1 A simple Gaussian location model

Take a simple Gaussian model with unknown mean and variance:

$$(y_i \mid \theta, \sigma^2) \sim N(\theta, \sigma^2) \, , \quad i = 1, \ldots, n \, . \tag{1}$$

Let $\mathbf{y}$ be the vector of observations $\mathbf{y} = (y_1, \ldots, y_n)^T$.

Suppose we place conjugate normal and inverse-gamma priors on θ and σ^2, respectively:

$$\begin{aligned} (\theta \mid \sigma^2) &\sim N(\mu, \tau^2\sigma^2) \\ \sigma^2 &\sim \text{Inv-Gamma}\left(\frac{d}{2}, \frac{\eta}{2}\right) \, , \end{aligned}$$

where μ, $\tau > 0$, $d > 0$ and $\eta > 0$ are fixed scalar hyperparameters.

Precisions are easier than variances. It's perfectly fine to work with this form of the prior, and it's easier to interpret this way. But it turns out that we can make the algebra a bit cleaner by working with the precisions $\omega = 1/\sigma^2$ and $\kappa = 1/\tau^2$ instead.

$$\begin{aligned} (\theta \mid \omega) &\sim N(\mu, (\omega\kappa)^{-1}) \\ \omega &\sim \text{Gamma}\left(\frac{d}{2}, \frac{\eta}{2}\right) \, . \end{aligned}$$

This means that the joint prior for (θ, ω) has the form

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

This is often called the *normal/gamma* prior for (θ, ω) with parameters (μ, κ, d, η), and it's equivalent to a normal/inverse-gamma prior for (θ, σ^2).

(A) By construction, we know that the marginal prior distribution $p(\theta)$ is a gamma mixture of normals. Show that this takes the form of a centered, scaled t distribution:

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

with center m, scale s, and degrees of freedom ν, where you fill in the blank for m, s^2, and ν in terms of the four parameters of the normal-gamma family.

We know that:

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

So,

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

We have shown that $p(\theta)$ takes the form of a centered, scaled t-distribution

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

where,

$$\begin{aligned}
m &= \mu \\
s^2 &= \frac{\eta}{\kappa d} \\
\nu &= d
\end{aligned}$$

(B) Assume the normal sampling model in Equation 1 and the normal-gamma prior in Equation 2. Calculate the joint posterior density $p(\theta, \omega \mid \mathbf{y})$, up to constant factors not depending on ω or θ. Show that this is also a normal/gamma prior in the same form as above:

$$p(\theta, \omega \mid \mathbf{y}) \propto \omega^{(d^\star+1)/2-1} \exp\left\{-\omega \cdot \frac{\kappa^\star(\theta-\mu^\star)^2}{2}\right\} \cdot \exp\left\{-\omega \cdot \frac{\eta^\star}{2}\right\} \tag{3}$$

We know that:

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

and

$$p(\mathbf{y} \mid \theta, \omega) \propto \omega^{\frac{n}{2}} exp\left\{-\frac{\omega}{2}\sum_{i=1}^{n}(y_i-\theta)^2\right\}.$$

Then,

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

So, we have calculated the joint density and shown that it takes the form,

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

where,

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

(C) From the joint posterior you just derived, what is the conditional posterior distribution $p(\theta \mid \mathbf{y}, \omega)$?

In the previous problem we showed that,

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

where,

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

We can rewrite the posterior as,

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

Since

$$p(\theta, \omega \mid \mathbf{y}) = p(\theta \mid \mathbf{y}, \omega) p(\omega \mid \mathbf{y})$$

we have that,

$$\begin{aligned}
p(\theta \mid \mathbf{y}, \omega) &\propto \omega^{\frac{1}{2}} \exp\left\{ -\omega \cdot \frac{\kappa^\star(\theta-\mu^\star)^2}{2} \right\} \\
p(\omega \mid \mathbf{y}) &\propto \omega^{(d^\star)/2-1} \exp\left\{ -\omega \cdot \frac{\eta^\star}{2} \right\}.
\end{aligned}$$

So,

$$p(\theta \mid \mathbf{y}, \omega) \sim N(\mu^\star, (\omega\kappa^\star)^{-1}).$$

(D) What is the marginal posterior distribution $p(\omega \mid \mathbf{y})$?

