Monte Carlo Method for Solving Definite Integrals

Anyi Rao

141180092

(2014', School of Electronic Science and Engineering, Nanjing University, Nanjing 210023)

May 17th 2016

Abstract: Monte Carlo methods (or Monte Carlo experiments) are a broad class of computational algorithms that rely on repeated random sampling to obtain numerical results. By the law of large numbers, definite integrals described by the expected value of some random variable can be approximated by taking the empirical mean (a.k.a. the sample mean) of independent samples of the variable. This paper compares it with numerical methods and considers to improve it. The appendixes supplement the understanding and the verification of central limit theorem.

Keywords: Monte Carlo method; numerical method; the law of large numbers

Contents

1 Theoretical Analyses

1.1 Monte Carlo method

Choose n data casually in [a,b] Obviously,

$$E[g(X)] = \frac{1}{b-a}\int_a^b g(x)\,dx \tag{1}$$

According to the law of large numbers,

$$\lim_{n\to+\infty} P\{|\frac{1}{n}\sum_{i=1}^{n} g(X_i) - \frac{1}{b-a}\int_a^b g(x)\,dx < \varepsilon|\} = 1 \tag{2}$$

when n$\to\infty$

$$\int_a^b g(x)\,dx \approx \frac{b-a}{n}\sum_{i=1}^{n} g(X_i) \tag{3}$$

1.2 Numerical methods

Trapziod method The area of the curved side trapzoid can be estimated as straight side trapzoid.

$$S_i \approx \frac{y_{i-1}+y_i}{2}\triangle x_i \tag{4}$$

Thus

$$\begin{aligned} S &= \int_a^b f(x)\,dx \\ &= \sum_{i=1}^{n} S_i \\ &\approx \sum_{i=1}^{n} \frac{y_{i-1}+y_i}{2}\triangle x_i \end{aligned} \tag{5}$$

Parabolic method Divide the region into 2n points equivalently.

$$h_1 = \frac{b-a}{2n} \tag{6}$$

Thus

$$\begin{aligned} \int_a^b f(x)\,dx &= \sum_{i=1}^{n}\int_{x_{2i-1}}^{x_{2i}} f(x)\,dx \\ &= \frac{b-a}{6n}[f(a)+f(b)+4\sum_{i=1}^{n/2} f(x_{2i-1}) + 2\sum_{i=1}^{n/2-1} f(x_{2i})] \\ &\approx \frac{b-a}{6n}[y_0 + y_{2n} + 4(y_1+y_3+\cdots) + 2(y_2+y_4+\cdots)] \end{aligned} \tag{7}$$

The equation. 7 is also called Simpson formula.

2 Experiment

2.1 Monte Carlo method

Using C language to simulate the Monte Carlo method. The main function sums the function values whose variables coming from random generation function, which is shown in figure.1. The random generation function uses mapping relationship to the interval, which is shown in figure.2.

```
#include<stdio.h>
  double uniform(double,double,long int *);
  int main()
  {
        long unsigned int n;
        double a,b,x,sum=0,integral;
        int i,j;
        long int s;
        printf("请输入积分下限a：");
        scanf("%lf",&a);
        printf("请输入积分上限b：");
        scanf("%lf",&b);
        printf("请输入取随机数的个数n：");
        scanf("%lu",&n);
        s=13579;
        for(i=0;i<n;i++)
         {
                x=uniform(a,b,&s);
                sum+=x*x;

          }
        integral=(b-a)/(float)(n)*sum;
        printf("函数y=x^2在%lf到%lf之间积分的值为%lf\n",a,b,integral);
 }
```

Figure 1: Main function

```
double uniform(double a,double b ,long int *seed)
{
        double t;
        *seed=2045*(*seed)+1;
        *seed=*seed-(*seed/1048576)*1048576;
        t=(*seed)/1048576.0;
        t=a+(b-a)*t;
        return (t);
}
```

Figure 2: random generation function

2.2 Numerical methods

Using MATLAB's trapziod and parabolic numerical calculating function to simulate the numerical methods. These function is inherent in MATLAB's function package.

