Problem Set 2 CS265, Winter 2025

Due: 1/31 (Friday) at 11:59pm on Gradescope

Please follow the homework policies on the course website.

---

1. **(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 $\exp(-\Omega(\sqrt{m}))$.

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

2. **(12 pt.) [Tightness of Markov's and Chebyshev's Inequalities]**

   (a) **(4 pt.)** Show that Markov's inequality is tight. Specifically, for each value $c > 1$, describe a distribution D_c supported on non-negative real numbers such that if the random variable X is drawn according to D_c then

   (1) $\mathbb{E}[X] > 0$, and

   (2) $\Pr[X \geq c\mathbb{E}[X]] = 1/c$.

   (b) **(4 pt.)** Show that Chebyshev's inequality is tight. Specifically, for each value $c > 1$, describe a distribution D_c supported on real numbers such that if the random variable X is drawn according to D_c then

   (1) $\mathbb{E}[X] = 0$ and $\mathrm{Var}[X] = 1$, and

   (2) $\Pr[|X - \mathbb{E}[X]| \geq c\sqrt{\mathrm{Var}[X]}] = 1/c^2$.

   (c) **(4 pt.) [One-sided version of Chebyshev's Inequality]** Consider the following statement: Let X be a random variable with $\mathrm{Var}[X] = 1$. For all $t \geq 0$,

   $$\Pr[X - \mathbb{E}[X] \geq t] \leq \frac{1}{1+t^2}.$$

   i. Prove this statement.

   ii. Show that this is tight: For any $t \geq 0$, come up with a random variable X with distribution D_t and variance 1 for which $\Pr[X - \mathbb{E}[X] \geq t] = \frac{1}{1+t^2}$.

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 $\Pr\left[\sum_{i=1}^{n} X_i \ge n/2\right]$.

(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. $\Pr[X_i = 1] \le 1/3$ *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,

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

Thus, the probability of a crash is exponentially small in n.

[**HINT:** *Feel free to use (without proof) any of the Chernoff bounds in lecture note #5 (including Theorem 2 and Corollaries 5 and 6) and also the inequality* $\frac{e^{\delta}}{(1+\delta)^{1+\delta}} \le e^{-\delta^2/3}$ *for* $\delta \in [0, 1]$.]

(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 $\Pr[X_i = 1] \le 1/3$ (and not the independence), the probability of crashing can be as large as $\Omega(1)$. In particular, prove that for any $n \ge 2$, there exist random variables $X_1, \ldots, X_n$ that satisfy: (1) $\Pr[X_i = 1] \le 1/3$ for every $i \in [n]$; (2) $\Pr\left[\sum_{i=1}^{n} X_i \ge n/2\right] \ge 1/3$.

(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 (1) 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* $\mathbb{E}[X_i X_j X_\ell] \le \mathbb{E}[Y_i Y_j Y_\ell]$ *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* $\mathbb{E}\left[X^k\right]$ *and* $\mathbb{E}\left[Y^k\right]$ *for integer* $k \ge 0$? *Considering the identity* $e^z = \sum_{k=0}^{+\infty} \frac{z^k}{k!}$, *what can we say about* $\mathbb{E}\left[e^{tX}\right]$ *and* $\mathbb{E}\left[e^{tY}\right]$ *for any* $t > 0$?]

(e) **(0 pt.) [Optional: this won't be graded.]** Can you construct counterexamples for 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* $n < p < 2n$.]