ỨNG DỤNG CỦA PHƯƠNG PHÁP MÔ PHỎNG MONTE CARLO VỚI XÁC SUẤT HÌNH HỌC

Nguyễn Thị Quyên
Khoa Toán và Khoa học tự nhiên
Email: quyennt@dhhp.edu.vn

Ngày nhận bài: 25/11/2021
Ngày PB đánh giá: 03/01/2022
Ngày duyệt đăng: 07/01/2022

TÓM TẮT: Trong bài này, bằng sự trợ giúp của phần mềm R, chúng tôi sử dụng nghiên cứu mô phỏng Monte Carlo để tính xấp xỉ xác suất một số biến cố - dựa vào định nghĩa xác suất hình học. Từ đó, chúng tôi đưa ra cách xấp xỉ một số tích phân, số siêu việt.

Từ khóa : xác suất hình học, Monte Carlo, mô phỏng, phần mềm R

APPLICATIONS OF MONTE CARLO SIMULATION METHOD WITH GEOMETRIC PROBABILITY

ABSTRACT: In this paper, by using R softwave, we use Monte Carlo simulation to approximate probability of events - based on geometric probability. Since then, we give an approximation of some integral values and transcendental numbers

Keywords: geometric probability, Monte Carlo, simulation, R software.

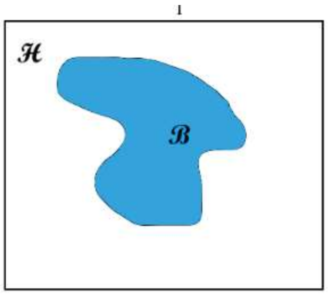
1. GIỚI THIỆU

Trong khoa học, để kiểm chứng một đề xuất mới nào đó, người ta sử dụng các luận cứ để chứng minh về mặt lý thuyết. Tuy nhiên, trong phân tích số liệu, ngoài việc chứng minh tính đúng đắn về mặt lý thuyết, người ta thường phân tích, minh họa kết luận bằng cách sinh ra bộ số liệu ngẫu nhiên đảm bảo giả thiết đã nêu, sau đó kiểm chứng lại bằng cách đếm số lần "thành công" (kết luận đúng) trong tổng số lần thực hiện – đó là phương pháp mô phỏng Monte Carlo.

Phương pháp này được giới thiệu trong [2], được hai nhà toán học người Mỹ gốc Hungari là John von Neumann (3/12/1903 – 8/2/1957) và Stanislaw Ulam (sinh năm 1909) nêu thành một phương pháp có cơ sở toán học chặt chẽ vào năm 1949. Nó là phương pháp mô hình hóa rất đặc dụng trong việc giải quyết nhiều bài toán khoa học kỹ thuật, nhất là trong dự báo các quá trình diễn biến phức tạp xảy ra trong thiên nhiên cũng như trong xã hội. Đặc biệt, từ khi có máy tính điện tử, thì nó này trở nên có hiệu quả và được sử dụng rộng rãi nhờ các mô hình hóa quá trình mô phỏng trên máy tính điện tử.

Những ý tưởng đầu tiên dẫn đến phương pháp Monte Carlo rất đơn giản. Chúng ta hãy xét bài toán nhỏ sau đây:

Cho một miền bất kì $\mathcal{B}$ trong hình vuông $\mathcal{H}$ có cạnh bằng 1 đơn vị độ dài. Ta cần tính diện tích miền $\mathcal{B}$, được minh họa bởi Hình 1.



Hình 1.

Để tính diện tích miền $\mathcal{B}$, ta có thể sử dụng các cách như phân chia theo các hình vuông đơn vị nhỏ hơn, thực hiện đếm và xấp xỉ lượng các hình vuông đó; hoặc có thể biểu diễn xấp xỉ chu tuyến $\mathcal{B}$ theo một phương trình toán học rồi tính bằng tích phân. Tuy nhiên, ta sẽ xét cách xấp xỉ diện tích $\mathcal{B}$ bằng phương pháp mô phỏng Monte Carlo như sau.

