SDS 383D Ex 02:

Bayes and the Gaussian Linear Model

February 4, 2016

Jennifer Starling

A Simple Gaussian Location Model

Part A

The marginal prior $p(\theta)$ is a gamma mixture of normals. Show it takes the form of a centered, scaled t distribution: $p(\theta) \propto \left(1 + \frac{1}{\nu} \cdot \frac{(x-m)^2}{s^2}\right)^{-\frac{\nu+1}{2}}$.

$$(\theta|\omega) \sim N(\mu, \omega^{-1}k^{-1}) and \omega \sim gamma(\frac{d}{2}, \frac{\eta}{2})$$

Joint prior for $p(\theta, \omega)$ has form

$$p(\theta,\omega) = p(\theta|\omega)p(\omega) \propto \omega^{\left(\frac{(d+1)}{2}-1\right)} \exp[-\omega \cdot \frac{k(\theta-\mu)^2}{2}] \exp[-\omega\frac{\eta}{2}]$$

Marginal prior for θ is calculated as follows:

$$\begin{aligned}
p(\theta) &= \int_0^\infty p(\theta,\omega)\delta\omega \\
&= \int_0^\infty \omega^{\left(\frac{(d+1)}{2}-1\right)} \exp[-\omega \cdot \frac{k(\theta-\mu)^2}{2}] \exp[-\omega\frac{\eta}{2}]\delta\omega \\
&= \int_0^\infty \omega^{\left(\frac{(d+1)}{2}-1\right)} e^{-\omega\left(\frac{k(\theta-\mu)^2}{2}+\frac{\eta}{2}\right)}\delta\omega
\end{aligned}$$

This is the kernel of a *gamma* $\left(\frac{(d+1)}{2}, \frac{k(\theta-\mu)^2}{2}+\frac{\eta}{2}\right)$ distribution.
The integral will therefore equal $\frac{1}{c}$, where c is the constant of proportionality for this gamma density, $\frac{1}{\beta^\alpha/\Gamma(\alpha)} = \frac{\Gamma(\alpha)}{\beta^\alpha}$.

$$\begin{aligned}
&= \Gamma(\frac{d+1}{2})\left[\frac{k(\theta-\mu)^2}{2}+\frac{\eta}{2}\right]^{-\frac{d+1}{2}} \\
&\propto \left[\left(\frac{\eta}{2}\right)\left(1+\frac{k(\theta-\mu)^2}{\eta}\right)\right]^{-\frac{d+1}{2}} \\
&= \left(\frac{\eta}{2}\right)^{-\frac{d+1}{2}}\left[\left(1+\frac{k(\theta-\mu)^2}{\eta}\right)\right]^{-\frac{d+1}{2}}
\end{aligned}$$

The left term cancels, as it is constant wrt θ.

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

Multiply numerator and denominator by d.

$$\begin{aligned}
&= \left[\left(1+\frac{dk(\theta-\mu)^2}{d\eta}\right)\right]^{-\frac{d+1}{2}} \\
&= \left[\left(1+\left(\frac{1}{d}\right)\frac{(\theta-\mu)^2}{\frac{\eta}{dk}}\right)\right]^{-\frac{d+1}{2}}
\end{aligned}$$

This has the desired central-scaled t form, with mean $m = \mu$, df $\nu = d$, and scale $s^2 = \frac{\eta}{dk}$.

Part B

Assume the sampling model $(y_i|\theta,\sigma^2) \sim N(\theta,\sigma^2)$ for $i = 1...n$, so $y = (y_1, ..., y_n)^T$. Work with precision $\omega = 1/\sigma^2$. Calculate the joint posterior up to a constant of proportionality (factors not dependent on θ or ω.) Show the joint posterior is also normal-gamma.

The model is as follows.

$$\begin{aligned} y_i &\sim N(\theta,\sigma^2) \\ \theta &\sim N\left(\mu,(\omega k)^{-1}\right) \\ \omega &\sim Gamma\left(\frac{d}{2},\frac{\eta}{2}\right) \end{aligned}$$

The prior has (normal-gamma) form

$$\begin{aligned} p(\theta,\omega) &= p(\theta|\omega)\cdot p(\omega) \\ &\propto \omega^{\left(\frac{(d+1)}{2}-1\right)} \exp[-\omega\cdot\frac{k(\theta-\mu)^2}{2}] \exp[-\omega\frac{\eta}{2}] \end{aligned}$$

The sampling model has form

$$\begin{aligned} p(y|\theta,\omega) &\propto \exp\left[-\frac{\omega}{2}\sum_{i=1}^{n}(y_i-\theta)^2\right] \\ &= \exp\left[-\frac{\omega}{2}\sum_{i=1}^{n}(y_i-\bar{y}+\bar{y}\theta)^2\right] \text{, using trick of adding/subtracting } \bar{y} \\ &= \exp\left[-\frac{\omega}{2}\sum_{i=1}^{n}\left\{(y_i-\bar{y})^2+n(\bar{y}-\theta)^2+2(y-\theta)\sum_{i=1}^{n}(y_i-\bar{y})\right\}\right] \\ &= \exp\left[-\frac{\omega}{2}\left\{S_y+n(\bar{y}-\theta)^2\right\}\right] \text{, with } S_y=\sum_{i=1}^{n}(y_i-\bar{y})^2 \text{ (since the last term equals zero)} \end{aligned}$$

Sampling model (from the normal likelihood), using $S_y = \sum_{i=1}^{n}(y_i-\bar{y})^2$, is
The posterior $\propto$ sampling model * prior, so

$$p(\theta,\omega|y) \propto p(y|\theta,\omega)p(\theta,\omega)$$

Therefore, the joint posterior is

$$\begin{aligned} p(\theta,\omega|y) &\propto \omega^{\left(\frac{(d+1)}{2}-1\right)} \exp[-\omega\cdot\frac{k(\theta-\mu)^2}{2}] \exp[-\omega\frac{\eta}{2}]\cdot\omega^{\frac{n}{2}} \exp\left[-\frac{\omega}{2}\left(S_y+n(\bar{y}-\theta)^2\right)\right] \\ &= \omega^{\left(\frac{n+d+1}{2}-1\right)}\cdot\exp\left[-\frac{\omega}{2}\left(S_y+n(\bar{y}-\theta)^2+k(\theta-\mu)^2+\eta\right)\right] \end{aligned}$$

Simplify the exponential term multiplied by $-\frac{\omega}{2}$.

