Homework 7: Central Limit Theorem

Version 2: We added expectations for work shown for integrals.

For each problem, remember you must briefly explain/justify how you obtained your answer, as correct answers without an explanation will not receive full credit. Moreover, in the event of an incorrect answer, we can still try to give you partial credit based on the explanation you provide.

In general, your goal in an explanation is to write enough that a student from class who has attended lecture, but not read the problem yet, could understand your approach, verify your reasoning, and believe your answer is correct. While we do not usually need to see arithmetic, you must include enough work that in principle one could rederive your answer with only a scientific calculator.

Unless a problem states otherwise, you should leave your answer in terms of factorials, combinations, etc., for instance 26^7 or $26!/7!$ or $26 \cdot \binom{26}{7}$ are all good forms for final answers.

Instructions as to how to upload your solutions to gradescope are on the course web page.

Remember that you must tag your written problems on Gradescope.

Submission: You must upload a **pdf** of your written solutions to Gradescope under "HW 7 [Written]". (Instructions as to how to upload your solutions to gradescope are on the course web page.) The use of latex is *highly recommended*. (Note that if you want to hand-write your solutions, you'll need to scan them. We will take off points for hand-written solutions that are difficult to read due to poor handwriting and neatness.)

Due Date: This assignment is due at 11:59 PM Wednesday May 19 (Seattle time, i.e. GMT-7).
You will submit the written problems as a PDF to gradescope. Please put each numbered problem on its own page of the pdf (this will make selecting pages easier when you submit), and ensure that your pdfs are oriented correctly (e.g. not upside-down or sideways). The coding problem will also be submitted to gradescope.

Collaboration: Please read the full collaboration policy. If you work with others (and you should!), you must still write up your solution independently and name all of your collaborators somewhere on your assignment.

For calculations that require evaluating integrals (unless we indicate otherwise), you must

(a) Show the integral to evaluate (e.g., $\int_0^2 z \cdot 2dz$)

(b) Show an antiderivative and the values to evaluate at (e.g., $z^2|_0^2$)

(c) Plug in the values and simplify (e.g., $2^2 - 0^2 = 4$)

1. Confidence intervals in "real life" [24 points]

In all of the following use the Central Limit Theorem. Use continuity correction if (and only if) you're approximating a discrete random variable.

(a) You've decided to scale up your rap-battle-betting ventures. Now, you've enlisted 20 friends who each bet on some amount of rap battles independently and you owe each friend an average of \$127 with a standard deviation of \$35. You currently do not have any money, but luckily, you are in the process of taking out a second mortgage in order to repay your friends. How large of a loan should you take out in order to repay your friends with probability at least 99%. You should treat these amounts of money that you owe as continuous.

(b) After failing to pay back your mortgage, you've decided to turn to other means in order to repay your friends: cryptocurrency. After a bit of research, you've narrowed down your focus to 8 different currencies. You're thinking of investing \$12000 in every currency. You know that each currency will go 'to the moon' with probability p and you will gross \$50000 (for a net profit of \$50000 - \$12000 = \$38000), and with probability $1 - p$ this currency will not be a trading at a price you will want to sell, so you will make a net profit of -\$12000. As cryptocurrency trading is based primarily on hype, the cryptocurrencies are independent. Your local bank has agreed to loan you the money as long as your net return from these investments is positive with at least 99% probability. What is the condition on p that you need to satisfy in order to secure the loan?

You should treat the amount of revenue you'll get from these investments as discrete (since you're only ever adding up multiples of \$38000 and -\$12000).

(c) You've decided to try a less-risky occupation and you were just put in charge of the TA-hiring process and told that we need to hire 312 TA's. To hire these 312 TA's, you interview every person, and they pass the interview (independently) with probability 0.75. How many people should you interview to guarantee that you will definitely have at least 312 TA's with probability 95%.

2. Polling again [6 points]

In class, we discussed an idealized polling procedure and analysis to estimate the fraction p of a population that supports your run for SAC.

Specifically, we analyzed how we could choose the number n of people to sample in order to guarantee that 95% of the time, it will be the case that $\hat{p} \in [p - .02, p + 0.02]$.

You probably know that many recent polls (that follow similar methodology) have been **way off**. Briefly discuss two reasons that real-world polling may not work as well as the idealized polling that we discussed.

Note: This question is much more like some of the "real-world" assignments than our usual homework questions. There can be many different answers. We expect most reasons will be 1-2 sentences.

3. Exponential in all directions [20 points]

A continuous random variable X has a density function with parameter λ given by:

$$f_X(x) = ce^{-2\lambda|x|} \qquad -\infty < x < \infty,$$

for some constant c.

(a) If λ is equal to 0 or negative, this is not a valid density function. Explain what property of pdfs is violated when $\lambda \leq 0$. [6 points] We recommend you graph the function on wolframalpha, desmos, or some other graphing calculator for a few values of λ before starting on this question.

For the rest of this problem, assume $\lambda > 0$.

(b) Compute the constant c in terms of λ. [4 points]

(c) Compute the mean and variance of X in terms of λ. for this problem, fully evaluating integrals would require integration by parts, for this part, you may skip steps b and c of the standard integral directions. I.e., you may write the integral to evaluate, and skip to the evaluation given by a calculator [5 points]

(d) Compute $Pr(X \geq x)$ in terms of x and λ. (Note that x can be positive or negative or 0. Consider all cases.) [5 points]

4. Distinct Elements Analysis [5 points]

In this problem you will do some theoretical analysis for the code you wrote in HW6.

Recall the setup for the problem: YouTube wants to count the number of ***distinct*** views for a video, but doesn't want to store all the user ID's. In class we described a way for them to get a good estimate of this number without storing everything.

We modelled the problem as follows: we see a **stream** of 8-byte integers (user ID's), $x_1, x_2, \ldots, x_N$, where x_i is the user ID of the i-th view to a video, but there are only n *distinct* elements ($1 \leq n \leq N$), since some people rewatch the video, even multiple times. We don't know what the number of views N is; we can't even store the number n of distinct views (i.e., the number of distinct views).

Let $U_1, \ldots, U_m$ be m iid samples from the continuous `Unif`$(0, 1)$ distribution, and let $X = \min\{U_1, \ldots, U_m\}$. We know from lecture (and the textbook) that $\mathbb{E}[X] = \frac{1}{m+1}$. Compute $\text{Var}(X)$. For this problem, you may start with the pdf of X found on page 338 of the textbook.