From the previous problem, we know that

$$p(\omega \mid \mathbf{y}) \propto \omega^{(d^\star)/2-1} \exp\left\{-\omega \cdot \frac{\eta^\star}{2}\right\}.$$

So,

$$p(\omega \mid \mathbf{y}) \sim Gamma(\frac{d^\star}{2}, \frac{\eta^\star}{2})$$

(E) Show that the marginal posterior $p(\theta \mid \mathbf{y})$ takes the form of a centered, scaled t distribution and express the parameters of this t distribution in terms of the four parameters of the normal-gamma posterior for (θ, ω).

In problem (B) we showed that,

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

Then,

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

Since a centered, scaled t distribution can be expressed as

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

we have shown that $p(\theta \mid \mathbf{y})$ takes the form of a centered, scaled t-distribution where,

$$\begin{aligned} m &= \mu^\star \\ s^2 &= \frac{\eta^\star}{\kappa^\star d^\star} \\ \nu &= d^\star \end{aligned}$$

(F) True or false: in the limit as the prior parameters κ, d, and η approach zero, the priors $p(\theta \mid \omega)$ and $p(\omega)$ are valid probability distributions.

We know that

$$\begin{aligned} (\theta \mid \omega) &\sim N(\mu, (\omega\kappa)^{-1}) \\ \omega &\sim \text{Gamma}\left(\frac{d}{2}, \frac{\eta}{2}\right). \end{aligned}$$

So,

$$\begin{aligned} p(\theta \mid \omega) &\propto (\omega\kappa)^{\frac{1}{2}} \exp\{\frac{\omega\kappa}{2}(\theta - \mu)^2\} \\ p(\omega) &\propto \omega^{\frac{d}{2}-1} e^{-\frac{\eta}{2}\omega}. \end{aligned}$$

As, κ, d, and η approach zero, this becomes,

$$\begin{aligned} p(\theta \mid \omega) &\propto 1 \\ p(\omega) &\propto \omega^{-1}. \end{aligned}$$

Neither of these are valid probability distributions as $\int_{-\infty}^{\infty} 1 d\theta = \infty$ and $\int_0^{\infty} \omega^{-1} d\omega = \infty$, therefore they are improper integrals. So, it is False that they are valid probability distributions.

(G) True or false: in the limit as the prior parameters κ, d, and η approach zero, the posteriors $p(\theta \mid \omega \mathbf{y})$ and $p(\omega \mid \mathbf{y})$ are valid probability distributions.

As κ, d, and η approach zero, :

$$\begin{aligned}
\mu^\star &= \frac{(\sum_{i=1}^n y_i - \kappa\mu)}{\kappa + n} \quad \longrightarrow \quad \frac{\sum_{i=1}^n y_i}{n} \\
\kappa^\star &= \kappa + n \quad \longrightarrow \quad n \\
d^\star &= n + d \quad \longrightarrow \quad n \\
\eta^\star &= \sum_{i=1}^n y_i^2 + \kappa\mu^2 + \eta - \frac{(\sum_{i=1}^n y_i - \kappa\mu)^2}{\kappa + n} \quad \longrightarrow \quad \sum_{i=1}^n y_i^2 - \frac{(\sum_{i=1}^n y_i)^2}{n}
\end{aligned}$$

We have already shown that

$$p(\theta \mid \mathbf{y}, \omega) \sim N(\mu^\star, (\omega\kappa^\star)^{-1})$$

and

$$p(\omega \mid \mathbf{y}) \sim Gamma(\frac{d^\star}{2}, \frac{\eta^\star}{2}).$$

So as κ, d, and η approach zero, these become:

$$p(\theta \mid \mathbf{y}, \omega) \sim N\left(\frac{\sum_{i=1}^n y_i}{n}, (n\omega)^{-1}\right)$$

and

$$p(\omega \mid \mathbf{y}) \sim Gamma\left(\frac{n}{2}, \frac{\sum_{i=1}^n y_i^2 - \frac{(\sum_{i=1}^n y_i)^2}{n}}{2}\right)$$

which are certainly valid distributions.

This means it is true, the posteriors are valid probability distributions. So, even though you start with improper priors you end up with valid posterior distributions that line up with frequentist notions.

