Lasso & Bayesian Lasso
Readings Chapter 15 Christensen

STA721 Linear Models Duke University

Merlise Clyde

October 17, 2017
Lasso

Tibshirani (JRSS B 1996) proposed estimating coefficients through $L_1$ constrained least squares “Least Absolute Shrinkage and Selection Operator”

- Control how large coefficients may grow
Lasso

Tibshirani (JRSS B 1996) proposed estimating coefficients through $L_1$ constrained least squares “Least Absolute Shrinkage and Selection Operator”

- Control how large coefficients may grow

$$\min_{\beta} (Y^c - X^c \beta^c)^T(Y^c - X^c \beta^c)$$

subject to

$$\sum |\beta_j^c| \leq t$$
Lasso

Tibshirani (JRSS B 1996) proposed estimating coefficients through $L_1$ constrained least squares “Least Absolute Shrinkage and Selection Operator”

- Control how large coefficients may grow

$$\min_{\beta} (Y^c - X^c \beta^c)^T (Y^c - X^c \beta^c)$$

subject to

$$\sum |\beta_j^c| \leq t$$

- Equivalent Quadratic Programming Problem for “penalized” Likelihood

$$\min_{\beta^c} \|Y^c - X^c \beta^c\|^2 + \lambda \|\beta^c\|_1$$
Lasso

Tibshirani (JRSS B 1996) proposed estimating coefficients through $L_1$ constrained least squares “Least Absolute Shrinkage and Selection Operator”

- Control how large coefficients may grow

$$
\min_{\beta} (Y^c - X^c \beta^c)^T (Y^c - X^c \beta^c)
$$

subject to

$$
\sum |\beta^c_j| \leq t
$$

- Equivalent Quadratic Programming Problem for “penalized” Likelihood

$$
\min_{\beta^c} \|Y^c - X^c \beta^c\|^2 + \lambda\|\beta^c\|_1
$$

- Posterior mode

$$
\max_{\beta^c} -\frac{\phi}{2}\left\{ \|Y^c - X^c \beta^c\|^2 + \lambda^*\|\beta^c\|_1 \right\}
$$
R Code

The entire path of solutions can be easily found using the “Least Angle Regression” Algorithm of Efron et al (Annals of Statistics 2004)

```r
> library(lars)
> longley.lars = lars(as.matrix(longley[, -7]), longley[, 7],
                     type="lasso")
> plot(longley.lars)
```

![Graph showing the path of solutions using the LASSO method.](duke.eps)
> round(coef(longley.lars),5)

<table>
<thead>
<tr>
<th></th>
<th>GNP.deflator</th>
<th>GNP</th>
<th>Unemployed</th>
<th>Armed.Forces</th>
<th>Population</th>
<th>Year</th>
</tr>
</thead>
<tbody>
<tr>
<td>[1,]</td>
<td>0.00000</td>
<td>0.00000</td>
<td>0.00000</td>
<td>0.00000</td>
<td>0.00000</td>
<td>0.00000</td>
</tr>
<tr>
<td>[2,]</td>
<td>0.00000</td>
<td>0.03273</td>
<td>0.00000</td>
<td>0.00000</td>
<td>0.00000</td>
<td>0.00000</td>
</tr>
<tr>
<td>[3,]</td>
<td>0.00000</td>
<td>0.03623</td>
<td>-0.00372</td>
<td>0.00000</td>
<td>0.00000</td>
<td>0.00000</td>
</tr>
<tr>
<td>[4,]</td>
<td>0.00000</td>
<td>0.03717</td>
<td>-0.00459</td>
<td>-0.00099</td>
<td>0.00000</td>
<td>0.00000</td>
</tr>
<tr>
<td>[5,]</td>
<td>0.00000</td>
<td>0.00000</td>
<td>-0.01242</td>
<td>-0.00539</td>
<td>0.00000</td>
<td>0.90681</td>
</tr>
<tr>
<td>[6,]</td>
<td>0.00000</td>
<td>0.00000</td>
<td>-0.01412</td>
<td>-0.00713</td>
<td>0.00000</td>
<td>0.94375</td>
</tr>
<tr>
<td>[7,]</td>
<td>0.00000</td>
<td>0.00000</td>
<td>-0.01471</td>
<td>-0.00861</td>
<td>-0.15337</td>
<td>1.18430</td>
</tr>
<tr>
<td>[8,]</td>
<td>-0.00770</td>
<td>0.00000</td>
<td>-0.01481</td>
<td>-0.00873</td>
<td>-0.17076</td>
<td>1.22888</td>
</tr>
<tr>
<td>[9,]</td>
<td>0.00000</td>
<td>-0.01212</td>
<td>-0.01663</td>
<td>-0.00927</td>
<td>-0.13029</td>
<td>1.43192</td>
</tr>
<tr>
<td>[10,]</td>
<td>0.00000</td>
<td>-0.02534</td>
<td>-0.01869</td>
<td>-0.00989</td>
<td>-0.09514</td>
<td>1.68655</td>
</tr>
<tr>
<td>[11,]</td>
<td>0.01506</td>
<td>-0.03582</td>
<td>-0.02020</td>
<td>-0.01033</td>
<td>-0.05110</td>
<td>1.82915</td>
</tr>
</tbody>
</table>
Cp Solution

