Problem Set 3 CS265/CME309, Winter 2026

Due: Friday 1/30, 11:59pm on Gradescope

Please follow the homework policies on the course website.

Note: Throughout this problem set (and in general in this class unless otherwise stated), you are allowed to use as a black box anything we have stated in the lecture notes. In particular, there are several statements of Chernoff and Chernoff-like (Hoeffding, Bernstein) bounds in the lecture notes, and a few corollaries. In this problem set you can make your life a lot easier by finding the "right" one for the problem!

1. **(5 pt.) [The Return of Median-of-means]** Let $X_1, X_2, \ldots, X_n$ be i.i.d. samples of a random variable X with $\mathbb{E}[X] = \mu$ and $\mathrm{Var}[X] \le 1$.

   Suppose that k divides n, let $m = n/k$. Consider the following algorithm for estimating μ from n samples, which we studied in HW2. First, we divide the n samples into k groups of size m:

   $$A_1 = \{X_1, \ldots, X_m\}, A_2 = \{X_{m+1}, \ldots, X_{2m}\}, \cdots, A_k = \{X_{(k-1)m+1}, \ldots, X_n\}.$$

   Then let Y_i be the mean of A_i for each $i = 1, \ldots, k$. Finally, let

   $$\hat{\mu} = \mathrm{Median}(Y_1, \ldots, Y_k).$$

   Let $\epsilon, \delta \in (0, 1)$. Our goal in this problem will be to pick k and m (and hence $n = mk$) in terms of ϵ and δ so that

   $$\Pr[|\hat{\mu} - \mu| > \epsilon] \le \delta. \tag{1}$$

   (a) **(5 pt.)** In HW2, you showed that $n = O\left(\frac{1}{\epsilon^2 \delta}\right)$ samples sufficed to achieve the bound (1) above with the median-of-means scheme. Using a Chernoff bound, show that in fact this scheme (for appropriate values of k and m) can achieve (1) with $n = O\left(\frac{\log(1/\delta)}{\epsilon^2}\right)$, much smaller than the n from HW2.
   You may use the statement from part (a) of HW2's problem: $\Pr[|Y_j - \mu| > \epsilon] \le \frac{1}{m\epsilon^2}$.
   In your solution, make sure you clearly state how to choose m and k, in terms of ϵ and δ.
   [**HINT:** *You may want to look back at the structure of the proof from HW2.*]

   (b) **(0 pt.) [Optional: this part won't be graded.]** Your friend thinks that all this median-of-means thing is over-rated. They say:

   "Suppose that we define $\hat{\mu} = \frac{1}{n}\sum_{i=1}^{n} X_i$ instead. Then since the X_i are fully independent, we can apply some Chernoff/Hoeffding/Bernstein bound to conclude that

   $$\Pr[|\hat{\mu} - \mu| \ge \epsilon] \le \exp(-\Omega(\epsilon^2 n)).$$

   In particular, if we choose n on the order of $\log(1/\delta)/\epsilon^2$, then the right-hand side is at most δ, and we achieve (1). So we can just use the mean instead of median-of-means and get the same thing."

   Is this argument correct (in spirit)? If so, make it rigorous. If not, state an additional assumption on the X_i's that would be needed to make it correct, and then make it rigorous.

2. **(11 pt.) [Graph Coloring]**

The vertices of a simple graph (a graph with no self loops or multiple edges) $G = (V, E)$ are each independently assigned one of three colors: red, green, or blue, chosen uniformly at random.

(a) **(3 pt.)** Let $n = |V|$ be the number of vertices in the graph. Show that the probability that more than half of the vertices are red is $\exp(-\Omega(n))$.

(b) **(4 pt.)** Let $m = |E|$ be the number of edges in the graph. Show that the probability that more than half of the edges are monochromatic (i.e., both endpoints have the same color) is $O(1/m)$.

(c) **(4 pt.)** Suppose G is a complete graph (that is, all of the $\binom{n}{2}$ possible edges are in the graph). Show that the probability that more than half of the edges are monochromatic is $\exp(-\Omega(\sqrt{m}))$.

[**HINT:** *Use similar reasoning as you did in part (a) to say something about the vertices.*]

(d) **(0 pt.) [Optional: This won't be graded.]** Improve the bound from part (b).

3. **(11 pt.) [Concentration without Independence]**

A computer system has n different failure modes, each of which happens with a small probability. Fortunately, the system is designed to be robust in the following sense: As long as less than half of the failure modes occur, things are fine; otherwise, a large-scale crash will happen. We want to make sure that the probability of crashing is small enough.

To model the above scenario, we define n Bernoulli random variables $X_1, \ldots, X_n$. Each X_i is the indicator of the i-th failure mode, i.e., $X_i = 1$ if failure i occurs and $X_i = 0$ otheriwse. Our goal is to upper bound the probability $\Pr\left[\sum_{i=1}^{n} X_i \geq n/2\right]$.

