Computer Intensive Statistics and Applications
Chapter 3: Bootstrap

Shaobo Jin

Department of Mathematics

Bootstrap Principle

The main principle of the bootstrap is that our data can be used to approximate the population. A large number of new observations drawn from this approximated population can be used to estimate distribution properties.

1. $\boldsymbol{X} = (X_1, ..., X_n) \sim \boldsymbol{P}$ (not necessarily iid sample) with distribution function $\boldsymbol{F}$ and the statistical model $\boldsymbol{P} \in \{\boldsymbol{P}_\theta; \theta \in \Theta\}$. X_i can be unidimensional or multidimensional.
2. We are interested in the parameter $\theta = \theta(\boldsymbol{F})$, the distribution properties of a given estimator $\hat{\theta} = \hat{\theta}(\boldsymbol{X})$, and the distribution properties of a given function $T(\boldsymbol{X}, \boldsymbol{F})$.
3. Our data is $\boldsymbol{x} = (x_1, ..., x_n)$ from $\boldsymbol{X} = (X_1, ..., X_n)$.
4. Estimate $\boldsymbol{F}$ or $\boldsymbol{P}$ by some estimator $\hat{\boldsymbol{F}}_n$ or $\boldsymbol{P}^*$. $\boldsymbol{P}^*$ is called the bootstrap distribution.

Bootstrap Principle

5 Draw independently B random samples $\boldsymbol{x}^*_{(j)}$ of size n from $\boldsymbol{P}^*$, $j = 1, ..., B$, that is $\boldsymbol{X}^*_{(j)} \sim \boldsymbol{P}^*$. The samples $\boldsymbol{x}^*_{(j)}$ are called the bootstrap samples.

6 For each j, estimate θ using the bootstrap sample $\boldsymbol{x}^*_{(j)}$ as $\hat{\theta}^*_j = \hat{\theta}\left(\boldsymbol{x}^*_{(j)}\right)$. The bootstrap replications of $\hat{\theta}$ are $\hat{\theta}^*_1$, ..., $\hat{\theta}^*_B$. The bootstrap replications of the function $T(\boldsymbol{X}, \boldsymbol{F})$ are $T^*_{(j)} = T\left(\boldsymbol{X}^*_{(j)}, \hat{\boldsymbol{F}}_n\right)$, $j = 1, ..., B$.

7 Pretend that the bootstrap replications are an iid sample from the distribution of $\hat{\theta}$ or $T(\boldsymbol{X}, \boldsymbol{F})$.

The main idea is that when considering a function $Q\left(\hat{\theta}, \theta\right)$ using the observed data, the bootstrap treats $\hat{\theta}$ as the truth and $\hat{\theta}^*_j$ as the estimator, i.e., $Q\left(\hat{\theta}^*_j, \hat{\theta}\right)$.

Bootstrap: An Example

Suppose that we have observed iid data x_1, ..., x_n from a statistical model $X \sim P$ with distribution function F. Our focus is the median of this distribution, denoted by $m(F)$.

- In classical inference, we can use the sample median $m_{(n)}$ to estimate the population median $m(F)$.
- The "bias" can be computed by $T(\boldsymbol{X}, \boldsymbol{F}) = m_{(n)} - m(F)$.
- However, we do not know the true value $m(F)$.

In bootstrap, the first steps are the same as classical inference:

2. The parameter is $m(F)$. We are interested in $T = m_{(n)} - m(F)$.
3. Our data is $\boldsymbol{x} = (x_1, ..., x_n)$ from $\boldsymbol{X} = (X_1, ..., X_n)$.

Bootstrap: An Example

4 If we choose the nonparametric approach, we estimate $\boldsymbol{F}$ by some estimator $\hat{\boldsymbol{F}}_n$ using $\boldsymbol{x}$ as

$$\hat{F}_n(x) = \frac{1}{n}\sum_{i=1}^{n} 1(x_i \leq x).$$

The estimated distribution P^* is $\mathrm{P}(X = x_i) = n^{-1}$ for any i. The estimator of $m(F)$ is $m_{(n)} = m\left(\hat{F}_n\right)$.

Bootstrap: An Example

From step 5, we need Monte Carlo simulation.

5 Given step 4 in our example, draw independently B bootstrap samples $\boldsymbol{x}^*_{(j)} = \left(x^*_{1j}, ..., x^*_{nj}\right)$ of size n from P^*, $j = 1, ..., B$.

6 For each j, estimate $m\left(\hat{F}_n\right)$ using the bootstrap sample $\boldsymbol{x}^*_{(j)}$ as $m_{(n),j}$, where $m_{(n),j}$ is the sample median of $x^*_{1j}, ..., x^*_{nj}$. Compute the difference

$$T_j = m_{(n),j} - m_{(n)},$$

since $m_{(n)}$ is the median of the distribution P^*.

7 View $m_{(n),j} - m_{(n)}$, $j = 1, ..., B$, as an iid sample from the distribution of $m_{(n)} - m\left(F\right)$. For example, the bias of the estimator $m_{(n)}$ can be approximated by

$$\frac{1}{B}\sum_{j=1}^{B} m_{(n),j} - m_{(n)}.$$

Parametric and Nonparametric Bootstrap

4 Estimate $\boldsymbol{F}$ by some estimator $\hat{\boldsymbol{F}}_n$ (e.g., parametric or nonparametric). Denote the estimated distribution by $\boldsymbol{P}^*$. $\boldsymbol{P}^*$ is called the bootstrap distribution.

