

Statistical Inference 2 Assignment 2

Master of Statistics (M.Stat.) Year II, Session 2020-21

Name: Subhrajyoty Roy
Roll: MB1911

January 12, 2021

Problem 1. Consider the linear regression model $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$ where $\mathbf{y} = (y_1, \dots, y_n)'$ is the vector of observations on the “dependent” variable, $\mathbf{X} = ((x_{ij}))_{n \times p}$ is of full rank, x_{ij} being the values of the nonstochastic regressor variables, $\boldsymbol{\beta} = (\beta_1, \dots, \beta_p)'$ is the vector of regression coefficients and the components of $\boldsymbol{\epsilon}$ are independent, each following $\mathcal{N}(0, \sigma^2)$. Consider the noninformative prior $\pi(\boldsymbol{\beta}, \sigma^2) \propto \sigma^{-2}, \boldsymbol{\beta} \in \mathbb{R}^p, \sigma^2 > 0$.

- (a) Find the $100(1 - \alpha)\%$ HPD credible set for $\boldsymbol{\beta}$.
- (b) Find the marginal posterior distribution of a particular β_j ($j = 1, \dots, p$) and use it to find the $100(1 - \alpha)\%$ HPD credible set for β_j .

Solution. (a) Since, $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$ and each component of $\boldsymbol{\epsilon}$ are independent, following $\mathcal{N}(0, \sigma^2)$, clearly \mathbf{y} has a n -variate multivariate normal distribution with mean parameter $\mathbf{X}\boldsymbol{\beta}$ and dispersion matrix $\sigma^2 \mathbf{I}_n$, where \mathbf{I}_n is the identity matrix of order n .

We start by writing the joint posterior of $(\boldsymbol{\beta}, \sigma^2)$ as,

$$\begin{aligned} \pi(\boldsymbol{\beta}, \sigma^2 \mid \mathbf{X}, \mathbf{y}) &\propto \pi(\boldsymbol{\beta}, \sigma^2) \mathcal{L}(\mathbf{y} \mid \mathbf{X}, \boldsymbol{\beta}, \sigma^2), \quad \text{since, } \mathbf{X} \text{ is nonstochastic} \\ &\propto \frac{1}{\sigma^2} \times \frac{1}{\det(\sigma^2 \mathbf{I}_n)^{1/2}} e^{-\frac{1}{2\sigma^2} (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})^\top (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})} \\ &= \sigma^{-(n+2)} \exp \left[-\frac{1}{2\sigma^2} (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})^\top (\mathbf{y} - \mathbf{X}\boldsymbol{\beta}) \right] \end{aligned}$$

Denoting $\hat{\boldsymbol{\beta}}$ as the ordinary least square regression estimate of $\boldsymbol{\beta}$, the quantity $(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})^\top (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})$ can be decomposed as,

$$\begin{aligned} (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})^\top (\mathbf{y} - \mathbf{X}\boldsymbol{\beta}) &= (\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}} + \mathbf{X}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}))^\top (\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}} + \mathbf{X}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})) \\ &= (\mathbf{y} - \hat{\mathbf{y}})^\top (\mathbf{y} - \hat{\mathbf{y}}) + 2(\mathbf{y} - \hat{\mathbf{y}})^\top \mathbf{X}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}) + (\boldsymbol{\beta} - \hat{\boldsymbol{\beta}})^\top \mathbf{X}^\top \mathbf{X}(\boldsymbol{\beta} - \hat{\boldsymbol{\beta}}) \\ &\quad \text{where, } \hat{\mathbf{y}} = \mathbf{X}\hat{\boldsymbol{\beta}} \\ &= (\mathbf{y} - \hat{\mathbf{y}})^\top (\mathbf{y} - \hat{\mathbf{y}}) + (\boldsymbol{\beta} - \hat{\boldsymbol{\beta}})^\top \mathbf{X}^\top \mathbf{X}(\boldsymbol{\beta} - \hat{\boldsymbol{\beta}}) \end{aligned}$$

since, the residual $(\mathbf{y} - \hat{\mathbf{y}})$ belongs to the orthogonal complement of the column space of \mathbf{X} . Denoting the first term $SSE = (\mathbf{y} - \hat{\mathbf{y}})^\top (\mathbf{y} - \hat{\mathbf{y}})$, we have

$$\pi(\boldsymbol{\beta}, \sigma^2 \mid \mathbf{X}, \mathbf{y}) \propto \sigma^{-(n+2)} \exp \left[-\frac{1}{2\sigma^2} (SSE + (\boldsymbol{\beta} - \hat{\boldsymbol{\beta}})^\top \mathbf{X}^\top \mathbf{X}(\boldsymbol{\beta} - \hat{\boldsymbol{\beta}})) \right]$$

Therefore, the posterior distribution of $\boldsymbol{\beta}$ can be obtained as,

$$\begin{aligned}
\pi(\boldsymbol{\beta}, | \mathbf{X}, \mathbf{y}) &= \int_0^\infty \pi(\boldsymbol{\beta}, \sigma^2 | \mathbf{X}, \mathbf{y}) d(\sigma^2) \\
&\propto \int_0^\infty \sigma^{-(n+2)} \exp \left[-\frac{1}{2\sigma^2} (SSE + (\boldsymbol{\beta} - \hat{\boldsymbol{\beta}})^\top \mathbf{X}^\top \mathbf{X} (\boldsymbol{\beta} - \hat{\boldsymbol{\beta}})) \right] d(\sigma^2) \\
&\propto \left[SSE + (\boldsymbol{\beta} - \hat{\boldsymbol{\beta}})^\top \mathbf{X}^\top \mathbf{X} (\boldsymbol{\beta} - \hat{\boldsymbol{\beta}}) \right]^{-n/2} \\
&\propto \left[1 + \frac{(\boldsymbol{\beta} - \hat{\boldsymbol{\beta}})^\top \mathbf{X}^\top \mathbf{X} (\boldsymbol{\beta} - \hat{\boldsymbol{\beta}})}{SSE} \right]^{-n/2} \\
&= \left[1 + \frac{(\boldsymbol{\beta} - \hat{\boldsymbol{\beta}})^\top \mathbf{X}^\top \mathbf{X} (\boldsymbol{\beta} - \hat{\boldsymbol{\beta}})}{(n-p)s^2} \right]^{-((n-p)+p)/2}, \text{ where } s^2 = \frac{(\mathbf{y} - \hat{\mathbf{y}})^\top (\mathbf{y} - \hat{\mathbf{y}})}{(n-p)}
\end{aligned}$$

Denoting $\Sigma_X = \mathbf{X}^\top \mathbf{X}$, we have the posterior distribution of $\boldsymbol{\beta}$ is a multivariate t distribution with $(n-p)$ d.f. (p is the number of predictors), location parameter $\hat{\boldsymbol{\beta}}$ (the OLS estimate of $\boldsymbol{\beta}$) and the scale or dispersion parameter $\Sigma_X^{-1/2}s$. In other words, the quantity

$$F(\boldsymbol{\beta}) = \frac{\frac{(\boldsymbol{\beta} - \hat{\boldsymbol{\beta}})^\top \mathbf{X}^\top \mathbf{X} (\boldsymbol{\beta} - \hat{\boldsymbol{\beta}})}{p}}{\frac{(n-p)s^2}{(n-p)}} = \frac{(\boldsymbol{\beta} - \hat{\boldsymbol{\beta}})^\top \mathbf{X}^\top \mathbf{X} (\boldsymbol{\beta} - \hat{\boldsymbol{\beta}})}{ps^2}$$

has a $F_{p, (n-p)}$ distribution. Therefore, an obvious $100(1-\alpha)\%$ HPD credible set can be given based on the ellipsoid

$$\{\boldsymbol{\beta} : F(\boldsymbol{\beta}) \leq F_{p, n-p, \alpha}\} = \left\{ \boldsymbol{\beta} : (\boldsymbol{\beta} - \hat{\boldsymbol{\beta}})^\top \mathbf{X}^\top \mathbf{X} (\boldsymbol{\beta} - \hat{\boldsymbol{\beta}}) \leq ps^2 F_{p, n-p, \alpha} \right\}$$

where $F_{p, n-p, \alpha}$ is the $(1-\alpha)$ -th quantile of the $F_{p, n-p}$ distribution. (The HPD property follows from the unimodality and symmetricity of the multivariate t distribution).

