# The Approximation of the Distribution by MC

Guangzi Liu

2020/3/27

## Abstract

This report considers the approximation of the distribution function $N(0,1)$ by using the Monte Carlo methods. Then use rmarkdown to produce a report which includes math equations,tables, graphs, and R code. Further, repeat the experiment 100 times and draw box plots of the 100 approximation errors at each t using ggplot2 for each n. ***# Math Equations In this part, I consider approximation of the distribution function of N(0,1):***

$$\Phi(t) = \int_{-\infty}^{t} \frac{1}{\sqrt{2\pi}} e^{(-y^2)/2} \, \mathrm{d}y, \tag{1}$$

***by the Monte Carlo methods:***

$$\hat{\Phi}(t) = \frac{1}{n} \sum_{i=1}^{n} I(X_i \le t), \tag{2}$$

***Where $X_i$ is a random sample from $N(0,1)$, $I(\cdot)$ is the indicator function. Experiment with the approximation at $n \in \{10^2, 10^3, 10^4\}$ at $t \in \{0.0, 0.67, 0.84, 1.28, 1.65, 2.32, 2.58, 3.09, 3.72\}$ to form a table.***

## Table

The approximation results are shown in the following table:

```
library(tidyverse)
t=c(0.0,0.67, 0.84,1.28,1.65,2.32,2.58,3.09,3.72)
x=pnorm( t, mean = 0, sd = 1)

n1=10^2
z1=c(rep(0,9))
w1=matrix(0,9,n1)
y1=c(rnorm(n1,mean=0,sd=1))
for(k in 1:9)
{
  for(j in 1:n1)
  {w1[k,j]=sign(y1[j]<=t[k])}
  z1[k]=sum(w1[k,])/n1}

n2=10^3
z2=c(rep(0,9))
w2=matrix(0,9,n2)
y2=c(rnorm(n2,mean=0,sd=1))
```

```
for(k in 1:9)
{
  for(j in 1:n2)
  {w2[k,j]=sign(y2[j]<=t[k])}
  z2[k]=sum(w2[k,])/n2}


n3=10^4
z3=c(rep(0,9))
w3=matrix(0,9,n3)
y3=c(rnorm(n3,mean=0,sd=1))
for(k in 1:9)
{
  for(j in 1:n3)
  {w3[k,j]=sign(y3[j]<=t[k])}
  z3[k]=sum(w3[k,])/n3}


tb<-tibble(
  t=t,
  true=x,
  '100'=z1,
  '1000'=z2,
  '10000'=z3
)
```

```
knitr::kable(head(tb), booktabs = TRUE,
             caption = 'table')
```

Table 1: table

| t | true | 100 | 1000 | 10000 |
|---|---|---|---|---|
| 0.00 | 0.5000000 | 0.58 | 0.476 | 0.5005 |
| 0.67 | 0.7485711 | 0.76 | 0.720 | 0.7448 |
| 0.84 | 0.7995458 | 0.81 | 0.777 | 0.7979 |
| 1.28 | 0.8997274 | 0.87 | 0.884 | 0.9010 |
| 1.65 | 0.9505285 | 0.92 | 0.945 | 0.9528 |
| 2.32 | 0.9898296 | 0.97 | 0.989 | 0.9908 |