$$\begin{aligned} &\left(S_y+n(\bar{y}-\theta)^2+k(\theta-\mu)^2+\eta\right) \\ &= S_y+n(\bar{y}-\theta)(\bar{y}-\theta)+k(\theta-\mu)(\theta-\mu)+\eta \\ &= S_y+n\bar{y}^2-2n\bar{y}\theta+n\theta^2+k\theta^2-2k\theta\mu+k\mu^2+\eta \\ &= (n+k)\theta^2-2(n\bar{y}+k\mu)\theta+\left(S_y+n\bar{y}^2+k\mu^2+\eta\right) \end{aligned}$$

This has the form $ax^2 - 2bx + c$. We can complete the square as follows:

$$\begin{aligned}
&ax^2 - 2bx + c \\
&\quad = a\left[x^2 - 2\frac{b}{a}x + \frac{c}{a}\right] \\
&\quad = a\left[x^2 - 2\frac{b}{a}x + (\frac{b}{a})^2 - (\frac{b}{a})^2 + \frac{c}{a}\right] \\
&\quad = a\left[(x - \frac{b}{a})^2 - (\frac{b}{a})^2 + \frac{c}{a}\right] \\
&\quad = a(x - \frac{b}{a})^2 - \frac{b^2}{a} + c
\end{aligned}$$

Plugging in the appropriate a, b and c terms yields:

$$\begin{aligned}
&= (n+k)\left[\theta - \frac{n\bar{y} + k\mu}{n+k}\right]^2 - \frac{(n\bar{y} + k\mu)^2}{n+k} + S_y + n\bar{y}^2 + k\mu^2 + \eta \\
&= (n+k)\left[\theta - \frac{n\bar{y} + k\mu}{n+k}\right]^2 + \frac{nk(\bar{y} - \mu)^2}{(n+k)} + S_y + \eta
\end{aligned}$$

Plug this rearranged term back into the entire posterior:

$$\begin{aligned}
p(\theta, \omega|y) &\propto \omega^{\left(\frac{n+d+1}{2} - 1\right)} \cdot \exp\left[-\frac{\omega}{2}\left((n+k)\left[\theta - \frac{n\bar{y} + k\mu}{n+k}\right]^2 + \frac{nk(\bar{y} - \mu)^2}{(n+k)} + S_y + \eta\right)\right] \\
p(\theta, \omega|y) &\propto \omega^{\left(\frac{n+d+1}{2} - 1\right)} \cdot \exp\left[-\frac{\omega}{2}\left((n+k)\left[\theta - \frac{n\bar{y} + k\mu}{n+k}\right]^2\right)\right] \cdot \exp\left[-\frac{\omega}{2}\left(\frac{nk(\bar{y} - \mu)^2}{(n+k)} + S_y + \eta\right)\right]
\end{aligned}$$

Posterior $p(\theta, \omega|y)$ has form of the normal-gamma distribution, with posterior parameters as follows.

$$d^* = (n+d) \tag{1}$$

$$k^* = (n+k) \tag{2}$$

$$\mu^* = \frac{n\bar{y} + k\mu}{(n+k)} \tag{3}$$

$$\eta^* = \frac{nk(\bar{y} - \mu)^2}{(n+k)} + S_y + \eta \tag{4}$$

Part C

The conditional posterior of $p(\theta|y,\omega)$ is

$$p(\theta|y,\omega) \propto \exp\left[-\frac{\omega}{2}\cdot(n+k)\left(\theta - \frac{n\bar{y}+k\mu}{(n+k)}\right)^2\right]$$

This is the Normal distribution form, parameterized with precision.

$$\begin{aligned} p(\theta|y,\omega) &\sim N\left(\frac{n\bar{y}+k\mu}{(n+k)}, (-\omega(n+k))^{-1}\right) \\ &\sim N\left(\mu^*, (\omega k^*)^{-1}\right) \end{aligned}$$

Part D

From the joint posterior in Part A, what is the marginal posterior $p(\omega|y)$?

Note: Will indicate posterior parameters with a *, for notational simplicity, and will substitute values of posterior parameters at end.

$$\begin{aligned} p(\omega|y) &= \int_{-\infty}^{\infty} p(\theta,\omega|y)\delta\theta \\ &\propto \int_{-\infty}^{\infty} \omega^{\left(\frac{(d^*+1)}{2}-1\right)} \exp[-\frac{\omega}{2}k^*(\theta-\mu^*)^2]\exp[-\omega\frac{\eta^*}{2}]\delta\theta \\ &= \omega^{\left(\frac{(d^*+1)}{2}-1\right)} e^{\left(-\omega\frac{\eta^*}{2}\right)} \int_{-\infty}^{\infty} \exp[-\frac{\omega}{2}k^*(\theta-\mu^*)^2]\delta\theta \end{aligned}$$

The integral is a normal kernel: $N\left(\mu^*, (\omega k^*)^{-1}\right)$. It integrates to $\frac{1}{c}$, where c is the constant of proportionality for the normal density. The integral $= \frac{1}{\omega^{1/2}} = \omega^{-\frac{1}{2}}$.

$$\begin{aligned} &\propto \omega^{\left(\frac{(d^*+1)}{2}-1\right)} e^{\left(-\omega\frac{\eta^*}{2}\right)} \omega^{-\frac{1}{2}} \\ &= \omega^{\left(\frac{d^*}{2}-1\right)} e^{\left(-\omega\frac{\eta^*}{2}\right)} \end{aligned}$$

This is the kernel for *gamma* $\left(\frac{d^*}{2}, \frac{\eta^*}{2}\right)$, where

$$\begin{aligned} d^* &= (n+d) \\ \eta^* &= \frac{nk(\bar{y}-\mu)^2}{(n+k)} + S_y + \eta \end{aligned}$$

Part E

From C and D, we know the marginal posterior $p(\theta|y)$ is a gamma mixture of normals. Show this takes the form of a centered, scaled t distribution of form $p(\theta) \propto \left(1 + \frac{1}{\nu} \cdot \frac{(x-m)^2}{s^2}\right)^{-\frac{\nu+1}{2}}$ and state what m, s and v are.

Joint posterior $p(\theta, \omega|y)$:

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

Marginal posterior $p(\theta|y)$:

$$\begin{aligned} p(\theta|y) &= \int_0^\infty p(\theta,\omega|y)\delta\omega \\ &= \int_0^\infty \omega^{\frac{d^*+1}{2}-1} \cdot \exp\left[-\frac{\omega}{2}\left(k^*(\theta-\mu^*)^2 + \frac{\eta^*}{2}\right)\right]\delta\omega \end{aligned}$$

This is the kernel of a *gamma* $\left(\frac{(d^*+1)}{2}, \frac{k^*(\theta-\mu^*)^2+\eta^*}{2}\right)$ distribution.

