Portfolio Assignment 2

GROUP 8: JAKOB GRØHN DAMGAARD, HANIA MATERA, BERTRAM HØJER
AND ORLA PATRICIA MALLON

March
2020

4th semester
BA Cognitive Science

Part 1

What is Riccardo’s knowledge of CogSci?

First we used grid approximation to calculate a posterior distribution for Riccardo’s knowledge. This was done using a grid of a 10000 points and a sceptical (assumes chance level of answering correct), normally distributed prior (mean = 0.5, SD = 0.2) with the specified data - 6 questions answered and 3 correct. Plotting the posterior distribution we can see that it is a Gaussian distribution centered around a mean of 0.5 (chance level). Hereby, the model suggests that it is most likely that Riccardo has no apparent knowledge of CogSci. The distribution has a SD of 0.13 which indicates that the model is 95% certain that the actual mean is between - approximately - 0.24 and 0.76. Thus, by conditioning the model on the data, the model has become more certain that Riccardo’s knowledge is around chance level (SD of posterior (0.13) < SD of prior (0.2). Identical results were produced by calculating Ricardo’s knowledge using quadratic approximation on the same prior.

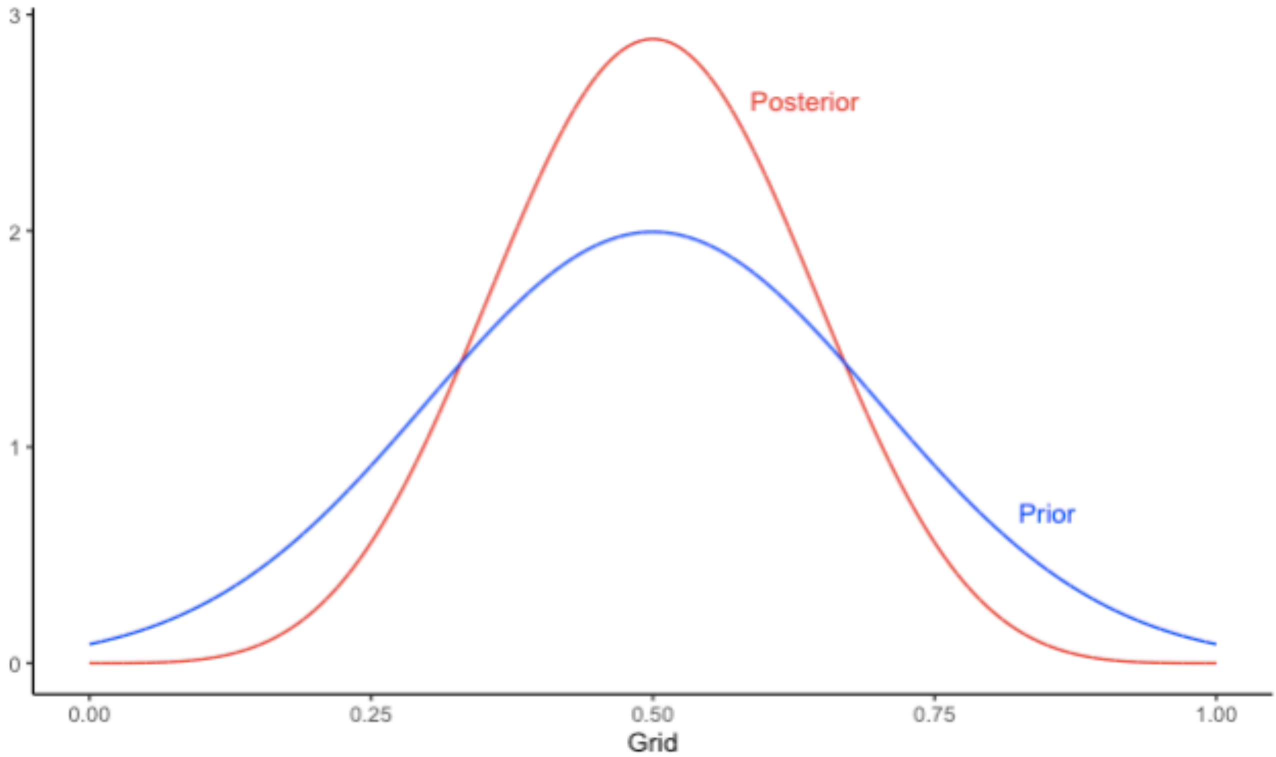


Figure 1 - Prior and posterior probability distributions for Riccardo

Drawing 10000 samples from the posterior distribution enables us to calculate the probability of Riccardo scoring above chance. This is done by simply summing all sampled values that are higher than 0.5. This turns out to be 0.5012 for a specific sample suggesting that there is a 50,12% probability that Riccardo will end up with a score better than chance.

