AGU International Journal of Sciences – 2026, Vol. 12 (1), 59 – 62

AGU INTERNATIONAL JOURNAL OF SCIENCES
p-ISSN 3093 - 3536

MONTE CARLO INTEGRATION METHOD AND APPLICATIONS

Pham Thi Thu Huong[1] ,Pham Thi Thu Hoa[2]
[1]*An Giang University, VNU-HCM*

Information:
Received:17/8/2025
Accepted:23/9/2025
Published:12/2025

Keywords:
MC integration, Bayesian statistics, infinite integrated regions, random generation, modification of MC integration

ABSTRACT

Effective integration methods are very necessary for making inferences in statistics. These methods can compute approximations for intractable integrals to calculate for probabilities, expectations, variances, etc. For deterministic numerical methods, quadrature forms are used to approximate for the region beneath the functions. These methods depend on the number of partitions and the finite of the integrated regions. Monte Carlo (MC) integration methods can solve for this problem and provide accurate approximations using random sample generation techniques in statistics. The aims of this paper are to weaken assumptions of the deterministic numerical methods for infinite integrated regions by using the MC integration method. The way to control the accuracy of the MC integration method is also proposed. This MC integration method has been implemented for particular examples with finite and infinite integrated regions. The results show that the proposed methods of applying MC integration can produce better calculations than the deterministic numerical methods and can handle for infinite integrated regions.

1. INTRODUCTION

In statistical practice, it is frequent to encounter situations with difficult and intractable integrals. The integrations integration methods are used for things like: computing probabilities, finding expectations and variances and calculating marginal distributions (Gelman, 1995; Robert & Casella, 2004; Wakefield, 2013).

In particular, in Bayesian analysis, unknown parameters are considered as random variables and they have distributions. These distributions are called posterior distribution or target distribution in model performance. We can make inferences of unknown parameters by drawing probability distributions (Choi & Hobert, 2013; Hastings, 1970; Wakefield, 2013). However, these distributions usually have complex forms and its integrals are intractable. Therefore, to make inference for the target distribution, effective integration methods need to be applied.

The process of taking derivatives of a function can be easily automated. However, to compute integrals, it is much more difficult. If the functions have the close-form solutions, there are a few rules to compute these integrals and they do not cover a lot of cases (David & Rabinowitz, 1984). Many situations either require repeated trial and error or are simply intractable. In these cases, we need to implement mathematical algorithms to approximate the integrals.

Given an integral that is analytically intractable and lacks a close-form solution, deterministic numerical methods are often used. Some quadrature forms are used to calculate the approximation of the region beneath the function (James, 2013). The accuracy of these methods is based on the number of partitions of the region and the computing finite volumes over regular subspaces. Moreover, these methods have to deal with a lot of problems when computing for the integrals with infinite integrated region.

The MC integration method is introduced to solve intractable integrals in this paper. By generating random samples for a distribution, Monte Carlo integration method approximated the integrals by the volume under the integrated function (Cox, 1979; Robert & Casella, 2004). This method can deal with the infinite integrated regions and has precise and solid calculating results (Gelman, et al., 1995; Wakefield, 2013). In this paper MC integration methods are proposed with finite and infinite integration regions in a special case of normal distribution. The syntaxes for the MC integration methods for particular examples are proposed. These algorithms can not only solve for integrations with infinite regions but also can deal with the integrations having small value. Moreover, the way to control the accuracy of the methods is presented to get more precise integration results.

The paper is organized into 5 sections as follows. The fundamental theories of deterministic numerical methods and the Monte Carlo integration method have been presented in Section 2 and Section 3. In Section 4, examples with finite and infinite integrated regions are proposed by implementing the Monte Carlo integration method with the detailed algorithms. The way to control the accuracy of the Monte Carlo integration methods is introduced in Section 4.2. The conclusion is discussed in Section 5.

