FIT 2086 Assignment 1

Sia Yi Bin

33363129

Q1

1. Generative AI systems, the task involves generating a novel image of a gorilla riding a motorcycle, which aligns with the capabilities of generative AI systems to create imaginative and creative content.

2. Recommendation systems, recommendation systems are commonly used to suggest products or items to users based on their historical behavior, which aligns with the goal of understanding and predicting shopping preferences.

3. Forecasting, because this problem involves predicting future values based on historical data.

4. Risk predictions, risk prediction techniques are used to estimate the probability of future outcomes based on current and historical data.

Q2

1.

P(W=1) = P(W=1, H=0, P=0) + P(W=1, H=0, P=1) + P(W=1, H=1, P=0) + P(W=1, H=1, P=1)

= 0.235 + 0.117 + 0.058 + 0.178

= 0.588

2.

We need to calculate the denominator, P(P=0) first,

P(P=0) = 0.176 + 0.235 + 0.117 + 0.058 = 0.586

P(W=1 | P=0) = P(W=1, P=0)/P(P=0)

= (P(W=1, H=0, P=0) + P(W=1, H=1, P=0))/P(P=0)

= (0.235 + 0.058)/0.586

= 0.5

3.

We need to calculate the denominator, P(P=1) first,

P(P=1) = 0.06 + 0.117 + 0.059 + 0.178 = 0.414

P(W=1 | P=1) = P(W=1, P=1)/P(P=1)

= (P(W=1, H=0, P=1) + P(W=1, H=1, P=1))/P(P=1)

= (0.117 + 0.178)/0.414

= 0.713

4.

Yes, based on the data and the calculation above, there is a higher probability of the team winning a game if they won their previous game.

5.

To find the probability of win at least 1 game in the next 2 games given previous game has won, we can use 1 minus probability of not winning any of the 2 games.

$$P(Win \geq 1) = 1 - P(Win = 0)$$

$$= 1 - (P(W = 0 \mid H = 1, P = 1) * P(W = 0 \mid H = 0, P = 0))$$

$$= 1 - \frac{0.059}{0.059+0.178} * \frac{0.176}{0.176+0.235}$$

$$= 0.8934$$

The probability of winning at least 1 in the next 2 games given that previous game won is 0.8934.

<u>Q3</u>

1.

E[X_1] = $\sum X_1$ P(X_1)

= $\frac{1}{6}$ (1+2+3+4+5+6)

= 3.5

E[Y_1] = $\sum Y_1$ P(Y_1)

= $\frac{1}{4}$ (1+2+3+4)

= 2.5

E[S] = E[$2X_1 - Y_1$]

= 2E[X_1] - E[Y_1]

= 2(3.5) – 2.5

= 4.5

2.

V[S] = V[$2X_1 - Y_1$]

$= V\left[2{X_1}^2\right] - V[2X_1]^2 + V\left[{Y_1}^2\right] - V[Y_1]^2$

$= (2^2 + 4^2 + 6^2 + 8^2 + 10^2 + 12^2)\frac{1}{6} - \left[\frac{2+4+6+8+10+12}{6}\right]^2 + \frac{1^2+2^2+3^2+4^2}{4} - \left[\frac{1+2+3+4}{4}\right]^2$

$= \frac{155}{12}$

3.

Possible outcomes of S:

$S = -2$	$(X_1 = 1, Y_1 = 4)$
$S = -1$	$(X_1 = 1, Y_1 = 3)$
$S = 0$	$(X_1 = 1, Y_1 = 2), (X_1 = 2, Y_1 = 4)$
$S = 1$	$(X_1 = 1, Y_1 = 1), (X_1 = 2, Y_1 = 3)$
$S = 2$	$(X_1 = 2, Y_1 = 2), (X_1 = 3, Y_1 = 4)$
$S = 3$	$(X_1 = 2, Y_1 = 1), (X_1 = 3, Y_1 = 3)$
$S = 4$	$(X_1 = 3, Y_1 = 2), (X_1 = 4, Y_1 = 4)$
$S = 5$	$(X_1 = 3, Y_1 = 1), (X_1 = 4, Y_1 = 3)$
$S = 6$	$(X_1 = 4, Y_1 = 2), (X_1 = 5, Y_1 = 4)$
$S = 7$	$(X_1 = 4, Y_1 = 1), (X_1 = 5, Y_1 = 3)$
$S = 8$	$(X_1 = 5, Y_1 = 2), (X_1 = 6, Y_1 = 4)$
$S = 9$	$(X_1 = 5, Y_1 = 1), (X_1 = 6, Y_1 = 3)$
$S = 10$	$(X_1 = 6, Y_1 = 2)$
$S = 11$	$(X_1 = 6, Y_1 = 1)$

