Método de Monte Carlo. Técnica de redução da variância

Monte Carlo Method. Variance reduction technique

Ana Vergínia Libos Messetti*
Simone de Castro Queiroz**

* Mestre em Agronomia (Energia na Agricultura, UNESP/Botucatu)
Docente do Departamento de Matemática Aplicada do Centro de Ciências Exatas da Universidade Estadual de Londrina (UEL)
** Doutora em Agronomia (Energia na Agricultura, UNESP/Botucatu)
Docente do Departamento de Matemática Aplicada do Centro de Ciências Exatas da Universidade Estadual de Londrina (UEL)
e-mail: <siqueiroz@uol.com.br>

Resumo

O presente trabalho tem como objetivo mostrar que o método Monte Carlo ou Simulação é um método numérico auxiliar, para resolver problemas de cálculo de integração no qual não é possível obter solução de forma analítica, ou mesmo usando solução numérica, se torna inviável. Na introdução apresenta-se a condição necessária para aplicação do método de Monte Carlo que para estimar área aproximada com certa precisão deve-se trabalhar com números aleatórios e uniformemente distribuídos. Por ser processo de geração de números aleatórios, uma parte fundamental na aplicação do método de Monte Carlo, foi realizado em materiais e métodos um exemplo do cálculo de uma integral, no qual possui solução analítica para que se possa comparar o resultado do método com o verdadeiro valor da área, utilizando duas técnicas de redução de variância. Finalmente concluiu-se que o Método Monte Carlo é eficiente, de fácil aplicação e deve ser utilizado sempre que não for possível estimar uma área analiticamente. Observa-se que com os conceitos da estatística básica, é possível fazer a comparação entre as duas técnicas utilizadas na aplicação, e o resultado apresentou que o erro, ocorrido no Método Amostragem Estratificada para atingir a verdadeira área, foi muito menor em relação ao erro do Método Primitivo.

Palavras-chave: Monte Carlo, simulação, integração, números aleatórios, redução de variância.

Abstract

The aim of the present work was to demonstrate that Monte Carlo Method or Simulation is an auxiliary numerical method to solve problems of integration when it is not possible to obtain a solution in analytical way or through numerical solution. In the introduction, was established a necessary condition for its application in estimating an area with certain precision. This condition was to deal with aleatory numbers evenly distributed. Being a process of random number generation, an important part of the experiment was to calculate the integral, on which the analytical solution could be compared to the result obtained by the method with the true value of the area, using two techniques of variance reduction. Thus, it was demonstrated that Monte Carlo Method is efficient, of easy application and it can be used whenever it is not possible to estimate some area analytically. It might be noticed that, with a few basic statistics concepts, it was possible to compare the two techniques, and the result demonstrated that the error, detected by using the Sampling Stratified Method to obtain the true area, was much smaller than the one resulting from the primitive method.

***Key words**: Monte Carlo, simulation, integration, aleatory numbers, variance reduction.*

Introdução

O nome e o desenvolvimento sistemático do método de Monte Carlo datam de aproximadamente 1944, embora, de modo assistemático e não nominado, a referida teoria já tivesse sido empregada em casos isolados de resolução de problemas de ordem determinística e probabilística.

O nome de “MONTE CARLO” se deve a cidade no principado de Mônaco, por causa de uma roleta, um gerador de números aleatórios. Em 1899, Lorde Rayleigh mostrou que uma variável aleatória satisfaz sem problema uma solução aproximada para uma equação diferencial parabólica. Em 1931, Kolmogorov mostrou a relação entre o processo estocástico de Markov e certas equações diferencial integral.

O uso real dos métodos de Monte Carlo, como uma ferramenta de pesquisa, originam de trabalho na bomba atômica durante a segunda Guerra Mundial. Este trabalho envolveu uma simulação direta dos problemas de probabilidade, tendo a ver com a difusão de nêutron aleatório em material fundível; mas até mesmo em estágios anteriores a essas investigações, Von Neumann e Ulam refinaram essa particular “roleta russa” e “intensos” métodos.

