GMM, Indirect Inference and Bootstrap

Bootstrap

Willi Mutschler

TU Dortmund

Winter 2015/2016

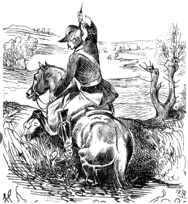
Bootstrap

Meaning

verb tr.: To help oneself with one's own initiative and no outside help.
noun: Unaided efforts.
adjective: Reliant on one's own efforts.
(http://wordsmith.org/words/bootstrap.html)

Etymology

While pulling on bootstraps may help with putting on one's boots, it's impossible to lift oneself up like that. Nonetheless the fanciful idea is a great visual and it gave birth to the idiom "to pull oneself up by one's (own) bootstraps", meaning to better oneself with one's own efforts, with little outside help. It probably originated from the tall tales of Baron Münchausen who claimed to have lifted himself (and his horse) up from the swamp by pulling on his own hair.
(http://wordsmith.org/words/bootstrap.html)



Point of departure

- Unknown distribution function F
- Simple random sample $x_1, \dots, x_n$ from F
- Make inference about a population characteristic θ, using a statistic T, whose value is t in the sample
  - What are bias, standard error or quantiles of T?
  - What are likely values under a certain null hypothesis?
  - How do we compute confidence intervals?

Key idea

- Resample from original data either directly or via a fitted model
- Assess variability of quantities of interest from replicate datasets without (long-winded and error-prone) analytical calculation

- **Basic bootstrap idea:** Approximate the unknown distribution of

$$T(X_1, \ldots, X_n) \text{ for } X_1, \ldots, X_n \text{ i.i.d. from } F$$

by the distribution of

$$T(X_1^*, \ldots, X_n^*) \text{ for } X_1^*, \ldots, X_n^* \text{ i.i.d. from } \hat{F}$$

- The distribution of T under $\hat{F}$ is usually found by Monte-Carlo simulations based on resamples (pseudo sample)

- How is F estimated?
  - parametric bootstrap: $\hat{F}$ is based on a fitted parametric distribution depending on parameters ψ, then $\hat{F}_{\hat{\psi}}$ and F_ψ have same form
  - nonparametric bootstrap: $\hat{F}$ is based on the empirical distribution function F_n
  - smooth bootstrap: $\hat{F}$ is based on a smoothed empirical distribution function with a kernel and bandwidth
  - model based: $\hat{F}$ is based on simulated values generated from a fitted model
- Applications of bootstrap
  - bias and standard errors
  - confidence intervals
  - hypothesis tests
  - check robustness and relax assumptions
  - check adequacy of theoretical properties and measures
  - get quick approximate solutions, if theoretical calculations are too complex or untrustworthy

- Nonparametric bootstrap of the standard error of

$$T = \bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$$

- Simple random sample from $X_1, \ldots, X_n$ iid
- Estimation of the unknown cdf F by the empirical distribution function

$$F_n(x) = \frac{1}{n}\sum_{i=1}^{n} 1\left(X_i \leq x\right)$$

- How is $\bar{X}$ distributed under F ?
- How is $\bar{X}$ distributed under $\hat{F} = F_n$?
- Estimation of the distribution of $\bar{X}$ under F_n by Monte-Carlo simulation
- Calculation of the standard deviation of $\bar{X}$ under F_n
- The distribution of $\bar{X}$ under F_n is an approximation of the distribution of $\bar{X}$ under F
  - Glivenko-Cantelli theorem: The edf $\hat{F}$ converges uniformly to the cdf F.
  - $\theta = t(F)$ and $T = t(\hat{F})$ means that T converges to θ as $n \to \infty$ with t continuous

1. Draw a random sample $x_1^*, \ldots, x_n^*$ from F_n (resampling)
2. Compute
$$\bar{X}^* = \frac{1}{n}\sum_{i=1}^{n} x_i^*$$
3. Repeat steps 1 and 2 a large number R of times, save the results as $\bar{X}_1^*, \ldots, \bar{X}_R^*$
4. Compute the standard error

bootex1.R

$$SE(\bar{X}) = \sqrt{\frac{1}{R-1}\sum_{r=1}^{R}\left(\bar{X}_r^* - \overline{\bar{X}^*}\right)^2}$$

with $\overline{\bar{X}^*} = \frac{1}{R}\sum_{r=1}^{R} \bar{X}_r^*$

- Parametric bootstrap of the bias of

$$T = \hat{\lambda} = \frac{1}{\bar{X}}$$

for the exponential distribution $X \sim Exp(\lambda)$ with cdf

$$F_\lambda(x) = 1 - \exp(-\lambda x)$$

- Simple random sample $X_1, \ldots, X_n$ with n small, e.g. $n = 8$
- Estimation of the unknown distribution function F by

