Bootstapping Lasso Estimators

Alberto Quaini

June 2017

Presentation

- *Presentation of:*
  "Bootstrapping Lasso Estimator" – A. Chatterjee, S. N. Lahiri [2011], JASA.
- *for PEF UNISG course:*
  "Resampling methods and forecasting" – L. Camponovo
- *Additional literature:*
  1
  2

Outline

Introduction

Linear regression model with iid errors:

$$y_i = x_i^T \beta + \epsilon_i, \qquad i = 1, ..., n \tag{1}$$

Lasso estimator:

$$\hat{\beta}_n = argmin_{u \in R^p} \sum_{i=1}^{n} (y_i - x_i^T u)^2 + \lambda_n \sum_{j=1}^{p} |u_j| \tag{2}$$

- estimation and variable selection method (Tibshirani [1996])
- computationally feasible (Friedman et Al. [2007])
- model consistency (Wainwright [2006], Zhao and Yu [2006] and Zou [2006])
- estimation consistency (Knight and Fu [2000])

Introduction

Problems:

1. Consistency
   - Knight and Fu [2000] show that the limiting distribution of the Lasso estimator is complicated
   - in practice alternative approximations are needed to carry on inference for the Lasso
   - The two autors consider the residual-based Bootstrap method
   - Chatterjee and Lahiri [2010] show that the Bootstrapped Lasso estimator is inconsistent whenever at least one component of the parameter vector is zero
2. Confidence intervals and testing
   - proposals of Tibshirani [1996] and Osborne et Al. [2000] have the drawback of considering the Lasso an approximately linear transformation
   - proposals of Tibshirani [1996], Fan and Li [2001] and Fan and Peng [2004] only provide CI for underlying non-zero parameters

Introduction

Results and proposals in Chatterjee and Lahiri [2011]:

1. Consistency
   - construct a suitable modification to the residual-based Bootstrap
   - show consistency under mild regularity conditions even when some of the underlying parameters are zero
2. Confidence interval and testing
   - the modified Bootstrap method provide consistent estimate of the variance of the Lasso estimator for both zero and non-zero parameter components

Introduction

3. choice of the regularization parameter λ_n
    - accuracy of the lasso critically depends on the regularization parameter
    - the modified Bootstrap is consistent for the MSE of the Lasso
    - the modified Bootstrap estimator of the MSE can be used for the choice of λ_n
4. Adaptive Lasso estimator (Zou [2006])
    - adaptive weights are used for penalizing different coefficients in the L_1 penalty
    - it enjoys the oracle property, i.e. performs as well as if the true underlying model were given in advance
    - the authors show that the simple residual Bootstrap can consistently estimate the distribution of the adaptive Lasso estimator

The Modified Bootstrap method

background and motivation

The residual Bootstrap method (standard in linear regression setting with nonrandom x_i, see Efron [1979], Freedman [1981]) proceeds as follows in the context of the Lasso (Knight and Fu [2000]):

1. Consider the set of centered residuals $E = \{e_i = \tilde{e}_i - \bar{e}, \text{for } i = 1, ..., n\}$, where $\bar{e} = n^{-1} \sum_i \tilde{e}_i$ and $\tilde{e}_i$'s are the residuals of the Lasso fit on the original sample.
2. Construct B bootstrap samples of size n selecting with replacement form E: $E_b^* = \{e_{i,b}^* : i = 1, ..., n\}$ and compute $y_{i,b}^* = x_i^T \hat{\beta}_n + e_{i,b}^*$, for $i = 1, ..., n$, and $b = 1, ..., B$, where $\hat{\beta}_n$ is the Lasso estimator for the original sample.

The Modified Bootstrap method

3. Compute the bootstrap version of $T_n = n^{1/2}(\hat{\beta}_n - \beta)$, i.e. $T_n^* = n^{1/2}(\hat{\beta}^*_{n,b} - \hat{\beta}_n)$, where $\hat{\beta}^*_{n,b}$ is the Lasso estimator for bootstrap sample b.
4. The residual Bootstrap estimator of the distribution G_n of T_n is $\hat{G}_n(B) = P_*(T_n^* \in B)$, where $B \in \mathscr{B}(R^p)$ and P_* is the probability of T_n^* given errors ϵ_i's.

