Winter 21 – STAT 206B Homework 3

Due: Feb. 26th Friday

1. Let $X_1, \ldots, X_n$ be an i.i.d. sample such that $X_i \mid \theta, \sigma^2 \overset{iid}{\sim} \mathrm{N}(\theta, \sigma^2)$, where σ^2 is known and θ is unknown. Also, let your prior for θ be a mixture of conjugate priors, i.e.,

$$\pi(\theta) = \sum_{\ell=1}^{K} w_\ell \phi(\theta \mid \mu_\ell, \tau^2)$$

where $\phi(\theta \mid \mu_\ell, \tau^2)$ denotes the Gaussian density with mean μ_ℓ and variance τ^2 and mixture weights $0 < w_\ell < 1$ for all $\ell = 1, \ldots, K$ with $\sum_{\ell=1}^{K} w_\ell = 1$.

Note: This questions is challenging. Use the results from the class example with $X_i \mid \theta \overset{iid}{\sim} \mathrm{N}(\theta, \sigma^2)$ and $\theta \sim \mathrm{N}(\mu, \tau^2)$.

(a) Find the posterior distribution for θ based on this prior.

First note that with $\sigma^2 = 1$, $\bar{x} = \sum_{i=1}^{n} x_i$, and $s^2 = \sum_{i=1}^{n}(x_i - \bar{x})^2$, we write

$$f(\boldsymbol{x} \mid \theta) = (2\pi\sigma^2)^{-n/2} \exp\left(-\frac{s^2}{2\sigma^2} - \frac{n(\bar{x} - \theta)^2}{2\sigma^2}\right).$$

Denote π_ℓ the density of the ℓ-th component of the prior and $m_\ell(\boldsymbol{x})$ the corresponding marginal distribution. Then,

$$\begin{aligned}
\pi(\theta \mid \boldsymbol{x}) &\propto f(\boldsymbol{x} \mid \theta)\pi(\theta) = f(\boldsymbol{x} \mid \theta) \sum_{\ell=1}^{K} w_\ell \pi_\ell(\theta) = \sum_{\ell=1}^{K} w_\ell f(\boldsymbol{x} \mid \theta)\pi_\ell(\theta) \\
&= \sum_{\ell=1}^{K} w_\ell m_\ell(\boldsymbol{x}) \pi_\ell(\theta \mid \boldsymbol{x}).
\end{aligned}$$

The normalizing constant is

$$C = \int_{-\infty}^{\infty} \sum_{\ell=1}^{K} w_\ell m_\ell(\boldsymbol{x}) \pi_\ell(\theta \mid \boldsymbol{x}) d\theta = \sum_{\ell=1}^{K} w_\ell m_\ell(\boldsymbol{x}).$$

After normalization, we have

$$\begin{aligned}
\pi(\theta \mid \boldsymbol{x}) &= \sum_{\ell=1}^{K} \left\{ \frac{w_\ell m_\ell(\boldsymbol{x}) \pi_\ell(\theta \mid \boldsymbol{x})}{\sum_{j=1}^{K} w_j m_j(\boldsymbol{x})} \right\} \\
&= \sum_{\ell=1}^{K} w'_\ell(\boldsymbol{x}) \pi_\ell(\theta \mid \boldsymbol{x}), \quad \text{with } w'_\ell(\boldsymbol{x}) = \frac{w_\ell m_\ell(\boldsymbol{x})}{\sum_{j=1}^{K} w_j m_j(\boldsymbol{x})}. \qquad (1)
\end{aligned}$$

From class, $\pi_\ell(\theta \mid \boldsymbol{x})$ is $\mathrm{N}(\mu_{1\ell}, \tau_1^2)$, where

$$\tau_1^2 = \left(\frac{1}{\tau^2} + \frac{n}{\sigma^2}\right)^{-1} \quad \text{and} \quad \mu_{1\ell} = \tau_1^2\left(\frac{n\bar{x}}{\sigma^2} + \frac{\mu_\ell}{\tau^2}\right).$$

That is, the posterior of θ in (1) is also a mixture of K normal distributions. We next find $m_\ell(\boldsymbol{x})$;

$$\begin{aligned} m_\ell(\boldsymbol{x}) &= \int_{\mathbb{R}} \frac{1}{(\sqrt{2\pi\sigma})^{n/2}} \exp\left(-\frac{s^2}{2\sigma^2} - \frac{n(\theta-\bar{x})^2}{2\sigma^2}\right) \frac{1}{\sqrt{2\pi\tau^2}} \exp\left(-\frac{(\theta-\mu_\ell)^2}{2\tau^2}\right) d\theta \\ &= \frac{\sqrt{2\pi\tau_1^2}}{(\sqrt{2\pi\sigma})^{n/2}\sqrt{2\pi\tau^2}} \exp\left(-\frac{s^2}{2\sigma^2} - \frac{(\bar{x}-\mu_{1\ell})^2}{2(\tau^2+\sigma^2/n)}\right). \end{aligned}$$

We plug in $m_\ell(\boldsymbol{x})$ and find mixture weights in (1) as follows;

$$w'_\ell(\boldsymbol{x}) = \frac{w_\ell \exp\left(-\frac{(\bar{x}-\mu_{1\ell})^2}{2(\tau^2+\sigma^2/n)}\right)}{\sum_{j=1}^K w_j \exp\left(-\frac{(\bar{x}-\mu_{1j})^2}{2(\tau^2+\sigma^2/n)}\right)}.$$

(b) Find the posterior mean.

$$\mathrm{E}(\theta \mid \boldsymbol{x}) = \sum_{\ell=1}^K w'_\ell(\boldsymbol{x})\mu_{1\ell}.$$

(c) Find the prior predictive distribution associated with this model.

$$\begin{aligned} m(\boldsymbol{x}) &= \int_{-\infty}^{\infty} f(\boldsymbol{x} \mid \theta) \sum_{\ell=1}^K w_\ell \pi_\ell(\theta) d\theta \\ &= \sum_{\ell=1}^K w_\ell \int_{-\infty}^{\infty} f(\boldsymbol{x} \mid \theta)\pi_\ell(\theta) d\theta \\ &= \sum_{\ell=1}^K w_\ell m_\ell(\boldsymbol{x}). \end{aligned}$$

$m_\ell(\boldsymbol{x})$ is given in (a).

(d) Find the posterior predictive distribution associated with this model.