(a) trapziod (b) paracurve

Figure 3: Numerical calculation

2.3 Results and Analyses

In order to prove the validity of the simulation of the methods above, it uses three integrals to verify. The equations are shown in below, where equation. 8 can be solved in mathmathetic way with accurate value $\frac{1}{3}$, while equation. 9. can't.

$$y = \int_0^1 x^2 \, dx \tag{8}$$

$$y = \int_0^1 \sqrt{sin(x)} \, dx \tag{9}$$

$$y = \int_0^1 x^x \, dx \tag{10}$$

Change the partition number in the interval [0,1], and get the table. 1.2.3 corresponding to the function figure.

Since $x^2$is a para curve inherently, the para-curve estimation is accurate absolutely. As the partition number enlarge, the results converge to to a certain value. The parabolic method convergence rate is the fast, while trapezoid method secondly, Monte Carlo method thirdly. From the function figure, it's seen that the integrand(function) is smooth. Accordingly, it's no wonder to get the result of convergence rate above.

Partition number	10	100	1000	10000	100000
Trapz	0.3350	0.3333	0.3333	0.3333	0.3333
Para	0.3333	0.3333	0.3333	0.3333	0.3333
MonteCarlo	0.596604	0.337846	0.349337	0.334427	0.333397

Table 1: The calculation of x^2

3 Transcendental Integrals

Transcendental Integrals are those integrals that can't be solved in general way. Such as,

$$\int_0^1 sin(\frac{1}{x}) \, dx \tag{11}$$

Partition number	10	100	1000	10000	100000
Trapz	0.6366	0.6428	0.6430	0.6430	0.6430
Para	0.6417	0.6417	0.6417	0.6417	0.6417
Monte Carlo	0.824023	0.647930	0.654688	0.644764	0.643579

Table 2: The calculation of $\sqrt{sin(x)}$

Partition number	10	100	1000	10000	100000
Trapz	0.7877	0.7835	0.7834	0.7834	0.7834
Para	0.7837	0.7837	0.7837	0.7837	0.7837
Monte Carlo	0.825816	0.783326	0.785504	0.783072	0.783109

Table 3: The calculation of x^x

$$\int_0^2 \frac{x}{lnx}\,dx \tag{12}$$

$$\int_0^1 \frac{1}{\sqrt{2\pi}} e^{\frac{-x^2}{2}}\,dx \tag{13}$$

$$\int_0^1 \frac{1}{\sqrt{2\pi}} x e^{\frac{-x^2}{2}}\,dx \tag{14}$$

$$\int_0^1 \frac{sinx}{x}\,dx \tag{15}$$

$$\int_0^1 sin(x^2)\,dx \tag{16}$$

$$\int_0^1 (sinx)^{0.5}\,dx \tag{17}$$

$$\int_0^1 \frac{1}{\sqrt{1+x^4}}\,dx \tag{18}$$

The corresponding function figure is shown in figure.7.

Notice: the equation.12 diverges at x=1, the integral value is infinity according to the Calculus. Nonetheless, the Monte Carlo method gets a finite value. It's a fault and will be discussed in the section Deficiencies analysis.

Then, it uses Monte Carlo method and numerical methods to calculate them. Set the partition points as 100 thousand(1×10^5) to ensure the precision. The result is shown in table.4.

4 Double Integral

The Monte Carlo method can be also applied in double integral, such as,

$$\int_{-8}^{8}\int_{-8}^{8} \frac{sin\sqrt{x^2+y^2}}{\sqrt{x^2+y^2}}\,dxdy \tag{19}$$

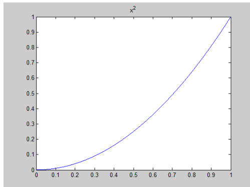


Figure 4: x^2

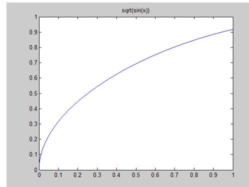


Figure 5: $\sqrt{sin(x)}$

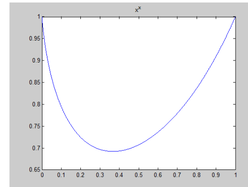


Figure 6: x^x

f(x)	Trapz	Para	Monte Carlo
$\int_0^1 sin(\frac{1}{x})\,dx$	NaN	0.5503	0.506946
$\int_0^2 \frac{x}{lnx}\,dx$	inf	inf	4.455943
$\int_0^1 \frac{1}{\sqrt{2\pi}}e^{\frac{-x^2}{2}}\,dx$	0.3413	0.3413	0.341316
$\int_0^1 \frac{1}{\sqrt{2\pi}}xe^{\frac{-x^2}{2}}\,dx$	0.1570	0.1570	0.157194
$\int_0^1 \frac{sinx}{x}\,dx$	NaN	0.9416	0.941326
$\int_0^1 sin(x^2)\,dx$	0.3103	0.3103	0.319316
$\int_0^1 (sinx)^{0.5}\,dx$	0.6430	0.6417	0.641316
$\int_0^1 \frac{1}{\sqrt{1+x^4}}\,dx$	0.9270	0.9270	0.902816

