Validity First, Optimality Next

An Approach to Valid Post-selection Inference in Assumption-lean Linear Regression

Arun Kumar Kuchibhotla[1]

Department of Statistics
University of Pennsylvania

WHOA-PSI 2017

Working Draft Available at:
https://statistics.wharton.upenn.edu/profile/arunku/#research

[1] Joint work with Wharton Group, including, Larry Brown, Andreas Buja, Richard Berk, Ed George, Linda Zhao.

Outline

1. Introduction
   - Assumption-lean Linear Regression
   - Some Notation and Uniform Results
2. Post-selection Inference Problem
   - PoSI-Dantzig
   - PoSI-Lasso
   - PoSI-SqrtLasso
3. Multiplier Bootstrap
4. Some Comments and Extensions

Outline

1. Introduction
   - Assumption-lean Linear Regression
   - Some Notation and Uniform Results
2. Post-selection Inference Problem
   - PoSI-Dantzig
   - PoSI-Lasso
   - PoSI-SqrtLasso
3. Multiplier Bootstrap
4. Some Comments and Extensions

Linear Regression

- Suppose $(X_i, Y_i) \in \mathbb{R}^p \times \mathbb{R}$ $(1 \leq i \leq n)$ are random vectors available as data.
- We apply linear regression on this data.
- The least squares estimator of "slope" vector is $\hat{\beta}$ given by

$$\hat{\beta} = \left(\mathbb{E}_n[XX^\top]\right)^{-1}(\mathbb{E}_n[XY]).$$

  ($\mathbb{E}_n$ represents empirical mean.)
- If the random vectors satisfy weak law of large numbers, then this estimator converges (in probability) to

$$\beta_0 = \left(\mathbb{E}[XX^\top]\right)^{-1}(\mathbb{E}[XY]).$$

- β_0 has an interpretation as best linear predictor (slope) in terms of squared error loss.

Linear Regression

- Suppose $(X_i, Y_i) \in \mathbb{R}^p \times \mathbb{R}$ $(1 \leq i \leq n)$ are random vectors available as data.
- We apply linear regression on this data.
- The least squares estimator of "slope" vector is $\hat{\beta}$ given by

$$\hat{\beta} = \left(\mathbb{E}_n[XX^\top]\right)^{-1}(\mathbb{E}_n[XY]).$$

  ($\mathbb{E}_n$ represents empirical mean.)
- If the random vectors satisfy weak law of large numbers, then this estimator converges (in probability) to

$$\beta_0 = \left(\mathbb{E}[XX^\top]\right)^{-1}(\mathbb{E}\,[XY]).$$

- β_0 has an interpretation as best linear predictor (slope) in terms of squared error loss.

Linear Regression

- Suppose $(X_i, Y_i) \in \mathbb{R}^p \times \mathbb{R}$ $(1 \le i \le n)$ are random vectors available as data.
- We apply linear regression on this data.
- The least squares estimator of "slope" vector is $\hat{\beta}$ given by

$$\hat{\beta} = \left(\mathbb{E}_n[XX^\top]\right)^{-1}(\mathbb{E}_n[XY]).$$

  ($\mathbb{E}_n$ represents empirical mean.)
- If the random vectors satisfy weak law of large numbers, then this estimator converges (in probability) to

$$\beta_0 = \left(\mathbb{E}[XX^\top]\right)^{-1}(\mathbb{E}[XY]).$$

- β_0 has an interpretation as best linear predictor (slope) in terms of squared error loss.

Linear Regression

- Suppose $(X_i, Y_i) \in \mathbb{R}^p \times \mathbb{R}$ $(1 \le i \le n)$ are random vectors available as data.
- We apply linear regression on this data.
- The least squares estimator of "slope" vector is $\hat{\beta}$ given by

$$\hat{\beta} = \left(\mathbb{E}_n[XX^\top]\right)^{-1}(\mathbb{E}_n[XY]).$$

($\mathbb{E}_n$ represents empirical mean.)

- If the random vectors satisfy weak law of large numbers, then this estimator converges (in probability) to

$$\beta_0 = \left(\mathbb{E}[XX^\top]\right)^{-1}(\mathbb{E}[XY]).$$

- β_0 has an interpretation as best linear predictor (slope) in terms of squared error loss.

Linear Regression

- Suppose $(X_i, Y_i) \in \mathbb{R}^p \times \mathbb{R}$ $(1 \le i \le n)$ are random vectors available as data.
- We apply linear regression on this data.
- The least squares estimator of "slope" vector is $\hat{\beta}$ given by

$$\hat{\beta} = \left(\mathbb{E}_n[XX^\top]\right)^{-1}(\mathbb{E}_n[XY]).$$

  ($\mathbb{E}_n$ represents empirical mean.)
- If the random vectors satisfy weak law of large numbers, then this estimator converges (in probability) to

$$\beta_0 = \left(\mathbb{E}[XX^\top]\right)^{-1}(\mathbb{E}\,[XY]).$$

- β_0 has an interpretation as best linear predictor (slope) in terms of squared error loss.

Some Comments

- It is important to recognize that to use linear regression, we do NOT need any model or distributional assumptions.
- We do NOT even require independence or identical distributions for random vectors. Existence of LLN is enough.
- This perspective also works if the covariate vectors X_i are treated as fixed.
- The target β_0 coincides with the usual model parameter under a true linear model.
- CLT: If the random vectors (X_i, Y_i) are independent with finite fourth moments, then

$$n^{1/2}\left(\hat{\beta} - \beta_0\right) \xrightarrow{\mathcal{L}} N\left(0, J^{-1}KJ^{-1}\right), \quad \leftarrow \textit{efficiency}$$

with $J = \mathbb{E}[XX^\top]$ and $K = \mathbb{E}\left[XX^\top(Y - X^\top\beta_0)^2\right]$.

Some Comments

- It is important to recognize that to use linear regression, we do NOT need any model or distributional assumptions.
- We do NOT even require independence or identical distributions for random vectors. Existence of LLN is enough.
- This perspective also works if the covariate vectors X_i are treated as fixed.
- The target β_0 coincides with the usual model parameter under a true linear model.
- CLT: If the random vectors (X_i, Y_i) are independent with finite fourth moments, then

$$n^{1/2}\left(\hat{\beta} - \beta_0\right) \xrightarrow{\mathcal{L}} N\left(0, J^{-1}KJ^{-1}\right), \quad \leftarrow \textit{efficiency}$$

with $J = \mathbb{E}[XX^\top]$ and $K = \mathbb{E}\left[XX^\top(Y - X^\top\beta_0)^2\right]$.

