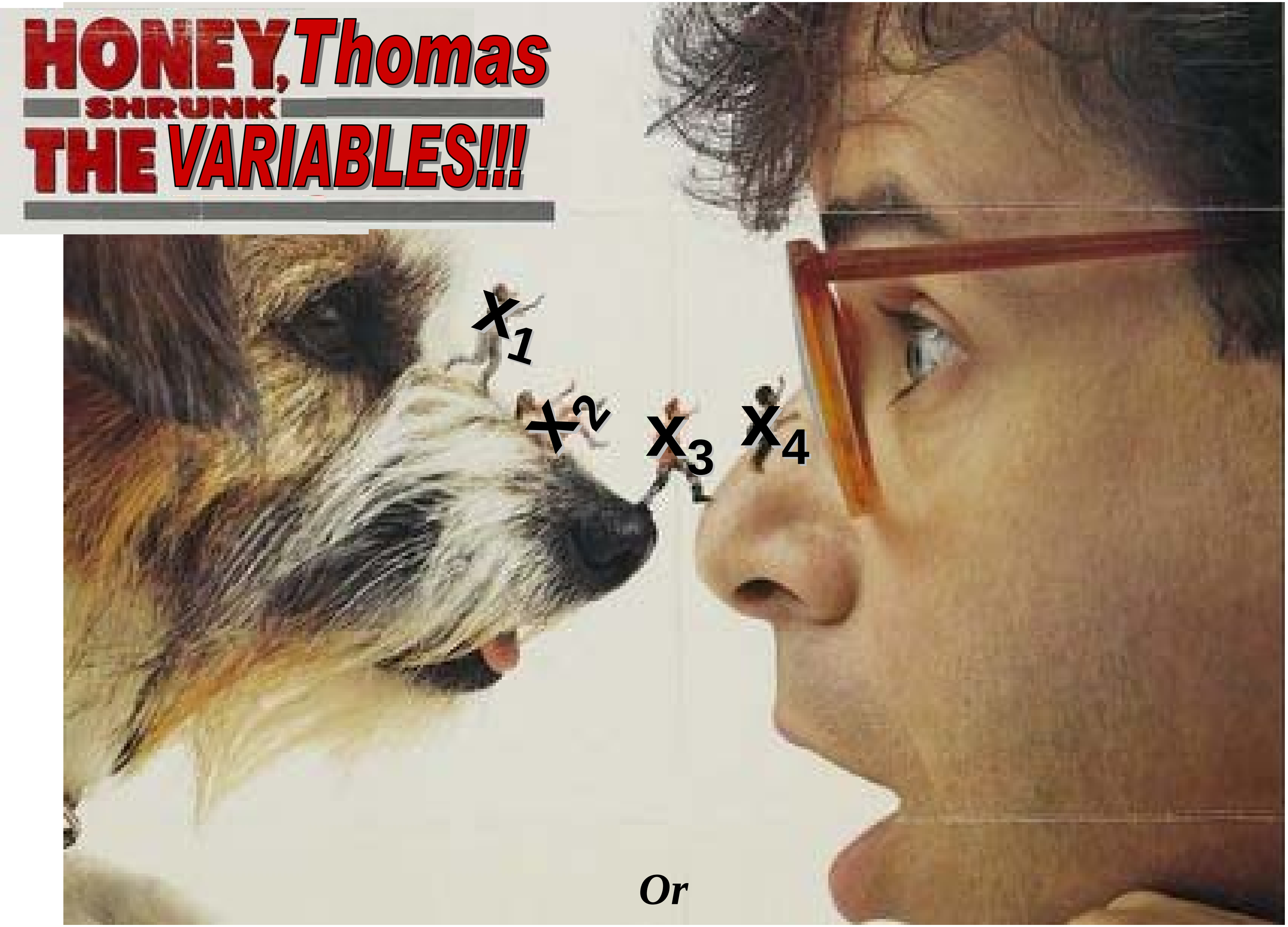
HONEY, Thomas SHRUNK THE VARIABLES!!!

