

3 Revisão de Probabilidade e Estatística

O nosso estudo dos sistemas microscópicos e seus reflexos nos sistemas macroscópicos evidenciou a necessidade de realizarmos um tratamento estatístico para o problema. Não há a menor possibilidade de calcularmos a dinâmica das partículas individuais. O grande número de partículas envolvidas, o grande número de eventos (p.ex., colisões) em um certo intervalo de tempo e a perda de informação (precisão) em qualquer cálculo envolvendo essas partículas e eventos torna impossível qualquer tentativa de acompanhar a dinâmica das partículas individualmente. Temos, portanto, que trabalhar com médias estatísticas e conceitos de probabilidade tornam-se importantes (como já mencionamos, Maxwell chegou a essa conclusão em torno de 1850 quando começou a buscar uma descrição para a termodinâmica em termos das partículas elementares). Para nos equiparmos para essa tarefa, será útil revisarmos alguns conceitos e elementos da teoria de probabilidade e estatística elementar. O livro do Salinas (ref. 1), os dois livros do Reif (refs. 2 e 3) e o livro da S. Vauclair trazem uma revisão do assunto. Novamente, o projeto STP (ref.) trás um capítulo sobre o assunto e algumas simulações que utilizaremos para ilustrar os conceitos. Para uma consulta mais detalhada sobre probabilidade e estatística, ver a ref. 6.

3.1 Probabilidade

O primeiro conceito que temos que definir é o de *probabilidade*. Para definirmos a probabilidade que um evento ocorra, é necessário a existência de uma *amostragem*, ou seja, um conjunto finito de eventos possíveis. Como exemplo, consideremos uma moeda, com duas possibilidades de resultados após ser jogada ao ar: cara ou coroa. A *amostragem* tem dois elementos. Ou um dado normal, onde a amostragem possui seis elementos, as seis faces do dado. A definição de probabilidade é *o número que expressa de forma quantitativa a possibilidade que o evento escolhido ocorra no decorrer de uma experiência*. Consideremos uma amostragem com n possíveis resultados do experimento, ou ventos. Os resultados do

experimento são denominados *variável aleatória* ou *variável estocástica*. Vamos associar para cada evento i uma probabilidade $P(i)$ de que ele ocorra quando realizamos um experimento. O conjunto de valores de $P(i)$ é chamado de *distribuição de probabilidades da variável estocástica*. Vamos assumir que cada evento é *mutuamente exclusivo*, isto é, a ocorrência de um evento impede a ocorrência de outro evento no mesmo experimento. O conjunto de A probabilidade $P(i)$ deve satisfazer algumas condições:

$$P(i) \geq 0 \quad (1)$$

e,

$$\sum_i P(i) = 1 \quad (2)$$

onde $P(i) = 0$ significa que o evento nunca ocorre e $P(i) = 1$ que o evento sempre ocorre. A normalização (eq. 2) nos diz que a soma das probabilidades de todos os eventos mutualmente exclusivos é 1.

Utilizando o exemplo da moeda ou do dado podemos facilmente encontrar algumas regras de operação da probabilidade. Para eventos mutualmente exclusivos, se queremos conhecer a probabilidade que um experimento seja ou i ou j , então,

$$P(i \text{ ou } j) = P(i) + P(j) \text{ (regra de adição)} \quad (3)$$

Essa operação pode ser facilmente generalizada para mais do que dois eventos. Dessa regra, extraímos que

$$P(\text{não ocorrer } i) = 1 - P(i) \quad (4)$$

Para o caso de um dado, com seis faces, assumindo a *simetria* entre todos os eventos (isto é, eles são igualmente prováveis), a probabilidade que ocorra uma das faces é $P(i) = 1/6$.

Então, a probabilidade de tirarmos, em um único experimento, 3 ou 6, por exemplo, é $P(3 \text{ ou } 6) = 1/6 + 1/6 = 1/3$.

Se considerarmos agora dois experimentos e assumindo que a ocorrência dos eventos são independentes, a probabilidade que no primeiro experimento tenhamos o evento i e no segundo experimento o evento j , é,

$$P(i \text{ e } j) = P(i)P(j) \text{ (regra de multiplicação)} \quad (5)$$

Para que essa regra seja válida, é importante enfatizar a independência entre os eventos, ou seja, o resultado do segundo experimento não pode depender do resultado do primeiro experimento. Isso é facilmente verificável no caso de uma moeda ou do dado. No entanto, vamos considerar um exemplo diferente. O objetivo é determinar a probabilidade de encontrar, aleatoriamente, uma mulher de mais de 1,8 m de altura. Vamos supor que a probabilidade de encontrar uma pessoa de 1,8 m seja $P(1,8) = 1/10$. A probabilidade que uma pessoa seja mulher é $P(m) = 1/2$. Então, da eq. 5 temos $P(\text{mulher} + 1,8 \text{ m}) = P(m)P(1,8) = \frac{1}{2} \times \frac{1}{10} = \frac{1}{20}$. O mesmo cálculo valeria para um homem. No entanto, sabemos, intuitivamente, pelo menos, que esse resultado está errado. Isso porque a ocorrência dos dois eventos - uma mulher e ter mais de 1,8 m de altura - não são independentes. Por outro lado, se quiséssemos encontrar uma mulher que tivesse nascido em um determinado dia do ano, teríamos $P(m + \text{dia do ano}) = P(m)P(\text{dia do ano}) = \frac{1}{2} \times \frac{1}{365} = \frac{1}{730}$, uma vez que as ocorrências são independentes (pelo menos, dentro do nosso melhor conhecimento).

Precisamos ainda encontrar uma forma de estimar ou determinar o valor das probabilidades. Consideremos dois eventos, E_1 e E_2 . Se E_1 é mais provável de ocorrer que E_2 , então, $P(E_1) > P(E_2)$. A probabilidade está associada a um certo grau de confiança que o evento vá ocorrer e trás consigo um grau de incerteza. Não temos, no entanto, uma única regra para determinarmos o valor de $P(E)$. Há várias formas de buscarmos uma associação entre nosso conhecimento sobre a possibilidade de ocorrer os eventos e encontrarmos um valor para a probabilidade. Vamos considerar o exemplo de uma moeda e tirarmos cara ou coroa

(ou o exemplo do dado). Uma forma de designarmos um valor para $P(E)$ é utilizarmos a *simetria*. Essencialmente, não temos, *a priori*, nenhuma razão que a probabilidade dos eventos ocorrerem seja diferentes e, portanto, associamos a mesma probabilidade para cada um dos eventos, ou seja, $P(i) = 1/n$, onde i é o evento e n a amostragem. No exemplo da moeda, $P(\text{cara}) = P(\text{coroa}) = 1/2$. Uma outra forma de designarmos os valores de $P(E)$ é pela *frequência*, ou seja, por meio do conhecimento ou informação adquirida de experimentos já realizados. Nesse caso, para uma amostragem independente, mutualmente exclusivos, se realizarmos N experimentos, a probabilidade $P(i)$ de ocorrer o evento i é expressa por,

$$P(i) = \lim_{N \rightarrow \infty} \frac{N_i}{N} \quad (6)$$

onde N_i representa o número de vezes em que o evento i ocorreu em N experiências. Essa definição de probabilidade é bastante comum e nos será útil. Observe que a probabilidade é medida por meio de eventos já realizados, ou seja, utilizando um conhecimento que já possuímos. Uma outra forma, ainda, é por meio de *simulações*. Por exemplo, vamos utilizar o programa *MultipleCoins*. A figura mostra o resultado obtido lançando 50 moedas cem vezes. O resultado pode nos dar uma estimativa para $P(E)$, da mesma forma que o resultado obtido com a frequência (eq. 6). Podemos ainda encontrar $P(E)$ por meio de *cálculos teóricos*. O valor da probabilidade, portanto, depende da forma como a determinamos, de quem a determina e do conhecimento existente anteriormente. Portanto, o que podemos associar é a probabilidade do evento E tendo a informação I , $P(E|I)$. E, dependendo da amostragem, diferentes observadores podem associar diferentes valores das probabilidades. Talvez o caso mais ilustrativo seja o caso dos operadores de ações na bolsa de valores.

Nos últimos anos, a teoria de informação bayesiana tem sido utilizada para designar o valor da probabilidade. Essa teoria, baseada em um trabalho do clérigo e matemático amador Thomas Bayes de 1763 (T. Bayes morreu em 1761, o artigo foi encontrado entre a documentação que ele deixou e foi publicado postumamente) e procura determinar uma forma de associar valores de probabilidade para probabilidades condicionadas (para uma discussão

sobre a inferência bayesiana ver a ref. 5).

A definição baseada nas frequências, eq. 6, é uma das mais utilizadas na prática. No entanto, é importante não esquecermos que ela é uma estimativa, uma vez que depende de uma decisão sobre a forma de fazermos a designação a probabilidade. É importante observar que *não há uma forma de realizarmos um experimento para medirmos o valor de $P(E)$* . No caso da frequência, os dados utilizados, resultados de experimentos, são eventos realizados com certeza e, portanto, os conceitos de probabilidade não se aplicam. Na designação da probabilidade há, inherentemente, uma incerteza. Vamos explorar esse conceito um pouco mais.

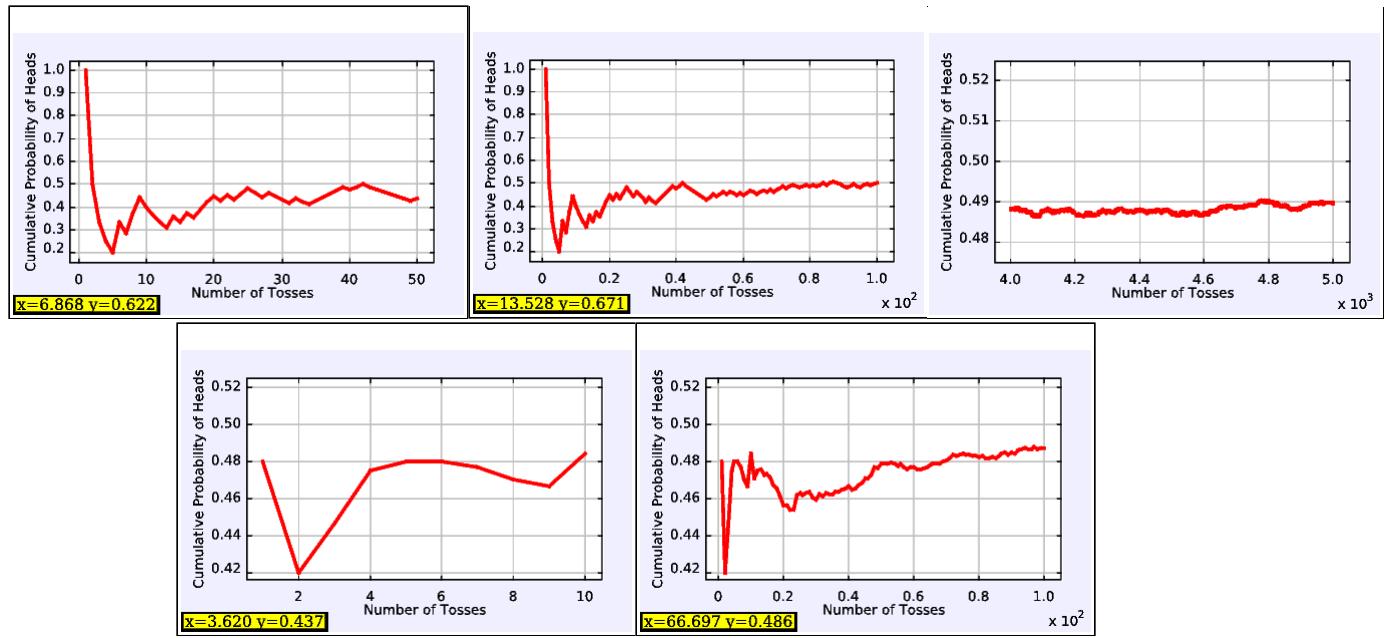


Figura 1: (Superior) Resultados de simulação de obtermos cara ao lançarmos uma moeda 50 vezes, 100 vezes e 5.000 vezes cem vezes e de lançarmos simultaneamente 50 moedas 10 vezes e 100 vezes. (simulação baseada no programa *MultipleCoins*, ref. 5).

