Universidade de São Paulo
Instituto de Física de São Carlos

Projeto 2: Introdução à física computacional

Julia Martins Simão - 13694997

São Carlos
2025

Sumário

1 Introdução

A utilização de números pseudo-aleatórios é fundamental para a realização de simulações computacionais de fenômenos naturais, como os processos termodinâmicos. A análise microscópica direta de sistemas compostos por um número extremamente grande de partículas — $6.02 \cdot 10^{23}$, no mínimo — é inviável na prática. Ainda assim, existe uma área da física dedicada justamente a esse tipo de estudo: a física estatística. Suas teorias conectam as propriedades macroscópicas e microscópicas de sistemas com muitas partículas, apoiando-se em teorias de probabilidade e estatística. Nesse contexto, o uso de números pseudo-aleatórios torna-se indispensável, pois permite explorar o comportamento coletivo de tais sistemas quando a verificação experimental direta não é possível. Apesar de não ser possível obter resultados exatos a partir de estudos probabilísticos, quanto maior a amostragem utilizada para o estudo, mais precisas serão as previsões. Por isso, nesse relatório, os valores dos parâmetros utilizados são altos.

Embora técnicas estocásticas sejam amplamente aplicadas para compreender fenômenos determinísticos de difícil caracterização, é importante destacar que os únicos processos verdadeiramente probabilísticos na natureza têm origem na mecânica quântica; e ela está em todo lugar.

Tarefa 1

Momentos de distribuição

A fim de testarmos o gerador de números aleatórios, calculemos alguns momentos da distribuição "aleatória"gerada, isto é:

$$\langle x^n \rangle \text{ , para } n = 1, 2, 3, 4$$

Faça a média acima gerando um número grande N de números aleatórios (escolha apropriadamente N). Que resultado você esperaria? Compare com os resultados esperados e explique os obtidos.

Do cálculo, tem-se que a média de uma função $f(x)$ é dada por

$$\langle f(x) \rangle = \frac{1}{b-a} \int_a^b f(x)\, dx$$

No caso da tarefa 1, $f(x) = x^n$, portanto

$$\langle f(x) \rangle = \langle x^n \rangle = \frac{1}{b-a} \int_a^b x^n\, dx = \frac{1}{b-a} \cdot \frac{x^{n+1}}{n+1}$$

A função *rand()* gera números entre 0 e 1, o que implica que o intervalo de integração é $[a, b] = [0, 1]$. Finalmente,

$$\langle x^n \rangle = \frac{1}{n+1} \tag{1}$$

Ou seja, para cada n, a média da função (ou momento de distribuição) será dada pela relação 1. Como $n = 1, 2, 3, 4$, tem-se

- $n = 1$: $\langle x \rangle = \frac{1}{2}$;
- $n = 2$: $\langle x^2 \rangle = \frac{1}{3}$;
- $n = 3$: $\langle x^3 \rangle = \frac{1}{4}$;
- $n = 4$: $\langle x^4 \rangle = \frac{1}{5}$.

O código foi testado para dois valores de N, de modo que fosse possível notar que quanto maior for a quantidade de números gerados, mais precisos serão os valores calculados pelo código. Os valores obtidos com o programa são compatíveis com os valores analíticos mostrados acima.

Exemplos:

```
digite N:
1000
a media de x eh:  0.497961760
a media de x**2 eh:  0.326714516
a media de x**3 eh:  0.240649104
a media de x**4 eh:  0.189388528


digite N:
10000
a media de x eh:  0.501827717
a media de x**2 eh:  0.335473716
a media de x**3 eh:  0.252227634
a media de x**4 eh:  0.202283651
```

O funcionamento do programa é simples: o usuário fornece um valor de números aleatórios a serem gerados N e são inicializadas 4 variáveis de soma, cada uma referente à um dos valores de n. Por meio de um loop, gera-se os N números aleatórios x com a função *rand()* e cada uma das somas é acrescida de x^n para n indo de 1 à 4. Por fim, fora do loop, são tomadas as médias.