- (b) For this, we shall use the following result that if $\mathbf{x}_{n \times 1}$ is a random variable following a multivariate t distribution of r d.f. with location parameter $\boldsymbol{\mu}$ and dispersion matrix Σ , then a linear combination of its components, i.e. for any nonstochastic vector $\mathbf{a}_{n \times 1}$, the random variable $y = \mathbf{a}^\top \mathbf{x}$ has a univariate t distribution of r d.f., with location parameter $\mathbf{a}^\top \boldsymbol{\mu}$ and scale parameter $\mathbf{a}^\top \Sigma \mathbf{a}$.

For $j = 1, 2, \dots, p$, let, \mathbf{e}_j be the column vector of size p such that only its j -th coordinate is equal to 1 and all other coordinates are equal to 0. Using this result for the posterior distribution of $\boldsymbol{\beta}$, we obtain $\beta_j = \mathbf{e}_j^\top \boldsymbol{\beta}$, has the posterior distribution as a univariate t distribution with $(n-p)$ d.f., but with the location parameter $\mathbf{e}_j^\top \hat{\boldsymbol{\beta}} = \hat{\beta}_j$ (which is the j -th coordinate of the OLS estimate $\hat{\boldsymbol{\beta}}$), and the scale parameter,

$$\mathbf{e}_j^\top (\Sigma_X^{-1/2}s) \mathbf{e}_j = s\sqrt{d_{jj}}$$

where d_{jj} is the diagonal element of the matrix $(\mathbf{X}^\top \mathbf{X})^{-1}$. Therefore, we have,

$$\frac{\beta_j - \widehat{\beta}_j}{s\sqrt{d_{jj}}} \mid \mathbf{X}, \mathbf{y} \sim t_{n-p}$$

Again using the unimodality and symmetricity of the t distribution, it follows that a $100(1-\alpha)\%$ HPD credible set for β_j would be given by

$$\left[\widehat{\beta}_j - t_{n-p, \alpha/2} \sqrt{d_{jj}} s, \widehat{\beta}_j + t_{n-p, \alpha/2} \sqrt{d_{jj}} s \right]$$

where $t_{n-p, \alpha/2}$ is the $(1 - \alpha/2)$ -th quantile of t_{n-p} distribution.

Problem 2. Let, X_1, \dots, X_n be i.i.d $\mathcal{N}(\mu, \sigma^2)$ variables where μ and σ^2 are both unknown. Consider the prior $\pi(\mu, \sigma^2) \propto 1/\sigma^2$. Show that the posterior predictive distribution of a future observation X_{n+1} is a t distribution with $n-1$ d.f., location \bar{X}_n and scale $(1+1/n)^{1/2}s$ where $s^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X}_n)^2$.

Solution. We compute the predictive distribution of $X_{(n+1)}$ conditional on the observed data X_1, \dots, X_n . To derive this, note that

$$\begin{aligned}
& f(X_{n+1} \mid X_1, \dots, X_n) \\
&= \int_0^\infty \int_{-\infty}^\infty f(X_{n+1} \mid \mu, \sigma^2, X_1, \dots, X_n) \pi(\mu, \sigma^2 \mid X_1, \dots, X_n) d\mu d\sigma^2 \\
&= \int_0^\infty \int_{-\infty}^\infty f(X_{n+1} \mid \mu, \sigma^2) \pi(\mu, \sigma^2 \mid X_1, \dots, X_n) d\mu d\sigma^2, \text{ since, } X_1, \dots, X_{n+1} \text{ are independent} \\
&\propto \int_0^\infty \int_{-\infty}^\infty f(X_{n+1} \mid \mu, \sigma^2) \pi(\mu, \sigma^2) f(X_1, \dots, X_n \mid \mu, \sigma^2) d\mu d\sigma^2 \\
&\propto \int_0^\infty \int_{-\infty}^\infty \frac{1}{\sigma} e^{-\frac{1}{2\sigma^2}(X_{n+1}-\mu)^2} \times \frac{1}{\sigma^2} \times \frac{1}{\sigma^n} \exp \left[-\frac{1}{2\sigma^2} \sum_{i=1}^n (X_i - \mu)^2 \right] d\mu d\sigma^2 \\
&= \int_0^\infty \int_{-\infty}^\infty \frac{1}{\sigma^{(n+3)}} \exp \left[-\frac{1}{2\sigma^2} \sum_{i=1}^{n+1} (X_i - \mu)^2 \right] d\mu d\sigma^2 \\
&= \int_0^\infty \frac{1}{\sigma^{(n+3)}} \exp \left[-\frac{\sum_{i=1}^{(n+1)} X_i^2}{2\sigma^2} \right] \int_{-\infty}^\infty \exp \left[-\frac{1}{2\sigma^2} \{ (n+1)\mu^2 - 2(n+1)\bar{X}_{(n+1)} \} \right] d\mu d\sigma^2 \\
&= \int_0^\infty \frac{1}{\sigma^{(n+3)}} \exp \left[-\frac{\sum_{i=1}^{(n+1)} X_i^2}{2\sigma^2} \right] \frac{\sqrt{2\pi}\sigma}{\sqrt{(n+1)}} \exp \left[-\frac{(n+1)}{2\sigma^2} \bar{X}_{(n+1)}^2 \right] d\sigma^2 \\
&\propto \int_0^\infty (\sigma^2)^{-(n+2)/2} \exp \left[-\frac{1}{2\sigma^2} \left\{ \sum_{i=1}^{(n+1)} X_i^2 - (n+1)\bar{X}_{(n+1)}^2 \right\} \right] d\sigma^2 \\
&\propto \left[\sum_{i=1}^{(n+1)} X_i^2 - (n+1)\bar{X}_{(n+1)}^2 \right]^{-n/2} \\
&= \left[X_{(n+1)}^2 + \sum_{i=1}^n X_i^2 - \frac{(n\bar{X}_n + X_{(n+1)})^2}{(n+1)} \right]^{-n/2} \\
&= \left[X_{(n+1)}^2 \frac{n}{(n+1)} - 2X_{(n+1)} \frac{n\bar{X}_n}{(n+1)} + \sum_{i=1}^n X_i^2 - \frac{n^2\bar{X}_n^2}{(n+1)} \right]^{-n/2} \\
&\propto \left[X_{(n+1)}^2 - 2X_{(n+1)}\bar{X}_n + (1+1/n) \sum_{i=1}^n X_i^2 - n\bar{X}_n^2 \right]^{-n/2} \\
&= \left[(X_{(n+1)} - \bar{X}_n)^2 + (1+1/n)(n-1)s^2 \right]^{-n/2}
\end{aligned}$$

$$\propto \left[1 + \frac{(X_{(n+1)} - \bar{X}_n)^2}{(1 + 1/n)(n-1)s^2} \right]^{-n/2}$$

This shows that the predictive distribution of X_{n+1} is a t distribution with $(n-1)$ d.f. having the location parameter \bar{X}_n and the scale parameter $\left(1 + \frac{1}{2}\right)^{1/2} s$, where $(n-1)s^2 = \sum_{i=1}^n (X_i - \bar{X}_n)^2 = \sum_{i=1}^n X_i^2 - n\bar{X}_n^2$.

Problem 3. Consider the setup of the theorem on asymptotic normality of posterior distribution (Theorem 4.2 of the book by Ghosh et al. (2006)), proved in the class for a proper prior.