3.2 Informação e incerteza

Esta seção está baseada na ref. 5.

Discutimos as ideias básicas de probabilidade e o fato que sempre temos suposições associadas às designações de valores para as probabilidades. É útil discutirmos uma forma de

associarmos o grau de incerteza existente nos experimentos. Para isso, consideremos dois experimentos com dois eventos possíveis, E_1 e E_2 com probabilidades P_1 e P_2 , respectivamente. O experimento pode ser o de jogar uma moeda para o alto. No primeiro experimento vamos assumir eventos simétricos e temos então $P_1 = P_2 = 1/2$. No segundo experimento, por alguma razão (a moeda entortou) a simetria foi perdida e temos $P_1 = 1/5$ e $P_2 = 4/5$. Qual experimento temos maior incerteza? Intuitivamente, diremos que o primeiro, uma vez que podemos estimar com maior segurança o resultado no segundo experimento.

Seguindo esse raciocínio, consideremos mais dois experimentos. O terceiro tem quatro eventos, simétricos, com $P_1 = P_2 = P_3 = P_4 = 1/4$ e o segundo seis eventos simétricos com $P_1 = P_2 = P_3 = P_4 = P_5 = P_6 = 1/6$. Novamente, nossa intuição nos diz que o quarto experimento é o mais incerto uma vez que temos um maior número de resultados possíveis, igualmente prováveis. No cômputo geral, o quarto experimento é o mais incerto e o segundo é o menos incerto.

Para avançarmos, precisamos associar uma medida ao conceito de incerteza. Vamos designar a função $S(P_1, P_2, \dots, P_i, \dots)$, onde P_i é a probabilidade do evento i , como sendo a *função incerteza*. Consideremos o caso simétrico, onde todas as probabilidades são iguais, $P_1 = P_2 = \dots = P_i = \dots = 1/\Omega$, onde Ω é o número de eventos possíveis. Nesse caso, podemos escrever simplesmente

$$S(1/\Omega, 1/\Omega, \dots) = S(\Omega) \quad (7)$$

A função $S(\Omega)$ deve satisfazer algumas condições simples.

1. Se temos apenas um evento possível, isto é, $\Omega = 1$, então

$$S(\Omega = 1) = 0 \quad (8)$$

2. Temos também,

$$S(\Omega_1) > S(\Omega_2) \text{ se } \Omega_1 > \Omega_2 \quad (9)$$

3. Consideremos agora a forma de S para múltiplos eventos. Por exemplo, atirar uma moeda e jogar um dado, com Ω_1 e Ω_2 possíveis eventos, respectivamente. O número total de resultados possíveis ao realizar os dois experimentos é $\Omega_1\Omega_2$. Imaginemos agora que conhecemos o resultado de atirar a moeda. Essa incerteza é zero, portanto. No entanto, permanece a incerteza de jogarmos o dado. Da mesma forma, se conhecemos o resultado de atirar o dado, permanece a incerteza de jogarmos a moeda. Essas considerações sugerem S ter a forma

$$S(\Omega_1\Omega_2) = S(\Omega_1) + S(\Omega_2) \quad (10)$$

Questão: você consegui imaginar outra forma para S para a condição 3? Tente.

As três condições discutidas acima permitem apenas uma única forma funcional. Para encontrá-la, vamos escrever a eq. 10 utilizando variáveis contínuas:

$$S(xy) = S(x) + S(y) \quad (11)$$

onde x, y são variáveis contínuas (essa simplificação não altera o resultado). Vamos chamar $z = xy$ e calcular as derivadas parciais de S :

$$\frac{\partial S(z)}{\partial x} = \frac{\partial z}{\partial x} \frac{dS(z)}{dz} = y \frac{dS(z)}{dz} \quad (12)$$

$$\frac{\partial S(z)}{\partial y} = \frac{\partial z}{\partial y} \frac{dS(z)}{dz} = x \frac{dS(z)}{dz} \quad (13)$$

Da eq. 10, temos,

$$\frac{\partial S(z)}{\partial x} = \frac{dS(x)}{dx} \quad (14)$$

$$\frac{\partial S(z)}{\partial y} = \frac{dS(y)}{dy} \quad (15)$$

Comparando os lados direitos das equações 12-13 e 14-15, temos,

$$\frac{dS(x)}{dx} = y \frac{dS(z)}{dz} \quad (16)$$

$$\frac{dS(y)}{dy} = x \frac{dS(z)}{dz} \quad (17)$$

Multiplicando a eq. 16 por x e a eq. 17 por y , temos,

$$x \frac{dS(x)}{dx} = y \frac{dS(y)}{dy} = z \frac{dS(z)}{dz} \quad (18)$$

O primeiro termo na eq. 18 depende somente de x enquanto que o segundo somente de y . Como x e y são variáveis independentes, os três termos na eq. 18 devem ser iguais a uma constante:

$$x \frac{dS(x)}{dx} = y \frac{dS(y)}{dy} = A \quad (19)$$

Integrando a equação 19, temos,

$$S(x) = A \ln x + B \quad (20)$$

A constante de integração B deve ser nula para satisfazer a condição 1) (eq. 8). A constante A é arbitrária e podemos escolher $A = 1$. A incerteza S para uma amostragem simétrica escreve-se na forma,

$$S(\Omega) = \ln\Omega \quad (21)$$

A expressão 21 foi deduzida para amostragem simétrica. Como fica o caso geral? Não vamos deduzir aqui mas para quem estiver interessado, ver o Apêndice F da ref. 7. O resultado é,

$$S = - \sum_i P_i \ln P_i \quad (22)$$

Vamos analisar o caso que deduzimos. Se todas as probabilidades são iguais, temos,

$$P_i = \frac{1}{\Omega} \text{ (para todos } i) \quad (23)$$

e então,

$$S = - \sum_i \frac{1}{\Omega} \ln \frac{1}{\Omega} = \Omega \frac{1}{\Omega} \ln \Omega = \ln \Omega \quad (24)$$

É fácil verificar que se o resultado j é uma certeza absoluta, então, $P_j = 1$ e $P_i = 0$ se $i \neq j$ e então $S = -1\ln 1 = 0$, como era de se esperar: *tendo certeza do resultado, não há falta de informação.*

O resultado da eq. 22 nada mais é que a entropia de Shannon, que introduzimos antes. Não fizemos (ainda) nenhuma demonstração em termos de entropia nem identificamos a constante que precede a expressão para a incerteza S . Apenas derivamos - ou melhor, indicamos - um resultado para a incerteza em termos das probabilidades de transmissão da informação.

Sabemos agora como calcular a incerteza ou falta de informação S se conhecemos as probabilidades P_i . Em geral, nosso problema é conhecer as probabilidades. Consideremos o caso mais simples, de uma moeda jogada. Nesse caso, por simetria, ou intuição (assumir que a moeda é simétrica e não viciada), esperamos o resultado $P_1(\text{cara}) = P_2(\text{coroa}) = 1/2$. Só alteraríamos essa expectativa se tivéssemos alguma informação a mais. Na prática, utilizamos

intuitivamente o que se conhece como princípio de *menor viés* ou *máxima incerteza*. Vejamos como se aplica para esse caso, lembrando que $P_1 + P_2 = 1$. O valor de S é

$$\begin{aligned} S &= - \sum_i P_i \ln P_i = -(P_1 \ln P_1 + P_2 \ln P_2) \\ &= -[P_1 \ln P_1 + (1 - P_1) \ln(1 - P_1)] \end{aligned} \quad (25)$$

Para maximizarmos a incerteza, temos que ter,

$$\begin{aligned} \frac{dS}{dP_1} &= -[\ln P_1 + 1 - \ln(1 - P_1) - 1] = -\ln \frac{P_1}{1 - P_1} = 0 \\ \Rightarrow \frac{P_1}{1 - P_1} &= 1 \\ \Rightarrow P_1 &= \frac{1}{2} \end{aligned} \quad (26)$$

Para verificarmos que temos um máximo, basta calcular a segunda derivada:

$$\frac{d^2S}{dP_1^2} = -\left[\frac{1}{P_1} + \frac{1}{1 - P_1}\right] = -4 < 0 \quad (27)$$

Exercício: Considere um experimento de jogar um dado de três faces com três possíveis resultados (eventos): E_1 , E_2 e E_3 , com faces 1, 2 e 3, respectivamente. Após jogar várias vezes o dado, sabemos que o número médio de pontos é $f = 1,9$, mas não conhecemos as probabilidades individuais. Quais são os valores de P_1 , P_2 e P_3 que maximizam a incerteza e que seja consistente com a informação que $f = 1,9$? (ver ref. 5)

3.3 Valores médios

A especificação da *distribuição das probabilidades* $P(1), P(2), \dots, P(n)$ para os n possíveis valores da variável estocástica (aleatória) é a descrição mais completa possível do sistema. No entanto, em muitos casos, é preferível descrever a distribuição dos possíveis valores da variável com menos detalhes. A forma mais comum é o *número médio* de x , que podemos chamar de $\bar{x} = \langle x \rangle$:

$$\begin{aligned}\bar{x} &\equiv x_1 P(1) + x_2 P(2) + \dots + x_n P(n) \\ &= \sum_{i=1}^n x_i P(i)\end{aligned}\tag{28}$$

Da mesma forma, uma função $f(x)$ de x tem sua média calculada por,

$$\overline{f(x)} = \sum_{i=1}^n f(x_i) P(i)\tag{29}$$

É fácil verificar que,

$$\overline{f(x) + g(x)} = \overline{f(x)} + \overline{g(x)}\tag{30}$$

$$\overline{cf(x)} = c\overline{f(x)}\tag{31}$$

onde f e g são duas funções de x e c é uma constante.

Definimos os *momentos* da distribuição de probabilidades P ,

$$\overline{x^m} \equiv \sum_{i=1}^n x_i^m P(i)\tag{32}$$

onde fizemos simplesmente $f(x) = x^m$. Obviamente, o valor médio da distribuição de probabilidades é o seu primeiro momento.

O desvio dos valores do experimento em relação ao seu valor médio tem grande importância na estatística. O desvio de x é definido como

$$\Delta x \equiv x - \bar{x} \quad (33)$$

e, obviamente,

$$\overline{\Delta x} = \overline{(x - \bar{x})} = \bar{x} - \bar{x} = 0 \quad (34)$$

A largura média de uma distribuição onde o evento j seja o único possível, isto é, $P(i) = 1$, $i = j$, e zero em todos os outros casos, tem uma largura nula. Podemos expressar a “largura” da distribuição dos valores dos experimentos em torno da média calculando o *desvio quadrático*:

$$\begin{aligned} \overline{\Delta x^2} &\equiv \overline{(x - \bar{x})^2} \\ &= \overline{(x^2 - 2x\bar{x} + \bar{x}^2)} \\ &= \overline{x^2} - 2\overline{x\bar{x}} + \overline{x^2} \\ &= \overline{x^2} - \bar{x}^2 \end{aligned} \quad (35)$$

Essa grandeza também é conhecida como *dispersão* ou *variância* e a sua raiz quadrada é o *desvio padrão*. Podemos também definir os momentos de ordem m em relação a média,

$$\overline{x^m} = \overline{(x - \bar{x})^m} \quad (36)$$

É possível reconstruir a distribuição de probabilidades se conhecemos momentos de todas as ordens.

3.4 O processo de Bernoulli e a distribuição binomial: analisando um conjunto de spins não-interagentes ou o “caminho do bêbado”

Nosso objetivo agora é aprender como calcular probabilidades e avançarmos na compreensão dos sistemas com muitas partículas. Vamos estudar o caso fundamental da *distribuição binomial*. Esse caso aplica-se para sistemas estatísticos (ou físicos) para os quais existe apenas dois resultados possíveis para o experimento. Consideremos alguns exemplos.

