Horseshoe, Lasso and Related Methods

Readings Chapter 14 Christensen

STA721 Linear Models Duke University

Merlise Clyde

September 28, 2014

Model for Bayesian Lasso

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

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

- $\tau^2 \mid \lambda \overset{\text{iid}}{\sim} \mathrm{Exp}(\lambda^2/2)$
- $p(\alpha) \propto 1$
- $\phi \sim \mathrm{Gamma}(a, b)$
- $\lambda^2 \sim \mathrm{Gamma}(r, \delta)$

Park & Casella (2008), Hans (2009) and earlier Fernandez and Steel (2000)!

Gibbs Sampling

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

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

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

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

Horseshoe

Carvalho, Polson & Scott propose

- Prior Distribution on

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

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

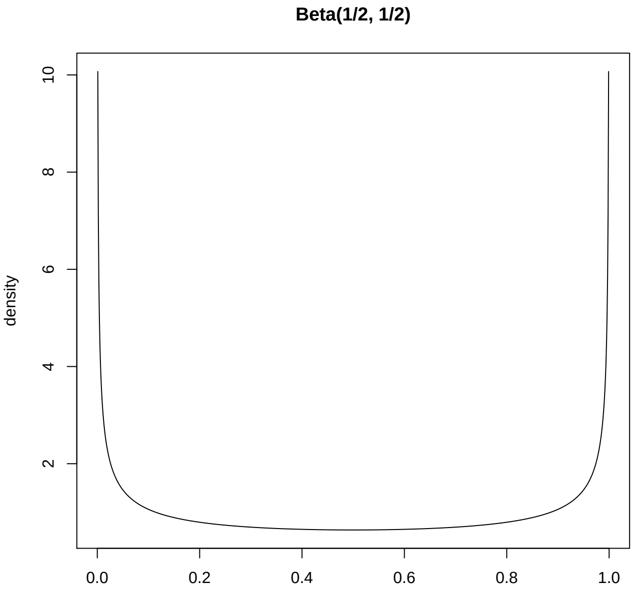
In the case $\lambda = \phi = 1$ and with $\mathbf{X}^t\mathbf{X} = \mathbf{I}$ $\mathbf{Y}^* = \mathbf{X}^T\mathbf{Y}$

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

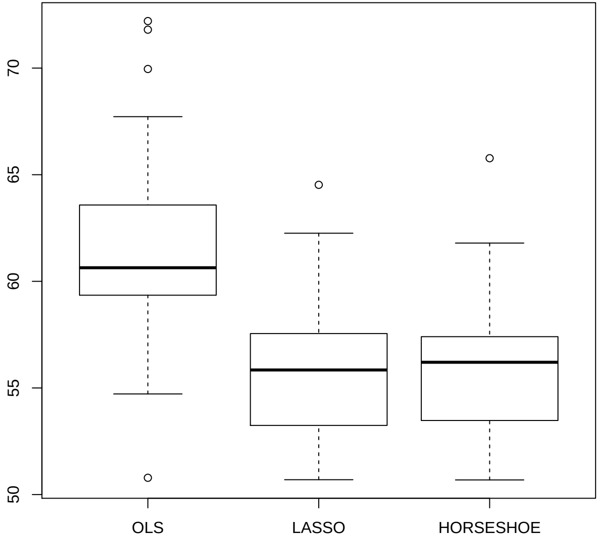
where $\kappa_i = 1/(1 + \tau_i^2)$ shrinkage factor

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

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

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

  see `http://arxiv.org/pdf/1104.0861.pdf`
- Normal-Exponenetial-Gamma (Griffen & Brown 2005)
  $\lambda^2 \sim \mathsf{Gamma}(\alpha, \eta)$
- 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)

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