

Contents

8	Bootstrap Methods	2
8.1	Bootstrap Estimate of MSE, Variance, Bias	3
8.2	Bootstrap Confidence Intervals	7
8.3	Bootstrap for Regression	21
8.4	Hypothesis Testing with Bootstrap	24

8 Bootstrap Methods

- Permutation: compare an observed test statistic value with the permutation distribution of the test statistic obtained by permuting the “membership” of the data (satisfying H_0) in some appropriate way.
- Bootstrap: use the observed sample to approximate the distribution from which the sample is generated from. E.g. we obtain a sample of size 20. Randomly drawing 20 observations with replacement from the observed data gives a bootstrap sample of size 20.
- The usage of bootstrap:
 - construct confidence intervals;
 - hypothesis testing.

Paradigm of Bootstrap:

<p>Unknown dist. $F(\cdot)$ generates sample $\{x_1, \dots, x_n\}$</p> <p>Test statistic $\hat{\theta} = t(x_1, \dots, x_n)$ based on $\{x_i\}$</p> <p>one $\hat{\theta}$</p>	<p>Empirical dist. $\hat{F}_n(\cdot)$ generates x_i^* by resampling w/ replacement</p> <p>bootstrap estimate $\hat{\theta}^* = t(x_1^*, \dots, x_n^*)$ based on $\{x_i^*\}$</p> <p>repeat and get many B $\hat{\theta}^*$'s</p>
--	---

- Efron's idea: assess the variability of $\hat{\theta}$ about the unknown parameter θ by the variability of $\hat{\theta}^*$ about $\hat{\theta}$.
- $Bias(\hat{\theta}) = E(\hat{\theta} - \theta) \approx E(\hat{\theta}^*) - \hat{\theta} \approx \frac{1}{B} \sum_{j=1}^B \hat{\theta}_j^* - \hat{\theta}$.

8.1 Bootstrap Estimate of MSE, Variance, Bias

Recall for an estimator $\hat{\theta}$ of θ ,

$$MSE(\hat{\theta}) = E(\hat{\theta} - \theta)^2 = \{Bias(\hat{\theta})\}^2 + Var(\hat{\theta}),$$

where $Bias(\hat{\theta}) = E(\hat{\theta}) - \theta$.

Bootstrap procedure

1. Calculate $\hat{\theta}$ based on the observed sample x_1, \dots, x_n .
2. Draw a bootstrap sample of size n by randomly drawing one from the observed sample with replacement for n times. Note that for each bootstrap, some observations may be drawn multiple times, while some are not drawn at all. In R: use line
`boot.sample = sample(x, replace=TRUE)`
3. Calculate the statistic $\hat{\theta}^*$ based on the bootstrap sample
4. Repeat Steps 2-3 B times, resulting in B bootstrap statistics

$$\hat{\theta}_1^*, \dots, \hat{\theta}_B^*.$$

5. Calculate bootstrap MSE, Bias and Variance as follows

$$(a) \widehat{MSE}(\hat{\theta}) = \frac{1}{B} \sum_{b=1}^B (\hat{\theta}_b^* - \hat{\theta})^2$$

$$(b) \widehat{Bias}(\hat{\theta}) = \frac{1}{B} \sum_{b=1}^B \hat{\theta}_b^* - \hat{\theta}$$

$$(c) \widehat{Var}(\hat{\theta}) = \frac{1}{B} \sum_{b=1}^B (\hat{\theta}_b^* - \bar{\hat{\theta}}^*)^2, \text{ where } \bar{\hat{\theta}}^* = 1/B \sum_{b=1}^B \hat{\theta}_b^*.$$

6. Choice of B : larger B leads to more accuracy. Booth and Sarkar (1998): $B \geq 800$ for estimating variance.

Example 8.1.1 *Testing of electrical and mechanical devices often involves an action such as turning a device on and off or opening and closing a device many times. The interest is in the distribution of the number of on-off or open-close cycles that occur before the device fails. The following data set consists of the number of cycles (in thousands) that it takes for 200 door latches to fail:*

7	11	15	16	20	22	24	25	29	33	34	37	41	42	49	57	66	71	84	90
---	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----	----

Suppose we are interested in estimating the median of the latch failure time, and define $\hat{\theta}$ = sample median of the 20 data points. Use bootstrap to estimate the bias, variance and MSE of $\hat{\theta}$.