Table 4: Values of Transcendental Integrals

The integral is shown in figure. 8 Comparing the value of equation.19 got from Monte Carlo method and MATLAB DoublePara function, which is shown in table.5. The sampling points of both method is 1 million(1×10^6), cause the integral interval is 100 times bigger than the previous one.

Double integral f(x)	MATLAB DoublePara	MonteCarlo	Relative Error
$\int_{-8}^{8}\int_{-8}^{8} \frac{sin\sqrt{x^2+y^2}}{\sqrt{x^2+y^2}}\,dxdy$	9.5936	9.6946	0.010

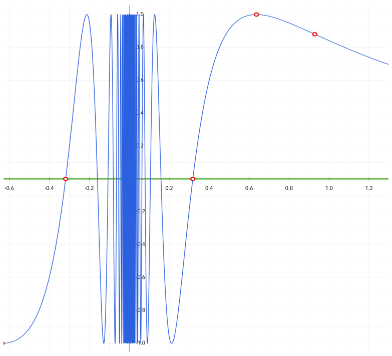
Table 5: Result of Double integral

5 Error and Deficiencies Analyses

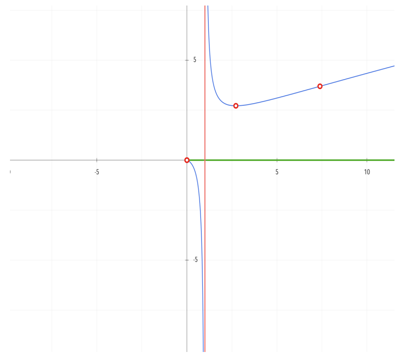
Taking $\int_0^1 \frac{4}{1+x^2}\,dx$, the accurate value is 3.1416. Comparing the results getting from different methods with the accurate value. Then calculate the relative error.

Reasons of error

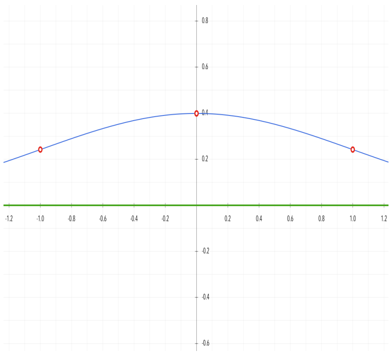
1. Only when n goes to infinity, can Monte Carlo method takes advantage of the law of large numbers to get accurate integration results. However in the program, n only can take finite number. Therefore this will produce error.



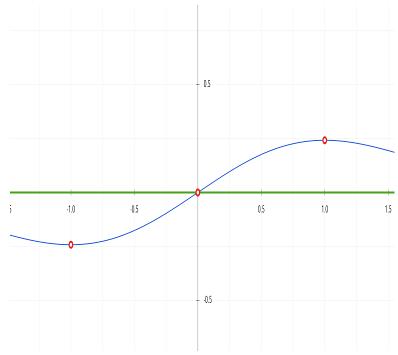
(a) $\int_0^1 sin(\frac{1}{x})\,dx$



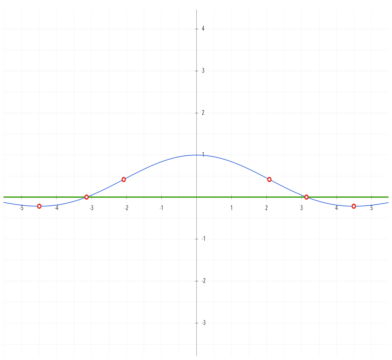
(b) $\int_0^2 \frac{x}{lnx}\,dx$



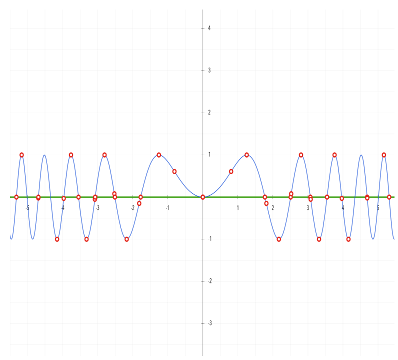
(c) $\int_0^1 \frac{1}{\sqrt{2\pi}}e^{\frac{-x^2}{2}}\,dx$



