Computational Modeling - Assignment 2

Josephine, Mie and Tobias

10/03/2020

Github link for the code: https://github.com/tobiasfrisenborg/A2_CogSci-knowledge.git

In this assignment we learn how to assess rates from a binomial distribution, using the case of assessing your teachers' knowledge of CogSci.

First part

We want to assess your teachers' knowledge of cognitive science.

To keep things simple (your teachers should not be faced with too complicated things):

- You created a pool of equally challenging questions on CogSci
- Each question can be answered correctly or not (we don't allow partially correct answers, to make our life simpler).
- Knowledge of CogSci can be measured on a scale from 0 (negative knowledge, all answers wrong) through 0.5 (random chance) to 1 (awesome CogSci superpowers)

This is the data:

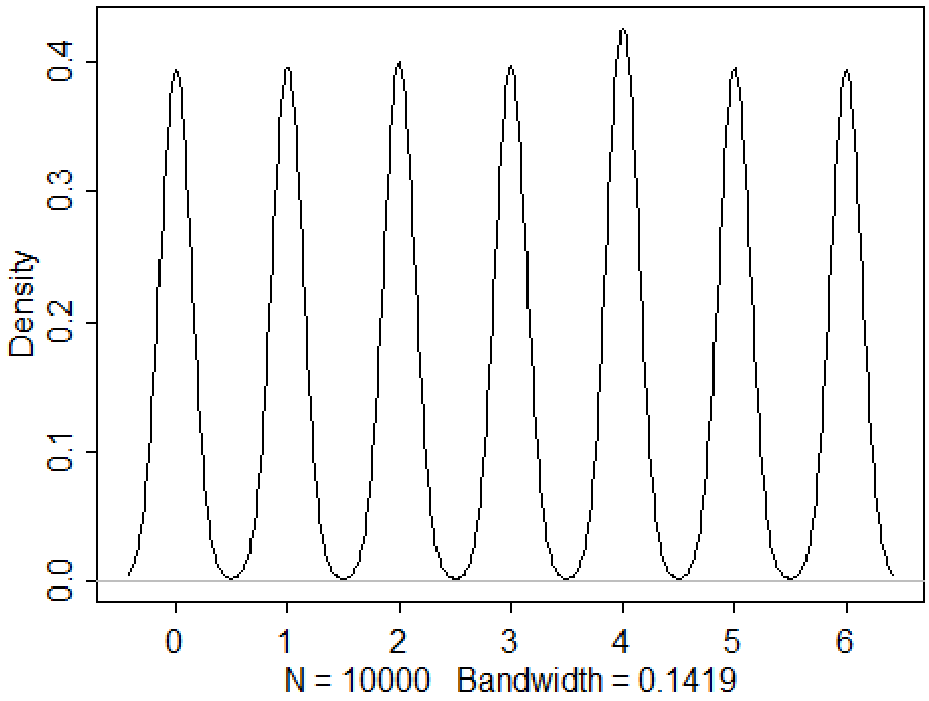
- Riccardo: 3 correct answers out of 6 questions
- Kristian: 2 correct answers out of 2 questions (then he gets bored)
- Josh: 160 correct answers out of 198 questions (Josh never gets bored)
- Mikkel: 66 correct answers out of 132 questions

Questions:

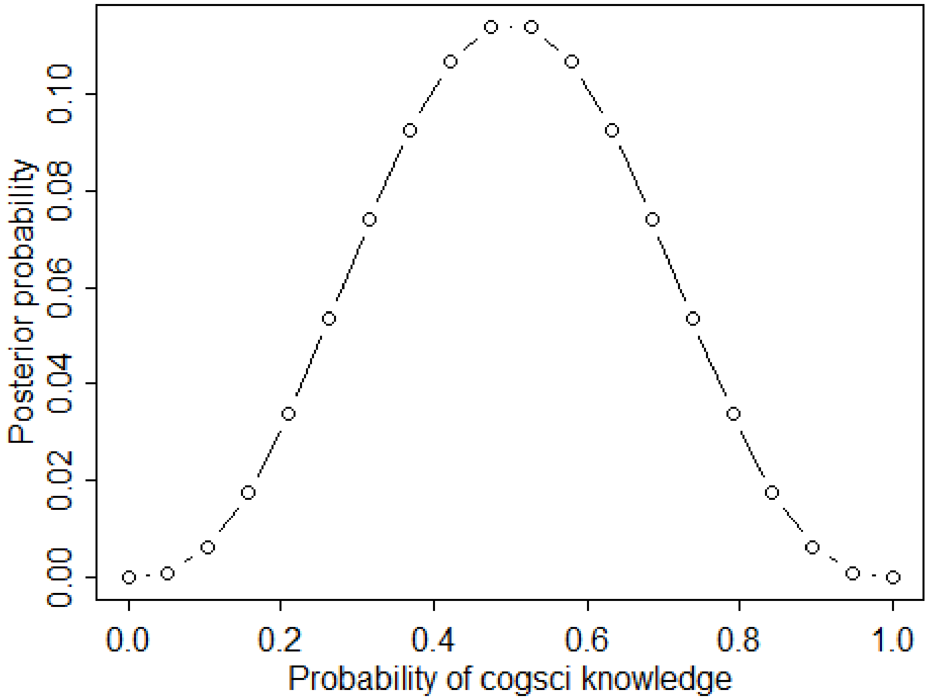
1. What's Riccardo's estimated knowledge of CogSci? What is the probability he knows more than chance (0.5)?

```
### Implement grid apporximation
dens <- 20  # define density (number of points to calculate)
p_grid <- seq(from = 0, to = 1, length.out = dens)  # define grid
```

```
prior <- rep(1, dens)  # define prior
dens(rbinom(1e4, 6, runif(1e4, 0, 1)))  # test the prior, it looks reasonable
```

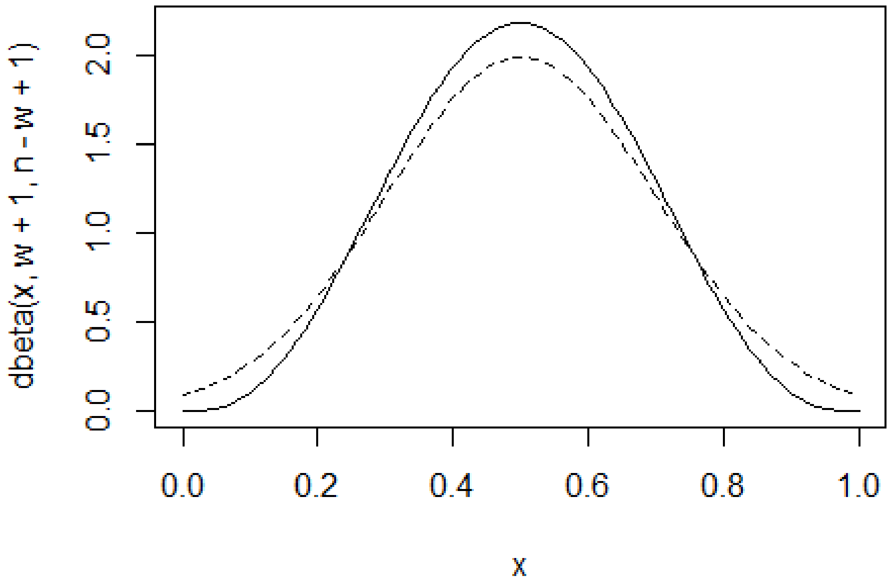


```
likelihood <- dbinom(3, size = 6, prob = p_grid)  # compute likelihood at each
value in grid
unstd.posterior <- likelihood * prior  # compute product of likelihood and prio
posterior <- unstd.posterior / sum(unstd.posterior)  # standardize the posterior,
so it sums to 1 - this is equivalent to bayes theorem

# plot the posterior probability
plot(p_grid, posterior,
     type = "b",
     xlab = "Probability of cogsci knowledge",
     ylab = "Posterior probability")
```



```
# add up posterior probability where p < 0.5, this is equivalent to calculating
the probability that Riccardo's knowledge is above chance
sum(posterior[p_grid > 0.5])

## [1] 0.5

### Implement quadratic apporximation
cogsci_qa <- rethinking::map(
  alist(
    w ~ dbinom(6,p), # binomial distribution
    p ~ dunif(0,1)),  # uniform prior
  data = list(w = 3))

# display summary of quadratic approximation
precis(cogsci_qa)

##   Mean StdDev 5.5% 94.5%
## p  0.5    0.2 0.17  0.83

# specify the results
w <- 3
n <- 6

# plot the actual distribution
curve(dbeta(x, w + 1, n - w + 1), from = 0 , to = 1)
```

```
# plot the quadratic approximation
curve(dnorm(x, 0.5, 0.2), lty = 2, add = TRUE)
```

