Readme file: PBSR Assignment 2

Anna Maria,Sneha KK, Sarvesh Bhandary

2022-11-15

Group Members:
Anna Maria KV (MDS202209)
Sarvesh Bhandary (MDS202253)
Sneha KK(MDS202240)

Question 1: The given probability distribution model is $Geom(1-r)$ under the appropriate conditions, hence the mean and variance exist. The parameters do not tally for the given values of mean and variance, so we used the mean to build the probability models with Geometric and Poisson as underlying distributions. We compared the variances with the observed variance: Geometric : r= 0.6 mean =1.5(observed) variance=3.5 Poisson: λ= 1.5= mean = variance Observed variance=2.25 Hence we can conclude that poisson distribution is a better fit to the data than geometric.

Question 2: We sample n points from a gamma distribution and try to estimate its parameter values using MLE. We vary the sample size (n) and observe that the gap between 2.5 and 97.5-percentile points is shrinking as sample size n is increasing:
n=20: Gap between the percentiles =1.1844
n=40: Gap between the percentiles=0.8332107
n=100: Gap between the percentiles= 0.5120706

Question 3: We tried to fit three different bimodal probability models on the faithful geyser dataset and compare their AIC values: Model 1:

$$f(x) = p * Gamma(x|\alpha, \sigma_1) + (1-p)N(x|\mu, \sigma_2^2), \quad 0 < p < 1$$

AIC value: 2076.18

Model 2:

$$f(x) = p * Gamma(x|\alpha_1, \sigma_1) + (1-p)Gamma(x|\alpha_2, \sigma_2), \quad 0 < p < 1$$

AIC value: 2076.116

Model 3:

$$f(x) = p * logNormal(x|\mu_1, \sigma_1^2) + (1-p)logNormal(x|\mu_1, \sigma_1^2), \quad 0 < p < 1$$

AIC value: 2075.42
Since the AIC value for the third model is lower, the log likelihood for third model will be the highest, and hence model 3 is the best.

Question 4: We try to predict Holders from Claims in the Insurance dataset using regression by varying the underlying probability distribution and compare the BIC values.
Normal: BIC= 510.7587
Laplace: BIC=498.6869
Lognorm: BIC=567.9726
Gamma: 664.9584
Laplace and normal distributions result in better fits than Gamma and Lognormal.

Question 5: We try to predict the return values of TCS stocks from Nifty stocks using linear regression. We compute our results using MOM plug in estimator and lm function in R :

Method of moments plug in estimator of theta is: (0.0004628222 , 0.7436844 , 0.01618466)
Estimator of theta using lm function is: (0.0004628222 , 0.7436844 , 0.01618685)

Our results also told us that as the current value of Nifty is 18000 and it goes up to 18200, the current value of TCS stock is expected to go up from 3200 to 3227.898.