Probability and Statistics with R

Assignment 2

Submission Nov 16-2022 (Wednesday)

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 necessary conditions and define the probability distribution model using S.
   Solution:
   Firstly we should have $0 < a < 1$ and
   We must have

   $$\sum_{i\geq 0} ar^i = 1 \therefore\ 1 + r + r^2 + \cdots = \frac{1}{a} \therefore\ \frac{1}{1-r} = \frac{1}{a} \implies a = 1 - r$$

   We also have $0 < r < 1$
   So the required conditions are:
   $0 < a < 1, 0 < r < 1$
   $a + r = 1$
   So $S = \{1-r, r(1-r), r^2(1-r), \dots\}$ Let $p = 1 - r$ then $S = \{p, p(1-p), p(1-p)^2, \dots\}$ This is the geometric distribution as we know it
2. Check if mean and variance exists for the probability model.
   Solution:

   $$\text{Mean} = a \sum_{i\geq 0} ir^i = a \sum_{i\geq 0} \left((i+1)r^i - r^i\right)$$

   $$= a\left[\sum_{i\geq 0}(i+1)r^i - \sum_{i\geq 0} r^i\right] = a\left[\frac{1}{(1-r)^2} - \frac{1}{1-r}\right]$$

   $$= a\left[\frac{r}{(1-r)^2}\right] = \frac{r}{1-r}$$

   $$\text{Variance} = E[X^2] - E[X]^2 = \sum_{i\geq 0} i^2 r^i - \left(\frac{r}{1-r}\right)^2$$

To calculate $\sum_{i\geq 0} i^2 r^i$:

$$\text{We know that} \sum_{i\geq 0} i r^i = \frac{r}{(1-r)^2}$$

Differentiating both sides and then multiplying by r, we get

$$\sum_{i\geq 0} i^2 r^i = \frac{r^2+r}{(1-r)^2}$$

$$\therefore \text{Variance} = \frac{r^2+r}{(1-r)^2} - \left(\frac{r}{1-r}\right)^2 = \frac{r}{(1-r)^2}$$

3. Can you find the analytically expression of mean and variance.
Solution:
Yes. We have found the analytically closed expression of mean and variance above 4. From historical data we found the following summary statistics

mean	median	variance	total number of matches
1.5	1	2.25	380

4. Using the summary statistics and your newly defined probability distribution model find the following:
   Using Method of Moments, if we equate the first moment, i.e the mean
   We get $\frac{r}{1-r} = 1.5 \implies r = 0.6$

a. What is the probability that home team will score at least one goal?
   Solution:
   P(at least 1 goal)=1-P(no goals)
   =1-(1-r)=r=0.6

b. What is the probability that home team will score at least one goal but less than four goal? Solution:
   $\text{P}(1 \leq \text{goals} < 4) = 0.24 + .0144 + 0.0864 = 0.4704$

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., If we want to fit Poisson model then we know mean=λ
   So we take $\lambda = 1.5$

a. What is the probability that home team will score at least one goal?
   Solution:
   P(at least 1 goal)=1-P(no goals)
   =1-$e^{-1.5}(1.5)^0\frac{1}{0!} = 1 - 0.22313 = 0.77687$

b. What is the probability that home team will score at least one goal but less than four goal?
   Solution:
   $$\text{P}(1 \leq \text{goals} < 4) = e^{-1.5}\left(\frac{(1.5)^1}{1!} + \frac{(1.5)^2}{2!} + \frac{(1.5)^3}{3!}\right) = 0.22313\,(1.5 + 1.125 + 0.5625) = 0.74310$$

6. Which probability model you would prefer over another?
   Solution:
   For both the models, the value of variance of the model and the available data does not match and hence both are not perfect. However if we had to choose one, we would choose the Poisson model because its actual variance (1.5) is closer than Geometric distribution's actual variance (3.75) to the given variance (2.25). Also intuitively Poisson model makes more sense than the Geometric model since geometric is generally used for number of failures required to get one success whereas here we are modelling for the number of goals.

7. Write down the likelihood functions of your newly defined probability models and Poisson models. Clearly mention all the assumptions that you are making.
   Solutions:
   Geometric Distribution: If $X_1, X_2, ..., X_n$ are i.i.d random variables following Geometric distribution,
   Likelihood function:
   $L(p|x) = \Pi_{i=1}^{n} p(1-p)_i^x$
   where x_i take non-negative integer values

Poisson Distribution:
If $X_1, X_2, ..., X_n$ are i.i.d random variables following Poisson distribution,
Likelihood function:
$L(\lambda|x) = \Pi_{i=0}^{n} \frac{e^{-\lambda}\lambda^{x_i}}{x_i!}$ where x_i take non-negative integer values