Some Comments

- It is important to recognize that to use linear regression, we do NOT need any model or distributional assumptions.
- We do NOT even require independence or identical distributions for random vectors. Existence of LLN is enough.
- This perspective also works if the covariate vectors X_i are treated as fixed.
- The target β_0 coincides with the usual model parameter under a true linear model.
- CLT: If the random vectors (X_i, Y_i) are independent with finite fourth moments, then

$$n^{1/2}\left(\hat{\beta} - \beta_0\right) \xrightarrow{\mathcal{L}} N\left(0, J^{-1}KJ^{-1}\right), \quad \leftarrow \textit{efficiency}$$

with $J = \mathbb{E}[XX^\top]$ and $K = \mathbb{E}\left[XX^\top(Y - X^\top\beta_0)^2\right]$.

Some Comments

- It is important to recognize that to use linear regression, we do NOT need any model or distributional assumptions.
- We do NOT even require independence or identical distributions for random vectors. Existence of LLN is enough.
- This perspective also works if the covariate vectors X_i are treated as fixed.
- The target β_0 coincides with the usual model parameter under a true linear model.
- CLT: If the random vectors (X_i, Y_i) are independent with finite fourth moments, then

$$n^{1/2}\left(\hat{\beta} - \beta_0\right) \xrightarrow{\mathcal{L}} N\left(0, J^{-1}KJ^{-1}\right), \quad \leftarrow \textit{efficiency}$$

with $J = \mathbb{E}[XX^\top]$ and $K = \mathbb{E}\left[XX^\top(Y - X^\top\beta_0)^2\right]$.

Some Comments

- It is important to recognize that to use linear regression, we do NOT need any model or distributional assumptions.
- We do NOT even require independence or identical distributions for random vectors. Existence of LLN is enough.
- This perspective also works if the covariate vectors X_i are treated as fixed.
- The target β_0 coincides with the usual model parameter under a true linear model.
- CLT: If the random vectors (X_i, Y_i) are independent with finite fourth moments, then

$$n^{1/2}\left(\hat{\beta} - \beta_0\right) \xrightarrow{\mathcal{L}} N\left(0, J^{-1}KJ^{-1}\right), \quad \leftarrow \textit{efficiency}$$

with $J = \mathbb{E}[XX^\top]$ and $K = \mathbb{E}\left[XX^\top(Y - X^\top\beta_0)^2\right]$.

Outline

1. Introduction
   - Assumption-lean Linear Regression
   - Some Notation and Uniform Results
2. Post-selection Inference Problem
   - PoSI-Dantzig
   - PoSI-Lasso
   - PoSI-SqrtLasso
3. Multiplier Bootstrap
4. Some Comments and Extensions

Some Notation

- We have p covariates and *model selection* refers to choosing a subset $M \subseteq \{1, 2, \ldots, p\}$ stating which covariates to consider in the model.
- For any vector $v \in \mathbb{R}^p$ and model M, the vector $v(M) \in \mathbb{R}^{|M|}$ denotes the sub-vector of v with indices in M.
- For instance, if $v = (1, 0, 5, -1)$, $M = \{2, 4\}$ and $p = 4$, then $v(M) = (0, -1)$.
- Let for any model $M \subseteq \{1, 2, \ldots, p\}$,

$$\Sigma_n(M) := \mathbb{E}_n\left[X(M)X^\top(M)\right] \quad \text{and} \quad \Sigma(M) := \mathbb{E}\left[X(M)X^\top(M)\right].$$

$$\Gamma_n(M) := \mathbb{E}_n\left[X(M)Y\right] \quad \text{and} \quad \Gamma(M) := \mathbb{E}\left[X(M)Y\right].$$

Some Notation

- We have p covariates and *model selection* refers to choosing a subset $M \subseteq \{1, 2, \ldots, p\}$ stating which covariates to consider in the model.
- For any vector $v \in \mathbb{R}^p$ and model M, the vector $v(M) \in \mathbb{R}^{|M|}$ denotes the sub-vector of v with indices in M.
- For instance, if $v = (1, 0, 5, -1)$, $M = \{2, 4\}$ and $p = 4$, then $v(M) = (0, -1)$.
- Let for any model $M \subseteq \{1, 2, \ldots, p\}$,

$$\Sigma_n(M) := \mathbb{E}_n\left[X(M)X^\top(M)\right] \quad \text{and} \quad \Sigma(M) := \mathbb{E}\left[X(M)X^\top(M)\right].$$

$$\Gamma_n(M) := \mathbb{E}_n\left[X(M)Y\right] \quad \text{and} \quad \Gamma(M) := \mathbb{E}\left[X(M)Y\right].$$

Some Notation

- We have p covariates and *model selection* refers to choosing a subset $M \subseteq \{1, 2, \ldots, p\}$ stating which covariates to consider in the model.
- For any vector $v \in \mathbb{R}^p$ and model M, the vector $v(M) \in \mathbb{R}^{|M|}$ denotes the sub-vector of v with indices in M.
- For instance, if $v = (1, 0, 5, -1)$, $M = \{2, 4\}$ and $p = 4$, then $v(M) = (0, -1)$.
- Let for any model $M \subseteq \{1, 2, \ldots, p\}$,

$$\Sigma_n(M) := \mathbb{E}_n \left[X(M) X^\top(M) \right] \quad \text{and} \quad \Sigma(M) := \mathbb{E} \left[X(M) X^\top(M) \right].$$

$$\Gamma_n(M) := \mathbb{E}_n \left[X(M) Y \right] \quad \text{and} \quad \Gamma(M) := \mathbb{E} \left[X(M) Y \right].$$

Some Notation

- We have p covariates and *model selection* refers to choosing a subset $M \subseteq \{1, 2, \ldots, p\}$ stating which covariates to consider in the model.
- For any vector $v \in \mathbb{R}^p$ and model M, the vector $v(M) \in \mathbb{R}^{|M|}$ denotes the sub-vector of v with indices in M.
- For instance, if $v = (1, 0, 5, -1)$, $M = \{2, 4\}$ and $p = 4$, then $v(M) = (0, -1)$.
- Let for any model $M \subseteq \{1, 2, \ldots, p\}$,

$$\Sigma_n(M) := \mathbb{E}_n\left[X(M)X^\top(M)\right] \quad \text{and} \quad \Sigma(M) := \mathbb{E}\left[X(M)X^\top(M)\right].$$

$$\Gamma_n(M) := \mathbb{E}_n\left[X(M)Y\right] \quad \text{and} \quad \Gamma(M) := \mathbb{E}\left[X(M)Y\right].$$

Notation Contd.