Em aproximadamente 1970, a recente teoria em desenvolvimento de complexidade computacional começou a prover uma razão mais precisa e persuasiva para empregar o método de Monte Carlo. Podemos dizer que o método Monte Carlo é um método numérico auxiliar, para resolver problemas matemáticos mediante simulação de variáveis aleatórias, na qual não é possível obter solução de forma analítica, ou mesmo usando solução numérica se torna inviável. Porém, nem todos os problemas são solucionados por Monte Carlo, pois nem sempre fornece valores que se encaixam dentro de limites razoáveis.

Para termos uma idéia mais clara, consideremos o exemplo: seja a Figura 1 compreendida dentro de um quadrado de lado 1.

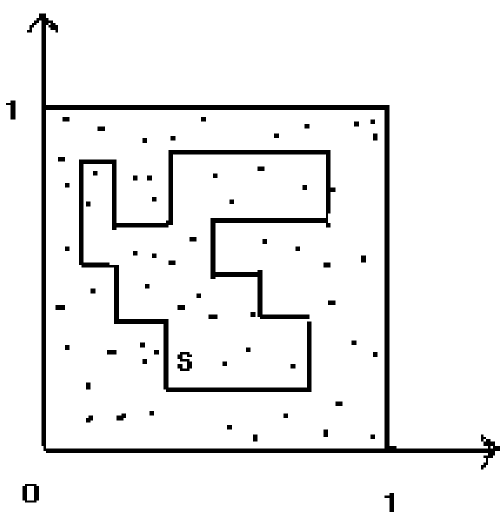

Figura 1

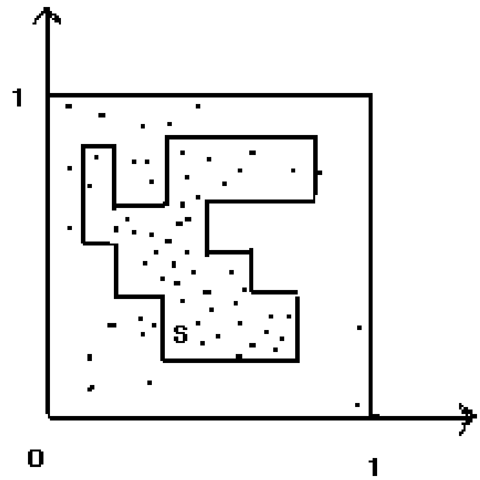

Figura 2

Para obtermos a área da Figura 1, tomemos N pontos aleatórios dentro do quadrado. Seja N’ o número de pontos dentro da figura S. Geometricamente, a área de S é aproximadamente N’/N. Quando N for muito grande, mais próximos estaremos da área verda-deira, isto é, menos erro estaremos cometendo.

Pelo exemplo da figura anterior, N = 60 pontos e N’= 18 pontos, assim obtemos uma área aproximada de N’/N = 18/60 = 030, onde a área exata é 0,36, obtendo um bom resultado.

Seja o experimento da Figura 1. Tomemos um bom atirador situado a uma certa distância dispara N vezes dentro do quadrado. Pela Figura 2, se ele atira 60 vezes e acerta em S 42 vezes, então N’/N= 42/60= 0,7, ocorrendo um grande erro na área de S. Assim, para obtermos uma boa estimativa da área S, devemos trabalhar somente com pontos aleatórios, que por serem aleatórios, estarão distribuídos uniformemente dentro do quadrado.

1 Aplicação

1.1 Aplicação do método de Monte Carlo para cálculo de integrais

O problema estatístico mais comum que requer integração é de encontrar algum parâmetro que descreva uma distribuição de probabilidade. Geral-mente não é simples descrever a função densidade de probabilidade. Assim, simulação, ou Método de Monte Carlo, pode solucionar tal problema, com certa precisão. Sempre devemos levar em conta que essa precisão seja mínima, ou que a variância seja mínima.