5 Draw independently B random samples $\boldsymbol{x}^*_{(j)}$ of size n from $\boldsymbol{P}^*$, $j = 1, ..., B$, that is $\boldsymbol{X}^*_{(j)} \sim \boldsymbol{P}^*$. The samples $\boldsymbol{x}^*_{(j)}$ are called the bootstrap samples.

Parametric and nonparametric approaches can make some difference.

1. The parametric bootstrap assumes a parametric form of the statistical model $\boldsymbol{P}$ index by some parameter θ.
   - For example, we assume $X \sim N\left(\mu, \sigma^2\right)$. The bootstrap samples are drawn from $N\left(\hat{\mu}, \hat{\sigma}^2\right)$.
2. The nonparametric bootstrap estimates $\boldsymbol{F}$ by the empirical distribution function.
   - The estimated distribution is $\mathrm{P}\left(X = x_i\right) = n^{-1}$ for any i.

Bootstrap: Median Example Again

If we want to use parametric bootstrap, we need to assume some parametric form.

4 If we assume our data are iid draws from an exponential distribution with mean λ, the maximum likelihood estimator of λ is $\hat{\lambda} = \bar{X}$. The estimated distribution P^* is an exponential distribution with mean $\bar{X}$, denoted by $\text{Exp}\left(\bar{X}\right)$. Our estimator of the median is $\bar{X}\log 2$.

5 Given step 4 in our example, draw independently B bootstrap samples $\boldsymbol{x}^*_{(j)} = \left(x^*_{1j}, ..., x^*_{nj}\right)$ of size n from $\text{Exp}\left(\bar{X}\right)$, $j = 1, ..., B$.

Empirical Distribution Function

For notation simplicity, we consider $X \in \mathbb{R}$.

Definition

Let x_1, ..., x_n be a sequence of measurements, then the fraction

$$\hat{F}_n(x) = \frac{1}{n}\sum_{i=1}^{n} 1(x_i \leq x)$$

is called the empirical cumulative distribution function (ecdf) or the empirical distribution function.

Example

Find the ecdf of the data 1, 1, 2, 3.

Plug-In Principle in Nonparametric Bootstrap

Real Word	Bootstrap World
cdf F	ecdf $\hat{F}_n$
$X \sim F$	$X^* \sim \hat{F}_n$
$\theta(F)$	$\theta\left(\hat{F}_n\right)$
$\hat{\theta}(\boldsymbol{X})$	$\hat{\theta}^* = \hat{\theta}(\boldsymbol{X}^*)$
$T(\boldsymbol{X}, F)$	$T\left(\boldsymbol{X}^*, \hat{F}_n\right)$

For example, if $\theta(F) = \mathrm{E}(X^r)$, then

$$\theta\left(\hat{F}_n\right) = \int x^r d\hat{F}_n = n^{-1}\sum_{i=1}^{n} x_i^r,$$

the sample moments.

Why Does Bootstrap Work? An Example

For simplicity, suppose that X_1, ..., X_n is an iid sample from the distribution P, that the distribution function F is continuous, and that $\sigma^2 = \mathrm{Var}\left(X\right) < \infty$ is known. Our parameter of interest is $\mu = \mathrm{E}\left[X\right]$. We estimate it by $\mu\left(\hat{F}_n\right) = \bar{x}$.

- By the central limit theorem,

$$Z = \frac{\sqrt{n}\left(\bar{X} - \mu\right)}{\sigma} \quad \stackrel{d}{\to} \quad N\left(0, 1\right).$$

- In the bootstrap world, given $\hat{F}_n$,

$$Z^* = \frac{\sqrt{n}\left(\bar{X}^* - \bar{x}\right)}{\sigma} \quad \stackrel{d}{\to} \quad N\left(0, 1\right).$$

- Hence,

$$\mathrm{P}\left(Z < z\right) - \mathrm{P}\left(Z^* < z \mid \hat{F}_n\right) \quad = \quad o_P\left(1\right).$$

More Generally: Consistency

Bootstrap aims to approximate the exact distribution of $T(\boldsymbol{X}, \boldsymbol{F})$ using the distribution of $T\left(\boldsymbol{X}^*, \hat{\boldsymbol{F}}_n\right)$.

- For example, we are interested in the distribution of $T(\boldsymbol{X}, \boldsymbol{F}) = \left(\hat{\theta} - \theta\right)/\hat{\sigma}$.
- The bootstrap counterpart is $\left(\hat{\theta}^* - \hat{\theta}\right)/\hat{\sigma}^*$.
- We basically use the distribution of $\left(\hat{\theta}^* - \hat{\theta}\right)/\hat{\sigma}^*$ given $\hat{\boldsymbol{F}}_n$ to approximate the distribution of $\left(\hat{\theta} - \theta\right)/\hat{\sigma}$.

The bootstrap is consistent if

$$\sup_t \left| \mathrm{P}\left(T(\boldsymbol{X}, \boldsymbol{F}) \le t\right) - \mathrm{P}\left(T\left(\boldsymbol{X}^*, \hat{\boldsymbol{F}}_n\right) \le t \mid \hat{\boldsymbol{F}}_n\right) \right| \xrightarrow{P} 0,$$

where the second probability is still a random variable.

