Randomized Algorithms **NYU, Fall 2025**
HW #2 (out Wednesday Oct 1, 2025) **Due:** Monday Oct 13, 2025

Please submit solutions only to the problems, not to exercises. **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 and also in the last problem.** 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. **(Coins from coins.)** We often say "sample each item with probability q". It's interesting to ask: how do we do this sampling?
   (a) Suppose we are given an unbiased coin U (of bias $1/2$). You can clearly simulate a coin of bias $1/4$ by flipping U twice, and saying "Heads" if you see HH, and "Tails" otherwise. Indeed, you can use i flips to simulate any $p \in [0,1]$ of the form $p = \frac{K}{2^i}$. However, show that no finite number of coin flips suffice to simulate a coin of bias $1/3$.
   (b) Complement the above result by giving a protocol that simulates a bias-$1/3$ coin using an *expected* constant number of flips of U.
   Extend your result to give a protocol simulating a coin of any bias p, e.g., $p = 1/\pi$, or $p = 1/\sqrt{17}$, again using *expected* constant flips. (Hint: Assume you can generate p's binary representation efficiently, say $1/\pi = 0.01010001011111001100000011011011....$)
   (c) Now let's do the other way. You are given a (possibly) biased coin B with some constant bias $p \in (0,1)$. *You do not know p.* Show how to generate simulate an unbiased coin (i.e., one with bias $1/2$) using an expected $O(1)$ flips of B—here the constant can depend on p.

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

3. **(Low-Discrepancy Colorings.)** Let $U = \{1, \ldots, n\}$ be a universe of elements, and let $\mathcal{S} = \{S_1, \ldots, S_m\}$ be a collection of m subsets of U. We want to find a coloring $\chi : U \to \{-1, +1\}$ that minimizes the maximum imbalance (discrepancy) of the collection. The imbalance of a set S_i is defined as

$$\mathrm{Imb}(S_i) = \left| \sum_{j \in S_i} \chi(j) \right|.$$

   Consider a randomized coloring where each element j independently chooses $\chi(j) = +1$ or $\chi(j) = -1$ with probability 1/2 each.
   (a) Focus on a single set S_i. Let $Z_i = \sum_{j\in S_i} \chi(j)$. What is $\mathbb{E}[Z_i]$?
   (b) Show that for any $\Delta > 0$:

$$\mathbf{Pr}[|\mathrm{Imb}(S_i)| \ge \Delta] \le 2e^{-\Delta^2/(2|S_i|)}.$$

(c) Use the probabilistic method to prove that there must exist a coloring χ such that the maximum imbalance across all sets is bounded by $O(\sqrt{n \log m})$.

Remark: A stronger argument shows that there is a coloring with maximum imbalance $O(\sqrt{n \log(m/n)})$ when $m \geq n$. We aim to cover this later in class. This is known as "Six standard deviations suffice," which is a result by Spencer from 1985. For $m = n$, Spencer's theorem says that there is a coloring with discrepancy at most $6\sqrt{n}$, i.e., six standard deviations.

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. **If you use any concentration bounds, please clearly make sure you argue that the conditions are satisfied.**

1. **Only Connect!** Given an undirected (unweighted) graph $G = (V, E)$, let $G(p)$ be the random graph where we retain each edge of G independently with probability p. In lecture #4, we saw that setting $p \geq c\frac{\log n}{\lambda}$, where λ is the min-cut value in G, the graph $G(p)$ is a cut-approximator for G with probability $1 - o(1)$. In particular, we get the simpler fact: if G is connected, then G_p is also connected whp. Let's prove this simpler fact in a different way that does not use the cut-counting lemma. Consider the following process:

   Initialize $G_0 = G$, and define $L = 100 \log n$. For each $i = 1, 2, \ldots, L$, let S_i be a set where we pick each edge in G_{i-1} independently with probability $1/\lambda$. Contract all the edges from S_i in the graph G_{i-1} (and remove self-loops) to get G_i.

   (a) For any vertex v in G_{i-1}, let $\mathcal{G}_{v,i}$ be the event that the set S_i contains at least one edge incident to v. Show that $\mathbf{Pr}[\mathcal{G}_{v,i}] \geq 1 - {}^{1}\!/_{e}$ assuming that G_{i-1} contains at least two vertices. Btw, are $G_{v,i}$ and $G_{u,i}$ independent?

