

# STK4051/STK9051 Computational Statistics

Riccardo De Bin

debin@math.uio.no

## Outline of the lecture I

- The Bootstrap Principle
- Basic Methods
  - Non-parametric Bootstrap
  - Parametric Bootstrap
  - Bootstrapping Regression
  - Bootstrap Bias Correction
- Bootstrap Inference
  - Percentile Method
  - Pivoting
    - Accelerated Bias-Corrected Percentile Method,  $BC_a$
    - The Bootstrap  $t$
    - Empirical Variance Stabilization
    - Nested Bootstrap and Pre-pivoting

## Bootstrap: Introduction

Bootstrapping is a computational intensive method that allows researchers to simulate the distribution of a statistic:

- main idea: repeatedly resample the observed data, producing an empirical distribution function from the resampled data;
- compute a new value of the statistic, and the collection of these values provides an estimate of the sampling distribution;
- the method allows you to “pull yourself up by your bootstraps” (Raspe, 1785);
- bootstrapping is nonparametric by nature.

## The Bootstrap Principle: Introduction

Consider a quantity of interest  $\theta = T(F)$ ,

- functional of a distribution function  $F$ ;
- e.g.,  $\theta = T(F) = \int z dF(z)$  is the mean.

Let  $x_1, \dots, x_n$  come from i.i.d. random variables  $X_1, \dots, X_n$

- where  $X_i, i = 1, \dots, n$  has cumulative distribution function  $F$ ;
- with abuse of notation, in the book  $X_i \sim F$ .

An estimate of  $\theta$ ,  $\hat{\theta}$ , can be obtained as  $T(\hat{F})$ ,

- $\hat{F}$  is the empirical distribution function;
- in the case of the mean,  $\hat{\theta} = \int z d\hat{F}(z) = \frac{\sum_{i=1}^n X_i}{n}$ .

## The Bootstrap Principle: Introduction

Statistical inference questions are usually based on some  $R(\mathcal{X}, F)$ ,

- e.g.  $R(\mathcal{X}, F) = \frac{T(\hat{F}) - T(F)}{S(\hat{F})}$ ,
  - $S(\cdot)$  estimates the standard deviation of  $T(\hat{F})$ ;
- often such that the distribution of  $R(\mathcal{X}, F)$  is available;
- when it is not, bootstrap provides an approximation.

Let  $\mathcal{X}^*$  denote a bootstrap sample (pseudo-dataset):

- $\mathcal{X}^* = \{X_1^*, \dots, X_n^*\}$  are i.i.d. r.v. with distribution  $\hat{F}$ ;
- the bootstrap strategy is to examine  $R(\mathcal{X}, \hat{F})$ ;
- i.e., the random variable formed by applying  $R$  to  $\mathcal{X}^*$ .

**The bootstrap principle is to equate the distributions of  $R(\mathcal{X}, F)$  and  $R(\mathcal{X}^*, \hat{F})$ .**

## The Bootstrap Principle: Example

Consider:

- $n = 3$  data points, namely  $\{x_1, x_2, x_3\} = \{1, 2, 6\}$ ;
- realizations of i.i.d. r.v. with distribution  $F$ , that has mean  $\theta$ ;
- $\hat{F}$  places mass  $1/3$  on each point;
- we want to bootstrap the estimator of the sample mean,  $\hat{\theta}$ ;
- $T(\hat{F}) = n^{-1} \sum_{i=1}^n X_i = R(\mathcal{X}, \hat{F}) = R(\mathcal{X})$ ;

Let  $\mathcal{X}^* = \{X_1^*, X_2^*, X_3^*\}$ ,

- there are  $3^3 = 27$  possible outcomes;
- since  $\hat{\theta}$  does not depend on the ordering of the data, only 10 distinct possible outcomes:  $(1, 1, 1), (1, 1, 2), (1, 2, 2), (2, 2, 2), (1, 1, 6), (1, 2, 6), (2, 2, 6), (1, 6, 6), (2, 6, 6), (6, 6, 6)$ .

