SDS 383D: Exercises 2 – Bayes and the Gaussian linear model

February 13, 2017

Professor Scott

Spencer Woody

Problem 1

The conjugate Gaussian linear model

We have a model where

$$(y_i|\theta, \sigma^2) \overset{\text{iid}}{\sim} \mathcal{N}(\theta, \omega^{-1}),\ i = 1, \dots n,$$
$$(\theta|\sigma) \sim \mathcal{N}(\mu, \tau^2\sigma^2),$$
$$\sigma^2 \sim \text{IG}(\frac{d}{2}, \frac{\eta}{2}).$$

Define $\omega = \frac{1}{\sigma^2}$ and $\kappa = \frac{1}{\tau^2}$, so now

$$(\theta|\omega) \sim \mathcal{N}(\mu, (\omega\kappa)^{-1}),$$

and

$$\omega \sim \text{Gamma}\left(\frac{d}{2}, \frac{\eta}{2}\right).$$

(A) Taking advantage that (1) below is the integral of the kernel of a gamma distribution, the marginal prior distribution of θ is

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

which is the kernel of the Student's t-distribution with degrees of freedom d, center parameter μ, and scale paramater $\sqrt{\frac{\eta}{d\kappa}}$.

(B) In this exercise, recall that the likelihood may be written as

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

where $S_y = \sum_{i=1}^n (y_i - \bar{y})^2 = \sum_{i=1}^n y_i^2 - n\bar{y}^2$ is the sum of squares. Here we calculate the joint posterior density,

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

This expression looks very much like the joint prior, but now with

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

We can also rewrite the joint posterior as

$$
\begin{aligned}
p(\theta,\omega|y) &\propto \omega^{(d+n+1)/2-1}\exp\left[-\frac{1}{2}\omega(n+\kappa)\left(\theta-\frac{n\bar{y}+\mu\kappa}{n+\kappa}\right)^2\right]\exp\left[-\frac{1}{2}\left(\eta+\sum_{i=1}^n y_i^2+\kappa\mu^2-\frac{(n\bar{y}+\mu\kappa)^2}{n+\kappa}\right)\omega\right] \\
&= \underbrace{\omega^{1/2}\exp\left[-\frac{1}{2}\omega(n+\kappa)\left(\theta-\frac{n\bar{y}+\mu\kappa}{n+\kappa}\right)^2\right]}_{p(\theta|\omega,y)\sim\mathcal{N}\left(\frac{n\bar{y}+\mu\kappa}{n+\kappa},[(n+\kappa)\omega]^{-1}\right)}\underbrace{\omega^{(d+n)/2-1}\exp\left[-\frac{1}{2}\left(\eta+\sum_{i=1}^n y_i^2+\kappa\mu^2-\frac{(n\bar{y}+\mu\kappa)^2}{n+\kappa}\right)\omega\right]}_{p(\omega|y)\sim\text{Gamma}\left(\frac{d+n}{2},\frac{1}{2}\left[\eta+\sum_{i=1}^n y_i^2+\kappa\mu^2-\frac{(n\bar{y}+\mu\kappa)^2}{n+\kappa}\right]\right)}.
\end{aligned}
$$

(C) It is easy to read off the the conditional posterior of θ,

$$
p(\theta|\omega,y)\sim\mathcal{N}\left(\frac{n\bar{y}+\mu\kappa}{n+\kappa},[(n+\kappa)\omega]^{-1}\right).
$$

(D) Getting the marginal posterior is not too much more difficult. We have separated the joint posterior into two kernels, and once we integrate out the normal kernel to get

$$p(\omega|y) \sim \text{Gamma}\left(\frac{d+n}{2}, \frac{1}{2}\left[\eta + \sum_{i=1}^{n} y_i^2 + \kappa\mu^2 - \frac{(n\bar{y}+\mu\kappa)^2}{n+\kappa}\right]\right).$$

(E) The marginal posterior distribution $p(\theta|y)$ will follow a Student's t-distribution, following a similar argument used in (A), but now with updated parameters. The degrees of freedom is

$$\nu' = d + n,$$

the center parameter is

$$m' = \frac{n\bar{y} + \mu\kappa}{n+\kappa},$$

and the scale parameter is

$$s' = \sqrt{\frac{\eta + \sum_{i=1}^{n} y_i^2 + \kappa\mu^2 - \frac{(n\bar{y}+\mu\kappa)^2}{n+\kappa}}{(d+n)(\kappa+n)}}.$$

(F) Suppose $\kappa \to 0$, $d \to 0$, and $\eta \to 0$. The priors of $\theta|\omega$ and ω become

$$\begin{aligned} p(\theta|\omega) &\to \mathcal{N}(\mu, 0^{-1}) \\ p(\omega) &\to \text{Gamma}\,(0,0)\,, \end{aligned}$$

neither of which is a proper PDF.