The integral will therefore equal $\frac{1}{c}$, where c is the constant of proportionality for this gamma density, $\frac{1}{\beta^\alpha/\Gamma(\alpha)} = \frac{\Gamma(\alpha)}{\beta^\alpha}$.

$$\begin{aligned} &= \Gamma(\frac{d^*+1}{2})\left[\frac{k^*(\theta-\mu^*)^2}{2} + \frac{\eta^*}{2}\right]^{-\frac{d^*+1}{2}} \\ &\propto \left[\left(\frac{\eta^*}{2}\right)\left(1 + \frac{k^*(\theta-\mu^*)^2}{\eta^*}\right)\right]^{-\frac{d^*+1}{2}} \\ &= \left(\frac{\eta^*}{2}\right)^{-\frac{d^*+1}{2}}\left[\left(1 + \frac{k^*(\theta-\mu^*)^2}{\eta^*}\right)\right]^{-\frac{d^*+1}{2}} \end{aligned}$$

The left term cancels, as it is constant wrt θ.

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

Multiply numerator and denominator by d^*.

$$\begin{aligned} &= \left[\left(1 + \frac{d^*k^*(\theta-\mu^*)^2}{d\eta^*}\right)\right]^{-\frac{d^*+1}{2}} \\ &= \left[\left(1 + \left(\frac{1}{d^*}\right)\frac{(\theta-\mu^*)^2}{\frac{\eta^*}{d^*k^*}}\right)\right]^{-\frac{d^*+1}{2}} \end{aligned}$$

This has the desired central-scaled t form, with mean $m = \mu^*$, df $\nu = d^*$, and scale $s^2 = \frac{\eta^*}{d^*k^*}$.

Part F

True or False? In the limit, as prior parameters $k, d, \eta \to 0$, priors $p(\theta)$ and $p(\omega)$ are valid probability distributions. (Must integrate to 1, or something finite so can be normalized over their domains.)

<u>$p(\theta)$:</u>
The prior $p(\theta)$ is a gamma mixture of normals, which can arrange as a scaled, centered t.

$$\lim_{d,k,\eta \to 0} \Gamma\left(\frac{d+1}{2}\right)\left[\frac{k(\theta-\mu)^2}{2}+\frac{\eta}{2}\right]^{-\frac{d+1}{2}}$$

Goes to:

$$\Gamma\left(\frac{1}{2}\right)\left[\frac{0(\theta-\mu)^2}{2}+\frac{0}{2}\right]^{-\frac{1}{2}} \to 0$$

This is not a valid probability distribution. It is a point mass at 0. It is an improper prior when the prior parameters go to zero.

<u>$p(\omega)$:</u>
The prior $p(\omega)$ is $gamma(\frac{d}{2}, \frac{\eta}{2})$.

$$p(\omega) = \frac{(\frac{\eta}{2})^{(\frac{d}{2})}}{\Gamma(\frac{d}{2})}\omega^{\frac{d}{2}-1}e^{-\omega\frac{\eta}{2}}$$

$$\lim_{d,k,\eta \to 0} \frac{(\frac{\eta}{2})^{(\frac{d}{2})}}{\Gamma(\frac{d}{2})}\omega^{\frac{d}{2}-1}e^{-\omega\frac{\eta}{2}}$$

Goes to:

$$(1)(\omega^{-1})(1) = \omega^{-1} = \omega^{0-1}e^{-0\omega}$$

This goes to ω^{-1}, which has an infinite integral in the area of zero. This is also the form of $gamma(0,0)$, which is not a valid probability distribution, since gamma limits both parameters to values greater than zero. This is also an improper prior.

The answer is False for both prior distributions; neither are proper probability distributions.

Part G

True or False? In the limit, as prior parameters $k, d, \eta \to 0$, priors $p(\theta|y|)$ and $p(\omega|y|)$ are valid probability distributions. (Must integrate to 1, or something finite so can be normalized over their domains.)

As the hyperparameters go to zero, the posterior parameters are as follows.

$$d^* = (n+d) \to d^* = n \tag{5}$$
$$k^* = (n+k) \to k^* = n \tag{6}$$
$$\mu^* = \frac{n\bar{y} + k\mu}{(n+k)} \to \mu^* = \bar{y} \tag{7}$$
$$\eta^* = \frac{nk(\bar{y}-\mu)^2}{(n+k)} + S_y + \eta \to \eta^* = S_y \tag{8}$$

<u>$p(\theta|y)$:</u>

The marginal posterior $p(\theta|y)$ is also a gamma mixture of normals, which can be written as a scaled, centered t. Plug the updated posterior parameters into the central t form.

$$\left[\left(1 + \left(\frac{1}{d^*}\right)\frac{(\theta-\mu^*)^2}{\frac{\eta^*}{d^*k^*}}\right)\right]^{-\frac{d^*+1}{2}} \quad \to \quad \left[\left(1 + \left(\frac{1}{n}\right)\frac{(\theta-\bar{y})^2}{\frac{S_y}{n^2}}\right)\right]^{-\frac{n+1}{2}}$$

This is a valid probability distribution. It has form of the centered, scaled t distribution, with parameters

$$m = \bar{y}$$
$$s^2 = \frac{S_y}{n^2} = \frac{\sum_{i=1}^n (y-\bar{y})^2}{n^2} = \frac{S}{n}, \text{ where } S^2 = \hat{\sigma^2} = \frac{\sum_{i=1}^n (y_i-\bar{y})^2}{n} \text{ is the sample variance}$$
$$\nu = n$$

<u>$p(\omega|y)$:</u>

The marginal posterior $p(\omega|y)$ is *gamma* $\left(\frac{d^*}{2}, \frac{\eta^*}{2}\right)$. Plug the updated posterior parameters into the gamma density.

$$p(\omega|y|) = \frac{(\frac{\eta^*}{2})^{(\frac{d^*}{2})}}{\Gamma(\frac{d^*}{2})}\omega^{\frac{d^*}{2}-1}e^{-\omega\frac{\eta^*}{2}} \quad \to \quad p(\omega|y|) = \frac{(\frac{S_y}{2})^{(\frac{n}{2})}}{\Gamma(\frac{n}{2})}\omega^{\frac{n}{2}-1}e^{-\omega\frac{S_y}{2}}$$

This is a valid probability distribution. It is *Gamma* $\left(\frac{n}{2}, \frac{S_y}{2}\right)$.

The answer is True for both posterior marginal distributions. Both are valid probability distributions. Morevoer, their parameters are intuitive based on priors assuming zero information.

Part H

The Bayesian Credible Interval has the form $\theta \in m \pm t^* \cdot s$ where m and s are the posterior mean and scale, and t^* is the appropriate t critical value using the posterior degrees of freedom: $t^* = t_{\nu^*, 1-\frac{\alpha}{2}}$.