(H) Your result in (E) implies that a Bayesian credible interval for θ takes the form

$$\theta \in m \pm t^\star \cdot s\,,$$

where m and s are the posterior center and scale parameters from (E), and $t^\star$ is the appropriate critical value of the t distribution for your coverage level and degrees of freedom (e.g. it would be 1.96 for a 95% interval under the normal distribution). True or false: In the limit as the prior parameters κ, d, and η approach zero, the Bayesian credible interval for θ becomes identical to the classical (frequentist) confidence interval for θ at the same confidence level.

$$\theta \in m \pm t^\star \cdot s$$
$$\theta \in \mu^\star \pm t^\star \cdot \sqrt{\frac{\eta^\star}{\kappa^\star d^\star}}$$

As κ, d, and η approach zero, this becomes,

$$\theta \in \frac{\sum_{i=1}^n y_i}{n} \pm t^\star \cdot \sqrt{\frac{\sum_{i=1}^n y_i^2 - \frac{(\sum_{i=1}^n y_i)^2}{n}}{n^2}}$$
$$\theta \in \bar{y} \pm t^\star \cdot \sqrt{\frac{\sum_{i=1}^n y_i^2 + \bar{y}^2 - \bar{y}^2 - \frac{(\sum_{i=1}^n y_i)^2}{n}}{n^2}}$$
$$\theta \in \bar{y} \pm t^\star \cdot \sqrt{\frac{\sum_{i=1}^n y_i^2 - 2(\bar{y}\sum_{i=1}^n y_i) + \bar{y}^2}{n^2}}$$
$$\theta \in \bar{y} \pm t^\star \cdot \sqrt{\frac{\sum_{i=1}^n (y_i - \bar{y})^2}{n^2}}$$
$$\theta \in \bar{y} \pm t^\star \cdot \frac{s}{\sqrt{n}}$$

Where s is the sample variance. Note that this is exactly the classical (frequentist) confidence interval for θ, so this is true.

2 The conjugate Gaussian linear model

Now consider the Gaussian linear model

$$(\mathbf{y} \mid \boldsymbol{\beta}, \omega) \sim \mathrm{N}(X\boldsymbol{\beta}, (\omega\Lambda)^{-1}) \, ,$$

where $\mathbf{y}$ is an n vector of responses, X is an $n \times p$ matrix of features, and $\omega = 1/\sigma^2$ is the error precision, and Λ is some known matrix. A typical setup would be $\Lambda = I$, the $n \times n$ identity matrix, so that the residuals of the model are i.i.d. normal with variance σ^2. But we'll consider other setups as well, so we'll leave a generic Λ matrix in the sampling model for now.

Note that when we write the model this way, we typically assume one of two things: either (1) that both the y variable and all the X variables have been centered to have mean zero, so that an intercept is unnecessary; or (2) that X has a vector of 1's as its first column, so that the first entry in β is actually the intercept.

We'll again work in terms of the precision $\omega = \sigma^2$, and consider a normal–gamma prior for β:

$$(\boldsymbol{\beta} \mid \omega) \sim N(m, (\omega K)^{-1}) \tag{4}$$

$$\omega \sim \mathrm{Gamma}(d/2, \eta/2) \, . \tag{5}$$

Here K is a $p \times p$ precision matrix in the multivariate normal prior for β, which we assume to be known.

The items below follow a parallel path to the derivations you did for the Gaussian location model—except for the multivariate case. Don't reinvent the wheel if you don't have to: you should be relying heavily on your previous results about the multivariate normal distribution.[1]

[1] That is, if you find yourself completing the square over and over again with matrices and vectors, you should stop and revisit your Exercises 1 solutions.

2.1 Basics

(A) Derive the conditional posterior $p(\boldsymbol{\beta} \mid \mathbf{y}, \omega)$.

