Final Exam

Andrea Hannah
Mathematics and Statistics Theory

December 7, 2024

Question 1 (15pts)

An experiment consists of throwing a biased coin three (3) times. The coin has a probability of 0.6 of landing heads and 0.4 of landing tails. Find the frequency function (PMF) and the cumulative distribution function (CDF) of the following random variables:

1. The number of heads before the first tail (if there is no tail in three throws, you can assume X = 3).

Let X be the number of heads before the first tail. The possible values of X are 0, 1, 2, and 3 (if there is no tail in three throws).

PMF of X

$$P(X = x) = \begin{cases} 0.4 & \text{if } x = 0 \\ 0.24 & \text{if } x = 1 \\ 0.144 & \text{if } x = 2 \\ 0.216 & \text{if } x = 3 \end{cases}$$

CDF of X

$$F_X(x) = P(X \leq x) = \begin{cases} 0 & \text{if } x < 0 \\ 0.4 & \text{if } 0 \leq x < 1 \\ 0.64 & \text{if } 1 \leq x < 2 \\ 0.784 & \text{if } 2 \leq x < 3 \\ 1 & \text{if } x \geq 3 \end{cases}$$

Possible values of X:

- $X = 0$: Tails on the first throw.
- $X = 1$: Heads on the first throw, tails on the second.
- $X = 2$: Heads on the first two throws, tails on the third.

- $X = 3$: Heads on all three throws.

2. The number of heads following the first tail (if there is no tail in three throws, you can assume the Y = 0)

Let Y be the number of heads following the first tail. The possible values of Y are 0, 1, and 2.

PMF of Y

$$P(Y = y) = \begin{cases} 0.064 & \text{if } y = 0 \\ 0.24 & \text{if } y = 1 \\ 0.144 & \text{if } y = 2 \end{cases}$$

CDF of Y

$$F_Y(y) = P(Y \leq y) = \begin{cases} 0 & \text{if } y < 0 \\ 0.064 & \text{if } 0 \leq y < 1 \\ 0.304 & \text{if } 1 \leq y < 2 \\ 1 & \text{if } y \geq 2 \end{cases}$$

Possible values of Y:

- $Y = 0$: No tails in three throws or tails on the last throw.
- $Y = 1$: One head after the first tail.
- $Y = 2$: Two heads after the first tail.

3. Number of Heads Minus the Number of Tails (Z).

Let Z be the number of heads minus the number of tails. The possible values of Z are -3, -1, 1, and 3.

PMF of Z

$$P(Z = z) = \begin{cases} 0.064 & \text{if } z = -3 \\ 0.432 & \text{if } z = -1 \\ 0.432 & \text{if } z = 1 \\ 0.216 & \text{if } z = 3 \end{cases}$$

CDF of Z

$$F_Z(z) = P(Z \leq z) = \begin{cases} 0 & \text{if } z < -3 \\ 0.064 & \text{if } -3 \leq z < -1 \\ 0.496 & \text{if } -1 \leq z < 1 \\ 0.928 & \text{if } 1 \leq z < 3 \\ 1 & \text{if } z \geq 3 \end{cases}$$

Possible values of Z:

- $Z = -3$: Three tails and no heads (TTT).
- $Z = -1$: More tails than heads (HTT, THT, TTH).
- $Z = 1$: More heads than tails (HHT, HTH, THH).
- $Z = 3$: All heads and no tails (HHH).

Question 2 (15 pts)

Suppose a factory produces $n = 10000$ items, and the probability that any given item is defective is $p = 0.0005$. What is the probability that:

1. Exactly 3 defective items are produced?

Formulas: Using Binomial Distribution:

$$P(X = 3) = \binom{n}{3} p^3 (1-p)^{n-3}$$

Using Poisson Distribution:

$$P(X = 3) = \frac{\lambda^3 e^{-\lambda}}{3!}$$

where $\lambda = np$.

