Monte Carlo Method for Double Integration: Good or Not

A.K. Saxena
Department of Mathematics, Bareilly College, Bareilly, Uttar Pradesh, India
aksaxena.saxena61@gmail.com

Abstract - Monte Carlo method was first developed as a method for estimating integrals that could not be evaluated analytically. It numerically computes a definite integral using random numbers. Role of random numbers in case of Monte Carlo Integration (one dimension) has already been discussed with several satisfactory conclusions. In this paper we are going to make an error analysis for the evaluation of double integrals using the same method. It will point the reader's attention to the uncertain pattern of error and they will be able to decide whether Monte Carlo method for double integration works well for their purpose or not.

Keywords: Monte Carlo Method, Two Dimensional Numerical Integration, Random Numbers, Reliability of random numbers

1. INTRODUCTION

Since Monte Carlo method was based on random numbers therefore equal attention of scientists were attracted towards the random numbers. In early times a great work has accomplished regarding random numbers, randomness and their use. A random number is a number generated by a process, whose outcome is unpredictable, and which cannot be sub sequentially reliably reproduced. This definition works fine provided that one has some kind of a black box usually called a random number generator that fulfills this task. However, if one were to be given a number, it is simply impossible to verify whether it was produced by a random number generator or not. In order to study the randomness of the output of such a generator, it is hence absolutely essential to consider sequences of numbers. A detailed study of random numbers and various techniques (Random Number Generators) to produce random numbers may be found in the book by James E.Gentle [5]. The types of random numbers and randomness are clearly defined and illustrated in an article by Eddelbuettel [2]. In a research paper by P. Hellekalek [9] he mentioned that what should be the properties of good random number generators. To understand the wide range of problem solving techniques which use random numbers and statistics of probability, the term "Monte Carlo Method is used.

The problem for Monte Carlo method for numerical integration has attracted considerable interest during the last few decades because of it's simplicity and time saving characteristic. This Method was first used by Stanslaw Ulam for simulations in Physics and other fields that require solutions for problems that are impractical or impossible to solve by traditional analytical or numerical methods. The beginning of Monte Carlo method and its origination is discussed in a scholarly article by N.Metropolis [8] and Rodger Eckhardt in his scholarly article [11] mentioned the problem, which as a solution discovered this technique. In the same article Rodger also published the letters showing the personal conversation of the two pioneers, Stan Ulam and John Von Neumann regarding Monte Carlo method. This is a method for solving problems using random variables. It is a powerful tool in many fields of mathematics, physics and engineering. A comprehensive review of literature concerning Monte Carlo method may be found in the book by Malving H. Kalos and Paula A. Whitlock [7].The method is being used more and more in recent years, especially in those cases where the number of factors included in the problem is so large that an analytical solution is impossible for example Numerical Integration for higher orders.

A chapter in encyclopedia of Biostatistics by Gordon K. Smith [6] throws light on all the techniques available to solve one or two dimensional integration including Monte Carlo technique. In 2001 Harvey Gold in his book [3] studied all the techniques of numerical integration and mentioned their error analysis. In 2002 Guan Yang discussed all other methods and techniques of numerical integration and stated the benefits of Monte Carlo integration over the other methods available in an extended essay in mathematics [4] during his IB diploma program. Early studies on Monte Carlo integration were mainly concerned with the problem of improving the randomness of numbers used. The non-parametric tests to check the randomness of numbers may be referred from Lal Mohan Bhar [1]. Just to avoid the inherent errors of random numbers a long back Stephen K. Park & Keith W. Miller suggested to follow the minimal standards [12] for random number generators. To get rid of this situation of ambiguity that whether the numbers in use are true random or not, concept of quasi random numbers was coined and discussed by Russel E. Caflisch [10]. In this paper the author has performed the reliability test of random numbers for Monte Carlo Integration in two dimensions.

2. MONTE CARLO INTEGRATION

Let we wish to evaluate

$$\mathrm{I} = \int_{\alpha}^{\beta}\int_{a}^{b} f(x,y)\,dx\,dy \qquad \text{.............(1)}$$

Now the problem of evaluating I is now in a form that can be solved efficiently using Monte Carlo method.
Monte Carlo method for such double integral is being performed in following two ways.