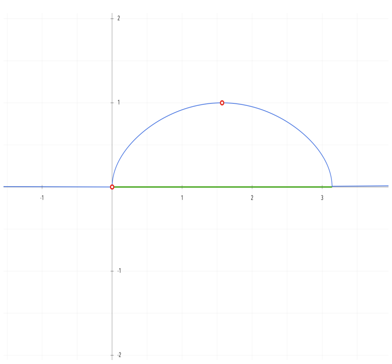
(d) $\int_0^1 \frac{1}{\sqrt{2\pi}}xe^{\frac{-x^2}{2}}\,dx$



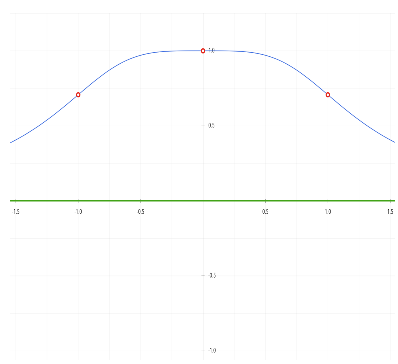
(e) $\int_0^1 \frac{sinx}{x}\,dx$



(f) $\int_0^1 sin(x^2)\,dx$



(g) $\int_0^1 (sinx)^{0.5}\,dx$



(h) $\int_0^1 \frac{1}{\sqrt{1+x^4}}\,dx$

Figure 7: Images of Transcendental function

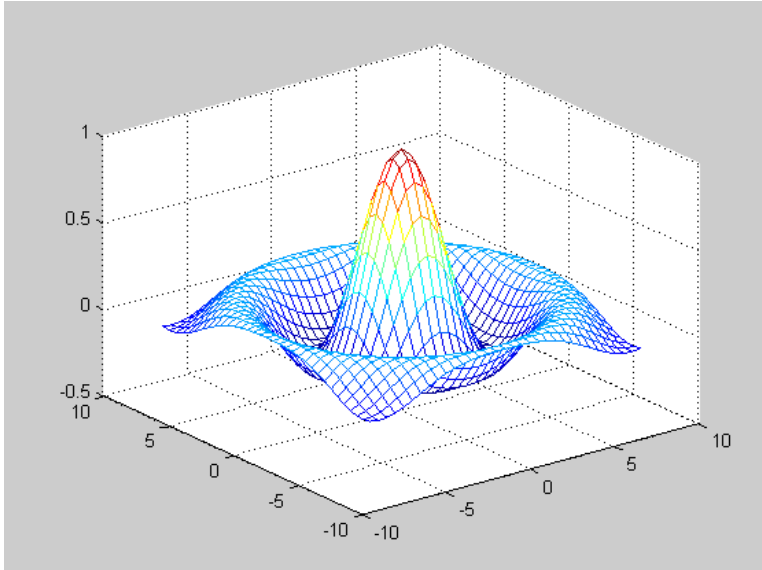


Figure 8: $sin\sqrt{x^2+y^2}/\sqrt{x^2+y^2}$

Partition number	method	estimate	error
10	Trap	3.1399398	1.7×10^{-3}
	Para	3.14155	4.1×10^{-8}
	Monte Carlo	2.545980	-1.9×10^{-1}
100	Trap	3.1416496	1.7×10^{-5}
	Para	3.14160128	4.0×10^{-11}
	Monte Carlo	3.129750	-3.7×10^{-3}
1000	Trap	3.1416001	1.6×10^{-8}
	Para	3.141600002	4.2×10^{-14}
	Monte Carlo	3.136207	-1.6×10^{-3}

Table 6: The error comparison of different methods

2. Due to the limitation of the accuracy of the C language and MATLAB data type, it will result in the loss of the accuracy of the calculation and output, therefore resulting in a certain degree of error.

