STAT 510: Homework 03

David Dalpiaz

Due: Friday, September 18, 11:59 PM

General Directions

This assignment is worth 10 points. For each exercise, you may obtain a score of 0, 0.5, or 1.

- To obtain a score of **1**, your answer must be correct, contain valid supporting work, and be reasonably formatted up to and including boxing your answer when possible.
- A score of **0.5** will be given to solutions which show reasonabe effort, but contain errors. (A score of **1** may be granted to some solutions containing errors if they are extremely minor.)
- A score of **0** will be given to a blank solution or a solution that shows no reasonable progress towards the correct solution. Note that if you do not indicate a page for a problem on Gradescope, it will be considered blank.

Please submit your assignment to Gradescope by the due date listed above. You may submit up to 48 hours late with a two point late penalty. After that, no late work will be accepted.

Any grade disputes must be petitioned through Gradescope within one week of receiving a grade.

Please attempt to submit your work as a single PDF and complete the process of indicating which problem is on which page. You may need to merge together PDF files from various sources and scans. We will keep track of best practice for submitting to Gradescope in this Piazza thread.

Homework assignments are meant to be learning experiences. You may discuss the exercises with other students, but you must write the solutions on your own. Directly sharing or copying any part of a homework solution is an infraction of the University's rules on academic integrity. Any violation will be punished as severely as possible.

For this, and all homework assignments, you may use any computational tools that you wish, such as a statistical computing enviroment or integral solver. The course staff is most familiar with `R`, so we will be able to best support `R` users, but you may use any software that you like.

Practice Exercises

The following exercises from Evans and Rosenthal contain back-of-the-book solutions.

- Section 3.6: 1, 3, 5, 7, 9, 11, 13
- Section 4.2: 1, 3, 5, 7, 9. 11
- Section 4.4: 1, 3, 5, 7, 9, 11, 13

You do not *need* to do each of these exercises. However, we expect that you are able to complete each of these exercises based on previous coursework. While exam problems will generally be most similar to graded exercises, some exercises similar to practice exercises in difficulty may appear on the exam to stabalize the variance of scores.

Graded Exercises

Exercise 1 (Expectation Review)

Let X_1, X_2, and X_3 be independent Uniform$(0, 1)$ random variables. Define $Y = X_1 - 3X_2 + 2X_3$. Provide an upper bound for $P(|Y| \geq 2)$ using Chebyshev's inequality.

Exercise 2 (Creating a Confidence Interval)

(Based on **LW** 4.4) Let $X_1, X_2, \ldots X_n \sim$ Bernoulli(p). Let $\alpha > 0$ and define

$$\epsilon_n = \sqrt{\frac{1}{2n} \log\left(\frac{2}{\alpha}\right)}.$$

Define $\hat{p}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$ and

$$C_n = (\hat{p}_n - \epsilon_n, \hat{p}_n + \epsilon_n).$$

Show that

$$P(C_n \text{ contains } p) \geq 1 - \alpha.$$

Exercise 3 (Decreasing Rate Poissons)

(Based on **LW** 5.7) Let $\lambda_n = 1/n$ for $n = 1, 2, \ldots$ and let $X_n \sim$ Poisson(λ_n).

Also define $Y_n = nX_n$. Show that

$$Y_n \xrightarrow{p} 0.$$

Exercise 4 (More Classic Setup)

(**LW** 5.3) Let $X_1, X_2, \ldots, X_n$ be independent and identically distributed and $\mu = \mathbb{E}[X_1]$. Give that the variance is finite, show that

$$\bar{X}_n \xrightarrow{qm} \mu.$$

Exercise 5 (The Sample Variance)

(**LW** 5.3) Let $X_1, X_2, \ldots, X_n$ be independent and identically distributed and finite mean $\mu = \mathbb{E}[X_1]$ and finite variance $\sigma^2 = \mathbb{V}[X_1]$. Let $\bar{X}_n$ be the sample mean and let S_n^2 be the sample variance. Show that

$$S_n^2 \xrightarrow{p} \sigma^2.$$

Exercise 6 (Normal Approximations with the CLT)

(Based **LW** 2.8) Suppose we have a computer program consisting of $n = 1000$ lines of code. (And somehow, someone wrote it without debugging along the way.) Let X_i be the number of errors on the i-th line of code. Suppose that the X_i are Poisson with mean 0.01 and that they are independent. Let Y be the sum of the X_i, that is, the total errors. Use the CLT to approximate the probability that there are 5 errors or less. Compare this to the exact probability.

Exercise 7 (CLT with Sample Variance)

Assuming the same conditions as the CLT, and knowing that the CLT exists, show that

$$\frac{\sqrt{n}(\bar{X}_n - \mu)}{S_n} \xrightarrow{D} N(0,1).$$

where S_n^2 is the sample variance.

Exercise 8 (Clever Titles are Hard)

(**LW** 2.14) Let $X_1, \ldots, X_n \sim \text{Uniform}(0,1)$. Let $Y_n = \bar{X}_n^2$. Find the limiting distribution of Y_n.

Exercise 9 (Coverage)

(Based on **LW** 4.4) Return to the results from Exercise 2. Set $\alpha = 0.2$ and $p = 0.4$. Use a simulation study to see how often this interval contains p. We call this quantity the interval's *coverage*. Do this for various sample sizes, n, between 1 and 10,000. Plot the coverage versus n. Note, for each n you will need to perform multiple simulations. Use enough values of n, and enough simulations for each, to create a reasonable looking plot.

Exercise 10 (Rate of Convergence)

So far, we have only been concerned with **if** a random variable converges, and to an extent, **how** a random variable converges, but we have not looked at the **rate** of convergence. To investigate this idea, consider random samples from two different distributions.

1. A Bernoulli like distribution with $P(X = -0.2) = P(X = 0.2) = 0.5$.
2. A t distribution with 2 degrees of freedom.

Note that both of these distributions have mean 0.

Generate a sample of size 10,000 from both and plot the sample mean against the sample size. Repeat this process three times and arrange the plots side-by-side. Comment on which distribute you believe converges faster.

Exercise 11 (BYOQ: Bring Your Own Question)

Submit your own question with a solution! If accepted, you will receive one **buffer point**. To be accepted:

- Your question must be reasonably challenging. It should be at least as challenging as the "average" question on this assignment.
- Is must be **original**. If, based on some quick searching we can find the *exact* question (that is, you just copy-pasted a question), you will **lose** a point instead of receiving a buffer point. If we find that it is simply a derivative of another exercise (for example, just changing a constant) you will not receive a point.

The instructor may create videos solving these, or they may be circulated for additional practice.