Lasso and Bayesian Lasso

Chapter 6 ISL

STA521 Linear Models Duke University

Merlise Clyde

September 26, 2014

Ridge Regression Model

Write $\mathbf{Y} = \mathbf{1}_n\alpha + \mathbf{X}^c\beta + \epsilon$ where $\mathbf{X}^c$ has been centered and standardized

- Prior Distribution on

$$\beta \mid \phi \sim \mathrm{N}(\mathbf{0}_p, \frac{1}{\phi k}\mathbf{I}_p)$$

Ridge Regression Model

Write $\mathbf{Y} = \mathbf{1}_n\alpha + \mathbf{X}^c\boldsymbol{\beta} + \boldsymbol{\epsilon}$ where $\mathbf{X}^c$ has been centered and standardized

- Prior Distribution on

$$\boldsymbol{\beta} \mid \phi \sim \mathrm{N}(\mathbf{0}_p, \frac{1}{\phi k}\mathbf{I}_p)$$

- Posterior mean

$$\tilde{\boldsymbol{\beta}} = (\mathbf{X}^{cT}\mathbf{X}^c + k\mathbf{I})^{-1}\mathbf{X}^{cT}\mathbf{X}^c\hat{\boldsymbol{\beta}}$$

Ridge Regression Model

Write $\mathbf{Y} = \mathbf{1}_n\alpha + \mathbf{X}^c\boldsymbol{\beta} + \boldsymbol{\epsilon}$ where $\mathbf{X}^c$ has been centered and standardized

- Prior Distribution on

$$\boldsymbol{\beta} \mid \phi \sim \mathrm{N}(\mathbf{0}_p, \frac{1}{\phi k}\mathbf{I}_p)$$

- Posterior mean

$$\tilde{\beta} = (\mathbf{X}^{cT}\mathbf{X}^c + k\mathbf{I})^{-1}\mathbf{X}^{cT}\mathbf{X}^c\hat{\beta}$$

- Prior distribution on k?

Priors on k

Reparameterization to canonical representation:

$$\mathbf{Y} = \mathbf{1}\alpha + (\mathbf{I} - \mathsf{P}_{\mathbf{1}})\mathbf{X}SS\boldsymbol{\beta} + \boldsymbol{\epsilon}$$

Priors on k

Reparameterization to canonical representation:

$$
\begin{aligned}
\mathbf{Y} &= \mathbf{1}\alpha + (\mathbf{I} - \mathsf{P}_{\mathbf{1}})\mathbf{X}SS\boldsymbol{\beta} + \boldsymbol{\epsilon} \\
&= \mathbf{1}\alpha + \mathbf{X}^c\boldsymbol{\beta}^c + \boldsymbol{\epsilon}
\end{aligned}
$$

Hierarchical prior

Priors on k

Reparameterization to canonical representation:

$$\begin{aligned} \mathbf{Y} &= \mathbf{1}\alpha + (\mathbf{I} - \mathsf{P}_{\mathbf{1}})\mathbf{X}SS\boldsymbol{\beta} + \boldsymbol{\epsilon} \\ &= \mathbf{1}\alpha + \mathbf{X}^c\boldsymbol{\beta}^c + \boldsymbol{\epsilon} \end{aligned}$$

Hierarchical prior

- $p(\alpha \mid \phi, \gamma, \kappa) \propto 1$

Priors on k

Reparameterization to canonical representation:

$$\begin{aligned} \mathbf{Y} &= \mathbf{1}\alpha + (\mathbf{I} - \mathsf{P}_{\mathbf{1}})\mathbf{X}SS\boldsymbol{\beta} + \boldsymbol{\epsilon} \\ &= \mathbf{1}\alpha + \mathbf{X}^c\boldsymbol{\beta}^c + \boldsymbol{\epsilon} \end{aligned}$$