2.1 Using Random Nodes

The very first step of Monte Carlo Problem is to set up a simulation to approximate the value of E$\{f(x,y)\}$.This expectation can be evaluated by taking the mean of the functional value at n randomly selected points emanating from a sequence of pseudo random numbers generated by some PRNG generator which provides values in the region $a \leq x \leq b, \alpha \leq y \leq \beta$ i.e.

$$E\{f(x,y)\} = \frac{1}{n}\sum_{i}^{n} f(x_i, y_i) \qquad \ldots\ldots\ldots\ldots\ (2)$$

Hence from (2.1) & (2.1.1) I can be approximated by

$$I \approx \frac{(b-a)(\beta-\alpha)}{n}\sum_{i}^{n} f(x_i, y_i) \qquad \ldots\ldots\ldots\ldots\ (3)$$

On account of the fact that $(b-a)$ is eaxact, the uncertainty of the algorithm is bounded by the uncertainty in our estimate of E$\{f(x,y)\}$.If our estimate for E$\{f(x,y)\}$ is precise then the sum of the volume between $f(x,y)$ and E$\{f(x,y)\}$ should be exactly equal to zero.

2.2 Using Random Nodes

While evaluating the expected value using equispaced points we take n $*$ n uniformly distributed points emanating from a sequence of equidistant numbers obtained by dividing the range of x and y in n parts which provides values in the region $a \leq x \leq b, \alpha \leq y \leq \beta$. i.e.

$$E\{f(x,y)\} = \frac{1}{n*n}\sum_{j}^{n}\sum_{i}^{n} f(x_i, y_j) \qquad \ldots..(4)$$

Hence from (2.1) & (2.2.1) I can be approximated by

$$I \approx \frac{(b-a)(\beta-\alpha)}{n*n}\sum_{j}^{n}\sum_{i}^{n} f(x_i, y_j) \qquad \ldots..(5)$$

3. SOURCE OF RANDOM NUMBERS

The random numbers which are used by author in this research paper may be obtained from [13]. Random numbers were taken from two different sources.

3.1 Using computer program (random number generator) the files of random numbers are stored as **".dat"** files of size 1000, 2000, 3000, 4000, 5000 with name

Int_1_1,	Int_2_1,	Int_3_1 ,
Int_1_2,	Int_2_2,	Int_3_2,
Int_1_3,	Int_2_3 ,	Int_3_3,
Int_1_4,	Int_2_4,	Int_3_4,
Int_1_5,	Int_2_5,	Int_3_5,

Here the first three letters "Int" refers to the word integral, last numeral refers to the file size (1000, 2000, 3000, 4000, 5000) and first numeral refers to n^{th} times execution of same program for same size.

3.2 Using three online generators
random.org (http://www.random.org/decimalfractions)
research randomizer
(http://www.randomizer.org/form.htm)
graph pad software
(http://www.graphpad.com/quickcalcs/randomn1.cfm)
five data files of random numbers from each of the above noted sites are saved as under

File Name	Site Address	Size of Data file
olrr1.dat	Research Randomizer	1000
olrr1.dat	Research Randomizer	2000
olrr1.dat	Research Randomizer	3000
olrr1.dat	Research Randomizer	4000
olrr1.dat	Research Randomizer	5000
olrorg1.dat	Random.Org	1000
olrorg2.dat	Random.Org	2000
olrorg3.dat	Random.Org	3000
olrorg4.dat	Random.Org	4000
olrorg5.dat	Random.Org	5000
olgp1.dat	Graph Pad	1000
olgp2.dat	Graph Pad	2000
olgp3.dat	Graph Pad	3000
olgp4.dat	Graph Pad	4000
olgp5.dat	Graph Pad	5000

Table 1: Nomenclature of Online Data Files

As far as the notation and nomenclature of these files are concerned it should be noted, ol stands for ONLINE , gp stands for GRAPH PAD , rorg stands for RANDOM.ORG , rr stands

for RESEARCH RANDOMIZER and the last numeral n stands for file size multiplied by 1000. Since it is very cumbersome to present all these numbers in this research paper therefore these may be seen and accessed from the web address*:
http://www.4shared.com/folder/BAytR7eW/data_files.html

All the random numbers in the above noted files are distinct and have no correlation with each other. Before using these numbers in the research paper these numbers have gone through four methods to test their independence and these methods are Poker Test, Run Test, Frequency Test and Frequency Monobit Test. The detailed study of the tests applied may be obtained from the web address*:

http://www.4shared.com/folder/311UvmlP/test_for_randomness_of_data_fi.html

4. INTEGRAL EVALUATION

In the present work, three types of two dimensional integrals are taken into consideration and are evaluated by Monte Carlo method using random points (computer generated & online generated) as well as equispaced points

