MATH 318 Homework 3

Joshua Lim (26928218)

February 3, 2023

Problem 1

Part A

If a team has to win 4 games in a series, then there will have to be 4-7 games played in the series with the last game being a team's 4th win. We can calculate the probability of each as follows:

Number of Games	Scores	Probability
4	4-0 0-4	$\text{p}^4 + (1-p)^4$
5	4-1 1-4	$\binom{4}{3}$ $(\text{p}^4(1-p) + p(1-p)^4)$
6	4-2 2-4	$\binom{5}{3}$ $(\text{p}^4(1-p)^2 + p^2(1-p)^4)$
7	4-3 3-4	$\binom{6}{3}$ $(\text{p}^4(1-p)^3 + p^3(1-p)^4)$

Hence, the probability mass function is:

$$p(x=a) = \begin{cases} \binom{x-1}{3}((p^4(1-p)^x) + (1-p)^4 p^x)) & x \in 4,5,6,7 \\ 0 & \text{otherwise} \end{cases} \tag{1}$$

Part B

$$P(A_{win}|X=4) = \frac{P(A_{win} \cap 4)}{P(4)} = \frac{p^4}{p^4 + (1-p)^4}$$

Part C

$$P(A_{win}|X=7) = \frac{P(A_{win} \cap 7)}{P(7)} = \frac{\binom{6}{3}}{\binom{6}{3}} \frac{p^4(1-p)^3}{p^4(1-p)^3 + p^3(1-p)^4} = \frac{1}{1+\frac{1}{p}(1-p)} = p$$

Problem 2

Part A

We consider the 4 possible cases (there are 2 cases in 2 distinct results but they share the same outcome). We count how many elements in the sample space are part of the subspace of each outcome and divide by the total, 6^4 for the four dice.

Distinct Results	Count in Sample Space	Probability
1	$\binom{6}{1}\binom{4}{4}$	$\frac{1}{216}$
2	$\binom{6}{1}\binom{4}{3}\binom{5}{1}\binom{4}{1}+\binom{6}{2}\binom{4}{2}\binom{2}{2}$	$\frac{35}{216}$
3	$\binom{6}{1}\binom{4}{2}\binom{5}{2}\binom{2}{1}\binom{1}{1}$	$\frac{5}{9}$
4	$\binom{6}{4}\binom{4}{1}\binom{3}{1}\binom{2}{1}\binom{1}{1}$	$\frac{5}{18}$

Hence, the probability mass function is:

$$p(x=a)=\begin{cases}\frac{1}{216} & x=1\\ \frac{35}{216} & x=2\\ \frac{5}{9} & x=3\\ \frac{5}{18} & x=4\\ 0 & \text{otherwise}\end{cases} \tag{2}$$

Additionally, the expectation of Y is:

$$<x>=\sum_i x\cdot p(x)=1\cdot\frac{1}{216}+2\cdot\frac{35}{216}+3\cdot\frac{5}{9}+4\cdot\frac{5}{18}$$
$$<x>=\frac{1}{216}+\frac{70}{216}+\frac{15}{9}+\frac{20}{18}=\frac{671}{216}\approx\mathbf{3.106}$$

Part B

It is easiest to find the cumulative probability distribution function, F(x), and then the probability mass function. To find the cumulative probability distribution function, we consider dice rolls with a lower bound restriction:

$$P(x_i\geq n)=(\frac{6-n+1}{6})^4\Rightarrow P(x=n)=\frac{(6-n+1)^4-(6-n)^4}{6^4}$$

Minimum Value	Probability
1	$\frac{671}{1296}$
2	$\frac{369}{1296}$
3	$\frac{175}{1296}$
4	$\frac{65}{1296}$
5	$\frac{15}{1296}$
6	$\frac{1}{1296}$
other	0

$$< x >= \sum_i x \cdot p(x) = \frac{1}{1296}(671 + 2 \cdot 369 + 3 \cdot 175 + 4 \cdot 65 + 5 \cdot 15 + 6 \cdot 1) = \frac{2275}{1296} \approx \mathbf{1.756}$$

Problem 3

Part A

At time t=mδ, total number of trials is $\frac{t}{\delta} = \frac{m\delta}{\delta} = m$.
By CDF of geometric distribution, the probability of no success until m trials is:

$$P(Y > m) = (1 - p)^m$$

Part B

Proof:

$$P(Y > m) = (1 - p)^m$$
$$\lambda = \frac{t}{\delta} \Rightarrow (1 - p)^m = (1 - p)^{\frac{t}{\delta}}$$
$$n = \frac{1}{-\lambda\delta} \Rightarrow (1 - \lambda\delta)^m = (1 - \frac{1}{n})^{-\lambda n t}$$
$$\lim_{\delta \to 0}(1 + \frac{1}{n})^{-\lambda n t} = \lim_{n \to \infty}((1 + \frac{1}{n})^n)^{-\lambda t} = e^{-\lambda t}$$
$$P(Y > m) = e^{-\lambda t}$$

Part C

We can reverse the CDF of the exponential distribution to get the PDF via FTC:

$$P(Y \leq t) = 1 - P(Y > t) = 1 - e^{-\lambda t}$$

$$1 - P(Y > t) = \int_0^t f(x)\,dx \; = 1 - e^{-\lambda t}$$

$$\frac{d}{dt}\int_0^t f(x)\,dx \; = \frac{d}{dt}(1 - e^{-\lambda t}$$

$$f(t) = \lambda e^{-\lambda t}$$

At last, we can conclude that the limiting case for the time of the first success is an exponential random variable with parameter λ.

