Assignment 3

Course: Bayesian Statistics

Emily Koskinen, 14379791

2024

- Please answer all assignment questions.
- Upload the answers as an `R` or `Rmd` document to canvas. Please additionally upload the HTML or PDF output (dependent on whether you "knit" it to PDF or HTML).

- **Deadline**: Monday, June 24, 9:00 AM.

This assignment is about the marginal likelihood and Bayesian updating.

1. Should Mathilde be added to the National Soccer Team?

Let's use Mathilde's sensational soccer (football?) abilities to learn about the marginal likelihood. Say we want to model Mathilde's performance in shooting twelve penalty shots. By now, hopefully everyone would choose a binomial model for this:

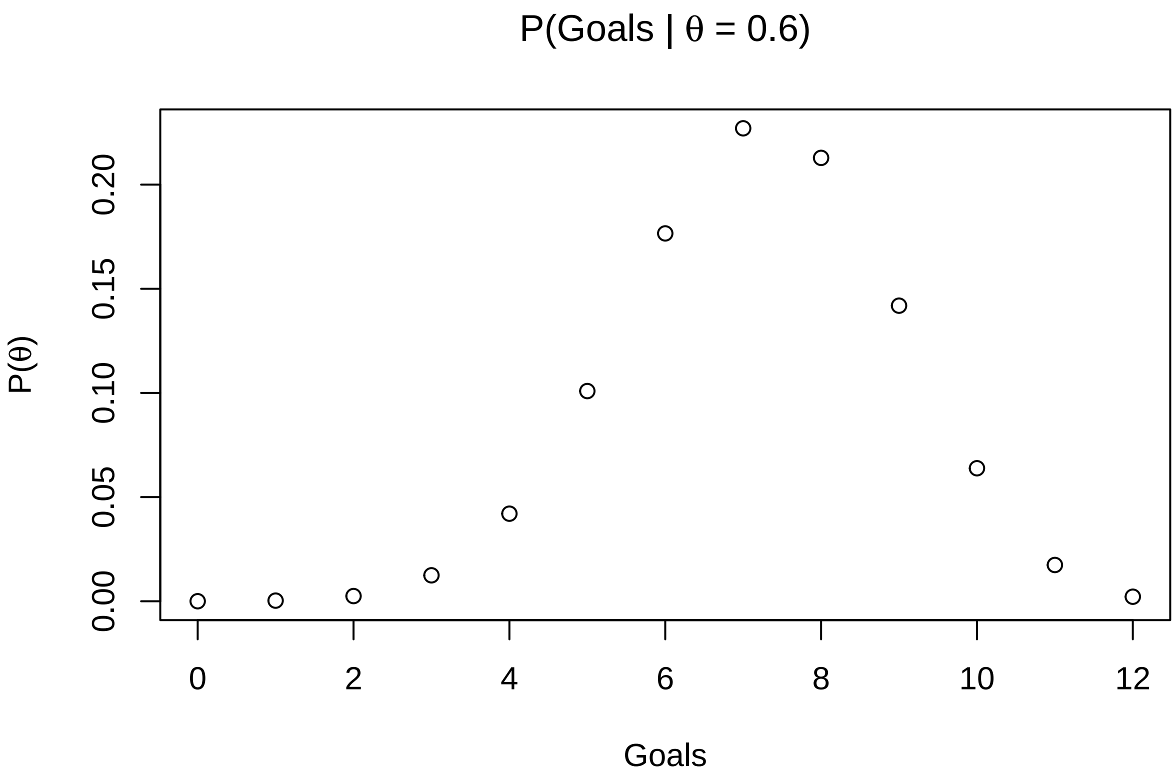
$$Y \sim \text{Binomial}(\theta, 12),$$

where Y are the number of goals out of $N = 12$ that Mathilde makes, and θ is the probability of making any one goal.

a. Suppose $\theta = 0.6$. Plot $P(Y \mid \theta = 0.6)$. (0.5pts)

```r
n <- 12
theta <- 0.6
Y <- 0:n

# Plot the likelihood
probabilities <- dbinom(Y, n, theta)
```

