Monte Carlo Simulation with R

Han Ruixin

2020/3/25

Abstract

In this paper, Monte Carlo method is used to calculate the approximate value of the standard normal distribution. At the same time, in the given array, the error of the method is represented by the box graph.

Math Equations

Consider approximation of the distribution function of $N(0,1)$,

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

by the Monte Carlo methods:

$$\hat{\Phi}(t) = \frac{1}{n} \sum_{i=1}^{n} I(X_i \leq t)$$

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

Packages

The following packages were used in this test:"ggplot2" "lattice" "plyr" "Rmisc".

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

True value

The table includes the true value for comparison.

	t=0	t=0.67	t=0.84	t=1.28	t=1.65	t=2.32	t=2.58	t=3.09	t=3.72
n=100	0.5000	0.8000	0.8300	0.9100	0.9800	1.0000	1.0000	1.0000	1.0000
n=1000	0.5390	0.7460	0.8150	0.9040	0.9420	0.9910	0.9940	0.9990	1.0000
n=10000	0.4911	0.7518	0.7963	0.8986	0.9489	0.9903	0.9946	0.9992	1.0000
True Value	0.5000	0.7486	0.7995	0.8997	0.9505	0.9898	0.9951	0.9990	0.9999

Monte Carlo Simulation

In the case of $n \in \{10^2, 10^3, 10^4\}$, $t \in \{0.0, 0.67, 0.84, 1.28, 1.65, 2.32, 2.58, 3.09, 3.72\}$, the approximate values of the standard normal distribution are calculated by Monte Carlo method. Take n = 100 as an example.

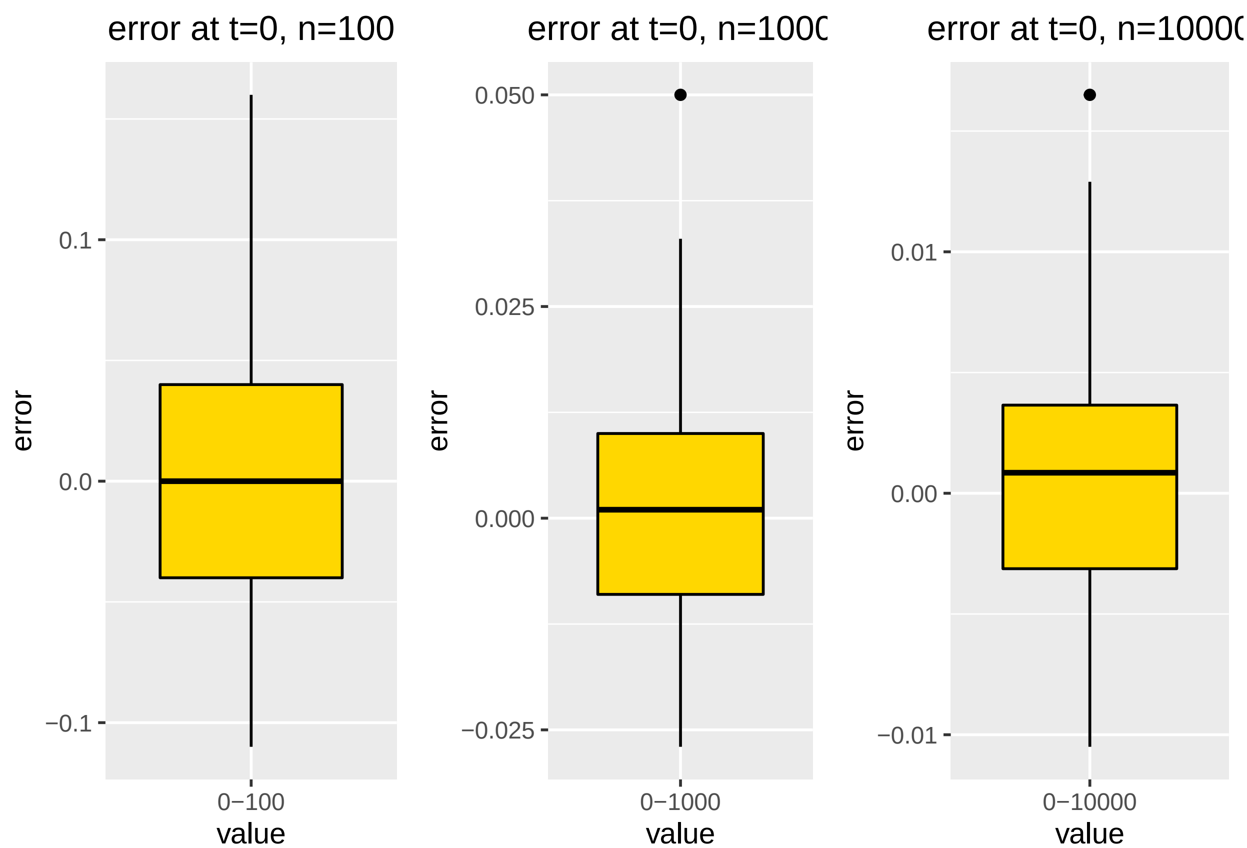
First, we define a matrix of three rows and nine columns as p.Then generate 100 random numbers that obey the standard normal distribution, and record them as num. According to the Monte Carlo formula, an indicator function is defined, and 100 random numbers in the matrix are compared with ti. If num < ti,it means true, 1 point will be recorded; otherwise, it means false, 0 point will be recorded. Take the mean value of the final score and record it as p [i, j].

