Monte Carlo: Como Estimar π e Calcular Integrais

Introdução a Física Estatística e Computacional - IFEC

1 O que é Monte Carlo?

Sistemas de relativamente poucas variáveis tais como o pêndulo duplo e o modelo logístico discutidos anteriormente frustram nossa capacidade de previsão mesmo quando resolvemos esses sistemas numericamente. Assim, para estudar sistemas cada vez mais complexos, é necessário recorrer ao uso de métodos mais poderosos. Em sistemas que contenham um número muito grande de entes ou onde o comportamento não seja completamente determinístico, o uso de métodos estatísticos se torna imprescindível. Talvez, o mais conhecido deles é o **método de Monte Carlo** que faz uso da lei dos grandes números.

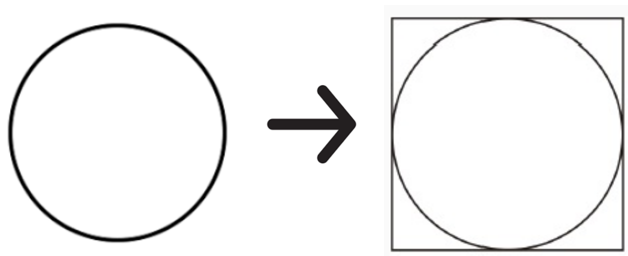
A origem do método de Monte Carlo remonta, em sua versão mais simples, à determinação do número π por Buffon no século XVIII. Avanços mais significativos e a formalização do método surgiram na década de 40 em conexão com o projeto Manhattan (projeto de desenvolvimento da bomba atômica pelos EUA). Quem cunhou o nome Monte Carlo para o método que utilizava números aleatórios para amostrar distribuições de probabilidades foi Nicholas Metropolis para fornecer um codinome ao projeto secreto que vinha sendo desenvolvido por Ulam e von Neumann. O nome é uma referência ao famoso cassino de Monte Carlo no principado de Mônaco, onde a aleatoriedade desempenha um papel relevante nos ganhos dos apostadores. Hoje em dia o termo Monte Carlo é empregado para designar uma imensa gama de métodos estatísticos baseados no uso de números aleatórios.

Algoritmos de Monte Carlo costumam resolver três tipos de problema: integração, otimização e amostragens de distribuições de probabilidades. Todavia, eles partem de um mesmo princípio: **usar muitas configurações aleatórias** chamadas *samples* com o auxílio de números aleatórios. Ao longo do curso exploraremos diferentes tipos de métodos de Monte Carlo. Discutiremos primeiro formas simples de se calcular integrais usando este método.

2 Estimando π (amostragem direta)

Não é novidade para ninguém que $\pi = 3,14159265...$ que por sua vez é a razão entre o perímetro e o diâmetro de um círculo. Resultados cada vez mais precisos podem ser obtidos por medidas cada vez mais precisas. No entanto, não podemos nos esquecer que qualquer processo de medida está sujeito à incertezas, que se propagam para os valores calculados. Mas, se não soubéssemos que $\pi = C/D$, haveria outras formas de determiná-lo?

Olhar para a área do círculo pode ser mais adequado do que olhar para o seu comprimento se quisermos estimar π. Sabemos que $A = \pi R^2$ onde R é o raio do círculo. Se consideramos um círculo unitário, sua área é exatamente π. A primeira coisa que podemos fazer é inscrevê-lo em um quadrado.



Como o círculo tem raio unitário, o lado do quadrado será 2 e a área do quadrado é 4 (estamos considerando unidades arbitrárias). O próximo passo é "lançar" dentro do quadrado aleatoriamente um número muito grande de “dardos”, i.e., escolher um numero grande N_{total} de pares (x, y) dentro do quadrado e contabilizar quais deles caíram dentro do círculo. Pela lei dos grandes números podemos afirmar o seguinte sobre a proporção das áreas:

$$\frac{A_{circulo}}{A_{quadrado}} = \frac{\pi}{4} \approx \frac{N_{dentro}}{N_{total}} \longrightarrow \pi \approx 4 \times \frac{N_{dentro}}{N_{total}}.$$

É claro que se o número N_{total} for baixo a nossa estimativa de π é grosseira. Considere, por exemplo, $N_{total} = 2$, se um ponto for sorteado dentro e outro fora do círculo teríamos que $\pi \approx 2$, e se ambos forem sorteados dentro teríamos que $\pi \approx 4$. Mas observe o que acontece para N_{total} a partir de 100 na Figura 1 Obtivemos $\pi \approx 3.36$ para $N_{total} = 100$, $\pi \approx 3.152$ para $N_{total} = 1000$ e $\pi = 3.148$ para $N_{total} = 10000$.

