

Lab4: Fitting Bayesian Models in JAGS

STA721 Linear Models Duke University

October 10, 2019

2 Block g -prior (Normal-Jeffreys)

Model in centered parameterization (review)

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

Log Likelihood (show)

$$\mathcal{L}(\beta_0, \beta_1, \phi) \propto \frac{n}{2} \log(\phi) - \frac{\phi}{2} (n(\beta_0 - \bar{y})^2 + (\mathbf{Y}_c - \mathbf{X}_c\boldsymbol{\beta})^T(\mathbf{Y}_c - \mathbf{X}_c\boldsymbol{\beta}))$$

Since

$$\mathbf{Y} = (\mathbf{I} - \mathbf{P}_1)\mathbf{Y} + \mathbf{P}_1\mathbf{Y}$$

Bayesian Estimation with g prior (2 groups)

$$\mathbf{Y} = \mathbf{1}\beta_0 + (\mathbf{I}_n - \mathbf{P}_1)\mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$$

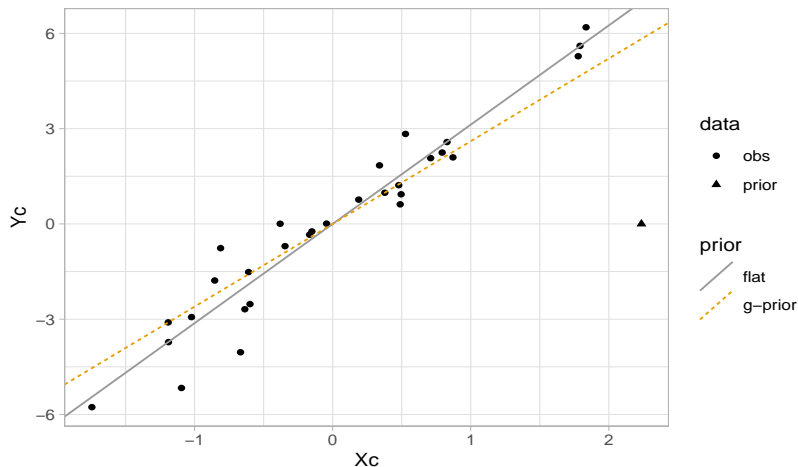
$$p(\beta_0, \phi) \propto 1/\phi$$

$$\boldsymbol{\beta} \mid \phi \sim \text{N}(\mathbf{0}, \frac{g}{\phi}(\mathbf{X}^T(\mathbf{I}_n - \mathbf{P}_1)\mathbf{X})^{-1})$$

Example from class: $g=5$, $n=30$

In SLR g prior contribution is like adding an extra $Y_0 = 0$ at

$$\mathbf{X}_0 = \sqrt{\frac{SS_x}{g}}:$$



Markov Chain Monte Carlo

- ▶ We know that $\beta_0, \beta, \phi \mid \mathbf{Y}, g = 1/\tau$ has a Normal-Gamma distribution
- ▶ Derive full conditional for $\theta = [\beta_0, \beta, \phi]$ as a sequence of distributions β_0 and β given ϕ, \mathbf{Y} (are independent) and $\phi \mid \mathbf{Y}$ [Successive Substitution Sampling] (in HW)
- ▶ Let $\tau = 1/g$ have a Gamma distribution $G(1/b, n/2)$ We can show that $\tau \mid \beta_0, \beta, \phi, \mathbf{Y}$ has a Gamma distribution (derive HW)

$$p(\tau \mid \beta, \phi, \mathbf{Y}) \propto \mathcal{L}(\beta_0, \beta, \phi) \tau^{p/2} e^{(-\tau \frac{\phi}{2} \beta^T (\mathbf{X}^T \mathbf{X}) \beta)} \tau^{1/2-1} e^{-\tau n/2}$$

- ▶ alternate sampling from two blocks of full conditional distributions given current values of other parameters.
- ▶ implement in JAGS or STAN

JAGS Code: library(R2jags)

```
model = function(){  
  for (i in 1:n) {  
    Y[i] ~ dnorm(beta0+ (X[i] -Xbar)*beta, phi)  
  }  
  beta0 ~ dnorm(0, .000001*phi) #precision is 2nd arg  
  beta ~ dnorm(0, phi*tau*SSX) #precision is 2nd arg  
  phi ~ dgamma(.001, .001) # approximate independent Jeffreys  
  tau ~ dgamma(.5, .5*n)  
  g <- 1/tau  
  sigma <- pow(phi, -.5)  
}  
data = list(Y=Y, X=X, n=length(Y), SSX=sum(Xc^2),  
            Xbar=mean(X))  
ZSout = jags(data, inits=NULL,  
             parameters.to.save=c("beta0", "beta", "g",  
                                  "sigma"),  
             model=model, n.iter=10000)
```

