Bootstrap Confidence Intervals

Max Turgeon

STAT 3150–Statistical Computing

Lecture Objectives

- Learn how to compute the different bootstrap confidence intervals.
- Understand their theoretical properties.

Motivation

- So far, we've mostly built CIs using the CLT, and we can certainly do so with bootstrap.
- But since we are (sort of) sampling from the sampling distribution, we can actually do better.

Bootstrap confidence intervals

- There are several ways to construct confidence intervals in bootstrap:
    - Standard normal bootstrap
    - Bootstrap percentile
    - Basic bootstrap
    - Student bootstrap
    - BCa interval
- They all have different properties, and they can all be useful depending on the context.

Standard normal bootstrap CI i

- This is similar to what we've been doing until now.
- It relies on the Central Limit Theorem:

$$\frac{\hat{\theta} - E(\hat{\theta})}{SE(\hat{\theta})} \to N(0, 1).$$

- If we estimate $\widehat{bias}(\hat{\theta})$ and $SE(\hat{\theta})$ using bootstrap, then we can construct an approximate $100(1-\alpha)\%$ confidence interval for θ via

$$\hat{\theta} - \widehat{bias}(\hat{\theta}) \pm z_{\alpha/2} SE(\hat{\theta}).$$

Standard normal bootstrap CI ii

- This interval is easy to compute, but it assumes that the sampling distribution is approximately normal.
    - Works well for estimators $\hat{\theta}$ that can be expressed as a sample mean (e.g. Monte Carlo integration)
    - Doesn’t work well when the sampling distribution is skewed.

Bootstrap percentile CI

- Let $\hat{\theta}^{(b)}$, $b = 1, \ldots, B$ be the bootstrap estimates.
- The **bootstrap percentile confidence interval** is the interval of the form $(\hat{\theta}_{\alpha/2}, \hat{\theta}_{1-\alpha/2})$, where $\hat{\theta}_{\alpha/2}$ and $\hat{\theta}_{1-\alpha/2}$ are the $\alpha/2$-th and $1 - \alpha/2$-th sample quantiles of the bootstrap estimates, respectively.
- **Don't be fooled by its simplicity!** Its validity actually requires strong assumptions (see notes on UM Learn).
    - In particular, when bias is large, you can get unrealistic CIs.

Basic bootstrap CI i

- This is also known as the **pivotal bootstrap CI**.
- It is very similar to the bootstrap percentile approach, but instead of taking the sample quantiles of $\hat{\theta}^{(b)}$, $b = 1, \ldots, B$, we take the sample quantiles of the *pivot quantities* $\hat{\theta}^{(b)} - \hat{\theta}$, $b = 1, \ldots, B$.
- Note that the β-th quantile of $\hat{\theta}^{(b)} - \hat{\theta}$ is equal to $\hat{\theta}_\beta - \hat{\theta}$, where $\hat{\theta}_\beta$ is the β-th quantile of $\hat{\theta}^{(b)}$.
- To build the basic bootstrap CI, we take $\hat{\theta}$ minus some critical values. But instead of using the critical values of the standard normal, we take our critical values from the *pivot quantities*:

$$\hat{\theta} - (\hat{\theta}_\beta - \hat{\theta}) = 2\hat{\theta} - \hat{\theta}_\beta.$$

Basic bootstrap CI ii

- Therefore, the **basic bootstrap** $100(1-\alpha)$% confidence interval for θ is

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

- **Why use basic over percentile?** It turns out the basic bootstrap CI has better theoretical properties and stronger convergence guarantees.

Example i

We will compute the above 3 types of confidence intervals for the correlation between LSAT and GPA scores.

```
library(bootstrap)
B <- 5000
n <- nrow(law)
boot_rho <- replicate(B, {
  # Sample with replacement
  indices <- sample(n, n, replace = TRUE)
  cor(law$LSAT[indices], law$GPA[indices])
})
```

Example ii

```
rho_hat <- cor(law$LSAT, law$GPA)
bias <- mean(boot_rho) - rho_hat
se <- sd(boot_rho)
```

```
# 1. Standard normal
c(rho_hat - bias - 1.96*se,
  rho_hat - bias + 1.96*se)
```

```
## [1] 0.5200736 1.0418491
```

Example iii

```
# 2. Bootstrap percentile
quantile(boot_rho,
         probs = c(0.025, 0.975))
```

```
##      2.5%     97.5%
## 0.4553849 0.9628290
```

