Assignment 2:

CogSci Teacher Knowledge

Computational Modeling for Cognitive Science

By Astrid Rybner, Kata Molnar, Nicole Dwenger and Sofie Rødkjær

March 12, 2020

[GITHUB!!!]

Part 1

1. What's Riccardo's estimated knowledge of CogSci? What is the probability he knows more than chance (0.5)? Implement a grid approximation and a quadratic approximation with a uniform prior. Calculate the posterior and plot the results.

Using a grid approximation and a uniform prior to calculate the posterior probability resulted in the probability distribution in Figure 1. The distribution is centered around 0.5 and is very symmetrical. Therefore it is assumed that there is 50% probability that Riccardo knows more than chance (0.5). Using a quadratic approximation revealed similar results and also suggest that there is 50% chance that Riccardo knows more about Cognitive Science than chance (0.5).

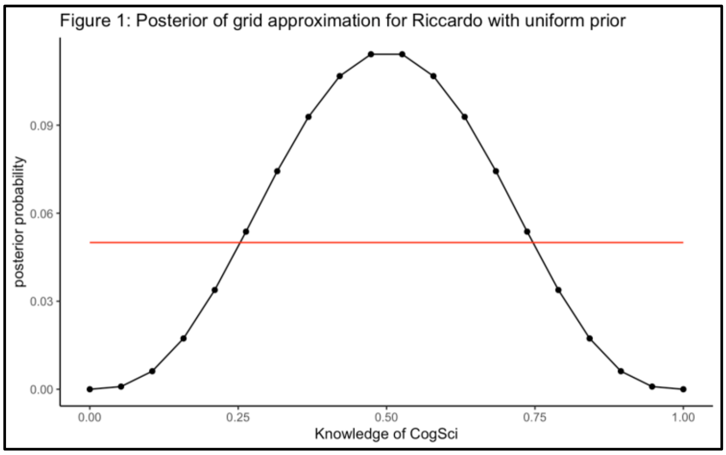

Figure 1. The posterior distribution (using grid approximation) with a uniform prior) of how many answers Ricardo is likely to get correct in a new test, given the fact that he had 3/6 correct answers in this test.

2. Estimate all the teachers' knowledge of CogSci. Who's best? Use grid approximation. Comment on the posteriors of Riccardo and Mikkel. They are both symmetrical, only Mikkel's distribution is narrower, so it is more sure or something.

Looking at the plots (Figure 2), Josh seems to be the most knowledgeable of the 4 teachers. He posterior distribution is centered above 0.75 and seems to have a very small standard deviation.

Riccardo and Mikkel both have posterior plots that center around 0.5 and that are symmetrical. However, the standard deviation in Mikkels posterior distribution is much smaller than Riccardo's. This is due to Mikkel having answered more questions than Riccardo and therefore he has more data. The small standard deviation of Mikkel's posterior suggests that the model is more certain (has less variance) of its mean than in Riccardo's posterior.