Código:

```
        write(*,*) "digite N:"
        read(*,*) n

        soma1=0.0
        soma2=0.0
        soma3=0.0
        soma4=0.0

        do i=1,n
                x=rand()
                soma1=soma1+x
                soma2=soma2+x**2
                soma3=soma3+x**3
                soma4=soma4+x**4
        go to 10
10      end do

        xmedia1=soma1/n
        xmedia2=soma2/n
        xmedia3=soma3/n
        xmedia4=soma4/n

        write(*,*) "a media de x eh:", xmedia1
        write(*,*) "a media de x**2 eh:", xmedia2
        write(*,*) "a media de x**3 eh:", xmedia3
        write(*,*) "a media de x**4 eh:", xmedia4
        end
```

Tarefa 2

Andarilhos aleatórios 1D

Vamos considerar agora o problema de andarilhos aleatórios em uma dimensão. Aqui, em cada unidade de tempo, cada caminhante, independentemente de onde esteja, dá um passo à direita (ou à esquerda) com probabilidade p ($q = 1 - p$). O caso $p = q = \frac{1}{2}$ corresponde a um caminhante tão desnorteado que ele não se lembra de onde veio e nem qual o rumo certo a tomar. O caso em que $p \neq q$ corresponde ao viajante aleatório em uma ladeira. A questão nesta tarefa é calcular $\langle x \rangle$ e $\langle x^2 \rangle$ após um certo número N de passos.

1.0.1 Cadeias de Markov

Uma Cadeia de Markov é um processo estocástico cujo princípio fundamental é que o que acontece a seguir depende unicamente do estado atual do sistema em estudo. Por depender de variáveis aleatórias, é impossível prever com exatidão o estado futuro de uma Cadeia de Markov e, por isso, é necessário que se faça um estudo estatístico do porvir do sistema. No caso do problema dos andarilhos aleatórios, a direção na qual um andarilho irá caminhar independe da direção em que outro andarilho seguiu e, assim, pode ser entendida como uma Cadeia de Markov com tempo discreto.

A

Considere $p = q = \frac{1}{2}$ e um número grande M de andarilhos, todos partindo da origem ($x = 0$) no tempo inicial ($t = 0$). Após $N = 1000$ passos, faça um histograma do número de andarilhos $n(x)$ em função de x. Que tipo de curva você obteve? Calcule $\langle x \rangle$ e $\langle x^2 \rangle$.

O programa foi escrito tomando $M = 10000$ e $N = 1000$. Além disso, ele analisa somente os passos dados à direita, pois é possível obter a probabilidade de passos à esquerda fazendo $q = 1 - p = 1 - \frac{1}{2} = \frac{1}{2}$: ou seja, a probabilidade de passos à direita é a mesma que de passos à esquerda. Sendo p a probabilidade de passos serem dados à direita, tem-se que

$$\langle p \rangle = N \cdot p \tag{2}$$

Como $p = q = \frac{1}{2}$, espera-se que $\langle p \rangle \to \frac{N}{2} = 500$

e também

$$\langle x \rangle = \langle p \rangle - \langle q \rangle = 2 \cdot \langle p \rangle - N \tag{3}$$

Além disso, é possível obter a forma analítica de $\langle x^2 \rangle$ a partir da relação

$$\langle x^2 \rangle = 4 \cdot \langle p^2 \rangle - 4 \cdot N \cdot \langle p \rangle + N^2 \tag{4}$$

Exemplo:

```
media p=   500.045807
media x=   9.16137695E-02
media de x**2=   1000.31250
```

Código:

```
c vetor que armazena os passos a direita
        dimension npd(0:10000)

        n=1000
        m=10000
        pm=0
        pm2=0

c arq que armazena os passos dos andarilhos
        open(unit=1, file='media_x.dat')

        do i=1,m
c nd (passo a direita) eh inicializado para cada andarilho
                nd=0
                do j=1,n
                        x=rand()
c considero que qualquer numero aleatorio gerado menor do que 0.5
c equivale a um passo a direita
                        if(x .lt. 0.5) then
                                nd=nd+1
                        end if
                end do
c se o passo for a direita, adiciono nd ao pm (media do numero de passos) e
c nd**2 ao pm2 (media quadrada do numero de passos)
                pm=pm+nd
                pm2=pm2+nd**2
c guardo o passo a direita atual no vetor de passos a direita
                npd(nd)=npd(nd)+1
        end do

c guardo o resultado dos passos de cada andarilho no arq
        do k=0,n
                write(1,*) k, npd(k)
        end do

c media de p
        pmedia=pm/m
        xmedia=(2*pmedia)-n
c media de p**2
        pmq=pm2/m
c formula da media de x**2
        xmq=4*pmq-4*n*pmedia+n**2

        write(*,*) "media p=", pmedia
        write(*,*) "media x=", xmedia
        write(*,*) "media de x**2=", xmq

        close(1)
        end
```