4.1 First integral (2 D)

The first integral under investigation is

$$I_2 = \int_0^1 \int_1^2 xy(1+x+y)dxdy \qquad \ldots\ldots..(7)$$

Exact value of which is 30.75

4.1.1 Using random nodes

We have 6 data files corresponding to each of the different sizes of 1000, 2000, 3000, 4000 and 5000 data (30 files in all).Out of these six data files of each we have to choose one data file for x series and one data file for y series of same size. Out of the six data files for 1000 data size there can be 15 combinations and 15 more combinations when data files for x series and y series are interchanged.
These 30 pairs of codes for 30 file combinations are

(1,2),(1,3),(1,4),(1,5),(1,6),(2,3),(2,4),(2,5),(2,6),
(3,4),(3,5),(3,6),(4,5),(4,6),(5,6),(2,1),(3,1),(4,1),
(5,1),(6,1),(3,2),(4,2),(5,2),(6,2),(4,3),(5,3),(6,3),
(5,4),(6,4),(6,5)

In order to evaluate our first integral by Monte Carlo Integration using the random nodes, we used a computer program and by repeated execution of the program to evaluate the integral for these 30 combinations of files for x and y series, each time we get the different values of the integral and the best combination of files for evaluation (i.e. minimum error in value of the integral) of the integral is being given corresponding to the different data size

S.No.	Data Size	Best Combination	Minimum Error
1	1000	olgp1 & int_3_1	0.02216
2	2000	int_1_2 & olrorg2	0.10327
3	3000	int_3_3 & olgp3	0.02228
4	4000	olrorg4 & olgp4	0.01607
5	5000	olgp5 & olrr5	0.00375

Table 2: Best combination of files for evaluation of first integral

4.1.2 Using equispaced nodes

In order to evaluate the value of the first integral by using equispaced nodes, we will make use of the same program and by repeated execution of the program for 200, 300,. . . 2000 divisions we get the following tabulated values.

S. No.	No. of Equispaced Nodes	Value of Integral	True Value	Error
1	100	31.26226	30.75	0.51226
2	200	31.00549	30.75	0.25549
3	300	30.92027	30.75	0.17027
4	400	30.87804	30.75	0.12804
5	500	30.85209	30.75	0.10209
6	600	30.83486	30.75	0.08486
7	700	30.82266	30.75	0.07266
8	800	30.81346	30.75	0.06346
9	900	30.80615	30.75	0.05615
10	1000	30.79992	30.75	0.04992
11	1100	30.7935	30.75	0.0435
12	1200	30.79036	30.75	0.04036
13	1300	30.7849	30.75	0.0349
14	1400	30.77986	30.75	0.02986
15	1500	30.7731	30.75	0.0231
16	1600	30.77006	30.75	0.02006
17	1700	30.76382	30.75	0.01382
18	1800	30.76133	30.75	0.01133
19	1900	30.75408	30.75	0.00408
20	2000	30.74195	30.75	0.00805

Table 3: First integral (2D) by equispaced nodes

The following graph displays the evaluated values of our first (2 D) integral using equispaced nodes.

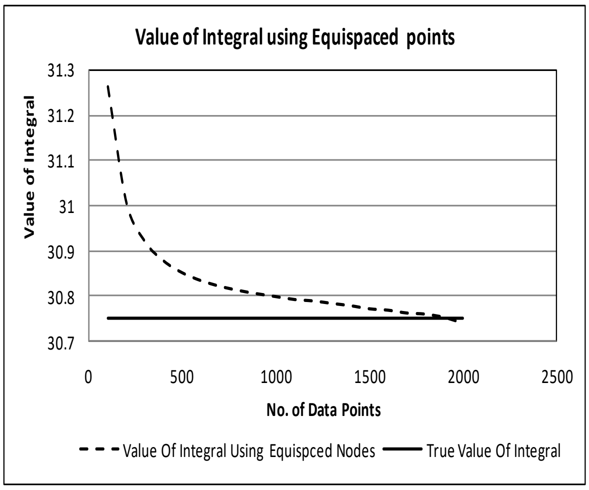

Fig. 1: Value of Integral using equispaced points

4.2 Second integral (2 D)

The second integral under investigation is

$$I_1 = \int_1^2 \int_3^4 (xy + e^y)dxdy \quad \text{........(6)}$$

Exact value of which is 39.7661

4.2.1 Using random nodes

For the evaluation of second integral with random nodes, we shall make use of the same program as in case of first integral with only a change in functional form, limits and the storage of number from data file as per the limits.

By repeated execution of the program to evaluate the second integral for the above mentioned 30 combinations of files for x and y series, each time we get the different values of the integral and the best combination of files for evaluation of the integral (i.e. minimum error in value of the integral) corresponding to the different data size is being given in the following table.