Due to the conditional independence,

$$\begin{aligned} f(y \mid \boldsymbol{x}) &= \int_{-\infty}^{\infty} f(y \mid \theta)\pi(\theta \mid \boldsymbol{x})d\theta \\ &= \sum_{\ell=1}^{K} w'_\ell(\boldsymbol{x}) \int_{-\infty}^{\infty} f(y \mid \theta)\phi(\theta \mid \mu_{1\ell}, \tau_1^2)d\theta \\ &= \sum_{\ell=1}^{K} w'_\ell(\boldsymbol{x}) \frac{1}{\sqrt{2\pi(\sigma^2+\tau_1^2)}} \exp\left(-\frac{(y-\mu_{1\ell})^2}{2(\sigma^2+\tau_1^2)}\right). \end{aligned}$$

(e) Simulate $n = 50$ i.i.d. observations from N$(4, 10)$. Let $K = 3$, $\tau^2 = 1$, $(\mu_1, \mu_2, \mu_3) = (-3, 0, 3)$ and $(\mu_1, \mu_2, \mu_3) = (1/3, 1/3, 1/3)$, and fit the above model to the simulated data. (i) draw samples of θ and make a histogram of the samples. (ii) draw samples of an unknown observable X from the posterior predictive distribution and make a histogram of the samples overlaid with the density of N$(4, 10)$. Choose option 'probability = TRUE' to make the histogram and the density plot comparable.

(f) Simulate $n = 50$ i.i.d. observations from N$(1.5, 10)$. Let $K = 3$, $\tau^2 = 1$, $(\mu_1, \mu_2, \mu_3) = (-3, 0, 3)$ and $(\mu_1, \mu_2, \mu_3) = (1/3, 1/3, 1/3)$, and fit the above model to the simulated data. (i) draw samples of θ and make a histogram of the samples. (ii) draw samples of an unknown observable X from the posterior predictive distribution and make a histogram of the samples overlaid with the density of N$(1.5, 10)$.

Panels (a) and (b) of Fig 1 illustrate the posterior distribution of θ and the posterior predictive distribution of an unobserved observation for Q1-(e), respectively. Panels (c) and (d) have the posterior distribution of θ and the posterior predictive distribution of an unobserved observation for Q1-(f), respectively. As shown in panels (a) and (c), the posterior distributions are concentrated around the true value of θ (blue dashed vertical lines) by updating the weights and locations of the mixture components. The posterior predictive distributions in panels (b) and (d) are close to the distribution of generating x, which indicates the model fits data well.

2. Let X be N$(0, \sigma^2)$. Assume that the unknown $1/\sigma^2$ has a gamma distribution with parameters $\alpha = r/2$ and $\beta = 2/r$, where r is a positive integer. Show that the marginal distribution of X is a $t-$distribution with r degrees of freedom.

Solution: We have $\sigma^2 \sim \text{IG}(r/2, r/2)$. We thus have

$$\begin{aligned} p(x, \sigma^2) &= \underbrace{\frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{x^2}{2\sigma^2}\right)}_{=p(x \mid \sigma^2)} \underbrace{\frac{(r/2)^{r/2}}{\Gamma(r/2)} (\sigma^2)^{-r/2-1} \exp\left(-\frac{r}{2\sigma^2}\right)}_{=p(\sigma^2)} \\ &= \frac{(r/2)^{r/2}}{\sqrt{2\pi}\Gamma(r/2)} (\sigma^2)^{-(r+1)/2-1} \exp\left(-\frac{x^2+r}{2\sigma^2}\right). \end{aligned}$$

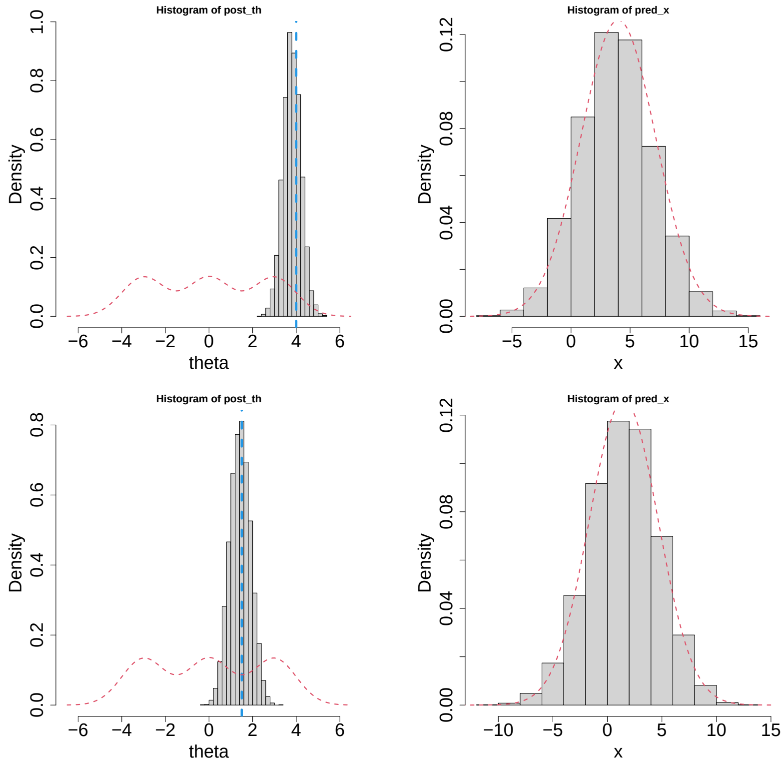


Figure 1: The posterior distribution of θ and the posterior predictive distribution of an unobserved observation for Q1-(e) are shown in panels (a) and (b), respectively. The posterior distribution of θ and the posterior predictive distribution of an unobserved observation for Q1-(f) are shown in panels (c) and (d), respectively.

We can obtain the marginal distribution of x by marginalizing over σ^2,

$$
\begin{aligned}
p(x) &= \int_{\mathbb{R}^+} p(x, \sigma^2) d\sigma^2 \\
&= \frac{(r/2)^{r/2}}{\sqrt{2\pi}\Gamma(r/2)} \int_{\mathbb{R}^+} \underbrace{(\sigma^2)^{-(r+1)/2-1} \exp\left(-\frac{x^2+r}{2\sigma^2}\right)}_{\text{a kernel for IG}((r+1)/2,(x^2+r)/2)} d\sigma^2 \\
&= \frac{\Gamma((r+1)/2)}{\sqrt{\pi r}\Gamma(r/2)} \left(1 + \frac{x^2}{r}\right)^{-(r+1)/2}.
\end{aligned}
$$

That is, $x \sim t_r$.

3. Let (X_1, X_2, X_3) have trinomial distribution with density

$$
f(x_1, x_2, x_3|\theta_1, \theta_2) \propto \theta_1^{x_1} \theta_2^{x_2} (1 - \theta_1 - \theta_2)^{x_3}.
$$

Derive Jeffreys prior for (θ_1, θ_2).

