



Universidade Federal de Uberlândia
Instituto de Física

Licenciatura em Física

ESTUDOS DAS FUNÇÕES ENTRÓPICAS

Keoma Hermenegildo Kurashima

Uberlândia-MG
2026

Keoma Hermenegildo Kurashima

ESTUDOS DAS FUNÇÕES ENTRÓPICAS

Trabalho de conclusão de curso apresentado à Coordenação do Curso de Licenciatura em Física como requisito parcial para obtenção do grau de Licenciado em Física.

Orientador: Juliano Gonçalves Oler

Uberlândia-MG

2026



**Universidade Federal de Uberlândia
Instituto de Física**

Coordenação do Curso de Licenciatura em Física

A banca examinadora, conforme abaixo assinado, certifica a adequação deste trabalho de conclusão de curso para obtenção do grau de Licenciado em Física.

Uberlândia, 19 de março de 2026.

BANCA EXAMINADORA

Juliano Gonçalves Oler

Marcel Novaes

Marcus Augusto Bronzi

**Uberlândia-MG
2026**

AGRADECIMENTOS

Dedico esta seção, primeiramente, ao meu orientador, Juliano Gonçalves Oler, por todos os anos de trabalho conjunto, por sua paciência para comigo ao longo desse período, bem como por sua compreensão e leveza durante nossa trajetória. Agradeço aos professores do curso de graduação em Física Licenciatura pela formação que me proporcionaram, com destaque especial ao professor Ricardo Kagimura, por evidenciar, com uma paciência inesgotável, o que significa ser um educador. Recordo-me de que ele dedicou horas a explicar-me ângulo em sua sala, um conteúdo considerado elementar e de conhecimento geral, permitindo que eu partisse apenas quando havia assimilado o tema, demonstrando sempre total disponibilidade para seus alunos. À professora Mariana Mieko Odashima, por acreditar em mim como poucos o fizeram.

Ao Programa de Educação Tutorial (PET), pela estrutura fornecida para o desenvolvimento deste trabalho, assim como para meus estudos. Não poderia deixar de agradecer a meu pai, Edson Hermenegildo Kurashima, por ser um exemplo ético e moral, e também por demonstrar, na prática, a resiliência necessária para não se abalar diante das adversidades da vida. Ele e minha mãe são os principais responsáveis pela pessoa que me tornei.

Cabe ainda mencionar todos aqueles que considereei amigos e que me apoiaram, estando ao meu lado nos momentos mais difíceis. Por fim, mas não menos importante, à República Taverna, onde residi todos esses anos, por ser um espaço de entretenimento, descanso e construção de laços nos instantes em que me sentia exaurido, em especial a Tequila, uma companheira fiel.

RESUMO

Este trabalho realiza uma caracterização matemática do conjunto $\mathcal{B}$ das funções de Boltzmann, formado pelas funções reais estritamente crescentes que satisfazem a aditividade multiplicativa $f(xy) = f(x) + f(y)$. O estudo investiga as propriedades analíticas e topológicas desse conjunto e estabelece uma conexão formal com a formulação informacional da entropia de Shannon. Utilizando métodos de análise real e teoria da probabilidade, demonstra-se que as funções em $\mathcal{B}$ são analíticas e de classe C^∞ , e que o conjunto é convexo e conexo, embora não seja aberto, fechado ou compacto na topologia da convergência uniforme em compactos. A caracterização fundamental revela que o domínio de $\mathcal{B}$ é o intervalo $(0, \infty)$ e que $\mathcal{B} = \{k \ln x \mid k > 0\}$. A partir dessa base, prova-se que a entropia de Shannon, restrita a variáveis aleatórias independentes, herda a estrutura matemática de $\mathcal{B}$ por meio de uma correspondência biunívoca. Conclui-se que a forma logarítmica da entropia é determinada unicamente por sua monotonicidade e aditividade, e que a formulação de Shannon generaliza a de Boltzmann-Gibbs ao acomodar sistemas com correlações arbitrárias, unificando os fundamentos da termodinâmica estatística e da teoria da informação.

Palavras-chave: Funções de Boltzmann, Entropia de Shannon, Caracterização matemática da entropia, Generalização entrópica.

ABSTRACT

This work conducts a mathematical characterization of the set $\mathcal{B}$ of Boltzmann functions, formed by strictly increasing real functions that satisfy the multiplicative additivity property $f(xy) = f(x) + f(y)$. The study investigates the analytical and topological properties of this set and establishes a formal connection with the informational formulation of Shannon's entropy. Using methods from real analysis and probability theory, it is demonstrated that the functions in $\mathcal{B}$ are analytic and of class C^∞ , and that the set is convex and connected, although it is not open, closed, or compact in the topology of uniform convergence on compact sets. The fundamental characterization reveals that the domain of $\mathcal{B}$ is the interval $(0, \infty)$ and that $\mathcal{B} = \{k \ln x \mid k > 0\}$. From this basis, it is proven that Shannon's entropy, when restricted to independent random variables, inherits the mathematical structure of $\mathcal{B}$ through a one-to-one correspondence. It is concluded that the logarithmic form of entropy is uniquely determined by its monotonicity and additivity, and that Shannon's formulation generalizes that of Boltzmann-Gibbs by accommodating systems with arbitrary correlations, unifying the foundations of statistical thermodynamics and information theory.

Keywords: Boltzmann functions, Shannon entropy, Mathematical characterization, Entropic generalization.

SUMÁRIO

Lista de Figuras	I
1 Introdução	1
2 Conhecimentos Prévios	5
2.1 Conceitos de Análise Real	5
2.1.1 Sequências e Séries	5
2.1.2 Limites e Continuidade	10
2.1.3 Cálculo Diferencial	12
2.2 Topologia dos Reais	15
2.2.1 Conjuntos Fundamentais	15
2.2.2 Propriedades Topológicas	16
2.3 Espaços de Funções	16
2.3.1 Estruturas de Espaços de Funções	17
2.3.2 Topologias em Espaços de Funções	17
2.3.3 Densidade e Aproximação	18
2.4 Teoria da Probabilidade	19
2.4.1 Variáveis Aleatórias Discretas	20
2.4.2 Probabilidade Condicional e Independência	20
2.4.3 Distribuições Diádicas	21
2.4.4 Relação de Refinamento	23
3 Conjunto das funções de Boltzmann	25
3.1 Propriedades fundamentais	25
3.2 Caracterização do domínio	26
3.3 Propriedades de regularidade	27
3.4 Propriedades topológicas do conjunto $\mathcal{B}$	30
3.5 Caracterização completa das funções de Boltzmann	33
4 Entropia de Shannon	35
5 Conclusões	43
Referências Bibliográficas	47

LISTA DE FIGURAS

1.1	Ilustração da propriedade extensiva da entropia. A entropia do sistema composto C , formado pela simples justaposição dos sistemas isolados A e B , é a soma das entropias de cada parte: $S(C) = S(A) + S(B)$	2
1.2	Ilustração da propriedade de aditividade para sistemas independentes. O número total de microestados do sistema composto $(A+B)$ é o produto do número de microestados de cada subsistema.	3

1. INTRODUÇÃO

Desde os primórdios da civilização, a humanidade busca compreender as leis fundamentais que regem a natureza e a estrutura do universo. Entre esses princípios, poucos são tão profundos e ubíquos quanto o conceito de entropia. Muito além de uma mera grandeza termodinâmica, a entropia emerge como uma narrativa poderosa sobre a seta do tempo, a desordem e a própria seta da complexidade que permeia desde a expansão cósmica até o surgimento da vida. Compreender a entropia é, portanto, desvendar um dos fios condutores que tecem a tapeçaria da realidade, respondendo à pergunta fundamental sobre por que os processos evoluem em uma direção e não em outra.

O papel central da entropia para a compreensão das leis fundamentais do universo foi expresso por Arthur Eddington, que em seu clássico “*A Natureza do Mundo Físico*” (1928), referiu-se à lei do aumento da entropia como ocupando “*a posição suprema entre as leis da Natureza*”. Para Eddington, um físico que não conseguisse compreender o significado profundo da entropia estaria fadado a uma visão incompleta da realidade física [3].

A concepção moderna de entropia emerge no século XIX a partir de dois pilares fundamentais: um termodinâmico e um estatístico.

O primeiro pilar foi estabelecido por **Rudolf Clausius**, que, ao formular a Segunda Lei da Termodinâmica, identificou uma grandeza que sempre aumenta em processos irreversíveis. Ele afirmou que “*a entropia do mundo tende a um máximo*” [2]. Esta declaração formaliza a observação física de que, espontaneamente, o calor sempre flui do corpo mais quente para o mais frio, nunca o inverso, tornando a entropia uma função que **sempre cresce** em sistemas isolados. Para entender isso, considere uma xícara de café quente deixada sobre a mesa em uma sala. Após algum tempo, o café esfria e a sala aquece ligeiramente, até que ambos atingem a mesma temperatura. Este processo espontâneo ilustra o crescimento contínuo da entropia.

Além de sempre aumentar, a entropia possui outra propriedade fundamental: ela é **extensiva**. Isso significa que seu valor total depende do tamanho ou da quantidade de matéria do sistema. Considere dois sistemas idênticos e isolados, cada um com sua própria entropia, que se chamam de $S(A)$ e $S(B)$.

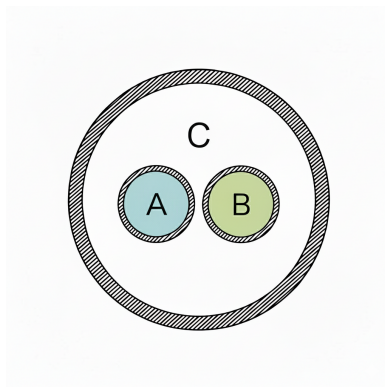


Figura 1.1: Ilustração da propriedade extensiva da entropia. A entropia do sistema composto C , formado pela simples justaposição dos sistemas isolados A e B , é a soma das entropias de cada parte: $S(C) = S(A) + S(B)$.

Se esses dois sistemas forem simplesmente juntados para formar um novo sistema C , sem permitir que troquem calor ou modifiquem suas estruturas internas, a entropia deste novo sistema será exatamente a **soma** das entropias dos sistemas originais: $S(C) = S(A) + S(B)$ [1]. Esta propriedade de aditividade é intuitiva e é crucial para a coerência da termodinâmica. Observa-se que duas xícaras de café quente, juntas, possuem o dobro da entropia de uma única xícara.

O segundo pilar foi erguido por **Ludwig Boltzmann**, que buscou uma explicação mecânica para essa irreversibilidade. Seu raciocínio pode ser ilustrado com um exemplo simples: considere um recipiente com gás. O **macroestado** é a descrição macroscópica de gás em equilíbrio, por exemplo: com uma pressão, temperatura e volume uniformes. O **microestado**, por sua vez, é a configuração instantânea de posições e velocidades de cada uma das inúmeras moléculas do gás. Boltzmann percebeu que um único macroestado, o gás uniformemente distribuído, pode ser resultado de um número astronomicamente grande de microestados diferentes. Enquanto isso, um macroestado muito ordenado, como todas as moléculas concentradas em um canto do recipiente, corresponde a um número muito pequeno de microestados. A evolução natural do sistema, portanto, é da configuração **menos provável** (ordenada) para a **mais provável** (desordenada). Esta tendência estatística para estados mais prováveis fornece assim uma explicação microscópica para o aumento da entropia descrito por Clausius: o crescimento da entropia corresponde à transição espontânea para macroestados compatíveis com um maior número de microestados.

A propriedade de aditividade da entropia também encontra uma fundamentação profunda nessa visão. Considere dois sistemas independentes, A e B , separados por uma parede. O número total de maneiras, microestados, de o sistema combinado ($A+B$) estar em um determinado estado é dado pelo **produto** do número de maneiras de A e de B estarem em seus respectivos estados, pois para cada configuração de A , há várias configurações possíveis para B .



Figura 1.2: Ilustração da propriedade de aditividade para sistemas independentes. O número total de microestados do sistema composto ($A+B$) é o produto do número de microestados de cada subsistema.

Como a entropia total do sistema composto deve ser a **soma** das entropias dos sistemas individuais, isso significa que a função que define a entropia a partir da probabilidade deve possuir a propriedade fundamental de converter uma **multiplicação** em uma **soma** [6]. Esta é a ponte conceitual entre o mundo microscópico das probabilidades e o comportamento macroscópico da entropia.