Hierarchical prior

- $p(\alpha \mid \phi, \gamma, \kappa) \propto 1$
- $\boldsymbol{\beta}^c \mid \phi, \kappa \sim \mathsf{N}(\mathbf{0}, \mathbf{I}(\phi\kappa)^{-1})$

Priors on k

Reparameterization to canonical representation:

$$\begin{aligned} \mathbf{Y} &= \mathbf{1}\alpha + (\mathbf{I} - \mathsf{P}_{\mathbf{1}})\mathbf{X}SS\boldsymbol{\beta} + \boldsymbol{\epsilon} \\ &= \mathbf{1}\alpha + \mathbf{X}^c\boldsymbol{\beta}^c + \boldsymbol{\epsilon} \end{aligned}$$

Hierarchical prior

- $p(\alpha \mid \phi, \gamma, \kappa) \propto 1$
- $\boldsymbol{\beta}^c \mid \phi, \kappa \sim \mathsf{N}(\mathbf{0}, \mathbf{I}(\phi\kappa)^{-1})$
- $p(\phi \mid \kappa) \propto 1/\phi$

Priors on k

Reparameterization to canonical representation:

$$\begin{aligned}\mathbf{Y} &= \mathbf{1}\alpha + (\mathbf{I} - \mathsf{P}_{\mathbf{1}})\mathbf{X}SS\boldsymbol{\beta} + \boldsymbol{\epsilon} \\ &= \mathbf{1}\alpha + \mathbf{X}^c\boldsymbol{\beta}^c + \boldsymbol{\epsilon}\end{aligned}$$

Hierarchical prior

- $p(\alpha \mid \phi, \boldsymbol{\gamma}, \kappa) \propto 1$
- $\boldsymbol{\beta}^c \mid \phi, \kappa \sim \mathsf{N}(\mathbf{0}, \mathbf{I}(\phi\kappa)^{-1})$
- $p(\phi \mid \kappa) \propto 1/\phi$
- prior on κ? Take $\kappa \mid \phi \sim \mathbf{G}(1/2, 1/2)$

Joint Distribution

- $\alpha, \boldsymbol{\beta}^c, \phi \mid \kappa, \mathbf{Y}$ Normal-Gamma family given $\mathbf{Y}$ and κ

Posterior Distributions

Joint Distribution

- $\alpha, \boldsymbol{\beta}^c, \phi \mid \kappa, \mathbf{Y}$ Normal-Gamma family given $\mathbf{Y}$ and κ
- $\kappa \mid \mathbf{Y}$ not tractable

Posterior Distributions

Joint Distribution

- $\alpha, \boldsymbol{\beta}^c, \phi \mid \kappa, \mathbf{Y}$ Normal-Gamma family given $\mathbf{Y}$ and κ
- $\kappa \mid \mathbf{Y}$ not tractable

Obtain marginal for $\boldsymbol{\beta}^c$ via

Posterior Distributions

Joint Distribution

- $\alpha, \boldsymbol{\beta}^c, \phi \mid \kappa, \mathbf{Y}$ Normal-Gamma family given $\mathbf{Y}$ and κ
- $\kappa \mid \mathbf{Y}$ not tractable

Obtain marginal for $\boldsymbol{\beta}^c$ via

- Numerical integration

Joint Distribution

- $\alpha, \boldsymbol{\beta}^c, \phi \mid \kappa, \mathbf{Y}$ Normal-Gamma family given $\mathbf{Y}$ and κ
- $\kappa \mid \mathbf{Y}$ not tractable

Obtain marginal for $\boldsymbol{\beta}^c$ via

- Numerical integration
- MCMC: Gibbs Sampling from Full Conditional Distributions
  - $\alpha, \boldsymbol{\gamma}, \phi \mid \kappa, \mathbf{Y}$ Normal-Gamma family given $\mathbf{Y}$ and κ