The knowledge of each CogSci teacher:

The above method was used for each teacher. We used a density value of a 10000 grid-points and a normal prior distribution (mean = 0.5, SD = 0.2). Despite identical answer rates (50% correct) there is an interesting difference between the posterior distributions of Riccardo and Mikkel. Riccardo has a 50,12% probability of scoring above chance and Mikkel has a 50.4% probability of scoring above chance. When investigating HPDI of both their scores a visible difference emerges. Riccardo has 89 percentile-intervals from 0.232 to 0.767 while Mikkel has 89 percentile-intervals from 0.429 to 0.567. This is also seen when plotting the distributions where Mikkel's posterior distribution is much more dense than Riccardo's (lower SD):

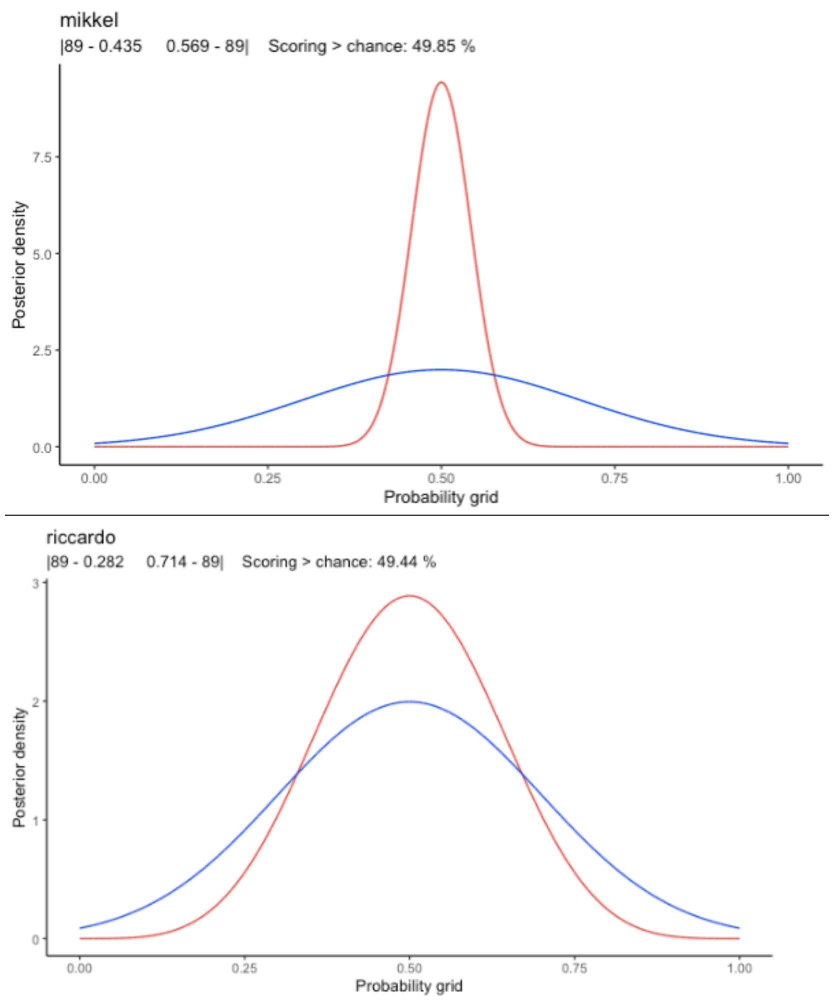


Figure 2 - The effect of more data

Thus, the more data the model is fed, the more certain it becomes of the actual mean.

Changing the prior with a higher appreciation for their knowledge:

By setting a much more positive prior (mean = 0.8, sd = 0.2) we see how the posterior distributions are influenced. The influence is strongest on the teachers with lowest sample sizes. The two people most strongly influenced are Riccardo (n = 6) and Kristian (n = 2). Riccardo and Mikkel originally had the same probability of scoring above chance - Riccardo now has a 84.7% of scoring above chance according to the model while Mikkel has a 62.2% of scoring above chance. This strongly shows how we can influence the behavior of the model by setting appropriate priors.

By increasing the amount of samples we can simulate how the posterior is then influenced by the data. Firstly, more data equals less uncertainty. Hence, when the amount of data is increased by a factor of 100, the standard deviations of the posterior distributions drops. Furthermore, the priors matter much less when the data is multiplied by a factor of 100 - this is good as we obviously want out model to be a good approximation of the data and therefore 'reality'. The prior still slightly influences the posterior distribution but in a much more subtle way.

