Homework 7: Multiple RVs, Tail bounds, and MLE

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. For each problem, make sure to explicitly define all random variables you use, and be clear about how they are related to each other using proper notation (conditionals, summations, etc.).

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 6 [Written]". 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.)

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).

Due Date: This assignment is due at 11:59 PM Monday August 16 (Seattle time, i.e. GMT-7).

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. Statistics Books [24 points]

Alice is going shopping for statistics books for H hours, where H is a random variable, equally likely to be 1, 2 or 3. The number of books B she buys is random and depends on how long she is in the store for. We are told that

$$\Pr(B = b|H = h) = \frac{c}{h}, \qquad \text{for } b = 1, \ldots, h,$$

for some constant c.
Hint: Remember that law of total expectation states that

$$E[A] = \sum_b E[A|B = b]P(B = b)$$

(a) Compute c. You may want to use one of the axioms of probability.

(b) Find the joint distribution of B and H using the chain rule.

(c) Find the marginal distribution of B.

(d) Find the conditional distribution of H given that $B = 1$ (i.e., $Pr(H = h|B = 1)$ for each possible h in 1,2,3). Use the definition of conditional probability and the results from previous parts.

(e) Suppose that we are told that Alice bought either 1 or 2 books. Find the expected number of hours she shopped conditioned on this event. Use the law of total expectation and conditional probability theorems.

(f) The cost of each book is a random variable with mean 3, and is independent of the number of books Alice buys. What is the expected amount of money Alice spends?
Warning: you might be tempted to skip some steps and assert that expected amount of money spent is 3 times the expected cost of each book. Even though the answer is intuitive, its formal derivation is a lot more involved. We expect you to show each step you used to get to that expression. Your work should involve the law of total expectation conditioning on the number of books bought, and make use of indicator random variables.

2. Coal Mining [24 points]

A miner is trapped in a mine containing 4 doors, and each door is equally likely to be chosen. The first door leads to a tunnel that will take him to safety after a number of hours which is Poisson with parameter 2. The second door leads to a tunnel that will take him to safety after a number of hours which is Geometric with parameter $\frac{1}{5}$. The third door leads to a tunnel that will take him to safety after a number of hours which is binomial with parameters $n = 100$ and $p = 1/20$. The fourth door leads to a tunnel which brings him back to where he started after 2 hours. Use the law of total expectation to compute the expected number of hours until the miner reaches safety.

3. Driving Requires Concentration [16 points]

Pascal is driving a car to a soccer game, 50 miles away. It takes an average of 1 minute for Pascal to cover a mile, with a variance of 0.1 minutes. The time to cover each mile is independent.

You should treat time as continuous for this problem.

Assume no time is taken to park the car at the game.

(a) What is the expectation of the total time to reach the game? [2 points]

(b) What is the variance of the total time to reach the game? [2 points]

(c) Pascal will have 65 minutes to reach and park between now and the soccer game. Use Markov's Inequality to bound the probability that Pascal parks before the game. Hint: you'll need to take a complement. [6 points]

(d) If Pascal reaches at least 30 minutes before the game starts, he can get a prime parking spot. Use Chebyshev's inequality to bound the probability that Pascal reaches the game after the prime parking ends, but in time for the game. [6 points]

4. Urns [21 points]

You have 2500 urns, that you place into a 50-by-50 grid. You will throw **60,000** balls (independently) toward the grid of urns, with equal probability for the ball to go in each urn.

You hope that at the end of the process, each urn will have at least 4 balls inside. You want to upperbound the probability that you will fail to get at least 4 balls into every urn.

(a) Use a Chernoff bound from class to bound the probability that the urn in the lower-left of the grid does not have at least 4 balls inside. [10 points]

(b) Is the probability of the lower-left urn having less than 4 balls independent of the probability that the urn in the upper-right has less than 4 balls? Briefly explain (you may give a formal derivation/calculation as an explanation or an informal one). [3 points]

(c) Bound the probability that any urn has fewer than 4 balls. Give the best bound you can from your answers in (a) and (b). [8 points]

5. MLE-1 [18 points]

Suppose you have a density

$$f_X(x) = \begin{cases} \frac{-|x|}{\theta^2} + \frac{1}{\theta} & \text{if } -\theta \le x \le \theta \\ 0 & \text{otherwise} \end{cases}$$

Let $x_1, x_2, \ldots, x_n$ be independent draws of a random variable from this distribution for some θ.

For this problem, *starting with part (b)* all of your work must be understandable without reference to any calculator (including wolframalpha). You may check your answers using calculators, but your explanation may not rely on them.

(a) Graph this density for a few values of θ. describe in a sentence or two what the shape is. The main goal of this part is for you to get intuition about the problem, you don't need to include the graphs [2 points]

(b) Write down the likelihood function for this problem; be careful. You should probably come back to this question after doing parts c and d to make sure you've handled edge cases. [3 points]

(c) Suppose you look at a value of θ such that $|x_1| \ge \theta$. What does the likelihood become in this case? Why? [3 points]

(d) In order to finish the problem without brutal algebra, we're going to assume that we've gotten exactly two independent samples x_1, x_2 and, moreover, assume that $x_1 = -x_2$.[1] Write the likelihood under this extra assumption. [3 points]

(e) Find the maximum likelihood estimator $\hat{\theta}$ under our extra assumptions. In taking a derivative, be sure to think about your answer in (c). You may skip the step of using the second derivative test to verify that your critical point is a maximizer (but you still must consider the piecewise part of the definition). [7 points]

6. MLE-2 [15 points]

In the final round of a rap battle, every member of the audience partakes in helping decide the winner between 3 rap battle contestants: Anna, Robbie and Kushal. Suppose that independently each member of the audience votes for Anna with probability θ_A, for Robbie with probability θ_R and for Kushal with probability $1 - \theta_A - \theta_R$. (Thus, $0 \le \theta_A + \theta_R \le 1$.) The parameters θ_A, θ_R are unknown. Suppose that $x_1, \ldots, x_n$ are n independent, identically distributed samples from this distribution. (Let n_A = number of x_i s equal to Anna (votes for Anna), let n_R = number of x_i s equal to Robbie (votes for Robbie), and let n_K = number of x_i s equal to Kushal (votes for Kushal).) What are the maximum likelihood estimates for θ_A and θ_R in terms of n_A, n_R, and n_K? In doing this problem, you do not need to do a second-derivative test (or any other test) to confirm you have a maximizer. You may assume any critical point you find is a maximizer
For this problem, all of your work must be understandable without reference to any calculator (including wolframalpha). You may check your answers using calculators, but your explanation may not rely on them.

[1] This is not a normal assumption to make for an MLE calculation, but the algebra is much worse in the general case, and this very simplified case will still be good practice.