Real Data Analysis

Group 6 (MA4740- Introduction to Bayesian Statistics)

Megavath Rajinikanth *ES20BTECH11018*

Mohd Saad *MA20BTECH11009*

Varunaditya Singhal *MA20BTECH11021*

Nikhil Kongara *MA20BTECH11011*

CONTENTS

I	Introduction	1
II	Prior Data	1
III	Gamma Distribution	1
IV	Data-Likelihood Distribution	1
V	Posterior Distribution	2
VI	Conclusion	2

***Abstract*—This project explores the idea of analyzing a real-world data set to demonstrate Poisson-Gamma Bayesian analysis. The project is part of the group project MA4740 - Introduction to Bayesian Statistics. The model is implemented in R.**

I. Introduction

The project includes collecting a real-life data set and applying the statistical methods while also performing a Poisson-Gamma Bayesian analysis. The real-life data set that we have taken is about the basketball matches played by Stephen Curry in one season i.e from 2022-2023. The data-set was taken from Stephen Curry 2022-23 Game Log The real-life data set contains attributes like Team, age, opponent team, minutes played, fields Goals,3 - pt field Goals, Free throws, rebounds, total points etc.

The parameter of interest is : **3 - Pointers scored per game**.

II. Prior Data

The prior data for the application of Poisson-Gamma distribution was taken from previous season i.e from 2021 - 2022.The attribute on which we will be working is 3-Pointers (3P). Gamma is the conjugate prior for the Poisson distribution.

The prior distribution is given by

$$\lambda \sim Gamma(\alpha, \beta) \tag{1}$$

where

- λ : Number of 3 - Pointers scored by Stephen Curry per game.
- α , β : Hyper-parameters .

III. Gamma Distribution

We try to fit the prior data distribution to a Gamma distribution.

In a Gamma distribution, we get $\lambda \sim$ Gamma(α, β), where,

$$mean = \frac{\alpha}{\beta} \tag{2}$$

$$variance = \frac{\alpha}{\beta^2} \tag{3}$$

We use the data of the first season i.e 2021 - 2022 for the prior distribution and use the Methods of Moments to estimates the hyper-parameters. The calculations yield us the results :

First moment = Mean = 4.507,
α = 11.17,
β = 2.47

Given below is the pdf of beta distribution of the prior data.

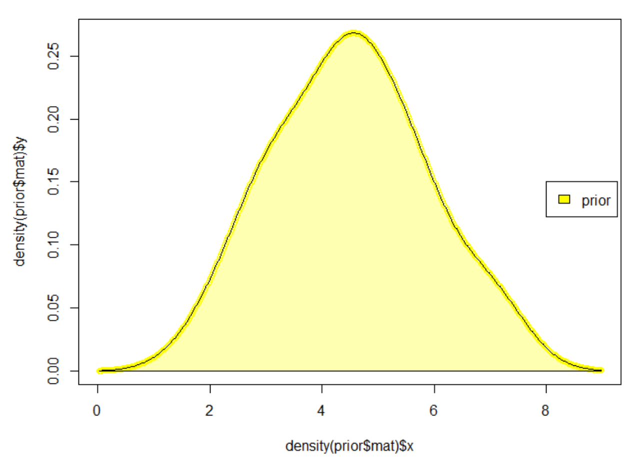

Fig. 1. Prior Distribution

IV. Data-Likelihood Distribution

To perform Poisson - Gamma analysis we will be requiring likelihood function.

We choose $Y_i \mid \pi \sim$ Poisson(y_i,λ) where,

L_i : The number of 3 - Pointers scored by Stephen during the 2nd season i.e 2022 - 2023.

λ : Rate of 3 - Pointers scored per game.

$$\begin{aligned} f(Y_1 = y_1, ..., Y_n = y_n \mid \lambda) &= f(y_1, ..., y_n \mid \lambda) \\ &= \prod_{i=1}^{n} f(x_i \mid \lambda) \\ &= \prod_{i=1}^{n} \frac{\lambda_i^x e^{-\lambda}}{x_i!} \\ &= \frac{\lambda^{n\overline{x}} e^{-n\lambda}}{n!} \end{aligned} \tag{4}$$

And the plot of the likelihood function and the prior distributions is as follows:

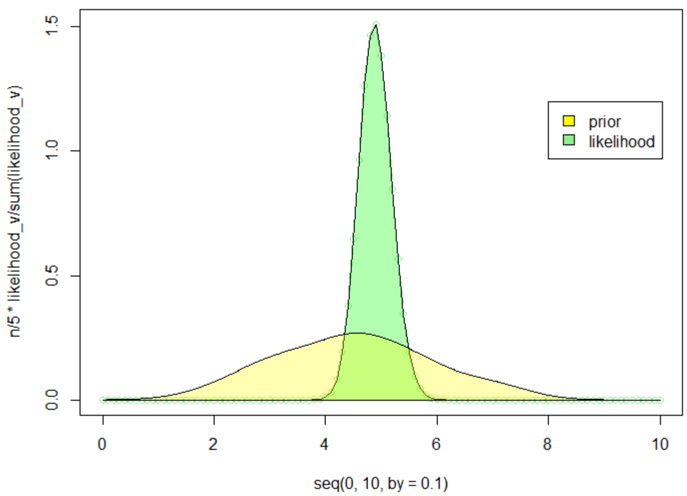

Fig. 2. Prior and Likelihood

It can be observed that the likelihood distribution is shifted to right of prior distribution. The data mean is found to be

$$\frac{1}{n}\sum_{i=1}^{n} y_i = 4.875$$

This concludes that Stephen scored more 3 - Pointer in the second season per game compared to average in first season.