Solution: Let $\boldsymbol{\theta} = (\theta_1, \theta_2)$. The log likelihood in this case is given by

$$\ell(\boldsymbol{\theta}) = \log(p(x \mid \boldsymbol{\theta})) \propto x_1 \log(\theta_1) + x_2 \log(\theta_2) + x_3 \log(1 - \theta_1 - \theta_2).$$

Thus,

$$\frac{\partial \ell(\boldsymbol{\theta})}{\partial \theta_1} = \frac{x_1}{\theta_1} - \frac{x_3}{1-\theta_1-\theta_2} \text{ and } \frac{\partial \ell(\boldsymbol{\theta})}{\partial \theta_2} = \frac{x_2}{\theta_2} - \frac{x_3}{1-\theta_1-\theta_2},$$

and the Fisher information matrix

$$\begin{aligned} I(\boldsymbol{\theta}) &= -\mathrm{E}\left(\begin{bmatrix} \frac{\partial^2 \ell(\boldsymbol{\theta})}{\partial \theta_1^2} & \frac{\partial^2 \ell(\boldsymbol{\theta})}{\partial \theta_1 \partial \theta_2} \\ \frac{\partial^2 \ell(\boldsymbol{\theta})}{\partial \theta_1 \partial \theta_2} & \frac{\partial^2 \ell(\boldsymbol{\theta})}{\partial \theta_2^2} \end{bmatrix}\right) = \begin{bmatrix} \frac{E(X_1)}{\theta_1^2} + \frac{E(X_3)}{(1-\theta_1-\theta_2)^2} & \frac{E(X_3)}{(1-\theta_1-\theta_2)^2} \\ \frac{E(X_3)}{(1-\theta_1-\theta_2)^2} & \frac{E(X_2)}{\theta_2^2} + \frac{E(X_3)}{(1-\theta_1-\theta_2)^2} \end{bmatrix} \\ &= \begin{bmatrix} \frac{n}{\theta_1} + \frac{n}{1-\theta_1-\theta_2} & \frac{n}{1-\theta_1-\theta_2} \\ \frac{n}{1-\theta_1-\theta_2} & \frac{n}{\theta_2} + \frac{n}{1-\theta_1-\theta_2} \end{bmatrix} \end{aligned}$$

Thus, the Jeffrey's Prior is $\pi^J(\boldsymbol{\theta}) \propto \sqrt{\det I(\boldsymbol{\theta})} \propto \theta_1^{-1/2}\theta_2^{-1/2}(1-\theta_1-\theta_2)^{-1/2}$. That is, a Dirichlet(1/2,1/2,1/2).

4. (Robert Problem 3.9) Let $x \mid \theta \sim \mathrm{Bin}(n, \theta)$ and $\theta \sim \mathrm{Be}(\alpha, \beta)$. Determine whether there exits values of α, β such that $\pi(\theta \mid x)$ is the uniform prior on [0, 1], even for a single value of x.

Solution: We have

$$\theta \mid x \sim \mathrm{Be}(x + \alpha, n - x + \beta).$$

since

$$f(\theta \mid x) \propto f(x \mid \theta)\pi(\theta) \propto \theta^{x+\alpha-1}(1-\theta)^{n+\beta-x-1}.$$

Then, the posterior would become uniform if $x + \alpha = 1$ and $n + \beta - x = 1$. But notice that x being a non-negative integer, along with $0 < \alpha = 1 - x$, implies that $x = 0$. In that case, $0 < \beta = 1 - n \leq 0$, a contradiction. Thus, there is no combination of $\alpha > 0$ and $\beta > 0$ for the posterior to be the uniform on 0 and 1.

5. (Robert Problem 3.10) Let $x \mid \theta \sim \mathrm{Pa}(\alpha, \theta)$, a Pareto distribution, and $\theta \sim \mathrm{Be}(\mu, \nu)$. Show that if $\alpha < 1$ and $x > 1$, a particular choice of μ and ν gives $\pi(\theta \mid x)$ as the uniform prior on [0,1].

Solution: The posterior satisfies

$$f(\theta \mid x) \propto f(x \mid \theta)\pi(\theta) \propto \theta^{\alpha+\mu-1}(1-\theta)^{\nu-1} I_{[\theta,\infty)}(x) I_{[0,1]}(\theta)$$

and, in particular, for $x > 1$

$$f(\theta \mid x) \propto \theta^{\alpha+\mu-1}(1-\theta)^{\nu-1} I_{[0,1]}(\theta),$$

That is, $\theta \mid x \sim \text{Be}(\alpha+\mu, \nu)$ with $x > 1$. It becomes Unif$(0,1)$ whenever $\nu = 1$ and $\mu = 1-\alpha$.

6. (Robert Problem 3.31) Consider $x \mid \theta \sim \text{N}(\theta, \theta)$ with $\theta > 0$.

   (a) Determine the Jeffreys prior $\pi^J(\theta)$.

   Solution: The likelihood in this case is given by

$$f(x \mid \theta) \quad = \quad (2\pi\theta)^{-\frac{1}{2}} \exp\left\{-\frac{(x-\theta)^2}{2\theta}\right\}.$$

   So

$$\ell(\theta) = \log(f(x \mid \theta)) = -1/2 \log(2\pi\theta) - \frac{(x-\theta)^2}{2\theta},$$

   and taking first and second derivatives,

$$\ell'(\theta) = -\frac{1}{2\theta} - \frac{\theta^2 - x^2}{2\theta^2} \qquad \text{and} \qquad \ell''(\theta) = \frac{1}{2\theta^2} - \frac{x^2}{\theta^3}.$$

   Then,

$$I(\theta) = -\text{E}[\ell''(\theta)] = -\frac{1}{2\theta^2} + \frac{E[X^2 \mid \theta]}{\theta^3} = -\frac{1}{2\theta^2} + \frac{\theta+\theta^2}{\theta^3} = \frac{2\theta+1}{2\theta^2},$$

   and

$$\pi^J(\theta) \propto \{I(\theta)\}^{1/2} = \left(\frac{2\theta+1}{\theta^2}\right)^{1/2}.$$

   (b) Say whether the distribution of x belongs to an exponential family and derive the conjugate priors on θ.

