Homework 7

10/29/2021

Question 1 (7.1)

Estimate the MSE of the level k trimmed means for random samples of size 20 generated from a standard Cauchy distribution. (The target parameter θ is the center or median; the expected value does not exist.) Summarize the estimates of MSE in a table for $k = 1, 2, ..., 9$.

```
m <- 1000 # number of replicates
n <- 20 # number of sample size

tmean <- numeric(m)
mse <- numeric(9)

for(k in 1:9){
  for(i in 1:m){
    x <- sort(rcauchy(n, 0, 1))
    tmean[i] <- sum(x[(k+1):(n-k)])/(n-2*k)
  }
  mse[k] <- mean(tmean^2)
}
```

-Ans

k	$\hat{MSE}$
1	1.3967671
2	0.4360118
3	0.2434097
4	0.176511
5	0.1385372
6	0.1418786
7	0.1320039
8	0.1268151
9	0.1412657

Question 2 (7.6)

Suppose a 95% symmetric $t - interval$ is applied to estimate a mean, but the sample data are non-normal. Then the probability that the confidence interval covers the mean is not necessarily equal to 0.95. Use a Monte Carlo experiment to estimate the coverage probability of the $t - interval$ for random samples of $\chi^2(2)$ data with sample size $n = 20$. Compare your $t - interval$ results with the simulation results in Example 7.4. (The $t - interval$ should be more robust to departures from normality than the interval for variance.)

```
m <- 1000
n <- 20
alpha <- 0.05
TL <- numeric(m)
TU <- numeric(m)
UCL <- numeric(m)
k <- 0

for (i in 1:m) {
  x <- rchisq(n,2)
  TL[i] <- mean(x) - qt(1-alpha, n-1)*sd(x)/sqrt(n)
  TU[i] <- mean(x) + qt(1-alpha, n-1)*sd(x)/sqrt(n)
}

for (i in 1:m) {
   #MC Exercise
  x <- rchisq(n,2)
  if (mean(TL) < mean(x) && mean(x) < mean(TU)) k <- k+1

  #Example 4
  y <- rnorm(n, 0, 2)
  UCL[i] <- ((n-1)*var(y))/qchisq(alpha,df=n-1)
}
```

-Ans: Using the Monte Carlo experiment the coverage probability is 0.901 and the coverage from example 4 is 0.956.

Question 3

Suppose that X_1, X_2 are *iid* from a standard normal distribution, and X_3, X_4 are *iid* from a $Gamma(3, 2)$ distribution. Obtain a Monte Carlo estimation of $\theta = E[g(X_1, X_2, X_3, X_4)] = E|(X_1 + X_2)^2 - (X_3 + X_4)^2|$ based on $m = 1,000$ replicates.

```
m <- 1000
g <- numeric(m)

for(i in 1:m){
  x1 <- rnorm(1)
  x2 <- rnorm(1)
  x3 <- rgamma(1, 3, 2)
  x4 <- rgamma(1, 3, 2)
  g[i] <- abs((x1 + x2)^2 - (x3 + x4)^2)
}

theta.hat <- mean(g)
```

-Ans: $\hat{\theta} = 9.0165038$.

Question 4

In introductory statistical courses, you have learned how to test or estimate proportions. This is implemented by *prop.test* in *R*. Read the help file of *prop.test* function and understand the usage of this function. Suppose $X_1, ..., X_{20}$ is a random sample from $Bernoulli(p = 0.6)$ distribution. (Note: You can use *rbinom(n,1,p)* to generate this random sample, $n = 20$ in this case). Construct a 90% confidence interval for the parameter p using the *prop.test* function. Use a Monte Carlo method to obtain an empirical estimate of the confidence level and check if it is close to the nominal level 90%.

```
#Prop.Test
n <- 20
m <- 1000
alpha <- 0.1
p0 <- 0.6
x <- sum(rbinom(n, 1, p0))
ptest <- prop.test(x, n, p = p0, conf.level= 1-alpha)

#Monte Carlo Estimate
UCL <- replicate(m, expr = {
  x <- rnorm(n,n*p0,sqrt(n*p0*(1-p0)))
  (n-1)*var(x)/qchisq(alpha,df=n-1)
})
```

-Ans: Confidence Interval from *prop.test*: [0.4423272,0.817166]. Confidence level from MC: 0.894.

Question 5

Suppose $X_1, ..., X_{20}$ is a random sample from $Bernoulli(p)$ distribution. Test $H_0 : p = 0.7$ *vs* $H_1 : p \neq 0.7$ at $\alpha = 0.05$. Under the null hypothesis,

$$T^* = \frac{\hat{p}}{\sqrt{\frac{0.7*(1-0.7)}{n}}} \sim N(0, 1).$$

Use a Monte Carlo method to compute an empirical probability of Type I Error and check that it is approximately equal to $\alpha = 0.05$. (Note: You can use the *prop.test* function in R).

```
n <- 20
alpha <- 0.05
mu0 <- 0.7

m <- 10000  #number of replicates
p <- numeric(m)  #storage for p-value
for (j in 1:m){
  x <- rbinom(n, 1, mu0)
  c <- sum(x)
  ptest <- prop.test(c, n, p = mu0, conf.level = 1-alpha)
  p[j] <- ptest$p.value
}

p.hat <- mean(p < alpha)
```

-Ans: P-value = 0.0466