S. No.	Data Size	Best Combination	Minimum Error
1	1000	olrorg1 & olrr1	0.0035
2	2000	int_1_2 & olrorg2	0.00281
3	3000	int_2_3 & int_1_3	0.0269
4	4000	olgp4 & int_3_4	0.0455
5	5000	int_3_5 & olrr5	0.00995

Table 4: Best combination of files for evaluation of second integral.

4.2.2 Using equispaced nodes

In order to evaluate the value of the integral by using equispaced nodes, we used another computer program and by repeated execution of the same for 200, 300...…, 2000 divisions we get the following tabulated values.

S. No.	No. of Equispaced Nodes	Value of Integral	True Value	Error
1	100	39.96033	39.7661	0.19423
2	200	39.86197	39.7661	0.09587
3	300	39.82852	39.7661	0.06242
4	400	39.81255	39.7661	0.04645
5	500	39.8028	39.7661	0.0367
6	600	39.79688	39.7661	0.03078
7	700	39.79083	39.7661	0.02473
8	800	39.78775	39.7661	0.02165
9	900	39.78392	39.7661	0.01782
10	1000	39.78723	39.7661	0.02113
11	1100	39.78639	39.7661	0.02029
12	1200	39.78155	39.7661	0.01545
13	1300	39.77782	39.7661	0.01172
14	1400	39.77043	39.7661	0.00433
15	1500	39.75865	39.7661	.00745
16	1600	39.74618	39.7661	.01992
17	1700	39.73696	39.7661	.02914
18	1800	39.72743	39.7661	.03867
19	1900	39.75777	39.7661	.00833
20	2000	39.79809	39.7661	0.03199

Table 4: First Integral (2D) by Equispaced Nodes

[The input (say 100) for number of division will divide the range of x & y in 100 equal parts. These points of division of x & y range are equispaced nodes for x and y series.]

The following graph displays the evaluated values of our first (2 D) integral using equispaced nodes.

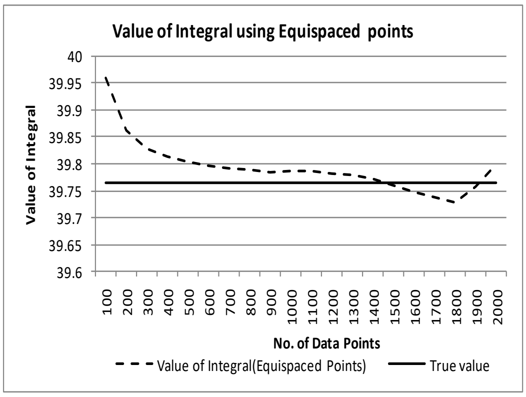

Fig. 2: Value of Integral using equispaced points

4.3 Third integral (2 D)

The third and last integral is

$I_3 = \int_2^4 \int_1^3 \left(\frac{x}{y} + \frac{y}{x}\right) dxdy$ …………..(8)

Exact value of which is 9.364262

4.3.1 Using random nodes

For the evaluation of third integral with random nodes, we shall make use of the same program as in case of first integral with only a change in functional form, limits and the storage of number from data file as per the limits.

By repeated execution of the program to evaluate the integral for the above mentioned 30 combinations of files for x and y series, each time we get the different values of the integral and the best combination of files for evaluation of the integral (i.e. minimum error in value of the integral) corresponding to the different data size is being given in the following table.

S. No.	Data Size	Best Combination	Minimum Error
1	1000	olrorg1 & int_3_1	0.001
2	2000	olrr2 & olgp2	0.00015
3	3000	olrorg3 & int_2_3	0.00035
4	4000	olrr4 & int_2_4	0.00036
5	5000	olrorg5 & int_2_5	0.00118

Table 6: Best combination of files for evaluation of third integral

Since it is very difficult to present all the values of integral corresponding to different combination of data files of random numbers, in this research paper therefore these may be seen and accessed from the web address*:

http://www.4shared.com/folder/dSsswzi/Table_for_double integrals.html

4.3.2 Using equispaced nodes

In order to evaluate the value of third integral using equispaced nodes, we will make use of the same program and by repeated execution of the program for 200, 300,...., 2000 divisions we get the following tabulated values.