# Graphs

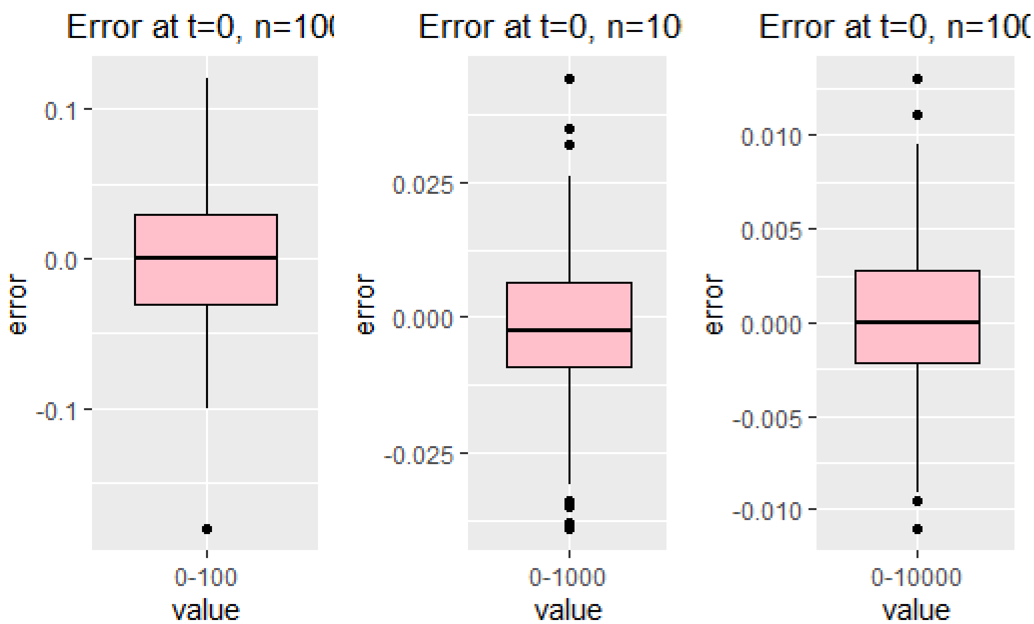
Repeat the experiment 100 times. And draw box plots of the 100 approximation errors at each $t$ using ggplot2 for each $n$. When t = 0 and n = 100,n=1000,n=10000, the graph showing the error of Monte Carlo method is as follows:

```
library(ggplot2)
library(lattice)
library(plyr)
library(Rmisc)

p1<-ggplot(data=plot_data,aes(y=V1,x=label1))+geom_boxplot(
  fill="pink",colour="black")+
  labs(title="Error at t=0, n=100",y="error",
```

```
       x="value")+theme(plot.title=element_text(size=13,hjust=0.5))
p2<-ggplot(data=plot_data,aes(y=V2,x=label2))+geom_boxplot(
  fill="pink",colour="black")+
  labs(title="Error at t=0, n=1000",y="error",
       x="value")+theme(plot.title=element_text(size=13,hjust=0.5))
p3<-ggplot(data=plot_data,aes(y=V3,x=label3))+geom_boxplot(
  fill="pink",colour="black")+
  labs(title="Error at t=0, n=10000",y="error",
       x="value")+theme(plot.title=element_text(size=13,hjust=0.5))
g1<-multiplot(p1,p2,p3,cols=3)
```

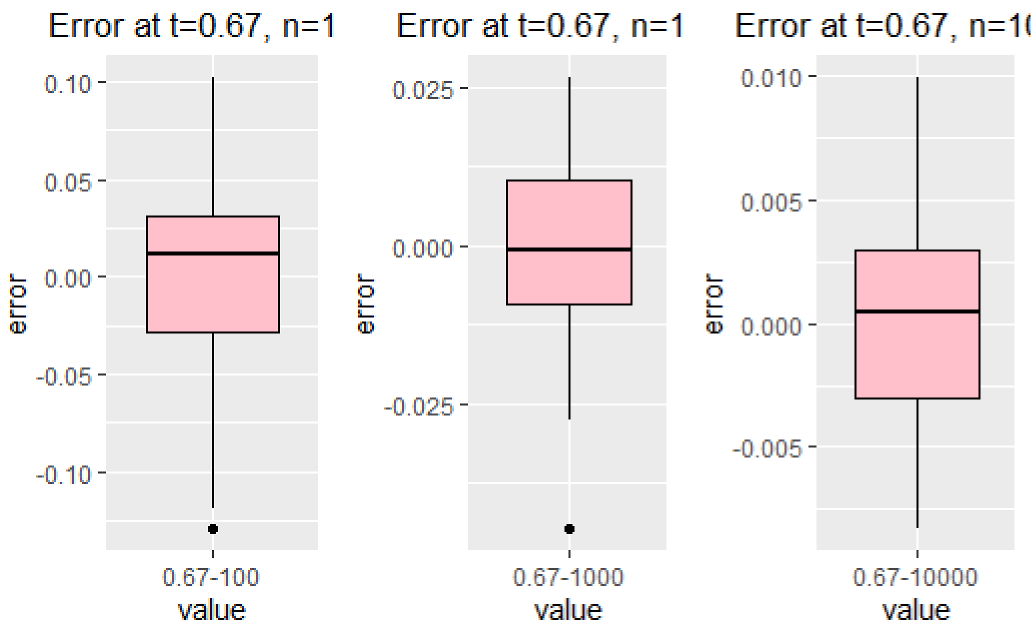
```
knitr::include_graphics("IMG/graph1.png",dpi=NA)
```