Min $C_p = \frac{SSE_p}{\hat{\sigma}_F^2} - n + 2p$
Cp Solution

\[
\text{Min } C_p = \frac{SSE_p}{\hat{\sigma}_F^2} - n + 2p
\]

> summary(longley.lars)

LARS/LASSO

Call: lars(x = as.matrix(longley[, -7]), y = longley[, 7], type = "lasso")

<table>
<thead>
<tr>
<th>Df</th>
<th>Rss</th>
<th>Cp</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>185.009</td>
<td>1976.7120</td>
</tr>
<tr>
<td>1</td>
<td>6.642</td>
<td>59.4712</td>
</tr>
<tr>
<td>2</td>
<td>3.883</td>
<td>31.7832</td>
</tr>
<tr>
<td>3</td>
<td>3.468</td>
<td>29.3165</td>
</tr>
<tr>
<td>4</td>
<td>1.563</td>
<td>10.8183</td>
</tr>
<tr>
<td>5</td>
<td>1.339</td>
<td>6.4068</td>
</tr>
<tr>
<td>6</td>
<td>1.024</td>
<td>5.0186</td>
</tr>
<tr>
<td>7</td>
<td>0.998</td>
<td>6.7388</td>
</tr>
<tr>
<td>8</td>
<td>0.907</td>
<td>7.7615</td>
</tr>
<tr>
<td>9</td>
<td>0.847</td>
<td>5.1128</td>
</tr>
<tr>
<td>10</td>
<td>0.836</td>
<td>7.0000</td>
</tr>
</tbody>
</table>
Cp Solution

\[ \text{Min } C_p = \frac{SSE_p}{\hat{\sigma}_F^2} - n + 2p \]

> summary(longley.lars)

LARS/LASSO

Call: lars(x = as.matrix(longley[, -7]), y = longley[, 7], type = "lasso")

<table>
<thead>
<tr>
<th>Df</th>
<th>Rss</th>
<th>Cp</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>185.009</td>
<td>1976.7120</td>
</tr>
<tr>
<td>1</td>
<td>6.642</td>
<td>59.4712</td>
</tr>
<tr>
<td>2</td>
<td>3.883</td>
<td>31.7832</td>
</tr>
<tr>
<td>3</td>
<td>3.468</td>
<td>29.3165</td>
</tr>
<tr>
<td>4</td>
<td>1.563</td>
<td>10.8183</td>
</tr>
<tr>
<td>5</td>
<td>1.339</td>
<td>6.4068</td>
</tr>
<tr>
<td>6</td>
<td>1.024</td>
<td>5.0186</td>
</tr>
<tr>
<td>7</td>
<td>0.998</td>
<td>6.7388</td>
</tr>
<tr>
<td>8</td>
<td>0.907</td>
<td>7.7615</td>
</tr>
<tr>
<td>9</td>
<td>0.847</td>
<td>5.1128</td>
</tr>
<tr>
<td>10</td>
<td>0.836</td>
<td>7.0000</td>
</tr>
</tbody>
</table>

GNP.deflator  GNP  Unemployed  Armed.Forces  Population  Year
[7,] 0.00000 0.00000 -0.01471 -0.00861 -0.15337 1.18430
Features

Combines shrinkage (like Ridge Regression) with Selection (like stepwise selection)
Features

Combines shrinkage (like Ridge Regression) with Selection (like stepwise selection)