Why Does Nonparametric Bootstrap Work?

By the Glivenko-Cantelli theorem, we know that

$$P\left(\sup_x \left|\hat{F}_n(x) - F(x)\right| > \epsilon\right) \rightarrow 0.$$

Hence for sufficiently large n, sampling from F is similar to sampling from $\hat{F}_n$.

However, we still need more assumptions in order for bootstrap to work. Suppose that $T(\boldsymbol{X}, \boldsymbol{F}) \xrightarrow{d} T_A$ with distribution function $F_A(t)$. We need to assume

1. For any sequence of distributions $\{G_n\}$ such that $\lim_{n\to\infty} G_n(t) = F(t)$, we have $\lim_{n\to\infty} P\left(T\left(\tilde{\boldsymbol{X}}, G_n\right) \leq t\right) = F_A(t)$.
2. $F_A(t)$ is a continuous function.

Why Does Nonparametric Bootstrap Work?

To simplify the assumptions a little, the bootstrap is consistent if, for every x,

$$\begin{aligned} \mathrm{P}\left(T\left(\boldsymbol{X}, \boldsymbol{F}\right) \leq x\right) &\rightarrow F_A\left(x\right), \\ \mathrm{P}\left(T\left(\boldsymbol{X}^*, \hat{\boldsymbol{F}}_n\right) \leq x \mid \hat{\boldsymbol{F}}_n\right) &\xrightarrow{P} F_A\left(x\right), \end{aligned}$$

and F_A is a continuous function.

Example

Consider again bootstrapping the mean. The central limit theorem implies that the distribution functions of

$$Z = \frac{\sqrt{n}\left(\bar{X} - \mu\right)}{\sigma} \quad \text{and} \quad Z^* = \frac{\sqrt{n}\left(\bar{X}^* - \bar{X}\right)}{\hat{\sigma}}$$

converge to $\Phi\left(\right)$, the distribution function of $N\left(0, 1\right)$.

Bootstrap Can Fail

Consider an iid sample X_1, ..., X_n from Uniform $[0, \theta]$. The maximum likelihood estimator of θ is

$$\hat{\theta} = \max_i X_i = X_{[n]}.$$

- We can show that $\sqrt{n}\left(X_{[n]} - \theta\right)$ does not converge to a normal distribution. In fact,

$$Q = \frac{n\left(\theta - X_{[n]}\right)}{\theta} \xrightarrow{d} \text{Exp}\left(1\right).$$

- If nonparametric bootstrap works, the distribution of $Q^* = n\left(x_{[n]} - X^*_{[n]}\right)/x_{[n]}$ should converge to distribution of Q.
- But

$$P\left(Q^* = 0\right) \rightarrow 1 - e^{-1}.$$

How Do We Know Bootstrap Works in Practice?

We can

- show that the conditions in the previous slides hold;
- or, run some simulations to get some ideas about whether it works.

In case, bootstrap does not work, we can try to modify the bootstrap algorithm.

Example

In the uniform distribution example, we taking a bootstrap sample of m instead of n, such as $m = n^{1/3}$ and $m = n^{1/2}$ and check whether it works.

Parametric Bootstrap

The main difference between parametric bootstrap and nonparametric bootstrap is that we simulate X^* from $F\left(x;\hat{\theta}\right)$, instead of $\hat{F}_n$, where $F\left(x;\theta\right)$ is the distribution function with parameter θ.

In order for the parametric bootstrap to work, we can establish similar conditions as in the nonparametric bootstrap, but with $F\left(x;\hat{\theta}\right)$, instead of $\hat{F}_n$.

An Extreme Example

One important property of bootstrap is that it can often achieve more accurate approximations than the usual asymptotic approaches.

Example

Consider again the uniform distribution but we apply parametric bootstrap. It is easy to show that

$$\begin{aligned} \mathrm{P}\left(\frac{n\left(x_{[n]}-X_{[n]}^{*}\right)}{x_{[n]}}>t \mid X_{[n]}\right) &= \left(1-\frac{t}{n}\right)^{n}, \\ \mathrm{P}\left(\frac{n\left(\theta-X_{[n]}\right)}{\theta}>t\right) &= \left(1-\frac{t}{n}\right)^{n}. \end{aligned}$$

But the asymptotic method relies on

$$\mathrm{P}\left(\frac{n\left(\theta-X_{[n]}\right)}{\theta}>t\right) \rightarrow e^{-t}.$$

Error of Approximation

In general, we are interested in the exact distribution of a random quantity $T(\boldsymbol{X}, \boldsymbol{F})$.

- Its exact distribution is difficult to derive, so we use its asymptotic distribution to approximate the exact distribution.

It is often the case that $T(\boldsymbol{X}, \boldsymbol{F}) = Z + O_P\left(n^{-1/2}\right)$, where $Z \sim N(0,1)$. Hence the error of approximation is $O_P\left(n^{-1/2}\right)$.

- For example, $\sqrt{n}\left(\hat{\theta} - \theta\right)$ often converges in distribution to a normal distribution.

