

Distribuições de probabilidade

Probability distributions

Marcel Novaes^{*1}

¹Universidade Federal de Uberlândia, Instituto de Física, 38408-100, Uberlândia, MG, Brasil.

Recebido em 08 de dezembro de 2021. Revisado em 25 de dezembro de 2021. Aceito em 27 de dezembro de 2021.

Apresentamos de forma pedagógica o tema das distribuições de probabilidade, que é central para o entendimento da mecânica estatística e da mecânica quântica, além de ser importante para todo tipo de descrição estatística dos fenômenos naturais.

Palavras-chave: Distribuições de probabilidade, Entropia, Momentos e cumulantes.

We present a pedagogical treatment of probability distributions, which is central for the understanding of statistical mechanics and quantum mechanics, apart from being importante for every kind of statistical description of natural phenomena.

Keywords: Probability distributions, Entropy, Moments and cumulants.

1. Introdução

O uso de variáveis aleatórias em física foi uma das grandes ideias da segunda metade do século XIX, quando Boltzmann, Maxwell e Gibbs desenvolveram a mecânica estatística e abriram caminho para um entendimento microscópico das leis da termodinâmica. A ideia fundamental por trás desse desenvolvimento é a de que um sistema muito complicado e imprevisível pode, mesmo estando submetido a leis determinísticas, exibir comportamento aparentemente aleatório e pode ser descrito de forma eficiente como se fosse realmente aleatório.

O conceito central dessa abordagem é o de distribuição de probabilidade. Esse conceito é indispensável para o entendimento tanto da mecânica clássica quanto da mecânica quântica. No primeiro caso, porque sistemas caóticos ou com muitos graus de liberdade se prestam muito bem a uma descrição probabilística; no segundo caso porque o módulo quadrado da função de onda é interpretado justamente como uma distribuição de probabilidade.

Apresentamos neste artigo um tratamento pedagógico das distribuições de probabilidade mais importantes, suas propriedades matemáticas e suas aplicações em problemas físicos.

2. Variáveis Aleatórias Discretas

Digamos que uma certa quantidade x assume valores a partir de um conjunto finito, $\{x_1, \dots, x_N\}$, e que a probabilidade de ocorrência do valor x_n é dada por $p(x_n)$. A probabilidade pode ser conhecida de antemão,

porque o processo de geração da variável aleatória é bem conhecido, ou pode ter de ser estimado a partir das frequências dos resultados em um número grande de realizações.

O exemplo mais simples é a distribuição uniforme, em que todos os valores são igualmente prováveis. É o caso de um dado ou de uma moeda, se acreditarmos que não estão viciados, ou dos números da Mega Sena (na prática, até novembro de 2021 a dezena mais sorteada, 53, apareceu 278 vezes, enquanto que a dezena menos sorteada, 26, apareceu apenas 203 vezes [1]).

Outra distribuição famosa é a de Zipf,

$$p_{\text{Zipf}}^{\alpha, N}(n) \propto \frac{1}{n^\alpha}, \quad (1)$$

que descreve a população dos países no mundo, ou a população das cidades de um país, ou a frequência de ocorrência das palavras de um idioma [2, 3]. A quantidade $p_{\text{Zipf}}^{\alpha, N}(n)$ é a resposta para as perguntas “quantos dentre os N brasileiros moram na n -ésima maior cidade do Brasil?” ou “qual a probabilidade de ocorrência da n -ésima palavra mais frequente em um texto de N palavras?”; o parâmetro α deve ser determinado empiricamente.

A distribuição de Poisson

$$p_{\text{Poisson}}^\lambda(n) = e^{-\lambda} \frac{\lambda^n}{n!}, \quad (2)$$

é útil na descrição do número de ocorrências independentes de um mesmo evento em um dado intervalo de tempo. Por exemplo, a quantidade $p_{\text{Poisson}}^\lambda(n)$ é a probabilidade de que aconteçam n decaimentos de uma amostra de elemento radioativo ao longo de um minuto, sendo λ o número médio de decaimentos por minuto.

* Endereço de correspondência: marcel.novaes@gmail.com

Se um evento tem probabilidade p de ocorrer e $1 - p$ de não ocorrer, então a probabilidade de que ele ocorra exatamente n vezes em N testes é dada por

$$p_{\text{Binomial}}^{N,p}(n) = \binom{N}{n} p^n (1-p)^{N-n}, \quad (3)$$

que é chamada de distribuição binomial.

Como mencionado na Introdução, a um conjunto de probabilidades $\{p(x_1), \dots, p(x_N)\}$ podemos associar uma entropia, que de forma adimensional é dada por

$$S = - \sum_n p(x_n) \log(p(x_n)). \quad (4)$$

De certa forma, a entropia mede a falta de informação a respeito do estado do sistema. Se soubermos que ele está no estado m , por exemplo, teremos $p_n(x_m) = \delta_{nm}$ e $S = 0$.

Se todas as probabilidades forem iguais, $p(x_n) = 1/N$, então $S = \log(N)$. No ensemble microcanônico da mecânica estatística, a entropia de Boltzmann é justamente dada por $S(E) = k \log(W)$, onde k é a constante de Boltzmann e W é o número de microestados com energia E . Já no ensemble canônico, a probabilidade depende da temperatura,

$$p(E_n) = \frac{e^{-E_n/kT}}{Z}, \quad (5)$$

sendo a função de partição dada por $Z = \sum_n e^{-E_n/kT}$, de modo que ficamos com

$$S = -k \sum_n p_n (-E_n/kT - \log(Z)) = \frac{\langle E \rangle}{T} + k \log(Z). \quad (6)$$

A identificação $Z = e^{-F/kT}$ entre a função de partição e a energia livre de Helmholtz leva à igualdade termodinâmica $F = \langle E \rangle - TS$.

3. Variáveis Aleatórias Contínuas

Seja x uma variável aleatória que assume valores contínuos. Dizemos que $\rho(x)$ é sua distribuição se a probabilidade de x estar no intervalo (a, b) é dada por $\int_a^b \rho(x) dx$. Toda distribuição satisfaz as propriedades de não-negatividade, $\rho(x) \geq 0$, e de normalização, $\int \rho(x) dx = 1$. Note que $\rho(x)$ não é uma probabilidade, e sim uma densidade de probabilidade (ou seja, é a sua integral que é uma probabilidade).