Probability distribution of S:

S		$P(X = S)$
-2	$\frac{1}{6} \times \frac{1}{4}$	$\frac{1}{24}$
-1	$\frac{1}{6} \times \frac{1}{4}$	$\frac{1}{24}$
0	$2(\frac{1}{6} \times \frac{1}{4})$	$\frac{1}{12}$
1	$2(\frac{1}{6} \times \frac{1}{4})$	$\frac{1}{12}$
2	$2(\frac{1}{6} \times \frac{1}{4})$	$\frac{1}{12}$
3	$2(\frac{1}{6} \times \frac{1}{4})$	$\frac{1}{12}$
4	$2(\frac{1}{6} \times \frac{1}{4})$	$\frac{1}{12}$
5	$2(\frac{1}{6} \times \frac{1}{4})$	$\frac{1}{12}$
6	$2(\frac{1}{6} \times \frac{1}{4})$	$\frac{1}{12}$
7	$2(\frac{1}{6} \times \frac{1}{4})$	$\frac{1}{12}$
8	$2(\frac{1}{6} \times \frac{1}{4})$	$\frac{1}{12}$
9	$2(\frac{1}{6} \times \frac{1}{4})$	$\frac{1}{12}$
10	$\frac{1}{6} \times \frac{1}{4}$	$\frac{1}{24}$
11	$\frac{1}{6} \times \frac{1}{4}$	$\frac{1}{24}$

4.

S^3	$P(X = S)$
-8	$\frac{1}{24}$
-1	$\frac{1}{24}$
0	$\frac{1}{12}$
1	$\frac{1}{12}$
8	$\frac{1}{12}$
27	$\frac{1}{12}$
64	$\frac{1}{12}$
125	$\frac{1}{12}$
216	$\frac{1}{12}$
343	$\frac{1}{12}$
512	$\frac{1}{12}$
729	$\frac{1}{12}$
1000	$\frac{1}{24}$
1331	$\frac{1}{24}$

E[S^3] = $-8\left(\frac{1}{24}\right) - \left(\frac{1}{24}\right) + 0 + \left(\frac{1}{12}\right) + 8\left(\frac{1}{12}\right) + 27\left(\frac{1}{12}\right) + 64\left(\frac{1}{12}\right) + 125\left(\frac{1}{12}\right) + 216\left(\frac{1}{12}\right) +$
$343\left(\frac{1}{12}\right) + 512\left(\frac{1}{12}\right) + 729\left(\frac{1}{12}\right) + 1000\left(\frac{1}{24}\right) + 1331\left(\frac{1}{24}\right)$

= 265.5

5.

$f(\mu) = \mu^3, f'(\mu) = 3\mu^2, f''(\mu) = 6\mu$

From $E[f(x)] \approx f(\mu) + \left(\frac{f''(\mu)}{2}\right)\sigma^2$,

$E[S^3]$ = $(E[S])^3 + (6E[S]/2)(V[S])$

= $(4.5)^3 + \left(\frac{6(4.5)}{2}\right)\left(\frac{155}{12}\right)$

= 265.5

6.

E$[Y_2]$ = $\sum Y_2$ P(Y_2)

= $\frac{1}{4}$ (1+2+3+4)

= 2.5

E$[(2X_1 - Y_1 + 2Y_2)^2]$

= $E[(2X_1)^2 \ + \ (Y_1)^2 + \ (2Y_2)^2 - \ 2\,(2X_1)(Y_1) - \ 2\,(2X_1)(2Y_2) + \ 2(Y_1)(2Y_2)]$

We have already calculated $E(X_1) = \ 3.5, E(Y_1) = \ 2.5, \ E(2Y_2) = \ 2 \ \times \ E(Y_2) = \ 2 \ \times 2.5 \ = \ 5.$

Now, we need to compute E$[(2X_1)(Y_1)]$, E$[(Y_1)(2Y_2)]$, E$[(2X_1)(2Y_2)]$, $E[X_1^2]$, $E[Y_1^2]$, $E[Y_2^2]$

