Travis Lilley
SDS 383D: Statistical Modeling II
Exercises 2

A simple Gaussian location model

(A)

To find the joint distribution of θ and σ^2, we do the following:

$$f(\theta, \omega) = f(\theta|\omega)f(\omega) = \left(\frac{\sqrt{\omega\kappa}}{\sqrt{2\pi}} \exp\left\{-\frac{\omega\kappa}{2}(\theta-\mu)^2\right\}\right)\left(\frac{\left(\frac{\eta}{2}\right)^{\frac{d}{2}}}{\Gamma\left(\frac{d}{2}\right)} \omega^{\frac{d}{2}-1} \exp\left\{-\frac{\eta}{2}\omega\right\}\right).$$

Then, to obtain the marginal (unconditional) distribution of θ, we have

$$f(\theta) = \int_\Omega f(\theta, \omega) d\omega$$

$$= \int_0^\infty \left[\frac{\sqrt{\omega\kappa}}{\sqrt{2\pi}} \exp\left\{-\frac{\omega\kappa}{2}(\theta-\mu)^2\right\} \cdot \frac{\left(\frac{\eta}{2}\right)^{\frac{d}{2}}}{\Gamma\left(\frac{d}{2}\right)} \omega^{\frac{d}{2}-1} \exp\left\{-\frac{\eta}{2}\omega\right\}\right] d\omega$$

$$= \frac{\left(\frac{\eta}{2}\right)^{\frac{d}{2}}\sqrt{\kappa}}{\Gamma\left(\frac{d}{2}\right)\sqrt{2\pi}} \int_0^\infty \left[\omega^{\frac{d}{2}-\frac{1}{2}} \exp\left\{-\omega\left[\frac{\kappa}{2}(\theta-\mu)^2 + \frac{\eta}{2}\right]\right\}\right] d\omega$$

$$= \frac{\left(\frac{\eta}{2}\right)^{\frac{d}{2}}\sqrt{\kappa}}{\Gamma\left(\frac{d}{2}\right)\sqrt{2\pi}} \cdot \frac{\Gamma\left(\frac{d}{2}+\frac{1}{2}\right)}{\left(\frac{\kappa}{2}(\theta-\mu)^2 + \frac{\eta}{2}\right)^{\frac{d}{2}+\frac{1}{2}}} \int_0^\infty \underbrace{\left[\frac{\left(\frac{\kappa}{2}(\theta-\mu)^2 + \frac{\eta}{2}\right)^{\frac{d}{2}+\frac{1}{2}}}{\Gamma\left(\frac{d}{2}+\frac{1}{2}\right)} \cdot \omega^{\frac{d}{2}-\frac{1}{2}} \exp\left\{-\omega\left[\frac{\kappa}{2}(\theta-\mu)^2 + \frac{\eta}{2}\right]\right\}\right]}_{\text{Gamma density}} d\omega$$

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

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

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

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

which we recognize as proportional to the "kernel" of a centered and scaled t distribution with parameters $m = \mu, s = \sqrt{\frac{\eta}{\kappa d}}$, and $v = d$.

(B)

The posterior is

$p(\theta, \omega | y_1, \dots, y_n) \propto f(y_1, \dots, y_n | \theta, \omega) p(\theta | \omega) p(\omega)$

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

$$\propto \omega^{\frac{n}{2}+\frac{d}{2}+\frac{1}{2}-1} \cdot \exp\left\{-\frac{\omega}{2}\left[S_y + n(\bar{y} - \theta)^2 + \kappa(\theta - \mu)^2\right]\right\} \exp\left\{-\omega \cdot \frac{\eta}{2}\right\}$$

$$= \omega^{\frac{n}{2}+\frac{d}{2}+\frac{1}{2}-1} \exp\left\{-\frac{\omega}{2}\left[n\bar{y}^2 - 2n\bar{y}\theta + n\theta^2 + \kappa\theta^2 - 2\kappa\mu\theta + \kappa\mu^2\right]\right\} \exp\left\{-\omega \cdot \frac{\eta}{2} - \omega \cdot \frac{S_y}{2}\right\}$$

$$= \omega^{\frac{n+d+1}{2}-1} \exp\left\{-\frac{\omega}{2}\left[n\theta^2 + \kappa\theta^2 - 2n\bar{y}\theta - 2\kappa\mu\theta\right]\right\} \exp\left\{-\omega \cdot \frac{\eta}{2} - \omega \cdot \frac{S_y}{2} - \frac{\omega}{2} n\bar{y}^2 - \frac{\omega}{2}\kappa\mu^2\right\}$$

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

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

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

$$= \omega^{\frac{n+d+1}{2}-1} \exp\left\{-\frac{\omega}{2}(n+\kappa)\left[\theta - \frac{n\bar{y} + \kappa\mu}{n+\kappa}\right]^2\right\} \exp\left\{-\frac{\omega}{2}\left(\eta + S_y + n\bar{y}^2 + \kappa\mu^2 - \frac{(n\bar{y} + \kappa\mu)^2}{n+\kappa}\right)\right\},$$

which we recognize as proportional to the kernel of a normal gamma distribution whose parameters are

$$\mu^* = \frac{n\bar{y} + \kappa\mu}{n+\kappa}$$

$$\kappa^* = n + \kappa$$

$$d^* = n + d$$

$$\eta^* = \eta + S_y + n\bar{y}^2 + \kappa\mu^2 - \frac{(n\bar{y} + \kappa\mu)^2}{n+\kappa} = \eta + S_y + \frac{n\kappa}{n+\kappa}(\mu - \bar{y})^2$$

(C)

