ALMA MATER STUDIORUM UNIVERSITÀ DI BOLOGNA

Fast and Robust Bootstrap

Salibián-Barrera, Van Aelst, Willems (2008)

Antonio Panico, Ivo Bonfanti

Department of Statistical Sciences
University of Bologna

May 29, 2025

Introduction

- Inference based on robust estimators' asymptotic distribution require assumptions such as symmetric or elliptical distribution of the data.
- Solution → bootstrap! Does not require stringent assumptions.
- However classical bootstrap has two problems: BOUNDS INF. FUNCTION
  1. Numerical instability (# OF OUTLIERS)
  2. Computational cost (ROB. ESTIMATES ~D NON-CONVEX OPT. PROB. (LOSS FUNCTION)) BURDEN DRASTICALLY INCREASES WITH THE # OF DIMENSIONS
- Goal: introduce a faster and more stable method → **Fast and Robust Bootstrap (FRB)**.
- Applications: robust regression, PCA, discriminant analysis.

FRB General Framework

$\hat{\theta}_n$ REPRESENTED AS A SOLUTION OF FIXED-POINT EQUATIONS

- Robust estimator computed on the original dataset:

$$\hat{\theta}_n = \mathbf{g}_n(\hat{\theta}_n) \tag{1}$$

- Bootstrap robust estimate has problems **1** and **2**.

$$\hat{\theta}_n^* = \mathbf{g}_n^*(\hat{\theta}_n^*) \tag{2}$$

- One step approximation solves **2**, however it is not consistent.

$$\hat{\theta}_n^{1*} = \mathbf{g}_n^*(\hat{\theta}_n) \tag{3}$$

- The **FRB** is the corrected approximation. Now also **1** is solved and the estimator is consistent.

COMES FROM THE TAYLOR EXPANSION ABOUT $\hat{\theta}_n$ LIMITING VALUE θ

$$\hat{\theta}_n^{R*} := \hat{\theta}_n + \left[\mathbf{I} - \nabla g_n(\hat{\theta}_n)\right]^{-1} (\hat{\theta}_n^{1*} - \hat{\theta}_n). \tag{4}$$

Robust Regression

$\sim f_\varepsilon$ (generic parametric distribution)

LINEAR REGRESSION MODEL: $y_i = \mathbf{x}_i^t \beta_0 + \sigma_0 \varepsilon_i, \quad i = 1, \ldots, n,$

where $\mathcal{Z}_n = \left\{ (y_1, \mathbf{z}_1^t)^t, \ldots, (y_n, \mathbf{z}_n^t)^t \right\}$ is a sample of independent random vectors and $\mathbf{x}_i = (1, \mathbf{z}_i^t)^t \in \mathbb{R}^p$.

NON-CONVEX FUNCTION (E.g. HUBER, TUKEY)

- SYMMETRIC
- TWICE CONTINUOUSLY DIFFERENTIABLE
- $\rho(0) = 0$
- STRICTLY INCR. $[0, c]$

AND CONSTANT ON (c, ∞)

(REGRESSION)

1. Compute S-estimate $\tilde{\beta}_n = \arg\min_{\beta} (\hat{\sigma}_n)$: IMPLICITELY DEFINED IN

$$\frac{1}{n} \sum_{i=1}^{n} \rho_0 \left(\frac{y_i - \mathbf{x}_i^t \beta}{\hat{\sigma}_n(\beta)} \right) = b \qquad (5)$$

b DETERMINES THE BDP

2. Compute **MM-estimate** $\hat{\beta}_n$ by solving:

F.O.C.

$$\frac{1}{n} \sum_{i=1}^{n} \rho_1' \left(\frac{y_i - \mathbf{x}_i^t \hat{\beta}_n}{\hat{\sigma}_n} \right) \mathbf{x}_i = \mathbf{0} \qquad (6)$$

$\hat{\sigma}_n$: SCALE S-ESTIMATE

FRB MM-Regression

Compute, for each bootstrap sample $\mathcal{Z}_n^*$, the one-step approximation estimate $\hat{\theta}_n^{1*} = (\hat{\beta}_n^{1*}, \tilde{\beta}_n^{1*}, \hat{\sigma}_n^{1*}) = \mathbf{g}_n^*(\hat{\theta}_n)$: Eq. (3)

fixed point eq. est.