2. DETERMINISTIC NUMERICAL INTEGRATION METHODS

The general idea for deterministic numerical integration is to use some forms of quadrature and base it on the approximation as (David & Rabinowitz, 1984; James, 2013)

$$\int_a^b f(x)dx \approx (b-a)\frac{f(a)+f(b)}{2}.$$

The integral is an approximation of the region beneath the function f(x) between a and b. It is obvious that this approximation is more accurate with smaller sections of the function. So, various schemes or partitioning are suggested. However, in general, the formula is expressed as

$$\int_a^b f(x)dx \approx \frac{1}{2}\sum_{k=1}^{N}\left(x_{k+1}-x_k\right)\left(f(x_{k+1})+f(x_k)\right),$$

where, $x_1 = a$ and $x_N = b$.

The error for this method is

$$error = \int_a^b f(x)dx - \frac{1}{2}\sum_{k=1}^{N}\left(x_{k+1}-x_k\right)\left(f(x_{k+1})+f(x_k)\right).$$

If the distances of the partitions are equal, the error term can be expressed as

$$error = \int_a^b f(x)dx - \frac{b-a}{N}\left[\frac{f(a)+f(b)}{2} + \sum_{k=1}^{N-1} f\left(a+k\frac{b-a}{N}\right)\right].$$

There exist numbers $a<\xi<b$ such that

$$error = -\frac{(b-a)^3}{12N^2} f''(\xi).$$

The accuracy of the approximation depends on the number of sub-regions N and how they are constructed.

3. MONTE CARLO INTEGRATION METHOD

The idea of the Monte Carlo integration method involves the derivation of the acceptance rate for rejection sampling for points generated in a region (Marin & Robert, 2014; Robert & Casella, 2004). The formula for this method is defined as

$$\int_X f(x)dx = volume\ of\ the\ region \times proportion\ of\ the\ volume\ in\ f(x).$$

For example, as in Figure 1, the integration is calculated as

$$\int_X f(x)dx = S_1 \frac{S_2}{S_1}.$$

Here, S_1 is the volume of the rectangle and S_2 is the volume under the function $f(x)$ from $-\infty$ to $+\infty$. The area S_1 is easy to calculate. To calculate the proportion S_2 / S_1, we can generate an uniform random sample from region X with N sample size and calculate the volume inside the region S_2 and outside of the region S_2. Assuming that the volume of the region S_1 which we are generating uniform random samples is known, the proportion can be approximated as

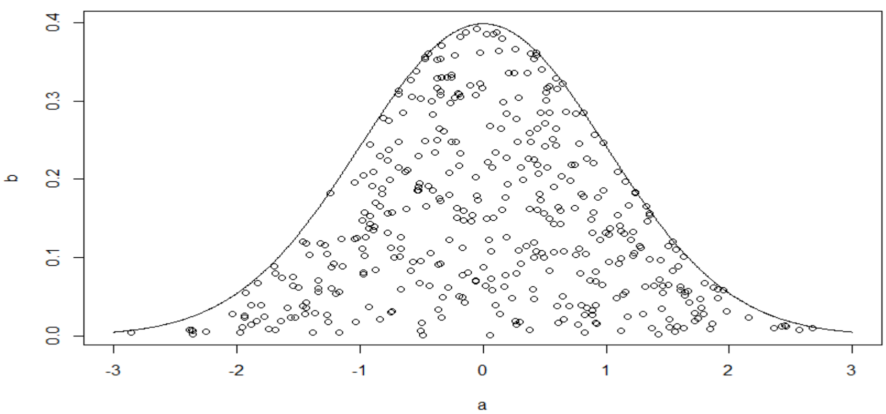

Figure 1. An example for the MC integration

$$\frac{S_2}{S_1} \approx \frac{1}{N*C}\sum_{i=1}^{n} f(x_i), x_i \sim U(X),$$

where, $x_i \sim U(X)$ means x_i follows an uniform distribution in the region X and $C = \max_{x_i \in X} f(x_i)$ is a constant.