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

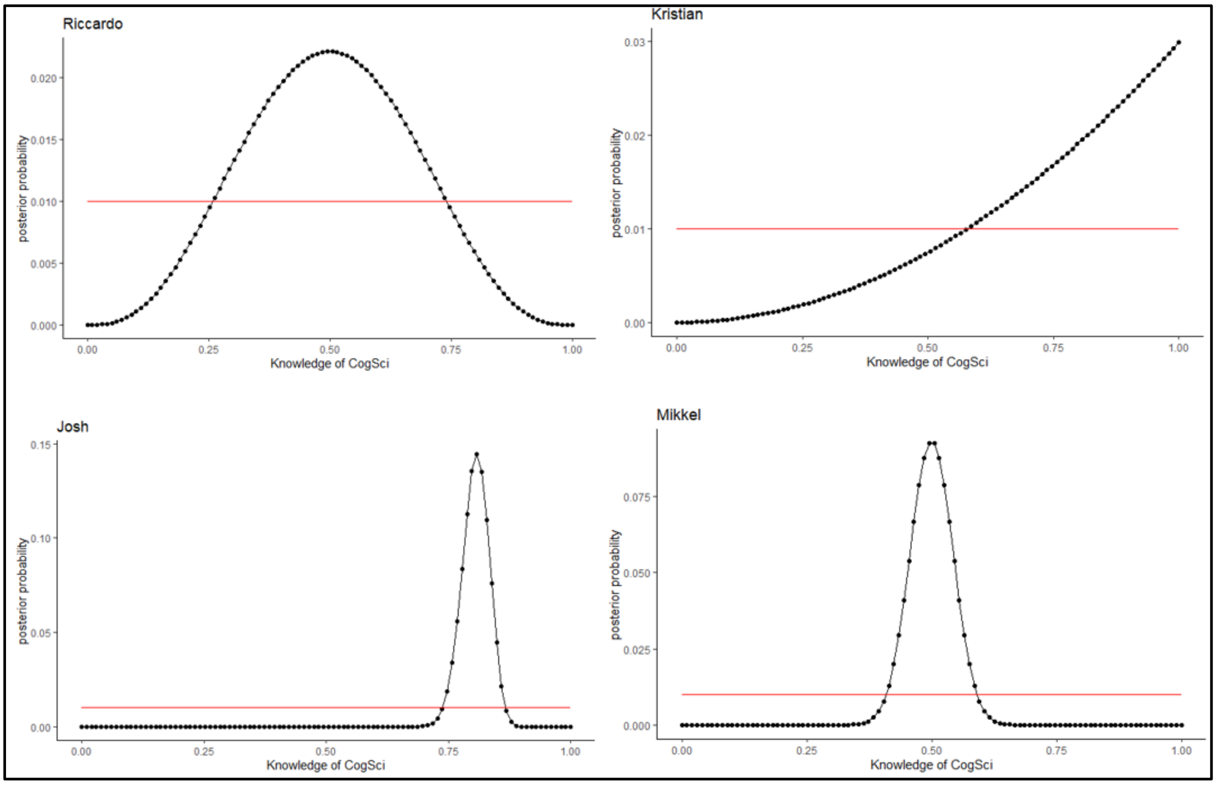

Figure 2. Posterior distributions (using grid approximation) with uniform priors. These plots show how how many correct answers the respective teachers are likely to get in a new test based on how they performed in the first test if we don't have any expectations of their performance.

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)?

Changing the priors to a normal distribution with a mean of 0.8 and a standard deviation of 0.2 changes mostly the posterior distribution for Riccardo and Kristian, meaning they are more influenced by the prior than Josh and Mikkel. More precisely, for Riccardo the peak of the distribution moves more to the right, while in Kristians case the peak moves more to the left, i.e. in both cases towards the mean of the prior (0.8). This suggests, that when we have very few data points (e.g. Riccardo, Kristian) the posterior is more influenced by the prior, than when we have a lot (e.g. Josh, Mikkel).