Uncertainty? Interval estimates?
Bayesian Lasso

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

\[
\begin{align*}
Y | \alpha, \beta, \phi & \sim \mathcal{N} \left( \frac{1}{n} \alpha + X \beta, \frac{1}{\phi} \right) \\
\beta | \alpha, \phi, \tau & \sim \mathcal{N} \left( 0, \frac{\text{diag}(\tau^2)}{\phi} \right) \\
\tau_1^2, \tau_2, \ldots, \tau_p | \alpha, \phi & \text{iid} \sim \text{Exp} \left( \frac{\lambda^2}{2} \right) \\
p(\alpha, \phi) & \propto \frac{1}{\phi}
\end{align*}
\]

Can show that \( \beta_j | \phi, \lambda \text{iid} \sim \text{DE} \left( \lambda \sqrt{\phi}, 2 \right) \)

\[
\int_{0}^{\infty} e^{-s/2} \left( \frac{1}{\lambda^2} \sqrt{\phi} \right)^{2} e^{-\lambda^2 s/2} ds = \frac{\lambda \sqrt{\phi}}{2} \]
Bayesian Lasso

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

\[ Y \mid \alpha, \beta, \phi \sim N(1_n \alpha + X^c \beta, I_n / \phi) \]
Bayesian Lasso

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

\[
Y \mid \alpha, \beta, \phi \sim N(1_n\alpha + X^c\beta, I_n/\phi)
\]
\[
\beta \mid \alpha, \phi, \tau \sim N(0, \text{diag}(\tau^2)/\phi)
\]
Bayesian Lasso

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

\[
\mathbf{Y} \mid \alpha, \beta, \phi \sim \mathcal{N}(\mathbf{1}_n\alpha + \mathbf{X}^c\beta, \mathbf{I}_n/\phi)
\]

\[
\beta \mid \alpha, \phi, \tau \sim \mathcal{N}(\mathbf{0}, \text{diag}(\tau^2)/\phi)
\]

\[
\tau_1^2, \ldots, \tau_p^2 \mid \alpha, \phi \overset{iid}{\sim} \text{Exp}(\lambda^2/2)
\]
Bayesian Lasso

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

\[
\begin{align*}
\mathbf{Y} \mid \alpha, \beta, \phi & \sim N(\mathbf{1}_n \alpha + \mathbf{X}^c \beta, \mathbf{I}_n / \phi) \\
\beta \mid \alpha, \phi, \tau & \sim N(\mathbf{0}, \text{diag}(\tau^2) / \phi) \\
\tau_1^2, \ldots, \tau_p^2 \mid \alpha, \phi & \text{iid} \sim \text{Exp}(\lambda^2 / 2) \\
p(\alpha, \phi) & \propto 1 / \phi
\end{align*}
\]
Bayesian Lasso

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

\[ \mathbf{Y} \mid \alpha, \beta, \phi \sim \mathcal{N}(\mathbf{1}_n \alpha + \mathbf{X}^c \beta, \mathbf{I}_n / \phi) \]
\[ \beta \mid \alpha, \phi, \tau \sim \mathcal{N}(\mathbf{0}, \text{diag}(\tau^2) / \phi) \]
\[ \tau^2_1, \ldots, \tau^2_p \mid \alpha, \phi \overset{iid}{\sim} \exp(\lambda^2 / 2) \]
\[ p(\alpha, \phi) \propto 1 / \phi \]

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

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

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

\[ \mathbf{Y} \mid \alpha, \beta, \phi \sim \mathcal{N}(\mathbf{1}_n \alpha + X^c \beta, \mathbf{I}_n/\phi) \]

\[ \beta \mid \alpha, \phi, \tau \sim \mathcal{N}(\mathbf{0}, \text{diag}(\tau^2)/\phi) \]

\[ \tau_1^2 \ldots, \tau_p^2 \mid \alpha, \phi \overset{iid}{\sim} \text{Exp}(\lambda^2/2) \]

\[ p(\alpha, \phi) \propto 1/\phi \]

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

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

Scale Mixture of Normals (Andrews and Mallows 1974)
Gibbs Sampling

- Integrate out $\alpha$: $\alpha | Y, \phi \sim N(\bar{y}, 1/(n\phi))$
Gibbs Sampling

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

Homework: Derive the full conditionals for $\beta, \phi, 1/\tau$.

