Portfolio 2

GitHub: <https://github.com/clolesen/methods4-portfolio2>

# PART 1

## 1.

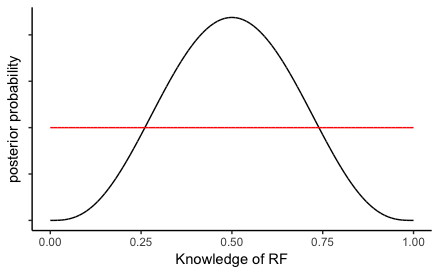
*What's Riccardo's estimated knowledge of CogSci? What is the probability he knows more than chance (0.5) [try figuring this out. if you can't peek into chapters 3.1 and 3.2 and/or the slides]?*

*- First implement a grid approximation (hint check paragraph 2.4.1!) with a uniform prior, calculate the posterior and plot the results*

*- Then implement a quadratic approximation (hint check paragraph 2.4.2!).*

*- N.B. for the rest of the exercise just keep using the grid approximation (we'll move to quadratic approximations in two classes)*

There is a 50% probability that RF’s CogSci knowledge is above chance. This was calculated by summing up all the posterior probabilities above 0.5 in the grid. This also makes perfect sense when looking at the plot below, where the distribution centers around 0.5 and looks normally distributed. The quadratic approximation estimates the mean to be 0.5 with a standard deviation of 0.2.



**Plot 1** - Plot showing the posterior distribution (black line) and the prior (red line) of the model for RF.

## 2.

*Estimate all the teachers' knowledge of CogSci. Who's best? Use grid approximation. Comment on the posteriors of Riccardo and Mikkel.*

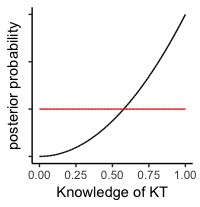
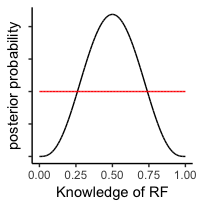
*2a. Produce plots of the prior, and posterior for each teacher.*

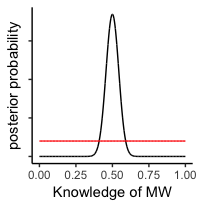
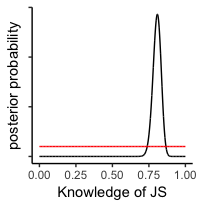
The percentage of posterior probabilities of a CogSci knowledge above 0.75 was calculated for each teacher and shown in table 1 below. The 0.75 threshold was chosen because it is the half way point between random chance and full CogSci knowledge and so this upper part should reflect who has the highest probability of being the most knowledgeable. This means that the model suggests that JS possesses the most CogSci Knowledge.

Since both RF and MW answered half of the questions right, both of their posterior probability distributions centers around 0.5. The difference is that we have a lot more data on MW than RF, which makes the curve of the distribution narrower. This means that the model is surer that MW’s knowledge is around random chance than RF’s knowledge, which is also reflected by the numbers in table 1.

|  |  |
| --- | --- |
| **Teacher** | **%** |
| RF | 7,1 |
| KT | 57,8 |
| JS | 97 |
| MW | <0,1 |

**Table 1** - Percentage of posterior probabilities above 0,75 knowledge per teacher.





**Plot 2** - Plot showing the posterior distribution (black line) and the prior (red line) of the model for each teacher.

## 3.

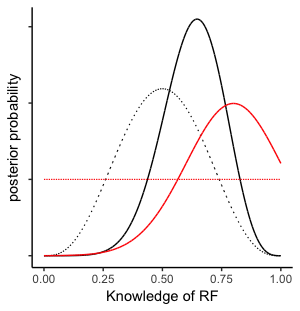
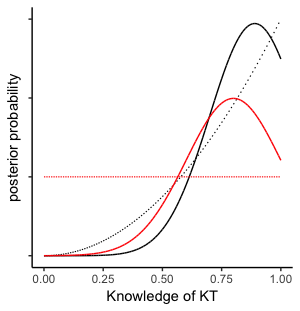
*Change the prior. Given your teachers have all CogSci jobs, you should start with a higher appreciation of their knowledge: the prior is a normal distribution with a mean of 0.8 and a standard deviation of 0.2. Do the results change (and if so how)?*

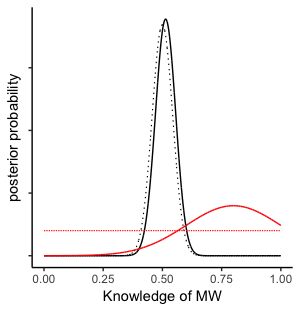
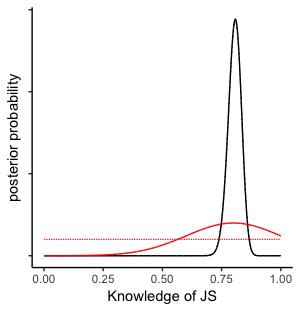
*3a. Produce plots of the prior and posterior for each teacher.*