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

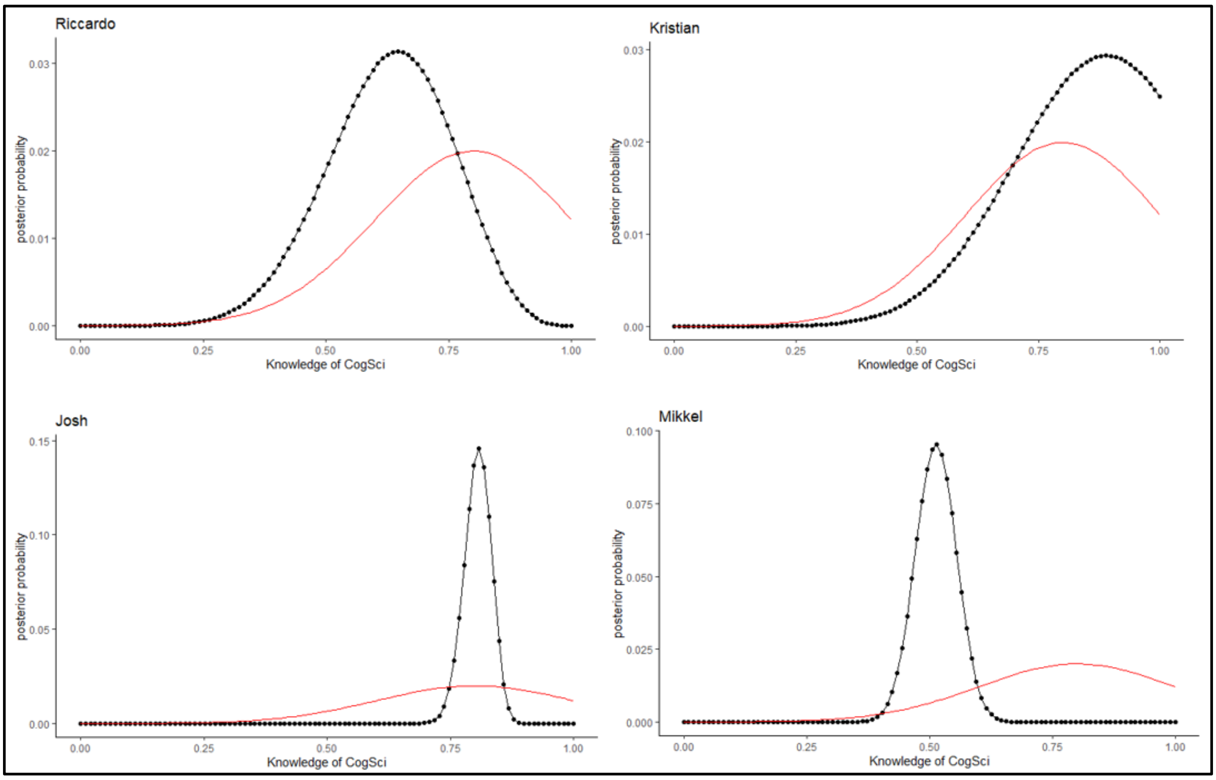

Figure 3. Posterior distributions (using grid approximation) with a normally distributed prior (mean = 0.8 and sd = 0.2). These plots show how how many correct answers the respective teachers are likely to get in a new test based on how they performed in the first test if we expect them to perform above chance.

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?

When increasing the amount of data, there seems to be no difference between using a uniform prior or a more informed (normally distributed) prior for estimation of the posterior. This suggest that with a lot of data the prior does not have a big influence on the posterior - the obtained data now has a bigger influence.