Com o código acima, foi feito um histograma que relaciona a quantidade de andarilhos pela posição x, que está representado na figura 1. Nota-se que o maior número de andarilhos está localizado ao redor de 500, ou seja, a “metade do caminho”. É possível perceber também que o histograma se aproxima do de uma distribuição Gaussiana com a origem deslocada.

1.0.2 Distribuição Gaussiana

Uma Distribuição Gaussiana de probabilidade possui dois parâmetros fundamentais:

- μ: o valor esperado da distribuição, que define onde está o ponto de máximo;

- σ: o desvio padrão, que define a largura da abertura da curva; $\sigma^2 = \langle x^2 \rangle - \langle x \rangle^2$.

Nessa tarefa, $\mu = \langle p \rangle = 500$. Utilizando a equação 3, é possível perceber que o resultado esperado para $\langle x \rangle$ é 0, que está de acordo com o resultado obtido pelo programa, no qual a média de $\langle x \rangle \propto 10^{-2}$ e, portanto, $\langle x \rangle \to 0$ e $\sigma = \sqrt{1000 - 0} \approx 32$.

A função densidade de probabilidade desse tipo de distribuição é dada por

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$$

Que, aqui, fica

$$f(x) = \frac{1}{32\sqrt{2\pi}} \exp\left(-\frac{(x-500)^2}{2048}\right)$$

Note que, para a posição $x = 500$, $f(x)$ é, realmente, máxima.

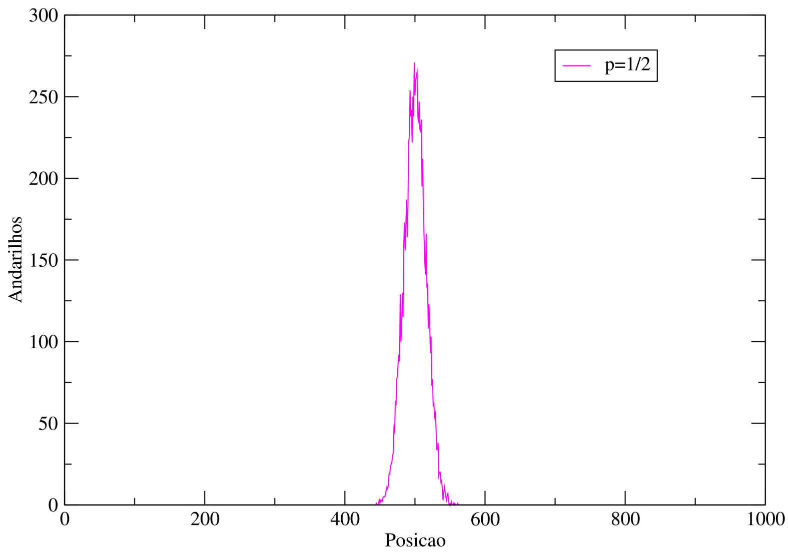

Figura 1: Número de andarilhos em função da posição para $p = \frac{1}{2}$.

B

Refaça o item anterior considerando $p = \frac{1}{3}, \frac{1}{4}, \frac{1}{5}$. Qual seria a forma analítica, em termos de N, p e q, para $\langle x \rangle$ e $\langle x^2 \rangle$?

O código anterior foi alterado três vezes ($p = \frac{1}{3}, \frac{1}{4}, \frac{1}{5}$) sendo que, para cada um dos casos, o *if* foi alterado para $x < 0.33$, $x < 0.25$ e $x < 0.20$, respectivamente. Foram criados também três arquivos para armazenar os passos em cada uma das situações (*media_033.dat*, *media_025.dat*, *media_02.dat*). A partir desses arquivos, foram plotados os histogramas mostrados nas figuras 2, 3 e 4, respectivamente. De forma a tornar o texto do relatório mais fluido, apenas um dos três códigos foi transcrito, mas os exemplos abrangem as saídas dos três programas.

