BIOS 779 Homework 2

Mingwei Fei

5th February 2023

1 Problem 1

Suppose that $y_1, \ldots y_n$ are independent Bernoulli random variables, with

$$P(y_i = 1|\beta) = \frac{\exp(\beta_0 + \beta_1 x_i)}{1 + \exp(\beta_0 + \beta_1 x_i)}$$

where $\beta = (\beta_0, \beta_1)'$ is a 2×1 vector of regression coefficients, and the x_i's are fixed covariate values, $i = 1, ..n$.

(a) Derive Jeffreys's prior for β and show it is of the form

$$\pi(\beta) \propto |X'V(\beta)X|^{1/2}$$

Proof: We can see that the canonical link is logit link, $logit(\pi_i) = x_i\beta$. And we can write the log-likelihood function of $l_n(\beta)$ as below

For Bernoulli distribution,

$$p(x|\pi) = \pi^x(1-\pi)^{1-x}$$
$$= \exp\{\log\left(\frac{\pi}{1-\pi}\right)x + \log(1-\pi)\}$$

We see that Bernoulli distribution is an exponential family distribution with

$$\theta = \log\left(\frac{\pi}{1-\pi}\right)$$
$$b(\theta) = -\log(1-\pi) = \log\left(1 + \exp(\theta)\right)$$
$$\phi = 1$$
$$p(y|\beta) = \exp\{(\beta_0 + \beta_1 x)y - \log\left(1 + \exp(\beta_0 + \beta_1 x)\right)\}$$
$$\dot{b(\theta)} = \frac{\exp(\theta)}{(1 + exp(\theta))}$$
$$\ddot{b(\theta)} = \frac{\exp(\theta)}{(1 + exp(\theta))^2}$$

The generalized estimation for Fisher Information

$$V(\beta) = \text{diag}\Big(v_1(\beta), \ldots, v_n(\beta)\Big)$$

$$\mu_i = b(\dot\theta) = \frac{\exp(\beta_0 + \beta_1 x_i)}{(1 + exp(\beta_0 + \beta_1 x_i))}$$

$$v_i = b(\ddot\theta) = \frac{\exp(\beta_0 + \beta_1 x_i)}{(1 + exp(\beta_0 + \beta_1 x_i)^2)}$$

$$e(\beta) = (y_1 - \mu_1(\beta), \ldots, y_n - \mu_n(\beta))'$$

$$D_\theta(\beta)' = \Big(\partial_\beta \beta_1(\beta), \ldots, \partial_\beta \beta_n(\beta)\Big)_{p \times n}$$

$$\partial_\beta \beta_i(\beta) = (1, x_i)^T$$

$$D(\beta)^T = \Big(\partial_\beta \mu_1(\beta), \ldots, \partial_\beta \mu_n(\beta)\Big)_{p \times n}$$

$$\dot{l}_n(\beta) = \phi D_\theta(\beta)^T e(\beta) = \phi D(\beta)' V(\beta)^{-1} e(\beta)$$

$$E\Big[-\ddot{l}_n(\beta)\Big] = \phi D_\theta(\beta)' V D_\theta(\beta) = \phi D(\beta)' V(\beta)^{-1} D(\beta)$$

Then Fisher Information of β

$$w_i = (1, x_i)^T$$

$$\theta_i = k\Big(w_i^T \beta\Big) = w_i^T \beta$$

$$\xi = (\beta, \phi)$$

$$ln(\xi) = \sum_{i=1}^{n} \phi\Big[y_i k\Big(w_i^T \beta\Big) - b\Big(k\Big(w_i^T \beta\Big)\Big) - c(y_i)\Big] - \frac{1}{2} s(y_i, \phi)$$

$$\dot{ln}(\beta) = \frac{\partial ln(\beta)}{\partial \beta} = \phi \sum_{i=1}^{n} \Big[y_i - \dot{b}\Big(k\Big(w_i^T \beta\Big)\Big)\Big] \dot{k}\Big(w_i^T \beta\Big) x_i$$

$$= \sum_{i=1}^{n} \Big[y_i - \mu_i\Big] w_i$$

$$\ddot{ln}(\beta) = \frac{\partial^2 ln(\beta)}{\partial \beta \, \partial \beta} = -\phi \sum_{i=1}^{n} \ddot{b}\Big(k(w_i^T \beta)\Big) \dot{k}(w_i^T \beta)^2 w_i w_i^T + \phi \sum_{i=1}^{n} \Big[y_i - \dot{b}(k(w_i^T \beta))\Big] \ddot{k}(w_i^T \beta) w_i w_i^T$$

$$E\Big[-\ddot{l}_n(\beta)\Big] = -\sum_{i=1}^{n} \ddot{b}\Big(\theta_i\Big) w_i w_i^T = -\sum_{i=1}^{n} V(\theta_i) w_i w_i^T, \qquad \partial_\beta^2 b(\theta_i) = V(\theta_i)$$

let $X = \{w_i^T\}_{n \times p}$,

$$V(\theta) = diag\{V(\theta_i)\} = diag\{v_i\},$$

$$v_i = b(\ddot\theta) = \frac{\exp(\beta_0 + \beta_1 x_i)}{(1 + exp(\beta_0 + \beta_1 x_i)^2)}$$

$$I_n = E\Big[-\ddot{l}_n(\beta)\Big] = D_\theta(\beta)' V D_\theta(\beta) = X' V(\theta) X$$

$$I(\beta) = \frac{1}{n} X' V(\theta) X$$

Thus, Jeffery's prior is

$$\pi(\theta) = I(\theta)^{1/2}$$
$$= \left| \frac{1}{n} X'V(\theta)X \right|^{1/2} \propto |X'V(\theta)X|^{1/2}$$

