## **Monte Carlo Simulation to Identify Independence and -distribution**

**Introduction**

The student’s theorem have stated that following the definition of a t-distributed RV, the independence of and with r degrees of freedom iff will lead to . Thus, it aroused our interest to investigate possible dependents of and such that . Different techniques such as Monte Carlo simulations and transformations methods are being applied while solving the questions. Results were being obtained throughout the exploration, which and are not independent as and are not independent.

**Student’s Theorem**

According to Student’s Theorem, if , the sample mean and sample variance are independent. The independence of and leads to . This follows the definition of a -distributed RV.

**The definition of a -distributed RV**

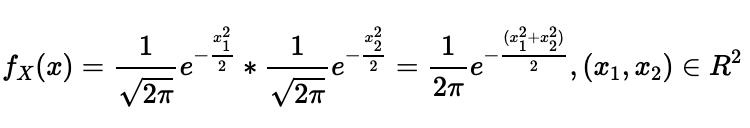
Let and such that and are independent. A RV is said to have a -distribution with degrees of freedom iif .

**Independence and -distribution**

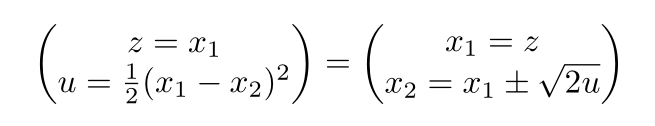
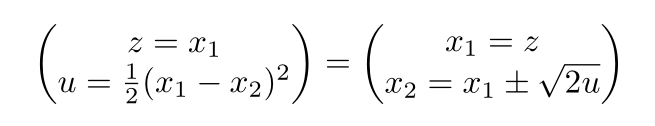
Suppose . Let and , so that and . It follows that and are not independent if and are not independent according to the theorem and definition explained above.

Suppose we consider the case , , so that and .

We first obtain the joint distribution of . Since and are standard normal distributions, we have the following equation.



In addition, suppose .

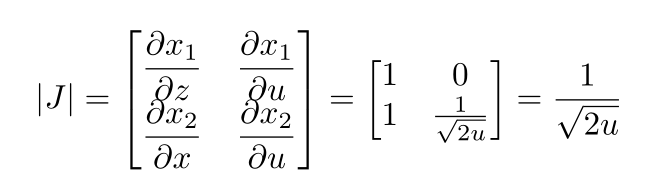
We use the transformation method by equating and arranging the vectors and with respect to and as follows;

⇔

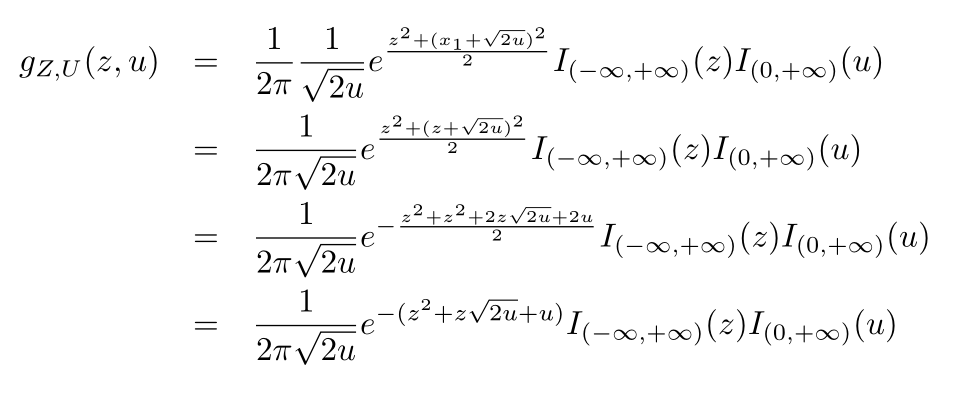
Thus, we have and .

Since we have two different values for , there are two ways to define range as following;

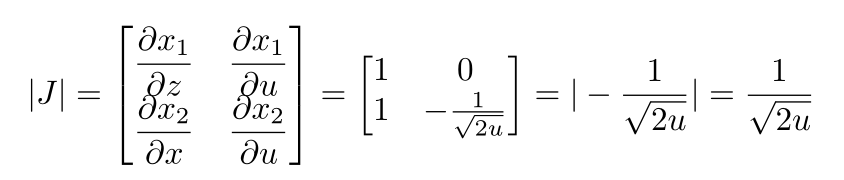
For , we have and where , hence, we obtain the following for .