   Solution:

$$\begin{aligned} f(x \mid \theta) &= (2\pi\theta)^{-\frac{1}{2}} \exp\left\{-\frac{(x-\theta)^2}{2\theta}\right\} \\ &= (2\pi)^{-\frac{1}{2}} \exp(x) \exp\left[-\frac{x^2}{2} \cdot \frac{1}{\theta} - \frac{1}{2}\left\{\log(\theta) + \theta\right\}\right], \end{aligned}$$

   which is a member of the exponential family with $h(x) = (2\pi)^{-\frac{1}{2}} \exp(x)$, $t(x) = \frac{-x^2}{2}$, $\eta(\theta) = \frac{1}{\theta}$ and $\psi(\eta) = \frac{1}{2}\left\{\frac{1}{\eta} - \log(\eta)\right\}$. Then, the conjugate prior satisfies

$$\pi(\eta|\mu, \lambda) \propto \exp\{\eta\mu - \lambda\psi(\eta)\}$$

and, transforming back to θ,

$$\begin{aligned}\pi(\theta \mid \mu, \lambda) &\propto \exp\left\{\frac{\mu}{\theta} - \frac{\lambda}{2}\left[\theta + \log(\theta)\right]\right\} \cdot \frac{1}{\theta^2} \\ &= \theta^{-\frac{\lambda}{2}-2} \exp\left(\mu\frac{1}{\theta} - \frac{\lambda}{2}\theta\right).\end{aligned}$$

(c) Use Proposition 3.3.14 to relate the hyperparameters of the conjugate priors with the mean of θ.

Solution: Let $z = t(x) = -\frac{x^2}{2}$ and we find that

$$\xi(\eta) = \mathbb{E}\left[z \mid \eta\right] = \frac{d\psi(\eta)}{d\eta} = -\frac{1}{2}\left(\frac{1}{\eta} + \frac{1}{\eta^2}\right).$$

Then by proposition 3.3.14 we have that:

$$\mathbb{E}\left[\xi(\eta) \mid z\right] = \mathbb{E}\left[-\frac{1}{2}\left(\frac{1}{\eta} + \frac{1}{\eta^2}\right) \mid z\right] = \frac{\mu + z}{\lambda + 1},$$

which implies that

$$\mathbb{E}[\theta + \theta^2 \mid x] = \frac{x^2 - 2\mu}{\lambda + 1}.$$

7. Consider a model of the form $x \mid \theta \sim \text{Bin}(n, \theta)$ and $\theta \sim \text{Be}(1/2, 1/2)$. Assume that you observe $n = 10$ and $x = 1$.

(a) Report an exact 95% (symmetric) posterior credible interval for θ (for example, you can use the `qbeta` function in `R`).

Solution: We have the posterior $\theta \mid x \sim \text{Be}(x + 1/2, n - x + 1/2)$.

```
> #### (a) exact posterior
> c(qbeta(.025, a_post, b_post),qbeta(.975, a_post, b_post))
[1] 0.01101167 0.38131477
```

(b) Report an approximate credible interval for θ using the Laplace approximation.

Solution: First, let $q(\theta) = \theta^{(a-1)}(1-\theta)^{(b-1)}$ and MAP θ_0 is such that $q'(\theta_0) = 0$ with $h(\theta) = \log(q(\theta))$. We find $\theta_0 = (a-1)/(a+b-2)$. We also find $A = -d^2(h(\theta))/d\theta^2|_{\theta=\theta_0} =$

$(a+b-2)^3/\{(a-1)(b-1)\}$. The Laplace approximation gives

$$\begin{aligned} p(\theta \mid x) &= \frac{1}{B(a,b)}\theta^{a-1}(1-\theta)^{b-1} \\ &\approx \frac{1}{B(a,b)}q(\theta_0)\sqrt{\frac{2\pi}{A}}\phi(\theta \mid \theta_0, A^{-1}), \end{aligned}$$

where $\phi(\theta \mid \theta_0, A^{-1})$ denotes a pdf of $\mathrm{N}(\theta_0, A^{-1})$.
We then find (ℓ_1, ℓ_2) such that

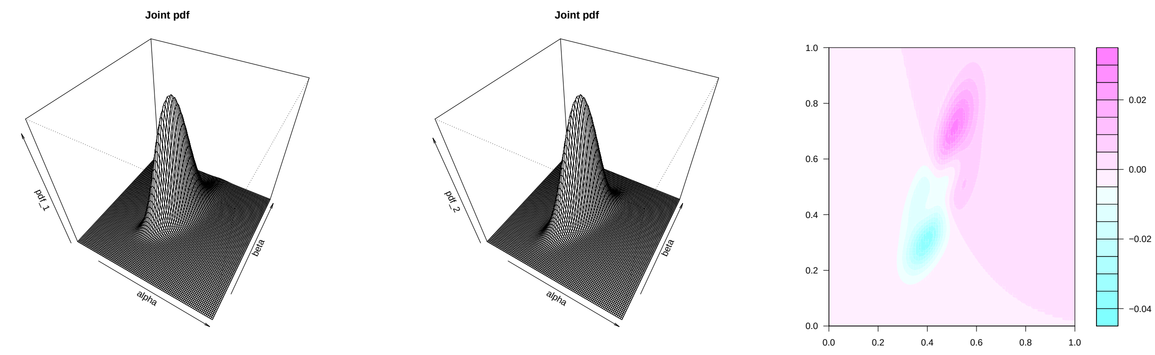
$$\frac{1}{B(a,b)}q(\theta_0)\sqrt{\frac{2\pi}{A}}\left\{\Phi(\ell_2 \mid \theta_0, A^{-1}) - \Phi(\ell_1 \mid \theta_0, A^{-1})\right\} = 0.95.$$

```
> #### (b) Laplace approximation
> # calculate the MAP - th0
> th0 <- (a_post - 1)/(a_post + b_post - 2)
>
> # evaluate the first and second derivaties at the MAP
> q0 <- exp((a_post - 1)*log(th0) + (b_post - 1)*log(1 - th0))
> A_inv <- (a_post - 1)*(b_post - 1)/(a_post + b_post - 2)^3
>
> #calculate the constant
> Const <- q0*sqrt(2*pi*A_inv)/beta(a_post, b_post)
>
>
> # compute the interval
> lower <- qnorm((1 - 0.95/Const)/2, th0, sqrt(A_inv))
> upper <- qnorm((1 - 0.95/Const)/2, th0, sqrt(A_inv), lower.tail=FALSE)
>
> c(lower, upper)
[1] -0.1736783  0.2847894
```

(c) Report an approximate credible interval for θ using Monte Carlo simulation.

Solution: A Monte Carlo sample of size 2000 is used to approximate 95% credible interval of $\mathrm{Be}(x+1/2, n-x+1/2)$.

```
> #### (c) Monte Carlo simulation
> th <- rbeta(2000, a_post, b_post)
> quantile(th, probs=c(0.025, 0.975))
      2.5%      97.5%
0.01185477 0.37301624
```



(a) Exact (b) Laplace Approx (c) Difference

Figure 2: (a) Exact evaluation of the posterior distribution; (b) Laplace approximation; (c) Difference, $p(\alpha, \beta \mid x) - \hat{p}(\alpha, \beta \mid x)$.