In more familiar terms we get the relationships

$$\int_a^b f(x)dx \approx (b-a)\frac{1}{N}\sum_{i=1}^{N} f(x_i),\ x_i \sim U(a,b). \qquad (3.1)$$

Here, $U(a,b)$ is an uniform distribution in the set (a,b).

We have another expression for (3.1) as

$$\int_a^b f(x)dx \equiv \int_X I_{(a,b)} f(x)dx.$$

So that

$$(b-a)\frac{1}{N}\sum_{i=1}^{N} f(x_i),\ x_i \sim U(a,b) \equiv \frac{1}{N}\sum_{i=1}^{N} I_{x_i \in (a,b)}, x_i \sim f(x). \quad (3.2)$$

Here, $x_i \sim f(x)$ means the density function of this variable is the function $f(x)$. If we can generate a sample from function *f*(x), we can evaluate the integral

$$\int_X h(x) f(x)dx = E(h(x)) \approx \overline{h}_N(x) = \frac{1}{N}\sum_{i=1}^{N} h(x_i),\ x_i \sim f(x), \quad (3.3)$$

where $h(x_i) = I_{x_i \in (a,b)}$.

In Huong and Hoa (2021) and Robert and Casella (2004), a random sample can be generated from any posterior function. The technique in Huong and Hoa (2021) can be applied to formula (3.3). This idea gives us a useful way to estimate integrals when we can generate random variables from our density *f*(x).

By the Strong Law of Large Numbers, we have

$$\overline{h}_N(x) \to h(x).$$

$E_f(h(x)^2) < +\infty$, we have

$$\mathrm{var}(\overline{h}_N(x)) = \frac{1}{N}\int_X \left(h(x) - E_f(h(x))\right)^2 f(x)dx.$$

Since $\overline{h}_N(x)$ is a random variable, the central limit theorem can be used to make inference about the accuracy of this estimation.

4. APPLING MC INTEGRATION AND CONTROLLING OF MC ESTIMATES

4.1 An example for MC integration

We consider the standard normal density function and using the MC methods as in (3.1) to calculate the integral having a form as

$$\int_0^1 f(x)dx = \int_0^1 \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx. \qquad (4.1)$$

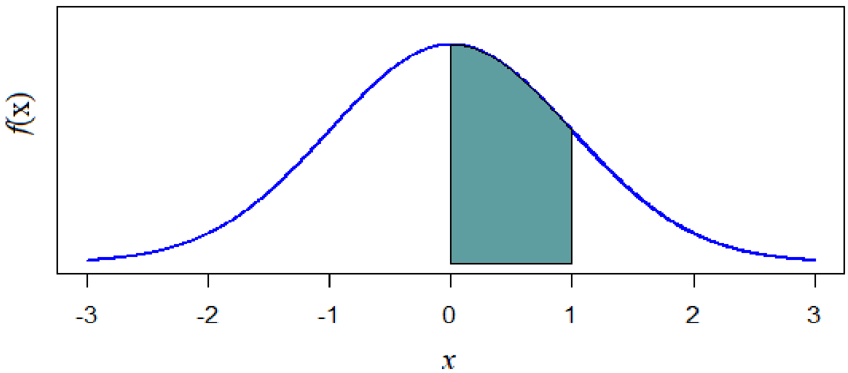

Figure 2. Region of the integration of the function in (4.1)

To use the MC integration method for integrating this function, the syntax is as follows:

Step 1: Generate N sample points from uniform distribution U(0,1)

Step 2: Calculate values of f(x) at these points

Step 3: Sum all value of $f(x_i)$ and divide by N.