When t = 0.67 and n = 100,n=1000,n=10000, the graph showing the error of Monte Carlo method is as follows:

```
p4<-ggplot(data=plot_data,aes(y=V4,x=label4))+geom_boxplot(
  fill="pink",colour="black")+
  labs(title="Error at t=0.67, n=100",y="error",
       x="value")+theme(plot.title=element_text(size=13,hjust=0.5))
p5<-ggplot(data=plot_data,aes(y=V5,x=label5))+geom_boxplot(
  fill="pink",colour="black")+
  labs(title="Error at t=0.67, n=1000",y="error",
       x="value")+theme(plot.title=element_text(size=13,hjust=0.5))
p6<-ggplot(data=plot_data,aes(y=V6,x=label6))+geom_boxplot(
  fill="pink",colour="black")+
  labs(title="Error at t=0.67, n=10000",y="error",
       x="value")+theme(plot.title=element_text(size=13,hjust=0.5))
g2<-multiplot(p4,p5,p6,cols=3)
```

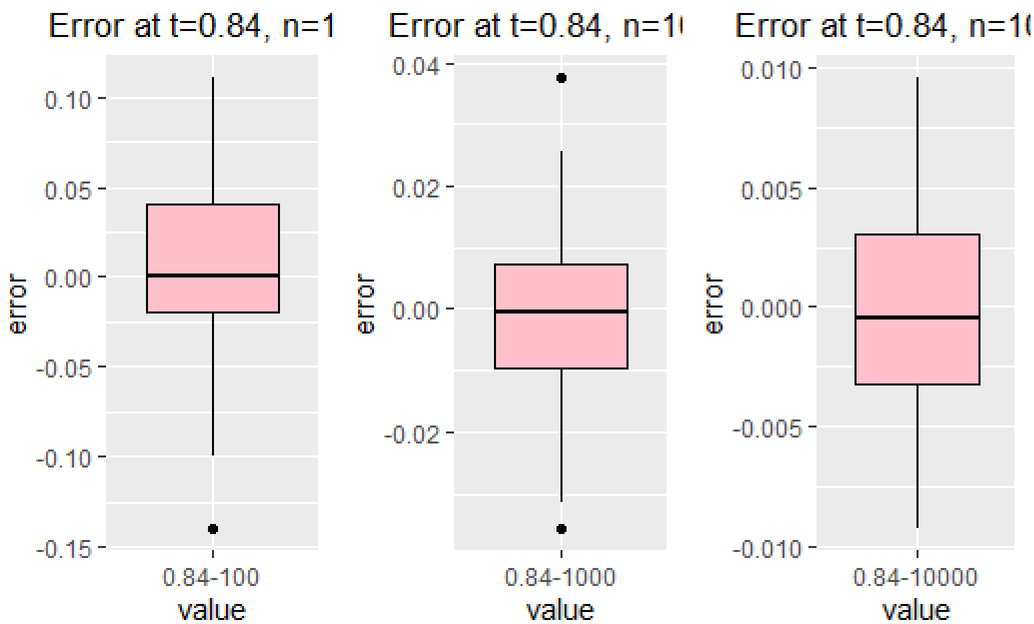
```
knitr::include_graphics("IMG/graph2.png",dpi=NA)
```