Na mecânica quântica, a função de onda $\Psi(x, t)$ satisfaz a equação de Schrödinger e portanto sua evolução temporal é determinística. Entretanto, seu módulo quadrado fornece a distribuição da variável aleatória que resulta do processo de medição da posição da partícula em questão no instante t , ou seja, $|\Psi(x, t)|^2 = \rho_t(x)$.

Na mecânica clássica, o uso de distribuições de probabilidade está ligado ao fato de que nenhum processo de medida tem precisão infinita, de modo que é sempre

necessário associar uma incerteza à posição (ou ao momento, ou a ambos) de uma partícula, mesmo que não estejamos levando em conta efeitos quânticos. No caso de sistemas caóticos, essa incerteza cresce exponencialmente com o tempo.

Se tivermos certeza absoluta de que a variável x assume um certo valor, digamos x_0 , então sua distribuição é uma função delta, $\delta(x - x_0)$. Se consideramos altamente provável que x esteja próximo de x_0 , com uma incerteza bem pequena, então usaremos uma distribuição próxima da delta; por exemplo, uma função que seja constante na vizinhança de x_0 e nula fora dessa vizinhança, ou algo mais suave como uma gaussiana.

Em analogia com a entropia de Gibbs, muitas vezes se define a “entropia” da distribuição $\rho(x)$ como sendo

$$S[\rho] = - \int \rho(x) \log[\rho(x)] dx, \quad (7)$$

e de fato a relação termodinâmica $F = \langle E \rangle - TS$ decorre da mesma forma que para variáveis discretas. Entretanto, essa não é uma boa definição de entropia. Basta notar que $\rho(x)$ pode ser maior do que 1, o que pode levar a uma “entropia” negativa. Ademais, essa entropia não tende a zero quando a incerteza na variável desaparece. Mesmo com esses pontos problemáticos, a quantidade acima costuma ser utilizada.

É importante saber mudar de variáveis. Se x possui distribuição de probabilidade $\rho(x)$ e temos uma variável que é função dela, $y = f(x)$, então quem será $\tau(y)$, a distribuição da variável y ? Para responder a essa questão, recorreremos à ideia de que a probabilidade deve se conservar. Supondo, por simplicidade, que $f(x)$ é injetora, esperamos intuitivamente que a chance de y ter o valor y_0 deve ser igual à chance de x ter o valor x_0 tal que $y_0 = y(x_0)$. Mais concretamente,

$$\int_a^b \rho(x) dx = \int_{y(a)}^{y(b)} \tau(y) dy$$

Se $b \rightarrow a$, temos

$$\tau(y) \lim_{b \rightarrow a} \frac{y(b) - y(a)}{b - a} = \tau(y) \frac{dy}{dx} = \rho(x) \quad (8)$$

e, portanto,

$$\tau(y) = \frac{\rho(f^{-1}(y))}{|df/dx|}. \quad (9)$$

Outra forma de proceder é usar a função delta de Dirac para escrever:

$$\tau(y) = \int \rho(x) \delta(y - f(x)) dx. \quad (10)$$

Note a presença de um módulo no denominador de (9), necessário para evitar a ocorrência de probabilidades negativas.

4. Variáveis Pseudo-Aleatórias

Com exceção daqueles que acontecem na escala quântica, os fenômenos naturais não são realmente aleatórios. Se jogarmos o nosso dado sempre da mesma maneira, exatamente da mesma maneira, sem variar o impulso, a altura da mão, a inclinação, o material da superfície onde ele vai cair, a densidade do ar, sem variar absolutamente nada, as leis de Newton garantem que o resultado vai ser sempre o mesmo. Na prática, isso é impossível, de modo que o resultado é pseudo-aleatório e usar uma abordagem probabilística faz sentido.

Em geral, comportamento pseudo-aleatório é comum em sistemas caóticos. Por exemplo, o famoso mapa logístico produz uma série de números entre 0 e 1 a partir da relação de recorrência $x_{n+1} = 4x_n(1 - x_n)$. Esses números são gerados a partir de um procedimento simples e determinístico, mas se distribuem erratically no intervalo (0,1) e são pseudo-aleatórios. A distribuição de probabilidade que descreve esse conjunto de pontos é $\frac{1}{\pi\sqrt{x(1-x)}}$, de tipo Beta. Um outro exemplo é o mapa

$$x_{n+1} = \begin{cases} \frac{2x}{1-x^2} & \text{for } 0 \leq x \leq \sqrt{2} - 1, \\ \frac{1-x^2}{2x} & \text{for } \sqrt{2} - 1 \leq x \leq 1, \end{cases} \tag{11}$$

que gera pontos distribuídos de acordo com a distribuição $\frac{4}{\pi(1+x^2)}$. Em geral é muito difícil determinar a distribuição de probabilidade associada a um mapa caótico [4]. A semelhança entre sistemas caóticos de poucos graus de liberdade e sistemas termodinâmicos é explorada, por exemplo, em [5].

A aplicação mais chocante de métodos probabilísticos a situações determinísticas talvez seja a que envolve os números primos. Por incrível que pareça, várias propriedades dos números primos sugerem que eles estão distribuídos, entre os inteiros, de maneira pseudo-aleatória. Essa hipótese, conhecida como modelo de Cramér, postula que a probabilidade de que um número n seja primo é aproximadamente $\frac{1}{\log n}$, o que funciona em algumas situações. Por exemplo, decorre daí que a quantidade de números primos menores que um valor qualquer $X \gg 2$ deve ser aproximadamente $\int_2^X \frac{dx}{\log x}$, o que está de fato correto.

Outro exemplo é teorema de Erdős-Kac [6], que diz que se $\omega(n)$ é o número de fatores primos distintos do número n , então para $n \gg 1$ a quantidade $(\omega(n) - \log \log n) / \sqrt{\log \log n}$ tem distribuição gaussiana com média 0 e variância 1. O comportamento dos ciclos de permutações de N números, com $N \gg 1$, também têm características pseudo-aleatórias muito semelhantes às dos números primos, envolvendo distribuições de Poisson [7-9].

