CONGRESSO BRASILEIRO DE METROLOGIA DAS RADIAÇÕES IONIZANTES 2018
26 a 28 de novembro de 2018 - Rio de Janeiro

Protagonismo de um gerador de números aleatórios no método de Monte Carlo

Matheus Rebello do Nascimento[1,2], **José Guilherme Pereira Peixoto**[2]

[1] Universidade do Estado do Rio de Janeiro (UERJ); [2] Instituto de Radioproteção e Dosimetria (IRD)

E-mail: rebellomatheus99@gmail.com, guilherm@ird.gov.br

Resumo: O método de Monte Carlo é uma ferramenta extremamente preciosa no cenário científico do século XXI, pois ele fornece uma solução diferente a problemas de determinação de parâmetros teóricos ou experimentais, devido a muitos graus de liberdade e ao desenvolvimento estocástico de um problema. Nesse sentido, este trabalho busca a compreensão mais aprofundada da centelha vital desse método que é o gerador de números aleatórios, tendo como foco o gerador que emprega o método linear congruencial.

Palavras-chave: Gerador de Números Aleatórios, Gerador Linear Congruencial, Método de Monte Carlo

Abstract: The Monte Carlo method is an extremely precious tool in the scientific scenario of the XXI century, because it gives another solution to problems that look for determine physical parameters theoretically or experimentally, due to many freedom degrees and the stochastic development of a problem. In this way, this work looks for a deeper comprehension of the vital spark of this method that is the random number generator, having the focus on the generator that deploys the linear congruential method.

Keywords: Random Number Generator, Linear Congruential Generator, Monte Carlo Method

1. INTRODUÇÃO

A geração de números aleatórios é o alicerce em simulações de situações físicas realizadas sobre a perspectiva do método de Monte Carlo (MC). Todavia, para compreender a sua relevância e necessidade no cenário de simulações é importante estar ciente de como, por linhas gerais, a metodologia acerca desse método é desenvolvida.

As simulações MC são vislumbradas dentro de um cenário em que o estágio inicial de desenvolvimento de um sistema não é conhecido e/ou a sua evolução é estocástica (está vinculada a probabilidades) que resulta em uma extrema dificuldade, ou ainda, impossibilidade de quantificar parâmetros físicos analiticamente ou de mensura-los experimentalmente. Nesse sentido, pode se ilustrar sua aplicabilidade e extrema importância com os exemplos a seguir: no treino de redes neurais especialistas que fazem seleção de elétrons no experimento ATLAS, localizado no LHC (Large Hadron Collider), uma proposta da COPPE-UFRJ implementada em 2017; o espalhamento de raios X em câmaras de ionização sendo muito útil para estimativa de coeficientes de atenuação em calibração, no estudo da difusão de nêutrons em materiais, uma das primeiras utilizações do método proposta feita

por Henrico Fermi; estimativa de dose absorvida em física médica, proporcionando a otimização da dose empregada que é um dos alicerces da radioproteção; estimativas de custo de uma obra, fornecendo um valor mais provável de custo dentro de um intervalo de confiança; entre outras diversas aplicações nos mais variados campos do saber.

2. EVOLUÇÃO HISTÓRICA

Historicamente, a humanidade utiliza há séculos instrumentos que fornecem números de modo aleatório, ainda que em seu princípio empreguem um método extremamente simplório. Como por exemplo, dados que foram encontrados há 5 000 anos no Iraque e no Iran. Essa abordagem rústica de números aleatórios começa a tomar uma densidade maior ainda antes de existirem os primeiros computadores. Nesse momento se calculavam os números aleatórios manualmente e eles eram dispostos em tabelas e livros, sendo Tippet o primeiro a publicar uma tabela que continha 41 600 números aleatórios.

Com o advento da introdução de computadores, em torno de 1940 e 1950, a corporação RAND fez uma tabela com um milhão de números aleatórios que seriam gravados em 20 000 cartões pela IBM. O que há séculos atrás se resumia a um dado com faces numeradas de um a seis, veio continuamente quebrando barreiras devido ao avanço computacional e atualmente transcende sequencias numéricas com mais de 2^{32} elementos. Não somente se calcula uma quantidade maior de números, como também os métodos evoluíram e aperfeiçoaram a maneira de produzir uma sequencia de números aleatórios.