Problem 4

One method we can calculate the probability that the message is being received correctly is using Bayes' Theorem. We can note that there is symmetry for the two cases (as if they represent two Gaussian distributions) and Z = $\frac{\mu - x}{\sigma} = \pm 2.5$. This means that the probability that whether a 0 or 1 is found below $\frac{1}{2}$ is 0.9938 and 0.0062 respectively.

$$P(0|x < \frac{1}{2}) = \frac{P(x < \frac{1}{2}|0)P(0)}{P(x < \frac{1}{2}|0)P(0) + P(x < \frac{1}{2}|0)P(0)} = \frac{0.9938 \cdot \frac{1}{2}}{0.9938 \cdot \frac{1}{2} + 0.0062 \cdot \frac{1}{2}} = \mathbf{0.9938}$$

Problem 5

The expectation value for a Poisson distribution is λ. The parameter predicts the average number of events in a given time frame. Over an infinite number of samples, this will amount to the actual expectation value. Since the average is unknown to us, we can only match the expectation value of the Poisson distribution to the expectation value of our unknown random variable.

Part A

Based on the given information that there were two murders in the first week, we can expect that $\lambda = 2$ because that is the expectation value of our limited sample space.

Part B

Since we now know that the second week had one murder, we can update our expectation value with the average of the results: $\lambda = \frac{2+1}{2} = 1.5$.

Part C

As we do more trials, we can expect that the expectation value will be $\lambda = \frac{\sum_k a_k}{k}$.

Problem 6

Part A

```
from math import comb

n = 430
p = 49/50

prob = sum(comb(n, k) * p**k * (1-p)**(n-k) for k in range(421, 431))

print(prob)
```

There is a **0.6405** probability that more than 420 passengers will show up.

Part B

We use the binomial approximation to the Poisson distribution. We have n = 430 trials and p = 0.02. Then $\lambda = np = \frac{430}{50} = 8.6$ and our random variable is approximated:

$$X = Bin(n, p) \approx Poisson(np) \Rightarrow P(X = k) = e^{-\lambda}\frac{\lambda^k}{k!}$$

There is also a **0.6405** probability that more than 420 passengers will show up.

Part C

```
import scipy.stats as stats
import numpy as np
import math

p = 1/50
n = 430
k = 9

samples = 50000
no_shows = np.random.binomial(n, p, samples)

# Plot histogram of number of no-shows
```

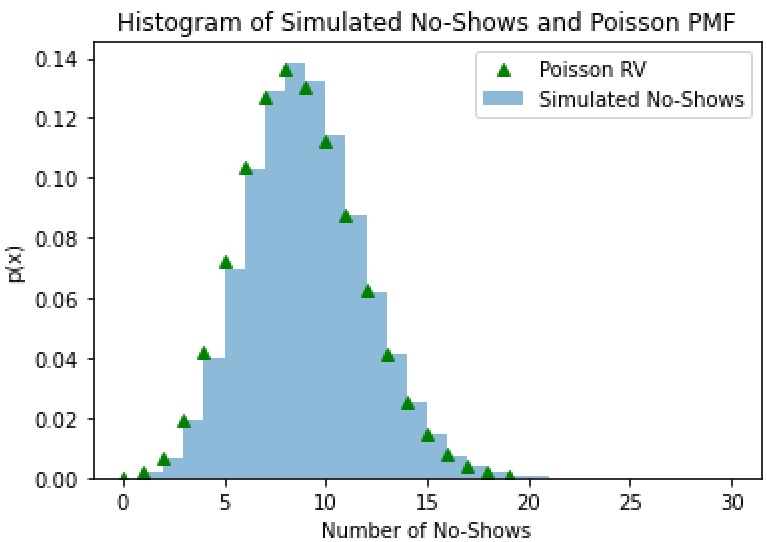
```
plt.hist(no_shows, bins=np.arange(0, 31), density=True, alpha=0.5, label='Simulated No-Shows')

# Plot Poisson PMF with the same parameters
count = 20
lam = n*p
pdf_pois = np.zeros(count)

for i in range(0, count):
    val = (lam**i)*math.exp(-lam)/math.factorial(i)
    pdf_pois[i]=val

cdf_pois = np.cumsum(pdf_pois)

plt.plot(np.array(range(0, count)), pdf_pois, 'g^', label='Poisson RV')
plt.xlabel('Number of No-Shows')
plt.ylabel('p(x)')
plt.title('Histogram of Simulated No-Shows and Poisson PMF')
plt.legend()
plt.show()
```



Part D

As n gets large, the average proportion of overbooked flights denoted by X_n/n, begins to asymptotically approach the actual expectation value of the Binomial distribution CDF, which is approximately 0.6405.

```
import scipy.stats as stats
import numpy as np
import math
```

```
p = 1/50
n = 430
k = 9

samples = 50000
no_shows = np.random.binomial(n, p, samples)

overbooked = no_shows < 10

trials = np.array(range(1,samples+1))
running_average = np.cumsum(overbooked)/trials

plt.plot(trials, running_average, 'b-', label='Running Average')

# Add a assymptotic line at expectation value
x_vals = [0, samples]
y_vals = [cdf[n-421], cdf[n-421]]
plt.plot(x_vals, y_vals, 'g--', label='Expectation Value')

# Add a legend and show the plot
plt.legend()
plt.show()
```