Statistical Modeling SDS 383D: Excercise 3

Khai Nguyen - kbn397
April 24, 2022

A simple Gaussian location model :
(A) We have

$$p(\theta, \omega) \propto \omega^{(d+1)/2-1} \exp\left\{-\omega \cdot \frac{\kappa(\theta-\mu)^2}{2}\right\} \cdot \exp\left\{-\omega \cdot \frac{\eta}{2}\right\}$$

Hence

$$\begin{aligned} p(\theta) &\propto \int_{-\infty}^{\infty} \omega^{(d+1)/2-1} \exp\left\{-\omega \cdot \frac{\kappa(\theta-\mu)^2}{2}\right\} \cdot \exp\left\{-\omega \cdot \frac{\eta}{2}\right\} d\omega \\ &\propto \int_{-\infty}^{\infty} \omega^{(d+1)/2-1} \exp\left\{-\omega \cdot \frac{\kappa(\theta-\mu)^2+\eta}{2}\right\} d\omega \\ &\propto \left(\frac{\kappa(\theta-\mu)^2+\eta}{2}\right)^{-\frac{d+1}{2}} \\ &\propto \left(1+\frac{1}{d}\frac{(\theta-\mu)^2}{\eta/(d\kappa)}\right)^{-\frac{d+1}{2}} \end{aligned}$$

So $v = d$, $s = \sqrt{\frac{\eta}{d\kappa}}$, $m = \mu$.
(B) We have the joint posterior:

$$\begin{aligned} p(\theta, \omega \mid \mathbf{y}) &\propto p(\mathbf{y} \mid \theta, \omega) p(\theta, \omega) \\ &\propto \prod_{i=1}^{n} \sqrt{\omega} \exp\left\{-\frac{\omega}{2}(y_i-\theta)^2\right\} \omega^{\frac{d+1}{2}-1} \exp\left\{-\omega\left(\frac{\kappa(\theta-\mu)^2}{2}+\frac{\eta}{2}\right)\right\} \\ &\propto \omega^{\frac{d+n+1}{2}-1} \exp\left\{-\frac{\omega}{2}\left[\sum_{i=1}^{n}\left(y_i^2-2y_i\theta+\theta^2\right)+\kappa\left(\theta^2-2\theta\mu+\mu^2\right)+\eta\right]\right\} \\ &\propto \omega^{\frac{d+n+1}{2}-1} \exp\left\{-\frac{\omega}{2}\left[\sum_{i=1}^{n} y_i^2-2n\bar{y}\theta+n\theta^2+\kappa\theta^2-2\kappa\theta\mu+\kappa\mu^2+\eta\right]\right\} \\ &\propto \omega^{\frac{d+n+1}{2}-1} \exp\left\{-\frac{\omega}{2}\left[\sum_{i=1}^{n} y_i^2+(\kappa+n)\theta^2-2(n\bar{y}+\kappa\mu)\theta+\kappa\mu^2+\eta\right]\right\} \\ &\propto \omega^{\frac{d+n+1}{2}-1} \exp\left\{-\frac{\omega}{2}\left[(\kappa+n)\theta^2-2(n\bar{y}+\kappa\mu)\theta+\left(\kappa\mu^2+\eta+\sum_{i=1}^{n} y_i^2\right)\right]\right\} \\ &\propto \omega^{\frac{d+n+1}{2}-1} \exp\left\{-\omega\left[\frac{(\kappa+n)\left(\theta-\frac{n\bar{y}+\kappa\mu}{\kappa+n}\right)^2}{2}\right]\right\} \exp\left\{-\omega\left(\frac{\kappa\mu^2+\eta+\sum_{i=1}^{n} y_i^2-\frac{(n\bar{y}+\kappa\mu)^2}{\kappa+n}}{2}\right)\right\} \\ &\equiv \omega^{\frac{d^*+1}{2}-1} \exp\left\{-\omega\frac{\kappa^*(\theta-\mu^*)^2}{2}\right\} \exp\left\{-\omega\frac{\eta^*}{2}\right\}, \end{aligned}$$

where

$$
\begin{aligned}
d^* &= d + n \\
\kappa^* &= \kappa + n \\
\mu^* &= \frac{n\bar{y} + \kappa\mu}{\kappa + n} \\
\eta^* &= \eta + \kappa\mu^2 + \sum_{i=1}^{n} y_i^2 - \frac{(n\bar{y} + \kappa\mu)^2}{\kappa + n} \\
&= \eta + S_y + \frac{n\kappa(\bar{y} - \mu)^2}{n + \kappa}
\end{aligned}
$$

