Exercises_4 Reading Course

Bayesian Methods

Achladianakis Minas

A report presented for the Reading Course on

Bayesian Methods

ΠΑΝΕΠΙΣΤΗΜΙΟ ΚΡΗΤΗΣ
UNIVERSITY OF CRETE

Department of Applied Mathematics

University of Crete

Greece, Dec 18, 2023

Abstract

This reading course is about Bayesian Statistical Methods.

Mostly using [DHo09] and [Cow13]

Declaration

It's Back to Bayes!

Task 1

1.1 G-L Ex_1

Let E_1, E_2, E_3 be events. Let I_1, I_2, I_3 be the corresponding indicators so that $I_1 = 1$ if E_1 occurs and $I_1 = 0$ otherwise.
i) Let $I_A = 1-(1-I_1)(1-I_2)$. Verify that I_A is the indicator for the event A, where A= ($E_1 \cup E_2$) and show that Pr(A)=Pr(E_1)+Pr(E_2)−Pr($E_1 \cup E_2$).

- I_A **is** the indicator of A:
  - If A, then E_1 **or** $E_2 \Rightarrow I_1 = 1$ **or** $I_2 = 1 \Rightarrow I_A = 1$
  - If $\neg$A, then $\neg E_1$ **and** $\neg E_2 \Rightarrow I_1 = 0$ **and** $I_2 = 0 \Rightarrow I_A = 0$
- Suppose A=E_1 and B=E_2, then, Pr($E_1 \cup E_2$)= Pr(E_1)+Pr(E_2)-Pr($E_1 \cap E_2$) holds:
  - A$\cup$B=(A\B)$\cup$(B\A)$\cup$(A$\cap$B), where (A\B)$\cap$(B\A), (B\A)$\cup$(A$\cap$B) and (A\B)$\cup$(A$\cap$B) are all empty $\Rightarrow$ Pr(A$\cup$B)=Pr(A\B)+Pr(B\A)+Pr(A$\cap$B)
  - A=(A\B)$\cup$(A$\cap$B) and B=(B\A)$\cup$(A$\cap$B), sets pairwise disjoint by nature $\Rightarrow$ Pr(A)=Pr(A\B)+Pr(A$\cap$B) and Pr(B)=Pr(B\A)+Pr(A$\cap$B) $\Rightarrow$ Pr(A)+Pr(B)=Pr(A\B)+Pr(B\A)+2Pr(A$\cap$B)=Pr(A$\cup$B)+Pr(A$\cap$B).

ii) Find a formula, in terms of I_1, I_2, I_3 for I_B, the indicator for the event B where B=$E_1 \cup E_2 \cup E_3$ and derive a formula for Pr(B) in terms of Pr(E_1), Pr(E_2), Pr(E_3), Pr($E_1 \cap E_2$), Pr($E_1 \cap E_3$), Pr($E_2 \cap E_3$), Pr($E_1 \cap E_2 \cap E_3$).

- Suppose I=1 − (1−I_1)(1−I_2)(1−I_3):
  - If B, then E_1 **or** E_2 **or** $E_3 \Rightarrow I_1 = 1$ **or** $I_2 = 1$ **or** $I_3 = 1 \Rightarrow$ I= 1
  - If $\neg$B, then $\neg E_1$ & $\neg E_2$ & $\neg E_3 \Rightarrow I_1 = 0$ & $I_2 = 0$ & $I_3 = 0 \Rightarrow$ I= 0

  Then I=I_B

- Suppose E_1=A, E_2=B and E_3=C, the formula for Pr(B) is.
  - A∪B∪C=(A\B∪C)∪(B\A∪C)∪(C\A∪A)∪(A∩B\C)∪(A∩C\B)∪ ∪(C∩B\A)∪(A∩B∩C), sets pairwise disjoint by construction, ⇒ Pr(A∪B∪C)=Pr(A\B∪C)+Pr(B\A∪C)+Pr(C\A∪B)+Pr(A∩B\C)+ +Pr(A∩C\B)+Pr(C∩B\A)+Pr(A∩B∩C)
  - A=(A\B∪C)∪(A∩C\B)∪(A∩B\C)∪(A∩B∩C), sets pairwise disjoint by nature, ⇒ Pr(A)=Pr(A\B∪C)+Pr(A∩C\B)+Pr(A∩B\C)+Pr(A∩B∩C), the equivalent holds for B & C. ⇒ Pr(A)+Pr(B)+Pr(C)= Pr(A∪B∪C)+ +Pr(A∩C\B)+Pr(A∩B\C)+Pr(B∩C\A)+Pr(A∩C∩B)
  - A∩B=(A∩B\C)∪(A∩B∩C), sets pairwise disjoint. ⇒ Pr(A∩B\C)= Pr(A∩B)-Pr(A∩B∩C), analogously for A∩C and B∩C.

⇒ Pr(A∩B∩C)=Pr(A)+Pr(b)+Pr(C)-Pr(A∩B)-Pr(A∩C)-Pr(B∩C)+Pr(A∩B∩C)

1.2 G-L Ex_2

[1]

A machine is built to make mass-produced items. Each item made by the machine has a probability p of being defective. Given the value of p, the items are independent of each other. Because of the way in which the machines are made, p could take one of several values. In fact $p = \frac{X}{100}$ where X has a discrete uniform distribution on the interval $[0, 5]$. The machine is tested by counting the number of items made before a defective is produced. Find the conditional probability distribution of X given that the first defective item is the thirteenth to be made.

Notation :

$$Item_i = I_i, \quad I_i = \begin{cases} 0, & Item_i \text{ is the first defective} \\ 1, & Item_i \text{ is } \boldsymbol{not} \end{cases}, \quad \mathrm{I}_i = \begin{cases} 1, & i \in \{1, \ldots, 12\} \\ 0, & i = 13 \end{cases}$$

Solution :

From Bayes rule :

$$posterior = \frac{prior * likelihood}{marginal} \tag{1.1}$$

- marginal= $\mathrm{p}(\mathrm{I}_{13} = 0) = \sum_{x=0}^{5} p(I_{13} = 0|X)p(X) = \frac{1}{6}\sum_{x=0}^{5}(1 - \frac{x}{100})^{12}\frac{x}{100}$

$$\Rightarrow marginal \approx 0.01614999747830071 64$$

- likelihood= $\mathrm{p}(\mathrm{I}_{13} = 0|\mathrm{X}=\mathrm{x}) = (1 - \frac{x}{100})^{12}\frac{x}{100}$
- prior= $\mathrm{p}(\mathrm{X}=\mathrm{x}) = \frac{1}{6}$, as $\mathrm{X} \sim$uniform([0,5])
- posterior= $\mathrm{p}(\mathrm{X}=\mathrm{x}|\mathrm{I}_{13} = 0) = \frac{\frac{1}{6}(1-\frac{x}{100})^{12}\frac{x}{100}}{0.016149974783007164} = \frac{\frac{1}{6}\frac{(100-x)^{12}x}{100^{13}}}{\frac{1}{6}\frac{\sum_{x=0}^{5}(100-x)^{12}x}{100^{13}}} = \frac{(100-x)^{12}x}{\sum_{x=0}^{5}(100-x)^{12}x}$

  $\Rightarrow$

x	Approximation of Posterior value
0	0
1	0.0914743297977957 5
2	0.1619644889601666
3	0.2148122118820556 3
4	0.2529249593522397
5	0.2788240100077422 6

[1] $\mathrm{Exercise}_3$ in our flyer.

1.3 G-L Ex$_3$

[2]

A crime has been committed. Assume that the crime was committed by exactly one person, that there are 1000 persons who would have committed the crime, and that, in the absence of any evidence, these people are all equally likely to be guilty of the crime.
A piece of evidence is found. It is judged that this evidence would have a probability of 0.99 of being observed if the crime was committed by a particular individual, A, and a probability of only 0.0001 of being observed if the crime was committed by any other individual.
Find the probability, given the evidence, that A committed the crime.

Notation :

$$\mathbf{E} = \begin{cases} 1, & \text{accepted} \\ 0, & \textbf{not} \end{cases} \quad \& \quad \mathbf{P}_i = \begin{cases} 1, & i \textbf{ is } \textit{the merderer} \\ 0, & i \textbf{ is not} \end{cases}, \text{ the evidence and person i.}$$

$\mathrm{P}_A \in \{\mathrm{P}_1, \ldots, \mathrm{P}_{1000}\}$, is the person A for whom E has 0.99 probability of acceptance.

Solution :

Using Bayes' rule 1.1 we compute the $\Pr_{A|E}(\mathrm{P}_A = 1|\mathrm{E})$

- marginal= $\Pr(\mathrm{E}|\mathrm{P}_A = 1)\Pr(\mathrm{P}_A{=}1)+\Pr(\mathrm{E}|\mathrm{P}_A = 0)\Pr(\mathrm{P}_A = 0)=$ $0.99\frac{1}{1000} + 0.0001(1 - \frac{1}{1000}) = 99 \cdot 10^{-5} + 999 \cdot 10^{-7}$
- prior= $\Pr(\mathrm{P}_A = 1) = \frac{1}{1000}$
- likelihood= $\Pr(\mathrm{E}|\mathrm{P}_A = 1)= 0.99$

The posterior yields:

$$\Pr_{A|E}(\mathrm{P}_A = 1|\mathrm{E}) = \frac{99 \cdot 10^{-5}}{99 \cdot 10^{-5} + 999 \cdot 10^{-7}} \approx 0.9083402146986$$

[2] Exercise$_5$ in our flyer.

1.4 G-L Ex_4

[3]

In a certain small town, there are n taxis which are clearly numbered $1, 2, \ldots, n$. Before we visit the town we do not know the value of n but our probabilities for the possible values of n are as follows:

Table of Values										
# Taxi	0	1	2	3	4	5	6	7	8	$\geq$9
Probability	0.00	0.11	0.12	0.13	0.14	0.14	0.13	0.12	0.11	0.00

On a visit to the town, we take a taxi which we assume would be equally likely to be any one of taxis 1, 2, . . . , n. It is the taxi number 5. Find our new probabilities for the value of n.

Solution : Using the Bayes' rule 1.1, we compute the probabilities.**Note:** if $n < 5 \Rightarrow p(y = 5) = 0$. So likelihood is $\frac{1}{n} \forall n \in \mathbf{N}$

- The prior is given from the table
- likelihood, $p(y|n) = \frac{1}{n}$
- marginal= $\mathrm{p}(y = 5) = \sum_n \mathrm{p}(n)\mathrm{p}(y = 5|n) = \sum_{n \geq 5} \mathrm{p}(n)\mathrm{p}(y = 5|n) = \frac{1}{5} \cdot 0.11 + \frac{1}{6} \cdot 0.12 + \frac{1}{7} \cdot 0.13 + \frac{1}{8} \cdot 0.14) = 0.08055952$

The result:

$$p(n|y = 5) = \frac{p(y = 5|n) \cdot p(n)}{p(y = 5)} = 12.41318 * p(y = 5|n) * p(n)$$

So for $n < 5$ and $n > 9$, $p(n|y = 5) = 0$ and the rest are:

Table of Values						
# Taxi	≤ 4	5	6	7	8	$\geq$9
Probability	0	0.3475691	0.2689523	0.2127974	0.1706813	0

[3] $Exercise_7$ in our flyer.

Task 2

2.1 Sensitivity Analysis

4

It is sometimes useful to express the parameters a and b in a beta distribution in terms of $\theta_0 = \frac{a}{(a+b)}$ and $n_0 = a + b$, so that $a = \theta_0 n_0$ and $b = (1 - \theta_0)n_0$. Reconsidering the sample survey data in the previous exercise for each combination of $\theta_0 \in \{0.1, 0.2, \ldots, 0.9\}$ and $n_0 \in \{1, 2, 8, 16, 32\}$ find the corresponding a, b values and compute $\Pr(\theta > 0.5 | \sum_1^{100} Y_i = 57)$ using a beta(a,b) prior distribution for θ. Display the results with a contour plot, and discuss how the plot could be used to explain to someone whether or not they should believe that $\theta > 0.5$, based on the data that $\sum_1^{100} Y_i = 57$

Solution :

To compute the probability, $\Pr(\theta > 0.5 | \sum_1^{100} Y_i = 57)$ we will use the conjugacy of the beta family priors when used upon the binomial model [5]:

$$\text{if } \begin{cases} \theta \sim beta(a, b) \\ Y \sim binomial(N, \theta) \end{cases} \text{, then } \{\theta | Y = y\} \sim beta(a + y, b + N - y) \qquad (2.2)$$

Based on the contour plot (2.1), we can easily persuade someone that $\theta > 0.5$. Image the threshold line at $\theta_0 = 0.4$ and observe that in cases $\theta_0 > 0.5$ the value of the probability that $\theta > 0.5$ based on our data is greater than 0.8, even for $\theta_0 < 0.5$ that value is still over 0.4 in most cases.

R code:

```
N = 100 ; th_0 = rev(seq(0.1, 0.9, by = 0.1))
n_0 = c(1,2,8,16,32)
y = 57
```

[4] Exercise 3.2 of [DHo09]

[5] A more analytical explanation can be found in [DHo09]pg.37

```
A <- list()
B <- list()
for (i in th_0){for (j in n_0){
  a <- i*j
  A <-append(A,a)
  b <- (1-i)*j
  B <-append(B,b)}}
results <- list()

for (i in 1:45){
  posterior_prob <-
    pbeta(0.5, shape1 = A[[i]] +
            y,shape2 = B[[i]] + N - y,lower.tail = FALSE)
  result <- data.frame(posterior_prob)
  results <- append(results, list(result))}

matrix<- matrix(unlist(results),
                nrow=9,ncol=5, byrow = TRUE )
library(plotly)
fig <- plot_ly(
  x = n_0,
  y = th_0,
  z = matrix,
  type = "contour" )
```

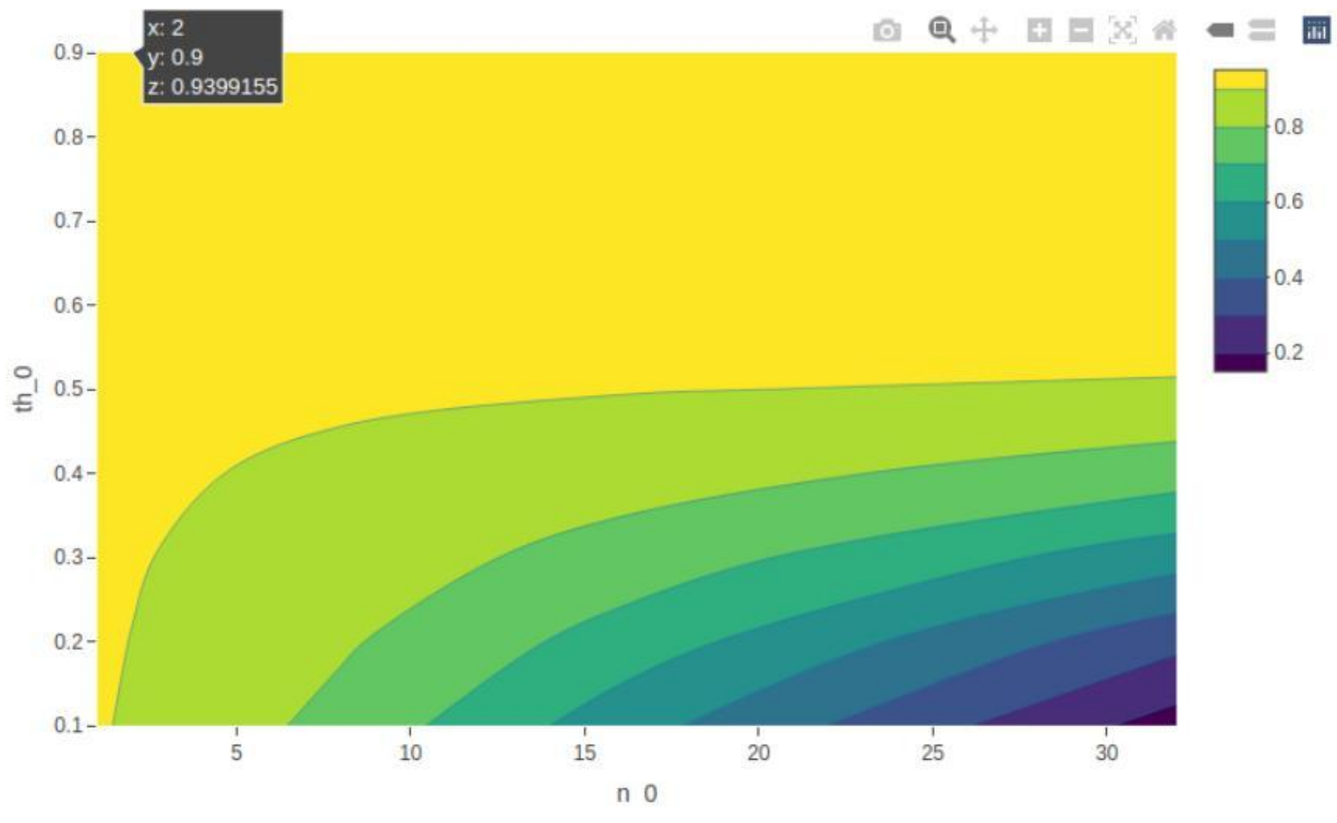


Figure 2.1: Contour plot of exercise 2 chapter 3 [DHo09]

2.2 Tumor Count

[6]

Tumor counts: A cancer laboratory is estimating the rate of tumorigenesis in two strains of mice, A and B. They have tumor count data for 10 mice in strain A and 13 mice in strain B. Type A mice have been well studied, and information from other laboratories suggests that type A mice have tumor counts that are approximately Poisson-distributed with a mean of 12. Tumor count rates for type B mice are unknown, but type B mice are related to type A mice. The observed tumor counts for the two populations are:

$$\begin{cases} y_A = (12, 9, 12, 14, 13, 13, 15, 8, 15, 6) \\ y_B = (11, 11, 10, 9, 9, 8, 7, 10, 6, 8, 8, 9, 7) \end{cases}$$

a) Find the posterior distributions, means, variances, and 95% quantile-based confidence intervals for θ_A and θ_B, assuming a Poisson sampling distribution for each group and the following prior distribution:
$\theta_A \sim \text{Gamma}(120, 10)$, $\theta_B \sim \text{Gamma}(12, 1)$, $p(\theta_A, \theta_B) = p(\theta_A) \times p(\theta_B)$.
b) Compute and plot the posterior expectation of θ_B under the prior distribution $\theta_B \sim gamma(12 \times n_0, n0)$ for each value of $n_0 \in \{1, 2, ..., 50\}$. Describe what sort of prior beliefs about θ_B would be necessary in order for the posterior expectation of θ_B to be close to that of θ_A.
c) Should knowledge about population A tell us anything about population B? Discuss whether or not it makes sense to have $p(\theta_A, \theta_B) = p(\theta_A) \times p(\theta_B)$.

Solution

a) Based on the information the exercise provides:

$$y_{A_1}, \ldots, y_{A_{10}} | \theta_A \sim Poisson(\theta_A) \ , \ \sum_{i=1}^{10} y_{a_i} = 117$$

$$y_{b_1}, \ldots, y_{b_{13}} | \theta_B \sim Poisson(\theta_B) \ , \ \sum_{j=1}^{13} y_{b_j} = 113$$

Gamma is conjugate prior for the poison model, thus: [7],

$$\theta_A \mid \mathbf{y}_a \sim \text{Gamma}(120 + 117, 10 + 10) = \text{Gamma}(237, 20)$$

$$\theta_B \mid \mathbf{y}_b \sim \text{Gamma}(12 + 113, 1 + 13) = \text{Gamma}(125, 14)$$

[6] [DHo09]pg228 Exercise 3.3
[7] For more information see [DHo09]pg.46

$$\mathbb{E}(\theta_A) = 237/20 = 11.85\ , \quad \mathbb{E}(\theta_B) = 125/14 = 8.92$$

$$\mathrm{Var}(\theta_A) = 237/400 = 0.593\ , \quad \mathrm{Var}(\theta_B) = 125/196 = 0.638$$

With 95% quantile-based confidence intervals 10.38924 13.40545 and 7.432064 10.560308, respectively (provided by the R function qgamma)

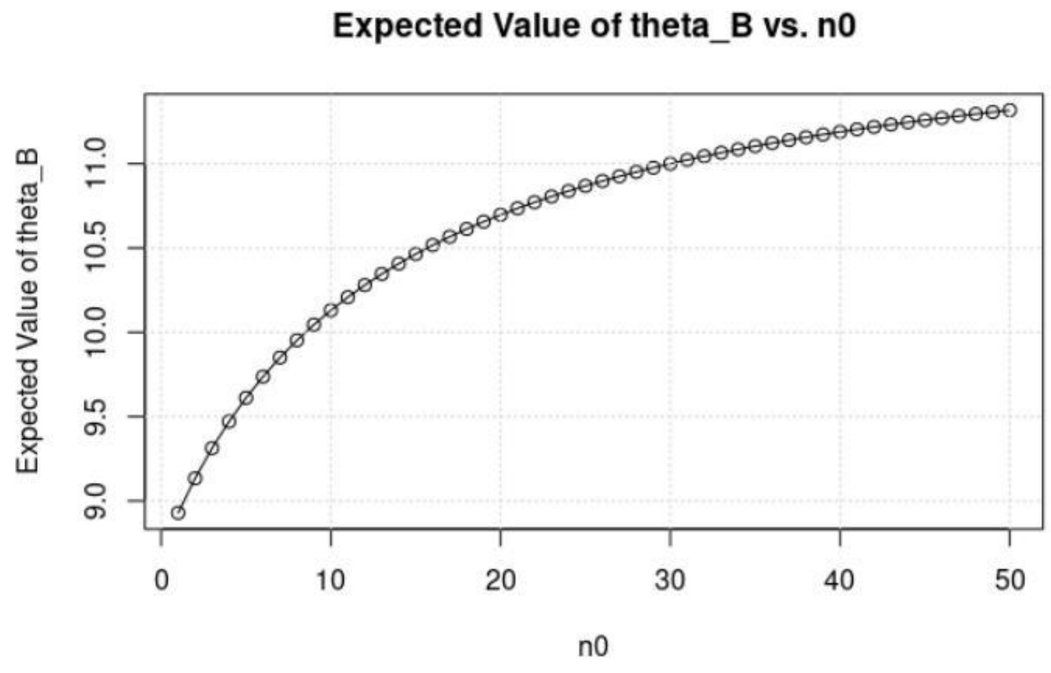
b)The posterior distributions of θ_B are:

$$\theta_b|y_b \sim Gamma(125, 14)\ , \quad \theta_b|y_b \sim Gamma(137, 15)$$
$$\theta_b|y_b \sim Gamma(149, 16)\ , \quad \theta_b|y_b \sim Gamma(161, 17)$$
$$\theta_b|y_b \sim Gamma(173, 18)\ , \quad \theta_b|y_b \sim Gamma(185, 19)$$
$$\theta_b|y_b \sim Gamma(197, 20)\ , \quad \theta_b|y_b \sim Gamma(209, 21)$$
$$\theta_b|y_b \sim Gamma(221, 22)\ , \quad \theta_b|y_b \sim Gamma(233, 23)$$
$$\theta_b|y_b \sim Gamma(245, 24)\ , \quad \theta_b|y_b \sim Gamma(257, 25)$$
$$\theta_b|y_b \sim Gamma(269, 26)\ , \quad \theta_b|y_b \sim Gamma(281, 27)$$
$$\theta_b|y_b \sim Gamma(293, 28)\ , \quad \theta_b|y_b \sim Gamma(305, 29)$$
$$\theta_b|y_b \sim Gamma(317, 30)\ , \quad \theta_b|y_b \sim Gamma(329, 31)$$
$$\theta_b|y_b \sim Gamma(341, 32)\ , \quad \theta_b|y_b \sim Gamma(353, 33)$$
$$\theta_b|y_b \sim Gamma(365, 34)\ , \quad \theta_b|y_b \sim Gamma(377, 35)$$
$$\theta_b|y_b \sim Gamma(389, 36)\ , \quad \theta_b|y_b \sim Gamma(401, 37)$$
$$\theta_b|y_b \sim Gamma(413, 38)\ , \quad \theta_b|y_b \sim Gamma(425, 39)$$
$$\theta_b|y_b \sim Gamma(437, 40)\ , \quad \theta_b|y_b \sim Gamma(449, 41)$$
$$\theta_b|y_b \sim Gamma(461, 42)\ , \quad \theta_b|y_b \sim Gamma(473, 43)$$
$$\theta_b|y_b \sim Gamma(485, 44)\ , \quad \theta_b|y_b \sim Gamma(497, 45)$$
$$\theta_b|y_b \sim Gamma(509, 46)\ , \quad \theta_b|y_b \sim Gamma(521, 47)$$
$$\theta_b|y_b \sim Gamma(533, 48)\ , \quad \theta_b|y_b \sim Gamma(545, 49)$$
$$\theta_b|y_b \sim Gamma(557, 50)\ , \quad \theta_b|y_b \sim Gamma(569, 51)$$
$$\theta_b|y_b \sim Gamma(581, 52)\ , \quad \theta_b|y_b \sim Gamma(593, 53)$$
$$\theta_b|y_b \sim Gamma(605, 54)\ , \quad \theta_b|y_b \sim Gamma(617, 55)$$
$$\theta_b|y_b \sim Gamma(629, 56)\ , \quad \theta_b|y_b \sim Gamma(641, 57)$$
$$\theta_b|y_b \sim Gamma(653, 58)\ , \quad \theta_b|y_b \sim Gamma(665, 59)$$
$$\theta_b|y_b \sim Gamma(677, 60)\ , \quad \theta_b|y_b \sim Gamma(689, 61)$$
$$\theta_b|y_b \sim Gamma(701, 62)\ , \quad \theta_b|y_b \sim Gamma(713, 63)$$

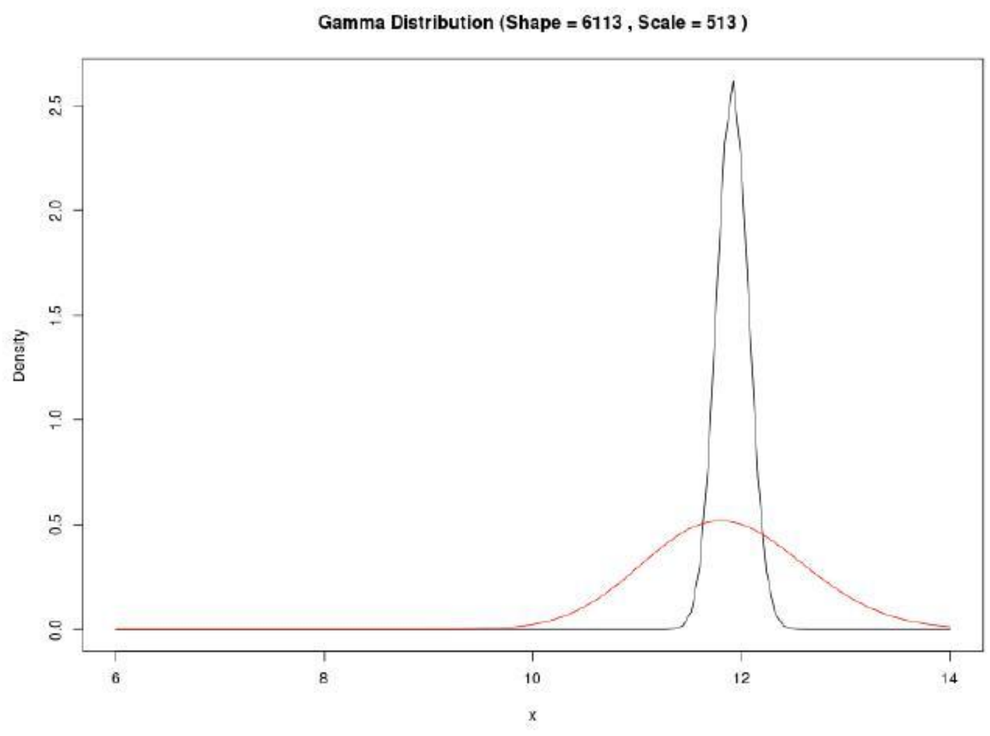
The plot (2.2a) is created with R:

```
plot(n0, shape / scale, type="o", xlab="n0",
  ylab="Expected Value of theta_B",
  main="Expected Value of theta_B vs. n0")
grid()
```

c) Knowing that the population of mice B are related to A, provides an insight and that knowledge should be used, thought in a mixture prior model and with relevant weight depending on the relation between them, so we can get the best from both words. Thus, $p(\theta_A, \theta_B) \approx p(\theta_A) \times p(\theta_B)$ could make sense only if their relationship is weak. In our case thought the independence feels natural judging by their posterior distributions for $n_0 = 50$, plot (2.2b) .