5. Momentos e Cumulantes

Seja $\rho(x)$ uma distribuição de probabilidade. As grandezas

$$M_n = \int x^n \rho(x) dx = \langle x^n \rangle \tag{12}$$

são seus momentos. Note que $M_0 = 1$ para toda distribuição, já que são sempre normalizadas. O primeiro momento corresponde ao valor médio da variável. A diferença entre o segundo momento e o quadrado do primeiro é a chamada variância, e mede o quanto a variável se distancia de sua média,

$$\sigma = M_2 - M_1^2 = \langle (x - \langle x \rangle)^2 \rangle. \tag{13}$$

A função geratriz dos momentos,

$$f(t) = \sum_{n=0}^{\infty} \frac{M_n t^n}{n!}, \tag{14}$$

chamada de função característica, se relaciona com a distribuição da seguinte maneira:

$$\begin{aligned} f(t) &= \sum_{n=0}^{\infty} \frac{t^n}{n!} \int x^n \rho(x) dx = \int \rho(x) \left(\sum_{n=0}^{\infty} \frac{(tx)^n}{n!} \right) dx \\ &= \int \rho(x) e^{tx} dx, \end{aligned} \tag{15}$$

ou seja, por uma espécie de transformada de Laplace (note que ainda não especificamos o espaço no qual a variável está definida), que talvez possa ser invertida para fornecer $\rho(x)$ dada $f(t)$. Em vez da função geratriz, muitas vezes se usa a chamada função característica, que é simplesmente

$$\varphi(t) = f(it) = \int \rho(x) e^{itx} dx. \tag{16}$$

O resultado acima é interessante porque muitas vezes os momentos podem ser obtidos experimentalmente, ou numericamente. Ou seja, podemos medir a média, a variância, etc. Tendo em mãos essas quantidades podemos, em tese, obter informações a respeito da distribuição que as gerou.

Outro conjunto de números que pode ser usado para caracterizar uma distribuição são seus cumulantes. Seja C_n o cumulante de ordem n e seja

$$g(t) = \sum_{n=1}^{\infty} \frac{C_n t^n}{n!} \tag{17}$$

a função geratriz desses números. Os cumulantes são definidos em termos dos momentos através de uma

relação entre as respectivas funções geratrizes:

$$g(t) = \log[f(t)], \quad f(t) = e^{g(t)}. \tag{18}$$

Como

$$\begin{aligned} &\log\left(1 + M_1t + \frac{M_2t^2}{2} + O(t^3)\right) \\ &= M_1t + (M_2 - M_1^2)\frac{t^2}{2} + O(t^3), \end{aligned} \tag{19}$$

vemos que o primeiro cumulante coincide com a média e o segundo é justamente a variância. O terceiro e o quarto cumulantes são chamados, respectivamente, de obliquidade e curtose. Nos casos mais simples, a obliquidade e a curtose medem, respectivamente, o quanto a distribuição é inclinada e o quanto seu pico é achatado.

Todos os cumulantes de ordem maior que 2 são nulos para uma distribuição gaussiana. Sendo assim, podemos dizer que, de certa forma, os cumulantes de uma distribuição $\rho(x)$ medem o quanto ela difere de uma gaussiana.

Os cumulantes possuem duas propriedades importantes, que não são compartilhadas pelos momentos: 1) os cumulantes da soma de variáveis aleatórias independentes são iguais às somas dos cumulantes individuais; 2) com exceção do primeiro, são invariantes por translação, ou seja, os cumulantes de $\rho(x+a)$ são iguais aos de $\rho(x)$.

Os quatro primeiros momentos são dados em função dos cumulantes por

$$M_1 = C_1, \tag{20}$$

$$M_2 = C_2 + C_1^2, \tag{21}$$

$$M_3 = C_3 + 3C_2C_1 + C_1^3, \tag{22}$$

$$M_4 = C_4 + 4C_3C_1 + 3C_2^2 + 6C_2C_1^2 + C_1^4. \tag{23}$$

Os coeficientes que aparecem nessa fórmulas são conhecidos como números de Faà di Bruno. Para descrevê-los, precisamos do conceito de partição. As partições de um inteiro positivo são as maneiras de escrevê-lo como soma de inteiros positivos. Por exemplo, o número 4 tem cinco partições: 4, 3+1, 2+2, 2+1+1 e 1+1+1+1. O número de partes da partição λ é denotado $\ell(\lambda)$, e $r_j(\lambda)$ é o número de vezes que a parte j aparece em λ . Por exemplo, $\ell(2, 1, 1) = 3$, $r_1(2, 1, 1) = 2$, $r_2(2, 1, 1) = 1$. Em geral, temos

$$M_n = \sum_{\lambda \vdash n} \frac{n!}{\prod_j \lambda_j! r_j(\lambda)!} \prod_{i \in \lambda} C_i, \tag{24}$$

onde a soma é sobre todas as partições de n . Por outro lado, os primeiros cumulantes em função dos momentos são:

$$C_1 = M_1, \tag{25}$$

$$C_2 = M_2 - M_1^2, \tag{26}$$

$$C_3 = M_3 - 3M_2M_1 + 2M_1^3, \tag{27}$$

$$C_4 = M_4 - 4M_3M_1 - 3M_2^2 + 12M_2M_1^2 - 6M_1^4, \tag{28}$$

e a fórmula geral é:

$$C_n = \sum_{\lambda \vdash n} \frac{(-1)^{\ell(\lambda)-1} (\ell(\lambda) - 1)! n!}{\prod_j \lambda_j! r_j(\lambda)!} \prod_{i \in \lambda} M_i. \tag{29}$$

A relação entre momentos e cumulantes possui uma interpretação interessante quando vista em termos gráficos. Representamos o cumulante C_n por um grafo cíclico que possui n vértices. O produto $C_n C_m$ corresponde a um grafo com $n + m$ vértices e dois ciclos, que visitam n e m vértices cada.

A Figura 1 mostra todas as possibilidades com 4 vértices. Temos 1 caso que consiste de quatro componentes triviais, de um vértice só, correspondendo a C_1^4 ; temos 6 casos que consistem de duas componentes triviais e uma componente de dois vértices, correspondendo a $C_1^2 C_2$; temos 3 casos que consistem de duas componentes de dois vértices, correspondendo a C_2^2 ; temos 4 casos que consistem de um componente trivial e uma de três vértices, correspondendo a $C_1 C_3$; e temos 1 caso totalmente conexo, que corresponde a C_4 . Note como cada conjunto de grafos na figura corresponde a um termo na expressão de M_4 em termos de cumulantes, Eq. (23).

De acordo com essa interpretação, dizemos que o cumulante C_n corresponde ao diagrama cíclico de n vértices conexo, enquanto o momento M_n corresponde a todos os diagramas de n vértices com componentes cíclicas, inclusive os que não são conexos.

