CMSC651 Assignment 2

Yancy Liao, Fan Yang, Bowen Zhi

February 2018

Problem 1

(a) $\Pr(X \geq \mu(1+\delta)) = \Pr\left(e^{tX} \geq e^{t\mu(1+\delta)}\right)$, for some $t \in \mathbb{R}$. Note that both sides are now strictly positive, thus by Markov:

$$\Pr(X \geq \mu(1+\delta)) \leq \frac{\mathrm{E}\left[e^{tX}\right]}{e^{t\mu(1+\delta)}} = e^{-t\mu(1+\delta)} \sum_k \left(e^{tk} e^{-\mu} \frac{\mu^k}{k!}\right) = e^{-\mu(t(1+\delta)+1)} \sum_k \left(\frac{(\mu e^t)^k}{k!}\right)$$
$$= e^{-\mu(t(1+\delta)+1)} e^{\mu e^t} = e^{\mu(e^t - t(1+\delta)-1)}.$$

Note that minimizing the above is the same as minimizing $g(t) = e^t - t(1+\delta) - 1$, with respect to t, and that $g'(t) = e^t - 1 - \delta = 0$ when $t = \ln(1+\delta)$. Moreover, $g''(t) = e^t > 0\ \forall t \in \mathbb{R}$, thus the bound is minimized for $t = \ln(1+\delta)$. Substituting this value of t into our bound, we have:

$$\Pr(X \geq \mu(1+\delta)) \leq e^{\mu(1+\delta-(1+\delta)\ln(1+\delta)-1)} = \left(\frac{e^\delta}{(1+\delta)^{(1+\delta)}}\right)^\mu,$$

which is exactly the same as the upper-tail Chernoff-Hoeffding bound derived in Lecture 15, Section 1.

(b) Y is a binomial distribution, thus $\Pr(Y=i) = \binom{n}{i}\left(\frac{\mu}{n}\right)^i\left(1-\frac{\mu}{n}\right)^{n-i}$. Thus we have:

$$\lim_{n\to\infty} \Pr(Y=i) = \lim_{x\to\infty} \frac{n!}{i!(n-i)!}\left(\frac{\mu}{n}\right)^i\left(1-\frac{\mu}{n}\right)^{n-i} = \lim_{n\to\infty} \frac{n(n-1)\ldots(n-i+1)}{i!}\left(\frac{\mu}{n}\right)^i\left(1-\frac{\mu}{n}\right)^{n-i}$$
$$= \frac{\mu^i}{i!}\lim_{n\to\infty} \frac{n(n-1)\ldots(n-i+1)}{n^i}\left(1-\frac{\mu}{n}\right)^{n-i} = \frac{\mu^i}{i!}\lim_{n\to\infty}\left(1-\frac{\mu}{n}\right)^{n-i},$$

where the middle fraction approaches 1 asymptotically since the numerator's expanded form consists of n^i plus lower-order powers of n. Now observe that $\left(1-\frac{\mu}{n}\right)^{n-i} = \exp\left((n-i)\ln\left(1-\frac{\mu}{n}\right)\right)$. Taking the Taylor expansion around 0 of $\ln(1-x)$, we thus have:

$$\lim_{n\to\infty} \Pr(Y=i) = \frac{\mu^i}{i!}\lim_{n\to\infty} \exp\left((n-i)\left(-\frac{\mu}{n} - \mathcal{O}(\frac{1}{n^2})\right)\right) = \frac{\mu^i}{i!}\lim_{n\to\infty} e^{-\mu(n-i)/n} = e^{-\mu}\frac{\mu^i}{i!}.$$

Problem 2

(a) $\Pr(Y_i) = 1 - \Pr(\neg Y_i) = 1 - (1-1/n)^u = 1 - (1-1/n)^{\frac{n}{2}\ln n}$. Since we are looking for the asymptotic behavior of this value, we can ignore the floor on u. Next, note that $\lim_{x\to\infty}(1-1/n)^n = e^{-1}$, thus:

$$\Pr(Y=i) = 1 - (1-1/n)^{\frac{n}{2}\ln n} = 1 - \Theta\left(e^{-\ln(n)/2}\right) = 1 - \Theta\left(1/\sqrt{n}\right).$$