Nesse contexto, podemos segregar a produção da sequência em dois campos diferentes. Uma feita empregando sistemas físicos e a outra por um algoritmo de implementação computacional. Quando se imagina dispositivos físicos, está se tratando de sistemas microscópicos que selecionam ruído elétrico de modo aleatório retornando como resultado do processo, números que são dependentes da intensidade do sinal. Um exemplo disso, que é uma das primeiras publicações do tema, fora feita por Cobine e Curry em 1947, em que se utilizava um tubo de gás permeado por um campo magnético. Todavia, esse tipo de técnica é demorada e complexa, colocando a geração de números aleatórios via dispositivo físico mais voltada para criptografia, jogos de azar, entre outros. Por esse motivo, o método que emprega um algoritmo de implementação computacional é utilizado em simulações, abarcando o método de MC, pois ele é mais conveniente dada sua simplicidade e reprodutibilidade, devido sua estrutura particular que será abordada ao longo do texto.

3. NÚMEROS ALEATÓRIOS

Visando tornar a interpretação dos números aleatórios mais clara, tomemos as colisões entre pacotes de prótons nos experimentos do LHC. Tipicamente, as colisões atuais ocorrem a uma taxa da ordem de grandeza de um bilhão de colisões por segundo, ou seja, há bilhões de prótons circulando no túnel do LHC que eventualmente colidem em 4 experimentos (ALICE, LHCb, CMS, ATLAS que são detectores delineados para estudo da física de partículas em seus eventos de interesse, como quark Top, quark Bottom, bóson de Higgs e busca por indicativos de famigerada nova física como a hipótese da supersimetria e do bóson gráviton).

Devido a essa quantidade impressionante de interações ocorrendo numa região de modo aleatório e das possíveis produções de um portfólio vasto de partículas como Higgs, W, Z, J/ψ entre outras, revela aos expectadores teóricos desses fenômenos um universo inteiro de densidade de probabilidades. Com isso, dado essa vastidão de possíveis caminhos de evolução do fenômeno físico, se torna factível compreender a necessidade de um artifício computacional agindo na simulação de uma situação física complexa em que o estado inicial de interação não é conhecido e o seu desenvolvimento é estocástico.

De modo geral, podemos olhar para uma simulação de Monte Carlo tendo a seguinte estrutura, figura 1:

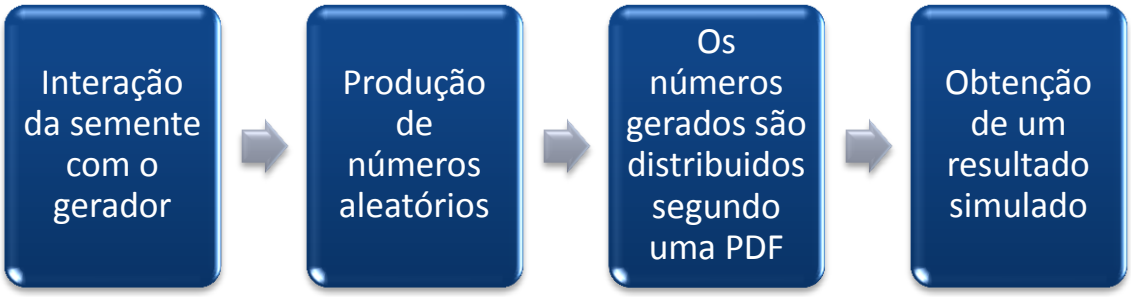

Figura 1: Nesta figura pode-se observar a estrutura elementar simplificada da simulação de MC em que se inicia pela definição de uma semente que começará a produzir os números aleatórios. Estes serão distribuídos, segundo métodos de amostragem, em uma função de densidade de probabilidade (PDF) e então será obtido um resultado oriundo da simulação do fenômeno físico.

3.1. Gerador de números aleatórios

Como está se tratando de variáveis cujo desenvolvimento dentro de uma situação física é regido por processos estocásticos, nada mais justo que utilizar uma ferramenta de simulação que tome como base algo com um certo grau de aleatoriedade. Esse algo se chama gerador de números pseudoaleatórios.

