A) we have: $\cdot\ y_i \sim N(\theta, w^{-1})$

$\cdot\ \theta|w \sim N(\mu, (wk)^{-1})$

$\cdot\ w \sim Ga\left(\frac{d}{2}, \frac{\eta}{2}\right)$

$$p(\theta) = \int p(\theta|w)\, p(w)\, dw$$

$$= \int \frac{1}{\sqrt{2\pi}} \sqrt{wk} \cdot \exp\left\{-\frac{wk}{2}(\theta-\mu)^2\right\} \cdot \frac{\left(\frac{\eta}{2}\right)^{\frac{d}{2}}}{\Gamma\left(\frac{d}{2}\right)} \cdot w^{\frac{d}{2}-1} \cdot \exp\left\{-\frac{w\eta}{2}\right\} dw$$

$$= \frac{1}{\sqrt{2\pi}} \frac{\left(\frac{\eta}{2}\right)^{\frac{d}{2}}}{\Gamma\left(\frac{d}{2}\right)} \cdot \underbrace{\int w^{\frac{d}{2}+\frac{1}{2}-1} \cdot \exp\left\{-\frac{w}{2}\left(\eta + k(\theta-\mu)^2\right)\right\} dw}_{\text{kernel of } Ga\left(\frac{d}{2}+\frac{1}{2},\ \frac{\eta+k(\theta-\mu)^2}{2}\right)}$$

$$= \frac{1}{\sqrt{2\pi}} \frac{\left(\frac{\eta}{2}\right)^{\frac{d}{2}}}{\Gamma\left(\frac{d}{2}\right)} \cdot \Gamma\left(\frac{d+1}{2}\right) \cdot \left(\frac{2}{\eta + k(\theta-\mu)^2}\right)^{\frac{d+1}{2}}$$

$$\propto \left(1 + \left(\frac{k^{1/2}(\theta-\mu)}{\eta^{1/2}}\right)^2\right)^{-\frac{d+1}{2}} = \left(1 + \frac{\left(\frac{k^{1/2}(\theta-\mu)}{(\eta/d)^{1/2}}\right)^2}{d}\right)^{-\frac{d+1}{2}}$$

$$\boxed{\begin{array}{l} \Rightarrow \theta \sim \text{t-student} \\ \text{with: } \cdot\ d = \text{degrees of freedom} \\ \qquad\quad \cdot\ \mu = \text{centrality parameter} \\ \qquad\quad \cdot\ \left(\frac{\eta}{dk}\right)^{1/2} = \text{scale parameter} \end{array}}$$

B) Calculate the joint posterior density

$$p(\theta,w|y) \propto p(y)\cdot p(\theta|w)\cdot p(w)$$

$$\propto w^{\frac{n}{2}}\cdot \exp\left\{-\frac{w}{2}\boxed{\sum^{n}(y-\mu)^2}\right\}\cdot w^{\frac{d+1}{2}-1}\cdot \exp\left\{-w\frac{k(\theta-\mu)^2}{2}\right\}\cdot \exp\left\{-w\frac{\eta}{2}\right\}$$

$$\sum^{n}(y_i-\bar{y}-\bar{y}-\mu)^2$$

$$=\sum^{n}(y_i-\bar{y})^2+2\sum(y_i-\bar{y})(\bar{y}-\mu)+n(\bar{y}-\mu)^2$$

$$=\sum^{n}(y_i-\bar{y})^2+2(\bar{y}-\mu)\underbrace{\sum(y_i-\bar{y})}_{0}+n(\bar{y}-\mu)^2$$

$$=\boxed{\sum^{n}(y_i-\bar{y})^2+n(\bar{y}-\mu)^2}$$

So that we have:

$$= w^{\frac{n+d+1}{2}-1} \cdot \exp\left\{-\frac{w}{2}\left(n(\mu-\bar{y})^2 + \underbrace{\boxed{\sum (y_i-\bar{y})^2}}_{S_y}\right)\right\}$$

$$\times \exp\left\{-\frac{w}{2}\left(k(\theta-\mu)^2+\eta\right)\right\}$$

$$= w^{\frac{n+d+1}{2}-1}\cdot\exp\left\{-w\left(\frac{S_y}{2}+\frac{\eta}{2}+\frac{n(\mu-\bar{y})^2+k(\theta-\mu)^2}{2}\right)\right\}$$

