

STP598-Assignment 3

Hao Wang
hwang306

October 26, 2017

1 Question 1

A monte carlo test for four bootstrap confidence intervals.

2 Question 2

For this question I checked t-test performance of three different distributions. $\alpha = 0.9$. The sample size I pick is 100.

2.1 chi-sq

```
# chi-sq distribution with 1 degrees of freedom
n <- 100 # sample size
alpha <- 0.1
mu0 <- 1
m <- 10000 # replication
p <- numeric(m)

for (i in 1:m) {
  x <- rchisq(n, 1)
  ttest <- t.test(x, alternative = "two.sided", mu = mu0)
  p[i] <- ttest$p.value
}
p.hat <- mean(p < alpha)
p.hat

## [1] 0.1163
```

3 Question 3

1. in bivariate normal, the Spearman correlation is less powerful than Pearson correlation
2. find an alternative where Spearman has better empirical power than Pearson correlation.

3.1 bivariate normal

```
n <- 1000 #sampling size
m <- 10000 #number of replications
cor0 <- -0.4 #my pick of alternative hypothesis
power.p <- numeric(m) # storage of test statistics of pearson
power.s <- numeric(m)

#generate bivariate normal distribution, assume mu = 0,
# covaraiance matrix is given by
Sigma <- matrix(c(1, -0.5, -0.5, 1),
                nrow = 2, byrow = TRUE) # covariance matrix
mu <- c(0,0) #mean
library(MASS)

# run the loop
for (i in 1:m) {
  bvn1 <- mvrnorm(n, mu = mu, Sigma = Sigma ) # from MASS package
  bvn1 <- as.data.frame(bvn1)
  colnames(bvn1) <- c("x","y")
  pearson <- cor.test(x, y, method = "pearson")
  spearmen <- cor.test(x, y, method = "spearmen")
}

bvn1 <- mvrnorm(n, mu = mu, Sigma = Sigma ) # from MASS package
bvn1 <- as.data.frame(bvn1)

colnames(bvn1) <- c("x","y")
attach(bvn1)

a <- cor.test(x, y, method = "pearson")
```

```
a <- a$p.value
b <- cor(x, y, method = "spearman")
```

To get the empirical power of the two tests,

4 Question 4

Monte Carlo integration

$$\theta = \int_0^{0.5} e^{-x} dx$$

let $f(x) = \frac{1}{0.5-0}$ then

$$\theta = (0.5 - 0) \int_0^{0.5} e^{-x} \frac{1}{0.5 - 0} dx = 0.5E[e^{-x}]$$

then

$$\hat{\theta} = 0.5Eg(X) = 0.5E(e^{-x})$$

The sample mean variance is

$$Var(\hat{\theta}) = 0.25/m * Var(g(x)) = \frac{0.25}{m} Var(e^{-x})$$

```
# generate unif(0, 0.5)
m <- 10000
theta <- numeric(m)
x <- runif(m, min = 0, max = 0.5)
theta.hat <- mean(exp(-x))*0.5
theta.hat
```

```
## [1] 0.3932094
```

```
# variance of theta.hat
for (i in 1:m) {
  theta[i] <- exp(-x[i])
}
variance <- 0.25/m * var(theta)
variance
```

```
## [1] 3.197632e-07
```

