Real Data Analysis

MA4740 - Introduction to Bayesian Statistics

GROUP - 6

April 21, 2023

Team Members

- Varunaditya Singhal - MA20BTECH11021
- Nikhil Kongara - MA20BTECH11011
- Megavath Rajinikanth - ES20BTECH11018
- Mohd Saad - MA20BTECH11009

Introduction

Introduction

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.

Introduction

Objective

The project includes:

- Real Data Analysis on Dataset:-
  - Stephen Curry 2021 - 2023 Game log.
- Performing Poisson Gamma Bayesian Analysis on the collected Dataset.

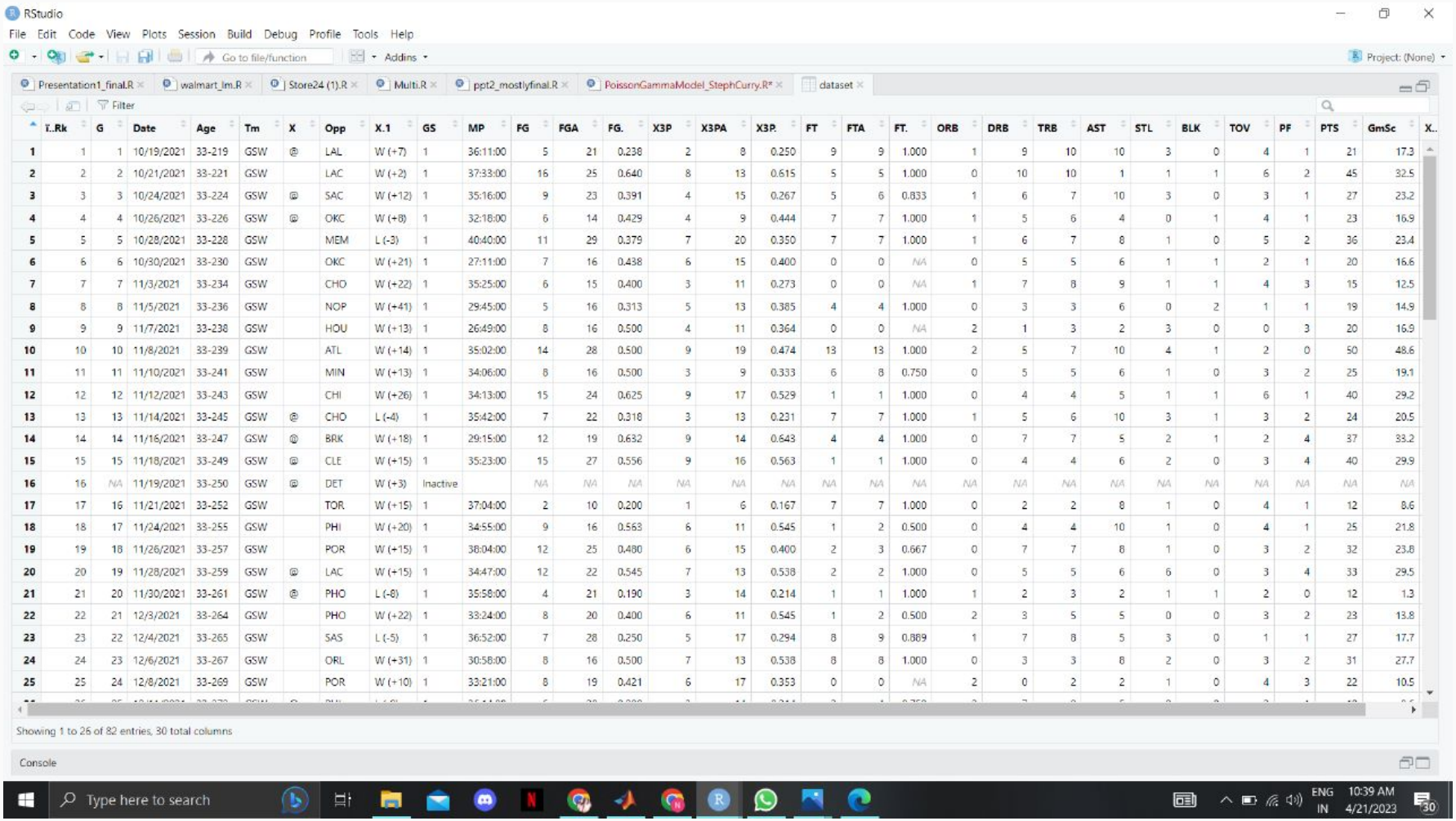
Data Collection

Data Collection

The Dataset

- The data-set was taken from Stephen Curry 2021-23 Game Log
- The data includes stats like number of field goal, number of 3 pointers, number of free throws, assists, blocks, total points,etc.
- Amongst which, we are interested in the **number of three pointers** he made in each game.

Glimpse of the Dataset



RStudio

File Edit Code View Plots Session Build Debug Profile Tools Help

Go to file/function | Addins | Project: (None)

Presentation1_final.R | walmart_lm.R | Store24 (1).R | Multi.R | ppt2_mostlyfinal.R | PoissonGammaModel_StephCurry.R* | dataset

Filter

	ï..Rk	G	Date	Age	Tm	X	Opp	X.1	GS	MP	FG	FGA	FG.	X3P	X3PA	X3P.	FT	FTA	FT.	ORB	DRB	TRB	AST	STL	BLK	TOV	PF	PTS	GmSc	X..
1	1	1	10/19/2021	33-219	GSW	@	LAL	W (+7)	1	36:11:00	5	21	0.238	2	8	0.250	9	9	1.000	1	9	10	10	3	0	4	1	21	17.3	
2	2	2	10/21/2021	33-221	GSW		LAC	W (+2)	1	37:33:00	16	25	0.640	8	13	0.615	5	5	1.000	0	10	10	1	1	1	6	2	45	32.5	
3	3	3	10/24/2021	33-224	GSW	@	SAC	W (+12)	1	35:16:00	9	23	0.391	4	15	0.267	5	6	0.833	1	6	7	10	3	0	3	1	27	23.2	
4	4	4	10/26/2021	33-226	GSW	@	OKC	W (+8)	1	32:18:00	6	14	0.429	4	9	0.444	7	7	1.000	1	5	6	4	0	1	4	1	23	16.9	
5	5	5	10/28/2021	33-228	GSW		MEM	L (-3)	1	40:40:00	11	29	0.379	7	20	0.350	7	7	1.000	1	6	7	8	1	0	5	2	36	23.4	
6	6	6	10/30/2021	33-230	GSW		OKC	W (+21)	1	27:11:00	7	16	0.438	6	15	0.400	0	0	NA	0	5	5	6	1	1	2	1	20	16.6	
7	7	7	11/3/2021	33-234	GSW		CHO	W (+22)	1	35:25:00	6	15	0.400	3	11	0.273	0	0	NA	1	7	8	9	1	1	4	3	15	12.5	