- Define the least squares estimator and target as

$$\hat{\beta}_M := \arg\min_{\theta \in \mathbb{R}^{|M|}} \mathbb{E}_n \left[(Y - X^\top(M)\theta)^2 \right]$$

$$\beta_M := \arg\min_{\theta \in \mathbb{R}^{|M|}} \mathbb{E} \left[(Y - X^\top(M)\theta)^2 \right]$$

- No matter whether M is fixed or random, $\hat{\beta}_M$ is estimating β_M. What it means is secondary.
- For any $1 \le k \le p$, define

$$\mathcal{M}(k) = \{M \subseteq \{1, 2, \ldots, p\} \text{ such that } |M| \le k\}.$$

- Define, finally,

$$\mathcal{D}_{2n} := \|\Sigma_n - \Sigma\|_\infty \quad \text{and} \quad \mathcal{D}_{1n} := \|\Gamma_n - \Gamma\|_\infty .$$

- All the rates about $\hat{\beta}_M - \beta_M$ can be bounded in terms of $\mathcal{D}_{1n}$ and $\mathcal{D}_{2n}$ (possibly sub-optimally).

Notation Contd.

- Define the least squares estimator and target as

$$\hat{\beta}_M := \arg\min_{\theta \in \mathbb{R}^{|M|}} \mathbb{E}_n \left[(Y - X^\top(M)\theta)^2 \right]$$

$$\beta_M := \arg\min_{\theta \in \mathbb{R}^{|M|}} \mathbb{E} \left[(Y - X^\top(M)\theta)^2 \right]$$

- No matter whether M is fixed or random, $\hat{\beta}_M$ is estimating β_M. What it means is secondary.
- For any $1 \le k \le p$, define

$$\mathcal{M}(k) = \{M \subseteq \{1, 2, \ldots, p\} \text{ such that } |M| \le k\}.$$

- Define, finally,

$$\mathcal{D}_{2n} := \|\Sigma_n - \Sigma\|_\infty \quad \text{and} \quad \mathcal{D}_{1n} := \|\Gamma_n - \Gamma\|_\infty.$$

- All the rates about $\hat{\beta}_M - \beta_M$ can be bounded in terms of $\mathcal{D}_{1n}$ and $\mathcal{D}_{2n}$ (possibly sub-optimally).

Notation Contd.

- Define the least squares estimator and target as

$$\hat{\beta}_M := \arg\min_{\theta \in \mathbb{R}^{|M|}} \mathbb{E}_n \left[(Y - X^\top(M)\theta)^2 \right]$$

$$\beta_M := \arg\min_{\theta \in \mathbb{R}^{|M|}} \mathbb{E} \left[(Y - X^\top(M)\theta)^2 \right]$$

- No matter whether M is fixed or random, $\hat{\beta}_M$ is estimating β_M. What it means is secondary.
- For any $1 \le k \le p$, define

$$\mathcal{M}(k) = \{M \subseteq \{1, 2, \ldots, p\} \text{ such that } |M| \le k\}.$$

- Define, finally,

$$\mathcal{D}_{2n} := \|\Sigma_n - \Sigma\|_\infty \quad \text{and} \quad \mathcal{D}_{1n} := \|\Gamma_n - \Gamma\|_\infty.$$

- All the rates about $\hat{\beta}_M - \beta_M$ can be bounded in terms of $\mathcal{D}_{1n}$ and $\mathcal{D}_{2n}$ (possibly sub-optimally).

Notation Contd.

- Define the least squares estimator and target as

$$\hat{\beta}_M := \arg\min_{\theta \in \mathbb{R}^{|M|}} \mathbb{E}_n \left[(Y - X^\top(M)\theta)^2 \right]$$

$$\beta_M := \arg\min_{\theta \in \mathbb{R}^{|M|}} \mathbb{E} \left[(Y - X^\top(M)\theta)^2 \right]$$

- No matter whether M is fixed or random, $\hat{\beta}_M$ is estimating β_M. What it means is secondary.
- For any $1 \le k \le p$, define

$$\mathcal{M}(k) = \{M \subseteq \{1, 2, \ldots, p\} \text{ such that } |M| \le k\}.$$

- Define, finally,

$$\mathcal{D}_{2n} := \|\Sigma_n - \Sigma\|_\infty \quad \text{and} \quad \mathcal{D}_{1n} := \|\Gamma_n - \Gamma\|_\infty .$$

- All the rates about $\hat{\beta}_M - \beta_M$ can be bounded in terms of $\mathcal{D}_{1n}$ and $\mathcal{D}_{2n}$ (possibly sub-optimally).

Notation Contd.

- Define the least squares estimator and target as

$$\hat{\beta}_M := \underset{\theta \in \mathbb{R}^{|M|}}{\arg\min}\ \mathbb{E}_n \left[(Y - X^\top(M)\theta)^2 \right]$$

$$\beta_M := \underset{\theta \in \mathbb{R}^{|M|}}{\arg\min}\ \mathbb{E} \left[(Y - X^\top(M)\theta)^2 \right]$$

- No matter whether M is fixed or random, $\hat{\beta}_M$ is estimating β_M. What it means is secondary.
- For any $1 \le k \le p$, define

$$\mathcal{M}(k) = \{M \subseteq \{1, 2, \ldots, p\} \text{ such that } |M| \le k\}.$$

- Define, finally,

$$\mathcal{D}_{2n} := \|\Sigma_n - \Sigma\|_\infty \quad \text{and} \quad \mathcal{D}_{1n} := \|\Gamma_n - \Gamma\|_\infty .$$

- All the rates about $\hat{\beta}_M - \beta_M$ can be bounded in terms of $\mathcal{D}_{1n}$ and $\mathcal{D}_{2n}$ (possibly sub-optimally).

Uniform (in Models) Results

Table: Uniform Rates of Convergence ($\bigvee_k \equiv \sup_{M \in \mathcal{M}(k)}$).

Quantity	Rate Bound	Sub-Gaussianity
$\bigvee_k \left\|\hat{\beta}_M - \beta_M\right\|_1$	$k(\mathcal{D}_{1n} + \mathcal{D}_{2n} S_{1,k})$	$k\sqrt{\log p/n}$
$\bigvee_k \left\|\hat{\beta}_M - \beta_M\right\|_2$	$k^{1/2}\mathcal{D}_{1n} + k\mathcal{D}_{2n} S_{2,k}$	$\sqrt{k \log p/n}$
$\bigvee_k \|\Sigma_n(M) - \Sigma(M)\|_{op}$	$k\mathcal{D}_{2n}$	$\sqrt{k \log p/n}$

Here

$$S_{1,k} := \sup_{M \in \mathcal{M}(k)} \|\beta_M\|_1 \quad \text{and} \quad S_{2,k} := \sup_{M \in \mathcal{M}(k)} \|\beta_M\|_2 .$$

Two Famous Quotes . . .

