Portfolio 2, part 1

Josephine Hillebrand Hansen

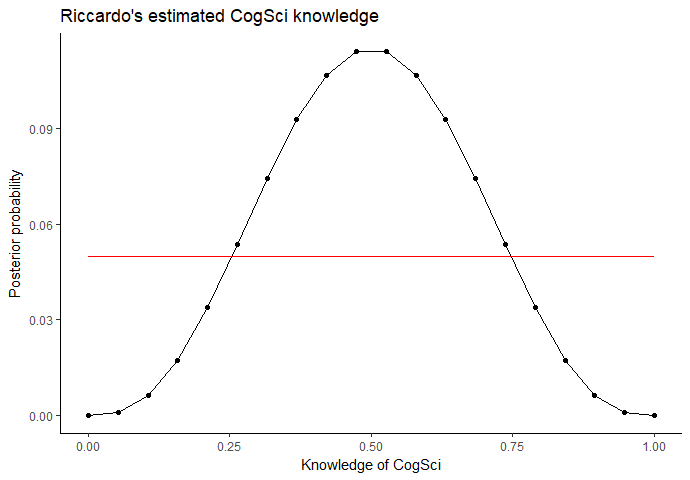
Script: https://github.com/JosephineHH/Assignment-2-2018/blob/master/Assignment2.Rmd

# Question 1) What is Riccardo’s estimated knowledge of CogSci?

We assume that we know nothing about Riccardo’s knowledge of CogSci. Therefore, we assume a flat prior, all outcomes are all equally likely. Riccardo answered 3 out of 6 questions correctly.

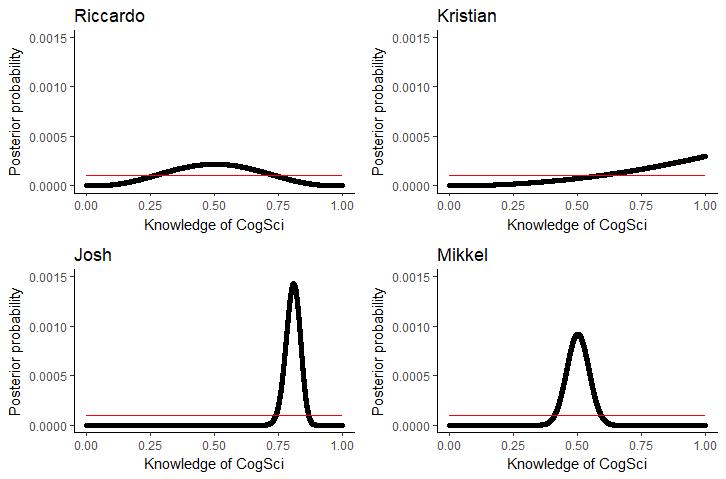
Using quadratic approximation, it was found that Riccardo’s posterior has a mean value of 0.5 with a standard deviation of 0.2 and the 89 % percentile interval is between 0.17 and 0.83. There is a 50 % probability of Riccardo knowing more than chance.

Riccardo’s estimated knowledge of CogSci is plotted below. The black line shows the posterior probability and the red line shows the prior probability:



# Question 2) Estimate all the teachers’ knowledge of CogSci

The teachers estimated knowledge of CogSci is plotted below. All approximations used a flat prior. The red line represents the prior, and the black the posterior distribution:



For Riccardo, the High posterior Density Interval for 50 % of the probability mass of knowledge is between 0.38 and 0.62. Thus, we estimate that there is a 50 percent change that Riccardo knows between 38 % and 62 % about cognitive science. Using quadratic approximation, it was found that Riccardo’s posterior has a mean value of 0.5 with a standard deviation of 0.2 and the 89 % percentile interval is between 0.17 and 0.83.