Do ponto de vista rigoroso, a entropia é uma grandeza que depende de toda a distribuição de probabilidade dos microestados de um sistema, caracterizando-se matematicamente como um funcional. Contudo, para desvendar sua estrutura essencial, é natural retornar à formulação original de Boltzmann, na qual todos os microestados acessíveis são igualmente prováveis.

Esse caso particular de probabilidades uniformes servirá de base para compreender a evolução histórica para as formulações de Gibbs e Shannon discutidas ao longo deste trabalho. Sob essa condição restrita, a entropia passa a depender exclusivamente do número total de microestados. Essa particularidade permite simplificar o modelo e representar a entropia de Boltzmann matematicamente através de uma função real de uma variável escalar $f(x)$.

A redução do conceito entrópico para o estudo de uma função real foi o verdadeiro ponto de partida desta pesquisa. Essa abordagem inicial funcionou como uma ponte conceitual para o desenvolvimento de todo o trabalho e teve seus primeiros resultados consolidados em um artigo publicado na Revista Eletrônica Paulista de Matemática [4]. Ao adotar essa representação focada no núcleo do comportamento empírico da entropia, exige-se que a função f satisfaça as duas propriedades matemáticas descritas a seguir.

- i) Se $x < y \Leftrightarrow f(x) < f(y)$ (f é estritamente crescente);
- ii) Se $f(x \cdot y) = f(x) + f(y)$ (f é aditiva).

O conjunto de todas as funções que satisfazem essas propriedades é definido por

$$\mathcal{B} = \{f \mid f \text{ satisfaz (i) e (ii)}\}. \quad (1.1)$$

A partir desta caracterização, emergem naturalmente questões profundas sobre a natureza deste conjunto:

- Qual o domínio natural das funções que pertencem ao conjunto $\mathcal{B}$?
- As funções em $\mathcal{B}$ são contínuas? Diferenciáveis? Analíticas?
- As funções em $\mathcal{B}$ são de classe C^r ? E de classe C^∞ ?
- O conjunto $\mathcal{B}$ é aberto? Fechado? Convexo? Compacto? Conexo?
- Quem são as funções que moram no conjunto $\mathcal{B}$?

A partir da caracterização inicial do conjunto $\mathcal{B}$, o presente trabalho dedica-se a investigar sistematicamente sua estrutura topológica e analítica. Será demonstrado que as propriedades impostas são tão restritivas que determinam completamente a forma funcional das funções que o compõem, revelando assim o núcleo matemático que fundamenta o conceito de entropia.

Além dessa caracterização intrínseca, o estudo busca situar o conjunto $\mathcal{B}$ no contexto mais amplo das teorias de entropia, estabelecendo conexões explícitas com formulações clássicas e contemporâneas. Em particular, propõe-se explorar de que maneira a estrutura desenvolvida para $\mathcal{B}$ serve como base unificadora para diferentes manifestações entrópicas, com especial atenção à entropia de Shannon [7].

Esta abordagem permite identificar as propriedades matemáticas essenciais compartilhadas por diversas definições de entropia, contribuindo para uma compreensão mais integrada e profunda de sua natureza universal. O presente texto, intitulado “*Estudos das Funções Entrópicas*”, visa compilar e organizar de forma didática essas investigações, reunindo os fundamentos analíticos, topológicos e probabilísticos necessários para tal empreendimento. Assim, busca-se não apenas apresentar resultados, mas também construir um material de referência estruturado, que possa servir como ponto de partida e estudo para aqueles interessados nos fundamentos matemáticos da entropia.

2. CONHECIMENTOS PRÉVIOS

Este capítulo tem como objetivo revisar os conceitos matemáticos que serão utilizados ao longo deste trabalho. A investigação do conjunto das funções de Boltzmann $\mathcal{B}$, definido no Capítulo 1, exige uma base sólida em análise real, topologia e espaços de funções para que se possa estudar suas propriedades de regularidade e estrutura global.

Além disso, para situar $\mathcal{B}$ no panorama mais amplo das teorias de entropia, em particular, para estabelecer sua conexão com a entropia de Shannon, objeto do Capítulo 4, é fundamental revisar conceitos de teoria da probabilidade. Tópicos como variáveis aleatórias, independência, distribuições diádicas e a relação de refinamento fornecerão a linguagem necessária para compreender e comparar a estrutura de $\mathcal{B}$ com a formulação informacional da entropia.

A seguir, apresentam-se de forma organizada e autocontida as ferramentas necessárias para a compreensão dos resultados desenvolvidos nos capítulos subsequentes. A escolha dos tópicos aqui abordados foi guiada por sua relevância direta tanto para a caracterização matemática de $\mathcal{B}$ quanto para a interpretação de suas propriedades no contexto da física estatística e da teoria da informação.

2.1 CONCEITOS DE ANÁLISE REAL

Nesta seção, são listados os conceitos fundamentais de análise real que serão frequentemente invocados ao longo deste trabalho. Tópicos como sequências, séries, limites são essenciais para a compreensão das propriedades de regularidade das funções de Boltzmann, bem como para a demonstração de resultados centrais como a continuidade e diferenciabilidade de tais funções. Além disso, ferramentas como o Princípio da Indução Matemática e a Série de Taylor serão utilizadas para estabelecer propriedades algébricas e analíticas do conjunto $\mathcal{B}$.

2.1.1 SEQUÊNCIAS E SÉRIES

Definição 1. *Uma sequência de números reais é uma função $x : \mathbb{N} \rightarrow \mathbb{R}$ que associa a cada número natural n um número real x_n . Denotamos uma sequência por $(x_1, x_2, x_3, \dots)$ ou simplesmente (x_n) . Cada x_n é chamado termo da sequência.*

Considere uma sequência como uma “lista ordenada infinita” de números. Cada número natural n (1, 2, 3, ...) está associado a um número real x_n que é o n -ésimo termo da sequência.

A notação (x_n) é mais compacta e enfatiza que o interesse recai sobre o comportamento dos termos x_n quando n torna-se muito grande, quando $n \rightarrow \infty$.

Exemplo 1. A sequência $(1, \frac{1}{2}, \frac{1}{3}, \frac{1}{4}, \dots)$ pode ser descrita pela função $x(n) = \frac{1}{n}$. Aqui

$$x_1 = 1, x_2 = \frac{1}{2}, x_3 = \frac{1}{3},$$

e assim por diante.

Exemplo 2. A sequência $(2, 4, 6, 8, \dots)$ dos números pares positivos é dada por $x(n) = 2n$. Note que esta sequência cresce indefinidamente.

Exemplo 3. A sequência constante $(5, 5, 5, 5, \dots)$ tem $x(n) = 5$ para todo $n \in \mathbb{N}$. Todos os termos são iguais.

Definição 2. Uma sequência (x_n) converge para o limite $L \in \mathbb{R}$, escrevemos $\lim_{n \rightarrow \infty} x_n = L$, se para todo $\epsilon > 0$ existe $n_0 \in \mathbb{N}$ tal que $|x_n - L| < \epsilon$ para todo $n \geq n_0$.

Intuitivamente, uma sequência converge para L se seus termos ficam “arbitrariamente próximos” de L a partir de um certo índice. O ϵ representa o “erro máximo permitido” e n_0 é o “índice a partir do qual” todos os termos estão dentro dessa margem de erro. Numa linguagem mais simples, dizer que $\lim_{n \rightarrow \infty} x_n = L$ significa que “para n suficientemente grande, x_n fica tão próximo de L quanto se queira”. A definição formal garante que não importa quão pequeno seja o $\epsilon > 0$ que for escolhido (0.1, 0.01, 0.001, ...), sempre existirá uma posição $n_0 > 0$ na sequência a partir da qual todos os termos estarão dentro da faixa $L - \epsilon$ até $L + \epsilon$.

Exemplo 4. Considere a sequência $x_n = \frac{1}{n}$. Ela converge para $L = 0$. Para que os termos fiquem a uma distância menor que $\epsilon = 0.1$ de 0, basta tomar $n_0 = 11$, pois para $n \geq 11$ tem-se $\frac{1}{n} \leq 0.0909 < 0.1$. Por outro lado, se $\epsilon = 0.001$, toma-se $n_0 = 1001$, e assim por diante. Sempre se encontra um n_0 que satisfaz a condição.

Exemplo 5. A sequência constante $x_n = 5$ converge para 5. Aqui, para qualquer $\epsilon > 0$, pode-se tomar $n_0 = 1$, pois $|5 - 5| = 0 < \epsilon$ para todo n .

Exemplo 6. A sequência $x_n = (-1)^n = (1, -1, 1, -1, \dots)$ não converge, pois os termos ficam alternando-se entre 1 e -1, não se estabilizando próximo de um único número.

Definição 3. Dada uma sequência (a_n) , a série numérica associada é a soma infinita

$$\sum_{n=1}^{\infty} a_n = a_1 + a_2 + a_3 + \dots$$

A sequência das somas parciais é definida por $s_n = a_1 + a_2 + \dots + a_n$.

Uma série numérica representa uma tentativa de somar infinitos termos de uma sequência. Como não é possível efetuar essa soma diretamente, estuda-se o comportamento das somas parciais s_n , que correspondem à soma dos primeiros n termos. Se à medida que n aumenta a soma

parcial s_n se aproxima de um valor limite finito, dizemos que a série converge; caso contrário, ela diverge. Essa abordagem permite atribuir um significado preciso a “somas infinitas”.

Exemplo 7. Considere a sequência $a_n = \frac{1}{2^n} = (\frac{1}{2}, \frac{1}{4}, \frac{1}{8}, \frac{1}{16}, \dots)$. A série associada é

$$\sum_{n=1}^{\infty} \frac{1}{2^n} = \frac{1}{2} + \frac{1}{4} + \frac{1}{8} + \frac{1}{16} + \dots$$

As somas parciais são

$$\begin{aligned} s_1 &= \frac{1}{2} = 0.5; \\ s_2 &= \frac{1}{2} + \frac{1}{4} = 0.75; \\ s_3 &= \frac{1}{2} + \frac{1}{4} + \frac{1}{8} = 0.875; \\ s_4 &= \frac{1}{2} + \frac{1}{4} + \frac{1}{8} + \frac{1}{16} = 0.9375. \end{aligned}$$

Exemplo 8. Para a sequência $a_n = 1 = (1, 1, 1, 1, \dots)$, a série é

$$\sum_{n=1}^{\infty} 1 = 1 + 1 + 1 + 1 + \dots$$

As somas parciais são $s_n = n$: $s_1 = 1$, $s_2 = 2$, $s_3 = 3$, etc. Esta soma cresce indefinidamente.

Definição 4. Dizemos que a série $\sum_{n=1}^{\infty} a_n$ converge se a sequência das somas parciais (s_n) converge. Caso contrário, a série diverge. Quando converge, o valor da série é o limite

$$\lim_{n \rightarrow \infty} s_n.$$

Teorema 1. Se a série $\sum_{n=1}^{\infty} a_n$ converge, então $\lim_{n \rightarrow \infty} a_n = 0$.

Demonstração. Seja $s_n = a_1 + \dots + a_n$. Se $\sum a_n$ converge, então (s_n) converge para algum L . Note que $a_n = s_n - s_{n-1}$, logo $\lim_{n \rightarrow \infty} a_n = L - L = 0$. $\square$

Exemplo 9 (Série geométrica). Considere a série $\sum_{n=0}^{\infty} r^n$. Sua soma parcial é

$$s_n = 1 + r + r^2 + \dots + r^n = \frac{1 - r^{n+1}}{1 - r}, \quad \text{para } r \neq 1.$$

Se $|r| < 1$, então $\lim_{n \rightarrow \infty} r^{n+1} = 0$, logo

$$\lim_{n \rightarrow \infty} s_n = \frac{1}{1 - r}.$$

Portanto, a série geométrica $\sum_{n=0}^{\infty} r^n$ converge para $\frac{1}{1-r}$ quando $|r| < 1$, e diverge quando $|r| \geq 1$.