. . . and two ways of invalidating inference!!

- John. W. Tukey:

  The greatest value of a picture is when it forces us to notice what we never expected to see.

  . . . more emphasis needed to be placed on using data to suggest hypotheses to test.

- George E. P. Box:

  all models are wrong, but some are useful.

Two Famous Quotes . . .

. . . and two ways of invalidating inference!!

- John. W. Tukey:

  The greatest value of a picture is when it forces us to notice what we never expected to see.

  . . . more emphasis needed to be placed on using data to suggest hypotheses to test.

- George E. P. Box:

  all models are wrong, but some are useful.

Two Famous Quotes . . .

. . . and two ways of invalidating inference!!

- John. W. Tukey:

  The greatest value of a picture is when it forces us to notice what we never expected to see.

  . . . more emphasis needed to be placed on using data to suggest hypotheses to test.

- George E. P. Box:

  all models are wrong, but some are useful.

Two Famous Quotes . . .

. . . and two ways of invalidating inference!!

- John. W. Tukey:

  The greatest value of a picture is when it forces us to notice what we never expected to see.

  . . . more emphasis needed to be placed on using data to suggest hypotheses to test.

- George E. P. Box:

  all models are wrong, but some are useful.

Two Famous Quotes . . .

. . . and two ways of invalidating inference!!

- John. W. Tukey:

  The greatest value of a picture is when it forces us to notice what we never expected to see.

  . . . more emphasis needed to be placed on using data to suggest hypotheses to test.

- George E. P. Box:

  all models are wrong, but some are useful.

The PoSI Problem

- We concentrate on the problem of constructing valid confidence regions which readily allows us to get valid tests of hypotheses.
- The PoSI problem is given by

Problem

Construct a set of confidence regions (depending on α)

$$\{\hat{\mathcal{R}}_M : M \in \mathcal{M}\}$$

for some non-random set of models, $\mathcal{M}$, such that for any random model $\hat{M}$ with $\mathbb{P}(\hat{M} \in \mathcal{M}) \to 1$,

$$\liminf_{n\to\infty} \mathbb{P}\left(\beta_{\hat{M}} \in \hat{\mathcal{R}}_{\hat{M}}\right) \geq 1 - \alpha.$$

The PoSI Problem

- We concentrate on the problem of constructing valid confidence regions which readily allows us to get valid tests of hypotheses.
- The PoSI problem is given by

Problem

Construct a set of confidence regions (depending on α)

$$\{\hat{\mathcal{R}}_M : \; M \in \mathcal{M}\}$$

for some non-random set of models, $\mathcal{M}$, such that for any random model $\hat{M}$ with $\mathbb{P}(\hat{M} \in \mathcal{M}) \to 1$,

$$\liminf_{n\to\infty} \; \mathbb{P}\left(\beta_{\hat{M}} \in \hat{\mathcal{R}}_{\hat{M}}\right) \geq 1 - \alpha.$$

The PoSI Problem

- We concentrate on the problem of constructing valid confidence regions which readily allows us to get valid tests of hypotheses.
- The PoSI problem is given by

Problem

Construct a set of confidence regions (depending on α)

$$\{\hat{\mathcal{R}}_M : M \in \mathcal{M}\}$$

for some non-random set of models, $\mathcal{M}$, such that for any random model $\hat{M}$ with $\mathbb{P}(\hat{M} \in \mathcal{M}) \to 1$,

$$\liminf_{n\to\infty} \mathbb{P}\left(\beta_{\hat{M}} \in \hat{\mathcal{R}}_{\hat{M}}\right) \geq 1 - \alpha.$$

The Basic Lemma

Lemma

For any set of confidence regions

$$\{\hat{\mathcal{R}}_M : M \in \mathcal{M}\},$$

and any random model $\hat{M}$ satisfying $\mathbb{P}(\hat{M} \in \mathcal{M}) = 1$, the following lower bound holds:

$$\mathbb{P}\left(\beta_{\hat{M}} \in \hat{\mathcal{R}}_{\hat{M}}\right) \geq \mathbb{P}\left(\bigcap_{M \in \mathcal{M}} \left\{\beta_M \in \hat{\mathcal{R}}_M\right\}\right).$$

- We will provide confidence regions such that the right hand probability is bounded below by $1 - \alpha$ (asymptotically).
- This Lemma is also the basis for valid PoSI in (fixed covariate) setting of Berk et al. (2013).

The Basic Lemma

Lemma

For any set of confidence regions

$$\{\hat{\mathcal{R}}_M : M \in \mathcal{M}\},$$

and any random model $\hat{M}$ satisfying $\mathbb{P}(\hat{M} \in \mathcal{M}) = 1$, the following lower bound holds:

$$\mathbb{P}\left(\beta_{\hat{M}} \in \hat{\mathcal{R}}_{\hat{M}}\right) \geq \mathbb{P}\left(\bigcap_{M \in \mathcal{M}} \left\{\beta_M \in \hat{\mathcal{R}}_M\right\}\right).$$

- We will provide confidence regions such that the right hand probability is bounded below by $1 - \alpha$ (asymptotically).
- This Lemma is also the basis for valid PoSI in (fixed covariate) setting of Berk et al. (2013).

The Basic Lemma

Lemma

For <u>any</u> set of confidence regions

$$\{\hat{\mathcal{R}}_M : M \in \mathcal{M}\},$$

and any random model $\hat{M}$ *satisfying* $\mathbb{P}(\hat{M} \in \mathcal{M}) = 1$*, the following lower bound holds:*

$$\mathbb{P}\left(\beta_{\hat{M}} \in \hat{\mathcal{R}}_{\hat{M}}\right) \geq \mathbb{P}\left(\bigcap_{M \in \mathcal{M}} \left\{\beta_M \in \hat{\mathcal{R}}_M\right\}\right).$$

- We will provide confidence regions such that the right hand probability is bounded below by $1 - \alpha$ (asymptotically).
- This Lemma is also the basis for valid PoSI in (fixed covariate) setting of Berk et al. (2013).