The Quantity $T(\boldsymbol{X}, \boldsymbol{F})$

The choice of $T(\boldsymbol{X}, \boldsymbol{F})$ can have effects on the accuracy of approximation.

Example

Consider again the uniform distribution but we apply parametric bootstrap and consider $T = n\left(\theta - X_{[n]}\right)$. It is easy to show that

$$\begin{aligned} \mathrm{P}\left(n\left(x_{[n]} - X_{[n]}^{*}\right) > t \mid X_{[n]}\right) &= \left(1 - \frac{t}{n x_{[n]}}\right)^{n}, \\ \mathrm{P}\left(n\left(\theta - X_{[n]}\right) > t\right) &= \left[1 - \frac{t}{n\theta}\right]^{n}. \end{aligned}$$

We achieve an approximation error of order $O_P\left(n^{-1}\right)$, same order as the asymptotic approximation.

Pivot

The "magic" behind some "good" choices of $T(\boldsymbol{X}, \boldsymbol{F})$ is that it is a pivot.

Definition

A quantity is a pivot if it is a function of $\boldsymbol{X}$ and the parameter but its distribution does not depend on the parameter. A quantity is an asymptotic pivot if its asymptotic distribution does not depend on the parameter, despite that its distribution depends on the parameter.

Example

In the uniform distribution example, $n\left(\theta - X_{[n]}\right)/\theta$ is a pivot but $n\left(\theta - X_{[n]}\right)$ is not.

Edgeworth Expansion

Suppose that S_n is a statistic with a limiting distribution $N(0,1)$. Then, the Edgeworth expansion means that

$$\mathrm{P}\left(S_n \leq s\right) = \Phi\left(s\right) + n^{-1/2} p_1\left(s\right)\phi\left(s\right) + O\left(n^{-1}\right),$$

where $p_1(x)$ is a polynomial of degree 2.

- In the special case where $S_n = \sqrt{n}\left(\bar{X} - \mu\right)/\sigma$ for iid data with $\sigma^2 = \mathrm{Var}\left(X\right)$,

$$p_1\left(s\right) = \frac{\mathrm{E}\left[\left(X-\mu\right)^3\right]}{6\sigma^3}\left(1-s\right)^2.$$

In the bootstrap world, the Edgeworth expansion yields

$$\mathrm{P}\left(S_n \leq s \mid \hat{F}\right) = \Phi\left(s\right) + n^{-1/2} p_1^*\left(s\right)\phi\left(s\right) + O\left(n^{-1}\right).$$

Second-Order Accuracy

Hence,

$$\mathrm{P}\left(S_n \leq s\right) - \mathrm{P}\left(S_n \leq s \mid \hat{F}\right) = n^{-1/2}\left[p_1\left(s\right) - p_1^*\left(s\right)\right]\phi\left(s\right) + O\left(n^{-1}\right).$$

If we are lucky such that $p_1\left(s\right) - p_1^*\left(s\right) = O_P\left(n^{-1/2}\right)$, we obtain the second-order accuracy as

$$\mathrm{P}\left(S_n \leq s\right) - \mathrm{P}\left(S_n \leq s \mid \hat{F}\right) = O_P\left(n^{-1}\right).$$

- In the special case where $S_n = \sqrt{n}\left(\bar{X} - \mu\right)/\sigma$ for iid data with $\sigma^2 = \mathrm{Var}\left(X\right)$, $\mathrm{E}\left[\left(X-\mu\right)^3\right]/\sigma^3$ is the skewness and it is asymptotically normal.

If S_n is not a pivot, then the leading $\Phi\left(\right)$ terms will not cancel each other, leading to $O_P\left(n^{-1/2}\right)$.

Regression

Consider a regression model

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

where $\mathrm{E}[\epsilon] = 0$, and $\epsilon = \begin{bmatrix}\epsilon_1 & \cdots & \epsilon_n\end{bmatrix}^T$ are iid from some distribution P. But $Y = \begin{bmatrix}Y_1 & \cdots & Y_n\end{bmatrix}^T$ are not iid. We only assume that they are independent.

- The estimator $\hat{\beta}$ of β can be estimated using X and Y. For example, the OLS estimator of β is $\hat{\beta} = \left(X^T X\right)^{-1} X^T y$.
- The residual vector is $\hat{\epsilon} = Y - X\hat{\beta}$.

Residual Bootstrap

Algorithm 1: Residual bootstrap

Fit the regression model using observed data ;
Obtain the residuals $\hat{\epsilon} = Y - X\hat{\beta}$;
Normalize the residuals, if needed, and obtain $\tilde{\epsilon}$ such that $n^{-1}\sum_{i=1}^{n}\tilde{\epsilon}_i = 0$;
for *each integer j from 1 to B* **do**
    Draw a random sample ϵ_j^* of size n from the empirical distribution of $\tilde{\epsilon}$, i.e., sample with replacement from $\tilde{\epsilon}$;
    Calculate the bootstrap response $Y_j^* = X\hat{\beta} + \epsilon_j^*$;
    Obtain the bootstrap OLS estimator $\hat{\beta}_j^* = \left(X^TX\right)^{-1}X^Ty_j^*$;
end

Casewise Bootstrap

Algorithm 2: Casewise bootstrap

for *each integer* j *from 1 to* B **do**

  Sample with replacement of size n from (y_i, x_i) ;

  Fit the model using the bootstrap sample (y_i^*, x_i^*), $i = 1, ..., n$;

  Obtain the bootstrap estimate $\hat{\beta}_j^*$;

end

Assumptions for Regression

When we perform regression analysis, we often assume that ϵ_1, ..., ϵ_n are iid from $N\left(0, \sigma^2\right)$. This assumption is ciritical for the t-test and F-test in the regression. This assumption actually contains two parts:

1. error terms are normally distributed,
2. the variances of the error terms are the same, i.e., homoscedasticity.