Chatterjee and Lahiri [2010] show that:

- the estimators of the zero parameters fail to capture the target sign value, which is zero
- hence $\hat{G}n$, instead of converging to the deterministic limit of Gn converges weakly to a random probability measure
- i.e. it fails to provide a valid approximation to G_n

The Modified Bootstrap method

A Modified Bootstrap method

Objective: capture the signs of the parameters, expecially the zero components, with probability tending to 1, as the sample size n goes to infinity.

Idea: force components of the Lasso estimator $\hat{\beta}_n$ to be exactly zero whenever they are close to zero using the fact that the Lasso estimator is root-n consistent.

To this end:

1. Form a sequence $\{a_n\}$ of real numbers such that $a_n + (n^{-1/2}log(n))a_n^{-1} \to 0$ asymptotically.
2. Threshold the components of the Lasso estimator β_n at a_n, and define the modified Lasso estimator

$$\tilde{\beta}_{n,j} = \beta_{n,j}\mathbb{1}(\beta_{n,j} \geq a_n), \text{for } j = 1, ..., p. \qquad (3)$$

The Modified Bootstrap method

Note that with probability tending to 1 (as $n \to \infty$):

- $|\hat{\beta}_{n,j}| = |\beta_j| + O(n^{-1/2}) > |\beta_j|/2 \geq a_n$, for a nonzero component β_j
- $|\hat{\beta}_{n,j}| = |\beta_j| + O(n^{-1/2}) = O(n^{-1/2}) \in [-a_n, a_n]$, for a zero component β_j

Then proceed as before:

3. Consider the set of centered residuals $R = \{r_i = \tilde{r}_i - \bar{r}, \text{for } i = 1, ..., n\}$, where $\bar{r} = n^{-1} \sum_i \tilde{r}_i$ and $\tilde{r}_i$'s are the residuals of the modified Lasso fit on the original sample.
4. Construct B bootstrap samples of size n selecting with replacement form R: $R_b^{**} = \{r_{i,b}^{**} : i = 1, ..., n\}$ and compute $y_{i,b}^{**} = x_i^T \tilde{\beta}_n + r_{i,b}^{**}$, for $i = 1, ..., n,$ and $b = 1, ..., B$, where $\tilde{\beta}_n$ is the modified Lasso estimator for the original sample.

The Modified Bootstrap method

5. Compute the bootstrap version of $T_n = n^{1/2}(\hat{\beta}_n - \beta)$, i.e. $T_n^{**} = n^{1/2}(\hat{\beta}_{n,b}^{**} - \tilde{\beta}_n)$, where $\hat{\beta}_{n,b}^{**}$ is the Lasso estimator for bootstrap sample b.
6. The residual Bootstrap estimator of the distribution G_n of T_n is $\tilde{G}_n(B) = P_{**}(T_n^{**} \in B)$, where $B \in \mathscr{B}(R^p)$ and P_{**} is the probability of T_n^{**} given errors ϵ_i's.

Remarks:

- Centering the residuals ensures the Bootstrap analogue of the condition $E[e_i] = 0$
- A rescaling factor $(1 - p/n) - 1/2$ is sometimes used (see Efron [1982]) to improve finite sample accuracy
- It is possible to replace $\hat{\beta}_n$ by any other $\sqrt{n}$-consistent estimator of β, e.g. least squares

Bootstrapping the Lasso estimator

Consistency and the distributional approximation

Theorem 1: Consistency of modified Bootstrap

Assume:

- (C1) $n^{-1} \sum_i x_i {x_i}^T \to C$, p.d. matrix. Furthemore $n^{-1} \sum_i \|x_i\|^3 \to O(1)$.
- (C2) $\lambda_n n^{-1/2} \to \lambda_0 \geq 0$.
- (C3) errors ϵ_i's are iid with $E[\epsilon_i] = 0$ and $VAR[\epsilon_i] = \sigma^2 < \infty$.

Then:

$$\mathscr{P}(\tilde{G}_n, G_n) \to 0, \text{as } n \to \infty, \text{with probability } 1,$$

where $\mathscr{P}(\cdot, \cdot)$ denotes the Prohorov probability metric.

Bootstrapping the Lasso estimator