Some More Notation

- Recall

$$\mathcal{D}_{1n} := \|\Gamma_n - \Gamma\|_\infty \quad \text{and} \quad \mathcal{D}_{2n} := \|\Sigma_n - \Sigma\|_\infty .$$

- Define $C_1(\alpha)$ and $C_2(\alpha)$ by

$$\mathbb{P}\left(\mathcal{D}_{1n} \le C_1(\alpha) \quad \text{and} \quad \mathcal{D}_{2n} \le C_2(\alpha)\right) \ge 1 - \alpha.$$

- In the following, we give three different confidence regions that satisfy the validity.
- If $k = o(\sqrt{n/\log p})$, then

$$\liminf_{n\to\infty} \mathbb{P}\left(\bigcap_{M\in\mathcal{M}(k)} \left\{\beta_M \in \hat{\mathcal{R}}_M\right\}\right) \ge 1 - \alpha$$

Some More Notation

- Recall

$$\mathcal{D}_{1n} := \|\Gamma_n - \Gamma\|_\infty \quad \text{and} \quad \mathcal{D}_{2n} := \|\Sigma_n - \Sigma\|_\infty .$$

- Define $C_1(\alpha)$ and $C_2(\alpha)$ by

$$\mathbb{P}\left(\mathcal{D}_{1n} \le C_1(\alpha) \quad \text{and} \quad \mathcal{D}_{2n} \le C_2(\alpha)\right) \ge 1 - \alpha.$$

- In the following, we give three different confidence regions that satisfy the validity.
- If $k = o(\sqrt{n/\log p})$, then

$$\liminf_{n\to\infty} \mathbb{P}\left(\bigcap_{M\in\mathcal{M}(k)} \left\{\beta_M \in \hat{\mathcal{R}}_M\right\}\right) \ge 1 - \alpha$$

Some More Notation

- Recall

$$\mathcal{D}_{1n} := \|\Gamma_n - \Gamma\|_\infty \quad \text{and} \quad \mathcal{D}_{2n} := \|\Sigma_n - \Sigma\|_\infty .$$

- Define $C_1(\alpha)$ and $C_2(\alpha)$ by

$$\mathbb{P}\left(\mathcal{D}_{1n} \le C_1(\alpha) \quad \text{and} \quad \mathcal{D}_{2n} \le C_2(\alpha)\right) \ge 1 - \alpha.$$

- In the following, we give three different confidence regions that satisfy the validity.
- If $k = o(\sqrt{n/\log p})$, then

$$\liminf_{n\to\infty} \mathbb{P}\left(\bigcap_{M\in\mathcal{M}(k)} \left\{\beta_M \in \hat{\mathcal{R}}_M\right\}\right) \ge 1 - \alpha$$

Some More Notation

- Recall

$$\mathcal{D}_{1n} := \|\Gamma_n - \Gamma\|_\infty \quad \text{and} \quad \mathcal{D}_{2n} := \|\Sigma_n - \Sigma\|_\infty .$$

- Define $C_1(\alpha)$ and $C_2(\alpha)$ by

$$\mathbb{P}\left(\mathcal{D}_{1n} \le C_1(\alpha) \quad \text{and} \quad \mathcal{D}_{2n} \le C_2(\alpha)\right) \ge 1 - \alpha.$$

- In the following, we give three different confidence regions that satisfy the validity.
- If $k = o(\sqrt{n/\log p})$, then

$$\liminf_{n\to\infty} \mathbb{P}\left(\bigcap_{M\in\mathcal{M}(k)} \left\{\beta_M \in \hat{\mathcal{R}}_M\right\}\right) \ge 1 - \alpha$$

Outline

1 Introduction
- Assumption-lean Linear Regression
- Some Notation and Uniform Results

2 Post-selection Inference Problem
- PoSI-Dantzig
- PoSI-Lasso
- PoSI-SqrtLasso

3 Multiplier Bootstrap

4 Some Comments and Extensions

PoSI-Dantzig

- Define for any model $M \subseteq \{1, 2 \ldots, p\}$,

$$\hat{\mathcal{R}}_M := \left\{\theta \in \mathbb{R}^{|M|} : \left\|\Sigma_n(M)\left(\hat{\beta}_M - \theta\right)\right\|_\infty \le C_1(\alpha) + C_2(\alpha)\left\|\hat{\beta}_M\right\|_1\right\}.$$

- Note the left hand side is

$$\Sigma_n(M)\left(\hat{\beta}_M - \theta\right) = \frac{1}{n}\sum_{i=1}^{n} X_i(M)\left(Y_i - X_i^\top(M)\theta\right).$$

- Thus, the confidence set is very similar to the constraint set of Dantzig selector and so the name PoSI-Dantzig.
- Validity does not require either independence or identical distribution for the random vectors.
- It allows for diverging number of covariates. If (X_i, Y_i) are independent and sub-Gaussian then for $k = o(\sqrt{n/\log p})$,

$$\textbf{Leb}\left(\hat{\mathcal{R}}_M\right) = O_p\left(\sqrt{\frac{|M|\log p}{n}}\right)^{|M|} \quad \text{uniformly for } M \in \mathcal{M}(k).$$

PoSI-Dantzig

- Define for any model $M \subseteq \{1, 2 \ldots, p\}$,

$$\hat{\mathcal{R}}_M := \left\{\theta \in \mathbb{R}^{|M|} : \left\|\Sigma_n(M)\left(\hat{\beta}_M - \theta\right)\right\|_\infty \le C_1(\alpha) + C_2(\alpha)\left\|\hat{\beta}_M\right\|_1\right\}.$$

- Note the left hand side is

$$\Sigma_n(M)\left(\hat{\beta}_M - \theta\right) = \frac{1}{n}\sum_{i=1}^{n} X_i(M)\left(Y_i - X_i^\top(M)\theta\right).$$

- Thus, the confidence set is very similar to the constraint set of Dantzig selector and so the name PoSI-Dantzig.
- Validity does not require either independence or identical distribution for the random vectors.
- It allows for diverging number of covariates. If (X_i, Y_i) are independent and sub-Gaussian then for $k = o(\sqrt{n/\log p})$,

$$\text{Leb}\left(\hat{\mathcal{R}}_M\right) = O_p\left(\sqrt{\frac{|M|\log p}{n}}\right)^{|M|} \quad \text{uniformly for } M \in \mathcal{M}(k).$$

PoSI-Dantzig

- Define for any model $M \subseteq \{1, 2 \ldots, p\}$,

$$\hat{\mathcal{R}}_M := \left\{\theta \in \mathbb{R}^{|M|} : \left\|\Sigma_n(M)\left(\hat{\beta}_M - \theta\right)\right\|_\infty \le C_1(\alpha) + C_2(\alpha)\left\|\hat{\beta}_M\right\|_1\right\}.$$

- Note the left hand side is

$$\Sigma_n(M)\left(\hat{\beta}_M - \theta\right) = \frac{1}{n}\sum_{i=1}^{n} X_i(M)\left(Y_i - X_i^\top(M)\theta\right).$$