See http://www.stat.ufl.edu/~casella/Papers/Lasso.pdf
Gibbs Sampling

- Integrate out $\alpha$: $\alpha | Y, \phi \sim \text{N}(\bar{y}, 1/(n\phi))$
- $\beta | \tau, \phi, \lambda, Y \sim \text{N}(, )$
- $\phi | \tau, \beta, \lambda, Y \sim \text{G}(, )$

Homework: Derive the full conditionals for $\beta$, $\phi$, $1/\tau^2$ see http://www.stat.ufl.edu/~casella/Papers/Lasso.pdf
Gibbs Sampling

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

- $\beta \mid \tau, \phi, \lambda, Y \sim N(, )$

- $\phi \mid \tau, \beta, \lambda, Y \sim G(, )$

- $1/\tau^2_j \mid \beta, \phi, \lambda, Y \sim \text{InvGaussian}(, )$
Gibbs Sampling

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

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

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

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

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

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

Homework: Derive the full conditionals for $\beta, \phi, 1/\tau^2$ 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
  ▶ Horseshoe (Carvalho, Polson & Scott)
Other Options

Range of other scale mixtures used

- Horseshoe (Carvalho, Polson & Scott)
- Generalized Double Pareto (Armagan, Dunson & Lee)
Other Options

Range of other scale mixtures used

- Horseshoe (Carvalho, Polson & Scott)
- Generalized Double Pareto (Armagan, Dunson & Lee)
- Normal-Exponential-Gamma (Griffen & Brown)
Other Options

Range of other scale mixtures used

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

Range of other scale mixtures used

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

Properties of Prior?
Other Options

Range of other scale mixtures used

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

Properties of Prior?
Horseshoe

Carvalho, Polson & Scott propose

- Prior Distribution on

\[ \beta | \phi \sim N(0_p, \frac{\text{diag}(\tau^2)}{\phi}) \]
Horseshoe

Carvalho, Polson & Scott propose

- Prior Distribution on

\[ \beta \mid \phi \sim N(0_p, \frac{\text{diag}(\tau^2)}{\phi}) \]

- \( \tau^2_j \mid \lambda \text{ iid } \sim C^+(0, \lambda) \)

where \( \kappa_i = \frac{1}{1 + \tau^2_i} \) shrinkage factor

Half-Cauchy prior induces a Beta(1/2, 1/2) distribution on \( \kappa \) a priori
Horseshoe

Carvalho, Polson & Scott propose

- Prior Distribution on

\[ \beta \mid \phi \sim N(0_p, \frac{\text{diag}(\tau^2)}{\phi}) \]

- \[ \tau_j^2 \mid \lambda \overset{\text{iid}}{\sim} C^+(0, \lambda) \]

- \[ \lambda \sim C^+(0, 1/\phi) \]
Horseshoe

Carvalho, Polson & Scott propose

- Prior Distribution on

\[
\beta \mid \phi \sim N(0_p, \frac{\text{diag}(\tau^2)}{\phi})
\]

- \( \tau_j^2 \mid \lambda \sim C^+(0, \lambda) \)
- \( \lambda \sim C^+(0, 1/\phi) \)
- \( p(\alpha, \phi) \propto 1/\phi \)

In the case \( \lambda = 1/\phi = 1 \) and with \( X_t X_t = I \), \( Y^* = X^T Y \)

\[
E[\beta_i \mid Y] = \int 1_{1-\kappa_i y_i^*} p(\kappa_i \mid Y) d\kappa_i = (1 - E[\kappa_i \mid y_i^*]) y_i^*
\]

where \( \kappa_i = 1/(1 + \tau_i^2) \) shrinkage factor

Half-Cauchy prior induces a Beta(1/2, 1/2) distribution on \( \kappa \) \text{a priori}
Horseshoe

Carvalho, Polson & Scott propose

- Prior Distribution on

\[
\beta \mid \phi \sim N(0_p, \frac{\text{diag}(\tau^2)}{\phi})
\]

- \( \tau_j^2 \mid \lambda \overset{\text{iid}}{\sim} C^+(0, \lambda) \)
- \( \lambda \sim C^+(0, 1/\phi) \)
- \( p(\alpha, \phi) \propto 1/\phi \)

In the case \( \lambda = \phi = 1 \) and with \( X^tX = I \), \( Y^* = X^T Y \)
Horseshoe

Carvalho, Polson & Scott propose

- Prior Distribution on

\[
\beta \mid \phi \sim N(0_p, \frac{\text{diag}(\tau^2)}{\phi})
\]

- \( \tau^2_j \mid \lambda \overset{\text{iid}}{\sim} C^+(0, \lambda) \)
- \( \lambda \sim C^+(0, 1/\phi) \)
- \( p(\alpha, \phi) \propto 1/\phi \)