Imagina-se que o leitor esteja se questionando acerca do prefixo pseudo. Esta nomenclatura ocorre, em sua essência, pelo fato de ser inimaginável delinear um gerador aleatório, pois isto é uma abstração matemática. Para entendermos o período anterior precisamos imaginar a seguinte situação:

Existe um sorteio de loteria em que uma pessoa retira 6 bolas de uma caixa, sem as colocar de volta no recipiente. Evidentemente esse evento é totalmente aleatório, pois a probabilidade de uma dada bola ser sorteada é igual as demais. Além do fato de a pessoa selecionar uma bola sem ser induzida visualmente ou ser instruída a fazer essa ação de forma a não haver privilégios.

Nesse momento imagine como um computador conseguiria repetir o procedimento mencionado. Isto é impossível, dado que um computador não consegue colocar a mão numa urna e retirar um número qualquer fruto de uma obra do acaso. O computador é apenas um conjunto complexo e belíssimo de capacitores e outros elementos eletrônicos que operam seguindo instruções claras e bem definidas. Então é necessário fornecer instruções as quais o computador será capaz de compreender. No nosso estágio atual de desenvolvimento tecnológico e conhecimento, essas instruções são dadas segundo linguagens de programação como Python, C++, Fortran, entre outras, que irão relatar instruções a serem seguidas pela máquina. Uma vez que é necessário instruir a máquina a seguir um meio de gerar um número, este procedimento não é aleatório, pois está sendo seguido uma lógica específica e consequentemente é inserido uma tendência de escolha.

Como então pode ser gerado esses números pseudoaleatórios e por que podemos trata-los como aleatórios? Respondendo a essa questão, inspirado em Hammerley e Handscomb (referência [1]), a preocupação acerca da sequência de números aleatórios não deve residir sobre o mecanismo de geração, mas sim se eles estão distribuídos corretamente, ou seja, estão fornecendo resultados em testes estatísticos que são compatíveis com o esperado por uma sequência de números aleatórios. Ao pensar em geração de números aleatórios o esperado é obter uma distribuição retangular em que cada

número gerado possua mesma probabilidade de ser sorteado, além de formar uma sequência de números não correlacionados.

Existem diversos geradores de números aleatórios seja ele linear ou não, a não linearidade de um gerador reside ao fato de ele possuir em sua estrutura um termo de grau maior que um. Embora o não linear consiga atingir o período máximo, eles são mais lentos que os lineares. Nesse sentido, será abordado o gerador mais empregado que é o linear congruencial que fora proposto por Lehmer em 1949.

3.2. Gerador linear congruencial aditivo e multiplicativo

Primariamente, o processo de geração desses números se baseia em uma equação de recorrência (expressa um termo da sequência em função de termos anteriores) que utiliza divisibilidade de números inteiros como artifício, exemplificado em (1):

$$u_{n+1} \equiv (au_n + c) \bmod (m) \tag{1}$$

Em que u é um número aleatório, a e c são constantes e m é o divisor.

Matematicamente falando, para gerar essa sequência de n termos de u ate u_n, chamada de sequencia pseudoaleatória, devemos atribuir um valor de partida para u, que na pratica computacional do problema seria a semente da recorrência. Logo em seguida deve se fazer $(au_n+c)/m$ e o resto da divisão desse termo será o u_{n+1}. Repetindo essa conta até atingirmos o valor de u_n, como intuitivamente se faz devido a definição matemática da recorrência, obtemos a anteriormente citada, sequencia pseudoaleatória.

Evidenciando o exposto a cima, tomemos a equação (1) atribuindo a semente com o valor de u_0=48, a=113, c=7 e m=93.

Ao realizar a divisão de $(113.48+7)/93$ obtém-se resto 37 que será o termo u_1 da sequência. Então será repetido o procedimento para o 37:

$$(113.37+7)/93$$

Obtendo resto três que será o termo u_2 da sequência. Isto deve ser repetido quantas vezes se deseja. Neste caso em particular o termo u_{30} é 48, o valor da semente, demonstrando que esse gerador de números pseudoaleatórios possui um período muito curto.