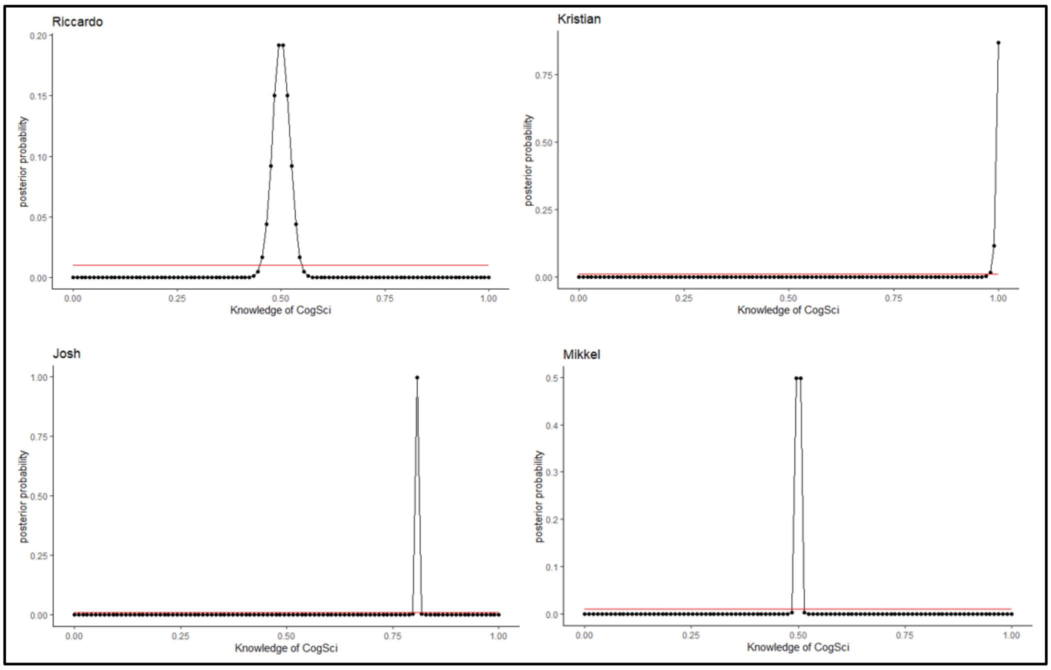

Figure 4. Posterior distributions (using grid approximation) with uniform priors, using increased amount of data (multiplied by 100).

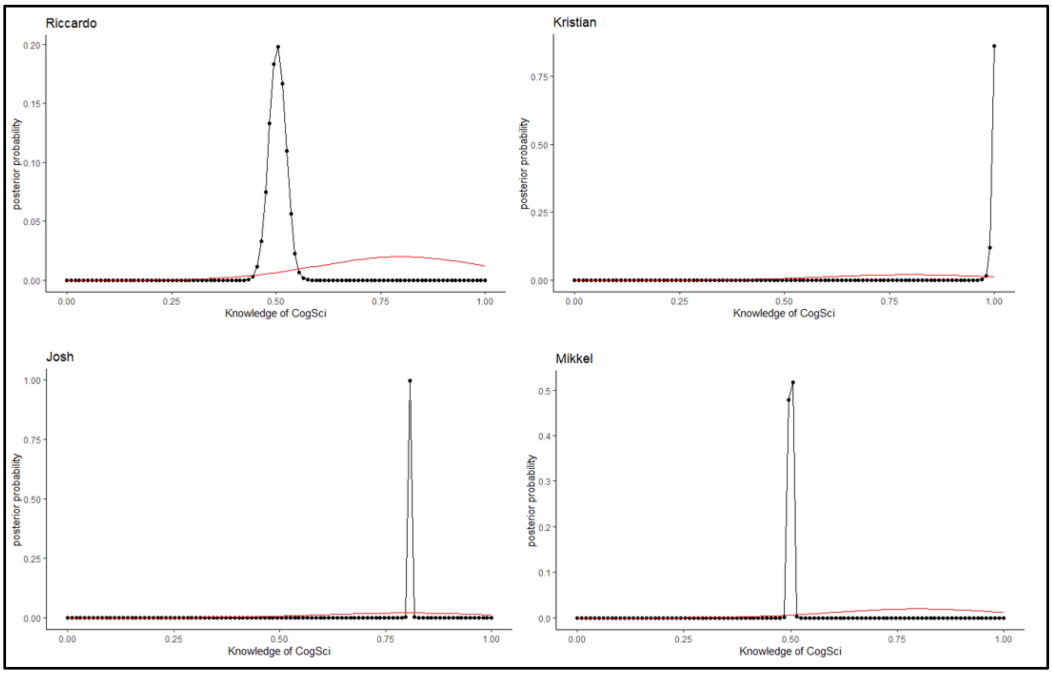

Figure 5. Posterior distributions (using grid approximation) with a normally distributed prior (mean = 0.8 and sd = 0.2), using increased amount of data (all multiplied by 100).

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?