- (a) Suppose that the prior is improper but there is an n_0 such that the posterior distribution of θ given x_1, \dots, x_{n_0} is proper for a.e. (x_1, \dots, x_{n_0}) . Show that the theorem hold also in this case.
- (b) In addition to the assumptions of Theorem 4.2, assume that the prior density $\pi(\theta)$ has a finite expectation. Proceeding as in the proof of Theorem 4.2, and using the assumption of finite expectation for π , show that

$$\int_{\mathbb{R}} |t| \left| \pi_n^*(t \mid X_1, \dots, X_n) - \frac{\sqrt{I(\theta_0)}}{\sqrt{2\pi}} e^{-\frac{1}{2}t^2 I(\theta_0)} \right| dt \rightarrow 0$$

with P_{θ_0} -probability one.

Solution. (a) Let π denotes the generic symbol for prior and posterior densities and let, $n > n_0$. Let, $\hat{\theta}_n^*$ be a strongly consistent maximum likelihood estimate of θ based on the samples $X_{n_0+1}, X_{n_0+2}, \dots, X_n$ and let $\hat{\theta}_n$ be a strongly consistent MLE of θ based on X_1, \dots, X_n .

Then, the posterior distribution conditional on the observations X_1, X_2, \dots, X_n is,

$$\begin{aligned} \pi_n(t^* \mid X_1, \dots, X_n) &= \pi_n(\hat{\theta}_n + \sqrt{n}t^* \mid X_1, \dots, X_n) \\ &= \frac{\prod_{i=1}^n f(X_i \mid \hat{\theta}_n + \sqrt{n}t^*) \pi(\hat{\theta}_n + \sqrt{n}t^*)}{\int \prod_{i=1}^n f(X_i \mid \hat{\theta}_n + \sqrt{n}t) \pi(\hat{\theta}_n + \sqrt{n}t) dt} \\ &= \frac{\prod_{i=1}^{n_0} f(X_i \mid \hat{\theta}_n + \sqrt{n}t^*) \prod_{i=n_0+1}^n f(X_i \mid \hat{\theta}_n + \sqrt{n}t^*) \pi(\hat{\theta}_n + \sqrt{n}t^*)}{\int \prod_{i=1}^{n_0} f(X_i \mid \hat{\theta}_n + \sqrt{n}t) \prod_{i=n_0+1}^n f(X_i \mid \hat{\theta}_n + \sqrt{n}t) \pi(\hat{\theta}_n + \sqrt{n}t) dt} \\ &= \frac{\prod_{i=n_0+1}^n f(X_i \mid \hat{\theta}_n + \sqrt{n}t^*) \frac{\prod_{i=1}^{n_0} f(X_i \mid \hat{\theta}_n + \sqrt{n}t^*) \pi(\hat{\theta}_n + \sqrt{n}t^*)}{\int \prod_{i=1}^{n_0} f(X_i \mid \hat{\theta}_n + \sqrt{n}u) \pi(\hat{\theta}_n + \sqrt{n}u) du}}{\int \prod_{i=1}^n f(X_i \mid \hat{\theta}_n + \sqrt{n}t) \frac{\prod_{i=1}^{n_0} f(X_i \mid \hat{\theta}_n + \sqrt{n}t) \pi(\hat{\theta}_n + \sqrt{n}t)}{\int \prod_{i=1}^{n_0} f(X_i \mid \hat{\theta}_n + \sqrt{n}u) \pi(\hat{\theta}_n + \sqrt{n}u) du} dt} \\ &= \frac{\prod_{i=n_0+1}^n f(X_i \mid \hat{\theta}_n + \sqrt{n}t^*) \pi(\hat{\theta}_n + \sqrt{n}t^* \mid X_1, \dots, X_{n_0})}{\int \prod_{i=n_0+1}^n f(X_i \mid \hat{\theta}_n + \sqrt{n}t) \pi(\hat{\theta}_n + \sqrt{n}t \mid X_1, \dots, X_{n_0}) dt} \end{aligned}$$

Now, as the theorem on asymptotic normality of posterior distribution holds for any choice of proper prior (as shown in the proof), it also holds true if we think the proper posterior distribution conditional on (x_1, \dots, x_{n_0}) as a new prior for the next samples. Also, we consider $\prod_{i=n_0+1}^n f(X_i \mid \hat{\theta}_n + \sqrt{n}t)$ as the new likelihood function instead of $\prod_{i=1}^n f(X_i \mid \hat{\theta}_n + \sqrt{n}t)$. Since, n_0 is finite and as $n \rightarrow \infty$, we have $(n - n_0) \rightarrow \infty$, and hence the theorem follows with

this new likelihood and new prior distribution once we show that the maximum likelihood estimates $\hat{\theta}_n$ and $\hat{\theta}_n^*$ satisfies $|\hat{\theta}_n - \hat{\theta}_n^*| \rightarrow 0$. This follows from the strong consistency of the likelihood equation which tells that both $\hat{\theta}_n$ and $\hat{\theta}_n^*$ converges to θ_0 under P_{θ_0} -probability one.

(b) We start by noting that it is enough to show that, $\int_{\mathbb{R}} |t| |g_n(t)| dt \rightarrow 0$, where

$$g_n(t) = \pi_n \left(\hat{\theta}_n + \frac{t}{\sqrt{n}} \right) \exp \left(L_n \left(\hat{\theta}_n + \frac{t}{\sqrt{n}} \right) - L_n(\hat{\theta}_n) \right) - \pi(\theta_0) e^{-\frac{t^2}{2} I(\theta_0)}$$

This is because from Theorem 4.2, it follows that

$$c_n = \int_{\mathbb{R}} \pi_n \left(\hat{\theta}_n + \frac{t}{\sqrt{n}} \right) \exp \left(L_n \left(\hat{\theta}_n + \frac{t}{\sqrt{n}} \right) - L_n(\hat{\theta}_n) \right) dt \rightarrow \pi(\theta_0) \sqrt{2\pi} / \sqrt{I(\theta_0)}$$

and the integrand

$$\begin{aligned} & \int_{\mathbb{R}} |t| \left| \pi_n \left(\hat{\theta}_n + \frac{t}{\sqrt{n}} \right) \exp \left(L_n \left(\hat{\theta}_n + \frac{t}{\sqrt{n}} \right) - L_n(\hat{\theta}_n) \right) - c_n \frac{\sqrt{I(\theta_0)}}{\sqrt{2\pi}} e^{-\frac{1}{2} t^2 I(\theta_0)} \right| dt \\ & \leq \int_{\mathbb{R}} |t| |g_n(t)| dt + \left| \pi(\theta_0) - c_n \frac{\sqrt{I(\theta_0)}}{\sqrt{2\pi}} \right| \int_{\mathbb{R}} |t| e^{-\frac{1}{2} t^2 I(\theta_0)} dt \end{aligned}$$

Clearly the integral $\int_{\mathbb{R}} |t| e^{-\frac{1}{2} t^2 I(\theta_0)} dt$ is a Gaussian integral and is finite. Also, since $c_n \rightarrow \pi(\theta_0) \sqrt{2\pi} / \sqrt{I(\theta_0)}$, it implies that the second integral goes to 0. Thus, to prove the result, it is enough to show that the first integral also goes to 0 as $n \rightarrow \infty$.

Similar to the proof of Theorem 4.2, we consider two sets, $A_1 = \{t : |t| \leq \delta_0 / \sqrt{n}\}$ and $A_2 = \{t : |t| > \delta_0 / \sqrt{n}\}$, where δ_0 is a suitably chosen positive number such that $2\delta_0 < \delta$ so that $|\theta - \theta_0| < \delta_0$ implies $\pi(\theta) < 2\pi(\theta_0)$ (this is possible since $\pi(\cdot)$ is continuous at θ_0 and is positive at θ_0), and also $\frac{\delta_0}{6} \mathbb{E}_{\theta_0}(M(X_1)) < \frac{I(\theta_0)}{8}$. Similar to the proof of Theorem 4.2, we shall show that $\int_{A_i} |t| |g_n(t)| dt \rightarrow 0$ for $i = 1, 2$.