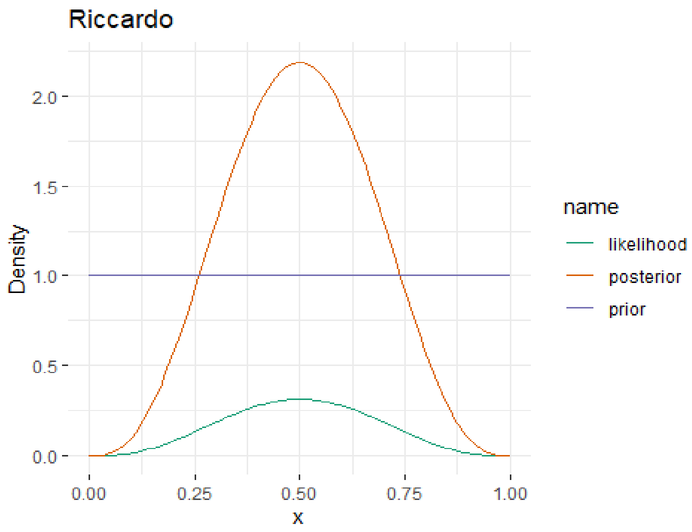


So from the plots and calculations seen above, we estimate that Riccardo has a 50% chance of having knowledge above chance (assuming that the questions are binary, and therefore chance will be at 50%). The results of the mean are the same for both quadratic approximation and the grid approximation, however, the curves are slightly different.

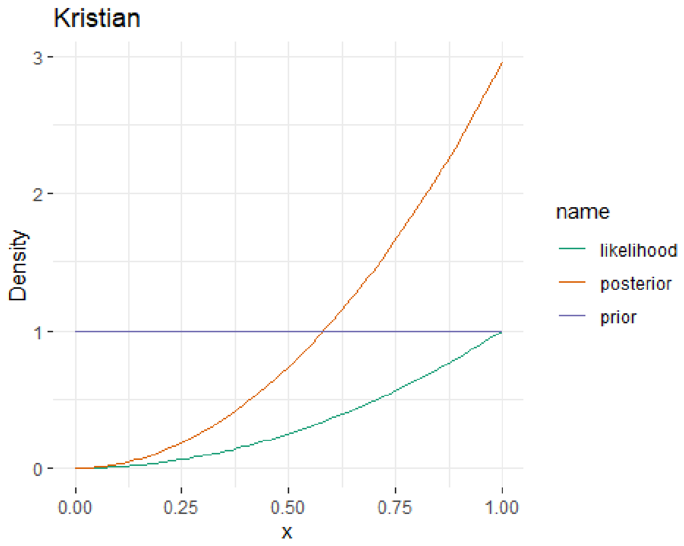
2. Estimate all the teachers' knowledge of CogSci. Who's best?

2a. Produce plots of the prior, and posterior for each teacher.

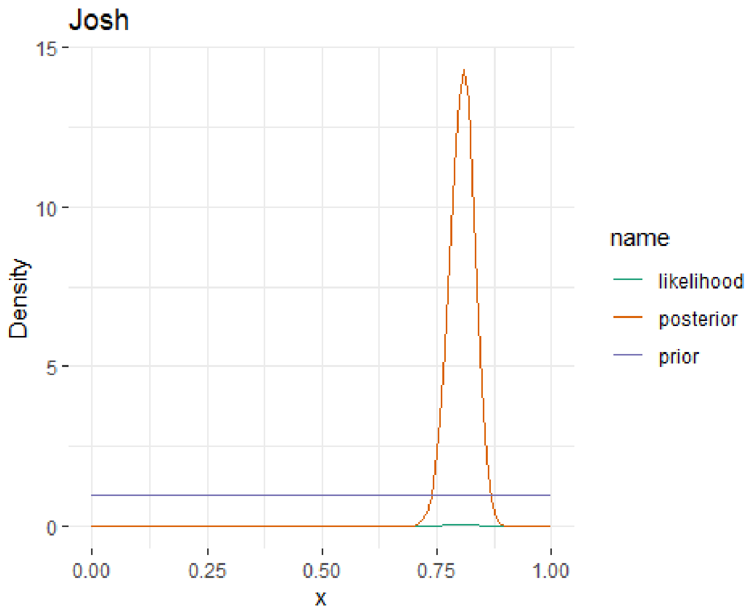
```
#Creating functions

calc_teacher <- function(n_correct, n_question, prior, length_out = 100){
  # this function calculate the posterior
  p_grid <- seq(from = 0, to = 1, length.out = length_out)
  likelihood <- dbinom(n_correct,
                       size = n_question,
                       prob = p_grid)
  unstd_posterior <- prior * likelihood
  bin_size <- abs(p_grid[1] - p_grid[2])
  posterior <- unstd_posterior/sum(unstd_posterior * bin_size)
  return(list(teacher_posterior = posterior,
              likelihood = likelihood,
              grid = p_grid))
}

pretty_plot <- function(p_grid, prior, likelihood, posterior, title = " "){
  # define data
  d <- tibble(p_grid = p_grid,
              prior = prior,
              likelihood = likelihood,
              posterior = posterior)

  # make to long format
  d <- d %>%
    pivot_longer(cols = c("prior", "likelihood", "posterior"), names_to = "name",
values_to = "value")

  # make a
  p <- ggplot(d, aes(x = p_grid, y = value, color = name)) +
    geom_line() +
    labs(x = "x", y = "Density", title = title) +
    theme_bw() +
    ggplot2::theme(panel.background = element_rect(fill = "white"),
                   panel.border = element_blank()) +
    scale_colour_brewer(palette = "Dark2", direction = 1)
  return(p)
}

subtract_plot <- function(posterior, title = " "){
  # define data
  d <- tibble(p_grid = p_grid,
              subtract_posterior = posterior)

  # make to long format
  d <- d %>%
    pivot_longer(cols = c("subtract_posterior"), names_to = "name", values_to =
"value")

  # make a
  p <- ggplot(d, aes(x = p_grid, y = value, color = name)) +
    geom_line() +
    labs(x = "x", y = "Density", title = title) +
    theme_bw() +
    ggplot2::theme(panel.background = element_rect(fill = "white"),
                   panel.border = element_blank()) +
    scale_colour_brewer(palette = "Dark2", direction = 1)
  return(p)
}

#Creating a uniform prior
uniform_prior <- rep(1, 100)

## Riccardo (3/6)
Riccardo <- calc_teacher(3,6, uniform_prior)
pretty_plot(Riccardo$grid, uniform_prior, Riccardo$likelihood,
Riccardo$teacher_posterior, title = "Riccardo")
```