Tung $n_{\mathcal{H}}$ hạt cát một cách ngẫu nhiên lên miền $\mathcal{H}$; khả năng hạt cát rơi vào bất cứ vị trí nào trong miền $\mathcal{H}$ đều như nhau và chắc chắn hạt cát chỉ rơi vào miền $\mathcal{H}$; ta xác định được $n_{\mathcal{B}}$ hạt cát rơi vào miền $\mathcal{B}$. Theo các xác định của xác suất thực nghiệm (định nghĩa xác suất bằng thống kê) xác suất để hạt cát rơi vào miền $\mathcal{B}$ là: $\frac{S_{\mathcal{B}}}{S_{\mathcal{H}}} \simeq \frac{n_{\mathcal{B}}}{n_{\mathcal{H}}}$. Vì $S_{\mathcal{H}} = 1$ nên ta có xấp xỉ $S_{\mathcal{B}} \simeq \frac{n_{\mathcal{B}}}{n_{\mathcal{H}}}$.

Rõ ràng, vấn đề "tung ngẫu nhiên" và "khả năng hạt cát rơi vào bất cứ vị trí nào trong miền $\mathcal{H}$ đều như nhau và chắc chắn hạt cát chỉ rơi vào miền $\mathcal{H}$ hoàn toàn được đảm bảo nếu ta mô hình hóa toán học bởi các điểm $M(x, y)$ trong miền $\mathcal{H}$ (là hình vuông có độ dài cạnh bằng 1 đơn vị dài). Khi đó, về mặt toán học, có thể hiểu x, y được sinh ngẫu nhiên từ phân bố đều trên [0,1]- việc này được thực hiện thông qua một số phần mềm thống kê như Matlab, SPSS hay R. Trong bài viết này, chúng tôi sử dụng phần mềm R (nguồn tại [5]) để thực hiện mô phỏng bởi tính đơn giản, dễ hiểu và tính "mở" của R. Việc tính diện tích miền $\mathcal{B}$ (và có thể một miền bất kì trong một không gian bất kì) theo cách xấp xỉ ở trên được dựa trên cơ sở xác suất hình học được trình bày sau đây.

Ta đã biết, để tính xác suất của một biến cố người ta có thể sử dụng một số định nghĩa xác suất: định nghĩa cổ điển, định nghĩa theo thống kê, định nghĩa bằng hình học. Mỗi cách định nghĩa được áp dụng trong điều kiện nhất định: định nghĩa xác suất cổ điển được sử dụng khi số kết quả của phép thử là hữu hạn và đồng khả năng; phương pháp định nghĩa bằng thống kê là phương pháp định nghĩa thực nghiệm của xác suất, được sử dụng để dự báo xác suất của biến cố. Định nghĩa này được dựa trên kết quả của luật số lớn: khi số phép thử lặp lại đủ lớn thì tần suất xuất hiện biến cố sẽ hội tụ hầu chắc chắn đến xác suất của biến cố đó. Tuy nhiên, trong trường hợp các kết quả là đồng khả năng nhưng vô hạn, ví dụ như ném một vật ra xa, tung một vật nào đó trong phạm vi miền nào đó, thì sử dụng xác suất hình học là một giải pháp, định nghĩa được phát biểu như sau (xem [1, 3, 4]).

Giả sử phép thử có các biến cố sơ cấp đồng khả năng, mỗi biến cố sơ cấp được biểu diễn bởi một điểm hình học trong một miền $\mathcal{H}$ nào đó, biến cố B được biểu diễn bởi miền $\mathcal{B}$ là một phần của miền $\mathcal{H}$. Khi đó, xác suất của B được định nghĩa

$$P(B) = \frac{m(\mathcal{B})}{m(\mathcal{H})},$$

ở đó $m(\mathcal{B})$, $m(\mathcal{H})$ là độ đo của miền $\mathcal{B}$, $\mathcal{H}$. Độ đo ở đây có thể được hiểu một cách đơn giản là độ dài (trong không gian một chiều), diện tích (trong không gian hai chiều) và thể tích (trong không gian ba chiều).