EMEIS + EMEIS Seminar Series

Or

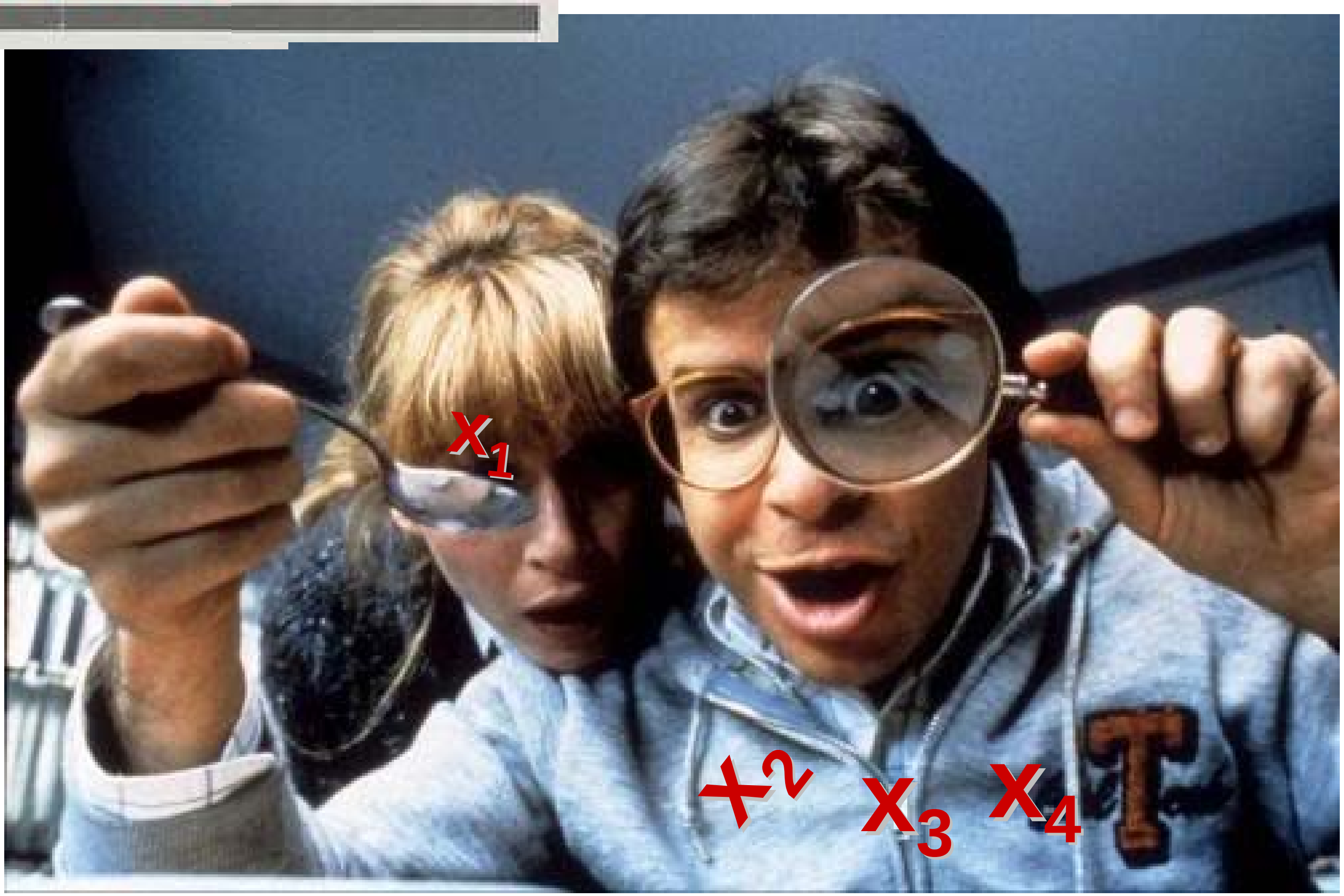
How to select variables Using Bayesian LASSO



Or

How to select variables Using Bayesian LASSO

HONEY, Thomas SHRUNK THE VARIABLES!!!



Or

How to select variables Using Bayesian LASSO

On Bayesian Variable Selection Using Lasso and the specification of the shrinkage parameter

Anastasia Lykou[1] and Ioannis Ntzoufras[2]

[1]Department of Mathematics and Statistics, Lancaster University, UK
[2]Department of Statistics, Athens University of Economics and Business, Greece

7 December 2010, Athens
Emeis kai Emeis

Outline

- Introduction
  - Motivation
- Bayesian variable selection and Lasso
  - Bayesian Lasso using Gibbs Sampling
  - Toy example
- Bayes factors in relation to shrinkage parameter
  - Bayes factors for simple Lasso Regression
  - Interpretation of BF
  - Specifying the shrinkage parameter using the Bayes factors
  - Bayes Factor for multiple Lasso regression
- Illustration
- Conclusion
- Further work
- References

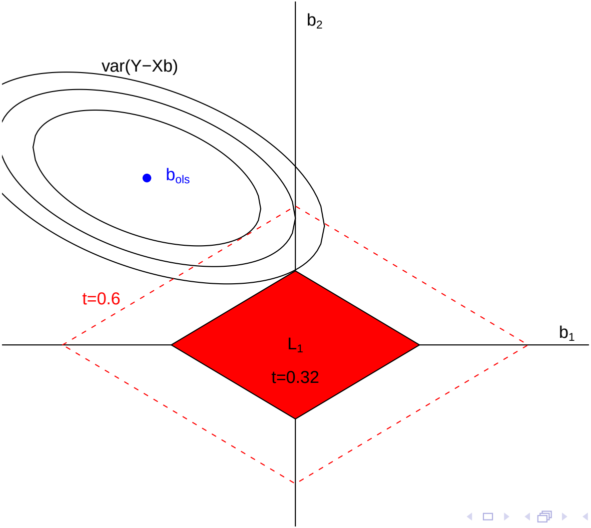
Lasso (least absolute shrinkage and selection operator)

The Lasso (Tibshirani, 1996) performs variable selection and shrinkage on the linear regression problems by imposing the L_1 norm.