5 Question 5

mixture of normal distribution

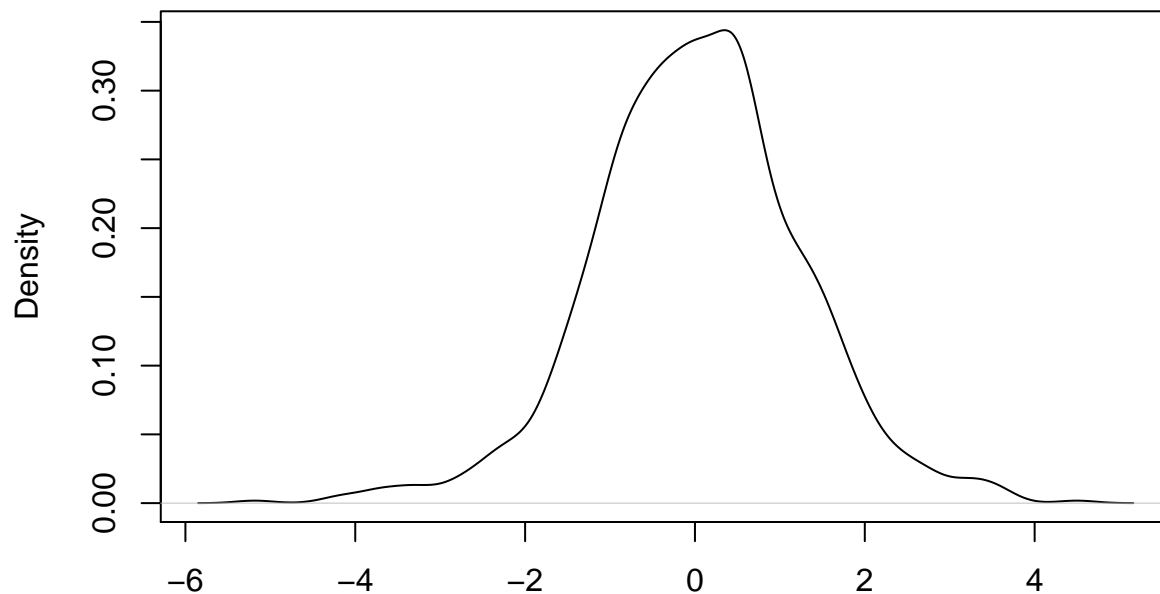
```

library(MASS)
# write function
loc.mix.0 <- function(n, p, mu1, mu2, Sigma1, Sigma2) {
  #generate sample from BVN location mixture
  X <- matrix(0, n, 2)
  for (i in 1:n) {
    k <- rbinom(1, size = 1, prob = p)
    if (k)
      X[i,] <- mvrnorm(1, mu = mu1, Sigma1) else
      X[i,] <- mvrnorm(1, mu = mu2, Sigma2)
    }
  return(X)
}

X <- loc.mix.0(1000, 0.75, 0, 0, 1, 3)
plot(density(X))

```

density.default(x = X)