For Kristian, the High Point Density Interval for 50 % of the probability of knowledge is between 0.84 and 1.00. There is thus a 50 % chance that Kristian knows between 85 % and 100 % about CogSci. Kristian’s posterior has a mean value of 0.77, with a standard deviation of 0.2 and an 89 % percentile interval between 0.37 and 1.[[1]](#footnote-1)

For Josh, the High Point Density Interval for 50 % of the probability mass of knowledge of CogSci is between 0.78 and 0.82. There is a 50 % chance that he knows between 78 % and 82 % of CogSci knowledge. Using quadratic approximation, it was found that Josh’s posterior has a mean value of 0.81 with a standard deviation of 0.03 and the 89 % percentile interval is between 0.76 and 0.85.

For Mikkel, the High Point Density Interval for 50 % of the probability mass of knowledge of CogSci is between 0.47 and 0.53. There is a 50 % chance that he knows between 47 % and 53 % of CogSci knowledge. Using quadratic approximation, it was found that Mikkel’s posterior has a mean value of 0.5 with a standard deviation of 0.04 and the 89 % percentile interval is between 0.43 and 0.57.

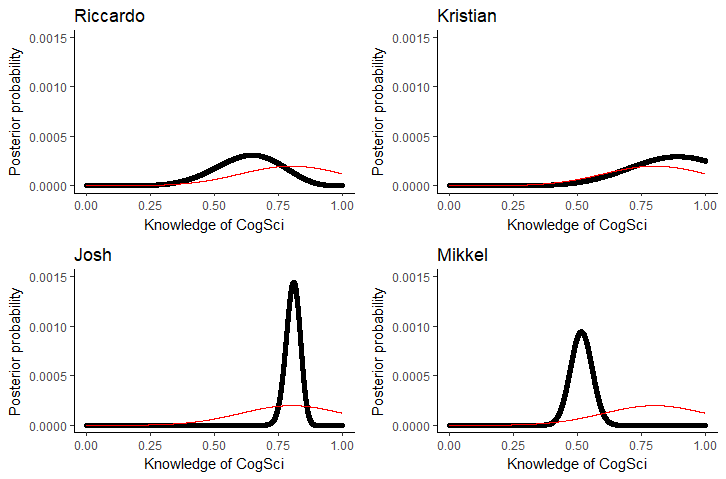
From the plots and intervals we now know the following about the teacher’s CogSci knowledge:

Riccardo’s and Mikkel’s posteriors for CogSci knowledge are both centered around 50 and they both have a 50 % probability of performing above chance level. However, the distribution of Riccardo’s possible knowledge is much broader, so he still has the possibility of performing worse than chance or better, whereas we can be more sure that Mikkel is centered around his peak distribution. E.g. we are pretty sure that Mikkel will only perform at chance level.

Both Kristian and Josh seem to perform above chance level. Kristian has a higher mean for his posterior distribution, but also a large standard deviation. On the other hand Josh might have a lower mean for his posterior probability, but the confidence interval is much lower. This shows that we are pretty confident in our estimation of Josh’s knowledge, whereas we aren’t very confident about our estimation of Kristian’s knowledge – however, Kristian still has the possibility of performing better than Josh, whereas Josh would be the safe choice.

# Question 3) Change the prior to a normal distribution with a mean of 0.8 and a standard deviation of 0.2. How does this change the distribution?

The prior was changed to a normal distribution with a mean of 0.8 and a standard deviation of 0.2. The plots for the 4 teachers can be seen below. The red line is the prior distribution and the black line is the posterior distribution:



Using quadratic approximation, it was found that Riccardo’s posterior has a mean value of 0.65 with a standard deviation of 0.13 and the 89 % percentile interval is between 0.43 and 0.86.

Kristian’s posterior has a mean value of 0.81, with a standard deviation of 0.14 and an 89 % percentile interval between 0.58 and 1.[[2]](#footnote-2)