Box plots

Use ggplot2 to draw the box line diagram, the horizontal axis represents value, and the vertical axis represents error. When t = 0 and n = 100,n=1000,n=10000, the box diagram showing the error of Monte Carlo method is as follows:

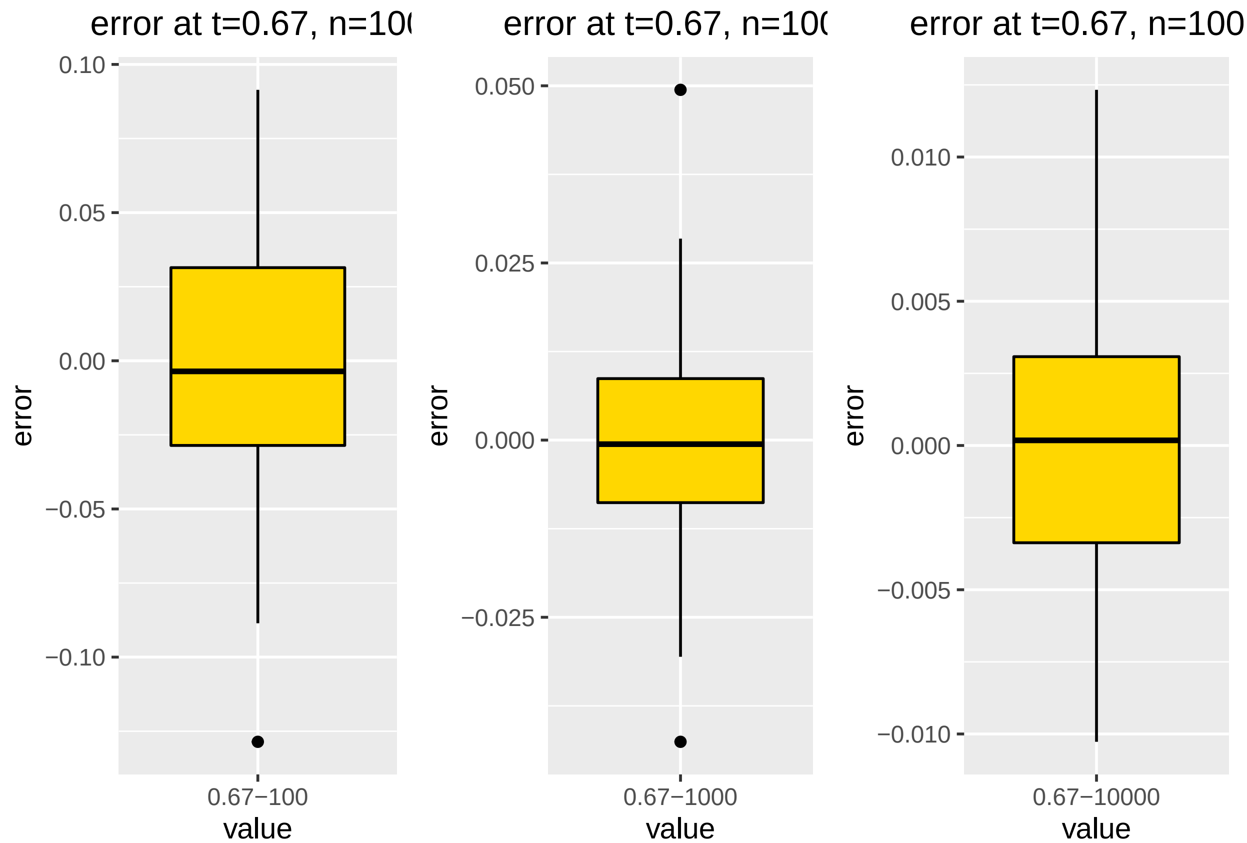
```
p1<-ggplot(data=plot_data,aes(y=V1,x=label1))+geom_boxplot(
  fill="gold",colour="black")+
  labs(title="error at t=0, n=100",y="error",
       x="value")+theme(plot.title=element_text(size=13,hjust=0.5))
p2<-ggplot(data=plot_data,aes(y=V2,x=label2))+geom_boxplot(
  fill="gold",colour="black")+
  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="gold",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)
```



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

```
p4<-ggplot(data=plot_data,aes(y=V4,x=label4))+geom_boxplot(
  fill="gold",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="gold",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="gold",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)
```



From these figures, we can see that the more calculations are made, the greater the probability of approaching future events when the model contains the largest amount of known data (according to the law of large numbers).

Summary and Discussion

When Monte Carlo method is used to find the approximate value of the function obeying the standard normal distribution, the more times the calculation is, the greater the probability of approaching future events is.