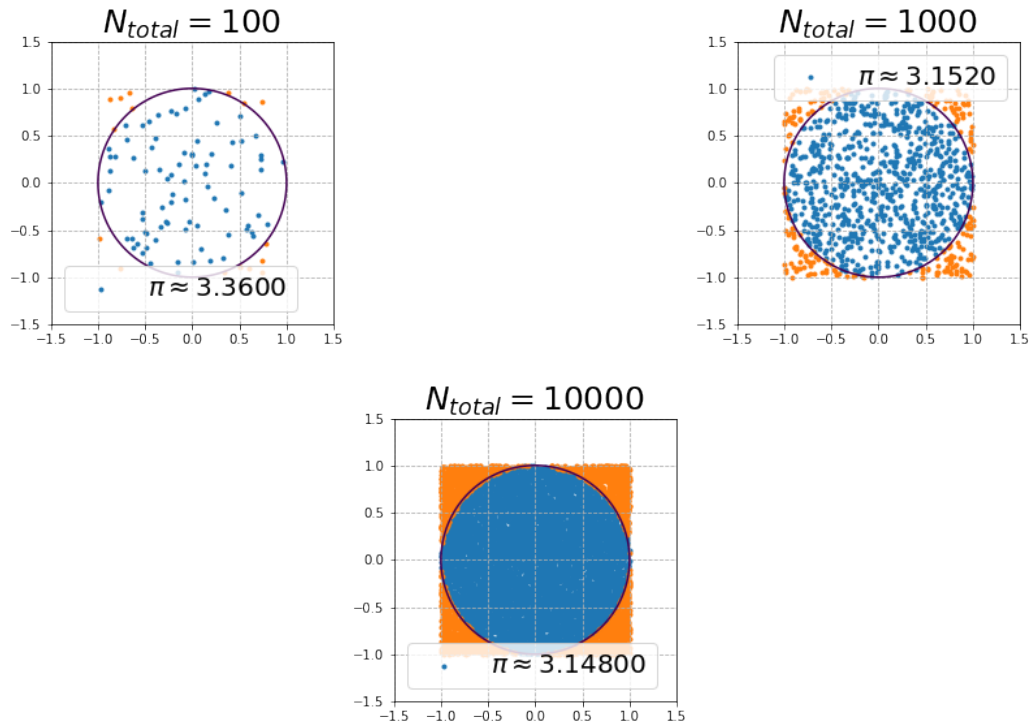


Figure 1: Estimativas de π para $N = 10^2, 10^3, 10^4$

Um fato que não podemos esquecer é que este é um método probabilístico. Se realizarmos novamente o experimento com $N_{total} = 100$, dificilmente obteríamos o mesmo valor, $\pi \approx 3.36$. Assim, temos que ter em mente que as estimativas obtidas estão sujeitas a erros estatísticos que podem ser estimados ao repetirmos um mesmo experimento muitas vezes. Além disso, quanto maior for N_{total} mais precisa deve ser a estimativa de π e estes fatos podem ser testados da seguinte forma:

1. Obtenha $N_{amos} = 10000$ estimativas diferentes de π para um valor fixo de $N_{total} = 100$ dardos jogados no círculo;
2. Faça um histograma desses valores;
3. Faça a mesma coisa para $N_{total} = 1000$ e $N_{total} = 10000$.