True or False? In the limit, as prior parameters $k, d, \eta \rightarrow 0$, the Bayes Credible Interval equals the frequentist confidence interval for θ.

This is true. We can use the marginal posterior of $\theta|y$ from Part E, which is centered and scaled t.

$$p(\theta|y) \sim t\left(m = \mu^*, \nu = d^*, s^2 = \frac{\eta^*}{d^* k^*}\right)$$

where, in the limit, the posterior parameters go to

$$\begin{aligned}
d^* &= (n+d) \rightarrow d^* = n \\
k^* &= (n+k) \rightarrow k^* = n \\
\mu^* &= \frac{n\bar{y} + k\mu}{(n+k)} \rightarrow \mu^* = \bar{y} \\
\eta^* &= \frac{nk(\bar{y}-\mu)^2}{(n+k)} + S_y + \eta \rightarrow \eta^* = S_y
\end{aligned}$$

The Bayes Credible Interval goes to

$$\begin{aligned}
&m^* \pm t^* s^* \\
&\mu^* \pm t^* \sqrt{\frac{\eta^*}{d^* k^*}} \\
&\bar{y} \pm t^* \sqrt{\frac{S_y}{n^2}} \\
&\bar{y} \pm t^* \frac{S}{\sqrt(n)}, \text{ where } S^2 = \frac{\sum_{i=1}^n (y_i - \bar{y})^2}{n}
\end{aligned}$$

The Bayesian Credible Interval converges to the frequentist confidence interval.

The Conjugate Gaussian Linear Model: Basics

Consider the Gaussian linear model

$$(y|\beta,\sigma^2) \sim N\left(X\beta,(\omega\Lambda)^{-1}\right)$$

where y is an n-vector of responses, X is a (nxp) feature matrix, and $\omega = 1/\sigma^2$ is the error precision, and Λ is some known matrix. Consider these priors:

$$\begin{aligned}(\beta|\omega) &\sim N(m,(\omega K)^{-1})\\ \omega &\sim Gamma(d/2,\eta/2)\end{aligned}$$

Part A

Derive the conditional posterior of $p(\beta|y,\omega)$.

Begin by deriving the joint posterior $p(\beta,\omega|y)$.

Joint posterior $\propto$ sampling distribution * joint prior.

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

Sampling distribution:

$$\begin{aligned}&(y|\beta,\sigma^2) \sim N\left(X\beta,(\omega\Lambda)^{-1}\right) \rightarrow\\ &p(y|\beta,\sigma^2) \propto \omega^{\frac{n}{2}}\exp\left[-\frac{\omega}{2}(y-X\beta)^T\Lambda(y-X\beta)\right]\end{aligned}$$

Joint prior:

$$\begin{aligned}&p(\beta,\omega) \propto p(\beta|\omega)\cdot p(\omega) \rightarrow\\ &p(\beta,\omega) \propto \omega^{\frac{p}{2}}\exp\left[-\frac{\omega}{2}(\beta-m)^TK(\beta-m)\right]\cdot\omega^{\frac{d}{2}-1}\exp\left[-\omega\frac{\eta}{2}\right]\end{aligned}$$

Joint posterior:

$$\begin{aligned}p(\beta,\omega|y) &\propto \omega^{\frac{n}{2}}\exp\left[-\frac{\omega}{2}(y-X\beta)^T\Lambda(y-X\beta)\right]\cdot\omega^{\frac{p}{2}}\exp\left[-\frac{\omega}{2}(\beta-m)^TK(\beta-m)\right]\cdot\omega^{\frac{d}{2}-1}\exp\left[-\omega\frac{\eta}{2}\right]\\ &= \omega^{\frac{n+p+d}{2}-1}\exp\left[-\frac{\omega}{2}\left\{(y-X\beta)^T\Lambda(y-X\beta)+(\beta-m)^TK(\beta-m)+\eta\right\}\right]\end{aligned}$$

Simplify the term multplied by $-\frac{\omega}{2}$ in the exponential term.

$$\begin{aligned}&\left\{(y-X\beta)^T\Lambda(y-X\beta)+(\beta-m)^TK(\beta-m)+\eta\right\}\\ &= y^T\Lambda y - 2\beta^TX^T\Lambda y + \beta^TX^T\Lambda X\beta + \beta^TK\beta - 2m^TK\beta + m^TKm + \eta\\ &= \beta^T\left(X^T\Lambda X + K\right)\beta - 2\beta^T\left(X^T\Lambda y + Km\right) + y^T\Lambda y + m^TKm + \eta\end{aligned}$$

This has form $x^TCx - 2x^Tb + a$, so we can complete the square in matrix form.

$$x^TCx{-}2x^Tb + a \rightarrow (x-m)^T M(x-m) + v, \text{ with}$$

$$\begin{aligned}
M &= C & &\rightarrow & M &= (X^T\Lambda X + K) \\
m &= -C^{-1}b & &\rightarrow & m &= -(X^T\Lambda X + K)^{-1}(X^T\Lambda y + Km) \\
v &= a - b^TC^{-1}b & &\rightarrow & v &= y^T\Lambda y + m^TKm + \eta - (m^TK + X^T\Lambda y)(X^T\Lambda X + K)^{-1}(X^T\Lambda y + Km)
\end{aligned}$$

Plug this result back into the entire joint posterior to obtain the normal-gamma mixture joint posterior:

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

where the posterior parameters are as follows.

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

We can read off the conditional posterior for β, which is Normal.

$$\begin{aligned}
p(\beta|\omega, y) &\sim N\left(m^*, (\omega K^*)^{-1}\right), \text{ with} \\
K^* &= (X^T\Lambda X + K) \\
m^* &= (K^*)^{-1}(X^T\Lambda y + Km)
\end{aligned}$$

Note regarding simplification of η^*:

$$\eta^* = y^T\Lambda y + m^TKm + \eta - (m^TK + X^T\Lambda y)(X^T\Lambda X + K)^{-1}(X^T\Lambda y + Km)$$

Write

$$\begin{aligned}
&(m^TK + X^T\Lambda y)(X^T\Lambda X + K)^{-1}(X^T\Lambda y + Km) \\
&= (Km + X^T\Lambda y)(X^T\Lambda X + K)^{-1}I(X^T\Lambda y + Km), \text{ and } I = (X^T\Lambda X + K)^{-1}(X^T\Lambda X + K) \\
&= (m^TK + X^T\Lambda y)(X^T\Lambda X + K)^{-1}\left(X^T\Lambda X + K\right)(X^T\Lambda X)^{-1}(X^T\Lambda y + Km) \\
&= m^{*T}K^*m^*
\end{aligned}$$

Then $\eta^* = y^T\Lambda y + m^TKm + \eta - (m^TK + X^T\Lambda y)(X^T\Lambda X)^{-1}(X^T\Lambda y + Km)$.