In the proof of Theorem 4.2, we showed that, $g_n(t) \rightarrow 0$ as $n \rightarrow \infty$ for any $t \in A_1$. Therefore, for any fixed $t \in A_1$, $|t| |g_n(t)| \rightarrow 0$. It was also shown that,

$$|g_n(t)| \leq 2\pi(\theta_0) e^{-\frac{1}{8} t^2 I(\theta_0)} + \pi(\theta_0) e^{-\frac{1}{2} t^2 I(\theta_0)}$$

i.e.,

$$|t| |g_n(t)| \leq 2\pi(\theta_0) |t| e^{-\frac{1}{8} t^2 I(\theta_0)} + \pi(\theta_0) |t| e^{-\frac{1}{2} t^2 I(\theta_0)}$$

which is again integrable. Therefore, an application of the Dominated Convergence Theorem (DCT) yields that $\int_{A_1} |t| |g_n(t)| dt \rightarrow 0$.

Turning to the integral over the set A_2 , we consider the chain of inequalities,

$$\begin{aligned}
& \int_{A_2} |t| |g_n(t)| dt \\
& \leq \int_{A_2} |t| \pi \left(\hat{\theta}_n + \frac{t}{\sqrt{n}} \right) e^{-n\epsilon/2} dt + \int_{A_2} |t| \pi(\theta_0) e^{-\frac{1}{2}t^2 I(\theta_0)} dt \\
& = \int_{A_2} \sqrt{n} |\theta - \hat{\theta}_n| \pi(\theta) e^{-n\epsilon/2} dt + \int_{A_2} |t| \pi(\theta_0) e^{-\frac{1}{2}t^2 I(\theta_0)} dt \\
& = I_1 + I_2
\end{aligned}$$

Note that, for sufficiently large n , as $|\hat{\theta}_n - \theta_0| < 1$ by strong consistency of $\hat{\theta}_n$, and by consistency of the posterior distribution, $\pi_{\theta_0|X_1, \dots, X_n}(|\theta - \theta_0| < 1) = 1$ for large n . Therefore, with P_{θ_0} -probability one, for large n , $|\hat{\theta}_n - \theta| < 2$. Hence, for sufficiently large n ,

$$I_1 \leq \int_{A_2} 2n\pi(\theta) e^{-n\epsilon/2} d\theta \leq 2ne^{-n\epsilon/2} \int_{\mathbb{R}} \pi(\theta) d\theta = 2ne^{-n\epsilon/2} \rightarrow 0$$

On the other hand, $I_2 = \pi(\theta_0) \int_{A_2} |t| e^{-\frac{1}{2}t^2 I(\theta_0)} dt = \frac{\pi(\theta_0) \sqrt{2\pi}}{\sqrt{I(\theta_0)}} \mathbb{E}(|T| \mathbf{1}_{\{T \in A_2\}})$, where T is a random variable following normal distribution with location parameter 0 and variance parameter $I(\theta_0)^{-1}$. Note that, $A_1 \uparrow \mathbb{R}$ as $n \rightarrow \infty$, and as $\mathbb{E}(|T|) < \infty$, it follows that by an application of DCT, $\mathbb{E}(|T| \mathbf{1}_{\{T \in A_1\}}) \uparrow \mathbb{E}(|T|)$. But, $\mathbb{E}(|T|) = \mathbb{E}(|T| \mathbf{1}_{\{T \in A_1\}}) + \mathbb{E}(|T| \mathbf{1}_{\{T \in A_2\}})$ as A_1 and A_2 are complement sets. Therefore, $\mathbb{E}(|T| \mathbf{1}_{\{T \in A_2\}}) \rightarrow 0$ as $n \rightarrow \infty$. This shows that $I_2 \rightarrow 0$.

This shows that, $\int_{A_2} |t| |g_n(t)| dt \rightarrow 0$. This completes the proof.

Problem 4. Suppose we have observations X_1, \dots, X_n . Under model M_0 , X_i are i.i.d. $\mathcal{N}(0, 1)$ and under model M_1 , X_i are i.i.d. $\mathcal{N}(\theta, 1)$, $\theta \in \mathbb{R}$. Consider the noninformative prior $g_1(\theta) \equiv 1$ for θ under M_1 . Show that if we use training samples of size 2, and calculate the corresponding AIBF, the corresponding intrinsic prior will be $\mathcal{N}(0, 1)$.

Solution. For the given setup, the objective Bayes factor based on a training sample $(X_i, X_j), i \neq j$ of size 2 is given by

$$B_{10}(X_i, X_j) = \frac{m_0(X_i, X_j)}{m_1(X_i, X_j)} \frac{\int f_1(X_1, \dots, X_n | \theta) g_1(\theta) d\theta}{f_0(X_1, \dots, X_n | \theta = 0)}$$

where $m_k(X_i, X_j)$, for $k = 0, 1$ denotes the marginal distribution of the training sample (X_i, X_j) under the model M_k . Hence,

$$\begin{aligned} \int f_1(X_1, \dots, X_n | \theta) g_1(\theta) d\theta &= \int_{-\infty}^{\infty} \frac{1}{(\sqrt{2\pi})^n} \exp \left[-\frac{1}{2} \sum_{i=1}^n (X_i - \theta)^2 \right] d\theta \\ &= \frac{1}{(\sqrt{2\pi})^n} \exp \left[-\frac{\sum_{i=1}^n X_i^2}{2} \right] \int_{-\infty}^{\infty} \exp \left[-\frac{1}{2} (n\theta^2 - 2n\bar{X}_n \theta) \right] d\theta \\ &= \frac{1}{(\sqrt{2\pi})^n} \exp \left[-\frac{\sum_{i=1}^n X_i^2}{2} \right] \frac{\sqrt{2\pi}}{\sqrt{n}} \exp \left[\frac{n\bar{X}_n^2}{2} \right] \\ &= \frac{1}{(\sqrt{2\pi})^{(n-1)} \sqrt{n}} \exp \left[-\frac{\sum_{i=1}^n X_i^2}{2} \right] \exp \left[\frac{n\bar{X}_n^2}{2} \right] \end{aligned}$$

Also, the marginal distribution of the training sample (X_i, X_j) under model M_1 is similarly given by,

$$\begin{aligned} m_1(X_i, X_j) &= \int f_1(X_i, X_j | \theta) g_1(\theta) d\theta \\ &= \int_{-\infty}^{\infty} \frac{1}{(\sqrt{2\pi})^2} \exp \left[-\frac{1}{2} [(X_i - \theta)^2 + (X_j - \theta)^2] \right] d\theta \\ &= \frac{1}{\sqrt{2\pi}\sqrt{2}} \exp \left[-\frac{X_i^2 + X_j^2}{2} \right] \exp \left[\frac{(X_i + X_j)^2}{4} \right] \end{aligned}$$

Therefore, the objective Bayes factor for the training sample (X_i, X_j) is,

$$\begin{aligned} B_{10}(X_i, X_j) &= \frac{(2\pi)^{-1} \exp\{-(X_i^2 + X_j^2)/2\}}{(2\pi)^{-1/2} 2^{-1/2} \exp\{-(X_i^2 + X_j^2)/2\} \exp\{(X_i + X_j)^2/4\}} \times \\ &\quad \frac{(2\pi)^{-(n-1)/2} n^{-1/2} \exp\{-\sum_{k=1}^n X_k^2/2\} \exp\{n\bar{X}_n^2/2\}}{(2\pi)^{-n/2} \exp\{-\sum_{k=1}^n X_k^2/2\}} \\ &= \frac{\sqrt{2}}{\sqrt{n}} \exp \left[\frac{1}{2} \left\{ n\bar{X}_n^2 - \frac{(X_i + X_j)^2}{2} \right\} \right] \end{aligned}$$

