Lasso & Bayesian Lasso

Readings Chapter 15 Christensen

STA721 Linear Models Duke University

Merlise Clyde

October 23, 2019

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

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

subject to

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

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

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

  subject to

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

- Equivalent Quadratic Programming Problem for "penalized" Likelihood

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

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

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

subject to

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

- Equivalent Quadratic Programming Problem for "penalized" Likelihood

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

- Posterior mode

$$\max_{\boldsymbol{\beta}^s} -\frac{\phi}{2} \{\|\mathbf{Y}^c - \mathbf{X}^s \boldsymbol{\beta}^s\|^2 + \lambda^* \|\boldsymbol{\beta}^s\|_1\}$$

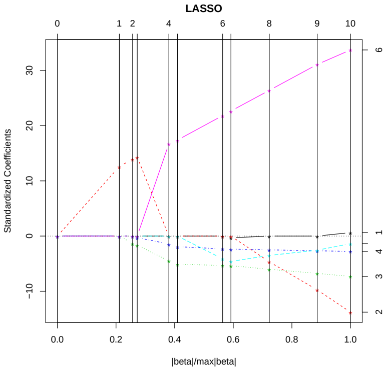
Picture

duke.eps

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)

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



Solutions

```
> round(coef(longley.lars),5)
      GNP.deflator      GNP Unemployed Armed.Forces Population    Year
 [1,]      0.00000  0.00000    0.00000      0.00000    0.00000 0.00000
 [2,]      0.00000  0.03273    0.00000      0.00000    0.00000 0.00000
 [3,]      0.00000  0.03623   -0.00372      0.00000    0.00000 0.00000
 [4,]      0.00000  0.03717   -0.00459     -0.00099    0.00000 0.00000
 [5,]      0.00000  0.00000   -0.01242     -0.00539    0.00000 0.90681
 [6,]      0.00000  0.00000   -0.01412     -0.00713    0.00000 0.94375
 [7,]      0.00000  0.00000   -0.01471     -0.00861   -0.15337 1.18430
 [8,]     -0.00770  0.00000   -0.01481     -0.00873   -0.17076 1.22888
 [9,]      0.00000 -0.01212   -0.01663     -0.00927   -0.13029 1.43192
[10,]      0.00000 -0.02534   -0.01869     -0.00989   -0.09514 1.68655
[11,]      0.01506 -0.03582   -0.02020     -0.01033   -0.05110 1.82915
```

Cp Solution

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

Cp Solution

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

```
> summary(longley.lars)
LARS/LASSO
Call: lars(x = as.matrix(longley[, -7]), y = longley[, 7], type
   Df     Rss        Cp
0   1 185.009 1976.7120
1   2   6.642   59.4712
2   3   3.883   31.7832
3   4   3.468   29.3165
4   5   1.563   10.8183
5   4   1.339    6.4068
6   5   1.024    5.0186
7   6   0.998    6.7388
8   7   0.907    7.7615
9   6   0.847    5.1128
10  7   0.836    7.0000
```