A semente, citada a cima, é um valor que exerce influência sobre a qualidade do gerador de números aleatórios, entretanto não há suporte teórico que aponte uma boa semente para determinado gerador, essa busca é empírica.

3.2.1 Exemplos teóricos de geradores pseudoaleatórios

O método de geração linear congruencial se subdivide em conguencial aditivo e congruencial multiplicativo. Todavia, o método aditivo apresenta uma superioridade frente ao multiplicativo uma vez que é possível atingir o seu período completo -entende-se por período completo a sequência pseudoaleatória que possuiu m termos até que um número desta se repita- satisfazendo três condições básicas. Por sua vez, o multiplicativo possui um período sempre menor que m e para alcançar o maior período possível, diferente de m, exige uma gama de condições mais complexas.

Congruencial multiplicativo, equação (2):

$$u_{n+1} \equiv (au_n) \bmod (m) \tag{2}$$

Condição para o período máximo podem ser encontrados na referência [1].

Congrunecial aditivo, equação (3):

$$u_{n+1} \equiv (au_n + c) \bmod (m) \qquad (3)$$

Condição para o período máximo:

- "a" e "c" não podem possuir divisores em comum
- $a \equiv 1 \bmod (p)$ para todos fatores primos de m, lembrando que todos números podem ser decompostos em parcelas primas.
- $a \equiv 1 \bmod (4)$ se m for múltiplo de 4

3.2.2 Exemplos práticos de geradores pseudoaleatórios

Outra forma de proporcionar uma boa sequência de números aleatórios, conforme sugere a literatura, é misturando geradores de propriedades aritméticas diferentes, ou seja, combinando os geradores, usualmente por soma ou subtração, como podemos observar alguns exemplos abaixo.

$x_n = \left(69069 x_{n-1} + c\right) \bmod 2^{32}$ apresenta período aproximado de 2^{32}

$x_n = \left(x_{n-8} - x_{n-10} - c\right) \bmod (2^{31} - 5)$ apresenta período aproximado de 2^{307}

$x_n = \left(x_{n-1} * x_{n-2}\right) \bmod 2^{32}$ apresenta período aproximado de 2^{31}

Outros geradores, seguindo os exemplos a cima, e mais detalhes podem ser encontrados na referência [2].

Uma implementação simples que também facilita a compreenção do funcionamento do gerador de números pseudoaleatórios é a implementação feita abaixo, em Fortran, utilizando a função mod (x, y) que retorna o resto da divisão de x por y, justamente o que é feito em um gerador de números aleatórios de modo geral. Possuindo um computador que dispõe de compilador gfortran pode-se verificar como a longevidade do período depende das escolhas de a, c, m e da semente.

```
program random
  implicit none
    integer:: semente !Número que inicia a sequência
    integer:: resto!Resto da divisão retornada pela função mod()
    integer:: k !contador
    logical:: presente !indica se a semente foi repetida
    integer, dimension(93):: numero !lista com a sequência pseudoaleatória

     semente=48
     resto=mod(113*semente+7,93)
     numero(1)=resto
     print*,resto
     presente=.false.
     do while (presente.eqv..false.)
          do k=1,93,1
```

```
            numero(k+1)=mod(7+113*numero(k),93)
                    if (numero(k+1)==seed) then
                            presente=.true.
                            exit !Força a parada do "do k=1,93,1"
                    else
                            print*, numero(k+1)
                    end if
            end do
        end do
  print*, 'o período é de', k
end program
```

O código a cima é feito para não somente calcular o resto das divisões e coloca-las em uma lista, como também realizar a contagem do período. No programa pode-se observar que no exemplo a cima o período atingido foi de 29 números. Todavia, ao substituir o gerador aditivo mod(113*seed+7,93) pelo gerador multiplicativo mod(113*seed,93) o período reduz para 14. Ou seja, este caso evidencia mais uma vez a superioridade teórica mencionada anteriormente que aponta para o gerador de linear congruencial aditivo como superior frente ao multiplicativo.