We know from the prior that, conditional on ω, θ is normally distributed with mean μ and precision $\omega\kappa$. Thus, using the conjugacy of the prior, we know the posterior distribution of θ conditional on ω must also be normal with mean μ^* and precision $\omega\kappa^*$. That is,

$$\theta|\boldsymbol{y},\omega \sim N\left(\frac{n\bar{y}+\kappa\mu}{n+\kappa},\frac{1}{\omega(n+\kappa)}\right)$$

(D)

Actually, this one basically does come for free. We know that the prior marginal distribution of ω is gamma with parameters $\frac{d}{2}$ and $\frac{\eta}{2}$. Using the conjugacy of the model, the posterior marginal distribution of ω must also be gamma, but with parameters d^* and η^*. That is,

$$\omega|\boldsymbol{y} \sim \mathrm{Gamma}\left(\frac{n+d}{2},\frac{1}{2}\left[\eta+S_y+n\bar{y}^2+\kappa\mu^2-\frac{(n\bar{y}+\kappa\mu)^2}{n+\kappa}\right]\right)$$

(E)

Once more we appeal to conjugacy. We saw that the prior marginal distribution for θ was a centered, scaled t with parameters $m=\mu, s=\sqrt{\frac{\eta}{\kappa d}}$, and $v=d$. This means that the posterior marginal distribution for θ will also be a centered and scaled t, but now with parameters

$$m^*=\mu^*=\frac{n\bar{y}+\kappa\mu}{n+\kappa}$$

$$s^*=\sqrt{\frac{\eta^*}{\kappa^* d^*}}=\sqrt{\frac{1}{(n+\kappa)(n+d)}\left[\eta+S_y+n\bar{y}^2+\kappa\mu^2-\frac{(n\bar{y}+\kappa\mu)^2}{n+\kappa}\right]}$$

$$v^*=d^*=n+d.$$

(F)

Neither $p(\theta)$ nor $p(\omega)$ will be valid densities as κ, d, and η go to zero. If d goes to zero, then $p(\theta)$ becomes a t distribution with zero degrees of freedom, which is not a proper density. Moreover, if d goes to zero, $p(\omega)$ becomes a gamma whose first parameter is zero, which is also not a proper density.

(G)

Both $p(\theta|y)$ and $p(\omega|y)$ are valid densities as κ, d, and η go to zero. We see that $p(\theta|y)$ becomes a centered and scaled distribution with parameters $m^*=\bar{y}$, $s^*=\frac{1}{n}\sqrt{S_y}$, and $v^*=n$. Also, $p(\omega|y)$ approaches a gamma with parameters $\frac{n}{2}$ and $\frac{S_y}{2}$. Both of these are valid distributions.

(H)

Using our result from (E), we could form the 95% credible interval

$$\theta \in \frac{n\bar{y} + \kappa\mu}{n + \kappa} \pm t^* \cdot \sqrt{\frac{1}{(n+\kappa)(n+d)}\left[\eta + S_y + n\bar{y}^2 + \kappa\mu^2 - \frac{(n\bar{y} + \kappa\mu)^2}{n + \kappa}\right]}.$$

If n is sufficiently large, so that the t distribution is approximately normal, we can take t^* to be about 1.96.

If we let κ, d, and η approach zero, then the credible interval reduces to

$$\theta \in \bar{y} \pm t^* \cdot \sqrt{\frac{S_y}{n^2}} = \bar{y} + \frac{t^*}{\sqrt{n}}\sqrt{\frac{\sum(y_i - \bar{y})^2}{n}},$$

which is essentially identical to the usual frequentist confidence interval, except that it typically has the denominator of $n - 1$ under $\sum(y_i - \bar{y})^2$. For moderate sized n, however, the two types of intervals will essentially be the same.

The conjugate Gaussian linear model

(A)

$$p(\boldsymbol{\beta}, \omega|\boldsymbol{y}) \propto p(\boldsymbol{y}|\boldsymbol{\beta}, \omega)p(\boldsymbol{\beta}|\omega)p(\omega)$$

$$= \frac{1}{(2\pi)^{\frac{p}{2}}|(\omega\boldsymbol{\Lambda})^{-1}|^{\frac{1}{2}}}\exp\left\{-\frac{1}{2}(\boldsymbol{y}-\boldsymbol{X\beta})^T(\omega\boldsymbol{\Lambda})(\boldsymbol{y}-\boldsymbol{X\beta})\right\}\cdot\frac{1}{(2\pi)^{\frac{p}{2}}|(\omega\boldsymbol{K})^{-1}|^{\frac{1}{2}}}\exp\left\{-\frac{1}{2}(\boldsymbol{\beta}-\boldsymbol{m})^T(\omega\boldsymbol{K})(\boldsymbol{\beta}-\boldsymbol{m})\right\}$$

$$\cdot\frac{\left(\frac{\eta}{2}\right)^{\frac{d}{2}}}{\Gamma\left(\frac{d}{2}\right)}\omega^{\frac{d}{2}-1}\exp\left\{-\omega\cdot\frac{\eta}{2}\right\}$$

$$\propto \omega^{\frac{n+p+d}{2}-1}\exp\left\{-\frac{1}{2}(\boldsymbol{y}-\boldsymbol{X\beta})^T\omega\boldsymbol{\Lambda}(\boldsymbol{y}-\boldsymbol{X\beta})-\frac{1}{2}(\boldsymbol{\beta}-\boldsymbol{m})^T(\omega\boldsymbol{K})(\boldsymbol{\beta}-\boldsymbol{m})\right\}\exp\left\{-\omega\cdot\frac{\eta}{2}\right\}$$