8	8	8	11/5/2021	33-236	GSW		NOP	W (+41)	1	29:45:00	5	16	0.313	5	13	0.385	4	4	1.000	0	3	3	6	0	2	1	1	19	14.9	
9	9	9	11/7/2021	33-238	GSW		HOU	W (+13)	1	26:49:00	8	16	0.500	4	11	0.364	0	0	NA	2	1	3	2	3	0	0	3	20	16.9	
10	10	10	11/8/2021	33-239	GSW		ATL	W (+14)	1	35:02:00	14	28	0.500	9	19	0.474	13	13	1.000	2	5	7	10	4	1	2	0	50	48.6	
11	11	11	11/10/2021	33-241	GSW		MIN	W (+13)	1	34:06:00	8	16	0.500	3	9	0.333	6	8	0.750	0	5	5	6	1	0	3	2	25	19.1	
12	12	12	11/12/2021	33-243	GSW		CHI	W (+26)	1	34:13:00	15	24	0.625	9	17	0.529	1	1	1.000	0	4	4	5	1	1	6	1	40	29.2	
13	13	13	11/14/2021	33-245	GSW	@	CHO	L (-4)	1	35:42:00	7	22	0.318	3	13	0.231	7	7	1.000	1	5	6	10	3	1	3	2	24	20.5	
14	14	14	11/16/2021	33-247	GSW	@	BRK	W (+18)	1	29:15:00	12	19	0.632	9	14	0.643	4	4	1.000	0	7	7	5	2	1	2	4	37	33.2	
15	15	15	11/18/2021	33-249	GSW	@	CLE	W (+15)	1	35:23:00	15	27	0.556	9	16	0.563	1	1	1.000	0	4	4	6	2	0	3	4	40	29.9	
16	16	NA	11/19/2021	33-250	GSW	@	DET	W (+3)	Inactive		NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	
17	17	16	11/21/2021	33-252	GSW		TOR	W (+15)	1	37:04:00	2	10	0.200	1	6	0.167	7	7	1.000	0	2	2	8	1	0	4	1	12	8.6	
18	18	17	11/24/2021	33-255	GSW		PHI	W (+20)	1	34:55:00	9	16	0.563	6	11	0.545	1	2	0.500	0	4	4	10	1	0	4	1	25	21.8	
19	19	18	11/26/2021	33-257	GSW		POR	W (+15)	1	38:04:00	12	25	0.480	6	15	0.400	2	3	0.667	0	7	7	8	1	0	3	2	32	23.8	
20	20	19	11/28/2021	33-259	GSW	@	LAC	W (+15)	1	34:47:00	12	22	0.545	7	13	0.538	2	2	1.000	0	5	5	6	6	0	3	4	33	29.5	
21	21	20	11/30/2021	33-261	GSW	@	PHO	L (-8)	1	35:58:00	4	21	0.190	3	14	0.214	1	1	1.000	1	2	3	2	1	1	2	0	12	1.3	
22	22	21	12/3/2021	33-264	GSW		PHO	W (+22)	1	33:24:00	8	20	0.400	6	11	0.545	1	2	0.500	2	3	5	5	0	0	3	2	23	13.8	
23	23	22	12/4/2021	33-265	GSW		SAS	L (-5)	1	36:52:00	7	28	0.250	5	17	0.294	8	9	0.889	1	7	8	5	3	0	1	1	27	17.7	
24	24	23	12/6/2021	33-267	GSW		ORL	W (+31)	1	30:58:00	8	16	0.500	7	13	0.538	8	8	1.000	0	3	3	8	2	0	3	2	31	27.7	
25	25	24	12/8/2021	33-269	GSW		POR	W (+10)	1	33:21:00	8	19	0.421	6	17	0.353	0	0	NA	2	0	2	2	1	0	4	3	22	10.5	