Hence, the average intrinsic Bayes factor will be

$$AIBF_{10} = \frac{\sqrt{2}}{n^{3/2}(n-1)} \exp \left[\frac{n\bar{X}_n^2}{2} \right] \sum_{1 \leq i \neq j \leq n} \exp \left[- \left(\frac{X_i + X_j}{2} \right)^2 \right] \quad (1)$$

Note that,

$$U = \frac{1}{n(n-1)} \sum_{1 \leq i \neq j \leq n} \exp \left[- \left(\frac{X_i + X_j}{2} \right)^2 \right]$$

is a U-statistic. Hence, by asymptotic theory of the U-statistic and strong law of large numbers, it follows that as $n \rightarrow \infty$, $U \rightarrow \mathbb{E}_\theta [\exp(-(X_1 + X_2)^2/4)]$, under the larger model M_1 . Since, under M_1 , X_1 and X_2 has $\mathcal{N}(\theta, 1)$ distribution, hence $(X_1 + X_2)/\sqrt{2} \sim \mathcal{N}(\sqrt{2}\theta, 1)$, hence $(X_1 + X_2)^2/2$ has a non-central chi-square distribution with 1 d.f. and noncentrality parameter $2\theta^2$. Hence, $\mathbb{E}_\theta [\exp(-(X_1 + X_2)^2/4)]$ is simply the value of the MGF (moment generating function) of such non-central chi-square distribution evaluated at $t = -1/2$. Therefore,

$$\mathbb{E}_\theta \left[\exp \left(- \frac{(X_1 + X_2)^2}{4} \right) \right] = (1 - 2t)^{-1/2} \exp \left(\frac{2\theta^2 t}{1 - 2t} \right) \Big|_{t=-1/2} = \frac{1}{\sqrt{2}} e^{-\theta^2/2}$$

Hence, for large n , the AIBF given in Eq. (1) can be approximated as

$$AIBF_{10} \approx \frac{1}{\sqrt{n}} \exp \left[\frac{n\bar{X}_n^2}{2} \right] e^{-\theta^2/2}$$

Finally, in order to obtain the intrinsic prior, the suggestion by Berger and Perichi can be employed. They suggests of using $\pi_1(\theta) = g_1(\theta)B_1^*(\theta)$, where $\bar{B}_{10} = \frac{1}{n(n-1)} \sum_{1 \leq i \neq j \leq n} \frac{m_0(X_i, X_j)}{m_1(X_i, X_j)} \rightarrow B_1^*(\theta)$. Based on the calculation shown above, we obtain

$$\bar{B}_{10} = \frac{\sqrt{2}}{\sqrt{2\pi}} U \rightarrow \frac{1}{\sqrt{2\pi}} e^{-\theta^2/2}$$

with P_θ -probability one, for any $\theta \in \mathbb{R}$. Since, $g_1(\theta) = 1$, we have intrinsic prior as

$$\pi_1(\theta) = \frac{1}{\sqrt{2\pi}} e^{-\theta^2/2}$$

which is the standard normal density $\mathcal{N}(0, 1)$.

Problem 5. In the Bayesian variable selection problem based on g -prior in normal linear regression models, Find the closed form expression for the marginal likelihood $m_\alpha(\mathbf{y}_n)$ under the model

$$M_\alpha : \boldsymbol{\mu}_n = \mathbf{1}_n \beta_0 + \mathbf{X}_{n\alpha} \boldsymbol{\beta}_\alpha$$

Solution. We have the parameter $\boldsymbol{\theta}_\alpha = (\beta_0, \boldsymbol{\beta}_\alpha, \sigma^2)$, with the g -prior given by,

$$p(\beta_0, \sigma^2 \mid M_\alpha) = \frac{1}{\sigma^2} \\ \boldsymbol{\beta}_\alpha \mid (\beta_0, \sigma^2, M_\alpha) \sim \mathcal{N}_{p(\alpha)}(\mathbf{0}, g\sigma^2(\mathbf{X}_{n\alpha}^\top \mathbf{X}_{n\alpha})^{-1})$$

Also, the response variable $\mathbf{y}_n \sim \mathcal{N}(\boldsymbol{\mu}_n, \sigma^2 I_n)$. Therefore, the marginal likelihood of \mathbf{y}_n under the model M_α is given by,

$$\begin{aligned} m_\alpha(\mathbf{y}_n) &= \int p(\mathbf{y}_n \mid \beta_0, \boldsymbol{\beta}_\alpha, \sigma^2) p(\boldsymbol{\beta}_\alpha \mid \beta_0, \sigma^2) p(\beta_0, \sigma^2) d\beta_0 d\boldsymbol{\beta}_\alpha d\sigma^2 \\ &\propto \int \frac{1}{\sigma^n} \exp \left[-\frac{1}{2\sigma^2} (\mathbf{y}_n - \beta_0 \mathbf{1}_n - \mathbf{X}_{n\alpha} \boldsymbol{\beta}_\alpha)^\top (\mathbf{y}_n - \beta_0 \mathbf{1}_n - \mathbf{X}_{n\alpha} \boldsymbol{\beta}_\alpha) \right] \times \\ &\quad \frac{1}{(\sqrt{g}\sigma)^{p(\alpha)} \det((\mathbf{X}_{n\alpha}^\top \mathbf{X}_{n\alpha})^{-1})} \exp \left[-\frac{\boldsymbol{\beta}_\alpha^\top \mathbf{X}_{n\alpha}^\top \mathbf{X}_{n\alpha} \boldsymbol{\beta}_\alpha}{2g\sigma^2} \right] \times \frac{1}{\sigma^2} d\beta_0 d\boldsymbol{\beta}_\alpha d\sigma^2 \\ &\propto \int (\sigma^2)^{-\frac{n+p(\alpha)}{2}-1} \exp \left[-\frac{A}{2\sigma^2} \right] d\beta_0 d\boldsymbol{\beta}_\alpha d\sigma^2 \end{aligned}$$

where,

$$A = (\mathbf{y}_n - \beta_0 \mathbf{1}_n - \mathbf{X}_{n\alpha} \boldsymbol{\beta}_\alpha)^\top (\mathbf{y}_n - \beta_0 \mathbf{1}_n - \mathbf{X}_{n\alpha} \boldsymbol{\beta}_\alpha) + \frac{1}{g} \boldsymbol{\beta}_\alpha^\top \mathbf{X}_{n\alpha}^\top \mathbf{X}_{n\alpha} \boldsymbol{\beta}_\alpha$$

Now, note that, A can be rewritten as,

$$A = \frac{1}{g} \boldsymbol{\beta}_\alpha^\top \mathbf{X}_{n\alpha}^\top \mathbf{X}_{n\alpha} \boldsymbol{\beta}_\alpha + (\mathbf{y}_n - \mathbf{X}_{n\alpha} \boldsymbol{\beta}_\alpha)^\top (\mathbf{y}_n - \mathbf{X}_{n\alpha} \boldsymbol{\beta}_\alpha) - 2\beta_0 \mathbf{1}_n^\top (\mathbf{y}_n - \mathbf{X}_{n\alpha} \boldsymbol{\beta}_\alpha) + n\beta_0^2$$

which is a quadratic in β_0 . Therefore, performing the integration with respect to β_0 using the technique of completing the squares and Gaussian integral, we have,