(G) Suppose $\kappa \to 0$, $d \to 0$, and $\eta \to 0$. The marginal posteriors $p(\theta|y)$ and $p(\omega|y)$ become

$$\begin{aligned} p(\theta|\omega, y) &\to \mathcal{N}\left(\bar{y}, (n\omega)^{-1}\right) \\ p(\omega|y) &\to \text{Gamma}\left(\frac{n}{2}, \frac{1}{2}\left[\sum_{i=1}^{n} y_i^2 - n\bar{y}^2\right]\right) \\ &= \text{Gamma}\left(\frac{n}{2}, \frac{S_y}{2}\right) \end{aligned}$$

(H) A Bayesian credible interval for θ, having observed y, is

$$\theta \in m' \pm t^\star s',$$

for some critical value $t^\star$ from the Student's t-distribution with ν' degrees of freedom. Suppose $\kappa \to 0$, $d \to 0$, and $\eta \to 0$. The parameters of this Student's t-distribution become

$$\begin{aligned} \nu' &\to n \\ m' &\to \bar{y} \\ s' &\to \sqrt{\frac{S_y}{n^2}}, \end{aligned}$$

so the credible interval becomes identical to the frequentist confidence interval at the same confidence level.

Problem 2

The conjugate Gaussian linear model

(A) We have the Gaussian linear model,

$$(y|\beta,\omega) \sim \mathcal{N}_n(X\beta, (\omega\Lambda)^{-1})$$

with priors

$$\begin{aligned}(\beta|\omega) &\sim \mathcal{N}_p(m, (\omega K)^{-1}) \\ \omega &\sim \text{Gamma}\left(\frac{d}{2}, \frac{\eta}{2}\right).\end{aligned}$$

Having observed y, the joint posterior for β and ω is

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

Now, let

$$\begin{aligned}
A &= X^T\Lambda X+K \\
b^T &= y^T\Lambda X+m^TK \\
\Rightarrow b &= X^T\Lambda y+Km \\
c &= y^T\Lambda y+m^TKm+\eta,
\end{aligned}$$

so now the expression in (i) becomes, once the square is completed,

$$\begin{aligned}
\beta^TA\beta-2b^T\beta+c &= \beta^TA\beta-2b^T\beta+b^TA^{-1}b-b^TA^{-1}b+c \\
&= (\beta-A^{-1}b)^TA(\beta-A^{-1}b)-b^TA^{-1}b+c,
\end{aligned}$$

Now let

$$\begin{aligned}
m^\star &= A^{-1}b = (X^T\Lambda X+K)^{-1}(X^T\Lambda y+Km) \\
K^\star &= A = X^T\Lambda X+K,
\end{aligned}$$

and we can also simplify the term

$$\begin{aligned}
b^TA^{-1}b &= b^TIA^{-1} \\
&= b^TA^{-1}AA^{-1}b \\
&= m^{\star T}K^\star m^\star,
\end{aligned}$$

and finally let

$$\begin{aligned}\eta^\star &= c - m^{\star T} K^\star m^\star \\ &= \eta + y^T \Lambda y + m^T K m - m^{\star T} K^\star m^\star\end{aligned}$$

We can at last express (i) as

$$(\beta - m^\star)^T K^\star (\beta - m^\star) + \eta^\star.$$

Now the joint posterior distribution may be written as

$$\begin{aligned}p(\beta, \omega | y) &\propto \omega^{(d+p+n)/2-1} \exp\left(-\frac{1}{2}\omega\left[(\beta - m^\star)^T K^\star (\beta - m^\star) + \eta^\star\right]\right) \\ &= \underbrace{\omega^{p/2} \exp\left[-\frac{1}{2}(\beta - m^\star)^T \omega K^\star (\beta - m^\star)\right]}_{p(\beta|y,\omega) \sim \mathcal{N}(m^*, (\omega K^\star)^{-1})} \underbrace{\omega^{d^\star/2-1} \exp\left[-\frac{1}{2}\eta^\star \omega\right]}_{p(\omega|y) \sim \text{Gamma}\left(\frac{d^\star}{2}, \frac{\eta^\star}{2}\right)},\end{aligned}$$

with

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

(B) $p(\omega|y) \sim \text{Gamma}\left(\frac{d^\star}{2}, \frac{\eta^\star}{2}\right)$

(C) The marginal posterior for β may be found with

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

which we may recognize as the p-dimensional Student's t-distribution with $d+n$ degrees of freedom, mean vector $m^\star$, and covariance matrix $\frac{d+n}{\eta^\star}K^\star$.