When t = 0.84 and n = 100,n=1000,n=10000, the graph showing the error of Monte Carlo method is as follows:

```
p7<-ggplot(data=plot_data,aes(y=V7,x=label7))+geom_boxplot(
  fill="pink",colour="black")+
  labs(title="Error at t=0.84, n=100",y="error",
       x="value")+theme(plot.title=element_text(size=13,hjust=0.5))
p8<-ggplot(data=plot_data,aes(y=V8,x=label8))+geom_boxplot(
  fill="pink",colour="black")+
  labs(title="Error at t=0.84, n=1000",y="error",
       x="value")+theme(plot.title=element_text(size=13,hjust=0.5))
p9<-ggplot(data=plot_data,aes(y=V9,x=label9))+geom_boxplot(
  fill="pink",colour="black")+
  labs(title="Error at t=0.84, n=10000",y="error",
       x="value")+theme(plot.title=element_text(size=13,hjust=0.5))
g3<-multiplot(p7,p8,p9,cols=3)
```

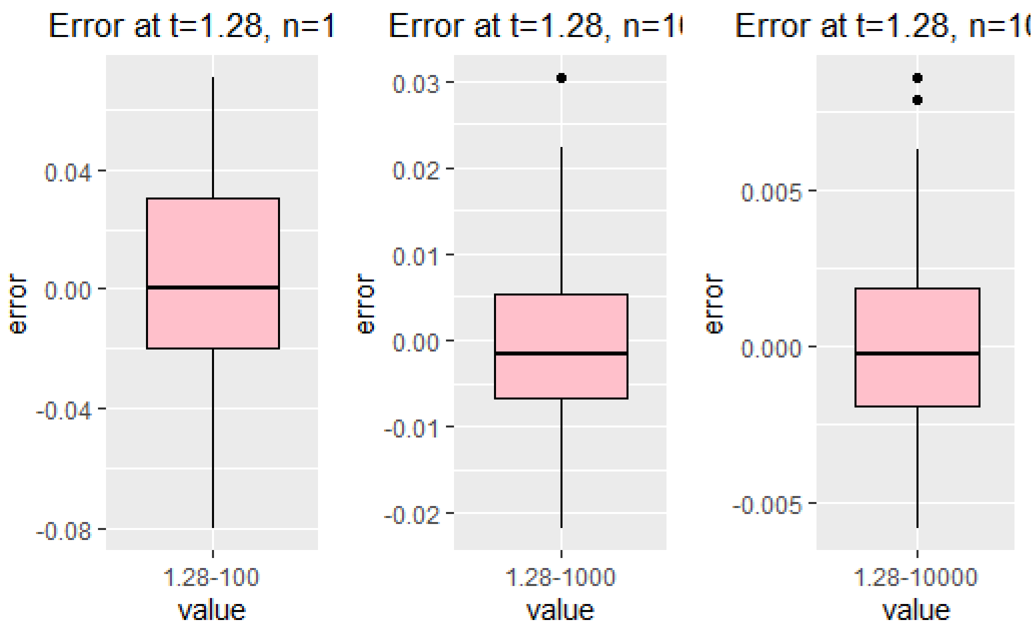
```
knitr::include_graphics("IMG/graph3.png",dpi=NA)
```




When t = 1.28 and n = 100,n=1000,n=10000, the graph showing the error of Monte Carlo method is as follows:

```
p10<-ggplot(data=plot_data,aes(y=V10,x=label10))+geom_boxplot(
  fill="pink",colour="black")+
  labs(title="Error at t=1.28, n=100",y="error",
       x="value")+theme(plot.title=element_text(size=13,hjust=0.5))
p11<-ggplot(data=plot_data,aes(y=V11,x=label11))+geom_boxplot(
  fill="pink",colour="black")+
  labs(title="Error at t=1.28, n=1000",y="error",
       x="value")+theme(plot.title=element_text(size=13,hjust=0.5))
p12<-ggplot(data=plot_data,aes(y=V12,x=label12))+geom_boxplot(
  fill="pink",colour="black")+
  labs(title="Error at t=1.28, n=10000",y="error",
       x="value")+theme(plot.title=element_text(size=13,hjust=0.5))
g4<-multiplot(p10,p11,p12,cols=3)
```