2. ỨNG DỤNG CỦA PHƯƠNG PHÁP MONTE CARLO VỚI XÁC SUẤT HÌNH HỌC

Từ định nghĩa xác suất hình học, nếu ta biết được độ đo $m(\mathcal{H})$ của miền $\mathcal{H}$, thì xác suất P(B) và độ đo $m(\mathcal{B})$ hoàn toàn được xác định qua nhau. Trong mục này, ta sẽ xét một số ví dụ để xác định xấp xỉ $m(\mathcal{B})$ thông qua việc xấp xỉ P(B) và ngược lại.

Trước hết, bằng phương pháp Monte Carlo, ta đếm được số "hạt cát" $n_{\mathcal{B}}$ sẽ rơi vào miền $\mathcal{B}$ khi tung ngẫu nhiên $n_{\mathcal{H}}$ "hạt cát" vào miền $\mathcal{H}$ thì

$$\frac{n_{\mathcal{B}}}{n_{\mathcal{H}}} \simeq \frac{m(\mathcal{B})}{m(\mathcal{H})},$$

nên $m(\mathcal{B}) \simeq m(\mathcal{H}).\dfrac{n_{\mathcal{B}}}{n_{\mathcal{H}}}$.

Như vậy, bằng việc "lựa chọn" miền $\mathcal{H}$ sao cho dễ "đo được" nhất, ta hoàn toàn xấp xỉ được độ đo của miền $\mathcal{B}$ thông qua việc kiểm đếm các "hạt cát". Việc kiểm đếm này được thực hiện trên phần mềm R.

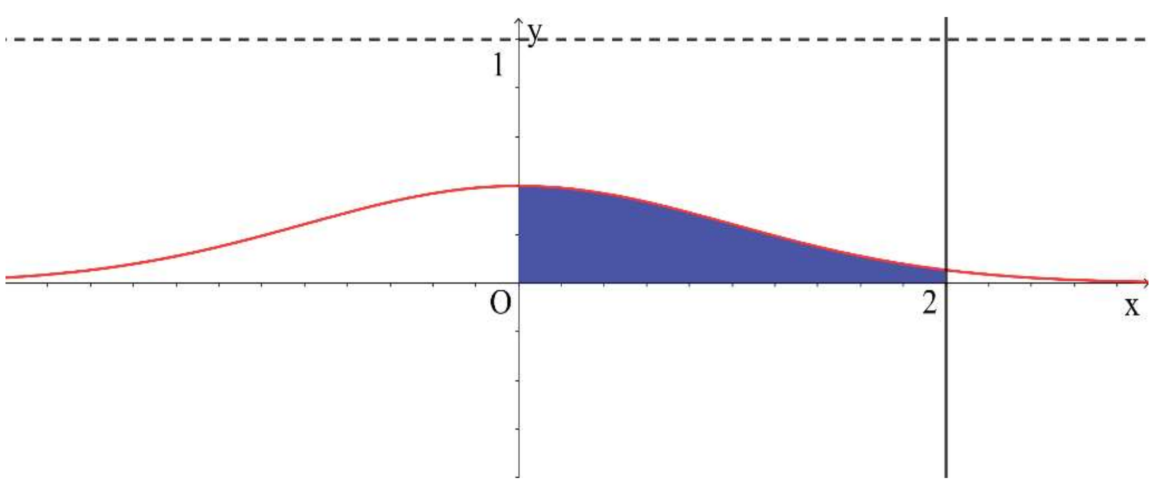
Trước hết, chúng ta đã biết, phân bố chuẩn là phân bố khá thông dụng trong thực tiễn, nó chính là phân bố giới hạn của lượng lớn các phân bố bất kì thỏa mãn điều kiện luật số lớn. Tuy nhiên, việc tính xác suất của phân bố chuẩn lại phải thông qua hàm Laplace $\Phi_0(u) = \dfrac{1}{2\sqrt{\pi}}\int_0^u e^{-\frac{x^2}{2}}dx$. Rõ ràng tích phân này không thể thực hiện bằng công thức Newton-Leibniz. Ta xét ví dụ sau.

Ví dụ 1. Tính xấp xỉ $I = \dfrac{1}{\sqrt{2\pi}}\int_0^2 e^{-\frac{x^2}{2}}dx$.