## The Bootstrap Principle: Example

| $\mathcal{X}^*$ | $\hat{\theta}^*$ | $P^*[\hat{\theta}^*]$ | Observed Frequency |
|-----------------|------------------|-----------------------|--------------------|
| 1 1 1           | 3/3              | 1/27                  | 36/1000            |
| 1 1 2           | 4/3              | 3/27                  | 101/1000           |
| 1 2 2           | 5/3              | 3/27                  | 123/1000           |
| 2 2 2           | 6/3              | 1/27                  | 25/1000            |
| 1 1 6           | 8/3              | 3/27                  | 104/1000           |
| 1 2 6           | 9/3              | 6/27                  | 227/1000           |
| 2 2 6           | 10/3             | 3/27                  | 131/1000           |
| 1 6 6           | 13/3             | 3/27                  | 111/1000           |
| 2 6 6           | 14/3             | 3/27                  | 102/1000           |
| 6 6 6           | 18/3             | 1/27                  | 40/1000            |

(Givens & Hoeting, 2013, table 9.1)

## The Bootstrap Principle: Example

Where:

- $P^*(\hat{\theta}^*)$  is the probability distribution for  $\hat{\theta}^*$  with respect to the bootstrap experiment of drawing  $\mathcal{X}^*$  conditional on the original observations;
- e.g.,  $P^*(\hat{\theta}^* \leq 6/3) = 8/27$ ;

In this simple example,

- “equate the distributions of  $R(\mathcal{X}, F)$  and  $R(\mathcal{X}^*, \hat{F})$ ” means to base inference on the distribution of  $\hat{\theta}^*$ ;
- e.g., an  $\approx 93\%$  (25/27) confidence interval for  $\theta$  is  $(4/3, 14/3)$ ;
- the point estimate remains  $\hat{\theta} = 9/3$  (from the original sample).

## Non-parametric Bootstrap

For realistic sample sizes:

- the number of possible pseudo-datasets is very large;
- complete enumeration of the possibilities is not feasible;

Instead:

- $B$  independent bootstrap pseudo-datasets are drawn from  $\hat{F}$ ;
- $\mathcal{X}_i^* = \{X_{i1}^*, \dots, X_{in}^*\}$ ,  $i = 1, \dots, B$ ;
- The empirical distribution of the  $R(\mathcal{X}_i^*, \hat{F})$  is used to approximate the distribution of  $R(\mathcal{X}, F)$ .

Note:

- the simulation error decreases with  $B$ ;
- inference without any parametric assumption;
- answers to problems without analytic solutions;
- accurate (may be more than standard parametric theory).

## Parametric Bootstrap: Introduction

When  $X_1, \dots, X_n$  i.i.d. with cumulative distribution  $F(x, \theta)$ ,

- data modelled to originate from a parametric distribution;

another estimate of  $F$  may be employed:

- the observed data are used to estimate  $\theta$ ;
- each pseudo-dataset is generated by drawing from  $F(x, \hat{\theta})$ .

Note:

- when the model is a good representation of reality, the parametric bootstrap can be a powerful tool,
  - allowing inference in otherwise intractable situations;
  - producing confidence intervals more accurate than those produced by standard asymptotic theory;
- when the model does not really approximate the mechanism generating the data, bootstrap can lead to bad inference.

## Non-parametric Bootstrap: Example

Continuing with the dataset before:

- recall that  $\hat{F}$  places mass  $1/3$  on 1, 2, and 6;
- a non-parametric bootstrap generates  $\mathcal{X}_i^*$  by sampling  $X_{i1}^*$ ,  $X_{i2}^*$ , and  $X_{i3}^*$  from  $\hat{F}$ ;
- basically, draw  $X_{ij}^*$  from  $\{1, 2, 6\}$  with replacement;
- for each  $\mathcal{X}_i^*$ , compute  $\hat{\theta}^*$ ;
- the relative frequencies approximates  $P(\hat{\theta}^*)$ .

Note:

- this hold only because  $X_1$ ,  $X_2$ , and  $X_3$  are i.i.d.;
- otherwise  $R(\mathcal{X}_i^*, \hat{F})$  does not approximate  $R(\mathcal{X}, F)$ ;
- more on this issue later.

## Bootstrapping Regression: naive approach

Consider the regression model,  $Y_i = x_i^T \beta + \epsilon_i$ ,  $i = 1, \dots, n$ :

- $\epsilon_i$  i.i.d. with 0 mean and constant variance;
- $x_i$  is a  $p$ -dimensional vector of predictors (covariates);
- $\beta$  a  $p$ -dimensional vector of parameters.