Exemplo 10 (Série geométrica alternada). *Considere a série $\sum_{n=0}^{\infty} (-1)^n r^n$. Esta é uma série geométrica com razão $-r$. Pelo exemplo anterior, se $|-r| = |r| < 1$, a série converge para*

$$\sum_{n=0}^{\infty} (-1)^n r^n = \frac{1}{1 - (-r)} = \frac{1}{1 + r}.$$

Teorema 2 (Teste da Razão). *Seja $\sum a_n$ uma série com $a_n \neq 0$. Se existir o limite*

$$L = \lim_{n \rightarrow \infty} \left| \frac{a_{n+1}}{a_n} \right|,$$

então:

- *Se $L < 1$, a série converge absolutamente;*
- *Se $L > 1$, a série diverge;*
- *Se $L = 1$, o teste é inconclusivo.*

Demonstração. Caso 1: $L < 1$. Escolha r tal que $L < r < 1$. Existe $n_0 \in \mathbb{N}$ tal que para todo $n \geq n_0$

$$\left| \frac{a_{n+1}}{a_n} \right| < r.$$

Então

$$\begin{aligned} |a_{n_0+1}| &< r|a_{n_0}|, \\ |a_{n_0+2}| &< r|a_{n_0+1}| < r^2|a_{n_0}|, \\ &\vdots \\ |a_{n_0+k}| &< r^k|a_{n_0}|. \end{aligned}$$

A série $\sum |a_{n_0+k}| r^k$ converge (pois é uma série geométrica com $r < 1$). Como $0 \leq |a_{n_0+k}| < |a_{n_0}| r^k$, segue pelo Teste da Comparação que a série $\sum |a_{n_0+k}|$ também converge. Consequentemente, a série $\sum |a_n|$ converge, o que significa que a série original converge absolutamente.

Caso 2: $L > 1$. Existe n_0 tal que para $n \geq n_0$

$$\left| \frac{a_{n+1}}{a_n} \right| > 1,$$

logo $|a_{n+1}| > |a_n|$. Portanto, $|a_n|$ não tende a zero, e a série diverge.

Caso 3: $L = 1$. Considere $\sum \frac{1}{n}$ (diverge) e $\sum \frac{1}{n^2}$ (converge). Em ambos, $L = 1$. □

Definição 5 (Raio de Convergência). *Para uma série de potências a qual, por definição, tem a forma $\sum_{n=0}^{\infty} c_n(x-a)^n$, o raio de convergência R é o número tal que a série converge quando $|x-a| < R$ e diverge quando $|x-a| > R$.*

Quando o limite existe, o raio de convergência pode ser calculado por

$$R = \lim_{n \rightarrow \infty} \left| \frac{c_n}{c_{n+1}} \right|.$$

Exemplo 11. Considere a série $\sum_{n=1}^{\infty} \frac{x^n}{n}$. Pelo teste da razão

$$\lim_{n \rightarrow \infty} \left| \frac{a_{n+1}}{a_n} \right| = \lim_{n \rightarrow \infty} \left| \frac{x^{n+1}/(n+1)}{x^n/n} \right| = |x|.$$

Portanto, a série converge se $|x| < 1$ e diverge se $|x| > 1$. O raio de convergência é $R = 1$.

Definição 6. O número e é definido pela série convergente

$$e = \sum_{n=0}^{\infty} \frac{1}{n!} = 1 + 1 + \frac{1}{2!} + \frac{1}{3!} + \cdots.$$

Teorema 3 (Limite fundamental).

$$\lim_{n \rightarrow \infty} \left(1 + \frac{1}{n} \right)^n = e.$$

Demonstração via série binomial. Pela fórmula do binômio de Newton

$$\begin{aligned} \left(1 + \frac{1}{n} \right)^n &= \sum_{k=0}^n \binom{n}{k} \frac{1}{n^k} \\ &= \sum_{k=0}^n \frac{1}{k!} \left(1 - \frac{1}{n} \right) \left(1 - \frac{2}{n} \right) \cdots \left(1 - \frac{k-1}{n} \right). \end{aligned}$$

Para cada k fixo, quando $n \rightarrow \infty$

$$\left(1 - \frac{1}{n} \right) \left(1 - \frac{2}{n} \right) \cdots \left(1 - \frac{k-1}{n} \right) \rightarrow 1.$$

Portanto,

$$\lim_{n \rightarrow \infty} \left(1 + \frac{1}{n} \right)^n = \sum_{k=0}^{\infty} \frac{1}{k!} = e.$$

□

Corolário 1 (Limite exponencial).

$$\lim_{x \rightarrow 0} \frac{e^x - 1}{x} = 1.$$

Demonstração. Pelo Teorema do Limite Fundamental, tem-se $\lim_{n \rightarrow \infty} \left(1 + \frac{1}{n} \right)^n = e$. Fazendo a substituição $x = \frac{1}{n}$, quando $n \rightarrow \infty$ tem-se $x \rightarrow 0^+$, logo

$$e = \lim_{x \rightarrow 0^+} (1 + x)^{1/x}.$$

Tomando logaritmo natural em ambos os lados

$$1 = \lim_{x \rightarrow 0^+} \frac{\ln(1+x)}{x}.$$

Fazendo a mudança $y = \ln(1+x)$, ou equivalentemente $x = e^y - 1$, obtém-se

$$1 = \lim_{y \rightarrow 0^+} \frac{y}{e^y - 1} \Rightarrow \lim_{y \rightarrow 0^+} \frac{e^y - 1}{y} = 1.$$

Um argumento similar vale para $x \rightarrow 0^-$. Quando $x \rightarrow 0$, obtém-se

$$\lim_{x \rightarrow 0} \frac{e^x - 1}{x} = 1.$$

□

Corolário 2 (Limite de raízes n-ésimas). *Para todo $a > 0$, tem-se $\lim_{n \rightarrow \infty} a^{1/n} = 1$.*

Demonstração. Pelo Teorema do Limite Fundamental, $\lim_{n \rightarrow \infty} (1 + \frac{1}{n})^n = e$. Para $a > 0$, tem-se $a^{1/n} = e^{\frac{\ln a}{n}} \rightarrow e^0 = 1$. □

2.1.2 LIMITES E CONTINUIDADE

Definição 7. *Um ponto $a \in \mathbb{R}$ é ponto de acumulação de $X \subset \mathbb{R}$ se todo intervalo aberto contendo a contém algum ponto de X diferente do próprio a . Isto é, para todo $\epsilon > 0$, existe $x \in X$ com $0 < |x - a| < \epsilon$.*

Definição 8. *Seja $f : D \subset \mathbb{R} \rightarrow \mathbb{R}$ e a um ponto de acumulação de D . Dizemos que $\lim_{x \rightarrow a} f(x) = L$ se para todo $\epsilon > 0$ existe $\delta > 0$ tal que se $0 < |x - a| < \delta$ então $|f(x) - L| < \epsilon$.*

Esta definição formaliza a ideia intuitiva de que $f(x)$ fica arbitrariamente próximo de L quando x se aproxima de a . O ϵ controla quão próximo deseja-se que $f(x)$ esteja de L , e o δ controla quão próximo x precisa estar de a para garantir isso.

Exemplo 12. *Se $f(x) = 2x + 1$, então $\lim_{x \rightarrow 2} f(x) = 5$.*

Solução. Dado $\epsilon > 0$, busca-se encontrar $\delta > 0$ tal que se $0 < |x - 2| < \delta$ então $|(2x+1)-5| < \epsilon$. Note que $|(2x+1)-5| = |2x-4| = 2|x-2|$. Portanto, se for tomado $\delta = \frac{\epsilon}{2}$, tem-se

$$0 < |x - 2| < \delta \Rightarrow |f(x) - 5| = 2|x - 2| < 2 \cdot \frac{\epsilon}{2} = \epsilon.$$

Definição 9. *Uma função f é contínua em a se*

1. $a \in D(f)$;
2. $\lim_{x \rightarrow a} f(x)$ existe;
3. $\lim_{x \rightarrow a} f(x) = f(a)$.

Intuitivamente, uma função é contínua em a se não há “saltos” ou “buracos” no gráfico quando passamos por $x = a$. Pode-se desenhar o gráfico sem levantar o lápis do papel nesse ponto.

Exemplo 13. Se $f(x) = x^2$, então f é contínua para todo $a \in \mathbb{R}$. Para qualquer $a \in \mathbb{R}$, tem-se

$$\lim_{x \rightarrow a} x^2 = a^2 = f(a).$$

Solução. Dado $\epsilon > 0$, busca-se encontrar $\delta > 0$ tal que se $|x - a| < \delta$, então $|x^2 - a^2| < \epsilon$.

Note que

$$|x^2 - a^2| = |x - a||x + a|.$$

Pode-se controlar $|x + a|$ observando que se $|x - a| < 1$, então $|x| < |a| + 1$, logo $|x + a| \leq |x| + |a| < 2|a| + 1$.

Escolhendo $\delta = \frac{\epsilon}{2|a|+1}$, se $|x - a| < \delta$, então

$$|x^2 - a^2| = |x - a||x + a| < \delta(2|a| + 1) = \frac{\epsilon}{2|a|+1}(2|a| + 1) = \epsilon.$$

Portanto, f é contínua em a para todo $a \in \mathbb{R}$.

Exemplo 14. A função $g(x) = \frac{1}{x}$ é contínua em $\mathbb{R} \setminus \{0\}$. Em $x = 0$ a função não está definida, portanto não pode ser contínua.

Teorema 4 (Critério de continuidade via sequências). Uma função f é contínua em a se, e somente se, para toda sequência (x_n) com $x_n \rightarrow a$ tem-se $f(x_n) \rightarrow f(a)$.

Demonstração. A equivalência será demonstrada em duas partes

Parte 1: Se f é contínua em a e $x_n \rightarrow a$, então $f(x_n) \rightarrow f(a)$. Por hipótese, f é contínua em a , logo para todo $\epsilon > 0$ existe $\delta > 0$ tal que se $|x - a| < \delta$ então $|f(x) - f(a)| < \epsilon$. Como $x_n \rightarrow a$, para este δ existe $n_0 \in \mathbb{N}$ tal que para todo $n \geq n_0$ tem-se $|x_n - a| < \delta$. Combinando as duas afirmações: para $n \geq n_0$ tem-se $|x_n - a| < \delta$, logo $|f(x_n) - f(a)| < \epsilon$. Portanto, $f(x_n) \rightarrow f(a)$.

Parte 2: Se para toda sequência (x_n) com $x_n \rightarrow a$ vale $f(x_n) \rightarrow f(a)$, então f é contínua em a . Prova-se por contradição. Suponha que f não seja contínua em a . Então existe $\epsilon > 0$ tal que para todo $\delta > 0$ existe algum x com $|x - a| < \delta$ mas $|f(x) - f(a)| \geq \epsilon$. Em particular, para cada $n \in \mathbb{N}$, tomando $\delta = \frac{1}{n}$, existe x_n tal que

$$0 < |x_n - a| < \frac{1}{n} \quad \text{mas} \quad |f(x_n) - f(a)| \geq \epsilon.$$

Constrói-se assim uma sequência (x_n) com $x_n \rightarrow a$, pois $|x_n - a| < \frac{1}{n} \rightarrow 0$. Porém, $f(x_n) \not\rightarrow f(a)$, pois $|f(x_n) - f(a)| \geq \epsilon$ para todo n . Isso contradiz a hipótese de que para toda sequência convergindo para a tem-se $f(x_n) \rightarrow f(a)$. Como a negação leva a contradição, conclui-se que f deve ser contínua em a . $\square$

Definição 10. Uma função $f : D \rightarrow \mathbb{R}$ é monótona crescente se para todos $x, y \in D$ com $x < y$ tem-se $f(x) \leq f(y)$. Se $f(x) < f(y)$ para $x < y$, dizemos que f é estritamente crescente.

Observação 1. Funções monótonas preservam a ordem, mais precisamente, se os valores de x aumentam, os valores de $f(x)$ também aumentam, ou pelo menos não diminuem.

2.1.3 CÁLCULO DIFERENCIAL

Definição 11. Uma função $f : D \subset \mathbb{R} \rightarrow \mathbb{R}$ é diferenciável em $a \in D$ se existe o limite

$$f'(a) = \lim_{h \rightarrow 0} \frac{f(a+h) - f(a)}{h}.$$

Quando este limite existe, dizemos que $f'(a)$ é a derivada de f em a .