(a) **(2 pt.)** Let's first assume that the n failure events are independent and the probability of each failure is at most $1/3$. Formally, we have:

Assumption 1. $\Pr[X_i = 1] \leq 1/3$ *for every* $i \in [n]$ *and* $X_1, \ldots, X_n$ *are independent.*

Prove that under Assumption 1, for some constant $C > 0$ that does not depend on n,

$$\Pr\left[\sum_{i=1}^{n} X_i \geq n/2\right] \leq e^{-Cn}. \tag{2}$$

Thus, you are to show that the probability of a crash is exponentially small in n.

[**HINT:** *The inequality* $\frac{e^\delta}{(1+\delta)^{1+\delta}} \leq e^{-\delta^2/3}$ *for* $\delta \in [0, 1]$, *which we used in the lecture notes, might be useful if you use Corollary 6 from the lecture notes.*]

(b) **(1 pt.)** Now we decide that Assumption 1 is too unrealistic, since many of the failure modes are known to be strongly correlated. Show that only assuming $\Pr[X_i = 1] \leq 1/3$ (and not the independence), the probability of crashing can be as large as $\Omega(1)$. In particular, prove that for any $n \geq 2$, there exist random variables $X_1, \ldots, X_n$ that satisfy: (1) $\Pr[X_i = 1] \leq 1/3$ for every $i \in [n]$; (2) $\Pr\left[\sum_{i=1}^{n} X_i \geq n/2\right] \geq 1/3$.

(c) **(2 pt.)** Let's try the following relaxation of Assumption 1, which states that the probability for k different failures to occur simultaneously is exponentially small in k:

Assumption 2. *For any* $S \subseteq [n]$, $\Pr[X_i = 1 \text{ for all } i \in S] \le (1/3)^{|S|}$.

Show that Assumption 2 is strictly weaker than Assumption 1 by proving:

i. Assumption 1 implies Assumption 2; and
ii. the implication on the other direction does not hold, i.e., there exist some n and $X_1, \ldots, X_n$ that satisfy Assumption 2 but not Assumption 1.

[**HINT:** *For (ii), there exists a counterexample for* $n = 2$.]

(d) **(6 pt.)** Prove that under Assumption 2, inequality (2) holds for some constant $C > 0$. In your proof, you can appeal to the proof of the Chernoff bounds from lecture videos/notes if you need to write it out verbatim at some point. For example, if you manage to upper bound $\Pr\left[\sum_{i=1}^{n} X_i \ge n/2\right]$ by an expression involving the moment-generating function of some random variable Y that is the sum of n independent Bernoulli random variables, you can simply say that "the rest of the proof is exactly the proof of Theorem 2 from Lecture #5".

[**HINT:** *Consider independent Bernoulli random variables* $Y_1, \ldots, Y_n$ *with* $Pr[Y_i = 1] = 1/3$ *for each* $i \in [n]$. *For distinct indices* $i, j, \ell \in [n]$, *does* $\mathbb{E}X_iX_jX_\ell \le \mathbb{E}Y_iY_jY_\ell$ *hold? Can you extend your proof of the inequality to the case with repeating indices?*]

[**HINT:** *Let* $X = \sum_{i=1}^{n} X_i$ *and* $Y = \sum_{i=1}^{n} Y_i$. *What can we say about* $\mathbb{E}X^k$ *and* $\mathbb{E}Y^k$ *for integer* $k \ge 0$? *Considering the identity* $e^z = \sum_{k=0}^{+\infty} \frac{z^k}{k!}$, *what can we say about* $\mathbb{E}e^{tX}$ *and* $\mathbb{E}e^{tY}$ *for any* $t > 0$?]

(e) **(0 pt.)** **[Optional: this won't be graded.]** Can you construct counterexamples, as in Part 3b, that satisfy *pairwise independence* but have a crashing probability of $\Omega(1/n)$? Formally, prove that there exists $C > 0$ such that for any $n \ge 2$, there exist $X_1, \ldots, X_n$ that satisfy: (1) $\Pr[X_i = 1] \le 1/3$; (2) X_i and X_j are independent for distinct $i, j \in [n]$; (3) $\Pr\left[\sum_{i=1}^{n} X_i \ge n/2\right] \ge C/n$.

Note: *This shows that unlike Chebyshev's inequality, Chernoff bounds do not hold if we only assume pairwise independence.*

[**HINT:** *Recall pairwise independent hash functions if you have seen them before. You can use the Bertrand-Chebyshev theorem, which states that for any integer* $n \ge 1$, *there exists a prime number* p *with* $n < p < 2n$.]