(ii) Formally show that Jeffery's prior for β for the Bernoulli model is proper.

The Jeffery's prior for β

$$|X'V(\theta)X|^{1/2} = \left| \begin{pmatrix} \frac{1}{n}\sum_{i=1}^{n} \frac{\exp(\beta_0+\beta_1 x_i)}{(1+\exp(\beta_0+\beta_1 x_i)^2)} & 0 \\ 0 & \frac{1}{n}\sum_{i=1}^{n} \frac{\exp(\beta_0+\beta_1 x_i)}{(1+\exp(\beta_0+\beta_1 x_i)^2)}x_i^2 \end{pmatrix} \right|^{1/2}$$
$$= \frac{1}{n}\left[\sum_{i=1}^{n} \frac{\exp(\beta_0+\beta_1 x_i)}{(1+\exp(\beta_0+\beta_1 x_i)^2)} \sum_{i=1}^{n} \frac{\exp(\beta_0+\beta_1 x_i)}{(1+\exp(\beta_0+\beta_1 x_i)^2)}x_i^2 \right]^{1/2}$$

To show that Jeffery's prior is proper, then we need to show the integral of joint distribution is finite by the theorem in paper "On Bayesian Analysis of Generalized Linear Models Using Jeffery's prior",

$$L(\beta|X,y) = \prod_{i=1}^{n} p(y_i|\beta_i)$$
$$= \prod_{i=1}^{n} \exp\left((\beta_0+\beta_1 x_i)y_i - \log\left[1+\exp(\beta_0+\beta_1 x_i)\right] \right)$$
$$p(\beta|X,Y) \propto L(\beta|X,y)|X'V(\theta)X|^{1/2}$$
$$\propto \prod_{i=1}^{n} \exp\left((\beta_0+\beta_1 x_i)y_i - \log\left[1+\exp(\beta_0+\beta_1 x_i)\right] \right)$$
$$\frac{1}{n}\left[\sum_{i=1}^{n} \frac{\exp(\beta_0+\beta_1 x_i)}{(1+\exp(\beta_0+\beta_1 x_i))^2} \sum_{i=1}^{n} \frac{\exp(\beta_0+\beta_1 x_i)}{(1+\exp(\beta_0+\beta_1 x_i))^2}x_i^2 \right]^{1/2}$$

The below integral is finite:

$$\int \sum_{i=1}^{n} \exp\left((\beta_0+\beta_1 x_i)y_i - \log\left[1+\exp(\beta_0+\beta_1 x_i)\right] \right) \frac{\exp(\beta_0+\beta_1 x_i)}{(1+\exp(\beta_0+\beta_1 x_i))^2} d\beta_0 d\beta_1$$
$$= \int \sum_{i=1}^{n} \pi_i^{y}(1-\pi_i)^{1-y_i}\pi_i(1-\pi_i)d\pi_i$$
$$\int \sum_{i=1}^{n} \pi_i^{1-y_i}(1-\pi_i)^{2-y_i}d\pi_i < M$$

From the above integral, we can see that it is finite. So the Jeffery's prior is proper.

2 Problem 2

Suppose $x \sim \text{Pareto}(k, \theta)$, and

$$p(x|\theta) = \frac{\theta k^\theta}{x^{\theta+1}}, \quad x > k, \theta > 0 \tag{1}$$

$$\tag{2}$$

(a) Suppose that k is known and θ unknown. Derive a class of conjugate priors for θ.

The CDF of Pareto distribution is exponential distribution when k is known and θ unknown (ie. let $y = \log\frac{x}{k}$).

$$F(x) = 1 - \left(\frac{x}{k}\right)^{-\theta} = 1 - \left(\exp(\log\frac{x}{k})\right)^{-\theta}, \quad x > k, \theta > 0$$

$$p(x|\theta) = \frac{\theta k^\theta}{x^{\theta+1}}$$

$$p(y|\theta) = \frac{\theta k^\theta}{(ke^y)^{\theta+1}} ke^y = \theta \exp(-y\theta)$$

So Pareto distribution is an exponential distribution, so Gamma distribution is the conjugate prior for Exponential distribution.

$$p(\theta) \sim \text{Gamma}(\alpha, \beta)$$

$$p(\theta|Y) \propto \theta \exp(-y\theta)\theta^{\alpha-1} \exp\left(-\beta\theta\right)$$

$$\propto \theta^\alpha \exp\left(-(y+\beta)\theta\right)$$

$$\propto \theta^\alpha \exp\left(-(ln(x) - ln(k) + \beta)\theta\right)$$

The posterior distribution is also gamma distribution $\Gamma(\alpha+1, ln(x) - ln(k) + \beta)$, so the class of conjugate prior are Gamma distribution $\Gamma(\alpha, \beta)$.

