Homework 9: Bootstrapping (Solutions)

Oliver d’Pug

## Create a **b.var** function

I’m lazy and will use the **boot** package to run my bootstrap. I need a function that computes the variance that can be passed to the **boot** function.

 b.var = function(d, i){
 var(d[i])
 }

## Create a dataset of 100 normal observations

 args(rnorm)

function (n, mean = 0, sd = 1)
NULL

 n = 100
 mu = 2
 sig = 7
 normal\_data = rnorm(n, mu, sig)
 write.csv(normal\_data, "hw9\_normal\_data.csv")
 c(summary(normal\_data), var\_x=var(normal\_data), sd\_x=sd(normal\_data))

 Min. 1st Qu. Median Mean 3rd Qu. Max. var\_x
-21.993247 -2.073824 3.010643 3.130375 8.066483 17.262490 52.399774
 sd\_x
 7.238769

### Bootstrap the variance using **b.var**

 ### Use the boot function to run the bootstrap
 normal\_b.var = boot(normal\_data, b.var, R=9999)
 normal\_b.var

ORDINARY NONPARAMETRIC BOOTSTRAP

Call:
boot(data = normal\_data, statistic = b.var, R = 9999)

Bootstrap Statistics :
 original bias std. error
t1\* 52.39977 -0.524394 8.152271

### How much bias does the estimate have? Is the bootstrap distribution normal or chisquare?

 attributes(summary(normal\_b.var))

$names
[1] "R" "original" "bootBias" "bootSE" "bootMed"

$row.names
[1] 1

$class
[1] "summary.boot" "data.frame"

 summary(normal\_b.var)

 R original bootBias bootSE bootMed
1 9999 52.4 -0.52439 8.1523 51.345

 plot(normal\_b.var)



 qqPlot(normal\_b.var$t, distribution="norm")



[1] 7840 1397

 qqPlot(normal\_b.var$t, distribution="chisq", df=n-1)



[1] 7840 1397

The variance estimator appears to have bias=-0.524394 and SE=8.1522711 when the data consist of 100 observations randomly drawn from a normal with mean $μ=$ 2 and variance $σ^{2}=$ 49. The variance appears to be neither normally nor chisquare distributed, although the chisquare may fit a bit better.

## Create a dataset of 100 standard normal observations

 args(rnorm)

function (n, mean = 0, sd = 1)
NULL

 n = 100
 mu = 0
 sig = 1
 std\_normal\_data = rnorm(n, mu, sig)
 write.csv(std\_normal\_data, "hw9\_std\_normal\_data.csv")
 c(summary(std\_normal\_data), var\_x=var(std\_normal\_data), sd\_x=sd(std\_normal\_data))

 Min. 1st Qu. Median Mean 3rd Qu. Max.
-3.33324516 -0.62310758 0.02923251 -0.01607254 0.66450345 2.78512838
 var\_x sd\_x
 1.10213077 1.04982416

### Bootstrap the variance using **b.var**

 ### Use the boot function to run the bootstrap
 std\_normal\_b.var = boot(std\_normal\_data, b.var, R=9999)
 std\_normal\_b.var

ORDINARY NONPARAMETRIC BOOTSTRAP

Call:
boot(data = std\_normal\_data, statistic = b.var, R = 9999)

Bootstrap Statistics :
 original bias std. error
t1\* 1.102131 -0.008390659 0.1827047

### How much bias does the estimate have? Is the bootstrap distribution normal or chisquare?

 summary(std\_normal\_b.var)

 R original bootBias bootSE bootMed
1 9999 1.1021 -0.0083907 0.1827 1.0841

 plot(std\_normal\_b.var)



 qqPlot(std\_normal\_b.var$t, distribution="norm")



[1] 541 5865

 qqPlot(std\_normal\_b.var$t, distribution="chisq", df=n)



[1] 541 5865

The variance estimator appears to have bias=-0.0083907 and SE=0.1827047 when the data consist of 100 observations randomly drawn from a normal with mean $μ=$ 0 and variance $σ^{2}=$ 1. The variance appears not to be normally distributed. However, it appears to have a chisquare distribution. The MGF of the sum of squared standard normals supports this contention.

## Create a dataset of 100 U(0,1)

 args(rnorm)

function (n, mean = 0, sd = 1)
NULL

 n = 100
 a = 0
 b = 1
 unif\_data = runif(n, a, b)
 write.csv(unif\_data, "hw9\_unif\_data.csv")
 c(summary(unif\_data), var\_x=var(unif\_data), sd\_x=sd(unif\_data))

 Min. 1st Qu. Median Mean 3rd Qu. Max.
0.001091276 0.173983906 0.424912276 0.442482873 0.679215578 0.973776141
 var\_x sd\_x
0.090396014 0.300659298

### Bootstrap the variance using **b.var**

 ### Use the boot function to run the bootstrap
 unif\_b.var = boot(unif\_data, b.var, R=9999)
 unif\_b.var

ORDINARY NONPARAMETRIC BOOTSTRAP

Call:
boot(data = unif\_data, statistic = b.var, R = 9999)

Bootstrap Statistics :
 original bias std. error
t1\* 0.09039601 -0.0009440588 0.007664379

### How much bias does the estimate have? Is the bootstrap distribution normal or chisquare?

 summary(unif\_b.var)

 R original bootBias bootSE bootMed
1 9999 0.090396 -0.00094406 0.0076644 0.089403

 plot(unif\_b.var)



 qqPlot(unif\_b.var$t, distribution="norm",)



[1] 6016 1394

 qqPlot(unif\_b.var$t, distribution="unif", min=a, max=b)



[1] 6016 1394

 qqPlot(unif\_b.var$t, distribution="beta", shape1=(3\*n-1)/2, shape2=(3\*n-1)/2)



[1] 6016 1394

When the data consist of 100 observations randomly drawn from a U( 0 , 1), the variance estimator appears to be unbiased as bias=-9.4405878^{-4}. The standard error is SE=0.0076644. The variance appears to be somewhat normally distributed, but it is not U( 0 , 1) distributed. For appropriately chosen shape parameters, $α$ and $β$, the distribution of the variance appears to be almost beta.