$$= \omega^{\frac{n+p+d}{2}-1}\exp\left\{-\frac{\omega}{2}(\boldsymbol{y}^T-\boldsymbol{\beta}^T\boldsymbol{X}^T)\boldsymbol{\Lambda}(\boldsymbol{y}-\boldsymbol{X\beta})-\frac{\omega}{2}(\boldsymbol{\beta}^T-\boldsymbol{m}^T)\boldsymbol{K}(\boldsymbol{\beta}-\boldsymbol{m})\right\}\exp\left\{-\omega\cdot\frac{\eta}{2}\right\}$$

$$= \omega^{\frac{n+p+d}{2}-1}\exp\left\{-\frac{\omega}{2}[(\boldsymbol{y}^T-\boldsymbol{\beta}^T\boldsymbol{X}^T)\boldsymbol{\Lambda}(\boldsymbol{y}-\boldsymbol{X\beta})+(\boldsymbol{\beta}^T-\boldsymbol{m}^T)\boldsymbol{K}(\boldsymbol{\beta}-\boldsymbol{m})]\right\}\exp\left\{-\omega\cdot\frac{\eta}{2}\right\}$$

$$= \omega^{\frac{n+p+d}{2}-1}\exp\left\{-\frac{\omega}{2}[\boldsymbol{y}^T\boldsymbol{\Lambda}\boldsymbol{y}-2\boldsymbol{y}^T\boldsymbol{\Lambda}\boldsymbol{X\beta}+\boldsymbol{\beta}^T\boldsymbol{X}^T\boldsymbol{\Lambda}\boldsymbol{X\beta}+\boldsymbol{\beta}^T\boldsymbol{K\beta}-2\boldsymbol{m}^T\boldsymbol{K\beta}+\boldsymbol{m}^T\boldsymbol{K\beta m}]\right\}\exp\left\{-\omega\cdot\frac{\eta}{2}\right\}$$

$$= \omega^{\frac{n+p+d}{2}-1}\exp\left\{-\frac{\omega}{2}[\boldsymbol{\beta}^T\boldsymbol{X}^T\boldsymbol{\Lambda}\boldsymbol{X\beta}+\boldsymbol{\beta}^T\boldsymbol{K\beta}-2\boldsymbol{y}^T\boldsymbol{\Lambda}\boldsymbol{X\beta}-2\boldsymbol{m}^T\boldsymbol{K\beta}]\right\}\exp\left\{-\omega\cdot\frac{\eta}{2}-\frac{\omega}{2}\boldsymbol{y}^T\boldsymbol{\Lambda}\boldsymbol{y}-\frac{\omega}{2}\boldsymbol{m}^T\boldsymbol{K}\boldsymbol{m}\right\}$$

$$= \omega^{\frac{p}{2}}\exp\left\{-\frac{\omega}{2}[\boldsymbol{\beta}-(\boldsymbol{X}^T\boldsymbol{\Lambda}\boldsymbol{X}+\boldsymbol{K})^{-1}(\boldsymbol{X}^T\boldsymbol{\Lambda}\boldsymbol{y}+\boldsymbol{K}\boldsymbol{m})]^T(\boldsymbol{X}^T\boldsymbol{\Lambda}\boldsymbol{X}+\boldsymbol{K})[\boldsymbol{\beta}-(\boldsymbol{X}^T\boldsymbol{\Lambda}\boldsymbol{X}+\boldsymbol{K})^{-1}(\boldsymbol{X}^T\boldsymbol{\Lambda}\boldsymbol{y}+\boldsymbol{K}\boldsymbol{m})]\right\}$$

$$\times\exp\left\{-\frac{\omega}{2}[\eta+\boldsymbol{y}^T\boldsymbol{\Lambda}\boldsymbol{y}+\boldsymbol{m}^T\boldsymbol{K}\boldsymbol{m}-(\boldsymbol{X}^T\boldsymbol{\Lambda}\boldsymbol{y}+\boldsymbol{K}\boldsymbol{m})^T(\boldsymbol{X}^T\boldsymbol{\Lambda}\boldsymbol{X}+\boldsymbol{K})^{-1}(\boldsymbol{X}^T\boldsymbol{\Lambda}\boldsymbol{y}+\boldsymbol{K}\boldsymbol{m})]\right\}\cdot\omega^{\frac{n+d}{2}-1},$$

which we recognize as the kernel of a multivariate normal gamma distribution with parameters

$$\begin{aligned}\boldsymbol{m}^* &= (\boldsymbol{X}^T\boldsymbol{\Lambda}\boldsymbol{X}+\boldsymbol{K})^{-1}(\boldsymbol{X}^T\boldsymbol{\Lambda}\boldsymbol{y}+\boldsymbol{K}\boldsymbol{m})\\ \boldsymbol{K}^* &= \boldsymbol{X}^T\boldsymbol{\Lambda}\boldsymbol{X}+\boldsymbol{K}\\ d^* &= n+d\\ \eta^* &= \eta+\boldsymbol{y}^T\boldsymbol{\Lambda}\boldsymbol{y}+\boldsymbol{m}^T\boldsymbol{K}\boldsymbol{m}-(\boldsymbol{X}^T\boldsymbol{\Lambda}\boldsymbol{y}+\boldsymbol{K}\boldsymbol{m})^T(\boldsymbol{X}^T\boldsymbol{\Lambda}\boldsymbol{X}+\boldsymbol{K})^{-1}(\boldsymbol{X}^T\boldsymbol{\Lambda}\boldsymbol{y}+\boldsymbol{K}\boldsymbol{m})\end{aligned}$$

Therefore, we see that $p(\boldsymbol{\beta}|\boldsymbol{y},\omega) = \mathcal{N}_p((\boldsymbol{X}^T\boldsymbol{\Lambda}\boldsymbol{X}+\boldsymbol{K})^{-1}(\boldsymbol{X}^T\boldsymbol{\Lambda}\boldsymbol{y}+\boldsymbol{K}\boldsymbol{m}),(\omega\boldsymbol{X}^T\boldsymbol{\Lambda}\boldsymbol{X}+\omega\boldsymbol{K})^{-1})$.