$$\begin{aligned} m_\alpha(\mathbf{y}_n) &\propto \int (\sigma^2)^{-\frac{n+p(\alpha)}{2}-1} \exp \left[-\frac{\boldsymbol{\beta}_\alpha^\top \mathbf{X}_{n\alpha}^\top \mathbf{X}_{n\alpha} \boldsymbol{\beta}_\alpha}{2g\sigma^2} \right] \exp \left[-\frac{(\mathbf{y}_n - \mathbf{X}_{n\alpha} \boldsymbol{\beta}_\alpha)^\top (\mathbf{y}_n - \mathbf{X}_{n\alpha} \boldsymbol{\beta}_\alpha)}{2\sigma^2} \right] \\ &\quad \times \sqrt{2\pi}\sigma \exp \left[\frac{1}{2n\sigma^2} \{(\mathbf{y}_n - \mathbf{X}_{n\alpha} \boldsymbol{\beta}_\alpha)^\top \mathbf{1}_n \mathbf{1}_n^\top (\mathbf{y}_n - \mathbf{X}_{n\alpha} \boldsymbol{\beta}_\alpha)\} \right] d\boldsymbol{\beta}_\alpha d\sigma^2 \\ &\propto \int (\sigma^2)^{-\frac{n+p(\alpha)+1}{2}} \exp \left[-\frac{B}{2\sigma^2} \right] d\boldsymbol{\beta}_\alpha d\sigma^2 \end{aligned}$$

where (let $\frac{1}{n}\mathbf{1}_n\mathbf{1}_n^\top$ be denoted by \mathbf{J}_n), B can be represented as,

$$\begin{aligned} B &= \frac{1}{g}\beta_\alpha^\top \mathbf{X}_{n\alpha}^\top \mathbf{X}_{n\alpha} \beta_\alpha + (\mathbf{y}_n - \mathbf{X}_{n\alpha} \beta_\alpha)^\top (\mathbf{y}_n - \mathbf{X}_{n\alpha} \beta_\alpha) - (\mathbf{y}_n - \mathbf{X}_{n\alpha} \beta_\alpha)^\top \mathbf{J}_n (\mathbf{y}_n - \mathbf{X}_{n\alpha} \beta_\alpha) \\ &= \beta_\alpha^\top \mathbf{X}_{n\alpha}^\top \left(\left(\frac{1}{g} + 1 \right) \mathbf{I}_n - \mathbf{J}_n \right) \mathbf{X}_{n\alpha} \beta_\alpha - 2\beta_\alpha^\top \mathbf{X}_{n\alpha}^\top (\mathbf{I}_n - \mathbf{J}_n) \mathbf{y}_n + \mathbf{y}_n^\top (\mathbf{I}_n - \mathbf{J}_n) \mathbf{y}_n \end{aligned}$$

Since the above is a quadratic form in β_α , again applying the method of square completion and formula of Gaussian integral, we can perform the integration with respect to β_α , to obtain that

$$\begin{aligned} m_\alpha(\mathbf{y}_n) &\propto \int (\sigma^2)^{-\frac{n+p(\alpha)+1}{2}} \exp \left[-\frac{\mathbf{y}_n^\top (\mathbf{I}_n - \mathbf{J}_n) \mathbf{y}_n}{2\sigma^2} \right] \times (\sqrt{2\pi}\sigma)^{p(\alpha)} \\ &\quad \exp \left[\frac{1}{2\sigma^2} \mathbf{y}_n^\top (\mathbf{I}_n - \mathbf{J}_n)^\top \mathbf{X}_{n\alpha} \left(\mathbf{X}_{n\alpha}^\top \left(\left(\frac{1}{g} + 1 \right) \mathbf{I}_n - \mathbf{J}_n \right) \mathbf{X}_{n\alpha} \right)^{-1} \mathbf{X}_{n\alpha}^\top (\mathbf{I}_n - \mathbf{J}_n) \mathbf{y}_n \right] d\sigma^2 \\ &\propto \int (\sigma^2)^{-\frac{n-1}{2}-1} \exp \left[-\frac{C}{2\sigma^2} \right] d\sigma^2 \\ &\propto \left[\frac{C}{2} \right]^{-(n-1)/2}, \quad \text{comparing with inverse gamma density} \end{aligned}$$

where,

$$C = \mathbf{y}_n^\top (\mathbf{I}_n - \mathbf{J}_n) \mathbf{y}_n - \mathbf{y}_n^\top (\mathbf{I}_n - \mathbf{J}_n)^\top \mathbf{X}_{n\alpha} \left(\mathbf{X}_{n\alpha}^\top \left(\left(\frac{1}{g} + 1 \right) \mathbf{I}_n - \mathbf{J}_n \right) \mathbf{X}_{n\alpha} \right)^{-1} \mathbf{X}_{n\alpha}^\top (\mathbf{I}_n - \mathbf{J}_n) \mathbf{y}_n$$

Now, since \mathbf{X}_n is centered, hence $\mathbf{X}_{n\alpha}^\top \mathbf{1}_n = \mathbf{0}$. Therefore, $\mathbf{X}_{n\alpha}^\top \mathbf{J}_n = \mathbf{0}$. Consequently, C is simplified as,

$$\begin{aligned} C &= \mathbf{y}_n^\top (\mathbf{I}_n - \mathbf{J}_n) \mathbf{y}_n - \left(\frac{g}{g+1} \right) \mathbf{y}_n^\top \mathbf{X}_{n\alpha} (\mathbf{X}_{n\alpha}^\top \mathbf{X}_{n\alpha})^{-1} \mathbf{X}_{n\alpha} \mathbf{y}_n \\ &= \frac{1}{(g+1)} \mathbf{y}_n^\top (\mathbf{I}_n - \mathbf{J}_n) \mathbf{y}_n + \frac{g}{g+1} \mathbf{y}_n^\top \left(\mathbf{I}_n - \frac{1}{n} \mathbf{1}_n \mathbf{1}_n^\top - \mathbf{X}_{n\alpha} (\mathbf{X}_{n\alpha}^\top \mathbf{X}_{n\alpha})^{-1} \mathbf{X}_{n\alpha} \right) \mathbf{y}_n \\ &= \frac{1}{(g+1)} \sum_{i=1}^n (y_i - \bar{y})^2 + \frac{g}{(g+1)} \mathbf{y}_n^\top \left(\mathbf{I}_n - [\mathbf{1}_n \quad \mathbf{X}_{n\alpha}] \begin{bmatrix} 1/n & \mathbf{0}^\top \\ \mathbf{0} & (\mathbf{X}_{n\alpha}^\top \mathbf{X}_{n\alpha})^{-1} \end{bmatrix} \begin{bmatrix} \mathbf{1}_n^\top \\ \mathbf{X}_{n\alpha}^\top \end{bmatrix} \right) \mathbf{y}_n \\ &= \frac{1}{(g+1)} \sum_{i=1}^n (y_i - \bar{y})^2 + \frac{g}{(g+1)} \mathbf{y}_n^\top (\mathbf{I}_n - P_n(\alpha)) \mathbf{y}_n \end{aligned}$$

where $P_n(\alpha)$ is the projection matrix $\mathbf{Z}_{n\alpha} (\mathbf{Z}_{n\alpha}^\top \mathbf{Z}_{n\alpha})^{-1} \mathbf{Z}_{n\alpha}^\top$, where $\mathbf{Z}_{n\alpha} = [\mathbf{1}_n \quad \mathbf{X}_{n\alpha}]$.

Now, we note that the marginal distribution $m_\alpha(\mathbf{y}_n)$ is given by, $C^{-(n-1)/2}$ upto a proportionality constant, and the quantity C is simply a quadratic form of \mathbf{y}_n . Therefore, $m_\alpha(\mathbf{y}_n)$ is a multivariate (n -variate) student's t distribution. Comparing it with the density of t-distribution, we obtain that,

$$m_{\alpha}(\mathbf{y}_n) = \frac{\Gamma((n-1)/2)}{\pi^{(n-1)/2} \sqrt{n} (1+g)^{p(\alpha)/2}} \times \left[(1-a) \sum_{i=1}^n (y_i - \bar{y})^2 + a \mathbf{y}_n^{\top} (\mathbf{I}_n - P_n(\alpha)) \mathbf{y}_n \right]^{-(n-1)/2}$$

where $a = g/(g+1)$. This completes the answer.

Problem 6. Consider the example of heirarchical Bayesian analysis of the usual one-way ANOVA.