Showing 1 to 26 of 82 entries, 30 total columns

Console

Type here to search

ENG IN 10:39 AM 4/21/2023

Figure 1: Stephen Curry Game Log

Defining variables

The below representations of the data we used regarding the variables we learned in class.

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

$$Y \mid \lambda \sim Poisson(\lambda) \tag{2}$$

$$\lambda \mid Y \propto f(Y \mid \lambda) * f(\lambda) \tag{3}$$

Where,

λ : Average number of 3 - Pointers scored by Stephen Curry per game.

$Y =$ Data of 3 - Pointers over the n entire season

(Realised values of Y of size n, $Y_i \mid \lambda \sim Poisson(\lambda)$)

Poisson Gamma Bayesian Data Analysis

Prior Distribution

Prior Distribution

- The prior data is taken from the dataset from the years 2021 to 2022.
- We try to fit the prior data to a Gamma Distribution.
  We get $\lambda \sim \mathit{Gamma}(\alpha, \beta)$, where,

Formula

$$\mu = \frac{\alpha}{\beta}, \quad \sigma = \frac{\alpha}{\beta^2}$$

Where, α is the shape hyper-parameter and β is the scale hyper-parameter of $\mathit{Gamma}(\alpha, \beta)$.

Prior Distribution

- From the season 2021-22,Bayesian Analysis has been performed on stephen curry's performance.
- After performing the necessary calculations, we get the values of the hyper-parameters α and β for these months as