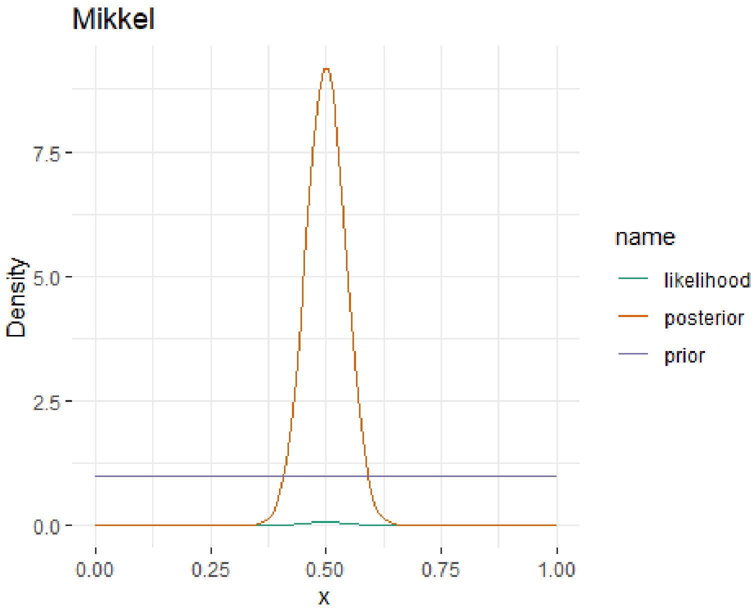
```
## Kristian (2/2)
Kristian <- calc_teacher(2,2, uniform_prior)
pretty_plot(Kristian$grid, uniform_prior, Kristian$likelihood,
Kristian$teacher_posterior, title = "Kristian")
```



```r
# Josh (160 / 198)
Josh <- calc_teacher(160,198, uniform_prior)
pretty_plot(Josh$grid, uniform_prior, Josh$likelihood, Josh$teacher_posterior,
title = "Josh")
```



```r
# Mikkel (66 / 132)
Mikkel <- calc_teacher(66,132, uniform_prior)
pretty_plot(Mikkel$grid, uniform_prior, Mikkel$likelihood,
Mikkel$teacher_posterior, title = "Mikkel")
```



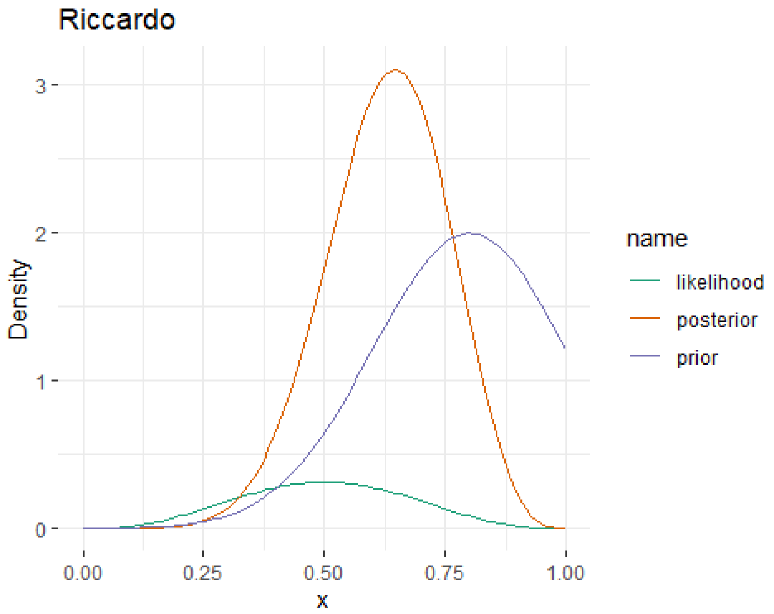
The posteriors of Riccardo and Mikkel are both centered around the mean of 0.5. However, as Mikkel have answered more questions, we are more certain that his actual knowledge of cogsci is around 0.5, which is evident from the curve of his posterior probability being more narrow than the one for Riccardo.

To determine who has greater CogSci knowledge, we can assess the plots to indentify which posterior distribution has the maximum a posteriori closest to one. If using this criteria, Kristian has the greatest knowledge of CogSci. However, we also have to take into account the certainty of the curve which we determine from eyeballing the area under the curve. Therefore, the teacher with most knowledge is Josh.

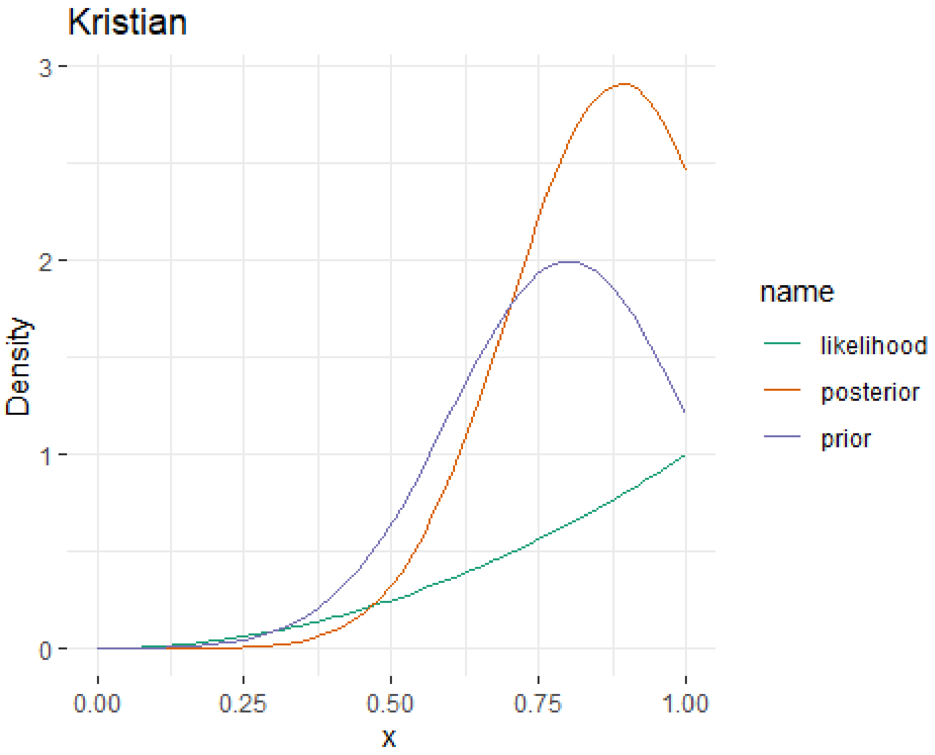
3. **Change the prior. Given your teachers have all CogSci jobs, you should start with a higher appreciation of their knowledge: the prior is a normal distribution with a mean of 0.8 and a standard deviation of 0.2. Do the results change (and if so how)?**

3a. Produce plots of the prior and posterior for each teacher.