(a) Plot of expected values



(b) Posterior relationship for large n_0

2.3 Posterior prediction

[8]

Consider a pilot study in which $n_1 = 15$ children enrolled in special education classes were randomly selected and tested for a certain type of learning disability. In the pilot study, $y_1 = 2$ children tested positive for the disability.

a) Using a uniform prior distribution, find the posterior distribution of θ, the fraction of students in special education classes who have the disability. Find the posterior mean, mode and standard deviation of θ, and plot the posterior density.

Researchers would like to recruit students with disabilities to participate in a long-term study, but first, they need to make sure they can recruit enough students. Let $n_2 = 278$ be the number of children in special education classes in this particular school district, and let Y_2 be the number of students with a disability.

b) Find $\Pr(Y_2 = y_2|Y_1 = 2)$, the posterior predictive distribution of Y_2, as follows:

- Discuss what assumptions are needed about the joint distribution of (Y_1, Y_2) such that the following is true:

$$\Pr(Y_2 = y_2|Y_1 = 2) = \int_0^1 \Pr(Y_2 = y_2|\theta)p(\theta|Y_1 = 2)d\theta. \tag{2.3}$$

- Now plug in the forms for $\Pr(Y_2 = y_2|\theta)$ and $\mathrm{p}(\theta|Y_1 = 2)$ in the above integral.
- Figure out what the above integral must be by using the calculus result discussed in Section 3.1 [DHo09].

c) Plot the function $\Pr(Y_2 = y_2|Y_1 = 2)$ as a function of y_2. Obtain the mean and standard deviation of Y2, given Y1 = 2.

d)The posterior mode and the MLE (maximum likelihood estimate, see Exercise 3.14[DHo09]) of θ, based on data from the pilot study, are both $\hat{\theta} = \frac{2}{15}$. Plot the distribution $\Pr(Y_2 = y_2|\theta = \hat{\theta})$, and find the mean and standard deviation of Y_2 given $\theta = \hat{\theta}$. Compare these results to the plots and calculations in c) and discuss any differences. Which distribution for Y2 would you use to make predictions, and why?

Solution

a) Assuming a uniform prior $p(\theta) = 1$ and the binomial likelihood the posterior is

[8] [DHo09]pg.229 Exercise 3.7

the beta$(1+y, 16-y)$. As the uniform is beta$(1,1)$ and the beta is a conjugate prior to the binomial model.

There is a closed form for calculating the mean $\mathbb{E}(\theta) = \frac{\alpha}{\alpha+\beta} = \frac{1+y}{1+N} = \frac{3}{17}$, the mode= $\frac{\alpha-1}{\alpha+\beta-2} = \frac{2}{15}$ and the standard deviation is $sd = \sqrt{\frac{\alpha\cdot\beta}{(\alpha+\beta)^2\cdot(\alpha+\beta+1)}} = 0.08985443$ for any beta distribution.

Solution for b):

To find $\Pr(Y_2 = y_2|Y_1 = 2)$, the posterior predictive distribution of Y_2 you can follow these steps. However, you'll need to make the assumption of Conditional Independence, which implies that:

$$\Pr(Y_2 = y_2 \mid \theta, Y_1 = 2) = \Pr(Y_2 = y_2 \mid \theta).$$

So that:

$$\Pr(Y_2 = y_2|Y_1 = 2) = \int_0^1 \Pr(Y_2 = y_2|\theta)p(\theta|Y_1 = 2)d\theta$$

Computing $\Pr(Y_2 = y_2|\theta)$:

$$\Pr(Y_2 = y_2|\theta) \sim \text{dbinom}(y_2, n_2 = 278, \theta) \Rightarrow$$
$$\Pr(Y_2 = y_2|\theta) = \binom{278}{y_2}\theta^{y_2}(1-\theta)^{278-y_2}$$

Thus the integral is:

$$\begin{aligned}
\Pr(Y_2 = y_2|\theta) &= \int_0^1 \binom{278}{y_2}\theta^{y_2}(1-\theta)^{278-y_2} \cdot \frac{\Gamma(17)}{\Gamma(3)\cdot\Gamma(14)} \cdot \theta^2(1-\theta)^{13}d\theta \\
&= \int_0^1 \binom{278}{y_2} \cdot \frac{\Gamma(17)}{\Gamma(3)\cdot\Gamma(14)} \cdot \theta^{y_2+2}(1-\theta)^{291-y_2}d\theta \\
&= \binom{278}{y_2} \cdot \frac{\Gamma(17)}{\Gamma(3)\cdot\Gamma(14)} \int_0^1 \theta^{y_2+2}(1-\theta)^{291-y_2}d\theta \\
&= \binom{278}{y_2} \cdot \frac{\Gamma(17)}{\Gamma(3)\cdot\Gamma(14)\cdot\binom{289}{y_2+2}} \int_0^1 \binom{289}{y_2+2}\theta^{y_2+2}(1-\theta)^{289-(y_2+2)}d\theta \\
&\Rightarrow \Pr(Y_2 = y_2|\theta) = \frac{\binom{278}{y_2}}{\binom{289}{y_2+2}} \cdot \frac{\Gamma(17)}{\Gamma(3)\cdot\Gamma(14)} \cdot 1
\end{aligned}$$

Then we can proceed with either of these methods:

- **The Analytical**: The integral of the binomial distribution formed is 1.

$$\Pr(Y_2 = y_2|\theta) = \binom{278}{y_2} \cdot \frac{\Gamma(17)}{\Gamma(3)\cdot\Gamma(14)\cdot\binom{289}{y_2+2}} \int \binom{289}{y_2+2} \theta^{y_2+2}(1-\theta)^{289-(y_2+2)} d\theta$$

$$\Rightarrow \Pr(Y_2 = y_2|\theta) = \frac{\binom{278}{y_2}}{\binom{289}{y_2+2}} \cdot \frac{\Gamma(17)}{\Gamma(3)\cdot\Gamma(14)} \cdot 1$$

$$= \frac{(y_2+2)(y_2+1)(291-y_2)\ldots(279-y_2)}{289\times\cdots\times 279} \cdot \frac{\Gamma(17)}{\Gamma(3)\cdot\Gamma(14)}$$

- **The implementation method**[9]:

$$\text{Posterior predictive model} = \text{model(posterior distribution)} \tag{2.4}$$

c) To find the mean and standard deviation of $Y_2|Y_1 = 2$ we can use the following formulas:

For the mean value:

$$\mathrm{E}[Y_2|Y_1 = 2] = \sum_{y_2=0}^{278} y_2 \cdot \Pr(Y_2 = y_2|Y_1 = 2)$$

For the variance:

$$\mathrm{Var}[Y_2|Y_1 = 2] = \sum_{y_2=0}^{278} (y_2 - \mathrm{E}[Y_2|Y_1 = 2])^2 \cdot \Pr(Y_2 = y_2|Y_1 = 2)$$

For the standard deviation:

$$\mathrm{SD}[Y_2|Y_1 = 2] = \sqrt{\mathrm{Var}[Y_2|Y_1 = 2]}$$

The plot 2.3, mean, sd and variance are calculated in R the Monte Carlo approximation method, utilizing the implementation method.

```
# Load necessary library
library(ggplot2)

# Define the parameters for the posterior distribution
alpha <- 3; beta <- 14
```

[9] I am still uncertain of how and why this holds, I would very much appreciate any reading material you can recommend that could help me understand why that holds.

```
post_pred_dist <- function(y2) {
  # The likelihood function (binomial distribution)
  likelihood <- function(theta) {
    dbinom(y2, size = 278, prob = theta)}

  posterior <- function(theta) {
    dbeta(theta, shape1 = alpha, shape2 = beta)}

  integrand <- function(theta) {
    likelihood(theta) * posterior(theta)}

  #posterior predictive probability
  integrate(integrand, lower = 0, upper = 1)$value}

y2_values <- 0:278

post_pred_probs <- sapply(y2_values, post_pred_dist)

ggplot(data = data.frame(
  y2 = y2_values,
  prob = post_pred_probs), aes(x = y2, y = prob)) +
  geom_line() +
  labs(title = "Posterior Predictive Distribution",
       x = expression(Y[2]), y = "Probability") +
  theme_minimal()
```

The approximating results yield, mean$\approx$ 49.05882, variance$\approx$ 662.1337, and sd = 25.73196.

d) The plot 2.4 shows a shift towards lower values of y_2 as we expected due to the nature of θ_{hat}, which has a low value and is more concentrated around it's mode as we expected and it is provided by the R code:

```
theta_hat <- 2 / 15; n2 <- 278
y2_values <- 0:n2
```

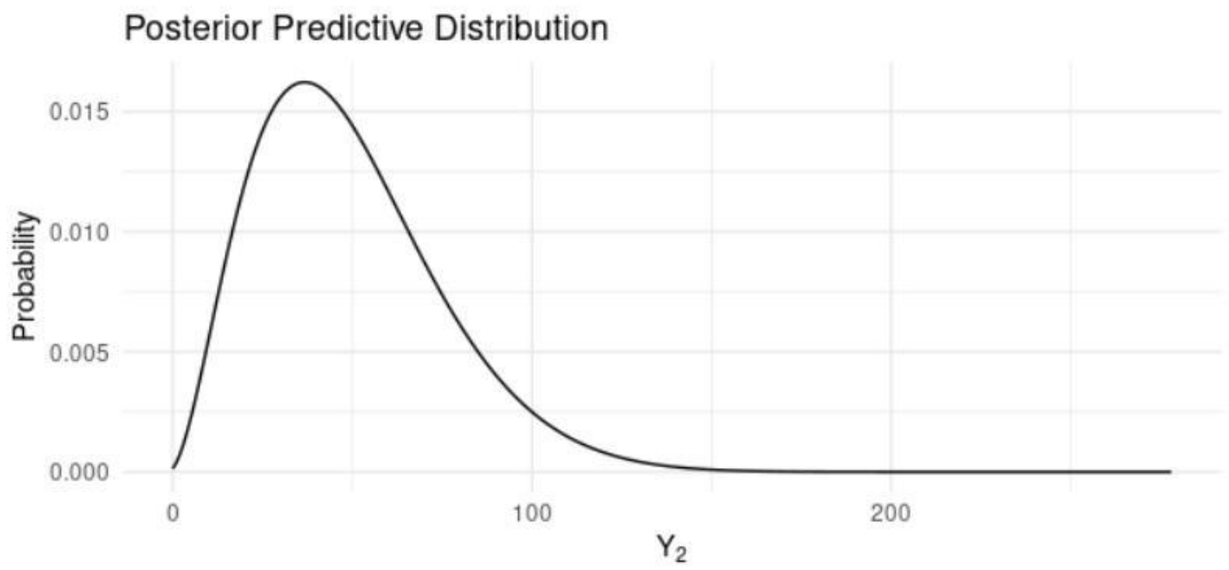


Figure 2.3: Posterior Predictive Distribution

```
probs <- dbinom(y2_values, n2, theta_hat)

mean_Y2_given_theta_hat <- n2 * theta_hat
sd_Y2_given_theta_hat <- sqrt(
  n2 * theta_hat * (1 - theta_hat))

plot(y2_values, probs, type = "h", lwd = 2,
col = "blue",
main = "Distribution Pr(Y2 = y2 | theta = theta_hat)",
xlab = "y2", ylab = "Probability")
```

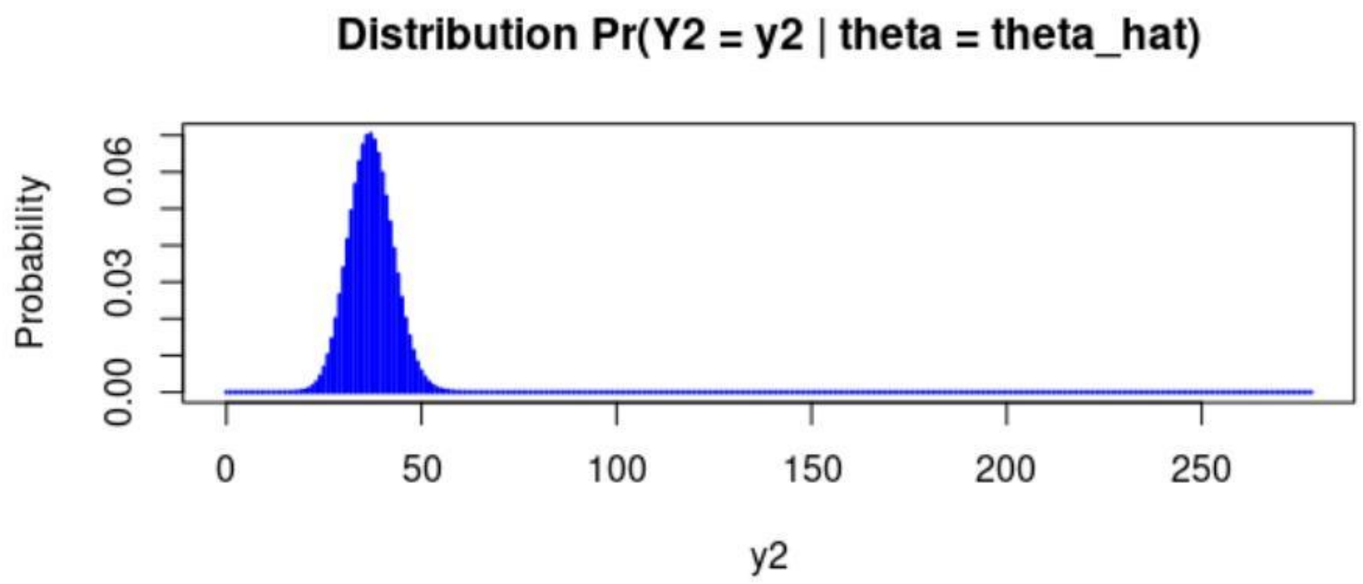


Figure 2.4: Probability plot.

2.4 Coins

10

Diaconis and Ylvisaker (1985) suggest that coins spun on a flat surface display long-run frequencies of heads that vary from coin to coin. About 20% of the coins behave symmetrically, whereas the remaining coins tend to give frequencies of $\frac{1}{3}$ or $\frac{2}{3}$.

a) Based on the observations of Diaconis and Ylvisaker, use an appropriate mixture of beta distributions as a prior distribution for θ, the long-run frequency of heads for a particular coin. Plot your prior.

b) Choose a single coin and spin it at least 50 times. Record the number of heads obtained. Report the year and denomination of the coin.

c) Compute your posterior for θ, based on the information obtained in b).

d) Repeat b) and c) for a different coin, but possibly using a prior for θ that includes some information from the first coin. Your choice of a new prior may be informal but needs to be justified. How the results from the first experiment influence your prior for the θ of the second coin may depend on whether or not the two coins have the same denomination, have a similar year, etc. Report the year and denomination of this coin.

Solution for a):

Creating and Plotting the Prior Distribution. Based on the observations of Diaconis and Ylvisaker, we can construct a prior mixture of beta distributions as:

$$\text{Prior}(\theta) = 0.20 \cdot \text{Beta}(1,1) + 0.40 \cdot \text{Beta}(2,1) + 0.40 \cdot \text{Beta}(2,3)$$

Where:

- $0.20 \cdot \text{Beta}(1,1)$ represents the 20% of coins that behave symmetrically (uniform distribution).
- $0.40 \cdot \text{Beta}(2,1)$ represents the 40% of coins that give frequencies of $\frac{1}{3}$ for heads.
- $0.40 \cdot \text{Beta}(2,3)$ represents the 40% of coins that give frequencies of $\frac{2}{3}$ for heads.

We will plot the prior in R.

```
library(ggplot2)
```

```
alpha1 <- 1; beta1 <- 1; alpha2 <- 2
```

[10] [DHo09]pg.231 Exercise 3.8

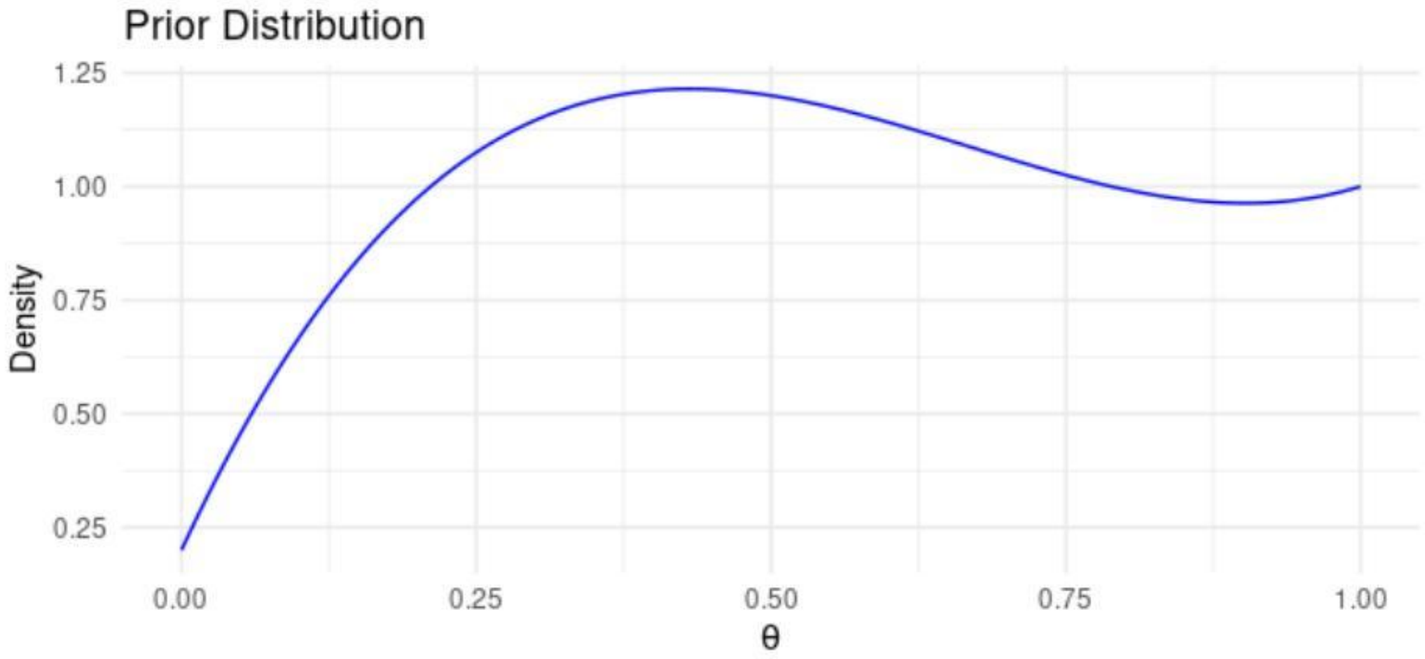


Figure 2.5: Prior plot.

```
beta2 <- 1; alpha3 <- 2; beta3 <- 3

theta <- seq(0, 1, length.out = 1000)

density1 <- 0.20 * dbeta(theta, alpha1, beta1)
density2 <- 0.40 * dbeta(theta, alpha2, beta2)
density3 <- 0.40 * dbeta(theta, alpha3, beta3)

mixture_density <- density1 + density2 + density3

data <- data.frame(theta, mixture_density)

ggplot(data, aes(x = theta, y = mixture_density)) +
  geom_line(color = "blue") +
  labs(title = "Prior Distribution",
  x = expression(theta), y = "Density") +
  theme_minimal()
```

b) I chose a single 2-euro coin minted in the year 2007 and spun it 50 times. The outcome was the coin landed heads 21 times. This data will be used in the next part **c)** to update our beliefs about the parameter θ, which represents the long-run frequency of obtaining heads for this particular coin.

c) To find the posterior distribution for θ, we update our prior distribution with the data from the experiment (21 heads in 50 spins). The likelihood function, which follows a binomial distribution, is given by:

$$\text{likelihood}(\theta) = \text{dbinom}(21, 50, \theta)$$

We then calculate the unnormalized posterior values at each point in a grid of θ values ranging from 0 to 1. The unnormalized posterior is given by:

$$\text{posterior}(\theta|y) \propto \sum_{i=1} w_i \cdot \text{dbeta}(\theta, a_i, b_i) \cdot \text{likelihood}(y|\theta)$$

$$\text{posterior}(\theta|y) \propto \sum_{i=1}^{3} w_i^{'} \cdot \text{dbeta}(\theta, a_i + y, b_i + n - y)$$

Where w_i are the initial weights and (a_i, b_i) are defined as per the prior distribution Diaconis and Ylvisaker suggested. The $w_i^{'}$ are the weight of these beta distributions as described in the later Task.

Finally, we plot the posterior distribution using a line plot where the x-axis represents θ and the y-axis represents the density.

The R code used to the plot (2.6):

```
library(ggplot2)

w_i=c(0.2,0.4,0.4); c_i=c(1,2,12)
K_i=c(factorial(51)/(factorial(21)*factorial(29)),
      factorial(52)/(factorial(22)*factorial(29)),
      factorial(54)/(factorial(22)*factorial(31)))

w_inew=w_i*c_i/K_i

w_inew_norm=w_inew/sum(w_inew)

set.seed(123)
f_w=w_inew_norm

post=f_w[1]*rbeta(50000,22,30)+
  f_w[2]*rbeta(50000,23,30)+
  f_w[3]*rbeta(50000,23,32)
```

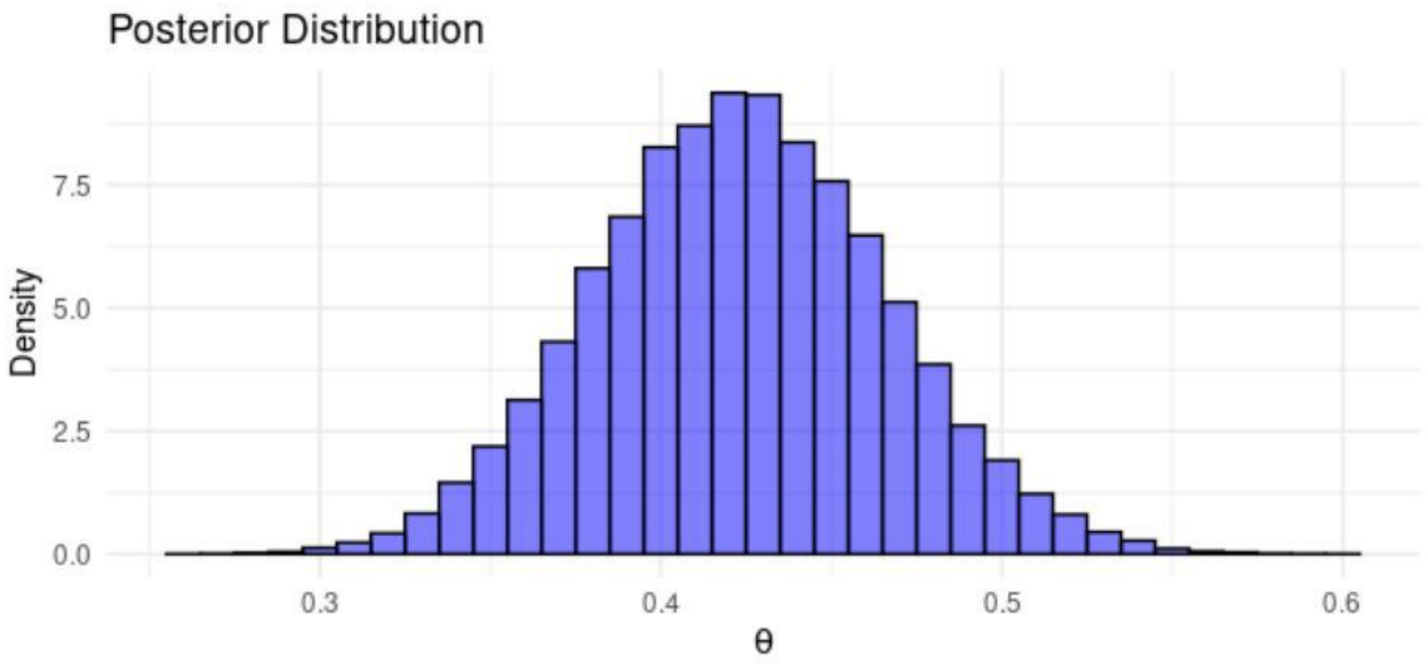


Figure 2.6: Posterior plot.

```
data_post <- data.frame(theta = post)


ggplot(data = data_post, aes(x = theta)) +
  geom_histogram(binwidth = 0.01,
                 fill = "blue", alpha = 0.5,
                 color = "black", aes(y = ..density..)) +
  labs(title = "Posterior Distribution", x = expression(theta),
       y = "Density") +
  theme_minimal()
```

d) Using a 10-cent coin, minted in the year 2020 was spun 50 times, the outcome was 31 heads. With this data, we can proceed to compute the posterior distribution for θ, the long-run frequency of heads for this particular coin, using a suitable prior which might incorporate information gleaned from the first experiment. Based on the coin being immensely different in weight and metals the knowledge from the first could be incorporated, though with great uncertainty thus I propose a weight of 0.1 and keep the rest mixture with the weights redefined to sum up in 1.