After running 10.000 sample points for this syntax, we have the result

$$\int_0^1 \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx = 0.3413447.$$

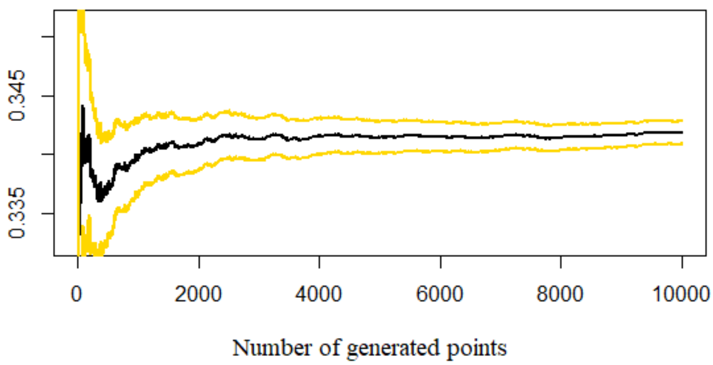

Figure 3. Estimates for the integral and 95% confident interval range

The trace of estimates of the integration is shown in Figure 3. The black line represents the estimates of the integral corresponding with the number of generated points of the function in (4.1) and the yellow lines are 95% ranges for estimating of the integral. It can be seen that when the sample size of the simulation is small, the estimates are more fluctuated. When the sample size is large than 1,000 sample points, the estimates are more stable and converged. It means that the Monte Carlo estimates are more accurate when the sample size of the simulation is large.

4.2 Controlling of Monte Carlo estimates

The Monte Carlo estimates depend on the sample size of the generation. The more samples the less variance in the MC estimates. However, to get the best answer with the fewest number of samples, the natural solution is that we conduct multiple MC sequences at the same time for estimating the integral of a function.

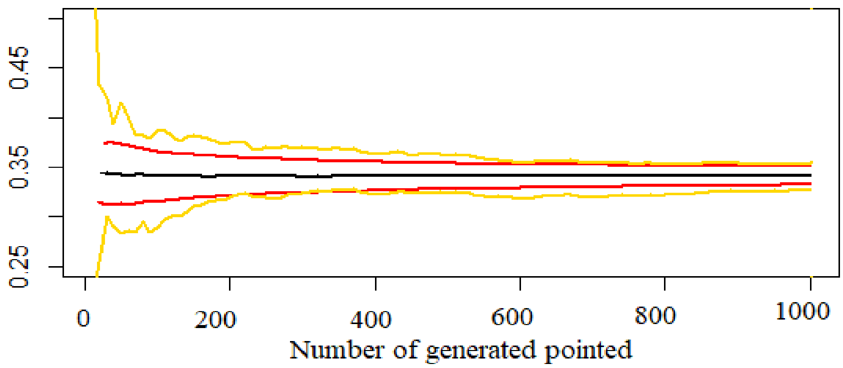

Figure 4. Mean and 95% error ranges for single MC sequence and multi MC sequences

The 100 MC sequences with sample size of 1,000 points in each sequence have been conducted to calculate the integration in (4.1). The results show in Figure 4. The black line shows the mean of the estimated integral of function (3.1). The yellow lines show 95% error ranges for a single MC sequence estimating the integral over (0,1) and the red lines show 95% error ranges for 100 MC sequences estimating the integral over (0,1).

It is clear that the multiple sequences converge to the true value of the integral much faster when the black line is very stable from the beginning of simulated points. The 95% ranges of a single MC sequence are larger than the 95% ranges of 100 MC sequences. This means that when more MC sequences conducted to estimate for the integral the result is more accurate with fewer generated points.

4.3 A modification of MC integration

In this case, the region of our interest is infinite and difficult to sample from. The difficulty is increased when the integration has a small value under infinite integration regions. An importance distribution can be used along with MC methods to solve this problem. The theory here is that

$$\int_X f(x)dx = \int_X \frac{f(x)}{g(x)} g(x)dx \approx \frac{1}{N}\sum_{i=1}^{N} \frac{f(x_i)}{g(x_i)},\ x_i \sim g(x). \quad (4.2)$$

For example, let $X \sim N\ (0,1)$, we want to estimate the integration

$$P(X > 4.5) = \int_{4.5}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx. \quad (4.3)$$

This integration can be estimated directly using MC methods. However, there are problems when we simulate a sample from a uniform distribution or normal distribution in the region of $(4.5;\infty)$. If we generate samples from a uniform distribution in this large area of $(4.5;\infty)$, it is required a lot of points. If we generate samples from normal distribution in this area, we need to generate even more points to reach over the value of 4.5 because $P(X > 4.5) = 3.397673 \times 10^{-6}$ is very small.