Calculation using Poisson Distribution:

$$\lambda = 10000 \times 0.0005 = 5$$

$$P(X = 3) = \frac{5^3 e^{-5}}{3!} = \frac{125e^{-5}}{6}$$

Here is the Python code:

```
import math

# Parameters
n = 10000
p = 0.0005
lambda_ = n * p
x = 3

# Poisson probability
P_X_3 = (lambda_ ** x * math.exp(-lambda_)) / math.factorial(x)
print(f"P(X = 3) = {P_X_3:.6f}")

#Output
P(X = 3) = 0.140374
```

2. At most 3 defective items are produced? For this part, you just need to use Poisson distribution to calculate the results through the summation of the Poisson probabilities for $k = 0, 1, 2$, and 3:

$$P(X \leq 3) = \sum_{k=0}^{3} \frac{\lambda^k e^{-\lambda}}{k!}$$

Where $\lambda = n \cdot p$.

For this scenario:

$$\lambda = 10000 \times 0.0005 = 5$$

$$P(X \leq 3) = \sum_{k=0}^{3} \frac{5^k e^{-5}}{k!}$$

Python Code

```
import math

# Parameters
n = 10000
p = 0.0005

# Poisson distribution parameter
lambda_ = n * p

# Function to calculate Poisson probability
def poisson_probability(k, lambda_):
    return (lambda_ ** k * math.exp(-lambda_)) / math.factorial(k)

# 1. Probability of exactly 3 defective items
P_X_3 = poisson_probability(3, lambda_)
print(f"Probability of exactly 3 defective items: {P_X_3}")

## Output
# Here are the results:
# Probability of exactly 3 defective items: 0.1403738958142805

# 2. Probability of at most 3 defective items
P_X_leq_3 = sum(poisson_probability(k, lambda_) for k in range(4))
print(f"Probability of at most 3 defective items: {P_X_leq_3}")

## Output
# Here are the results:
# Probability of at most 3 defective items: 0.26502591529736164
```

Question 3 (10 pts)

Let X be a normal random variable with mean $\mu = 3$ and standard deviation $\sigma = 5$. Find the following probabilities: 1. $P(X > 8)$:

1. $P(X > 8)$:

First, convert X to a standard normal variable Z:

$$Z = \frac{X - \mu}{\sigma} = \frac{X - 3}{5}$$

Then,

$$P(X > 8) = P\left(Z > \frac{8 - 3}{5}\right) = P(Z > 1)$$

Using the cumulative distribution function of the standard normal distribution,

$$P(Z > 1) = 1 - P(Z \leq 1) \approx 1 - 0.8413 = 0.1587$$

2. $P(-2 < X < 3)$:

We convert X to Z:

$$Z = \frac{X - \mu}{\sigma} = \frac{X - 3}{5}$$

Therefore,

$$P(-2 < X < 3) = P\left(\frac{-2 - 3}{5} < Z < \frac{3 - 3}{5}\right) = P(-1 < Z < 0)$$

Now with using the cumulative distribution function,

$$P(-1 < Z < 0) = P(Z < 0) - P(Z \leq -1) \approx 0.5 - 0.1587 = 0.3413$$

Question 4 (10 pts)

A company is producing batteries for electronic gadgets. The lifespans of these batteries (measured in hours) adhere to a Gamma distribution with shape parameter $k = 2$ and scale parameter $\theta = 10$ hours.

1. What is the average lifespan and the variance of the batteries?

The mean of a Gamma-distributed random variable is calculated as:

$$E[X] = k\theta = 2 \times 10 = 20 \, \text{hours}$$

The variance is provided by:

$$\text{Var}(X) = k\theta^2 = 2 \times 10^2 = 200 \, \text{hours}^2$$

2. Find the probability that a battery will exceed 24 hours in lifespan:

Using the cumulative distribution function (CDF) of the Gamma distribution, we can determine this probability. Here's the Python code to compute it:

Here is the Python code:

```
from scipy.stats import gamma

# Parameters
k = 2
theta = 10
x = 24

# Probability that a battery lasts more than 24 hours
probability_greater_24 = 1 - gamma.cdf(x, a=k, scale=theta)
print(f"P(X > 24) = {probability_greater_24:.4f}")

# Output
P(X > 24) = 0.2032
```