My prior:

$$\text{prior} = \frac{1}{6}\text{beta}(1,1) + \frac{11}{30}\text{beta}(2,1) + \frac{11}{30}\text{beta}(2,3) + 0.1\text{post}_{coin_1}$$

$$\Rightarrow \text{prior} = \frac{1}{6}\text{beta}(1,1) + \frac{11}{30}\text{beta}(2,1) + \frac{11}{30}\text{beta}(2,3)$$

$$+0.1(\frac{w_1^{'}}{\sum_{i=1}^{3} w_i^{'}} beta(22,30) + \frac{w_2^{'}}{\sum_{i=1}^{3} w_i^{'}} beta(23,30) + \frac{w_3^{'}}{\sum_{i=1}^{3} w_i^{'}} beta(23,32))$$

The R code for the plot (2.7):

```
# post and f_w from before:

w_i=c(1/6,11/30,11/30,f_w[1],f_w[2],f_w[3])
c_i=c(1,2,12,
      factorial(51)/(factorial(21)*factorial(29)),
      factorial(52)/(factorial(22)*factorial(29)),
      factorial(54)/(factorial(22)*factorial(31)))

K_i=c(factorial(51)/(factorial(31)*factorial(19)),
      factorial(52)/(factorial(32)*factorial(19)),
      factorial(54)/(factorial(32)*factorial(21)),
      factorial(101)/(factorial(52)*factorial(48)),
      factorial(102)/(factorial(53)*factorial(48)),
      factorial(104)/(factorial(53)*factorial(50)))

w_inew=w_i*c_i/K_i
w_inew_norm=w_inew/sum(w_inew)
ff_w=w_inew_norm

set.seed(123)
post=ff_w[1]*rbeta(50000,32,20)+
  ff_w[2]*rbeta(50000,33,20)+
  ff_w[3]*rbeta(50000,33,22)+
  ff_w[4]*rbeta(50000,53,49)+
  ff_w[5]*rbeta(50000,54,49)+
  ff_w[6]*rbeta(50000,54,51)

data <- data.frame(posterior = post)

ggplot(data, aes(x = posterior)) +
  geom_histogram(binwidth = 0.01, fill = "blue",
                 alpha = 0.5, color = "black", aes(y = ..density..)) +
  labs(title = "Posterior Distribution",
       x = expression(theta), y = "Density") +
  theme_minimal()
```

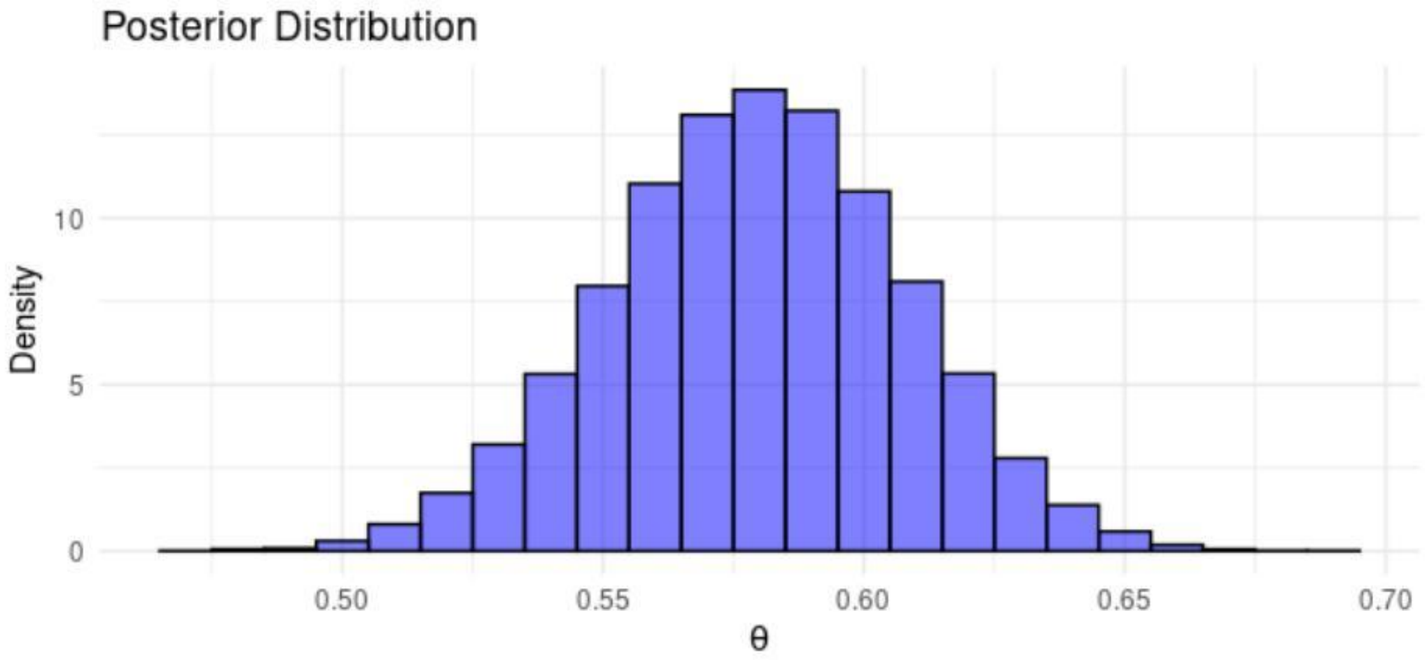


Figure 2.7: Posterior plot utilizing the posterior of coin1.

Note: It would be much more informative if I had used a different coin with the same denominator and year of mint, as the materials and their weight would have been almost identical.

Task 3

3.1 Galenshore distribution

[11]

An unknown quantity Y has a Galenshore(a, θ) distribution if its density is given by

$$p(y) = \frac{2}{\Gamma(a)}\theta^{2a}y^{2a-1}e^{-\theta^2y^2}$$

for $y > 0$, $\theta > 0$ and $a > 0$. Assume for now that a is known. For this density,

$$E[Y] = \frac{\Gamma(a+\frac{1}{2})}{\theta\Gamma(a)}, \quad E[Y^2] = \frac{a}{\theta^2}$$

a) Identify a class of conjugate prior densities for θ. Plot a few members of this class of densities.

b) Let $Y_1, \ldots, Y_n \sim$ i.i.d. Galenshore(a, θ). Find the posterior distribution of θ given $Y_1, \ldots, Y_n$, using a prior from your conjugate class.

c) Write down $\mathrm{p}(\theta_a|Y_1, \ldots, Y_n)/p(\theta_b|Y_1, \ldots, Y_n)$ and simplify. Identify a sufficient statistic.

d) Determine $E[\theta|y_1, \ldots, y_n]$. **e)** Determine the form of the posterior predictive density $\mathrm{p}(\tilde{(y)}|y_1, \ldots, y_n)$.

Solution :

a) I will start by underlining that y^2 is a sufficient statistic as described in the below ratio equation all the information about y comes from y^2 and additionally $y > 0$ so knowing about y^2 is as if we know about y.

$$\frac{p(\theta_a|y)}{p(\theta_b|y)} = \frac{p(y|\theta_a)p(\theta_a)}{p(y|\theta_b)p(\theta_b)} = \frac{\theta_a^{2a}}{\theta_b^{2a}}\left(\frac{e^{\theta_a^2}}{e^{\theta_b^2}}\right)^{-y^2}\frac{p(\theta_a)}{p(\theta_b)}$$

Now that we proved that $-y^2$ is a sufficient statistic all the requirements have been

[11] [DHo09]pg.231 Exercise 3.9

met for the Galenshore distribution to be expressed as an exponential family model.

$$p(y|\phi) = h(y)c(\phi)e^{\phi t(y)}$$

where, $h(y) = \frac{2}{\Gamma(a)}y^{2a-1}$, $c(\phi) = \phi^{2a}$, $t(y) = -y^2$, $\phi = \theta$.
Now we can use our knowledge about the form of exponential family's priors which are:

$$p(\phi|n_0, t_0) = K(n_0, t_0)c(\phi)_0^n e^{n_0 t_0 \phi} = K(n_0, t_0)(\theta^2)^{an_0} e^{-n_0 t_0 \theta^2} \tag{3.5}$$

where for a weak prior we choose $n_0 = 1$ and t_0 represents our prior beliefs about t, for which I am still concerned about how it should be chosen with my information is only that it should represent our prior belief about the expected value of the sufficient statistic t(y) and is chosen based on prior knowledge about the parameter in question. To answer the question, thus our priors have the form:

$$p(\phi|t_0) = K(1, t_0)\theta^{2a} e^{-t_0\theta^2}$$

For the sake of this exercise, I will arbitrarily choose some t_0 values as well as a range of θ values[12] to produce the plot (3.8).

```
library(ggplot2)
```

```
prior_density <- function(theta, a, t0) {
```

[12] In the context of the Galenshore distribution, the parameter θ is not a probability parameter, which would be constrained to the interval $[0, 1]$. Instead, it is a scale parameter that influences the spread and shape of the distribution.

In the density function of the Galenshore distribution given by

$$p(y) = \frac{2}{\Gamma(a)}\theta^{2a}y^{2a-1}e^{-\theta^2 y^2},$$

we see that θ appears in two places: in the power term θ^{2a} and in the exponential term $e^{-\theta^2 y^2}$. This means that θ has a significant role in determining the behavior of the distribution.
A larger value of θ will result in a distribution that is more spread out, with a lower peak, as the exponential term will decay more quickly. Conversely, a smaller value of θ will result in a distribution that is more peaked, with a higher peak, as the exponential term will decay more slowly.
Therefore, θ can take any positive value, allowing it to control the scale and shape of the distribution to a significant degree, and it is not limited to the interval $[0, 1]$.

```
  (theta^2)^a * exp(-t0 * theta^2)}

theta_seq <- seq(0, 3, length.out = 1000); a <- 1

# Create a data frame
data <- data.frame(
  theta = theta_seq,
  t0_1 = prior_density(theta_seq, a, 1),
  t0_2 = prior_density(theta_seq, a, 2),
  t0_3 = prior_density(theta_seq, a, 3))

# Convert the data from wide format to long format for ggplot
data_long <- tidyr::pivot_longer(
  data, cols = starts_with("t0"),
  names_to = "t0", values_to = "density")

ggplot(data_long, aes(x = theta, y = density, color = t0)) +
  geom_line() +
labs(title = "Prior Densities for Different t0 Values",
     x = expression(theta), y = "Density") +
theme_minimal()
```

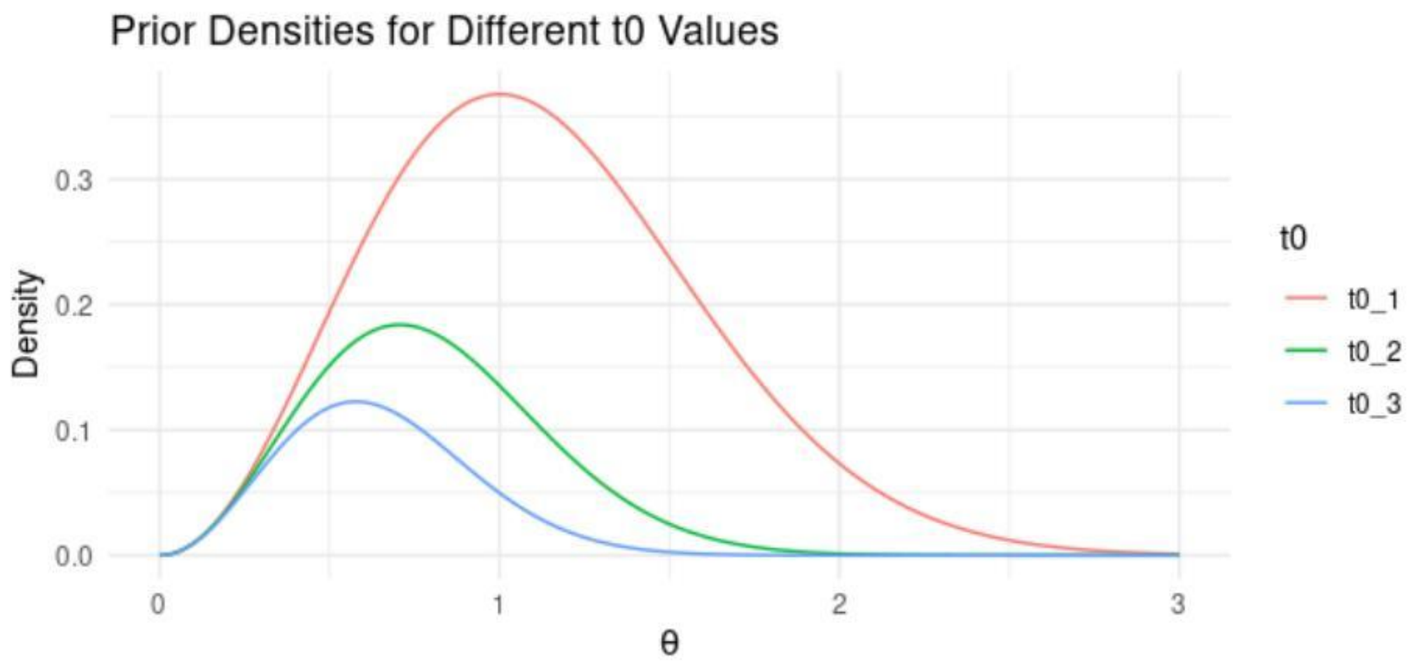


Figure 3.8: Prior plots for various t_0

b) Given the model function for a single observation y_i:

$$p(y_i|\phi) = h(y_i)c(\phi)e^{\phi t(y_i)}$$

,we can express the likelihood function $Y = \{y_1, y_2, \ldots, y_n\}$ as the product of the models:

$$p(Y|\phi) = \prod_{i=1}^{n} p(y_i|\phi) = \prod_{i=1}^{n} h(y_i)c(\phi)e^{\phi t(y_i)} = \left(\prod_{i=1}^{n} h(y_i)\right) c(\phi)^n e^{\phi \sum_{i=1}^{n} t(y_i)}$$

Now, using Bayes' theorem, we can find the posterior distribution of ϕ given the data Y:

$$p(\phi|Y) \propto p(Y|\phi)p(\phi) = \left(\prod_{i=1}^{n} h(y_i)\right) c(\phi)^n e^{\phi \sum_{i=1}^{n} t(y_i)} \cdot p(\phi)$$

Substituting the expression for the prior distribution $p(\phi|n_0, t_0)$:

$$p(\phi|n_0, t_0) = K(n_0, t_0)c(\phi)^{n_0} e^{n_0 t_0 \phi} \Rightarrow$$

$$p(\phi|Y) \propto \left(\prod_{i=1}^{n} h(y_i)\right) c(\phi)^{n+n_0} e^{\phi(\sum_{i=1}^{n} t(y_i) + n_0 t_0)}$$

This shows that the posterior distribution is in the same family as the prior, confirming that we have a conjugate prior. To find the exact form of the posterior, we identify the new hyperparameters as:

$$n_{\text{new}} = n + n_0, \quad t_{\text{new}} = \frac{1}{n_{\text{new}}} \left(\sum_{i=1}^{n} t(y_i) + n_0 t_0\right)$$

Thus, the posterior distribution is given by [13]:

[13] In the R code used to plot the priors, we chose to set the normalizing constant $K(n_0, t_0)$ to 1. This choice was made to facilitate the visualization of the prior distributions without affecting the relative shapes of the distributions, which are the main focus of our analysis at this stage.

In a more formal analytical context, $K(n_0, t_0)$ would be determined to ensure that the prior distribution integrates to 1, satisfying the probability density function properties. However, determining $K(n_0, t_0)$ analytically can be a complex process involving integration over the entire parameter space, which might not always yield a closed-form solution.

Therefore, in practice, especially in the context of plotting or simulation studies, it is common to leave $K(n_0, t_0)$ undetermined or set it to a convenient value such as 1, with the understanding that a proper normalization would be carried out when necessary, such as when exact probability

$$p(\phi|Y) = K(n_{\text{new}}, t_{\text{new}})c(\phi)^{n_{\text{new}}} e^{n_{\text{new}} t_{\text{new}} \phi}$$

c) In this part, we are tasked with finding the ratio

$$\frac{p(\theta_a|Y_1, \ldots, Y_n)}{p(\theta_b|Y_1, \ldots, Y_n)}$$

and simplifying it to identify a sufficient statistic. To find this ratio, we can use the posterior distributions we found in part b).
The posterior distribution was of the form:

$$p(\theta|Y_1, \ldots, Y_n) = K(n_0 + n, t_0 + t^{'}(y))\theta^{2a(n_0+n)} e^{-(n_0 t_0 + n t^{'}(y))\theta^2}$$

where n is the number of observations and $t(y)$ is the sum of the squares of the observations.
So the ratio becomes:

$$\frac{K(n_0 + n, t_0 + t^{'}(y))\theta_a^{2a(n_0+n)} e^{-(n_0 t_0 + n t^{'}(y))\theta_a^2}}{K(n_0 + n, t_0 + t^{'}(y))\theta_b^{2a(n_0+n)} e^{-(n_0 t_0 + n t^{'}(y))\theta_b^2}}$$

We can see that many terms will cancel out, leaving us with:

$$\left(\frac{\theta_a}{\theta_b}\right)^{2a(n_0+n)} \exp\left(-(n_0 t_0 + n t^{'}(y))(\theta_a^2 - \theta_b^2)\right)$$

Now, we can see that the only term involving the data $Y_1, \ldots, Y_n$ is $t^{'}(y)$, which is the sum of the squares of the observations.
This indicates that $t^{'}(y) = -\sum_1^n y_i^2$ is a sufficient statistic.

d) To find the expected value $E[\theta|y_1, y_2, \ldots, y_n]$, we will integrate the posterior

---

values are required or when performing Bayesian updates to obtain the posterior distribution.

In our case, since we are primarily interested in the shape and characteristics of the prior distributions, and not the exact probability values, setting $K(n_0, t_0) = 1$ in the code serves our purpose adequately. Later, when we derive the posterior distribution, we will determine $K(n_0, t_0)$ appropriately to ensure that the posterior distribution satisfies the properties of a probability density function.

distribution of θ multiplied by θ over all possible values of θ from 0 to infinity. The integral we are looking to solve is:

$$E[\theta|y_1, y_2, \ldots, y_n] = \int_0^\infty \theta \cdot K \cdot (\theta^2)^{a(n+1)-1} e^{-\left(1+\sum_{i=1}^n y_i^2\right)\theta^2} d\theta.$$

To find this expectation,we notice the integral's similarities with the gamma distribution density and use the property of its expected value:

$$E[\theta] = \frac{\alpha}{\beta}.$$

Substituting the values we have for α and β, which are $\alpha = a(n+1)$ and $\beta = 1 + \sum_{i=1}^n y_i^2$, we find:

$$E[\theta|y_1, y_2, \ldots, y_n] = \frac{\Gamma(a(n+1))}{\beta^a} \frac{a(n+1)}{1+\sum_{i=1}^n y_i^2}.$$

This gives us the expected value of θ given the observations $y_1, y_2, \ldots, y_n$.

e) To find the posterior predictive density $\mathrm{p}(y|y_1, \ldots, y_n)$, we need to integrate out the parameter θ from the joint distribution of y and θ given the data. This involves integrating the product of the likelihood function $\mathrm{p}(\tilde{y}|\theta)$ and the posterior distribution $\mathrm{p}(\theta|y_1, \ldots, y_n)$ with respect to θ. This could be done by Monte Carlo sampling which is in the next chapteror by integrating the posterior data samples of $\theta|y_1, \ldots, y_n$ into our model (to approximate the predictive density). In the case of prior conjugacy, an analytical solution can be found. In this direction, substituting the expressions for $p(\tilde{y}|\theta, y_1, \ldots, y_n)$ and $p(\theta|y_1, \ldots, y_n)$ we derived in previous parts, we have:

$$p(\tilde{y}|y_1, \ldots, y_n) = \int \frac{2}{\Gamma(a)} \theta^{2a} \tilde{y}^{2a-1} e^{-\theta^2\tilde{y}^2} K \theta^{2a(n+1)} e^{-n_0 t_0 \theta^2} d\theta$$

To find the normalization constant K, we set the integral of the density equal to 1:

$$1 = \int p(\tilde{y}|y_1, \ldots, y_n) d\tilde{y}$$

This gives us an integral equation for K. Solving this equation will give us the normalization constant, and hence the full expression for the posterior predictive density.

After finding K, we can write down the final expression for the posterior predictive density as a function of $\tilde{y}$ and the data $y_1, \ldots, y_n$.

$$p(\tilde{y}|y_1, \ldots, y_n) = \frac{2}{\Gamma(a)} \tilde{y}^{2a-1} K \left(\frac{1}{2(\tilde{y}^2 + 1 + \sum_{i=1}^{n} y_i^2)} \right)^{\frac{2a(n+1)-1}{2}} \Gamma\left(a(n+1)\right)$$

This is the expression for the posterior predictive density. It is a function of $\tilde{y}$ and depends on the observed data $y_1, \ldots, y_n$ through the term $\sum_{i=1}^{n} y_i^2$.

3.2 Change of variables

14

Let $\psi = g(\theta)$, where g is a monotone function of θ, and let h be the inverse of g so that $\theta = h(\psi)$. If $p_\theta(\theta)$ is the probability density of θ, then the probability density of ψ induced by p_θ is given by:

$$p_\psi(\psi) = p_\theta(h(\psi)) \times \left| \frac{dh}{d\psi} \right| .$$

a) Let $\theta \sim \text{beta}(a, b)$ and let $\psi = \log\left[\frac{\theta}{1-\theta}\right]$. Obtain the form of p_ψ and plot it for the case that $a = b = 1$.
b) Let $\theta \sim \text{gamma}(a, b)$ and let $\psi = \log\theta$. Obtain the form of p_ψ and plot it for the case that $a = b = 1$.

Solution :

a) In a straightforward manner, we can find the $h(\psi)$, it's derivative, and use the density function of θ for that $h(\psi)$:

$$p_\psi(\psi) = p_\theta(h(\psi)) \times \left| \frac{dh}{d\psi} \right| \tag{3.6}$$

$$= \left[\frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} \left(\frac{e^\psi}{1+e^\psi} \right)^{a-1} \left(1 - \frac{e^\psi}{1+e^\psi} \right)^{b-1} \right] \times \frac{e^\psi}{(e^\psi+1)^2} \tag{3.7}$$

$$= \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} \left[\left(\frac{e^\psi}{1+e^\psi} \right)^{a-1} \left(\frac{1}{1+e^\psi} \right)^{b-1} \right] \times \frac{e^\psi}{(e^\psi+1)^2} \tag{3.8}$$

$$= \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} \left[\left(\frac{e^\psi}{1+e^\psi} \right)^{a} \left(\frac{1+e^\psi}{e^\psi} \right) \left(\frac{1}{1+e^\psi} \right)^{b} \left(\frac{1+e^\psi}{1} \right) \right] \times \frac{e^\psi}{(e^\psi+1)^2} \tag{3.9}$$

$$= \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} \left[\left(\frac{e^\psi}{1+e^\psi} \right)^{a} \left(\frac{1}{1+e^\psi} \right)^{b} \frac{(e^\psi+1)^2}{e^\psi} \right] \times \frac{e^\psi}{(e^\psi+1)^2} \tag{3.10}$$

$$= \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} \left(\frac{e^\psi}{1+e^\psi} \right)^{a} \left(\frac{1}{1+e^\psi} \right)^{b} \tag{3.11}$$

Supposing a=b=1 the plot (3.9) was made in R:

```
a <- 1; b <- 1
psi <- seq(-6, 6, length.out = 1000)
theta <- exp(psi)/(1 + exp(psi))
```

[14] [DHo09]pg.232 Exercise 3.10

```
p_theta <- gamma(a + b)/
        (gamma(a) * gamma(b)) * theta^(a - 1) *
                                (1 - theta)^(b - 1)
jacobian <- exp(psi)/((1 + exp(psi))^2)

p_psi <- p_theta * jacobian
data <- data.frame(psi, p_psi)

ggplot(data, aes(x = psi, y = p_psi)) +
  geom_line() +
  labs(title = "Density function of ψ", x = "ψ", y = "pψ(ψ)") +
  theme_minimal()
```

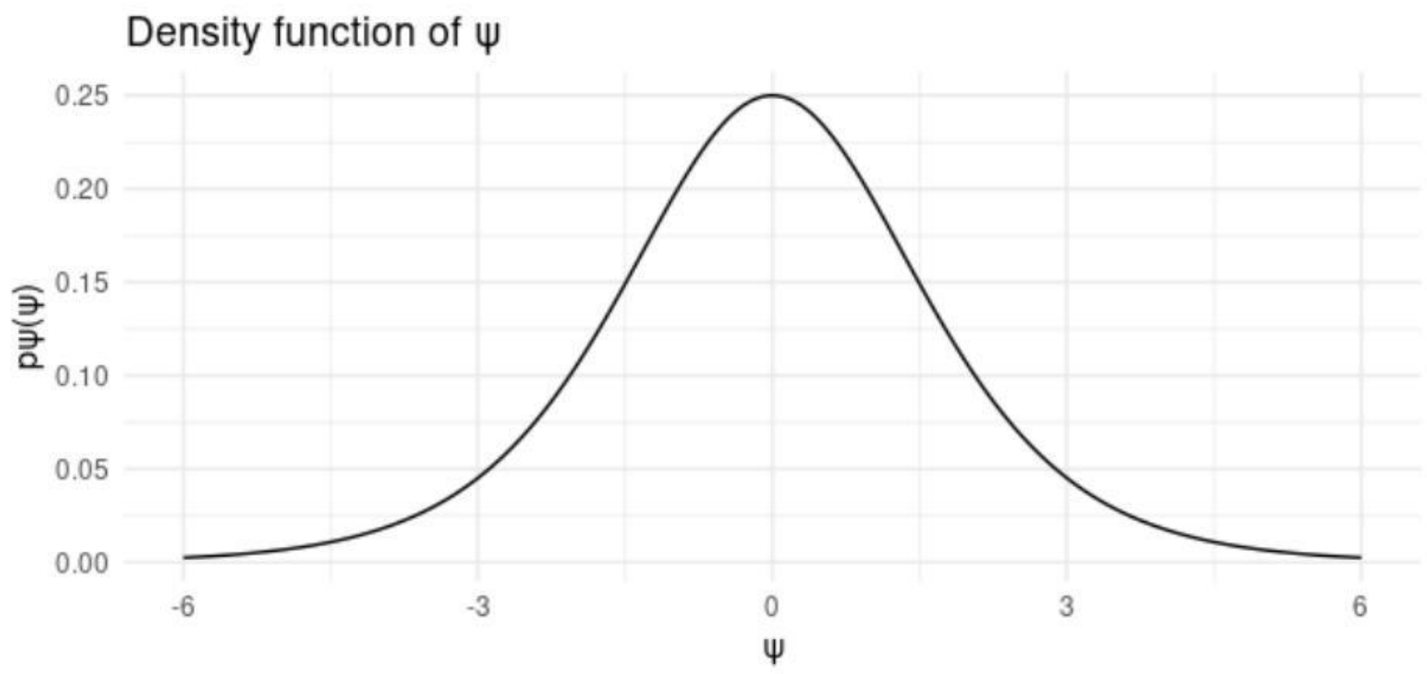


Figure 3.9: Plot for of ψ density where $a = b = 1$

b) For $\psi = g(\theta) = \log\theta$, the $\theta = h(\psi) = e^{\psi}$. Then,

$$p_{\psi}(\psi) = p_{\theta}(h(\psi)) \times \left|\frac{dh}{d\psi}\right| \tag{3.12}$$

$$= \left[\frac{b^a}{\Gamma(a)}\exp\left(\psi(a-1)\right)\exp\left(-be^{\psi}\right)\right] \times \exp\left(\psi\right) \tag{3.13}$$

$$= \frac{b^a}{\Gamma(a)}\exp\left(a\psi - \psi - be^{\psi} + \psi\right) \tag{3.14}$$

$$= \frac{b^a}{\Gamma(a)}\exp\left(a\psi - be^{\psi}\right) \tag{3.15}$$

For the case $a = b = 1$ the plot (3.10) was produced in R:

```r
library(ggplot2)
a <- 1; b <- 1
psi <- seq(-6, 6, length.out = 1000)
p_psi <- (b^a/gamma(a)) * exp(a*psi - b*exp(psi))
data <- data.frame(psi, p_psi)

ggplot(data, aes(x = psi, y = p_psi)) +
geom_line() +
labs(
title = "Density function of ψ",
x = "ψ", y = expression(p[ψ](ψ))) +
theme_minimal()
```