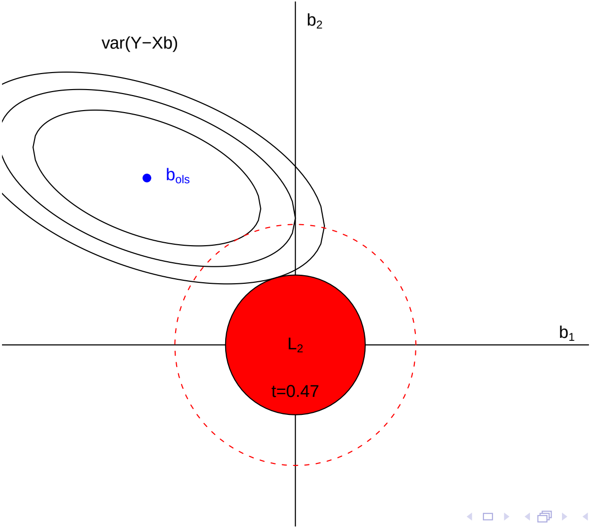
$$\operatorname{argmin}_{\boldsymbol{\beta}} \{ \mathsf{var}(\mathbf{y} - \mathbf{X}\beta) \} \quad \text{subject to} \quad \sum_{j=1}^{p} |\boldsymbol{\beta}_j| \leq t.$$

- The L_1 constraint shrinks the coefficients towards zero and sets some of them to zero if t is small enough.
- The shrinkage parameter takes values between 0 and $t_0 = \sum |\beta_{ols}|$.
- The shrinkage level is usually controlled through $s = \frac{t}{t_0}$.
- Generalized cross validation methods or the C_p criterion are used to tune the shrinkage parameter s.

Geometry of the Lasso



Geometry of the Ridge Regression



Lasso (cont.)

Lasso is treated as an optimization problem

$$\widehat{\beta}_{\text{lasso}} = \text{argmin}_{\beta} \left\{ \text{var}(\mathbf{y} - \mathbf{X}\beta) + \lambda ||\beta||_1 \right\},$$

- one to one correspondence between λ and t (or s),
- for $\lambda = 0 \Rightarrow \widehat{\beta}_{\text{ols}}$,
- the coefficients shrink as λ increases.
- Lasso is intensively used in the literature as a variable selection method,
- extensions and improvements of the method; (Efron et al., 2004, Zou and Hastie, 2005, Zou, 2006, Meier et al., 2008, Lykou and Whittaker, 2010).

Regularization plot

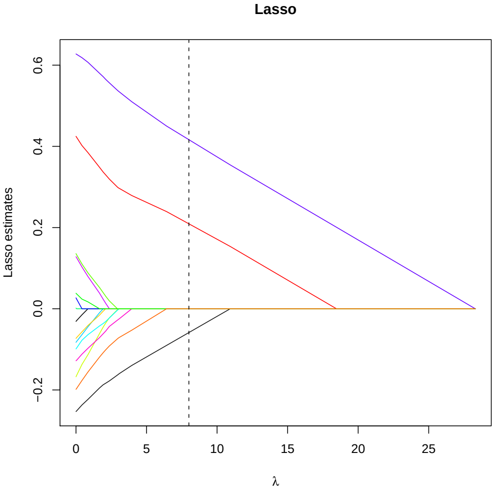


Figure: Lasso estimates against λ for a simulated data set of 15 covariates and $n = 50$.

Bayesian perspective

The Lasso estimates can be considered as the posterior modes under independent double-exponential priors.

$$\begin{aligned} \mathbf{y}|\boldsymbol{\beta} &\sim \mathrm{N}_n(\mathbf{X}\boldsymbol{\beta}, \tau^{-1} I_n), \\ \beta_j &\sim \mathrm{DE}\left(0, \frac{1}{\tau\lambda}\right), \quad \text{for} \;\; j = 1, \ldots, p, \end{aligned}$$

The posterior mode is

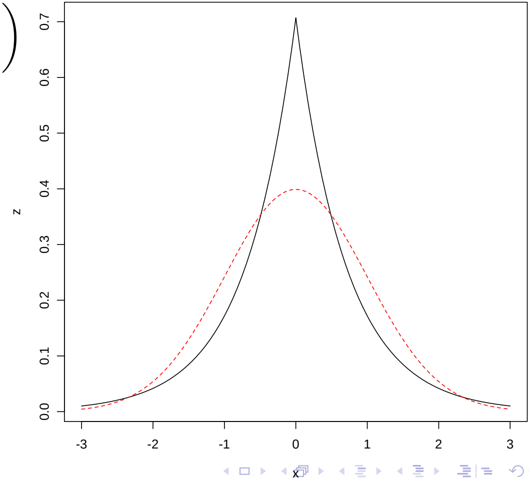
$$\widehat{\boldsymbol{\beta}}_{\text{lasso}} = (\mathbf{X}^T\mathbf{X})^{-1}(\mathbf{X}^T\mathbf{Y} - \lambda\mathbf{s}),$$

where $\mathbf{s}$ is the sign of $\widehat{\boldsymbol{\beta}}_{\text{lasso}}$.

The Double Exponential Distribution

$$f(y|\mu,\lambda) = \frac{1}{2\lambda}\exp\left(-\frac{|y-\mu|}{\lambda}\right)$$

$$\begin{aligned} E(Y) &= \mu \\ V(Y) &= 2\lambda^2 \end{aligned}$$



Background

The double-exponential prior distribution can be represented as a mixture of normal distributions (Andrews and Mallows, 1974).

- Park and Casella (2008): the method lacks of direct variable selection.
- Balakrishnan and Madigan (2009): propose the demi-Bayesian Lasso, where the mixing parameter is found by maximizing the marginal data likelihood and its zero values control the variables excluded from the model.
- Griffin and Brown (2010): adopt the Normal-Gamma prior and shrink the posterior expectation to values very close to zero.