- Thus, the confidence set is very similar to the constraint set of Dantzig selector and so the name PoSI-Dantzig.
- Validity does not require either independence or identical distribution for the random vectors.
- It allows for diverging number of covariates. If (X_i, Y_i) are independent and sub-Gaussian then for $k = o(\sqrt{n/\log p})$,

$$\textbf{Leb}\left(\hat{\mathcal{R}}_M\right) = O_p\left(\sqrt{\frac{|M|\log p}{n}}\right)^{|M|} \quad \text{uniformly for } M \in \mathcal{M}(k).$$

PoSI-Dantzig

- Define for any model $M \subseteq \{1, 2 \ldots, p\}$,

$$\hat{\mathcal{R}}_M := \left\{\theta \in \mathbb{R}^{|M|} : \left\|\Sigma_n(M)\left(\hat{\beta}_M - \theta\right)\right\|_\infty \le C_1(\alpha) + C_2(\alpha)\left\|\hat{\beta}_M\right\|_1\right\}.$$

- Note the left hand side is

$$\Sigma_n(M)\left(\hat{\beta}_M - \theta\right) = \frac{1}{n}\sum_{i=1}^{n} X_i(M)\left(Y_i - X_i^\top(M)\theta\right).$$

- Thus, the confidence set is very similar to the constraint set of Dantzig selector and so the name PoSI-Dantzig.
- Validity does not require either independence or identical distribution for the random vectors.
- It allows for diverging number of covariates. If (X_i, Y_i) are independent and sub-Gaussian then for $k = o(\sqrt{n/\log p})$,

$$\textbf{Leb}\left(\hat{\mathcal{R}}_M\right) = O_p\left(\sqrt{\frac{|M|\log p}{n}}\right)^{|M|} \quad \text{uniformly for } M \in \mathcal{M}(k).$$

PoSI-Dantzig

- Define for any model $M \subseteq \{1, 2 \ldots, p\}$,

$$\hat{\mathcal{R}}_M := \left\{\theta \in \mathbb{R}^{|M|} : \left\|\Sigma_n(M)\left(\hat{\beta}_M - \theta\right)\right\|_\infty \le C_1(\alpha) + C_2(\alpha)\left\|\hat{\beta}_M\right\|_1\right\}.$$

- Note the left hand side is

$$\Sigma_n(M)\left(\hat{\beta}_M - \theta\right) = \frac{1}{n}\sum_{i=1}^{n} X_i(M)\left(Y_i - X_i^\top(M)\theta\right).$$

- Thus, the confidence set is very similar to the constraint set of Dantzig selector and so the name PoSI-Dantzig.
- Validity does not require either independence or identical distribution for the random vectors.
- It allows for diverging number of covariates. If (X_i, Y_i) are independent and sub-Gaussian then for $k = o(\sqrt{n/\log p})$,

$$\textbf{Leb}\left(\hat{\mathcal{R}}_M\right) = O_p\left(\sqrt{\frac{|M|\log p}{n}}\right)^{|M|} \quad \text{uniformly for } M \in \mathcal{M}(k).$$

Outline

1 Introduction
- Assumption-lean Linear Regression
- Some Notation and Uniform Results

2 Post-selection Inference Problem
- PoSI-Dantzig
- PoSI-Lasso
- PoSI-SqrtLasso

3 Multiplier Bootstrap

4 Some Comments and Extensions

PoSI-Lasso

- Define for $M \subseteq \{1, 2, \ldots, p\}$,

$$\hat{\mathcal{R}}_M := \left\{ \theta : \hat{\mathcal{L}}_M(\theta) \le \hat{\mathcal{L}}_M(\hat{\beta}_M) + 4C_1(\alpha) \left\| \hat{\beta}_M \right\|_1 + 2C_2(\alpha) \left\| \hat{\beta}_M \right\|_1^2 \right\},$$

where

$$\hat{\mathcal{L}}_M(\theta) = \mathbb{E}_n (Y - X^\top(M)\theta)^2.$$

- Note $C_2(\alpha)$ is related to the quantile of $\|\Sigma_n - \Sigma\|_\infty$. This is exactly zero, if covariates are treated fixed.

- In this case, the right hand side is exactly the same as the lasso objective function. Hence the name PoSI-Lasso.

- Validity does not require either independence or identical distributions for the random vectors.

PoSI-Lasso

- Define for $M \subseteq \{1, 2, \ldots, p\}$,

$$\hat{\mathcal{R}}_M := \left\{\theta : \hat{\mathcal{L}}_M(\theta) \le \hat{\mathcal{L}}_M(\hat{\beta}_M) + 4C_1(\alpha)\left\|\hat{\beta}_M\right\|_1 + 2C_2(\alpha)\left\|\hat{\beta}_M\right\|_1^2\right\},$$

  where

$$\hat{\mathcal{L}}_M(\theta) = \mathbb{E}_n(Y - X^\top(M)\theta)^2.$$

- Note $C_2(\alpha)$ is related to the quantile of $\|\Sigma_n - \Sigma\|_\infty$. This is exactly zero, if covariates are treated fixed.
- In this case, the right hand side is exactly the same as the lasso objective function. Hence the name PoSI-Lasso.
- Validity does not require either independence or identical distributions for the random vectors.

PoSI-Lasso

- Define for $M \subseteq \{1, 2, \ldots, p\}$,

$$\hat{\mathcal{R}}_M := \left\{\theta : \hat{\mathcal{L}}_M(\theta) \le \hat{\mathcal{L}}_M(\hat{\beta}_M) + 4C_1(\alpha)\left\|\hat{\beta}_M\right\|_1 + 2C_2(\alpha)\left\|\hat{\beta}_M\right\|_1^2\right\},$$

  where

$$\hat{\mathcal{L}}_M(\theta) = \mathbb{E}_n(Y - X^\top(M)\theta)^2.$$

- Note $C_2(\alpha)$ is related to the quantile of $\|\Sigma_n - \Sigma\|_\infty$. This is exactly zero, if covariates are treated fixed.

- In this case, the right hand side is exactly the same as the lasso objective function. Hence the name PoSI-Lasso.

- Validity does not require either independence or identical distributions for the random vectors.

PoSI-Lasso

- Define for $M \subseteq \{1, 2, \ldots, p\}$,

$$\hat{\mathcal{R}}_M := \left\{\theta : \hat{\mathcal{L}}_M(\theta) \le \hat{\mathcal{L}}_M(\hat{\beta}_M) + 4C_1(\alpha)\left\|\hat{\beta}_M\right\|_1 + 2C_2(\alpha)\left\|\hat{\beta}_M\right\|_1^2\right\},$$

  where

$$\hat{\mathcal{L}}_M(\theta) = \mathbb{E}_n(Y - X^\top(M)\theta)^2.$$

- Note $C_2(\alpha)$ is related to the quantile of $\|\Sigma_n - \Sigma\|_\infty$. This is exactly zero, if covariates are treated fixed.
- In this case, the right hand side is exactly the same as the lasso objective function. Hence the name PoSI-Lasso.
- Validity does not require either independence or identical distributions for the random vectors.