$$\hat{\beta}_n^{1*} = \left(\sum_{i=1}^n \omega_i^* \, \mathbf{x}_i^* \mathbf{x}_i^{*t} \right)^{-1} \sum_{i=1}^n \omega_i^* \, \mathbf{x}_i^* y_i^*,$$

$$\hat{\sigma}_n^{1*} = \sum_{i=1}^n \nu_i^* \left(y_i^* - \mathbf{x}_i^{*t} \tilde{\beta}_n \right),$$

$$\tilde{\beta}_n^{1*} = \left(\sum_{i=1}^n \tilde{\omega}_i^* \, \mathbf{x}_i^* \mathbf{x}_i^{*t} \right)^{-1} \sum_{i=1}^n \tilde{\omega}_i^* \, \mathbf{x}_i^* y_i^*,$$

WEIGHTED LEAST SQUARE FIT

where residuals and weights are defined as follow:

$$r_i^* = y_i^* - \mathbf{x}_i^{*t} \hat{\beta}_n, \; ! \qquad \omega_i^* = \rho_1' \left(r_i^* / \hat{\sigma}_n \right) / r_i^*, \qquad \tilde{W}_i^* = \rho_0' (\tilde{r}_i^* / \hat{\sigma}_n) / \tilde{r}_i^*$$

$$\tilde{r}_i^* = y_i^* - \mathbf{x}_i^{*t} \tilde{\beta}_n, \qquad \nu_i^* = \frac{\hat{\sigma}_n}{nb} \rho_0 \left(\tilde{r}_i^* / \hat{\sigma}_n \right) / \tilde{r}_i^*.$$

Finally, calculate the correction to obtain the **FRB** estimate:

$$\hat{\beta}_n^{R*} = \hat{\beta}_n + \mathbf{M}_n\left(\hat{\beta}_n^{1*} - \hat{\beta}_n\right) + \mathbf{d}_n\left(\hat{\sigma}_n^{1*} - \hat{\sigma}_n\right) + \tilde{\mathbf{M}}_n\left(\tilde{\beta}_n^{1*} - \tilde{\beta}_n\right),$$

where M_n, d_n and $\tilde{M}_n$ are computed only once on the original dataset.
(lower computational cost)

- Under certain regularity conditions the FRB estimate is **consistent**:

$$\sqrt{n}\left(\hat{\beta}_n^{R*} - \hat{\beta}_{\boldsymbol{n}}\right) \sim \sqrt{n}\left(\hat{\beta}_n - \beta\right).$$

- **Confidence intervals** based on FRB MM-estimator:

MM estimate computed on the original dataset

$$\hat{\beta}_{n,j} \pm z_{\alpha/2}\hat{\sigma}_j/\sqrt{n},$$

(empirical)
where the standard error estimate $\hat{\sigma}_j$ is provided by the empirical standard deviation of $\hat{\beta}_{n,j}^{R*}$. • higher BDP of the classical bootstrap

Mean Response C.I. and Prediction Interval

DATA A FIXED POINT

- Similarly, the confidence interval for the **mean response** is:

$$x_0^\top \hat{\beta}_n \pm z_{\alpha/2}\sqrt{x_0^\top \hat{\Sigma} x_0} \qquad (1-\alpha)\%\ \text{confidence}$$

ORIGINAL MM-EST.

FAST

where $\hat{\Sigma}$ is a bootstrap estimate of the covariance matrix of $\hat{\beta}_n$.

FROM THE SAMPLE COVARIANCE OF THE REPLICATES $\hat{\beta}_n^{R*(i)}$

- While the **prediction interval**: (POINT DRAWN FROM THE POPULATION)

INCLUDE NOISE FROM THE DGP.

$$x_0^\top \hat{\beta}_n \pm z_{\alpha/2}\sqrt{x_0^\top \hat{\Sigma} x_0 + \hat{\sigma}_n^2}$$

where $\hat{\sigma}_n$ is the scale S-estimate.

