Modelling goals in a football match

Problem 1

Problem solver: Rohan Karthikeyan

Reviewer: Sulagna Barat

Part 1

We are given a random variable X with a supposed prob. distribution $\{a, ar, ar^2, ...\}$. Hence, we can write:

$$\mathbb{P}(X = i) = ar^i$$

Now, because $\sum_i \mathbb{P}(X = i) = 1$ for a probability distribution, we must have

$$\begin{aligned} \sum_i \mathbb{P}(X = i) &= 1 \\ \implies \sum_i ar^i &= 1 \\ \implies \frac{a}{1-r} &= 1 \\ \implies a &= 1 - r \end{aligned}$$

We have ***assumed*** that $|r| < 1$, but because probabilities are non-negative, it's sufficient to have: $0 \leq r < 1$. Hence, we finally have

$$\mathbb{P}(X = i) = (1 - r)r^i, \quad 0 \leq r < 1$$

Parts 2 and 3

Once again, we ***assume*** $|r| < 1$, but because probabilities are non-negative, it's sufficient to have: $0 \leq r < 1$. We now have:

$$\begin{aligned} \mathbb{E}(X) &= \sum_i i\mathbb{P}(X = i) \\ &= \sum_i i(1 - r)r^i \\ \implies \frac{\mathbb{E}(X)}{1 - r} &= \sum_i ir^i \\ \implies \frac{\mathbb{E}(X)}{1 - r} &= \frac{r}{(1 - r)^2} \\ \implies \mathbb{E}(X) &= \frac{r}{1 - r} \end{aligned}$$

Observe that

$$\begin{aligned}\mathbb{E}(X^2) &= \sum_i i^2 \mathbb{P}(X=i) \\ &= \sum_i i^2 (1-r) r^i\end{aligned}$$

Calculating the sum, we obtain:

$$\mathbb{E}(X^2) = \frac{r(r+1)}{(1-r)^2}$$

To calculate the variance, now:

$$\begin{aligned}\mathrm{Var}(X) &= \mathbb{E}(X^2) - [\mathbb{E}(X)]^2 \\ &= \frac{r(r+1)}{(1-r)^2} - \frac{r^2}{(1-r)^2} \\ &= \frac{r}{(1-r)^2}\end{aligned}$$

Part 4

We are given that the mean is 1.5 and the variance is 2.25. This amounts to solving the following set of equations:

$$\frac{r}{1-r} = 1.5 \tag{1}$$

$$\frac{r}{(1-r)^2} = 2.25 \tag{2}$$

But, one can easily verify that this system of equations is inconsistent. So, we cannot solve for the value of r using this method. To resolve this issue, we are going to find the ***maximum likelihood estimate*** of r.

Denote by $\mathbf{x}$ the observed values in a random sample $x_1, x_2, \cdots, x_n$. The likelihood function for the geometric distribution can be expressed as:

$$L(r|\mathbf{x}) = \prod_{i=1}^{n} (1-r) r^{x_i} = (1-r)^n \, r^{\sum_{i=1}^{n} x_i}$$

Taking the natural logarithm of the likelihood function gives:

$$\ln L(r|\mathbf{x}) = \ln \left[(1-r)^n \, r^{\sum_{i=1}^{n} x_i}\right] = n \ln(1-r) + \ln(r) \sum x_i \tag{a}$$

Let's take the first-order partial derivative of $\ln L(r|\mathbf{x})$ with respect to r and set the answer equal to zero:

$$\frac{\partial \ln L(r|\mathbf{x})}{\partial r} = -\frac{n}{1-r} + \frac{\sum x_i}{r} \overset{set}{=} 0$$

The solution is given by $\hat{r} = \frac{\sum x_i}{\sum x_i + n}$. It's easy to check that the second-order partial derivative of the log-likelihood function is negative at $r = \hat{r}$.

For our problem, let's find this value from the summary of data we're given:

$$\hat{r} = \frac{380 * 1.5}{380 * 1.5 + 380} = \frac{3}{5}$$

Consequently,

a. $\mathbb{P}(X \geq 1) = 1 - \mathbb{P}(X=0) = 1 - (1-r) = 0.6$;
b. $\mathbb{P}(1 \leq X < 4) = \mathbb{P}(X=1) + \mathbb{P}(X=2) + \mathbb{P}(X=3) = (1-r)\big[r + r^2 + r^3\big] = 0.4704$

Part 5: What is λ?

For a Poisson distribution, the mean and variance is equal to λ. Furthermore, we know that the ***maximum likelihood estimate*** for λ is the sample mean. So, we will take $\lambda = 1.5$.

a. $\mathbb{P}(X \geq 1) = 1 - \mathbb{P}(X = 0) = 1 - e^{-\lambda} = 0.78$;
b. $\mathbb{P}(1 \leq X < 4) = \mathbb{P}(X = 1) + \mathbb{P}(X = 2) + \mathbb{P}(X = 3) = e^{-\lambda}\left[\lambda + \frac{\lambda^2}{2} + \frac{\lambda^3}{6}\right] = 0.71$

Part 6: Which is better?

Looking at the observed probabilities for both (a) and (b) in Parts 4 and 5, we are inclined towards the Poisson model as likely to better fit the data.

Part 7: The likelihood functions

For the Poisson distribution:

$$L(\lambda|\mathbf{x}) = \prod_{i=1}^{n} \frac{e^{-\lambda}\lambda^{x_i}}{x_i!} = e^{-n\lambda} \prod_{i=1}^{n} \frac{\lambda^{x_i}}{x_i!}$$

For the geometric distribution:

$$L(r|\mathbf{x}) = \prod_{i=1}^{n} (1-r) r^{x_i} = (1-r)^n \, r^{\sum_{i=1}^{n} x_i}$$