(b) Note that this scenario is equivalent to that of the "balls and bins" problem with $m = u$. Moreover, note that $\Pr(Y_i) = \Pr(X_i \geq 1)$. Thus:

$$\Pr(W \leq u) = \Pr\left(\bigwedge_{i=1}^n Y_i\right) = \Pr\left(\bigwedge_{i=1}^n (X_i \geq 1)\right) \leq \prod_{i=1}^n \Pr(X_i \geq 1) = \prod_{i=1}^n \Theta\left(1-1/\sqrt{n}\right)$$
$$= \Theta\left((1-1/\sqrt{n})^n\right) = \Theta\left(\left((1-1/\sqrt{n})^{\sqrt{n}}\right)^{\sqrt{n}}\right) = \Theta\left(e^{-\sqrt{n}}\right).$$

Problem 3

(a) Yes. Suppose $x \in L_1 \cup L_2$. Then there exist polynomial-time verifiers $C_1(s,t)$ and $C_2(s,t)$ for L_1 and L_2, respectively, and certificate t_i which enables x to be accepted by C_1 or C_2. Then we can create a new verifier, $C_\cup(s,t)$, which we define as taking in a string and certificate, feeding this input into C_1, returning "yes" if C_1 returns "yes", and otherwise feeding its input into C_2 and returning whatever C_2 does. We see that so long as t_i is a certificate for x for either one of C_1 or C_2, it will also be a certificate of x for $C_\cup$, and so $C_\cup$ is a polynomial-time verifier for $L_1 \cup L_2$, and therefore $L_1 \cup L_2$ is in NP.

(b) No not necessarily, because if this were the case, we could simply take $\emptyset \cup \overline{L_2} = \overline{L_2}$ to be in NP, and so NP would be equal to co-NP, which is a known open problem.

Problem 4

(a) Let I denote the indices i of the subsets in the solution, S_i denote the ith subset, and $\hat{S}_i$ denote the modified subset S_i (i.e., S_i with some elements removed). The algorithm 1.2 from Williamson-Shmoys can be slightly modified to give the algorithm below:

$I \leftarrow \emptyset$
$\hat{S}_j \leftarrow S_j \quad \forall j$
while $|\bigcup_{j \in I} S_j| < p|E|$ **do**
  $l \leftarrow \arg\min_{j : \hat{S}_j \neq \emptyset} \frac{w_j}{|\hat{S}_j|}$
  $I \leftarrow I \cup \{l\}$
  $\hat{S}_j \leftarrow \hat{S}_j - \hat{S}_l \quad \forall j$

Let n = $|E|$. In the following analysis and in problem (b) as well, we assume pn is an integer (this can be achieved by simply taking the ceiling of pn if pn is a fractional). The algorithm can run no more than pn rounds, and in each round it computes $\mathcal{O}(m)$ ratios, each in constant time, thus it finishes in polynomial time.

It is proved in the textbook that the kth covered element in the solution has cost no more than $\text{OPT}/(n-k+1)$, therefore the total cost of the first pn added elements is:

$$\begin{aligned}\sum_{i=1}^{pn} cost(e_i) &\leq \sum_{i=1}^{pn} \frac{\text{OPT}}{n-i+1} \\ &= \text{OPT} \sum_{j=n-pn+1}^{n} \frac{1}{j} \\ &= (H_n - H_{n-pn})\text{OPT} \\ &\leq \big(1 + \ln n - \ln{(n-pn)}\big)\text{OPT} \\ &= \Big(1 + \ln\Big(\frac{1}{1-p}\Big)\Big)\text{OPT}\end{aligned}$$

where the final inequality follows because $\ln n < H_n < 1 + \ln n$.