Part B

Derive the marginal posterior of $p(\omega|y)$.

Derive the marginal posterior $p(\omega|y)$ by integrating the joint posterior derived in the previous section, $p(\beta,\omega|y)$ wrt ω.

$$\begin{aligned}
p(\omega|y) &\propto \int_{-\infty}^{\infty} p(\beta,\omega|y)\delta\beta \\
&= \int_{-\infty}^{\infty} \omega^{\frac{d^*}{2}-1} \underbrace{\omega^{\frac{p}{2}} \exp\left[-\frac{\omega}{2}(\beta-m^*)^T K^*(\beta-m^*)\right]}_{\text{(i)}} \exp\left[-\omega\frac{\eta^*}{2}\right]\delta\beta
\end{aligned}$$

The term **(i)** is the kernel of multivariate normal $N\left(m^*, (\omega K^*)^{-1}\right)$, so integrates to the inverse of the constant of proportionality, $(\omega K^*)^{-\frac{p}{2}}$.

$$\begin{aligned}
&= \omega^{\frac{d^*+p}{2}-1}\omega^{-\frac{p}{2}} \exp\left[-\omega\frac{\eta^*}{2}\right] \\
&= \omega^{\frac{d^*}{2}-1} \exp\left[-\omega\frac{\eta^*}{2}\right]
\end{aligned}$$

The marginal posterior $p(\omega|y) \sim Gamma\left(\frac{d^*}{2}, \frac{\eta^*}{2}\right)$, with

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

Part C

Derive the marginal posterior $p(\beta|y)$. Can do this by integrating the joint posterior over ω.

$$
\begin{aligned}
p(\beta|y) &\propto \int_0^\infty p(\beta,\omega|y)\delta\omega \\
&= \int_0^\infty \omega^{\frac{d^*}{2}-1}\omega^{\frac{p}{2}} \exp\left[-\frac{\omega}{2}(\beta-m^*)^T K^*(\beta-m^*)\right] \exp\left[-\omega\frac{\eta^*}{2}\right]\delta\omega \\
&= \int_0^\infty \omega^{\frac{d^*+p}{2}-1} \exp\left[-\omega\left\{\frac{\{(\beta-m^*)^T K^*(\beta-m^*)+\eta^*\}}{2}\right\}\right]\delta\omega
\end{aligned}
$$

This is the kernel of a $Gamma\left(\frac{d^*+p}{2}, \frac{(\beta-m^*)^T K^*(\beta-m^*)+\eta^*}{2}\right)$, so the integral equals $1/c = \frac{\Gamma(a)}{b^a}$.

$$
\begin{aligned}
&= \Gamma\left(\frac{d^*+p}{2}\right)\left[\frac{(\beta-m^*)^T K^*(\beta-m^*)+\eta^*}{2}\right]^{-\left(\frac{d^*+p}{2}\right)} \\
&\propto \left[\frac{\eta^*}{2}\left(1+\frac{(\beta-m^*)^T K^*(\beta-m^*)}{\eta^*}\right)\right]^{-\left(\frac{d^*+p}{2}\right)} \\
&\propto \left[1+\left(\frac{1}{d^*}\right)\frac{(\beta-m^*)^T K^*(\beta-m^*)}{\frac{\eta^*}{d^*}}\right]^{-\left(\frac{d^*+p}{2}\right)}
\end{aligned}
$$

This is the form of the multivariate centered, scaled T distribution, with

$$
\begin{aligned}
\text{mean} &= m^* \\
\text{df} &= d^* \\
s^2 &= \left(\frac{\eta^*}{d^*}\right)(K^*)^{-1}
\end{aligned}
$$

Part D

I fit the Bayesian linear model, choosing Λ and (vague) prior parameters as follows.

$$\begin{aligned}
\Lambda &= I_n \\
m &= rep(0, p), \text{ the zero vector} \\
d &= .02 \\
\eta &= .02 \\
K &= \begin{bmatrix} .02 & 0 \\ 0 & .02 \end{bmatrix}
\end{aligned}$$

The figure below compares the frequentist linear model with the Bayesian version.

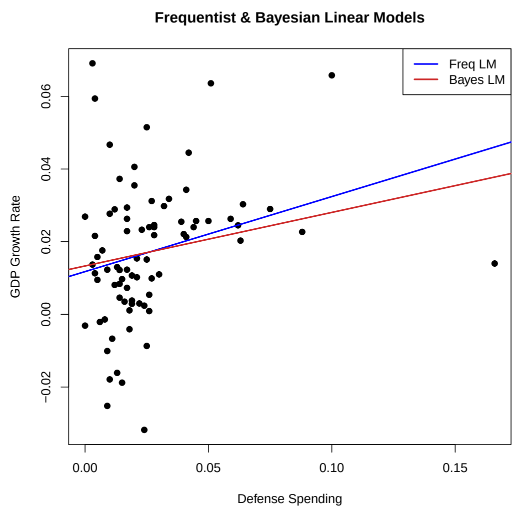

Figure 1: Frequentist and Bayesian Linear Models

These models are practically identical, which is expected given the non-informative priors. I am unhappy with both of these models, as both look to be influenced by outliers. The Bayesian line is slightly less influenced than the frequentist line. (If choose priors as .01 instead of .02, the lines are even closer.)

The Conjugate Gaussian Linear Model: A Heavy-Tailed Error Model

The full model is as follows.

$$
\begin{aligned}
(y|\beta,\omega,\Lambda) &\sim N(X\beta,(\omega\Lambda)^{-1}) \\
\Lambda &= \text{diag}\,(\lambda_1,...,\lambda_n) \\
\lambda_i &\overset{iid}{\sim} Gamma(\frac{h}{2},\frac{h}{2}) \\
(\beta|\omega) &\sim N\left(m,(\omega K)^{-1}\right) \\
\omega &\sim Gamma\left(\frac{d}{2},\frac{\eta}{2}\right)
\end{aligned}
$$

where h is a fixed hyperparameter.

Part A

What is the implied conditional distribution of $p(y_i|\beta,\omega)$, with λ_i marginalized out? This should look familiar.