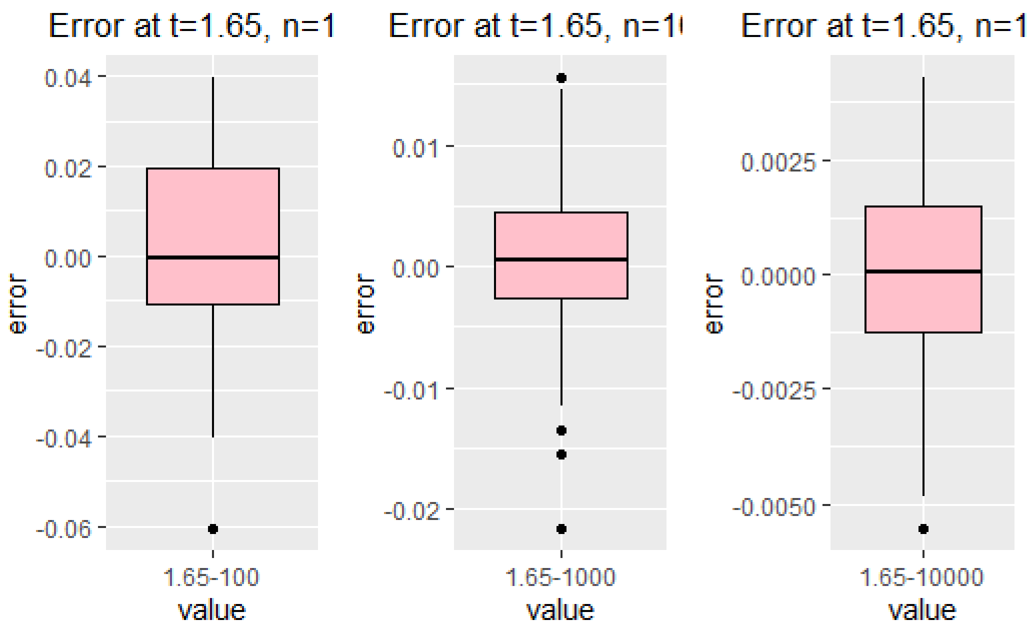
```
knitr::include_graphics("IMG/graph4.png",dpi=NA)
```




When t = 1.65 and n = 100,n=1000,n=10000, the graph showing the error of Monte Carlo method is as follows:

```
p13<-ggplot(data=plot_data,aes(y=V13,x=label13))+geom_boxplot(
  fill="pink",colour="black")+
  labs(title="Error at t=1.65, n=100",y="error",
       x="value")+theme(plot.title=element_text(size=13,hjust=0.5))
p14<-ggplot(data=plot_data,aes(y=V14,x=label14))+geom_boxplot(
  fill="pink",colour="black")+
  labs(title="Error at t=1.65, n=1000",y="error",
       x="value")+theme(plot.title=element_text(size=13,hjust=0.5))
p15<-ggplot(data=plot_data,aes(y=V15,x=label15))+geom_boxplot(
  fill="pink",colour="black")+
  labs(title="Error at t=1.65, n=10000",y="error",
       x="value")+theme(plot.title=element_text(size=13,hjust=0.5))
g5<-multiplot(p13,p14,p15,cols=3)
```