(d) Repeat the previous calculations with $n = 100, x = 10$ and with $n = 1000, x = 100$. Comment on the difference between all 9 situations.

Solution: The results are summarized in the table below;

(n, x)	$(10, 1)$	$(100, 10)$	$(1000, 100)$
Exact	(0.011, 0.381)	(0.053, 0.170)	(0.083, 0.120)
Laplace approx.	(-0.174, 0.285)	(0.038, 0.154)	(0.081, 0.118)
Monte Carlo approx.	(0.012, 0.373)	(0.053, 0.168)	(0.083, 0.120)

The Monte Carlo (MC) approximation with a fairly large MC sample size (2000 is used) gives accurate estimates of the credible interval (CI) for all cases with $n = 10, 100, 1000$. As n increase, the posterior distribution becomes centered at 0.1 more sharply. Thus, the approximate intervals, especially the interval by the Laplace approximation, are improved significantly.

8. Let $x_1, \ldots, x_n$ be an i.i.d. sample from a Gumbel type-II distribution with density

$$p(x \mid \alpha, \beta) = \alpha\beta x^{-\alpha-1} \exp(-\beta x^{-\alpha}),$$

with $\alpha, \beta > 0$. Let $\pi(\alpha, \beta) = 1$ for $\alpha, \beta > 0$ be the prior distribution. Simulate $n = 500$ i.i.d. observations from the Gumbel type-II distribution with $\alpha = \beta = 5$ (may use the inverse CDF method to generate random draws).

(a) Find the posterior $p(\alpha, \beta \mid x_1, \ldots, x_n)$. Use the simulated dataset and evaluate the posterior density in the grid of (α, β). You may find the normalizing constant numerically.

Solution: The evaluation of the posterior $p(\alpha, \beta \mid \boldsymbol{x})$ is shown in Figure 2(a). See the code in a separately attached.

(b) Find the Laplace approximation to the posterior $p(\alpha, \beta \mid x_1, \ldots, x_n)$. Observe the mode can be found using some numerical method. Use the same simulated data and evaluate the approximated posterior on the grid of (α, β).

Solution: The evaluation of the posterior $p(\alpha, \beta \mid \boldsymbol{x})$ using the Laplace approximation is shown in Figure 2(b). See the code in a separately attached.

(c) Compare (a) and (b) and comment.

Solution: The difference $p(\alpha, \beta \mid \boldsymbol{x}) - \hat{p}(\alpha, \beta \mid \boldsymbol{x})$ is illustrated in Figure 2(c), where $\hat{p}(\alpha, \beta \mid \boldsymbol{x})$ denotes the evaluation using the Laplace approximation. Since a normal distribution is used for the Laplace approximation, as (α, β) far from their model, $\hat{p}(\alpha, \beta \mid \boldsymbol{x})$ is smaller.

9. Let $x_1, \ldots, x_n$ be an i.i.d. sample such that $x_i \mid \theta, \sigma^2 \sim \mathrm{N}(\theta, \sigma^2)$ with θ and σ^2 unknown. Assume a conjugate normal-inverse-gamma prior on (θ, σ^2) such that $\theta \mid \sigma^2 \sim \mathrm{N}(\theta_0, \kappa_0 \sigma^2)$ and $\sigma^2 \sim \mathrm{IG}(a, b)$ with θ_0, κ_0, a and b known.

(a) Find the joint posterior $p(\theta, \sigma^2 \mid x_1, \ldots, x_n)$ (up to proportionality).

Solution: We have

$$\begin{aligned} p(\theta, \sigma^2 \mid \boldsymbol{x}) &\propto (\sigma^2)^{-n/2} \exp\left\{-\frac{s^2}{2\sigma^2} - \frac{n(\bar{x}-\theta)^2}{2\sigma^2}\right\} \\ &\quad \times (\sigma^2)^{-1/2} \exp\left\{-\frac{(\theta-\theta_0)^2}{2\kappa_0\sigma^2}\right\} (\sigma^2)^{-a-1} \exp\left(-\frac{b}{\sigma^2}\right) \\ &= (\sigma^2)^{-(n+1)/2-a-1} \exp\left\{-\frac{s^2}{2\sigma^2} - \frac{n(\bar{x}-\theta)^2}{2\sigma^2} - \frac{(\theta-\theta_0)^2}{2\kappa_0\sigma^2} - \frac{b}{\sigma^2}\right\}, \end{aligned}$$

where $s^2 = \sum_{i=1}^n (x_i - \bar{x})^2$ and $\bar{x} = 1/n \sum_{i=1}^n x_i$.

(b) Find $p(\theta \mid \sigma^2, x_1, \ldots, x_n)$.

Solution: We have

$$\begin{aligned} p(\theta \mid \sigma^2, \boldsymbol{x}) &\propto \exp\left\{-\frac{n(\bar{x}-\theta)^2}{2\sigma^2} - \frac{(\theta-\theta_0)^2}{2\kappa_0\sigma^2}\right\} \\ &\propto \exp\left[-\frac{1}{2}\left\{\left(\frac{n}{\sigma^2} + \frac{1}{\kappa_0\sigma^2}\right)\theta^2 - \frac{2\theta}{\sigma^2}\left(n\bar{x} + \frac{\theta_0}{\kappa_0}\right)\right\}\right]. \end{aligned}$$

We thus have

$$\theta \mid \sigma^2, \boldsymbol{x} \sim \mathrm{N}\left(\left(n + \frac{1}{\kappa_0}\right)^{-1}\left(n\bar{x} + \frac{\theta_0}{\kappa_0}\right), \left(\frac{n}{\sigma^2} + \frac{1}{\kappa_0\sigma^2}\right)^{-1}\right).$$

(c) Find $p(\sigma^2 \mid x_1, \ldots, x_n)$.