Part 2

Prediction error assessment

Model prediction assessment in the Bayesian framework is popularly approached via two different methods. The first method has visible similarities to the common frequentist approach. Prediction assessment in the frequentist framework revolves around dismissing subparts of your data during model conditioning and then assessing the predictive power of the model on the left out data. This assessment can be performed by e.g. analysing, how much the model predictions vary from the actual data (calculating RMSE). In the case of our specific task of assessing the CogSci knowledge of our teachers, we could - given the new data from 2020 - perform a prediction error assessment similar to the frequentist method.

Instead of conditioning our defined models on the new data, we compute predictive posterior distributions based on the initial data and simulate the number of questions corresponding to the data provided for 2020. Hereby, we are able to simulate a distribution of how many correct answers the 2019 models predict each teacher to get in 2020. This was done by producing binomial distributions by simulating the number of new questions given to each teacher 10000 times using their 2019 posterior probability distributions as the correctness probabilities. This yielded an actual distribution of predictions of each of the teacher's 2020 results based on their 2019 models. After doing this we were able to plot the prediction error distributions by subtracting the predicted number of correct answers from the simulated prediction and plotting a histogram of the remaining errors. These plot can be seen in figure 3:

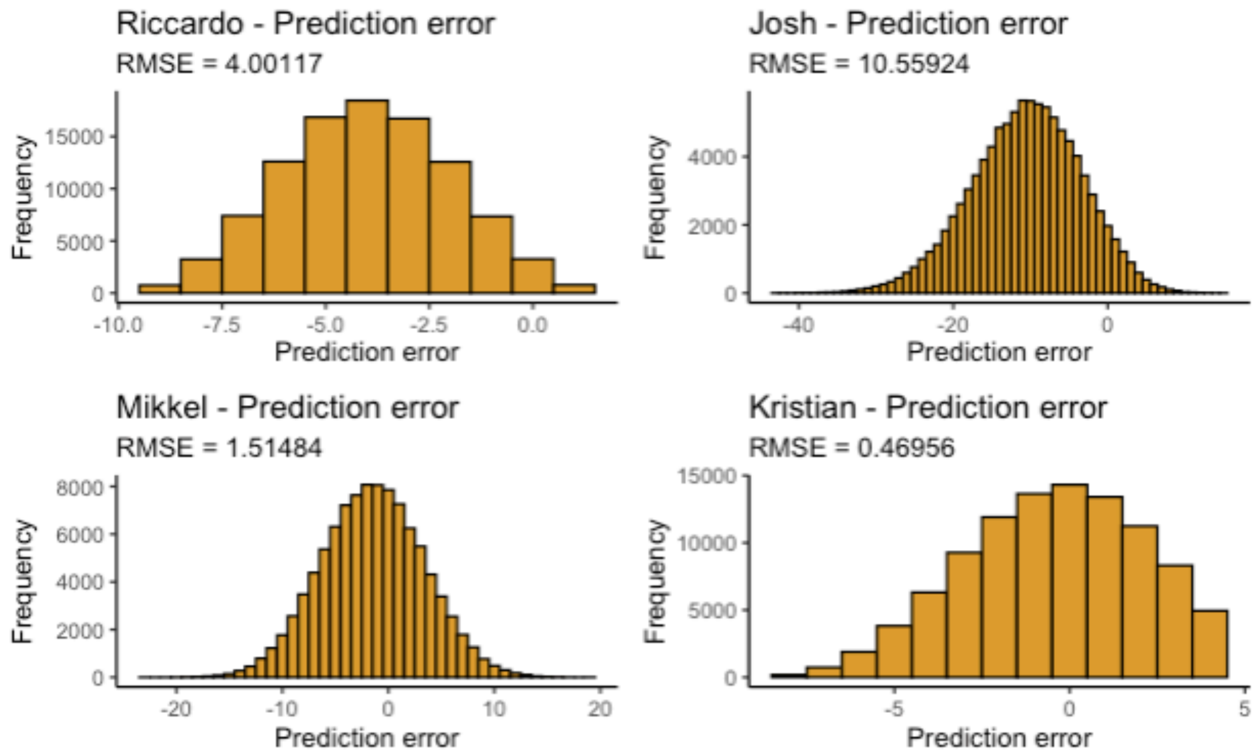
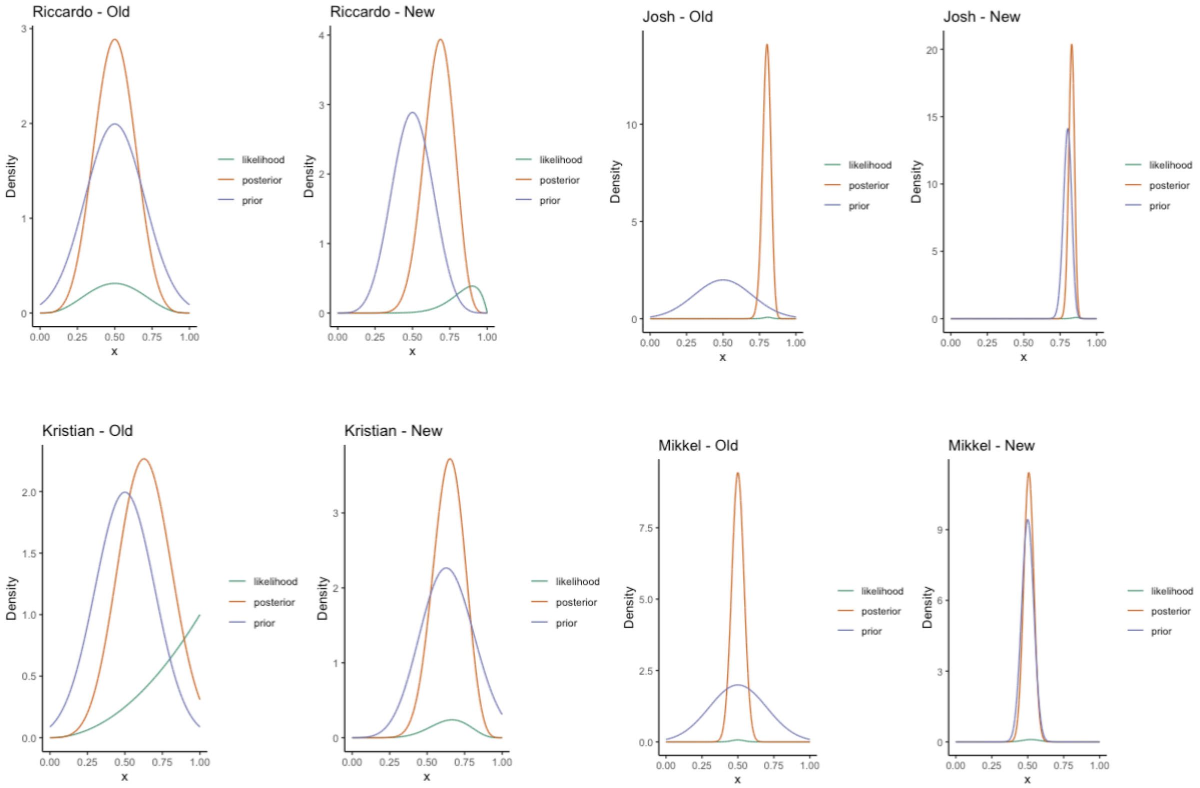


