STAT 207, Spring 2026
Homework # 3
Released Apr 30. Due May 13 on Canvas

Instructions: You may discuss the homework problems in small groups, but you must write up the final solutions and codes yourself. You can either type your work in LaTeX or write down on papers and scan. Please make sure that all handwriting are visible and please combine all pages into a single PDF file (in the correct order).

For any problems that involve coding, you must provide written answers and also include your codes. You can either include your codes at the end of the homework and label which questions they correspond to, or include as part of the answer (e.g., in the R Markdown style). You will receive no credit if you submit only codes or only written answers.

1. Let $\boldsymbol{Y}$ be a $n \times p$ data matrix. Let O_{ij} denote if Y_{ij} is missing, i.e., $Y_{ij} = \texttt{NA}$ if $O_{ij} = 0$. Consider the model for the complete data $\boldsymbol{Y}$ that

$$\boldsymbol{Y}_i \sim N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$$

with priors $\boldsymbol{\mu} \sim N(\boldsymbol{\mu}_0, \boldsymbol{\Lambda}_0)$ and $\boldsymbol{\Sigma} \sim InvWishart(v, \boldsymbol{I}_p)$.

   (a) Describe how to implement a Gibbs sampler to make inference about the parameters of the model. *Hint: conditional distribution of an element in a multivariate normal is univariate normal.*

   (b) If we also know that

$$p(O_{ij} = 0 | Y_{ij}) = \begin{cases} 0.5 & Y_{ij} > 0 \\ 0 & Y_{ij} \leqslant 0 \end{cases}$$

   How would you change your sampler above?

2. Consider a scale that only shows weights at fixed increments. That is, for an object with true weight y, it shows that the object is within the interval $(x_{j-1}, x_j]$ if $x_{j-1} < y \leqslant x_j$. The $(x_0, ..., x_K)$ points are fixed with $x_0 = 0, x_K = \infty$. Consider n objects with weights $y_1, ..., y_n$ follow independent Half-Normal(σ^2) distribution, i.e., truncated normal with mean 0 on the support of $(0, \infty)$. The scale shows that n_j objects have weights in interval $(x_{j-1}, x_j]$, with $n_1 + n_2 + ..., +n_K = n$. Assume that the prior distribution is $p(\sigma^2) \propto 1/\sigma^2$.

   (a) Write explicitly the complete data likelihood, the observed data likelihood, and the posterior distribution of σ^2.

   (b) Write the full conditionals required to obtain samples of σ^2, and the latent variables using a Gibbs sampler.

3. Consider a random sample of size $n + k$ of random variables following a Poisson distribution with mean λ. Suppose any values **below** C are censored. Denote the first n observations to

be observed and the next k observations are censored. That is, if we denote the sample as $x_1, ..., x_n, x^*_{n+1}, ..., x^*_{n+k}$, then the recorded sample $x^*_i = C$ if the true value is **less than or equal to** C.

(a) Assume that the prior is given as $\lambda \sim Ga(\alpha, \beta)$. Use the missing data formulation to obtain the posterior distribution $p(\lambda|X, X^*)$.

(b) In order to facilitate the estimation of λ using MCMC, introduce latent variables that indicate the censored observationsand write the explicit steps for sampling λ and the latent variables.

(c) Suppose we observe J batches of data, i.e.,

$$x_{j,1}, ..., x_{j,n_j}, x^*_{j,n_j+1}, ..., x^*_{j,n_j+k_j}$$

for $j = 1, ..., J$ with the same censoring point C. Each batch of the data correspond to a Poisson distribution with mean λ_j. Consider a hierarchical model assuming $\lambda_j \sim_{iid} Gamma(\alpha, \beta)$. Introduce appropriate latent variables and find conditional distributions of all model parameters.

(d) Suppose now that the censoring point is unknown and different in each batch of the data, i.e., $x^*_{j,\cdot}$ will not be recorded if the true value is below an unknown censoring point C_j. So we need to treat $C_1, ..., C_J$ as random variables. Introduce appropriate priors for C_j and find conditional distributions of all model parameters.