Jogar uma moeda Já vimos utilizando esse exemplo e é o mais simples de todos. Temos dois resultados possíveis, cara ou coroa. Cada vez que atiramos a moeda, o resultado é independente dos resultados anteriores. Chamemos de p a probabilidade de sair cara e q a probabilidade de sair coroa. Para uma moeda não-viciada, temos $p = 1/2$ e $q = 1 - p = 1/2$. O que queremos determinar é, jogando N vezes a moeda, qual a probabilidade de sair n vezes coroa, por exemplo.

Momentos magnéticos não-interagentes Esse é um problema comum na física. Consideremos um sistema com N dipolos magnéticos não-interagentes com momento magnético μ na presença de um campo magnético externo B . Orientemos o eixo de coordenadas tal que $\vec{B}/|\vec{B}| = \hat{z}$. Da mecânica quântica, sabemos que a energia do momento de dipolo é paralelo ou anti-paralelo ao campo magnético. A energia de interação do momento magnético é $E = \mp\mu B$, onde o sinal superior (inferior) refere-se ao momento de dipolo paralelo (anti-paralelo). Podemos escrever na forma $E = -s\mu B$, onde $s = \pm 1$, onde o sinal $+(-)$ refere-se ao momento de dipolo paralelo (anti-paralelo). Vamos chamar s de *spin* do momento magnético.

Nota: Em sistemas reais, os momentos magnéticos interagem entre si. O modelo que estamos discutindo aqui ilustra um caso extremo e permite aplicar a distribuição binomial. Se estivéssemos considerando um gás de elétrons, o momento magnético seria $\mu_B =$ magnetão de Bohr $= e\hbar/2m_e$, a energia de interação é $E = -g_s\mu_B\vec{s} \cdot \vec{B}$, onde $g_s \approx 2$ e $\vec{s} = \pm \frac{1}{2}\hat{z}$.

Vamos chamar de p a probabilidade do spin ser $+1$ e q a probabilidade de ser -1 , de forma que $p+q = 1$. Na ausência de um campo magnético, devemos esperar uma distribuição aleatória, ou seja, $p = q = 1/2$. O que queremos saber é conhecer a probabilidade que a magnetização do sistema N momentos magnéticos tenha uma certa magnetização M ,

$$M = \mu(s_1 + s_2 + \dots + s_N) = \mu \sum_{i=1}^N s_i \quad (37)$$

O problema é equivalente ao problema de atirarmos uma moeda.

Caminho aleatório ou o “andar de bêbado” (*drunkard walk*) Consideremos um “bêbado” que inicia sua caminhada a partir de um poste de luz e segue erraticamente, realizando passos sucessivos em direções aleatórias. A questão que se faz é *a que distância ele encontra-se do poste após N passos?* Vamos considerar um caso ideal, simplificado, onde o “bêbado” pode andar apenas em linha reta (uma única dimensão). Assumimos que cada passo seja igual em distância, l e que a cada intervalo de tempo (idêntico, Δt) ele dá um passo ou para a direita com probabilidade p ou para a esquerda com probabilidade $q = 1 - p$. A direção de cada novo passo independe dos passos anteriores. Após um certo intervalo de tempo, tendo realizado N passos no total, ele terá dado n passos a direita e n' passos a esquerda, tais que $n + n' = N$. (Para quem tiver interesse, a história desse problema está descrita na Parte I da ref. 8).

Os três problemas mencionados acima tem a mesma estrutura matemática e, portanto, a mesma solução. Esse tipo de problema é conhecido como *processo de Bernoulli* (do matemático Jacob Bernoulli, 1654-1705). O que queremos conhecer é qual a probabilidade $P_N(n)$ que o sistema - vamos utilizar como exemplo o caso do “caminhar do bêbado” - tenha dado n passos a direita. Podemos escrever $P_N(n)$ como sendo,

$$P_N(n) = W_N(n, n') p^n q^{n'} \quad (38)$$

Há várias formas de determinarmos o valor de $W_N(n, n')$, que nada mais é do que o número de *microestados* de N passos com n à direita e n' à esquerda. Ou, se recordarmos o exemplo das partículas na caixa com uma partição central, n partículas no compartimento da direita e n' no compartimento da esquerda. Já estudamos o caso particular de $N = 2$ e $N = 16$ (ver Capítulo 2). Para o caso geral, vamos resolver o problema inicialmente utilizar uma relação de recorrência entre W_N e W_{N-1} . Um total de n passos à direita e n' passos à esquerda de um total de N passos, pode ser encontrado adicionando um passo a $N - 1$ passos. Sequências em que a probabilidade em um dado instante depende apenas dos valores das probabilidades no instante anterior são conhecidas como *cadeias de Markov* e têm grande importância na física. Há duas possibilidades para o próximo passo:

- 1) um passo à direita, se já temos $(n - 1)$ passos à direita e n' passos à esquerda;
- 2) um passo à esquerda, se já temos n passos à direita e $(n' - 1)$ passos à esquerda.

Temos $W_N(n - 1, n')$ possíveis maneiras de estarmos no primeiro caso e $W_{N-1}(n, n' - 1)$ de estarmos no segundo caso. Logo,

$$W_N(n, n') = W_{N-1}(n - 1, n') + W_{N-1}(n, n' - 1) \quad (39)$$

Os valores “iniciais” são conhecidos: $W_0(0, 0) = 0$, $W_1(1, 0) = W_1(0, 1) = 1$. A partir deles, podemos construir $W_N(n, n')$ para qualquer valor de N :

$$\begin{aligned} W_2(2, 0) &= W_1(1, 0) + W_1(2, -1) = 1 + 0 = 1 \\ W_2(1, 1) &= W_1(0, 1) + W_1(1, 0) = 1 + 1 = 2 \\ W_2(0, 2) &= W_1(-1, 2) + W_1(0, 2) = 0 + 1 = 1 \end{aligned} \quad (40)$$

Podemos continuar essa sequência indefinidamente. Arranjando-a geometricamente, temos uma estrutura piramidal ou como é conhecido, o triângulo de Pascal, representado na figura, onde a $N - \text{ésima}$ linha corresponde aos valores de $W_N(n, n')$, iniciando com $W_N(N, 0)$, à

esquerda e indo até $W_N(0, N)$ à direita.

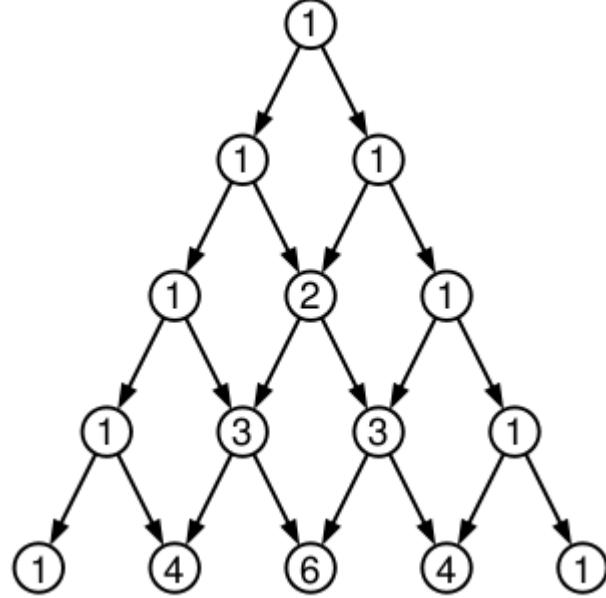


Figura 2: Construção do triângulo de Pascal que reproduz os valores dos coeficientes $W_N(n, n')$. Cada componente é a soma dos dois componentes acima, à esquerda e à direita. A N – éssima linha corresponde aos valores de $W_N(n, n')$, iniciando com $W_N(N, 0)$, à esquerda e indo até $W_N(0, N)$ à direita. (extraído da ref. 5)

Por indução, a solução geral se escreve,

$$W_N(n, n') = \frac{N!}{n!n'} = \frac{N!}{n!(N-n)!} \quad (41)$$

onde utilizamos $0! = 1$. O resultado para $P_N(n)$ é,

$$P_N(n) = \frac{N!}{n!(N-n)!} p^n q^{(N-n)} \quad (42)$$

que é a *distribuição binomial*. Para o caso particular em que $p = q = 1/2$, temos,

$$P_N(n) = \frac{N!}{n!(N-n)!} 2^{-N} \quad (43)$$

Análise combinatória Poderíamos ter chegado ao mesmo resultado partindo da análise combinatória. Já vimos que se temos dois estados possíveis para cada experimento, o número de estados possíveis para um sistema com N elementos (experimentos) é 2^N e, considerando todos os estados igualmente prováveis, a probabilidade do sistema encontrar-se em um desses estados é $1/2^N$. A *permutação* de N elementos é o arranjo desses elementos em uma certa ordem. O número de permutações se calcula facilmente: temos N possibilidades para o primeiro elemento (resultado), $N - 1$ possibilidades para o segundo elemento, $N - 2$ para o terceiro, e assim por diante. O número total de permutações será $N!$. Consideremos o caso em que queremos ordenar apenas n elementos. Então, temos N possibilidades para o primeiro elemento, $N - 1$ para o segundo, e assim por diante, até $N - n + 1$ possibilidades para o n -ésimo elemento. O número de arranjos n à n de N elementos é,

$$A_N^n = \frac{N!}{(N - n)!} \quad (44)$$

Consideremos agora que os n elementos sejam indistinguíveis (como é o caso no nosso exemplo - tanto faz qual a ordem dos elementos, o resultado “macroscópico” final é o mesmo). Temos que calcular o número possível de sub-conjuntos de N contendo n elementos. Chamaremos esse número de *combinações* possíveis de n elementos em N : C_N^n . O número que encontramos anteriormente, A_N^n é o número de combinações C_N^n multiplicado pelo número de permutações possíveis dos n elementos do sub-conjunto considerado (ou seja, o número de variações possíveis dos objetos indistinguíveis):

$$A_N^n = C_N^n n! \quad (45)$$

e então,

$$C_N^n = \frac{N!}{n!(N - n)!} \quad (46)$$

que é o número de estados possíveis com n passos à direita em N passos, por exemplo.

O resultado que obtivemos tem seu nome no fato que o resultado é o mesmo que obtemos em um *desenvolvimento binomial*:

$$\begin{aligned}
 (x+y)^N &= x^N + Nx^{N-1}y + \frac{1}{2}N(N-1)x^{N-2}y^2 + \dots + \frac{N!}{n!(N-n)!}x^{N-n}y^n + \dots + y^N \\
 &= \sum_{n=0}^N \frac{N!}{n!(N-n)!}x^{N-n}y^n
 \end{aligned} \tag{47}$$

Escolhendo, $x = y = 1$, temos,

$$\sum_{n=0}^N C_N^n = 2^N \tag{48}$$

que é o número de estados possíveis do sistema.

A equação 47 nos mostra também a normalização de $W_N(n, n')$:

$$(p+q)^N = \sum_{n=0}^N \frac{N!}{n!(N-n)!} p^n q^{N-n} = 1 \tag{49}$$

onde utilizamos $p + q = 1$.

Antes de avançarmos na nossa análise, vamos examinar alguns resultados de simulações.

Para isso, vamos utilizar o programa *stp_binomial* da ref. 5. A figura 3 mostra o resultado da simulação para $N = 20, 40, 60, 80, 100$ e $p = 0, 5$ e $0, 7$.