The implied conditional distribution of $p(y_i|\beta,\omega)$ is

$$
\begin{aligned}
p(y_i|\beta,\omega) &= \int p(y_i|X,\beta,\omega,\lambda_i)p(\lambda_i|X,\beta,\omega)p(\beta|\omega)p(\omega)\delta\lambda_i \\
&\propto \int p(y_i|\lambda_i,X,\beta,\omega)p(\lambda_i)\delta\lambda_i \\
&\propto \int_0^\infty \omega^{\frac{1}{2}}\lambda_i^{\frac{1}{2}}\exp\left[-\frac{\omega\lambda_i}{2}\left(y_i-x_i'\beta\right)^2\right]\cdot\lambda_i^{\frac{h}{2}-1}\exp\left[-\lambda_i\frac{h}{2}\right]\delta\lambda_i \\
&\propto \int_0^\infty \lambda_i^{\frac{h+1}{2}-1}\exp\left[-\lambda_i\left\{\frac{\omega\left(y_i-x_i'\beta\right)^2+h}{2}\right\}\right]\delta\lambda_i
\end{aligned}
$$

This is the kernel of $Gamma\left(\frac{h+1}{2},\frac{\left(y_i-x_i'\beta\right)^2+h}{2}\right)$, so integrates to $\frac{1}{c}=\frac{\Gamma(a)}{b^a}$.

$$
\begin{aligned}
&= \Gamma\left(\frac{h+1}{2}\right)\left[\frac{\omega\left(y_i-x_i'\beta\right)^2+h}{2}\right]^{-\left(\frac{h+1}{2}\right)} \\
&\propto \left[\frac{h}{2}\left(1+\frac{\omega\left(y_i-x_i'\beta\right)^2}{h}\right)\right]^{-\left(\frac{h+1}{2}\right)} \\
&\propto \left[1+\left(\frac{1}{h}\right)\frac{(y_i-x_i'\beta)^2)}{\left(\frac{1}{\omega}\right)}\right]^{-\left(\frac{h+1}{2}\right)}
\end{aligned}
$$

Therefore,

$$
p(y_i|X,\beta,\omega)\sim t_h\left(m=x_i'\beta,s^2=\frac{1}{\omega}\right)
$$

Part B

What is the conditional posterior distribution of $p(\lambda_i|y,\beta,\omega)$?

Note: Λ diagonal, each y_i depends only on its corresponding λ_i.

The conditional posterior $p(\lambda_i|y,\beta,\omega)$ is

$$\begin{aligned}
p(\lambda_i|y,\beta,\omega) &\propto p(y|\lambda_i,\beta,\omega)p(\lambda_i|\beta,\omega)p(\beta,\omega) \\
&\propto p(y_i|\lambda_i,\beta,\omega)p(\lambda_i) \\
&\propto \omega^{\frac{1}{2}}\lambda_i^{\frac{1}{2}} \exp\left[-\frac{\omega\lambda_i}{2}(y_i - x_i'\beta)^2\right] \cdot \lambda_i^{\frac{h}{2}-1} \exp\left[-\lambda_i \frac{h}{2}\right] \\
&\propto \lambda_i^{\frac{h+1}{2}-1} \exp\left[-\lambda_i \left\{\frac{\omega\left(y_i - x_i'\beta\right)^2 + h}{2}\right\}\right]
\end{aligned}$$

Therefore,

$$p(\lambda_i|y,\beta,\omega) \sim Gamma\left(\frac{h+1}{2}, \frac{\omega\left(y_i - x_i'\beta\right)^2 + h}{2}\right)$$

Part C

Combining the results from the **Basics** subsection, code a Gibbs Sampler that cycles repeatedly through the following three sets of posterior conditional distributions.

- $p(\beta|y,\omega,\Lambda) \sim N\left(m^*, (\omega K^*)^{-1}\right)$
- $p(\omega|y,\Lambda) \sim Gamma\left(\frac{d^*}{2}, \frac{\eta^*}{2}\right)$
- $p(\lambda_i|y,\beta,\omega) \sim Gamma\left(\frac{h+1}{2}, \frac{\omega\left(y_i - x_i'\beta\right)^2 + h}{2}\right)$

Recap of posterior parameters.

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

We can use the same posterior parameters as the previous section, and set h=1. I ran the Gibbs sampler for 11,000 iterations, throwing away the first 1,000 as a burn-in period, and thinning every second observation to decrease autocorrelation, leaving 5,000 observations sampled from the joint posterior. Figure 2 shows that the beta estimates have converged.

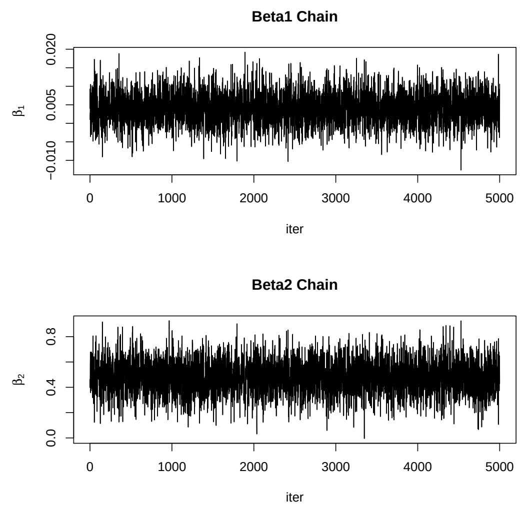

Figure 2: Convergence of the beta regression parameters

The heavy-tailed Bayes regression appears more robust to outliers than the previous two (regular Bayesian and frequentist) models. The posterior means of the λ_i values represent the posterior precision for each country, and we can see that there are countries with precisions smaller than the others, ie variances larger than the others.

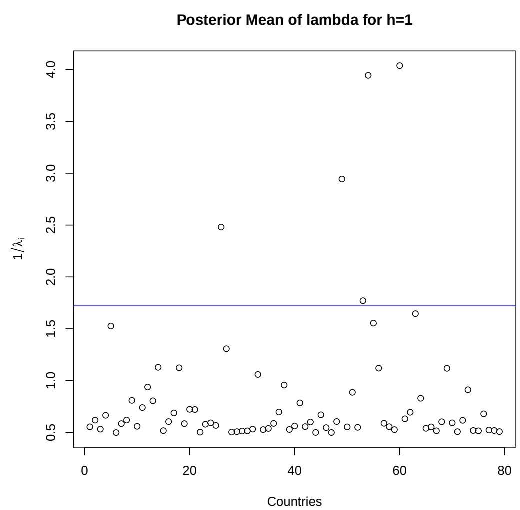

Figure 3: Plot of variances $1/\lambda_i$ by Country

We can highlight these points on the scatterplot of the data and the posterior regression line, and identify them as outliers. This hierarchical model setup could almost be thought of as an outlier detection procedure.

