You are allowed to discuss with others but not allowed to use any references except the course notes and the books "Probability and Computing" and "Randomized Algorithms". Please list your collaborators for *each* question. This will not affect your marks. In any case, you must write your own solutions.

There are totally 50 marks, and the full mark is 40. This homework is counted 8% of the course. The extra marks will not be carried to other parts 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. **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.)

4. **Uniform Sampling for Graph Sparsification**

   (10 marks) Suppose we run the uniform sampling algorithm for graph sparsification in L03 on the complete graph on n vertices. Prove that if we set the sampling probability $p := \frac{c \ln n}{n}$ for a constant $c < 1$, then for any constant $\epsilon > 0$ and for n sufficiently large, the graph has isolated vertices with probability at least $1 - \epsilon$.

   (Remark: This implies that the uniform sampling algorithm for graph sparsification in L03 requires $\Omega(n \ln n)$ edges to form a good sparsifier if the input graph is a complete graph.)

5. **Chernoff Bound** (Bonus)

   (10 marks) Recall that X is a standard normal random variable if its probability density function is $f(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2}$.

   (a) Compute the moment generating function $M_{X^2}(t) := E[e^{tX^2}]$. You may use the fact that $\int_{-\infty}^{\infty} f(x) = 1$ without providing a proof.

   (b) Use (a), or otherwise, to compute $E[X^4]$.

   (c) Let $X_1, \ldots, X_k$ be independent standard normal random variables, and $Y = \frac{1}{k} \sum_{i=1}^{k} X_i^2$. Derive a Chernoff bound to prove that $\Pr(Y > 1 + \epsilon) \leq e^{-k\epsilon^2/8}$ for $0 \leq \epsilon \leq 1$. You may use the Taylor expansion $\ln(1 - x) = -\sum_{i=1}^{\infty} x^i/i$ for $-1 \leq x \leq 1$.