Modules 5: Exercises with answers

Davide Ferrari
Free University of Bozen/Bolzano

Bootstrap

One of the great advantages of the bootstrap approach is that it can be applied in almost all situations. No complicated mathematical calculations are required. Performing a bootstrap analysis in R entails only two steps. First, we must create a function that computes the statistic of interest. Second, we use the boot() function, which is part of the boot library, to perform the bootstrap by repeatedly sampling observations from the data set with replacement.

```
library(boot)
library(ISLR)
data(Portfolio)

alpha.fn = function( data , index )
{
X = data$X[ index ]
Y = data$Y[ index ]
out = ( var(Y) - cov(X,Y))/(var (X) + var(Y) - 2*cov(X,Y))
return(out)
}
```

This function returns, or outputs, an estimate for based on applying a statistic of interest to the observations indexed by the argument index. For instance, the following command tells R to estimate α using the first 50 observations.

```
alpha.fn( Portfolio ,1:50)

## [1] 0.6308955
```

The next command uses the sample() function to randomly select 100 ob- servations from the range 1 to 100, with replacement. This is equivalent to constructing a new bootstrap data set and recomputing based on the new data set.

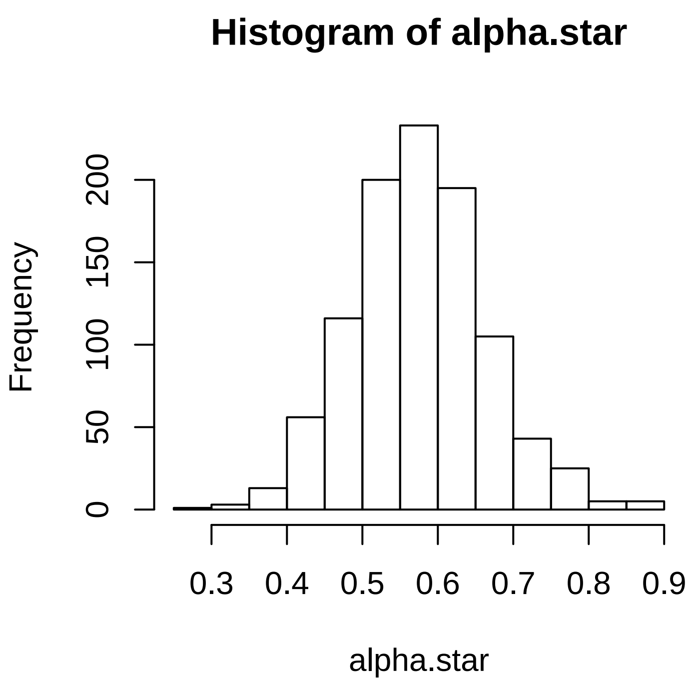
```
set.seed (1)
alpha.fn(Portfolio , sample(100 ,100 , replace =T ) )

## [1] 0.5963833
```

We can implement a bootstrap analysis by performing this command many times, recording all of the corresponding estimates for α, and computing the resulting standard deviation. However, the boot() function automates this approach. Below we produce R = 1000 bootstrap estimates for α

```
boot.result =  boot(Portfolio , alpha.fn , R =1000)
boot.result

##
## ORDINARY NONPARAMETRIC BOOTSTRAP
##
##
## Call:
## boot(data = Portfolio, statistic = alpha.fn, R = 1000)
##
##
## Bootstrap Statistics :
##      original        bias    std. error
## t1* 0.5758321 -7.315422e-05  0.08861826

names(boot.result)

##  [1] "t0"        "t"         "R"         "data"      "seed"
##  [6] "statistic" "sim"       "call"      "stype"     "strata"
## [11] "weights"
```

The final output shows that using the original data, $\hat{\alpha} = 0.5758$, and that the bootstrap estimate for $SE(\hat{\alpha})$ is 0.0886. The following creates a histogram for the bootstrapped statistics and computes a 95% bootstrap confidence interval based on their distribution.

```
alpha.star = boot.result$t
hist(alpha.star)
```




```
quantile(alpha.star, prob=c(0.025, 0.975))

##      2.5%     97.5%
## 0.4102428 0.7575927
```