Naive (wrong) bootstrapping:

- resample from the collection of response values a new pseudo-response, say  $Y_i^*$ , for each observed  $x_i$ ;
- generating a new regression dataset;
- a bootstrap parameter vector estimate,  $\hat{\beta}^*$  is calculated from these pseudo-data;
- the empirical distribution of  $\hat{\beta}^*$  is used for inference about  $\beta$ .

## Bootstrapping Regression: correct approach

The mistake is that the  $Y_i|x_i$  are not i.i.d., they have different conditional means.

Given the model,  $\epsilon_i$  are i.i.d.  $\Rightarrow$  bootstrap the residuals:

- fit the regression model to the observed data,
  - obtaining  $\hat{y}_i$  (fitted responses);
  - and the residuals  $\hat{\epsilon}_i$ ;
- sample a random bootstrap set of residuals  $\{\hat{\epsilon}_1^*, \dots, \hat{\epsilon}_n^*\}$ ,
  - sampling with replacement from  $\{\hat{\epsilon}_1, \dots, \hat{\epsilon}_n\}$ ;
- create a bootstrap set of pseudo-responses,  $Y_i^* = \hat{y}_i + \hat{\epsilon}_i^*$ ,  $i = 1, \dots, n$ ;
- regress  $Y^*$  on  $x$  to obtain a bootstrap parameter estimate  $\hat{\beta}^*$ ;
- use the empirical distribution of  $\hat{\beta}^*$  for inference.

## Bootstrapping Regression: paired bootstrap

This approach is sometimes called the paired bootstrap:

- good if any doubts about the adequacy of the model;
- or on any regression assumption;
- less sensitive to violations in the assumptions than bootstrapping the residuals;
- more directly mirrors the original data generation mechanism.

## Bootstrapping Regression: paired bootstrap

We are assuming:

- the  $x_i$  are fixed (experimental design);
- homoscedasticity;
- the chosen model provides an appropriate fit to the data.

When both response and predictors are random:

- e.g., data from an observational study;
- the data pairs  $z_i = (x_i, y_i)$  are values from i.i.d. r.v.  $Z_i = (X_i, Y_i)$ , with a joint response-predictors distribution;
- sample with replacement from  $\{z_1, \dots, z_n\}$ ;
- apply the regression model to the pseudo-dataset  $\rightarrow \hat{\beta}^*$ ;
- repeat these steps many times, then proceed to inference.

## Bootstrapping Regression: Example

Consider the book example (Copper-Nickel Alloy):

- 13 measurements of corrosion loss ( $y_i$ ) in copper-nickel alloys;
- each with a specific iron content ( $x_i$ );

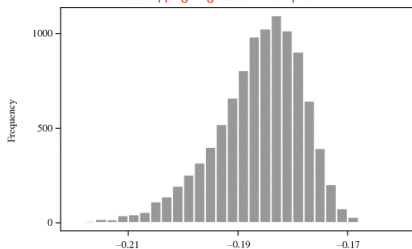
|       |       |       |       |       |       |       |       |
|-------|-------|-------|-------|-------|-------|-------|-------|
| $x_i$ | 0.01  | 0.48  | 0.71  | 0.95  | 1.19  | 0.01  | 0.48  |
| $y_i$ | 127.6 | 124.0 | 110.8 | 103.9 | 101.5 | 130.1 | 122.0 |
| $x_i$ | 1.44  | 0.71  | 1.96  | 0.01  | 1.44  | 1.96  |       |
| $y_i$ | 92.3  | 113.1 | 83.7  | 128.0 | 91.4  | 86.2  |       |

- of interest: change in corrosion loss in the alloys as the iron content increases, relative to the corrosion loss without iron;
- in formula,  $\theta = \beta_1/\beta_0$ .

Results:

- paired bootstrapping approach on  $z_i = (x_i, y_i)$ ,  $i = 1, \dots, 13$ ;
- from the observed data,  $\hat{\theta} = \hat{\beta}_1/\hat{\beta}_0 = -0.185$ ;
- distribution of  $\hat{\theta}$  based on 10,000 bootstrap samples.