Utilizando as equações 2 e 3, é possível obter os valores de $\langle x \rangle$ para os três casos, sendo eles:

Exemplos:

- $p = \frac{1}{3}$: analiticamente, $\langle p \rangle = 1000 \cdot \frac{1}{3} = \frac{1000}{3} \approx 333.33$, $\langle x \rangle = 2 \cdot 333.33 - 1000 = -333.34$

```
media x=  -339.274780
media de x**2=   115968.375
```

- $p = \frac{1}{4}$: analiticamente, $\langle p \rangle = 1000 \cdot \frac{1}{4} = \frac{1000}{4} = 250$, $\langle x \rangle = 2 \cdot 250 - 1000 = -500$

```
media x=  -499.392609
media de x**2=   250128.500
```

- $p = \frac{1}{5}$: analiticamente, $\langle p \rangle = 1000 \cdot \frac{1}{5} = \frac{1000}{5} = 200$, $\langle x \rangle = 2 \cdot 200 - 1000 = -600$

```
media x=  -599.323975
media de x**2=   359703.875
```

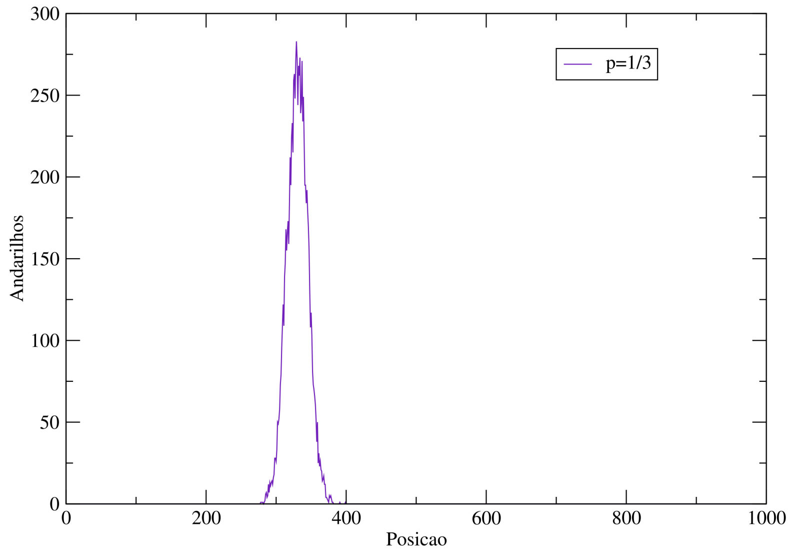

Figura 2: Número de andarilhos em função da posição para $p = \frac{1}{3}$.

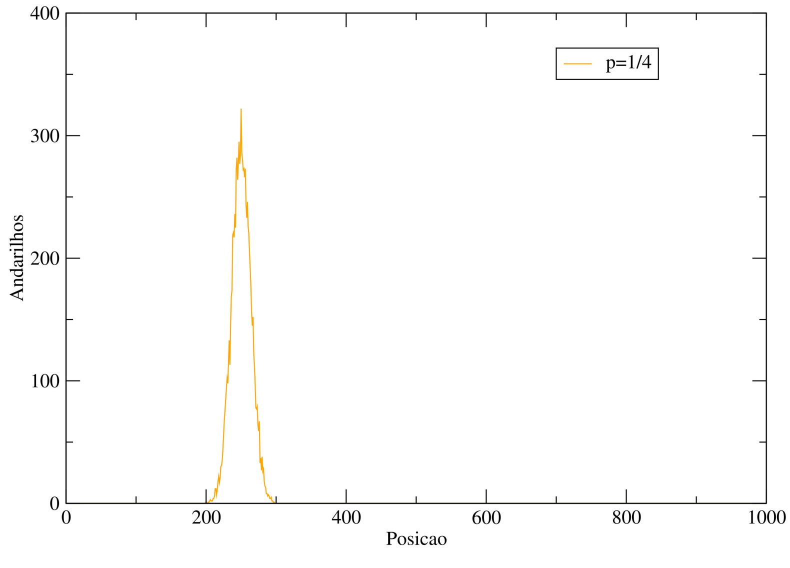

Figura 3: Número de andarilhos em função da posição para $p = \frac{1}{4}$.