Solution: We have

$$\begin{aligned}
p(\sigma^2 \mid \boldsymbol{x}) &= \int_{\mathbb{R}} p(\theta, \sigma^2 \mid \boldsymbol{x}) d\theta \\
&\propto \int_{\mathbb{R}} (\sigma^2)^{-(n+1)/2-a-1} \exp\left\{-\frac{s^2}{2\sigma^2} - \frac{n(\bar{x}-\theta)^2}{2\sigma^2} - \frac{(\theta-\theta_0)^2}{2\kappa_0\sigma^2} - \frac{b}{\sigma^2}\right\} d\theta \\
&= (\sigma^2)^{-(n+1)/2-a-1} \exp\left\{-\frac{s^2}{2\sigma^2} - \frac{b}{\sigma^2}\right\} \int_{\mathbb{R}} \exp\left\{-\frac{n(\bar{x}-\theta)^2}{2\sigma^2} - \frac{(\theta-\theta_0)^2}{2\kappa_0\sigma^2}\right\} d\theta \\
&= (\sigma^2)^{-n/2-a-1} \exp\left\{-\frac{s^2}{2\sigma^2} - \frac{b}{\sigma^2} - \frac{(\bar{x}-\theta_0)^2}{2(n\kappa_0+1)\sigma^2/n}\right\},
\end{aligned}$$

We thus have

$$\sigma^2 \mid \boldsymbol{x} \sim \mathrm{IG}\left(\frac{n}{2} + a, \frac{s^2}{2} + \frac{n(\bar{x}-\theta_0)^2}{2(n\kappa_0+1)} + b\right)$$

(d) Find $p(\theta \mid x_1, \ldots, x_n)$.

Solution: We have

$$\begin{aligned}
p(\theta \mid \boldsymbol{x}) &= \int_{\mathbb{R}+} p(\theta, \sigma^2 \mid \boldsymbol{x}) d\sigma^2 \\
&\propto \int_{\mathbb{R}+} (\sigma^2)^{-(n+1)/2-a-1} \exp\left\{-\frac{s^2}{2\sigma^2} - \frac{n(\bar{x}-\theta)^2}{2\sigma^2} - \frac{(\theta-\theta_0)^2}{2\kappa_0\sigma^2} - \frac{b}{\sigma^2}\right\} d\sigma^2 \\
&\propto \left\{\frac{s^2}{2} + \frac{n(\bar{x}-\theta)^2}{2} + \frac{(\theta-\theta_0)^2}{2\kappa_0} + b\right\}^{-\frac{n+1}{2}-a} \\
&\propto \left[1 + \frac{n+1/\kappa_0}{s^2 + 2b + \frac{n(\bar{x}-\theta_0)^2}{(n\kappa_0+1)}}\left\{\theta - \left(n + \frac{1}{\kappa_0}\right)^{-1}\left(n\bar{x} + \frac{\theta_0}{\kappa_0}\right)\right\}^2\right]^{-\frac{n+2a+1}{2}}.
\end{aligned}$$

That is, $p(\theta \mid \boldsymbol{x})$ is a t-distribution with d.f. $n+2a$, where the location and scale parameters are

$$\left(n + \frac{1}{\kappa_0}\right)^{-1}\left(n\bar{x} + \frac{\theta_0}{\kappa_0}\right), \text{ and } \left\{\frac{(2a+n)(n+1/\kappa_0)}{s^2 + 2b + \frac{n(\bar{x}-\theta_0)^2}{(n\kappa_0+1)}}\right\}^{-1/2}, \text{ respectively.}$$

(e) Simulate $n = 1000$ i.i.d. observations from a $\mathrm{N}(5, 1)$. Fit the above model to these data assuming the following prior scenarios: (i) fairly informative priors around the true values of both parameters, (ii) informative prior on θ and vague on σ^2 (iii) informative prior on

σ^2 and vague on θ (iv) vague on both parameters. Specify the form of your posteriors in each case.

Solution: We specify values of the fixed hyperparameters μ_0, κ_0, a and b to express strong prior information or weak prior information. For more explanation on strong prior and vague (weakly informative) prior, read `https://statmodeling.stat.columbia.edu/2007/07/18/informative_and/`. An example is the values below;

cases	θ_0	κ_0	a	b
i	5	0.01	1001	1000
ii	5	0.01	0.1	0.1
iii	5	100	1001	1000
v	5	100	0.1	0.1

The prior means of θ and σ^2 are at their true values, $\theta^{\text{true}} = 5$ and $\sigma^{2,\text{true}} = 1$. The prior variances (i.e., prior uncertainty about the prior center) are assumed to be large for vague priors, and small for informative priors.

(f) Assume that you are interested in estimating $\eta = \theta/\sigma$. Develop a Monte Carlo algorithm for computing the posterior mean and a 95% credible interval for η. Use the algorithm to compute such quantities under all the prior scenarios described above.

Solution: For each of the prior specifications, a 95% credible interval of $\eta = \theta/\sigma$ is computed;

cases	θ_0	κ_0	a	b	Post. mean	Post. 95% CI
i	5	0.01	1001	1000	5.009	(4.867, 5.146)
ii	5	0.01	0.1	0.1	4.978	(4.760, 5.208)
iii	5	100	1001	1000	5.011	(4.870, 5.150)
v	5	100	0.1	0.1	5.012	(4.874, 5.149)

As can be seen there, because the sample size ($n = 1000$) is fairly large, the the prior distribution does not have a huge impact on the posterior. Having said that, the posterior mean of η is close to its true value, $\eta^{\text{true}} = \theta^{\text{true}}/\sigma^{2,\text{true}} = 5$, under the four different prior calibration. Interestingly, when the prior distribution on σ^2 is informative, the credible interval is shorter. This is due to the fact that σ^2 also controls the dispersion of the prior of θ.

The simulation was re-ran with $n = 10$;

cases	θ_0	κ_0	a	b	Post. mean	Post. 95% CI
i	5	0.01	1001	1000	5.019	(4.777, 5.256)
ii	5	0.01	0.1	0.1	4.049	(2.390, 5.962)
iii	5	100	1001	1000	5.255	(4.628, 5.893)
v	5	100	0.1	0.1	5.260,	(4.628, 5.905)

As shown the table above, when the sample size is small, the prior affects the posterior. When any of the priors is vague, the CI tends to be wider and the posterior mean to be away of its true value.

10. Consider the usual regression model, $y_i = \boldsymbol{x}_i\boldsymbol{\beta} + \epsilon_i$, where $\epsilon_i \overset{iid}{\sim} \mathrm{N}(0, \sigma^2)$, covariates $\boldsymbol{x}_i = (x_{i1}, \ldots, x_{ip})$ and a coefficient vector $\boldsymbol{\beta} \in \mathbb{R}^p$.

Fact: a sufficient statistic is $\hat{\boldsymbol{\beta}} = (\boldsymbol{X}^t\boldsymbol{X})^{-1}\boldsymbol{X}^t\boldsymbol{y}$, where $\boldsymbol{X}$ is a $n \times p$ matrix whose ith row is $\boldsymbol{x}_i$, and $\boldsymbol{y}$ is the $n-$dimensional column vector of $(y_1, \ldots, y_n)$. It is also the maximum likelihood estimator and the least-squares estimator of $\boldsymbol{\beta}$.