(B)

From our results in (A), and the conjugacy of our model, we immediately deduce that

$$p(\omega|\boldsymbol{y}) = \text{Gamma}\left(\frac{n+d}{2},\frac{1}{2}[\eta+\boldsymbol{y}^T\boldsymbol{\Lambda}\boldsymbol{y}+\boldsymbol{m}^T\boldsymbol{K}\boldsymbol{m}-(\boldsymbol{X}^T\boldsymbol{\Lambda}\boldsymbol{y}+\boldsymbol{K}\boldsymbol{m})^T(\boldsymbol{X}^T\boldsymbol{\Lambda}\boldsymbol{X}+\boldsymbol{K})^{-1}(\boldsymbol{X}^T\boldsymbol{\Lambda}\boldsymbol{y}+\boldsymbol{K}\boldsymbol{m})]\right).$$

(C)

We will use the conjugacy of our model. First, we will find the marginal distribution $\boldsymbol{\beta}$ in the prior. Then, by appealing to conjugacy, we will immediately have the marginal posterior of $\boldsymbol{\beta}$.

$$p(\boldsymbol{\beta}, \omega) = p(\boldsymbol{\beta}|\omega)p(\omega)$$

$$= \frac{1}{(2\pi)^{\frac{p}{2}}|(\omega \boldsymbol{K})^{-1}|^{\frac{1}{2}}} \exp\left\{-\frac{1}{2}(\boldsymbol{\beta}-\boldsymbol{m})^T(\omega\boldsymbol{K})(\boldsymbol{\beta}-\boldsymbol{m})\right\} \cdot \frac{\left(\frac{\eta}{2}\right)^{\frac{d}{2}}}{\Gamma\left(\frac{d}{2}\right)} \omega^{\frac{d}{2}-1} \exp\left\{-\frac{\omega\eta}{2}\right\},$$

so that

$$p(\boldsymbol{\beta}) = \int p(\boldsymbol{\beta}|\omega)p(\omega)d\omega$$

$$\propto \int \underbrace{\omega^{\frac{p}{2}+\frac{d}{2}-1} \exp\left\{-\frac{\omega}{2}[(\boldsymbol{\beta}-\boldsymbol{m})^T\boldsymbol{K}(\boldsymbol{\beta}-\boldsymbol{m})+\eta]\right\}}_{\text{Gamma kernel}} d\omega$$

$$= \frac{\Gamma\left(\frac{p}{2}+\frac{d}{2}\right)}{\left(\frac{\eta}{2}+\frac{1}{2}(\boldsymbol{\beta}-\boldsymbol{m})^T\boldsymbol{K}(\boldsymbol{\beta}-\boldsymbol{m})\right)^{\frac{p}{2}+\frac{d}{2}}}$$

$$\propto \left[1+\frac{1}{\eta}(\boldsymbol{\beta}-\boldsymbol{m})^T\boldsymbol{K}(\boldsymbol{\beta}-\boldsymbol{m})\right]^{-\frac{p+d}{2}}$$

$$= \left[1+\frac{1}{d}(\boldsymbol{\beta}-\boldsymbol{m})^T\left(\frac{d}{\eta}\boldsymbol{K}\right)(\boldsymbol{\beta}-\boldsymbol{m})\right]^{-\frac{p+d}{2}},$$

which we recognize as a multivariate t distribution with parameters

$$\boldsymbol{\Sigma} = \frac{\eta}{d}\boldsymbol{K}^{-1}$$
$$\boldsymbol{\mu} = \boldsymbol{m}$$
$$\nu = d.$$

We can apply conjugacy to see that the posterior marginal of $p(\boldsymbol{\beta}|\boldsymbol{y})$ must also be multivariate t, but with parameters

$$\boldsymbol{\Sigma}^* = \frac{\eta^*}{d^*}(\boldsymbol{K}^*)^{-1}$$
$$\boldsymbol{\mu}^* = \boldsymbol{m}^*$$
$$\nu^* = d^*,$$

where $\eta^*, d^*, \boldsymbol{m}^*$, and $\boldsymbol{K}^*$ are defined as in part (A).

(D)

We use the following approach to model the data. From our earlier results, we expect that the distribution of $\boldsymbol{\beta}$, given the data, is multivariate t with mean parameter

$$\boldsymbol{m}^* = (\boldsymbol{X}^T\boldsymbol{\Lambda}\boldsymbol{X} + \boldsymbol{K})^{-1}(\boldsymbol{X}^T\boldsymbol{\Lambda}\boldsymbol{y} + \boldsymbol{K}\boldsymbol{m}).$$

Once we compute $\boldsymbol{m}^*$, we will have the mean values for β_0 and β_1, conditional on the data, which we call $\hat{\beta}_0$ and $\hat{\beta}_1$. We will then use the equation $\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_i$ as our "best-fit" line and then overlay this line on top of the actual data, where $\hat{y}_i$ represents our prediction. (Recall that our response variable is y_i, the growth rate, and our predictor is x_i, the country's defense spending.)

