Efficient sampling for Gaussian linear regression with arbitrary priors

P. Richard Hahn (Arizona State), Jingyu He (Chicago Booth),
Hedibert Lopes (INSPER)

November 21, 2018

Motivation

Goal: run a Gaussian linear regression and a new prior

- Do math, try to write a new Gibbs sampler.
- Maybe hard to find easy conditional distributions.
- Probably require data augmentation, add lots of latent variables.

Why not take advantage of the **Gaussian** likelihood?

Elliptical slice sampler can be extended to arbitrary priors, as long as we can evaluate the prior density.

Outline

Brief review of (Bayesian) regularization
- Ridge and Lasso regressions
- Shrinkage priors

Algorithm 1: Elliptical slice sampler

Algorithm 2: Elliptical slice-within-Gibbs Sampler

Other issues
- Computational considerations
- Rank deficiency
- Comparison metrics

Simulation exercise
- $n > p$
- $n < p$

Empirical illustrations
- Illustration 1: Beauty and course evaluations
- Illustration 2: Diabetes dataset
- Illustration 3: Motorcycle data

References

Brief review of (Bayesian) regularization

Consider the Gaussian linear model

$$(y|X, \beta, \sigma^2) \sim N(X\beta, \sigma^2 I_n),$$

where β is p-dimensional.

Ridge Regression: ℓ_2 penalty on β:

$$\hat{\beta}_R = \arg\min_{\beta}\{||y - X\beta||^2 + \lambda||\beta||_2^2\}, \qquad \lambda \geq 0,$$

leading to $\hat{\beta}_{ridge} = (X'X + \lambda I)^{-1}X'y$.

LASSO Regression: ℓ_1 penalty on β:

$$\hat{\beta}_L = \arg\min_{\beta}\{||y - X\beta||^2 + \lambda||\beta||_1\}, \qquad \lambda \geq 0,$$

which can be solved by using quadratic programming techniques such as a *coordinate gradient descent* algorithm.

Elastic net

The Elastic net combines the Ridge and the LASSO approaches:

$$\hat{\beta}_{EN} = \underset{\beta}{\arg min}\{||y - X\beta||^2 + \lambda_1||\beta||_1 + \lambda_2||\beta||_2^2\},\ \lambda_1 \geq 0, \lambda_2 \geq 0,$$

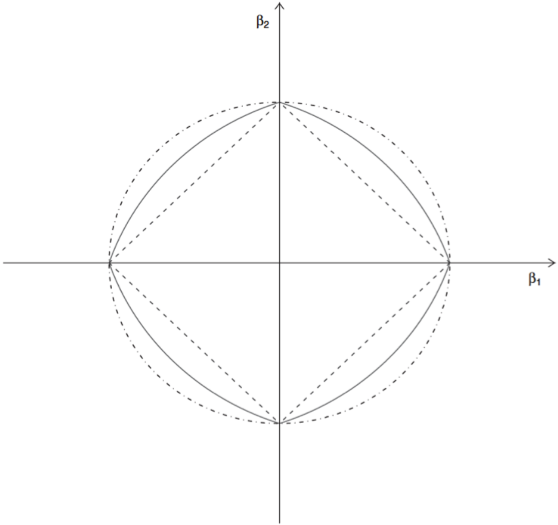
The ℓ_1 part of the penalty generates a sparse model.

The ℓ_2 part of the penalty

- Removes the limitation on the number of selected variables;
- Encourages grouping effect;
- Stabilizes the ℓ_1 regularization path.

`R` package `elasticnet`

Two dimension contour plots of the three penalty functions



Ridge (dot-dashed), LASSO (dashed) and Elastic net (solid)

Bayesian regularization

- Regularization and variable selection are done by assuming independent prior distributions from a scale mixture of normals (SMN) class:

$$\beta|\psi \sim \mathcal{N}(0, \psi) \qquad \text{and} \qquad \psi \sim p(\psi),$$

- The posterior mode or the maximum a posteriori (MAP) is

$$\arg\max_{\beta} \; \{\log p(y|\beta) + \log p(\beta|\psi)\}$$

which is equivalent to penalizing the log-likelihood

$$\log p(y|\beta)$$

with penalty equal to the log prior

$$\log p(\beta|\psi)$$

when ψ is held fixed.

Bayesian regularization in linear regression problems

The marginal prior distribution of β

$$p(\beta) = \int_0^\infty p_N(\beta, 0, \psi) p(\psi) d\psi$$

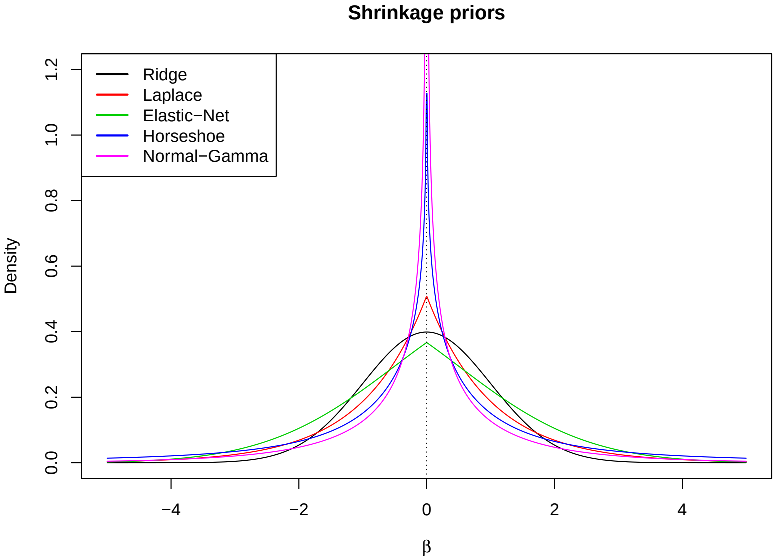
can assume many forms depending on the mixing distribution $p(\psi)$:

	$p(\psi)$	$p(\beta)$
Ridge	$\mathcal{IG}(\alpha, \delta)$	Student's t
Lasso	$\mathcal{E}(\lambda^2/2)$	Laplace
NG prior	$\mathcal{G}(\lambda, 1/(2\gamma^2))$	Normal-Gamma
Horseshoe	$\sqrt{\psi} \sim C^+(0,1)$	No closed form