Background (cont.)

Other prior distributions for the Bayesian Lasso.

- Hans (2009) describes also the Bayesian Lasso regression by imposing directly the double-exponential prior. He focuses on the problem of predicting future observations rather than the variable selection problem.
- Hans (2010) addresses model uncertainty for the Bayesian Lasso regression by computing the marginal posterior probabilities for small model space. He handles the cases of large model space by imposing the prior as a mixture of a mass at zero and of the double-exponential and sample the posterior inclusion probabilities by a Gibbs sampler.
- Yuan and Lin (2005) prove that the model with the highest posterior probability is the Lasso solution.

Background (cont.)

Other prior distributions for the Bayesian shrinkage and sparsity.

- Carvalho, Polson and Scott (2010), *Biometrika*. Horseshoe Estimator (Half Cauchy Hyper Prior for normal s.d.)
- Armagan, Dunson and Lee (2010). Double Generalized Pareto

Choice of the shrinkage parameter

The choice of the shrinkage parameter is always crucial in sparse methods.

Most publications in Bayesian versions of Lasso impose a hyperprior on the shrinkage parameter or use empirical methods.

- Lack of interpretation of the chosen values of λ.
- The range of values of λ for each data set is not known.

Our contribution

- We use the variable selection method that introduces an indicator parameter, which identifies the variables included in the model formulation.

  Quantify model uncertainty by deriving the posterior model probabilities.

- Propose a method that controls the shrinkage levels using simple arguments based on traditional correlation coefficients that are widely understood.

  Examine the sensitivity of the variable selection problem on the various shrinkage values.

  Propose values for the shrinkage parameter that we can interpret.

Bayesian Lasso

We use the following formulation

$$
\begin{aligned}
\mathbf{y}|\boldsymbol{\beta},\tau,\boldsymbol{\gamma} &\sim \mathrm{N}_n(\mathbf{X}\mathbf{D}_{\gamma}\boldsymbol{\beta},\tau^{-1}I_n), \quad \text{where} \quad \mathbf{D}_{\gamma} = \mathrm{diag}(\gamma_1,\dots,\gamma_p), \\
\boldsymbol{\beta}_j &\sim \mathrm{DE}\left(0,\frac{1}{\tau\lambda}\right), \quad \text{for} \quad j=1,\dots,p, \qquad (1)\\
\gamma_j &\sim \mathrm{Bernoulli}(\pi_j), \\
\tau &\sim \mathrm{Gamma}(a,d),
\end{aligned}
$$

where λ is the shrinkage parameter which controls the prior variance given by $2/(\lambda\tau)^2$.

- Inference is based on the posterior medians of $\beta_j^* = \gamma_j\beta_j$ for $j=1,\dots,p$.
- We estimate $f(\boldsymbol{\beta}|\mathbf{y},\cdot)$ and $f(\boldsymbol{\gamma}|\mathbf{y},\cdot)$ with Kuo and Mallick (1998) Gibbs sampler for variable selection.
- Any equivalent such as GVS (Dellaportas et al , 2002) or RJMCMC (Green, 1995) will provide similar results.

A Gibbs Sampler for Bayesian Lasso

- If $\gamma_j = 1$,
  - Generate ω_j from Bernoulli(w_j)
  - Generate β_j from $\begin{cases} TN(m_j^-, s_j^2, \beta_j < 0), & \text{if } \omega_j = 1 \\ TN(m_j^+, s_j^2, \beta_j \geq 0), & \text{if } \omega_j = 0 \end{cases}$,

where
$TN(\mu, \sigma^2, A)$ is the normal distribution truncated in the subset $A \subset \Re$,

ω_j is binary parameter specifying whether β_j is less than 0 or not with probability of success

$$w_j = \frac{\Phi(-m_j^-/s_j)/f_N(0\,; m_j^-, s_j^2)}{\Phi(-m_j^-/s_j)/f_N(0\,; m_j^-, s_j^2) + \Phi(m_j^+/s_j)/f_N(0\,; m_j^+, s_j^2)},$$

$m_j^- = \frac{c_j + \lambda}{||X_j||^2}$, $m_j^+ = \frac{c_j - \lambda}{||X_j||^2}$, $c_j = X_j^T(e + \beta_j X_j)$, $s_j^2 = \frac{1}{\tau ||X_j||^2}$
and $\mathbf{X}_j$ is the jth column of matrix $\mathbf{X}$ and $e = \mathbf{y} - \eta$ is the vector of residuals.

A Gibbs Sampler for Bayesian Lasso (cont.)