De fato, essa relação entre diagrams conexos/desconexos é mais geral do que a relação entre cumulantes/momentos. Seja f uma função que associa a qualquer grafo um número real. Suponhamos que a função seja multiplicativa, no sentido de que seu valor em um grafo

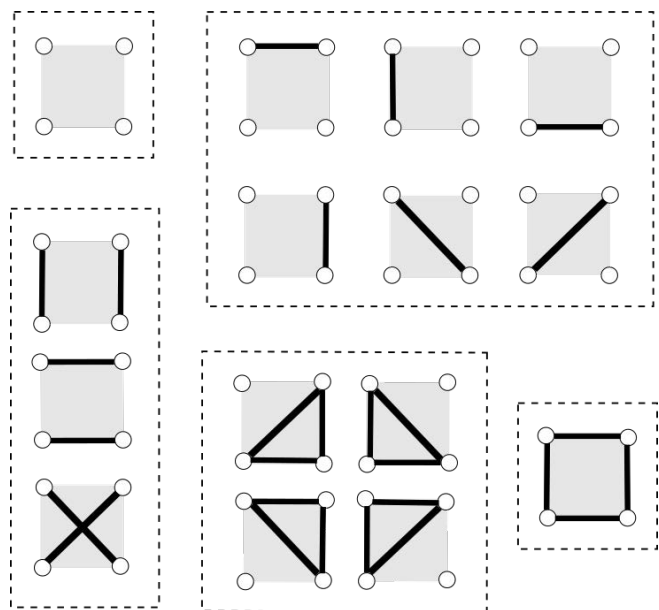


Figura 1: Relação gráfica entre momentos e cumulantes: decomposição de grafos em componentes conexas.

desconexo é dado pelo produto dos valores calculados sobre as componentes conexas: $F(A \cup B) = F(A)F(B)$. Então, se

$$s = \sum_{g \in \text{todos}} F(g) \tag{30}$$

é uma soma de F sobre todos os grafos com componentes conexas de um certo tipo (cíclicas, por exemplo), podemos obter a mesma soma porém restrita a grafos conexos do mesmo tipo tomando o logaritmo:

$$\log(s) = \sum_{g \in \text{conexos}} F(g). \tag{31}$$

6. Distribuições Beta, Gamma e Gaussiana

Suponha que temos um problema físico que queremos descrever de maneira estatística. Temos uma variável aleatória x e queremos associar a ela alguma distribuição. Qual a distribuição mais natural? A resposta depende da natureza do espaço no qual a variável está definida: ele é finito, semi-infinito ou duplamente infinito?

6.1. Intervalo finito, Beta

Algumas variáveis aleatórias têm seu valor restrito a um intervalo finito. A posição de uma partícula dentro de uma caixa, por exemplo, ou a posição de um oscilador com energia finita, ou a condutância de um sistema eletrônico.

A distribuição uniforme é a mais simples que podemos imaginar nessa situação. Em um intervalo finito, (a, b) , ela é dada por $\rho(x) = 1/(b - a)$. No caso do intervalo $(0, 1)$, temos $\rho(x) = 1$. Essa distribuição tem momentos dados por

$$\int_0^1 x^n dx = \frac{1}{n + 1} \tag{32}$$

e função geratriz $f(t) = \frac{1}{t}(e^t - 1)$.

Existe uma generalização da distribuição constante no intervalo finito $(0, 1)$, chamada de distribuição Beta, que é especificada por dois parâmetros positivos:

$$\rho_{\text{Beta}}^{\alpha, \beta}(x) = Bx^{\alpha-1}(1-x)^{\beta-1}. \tag{33}$$

Para que ela seja corretamente normalizada, é preciso que

$$\frac{1}{B} = \int_0^1 x^{\alpha-1}(1-x)^{\beta-1}, \tag{34}$$

e essa integral é conhecida como a primeira integral de Euler ou integral Beta, daí o nome da distribuição. Ela é dada em termos da função Gamma (veja na próxima subseção):

$$B = \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)}. \tag{35}$$

Podemos generalizar a distribuição Beta para um intervalo genérico (a, b) , usando a forma $(x - a)^{\alpha-1}(b - x)^{\beta-1}$.

A distribuição constante corresponde ao caso particular $\alpha = \beta = 1$. Para que a distribuição seja normalizável temos que ter $\alpha, \beta \geq 1/2$.

Se $\alpha = \beta$, a distribuição é simétrica. Se $\alpha > 1$, valores muito pequenos de x são pouco prováveis, e o contrário acontece se $\alpha < 1$ (a situação na vizinhança de $x = 1$ é análoga, com β em lugar de α). Os momentos da distribuição Beta são dados por

$$\int_0^1 Bx^n x^{\alpha-1}(1-x)^{\beta-1} dx = \frac{\Gamma(\alpha + n)\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\alpha + \beta + n)}. \tag{36}$$

Um exemplo importante da distribuição Beta é o seguinte. Imagine que temos um oscilador harmônico que satisfaz $mv^2 + kx^2 = kx_0^2$, ou seja, tem frequência angular ω e amplitude x_0 , mas não temos nenhuma informação sobre onde ele está. Se medirmos sua posição num instante qualquer, é razoável supor que a probabilidade de que esteja num intervalo de tamanho dx é proporcional ao tempo que ele fica nesse intervalo, portanto inversamente proporcional à velocidade que tem quando passa por ali, ou seja,

$$\frac{dx}{v(x)} = \frac{dx}{\omega \sqrt{x_0^2 - x^2}}. \tag{37}$$

Essa é uma distribuição Beta no intervalo $(-x_0, x_0)$ com parâmetros $\alpha = \beta = \frac{1}{2}$. Esse tipo de densidade de probabilidade aparece também em aproximações semiclássicas da mecânica quântica.

Se distribuirmos pontos de maneira uniforme sobre a circunferência S^1 , ou seja se tivermos o ângulo θ com distribuição uniforme, então a projeção desses pontos sobre o eixo horizontal $x = \cos \theta$ será $\frac{1}{|dx/d\theta|} = \frac{1}{|\sin(\theta)|}$, que em termos de $y = x^2$ é a distribuição Beta em $(0, 1)$ com parâmetros $\alpha = \beta = \frac{1}{2}$. De forma mais geral, como discutido em [10], a projeção da esfera em \mathbb{R}^N sobre \mathbb{R}^M leva a uma distribuição Beta para o quadrado da distância até a origem, com parâmetros $\alpha = M/2$ e $\beta = (N - M)/2$.