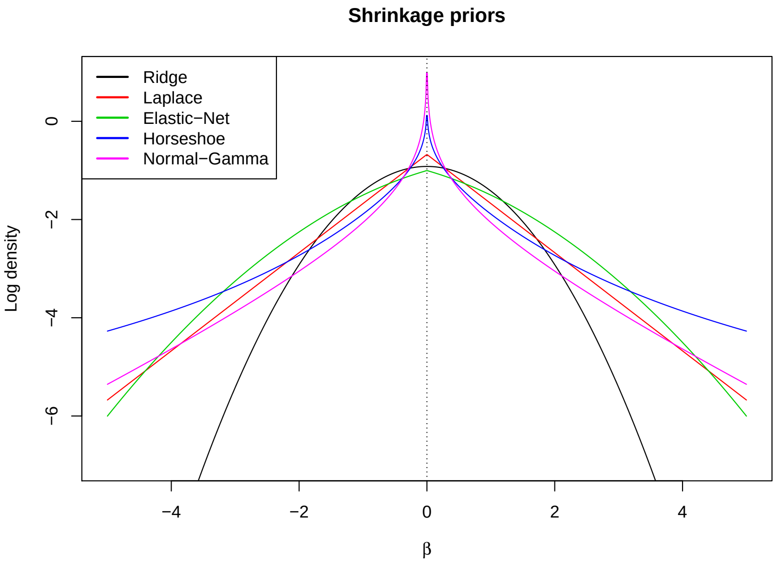
where,

$$\begin{aligned} p_{NG}(\beta|\lambda,\gamma) &= \frac{1}{\sqrt{\pi} 2^{\lambda-1/2} \gamma^{\lambda+1/2} \Gamma(\lambda)} |\beta|^{\lambda-1/2} K_{\lambda-1/2}(|\beta|/\gamma) \\ \log p_H(\beta) &\approx \log\left(1 + \frac{4}{\beta^2}\right) \end{aligned}$$

Comparing shrinkage priors



Comparing shrinkage priors



Algorithm 1: Elliptical Slice Sampler

- The original elliptical slice sampler (Murray et. al. [2010]) was designed to sampling from a posterior arising from a normal prior and a general likelihood.

- It can also be used with a normal likelihood and general prior such as shrinkage priors.

- Advantages:
  - Flexible : It only requires evaluating the prior density or an approximation (no special samplers are required).

  - Fast : Sample all coefficients simultaneously. Not necessary to loop over variables.

Advantages of our sampling scheme

Flexibility: $\pi(\beta)$ evaluated up to a normalizing constant.

MC efficiency: In each MC iteration, single multivariate Gaussian draw and several univariate uniform draws.

Acceptance rate: The size of the sampling region for θ shrinks rapidly with each rejected value and is guaranteed to eventually accept.

Single/block move: The basic strategy of the elliptical slice sampler can be applied to smaller blocks.

Algorithm 1: Elliptical slice sampler

Goal: Sample Δ from

$$p(\Delta) \propto p_N(\Delta; 0, V) L(\Delta)$$

Key idea: For v_0 and v_i iid $N(0, V)$ and any $\theta \in [0, 2\pi]$, it follows that

$$\Delta = v_0 \sin\theta + v_1 \cos\theta \sim N(0, V).$$

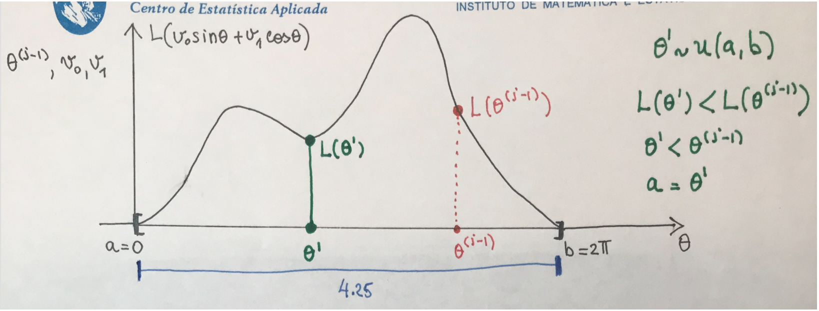
Parameter-expansion: Sampling from

$$p(v_0, v_1, \Delta, \theta) \propto p_N(0, \Sigma_\theta) L(v_0 \sin\theta + v_1 \cos\theta)$$

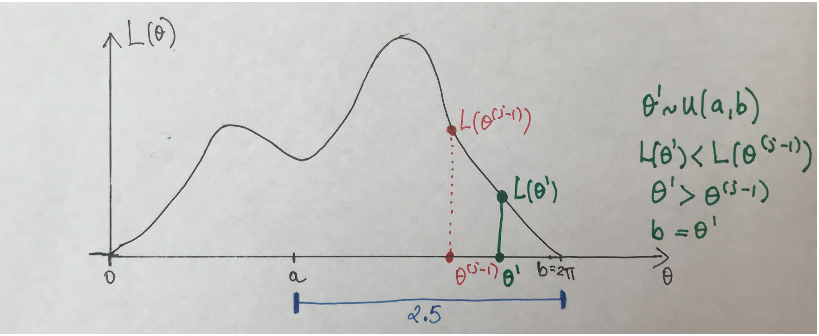
can be done via two-block Gibbs sampler:

- Sample from $p(v_0, v_1 | \Delta, \theta)$
  - Sample v from $N(0, V)$
  - Set $v_0 = \Delta \sin\theta + v \cos\theta$ and $v_1 = \Delta \cos\theta - v \sin\theta$
- Sample from $p(\Delta, \theta | v_1, v_2)$
  - Slice sampling from $p(\theta | v_0, v_1) \propto L(v_0 \sin\theta + v_1 \cos\theta)$
  - Set $\Delta = v_0 \sin\theta + v_1 \cos\theta$

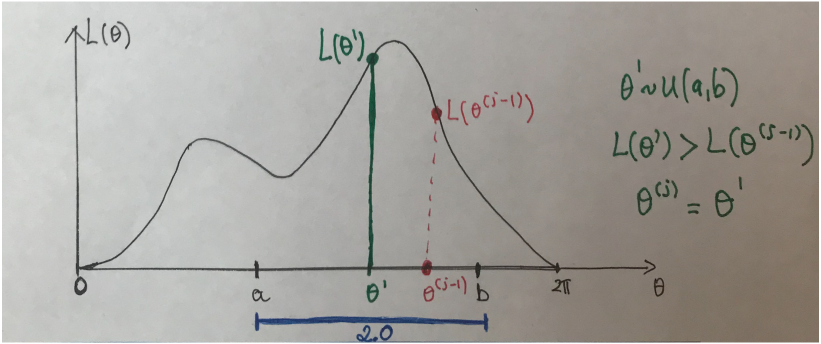
Slice sampling from $p(\theta|v_0, v_1) \propto L(v_0 \sin\theta + v_1 \cos\theta)$



Second draw



Third draw



Elliptical slice sampler for Gaussian linear regression

- Regression model:

$$y|X, \beta, \sigma^2 \sim N(X\beta, \sigma^2 I_n),$$

- Posterior

$$p(\beta|X, y, \sigma^2) \propto \underbrace{f(y \mid X, \beta, \sigma^2)}_{\text{normal}} \quad \underbrace{\pi(\beta)}_{\text{arbitrary prior}}$$

- $f(y \mid X, \beta, \sigma^2)$ can be rewritten as (based on OLS estimates)

$$\pi_0(\beta|X, y, \sigma^2) \propto \exp\left\{-\frac{1}{2\sigma^2}(\beta - \widehat{\beta})' X' X (\beta - \widehat{\beta})\right\}.$$

- The slice sampler of Murray et al (2010) can be applied directly, using $\pi_0(\beta)$ as the Gaussian "prior" and $\pi(\beta)$ as the arbitrary "likelihood"

- We actually sample $\Delta = \beta - \widehat{\beta}$, which is centered around zero.