Consideremos a função g(x),definida no intervalo (a, b). Queremos calcular o valor aproximado da integral

$$\theta = \int_a^b g(x)dx$$

Seja uma função $f_\xi(x)$ de densidade de probabilidade em (a, b) , onde ξ é uma variável aleatória, definida por $\eta = g(\xi) / f_\xi(\xi)$

Assim temos,

$$E(\eta) = \int_a^b \left[g(x) / f_\xi(x)\right] f_\xi(x)dx = \theta$$

Consideremos agora n variáveis aleatórias $\eta_1, \eta_2, ..., \eta_n$ independentes e uniformemente distribuídas em (a, b), e aplicamos a soma dessas variáveis. O teorema do limite central afirma que qualquer que seja o intervalo (a, b) se tem para valores grandes de n

$$P\left[\left|\frac{1}{n}\sum_{j=1}^{n}\eta_j - \theta\right| < 3\sqrt{\frac{V(\eta)}{n}}\right] \approx 0{,}997$$

Escolhendo n valores $\xi_1, _2, \xi..., \xi_n$ (onde cada um deles é igual aos números x_1, x_2, ...x_n). Para n suficientemente grande, temos

$$\frac{1}{n}\sum_{j=1}^{n}\frac{g(\xi_j)}{f_\xi(\xi_j)} \approx \theta$$

Temos que essa aproximação possui uma elevada probabilidade de erro que não passa de

$$3\sqrt{V(\eta)/n}$$

Para escolher o esquema de cálculo, tomemos uma variável aleatória qualquer ξ definida, no intervalo (a, b). Assim em todos os casos temos

$$E(\eta) = E\left[g(\xi) / f_\xi(\xi)\right] = \theta$$

Pela variância $V(\eta)$ e estimativa de θ o erro depende de como tomemos a variável ξ . Com efeito

$$V(\eta) = E(\eta^2) - \theta^2 = \int_a^b \left[g^2(x) / f_\xi(x)\right]dx - \theta^2$$

onde essa expressão será mínima quando $f_\xi(x)$ seja proporcional a $|g(x)|$.

Para reduzir a variância usamos algumas técnicas de redução da variância, que, como já vimos podem depender do modelo matemático e das técnicas de geração de variável aleatória.

As principais técnicas de redução de variância são as seguintes:

– Monte Carlo primitivo (grude Monte Carlo)
– Monte Carlo Certo-ou-Errado
– Amostragem Estratificada
– Amostragem de Importância
– Variáveis de Controle
– Variáveis Antitéticas
– Roleta Russa e fracionamento.

Iremos exemplificar apenas duas técnicas a seguir.

Consideremos o problema de calcular numericamente uma integral da forma

$\theta = \int_0^1 g(x)dx$ onde $g(x) = \sqrt{1-x^2}$, ou seja, a área hachurada da Figura 3. Assim analiticamente

$$\theta = \int_0^1 \sqrt{1-x^2}dx = \left(\frac{x}{2}\sqrt{1-x^2} + \frac{1}{2}\operatorname{arcsen} x\right)\Bigg|_0^1 = \frac{\pi}{4}$$

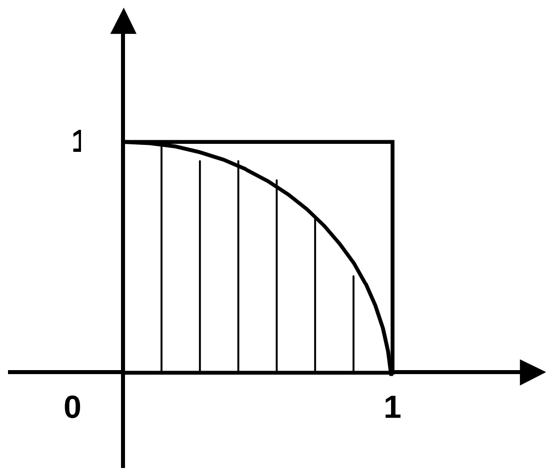

Figura 3

2 Monte Carlo Primitivo (Grude Monte Carlo)

Consideremos os números aleatórios x_1, x_2, ..., x_n independentes e uniformemente distribuídos entre 0 e 1. Então, os valores $g(x_j)$, j=1, 2, ..., n, são variáveis aleatórias independentes com esperança

$$E(g(x)) = \int_0^1 g(x)f(x)dx = \int_0^1 g(x)dx = \theta$$

onde $f(x) = \dfrac{1}{1-0}$ ou seja f(x) é uniforme em (0, 1).