We have,

$$
\begin{aligned}
p(\boldsymbol{\beta}, \omega \mid \mathbf{y}) &\propto p(\mathbf{y} \mid \boldsymbol{\beta}, \omega) p(\boldsymbol{\beta} \mid \omega) p(\omega) \\
&\propto |\omega\Lambda|^{\frac{1}{2}} \exp\left\{-\frac{1}{2}(\mathbf{y} - X\boldsymbol{\beta})^T \omega\Lambda(\mathbf{y} - X\boldsymbol{\beta})\right\} |\omega K|^{\frac{1}{2}} \exp\left\{-\frac{1}{2}(\boldsymbol{\beta} - \mathbf{m})^T \omega K(\boldsymbol{\beta} - \mathbf{m})\right\} \\
&\qquad \omega^{\left(\frac{d}{2}-1\right)} \exp\left\{-\frac{\eta}{2}\omega\right\} \\
&\propto \omega^{\frac{n}{2}} \exp\left\{-\frac{1}{2}(\mathbf{y} - X\boldsymbol{\beta})^T \omega\Lambda(\mathbf{y} - X\boldsymbol{\beta})\right\} \omega^{\frac{p}{2}} \exp\left\{-\frac{1}{2}(\boldsymbol{\beta} - \mathbf{m})^T \omega K(\boldsymbol{\beta} - \mathbf{m})\right\} \\
&\qquad \omega^{\left(\frac{d}{2}-1\right)} \exp\left\{-\frac{\eta}{2}\omega\right\} \\
&= \exp\left\{-\frac{\omega}{2}\Big(\, (X\boldsymbol{\beta} - \mathbf{y})^T \Lambda (X\boldsymbol{\beta} - \mathbf{y}) + (\boldsymbol{\beta} - \mathbf{m})^T K(\boldsymbol{\beta} - \mathbf{m})\,)\right\} \\
&\qquad \omega^{\left(\frac{n+p+d}{2}-1\right)} \exp\left\{-\frac{\eta}{2}\omega\right\} \\
&= \exp\left\{-\frac{\omega}{2}\left(\boldsymbol{\beta}^T (X^T\Lambda X)\boldsymbol{\beta} - 2\mathbf{y}^T\Lambda X\boldsymbol{\beta} + \mathbf{y}^T\Lambda\mathbf{y} + \boldsymbol{\beta}^T K\boldsymbol{\beta} - 2\mathbf{m}^T K\boldsymbol{\beta} + \mathbf{m}^T K\mathbf{m}\right)\right\} \\
&\qquad \omega^{\left(\frac{n+p+d}{2}-1\right)} \exp\left\{-\frac{\eta}{2}\omega\right\} \\
&= \exp\left\{-\frac{\omega}{2}\left(\boldsymbol{\beta}^T (X^T\Lambda X + K)\boldsymbol{\beta} - 2(\mathbf{y}^T\Lambda X + \mathbf{m}^T K)\boldsymbol{\beta}\right)\right\} \\
&\qquad \omega^{\left(\frac{n+p+d}{2}-1\right)} \exp\left\{-\frac{\omega}{2}\left(\eta + \mathbf{y}^T\Lambda\mathbf{y} + \mathbf{m}^T K\mathbf{m}\right)\right\} \\
&= \omega^{\frac{p}{2}} \exp\{-\frac{\omega}{2}\left(\boldsymbol{\beta} - \left((X^T\Lambda X + K)^{-1}\right)^T (X^T\Lambda\mathbf{y} + K^T\mathbf{m})\right)^T (X^T\Lambda X + K) \\
&\qquad \left(\boldsymbol{\beta} - \left((X^T\Lambda X + K)^{-1}\right)^T (X^T\Lambda\mathbf{y} + K^T\mathbf{m})\right)\} \omega^{\left(\frac{n+d}{2}-1\right)} \\
&\qquad \exp\left\{-\frac{\omega}{2}\left(\eta + \mathbf{y}^T\Lambda\mathbf{y} + \mathbf{m}^T K\mathbf{m} - (\mathbf{y}^T\Lambda X + \mathbf{m}^T K)^T (X^T\Lambda X + K)^{-1}(\mathbf{y}^T\Lambda X + \mathbf{m}^T K)\right)\right\} \\
&= \omega^{\frac{p}{2}} \exp\left\{-\frac{1}{2}\left(\boldsymbol{\beta} - \mu^\star\right)^T \omega\Lambda^\star \left(\boldsymbol{\beta} - \mu^\star\right)\right\} \omega^{\left(\frac{n+d}{2}-1\right)} \exp\left\{-\frac{\omega}{2}\left(\eta^\star\right)\right\}
\end{aligned}
$$

where:

$\mu^\star = \left((X^T\Lambda X + K)^{-1}\right)^T (X^T\Lambda\mathbf{y} + K^T\mathbf{m})$

$\Lambda^\star = X^T\Lambda X + K$

$\eta^\star = \eta + \mathbf{y}^T\Lambda\mathbf{y} + \mathbf{m}^T K\mathbf{m} - (\mathbf{y}^T\Lambda X + \mathbf{m}^T K)^T (X^T\Lambda X + K)^{-1}(\mathbf{y}^T\Lambda X + \mathbf{m}^T K).$

Since $p(\boldsymbol{\beta}, \omega \mid y) = p(\boldsymbol{\beta} \mid y, \omega) p(\omega \mid y)$, we can see that:

$$p(\boldsymbol{\beta} \mid \mathbf{y}, \omega) \propto \omega^{\frac{p}{2}} \exp\left\{-\frac{1}{2}\left(\boldsymbol{\beta} - \mu^\star\right)^T \omega\Lambda^\star \left(\boldsymbol{\beta} - \mu^\star\right)\right\}$$

and

$$p(\omega \mid \mathbf{y}) \propto \omega^{\left(\frac{n+d}{2}-1\right)} \exp\left\{-\frac{\omega}{2}\left(\eta^\star\right)\right\}$$

So, $\boldsymbol{\beta} \mid \mathbf{y}, \omega \sim N(\mu^\star, (\omega\Lambda^\star)^{-1})$.

(B) Derive the marginal posterior $p(\omega \mid \mathbf{y})$.

From part (A) we know that :

$$p(\omega \mid \mathbf{y}) \propto \omega^{\left(\frac{n+d}{2}-1\right)} \exp\left\{-\frac{\omega}{2}\left(\eta^\star\right)\right\}$$

So, $\omega \mid \mathbf{y} \sim Gamma(\frac{n+d}{2}, \frac{\eta^\star}{2})$.

(C) Putting these together, derive the marginal posterior $p(\boldsymbol{\beta} \mid \mathbf{y})$.

From part (A),

$$p(\boldsymbol{\beta}, \omega \mid \mathbf{y}) \propto \omega^{\frac{p}{2}} \exp\left\{-\frac{1}{2}\left(\boldsymbol{\beta}-\mu^\star\right)^T \omega\Lambda^\star \left(\boldsymbol{\beta}-\mu^\star\right)\right\} \omega^{\left(\frac{n+d}{2}-1\right)} \exp\left\{-\frac{\omega}{2}\left(\eta^\star\right)\right\}.$$

So,

$$\begin{aligned}
p(\boldsymbol{\beta} \mid \mathbf{y}) &= \int_\omega p(\boldsymbol{\beta}, \omega \mid \mathbf{y}) d\omega \\
&\propto \int_\omega \omega^{\frac{n+p+d}{2}-1} \exp\left\{-\frac{\omega}{2}\left(\left(\boldsymbol{\beta}-\mu^\star\right)^T \Lambda^\star \left(\boldsymbol{\beta}-\mu^\star\right)+\eta^\star\right)\right\} d\omega \qquad (6) \\
&\propto \left(-\frac{1}{2}\left(\left(\boldsymbol{\beta}-\mu^\star\right)^T \Lambda^\star \left(\boldsymbol{\beta}-\mu^\star\right)+\eta^\star\right)\right)^{-\frac{n+p+d}{2}} \\
&\propto \left(1+\frac{1}{n+d}\left(\boldsymbol{\beta}-\mu^\star\right)^T \frac{n+d}{\eta^\star}\Lambda^\star \left(\boldsymbol{\beta}-\mu^\star\right)\right)^{-\frac{(n+d)+p}{2}} \\
&= \left(1+\frac{1}{\nu^\star}\left(\boldsymbol{\beta}-\mu^\star\right)^T \Sigma^\star \left(\boldsymbol{\beta}-\mu^\star\right)\right)^{-\frac{\nu^\star+p}{2}}.
\end{aligned}$$

This is a multivariate t distribution with $\nu^\star$ degrees of freedom, location parameter $\mu^\star$, and covariance $\Sigma^\star$, where $\mu^\star$ is defined in part (A) and,

$$\begin{aligned}
\nu^\star &= n+d \\
\Sigma^\star &= \frac{n+d}{\eta^\star}\Lambda^\star
\end{aligned}$$

Note: In line 6 we recognize the kernal of a Gamma distribution, and use that information to solve the integral.

