1. **[True/False]** For each of the questions, answer TRUE or FALSE. [No need to justify.]

(a) If $X \sim \text{Geo}(p)$, then $\mathbb{E}[X + m | X > n] = m + n + \mathbb{E}[X]$.

(b) If $X \sim \text{Geo}(p)$, then for $i \geq j$, $\mathbb{P}[X > i | X > j] = (1 - p)^{i-j}$

**Solution:**

(a) True. It is $m + \mathbb{E}[X | X > n] = m + n + \mathbb{E}[X]$ since the geometric distribution is memoryless.

(b) True. We have

$$\mathbb{P}[X > i | X > j] = \frac{\mathbb{P}[X > i, X > j]}{\mathbb{P}[X > j]}.$$ 

Since $i \geq j$, $\mathbb{P}[X > i, X > j] = \mathbb{P}[X > i]$. So

$$\mathbb{P}[X > i | X > j] = \frac{\mathbb{P}[X > i]}{\mathbb{P}[X > j]} = \frac{(1-p)^i}{(1-p)^j} = (1-p)^{i-j}.$$ 

2. **[Short Answer]** What distribution would best model each of the following scenarios? Choose from Binomial, Poisson, Geometric. No justification needed.

(a) Number of taxis passing the corner of Euclid Ave and Hearst Ave between 5 pm and 6 pm on a weekday.

(b) Number of customers who purchase a lottery ticket before someone hits the jackpot.

(c) Number of girls in a family with 6 kids.

**Solution:**

(a) Poisson. We can model this with a Poisson random variable $X$ with $\lambda$ as the average number of taxis passing during an hour.

(b) Geometric. We can model this with a Geometric random variable $X$ with $p$ as the probability of hitting the jackpot by a lottery ticket.

(c) Binomial. We can model this with a binomial distribution $\text{Bin}(6, p)$, where $p$ is the probability of a child being a girl.
3. **[Long Answer]** CS70 teaching staff are trying to infer how many TAs would need to be present at the final exam to be able to answer questions students might have on the exam. Each TA recorded the statistics of how often students enrolled in their sections ask questions during the last midterm exams. You can assume the students ask questions independently from each other, and the number of questions students ask follows Poisson distribution.

- Edward’s students ask $N_1$ number of questions per every 60 minutes on average.
- Yimings’s students ask $N_2$ number of questions per every 90 minutes on average.
- Jonathan’s students ask $N_3$ number of questions per every 30 minutes on average.

Note that there are no other students outside of these three sections.

(a) What is the probability that the students ask at least two questions throughout three hour (180 min) final exam?

(b) If we let $Z$ be the number of questions asked by students from one of the three sections with the least number of questions, what is the probability that $Z \geq 2$?

**Solution:**

(a) Let the poisson parameters $\lambda_1$, $\lambda_2$, $\lambda_3$ denote the average number of questions from Edward, Yiming, and Jonathan’s section during the three hour final exam. Then $\lambda_1 = 3N_1$, $\lambda_2 = 2N_2$, $\lambda_3 = 6N_3$. Also let the random variables $X_1, X_2, X_3$ denote the number of questions from the three groups during the final exam. We know that

$$X_i \sim \text{Poisson}(\lambda_i).$$

The question is asking us to compute $\mathbb{P}[X_1 + X_2 + X_3 \geq 2]$.

$$\mathbb{P}[X_1 + X_2 + X_3 \geq 2] = 1 - \mathbb{P}[X_1 + X_2 + X_3 = 0] - \mathbb{P}[X_1 + X_2 + X_3 = 1]$$

$$= 1 - \mathbb{P}[X_1 = 0, X_2 = 0, X_3 = 0] - \mathbb{P}[X_1 = 1, X_2 = 0, X_3 = 0]$$

$$- \mathbb{P}[X_1 = 0, X_2 = 1, X_3 = 0] - \mathbb{P}[X_1 = 0, X_2 = 0, X_3 = 1].$$

From independence,

$$\mathbb{P}[X_1 + X_2 + X_3 \geq 2] = 1 - e^{-\lambda_1}e^{-\lambda_2}e^{-\lambda_3} - \lambda_1e^{-\lambda_1}e^{-\lambda_2}e^{-\lambda_3} - \lambda_2e^{-\lambda_1}e^{-\lambda_2}e^{-\lambda_3} - \lambda_3e^{-\lambda_1}e^{-\lambda_2}e^{-\lambda_3}$$

$$= 1 - e^{-\lambda_1}e^{-\lambda_2}e^{-\lambda_3}(1 + \lambda_1 + \lambda_2 + \lambda_3)$$

$$= 1 - e^{-(\lambda_1 + \lambda_2 + \lambda_3)}(1 + \lambda_1 + \lambda_2 + \lambda_3)$$

$$= 1 - e^{-(3N_1 + 2N_2 + 6N_3)}(1 + 3N_1 + 2N_2 + 6N_3).$$

**Alternatively:** We know that the summation of Poisson random variables is a Poisson random variable. Hence, $Y = X_1 + X_2 + X_3 \sim \text{Poisson}(\lambda_1 + \lambda_2 + \lambda_3)$.

$$\mathbb{P}[Y = y] = \frac{(\lambda_1 + \lambda_2 + \lambda_3)^y}{y!}e^{-(\lambda_1 + \lambda_2 + \lambda_3)}.$$
The question is asking us to compute $\mathbb{P}[Y \geq 2]$.

\[
\mathbb{P}[Y \geq 2] = 1 - \mathbb{P}[Y = 0] - \mathbb{P}[Y = 1]
\]
\[
= 1 - \frac{(\lambda_1 + \lambda_2 + \lambda_3)^0}{0!} e^{-(\lambda_1 + \lambda_2 + \lambda_3)} - \frac{(\lambda_1 + \lambda_2 + \lambda_3)^1}{1!} e^{-(\lambda_1 + \lambda_2 + \lambda_3)}
\]
\[
= 1 - e^{-(\lambda_1 + \lambda_2 + \lambda_3)}(1 + \lambda_1 + \lambda_2 + \lambda_3)
\]
\[
= 1 - e^{-(3N_1 + 2N_2 + 6N_3)}(1 + 3N_1 + 2N_2 + 6N_3).
\]

(b) Since $Z$ is the least number of questions among all sections, $Z \sim \min(X_1, X_2, X_3)$. The question is then just asking for $\mathbb{P}[Z \geq 2]$. Again from independence,

\[
\mathbb{P}[Z \geq 2] = \mathbb{P}[\min(X_1, X_2, X_3) \geq 2]
\]
\[
= \mathbb{P}[X_1 \geq 2 \text{ and } X_2 \geq 2 \text{ and } X_3 \geq 2]
\]
\[
= \mathbb{P}[X_1 \geq 2] \mathbb{P}[X_2 \geq 2] \mathbb{P}[X_3 \geq 2]
\]
\[
= (1 - e^{-\lambda_1 - \lambda_1 e^{-\lambda_1}})(1 - e^{-\lambda_2 - \lambda_2 e^{-\lambda_2}})(1 - e^{-\lambda_3 - \lambda_3 e^{-\lambda_3}})
\]
\[
= (1 - e^{-3N_1 - 3N_1 e^{-3N_1}})(1 - e^{-2N_2 - 2N_2 e^{-2N_2}})(1 - e^{-6N_3 - 6N_3 e^{-6N_3}})
\]