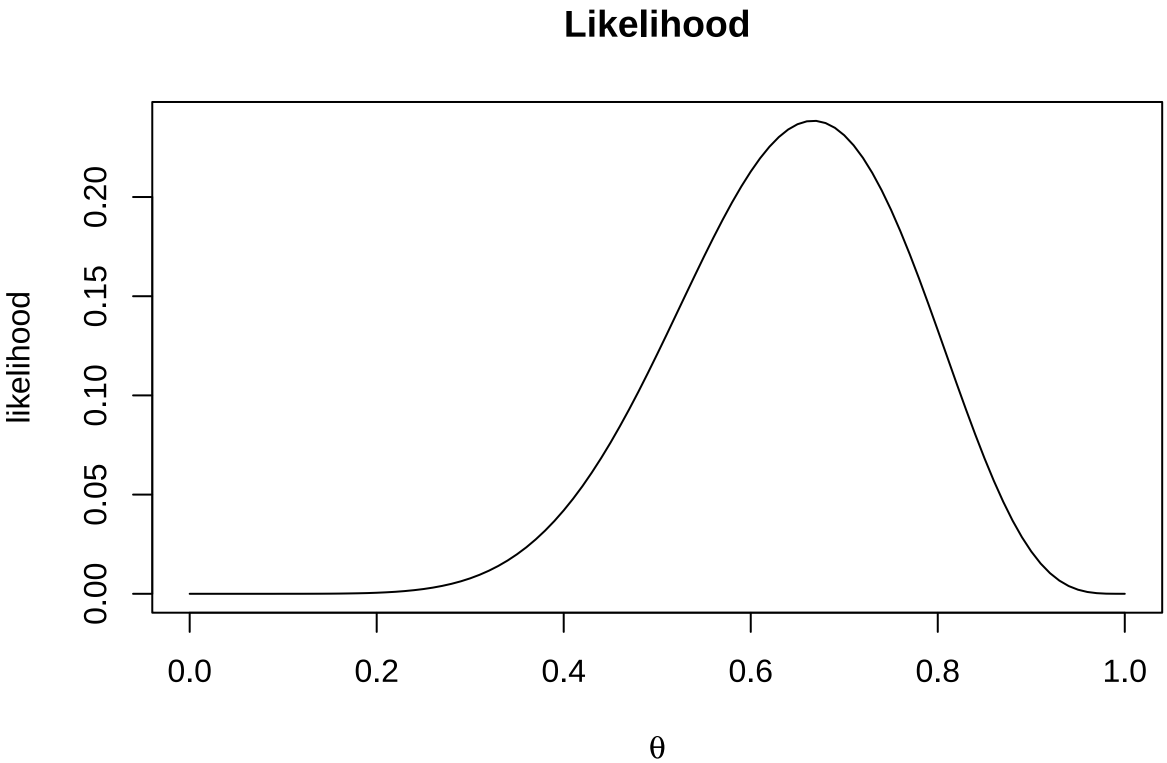
```r
plot(
  Y,
  probabilities,
  ylab = expression(paste("P(", theta, ")")),
  xlab = "Goals",
  main = expression(paste("P(Goals | ", theta, " = 0.6)")
  )
)
```



b. Suppose $Y = 8$. Plot the likelihood of θ for $Y = 8$. (0.5pts)

```r
Y <- 8
theta <- seq(0, 1, 0.01)

# Likelihood function
likelihood <- dbinom(Y, n, theta)

plot(
  theta,
  likelihood,
  xlab = expression(theta),
```