Sequência gerada pelo método aditivo: 37, 3, 67, 45, 70, 12, 61, 18, 88, 0, 7, 54, 64, 78, 79, 6, 34, 36, 76, 39, 43, 30, 49, 57, 31, 69, 85, 33, 16

Sequência gerada pelo método multiplicativo: 30, 42, 3, 60, 84, 6, 27, 75, 12, 54, 57, 24, 15, 21

Estas sequências geradas, ao plotar um gráfico com os 14 primeiros números dos dois geradores formam o seguinte gráfico:

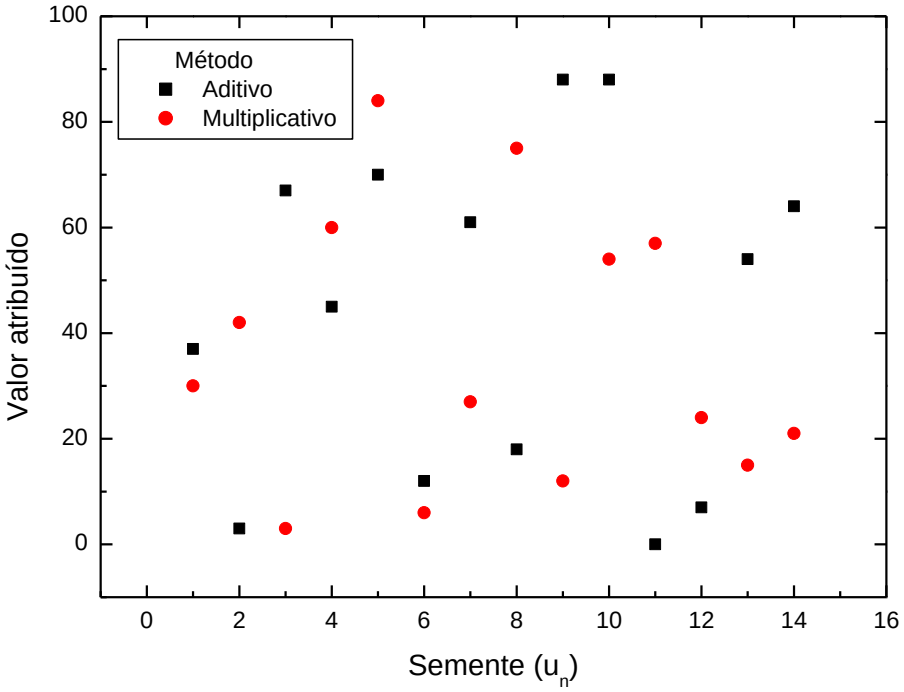

Figura 2: Na imagem constam os 14 primeiros números produzidos pelos dois geradores de números aleatórios. Foram colocados apenas 14, pois é o período do gerador congruencial multiplicativo, desse modo a comparação entre os geradores pode ser estabelecida sem a defasagem entre os períodos.

Outra maneira de gerar números aleatórios é utilizar o software do EXCEL, amplamente difundido e presente nos computadores de qualquer instituição ou pessoa física. Todavia, a sua estrutura matemática não é conhecida e deve ser evitada para trabalhos científicos. Uma particularidade desse gerador é que ele se apoia em uma distribuição que pode ser escolhida ao gerar o número. Por exemplo, podemos gerar números aleatórios segundo algumas distribuições (retangular, Poisson, normal, entre outras), que é justamente a próxima etapa do método de MC após gerar os números aleatórios, exemplificado na figura 3.

A fim de verificar essa funcionalidade, deve-se clicar na aba "Dados" (certifique-se de que este complemento de análise de dados esteja habilitado) e na seção "análise" clicar em "análise de dados". Então aparecerá uma gama de opções de análise. Portanto, deve-se procurar por "geração de número aleatório". Logo, aparecerá:

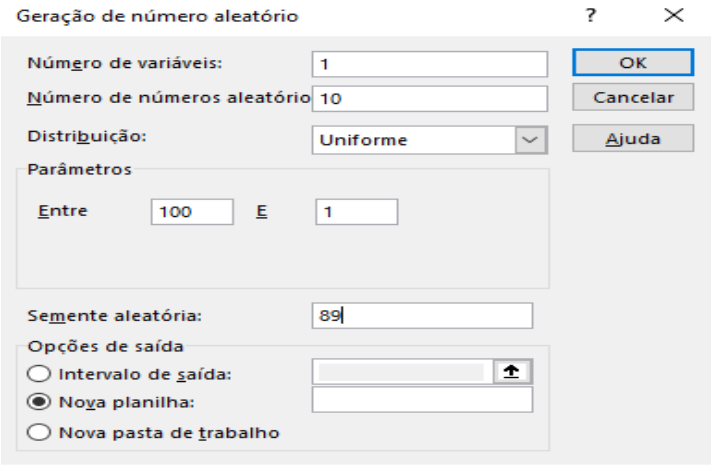

Figura 3: Nesta imagem aparecem as opções de escolha para gerar o número. Pode variar a distribuição, os limites em que se deseja gerar os números que no exemplo é no intervalo [1,100], escolher o valor da semente e a quantidade dos valores de que se deseja produzir ao modificar o parâmetro número de variáveis e número de números aleatório.

O exemplo à cima gerou a seguinte saída: 1,994018; 43,46483; 17,66869; 43,7579; 29,24342; 3,501663; 71,18253; 4,426191; 17,38166; 24,33073

4. CONCLUSÃO

Ao longo deste artigo, foi possível observar que a geração de números aleatórios é algo que permeia a humanidade há séculos e a sua evolução residiu ao modo que se produz informação aleatória. Nesse contexto, jamais se cogitaria fazer simulações via método de Monte Carlo se não existisse uma maneira mais conveniente de gerar números aleatórios, com rapidez, qualidade e em grande quantidade que hoje existe devido ao avanço tecnológico. Utilizar moedas com um barco numa face e imperador na outra, como era no Império Romano, parece um tanto inapropriado no universo das simulações.

No que tange os métodos de produzir números aleatórios, foi exposto de modo teórico o fato de que o gerador linear congruencial aditivo possui uma superioridade em comparação com o multiplicativo. Esta afirmação foi colocada em prova no caso particular no código em Fortran. Ainda que se tenha

utilizado um gerador com um período muito curto, pode se observar que os números produzidos por um gerador não foram iguais na sequencia produzida pelo outro, observando a figura 2. Nesse sentido, como perspectiva futura, deve se buscar aumentar o período de ambos geradores e com isso entender com maior clareza se os números de uma sequencia, para a mesma semente, irão coincidir com o da outra sequência e a que taxa isso ocorre, além do fato de poder observar se haverá a formação de algum padrão gráfico.

Agradecimentos

Agradeço ao CNPq pelo apoio financeiro através da bolsa PIBIC/PROBIC. Também cabe agradecer a minha família que proporciona suporte psicológico e financeiro sempre auxiliando minha caminhada acadêmica. Em especial ao Agostinho Villanova pelo transporte para o IRD e para a UERJ visando minha segurança e rapidez de locomoção. No cenário acadêmico cabe ressaltar o protagonismo que a 2° Jornada de Física Médica da UFRJ teve, expandindo meus horizontes para um novo campo do saber em coadunação com o curso de Fundamentos de Radioproteção e Metrologia do IRD. Neste curso, o conhecimento adquirido foi decisivo para a vontade de trabalhar nessa instituição. Em adição a isto, agradeço a Rafaella Carvalho pela indicação do professor José Guilherme, proporcionando firmar essa parceria e desenvolver um trabalho de iniciação científica.

REFERÊNCIAS

[1]-HAMMERSLEY, J. M. **Monte Carlo methods.** c1964, London: Chapman & Hall. 178 p. (Monographs on applied probability and statistics).

[2]-MARSAGLIA G, ZAMAN A., **Some portable very-long period random number generators**. c1994 Comput. Phys**. 8 117-21**;

[3]-PIERRE L'ECUYER, **History of uniform random number generation**. WSC 2017-Winter Simulation Conference, Dec 2017, Las Vegas, United States. 2017.

[4]-YORIYA H., **Método de Monte Carlo: princípios e aplicações em Física Médica**. Ago 2009. Revista Brasileira de Física Médica, v. 01, p. 141-149.

[5]-*ZAID H, SGOUROS G.*, **Therapeutics Applications of Monte Carlo Calculations in Nuclear Medicine.** c2003, London: Institute of Physics Publishing. 364 p.