Horseshoe, Lasso and Related Shrinkage Methods

Readings Chapter 15 Christensen

STA721 Linear Models Duke University

Merlise Clyde

October 14, 2015

Bayesian Lasso

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

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

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

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

Scale Mixture of Normals (Andrews and Mallows 1974)

Gibbs Sampling

Prior $\lambda^2 \sim \text{Gamma}(r, \delta)$ Integrate out α: $\alpha \mid \mathbf{Y}, \phi \sim \text{N}(\bar{y}, 1/(n\phi)$
Full Conditionals

- $\boldsymbol{\beta}^s \mid \boldsymbol{\tau}, \phi, \lambda, \mathbf{Y} \sim \text{N}(,)$
- $\phi \mid \boldsymbol{\tau}, \boldsymbol{\beta}^s, \lambda, \mathbf{Y} \sim \mathbf{G}(,)$
- $\lambda^2 \mid \boldsymbol{\beta}^s, \phi, \tau^2, \mathbf{Y} \sim \mathbf{G}(,)$
- $1/\tau_j^2 \mid \boldsymbol{\beta}^s, \phi, \lambda, \mathbf{Y} \sim \text{InvGaussian}(,)$

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

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

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

Horseshoe

Carvalho, Polson & Scott propose

- Prior Distribution on

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

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

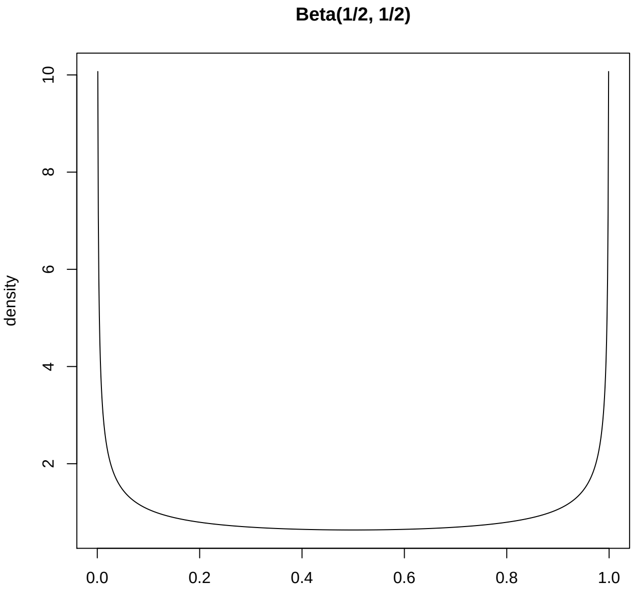
In the case $\lambda = \phi = 1$ and with canonical representation $\mathbf{Y} = \mathbf{I}\boldsymbol{\beta} + \boldsymbol{\epsilon}$

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

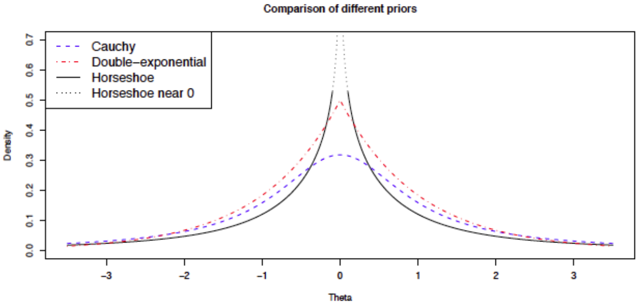
where $\kappa_i = 1/(1 + \tau_i^2)$ shrinkage factor

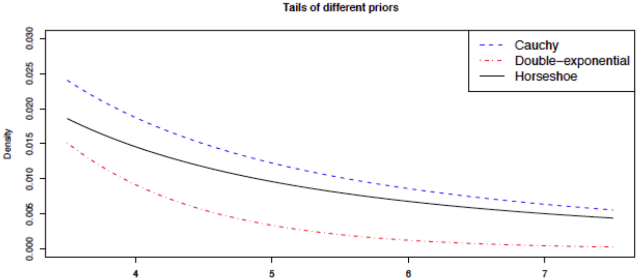
Half-Cauchy prior induces a Beta(1/2, 1/2) distribution on κ_i a priori

Horseshoe



Prior Comparison (from PSC)





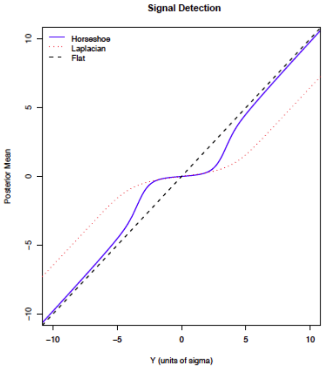
Bounded Influence

Normal means case $Y_i \overset{\text{iid}}{\sim} N(\beta_i, 1)$ (Equivalent to Canonical case)

- Posterior mean $E[\beta \mid y] = y + \frac{d}{dy} \log m(y)$ where $m(y)$ is the predictive denisty under the prior (known λ)
- HS has Bounded Influence:

$$\lim_{|y| \to \infty} \frac{d}{dy} \log m(y) = 0$$

- $\lim_{|y| \to \infty} E[\beta \mid y) \to y$ (MLE)
- DE is also bounded influence, but bound does not decay to zero in tails



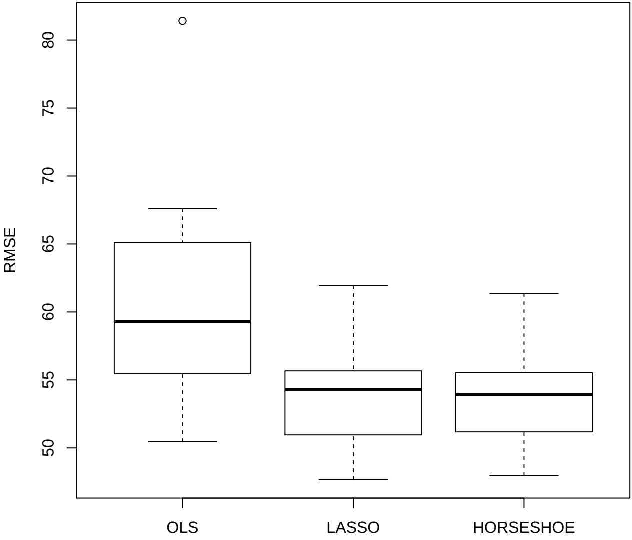
R packages

The `monomvn` package in R includes

- blasso
- bhs

See Diabetes.R code

Simulation Study with Diabetes Data



Other Options

Range of other scale mixtures used

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

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

  see http://arxiv.org/pdf/1104.0861.pdf
- Normal-Exponenetial-Gamma (Griffen & Brown 2005)
  $\lambda^2 \sim \mathsf{Gamma}(\alpha, \eta)$
- 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)
- Posterior Mean (no selection, just shrinkage)

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

- 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 $\beta_j^s = 0$ and combine with decision problem

Remember all models are wrong, but some may be useful!