Outline

1. Introduction
   - Assumption-lean Linear Regression
   - Some Notation and Uniform Results
2. Post-selection Inference Problem
   - PoSI-Dantzig
   - PoSI-Lasso
   - PoSI-SqrtLasso
3. Multiplier Bootstrap
4. Some Comments and Extensions

PoSI-SqrtLasso

- Define the confidence set

$$\hat{\mathcal{R}}_M := \left\{\theta : \hat{\mathcal{L}}_M^{\frac{1}{2}}(\theta) \le \hat{\mathcal{L}}_M^{\frac{1}{2}}(\hat{\beta}_M) + 2C(\alpha)\left\|\hat{\beta}_M\right\|_1\right\},$$

  where

$$C^2(\alpha) = \max\{C_1(\alpha), C_2(\alpha)\}.$$

- The right hand side is exactly the objective function of square-root lasso. Hence the name PoSI-SqrtLasso.
- Again the validity does not require either independence or identical distribution for the random vectors.
- All these three confidence regions are valid as long as $C_1(\alpha)$ and $C_2(\alpha)$ are valid quantiles (or their estimators).

PoSI-SqrtLasso

- Define the confidence set

$$\hat{\mathcal{R}}_M := \left\{\theta : \hat{\mathcal{L}}_M^{\frac{1}{2}}(\theta) \le \hat{\mathcal{L}}_M^{\frac{1}{2}}(\hat{\beta}_M) + 2C(\alpha)\left\|\hat{\beta}_M\right\|_1\right\},$$

where

$$C^2(\alpha) = \max\{C_1(\alpha), C_2(\alpha)\}.$$

- The right hand side is exactly the objective function of square-root lasso. Hence the name PoSI-SqrtLasso.
- Again the validity does not require either independence or identical distribution for the random vectors.
- All these three confidence regions are valid as long as $C_1(\alpha)$ and $C_2(\alpha)$ are valid quantiles (or their estimators).

PoSI-SqrtLasso

- Define the confidence set

$$\hat{\mathcal{R}}_M := \left\{\theta : \hat{\mathcal{L}}_M^{\frac{1}{2}}(\theta) \le \hat{\mathcal{L}}_M^{\frac{1}{2}}(\hat{\beta}_M) + 2C(\alpha)\left\|\hat{\beta}_M\right\|_1\right\},$$

where

$$C^2(\alpha) = \max\{C_1(\alpha), C_2(\alpha)\}.$$

- The right hand side is exactly the objective function of square-root lasso. Hence the name PoSI-SqrtLasso.
- Again the validity does not require either independence or identical distribution for the random vectors.
- All these three confidence regions are valid as long as $C_1(\alpha)$ and $C_2(\alpha)$ are valid quantiles (or their estimators).

PoSI-SqrtLasso

- Define the confidence set

$$\hat{\mathcal{R}}_M := \left\{\theta : \hat{\mathcal{L}}_M^{\frac{1}{2}}(\theta) \le \hat{\mathcal{L}}_M^{\frac{1}{2}}(\hat{\beta}_M) + 2C(\alpha)\left\|\hat{\beta}_M\right\|_1\right\},$$

  where

$$C^2(\alpha) = \max\{C_1(\alpha), C_2(\alpha)\}.$$

- The right hand side is exactly the objective function of square-root lasso. Hence the name PoSI-SqrtLasso.
- Again the validity does not require either independence or identical distribution for the random vectors.
- All these three confidence regions are valid as long as $C_1(\alpha)$ and $C_2(\alpha)$ are valid quantiles (or their estimators).

Multiplier Bootstrap

- All the three confidence regions depend on $C_1(\alpha)$ and $C_2(\alpha)$ for $\alpha \in (0, 1)$.
- For the cases of independent random vectors and certain dependence settings, (with $\log p = o(n^{1/5})$) multiplier bootstrap works and gives correct estimators of $(C_1(\alpha), C_2(\alpha))$.
- If $w_1, w_2, \ldots, w_n$ are mean zero independent random variables with $\mathbb{E}w_i^2 = \mathbb{E}w_i^3 = 1$, then it can be proved that with high probability,

$$\mathbb{P}\left(\|\mathbb{E}_n[Z] - \mathbb{E}[Z]\|_\infty \le t\right) \le \mathbb{P}\left(\|\mathbb{E}_n\left[w\left(Z - \mathbb{E}_nZ\right)\right]\|_\infty \le t \middle| Z_1^n\right) + O_p\left(\frac{(\log p)^5}{n}\right)^{1/6}.$$

- Here $Z_1, Z_2, \ldots, Z_n$ are independent random vectors in $\mathbb{R}^p$. The result applies for dependent variables with some changes in rates.

Multiplier Bootstrap

- All the three confidence regions depend on $C_1(\alpha)$ and $C_2(\alpha)$ for $\alpha \in (0, 1)$.
- For the cases of independent random vectors and certain dependence settings, (with $\log p = o(n^{1/5})$) multiplier bootstrap works and gives correct estimators of $(C_1(\alpha), C_2(\alpha))$.
- If $w_1, w_2, \ldots, w_n$ are mean zero independent random variables with $\mathbb{E}w_i^2 = \mathbb{E}w_i^3 = 1$, then it can be proved that with high probability,

$$\mathbb{P}\left(\|\mathbb{E}_n[Z] - \mathbb{E}[Z]\|_\infty \le t\right) \le \mathbb{P}\left(\|\mathbb{E}_n\left[w\left(Z - \mathbb{E}_nZ\right)\right]\|_\infty \le t \middle| Z_1^n\right) + O_p\left(\frac{(\log p)^5}{n}\right)^{1/6}.$$

- Here $Z_1, Z_2, \ldots, Z_n$ are independent random vectors in $\mathbb{R}^p$. The result applies for dependent variables with some changes in rates.