Example iv

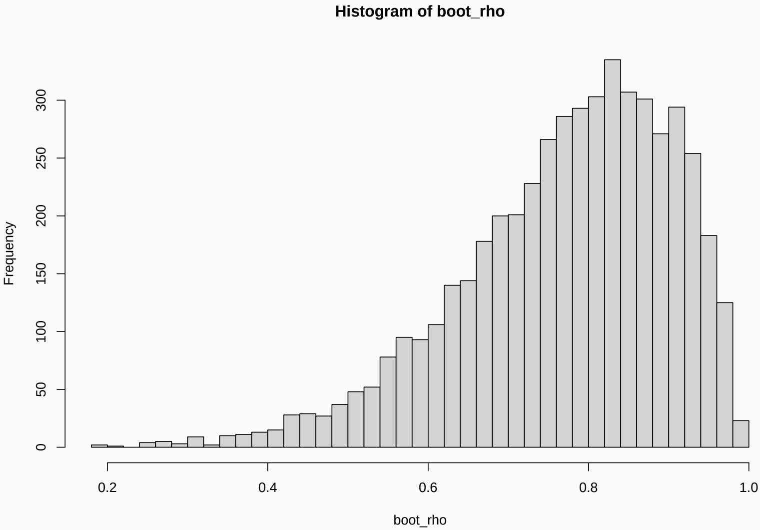
```r
# 3. Basic bootstrap
crit_vals <- quantile(boot_rho,
                      probs = c(0.025, 0.975))
c(2*rho_hat - crit_vals[2],
  2*rho_hat - crit_vals[1],
  use.names = FALSE)
```

```
## [1] 0.589920 1.097364
```

Table 1: Only the percentile method gives a sensible confidence interval, i.e. a CI that is contained within the interval $(-1, 1)$.

Method	95% CI
Standard Normal	(0.52, 1.04)
Percentile	(0.46, 0.96)
Basic Bootstrap	(0.59, 1.1)

Example vi



Histogram of boot_rho
Frequency
0
50
100
150
200
250
300
0.2
0.4
0.6
0.8
1.0
boot_rho

Student bootstrap CI i

- This confidence interval accounts for the fact we have to estimate the standard error.
- However, it is much more involved: we can construct an approximate $100(1-\alpha)\%$ confidence interval for θ via

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

where $t^*_{1-\alpha/2}$ and $t^*_{\alpha/2}$ are computed using a **double bootstrap**, and where $SE(\hat{\theta})$ is the usual bootstrap estimate of the standard error.

Student bootstrap CI ii

Algorithm

1. For each bootstrap sample estimate $\hat{\theta}^{(b)}$, compute a “t-type” statistic $t^{(b)} = \frac{\hat{\theta}^{(b)} - \hat{\theta}}{SE(\hat{\theta}^{(b)})}$, where $SE(\hat{\theta}^{(b)})$ is specific to the b-th sample, and it can be computed using bootstrap on the samples $X_1^{(b)}, \ldots, X_n^{(b)}$.
2. From the sample $t^{(b)}, b = 1, \ldots, B$, let $t^*_{1-\alpha/2}$ and $t^*_{\alpha/2}$ be the $1 - \alpha/2$-th and $\alpha/2$-th sample quantiles.

This confidence interval is more accurate than the standard normal bootstrap CI, but this accuracy comes with a large computational cost.

Example (cont'd) i

```
# 4. Student bootstrap
boot_rho_t <- replicate(B, {
  indices <- sample(n, n, replace = TRUE)
  rho_b <- cor(law$LSAT[indices], law$GPA[indices])
  double_boot <- replicate(100, {
    double_ind <- sample(indices, n, replace = TRUE)
    cor(law$LSAT[double_ind], law$GPA[double_ind])
  })
  tb <- (rho_b - rho_hat)/sd(double_boot)
  return(c(rho_b, tb))
})
```

Example (cont'd) ii

```
# The output has two rows:
# First row: rho_b values
# Second row: tb values
str(boot_rho_t)
```

```
## num [1:2, 1:5000] 0.7932 0.1144 0.7741 -0.0215
0.5658 ...
```

```
# SE estimated using rho_b values
SE <- sd(boot_rho_t[1,])
```

Example (cont'd) iii

```
# t critical values
tcrit_vals <- quantile(boot_rho_t[2,],
                       probs = c(0.025, 0.975))
```

```
c(rho_hat - tcrit_vals[2]*SE,
  rho_hat - tcrit_vals[1]*SE,
  use.names = FALSE)
```

```
## [1] -0.3212171  0.9898829
```

- This is a valid confidence interval, but it is much wider than the other three!

