STAT400 - Homework 9

Your Name

Due 11/12/2020 by 4pm

Be sure to `set.seed(400)` at the beginning of your homework.

```r
#reproducibility
set.seed(400)
```

```r
# useful libraries
library(tidyverse)
```

1. Use the Monte Carlo simulation to investigate whether the empirical Type I error rate of the t-test is approximately equal to the nominal significance level when the sampled population is non-normal.

   a. For $n = 5, 10, 30, 100, 500, 1000$, investigate the empirical type I error for a test of $H_0 : \mu = 1$ vs. $H_a : \mu \neq 1$ when $X_1, \dots, X_n \sim \chi^2(1)$ with $m = 2000$ Monte Carlo samples with nominal $\alpha = .05$.

   b. For $n = 5, 10, 30, 100, 500, 1000$, investigate the empirical type I error for a test of $H_0 : \mu = 1$ vs. $H_a : \mu \neq 1$ when $X_1, \dots, X_n \sim Unif[0, 2]$ with $m = 2000$ Monte Carlo samples with nominal $\alpha = .05$.

   c. For $n = 5, 10, 30, 100, 500, 1000$, investigate the empirical type I error for a test of $H_0 : \mu = 1$ vs. $H_a : \mu \neq 1$ when $X_1, \dots, X_n \sim Exponential(1)$ with $m = 2000$ Monte Carlo samples with nominal $\alpha = .05$.

   d. Compare your results in a.-c. in a table. What can you say about the departures from Normality as they relate to the Type I error rate of the t-test?

```r
# function to compute empirical alpha based on n, alpha, m, and the generating distn

# a. \chi^2(1)
type1chi <- function(n,m,alpha){

 ind <- rep(NA, m)
    for(i in seq_len(m)){
     sample<-rchisq(n,1)
     tstar<-(mean(sample)-1)/sd(sample)

     ind[i] <- (abs(tstar) >= qnorm(1 - alpha/2))
    }
 mean(ind)
   }




# b. Unif[0, 2]\
type1u <- function(n,m,alpha){
```

```r
 ind <- rep(NA, m)
    for(i in seq_len(m)){
     sample<-runif(n,0,2)
     tstar<-(mean(sample)-1)/sd(sample)

     ind[i] <- (abs(tstar) >= qnorm(1 - alpha/2))
    }
 mean(ind)
   }

# c. Exp(1)
type1e <- function(n,m,alpha){

 ind <- rep(NA, m)
    for(i in seq_len(m)){
     sample<-rexp(n,1)
     tstar<-(mean(sample)-1)/sd(sample)

     ind[i] <- (abs(tstar) >= qnorm(1 - alpha/2))
    }
 mean(ind)
   }

# make table to compare results
n<-c(5, 10, 30, 100, 500, 1000)
table <-c()
for (i in n){
    table <- c(table,c(type1chi(n,2000,0.05),type1u(n,2000,0.05),type1e(n,2000,0.05)))
}
table <- matrix(table,ncol=3,byrow = TRUE)
colnames(table)<-c("CHI","Unif","exp")
rownames(table)<-c("n=5","n=10","n=30","n=100","n=500","n=1000")
table <- as.table(table)
table
```

```
##            CHI   Unif    exp
## n=5     0.0820 0.0100 0.0325
## n=10    0.0750 0.0110 0.0385
## n=30    0.0750 0.0070 0.0335
## n=100   0.0725 0.0100 0.0370
## n=500   0.0800 0.0125 0.0330
## n=1000  0.0725 0.0100 0.0370
```

We can see that the further the distributions is from the normal the worse the empirical type 1 error coverage is, i.e. where the $\chi^2(1)$ is the worst one and the U(0,1) is the best one in this case.