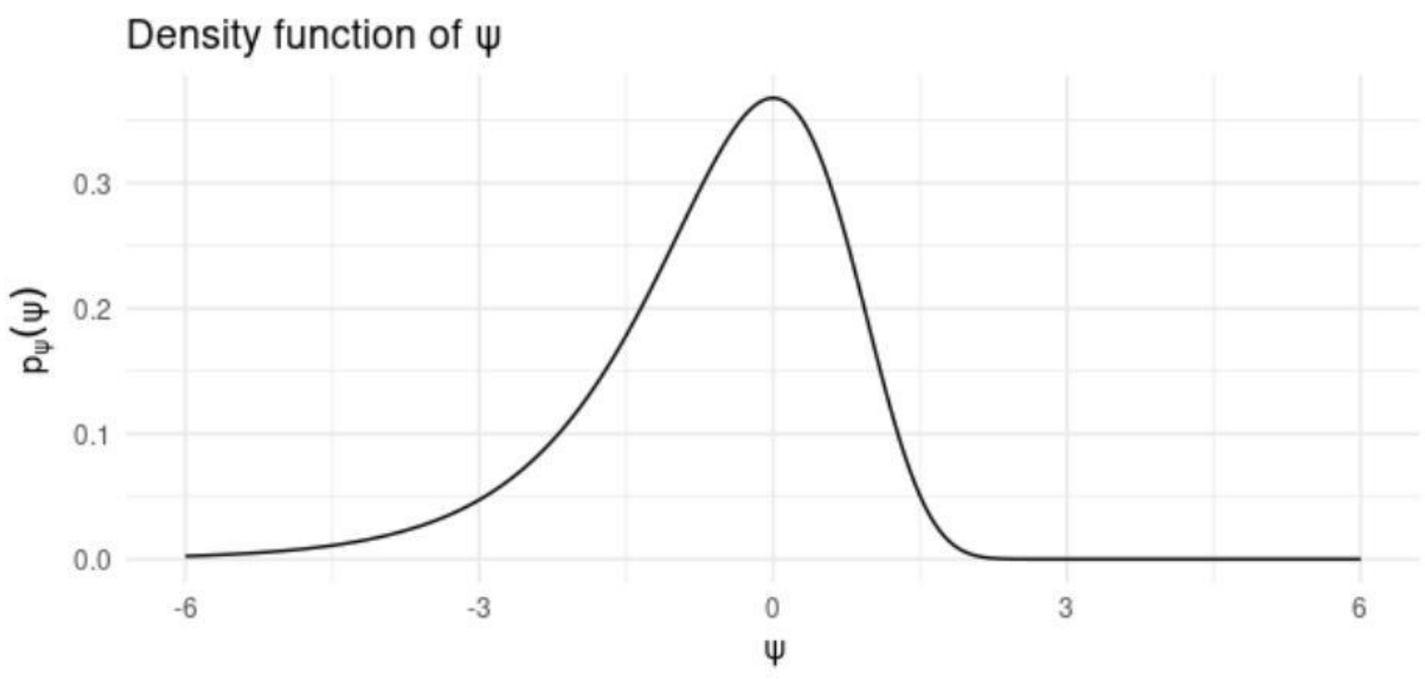


Figure 3.10: Plot for of ψ density where $a = b = 1$

3.3 Jeffreys' prior

15

Jeffreys (1961) suggested a default rule for generating a prior distribution of a parameter θ in a sampling model $p(y|\theta)$. Jeffreys' prior is given by $p_J(\theta) \propto \sqrt{I(\theta)}$, where $I(\theta) = -E\left[\frac{\partial^2 \log p(Y|\theta)}{\partial \theta^2}\middle|\theta\right]$ is the Fisher information.

a) Let $Y \sim$ binomial(n, θ). Obtain Jeffreys' prior distribution $p_J(\theta)$ for this model.

b) Reparameterize the binomial sampling model with $\psi = \log \frac{\theta}{1-\theta}$, so that $p(y|\psi) = \binom{n}{y} e^{\psi y}(1 + e^{\psi})^{-n}$. Obtain Jeffreys' prior distribution $p_J(\psi)$ for this model.

c) Take the prior distribution from a) and apply the change of variables formula from Exercise 3.10 to obtain the induced prior density on ψ. This density should be the same as the one derived in part b) of this exercise. This consistency under reparameterization is the defining characteristic of Jeffrey's prior.

Solution :

a) $Y \sim$ Binomial, then $p(y \mid \theta) = \binom{n}{y}\theta^y(1-\theta)^{n-y}$. and $I(\theta) = -\mathbb{E}(\partial^2 \ell(y \mid \theta)/\partial\theta^2)$ where $\ell(y \mid \theta) = \log p(y \mid \theta)$. So,

$$\ell(y \mid \theta) = \log p(y \mid \theta) \tag{3.16}$$

$$= \log\left(\binom{n}{y}\theta^y(1-\theta)^{n-y}\right) \tag{3.17}$$

$$= \log\left(\binom{n}{y}\right) + y\log(\theta) + (n-y)\log(1-\theta) \tag{3.18}$$

$$\ell_\theta(y \mid \theta) = \frac{y}{\theta} - \frac{n-y}{1-\theta} \tag{3.19}$$

$$\ell_{\theta\theta}(y \mid \theta) = -\frac{y}{\theta^2} - \frac{n-y}{(1-\theta)^2} \tag{3.20}$$

Thus,

$$I(\theta) = -\mathbb{E}\left(-\frac{y}{\theta^2} - \frac{n-y}{(1-\theta)^2}\right) \tag{3.21}$$

$$= -\left(-\frac{1}{\theta^2}\mathbb{E}(y) - \frac{1}{(1-\theta)^2}\mathbb{E}(n-y)\right) \tag{3.22}$$

$$= \frac{n\theta}{\theta^2} + \frac{n - n\theta}{(1-\theta)^2} \tag{3.23}$$

$$= \frac{n}{\theta} + \frac{n}{1-\theta} \tag{3.24}$$

$$= \frac{n}{\theta(1-\theta)} \tag{3.25}$$

And finally, the Jeffreys' prior distribution is:

[15] [DHo09]pg.231 Exercise 3.12

$$p_J(\theta) = c\sqrt{\frac{n}{\theta(1-\theta)}} \tag{3.26}$$

$$\tag{3.27}$$

b) Again similarly, for $\psi = \log \frac{\theta}{1-\theta}$, where $p(y \mid \psi) = \binom{n}{y} e^{\psi y}(1+e)^{-n}$ and $\ell(y \mid \psi) = \log p(y \mid \psi)$.

$$\begin{aligned}
\ell(y \mid \psi) &= \log p(y \mid \psi) && (3.28)\\
&= \log\left(\binom{n}{y} e^{\psi y}(1+e^{\psi})^{-n}\right) && (3.29)\\
&= \log\binom{n}{y} + \psi y - n\log\left(1+e^{\psi}\right) && (3.30)\\
\ell_{\psi}(y \mid \psi) &= y - \frac{ne^{\psi}}{e^{\psi}+1} && (3.31)\\
\ell_{\psi\psi}(y \mid \psi) &= -\frac{ne^{\psi}}{(e^{\psi}+1)^2} && (3.32)
\end{aligned}$$

And

$$\begin{aligned}
I(\psi) &= -\mathbb{E}\left(-\frac{ne^{\psi}}{(e^{\psi}+1)^2}\right) && (3.33)\\
&= \frac{ne^{\psi}}{(e^{\psi}+1)^2} && (3.34)\\
& && (3.35)
\end{aligned}$$

Thus, Jeffrey's prior is

$$\begin{aligned}
p_J(\psi) &\propto \sqrt{\frac{ne^{\psi}}{(e^{\psi}+1)^2}} && (3.36)\\
&\propto \frac{\sqrt{ne^{\psi}}}{e^{\psi}+1} && (3.37)
\end{aligned}$$

c) If $\psi = g(\theta) = \log \frac{\theta}{1-\theta}$, then let $\theta = h(\psi) = \frac{e^{\psi}}{1+e^{\psi}}$. Then, by the change of variables formula,

$$p_\psi(\psi) \propto p_\theta(h(\psi)) \times \left|\frac{dh}{d\psi}\right| \tag{3.38}$$

$$\propto \sqrt{\frac{n}{\frac{e^\psi}{1+e^\psi}\left(1-\frac{e^\psi}{1+e^\psi}\right)}} \times \frac{e^\psi}{(e^\psi+1)^2} \tag{3.39}$$

$$\propto \sqrt{\frac{n(e^\psi+1)^2}{e^\psi}} \times \frac{e^\psi}{(e^\psi+1)^2} \tag{3.40}$$

$$\propto \frac{\sqrt{n}(e^\psi+1)}{\sqrt{e^\psi}} \times \frac{e^\psi}{(e^\psi+1)^2} \tag{3.41}$$

$$\propto \frac{\sqrt{n}\sqrt{e^\psi}}{e^\psi+1} \tag{3.42}$$

$$\propto p_J(\psi). \tag{3.43}$$

In this case, it has been demonstrated that Jeffreys' prior is invariant under monotone transformation.

3.4 Some excersises sent to our mail.

16

3.4.1 Photo Exercise from mail 1

In a medical experiment, patients with a chronic condition are asked to say which of the two treatments, A,B, they prefer. (You may assume for the purpose of this question that every patient will express a preference one way or the other). Let the population proportion who prefer A be θ. We observe a sample of n patients. Given θ the n responses are independent and the probability that a particular patient prefers A is θ.

Our prior distribution for θ is a beta(a, a) distribution with a standard deviation of 0.25.

a) Find the value of a.

b) We observed n=30 patients of whom 21 prefer treatment A. Find the posterior distribution of θ.

c) Find the posterior mean and standard deviation of θ.

d) Using R or otherwise, find a symmetric 95% posterior probability interval for θ. (**Hint:** The R command qbeta$(0,025, a, b)$ will give the 2.5% point of a beta(a,b) distribution)

Solution :

a) Pretty straight forward:

$$0.0625 = \text{sd}(\theta)^2 = \text{Var}[\theta] = \frac{ab}{(a+b+1)(a+b)^2} = \frac{a^2}{(2a+1)4a^2} = \frac{1}{8a+4} \Rightarrow a = 1.5$$

b) Given the nature of the problem our model is binomial(y, n, θ) = binomial$(21, 30, \theta)$,thus our posterior is a beta$(a + y, a + n - y)$ = beta$(22.5, 10.5)$

c) The posterior mean and sd is the mean of our beta$(22.5, 10.5)$, so:

$$E[\theta|a_{new}, b_{new}] = \frac{a_{new}}{a_{new} + b_{new}} = \frac{22.5}{30} = 0.75$$

d) This part was done in R using the below code:

```
library(stats)
```

[16] Two exercises from pictures sented to our mails...

```
a_new <- 22.5; b_new <- 10.5

lower_bound <- qbeta(0.025, a_new, b_new)
upper_bound <- qbeta(0.975, a_new, b_new)

cat("95% credible interval for θ: [", lower_bound, ", ", upper_bound, "]\n")
```

The provided "95% credible interval for θ was given as: $[0.5161281, 0.8266448]$"

3.4.2 Photo Exercise from mail 2

Our prior distribution for λ is a gamma(a,b) distribution.

a) Our prior mean and standard deviation for λ are 16 and 8 respectively. Find the values of a and b.

b) The observed number of sales are as follows.

14 19 14 21 22 33 15 13 16 19 27 22 27 21 16 25 14 23 22 17

Find the posterior distribution of λ.

c) Using R or otherwise, plot a graph showing both the prior and posterior probability density functions of λ.

d) Using R or otherwise, find a 95% posterior hpd interval for λ.

(**Note:** The R function hpdgamma is available from Module Web Page).

Solution :

a) $\lambda \sim \text{gamma}(a, b)$ with density function:

$$p(\lambda|a, b) = \frac{b^a}{\Gamma(a)}\lambda^{a-1}e^{-\lambda b}$$

The mean and standard deviation come from:

$$E[\lambda|a, b] = \frac{a}{b} = 8 \Rightarrow a = 16b$$

and

$$\text{Var}[\lambda|a, b] = \frac{a}{b^2} \Rightarrow 64b^2 = 16b \Rightarrow b = \frac{1}{4}$$

so $a = 4$.

b) Through conjugacy we know that, Given the observed data, the sum of the sales

$Y = \sum_i y_i = 432$ and the number of observations n=20, the posterior parameters are updated as follows:

$$a_{new} = a + Y = 436$$

$$b_{new} = b + n = 20.25$$

Thus,

$$\lambda|Y_1, \ldots, Y_20 \sim \text{gamma}(436, 20.25)$$

c) To re-create the plot, you can run the below R code:

```
library(ggplot2)

a_prior <- 4; b_prior <- 1/4
a_post <- 436; b_post <- 20.25

lambda <- seq(0, 50, length.out = 1000)
prior_density <- dgamma(lambda, shape = a_prior, rate = b_prior)
posterior_density <- dgamma(lambda, shape = a_post, rate = b_post)

data <- data.frame(lambda, prior_density, posterior_density)

ggplot(data, aes(x = lambda)) +
  geom_line(aes(y = prior_density), color = "blue") +
  geom_line(aes(y = posterior_density), color = "red") +
  labs(title = "Prior and Posterior Distributions", x = "Lambda", y = "Density") +
  theme_minimal()
```

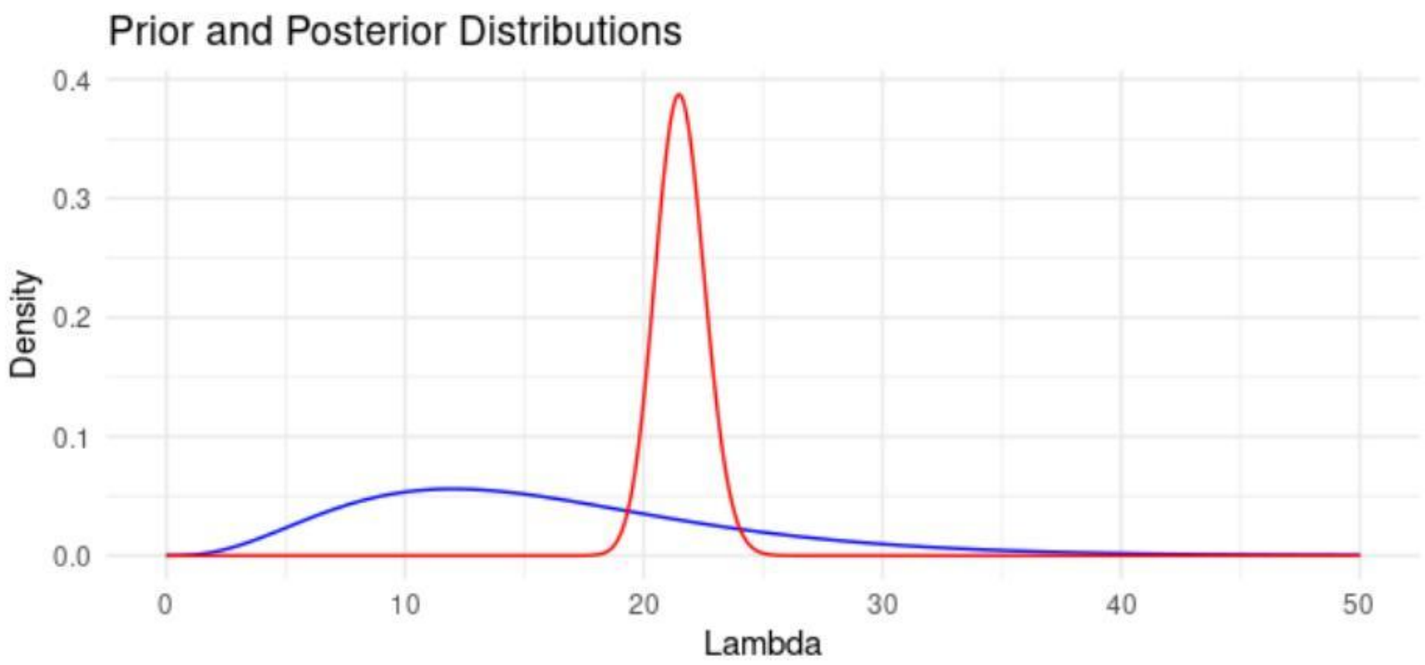


Figure 3.11: Plotprior and posterior.

d) For this part hpdgamma lies on the edge of the unknown, thus[17] I will use the hpd (19.55704, 23.59822) as shown in the code below.

```
a_post <- 436; b_post <- 20.25

hpd_interval <- qgamma(c(0.025, 0.975), shape = a_post, rate = b_post)
print(hpd_interval)
```

[17] I believe no-human, machinery or man-made AI know how to get access to the hpdgamma function or as I call it my white whale

Task 4

4.1 Tumor count comparisons

18

a) For the prior distribution given in part a) of the exercise (2.2), obtain $\Pr(\theta_B < \theta_A | y_A, y_B)$ via Monte Carlo sampling.
b) For a range of values of n_0, obtain $\Pr(\theta_B < \theta_A | y_A, y_B)$ for $\theta_A \sim gamma(120, 10)$ and $\theta_B \sim gamma(12 \times n_0, n_0)$. Describe how sensitive the conclusions about the event $\{\theta_B < \theta_A\}$ are to the prior distribution on θ_B.
c) Repeat parts a) and b), replacing the event $\{\theta_B < \theta_A\}$ with the event $\{Y_B < Y_A\}$, where Y_A and Y_B are samples from the posterior predictive distribution.

Solution):

a) For each run the result might differ slightly.

```
theta.a = rgamma(500, 237, 20)
theta.b = rgamma(500, 125, 14)
mean(theta.b < theta.a)
#[1] 0.9956
```

By the Monte Carlo sampling method we get 500 values for θ_A and θ_B, from their posterior distributions. Then we evaluate their relationship by assigning the value 1 if $\theta_B < \theta_A$ and 0 otherwise. Finally the result is provided through the mean() function which provides the mean value of these 1's and 0's. To further aid our understanding of the result we can observe the plot (4.12), made in R as:

```
library(ggplot2)

# Set the seed for reproducibility,
```

[18] [DHo09] pg. 232 Exercise 4.2

```r
set.seed(123)
thetaA <- rgamma(500, 237, 20)
thetaB <- rgamma(500, 125, 14)

df <- data.frame(thetaA = thetaA,thetaB = thetaB)

ggplot(
  df, aes(x = thetaA, y = thetaB)) + geom_point() +
  geom_abline(intercept=0,slope=1,
              linetype = "dashed",color="red") +
  labs(x = expression(paste(
    "gamma(", theta[A], ", 237, 20)")),
    y = expression(paste(
      "gamma(", theta[B], ", 125, 14)"))) +
  theme_minimal()

prop_thetaB_less_than_thetaA <- mean(thetaB < thetaA)
cat("Proportion where thetaB<thetaA:",
    prop_thetaB_less_than_thetaA,"\n")
```

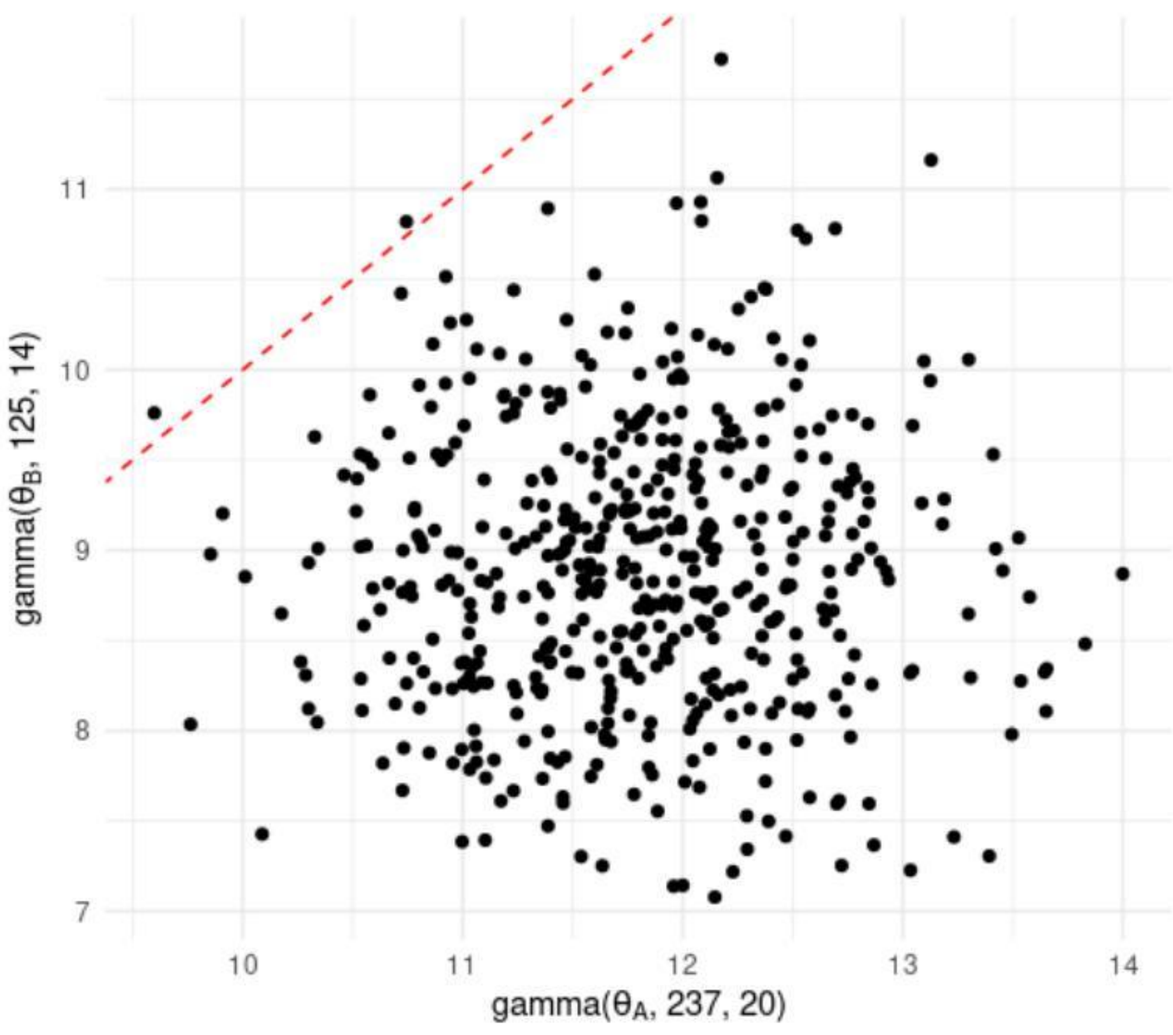
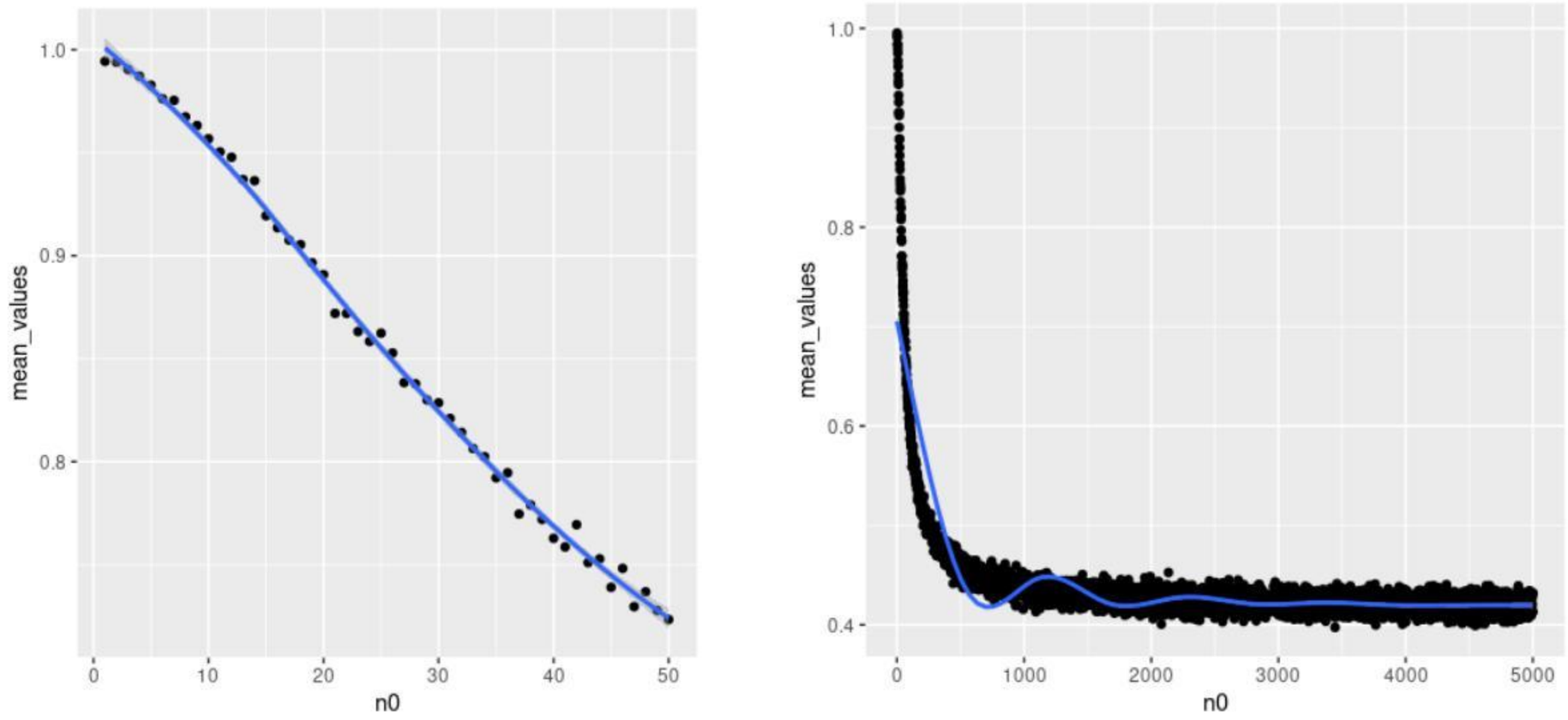


Figure 4.12: Scatter plot comparing θ_A and θ_B

b) My interpretation of this task is to figure the relationship between θ_A and θ_B



(a) $n_0 = [1 : 50]$ (b) $n_0 = [1 : 5000]$

Figure 4.13: Comparing means of $\theta_A > \theta_B$ for various n_0

for the gamma variations in 2.2 in hopes that for a n_0 it will converge to 0.5, recall Tumor count objective, which would mean that $\theta_A > \theta_B$ about half of the sampled times. Using R, we can visualize these differences as n_0 increases (see the plot 4.13).

