Name: ______________________________ Roll no. ______________

CSL 7030: Algorithms for Big Data   Quiz 1

Date: 22/01/2025   Time: 10 minutes

Instructions: Each question is worth 2 points.
For all questions including MCQs, you need to give the correct answer **and** a correct explanation for getting full points. Non-MCQs *may* get partial credit for steps.
MCQs have **negative marking**: a wrong answer, or the correct answer without explanation, or the correct answer with a wrong explanation gives -1 point.

1. Let A, B be two events with $\Pr(A) = 0.3$, $\Pr(B) = 0.4$ and $\Pr(A \cap B) = 0.1$. What is $\Pr(A \cup B)$?
   Ans. $\Pr(A \cup B) = \Pr(A) + \Pr(B) - \Pr(A \cap B) = 0.3 + 0.4 - 0.1 = 0.6$.

2. Let M be a matrix whose dimension is $n^{1/2} \times n^{1/4}$. Among the following options, which running time would make an algorithm operating on M a *sublinear-time* algorithm?
   (a) $3n/4$ (b) $n^{3/4}$ (c) $100\sqrt{n}\log n$ (d) $n^{4/3}$
   Ans. Total input size is $n^{1/2} \cdot n^{1/4} = n^{3/4}$. So sublinear time complexity should be $o(n^{3/4})$. The only option satisfying is $100\sqrt{n}\log n$.
   This can be seen as follows: $\lim_{n\to\infty} \frac{100\sqrt{n}\log n}{n^{0.75}} = \lim_{n\to\infty} \frac{100}{n^{0.25}/\log n} = 0$.

3. There is a biased coin that has a 10% chance of showing heads. We toss this coin 1000 times.
   Use Markov's inequality to find a lower bound on the probability that the number of heads is less than 500.
   That is, Pr(number of heads is less than 500) $\geq$ ______
   Ans. Let X_i be a random variable that the ith coin toss is heads. $\Pr(X_i = 1) = \mathrm{E}[X_i] = \frac{1}{10}$. Let X denote the total number of heads $= \sum_{i=1}^{1000} X_i$. Hence $\mathrm{E}[X] = 1000 \cdot 1/10 = 100$.
   By Markov's inequality, $\Pr(X \geq 500) \leq \frac{\mathrm{E}[X]}{500} = \frac{1}{5}$.
   Therefore, $\Pr(X < 500) = \Pr(\overline{X \geq 500}) = 1 - \Pr(X \geq 500) \geq \frac{4}{5}$.

4. Let $h : \mathcal{U} \to \mathcal{R}$ be a hash function chosen from the 2-universal hash family we saw in the class, where $\mathcal{U} = \{0, 1, \ldots, 9999\}$, and $\mathcal{R} = \{0, 1, \ldots, 36\}$. How many bits are required to store h?
   (a) $10000 \cdot \lceil\log_2(37)\rceil$ (b) $\lceil\log_2(10000)\rceil$ (c) $37 \cdot \lceil\log_2(10000)\rceil$ (d) $3\lceil\log_2(37)\rceil$
   Ans. 2-universal hash family from class contains functions $h_{a,b}$ where $h_{a,b}(i) = a \cdot i + b \mod p$. Hence, to store such a hash function, we only need to store three integers a, b, p, each of which takes $\lceil\log_2(p)\rceil$ bits. Here $p = 37$, so the answer is $3 \cdot \lceil\log_2(37)\rceil$ bits.

5. Let $Z_1, Z_2, \ldots, Z_{80}$ be *pairwise independent* 0/1 random variables, such that $\mathrm{E}[Z_i] = \frac{1}{4}$ if i is odd, and $\mathrm{E}[Z_i] = \frac{3}{4}$ if i is even. Let $Z = \sum_{i=1}^{80} Z_i$. Find an upper bound on $\Pr\big((Z \leq 25) \cup (Z \geq 55)\big)$.
   (Note: 0/1 random variable means that the random variable can only take value either 0 or 1.)
   Ans. If X is a 0/1 random variable with $\Pr(X = 1) = \mathrm{E}(X) = p$. For such a random variable, $\mathrm{Var}[X] = p(1-p)$.
   Here, for even i, $\mathrm{Var}[Z_i] = \frac{1}{4} \cdot \frac{3}{4} = \frac{3}{16}$. Turns out that this is also the variance for the odd i's by symmetry. By linearity of expectations, $\mathrm{E}[Z] = \sum_{\text{odd } i} \mathrm{E}[Z_i] + \sum_{\text{even } i} \mathrm{E}[Z_i] = 40 \cdot \frac{3}{4} + 40 \cdot \frac{1}{4} = 40$.
   Since Z_j's are pairwise independent, $\mathrm{Var}[Z] = \sum_{i=1}^{80} \mathrm{Var}[Z_i] = 80 \cdot \frac{3}{16} = 15$.
   Then by Chebyshev's inequality,
   $$\Pr\left((Z \leq 25) \cup (Z \geq 55)\right) = \Pr\left(|Z - 40| \geq 15\right) \leq \frac{\mathrm{Var}[Z]}{15^2} = \frac{1}{15}.$$