Bootstrap of constraint estimators with application to rank estimation.

François Portier

IRMAR-University of Rennes 1

April 17, 2012

Table of contents

Introduction to Bootstrap

- Goal: To reproduce the asymptotic behavior of some estimators.
- Means: Creation of a new sample which "look like" the previous one.

• Bootstrap of Efron [Efron(1982)] :
Suppose that $(X_1, ..., X_n)$ i.i.d. with law P. We draw $(X_1^*, ..., X_n^*)$ with respect to the law

$$\widehat{P} = n^{-1} \sum_{i=1}^{n} \delta_{X_i}.$$

Define $\theta_0 = \mathbb{E}[X]$, $\quad \overline{X} = \frac{1}{n} \sum_{i=1}^{n} X_i$ and

$$\overline{X^*} = \frac{1}{n} \sum_{i=1}^{n} X_i^* = \frac{1}{n} \sum_{i=1}^{n} N_i X_i \quad \text{with } N_i \sim \text{ mult(1/n)}$$

• Another bootstrap method:

$$\overline{X^*} = \overline{X} + \frac{1}{n} \sum_{i=1}^{n} \epsilon_i X_i = \frac{1}{n} \sum_{i=1}^{n} (\epsilon_i + 1) X_i$$

with ϵ_i any i.i.d. sequence standard random variable.

Introduction to Bootstrap

Example of results with both previous bootstrap methods

If ϕ is continuously differentiable on a neighborhood of $\theta_0 = \mathbb{E}[X]$, if P has a finite second order moment 2, then

$$\sqrt{n}(\phi(\overline{X^*}) - \phi(\overline{X})) \quad \text{bootstrap} \quad \sqrt{n}(\phi(\overline{X}) - \phi(\theta_0)),$$

i.e.

$$\mathcal{L}(\sqrt{n}(\phi(\overline{X^*}) - \phi(\overline{X}))|\widehat{P}) \overset{n\to\infty}{=} \mathcal{L}(\sqrt{n}(\phi(\overline{X}) - \phi(\theta_0))|P).$$

Why the bootstrap ? Alternative to the use of the asymptotic law ([Hall(1992)]) for

- Building confidence interval
- Hypothesis testing (for the choice of quantile)

Test of equal means: classical bootstrap works

Assume $\theta_0 \in \mathbb{R}$,

$$H_0 : \theta_0 = \theta \qquad \text{against} \qquad H_1 : \theta_0 \neq \theta$$

To arbitrate:

$$\|\sqrt{n}(\overline{X} - \theta)\|^2 \text{ is compared to } \begin{cases} q_\alpha^\infty \text{ a quantile of the limiting distribution} \\ q_\alpha^* \text{ a quantile of the bootstrap statistic} \end{cases}$$

Level and power

$$\mathbb{P}_{H_0}(\|\sqrt{n}(\overline{X} - \theta)\|^2 > q_\alpha^\infty \text{ or } q_\alpha^*) \qquad \text{and} \qquad \mathbb{P}_{H_1}(\|\sqrt{n}(\overline{X} - \theta)\|^2 > q_\alpha^\infty \text{ or } q_\alpha^*)$$

- For q_α^∞: OK.
- For q_α^*: $\underbrace{\sqrt{n}(\overline{X^*} - \overline{X})}_{\text{do not depend on } H_0 \text{ or } H_1}$ bootstrap $\sqrt{n}(\overline{X} - \theta_0)$ $\Rightarrow$ OK.

In general classical bootstrap fails

Assume $\theta_0 \in \mathbb{R}^2$ and $\mathcal{C}$ is the unit circle,

$$H_0 : \ \theta_0 \in \mathcal{C} \qquad \text{against} \qquad H_1 : \ \theta_0 \notin \mathcal{C}$$

$\Rightarrow$ Constraint estimators :

$$\widehat{T}_n = n \min_{g(\theta)=0} \|\overline{X} - \theta\|^2$$

Does the classical Bootstrap works ?

Under H_0, $\qquad \widehat{T}_n = |\sqrt{n}(\phi(\overline{X}) - \phi(\theta_0))|^2 \qquad$ with $\phi : x \to \min_{g(\theta)=0} \|x - \theta\|$

Bootstrap candidate:

$$T_n^* = |\sqrt{n}(\phi(\overline{X^*}) - \phi(\overline{X}))|^2$$

Can not work because ϕ is not $\mathcal{C}^1$.

