

Gov 2002: Problem Set 3

Fall 2023

Problem Set Instructions

This problem set is due on **September 27, 11:59 pm** Eastern time. Please upload a PDF of your solutions to Gradescope. We will accept hand-written solutions but we strongly advise you to typeset your answers in Rmarkdown. Please list the names of other students you worked with on this problem set.

Question 1 (15 points)

A common approximation for a Binomial random variable $X \sim \text{Bin}(n, p)$ is a Poisson random variable $Y \sim \text{Pois}(\lambda)$ where $\lambda = n \cdot p$. One of the reasons for this approximation is that the expectations of Binomial and Poisson random variables match (both np , in our case). We will explore this a bit further.

- Holding n fixed, would Y better approximate X if $p = \frac{1}{10}$ or if $p = \frac{1}{2}$? Compare the variances of X and Y by checking the difference and ratio of the two variances.
- Holding p fixed, would Y better approximate X if $n = 1,000$ or if $n = 100,000$? Compare the variances of X and Y by checking the difference and ratio of the two variances.

Question 2 (15 points)

Let n be the number of students in a certain graduate program, and c be the number of courses that the department will offer next semester. Suppose that each student chooses their course schedule by randomly choosing 4 courses to take in the department (with all sets of 4 courses equally likely), independent of other students' choices.

- Assume for this part that simultaneous enrollment is allowed (i.e., a student can enroll in two or more courses that have overlapping meeting times) and there are no enrollment caps (so students can enroll in whatever courses they want). Find the expected number of pairs of students such that the two students in the pair will take exactly the same set of courses next semester. (HINT: use the fundamental bridge and consider one pair of students to start.)
- Now suppose instead that simultaneous enrollment is *not* allowed (which explains the low workshop enrollment in said graduate program). Fortunately, there are still no enrollment caps. There are 8 different time slots in which a course can be offered (the

time slots are non-overlapping). Suppose that $c = 8k$ for some positive integer k , and that there are exactly k courses in each of the 8 time slots next semester. As before, each student chooses their courses randomly, but now their chosen schedule will not be allowed if there are time conflicts between the courses. Find the expected number of students who will be able to take their initially chosen set of courses next semester.

Question 3 (30 points)

Let X and Y be $\text{Pois}(\lambda)$ r.v.s, and $T = X + Y$.

- Using LOTUS, find $\mathbb{E}[2^X]$ and $\mathbb{E}[e^Y]$ (Hint: Recall that the Taylor series for e^x is $e^x = \sum_{k=0}^{\infty} x^k/k! = 1 + x + \frac{x^2}{2!} + \frac{x^3}{3!} + \frac{x^4}{4!} + \frac{x^5}{5!} + \dots$)
- Prove that if X and Y are independent, then $T \sim \text{Pois}(2\lambda)$ (Hint: derive the PMF of T from the PMFs of X and Y and compare this to the PMF of $\text{Pois}(2\lambda)$.)
- Formally or informally, show that if $X = Y$, then T cannot follow a $\text{Pois}(2\lambda)$ distribution.

Question 4 (20 points)

In many problems about modeling count data, it is found that values of zero in the data are far more common than can be explained well using a Poisson model (e.g., the onset of major conflicts between great powers). The Zero-Inflated Poisson distribution is a modification of the Poisson to address this issue, making it easier to handle frequent zero values gracefully. A Zero-Inflated Poisson r.v. X with parameters p and λ can be generated as follows:

First, flip a coin with probability of p of Heads. Given that the coin lands Heads, $X = 0$. Given that the coin lands Tails, X is distributed $\text{Pois}(\lambda)$. Note that if $X = 0$ occurs, there are two possible explanations: the coin could have landed Heads (in which case the zero is called a structural zero), or the coin could have landed Tails but the Poisson r.v. turned out to be zero anyway. For example, if X is the number of atom bomb a random country detonated in another country within a year, then $X = 0$ for all non-nuclear powers (this is a structural zero), but a nuclear power could still have $X = 0$ occur.

- Find the PMF of a Zero-Inflated Poisson r.v. X .
- Explain why X has the same distribution as $(1 - I)Y$, where $I \sim \text{Bern}(p)$, $Y \sim \text{Pois}(\lambda)$, and $I \perp\!\!\!\perp Y$.
- Find $\mathbb{E}[X]$ (Hint: You can either derive this using the definition of expectation, or the result from part (b))
- Find $\mathbb{V}[X]$ (Hint: You can either derive this using the definition of variation, or the result from part (b). Also, you can assume that if random variables A and B are independent, then A^2 and B^2 are also independent.)

Question 5 (20 points)

You are a manager of a football team in your town. Each season has 10 games, and assume that each season and each game is played independently. The probability of winning one game is $p \in (0, 1)$.

- (a) What is the probability that your team wins 8 out of 10 games in one season?
- (b) What is the probability that your team wins at least 8 out of 10 games in one season?
- (c) There are $n = 20$ upcoming seasons. What is the expected number of *seasons* of your team winning at least 8 games?
- (d) Finally, let's practice R coding given the result in (c). Assume that $p = 0.8$ and calculate the expectation in (c) numerically, using the `dbinom` function.

Hint for (d): The `dbinom` function gives you the PMF of Binomial. For example, if you want to calculate $P(X = 30)$, where $X \sim \text{Bin}(n = 100, p = 0.3)$, then you can code it by typing: `dbinom(x = 30, size = 100, prob = 0.3)`.