The last set added to the solution may introduce more covered elements than necessary, incurring a total cost greater than the bound calculated above. However, the extra cost cannot be more than OPT. Thus, for the value of our final solution, i.e., $\sum_{j \in I} w_j$, we have:

$$\sum_{j \in I} w_j \leq \sum_{i=1}^{pn} cost(e_i) + \text{OPT} = \Big(1 + \ln\Big(\frac{1}{1-p}\Big)\Big)\text{OPT} + \text{OPT} = \Big(2 + \ln\Big(\frac{1}{1-p}\Big)\Big)\text{OPT}$$

(b) The algorithm proposed in (a) can be slightly modified to give an $f(p)$-approximation algorithm for the partial cover problem.

Let m denote the number of elements that remain to be covered at the start of the current iteration, i.e., $m = np - |\sum_{j\in I} S_j|$ where I is the partial solution. In this modified algorithm, instead of taking the set corresponding to the minimum of $\frac{w_j}{|\hat{S}_j|}$, we will take the set corresponding to the minimum of $\frac{w_j}{\min(m, |\hat{S}_j|)}$. This modification is made to ensure that only np elements are covered in the final solution. The modified algorithm is shown below:

$I \leftarrow \emptyset$
$\hat{S}_j \leftarrow S_j \quad \forall j$
$m \leftarrow np$
while $m > 0$ **do**
  $l \leftarrow \arg\min_{j:\hat{S}_j \neq \emptyset} \frac{w_j}{\min(m, |\hat{S}_j|)}$
  $I \leftarrow I \cup \{l\}$
  $m \leftarrow m - |\hat{S}_l|$
  $\hat{S}_j \leftarrow \hat{S}_j - \hat{S}_l \quad \forall j$

Let OPT* denote the value of an optimal solution to the partial cover problem. At each iteration, the optimal solution must contain at least m elements that currently are not in our greedy selection, therefore, at each iteration we can always find a subset that covers its elements with an average weight of at most OPT^*/m. Let m_k denote the value of m at kth iteration. For the set j chosen in the kth iteration, we have:

$$w_j \le (m_k - m_{k+1})\frac{\mathrm{OPT}^*}{m_k} = \frac{m_k - m_{k+1}}{m_k}\mathrm{OPT}^*$$

Let $l =$ the number of sets in our final solution. We have:

$$\begin{aligned}
\sum_{j\in I} w_j &\le \sum_{k=1}^{l} \frac{m_k - m_{k+1}}{m_k}\mathrm{OPT}^* \\
&\le \mathrm{OPT}^* \cdot \sum_{k=1}^{l} \Big(\frac{1}{m_k} + \frac{1}{m_k - 1} + \cdots + \frac{1}{m_{k+1}+1}\Big) \\
&= \mathrm{OPT}^* \cdot \sum_{i=1}^{np} \frac{1}{i} \\
&= H_{pn} \cdot \mathrm{OPT}^*
\end{aligned}$$

where the second inequality follows from the fact that $\frac{1}{m_k} \le \frac{1}{m_k - i}$ for each positive i.

H_{np} is nondecreasing in p. When $p = 1$, $H_{np} = H_n = H_{|E|}$.

Problem 5

We first show a Karp reduction from the unweighted subset problem to the Steiner directed subtree problem. We start by making a root vertex. For each element in the set we are trying to cover, make a terminal vertex. For each subset we have, make a non-terminal vertex. Connect the root vertex to each of the subset vertices with an arc of weight 1, which represents the cost of 1 for choosing to go through that particular subset. Then connect each subset vertex with each of its element vertices using arcs of weight 0, which represents that if we chose to go through a particular subset vertex (incurring the cost of 1), we get all of its element terminal vertices for free. So if we find a minimal subtree that covers every terminal vertex, we must have gone through a minimal number of subset vertices, and this exactly corresponds to a selection of the minimal number of subsets which covers every element of our set in the subset problem.

Hence, a $c \ln |T|$-approximation to the Steiner directed subtree problem could thus be used to create a $c \ln |T|$-approximation to the unweighted subset problem, and vice versa. But if such an approximation exists, then, as Theorem 1.14 from Williamson-Shmoys states, $P = NP$.