We need to specify the prior information. We will take $\boldsymbol{\Lambda} = \boldsymbol{I}$. For $\boldsymbol{K}$, we will use a vague prior by letting

$$\boldsymbol{K} = \begin{pmatrix} 0.001 & 0 \\ 0 & 0.001 \end{pmatrix},$$

so that our precision is low or "vague." For m, we will use the maximum likelihood estimates of β_0 and β_1, which, as we saw from the previous exercise, are given by $(\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{y}$. The MLE of $\boldsymbol{\beta}$ is then

$$\hat{\boldsymbol{\beta}}_{\text{MLE}} = \begin{pmatrix} 0.0118 \\ 0.2065 \end{pmatrix}.$$

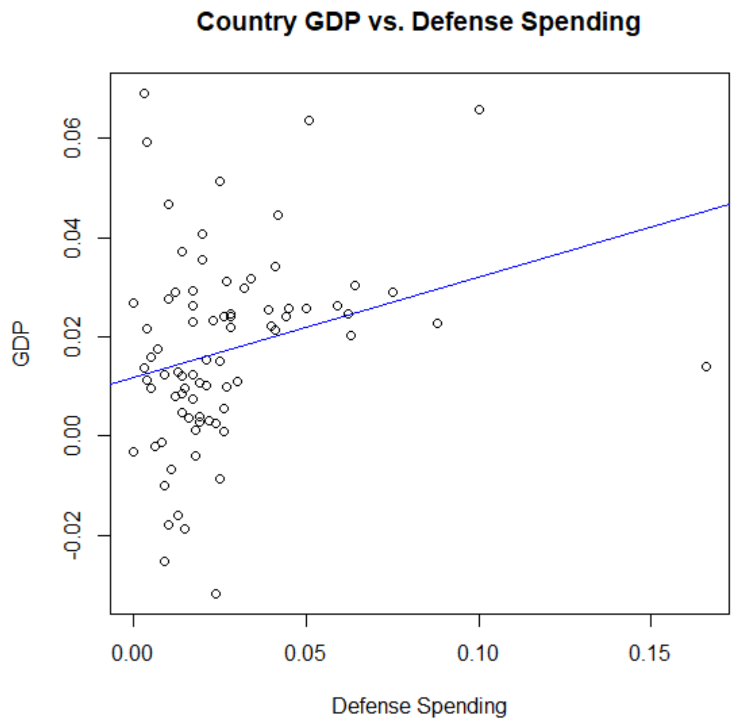
We then compute m^* using the formula above, and obtain the following posterior means for β:

$$\boldsymbol{\beta}_{\text{post}} = \begin{pmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \end{pmatrix} = \begin{pmatrix} 0.0118 \\ 0.2024 \end{pmatrix}.$$

We then examined what would happen if we had used a very noninformative prior for the prior mean of $\boldsymbol{\beta}$ by taking $\boldsymbol{m} = (0, 0)^T$. However, because the elements in K are so small, the prior mean has very little bearing on the posterior mean of $\boldsymbol{\beta}$, and our posterior means for $\boldsymbol{\beta}$ were essentially the same. This is somewhat reassuring, for it suggests that our prior choices for the parameters are not unduly influencing the model. We also note that our Bayesian analysis, even with vague priors, produces estimates very similar to the MLEs. It's always nice when frequentist and Bayesian analyses agree.

Below is the plot of the data with the line $\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_i$, overlaid, where $\hat{\beta}_0 = 0.0117$, and $\hat{\beta}_1 = 0.2106$.

There are quite a few concerns we should have with the plot below. It appears that we have some very influential points. In particular, the point located at about (0.15, 0.02) is strongly pulling the best fit line down and to the right. We should probability investigate this point, and see what happens when it is removed. There also appears to be some nonconstant variance as well, and perhaps even some curvature. Overall, it seems our model is misspecified. We probably need to go back to our original model and reevaluate.


Country GDP vs. Defense Spending
GDP
Defense Spending
0.06
0.04
0.02
0.00
-0.02
0.00
0.05
0.10
0.15

A heavy-tailed error model

(A)

First, we obtain the joint distribution of $p(y_i, \boldsymbol{\beta}, \omega, \lambda_i)$. We assume that $\boldsymbol{\beta}$ and ω are independent of the λ_i.

$$\begin{aligned}
p(y_i, \boldsymbol{\beta}, \omega, \lambda_i) &= p(y_i|\boldsymbol{\beta}, \omega, \lambda_i)p(\boldsymbol{\beta}, \omega, \lambda_i) \\
&= p(y_i|\boldsymbol{\beta}, \omega, \lambda_i)p(\boldsymbol{\beta}, \omega)p(\lambda_i) \\
&= p(y_i|\boldsymbol{\beta}, \omega, \lambda_i)p(\boldsymbol{\beta}|\omega)p(\omega)p(\lambda_i) \\
&\propto \sqrt{\omega\lambda_i}\exp\left\{-\frac{\omega\lambda_i}{2}\left(y_i - \boldsymbol{x}_i^T\boldsymbol{\beta}\right)^2\right\} \cdot \frac{1}{|(\omega\boldsymbol{K})^{-1}|^{\frac{1}{2}}}\exp\left\{-\frac{1}{2}(\boldsymbol{\beta}-\boldsymbol{m})^T(\omega\boldsymbol{K})(\boldsymbol{\beta}-\boldsymbol{m})\right\} \cdot \omega^{\frac{d}{2}-1}\exp\left\{-\frac{\omega\eta}{2}\right\} \\
&\quad \cdot \lambda_i^{\frac{h}{2}-1}\exp\left\{-\frac{\lambda_i h}{2}\right\} \\
&= \omega^{\frac{d}{2}+\frac{p}{2}+\frac{1}{2}-1}\lambda_i^{\frac{1}{2}+\frac{h}{2}-1}\exp\left\{-\lambda_i\left(\frac{\omega}{2}\left(y_i - \boldsymbol{x}_i^T\boldsymbol{\beta}\right)^2 + \frac{h}{2}\right)\right\}\exp\left\{-\frac{1}{2}(\boldsymbol{\beta}-\boldsymbol{m})^T(\omega\boldsymbol{K})(\boldsymbol{\beta}-\boldsymbol{m}) - \frac{\omega\eta}{2}\right\}
\end{aligned}$$