Calculations

$$\alpha = 11.17, \quad \beta = 2.47, \quad \textit{PriorMean} = 4.507$$

Prior Distribution

The corresponding Prior Distribution graphs are as follows

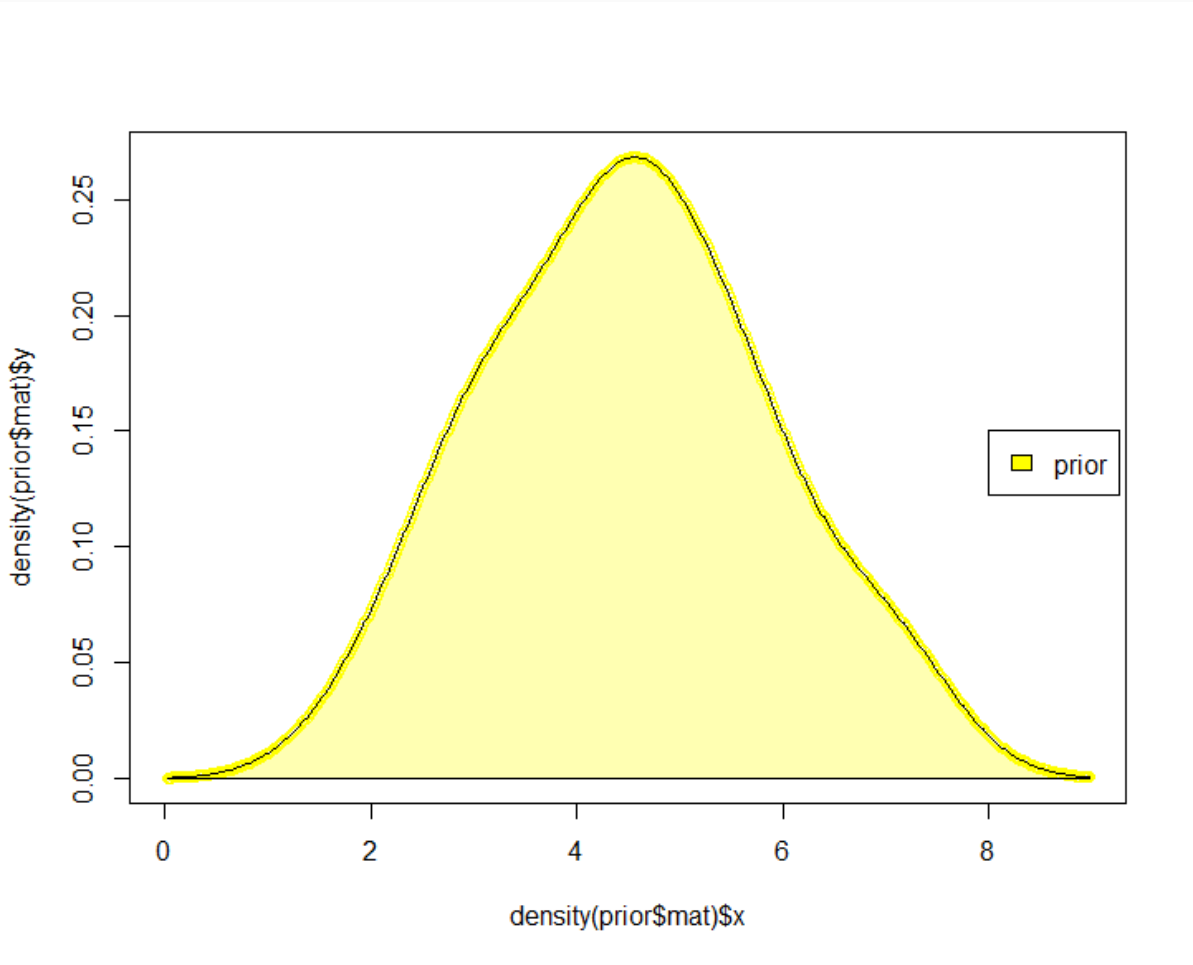


Figure 2: Prior Distribution

Data-Likelihood Function

From our likelihood data we get that

Result

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