Once the assumption is violated, the OLS estimator often remains valid, but the t-test and the F-test are invalid.

- We can use bootstrap to overcome the violated assumptions.

Residual Bootstrap or Casewise Bootstrap?

In residual bootstrap, we sample the residuals from the empirical distribution (resample with replacement). Hence,

- it still requires homoscedasticity.
- it also assumes Y is random, but X is not random
  - For example, we want to compare the caffeine content of Lindvalls Mörkrost och Lindvalls Brygg.

In casewise bootstrap, we resample the pair (y_i, x_i). Hence,

- it does not require homoscedasticity,
- it requires both Y and X are random.

method = "case" eller method = “residual”?

Boot {car} R Documentation

Bootstrapping for regression models

Description

This function provides a simple front-end to the `boot` function in the **boot** package that is tailored to bootstrapping based on regression models. Whereas `boot` is very general and therefore has many arguments, the `Boot` function has very few arguments.

Usage

```
Boot(object, f=coef, labels=names(f(object)), R=999,
  method=c("case", "residual"), ncores=1, ...)
```

Autocorrelation

A common assumption in regression analysis is that our data are independent.

- The independence assumption is often violated in time series data due to autocorrelation.

When we resample time series data, the bootstrapped time series should keep their dependence as much as possible.

- When we deal with iid data previously, the bootstrap sample is iid conditional on the observed data.

Blockwise Bootstrap

One idea is that we can divide the time series into several blocks. If the dependence is weak and the blocks are long, the dependence is preserved.

Algorithm 3: Blockwise Bootstrap

Specify a fixed block length L such that $n = KL$;
Divide the time series into K non-overlapping blocks of length L. Denote the blocks by B_1, ..., B_K ;
for *each integer j from 1 to B* **do**
    Sample with replacement k_1^*, ..., k_K^* from 1, ..., K ;
    The bootstrapped time series is $B_{k_1^*}$, ..., $B_{k_K^*}$;
end

Blockwise Bootstrap: Some Details

- It is common to choose a length of order n^γ for $\gamma \in (0, 1)$ such that the length tends to infinity and the number of blocks also tends to infinity as $n \to \infty$.
- A limitation is that the bootstrapped series is not stationary, because the the joint distribution of resampled observations close to a join between blocks differs from that in the centre of a block.

Stationary Bootstrap

The stationary bootstrap takes blocks whose lengths L are geometrically distributed with density

$$P(L = j) = (1-p)^{j-1} p, \quad j = 1, 2, ...$$

The resampled series are stationary with mean block length p^{-1}.

Algorithm 4: Stationary Bootstrap

Specify a tuning parameter $0 < p < 1$;
for *each integer j from 1 to B* **do**
  Generate independently L_1, L_2 ... from a geometric distribution with parameter p ;
  Draw independently I_1, I_2, ... uniformly distributed on 1, ..., n ;
  The bootstrapped time series is $X_j^* = (B_{I_1,L_1}, B_{I_2,L_2}, ...)$, where $B_{a,b} = (X_a, ..., X_{a+b-1})$;
end

Stationary Bootstrap: Some Details

- If $j > n$, then we let $X_j = X_i$, where $i = j \bmod n$.
- How many L_k and I_k are simulated depends on when X_j^* reaches length n. We stop once n observations in the pseudo-time series are generated.
- Conditional on $X = (X_1, ..., X_n)$, X^* is stationary.

Confidence Set

One of the most important tasks of bootstrap is to construct confidence sets/intervals. Consider an iid sample X_1, ..., X_n from P_θ. We want a confidence set $C(\boldsymbol{X})$ for $g(\theta)$ such that

$$\mathrm{P}\left(g(\theta) \in C(\boldsymbol{X})\right) = 1 - \alpha.$$

We will introduce

1. percentile bootstrap confidence interval,
2. bootstrap pivotal method,
    1. basic bootstrap confidence interval,
    2. normal bootstrap confidence interval,
    3. studentized bootstrap confidence interval,
3. BC_a-Confidence Interval.