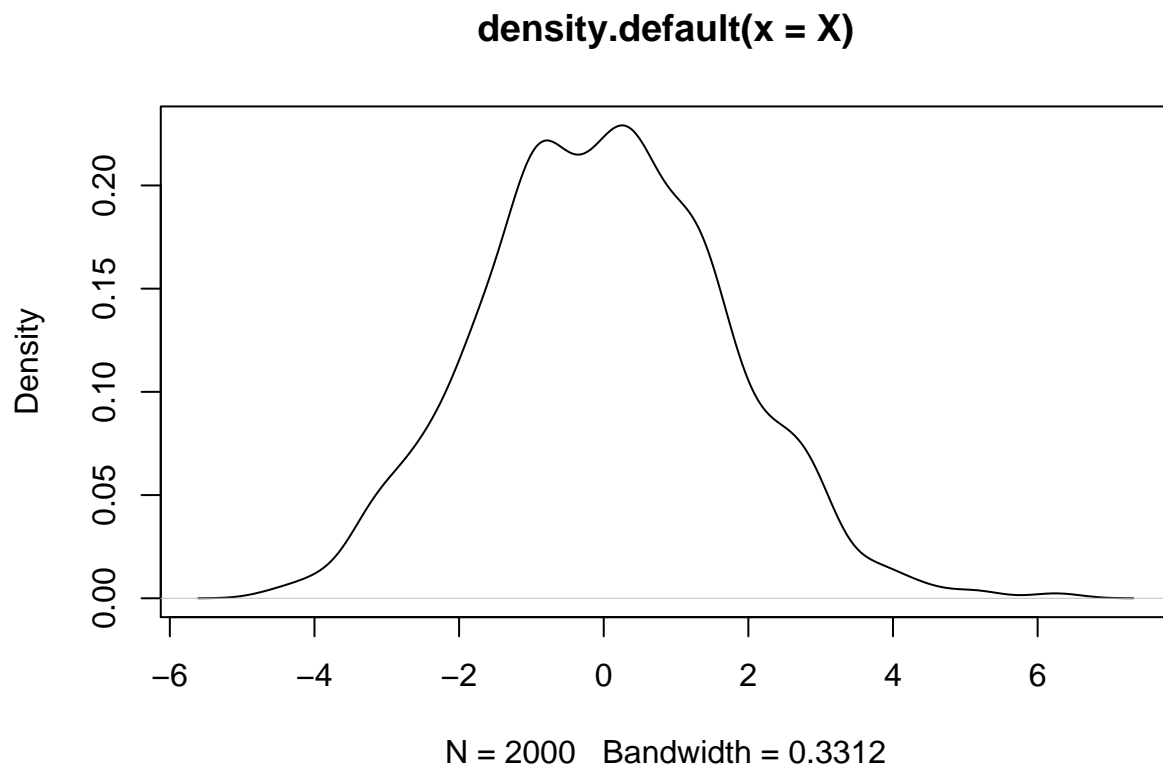


N = 2000 Bandwidth = 0.2216

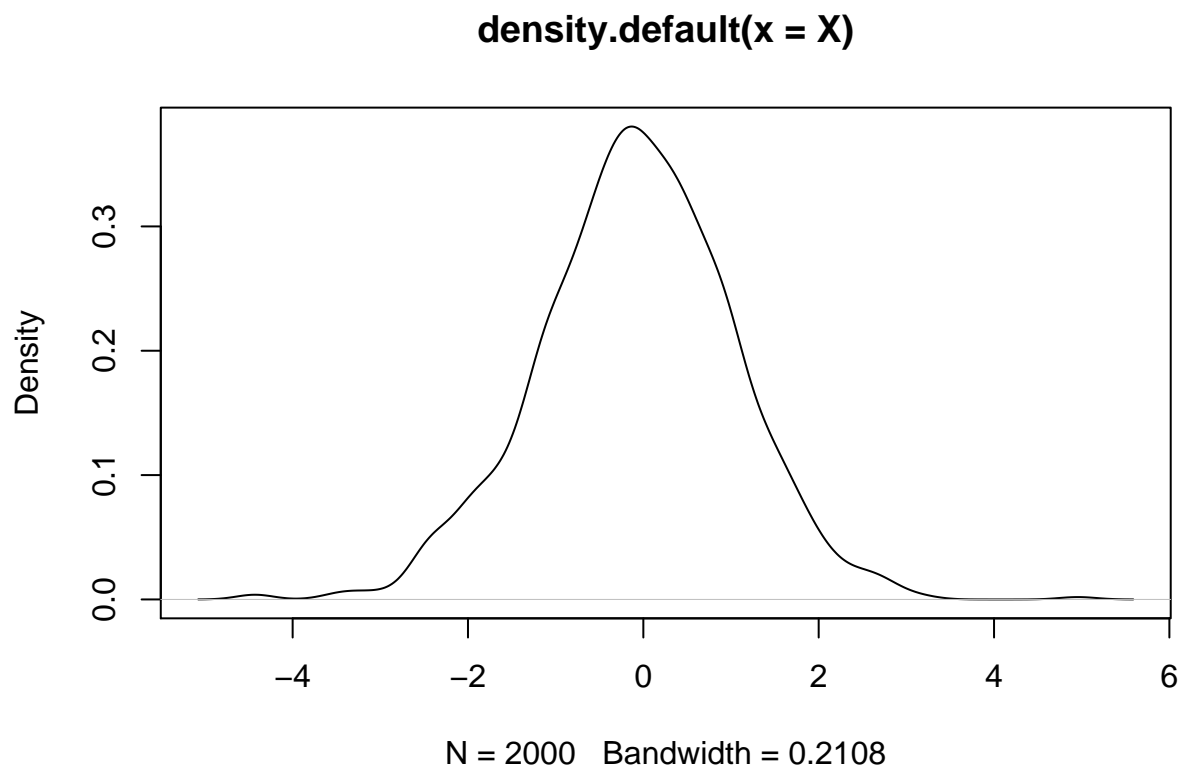
```

X <- loc.mix.0(1000, 0.1, 0, 0, 1, 3)
plot(density(X))