$\sum_{i=1}^{n} y_i = 273$ and n = 56

Data-Likelihood Function

The corresponding Likelihood Function graphs are as follows

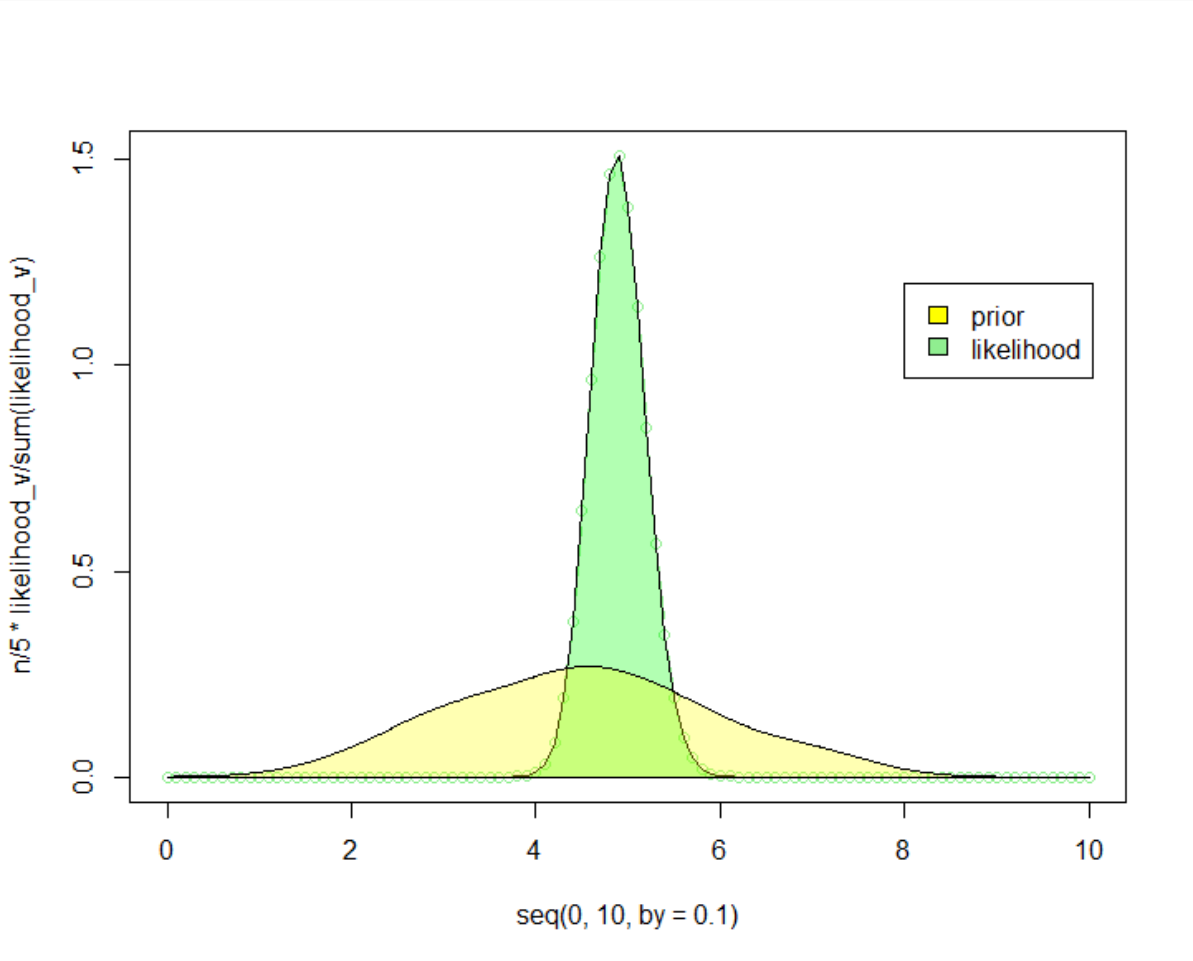


Figure 3: Prior and Likelihood

Posterior Distribution

Thus, after performing the necessary calculations, we get the following values for $Gamma(\alpha', \beta')$