Assim temos um estimador não-viesado de θ

$$\bar{g} = \frac{1}{n}\sum_{j=1}^{n} g(x_j)$$

e sua variância é

$$\frac{\sigma^2}{n} = \frac{1}{n}\int_0^1 [g(x)-\theta]^2 dx$$

e analiticamente

$$\frac{\sigma^2}{n} = \frac{1}{n}\int_0^1\left[\left(g(x)\right)^2 - \frac{\pi}{2}g(x) + \left(\frac{\pi}{4}\right)^2\right] =$$

$$\frac{1}{n}\left[\left(x - \frac{x^3}{3}\right)\Bigg|_0^1 - \frac{\pi}{2}\left(\frac{\pi}{4}\right) + \left(\frac{\pi}{4}\right)^2 x\Bigg|_0^1\right] = 0{,}0498/n$$

E tomando-se 100 números aleatórios (x_j), independentes e uniformemente distribuídos entre (0,1), conforme a Tabela 2.

Tabela 1 – Números aleatórios x_i, com as respectivas variáveis aleatórias g(x).

Ordem i	Xi	g(x)
1	0,030	1,00
2	0,990	0,141
3	0,110	0,994
4	0,040	0,999
5	0,610	0,792
6	0,930	0,368
7	0,710	0,704
8	0,610	0,792
9	0,680	0,733
10	0,940	0,341
11	0,980	0,199
12	0,950	0,312
13	0,370	0,929
14	0,320	0,947
15	0,310	0,951
16	0,390	0,921
17	0,520	0,854
18	0,870	0,493
19	0,240	0,971
20	0,840	0,543
21	0,820	0,572
22	0,470	0,883
23	0,420	0,908
24	0,550	0,835
25	0,930	0,368
26	0,240	0,971
27	0,120	0,993
28	0,260	0,966
29	0,650	0,760
30	0,910	0,415
31	0,270	0,963
32	0,690	0,724
33	0,900	0,436
34	0,640	0,768
35	0,940	0,341
36	0,030	1,000
37	0,150	0,989
38	0,210	0,978
39	0,910	0,415
40	0,210	0,978
41	0,340	0,940
42	0,410	0,912
43	0,480	0,877
44	0,210	0,978
45	0,570	0,822
46	0,860	0,510
47	0,880	0,475
48	0,750	0,661
49	0,500	0,866
50	0,870	0,493
51	0,880	0,475
52	0,610	0,792
53	0,810	0,586
54	0,910	0,415
55	0,610	0,792
56	0,410	0,912
57	0,290	0,957
58	0,060	0,998
59	0,730	0,683
60	0,120	0,993
59	0,730	0,683
60	0,120	0,993
61	0,740	0,673
62	0,850	0,527
63	0,710	0,704
64	0,590	0,807
65	0,570	0,822
66	0,480	0,877
67	0,540	0,842
70	0,470	0,883
71	0,180	0,984
72	0,610	0,792
73	0,910	0,415
74	0,360	0,933
75	0,740	0,673
76	0,900	0,436
77	0,890	0,456
78	0,970	0,243
79	0,570	0,822
80	0,540	0,842
81	0,600	0,800
82	0,830	0,558
83	0,320	0,947
84	0,590	0,807
85	0,830	0,558
86	0,010	1,000
87	0,290	0,957
88	0,140	0,990
89	0,130	0,992
90	0,490	0,872
91	0,190	0,982
92	0,150	0,989
93	0,200	0,980
94	0,020	1,000
95	0,230	0,973
96	0,120	0,993
97	0,300	0,954
98	0,280	0,960
99	0,070	0,998
100	0,830	0,557

Tabela 2 – Resultados da análise estatística Método Monte Carlo Primitivo.

Amostra N	$\mid g(x) - \theta \mid$	Variância	Erro padrão
100	0,0183	0,052	0,022

3 Monte Carlo por Amostragem Estratificada

Uma forma de melhorar o resultado do método de Monte Carlo Primitivo é subdividindo o intervalo (0, 1) em vários subintervalos (α_{j-1}, α_j), onde $0 = \alpha_1 < \alpha_2 < ... < \alpha_k = 1$, e a cada subintervalo aplicar o método.

Tomamos novamente 100 números aleatórios, mas subdivididos em 4 intervalos $\alpha_1 = 0$, $\alpha_2 = 0{,}25$, $\alpha_3 = 0{,}5$, $\alpha_4 = 0{,}75$ e $\alpha_5 = 1{,}0$. Assim, 25 números aleatórios em cada subintervalo e aplicamos o Método de Monte Carlo Primitivo, como mostra a Tabela 3.

