SDS 383C: Statistical Modeling I, Fall 2022
HW 4, 110 points, Due Nov 09, 12:00 Noon

Instructor: Abhra Sarkar (abhra.sarkar@utexas.edu)
Teaching Assistant: Preston Biro (prestonbiro@utexas.edu)
Department of Statistics and Data Sciences
The University of Texas at Austin
2317 Speedway D9800, Austin, TX 78712-1823, USA

All homework must be submitted typed-in as a single pdf file. Name the file "firstname-lastname-SDS383C-HW-4.pdf" Submit this file without compression such as zip or rar. Figures accompanying the solutions must be presented close to the actual solution. Computer codes will be rarely evaluated but must still be submitted separately from the main file. Codes must be commented properly and should run easily on other machines. Precise, concise, clear, innovative solutions may be rewarded with bonus points. Explain your answer with logic reasoning and/or mathematical proofs. Organize your solutions in the same order as they were presented. If you can solve a problem using multiple techniques, present only your best solution.

1. (5 points) For a normal likelihood model with conjugate Normal-Inverse-Gamma prior on (μ, σ^2), show that a-priori and a-posteriori μ and σ^2 are dependent but uncorrelated.

   [Hints: No need to explicitly find the joint expectations - prove the result using law of iterated expectations $\mathbb{E}(\mu \cdot \sigma^2) = \mathbb{E}_{\sigma^2}\mathbb{E}_{\mu|\sigma^2}(\mu \cdot \sigma^2 \mid \sigma^2)$ for general NIG families. For the posterior result, appeal to conjugacy.]

2. (20 points) Consider again a normal likelihood model but with NIP $p(\mu, \sigma^2) \propto \sigma^{-2}$.

   (a) Show that a-posteriori μ and σ^2 are dependent but uncorrelated.

   (b) Find out the marginal posteriors $p(\mu \mid \mathbf{y}_{1:n})$ and $p(\sigma^2 \mid \mathbf{y}_{1:n})$.

   (c) Find out the predictive distribution $p(y_{new} \mid \mathbf{y}_{1:n})$.

3. (5 points) For a multinomial likelihood model with K categories, show that the Jeffreys' prior for the category probabilities is $\text{Dir}(1/2, \ldots, 1/2)$.

4. (35 points) The 'galaxies' dataset from package 'MASS' in R gives the velocities in kms/sec of 82 galaxies (export this dataset from R if you are using a different programming language). Divide all these values by 1000. Using these scaled values as your data points, do the following problems.

   (a) Using the EM algorithm, fit location mixtures of normals

   $$f(y) = \sum_{k=1}^{K} \pi_k \text{ Normal}(y \mid \mu_k, \sigma^2)$$

   with $K = 4, 6, 8, 11, 15, 20$ components. Summarize your results by showing the fitted densities superimposed over a histogram of the data points in a single figure with 3×2 panels.

[To facilitate convergence, keep the means μ_k's fixed at initial values set at k points spread over the range of values of y for the first 20 or so iterations and only update π_k's and σ^2. After 20 or so iterations, update all parameters.]

(b) Tabulate AIC and BIC values for each case and report the 'best' model(s).

(c) Next, fit location-scale mixtures of normals

$$f(y) = \sum_{k=1}^{K} \pi_k \text{Normal}(y \mid \mu_k, \sigma_k^2)$$

with $K = 3, 4, 5, 6, 7, 8$ components. Summarize your results by showing the fitted densities superimposed over a histogram of the data points in a single figure with 3×2 panels.

(d) Tabulate AIC and BIC values for each case and report the 'best' model(s).

(e) Summarize your general findings.

5. (25 points) Repeat everything you did in Problem No 3 above but this time using the stochastic EM algorithm.

6. (20 points) The 'faithful' dataset from package 'datasets' in R gives eruption and waiting times of the old faithful geyser in Yellowstone national park (export this dataset from R if you are using a different programming language).

   (a) Using the EM algorithm, fit location-scale mixtures of bivariate normals

$$f(\mathbf{y}) = \sum_{k=1}^{K} \pi_k \ \text{MVN}_2(\mathbf{y} \mid \boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k)$$

   to this dataset with $K = 2, 3, 4, 5$ components. Summarize your results by showing contours of the fitted densities superimposed over a scatterplot of the data points in a single figure with 2×2 panels.

   (b) Tabulate AIC and BIC values for each case and report the 'best' model(s).