$$n(\mu-\bar{y})^2+k(\theta-\mu)^2 = n(\bar{y}^2-2\bar{y}\theta+\theta^2)+k(\theta^2-2\theta\mu+\mu^2)$$

$$= n\bar{y}^2 - 2n\bar{y}\theta + n\theta^2 + k\theta^2 - 2\theta k\mu + \mu^2 k$$

$$= n\bar{y}^2 + \theta^2(n+k) - 2\theta(\bar{y}n + kn) + \mu^2 k$$

$$= n\bar{y}^2 + (n+k)\left(\theta^2 - 2\theta\left(\frac{\bar{y}n + k\mu}{n+k}\right)\right) + \mu^2 k$$

completing the square:

$$= (n+k)\left(\theta-\left(\frac{\bar{y}n + kn}{n+k}\right)\right)^2 - (n+k)\left(\frac{\bar{y}n + k\mu}{n+k}\right)^2 + n\bar{y}^2 + \mu^2 k$$

$$= (n+k)\left(\theta-\left(\frac{\bar{y}n+kn}{n+k}\right)\right)^2 - \frac{\bar{y}^2 n^2 + k^2\mu^2 + 2\bar{y}nk\mu + (n\bar{y}^2+\mu^2 k)(n+k)}{n+k}$$

$$= (n+k)\left(\theta - \left(\frac{\bar{y}n + kn}{n+k}\right)\right)^2 + \frac{\left(-\bar{y}^2n^2 - k^2\mu^2 - 2\mu k n\bar{y} + n^2\bar{y}^2 + \mu^2 kn + kn\bar{y}^2 + k^2\mu^2\right)}{n+k}$$

$$= (n+k)\left(\theta - \frac{\bar{y}n + kn}{n+k}\right)^2 + \frac{nk\left(\bar{y}^2 + \mu^2 - 2\mu\bar{y}\right)}{n+k}$$

$$= (n+k)\left(\theta - \left(\frac{\bar{y}n + kn}{n+k}\right)\right)^2 + \frac{nk(\bar{y}-\mu)^2}{n+k}$$

So that going back to :

$$p(\theta, w \mid y) \propto w^{\frac{d+n+1}{2} - 1}$$
$$\times \exp\left\{ -w\left(\frac{S_y}{2} + \frac{\eta}{2} + \frac{(n+k)}{2}\left(\theta - \left(\frac{\bar{y}n + kn}{n+k}\right)\right)^2 + \frac{nk(\bar{y}-\mu)^2}{2(n+k)} \right) \right\}$$

which can be written as:

$$p(\theta, w \mid y) \propto w^{\frac{d^*+1}{2} - 1} \cdot \exp\left\{ -w \frac{k^*(\theta - \mu^*)^2}{2} \right\} \cdot \exp\left\{ -w \frac{\eta^*}{2} \right\}$$

which is again a normal-gamma with parameters:

- $d^* = n + d$
- $k^* = n + k$
- $\mu^* = \dfrac{k\mu + n\bar{y}}{k+n}$
- $\eta^* = \eta + S_y + \dfrac{n \cdot k(\bar{y}-\mu)^2}{k+n}$

c) Derive the full conditional posterior $p(\theta|w,y)$

We can start from the joint posterior that we just derived and dropping all the elements that do not depend on w, we will easily get:

$$p(\theta|w,y) \propto \exp\left\{-w \cdot \frac{t^*(\theta-\mu^*)^2}{2}\right\}$$

which is the kernel of a normal, so that:

$$\boxed{\theta|w,y \sim \mathcal{N}\left(\mu^*, (w \cdot t^*)^{-1}\right)}$$

with μ^* and t^* as defined above.

