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/2_likelihoods_priors_posteriors/2_script.R

We begin by loading the fitted model and data from Activity 1, so make sure those were saved. It may be best (but not required) if you start this activity with results based on the default simulation settings from Activity 1.

What's in the script

In **Section 1 of the script**, we will create a couple of functions to help with this activity:

- Lfun() manually calculates the log-likelihood for our data with given parameter values. Notice that this actually a very simple process. We use dnorm() to get the probability density (on the log scale) for a single data point assuming a normal distribution with a fixed mean and standard deviation. Repeat this for each data point in our sample and then sum all the log-densities together to get the log-likelihood.
- rescale_log_likelihoods() rescales our log-likelihoods to be on the same scale as our Bayesian posteriors so that we can plot them together. YOU CAN HONESTLY IGNORE THE DETAILS OF THIS FUNCTION (which came mostly from ChatGPT).

Q: Why did we choose dnorm() as the basis for our function Lfun() to calculate the likelihoods rather than another distribution like dlnorm() or dunif() or dbeta()?

In **Section 2 of the script**, we create the data that we need to plot the distribution of the posterior, prior, and likelihood for the mean μ. The details of each of these steps are not critical, but I've left a couple of notes in the code where you may want/need to specify your own values if you are not working with the default model from activity 1.

In **Section 3 of the script**, we plot the posterior, prior, and likelihood together on the same plot along with the true simulated value for this parameter.

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 modifying and re-running code from Activity 1) to have an informative prior that is different than the simulated value of the μ parameter: e.g. mu ~ normal(3, 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 with an uninformative prior: mu ~ normal(0, 100)

Q: What happened to our accuracy estimating the parameter?

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)