E$[(2X_1)(Y_1)] = 2\mathrm{E}[(X_1)(Y_1)]$, E$[(2X_1)(2Y_2)]$= $4\mathrm{E}[(X_1)(Y_2)]$, E$[(Y_1)(2Y_2)]$ = 2E$[(Y_1)(Y_2)]$

Since X1 and Y1 are independent, the joint probability distribution is simply the product of their individual probabilities:

$$2\mathrm{E}[(X_1)(Y_1)] = 2(\frac{1}{24}(1+2+3+4+5+6))$$

$$= \ 2(\frac{21}{24})$$

$$= \frac{21}{12}$$

Since X1 and Y2 are independent, the joint probability distribution is simply the product of their individual probabilities:

$$4\mathrm{E}[(X_1)(Y_2)] = 4(\frac{1}{24}(1+2+3+4+5+6))$$

$$= \ 4(\frac{21}{24})$$

$$= \frac{21}{6}$$

Since Y1 and Y2 are independent, the joint probability distribution is also simply the product of their individual probabilities:

2E$[(Y_1)(Y_2)]$ = $2(\frac{1}{16}(1+2+3+4))$

= $2(\frac{10}{16})$

= $\frac{5}{4}$

$E[X_1^2]$ = $\frac{1}{6}(1+4+9+16+25+36)$

= $\frac{91}{6}$

$E[Y_1^2]$ = $\frac{1}{4}(1+4+9+16)$

= $\frac{15}{2}$

$E[Y_2^2] = \frac{1}{4}(1 + 4 + 9 + 16)$

$= \frac{15}{2}$

$E[(2X_1 - Y_1 + 2Y_2)^2]$

$= E[(2X_1)^2 + (Y_1)^2 + (2Y_2)^2 - 2(2X_1)(Y_1) - 2(2X_1)(2Y_2) + 2(Y_1)(2Y_2)]$

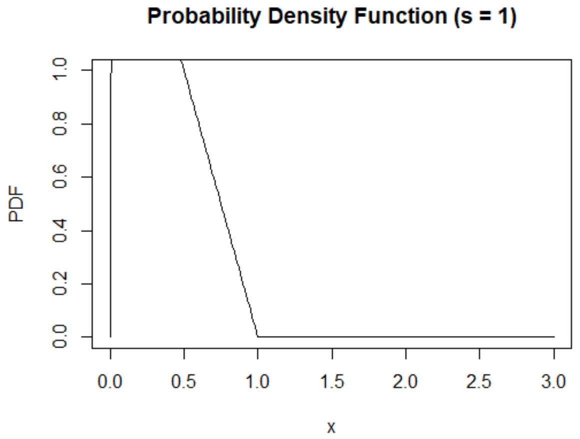
$= 2E[X_1^2] + E[Y_1^2] + 2E[Y_2^2] - 2E[(2X_1)(Y_1)] - 2E[(2X_1)(2Y_2)] + 2E[(Y_1)(2Y_2)]$

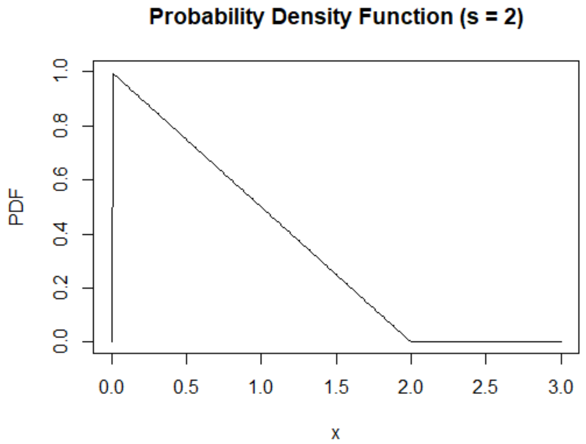
$= 2(\frac{91}{6}) + (\frac{15}{2}) + 2\left(\frac{15}{2}\right) - 2\left(\frac{21}{12}\right) - 2\left(\frac{21}{6}\right) + 2\left(\frac{5}{4}\right)$

$= \frac{269}{6}$

<u>Q4</u>

1.





2.

$E[X] = \int_0^s (xp(X = x \mid s))dx$

$= \int_0^s (x(2(s-x)/s^2)dx$

$= \frac{2}{s^2} \int_0^s (x(s-x)dx$

$= \frac{2}{s^2} \int_0^s (sx - x^2)dx$

$= [\frac{2}{s^2} \left(\frac{s}{2} x^2 - \frac{x^3}{3}\right)]_0^s$

$=[\frac{1}{s} x^2 - \frac{2}{3s^2} x^3]_0^s$

$= s - \frac{2}{3} s - 0$

$= \frac{1}{3} s$

3.