- If $\gamma_j = 0$, generate β_j from its prior, that is

$$\beta_j|\mathbf{y}, \tau, \boldsymbol{\beta}_{\setminus j}, \boldsymbol{\gamma}_{\setminus j}, \gamma_j = 0 \quad \sim \quad \mathrm{DE}\left(0, \frac{1}{\tau\lambda}\right)$$

- Generate τ from $G\left(\frac{n}{2} + p + \alpha, \frac{||Y - \mathbf{X}\mathbf{D}_{\boldsymbol{\gamma}}\boldsymbol{\beta}||^2}{2} + \lambda||\boldsymbol{\beta}|| + d\right)$.
- Generate γ_j from Bernoulli with probability $O_j/(1 + O_j)$ with

$$O_j = \frac{f(\mathbf{y}|\boldsymbol{\beta}, \tau, \boldsymbol{\gamma}_{\setminus j}, \gamma_j = 1)}{f(\mathbf{y}|\boldsymbol{\beta}, \tau, \boldsymbol{\gamma}_{\setminus j}, \gamma_j = 0)} \frac{\pi(\boldsymbol{\gamma}_{\setminus j}, \gamma_j = 1)}{\pi(\boldsymbol{\gamma}_{\setminus j}, \gamma_j = 0)}.$$

Example

A simulated dataset of Dellaportas et al. (2002), consists of $n = 50$ observations and $p = 15$ covariates generated from a standardised normal distribution and the response from

$$y_i \sim \mathrm{N}(X_{i4} + X_{i5}, 2.5^2), \quad \text{for} \quad i = 1, \ldots, 50.$$

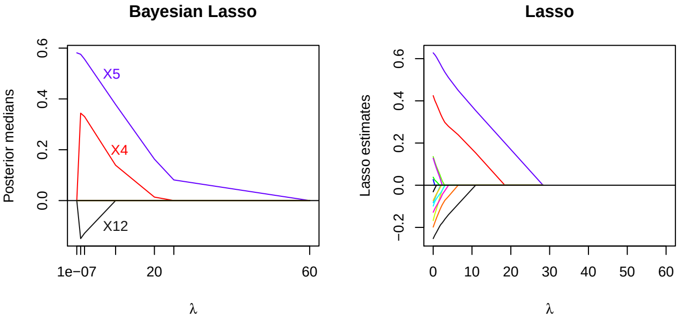


Figure: Posterior medians of $\beta_j^* = \gamma_j \beta_j$ and usual Lasso estimates against λ.

Bayes factors for simple Lasso Regression

The Bayes factors for the comparison between two simple models:

$$m_j : \mathbf{y}|\boldsymbol{\beta}, \tau, m_j \sim \mathrm{N}_n(X_j\beta_j, \tau^{-1}I_n), \qquad m_0 : \mathbf{y}|\boldsymbol{\beta}, \tau, m_0 \sim \mathrm{N}_n(0, \tau^{-1}I_n).$$

Assuming standardized data, $\mathrm{BF}_{\mathrm{un}}$ can be expressed

$$\mathrm{BF}_{\mathrm{un}} = \frac{f(\mathbf{y}|m_j)}{f(\mathbf{y}|m_0)} = \frac{\lambda c}{n-1}\left\{h_1(n, \lambda, \rho) + h_2(n, \lambda, \rho)\right\}$$

- h_1, h_2 are function of n, λ, ρ.
- ρ_j is the sample Pearson correlation between $\mathbf{y}$ and $\mathbf{X}_j$. Without loss of generality we assume that is positive.

For fixed sample size, the $\mathrm{BF}_{\mathrm{un}}$ is a function of ρ and λ.

Interpretation of BF

According to the Kass and Raftery (1995) interpretation tables

BF	Evidence against H_0
1 to 3	Not worth more than a bare mention
3 to 20	Substantial
> 20	More than strong

- As ρ increases, we expect that BF_{un} will become stronger.
- As shown in the Bayesian regularization path, there are variables never included in the model.
- There are values of ρ that correspond to $\text{BF}_{\text{un}} < 3$ for all the values of λ.
- Variables with such correlations will not be supported in the simple regression model.

Graphical representation

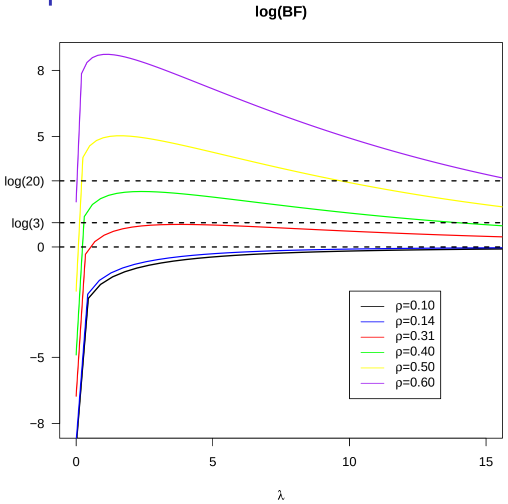


Figure: Bayes factor $\mathrm{BF}_{\mathrm{un}}$ against the values of λ, ρ; sample size is fixed to $n = 50$.

Graphical interpretation (cont.)

- For $n = 50$, the BF_{un} never provides strong evidence against the H_0 for X_j which is associated with Y with $\rho \leq 0.31$.

  The choice of λ will not affect the variable selection process when $\rho \leq 0.31$.
- As ρ increases, there are values of λ that correspond to $\text{BF}_{\text{un}} > 3$.

Graphical interpretation (cont.)

- For any given sample size *n*, there is a range of sample correlations that make $\text{BF}_{\text{un}} < 3$.

  We call this set of values of ρ as "non-important" set and it is defined as

  $$\rho \in \{\text{BF}_{\text{un}}(\rho, \lambda) < 3\}, \text{ for all the values of } \lambda.$$

- An iterative algorithm derives the non-important set of ρ, for a given sample size *n*.