Elliptical slice sampler for Gaussian linear regression

For initial value β, $\Delta = \beta - \hat{\beta}$, σ^2 fixed, and $\theta \in [0, 2\pi]$:

---

1. *Draw $v \sim N(0, \sigma^2(X'X)^{-1})$.*
   Set $v_0 = \Delta \sin\theta + v \cos\theta$
   Set $v_1 = \Delta \cos\theta - v \sin\theta$.

2. *Draw ℓ from $U[0, \pi(\hat{\beta} + v_0 \sin\theta + v_1 \cos\theta)]$.*
   Initialize $a = 0$ and $b = 2\pi$.
   2.1 *Sample θ' from $U[a, b]$.*
   2.2 *If $\pi(\hat{\beta} + v_0 \sin\theta' + v_1 \cos\theta') > \ell$*
   Then: Set $\theta \leftarrow \theta'$. Go to step 3.
   Else: If $\theta' < \theta$, set $a \leftarrow \theta'$, else set $b \leftarrow \theta'$. Go to step 2.1.
3. *Return $\Delta = v_0 \sin\theta + v_1 \cos\theta$ and $\beta = \hat{\beta} + \Delta$.*

Algorithm 2: Elliptical Slice-within-Gibbs Sampler

One problem of naive slice sampler: If the number of regresson coefficients p is large, the green slice region is so tiny thus too many rejections before one acceptable sample.

Solution:

- β has a jointly Gaussian likelihood and independent priors, it's natural to write a Gibbs sampler, update a subset β^k given other coefficients β^{-k} in each MCMC iteration.
- Apply elliptical slice sampler to conditional likelihood $\beta^k \mid \beta^{-k}$ instead of full likelihood.
- Thus it is a "slice-within-Gibbs sampler"

Update a subset $\beta^k \mid \beta^{-k}$

Let us assume that β can be split into β^k and β^{-k}.

$$\begin{bmatrix} \beta^k \\ \beta^{-k} \end{bmatrix} \sim N\left(\begin{bmatrix} \widehat{\beta}^k \\ \widehat{\beta}^{-k} \end{bmatrix}, \sigma^2 \begin{bmatrix} \Sigma_{k,k} & \Sigma_{k,-k} \\ \Sigma_{-k,k} & \Sigma_{-k,-k} \end{bmatrix} \right) \tag{1}$$

where

$$\begin{bmatrix} \widehat{\beta}^k \\ \widehat{\beta}^{-k} \end{bmatrix} = \widehat{\beta} \quad \text{and} \quad \begin{bmatrix} \Sigma_{k,k} & \Sigma_{k,-k} \\ \Sigma_{-k,k} & \Sigma_{-k,-k} \end{bmatrix} = (X'X)^{-1}.$$

Therefore, the conditional distribution of β^k given β^{-k} is $N(\tilde{\beta}^k, \tilde{\Sigma}^k)$:

$$\tilde{\beta}^k = \widehat{\beta}^k + \Sigma_{k,-k}\Sigma_{-k,-k}^{-1}(\beta^{-k} - \widehat{\beta}^{-k}) \tag{2}$$

$$\tilde{\Sigma}^k = \sigma^2 \left(\Sigma_{k,k} - \Sigma_{k,-k}\Sigma_{-k,-k}^{-1}\Sigma_{-k,k} \right). \tag{3}$$

Simulation shows that **let $k = 1$, sampling one element of β at a time** has highest effective sample size per second!

Algorithm 2: Slice-within-Gibbs for linear regression

For each k from 1 to K.

1. *Construct $\tilde{\beta}^k$ and $\tilde{\Sigma}^k$ as in expressions (2) and (3).*
   Set $\Delta^k = \beta^k - \tilde{\beta}^k$.
   Draw $v \sim N(0, \tilde{\Sigma}^k)$.
   Set $v_0 = \Delta^k \sin\theta^k + v\cos\theta^k$
   Set $v_1 = \Delta^k \cos\theta^k - v\sin\theta^k$.
2. *Draw ℓ uniformly on $[0, \pi(\Delta^k + \tilde{\beta}^k)]$.*
   Initialize $a = 0$ and $b = 2\pi$.
   - 2.1 *Sample θ' uniformly on $[a, b]$.*
   - 2.2 *If $\pi(\tilde{\beta}^k + v_0 \sin\theta' + v_1\cos\theta') > \ell$, set $\theta^k \leftarrow \theta'$. Go to step 3.*
     Otherwise, shrink the support of θ' (if $\theta' < \theta^k$, set $a \leftarrow \theta'$; if $\theta' > \theta^k$, set $b \leftarrow \theta'$), and go to step 2.1.
3. *Return $\Delta^k = v_0 \sin\theta^k + v_1 \cos\theta^k$ and $\beta^k = \tilde{\beta}^k + \Delta^k$.*

Computational considerations

Question: Is slice-within-Gibbs sampler (updating one element of β at a time) better than regular standard Gibbs sampler?

Answer: **Yes!**, All the conditional covariance are fixed in MCMC iterations. We can precompute all the conditional likelihoods.

We can precompute $\Sigma_{k,-k}\Sigma_{-k,-k}^{-1}$, $\Sigma_{k,k} - \Sigma_{k,-k}\Sigma_{-k,-k}^{-1}\Sigma_{-k,k}$, and Cholesky factors L_k, with $L_k L_k^T = \Sigma_{k,k} - \Sigma_{k,-k}\Sigma_{-k,-k}^{-1}\Sigma_{-k,k}$, for each $k = 1, \ldots, K$

By contrast, regular Gibbs samplers have full conditional updates of the form

$$(\beta \mid X, y, \sigma^2) \sim \mathrm{N}((X'X + D)^{-1}X'y, \sigma^2(X'X + D)^{-1}), \qquad (4)$$

which require costly Cholesky or eigenvalue decompositions of the matrix $(X'X + D)^{-1}$ at each iteration as D is updated

Rank deficient case

Question: What if $X'X$ is not invertible?