Os resultados são mostrados na Figura 2:

Observe que quanto maior o número de dardos jogados, mais centrada é a distribuição de estimativas em torno do valor real de π. Além disso, a largura da distribuição também diminui, indicando uma redução no erro estatístico associado ao valor calculado. Quanto maior for o número de experimentos realizados, menor

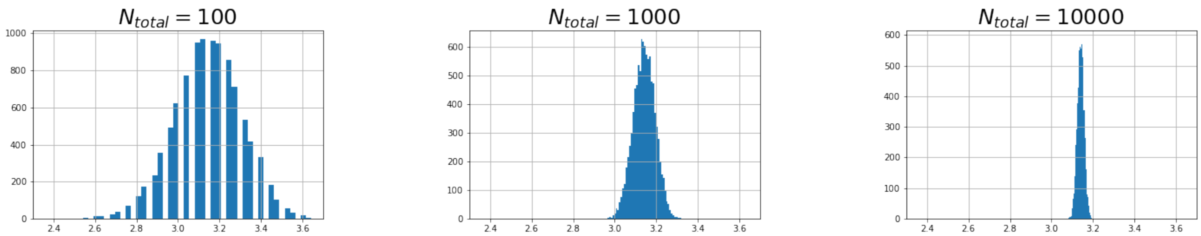


Figure 2: Histogramas de estimativas de π para $N = 10^2, 10^3, 10^4$

será a incerteza estatística no valor obtido. O cálculo de π mostrado aqui é um dos exemplos mais didáticos para se introduzir o método de **Monte Carlo** em sua versão mais simples, a amostragem direta (Direct Sampling), e pode ser expandido para calcular integrais em geral em quantas dimensões for necessário.

3 Calculando integrais por amostragem direta - Método 1

O mesmo procedimento que foi feito para calcular π também pode ser feito para calcular numericamente integrais definidas, $\int_{a_1}^{a_2} f(x)$. A diferença é que ao invés de escolher pontos (x, y) aleatoriamente em um quadrado de área 4 e contar os $(x, y) \quad \text{t.q} \quad x^2 + y^2 < 1$, você os atirará em um retângulo de lados $a_2 - a_1$ e $\max(f(x)), \quad a_1 < x < a_2$ e contará os pontos $(x, y) \quad \text{t.q} \quad y < f(x)$. Ao usar esse método, um cuidado especial deve ser tomado com funções que apresentam valores negativos no domínio considerado. Neste caso, pode-se escolher pontos também no eixo y negativo, tomando o cuidado de subtrair os pontos que ficaram entre a curva da função e o eixo x. Outro ponto muito relevante é a eficiência do método, que depende fortemente da forma da função e da escolha feita para o retângulo onde os pontos serão sorteados. De fato, se a função apresenta picos ou vales agudos, pode ocorrer de uma fração muito grande dos pontos caírem sempre dentro ou sempre fora da região de interesse, tornando o método pouco eficiente e pouco preciso.

4 Calculando integrais por amostragem direta - Método 2

Tomando como inspiração o método apresentado anteriormente e o método do trapézio para cálculo numérico de integrais definidas, podemos definir um método de Monte Carlo alternativo para o cálculo de integrais. A ideia é partir da própria definição do valor médio de uma função:

$$\langle f(x) \rangle = \frac{1}{a_2 - a_1} \int_{a_1}^{a_2} f(x) dx. \tag{1}$$

Desta forma, bastaria encontrar o valor médio da função, $\langle f(x) \rangle$, fazendo uma amostragem aleatória dos seus valores. Assim, considerando um conjunto de N valores de x escolhidos de forma aleatória e uniformemente distribuídos no intervalo (a_1, a_2), $\{x_i\}$, podemos estimar o valor médio da função e, consequentemente o valor da integral fazendo:

$$\int_{a_1}^{a_2} f(x) dx = (a_2 - a_1) \langle f(x) \rangle = \frac{a_2 - a_1}{N} \sum_{i=1}^{N} f(x_i). \tag{2}$$

Desta forma, evitamos alguns dos problemas levantados anteriormente. Ainda poderíamos melhorar o método fazendo uma escolha dos valores de x que leve em consideração, de alguma forma, os valores da função, amostrando de forma mais efetiva as regiões que contribuem mais para a integral. Esse método, no entanto, está fora do nosso escopo inicial e será discutido mais a frente.