Tabela 4 – Resultados da Análise Estatística. Método Monte Carlo – Amostragem Estratificada.

Amostra N	$\mid g(x) - \theta \mid$	Variância	Erro padrão
100	0,003	0,056	0,024

4 Resultados e Discussão

Os problemas de Monte Carlo e das estimativas estatísticas diferem na complexidade do modelo. Seria aconselhável trabalharmos com um modelo que não pode ser resolvido analiticamente, pois Monte Carlo é um método para obter respostas quando quase todos os meios analíticos são complexos e de difíceis soluções. Mas para confirmarmos realmente a redução da variância, consideramos uma função de área conhecida e, através de algumas análises estatísticas, verificamos os resultados dos parâmetros para comprovar se o método é realmente eficiente.

Como a diferença dos valores das áreas encontradas pelos dois métodos foi um erro muito pequeno, concluímos que o Método Monte Carlo é eficiente e deve ser utilizado sempre que não for possível estimar a área de uma função complexa.

Para comparar os dois métodos utilizados na aplicação, no primeiro momento tomamos uma amostra de 100 números aleatórios para cada método. O erro ocorrido no Método Amostragem Estratificada para atingir a verdadeira área foi muito menor em relação ao erro do Método Primitivo. Num segundo momento, e para enriquecer a afirmação, aplicamos o teste para diferença das duas médias amostrais e houve diferença significativa no nível de significância de 0,05. Logicamente, o Método Amostragem Estratificada foi mais eficiente, o que já era esperado, porque, reduzindo o espaço amostral em cada intervalo, se reduz também o erro.

4.1 Monte Carlo primitivo

– Utilizando a técnica de redução de variância – Monte Carlo Primitivo – o valor da esperança da variável aleatória 0 foi igual a 0,78539. O estimador não viesado da esperança é a média amostral g(x) com valor igual 0,7667. A diferença entre os dois valores é considerado o erro das médias 0,0187.

– Como foi demonstrado na metodologia, o valor da variância da variável aleatória foi igual a 0,0498 / n. Extraindo a raiz quadrada obtemos o erro padrão teórico e igual à 0,022.

– Comparando o resultado do item1 com resultado do item 2, concluímos que o Método Monte Carlo é eficiente quanto a aproximação do valor final da área.

Pode-se escrever o resultado de uma estimativa de g(x), num intervalo de confiança para a média da variável aleatória com o nível de significância de 5%:

$$P\ [\ 0,721198 < \theta < 0,81239\ ] = 95\%$$

4.2 Monte Carlo – Amostragem estratificada

– Utilizando a técnica de redução de variância, Monte Carlo - Amostragem Estratificada, dividimos o intervalo (0, 1) em 4 partes iguais, e a média amostral da variável aleatória foi igual a 0,7772.

– Como foi também demonstrado na metodologia, o valor da variância da variável aleatória nesse método foi igual a 0,0498. O erro padrão teórico foi igual a 0,0243.

– Comparando os itens 1 e 2, concluímos que o valor final da área está bem próximo do valor real, com o erro muito pequeno e melhorando muito o resultado final.

Pode-se escrever o resultado de uma estimativa de g(x), num intervalo de confiança para a média da variável aleatória com nível de significância de 5%:

$$P\ [\ 0,72890 < \theta < 0,82567\ ] = 95\%$$

4.3 Teste de significância para a igualdade das médias

– Hipóteses
H0: $\mu 1 = \mu 2$
H1: $\mu 1 \,^{1}\, \mu 2$

– Nível de significância = 0,05

– Estatística
Cálculo t = (0,7667 - 0,7772) / 0,2323* 0,1414 = -0,320

– Região Crítica
T tabelado (198 ; 5%) = ± 1,96

Tabela 3 – Números aleatórios xi, e as respectivas variáveis aleatórias g(x).