$\Rightarrow$ Even if we can bootstrap $\sqrt{n}(\overline{X} - \theta_0)$, it is not clear we are able to bootstrap some constraint estimators.

From now $\theta_0 \in \mathbb{R}^p$ (parameter of interest), it exists $\widehat{\theta}$ some consistent estimators of θ_0. Define the random function

$$\widehat{Q}_n(\theta) = (\widehat{\theta} - \theta)^T \widehat{S} (\widehat{\theta} - \theta).$$

Question

If we can bootstrap $\sqrt{n}(\widehat{\theta} - \theta_0)$, does the **under** H_0-law of

$$\sqrt{n}(\widehat{\theta}_c - \theta_0) \quad \text{with} \quad \widehat{\theta}_c = \underset{g(\theta)=0}{\operatorname{argmin}} \ \widehat{Q}(\theta)$$

can be bootstrapped ?

Applications

Statistics of the kind $\min_{g(\theta)=0} Q(\theta)$ to arbitrate between

$$H_0 : \ g(\theta_0) = 0 \qquad \text{and} \qquad H_1 : g(\theta_0) \neq 0$$

Intuitively

Define

$$\theta_c^* = \underset{g(\theta)=0}{\text{argmin}}\ Q^*(\theta) \quad \text{and} \quad Q^*(\theta) = (\theta^* - \theta)^T S^*(\theta^* - \theta).$$

As traditional bootstrap: we expect results such as

$$\sqrt{n}(\theta_c^* - \widehat{\theta}_c) \text{ bootstrap } \sqrt{n}(\widehat{\theta}_c - \theta_0)$$

The idea : A good choice of θ^*

We try to "reproduce" H_0 with

$$\theta^* = \widehat{\theta}_c + \text{ "something going to 0 with good speed and variance"}$$

$$\widehat{\theta}_c = \underset{g(\theta)=0}{\operatorname{argmin}} \, (\widehat{\theta} - \theta)^T \widehat{S} (\widehat{\theta} - \theta) \quad \text{and} \quad \theta_c^* = \underset{g(\theta)=0}{\operatorname{argmin}} \, (\theta^* - \theta)^T S^* (\theta^* - \theta)$$

Assumptions :

1. $\widehat{S} \xrightarrow{\mathbb{P}} S$ and $S^* \xrightarrow{\mathbb{P}} S$.
2. S is full rank.
3. $g : \mathbb{R}^p \to \mathbb{R}^q$ is $\mathcal{C}^1$ on a neighborhood of θ_0 and $J_g(\theta_0)$ is of full rank.

Theorem

Under H_0, if $\sqrt{n}(\theta^* - \widehat{\theta}_c)$ bootstrap $\sqrt{n}(\widehat{\theta} - \theta_0) \xrightarrow{d} X$ (Gaussian) we have

$$\sqrt{n}(\widehat{\theta}_c^* - \widehat{\theta}_c) \text{ bootstrap } \underbrace{\sqrt{n}(\widehat{\theta}_c - \theta_0) \textbf{ under } H_0.}_{\text{(Gaussian limit)}} \tag{1.1}$$

Under H_1, we additionally need to assume the existence of θ_c such as $\widehat{\theta}_c \xrightarrow{a.s.} \theta_c$ with $g(\theta_c) = 0$, to get (1.1).

$\Rightarrow \quad \theta^* = \widehat{\theta}_c + (\theta^*_{\text{classical}} - \widehat{\theta})$ with $\theta^*_{\text{classical}}$ comes from any methods of classical bootstrap.

What about T^* ?

Corollary

Under the previous set of assumptions under H_0 and H_1,

$$T^* = \underset{g(\theta)=0}{\text{argmin}}\ (\theta^* - \theta)^T S^* (\theta^* - \theta) \quad \textit{bootstrap} \qquad \underbrace{\widehat{T}\ \textbf{under}\ H_0}_{\textit{weighted Chi-squared limit}}$$

Problem

The assumption for convergence under H_1: $\widehat{\theta}_c \xrightarrow{a.s.} \theta_c$ need to be check for each case.

