Bootstrap and Jackknife

Anaranya Basu

November 2023

Problem Statement:
Suppose (X,Y) $\sim N_2$(0,0,1,1,0.3). Fix n=10. Draw (X_i, Y_i), i=1,2,...10. Find the var(r), Bias(r), Median(r) via bootstrap. Repeat using jackknife too.
Solution:
The Bootstrap samples $x^* = (x_1^*, x_2^*, ..., x_n^*)$) are drawn with replacement from the original sample x=$(x_1, x_2, ...x_n)$. The bootstrap algorithm for estimate the standard error of T(X) is as follows:

- B independent bootstrap samples $x^{*1}, x^{*2}, ...x^{*B}$ are generated from the original dataset x:-$(x_1, x_2, ...x_n)$, where, $x^{*b} = (x_1^{*b}, x_2^{*b}, ..., x_n^{*b})$
- Bootstrap replicates $T(x^{*1}), T(x^{*2}), ..., T(x^{*B})$ are obtained by calculating the value of the statistic of our interest T(x) on each bootstrap samples.
- Finally, the standard deviation of the values $T(x^{*1}), T(x^{*2}), ..., T(x^{*B})$ is our estimate of the standard error of T(x).

$$s.\hat{e}_B = \frac{1}{B-1} \sum_{b=1}^{B} ((T(x^{*b}) - T(.))^2)^{\frac{1}{2}}$$

where $T(.) = \frac{\sum_{b=1}^{B} T(x^{*b})}{B}$

- Hence, the bootstrap estimate of variance is given by -

$$V\hat{a}r_B = s.\hat{e}_B{}^2 = \frac{1}{B-1} \sum_{b=1}^{B} ((T(x^{*b}) - T(.))^2)$$

Computation:
We generated observations from Bivariate Normal distribution as mentioned in the problem in R

```
                [,1]        [,2]
 [1,] -1.176046470  0.27230665
 [2,] -0.398443756  0.02729371
 [3,]  1.019571231  1.49377041
 [4,] -0.008635095  0.12232646
```

```
 [5,]  0.433075154 -0.22460495
 [6,]  0.325577530  2.43988167
 [7,]  0.077070210  0.66613484
 [8,]  0.143541426 -2.18339137
 [9,] -0.968686220 -0.13883073
[10,] -0.079596998 -0.63901134
```

Now fixing B=1000 we again draw random samples with replacement, i.e. $(x_1^{*b}, x_2^{*b}, ..., x_n^{*b}) \sim \hat{F}$
The simulated observations are as below-

```
    [1] 0.30306355 0.37334016 0.56568709 0.41231025 0.58589678
   [6] 0.16685906 0.26628431 0.33486549 0.26325855 0.39769976
  [11] 0.45883012 0.51228843 0.50060127 0.63092709 0.57412467
  [16] 0.45557759 0.25065712 0.68606308 0.33830140 0.19368714
  [21] 0.69089943 0.27813700 0.41013538 0.31138707 0.27874609
  [26] 0.16907330 0.43030923 0.62186782 0.46076887 0.38940206
  [31] 0.38546104 0.60894862 0.34983801 0.59494152 0.28485140
  [36] 0.37582036 0.38579818 0.61825948 0.44789845 0.22730949
  [41] 0.38004009 0.31450200 0.42448867 0.40756578 0.64496169
  [46] 0.25447246 0.39687184 0.24582151 0.42843154 0.44729788
  ...........................................................
  ...........................................................
  ............................................................
  [981] 0.50515327 0.35833152 0.67352938 0.49124025 0.26112305
 [986] 0.13444140 0.08207273 0.03273771 0.15023787 0.59451006
 [991] 0.21424516 0.26986539 0.63331742 0.20095496 0.58747373
 [996] 0.24233122 0.55844543 0.33251879 0.32894024 0.38217006
```

Therefore the estimated variance, $\hat{VAR}_{BOOT}T(X)$= **0.02672302**
Estimated Bias = -0.03398355
Estimated Median = 0.3770298

Jackknife Estimate:
The jackknife estimator of the variance of T(X) is obtained as,

- The jackknife focuses on the samples that leave out one observation at a time:

  $x_(i) = (x_1, x_2, ...x_{i-1}, x_{i+1}, ..., x_n)$. These $x_i^{'}$ s are called jackknife samples.

- Jackknife replicates $T_{n-1,s}$'s, i=1,2,...n are obtained by calculating the value of the statistic of our interest T(x) on each jackknife samples.

- The jackknife estimate of variance of T (X) is obtained by

$$V\hat{AR}_{JACK} = \frac{n-1}{n} \sum_{i=1}^{n} (T_{n-1,i} - \frac{1}{n} \sum_{j=1}^{n} T_{n-1,j})^2 \tag{1}$$

Computation:

$Bias_{JACK} = (n-1)(\hat{T_n} - T_n) = 0$
$Var_{JACK} = 0.03871397$

Appendix

```
set.seed(123)
mean_x <- 0
mean_y <- 0
var_x <- 1
var_y <- 1
correlation <- 0.3
# Generate bivariate normal data
n <- 10
data <- MASS::mvrnorm(n, mu = c(mean_x, mean_y), Sigma = matrix(c(var_x, correlation, correl
# Function to calculate the sample statistic of interest
calculate_statistic <- function(data) {
  return(var(data[, 1]))
}
# Bootstrap function
bootstrap <- function(data, B) {
  statistics <- numeric(B)
  for (i in 1:B) {

    resampled_data <- data[sample(1:n, replace = TRUE), ]

    statistics[i] <- calculate_statistic(resampled_data)
  }
```

```
  return(statistics)
}
# Number of bootstrap samples
B <- 1000
bootstrap_results <- bootstrap(data, B)
original_statistic <- calculate_statistic(data)
bias <- mean(bootstrap_results) - original_statistic
variance <- var(bootstrap_results)
median_estimate <- median(bootstrap_results)
#######################
# Function to calculate the sample statistic of interest
calculate_statistic <- function(data) {
  return(var(data[, 1]))
}
# Jackknife function
jackknife <- function(data) {
  n <- nrow(data)
  statistics <- numeric(n)
  for (i in 1:n) {
        subset_data <- data[-i, ]
    statistics[i] <- calculate_statistic(subset_data)
  }
  return(statistics)
}
jackknife_results <- jackknife(data)
original_statistic <- calculate_statistic(data)
bias_jackknife <- (n - 1) * (mean(jackknife_results) - original_statistic)
variance_jackknife <- ((n - 1) / n) * sum((jackknife_results - mean(jackknife_results))^2)
```