Ordem i	x_i	g(x)
1	0,010	1,000
2	0,020	1,000
3	0,030	1,000
4	0,030	1,000
5	0,030	1,000
6	0,040	0,999
7	0,050	0,999
8	0,060	0,998
9	0,100	0,995
10	0,100	0,995
11	0,100	0,995
12	0,110	0,994
13	0,140	0,990
14	0,140	0,990
15	0,150	0,989
16	0,150	0,989
17	0,150	0,989
18	0,170	0,985
19	0,180	0,984
20	0,180	0,984
21	0,190	0,982
22	0,190	0,982
23	0,190	0,982
24	0,200	0,980
25	0,220	0,975
26	0,260	0,966
27	0,260	0,966
28	0,270	0,963
29	0,270	0,963
30	0,290	0,957
31	0,300	0,954
32	0,300	0,954
33	0,320	0,947
34	0,320	0,947
35	0,320	0,947
36	0,330	0,944
37	0,330	0,944
38	0,340	0,940
39	0,340	0,940
40	0,340	0,940
41	0,380	0,925
42	0,410	0,912
43	0,410	0,912
44	0,420	0,908
45	0,440	0,898
46	0,450	0,893
47	0,450	0,893
48	0,480	0,877
49	0,480	0,877
50	0,480	0,877
51	0,530	0,848
52	0,540	0,842
53	0,540	0,842
54	0,540	0,842
55	0,540	0,842
56	0,550	0,835
57	0,550	0,835
58	0,590	0,807
59	0,600	0,800
60	0,610	0,792
61	0,610	0,792
62	0,610	0,792
63	0,610	0,792
64	0,620	0,785
65	0,630	0,777
66	0,640	0,768
67	0,650	0,760
68	0,670	0,742
69	0,670	0,742
70	0,700	0,714
71	0,720	0,694
72	0,720	0,694
73	0,720	0,694
74	0,740	0,673
75	0,740	0,673
76	0,750	0,661
77	0,770	0,638
78	0,780	0,626
79	0,780	0,626
80	0,800	0,600
81	0,810	0,586
82	0,820	0,572
83	0,840	0,543
84	0,840	0,543
85	0,860	0,510
86	0,870	0,493
87	0,890	0,456
88	0,920	0,392
89	0,930	0,368
90	0,930	0,368
91	0,930	0,368
92	0,950	0,312
93	0,950	0,312
94	0,960	0,280
95	0,970	0,243
96	0,980	0,199
97	0,980	0,199
98	0,990	0,141
99	0,990	0,141
100	0,990	0,141

Conclusão

Como t calculado < t tabelado, rejeita-se H0 e confirmamos que há diferença significativa entre os dois métodos.

1- MONTE CARLO PRIMITIVO - CRUDE

Amostra N	| g(x) - θ |	Variância	Erro padrão
100	0,0183	0,052	0,022
200	0,0066	0,049	0,015
300	0,0073	0,046	0,012
400	0,0003	0,046	0,010
500	0,0001	0,046	0,009

2- MONTE CARLO - AMOSTRAGEM ESTRATIFICADA

Amostra N	| g(x) - θ |	Variância	Erro padrão
100	0,003	0,056	0,024
200	0,001	0,052	0,016
300	0,003	0,054	0,013
400	0,009	0,056	0,011
500	0,005	0,055	0,010

1- MONTE CARLO PRIMITIVO - CRUDE

Amostra N	| g(x) - 0 |	Variância	Erro padrão
100	0,0183	0,052	0,022
200	0,0066	0,049	0,015
300	0,0073	0,046	0,012
400	0,0003	0,046	0,010
500	0,0001	0,046	0,009

2- MONTE CARLO - AMOSTRAGEM ESTRATIFICADA

Amostra N	| g(x) - 0 |	Variância	Erro padrão
100	0,003	0,056	0,0243
200	0,001	0,052	0,0162
300	0,003	0,054	0,0136
400	0,009	0,056	0,0140
500	0,005	0,055	0,0104

Referências

HAMMERSLEY, J.M.; HANDSCOMB, D.C. *Les Méthodes de monte Carlo*. Paris: Dunod, 1967.

SHIMIZU, T. *Método Monte Carlo*. São José dos Campos, 1996. Publicação Interna do laboratório de Processamento de Dados do ITA.

SÓBOL, I. M. *Método de Montecarlo*. Moscou -URSS: Ed. Mir, 1976.

THISTESTED, R. A. *Elements of statistical computing*. New York: Chapman and Hall, 1988.