Answer: Recall that the slice sampler draws from

$$p(\beta \mid y, X, \sigma) \propto \mathrm{N}_Y(X\beta, \sigma^2)\pi(\beta) \propto \mathrm{N}_\beta(\hat{\beta}, \sigma^2(X'X)^{-1})\pi(\beta), \tag{5}$$

It can be written as

$$\begin{aligned} p(\beta \mid y, X, \sigma) &\propto \mathrm{N}_Y(X\beta, \sigma^2)\mathrm{N}_\beta(0, c\sigma^2 I)\frac{\pi(\beta)}{\mathrm{N}_\beta(0, c\sigma^2 I)} \\ &\propto \mathrm{N}_\beta(\bar{\beta}, \sigma^2(X'X + c^{-1}I)^{-1})\frac{\pi(\beta)}{\mathrm{N}_\beta(0, c\sigma^2 I)}. \end{aligned} \tag{6}$$

$c > 0$ and makes $(X'X + c^{-1}I)$ invertible, then we evaluate $\frac{\pi(\beta)}{\mathrm{N}_\beta(0, c\sigma^2 I)}$ rather than $\pi(\beta)$. Pick small c around 1 works fine in practice.

Comparison metrics

Effective sample size (ESS) per second: Letting N denote the Monte Carlo sample size, then ESS for parameter β_j is

$$ESS(\beta_j) = \frac{N}{1 + 2\sum_{k=1}^{\infty} \rho_k}, \tag{7}$$

where $\rho_k = \text{corr}\left(\beta_j^{(0)}, \beta_j^{(k)}\right)$ is the autocovariance of lag k. We divide ESS by running time in seconds to compute ESS per second as a measure of efficiency of each sampler.

Root mean square error: Suppose $\{\bar{\beta}_j\}$ are posterior means of each variable and $\{\beta_j\}$ are true values. The estimation error is measured by

$$\text{error} = \sqrt{\frac{\sum_{j=1}^{p}(\bar{\beta}_j - \beta_j)^2}{\sum_{j=1}^{p} \beta_j^2}}. \tag{8}$$

Simulation exercise

1. Draw elements of β from a "sparse Gaussian" where $\lceil p \rceil$ entries of β are non-zero, drawn from a standard Gaussian distribution, and all other entries are zero.
2. Generate the regressors matrix X in one of two ways.
   - Independent regressor: X_{ij} are iid from standard Gaussian.
   - Factor structure: Suppose there are $k = p/5$ factors. Factors are iid $F \sim N(0, 1)$. The factor loading matrix, B, has exactly five ones in each column and a single 1 in each row, all others 0, so that BB' is block diagonal, with blocks of all ones and all other elements being zero. The regressors are then set as $X = F'B' + \varepsilon$ where ε_{ij} are iid $N(0, 0.01)$.
3. Set $\sigma = \kappa\sqrt{\sum_{j=1}^{p} \beta_j^2}$, where κ controls noise level.
4. Draw $y_i = x_i'\beta + \epsilon_i, \epsilon_i \sim \mathrm{N}(0, \sigma^2)$ for $i = 1, \ldots, n$.

Additionally, we vary the noise level, letting $\kappa = 1$ or $\kappa = 2$, corresponding to signal-to-noise ratios of 1 and 1/2, respectively.

$n > p$, regressors are independent

$\kappa = 1$, signal-to-noise ratio is 1:1. The regressors are independent. We show effective sample size per second and RMSE of estimation.

Prior	p	RMSE				ESS per second		
		OLS	slice	mono	Gibbs	slice	mono	Gibbs
Horseshoe	100	3.38%	1.52%	1.51%	1.51%	**1399**	613	567
	1000	1.05%	0.27%	0.27%	0.27%	**91**	5	5
Laplace	100	3.38%	2.39%	2.38%	–	**2362**	809	–
	1000	1.04%	0.63%	0.63%	–	**168**	8	–
Ridge	100	3.38%	3.20%	3.20%	–	**3350**	959	–
	1000	1.06%	0.99%	0.99%	–	**178**	5	–

We compare elliptical slice sampler, Gibbs sampler in `R` package `monomvn` (column mono) and our own implementation of Gibbs sampler for horseshoe regression (column Gibbs).

The elliptical slice sampler has similar error to the Gibbs sampler, but much higher effective sample size per second.

$n > p$, regresssors have factor structure

$\kappa = 1$, signal-to-noise ratio is 1:1. The regressors have underlying factor structure where every 5 regressors are highly correlated.

Prior	p	Error				ESS per second		
		OLS	1block	mono	Gibbs	1block	mono	Gibbs
Horseshoe	100	16.47%	6.06%	6.04%	6.03%	**387**	747	792
	1000	6.85%	1.64%	1.64%	1.64%	**36**	4	4
Laplace	100	17.06%	7.21%	7.15%	–	**573**	1257	–
	1000	6.77%	1.95%	1.94%	–	**38**	5	–
Ridge	100	16.90%	8.50%	8.75%	–	**669**	1668	–
	1000	6.85%	2.93%	3.09%	–	**38**	6	–

We compare elliptical slice sampler, Gibbs sampler in `R` package `monomvn` (column mono) and our own implementation of Gibbs sampler for horseshoe regression (column Gibbs).

The elliptical slice sampler has similar error to the Gibbs sampler, but much higher effective sample size per second when $p = 1,000$.

$p > n$

Compare with Johndrow and Orenstein [2017] (denoted J&O) for the $p > n$ case. $12,000$ posterior draws with the first $2,000$ as burn-in.

			Running Time		RMSE		ESS per second	
p	n	κ	J&O	Slice	J&O	Slice	J&O	Slice
1000	300	0.25	119.11	91.50	0.0041	0.0038	46.71	43.19
1000	600	0.25	394.02	88.61	0.0028	0.0026	14.68	47.26
1000	900	0.25	905.36	88.91	0.0021	0.0020	6.60	48.85
1000	300	1	127.33	90.19	0.0189	0.0189	43.92	39.25
1000	600	1	399.50	91.17	0.0129	0.0129	14.39	44.12
1000	900	1	927.96	91.58	0.0098	0.0099	6.35	46.09
1500	450	0.25	346.37	187.91	0.0029	0.0027	16.37	21.26
1500	900	0.25	1073.28	185.57	0.0022	0.0021	5.50	23.08
1500	1350	0.25	2629.52	183.68	0.0018	0.0017	2.27	24.04
1500	450	1	326.63	183.66	0.0164	0.0164	17.39	20.28
1500	900	1	1021.47	174.52	0.0100	0.0101	5.73	23.72
1500	1350	1	2515.37	176.51	0.0071	0.0071	2.36	24.78
3000	100	0.25	85.95	985.68	0.0067	0.0075	**69.72**	3.89
3000	500	0.25	575.92	983.64	0.0024	0.0022	9.85	4.17

Similar RMSE. The elliptical slice sampler has higher ESS per second in most cases considered here, especially when $p \approx n$. The Johndrow et al. sampler is much more efficient only when $p \gg n$, such as $p = 3000$ and $n = 100$.

Empirical illustration 1: beauty and course evaluations

Use data from Hamermesh and Parker [2005].

Course evaluations from the University of Texas at Austin between 2000 and 2002 and additional information of class and instructor.

We want to study how the beauty score of instructor affect his or her course evaluation scores (on 1 to 5 scale, 5 is the best).

We fit regression model with horseshoe, ridge, Laplace (lasso) and two exotic priors.

The purpose is not to argue advantage of exotic priors, but just to show we are able to fit them easily.