(C)

$$
\begin{aligned}
p(\theta \mid \mathbf{y}, \omega) &\propto \exp\left\{-\frac{\omega\kappa^*}{2}(\theta - \mu^*)^2\right\} \\
&\equiv \text{Normal}\left(\mu^*, (\omega\kappa^*)^{-1}\right)
\end{aligned}
$$

(D)

$$
\begin{aligned}
p(\omega \mid \mathbf{y}) &= \int p(\omega, \theta \mid \mathbf{y}) d\theta \\
&\propto w^{\frac{d^*+1}{2}-1} \exp\left\{-\omega\frac{\eta^*}{2}\right\} \underbrace{\int \exp\left\{-\frac{\omega\kappa^*}{2}(\theta - \mu^*)^2\right\} d\theta}_{\text{kernel of Normal } \left(\mu^*, (\omega\kappa^*)^{-1}\right)} \\
&\propto \omega^{\frac{d^*+1}{2}-1} \exp\left\{-\omega\frac{\eta^*}{2}\right\} \left(\sqrt{\omega\kappa^*}\right)^{-1} \\
&\propto \omega^{\frac{d^*}{2}-1} \exp\left\{-\omega\frac{\eta^*}{2}\right\}
\end{aligned}
$$

which is Gamma $\left(\frac{d^*}{2}, \frac{\eta^*}{2}\right)$

(E)

$$
p(\theta \mid \mathbf{y}) \propto \left(1 + \frac{1}{v} \cdot \frac{(x - m)^2}{s^2}\right)^{\frac{v+1}{2}},
$$

where $v = d^*, m = \mu^*,$ and $s^2 = \frac{\eta^*}{d^*\kappa^*}$.

(F) FALSE; both priors will be undefined if their hyperparameters are 0

(G) TRUE: In this case, $d^*, \kappa^* \to n, \mu^* \to \bar{y}$, and $\eta^* \to S_y$.Therefore, the parameters in both marginal posteriors are defined and the probability distributions are valid.

(H) TRUE:

$$\begin{aligned}
s &= \sqrt{\frac{\eta^*}{d^*\kappa^*}} \\
&\to \sqrt{\frac{S_y}{n^2}} \\
&= \frac{\sqrt{S_y}}{n} \\
m &= \mu^* \\
&\to \bar{y}
\end{aligned}$$

So $m \pm t^\star \cdot s \to \bar{y} \pm t^* \frac{\sqrt{S_y}}{n}$.

The conjugate Gaussian linear model :
We first consider the joint distribution:

$$\begin{aligned}
p(\beta, \omega, \boldsymbol{y}) &= p(\omega)p(\beta|\omega)p(\boldsymbol{y}|\beta, \omega) \\
&\propto \omega^{\frac{d}{2}-1} \exp\left(-\frac{\eta}{2}\omega\right) \frac{1}{\sqrt{|(\omega K)^{-1}|}} \exp\left(-\frac{1}{2}(\beta - m)^\top \omega K(\beta - m)\right) \\
&\quad \frac{1}{\sqrt{|(\omega\Lambda)^{-1}|}} \exp\left(-\frac{1}{2}(\boldsymbol{y} - X\beta)^\top \omega\Lambda(\boldsymbol{y} - X\beta)\right)
\end{aligned}$$

(A) We have

$$\begin{aligned}
p(\beta|\boldsymbol{y}, \omega) &\propto \exp\left(-\frac{1}{2}(\beta - m)^\top \omega K(\beta - m)\right) \exp\left(-\frac{1}{2}(\boldsymbol{y} - X\beta)^\top \omega\Lambda(\boldsymbol{y} - X\beta)\right) \\
&\propto \exp\left(-\frac{1}{2}\left(\beta^\top(\omega K + X^\top\omega\Lambda X)\beta - 2(m^\top\omega K + \boldsymbol{y}^\top\omega\Lambda X)\beta\right)\right)
\end{aligned}$$

Let $A = (\omega K + X^\top\omega\Lambda X)$ and $b^\top = (m^\top\omega K + \boldsymbol{y}^\top\omega\Lambda X)$ then

$$\begin{aligned}
p(\beta|\boldsymbol{y}, \omega) &= N(A^{-1}b, A^{-1}) \\
&= N((K + X^\top\Lambda X)^{-1}(K^\top m + X^\top\Lambda^\top\boldsymbol{y}), A^{-1})
\end{aligned}$$