Approach 1: Percentile Bootstrap Interval

From the bootstrap sample we can obtain the bootstrap replications $\hat{\theta}_1^*$, ..., $\hat{\theta}_B^*$. Let the $100\alpha/2$th and $100\,(1-\alpha/2)$th percentiles of the bootstrap replications be L^* and U^*, respectively, such that

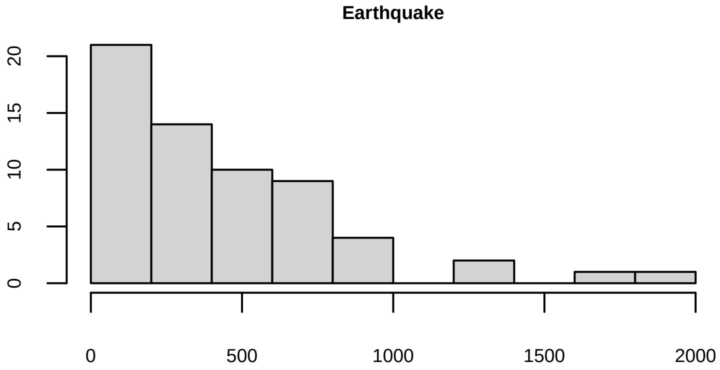
$$\mathrm{P}^*\left(L^* \leq \hat{\theta}^* \leq U^*\right) \;=\; 1-\alpha.$$

Hereafter, we use P^* to denote that the probability is evaluated for the bootstrap sample, sampled from some $\hat{F}$ (e.g., ecdf or $\mathrm{P}_{\hat{\theta}}$).
The percentile bootstrap confidence interval is

$$\{\theta :\ L^* \leq \theta \leq U^*\}\,.$$

Example: Earthquake Data

The time intervals in days between 63 successive serious earthquakes world-wide have been recorded.



We assume that our data are iid from $\mathrm{Exp}(\theta)$ with mean θ. We want to estimate $p = P(X > 1500) = \exp\left(-\frac{1500}{\theta}\right)$.

Approach 2: Bootstrap Pivotal Method

Let the random variable $Q(\boldsymbol{X}, \theta)$ be a pivot. Deriving the exact distribution of $Q(\boldsymbol{X}, \theta)$ may not be easy, but we can estimate the quantiles of the distribution of $Q(\boldsymbol{X}, \theta)$ by boostrapping.

- We can find L^* and U^* from the bootstrap sample such that

$$\mathrm{P}^*\left(L^* \leq Q\left(\boldsymbol{X}^*, \hat{\theta}\right) \leq U^*\right) = 1 - \alpha.$$

- Then we approximate

$$\mathrm{P}_\theta\left(L^* \leq Q(\boldsymbol{X}, \theta) \leq U^*\right) = 1 - \alpha.$$

The bootstrap confidence set is

$$\{\theta : L^* \leq Q(\boldsymbol{X}, \theta) \leq U^*\}$$

Basic Bootstrap Confidence Interval

Let $Q(\boldsymbol{X}, \theta) = \hat{\theta} - \theta$, although it may not be a pivot. The bootstrap replicates are $Q^*_{(j)} = \hat{\theta}^*_{(j)} - \hat{\theta}$. Then,

$$1 - \alpha = \mathrm{P}^* \left(Q^*_L \leq \hat{\theta}^* - \hat{\theta} \leq Q^*_U \right).$$

Hence,

$$\begin{aligned} 1 - \alpha &= \mathrm{P} \left(Q^*_L \leq \hat{\theta} - \theta \leq Q^*_U \right) \\ &= \mathrm{P} \left(\hat{\theta} - Q^*_U \leq \theta \leq \hat{\theta} - Q^*_L \right). \end{aligned}$$

The basic bootstrap confidence interval is

$$\left\{ \theta : \ \hat{\theta} - Q^*_U \leq \theta \leq \hat{\theta} - Q^*_L \right\}.$$

Basic Bootstrap Interval

Suppose that θ_L^* and θ_U^* are the quantiles such that

$$\mathrm{P}^*\left(\theta_L^* \leq \hat{\theta}^* \leq \theta_U^*\right) \;=\; 1-\alpha.$$

Then,

$$\mathrm{P}^*\left(Q_L^* \leq \hat{\theta}^* - \hat{\theta} \leq Q_U^*\right) \;=\; 1-\alpha$$

means that

$$Q_L^* = \hat{\theta}_L^* - \hat{\theta}, \qquad Q_U^* = \hat{\theta}_U^* - \hat{\theta}.$$

Hence, the basic bootstrap confidence interval can also be expressed as

$$\left\{\theta :\; 2\hat{\theta} - \hat{\theta}_U^* \leq \theta \leq 2\hat{\theta} - \hat{\theta}_L^*\right\}.$$

Normal Bootstrap Confidence Interval

Let

$$Q(\boldsymbol{X}, \theta) = \frac{\hat{\theta} - \theta}{\sqrt{\hat{V}}},$$

where $\hat{V}$ is an estimator of $V = \mathrm{Var}\left(\hat{\theta}\right)$, i.e., $\sqrt{\hat{V}}$ is the standard error. If the distribution of Q is $N(0,1)$, the usual asymptotic confidence interval is

$$1 - \alpha = \mathrm{P}\left(\hat{\theta} - \lambda_{1-\alpha/2}\sqrt{\hat{V}} \leq \theta \leq \hat{\theta} + \lambda_{1-\alpha/2}\sqrt{\hat{V}}\right),$$

where λ_α is the quantile of $N(0,1)$. The bootstrap counterpart is

$$1 - \alpha = \mathrm{P}^*\left(\hat{\theta}^* - \lambda_{1-\alpha/2}\sqrt{\hat{V}^*} \leq \hat{\theta} \leq \hat{\theta}^* + \lambda_{1-\alpha/2}\sqrt{\hat{V}^*}\right),$$

