Statistical Modeling 2
Exercise 2

February 16, 2017

1 A simple Gaussian location model

A

The joint prior over the mean parameter θ and precision parameter ω is:

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

To get the marginal prior, we integrate out the parameter ω:

$$\begin{aligned}
p(\theta) &\propto \int_0^\infty \omega^{(d+1)/2-1} \exp\left\{-\omega \frac{\kappa(\theta-\mu)^2+\eta}{2}\right\} \\
&\propto \left(\frac{\kappa(\theta-\mu)^2+\eta}{2}\right)^{-(d+1)/2} \\
&= \left(\frac{\eta}{2} + \frac{\kappa(\theta-\mu)^2}{2}\right)^{-(d+1)/2} \\
&= \left(1 + \frac{\kappa(\theta-\mu)^2}{\eta}\right)^{-(d+1)/2} \left(\frac{\eta}{2}\right)^{-(d+1)/2} \\
&\propto \left(1 + \frac{\kappa(\theta-\mu)^2}{\eta}\right)^{-(d+1)/2} \\
&= \left(1 + \frac{1}{d}\frac{d\kappa(\theta-\mu)^2}{\eta}\right)^{-(d+1)/2}
\end{aligned}$$

Let $\nu = d, m = \mu$ and $s = \sqrt{\eta/(d\kappa)}$, we have a Student t distribution with ν degrees of freedom and scale s:

$$p(\theta) \propto \left(1 + \frac{1}{\nu}\frac{(\theta-m)^2}{s^2}\right)^{-(\nu+1)/2}$$

B

The sampling model is:

$$(y_i \mid \theta, \omega) \sim N(\theta, 1/\omega)$$

where $y_1, \ldots, y_n$ are the datapoints, θ is the mean and ω is the precision. We have that the likelihood for all the datapoints can be written as:

$$\begin{aligned}
p(\mathbf{y} \mid \theta, \omega) &\propto \prod_{i=1}^{n} \omega^{1/2} \exp\left\{-\frac{1}{2}\omega(y_i - \theta)^2\right\} \\
&= \omega^{n/2} \exp\left\{-\frac{1}{2}\omega \sum_{i=1}^{n}(y_i - \theta)^2\right\} \\
&= \omega^{n/2} \exp\left\{-\frac{1}{2}\omega \left(\sum_{i=1}^{n} y_i^2 + \sum_{i=1}^{n} \theta^2 - 2\sum_{i=1}^{n} y_i\theta\right)\right\} \\
&= \omega^{n/2} \exp\left\{-\frac{1}{2}\omega \left(\sum_{i=1}^{n} y_i^2 + n\theta^2 - 2n\overline{y}\theta + n\overline{y}^2 - n\overline{y}^2\right)\right\}
\end{aligned}$$

where $\overline{y} = \left(\sum_{i=1}^{n} y_i\right)/n$. Let $S_y = \sum_{i=1}^{n}(y_i - \overline{y})^2$, we have:

$$\begin{aligned}
S_y &= \sum_{i=1}^{n} y_i^2 + n\overline{y}^2 - 2\sum_{i=1}^{n} y_i\overline{y} \\
&= \sum_{i=1}^{n} y_i^2 - n\overline{y}^2
\end{aligned}$$

Therefore, the likelihood is:

$$\begin{aligned}
p(\mathbf{y} \mid \theta, \omega) &= \omega^{n/2} \exp\left\{-\frac{1}{2}\omega \left[S_y + n(\theta^2 - 2\overline{y}\theta + \overline{y}^2)\right]\right\} \\
&= \omega^{n/2} \exp\left\{-\frac{1}{2}\omega \left[S_y + n(\overline{y} - \theta)^2\right]\right\}
\end{aligned}$$

The posterior is proportional to the product of the likelihood and the prior:

$$\begin{aligned}
p(\theta, \omega \mid \mathbf{y}) \propto\ & \omega^{(d+1)/2-1} \exp\left\{-\omega \frac{\kappa(\theta - \mu)^2}{2}\right\} \exp\left\{-\omega\frac{\eta}{2}\right\} \\
& \omega^{n/2} \exp\left\{-\frac{1}{2}\omega \left[S_y + n(\overline{y} - \theta)^2\right]\right\} \\
&= \omega^{(d+n+1)/2-1} \exp\left\{-\omega \frac{\kappa(\theta - \mu)^2 + n(\theta - \overline{y})^2}{2}\right\} \exp\left\{-\omega\frac{\eta + S_y}{2}\right\}
\end{aligned}$$