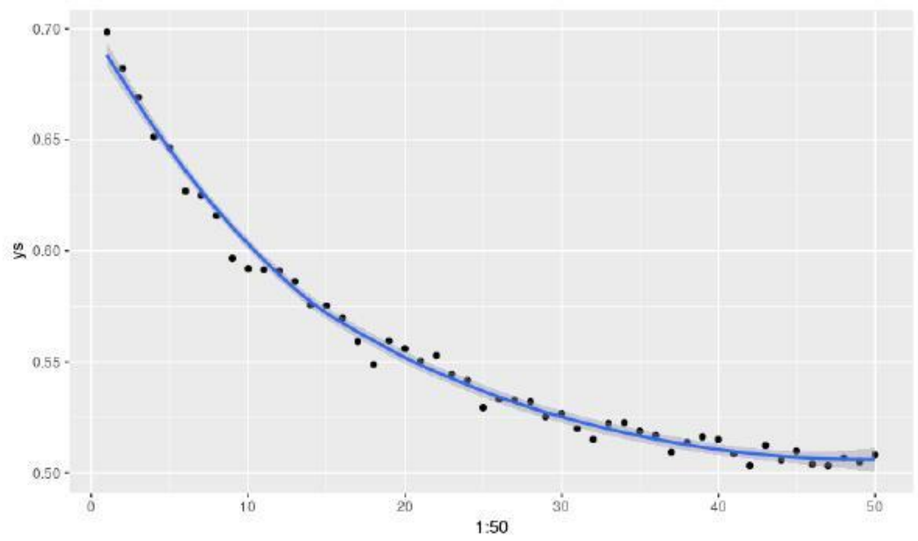
```
n0 = 1:50
mean_values = sapply(n0, function(n) {
  mean(rgamma(5000, (12 * n) + 113, n + 13) < rgamma(5000, 237, 20))
})
qplot(n0, mean_values, geom = c('point', 'smooth'))

#for the second plot just change n0= 1:5000
```

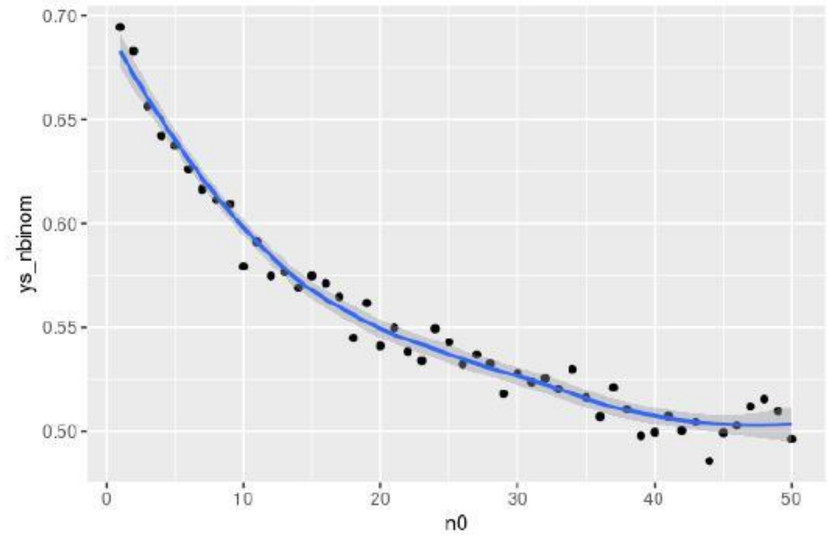
The result becomes increasingly evident, once you carefully analyze the meaning of our plots, as the posterior means are getting closer and closer, even though we cannot be certain that the means will converge, based on these plots, as $n_0 \to \infty$, I am confident in this interpretation.

c) The posterior predictive distribution after a Gamma(a, b) prior is a negative binomial distribution NB$(a + \sum y_i, b, n)$. (The posterior predictive parameters are the same as those of the posterior distribution itself). However, we can also simulate it by sampling from the Poisson distribution implied by the sampled theta (see 4.14)

```
n0 = 1:50; N = 5000
ys = sapply(n0, function(n) {
```



(a) Sampling from a Poisson



(b) Sampling from a negative binomial

Figure 4.14: Comparing means of $Y_A > Y_B$

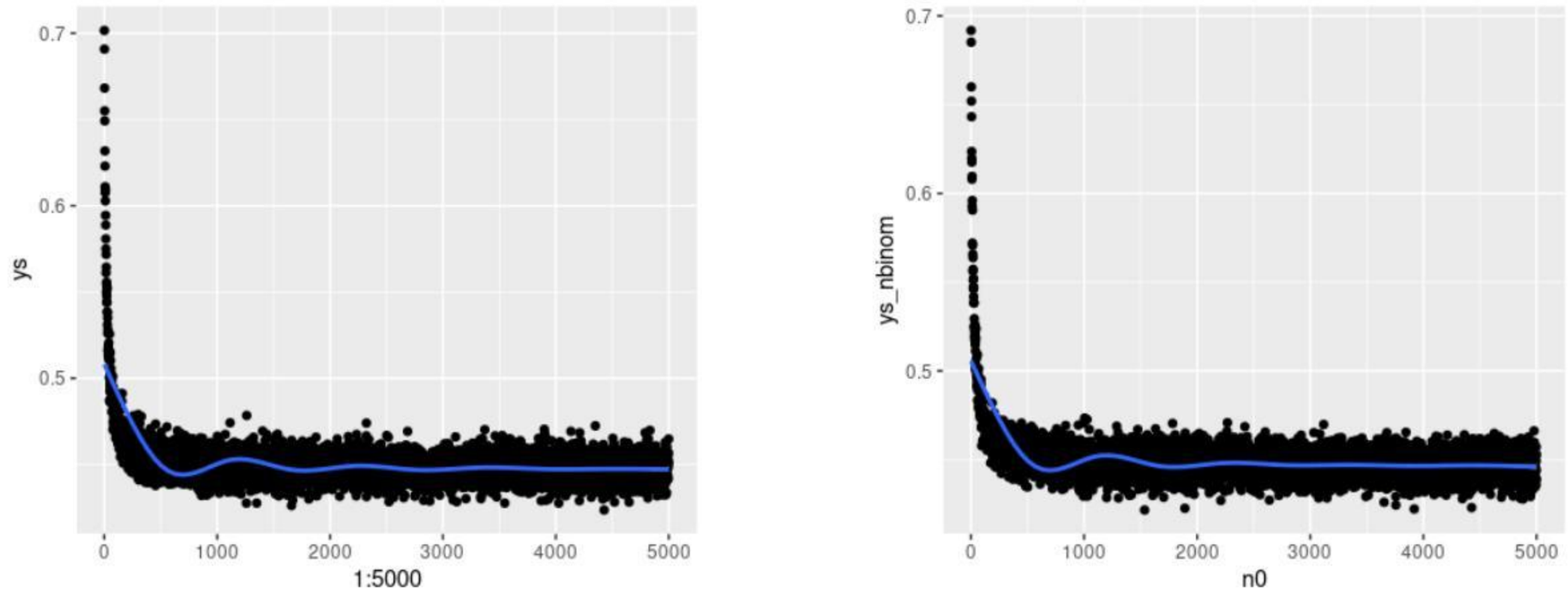
```
  theta.a = rgamma(N, 237, 20)
  theta.b = rgamma(N, (12 * n) + 113, n + 13)
  y.a = rpois(N, theta.a)
  y.b = rpois(N, theta.b)
  mean(y.b < y.a)})

qplot(1:50, ys, geom = c('point', 'smooth'))

#For negative binomial
n0 = 1:50; N = 5000
ys_nbinom = sapply(n0, function(n) {
  theta.a = rgamma(N, 237, 20)
  theta.b = rgamma(N, (12 * n) + 113, n + 13)
  y.a = rnbinom(N, size = 237, mu = theta.a)
  y.b = rnbinom(N, size = (12 * n) + 113, mu = theta.b)
  mean(y.b < y.a)})

qplot(n0, ys_nbinom, geom = c('point', 'smooth'))
```

As n_0 grows larger, both seem to converge to a number near 0.25 (see the 4.15 plot).



(a) Sampled from a Poisson

(b) Sampled from a negative binomial

Figure 4.15: Comparing means for $n_0 = [1 : 5000]$

4.2 Posterior predictive checks

[19] Let's investigate the adequacy of the Poisson model for the tumor count data. Following the example in [DHo09]Section 4.4, generate posterior predictive datasets $y_A^{(1)}, \ldots y_A^{(1000)}$. Each $y_A^{(s)}$ is a sample of size $n_A = 10$ from the Poisson distribution with parameter $\theta_A^{(s)}$, $\theta_A^{(s)}$ is itself a sample from the posterior distribution $p(\theta_A|y_A)$, and y_A is the observed data.

a) For each s, let $t^{(s)}$ be the sample average of the 10 values of $y_A^{(s)}$, divided by the sample standard deviation of $y_A^{(s)}$. Make a histogram of $t^{(s)}$ and compare it to the observed value of this statistic. Based on this statistic, assess the fit of the Poisson model for these data.

b)Repeat the above goodness of fit evaluation for the data in population B.

Solution):

a) Suppose the observed value of the statistic $t_{obs} = \frac{mean_{obs}}{sd_{obs}} \approx 3.82$. Using R, as presented below, we could make the histogram and see t_{obs} position in it.

```
ya = c(12, 9, 12, 14, 13, 13, 15, 8, 15, 6)
t_obs=mean(ya)/sd(ya)

theta1 = rgamma(1000, 2 + sum(ya), 1 + length(ya))
ts = sapply(theta1, function(theta) {
  ys = rpois(10, theta)
```

[19] [DHo09]pg 232 Exercise 4.3

```
  t = mean(ys) / sd(ys)
  t})

ggplot(data.frame(ts = ts), aes(x = ts)) +
  geom_histogram() +
  geom_vline(xintercept = mean(ya) / sd(ya))
```

Since our initial t_{obs} is centered well within the standard spread of the observed $t^{(s)}$ see 4.16, we have no reason to change our initial Poisson model.Since the observed statistic seems to be an outlier for the range of statistics.

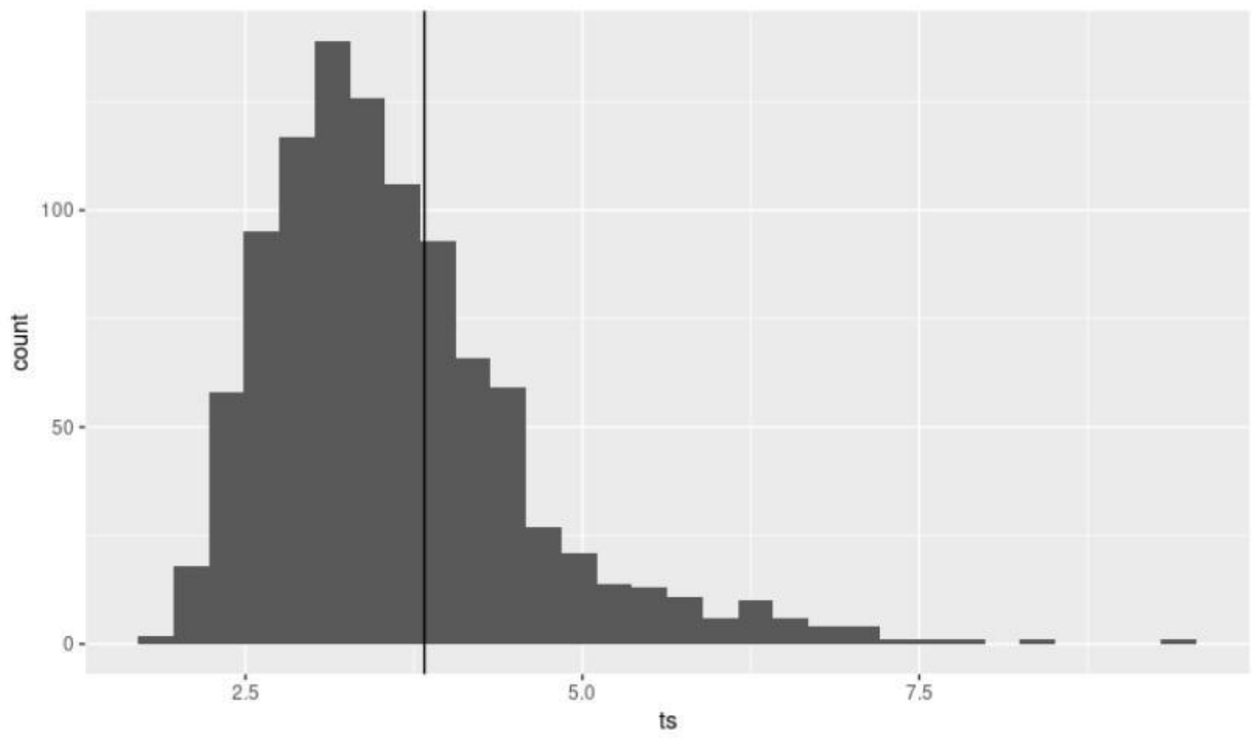


Figure 4.16: Histogram for Post. Predictive check of type A mice.

b) Following the same procedure for the type B mice, we get $t_{obs} \approx 2.84$, in this type of mice is obvious that our initial beliefs should change as they don't represent our population based on the given statistic as presented in (4.17).

```
yb = c(11, 11, 10, 9, 9, 8, 7, 10, 6, 8, 8, 9, 7)
t_obs=mean(yb)/sd(ya)
theta2 = rgamma(1000, 2 + sum(yb), 1 + length(yb))
ts = sapply(theta2, function(theta) {
  ys = rpois(10, theta)
  t = mean(ys) / sd(ys)
  t})

ggplot(data.frame(ts = ts), aes(x = ts)) +
  geom_histogram() +
  geom_vline(xintercept = mean(yb) / sd(yb))
```

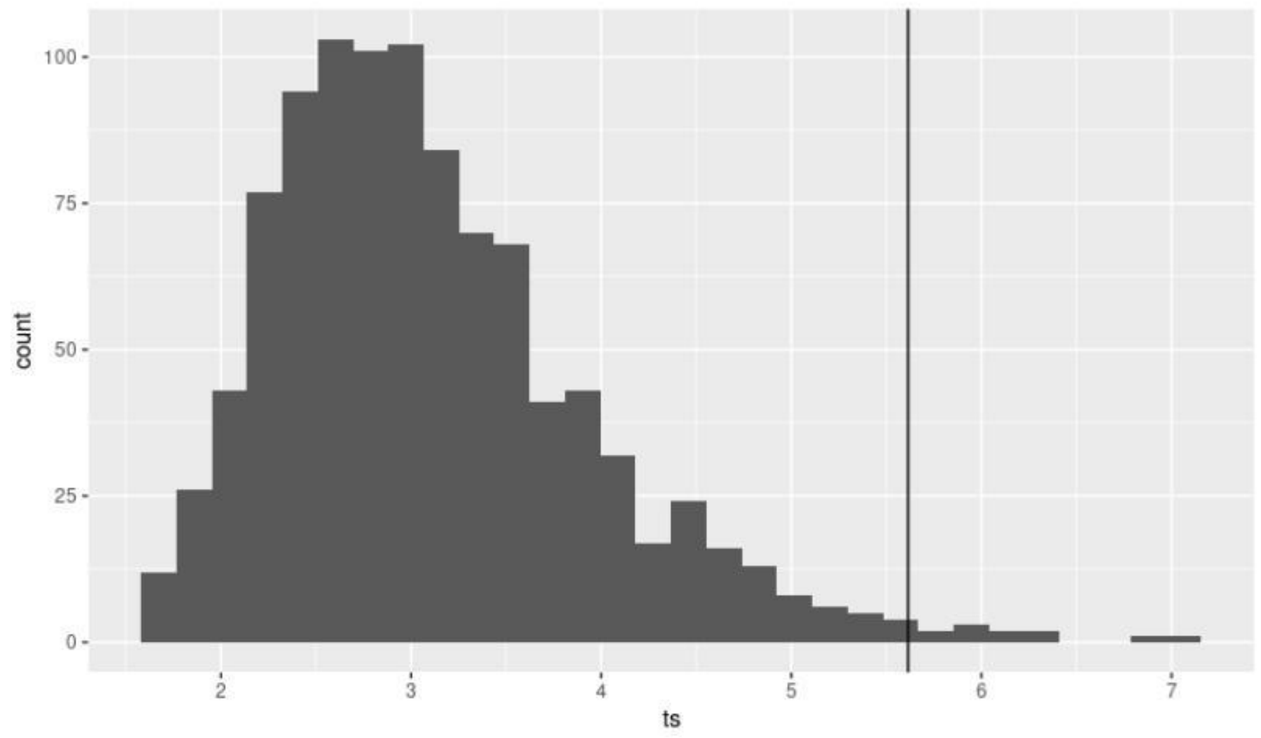


Figure 4.17: Histogram for Post. Predictive check of type B mice.

4.3 Non-informative prior distributions

20

Suppose for a binary sampling problem we plan on using a uniform, or beta$(1, 1)$, prior for the population proportion θ. Perhaps we reason that this represents "no prior information about θ." However, some people like to look at proportions on the log-odds scale, that is, they are interested in $\gamma = log\frac{\theta}{1-\theta}$. Via Monte Carlo sampling or otherwise, find the prior distribution for γ that is induced by the uniform prior for θ. Is the prior informative about γ?

Solution:

Using Monte Carlo sampling we get our θ from the uniform distribution and then we compute our γ based on the sampled θ and represent our γ's distribution in the form of a histogram 4.18, as implemented by the following R code.

```
library(ggplot2)


theta.mc <- runif(1000)
gamma.mc <- sapply(theta.mc, function(i) log(i / (1 - i)))
normal.mc <- rnorm(1000, mean = 0, sd = 1.5)

# Combine all the data
data_df <- data.frame(
  value = c(gamma.mc, theta.mc, normal.mc),
  type = rep(c("Gamma", "Theta", "Normal"), each = 1000))
```

[20] [DHo09]pg 234 Exercise 4.6

```r
# Create a histogram
ggplot(data_df, aes(x = value, fill = type)) +
  geom_histogram(
    binwidth = 0.5, position = "identity",
    alpha = 0.5, aes(y = ..density..)) +
  scale_fill_manual(
    values = c("Gamma" = "blue",
               "Theta" = "red",  "Normal" = "green")) +
labs(y = "Density") +
xlim(min(gamma.mc), max(gamma.mc))
```

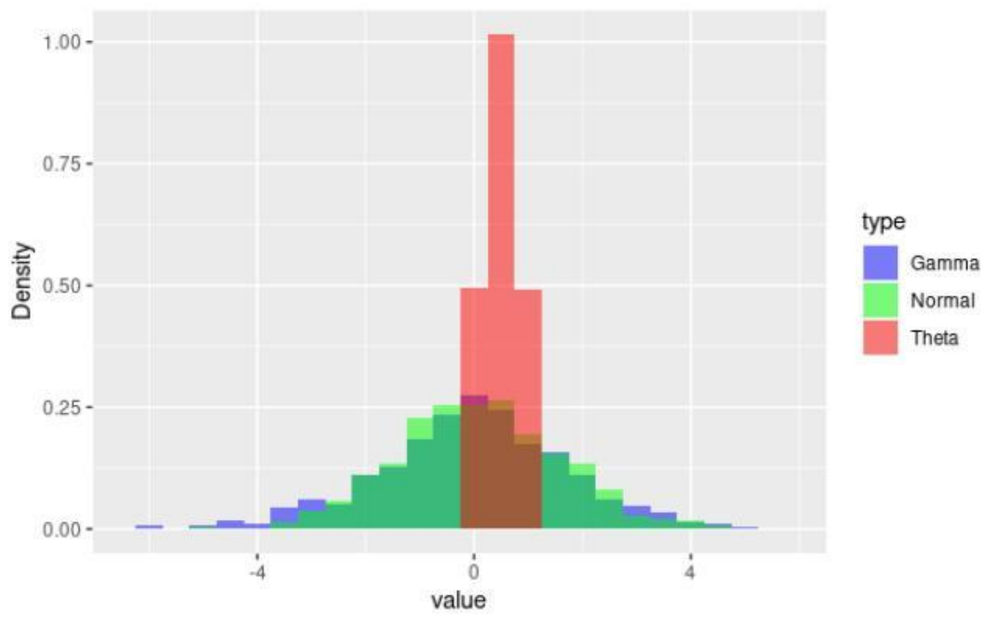


Figure 4.18: Histogram of γ distribution.

Based on our graph it appears as if by sampling our θ from a uniform and computing γ, forms a distribution similar to the normal with mean$= 0$ and standard deviation ≈ 1.5. So to the question of informative prior for γ, I say yes. A normal prior distribution can be seen as an informative distribution, [21] because to the values which are closer to the mean, the probability is greater and that could be seen as a prior bias from our part. The same logic applies to our distribution (see the plot 4.18)

[21] I would love some feedback on this...

4.4 Mixture models

22

After a posterior analysis on data from a population of squash plants, it was determined that the total vegetable weight of a given plant could be modeled with the following distribution:

$$\mathrm{p}(y|\theta,\sigma^2) = 0.31 \cdot \mathrm{dnorm}(y,\theta,\sigma) + 0.46 \cdot \mathrm{dnorm}(2\theta_1, 2\sigma) + 0.23 \cdot \mathrm{dnorm}(y, 3\theta_1, 3\sigma) \tag{4.44}$$

where the posterior distributions of the parameters have been calculated as $\frac{1}{\sigma^2} \sim$ gamma$(10, 2.5)$, and $\theta|\sigma^2 \sim$ normal$(4.1, \frac{\sigma^2}{20})$.

a) Sample at least 5.000 y values from the posterior predictive distribution.

b) Form a 75% quantile-based confidence interval for a new value of Y.

i. Compute estimates of the posterior density of Y using the density command in R and then normalize the density values so they sum to 1.

ii. Sort these discrete probabilities in decreasing order.

iii. Find the first probability value such that the cumulative sum of the sorted values exceeds 0.75. Your HPD region includes all values of y which have a discretized probability greater than this cutoff. Describe your HPD region, and compare it to your quantile-based region.

c) Can you think of a physical justification for the mixture sampling distribution of Y?

Solution:

a) Suppose the model is the predictive posterior distribution of which I need to sample from. The below code produces the plot (4.19):

```
N = 10000
sigma2.mc = 1 / rgamma(N, 10, 2.5)
theta.mc = rnorm(N, 4.1, sigma2.mc / 20)
sigma.mc = sqrt(sigma2.mc)
ynew.mc = .31 * rnorm(N, theta.mc, sigma.mc) +
  .46 * rnorm(N, 2 * theta.mc, 2 * sigma.mc) +
  .23 * rnorm(N, 3 * theta.mc, 3 * sigma.mc)
```

[22] [DHo09]pg.234 Exercise 4.7

```r
# Create a histogram for ynew.mc using ggplot2
ggplot(data.frame(ynew.mc), aes(x = ynew.mc)) +
  geom_histogram(binwidth = 0.1, fill = "blue",
                 color = "black") +
  labs(x = "ynew.mc", y = "Frequency",
       title = "Histogram of ynew.mc")
```

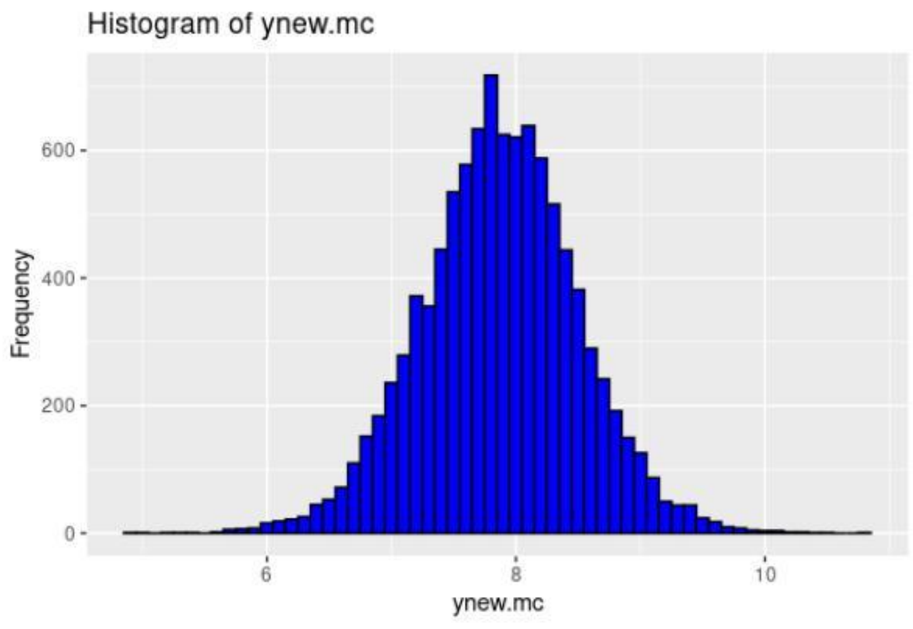


Figure 4.19: Histogram of y_{new} simple version.