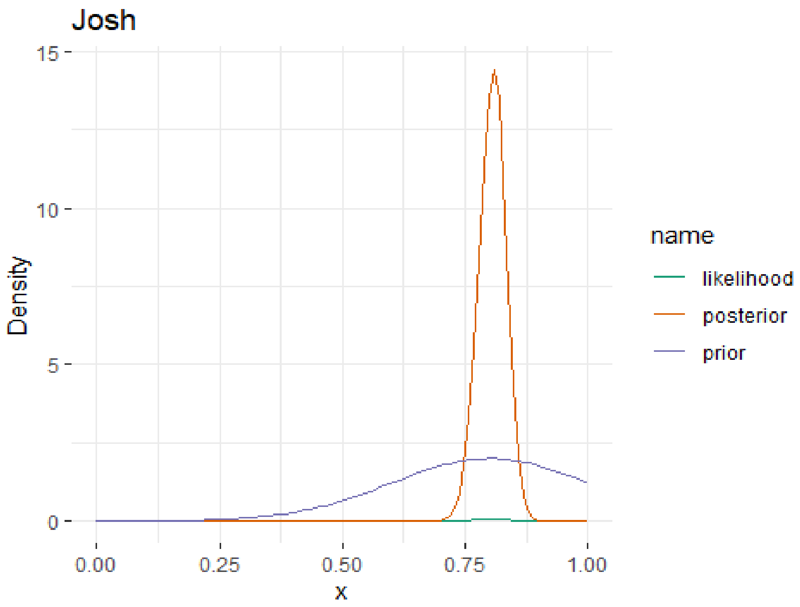
```
#creating a new prior

p_grid <- seq(from = 0, to = 1, length.out = 100)
sensibly_centered_at_0.8 <- dnorm(p_grid, 0.8, 0.2)

# Riccardo (3 / 6)
Riccardo_0.8<- calc_teacher(3,6,sensibly_centered_at_0.8)
pretty_plot(Riccardo_0.8$grid, sensibly_centered_at_0.8, Riccardo_0.8$likelihood,
Riccardo_0.8$teacher_posterior, title = "Riccardo")
```



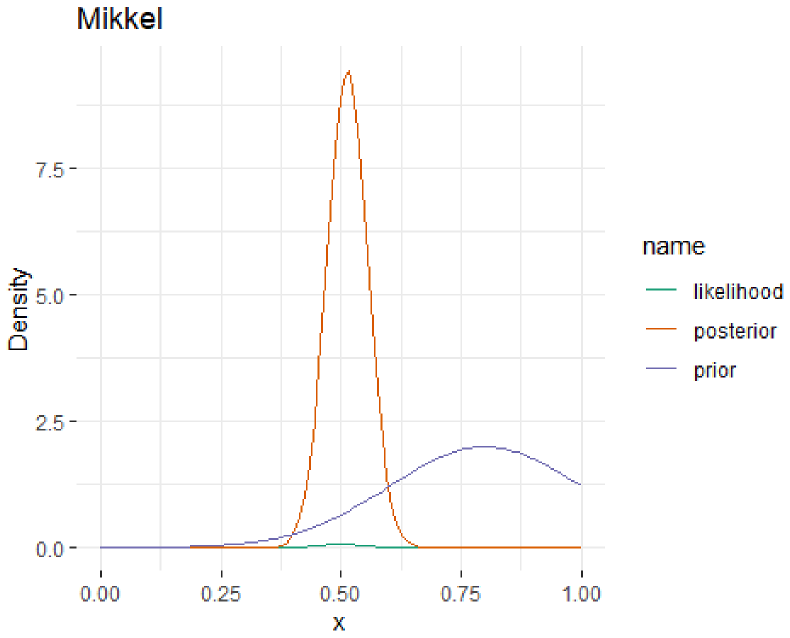
```
## Kristian (2/2)
Kristian_0.8<- calc_teacher(2,2,sensibly_centered_at_0.8)
pretty_plot(Kristian_0.8$grid, sensibly_centered_at_0.8, Kristian_0.8$likelihood,
Kristian_0.8$teacher_posterior, title = "Kristian")
```



```r
# Josh (160 / 198)
Josh_0.8<- calc_teacher(160,198,sensibly_centered_at_0.8)
pretty_plot(Josh_0.8$grid, sensibly_centered_at_0.8, Josh_0.8$likelihood,
Josh_0.8$teacher_posterior, title = "Josh")
```



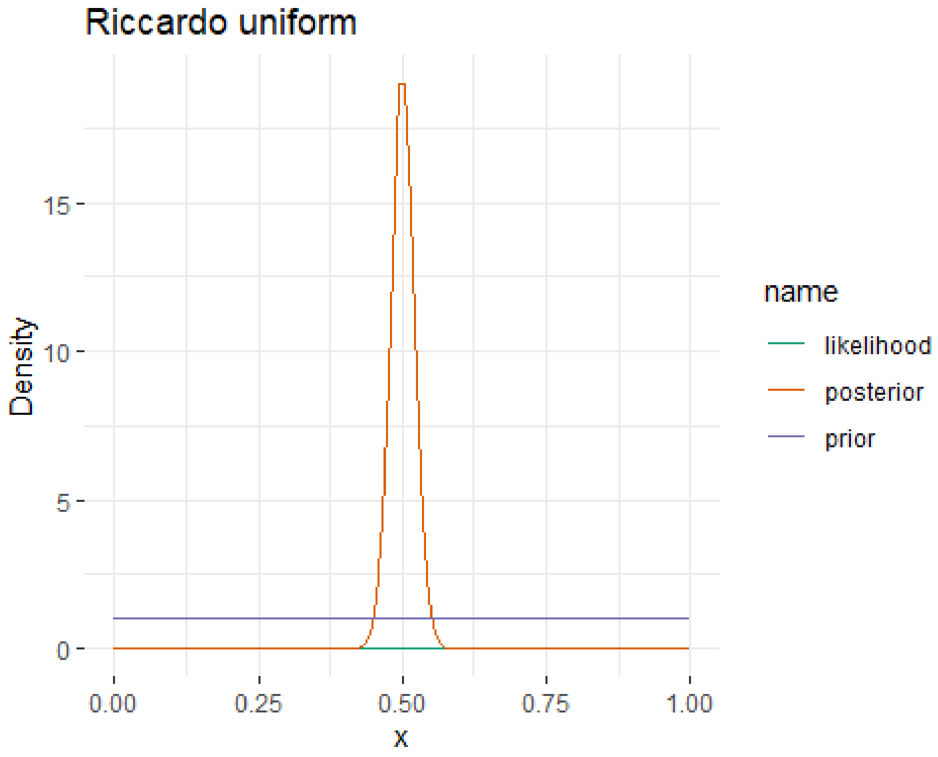
```r
# Mikkel (66 / 132)
Mikkel_0.8<- calc_teacher(66,132,sensibly_centered_at_0.8)
pretty_plot(Mikkel_0.8$grid, sensibly_centered_at_0.8, Mikkel_0.8$likelihood,
Mikkel_0.8$teacher_posterior, title = "Mikkel")
```