Specification of λ

We choose some values for the correlations higher than this set and identify for which λ the $\mathrm{BF_{un}} = 1$. We call these values threshold correlations.

For example, for $n = 50$, we select the threshold correlation to be $\rho_t = 0.4$ and find the λ that makes the corresponding $\mathrm{BF_{un}} = 1$.

This procedure provides with a value of λ that gives 50% posterior probability to the model with a covariate of such correlation and 50% to the constant model.

n $\rho : \mathrm{BF_{un}} < 3$	50 (0,0.31)	100 (0,0.22)	500 (0,0.10)
$BF = 1$	$\rho_t = 0.35,\ \lambda = 0.218$	$\rho_t = 0.25,\ \lambda = 0.335$	$\rho_t = 0.15,\ \lambda = 0.060$
$BF = 1$	$\rho_t = 0.40,\ \lambda = 0.067$	$\rho_t = 0.30,\ \lambda = 0.069$	$\rho_t = 0.20,\ \lambda = 7 \times 10^{-4}$
$BF = 1$	$\rho_t = 0.50,\ \lambda = 0.004$	$\rho_t = 0.40,\ \lambda = 0.001$	$\rho_t = 0.30,\ \lambda = 5 \times 10^{-6}$

Specification of λ (cont.)

The advantages of the chosen values of λ are

- We know how these values affect the BF_{un}.
- The chosen λ is easily interpreted through the sample correlation between the candidate variable and the response.
- We control how strict we want to be while choosing the threshold values.

However, we have seen the interpretation of the chosen shrinkage values on the simple Lasso regression.

We check what happens when these shrinkage levels are imposed on the multiple Lasso regression.

Bayes Factor for multiple Lasso regression

We compare the model m_A, where A is the set of the variables included in the model with the model $m_{A\setminus j}$, where the j_{th} variable is excluded.

We use the Laplace approximation to derive the corresponding BF.

$$\mathrm{BF_m} \approx \frac{\lambda c}{n-1}\left[1-\ \mathrm{corr}^{(\mathrm{lasso})2}(\mathbf{y},\mathbf{X}_j|\mathbf{X}_{A\setminus j})\right]^{-df/2}\frac{1}{\sqrt{(1-R^2_{\mathbf{X}_j|\mathbf{X}_{A\setminus j}})(1-R^{(\mathrm{lasso})2}_{\mathbf{y}|\mathbf{X}_{A\setminus j}})}}.$$

where

- $\mathrm{corr}^{(\mathrm{lasso})2}(\mathbf{y},\mathbf{X}_j|\mathbf{X}_{A\setminus j})$ is the Lasso version of the partial correlation,
- $R^2_{\mathbf{X}_j|\mathbf{X}_{A\setminus j}}$ is the multiple correlation coefficient when regressing $\mathbf{X}_j$ on $\mathbf{X}_{A\setminus j}$,
- $R^{(\mathrm{lasso})2}_{\mathbf{y}|\mathbf{X}_{A\setminus j}}$ is the Lasso version of the multiple correlation coefficient when regressing $\mathbf{y}$ on $\mathbf{X}_{A\setminus j}$.

Shrinkage parameter and partial correlation

We interpret the chosen shrinkage values with respect to the partial correlation coefficient.

- We perform multiple Lasso regression using the shrinkage parameter derived from the $\mathrm{BF}_{\mathrm{un}}$ for a given threshold ρ_t.
- We find the partial correlation that corresponds to $\mathrm{BF}_m = 1$ for the chosen λ.
- We identify the threshold value of the partial correlation that is actually used when imposing the chosen λ.

Thus,

$$(1-pr_{j,t}^2)\left[(1-R^2_{\mathbf{X}_j|\mathbf{X}_{A\setminus j}})(1-R^{(lasso)2}_{\mathbf{y}|\mathbf{X}_{A\setminus j}})\right]^{1/df} = 1-\left(\rho_t-\frac{\lambda}{n-1}sg\right)^2,$$

where $pr_{j,t}^2$ is the threshold for $\mathrm{corr}^{(\mathrm{lasso})2}(\mathbf{y},\mathbf{X}_j|\mathbf{X}_{A\setminus j})$ and sg is the sign of the $\widehat{\beta}_j$ on the simple linear model.

Shrinkage parameter and partial correlation (cont.)

From the above it is deduced that

- the threshold value for the Lasso partial correlation is upper bounded by a penalized expression of the corresponding threshold value of the Pearson correlation.

$$pr_{j,t}^2 \leq \left(\rho_t - \frac{\lambda}{n-1} sg \right)^2 ,$$

- the two thresholds are approximately equal as $n \rightarrow \infty$ (i.e. for large sample sizes)

$$pr_{j,t}^2 \xrightarrow{n \rightarrow \infty} \rho_t^2 .$$

What is next?

- Perform multiple Lasso regression by imposing the chosen values for the shrinkage parameters.

- Define areas for the shrinkage levels around the chosen values and impose hyperiors on these areas.

- Impose different shrinkage levels for each variable (Adaptive Lasso).

Simulation study 1

We perform the Bayesian Lasso on the simulated data from Dellaportas et. al. (2002) for specific values of λ that have been chosen through the univariate Bayes factor.

