NATIONAL UNIVERSITY OF SINGAPORE
SCHOOL OF COMPUTING

Midterm Assessment for CS5330 – Randomized Algorithms

3 March 2020 Time Allowed: 180 minutes

INSTRUCTIONS:

- This paper consists of **SIX** questions.
- This is an **OPEN BOOK/NOTES** examination.
- Do **NOT** use the internet for help while taking the examination. Any evidence otherwise will result in a deduction of **at least 20%** for that problem.
- Do **NOT** collaborate while taking the examination. Any evidence otherwise will result in a deduction of **at least 20%** for that problem.
- Either type or hand-write your solutions. Please submit your solutions in PDF format to the LumiNUS submission folder before **10:00 PM**.
- Please write your student number on the top of each sheet that you use. Do **NOT** write your name.

QUESTIONS:

1. **(20 points)** A random variable X takes values over the non-negative integers. X satisfies $\mathbb{E}[X] = 1$ and $\mathsf{Var}[X] = 2$. Decide if each of the following **must** be true about X. If the statement is true, write why. If it is false, show a counterexample.

   (a) $\Pr[X \geq 5] \leq \frac{1}{5}$

   (b) $\Pr[X \geq 2] > 0$

   (c) $\Pr[X = 1] > 0$

   (d) $\Pr[X \leq 1] \leq \Pr[X \geq 1]$.

2. **(10 points)** In the fictional game *mossix*, you repeatedly toss a (fair, six-sided) die. If it ever turns up to be an odd number, you lose. Otherwise, your score is the number of tosses that it takes you to toss the first '6'. What is the probability that your score is 1 conditioned on the event that you don't lose? Explain your reasoning.

3. **(20 points)** Recall Karger's algorithm that repeatedly contracts random edges.

   (a) **(5 points)** Why does Karger's algorithm return the min-cut with probability 1 when it's run on trees?

   (b) **(15 points)** Suppose we modify Karger's algorithm to contract random pairs of vertices at each step (instead of random edges). Describe a graph on n vertices for which this modified algorithm returns a minimum cut with probability $2^{-\Omega(n)}$. Explain why.

4. **(15 points)** Suppose $G = (V, E)$ is an undirected graph with vertex set V and edge set E. Let $n = |V|$ and $m = |E|$.

   Order the vertices in V randomly. Let I consist of the set of vertices v whose neighbors all occur after v in the order. Why is I an independent set (i.e., no edges between vertices in I)? Show that:

$$\mathbb{E}[|I|] = \sum_{v \in V} \frac{1}{\deg(v) + 1}$$

   where $\deg(v)$ is the degree of vertex v.

5. **(15 points)** Your friend runs up to you and breathlessly exclaims that she has found a randomized algorithm that compresses by 10% (in expectation) any binary string of length 10^6. She claims that her compression scheme is such that the original string is always recoverable from the compression. Prove to her that her claims cannot be true by using Yao's minimax principle.

6. **(20 points)** You are in front of a huge playpen filled with balls of different colors. These balls are of n colors, and you want to estimate p_i, the fraction of balls in the pen with color i.

   You sample m balls, each uniformly at random and independently. Let $q_i^{(m)}$ be the fraction of balls in the sample colored i.

   (a) **(6 points)** Show that if $m = O(\varepsilon^{-2} \log \delta^{-1})$, with probability at least $1 - \delta$, $|q_1^{(m)} - p_1| \leq \varepsilon$.

   (b) **(7 points)** Show that if $m = O(\varepsilon^{-2}(n + \log \delta^{-1}))$, with probability at least $1 - \delta$:

$$\max_{S \subseteq [n]} \left| \sum_{i \in S} q_i^{(m)} - \sum_{i \in S} p_i \right| \leq \varepsilon.$$

   (c) **(7 points)** Show that if $m = O(\varepsilon^{-2} \log(n\delta^{-1}))$, with probability at least $1 - \delta$:

$$\max_{i \in [n]} |q_i^{(m)} - p_i| \leq \varepsilon.$$

   (d) **[Optional Bonus! (+10 points)]** For part (c), actually $m = O(\varepsilon^{-2} \log \delta^{-1})$ samples suffice, without any dependence on n. Can you show this?