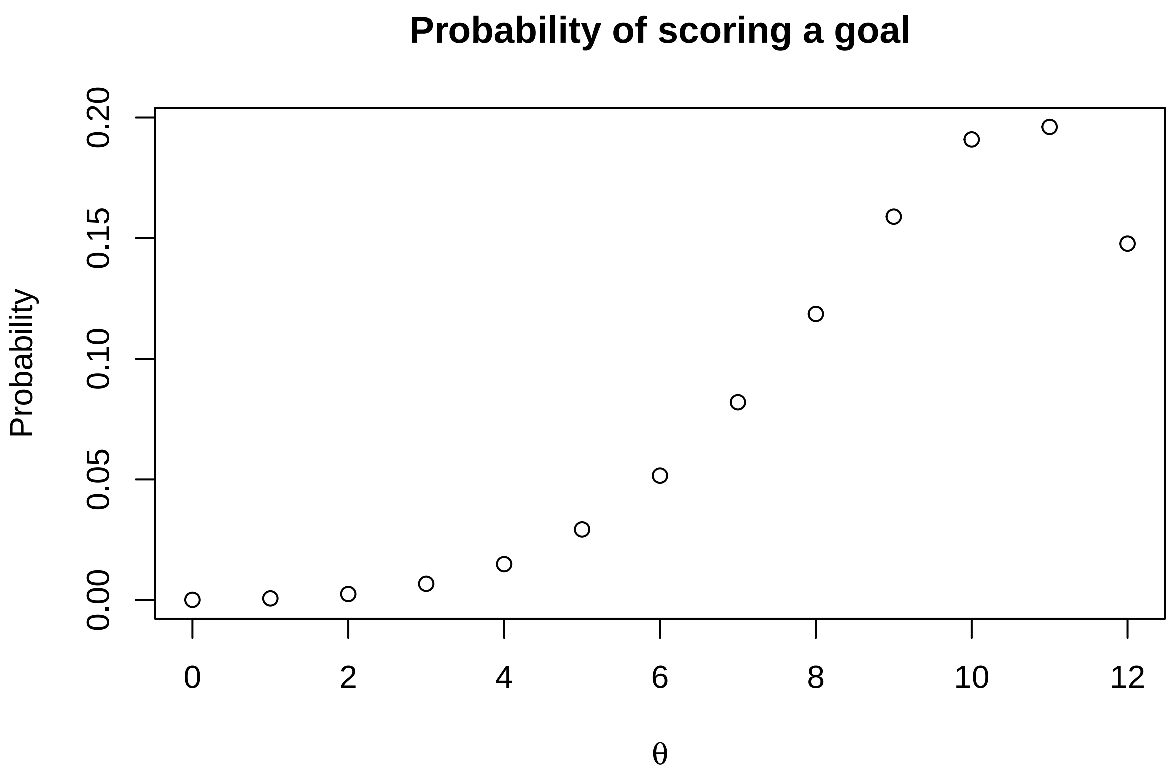
```r
  ylab = "likelihood",
  type = "l",
  main = "Likelihood"
)
```



c. Suppose we are somewhat optimistic about Mathilde's performance and choose a prior of $\theta \sim \text{Beta}(7, 2)$ for the probability of scoring a goal. Plot the implied predictions for all possible counts of made goals from this prior. Plot these predictions in the same manner as your plot from 1a. (1pts)

```r
# Prior parameters
alpha <- 7
beta <- 2

n <- 12

# sampling method
out <- matrix(0, nrow = 1e6, ncol = 13)

for (sample in 1:1e6) {
  theta = rbeta(n = 1,
                shape1 = alpha,
```

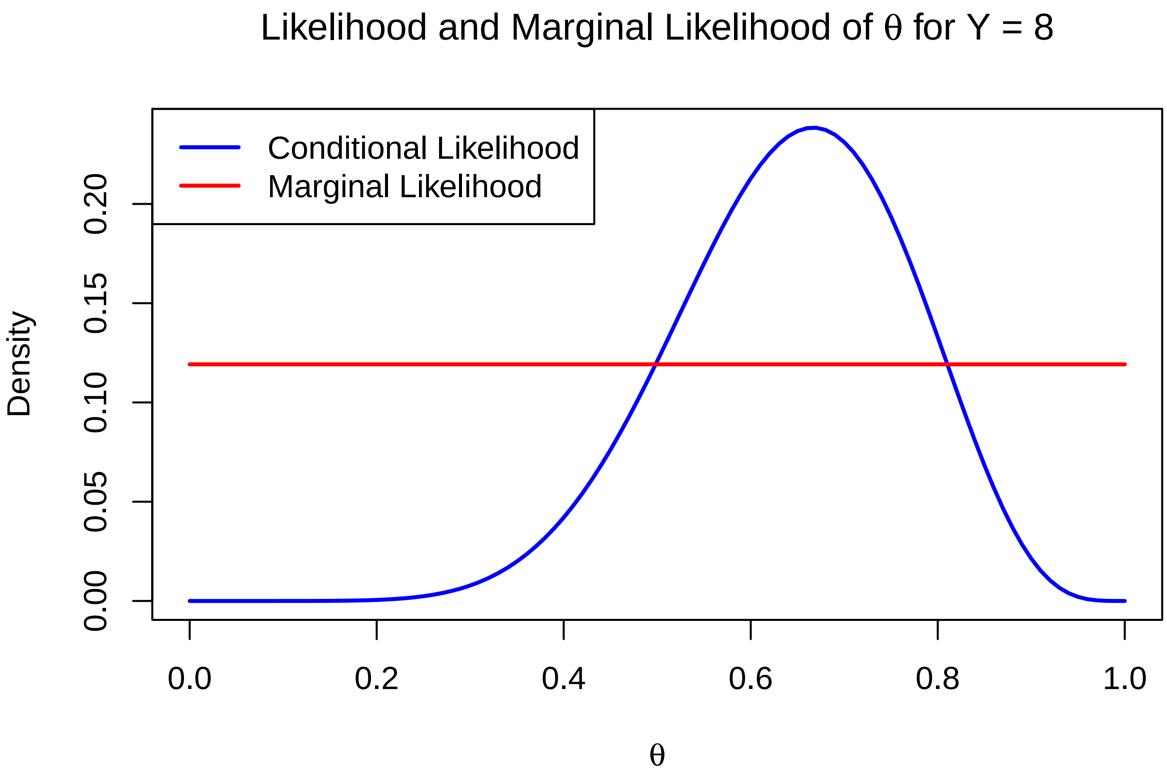
```r
                shape2 = beta)
  x = rbinom(n = 1, size = n, prob = theta)
  out[sample, ]= tabulate(x + 1, nbins = 13)
}

probability <- colMeans(out)
plot(
  0:12,
  probability,
  xlab = expression(theta),
  ylab = "Probability",
  main = "Probability of scoring a goal"
)
```



