269
Homework 2 – Solutions

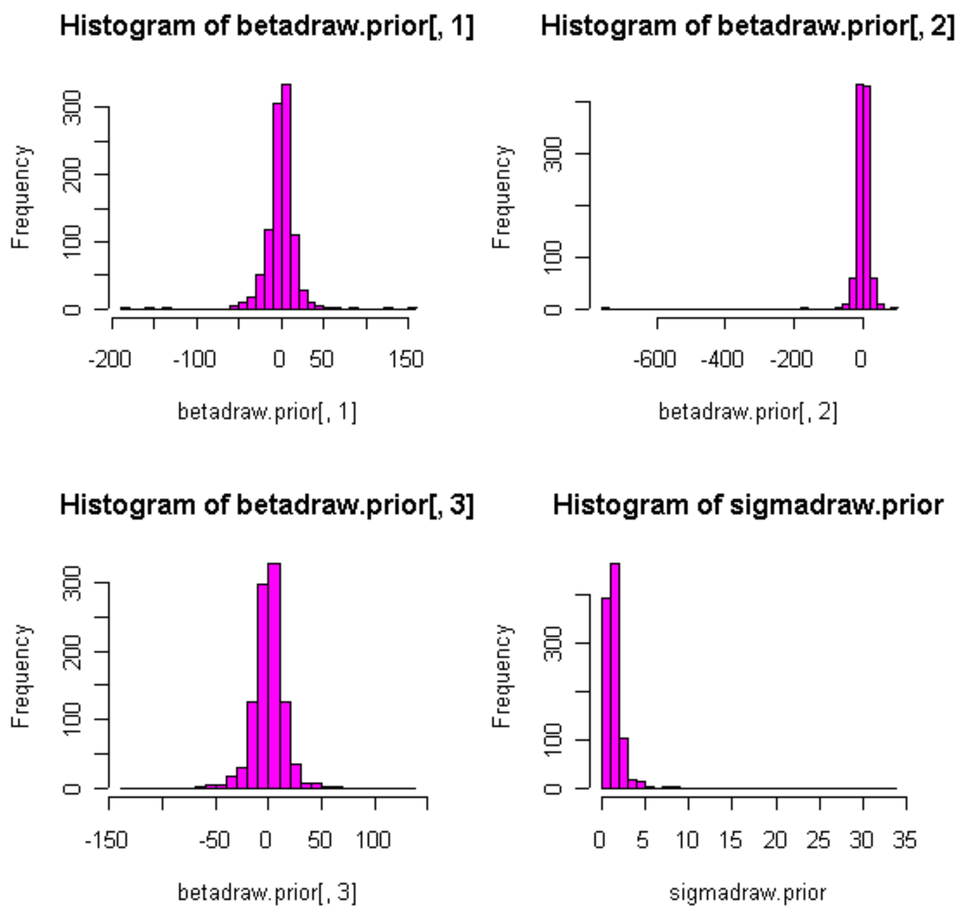
1. Univariate Bayes Regression

a. - see the R code in the file "univariate-regression.R" in the homework directory. The values of beta and sigma-squared from the regression are