We also have:

$$\begin{aligned}
\kappa(\theta-\mu)^2+n(\theta-\overline{y})^2 &= \kappa\theta^2+\kappa\mu^2-2\kappa\theta\mu+n\theta^2+n\overline{y}^2-2n\theta\overline{y} \\
&= (\kappa+n)\theta^2-2\theta(\kappa\mu+n\overline{y})+(\kappa\mu^2+n\overline{y}^2) \\
&= (\kappa+n)\left(\theta^2-2\theta\frac{\kappa\mu+n\overline{y}}{\kappa+n}+\frac{(\kappa\mu+n\overline{y})^2}{(\kappa+n)^2}\right)-\frac{(\kappa\mu+n\overline{y})^2}{\kappa+n}+(\kappa\mu^2+n\overline{y}^2) \\
&= (\kappa+n)\left(\theta^2-\frac{\kappa\mu+n\overline{y}}{\kappa+n}\right)^2-\frac{(\kappa\mu+n\overline{y})^2}{\kappa+n}+(\kappa\mu^2+n\overline{y}^2)
\end{aligned}$$

Therefore, the posterior is:

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

where:

$$\begin{aligned}
d^* &= d+n \\
\kappa^* &= \kappa+n \\
\mu^* &= \frac{\kappa\mu+n\overline{y}}{\kappa+n}
\end{aligned}$$

and

$$\begin{aligned}
\eta^* &= \eta+S_y-\frac{(\kappa\mu+n\overline{y})^2}{\kappa+n}+(\kappa\mu^2+n\overline{y}^2) \\
&= \eta+S_y+\frac{(\kappa+n)(\kappa\mu^2+n\overline{y}^2)-\kappa^2\mu^2-n^2\overline{y}^2+2\kappa n\mu\overline{y}}{\kappa+n} \\
&= \eta+S_y+\frac{\kappa n\mu^2+\kappa n\overline{y}^2+2\kappa n\mu\overline{y}}{\kappa+n} \\
&= \eta+S_y+\frac{\kappa n(\mu+\overline{y})^2}{\kappa+n}
\end{aligned}$$

C

The conditional distribution is:

$$p(\theta\mid\mathbf{y},\omega)\propto\exp\left\{-\omega\frac{\kappa^*(\theta-\mu^*)^2}{2}\right\}$$

We see that this is a Normal distribution with mean μ^* and variance $1/(\omega\kappa^*)$.

D

The marginal posterior of ω is:

$$
\begin{aligned}
p(\omega \mid \mathbf{y}) &= \int_{-\infty}^{\infty} p(\omega, \theta \mid \mathbf{y}) d\theta \\
&\propto \omega^{(d^*+1)/2-1} \exp\left\{-\omega \frac{\eta^*}{2}\right\} \int_{-\infty}^{\infty} \exp\left\{-\omega \frac{\kappa^*(\theta-\mu^*)^2}{2}\right\} d\theta \\
&\propto \omega^{(d^*+1)/2-1} \exp\left\{-\omega \frac{\eta^*}{2}\right\} \qquad \text{(Gaussian integral)}
\end{aligned}
$$

We see that this marginal is a Gamma distribution with parameter $(d^*/2, \eta^*/2)$.

E

The marginal posterior of θ is:

$$
\begin{aligned}
p(\theta \mid \mathbf{y}) &= \int_{0}^{\infty} p(\theta, \omega \mid \mathbf{y}) d\omega \\
&= \int_{0}^{\infty} \omega^{(d^*+1)/2-1} \exp\left\{-\omega \frac{\kappa^*(\theta-\mu^*)^2 + \eta^*}{2}\right\}
\end{aligned}
$$

This is the same integral in part A. By the results in A, we can see that this marginal is a Student t distribution with parameters $\nu = d^*, m = \mu^*$ and $s = \sqrt{\eta^*/(\kappa^* d^*)}$.

F

FALSE. As κ approaches 0, the Normal prior on θ approaches a point distribution but the density at that point is infinite. As d and η approach 0, the Gamma prior on ω also approach a point distribution with infinite density.

G

TRUE. By the results in D and E, we see that when the prior parameters approach 0, the posterior parameters are not 0 then $p(\theta \mid \mathbf{y})$ and $p(\omega \mid \mathbf{y})$ are valid distribution.

H

The classical frequentest confidence interval for θ is:

$$
\overline{y} \pm t^* \frac{\sqrt{S_y}}{\sqrt{n(n-1)}}
$$

As the prior parameters κ, d, η approach 0, we have the Bayesian credible interval for θ is:

$$m \pm t^* s$$

from the results in B and E, we have

$$m = \mu^* = \overline{y}$$

and

$$s = \sqrt{\eta^*/(\kappa d^*)} = \sqrt{\frac{S_y}{n^2}}$$

We see that this is different from the classical confidence interval.