Using the fact above, we write the likelihood as follows;

$$
\begin{aligned}
p(\boldsymbol{y} \mid \boldsymbol{\beta}, \boldsymbol{X}, \sigma^2) &= \frac{1}{\sqrt{2\pi}(\sigma^2)^{n/2}} \exp\left\{-\frac{1}{2\sigma^2}(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta})^t(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta})\right\} \\
&= \frac{1}{\sqrt{2\pi}(\sigma^2)^{n/2}} \exp\left\{-\frac{1}{2\sigma^2}(\boldsymbol{y} \pm \boldsymbol{X}\hat{\boldsymbol{\beta}} - \boldsymbol{X}\boldsymbol{\beta})^t(\boldsymbol{y} \pm \boldsymbol{X}\hat{\boldsymbol{\beta}} - \boldsymbol{X}\boldsymbol{\beta})\right\} \\
&= \frac{1}{\sqrt{2\pi}(\sigma^2)^{n/2}} \exp\left\{-\frac{1}{2\sigma^2}\underbrace{(\boldsymbol{y} - \boldsymbol{X}\hat{\boldsymbol{\beta}})^t(\boldsymbol{y} - \boldsymbol{X}\hat{\boldsymbol{\beta}})}_{=s^2} - \frac{1}{2\sigma^2}(\boldsymbol{\beta} - \hat{\boldsymbol{\beta}})^t(\boldsymbol{X}^t\boldsymbol{X})(\boldsymbol{\beta} - \hat{\boldsymbol{\beta}})\right\} \\
&= \frac{1}{\sqrt{2\pi}(\sigma^2)^{n/2}} \exp\left\{-\frac{s^2}{2\sigma^2} - \frac{1}{2\sigma^2}(\boldsymbol{\beta} - \hat{\boldsymbol{\beta}})^t(\boldsymbol{X}^t\boldsymbol{X})(\boldsymbol{\beta} - \hat{\boldsymbol{\beta}})\right\},
\end{aligned}
$$

where $\boldsymbol{y}$ is a $n-$dim vector of y_i and $\boldsymbol{X}$ a $n \times p$ dim matrix.

(a) Consider the following priors on $(\boldsymbol{\beta}, \sigma^2)$;

$$
\begin{aligned}
\boldsymbol{\beta} \mid \sigma^2 &\sim \mathrm{N}_p\left(\boldsymbol{\beta}_0, \frac{\sigma^2}{n_0}(\boldsymbol{X}^t\boldsymbol{X})^{-1}\right), \\
\sigma^2 &\sim \mathrm{IG}(\nu/2, s_0^2/2).
\end{aligned}
$$

i. Find an expression of the joint posterior distribution as follows;

$$\pi(\boldsymbol{\beta}, \sigma^2 \mid \boldsymbol{y}, \boldsymbol{X}) = \pi_1(\boldsymbol{\beta} \mid \sigma^2, \boldsymbol{y}, \boldsymbol{X})\pi_2(\sigma^2 \mid \boldsymbol{y}, \boldsymbol{X}).$$

Also, identify $\pi_1(\boldsymbol{\beta} \mid \sigma^2, \boldsymbol{y}, \boldsymbol{X})$ and $\pi_2(\sigma^2 \mid \boldsymbol{y}, \boldsymbol{X})$.

Solution: The joint posterior distribution is

$$
\begin{aligned}
p(\boldsymbol{\beta}, \sigma^2 \mid \boldsymbol{y}, \boldsymbol{X}) &\propto p(\boldsymbol{y} \mid \boldsymbol{\beta}, \boldsymbol{X}, \sigma^2)\pi(\boldsymbol{\beta}, \sigma^2) \\
&\propto (\sigma^2)^{-n/2} \exp\left\{-\frac{s^2}{2\sigma^2} - \frac{1}{2\sigma^2}(\boldsymbol{\beta} - \hat{\boldsymbol{\beta}})^t(\boldsymbol{X}^t\boldsymbol{X})(\boldsymbol{\beta} - \hat{\boldsymbol{\beta}})\right\} \\
&\quad \times (\sigma^2)^{-p/2} \exp\left\{-\frac{1}{2\sigma^2}(\boldsymbol{\beta} - \boldsymbol{\beta}_0)^t(\boldsymbol{X}^t\boldsymbol{X})(\boldsymbol{\beta} - \boldsymbol{\beta}_0)\right\} \\
&\quad \times (\sigma^2)^{-\nu/2-1} \exp\left(-\frac{s_0^2}{2\sigma^2}\right)
\end{aligned}
$$

We then have

$$
\begin{aligned}
\pi_1(\boldsymbol{\beta} \mid \sigma^2, \boldsymbol{y}, \boldsymbol{X}) &\propto p(\boldsymbol{\beta}, \sigma^2 \mid \boldsymbol{y}, \boldsymbol{X}) \\
&\propto \exp\left\{-\frac{1}{2\sigma^2}(\boldsymbol{\beta} - \hat{\boldsymbol{\beta}})^t(\boldsymbol{X}^t\boldsymbol{X})(\boldsymbol{\beta} - \hat{\boldsymbol{\beta}})\right\} \\
&\quad \times \exp\left\{-\frac{1}{2\sigma^2}(\boldsymbol{\beta} - \boldsymbol{\beta}_0)^t(\boldsymbol{X}^t\boldsymbol{X})(\boldsymbol{\beta} - \boldsymbol{\beta}_0)\right\}.
\end{aligned}
$$

So we have

$$\boldsymbol{\beta} \mid \sigma^2, \boldsymbol{y}, \boldsymbol{X} \sim \mathrm{N}_p\left(\frac{n_0\boldsymbol{\beta}_0 + \hat{\boldsymbol{\beta}}}{n_0+1}, \frac{\sigma^2}{n_0+1}(\boldsymbol{X}^t\boldsymbol{X})^{-1}\right).$$

We marginalize $\boldsymbol{\beta}$ out and find the marginal posterior of σ^2,

$$\begin{aligned}
\pi_2(\sigma^2 \mid \boldsymbol{y}, \boldsymbol{X}) &= \int \pi(\boldsymbol{\beta}, \sigma^2 \mid \boldsymbol{y}, \boldsymbol{X}) d\boldsymbol{\beta} \\
&\propto \int (\sigma^2)^{-n/2} \exp\left\{-\frac{s^2}{2\sigma^2} - \frac{1}{2\sigma^2}(\boldsymbol{\beta} - \hat{\boldsymbol{\beta}})^t(\boldsymbol{X}^t\boldsymbol{X})(\boldsymbol{\beta} - \hat{\boldsymbol{\beta}})\right\} \\
&\quad \times (\sigma^2)^{-p/2} \exp\left\{-\frac{1}{2\sigma^2}(\boldsymbol{\beta} - \boldsymbol{\beta}_0)^t(\boldsymbol{X}^t\boldsymbol{X})(\boldsymbol{\beta} - \boldsymbol{\beta}_0)\right\} \\
&\quad \times (\sigma^2)^{-\nu/2-1} \exp\left(-\frac{s_0^2}{2\sigma^2}\right) d\boldsymbol{\beta} \\
&\propto (\sigma^2)^{-(n+\nu)/2-1} \exp\left\{-\frac{1}{2\sigma^2}\left(s_0^2 + s^2 + \frac{n_0}{n_0+1}(\boldsymbol{\beta}_0 - \hat{\boldsymbol{\beta}})^t(\boldsymbol{X}^t\boldsymbol{X})(\boldsymbol{\beta}_0^t - \hat{\boldsymbol{\beta}})\right)\right\}.
\end{aligned}$$