5 Cálculo de integrais múltiplas

Pelo exposto na seção anterior, pode-se notar que não há elementos suficientes para acreditarmos que o método 2, mesmo sendo mais eficiente que o método 1, seja muito superior à regra do trapézio ou 3/8 de

Simpson, por exemplo. Então, por que deveríamos nos preocupar em aprender métodos de Monte Carlo na resolução de Integrais? A resposta é relativamente simples: Enquanto para integrais unidimensionais, i.e., aquelas que envolvem apenas uma variável de integração, a vantagem do método de Monte Carlo não seja evidente, para integrais multidimensionais a situação é diferente. A amostragem aleatória em integrais multidimensionais é capaz de fornecer resultados precisos a um custo computacional muito menor que o necessário ao usar métodos como a regra do trapézio ou 3/8 de Simpson. Neste caso, são escolhidos N pontos aleatórios em d dimensões e o valor médio da função pode então ser calculado e o valor da integral estimado. Matematicamente, considerando uma integral em 3 dimensões,

$$\int_{a_1}^{a_2}\int_{b_1}^{b_2}\int_{c_1}^{c_2} f(x_1,x_2,x_3)dx_1dx_2dx_3 = (a_2-a_1)(b_2-b_1)(c_2-c_1)\langle f(x_1,x_2,x_3)\rangle \tag{3}$$

$$\langle f(x_1,x_2,x_3)\rangle = \frac{1}{N}\sum_{i=1}^{N} f(x_1^i,x_2^i,x_3^i). \tag{4}$$

A generalização para problemas envolvendo d dimensões é imediata.

5.1 Tarefa:

Resolva, usando os métodos de Monte Carlo descritos aqui, as seguintes integrais:

1. $$\int_0^1 \left(1-x^2\right)\mathrm{d}x, \quad \text{Valor analítico: } \frac{2}{3}$$

2. $$\int_0^1 e^x\mathrm{d}x, \quad \text{Valor analítico: } e-1$$

3. $$\int_0^\pi \sin^2(x)\mathrm{d}x, \quad \text{Valor analítico: } \frac{\pi}{2}$$

4. $$\int_0^1\cdots\int_0^1 \frac{1}{(\vec{r}_1+\vec{r}_2)\cdot\vec{r}_3}d\vec{r}_1d\vec{r}_2d\vec{r}_3 =$$
$$\int_0^1\cdots\int_0^1 \frac{1}{((x_1+x_2)x_3+(y_1+y_2)y_3+(z_1+z_2)z_3)}dx_1dy_1dz_1dx_2dy_2dz_2dx_3dy_3dz_3$$

Calcule as integrais 1, 2 e 3 pelos métodos 1 e 2. Além do valor obtido para a integral, apresente os histogramas para $N_{total} = 100$, 1000 e 10000 pontos lançados considerando $N_{amos} = 1000$ amostras diferentes, ou seja, obtenha 1000 estimativas diferentes para o valor da integral considerando, por exemplo, que $N_{total} = 100$ pontos são sorteados em cada uma destas amostras. Comente os resultados, em especial, comente sobre erros estatísticos relacionados aos valores obtidos. **Bônus:** Você é capaz de estimar estes erros?

Calcule a integral 4 usando o método 2. Você deve fazer alguns testes, que não precisam ser apresentados, para determinar um número razoável de pontos a serem sorteados (N_{total}) e de amostras a serem usadas (N_{amos}).

6 Referências:

- D.P. Landau e K. Binder, A guide to Monte Carlo Simulations in Statistical Physics, Cambridge University Press (2014)
- W. Krauth, Statistical Mechanics Algorithms and Computations, Oxford (2006)

- Approximating Pi (Monte Carlo integration) — animation https://www.youtube.com/watch?v=ELetCV_wX_c
- Monte Carlo Integration In Python https://www.youtube.com/watch?v=WAf0rqwAvgg&t=324s
- Monte Carlo Method https://en.wikipedia.org/wiki/Monte_Carlo_method