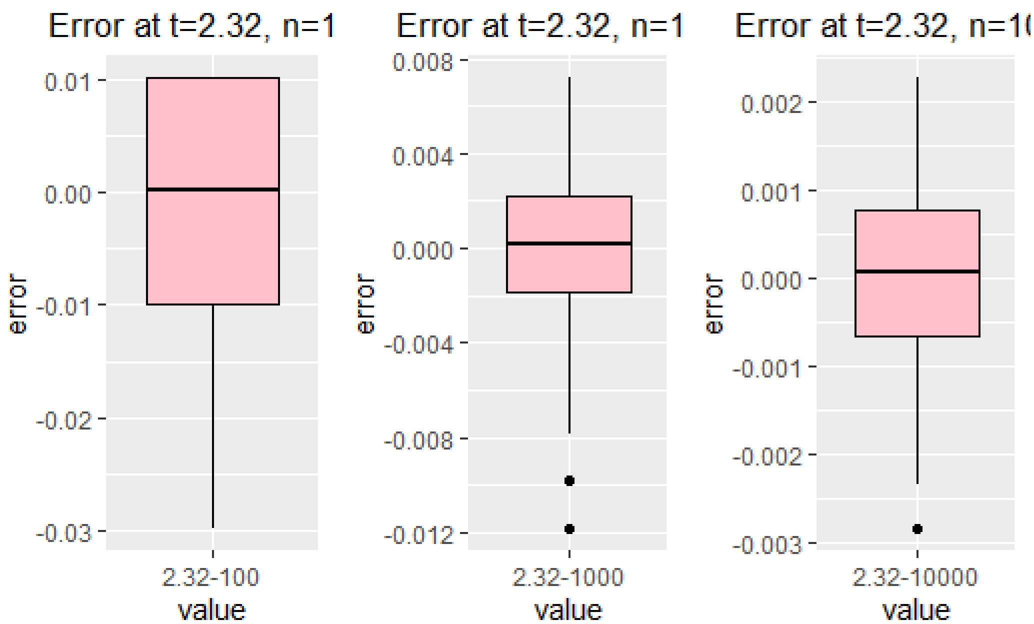
```
knitr::include_graphics("IMG/graph5.png",dpi=NA)
```




When t = 2.32 and n = 100,n=1000,n=10000, the graph showing the error of Monte Carlo method is as follows:

```
p16<-ggplot(data=plot_data,aes(y=V16,x=label16))+geom_boxplot(
  fill="pink",colour="black")+
  labs(title="Error at t=2.32, n=100",y="error",
       x="value")+theme(plot.title=element_text(size=13,hjust=0.5))
p17<-ggplot(data=plot_data,aes(y=V17,x=label17))+geom_boxplot(
  fill="pink",colour="black")+
  labs(title="Error at t=2.32, n=1000",y="error",
       x="value")+theme(plot.title=element_text(size=13,hjust=0.5))
p18<-ggplot(data=plot_data,aes(y=V18,x=label18))+geom_boxplot(
  fill="pink",colour="black")+
  labs(title="Error at t=2.32, n=10000",y="error",
       x="value")+theme(plot.title=element_text(size=13,hjust=0.5))
g6<-multiplot(p16,p17,p18,cols=3)
```

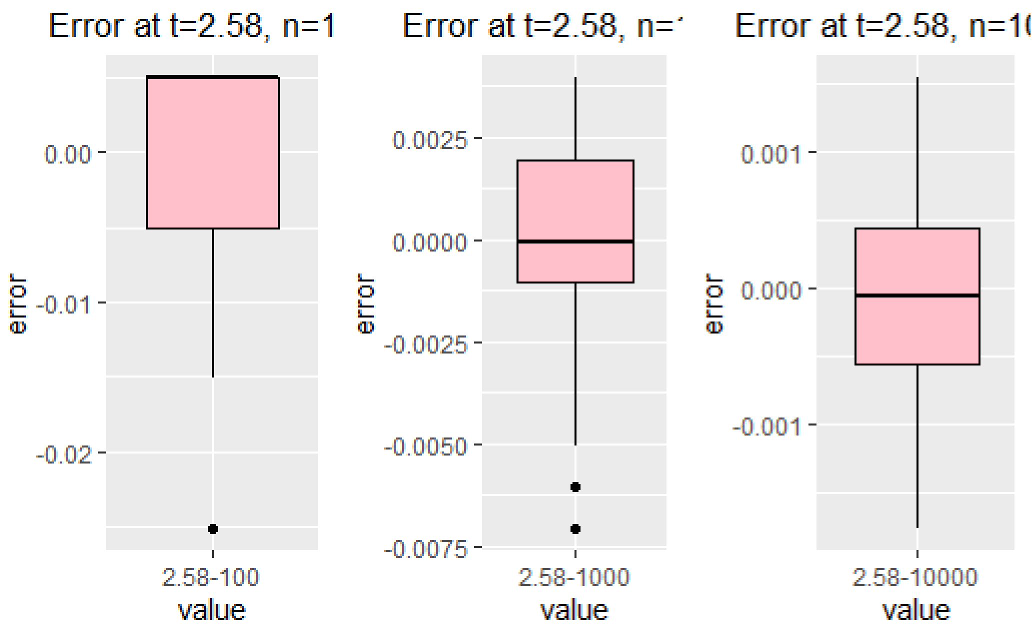
```
knitr::include_graphics("IMG/graph6.png",dpi=NA)
```