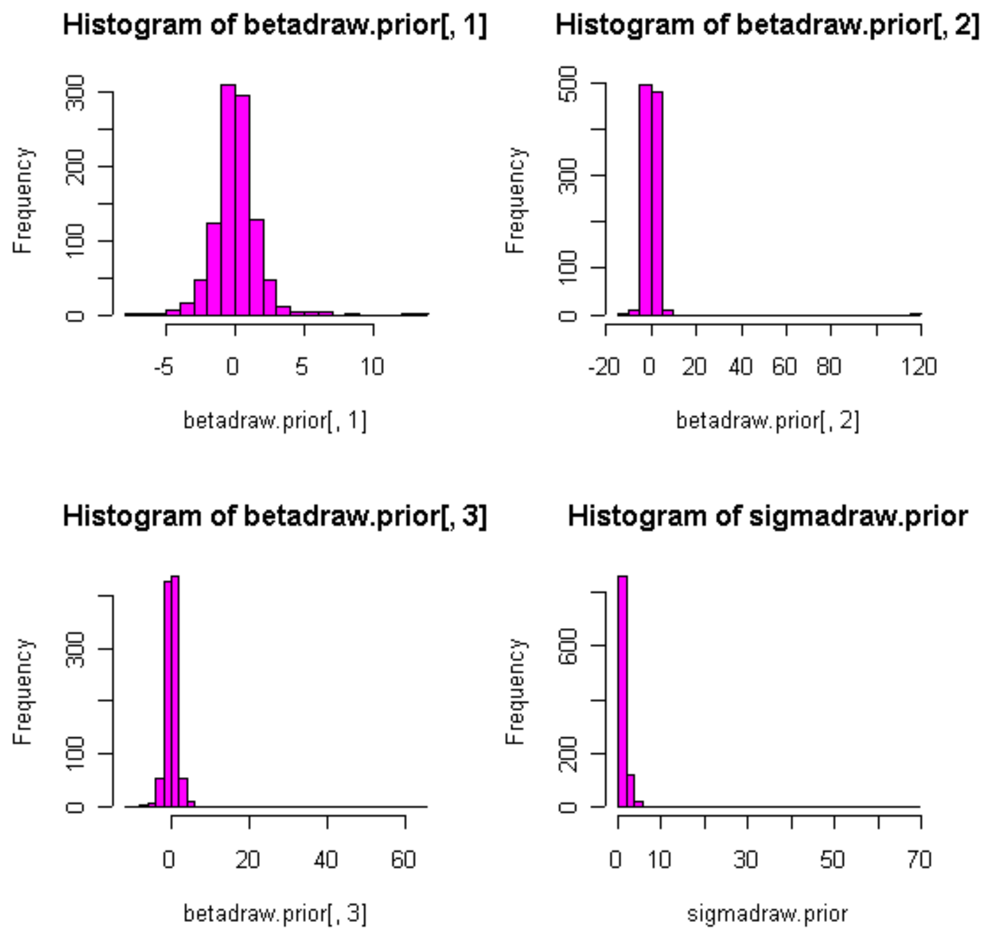
Parameter	Regression Estimate
beta0: intercept	9.3712
beta1: log(PRICE) coefficient	-1.2576
beta2: DISP coefficient	0.5132

b. See the R code.

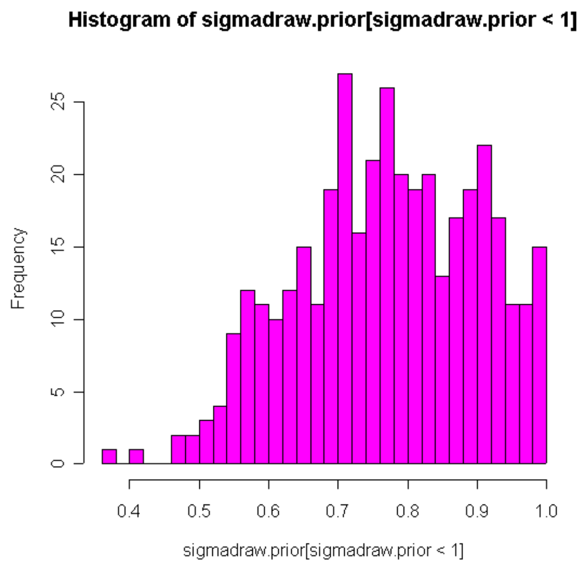
c. Prior distributions for the standard diffuse prior