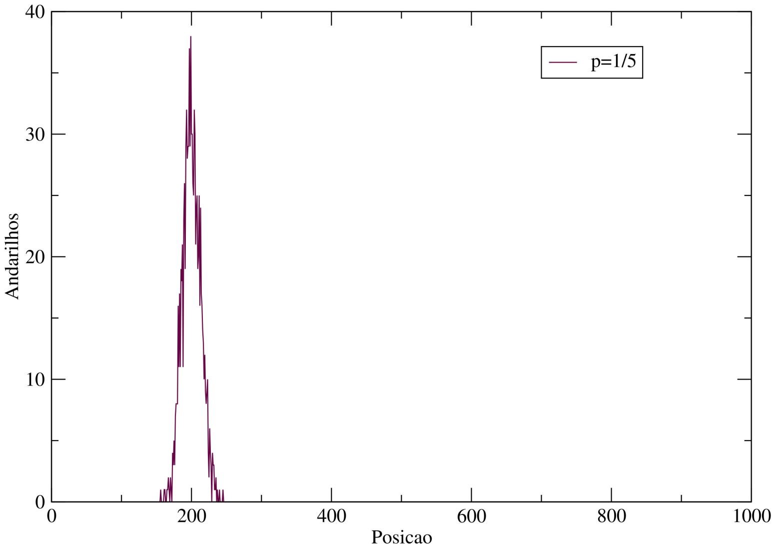

Figura 4: Número de andarilhos em função da posição para $p = \frac{1}{5}$.

Código:

```
dimension npd(0:10000)

n=1000
m=10000
pm=0
pm2=0

open(unit=1, file='media_033.dat')

do i=0,m
        nd=0
        do j=0,n
                x=rand()
                if(x .lt. 0.33) then
                        nd=nd+1
                end if
        end do
        pm=pm+nd
        pm2=pm2+nd**2

        npd(nd)=npd(nd)+1
end do
```

```
      do k=0,n
              write(1,*) k, npd(k)
      end do

c media de p
      pmedia=pm/m
      xmedia=(2*pmedia)-n
c media de p**2
      pmq=pm2/m
c formula da media de x**2
      xmq=4*pmq-4*n*pmedia+n**2

      write(*,*) "media x=", xmedia
      write(*,*) "media de x**2=", xmq

      close(1)
      end
```

Tarefa 3

Andarilhos aleatórios 2D

Considere agora o caso do andarilho bidimensional não enviesado, i.e., com iguais chances ($\frac{1}{4}$) de dar um passo em qualquer direção dos pontos cardeais: norte, sul, leste e oeste. Calcule $\langle \vec{r} \rangle$ e $\Delta^2 = \langle \vec{r} \cdot \vec{r} \rangle - \langle \vec{r} \rangle \cdot \langle \vec{r} \rangle$. Repare que estes andarilhos perfazem o mesmo tipo de movimento que moléculas no processo de difusão, como, por exemplo, a difusão de um pingo de leite numa xícara de café. Faça um diagrama das posições das moléculas após um número N de passos ($N = 10, 10^2, 10^3, 10^4, 10^5, 10^6$).

O código produzido para essa tarefa é uma adaptação dos códigos anteriores para o caso bidimensional: agora, existem 4 possibilidades para o passo de cada andarilho, cada uma delas com $p = \frac{1}{4}$, e os passos possuem coordenadas (x, y). Pelo caráter bidimensional do problema, foi necessária a definição de um vetor *npb* com dimensões M e 2, sendo definido 1 = coordenada x e 2 = coordenada y. No enunciado, Δ^2 equivale ao σ^2 discutido anteriormente (desvio padrão da distribuição) e, no código, *delta* representa Δ^2. É esperado que Δ seja da ordem de $\sqrt{N}$ e, portanto, *delta* seja da ordem de N. O valor de M foi escolhido como 1000.

