Homework 4

PSTAT 115, Winter 2025

Due on March 10th, 2025 at 11:59 pm___

Problem 1. Frequentist Coverage of The Bayesian Posterior Interval.

Suppose that $y_1, .., y_n$ is an IID sample from a $Normal(\mu, 1)$. We wish to estimate μ.

1a. For Bayesian inference, we will assume the prior distribution $\mu \sim Normal(0, \frac{1}{\kappa_0})$ for all parts below. Remember, from lecture that we can interpret κ_0 as the pseudo-number of prior observations with sample mean $\mu_0 = 0$. State the posterior distribution of μ given $y_1, .., y_n$. Report the lower and upper bounds of the 95% quantile-based posterior credible interval for μ, using the fact that for a normal distribution with standard eviation σ, approximately 95% of the mass is between $\pm 1.96\sigma$.

1b. Plot the length of the posterior credible interval as a function of κ_0, for $\kappa_0 = 1, 2, ..., 25$ assuming $n = 10$. Report how this prior parameter effects the length of the posterior interval and why this makes intuitive sense.

1c. Now we will evaluate the *frequentist coverage* of the posterior credible interval on simulated data. Generate 1000 data sets where the true value of $\mu = 0$ and $n = 10$. For each dataset, compute the posterior 95% interval endpoints (from the previous part) and see if it the interval covers the true value of $\mu = 0$. Compute the frequentist coverage as the fraction of these 1000 posterior 95% credible intervals that contain $\mu = 0$. Do this for each value of $\kappa_0 = 1, 2, ..., 25$. Plot the coverage as a function of κ_0.

1d. Repeat the 1c but now generate data assuming the true $\mu = 1$.

1e. Explain the differences between the coverage plots when the true $\mu = 0$ and the true $\mu = 1$. For what values of κ_0 do you see closer to nominal coverage (i.e. 95%)? For what values does your posterior interval tend to overcover (the interval covers the true value more than 95% of the time)? Undercover (the interval covers the true value less than 95% of the time)? Why does this make sense?

Problem 2. Rstan Warm up for Women's World Cup

Chinese Women soccer team has won AFC Woman's Asian Cup recently. Suppose you are interested in the following World Cup performance of Chinese women soccer team. Let λ be the be the average number of goals scored of Chinese Women team. We will analyze λ by Gamma-Poisson model where data Y_i is the observed number of goals scored in World Cup games. ie. we have $Y_i|\lambda \sim Pois(\lambda)$ and $\lambda \sim Gamma(a, b)$. According to a sport analyst, they believe that λ follows a Gamma distribution with $a = 1$ and $b = 0.25$.

2a. Compute the theoretical posterior parameters a, b, and also posterior mean mu.

2b. Create a stan file named `women_cup.stan`, use Rstan to Report and estimate of the posterior mean of the scoring rate by computing the sample average of all Monte Carlo samples of λ. **2c.** Produce histogram for simulated lambda and density plot for theoretical posterior distribution of lambda. Does the simulated results coincide with the theoretical ones? Briefly explain your answer.

2d. Right now, we have lambda samples generated by Rstan. Use them as samples from posterior distribution to compute the mean of predicative posterior distribution to estimate the possible goal scored for next game for Chinese women soccer team.