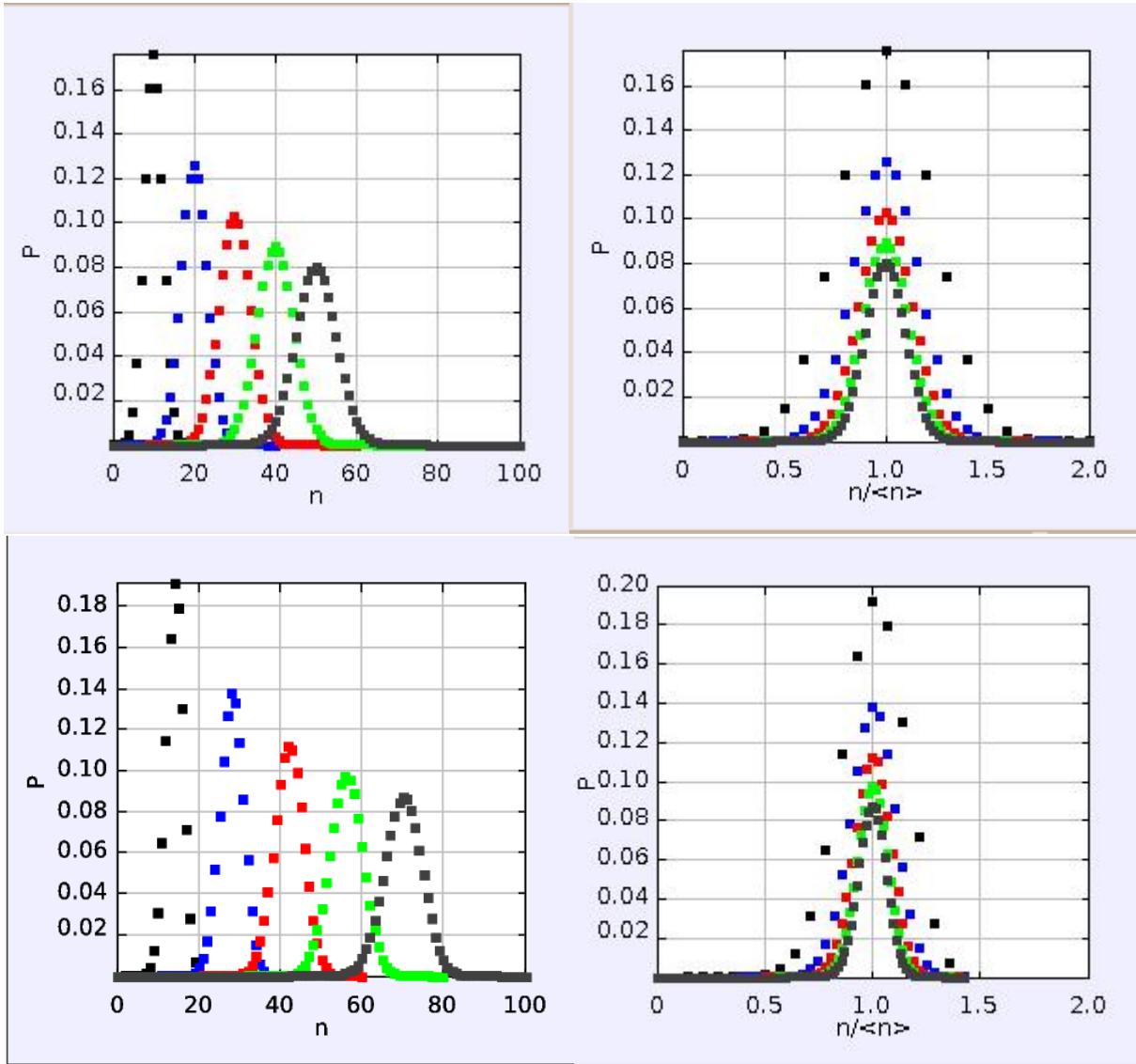


Figura 3: Resultados de simulações da distribuição binomial, $P_N(n)$ em função de n e n/\bar{n} , para $N = 20$ (preto), 40(azul), 60(vermelho), 80(verde), 100(cinza), e $p = 0,5$ (superior) e 0,7 (inferior). Calculadas com o programa *stp-binomial* da ref. 5.

Vamos analisar os resultados. Olhando os dados para $p = 0,5$, observamos que o valor máximo desloca-se para a direita com N , o que é de se esperar. Ao mesmo tempo, a largura da distribuição aumenta com N . Quando observamos em função de n/\bar{n} , a largura diminui com N . O valor máximo de $P_N(n)$ diminui com N . Quando alteramos a distribuição de uma distribuição simétrica para $p = 0,7$, vemos um deslocamento do valor máximo para a direita ao mesmo tempo que o valor da largura diminui quando graficado em função de n/\bar{n}

em relação aos valores para $p = 0,5$. O resultado para $p = 0,7$ pode ser interpretado, no nosso exemplo de spins não-interagentes, como sendo o conjunto de spins na presença de um campo magnético.

Podemos analisar esses resultados quantitativamente, uma vez que conhecemos a distribuição de probabilidades. É o que faremos agora.

Valores médios e desvio padrão do “caminhar do bêbado” Vamos analisar um pouco os resultados obtidos até agora. De posse do valor de $P_N(n)$ podemos calcular alguns valores que caracterizam o sistema físico desejado (ou melhor, qualquer um dos problemas mencionados inicialmente).

Inicialmente, vamos calcular o valor médio, \bar{n} :

$$\bar{n} = \sum_{n=0}^N n P_N(n) = \sum_{n=0}^N n \frac{N!}{n!(N-n)!} p^n q^{N-n} \quad (50)$$

Para isso, vamos utilizar a relação,

$$p \frac{d}{dp} p^n = np^n \quad (51)$$

e então,

$$\begin{aligned} \bar{n} &= \sum_{n=0}^N \frac{N!}{n!(N-n)!} \left(p \frac{\partial}{\partial p} p^n \right) q^{N-n} \\ &= p \frac{\partial}{\partial p} \left[\sum_{n=0}^N \frac{N!}{n!(N-n)!} p^n q^{N-n} \right] \\ &= p \frac{\partial}{\partial p} (p+q)^N \\ &= pN(p+q)^{N-1} \\ &= pN \quad (52) \end{aligned}$$

e, da mesma forma,

$$\overline{n'} = qN = (1-p)N \quad (53)$$

Para determinar o desvio padrão, procedemos da mesma forma.

$$\begin{aligned}
\overline{n^2} &= \sum_{n=0}^N n^2 \frac{N!}{n!(N-n)!} p^n q^{N-n} \\
&= \sum_{n=0}^N \frac{N!}{n!(N-n)!} \left(p \frac{\partial}{\partial p} \right)^2 p^n q^{N-n} \\
&= \left(p \frac{\partial}{\partial p} \right)^2 \left[\sum_{n=0}^N \frac{N!}{n!(N-n)!} p^n q^{N-n} \right] \\
&= \left(p \frac{\partial}{\partial p} \right)^2 (p+q)^N \\
&= p \frac{\partial}{\partial p} [pN(p+q)^{N-1}] \\
&= p [N(p+q)^{N-1} + pN(N-1)(p+q)^{N-2}]
\end{aligned} \quad (54)$$

e, com $(p+q) = 1$,

$$\begin{aligned}
\overline{n^2} &= p [N + pN(N-1)] \\
&= p [pN^2 + N(1-p)] \\
&= (pN)^2 + p(1-p)N \\
&= \overline{n^2} + pqN
\end{aligned} \quad (55)$$

A variância é,

$$\sigma_n^2 = \overline{(\Delta n)^2} = \overline{n^2} - \overline{n}^2 = pqN \quad (56)$$

e o desvio padrão é,

$$\sigma_n = \sqrt{pq}\sqrt{N} \quad (57)$$

que é o resultado conhecido, proporcional a \sqrt{N} . A probabilidade relativa, σ_n/\bar{n} é,

$$\frac{\sigma_n}{\bar{n}} = \frac{\sqrt{pqN}}{pN} = \left(\frac{q}{p}\right)^{1/2} \frac{1}{\sqrt{N}} \quad (58)$$

A distribuição binomial torna-se muito fina a medida que N cresce, centrando em torno do valor médio \bar{n} .

Podemos comparar os resultados das eqs. 52, 57 e 58 com os resultados da figura 3. Vemos que, efetivamente, \bar{n} aumenta com N enquanto que σ_n aumenta com \sqrt{N} . Por outro lado, σ_n/\bar{n} diminui com $1/\sqrt{N}$. Da mesma forma, vemos que \bar{n} aumenta com p e $\sigma_n = 0,5\sqrt{N}$ para $p = 0,5$ e $\sigma_n = 0,458\sqrt{N}$ para $p = 0,7$. Da mesma forma, $\sigma_n/\bar{n} = 1/\sqrt{N}$ ($p = 0,5$) e $= 0,655/\sqrt{N}$ ($p = 0,7$). A diminuição das larguras de σ_n e σ_n/\bar{n} de $p = 0,7$ quando comparado com $p = 0,5$ é visível na figura 3.

De volta ao “caminhar do bêbado”. Vamos retornar a um dos nossos problemas. As perguntas que gostaríamos de responder são do tipo “qual a distância média percorrida pelo bêbado após um número N de passos?”, ou “qual a probabilidade de, após executar N passos o bêbado encontrar-se a uma distância igual a x da origem?”. Como conhecemos agora a distribuição de probabilidades, podemos responder facilmente a essas questões (*Exercício: responda!*).

Vamos utilizar a simulação *RandomWalk1D* da ref. 5 para exercitarmos um pouco os resultados desse problema. Essa simulação é baseada no *Método Monte Carlo*. A cada passo, o caminhante dá um passo à direita com probabilidade p ou para a esquerda com probabilidade q . Cada passo em a mesma distância e não depende dos passos anteriores. Cada passo é decidido baseado na geração de um número aleatório entre 0 e 1. Se o número aleatório for inferior a p então o passo é à direita, se for superior então é à esquerda, e assim sucessivamente. A simulação realiza vários caminhos aleatórios de N passos e calcula a

distância final percorrida $x = n - n'$ e o valor de $\langle x^2 \rangle$. A figura 4 mostra alguns exemplos.

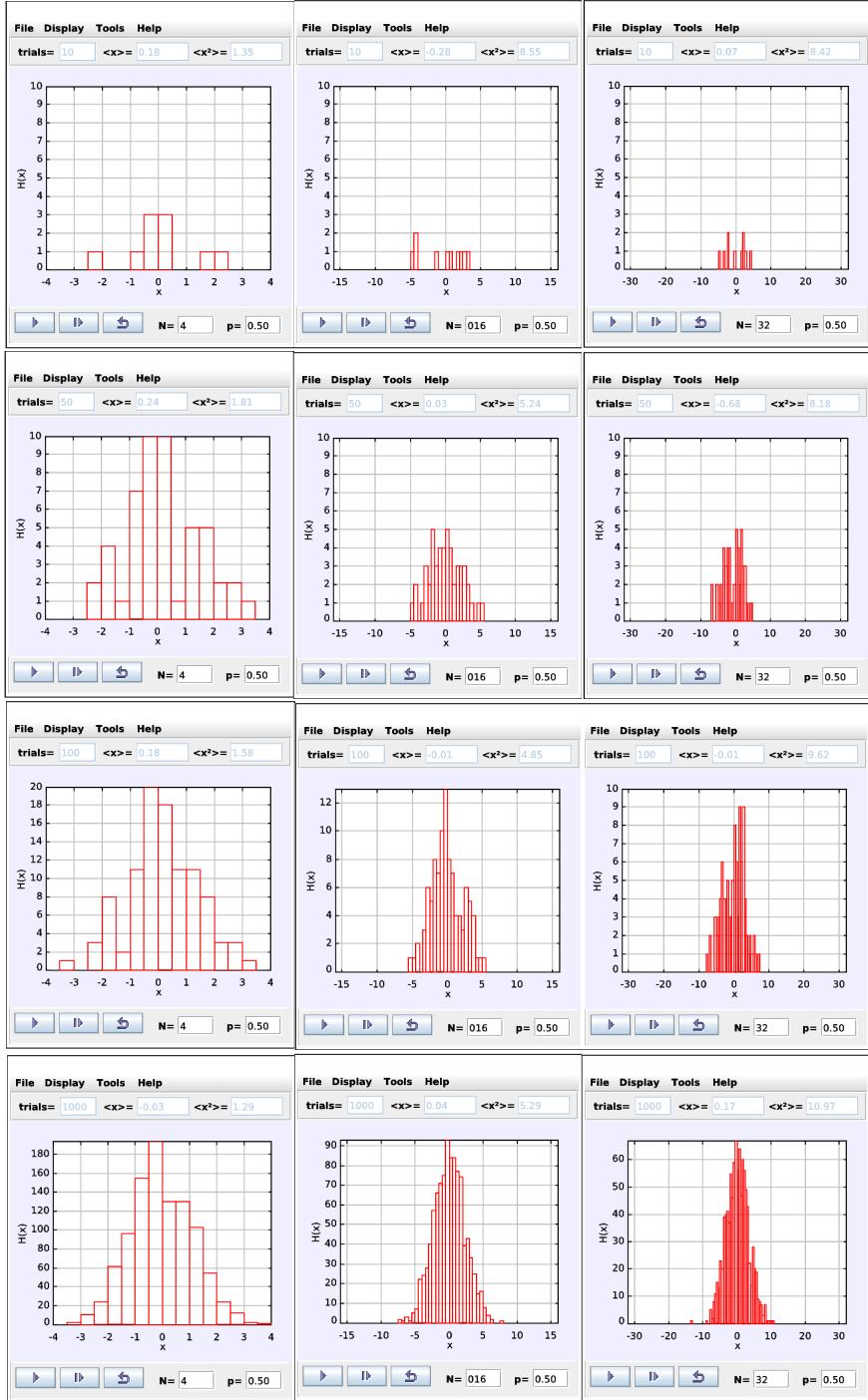


Figura 4: Histogramas obtidos das simulações do “caminho de bêbado” baseadas no programa *RandomWalk1D* (ref.) para $N = 4, 16, 32$ (da esquerda para direita) com $10, 50, 100, 1000$ (de cima para baixo).

