Q1) 1. Generative AI systems. 2. Recommendation systems 3. Forecasting. 4. Risk prediction

Q2)

1. Pr(W) = .588
2. Pr(W|P’) = Pr(W ^ P’) / Pr(P’) = (.235 + .058) + / 0.586 = 0.5
3. 1 - Pr(W|P’) = .5
4. No. Same probability for each outcome

Let X be the number of wins in a sequence

1. Pr ( X > 0 ) = 1- Pr (X = 0)

Over 2 games where won previous, and conditional that PH then.

We are interested in the case where the team loses both games, to take the inverse. In the first game they need to lose conditional to being at home and having won the first game, then they need to lose again this time they will have lost the last game and be away. Therefore:

Pr(X>0) = 1 – (Pr(W’| P,H) \* Pr(W’|P’,H’))

Pr(X>0) = 1 – ( ( .059/(.178+.059) ) \* (.176/(.235+.176)) ) = .893

Q3)

1)

S = 2 \* X1 – Y1

E[S] = 2\*E[X1] -E[Y1] since L.I.E

E[X1] = p \* Sum x = 1/6 \* (1+2+3+4+5+6) = 3.5

E[Y1] = p \* Sum y = ¼ \* (1+2+3+4) = 2.5

E[S] = 2\*3.5 – 2.5 = 4.5

2)

V[S] = Var[2X1] + Var[-Y1] since both die rolls are independent ie Cov(X1,Y1) = 0

V[S] = 4\* Var[X] + 1\*Var[Y1] since Var is a ^2 operator

Var[X1] = E[ (X1 – (4.5) )2]

= 1/6 \* ( (1-3.5)2 + (2-3.5)2 + (3-3.5)2 + (4-3.5)2 +(5-3.5)2 + (6-.3.5)2 )

= 35/12

Var[Y1] = E[ (Y1 – (2.5) )2]

Var[Y1] = ¼ \* ( (1-2.5)^2 + (2-2.5)^2 + (3-2.5)^2 + (4-2.5)^2 )

= 5/4

V[S] = 4\*(35/12) + ( 5/4 ) = 12.917

3)

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| x | -2 | -1 | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
| Pr(X=x) | 1/24 | 1/24 | 2/24 | 2/24 | 2/24 | 2/24 | 2/24 | 2/24 | 2/24 | 2/24 |

|  |  |  |  |
| --- | --- | --- | --- |
| 8 | 9 | 10 | 11 |
| 2/24 | 2/24 | 1/24 | 1/24 |

This was done by the probability of one outcome = 1/6 \* ¼ = 1/24, multiplied by the number of ways to get that outcome.

4)

E [S 3] = 186.819444

= SUM ( S3 \* Pr(X=x) ) = 265.5

A screenshot of a calculator

Description automatically generated

5)

E[S3] using Taylor Series aprox equal to:

f( E[S] ) + d2f(x)/dx2|x=u \* ó2/2

E[S]^3 = 4.53 = 91.125

ó2/2 = 10.42 / 2 = 6.45833333

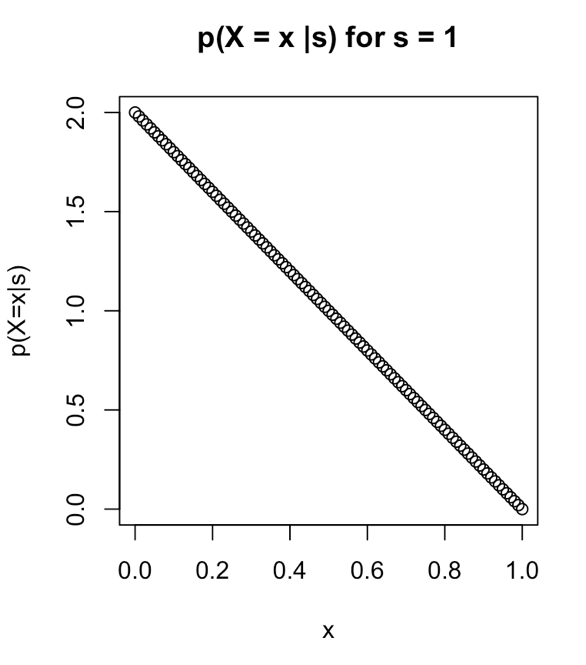
d2f(x)/dx2|x=u

d2f(x)/dx2 = 6S at x = u = 6 \* 4.5 = 27

E[S3] using Taylor Series approx. equal to 91.125 + 6.45833333.