We recognize

$$\sigma^2 \mid \boldsymbol{y}, \boldsymbol{X} \sim \mathrm{IG}\left(\frac{n+\nu}{2}, \frac{1}{2}\left(s_0^2 + s^2 + \frac{n_0}{n_0+1}(\boldsymbol{\beta}_0 - \hat{\boldsymbol{\beta}})^t(\boldsymbol{X}^t\boldsymbol{X})(\boldsymbol{\beta}_0^t - \hat{\boldsymbol{\beta}})\right)\right).$$

ii. Is the prior conjugate? Explain.

Solution: The joint posterior of $\boldsymbol{\beta}$ and σ^2 is decomposed into $\pi_1(\boldsymbol{\beta} \mid \sigma^2, \boldsymbol{y}, \boldsymbol{X})$ and $\pi_2(\sigma^2 \mid \boldsymbol{y}, \boldsymbol{X})$, where each has the same distribution family. So the prior is conjugate.

(b) Consider the following priors on $(\boldsymbol{\beta}, \sigma^2)$;

$$\begin{aligned}
\boldsymbol{\beta} &\sim \mathrm{N}_p(\boldsymbol{\beta}_0, \Sigma_0), \\
\sigma^2 &\sim \mathrm{IG}(\nu/2, s_0^2/2).
\end{aligned}$$

i. Find the conditional posterior distribution of $\boldsymbol{\beta}$ given σ^2, $\pi_1(\boldsymbol{\beta} \mid \sigma^2, \boldsymbol{y}, \boldsymbol{X})$.

Solution: The joint posterior distribution is

$$\begin{aligned}
p(\boldsymbol{\beta}, \sigma^2 \mid \boldsymbol{y}, \boldsymbol{X}) &\propto p(\boldsymbol{y} \mid \boldsymbol{\beta}, \boldsymbol{X}, \sigma^2)\pi(\boldsymbol{\beta}, \sigma^2) \\
&\propto (\sigma^2)^{-n/2} \exp\left\{-\frac{s^2}{2\sigma^2} - \frac{1}{2\sigma^2}(\boldsymbol{\beta} - \hat{\boldsymbol{\beta}})^t(\boldsymbol{X}^t\boldsymbol{X})(\boldsymbol{\beta} - \hat{\boldsymbol{\beta}})\right\} \\
&\quad \times \exp\left\{-\frac{1}{2}(\boldsymbol{\beta} - \boldsymbol{\beta}_0)^t\Sigma_0^{-1}(\boldsymbol{\beta} - \boldsymbol{\beta}_0)\right\} \\
&\quad \times (\sigma^2)^{-\nu/2-1} \exp\left(-\frac{s_0^2}{2\sigma^2}\right).
\end{aligned}$$

We then have

$$\begin{aligned} p(\boldsymbol{\beta} \mid \sigma^2, \boldsymbol{y}, \boldsymbol{X}) &\propto p(\boldsymbol{\beta}, \sigma^2 \mid \boldsymbol{y}, \boldsymbol{X}) \\ &\propto \exp\left\{-\frac{1}{2\sigma^2}(\boldsymbol{\beta} - \hat{\boldsymbol{\beta}})^t(\boldsymbol{X}^t\boldsymbol{X})(\boldsymbol{\beta} - \hat{\boldsymbol{\beta}}) - \frac{1}{2}(\boldsymbol{\beta} - \boldsymbol{\beta}_0)^t\Sigma_0^{-1}(\boldsymbol{\beta} - \boldsymbol{\beta}_0)\right\}. \end{aligned}$$

We recognize a kernel of a normal distribution, and obtain

$$\boldsymbol{\beta} \mid \sigma^2, \boldsymbol{y}, \boldsymbol{X} \sim \mathrm{N}_p\left(\left(\frac{1}{\sigma^2}(X^tX) + \Sigma_0^{-1}\right)^{-1}\left(\frac{1}{\sigma^2}(X^tX)\hat{\boldsymbol{\beta}} + \Sigma_0^{-1}\boldsymbol{\beta}_0\right), \left(\frac{1}{\sigma^2}(X^tX) + \Sigma_0^{-1}\right)^{-1}\right).$$

ii. Find the conditional posterior distribution of σ^2 given $\boldsymbol{\beta}$, $\pi_2(\sigma^2 \mid \boldsymbol{\beta}, \boldsymbol{y}, \boldsymbol{X})$.

Solution: Similar to the solution of part (i), we have

$$\begin{aligned} p(\sigma^2 \mid \boldsymbol{\beta}, \boldsymbol{y}, \boldsymbol{X}) &\propto p(\boldsymbol{\beta}, \sigma^2 \mid \boldsymbol{y}, \boldsymbol{X}) \\ &\propto (\sigma^2)^{-n/2}\exp\left\{-\frac{s^2}{2\sigma^2} - \frac{1}{2\sigma^2}(\boldsymbol{\beta} - \hat{\boldsymbol{\beta}})^t(\boldsymbol{X}^t\boldsymbol{X})(\boldsymbol{\beta} - \hat{\boldsymbol{\beta}})\right\} \\ &\quad \times(\sigma^2)^{-\nu/2-1}\exp\left(-\frac{s_0^2}{2\sigma^2}\right). \end{aligned}$$

We recognize a kernel of an inverse gamma distribution, and we obtain

$$\sigma^2 \mid \boldsymbol{\beta}, \boldsymbol{y}, \boldsymbol{X} \sim \mathrm{IG}\left(\frac{n+\nu}{2}, \frac{1}{2}\left(s_0^2 + s^2 + (\boldsymbol{\beta} - \hat{\boldsymbol{\beta}})^t(\boldsymbol{X}^t\boldsymbol{X})(\boldsymbol{\beta}^t - \hat{\boldsymbol{\beta}})\right)\right).$$