Exercício. Baseado nessa figura, discuta as seguintes questões:

- (a) Por quê realizar várias simulações? Qual o significado de uma única simulação? (Uma simulação representa uma “caminhada” com N passos)
 - (b) A simulação Monte Carlo reproduz a distribuição de probabilidade?
 - (c) Qual o valor mais provável de x em cada caso?
 - (d) Qual é aproximadamente a largura da distribuição (largura a meia-altura)? Como esse valor muda para p fixo em função de N ?
 - (e) Como o histograma muda em função do número de simulações/caminhadas para N fixo e em função de N ?
-

É importante sempre enfatizarmos o que estamos calculando e o que podemos calcular. Na simulação, realizamos para um determinado número de passos, N , um grande número de experimentos. Se realizássemos apenas um único “caminhar”, não temos como saber o resultado final, embora possamos calcular a probabilidade dos possíveis resultados. No entanto, podemos calcular o resultado “médio” de um grande número de experimentos. Quanto maior for o número de experimentos, mais nos aproximamos do valor médio, que é previsível se conhecemos a distribuição de probabilidades. O que é possível conhecer, portanto, são valores estatísticos, como a média, em relação a um grande número de experimentos ou conjunto de experimentos.

Nota avançada: caminho aleatório bidimensional

A figura mostra de um “caminho aleatório” (o “andar de bêbado”) em duas dimensões, extraído da ref. 9. Os caminhos aleatórios representam a evolução em função do número de passos, desde 31 passos até 32.000 passos. A forma de obter as diferentes trajetórias é tomar o primeiro quarto do caminho anterior e ampliar por um fator dois. Em todos os casos, observamos um caminho entrecortado. O objetivo dessa sequência é indicar uma auto-similaridade do sistema, isto é, os caminhos aleatórios parecem similares em todas as escalas,

ou seja são *invariante na escala* (comente!). A figura (esquerda) mostra o caminho aleatório após 128.000 passos. Novamente, o caminho parece completamente entrecortado. O ponto importante para nós é a impossibilidade de prevermos um caminho aleatório qualquer. No entanto, podemos entender (e fazer previsões sobre) o comportamento de um conjunto desses caminhos. A figura (direita) mostra os pontos finais de um grande número de caminhos aleatórios todos partindo da origem. Você pode exercitar sua compreensão sobre isso com a simulação *stp – RandomWalk2D* da ref. 5.

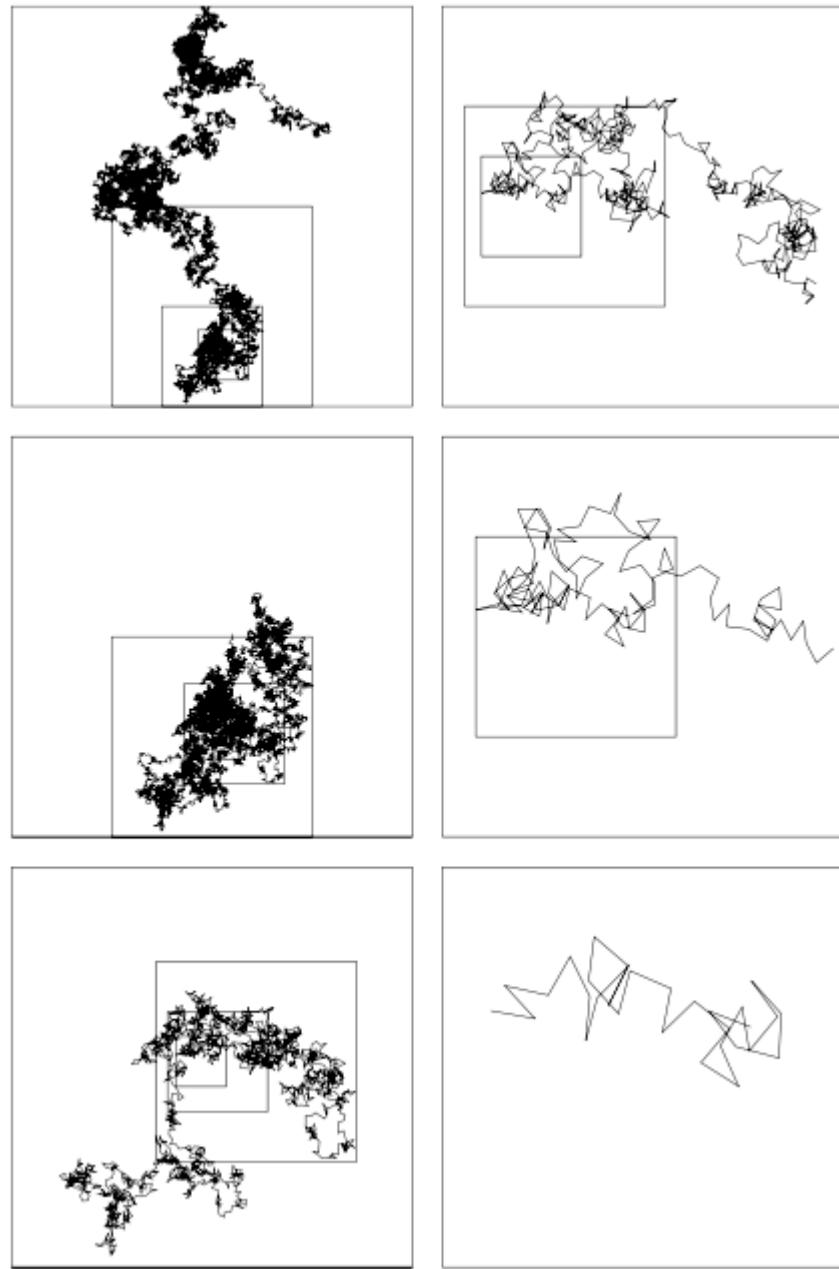


Figura 5: Caminho aleatório bidimensional. Cada caminho é obtido do anterior tomando-se o primeiro quarto e ampliando por um fator dois. O caminho mais curto tem 31 passos e o caminho mais longo 32.000 passos. (ref. 9)

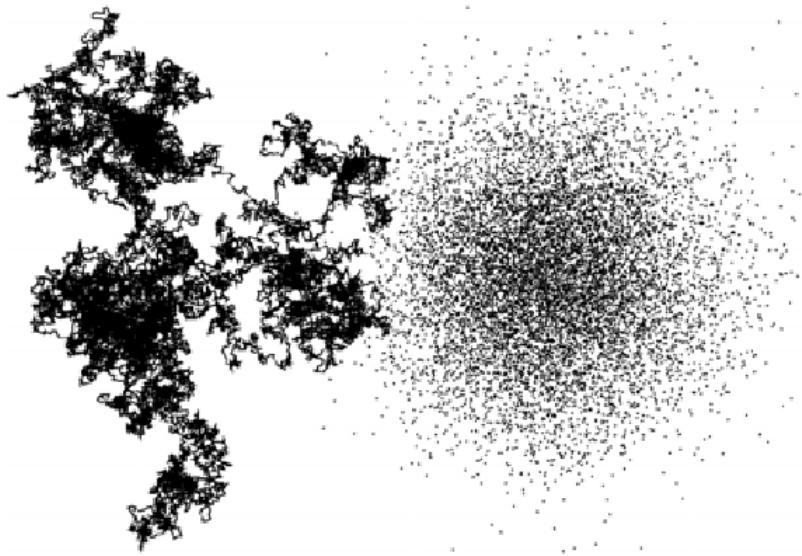


Figura 6: (Esquerda) Caminho aleatório bidimensional com 128.000 passos. (Direita) Distribuição do ponto de chegada a partir da origem para um grande número de caminhos aleatórios. (ref. 9)

Distribuição binomial para valores grandes de N A figura 3 mostra que a distribuição binomial adquire uma forma suave, quase contínua, a medida que N aumenta. Vamos chamar de \tilde{n} o valor máximo de $P_N(n)$. Vemos que P_N varia rapidamente em torno de \tilde{n} para $N \gg 1$. É mais fácil trabalhar com o $\ln P_N$ que deve ser uma função lentamente variável. Efetivamente, utilizando a simulação *stp-binomial*, podemos calcular $\ln P_N \times n$, que está representada na figura 7. É fácil visualizar que essa curva parece uma parábola. Podemos esperar poder aproxima-la por uma relação do tipo $\ln P_N = A + B(n - \bar{n})^2$, onde A e B são parâmetros de ajuste.

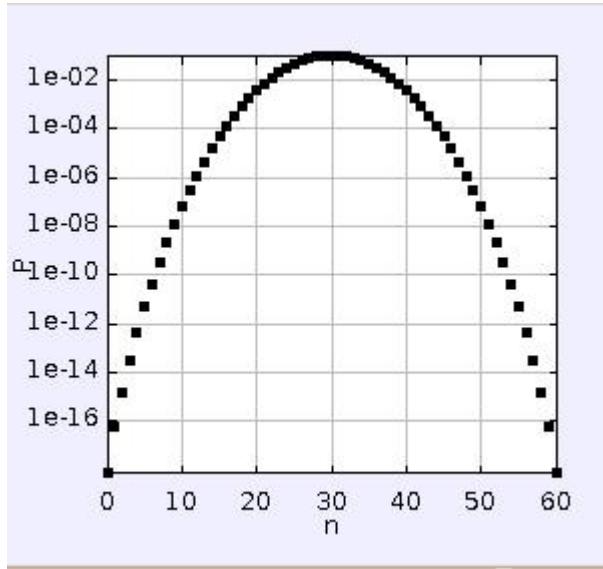


Figura 7: Gráfico de $\ln P_N$ em função de n para $N = 60$. Calculado com a simulação *stp-binomial* (ref. 5).

Para obter uma expressão quantitativa para $\ln P_N$, vamos expandi-la em série de Taylor, em torno de $n = \tilde{n}$:

$$\ln P_N(n) = \ln P_N(n = \tilde{n}) + (n - \tilde{n}) \frac{d \ln P_N(n)}{dn} \Big|_{n=\tilde{n}} + \frac{1}{2} (n - \tilde{n})^2 \frac{d^2 \ln P_N(n)}{dn^2} \Big|_{n=\tilde{n}} + \dots \quad (59)$$

Pelo resultado da figura 7, a primeira derivada deve se anular (ponto de extremo) e a segunda derivada deve ser negativa (valor máximo). Se desprezarmos as derivadas de ordem superior e escrevermos,

$$\begin{aligned} A &= \ln P_N(n = \tilde{n}) \\ B &= -\frac{d^2 \ln P_N(n)}{dn^2} \Big|_{n=\tilde{n}} \end{aligned} \quad (60)$$

podemos escrever,

$$\ln P_N(n) \approx A - \frac{1}{2}B(n - \tilde{n})^2 \quad (61)$$

e,

$$P_N(n) \approx Ae^{-\frac{1}{2}B(n - \tilde{n})^2} \quad (62)$$

Para avançarmos, temos que encontrar os valores das expressões em 60. Para isso, vamos utilizar a aproximação de Stirlig para $\ln N!$ para valores grandes de N :

$$\ln N! \approx N \ln N - N + \frac{1}{2} \ln(2\pi N) \text{ (aproximação de Stirling)} \quad (63)$$

Nota: o correto é chamarmos essa aproximação de *de Moivre-Stirling*. Abraham de Moivre (1617-1754), matemático francês chegou até a expressão $n! \approx c\sqrt{n}n^n/e^n$. A contribuição de James Stirling (matemático escocês, 1692-1770) foi determinar o valor da constante $c = 2\pi$. No entanto, a expressão 63 é conhecida simplesmente por aproximação de Stirling.