Exotic priors

Consider "sharkfin" prior

$$\pi(\beta) = \begin{cases} 2qf(\beta) & \beta \leq 0 \\ 2f(\beta/s)(1-q)/s & \beta > 0 \end{cases}, \qquad (9)$$

where $f(x) = \frac{1}{\pi(1+x^2)}$ is the density

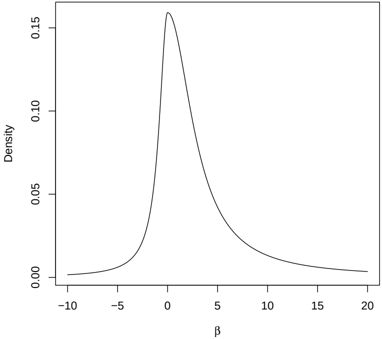


Figure: Density of sharkfin prior.

Exotic priors

"non-local" prior is a mixture of Cauchy priors, which is anti-sparse

$$\pi(\beta) = 0.5t(\beta; -1.5, 1) + 0.5t(\beta; 1.5, 1) \tag{10}$$

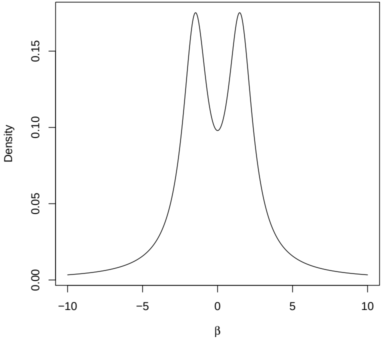


Figure: Density of non-local prior.

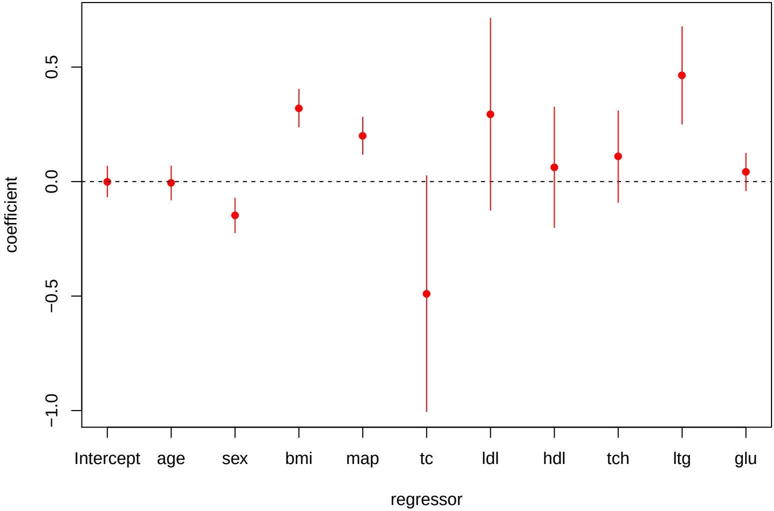
Results of beauty and course evaluations

Table: Posterior points estimates of regression coefficients under each prior; those whose posterior 95% credible intervals exclude zero are shown in bold.

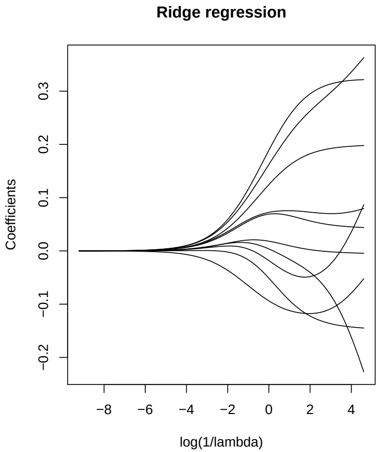
variable name	horseshoe	lasso	ridge	sharkfin	non-local
class size 61 to 150	−0.13	−**0.19**	−**0.20**	−**0.14**	−**0.22**
class size 151 to 600	−**0.36**	−**0.41**	−**0.43**	−**0.36**	−**0.46**
tenure track	0.22	**0.29**	0.31	**0.27**	**0.40**
non-minority	**0.65**	**0.65**	**0.53**	**0.63**	**0.71**
highly beautiful	0.14	**0.36**	**0.54**	0.25	**0.38**

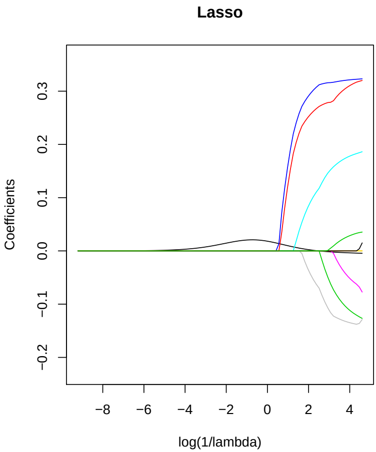
Empirical illustration 2: Diabetes

The data consist of $p = 10$ baseline measurements on $n = 442$ diabetic patients; the response variable is a numerical measurement of disease progression (Efron et al., 2004). Below are the OLS estimates.

