Bayesian Analysis I, FALL 2024

Lab Exercise 6: on **Monday Dec. 2, 12-14**

1. Two analysts in the same laboratory made repeated determinations of the percentage of fiber in soya cotton cake, the results being as shown below:

 Analyst A: 12.38 12.53 12.25 12.37 12.48 12.58 12.43 12.43 12.30

 Analyst B: 12.25 12.45 12.31 12.31 12.30 12.20 12.25 12.25 12.26 12.42 12.17 12.09

Investigate the mean discrepancy θ between their mean determinations and in particular give an interval in which you are 90% sure that it lies (i.e. 90% HPI):

 a) Assuming that it is known from past experience that the standard deviation of both sets of observations is 0.1. Use a noninformative prior for μ_i.

 b) Assuming simply that it is known that the standard deviation of the two sets of observations are equal. Take a NIC $(-2, 0, \boldsymbol{m}, \mathbf{0}^{-1})$ prior.

2. Refer to the lecture slides (p 15-16) for Normal sample with two unknown parameters. Consider using the joint prior given by $p(\mu, \tau) = p(\mu|\tau)p(\tau)$ where

$$\mu|\tau \sim Normal(\delta, \tau/s) \quad \Rightarrow p(\mu|\tau) \propto (\tau/s)^{-1/2} e^{-\frac{s}{2\tau}(\mu-\delta)^2}$$

$$\tau \sim \text{IC}(a, b) \quad \Rightarrow p(\tau) \propto \tau^{-(a/2+1)} e^{-\frac{b}{2\tau}}$$

Derive the posterior distributions and show how those parameters $(\delta^*, s^*, a^*, b^*)$ are found: $\mu|\tau, \boldsymbol{X} \sim Normal(\delta^*, \tau/s^*)$ $\quad \tau|\boldsymbol{X} \sim \text{IC}(a^*, b^*)$ where $s^* = s + n, \delta^* = \frac{\delta s + n\overline{x}}{s+n}, a^* = a + n, b^* = b + S + \frac{sn}{s+n}(\overline{x} - \delta)^2$.

(Note: these results correspond to main posterior results with the NIC prior distribution.)

3. Suppose that $\boldsymbol{Y}|\boldsymbol{\theta}$ has an n dimensional multivariate normal distribution with mean $A\boldsymbol{\theta}$ and a known dispersion matrix V. Its density is

$$p(\boldsymbol{y}|\boldsymbol{\theta}) = (2\pi)^{-n/2}|V|^{-1/2} \exp\left(-\frac{1}{2}(\boldsymbol{y} - A\boldsymbol{\theta})^T V^{-1}(\boldsymbol{y} - A\boldsymbol{\theta})\right)$$

where A is known. We assume a locally uniform prior for $\boldsymbol{\theta}$, $p(\boldsymbol{\theta}) \approx$ constant. Using the Bayes theorem, derive the posterior distribution of $\boldsymbol{\theta}|\boldsymbol{Y}$.

4. a) Consider a multivariate Normal data with

$$\begin{aligned} y_1|\theta,\gamma &\sim N(\theta+\gamma,1) \\ y_2|\theta,\gamma &\sim N(\theta-\gamma,1) \end{aligned}$$

where y_1 and y_2 are independent and the prior for (θ,γ) is given by

$$\begin{pmatrix}\theta\\ \gamma\end{pmatrix} \sim N\left(\begin{pmatrix}0\\ m\end{pmatrix}, \begin{pmatrix}3 & 0\\ 0 & 3\end{pmatrix}\right)$$

Obtain the posterior distribution θ and γ using the theorem by Lindley and Smith.

b) A three stage linear model is given by

$$\begin{aligned} \boldsymbol{Y}|\boldsymbol{\theta_1} &\sim N(A_1\boldsymbol{\theta_1}, C_1) \\ \boldsymbol{\theta_1}|\boldsymbol{\theta_2} &\sim N(A_2\boldsymbol{\theta_2}, C_2) \\ \boldsymbol{\theta_2} &\sim N(\boldsymbol{\mu}, C_3) \end{aligned}$$

where $A_1, A_2, C_1, C_2, C_3, \boldsymbol{\mu}$ are known.

Using the theorem by Lindley and Smith, show that the posterior distribution of $\boldsymbol{\theta_1}|\boldsymbol{Y}$ is N(Dd, D) where

$$\begin{aligned} D^{-1} &= A_1^T C_1^{-1} A_1 + (C_2 + A_2 C_3 A_2^T)^{-1} \\ d &= A_1^T C_1^{-1}\boldsymbol{Y} + (C_2 + A_2 C_3 A_2^T)^{-1} A_2 \boldsymbol{\mu} \end{aligned}$$

Hint: Use the L-S theorem to second and third stages of the above model to get the marginal distribution of $\boldsymbol{\theta_1}$ and apply the theorem again.

5. In SIFO's monthly opinion poll for December, 1631 persons answered which political party they would vote for if there was an election now. Of these, 40.8% said they would vote for the government alliance, 52.2% said that they would vote for the opposition and the remaining 7.0% said that they would vote for another party.

Let $\boldsymbol{\theta} = (\theta_1, \theta_2, \theta_3)$ denote the probabilities of the three alternatives and $\boldsymbol{y} = (y_1, y_2, y_3)$ the observed counts.

a) Use the non-informative prior $p(\boldsymbol{\theta}) = 1$ and calculate the posterior distribution of $\boldsymbol{\theta}$.

b) Calculate the posterior mean and variance of $\theta_1 - \theta_2$ the difference between the two political blocks. (Refer to the formulas for mean and variance given during the lecture).

c) Give the 95% posterior credible interval for $\theta_1 - \theta_2$ by approximating the posterior distribution of $\theta_1 - \theta_2$ using a normal distribution.

6. Consider the hierarchical Poisson model

$$\begin{aligned} Y_j|\theta_j &\sim \text{Poiss}(\theta_j) \quad j = 1, 2, ..., J \\ \theta_j|\alpha, \beta &\sim \text{Gamma}(\alpha, \beta) \\ p(\alpha, \beta) &\propto \frac{1}{\alpha\beta}. \end{aligned}$$

a) Calculate the joint posterior distribution $p(\boldsymbol{\theta}, \alpha, \beta|Y)$, where $Y = (Y_1, ..., Y_n)$ (up to a proportionality constant).

b) Calculate the marginal conditional density $p(\boldsymbol{\theta}|\alpha, \beta, Y)$.

c) Calculate the marginal posterior density $p(\alpha, \beta|Y)$ (up to a proportionality constant).