

Estimation & Decisions

STA721 Linear Models Duke University

Merlise Clyde

September 24, 2014

Model

$$\mathbf{Y} \sim N(\mathbf{X}\boldsymbol{\beta}, \mathbf{I}_n/\phi)$$

with precision $\phi = 1/\sigma^2$.

More Prior Choices:

- More on g-priors
- Zellner-Siow Cauchy Prior
- Utility and choice of Estimators

Another Version of Zellner's g -prior

$$\begin{aligned}\mathbf{Y} &= \mathbf{1}\beta_0 + \mathbf{X}_1\boldsymbol{\beta} + \boldsymbol{\epsilon} \\ p(\beta_0, \phi) &\propto 1 \\ \boldsymbol{\beta} \mid \phi &\sim N(\mathbf{0}, \frac{g}{\phi}(\mathbf{X}^T(\mathbf{I}_n - \mathbf{P}_1)\mathbf{X})^{-1})\end{aligned}$$

Note

$$(\mathbf{X}^T(\mathbf{I}_n - \mathbf{P}_1)\mathbf{X}) = (\mathbf{X}^T(\mathbf{I}_n - \mathbf{P}_1)^T(\mathbf{I}_n - \mathbf{P}_1)\mathbf{X}) = (\mathbf{X} - \mathbf{1}_n\bar{\mathbf{X}}^T)^T(\mathbf{X} - \mathbf{1}_n\bar{\mathbf{X}})$$

Let $(\mathbf{X} - \mathbf{1}_n\bar{\mathbf{X}}^T)^T(\mathbf{X} - \mathbf{1}_n\bar{\mathbf{X}}) = \text{SS}_X = \mathbf{U}^T\mathbf{U}$ Contribution quadratic to the log likelihood from prior after integrating out β_0

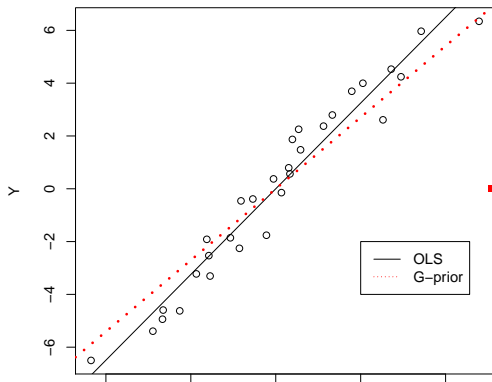
$$(\mathbf{Y}_c - \mathbf{X}_c\boldsymbol{\beta})^T(\mathbf{Y}_c - \mathbf{X}_c\boldsymbol{\beta}) + (\boldsymbol{\beta}^T \frac{\mathbf{U}^T\mathbf{U}}{g}\boldsymbol{\beta})$$

$$(\mathbf{Y}_c - \mathbf{X}_c\boldsymbol{\beta})^T(\mathbf{Y}_c - \mathbf{X}_c\boldsymbol{\beta}) + (\mathbf{0}_p - \frac{\mathbf{U}}{\sqrt{g}}\boldsymbol{\beta})^T(\mathbf{0}_p - \frac{\mathbf{U}}{\sqrt{g}}\boldsymbol{\beta})$$

Example

In SLR it is like an extra $Y_0 = 0$ at $\mathbf{X}_0 = \sqrt{\frac{SS_x}{g}}$:

$$(\mathbf{Y}_c - \mathbf{X}_c\beta)^T(\mathbf{Y}_c - \mathbf{X}_c\beta) + (0 - \sqrt{\frac{SS_x}{g}}\beta)^T(0 - \sqrt{\frac{SS_x}{g}}\beta)$$



Disadvantages of Conjugate Priors

Disadvantages:

- Results may have be sensitive to prior “outliers” due to linear updating
- Problem potentially with all Normal priors, not just the g -prior.
- Cannot capture all possible prior beliefs
- Mixtures of Conjugate Priors

Mixtures of Conjugate Priors

Theorem (Diaconis & Ylvisaker 1985)

Given a sampling model $p(y \mid \theta)$ from an exponential family, any prior distribution can be expressed as a mixture of conjugate prior distributions