A eq. 63 pode ser aproximada, para valores muito grandes de N , na forma,

$$\ln N! \approx N \ln N - N \quad (64)$$

Demonstração da aproximação de Stirling:

Essa demonstração segue os passos da ref. 4. Inicialmente, vamos demonstrar que

$$n! = \int_0^\infty dx e^{-x} x^n \quad (65)$$

Para isso, vamos integrar por partes,

$$\begin{aligned}
\int_0^\infty dx e^{-x} x^n &= - \int_0^\infty d(e^{-x}) x^n = - [x^n e^{-x}] \Big|_{x=0}^{x=\infty} + n \int_0^\infty dx e^{-x} x^{n-1} \\
&= n \int_0^\infty dx e^{-x} x^{n-1}
\end{aligned} \tag{66}$$

e, por recorrência, temos a eq. 65.

Analizando a função do integrando,

$$f(x) = x^n e^{-x} \tag{67}$$

vemos que enquanto x^n é rapidamente crescente com x , e^{-x} é rapidamente decrescente. Logo, a função $f(x)$ tem um ponto de máximo em algum valor x_0 . Para encontrá-lo, vamos calcular a derivada do logarítmico da função f ,

$$\begin{aligned}
\frac{d [\ln f(x)]}{dx} &= 0 \\
\implies \frac{d}{dx} (n \ln x - x) &= \frac{n}{x} - 1 = 0
\end{aligned} \tag{68}$$

ou seja,

$$x_0 = n \tag{69}$$

Vamos expandir o logarítmico do integrando em torno do ponto de máximo, $x = n + \xi$,

$$\ln f_n(\xi) = n \ln(n + \xi) - (n + \xi) \tag{70}$$

Expandindo o termo logarítmico,

$$\begin{aligned}
\ln(n + \xi) &\approx \ln n + \ln \left(1 + \frac{\xi}{n}\right) \\
&= \ln n + \frac{\xi}{n} - \frac{1}{2} \frac{\xi^2}{n^2} + \dots
\end{aligned} \tag{71}$$

onde utilizamos o resultado,

$$\begin{aligned}
\ln x &= (x - 1) + \frac{1}{2}(x - 1)^2 + \frac{1}{3}(x - 1)^3 + \dots \\
&= \sum_{n=1}^{\infty} (-1)^{n-1} \frac{1}{n} (x - 1)^n \quad (\text{válido para } 0 < x \leq 2)
\end{aligned} \tag{72}$$

Temos então,

$$\ln f_n(\xi) \approx n \ln n - n - \frac{1}{2} \frac{\xi^2}{n} \tag{73}$$

e,

$$f_n(\xi) \approx n^n e^{-n} e^{-\xi^2/2n} \tag{74}$$

Podemos agora calcular aproximadamente a equação 65, substituindo $f_n(x)dx$ por $f_n(\xi)d\xi$ e adaptando os limites de integração, $x \in [0, \infty] \longrightarrow \xi \in [-n, \infty]$,

$$n! \simeq \int_{-n}^{\infty} d\xi n^n e^{-n} e^{-\xi^2/2n} \tag{75}$$

Como a função $f_n(\xi)$ é praticamente desprezível para $\xi < -n$, podemos substituir o limite inferior da integração por $-\infty$,

$$\begin{aligned}
n! &\simeq n^n e^{-n} \int_{-\infty}^{\infty} d\xi e^{-\xi^2/2n} \\
&= n^n e^{-n} \sqrt{2\pi n}
\end{aligned} \tag{76}$$

que é a aproximação de Stirling.

A aproximação de Stirling, na sua forma forte (eq. 63) tem uma precisão melhor que 2% para $N > 5$ e, na sua forma fraca (eq. 64) uma precisão melhor que 2% para $N > 50$.

Podemos agora calcular os coeficientes.

$$\ln P_N(n) = \ln N! - \ln n! - \ln(N-n)! + n \ln p + (N-n) \ln p \tag{77}$$

e,

$$\frac{d(\ln P_N(n))}{dn} = -\ln n + \ln(N-n) + \ln p - \ln q \tag{78}$$

O valor máximo de n (\tilde{n}) é encontrado com a condição $d\ln P_N(n)/dn = 0$. Então,

$$\frac{N - \tilde{n}}{\tilde{n}} = \frac{q}{p} \tag{79}$$

Utilizando a relação $(p + q) = 1$,

$$\begin{aligned}
(N - \tilde{n})p &= \tilde{n}q \\
\implies Np &= \tilde{n}(p + q) = \tilde{n}
\end{aligned} \tag{80}$$

que era o resultado esperado (por quê?), ou seja, $\tilde{n} = \bar{n}$, o valor de n para o qual $P_N(n)$ é o máximo é também o seu valor médio.

Para a segunda derivada temos,

$$\frac{d^2(\ln P_N(n))}{dn^2} = -\frac{1}{n} - \frac{1}{N-n} \quad (81)$$

e o coeficiente B é,

$$\begin{aligned} B &= -\frac{d^2(\ln P_N(n))}{dn^2}|_{n=\tilde{n}} = \frac{1}{\tilde{n}} + \frac{1}{N-\tilde{n}} = \frac{N}{\tilde{n}(N-\tilde{n})} \\ &= -\frac{N}{(Np)N(1-p)} = \frac{1}{Npq} \end{aligned} \quad (82)$$

e da eq. 56, temos

$$B = \frac{1}{\sigma_n^2} \quad (83)$$

Para calcularmos o valor de A , temos que utilizar a eq. 60 e a aproximação de Stirling para $n = \tilde{n}$ e usando $\tilde{n} = pN$,

$$\begin{aligned} \ln A &= \ln P_N(\tilde{n}) = \ln N! - \ln \tilde{n}! - \ln(N-\tilde{n})! + \tilde{n} \ln p + (N-\tilde{n}) \ln q \\ &= N \ln N - N + \frac{1}{2} \ln(2\pi N) - \tilde{n} \ln \tilde{n} + \tilde{n} - \frac{1}{2} \ln(2\pi \tilde{n}) \\ &\quad - (N-\tilde{n}) \ln(N-\tilde{n}) + (N-\tilde{n}) - \frac{1}{2} \ln[2\pi(N-\tilde{n})] + \tilde{n} \ln p + (N-\tilde{n}) \ln q \\ &= N \ln N + \frac{1}{2} \ln(2\pi N) - pN \ln(pN) - \frac{1}{2} \ln(2\pi pN) \\ &\quad - Nq \ln(qN) - \frac{1}{2} \ln(2\pi qN) + pN \ln p + Nq \ln q \\ &= N \ln N + \frac{1}{2} \ln(2\pi N) - pN \ln p - pN \ln N - \frac{1}{2} \ln(2\pi pN) \\ &\quad - qN \ln q - qN \ln N - \frac{1}{2} \ln(2\pi qN) + pN \ln p + Nq \ln q \\ &= \frac{1}{2} \ln \frac{2\pi N}{(2\pi)^2 pq N^2} = \ln \frac{1}{\sqrt{2\pi N pq}} \\ \implies A &= \frac{1}{\sqrt{2\pi N pq}} = \frac{1}{(2\pi \sigma^2)^{1/2}} \end{aligned} \quad (84)$$

A probabilidade da distribuição binomial identifica-se com a *distribuição Gaussiana de probabilidades*,

$$P_N(n) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-(n-\bar{n})^2/2\sigma^2} \quad (85)$$

É interessante observar que a aproximação “fraca” de Stirling, eq. 64, daria simplesmente $A = 0$. É necessário utilizar a aproximação em toda a sua expressão, i.e., eq. 63.

Exercício: Confira a normalização da equação 85, isto é, $\sum_{n=0}^N P_N(n) = 1$.

Vamos analisar a validade da aproximação gaussiana. Em princípio, ela só deveria ser válida para valores de N grandes e em torno de $n = \bar{n}$. No entanto, os resultados são bons mesmo para valores relativamente pequenos de N e para a maioria dos valores de n . A principal característica da probabilidade gaussiana está na dependência de sua largura relativa, σ_n/\bar{n} que decresce com $N^{-1/2}$, da mesma forma que a distribuição binomial. A distribuição gaussiana e a distribuição binomial possuem o mesmo valor médio e a mesma variância. Para verificarmos a validade da aproximação, vamos calcular a terceira derivada de $P_N(n)$:

$$\frac{d^3 P_N(n)}{dn^3} = \frac{1}{n^2} - \frac{1}{(N-n)^2} + 0\left(\frac{1}{n^3}\right) \quad (86)$$

e, em $n = \bar{n}$, para N grande,

$$\begin{aligned} \frac{d^3 P_N(n)}{dn^3} &= \frac{1}{p^2 N^2} - \frac{1}{q^2 N^2} = \frac{N^2(p+q)(q-p)}{p^2 q^2 N^4} \\ &= \frac{q-p}{p^2 q^2 N^2} \end{aligned} \quad (87)$$

A aproximação deve ser suficiente sempre que o termo de terceira ordem for muito inferior

ao termo de segunda ordem, ou seja,

$$\begin{aligned} \frac{1}{2Npq}(n - \bar{n})^2 &\gg \frac{|q - p|}{6N^2p^2q^2}(n - \bar{n})^3 \\ \implies (n - \bar{n}) &\ll \frac{3Npq}{|q - p|} \end{aligned} \quad (88)$$

que é a condição de validade da aproximação gaussiana.

A Tabela mostra uma comparação entre a distribuição binomial e a aproximação gaussiana para $P_{10}(n)$ e $p = q = 1/2$.

n	$P_{10}(n)$	Gaussian approximation
0	0.000977	0.001700
1	0.009766	0.010285
2	0.043945	0.041707
3	0.117188	0.113372
4	0.205078	0.206577
5	0.246094	0.252313

Tabela 1: Comparação entre os valores exatos, distribuição binomial, e a aproximação gaussiana para $P_{10}(n)$ para $p = q = 1/2$. Tabela extraída da ref. 5.

Exercício: O que acontece com a condição da eq. 88 quando $p = q = 1/2$? Em que etapa da derivação fizemos uma “inconsistência” nesse caso? Encontre a condição de validade da aproximação gaussiana nesse caso. Mostre, baseado nesse resultado e no da equação 88 que a aproximação gaussiana é justificável se $|n - \bar{n}| \ll Npq$. Explique porque parando em segunda ordem é justificável se $Npq \gg 1$.

3.5 Distribuições contínuas de probabilidade

Em muitos casos práticos, as variáveis aleatórias (estocásticas) são contínuas. O exemplo mais típico para nós, por exemplo, é a posição e velocidade das partículas de um gás. Na simulação que fizemos no capítulo 2 do gás interagente distribuindo-se em duas metades de uma caixa, utilizamos a posição em uma ou outra metade como a informação desejada e, dessa forma, discreta. No entanto, se estamos interessados em grandezas termodinâmicas, como a pressão (impossível na nossa simulação devido às condições de contorno periódicas) ou a temperatura, as variáveis microscópicas aleatórias deveriam ser a posição e velocidade de cada partícula (seis variáveis independentes). Dos três exemplos que consideramos para o desenvolvimento da distribuição binomial, o “caminhar do bêbado” na verdade é um caso de variável aleatória contínua. Mesmo com a simplificação unidimensional, fizemos uma outra simplificação mais drástica, que foi considerar os passos idênticos em tamanho. Isso permitiu tratar o problema discretamente e como uma distribuição binomial. Na realidade, devemos esperar que os passos sejam de comprimento variável e contínuos. (Os outros dois casos, momentos de dipolos magnéticos e jogar uma moeda, são difíceis de se imaginar como variáveis contínuas).

