Assignment 4 Question 1 d)

Sheen Thusoo

30/11/2021

For each of the two attributes of interest estimate the coverage probability when using the percentile method and give the standard error of your estimate. For the simulation, choose an appropriate number of samples and number of bootstrap samples. In addition, provide a conclusion about the procedure.

```
data <- read.csv('EconomicMobility.csv')
```

```
VarIQR <- function(pop, N) {
  variance <- var(pop) * (N-1)/N
  iqr <- IQR(pop)
  c(variance, iqr)
}

var.pop <- VarIQR(data$Mobility, length(data$Mobility))[1]
iqr.pop <- VarIQR(data$Mobility, length(data$Mobility))[2]
```

```
M <- 100 # number of samples
B <- 1000 # number of bootstrap samples
n <- 100 # size of samples

var_boolean = c()
iqr_boolean = c()
var_result = matrix(0,100,2)
iqr_result = matrix(0,100,2)

set.seed(341)

for(i in 1:M){
  indx = sample(data$Mobility, n, replace=FALSE)
  var=c()
  iqr =c()
  for(j in 1:B){
      Bindx = sample(indx,n,replace=TRUE)
  var[j]=VarIQR(Bindx, length(Bindx))[1]
  iqr[j]=VarIQR(Bindx, length(Bindx))[2]

  }
  var_result[i,] = quantile(var,c(0.025,0.975))
  iqr_result[i,] = quantile(iqr,c(0.025,0.975))
```

```
  var_boolean[i] = (var.pop>var_result[i,1]) & (var.pop<var_result[i,2])
  iqr_boolean[i] = (iqr.pop>iqr_result[i,1]) & (iqr.pop<iqr_result[i,2])
}
```

```
cat("The coverage probability of the variance is: ", mean(var_boolean))
```

```
## The coverage probability of the variance is:  0.84
```

```
cat("\nThe coverage probability of IQR is: ", mean(iqr_boolean))
```

```
##
## The coverage probability of IQR is:  0.95
```

```
sdn <- function(p) {
  sqrt( sum(( p - mean(p))^2) / length(p))
}
var_SE <-  sdn(var_boolean) / sqrt(M)
iqr_SE <-  sdn(iqr_boolean) / sqrt(M)

cat("Standard error of Variance is ", var_SE)
```

```
## Standard error of Variance is  0.03666061
```

```
cat("\nStandard error of IQR is ", iqr_SE)
```

```
##
## Standard error of IQR is  0.02179449
```

Conclusion

The procedure for estimating the coverage probability of the attributes involves first generating M (in this case $M = 100$) samples from the population (size N = 729), which is the Mobility variate. For each of these samples, B bootstrap samples of size n are generated and the attribute is calculated on them:

$$a_b = a(S_b^*)$$

Next, from the values $a_1, ..., a_B$, we find $a_{lower} = Q_a(0.025)$ and $a_{upper} = Q_a(0.975)$. The 95 confidence interval for each sample is then $[a_{lower}, a_{upper}]$. For each sample M, a confidence interval is generated. In this case, this meant that 100 confidence intervals were generated. To find the coverage probability, we calculate how many of these intervals contain the population attribute value $a(P)$; we set a value of True if the interval contains the population attribute value and False if it does not. From this list of Booleans, we can calculate the number of intervals containing $a(P)$ divided by the total number of intervals to get the coverage probability.

From the results, the coverage probability for the Variance attribute is 0.84 with a standard error of 0.037, whereas the coverage probability for the IQR attribute is higher at 0.95 with a standard error of 0.022. The lower coverage probability of the variance attribute is likely due variation caused by random sampling. If we had generated all possible confidence intervals instead, we would see an exact coverage probability of 0.95. As we increase the size of the samples n, we see that the coverage probability approaches the expected coverage probability of 0.95.