			Var. incl.	Post. Incl. Prob	Prob. of model	
ρ_t	BF	λ		$\mathbf{X}_4,\ \mathbf{X}_5,\ \mathbf{X}_{12}$	MAP	true
0.35	1	0.217	$\mathbf{X}_4, \mathbf{X}_5$	0.96, 1.00, 0.38	26.22%	
0.40	1	0.067	$\mathbf{X}_4, \mathbf{X}_5$	0.85, 1.00, 0.15	55.49%	
0.50	1	0.004	$\mathbf{X}_5$	0.45, 0.96, 0.01	50.45%	43.33%

The observed Pearson and partial correlation coefficients for this data set (in absolute value)

	$\mathbf{X}_2$	$\mathbf{X}_4$	$\mathbf{X}_5$	$\mathbf{X}_6$	$\mathbf{X}_8$	$\mathbf{X}_9$	$\mathbf{X}_{10}$	$\mathbf{X}_{11}$	$\mathbf{X}_{12}$	$\mathbf{X}_{15}$
$\text{corr}(\mathbf{y}, \mathbf{X}_j)$	0.03	**0.38**	**0.58**	0.01	0.08	0.06	0.02	0.03	0.22	0.10
$\text{corr}^{(\text{lasso})}(\mathbf{y}, \mathbf{X}_j \mid \mathbf{X}_{\setminus j})$	0.11	**0.51**	**0.68**	0.18	0.16	0.22	0.13	0.16	0.34	0.28

Simulation study 2

- Consists of 15 covariates and 50 observations (Nott and Kohn, 2005).
- The first 10 variables follow independent $N(0,1)$.
- The last 5 are generated using the following scheme

$$(\mathbf{X}_{11},\dots,\mathbf{X}_{15}) = (\mathbf{X}_1,\dots,\mathbf{X}_5)\times(0.3,0.5,0.7,0.9,1.1)^T\times(1,1,1,1,1)+\mathbf{E},$$

  where E consists of 5 independent $N(0,1)$.
- The response is generated as

$$\mathbf{y} = 2\mathbf{X}_1 - \mathbf{X}_5 + 1.5\mathbf{X}_7 + \mathbf{X}_{11} + 0.5\mathbf{X}_{13} + \epsilon, \quad \text{where} \quad \epsilon \sim N(0, 2.5^2 I).$$

			Var. incl.	Post. Incl. Prob	Prob. of model	
ρ_t	BF	λ		$\mathbf{X}_1$, $\mathbf{X}_5$, $\mathbf{X}_7$, $\mathbf{X}_{11}$, $\mathbf{X}_{13}$	MAP	true
0.35	1	0.217	$\mathbf{X}_1,\mathbf{X}_7,\mathbf{X}_{11}$	$1.00,(0.24),1.00,0.97,(0.19)$	20.07%	1.47%
0.40	1	0.067	$\mathbf{X}_1,\mathbf{X}_7,\mathbf{X}_{11}$	$1.00,(0.09),1.00,0.96,(0.08)$	57.38%	0.45%
0.50	1	0.004	$\mathbf{X}_1,\mathbf{X}_7,\mathbf{X}_{11}$	$1.00,(0.01),0.99,0.91,(0.04)$	88.50%	0.00%

The observed Pearson and partial correlation coefficients for this data set

	$\mathbf{X}_1$	$\mathbf{X}_5$	$\mathbf{X}_7$	$\mathbf{X}_{11}$	$\mathbf{X}_{13}$
$\text{corr}(\mathbf{y},\mathbf{X}_j)$	**0.56**	0.15	**0.34**	**0.56**	0.50
$\text{corr}^{(\text{lasso})}(\mathbf{y},\mathbf{X}_j\|\mathbf{X}_{\setminus j})$	**0.50**	0.27	**0.67**	**0.49**	0.18

Simulation study 2 (cont.)

We have simulated 100 samples and we perform the Bayesian Lasso regression for the chosen levels of shrinkage.

The average posterior inclusion probabilities for each variable are:

			Post. Incl. Prob
ρ_t	BF	λ	$X_1,\ X_5,\ X_7,\ X_{11},\ X_{13}$
$\rho = 0.35$	1	0.218	$0.99, 0.37, 0.91, 0.83, 0.40$
$\rho = 0.40$	1	0.067	$0.98, 0.24, 0.84, 0.78, 0.27$
$\rho = 0.50$	1	0.004	$0.83, 0.07, 0.51, 0.57, 0.18$

The following Table shows the frequency that three selected models are the MAP model in our Bayesian Lasso procedure.

ρ_t	BF	λ	X_1, X_7, X_{11}	X_1, X_5, X_7, X_{11}	$X_1, X_5, X_7, X_{11}, X_{13}$
0.35	1	0.218	27%	11%	6%
0.40	1	0.067	43%	9%	4%
0.50	1	0.004	30%	3%	0%

Diabetes example

- Contains 10 baseline variables, age, sex, body mass index, average blood pressure and six blood serum measurements.
- The response is a one year measure of disease progression for 442 diabetes patients.
- For $n = 442$, the non-important set of the ρ is $(0, 0.11)$.

Bayesian Lasso regression for various choices of λ.

