STAT 157/STAT 260 - Bayesian Statistics
Homework Three
Due on 30 October 2020 (by 11:59 pm)

Fall 2020, UC Berkeley

Aditya Guntuboyina

October 18, 2020

1 Instructions

1. I will only accept typed solutions.
2. You need to upload your homework solution as a single pdf file on bCourses by 11:59 pm on 30 October 2020.
3. Total points: 36 for students taking 157 and 44 for students taking 260

2 Questions

1. Describe the Gibbs sampling algorithm to output a large number N of samples from the joint density: (**4 points**)

$$\pi(x_1, x_2, x_3) \propto \exp\left(-2.25x_1^2 - 5x_2^2 - x_3^2 + 5x_1x_2 + 4x_2x_3 - 2x_1x_3\right).$$

2. This problem asks you to describe the Gibbs sampling algorithm for the *tobit* model. We observe data $(y_1, x_1), \dots, (y_n, x_n)$ with $y_1, \dots, y_n$ taking values in $[0, \infty)$ and $x_1, \dots, x_n \in \mathbb{R}^p$. $x_1, \dots, x_n$ are treated non-random and $y_1, \dots, y_n$ random. The parameter is $\beta \in \mathbb{R}^p$. The tobit regression model assumes that, conditional on β, the random variables $y_1, \dots, y_n$ are independent with

$$y_i = \max(0, w_i) \qquad \text{where } w_i \sim N(x_i'\beta, 1).$$

Take the improper uniform prior for β on $\mathbb{R}^p$. Describe the Gibbs sampling algorithm for generating a large number of samples from the posterior distribution of β given $y_1, \dots, y_n$. (**6 points**).

3. Consider the linear regression model where we observe data $(y_1, x_1), \dots, (y_n, x_n)$ where $y_1, \dots, y_n$ are real-valued and $x_1, \dots, x_n \in \mathbb{R}^p$. As always, we treat $x_1, \dots, x_n$ as non-random and $y_1, \dots, y_n$ are random. The parameter is $\theta \in \mathbb{R}^{p+1}$ with $\theta = (\beta, \sigma)$. Conditional on θ, we assume that

$$Y \sim N_n(X\beta, \sigma^2 I_n)$$

where Y is the $n \times 1$ vector with entries $y_1, \ldots, y_n$ and X is the $n \times p$ matrix with rows $x_1', \ldots, x_n'$. Also $N_k(\mu, \Sigma)$ denotes the k-variate normal distribution with mean vector μ and covariance matrix Σ. Consider the following prior for $\theta = (\beta, \sigma)$:

$$\beta | \sigma \sim N_p(\beta_0, \sigma^2 \Sigma_0) \quad \text{and} \quad \sigma^{-2} \sim \text{Gamma}(a, b)$$

for some values of β_0, Σ_0, a, b that are assumed to be available for this problem. Calculate the posterior distribution of β given the data $y_1, \ldots, y_n$. (**6 points**).

4. Consider the baseball dataset that we have studied in class. Use JAGS or NIMBLE to fit the hierarchical model:

$$X_i | \Theta_1, \ldots, \Theta_n, \mu, \tau \overset{\text{independent}}{\sim} \text{Bin}(n_i, \Theta_i) \text{ with } m_i \text{ known}, i = 1, \ldots, n$$

$$\log \frac{\Theta_1}{1 - \Theta_1}, \ldots, \log \frac{\Theta_n}{1 - \Theta_n} \Big| \mu, \tau \overset{\text{i.i.d}}{\sim} N(\mu, \tau^2)$$

$$\mu \sim N(\mu_0, \eta_0^2) \quad \text{and} \quad \tau^{-2} \sim \text{Gamma}(\alpha, \beta) \text{ are independent.}$$

Here X_i denotes the number of hits for the i^{th} player and m_i denotes the number of atbats for the i^{th} player (in this dataset $m_i = 45$ for each i). Put in some standard noninformative values for $\mu_0, \eta_0, \alpha, \beta$. Obtain point estimates of each Θ_i and compare them with the End-of-season averages. Compare your solution with that obtained from the analysis of this data in Lecture 14 where I used the arcsin transformation and worked with normal likelihoods. (**6 points**).

