Empirical properties of the Lasso by simulations Part 1

Franck Picard*

**Laboratoire de Biométrie et Biologie Évolutive*, Univ. Lyon 1

ECAS, High Dimensional Statistics Course, October 2017

Objectives

- Study empirical properties of the Lasso using simulations
- Learn how to set up numerical simulations, often needed when using or developing methods
- Simulations are essential:
  i) study the properties of a method in a controlled numerical environment,
  ii) assess the robustness to hypothesis,
  iii) fairly compare methods

When (can) should we use the Lasso ? Is it the "best" method ? A "better" method ?

Numerical simulations in a controlled environment

- What are the "empirical properties" that we want to study ?
  - $\rightarrow$ Estimation quality,
  - $\rightarrow$ Prediction accuracy
  - $\rightarrow$ Model selection accuracy
- What is a controlled environment ?
  - $\rightarrow$ List hypothesis on which the method is built
  - $\rightarrow$ Ex: distribution, dimension, dependencies
  - $\rightarrow$ Turn the buttons using an experimental design

Practice, Part 1

1. Generate simulated data
2. Compute the Lasso using glmnet
3. Assess the performance of the Lasso
4. Set up an experimental design for simulations
5. Run first simulations and interpret the results

Simulations set up

1) Simulation of observations

- We consider the Gaussian regression model such that

$$Y_i = x_i^T \beta^* + \varepsilon_i,\ \varepsilon_i \sim \mathcal{N}(0, \sigma^2),$$

- $Y \in \mathbb{R}^n$, et $\beta^* \in \mathbb{R}^p$ with p_0 non null elements.
- J_0: the set of non null positions in β^*:

$$J_0 = \{j \in \{1, ..., p\}, \beta_j^* \neq 0\}$$

- For the sake of simplicity, only one distinct non null value in β^*:

$$\beta^* = \beta_0^* \times (1, \ldots, 1, 0, \ldots 0).$$

- We consider independent covariates with variance σ_X^2

The Lasso estimator in practice with glmnet

glmnet package page

- By Friedman, Hastie, Tibshirani and Simon (Stanford University)
- Gaussian regression, generalized linear models and cox models
- Lasso ($\alpha = 1$), Ridge Regression ($\alpha = 0$), Elastic Net ($\alpha \in]0, 1[$)

$$\widehat{\beta}_\lambda = \min_{\beta_0, \beta} \frac{1}{N} \sum_{i=1}^{N} w_i l(y_i, \beta_0 + \beta^T x_i) + \lambda \left[(1-\alpha)||\beta||_2^2/2 + \alpha||\beta||_1 \right],$$

- λ values are chosen within a sequence $[\lambda_{\text{min}}, \lambda_{\text{max}}]$
- λ_{max}: the smallest λ above which $\widehat{\beta}_\lambda = 0_p$
- λ will be chosen by cross-validation

2) Computing the Lasso with glmnet

Assessing the performance of the Lasso

▸ 3) Quality of estimation

- **Estimation** : Bias and Mean Square Error $\mathbb{E}\left[\widehat{\beta}_\lambda\right]$, $\mathbb{E}\left[(\widehat{\beta}_\lambda - \beta^*)^2\right]$
- **Model Selection** : using binary rule
  - a "positive" coefficient is a non-null coefficient
  - a "negative" coefficient is a null coefficient
  - criteria based on J_0 and $\widehat{J_0}$
- **Prediction** $\mathbb{E}\left(\|Y_0 - X\widehat{\beta}_\lambda\|^2\right)$, with Y_0 a new observation

Computing Bias and Variance

- if `nbsimul` simulations are performed, with new observations $Y^{(h)}$ and a new estimator $\widehat{\beta}_\lambda^{(h)}$
- Bias and Mean Square Error are estimated by:

$$
\begin{aligned}
\widehat{\mathrm{B}} &= \frac{1}{\mathrm{nbsimul}} \sum_{h=1}^{\mathrm{nbsimul}} \sum_{j=1}^{p} \left(\widehat{\beta}_{j,\lambda^{(h)}} - \beta_j^* \right) \\
\widehat{\mathrm{MSE}} &= \frac{1}{\mathrm{nbsimul}} \sum_{h=1}^{\mathrm{nbsimul}} \sum_{j=1}^{p} \left(\widehat{\beta}_{j,\lambda^{(h)}} - \beta_j^* \right)^2
\end{aligned}
$$

Model selection ‣ 4) Quality of selection

- **Estimated dimension**: $|\widehat{J_0}|$
- **Accuracy**: proportion of correctly selected coefficients

$$\frac{|J_0 \cap \widehat{J_0}| + |J_0^c \cap \widehat{J_0^c}|}{p}$$

- **Sensitivity**: proportion of true positives (selected) among positives

$$\frac{|J_0 \cap \widehat{J_0}|}{|J_0 \cap \widehat{J_0}| + |J_0 \cap \widehat{J_0^c}|}$$

- **Specificity**: proportion of true negatives (non selected) among negatives

$$\frac{|J_0^c \cap \widehat{J_0^c}|}{|J_0^c \cap \widehat{J_0^c}| + |J_0^c \cap \widehat{J_0}|}$$

Prediction Error

5) Quality of prediction

- In practice there is no "new" observation, so we need to use cross-validation
- K-fold CV: randomly split $\{1, \ldots, n\}$ into K independent samples `split(sample(n),seq(1,n,by=K))`
- Run the Lasso on the subsample: $\widehat{\beta}_{\lambda}^{(-k)}$
- Assess the prediction error using the data that were not used for estimation Y^k

Influencing Factors

5) Running first simulations

- n: to study the asymptotic / non asymptotic frameworks
- p: to study high dimension or low-dimension performance