4. In this problem, we consider a different formulation of the hierarchical normal problem. Consider observations from $i = 1, ..., m$ groups

$$y_{ij}|\theta_i, \sigma^2 \sim N(\theta_i, \sigma^2), \quad i = 1, ..., m, j = 1, ..., n_i$$

Consider the same normal prior for θ_i:

$$\theta_i|\tau \sim N(\mu, \tau^2), \quad i = 1, ..., m$$

We use the uniform prior on μ and a half-Cauchy prior on τ, i.e., $p(\mu) \propto 1$ and $\tau \sim C^+(0, 1)$, where the half-Cauchy distribution $X \sim C^+(0, 1)$ is defined by

$$f(x) = \frac{2}{\pi(1 + x^2)} 1_{x>0}$$

Finally, we complete the model with the prior on σ^2 to be $p(\sigma^2) \propto \sigma^{-2}$.

(a) Write down the posterior conditional distributions for $(\mu, \theta_i, \tau, \sigma)$. If the distribution is not in the standard form, just write down the unnormalized distribution.

(b) In order to sample τ more easily, we will use a data augmentation approach by introducing auxiliary variables. As the first step, derive and prove the following statement: if $U \sim InvGamma(1/2, 1)$ and $X^2 \sim InvGamma(1/2, 1/U)$, then $X \sim C^+(0, 1)$.

(c) Using the identity proved in the previous part, we can write the prior for θ_i to be

$$\begin{aligned} \theta_i|\tau &\sim N(\mu, \tau^2) \\ \tau^2 &\sim InvGamma(1/2, 1/\phi) \\ \phi &\sim InvGamma(1/2, 1) \end{aligned}$$

Write down how you modify the posterior conditional distributions in part (a) to construct a Gibbs sampler.

5. Consider the normal mean model for estimated causal effect of an education intervention on test scores of students in m schools. Let x_{ij} denote the change of GPA after attending the program, for the j-th student from the i-th school. Assume there are n students from each school, and the same data likelihood as in the previous problem,

$$y_{ij} = \theta_i + \epsilon_{ij}, \quad \epsilon_{ij} \sim N(0, \sigma^2), \quad i = 1, ..., m, \;\; j = 1, ..., n$$

Assume the prior for θ_i is given as

$$p(\theta_i|\pi, \tau_0, \tau_1) = \pi\psi(\theta_i; \tau_1) + (1-\pi)\psi(\theta_i; \tau_0)$$

for given constants $0 < \tau_1 \ll \tau_0$ and the θ_i's are independent a priori. $\psi(\theta_i; \tau)$ denotes the double exponential density

$$\psi(\theta; \tau) = \frac{\tau}{2} e^{-\tau|\theta|}.$$

and prior for π is

$$\pi \sim \text{Beta}(a, b)$$

(a) Consider a fixed π, a small $\tau_1 \approx 0$ and a large τ_0, give an interpretation of the model formulation.

(b) A double exponential density can be written as a scale mixture of normals due to the following identity:

$$\frac{a}{2} e^{-a|z|} = \int_0^\infty \frac{e^{-z^2/(2s^2)}}{\sqrt{2\pi s^2}} \frac{a^2}{2} e^{-a^2 s^2/2} ds^2$$

where the right hand side is an integral over a product of $N(z; 0, s^2)$ and $Exp(s^2; a^2/2)$ densities. Using the above identify to introduce latent variables that provide a hierarchical formulation of the prior for θ_i.

(c) Introduce indicator variables for the mixture components to obtain a multiplicative representation of the mixture that is computationally efficient.

(d) Consider fixed τ_0 and τ_1, and let $p(\sigma^2) \propto 1/\sigma$. Obtain and identify the full conditionals for all model parameters and latent variables that are needed to sample from the joint posterior distribution using a Gibbs sampler. *(Depending on your parameterization, there may be posterior conditionals that do not seem to take standard distribution forms. You can leave them up to a proportional constant in your answer.)*

Optional practice problems. You do not need to turn in these problems, but they are good practice on the topic of mixture model that is not covered here: 22.2, 22.4, 22.5