(b) Suppose $x_1, \ldots x_n$ are i.i.d. random variables from the density in (2), and let z be a future observation with the same sampling density as (2). Derive the predictive distribution of z based on the prior obtained in part (a).

The posterior distribution of θ based on $x_1, \ldots x_n$, as the gamma distribution is conjugate for Pareto distribution, so the predictive distribution z is also a gamma distribution.

$$p(\theta|Y) \propto \theta^{\alpha+n} \exp\left(-(\sum y_i + \beta)\theta\right)$$

$$= \Gamma(\alpha+n, \sum_{i=1}^{n} y_i + \beta)$$

Then the predictive distribution of $z' = ln(\frac{z}{k})$

$$p(z'|Y) = \int p(z'|\theta)p(\theta|Y)d\theta$$

$$\propto \int \theta \exp(-z'\theta)\theta^{\alpha+n-1} \exp\left(-\theta(\sum_{i=1}^{n} y_i + \beta)\right)d\theta$$

$$= \frac{(\alpha+n)(\sum ln(x_i) - nln(k) + \beta)^{\alpha+n}}{\left(ln(z) + \sum ln(x_i) - (n+1)ln(k) + \beta\right)^{\alpha+n+1}}, \qquad \sum ln(x_i) - nln(k) + \beta > 0$$

where $\sum ln(x_i) - nln(k) + \beta > 0$.

(c) Derive a 95% credible interval for θ.

The posterior distribution of θ is Gamma distribution, and 95% credible interval with $\alpha = 0.05$

We transform the gamma distribution to chi-square distribution, let $\theta = \frac{w}{2(\sum y_i + \beta)}, w = 2\theta(\sum y_i + \beta)$, the Jacobian transformation

$$\left|\frac{d\theta}{dw}\right| = 2|(\sum y_i + \beta)| = 2(\sum y_i + \beta), \qquad \sum y_i + \beta > 0$$

$$w = 2\theta(\sum y_i + \beta) = \frac{\left(\sum y_i + \beta\right)^{\alpha+n}}{\Gamma(\alpha)}\left[\frac{w}{2(\sum y_i + \beta)}\right]^{\alpha+n-1} \exp\left(-\frac{(\sum y_i + \beta)w}{2(\sum y_i + \beta)}\right)$$

$$= \frac{1}{2^{\alpha+n}\Gamma(\alpha+n)}w^{\alpha+n-1}e^{-\frac{w}{2}}$$

So we have the 95% credible interval for w,

$$\chi^2\left(\alpha/2, 2(\alpha+n)\right) \le w \le \chi^2\left(1 - \alpha/2, 2(\alpha+n)\right)$$

Then 95% credible interval for θ,

$$\left\{\theta : 2(\sum y_i + \beta)\chi^2\left(\alpha/2, 2(\alpha+n)\right) \le \theta \le 2(\sum y_i + \beta)\chi^2\left(1 - \alpha/2, 2(\alpha+n)\right)\right\}$$

(d) Now suppose that θ, k are both unknown. Let $\gamma = log(k)$, and suppose that $\gamma \sim N(\mu_0, \sigma_0^2)$ where (μ_0, σ_0^2) are specified hyperparameters. Suppose that the prior is the same as in part (a) and that (γ, θ) are independent a priori.

(i) Derive the conditional posterior of $(\gamma|\theta, x)$ in closed form, where $x = (x_1, ..x_n)$.
As the gamma prior of θ is conjugate priors for exponential distribution, so the posterior distribution of θ is also a gamma distribution. Based on the independent priors of θ and γ,

$$p(X|\theta, k) = \frac{\theta k^\theta}{x^{\theta+1}}$$

$$= \theta \left[\exp\left(\log\frac{k}{x} \right) \right]^\theta x$$

$$L(X|\theta, k) = \theta^n \left[\exp \sum_{i=1}^n \left(\log\frac{k}{x} \right)\theta \right] \exp\left(\sum \log(x_i) \right)$$

$$p(\gamma, \theta|X) \propto L(Y|\theta, \gamma)p(\gamma)p(\theta)$$

$$\propto \theta^{\alpha+n-1} \exp\left(-\left(\sum_{i=1}^n \log(x_i) - n\gamma + \beta \right)\theta \right) (\sigma_0^2)^{-1/2} \exp\left(-\frac{(\gamma-\mu_0)^2}{2\sigma_0^2} \right)$$

$$\sum_{i=1}^n \log(x_i) - n\gamma + \beta > 0$$

Then we can write the posterior distribution $p(\gamma, \theta|X)$ in terms of the conditional distribution $p(\gamma|\theta, x)$ and $p(\theta|X)$