See the R code.

8.2 Bootstrap Confidence Intervals

Approach 1: normality-based bootstrap CI

Requirements:

1. $\hat{\theta}$ is approximately normal
2. $\hat{\theta}$ is unbiased
3. bootstrap sampling gives a good estimate of $Var(\hat{\theta})$.

Bootstrap $100(1 - \alpha)\%$ CI for θ :

$$\hat{\theta} - z_{\alpha/2} \sqrt{\widehat{Var}(\hat{\theta})} \leq \theta \leq \hat{\theta} + z_{\alpha/2} \sqrt{\widehat{Var}(\hat{\theta})},$$

where $\widehat{Var}(\hat{\theta})$ is the bootstrap estimator of $Var(\hat{\theta})$.

Approach 2: normality-based bootstrap CI with bias adjustment

Recall $Bias(\hat{\theta}) = E(\hat{\theta}) - \theta$ can be estimated by the bootstrap Bias,

$$\widehat{Bias}(\hat{\theta}) = 1/B \sum_{b=1}^B \hat{\theta}_b^* - \hat{\theta}.$$

Therefore, we can obtain the bias-adjusted estimator

$$\hat{\theta}_{adj} = \hat{\theta} - \widehat{Bias}(\hat{\theta}) = 2\hat{\theta} - \frac{1}{B} \sum_{b=1}^B \hat{\theta}_b^*$$

The bias-adjusted bootstrap CI (assuming normality):

$$\hat{\theta}_{adj} - z_{\alpha/2} \sqrt{\widehat{Var}(\hat{\theta})} \leq \theta \leq \hat{\theta}_{adj} + z_{\alpha/2} \sqrt{\widehat{Var}(\hat{\theta})}.$$

Approach 3: percentile bootstrap CI

1. Draw a bootstrap sample of size n by resampling with replacement, and calculate the test statistic $\hat{\theta}_b^*$ based on the bootstrap sample.
2. Repeat the first step B times, resulting in $\hat{\theta}_b^*, b = 1, \dots, B$.
3. Calculate $100(1 - \alpha)\%$ confidence interval of θ :

$$(\hat{\theta}_{\alpha/2}^*, \hat{\theta}_{1-\alpha/2}^*),$$

where $\hat{\theta}_{\alpha}^*$ is the $100 \times \alpha$ th percentile of $\hat{\theta}_b^*, b = 1, \dots, B$.

Approach 4: residual method bootstrap CI

Residual method is based on obtaining a bootstrap sample of the residuals $\epsilon = \hat{\theta} - \theta$. Note that

$$\begin{aligned} P(\epsilon_{0.025} \leq \hat{\theta} - \theta \leq \epsilon_{0.975}) &= 0.95 \\ \Leftrightarrow P(\hat{\theta} - \epsilon_{0.975} \leq \theta \leq \hat{\theta} - \epsilon_{0.025}) &= 0.95, \end{aligned}$$

and the distribution of $e_b = \hat{\theta}_b^* - \hat{\theta}$ mimics that of $\epsilon = \hat{\theta} - \theta$.

1. Compute $\hat{\theta}$ based on the observed sample.
2. Draw a bootstrap sample of size n from the data, and calculate $\hat{\theta}_b^*$ and the residual $e_b = \hat{\theta}_b^* - \hat{\theta}$.
3. Repeat the first step B times, resulting in $\hat{\theta}_b^*, e_b, b = 1, \dots, B$.
4. Calculate the $\alpha/2$ and $1 - \alpha/2$ quantiles of e_b , denoted by $e_{b,\alpha/2}$ and $e_{b,1-\alpha/2}$.

5. Calculate $100(1 - \alpha)\%$ confidence interval of θ :

$$\begin{aligned} & (\hat{\theta} - e_{b,1-\alpha/2}, \hat{\theta} - e_{b,\alpha/2}) \\ &= (2\hat{\theta} - \hat{\theta}_{b,1-\alpha/2}^*, 2\hat{\theta} - \hat{\theta}_{b,\alpha/2}^*), \end{aligned}$$

where $\hat{\theta}_{b,\alpha}^*$ is the α th quantile of $\{\hat{\theta}_b^*, b = 1, \dots, B\}$.