```



```
X <- loc.mix.0(1000, 0.9, 0, 0, 1, 3)
plot(density(X))
```



I changed multiple p values, this does not look like a bimodal distribution. I think it is

because the $\mu_1 = \mu_2 = 0$. The parts with highest probability are overlapping, thus the mixture looks like a normal distribution.

6 Question 6

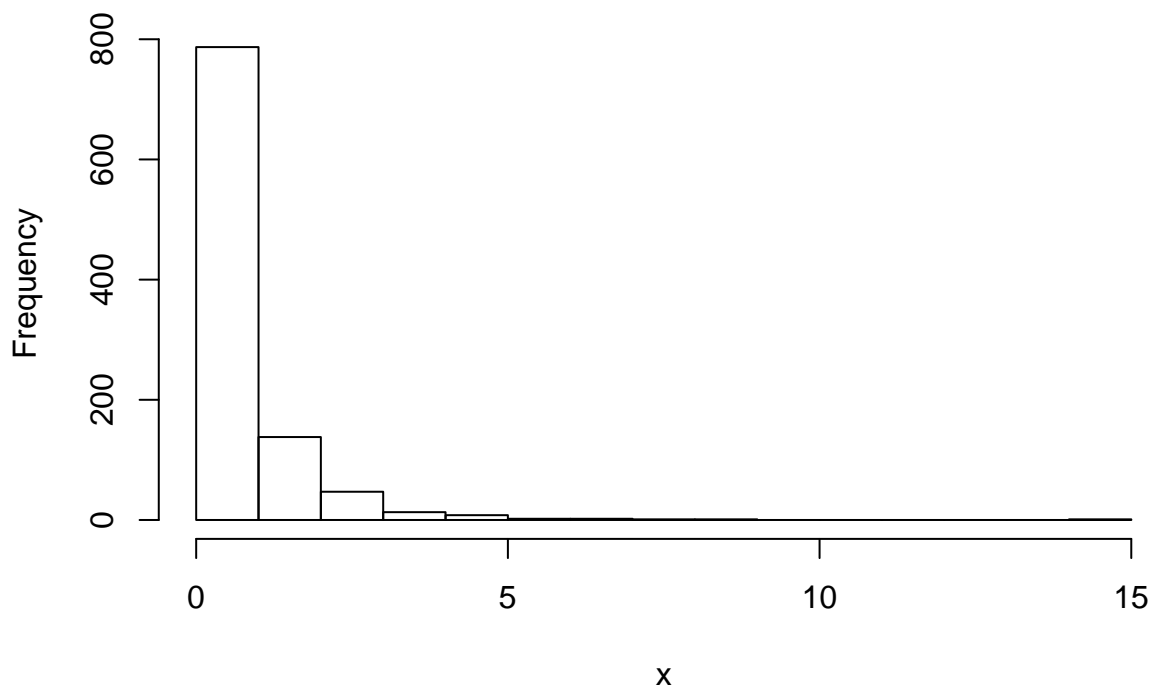
Gamma-Exponential mixture.

$$\Lambda \sim \Gamma(r, \beta)$$
$$(Y|\Lambda = \lambda) \sim \exp(\Lambda)$$

```
# generate distributions of Lambda
n <- 1000 #1000 random variables
r <- 4
beta <- 2
lambda <- rgamma(n, r, beta)

#now apply the sample of lambda as the exponential
x <- rexp(n, lambda)
hist(x)
```