## Bootstrapping Regression: Example



(Givens & Hoeting, 2013, figure 9.1)

STK4051: Lecture 7

17/ 43 STK4051: Lecture 7

18/ 43

## Bootstrap Inference: Percentile Method

Inference about a univariate  $\theta$  using bootstrap simulations:

- construct a confidence interval using the percentile method,
  - read the percentiles of the histogram of  $\hat{\theta}_i^*$ ;
- perform hypothesis testing based on the p-value on a bootstrap confidence interval,
  - if the  $(1 - \alpha)100\%$  bootstrap confidence interval does not cover the null value, then the null hypothesis is rejected;
  - often sacrifices statistical power;
  - techniques to mitigate this issue.

Note:

- the percentile method is simple, but prone to bias and inaccurate coverage probabilities;
- the bootstrap works better when  $\theta$  is a location parameter;
- bootstrapped statistic should be approximately pivotal,
  - its distribution should not depend on the true value of  $\theta$ ;
  - variance-stabilizing transformations help.

STK4051: Lecture 7

19/ 43 STK4051: Lecture 7

## Bootstrap Bias Correction

A particularly interesting choice for bootstrap analysis when  $T(F) = \theta$  is the quantity  $R(\mathcal{X}, F) = T(\hat{F}) - T(F)$ :

- represents the bias of  $T(\hat{F}) = \hat{\theta}$ ;
- it has mean equal to  $E[\hat{\theta}] - \theta$ .
- the bootstrap estimate is  $\sum_{i=1}^B (\hat{\theta}_i^* - \hat{\theta})/B = \bar{\theta}^* - \hat{\theta}$ ;
- one may consider  $T(\bar{F}^*)$  instead of  $\hat{\theta}$ .

Example of the book (Copper-Nickel Alloy, continued):

- the mean value of  $\hat{\theta}^* - \hat{\theta}$  among the pseudo-datasets is  $-0.00125$ ;
- the bias-corrected bootstrap estimate of  $\beta_1/\beta_0$  is  $-0.18507 - (-0.00125) = -0.184$ .

## Bootstrap Inference: Justification for the Percentile Method

Consider a continuous, strictly increasing transformation  $\phi$  and a distribution function  $H$

- continuous and symmetric, namely  $H(z) = 1 - H(-z)$ ;
- $P[h_{\alpha/2} \leq \phi(\hat{\theta}) - \phi(\theta) \leq h_{1-\alpha/2}] = 1 - \alpha$ 
  - where  $h_\alpha$  is the  $\alpha$  quantile of  $H$ ;
  - e.g., with  $\phi$  is a normalizing transformation, then  $H$  is the standard normal distribution.
- when  $F$  is continuous, we may transform any random variable  $X \sim F$  to have any desired distribution  $G$ , using the monotone transformation  $G^{-1}(F(X))$ ;
- the remarkable aspect of the percentile approach is that we are never actually required to specify explicitly  $\phi$  or  $H$ ;

20/ 43

## Bootstrap Inference: Justification for the Percentile Method

Applying the bootstrap principle, we have

$$\begin{aligned} 1 - \alpha &\approx P[h_{\alpha/2} \leq \phi(\hat{\theta}^*) - \phi(\hat{\theta}) \leq h_{1-\alpha/2}] \\ &= P[h_{\alpha/2} + \phi(\hat{\theta}) \leq \phi(\hat{\theta}^*) \leq h_{1-\alpha/2} + \phi(\hat{\theta})] \\ &= P[\underbrace{\phi^{-1}(h_{\alpha/2} + \phi(\hat{\theta}))}_{\xi_{\alpha/2}} \leq \hat{\theta}^* \leq \underbrace{\phi^{-1}(h_{1-\alpha/2} + \phi(\hat{\theta}))}_{\xi_{1-\alpha/2}}] \end{aligned}$$

where

- $\xi_{\alpha}$  denotes the  $\alpha$  quantile of the empirical distribution of  $\hat{\theta}^*$ ;
- known from the bootstrap distribution.

## Bootstrap Inference: Justification for the Percentile Method

Since  $P[h_{\alpha/2} \leq \phi(\hat{\theta}) - \phi(\theta) \leq h_{1-\alpha/2}] = 1 - \alpha$  similarly leads to

$$P[\phi^{-1}(h_{\alpha/2} + \phi(\hat{\theta})) \leq \theta \leq \phi^{-1}(h_{1-\alpha/2} + \phi(\hat{\theta}))] = 1 - \alpha,$$

using the fact that  $h_{\alpha/2} = -h_{1-\alpha/2}$ , and the two expressions are asymptotically equivalent,

then

we can simply read off the quantiles from the bootstrap distribution and use these as the confidence limits for  $\theta$ .