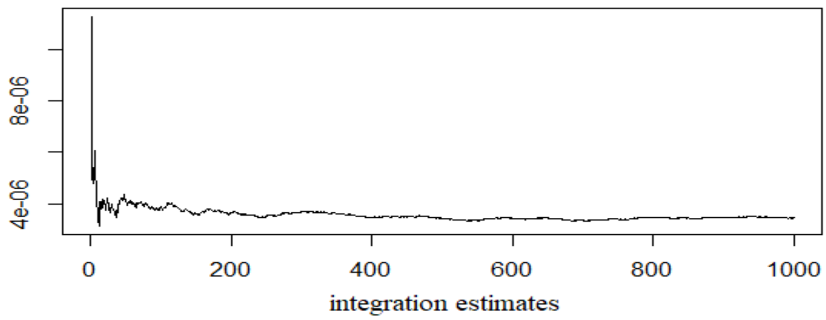

Figure 5. The integration estimates for the function in (4.3)

Let consider $X \sim Exp(1)$ and a truncated exponential function has the form of

$$g(x) = \frac{e^{-x}}{\int_{4.5}^{\infty} e^{-x} dx} = e^{-(x-4.5)}.$$

We use formula in (4.2) to calculate the integration in (4.3) with the syntax

Step1: Generate N sample points following $g(x)$ distribution

Step 2: Calculate the values of $f(x)/g(x)$ at these points

Step 3: Sum all values of $f(x_i)/g(x_i)$ and divided by N

After running this syntax for 1.000 sample points, we have the result as

$$\int_{4.5}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx \approx \frac{1}{N} \sum_{i=1}^{N} \frac{f(x_i)}{g(x_i)} = 0.000003374502.$$

In Figure 5, it can be seen that with just about 200 random values sample from $g(x)$ distribution, value of the estimate is stable and converge to the integration value.

5. CONCLUSION

Calculating intractable integrals is an important task for making inferences in statistics. These integration methods can help to compute probabilities for complex target distributions. Among these methods, the MC integration method shows its advantages in handling infinite integration regions and improving the accuracy of calculations. To have more precise results, we need to conduct Monte Carlo sequences in parallel. By doing this, the estimated integrals will converge quickly with fewer samples. Therefore, when calculating for intractable integrals we should use the proposed methods. In Bayesian statistics, the target posterior distributions usually have complex forms corresponding with hierarchical models. Applying the proposed MC integration method for Bayesian statistics with specific models is our future work.

REFERENCE

Cox, D.R., & Hinkley, D.V. (1979). Theoretical Statistics, Chapman & Hall, New York.

Choi, H. M., & Hobert, J. P. (2013). The polya-gamma gibbs sampler for Bayesian logistic regression is uniformly ergodic, *Electronic journal of Statistics*, 7, 2054-2064.

David, J. and Rabinowitz, P (1984). Methods of Numerical integration. Elsevier, United states of America.

Gelman, A., Carlin, J.B., Stern, H. S., & Rubin, D.B. (1995). Bayesian Data Analysis, Chapman & Hall, London.

Hoa, Pham., & Huong, T. T. Pham. (2021). Generating random numbers for standard and non-standard posterior distributions. *AGU International Journal of Sciences*.

Hastings, W.K. (1970). Monte Carlo sampling methods using markov chains and their applications, Biometrika 57(1): 97–109.

James, F.E. (2013). An introduction to numerical methods and analysis. Wiley.and Son.

Marin, J., & Robert, C. P. (2014). Bayesian essentials with R, Springer New York. London.

Robert, C., & Casella, G. (2004). Monte Carlo Statistical Methods, Springer-Verlag, NewYork.

Wakefield, J. (2013). Bayesian and Frequentist Regression Methods, Springer Science & Business Media, New York.