Histogram of x

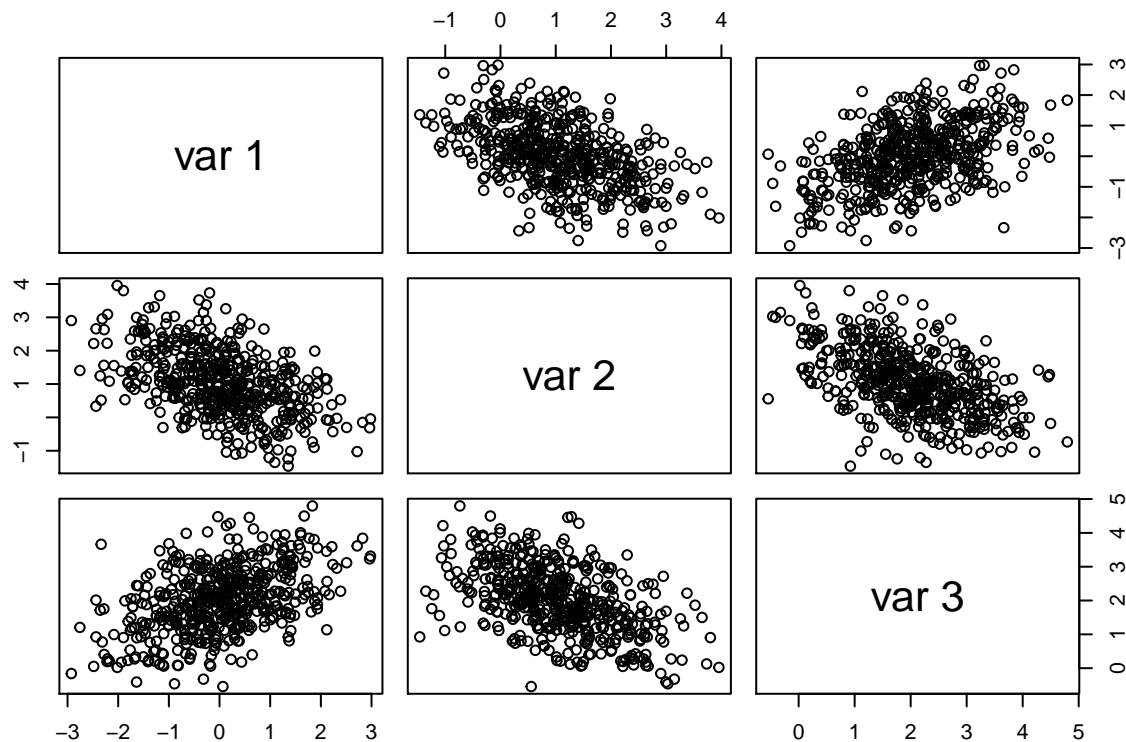


7 Question 7

requires to generate 500 random observations of the 3-dimensional MVN, using Choleski factorization method.

write a function of factorization method

```
rmvn.Choleski <-  
function(n, mu, Sigma){  
  # generate n random vectors from MVN(mu, Sigma)  
  # dimension is inferred from mu and Sigma  
  d <- length(mu)  
  Q <- chol(Sigma) # Choleski factorization of Sigma  
  Z <- matrix(rnorm(n*d), nrow = n, ncol = d)  
  X <- Z %*% Q + matrix(mu, n, d, byrow = TRUE)  
  X  
}  
# write input mean matrix and covariance matrix  
  
mu <- c(0,1,2)  
Sigma <- matrix(c(1, -0.5, 0.5, -0.5, 1, -0.5, 0.5, -0.5, 1),  
               nrow = 3, byrow = TRUE)  
  
X <- rmvn.Choleski(500, mu, Sigma)  
pairs(X)
```



From the graph, it satisfies the theoretical expectation. The joint distribution of each pair of marginal distributions is theoretically bivariate normal. var1 and var2 are negatively correlated, and var1 and var 3 are postively correlated.