(B) We also have

$$\begin{aligned}
p(\omega \mid \mathbf{y}) &\propto \int_\beta p(\omega, \beta \mid \mathbf{y}) d\beta \\
&\propto \omega^{(n+p+d)/2-1} \exp\left(-\frac{\omega}{2}\left[\mathbf{y}^T\Lambda\mathbf{y} + m^T K m + \eta\right]\right) \\
&\quad \int_\beta \exp\left\{-\frac{1}{2}\left[-2\left(\mathbf{y}^T\omega\Lambda X + m^T\omega K\right)\beta + \beta^T\left(X^T\omega\Lambda X + \omega K\right)\beta\right]\right\} d\beta \\
&\propto \omega^{(n+d)/2-1} \exp\left\{-\frac{\omega}{2}\left[\mathbf{y}^T\Lambda\mathbf{y} + m^T K m + \eta\right]\right\} \\
&\quad \cdot \exp\left\{\frac{\omega}{2}\left[\left(\mathbf{y}^T\Lambda X + m^T K\right)^T\left(X^T\Lambda X + K\right)^{-1}\left(\mathbf{y}^T\Lambda X + m^T K\right)\right]\right\}
\end{aligned}$$

Therefore,

$$p(\omega \mid \mathbf{y}) = \text{Gamma}\left(\frac{n+d}{2}, \frac{\eta^*}{2}\right)$$
$$\eta^* = \eta + \mathbf{y}^T \Lambda \mathbf{y} + m^T K m - \left(\mathbf{y}^T \Lambda X + m^T K\right)^T \left(X^T \Lambda X + K\right)^{-1} \left(\mathbf{y}^T \Lambda X + m^T K\right)$$

(C) Using the result from previous sections, we know that the marginal posterior

$$p(\beta|\boldsymbol{y}) \propto \left(\frac{1}{v^*}\left(\beta - \mu^*\right)^T \Sigma^* \left(\beta - \mu^*\right) + 1\right)^{-\frac{v^*+p}{2}}$$

where $v^* = n + d, \Lambda^* = X^T \Lambda X + K, \mu^* = (\Lambda^*)^{-1}\left(X^T \Lambda \mathbf{y} + K^T m\right)$ and $\Sigma^* = \frac{v^*}{\eta^*}\Lambda^*$. So, the posterior is a t-distribution.

(D) I set $d = 4$, $\eta = 4$. The 95% Bayesian credible intervals that I got from $p(\beta|\boldsymbol{y})$ is [0.94044737 1.63745181]. The conventional 95% confident interval is [0.405, 2.112]. The histogram of the residual is given in Figure 1. The histogram is heavy tail, so our model should not assume normality. Also,

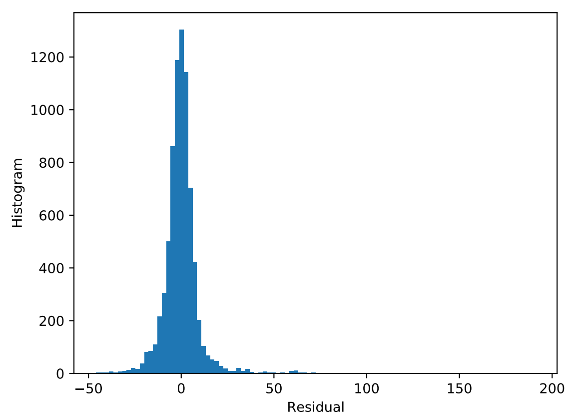

Figure 1: Histogram of the residual.

our assumptions on the common variance are not very true.

A heavy-tailed error model :

(A) We have

$$\begin{aligned}
p(y_i|X, \beta, \omega) &= \int p(y_i, \lambda_i|X, \beta, \omega) d\lambda_i \\
&= \int p(\lambda_i) p(y_i|X, \beta, \omega, \lambda_i) d\lambda_i \\
&= \int \frac{(h/2)^{h/2}}{\Gamma(h/2)} \lambda_i^{\frac{h}{2}-1} \exp\left(-\frac{h}{2}\lambda_i\right) \frac{(\omega\lambda_i)^{\frac{1}{2}}}{\sqrt{2\pi}} \exp\left(-\frac{\omega\lambda_i}{2}(y_i - x_i^\top \beta)^2\right) d\lambda_i \\
&= \omega^{\frac{1}{2}} \frac{(h/2)^{h/2}}{\Gamma(h/2)} \int \lambda_i^{\frac{h}{2}+\frac{1}{2}-1} \exp\left(-\lambda_i \frac{\omega(y_i - x_i^\top \beta)^2 + h}{2}\right) d\lambda_i
\end{aligned}$$