$$p(\gamma, \theta|X) = p(\gamma|\theta, x)p(\theta|x)$$

$$\propto \theta^{\alpha+n-1} \exp\left(-\left(\sum_{i=1}^n \log(x_i) + \beta \right)\theta \right) \exp\left(-\frac{(\gamma-\mu_0)^2}{2\sigma_0^2} - n\gamma\theta \right),$$

$$\log(x_i) > \gamma, \qquad \sum_{i=1}^n \log(x_i) - n\gamma + \beta > 0$$

Also, the normal prior for γ is also a conjugate for

$$p(\gamma|\theta, x) \propto \exp\left(-\frac{(\gamma-\mu_0)^2}{2\sigma_0^2} - n\gamma\theta \right)$$

$$\propto \exp\left(-\frac{\left(\gamma - \mu_0 + n\theta\sigma_0^2 \right)^2}{2\sigma_0^2} \right)$$

Because normal prior is conjugate to normal distribution, so the $\gamma|\theta, x \sim N(\mu_0 + n\theta\sigma_0^2, \sigma_0^2)$.

(ii) Use Part (i) to derive a 95% HPD interval for γ.

The 95% HPD interval for γ is

$$\left\{ \gamma : \mu_0 + n\theta\sigma_0^2 + z_{\alpha/2}\sigma_0 \le \gamma \le \mu_0 + n\theta\sigma_0^2 + z_{1-\alpha/2}\sigma_0 \right\}, \quad \log(x_i) > \gamma$$

3 Problem 3

Let $U(a, b)$ denote the uniform distribution on the interval (a, b). Given θ, suppose $X_1, \ldots X_n$ follows $U(\theta, \theta+1)$, and suppose that $\theta \sim U(a_0, b_0)$, a_0, b_0 are hyperparameters.

(a) find 95% credible interval for posterior distribution of θ We get from last homework that, the posterior distribution

$$p(\theta|X) = \frac{\frac{1}{b_0-a_0}I\left(\max(a_0, X_{(n)}-1) \le \theta \le \min(X_{(1)}, b_0)\right)}{\frac{\min(X_{(1)},b_0)-\max(a_0,X_{(n)}-1)}{(b_0-a_0)}}$$

$$= \frac{I\left(\max(a_0, X_{(n)}-1) \le \theta \le min(X_{(1)}, b_0)\right)}{\min(X_{(1)}, b_0) - max(a_0, X_{(n)}-1)}$$

The 95% credible interval of θ falls between θ_U and θ_L,

$$\begin{cases} p(\theta \ge \theta_U|X_1,..X_n) = 0.025, & \rightarrow \frac{\min(X_{(1)},b_0)-\theta_U}{\min(X_{(1)},b_0)-\max(a_0,X_{(n)}-1)} = 0.025 \\ p(\theta \le \theta_L|X_1,..X_n) = 0.025 & \rightarrow \frac{\theta_L-\max(a_0,X_{(n)}-1)}{\min(X_{(1)},b_0)-\max(a_0,X_{(n)}-1)} = 0.025 \end{cases}$$

Let $a_1 = \max(a_0, X_{(n)}-1), b_1 = \min(X_{(1)}, b_0)$ So the posterior θ 95% credible interval is

$$\left\{ \theta|X : a_1 + 0.025\{b_1 - a_1\} \le \theta \le b_1 - 0.025\{b_1 - a_1\} \right\}$$

(ii) Let z be a future observation with the same sampling density as the x_i's. Derive the predictive density of z.

$$p(z|X) = \int p(z|\theta)p(\theta|Y)d\theta$$

$$= \int_{a_1}^{b_1} I(\theta < z < \theta + 1)\frac{1}{b_1 - a_1}d\theta, \qquad \max(a_1, z-1) < \theta < \min(b_1, z)$$

$$= \int_{\max(a_1,z-1)}^{\min(b_1,z)} \frac{1}{b_1 - a_1}d\theta$$

$$= \frac{\min(b_1, z) - \max(a_1, z-1)}{b_1 - a_1}$$

(iii) Derive a 95% credible interval for z.

The 95% credible interval of z falls between z_U and z_L,

$$\begin{cases} p(z \ge z_U|X_1,..X_n) = 0.025, & \rightarrow \frac{\min(b_1,z_U)-z_U}{b_1-a_1} = 0.025 \\ p(z \le z_L|X_1,..X_n) = 0.025 & \rightarrow \frac{z_L-\max(a_1,z_L-1)}{b_1-a_1} = 0.025 \end{cases}$$

So the 95% credible interval is

$$\left\{ z|X : a_1 + 0.025\{b_1 - a_1\} \le \theta \le b_1 - 0.025\{b_1 - a_1\} \right\}$$

4   Problem 4

Consider linear model with error i.i.d. $\epsilon_i \sim N(0, \sigma^2), \tau = 1/\sigma^2$.