We assumed that the questions asked were not multiple-choice questions, but that the teachers had to come up with the answer on their own. Therefore, we changed the prior to be below chance, with a mean of 0.2 and standard deviation of 0.2. This embodies our assumption, that they know next to nothing about Cognitive Science and therefore will only get very few of the questions correct (and not below 0 correct). When using the normal, original data, we can see that Riccardo and Kristian are again more influenced by the prior, and their posterior is now dragged to the left, i.e. the peak of both distributions is below chance (0.5). For both Josh and Mikkel (because they had so much data) the posterior distribution is still similar to the one obtained when we used other priors (see Figure 2, Figure 3).

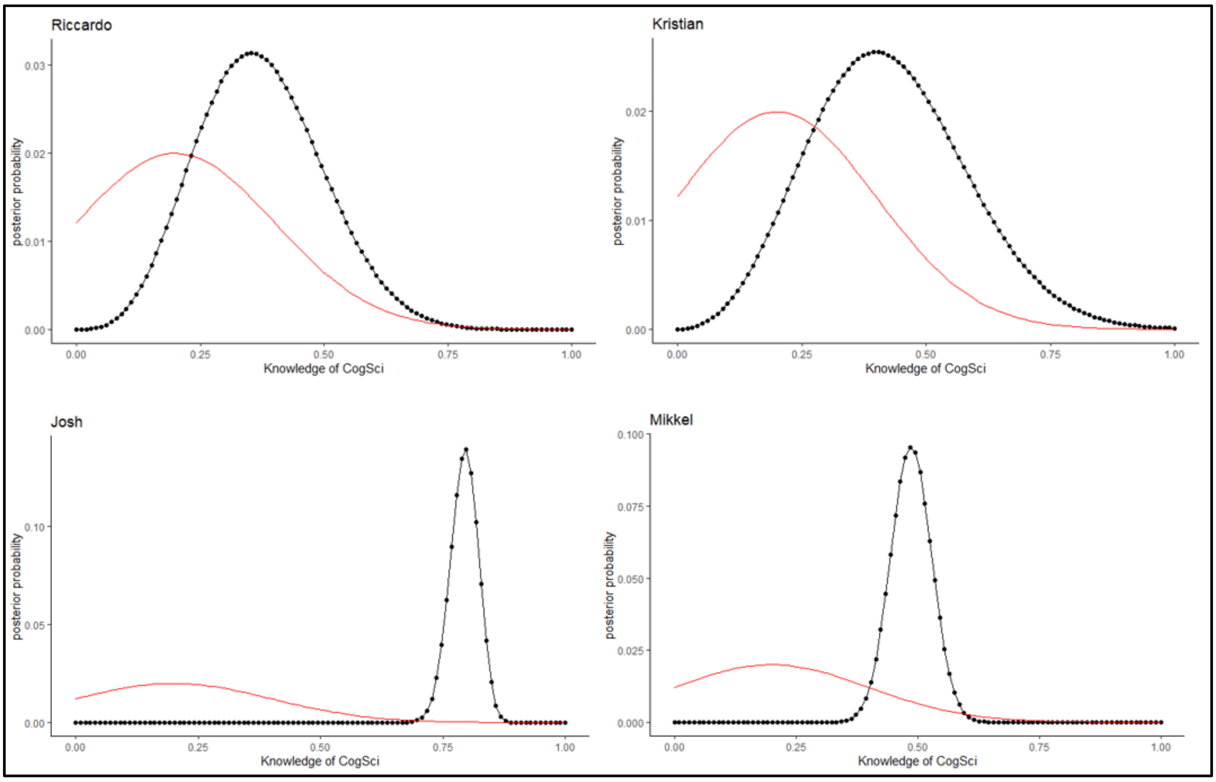

Figure 6. Posterior distributions (using grid approximation) with a normally distributed prior (mean = 0.2 and sd = 0.2).

Part 2

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

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