which is a t-distribution with $\nu = h, m = x_i^\top \beta, s^2 = 1/\omega$.
(B) We have

$$\begin{aligned}
p(\lambda_i|\boldsymbol{y},\beta,\omega) &\propto p(\lambda_i)p(y_i|X,\beta,\omega,\lambda_i) \\
&\propto \lambda_i^{\frac{h}{2}+\frac{1}{2}-1}\exp\left(-\lambda_i\frac{\omega(y_i-x_i^\top\beta)^2+h}{2}\right) \\
&= Gamma\left(\frac{h+1}{2},\frac{\omega(y_i-x_i^\top\beta)^2+h}{2}\right)
\end{aligned}$$

(C) We have

$$\begin{aligned}
p(\beta|\boldsymbol{y},\omega,\Lambda) &= p(\beta|\omega)p(\boldsymbol{y}|\beta,\omega,\Lambda) \\
&= N((K+X^\top\Lambda X)^{-1}(K^\top m + X^\top\Lambda^\top\boldsymbol{y}), A^{-1}) \\
p(\omega \mid \Lambda,\mathbf{y}) &= \text{Gamma}\left(\frac{n+d}{2},\frac{\eta^*}{2}\right) \\
\eta^* &= \eta + \mathbf{y}^T\Lambda\mathbf{y} + m^TKm - \left(\mathbf{y}^T\Lambda X + m^TK\right)^T\left(X^T\Lambda X + K\right)^{-1}\left(\mathbf{y}^T\Lambda X + m^TK\right)
\end{aligned}$$

I set $d = 4$, $\eta = 4$, and $h = 4$. 95% credible regions of the old model and the new model is given in Table 1. Based on Figure 2, we can see that for data points with smaller values, the relative variance could be significantly smaller than the variance for data points of higher values The comparison between previous model and the new model in terms of residual histogram is given in Figure 3.

β	Old	New
β_1	[-8.37695496, -7.28299482]	[-5.64635487, -4.25242018]
β_2	[0.89236421, 1.62210089]	[0.84528484, 1.85536977]
β_3	[1.01320204, 1.01369265]	[0.89993717, 0.93106667]
β_4	[-0.00159388, -0.00151613]	[-0.01721082, -0.00719019]
β_5	[6.86655076, 7.5869833]	[5.98062445, 6.98871499]
β_6	[3.00068415,3.55055834]	[2.76915875, 3.69882239]]

Table 1: 95% credible regions

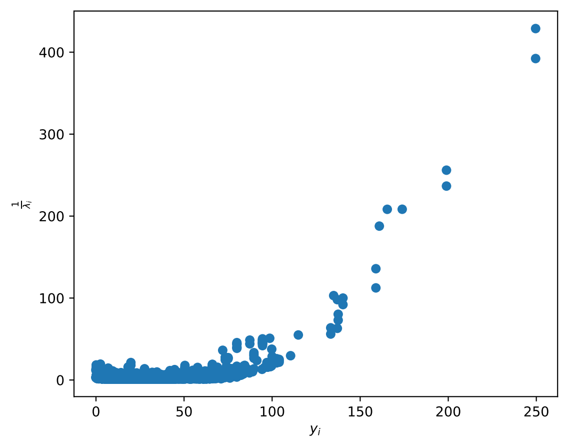

Figure 2: $\frac{1}{\lambda_i}$ and y_i.

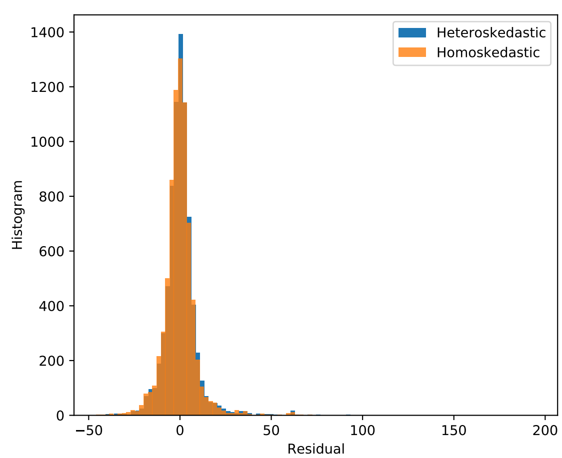

Figure 3: Histogram of the residual.