The posterior predictive distribution is indeed defined as:

$$p(y^* \mid y_1, \ldots, y_n) = \int_0^\infty p(y^* \mid \theta, y_1, \ldots, y_n) \cdot p(\theta \mid y_1, \ldots, y_n)\, d\theta$$

Now, if you assume conditional independence, which means that y^* is conditionally independent of θ given the observed data $y_1, \ldots, y_n$, then you can simplify this to:

$$p(y^* \mid y_1, \ldots, y_n) = \int_0^\infty p(y^* \mid \theta) \cdot p(\theta \mid y_1, \ldots, y_n)\, d\theta$$

This means that the predictive distribution for y^*, after observing $y_1, \ldots, y_n$, is a result of integrating over all possible values of θ, weighted by the likelihood $p(y^* \mid \theta)$ and the posterior distribution of θ given the observed data $p(\theta \mid y_1, \ldots, y_n)$.
This simplified expression is essentially equal to the model $p(y^* \mid \theta)$ integrated over the posterior distribution of θ, $p(\theta \mid y_1, \ldots, y_n)$. So, when you sample from the model with θ values drawn from their posterior distribution, you are indeed sampling from the posterior predictive distribution for y^* after observing $y_1, \ldots, y_n$.
b)i) The result of a 75% quantile-based confidence interval for a new value of Y, based on our sampled y_{new} is $7.155779 - 8.592139$, found using an R function:

```
> quantile(ynew.mc, c(.125, .875))
```

To normalize the density I used the function:

```
y_density = density(ynew.mc)
y_density$y = y_density$y / sum(y_density$y)
```

ii) The request is fulfilled by running the R code:

```
sorted_indices = order(y_density$y, decreasing = TRUE)
sorted_probs = y_density$y[sorted_indices]
sorted_y_values = y_density$x[sorted_indices]
```

iii) To answer this question we just need to run the below lines of R code:

```
# Find the first probability value exceeding 0.75
cumsum_probs = cumsum(sorted_probs)
cutoff_index = which(cumsum_probs > 0.75)[1]
cutoff_value = sorted_y_values[cutoff_index]

# HPD region
hpd_region = sorted_y_values[1:cutoff_index]

#Comparison with the quantile based:
quantile_interval = quantile(y.new.mc, probs = c(0.125, 0.875))

##################################################
> cat("75% Quantile-Based Confidence Interval:",
    quantile_interval, "\n")
#75% Quantile-Based Confidence Interval: 7.312753 8.409216
> cat("HPD Region:", range(hpd_region), "\n")
#HPD Region: 7.315658 8.424075
##################################################

# Visualization
hist(y.new.mc, probability = TRUE,
```

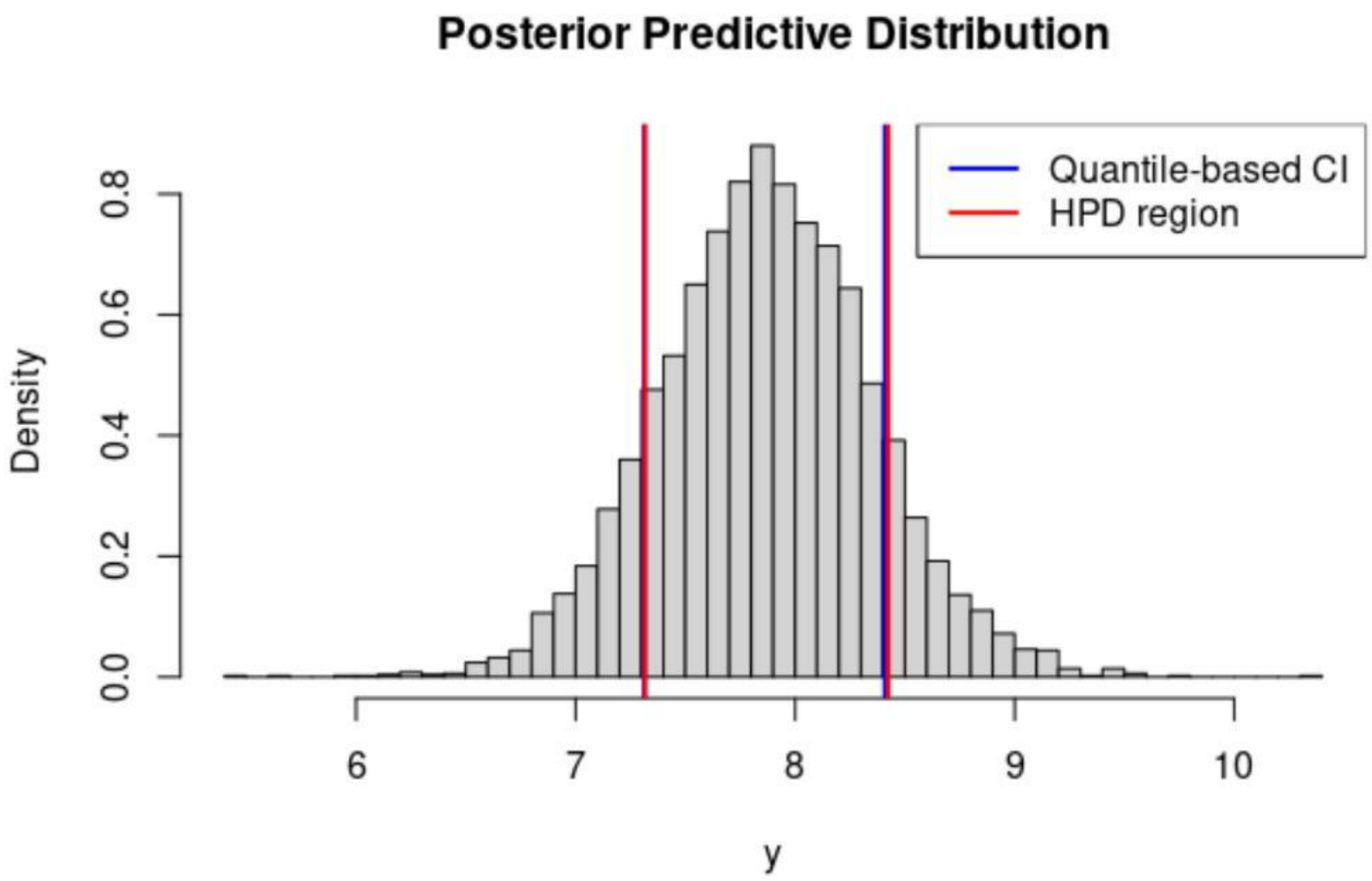


Figure 4.20: Visualization

```
  main = "Posterior Predictive Distribution",
  xlab = "y", breaks = 50, col = "lightgray")
abline(v = quantile_interval, col = "blue", lwd = 2)  # Q-based
abline(v = range(hpd_region), col = "red", lwd = 2)  # HPD
legend("topright", legend = c(
  "Quantile-based CI", "HPD region"), col = c("blue", "red"),
  lwd = 2)
```

d) I suppose it makes sense for the weight of the vegetables to be uniformly distributed and then assign a weight probably based on the possibility of it being the one you picked. Like if I had 40 tomatoes and 60 carrots I could understand a model that weights $0.4\times$uniform(tomato) + $0.6\times$ uniform(carrot). If the question was if it makes sense for the quantile-based and the HPD regions to be almost identical, it also does because we have a distribution with one bell.

4.5 More posterior predictive checks

[23]

Let θ_A and θ_B be the average number of children of men in their 30_s with and without bachelor's degrees, respectively.

a) Using a Poisson sampling model, a gamma$(2,1)$ prior for each θ and the data in the files menchild30bach.dat and menchild30nobach.dat, obtain 5.000 samples of $\widetilde{Y_A}$ and $\widetilde{Y_B}$ from the posterior predictive distribution of the two samples. Plot the Monte Carlo approximations to these two posterior predictive distributions.

b) Find 95% quantile-based posterior confidence intervals for $\theta_B - \theta_A$ and $\widetilde{Y_A} - \widetilde{Y_B}$. Describe in words the differences between the two populations using these quantities and the plots in a), along with any other results that may be of interest to you.

c) Obtain the empirical distribution of the data in group B. Compare this to the Poisson distribution with mean $\tilde{\theta} = 1.4$. Do you think the Poisson model is a good fit? Why or why not?

d) For each of the 5.000 θ_B-values you sampled, sample $n_B = 218$ Poisson random variables and count the number of 0s and the number of 1_s in each of the 5.000 simulated datasets. You should now have two sequences of length 5.000 each, one sequence counting the number of people having zero children for each of the 5.000 posterior predictive datasets, the other counting the number of people with one child. Plot the two sequences against one another (one on the x-axis, one on the y-axis). Add to the plot a point marking how many people in the observed dataset had zero children and one child. Using this plot, describe the adequacy of the Poisson model.

Solution):

a) Using the poison model and the prior gamma$(2,1)$ we acquire posterior $\theta \sim$ gamma$(a = 2+\sum_{i=1}^{n_0}, b = 1+n_0)$, conjugate prior. Then we will infuse our posterior sampled data inside the model and the sampled and plotted 4.21 as we did in 4.4 to obtain the 5.000.

```
m30b=scan(url(
  'http://www2.stat.duke.edu/~pdh10/FCBS/Exercises/menchild30bach.dat'
  ))
m30nob =scan(url(
  'http://www2.stat.duke.edu/~pdh10/FCBS/Exercises/menchild30nobach.dat'
```

[23] [DHo09] pg. 235 Exercise 4.8

```
  ))

N = 5000
thetaA = rgamma(N, 2 + sum(m30b), 1 + length(m30b))
thetaB = rgamma(N, 2 + sum(m30nob), 1 + length(m30nob))

ynewA = rpois(N, thetaA)
ynewB = rpois(N, thetaB)

df=data.frame(
  ynew = c(ynewA, ynewB),
  dist = factor(rep(c('bach', 'nobach'),
                    each = N), levels = c('bach', 'nobach')),
  type = 'posterior'
)
ggplot(df, aes(x = ynew, group = dist)) +
  geom_histogram() +
  facet_grid(. ~ dist)
```

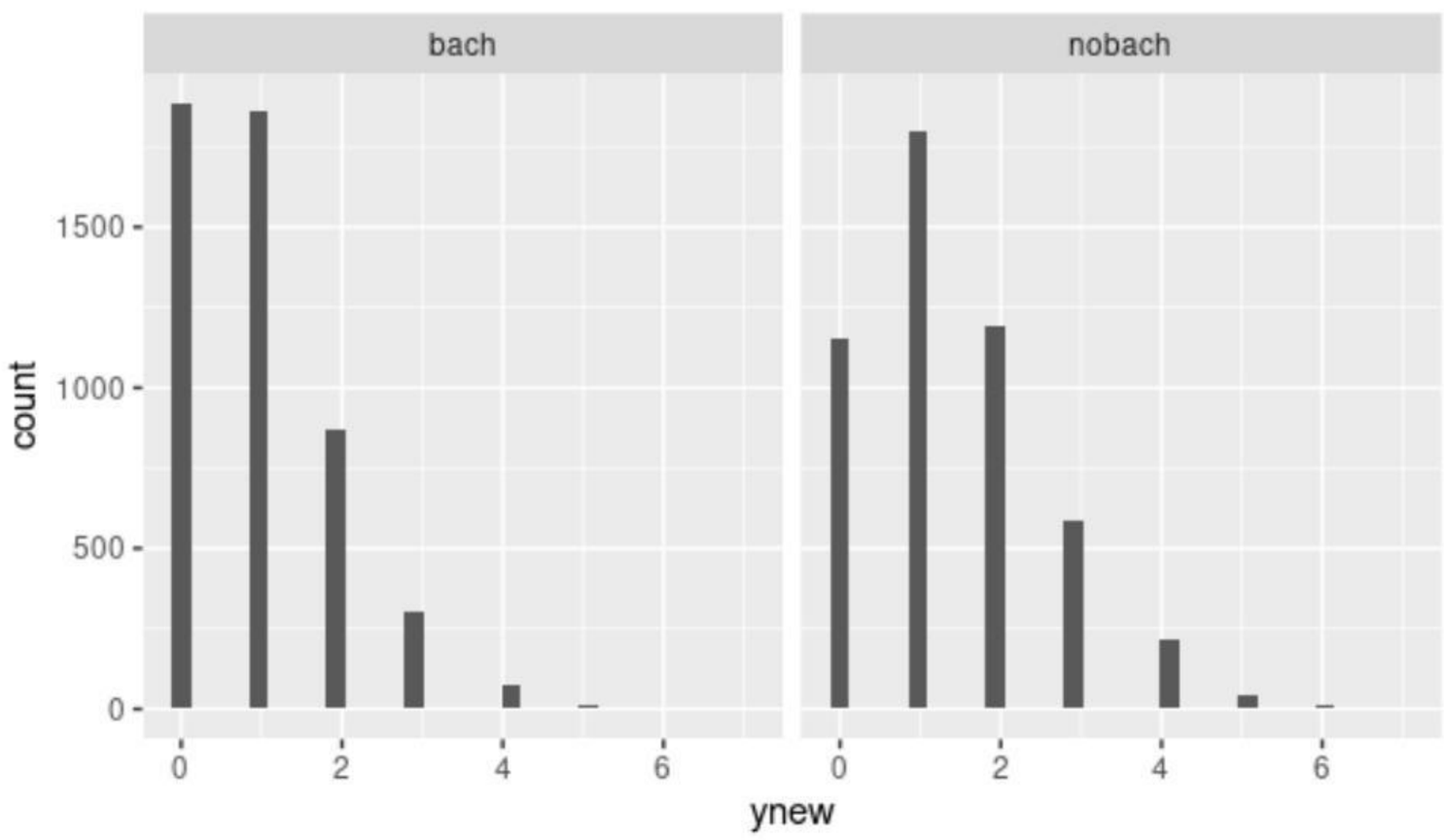


Figure 4.21: Histogram of frequencies of y_{new_A} and y_{new_B} respectively

b) Based on the 4.21 plot if you choose a random θ_B there is a high probability of it being larger than the corresponding random choice of a θ_A, but based on our observed data the posterior gamma of θ_b will have less uncertainty as $a >> b$ and

larger values than the corresponding gamma of θ_A. Since the 95% quantile-based posterior confidence interval $0.025 - 0.975$ is very small, it provides certainty and confirms our earlier observation that θ_B are most likely larger, also the posterior mean$\theta_B \approx 1.4$ is greater than mean$\theta_A \approx 0.95$.
In the case of $y_{new_B} - y_{new_B}$ the values fluctuate from -2 to 3 based on our 95% quantile-based posterior confidence interval, which in first look does not correspond with the frequencies and the values in the plot (4.21) , where is appears as if it would have a higher probability of being greater than a random choice of y_{new_A}. Maybe this relationship would be more obvious if you choose a smaller confidence interval, though based on this I have to note that it holds great uncertainty about the value of $y_{new_B} - y_{new_B}$, even though our "parameters", posterior samples of θ seem set in stone.

```
> theta.d = thetaB - thetaA
> print(quantile(theta.d, c(0.025, 0.975)))
     2.5%     97.5%
0.1554364 0.7347836
>
> ynew.d = ynewB - ynewA
> print(quantile(ynew.d, c(0.025, 0.975)))
 2.5% 97.5%
   -2     3
```

c)The empirical comes from using our data to sample from the ep model, So for the Poisson model with a mean($\theta = 1.4$).Based on the plot 4.22 there is an obvious difference between the empirical and the posterior, mostly in their estimates about 0, 1 and 6 children. Is it a good fit? It depends on what you hope to accomplish, a dissimilarity is obvious, for sure, but to measure this model's usefulness we need a statistic, to base our answer, if our statistic is similar to [DHo09]pg. 63 then yes our model is not a good fit, based on the posterior distribution for $\tilde{Y}_B$ peaking around 1, while the empirical peaks strongly around 0. To conclude it it comes down to what we wish to accomplish using this model.

```
#Reminder
ynewB = rpois(N, thetaB)


df = data.frame(
```

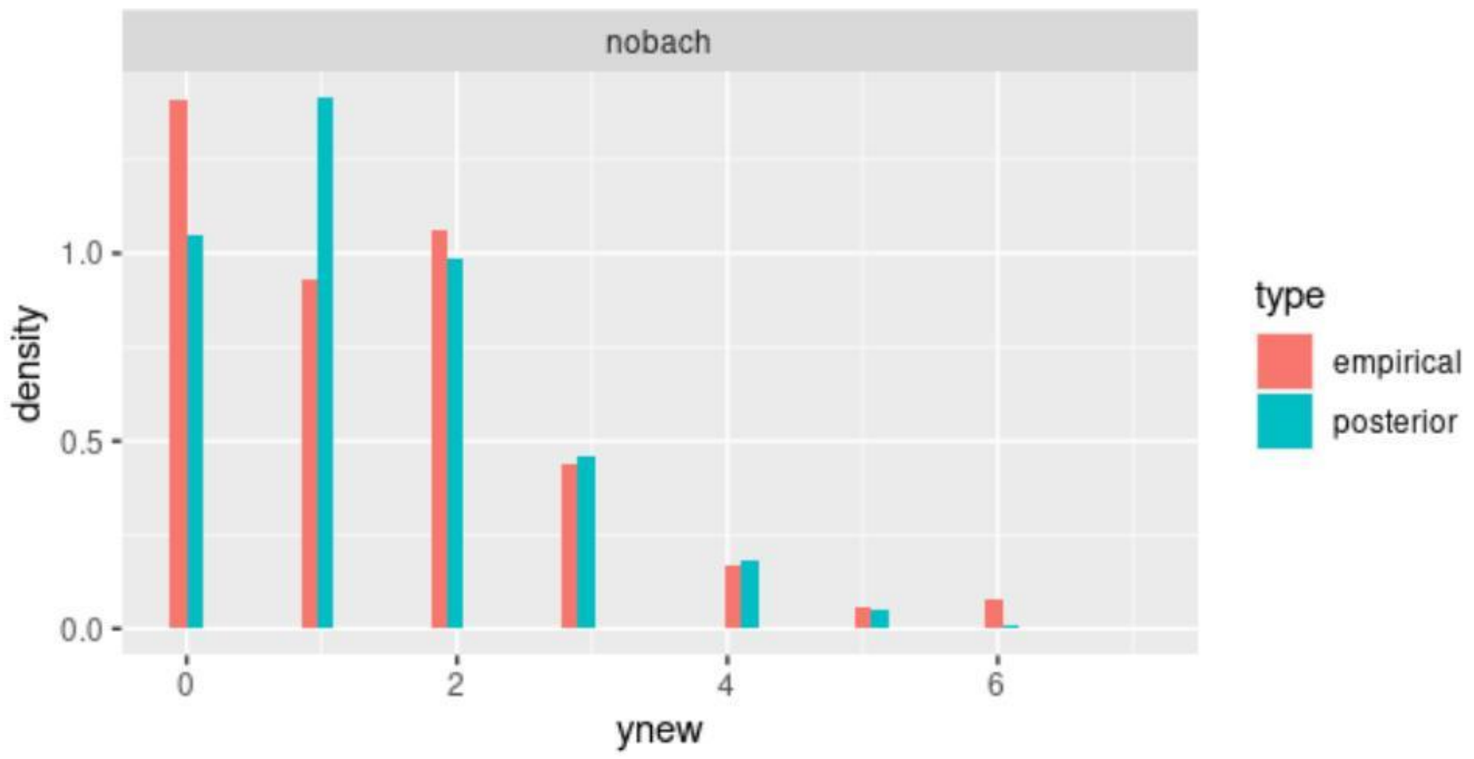


Figure 4.22: Plot of the empirical and the Poisson, specified in c)

```
  ynew = ynewB,
  dist = factor(rep('nobach', N)),
  type = 'posterior')

# Create an empirical
emp.df = data.frame(
  ynew = m30nob,
  dist = factor(rep('nobach', length(m30nob))),
  type = 'empirical')

# Combine the data frames
total.df = rbind(df, emp.df)

# The histogram
ggplot(total.df, aes(x = ynew, y = ..density..,
                     group = type, fill = type)) +
  geom_histogram(position = 'dodge') +
  facet_grid(. ~ dist)
```

d)I will start by providing the code for the exercise's plot 4.23.

```
N = 5000
count_zero_children = numeric(N)
count_one_child = numeric(N)
```

```
for (i in 1:N) {
  # Sample thetaB from the posterior distribution
  thetaB = rgamma(1, 2 + sum(m30nob), 1 + length(m30nob))

  simulated_data = rpois(218, thetaB)

  count_zero_children[i] = sum(simulated_data == 0)
  count_one_child[i] = sum(simulated_data == 1)
}

counts_df = data.frame(ZeroChildren = count_zero_children,
                       OneChild = count_one_child)

# The scatter plot
ggplot(counts_df, aes(x = ZeroChildren, y = OneChild)) +
  geom_point() +
  geom_point(aes(x = sum(m30nob == 0), y = sum(m30nob == 1)),
             color = "red", size = 3) +
  labs(x = "Count of Zero Children", y = "Count of One Child",
       title = "Counts in Simulated Datasets vs. Observed Data")
```

I decided to answer this question using the plot 4.24 which was produced with the code below.

```
N = 5000
thetaB = rgamma(N, 2 + sum(m30nob), 1 + length(m30nob))

count_zero_children <- numeric(N)
count_one_child <- numeric(N)

for (i in 1:N) {

  simulated_data = rpois(218, thetaB[i])

  count_zero_children[i] = sum(simulated_data == 0) / 218
```

```
  count_one_child[i] = sum(simulated_data == 1) / 218  }

count_ones<-sum(m30nob== 1)/218
count_zero<-sum(m30nob== 0)/218

sim_data <- data.frame(
  Simulation_Number = 1:N,
  Count_Zero_Children = count_zero_children,
  Count_One_Child = count_one_child)

# The plot
plot <- ggplot(sim_data, aes(x = Simulation_Number)) +
  geom_line(aes(y = Count_Zero_Children,
                color = "Zero Children")) +
  geom_line(aes(y = Count_One_Child,
                color = "One Child")) +
  geom_hline(yintercept = count_zero, acolor = "red",
             linetype = "dotted", size = 1) +
  geom_hline(yintercept = count_ones, color = "red",
             linetype = "dotted", size = 1) +
  scale_color_manual(values = c("Zero Children" = "blue",
                                "One Child" = "green")) +
  labs(x = "Simulation Number", y = "Probability",
       title = "Simulation Results") +
  theme_minimal()

print(plot)
```

In the first plot, we can see the count of zero children and one child spread across a scatter plot, with the red dot as the value provided by the observed data and it could only be an outlier value based on the poison model, thus making it unrelatable with the reality of our population making it less worthy. I took the liberty to produce a plot representing their probabilities 4.24, in which at first look everything appears to be fine, as the observed probability values noted with the red break-line appear to be some sort of mean value to the one child and zero children simulated data vector, where the first break line corresponds to the zero children and the second to

the one's. The poison distribution is not a good fit based on this statistics and that is more easily observed in the first graph 4.23

```
> count_ones<-sum(m30nob== 1)/218
[1] 0.2247706
> count_zero<-sum(m30nob== 0)/218
[1] 0.3394495
> d_zero<- mean(count_zero_children)- count_zero
[1] -0.0930633
> d_one<- mean(count_one_child)-count_ones
[1] 0.1205183
```

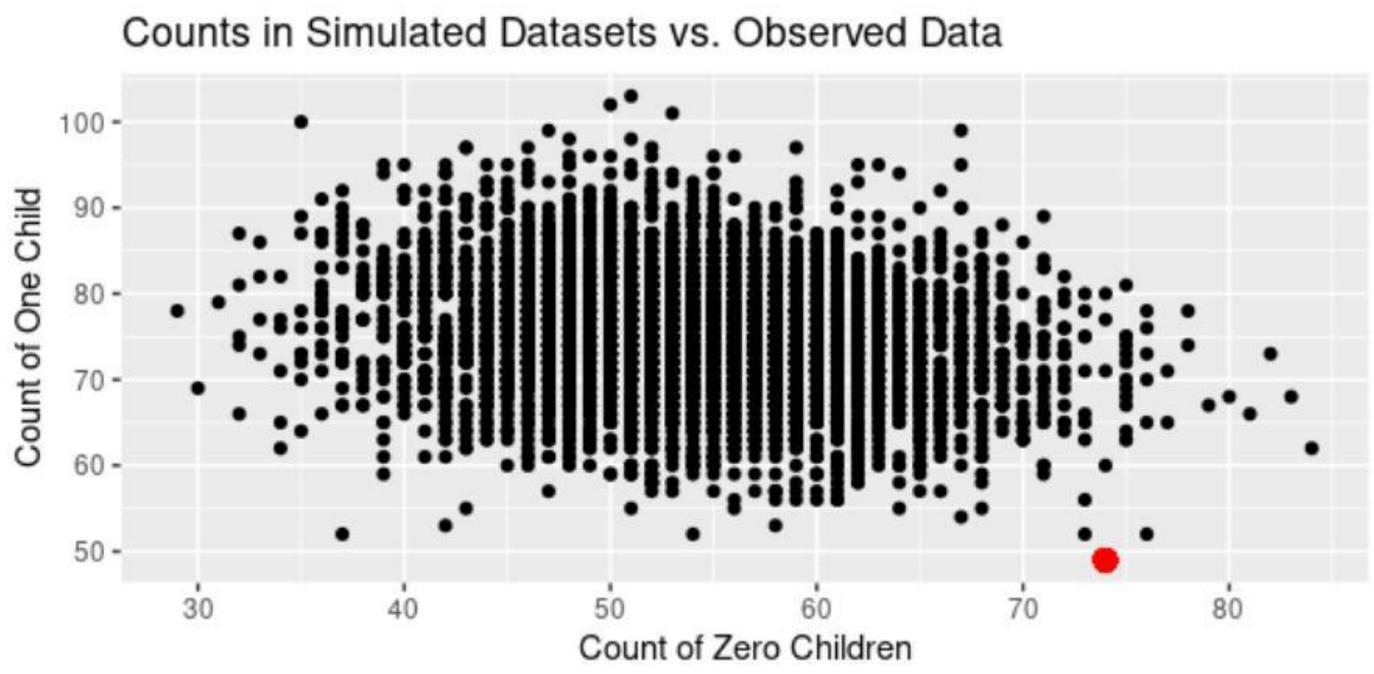


Figure 4.23: Plot representing the relationship between sim and observed data.

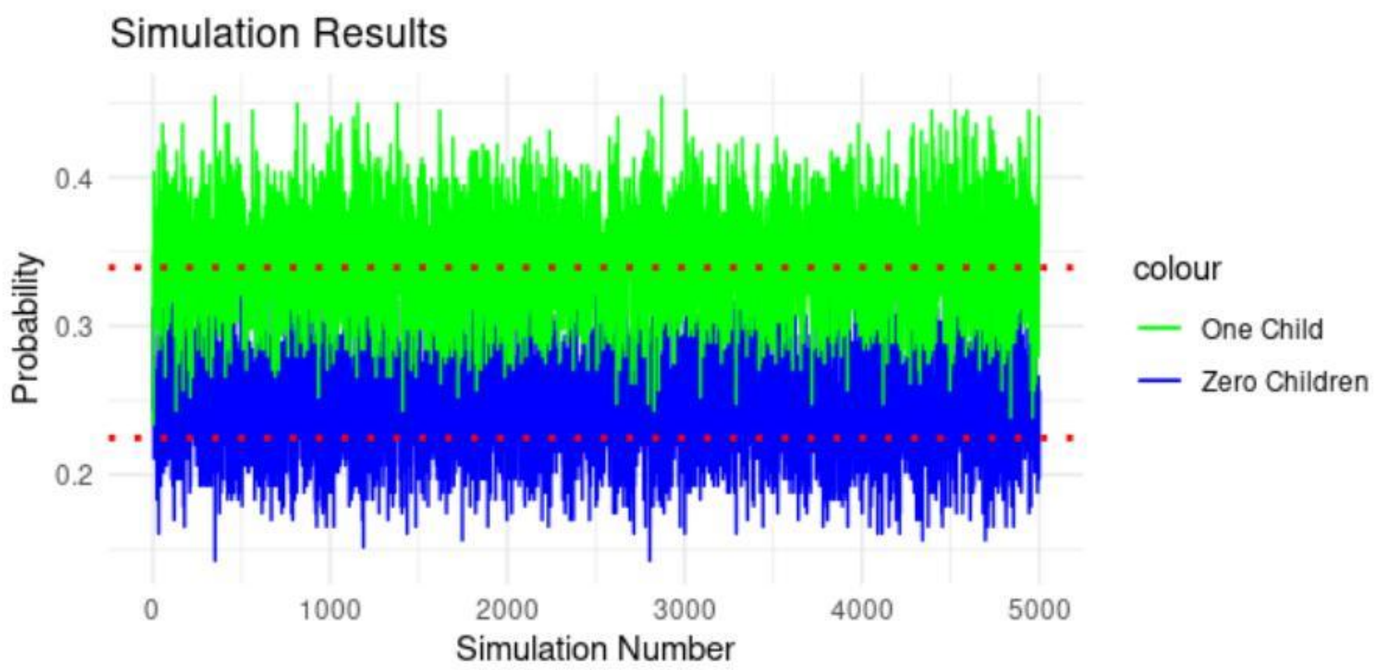


Figure 4.24: Plot representing the relationship between sim and observed data.