When t = 2.58 and n = 100,n=1000,n=10000, the graph showing the error of Monte Carlo method is as follows:

```
p19<-ggplot(data=plot_data,aes(y=V19,x=label19))+geom_boxplot(
  fill="pink",colour="black")+
  labs(title="Error at t=2.58, n=100",y="error",
       x="value")+theme(plot.title=element_text(size=13,hjust=0.5))
p20<-ggplot(data=plot_data,aes(y=V20,x=label20))+geom_boxplot(
  fill="pink",colour="black")+
  labs(title="Error at t=2.58, n=1000",y="error",
       x="value")+theme(plot.title=element_text(size=13,hjust=0.5))
p21<-ggplot(data=plot_data,aes(y=V21,x=label21))+geom_boxplot(
  fill="pink",colour="black")+
  labs(title="Error at t=2.58, n=10000",y="error",
       x="value")+theme(plot.title=element_text(size=13,hjust=0.5))
g7<-multiplot(p19,p20,p21,cols=3)
```

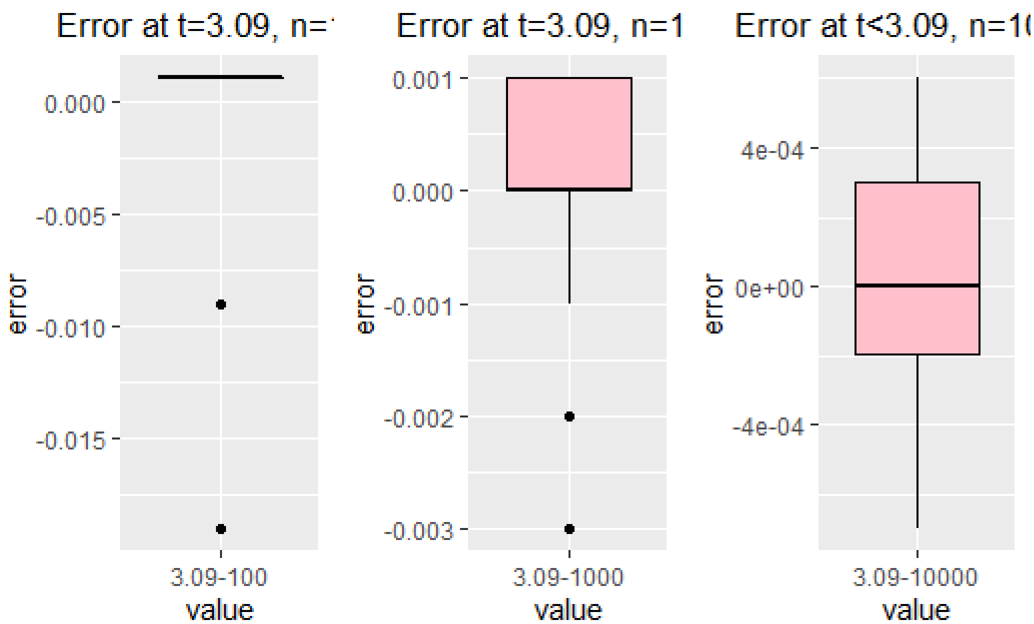
```
knitr::include_graphics("IMG/graph7.png",dpi=NA)
```