Take $k = 10$ and $n_i = 25$ for all i . This means to generate 10 samples, each of size 25, from 10 normal populations. Choose 10 different values of the population means $\theta_1, \dots, \theta_{10}$ and a common value of the population variance. Choose the values of the hyperparameters $a_1, a_2, b_1, b_2, \mu_0, \sigma_0^2$ so that the corresponding priors are not very informative. Use the Gibbs sampling procedure to estimate the population means. Do the same for five different choices of $(a_1, a_2, b_1, b_2, \mu_0, \sigma_0^2)$.

Do the above procedure (a) using WinBUGS and also (b) using your own code written in your favourite programming language such as R or Python.

Solution. The complete heirarchical Bayesian setup of the usual one-way ANOVA model is given by,

$$\begin{aligned} y_{ij} &= \theta_i + \epsilon_{ij}, \quad j = 1, 2, \dots, 25, \quad i = 1, 2, \dots, 10 \\ \epsilon_{ij} &\sim \mathcal{N}(0, \sigma^2) \quad j = 1, 2, \dots, 25, \quad i = 1, 2, \dots, 10 \\ \theta_i &\sim \mathcal{N}(\mu_\pi, \sigma_\pi^2), \quad i = 1, 2, \dots, 10 \\ \sigma^2 &\sim \text{InverseGamma}(a_1, b_1), \\ \mu_\pi &\sim \mathcal{N}(\mu_0, \sigma_0^2) \\ \sigma_\pi^2 &\sim \text{InverseGamma}(a_2, b_2) \end{aligned}$$

where $a_1, a_2, b_1, b_2, \mu_0, \sigma_0^2$ are all specified constants.

We consider the true value of the population means as $\theta_i = 2i$, for $i = 1, 2, \dots, 10$, i.e. $\theta_1 = 2, \theta_2 = 4$ and so on. Also, the common value for the population variance is chosen to be $\sigma^2 = 1$. The following function generates the data.

```
1 theta <- 2 * c(1:10)      # different means
2 sigma.sq <- 1
3 n <- 25
4 k <- 10
5
6 set.seed(1911)           # set a seed for reproducibility
7 errors <- rnorm(n * k, mean = 0, sd = sqrt(sigma.sq))
8 y <- matrix(rep(theta, each = n) + errors, nrow = k, ncol = n, byrow = TRUE)
9
```

To perform Gibbs sampling, the full conditional distributions obtained can be described as follows,

1. $\boldsymbol{\theta} \mid \mu_\pi, \sigma_\pi^2, \sigma^2, \mathbf{y} \sim \mathcal{N}_{10}(\boldsymbol{\mu}_1, \Sigma_1)$, where,

$$\mu_{1i} = \frac{\sigma_\pi^2}{\sigma_\pi^2 + \sigma^2/n_i} \bar{y}_i + \frac{\sigma^2/n_i}{\sigma_\pi^2 + \sigma^2/n_i} \mu_\pi$$

and Σ_1 is a diagonal matrix with

$$(\Sigma_1)_{ii} = \frac{\sigma_\pi^2 \sigma^2 / n_i}{\sigma_\pi^2 + \sigma^2 / n_i}$$

2. $\sigma^2 \mid \mu_\pi, \sigma_\pi^2, \boldsymbol{\theta}, \mathbf{y}$ follows Inverse Gamma distribution with shape parameter $(a_1 + n_i/2)$ and scale parameter $(b_1 + S^2/2)$, where $S^2 = \sum_{i=1}^k \sum_{j=1}^{n_i} (y_{ij} - \theta_i)^2$.
3. $\mu_\pi \mid \sigma_\pi^2, \boldsymbol{\theta}, \mathbf{y}, \sigma^2$ follows a normal distribution

$$\mathcal{N}\left(\frac{\sigma_0^2}{\sigma_0^2 + \sigma_\pi^2/k} \bar{\theta} + \frac{\sigma_\pi^2/k}{\sigma_0^2 + \sigma_\pi^2/k} \mu_0, \frac{\sigma_0^2 \sigma_\pi^2/k}{\sigma_0^2 + \sigma_\pi^2/k}\right)$$

4. $\sigma_\pi^2 \mid \mu_\pi, \boldsymbol{\theta}, \mathbf{y}, \sigma^2$ follows an Inverse Gamma distribution with shape $(a_2 + k/2)$ and scale parameter $(b_2 + \sum_{i=1}^k (\theta_i - \mu_\pi)^2/2)$.

The following custom function written in R performs the above Gibbs sampling. We shall set a burn-in period equal to 1000 to ensure better convergence and a thinning interval as 10 (i.e. take every 10-th sample from the chain) to ensure that the successive posterior samples remain uncorrelated. In this way, we generate 1000 posterior samples from the joint posterior distribution of $(\boldsymbol{\theta}, \sigma^2, \mu_\pi, \sigma_\pi^2)$, and then obtain the estimate of population mean as a Bayes estimator with respect to the squared error loss, i.e. by obtaining the posterior average of the obtained samples.

```

1 do.Gibbs <- function(y, a1, a2, b1, b2, mu0, sigma0.sq, n_samples = 100, burnin
  = 1000, nthin = 10) {
2   k <- nrow(y)
3   n <- ncol(y)
4
5   # initialize Gibbs sampler
6   theta <- rep(1, k)
7   mu_pi <- 0
8   sigma_pi.sq <- 1
9   sigma.sq <- 1
10
11  nchain = burnin + (n_samples * nthin)
12  theta_post <- matrix(NA, nrow = n_samples, ncol = k)
13  pb <- txtProgressBar(max = nchain, style = 3) # create a progress bar
14
15  for (t in 1:nchain) {
16    # generate theta
17    mu1 <- (sigma_pi.sq / (sigma_pi.sq + sigma.sq / n)) * rowMeans(y) +
18      ((sigma.sq / n) / (sigma_pi.sq + sigma.sq / n)) * mu_pi
19    Sigma1 <- diag(k) * (sigma_pi.sq * sigma.sq / n) / (sigma_pi.sq + sigma.
sq / n)
20    theta <- MASS::mvrnorm(mu = mu1, Sigma = Sigma1)
21
22    # generate sigma^2
23    ss <- sum((y - matrix(theta, nrow = k, ncol = n, byrow = FALSE))^2)
24    sigma.sq <- 1 / rgamma(n = 1, shape = (a1 + n/2), rate = (b1 + ss/2))
25
26    # generate mu_pi
27    mu2 <- (sigma0.sq/(sigma0.sq + sigma_pi.sq/k)) * mean(theta) +
28      ((sigma0.sq/k)/(sigma0.sq + sigma_pi.sq/k)) * mu0
29    sigma2 <- (sigma0.sq * sigma_pi.sq / k) / (sigma0.sq + sigma_pi.sq / k)
30    mu_pi <- rnorm(n = 1, mean = mu2, sd = sqrt(sigma2))
31
32    # generate sigma_pi.sq

```



```

33     ss2 <- sum((theta - mu_pi)^2)
34     sigma_pi.sq <- 1 / rgamma(n = 1, shape = (a2 + k/2), rate = (b2 + ss2/2)
35   )
36   if (t > burnin) {
37     if ((t - burnin) %% nthin == 0) {
38       index <- (t - burnin) / nthin
39       theta_post[index, ] <- theta
40     }
41   }
42   setTxtProgressBar(pb, value = t)
43 }
44
45 close(pb)
46
47 return(theta_post)
48 }
49
50

```

To ensure that the quantities $(a_1, a_2, b_1, b_2, \mu_0, \sigma_0^2)$ are chosen such that the corresponding priors are not very informative, we choose $\mu_0 = 0$, and all of $a_1, a_2, b_1, b_2, \sigma_0^2$ as some high values (say 100 or 1000). We shall experiment with 5 different choices of these parameters.

```

1 set.seed(1234)
2 res <- do.Gibbs(y, 100, 100, 100, 100, 0, 100, n_samples = 1000, burnin = 1000,
3   nthin = 10)
4 colMeans(res)

```

The estimates of population means are summarized in the following table.