Posterior Distributions

Joint Distribution

- $\alpha, \boldsymbol{\beta}^c, \phi \mid \kappa, \mathbf{Y}$ Normal-Gamma family given $\mathbf{Y}$ and κ
- $\kappa \mid \mathbf{Y}$ not tractable

Obtain marginal for $\boldsymbol{\beta}^c$ via

- Numerical integration
- MCMC: Gibbs Sampling from Full Conditional Distributions
  - $\alpha, \boldsymbol{\gamma}, \phi \mid \kappa, \mathbf{Y}$ Normal-Gamma family given $\mathbf{Y}$ and κ
  - $\kappa \mid \alpha, \boldsymbol{\gamma}, \phi, \mathbf{Y}$ is tractable!

Full Conditional for k

MCMC Algorithm

- Pick initial values $\alpha^{(0)}, \boldsymbol{\beta}^{(0)}, \phi^{(0)}$,

MCMC Algorithm

- Pick initial values $\alpha^{(0)}, \boldsymbol{\beta}^{(0)}, \phi^{(0)}$,
  Set $t = 1$
  1. Sample $\kappa^{(t)} \sim p(\kappa \mid \alpha^{(t-1)}, \boldsymbol{\gamma}^{(t-1)}, \phi^{(t-1)}, \mathbf{Y})$

MCMC Algorithm

- Pick initial values $\alpha^{(0)}, \boldsymbol{\beta}^{(0)}, \phi^{(0)}$,
  Set $t = 1$
  1. Sample $\kappa^{(t)} \sim p(\kappa \mid \alpha^{(t-1)}, \boldsymbol{\gamma}^{(t-1)}, \phi^{(t-1)}, \mathbf{Y})$
  2. Sample $\alpha^{(t)}, \boldsymbol{\gamma}^{(t)}, \phi^{(t)} \mid \kappa^{(t)}, \mathbf{Y}$

MCMC Algorithm

- Pick initial values $\alpha^{(0)}, \boldsymbol{\beta}^{(0)}, \phi^{(0)}$,
  Set $t = 1$
  1. Sample $\kappa^{(t)} \sim p(\kappa \mid \alpha^{(t-1)}, \boldsymbol{\gamma}^{(t-1)}, \phi^{(t-1)}, \mathbf{Y})$
  2. Sample $\alpha^{(t)}, \boldsymbol{\gamma}^{(t)}, \phi^{(t)} \mid \kappa^{(t)}, \mathbf{Y}$
  3. Set $t = t + 1$ and repeat until $t > T$

MCMC Algorithm

- Pick initial values $\alpha^{(0)}, \boldsymbol{\beta}^{(0)}, \phi^{(0)}$,
  Set $t = 1$
  1. Sample $\kappa^{(t)} \sim p(\kappa \mid \alpha^{(t-1)}, \boldsymbol{\gamma}^{(t-1)}, \phi^{(t-1)}, \mathbf{Y})$
  2. Sample $\alpha^{(t)}, \boldsymbol{\gamma}^{(t)}, \phi^{(t)} \mid \kappa^{(t)}, \mathbf{Y}$
  3. Set $t = t + 1$ and repeat until $t > T$

  Use Samples $\alpha^{(t)}, \boldsymbol{\gamma}^{(t)}, \phi^{(t)}, \kappa^{(t)}$ for $t = B, \ldots, T$ for inference
  (B = Burnin)

JAGS

Just Another Gibbs Sampler (and more)

- Stanalone Program (for MAC, Windows, Unix)
- R interface
- Easy scripting language (no full conoditional derivation!)
- CODA Diagnostics
- Wide Range of Models/Priors!