Foram realizados testes para $N = 10^1, 10^2, 10^3, 10^4, 10^5, 10^6$ e com eles foram plotados os histogramas mostrados nas figuras 6 e 5. Ademais, os resultados de dois deles ($N = 100$ e $N = 10000$) gerados no terminal são mostrados detalhadamente abaixo:

Exemplos:

```
digite N:
100
r=   267.738678
media=  0.267738670
media quadrada=   7.16839954E-02
quadrado da media=   94.5839996
delta=   94.5123138
```

```
digite N:
100000
r=   5528.42188
media=   5.52842188
media quadrada=   30.5634480
quadrado da media=   98396.4453
delta=   98365.8828
```

O valor N foi alterado no terminal para cada um dos valores pedidos no enunciado, e os resultados obtidos pelo código foram guardados nos arquivos *media_r.dat*, *media_r1.dat*, *media_r2.dat*, *media_r3.dat*, *media_r4.dat* e *media_r5.dat*. A dispersão dos andarilhos no espaço mostra que, quanto maior o número de passos, mais "espalhados"eles estão. Esse fenômeno é resultado do comportamento usual de moléculas (aqui, andarilhos) de busca por equilíbrio a partir da saída de uma região de alta concentração para uma de menor concentração, evidenciando a presença de um gradiente de concentração de moléculos quando o movimento das mesmas é aleatório.

Apesar das moléculas ficarem cada vez mais dispersas, é visível que há uma maior concentração delas na origem, fato que decorre da média de r $\langle r \rangle$ ser muito pequena em comparação com N.

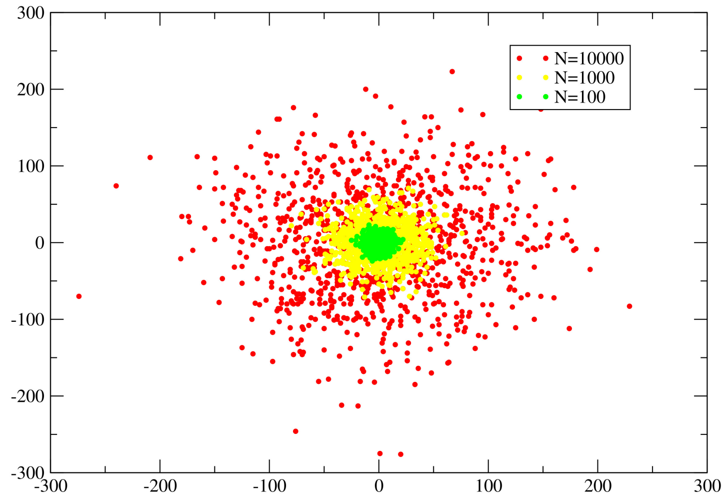

Figura 5: Dispersão dos andarilhos no espaço para $N = 100, 1000$ e 10000.

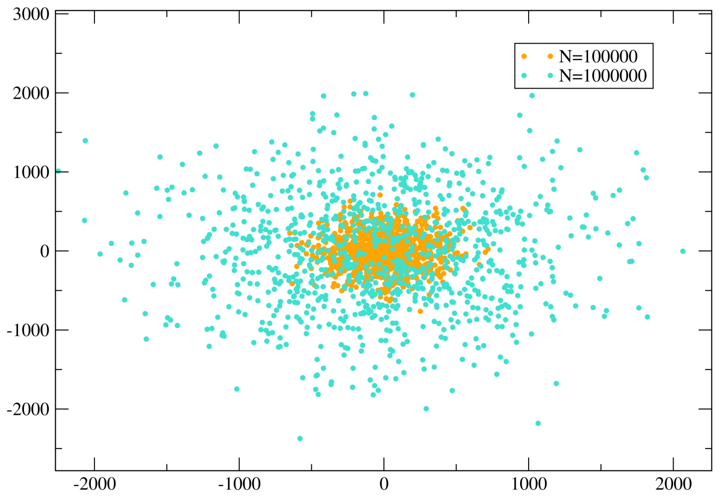

Figura 6: Dispersão dos andarilhos no espaço para $N = 100000$ e 1000000.