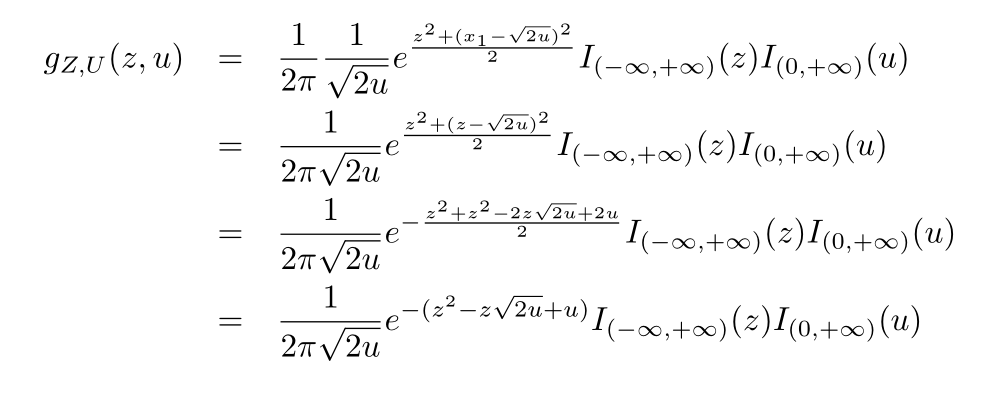


and the following joint distribution.

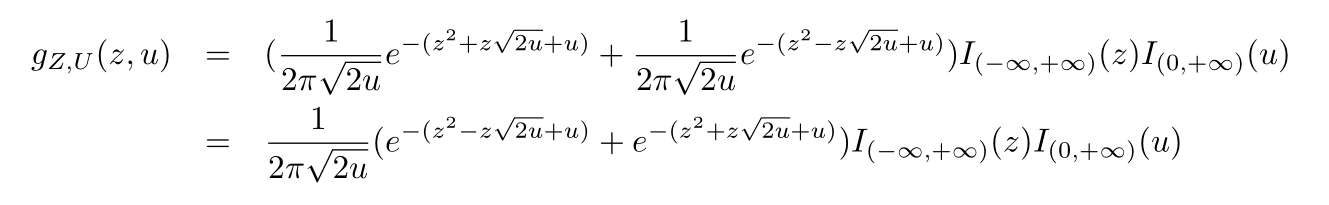


Similarly, for , we obtain and joint distribution as follows.

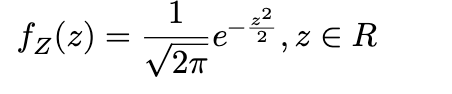


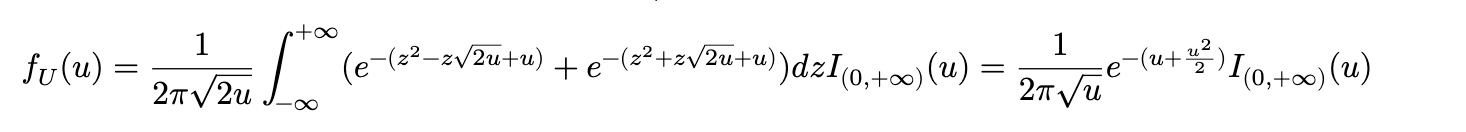


Therefore, we obtain the the joint distribution of and as follows;