- Prior $p(\theta) = \int p(\theta \mid \omega)p(\omega) d\omega$
- Posterior

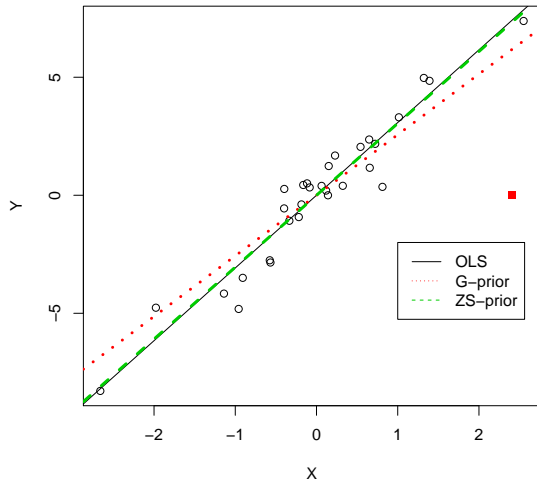
$$\begin{aligned} p(\theta \mid \mathbf{Y}) &\propto \int p(\mathbf{Y} \mid \theta)p(\theta \mid \omega)p(\omega) d\omega \\ &\propto \int \frac{p(\mathbf{Y} \mid \theta)p(\theta \mid \omega)}{p(\mathbf{Y} \mid \omega)} p(\mathbf{Y} \mid \omega)p(\omega) d\omega \\ &\propto \int p(\theta \mid \mathbf{Y}, \omega)p(\mathbf{Y} \mid \omega)p(\omega) d\omega \\ p(\theta \mid \mathbf{Y}) &= \frac{\int p(\theta \mid \mathbf{Y}, \omega)p(\mathbf{Y} \mid \omega)p(\omega) d\omega}{\int p(\mathbf{Y} \mid \omega)p(\omega) d\omega} \end{aligned}$$

Zellner's g -prior $\beta \mid \phi \sim N(\mathbf{b}_0, gn(\mathbf{X}^T \mathbf{X})^{-1}/\phi)$

- Choice of g ?
- $\frac{g}{1+g}$ weight given to the data
- Let $\tau = 1/g$ assign $\tau \sim G(1/2, 1/2)$
- Find prior distribution
- Can express posterior as a mixture of g -priors

Example Again

From JAGS:




```
model = function(){  
  for (i in 1:n) {  
    Y[i] ~ dnorm(X[i]*beta, phi)  
  }  
  beta ~ dnorm(0, SSX/(n*phi*lambda))  
  phi ~ dgamma(.05, .05)  
  lambda ~ dgamma(.5, .5)  
}
```

```
write.model(model, "ZSmodel")  
model.file="ZSmodel"  
data = list(Y=Y, X=X, n =length(Y), SSX=sum(X^2) )  
ZSout = jags(data, inits=NULL,  
             parameters.to.save=c("beta", "lambda", "phi"),  
             model=model.file, n.iter=10000)
```

How Good are these Estimators?

Quadratic loss for estimating β using estimator \mathbf{a}

$$L(\beta, \mathbf{a}) = (\beta - \mathbf{a})^T (\beta - \mathbf{a})$$

- \mathbf{a} equals the Posterior mean $E[\beta \mid \mathbf{Y}]$ minimizes Posterior expected loss.
- Consider our expected loss (before we see the data) of taking an “action” \mathbf{a}
- Under OLS or the Reference prior the Expected Mean Square Error

$$\begin{aligned} E_{\mathbf{Y}}[(\beta - \hat{\beta})^T (\beta - \hat{\beta})] &= \sigma^2 \text{tr}[(\mathbf{X}^T \mathbf{X})^{-1}] \\ &= \sigma^2 \sum_{j=1}^p \lambda_j^{-1} \end{aligned}$$

where λ_j are eigenvalues of $\mathbf{X}^T \mathbf{X}$.

- If smallest $\lambda_j \rightarrow 0$ then $\text{MSE} \rightarrow \infty$
- Note: estimate is unbiased!

Is the g -prior better?

Explore Frequentist properties of using a Bayesian estimator

$$E_{\mathbf{Y}}[(\boldsymbol{\beta} - \hat{\boldsymbol{\beta}}_g)^T (\boldsymbol{\beta} - \hat{\boldsymbol{\beta}}_g)]$$

but now $\hat{\boldsymbol{\beta}}_g = g/(1+g)\hat{\boldsymbol{\beta}}$ for g prior.

when is the g prior better than the Reference prior of OLS?

Estimator Properties

- Bias
- Variability
- $MSE = \text{Bias}^2 + \text{Variance}$ (multivariate analogs)
- Problems with OLS, g -priors and mixtures of g -priors with collinearity
- Solutions:
 - removal of terms
 - other shrinkage estimators