Question 5 (15 pts)

A bus arrives at a station every 20 minutes, and the waiting time for the bus is uniformly distributed between 0 and 20 minutes. A commuter records the waiting time for 25 buses. Use the Central Limit Theorem to:

1. Find the probabilities that the average waiting time for these 25 buses is less than 9 minutes.

Since the waiting time X is uniformly distributed between 0 and 20 minutes, we have:

$$E[X] = \frac{a+b}{2} = \frac{0+20}{2} = 10\,\text{minutes}$$

$$\text{Var}(X) = \frac{(b-a)^2}{12} = \frac{(20-0)^2}{12} = \frac{400}{12} = \frac{100}{3}\,\text{minutes}^2$$

For the average waiting time $\bar{X}$ for 25 buses,

$$E[\bar{X}] = E[X] = 10\,\text{minutes}$$

$$\text{Var}(\bar{X}) = \frac{\text{Var}(X)}{n} = \frac{\frac{100}{3}}{25} = \frac{4}{3}\,\text{minutes}^2$$

$$\text{SD}(\bar{X}) = \sqrt{\text{Var}(\bar{X})} = \sqrt{\frac{4}{3}} = \frac{2}{\sqrt{3}} \approx 1.1547\,\text{minutes}$$

Using the Central Limit Theory, the probability that the average waiting time is less than 9 minutes:

$$P(\bar{X} < 9) = P\left(Z < \frac{9-10}{\frac{2}{\sqrt{3}}}\right) = P\left(Z < \frac{-1}{1.1547}\right) = P(Z < -0.866)$$

Using the standard normal distribution table,

$$P(Z < -0.866) \approx 0.1931$$

2. Write a Python program to simulate the waiting time for 25 buses and calculate the probabilities. Note that this is a simulation program. You should use a random number generator to generate the waiting time for each bus and calculate the average waiting time for 25 buses and do the simulation for 10000 times.

Here is the Python code:

```
import numpy as np

# Parameters
num_buses = 25
num_simulations = 10000
lower_bound = 0
upper_bound = 20
threshold = 9

# Simulation
average_waiting_times = [ ]
for _ in range(num_simulations):
    waiting_times = np.random.uniform(lower_bound, upper_bound, num_buses)
    average_waiting_time = np.mean(waiting_times)
    average_waiting_times.append(average_waiting_time)

# Calculate probability
count_less_than_threshold = np.sum(np.array(average_waiting_times) < threshold)
probability = count_less_than_threshold / num_simulations

print(f"P(average waiting time < 9 minutes) = {probability:.4f}")

#Output
P(average waiting time < 9 minutes) = appx. 0.1864
```

Question 6 (15 pts)

If X is a normal random variable ($X \sim N(1,4)$) with mean $\mu = 1$ and variance $\sigma^2 = 4$, and $Y = 2X + 4$ is a linear transformation of X:

1. Find the mean and variance of Y:

The mean of Y:

$$E[Y] = E[2X + 4] = 2E[X] + 4 = 2(1) + 4 = 6$$

The variance of Y:

$$\mathrm{Var}(Y) = \mathrm{Var}(2X + 4) = 2^2\mathrm{Var}(X) = 4 \times 4 = 16$$

2. Find the probability density function (pdf) of Y:

Since Y is a linear transformation of a normally distributed variable X, Y is also normally distributed. The pdf of Y, $f_Y(y)$, is:

$$Y \sim N(6, 16)$$

$$f_Y(y) = \frac{1}{\sqrt{2\pi \cdot 16}} \exp\left(-\frac{(y-6)^2}{2 \cdot 16}\right) = \frac{1}{4\sqrt{2\pi}} \exp\left(-\frac{(y-6)^2}{32}\right)$$

3. Find the probability that Y is less than 6:

Standardize Y by using the standard normal distribution:

$$P(Y < 6) = P\left(Z < \frac{6-6}{\sqrt{16}}\right) = P(Z < 0)$$

Using the standard normal distribution table,

$$P(Z < 0) = 0.5$$

Question 7 (20 pts)

