Activity 2: Exploring posteriors, likelihoods, and priors

In this activity, we will re-use our simulated data and fitted model of the mean from the previous activity to explore how priors and likelihood interact to produce a Bayesian posterior.

You will hopefully prove a few things to yourself through this activity:

1. Bayesian posteriors are nearly equivalent to likelihoods when the priors are uninformative.
2. Likelihood-based estimates are based on the data and the model, whereas Bayesian posteriors incorporate additional information from the priors.
3. Good informative priors can strengthen inferences from small data sets. Bad informative priors can ruin inferences from even the best data sets.

Source files

1. R script to simulate data and run MCMC: `./practicals/1_model_of_the_mean/1_script.R`
2. Stan code for the model of the mean: `./practicals/1_model_of_the_mean/1_model.stan`
3. R function to make a plot comparing the likelihood, prior, and posterior for the "mu" parameter from our model of the mean: `./practicals/1_model_of_the_mean/1_script_fun.R`
   Note: This function will be automatically loaded into the script above, so you don't necessarily need to dive into its source code.

What's in the script

We are re-using the script from our previous activity building the model of the mean. At the very end of that script, we made a plot that compared the likelihood, prior, and posterior estimates of the "mu" parameter from our model. Now, we are going to modify a few things and see how the likelihood and posterior compare with various priors.

Before we change the priors, let's just check a few things from our original model that used vague priors.

Q: Are there differences between the Bayesian posterior and the likelihood?

Q: Is the prior flat within the range of parameter values that you are assessing, or is it more informative? What would an informative prior look like in this plot?

Exercises to explore

Modify and refit the model (by editing the Stan model code) to have an informative prior that is different than the simulated value of the μ parameter: e.g. `mu ~ normal(6.5, 0.1)`. Re-plot the posterior, prior, and likelihood.

Q: Are there differences between the Bayesian posterior and the likelihood? Why or why not?

Q: Is the prior flat within the range of parameter values that you are assessing, or is it more informative?

Now, modify the simulated data so that the sample size is VERY small (e.g. n=10) and rerun the model:

Q: Did this change the degree of influence that the prior had on the posterior estimates? Why or why not?

Q: How might this change if we made the sample size VERY large (e.g. n=1e5)? Give that a try.

—-

Now, modify the simulated data so that the sample size is VERY small (e.g. n=10) and rerun the model with an uninformative prior: mu ~ normal(0, 100)

Q: What happened to our accuracy estimating the parameter? What if we run that again without changing any parameters (i.e. just re-run the script to generate a new stochastic simulated data set)?

Q: How is accuracy affected if we use a good informative prior that is centred on the true value of the parameter with a relatively small standard deviation? mu ~ normal(5, 1)

Q: What happens to the uncertainty in our posterior parameter estimate if you further reduce the standard deviation of the informative prior? mu ~ normal(5, 0.1)