A derivada representa a taxa de variação instantânea da função. Geometricamente, ela corresponde ao coeficiente angular da reta tangente ao gráfico de f no ponto $(a, f(a))$. Como a derivada é definida formalmente por meio de um limite, sua existência exige que a função seja contínua no ponto, de acordo com os critérios estabelecidos no Teorema 4. Contudo, a continuidade por si só não garante a diferenciabilidade, sendo estritamente necessário que o limite da razão incremental exista. Para que isso ocorra, os limites laterais, representados pelas aproximações pela direita ($h \rightarrow 0^+$) e pela esquerda ($h \rightarrow 0^-$), devem resultar no mesmo valor. Caso esses limites laterais sejam distintos, conclui-se que a função não é diferenciável no ponto em questão.

Exemplo 15. A função $f(x) = x^2$ é diferenciável em todo $\mathbb{R}$.

Solução.

$$f'(x) = \lim_{h \rightarrow 0} \frac{(x+h)^2 - x^2}{h} = \lim_{h \rightarrow 0} \frac{2xh + h^2}{h} = \lim_{h \rightarrow 0} (2x + h) = 2x.$$

Exemplo 16. A função $g(x) = |x|$ não é diferenciável em $x = 0$, pois os limites laterais são diferentes.

Solução. pois,

$$\lim_{h \rightarrow 0^+} \frac{|h|}{h} = 1 \quad e \quad \lim_{h \rightarrow 0^-} \frac{|h|}{h} = -1.$$

Definição 12. Seja c um ponto no domínio de uma função f . Dizemos que c é um ponto crítico de f se $f'(c) = 0$ ou se a derivada $f'(c)$ não existir.

Essa definição consolida os locais de maior interesse na análise do comportamento de uma função. A primeira condição identifica os instantes em que a reta tangente é horizontal, o que auxilia diretamente na localização de picos e vales no gráfico. A segunda condição incorpora os casos de não diferenciabilidade recém-discutidos, nos quais a mudança abrupta da taxa de variação também configura um ponto analítico relevante.

Pode-se ilustrar essa classificação retornando aos casos anteriores. No Exemplo 15, a função $f(x) = x^2$ possui derivada $f'(x) = 2x$. Ao igualar essa expressão a zero, encontra-se o ponto crítico $x = 0$, que corresponde ao vértice da parábola. Já no Exemplo 16, foi demonstrado que a função $g(x) = |x|$ não é diferenciável na origem. Por se enquadrar na segunda exigência da definição, o ponto $x = 0$ também é formalmente classificado como um ponto crítico para a função módulo.

Teorema 5 (Princípio da Indução Matemática). *Seja $P(n)$ uma propriedade sobre os números naturais. Se,*

1. $P(1)$ é verdadeira, e

2. Para todo $k \in \mathbb{N}$, se $P(k)$ é verdadeira então $P(k+1)$ é verdadeira,

então $P(n)$ é verdadeira para todo $n \in \mathbb{N}$.

Demonstração. Suponha que as condições (1) e (2) valham mas exista algum n para o qual $P(n)$ é falsa. Seja m o menor tal número. Por (1), $m > 1$. Então $P(m-1)$ é verdadeira, mas por (2), $P(m)$ também seria verdadeira, contradição. $\square$

Exemplo 17. Para todo $n \in \mathbb{N}$, tem-se, $1 + 2 + \cdots + n = \frac{n(n+1)}{2}$.

Solução. Para $n = 1$ tem-se $1 = \frac{1(1+1)}{2} = 1$. Suponha válido para k . Para $k+1$

$$\begin{aligned} 1 + 2 + \cdots + k + (k+1) &= \frac{k(k+1)}{2} + (k+1) \\ &= \frac{k(k+1) + 2(k+1)}{2} \\ &= \frac{(k+1)(k+2)}{2}. \end{aligned}$$

A fórmula vale para $k+1$, portanto, pelo Princípio da Indução, vale para todo $n \in \mathbb{N}$.

Teorema 6 (Homogeneidade racional). *Se $f : \mathbb{R} \rightarrow \mathbb{R}$ é uma função que tem a propriedade $f(x+y) = f(x) + f(y)$ para todos $x, y \in \mathbb{R}$ e é contínua, então*

$$f(qx) = qf(x).$$

para todo $q \in \mathbb{Q}$, $x \in \mathbb{R}$.

Demonstração. Para $n \in \mathbb{N}$, prova-se por indução que $f(nx) = nf(x)$. Para $n = 1$, trivialmente $f(1 \cdot x) = f(x) = 1 \cdot f(x)$. Suponha $f(kx) = kf(x)$. Então, para $k+1$, tem-se

$$\begin{aligned} f((k+1)x) &= f(kx+x) \\ &= f(kx) + f(x) \\ &= kf(x) + f(x) \\ &= (k+1)f(x). \end{aligned}$$

Para $f(0)$, tem-se que $f(0) = f(0+0) = f(0) + f(0)$, logo $f(0) = 2f(0)$, o que implica $f(0) = 0$. Para inteiros negativos, $f(-x) = -f(x)$, pois

$$\begin{aligned} f(x) + f(-x) &= f(x-x) \\ &= f(0) \\ &= 0. \end{aligned}$$

Estendendo para $q = \frac{m}{n} \in \mathbb{Q}$ com $m \in \mathbb{Z}$, $n \in \mathbb{N}$. Seja $y = f\left(\frac{x}{n}\right)$, tem-se

$$\begin{aligned} f(x) &= f\left(n \cdot \frac{x}{n}\right) \\ &= nf\left(\frac{x}{n}\right) \\ &= ny \Rightarrow y = \frac{f(x)}{n}. \end{aligned}$$

Logo,

$$\begin{aligned} f\left(\frac{m}{n}x\right) &= f\left(m \cdot \frac{x}{n}\right) \\ &= mf\left(\frac{x}{n}\right) \\ &= m \cdot \frac{f(x)}{n} \\ &= \frac{m}{n}f(x). \end{aligned}$$

□

Definição 13 (Série de Taylor). *Se f é uma função infinitamente diferenciável em a , a série de Taylor de f em torno de a é dada por*

$$\sum_{n=0}^{\infty} \frac{f^{(n)}(a)}{n!} (x-a)^n,$$

onde $f^{(n)}(a)$ denota a n -ésima derivada de f em a .

Exemplo 18. A série de Taylor da função exponencial e^x em torno de 0 é

$$e^x = \sum_{n=0}^{\infty} \frac{x^n}{n!} = 1 + x + \frac{x^2}{2!} + \frac{x^3}{3!} + \cdots.$$

Definição 14. Uma função $f : D \rightarrow \mathbb{R}$ é de classe C^r , $r \geq 0$, se todas as suas derivadas até ordem r existem e são contínuas em D . Em particular:

- C^0 : funções contínuas;
- C^1 : funções continuamente diferenciáveis;
- C^∞ : funções infinitamente diferenciáveis, suaves.

Funções de classe C^∞ são “suaves”, não têm cantos, bicos ou mudanças bruscas de curvatura. A maioria das funções elementares, polinômios, exponenciais, senos, cossenos, são de classe C^∞ .

Exemplo 19. $f(x) = e^x$ é C^∞ em $\mathbb{R}$, pois todas as suas derivadas existem e são contínuas, $f^{(n)}(x) = e^x$ para todo n .

Exemplo 20. $g(x) = x|x|$ é C^1 mas não C^2 em $\mathbb{R}$, pois sua segunda derivada não é contínua em $x = 0$.

2.2 TOPOLOGIA DOS REAIS

A topologia dos números reais desempenha um papel crucial na análise do conjunto $\mathcal{B}$ como um subconjunto de um espaço de funções. Nesta seção, revisam-se noções como conjuntos abertos, fechados, compactos, convexos e conexos, que serão utilizadas para investigar a estrutura topológica de $\mathcal{B}$ no Capítulo 3. Propriedades como convexidade e conexidade, em particular, serão demonstradas para $\mathcal{B}$, revelando aspectos importantes de sua organização como espaço funcional.

2.2.1 CONJUNTOS FUNDAMENTAIS

Definição 15. Um conjunto $A \subset \mathbb{R}$ é aberto se para todo $x \in A$ existe $\epsilon > 0$ tal que o intervalo $(x - \epsilon, x + \epsilon)$ está contido em A .

Observação 2. Intuitivamente, um conjunto é aberto se todos os seus pontos são “pontos interiores” em torno de cada ponto pode-se encontrar uma pequena vizinhança totalmente contida no conjunto.

Exemplo 21. O conjunto $\mathbb{R}$ é aberto. Para qualquer $x \in \mathbb{R}$, pode-se tomar $\epsilon = 1$, ou qualquer outro número positivo, e tem-se $(x - 1, x + 1) \subset \mathbb{R}$.

Exemplo 22. O conjunto vazio $\emptyset$ é aberto por vacuidade, não há pontos para verificar a condição.

Exemplo 23. O intervalo $[0, 1]$ não é aberto. Para $x = 0$, qualquer $\epsilon > 0$ faz com que $(-\epsilon, \epsilon)$ contenha pontos negativos que não pertencem a $[0, 1]$.

Definição 16. Um conjunto $F \subset \mathbb{R}$ é fechado se seu complementar $\mathbb{R} \setminus F$ é aberto.

Observação 3. Equivalentemente, F é fechado se contém todos os seus pontos de acumulação. Um ponto de acumulação é um ponto que pode ser “aproximado arbitrariamente” por outros pontos do conjunto. Um conjunto pode ser aberto, fechado, ambos ou nenhum dos dois.

Exemplo 24. O intervalo $[0, 1]$ é fechado. Seu complementar é $(-\infty, 0) \cup (1, \infty)$, é aberto.

Exemplo 25. O conjunto $\{0\} \cup \{1/n : n \in \mathbb{N}\}$ é fechado, pois contém seu único ponto de acumulação 0.

Definição 17. Um conjunto $C \subset \mathbb{R}$ é convexo se para todos $x, y \in C$ e $t \in [0, 1]$, tem-se $(1 - t)x + ty \in C$.

Geometricamente, um conjunto é convexo se o segmento de reta ligando quaisquer dois de seus pontos está totalmente contido no conjunto.

Exemplo 26. Todo intervalo da reta real é convexo. Os intervalos $[a, b]$, (a, b) , $[a, b)$, etc, são convexos.

Exemplo 27. O conjunto $\mathbb{Q}$ não é convexo, pois existem números racionais x, y tais que o segmento $[x, y]$ contém números irracionais.

Definição 18. Um conjunto $K \subset \mathbb{R}$ é compacto se é fechado e limitado.

Observação 4. Em $\mathbb{R}$, compacto significa “fechado e limitado”. Esta é uma das caracterizações mais importantes da reta real, Teorema de Heine-Borel ou Teorema de Borel-Lebesgue [5].

Exemplo 28. O intervalo $[0, 1]$ é compacto, pois é fechado e limitado.

Exemplo 29. O intervalo $(0, 1)$ não é compacto, visto que é limitado mas não é fechado.

Definição 19. Um conjunto $X \subset \mathbb{R}$ é conexo se não pode ser escrito como união de dois abertos disjuntos e não vazios.

Intuitivamente, um conjunto é conexo se é “feito de uma única peça”, não pode ser separado em duas partes que não se tocam.

Exemplo 30. O intervalo $[0, 1]$ é conexo. Não é possível separá-lo em dois abertos disjuntos.

Definição 20. Um conjunto $X \subset \mathbb{R}$ é conexo por caminhos se para quaisquer $x, y \in X$ existe uma função contínua $\gamma : [0, 1] \rightarrow X$ tal que $\gamma(0) = x$ e $\gamma(1) = y$.

Observação 5. Conexo por caminhos é uma noção mais forte que conexo. Significa que é possível “caminhar” de qualquer ponto a qualquer outro ponto dentro do conjunto.

Exemplo 31. Todo intervalo da reta é conexo por caminhos. Pode-se sempre usar o segmento de reta $\gamma(t) = (1 - t)x + ty$.

2.2.2 PROPRIEDADES TOPOLÓGICAS

Definição 21. Seja $A \subset \mathbb{R}$ não vazio e limitado superiormente. O supremo de A , denotado $\sup A$, é a menor das cotas superiores de A .

Definição 22. Seja $A \subset \mathbb{R}$ não vazio e limitado inferiormente. O ínfimo de A , denotado $\inf A$, é a maior das cotas inferiores de A .