\*27 = 265.499

Q4) 1)

A graph of a line

Description automatically generated

2) E[X] =  ∫s0 x P ( X=x|s ) d x

=  ∫s0 x (2(s-x)/s2) d x

Just expanding and factorising the middle term to substitute later

x (2(s-x)/s2) = 2x(s-x)/s2 = (2xs -2x2)/s2

Then anti differentiate with respect to x

= ( x2s – (2/3 \* x3 ) )/s2

Then calculate the difference once we substitute out values for s and 0

∫s0 x (2(s-x)/s2) d x =([s]2s – (2/3 \* [s]3 ) )/s2 - ( [0]2s – (2/3 \* [0]3 ) )/s2

= ( s3 – 2s3/3) / s2 - 0

E[X] = (s3/3) / s2 = s/3

A paper with writing on it

Description automatically generated

3) E[√X]

A math equation with black text

Description automatically generated

A notebook with writing on it

Description automatically generated

Q4)

A notebook with math equations on it

Description automatically generated

Q5)

A notebook with math equations on it

Description automatically generated

Question 5)

1. The estimated parameter value for the rate in a Poisson distribution is 15.556

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

x = mean(data$Days)

* 1. ppois(10, lambda=x)  
       
     Pr(X <= 10 ) = 0.0938464
  2. 14,15,16
  3. ans =ppois(80/5, lambda=x) - ppois(60/5, lambda=x)

Using the additive characteristic of Poisson distribution will be the same as one individual recovering in 60/5 to 80/5 days   
Pr(X <= 80/5 ) - Pr(X <= 60/5 ) = 0.3857989

* 1. What is the probability that three or more of these five patients will recover on or after day 14

Probability that one person recovers on or after day 14 = 0.5900841

ppois(14, lambda=x, lower = FALSE)

Becomes a Bernoulli sequence with P success 0.5900841

5 choose 3

+

5 choose 4

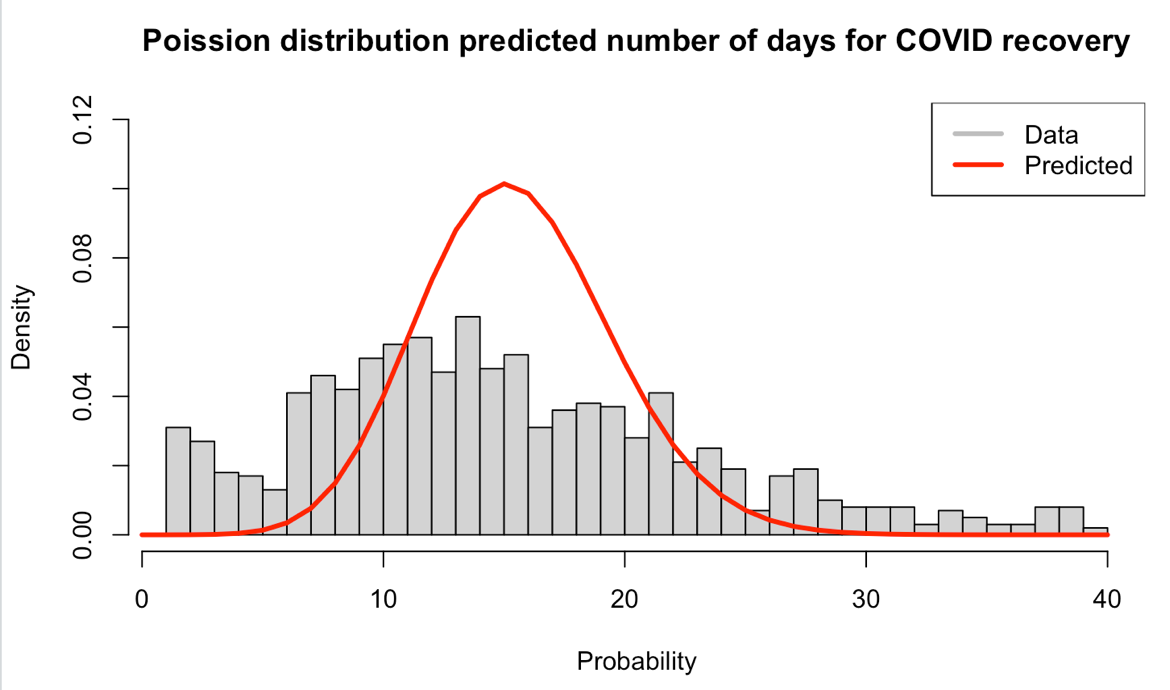
+

5 choose 5

dbinom(3, 5, 0.5900841) + dbinom(4, 5, 0.5900841) + dbinom(5, 5, 0.5900841)

=

0.6652881



recovery\_days <- 0:40

probability = dpois(recovery\_days, lambda=x )

RecoveryDays =data$Days

hist(RecoveryDays, freq = FALSE, breaks = 40,

main = "Poission distribution predicted number of days for COVID recovery",

ylim = c(0,.12),

xlab = "Probability")

lines(recovery\_days, probability,col="red",lwd = 3)

legend("topright", legend = c("Data", "Predicted"),

lwd = 3, col = c("grey", "red"))

I do not believe the poission model is a good predictor for the population data since it seems the population is more spread than the poisssion predicted distribution. There are several extreme values in the actual data set that occur significantly more than the poission model predicts, and the central number of recover days is over predicted by the poisssion model meaning it predicts they those number of recovery days will occur more than they actually do. Put simply, the poission model does not measure the kurtosis of the data well.