Example

```
  rr.model = function() {
  for (i in 1:n) {
      mu[i] <- inprod(X[i,], alpha)
      Y[i] ~ dnorm(mu[i], phi)
  }
  phi ~ dgamma(1.0E-6, 1.0E-6)
  sigma <- pow(phi, -.5)
  alpha[1] ~ dnorm(0, 1.0E-10)

  lambda.beta ~ dgamma(.5, .5)
  for (j in 2:p) {
      prec.beta[j] <- lambda.beta*phi
      alpha[j] ~ dnorm(0, prec.beta[j])
      beta[j] <- alpha[j]/scales[j]    # rescale
  }
  beta[1] <- alpha[1] - inprod(beta[2:p], Xbar)
}
```

Initial Values

```
 rr.inits = function() {
    bf.lm <- lm(data$Y ~ -1 + data$X)
    coefs = coef(bf.lm)
    alpha= coefs
    phi  = (1/summary(bf.lm)$sigma)^2
return(list(alpha=alpha, phi=phi))
  }
```

Running

```
rr.model.file = paste(getwd(),"rr-model.txt", sep="/")
write.model(rr.model, rr.model.file)

parameters = c("beta","sigma","lambda.beta")

bf.sim = jags(data, inits=rr.inits, parameters,
              model.file=rr.model.file,  n.iter=10000)
plot(bf.sim)
bf.sim
```

Output

```
> bf.sim
Inference for Bugs model at "/home/fac/clyde/Dropbox/Sta521/Lectures/ridge/rr-model.txt", fit using jags,
 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 n.eff
beta[1]     -464.642 415.597 -1296.740  -743.015  -459.517  -170.197   321.255 1.001  3000
beta[2]     -497.325 137.834  -761.888  -590.347  -497.310  -404.987  -221.117 1.001  3000
beta[3]        1.567   0.042     1.488     1.538     1.567     1.596     1.647 1.002  1300
beta[4]       -0.812   0.182    -1.167    -0.935    -0.812    -0.686    -0.449 1.001  3000
beta[5]       49.125   5.684    38.420    45.290    49.024    53.027    60.348 1.001  3000
beta[6]      -13.752   4.615   -22.989   -16.781   -13.719   -10.688    -4.698 1.001  3000
beta[7]        0.053   0.031    -0.009     0.032     0.054     0.075     0.114 1.001  3000
beta[8]        0.044   0.032    -0.020     0.023     0.044     0.066     0.108 1.001  3000
beta[9]       -0.084   0.019    -0.121    -0.096    -0.084    -0.071    -0.047 1.002  1600
beta[10]       0.153   0.048     0.057     0.122     0.154     0.185     0.248 1.001  3000
beta[11]       0.026   0.244    -0.458    -0.137     0.024     0.193     0.498 1.001  3000
beta[12]       0.032   0.063    -0.089    -0.012     0.032     0.075     0.156 1.001  3000
beta[13]      -8.507   4.562   -17.402   -11.596    -8.471    -5.468     0.272 1.001  3000
beta[14]      -3.648   5.032   -13.822    -6.907    -3.557    -0.123     5.761 1.002  1900
beta[15]      15.061  12.816    -9.761     6.503    14.759    23.718    40.284 1.001  3000
beta[16]      -0.070   4.062    -7.787    -2.917    -0.096     2.624     7.942 1.001  2300
beta[17]       0.078   0.012     0.053     0.069     0.078     0.086     0.102 1.001  3000
beta[18]       8.690   2.925     2.974     6.745     8.739    10.751    14.306 1.001  3000
lambda.beta    1.102   0.380     0.487     0.827     1.054     1.325     1.923 1.001  3000
sigma       1042.370  26.961   992.679  1023.561  1041.502  1060.475  1094.968 1.002  1200
deviance   13003.732   6.262 12993.517 12999.206 13003.168 13007.578 13017.910 1.001  3000
```

Bayesian Lasso

Park & Casella (JASA 2008) and Hans Biometrika (2010) propose Bayesian versions of the Lasso

Bayesian Lasso

Park & Casella (JASA 2008) and Hans Biometrika (2010) propose Bayesian versions of the Lasso

$$\mathbf{Y} \mid \alpha, \boldsymbol{\beta}, \phi \quad \sim \quad \mathsf{N}(\mathbf{1}_n\alpha + \mathbf{X}^c\boldsymbol{\beta}, \mathbf{I}_n/\phi)$$