$$F_{\hat{\lambda}}(x) = 1 - \exp\left(-\hat{\lambda} x\right)$$

- How is $\hat{\lambda}$ distributed under $F = F_\lambda$?
- How is $\hat{\lambda}$ distributed under $\hat{F} = F_{\hat{\lambda}}$?
- Estimation of the distribution of $\hat{\lambda}$ under $F_{\hat{\lambda}}$ by Monte-Carlo simulation
- Find the expectation of $\hat{\lambda}$ under $F_{\hat{\lambda}}$
- The distribution of $\hat{\lambda}$ under $F_{\hat{\lambda}}$ approximates the distribution of $\hat{\lambda}$ under F

1. Compute $\hat{\lambda} = 1/\bar{X}$ from original small (e.g. $n = 8$) sample $X_1, \ldots, X_n$
2. Draw a simple random sample $X_1^*, \ldots, X_n^*$ from $F_{\hat{\lambda}}$
3. Compute $\hat{\lambda}^* = 1/\bar{X}^*$
4. Repeat steps 1 and 2 a large number R of times, save the results as $\hat{\lambda}_1^*, \ldots, \hat{\lambda}_R^*$
5. Estimate the bias by

bootex2.R

$$\left(\frac{1}{R}\sum_r \hat{\lambda}_r^*\right) - \hat{\lambda}$$

Bootstrap

General approach for bootstrap standard errors

$$
\begin{array}{c} \text{original} \\ \text{sample} \\ X_1, \dots, X_n \end{array}
\longrightarrow
\begin{array}{c} \text{edf} \\ \hat{F} = F_n \\ \text{or} \\ \hat{F} = F_{\hat{\psi}} \end{array}
\longrightarrow
\left\langle
\begin{array}{l}
1.\ \text{resample: } X_1^*, \dots, X_n^* \to T_1^* \\
2.\ \text{resample: } X_1^*, \dots, X_n^* \to T_2^* \\
\qquad \vdots \\
R.\ \text{resample: } X_1^*, \dots, X_n^* \to T_R^*
\end{array}
\right]
$$

$$
\longrightarrow SE(T) = \sqrt{\frac{1}{R-1} \sum_{r=1}^{R} \left(T_r^* - \overline{T^*} \right)^2}
$$

with $\overline{T^*} = \frac{1}{R} \sum_{r=1}^{R} T_r^*$

- General definition: An interval

$$\left[T_{low}\left(X_1, \ldots, X_n\right);\ T_{high}\left(X_1, \ldots, X_n\right)\right]$$

is called $(1-\alpha)$-confidence interval if

$$Pr\left(T_{low} \leq \theta \leq T_{high}\right) = 1 - \alpha$$

- If the equality holds only asymptotically, the interval is called asymptotic $(1-\alpha)$-confidence interval
- Note: The interval limits are random variables

- The naive confidence intervals are sometimes called the “other” percentile method
- Generate a large number R of resamples and compute $T_1^*, \ldots, T_R^*$
- Let $T_{(1)}^* \leq T_{(2)}^* \leq \ldots \leq T_{(R)}^*$ be the order statistic
- The naive $(1-\alpha)$-confidence interval is

$$\left[T^*_{((\alpha/2)R)};\ \ T^*_{((1-\alpha/2)R)}\right]$$

- Why is this approach often problematic? bootnaiv.R

- To determine confidence intervals we look at the distribution of

$$T - \theta$$

- Let c_1 and c_2 be the $\alpha/2$- and $(1-\alpha/2)$-quantiles, i.e.

$$Pr\left(c_1 \leq T - \theta \leq c_2\right) = 1 - \alpha$$

- Then

$$[T - c_2,\ T - c_1]$$

is the $(1-\alpha)$-confidence interval

- Approximate the distribution of $T-\theta$ by bootstrapping

$$T^* - T$$

- Let c_1^* and c_2^* be the $\alpha/2$- and $(1-\alpha/2)$-quantiles, i.e.

$$Pr\left(c_1^* \leq T^* - T \leq c_2^*\right) = 1-\alpha$$

- We obtain $c_1^* = T^*_{((\alpha/2)R)} - T$ and $c_2^* = T^*_{((1-\alpha/2)R)} - T$ and

$$[T - c_2^*,\ T - c_1^*] = \left[2T - T^*_{((1-\alpha/2)R)};\ 2T - T^*_{((\alpha/2)R)}\right]$$

Algorithm of the percentile method:

- Compute T from the original sample $X_1, \ldots, X_n$
- Generate a large number R of resamples and compute $T_1^*, \ldots, T_R^*$
- Let $T_{(1)}^* \leq T_{(2)}^* \leq \ldots \leq T_{(R)}^*$ be the order statistics
- The bootstrap $(1-\alpha)$-confidence interval is

$$\left[2T - T^*_{((1-\alpha/2)R)};\ 2T - T^*_{((\alpha/2)R)}\right]$$

- Parametric bootstrap 0.95-confidence interval for λ of an exponential distribution
- Simple random sample $X_1, \ldots, X_n$ with n small, e.g. $n = 8$
- Estimate λ by $\hat{\lambda} = 1/\bar{X}$
- Estimate the unknown distribution function F by

$$F_{\hat{\lambda}}(x) = 1 - \exp\left(-\hat{\lambda}x\right)$$

The algorithm bootex3.R

1. Compute $\hat{\lambda} = 1/\bar{X}$ from $X_1, \ldots, X_n$
2. Draw a simple random sample $X_1^*, \ldots, X_n^*$ from $F_{\hat{\lambda}}$
3. Compute $\hat{\lambda}^* = 1/\bar{X}^*$
4. Repeat steps 1 and 2 a large number R of times, save the results as $\hat{\lambda}_1^*, \ldots, \hat{\lambda}_R^*$
5. The bootstrap 0.95-confidence interval is

$$\left[2\hat{\lambda} - \hat{\lambda}^*_{((1-\alpha/2)R)};\ 2\hat{\lambda} - \hat{\lambda}^*_{((\alpha/2)R)}\right]$$

- Test the hypotheses

$$
\begin{aligned}
H_0 &: \quad \theta = \theta_0 \\
H_1 &: \quad \theta \neq \theta_0
\end{aligned}
$$

  at significance level α

- Assumption: Random sample (univariate or multivariate), estimator $\hat{\theta}$
- Test statistic

$$T = \hat{\theta} - \theta_0$$

- Reject H_0 if the value of the test statistic is less than the $\alpha/2$-quantile of T or greater than the $(1-\alpha/2)$-quantile of T
- The p-value of the test is $Pr(|T| > |t|)$
- How can we estimate the distribution of T under H_0 ?

- Wald approach: bootstrap distribution

$$T^* = \hat{\theta}^* - \hat{\theta}$$

- $\hat{\theta}^*$ is calculated from resamples drawn under the alternative hypothesis

- Lagrange multiplier approach: bootstrap distribution

$$T^{\#} = \hat{\theta}^{\#} - \theta_0$$

- $\hat{\theta}^{\#}$ is calculated from resamples drawn under the null hypothesis!
- This approach is particularly suitable for the parametric bootstrap (but can also be used for other bootstraps)

Bootstrap

Hypothesis testing: General algorithm

1. Compute test statistic T from $X_1, \dots, X_n$
2. Draw a resample under the null hypothesis, $X_1^{\#}, \dots, X_n^{\#}$, or draw a resample under the alternative hypothesis, $X_1^*, \dots, X_n^*$
3. Compute the test statistic T^* or $T^{\#}$ for the resample
4. Repeat steps 2 and 3 a large number R of times; save the results as $T_1^{\#}, \dots, T_R^{\#}$ or $T_1^*, \dots, T_R^*$
5. Calculate the $\alpha/2$-quantile $c_1^{\#}$ (or c_1^*) and the $(1-\alpha/2)$-quantile $c_2^{\#}$ (or c_2^*)
6. Reject H_0 if the test statistic T is less than $c_1^{\#}$ (or c_1^*) or greater than $c_2^{\#}$ (or c_2^*)

- Parametric bootstrap for the parameter λ of an exponential distribution $X \sim Exp(\lambda)$
- Random sample $X_1, \dots, X_n$ and maximum likelihood estimator $\hat{\lambda}$
- Hypotheses $H_0 : \lambda = \lambda_0 = 2$ against $H_1 : \lambda \neq \lambda_0$ (at level $\alpha = 0.05$)
- Test statistic
$$T = \hat{\lambda} - 2$$
- Bootstrap of the distribution of T under the alternative hypothesis (Wald approach)

bootex4a.R

- Bootstrap of the distribution of T under the null hypothesis (LM approach) bootex4b.R
- Under the null hypothesis, $X^{\#} \sim Exp(\lambda_0)$ with $\lambda_0 = 2$
- Hence, the distribution of $T^{\#}$ is found by an ordinary Monte-Carlo simulation!
- If $T < T^{\#}_{(\alpha/2B)}$ or $T > T^{\#}_{((1-\alpha/2)B)}$, reject H_0