2 The conjugate Gaussian linear model

A

$$\begin{aligned}
p(\beta, \mathbf{y} \mid \omega) &= p(\beta \mid \omega) p(\mathbf{y} \mid \beta, \omega) \\
&= N(\beta; m, (\omega K)^{-1}) N(\mathbf{y}; X\beta, (\omega\Lambda)^{-1}) \\
&\propto \omega^{p/2} K^{p/2} \exp\left\{\frac{1}{2}(\beta - m)^T \omega K (\beta - m)\right\} \\
&\omega^{n/2} \Lambda^{n/2} \exp\left\{\frac{1}{2}(\mathbf{y} - X\beta)^T \omega \Lambda (\mathbf{y} - X\beta)\right\} \\
&= \omega^{(n+p)/2} K^{p/2} \Lambda^{n/2} \exp\left\{\frac{1}{2}\omega E\right\}
\end{aligned}$$

where we have let E be the sum of two quadratic forms inside the exponent:

$$\begin{aligned}
E &= \beta^T K \beta - 2\beta^T K m + m^T K m + \mathbf{y}^T \Lambda \mathbf{y} - 2\beta^T X^T \Lambda \mathbf{y} + \beta^T X^T \Lambda X \beta \\
&= \begin{pmatrix} \beta \\ \mathbf{y} \end{pmatrix}^T \begin{pmatrix} K + X^T \Lambda X & -X^T \Lambda \\ -\Lambda^T X & \Lambda \end{pmatrix} \begin{pmatrix} \beta \\ \mathbf{y} \end{pmatrix} + \text{const}
\end{aligned}$$

where we have ignored linear terms in β and $\mathbf{y}$. We then have that (given ω) β and $\mathbf{y}$ are jointly Normal:

$$\begin{pmatrix} \beta \\ \mathbf{y} \end{pmatrix} \sim N\left(\begin{pmatrix} m \\ Xm \end{pmatrix}, \begin{pmatrix} \omega K + \omega X^T \Lambda X & -\omega X^T \Lambda \\ -\omega \Lambda^T X & \omega \Lambda \end{pmatrix}^{-1}\right)$$

Using the conditional Normal equations, we have that $p(\beta \mid \mathbf{y}, \omega)$ is Normal with mean:

$$\begin{aligned}
m^* &= m + (\omega K + \omega X^T \Lambda X)^{-1}(-\omega X^T \Lambda)(Xm - \mathbf{y}) \\
&= m + (K + X^T \Lambda X)^{-1}(-X^T \Lambda)(Xm - \mathbf{y})
\end{aligned}$$

and variance:

$$(\omega K^*)^{-1} = (\omega K + \omega X^T \Lambda X)^{-1}$$

B

We have that $p(\mathbf{y} \mid \omega)$ is Normal with mean Xm and variance:

$$\begin{aligned} \text{var}(X\beta + \epsilon) &= X\text{var}(\beta)X^T + (\omega\Lambda)^{-1} \\ &= X(\omega K)^{-1}X^T + (\omega\Lambda)^{-1} \end{aligned}$$

We then have:

$$\begin{aligned} p(\omega \mid \mathbf{y}) &\propto p(\omega)p(\mathbf{y} \mid \omega) \\ &\propto \omega^{d/2-1} \exp\left\{-\omega\frac{\eta}{2}\right\} \omega^{n/2} \exp\left\{-\omega\frac{1}{2}(\mathbf{y} - Xm)^T(XK^{-1}X^T + \Lambda^{-1})^{-1}(\mathbf{y} - Xm)\right\} \end{aligned}$$

We see that this distribution is Gamma with parameter $d^*/2$ and $\eta^*/2$ where:

$$\begin{aligned} d^* &= d + n \\ \eta^* &= \eta + (\mathbf{y} - Xm)^T(XK^{-1}X^T + \Lambda^{-1})^{-1}(\mathbf{y} - Xm) \end{aligned}$$

C

$$\begin{aligned} p(\beta \mid \mathbf{y}) &= \int_0^\infty p(\beta \mid \mathbf{y}, \omega)p(\omega \mid \mathbf{y})d\omega \\ &\int_0^\infty \omega^{(d+n+p)/2} \exp\left\{-\frac{\omega}{2}(\beta - m^*)^T K^*(\beta - m^*)\right\} \exp\left\{-\frac{\omega}{2}\eta^*\right\} \end{aligned}$$

This is an integral over a Normal gamma distribution. From previous results, we see that it is a Student t distribution:

$$p(\beta \mid \mathbf{y}) = \left(1 + \frac{1}{\nu}\frac{(\beta - \mu)^2}{s^2}\right)^{-(\nu+1)/2}$$

where

$$\begin{aligned} \nu &= d + n + p \\ \mu &= m^* \\ s &= \sqrt{\eta^*} \end{aligned}$$