Bayesian Lasso

Park & Casella (JASA 2008) and Hans Biometrika (2010) propose Bayesian versions of the Lasso

$$\begin{aligned} \mathbf{Y} \mid \alpha, \boldsymbol{\beta}, \phi &\sim \mathsf{N}(\mathbf{1}_n\alpha + \mathbf{X}^c\boldsymbol{\beta}, \mathbf{I}_n/\phi) \\ \boldsymbol{\beta} \mid \alpha, \phi, \boldsymbol{\tau} &\sim \mathsf{N}(\mathbf{0}, \mathrm{diag}(\boldsymbol{\tau}^2)/\phi) \end{aligned}$$

Bayesian Lasso

Park & Casella (JASA 2008) and Hans Biometrika (2010) propose Bayesian versions of the Lasso

$$\begin{array}{rcl}\mathbf{Y} \mid \alpha, \boldsymbol{\beta}, \phi & \sim & \mathsf{N}(\mathbf{1}_n\alpha + \mathbf{X}^c\boldsymbol{\beta}, \mathbf{I}_n/\phi) \\ \boldsymbol{\beta} \mid \alpha, \phi, \boldsymbol{\tau} & \sim & \mathsf{N}(\mathbf{0}, \mathrm{diag}(\boldsymbol{\tau}^2)/\phi) \\ \tau_1^2 \ldots, \tau_p^2 \mid \alpha, \phi & \textit{i.i.d.} & \mathsf{Exp}(\lambda^2/2)\end{array}$$

Park & Casella (JASA 2008) and Hans Biometrika (2010) propose Bayesian versions of the Lasso

$$\begin{aligned}
\mathbf{Y} \mid \alpha, \boldsymbol{\beta}, \phi &\quad \sim \quad \mathsf{N}(\mathbf{1}_n \alpha + \mathbf{X}^c \boldsymbol{\beta}, \mathbf{I}_n/\phi) \\
\boldsymbol{\beta} \mid \alpha, \phi, \boldsymbol{\tau} &\quad \sim \quad \mathsf{N}(\mathbf{0}, \operatorname{diag}(\boldsymbol{\tau}^2)/\phi) \\
\tau_1^2 \ldots, \tau_p^2 \mid \alpha, \phi &\quad \textit{i.i.d.} \quad \mathsf{Exp}(\lambda^2/2) \\
p(\alpha, \phi) &\quad \propto \quad 1/\phi
\end{aligned}$$

Bayesian Lasso

Park & Casella (JASA 2008) and Hans Biometrika (2010) propose Bayesian versions of the Lasso

$$
\begin{aligned}
\mathbf{Y} \mid \alpha, \boldsymbol{\beta}, \phi &\quad \sim \quad \mathrm{N}(\mathbf{1}_n \alpha + \mathbf{X}^c \boldsymbol{\beta}, \mathbf{I}_n/\phi) \\
\boldsymbol{\beta} \mid \alpha, \phi, \boldsymbol{\tau} &\quad \sim \quad \mathrm{N}(\mathbf{0}, \mathrm{diag}(\boldsymbol{\tau}^2)/\phi) \\
\tau_1^2 \ldots, \tau_p^2 \mid \alpha, \phi &\quad \textit{i.i.d.} \quad \mathrm{Exp}(\lambda^2/2) \\
p(\alpha, \phi) &\quad \propto \quad 1/\phi
\end{aligned}
$$

Can show that $\beta_j \mid \phi, \lambda \overset{\text{iid}}{\sim} DE(\lambda\sqrt{\phi})$

$$
\int_0^\infty \frac{1}{\sqrt{2\pi s}} e^{-\frac{1}{2}\phi\frac{\beta^2}{s}} \frac{\lambda^2}{2} e^{-\frac{\lambda^2 s}{2}}\, ds = \frac{\lambda\phi^{1/2}}{2} e^{-\lambda\phi^{1/2}|\beta|}
$$