Cp Solution

$$\text{Min } C_p = SSE_p/\hat{\sigma}^2_F - n + 2p$$

```
> summary(longley.lars)
LARS/LASSO
Call: lars(x = as.matrix(longley[, -7]), y = longley[, 7], type
   Df     Rss        Cp
0   1 185.009 1976.7120
1   2   6.642   59.4712
2   3   3.883   31.7832
3   4   3.468   29.3165
4   5   1.563   10.8183
5   4   1.339    6.4068
6   5   1.024    5.0186
7   6   0.998    6.7388
8   7   0.907    7.7615
9   6   0.847    5.1128
10  7   0.836    7.0000

     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 in penalty?

Features

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

Uncertainty in penalty?

Interval estimates?

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

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

Bayesian Lasso

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

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

Bayesian Lasso

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

$$
\begin{aligned}
\mathbf{Y} \mid \alpha, \boldsymbol{\beta}^s, \phi &\sim \mathsf{N}(\mathbf{1}_n\alpha + \mathbf{X}^s\boldsymbol{\beta}^s, \mathbf{I}_n/\phi) \\
\boldsymbol{\beta}^s \mid \alpha, \phi, \boldsymbol{\tau} &\sim \mathsf{N}(\mathbf{0}, \mathrm{diag}(\boldsymbol{\tau}^2)/\phi) \\
\tau_1^2 \ldots, \tau_p^2 \mid \alpha, \phi &\overset{\text{iid}}{\sim} \mathsf{Exp}(\lambda^2/2)
\end{aligned}
$$

Bayesian Lasso

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

$$
\begin{aligned}
\mathbf{Y} \mid \alpha, \boldsymbol{\beta}^s, \phi &\sim \mathsf{N}(\mathbf{1}_n \alpha + \mathbf{X}^s \boldsymbol{\beta}^s, \mathbf{I}_n/\phi) \\
\boldsymbol{\beta}^s \mid \alpha, \phi, \boldsymbol{\tau} &\sim \mathsf{N}(\mathbf{0}, \mathrm{diag}(\boldsymbol{\tau}^2)/\phi) \\
\tau_1^2 \ldots, \tau_p^2 \mid \alpha, \phi &\overset{\text{iid}}{\sim} \mathsf{Exp}(\lambda^2/2) \\
p(\alpha, \phi) &\propto 1/\phi
\end{aligned}
$$

Bayesian Lasso

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

$$\begin{aligned}
\mathbf{Y} \mid \alpha, \boldsymbol{\beta}^s, \phi &\sim \mathsf{N}(\mathbf{1}_n\alpha + \mathbf{X}^s\boldsymbol{\beta}^s, \mathbf{I}_n/\phi) \\
\boldsymbol{\beta}^s \mid \alpha, \phi, \boldsymbol{\tau} &\sim \mathsf{N}(\mathbf{0}, \mathrm{diag}(\boldsymbol{\tau}^2)/\phi) \\
\tau_1^2 \ldots, \tau_p^2 \mid \alpha, \phi &\overset{\text{iid}}{\sim} \mathsf{Exp}(\lambda^2/2) \\
p(\alpha, \phi) &\propto 1/\phi
\end{aligned}$$

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

$$
\begin{aligned}
\mathbf{Y} \mid \alpha, \boldsymbol{\beta}^s, \phi & \sim \mathsf{N}(\mathbf{1}_n \alpha + \mathbf{X}^s \boldsymbol{\beta}^s, \mathbf{I}_n/\phi) \\
\boldsymbol{\beta}^s \mid \alpha, \phi, \boldsymbol{\tau} & \sim \mathsf{N}(\mathbf{0}, \mathrm{diag}(\boldsymbol{\tau}^2)/\phi) \\
\tau_1^2 \ldots, \tau_p^2 \mid \alpha, \phi & \stackrel{\text{iid}}{\sim} \mathsf{Exp}(\lambda^2/2) \\
p(\alpha, \phi) & \propto 1/\phi
\end{aligned}
$$

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

Gibbs Sampling

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

Gibbs Sampling

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

Gibbs Sampling

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

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

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

Gibbs Sampling

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

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

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

Homework: Derive the full conditionals for $\boldsymbol{\beta}^s$, 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

- 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-Exponenetial-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-Exponenetial-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-Exponenetial-Gamma (Griffen & Brown)
- Bridge - Power Exponential Priors

Properties of Prior?

Horseshoe

Carvalho, Polson & Scott propose

- Prior Distribution on

$$\boldsymbol{\beta} \mid \phi \sim \mathsf{N}(\mathbf{0}_p, \frac{\text{diag}(\tau^2)}{\phi})$$

Horseshoe

Carvalho, Polson & Scott propose

- Prior Distribution on