Assumptions :

1. $\sqrt{n}(\theta^* - \widehat{\theta}_c)$ bootstrap $\sqrt{n}(\widehat{\theta} - \theta_0) \xrightarrow{d} X$ (Gaussian)
2. $\widehat{S} \xrightarrow{\mathbb{P}} S$ and $S^* \xrightarrow{\mathbb{P}} S$
3. $g : \mathbb{R}^p \to \mathbb{R}^q$ is $\mathcal{C}^1$ on a neighborhood of θ_0 and $J_g(\theta_0)$ is of full rank.
4. S is full rank.

Corollary 2

The test with null hypothesis

$$H_0 : \; g(\theta_0) = 0 \text{ against } H_1 : \; g(\theta_0) \neq 0$$

and associated statistic $\widehat{T}$ with bootstrap calculation of quantile is consistent.

For the test procedure, one can draw

$$T_1^*, ..., T_B^* \text{ to estimate } q_\alpha^*$$

and we do not reject H_0 if $\widehat{T} \leq q^*$, or reject H_0 if not.
In other words Corollary 2 means:

$\Rightarrow$The asymptotic level of the test is α.

$\Rightarrow$The power of the test goes to 1.

Rk: This kind of test is pivotal (Chi-squared) when $S = Var(X)^{-1}$.

Application to rank estimation

Framework and notation

Goal: Estimation of the rank of a matrix **Means:** Hypothesis testing.

Assumptions

$\widehat{M}$ and M are matrices $\mathbb{R}^{p\times H}$ such that

- $\sqrt{n}(\vec{(}\widehat{M}) - \text{vec}(M)) \xrightarrow{\text{d}} \mathcal{N}(0, \Gamma)$
- $\widehat{\Gamma} \xrightarrow{\mathbb{P}} \Gamma$
- Nothing more **or** Γ invertible **or** $\Gamma = FF^T \otimes GG^T$ invertible.

Notations: $\text{rank}(M) = d_0$,

SVD of M and $\widehat{M}$:

$$M = (U_1 U_0) \begin{pmatrix} D_1 & 0 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} V_1^T \\ V_0^T \end{pmatrix} \quad \text{and} \quad \widehat{M} = (\widehat{U}_1 \widehat{U}_0) \begin{pmatrix} \widehat{D}_1 & 0 \\ 0 & \widehat{D}_0 \end{pmatrix} \begin{pmatrix} \widehat{V}_1^T \\ \widehat{V}_0^T \end{pmatrix}$$

$P_1 = U_1 U_1^T, Q_1 = U_0 U_0^T, P_2 = V_1 V_1^T, Q_2 = V_0 V_0^T$, $\widehat{P}_1$, $\widehat{P}_2$, $\widehat{Q}_1$ $\widehat{Q}_2$.
$(\widehat{\lambda}_1, ..., \widehat{\lambda}_p)$, (resp. $(\lambda_1, ..., \lambda_p)$) singularvalues of $\widehat{M}$ (resp. M) in ascending order.

Short review

For $d = 0, ..., d_0$, we test

$$H_0 : \quad d_0 = d \qquad \text{against} \qquad H_1 : \quad d_0 > d$$

Some statistics

$$\text{[Li(1991)]} \qquad \widehat{T}_1 = n \sum_{k=1}^{p-d} \widehat{\lambda}_k^2 \qquad (= n \| \operatorname{vec}(\widehat{Q}_1 \widehat{M} \widehat{Q}_2) \|^2)$$

$$\text{[Bura and Yang(2011)]} \qquad \widehat{T}_2 = n \operatorname{vec}(\widehat{Q}_1 \widehat{M} \widehat{Q}_2)^T \widehat{\Gamma}^+ \operatorname{vec}(\widehat{Q}_1 \widehat{M} \widehat{Q}_2)$$

$$\text{[Cragg and Donald(1997)]} \qquad \widehat{T}_3 = n \min_{\operatorname{rank}(M)=d} \operatorname{vec}(\widehat{M} - M)^T \widehat{\Gamma}^{-1} \operatorname{vec}(\widehat{M} - M)$$