d. Remember from question 1b that $Y = 8$. Add the marginal likelihood for these data as a curve to the likelihood plot from 1b. Make sure to label conditional and marginal likelihood in the plot. Which values of θ are more likely than average and which are less likely than average for $Y = 8$? (1pt)

```r
Y <- 8
alpha <- 7
```

```r
beta <- 2
theta <- seq(0, 1, 0.01)

# Calculate the marginal likelihood
f <- function(theta) {
  dbinom(8, size = 12, prob = theta) * dbeta(theta, shape1 = alpha, shape2 = beta)
}

integral <- integrate(f, lower = 0, upper = 1)

marginal_likelihood <- rep(integral$value, 101)

# Plot the likelihood
plot(
  theta,
  likelihood,
  type = 'l',
  col = 'blue',
  lwd = 2,
  xlab = expression(theta),
  ylab = 'Density',
  main = expression(paste("Likelihood and Marginal Likelihood of ", theta, " for Y = 8")
)

# Add the marginal likelihood to the plot
lines(theta,
      marginal_likelihood,
      col = 'red',
      lwd = 2)

legend(
  'topleft',
  legend = c('Conditional Likelihood', 'Marginal Likelihood'),
  col = c('blue', 'red'),
  lty = 1,
  lwd = 2
)
```



The values above the line of the marginal likelihood are on average likely and the values below it are on average less likely. This is because the marginal likelihood is the average value for the conditional likelihood.

2. Quizzing ChatGPT

a. Ask ChatGPT what a confidence interval is, and how it is different from a credible interval. Provide prompt and answer (0.25pts).

Prompt:

What is a confidence interval, and how it is different from a credible interval?

Answer:

Confidence Interval

Definition: A confidence interval (CI) is a range of values, derived from sample data, that is likely to contain the true value of a population parameter with a specified confidence level (e.g., 95%).

Frequentist Interpretation:

- A 95% confidence interval means that if we repeated the sampling process many times and calculated the interval each time, approximately 95% of those intervals would contain the true parameter value.

- The true parameter value is fixed, but the interval can vary from sample to sample.
- For a given sample, we do not say that there is a 95% probability that the parameter lies within the interval; rather, we say that 95% of such constructed intervals from repeated samples would contain the parameter.

Example: Suppose we are estimating the mean height of a population. From a sample, we calculate a 95% confidence interval for the mean to be [160 cm, 170 cm]. This means that if we repeated the sampling process many times, 95% of the confidence intervals calculated would contain the true mean height.

Credible Interval

Definition: A credible interval (CrI) is a range of values within which an unknown parameter value lies with a certain probability, given the observed data and prior distribution.

Bayesian Interpretation:

- A 95% credible interval means that there is a 95% probability that the true parameter value lies within this interval, given the observed data and prior distribution.
- Both the parameter and the interval are considered random variables in the Bayesian framework.
- The interval reflects the updated beliefs about the parameter after considering both the data and prior information.