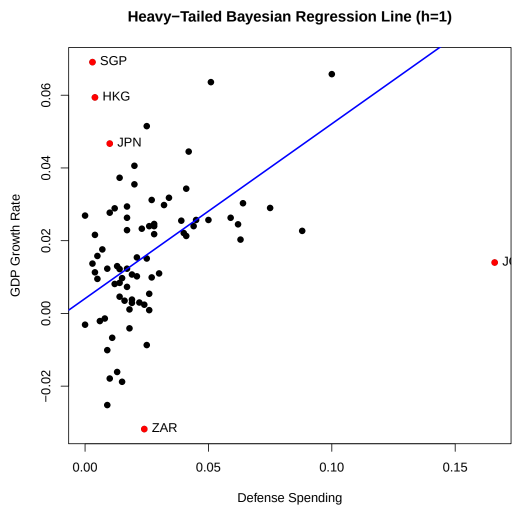

Figure 4: Heavy tailed Bayesian regression line with outliers highlighted

I ran the same Gibbs sampler for a variety of h values, to analyze the impact of the h selection. The h value selected does impact which of the countries are considered outliers. Larger h means the variance of the λ_i values is larger, since $var = a/b^2$ for a gamma distribution, ie $\frac{h/2}{h^2/2^2} = \frac{2}{h}$ in this case. Larger h means smaller variance in the λ_i values. This is seen on the y-axis of the plots below. Outliers do still float to the top, regardless of the choice of h. The same most-extreme countries are seen consistently, but the borderline countries may vary slightly depending on h.

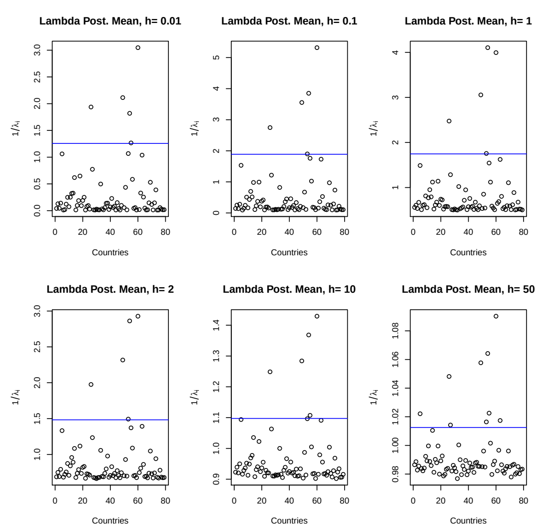

Figure 5: Varying h values under vague priors

Appendix: R Code

Frequentist and Bayesian Regression Comparison

```
#Stats Modeling 2 — Exercise 2
#Jennifer Starling
#Feb 6, 2017

#Housekeeping
rm(list=ls())

#Read data file.
data = read.csv('/Users/jennstarling/UTAustin/2017S_Stats Modeling 2/statsmod/Course-
    Data/gdpgrowth.csv',header=T)

#Load libraries.
library(mvtnorm)

#Consider a linear model with intercept for GDP growth rate (GR6096) versus
#its level of defense spending as a fraction of GDP (DEF60).

#Fit the Bayesian linear model to this data set, choosing Lambda = I, and
#something diagonal and vague for the prior precision matrix K = diag(k1,k2).

#Inspect the fitted line (graphically).  Are you happy with the fit?  Why/not?

#————————————————————————————————————————
#Set up model variables.
n = nrow(data)
X = cbind(rep(1,n),data$DEF60)
y = data$GR6096
p = ncol(X)

#####################################
###   Frequentist linear model.   ###
#####################################

freq_lm = lm(y~X-1)
summary(freq_lm)
beta_hat = freq_lm$coefficients

#Visually inspect fit.
#pdf(file='/Users/jennstarling/UTAustin/2017S_Stats Modeling 2/Exercise—02/LaTeX Files
    /Figures/Frequentist_Linear_Model.pdf')
plot(X[,2],y,pch=19,col='black',xlab='X',ylab='y',main='Frequentist Linear Model')
abline(a=beta_hat[1], b=beta_hat[2],lwd=2,col='blue')
#dev.off()

################################################
###   Bayesian linear model (vague prior).   ###
################################################

#   (y | B, sigma2) ~ N(XB, (wL)^{—1})
#   (B|w) ~ N(m, (wK)^{—1})
#   w ~ Gamma(d/2,eta/2)

#Set up L matrix and some vague prior parameters.
L = diag(n) #Begin with Lambda = I.
```

```
m <- rep(0, p)
d = .02
eta = .02
K = diag(c(.02,.02))

#Precache X^T %*% L %*% X for use in hyperparameter calculation.
XtLX = t(X) %*% L %*% X

#Update posterior parameters.
d_star = d+n
K_star = XtLX + K
m_star = solve(K_star) %*% (t(X) %*% L %*% y + K %*% m)
eta_star = eta + t(y) %*% y + t(m) %*% K %*% m - t(m_star) %*% K_star %*% m_star

# Calculate posterior beta_hat estimate.  This is posterior mean of beta.
# From Part C, p(beta|y) ~ t(m*,d*,K)
beta_hat_post = m_star

#Plot frequenstist and Bayesian lm results over data for comparison.
pdf(file='/Users/jennstarling/UTAustin/2017S_Stats Modeling 2/Exercise-02/LaTeX Files/
    Figures/Freq_Bayes_LM_Compare.pdf')
plot(X[,2],y,pch=19,col='black',
    main='Frequentist & Bayesian Linear Models',
    xlab='Defense Spending',ylab='GDP Growth Rate')
abline(a=beta_hat[1], b=beta_hat[2],lwd=2,col='blue')
abline(a=beta_hat_post[1],b=beta_hat_post[2],lwd=2,col='firebrick3')
legend('topright',legend=c("Freq LM","Bayes LM"),lwd=2,lty=1,col=c('blue','firebrick3'
    ))
dev.off()
```