S. No.	No. of Equispaced Nodes	Value of Integral	True Value	Error
1	100	9.350109	9.364262	0.014153
2	200	9.357004	9.364262	0.007258
3	300	9.358994	9.364262	0.005268
4	400	9.359395	9.364262	0.004867
5	500	9.358871	9.364262	0.005391
6	600	9.357507	9.364262	0.006755
7	700	9.35608	9.364262	0.008182
8	800	9.351929	9.364262	0.012333
9	900	9.350745	9.364262	0.013517
10	1000	9.336454	9.364262	0.027808
11	1100	9.329106	9.364262	0.035156
12	1200	9.325695	9.364262	0.038567
13	1300	9.323879	9.364262	0.040383
14	1400	9.319521	9.364262	0.044741
15	1500	9.305141	9.364262	.059121
16	1600	9.280134	9.364262	0.084128
17	1700	9.272767	9.364262	0.091495
18	1800	9.270508	9.364262	0.093754
19	1900	9.269471	9.364262	0.094791
20	2000	9.267021	9.364262	0.097241

Table 7: Third integral (2 D) by equispaced nodes

The graph 6.5.1 displays the evaluated values of our first (2 D) integral using equispaced nodes.

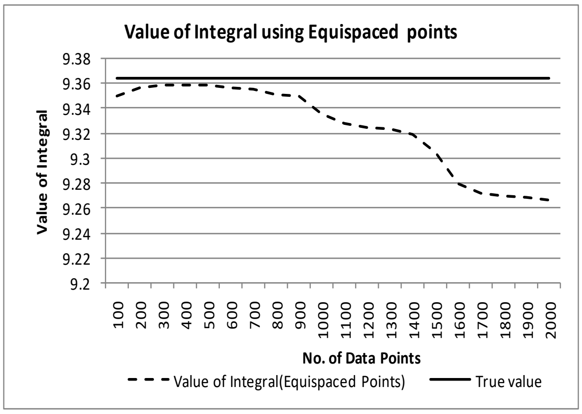

Fig. 3: Value of integral using equispaced points

5. OBSERVATIONS

The best combinations to evaluate a particular integral corresponding to the file of size 1000 is not giving the same or better accuracy corresponding to the files of size 2000, 3000, 4000 and 5000 [see table 2, 4 and 6] and the error in the values of all the double integrals don't follow any pattern corresponding to different size of random numbers and also seems to be random in nature whereas using equispaced nodes the error follow a pattern and steadily approaches to zero [see Fig. 1, 2 and 3] which makes our conclusion as.

CONCLUSION

- Whatever may be the extent of randomness of random the numbers (True or Pseudo) that are used in Monte Carlo integration, it is not necessary that the set of random numbers which gives the best approximation (minimum error) of one integral (Double) will also yield the same accuracy in the evaluation of other integral whereas if we use equispaced numbers in a given range of integration then we get almost smooth curve corresponding to the error in the values of integral having a regular decrement.
- It is also not necessary that corresponding to different size of random numbers obtained through same source will give the value of integral more and more approximated i.e. increase in the random numbers doesn't give the assurance for regular increment in accuracy of the same integral.

  Since one can't be assured about accuracy of integral by assuring the randomness of numbers or size of the numbers therefore in total we can say

  "Random numbers may not be considered as of much reliability in case of 2 dimensional numerical integration."

  Although the decision may vary as per the accuracy required for a particular purpose but those cases where accuracy is of utmost importance and stored data of random numbers are used we can easily accept the above mentioned conclusion.

REFERENCES

1. Bhar L., "*Non-parametric Test"*, Indian Agricultural Statistics Research Institute, LIbrary Avenue, New Delhi.
2. D. Eddelbuettel, Random: *True random numbers using random.org*, 2007, http://www.random.org.
3. Gould, Harvey, Jan Tobochnik, and Wolfgang Christian, "*Numerical Integration and Monte Carlo Method*
4. Guan Yang, "*A Monte Carlo Method of Integration*", Extended Essay in Mathematics, Denmark (2002)
5. J. E.Gentle,"Random Number Generation and Monte Carlo Methods",Springer, New York, second edition (2003)
6. K. Gordon, "*Numerical Integration*", John Wiley & Sons, Ltd, (ISBN 0471975761) Edited by, Peter Armitage and Theodore Colton, Chichester, (1998).
7. Malvin H. Kalos and Paula A. Whitlock, "*Monte Carlo Methods*", Wiley -VCH. ISBN 9783527407606, (2008).
8. N.Metropolis, *"The beginning of the Monte Carlo method"*, Los Alamos Science, Special Issue (1987) 125-130.
9. Peter Hellekalek, "*Good random number generators are (not so) easy to find"*, Mathematics and Computers in Simulation, 46 (1998) 485-505.
10. R. E. Caflisch, "*Monte Carlo and quasi-Monte Carlo methods*", Acta Numerica. **7**. (1998)1–49.
11. Roger Eckhardt, "*Stan Ulam, John von Neumann and the Monte Carlo method"*, Los Alamos Science, Special Issue (15) (1995) 131-137.
12. S. K. Park and K. W. Miller," *Random number generators: Good ones are hard to find*", Communications of the ACM, 31(1988)1192-1201.