Discrete Distributions Sagnik

1 What is the Distribution?

What distribution would best model each of the following scenarios? Choose from Binomial, Poisson, Geometric. No justification needed.

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

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

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

Solution:

(a) Poisson. We can model this with a Poisson random variable X with λ 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.

2 Poisson Quiz

In a semester when we got to meet in person, CS70 teaching staff were trying to infer how many GSIs would need to be present at the final exam to be able to answer questions students might have on the exam. Each GSI 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.

- Sagnick's students ask N_1 number of questions per every 60 minutes on average.
- Agnibho's students ask N_2 number of questions per every 90 minutes on average.
- Shahzar'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 λ_1, λ_2, λ_3 denote the average number of questions from Hyun Oh, Victor and Dapo'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 $\mathbb{P}[X_1 + X_2 + X_3 \geq 2]$.

$$\begin{aligned}
\mathbb{P}[X_1 + X_2 + X_3 \geq 2] &= 1 - \mathbb{P}[X_1 + X_2 + X_3 = 0] - \mathbb{P}[X_1 + X_2 + X_3 = 1] \\
&= 1 - \mathbb{P}[X_1 = 0, X_2 = 0, X_3 = 0] - \mathbb{P}[X_1 = 1, X_2 = 0, X_3 = 0] \\
&\quad - \mathbb{P}[X_1 = 0, X_2 = 1, X_3 = 0] - \mathbb{P}[X_1 = 0, X_2 = 0, X_3 = 1].
\end{aligned}$$

From independence,

$$\begin{aligned}
\mathbb{P}[X_1 + X_2 + X_3 \geq 2] &= 1 - e^{-\lambda_1} e^{-\lambda_2} e^{-\lambda_3} - \lambda_1 e^{-\lambda_1} e^{-\lambda_2} e^{-\lambda_3} - \lambda_2 e^{-\lambda_1} e^{-\lambda_2} e^{-\lambda_3} - \lambda_3 e^{-\lambda_1} e^{-\lambda_2} e^{-\lambda_3} \\
&= 1 - e^{-\lambda_1} e^{-\lambda_2} e^{-\lambda_3} (1 + \lambda_1 + \lambda_2 + \lambda_3) \\
&= 1 - e^{-(\lambda_1 + \lambda_2 + \lambda_3)} (1 + \lambda_1 + \lambda_2 + \lambda_3) \\
&= 1 - e^{-(3N_1 + 2N_2 + 6N_3)} (1 + 3N_1 + 2N_2 + 6N_3).
\end{aligned}$$

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

$$\mathbb{P}[Y = y] = \frac{(\lambda_1 + \lambda_2 + \lambda_3)^y}{y!} e^{-(\lambda_1 + \lambda_2 + \lambda_3)}.$$

The question is asking us to compute $\mathbb{P}[Y \geq 2]$.

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

(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 $\mathbb{P}[Z >= 2]$. Again from independence,

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

3 A roll of the dice

Consider a single roll of two dice, one red and one blue.

1. Let R be the value of the red die. What is the distribution of R? What is the expectation of R?
2. Let M be the maximum of the numbers on the two dice. What is the distribution of M? What is the expectation of M?
3. How do the distribution and expectation of M compare to that of R?

Solution:

(a) Assuming the dice are fair dice, there are 6 possible outcomes of the red die roll, and they are all equally likely. Therefore

$$
\begin{aligned}
Pr[R = 1] &= 1/6 \\
Pr[R = 2] &= 1/6 \\
Pr[R = 3] &= 1/6 \\
Pr[R = 4] &= 1/6 \\
Pr[R = 5] &= 1/6 \\
Pr[R = 6] &= 1/6,
\end{aligned}
$$

and thus

$$\mathbb{E}(R) = \sum_{i=1}^{6} i \times Pr[R = i] = \sum_{i=1}^{6} i \times \frac{1}{6} = \frac{1}{6}\sum_{i=1}^{6} i = \frac{21}{6} = \frac{7}{2} = 3.5.$$

(b)
- 1 way to get maximum of 1: (1,1)
- 3 ways to get maximum of 2: (1,2),(2,1),(2,2)
- 5 ways to get maximum of 3: (1,3),(2,3),(3,1),(3,2),(3,3)
- 7 ways to get maximum of 4: (1,4),(2,4),(3,4),(4,1),(4,2),(4,3),(4,4)
- 9 ways to get maximum of 5: (1,5),(2,5),(3,5),(4,5),(5,1),(5,2),(5,3),(5,4),(5,5)

- 11 ways to get maximum of 6: (1,6),(2,6),(3,6),(4,6),(5,6),(6,1),(6,2),(6,3),(6,4),(6,5),(6,6)

Therefore, since the sample space of possible outcomes of two dice rolls is 36, we have

$$\begin{aligned}
Pr[M=1] &= 1/36 \\
Pr[M=2] &= 3/36 \\
Pr[M=3] &= 5/36 \\
Pr[M=4] &= 7/36 \\
Pr[M=5] &= 9/36 \\
Pr[M=6] &= 11/36,
\end{aligned}$$

and therefore

$$\mathbb{E}(M) = \sum_{i=1}^{6} i \times Pr[M=i] = \frac{161}{36} \approx 4.47.$$

(c) The distribution of M seems to have more mass closer to 6, and hence has a higher expected value than the distribution of R.

4 Dice Distributions

A fair die with k faces ($k \geq 2$), numbered $1,\ldots,k$, is rolled n times, with each roll being independent of all other rolls. Let X_i be a random variable for the number of times the ith face shows up. For each of the following parts, your answers should be in terms of n and k.

(a) What is the size of the sample space? (How many possible outcomes are there?)

(b) What is the distribution of X_i, for $1 \leq i \leq k$?

(c) What is the joint distribution of $X_1, X_2, \ldots, X_k$?

(d) Are X_1 and X_2 independent random variables?

Solution:

(a) For each of the n rolls, the die could land on anything from 1 to k, so by the first counting principle, there are k^n possible outcomes.

(b) $X_i \sim Bin(n, \frac{1}{k})$. For each of the n independent rolls, there is a $\frac{1}{k}$ probability the ith face shows up, and a $1-\frac{1}{k}$ probability a different face shows up.

(c) If $x_1 + x_2 + \ldots + x_k = n$, then we can use a counting argument to show that $\mathbb{P}(X_1 = x_1, X_2 = x_2, \ldots, X_k = x_k) = \frac{n!}{x_1!x_2!\ldots x_k! \cdot k^n}$. Otherwise, $\mathbb{P}(X_1 = x_1, X_2 = x_2, \ldots, X_k = x_k) = 0$.

(d) No, they are dependent random variables. For example, $\mathbb{P}(X_1 = n, X_2 = n) = 0$, but $\mathbb{P}(X_1 = n) \cdot \mathbb{P}(X_2 = n) = \left(\frac{1}{k}\right)^n \left(\frac{1}{k}\right)^n \neq 0$.