- Nonparametric test for equality of two expectations
- Two independent variables X and Y with expectations μ_X, μ_Y and unknown variances σ_X^2, σ_Y^2
- Hypotheses $H_0 : \mu_X = \mu_Y$ against $H_1 : \mu_X \neq \mu_Y$
- Samples $X_1, \ldots, X_{n_x}$ and $Y_1, \ldots, Y_{n_y}$
- Test statistic

$$T = \frac{\hat{\mu}_X - \hat{\mu}_Y}{\sqrt{\hat{\sigma}_X^2 + \hat{\sigma}_Y^2}}$$

Bootstrap

Example 5 (contd)

- Case I: resampling under the alternative hypothesis bootex5a.R
- Draw $X_1^*, \ldots, X_m^*$ with replacement from $X_1, \ldots, X_m$ and $Y_1^*, \ldots, Y_n^*$ from $Y_1, \ldots, Y_n$
- Compute the test statistic T^*
- Repeat this R times; calculate the quantile of T^*
- Reject H_0 at level $\alpha = 0.05$ if $T < T^*_{(0.025R)}$ or $T > T^*_{(0.975R)}$

Bootstrap

Example 5 (still contd)

- Case II: resampling under the null hypothesis bootex5b.R
- Estimate the joint expectation by

$$\hat{\mu} = \frac{n_x \hat{\mu}_X + n_y \hat{\mu}_Y}{n_x + n_y}$$

- Transform $X_1, \ldots, X_{n_x}$ such that their mean is $\hat{\mu}$
- Transform $Y_1, \ldots, Y_{n_y}$ such that their mean is $\hat{\mu}$
- Resample from the transformed data (i.e. under the null hypothesis); then continue as before

- Nonparametric bootstrap for independence
- Bivariate distribution (X, Y)
- Hypothesis $H_0 : X$ and Y are stochastically independent
- Sample $(X_1, Y_1), \ldots, (X_n, Y_n)$
- Test statistic: Empirical coefficient of correlation

$$T = \widehat{Corr}(X, Y) = \frac{\sum \left(X_i - \bar{X}\right)\left(Y_i - \bar{Y}\right)}{\sqrt{\sum \left(X_i - \bar{X}\right)^2 \sum \left(Y_i - \bar{Y}\right)^2}}$$

- Resampling under the null hypothesis bootex6.R
- Draw $X_1^{\#}, \ldots, X_n^{\#}$ with replacement from $X_1, \ldots, X_n$
- Independently, draw $Y_1^{\#}, \ldots, Y_n^{\#}$ with replacement from $Y_1, \ldots, Y_n$
- Bootstrap distribution of

$$T^{\#} = \widehat{Corr}(X^{\#}, Y^{\#})$$

- Reject H_0 if $T < T^{\#}_{(0.025R)}$ or $T > T^{\#}_{(0.975R)}$

Parametric bootstrap under the alternative hypothesis

1. Estimate $\hat{\psi}$ from the original data $X_1, \dots, X_n$
2. The estimated distribution function is $\hat{F} = F_{\hat{\psi}}$
3. Draw $X_1^*, \dots, X_n^*$ from $F_{\hat{\psi}}$ and compute $\hat{\psi}^*$
4. Repeat step 3 a large number of times to determine the required distribution

Parametric bootstrap under the null hypothesis

1. The estimated distribution function is $\hat{F} = F_{\psi_0}$

   If the distribution function is not completely specified by ψ_0, choose $\hat{F}$ "as close as possible" to $\hat{\psi}$
2. Draw $X_1^{\#}, \dots, X_n^{\#}$ from F_{ψ_0} and compute $\hat{\psi}^{\#}$
3. Repeat step 2 a large number of times to determine the required distribution

Nonparametric bootstrap under the alternative hypothesis

1. The estimated distribution function is $\hat{F} = F_n$ (empirical distribution function)
2. Draw $X_1^*, \ldots, X_n^*$ with replacement from $X_1, \ldots, X_n$ and compute T^*
3. Repeat step 2 a large number of times to determine the required distribution

Nonparametric bootstrap under the null hypothesis

1. The estimated distribution function $\hat{F}$ is a weighted empirical distribution function
2. Draw $X_1^{\#}, \ldots, X_n^{\#}$ with replacement (but with different probabilities) from $X_1, \ldots, X_n$

   The probabilities are chosen such that $\hat{F}$ satisfies H_0. If not unique, choose an optimality criterion, e.g. maximal entropy
3. Repeat step 2 a large number of times to determine the required distribution