Bayesian Lasso

Park & Casella (JASA 2008) and Hans Biometrika (2010) propose Bayesian versions of the Lasso

$$
\begin{aligned}
\mathbf{Y} \mid \alpha, \boldsymbol{\beta}, \phi &\sim \mathsf{N}(\mathbf{1}_n \alpha + \mathbf{X}^c \boldsymbol{\beta}, \mathbf{I}_n/\phi) \\
\boldsymbol{\beta} \mid \alpha, \phi, \boldsymbol{\tau} &\sim \mathsf{N}(\mathbf{0}, \operatorname{diag}(\boldsymbol{\tau}^2)/\phi) \\
\tau_1^2 \ldots, \tau_p^2 \mid \alpha, \phi &\ \ i.i.d.\ \ \mathsf{Exp}(\lambda^2/2) \\
p(\alpha, \phi) &\propto 1/\phi
\end{aligned}
$$

Can show that $\beta_j \mid \phi, \lambda \stackrel{\text{iid}}{\sim} DE(\lambda\sqrt{\phi})$

$$
\int_0^\infty \frac{1}{\sqrt{2\pi s}} e^{-\frac{1}{2}\phi\frac{\beta^2}{s}} \frac{\lambda^2}{2} e^{-\frac{\lambda^2 s}{2}}\, ds = \frac{\lambda \phi^{1/2}}{2} e^{-\lambda\phi^{1/2}|\beta|}
$$

Scale Mixture of Normals (Andrews and Mallows 1974)

Gibbs Sampling

- Integrate out α: $\alpha \mid \mathbf{Y}, \phi \sim \mathsf{N}(\bar{y}, 1/(n\phi))$

Gibbs Sampling

- Integrate out α: $\alpha \mid \mathbf{Y}, \phi \sim \mathsf{N}(\bar{y}, 1/(n\phi))$
- $\boldsymbol{\beta} \mid \boldsymbol{\tau}, \phi, \lambda, \mathbf{Y} \sim \mathsf{N}(,)$

Gibbs Sampling

- Integrate out α: $\alpha \mid \mathbf{Y}, \phi \sim \mathsf{N}(\bar{y}, 1/(n\phi))$
- $\boldsymbol{\beta} \mid \boldsymbol{\tau}, \phi, \lambda, \mathbf{Y} \sim \mathsf{N}(,)$
- $\phi \mid \boldsymbol{\tau}, \boldsymbol{\beta}, \lambda, \mathbf{Y} \sim \mathbf{G}(,)$

Gibbs Sampling

- Integrate out α: $\alpha \mid \mathbf{Y}, \phi \sim \mathsf{N}(\bar{y}, 1/(n\phi))$
- $\boldsymbol{\beta} \mid \boldsymbol{\tau}, \phi, \lambda, \mathbf{Y} \sim \mathsf{N}(,)$
- $\phi \mid \boldsymbol{\tau}, \boldsymbol{\beta}, \lambda, \mathbf{Y} \sim \mathbf{G}(,)$
- $1/\tau_j^2 \mid \boldsymbol{\beta}, \phi, \lambda, \mathbf{Y} \sim \mathsf{InvGaussian}(,)$

Gibbs Sampling

- Integrate out α: $\alpha \mid \mathbf{Y}, \phi \sim \mathsf{N}(\bar{y}, 1/(n\phi))$
- $\boldsymbol{\beta} \mid \boldsymbol{\tau}, \phi, \lambda, \mathbf{Y} \sim \mathsf{N}(,)$
- $\phi \mid \boldsymbol{\tau}, \boldsymbol{\beta}, \lambda, \mathbf{Y} \sim \mathbf{G}(,)$
- $1/\tau_j^2 \mid \boldsymbol{\beta}, \phi, \lambda, \mathbf{Y} \sim \mathsf{InvGaussian}(,)$