Observação 6. O supremo, ou ínfimo, pode ou não pertencer ao conjunto. Se pertence é o máximo, ou mínimo, do conjunto.

Exemplo 32. Para $A = (0, 1)$, tem-se $\sup A = 1$ e $\inf A = 0$. Note que $1 \notin A$ e $0 \notin A$.

Exemplo 33. Para $B = [0, 1]$, tem-se $\sup B = 1 \in B$ e $\inf B = 0 \in B$.

2.3 ESPAÇOS DE FUNÇÕES

Nesta seção, introduz-se a estrutura dos espaços de funções contínuas e suaves, que constituem o ambiente natural para o estudo do conjunto $\mathcal{B}$. Conceitos como convergência uniforme, topologias em espaços funcionais e densidade de subconjuntos serão fundamentais para analisar propriedades topológicas de $\mathcal{B}$ e compreender seu comportamento sob limites e aproximações. Essas ferramentas permitirão investigar, no Capítulo 3, questões sobre a estrutura global do conjunto das funções de Boltzmann.

2.3.1 ESTRUTURAS DE ESPAÇOS DE FUNÇÕES

Definição 23. O conjunto de todas as funções contínuas $f : (0, \infty) \rightarrow \mathbb{R}$ é denotado por $C((0, \infty), \mathbb{R})$.

Observação 7. Este conjunto forma um espaço vetorial sobre $\mathbb{R}$. Pode-se somar funções contínuas e multiplicá-las por escalares reais, e o resultado permanece contínuo.

A notação C^0 enfatiza que consideramos funções contínuas, derivada de ordem zero. O conjunto das funções suaves, infinitamente diferenciáveis, é denotado por $C^\infty((0, \infty), \mathbb{R})$.

Exemplo 34. As funções $f(x) = x^2$, $g(x) = \sin x$ e $h(x) = e^{-x}$ pertencem a $C((0, \infty), \mathbb{R})$.

Exemplo 35. A função $k(x) = \frac{1}{x}$ pertence a $C((0, \infty), \mathbb{R})$, pois é contínua em todo seu domínio.

2.3.2 TOPOLOGIAS EM ESPAÇOS DE FUNÇÕES

Definição 24. Uma sequência de funções, f_n , converge uniformemente para f em X se para todo $\epsilon > 0$ existe $n_0 \in \mathbb{N}$ tal que $|f_n(x) - f(x)| < \epsilon$ para todo $n \geq n_0$ e todo $x \in X$.

A palavra “uniformemente” é crucial aqui. Significa que a convergência acontece “do mesmo jeito” em todos os pontos do domínio simultaneamente. Escolhendo uma “tolerância de erro”, um $\epsilon = 0,01$ ou $\epsilon = 0,001$, a convergência uniforme garante que existe um índice $n_0 > 0$ que funciona para todos os pontos $x \in X$ ao mesmo tempo. Ou seja, partir de n_0 , todas as funções f_n estarão dentro da faixa $[f(x) - \epsilon, f(x) + \epsilon]$, para todo $x \in D(f)$. Isso é muito mais forte que convergência pontual, onde cada ponto pode ter seu próprio n_0 .

Exemplo 36. Considere $f_n(x) = \frac{x}{n}$ em $X = [0, 1]$. Deseja-se mostrar que $f_n \rightarrow 0$ uniformemente.

Solução. Para qualquer $x \in [0, 1]$, tem-se

$$|f_n(x) - 0| = \frac{x}{n} \leq \frac{1}{n}.$$

Dado $\epsilon > 0$, escolha n_0 tal que $\frac{1}{n_0} < \epsilon$. Para $n \geq n_0$ e qualquer $x \in [0, 1]$

$$|f_n(x)| \leq \frac{1}{n} \leq \frac{1}{n_0} < \epsilon.$$

A partir do índice n_0 , o gráfico de f_n está inteiramente dentro de uma “faixa” de largura 2ϵ em torno da função zero, independentemente do ponto x considerado.

Exemplo 37. Agora considere $g_n(x) = x^n$ em $X = [0, 1]$. Esta sequência não converge uniformemente para $g(x) = 0$.

Solução. Para x próximo de 0, $x^n \rightarrow 0$ rapidamente, por exemplo $0.1^{10} = 0.0000000001$. Para x próximo de 1, $x^n \rightarrow 0$ muito lentamente, por exemplo $0.99^{100} \approx 0.366$. Fixando-se $\epsilon = 0.1$,

não existe n_0 que funcione para todos os pontos simultaneamente. Para $x = 0.99$ é preciso $n > 229$ para ter $0.99^n < 0.1$. Para $x = 0.999$ é preciso $n > 2301$ para ter $0.999^n < 0.1$. Cada ponto tem sua própria “velocidade” de convergência, portanto não há convergência uniforme.

Definição 25. A topologia da convergência uniforme em compactos em $C((0, \infty), \mathbb{R})$ é definida pela convergência: $f_n \rightarrow f$ se para todo compacto $K \subset (0, \infty)$, $f_n \rightarrow f$ uniformemente em K .

Esta definição resolve um problema importante: Em domínios infinitos, como $(0, \infty)$, a convergência uniforme é muito restritiva. A topologia da convergência uniforme em compactos é um “meio-termo” mais razoável. Um compacto em $(0, \infty)$ é basicamente um intervalo fechado e limitado como $[a, b]$ com $a > 0$. Dizemos que $f_n \rightarrow f$ se para cada intervalo $[a, b] \subset (0, \infty)$, a convergência é uniforme nesse intervalo. Em outras palavras: A convergência é “localmente uniforme”, funciona igualmente bem em cada pedaço finito do domínio. Esta é a topologia natural para estudar funções em domínios não limitados.

Exemplo 38. Considere $f_n(x) = \frac{\sin(nx)}{n}$ em $(0, \infty)$. Mostre que $f_n \rightarrow 0$ na topologia da convergência uniforme em compactos.

Solução. Tome qualquer compacto $K = [a, b] \subset (0, \infty)$. Para $x \in [a, b]$, tem-se

$$|f_n(x) - 0| = \left| \frac{\sin(nx)}{n} \right| \leq \frac{1}{n}.$$

Dado $\epsilon > 0$, escolha n_0 tal que $\frac{1}{n_0} < \epsilon$. Para $n \geq n_0$ e todo $x \in [a, b]$: $|f_n(x)| \leq \frac{1}{n} < \epsilon$. Em cada intervalo $[a, b]$, a convergência é uniforme. Portanto, $f_n \rightarrow 0$ na topologia desejada.

Exemplo 39. Agora considere $g_n(x) = x^n$ em $(0, \infty)$. Esta sequência não converge na topologia da convergência uniforme em compactos.

Solução. Considere o compacto $K = [0.9, 1]$. Para x próximo de 1, por exemplo $x = 0.99$, $x^n \rightarrow 0$ muito lentamente. Para $x = 1$, tem-se $g_n(1) = 1$ para todo n . Não há uma função limite que funcione uniformemente em intervalos que contenham pontos próximos de 1. O comportamento é muito irregular próximo de $x = 1$, portanto não há convergência uniforme em compactos que se aproximem desse ponto.

2.3.3 DENSIDADE E APROXIMAÇÃO

Definição 26. Um conjunto $D \subset \mathbb{R}$ é denso em $\mathbb{R}$ se para todo $x \in \mathbb{R}$ e todo $\epsilon > 0$, existe $d \in D$ tal que $|x - d| < \epsilon$.

Um conjunto é denso se seus pontos estão “espalhados por toda parte”, pode-se aproximar qualquer número real com precisão arbitrária usando elementos do conjunto.

Exemplo 40. $\mathbb{Q}$ é denso em $\mathbb{R}$. Entre quaisquer dois números reais existe um número racional.

Exemplo 41. $\mathbb{Z}$ não é denso em $\mathbb{R}$. Não é possível aproximar $\frac{1}{2}$ com precisão 0.1 usando apenas números inteiros.

Teorema 7 (Densidade dos racionais). *O conjunto $\mathbb{Q}$ dos números racionais é denso em $\mathbb{R}$.*

Demonstração. Seja $x \in \mathbb{R}$ e $\epsilon > 0$ arbitrários. Busca-se encontrar $q \in \mathbb{Q}$ tal que $|x - q| < \epsilon$.

Pela propriedade Arquimediana (veja [5], Capítulo 2, Teorema 2), existe $n \in \mathbb{N}$ tal que $n > \frac{1}{\epsilon}$, ou equivalentemente, $\frac{1}{n} < \epsilon$. Considere o número real nx . Pela definição de função piso, parte inteira, (veja [5], Capítulo 2, Seção 4), existe um único inteiro $m \in \mathbb{Z}$ tal que

$$m \leq nx < m + 1.$$

Este inteiro m é denotado por $\lfloor nx \rfloor$.

Defina $q = \frac{m}{n} = \frac{\lfloor nx \rfloor}{n}$. Claramente $q \in \mathbb{Q}$, pois é razão de dois inteiros. Das desigualdades anterior, tem-se

$$m \leq nx < m + 1.$$

Dividindo por $n > 0$

$$\frac{m}{n} \leq x < \frac{m+1}{n}.$$

Ou seja

$$q \leq x < q + \frac{1}{n}.$$

Portanto

$$0 \leq x - q < \frac{1}{n}.$$

O que implica

$$|x - q| < \frac{1}{n}.$$

Como, $\frac{1}{n} < \epsilon$, logo

$$|x - q| < \frac{1}{n} < \epsilon.$$

Como x e ϵ eram arbitrários, conclui-se que $\mathbb{Q}$ é denso em $\mathbb{R}$. □

Observação 8. *Este teorema é fundamental, significa que é possível aproximar qualquer número real, ou mesmo irracional, com precisão arbitrária usando números racionais.*

Exemplo 42. *Para aproximar $\pi \approx 3.14159\dots$ com erro menor que 0.001, pode-se tomar $\frac{3141}{1000} = 3.141$.*

Exemplo 43. *Para aproximar $\sqrt{2} \approx 1.41421\dots$ com erro menor que 0.01, pode-se tomar $\frac{141}{100} = 1.41$.*

2.4 TEORIA DA PROBABILIDADE

Nesta seção, revisam-se os conceitos de probabilidade que são fundamentais para estabelecer, no Capítulo 4, a relação entre o conjunto das funções de Boltzmann $\mathcal{B}$ e a entropia de Shannon. Introduzimos variáveis aleatórias discretas, probabilidade condicional, independência, distribuições diádicas e a relação de refinamento entre variáveis. Estes conceitos permitirão

entender como a formulação informacional da entropia se relaciona com o conjunto $\mathcal{B}$ quando aplicada a sistemas estatisticamente independentes, conectando a termodinâmica estatística com a teoria da informação.

2.4.1 VARIÁVEIS ALEATÓRIAS DISCRETAS

Definição 27. Uma variável aleatória discreta X é uma função que associa a cada resultado de um espaço amostral Ω um valor numérico em um conjunto finito ou enumerável $\mathcal{X}$. A função de probabilidade de X é definida por

$$p(x) = P(X = x) \quad \text{para } x \in \mathcal{X}.$$

Pode-se interpretar uma variável aleatória como um “medidor” que associa um número a cada resultado possível de um experimento. Por exemplo, ao lançar um dado, pode-se definir uma variável aleatória que associa o número da face superior a cada resultado.

Exemplo 44. Considere o lançamento de uma moeda justa. Pode-se definir uma variável aleatória X onde: $X = 0$ se sair “cara” e $X = 1$ se sair “coroa”

A função de probabilidade é $p(0) = \frac{1}{2}$ e $p(1) = \frac{1}{2}$.

Exemplo 45. No lançamento de um dado honesto de 6 faces, define-se Y como o número da face superior. Então

$$p(1) = p(2) = p(3) = p(4) = p(5) = p(6) = \frac{1}{6}.$$

Definição 28. Dadas duas variáveis aleatórias X e Y , a função de probabilidade conjunta é definida por

$$p(x, y) = P(X = x, Y = y).$$

A probabilidade conjunta permite estudar como duas variáveis aleatórias se relacionam. Ela captura a probabilidade de eventos envolvendo ambas as variáveis simultaneamente.