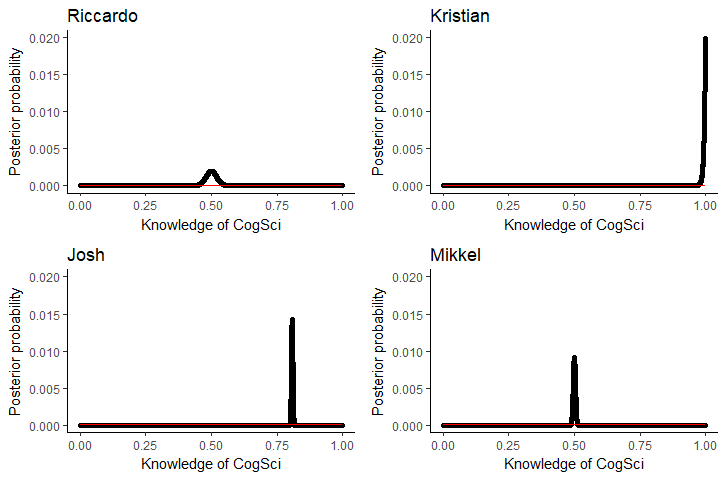
Using quadratic approximation, it was found that Josh’s posterior has a mean value of 0.81 with a standard deviation of 0.03 and the 89 % percentile interval is between 0.76 and 0.85.

Using quadratic approximation, it was found that Mikkel’s posterior has a mean value of 0.51 with a standard deviation of 0.04 and the 89 % percentile interval is between 0.45 and 0.58.

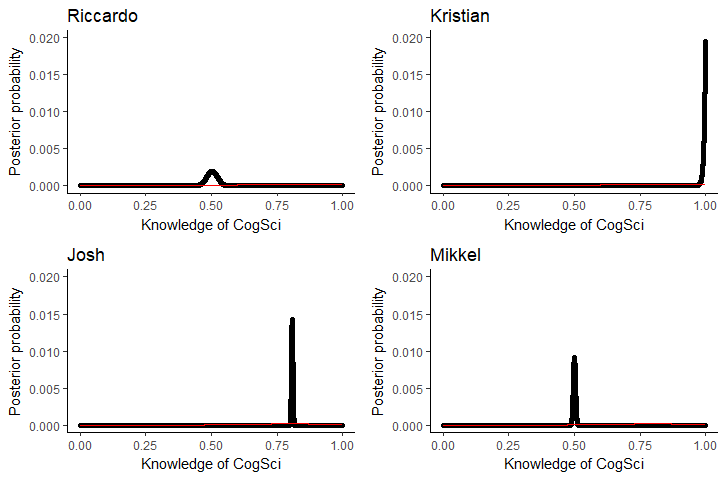
From the plots can be seen that the teachers who answered many questions (Josh and Mikkel) are not affected that much by the prior. Actually, their mean and standard deviation stays the same. However, the teachers with little data collected, Kristian and Riccardo has their posterior probabilities shifted a lot more due to the change in prior. This shows that the prior matters a lot more for the posterior, when we have little actual evidence.

# Question 4) Let us go and collect more data. Do we still see a difference between the results?

Below is showed the flat prior and the posterior for a hundred times as many data points.



Below is showed the optimistic prior and the posterior for a hundred times as many data points.



The prior hardly makes any difference when we collect this much data, looking at the plots. However, quadratic approximations were also made to determine the shape of the posterior. The results can be seen in the table below:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Teacher | Mean | SD | Lower 89 % | Upper 89 % |
| Riccardo flat | 0.5 | 0.02 | 0.47 | 0.54 |
| Riccardo optimist | 0.5 | 0.02 | 0.47 | 0.54 |
| Kristian flat[[3]](#footnote-3) | 1 | 0 | 0.99 | 1 |
| Kristian optimist | 1 | 0.01 | 0.99 | 1 |
| Josh flat | 0.81 | 0 | 0.8 | 0.81 |
| Josh optimist | 0.81 | 0 | 0.8 | 0.81 |
| Mikkel flat | 0.5 | 0 | 0.49 | 0.51 |
| Mikkel optimist | 0.5 | 0 | 0.49 | 0.51 |

Using quadratic approximation, it was found that when collecting this much data, changing the prior doesn’t make a difference for the posterior. Also, collecting more data makes the distribution much narrower, which is also reflected in the very small standard deviations.

This intuitively also makes sense, as there is a more data collected will give more weight to the data, than to the prior. The more data, the more we can rely on the data.

# Question 5) Imagine that you are a skeptic and think that your teachers do not know anything about CogSci. How would you operationalize that belief?