## Pivoting: Accelerated Bias-Corrected Percentile Method, $BC_a$

The accelerated bias-corrected percentile method,  $BC_a$ ,

- substantial improvement over the simple percentile approach;
- we saw that for the simple approach to work:
  - the transformed estimator must be unbiased;
  - with variance that does not depend on  $\theta$ ;
- $BC_a$  augments  $\phi$  with two parameters  $a$  and  $b$ 
  - to better meet these conditions;
  - thereby ensuring an approximate pivot.

## Pivoting: Accelerated Bias-Corrected Percentile Method, $BC_a$

Assume  $\exists \phi$  and constants  $a$  and  $b$  such that

$$U = \frac{\phi(\hat{\theta}) - \phi(\theta)}{1 + a\phi(\theta)} + b \sim N(0, 1)$$

- $1 + a\phi(\theta) > 0$ ;
- if  $a = b = 0$ , back to the simple percentile method.

By the bootstrap principle,

$$U^* = \frac{\phi(\hat{\theta}^*) - \phi(\hat{\theta})}{1 + a\phi(\hat{\theta})} + b \sim N(0, 1),$$

so, denoting by  $z_{\alpha}$  a quantile of the standard normal distribution,

$$\alpha \approx P^*[U^* \leq z_{\alpha}] = P^*[\underbrace{\hat{\theta}^* \leq \phi^{-1}(\phi(\hat{\theta}) + (z_{\alpha} - b)[1 + a\phi(\hat{\theta})])}_{\xi_{\alpha}}]$$

where  $\xi_{\alpha}$  is observable from the bootstrap distribution.

### Pivoting: Accelerated Bias-Corrected Percentile Method, $BC_a$

To use this fact, go back to  $U$ .

$$1 - \alpha = P[U > z_\alpha] = P[\theta < \phi^{-1}(\phi(\hat{\theta}) + u(a, b, \alpha)[1 + a\phi(\hat{\theta})])]$$

where  $u(a, b, \alpha) = (b - z_\alpha)/[1 - a(b - z_\alpha)]$ .

We need to find the  $\beta$  such that  $u(a, b, \alpha) = z_\beta - b$ , so we can apply the bootstrap principle to conclude that  $\theta < \beta$  approximates a  $1 - \alpha$  upper confidence limit. After some algebra,

$$\beta = \Phi(b + u(a, b, \alpha)) = \Phi\left(b + \frac{b + z_{1-\alpha}}{1 - a(b + z_{1-\alpha})}\right)$$

where  $\Phi(\cdot)$  is the standard normal cumulative function.

### Pivoting: Accelerated Bias-Corrected Percentile Method, $BC_a$

As with the percentile method:

- explicit specification of the transformation  $\phi$  is not necessary;
- $BC_a$  approach merely corrects the percentile levels determining the confidence interval endpoints to be read from the bootstrap distribution;
- it shares the transformation-respecting property of the simple percentile method.

### Pivoting: Accelerated Bias-Corrected Percentile Method, $BC_a$

If we knew a suitable  $a$  and  $b$ , then to find a  $1 - \alpha$  upper confidence limit:

- compute  $\beta$ ;
- find the  $\beta$ -th quantile of the empirical distribution of  $\hat{\theta}^*$ ,  $\xi_\beta$ .

Similarly, for a two-sided  $1 - \alpha$  interval,  $P[\xi_{\beta_1} \leq \theta \leq \xi_{\beta_2}] \approx 1 - \alpha$ ,

- compute

$$\beta_1 = \Phi\left(b + \frac{b + z_{\alpha/2}}{1 - a(b + z_{\alpha/2})}\right)$$

and

$$\beta_2 = \Phi\left(b + \frac{b + z_{1-\alpha/2}}{1 - a(b + z_{1-\alpha/2})}\right)$$

- find the corresponding  $\xi_{\beta_1}$  and  $\xi_{\beta_2}$ .