INCLUDE BOTH COEFF. UNCERTAINTY + RANDOM NOISE

Performance assessed through a simulation study:

ORIGINAL DATASET

- Symmetric outliers: High coverage in all settings.
- Asymmetric outliers: High coverage except in challenging situations (at the edge of the predictor space with high contamination).

FRB Score Hypothesis Test

- Test $H_0 : \beta_{0,2} = 0$ vs $H_a : \beta_{0,2} \neq 0$, where $\beta_0 = (\beta_{0,1}^t, \beta_{0,2}^t)^t$ with $\beta_{0,1} \in \mathcal{R}^q$ and $\beta_{0,2} \in \mathcal{R}^{p-q}$ is a partition of the coefficients vector.
- Partition in a similar way the covariates $x_i = (x_{i(1)}^t, x_{i(2)}^t)$, define the residuals $r_i^{(a)} = y_i - \mathbf{x}_i^t \hat{\beta}_n^{(a)}$, $r_i^{(0)} = y_i - \mathbf{x}_{i(1)}^t \hat{\beta}_n^{(0)}$ and the estimates $\hat{\beta}_n^{(0)} \in \mathcal{R}^{p-q}$ and $\hat{\beta}_n^{(a)} \in \mathcal{R}^p$.
- The **score test** statistic is defined as follow:

$$W_n^2 = n^{-1} \left[\Sigma_n^{(0)} \right]^t \hat{\mathbf{U}}^{-1} \left[\Sigma_n^{(0)} \right] \tag{7}$$

  where $\Sigma_n^{(0)} = \sum_{i=1}^n \rho_1' \left(\frac{r_i^{(0)}}{\hat{\sigma}_n^{(a)}} \right) \mathbf{x}_{i(2)}$ and $\hat{\mathbf{U}}$ is a robust estimate of the covariance.
- If the errors are symmetric, for large n we have $W_n^2 \sim \chi_q^2$.

If the errors are not symmetric, we cannot rely on the statistic's asymptotic distribution. However we can exploit FRB to estimate it:

1. First we have to draw bootstrap samples from data that follow H_0:

ensures H_0 data structure

$$\tilde{y}_i = \mathbf{x}_i^t \hat{\beta}_n^{(0)} + r_i^{(a)}$$

preserve error structure

2. Compute the MM-estimates $\ddot{\beta}_n^{(0)}$ and $\ddot{\beta}_n^{(a)}$ on $(\tilde{y}, x)$ null data.
3. For each bootstrap sample $(\tilde{y}^*, x^*)$ compute FRB estimates under both the null and the alternative $\ddot{\beta}_n^{R*(0)}$ $\ddot{\beta}_n^{R*(a)}$.
4. The **FRB score test** statistic is, for each sample:

$$W_n^{2R^*} = n^{-1} \Sigma_n^{R^*} (\ddot{\beta}_n^{R^*(0)})^t \mathbf{U}^{R^*-1} \Sigma_n^{R^*} (\ddot{\beta}_n^{R^*(0)}) \tag{8}$$

where $\Sigma_n^{R^*}(\beta) = \sum_{i=1}^n \rho_1' \left(\frac{\tilde{y}_i^* - \hat{\beta}_n^{(0)t} \mathbf{x}_i}{\ddot{\sigma}_n^{(a)}} \right) \mathbf{x}_{i(2)}$ and $\mathbf{U}^{R*}$ the corresponding matrix.

5. Finally compute the p-value as:

$$\hat{p} = \#\left\{W_n^{2R*} > W_n^2\right\}/\mathcal{B} \tag{9}$$

where $\mathcal{B}$ is the number of bootstrap samples. If $\hat{p} < 0.05$ it means that the observed statistic W_n^2 is rare under H_0. Reject it!