5. The following data represent the mortality rates from 12 English hospitals carrying out heart surgery on children under 1 year old between 1991 and 1995 (for example, in the Bristol hospital, there were 143 operations of which 41 resulted in deaths).

Hospital	Operations	Deaths
Bristol	143	41
Leicester	187	25
Leeds	323	24
Oxford	122	23
Guys	164	25
Liverpool	405	42
Southampton	239	24
Great Ormond St	482	53
Newcastle	195	26
Harefield	177	25
Birmingham	581	58
Brompton	301	31

Let m_i and X_i denote the number of operations and deaths for the i^{th} hospital. Assume that the $m_i's$ are non-random. Model X_i as $\text{Bin}(m_i, \Theta_i)$ independently across i. First obtain point estimates and 95% confidence intervals for Θ_i using only X_i and m_i. Then, using JAGS or NIMBLE, fit the hierarchical model of the previous problem with noninformative values of $\mu_0, \eta_0, \alpha, \beta$. Use the drawn MCMC samples to form point estimates and 95% credible intervals for $\Theta_1, \ldots, \Theta_n$. Compare these estimates and intervals with the intervals obtained previously (using only X_i and m_i). What differences do you see between the two sets of intervals? (**6 points**).

6. Consider the following data on the math scores for a sample of 10th grade students from two US public high schools:

School One (31 students): 52.11, 57.65, 66.44, 44.68, 40.57, 35.04, 50.71, 66.17, 39.43, 46.17, 58.76, 47.97, 39.18, 64.63, 69.38, 32.38, 29.98, 59.32, 43.04, 57.83, 46.07, 47.74, 48.66, 40.80, 66.32, 53.70, 52.42, 71.38, 59.66, 47.52, 39.51

School Two (22 students): 57.57, 42.40, 41.41, 55.22, 43.90, 53.04, 49.00, 62.45, 53.78, 49.08, 40.25, 43.08, 52.43, 21.73, 53.68, 41.45, 45.47, 34.06, 33.45, 60.78, 35.92, 52.40

To this data, fit the model

$$\begin{aligned} Y_{i1} &= \mu + \delta + \epsilon_{i1} \\ Y_{i2} &= \mu - \delta + \epsilon_{i2} \\ \{\epsilon_{ij}\} &\overset{\text{i.i.d}}{\sim} N(0, \sigma^2). \end{aligned}$$

with parameters μ, δ, σ^2 along with the prior:

$$\mu, \delta, \sigma \text{ independent with } \mu \sim N(50, 625), \delta \sim N(0, 625) \text{ and } \sigma^{-2} \sim \text{Gamma}(0.5, 50).$$

Use the Gibbs sampler to draw posterior samples for the parameters μ, δ, σ. Implement the Gibbs sampler and check the correctness of your implementation with the results of JAGS or NIMBLE.

Use your samples to calculate the posterior probability that $\delta > 0$. Also calculate the probability that a random sampled tenth grader from the first school will have a higher math score than one sampled from the second school. (**8 points**).

3 Additional Question for 260 Students

1. Consider the problem of estimating the parameters μ and σ of the univariate $t_v(\mu, \sigma^2)$ distribution. Recall that this distribution has the density:

$$x \mapsto \frac{\Gamma((v+1)/2)}{\sqrt{v\pi}\Gamma(v/2)\sigma} \left(1 + \frac{1}{v}\left(\frac{x-\mu}{\sigma}\right)^2\right)^{-(v+1)/2}.$$

Suppose we observe $X_1, \ldots, X_n$ that (conditional on μ, σ) are i.i.d having the $t_v(\mu, \sigma^2)$ distribution with a known v. Suppose we have the noninformative prior distribution:

$$\pi(\mu, \sigma) \propto \frac{1}{\sigma} I\{\sigma > 0\}$$

on μ, σ. Describe a Gibbs sampling algorithm for obtaining samples from the posterior distribution of μ, σ. For designing an appropriate data augmentation scheme for the Gibbs sampler, you might find the following fact helpful:

$$X|\tau \sim N(a, \tau^2) \text{ and } \tau^{-2} \sim \text{Gamma}(v/2, b^2/2) \implies X \sim t_v(a, b^2).$$

(**8 points**).