4.6 Problem 3.4

24

You are the assistant coach of the women's softball team at a Midwestern college. The head coach has asked you to assess a new first year player who is joining the team. As a high school student, she was at bat 120 times and got 40 hits. You wish to estimate θ, her underlying true probability of getting a hit in any at bat as a college-level player.

1. Specify a beta prior that seems appropriate to capture your knowledge or uncertainty about θ before the new player plays in any college-level games. Use any information you have that seems important (her high school record), anything you know about college-level women's softball, etc. Use R functions as needed. Explain in a few sentences (supplemented with plots and/or R output) how you chose the values of α and β. There is no one right answer here (I want to see how you think about this and what procedure you use).

2. Specify a beta prior that you think might reflect the player's mother's beliefs about θ. This may be similar to, or quite different from, your prior. Again, justify your choice with graphical or numeric R output.

3. Suppose the player now plays eight college-level games, has thirty at bats, and gets 5 hits. Thus, the data are

$$y = 5,\ n = 30 \tag{4.45}$$

We will use a binomial likelihood for these data. This requires the assumption that, conditional on θ, each at bat is an independent Bernoulli trial with success probability θ. There are several reasons why independence might actually not be a reasonable assumption in this problem. Give one.

Note: For our present purposes, we'll use a binomial likelihood anyway. We'll come up with a better model when we talk about hierarchical models later in the semester.

4. Obtain the following characteristics of the posterior distribution p($\theta|y$)

1. Name of posterior distribution and its parameter values.

2. Posterior density plot

3. A plot showing the prior density, the likelihood, and the posterior density, all on the same axes.

[24] [Cow13] pg.46 Problem 3.4

based on the data from the college-level games under each of three priors:

1. Your prior from part 1
2. The mother's prior from part 2
3. A uniform (noninformative) prior

Note This problem will be continued at the end of Chap. 4.[Cow13]

Solution:

1. In the context of this problem, each at-bat can be considered a Bernoulli trial, where there are two possible outcomes, a hit (success) or not a hit (failure). When you have a series of Bernoulli trials, the number of successes in a fixed number of trials follows a binomial distribution as we have seen in [DHo09]ch.3. Furthermore the Beta distribution is the conjugate prior to the binomial likelihood, so the posterior distribution is also Beta. This property simplifies the computational and analytical process significantly because we can find the posterior distribution analytically. Thought our efforts must be in accuracy so a question arose "Is the beta suitable for modeling an athlete's hit rate over time?". The Beta distribution is indeed suitable cause it's flexible and can model various shapes of distributions, depending on the α and β parameters, which allows it to represent different levels of belief about the athlete's true hit rate.

In the Beta distribution, both α and β parameters influence the shape of the distribution, and hence, our belief about the parameter θ, α represents the number of successes, and increasing it will shift the distribution towards 1, indicating a belief in a higher success rate, while β represents the number of failures + 1 and increasing it will shift the distribution towards 0, indicating a belief in a lower success rate. An obvious choice would be beta(40,80) as it has mean value = $\frac{40}{120}$, which is the hit rate, though increasing both α and β while keeping their ratio constant will increase the concentration of the distribution around the mean, reducing the uncertainty. Thought in my hamble opinion we should use the beta(1,2) prior as it increases our uncertainty and has the same mean value, moreover, beta$(1, 2)$ distribution has a much wider spread, indicating that we are considering a wide range of possible values for the player's true hit rate, which seems reasonable given that players' performances can fluctuate, especially when transitioning from high school to college

level, where the competition is generally tougher. So, adopting a beta(1,2) seems like a thoughtful choice as it allows significant change in the player's performance and it gives her the benefit of the doubt, so to speak, while still leaning slightly towards a lower hit rate based on her high school performance as presented in the plot 4.25.

$$f(\theta;\alpha,\beta) = \frac{\theta^{\alpha-1}(1-\theta)^{\beta-1}}{B(\alpha,\beta)} \Rightarrow$$
$$f(\theta;1,2) = \theta^{1-1}(1-\theta)^{2-1} = 2(1-\theta)$$

```
# The Beta(1, 2) prior
beta_1_2 <- function(theta) {
  dbeta(theta, 1, 2)}

theta_vals <- seq(0, 1, length.out = 1000)

# The corresponding density values
p_vals_beta_1_2 <- beta_1_2(theta_vals)

data <- data.frame(
  theta = theta_vals,
  p_beta_1_2 = p_vals_beta_1_2
)

# The Plot
ggplot(data) +
  geom_line(aes(x = theta, y = p_beta_1_2, colour =
                  "Beta(1,2) Prior")) +
  labs(title = "Beta(1,2) Prior Distribution",
       x = expression(theta),
       y = "Density") +
  theme_minimal()
```

2. To construct a prior that might reflect the player's mother's beliefs about θ, we propose to use a mixed prior constructed from a weighted sum of different beta distributions. This approach allows us to encapsulate a range of potential attitudes a

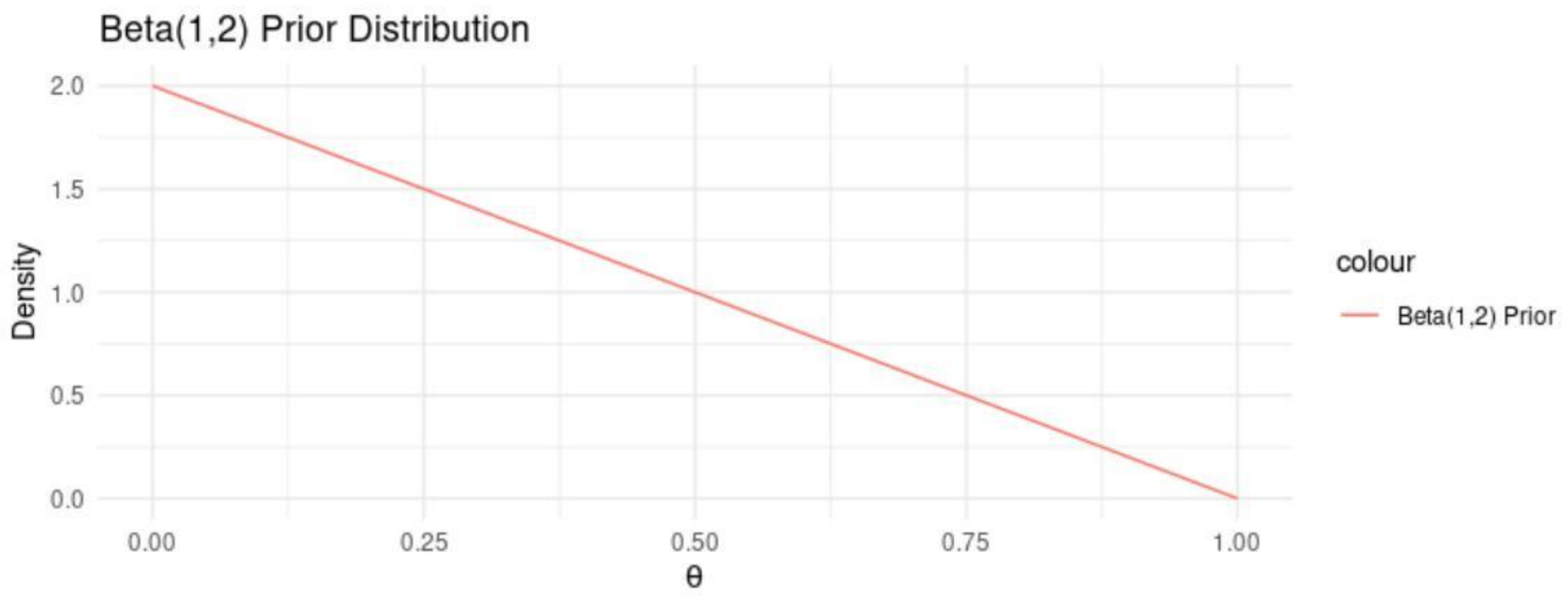


Figure 4.25: Plot representing my pessimistic prior which is linear.

mother might have towards her child's abilities in a single prior distribution. The rationale behind choosing this approach is to represent different perspectives, including a not-supportive perspective possibly influenced by jealousy, a truthful perspective based on the player's high school performance, and a supportive perspective where the mother is overly optimistic about her child's abilities.

We define the mixed prior as follows:

$$p(\theta) = w_1 \cdot \text{Beta}(\theta; \alpha_1, \beta_1) + w_2 \cdot \text{Beta}(\theta; \alpha_2, \beta_2) + w_3 \cdot \text{Beta}(\theta; \alpha_3, \beta_3)$$

,where the weights w_1, w_2, and w_3 satisfy $w_1 + w_2 + w_3 = 1$. The parameters of the beta distributions are chosen to represent different attitudes as described below:

1. **Not Supportive Mother** (Possibly influenced by jealousy):
   - Beta distribution parameters: $\alpha_1 = 2$, $\beta_1 = 6$ (mean = 0.25)
   - Weight (w_1): 0.1
2. **Truthful Mother** (Based on the player's high school performance):
   - Beta distribution parameters: $\alpha_2 = 40$, $\beta_2 = 80$ (mean = 0.333)
   - Weight (w_2): 0.5
3. **Supportive Mother** (Possibly overestimating the player's abilities):
   - Beta distribution parameters: $\alpha_3 = 50$, $\beta_3 = 50$ (mean = 0.5)
   - Weight (w_3): 0.4

To visualize this mixed prior alongside the individual beta distributions and have a comprehensive view of the constructed prior as well as the plots it originated from 4.26 we can use the code:

```
theta_seq <- seq(0, 1, length.out = 1000)

# Define the individual beta distributions
beta_not_supportive <- dbeta(theta_seq, 2, 6)
beta_truthful <- dbeta(theta_seq, 40, 80)
beta_supportive <- dbeta(theta_seq, 50, 50)

# Define the mixed prior
mixed_prior <- 0.1 * beta_not_supportive +
  0.5 * beta_truthful + 0.4 * beta_supportive

data <- data.frame(
  theta = rep(theta_seq, 4),
  density = c(beta_not_supportive,
              beta_truthful, beta_supportive, mixed_prior),
  distribution = rep(c(
    "Not Supportive", "Truthful", "Supportive",
    "Mixed Prior"), each = 1000)
)

# Plot the distributions
ggplot(data, aes(x = theta, y = density,
                 color = distribution)) +
  geom_line() +
  labs(title = "Mother's Prior Beliefs",
       x = expression(theta), y = "Density") +
  theme_minimal()
```

I would like to show the smoothened mixed prior using the Monte Carlo sampling from each beta separately in comparison to our mixed (analytical) prior 4.27.

```
n_samples <- 10000; weights <- c(0.1, 0.5, 0.4)
```

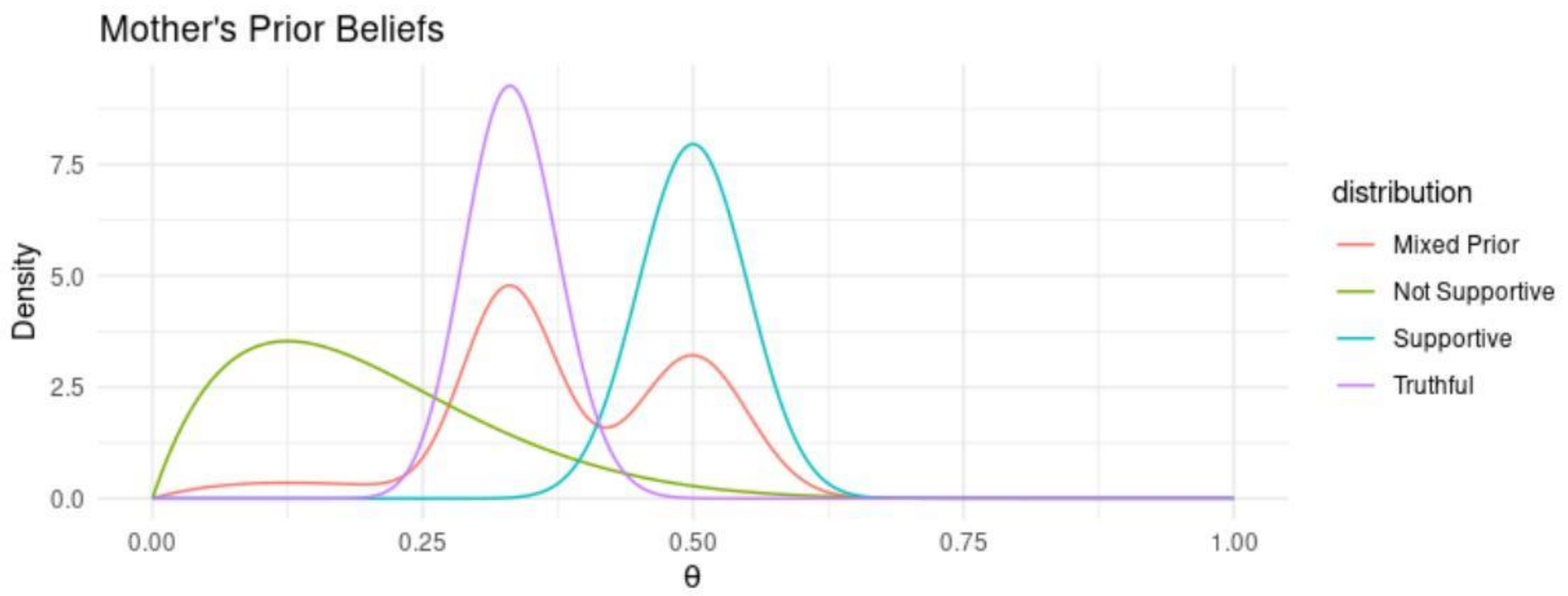


Figure 4.26: Plot representing all the priors.

```r
samples <- c(
  rbeta(n_samples * weights[1], 2, 6),
  rbeta(n_samples * weights[2], 40, 80),
  rbeta(n_samples * weights[3], 50, 50))

data <- data.frame(theta = samples)
theta_vals <- seq(0, 1, length.out = 1000)

# The mixed prior function
mixed_prior <- function(theta) {
  weights[1] * dbeta(theta, 2, 6) +
    weights[2] * dbeta(theta, 40, 80) +
    weights[3] * dbeta(theta, 50, 50)}

density_vals <- mixed_prior(theta_vals)

mixed_prior_data <- data.frame(
  theta = theta_vals, density = density_vals)

# The plot
ggplot() +
  geom_density(data = data,
               aes(x = theta), fill = "blue",
               alpha = 0.5, color = "blue") +
```

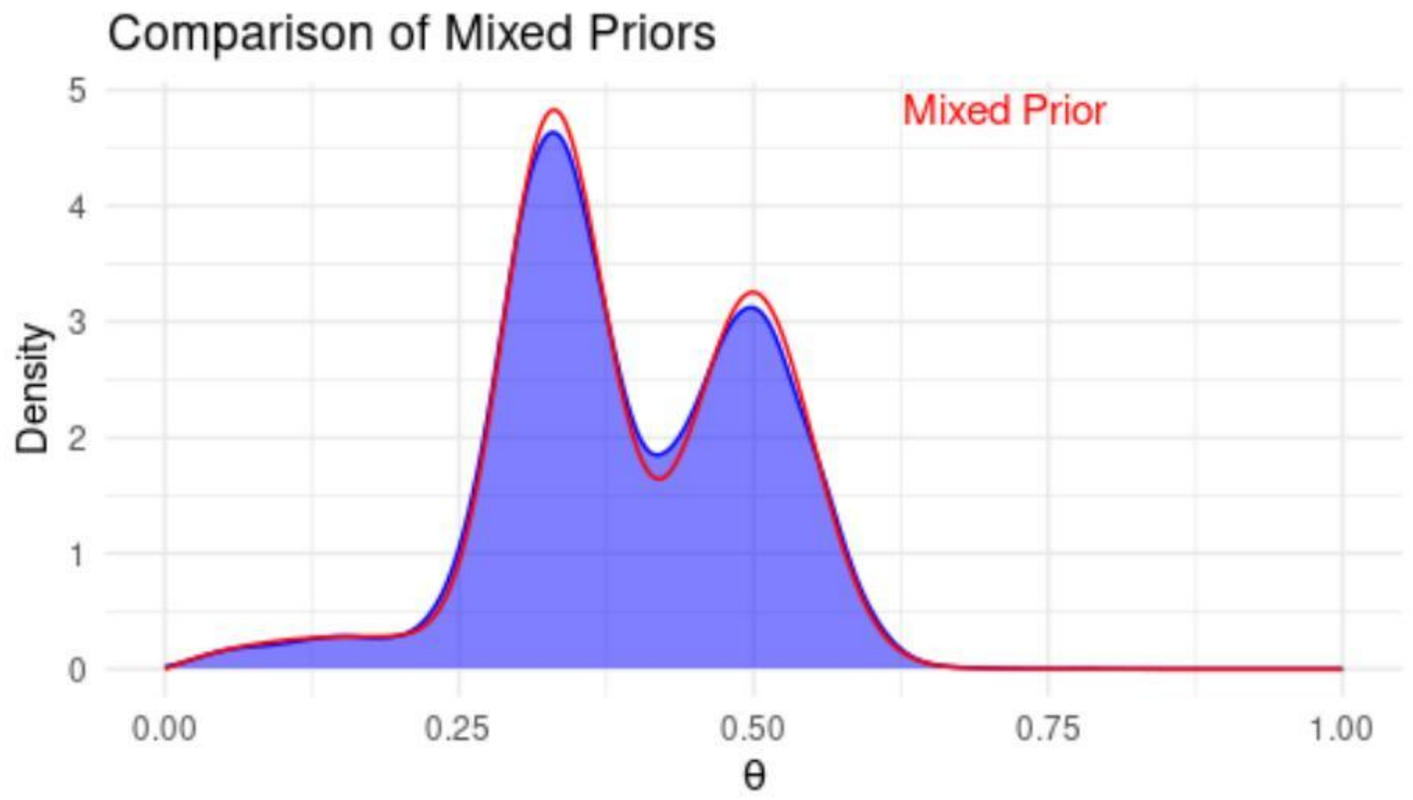


Figure 4.27: Comparing Monte Carlo sampling with analytical.

```
geom_line(
  data = mixed_prior_data,
  aes(x = theta, y = density), color = "red") +
geom_text(data = data.frame(
  x = 0.8, y = max(mixed_prior_data$density),
  label = "Mixed Prior"),
            aes(x = x, y = y, label = label),
  color = "red", size = 4, hjust = 1) +
labs(title = "Comparison of Mixed Priors",
     x = expression(theta), y = "Density") +
theme_minimal()
```

3. In this part, the binomial model assumes that each trial (or in this case, each at-bat) is independent of the others and that each trial has the same probability of success (θ). However, in real-life scenarios, this might not be the case due to a variety of factors. one of which could be the player's Mental State as a player might be more confident after a series of successful hits, which could potentially increase her chance of getting a hit in the subsequent games or the reverse like in our case as from $\frac{40}{120} = \frac{1}{3}$ went to $\frac{5}{30} = \frac{1}{6}$

4. Given a Beta prior distribution for θ characterized by parameters α and β, and a binomial likelihood function, the posterior distribution for θ can be derived using the following update rule 2.2. Therefore, the posterior distribution is also a

Beta distribution, given by:

$$\text{Beta}(\alpha + y, \beta + n - y)$$

Thus using the pessimistic prior, our posterior is $\text{Beta}(6, 27)$ which is represented in the plot 4.28. As created using R:

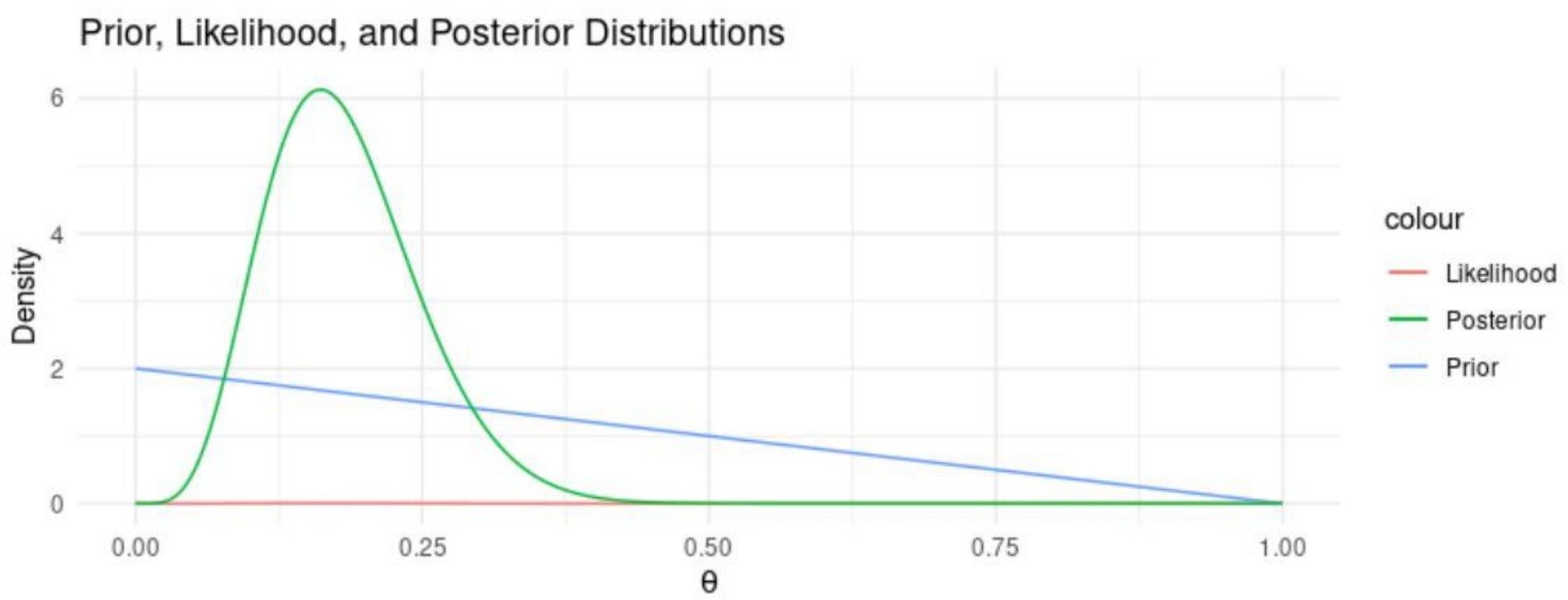


Figure 4.28: Posterior vs Prior× Likelihood.

```
alpha_prior <- 1; beta_prior <- 2
y <- 5; n <- 30

theta_vals <- seq(0, 1, length.out = 1000)

prior <- dbeta(theta_vals, alpha_prior, beta_prior)
likelihood <- dbinom(y, size = n, prob = theta_vals) /
  sum(dbinom(y, size = n, prob = theta_vals))
posterior <- dbeta(theta_vals, alpha_prior + y,
                   beta_prior + n - y)

data <- data.frame(
  theta = theta_vals,
  prior = prior,
  likelihood = likelihood,
  posterior = posterior)

#The plot
ggplot(data) +
```

```
  geom_line(aes(x = theta, y = prior, colour = "Prior")) +
  geom_line(aes(x = theta, y = likelihood, colour = "Likelihood")) +
  geom_line(aes(x = theta, y = posterior, colour = "Posterior")) +
  labs(title = "Prior, Likelihood, and Posterior Distributions",
       x = expression(theta),
       y = "Density") +
  theme_minimal()
```

Now for the uniform prior, with alpha_prior= 1 =beta_prior represented in the plot 4.29.

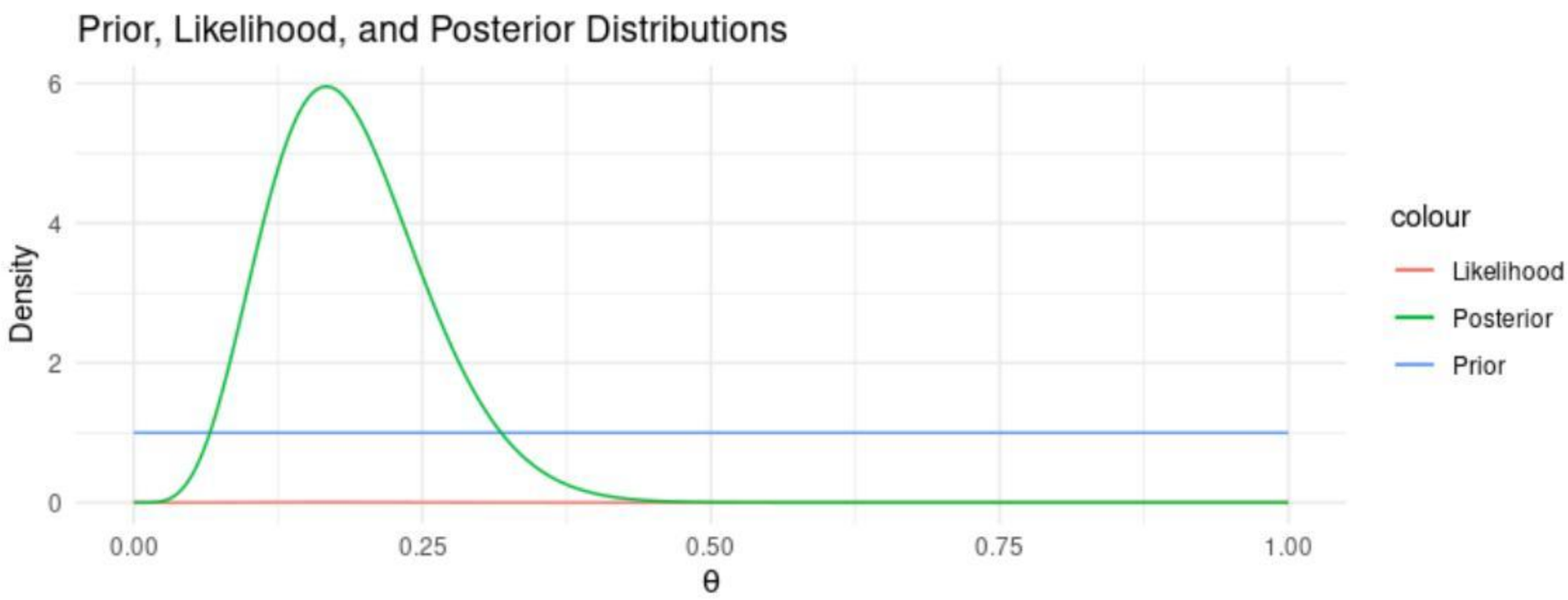


Figure 4.29: All for uniform.