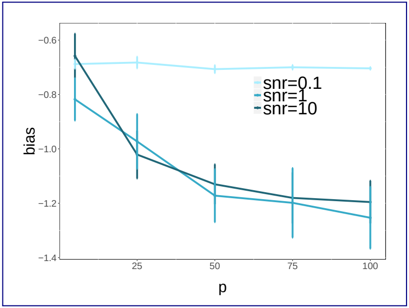
$\rightarrow$ We fix n and modify the n/p ratio

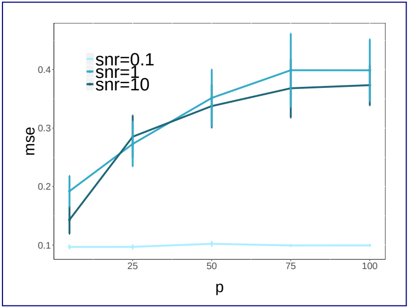
- p_0: the sparsity of β^*, that is unknown in practice
- β^* and σ to assess the impact of the signal to noise ratio

$$\mathrm{SNR} = \frac{\mathbb{E}_X\left(\mathbb{E}(Y_i|\mathbf{x}_i)^2\right)}{\mathbb{E}_X\left(\mathbb{V}(Y_i|\mathbf{x}_i)\right)} = \left(\frac{\sigma_X}{\sigma_\varepsilon}\beta_0^*\right)^2 p_0.$$

Consider Low/High SNR with growing dimension p, p_0 fixed

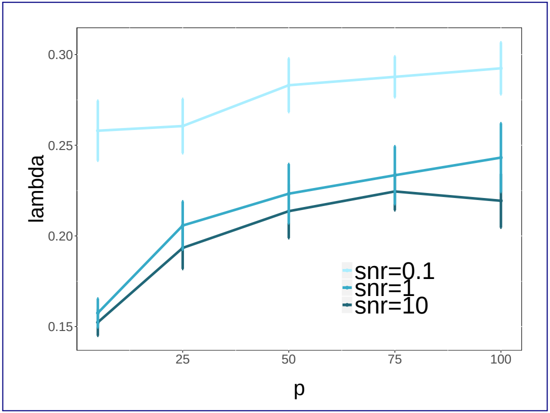
Estimation quality





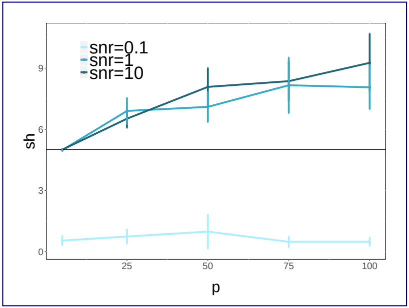
The quality of estimation decreases when p increases

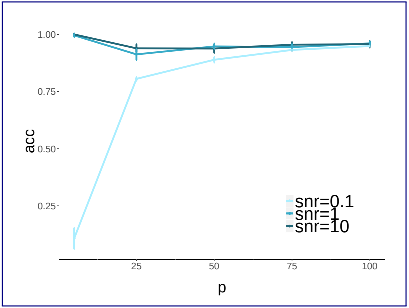
Shrinkage



Shrinkage (λ) increases when p increases

Model selection and accuracy

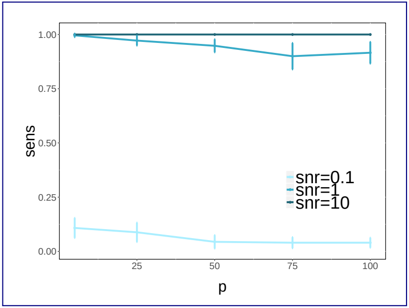


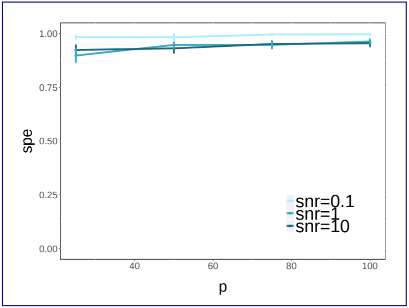


The Lasso over-estimates the dimension of the model when λ is calibrated by cross validation. The overall accuracy of selection is excellent (when there is some signal)

S

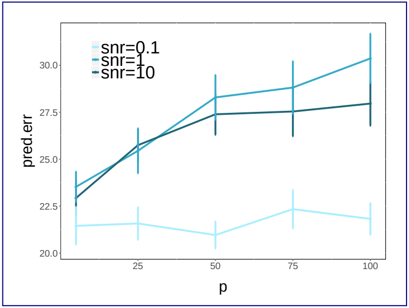
Sensitivity and Specificity

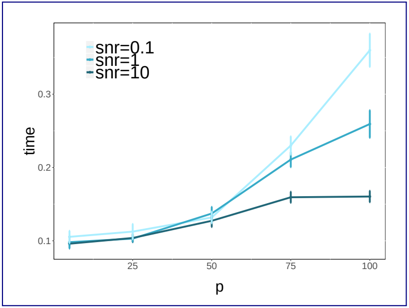




The sensitivity of detection (with signal) decreases when p increases (false positives issue). The overall specificity is excellent

Prediction and Time of execution





The prediction error increases with p as well as the computational complexity (which depends on the optimization algorithm)

Conclusion Part 1

- The performance of the Lasso are "good" when $p < n$
- The bias should be corrected for interpretation purposes
- Model selection could be also improved
- The quality of the detection is good (Acc, Sens, Spe)
- Prediction should be used cautiously, regarding p

⋆ The Lasso, the "best" ? What about competing methods ?

⋆ All results depend on the calibration of λ