Exemplo 46. Considere o lançamento de dois dados. Seja X o resultado do primeiro dado e Y do segundo. A probabilidade conjunta $p(x, y)$ para $x, y \in \{1, 2, 3, 4, 5, 6\}$ é

$$p(x, y) = \frac{1}{36} \quad \text{para todo } x, y.$$

2.4.2 PROBABILIDADE CONDICIONAL E INDEPENDÊNCIA

Definição 29. A probabilidade condicional de Y dado X é

$$p(y | x) = \frac{p(x, y)}{p(x)}, \quad \text{desde que } p(x) > 0.$$

A probabilidade condicional mede “o quanto se sabe sobre Y quando se conhece o valor de X ”. É como atualizar as crenças sobre Y com nova informação sobre X .

Exemplo 47. No lançamento de dois dados, qual a probabilidade de Y ser par sabendo que $X = 3$?

Tem-se $p(3, y) = \frac{1}{36}$ para cada y , e $p(3) = \frac{1}{6}$. Os valores pares de Y são 2, 4, 6, então

$$\begin{aligned} p(Y \text{ par} \mid X = 3) &= \frac{p(3, 2) + p(3, 4) + p(3, 6)}{p(3)} \\ &= \frac{3/36}{1/6} \\ &= \frac{1}{2}. \end{aligned}$$

Definição 30. Duas variáveis aleatórias X e Y são independentes se

$$p(x, y) = p(x) \cdot p(y) \quad \text{para todos } x \in \mathcal{X}, y \in \mathcal{Y}.$$

Independência significa que conhecer o valor de uma variável não fornece nenhuma informação sobre a outra. São como “mundos separados” que não interferem entre si.

Exemplo 48. No lançamento de dois dados honestos, X e Y são independentes, pois

$$\begin{aligned} p(x, y) &= \frac{1}{36} \\ &= \frac{1}{6} \cdot \frac{1}{6} \\ &= p(x) \cdot p(y). \end{aligned}$$

Exemplo 49. Considere retirar duas cartas de um baralho sem reposição. Se X é a primeira carta e Y a segunda, então X e Y não são independentes, pois saber X afeta as probabilidades para Y .

Observação 9. Quando X e Y são independentes, a probabilidade condicional se reduz à probabilidade marginal

$$\begin{aligned} p(y \mid x) &= \frac{p(x, y)}{p(x)} \\ &= \frac{p(x)p(y)}{p(x)} \\ &= p(y). \end{aligned}$$

Isso reflete a intuição: Se são independentes, conhecer X não altera a informação sobre Y .

2.4.3 DISTRIBUIÇÕES DIÁDICAS

Definição 31. Uma distribuição de probabilidade $(p_1, p_2, \dots, p_n)$ é chamada de diádica se cada p_i é um número racional da forma $p_i = \frac{m_i}{2^M}$, onde m_i são inteiros não-negativos e M é um inteiro positivo tal que $\sum_{i=1}^n m_i = 2^M$.

As distribuições diádicas são aquelas que podem ser obtidas dividindo um “todo” em 2^M partes iguais e distribuindo essas partes entre os diferentes eventos. O número 2^M aparece naturalmente em contextos de teoria da informação e computação, onde frequentemente se trabalha com bits e potências de 2.

Exemplo 50. A distribuição uniforme $\left(\frac{1}{2}, \frac{1}{2}\right)$ é diádica, pois pode-se escrever $\frac{1}{2} = \frac{1}{2^1}$. Aqui $M = 1$ e $m_1 = m_2 = 1$.

A distribuição $\left(\frac{1}{4}, \frac{1}{4}, \frac{1}{2}\right)$ também é diádica, pois:

$$\begin{aligned}\frac{1}{4} &= \frac{1}{2^2} = \frac{1}{4}, \\ \frac{1}{2} &= \frac{2}{2^2} = \frac{2}{4}.\end{aligned}$$

Tem-se $M = 2$, $m_1 = 1$, $m_2 = 1$, $m_3 = 2$, e $1 + 1 + 2 = 4 = 2^2$.

Exemplo 51. A distribuição $\left(\frac{1}{3}, \frac{2}{3}\right)$ não é diádica, pois não existem inteiros m_1, m_2 e M tais que:

$$\begin{aligned}\frac{m_1}{2^M} &= \frac{1}{3}, \\ \frac{m_2}{2^M} &= \frac{2}{3}, \\ m_1 + m_2 &= 2^M.\end{aligned}$$

Isso exigiria que 2^M fosse múltiplo de 3, o que nunca acontece para potências de 2.

Observação 10. As distribuições diádicas são densas no conjunto de todas as distribuições de probabilidade. Isso significa que qualquer distribuição de probabilidade pode ser aproximada arbitrariamente bem por uma sequência de distribuições diádicas. Esta propriedade será crucial no Capítulo 4 para estender resultados das distribuições diádicas para distribuições arbitrárias usando argumentos de limite.

Exemplo 52. Para aproximar a distribuição $\left(\frac{1}{3}, \frac{2}{3}\right)$ com $M = 3$ ($2^3 = 8$ partes), pode-se tomar:

$$\begin{aligned}p_1^{(3)} &= \frac{2}{8} = 0.25, \\ p_2^{(3)} &= \frac{6}{8} = 0.75.\end{aligned}$$

Com $M = 5$ (32 partes), tem-se uma aproximação melhor:

$$\begin{aligned}p_1^{(5)} &= \frac{10}{32} = 0.3125, \\ p_2^{(5)} &= \frac{22}{32} = 0.6875.\end{aligned}$$

À medida que M aumenta, as aproximações convergem para $\left(\frac{1}{3}, \frac{2}{3}\right)$.

2.4.4 RELAÇÃO DE REFINAMENTO

Definição 32. Dizemos que Y é um refinamento de X se existe uma função f tal que $X = f(Y)$. Neste caso, X é completamente determinado por Y .

Um refinamento fornece uma visão “mais detalhada” ou “mais granular” da mesma informação. Conhecendo-se Y , é possível calcular exatamente X , mas o contrário nem sempre é verdade.

Exemplo 53. Se Y representa o resultado de lançar um dado, valores 1 a 6, e X representa se o resultado é par ou ímpar, então Y é um refinamento de X . A função f é

$$f(y) = \begin{cases} \text{par}, & \text{se } y \in \{2, 4, 6\} \\ \text{ímpar}, & \text{se } y \in \{1, 3, 5\}. \end{cases}$$

Conhecendo Y , sabe-se exatamente se é par ou ímpar, mas conhecendo apenas que é par, não se sabe qual dos três valores pares ocorreu.

Exemplo 54. Se X representa a temperatura em Celsius e Y representa a temperatura em Fahrenheit, então X e Y são refinamentos um do outro, pois existem funções bijetoras $f(Y) = \frac{5}{9}(Y - 32)$ e $g(X) = \frac{9}{5}X + 32$ que permitem converter entre as escalas.

Exemplo 55. Se Y é um vetor (Y_1, Y_2) e $X = Y_1 + Y_2$, então Y é um refinamento de X . Conhecendo ambas as componentes, sabemos exatamente a soma, mas conhecendo apenas a soma, não sabemos como ela foi distribuída entre as componentes.

Observação 11. A relação de refinamento é transitiva: se Z é refinamento de Y e Y é refinamento de X , então Z é refinamento de X . Isso cria uma hierarquia natural de “níveis de detalhe” na informação.

Observação 12. Em termos de informação, um refinamento sempre contém pelo menos tanta informação quanto o original, e geralmente contém mais. Esta ideia será fundamental no Capítulo 4 quando estudarmos entropia de Shannon.

3. CONJUNTO DAS FUNÇÕES DE BOLTZMANN

Conforme introduzido no Capítulo 1, consideremos o conjunto $\mathcal{B}$, das funções de Boltzmann, caracterizado pelas propriedades de monotonicidade estrita e aditividade multiplicativa. Neste capítulo, investiga-se sistematicamente a estrutura matemática deste conjunto, respondendo às questões fundamentais levantadas na introdução sobre suas propriedades de regularidade, topológicas e sua caracterização completa.

3.1 PROPRIEDADES FUNDAMENTAIS

Proposição 1 (Propriedades elementares). *Se $f \in \mathcal{B}$, então*

1. $f(1) = 0$;
2. $f(x^n) = n f(x)$ para todo $n \in \mathbb{Z}$ e $x > 0$.

Demonstração. Demonstra-se cada item separadamente.

Item (1). Seja $f \in \mathcal{B}$. Pela propriedade de aditividade multiplicativa, tem-se que para quaisquer $x, y > 0$

$$f(xy) = f(x) + f(y).$$

Em particular, tomando $x = y = 1$, obtém-se

$$f(1 \cdot 1) = f(1) + f(1) \Rightarrow f(1) = 2f(1).$$

Subtraindo $f(1)$ de ambos os lados, tem-se que

$$0 = f(1).$$

Item (2). Dividiremos a demonstração em três casos.

Caso 1:

Para $n = 0$, tem-se que $x^0 = 1$, e pelo item (1) sabe-se que $f(1) = 0$. Portanto

$$f(x^0) = 0 \cdot f(x).$$

Caso 2:

Para $n > 0$, provaremos por indução em n . Observe que a igualdade é válida para $n = 1$

$$f(x^1) = 1 \cdot f(x).$$

Suponha que para $n = k$ a propriedade seja válida, isto é, $f(x^k) = kf(x)$. Prova-se que a igualdade é válida para $n = k + 1$

$$\begin{aligned} f(x^{k+1}) &= f(x^k \cdot x) \\ &= f(x^k) + f(x) \\ &= kf(x) + f(x) \\ &= (k + 1)f(x), \end{aligned}$$

onde usamos a hipótese de indução e a aditividade multiplicativa. Pelo Princípio da Indução Matemática, Teorema 5, a propriedade vale para todo $n \in \mathbb{N}$.

Caso 3. Para $n < 0$.

Seja $n = -m$ com $m > 0$. Primeiro, prova-se que $f(x^{-1}) = -f(x)$ para todo $x > 0$. Note que

$$f(x \cdot x^{-1}) = f(x) + f(x^{-1}) \Rightarrow f(1) = f(x) + f(x^{-1}) \Rightarrow 0 = f(x) + f(x^{-1}) \Rightarrow f(x^{-1}) = -f(x).$$

Agora, usando este resultado e o caso para expoentes positivos

$$\begin{aligned} f(x^n) &= f(x^{-m}) \\ &= f((x^m)^{-1}) \\ &= -f(x^m) \\ &= -mf(x) \\ &= nf(x). \end{aligned}$$

Portanto, a propriedade vale para todo $n \in \mathbb{Z}$. □

3.2 CARACTERIZAÇÃO DO DOMÍNIO

Teorema 8 (Domínio das funções em $\mathcal{B}$). *O domínio natural das funções que pertencem ao conjunto $\mathcal{B}$ é $(0, \infty)$. Mais precisamente, se $f \in \mathcal{B}$, então f não pode ser estendida continuamente a 0 nem a números negativos mantendo as propriedades (i) e (ii).*

Demonstração. A demonstração será dividida em duas partes.

Parte 1: f não está definida em 0.

Suponha, por contradição, que $f(0)$ esteja definida. Então para qualquer $x > 0$

$$f(0) = f(x \cdot 0) = f(x) + f(0),$$

o que implica que $f(x) = 0$, mas isso contradiz a propriedade de que f é estritamente crescente. Logo, $f(0)$ não pode ser definido.

Parte 2: f não está definida em números negativos.

Suponha que f esteja definida em $\mathbb{R} \setminus \{0\}$ e satisfaça (i) e (ii). Pela Proposição 1, tem-se $f(1) = 0$. Além disso,

$$\begin{aligned} 0 &= f(1) \\ &= f((-1)(-1)) \\ &= f(-1) + f(-1) \\ &= 2f(-1), \end{aligned}$$

o que mostra que $f(-1) = 0$. Agora, para $x > 0$

$$\begin{aligned} f(-x) &= f((-1) \cdot x) \\ &= f(-1) + f(x) \\ &= f(x). \end{aligned}$$

Logo, $f(-x) = f(x)$, $\forall x \in \mathbb{R} \setminus \{0\}$. Mas isso é uma contradição, pela monotonicidade estrita. Portanto, f não pode estar definida em números negativos. $\square$

3.3 PROPRIEDADES DE REGULARIDADE

Teorema 9 (Continuidade). *Toda função $f \in \mathcal{B}$ é contínua em $(0, \infty)$.*

Demonstração. A demonstração será realizada em duas etapas.

