This homework is counted 20% of the whole course. There are total 120 marks. Your final score will be min{110, your actual score}. There are many questions, and so you should start early.

You are allowed to discuss with others and use any references, but if you do so please list your collaborators and cite your references for each question. This will not affect your marks. In any case, you must write your own solutions (meaning that after you understand the solutions either by your own or from discussions with others or from references, you must write the solutions in your own words without any help from others and without any help from the references). Not writing your own solutions or not listing your collaborators or not citing your references may be considered plagiarism.

Hope you will enjoy solving the problems. If you have any questions or comments please let me (chi@cse.cuhk.edu.hk) and/or your TA Chiu (tckwok@cse.cuhk.edu.hk) know. We will be happy to help.

---

1. **Questionnaire**

   (20 marks) Please complete the questionnaire. You don't need to do it again if you have already done so. You will get full marks as long as you have answered all the required questions (i.e. you can submit it successfully).

2. **Sampling**

   (5 marks) Suppose we have a long sequence of numbers coming one at a time. We would like to maintain a sample of one number with the property that it is uniformly distributed over all the items that we have seen at each step. We want to accomplish this without knowing the total number of items in advance or storing all of the items that we see.

   Prove that the following simple algorithm works. When the first item comes, it is stored in the memory. When the k-th item appears, it replaces the item in memory with probability $1/k$.

3. **Minimum Cut**

   (a) (10 marks) Generalizing on the notion of a cut-set, we define an k-way cut-set in an undirected graph as a set of edges whose removal breaks the graph into k or more connected components. Explain how the randomized min-cut algorithm can be used to find minimum k-way cut sets. Bound the probability that it succeeds in one iteration and bound the total running time for it to have success probability at least $1 - 1/n$ where n is the number of vertices in the graph.

   (b) (5 marks) Given an undirected graph G with minimum-cut size c, prove that G has at most $O(n^{2\alpha})$ cuts with at most αc edges.

4. **Coupon Collectors**

   (10 marks) Now you know that it requires $\Theta(n \log n)$ coupons to collect n different types of coupons, with at least one coupon per type. You wonder whether it would be more efficient if a group of k people cooperate, such that each person buys cn coupons and exchange with each other.

   Prove the best bound you can on k to ensure that, with probability at least 0.9, each person only needs to buy at most $10n$ coupons.

5. **Random Permutations**

   A permutation $\pi : [n] \to [n]$ can be represented as a set of cycles as follows. Let there be one vertex for each number i for $1 \le i \le n$. If the permutation maps the number i to the number $\pi(i)$, then a directed arc is drawn from vertex i to vertex $\pi(i)$. This leads to a graph that is a set of disjoint cycles. Notice that some of the cycles could be self-loops.

   (a) (10 marks) What is the expected number of cycles in a random permutation of n numbers?

   (b) (10 marks) Let X be the number of self-loops in a random permutation of n numbers. What are $\mathbb{E}[X]$ and $\mathrm{Var}[X]$?

6. **Online Hiring**

   You need to hire a new staff. There are n applicants for this job. Assume that you will know how good they are (as a score) when you interview them, and the score for each applicant is different. So there is a unique candidate with the highest score, but you don't know that the applicant is the best when you interview him/her until you have interviewed all the applicants. The difficulty is that after you interview one applicant, you 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 you 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 you interview who is better than all of the previous applicants that you have interviewed.

   (a) (10 marks) Let E be the event that you hire the best applicant. Let E_i be the event that the i-th applicant is the best and you hire him/her. Compute $\Pr(E_i)$ and show that

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

   (b) (5 marks) Prove that $\Pr(E) \ge 1/e$ for an appropriate choice of m.

7. **Graph Drawing**

   (10 marks) A graph is *planar* if it can be drawn on the plane such that the edges do not intersect with each other. It is a well-known result that a simple planar graph with n vertices can have at most $3n-6$ edges. We say a graph G has intersecting number k if k is the maximum number such that any drawing of G on the plane has at least k pairs of edges intersecting. By the above result, a simple bound is that the intersecting number is at least t if the graph has at least $3n-6+t$ edges. Prove that the intersecting number is at least $m^3/(64n^2)$ for any simple graph with at least $m \ge 4n$ edges. (Hint: do random sampling and apply the simple bound.)

8. **Error Correcting Code**

We will prove the existence of an error correcting code with *zero* decoding error probability. The model of the noisy channel is slightly different: if we send n bits through the channel, then we are guaranteed that at most pn bits will be flipped (but an "adversary" can decide to flip an arbitrary subset of at most pn bits). We would like to find an encoding function $f : \{0,1\}^m \to \{0,1\}^n$ so that m is as large as possible, while any arbitrary pn errors on the codeword can be tolerated for all m-bit messages.

Consider the following coding scheme. We generate an $n \times m$ matrix A where each entry of A is 0 with probability 1/2 and 1 with probability 1/2. To encode a message x of m bits, we compute the codeword $y = Ax$ where we use arithmetic modulo two so that y is an n-bit string. We then send the codeword y to the receiver through the noisy channel.

(a) (10 marks) Prove that the receiver can recover the message with probability one if and only if Ax has at least $2pn + 1$ nonzero bits for all $x \neq 0$.

(b) (10 marks) Prove that for any $\epsilon > 0$, we can achieve $m \geq (1 - H(2p) - \epsilon)n$ while the receiver can always recover the message for large enough n.

(c) (5 marks) Prove that for any $\epsilon > 0$, we *cannot* achieve $m \geq (1 - H(p) + \epsilon)n$ while the receiver can always recover the message for large enough n.