(a) Calculate based on given information

(i) Compute $P(\beta^2 > .25|y)$. We have

$$y \sim N(X\beta, \sigma^2 I_{4\times 4}), \qquad \tau = 1$$
$$\beta \sim N(\mu_0, \sigma_0^2), \qquad \mu_0 = 0, \sigma_0^2 = 1000$$

Then the posterior distribution of β is also a normal distribution due to conjugate prior

$$P(\beta|y) \propto (\tau)^{\frac{n}{2}} \exp\left(-\frac{\tau(y-X\beta)^T(y-X\beta)}{2}\right)(\sigma_0^2)^{-\frac{1}{2}} \exp\left(-\frac{(\beta-\mu_0)^2}{2\sigma_0^2}\right)$$

$$\propto (\tau)^{\frac{n}{2}} \exp\left(-\frac{\tau(\beta-\hat{\beta})^T X'X(\beta-\hat{\beta})}{2}\right)(\sigma_0^2)^{-\frac{1}{2}} \exp\left(-\frac{(\beta-\mu_0)^2}{2\sigma_0^2}\right)$$

$$\propto \exp\left\{-\frac{\tau}{2}\left[(X'X)\beta^2 - 2\hat{\beta}(X'X)\beta + \hat{\beta}^2\right] - \frac{(\beta-\mu_0)^2}{2\sigma_0^2}\right\}$$

$$\propto \exp\left\{-\frac{1}{2}\left((X'X)\tau + \frac{1}{\sigma_0^2}\right)\beta^2 + \left((X'X)\tau\hat{\beta} + \frac{\mu_0}{\sigma_0^2}\right)\beta\right\}$$

$$\propto \exp\left\{-\frac{1}{2}\left((X'X)\tau + \frac{1}{\sigma_0^2}\right)\left(\beta - \frac{\tau X'Y + \frac{\mu_0}{\sigma_0^2}}{\tau(X'X) + \sigma_0^{-2}}\right)\right\}$$

where we have the posterior distribution

$$P(\beta|y) \sim N\left\{\frac{\tau X'Y + \frac{\mu_0}{\sigma_0^2}}{\tau(X'X) + \sigma_0^{-2}}, \left((X'X)\tau + \frac{1}{\sigma_0^2}\right)^{-1}\right\}$$

$$X'Y = 6.535, \qquad X'X = 25.5, \qquad \tau = 1, \qquad \mu_0 = 0$$

$$P(\beta|y) \sim N\left\{\frac{6.535}{25.5 + 0.001}, \frac{1}{25.5 + 0.001}\right\}$$

$$P(\beta^2 > .25|y) = P(\beta > .5|y) + P(\beta < -.5|y)$$

$$=\sim N\left\{\frac{6.535}{25.5 + 0.001}, \frac{1}{25.5 + 0.001}\right\} = N\{0.256, 0.039\}$$

$$Z_1 = \frac{0.5 - 0.256}{\sqrt{0.039}} = 1.24, \qquad P(Z > 1.24) = 0.107$$

$$Z_2 = \frac{-0.5 - 0.256}{\sqrt{0.039}} = -3.83, \qquad P(Z < -3.83) = 0.00006$$

$$P(\beta^2 > .25|y) = 0.107 + 0.00006 = 0.10706$$

(ii) Construct a 95% HPD interval .

$$\left\{ \beta : Z_{\alpha/2}\sqrt{0.039} + 0.256 \leq \beta \leq Z_{1-\alpha/2}\sqrt{0.039} + 0.256 \right\}$$

$$\left\{ \beta : -1.96\sqrt{0.039} + 0.256 \leq \beta \leq 1.96\sqrt{0.039} + 0.256 \right\} = \left\{ \beta \ : -0.131 \leq \beta \leq 0.643 \right\}$$

(iii) Let $z = (z_1, z_2, z_3)$ be a 3×1 vector of future observations. Find the predictive distribution of z at $X_f = (1, 2, 3)'$.

The predictive distribution is a t distribution,

$$p(z|X, Y) = \int p(\beta|Y)p(z|\beta)d\beta$$

$$\propto \int \exp\left(-\frac{(\beta - 0.256)^2}{2 \times 0.039} \right) \exp\left(-\frac{(z - X_f\beta)^T(z - X_f\beta)}{2} \right) d\beta$$

$$\propto \int \exp\left(-\frac{1}{2}\left[\frac{1}{0.039} + x_f^T x_f \right] \left(\beta^2 - 2\frac{\frac{0.256}{0.039} + z^T x_f}{\frac{1}{0.039} + x_f^T x_f} \right) \right) d\beta$$

$$= S_3(n - 1, X_f\hat{\beta}, s^2(1 + \frac{1}{n}))$$

The mean of the predictive distribution is the same as the frequentist point estimate of z, and note that $s^2(1 + \frac{1}{n})$ is the frequentist estimate of $Var(z - \bar{X})$.