Outro exemplo notável é distribuição de Wigner, conhecida como lei do semi-círculo, dada por $\frac{1}{2\pi} \sqrt{4 - x^2}$. Ela pode ser vista como um caso particular da Beta no intervalo $(-2, 2)$ com parâmetros $\alpha = \beta = 3/2$. Essa distribuição descreve os autovalores de matrizes hermitianas e com elementos independentes e distribuídos de forma gaussiana [11], no regime de grandes dimensões. Seus momentos pares são os chamados números de Catalão,

$$\frac{1}{2\pi} \int_{-2}^2 x^{2n} \sqrt{4 - x^2} = \frac{1}{n + 1} \binom{2n}{n}. \tag{38}$$

Podemos mencionar ainda o contexto de transporte eletrônico através de cavidades quânticas com dinâmica clássica caótica [12]. Se as guias de entrada e saída

têm apenas um canal de transmissão aberto cada uma, então a condutância do sistema, $0 \leq g \leq 1$ em unidades apropriadas, tem distribuição de probabilidade constante se não houver simetria de reversão temporal e proporcional a $g^{-1/2}$ se houver essa simetria (ambas distribuições Beta).

6.2. Intervalo semi-infinito, Gamma

Algumas variáveis aleatórias podem estar restritas a um intervalo semi-infinito, tipicamente $(0, \infty)$. É o caso, por exemplo, da coordenada radial de uma partícula em duas ou três dimensões, ou do intervalo de tempo entre dois decaimentos radioativos do mesmo átomo, ou duas colisões no modelo de Drude.

No modelo de Drude, um elétron se move livremente exceto por colisões com impurezas imóveis. Digamos que $P(t)$ é a probabilidade de nenhuma colisão ter ocorrido durante um intervalo de tempo t . É razoável supor que duas colisões são eventos independentes e portanto a probabilidade de não ocorrer colisão em $(0, t_1 + t_2)$ seja igual à probabilidade de não ocorrer colisão em $(0, t_1)$ multiplicada pela probabilidade de não ocorrer colisão em $(t_1, t_1 + t_2)$. Supondo ainda que a probabilidade depende apenas do intervalo de tempo e não dos instantes inicial e final, queremos que $P(t_1 + t_2) = P(t_1)P(t_2)$. Concluímos imediatamente que $P(t) = e^{-t/\tau}$ para algum parâmetro τ .

A distribuição exponencial $\frac{1}{\xi}e^{-x/\xi}$ é o caso mais simples de distribuição Gamma. Seu valor médio é ξ e sua variância é ξ^2 .

Por exemplo, se x e y são coordenadas no plano e têm ambas a mesma distribuição gaussiana, proporcional a e^{-x^2} e a e^{-y^2} , respectivamente, então a distribuição da distância até a origem, $r = \sqrt{x^2 + y^2}$, é proporcional a re^{-r^2} . Se definirmos $z = r^2$, então a variável z tem distribuição exponencial.

A distribuição Gamma mais geral contém dois parâmetros reais e é dada por

$$\rho_{\text{Gamma}}^{\alpha, \xi}(x) = Gx^{\alpha-1}e^{-x/\xi}. \quad (39)$$

A normalização exige que

$$\frac{1}{G} = \int_0^{\infty} x^{\alpha-1}e^{-x/\xi} dx, \quad (40)$$

que é conhecida como a segunda integral de Euler ou integral Gamma. Sua solução é justamente a função Gamma,

$$\int_0^{\infty} x^{\alpha-1}e^{-x} dx = \Gamma(\alpha), \quad (41)$$

uma generalização do fatorial, pois se α é um número inteiro, então $\Gamma(\alpha) = (\alpha - 1)!$. A constante de normalização é

$$G = \frac{1}{\xi^\alpha \Gamma(\alpha)}. \quad (42)$$

Por exemplo, a função de onda do estado fundamental do átomo de hidrogênio decai exponencialmente a partir da origem, sendo proporcional a e^{-r/a_0} onde a_0 é a raio de Bohr. A densidade de probabilidade correspondente, em coordenadas esféricas, é proporcional a $r^2 e^{-2r/a_0}$ (note o fator r^2 que vem do elemento de volume). Essa é uma distribuição Gamma com $\alpha = 3$ e $\xi = a_0/2$.

Se $\alpha < 1$, a distribuição Gamma favorece o valor $x = 0$. Se $\alpha > 1$, então $x = 0$ passa a ser improvável (é necessário que $\alpha \geq 1/2$).

Os momentos da distribuição Gamma são

$$\int_0^{\infty} x^n \rho_{\text{Gamma}}^{\alpha, \xi}(x) dx = \xi^n \frac{\Gamma(\alpha + n)}{\Gamma(\alpha)}. \quad (43)$$

A função geratriz desses momentos é

$$f(t) = \frac{1}{(1 - \xi t)^\alpha}, \quad (44)$$

que realmente satisfaz

$$f(t) = \int_0^{\infty} \rho_{\text{Gamma}}^{\alpha, \xi}(x) e^{tx} dx. \quad (45)$$

Já os cumulantes são dados simplesmente por $C_n = \alpha(n - 1)!\xi^n$, já que a sua função geratriz é

$$g(t) = -\alpha \log(1 - \xi t). \quad (46)$$

É curioso que, quando $\alpha = n + 1$ é um número inteiro e $\xi = 1$, a distribuição Gamma coincide com a distribuição de Poisson, $\rho_{\text{Gamma}}^{n+1, 1}(x) = p_{\text{Poisson}}^x(n)$.

É interessante notar que a distribuição Gamma pode ser encarada como um caso limite da distribuição Beta em $(0, 1)$: se temos $x^{\alpha-1}(1-x)^\beta$ e fazemos $x = y/(\xi\beta)$ e depois tomamos $\beta \rightarrow \infty$, o resultado é proporcional a $y^{\alpha-1}e^{-y/\xi}$.