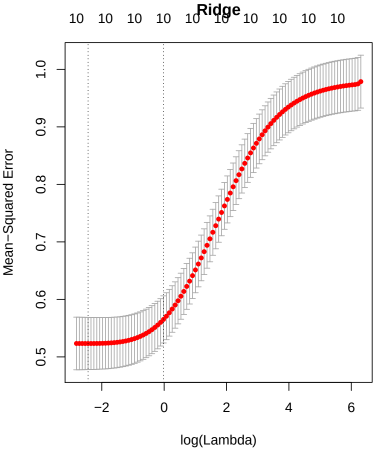


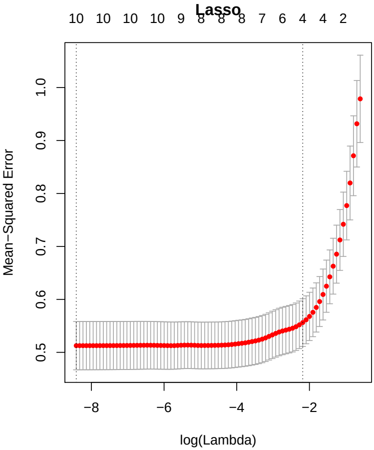
Ridge and Lasso penalties



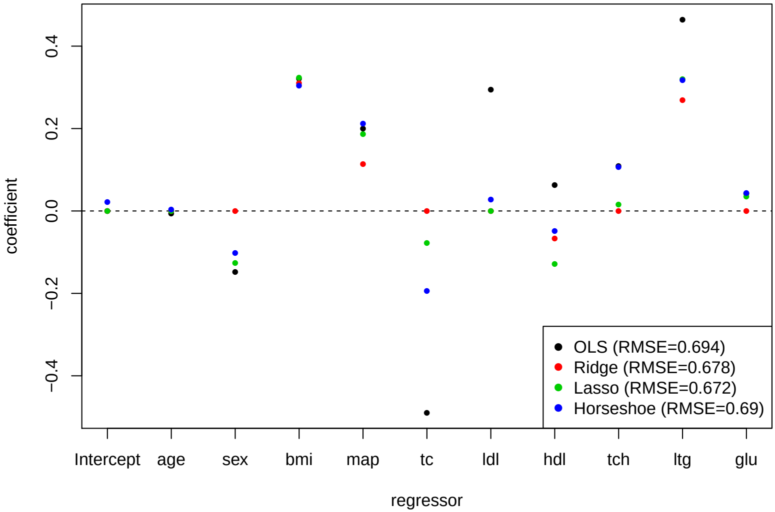


Cross validation

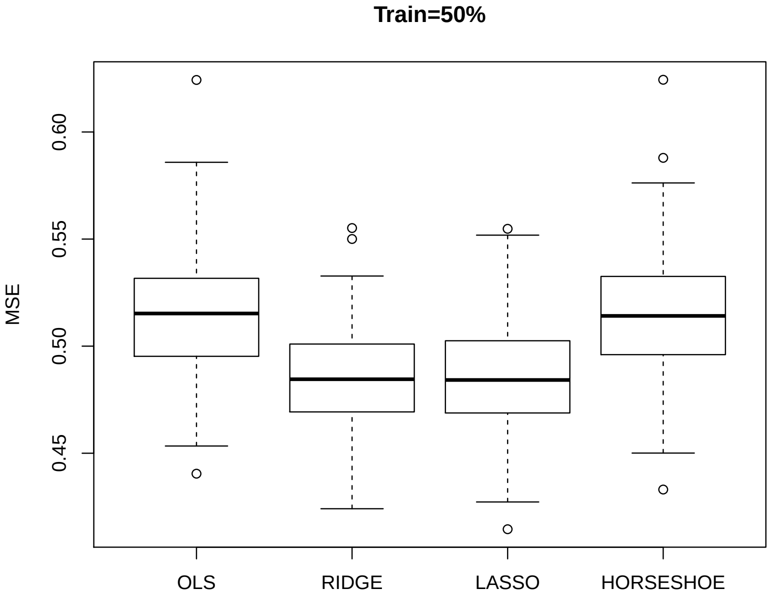




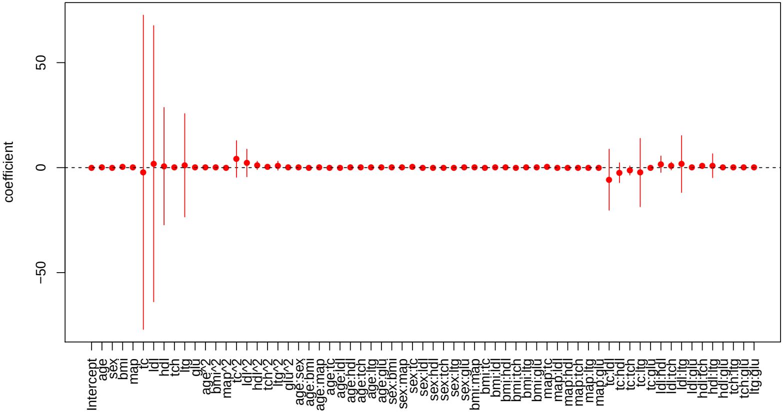
Out-of-sample Root MSE



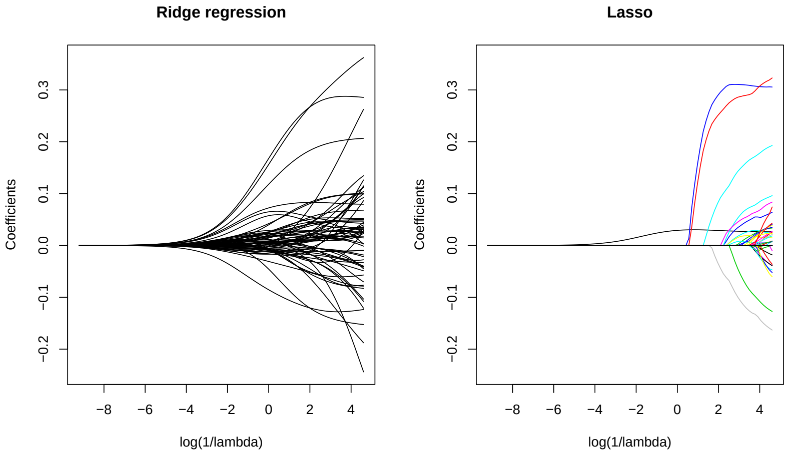
Out-of-sample Root MSE: replications



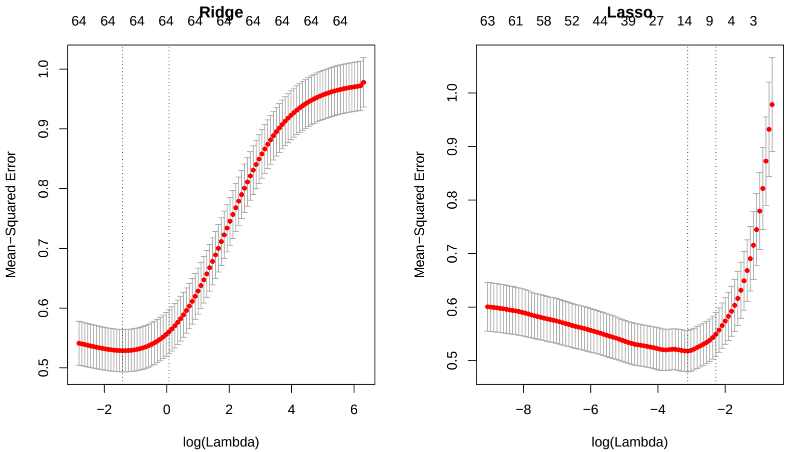
OLS: Including squares and interactions ($p = 64$)



