ECON 3640–001

Problem Set 3

Marcio Santetti

Spring 2022

Problem 1

(a) Simulate 1,000 values drawn from a **Poisson**-distributed random variable, with rate parameter 16. Do not forget to set a **seed** for reproducibility.

(b) Generate 10,000 values drawn from a **Beta**-distributed random variable, with shape parameters 10 and 15. Do not forget to set a **seed** for reproducibility.

(c) Generate 1,000 values drawn from a **uniformly**-distributed random variable, ranging from 0 to 20. Do not forget to set a **seed** for reproducibility.

(d) Repeat part (c), now simulating 10,000 values. Produce a histogram for both parts, and compare your results.

(e) Generate 5,000 draws from a Normal distribution, with mean 5 and standard deviation 2. Generate these simulated data's density curve, and put, on the same plot, its long-run probability density function (PDF) curve. Compare your results. Do not forget to set a **seed** for reproducibility.

Problem 2

Consider the following model:

$$\begin{aligned} Y|\theta &\sim \mathrm{Binomial}(n, \theta) \\ \theta &\sim \mathrm{Beta}(3, 8) \end{aligned}$$

Approximate the posterior distribution, $\theta|Y$, using **grid approximation**. How "thin" the grid will be is up to you, but make sure to play around with different values.

Lastly, assume that you observe $n = 100$ trials, with $Y = 5$ successes.

Problem 3

Use **grid approximation** to estimate the posterior distribution for a Gamma-Poisson model, where you observe $n = 3$ independent data points $(Y_1, Y_2, Y_3) = (0, 1, 0)$. The likelihood and prior are the following:

$$\begin{aligned} Y_i|\lambda &\sim \mathrm{Poisson}(\lambda) \\ \lambda &\sim \mathrm{Gamma}(20, 5) \end{aligned}$$

Problem 4

Using `stan` syntax, set up the `model{}` block of a model string for an MCMC simulation for the following priors and likelihood functions:

(a) $Y|\theta \sim \text{Binomial}(20, \theta)$ and $\theta \sim \text{Beta}(1, 1)$;

(b) $Y|\lambda \sim \text{Poisson}(\lambda)$ and $\lambda \sim \text{Gamma}(4, 2)$;

(c) $Y|\mu \sim \mathcal{N}(\mu, 1)$ and $\mu \sim \mathcal{N}(0, 10)$;

(d) $Y|\mu \sim \mathcal{N}(\mu, \sigma)$, with $\mu \sim \mathcal{N}(0, 10)$ and $\sigma \sim \text{Uniform}(0, 100)$.

Problem 5

Using the information from **Problem 2**, set up an **MCMC** simulation using `rstan` with 4 chains and 10,000 iterations. Then, do the following:

(a) Generate trace plots for your Markov chains. Did they converge?

(b) Generate the density plots of your parameter. Are these densities similar to each other?

(c) What is the posterior mean of your parameter?

(d) Interpret the 95% credibility interval you've obtained from your simulation.

Problem 6

Using the `howell.csv` data set (available on Canvas), run an MCMC simulation to approximate the posterior distribution for the average *height*, but only for adults (i.e., individuals older than 18 years old). In other words, make sure to apply the `filter()` function to your data set first.

You may assume, for simplicity, that the standard deviation parameter, σ, is constant and known. You may use any value you want, as long as it makes sense. Recall that, if your data are normally distributed, approximately 95% of the observations lie within 2 standard deviations of the mean, or $\mu \pm 2 \cdot \sigma$. Take this detail into account when setting up your Bayesian model.

Answer the same questions from Problem 5.

Problem 7

Still using the `howell.csv` data set, now estimate a Bayesian model for the posterior distributions for *weights* (in kilograms). This time, you will have to set a prior for σ as well.

Here, make sure to play around with models including every individual in the sample, but also for adults only.

As a bonus, you may estimate models also based on `gender` since, these data are also available in this data set. With different research questions in mind, make sure to think about your priors and adjust your data accordingly.