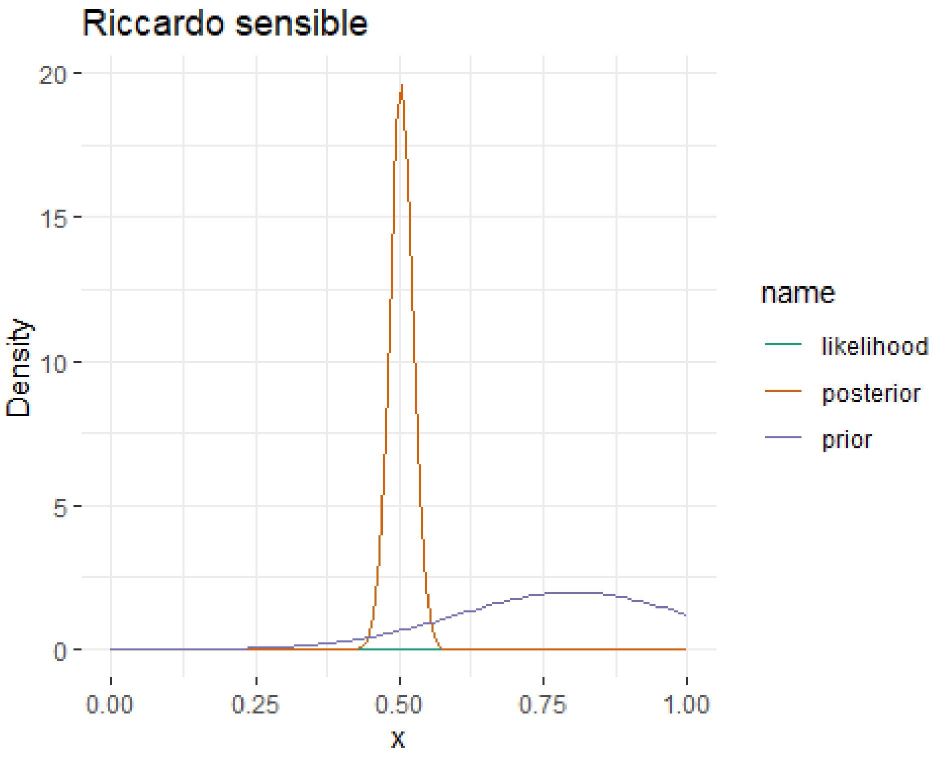
The results change for all teachers. However, the degree of change varies due to the amount of data. Previously, the probability distributions of Mikkel's and Riccardo's knowledge were similar. However, since Riccardo has fewer datapoints his posterior probability is more susceptable to change because of the prior. In this case, it increases his posterior probability to a higher degree than Mikkel. The same applies to Kristian, again due to the few datapoints. Josh's results are consistent, which is sensible since he answered more questions.

4. You go back to your teachers and collect more data (multiply the previous numbers by 100). Calculate their knowledge with both a uniform prior and a normal prior with a mean of 0.8 and a standard deviation of 0.2. Do you still see a difference between the results? Why?

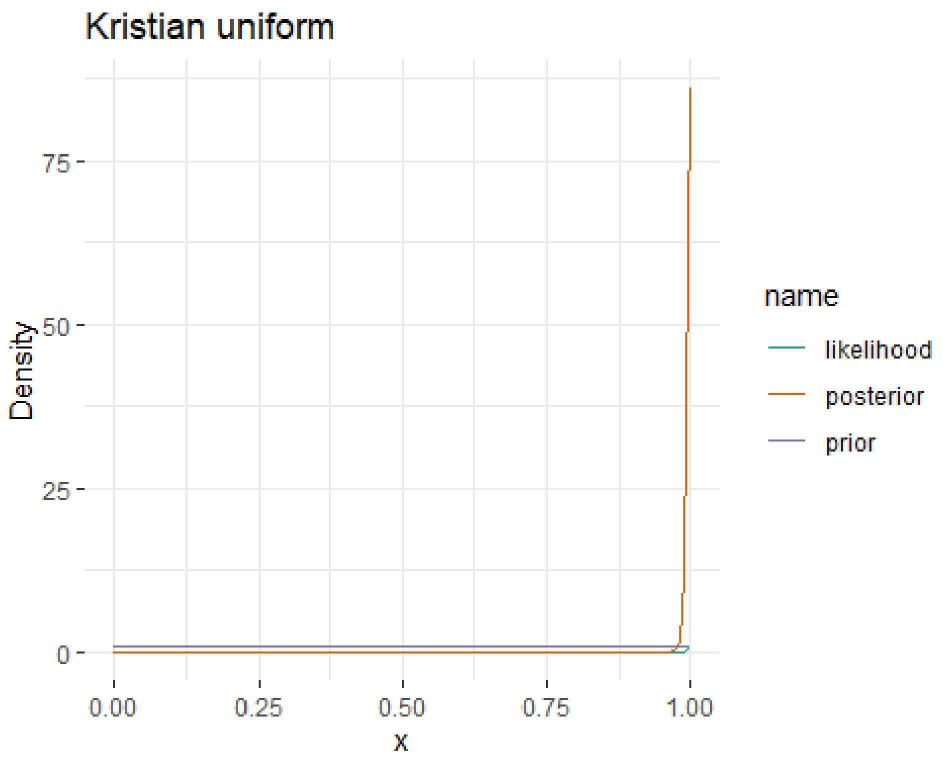
```r
# Riccardo (300 / 600)
Riccardo <- calc_teacher(300,600, uniform_prior)
pretty_plot(Riccardo$grid, uniform_prior, Riccardo$likelihood,
Riccardo$teacher_posterior, title = "Riccardo uniform")
```



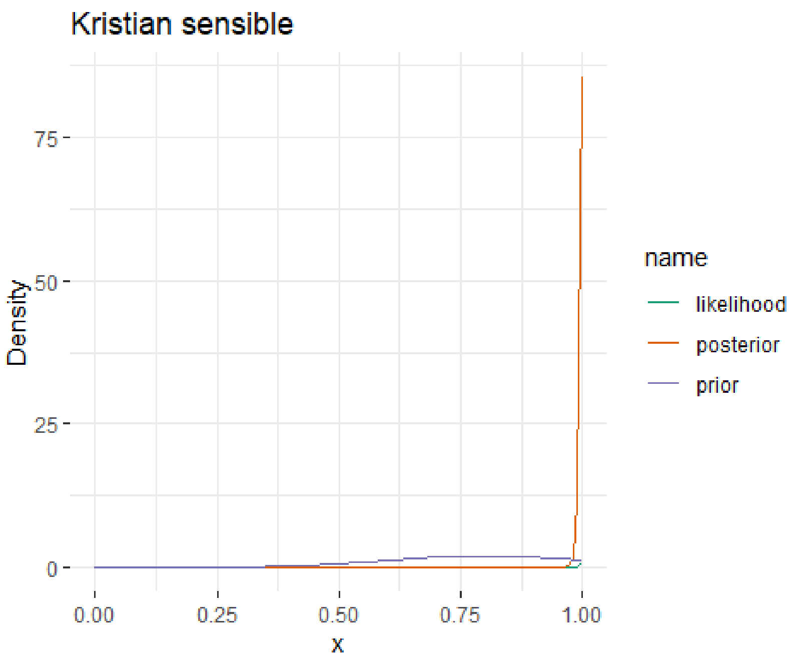
```
Riccardo_0.8<- calc_teacher(300,600,sensibly_centered_at_0.8)
pretty_plot(Riccardo_0.8$grid, sensibly_centered_at_0.8, Riccardo_0.8$likelihood,
Riccardo_0.8$teacher_posterior, title = "Riccardo sensible")
```