Vamos estender nossa análise da distribuição binomial para o caso das variáveis contínuas e utilizar o “caminhar do bêbado” como exemplo. Novamente, consideramos que ele tem uma probabilidade p de dar um passo à direita, mas agora a distância percorrida é um valor aleatório entre zero e uma distância máxima a . O deslocamento x é, portanto, uma variável contínua. Consideremos uma simulação do caminhar, similar a que fizemos na figura 4, mas agora com passos variáveis, anotando o número de vezes $H(x)$ que o caminhante desloca-se a partir da origem, após N passos, até uma distância com intervalo Δx , entre x e $x + \Delta x$. A figura 8 exemplifica esse caso para $\Delta x = 0,5$ e $N = 16$ e $\Delta x = 1,0$ e $N = 32$ e foi obtida a partir do programa *stp-RandomWalk1DContinuous* da ref. 5. O valor do histograma $H(x)$ é proporcional a probabilidade estimada que o caminhante encontra-se no intervalo Δx a distância x da origem após N passos. Para obter a probabilidade, dividimos

$H(x)$ pelo número de passos N . Aqui é importante observar que a escolha do intervalo Δx parece arbitrária mas possui algumas restrições. O valor não pode ser muito pequeno porque nesse caso, para um número finito de caminhadas, a posição pode se encontrar “vazia”, isto é, nenhuma das caminhadas finalizar naquele intervalo e a estimativa do número de caminhadas em cada intervalo ficaria imprecisa. Por outro lado, não pode ser muito grande, caso contrário as características do histograma ficam perdidas. Como devemos esperar que o número de caminhadas em um certo intervalo é proporcional ao intervalo, escrevemos,

$$p(x)\Delta x = \frac{H(x)}{N} \quad (89)$$

e a grandeza $p(x)$ é conhecida como a *densidade de probabilidade*.

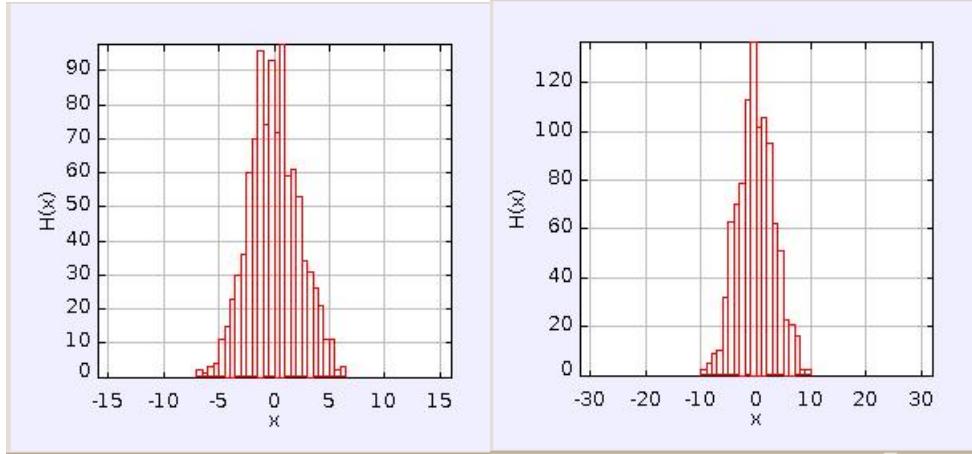


Figura 8: Histograma $H(x)$ do número de vezes que o deslocamento de um caminhante aleatório tem seu deslocamento, após $N = 16$ (esquerda) e $N = 32$ (direita) passos, entre x e $x + \Delta x$, onde $\Delta = 0,5$ e $\Delta = 1,0$, respectivamente. A distância de cada passo é escolhida aleatoriamente entre zero e um. Cada histograma foi gerado com 1000 caminhadas, utilizando a simulação *stp-RandomWalk1DContinuous* da ref. 5. Os valores médios estatísticos foram $\bar{x} = 0,00$ (isto é, $\bar{x} \lesssim 0(10^{-3})$) e $\bar{x^2} = 5,18$ para $N = 16$ e $\bar{x} = 0,02$ e $\bar{x^2} = 10,7$ para $N = 32$.

No limite $\Delta x \rightarrow 0$, $H(x)$ torna-se uma função contínua de x e a probabilidade do deslocamento x do caminhante encontrar-se entre a e b (ver figura), é,

$$P(a < x < b) = \int_a^b dx p(x) \quad (90)$$

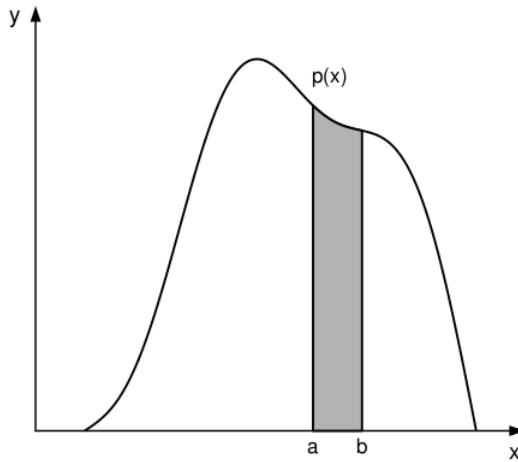


Figura 9: A Probabilidade que x encontre-se entre a e b é determinada pela área da região sombreada. (extraído da ref. 5)

A densidade de probabilidade $p(x)$ é não-negativa e tem unidades do inverso da dimensão da distância.

As propriedades da densidade de probabilidade podem ser obtidas generalizando o caso discreto. A normalização se escreve na forma,

$$\int_{-\infty}^{\infty} dx p(x) = 1 \quad (91)$$

O valor médio da função $f(x)$ é,

$$\bar{f} = \int_{-\infty}^{\infty} dx f(x)p(x) \quad (92)$$

e assim por diante.

3.6 Teorema do Limite Central

Até agora, discutimos as estimativas das probabilidades de uma amostragem de resultados sem entrar em maiores detalhes, isto é, empíricamente. Nas figuras 4 e 8 mostramos histogramas de um conjunto de experimentos (no caso, caminhadas aleatórias) e a altura do histograma foi identificada com a probabilidade de ocorrer aquele resultado (a menos de fa-

tores). Esperamos que quanto maior o número de experimentos mais próximo será a média obtida do resultado exato. Essa ideia é o que conhecemos como *lei dos grandes números*. Sabemos que um resultado particular, no entanto, pode diferir desse valor médio. Situação similar já havia sido discutida por ocasião das simulações das partículas na caixa com divisória (capítulo 2). Naquele caso, observamos a evolução temporal do sistema e sua tendência a convergir para uma situação de equilíbrio (a menos de flutuações). Quando executamos uma medida experimental, a medida é realizada durante um certo intervalo de tempo e o sistema evolue rapidamente para diferentes configurações nesse tempo. Como consequência, o resultado da medida é uma média sobre essas várias configurações. Utilizamos o princípio ergódico para identificar uma média temporal sobre um tempo suficientemente longo com a média sobre o conjunto (*ensemble*) de configurações microscópicas do sistema. Isso equivale a fazermos uma média sobre vários caminhos aleatórios, como comentamos acima. Até aqui, esse resultado tem sido adotado de forma empírica. No entanto, podemos avançar e encontrar a forma da distribuição de probabilidades que mede o quanto uma medida específica (um caminho executado pelo caminhante ou uma certa medida de um gás) difere do seu valor médio exato. A forma dessa distribuição de probabilidades é o *teorema do limite central*. Para simplificar a discussão, vamos considerar um exemplo simples e depois desenvolveremos para o caso do caminhante aleatório.

Nosso objetivo é estimar a probabilidade de, ao jogar um dado, encontrar, por exemplo, a face com o número 5. Sabemos que o resultado $1/6$ significa que, se realizarmos o experimento N vezes, deveremos ter, *aproximadamente*, $N/6$ vezes o número 5. Nossa questão é determinar o significado desse *aproximadamente*. Seja S número total de vezes que o 5 aparece em N experimentos. Temos,

$$S = \sum_{i=1}^N s_i \tag{93}$$

onde,

$$s_i = \begin{cases} 1 \text{ se a } i\text{-ésima jogada der 5} \\ 0 \text{ em qualquer outra situação} \end{cases} \quad (94)$$

Vimos nas simulações que a razão S/N aproxima-se de $1/6$ para valores grandes de N , como esperaríamos intuitivamente. A questão que nos colocamos é *qual a taxa na qual esse limite se aproxima?* A forma de respondermos essa pergunta, experimentalmente, é medirmos M vezes (como fizemos nos histograma). *Atenção:* cada medida significa realizar N jogadas do dado. Como S é uma variável aleatória em si mesma, sabemos que as medidas de S não serão idênticas. A figura 10 mostra os resultados para $M = 10.000$ medidas de S para $N = 100$ e $N = 800$. Vemos que a forma aproximada da distribuição dos valores de S é gaussiana. (Podemos fazer a mesma afirmação em relação a figura 4?) (Qual o valor relativo da largura da curva em cada caso?).

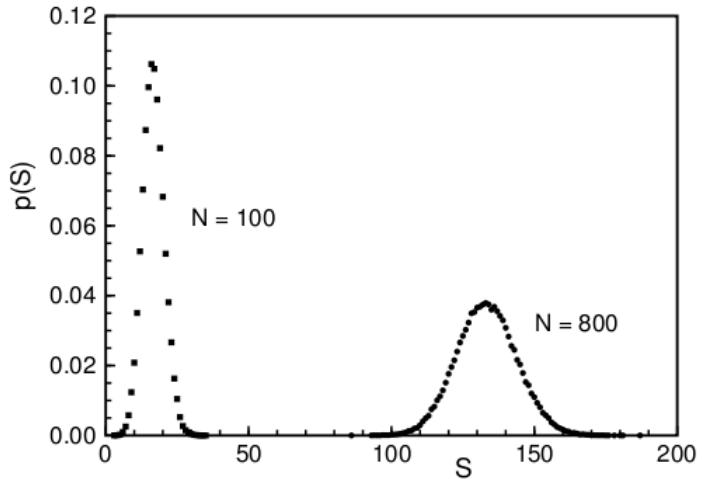


Figura 10: Distribuição de $M = 10.000$ medidas diferentes da soma S para $N = 100$ e $N = 800$ termos na soma. A grandeza S é o número de vezes que a face 5 aparece em N jogadas de um dado. Para $N = 100$, os valores medidos são $\bar{S} = 16,67$, $\bar{S}^2 = 291,96$ e $\sigma_S = 3,74$. Para $N = 800$, temos $\bar{S} = 133,31$, $\bar{S}^2 = 17881,2$ e $\sigma_S = 10,52$. (extraído da ref. 5).

Podemos encontrar uma expressão para $p(S)$, isto é, a densidade de probabilidade de medirmos o valor S após N jogadas do dado para obtermos um resultado específico em cada jogada? Ou seja, é possível demonstrar que $p(S)$ é uma distribuição gaussiana? A resposta

é positiva e uma demonstração pode ser encontrada na seção 3.11.2 da ref. 5. Vamos apenas discutir aqui o resultado e detalhar para o caso específico do caminho aleatório. Como dissemos, pode-se mostrar que, *no limite* $N \rightarrow \infty$, a densidade de probabilidade $p(S)$ é,

$$p(S) = \frac{1}{\sqrt{2\pi\sigma_S^2}} e^{-(S-\bar{S})^2/2\sigma_S^2} \quad (95)$$

onde,

$$\bar{S} = N\bar{s} \quad (96)$$

$$\sigma_S^2 = N\sigma^2 \quad (97)$$

onde,

$$\sigma^2 = \bar{s}^2 - \bar{s}^2 \quad (98)$$

A grandeza $p(S)\Delta S$ é a probabilidade que o valor da soma $\sum_{i=1}^N s_i$ esteja entre S e $S + \Delta S$. A equação 95 é o *teorema do limite central*. A forma gaussiana vale apenas para valores grandes de N e para valores de S próximos do valor mais provável (valor médio). A origem do nome é significativa: *central* indica a validade para valores próximos da média enquanto que *limite* indica a validade para N muito grande.