b) Derive the marginal posterior $p(w|y)$:

$$\int w^{\frac{d^*+1}{2}-1} \cdot \exp\left\{-w \frac{t^*(\theta-\mu^*)^2}{2}\right\} \cdot \exp\left\{-w \frac{\eta^*}{2}\right\} d\theta$$

$$= w^{\frac{d^*+1}{2}-1} \cdot \exp\left\{-w \frac{\eta^*}{2}\right\} \cdot \underbrace{\int \exp\left\{-\frac{wt^*}{2}(\theta-\mu^*)^2\right\} d\theta}_{\text{kernel of } N(\mu^*,\ (wt^*)^{-1})}$$

$$\propto w^{\frac{d^*+1}{2}-1} \cdot \exp\left\{-w \frac{\eta^*}{2}\right\} \cdot \frac{1}{\sqrt{wt^*}}$$

$$= w^{\frac{d^*+\cancel{1}-\cancel{1}}{2}-1} \cdot \exp\left\{-w \frac{\eta^*}{2}\right\}$$

$$\Rightarrow \boxed{w \sim \text{Gamma}\left(\frac{d^*}{2}, \frac{\eta^*}{2}\right)}$$

E) Derive the marginal posterior $p(\theta|y)$

We start by recalling the previous result:

$p(\theta, w) =$ Normal-Gamma with d, k, μ, η $\Rightarrow$ $p(\theta) =$ t-student with $d, \mu, \left(\frac{\eta}{kd}\right)^{1/2}$ as parameters

$\Downarrow$

Similarly, we can say, using the result in B):

$p(\theta, w|y) \sim$ Normal-Gamma $\Rightarrow$ with parameters:

d^*, k^*, μ^*, η^* as defined above

$p(\theta|y) \sim$ t-student with:

- $d^* =$ degrees of freedom
- $\mu^* =$ centrality parameter
- $\left(\frac{\eta^*}{k^* d^*}\right)^{1/2} =$ scale parameter

F) Both priors will undefined if their hyperparameters are $\emptyset$.

($\Rightarrow$ They will not integrate to 1)

G) The parameters of the POSTERIORS, thanks to the contribution given by the likelihood, will still be such that the posteriors are defined.

H) The Bayesian credible interval for θ takes the form

$$\theta \in m \pm t^* \cdot s$$

Using the results we derived above, we have:

- $s = \sqrt{\dfrac{n^*}{d^* t^*}} \rightarrow \sqrt{\dfrac{S_y}{n^2}} = \sqrt{\dfrac{S_y}{n}}$

- $m = \mu^* \rightarrow \bar{y}$

so that the interval becomes

$$m \pm t^* \cdot s \rightarrow \bar{y} \pm t^* \frac{\sqrt{S_y}}{n}$$

which coincide with the confidence intervals for θ in a frequentist setting (up to a small discrepancy in the degrees of freedom)

THE CONJUGATE GAUSSIAN LINEAR MODEL

- $y \mid \beta, \sigma^2 \sim N(X\beta, (w\Lambda)^{-1})$ now not independent anymore, has to be seen as a vector

- $\beta \mid w \sim N(m, (wK)^{-1})$ same, not independent anymore (also not a unique parameter anymore)

- $w \sim Ga\left(\frac{d}{2}, \frac{\eta}{2}\right)$

A) We have:

$$p(\beta \mid w, y) \propto \exp\left\{-\frac{w}{2}\left[(y-X\beta)'\Lambda(y-X\beta)+(\beta-m)'K(\beta-m)\right]\right\}$$

$$= y'\Lambda y - \beta'X'\Lambda y + \beta'X'\Lambda X\beta - y'\Lambda X\beta + \beta'K\beta - m'K\beta - \beta'Km + m'Km$$

$$\propto -2\beta'X'\Lambda y + \beta'X'\Lambda X\beta + \beta'K\beta - 2\beta'Km$$

$$= -2\beta'(X'\Lambda y + Km) + \beta'(X'\Lambda X + K)\beta$$

completing the square:

$$= (\beta-m^*)'K^*(\beta-m^*) - m^{*\prime}K^*m$$

$$\propto (\beta-m^*)'K^*(\beta-m^*)$$

with: $m^* = (X'\Lambda X + K)^{-1}(X'\Lambda y + Km)$

$K^* = (X'\Lambda X + K)$

So that:

$$p(\beta|\omega,y) \propto \exp\left\{-\frac{\omega}{2}(\beta-m^*)'K^*(\beta-m^*)\right\}$$

which is the kernel of a Normal:

$$\Rightarrow \beta|\omega,y \sim N(m^*,(\omega K^*)^{-1})$$

B) $$p(\omega|y) \propto \int_{\beta} p(\omega,\beta|y)\, d\beta$$

$$\propto \omega^{(n+p+d)/2-1} \exp\left\{-\frac{\omega}{2}(y'\Lambda y + m'Km + \eta)\right\}$$