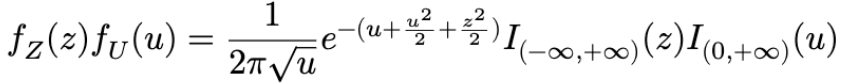


Suppose we have independent and such as





Then, the joint distribution would be

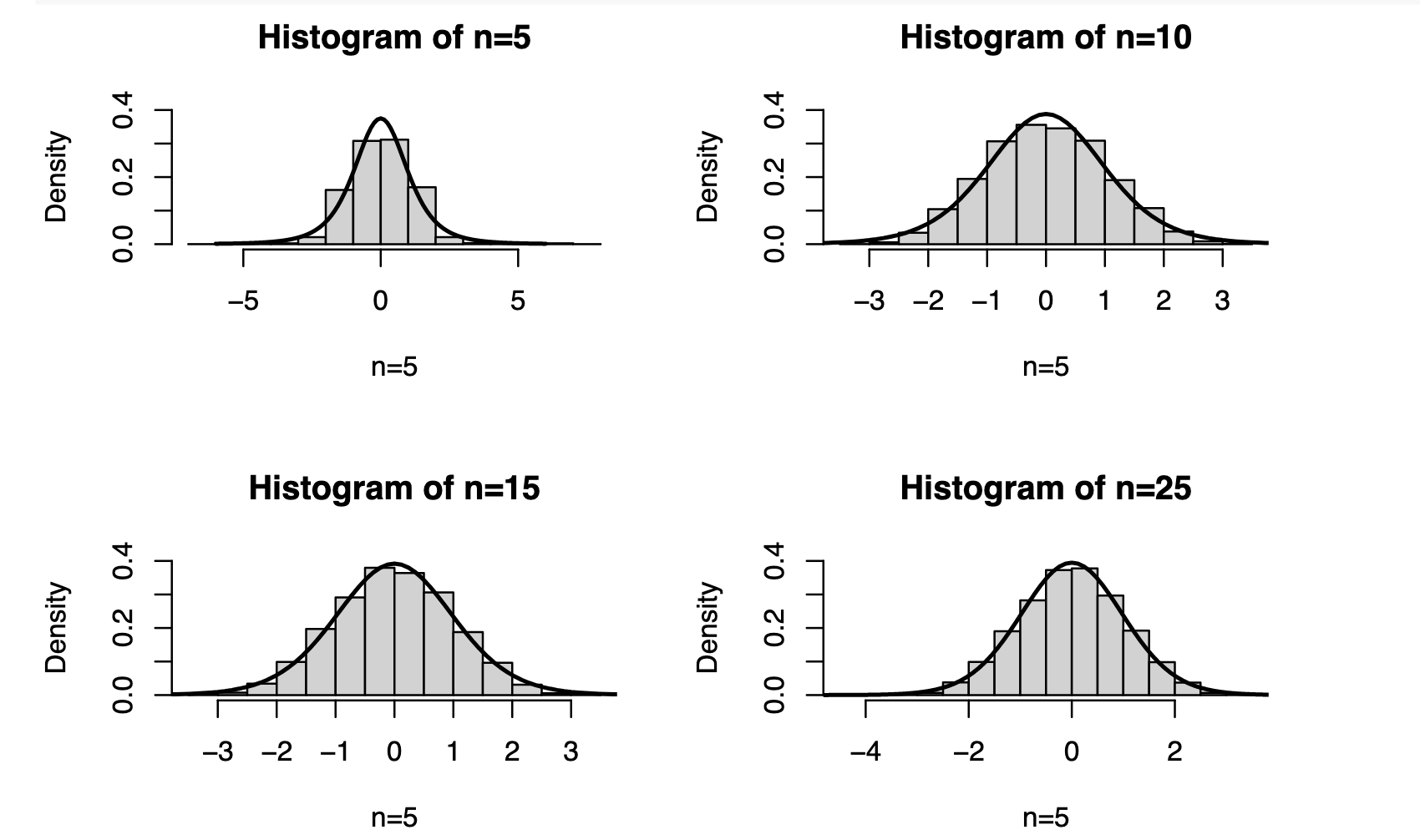


which does not equal to .

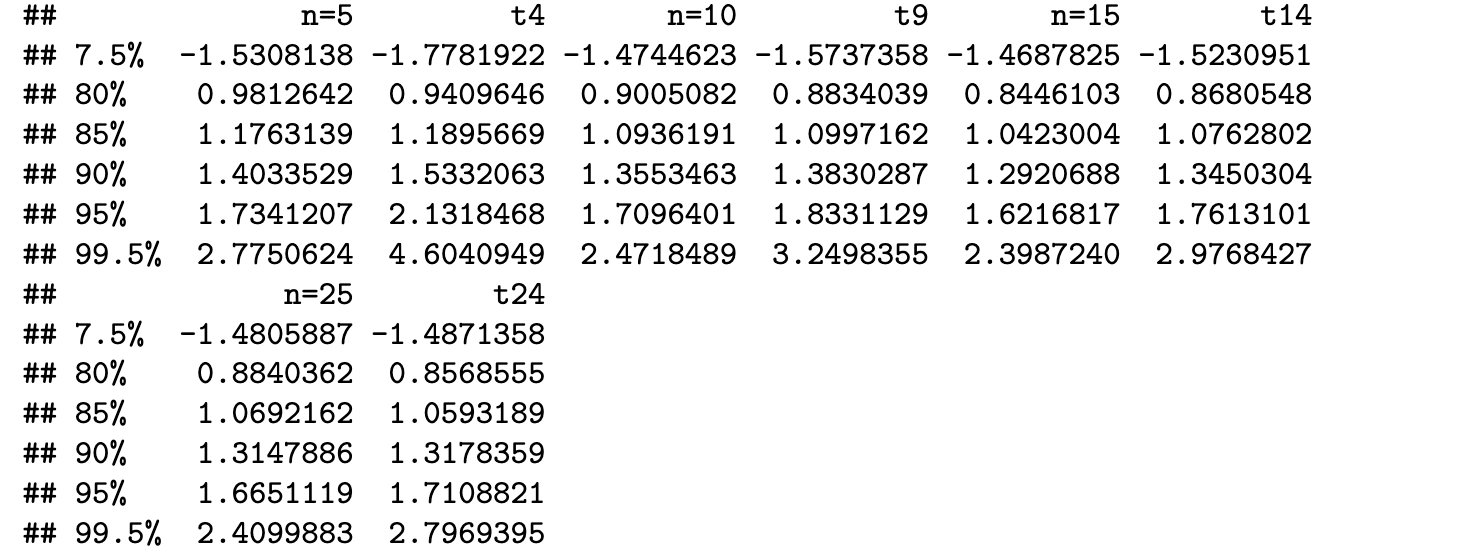
Therefore, and are not independent as and are not independent.