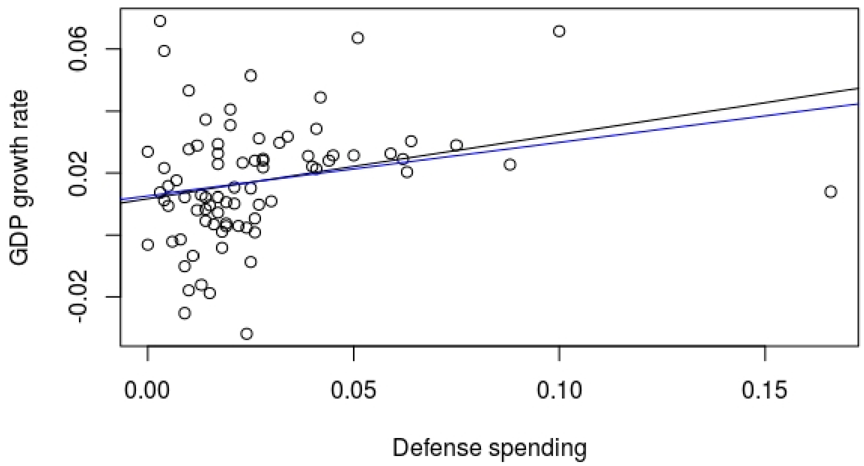

Figure 1: Frequentest regression line in black and Bayesian regression line in blue.

D

CODE: BLR.R

Figure 1: The two lines (Frequentest and Bayesian) are close on the left (Defense spending from 0.00 to 0.10), where most of the data points are. On the right (Defense spending from 0.10), there is a greater difference, likely due to the few points available.

A heavy-tailed error model

A

With the Gamma parameters λ integrated out, the marginal distribution over y_i is Student t. It has a wider tail and is more robust to outliers.

B

$$
\begin{aligned}
p(\lambda_i \mid \mathbf{y}, \beta, \omega) &\propto p(\lambda_i, y_i, \beta, \omega) \\
&= p(\omega)p(\lambda_i)p(\beta \mid \omega)p(y_i|\beta, \omega, \lambda_i) \\
&\propto p(\lambda_i)p(y_i|\beta, \omega, \lambda_i) \\
&\propto \lambda_i^{h/2-1} \exp\left(-\lambda_i h/2\right) (\omega\lambda_i)^{1/2} \exp\left[-\frac{1}{2}\omega\lambda_i(X_i^T\beta - y_i)^2\right] \\
&\propto \text{Gamma}\left(\frac{h+1}{2}, \frac{1}{2}[h + \omega(X_i^T\beta - y_i)^2]\right)
\end{aligned}
$$

C

CODE: HEAVY_TAIL.R

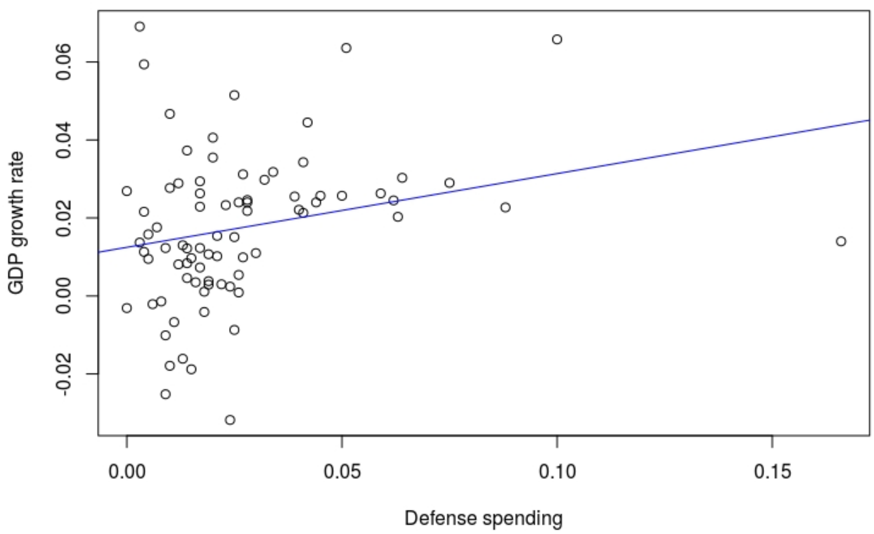

Figure 2: Regression line fitted by Gibbs sampling.

Figure 2: Compare to the lines in Figure 1, this line seems to be less affected by the outliers.