Randomized Algorithms **NYU, Fall 2025**
HW #1 (out Friday Sept 19, 2025) **Due:** Monday Sept 29, 2025

Please collaborate in groups of 2 (or at most 3). Write your own solutions, no sharing of written content. Put down names of your collaborator(s) on the front page. Submissions will be via gradescope, and the link will appear on the course webpage and on Brightspace. Also, changes, corrections, and clarifications will also appear on the Ed discussion board, so please check it regularly.

Exercises

1. (Non-Musical Chairs) An airplane in Politesville has n seats, and n passengers assigned to these seats. The first passenger to board gets confused, and sits down at a uniformly random seat. The rest of the passengers do the following: when they board, if their assigned seat is free they sit in it, else (being too polite) they choose a uniformly random empty seat and sit in it. (a) Show that the last person to board sits in their assigned seat with probability $1/2$. (b) Show that the expected number of people who board to find their assigned seat already occupied is $H_n - 1 = 1/2 + 1/3 + \ldots + 1/n$.

2. (Number of Near Min-Cuts) An α-min-cut is a subset S such that the weight of ∂S is at most α times the weight of the minimum cut. Show that the probability that any particular α-min-cut is output is at least $1/n^{2\alpha}$. Hence infer that the number of α-min-cuts is at most $n^{2\alpha}$.

3. (Estimate the Coin's Bias.) You have a coin with some unknown bias q. To estimate p, you flip it $T := O\big(\frac{1}{\varepsilon^2 \delta}\big)$ times, and suppose it comes up heads K times. You output the estimate $Q := K/T$. Use Chebyshev's inequality to show that $\mathbf{Pr}[|Q - q| \leq \varepsilon] \geq 1 - \delta$.

4. (Large Cliques) In lecture, we claimed that for any $\varepsilon > 0$, the random graph $G(n, 1/2)$ has a clique of size $(2 - \varepsilon) \log_2 n$ with probability $1 - o(1)$. Prove this for yourself using the second-moment method. (Hint: if X_S is the indicator of the event that set $S \subseteq V$ is a clique, then the covariance $\mathbf{Cov}(X_S, X_T)$ may no longer be non-positive.)

Problems

Please write short and clear solutions to each of these problems. Use the language of probability to your advantage. Be clear what the events are, what probabilities and expectations you are reasoning about.

1. (Lonely Vertices) Consider the Erdős-Rényi random graph $G(n, p)$ and suppose $\varepsilon > 0$ is some constant. Show that:
   (a) if $p \geq \frac{(1+\varepsilon)\ln n}{n-1}$ then the graph has at no isolated vertices with probability $1 - o(1)$.
   (b) if $p \leq \frac{(1-\varepsilon)\ln n}{n-1}$ then the graph has at least one isolated vertex with probability $1 - o(1)$.

2. (Cutting it Fine.) Given a connected undirected (and unweighted, for simplicity) n-vertex graph G, a k-cut is a paritition of the vertex set into k non-empty parts $S_1, S_2, \ldots, S_k$. The size of the k-cut is the number of edges that cross between distinct parts in this partition. Consider the following algorithm:

- (Phase 1) As long as the number of vertices in G is more than $2k-2$, pick a random edge and contract it.
- (Phase 2) Now we have a graph on $2k-2$ vertices. For each of these vertices, choose a random label from $\{L_1, L_2, \ldots, L_k\}$, contract vertices with the same label, and output the resulting cut.

Show the following:

(a) If the min k-cut size is λ, then $\lambda \leq 2(k-1)\frac{m}{n}$.

(b) Any fixed min k-cut survives the first phase with probability at least $1/\binom{n}{2(k-1)}$.

(c) Conditioned on surviving the first phase, it is output in the second phase with probability at least $\frac{k!}{k^k} \cdot \frac{1}{k^{k-2}}$.

Hence, this gives an $\approx n^{2(k-1)}$-time algorithm for the k-cut problem.

3. (Cyclic Changes.) An ℓ-cycle in a graph is a cycle with *at most* ℓ nodes; please note the "at most". In this problem, we want to show there exist graphs with many edges, and no *short* cycles. It is easy to construct such graphs with $\Omega(n)$ edges—in fact, a tree has $n-1$ edges and no cycles at all! We want slightly denser graphs with no ℓ-cycles for any constant ℓ.

   (a) Consider the graph $G(n,p)$ for some $p \in [0,1]$. Calculate the probability that some sequence of ~~ℓ~~ k vertices is a cycle, and hence calculate the expected number of ℓ-cycles.

   (b) (Do not submit.) Note that setting $p \approx 1/n$ means the expected number of ℓ-cycles is $o(1)$, but the expected number of edges is $O(n)$—which is not very interesting!

   (c) Now consider the following two part algorithm: (i) first pick $G \sim G(n,p)$, and then (ii) for each ℓ-cycle with $k \leq \ell$ in G, delete an arbitrary edge on it. By construction this graph has no ℓ-cycles i.e., cycles of length at most ℓ. Show that setting $p = n^{\frac{2-\ell}{\ell-1}}$ ensures that the expected number of edges in the resulting graph is $O(m)$ with $m := n^{1+\frac{1}{\ell-1}}$. Hence, infer that there exist graphs with $m = \omega(n)$ edges and no ℓ-cycles.

4. (Taken from Johan Håstad's course "Theoreticians toolkit" at KTH.) Constructing a random 3-SAT formula with n variables and $m = \lceil dn \rceil$ clauses[1] is done as follows: Randomly take three different variables (all triples being equally likely). Choose one of the eight ways to negate these variables (uniformly at random) and make them into a clause. Repeat with independent randomness until you have m clauses.

   (a) For what value of d is the expected number of satisfying assignments $\Theta(1)$? Call this value d_0.

   (b) Prove that the formula is likely (with probability $1-o(1)$) to be unsatisfiable for any constant d such that $d \geq (d_0 + \varepsilon)$.

   (c) (This is more challenging) Prove that the formula remains at least somewhat likely to be unsatisfiable also in the case when d is slightly smaller than d_0. The difficulty of this problem is very much dependent on what we mean by "somewhat likely" and "slightly smaller". The exact formulation to prove to get a full score on this problem is that there is some constant $d_1 < d_0$ such that for $d = d_1$ the probability that the corresponding random formula is satisfiable is at most $1/2$. The size of $d_0 - d_1$ does not matter for your score on the problem and the main property of a solution to aim for is a mathematically correct argument. *Hint:* A satisfiable formula that does not depend on all its variables has many satisfying assignments.

[1] $\lceil x \rceil$ is the smallest integer that is at least x