Mathematical analysis of the Monte Carlo method According to central limit theorem, if Xi independent with each other, and expectation and variance exist. When n go to infinity, variable

$$Y = \frac{\overline{I} - I}{\sigma/\sqrt{n}} \sim N(0,1)$$

that is, for every $t > 0$,

$$P\{|Y| < T\} = P\{|\overline{I} - I| < \frac{t\sigma}{\sqrt{n}}\}$$

In this way, it means that the convergence rate is relative slow, the error is $O(\frac{1}{\sqrt{n}})$, which is corresponding to the experimental result.

Deficiencies of Monte Carlo method Deliberately, it calculates $\int_0^2 \frac{x}{lnx}\,dx$. The integrand will diverge. However Monte Carlo method get the finite result, which is obviously wrong.

6 Conclusions and Ameliorations

6.1 Conclusions

The trapezoidal method and parabolic method is accurate especially for less smooth integrand. Nevertheless, Monte Carlo method doesn't have this limit. As is shown in the table.4, the trapezoidal method can get a result when the integrand oscillate violently.

6.2 Ameliorations

It's shown that the convergence rate of Monte Carlo method is slow, hence considering combining it with numerical method. First of all, dividing the interval into n parts equivalently. Generating n random variables in the subinterval, and using Monte Carlo method in them.

$$\eta_i = a + \frac{b-a}{n}(\psi + i - 1) \tag{20}$$

Mapping η_i to subinterval.

$$a + \frac{i-1}{n}(b-a), a + \frac{i}{n}(b-a) \tag{21}$$

Thus the integral is,

$$I = \frac{b-a}{n}\sum_{i=1}^{n} f(\eta_i) \tag{22}$$

A Central Limit Theorem

x	-3.0	-2.5	-2.0	-1.5	-1.0	-0.5	0.0
$\Phi(x)$	0.001	0.006	0.023	0.067	0.159	0.309	0.500
$F_{30}(x)$	0.000	0.001	0.014	0.054	0.168	0.343	0.444
$F_{3000}(x)$	0.002	0.005	0.018	0.063	0.157	0.316	0.496
x	0.5	1.0	1.5	2.0	2.5	3.0	-
$\Phi(x)$	0.691	0.841	0.933	0.977	0.994	0.999	-
$F_{30}(x)$	0.663	0.814	0.920	0.967	0.993	0.997	-
$F_{3000}(x)$	0.688	0.842	0.924	0.964	0.993	0.999	-

Table 7: Comparison between F(x) and $\Phi(x)$

The central limit theorem (CLT) states that, given certain conditions, the arithmetic mean of a sufficiently large number of iterates of independent random variables, each with a well-defined expected value and well-defined variance, will be approximately normally distributed, regardless of the underlying distribution.

$$\sqrt{n}((\frac{1}{n}\sum_{i=1}^{n} X_i) - \mu) \to N(0, (\sigma)^2)$$

Now $X \sim \pi(0.5)$, the expectation is 0.5, and the variance is 0.5. Taking n repeated observation result

$$\overline{X} = \sum_{i=1}^{n} X_i$$

$$Z = \frac{\overline{X} - n \times 0.5}{\sqrt{n \times 0.5}}$$

Using MATLAB simulate the measurement 30 times and 3000 times, and comparing them with $\Phi(x)$ respectively.

```
y=poissrnd(0.5,1000,3000);
%A 1000*3000 martix, every line can be seen as a 3000 repeated measurements
xm=(mean(y,2)-0.5)*sqrt(6000);
x1=sum([xm<-3,xm<-2.5,xm<-2,xm<-1.5,xm<-1,xm<-0.5,xm<0])/1000;%Classify
x2=sum([xm<0.5,xm<1,xm<1.5,xm<2,xm<2.5,xm<3])/1000;%Classify
normcdf(-3:0.5:3,0,1);%Normal distribution
```

Figure 9: Simulation of central limit theorem

Then it's found that the F(x) is corresponding to $\Phi(x)$ to a certain extent, and F_{3000} behaves better, as shown in the table.

Acknowledgements

I would like to extend my sincere gratitude to my teacher for her constant encouragement and guidance on this paper and teaching assistants for their advice and help.

References

[1] Monte Carlo method $https://en.wikipedia.org/wiki/Monte_Carlo_method$, 2016-03-17