(D) Take a look at the data in "gdpgrowth.csv" from the class website, which has macroeconomic variables for several dozen countries. In particular, consider a linear model (with intercept) for a country's GDP growth rate (GR6096) versus its level of defense spending as a fraction of its GDP (DEF60).

Fit the Bayesian linear model to this data set, choosing $\Lambda = I$ and something diagonal and pretty vague for the prior precision matrix $K = \text{diag}(\kappa_1, \kappa_2)$. Inspect the fitted line (graphically). Are you happy with the fit? Why or why not?

We have already shown that:

$$\boldsymbol{\beta} \mid \mathbf{y}, \omega \sim N(\mu^\star, (\omega\Lambda^\star)^{-1})$$

$$\omega \mid \mathbf{y} \sim Gamma(\frac{n+d}{2}, \frac{\eta^\star}{2})$$

where:

$$\mu^\star = \left((X^T\Lambda X + K)^{-1}\right)^T (X^T\Lambda\mathbf{y} + K^T m)$$
$$\Lambda^\star = X^T\Lambda X + K$$
$$\eta^\star = \eta + \mathbf{y}^T\Lambda\mathbf{y} + \mathbf{m}^T K\mathbf{m} - (\mathbf{y}^T\Lambda X + \mathbf{m}^T K)^T (X^T\Lambda X + K)^{-1}(\mathbf{y}^T\Lambda X + \mathbf{m}^T K).$$

I chose $\kappa_1 = \kappa_2 = 0.1$, $\mathbf{m} = (0,0)^T$, and $\eta = d = 1$. Below is a graph of the fit obtained from 5,000 iterations, throwing out the first 1,000 as burn-in. I examined trace plots to see if the chain was mixing which are included in the code.

Overall this does not appear to be a great fit, mostly because I do not believe this data to be linear.

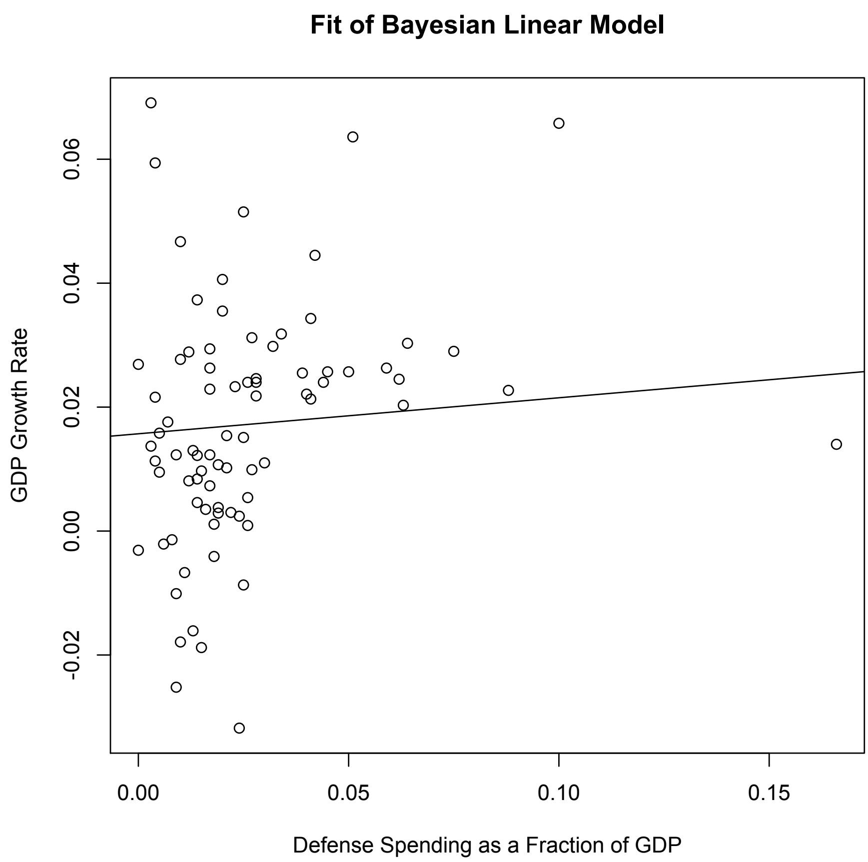

Figure 1: Plot of the fit of the Bayesian linear model.

