CS 3190: Foundations of Data Analysis **Due: Tuesday 9.27 at 11:50pm**

Homework 2: Convergence and Linear Algebra

Instructions: Your answers are due at 11:50pm submitted on canvas. You must turn in a pdf through canvas. I recommend using latex (`http://www.cs.utah.edu/~jeffp/teaching/latex/`, see also `http://overleaf.com`) for producing the assignment answers. If the answers are too hard to read you will lose points, entire questions may be given a 0 (e.g. **sloppy pictures with your phone's camera are not ok, but very careful ones are**)

Please make sure your name appears at the top of the page.

You may discuss the concepts with your classmates, but write up the answers entirely on your own. **Be sure to show all the work involved in deriving your answers! If you just give a final answer without explanation, you may not receive credit for that question.**

---

1. [**35 points**] Consider a pdf f so that a random variable $X \sim f$ has expected value $\mathbf{E}(X) = 10$ and variance $\mathbf{Var}(X) = 1.5$. Now consider $n = 10$ iid random variables $X_1, X_2, \ldots, X_{10}$ drawn from f. Let $\bar{X} = \frac{1}{10}\sum_{i=1}^{10} X_i$

   (a) What is $\mathbf{E}(\bar{X})$?

   (b) What is $\mathbf{Var}(\bar{X})$?

   (c) Given the information we have so far, which of the Concentration of Measure inequalities (Markov, Chebyshev, Chernoff-Hoeffding) can be applied? *Explain why.*

   Assume we know X is never smaller than 3 and never larger than 17

   (d) Use Markov inequality to upper bound $\mathbf{Pr}(\bar{X} > 15)$

   (e) Use Chebyshev inequality to upper bound $\mathbf{Pr}(\bar{X} > 15)$

   (f) Use Chernoff-Hoeffding inequality to upper bound $\mathbf{Pr}(\bar{X} > 15)$

   (g) Now suppose $n = 100$. Calculate the 3 bounds again. For this part, also make sure to report the *name* of the inequality that gives the tightest bound, the second tightest bound, and the third tightest bound respectively.

2. [**15 points**] Let X be a random variable that you know is in the range $[-2, 3]$ and you know has expected value of $\mathbf{E}(X) = .5$ and $\mathbf{Var}(X) = .02$. *(Hint: This question is about how to apply concentration of measure inequalities under change of variables)*

   (a) Use the Markov Inequality to upper bound $\mathbf{Pr}(X > 1)$

   (b) Let $Y = 2X - 1$. Use Chebyshev's inequality to bound $\mathbf{Pr}(Y > 2)$

3. [**25 points**] Consider the following 2 vectors in $\mathbb{R}^3$:

$$\begin{aligned} p &= (1, 2, \mathtt{x}) \\ q &= (\mathtt{y}, 4, 2\mathtt{x}) \end{aligned}$$

   Report the following:

(a) Choose the value `y` so that regardless of the value of `x`, p and q are linearly dependent

(b) Choose the value `y` as a function of `x` so that p and q are orthogonal

(c) if $\mathtt{x} = 1$, then choose a single value of `y` so that p and q are neither linearly dependent nor orthogonal

(d) Calculate $\|p\|_1$ if $\mathtt{x} = 1$

(e) Calculate $\|p\|_2^2$ if $\mathtt{x} = 1$

4. **[25 points]** Consider the following 3 matrices:

$$A = \begin{bmatrix} 1 & -2 & -1 \\ -2 & 3 & 1 \end{bmatrix} \quad B = \begin{bmatrix} 3 & 5 & 2 \\ 3 & 2 & -3 \\ 3 & 3 & 5 \end{bmatrix} \quad C = \begin{bmatrix} 0 & 2 & 0 \\ -1 & 0 & 0 \\ 0 & 0 & 4 \end{bmatrix}$$

Report the following (it is intended that you use Python for this question):

(a) AB

(b) $B + C^T$

(c) Which matrices are full rank?

(d) $\|C\|_F$

(e) C^{-1}