The table below is an updated version of table 1. Here the numbers calculated from the model with a non-flat prior is also shown. The new prior affects the results for RF and KT in a positive direction but for the others are barely affected. This is also clearly demonstrated in plot 3. The reason for this is that there is considerably less data available for RF and KT and therefore the prior information weights relatively more in these cases than the in the others. The model is then more likely to skew the posterior probability distribution towards the prior.

|  |  |  |
| --- | --- | --- |
| **Teacher** | **% (flat prior)** | **% (non-flat prior)** |
| RF | 7,1 | 17,1 |
| KT | 57,8 | 68,4 |
| JS | 97 | 97,1 |
| MW | <0,1 | <0,1 |

**Table 2** - Percentage of posterior probabilities above 0,75 knowledge per teacher and for models with different priors.





**Plot 3** - Plot showing the posterior distribution (black lines) and the prior (red lines) of the models for each teacher. The dotted lines represent the models with a flat prior as also shown in plot 2. The solid lines represent the models with a non-flat prior. Note that in the case of JS the two posterior distributions are so close that they appear to overlap completely in the plot.

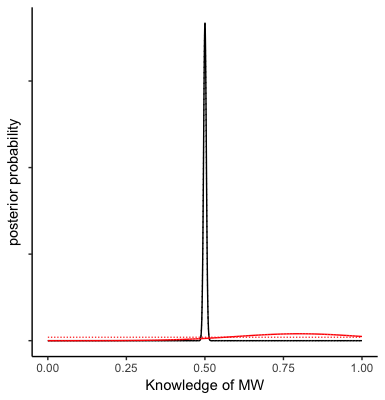
## 4.

*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?*

Table 3 resembles table 2 but the numbers are calculated from the new and bigger set of data. It is quite clear that the results are much more extreme than before and that the priors doesn’t have an effect. The posterior has become very narrow with the new data set (demonstration in plot 4). Again, the low difference between models with different priors and the narrowness of the posterior probability distribution curves are a result of more data. The model is now much more sure of the teachers respectable levels of knowledge.

|  |  |  |
| --- | --- | --- |
| **Teacher** | **% (flat prior)** | **% (non-flat prior)** |
| RF | <0,1 | <0,1 |
| KT | 100 | 100 |
| JS | 100 | 100 |
| MW | 0 | 0 |

**Table 3** - Percentage of posterior probabilities above 0,75 knowledge per teacher and for models with different priors. These numbers are calculated from a bigger data set.



**Plot 4** - Plot showing the posterior distribution (black lines) and the prior (red lines) of the models for MV. The dotted lines represent the models with a flat prior as also shown in plot 2. The solid lines represent the models with a non-flat prior. This plot is made from a bigger data set.

## 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?*

One way is to make the prior distribution a normal distribution with a mean of 0.5. This would reflect the belief that the teachers are at chance level and thus they must answer more questions right in order to convince us that they are worthy CogSci’s.

# PART 2

## 1.

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

In frequentist statistics assessment is based upon point estimates like root mean square error (RMSE), which is the average error between the model prediction and the observed data. These estimates can be assessed using model validation techniques like cross validation. This way it is relatively easy to compare model predictions, since they are usually represented by single numeric values.

In the Bayesian framework, models are assessed by comparing the posterior probability distribution to the observed data. This makes the job of comparing model predictions less straight forward than it can be in the frequentist tradition. On the other hand, the Bayesian model prediction is also less arbitrary to interpret since it captures the full uncertainty of the model.

## 2.

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

In plot 5 the models with a Gaussian prior is shown for each teacher. They are the same as those also shown in plot 3, but here there is a vertical dotted line representing newly collected data. For RF, KT and JS the model did not predict the new observations very well, since there is not much space under the posterior curve at the point of the observation. In the case of KT the model seems to do a lot better than RF or JS, but it is worth considering that the distribution for KT is very wide due to low sample size and one should therefor be skeptical when assessing anything from this. Common for all three is that the new observation isn’t to far of the prior distribution, which suggest that the teachers might actually meet our expectations of them. For MW it is another story. The new observation for MW is close to the predictive posterior mean and far from the mean of the prior. This suggests that MW does not meet our expectations of how knowledgeable a teacher of cognitive science should be but instead strengthens the notion that MW only perform at the level of chance.



**Plot 5** - Plot showing the posterior distribution (black line) and the prior (red line) of the model for each teacher. The dotted line represents the new observation (number of correct answers divided by number of questions).