Robust PCA and Fast and Robust Bootstrap (FRB)

Robust PCA:

- Classical PCA is based on eigenvalues/eigenvectors of the sample covariance matrix, sensitive to outliers.
- Robust PCA replaces the sample scatter matrix with a robust estimator, e.g., an **S-estimator**:

$$\min_{T,C} |C| \quad \text{s.t.} \quad \frac{1}{n}\sum_{i=1}^{n} \rho\left((x_i - T)^\top C^{-1}(x_i - T)\right) = b$$

- $\rho(\cdot)$ is a loss function satisfying robustness properties (bounded, non-decreasing); $T \in \mathbb{R}^p$, $C \in \mathrm{PDS}(p)$, and b is a constant chosen as

$$b = \mathbb{E}_\phi\left[\rho(||\mathbf{X}||)\right]$$

- The robust PCA is then based on the eigen-decomposition of the estimated scatter matrix $\widehat{\Sigma}_n$.

Robust PCA and Fast and Robust Bootstrap (FRB)

- Bootstrap estimates are computed starting from an S-estimates of location μ_n and scatter Σ_n.:

$$\widehat{\theta}_n = \left(\mu_n^\top, \text{vec}(\Sigma_n)^\top\right)^\top$$

- FRB replicates:

$$\hat{\theta}_n^{R*} = \hat{\theta}_n + \left[I - \nabla g_n(\hat{\theta}_n)\right]^{-1} \left(\hat{\theta}_n^{1*} - \hat{\theta}_n\right)$$

  where $\widehat{\theta}_n^{1*}$ is a one-step update from bootstrapped sample.
- Allows construction of:
  - CI for eigenvalues λ_i.
  - Stability of eigenvectors via angle distribution.
  - CI for proportion of variance explained:

$$p_k = \frac{\sum_{i=1}^{k} \lambda_i}{\sum_{i=1}^{p} \lambda_i}$$

Swiss Bank Notes Example

- $n = 100$ forged Swiss 1000-franc notes.
- $p = 6$ measurements: length, height, and other distances.
- Analyzed using robust PCA based on 50% breakdown S-estimators.

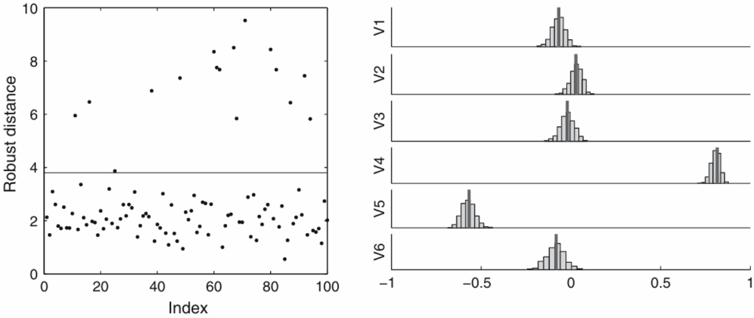


Figure: Swiss bank notes data; Left: robust distances based on S; Right: weights in the first PC, with FRB histograms.

Bootstrap Stability of PC1

To compare variability, compute: $\theta = \arccos(|\mathbf{v}_1^\top \mathbf{v}_1^*|)$

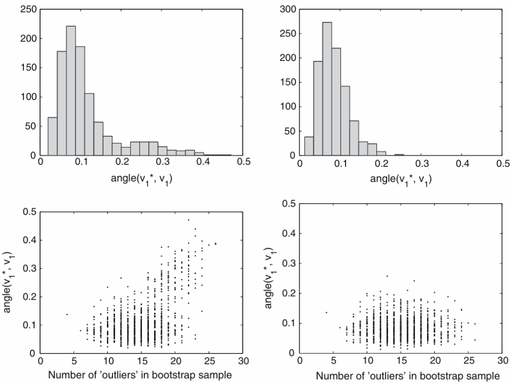


Figure: Bootstrap angle distributions: Classical vs FRB

Explained Variance and Confidence Intervals

- Proportion of variance:

$$p_1 = 72.0\%, \quad p_2 = 84.5\%, \quad p_3 = 91.6\%, \ldots$$

- FRB used to construct BCa 95% confidence intervals.
- Criteria: choose smallest k such that p_k exceeds cutoff.