In the case \( \lambda = \phi = 1 \) and with \( X^tX = I, Y^* = X^T Y \)

\[
E[\beta_i \mid Y] = \int_0^1 (1 - \kappa_i)y_i^* p(\kappa_i \mid Y) \ d\kappa_i = (1 - E[\kappa \mid y_i^*])y_i^*
\]

where \( \kappa_i = 1/(1 + \tau_i^2) \) shrinkage factor
Horseshoe

Carvalho, Polson & Scott propose

- Prior Distribution on

\[ \beta | \phi \sim N(0_p, \frac{\text{diag}(\tau^2)}{\phi}) \]

- \( \tau_j^2 | \lambda \overset{\text{iid}}{\sim} C^+(0, \lambda) \)
- \( \lambda \sim C^+(0, 1/\phi) \)
- \( p(\alpha, \phi) \propto 1/\phi \)

In the case \( \lambda = \phi = 1 \) and with \( X^tX = I, Y^* = X^T Y \)

\[ E[\beta_i | Y] = \int_0^1 (1 - \kappa_i)y_i^* p(\kappa_i | Y) d\kappa_i = (1 - E[\kappa | y_i^*])y_i^* \]

where \( \kappa_i = 1/(1 + \tau_i^2) \) shrinkage factor

Half-Cauchy prior induces a Beta(1/2, 1/2) distribution on \( \kappa_i \) a priori
Horseshoe

Beta\(1/2, 1/2\)
Simulation Study with Diabetes Data

![Box plot comparing OLS, LASSO, and HORSESHOE methods for RMSE]

- OLS
- LASSO
- HORSESHOE
Other Options

Range of other scale mixtures used

Generalized Double Pareto (Armagan, Dunson & Lee)

\[ \lambda \sim \Gamma(\alpha, \eta) \] then

\[ \beta_j \sim \text{GDP}(\xi = \eta/\alpha, \alpha) \]

\[ f(\beta_j) = \frac{1}{2\xi} (1 + |\beta_j|^\xi\alpha)^{-\frac{1}{\alpha} - 1} \]


Normal-Exponental-Gamma (Griffen & Brown 2005)

\[ \lambda_2 \sim \Gamma(\alpha, \eta) \]

Bridge - Power Exponential Priors (Stable mixing density)

See the monomvn package on CRAN

Choice of prior? Properties? Fan & Li (JASA 2001) discuss

Variable selection via nonconcave penalties and oracle properties
Other Options

Range of other scale mixtures used

- Generalized Double Pareto (Armagan, Dunson & Lee)

See the monomvn package on CRAN

Choice of prior? Properties? Fan & Li (JASA 2001) discuss

Variable selection via nonconcave penalties and oracle properties

Other Options

Range of other scale mixtures used

- Generalized Double Pareto (Armagan, Dunson & Lee)
  \( \lambda \sim \text{Gamma}(\alpha, \eta) \) then \( \beta_j \sim \text{GDP}(\xi = \eta/\alpha, \alpha) \)


- Normal-Exponental-Gamma (Griffen & Brown 2005)
- Bridge - Power Exponential Priors (Stable mixing density)

See the monomvn package on CRAN

Choice of prior? Properties? Fan & Li (JASA 2001) discuss

Variable selection via nonconcave penalties and oracle properties
Other Options

Range of other scale mixtures used

- Generalized Double Pareto (Armagan, Dunson & Lee)
  \( \lambda \sim \text{Gamma}(\alpha, \eta) \) then \( \beta_j \sim \text{GDP}(\xi = \eta/\alpha, \alpha) \)

  \[
  f(\beta_j) = \frac{1}{2\xi} \left( 1 + \frac{|\beta_j|}{\xi \alpha} \right)^{-(1+\alpha)}
  \]

Other Options

Range of other scale mixtures used

- Generalized Double Pareto (Armagan, Dunson & Lee)
  \( \lambda \sim \text{Gamma}(\alpha, \eta) \) then \( \beta_j \sim \text{GDP}(\xi = \eta/\alpha, \alpha) \)

  \[
  f(\beta_j) = \frac{1}{2\xi} (1 + \frac{|\beta_j|}{\xi\alpha})^{-(1+\alpha)}
  \]


- Normal-Exponential-Gamma (Griffen & Brown 2005)
  \( \lambda^2 \sim \text{Gamma}(\alpha, \eta) \)
Other Options

Range of other scale mixtures used

- Generalized Double Pareto (Armagan, Dunson & Lee)
  