Example: Using Bayesian methods, we estimate the mean height of a population. Incorporating prior beliefs and observed data, we compute the posterior distribution of the mean. A 95% credible interval for the mean might be [161 cm, 169 cm]. This means that, given our data and prior information, there is a 95% probability that the true mean height is between 161 cm and 169 cm.

Key Differences

1. **Statistical Paradigm:**
    - **Confidence Interval:** Frequentist.
    - **Credible Interval:** Bayesian.
2. **Interpretation:**
    - **Confidence Interval:** Refers to the long-run frequency of the interval containing the parameter across repeated samples.

- **Credible Interval:** Refers to the probability of the parameter lying within the interval given the data and prior information.

3. **Probability Statement:**
   - **Confidence Interval:** Does not provide a probability statement about the parameter for a given interval. Instead, it provides a statement about the intervals over repeated sampling.
   - **Credible Interval:** Provides a direct probability statement about the parameter for a given interval.

4. **Dependence on Prior:**
   - **Confidence Interval:** Does not incorporate prior information; relies solely on the sample data.
   - **Credible Interval:** Incorporates prior information through the prior distribution, combined with the likelihood of the observed data.

Summary

- **Confidence Interval:** A range derived from data that would contain the true parameter value in a certain percentage of repeated samples. It is a frequentist concept and does not make a direct probability statement about the parameter for a given sample.
- **Credible Interval:** A range within which the true parameter value lies with a certain probability, given the data and prior beliefs. It is a Bayesian concept and makes a direct probability statement about the parameter for a given sample.

b. Do you think the ChatGPT is right? Would you change or add something? (0.5pts)

It says that a confidence interval is something that says what is the probability of the interval containing the true parameter value but this is a common misconception as it is more the definition for the credible interval. I would change the definition and say that the confidence interval captures more the uncertainty about the interval itself rather than the paramater. It means that across all the hypothetical interval we calculate, the probability of them capturing the true parameter value is 95%.

The Frequentist Interpretation part and the Example have it correct but the definition could be more specific so that it is clear that the credibel interval and confidence interval are different.

For the credible interval, I would add a something about it being a statement of confidence about the location or value of the parameter. Additionally, I would highlight in the differences that credible interval uses an actual sample in contrast to the hypothetical samples used in the thinking behind critical intervals.

3. Testing the Sorting Hat

A new school year starts at the famous Hogwarts School of Witchcraft and Wizardry (for more information, see here). There is a great hustle and bustle around the castle to finish all preparations before the new students arrive. As every year, it is the duty of arithmancy professor Septima Vector to check the accuracy of the Sorting Hat, a loquacious magical utensil that helps sorting new students into four houses: Gryffindor, Ravenclaw, Hufflepuff, and Slytherin. Usually, the sorting hat sorts students into houses based on their personality. However, to check the accuracy of the Sorting Hat, Professor Vector lets the Sorting Hat allocate two types of magical creatures to houses: Boring Flobberworms and Malicious Boggarts. As magical research has shown that Flobberworms are too boring to have any personality, they should be sorted randomly into the four houses. Malicious Boggarts should have a high probability to be sorted into the house of Slytherin.

For her test, Professor Vector lets the Sorting Hat decide on ten Flobberworms and ten Boggarts (the sample size needs to be small because the Sorting Hat complains loudly when he has to sort anything else than witches and wizards).

a. Suppose Y_i is the count of creatures sorted into the ith house ($i = 1, 2, 3, 4$ for Gryffindor, Ravenclaw, Hufflepuff, and Slytherin, respectively). What would be a good model for these counts? *Hint:* See chapter 8. (1pt)

Multinomial distribution, because the data is discrete (4 outcomes/houses). It also includes multiple trials with a fixed number of total trials that are independent. Additionally, each trial has more than two outcomes.

. . . .

b. Formulate two prior distributions using a prior that is conjugate to the likelihood: A prior distribution for the Flobberworms, and a prior distribution for the Boggarts. Justify your choices. (1pt)

I chose the values for Boggarts based on the statement that they are more likely to put into Slytherin due to their personality. This is why the alpha value for getting into Slytherin is 6 and the others are lower and equally likely.