$(a_1, a_2, b_1, b_2, \mu_0, \sigma_0^2)$	Posterior mean of $(\theta_1, \dots, \theta_{10})$
$(10^2, 10^2, 10^2, 10^2, 0, 10^2)$	(2.328, 4.219, 6.26, 8.112, 10.072, 12.227, 13.931, 15.828, 17.224, 20.041)
$(10^3, 10^2, 10^3, 10^2, 0, 10^3)$	(2.219, 4.138, 6.189, 8.052, 10.043, 12.251, 13.988, 15.909, 17.307, 20.178)
$(10^2, 10^3, 10^2, 10^3, 0, 10^3)$	(2.662, 4.491, 6.441, 8.236, 10.085, 12.168, 13.835, 15.652, 16.991, 19.705)
$(10^3, 10^3, 10^3, 10^3, 0, 10^3)$	(2.401, 4.271, 6.286, 8.125, 10.061, 12.228, 13.934, 15.808, 17.174, 19.985)
$(10^3, 10^3, 10^3, 10^3, 10, 10^2)$	(2.428, 4.313, 6.334, 8.158, 10.099, 12.26, 13.965, 15.843, 17.213, 20.035)

Table 1: Estimates of population means based on Gibbs Sampling using own custom code

We see that the posterior estimates are very close to the true population means. Also, the effect of the choice of different values of $(a_1, a_2, b_1, b_2, \mu_0, \sigma_0^2)$ is negligible. Also, high values of a_1, a_2 (the shape parameter of the inverse gamma distributions) and low values of b_1, b_2 (the scale parameter of inverse gamma distributions) makes the posterior estimates less biased.

To perform the sample Gibbs sampling in WinBUGS software, we use the following WinBUGS code.

```

1 model {
2   for (i in 1:10) {
3     for (j in 1:25) {
4       y[i, j] ~ dnorm(theta[i], precision_sigma)
5     }
6   }
7 }

```

```

6   }
7
8   for (i in 1:10) {
9     theta[i] ~ dnorm(mu_pi, precision_sigma_pi)
10  }
11
12  precision_sigma ~ dgamma(100, 100)    # put values of (a1, b1)
13  precision_sigma_pi ~ dgamma(100, 100)  # put values of (a2, b2)
14  mu_pi ~ dnorm(0, 1/100) # put values of mu_0 and precision = (1 / sigma_0
15  ^2)
16 }

```

The WinBUGS software also outputs plots corresponding to the markov chains. Some of them are given in Figure 1.

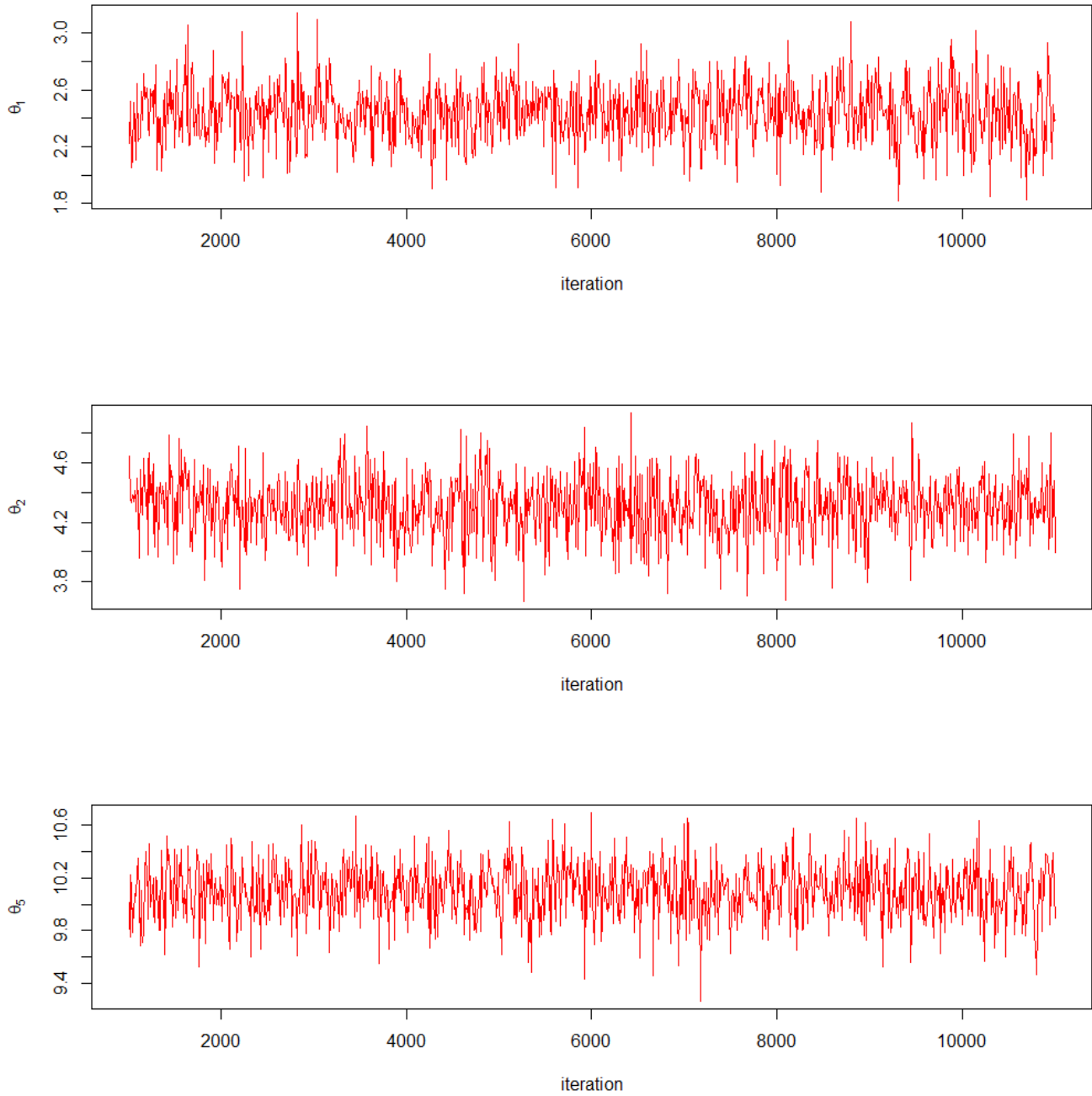
The estimates obtained from Gibbs sampling using WinBUGS software are given in the following table.

$(a_1, a_2, b_1, b_2, \mu_0, \sigma_0^2)$	Posterior mean of $(\theta_1, \dots, \theta_{10})$
$(10^2, 10^2, 10^2, 10^2, 0, 10^2)$	(2.345, 4.225, 6.254, 8.106, 10.043, 12.235, 13.947, 15.831, 17.201, 20.054)
$(10^3, 10^2, 10^3, 10^2, 0, 10^3)$	(2.221, 4.126, 6.188, 8.061, 10.049, 12.247, 13.983, 15.89, 17.311, 20.171)
$(10^2, 10^3, 10^2, 10^3, 0, 10^3)$	(2.684, 4.482, 6.437, 8.223, 10.084, 12.175, 13.835, 15.657, 16.968, 19.71)
$(10^3, 10^3, 10^3, 10^3, 0, 10^3)$	(2.394, 4.266, 6.285, 8.121, 10.061, 12.224, 13.919, 15.825, 17.177, 19.989)
$(10^3, 10^3, 10^3, 10^3, 10, 10^2)$	(2.437, 4.306, 6.326, 8.172, 10.098, 12.263, 13.958, 15.843, 17.216, 20.029)

Table 2: Estimates of population means based on Gibbs Sampling using WinBUGS

As expected, the results are quite similar to that obtained using the custom code for Gibbs sampling.

Figure 1: History of the MCMC chains in Gibbs sampling



Thank You