**Monte Carlo Simulations**

Now we are conducting Monte Carlo simulation to compute the -RV , where and . 10,000 random samples were generated from , … , for fixed = 5,10,15, and 25, each of the random samples is computing the distribution of . The following histograms are demonstrating the distribution of and the density curves of the distribution of each .



**Comparing quantiles of and distributions**



t4, t9, t14, and t24 stand for true quantiles of the distributions, respectively.

**Conclusion**

From the table above, we observe that does not follow distribution because there is a significant difference between estimated value and true value at p=0.995. Since there is a significant difference after 10000 simulations, the independence of and is both necessary and sufficient for to have a distribution. When increases, the difference of quantiles is not significant. There is a situation in which it would be reasonable to approximate the distribution of with the distribution, such as is greater than 25.

**APPENDIX**

The following is R-code used to conduct Monte Carlo simulation. What’s written after # is the explanation of the codes.

#Produce the same random values

set.seed(2022)

#Set the function

Tvalue = function(X,n){

Z <- X[[1]]

U <- (n-1)\*var(X)

t <- Z/sqrt(U/(n-1))

return(t)

}

#Do the simulation

R=10000

T5 <- rep(0,R)

T10 <- rep(0,R)

T15 <- rep(0,R)

T25 <- rep(0,R)

for (i in 1:R){

X5 <- rnorm(5,0,1)

X10 <- rnorm(10,0,1)

X15 <- rnorm(15,0,1)

X25 <- rnorm(25,0,1)

T5[i] <- Tvalue(X5,5)

T10[i] <- Tvalue(X10,10)

T15[i] <- Tvalue(X15,15)

T25[i] <- Tvalue(X25,25)

}

#Plot the histogram

x=seq(-5,5,0.01)

par(mfrow=c(2,2))

hist(T5,freq=FALSE,main="Histogram of n=5",xlab="n=5",ylim = c(0, 0.4))

curve(dt(x, 4), -6, 6, add=T,lwd = 2, ylim = c(0, 0.4))

hist(T10,freq=FALSE,main="Histogram of n=10",xlab="n=5",ylim = c(0, 0.4))

curve(dt(x, 9), -6, 6, add=T,lwd = 2, ylim = c(0, 0.4))

hist(T15,freq=FALSE,main="Histogram of n=15",xlab="n=5",ylim = c(0, 0.4))

curve(dt(x, 14), -6, 6, add=T,lwd = 2, ylim = c(0, 0.4))

hist(T25,freq=FALSE,main="Histogram of n=25",xlab="n=5",ylim = c(0, 0.4))

curve(dt(x, 24), -6, 6, add=T,lwd = 2, ylim = c(0, 0.4))

#Set the function

qvalue=function(x){

q <- c(quantile(x,0.075),quantile(x,0.8),quantile(x,0.85),quantile(x,0.9),quantile(x,0.95),quantile(x,0.995))

return(q)

}

q\_t = function(n){

q <- c(qt(0.075,n-1),qt(0.8,n-1),qt(0.85,n-1),qt(0.9,n-1),qt(0.95,n-1),qt(0.995,n-1))

return(q)

}

#Create the quantile table

q5 <- qvalue(T5)

q10 <- qvalue(T10)

q15 <- qvalue(T15)

q25 <- qvalue(T25)

q\_t5 <- q\_t(5)

q\_t10 <- q\_t(10)

q\_t15 <- q\_t(15)

q\_t25 <- q\_t(25)

q <-c('0.075','0.8','0.85','0.9','0.95','0.995')

q\_table<-data.frame(q5,q\_t5,q10,q\_t10,q15,q\_t15,q25,q\_t25)

colnames(q\_table)<-list('n=5','t4','n=10','t9','n=15','t14','n=25','t24')

print(q\_table)