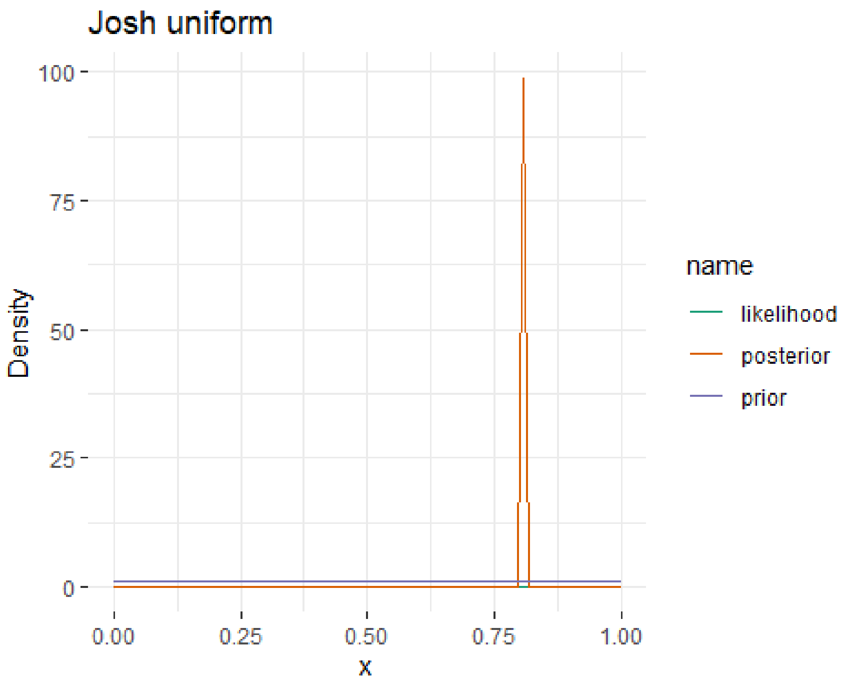
```
## Kristian (200/200)
Kristian <- calc_teacher(200,200, uniform_prior)
pretty_plot(Kristian$grid, uniform_prior, Kristian$likelihood,
Kristian$teacher_posterior, title = "Kristian uniform")
```



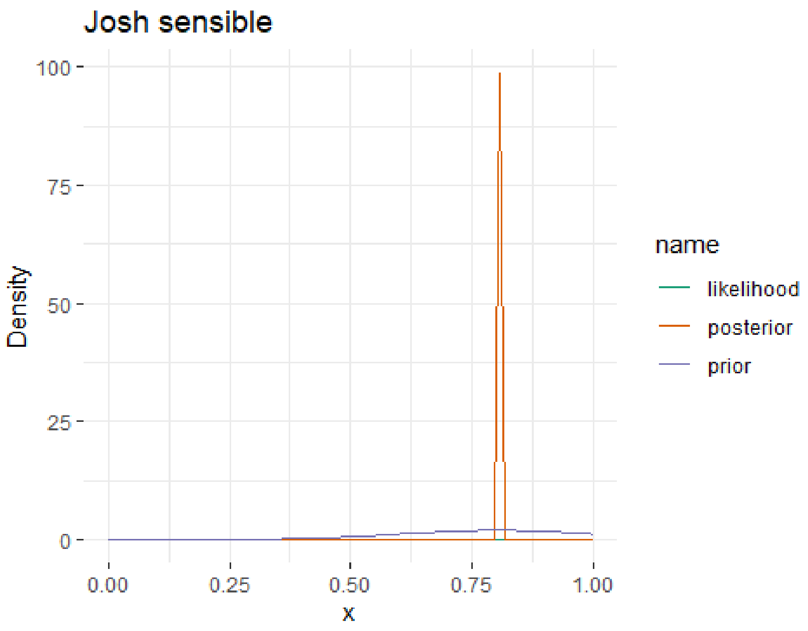
```
Kristian_0.8<- calc_teacher(200,200,sensibly_centered_at_0.8)
pretty_plot(Kristian_0.8$grid, sensibly_centered_at_0.8, Kristian_0.8$likelihood,
Kristian_0.8$teacher_posterior, title = "Kristian sensible")
```



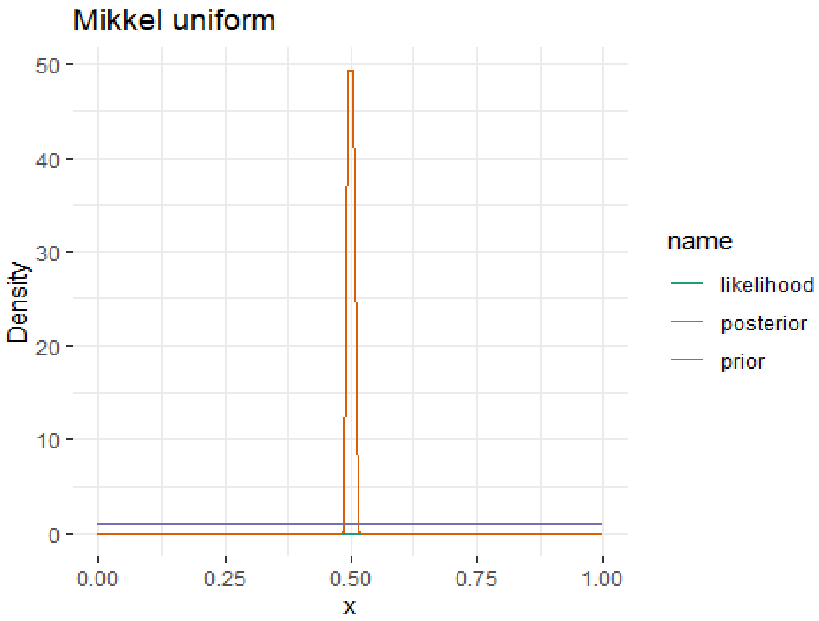
```r
# Josh (16000 / 19800)
Josh <- calc_teacher(16000,19800, uniform_prior)
pretty_plot(Josh$grid, uniform_prior, Josh$likelihood, Josh$teacher_posterior,
title = "Josh uniform")
```



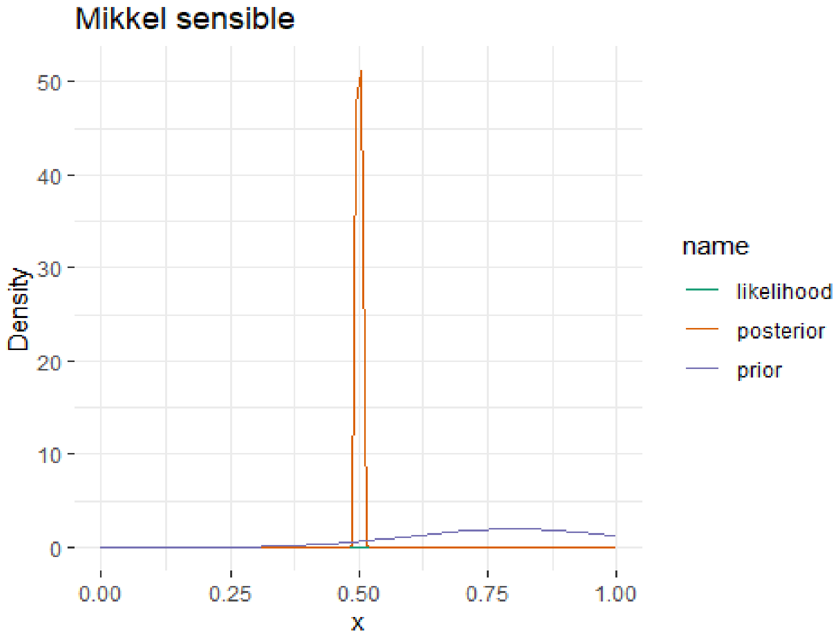
```r
Josh_0.8<- calc_teacher(16000,19800,sensibly_centered_at_0.8)
pretty_plot(Josh_0.8$grid, sensibly_centered_at_0.8, Josh_0.8$likelihood,
Josh_0.8$teacher_posterior, title = "Josh sensible")
```



```
# Mikkel (6600 / 13200)
Mikkel <- calc_teacher(6600,13200, uniform_prior)
pretty_plot(Mikkel$grid, uniform_prior, Mikkel$likelihood,
Mikkel$teacher_posterior, title = "Mikkel uniform")
```



```
Mikkel_0.8<- calc_teacher(6600,13200,sensibly_centered_at_0.8)
pretty_plot(Mikkel_0.8$grid, sensibly_centered_at_0.8, Mikkel_0.8$likelihood,
Mikkel_0.8$teacher_posterior, title = "Mikkel sensible")
```



After adding more data one can observe changes to the posterior. However, the posteriors are less affected by the prior than previously. For instance, this can be seen on the plots for Kristian and Josh. We also see that when the posteriors of the teachers are further away from the prior of 0.8 (Mikkel and Riccardo), they are more influenced by the prior.