ρ_t	BF	λ	age	sex	bmi	bp	tc	ldl	hdl	tch	ltg	glu
0.15	1	0.110	0.01	**0.89**	**1.00**	**1.00**	0.45	0.31	**0.59**	0.10	**1.00**	0.02
0.16	1	0.067	0.01	**0.78**	**1.00**	**0.99**	0.26	0.09	**0.71**	0.08	**1.00**	0.01
0.20	1	0.002	0.00	0.00	**1.00**	**0.62**	0.03	0.00	0.02	0.00	**1.00**	0.00
0.30	1	4.41×10^{-6}	0.00	0.00	**1.00**	0.00	0.00	0.00	0.00	0.00	**1.00**	0.00

*The variables in red are the ones included in the MAP model.

The sample Pearson and partial correlations are given below.

	age	sex	bmi	bp	tc	ldl	hdl	tch	ltg	glu
$\text{corr}(Y\mathbf{y}, \mathbf{X}_j)$	0.19	0.04	0.59	0.44	0.21	0.17	0.40	0.43	0.57	0.38
$\text{corr}^{(\text{lasso})}(\mathbf{y}, \mathbf{X}_j \vert \mathbf{X}_{\setminus j})$	0.01	**0.19**	**0.35**	**0.23**	0.09	0.07	**0.02**	0.05	**0.21**	0.05
$\text{corr}^{(\text{lasso})}(\mathbf{y}, \mathbf{X}_j \vert \mathbf{X}_{A\setminus j})$		**0.18**	**0.36**	**0.24**			**0.21**		**0.33**	

Conclusions

- We perform both shrinkage and variable selection using the Bayesian version of the Lasso.
- We can interpret the behaviour of our Bayesian Lasso based on levels of practical significance of correlations and partial correlations.
- We propose an approach to select the prior (shrinkage) parameter λ based on its effect on Bayes factors.
- We specify shrinkage values that we can interpret.
- We can select shrinkage values that do not depend on the sample size.

Further work

- We have specified an active area for λ, which truncates the range of λ in order to avoid the Bartlet-Lindley paradox and over-shrinkage.
- Extend to hierarchical model, where a hyperprior is imposed on the shrinkage parameter using its active area.
- Allow different shrinkage parameters for each covariate (Adaptive Lasso).
- Perform the proposed ideas on the Bayesian ridge regression.
- Extend them for GLMs.
- Examine the Bayesian implementation of other related methods such as Elastic net.

References

- Dellaportas, P. and Forster, J. and Ntzoufras, I. (2002). *On Bayesian Model and Variable Selection Using MCMC*. Statistics and Computing. 12:27-36.
- Efron, B., Hastie, T., Johnstone, I., and Tibshirani, R. (2004). *Least angle regression*. Annals of Statistics, 32:407499.
- Green, P. (1995) *Reversible Jump Markov Chain Monte Carlo Computation and Bayesian Model Determination*. Biometrika. 82:711-732.
- Hans, C. (2009) *Bayesian Lasso Regression*. Biometrika. 96:835-845.
- Hans, C. (2009) *Model uncertainty and variable selection in Bayesian lasso regression*. Statis- tics and Computing, 20:221-229.
- Kass, R. and Raftery, A.(1995). *Bayes Factors*. Journal of the American Statistical Association. 90: 773-795.
- Kuo, L. and Mallick, B.(1998). *Variable Selection for Regression Models*. The Indian Journal of Statistics. 60: 65-81.
- Lykou, A. and Whittaker, J. (2010). *Sparse canonical correlation analysis by using the lasso*. Computational Statistics and Data Analysis, 54:3144-3157.
- Meier, L., Van de Geer, S., and Bhlmann, P. (2008). *The group lasso for logistic regression*. Journal of the Royal Statistical Society Series B, 70:53-71.
- Nott, D. and Kohn, R. (2005) *Adaptive sampling for Bayesian variable selection*. Biometrika. 92:747-763.

References

Park, T. and Casella, G. (2008). *The Bayesian Lasso*. Journal of the American Statistical Association. 103:681-687.

Tibshirani, R. (1996). *Regression shrinkage and selection via the lasso*. J. Royal. Statist. Soc. B. 58:267-288.

Yuan, M. and Lin, Y. (2005) *Efficient Empirical Bayes Variable selection and Estimation in Linear Models*. Journal of American Statistical Association 100.

Zou, H. (2006). *The adaptive lasso and its oracle properties*. Journal of the American Statistical Association, 101(476):14181429.

Zou, J. and Hastie, T. (2005). *Regularization and variable selection via the elastic net*. Journal of the Royal Statistical Society Series B, 67:301320.

Specification of λ (cont.)

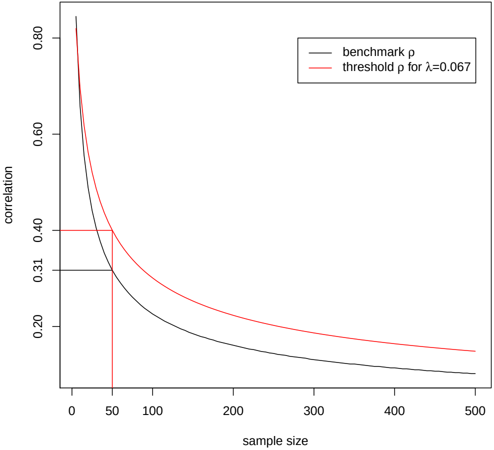


Figure: Benchmark and threshold correlations against the sample size.