6.2.1. Distribuição de Pareto

Podemos generalizar a distribuição Gamma usando um intervalo semi-infinito (a, ∞) com $a > 0$. A versão mais popular dessa situação é a que obtemos ao fazer $\xi \rightarrow \infty$. Nesse caso, temos a distribuição de Pareto, que corresponde a uma lei de potência:

$$\rho_{\text{Pareto}}^{a, \alpha}(x) = \frac{\alpha a^\alpha}{x^{\alpha+1}}. \quad (47)$$

Empiricamente, observa-se que essa distribuição descreve bem a distribuição de renda em um país (muita gente ganhando pouco, raros milionários) e a distribuição do tamanho das cidades (muitas cidades pequenas, poucas metrópoles), entre outros dados obtidos empiricamente. Leis de potência parecem ser muito comuns na natureza, ainda que sua real validade matemática seja discutível [13].

Como a distribuição de Pareto é muito concentrada, ela deu origem ao chamado “princípio de Pareto” de que 80% dos resultados correspondem a 20% das causas (ou

seja, 80% da população viveria em 20% das cidades, ou 80% da riqueza estaria nas mãos dos 20% mais ricos, 80% das vendas de uma loja seriam feitas para 20% dos clientes etc). Esse princípio só vale exatamente se a distribuição tiver parâmetro $\alpha = \log_4 5 \approx 1,16$, mas já foi observado (em proporções um pouco diferentes) em várias situações.

A distribuição de Pareto só tem os primeiros momentos,

$$M_n = \frac{\alpha a^n}{\alpha - n}, \quad n < \alpha. \tag{48}$$

Os momentos com ordem maior que α divergem, de modo que a função geratriz não existe.

Se x tem distribuição de Pareto no intervalo (a, ∞) , então a variável $y = \log(x/a)$ tem distribuição exponencial.

6.3. Intervalo duplamente infinito, Gaussiana

E se nossa variável aleatória não estiver contida em um intervalo finito mas, em vez disso, puder assumir qualquer valor real? Nesse caso, a distribuição gaussiana é de longe a mais popular:

$$\rho_{\text{Gauss}}^{\mu, \sigma}(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}. \tag{49}$$

Essa distribuição tem valor médio μ e variância σ^2 .

Sendo a distribuição uma função par em torno da média, os momentos ímpares em torno da média são nulos,

$$\int_{-\infty}^{\infty} (x - \mu)^{2n+1} \rho_{\text{Gauss}}^{\mu, \sigma}(x) dx = 0. \tag{50}$$

Por outro lado, os momentos pares em torno da média são dados por

$$\int_{-\infty}^{\infty} (x - \mu)^{2n} \rho_{\text{Gauss}}^{\mu, \sigma}(x) dx = \sigma^{2n} (2n - 1)!! = \sigma^{2n} \frac{(2n)!}{2^n n!}. \tag{51}$$

Os momentos originais, por sua vez, são um pouco mais complicados,

$$\int_{-\infty}^{\infty} x^n \rho_{\text{Gauss}}^{\mu, \sigma}(x) dx = \sum_{k=0}^{\lfloor n/2 \rfloor} \frac{n!}{k!(n-2k)! 2^k} \sigma^{2k} \mu^{n-2k}. \tag{52}$$

A função geratriz desses momentos é

$$f(t) = e^{t^2 \sigma^2 / 2 + t\mu}. \tag{53}$$

Neste caso, a função característica, Eq. (16), e a distribuição original podem ser relacionadas por uma transformada de Fourier,

$$\varphi(t) = e^{-t^2 \sigma^2 / 2 + it\mu} = \int_{-\infty}^{\infty} \rho_{\text{Gauss}}^{\mu, \sigma}(x) e^{itx} dx. \tag{54}$$

Por outro lado, a função geratriz dos cumulantes é simplesmente

$$g(t) = t^2 \sigma^2 / 2 + t\mu, \tag{55}$$

de modo que todos os cumulantes de ordem maior que 2 são nulos para uma distribuição gaussiana, como já comentamos anteriormente.

O estado fundamental do oscilador harmônico quântico, ou seja, a solução da equação de Schrödinger

$$-\frac{\hbar^2}{2m} \frac{d^2}{dx^2} \psi(x) + \frac{m\omega^2}{2} x^2 \psi(x) = \frac{\hbar\omega}{2} \psi(x), \tag{56}$$

dá origem a uma distribuição de probabilidade gaussiana,

$$|\psi(x)|^2 = \sqrt{\frac{m\omega}{\pi\hbar}} e^{-\frac{m\omega}{\hbar} x^2}, \tag{57}$$

com média 0 e variância proporcional a $\sqrt{\hbar}$.

A gaussiana é também a solução fundamental da equação de difusão, $\frac{\partial \psi}{\partial t} = D \frac{d^2 \psi}{dx^2}$, para $x \in (-\infty, \infty)$. Isso significa que, se a condição inicial for $\psi(x, 0) = \delta(x)$, a solução no tempo t será dada por

$$\psi(x, t) = \frac{1}{\sqrt{4\pi Dt}} e^{-\frac{x^2}{4Dt}}. \tag{58}$$

Parte da importância da distribuição gaussiana vem do chamado Teorema Central do Limite: a soma de um número grande de variáveis independentes e identicamente distribuídas, com variância finita, será sempre gaussiana, mesmo que essas variáveis não o sejam. Concretamente, digamos que x_i , com $1 \leq i \leq N$, são N variáveis aleatórias independentes, cada uma dela com distribuição $\rho(x)$, sendo que $\langle x \rangle = \mu$ e $\langle (x - \mu)^2 \rangle = \sigma^2$. Então, quando $N \rightarrow \infty$ a distribuição da variável $y = (x_1 + \dots + x_N)/N$ fica indistinguível da função $\rho_{\text{Gauss}}^{\mu, \sigma/\sqrt{N}}(y)$, qualquer que seja $\rho(x)$.

O exemplo mais famoso do teorema central do limite é o movimento Browniano. Em uma dimensão, se uma partícula se move de forma aleatória, tendo a cada passo uma certa probabilidade de se deslocar um pouco para esquerda ou um pouco para a direita, sua distância a partir de onde começou terá, depois de um tempo longo, distribuição gaussiana [14], com variância que cresce linearmente com o tempo (essencialmente, esse é o mesmo problema resolvido pela equação de difusão mencionada acima).