Remark:

- The CIs by the percentile bootstrap method might not be centered on the true parameter. Thus the CIs might have bias, and might be too narrow (or wide), especially when the distribution of $\hat{\theta}^*$ is not symmetric around $\hat{\theta}$.

Approach 5: BCA (bias corrected and accelerated) method

The idea of BCA

- Assume that there is a transformation of $\hat{\theta}$, $T(\hat{\theta})$, whose distribution is

$$N(g\{T(\theta)\}, \sigma^2\{T(\theta)\}),$$

where $g\{T(\theta)\} = T(\theta) - z_0\{1 + aT(\theta)\}$, and $\sigma\{T(\theta)\} = 1 + aT(\theta)$.

- Similar to the percentile method, endpoints of the BCA CI are percentiles of $\hat{\theta}_1^*, \dots, \hat{\theta}_B^*$.
- However, the percentiles are not necessarily symmetric, e.g. for 95% CI,
 - the end points are not necessarily the 2.5th and 97.5th percentiles;
 - the fraction of the bootstrap distribution between the two BCA end points is not necessarily 95%.

- The percentiles depend on
 - z_0 (bias correction): corrects for the median bias of the bootstrap estimator;
 - a (acceleration): a measurement of skewness of the data.
- BCA $100(1 - \alpha)\%$ CI of θ :

$$(\hat{\theta}_{\alpha_L}^*, \hat{\theta}_{\alpha_U}^*),$$

where

$$\begin{aligned}\alpha_U &= \Phi \left(\frac{z_0 + z_{\alpha/2}}{1 - a(z_0 + z_{\alpha/2})} + z_0 \right) \\ \alpha_L &= \Phi \left(\frac{z_0 - z_{\alpha/2}}{1 - a(z_0 - z_{\alpha/2})} + z_0 \right).\end{aligned}$$

- Estimation of z_0 :

$$\hat{z}_0 = \Phi^{-1} \left\{ \frac{1}{B} \sum_{b=1}^B I(\hat{\theta}_b^* < \hat{\theta}) \right\},$$

where $\Phi(\cdot)$ is the CDF of $N(0, 1)$.

- Estimation of a :

$$\hat{a} = \frac{\sum_{i=1}^n \left(\hat{\theta}_{()} - \hat{\theta}_{-i} \right)^3}{6 \left\{ \sum_{i=1}^n \left(\hat{\theta}_{()} - \hat{\theta}_{-i} \right)^2 \right\}^{3/2}},$$

where $\hat{\theta}_{-i}$ is the statistic based on the sample after leaving out the i th observation, and $\hat{\theta}_{()} = \frac{1}{n} \sum_{i=1}^n \hat{\theta}_{-i}$.

- Suppose the bootstrap distribution of $\hat{\theta}_b^*$ is centered at $\hat{\theta}$, specifically, with median $\hat{\theta}$, then $\hat{z}_0 = 0$.
- If 25% of $\hat{\theta}_b^*$'s are $< \hat{\theta}$, then $\hat{z}_0 = -0.675$.

- If the distribution of $\hat{\theta}_b^*$ is symmetric, $a = 0$.
- Setting $a = 0$ leads to a simpler version of the BCA interval, called the Bias Corrected (BC) interval.

Advantages of BIC intervals

- Transformation equivariant. For example, if a 95% BCA CI for $\hat{\theta}$ is (L, U) , then a 95% BCA CI for $g(\theta)$ is $(g(L), g(U))$ for any increasing function $g(\cdot)$.
- Second-order accurate $P(\hat{\theta}_{\alpha_1}^* < \theta < \hat{\theta}_{\alpha/2}^*) = \alpha + o(n^{-1})$, that is, errors go to zero at the rate of $1/n$.

Approach 6: pivotal approach

The bootstrapping may often be improved by using transformations related to pivotal quantities: quantities whose distribution do not depend on the unknown parameter of interest.

For instance, consider the CI of the mean μ .