Prior Distributions for a tighter prior : $A = I_{\dim(\beta)}$

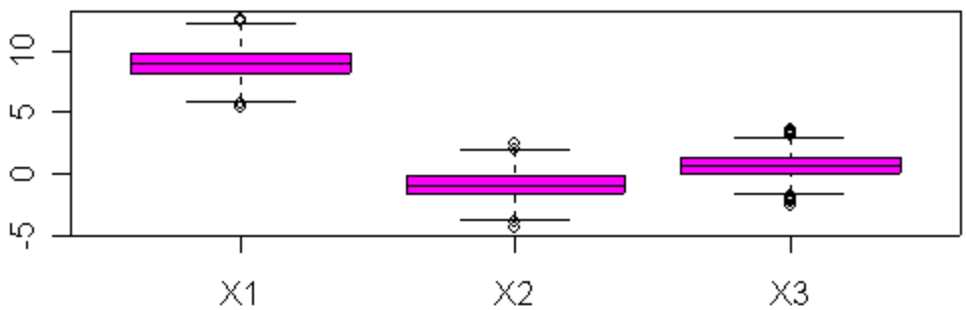


d. The prior distribution for the standard diffuse prior places very small (almost zero) mass on values below 0.5; thus the posterior will also place very low mass in this region as an artifact of the choice of the prior. This can be seen by looking at only a part of the distribution (sigma-sq < 1) on the following histogram.

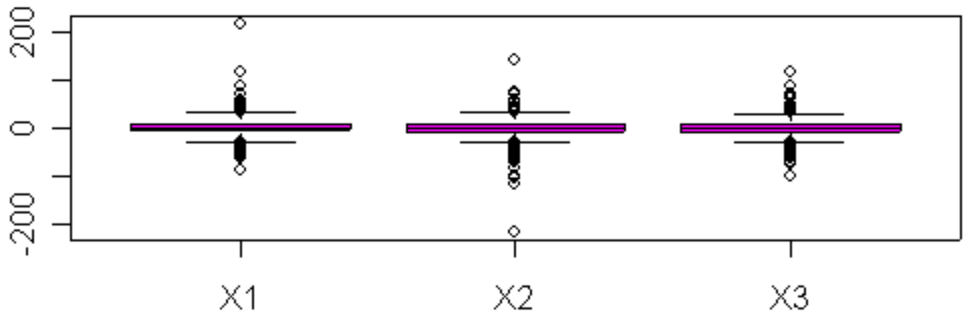


e.

Posterior Distributions for the three beta parameters, N=30



Prior Distributions for the three beta parameters, N=30

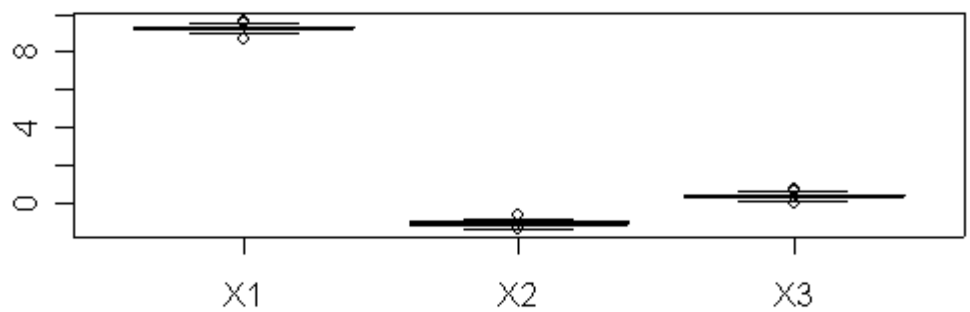


Parameter	Regression Estimate	Posterior	
		Mean	Variance
beta0: intercept	9.3712	8.9905	1.3836
beta1: log(PRICE) coefficient	-1.2576	-0.9830	1.0572
beta2: DISP coefficient	0.5132	0.6230	0.8166
N: sample size	30		

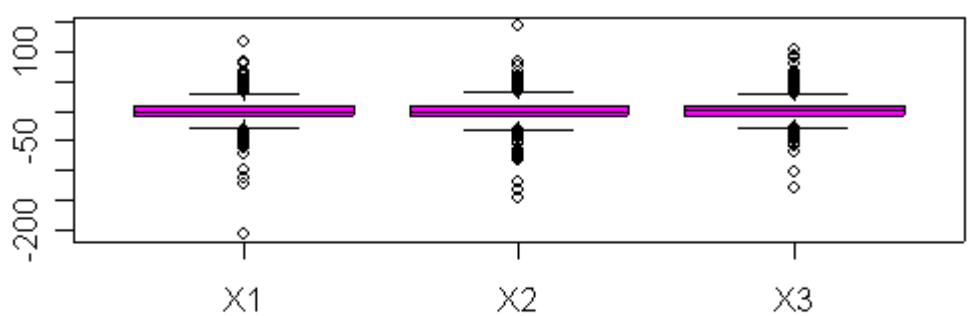
We note that the posterior distributions are not very tight, though the means of these distributions are reasonably close to the true values of the parameters. In fact, the priors have lower variance in this case than the posteriors.