$E\left[\sqrt{x}\right] = \int_0^s (\sqrt{x} p(X = x \mid s))dx$

$= \int_0^s (\sqrt{x}(2(s-x)/s^2)dx$

$= \frac{2}{s^2} \int_0^s (\sqrt{x}(s-x)dx$

$= \frac{2}{s^2} \int_0^s (sx^{1/2} - x^{3/2})dx$

$= [\frac{2}{s^2} \left(\frac{2s}{3} x^{3/2} - \frac{2x^{5/2}}{5}\right)]_0^s$

$= [\frac{4}{3s} x^{3/2} - \frac{4x^{5/2}}{5s^2}]_0^s$

$= \frac{4}{3} \sqrt{s} - \frac{4}{5} \sqrt{s}$

$= \frac{8}{15} \sqrt{s}$

4.

$V[X] = E[X^2] - E[X]^2$

$= \int_0^s (x^2 p(X = x \mid s))dx - E[X]^2$

$= \int_0^s (x^2(2(s-x)/s^2)dx - E[X]^2$

$= \frac{2}{s^2} \int_0^s (x^2(s-x)dx - E[X]^2$

$= \frac{2}{s^2} \int_0^s (sx^2 - x^3)dx - E[X]^2$

$= [\frac{2}{s^2} \left(\frac{s}{3} x^3 - \frac{x^4}{4}\right)]_0^s - E[X]^2$

$= [\frac{2}{3s} x^3 - \frac{x^4}{2s^2}]_0^s - E[X]^2$

$= \frac{2}{3} s^2 - \frac{1}{2} s^2 - \frac{1}{9} s^2$

$= \frac{1}{18} s^2$

5.

$$Median[X] \ = \ Q(P = 1/2)$$

$$\int_0^x p(x \ \leq x \mid s)dx = 1/2$$

$$\int_0^x 2(s - x')/s^2 dx' = \frac{1}{2}$$

$$\frac{2}{s^2}[\left(sx' - \frac{x'^2}{2}\right)]_0^x = \frac{1}{2}$$

$$-\frac{x^2}{s^2} + \frac{2x}{s} = \frac{1}{2}$$

$$2(2xs - x^2) = s^2$$

$$2x^2 - 4xs + s^2 = 0$$

By using the quadratic function,

$$x = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a}$$

$$x = \frac{4s \pm \sqrt{16s^2 - 4(2)s^2}}{2(2)}$$

$$x = \frac{4s \pm \sqrt{8s^2}}{4}$$

$$x = \frac{4s \pm 2\sqrt{2}s^2}{4}$$

$$x = \frac{2s \pm \sqrt{2}s^2}{2}$$

I would take the positive x as median because the median represents a value that is at the center of the dataset. Negative values wouldn't make sense in this context, as they don't correspond to the middle value of the data.

Median is equal to $\frac{2s + \sqrt{2}s^2}{2}$.

Q5

1.

By using the code below, I found that the estimated value is 15.556.

```
covid <- read.csv("covid.2023.csv")

calculate_estimates <- function(x) {

  n <- length(x)

  return(sum(x)/n)

}

lambda_est <- calculate_estimates(covid$Day)
```

```
> setwd("D:/Downloads")
> covid <- read.csv("covid.2023.csv")
> # 1.
> calculate_estimates <- function(x) {
+   n <- length(x)
+   return(sum(x)/n)
+ }
> lambda_est <- calculate_estimates(covid$Day)
> print(lambda_est)
[1] 15.556
```

2a.

The probability is 0.0938464.
Code:

```
ppois(10,lambda_est)
```

```
> # 2a.
> ppois(10,lambda_est)
[1] 0.0938464
```

2b.

11, 12 and 14 days
Code:

```
recovery_counts <- as.data.frame(table(covid$Days))

colnames(recovery_counts) <- c("recovery_days", "count")

recovery_counts <- recovery_counts[order(recovery_counts$count, decreasing = TRUE), ]

top_three_recovery <- head(recovery_counts, 3)

print(top_three_recovery)
```

```
> # 2b.
> recovery_counts <- as.data.frame(table(covid$Days))
> colnames(recovery_counts) <- c("recovery_days", "count")
> recovery_counts <- recovery_counts[order(recovery_counts$count, decreasing = TRUE), ]
> top_three_recovery <- head(recovery_counts, 3)
> print(top_three_recovery)
   recovery_days count
14            14    63
12            12    57
11            11    55
```

2c.