Given that 0 is negative knowledge of CogSci, 0.5 is random chance and 1 is awesome cognitive science super powers and someone performing at chance level will therefore perform at random chance. Thus, we would not expect them to perform below 0.5, which would give them negative knowledge. Therefore, I would operationalize the belief by implementing a prior that is normally distributed with a mean of 0.5. If we expect them to know nothing, we expect them to perform at chance level, which is modelled by the normal distributed prior. We also expect very little chance of performing above and below, which would be normally distributed. Therefore, we would also set a narrow standard deviation of 0.1, to implement the sceptic belief of the teachers’ knowledge of CogSci. The result of the smaller standard deviation is that more evidence is needed to shift the posterior away from the prior.

Portfolio 2, part 2

Script: https://github.com/JosephineHH/Assignment-2-2018/blob/master/Assignment2\_Part2.Rmd

# Question 1) How is assessment of prediction performance different in Bayesian versus frequentist models

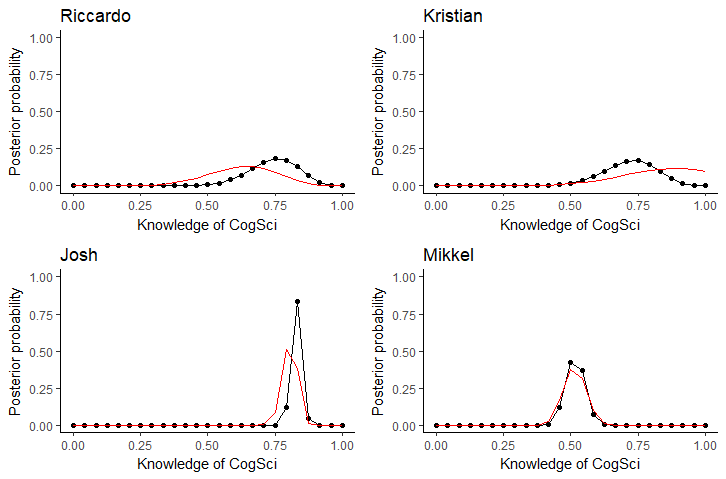
In frequentist models performance is predicted by taking the model and from the model predict values of new data. The prediction performance is then assessed by comparing the predicted values to actual values to evaluate how well the model performed.

In Bayesian models prediction performance is evaluated by putting the performance for last time as the prior and then the performance this time as the posterior and then evaluating how much the distribution changes. We therefore evaluate how well the model performs by looking at how much the distribution changes from the prior distribution to the posterior distribution.

In frequentist statistics we get a single predicted value, precise numbers, for each data point, that we can evaluate against the measurements. In Baysian statistics we get a distribution of likelihood of different outcomes. This does not mean that we cannot evaluate the results, but we need to do it in a different way. One way would be to measure the difference in distributions.

# Question 2) Provide at least one plot and one written line discussing prediction errors for each of the teachers

I believe (or want to believe (-; ) that the university would not hire stupid teachers, and therefore I will use the posterior created last time, using the optimal prior, as our new prior. Thus, the red line in the plot below is last years posterior, this years prior, created from an assumption that the teachers must know something about cognitive science, given that they were hired as professors. The black line is the new posterior after the teachers took the new quiz.



After the new round of tests, it seems that Mikkel obtains the same level of knowledge, more or less. Riccardo seem to become more likely to perform better than what we thought after the first test. Mikkel seem to perform at the same level as before. Kristian seem to perform worse than what was assumed last year. The new data has the largest impact on Riccardo’s and Kristian’s data, as they answered fewer questions in the first round of tests, and therefore our estimations of their knowledge was more unsure.

1. Had to be calculated manually using samples from the posterior, as the map function did not work for Kristian’s data (same number of correct observations as total observations). [↑](#footnote-ref-1)
2. Had to be calculated manually using samples from the posterior, as the map function did not work for Kristian’s data (same number of correct observations as total observations). [↑](#footnote-ref-2)
3. Had to be calculated manually using samples from the posterior, as the map function did not work for Kristian’s data (same number of correct observations as total observations). [↑](#footnote-ref-3)