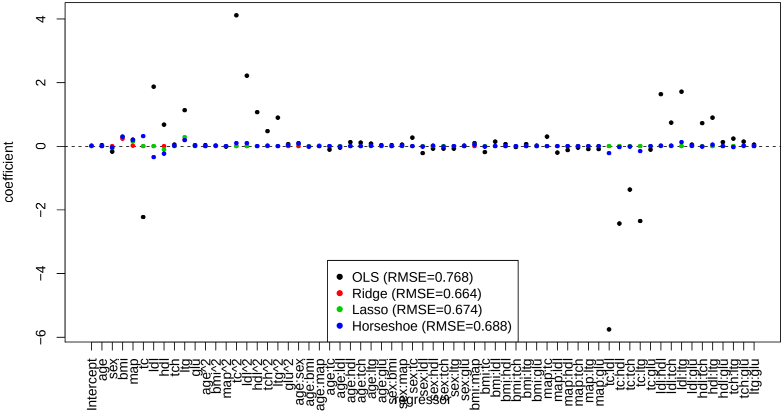
Ridge and Lasso penalties



Cross validation



Out-of-sample Root MSE



Out-of-sample Root MSE: replications

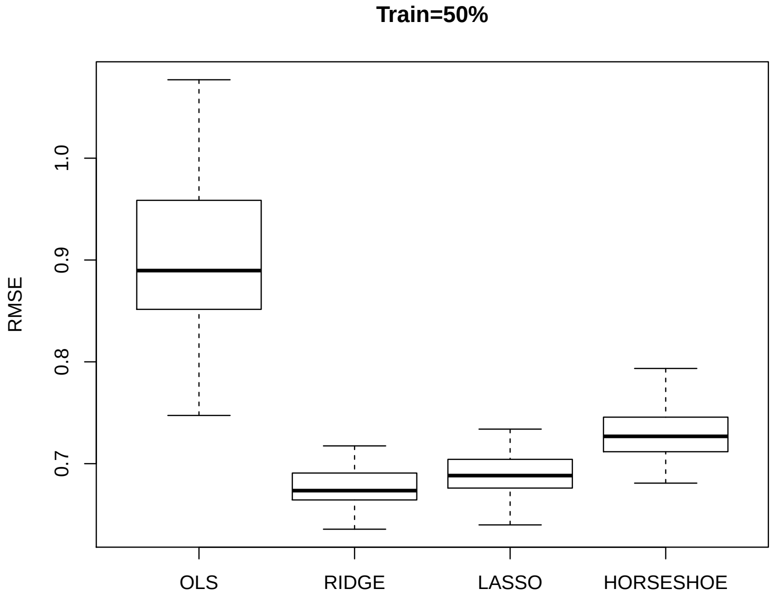


Illustration 3: Motorcycle data

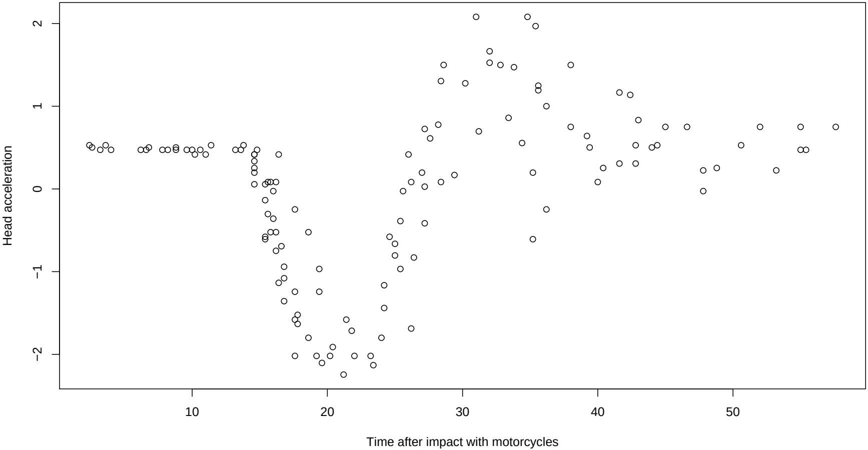
Description
This table gives the results of 133 simulations showing the effects of motorcycle crashes on victims heads: time after a simulated impact with motorcycles and head acceleration of a PTMO (post mortem human test object) were recorded.

Usage
data(motorcycledata)

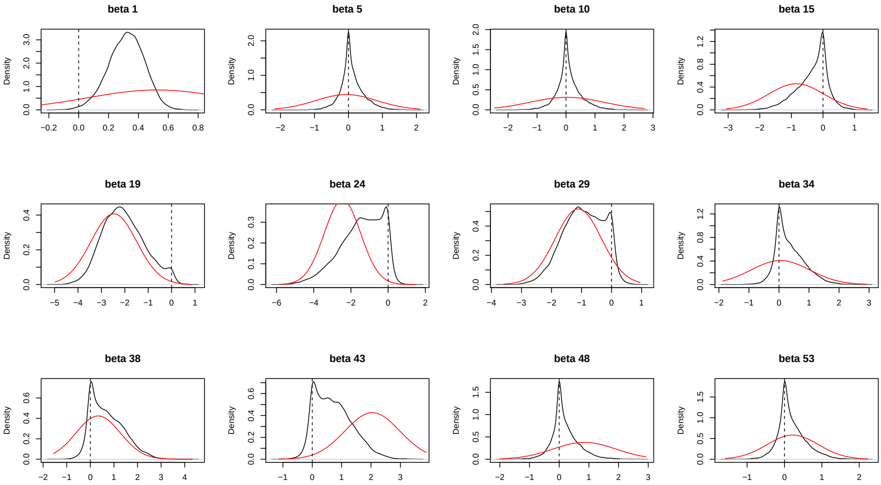
Format
A 133 by 2 data frame.

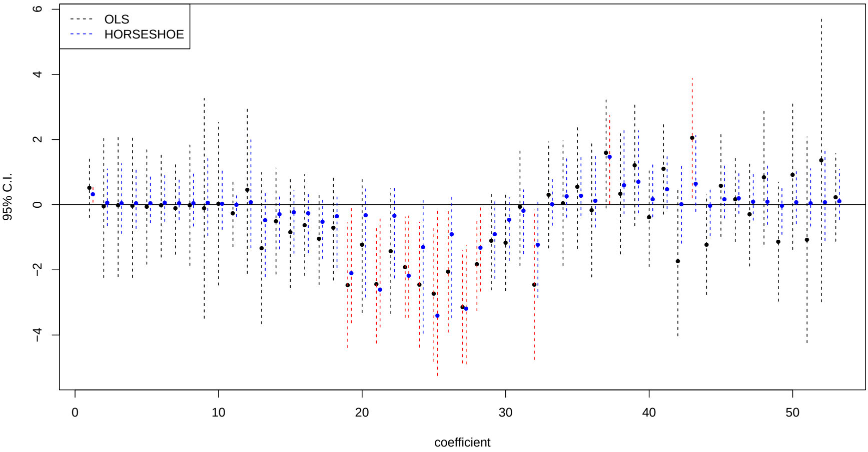
References
Hardle, W. (1990) Applied Nonparametric Regression. Cambridge University Press.

