REPORT

Probability and Statistics with R- Assignment 2

P Keshavan| D Siva Manoj | Bhumika Chawla

# Question 1

We had to fit the number of goals scored by the home team according two models. In the first model, a geometric sequence was given; we had to define the function and establish various assumptions and conditions necessary to convert it into a valid probability distribution. We showed the existence and expressed the analytical form of mean and variance. We determined the distribution's parameters and the probability for the given issues using the provided historical data. Poisson distribution was the second model used to fit the number of goals. Again using the given historical data we calculated the parameter value and calculated the probability for the given problems. We defined the likelihood function for each of the models with assumptions. We compared the two models and concluded that; poisson model fits better for the given data.

# Question 2

The distribution in question was Gamma (α, σ=2.2). We had to find the Maximum Likelihood estimates of log(α) for different sample sizes. For different sizes of the sample we observed the sampling distribution of the Maximum Likelihood Estimator using a histogram. The histogram was also marked with the true value of parameter. We concluded that the interquartile range of the sampling distribution decreases as the sample size increases.

# Question 3

The data in the question was about the 'Faithful' dataset in R. It's a built-in dataset about the 'Old faithful geyser in Yellowstone National Park in Wyoming, United States. In this question, we had to fit a probability model for the waiting times for eruptions at the geyser. We tried 3 different mixed probability models, Gamma-Normal, Gamma-Gamma and Lognormal-Lognormal. We calculated the Akaike's Information Criterion of these 3 probability models. We found that the best model among them was Lognormal-Lognormal using the lowest Akaike's Information Criterion (aka AIC).

# Question 4

The data in question was about the ‘Insurance’ dataset that consists of the numbers of policyholders of an insurance company who were exposed to risk, and the numbers of car insurance claims made by those policyholders in the third quarter of 1973. We want to fit ‘Claims’ as function of ‘Holders’. The different distributions considered are Normal Distribution, Laplace Distribution, Lognormal Distribution and Gamma Distribution. For each of these we calculated Log Likelihood function and used the ‘optim’ function to estimate the maximum log likelihood estimates of the parameters. We plotted the scatter plot and the regression line for all the above models. We calculated the Bayesian Information Criterion (BIC) for each of the model and concluded that Gamma distribution (has the least BIC value) is the best model.

# Question 5

The data used in the question is the NSEI and TCS prices available in the ‘quantmod’ library. We want to model TCS (Adjusted) log returns as a model of Nifty50 (Adjusted) log returns. We estimated the parameters of the model using Method of methods and using the in-built ‘linear model- lm()’ function. In both the cases we received the same parameter values. We used this model to estimate what changes we can expect in TCS prices when Nifty50 prices change.