Math 526 – Statistics II, Spring 2023

Midterm exam

Due: Tuesday, April 18, by 8:00PM

Problem: The dataset provided consists of independent and identically acquired datapoints w. Each w is generated by one out of a total of *4 clusters* and follows the statistics

$$w|s \sim \text{Normal}\left(\mu_s, \frac{1}{\tau}\right).$$

The cluster locations μ_{σ_m} are known and have the values

$$\mu_{\sigma_1} = 1, \qquad \mu_{\sigma_2} = 2, \qquad \mu_{\sigma_3} = 3, \qquad \mu_{\sigma_4} = 4.$$

The precision τ is the same for all clusters and has an unknown value.

1. Formulate a Bayesian model for cluster analysis. Make your own choices for the necessary priors and briefly reason on your selection.

2. Describe a Markov chain Monte Carlo scheme to sample from your model's posterior.

3. Implement your Markov chain Monte Carlo scheme and generate samples to characterize the joint posterior probability distribution of your model.

4. Estimate the probability that datapoints w_{15} and w_{25} stem from the same cluster.