- Suppose X_1, \dots, X_n are *i.i.d.* from $N(\mu, \sigma^2)$. Then

$$T = \frac{\bar{X} - \mu}{S/\sqrt{n}} \sim t_{n-1}$$

and thus is a pivotal quantity.

- When the distribution assumption is true,

$$P \left(t_{0.025} \leq \frac{\bar{X} - \mu}{S/\sqrt{n}} \leq t_{0.975} \right) = 0.95,$$

and the 95% CI of μ can be constructed as

$$\bar{X} - t_{0.975} \frac{S}{\sqrt{n}} < \mu < \bar{X} - t_{0.025} \frac{S}{\sqrt{n}},$$

where t_α is the α th quantile of t_{n-1} distribution. Recall $t_\alpha = -t_{1-\alpha}$ since t_{n-1} is symmetric around 0.

- When the normal distribution is violated, $T = \frac{\bar{X} - \mu}{S/\sqrt{n}}$ no longer follows t_{n-1} distribution, but $t_{0.025}$ and $t_{0.975}$ can be obtained by the percentiles of the distribution of the bootstrap version of the pivotal quantity.

Steps for the bootstrap-t method based on the t -pivotal quantity:

1. Calculate \bar{x} and s based on the observed data.
2. Draw a bootstrap sample of size n from the observed data with replacement.

3. Based on the bootstrap sample x_1^*, \dots, x_n^* , calculate the t -statistic

$$t_b^* = \frac{\bar{x}_b^* - \bar{x}}{s_b^* / \sqrt{n}},$$

where $\bar{x}_b^* = 1/n \sum_{i=1}^n x_i^*$ and s_b^* are the sample means and sample variances based on the b th bootstrap sample.

4. Repeat steps 2-3 B times resulting in t_b^* , $b = 1, \dots, B$.
5. Calculate $100(1 - \alpha)\%$ CI of μ as

$$\left(\bar{x} - t_{1-\alpha/2}^* \frac{s}{\sqrt{n}}, \bar{x} - t_{\alpha/2}^* \frac{s}{\sqrt{n}} \right),$$

where t_{α}^* is the α th quantile of $\{t_b^*, b = 1, \dots, B\}$.

Comparison of different approaches

	Transformation equivariance	2nd order accuracy
1. Normality-based	No	No
2. Percentile	Yes	No
3. BCA	Yes	Yes
4. Bootstrap-t	No	Yes

Example 8.2.1 Recall the latch failure data 8.1.1. Construct 95% CI of the mean using different approaches.

```
mymean.func = function(x, d)
{
  #d is the index of resampled observations
  mean(x[d])
}

### bootstrap median estimates based on R bootstrap samples
mean.boot = boot(x, mymean.func, R=1000)
boot.ci(mean.boot)
```

8.3 Bootstrap for Regression

Suppose we observed data $(y_i, x_i), i = 1, \dots, n$, where $x_i = (x_{i1}, \dots, x_{ip})$ includes p predictors. Consider the multiple linear regression model:

$$y_i = \beta_0 + \beta_1 x_{i1} + \dots + \beta_p x_{ip} + e_i.$$

We might be interested in testing

$$H_0 : \beta_j = 0$$

or

$$H_0 : \beta_1 = \beta_2 = \dots = \beta_p = 0$$

or constructing CI for β_j .

- **Approach 1: paired bootstrap.** Consider each subject as a unit, and bootstrap the subjects, that is, the paired data $(y_1, x_1), \dots, (y_n, x_n)$.

- **Approach 2: residual bootstrap.**

- Fit the regression to the observed data and obtain the residuals

$$e_i = y_i - (\hat{\beta}_0 + \hat{\beta}_1 x_{i1} + \cdots + \hat{\beta}_p x_{ip}).$$

- Resample $\{e_i, i = 1, \dots, n\}$ with replacement to obtain a bootstrap sample $\{e_i^*, i = 1, \dots, n\}$. Then obtain the bootstrap data

$$y_i^* = \hat{\beta}_0 + \hat{\beta}_1 x_{i1} + \cdots + \hat{\beta}_p x_{ip} + e_i^*.$$

- Calculate the coefficient estimator $\hat{\beta}_j^*, j = 0, 1, \dots, p$ based on the bootstrap sample (y_i^*, x_i) .
- Repeat the above steps 2-3 B times, and obtain the bootstrap coefficient estimates $\hat{\beta}_{jb}^*, b = 1, \dots, B$.