### Pivoting: Accelerated Bias-Corrected Percentile Method, $BC_a$

The remaining question is the choice of  $a$  and  $b$ . The simplest non-parametric choices are:

$$b = \Phi^{-1}\left(\hat{F}^*(\hat{\theta})\right)$$

and

$$a = \frac{1}{6} \sum_{i=1}^n \psi_i^3 / \left( \sum_{i=1}^n \psi_i^2 \right)^{3/2}$$

where  $\psi_i = \hat{\theta}_{(i)} - \hat{\theta}_{(-i)}$ ,

- $\hat{\theta}_{(-i)}$  is the statistic computed omitting the  $i$ -th observation;
- $\hat{\theta}_{(i)} = n^{-1} \sum_{i=1}^n \hat{\theta}_{(-i)}$ .

## Pivoting: Example

In the book example (Copper-Nickel Alloy, continued),

- $a = 0.0486$ ;
- $b = 0.00802$ ;

The adjusted quantiles are therefore:

- $\beta_1 = 0.038$ ;
- $\beta_2 = 0.986$ .

The main effect of  $BC_a$  is therefore to shift the confidence interval slightly to the right. The resulting interval is  $(-0.203, -0.172)$ .

## Pivoting: The Bootstrap t

The bootstrap t method,

- approximate pivot quite easy to implement;
- also called the studentized bootstrap.

Suppose  $\theta = T(F)$  is to be estimated using  $\hat{\theta} = T(\hat{F})$ ,

- if  $V(\hat{F})$  estimates the variance of  $\hat{\theta}$ ,
- $R(\mathcal{X}, F) = \frac{T(\hat{F}) - T(F)}{V(\hat{F})}$  is roughly pivotal;
- bootstrapping  $R(\mathcal{X}, F)$  yields  $R(\mathcal{X}^*, \hat{F})$ .

## Pivoting: The Bootstrap t

Denote by  $\hat{G}$  and  $G$  the distributions of  $R(\mathcal{X}, F)$  and  $R(\mathcal{X}^*, \hat{F})$ ,

$$\begin{aligned} P[\xi_{\alpha/2}(\hat{G}) \leq R(\mathcal{X}, F) \leq \xi_{1-\alpha/2}(\hat{G})] \\ = P[\hat{\theta} - \sqrt{V(\hat{F})}\xi_{1-\alpha/2}(\hat{G}) \leq \theta \leq \hat{\theta} - \sqrt{V(\hat{F})}\xi_{\alpha/2}(\hat{G})] = 1 - \alpha \end{aligned}$$

where  $\xi_{\alpha}(\hat{G})$  is the  $\alpha$  quantile of  $\hat{G}$ .

The  $\xi_{\alpha}(\hat{G})$  are unknown but invoking the bootstrap principle,  $\xi_{\alpha}(\hat{G}) \approx \xi_{\alpha}(\hat{G}^*) \forall \alpha$ . Therefore the confidence interval is

$$\left( T(\hat{F}) - \sqrt{V(\hat{F})}\xi_{1-\alpha/2}(\hat{G}^*), T(\hat{F}) - \sqrt{V(\hat{F})}\xi_{\alpha/2}(\hat{G}^*) \right)$$

where  $\xi_{\alpha}(\hat{G}^*)$  are taken from the histogram of  $R(\mathcal{X}^*, \hat{F})$  values :

- these percentiles are in the tail of the distribution;
- several thousand bootstrap pseudo-datasets are needed for adequate precision.

## Pivoting: The Bootstrap t

Example from the book (Copper-Nickel Alloy, continued):

- an estimator  $V(\hat{F})$  of the variance of  $\hat{\beta}_1/\hat{\beta}_0$  is

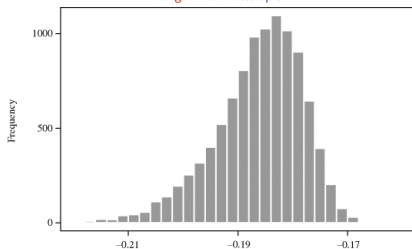
$$\left( \frac{\hat{\beta}_1}{\hat{\beta}_0} \right)^2 \left( \frac{\widehat{\text{var}}[\hat{\beta}_1]}{\hat{\beta}_1^2} + \frac{\widehat{\text{var}}[\hat{\beta}_0]}{\hat{\beta}_0^2} - \frac{2\widehat{\text{cov}}[\hat{\beta}_0, \hat{\beta}_1]}{\hat{\beta}_0\hat{\beta}_1} \right)$$

- based on the delta method;

Results:

- $\xi_{0.025}(\hat{G}^*) = -5.77$ ;
- $\xi_{0.975}(\hat{G}^*) = 4.44$ ;
- $\sqrt{V(\hat{F})} = 0.00273$ ;
- the resulting confidence interval is  $(-0.197, -0.169)$ .