Figure 3 - Prediction error distributions

As the number of questions given to each teacher varies to a high degree it is arguable that the pure RMSE scores for each teacher are incomparable (e.g. as Riccardo was only given 10 question he would never be able to yield a RMSE of 10.56 like Josh). Instead, it would be more optimal to calculate a standardized prediction error value for each teacher e.g. mean percentage error. Nonetheless, it seems the 2019 models for Mikkel and Kristian generally do a good job in predicting his results for 2020. For both Josh and Riccardo, it is, however, apparent that their predictions are negatively skewed meaning. Thus, they perform better in 2020 than predicted by the data from 2019.

Parameter estimates updating - conditioning old model on new data

Despite being highly comparable to prediction assessment in a frequentist perspective, the latter approach disobeys one of the core features of the Bayesian statistics namely a fact that it is always optimal to inform your model of previous data. Thus, another way of analysing prediction is to condition the old models on the new data (using 2019 posteriors as the new priors) and then subsequently assessing whether the new data updates the model estimates. This process yielded the following update posterior distributions:



The new plots that display both the updated posterior and the 2019 posterior distributions as priors provide a good visual of how that data has updated the model. Quantitatively, there are multiple ways of comparing the different distributions (overlapping AUC, HDPI etc.), and we cannot just compare means like we potentially would in a frequentist framework. One way of comparing the updated distribution to the initial posterior is by calculating the probability that the updated model yields a higher performance. In other words, we can calculate the likelihood that the updated model believes the teacher to know more or less about CogSci than the first model. This is simply done by sampling 10000 points from each distribution and calculating the percentage of sampled values from the new model that are higher than the values from the other. By doing that we get the following table:

```
"Likelihood of being better acc. new model: Riccardo =  85.388 %    Josh =  79.312 %    Kristian =  52.855 %    Mikkel =  55.364 %"
```

From this table it is apparent that it seems that the new data has 'significantly' updated the model's beliefs of both Josh' and Riccardo's knowledge of CogSci. The updated models seem confident that Josh and Riccardo have a higher knowledge of CogSci than initially estimated. This statement is further confirmed by eyeballing the plots and in full accordance with the findings from the first approach.

Furthermore, it seems that the updated models for Kristian and Mikkel have simply grown more confident in their estimations of their knowledge (same distribution shape but lower SDs (more narrow distribution)), which is also what is expected when more (similar) data is added to the model.