Etapa 1: Continuidade em $x = 1$.

Seja (x_n) uma sequência em $(0, \infty)$ com $x_n \rightarrow 1$. Deseja-se mostrar que $f(x_n) \rightarrow f(1) = 0$. Fixado $t > 1$, pelo Corolário 2 tem-se

$$\left(\frac{1}{t}\right)^{1/n} \rightarrow 1 \quad \text{e} \quad t^{1/n} \rightarrow 1.$$

Como $x_n \rightarrow 1$ e ambas as sequências laterais convergem para 1, existe $N \in \mathbb{N}$ tal que para todo $n \geq N$

$$\left(\frac{1}{t}\right)^{1/n} \leq x_n \leq t^{1/n}.$$

Pela monotonicidade estrita de f , propriedade (i) da definição de $\mathcal{B}$, tem-se

$$f\left(\left(\frac{1}{t}\right)^{1/n}\right) \leq f(x_n) \leq f(t^{1/n}).$$

Usando a propriedade aditiva multiplicativa, propriedade (ii) da definição de $\mathcal{B}$, e a Proposição

1

$$f(t^{1/n}) = \frac{1}{n}f(t) \quad \text{e} \quad f\left(\left(\frac{1}{t}\right)^{1/n}\right) = -\frac{1}{n}f(t).$$

Dessa forma,

$$-\frac{f(t)}{n} \leq f(x_n) \leq \frac{f(t)}{n} \Rightarrow |f(x_n)| \leq \frac{|f(t)|}{n} \rightarrow 0.$$

Logo, $f(x_n) \rightarrow 0 = f(1)$. Pelo Teorema 4 (Critério de continuidade via sequências), f é contínua em 1.

Etapa 2: Continuidade em ponto arbitrário.

Seja $a > 0$ e (x_n) com $x_n \rightarrow a$. Escreva $x_n = ar_n$, onde $r_n = \frac{x_n}{a} \rightarrow 1$. Então

$$\begin{aligned} f(x_n) &= f(ar_n) \\ &= f(a) + f(r_n). \end{aligned}$$

Como f é contínua em 1, tem-se $f(r_n) \rightarrow f(1) = 0$. Logo

$$f(x_n) \rightarrow f(a) + 0 = f(a).$$

Portanto, f é contínua em todo $a > 0$. □

Teorema 10 (Diferenciabilidade). *Toda função $f \in \mathcal{B}$ é diferenciável em $(0, \infty)$.*

Demonstração. Defina $g : \mathbb{R} \rightarrow \mathbb{R}$ por $g(u) = f(e^u)$. Para $u, v \in \mathbb{R}$

$$\begin{aligned} g(u+v) &= f(e^{u+v}) \\ &= f(e^u e^v) \\ &= f(e^u) + f(e^v) \\ &= g(u) + g(v). \end{aligned}$$

Logo, g é aditiva. Além disso, g é estritamente crescente pois e^u é crescente e f é crescente.

Pelo Teorema 6 (Homogeneidade racional) e usando a continuidade de g , que segue da continuidade de f , tem-se que existe $k \in \mathbb{R}$ tal que $g(u) = ku$ para todo $u \in \mathbb{R}$.

Agora, para $a > 0$, calcula-se a derivada de f . Seja $r > 0$ e escreva $r = e^t$. Então

$$\begin{aligned} \frac{f(ar) - f(a)}{ar - a} &= \frac{f(r)}{a(r-1)} \\ &= \frac{g(t)}{a(e^t - 1)} \\ &= \frac{kt}{a(e^t - 1)}. \end{aligned}$$

Pelo Corolário 1, tem-se

$$\lim_{t \rightarrow 0} \frac{e^t - 1}{t} = 1 \Rightarrow \lim_{t \rightarrow 0} \frac{kt}{a(e^t - 1)} = \frac{k}{a}.$$

Logo,

$$\begin{aligned} f'(a) &= \lim_{r \rightarrow 1} \frac{f(ar) - f(a)}{ar - a} \\ &= \frac{k}{a}. \end{aligned}$$

Portanto, f é diferenciável em $(0, \infty)$ e $f'(x) = \frac{k}{x}$. □

Teorema 11 (Analiticidade). *Toda função $f \in \mathcal{B}$ é analítica em $(0, \infty)$.*

Demonstração. Do Teorema 10, sabe-se que $f'(x) = \frac{k}{x}$ para algum $k \in \mathbb{R}$. Por indução, pode-se mostrar que

$$f^{(n)}(x) = (-1)^{n-1}(n-1)! \frac{k}{x^n} \quad \text{para } n \geq 1.$$

Segue da Definição 13 que a série de Taylor de f em torno de $a > 0$ é

$$\begin{aligned} T_a(x) &= f(a) + \sum_{n=1}^{\infty} \frac{f^{(n)}(a)}{n!} (x-a)^n \\ &= f(a) + \sum_{n=1}^{\infty} (-1)^{n-1} \frac{k}{na^n} (x-a)^n. \end{aligned}$$

Pelo Teorema 2, o raio de convergência desta série é $R = a$ (veja Definição 5). Para mostrar que a série de Taylor coincide com $f(x)$ no intervalo de convergência, consideramos $|x-a| < a$. Neste intervalo, pode-se escrever

$$\begin{aligned} f(x) - f(a) &= \int_a^x f'(s) ds \\ &= k \int_a^x \frac{1}{s} ds \\ &= k \int_a^x \frac{1}{a + (s-a)} ds. \end{aligned}$$

Expandindo em série geométrica, Exemplo 9, para $|s-a| < a$

$$\frac{1}{a + (s-a)} = \frac{1}{a} \sum_{m=0}^{\infty} (-1)^m \left(\frac{s-a}{a} \right)^m.$$

Como a série de potências converge uniformemente em qualquer intervalo compacto contido em seu raio de convergência, a integração termo a termo é válida [5]. Logo, obtém-se

$$\begin{aligned} f(x) - f(a) &= \frac{k}{a} \sum_{m=0}^{\infty} (-1)^m \frac{(x-a)^{m+1}}{(m+1)a^m} \\ &= \sum_{n=1}^{\infty} (-1)^{n-1} \frac{k}{na^n} (x-a)^n. \end{aligned}$$

Que coincide com a série de Taylor. Portanto, f é analítica em $(0, \infty)$. □

Corolário 3 (Classe de diferenciabilidade). *Toda função $f \in \mathcal{B}$ é de classe C^∞ em $(0, \infty)$.*

Demonstração. Segue imediatamente do Teorema 11, pois funções analíticas são infinitamente diferenciáveis, Definição 14. $\square$

3.4 PROPRIEDADES TOPOLÓGICAS DO CONJUNTO $\mathcal{B}$

Nesta seção, considera-se $\mathcal{B}$ como subconjunto do espaço $C((0, \infty), \mathbb{R})$ com a topologia da convergência uniforme em compactos, Definição 25.

Teorema 12 (Não abertura). *O conjunto $\mathcal{B}$ não é aberto em $C((0, \infty), \mathbb{R})$.*

Demonstração. Seja $f \in \mathcal{B}$. Para provar que o conjunto não é aberto, será construída uma função g_ϵ que está arbitrariamente próxima de f na topologia da convergência uniforme em compactos, mas que não pertence a $\mathcal{B}$.

Tome um ponto $x_0 \in (0, \infty)$ com $x_0 > 1$. Dado um raio $\epsilon > 0$ para a vizinhança na topologia em questão, escolha um $\delta > 0$ suficientemente pequeno tal que o intervalo fechado $[x_0 - \delta, x_0 + \delta]$ não contenha o ponto x_0^2 . Isso é sempre possível pois $x_0 \neq x_0^2$ para $x_0 > 1$, bastando tomar $\delta < |x_0^2 - x_0|$.

Defina a função perturbação $b_\epsilon : (0, \infty) \rightarrow \mathbb{R}$ por

$$b_\epsilon(x) = \begin{cases} \frac{\epsilon}{2} \left(1 - \frac{|x - x_0|}{\delta}\right), & \text{se } |x - x_0| \leq \delta \\ 0, & \text{se } |x - x_0| > \delta. \end{cases}$$

Note que b_ϵ é contínua em todo o seu domínio, pois nos pontos de fronteira $x_0 - \delta$ e $x_0 + \delta$ a função se anula continuamente.

Antes de encontrar o valor máximo dessa perturbação, verifica-se o comportamento da sua derivada no ponto x_0 . Calculando os limites laterais da razão incremental em x_0 , obtém-se

$$\begin{aligned} b'_\epsilon(x_0) &= \lim_{h \rightarrow 0} \frac{b_\epsilon(x_0 + h) - b_\epsilon(x_0)}{h} \\ &= \lim_{h \rightarrow 0} \frac{\frac{\epsilon}{2} - \frac{\epsilon}{2\delta}|x_0 + h - x_0| - \frac{\epsilon}{2}}{h} \\ &= \lim_{h \rightarrow 0} \left(-\frac{\epsilon}{2\delta} \frac{|h|}{h} \right). \end{aligned}$$

Analisando os limites laterais, tem-se que o limite pela direita ($h \rightarrow 0^+$) resulta em $-\frac{\epsilon}{2\delta}$, enquanto o limite pela esquerda ($h \rightarrow 0^-$) resulta em $\frac{\epsilon}{2\delta}$. Como os limites laterais são distintos, a derivada $b'_\epsilon(x_0)$ não existe. Consequentemente, de acordo com a Definição 12, x_0 é um ponto crítico da função b_ϵ .

Agora, aplica-se o Teste da Primeira Derivada para classificar esse ponto crítico. Para $x \in (x_0 - \delta, x_0)$, tem-se $x - x_0 < 0$, o que implica $|x - x_0| = -(x - x_0)$. Substituindo na expressão da função, deriva-se e obtém-se $b'_\epsilon(x) = \frac{\epsilon}{2\delta} > 0$. Logo, a função é estritamente crescente no intervalo $(x_0 - \delta, x_0)$. Por outro lado, para $x \in (x_0, x_0 + \delta)$, tem-se $x - x_0 > 0$,

de modo que $|x - x_0| = x - x_0$. A derivada nesse trecho é $b'_\epsilon(x) = -\frac{\epsilon}{2\delta} < 0$, indicando que a função é estritamente decrescente em $(x_0, x_0 + \delta)$.

Como b_ϵ é crescente à esquerda de x_0 e decrescente à direita, conclui-se que x_0 é um ponto de máximo absoluto. O valor desse máximo é

$$b_\epsilon(x_0) = \frac{\epsilon}{2} - \frac{\epsilon}{2\delta}|x_0 - x_0| = \frac{\epsilon}{2}.$$

Por fim, constrói-se a função $g_\epsilon(x) = f(x) + b_\epsilon(x)$. Observa-se que g_ϵ está dentro da vizinhança de raio ϵ em torno de f , pois a distância máxima entre elas ocorre no pico da perturbação

$$\sup_{x \in [x_0 - \delta, x_0 + \delta]} |g_\epsilon(x) - f(x)| = \sup_{x \in [x_0 - \delta, x_0 + \delta]} |b_\epsilon(x)| = \frac{\epsilon}{2} < \epsilon.$$

No entanto, g_ϵ não satisfaz a propriedade multiplicativa que caracteriza o conjunto $\mathcal{B}$. Avaliando a função no ponto x_0^2 , tem-se

$$g_\epsilon(x_0 \cdot x_0) = g_\epsilon(x_0^2) = f(x_0^2) + b_\epsilon(x_0^2).$$

Pela escolha inicial de δ , o ponto x_0^2 está fora do intervalo $[x_0 - \delta, x_0 + \delta]$, logo $b_\epsilon(x_0^2) = 0$. Sabendo que $f \in \mathcal{B}$, tem-se $f(x_0^2) = 2f(x_0)$, o que resulta em $g_\epsilon(x_0 \cdot x_0) = 2f(x_0)$.