## Pivoting: The Bootstrap t



(Givens & Hoeting, 2013, figure 9.2)

## Pivoting: Empirical Variance Stabilization

A variance-stabilizing transformation:

- is often the basis for a good pivot;
- make the sampling variance of the transformed estimator independent of  $\theta$ ;
- usually a variance-stabilizing transformation of the statistic to be bootstrapped is unknown;
- it can be estimated using the bootstrap.

## Pivoting: Empirical Variance Stabilization

Start by drawing  $B_1$  bootstrap pseudo-datasets  $\mathcal{X}_j^*$ ,  $j = 1, \dots, B_1$ .

For each of them

- calculate  $\hat{\theta}_j^*$ ;
- call  $\hat{F}_j^*$  the empirical distribution of  $\mathcal{X}_j^*$ .

For each  $\mathcal{X}_j$ ,

- draw  $B_2$  pseudo-datasets  $\mathcal{X}_{jk}^{**}$  from  $\hat{F}_j^*$ ;
- compute the  $\hat{\theta}_{jk}^{**}$ ;
- let  $\bar{\theta}_{jk}^{**}$  the mean of  $\hat{\theta}_{jk}^{**}$ .

Then an estimator of the se of  $\hat{\theta}$  given  $\theta = \hat{\theta}_j^*$  is

$$\hat{s}(\hat{\theta}_j^*) = \frac{1}{B_2 - 1} \sum_{k=1}^{B_2} (\hat{\theta}_{jk}^{**} - \bar{\theta}_{jk}^{**}).$$

## Pivoting: Empirical Variance Stabilization

Fit a curve to the set of points  $\{\hat{\theta}_j^*, \hat{s}(\hat{\theta}_j^*)\}$ ,  $j = 1, \dots, B_1$ ,

- e.g., using local density estimation;
- the curve estimates the relationship between  $\hat{s}(\hat{\theta}_j^*)$  and  $\theta$ ;
- find a transformation  $g$  that neutralises this relationship.

After the variance-stabilizing transformation has been estimated:

- draw  $B_3$  new bootstrap sample from  $\hat{F}$ ;
- apply bootstrap t to find a confidence interval for  $g(\hat{\theta})$ ;
- convert the endpoints of the interval back to the scale of  $\theta$  through  $g^{-1}$ .

## Pivoting: Nested Bootstrap and Pre-pivoting

The nested bootstrap:

- also called the iterated or double bootstrap;
- based on the Empirical Variance Stabilization idea.

Consider constructing a confidence interval or conducting a hypothesis test

- based on a test statistic  $R_0(\mathcal{X}, F)$ ;
- given observed data values  $x_1, \dots, x_n$ ;
- realizations of  $X_1, \dots, X_n$  i.i.d. with  $X_i \sim F$ ;

## Pivoting: Nested Bootstrap and Pre-pivoting

Use classical non-parametric bootstrap to compute:

- $R_0(\mathcal{X}^*, \hat{F})$ ;
- $\hat{F}_0(q, \hat{F}) = (1/B) \sum_{j=1}^B 1(R_0(\mathcal{X}_j, \hat{F}) \leq q)$ ;
- for the bootstrap principle,
  - $\hat{F}_0(q, \hat{F})$  estimates  $P^*[R_0(\mathcal{X}^*, \hat{F}) \leq q]$ ;
  - $P^*[R_0(\mathcal{X}^*, \hat{F}) \leq q]$  estimates  $P[R_0(\mathcal{X}, F) \leq q] = F_0(q, F)$ ;
- the confidence interval has upper limit  $\hat{F}_0^{-1}(1 - \alpha/2, \hat{F})$ ;
- reject  $H_0$  if  $R_0(\{x_1, \dots, x_n\}, F) > \hat{F}_0^{-1}(1 - \alpha, \hat{F})$ .