Heavy-Tailed Bayesian Regression

```
#Stats Modeling 2
#Exercise 02 - Gibbs Sampler for Bayesian Linear Model

#Housekeeping
rm(list=ls())

#Full hierarchical model:
#   (y | B,w,L) ~ N(XB,(wL)^{-1})
#   L = diag(l1,...,ln) #Lambda_i values.
#   li ~ iid Gamma(h/2,h/2)
#   (B|w) ~ N(m,(wK)^{-1})
#   w ~ Gamma(d/2,eta/2)

#Conditional posteriors:
#   p(B|y,w,L) ~ N(m_star,(wK_star)^{-1})
#   p(w|y,L) ~ Gamma(d_star/2,eta_star/2)
#   p(li | y,B,w) ~ Gamma( (h+1)/2, (w(y_i - x_i'B)^2 + h)/2 )

#Posterior parameters:
#   d_star = d + n
#   K_star = (X^T %*% L %*% X + K)
#   m_star = (K_star)^{-1} (X^T %*% L %*% y + K %*% m)
            = (X^T %*% L %*% X + K)^{-1} %*% (X^T %*% L %*% y + K %*% m)
#   eta_star = eta + y^T %*% y + m^T %*% K %*% M - m_star^T %*% K_star %*% m_star
```

```
#———————————————————————————————————————————
gibbs = function(iter=11000,burn=1000,thin=2){
    #PURPOSE: Gibbs Sampler for model described above.
    #INPUTS:    iter = number of posterior samples to generate.
    #           burn = number of burn—in values to discard.
    #           thin = thinning to reduce autocorr of chain.
    #OUTPUT:    gibbs = matrix of values sampled from posterior.  Rows = samples.

    #Require mvtnorm for sampling from multivariate normal.
    require(mvtnorm)

    # Set up objects to hold sampled posterior results.
    mat = matrix(0,nrow=iter, ncol = p + 1 + n)
        # Since lambda is a diagonal matrix, storing just diagonal
        # elements diag(lambda_1...lambda_n), as
        # part of mat instead of storing list of matrices.

    #Initialize each element of chain.
    lambda = rep(1,n)
    beta = rep(0,p)
    omega = 1

    #Iterate through sampler.
    for (i in 1:iter){

        #1. Update beta & relevant posterior parameters.
        XtLX = t(X) %*% diag(lambda) %*% X #Precache to save time.
        K_star = XtLX + K
        m_star = solve(K_star) %*% (t(X) %*% diag(lambda) %*% y + K %*% m)

        beta = rmvnorm(1,m_star,(1/omega) * solve(K_star))

        #2. Update omega & relevant posterior parameters.
        XtLX = t(X) %*% diag(lambda) %*% X #Precache to save time.
        d_star = d + n
        eta_star = t(y) %*% diag(lambda) %*% y + eta + t(m) %*% K %*% m - t(m_star) %*
            % K_star %*% m_star
        #Note: m_star, K_star do not change from beta update; no need to update here.

        omega = rgamma(1,d_star/2,rate=eta_star/2)

        #3. Update lambda.
        lambda = rgamma(n,(h+1)/2, rate = (omega * (y-X %*% t(beta))^2 + h)/2)

        #Add values to results matrix.
        mat[i,] = c(beta,omega,lambda)
    }

    #Set up variable names for matrix of results.
    colnames(mat) = c(unlist(strsplit(paste("beta",1:p,sep=""),"  ")),
        "omega",
        unlist(strsplit(paste("lambda",1:n,sep=""),"  ")))

    #Keep only values after burn—in period.
    mat = mat[(burn+1):iter,]

    #Thin values.
    mat = mat[seq(1,nrow(mat),by=2),]
```

```
    #Return results.
    return(mat)
}
#————————————————————————————————————————

### Read data file & set up model variables (n,p,etc).
data = read.csv('/Users/jennstarling/UTAustin/2017S_Stats Modeling 2/statsmod/Course-Data/gdpgrowth.csv',header=T)

n = nrow(data)
X = cbind(rep(1,n),data$DEF60)
y = data$GR6096
p = ncol(X)

#Initialize prior parameters.
m <- rep(0, p)
d = .1
eta = .1
K = diag(c(.5,.5))

h = 1

### Run sampler to obtain samples from joint posterior p(beta,omega,lambda|y).
output = gibbs(iter=11000,burn=1000,thin=2)

### Analysis:

#Plot beta_hat values to show chain stabilized.
pdf(file='/Users/jennstarling/UTAustin/2017S_Stats Modeling 2/Exercise-02/LaTeX Files/Figures/Gibbs_beta_stabilized.pdf')
par(mfrow=c(2,1))
plot(1:nrow(output),output[,1],type='l',main='Beta1 Chain',xlab='iter',ylab=expression(beta[1]))
plot(1:nrow(output),output[,2],type='l',main='Beta2 Chain',xlab='iter',ylab=expression(beta[2]))
dev.off()

#Posterior mean of beta.
beta_post_mean = colMeans(output[,1:2])

#Posterior mean of sigma^2
sigma2_post_mean = 1/mean(output[,2])

### Lambda_i precision plot.

#Calculate posterior mean of 1/lambda_i for each country (each row X_i is a country).
#We notice that five of the 1/lambda_i values are larger than the others.
#These correspond to observations with lower precision, ie higher variance.
lam_post_mean_inv = 1/colMeans(output[,4:ncol(output)])
idx_outliers = which(lam_post_mean_inv >= sort(lam_post_mean_inv,decreasing=T)[5])
country_outliers = as.character(data$CODE[idx_outliers])

pdf(file='/Users/jennstarling/UTAustin/2017S_Stats Modeling 2/Exercise-02/LaTeX Files/Figures/Lambda_i_Inverse.pdf')
plot(1:n,lam_post_mean_inv,main= 'Posterior Mean of lambda for h=1',xlab='Countries',ylab=expression(1/lambda[i]))
abline(h=sort(lam_post_mean_inv,decreasing=T)[5] - .05,col='blue') #Plots h—line below top 5 points.
```

```
dev.off()

#Heavy-tailed Bayesian regression line w outliers labeled.
pdf(file='/Users/jennstarling/UTAustin/2017S_Stats Modeling 2/Exercise-02/LaTeX Files/Figures/Gibbs_HeavyTail_Bayes_LM.pdf')
plot(X[,2],y,pch=19,col='black',xlab='Defense Spending',ylab='GDP Growth Rate',main='Heavy-Tailed Bayesian Regression Line (h=1)')
abline(a=beta_post_mean[1], b=beta_post_mean[2],lwd=2,col='blue')
points(X[idx_outliers,2],y[idx_outliers],pch=19,col='red')
text(X[idx_outliers,2],y[idx_outliers],labels=country_outliers,pos=4,offset=.5)
dev.off()

#This plot confirms that these observations are those furthest from the regression line.
#The regression line is more robust to these outliers than the previous lines (Frequentiest and Bayesian).

### Analysis of choice of h under vague priors:

#How does choice h affect things?  Used h=1 previously.  Try a few different h values.

#Look at 1/lambda_i precision plots for a variety of h values.
pdf(file='/Users/jennstarling/UTAustin/2017S_Stats Modeling 2/Exercise-02/LaTeX Files/Figures/H_Compare.pdf')
par(mfrow=c(2,3))
H = c(.01,.1,1,2,5,10)
for (g in H){
    h = g
    output = gibbs(iter=11000,burn=1000,thin=2)
    plot(1:n,lam_post_mean_inv,main= paste('Lambda Post. Mean, h=',h),xlab='Countries',ylab=expression(1/lambda[i]))
    abline(h=sort(lam_post_mean_inv,decreasing=T)[5] - .01,col='blue') #Plots h-line below top 5 points.
}
dev.off()
```