A heavy-tailed model

Now we have a model with

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

(A) The conditional distribution $p(y_i|X,\beta,\omega)$, once the λ_i has been marginalized out, is

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

which follows the Student's t-distribution with h degrees of freedom, center parameter x_i^T, and scale parameter $\omega^{-1/2}$.

(B) The conditional posterior distribution of each λ_i is

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

(C) To run a Gibbs sampler with the heavy-tails model, we have the three conditional distributions

$$p(\beta|y,\omega) \sim \mathcal{N}(m^*, (\omega K^\star)^{-1}) \tag{1}$$

$$p(\omega|y) \sim \text{Gamma}\left(\frac{d^\star}{2}, \frac{\eta^\star}{2}\right) \tag{2}$$

$$p(\lambda_i|y,\beta,\omega) \sim \text{Gamma}\left(\frac{h+1}{2}, \frac{h + \omega(y_i - x_i^T\beta)^2}{2}\right). \tag{3}$$

The figure below shows the comparison between frequentist, Bayesian, and Bayesian heavy-tail models for the GDP growth data.

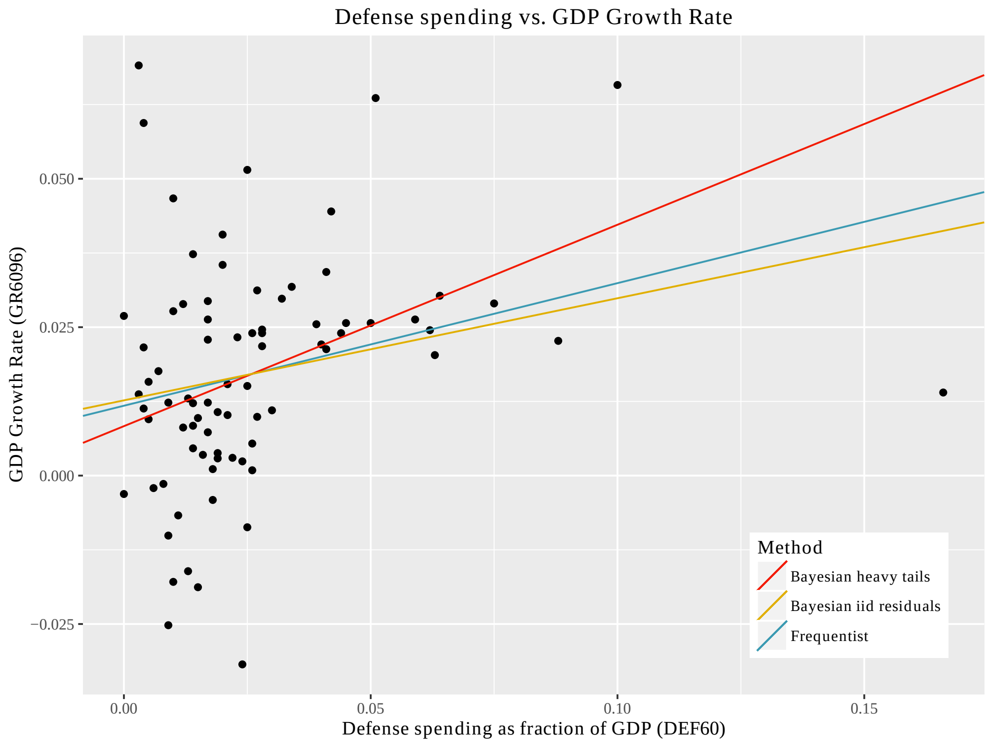

Figure 1: Comparison of three methods

R code for `myfuns02.R`

```
###########################################################
######### Created by Spencer Woody on 04 Feb 2017 #########
###########################################################


my.lm <- function(X, y) {
    #  Custom function for linear regression
    #
    #  Note: this function assumes that X already has an intercept term
    # (or doesn't, if we want to force OLS through the origin)
    #
    # INPUTS:
    # X is the design matrix
    # y is the response vector
    #
    # OUTPUTS
    # a list of...
    #  Beta.hat is a vector of estimates of the coefficients
    #  Beta.SE is a vector of the standard errors of the coefficients
    #  Beta.t is a vector of t-scores of the coefficients
    #  Beta.p is the p-value for each coefficient
    #  RSS is the residual sum of squares
    #  Var.hat is the estimated variance of homoscedastic residuals
    #  R.sq is the R-squared value
    #  R.sqadj is the adjusted R-squared value
```

```
    #

    N <- nrow(X)
