Bootstrapping

Irene Zapata Moran, Marco A Zuniga-GaMez
and
Nida Zafar

December 7, 2020

Bootstrapping

Introduction

The term comes from 'The Adventures of Baron Munchausen'

Bootstrapping

Introduction

The term comes from 'The Adventures of Baron Munchausen'

- *"to pull oneself up by the bootstraps"*

Bootstrapping

Introduction

The term comes from 'The Adventures of Baron Munchausen'

- *"to pull oneself up by the bootstraps"*
- A sample is all we have

Bootstrapping

Introduction

The term comes from 'The Adventures of Baron Munchausen'

- *"to pull oneself up by the bootstraps"*
- A sample is all we have
- Treat the sample as the population

Why Bootstrap?

- Instances when we have a small sample

Why Bootstrap?

- Instances when we have a small sample
- Assumed distribution can lead to incorrect hypothesis testing

Bootstrap methods

- Simple bootstrap

Bootstrap methods

- Simple bootstrap
- Methods for time series

Bootstrap methods

- Simple bootstrap
- Methods for time series
- Confidence Intervals

Confidence Intervals

- Construct a confidence interval at level $1 - \alpha$ and α for $\hat{\theta}$.

Confidence Intervals

- Construct a confidence interval at level $1 - \alpha$ and α for $\hat{\theta}$.
- If $\hat{\theta}$ is the OLS estimator of θ

$$\tau = \frac{\hat{\theta} - \theta_0}{\hat{s}_\theta}$$

Confidence Intervals

- Construct a confidence interval at level $1-\alpha$ and α for $\hat{\theta}$.
- If $\hat{\theta}$ is the OLS estimator of θ

$$\tau = \frac{\hat{\theta} - \theta_0}{\hat{s}_\theta}$$

- Percentile t method

$$[\hat{\theta} - \hat{s}\theta t^*_{(1-\alpha/2)(B+1)}, \hat{\theta} - \hat{s}\theta t^*_{(\alpha/2)(B+1)}]$$

Regression

- Resampling Residuals

Regression

- Resampling Residuals
- Resampling Cases

Regression Model

Consider a linear regression model

$$Y = X\beta + \epsilon$$

Y = (n x 1) vector of values of the dependent variable

X = matrix of (n x k) independent variable's data

β = (k x 1) vector of regression coefficients

ϵ = (n x 1) errors terms

Regression Model

Consider a linear regression model

$$Y = X\beta + \epsilon$$

Y = (n x 1) vector of values of the dependent variable

X = matrix of (n x k) independent variable's data

β = (k x 1) vector of regression coefficients

ϵ = (n x 1) errors terms

Regression Model

Consider a linear regression model

$$Y = X\beta + \epsilon$$

$Y = (n \text{ x } 1)$ vector of values of the dependent variable

$X =$ matrix of $(n \text{ x } k)$ independent variable's data

$\beta = (k \text{ x } 1)$ vector of regression coefficients

$\epsilon = (n \text{ x } 1)$ errors terms

Regression Model

Consider a linear regression model

$$Y = X\beta + \epsilon$$

$Y = (n \text{ x } 1)$ vector of values of the dependent variable

$X =$ matrix of $(n \text{ x } k)$ independent variable's data

$\beta = (k \text{ x } 1)$ vector of regression coefficients

$\epsilon = (n \text{ x } 1)$ errors terms

Resampling Residuals

Steps

Step 1: Take random data for X and Y from the population and estimate the $\hat{\beta}$s using OLS.

Resampling Residuals

Steps

Step 1: Take random data for X and Y from the population and estimate the $\hat{\beta}$s using OLS.

Step 2: Using these $\hat{\beta}$s and the observed Y values, we calculate the residuals,

$$\hat{\epsilon}_i = Y_i - \hat{Y}_i.$$

Step 3: Now, we re-sample these residuals with replacement. And create a bootstrapped vector of the independent variable

$$Y_b^* = \hat{Y} + \hat{\epsilon_b^*}$$

Step 3: Now, we re-sample these residuals with replacement. And create a bootstrapped vector of the independent variable

$$Y_b^* = \hat{Y} + \hat{\epsilon_b^*}$$

Step 4: These bootstrapped independent variables are then regressed on the previous (fixed), independent variables and a vector of bootstrapped $\hat{\beta_b^*}$ for this particular re-sample is obtained.

$$\hat{Y_b^*} = X\hat{\beta_b^*} + \hat{\epsilon}$$

Step 3 and 4 are repeated B times, to obtain a (B x k) matrix, with each row representing βs of one re-sample.

Resampling Cases

- Flexibility of resampling the rows of data matrix

Resampling Cases

- Flexibility of resampling the rows of data matrix
- Relaxing the non-stochastic regressors

Resampling Cases

- Flexibility of resampling the rows of data matrix
- Relaxing the non-stochastic regressors
- B samples of size n are generated

Resampling Cases

- Flexibility of resampling the rows of data matrix
- Relaxing the non-stochastic regressors
- B samples of size n are generated
- A regression model estimated for each resample

Resampling Cases

- Flexibility of resampling the rows of data matrix
- Relaxing the non-stochastic regressors
- B samples of size n are generated
- A regression model estimated for each resample
- (B x k) matrix of bootstrapped regression coefficients

Resampling Cases

Depending on the one's model

- If the model has stochastic component

Resampling Cases

Depending on the one's model

- If the model has stochastic component
- If the errors are heteroscedastic

Resampling Cases

Depending on the one's model

- If the model has stochastic component
- If the errors are heteroscedastic
- Survey research where independent variables are as random as dependent variable, data resampling works better

Resampling Cases

Depending on the one's model

- If the model has stochastic component
- If the errors are heteroscedastic
- Survey research where independent variables are as random as dependent variable, data resampling works better
- However, most theoretical statisticians suggest, residuals approach

MATLAB