Normal Bootstrap Confidence Interval

1. We can estimate $V = \mathrm{Var}\left(\hat{\theta}\right)$ using the variance of the bootstrap sample.
2. We can also estimate the bias of $\hat{\theta}$ by

$$\mathrm{Bias} = \frac{1}{B}\sum_{j=1}^{B}\hat{\theta}^*_{(j)} - \hat{\theta}.$$

Hence, we can approximate

$$1-\alpha = \mathrm{P}\left(\hat{\theta} - \mathrm{Bias} - \lambda_{1-\alpha/2}\sqrt{\hat{V}^*} \le \theta \le \hat{\theta} - \mathrm{Bias} + \lambda_{1-\alpha/2}\sqrt{\hat{V}^*}\right).$$

The normal bootstrap confidence interval is

$$\left\{\theta:\ \hat{\theta} - \mathrm{Bias} - \lambda_{1-\alpha/2}\sqrt{\hat{V}^*} \le \theta \le \hat{\theta} - \mathrm{Bias} + \lambda_{1-\alpha/2}\sqrt{\hat{V}^*}\right\}.$$

Studentized Bootstrap Confidence Interval

Let

$$Q(\boldsymbol{X}, \theta) = \frac{\hat{\theta} - \theta}{\sqrt{\hat{V}}},$$

where $\hat{V}$ is an estimator of $\mathrm{Var}\left(\hat{\theta}\right)$, i.e., $\sqrt{\hat{V}}$ is the standard error. Then,

$$1 - \alpha = \mathrm{P}^* \left(Q_L^* \leq \frac{\hat{\theta}^* - \hat{\theta}}{\sqrt{\hat{V}^*}} \leq Q_U^* \right).$$

The studentized bootstrap confidence interval is

$$\left\{ \theta : \hat{\theta} - Q_U^* \sqrt{\hat{V}} \leq \theta \leq \hat{\theta} - Q_L^* \sqrt{\hat{V}} \right\}.$$

Variance Estimator

In order to use the normal bootstrap CI and the studentized bootstrap CI, we need to estimate $\text{Var}\left(\hat{\theta}\right)$.

- We have used the sample variance of the bootstrap replicates in the normal boostrap CI.

Alternative approaches include the

- nonparametric delta method, a model-by-model approach,
- bootstrap the bootstrap sample.

Bootstrap Variances For Studentized Intervals

We can apply a nested bootstrap to approximate the variance.

1. Draw a bootstrap sample.
2. Estimate the statistic as usual. Denote the estimate by $\hat{\theta}^*_{(j)}$.
3. Bootstrap the bootstrap sample, and estimate the variance by using the variance of estimates across the bootstrapped estimates. Denote the estimate of the variance by $\hat{V}^*_{(j)}$.
4. Calculate $\hat{t}^*_{(j)} = \left(\hat{\theta}^*_{(j)} - \hat{\theta}\right) / \sqrt{\hat{V}^*_{(j)}}$.

Order of Accuracy

The above confidence intervals are often first-order accurate, i.e.,

$$\mathrm{P}\left(L \leq \theta \leq U\right) = 1 - \alpha + O\left(n^{-1/2}\right).$$

The bias-corrected and accelerated (BC_a) CI is second-order accurate, i.e.,

$$\mathrm{P}\left(L \leq \theta \leq U\right) = 1 - \alpha + O\left(n^{-1}\right).$$

Second-Order Accurate

Suppose that there exists a monotone transformation $\eta = h(\theta)$ such that $\hat{\eta} = h\left(\hat{\theta}\right)$ satisfies

$$\frac{\hat{\eta} - \eta}{\sqrt{(1 + a\eta)^2}} + w = N(0,1) + o_{\mathrm{P}}(1),$$

for bias w and acceleration coefficient a.

To achieve second-order accuracy, we choose w and a such that

$$\mathrm{P}\left(\frac{\hat{\eta} - \eta}{\sqrt{(1 + a\hat{\eta})^2}} + w < z\right) = \Phi(z) + O\left(n^{-1}\right).$$

This is often possible because of the pivot and Edgeworth expansion trick that we discussed above.

BC_a-Confidence Interval

For known a and w, define

$$\hat{\eta}_{\alpha/2} = \hat{\eta} + (1 + a\hat{\eta}) \frac{\lambda_{\alpha/2} + w}{1 - a\left(\lambda_{\alpha/2} + w\right)}.$$

Then, for a monotone h, we have

$$\mathrm{P}^* \left(\hat{\eta}^* < \hat{\eta}_{\alpha/2}\right) = \Phi\left(\frac{\lambda_{\alpha/2} + w}{1 - a\left(\lambda_{\alpha/2} + w\right)} + w\right) + O\left(n^{-1}\right).$$

The limits of the BC_a interval are the the $\Phi\left(\frac{\lambda_{\alpha/2}+w}{1-a\left(\lambda_{\alpha/2}+w\right)} + w\right)$ and $\Phi\left(\frac{\lambda_{1-\alpha/2}+w}{1-a\left(\lambda_{1-\alpha/2}+w\right)} + w\right)$ quantiles of the bootstrap replicates $\left\{\hat{\theta}_{(j)}\right\}$.