Note, however:

- the confidence interval does not have exact coverage  $1 - \alpha$ ;
- $\hat{F}_0$  a bootstrap approximation of the distribution of  $R_0(\mathcal{X}, F)$ ;
- similarly for the size of the hypothesis test.

## Pivoting: Nested Bootstrap and Pre-pivoting

Let  $F_0(q, F) = P[R_0(\mathcal{X}, F) \leq q]$ ,

- a two-sided confidence interval can be based on

$$P[F_0^{-1}(\alpha/2, F) \leq R_0(\mathcal{X}, F) \leq F_0^{-1}(1 - \alpha/2, F)] = 1 - \alpha;$$

- a hypothesis test based on

$$P[R_0(\mathcal{X}, F) \leq F_0^{-1}(1 - \alpha, F)] = 1 - \alpha.$$

These probability statements depend on the quantiles of  $F_0$ :

- are not known;
- we can use the bootstrap to approximate  $F_0$  and its quantiles;
- $R_0(\mathcal{X}, F)$  and  $F$  are also unknown ( $F$  is hypothesised in the hypothesis test case).

## Pivoting: Nested Bootstrap and Pre-pivoting

Define the random variable  $R_1(\mathcal{X}, F) = F_0(R_0(\mathcal{X}, F), F)$ ,

- is a pivot (uniform distribution independent of  $F$ );
- we do not have it because  $F_0$  unknown;
- use the bootstrap principle,  $\hat{F}_0 \approx F_0$ ;
- and  $\hat{R}_1(\mathcal{X}, F) = \hat{F}_0(R_0(\mathcal{X}, F), \hat{F})$ ;
- this allow inference, as  $\hat{R}_1(\mathcal{X}, F) \sim Unif$ .

Alternatively,

- we know that  $\hat{R}_1(\mathcal{X}, F) \sim F_1$ , for some  $F_1$ ;
- $\hat{F}_1(q, F) = P[\hat{R}_1(\mathcal{X}, F) \leq q]$ ;
- the correct size test rejects the null hypothesis if  $\hat{R}_1(\{x_1, \dots, x_n\}, F) > F_1^{-1}(1 - \alpha, F)$ ;
- confidence interval with the correct coverage probability,  $P[F_1^{-1}(\alpha/2, F) \leq \hat{R}_1(\mathcal{X}, F) \leq F_1^{-1}(1 - \alpha/2, F)]$ ;
- $F_1$  is unknown and must be estimated by bootstrap.

## Pivoting: Nested Bootstrap and Pre-pivoting

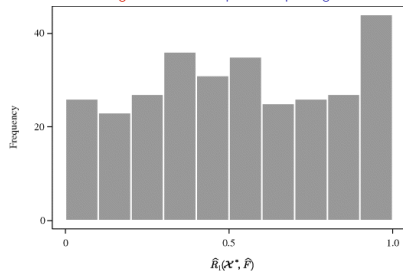
Now the randomness comes from two sources:

1. the observed data are random observations from  $F$ ;
2. given the data ( $\hat{F}$ ),  $\hat{R}_1$  is calculated by random resampling;
3. to capture both sources, nested bootstrap.

Example of the book (Copper-Nickel Alloy, continued),

- $R_0(\{x_1, \dots, x_n\}, F) = \hat{\beta}_1/\hat{\beta}_0 - \beta_1/\beta_0$ ;
- a histogram of  $\hat{R}_1$  obtained by the nested bootstrap;
- the nested bootstrap gave 0.025 and 0.975 quantiles of  $\hat{R}_1$  as 0.0316 and 0.990;
- the 3.16% and 99.0% percentiles of  $R_0(\mathcal{X}^*, \hat{F})$  are used for the confidence interval of  $\beta_1/\beta_0$ , namely  $(-0.197, -0.168)$ .

## Pivoting: Nested Bootstrap and Pre-pivoting



(Givens & Hoeting, 2013, figure 9.3)

## References |

GIVENS, G. H. & HOETING, J. A. (2013). *Computational Statistics*. Wiley.

RASPE, R. E. (1785). *Baron Munchausen's Narrative of His Marvellous Travels and Campaigns in Russia*.