$$\hat{\beta} = (X'X)^{-1}X'Y = 0.256$$
$$M = X(X'X)^{-1}X'$$
$$s^2 = \frac{Y'(I - M)Y}{n - 1} = 0.034$$
$$E(z) = X_f\hat{\beta} = (1, 2, 3)'0.256 = (0.256, 0.512, 0.768)'$$
$$Var(z - \bar{X}) = s^2(1 + \frac{1}{n}) = 0.034(1 + \frac{1}{4}) = 0.0425$$
$$p(z|Y) = t_3(n - 1, E(z), Var(z - \overline{(X)})) = t(3, (0.256, 0.512, 0.768)', 0.0425I_{3\times3})$$

The R codes for calculating estimates:

```
X= c(1, 1.5, 2.5, 4)
Y = c(0.45, 0.6, 0.65, 0.89)
beta = solve(t(X) %*% X) * t(X) %*% Y
Xf = c(1,2,3)
Ez = Xf * beta
inv = solve(t(X) %*% X)
M= X %*% inv %*% t(X)
s2 = t(Y) %*% Y - t(Y) %*% M %*% Y
```

(b) Suppose $\tau = 1$, consider the hypothesis

$$H_0 : \beta = 0$$
$$H_1 : \beta = 1$$

9

Derive a general expression for the Bayes factor in favor of H_0, and then compute the Bayes factor using x, y given above.

The Bayes factor under $H_0 : \beta = 0, vs. H_1 : \beta = 1$ could be simplified as following:

$$
\begin{aligned}
B &= \frac{p(x|H_0)}{p(x|H_1)} = \frac{\int p(x|H_0, \theta)\lambda(\theta|H_0)d\theta}{\int p(x|H_1, \theta)\lambda(\theta|H_1)d\theta} \\
&= \frac{p(x|\beta_0)}{p(x|\beta_1)} \\
&= (\tau)^{\frac{n}{2}} \exp\left(- \frac{\tau(y - X\beta_0)^T(y - X\beta_0)}{2} \right) \Big/ (\tau)^{\frac{n}{2}} \exp\left(- \frac{\tau(y - X\beta_1)^T(y - X\beta_1)}{2} \right) \\
&= \frac{\exp\left(- \frac{\tau(y - X\beta_0)^T(y - X\beta_0)}{2} \right)}{\exp\left(- \frac{\tau(y - X\beta_1)^T(y - X\beta_1)}{2} \right)} \\
&= \exp\left[- \frac{y^T y}{2} + \frac{(y - x)^T(y - x)}{2} \right] \\
&= 500.2
\end{aligned}
$$

Due to $B = 500$, there is strong evidence in favor of H_0.

(c) Suppose both (τ, β) are unknown, and suppose we wish to test

$$
\begin{aligned}
H_0 &: \beta = 0 \\
H_1 &: \beta \neq 0
\end{aligned}
$$

Under $H_0, \tau \sim gamma(\frac{\delta_0}{2}, \frac{\gamma_0}{2})$. Under H_1, the joint prior for (β, τ) is

$$
\begin{aligned}
\beta|\tau &\sim N(\mu_0, \tau^{-1}\sigma_0^2) \\
\tau &\sim gamma(\frac{\delta_0}{2}, \frac{\gamma_0}{2})
\end{aligned}
$$

Derive a general expression for the Bayes factor in favor of H_0, and then compute the Bayes factor using the data given in part (b).

The posterior distribution of β with τ unknown is

Under $H_0 : \beta = 0$,

$$
\begin{aligned}
\lambda(\theta|H_0) &= p(\tau) \\
&= \frac{\frac{\gamma_0}{2}^{\frac{\delta_0}{2}}}{\Gamma(\frac{\delta_0}{2})} \tau^{\frac{\delta_0}{2} - 1} \exp\left(-\tau\frac{\gamma_0}{2} \right)
\end{aligned}
$$

Under $H_1 : \beta \neq 0$,

$$
\begin{aligned}
\lambda(\theta|H_1) &= p(\beta|\tau)p(\tau) \\
&= (2\pi)^{-\frac{n}{2}} \tau^{\frac{n}{2}} \exp\left(- \frac{(\beta - \mu_0)^T(\beta - \mu_0)}{2\tau^{-1}\sigma_0^2} \right) \frac{\frac{\gamma_0}{2}^{\frac{\delta_0}{2}}}{\Gamma(\frac{\delta_0}{2})} \tau^{\frac{\delta_0}{2} - 1} \exp\left(-\tau\frac{\gamma_0}{2} \right)
\end{aligned}
$$

10

The Bayes factor is

$$B = \frac{p(x|H_0)}{p(x|H_1)} = \frac{\int p(x|\beta_0)p(\tau)d\tau}{\int\int p(x|\beta,\tau)p(\beta|\tau)p(\tau)d\beta d\tau}$$