$$\propto \omega^{(n+p+d)/2-1} \cdot \exp\left\{-\frac{\omega}{2}(y'\Lambda y + m'Km + \eta)\right)$$

$$\cdot \exp\left\{\frac{\omega}{2}(y'\Lambda X + m'K)'(X'\Lambda X + K)^{-1}(y\Lambda X + m'K)\right]\right\}$$

$$\Rightarrow p(\omega|y) \sim Ga\left(\frac{n+d}{2}, \eta^*\right) \text{ as above}$$

c) $p(\beta|y)$

$$p(\beta|y) \propto \int p(\omega, \beta|y)\, d\omega$$

$$\propto \int p(\beta|\omega, y) \cdot p(\omega|y)\, d\omega$$

$$\propto \int \omega^{(n+p+d)/2-1} \cdot \exp\left\{-\frac{1}{2}(y-X\beta)'\omega\Lambda(y-X\beta)\right\}$$

$$\cdot \underbrace{\exp\left\{-\frac{1}{2}(\beta-m)'\omega k(\beta-m)\right\}}_{\text{kernel of gamma}} \cdot \exp\left\{-\frac{\omega\eta}{2}\right\} d\omega$$

$\Rightarrow$ ... using results we derived at the beginning of this batch of exercise:

$\beta|y \sim$ t-distribution with

- $\nu^* = n+d$ degree of freedom

$\lambda^* = X'\Lambda X + k$

- $\mu^* = (\lambda^*)^{-1}(X'\Lambda y + k'm)$ location
- $\Sigma^* = \frac{\nu^*}{\eta^*}\Lambda^*$ covariance

D) for Gibbs sampler implementation and results see code & notebook

HEAVY - TAILED ERROR MODEL

The new model is:

- $y \mid \beta, w, \Lambda \sim N(X\beta, (w\Lambda)^{-1})$
- $\Lambda = \operatorname{diag}(\lambda_1 \dots \lambda_n)$

  $\lambda_i \sim \text{Gamma}\left(\frac{h}{2}, \frac{h}{2}\right)$

- $\beta \mid w \sim N(m, (wK)^{-1})$

- $w \sim \text{Gamma}\left(\frac{d}{2}, \frac{\eta}{2}\right)$

A) What does this model imply for $p(y_i \mid \beta, w)$?

we have:

$$p(y_i \mid \beta, w) = \int_{\Lambda} p(y_i \mid \lambda_i, \beta, w) \cdot p(\lambda_i) \, d\lambda_i$$

$$\propto \int_{\Lambda} \sqrt{\lambda_i}\, \exp\left\{ -\frac{w\lambda_i}{2} (y_i - x_i'\beta)^2 \right\} \cdot \lambda_i^{\frac{h}{2}-1} e^{-\frac{\lambda_i h}{2}} d\lambda_i$$

$$\propto \int_{\Lambda} \lambda_i^{\frac{h+1}{2}-1} \cdot \exp\left\{ -\lambda_i \left[\frac{w}{2} (y_i - x_i'\beta)^2 + \frac{h}{2} \right] \right\} d\lambda_i$$

$$\propto \left[\frac{w}{2} (y_i - x_i'\beta)^2 + \frac{h}{2} \right]^{-\frac{h+1}{2}}$$

$$\propto \left[\left(\frac{w}{2} (y_i - x_i'\beta) + \frac{h}{2} \right) \frac{2}{h} \right]^{-\frac{h+1}{2}} \left(\frac{2}{h} \right)^{h+1}$$

$$\propto \left[\frac{w}{h} (y_i - x_i'\beta)^2 + 1 \right]^{-\frac{h+1}{2}}$$

$\Rightarrow$ $y_i \mid \beta, w \sim$ t- distribution

with parameters:

- $\nu = h$
- $\mu = x_i'\beta$
- $\delta^2 = \frac{1}{w}$

c) ? $p(x_i \mid \beta, \omega)$

using the integrand expression we found at the previous point, we can say that:

$$y_i \mid \beta, \omega \sim Ga\left(\frac{h+1}{2}, \frac{\omega}{2}(y_i - x_i'\beta)^2 + \frac{h}{2}\right)$$

c) For implementation and results, see python code and notebooks.