Código:

```
      parameter (m=1000)
c vetor que controla cada andarilho
      dimension npb(m,2)

      write(*,*) "digite N:"
      read(*,*) n

      rx=0.0
      ry=0.0
      rx2=0.0
      ry2=0.0

      open(unit=1, file='media_r5.dat')

      do i=1,m
              nx=0
              ny=0
c npb(i,1)->controla os passos em x
c npb(i,2)->controla os passos em y
              npb(i,1)=0
              npb(i,2)=0
                              do j=1,n
                      r=rand()
c sao 4 casos possiveis, todos de igual probabilidade
                      if(r .lt. 0.5) then
c andou para leste
                              if(r .lt. 0.25) then
                                      nx=nx+1
                              else
c andou para oeste
                                      nx=nx-1
                              end if
                      else
c andou para o norte
                              if(r .lt. 0.75) then
                                      ny=ny+1
                              else
c andou para o sul
                                      ny=ny-1
                              end if
                      end if
              end do
c raios medios em x e y
              rx=rx+nx
              ry=ry+ny
              rx2=rx2+nx**2
              ry2=ry2+ny**2

c armazeno as coordenadas x e y do andarilho
              npb(i,1)=nx
              npb(i,2)=ny
              write(1,*) npb(i,1), npb(i,2)
      end do
```

```
c distancia entre as coordenadas
        d=sqrt(rx**2+ry**2)
c raios medios
        rm=d/m
        rm2=(rx2+ry2)/m

        write(*,*) "r=", d
        write(*,*) "media=", rm
        write(*,*) "media quadrada=", rm**2
        write(*,*) "quadrado da media=", rm2
        write(*,*) "delta=", abs(rm**2-rm2)

        close(1)
        end
```

Tarefa 4

Entropia

Vamos verificar o aumento da entropia e a flecha do tempo no exercício anterior. Calcule a entropia como função do número de passos N das moléculas (que é proporcional ao tempo $t = N\,\Delta t$, onde Δt é o intervalo de tempo médio entre passos). A entropia é dada por

$$S = -\sum_i P_i \ln P_i,$$

onde P_i é a probabilidade de se encontrar o sistema em um certo micro-estado i. Para se definir o micro-estado i, definimos um reticulado (muito maior que o tamanho de um passo) e contamos quantas moléculas encontramos em cada célula do reticulado.

1.0.3 Entropia de Shannon

A Entropia de Shannon é uma forma de descrever a incerteza de uma informação. Para uma variável aleatória, a Teoria da Informação mostra que a entropia é a forma de se analisar quais são os possíveis arranjos na comunicação de uma "mensagem". Matematicamente,

$$S = -\sum_i p_i \ln p_i \tag{5}$$

Sendo S a Entropia de Shannon e p_i as possibilidades de arranjos. A entropia usual da termodinâmica difere da equação 5 por uma constante, sendo, assim, inevitável correlacioná-las: na termodinâmica, p_i é a probabilidade do sistema ser encontrado em um microestado i. Ademais, sabe-se que a entropia tende a aumentar com o tempo em um sistema isolado.

Nessa tarefa, considerou-se um intervalo de tempo $\Delta t = 1$ entre cada passo dado por um andarilho, de modo que $N \cdot \Delta t = N \cdot 1 = N$. Além disso, p_i é a probabilidade de um andarilho ser encontrado dentro de um certo espaço previamente delimitado como um quadrado de lado muito maior do que o desvio padrão σ da distribuição (equivalente ao tamanho do passo). Sendo $\sigma \propto \sqrt{N}$ e N tendo sido escolhido como 2000, definiu-se um parâmetro *lado*= 100. Finalmente, tem-se que

$$p_i = \frac{\text{quantidade de andarilhos encontrados no interior do quadrado}}{M}$$

Com M sendo a quantidade total de andarilhos, que aqui foi definida como $M = 1000$.

Os resultados obtidos foram armazenados em um arquivo ($entropia.dat$), que foi utilizado para a produção do histograma mostrado na figura 7. A base do programa é bastante semelhante aos anteriores, mas algumas alterações importantes são:

- Para que o cálculo da entropia fosse realizado corretamente, foi necessária a criação de um "grid" de zeros com dimensões do lado do quadrado, para que fosse possível armazenar *isoma* (vetor de controle dos andarilhos que são encontrados no interior do quadrado) para cada um dos andarilhos;
- o loop mais externo agora é o de passos, pois é preciso que cada andarilho dê N passos;
- dois loops foram criados para o cálculo da entropia, cada um deles percorrendo todo o quadrado de interesse em uma direção (x e y).