The numerator is

$$p(y|H_0) = (2\pi)^{-\frac{n}{2}} \frac{\left(\frac{\gamma_0}{2}\right)^{\frac{\delta_0}{2}}}{\Gamma(\frac{\delta_0}{2})} \int \tau^{\frac{n}{2}} \exp\left(-\tau \frac{y^T y}{2}\right) \frac{\left(\frac{\gamma_0}{2}\right)^{\frac{\delta_0}{2}}}{\Gamma(\frac{\delta_0}{2})} \tau^{\frac{\delta_0}{2}-1} \exp\left(-\tau \frac{\gamma_0}{2}\right) d\tau$$

$$= (2\pi)^{-\frac{n}{2}} \frac{\left(\frac{\gamma_0}{2}\right)^{\frac{\delta_0}{2}}}{\Gamma(\frac{\delta_0}{2})} \int \tau^{\frac{n}{2}+\frac{\delta_0}{2}-1} \exp\left(-\tau(\frac{y^T y}{2} + \frac{\gamma_0}{2})\right) d\tau$$

$$= (2\pi)^{-\frac{n}{2}} \frac{\Gamma(\frac{n}{2}+\frac{\delta_0}{2})}{\Gamma(\frac{\delta_0}{2})} \frac{\frac{\gamma_0}{2}^{\frac{\delta_0}{2}}}{\left(\frac{y^T y}{2} + \frac{\gamma_0}{2}\right)^{\frac{n}{2}+\frac{\delta_0}{2}}}$$

$$= A$$

The denominator is

$$p(y|H_1) = (2\pi\sigma_0^2)^{-\frac{1}{2}}(2\pi)^{-\frac{n}{2}} \frac{\left(\frac{\gamma_0}{2}\right)^{\frac{\delta_0}{2}}}{\Gamma(\frac{\delta_0}{2})} \frac{\left(\frac{\gamma_0}{2}\right)^{\frac{\delta_0}{2}}}{\Gamma(\frac{\delta_0}{2})}$$

$$\int\int \tau^{1/2} \exp\left(-\frac{(\beta-\mu_0)^2}{2\tau^{-1}\sigma_0^2}\right) \tau^{\frac{n}{2}} \exp\left(-\tau \frac{(y-x\beta)^T(y-x\beta)}{2}\right)$$

$$\tau^{\frac{\delta_0}{2}-1} \exp\left(-\tau \frac{\gamma_0}{2}\right) d\tau d\beta$$

$$= S_n\left(\delta_0, X\mu_0, \frac{\gamma_0}{\delta_0}(I + \sigma_0^2 XX')\right)$$

$$= \left(\frac{\Gamma(\frac{n+\delta_0}{2})(\pi\delta_0)^{-\frac{n}{2}}(\frac{\gamma_0}{\delta_0})^{-\frac{n}{2}}\left|I + X\sigma_0^2 X'\right|}{\Gamma(\frac{\delta_0}{2})}\right)$$

$$\left[1 + \frac{1}{\gamma_0}(Y - X\mu_0)^T(I + X\sigma_0^2 X')^{-1}(Y - X\mu_0)\right]^{-\frac{n+\delta_0}{2}}$$

$$= B$$

Then Baye's factor is $A/B = 125$, which is a strong evidence in favor of H_0.

we have R codes as following:

```
X= c(1, 1.5, 2.5, 4)
Y = c(0.45, 0.6, 0.65, 0.89)
n = 4
delta0 = 4
gamma0 = 1
mu0 = 0
sigma0 = 1000
matrix = diag(n) + sigma0^2 * X %*% t(X)
a = (2* pi)^(-n/2) *  gamma(n/2 + delta0 /2) *(gamma0/2)^(delta0 / 2) /
    gamma(delta0 /2) / ( t(Y) %*% Y /2 + gamma0 / 2)^(n/2 + delta0/2)
b = gamma(n/2 + delta0 /2) *(pi * delta0 )^(-n/2) * (gamma0/ delta0)^(-
    n/2) * det(matrix)^(-1/2) * (1 + 1/gamma0 * t(Y- mu0 * X)  %*%
    solve(matrix) %*% (Y- mu0 * X) )^(-n/2 - delta0/2)
bf = a/b
```

(d) Consider the model space $M = \{m_1, m_2\}$, where m_1 is the model $y_i = \epsilon_i$, and m_2 is the model $y_i = \beta x_i + \epsilon_i$. Suppose the prior probabilities of the models are $p(m_1) = .2$ and $p(m_2) = 0.8$. Using the priors and data given in part c), compute the posterior model probabilities for m_1 and m_2.