   (b) Let N_i be the number of vertices in G_i, so that $N_0 = n$. Define the event $\mathcal{E}_i$ that is true if $N_i \leq N_{i-1} \cdot {}^{3}\!/_{4}$ or $N_i = 1$. Show that $\mathbf{Pr}[\mathcal{E}_i] \geq c$, for some absolute constant $c > 0$.

   (c) Use a Chernoff bound to show that $|N_L| = 1$ with probability at least $1 - 1/\text{poly}(n)$. Please clearly state what random variables are you summing over, and why they are independent and bounded.

   (d) Finally, define $S = \cup_{i=1}^{L} S_i$, and note that each edge in G belongs to S with probability at most ~~$L \cdot q$~~ L/λ. Infer that sampling each edge of G with probability $p := L/\lambda$ gives us a connected graph with high probability.

2. **Nearly Orthonormal Vectors.** Call a set of unit vectors "near-orthonormal" if the inner product of any two of them is close to zero. In this problem we will show that while there are at most d orthonormal vectors in $\mathbb{R}^d$, there can be exponentially many near-orthonormal vectors! For vectors $x, y \in \mathbb{R}^d$, we use $\langle x, y \rangle = \sum_{i=1}^{d} x_i y_i$ to denote the inner product.

   (a) Let $x = (x_1, x_2, \ldots, x_d)$ and $y = (y_1, y_2, \ldots, y_d)$ be two independently and uniformly chosen vectors in $\{-1, 1\}^d$. (I.e., each bit x_i and y_i in each vector is independently and uniformly chosen from $\{-1, 1\}$.) Show that

   $$\mathbf{Pr}[|\langle x, y \rangle| \geq \varepsilon d] \leq 2 \exp\left(-\varepsilon^2 d/6\right)$$

(b) Given parameter $\varepsilon > 0$, a set S of unit vectors is called *ε-orthonormal* if for all $\vec{x}, \vec{y} \in S$,

$$|\langle \vec{x}, \vec{y} \rangle| \leq \varepsilon.$$

Show that there exists a constant $c > 0$ and constant d_0, such that for any $\varepsilon \leq 1/2$ (say) and any $d \geq d_0$, if you sample $N := \exp(c\varepsilon^2 d)$ random vectors independently and uniformly from the set $\{-\frac{1}{\sqrt{d}}, +\frac{1}{\sqrt{d}}\}^d$, this sampled set is ε-orthonormal with probability at least $1/2$.

3. **An Approximate Counter, and the Median-of-Means Estimator.** Here is a way of maintaining an approximate counter. (Call this the *basic* counter.)

> Start with $X \leftarrow 0$. When an element arrives, increment X by 1 with probability 2^{-X}. When queried, return $N := 2^X - 1$.

(a) Suppose the actual count is n, show that $\mathbb{E}[N] = n$, and $\mathbf{Var}(N) = \frac{n(n-1)}{2}$.

Since its variance is large, average k independent basic counters $N_1, N_2, \ldots, N_k$, and output the sample average $\widehat{N} := \frac{1}{k}\sum_i N_i$. Call this the *$k$-mean counter.*

(b) Show that $\mathbf{Pr}[\widehat{N} \notin (1 \pm \varepsilon)n] \leq \frac{1}{2\varepsilon^2 k}$.

Hence using $k = \frac{1}{2\varepsilon^2\delta}$ counters can make the failure probability at most δ. (I.e., your error is less than εn with "confidence" $1 - \delta$.) Here's a way to use only $K = O(\frac{1}{\varepsilon^2}\log\frac{1}{\delta})$ counters to get the same answer (and the approach is useful in many different contexts beyond this one). We call this counter the *median-of-means counter.*

(c) Suppose Y is a real-valued random variable and let $I \subseteq \mathbb{R}$ denote an interval. Suppose $\mathbf{Pr}[Y \notin I] \leq 1/4$.

Now, take a collection of ℓ-many independent copies of Y and let M denote the median of $Y_1, \ldots, Y_\ell$. Show that by taking $\ell = \Theta(\log(1/\delta)$, we get $\mathbf{Pr}[M \notin I] \leq \delta$. *Hint: what must happen for the the median to be too high? What is the chance of that?*

(d) Using (c), conclude that by taking Y to be the k_0-mean counter from part (b) with $k_0 = \Theta(1/\epsilon^2)$, we have $\mathbf{Pr}[M \notin (1 \pm \epsilon)n] \leq \delta$.

4. **(Collaboration Acknowledgments.)** Please write down names of people you collaborated with, which online resources you used, and whether you used any LLMs/Chatbots for this problem (and if so, which aspects you used them for). If none, please say so explicitly; it will be useful for me to know this.