5. Imagine you're a skeptic and think your teachers do not know anything about CogSci, given the content of their classes. How would you operationalize that belief?

If we assume the teachers know nothing, it would be reasonable to define the prior to neglect all data points above .5, so that we are entirely sure that there is no chance that a teacher is above a .5 level. The code would look as follows:

```
prior <- ifelse(p_grid > 0.5 , 0 , 1)
```

Second part: Focusing on predictions

Last year you assessed the teachers (darned time runs quick!). Now you want to re-test them and assess whether your models are producing reliable predictions. In Methods 3 we learned how to do machine-learning style assessment of predictions (e.g. rmse on testing datasets). Bayesian stats makes things a bit more complicated. So we'll try out how that works. N.B. You can choose which prior to use for the analysis of last year's data.

1- Write a paragraph discussing how assessment of prediction performance is different in Bayesian vs. frequentist models

In frequentist models you could use a t-test to get a p-value indicating whether or not the means are significantly different. Contrarily, Bayesian models estimate how likely it is that the model has learned something. To check the model, one can apply predictive checks like running simulations or if new data is acquired, one can calculate a new posterior using the previous posterior as a prior.

2- Provide at least one plot and one written line discussing prediction errors for each of the teachers.

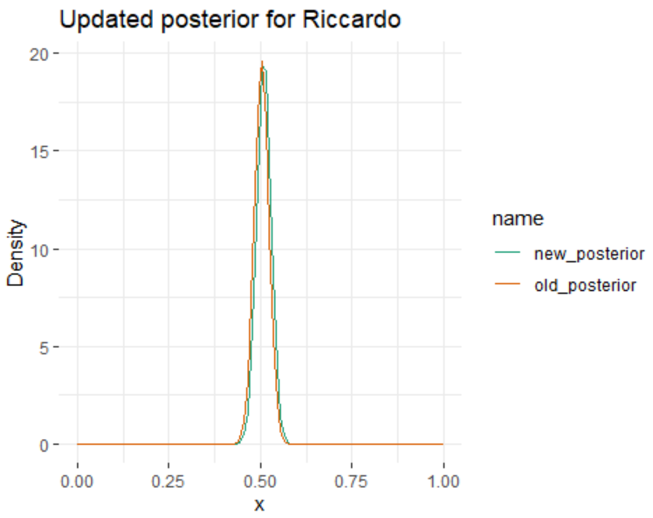
This is the old data:

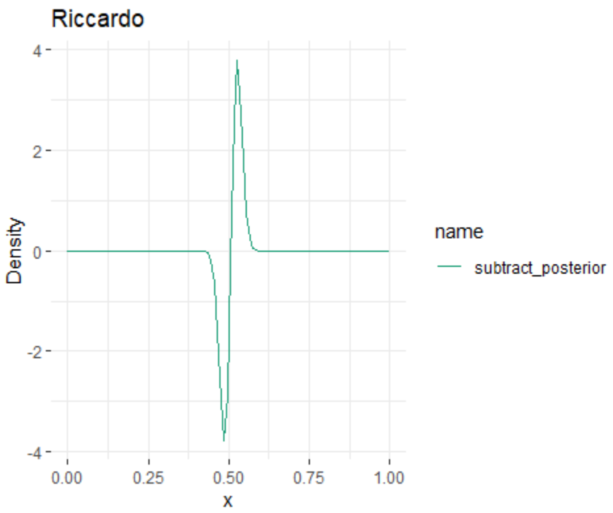
- Riccardo: 3 correct answers out of 6 questions
- Kristian: 2 correct answers out of 2 questions (then he gets bored)
- Josh: 160 correct answers out of 198 questions (Josh never gets bored)
- Mikkel: 66 correct answers out of 132 questions

This is the new data:

- Riccardo: 9 correct answers out of 10 questions (then he freaks out about teaching preparation and leaves)
- Kristian: 8 correct answers out of 12 questions
- Josh: 148 correct answers out of 172 questions (again, Josh never gets bored)
- Mikkel: 34 correct answers out of 65 questions

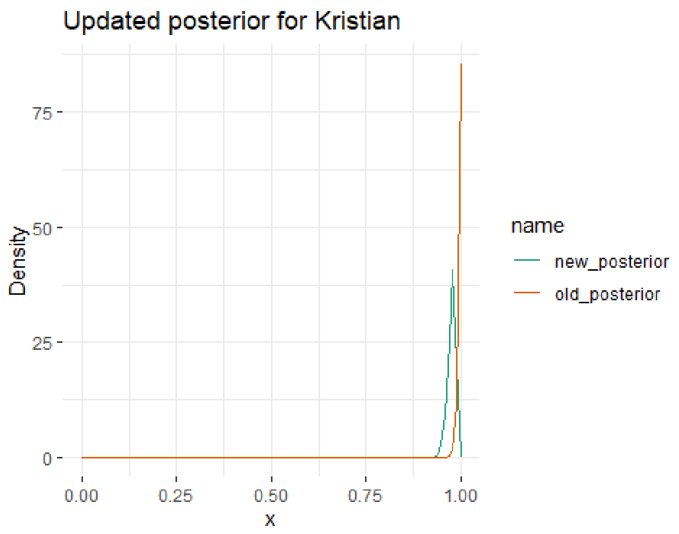
```r
# New function for plotting old posterior and new posterior
posterior_plot <- function(p_grid, prior, posterior, title = " "){
  # define data
  d <- tibble(p_grid = p_grid,
              old_posterior = prior,
              new_posterior = posterior)

  # make to long format
  d <- d %>%
    pivot_longer(cols = c("old_posterior", "new_posterior"), names_to = "name",
values_to = "value")

  # make a
  p <- ggplot(d, aes(x = p_grid, y = value, color = name)) +
    geom_line() +
    labs(x = "x", y = "Density", title = title) +
    theme_bw() +
    ggplot2::theme(panel.background = element_rect(fill = "white"),
                   panel.border = element_blank()) +
    scale_colour_brewer(palette = "Dark2", direction = 1)
  return(p)
}
#1 using the information from previous years and see the difference

#Riccardo
Riccardo_2020 <- calc_teacher(9,10,Riccardo_0.8$teacher_posterior)
posterior_plot(Riccardo_2020$grid, Riccardo_0.8$teacher_posterior,
Riccardo_2020$teacher_posterior, title = "Updated posterior for Riccardo")

subtract_plot(Riccardo_2020$teacher_posterior- Riccardo_0.8$teacher_posterior,
"Riccardo")
```

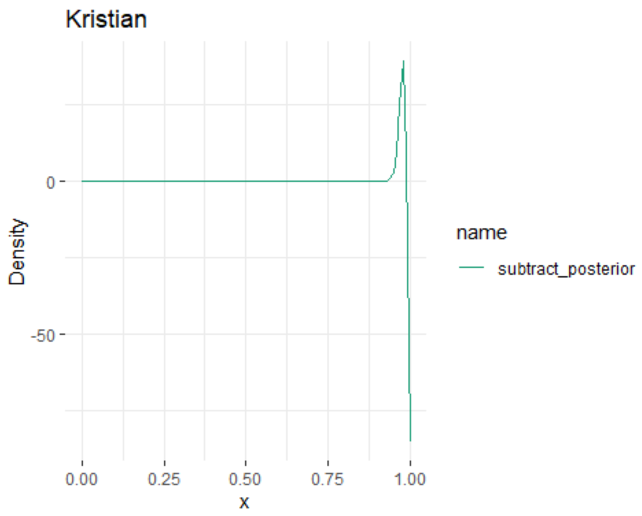




Riccardo: The posterior changed as we do not have a lot of data and the proportion of correct responses is greater for the new data. This has the implication that the new posterior is slightly shifted towards 1 on the x axis. The distribution is still fairly narrow, meaning that we are most certain that Riccardo's cogsci knowledge is around chance (0.5).

