Homework 4

Type up answers in LaTeX and complete according to your personal schedule. Write pseudocode in LaTeX and embed (e.g., R, Python, Julia) code within the same document.

1. **Normal-normal** Consider the setup

$$\begin{aligned}\mathbf{y} = y_1, \dots, y_{100} &\overset{iid}{\sim} \text{Normal}(\mu, \sigma^2) \\ \mu &\sim \text{Normal}(\mu_0 = 0, \sigma_0^2 = 100) \\ \sigma^2 &\sim \text{InverseGamma}(\alpha = 2, \beta = 10)\,,\end{aligned}$$

   where the second and third lines denote prior specifications, and μ and σ^2 are assumed to be independent *a priori*.

   After simulating $y_1, \dots, y_{100}$ with $\mu = 50$ and $\sigma^2 = 10$ fixed, sample the joint posterior distribution $p(\mu, \sigma^2 | y_1, \dots, y_{100})$ using

   (a) Metropolis-Hastings with a bivariate Gaussian proposal distribution truncated along the dimension corresponding to σ^2; and

   (b) Hamiltonian Monte Carlo over the transformed parameter vector $(\mu, \log \sigma^2)$ using 100 leapfrog iterations at each step. Note that this will require a change of variables that must be taken into account in both the posterior distribution and its gradient.

   Tune both samplers (i.e., M-H proposal variance and HMC leapfrog stepsize) to average a 50% acceptance rate. Run both samplers to obtain a minimum effective sample size (ESS) of 10,000 for either parameter. What is the average ESS per sample for each sampler? Create 2D contour plots for each of the 2 posteriors.

2. **Hierarchically modeling house prices**

   (a) Start by simulating home prices in 5 fictional states. Set the true global mean to be $\mu_0 = 13.37$ and sample $j = 1, \dots, 5$ state house true price means according to the distribution $\mu_j \sim \text{N}(\mu_0, 4^2)$. Let each state share the same true $\sigma_j = \sqrt{5.35}$. Next, sample $n_j = 100$ house prices from each state according to the distribution $y_{1j}, \dots, y_{n_j j} \overset{iid}{\sim} \text{N}(\mu_j, \sigma_j^2)$.

(b) Next, define the hierarchical model

$$\begin{aligned} y_{ij} &\overset{ind}{\sim} \mathrm{N}(\mu_j, \sigma^2) \\ \mu_j &\overset{ind}{\sim} \mathrm{N}(\mu_0, 25^2) \\ \mu_0 &\sim \mathrm{N}(0, 50^2) \\ \sigma^2 &\sim \text{Inv-}\chi^2(1, 0.5) \end{aligned}$$

and build a Gibbs sampler that uses conditional conjugacy to generate closed-form updates for

i. each μ_j given μ_0, σ^2 and all y_{ij};

ii. μ_0 given all μ_j; and

iii. σ^2 given all μ_j and all y_{ij}.

(c) Simulate the Gibbs sampler for 10,000 iterations. Obtain 95% credible intervals for each model parameter by taking the 2.5th and 97.5th percentiles. Plot the 7 credible intervals along with the true values.

3. **Rat tumors** An experiment features $J = 71$ groups of rats, each of which receive different dosages of a certain treatment. Each group features a different number of rats n_j. Denote the number of rats in each group that develop a tumor $y_j \leq n_j$. The data are available here: https://ucla-biostats-202c.github.io/code/ratTumor.txt.

(a) Define the hierarchical model

$$\begin{aligned} y_j &\overset{ind}{\sim} \mathrm{Binomial}(n_j, \theta_j) \\ \theta_j &\overset{iid}{\sim} \mathrm{Beta}(2, \beta) \\ \beta &\sim \mathrm{Gamma}(2, 1) \end{aligned}$$

and build a Gibbs sampler that uses

i. conditional conjugacy to generate closed-form updates for each θ_j given n_j, y_j and β; and

ii. Metropolis-Hastings to update β conditioned on all 71 θ_js. Note that you should tune the β update step to obtain a roughly 50% acceptance rate. We sometimes refer to this scheme as "Metropolis-within-Gibbs".

(b) Simulate your Gibbs sampler for 100,000 iterations and

i. report the mean ESS for the θs as well as the ESS for β;

ii. create a figure containing

A. a 95% credible interval for each θ_js;
B. the posterior mean for each θ_j;
C. the maximum likelihood estimate of each θ_j, i.e., $\hat{\theta}_j = y_j/n_j$.

(c) Based on the relationship between the posterior means and the maximum likelihood estimates, guess the meaning of the phrase "Bayesian shrinkage".

4. **Snoring and heart disease** The following is one example of generalized linear regression within the Bayesian paradigm. We model the association between snoring and heart disease using the data in Table 1. For each individual i, we model the binary outcome $y_i \in \{0, 1\}$, where $y_i = 0$ denotes no heart disease and $y_i = 1$ denotes heart disease. Define our vector of covariates $\mathbf{x}_i = (1, x_i)^T$, with first element corresponding to the intercept and x_i corresponding to the snoring level of the ith individual. Letting $\eta_i = \mathbf{x}_i^T \boldsymbol{\beta}$ for $\boldsymbol{\beta} = (\beta_0, \beta_1)^T$ the corresponding vector of regression coefficients, we have the logistic-regression likelihood

$$\Pr(\mathbf{y}|\boldsymbol{\beta}) \propto \prod_i \left(\frac{e^{\eta_i}}{1 + e^{\eta_i}}\right)^{y_i} \left(\frac{1}{1 + e^{\eta_i}}\right)^{1-y_i},$$

and the log-likelihood

$$\begin{aligned} \log \Pr(\mathbf{y}|\boldsymbol{\beta}) &\propto \sum_i y_i \eta_i - \log\left(1 + e^{\eta_i}\right) \\ &= \sum_{j \in \{0,2,4,5\}} y_j \eta_j - n_j \log\left(1 + \eta_j\right), \end{aligned}$$

where y_j are the group heart disease totals and n_j are the group sizes. If we place independent normal priors on the coefficients

$$\beta_0,\ \beta_1 \overset{iid}{\sim} \mathrm{N}(0, 10^2),$$

then their joint log-posterior is

$$\log p(\boldsymbol{\beta}|\mathbf{y}) \propto -\frac{1}{200} \boldsymbol{\beta}^T \boldsymbol{\beta} + \sum_{j \in \{0,2,4,5\}} y_j \eta_j - n_j \log\left(1 + \eta_j\right).$$

(a) Use the Metropolis algorithm (tuned to a 50% acceptance rate) with bivariate Gaussian proposals to infer the joint posterior distribution of β_0 and β_1. How many MCMC iterations are required to achieve a minimum ESS of 1,000?

Snoring level (x_j)	Number of people with heart disease (y_j)	Total number surveyed (n_j)
0	24	1355
2	35	603
4	21	192
5	30	224

Table 1: Data from 2484 subjects (reported by spouses). Snoring level $j = 5$ is the most severe, and $j = 0$ means no snoring.

(b) What are the posterior means and 95% credible intervals for β_0 and β_1? How do we interpret them?

(c) Are the results statistically significant? Why or why not?