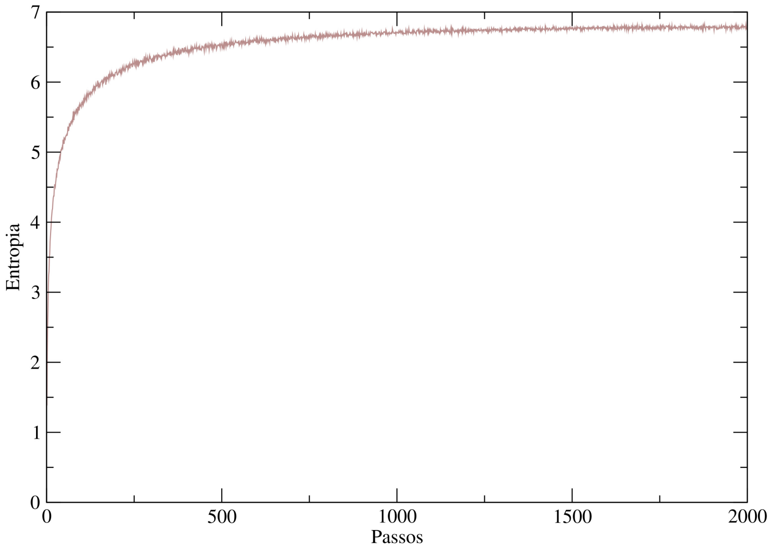

Figura 7: Entropia em função do número de passos.

Código:

```
      parameter(m=1000, n=2000, lado=100)
c vetor de controle de qnts andarilhos estao no interior do
c quadrado de lado=100
      dimension isoma(-lado:lado,-lado:lado)

      open(unit=1, file='entropia.dat')

c loop de passos
      do i=1,n
            s=0.0
c zera contador de posições para cada passo, criando um quadrado
            do ix=-lado,lado
                  do iy=-lado,lado
                        isoma(ix,iy)=0
                  end do
            end do

c loop de andarilhos
            do j=1,m
                  nx=0
                  ny=0
                  do k=1,i
                        r=rand()
```

```
c direcoes norte, sul, leste e oeste
                                if(r .lt. 0.25) then
                                        nx=nx+1
                                else if(r .lt. 0.50) then
                                        nx=nx-1
                                else if(r .lt. 0.75) then
                                        ny=ny+1
                                else
                                        ny=ny-1
                                end if
                        end do

c checo se as coordenadas do passo dado estao no interior do quadrado
                        if(abs(nx) .le. lado) then
                                if(abs(ny) .le. lado) then
                                        isoma(nx,ny)=isoma(nx,ny)+1
                                end if
                        end if
                end do
c calculo da entropia dos andarilhos no interior do quadrado
                do ix=-lado,lado
                        do iy=-lado,lado
                                if(isoma(ix,iy) .gt. 0) then
                                        p=float(isoma(ix,iy))/m
                                        s=s-p*alog(p)
                                end if
                        end do
                end do

                write(1,*) i, s
        end do

        close(1)
        end
```

Referências

[1] DataCamp. Gaussian distribution: A comprehensive guide. DataCamp Tutorial, 2024. URL: https://www.datacamp.com/tutorial/gaussian-distribution.

[2] Kevin Hartnett. How claude shannon's concept of entropy quantifies information. Quanta Magazine, September 2022. URL: https://www.quantamagazine.org/how-claude-shannons-concept-of-entropy-quantifies-information-20220906/.

[3] Wikipedia contributors. Entropy (information theory). Wikipedia, The Free Encyclopedia, 2025. URL: https://en.wikipedia.org/wiki/Entropy_(information_theory).

[4] Wikipedia contributors. Markov chain. Wikipedia, The Free Encyclopedia, 2025. URL: https://en.wikipedia.org/wiki/Markov_chain.

[5] Wikipedia contributors. Random walk. Wikipedia, The Free Encyclopedia, 2025. URL: https://en.wikipedia.org/wiki/Random_walk.