We integrate this last expression with respect to λ_i to obtain the joint density of $y_i, \boldsymbol{\beta}$, and ω.

$$\begin{aligned}
p(y_i, \boldsymbol{\beta}, \omega) &= \int\left[\omega^{\frac{d}{2}+\frac{p}{2}+\frac{1}{2}-1}\lambda_i^{\frac{1}{2}+\frac{h}{2}-1}\exp\left\{-\lambda_i\left(\frac{\omega}{2}\left(y_i - \boldsymbol{x}_i^T\boldsymbol{\beta}\right)^2 + \frac{h}{2}\right)\right\}\exp\left\{-\frac{1}{2}(\boldsymbol{\beta}-\boldsymbol{m})^T(\omega\boldsymbol{K})(\boldsymbol{\beta}-\boldsymbol{m}) - \frac{\omega\eta}{2}\right\}\right]d\lambda_i \\
&= \omega^{\frac{d}{2}+\frac{p}{2}+\frac{1}{2}-1}\exp\left\{-\frac{1}{2}(\boldsymbol{\beta}-\boldsymbol{m})^T(\omega\boldsymbol{K})(\boldsymbol{\beta}-\boldsymbol{m}) - \frac{\omega\eta}{2}\right\}\int\underbrace{\left[\lambda_i^{\frac{1}{2}+\frac{h}{2}-1}\exp\left\{-\lambda_i\left(\frac{\omega}{2}\left(y_i - \boldsymbol{x}_i^T\boldsymbol{\beta}\right)^2 + \frac{h}{2}\right)\right\}\right]}_{\text{Gamma kernel}}d\lambda_i \\
&= \omega^{\frac{d}{2}+\frac{p}{2}+\frac{1}{2}-1}\exp\left\{-\frac{1}{2}(\boldsymbol{\beta}-\boldsymbol{m})^T(\omega\boldsymbol{K})(\boldsymbol{\beta}-\boldsymbol{m}) - \frac{\omega\eta}{2}\right\} \cdot \frac{\Gamma\left(\frac{1}{2}+\frac{h}{2}\right)}{\left(\frac{\omega}{2}\left(y_i - \boldsymbol{x}_i^T\boldsymbol{\beta}\right)^2 + \frac{h}{2}\right)^{\frac{1}{2}+\frac{h}{2}}}.
\end{aligned}$$

To obtain the conditional distribution of $y_i|\boldsymbol{\beta}, \omega$, we regard the last expression as a function of y_i only, whereby $\boldsymbol{\beta}$ and ω are treated as constants

$$\begin{aligned}
p(y_i|\boldsymbol{\beta}, \omega) &\propto p(y_i, \boldsymbol{\beta}, \omega) \\
&= \omega^{\frac{d}{2}+\frac{p}{2}+\frac{1}{2}-1}\exp\left\{-\frac{1}{2}(\boldsymbol{\beta}-\boldsymbol{m})^T(\omega\boldsymbol{K})(\boldsymbol{\beta}-\boldsymbol{m}) - \frac{\omega\eta}{2}\right\} \cdot \frac{\Gamma\left(\frac{1}{2}+\frac{h}{2}\right)}{\left(\frac{\omega}{2}\left(y_i - \boldsymbol{x}_i^T\boldsymbol{\beta}\right)^2 + \frac{h}{2}\right)^{\frac{1}{2}+\frac{h}{2}}} \\
&\propto \left(\frac{\omega}{2}\left(y_i - \boldsymbol{x}_i^T\boldsymbol{\beta}\right)^2 + \frac{h}{2}\right)^{-\frac{h+1}{2}}
\end{aligned}$$

$$\propto \left(1 + \frac{1}{h}\frac{\left(y_i - \boldsymbol{x}_i^T\boldsymbol{\beta}\right)^2}{(1/\omega)}\right)^{-\frac{h+1}{2}},$$

and we recognize this last expression as the kernel of a location-scale t distribution with parameters

$$d = h$$
$$s = \omega^{-\frac{1}{2}}$$
$$m = \boldsymbol{x}_i^T\boldsymbol{\beta},$$

where $d, s,$ and m are the degrees of freedom, scale parameter, and location parameter, respectively.

(B)

In part (A) we derived $p(y_i, \boldsymbol{\beta}, \omega, \lambda_i)$. To obtain $p(\lambda_i|y_i, \boldsymbol{\beta}, \omega)$, we need only condition $p(y_i, \boldsymbol{\beta}, \omega, \lambda_i)$ on everything except for λ_i.

$$p(\lambda_i|y_i, \boldsymbol{\beta}, \omega) \propto p(y_i, \boldsymbol{\beta}, \omega, \lambda_i)$$

$$\propto \omega^{\frac{d}{2}+\frac{p}{2}+\frac{1}{2}-1}\lambda_i^{\frac{1}{2}+\frac{h}{2}-1} \exp\left\{-\lambda_i\left(\frac{\omega}{2}\left(y_i - \boldsymbol{x}_i^T\boldsymbol{\beta}\right)^2 + \frac{h}{2}\right)\right\} \exp\left\{-\frac{1}{2}(\boldsymbol{\beta} - \boldsymbol{m})^T(\omega\boldsymbol{K})(\boldsymbol{\beta} - \boldsymbol{m}) - \frac{\omega\eta}{2}\right\}$$

$$\propto \lambda^{\frac{h+1}{2}-1} \exp\left\{-\lambda_i\left[\frac{\omega}{2}\left(y_i - \boldsymbol{x}_i^T\boldsymbol{\beta}\right)^2 + \frac{h}{2}\right]\right\},$$

which we recognize as the kernel of a gamma distribution with shape and scale parameters

$$\alpha_0 = \frac{h+1}{2},$$
$$\alpha_1 = \frac{\omega}{2}\left(y_i - \boldsymbol{x}_i^T\boldsymbol{\beta}\right)^2 + \frac{h}{2}.$$