$$\boldsymbol{\beta} \mid \phi \sim \mathsf{N}(\mathbf{0}_p, \frac{\text{diag}(\tau^2)}{\phi})$$

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

Horseshoe

Carvalho, Polson & Scott propose

- Prior Distribution on

$$\boldsymbol{\beta} \mid \phi \sim \mathsf{N}(\mathbf{0}_p, \frac{\mathrm{diag}(\tau^2)}{\phi})$$

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

Horseshoe

Carvalho, Polson & Scott propose

- Prior Distribution on

$$\boldsymbol{\beta} \mid \phi \sim \mathsf{N}(\mathbf{0}_p, \frac{\mathrm{diag}(\tau^2)}{\phi})$$

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

Horseshoe

Carvalho, Polson & Scott propose

- Prior Distribution on

$$\boldsymbol{\beta} \mid \phi \sim \mathsf{N}(\mathbf{0}_p, \frac{\text{diag}(\tau^2)}{\phi})$$

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

In the case $\lambda = \phi = 1$ and with $\mathbf{X}^t\mathbf{X} = \mathbf{I}$, $\mathbf{Y}^* = \mathbf{X}^T\mathbf{Y}$

Horseshoe

Carvalho, Polson & Scott propose

- Prior Distribution on

$$\boldsymbol{\beta} \mid \phi \sim \mathsf{N}(\mathbf{0}_p, \frac{\mathrm{diag}(\tau^2)}{\phi})$$

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

In the case $\lambda = \phi = 1$ and with $\mathbf{X}^t\mathbf{X} = \mathbf{I}$, $\mathbf{Y}^* = \mathbf{X}^T\mathbf{Y}$

$$E[\beta_i \mid \mathbf{Y}] = \int_0^1 (1 - \kappa_i) y_i^* p(\kappa_i \mid \mathbf{Y}) \, d\kappa_i = (1 - \mathsf{E}[\kappa \mid y_i^*]) y_i^*$$

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

Horseshoe

Carvalho, Polson & Scott propose

- Prior Distribution on

$$\boldsymbol{\beta} \mid \phi \sim \mathrm{N}(\mathbf{0}_p, \frac{\mathrm{diag}(\tau^2)}{\phi})$$

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

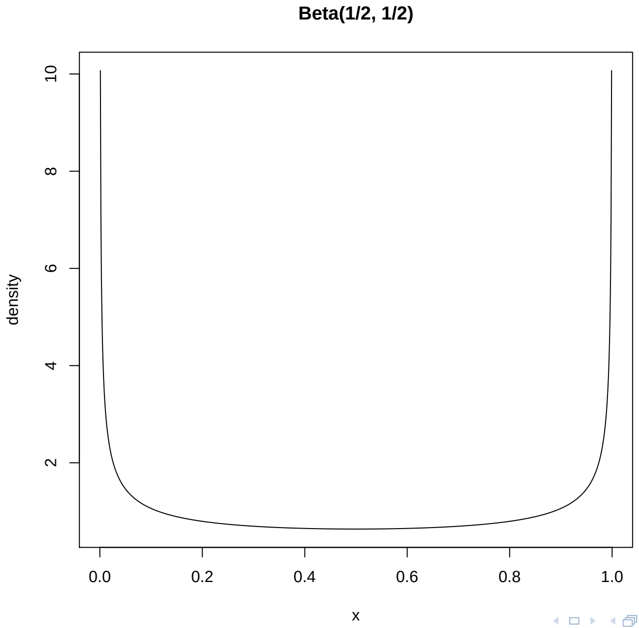
In the case $\lambda = \phi = 1$ and with $\mathbf{X}^t\mathbf{X} = \mathbf{I}$, $\mathbf{Y}^* = \mathbf{X}^T\mathbf{Y}$

$$E[\beta_i \mid \mathbf{Y}] = \int_0^1 (1 - \kappa_i) y_i^* p(\kappa_i \mid \mathbf{Y}) \, d\kappa_i = (1 - \mathrm{E}[\kappa \mid y_i^*]) y_i^*$$

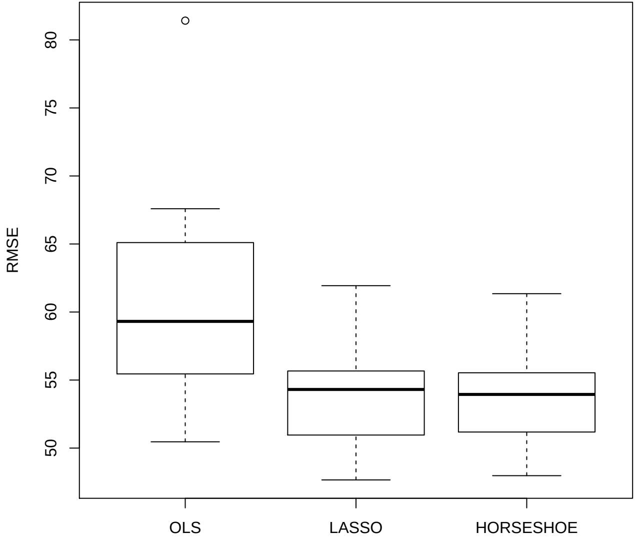
where $\kappa_i = 1/(1 + \tau_i^2)$ shrinkage factor

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

Horseshoe



Simulation Study with Diabetes Data



Other Options

Range of other scale mixtures used

Other Options

Range of other scale mixtures used

- Generalized Double Pareto (Armagan, Dunson & Lee)

Other Options

Range of other scale mixtures used

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

Other Options

Range of other scale mixtures used

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

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

see `http://arxiv.org/pdf/1104.0861.pdf`

Other Options

Range of other scale mixtures used

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

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

  see `http://arxiv.org/pdf/1104.0861.pdf`
- Normal-Exponenetial-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)
  $\lambda \sim$ Gamma(α, η) then $\beta_j \sim$ GDP$(\xi = \eta/\alpha, \alpha)$

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

  see `http://arxiv.org/pdf/1104.0861.pdf`
- Normal-Exponenetial-Gamma (Griffen & Brown 2005)
  $\lambda^2 \sim$ Gamma(α, η)
- Bridge - Power Exponential Priors (Stable mixing density)

Other Options

Range of other scale mixtures used

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

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

  see `http://arxiv.org/pdf/1104.0861.pdf`
- Normal-Exponenetial-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 $\beta_j \sim \text{GDP}(\xi = \eta/\alpha, \alpha)$

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

  see `http://arxiv.org/pdf/1104.0861.pdf`
- Normal-Exponenetial-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 $\beta_j = 0$

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`
- 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 `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

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`
- 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