2.2 A heavy-tailed error model

Now it's time for your first "real" use of the hierarchical modeling formalism to do something cool. Here's the full model you'll be working with:

$$\begin{aligned}
(\mathbf{y} \mid \boldsymbol{\beta}, \omega, \Lambda) &\sim N(X\boldsymbol{\beta}, (\omega\Lambda)^{-1}) \\
\Lambda &= \mathrm{diag}(\lambda_1, \ldots, \lambda_n) \\
\lambda_i &\overset{iid}{\sim} \mathrm{Gamma}(h/2, h/2) \\
(\boldsymbol{\beta} \mid \omega) &\sim N(\mathbf{m}, (\omega K)^{-1}) \\
\omega &\sim \mathrm{Gamma}(d/2, \eta/2)\,.
\end{aligned}$$

where h is a fixed hyperparameter.

(A) Under this model, what is the implied conditional distribution $p(y_i \mid$

$X, \boldsymbol{\beta}, \omega)$? Notice that λ_i has been marginalized out. This should look familiar.

Let $\mathbf{x_i}$ be the vector of covariates for observation i, then:

$$y_i \mid \boldsymbol{\beta}, \omega, \lambda_i \sim N(\mathbf{x_i}^T \boldsymbol{\beta}, (\omega \lambda_i)^{-1}).$$

So,

$$\begin{aligned}
p(y_i \mid \boldsymbol{\beta}, \omega) &\propto \int_{\lambda_i} p(y_i \mid \boldsymbol{\beta}, \omega, \lambda_i) p(\lambda_i) d\lambda_i \\
&\propto \int_{\lambda_i} (\omega\lambda_i)^{\frac{1}{2}} e^{-\frac{1}{2}\omega\lambda_i (y_i - \mathbf{x_i}^T\boldsymbol{\beta})^2} \lambda_i^{\frac{h}{2}-1} e^{-\lambda_i \frac{h}{2}} d\lambda_i \\
&\propto \int_{\lambda_i} \lambda_i^{\frac{h+1}{2}-1} e^{-\lambda_i \frac{1}{2}(\omega(y_i - \mathbf{x_i}^T\boldsymbol{\beta})^2 + h)} d\lambda_i \qquad (7) \\
&\propto \left(\frac{1}{2} (\omega(y_i - \mathbf{x_i}^T\boldsymbol{\beta})^2 + h) \right)^{-\frac{h+1}{2}} \\
&\propto \left(1 + \frac{1}{h} \omega (y_i - \mathbf{x_i}^T\boldsymbol{\beta})^2 \right)^{-\frac{h+1}{2}}
\end{aligned}$$

This is a centered, scaled t-distribution with scale parameter $\sqrt{\omega^{-1}}$, location parameter $\mathbf{x_i}^T\boldsymbol{\beta}$, and degrees of freedom h. Note: In line 7 we recognize the kernal of a Gamma distribution, and use that information to solve the integral.

(B) What is the conditional posterior distribution $p(\lambda_i \mid \mathbf{y}, \boldsymbol{\beta}, \omega)$?

We have:

$$\begin{aligned}
p(\lambda_i \mid \mathbf{y}, \boldsymbol{\beta}, \omega) &\propto p(y \mid \boldsymbol{\beta}, \omega, \lambda_i) p(\lambda_i) \\
&\propto (\omega\lambda_i)^{\frac{1}{2}} e^{-\frac{1}{2}\omega\lambda_i (y_i - \mathbf{x_i}^T\boldsymbol{\beta})^2} \lambda_i^{\frac{h}{2}-1} e^{-\lambda_i \frac{h}{2}} \\
&\propto \lambda_i^{\frac{h+1}{2}-1} e^{-\lambda_i \frac{1}{2}(\omega(y_i - \mathbf{x_i}^T\boldsymbol{\beta})^2 + h)}
\end{aligned}$$

We recognize this as the kernal of a Gamma distribution, so

$$\lambda_i \mid \mathbf{y}, \boldsymbol{\beta}, \omega \sim Gamma \left(\frac{h+1}{2}, \frac{1}{2}(\omega(y_i - \mathbf{x_i}^T\boldsymbol{\beta})^2 + h) \right).$$

(C) Code up a Gibbs sampler that repeatedly cycles through sampling the following three sets of conditional distributions, use your Gibbs sampler to fit this model to the GDP growth rate data for an appropriate choice of h.

