

Sardana_Module4 HW

2023-02-06

Problem Set1 1

```
n = 20
p = 0.7
E_x <- n*p
q <- (1-p)
V_x <- n*p*q
V_x

## [1] 4.2

# Probability of Y greater than 15
# There are 100 microRNA's independent of each other.

P_Y <- 1-pnorm(15, E_x, sd = sqrt(V_x)/sqrt(100))
P_Y

## [1] 5.317746e-07
```

Problem Set 2 # Require package mvtnorm

```
require(mvtnorm)

## Loading required package: mvtnorm

mean_x = 7
var_x = 3
mean_y <- 12
var_y <- 7
covar_xy <- 3

# No. of simulations = 1000
nsim <- 1000

XmeanLess.sim <- rep(NA, nsim)

for (i in 1:nsim) {

  data_simulation <- rmvnorm(100, mean = c(mean_x, mean_y), sigma = matrix(c(
var_x, covar_xy, covar_xy, var_y), ncol = 2)) # Generate a sample of size 100
from the bivariate normal distribution

  mean_simulation <- apply(data_simulation, 2, mean) # Find the mean of data
simulation
```

```

Xmean <- mean_simulation[1]

Ymean <- mean_simulation[2]

XmeanLess.sim[i] <- (Xmean + 0.5 < Ymean) # Logical statement
}

# Mean of all Monte Carlo simulations
mean(XmeanLess.sim)

## [1] 1

```

Sub part of Problem 2

```

# Monte carlo simulation for 95% confidence interval

mean(XmeanLess.sim) + c(-1,1) * 1.96 * sqrt(var(XmeanLess.sim) / nsim)

## [1] 1 1

```

Problem3 Calculate the mean of new random variable

```

n = 1000

# x1 follows chi square distribution
x1 <- rchisq(n,df = 8)
# x2 follows gamma distribution
x2 <- rgamma(n,shape = 1,scale = 2)
# x3 follows t test distribution
x3 <- rt(n,df = 5)

y <- sqrt(x1)*x2 + 4*(x3)^2
mean(y)

## [1] 12.21762

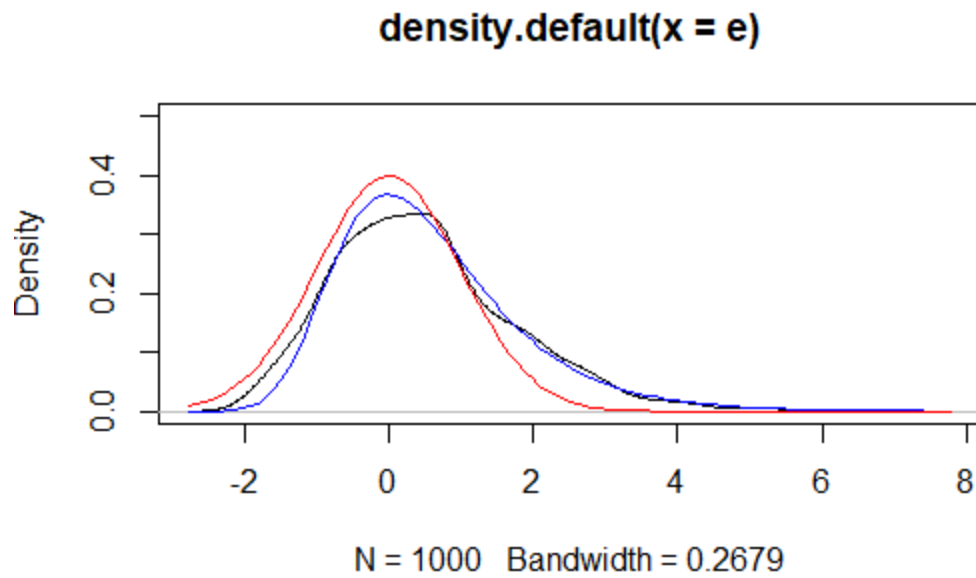
```

Problem4

```

an <- sqrt(2*log(n)) - 0.5*(log(log(n))+log(4*pi))*(2*log(n))^(1/2)
bn <- (2*log(n))^(1/2)
e <- double(); n <- 10000 # Serfling p.90
for (i in 1:1000) e[i] <- (max(rnorm(n))-an)/bn
plot(density(e),ylim=c(0,0.5))
f<-function(x){exp(-x)*exp(-exp(-x))}
curve(f,range(density(e)$x),add=TRUE,col = "blue")
curve(dnorm,add=TRUE,col = "red")

```



According to kernel density bandwidth selection, the big bandwidth over smooth the density curve, however a small bandwidth overfit the kernel density distribution.

In the above graph , black line is near to the original data produced.

The density of blue line fits much better to the black line than the red.