PBSR ASSig.2 Q.1

2022-11-14

Problem 1

Suppose X denote the number of goals scored by home team in premier league. We can assume X is a random variable. Then we have to build the probability distribution to model the probability of number of goals. Since X takes value in $\mathbb{N} = \{0, 1, 2, \cdots\}$, we can consider the geometric progression sequence as possible candidate model, i.e.,

$$S = \{a, ar, ar^2, ar^3, \cdots\}.$$

But we have to be careful and put proper conditions in place and modify S in such a way so that it becomes proper probability distributions.

1. Figure out the necesary conditions and define the probability distribution model using S.

Answer: In order to define the probability distribution model using S, following necessary conditions must be satisfied.

1. $0 \leq P(X = i) \leq 1, \forall\ i$.
2. $\Sigma_{i=0}^{\infty} P(X = i) = 1$.

According to the given model $P(X = i) = ar^i$.

Therefore,

$$0 \leq a \leq 1 \quad \& \quad 0 \leq r < 1$$

Now,

$$\Sigma_{i=0}^{\infty} P(X = i) = 1$$
$$\implies \Sigma_{i=0}^{\infty} ar^i = 1$$
$$\implies r = 1 - a$$

Therefore, $P(X = i) = a(1 - a)^i$.

2.Check if mean and variance exists for the probability model.

Answer: $E(X) = \Sigma_{i=0}^{\infty} iP(X = i) = \Sigma_{i=0}^{\infty} a(1 - a)^i = \frac{1-a}{a}$.

Therefore, mean $= E(x) = \frac{1-a}{a}$.

variance $= Var(X) = \Sigma_{i=0}^{\infty} (i - E(X))^i = \frac{1-a}{a^2}$.

Therefore, Variance $= \frac{1-a}{a^2}$.

Hence, mean and variance exists for the probability model.

3. Can you find the analytically expression of mean and variance.

Answer: Yes, we can analytically find the expression of mean and variance as follows.

mean $= E(x) = \frac{1-a}{a}$.

Variance $= \frac{1-a}{a^2}$.

4.

From historical data we found the following summary statistics

mean	median	variance	total number of matches
1.5	1	2.25	380

Using the summary statistics and your newly defined probability distribution model find the following:

a. What is the probability that home team will score at least one goal?

b. What is the probability that home team will score at least one goal but less than four goal?

Answer: According to given historical data,

$$
\begin{aligned}
mean &= 1.5 \\
\implies \frac{1-a}{a} &= 1.5 \\
\implies a &= 0.4
\end{aligned}
$$

Therefore, $X \sim geom(0.4)$.

a. probability that home team will score atleast one goal,

$= p(X \geq 1) = 1 - p(X = 0)$

```
p = 1 - dgeom(0,0.4)
print(p)
```

```
## [1] 0.6
```

b. probability that home team will score at least one goal but less than four goal,

$= p(1 \leq X < 4) = p(X = 1) + p(X = 2) + p(X = 3)$

```
p = dgeom(1,0.4) + dgeom(2,0.4) + dgeom(3,0.4)
print(p)
```

```
## [1] 0.4704
```

5. Suppose on another thought you want to model it with off-the shelf Poisson probability models. Under the assumption that underlying distribution is Poisson probability find the above probabilities, i.e.,

a. What is the probability that home team will score at least one goal?

b. What is the probability that home team will score at least one goal but less than four goal?

Answer: Now, we want to model above distribution with poisson probability model. Assume that, the underlying distribution is poisson probability distribution.

i.e Assume that, X~poisson(λ), where λ = average number of goals = mean.

According to summary statistics, mean = 1.5.

$\implies \lambda = 1.5.$

Thus, $X \sim poisson(1.5)$

a. probability that home team will score at least one goal,

$= p(X \geq 1) = 1 - p(X = 0)$

```
p = 1 - dpois(0,1.5)
print(p)
```

```
## [1] 0.7768698
```

b. probability that home team will score at least one goal but less than four goal,

$= p(1 \leq X < 4) = p(X = 1) + p(X = 2) + p(X = 3)$

```
p = dpois(1,1.5) + dpois(2,1.5) + dpois(3,1.5)
print(p)
```

```
## [1] 0.7112274
```

6. Which probability model you would prefer over another?

Answer: I prefer the poisson model over the newly defined probability model for this given data.

7. Write down the likelihood functions of your newly defined probability models and Poisson models. Clearly mention all the assumptions that you are making.

Answer:

i) Likelihood function for newly defined probability model.

Let $x_1, x_2, x_3, \cdots, x_n$ be a observed sample.

Then the likelihood function is given by,

$$
\begin{aligned}
f(x_1, x_2, x_3, \cdots, x_n|\hat{a}) &= f(x_1|\hat{a})f(x_2|\hat{a})f(x_3|\hat{a})\cdots f(x_n|\hat{a}) && (1)\\
&= \hat{a}(1-\hat{a})^{x_1}\hat{a}(1-\hat{a})^{x_2}\hat{a}(1-\hat{a})^{x_3}\cdots\hat{a}(1-\hat{a})^{x_n} && (2)\\
&= \hat{a}^n(1-\hat{a})^{\Sigma_{i=1}^{\infty}x_i} && (3)
\end{aligned}
$$

ii) Likelihood function for poisson model.

Let $x_1, x_2, x_3, \cdots, x_n$ be a observed sample.

Then the likelihood function is given by,

$$
\begin{aligned}
f(x_1, x_2, x_3, \cdots, x_n|\hat{\lambda}) &= f(x_1|\hat{\lambda})f(x_2|\hat{\lambda})f(x_3|\hat{\lambda})\cdots f(x_n|\hat{\lambda}) && (4)\\
&= \frac{e^{-\hat{\lambda}}\hat{\lambda}^{x_1}}{x_1!}\frac{e^{-\hat{\lambda}}\hat{\lambda}^{x_2}}{x_2!}\frac{e^{-\hat{\lambda}}\hat{\lambda}^{x_3}}{x_3!}\cdots\frac{e^{-\hat{\lambda}}\hat{\lambda}^{x_n}}{x_n!} && (5)\\
&= \frac{e^{-n\hat{\lambda}}\hat{\lambda}^{\Sigma_{i=1}^{n}x_i}}{\Pi_{i=1}^{n}x_i!} && (6)
\end{aligned}
$$

Assumptions:

- Number of goals scored in a match is independent of the number of goals scored in every other match.
- Number of goals in every match is identically distributed.
- Assume that no other factors is affecting the number of goals scored in a match.