(C)

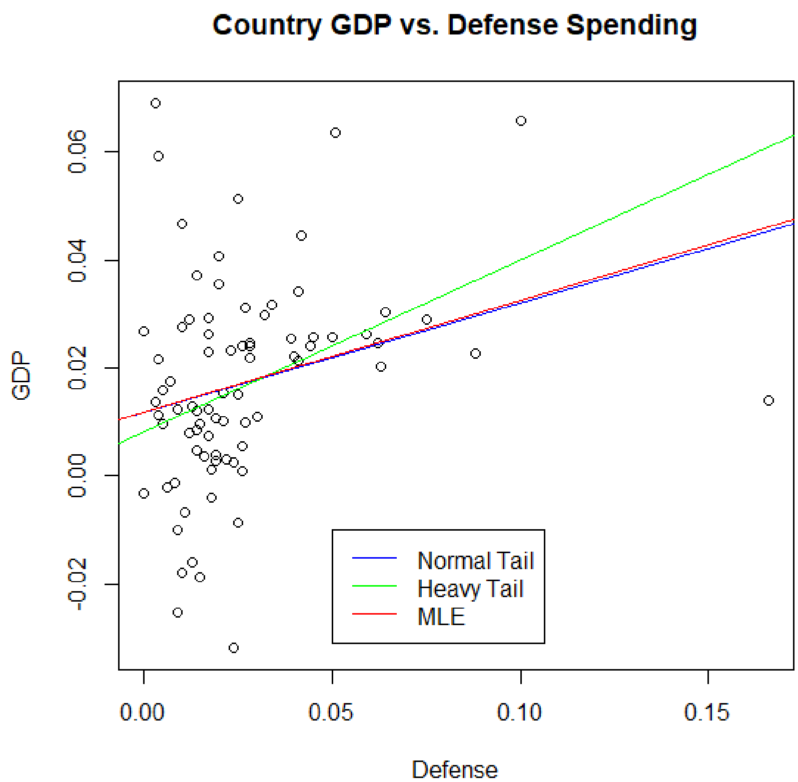
Combining our results from the last section with part (B) from this section, we have the following full conditionals:

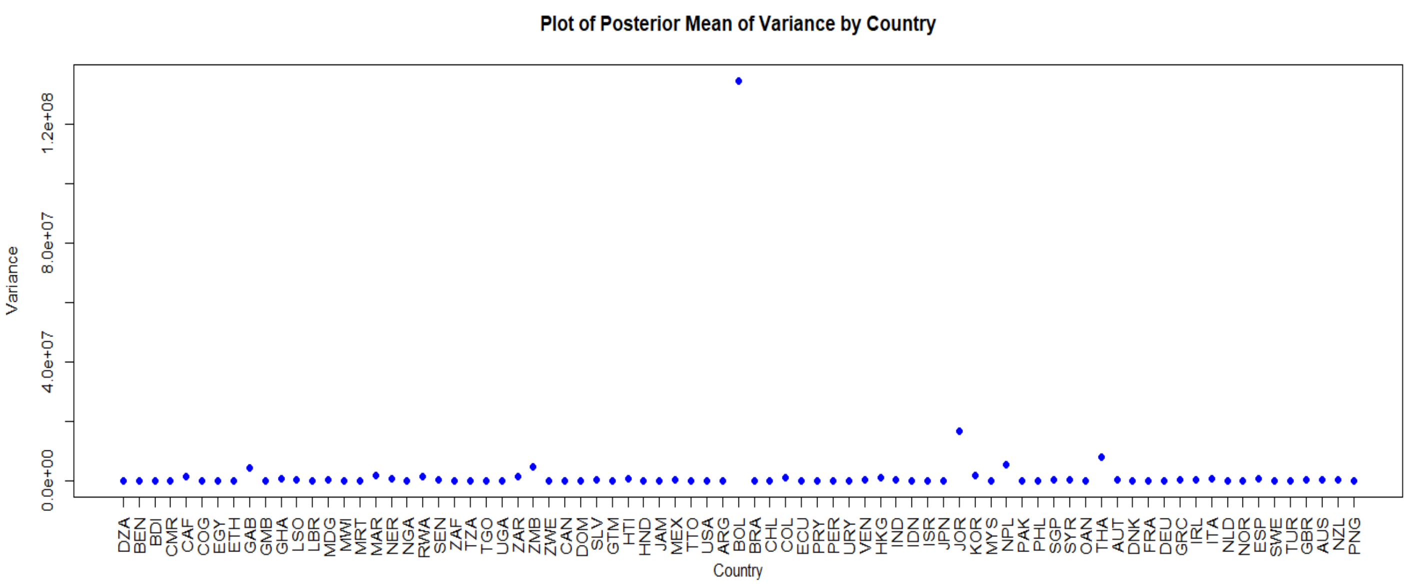
$$p(\boldsymbol{\beta}|\boldsymbol{y}, \omega, \boldsymbol{\Lambda}) = \mathcal{N}_p\left((\boldsymbol{X}^T\boldsymbol{\Lambda}\boldsymbol{X} + \boldsymbol{K})^{-1}(\boldsymbol{X}^T\boldsymbol{\Lambda}\boldsymbol{y} + \boldsymbol{K}\boldsymbol{m}), (\omega\boldsymbol{X}^T\boldsymbol{\Lambda}\boldsymbol{X} + \omega\boldsymbol{K})^{-1}\right)$$

