Instructions:

- Please submit your work to Gradescope by no later than the due date posted above.

- Be sure to show your work; correct answers with no supporting work will not be awarded full points.

- 2 randomly selected questions/parts will be graded, but you must still turn in your work for all problems in order to be eligible to earn full credit.

....

1. **The Multinomial Distribution.** Recall that the Binomial distribution arises in the context of tracking the number of successes across n independent Bernoulli(p) trials. Definitionally, then, we require a binary division; namely a well-defined notion of "success" and "failure." Oftentimes, in Statistical Modeling, this is too stringent of a restriction.

   Suppose our n independent trials each result in one of r outcomes; as a simple case, when $r = 3$, we might say that our outcomes are "success," "failure," and "neutral." Additionally, suppose that each trial results in outcome i with probability p_i, for $i = 1, \cdots, r$. Let X_i denote the number of outcomes of type i we see (again, for $i = 1, \cdots, r$); then the random vector $(X_1, \cdots, X_r)$ is said to follow the **Multinomial Distribution** with parameters n (total number of trials), r (number of possible outcomes on each trial), and $p_1, \cdots, p_r$ (the probability of each outcome). We denote this:

$$(X_1, \cdots, X_r) \sim \text{Multi}(n, r, p_1, \cdots, p_n)$$

   Over the next few parts, we will investigate the Multinomial distribution in greater detail.

   PART I: Deriving the Joint P.M.F.

   (a) Suppose that PSTAT 120A has 100 students and 4 Discussion Sections (we can call them Sections 1 through 4). Further suppose that section 1 must contain 15 students, section 2 must contain 35, Section 3 must contain 20, and Section 4 must contain 30. In how many ways can we divide the students among these 4 sections?

   (b) If n and r are positive integers, and $k_1, \cdots, k_r$ are nonnegative integers that sum to n (i.e. $k_1 + \cdots, k_r = n$), then the number of ways of assigning lables $1, 2, \cdots, r$ to n items so that, for each $i = 1, 2, \cdots, r$ exactly k_i items receive label i, is the **multinomial coefficient**

$$\binom{n}{k_1, k_2, \cdots, k_r} = \frac{n!}{(k_1!) \times (k_2!) \times \cdots \times (k_r)!}$$

   Rewrite your answer to part (a) using a multinomial coefficient.

   (c) Now, let's return to the Multinomial distribution. Find $p_{X_1, \cdots, X_r}(k_1, \cdots, k_r)$, the joint p.m.f. of $(X_1, \cdots, X_r)$. You may find it useful to revisit the methodology we used when deriving the p.m.f. of the Binomial distribution.

   (d) Speaking of the Binomial Distribution, show that the Multi($n, 2, p_1, p_2$) distribution is equivalent to the Binomial distribution.

   PART II: Using the Joint P.M.F. In all parts that follow, continue to take $(X_1, \cdots, X_r) \sim \text{Multi}(n, r, p_1, \cdots, p_r)$

(e) What is the marginal distribution of X_1? (No summations needed; just make an argument about exactly *what* X_1 measures.)

(f) Give an expression for $\text{Cov}(X_i, X_j)$, for $i, j = 1, \cdots, r$. **Hint:** There are two possible ways to solve this part.

(1) Consider the indicator defined by

$$
\mathbb{1}_{k,i} = \begin{cases} 1 & \text{if trial } k \text{ gives outcome } i \\ 0 & \text{if trial } I \text{ gives an outcome other than } i \end{cases}
$$

and express X_i as a suitable sum of these indicators.

(2) Alternatively, you can recognize the distribution of $(X_1 + X_2)$, compute its variance, and then use previously-derived results about variances of sums of random variables to obtain an equation involving $\text{Cov}(X_i, X_j)$ that you can solve for.

2. **Question 2 has been removed.**

3. Let X and Y be two continuous random variables with: $\mathbb{E}[X] = 6$, $\text{Var}(X) = 4$, $\mathbb{E}[Y] = 6$, $\text{Var}(Y) = 3$, and $\text{Cov}(X, Y) = -1$. Use Chebyshev's Inequality to provide a bound for $\mathbb{P}(9 \leq X + Y \leq 15)$; be sure to specify whether this bound is an *upper* or *lower* bound.