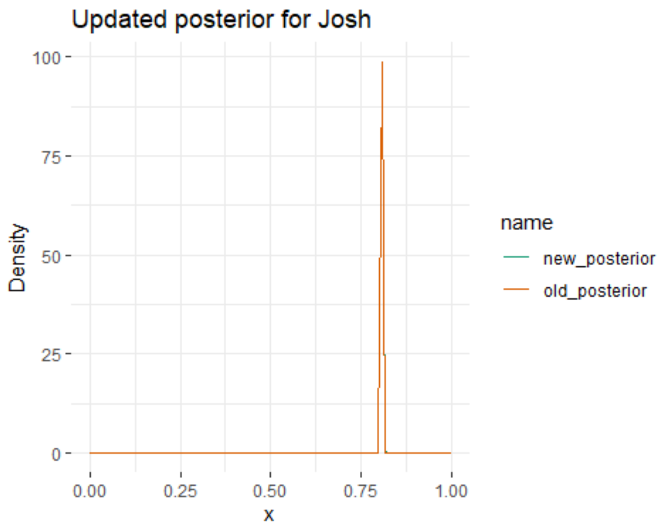
```
#Kristian
Kristian_2020 <- calc_teacher(8,12,Kristian_0.8$teacher_posterior)
posterior_plot(Kristian_2020$grid, Kristian_0.8$teacher_posterior,
Kristian_2020$teacher_posterior, title = "Updated posterior for Kristian")

subtract_plot(Kristian_2020$teacher_posterior- Kristian_0.8$teacher_posterior,
"Kristian")
```

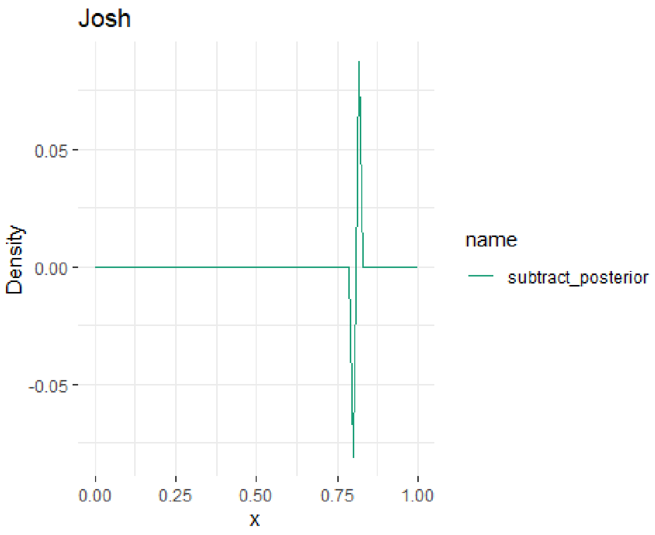




Kristian: The updated posterior is shifted away from 1, meaning that the model has learned from the new data. This is due to the fact that Kristian only answered 2 questions (both correctly) during the first trial and that the second trial had 12 questions - so the amount of data in the second trial is enough to update the posterior.

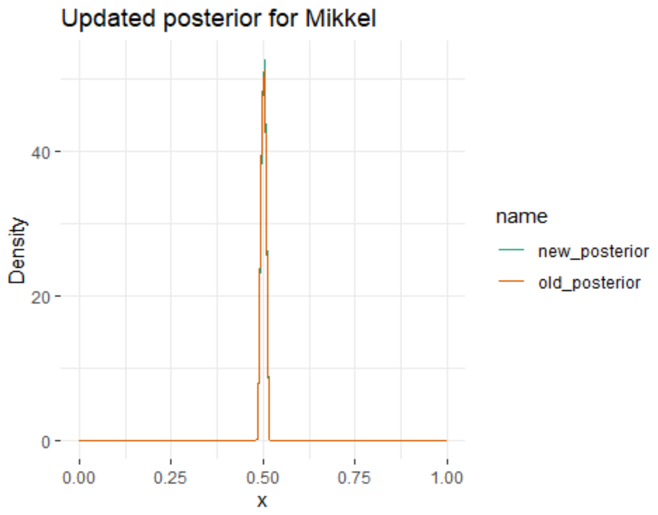
```r
#Josh
Josh_2020 <- calc_teacher(148,172,Josh_0.8$teacher_posterior)
posterior_plot(Josh_2020$grid, Josh_0.8$teacher_posterior,
Josh_2020$teacher_posterior, title = "Updated posterior for Josh")

subtract_plot(Josh_2020$teacher_posterior- Josh_0.8$teacher_posterior, "Josh")
```

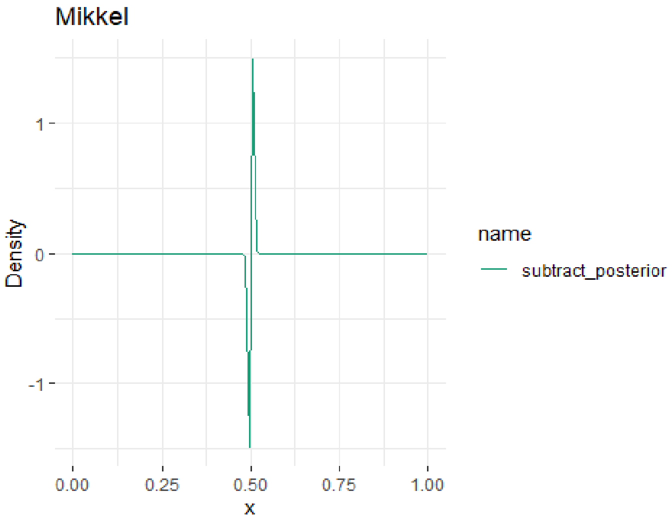




Josh: The updated posterior is almost identical to the old posterior (as can also be seen from the subtracted plot), so the model has not learned from the new data. Josh answered a lot of questions during the first trial, so the amount of questions during the second trial was not enough to change the posterior. Also, the new data is fairly consistent with the old data.

```r
#Mikkel
Mikkel_2020 <- calc_teacher(34,65,Mikkel_0.8$teacher_posterior)
posterior_plot(Mikkel_2020$grid, Mikkel_0.8$teacher_posterior,
Mikkel_2020$teacher_posterior, title = "Updated posterior for Mikkel")

subtract_plot(Mikkel_2020$teacher_posterior - Mikkel_0.8$teacher_posterior,
"Mikkel")
```





Mikkel: The posterior for Mikkel has only changed minimally. This is due to the amount of data for the second trial not being enough to change the posterior, as well as the proportion of correct responses in the second trial being somewhat consistent with the first trial.