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

$$p(\lambda_i|y, \boldsymbol{\beta}, \omega) = \text{Gamma}\left(\frac{h+1}{2}, \frac{\omega}{2}\left(y_i - \boldsymbol{x}_i^T\boldsymbol{\beta}\right)^2 + \frac{h}{2}\right)$$

The graph below shows the results of best fit lines that were made from the mean of 180000 Gibbs samples of $\boldsymbol{\beta}$ after burning and thinning. As we can see, the heavy-tailed model is less susceptible to the highly influential point on the far right. Including a component in the model that allowed us to estimate and update the precision of each observation allowed us to better model without succumbing to the large variance of a single point. However, it still seems that a simple line is not a good fit for this data. Perhaps a log-transformation on either the response or covariate (or both) would result in a graph that could be much better represented by a single line. The second graph is a plot of the posterior mean variance by country. It isn't hard to see that Bolivia has an incredibly high variance. Unsurprisingly, when we investigate, we see that Bolivia is the large leverage point that is throwing off the model. We should probably investigate why this point is so different from all the others.





R Code

```
# SDS 383D: Statistical Modeling II
# Exercise 2
# gdpgrowth.csv data

# Load data
data <- read.csv("C:\\Users\\trlil\\Google Drive\\UT Austin\\SDS 383D Statistical Modeling
II\\Homework\\Exercises 2\\gdpgrowth.csv", header = T)

# Response
y <- data$GR6096

# Design matrix
X <- cbind(1, data$DEF60)

# Number of obs
n <- length(growth)

# Likelihood precision
L <- diag(1, n)

# Prior precision of B|w
# Low precision means "vague"
K <- diag(c(0.001, 0.001))

# MLE
mle <- solve(t(X) %*% X) %*% t(X) %*% y

# Posterior mean with uninformative prior
m <- c(0, 0)
m_star <- solve(t(X) %*% L %*% X + K) %*%
           (t(X) %*% L %*% y + K %*% m)

# Posterior mean is almost the same,
# regardless of what our prior m is.
# This is because we have made K very
# uninformative. Also, the posterior
# estimates are similar to the MLEs.

# Plot data
plot(X[, 2], y,
     main = "Country GDP vs. Defense Spending",
     xlab = "Defense Spending",
     ylab = "GDP")

# Plot Bayesian regression line
abline(m_star, col = "blue")
```

```r
# SDS 383D: Statistical Modeling II
# Exercise 2
# GDP Growth Gibbs Sampler

# Add libraries
library(mvtnorm)

# Load data
data <- read.csv("C:\\Users\\trlil\\Google Drive\\UT Austin\\SDS 383D Statistical Modeling
II\\Homework\\Exercises 2\\gdpgrowth.csv",
                    header = T)

# Response
y <- data$GR6096

# Design matrix
X <- cbind(1, data$DEF60)

# Initialize prior parameters
d <- 1/100
eta <- 1/100
m <- c(0, 0)
K <- diag(c(0.001, 0.001))
h <- 1/100

# Number of Gibbs samples
G <- 10^6

# Number of obs
n <- length(y)

# Objects to store Gibbs samples

# Lambda
L <- matrix(0, nrow = G, n)

# Omega
w <- rep(0, G)

# Beta
B <- matrix(0, nrow = G, 2)

# Initialize Gibbs
L[1, ] <- rep(1, n)
w[1] <- 1
B[1, ] <- c(0, 0)

# Run Gibbs
for(i in 2:G){

	# Update location of sampler
	if(i %% (G/100) == 0){
		print(i)
	}

	# Update L
	for(j in 1:n){
		L[i, j] <- rgamma(1, h/2 + 1/2,
		                    w[i-1]/2*(y[j] - X[j, ] %*% B[i - 1, ])^2 + h/2)
	}

	# Diagonal matrix Lambda
	L2 <- diag(L[i, ])

	# Update w
	w[i] <- rgamma(1, n/2 + d/2,
	               (1/2)*(eta + t(y) %*% L2 %*% y + t(m) %*% K %*% m -
	               t(t(X) %*% L2 %*% y + K %*% m) %*%
	               solve(t(X) %*% L2 %*% X + K) %*%
	               (t(X) %*% L2 %*% y + K %*% m)))

	# Update B
	B[i, ] <- rmvnorm(1, solve(t(X) %*% L2 %*% X + K) %*%
	                     (t(X) %*% L2 %*% y + K %*% m),
```

```r
                                  solve(w[i] * t(X) %*% L2 %*% X + w[i] * K))
}

# Remove burn-in period
L <- L[(0.1*G):G, ]
w <- w[(0.1*G):G]
B <- B[(0.1*G):G, ]

# Thin
L <- L[seq(1, dim(L)[1], by = 5), ]
w <- w[seq(1, length(w), by = 5)]
B <- B[seq(1, dim(B)[1], by = 5), ]

# Posterior mean of B
colMeans(B)

# Plot data
plot(X[, 2], y,
     main = "Country GDP vs. Defense Spending",
     xlab = "Defense",
     ylab = "GDP")

# Best fit line part (1)
abline(c(0.0118, 0.2024), col = "blue")

# Best fit line part (2)
abline(colMeans(B), col = "green")

# MLE best fit line
abline(c(0.0118, 0.2065), col = "red")

# Add legend
legend(0.05, -0.01,
        c("Normal Tail", "Heavy Tail", "MLE"),
        col = c("blue", "green", "red"),
        lty = c(1, 1, 1))


# Plot of posterior means of 1/lambda
plot(colMeans(1/L),
     main = "Plot of Posterior Mean of Variance by Country",
     ylab = "Variance",
     xlab = "Country",
     pch = 20,
     cex = 1.5,
     col = "blue",
     xaxt = "n")

# Customize x-axis
axis(1, at = 1:79, labels = data$CODE[1:79], las = 2, ps = 5)
```