Calculations

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

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

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

$$\alpha' = 284.17, \quad \beta' = 58.47, \quad PosteriorMean = 4.85$$

Posterior Distribution

The graph of Posterior is as follows

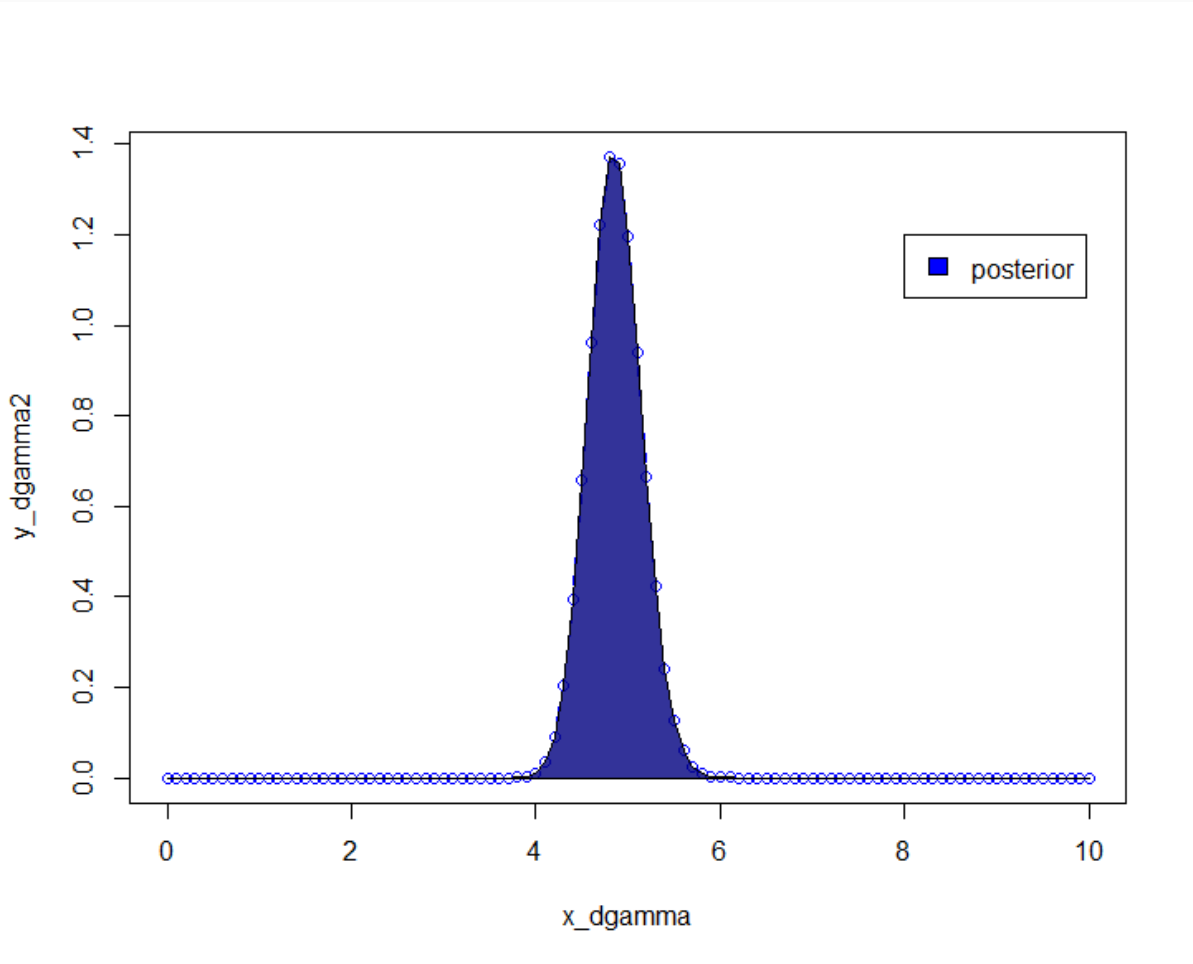


Figure 4: Posterior Distribution

Prior Likelihood and Posterior Distribution

The combined graphs of Prior, Likelihood, and Posterior are as follows