Estimating a and w

We do not need to find the exact form of the transformation h. We only need its existence. But we need to estimate a and w. For example,

1. We can estimate w by
$$\hat{w} = \Phi^{-1}\left(\frac{\#\left\{\hat{\theta}^*_{(j)} < \hat{\theta}\right\}}{B}\right).$$
2. We can estimate a by
$$\hat{a} = \frac{\sum_{i=1}^{n}\left(\hat{\theta}_{(\cdot)} - \hat{\theta}_{(i)}\right)^3}{6\left[\sum_{i=1}^{n}\left(\hat{\theta}_{(\cdot)} - \hat{\theta}_{(i)}\right)^2\right]^{3/2}},$$
where $\hat{\theta}_{(i)}$ is the jackknife estimate of θ and $\hat{\theta}_{(\cdot)}$ is the average of $\hat{\theta}_{(i)}$.

Transformed Confidence Interval

Suppose that we have obtained a CI for θ as

$$\{\theta : \ L \leq \theta \leq U\}.$$

Let $\eta = h(\theta)$ be a monotone smooth transformation. The confidence interval for η is then

$$\{\eta : \ h(L) \leq \eta \leq h(U)\}.$$

Confidence Interval and Hypothesis Test

1. Using bootstrap confidence interval: suppose that the bootstrap confidence interval is $L \leq \theta \leq U$. Consider H_0: $\theta = \theta_0$. If θ_0 lies in the bootstrap CI, then we cannot reject H_0.
2. We can also bootstrap to approximate a p-value.

Suppose that we observe an iid sample X_1, ..., X_n from $X \sim P_\theta$ and we are interested in $\gamma = g(\theta)$. In hypothesis testing,

1. Construct a test statistic $T = T(X_1, ..., X_n)$ and the observed value is $t_{\text{obs}} = T(x_1, ..., x_n)$.
2. Find the distribution of T: $T \sim P_\theta^{(T)}$.

Hypothesis Testing

3 Compute the p-value

$$p = \int I_{A_{\text{obs}}}(t)\, dP_0^{(T)},$$

where A_{obs} is the region in the direction of the extreme values, $I_A(t)$ is the indicator function, and $P_0^{(T)}$ is the null distribution of T, i.e., the distribution of T under H_0.

4 Reject H_0 if $p < \alpha$ for a given significance level α.

However, it is not always easy to obtain the closed form of the null distribution $P_0^{(T)}$. But we can use bootstrap to approximate the distribution and the p-value.

Parametric Bootstrap Hypothesis Test

Suppose that $X \sim P_{\theta,\lambda}$, where θ is the parameter of interest and λ is the nuisance parameter. The distribution of test statistic T depends on θ and λ.

- Consider H_0: $\theta = \theta_0$.
- The null distribution $P^{(T)}_{\theta_0,\lambda}$ is not completely known due to unknown λ.

Algorithm 5: Parametric bootstrap hypothesis test

Estimate $\hat{\lambda}$;
for $j = 1$ *in 1 : B* **do**
  Draw a random sample $\left\{x^*_{i,j}\right\}$ of size n from $P_{\theta_0,\hat{\lambda}}$;
  Calculate the test statistic $T^*_j = T\left(x^*_{1,j}, ..., x^*_{n,j}\right)$;
end
Approximate the p-value by $p_{boot} = \#\left\{j : T^*_j \in A_{\text{obs}}\right\}/B$

Parametric Bootstrap Hypothesis Test: Example

Example

Suppose that we have a sample from

$$X = U_1 Z_1 + (1 - U_1) Z_2,$$

where $U_1 \sim \text{Bernoulli}\left(\frac{1}{2}\right)$, $Z_1 \sim N(0,1)$, and $Z_2 \sim N(\mu, 1)$. We also have another sample from

$$Y = U_2 V_1 + (1 - U_2) V_2,$$

where $U_2 \sim \text{Bernoulli}\left(\frac{3}{4}\right)$, $V_1 \sim N(0,1)$, and $V_1 \sim N(\mu + \delta, 1)$. U_1, U_2, Z_1, Z_2, V_1, and V_2 are mutually independent. We want to test

$$H_0 : \delta = 0 \qquad H_1 : \delta > 0,$$

where μ is the nuisance parameter. From our data, $n = 20$, $\bar{X} = 2.576$, and $\bar{Y} = 2.211$.

Non-Parametric Bootstrap Hypothesis Test

If we prefer non-parametric bootstrap, the algorithm is similar to parametric bootstrap test. But we must have a resampling procedure that incorporates H_0.

Algorithm 6: Parametric bootstrap hypothesis test

Estimate $\hat{\lambda}$;
for $j = 1$ *in 1 : B* **do**
    Draw a random sample $\left\{x_{i,j}^{*}\right\}$ of size n under H_0 ;
    Calculate the test statistic $T_j^* = T\left(x_{1,j}^*, ..., x_{n,j}^*\right)$;
end
Approximate the p-value by $p_{boot} = \#\left\{j : T_j^* \in A_{\mathrm{obs}}\right\}/B$

Non-Parametric Bootstrap Hypothesis Test: Example

Example

Suppose that we have an iid sample X_1, ..., X_n from some distribution and another iid sample Y_1, ..., Y_n from the same distribution but shifted by δ. These two samples are independent. We want to test

$$H_0: \delta = 0 \qquad H_1: \delta \neq 0.$$