Problem Set 7 (due Friday, August 12, 11:59pm)

Directions: *For each problem, explain/justify how you obtained your answer, as correct answers without an explanation may receive **no credit**. Moreover, in the event of an incorrect answer, we can still try to give you partial credit based on the explanation you provide. Unless you are asked to, you should leave your answer in terms of factorials, combinations, etc., for instance* 26^7 *or* $26!/7!$ *or* $26 \cdot \binom{26}{7}$.

Submission*: You must upload a **pdf** of your solutions to Gradescope under "Pset 6 [Written]". The use of LaTex is highly recommended, and we have provided a template. (Note that if you want to hand-write your solutions, you'll need to scan them. If we cannot make out your writing, your work may be ungradable, so make sure it is legible.) There is no coding for this homework.*

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, or you will potentially receive **no credit** as mentioned in the syllabus. Please put each numbered problem on its own page in 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). As stated above, the coding problem will also be submitted to Gradescope.*

Collaboration*: This pset must be submitted **individually**. You are welcome and encouraged to discuss approaches with your fellow students, but everyone **must write up their own solutions.** Failure to do so is an instance of academic dishonesty.*

For this homework, there is no need to use continuity correction for CLT problems. You may have practiced this in section, but don't worry about it here.

1. Knitting Requires Concentration (20 points)

Aleks is slowly knitting a blanket, made of 100 squares. It takes an average of 1 hour for Aleks to knit a square, with a standard deviation of 0.4 hours. The time to knit each square is independent. (You should treat time as continuous for this problem.)

(a) (3 Points) What is the expectation of the total time to knit the blanket?

(b) (3 Points) What is the variance of the total time to knit the blanket?

(c) (7 Points) Aleks will have 150 hours to knit between now and when he needs the blanket to be finished to stay warm at a football game. Use Markov's Inequality to give a *lower bound* on the probability that Aleks finishes the blanket before the game.

(d) (7 Points) Can we improve the lower bound from c) using Chebyshev's inequality?

2. Normal, normal, normal (20 points)

Your answers should be correct to four decimal places.

(a) (10 Points) Apparently IQ is roughly normally distributed, with a mean of 100 and a standard deviation of about 15. What fraction of people would be classified as "genius" if that means IQ of 140 or above? (These numbers are made up.)

(b) (10 Points) The height of a group of people is approximately normally distributed with a mean of 69.2 inches (5' 9.2") and a standard deviation of 2.67 inches. Approximately what fraction of these people are 5'2" tall or less? If we form a basketball team by picking 5 people from this group at random, calculate the probability that at least one of them is over 6' 3" (75 inches) tall.

3. Centrally Limited (30 points)

Use the Central Limit Theorem to approximate the following probabilities, and **state explicitly at which step you invoke it**.

(a) (15 Points) The Internet fire marshal declared that if too many students show up to Zoom office hours, it would constitute a fire hazard. Consider an offering of this class in which there are 234 students registered in the class. On the day homework is due, each student comes to office hour with probability 1/6, independently of the other students. What is the approximate probability that between 37 and 41 people (inclusive) come to office hours? **Give your answer to 4 decimal places.**

(b) (15 Points) A fair 6-sided die is repeatedly rolled until the total sum of all the rolls reaches 180. What is the approximate probability that *at least* 49 rolls are necessary to reach a sum that reaches (or surpasses) 180? **Give your answer to 4 decimal places.** Hint: Try to come up with an equivalent statement where the number of die rolls is fixed and not random.

4. CLT in "real life" (30 points)

Use the Central Limit Theorem to approximate the following probabilities, and **state explicitly at which step you invoke it**.

(a) [15 Points] You've decided to try starting a company together with your 312 pset partner. After hours of brainstorming (in between solving 312 problems), you've cut your list of ideas down to 10, all of which you want to implement at the same time. An angel investor has agreed to back all 10 ideas, as long as your net return from implementing the ideas is positive with at least 95% probability.

Suppose that implementing an idea requires 50 thousand dollars, and your start-up then succeeds with probability p, generating 150 thousand dollars in revenue (for a net gain of 100 thousand dollars), or fails with probability $1 - p$ (for a net loss of 50 thousand dollars). The success of each idea is independent of every other. What is the condition on p that you need to satisfy in order to secure the angel investor funding? (To make your solution simpler, first figure out how many ideas need to succeed for the net return to be positive.)

(b) [15 Points] One of your start-ups uses error-correcting codes[1], which can recover the original message as long as at least 1000 packets are received (not erased). Each packet gets erased independently with probability 0.8. How many packets should you send such that you can recover the message with probability at least 99% You should treat the number of packets received as discrete.

[1]From the Wikipedia page: "In computing, telecommunication, information theory, and coding theory, an error correcting code (ECC) is used for controlling errors in data over unreliable or noisy communication channels. The central idea is the sender encodes the message with redundant information in the form of an ECC. The redundancy allows the receiver to detect a limited number of errors that may occur anywhere in the message, and often to correct these errors without retransmission."