Approximation of the distribution

liangqingqing

2020/3/26

##1.Introduction This report uses **rmarkdown** to produce .The report is about approximation of the distribution function of $N(0,1)$ by the Monte Carlo methods, and draw box plots of the 100 approximation errors at each t using **ggplot2** [@R-ggplot2] for each n.

##2.Math Equations Consider approximation of the distribution function of $N(0,1)$,

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

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's are a random sample from $N(0,1)$, and $I(\cdot)$ 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.

##3.The comparison between the experimental value and the true value 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,value1=z1,value2=z2,value3=z3)

##4.Repeat the experiment 100 times 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=100 z=matrix(0,100,9) w=matrix(0,9,n1) for(p in 1:100) { y=c(rnorm(n1,mean=0,sd=1)) for(k in 1:9) { for(j in 1:n1) {w[k,j]=sign(y[j]<=t[k])} z[p,k]=sum(w[k,])/n1}} z=as.data.frame(z) e100_1=z$V1-x[1]e100_2=z$V2-x[2] e100_3=z$V3-x[3]e100_4=z$V4-x[4] e100_5=z$V5-x[5]e100_6=z$V6-x[6] e100_7=z$V7-x[7]e100_8=z$V8-x[8] e100_9=z$V9-x[9]

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) n2=10^3 z=matrix(0,100,9) w=matrix(0,9,n2) for(p in 1:100) { y=c(rnorm(n2,mean=0,sd=1)) for(k in 1:9) { for(j in 1:n2) {w[k,j]=sign(y[j]<=t[k])} z[p,k]=sum(w[k,])/n2}} z=as.data.frame(z) e1000_1=z$V1-x[1]e1000_2=z$V2-x[2] e1000_3=z$V3-x[3]e1000_4=z$V4-x[4] e1000_5=z$V5-x[5]e1000_6=z$V6-x[6] e1000_7=z$V7-x[7]e1000_8=z$V8-x[8] e1000_9=z$V9-x[9]

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) n3=10000 z=matrix(0,100,9) w=matrix(0,9,n3) for(p in 1:100) { y=c(rnorm(n3,mean=0,sd=1)) for(k in 1:9) { for(j in 1:n3) {w[k,j]=sign(y[j]<=t[k])} z[p,k]=sum(w[k,])/n3}} z=as.data.frame(z) e10000_1=z$V1 - x[1]e10000_2 = z$V2-x[2] e10000_3=z$V3 - x[3]e10000_4 = z$V4-x[4] e10000_5=z$V5 - x[5]e10000_6 = z$V6-x[6] e10000_7=z$V7 - x[7]e10000_8 = z$V8-x[8] e10000_9=z$V9-x[9]

A=cbind.data.frame(e100_1,e100_2,e100_3,e100_4,e100_5,e100_6,e100_7,e100_8,e100_9,e1000_1,e1000_2,e1000_3,e100

library(ggplot2) plot1<-ggplot(data=A,aes(y=e100_1,x="0-100"))+geom_boxplot()+ labs(title="Boxplot of error at t=0, n=100",y="error", x=NULL) plot1

plot2<-ggplot(data=A,aes(y=e100_2,x="0.67-100"))+geom_boxplot()+ labs(title="Boxplot of error at t=0.67, n=100",y="error", x=NULL) plot2

plot3<-ggplot(data=A,aes(y=e100_3,x="0.84-100"))+geom_boxplot()+ labs(title="Boxplot of error at t=0.84, n=100",y="error", x=NULL) plot3

plot4<-ggplot(data=A,aes(y=e100_4,x="1.28-100"))+geom_boxplot()+ labs(title="Boxplot of error at t=1.28, n=100",y="error", x=NULL) plot4

plot5<-ggplot(data=A,aes(y=e100_5,x="1.65-100"))+geom_boxplot()+ labs(title="Boxplot of error at t=1.65, n=100",y="error", x=NULL) plot5

plot6<-ggplot(data=A,aes(y=e100_6,x="2.32-100"))+geom_boxplot()+ labs(title="Boxplot of error at t=2.32, n=100",y="error", x=NULL) plot6

plot7<-ggplot(data=A,aes(y=e100_7,x="2.58-100"))+geom_boxplot()+ labs(title="Boxplot of error at t=2.58, n=100",y="error", x=NULL) plot7