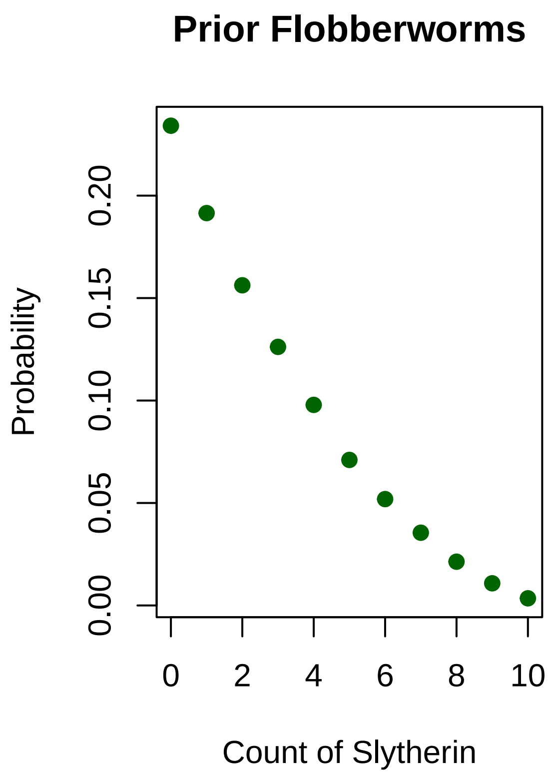
Flobberworms were said to be lazy and have equally likely outcome for each house, which is why I chose the uniform distribution for them.

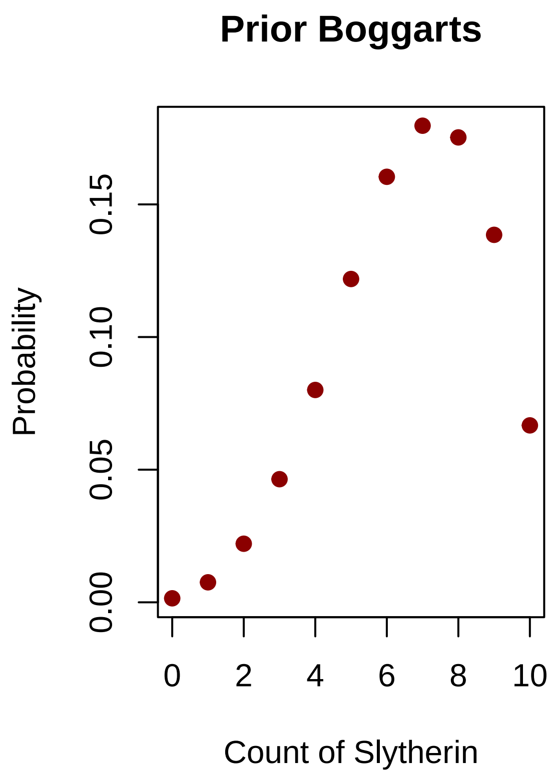
I chose the Dirichlet distribution for both of them as the prior as it is conjugate to the Multinomial likelihood.

c. Now it is time to make predictions! Use simulations to make predictions about the number of creatures to be sorted into Slytherin in a sample of 10. Plot the prior predictive distribution for the two prior distributions you formulated in 3b. Note that the x-axis should denote the potential counts for Slytherin, and the y-axis should denote the probability for each count. Use $M = 100.000$ simulation runs. (2pt)

```r
M <- 100000
n <- 10

# Sampling: flobberworms
alpha_f <- c(1, 1, 1, 1)
output_f <- matrix(0, nrow = M, ncol = 11)

for (sample in 1:M) {
  theta = rdirichlet(1, alpha_f)
  x = rmultinom(n = 1, size = n, prob = theta)
  output_f[sample, ] = tabulate(x[4, ] + 1, nbins = 11)
}

prob_f <- colMeans(output_f)

par(mfrow = c(1, 2))

plot(
  0:n,
  prob_f,
  col = "darkgreen",
  main = "Prior Flobberworms",
  xlab = "Count of Slytherin",
  ylab = "Probability",
  pch = 19
)

# Sampling: Boggarts
alpha_b <- c(1, 1, 1, 6)
output_b <- matrix(0, nrow = M, ncol = 11)
```

```r
for (sample in 1:M) {
  theta = rdirichlet(1, alpha_b)
  x = rmultinom(n = 1, size = n, prob = theta)
  output_b[sample, ] = tabulate(x[4,] + 1, nbins = 11)
}

prob_b <- colMeans(output_b)

plot(
  0:n,
  prob_b,
  col = "darkred",
  main = "Prior Boggarts",
  xlab = "Count of Slytherin",
  ylab = "Probability",
  pch = 19
)
```






```r
# Reset plotting space
par(mfrow = c(1, 1))
```

