CSCI 145 Problem Set 11

November 11, 2025

Submission Instructions

Please upload *your* work by **11:59pm Monday November 17, 2025.**

- You are encouraged to discuss ideas and work with your classmates. However, you **must acknowledge** your collaborators at the top of each solution on which you collaborated with others and you **must write** your solutions independently.
- Your solutions to theory questions must be written legibly, or typeset in LaTeX or markdown. If you would like to use LaTeX, you can import the source of this document (available from the course webpage) to Overleaf.
- I recommend that you write your solutions to coding questions in a Jupyter notebook using Google Colab.
- You should submit your solutions as a **single PDF** via the assignment on Gradescope.

Grading: The point of the problem set is for *you* to learn. To this end, I hope to disincentivize the use of LLMs by **not** grading your work for correctness. Instead, you will grade your own work by comparing it to my solutions. This self-grade is due the Friday *after* the problem set is due, also on Gradescope.

Problem 1: Concentration

Consider n *independent* random variables $X_1, \dots, X_n$. Each random variable is bounded between 0 and 1, i.e., $X_i \in [0, 1]$ Define the empirical mean

$$\bar{X} = \frac{1}{n} \sum_{i=1}^{n} X_i. \tag{1}$$

In this problem, we will analyze $\bar{X}$.

Part A: Variance

Prove that $\mathrm{Var}(\bar{X}) \leq \frac{1}{n}$.

Hint: Use the linearity of variance for independent random variables.

Part B: Concentration Inequalities

Apply Markov's, Chebyshev's, and Hoeffding's inequality to bound how $\bar{X}$ deviates from its mean $\mathbb{E}[\bar{X}]$. Rearrange the inequalities so that the probabilities are at most $\frac{1}{\epsilon}$, $\frac{1}{\epsilon^2}$, and $\frac{1}{e^\epsilon}$ respectively.

Note: Use part A when manipulating Chebyshev's inequality.

Part C: Union Bound

In the prior part, we saw that—except for the probability bound—Chebyshev's is quite similar to Hoeffding's. Consider m different empirical means $\bar{X}^{(1)}, \dots, \bar{X}^{(m)}$, each with their own n independent random variables $X_1^{(j)}, \dots, X_n^{(j)}$ for $j = 1, \dots, m$.

Use Hoeffding's and the Union Bound to upper bound the probability that, for *any* j, we have

$$\bar{X}^{(j)} - \mathbb{E}[\bar{X}^{(j)}] > \frac{\epsilon}{\sqrt{2n}}. \tag{2}$$

How should we set ϵ if we want this probability to be $\frac{1}{100}$?