Em um espaço de dimensão N , o quadrado da distância até a origem é dado por $r^2 = \sum_{i=1}^N x_i^2$. Se os x_i são variáveis aleatórias independentes, cada uma delas com distribuição $\rho_{\text{Gauss}}^{0,1}(x)$, então a distribuição da variável r é

$$\rho_{\text{Chi}}^N(r) = \frac{r^{N-1} e^{-r^2/2}}{2^{N/2-1} \Gamma(N/2)}, \tag{59}$$

que é chamada de distribuição Chi. Para $N = 2$ ela é conhecida como distribuição de Rayleigh e descreve,

por exemplo, o primeiro estado excitado do oscilador harmônico quântico e também os autovalores de matrizes circulares aleatórias de alta dimensão [15]. Para $N = 3$ ela é a distribuição de Maxwell-Boltzmann, que descreve a velocidade das partículas que compõem um gás.

A distribuição Chi também descreve a diferença $\lambda_2 - \lambda_1$ entre o maior e o menor autovalor de matrizes hermitianas 2×2 com distribuição gaussiana. Talvez valha a pena fazer essa conta explicitamente no caso de matrizes reais. A diferença entre os autovalores da matriz $M = \begin{pmatrix} a & b \\ b & c \end{pmatrix}$ é $d = \sqrt{(a-c)^2 + 4b^2}$. Tomamos a distribuição conjunta como sendo proporcional a

$$e^{-\frac{1}{2}\text{Tr}(M^2)} = e^{-a^2/2} e^{-c^2/2} e^{-b^2}. \quad (60)$$

Fazendo a mudança de variáveis $a - c = d \cos \theta$, $2b = d \sin \theta$ e $a + c = e$, temos um jacobiano que é proporcional a d e

$$\frac{a^2}{2} + \frac{c^2}{2} + b^2 = \frac{d^2 + e^2}{4}. \quad (61)$$

A integral sobre e é trivial e chegamos ao resultado (que corresponde à Chi com $N = 2$). O aspecto mais interessante desse cálculo é justamente o jacobiano, pois ele mostra que os autovalores se repelem: a variável d tem pouca chance de ser nula. Em outras palavras, matrizes aleatórias não costumam ser degeneradas [16].

Uma observação interessante é que, se r tem a distribuição Chi apresentada acima, a variável $z = r^2$ tem distribuição

$$\frac{z^{N/2-1} e^{-z/2}}{2^{N/2} \Gamma(N/2)}, \quad (62)$$

que é às vezes chamada de distribuição Chi-quadrado, mas que é simplesmente uma distribuição Gamma com $\alpha = N/2$ e $\xi = 2$.

Finalmente, é interessante notar que a distribuição gaussiana pode ser encarada como um caso limite da distribuição Beta simétrica em $(-1, 1)$: se temos $(1+x)^\beta (1-x)^\beta$ e fazemos $x = y/\sqrt{\beta}$ e depois tomamos $\beta \rightarrow \infty$, o resultado é exatamente e^{-y^2} .

6.4. Princípio variacional

Comentamos na Seção 3 que a quantidade

$$S[\rho] = - \int_{-\infty}^{\infty} \rho(x) \log[\rho(x)] dx \quad (63)$$

não é uma definição adequada de entropia. Não obstante, vale notar que a gaussiana é a distribuição que maximiza esse funcional, dados os três vínculos

$$\int_{-\infty}^{\infty} \rho(x) dx = 1, \quad (64)$$

$$\int_{-\infty}^{\infty} x \rho(x) dx = M_1 \quad (65)$$

e

$$\int_{-\infty}^{\infty} x^2 \rho(x) dx = M_2. \quad (66)$$

De fato, a equação de Euler-Lagrange para esse problema é dada por

$$\frac{\delta S}{\delta \rho} = -\log(\rho) - 1 - \nu_0 - \nu_1 x - \nu_2 x^2 = 0, \quad (67)$$

cujas soluções são imediatas e fornece justamente a gaussiana. Assim, de certa forma pode-se dizer que, ao supormos que uma variável aleatória definida em $(-\infty, \infty)$ tem distribuição gaussiana, estamos fazendo a escolha de “máxima entropia”, ou seja, a hipótese que pressupõe o mínimo de informação a respeito do problema.

No caso do intervalo semi-infinito $(0, \infty)$, podemos acrescentar uma penalidade para o valor $x = 0$, se quisermos tornar menos provável que a variável tenha esse valor. Uma maneira de fazer isso é acrescentar um termo logarítmico ao funcional e escrever

$$S[\rho] = - \int_0^{\infty} \{\rho(x) \log[\rho(x)] + (\alpha - 1)\rho(x) \log x\} dx. \quad (68)$$

A distribuição Gamma maximiza esse funcional, com os vínculos (64) e (65).

Finalmente, para uma variável aleatória no intervalo $(0, 1)$ introduzimos duas penalidades,

$$S[\rho] = - \int_0^1 \{\rho(x) \log[\rho(x)] + (\alpha - 1)\rho(x) \log x + (\beta - 1)\rho(x) \log(1 - x)\} dx. \quad (69)$$

A distribuição Beta maximiza esse funcional, com o vínculo (64).

7. Outras Distribuições Importantes

As distribuições Beta, Gamma e gaussiana são fundamentais, mas há várias outras distribuições de probabilidades importantes além delas. Mencionemos alguns exemplos.

7.1. Distribuição lorentziana

A distribuição lorentziana (também chamada distribuição de Cauchy):

$$\frac{a}{\pi((x - \mu)^2 + a^2)}, \quad -\infty < x < \infty. \quad (70)$$

Uma característica dessa distribuição é que, como ela decai muito devagar, não tem nenhum momento bem definido. Nem mesmo a média existe. A grandeza μ é a moda e a mediana, mas não é a média; por outro lado, a grandeza a é a largura à meia altura, mas não é o desvio padrão.

Em termos de um parâmetro complexo, $z = \mu + ia$, a distribuição lorentziana pode ser escrita como $\frac{1}{\pi} \text{Im} \left(\frac{1}{x-z} \right)$. Sua função característica é bastante simples,

$$\int_{-\infty}^{\infty} e^{itx} \frac{a}{\pi((x-\mu)^2 + a^2)} dx = e^{i\mu t - a|t|}. \quad (71)$$

Se $t > 0$, essa função é simplesmente e^{izt} .

A largura de linha dos estados excitados de um átomo (existentes por causa da interação com o vácuo eletromagnético e responsável pelo tempo de vida finito dos estados excitados) é uma lorentziana (a variável sendo a energia).

A lorentziana não está sujeita ao teorema central do limite, por não ter variância. Em vez disso, ela satisfaz sua própria versão do teorema: a soma de variáveis identicamente distribuídas com distribuição lorentziana também tem distribuição lorentziana.