O teorema do limite central é um dos resultados mais importantes da teoria de probabilidades. Ele estabelece, na sua forma mais simples, que a distribuição de probabilidades do valor de uma soma de um grande número de variáveis aleatórias é aproximadamente gaussiana. A aproximação melhor quanto maior for o número de variáveis na soma. É esse teorema que nos permite que o limite termodinâmico, clássico e macroscópico, faça sentido.

Vamos voltar agora ao caso do “caminhar do bêbado” ou o caminho aleatório. S representa o deslocamento do caminhante depois de N passos. σ^2 é o desvio padrão do deslocamento de

um único passo. No teorema do limite central, encontramos que a densidade de probabilidade do deslocamento ser uma gaussiana (ver figura 7 para o caso bidimensional). Esse resultado é equivalente ao que encontramos para caminhos aleatórios no limite de um número grande de passos N . O mesmo raciocínio pode ser feito para o caso de jogarmos moedas ou dos momentos magnéticos. O deslocamento do caminho aleatório depois de N passos e a magnetização de um sistema de N spins são exemplos de *processos aleatórios aditivos*. A distribuição dos caminhos aleatórios, dos spins ou das múltiplas jogadas de moeda é dada pela eq. 95, e o nosso trabalho reduz-se a encontrar as expressões para \bar{s} e σ^2 , para o problema de interesse, o que determina completamente a eq. 95.

Vamos desenvolver o caso particular do caminho aleatório (ver ref. 1). Novamente, consideramos uma sequência de N passos, assumindo que o deslocamento do i -ésimo passo tenha o comprimento aleatório s_i o qual ocorre com a probabilidade $\omega(s_i)ds_i$. Para uma sequência de passos N , a posição do caminhante é,

$$x = \sum_{i=1}^N s_i$$

A probabilidade de encontrar o caminhante entre x e $x + dx$ é dada por

$$p_N(x)dx = \int \dots \int_{x < s_1 + s_2 + \dots + s_N < x + dx} \omega(s_1)ds_1 \dots \omega(s_N)ds_N \quad (99)$$

onde as integrações são realizadas entre $(-\infty, +\infty)$ com as restrições explicitadas. podemos remover a restrição substituindo-a por funções δ de Dirac:

$$\delta(x) = \begin{cases} 1/\gamma \text{ para } -\gamma/2 < x < \gamma/2, \\ 0 \text{ para } |x| > \gamma/2 \end{cases} \quad (100)$$

e $\gamma \rightarrow 0$. Temos então,

$$p_N(x) = \int_{-\infty}^{+\infty} \dots \int_{-\infty}^{+\infty} \omega(s_1)ds_1 \dots \omega(s_N)ds_N \delta \left(x - \sum_{i=1}^N s_i \right) \quad (101)$$

Utilizando a representação integral da função δ ,

$$\delta(x) = \frac{1}{2\pi} \int_{-\infty}^{+\infty} dk e^{-ikx} \quad (102)$$

temos,

$$p_N(x) = \frac{1}{2\pi} \int_{-\infty}^{+\infty} dk e^{-ikx} [\hat{\omega}(k)]^N \quad (103)$$

onde,

$$\hat{\omega}(k) = \int_{-\infty}^{+\infty} ds \omega(s) e^{ikx} \quad (104)$$

Vamos obter agora $p_N(x)$ no limite de N muito grande. Como temos um fator oscilante, e^{+iks} , $\hat{\omega}(k)$ só é significativa na vizinhança de $k = 0$. Essa condição fica reforçada para $[\hat{\omega}(k)]^N$, com N grande. Podemos aproximar $\hat{\omega}(k)$ pela expansão,

$$\begin{aligned} \hat{\omega}(k) \int_{-\infty}^{+\infty} ds \omega(s) e^{ikx} &\approx \int_{-\infty}^{+\infty} ds \left\{ 1 + iks - \frac{1}{2} k^2 s^2 + \dots \right\} \\ &= 1 + ik\bar{s} - \frac{1}{2} k^2 \bar{s}^2 + \dots \end{aligned} \quad (105)$$

E temos então,

$$\begin{aligned} [\hat{\omega}(k)]^N &= \exp [N \ln \hat{\omega}(k)] \\ &= \exp \left\{ N \left[ik\bar{s} + \frac{1}{2} \left(k^2 \bar{s}^2 - \bar{s}^2 \right) + O(k^3) \right] \right\} \end{aligned} \quad (106)$$

Desprezando os termos $k^2 \bar{s}^2 - \bar{s}^2$ de ordem superior a k^2 , temos,

$$p(x) = \frac{1}{2\pi} \int_{-\infty}^{+\infty} dk \exp \left[-ikx + N k \bar{s} - \frac{1}{2} N \overline{(\Delta s)^2} k^2 \right] \quad (107)$$

que pode ser escrita na forma,

$$p(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp \left[-\frac{(x - \mu)^2}{2\sigma^2} \right] \quad (108)$$

onde,

$$\mu = N \bar{s} \quad (109)$$

e,

$$\sigma^2 = N \overline{(\Delta s)^2} \quad (110)$$

e recuperamos, para o caso do caminhante aleatório o resultado do *teorema do limite central* (eqs. 95-97).

Exercício: Verifique que podemos recuperar a distribuição binomial para um caminho aleatório em uma dimensão com passos de mesmo comprimento l , onde $\omega(s)$ é dada por,

$$\omega(s) = p\delta(s - l) + q\delta(s + l) \quad (111)$$

Podemos deduzir $p_N(x)$ a partir da generalização para passos variáveis e contínuos da equação estocástica de diferenças para caminhos aleatórios com passos iguais que utilizamos anteriormente para a distribuição binomial (eq. 39). A equação generalizada de recorrência se escreve na forma,

$$p_N(x) = \int_{-\infty}^{+\infty} ds p_{N-1}(x - s) \omega(s) \quad (112)$$

Introduzindo a transformada de Fourier,

$$\hat{p}_N(k) = \int_{-\infty}^{+\infty} dx p_N(x) e^{ikx} \quad (113)$$

e com $\hat{\omega}(k)$ dada pela eq. 104, temos,

$$\hat{p}_{N+1}(k) = \hat{p}_N(k) \hat{\omega}(k) \quad (114)$$

Como no instante inicial (i.e., para $N = 0$) o caminhante está na origem, ou seja, $p_0(x) = \delta(x)$,

3.7 Equação de difusão

No limite contínuo, podemos recuperar resultados macroscópicos considerando escalas de distância e tempo suficientemente longas. Nesse caso, um comportamento relativamente simples *emerge* a partir do movimento caótico dos estados microscópicos, isto é, do movimento das partículas ou do caminho errático do caminhante aleatório. Vamos derivar um caso simples, a partir dos resultados que trabalhamos até aqui, que é a *equação de difusão*. Partimos do limite contínuo de um conjunto (*ensemble*) de caminhadas aleatórias. Consideremos um caso geral, de caminhos aleatórios não-correlacionados - o que fizemos até agora - onde para cada passo temporal Δt a partícula na posição x muda de posição por um passo s :

$$x(t + \Delta t) = x(t) + s(t) \quad (115)$$

Vamos designar a distribuição de probabilidade para cada passo por $\omega(s)$ (ver eq. 111). Assumimos portanto que $\omega(s)$ tem média nula e desvio padrão σ . Os primeiros momentos de $\omega(s)$ são:

$$\begin{aligned}
\int ds \omega(s) &= 1 \\
\int ds \omega(s)s &= 0 \\
\int ds \omega(s)s^2 &= \sigma^2
\end{aligned} \tag{116}$$

Nossa tarefa é encontrar a distribuição de probabilidade $\rho(x, t + \Delta t)$ no próximo passo temporal uma vez conhecida a distribuição de probabilidade $\rho(x', t)$.

Para que a partícula mova-se de x' no tempo t para a posição x no tempo $t + \Delta t$, o passo $s(t)$ deve ser igual a $x - x'$. Isso acontece com a probabilidade $\omega(x - x')$ vezes a densidade de probabilidade $\rho(x', t)$ que ela tenha iniciado em x' . Integrando sobre as posições x' , temos,

$$\begin{aligned}
\rho(x, t + \Delta t) &= \int_{-\infty}^{+\infty} dx' \rho(x', t) \omega(x - x') \\
&= \int_{-\infty}^{+\infty} ds \rho(x - s, t) \omega(s)
\end{aligned} \tag{117}$$

Vamos assumir agora que ρ seja uma função bastante extendida no espaço, de tal forma que o tamanho do passo seja pequeno na escala na qual ρ varia. Podemos então fazer uma expansão em série de Taylor,

$$\begin{aligned}
\rho(x, t + \Delta t) &\approx \int ds \left[\rho(x, t) - s \frac{\partial \rho}{\partial x} + \frac{s^2}{2} \frac{\partial^2 \rho}{\partial x^2} \right] \omega(s) \\
&= \rho(x, t) \int ds \omega(s) - \frac{\partial \rho}{\partial x} \int ds s \omega(s) + \frac{1}{2} \frac{\partial^2 \rho}{\partial x^2} \int ds s^2 \omega(s) \\
&= \rho(x, t) + \frac{1}{2} \frac{\partial^2 \rho}{\partial x^2} \sigma^2
\end{aligned} \tag{118}$$

Assumindo também que ρ varie lentamente no intervalo do passo temporal, Δt , podemos aproximar,

$$\rho(x, t + \Delta t) - \rho(x, t) \approx \frac{\partial \rho}{\partial t} \Delta t \quad (119)$$

e então,

$$\frac{\partial \rho}{\partial t} = D \frac{\partial^2 \rho}{\partial x^2} \quad (120)$$

onde escrevemos,

$$D = \frac{\sigma^2}{2\Delta t} \equiv \text{coeficiente de difusão} \quad (121)$$

A equação de difusão aplica-se, para o comportamento macroscópico, de qualquer sistema físico semelhante ao caminho aleatório, guardada as condições que a função de distribuição de probabilidades varie lentamente na escala dos passos espacial e temporal.

Essa equação tem grande aplicação. Para o caso de partículas, a densidade de probabilidade descreve uma partícula individual a medida que ela evolue no espaço de forma aleatória. Para partículas não-interagentes, a probabilidade de distribuição de uma partícula descreve a densidade de todas as partículas.

Referências

- [1] Sílvio R. A. Salinas, **Introdução à Física Estatística**, EdUSP, 1997.
- [2] Federik Reif, **Fundamentals of Statistical and Thermal Physics**, McGraw-Hill, 1965.
- [3] Federik Reif, **Física Estadística**, Berkeley Physics Course vol. 5, Editorial Reverté.
- [4] Sylvie Vauclair, **Éléments de physique statistique: Hasard, organisation, évolution**, InterEditions, 1993.
- [5] Harvey Gould e Jan Tobochnik, **Statistical and Thermal Physics**, Princeton University Press, 2010 e <http://www.compadre.org/stp> (Statistical and Thermodynamic Project, apoiado pela National Science Foundations – EUA).
- [6] Paul L. Meyer, **Probabilidade: Aplicações a Estatística**, 2a. ed., LTC, 2000.
- [7] Arieh Ben-Naim, **Statistical Thermodynamics Based on Information: A Farewell to Entropy**, World Scientific, 2008.
- [8] Elliot W. Montroll e Michael F. Shlesinger, “On the Wonderful World of Random Walks”, Capítulo 1 em **Studies in Statistical Mechanics vol. XI Nonequilibrium Phenomena II From Stochastics to Hydrodynamics**, Eds. E. W. Montroll e J.L. Lebowitz, North-Holland Physics Publishing, 1984.
- [9] James P. Sethna, **Statistical Mechanics: Entropy, Order Parameters, and Complexity**, Oxford Master Series, 2006.