Consider two random variables X and Y with the following joint probability distribution:

$P(X,Y)$	$X=0$	$X=1$
$Y=0$	0.1	0.2
$Y=1$	0.3	0.4

1. Find the marginal distribution of X:

$$P(X = 0) = P(X = 0, Y = 0) + P(X = 0, Y = 1) = 0.1 + 0.3 = 0.4$$

$$P(X = 1) = P(X = 1, Y = 0) + P(X = 1, Y = 1) = 0.2 + 0.4 = 0.6$$

2. Find the entropy of X:

Entropy $H(X)$:

$$H(X) = -\sum_x P(X = x) \log_2 P(X = x)$$

$$H(X) = -[P(X = 0) \log_2 P(X = 0) + P(X = 1) \log_2 P(X = 1)]$$

$$H(X) = -[0.4 \log_2 0.4 + 0.6 \log_2 0.6]$$

$$H(X) \approx -[0.4(-1.3219) + 0.6(-0.7369)] \approx 0.97095$$

3. Compute the overall conditional entropy of Y given X:
The conditional entropy $H(Y|X)$ can be calculated using the formula:

$$H(Y|X) = \sum_x P(X = x)H(Y|X = x)$$

Firstly, find $H(Y|X = 0)$ and $H(Y|X = 1)$:

$$P(Y = 0|X = 0) = \frac{P(X = 0, Y = 0)}{P(X = 0)} = \frac{0.1}{0.4} = 0.25$$

$$P(Y = 1|X = 0) = \frac{P(X = 0, Y = 1)}{P(X = 0)} = \frac{0.3}{0.4} = 0.75$$

$$H(Y|X = 0) = -(0.25 \log_2 0.25 + 0.75 \log_2 0.75)$$

$$= -(0.25 \times -2 + 0.75 \times -0.415037) \approx 0.5 + 0.311278 = 0.811278$$

$$P(Y = 0|X = 1) = \frac{P(X = 1, Y = 0)}{P(X = 1)} = \frac{0.2}{0.6} = 0.3333$$

$$P(Y = 1|X = 1) = \frac{P(X = 1, Y = 1)}{P(X = 1)} = \frac{0.4}{0.6} = 0.6667$$

$$H(Y|X = 1) = -(0.3333 \log_2 0.3333 + 0.6667 \log_2 0.6667)$$

$$= -(0.3333 \times -1.585 + 0.6667 \times -0.584963) \approx 0.5287 + 0.389975 = 0.918675$$

Finally, we can calculate the overall conditional entropy:

$$H(Y|X) = P(X = 0)H(Y|X = 0) + P(X = 1)H(Y|X = 1)$$

$$= 0.4 \times 0.811278 + 0.6 \times 0.918675 \approx 0.324511 + 0.551205 = 0.875716$$

We see that the overall conditional entropy $H(Y|X)$ is approx. 0.876.

4. Mutual Information $I(X;Y)$ Between X and Y. You can calculate Mutual information $I(X;Y)$using the formula:

$$I(X;Y) = H(Y) - H(Y|X)$$

Firstly, calculate the entropy of Y:
Marginal distribution of Y:

$$P(Y = 0) = P(Y = 0, X = 0) + P(Y = 0, X = 1) = 0.1 + 0.2 = 0.3$$

$$P(Y = 1) = P(Y = 1, X = 0) + P(Y = 1, X = 1) = 0.3 + 0.4 = 0.7$$

$$H(Y) = -(0.3 \log_2 0.3 + 0.7 \log_2 0.7)$$

$$= -(0.3 \times -1.737 + 0.7 \times -0.514573) \approx 0.52105 + 0.3602 = 0.8813$$

Now, calculate the mutual information:

$$I(X;Y) = H(Y) - H(Y|X) \approx 0.8813 - 0.875716 = 0.005584$$

We see that the mutual information $I(X;Y)$ is appx. 0.0056.