Stat 260/CS 294 Homework Assignment 4 (due April 12)

1. Consider the following model:

$$\begin{aligned} x_i \mid \mu &\overset{iid}{\sim} \mathrm{Ca}(\mu) \\ \mu &\sim \mathrm{N}(0, 10), \end{aligned}$$

where the density of the Cauchy distribution, $\mathrm{Ca}(\mu)$, is:

$$p(x \mid \mu) = \frac{1}{\pi(1 + (x - \mu)^2)}.$$

(a) Given two data points, $x_1 = -4.3$ and $x_2 = 3.2$, plot the posterior distribution over the range $[-10, 10]$. (Let the scale on the y-axis be arbitrary; i.e., don't worry about the normalization).

(b) Show that the Cauchy distribution can be written as a scale mixture of a Gaussian distribution.

(c) Using the scale-mixture representation, develop a Gibbs sampler for the posterior distribution in part (a). I.e., find the conditional probabilities for μ and for the variable λ introduced in writing the Cauchy as a scale mixture.

(d) Implement the Gibbs sampler in R (or Matlab) and run the sampler for 10,000 iterations, with a burn-in of 1,000 iterations. Make a histogram of the resulting set of samples, and compare to the plot in part (a).

(e) Plot a time series of samples of values of μ for a few hundred iterations, chosen so as to show that the sampler is coping with the bimodal nature of this posterior.

2. The following well known data set consists of 15 pairs of numbers (GPA, LSAT) for a sample of American law schools:

SAT(x)	576	635	558	578	666	580	555	661	651	605	653	575	545	572	594
GPA(y)	3.39	3.3	2.81	3.03	3.44	3.07	3	3.43	3.36	3.13	3.12	2.74	2.76	2.88	2.96

Table 1: Data from Efron and Tibshirani Bootstrap Monograph.

The model is:

$$y_i = \beta_0 + \beta_1 x_i + \epsilon_i, i = 1, \ldots, n,$$

where the ϵ_i are $\mathrm{N}(0, 1/\tau)$. Using the prior $p(\beta_0, \beta_1, \tau) = 1/\tau, \tau > 0$ write down the joint posterior and find the conditionals needed for a Gibbs sampler.

(a) Implement the Gibbs sampler in R (or Matlab). Initializing with ($\beta_0 = 0, \beta_1 = 0, \tau = 1$), run the sampler for 10,000 iterations with a burn-in of 1,000 iterations. Plot traces for the three parameters. Do these appear to have mixed? Plot histograms for the samples of β_0, β_1 and τ. Plot a scatterplot of (β_0, β_1). You should see a strong posterior correlation.

(b) Consider an orthogonalization in which $(\beta_0, \beta_1, \tau) \to (\beta_0', \beta_1', \tau')$, where $\beta_0' = \beta_0 + \beta_1 \bar{x}$, $\beta_1' = \beta_1$, and $\tau' = \tau$. Redo the plots that you did for part (a). Has the mixing improved?

3. Stratified random sampling is a popular survey sampling scheme where the population of N units is first divided into J non-overlapping subpopulations, or strata, of $N_1, N_2, \ldots, N_J$ units so that $\sum_{j=1}^{J} N_j = N$. The population size N and the stratum sizes N_j are all assumed known. Once the strata have been determined, a simple random sample (here we assume without replacement) is drawn from within each stratum, the drawings being made independently across strata. The sample sizes within each strata are given by $n_1, \ldots, n_J$ and the total sample size is $n = \sum_{j=1}^{J} n_j$.

Stratum 1	Stratum 2		
324	180	130	101
797	314	172	121
507	238	153	116
748	296	163	119
457	235	138	113
381	192	132	104

Table 2: Inhabitants, in thousands, from a stratified random sample of 24 US cities.

Let Y_{ij} denote unit i of the population in stratum j and let $I = \{I_{ij}\}$ be the collection of inclusion indicators, where $I_{ij} = 1$ if the i-th population unit in stratum j is included in the sample, and $I_{ij} = 0$ otherwise. We also denote by $Y_I = \{y_{ij}\}$ the collection of sampled units, where y_{ij} represents unit i of the random sample from stratum j. Finally, let $D = (Y_I, I)$ be the observed data conditional upon which all posterior distributions will be evaluated.

(a) Let $\bar{y}_j = \frac{1}{n_j}\sum_{i=1}^{n_j} y_{ij}$ be the sample mean from stratum j. Prove that:

$$E\left[\sum_{j=1}^{J} \frac{N_j}{N}\bar{y}_j\right] = \bar{Y},$$

where $\bar{Y}$ is the population mean $\bar{Y} = \frac{1}{N}\sum_{j=1}^{J}\sum_{i=1}^{N_j} Y_{ij}$.

(b) Consider the following prior distribution independently for each stratum j:

$$Y_{ij} \mid \mu_j, \sigma_j^2 \overset{iid}{\sim} \mathrm{N}(\mu_j, \sigma_j^2), j = 1, \ldots, J.$$

Assuming that all the σ_j^2's are known, and a flat prior on each μ_j, show that $\mu_j \mid D, \sigma_j^2 \overset{indep}{\sim} \mathrm{N}(\bar{y}_j, \frac{\sigma_j^2}{n_j})$.

(c) Derive the posterior distribution, $p(\bar{Y}_j \mid D, \sigma_j^2)$, for each population stratum mean $\bar{Y}_j = \frac{1}{N_j}\sum_{j=1}^{N_j} Y_{ij}, j = 1, \ldots, J$. Next derive the posterior distribution, $p(\bar{Y} \mid D, \{\sigma_j^2\})$, for the mean $\bar{Y}$.

(d) Now assume that the σ_j^2's are all unknown and consider the noninformative prior $p(\mu_j, \sigma_j^2) \propto 1/\sigma_j^2$ for each j. Describe clearly an exact posterior sampling algorithm that will yield samples from the posterior distribution of the population stratum means $\bar{Y}_j$. You do not need to show your derivations for this part. Explain how the posterior samples of $\bar{Y}_j$ for $j = 1, \ldots, J$ will deliver the posterior samples of $\bar{Y}$.

(e) Table 2 presents data from one of the first stratified sampling exercises in the United States that was carried out in 1920 to estimate the number of inhabitants in the $N = 64$ largest cities of the time. These 64 cities were divided into $J = 2$ strata, where the first strata comprised $N_1 = 16$ of the largest cities and the second strata comprised the remaining $N_2 = 48$ cities. Within the first stratum $n_1 = 6$ cities were chosen, while $n_2 = 18$ cities were chosen from stratum 2. The total number of inhabitants, in thousands, for each of these $n = 24$ cities is given in Table 1.

Using the data in Table 1 write an R (or Matlab) program to compute the posterior estimates (mean, median and 95% credible intervals) for the total number of inhabitants in the 64 cities. Also present posterior estimates (mean, median and 95% credible intervals) for the total number of inhabitants in each stratum. Please supply your code (as an Appendix) with your answers.