  \( \lambda \sim \text{Gamma}(\alpha, \eta) \) then \( \beta_j \sim \text{GDP}(\xi = \eta/\alpha, \alpha) \)

  \[
  f(\beta_j) = \frac{1}{2\xi} (1 + \frac{|\beta_j|}{\xi\alpha})^{-(1+\alpha)}
  \]


- Normal-Exponenetial-Gamma (Griffen & Brown 2005)

  \( \lambda^2 \sim \text{Gamma}(\alpha, \eta) \)

- Bridge - Power Exponential Priors (Stable mixing density)
Other Options

Range of other scale mixtures used

- Generalized Double Pareto (Armagan, Dunson & Lee)
  \[ \lambda \sim \text{Gamma}(\alpha, \eta) \text{ then } \beta_j \sim \text{GDP}(\xi = \eta/\alpha, \alpha) \]
  \[
f(\beta_j) = \frac{1}{2\xi}(1 + \frac{|\beta_j|}{\xi \alpha})^{-(1+\alpha)}
\]

- Normal-Exponential-Gamma (Griffen & Brown 2005)
  \[ \lambda^2 \sim \text{Gamma}(\alpha, \eta) \]

- Bridge - Power Exponential Priors (Stable mixing density)
  See the monomvn package on CRAN
Other Options

Range of other scale mixtures used

- Generalized Double Pareto (Armagan, Dunson & Lee)
  \( \lambda \sim \text{Gamma}(\alpha, \eta) \) then \( \beta_j \sim \text{GDP}(\xi = \eta/\alpha, \alpha) \)

\[
f(\beta_j) = \frac{1}{2\xi} \left(1 + \frac{|\beta_j|}{\xi \alpha}\right)^{-1+\alpha}
\]


- Normal-Exponential-Gamma (Griffen & Brown 2005)
  \( \lambda^2 \sim \text{Gamma}(\alpha, \eta) \)

- Bridge - Power Exponential Priors (Stable mixing density)

See the monomvn package on CRAN

Choice of prior? Properties? Fan & Li (JASA 2001) discuss
Variable selection via nonconcave penalties and oracle properties
Choice of Estimator & Selection?

- Posterior Mode (may set some coefficients to zero)
Choice of Estimator & Selection?

- Posterior Mode (may set some coefficients to zero)
- Posterior Mean (no selection)
Choice of Estimator & Selection?

- Posterior Mode (may set some coefficients to zero)
- Posterior Mean (no selection)

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

Selection solved as a post-analysis decision problem

Selection part of model uncertainty ⇒ add prior probability that \( \beta_j = 0 \) and combine with decision problem

See article by Datta & Ghosh [http://ba.stat.cmu.edu/journal/forthcoming/datta.pdf]
Choice of Estimator & Selection?

▶ Posterior Mode (may set some coefficients to zero)
▶ Posterior Mean (no selection)

Bayesian Posterior does not assign any probability to $\beta_j = 0$
▶ selection based on posterior mode ad hoc rule - Select if $\kappa_i < .5$)
Choice of Estimator & Selection?

- Posterior Mode (may set some coefficients to zero)
- Posterior Mean (no selection)

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

- selection based on posterior mode ad hoc rule - Select if $\kappa_i < .5$

See article by Datta & Ghosh http://ba.stat.cmu.edu/journal/forthcoming/datta.pdf
Choice of Estimator & Selection?

- Posterior Mode (may set some coefficients to zero)
- Posterior Mean (no selection)

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

- Selection based on posterior mode ad hoc rule - Select if $\kappa_i < .5$
  
  See article by Datta & Ghosh [http://ba.stat.cmu.edu/journal/forthcoming/datta.pdf](http://ba.stat.cmu.edu/journal/forthcoming/datta.pdf)

- Selection solved as a post-analysis decision problem
Choice of Estimator & Selection?

- Posterior Mode (may set some coefficients to zero)
- Posterior Mean (no selection)

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

- selection based on posterior mode ad hoc rule - Select if $\kappa_i < .5$
  
  See article by Datta & Ghosh [link]

- Selection solved as a post-analysis decision problem
- Selection part of model uncertainty $\Rightarrow$ add prior
Choice of Estimator & Selection?

- Posterior Mode (may set some coefficients to zero)
- Posterior Mean (no selection)

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

- selection based on posterior mode ad hoc rule - Select if $\kappa_i < .5$
  See article by Datta & Ghosh [http://ba.stat.cmu.edu/journal/forthcoming/datta.pdf](http://ba.stat.cmu.edu/journal/forthcoming/datta.pdf)

- 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