If we increase the sample size from 30, we shall see the posterior distributions being much tighter. For instance, if we take a sample size of 2000, for which the posterior and prior distributions are plotted below, we note that the posteriors are much tighter than the priors.

Posterior Distributions for the three beta parameters, N=2000



Prior Distributions for the three beta parameters, N=2000



Parameter	Regression Estimate	Posterior	
		Mean	Variance
beta0: intercept	9.3712	9.2388	0.0109
beta1: log(PRICE) coefficient	-1.2576	-1.1249	0.0095
beta2: DISP coefficient	0.5132	0.3029	0.0110
N: sample size	2000		

2. Given Σ (i.e. with Σ known), we can write the conjugate prior for β as follows:

Rewrite the SUR model in the following form

$$y = X\beta + \varepsilon , \qquad (1)$$

$$y = \begin{bmatrix} y_1 \\ . \\ y_m \end{bmatrix}_{\text{N x 1}} , X = \begin{bmatrix} X_1 & . & 0 \\ . & . & . \\ 0 & . & X_m \end{bmatrix}_{\text{N x mk}} , \beta = \begin{bmatrix} \beta_1 \\ . \\ \beta_m \end{bmatrix}_{\text{mk x 1}} , \varepsilon = \begin{bmatrix} \varepsilon_1 \\ . \\ \varepsilon_m \end{bmatrix}_{\text{N x 1}}$$

$$\varepsilon \sim \text{MVN}(0, \Sigma \otimes I_n)$$

where

each of the y_i, ε_i are n x 1 vectors ($N = nm$), each of the β_i are k x 1 vectors and each of the X_i are n x k matrices

We can standardize the observations to remove the correlations as follows

$$\Sigma = LL'$$

Premultiplying by L^{-1} and postmultiplying by $\left(L^{-1}\right)'$, we get

$$L^{-1}\Sigma\left(L^{-1}\right)' = L^{-1}LL'\left(L^{-1}\right)' = L^{-1}LL'\left(L'\right)^{-1} = I_m$$

Now, if premultiply equation (1) by $\left(L^{-1} \otimes I_n\right)$, we get

$$\tilde{y} = \tilde{X}\beta + \tilde{\varepsilon} \qquad (2)$$

where

$$\tilde{y} = (L^{-1} \otimes I_n)y, \quad \tilde{X} = (L^{-1} \otimes I_n)X, \quad \tilde{\varepsilon} = (L^{-1} \otimes I_n)\varepsilon$$

$$\begin{aligned} \text{Var}(\tilde{\varepsilon}) &= E\left[\left(L^{-1} \otimes I_n\right)\varepsilon\varepsilon'\left(\left(L^{-1}\right)' \otimes I_n\right)\right] \\ &= \left(L^{-1} \otimes I_n\right)E\left[\varepsilon\varepsilon'\right]\left(\left(L^{-1}\right)' \otimes I_n\right) \\ &= \left(L^{-1} \otimes I_n\right)(\Sigma \otimes I_n)\left(\left(L^{-1}\right)' \otimes I_n\right) \\ &= I_m \otimes I_n = I_N \end{aligned}$$

Thus, the transformed equation has uncorrelated errors.

We can then write the likelihood as follows

$$p(\tilde{y} \mid \tilde{X}, \beta) \propto \exp\left(-\frac{1}{2} \left(\tilde{y} - \tilde{X}\beta \right)' \left(\tilde{y} - \tilde{X}\beta \right) \right)$$

and use the usual trick of the least squares projection, splitting up the resulting term into two parts and writing the posterior.