    p <- ncol(X)

    XtX <- crossprod(X)

    # Calculate beta.hat
    beta.hat <- solve(XtX, crossprod(X, y))

    # Calculate predicted values and residuals
    y.hat <- crossprod(t(X), beta.hat)
    res <- y - y.hat

    rss <- sum(res^2)

    # Calculate \hat{sigma^2}
    var.hat <- rss / (N - p)

    # Calculate covariance matrix of beta and SE's of beta
    var.beta <- var.hat * solve(XtX)
    beta.SE <- diag(var.beta) ^ 0.5

    # Calculate t-score of each beta
    beta.t <- beta.hat / beta.SE

    # Calculate p-values for coefficients
    beta.p <- 2 * (1 - pt(abs(beta.t), N - p))

    # Calculate r-squared and adjusted r-squared
    r.sq <- 1 - rss / sum((y - mean(y))^2)
    r.sqadj <- r.sq - (1 - r.sq) * (p - 1) / (N - p - 2)

    # Create a list of calculated values, return it back
    mylist <- list(Beta.hat = beta.hat, Beta.SE = beta.SE,
                   Beta.t = beta.t, Beta.p = beta.p, RSS = rss, Var.hat = var.hat,
                   R.sq = r.sq, R.sqadj = r.sqadj, Res = res)
    return(mylist)
}


my.mvn <- function(n, mu, Sigma) {
    #  Simulate n draws from MVN(mu, Sigma)
    #
    #  Note: this function assumes that X already has an intercept term
    # (or doesn't, if we want to force OLS through the origin)
    #
    # INPUTS:
    # n is the number of draws
    # mu is the mean vector
    # Sigma is the covariance matrix
    #
    # OUTPUT:
```

```
    # x is matrix of n draws from MVN(mu, Sigma) [with n rows, p columns]
    #

    # dimension of MVN
    p <- length(mu)

    # Check if inputs are valid (dimensions match, Sigma is square and p.s.d.)
    # if ( (ncol(Sigma) != p) | (nrow(Sigma) != p) | (max(eigen(Sigma)$values) <= 0) ) {
    #   return("Try again...")
    # }
    #
    # Generate n*p univariate standard normal variables
    z     <- matrix(rnorm(n*p), nrow = p)

    # Create a matrix containing copies of mu
    mumat <- matrix(rep(mu, n), nrow = p)

    # Decompose Sigma into Sigma = L %*% Lt
    Lt <- chol(Sigma)

    # Generate sample with affine transformation of z
    x <- crossprod(Lt, z) + mumat

    return(t(x))
}
```