Để xấp xỉ được tích phân trên, trước hết, ta "lựa chọn" miền $\mathcal{H}$

$$\mathcal{H} = \{(x, y) : 0 \le x \le 2,\ 0 \le y \le 1\}.$$

Khi đó, độ đo ở đây được hiểu là diện tích với $m(\mathcal{H}) = 2$ và tích phân I là diện tích của miền $\mathcal{B} = \left\{(x, y) \in \mathcal{H} : y < \dfrac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}\right\}$ được mô tả trong Hình 2.



Hình 2.

Bằng cách sinh ngẫu nhiên 10000 cặp (x, y) trong miền $\mathcal{H}$ và đếm số cặp (x, y) nằm trong miền $\mathcal{B}$ với code trong R:

```
> x=runif(10000,0,2)
> y=runif(10000,0,1)
> z=y-1/(sqrt(2*pi))*exp(-x^2/2)
> length(subset(z,z<=0))
 [1] 2344
```

Ta thu được 2344 cặp rơi vào miền $\mathcal{B}$ trong số 10000 cặp (x, y) được sinh ngẫu nhiên trong miền $\mathcal{H}$. Từ đó

$$\frac{m(\mathcal{B})}{m(\mathcal{H})} \simeq \frac{2344}{10000} = 0.2344 \Rightarrow m(\mathcal{B}) \simeq 0.4688$$

hay tích phân I được xấp xỉ bởi 0.4688.

Để tránh sai số, ta lặp lại số lần lấy mẫu và lấy giá trị trung bình với code:

```
>h=function(n, m)
+ {t=numeric(n)
+ for (i in 1: n)
+ {x=runif(m,0,2)
+ y=runif(m,0,1)
+ z=y-1/(sqrt(2*pi))*exp(-x^2/2)
+ t[i]= length(subset(z,z<=0))}
+ mean(t)}
> h(1000,10000)
 [1] 2387.621
```

Cho lặp lại 1000 lần lấy mẫu, kích thước mẫu mỗi lần là 10000 thu được kết quả:

```
> h(1000,10000)
 [1] 2387.621
```

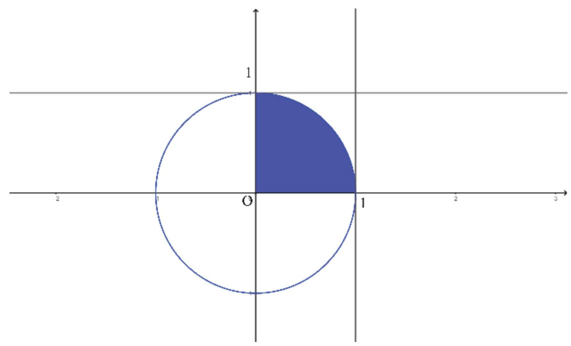
nên tích phân I được xấp xỉ 0.4775242. Kết quả này rất sát với kết quả bảng tra hàm Laplace vẫn được dùng trong xác suất thống kê (0.4772). Để được kết quả chính xác hơn, ta có thể cho lặp lại số lần lấy mẫu lớn hơn và kích thước mẫu được chọn đủ lớn.

Ví dụ 2. Tính xấp xỉ giá trị của π.

Để tính xấp xỉ này, ta sẽ sử dụng công cụ diện tích, cụ thể ở đây là diện tích hình tròn. Để đơn giản ta sẽ dựa vào tính xấp xỉ diện tích góc phần tư thứ nhất của hình tròn. Miền $\mathcal{H}$, $\mathcal{B}$ được chọn tương ứng và mô tả ở Hình 3. là

$$\mathcal{H} = \{(x, y) : 0 < x, y < 1\},$$

$$\mathcal{B} = \{(x, y) : x > 0,\ y > 0,\ x^2 + y^2 < 1\}.$$



Hình 3.