For the mixture prior computing the true posterior distribution is more complex compared to using a single prior. The true approach to finding the posterior in this scenario is to consider it as a hierarchical model. This means that there is uncertainty not only about the parameter of interest, denoted as θ, but also about which component of the mixture is the "true" one.

To find the true posterior distribution, it is necessary to integrate all possible values of the parameter, considering each component of the mixture. This results in a more complex posterior.

Mathematically, the posterior distribution can be expressed as:

$$p(\theta|y) = \sum_i w_i p(\theta|y, \text{component i}) p(\text{component i}|y)$$

where:

- $p(\theta|y, \text{component i})$ is the posterior distribution of θ given the data and assuming component i of the mixture is the true component, which would be a beta distribution.

- $p(\text{component i}|y)$ is the posterior probability that component i is the true component, given the data. Finding this is not trivial.

- w_i are the weights of each component in the mixture.

Finding $p(\text{component i}|y)$ generally requires using methods like Markov Chain Monte Carlo (MCMC) to sample from the posterior distribution, which goes beyond the scope of a simple analytical solution. In our case an analytical solution can be found due to conjugacy. Before moving into the calculations for the analytical solution, we can explore a numerical method to approximate the posterior distribution using Monte Carlo sampling, using the steps noted below.:

1. **Sampling from the Mixed Prior:** We start by sampling a large number of values from our mixed prior. This involves sampling values from each component of the mixture according to their respective weights and then combining these samples to get a sample from the mixed prior.

2. **Calculating the Likelihood for Each Sample:** For each sampled value from the mixed prior, we calculate the likelihood of observing the data, which in this case is the likelihood of observing 5 hits in 30 at-bats, following a binomial distribution.

3. **Calculating the Unnormalized Posterior:** We then multiply the likelihood values by the prior values (the values of the PDF of the mixed prior at the sampled points) to get unnormalized posterior values.

4. **Normalizing the Posterior:** To get the posterior distribution, we normalize the unnormalized posterior values obtained in the previous step. This involves dividing each unnormalized posterior value by the sum of all unnormalized posterior values, ensuring that the posterior distribution integrates to 1.

5. **Estimating the Posterior Distribution:** Finally, we estimate the posterior distribution by plotting a histogram of the sampled values, weighted by the normalized posterior values. This gives us a discrete approximation of the posterior distribution.

Mathematically, the procedure can be summarized with the equation:

$$\text{posterior}(\theta|y) \propto \text{likelihood}(y|\theta) \times \text{prior}(\theta)$$

Where:

- posterior$(\theta|y)$: The posterior distribution of θ given the data y.
- likelihood$(y|\theta)$: The likelihood of observing the data y given θ, which follows a binomial distribution.
- prior(θ): The mixed prior distribution of θ.

This approach gives us a numerical approximation of the posterior distribution. Although it is a heuristic method and might not yield the exact posterior distribution, it can provide a reasonably good approximation if a sufficiently large number of samples are used as presented in the below plot 4.30. Using the R code:

```
alpha1 <- 2; beta1 <- 6
alpha2 <- 40; beta2 <- 80
alpha3 <- 50; beta3 <- 50
weights <- c(0.1, 0.5, 0.4)

set.seed(123)
samples <- c(
  rbeta(50000, alpha1, beta1),
  rbeta(50000, alpha2, beta2),
  rbeta(50000, alpha3, beta3))

y <- 5; n <- 30
likelihoods <- dbinom(y, n, samples)

unnormalized_posterior <- likelihoods * (
  weights[1] * dbeta(samples, alpha1, beta1) +
    weights[2] * dbeta(samples, alpha2, beta2) +
    weights[3] * dbeta(samples, alpha3, beta3))

# Normalize
normalized_posterior <- unnormalized_posterior /
```

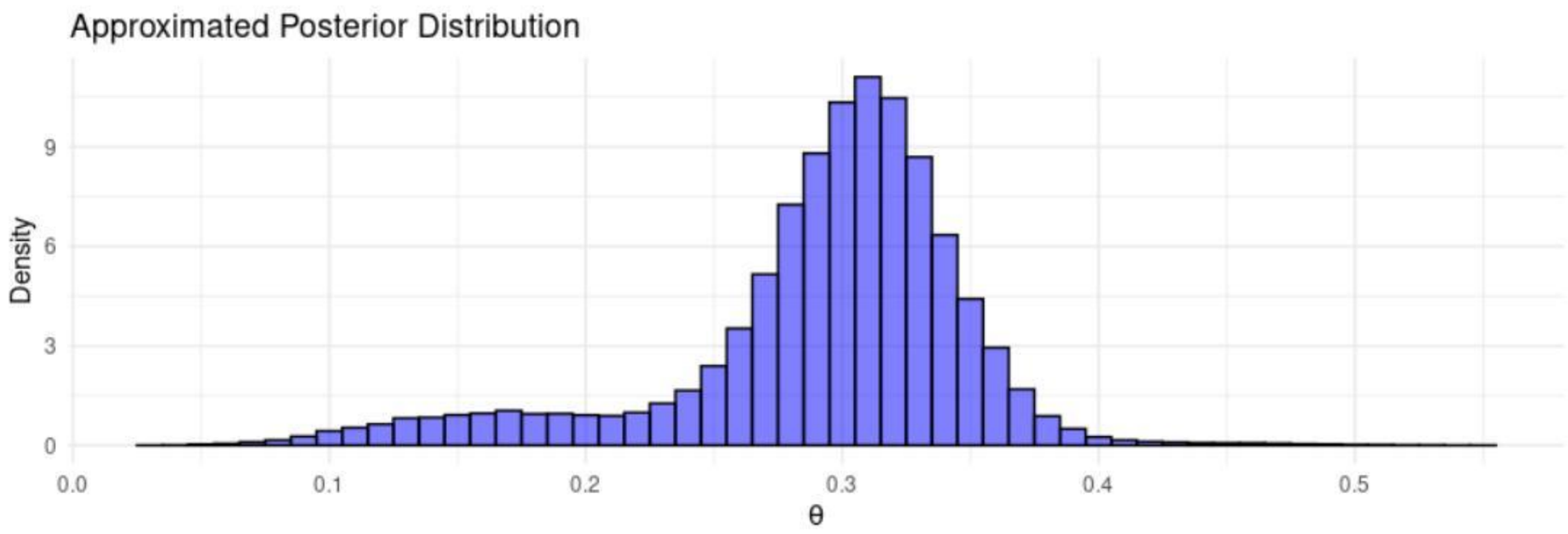


Figure 4.30: The mixed priors heuristic posterior.

```
  sum(unnormalized_posterior)

# The posterior
posterior_samples <- sample(
  samples, size = 100000,
  replace = TRUE, prob = normalized_posterior)

# The plot
ggplot(data = data.frame(posterior_samples),
       aes(x = posterior_samples)) +
  geom_histogram(binwidth = 0.01, fill = "blue",
                 alpha = 0.5, color = "black",
                 aes(y = ..density..)) +
  labs(title = "Approximated Posterior Distribution",
       x = expression(theta), y = "Density") +
  theme_minimal()
```

As opposed to the first heuristics argument plot 4.31. Given by the below R code:

```
# the weights
weights <- c(0.1, 0.5, 0.4)

alpha_params <- c(2, 40, 50)
beta_params <- c(6, 80, 50)

# Define the data
y <- 5; n <- 30
```

```r
theta_vals <- seq(0, 1, length.out = 1000)

posterior_density <- numeric(length(theta_vals))

for (i in 1:length(weights)) {
  posterior_density <- posterior_density + weights[i] *
    dbeta(theta_vals, alpha_params[i] +
            y, beta_params[i] + n - y)}

posterior_data <- data.frame(
  theta = theta_vals,
  density = posterior_density)

# The plot
ggplot(data = posterior_data, aes(x = theta, y = density)) +
  geom_line(color = "blue", size = 1) +
labs(title = "Posterior Distribution",
     x = expression(theta), y = "Density") +
theme_minimal()
```

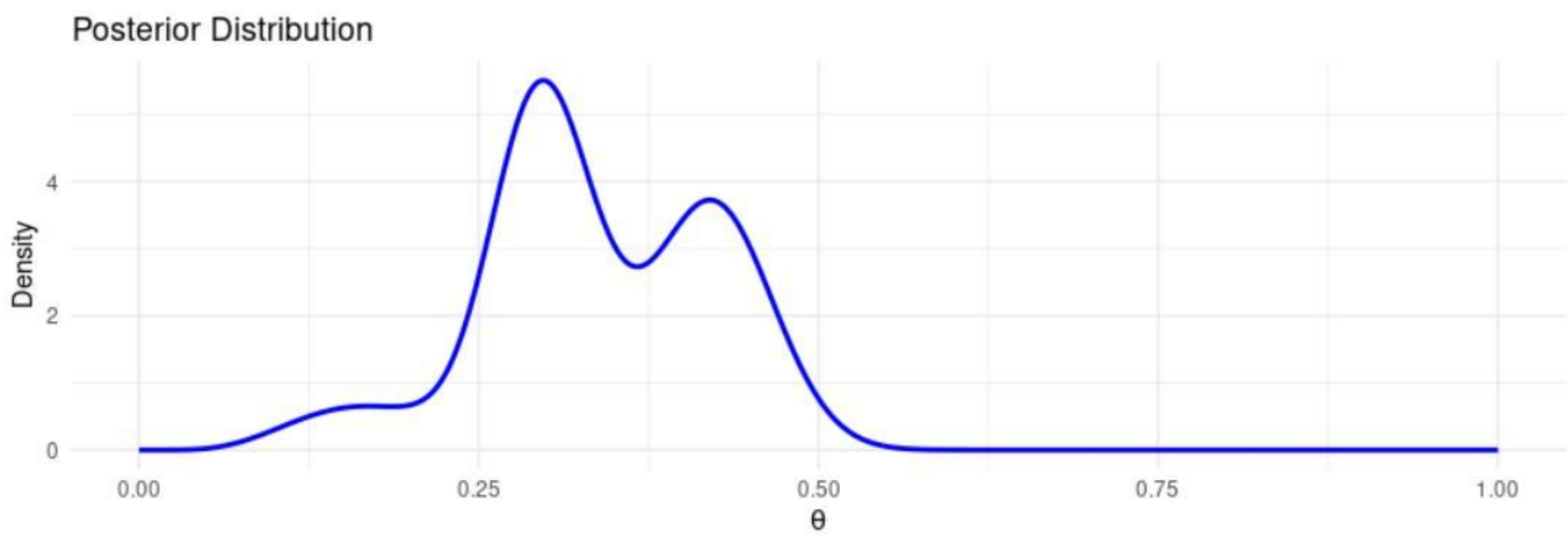


Figure 4.31: The mixed priors original huiristic posterior aproach.

As we can see from the plots 4.30 and 4.31 these different heuristic approaches provide different results and in conclusion, I would like to do some calculations as to what the posterior of the mixture prior might look like and see which heuristic approach better describes it.

$$p(\theta|y) \propto p(y|\theta)p(\theta) \Rightarrow$$

$$p(\theta|y) \propto \theta^y(1-\theta)^{n-y}(w_1 \cdot \text{Beta}(\alpha_1, \beta_1) + w_2 \cdot \text{Beta}(\alpha_2, \beta_2) + w_3 \cdot \text{Beta}(\alpha_3, \beta_3)) \Rightarrow$$

$$p(\theta|y) \propto \theta^5(1-\theta)^{25}(0.1 \cdot \frac{\Gamma(8)}{\Gamma(2)\cdot\Gamma(6)} \cdot \theta(1-\theta)^5 + 0.5 \cdot \frac{\Gamma(120)}{\Gamma(40)\cdot\Gamma(80)} \cdot \theta^{39}(1-\theta)^{79}$$

$$+ 0.5 \cdot \frac{\Gamma(100)}{\Gamma(50)\cdot\Gamma(50)} \cdot \theta^{49}(1-\theta)^{49}) \Rightarrow$$

$$p(\theta|y) \propto a \cdot \theta^6(1-\theta)^{30} + b \cdot \theta^{44}(1-\theta)^{104} + c \cdot \theta^{54}(1-\theta)^{74} \Rightarrow$$

$$p(\theta|y) \propto a_{new} \cdot \text{Beta}(7, 31) + b_{new} \cdot \text{Beta}(45, 105) + c_{new} \cdot \text{Beta}(55, 75)$$

We can do the calculations for a,b,c in R and also create there the $a_{new}, b_{new}, c_{new}$ by dividing with their sum, thus creating our sampled estimation of the posterior, but also the analytical.

```
# The weights
a <- factorial(7)/(factorial(5)*10)
b <- 5*factorial(119)/(factorial(39)*factorial(79)*10)
c <- 4*factorial(99)/(factorial(49)*factorial(49)*10)

# Normalize the weights
total_weight <- a + b + c
a_new <- a / total_weight
b_new <- b / total_weight
c_new <- c / total_weight

set.seed(123)
posterior <- a_new * rbeta(50000, 7, 31) +
  b_new * rbeta(50000, 45, 105) +
  c_new * rbeta(50000, 55, 75)

# The histogram of the posterior
ggplot(data = data.frame(posterior), aes(x = posterior)) +
  geom_histogram(binwidth = 0.01, fill = "blue",
                 alpha = 0.5, color = "black",
                 aes(y = ..density..)) +
```

```
# Add lines for the density functions
stat_function(fun = function(x) dbeta(x, 7, 31) * a_new,
              color = "red") +
stat_function(fun = function(x) dbeta(x, 45, 105) * b_new,
              color = "green") +
stat_function(fun = function(x) dbeta(x, 55, 75) * c_new,
              color = "purple") +

# Add a line for the analytical posterior
stat_function(
  fun = function(x) (dbeta(
    x,7,31) * a_new + dbeta(
      x,45,105) * b_new + dbeta(
        x,55,75) * c_new),
              color = "lightblue") +

labs(title = "Posterior Distribution", x = expression(theta),
     y = "Density") +
  theme_minimal()
```

As we can see from the plot 4.32 the posterior does not reflect our earlier heuristic approaches though it is closer to 4.30 than 4.31 and the reason is that in the first heuristic approach we sampled from our prior and then created our posterior by multiplying the likelihood with the priors according to their weights, thought taking 100000 samples from the 150000 might be the reason of this difference between them cause the probability of sampling form the first's and third's corresponding posterior is far less than of the middle one so our model failed to consider that. In conclusion, it seems the result could have been the same if we had just used beta(40,80) because its constant had at least 100 times more weight than the other beta constants.

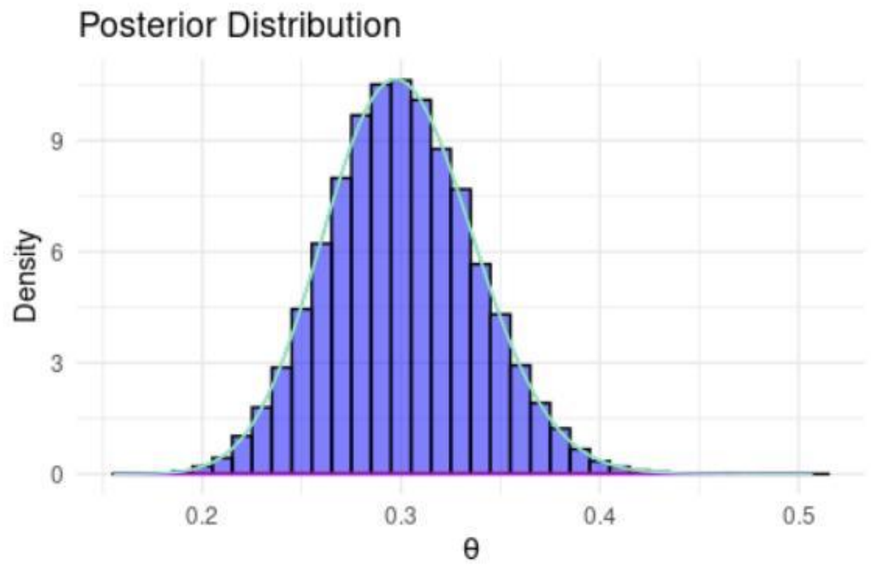


Figure 4.32: The mixed priors analytical post.

Suppose the prior distribution is a mixture of beta(a_j, b_j) priors with w_j weights, $j = 1, \ldots, m$ and the model is binomial. Additionally assume that $p(\theta|y) \propto$ model(binomial) $\times \sum_{j=1}^{m} w_j \cdot beta(a_j, b_j)$. Then:

$$p(\theta|y) \propto \sum_{j=1}^{m} w_j \cdot c_j \cdot \theta^{a_j+y-1}(1-\theta)^{b_j+n-y-1} \Rightarrow \tag{4.46}$$

$$p(\theta|y) \propto \sum_{j=1}^{m} \frac{w_j \cdot c_j}{K_j} \cdot \frac{\Gamma(a_j+b_j+n)}{\Gamma(a_j+y)\cdot\Gamma(b_j+n-y)} \theta^{a_j+y-1}(1-\theta)^{b_j+n-y-1} \tag{4.47}$$

Thus, the posterior is $\propto$ to the linear combinations of the beta$(a_j + y, b_j + n - y)$ with weights $\frac{w_j \cdot c_j}{K_j}$, $c_j = \frac{\Gamma(a_j+b_j)}{\Gamma(a_j)\cdot\Gamma(b_j)}$ and $K_j = \frac{\Gamma(a_j+b_j+n)}{\Gamma(a_j+y)\cdot\Gamma(b_j+n-y)}$
The corrected R code based on this is:

```
w_j=c(0.1,0.5,0.4)
c_j=c(factorial(7)/(factorial(1)*factorial(5)),
      factorial(119)/(factorial(39)*factorial(79)),
      factorial(99)/(factorial(49)*factorial(49)))
K_j=c(factorial(37)/(factorial(6)*factorial(30)),
      factorial(149)/(factorial(44)*factorial(104)),
      factorial(129)/(factorial(54)*factorial(74)))

a<-w_j[1]*c_j[1]/K_j[1]
b<-w_j[2]*c_j[2]/K_j[2]
c<-w_j[3]*c_j[3]/K_j[3]

total_weight <- a + b + c
a_new <- a / total_weight
b_new <- b / total_weight
```

```r
c_new <- c / total_weight

set.seed(123)
sample_size <- 500000

component_indices <- sample(1:3,
                            size = sample_size,
                            replace = TRUE,
                            prob = c(a_new, b_new, c_new))

posterior <- c(rbeta(sum(component_indices == 1), 7, 31),
               rbeta(sum(component_indices == 2), 45, 105),
               rbeta(sum(component_indices == 3), 55, 75))

# Histogram
ggplot(data = data.frame(posterior), aes(x = posterior)) +
  geom_histogram(binwidth = 0.01, fill = "blue",
                 alpha = 0.5, color = "black",
                 aes(y = ..density..)) +

# Add lines for the density functions of the mixture
stat_function(fun = function(x) dbeta(x, 7, 31) * a_new,
              color = "red") +
stat_function(fun = function(x) dbeta(x, 45, 105) * b_new,
              color = "green") +
stat_function(fun = function(x) dbeta(x, 55, 75) * c_new,
              color = "purple") +

# Add a line for the combined analytical posterior
  stat_function(
    fun = function(x) (dbeta(
      x,7,31) * a_new + dbeta(
        x,45,105) * b_new + dbeta(
          x,55,75) * c_new),
    color = "lightblue") +
```

```
labs(title = "Posterior Distribution", x = expression(theta),
     y = "Density") +
theme_minimal()
```

This code actually makes sense and the results are fairly similar to our first heuristic approach in the changes of the trend, not their exact values. The correct posterior for my mixture prior is represented by the plot 4.33. A significant finding is that the weight of the un-supportive mother goes up in the corresponding posterior as opposed to the supportive mother's, which significantly dropped. As we previously discussed this is due to the nature of beta distribution constants and is probably relative to the uncertainty of that beta, this is just a conjecture but given our beta, the ones with the most uncertainty thrived as the certain ones plummeted.

Last note/question For additional clarity I have uploaded 2 photos that describe the two different approaches (see 4.34)

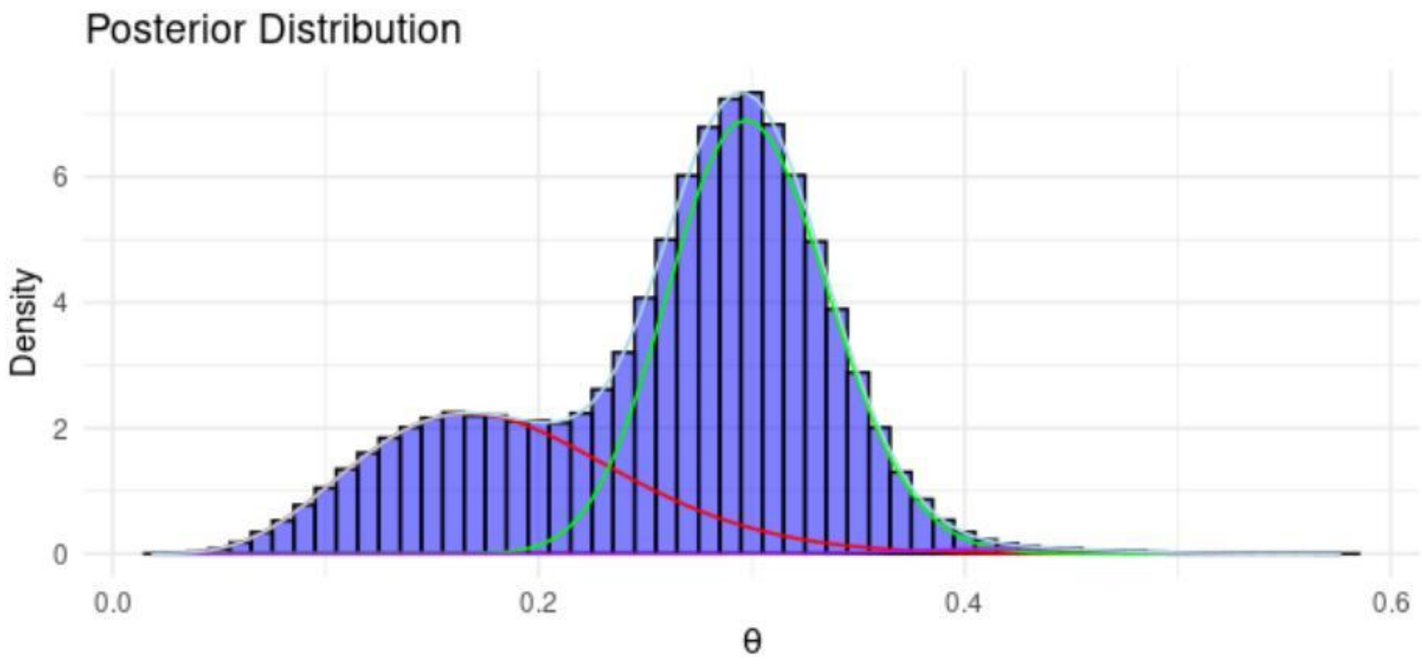
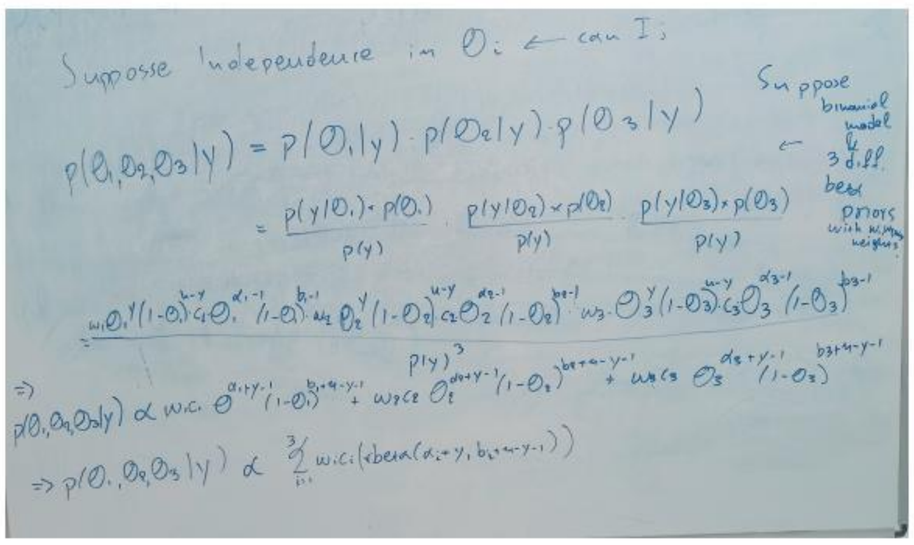
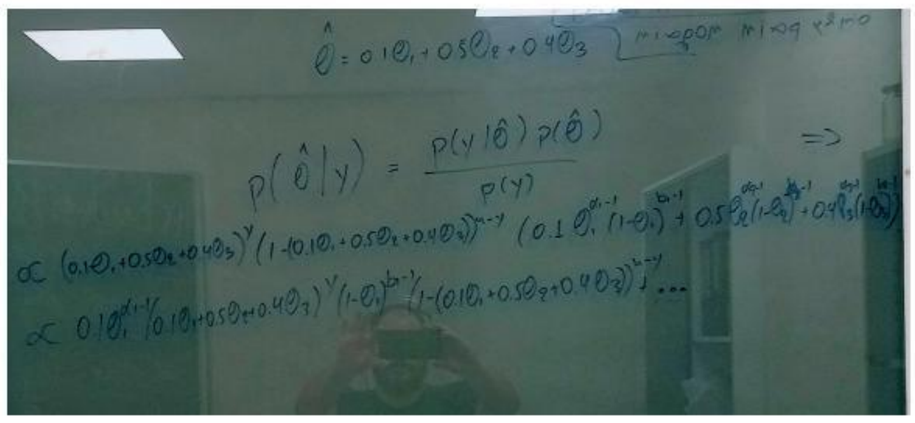


Figure 4.33: The true posterior of the mixture prior.



(a) In the white board.



(b) Everything can be a board.

Figure 4.34: Comparison of the two boards.