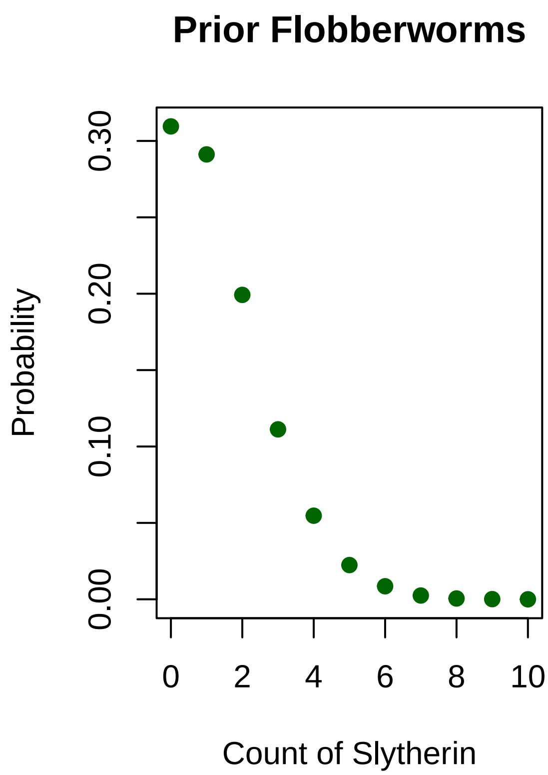
Hint: The R-package **LaplacesDemon** might be helpful.

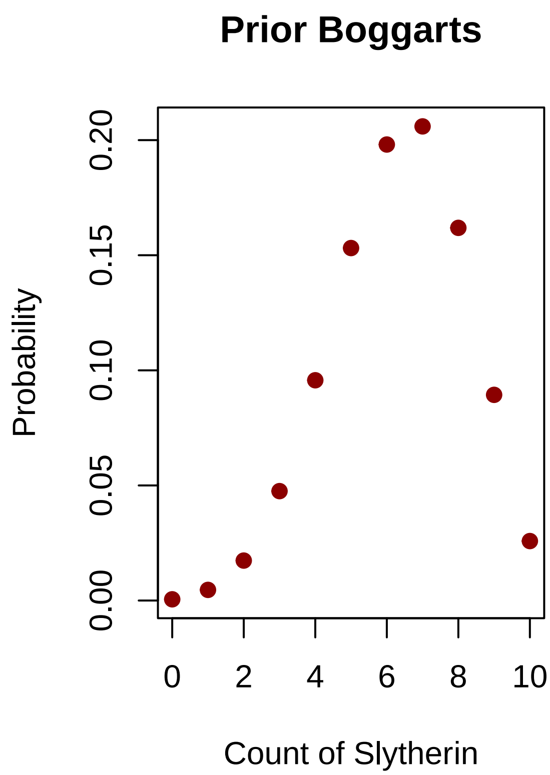
d. Interpret the two plots from the 3c. What is shown, how does it relate to your prior choices? (1pts)

The uniform prior of only 1's translates to a slanted posterior because the assignment into each group is equally likely but when these probabilities are put together, then lower values are more likely than high values for placement. This is shown by the slanted that has the count of 0 as the highest probability.

For Boggarts, I chose the alpha values that are highest for getting into Slytherin as this was described in their characteristics. The shape of the distribution for them shows that the peak is at the highest point around the count of 7 or 8, which reflects the high alpha value I chose for them.

e. Below, you can see the experimental data Professor Vector collected on Flobberworms and Boggarts. Use it to update your prior distributions to posterior distributions for these creatures! (1pt)





f. Compare your predictions from question 3c with the actual data for the house Slytherin. Would you say that your Flobberworm and Boggart models made good predictions? (1pt)

I would say that my models made fairly good predictions. The shape for both of the models is similar to my predictions and the peak has moved higher showing greater certainty for the probabilities. The most plausible probability is also almost in the same place as was for my predictions shown by 0 for Flobberworms and around 7 for Boggarts.