Theo lập luận trên, ta có

$$\frac{n_{\mathcal{B}}}{n_{\mathcal{H}}} \simeq \frac{S_{\mathcal{B}}}{S_{\mathcal{H}}} = \frac{\pi}{4} \Leftrightarrow \pi \simeq \frac{4n_{\mathcal{B}}}{n_{\mathcal{H}}}.$$

Sinh ra 10000 cặp (x, y) trong đó x, y được sinh ngẫu nhiên từ phân bố đều trên $(0, 1)$, ta đếm số lượng điểm (x, y) rơi vào một phần tư hình tròn thứ nhất bằng code trong R như sau:

```
> x=runif(10000,0,1)
> y=runif(10000,0,1)
> z=x^2+y^2
> length(subset(z,z<1))
[1] 7853
```

Ta nhận được kết quả: 7853, nên diện tích $\frac{1}{4}$ đường tròn được xấp xỉ bởi 0.7853

suy ra $\pi \simeq \frac{4 \times 7853}{10000} = 3.1412.$

Tương tự như Ví dụ 1, ta lặp lại 1000 lần để tránh sai số với code:

```
> t=numeric(1000)
> for (i in 1: 1000)
+ {x=runif(10000,0,1)
+ y=runif(10000,0,1)
+ z=x^2+y^2
+ t[i]=length(subset(z,z<1))/10000}
> mean(t)
[1] 0.785398
```

Nên $\pi \simeq 4 \times 0.785398 = 3.141592$.

Việc tính xấp xỉ xác suất bởi mô phỏng được thể hiện rõ trong những bài toán mà việc vẽ hình trở nên khó khăn. Cụ thể, ta xét ví dụ sau

Ví dụ 3. Lấy ngẫu nhiên ba đoạn thẳng, mỗi đoạn thẳng có độ dài không vượt quá số $a > 0$ cho trước. Tìm xác suất sao cho các đoạn thẳng đó là các cạnh: của một tam giác? của một tam giác tù?

Gọi x, y, z là độ dài các đoạn thẳng, không gian mẫu được mô tả bởi

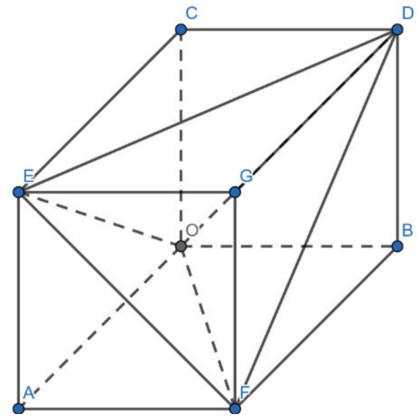
$$\mathcal{H} = \{(x, y, z) : 0 < x, y, z \le a\}.$$

Biến cố B: các đoạn thẳng đó là các cạnh của một tam giác, được mô tả

$$\mathcal{B} = \{(x, y, z) \in \mathcal{H} : x + y > z, y + z > x, z + x > y\}.$$

Nếu xem (x, y, z) là tọa độ của một điểm trong không gian ba chiều thì miền $\mathcal{H}$ là hình lập phương cạnh a, còn miền $\mathcal{B}$ là hình OEDFG (Hình 4.) nên

$$P(B) = \frac{m(\mathcal{B})}{m(\mathcal{H})} = \frac{V_{ODEFG}}{V_{OAFBCEGD}} = \frac{1}{2}.$$



Hình 4.

Nếu sử dụng phương pháp mô phỏng Monte Carlo, ta có thể coi a=1 và sinh bộ ba (x, y, z) từ phân bố đều trên $(0,1]$ sau đó kiểm đếm số bộ (x, y, z) thỏa mãn điều kiện tam giác. Cụ thể, với việc sinh ra 100000 bộ ta dùng R với code:

```
t=rep(0,100000)
> for (i in 1:100000)
+ {x=runif(3,0,1)
+ if((x[1]+x[2]>x[3])& (x[2]+x[3]>x[1])& (x[1]+x[3]>x[2])){t[i]=1}}
> sum(t)
[1] 49772
```

Như vậy, ta xấp xỉ được xác suất để ba đoạn thẳng tạo thành tam giác là 0.49772.

Cũng để tránh sai số, ta lặp lại 1000 lần và tính giá trị trung bình với code trong R

```
> u=numeric(100)
> for (j in 1:100)
+ {t=rep(0,100000)
+ for (i in 1:100000)
+ {x=runif(3,0,1)
+if((x[1]+x[2]>x[3])&(x[2]+x[3]>x[1])& (x[1]+x[3]>x[2])){t[i]=1}}
+ u[j]=sum(t)}
> mean(u)
[1] 50020
```

Từ đó, xác suất được xấp xỉ 0.50020 – giá trị này gần sát với giá trị tính trực tiếp.