Multiplier Bootstrap

- All the three confidence regions depend on $C_1(\alpha)$ and $C_2(\alpha)$ for $\alpha \in (0, 1)$.
- For the cases of independent random vectors and certain dependence settings, (with $\log p = o(n^{1/5})$) multiplier bootstrap works and gives correct estimators of $(C_1(\alpha), C_2(\alpha))$.
- If $w_1, w_2, \ldots, w_n$ are mean zero independent random variables with $\mathbb{E}w_i^2 = \mathbb{E}w_i^3 = 1$, then it can be proved that with high probability,

$$\mathbb{P}\left(\|\mathbb{E}_n[Z] - \mathbb{E}[Z]\|_\infty \le t\right) \le \mathbb{P}\left(\|\mathbb{E}_n\left[w\left(Z - \mathbb{E}_nZ\right)\right]\|_\infty \le t \middle| Z_1^n\right) + O_p\left(\frac{(\log p)^5}{n}\right)^{1/6}.$$

- Here $Z_1, Z_2, \ldots, Z_n$ are independent random vectors in $\mathbb{R}^p$. The result applies for dependent variables with some changes in rates.

Multiplier Bootstrap

- All the three confidence regions depend on $C_1(\alpha)$ and $C_2(\alpha)$ for $\alpha \in (0, 1)$.
- For the cases of independent random vectors and certain dependence settings, (with $\log p = o(n^{1/5})$) multiplier bootstrap works and gives correct estimators of $(C_1(\alpha), C_2(\alpha))$.
- If $w_1, w_2, \ldots, w_n$ are mean zero independent random variables with $\mathbb{E}w_i^2 = \mathbb{E}w_i^3 = 1$, then it can be proved that with high probability,

$$\mathbb{P}\left(\|\mathbb{E}_n[Z] - \mathbb{E}[Z]\|_\infty \le t\right)$$
$$\le \mathbb{P}\left(\|\mathbb{E}_n\left[w\left(Z - \mathbb{E}_n Z\right)\right]\|_\infty \le t \middle| Z_1^n\right) + O_p\left(\frac{(\log p)^5}{n}\right)^{1/6}.$$

- Here $Z_1, Z_2, \ldots, Z_n$ are independent random vectors in $\mathbb{R}^p$. The result applies for dependent variables with some changes in rates.

Summary and Extensions

- We proved some new uniform rates of convergence results in linear regression with NO model/distributional assumptions.
- These results apply to various structurally dependent random vectors with rates accordingly affected.
- We described three different confidence regions that have valid coverage no matter how the model is chosen.
- All these methods are computationally efficient and only require computation of $C_1(\alpha)$, $C_2(\alpha)$ once for the dataset.
- The construction of these confidence regions (for now) seems very specific to the quadratic structure of the objective function.
- We also developed two different approaches (based on t- and F-tests) that apply to most of the M-estimation problems, including, generalized linear models.

Summary and Extensions

- We proved some new uniform rates of convergence results in linear regression with NO model/distributional assumptions.
- These results apply to various structurally dependent random vectors with rates accordingly affected.
- We described three different confidence regions that have valid coverage no matter how the model is chosen.
- All these methods are computationally efficient and only require computation of $C_1(\alpha)$, $C_2(\alpha)$ once for the dataset.
- The construction of these confidence regions (for now) seems very specific to the quadratic structure of the objective function.
- We also developed two different approaches (based on t- and F-tests) that apply to most of the M-estimation problems, including, generalized linear models.

Summary and Extensions

- We proved some new uniform rates of convergence results in linear regression with NO model/distributional assumptions.
- These results apply to various structurally dependent random vectors with rates accordingly affected.
- We described three different confidence regions that have valid coverage no matter how the model is chosen.
- All these methods are computationally efficient and only require computation of $C_1(\alpha)$, $C_2(\alpha)$ once for the dataset.
- The construction of these confidence regions (for now) seems very specific to the quadratic structure of the objective function.
- We also developed two different approaches (based on t- and F-tests) that apply to most of the M-estimation problems, including, generalized linear models.

Summary and Extensions

- We proved some new uniform rates of convergence results in linear regression with NO model/distributional assumptions.
- These results apply to various structurally dependent random vectors with rates accordingly affected.
- We described three different confidence regions that have valid coverage no matter how the model is chosen.
- All these methods are computationally efficient and only require computation of $C_1(\alpha)$, $C_2(\alpha)$ once for the dataset.
- The construction of these confidence regions (for now) seems very specific to the quadratic structure of the objective function.
- We also developed two different approaches (based on t- and F-tests) that apply to most of the M-estimation problems, including, generalized linear models.

Summary and Extensions

- We proved some new uniform rates of convergence results in linear regression with NO model/distributional assumptions.
- These results apply to various structurally dependent random vectors with rates accordingly affected.
- We described three different confidence regions that have valid coverage no matter how the model is chosen.
- All these methods are computationally efficient and only require computation of $C_1(\alpha)$, $C_2(\alpha)$ once for the dataset.
- The construction of these confidence regions (for now) seems very specific to the quadratic structure of the objective function.
- We also developed two different approaches (based on t- and F-tests) that apply to most of the M-estimation problems, including, generalized linear models.

Summary and Extensions

- We proved some new uniform rates of convergence results in linear regression with NO model/distributional assumptions.
- These results apply to various structurally dependent random vectors with rates accordingly affected.
- We described three different confidence regions that have valid coverage no matter how the model is chosen.
- All these methods are computationally efficient and only require computation of $C_1(\alpha)$, $C_2(\alpha)$ once for the dataset.
- The construction of these confidence regions (for now) seems very specific to the quadratic structure of the objective function.
- We also developed two different approaches (based on t- and F-tests) that apply to most of the M-estimation problems, including, generalized linear models.

References

[1] Berk, R., Brown, L. D., Buja, A., Zhang, K., Zhao, L. (2013)
Valid post-selection inference.
Ann. Statist. 41, no. 2, 802–837.

[2] Deng, H., Zhang, C. H. (2017)
Beyond Gaussian Approximation: Bootstrap for Maxima of Sums of Independent Random Vectors
arxiv.org/abs/1705.09528.

Thank You!
Questions?

References

[1] Berk, R., Brown, L. D., Buja, A., Zhang, K., Zhao, L. (2013)
Valid post-selection inference.
Ann. Statist. 41, no. 2, 802–837.

[2] Deng, H., Zhang, C. H. (2017)
Beyond Gaussian Approximation: Bootstrap for Maxima of Sums of Independent Random Vectors
arxiv.org/abs/1705.09528.

Thank You!
Questions?

References

[1] Berk, R., Brown, L. D., Buja, A., Zhang, K., Zhao, L. (2013)
Valid post-selection inference.
Ann. Statist. 41, no. 2, 802–837.

[2] Deng, H., Zhang, C. H. (2017)
Beyond Gaussian Approximation: Bootstrap for Maxima of Sums of Independent Random Vectors
arxiv.org/abs/1705.09528.

Thank You!
Questions?