Data


Head acceleration
Time after impact with motorcycles
2
1
0
−1
−2
10
20
30
40
50

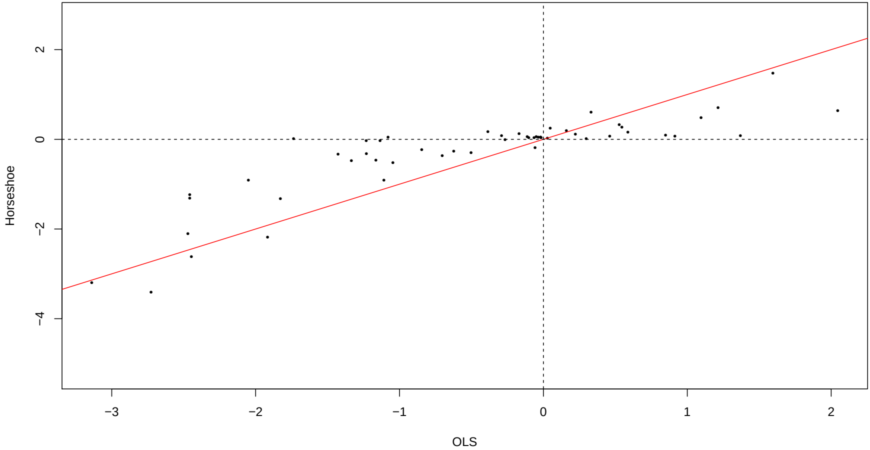
Spline regression: OLS and horseshoe



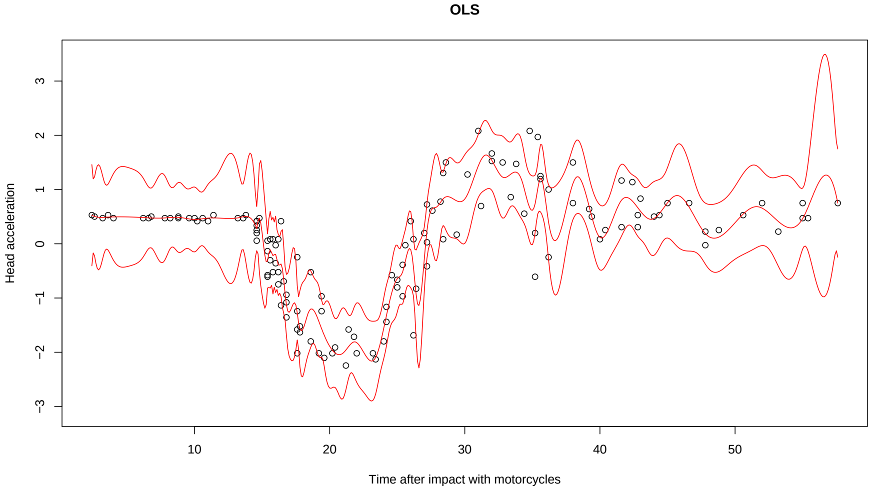


OLS
HORSESHOE
95% C.I.
coefficient
6
4
2
0
−2
−4
0
10
20
30
40
50

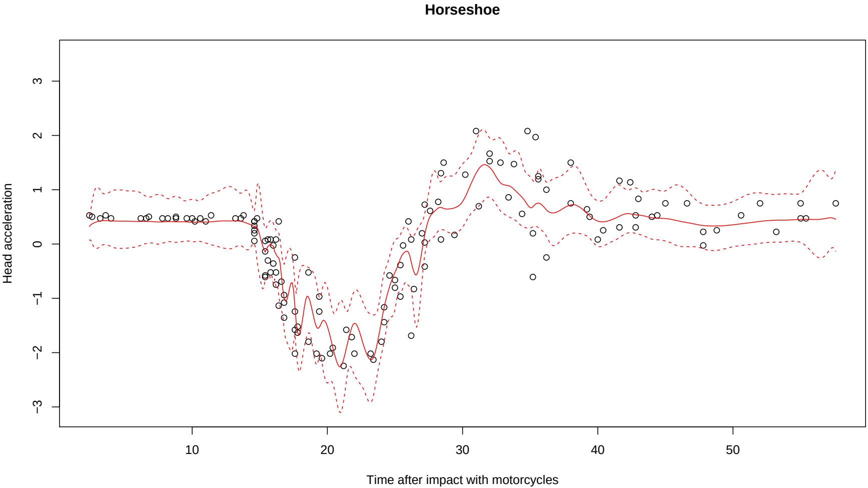
Shrinkage



OLS

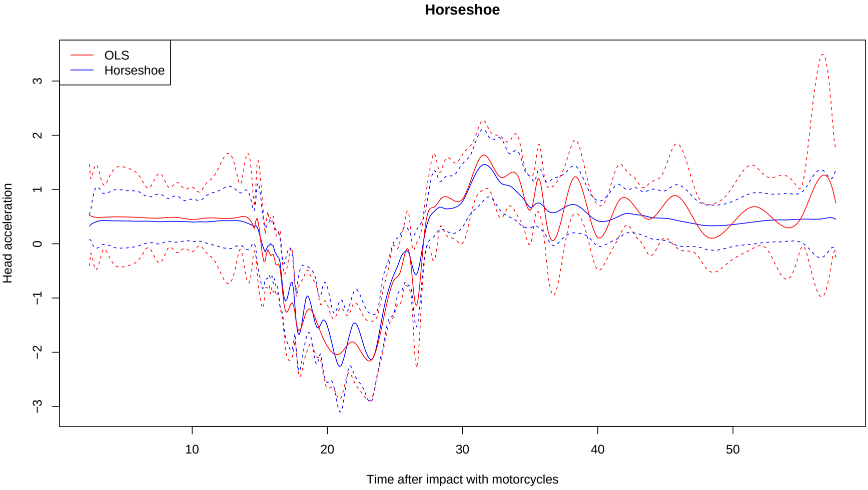

OLS
Head acceleration
Time after impact with motorcycles
−3
−2
−1
0
1
2
3
10
20
30
40
50

Horseshoe



Horseshoe
Head acceleration
Time after impact with motorcycles
3
2
1
0
−1
−2
−3
10
20
30
40
50

Comparison


Horseshoe
OLS
Horseshoe
Head acceleration
3
2
1
0
−1
−2
−3
10
20
30
40
50
Time after impact with motorcycles

R code

```
install.packages("adlift")
install.packages("bayeslm")
require(splines)
library("adlift")
library("bayeslm")

data(motorcycledata)
y = motorcycledata[,2]
y = (y-mean(y))/sqrt(var(y))
x = motorcycledata[,1]
n = length(x)

cuts = quantile(x,seq(0.02,0.98,by=0.02))
X    = bs(x,knots=cuts)
p    = ncol(X)

fit.ols <- lm(y~X)
fit.hs = bayeslm(y~X)
```

References

- Gramacy and Pantaleo (2009) Shrinkage regression for multivariate inference with missing data, and an application to portfolio balancing. *Bayesian Analysis*, 5(2), 237-262.

- Hahn, He and Lopes (2018) Bayesian factor model shrinkage for linear IV regression with many instruments, *Journal of Business & Economic Statistics*, 36(2), 278-287.

- Hahn, He and Lopes (2018) Efficient sampling for Gaussian linear regression with arbitrary priors. *Journal of Computational and Graphical Statistics*, forthcoming.

- Johndrow, Orenstein and Bhattacharya (2017) Scalable MCMC for Bayes shrinkage priors[J]. arXiv preprint `arXiv:1705.00841`.

- Murray, Adams and MacKay (2010) Elliptical slice sampling. In *JMLR Workshop and Conference Proceedings*, Volume 9, pp. 541-548. JMLR.