- Comparison

- Paired bootstrap is more robust than the residual bootstrap against the model misspecification.

- Paired bootstrap only requires the sample $(y_i, x_i)_{i=1}^n$ are randomly draws from the distribution.
- Residual bootstrap requires the validity of the multiple linear model:

$$y_i = \beta_0 + \beta_1 x_{i1} + \cdots + \beta_p x_{ip} + e_i,$$

where e_i are *i.i.d.*. By treating covariates as fixed, the bootstrap standard error reflects the variation associated with the observed covariates, and note that the standard errors of the regression coefficient estimates are affected by the variation of observed covariates.

8.4 Hypothesis Testing with Bootstrap

Suppose we are interested in testing $H_0 : \theta = \theta_0$.

- Permutation testing method: compare the observed test statistic value $\hat{\theta}$ with the permutation distribution obtained under the null hypothesis. If the sample is drawn from a distribution where H_0 is not true, then permutation/bootstrap test can be tricky.
- Two guidelines for bootstrap hypothesis testing:
 - Resample $\hat{\theta}^* - \hat{\theta}$ instead of $\hat{\theta} - \theta$
 \Rightarrow the resampling has to be done reflecting the null distribution.
 - Base the test on the distribution of $(\hat{\theta}^* - \hat{\theta})/\hat{\sigma}^*$ instead of $(\hat{\theta}^* - \hat{\theta})/\hat{\sigma}$ or $(\hat{\theta}^* - \hat{\theta})$.
 \Rightarrow whenever possible use a pivotal quantity in the test procedure.

- Bootstrap test: as in the permutation test, compare the observed test statistic $\hat{\theta}$ with the bootstrap distribution of $\hat{\theta}$ under the null hypothesis.
- Note that the bootstrap distribution of $\hat{\theta}_b^* - \hat{\theta}$ reflects the distribution of $\hat{\theta} - \theta$. Therefore, we should compare the test statistic $\hat{\theta} - \theta_0$ with the distribution of $\hat{\theta}_b^* - \hat{\theta}$.

For instance, let X_1, \dots, X_n be a random sample from a distribution with mean μ . We want to test $H_0 : \mu = \mu_0$ versus $H_0 : \mu > \mu_0$.

- Idea 1: let the test statistic be $t_1 = \bar{X} - \mu_0$, and compare t_1 to the distribution of

$$t_{1b}^* = \bar{X}_b^* - \mu_0, b = 1, \dots, B.$$

May be lack of power when H_0 is incorrect. Recall the distribution of \bar{X}_b^* is centered at \bar{X} , not μ_0 . Thus t_{1b}^* is not centered at 0, and the distribution of t_{1b}^* does not reflect the

distribution of the test statistic t_1 under the null hypothesis.

- Idea 2: Compare $t_2 = \bar{X} - \mu_0$ with the distribution of

$$t_{2b}^* = \bar{X}_b^* - \bar{X}, \quad b = 1, \dots, B.$$

Idea 2 is good but not ideal, as t_{2b}^* is dependent on the variation in the observed sample, not pivotal.

- Idea 3: based on the pivotal quantity

$$t_3 = \frac{\bar{X} - \mu_0}{S/\sqrt{n}}.$$

Compare t_3 with the distribution of

$$t_{3b} = \frac{\bar{X}_b^* - \bar{X}}{S_b^*/\sqrt{n}}.$$

The idea 3 often controls the Type I error closer to the nominal significance level better than idea 2.

Example 8.4.1 Recall the latch failure data 8.1.1. Test $H_0 : \mu = 30$ versus $H_a : \mu > 30$ using bootstrap.

Example 8.4.2 *grades. X_i : middle term exam score, Y_i : final exam score, $i = 1, \dots, 21$. Obtain bootstrap standard error of the linear regression slope coefficient estimator $\hat{\beta}_1$, and test $H_0 : \beta_1 = 0.5$.*