We have the posterior model probabilities for all models $m \in M$,

For model m_1, $\beta^{(m_1)} = 0$, we could get the $p(y|m_1)$ from part (c) that

$$p(m_1|y) = \frac{p(y|m_1)p(m_1)}{\sum_{m \in M} p(y|m)p(m)}$$

$$p(y|m_1) = \int \int p(y|X^{(m_1)}, \beta^{(m_1)}, \tau)\pi(\beta^{(m_1)}|\tau)\pi(\tau)d\beta^{(m_1)}d\tau$$

$$= (2\pi)^{-\frac{n}{2}} \frac{\Gamma(\frac{n}{2} + \frac{\delta_0}{2})}{\Gamma(\frac{\delta_0}{2})} \frac{\frac{\gamma_0}{2}^{\frac{\delta_0}{2}}}{(\frac{y^T y}{2} + \frac{\gamma_0}{2})^{\frac{n}{2} + \frac{\delta_0}{2}}} = A$$

For model m_2, $\beta^{(m_2)} \neq 0$, we could get the $p(y|m_2)$ from part (c) that

$$p(m_2|y) = \frac{p(y|m_2)p(m_2)}{\sum_{m \in M} p(y|m)p(m)}$$

$$p(y|m_2) = \int \int p(y|X^{(m_2)}, \beta^{(m_2)}, \tau)\pi(\beta^{(m_2)}|\tau)\pi(\tau)d\beta^{(m_2)}d\tau$$

$$= S_n\left(\delta_0, X\mu_0, \frac{\gamma_0}{\delta_0}(I + X\sigma_0^2 X')\right) = B$$

Then

$$p(m_1|y) = \frac{p(y|m_1)p(m_1)}{\sum_{m \in M} p(y|m)p(m)}$$

$$= \frac{0.2A}{0.2A + 0.8B} = 0.969$$

$$p(m_2|y) = \frac{p(y|m_2)p(m_2)}{\sum_{m \in M} p(y|m)p(m)}$$

$$= \frac{0.8B}{0.2A + 0.8B} = 0.031$$

Based on the posterior probabilities, it is in favor of model m_1 with $p(m_1|y) = 0.969$.

(e) Suppose that (β, τ) are both unknown, derive a general expression for the 95% HPD interval for β, and use the data in part (b) along with hyperparameters to construct the actual interval.

The posterior distribution for β as below

$$p(\beta|y) \propto p(y|\beta,\tau)\pi(\beta|\tau)\pi(\tau)$$

$$\propto \tau^{1/2}\exp\left(-\frac{(\beta-\mu_0)^T(\beta-\mu_0)}{2\tau^{-1}\sigma_0^2}\right)\tau^{\frac{n}{2}}\exp\left(-\tau\frac{(y-x\beta)^T(y-x\beta)}{2}\right)\tau^{\frac{\delta_0}{2}-1}\exp\left(-\tau\frac{\gamma_0}{2}\right)$$

$$\propto \tau^{\frac{n+\delta_0-1}{2}}\exp\left(-\frac{(\beta-\mu_0)^T(\beta-\mu_0)}{2\tau^{-1}\sigma_0^2}-\tau\frac{(y-x\beta)^T(y-x\beta)}{2}\right)\exp\left(-\tau\frac{\gamma_0}{2}\right)$$

$$= S_p\left(n+\delta_0,\tilde{\beta},\tilde{s}^2\left(X'X+(\sigma_0^2)^{-1}\right)^{-1}\right)$$

$$\tilde{\beta} = \Lambda\mu_0 + (I-\Lambda)\hat{\beta}$$

$$\Lambda = \left(X'X+(\sigma_0^2)^{-1}\right)^{-1} + (\sigma_0^2)^{-1}$$

$$\hat{\beta} = (X'X)^{-1}X'Y$$

$$\tilde{s}^2 = (n+\delta_0)^{-1}\left[Y'(I-M)Y + (\hat{\beta}-\mu_0)'(\Lambda'X'X)(\hat{\beta}-\mu_0) + \gamma_0\right]$$

So we have

$$\frac{\beta-\tilde{\beta}}{\sqrt{\tilde{s}^2\left(X'X+(\sigma_0^2)^{-1}\right)^{-1}\Big/n}} \sim t_{n+\delta_0}$$

Then the 95% HPD interval for β

$$\left\{\beta : \tilde{\beta} + t_{\frac{\alpha}{2},n+\delta_0}\sqrt{\tilde{s}^2\left(X'X+(\sigma_0^2)^{-1}\right)^{-1}\Big/n} \le \beta \le \tilde{\beta} + t_{1-\frac{\alpha}{2},n+\delta_0}\sqrt{\tilde{s}^2\left(X'X+(\sigma_0^2)^{-1}\right)^{-1}\Big/n}\right\}$$

$$\left\{\beta : -0.636 \le \beta \le 1.128\right\}$$

The R codes for calculating estimates:

```
X= c(1, 1.5, 2.5, 4)
Y = c(0.45, 0.6, 0.65, 0.89)
XX = t(X) %*% X
inv = solve(t(X) %*% X)
beta = solve(t(X) %*% X) * t(X) %*% Y
n = 4
delta0 = 4
gamma0 = 1
mu0 = 0
sigma0 = 1000
lambda = 1/(XX + 0.001) + 0.001
tbeta = (1 - lambda) * beta
M= X %*% inv %*% t(X)
s2 = t(Y) %*% Y - t(Y) %*% M %*% Y
ts2 = 1/(n+ delta0) *(s2 + beta^2*lambda*XX + gamma0)
LL = tbeta + qt(0.025, 8) *sqrt(ts2)
UU = tbeta + qt(0.975, 8) * sqrt(ts2)
```