Using the prior, $\beta \sim N(\bar{\beta}, A^{-1})$

the posterior can be written as

$$\beta \mid \Sigma, y, X \sim N\left(\tilde{\beta}, \left(\tilde{X}'\tilde{X} + A \right)^{-1} \right) \qquad \tilde{\beta} = \left(\tilde{X}'\tilde{X} + A \right)^{-1} \left(\tilde{X}'\tilde{y} + A\bar{\beta} \right)$$

3. $[(\tilde{B}-\overline{B})' A(\tilde{B}-\overline{B})]$ is missing in the expression for S (see equation 2.12.5). If A is small relative to then this term can often be tiny relative to the residual sum of squares matrix. For "standard" diffuse prior settigs, A = .01 and this would be hard to detect. The multivariate regression model will be used in hierarchical models as part of the prior distribution. In those cases, A will not be small and this term could be very important. This is another reason to use informative priors! When you test your code you should try very informative settings to make sure the prior to posterior mapping is working!

4. The method in rwishart is more efficient because the Gelman method involves drawing from a multivariate normal ν times, where ν represents the degrees of freedom. The greater the value of ν, the slower the Gelman method will be, relative to the method in rwishart (which involves only one set of draws irrespective of the value of ν). see rwishart2.R for implementation of the "Gelman" method.

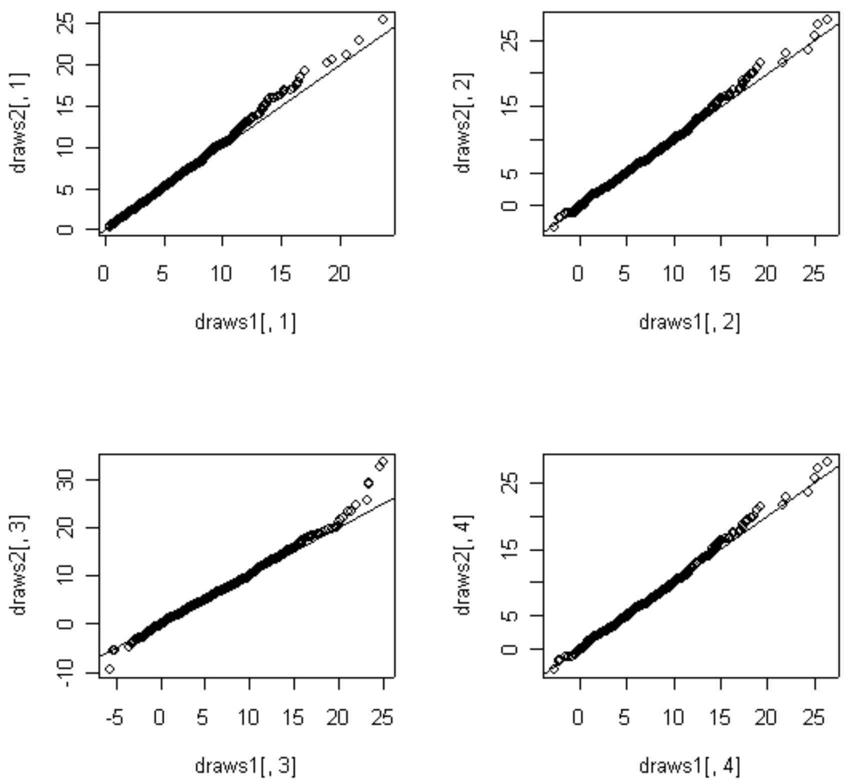
   More efficient code than (rwishart2) but harder to read would look something like:

```
Z=matrix(rnorm(m*nu),ncol=nu)
X=chol(V)%*%X
W=crossprod(t(X))
```

   In order to test our code for the Wishart draw using the Gelman method, we make a set of draws using the two different methods and compare them. For this exercise, let ν =6 and the positive definite matrix V be

$$V = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 2 & 2 \\ 1 & 2 & 3 \end{bmatrix}$$

   The q-q plots of the draws obtained using the two methods (draws1 for rwishart and draws2 for rwishart2) are shown below. The code for testing the two functions is in test-wishart.R


draws2[, 1]
draws1[, 1]
draws2[, 2]
draws1[, 2]
draws2[, 3]
draws1[, 3]
draws2[, 4]
draws1[, 4]