BCa confidence intervals i

- The BCa confidence interval is an improvement on the bootstrap percentile approach.
  - "BCa" stand for "bias-corrected" and "adjusted for acceleration".
- Let Φ be the CDF of the standard normal distribution.
- The **BCa confidence interval** is defined using quantiles of the bootstrap sample: $(\hat{\theta}_{\beta_1}, \hat{\theta}_{\beta_2})$, where

$$\beta_1 = \Phi\left(\hat{z}_0 + \frac{\hat{z}_0 + z_{\alpha/2}}{1 - \hat{a}(\hat{z}_0 + z_{\alpha/2})}\right),$$
$$\beta_2 = \Phi\left(\hat{z}_0 + \frac{\hat{z}_0 + z_{1-\alpha/2}}{1 - \hat{a}(\hat{z}_0 + z_{1-\alpha/2})}\right).$$

BCa confidence intervals ii

- The quantities $\hat{z}_0$ and $\hat{a}$ are correction factors for bias and skewness, respectively.
    - If we have $\hat{z}_0 = 0$ and $\hat{a} = 0$, then the formulas above simplify to $\beta_1 = \alpha/2$ and $\beta_2 = 1 - \alpha/2$, and the BCa interval then becomes the same as the bootstrap percentile.
- The bias correction factor is defined as

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

where Φ^{-1} is the *quantile function* from the standard normal distribution.

BCa confidence intervals iii

- Note that $\hat{z}_0 = 0$ if and only if $\hat{\theta}$ is the median of the bootstrap samples.
- The acceleration factor is estimated using jackknife:

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

where $\overline{\theta_{(\cdot)}}$ is the sample mean of the jackknife estimates $\hat{\theta}_{(i)}$.

- **Note**:for Student, we need a second level of bootstrap.
  - This leads to a total of $B_1 \cdot B_2$ iterations, where B_1 and B_2 are the number of bootstrap samples at each level.

BCa confidence intervals iv

- On the other hand, for the BCa interval, the bootstrap and the jackknife are done independently.
    - This leads to a total of $B + n$ iterations, which is typically less than $B_1 \cdot B_2$.

Example (cont'd) i

```r
# First estimate z0 hat
z0_hat <- qnorm(mean(boot_rho < rho_hat))
z0_hat
```

```
## [1] -0.1206099
```

Example (cont'd) ii

```r
# Next: Jackknife
rho_i <- numeric(n)

for (i in 1:n) {
  rho_i[i] <- cor(law$LSAT[-i], law$GPA[-i])
}
```

Example (cont'd) iii

```
# Then estimate a hat
rho_bar <- mean(rho_i)
ahat_num <- sum((rho_bar - rho_i)^3)
ahat_denom <- 6*sum((rho_bar - rho_i)^2)^(3/2)
(a_hat <- ahat_num/ahat_denom)
```

```
## [1] -0.07567156
```

Example (cont'd) iv

```
# Putting everything together
beta1 <- pnorm(z0_hat + (z0_hat - 1.96) /
                 (1 - a_hat*(z0_hat - 1.96)))
beta2 <- pnorm(z0_hat + (z0_hat + 1.96) /
                 (1 - a_hat*(z0_hat + 1.96)))
c(beta1, beta2)
```

```
## [1] 0.004798663 0.932417162
```

Example (cont'd) v

```
# BCa interval
quantile(boot_rho, probs = c(beta1, beta2))
```

```
## 0.4798663%  93.24172%
##  0.3202500  0.9396505
```

```
# Compare with percentile
quantile(boot_rho, probs = c(0.025, 0.975))
```

```
##      2.5%     97.5%
## 0.4553849 0.9628290
```

Theoretical properties i

- Two theoretical properties of interest:
  - Transformation invariant: If (a, b) is a confidence interval for a parameter θ, then for any monotone transformation m, the interval $(m(a), m(b))$ is a confidence interval for the parameter $m(\theta)$.
  - Accuracy: We say a confidence interval is *first-order* accurate if its error goes to zero at the same rate as $1/\sqrt{n}$; we say it is *second-order* accurate if its error goes to zero at the same rate as $1/n$ (so twice as fast).

Theoretical properties ii

	Transformation Invariant	Accuracy
Standard normal	No	First order
Percentile	Yes	First order
Basic Bootstrap	Yes	First order
Student CI	No	Second order
BCa interval	Yes	Second order

Theoretical properties iii

- The BCa interval is the **only one** of the five that is both transformation invariant and second-order accurate.
    - This comes with a steep computational price (we need a second level of resampling)
- **Recommendation**: Use BCa, unless computation time is an issue. In that case, use basic bootstrap.