Se x e y são duas variáveis independentes com distribuição gaussiana, de média 0 e variância 1, então a razão x/y tem distribuição lorentziana, com $\mu = 0$ e $a = 1$.

7.2. Distribuições de valores extremos

Seja uma variável aleatória x com distribuição de probabilidade $\rho(x)$. Sua distribuição cumulativa

$$P(x < X) = \int_{-\infty}^X \rho(x) dx \quad (72)$$

representa a probabilidade de que x tenha valor menor do que X .

Imaginemos uma nova variável aleatória \tilde{x} que é o maior valor de x dentre um sorteio de N realizações independentes,

$$\tilde{x} = \max(x_1, \dots, x_N). \quad (73)$$

É claro que $\tilde{x} \leq X$ se e somente se $x_i \leq X$ para todo i . Portanto,

$$P(\tilde{x} < X) = \prod_{i=1}^N P(x_i < X) = [P(x < X)]^N. \quad (74)$$

Por exemplo, se $\rho(x) = \frac{1}{\xi} e^{-x/\xi}$, então $P(x < X) = (1 - e^{-X/\xi})$. Subtraindo um fator $\xi \log(N)$, por conveniência, temos $P(\tilde{x} - \xi \log(N) < X) = P(\tilde{x} < X + \xi \log(N))$ e

$$P(\tilde{x} < X + \xi \log(N)) = \left(1 - \frac{e^{-X/\xi}}{N} \right)^N \rightarrow e^{-e^{-X/\xi}}, \quad (75)$$

onde tomamos o limite $N \rightarrow \infty$. Esse é o caso mais simples de uma variável com distribuição de Gumbel, cuja cumulativa é em geral dada por $e^{-e^{-(x-\mu)/\sigma}}$.

Temos aqui um caso de universalidade [17], porque o valor extremo deslocado, $\tilde{x} - q_N$, tem distribuição de

Gumbel, quando $N \rightarrow \infty$, para qualquer distribuição original $\rho(x)$ que decaia mais rápido do que qualquer potência quando $x \rightarrow \infty$. O valor do deslocamento é dado pela igualdade $P(x < q_N) = 1 - 1/N$.

A distribuição de Gumbel já foi usada para descrever a chuva mais forte do ano e outros fenômenos meteorológicos [18]. Matematicamente, ela descreve bem o módulo do maior autovalor de matrizes não-hermitianas [19].

Se $\rho(x)$ decai como uma potência, $\rho(x) \sim 1/x^{\alpha+1}$, como no caso da distribuição de Pareto, então o valor extremo tem distribuição de Fréchet, cuja cumulativa é do tipo e^{-N/X^α} .

Existe ainda uma terceira classe de universalidade, quando $\rho(x) = 0$ para $x > a$ e $\rho(x) \sim (x-a)^{\alpha-1}$ quando x tende a a pela esquerda (essa situação inclui uma distribuição constante num intervalo finito, por exemplo). Nesse caso, o valor extremo tem distribuição de Weibull, cuja cumulativa é do tipo $e^{-N(a-X)^\alpha}$.

8. Conclusão

Distribuições de probabilidade são um conceito fundamental em física e matemática, indispensáveis na compreensão da mecânica estatística e da mecânica quântica, sem falar no tratamento estatístico de dados que hoje em dia só cresce, com computadores cada vez mais rápidos processando quantidades de dados cada vez maiores.

Apresentamos neste artigo uma discussão acessível das principais distribuições de probabilidade que são encontradas na física, mencionando suas principais propriedades e exemplos de aplicações. Esperamos assim contribuir para a formação de estudantes de física, fornecendo um material de referência e de apoio àqueles que estão encontrando essas ideias pela primeira vez.

Agradeço o apoio financeiro do CNPq (306765/2018-7) e os comentários de um parecerista anônimo.

Referências

- [1] <http://www.numeromania.com.br/sena.html>.
- [2] S.K. Baekl, S. Bernhardsson e P. Minnhagen, *New J. Phys.* **13**, 043004 (2011).
- [3] M. Visser, *New J. Phys.* **15**, 043021 (2013).
- [4] P. Biswas, H. Shimoyama e L.R. Mead, *J. Phys. A: Math. Theor.* **43**, 125103 (2010).
- [5] C. Beck e F. Schlogl, *Thermodynamics of Chaotic Systems* (Cambridge University Press, Cambridge, 1993).
- [6] WIKIPEDIA, *Erdős-Kac Theorem*, disponível em: https://en.wikipedia.org/wiki/Erdős-Kac_theorem, acessado em 04/01/2022.
- [7] A. Granville, *The anatomy of integers and permutations*, disponível em: <https://dms.umontreal.ca/~andrew/MSI/AnatomyForTheBook.pdf>, acessado em 04/01/2022.
- [8] R. Arratia, A.D. Barbour e S. Tavaré, *Notices Am. Math. Soc.* **44**, 903 (1997).

- [9] J. Baez, *Random permutations*, disponível em: <https://math.ucr.edu/home/baez/permutations/>, acessado em 04/01/2022.
- [10] K. Życzkowski e H.J. Sommers, *J. Phys. A: Math. Gen.* **33**, 2045 (2000).
- [11] G. Livan, M. Novaes e P. Vivo, *Introduction to Random Matrices: Theory and Practice* (Springer, Cham, 2018).
- [12] C.W.J. Beenakker, *Rev. Mod. Phys.* **69**, 731 (1997).
- [13] A. Clauset, C.R. Shalizi e M.E.J. Newman, *SIAM Review* **51**, 661 (2009).
- [14] WIKIPEDIA, *Brownian motion*, disponível em: https://en.wikipedia.org/wiki/Brownian_motion, acessado em 04/01/2022.
- [15] K. Saha, *Spectral properties of large dimensional random circulant type matrices*. Tese de Doutorado, Indian Statistical Institute, Calcutá (2010).
- [16] F. Haake, S. Gnutzmann e M. Kuś, *Quantum Signatures of Chaos* (Springer, Berlim, 2018), 4^a ed.
- [17] A. Hansen, *Front. Phys.* **8**, 533 (2020).
- [18] D. Koutsoyiannis, em: *ESF LESC Exploratory Workshop* (Bologna, 2003).
- [19] M. Novaes, *Phys. Rev. E* **85**, 036202 (2012).