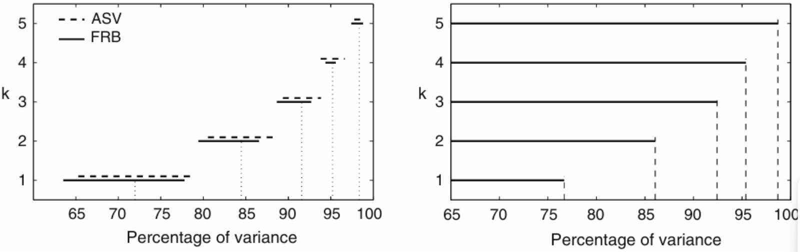


Figure: Left: FRB vs asymptotic intervals; Right: 95% one-sided CI

Classical & Robust Discriminant Analysis

Goal: Classify $x \in \mathbb{R}^p$ into π_1 or π_2 with means μ_1, μ_2, covariances Σ_1, Σ_2.
Linear Discriminant Rule (LDA): if $\Sigma_1 = \Sigma_2 = \Sigma$

$$d_j^L(x) = \mu_j^T \Sigma^{-1} x - \frac{1}{2}\mu_j^T \Sigma^{-1}\mu_j \qquad j = 1, 2$$

Assign x to class π_1 if $d_1^L(x) > d_2^L(x)$

Classical Limitation: Empirical estimates μ_j, Σ_j are **sensitive to outliers**.

Robust Approach:

- Use **S-estimators** μ_{jn}, Σ_{jn}.
- Robust LDA: $\Sigma_n = \frac{1}{2}(\Sigma_{1n} + \Sigma_{2n})$ or **two-sample S-estimators**:

$$\arg\min_{T_1, T_2, C} |C| \quad \text{s.t.} \quad \frac{1}{n_1}\sum \rho(d_{1i}) + \frac{1}{n_2}\sum \rho(d_{2i}) = b$$

$$d_{ji} = (x_i^j - T_j)^T C^{-1}(x_i^j - T_j)$$

Error Estimation: FRB and Comparison

Problem: Estimate misclassification error for robust rules.

- **Cross-validation (CV)** and **classical bootstrap:** accurate, slow.
- **Train/validation split:** faster, less stable.

FRB (Fast and Robust Bootstrap):

- Efficiently applies to S-estimates.
- Each bootstrap sample:
  - Recalculate robust estimates $\mu_1^{R*}, \mu_2^{R*}, \Sigma_1^{R*}, \Sigma_2^{R*}$.
  - Build rule, evaluate on out-of-bootstrap data.
- Average misclassification error:

$$\text{err}_{\text{boot}} = \frac{1}{B}\sum_{b=1}^{B} \text{error on out-of-bootstrap data}$$

Efron's 0.632 Estimator:

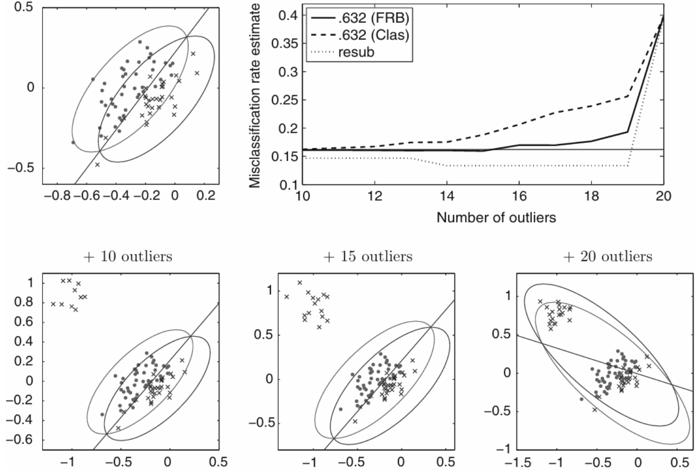
$$\text{err}_{0.632} = 0.632 \cdot \text{err}_{\text{boot}} + 0.368 \cdot \text{err}_{\text{resub}}$$

FRB Advantages: Faster, more stable, avoids full retraining.

Hemophilia Data Application

Data:

- $n_1 = 30$ normal women ($\times$), $n_2 = 45$ hemophilia A carriers ($\bullet$)



Hemophilia Data Application

Results:

- FRB and classical bootstrap errors ≈ 0.162 ($B = 100$).
- FRB **faster** and more stable under contamination.
- Robustness to 10–15 outliers; classical bootstrap becomes variable after 15 outliers.