By noting that

Lemma (From PCA)

$$\widehat{P}_1 \widehat{M} \widehat{P}_2 = \underset{\operatorname{rank}(M)=d}{\operatorname{argmin}} \| \widehat{M} - M \|_F^2 = \underset{\operatorname{rank}(M)=d}{\operatorname{argmin}} \| \operatorname{vec}(\widehat{M} - M) \|^2$$

we get

$$[\text{Li(1991)}] \qquad \widehat{T}_1 = n \min_{\text{rank}(M)=d} \| \text{vec}(\widehat{M} - M)\|^2$$

$$[\text{Bura and Yang(2011)}] \qquad \widehat{T}_2 = n\, \text{vec}(\widehat{M} - \widehat{M}_c)^T \widehat{\Gamma}^+ \text{vec}(\widehat{M} - \widehat{M}_c)$$

$$[\text{Cragg and Donald(1997)}] \qquad \widehat{T}_3 = n \min_{\text{rank}(M)=d} \text{vec}(\widehat{M} - M)^T \widehat{\Gamma}^+ \text{vec}(\widehat{M} - M)$$

with $\widehat{M}_c = \underset{\text{rank}(M)=d}{\text{argmin}} \|vecv(\widehat{M} - M)\|^2$.

Application of the results

- $\{\text{rank}(M) = d\}$ is a smooth submanifold.
- Example of sufficient conditions for bootstrap: It exists $\xi_1, ..., \xi_n$ i.i.d. with $\mathbb{E}[\|\xi_1\|_F^2] < +\infty$ such that $\widehat{M} = \frac{1}{n} \sum_{i=1}^n \xi_i$.

$$\Rightarrow \text{Example for } \widehat{T}_1 \text{ and } \widehat{T}_2\text{: } M^* = \widehat{P}_1 \widehat{M} \widehat{P}_2 + \frac{1}{n} \sum_{i=1}^n \epsilon_i \xi_i$$

Example under H_1 i.e. $d < d_0$ when $\widehat{\theta}_c$ does not converge

We need to ensure the a.s. convergence of

$$\widehat{\theta}_c = \underset{\mathrm{rank}(M)=d}{\mathrm{argmin}} \| \mathrm{vec}(\widehat{M}) - \mathrm{vec}(M)\|^2 = \widehat{P}_1 \widehat{M} \widehat{P}_2$$

$\Rightarrow$ problem of convergence of eigenprojectors

Riesz formula: $P_\lambda = \oint_{\mathcal{C}_\lambda} (Iz - M)^{-1} \mathrm{d}z.$

Suppose that M and $\widehat{M}$ are symetric with $\widehat{M} \overset{\text{a.s.}}{\to} M$, then

$$\widehat{P} = \oint_{\widehat{\mathcal{C}}} (Iz - \widehat{M})^{-1} \mathrm{d}z$$

Example under H_1 i.e. $d < d_0$ when $\widehat{\theta}_c$ does not converge

We need to ensure the a.s. convergence of

$$\widehat{\theta}_c = \underset{\text{rank}(M)=d}{\text{argmin}} \| \text{vec}(\widehat{M}) - \text{vec}(M)\|^2 = \widehat{P}_1 \widehat{M} \widehat{P}_2$$

$\Rightarrow$ problem of convergence of eigenprojectors

Riesz formula: $P_\lambda = \oint_{\mathcal{C}_\lambda} (Iz - M)^{-1} \mathrm{d}z.$

Suppose that M and $\widehat{M}$ are symetric with $\widehat{M} \overset{\text{a.s.}}{\to} M$, then

$$\widehat{P} = \oint_{\widehat{\mathcal{C}}} (Iz - \widehat{M})^{-1} \mathrm{d}z \underset{\text{from a certain rank}}{\overset{\text{if } \lambda_{p-d+1} \neq \lambda_{p-d}}{=}} \oint_{\mathcal{C}} (Iz - \widehat{M})^{-1} \mathrm{d}z$$

Example under H_1 i.e. $d < d_0$ when $\widehat{\theta}_c$ does not converge

We need to ensure the a.s. convergence of

$$\widehat{\theta}_c = \underset{\text{rank}(M)=d}{\text{argmin}} \| \text{vec}(\widehat{M}) - \text{vec}(M)\|^2 = \widehat{P}_1 \widehat{M} \widehat{P}_2$$

$\Rightarrow$ problem of convergence of eigenprojectors
Riesz formula: $P_\lambda = \oint_{\mathcal{C}_\lambda} (Iz - M)^{-1} \mathrm{d}z.$