V. Posterior Distribution

We combine the results from our prior data of first season scores and after finding an appropriate data likelihood function,we know that if

Prior: $\lambda \sim$ Gamma(α, β)
Data-Likelihood: $Y_i \mid \lambda \sim$ Poisson(y_i, λ)

Then

$$\lambda \mid (Y_1 = y_1, ...Y_n = y_n) \sim Gamma(\alpha', \beta')$$

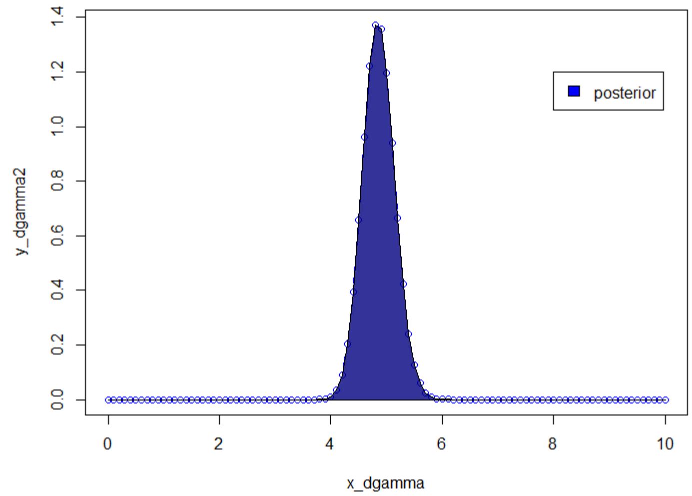

Fig. 3. Posterior Distribution

where

$$\alpha' = \alpha + \sum_{i=1}^{n} y_i \tag{5}$$

$$\beta' = \beta + n \tag{6}$$

After substituting the values for α and β calculated from R and substituting in the equation for posterior distribution,we obtain:

$$\lambda \mid (Y_1 = y_1, ...Y_n = y_n) \sim Gamma(\alpha + \sum_{i=1}^{n} y_i, \beta + n)$$

$$\lambda \mid (Y_1 = y_1, ...Y_n = y_n) \sim Gamma(284.17, 58.47)$$

where, y_i is the realised value of 3 - Pointers scored in the i-th game and n is the no of games Stephen scored 3 - Pointers.

$$E(\lambda \mid (Y_1 = y_1, ...Y_n = y_n)) = 4.85 \tag{7}$$

$$\text{Var}(\lambda) \mid (Y_1 = y_1, ...Y_n = y_n)) = 0.083 \tag{8}$$

VI. Conclusion

The final posterior mean obtained is equal to 4.85. It shows that there is a significant difference between prior and posterior mean.Posterior leans towards likelihood more and our posterior mean(4.859) is closer to data mean(4.875).

Hence,the newly obtained results imply that after performing Bayesian analysis on the data, We can conclude that the possibility of hitting 3 - Pointers in the next season are more compared to the last season.

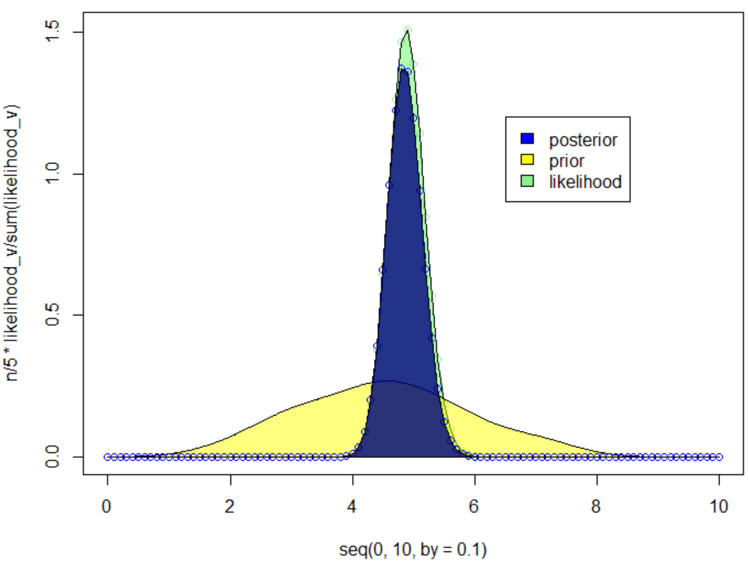

Fig. 4. Prior, Likelihood and Posterior Distribution