Please list your collaborators, including AI assistance, for *each* question. This will not affect your marks.

There are totally 60 marks, and the full mark is 40. This homework is counted 5% of the course. Please read the course outline for the late submission policy.

---

1. **Online Hiring**

   (10 marks) We need to hire a new staff. There are n applicants for this job. Assume that we will know how good they are (as a score) when we interview them, and the score for each applicant is different. So there is a unique candidate with the highest score, but we don't know that the applicant is the best when we interview him/her until we have interviewed all the applicants. The problem is that after we interview one applicant, we need to make an online decision to either give him/her an offer or forever lose the chance to hire that applicant. Suppose the applicants come in a random order (i.e. a uniformly random permutation), and we would like to come up with a strategy to hire the best applicant.

   Consider the following strategy. First, interview m applicants but reject them all. Then, after the m-th applicant, hire the first applicant we interview who is better than all of the previous applicants that we have interviewed.

   Let E be the event that we hire the best applicant. Let E_i be the event that the i-th applicant is the best and we hire him/her. Compute $\Pr(E_i)$ and show that

   $$\Pr(E) = \frac{m}{n} \sum_{j=m+1}^{n} \frac{1}{j-1}.$$

   Then, show that $\Pr(E) \geq \frac{m}{n}(\ln n - \ln m)$, and that $\Pr(E)$ can be arbitrarily close to $1/e$ for an appropriate choice of m when n tends to infinity.

2. **Minimum k-cut**

   (10 marks) Generalizing on the notion of a cut-set, we define a k-way cut-set in an undirected graph as a subset of edges whose removal leaves the graph into k or more connected components. Show that the randomized contraction algorithm can be modified to find a k-way cut-set with minimum cardinality in $n^{O(k)}$ time.

3. **Uncrossing Minimum Cuts**

   (10 marks) Let $G = (V, E)$ be an undirected graph. Let $X \subset V$ be a minimum s-t cut with $s \in X$ and $t \notin X$. Let $u, v \in X$. Prove that there is a minimum u-v cut Y such that $Y \subseteq X$.

   (Remark: This is a key step in constructing Gomory-Hu tree, a compact data structure to represent all minimum u-v cuts of an undirected graph.)

4. **Quicksort**

   (10 marks) The expected runtime of the randomized quicksort algorithm is at most $2n \ln n$ steps where n is the number of elements to be sorted (see Section 2.5 of "Probability and Computing" for a proof). Prove that the probability that the actual runtime is more than say $100n \ln n$ is at most inverse polynomial in n. (Hint: Bound the probability that the recursion depth is large.)

5. **Streaming Sampling**

   (10 marks) Suppose we have a long sequence of numbers coming one at a time. We would like to maintain a set of numbers of size k with the property that for each number we have seen so far, the probability that the number appears in the set are equal.

   We want to accomplish this without knowing the total number of items in advance or storing all of the items that we have seen.

   Prove that the following simple algorithm works. When the first k numbers come, we put it in the set. After that when the m-th number appears, with probability k/m, we replace a random number in the set with the m-th number.

6. **Approximate Median in Sublinear Time**

   (10 marks) We would like to find an approximate median of n distinct numbers in sublinear time. To do so, we sample $m \ll n$ numbers with replacement, find the median c of these m numbers, and report c as the approximate median of the n numbers. Let the sorted list of the n numbers be $\{x_1, x_2, \ldots, x_n\}$ and so the true median is $x_{n/2}$. The approximate median is said to be a $\pm k$-approximation if $c \in [x_{\frac{n}{2}-k}, x_{\frac{n}{2}+k}]$. Suppose we want the algorithm to succeed to find a $\pm k$-approximation with probability at least 0.9999. What is the tradeoff between m and k? How large should we set m if we want $k \leq \epsilon n$ for some small constant ϵ? How large should we set m if we want $k \leq n^{1-\eta}$ for $\eta \leq 1/2$?