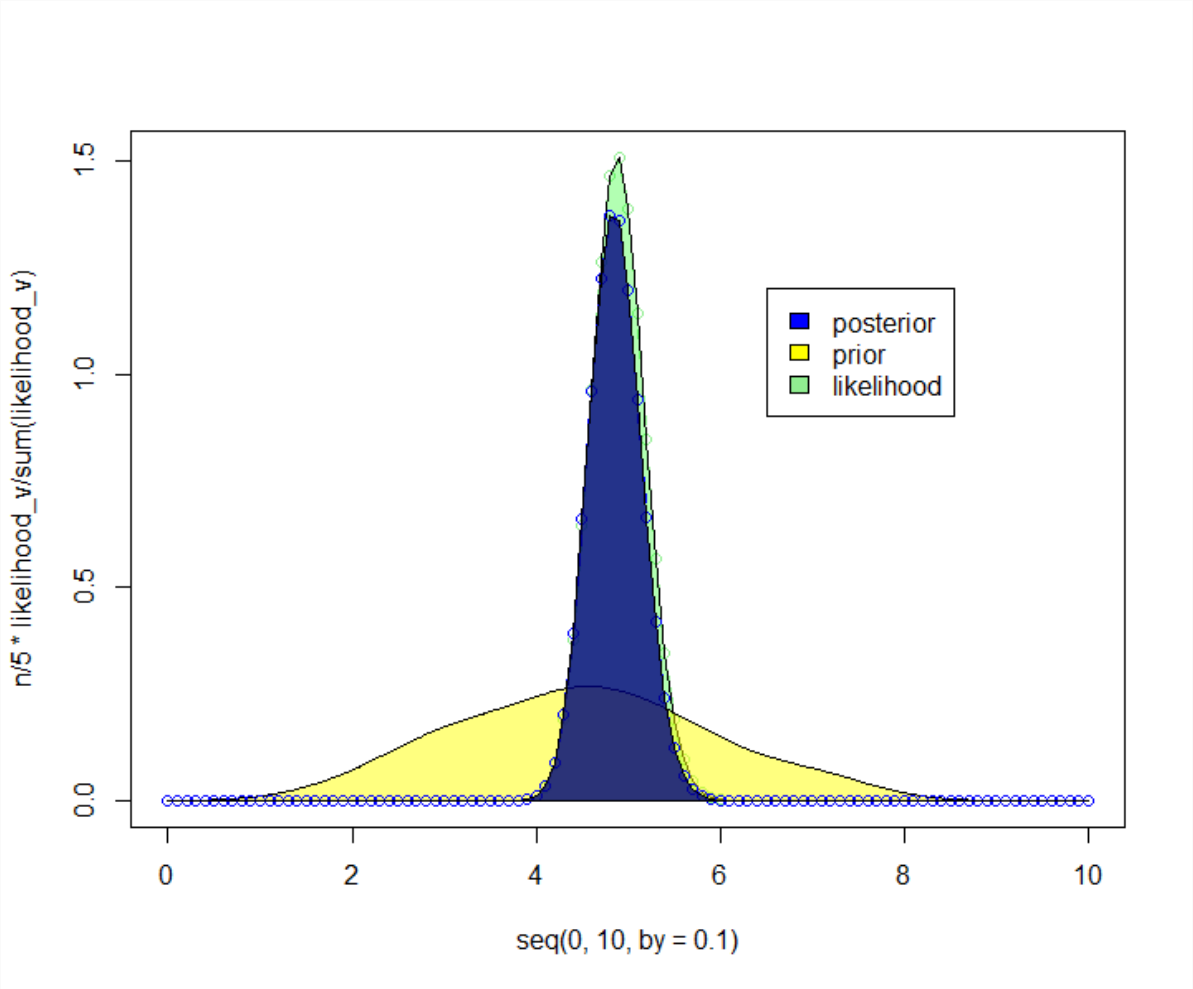


Figure 5: Posterior Distribution

Conclusion

From the above plots, we can see that the mean of our prior and posterior differ slightly whereas the variance of prior and posterior differ by a very large margin.We can thus interpret from this Bayesian model that our posterior distribution is much more accurate than the prior.

Stephen Curry is more likely to increase his 3P in the next season.

$$Prior\ std = 1.3453$$

$$Post\ std = 0.0298$$

Thank You