- $p(\boldsymbol{\beta} \mid \mathbf{y}, \omega, \Lambda) \sim N(\mu^\star, (\omega\Lambda^\star)^{-1})$
- $p(\omega \mid \mathbf{y}, \Lambda) \sim Gamma(\frac{n+d}{2}, \frac{\eta^\star}{2})$

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

where:

$\mu^\star = \left((X^T\Lambda X + K)^{-1}\right)^T (X^T\Lambda\mathbf{y} + K^T\mathbf{m})$
$\Lambda^\star = X^T\Lambda X + K$
$\eta^\star = \eta + \mathbf{y}^T\Lambda\mathbf{y} + \mathbf{m}^T K\mathbf{m} - (\mathbf{y}^T\Lambda X + \mathbf{m}^T K)^T(X^T\Lambda X + K)^{-1}(\mathbf{y}^T\Lambda X + \mathbf{m}^T K).$

I chose $\kappa_1 = \kappa_2 = 0.1$, $\mathbf{m} = (0,0)^T$, and $\eta = d = h = 1$. Below is a graph of the fit obtained from 5,000 iterations, throwing out the first 1,000 as burn-in. Since this model includes heavy-tailed error, the line is not drawn quite as far towards the potential outlier on the right-hand side, although the fit is still not satisfactory.

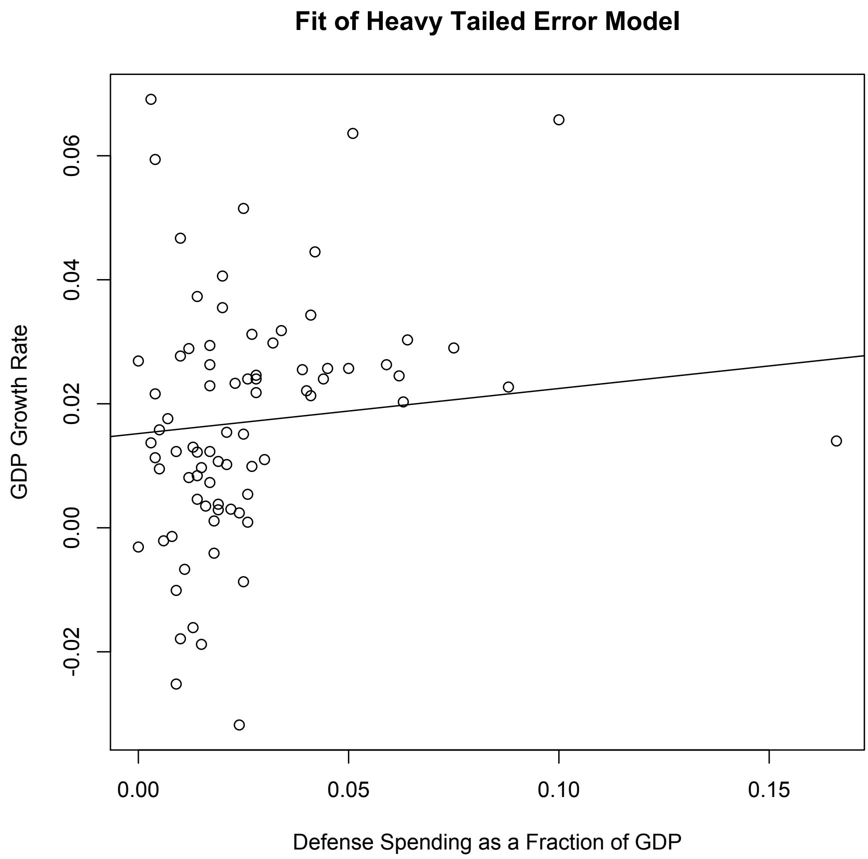

Figure 2: Plot of the fit of the Bayesian heavy-tailed error model.