Homework 3

PSTAT 115, Spring 2023

Due on May 26, 2022 at 11:59 pm

Problem 1. Posterior Credible Intervals

One way for us to learn more about posterior distributions is to find some credible intervals. Suppose we have a posterior density of a parameter λ, defined by $\lambda|y \sim \text{Gamma}(4,1)$.

a. (7pts) Plot this distribution. Construct the posterior middle 95% credible interval for λ. The middle 95% interval is such that 2.5% of the probability mass is left to the left and 2.5% to the right. Save the lower and upper endpoints in a vector of length 2, called `middle_95`. Using either line segments or shading, display this interval on the plot.

b. (5pts) Interpret the interval you found. How is it different from a frequentist confidence interval?

c. (8pts) Besides the middle 95% credible interval, we could also find the 95% highest posterior density (HPD) region. This region contains the 95% of posterior values with the highest posterior densities. The HPD region will always be the shortest credible interval for a given probability, since it by definition contains the values of λ with the highest probability of occurring. Use `HDInterval::hdi()` to construct the HPD region. Save the lower and upper endpoints of this region in a variable called `hdi_region`. Add this interval to the plot you made in part (a), making sure that both intervals are distinguishable on the plot.

d. (5pts) Based on your plot, how do the two kinds of 95% credible intervals differ? How long is the middle interval? The HDI interval?

Problem 2. Posterior predictive Beta-Binomial

As we saw in the lectures, it is sometimes possible to derive an exact posterior predictive model. In this exercise we will discuss one further example. To begin, suppose we observe $Y = y$ successes in n trials where $Y \mid \pi \sim Bin(n, \pi)$ and $\pi \sim Beta(\alpha, \beta)$.

a. (5pts) Identify the posterior probability density function (pdf) of π given the observed data $Y = y$, denoted by $f(\pi \mid y)$. This will depend on $(y, n, \alpha, \beta, \pi)$.

b. (5pts) Suppose we conduct n' *new* trials, wheren n' might differ from our original number of trials n) and let $Y' = y'$ be the observed number of successes in these new trials. Identify the conditional probability mass function (pmf) of Y' given π, denoted by $f(y' \mid \pi)$. This will depend on (y', n', π).

c. (10pts) Idenfity the posterior predictive pmf of Y', denoted by $f(y' \mid y)$. This will depend on $(y, n, y', n', \alpha, \beta)$.

d. (7pts) Suppose now that your posterior model for π is based on a prior model with $\alpha = 4$ and $\beta = 6$ and an observed $y = 14$ successes in $n = 100$ original trials. We plan to conduct $n' = 20$ *new* trials. Specify the posterior predictive pmf of Y', the number of successes we might observe in these 20 trials.

e. (8pts) Continuing part d., suppose instead we plan to conduct $n' = 4$ *new* trials. Specify and sketch the posterior predictive pmf of Y', the number of successes we might observe in these 4 trials.

Problem 3. Australian temperature

Let μ be the average 3 p.m. temperature in Perth, Australia. Not knowing much about Australian weather, your friend's prior understanding is that the average temperature is likely around 30 degrees Celsius, though might be anywhere between 10 and 50 degrees Celsius. To learn about μ, they plan to analyze 1000 days of temperature data. Letting Y_i denote the 3 p.m. temperature on day i, they'll assume that daily temperatures vary Normally around μ with a standard deviation of 5 degrees:

$$Y_i \sim N(\mu, 5^2).$$

a. (5pts) Tune and plot a Normal prior for μ that reflects your friend's understanding.

b. (7pts) The `weather_perth` data set in the folder `homework4/data/` includes 1000 daily observations of 3 p.m. temperatures in Perth (`temp3pm` column). Plot this data and discuss whether it's reasonable to assume a Normal model for the temperature data.

c. (8pts) Identify the posterior model of μ and verify your answer using the function `summarize_normal_normal()` saved in the folder `homework4/functions`.

d. (10pts) Plot the prior pdf, likelihood function, and posterior pdf of μ. Describe the evolution in your understanding of μ from the prior to the posterior.

Problem 4. Normal-Normal calculation

Prof. Abebe and Prof. Morales both recently finished their PhDs and are teaching their first statistics classes at Bayesian University. Their colleagues told them that the average final exam score across all students, μ, varies Normally from year to year with a mean of 80 points and a standard deviation of 4. Further, individual students' scores Y vary Normally around μ with a known standard deviation of 3 points.

a. (5pts) Prof. Abebe conducts the final exam and observes that his 32 students scored an average of 86 points. Calculate the posterior mean and variance of μ using the data from Prof. Abebe's class.

b. (5pts) Prof. Morales conducts the final exam and observes that her 32 students scored an average of 82 points. Calculate the posterior mean and variance of μ using the data from Prof. Morales' class.

c. (10pts) Next, use Prof. Abebe and Prof. Morales' combined exams to calculate the posterior mean and variance of μ. Interpret your results.