Probability is 0.6058841
Code:

prob_60_to_80 <- ppois(80,5*lambda_est) - ppois(60,5*lambda_est)

print(prob_60_to_80)

```
> # 2c.
> prob_60_to_80 <- ppois(80,5*lambda_est) - ppois(60,5*lambda_est)
> print(prob_60_to_80)
[1] 0.6058841
```

2d.

Let X be days to recover,

$P(X \geq 14) = 1 - P(X < 14)$

$= 0.6879$

```
> # 2d.
> prob_after_13 <- (1 - ppois(13,lambda_est))
> print(prob_after_13)
[1] 0.6878704
```

Code:

prob_after_13 <- (1 - ppois(13,lambda_est))

print(prob_after_13)

Probability of taking 14 or more than 14 days to recover is 0.6879.

Let Y be number of patients,

$P(Y \geq 3) = 1 - P(Y < 3)$

$P(Y \geq 3) = 1 - P(Y = 1) - \ P(Y = 2)$

$= 1 - 5C_1(0.6879)^1(1 - 0.6879)^4 - 5C_2(0.6879)^2(1 - 0.6879)^3$

$= 1 - 0.0326 - 0.1439$

$= 0.8235$

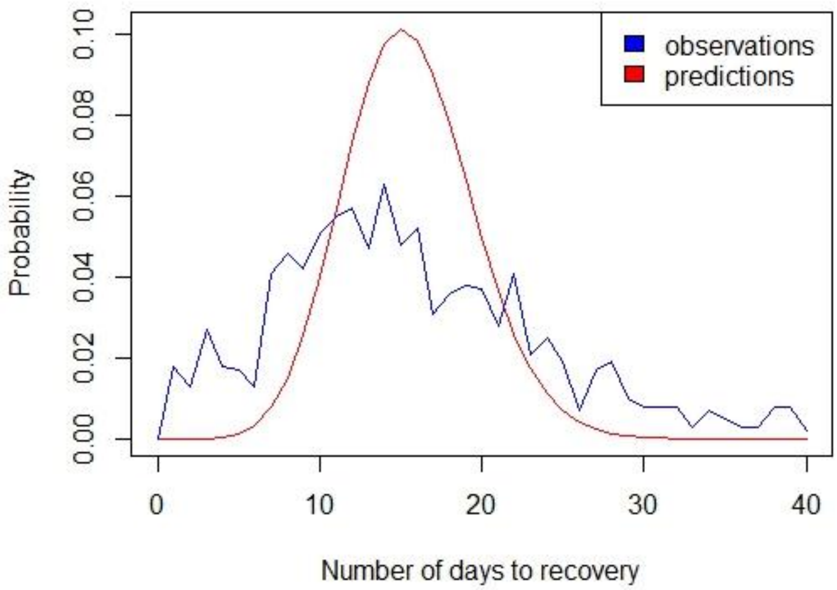
Probability of among 5 patients, 3 or more than 3 patients taking 14 or more than 14 days to recover is 0.8235.

3.

Code:

```
x <- 0:40

y_observed = rep(0,41)

n = length(covid$Days)

for(i in covid$Days){

  y_observed[i + 1] <- y_observed[i + 1] + (1 / n)

}

y_predict <- dpois(x,lambda_est)

plot(x,y_predict,"l",col="red",xlab = "Number of days to recovery", ylab = "Probability")

lines(x,y_observed, col = "blue")

legend(x= "topright", c("observations","predictions"),fill = c("blue","red"))
```

```
> # 3.
> x <- 0:40
> y_observed = rep(0,41)
> n = length(covid$Days)
> #Calculate accumulated probabilities of patients recovering on or before each day of the data
> for(i in covid$Days){
+   y_observed[i + 1] <- y_observed[i + 1] + (1 / n)
+ }
> y_predict <- dpois(x,lambda_est)
> plot(x,y_predict,"l",col="red",xlab = "Number of days to recovery", ylab = "Probability")
> lines(x,y_observed, col = "blue")
> legend(x= "topright", c("observations","predictions"),fill = c("blue","red"))
>
```



By observing the graph, we can see that there are significant discrepancies between the observed and predicted values, thus the Poisson distribution is not a appropriate model for the COVID recovery data.