HPD intervals

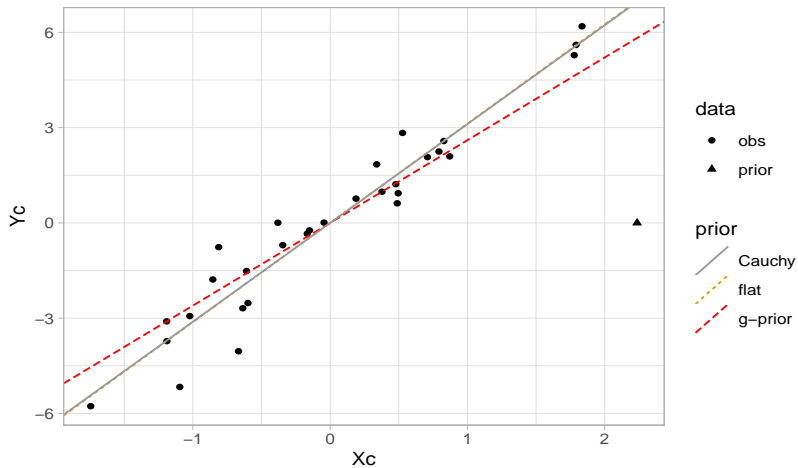
```
confint(lm(Y ~ Xc))
```

```
##                2.5 %    97.5 %  
## (Intercept) -0.3985359 0.2048303  
## Xc           2.7945824 3.4555162
```

```
HPDinterval(as.mcmc(ZSout$BUGSoutput$sims.matrix))
```

```
##                lower      upper  
## beta           2.7823047    3.4453690  
## beta0          -0.3764027    0.2095465  
## deviance       70.2043917    78.4813041  
## g              19.4503373 3782.7134974  
## sigma          0.6171029    1.0504892  
## attr(,"Probability")  
## [1] 0.95
```

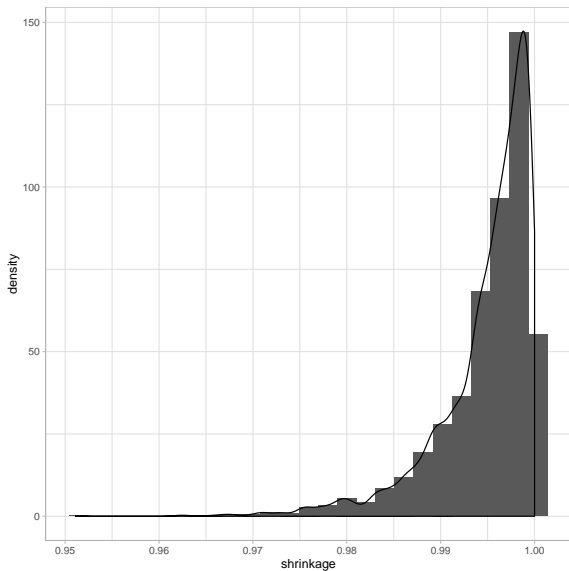
Compare



ZSout

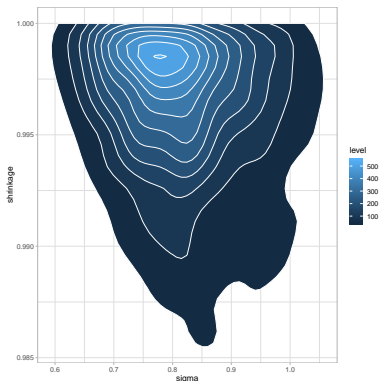
```
## Inference for Bugs model at "/var/folders/n4/nj1122xj6bn5_xgbptv7bml40000gp/T//RtmpRxs007/model17bbc1a
## 3 chains, each with 10000 iterations (first 5000 discarded), n.thin = 5
## n.sims = 3000 iterations saved
##          mu.vect   sd.vect   2.5%    25%    50%    75%    97.5%  Rhat
## beta          3.112     0.170   2.782   2.997   3.115   3.225   3.445  1.001
## beta0         -0.099     0.152  -0.384  -0.204  -0.099   0.001   0.204  1.002
## g            2263.147 38967.029 48.273 146.129 282.298 697.063 9018.709 1.001
## sigma          0.827     0.114   0.636   0.747   0.816   0.896   1.079  1.001
## deviance       73.347     2.563  70.390  71.458  72.680  74.500  79.882  1.002
##          n.eff
## beta          3000
## beta0         1200
## g             3000
## sigma         3000
## deviance      1600
##
## For each parameter, n.eff is a crude measure of effective sample size,
## and Rhat is the potential scale reduction factor (at convergence, Rhat=1).
##
## DIC info (using the rule, pD = var(deviance)/2)
## pD = 3.3 and DIC = 76.6
## DIC is an estimate of expected predictive error (lower deviance is better).
```

Posterior Distribution of shrinkage



Joint Distribution of σ and $g/(1+g)$

```
ggplot(postdf, aes(x=sigma, y=shrinkage) ) +  
  stat_density_2d(aes(fill = ..level..),  
                  geom = "polygon", colour="white") +  
  theme_light()
```



Cauchy Summary

- ▶ Cauchy rejects prior mean if it is an "outlier"
- ▶ robustness related to "bounded" influence (more later)
- ▶ requires numerical integration or Monte Carlo sampling (MCMC)