Smooth bootstrap under the alternative hypothesis

- Kernel density estimation (e.g. with Gaussian kernel ϕ and bandwidth h)

$$\hat{f}_X(x) = \frac{1}{nh}\sum_{i=1}^{n} \phi\left(\frac{x - X_i}{h}\right)$$

- Estimated distribution function $\hat{F}(x) = \int_{-\infty}^{x} \hat{f}_X(z)dz$
- Draw $X_1^*, \ldots, X_n^*$ from $\hat{F}(x)$
  1. Draw $Z_1, \ldots, Z_n$ with replacement from $X_1, \ldots, X_n$
  2. Draw $\varepsilon_1, \ldots, \varepsilon_n$ from a standard normal distribution
  3. For $i = 1, \ldots, n$, compute

$$X_i^* = Z_1 + h\varepsilon_i$$

- Smooth bootstrap: nonparametric bootstrap with additional noise

- The bootstrap approximates the distribution of T (or some transformation of interest) *if the model is correctly specified*
- Bias due to misspecification cannot be found by bootstrapping!
- Example: Errors-in-variables, omitted variables
- Bootstrap fails (or has to respecified) for e.g.
  - $T = \underset{1 \leq i \leq n}{max} X_i$ (non-smooth estimators)
  - $T = \bar{X}$ and $E(X^2) = \infty$ (heavy tails)
- The validity of the bootstrap approximation can usually be shown only asymptotically, i.e. for $B \to \infty$ **and** $n \to \infty$
- **However**, experience shows that the bootstrap often yields good approximations of the small-sample distribution of T

- Simple linear regression model

$$y_i = \alpha + \beta x_i + u_i$$

for $i = 1, \ldots, n$ with i.i.d. error terms u_i

- Let $E(u_i | x_i) = 0$ for all $i = 1, \ldots, n$
- OLS estimator of β is

$$\hat{\beta} = \frac{\sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{n} (x_i - \bar{x})^2}$$

- OLS estimator of α is $\hat{\alpha} = \bar{y} - \hat{\beta}\bar{x}$
- Fitted values
$$\hat{y}_i = \hat{\alpha} + \hat{\beta} x_i$$
- Residuals
$$\hat{u}_i = y_i - \hat{y}_i$$
- Estimated error term variance
$$\hat{\sigma}^2 = \frac{1}{n-2} \sum_{i=1}^{n} \hat{u}_i^2$$

- How can we construct a $(1-\alpha)$-confidence interval for β?
- Usual approach: Normal approximation

$$\left[\hat{\beta} - 1.96 \cdot SE(\hat{\beta});\ \hat{\beta} + 1.96 \cdot SE(\hat{\beta})\right]$$

  with standard errors $SE(\hat{\beta}) = \sqrt{\hat{\sigma}^2 / \sum (x_i - \bar{x})^2}$
- Alternative method (1): bootstrap the residuals
- Alternative method (2): bootstrap the observations (x_i, y_i)

Bootstrap the residuals

- The unknown distribution function F is the distribution function of the error terms
- The estimated distribution function $\hat{F}$ is the (parametrically or nonparametrically) estimated distribution function of the residuals $\hat{u}_1, \ldots, \hat{u}_n$
- The x-values are kept constant
- Only the error terms are resampled

Algorithm (nonparametric) bootregr1.R

1. Estimate the model ($\hat{\beta}$) from the data and calculate $\hat{u}_1, \ldots, \hat{u}_n$
2. Draw a resample $u_1^*, \ldots, u_n^*$ with replacement from $\hat{u}_1, \ldots, \hat{u}_n$
3. For $i = 1, \ldots, n$ generate

$$y_i^* = \hat{\alpha} + \hat{\beta} x_i + u_i^*$$

4. Compute $\hat{\beta}^*$ from $(x_1, y_1^*), \ldots, (x_n, y_n^*)$
5. Proceed as usual

Bootstrap of the observations

- The unknown distribution function F is the joint distribution function of (x_i, y_i)
- The estimated distribution function $\hat{F}$ is the (usually nonparametrically) estimated multivariate distribution function of the observations $(x_1, y_1), \dots, (x_n, y_n)$
- The x-values and corresponding y-values are different in each resample

Algorithm bootregr2.R

1. Estimate $\hat{\beta}$ from the data
2. Draw a resample $(x_1^*, y_1^*), \dots, (x_n^*, y_n^*)$ with replacement from $(x_1, y_1), \dots, (x_n, y_n)$
3. Compute $\hat{\beta}^*$ from $(x_1^*, y_1^*), \dots, (x_n^*, y_n^*)$
4. Proceed as usual