plot8<-ggplot(data=A,aes(y=e100_8,x="3.09-100"))+geom_boxplot()+ labs(title="Boxplot of error at t=3.09, n=100",y="error", x=NULL) plot8

plot9<-ggplot(data=A,aes(y=e100_9,x="3.72-100"))+geom_boxplot()+ labs(title="Boxplot of error at t=3.72, n=100",y="error", x=NULL) plot9

plot10<-ggplot(data=A,aes(y=e1000_1,x="0-100"))+geom_boxplot()+ labs(title="Boxplot of error at t=0, n=100",y="error", x=NULL) plot10

plot11<-ggplot(data=A,aes(y=e1000_2,x="0.67-100"))+geom_boxplot()+ labs(title="Boxplot of error at t=0.67, n=100",y="error", x=NULL) plot11

plot12<-ggplot(data=A,aes(y=e1000_3,x="0.84-100"))+geom_boxplot()+ labs(title="Boxplot of error at t=0.84, n=100",y="error", x=NULL) plot12

plot13<-ggplot(data=A,aes(y=e1000_4,x="1.28-100"))+geom_boxplot()+ labs(title="Boxplot of error at t=1.28, n=100",y="error", x=NULL) plot13

plot14<-ggplot(data=A,aes(y=e1000_5,x="1.65-100"))+geom_boxplot()+ labs(title="Boxplot of error at t=1.65, n=100",y="error", x=NULL) plot14

plot15<-ggplot(data=A,aes(y=e100_6,x="2.32-100"))+geom_boxplot()+ labs(title="Boxplot of error at t=2.32, n=100",y="error", x=NULL) plot15

plot16<-ggplot(data=A,aes(y=e1000_7,x="2.58-100"))+geom_boxplot()+ labs(title="Boxplot of error at t=2.58, n=100",y="error", x=NULL) plot16

plot17<-ggplot(data=A,aes(y=e1000_8,x="3.09-100"))+geom_boxplot()+ labs(title="Boxplot of error at t=3.09, n=100",y="error", x=NULL) plot17

plot18<-ggplot(data=A,aes(y=e1000_9,x="3.72-100"))+geom_boxplot()+ labs(title="Boxplot of error at t=3.72, n=100",y="error", x=NULL) plot18

plot19<-ggplot(data=A,aes(y=e10000_1,x="0-100"))+geom_boxplot()+ labs(title="Boxplot of error at t=0, n=100",y="error", x=NULL) plot19 plot20<-ggplot(data=A,aes(y=e10000_2,x="0.67-100"))+geom_boxplot()+ labs(title="Boxplot of error at t=0.67, n=100",y="error", x=NULL) plot20

plot21<-ggplot(data=A,aes(y=e10000_3,x="0.84-100"))+geom_boxplot()+ labs(title="Boxplot of error at t=0.84, n=100",y="error", x=NULL) plot21

plot22<-ggplot(data=A,aes(y=e10000_4,x="1.28-100"))+geom_boxplot()+ labs(title="Boxplot of error at t=1.28, n=100",y="error", x=NULL) plot22

plot23<-ggplot(data=A,aes(y=e10000_5,x="1.65-100"))+geom_boxplot()+ labs(title="Boxplot of error at t=1.65, n=100",y="error", x=NULL) plot23

plot24<-ggplot(data=A,aes(y=e10000_6,x="2.32-100"))+geom_boxplot()+ labs(title="Boxplot of error at t=2.32, n=100",y="error", x=NULL) plot24

plot25<-ggplot(data=A,aes(y=e10000_7,x="2.58-100"))+geom_boxplot()+ labs(title="Boxplot of error at t=2.58, n=100",y="error", x=NULL) plot25

plot26<-ggplot(data=A,aes(y=e10000_8,x="3.09-100"))+geom_boxplot()+ labs(title="Boxplot of error at t=3.09, n=100",y="error", x=NULL) plot26

plot27<-ggplot(data=A,aes(y=e10000_9,x="3.72-100"))+geom_boxplot()+ labs(title="Boxplot of error at t=3.72, n=100",y="error", x=NULL) plot27

##5.Summary In this report, I formed a table that shows the difference between approximate value and true value, and draw box plots of the 100 approximation errors at each t for each n.From the plots,we can see that the approximate value becomes smaller when n increases.