When t = 3.09 and n = 100,n=1000,n=10000, the graph showing the error of Monte Carlo method is as follows:

```
p22<-ggplot(data=plot_data,aes(y=V22,x=label22))+geom_boxplot(
  fill="pink",colour="black")+
  labs(title="Error at t=3.09, n=100",y="error",
       x="value")+theme(plot.title=element_text(size=13,hjust=0.5))
p23<-ggplot(data=plot_data,aes(y=V23,x=label23))+geom_boxplot(
  fill="pink",colour="black")+
  labs(title="Error at t=3.09, n=1000",y="error",
       x="value")+theme(plot.title=element_text(size=13,hjust=0.5))
p24<-ggplot(data=plot_data,aes(y=V24,x=label24))+geom_boxplot(
  fill="pink",colour="black")+
  labs(title="Error at t<3.09, n=10000",y="error",
       x="value")+theme(plot.title=element_text(size=13,hjust=0.5))
g8<-multiplot(p22,p23,p24,cols=3)
```

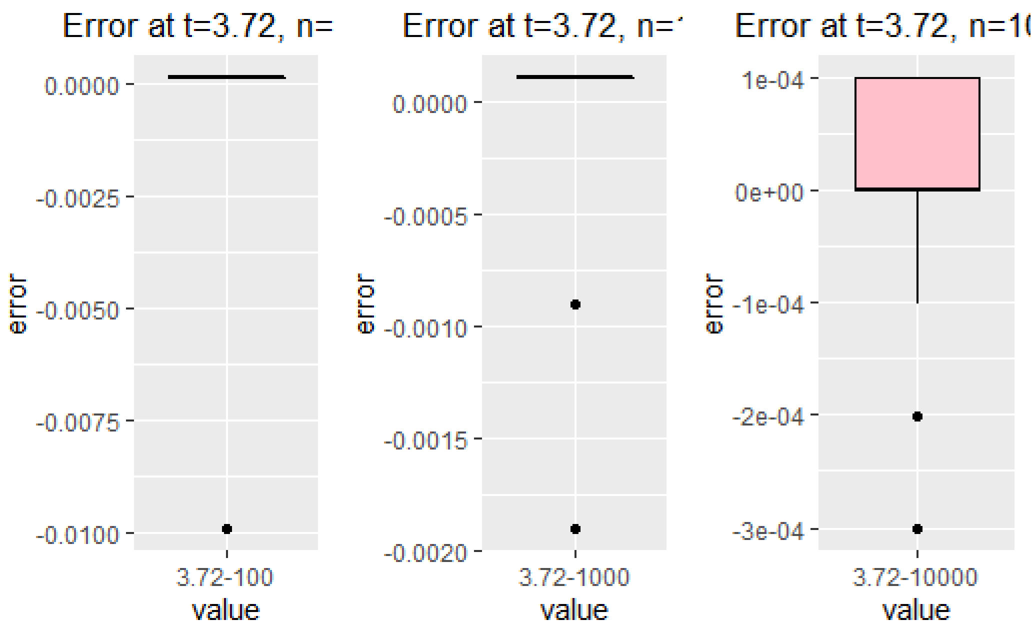
```
knitr::include_graphics("IMG/graph8.png",dpi=NA)
```




When t = 3.72 and n = 100,n=1000,n=10000, the graph showing the error of Monte Carlo method is as follows:

```
p25<-ggplot(data=plot_data,aes(y=V25,x=label25))+geom_boxplot(
  fill="pink",colour="black")+
  labs(title="Error at t=3.72, n=100",y="error",
       x="value")+theme(plot.title=element_text(size=13,hjust=0.5))
p26<-ggplot(data=plot_data,aes(y=V26,x=label26))+geom_boxplot(
  fill="pink",colour="black")+
  labs(title="Error at t=3.72, n=1000",y="error",
       x="value")+theme(plot.title=element_text(size=13,hjust=0.5))
p27<-ggplot(data=plot_data,aes(y=V27,x=label27))+geom_boxplot(
  fill="pink",colour="black")+
  labs(title="Error at t=3.72, n=10000",y="error",
       x="value")+theme(plot.title=element_text(size=13,hjust=0.5))
g9<-multiplot(p25,p26,p27,cols=3)
```

```
knitr::include_graphics("IMG/graph9.png",dpi=NA)
```




## Conclusion

When we consider the approximation of the distribution function $N(0, 1)$ by using the Monte Carlo methods, we can get a conclusion from the above boxplots, the more times the experiment is, the smaller error is, the closer the true value is.