Suppose that M and $\widehat{M}$ are symetric with $\widehat{M} \overset{\text{a.s.}}{\to} M$, then

$$\widehat{P} = \oint_{\widehat{\mathcal{C}}} (Iz - \widehat{M})^{-1} \mathrm{d}z \underset{\text{from a certain rank}}{\overset{\text{if } \lambda_{p-d+1} \neq \lambda_{p-d}}{=}} \oint_{\mathcal{C}} (Iz - \widehat{M})^{-1} \mathrm{d}z$$

If $\lambda_{p-d+1} = \lambda_{p-d}$ then P does not exists.

Rk: Application of the previous results to $\widehat{M}\widehat{M}^T$ and $\widehat{M}^T\widehat{M}$.

Conclusion

Concluding remarks

- We provide a general bootstrap procedure for constraint estimator associate to a quadratic function.
- The test procedure associate is consistent.
- Large application thanks to hypothesis testing.
- As an example, it can easily be applied to rank estimation.

Work in progress

- Alleviate the under H_1 assumption $\theta_c \overset{\text{a.s.}}{\to} \theta_c$ for the $\widehat{T}$ stat.
- Possibility to extend such results to $M-$estimator, $Z-$estimator.
- Simulation study : bootstrap vs asymptotic,
  and also constraint bootstrap vs traditional bootstrap.

E. Bura and J. Yang.
Dimension estimation in sufficient dimension reduction: a unifying approach.
J. Multivariate Anal., 102(1):130–142, 2011.

John G. Cragg and Stephen G. Donald.
Inferring the rank of a matrix.
J. Econometrics, 76(1-2):223–250, 1997.

Bradley Efron.
The jackknife, the bootstrap and other resampling plans, volume 38 of *CBMS-NSF Regional Conference Series in Applied Mathematics*.
Society for Industrial and Applied Mathematics (SIAM), Philadelphia, Pa., 1982.

Peter Hall.
The bootstrap and Edgeworth expansion.
Springer Series in Statistics. Springer-Verlag, New York, 1992.

Ker-Chau Li.
Sliced inverse regression for dimension reduction.
J. Amer. Statist. Assoc., 86(414):316–342, 1991.

SIR

Sufficient dimension reduction (SDR) introduit par [Li(1991)]: on suppose le modèle de régression suivant,

$$Y = g(PZ, \varepsilon), \quad Z \perp \varepsilon$$

où $Y \in \mathbb{R}$, $Z \in \mathbb{R}^p$, P est un projecteur orthogonal de rang d_0 et g est inconnue.

- But de la SDR : Estimation de P.
- Enjeux : Obtenir une meilleur vitesse lors de l'estimation de g.

L'inférence sur P se base sur

$$\mathbb{E}[Z|Y] \in \mathrm{Im}(P) \quad \text{p.s.}$$

SIR

On partitionne l'image de Y en H tranches appelées $I(h)$

Enjeux de SIR

Estimer l'espace engendré par les vecteurs

$$\mathbb{E}[Z|Y \in I(1)], \ldots, \mathbb{E}[Z|Y \in I(H)]$$

Procedure de SIR:
1/ Estimation de

$$C_h = \mathbb{E}[Z\mathbb{1}_{\{Y \in I(h)\}}] \in E_c \quad \text{pour } h = 1, ..., H.$$

2/ Extraire une base de $\text{span}(\widehat{C}_1, ..., \widehat{C}_H)$: Elements propres de la matrice

$$\widehat{M}_{SIR} = \sum_h \widehat{p}_h^{-1} \widehat{C}_h \widehat{C}_h^T$$

avec $p_h = \mathbb{P}(Y \in I(h))$.

Trouver la dimension

En notant $\widehat{\eta}_1, ..., \widehat{\eta}_p$ les vecteurs propres de $\widehat{M}_{SIR}$ dans l'ordre croissant des v.p., on peut estimer P de manière consistante par

$$\widehat{P} = \sum_{k=1}^{d_0} \widehat{\eta}_k \widehat{\eta}_k^T,$$

mais d_0 est inconnu.

Importance de bien estimer d_0

- Perte dans la valeur explicative du modèle.
- Vitesse non-paramétrique mauvaise.