$X \sim \mathsf{InvGaussian}(\mu, \lambda)$

$$f(x) = \sqrt{\frac{\lambda^2}{2\pi}} x^{-3/2} e^{-\frac{1}{2}\frac{\lambda^2 (x-\mu)^2}{\mu^2 x}} \qquad x > 0$$

Gibbs Sampling

- Integrate out α: $\alpha \mid \mathbf{Y}, \phi \sim \mathsf{N}(\bar{y}, 1/(n\phi))$
- $\boldsymbol{\beta} \mid \boldsymbol{\tau}, \phi, \lambda, \mathbf{Y} \sim \mathsf{N}(,)$
- $\phi \mid \boldsymbol{\tau}, \boldsymbol{\beta}, \lambda, \mathbf{Y} \sim \mathbf{G}(,)$
- $1/\tau_j^2 \mid \boldsymbol{\beta}, \phi, \lambda, \mathbf{Y} \sim \mathsf{InvGaussian}(,)$

$X \sim \mathsf{InvGaussian}(\mu, \lambda)$

$$f(x) = \sqrt{\frac{\lambda^2}{2\pi}} x^{-3/2} e^{-\frac{1}{2}\frac{\lambda^2 (x-\mu)^2}{\mu^2 x}} \qquad x > 0$$

Exercise for Energetic Student: Derive the full conditionals for $\boldsymbol{\beta}$, ϕ, $1/\tau^2$ see `http://www.stat.ufl.edu/~casella/Papers/Lasso.pdf`

Other Options

Range of other scale mixtures used

Other Options

Range of other scale mixtures used

- Bridge - Power Exponential Priors

Other Options

Range of other scale mixtures used

- Bridge - Power Exponential Priors
- Horseshoe (Carvalho, Polson & Scott)

Other Options

Range of other scale mixtures used

- Bridge - Power Exponential Priors
- Horseshoe (Carvalho, Polson & Scott)
- Generalized Double Pareto (Armagan, Dunson & Lee)

Other Options

Range of other scale mixtures used

- Bridge - Power Exponential Priors
- Horseshoe (Carvalho, Polson & Scott)
- Generalized Double Pareto (Armagan, Dunson & Lee)
- Normal-Exponenetial-Gamma (Griffen & Brown)

Other Options

Range of other scale mixtures used

- Bridge - Power Exponential Priors
- Horseshoe (Carvalho, Polson & Scott)
- Generalized Double Pareto (Armagan, Dunson & Lee)
- Normal-Exponenetial-Gamma (Griffen & Brown)

Properties of Prior?

Other Options

Range of other scale mixtures used

- Bridge - Power Exponential Priors
- Horseshoe (Carvalho, Polson & Scott)
- Generalized Double Pareto (Armagan, Dunson & Lee)
- Normal-Exponenetial-Gamma (Griffen & Brown)

Properties of Prior?

Selection?

Bayesian Posterior does not assign any probability to $\beta_j = 0$

- selection based on posterior mode ad hoc rule

Selection?

Bayesian Posterior does not assign any probability to $\beta_j = 0$

- selection based on posterior mode ad hoc rule
- Selection solved as a post-analysis decision problem

Selection?

Bayesian Posterior does not assign any probability to $\beta_j = 0$

- selection based on posterior mode ad hoc rule
- Selection solved as a post-analysis decision problem
- Selection part of model uncertainty $\Rightarrow$ add prior

Selection?

Bayesian Posterior does not assign any probability to $\beta_j = 0$

- selection based on posterior mode ad hoc rule
- Selection solved as a post-analysis decision problem
- Selection part of model uncertainty $\Rightarrow$ add prior probability that $\beta_j = 0$ and combine with decision problem