R code for exercises02.R

```
###########################################################
######### Created by Spencer Woody on 04 Feb 2017 #########
###########################################################

library(ggplot2)
library(mvtnorm)
library(wesanderson)
library(extrafont)

source("myfuns02.R")

# Prep color pallette
pal <- wes_palette("Zissou", 5)
col1 <- pal[1]
col2 <- pal[4]
col3 <- pal[5]

# ------------------------------------------------------------------------
# Read in the data -------------------------------------------------------
# ------------------------------------------------------------------------

GDP <- read.csv("gdpgrowth.csv", header = T)

y <- matrix(GDP$GR6096, ncol = 1)
X <- matrix(GDP$DEF60, ncol = 1)

n <- length(y)

X <- cbind(rep(1, n), X)
colnames(X) <- c("int", "DEF60")

# ------------------------------------------------------------------------
# Specify prior model ----------------------------------------------------
# ------------------------------------------------------------------------

freq.lm <- my.lm(X, y)
freq.beta <- freq.lm$Beta.hat

# ------------------------------------------------------------------------
# Specify prior model ----------------------------------------------------
# ------------------------------------------------------------------------

# Hyperparameters: K = diag(k1, k2), Lambda, d, eta, m

k1 <- 0.01
k2 <- 0.01

m <- matrix(c(0, 0), ncol = 1)
K <- diag(c(k1, k2))

Lambda <- diag(n)
```

```r
d <- 0.01
eta <- 0.01

# For heavy tails only
h <- 0.000001

# ----------------------------------------------------------------------------
# Obtain posterior model (no heavy tails) ------------------------------------
# ----------------------------------------------------------------------------

XtLambda <- crossprod(X, Lambda)

K.star   <- crossprod(t(XtLambda), X) + K
m.star   <- solve(K.star, crossprod(t(XtLambda), y) + crossprod(K, m))
eta.star <- eta + crossprod(Lambda %*% y, y) +
            crossprod(K %*% m, m) - crossprod(K.star %*% m.star, m.star)
d.star   <- d + n

m.star

# Prep for Gibbs sampler for model with heavy tails

nruns <- 1e4
burn  <- 2e3

beta.post <- matrix(nrow = nruns, ncol = 2)
omega.post <- rep(NA, nruns)
lambda.post <- matrix(nrow = nruns, ncol = n)

beta.post[1, ] <- c(0, 0)
omega.post[1] <- 1
lambda.post[1, ] <- rep(1, n)

# Gibbs sampler
for (i in 2:nruns) {
    # Create lambda matrix
    Lambda.i <- diag(lambda.post[i - 1, ])

    # Precache this crossproduct
    XtLambda.i <- crossprod(X, Lambda.i)

    # Update parameters of posterior conditionals
    K.star.i   <- crossprod(t(XtLambda.i), X) + K
    m.star.i   <- solve(K.star.i, crossprod(t(XtLambda.i), y) + crossprod(K, m))
    eta.star.i <- eta + crossprod(Lambda.i %*% y, y) +
            crossprod(K %*% m, m) - crossprod(K.star.i %*% m.star.i, m.star.i)

    # Draw from conditional posterior
    omega.post[i] <- rgamma(1, d.star / 2, eta.star.i / 2)
    beta.post[i, ] <- rmvnorm(1, m.star.i,
                              sigma = solve(omega.post[i] * K.star.i))
    lambda.post[i, ] <- rgamma(n, rep(h/2 + 1/2, n),
        h/2 + omega.post[i] * ( y- X %*% beta.post[i, ] )^2 / 2)
```

```
}

# Burn—out
beta.post <- beta.post[-(1:burn), ]
omega.post <- omega.post[-(1:burn)]
lambda.post <- lambda.post[-(1:burn), ]

# Mean posterior for betas
m.star2 <- colMeans(beta.post)

# Plot Frequentist, Bayesian, and Bayesian heavy—tail linear models
r <- ggplot(GDP, aes(DEF60, GR6096)) + geom_point() +
     geom_abline(mapping = aes(colour = "Frequentist",
     intercept = freq.beta[1],
     slope = freq.beta[2]),
     show.legend = T) +
     geom_abline(mapping = aes(colour = "Bayesian heavy tails",
     intercept = m.star2[1],
     slope = m.star2[2]),
     show.legend = T) +
     geom_abline(mapping = aes(colour = "Bayesian iid residuals",
     intercept = m.star[1],
     slope = m.star[2]),
     show.legend = T) +
     xlab("Defense spending as fraction of GDP (DEF60)") +
     ylab("GDP Growth Rate (GR6096)") +
     ggtitle("Defense spending vs. GDP Growth Rate") +
     scale_color_manual(name = "Method",
     values = c("Frequentist" = col1,
     "Bayesian iid residuals" = col2, "Bayesian heavy tails" = col3)) +
     theme(legend.position = c(0.85, 0.15),
     text = element_text(family="Palatino"))
r

# Save this plotr
pdf("img/compplot.pdf", width = 8, height = 6)
r
dev.off()

# Plot of reciprocals of lambdas
plot(sort(1 / colMeans(lambda.post)))

plot(sort(colMeans(lambda.post)))

# Plot gibbs sampler results
plot(omega.post,
    type = "l")

q <- qplot(1:length(beta.post[, 1]), beta.post[, 1], geom = "blank") +
     geom_line() +
     ggtitle("Traceplot of Gibbs sampler for intercept term") +
     xlab("iteration") +
     ylab("Intercept") +
```

```
    theme(text = element_text(family="Palatino"))

w <- qplot(1:length(beta.post[, 2]), beta.post[, 1], geom = "blank") +
    geom_line() +
    ggtitle("Traceplot of Gibbs sampler for DEF60 term") +
    xlab("iteration") +
    ylab("Intercept") +
    theme(text = element_text(family="Palatino"))
```