(X)^2 && (12)\\
&= \mathbf{E}\Big[\Big(\sum_{\{i,j\}\in E} X_{ij}\Big)^2\Big] - (0.5m)^2 && (13)\\
&= \mathbf{E}\Big[\sum_{\{i,j\}\in E} X_{ij}^2 + \sum_{\{i,j\}\in E,\{i,j\}\neq\{m,n\}} \sum_{\{m,n\}\in E} X_{ij}X_{mn}\Big] - (0.5m)^2 && (14)\\
&= \sum_{\{i,j\}\in E} \mathbf{E}(X_{ij}^2) + \sum_{\{i,j\}\in E,\{i,j\}\neq\{m,n\}} \sum_{\{m,n\}\in E} \mathbf{E}[X_{ij}X_{mn}] - 0.25m^2 && (15)\\
&= m/2 + \binom{m}{2}\frac{1}{4} - 0.25m^2 = m/4 && (16)
\end{aligned}$$

$\frac{4}{4}$

From (14) to (15), we used linearity of expectation, and from (15) to (16), we used the fact that X_{ij}^2 follows the same distribution as X_{ij}, thus $\mathbf{E}(X_{ij}^2) = \mathbf{E}(X_{ij})$ and that $\mathbf{E}(X_{ij}X_{mn}) = 1/4$ if the two edges are different edges and there are $\binom{m}{2}$ such pairs of edges.

(d)

Proof.

$$\begin{aligned}
p &= Pr[X \leq 0.49OPT] && (17)\\
&= 1 - Pr[Y \geq m - 0.49OPT] && (18)\\
&= 1 - Pr[Y - 0.5m \geq 0.5m - 0.49OPT] && (19)\\
&\geq 1 - Pr[|Y - \mathbf{E}(Y)| \geq 0.5m - 0.49OPT] && (20)\\
&\geq 1 - \frac{Var[Y]}{(0.5m - 0.49OPT)^2} && (21)\\
&\geq 1 - \frac{m}{4(0.01m)^2} && (22)\\
&= 1 - \frac{2500}{m} = 1 - 2500 \cdot \frac{1}{|E|} && (23)\\
p &= \Omega(1 - 2500 \cdot \frac{1}{|E|}) && (24)\\
&= 1 - O(1/|E|) && (25)
\end{aligned}$$

$\frac{2}{2}$

From (19) to (20), we used the fact that $0.5 - 0.49OPT > 0$. From (20) to (21), we used Chebyshev's inequality. From (21) to (22), we used $Var[Y] = Var[m - X] = Var[X]$ and $OPT \leq m$.

10/10

Peilun Dai

3. (Generalization of Randomized Median Algorithm)

(a) Solutions: The main idea is to replace 1/2 with k/n. So we need the following modifications

- line 3: Let d be the $\left(\lfloor \frac{k}{n} n^{3/4} - \sqrt{n} \rfloor\right)$th smallest element in the sorted set R.
- line 4: Let u be the $\left(\lceil \frac{k}{n} n^{3/4} + \sqrt{n} \rceil\right)$th smallest element in the sorted set R.
- line 6: If $\ell_d > k$ or $\ell_u > n - k$ then FAIL.
- line 8: Output the $(k - \ell_d + 1)$th element in the sorted order of C.

(b) Solutions: We list the running time in big-O notation for each of the time consuming steps as follows:

- Choosing set R from A assuming $O(1)$ time element access for A: $O(n^{3/4})$.
- Sorting the set R: $O(n^{3/4} \log n^{3/4}) \approx O(n)$.
- Partition set A based on the values of d and u: $O(n)$.
- Sorting set C if $|C| \leq 4n^{3/4}$: $O(n^{3/4} \log n) \approx O(n)$.

Thus the total running time of the modified algorithm is $O(n)$.

(c) Solutions:

$$\begin{array}{l} \mathcal{E}_1 : Y_1 = |\{r \in R | r \leq k\}| < \frac{k}{n} n^{3/4} - \sqrt{n} \\ \mathcal{E}_2 : Y_2 = |\{r \in R | r \geq k\}| < (1 - \frac{k}{n}) n^{3/4} - \sqrt{n} \\ \mathcal{E}_{3,1} : \text{at least } 2n^{3/4} \text{ elements of C are greater than the } k\text{th smallest element in } A \\ \mathcal{E}_{3,2} : \text{at least } 2n^{3/4} \text{ elements of C are smaller than the } k\text{th smallest element in } A \end{array} \tag{1}$$

(d) Solutions:

For $\mathcal{E}_1$, define a random variable X_i by

$$X_i = \begin{cases} 1 & \text{if the } i \text{ th sample is less than or equal to the } k\text{th smallest element.} \\ 0 & \text{otherwise.} \end{cases} \tag{2}$$

The X_i are independent. The probability that a sample is smaller than the kth smallest element is $Pr[X_i = 1] = k/n$. The event $\mathcal{E}_1$ is equivalent to

$$Y_1 = \sum_{i=1}^{n^{3/4}} X_i < \frac{k}{n} n^{3/4} - \sqrt{n} \tag{3}$$

Thus, $Pr[\mathcal{E}_1] = Pr[Y_1 < \frac{k}{n} n^{3/4} - \sqrt{n}]$.

For $\mathcal{E}_{3,1}$, let us bound the probability the first event occurs. If there are at least $2n^{3/4}$ elements of C above the kth smallest element, then the order of u in the sorted order of S was at least $k + 2n^{3/4}$ and thus the set R has at least $(1-k/n)n^{3/4}-\sqrt{n}$ samples among the $n-k-2n^{3/4}$ largest elements in A, the input array.

Let X be the number of samples among the $n-k-2n^{3/4}$ largest elements in A. Let $X=\sum_{i=1}^{n^{3/4}} X_i$.

$$X_i = \begin{cases} 1 & \text{if the } i \text{ th sample is among the } n-k-2n^{3/4} \text{ largest elements in } S, \\ 0 & \text{otherwise.} \end{cases} \tag{4}$$

Then $Pr[\mathcal{E}_{3,1}] = Pr\big[X \geq (1-k/n)n^{3/4} - \sqrt{n}\big]$