Ta nhận thấy, trong trường hợp này, việc vẽ hình và xác định miền $\mathcal{B}$ khá dễ dàng nên việc tính xác suất trực tiếp khá đơn giản. Tuy nhiên, nếu trong trường hợp các đoạn thẳng lấy ra là các cạnh của một tam giác tù, và đặc biệt nếu tam giác tạo thành có thêm nhiều điều kiện khác, thì việc vẽ hình không hề đơn giản. Việc tính xấp xỉ bằng mô phỏng Monte Carlo là một giải pháp.

Biến cố B': ba đoạn thẳng đó là các cạnh của một tam giác tù, được mô tả:

$$\mathscr{B}' = \{(x, y, z) \in \mathscr{B} : x^2 > y^2 + z^2 \text{ hoặc } y^2 > z^2 + x^2 \text{ hoặc } z^2 > x^2 + z^2\}.$$

Để xấp xỉ xác suất P(B') ta dùng code trong R

```
> h=function(n)
+ {t=rep(0,n)
+ for (i in 1:n)
+ { x=runif(3,0,1)
+ if((x[1]+x[2]>x[3])& (x[2]+x[3]>x[1])& (x[1]+x[3]>x[2])&
+ (( x[1]^2+x[2]^2<x[3]^2) || (x[2]^2+x[3]^2<x[1]^2) || (x[3]^2+x[1]^2<x[2]^2)))
+ {t[i]=1}}
+ sum(t)}
```

Với việc sinh ra mẫu kích thước n=100; 1000; 10000; 10000 ta có được các kết quả tương ứng là 25; 296; 2841; 28397 nên xác suất để ba đoạn thẳng tạo thành tam giác tù là: 0.25; 0.296; 0.2841; 0.28397.

Để dự đoán và xấp xỉ chính xác hơn, ta lặp lại số lần lấy mẫu với các kích thước mẫu đủ lớn bằng code trong R:

```
> h=function(n,m)
+ {u=numeric(n)
+ for (j in 1:n)
+ {t=rep(0,m)
+ for (i in 1:m)
+ {x=runif(3,0,1)
+ if((x[1]+x[2]>x[3])& (x[2]+x[3]>x[1])& (x[1]+x[3]>x[2])&
+ (( x[1]^2+x[2]^2<x[3]^2) || (x[2]^2+x[3]^2<x[1]^2) || (x[3]^2+x[1]^2<x[2]^2)))
+ {t[i]=1}}
+ u[j]=sum(t)/m}
+ mean(u)}
```

Các kết quả trả về xấp xỉ xác suất với 100 lần lấy mẫu với các kích thước mẫu trên lần lượt là 0.264; 0.28479; 0.2865, 0.2857; 0.28538. Như vậy, ta có thể xấp xỉ xác suất để ba đoạn thẳng tạo thành tam giác tù là 0.285.

3. KẾT LUẬN

Bằng việc sử dụng định nghĩa xác suất hình học và sự trợ giúp của phần mềm R, nghiên cứu mô phỏng Monte Carlo được sử dụng để tính xấp xỉ tích phân của những hàm không thể lấy nguyên hàm, số siêu việt π; xấp xỉ xác suất của các biến cố mà việc mô hình hóa nó bằng các hình hình học phức tạp. Các kết quả xấp xỉ khá gần với giá trị thực, nếu muốn độ chính xác lớn hơn, bạn đọc có thể sinh ra nhiều cặp hơn với số lần lặp lại nhiều hơn.

TÀI LIỆU THAM KHẢO

1. Nguyễn Văn Hữu, Hoàng Hữu Như, *Bài tập lý thuyết xác suất và thống kê toán*, Nhà xuất bản Đại học và Trung học chuyên nghiệp Hà Nội, 1976.

2. Nguyễn Quý Hỷ (2004), *Phương pháp mô phỏng số Monte Carlo*, NXB Đại học Quốc gia Hà Nội.

3. Đặng Hùng Thắng (1998), *Mở đầu về xác suất và các ứng dụng*, NXB Giáo dục.

4. Herbert Solomon (2016), *Geometric probability*, Published online by Cambridge University Press.

5. https://cran.r-project.org