Por outro lado, a soma das funções avaliadas individualmente resulta em

$$\begin{aligned} g_\epsilon(x_0) + g_\epsilon(x_0) &= 2(f(x_0) + b_\epsilon(x_0)) \\ &= 2\left(f(x_0) + \frac{\epsilon}{2}\right) \\ &= 2f(x_0) + \epsilon. \end{aligned}$$

Como $\epsilon > 0$, conclui-se que $g_\epsilon(x_0 \cdot x_0) \neq g_\epsilon(x_0) + g_\epsilon(x_0)$. Portanto, $g_\epsilon \notin \mathcal{B}$, provando que qualquer vizinhança de f contém elementos fora de $\mathcal{B}$ e, conseqüentemente, o conjunto $\mathcal{B}$ não é aberto. $\square$

Teorema 13 (Não fechamento). *O conjunto $\mathcal{B}$ não é fechado em $C((0, \infty), \mathbb{R})$.*

Demonstração. Pela definição 16, um conjunto é fechado se contém todos os seus pontos de acumulação. Será mostrado que $\mathcal{B}$ possui um ponto de acumulação que não pertence a $\mathcal{B}$.

Seja $f \in \mathcal{B}$ não identicamente nula. Defina a sequência $f_n = \frac{1}{n}f$. Pela linearidade da equação funcional, cada $f_n \in \mathcal{B}$.

Para qualquer compacto $K \subset (0, \infty)$, tem-se

$$\sup_{x \in K} |f_n(x) - 0| = \frac{1}{n} \sup_{x \in K} |f(x)| \rightarrow 0.$$

Logo, da definição 25 segue que $f_n \rightarrow 0$ na topologia da convergência uniforme em compactos.

Porém, a função nula não pertence a $\mathcal{B}$ pois não é estritamente crescente: para $x < y$ tem-se $0(x) = 0 = 0(y)$, violando a propriedade (i).

Portanto, foi encontrada uma sequência $(f_n) \subset \mathcal{B}$ que converge para um ponto fora de $\mathcal{B}$, logo $\mathcal{B}$ não é fechado. $\square$

Teorema 14 (Convexidade). *O conjunto $\mathcal{B}$ é convexo.*

Demonstração. Sejam $f, g \in \mathcal{B}$ e $\alpha \in [0, 1]$. Defina $h = \alpha f + (1 - \alpha)g$ pela definição 17, deve-se verificar que h satisfaz as propriedades (i) e (ii) que definem $\mathcal{B}$.

Para a aditividade multiplicativa, propriedade (ii), tem-se

$$\begin{aligned} h(xy) &= \alpha f(xy) + (1 - \alpha)g(xy) \\ &= \alpha[f(x) + f(y)] + (1 - \alpha)[g(x) + g(y)] \\ &= h(x) + h(y). \end{aligned}$$

Para a Monotonicidade estrita, propriedade (i), se $x < y$, então $f(x) < f(y)$ e $g(x) < g(y)$ pois $f, g \in \mathcal{B}$. Para $\alpha \in (0, 1)$

$$h(x) = \alpha f(x) + (1 - \alpha)g(x) < \alpha f(y) + (1 - \alpha)g(y) = h(y),$$

onde a desigualdade estrita segue da combinação de desigualdades estritas com coeficientes positivos. Os casos $\alpha = 0$ e $\alpha = 1$ são triviais, resultam em g e f respectivamente, que já estão em $\mathcal{B}$.

Portanto, $h \in \mathcal{B}$ e $\mathcal{B}$ é convexo. $\square$

Teorema 15 (Não compacidade). *O conjunto $\mathcal{B}$ não é compacto em $C((0, \infty), \mathbb{R})$.*

Demonstração. Pela Definição 18, um conjunto é compacto se é fechado e limitado. Será mostrado que $\mathcal{B}$ não é limitado.

Seja $f \in \mathcal{B}$ não nula. Considere a sequência $f_n = nf$. Cada $f_n \in \mathcal{B}$ pela linearidade da equação funcional. Para qualquer compacto $K \subset (0, \infty)$

$$\sup_{x \in K} |f_n(x)| = n \sup_{x \in K} |f(x)| \rightarrow \infty.$$

Logo, pela definição 25, $\mathcal{B}$ não é limitado na topologia da convergência uniforme em compactos.

Portanto, pela definição 18, $\mathcal{B}$ não pode ser compacto, pois não é limitado. $\square$

Teorema 16 (Conexidade). *O conjunto $\mathcal{B}$ é conexo por caminhos.*

Demonstração. Pela Definição 20, um conjunto é conexo por caminhos se para quaisquer dois pontos existe um caminho contínuo ligando-os dentro do conjunto.

Sejam $f, g \in \mathcal{B}$. Defina o caminho $\gamma : [0, 1] \rightarrow \mathcal{B}$ por

$$\gamma(t) = (1 - t)f + tg.$$

Pela demonstração do Teorema 14, cada $\gamma(t) \in \mathcal{B}$. Além disso, segue pela definição 25 que γ é contínua na topologia da convergência uniforme em compactos.

Como $\gamma(0) = f$ e $\gamma(1) = g$ tem-se que $\mathcal{B}$ é conexo por caminhos pela definição 20. $\square$

3.5 CARACTERIZAÇÃO COMPLETA DAS FUNÇÕES DE BOLTZMANN

Teorema 17 (Caracterização de $\mathcal{B}$). *Se $f \in \mathcal{B}$, então existe $k > 0$ tal que $f(x) = k \ln(x)$ para todo $x > 0$.*

Demonstração. Pelo Teorema 10, sabe-se que f é diferenciável e que $f'(x) = \frac{k}{x}$ para algum $k \in \mathbb{R}$. Como f é estritamente crescente, tem-se $k > 0$. Integrando de 1 a x

$$\begin{aligned} f(x) - f(1) &= \int_1^x f'(s) ds \\ &= \int_1^x \frac{k}{s} ds \\ &= k \ln x. \end{aligned}$$

Como $f(1) = 0$ pela Proposição 1, conclui-se que $f(x) = k \ln x$. □

Corolário 4 (Estrutura de $\mathcal{B}$). *O conjunto $\mathcal{B}$ é exatamente o conjunto dos múltiplos positivos do logaritmo natural, mais precisamente*

$$\mathcal{B} = \{k \ln x \mid k > 0\}.$$

Demonstração. Pelo Teorema 17, toda $f \in \mathcal{B}$ é da forma $k \ln x$ com $k > 0$. Reciprocamente, é fácil verificar que para qualquer $k > 0$, a função $f(x) = k \ln x$ pertence a $\mathcal{B}$. □

4. ENTROPIA DE SHANNON

Neste capítulo, avança-se da caracterização matemática do conjunto das funções de Boltzmann $\mathcal{B}$, desenvolvida no capítulo anterior, para o estudo sistemático da entropia de Shannon. O objetivo central é estabelecer uma conexão entre essas duas formulações de entropia, investigando como as propriedades fundamentais de $\mathcal{B}$ encontram um análogo ou uma generalização no contexto da teoria da informação estudado por Shannon.

Para isso, parte-se da definição clássica, proposta por Shannon em [7] e apresentam-se suas propriedades básicas, que serão utilizadas ao longo do capítulo. Em seguida, demonstra-se que, no caso particular de variáveis aleatórias estatisticamente independentes, a entropia de Shannon herda a estrutura matemática de $\mathcal{B}$. Essa análise não apenas unifica as perspectivas termodinâmica e informacional, mas também revela a hierarquia lógica que vai da formulação de Boltzmann à de Shannon.

A partir da definição fundamental da entropia de Shannon, são introduzidas de forma encadeada as noções de entropia conjunta e entropia condicional, que serão essenciais para formalizar a comparação com $\mathcal{B}$.

Definição 33 (Entropia de Shannon). *Seja X uma variável aleatória discreta (veja definição 27) com espaço amostral $\mathcal{X}$ e função de probabilidade $p(x)$. A entropia de Shannon de X é definida como*

$$H(X) = - \sum_{x \in \mathcal{X}} p(x) \log p(x).$$

Definição 34 (Entropia Conjunta). *Dadas duas variáveis aleatórias X e Y , a entropia conjunta é definida como*

$$H(X, Y) = - \sum_x \sum_y p(x, y) \log p(x, y),$$

sendo $p(x, y)$ é a função de probabilidade conjunta (veja definição 28).

Definição 35 (Entropia Condicional). *Dadas duas variáveis aleatórias X e Y , a entropia condicional de Y dado X é definida como*

$$H(Y | X) = - \sum_x \sum_y p(x, y) \log p(y | x),$$

com $p(y | x)$ probabilidade condicional (veja definição 29).

Proposição 2. *A entropia de Shannon é condicionalmente aditiva, isto é*

$$H(X, Y) = H(X) + H(Y | X).$$

Demonstração. Pela definição de entropia conjunta, tem-se que

$$H(X, Y) = - \sum_x \sum_y p(x, y) \log p(x, y).$$

Como pela definição 29 $p(x, y) = p(x)p(y | x)$, tem-se

$$H(X, Y) = - \sum_x \sum_y p(x, y) \log [p(x)p(y | x)].$$

Usando a propriedade do logaritmo do produto

$$H(X, Y) = - \sum_x \sum_y p(x, y) \log p(x) - \sum_x \sum_y p(x, y) \log p(y | x).$$

Na primeira soma, como $\log p(x)$ não depende de y

$$\begin{aligned} - \sum_x \sum_y p(x, y) \log p(x) &= - \sum_x \log p(x) \left(\sum_y p(x, y) \right) \\ &= - \sum_x \log p(x) \left(\sum_y p(x)p(y | x) \right) \\ &= - \sum_x p(x) \log p(x) \\ &= H(X). \end{aligned}$$

A segunda soma corresponde à definição de entropia condicional

$$- \sum_x \sum_y p(x, y) \log p(y | x) = H(Y | X).$$

Portanto

$$H(X, Y) = H(X) + H(Y | X).$$

□

Proposição 3. *A entropia de Shannon é não-decrescente por refinamento (veja definição 32), isto é, seja Y um refinamento de X , então*

$$H(X) \leq H(Y).$$

Demonstração. Se Y é um refinamento de X , existe uma função f tal que $X = f(Y)$.

Como X é completamente determinado por Y (através da função f), o par (X, Y) contém

exatamente a mesma informação que Y sozinho (veja a definição 32). Portanto:

$$H(X, Y) = H(Y).$$

Pela Proposição 2

$$H(X, Y) = H(X) + H(Y | X).$$

Combinando estas duas igualdades

$$H(Y) = H(X) + H(Y | X) \Rightarrow H(Y) - H(X) = H(Y | X).$$

Como a entropia condicional é não-negativa, $H(Y | X) \geq 0$, segue que

$$H(Y) - H(X) \geq 0 \Rightarrow H(Y) \geq H(X).$$

A igualdade ocorre se e somente se $H(Y | X) = 0$, o que significa que Y é determinística-mente determinado por X , ou seja, o refinamento é trivial. $\square$

Observação 13. *Para variáveis aleatórias arbitrárias X e Y que não são refinamentos uma da outra, a entropia não é necessariamente não-decrescente. Isto é, se Y não é refinamento de X e X não é refinamento de Y , então pode-se ter $H(X) > H(Y)$ ou $H(X) < H(Y)$, mesmo se a cardinalidade do espaço amostral de Y for maior que a de X .*

Demonstração. Seja $|\mathcal{X}|$ e $|\mathcal{Y}|$ denotam as cardinalidades dos espaços amostrais de X e Y , respectivamente. A prova será dividida em dois casos.

Caso 1: $H(X) > H(Y)$ com $|\mathcal{Y}| > |\mathcal{X}|$

Seja X uma variável com distribuição

$$P(X = 0) = 0.99, \quad P(X = 1) = 0.01.$$

$$H(X) \approx 0.08 \text{ bits.}$$

Seja Y uma variável com três resultados possíveis

$$P(Y = 0) = 0.999, \quad P(Y = 1) = 0.0005, \quad P(Y = 2) = 0.0005.$$

$$H(Y) \approx 0.01 \text{ bits.}$$

Tem-se $|\mathcal{Y}| = 3 > |\mathcal{X}| = 2$ mas $H(Y) < H(X)$.

Caso 2: $H(X) < H(Y)$ com $|\mathcal{X}| > |\mathcal{Y}|$

Seja X uma variável com três resultados

$$P(X = 0) = 0.8, \quad P(X = 1) = 0.1, \quad P(X = 2) = 0.1.$$

$$H(X) \approx 0.92 \text{ bits.}$$

Seja Y uma variável Bernoulli com distribuição uniforme

$$P(Y = 0) = 0.5, \quad P(Y = 1) = 0.5.$$

$$H(Y) = 1 \text{ bit.}$$

Tem-se $|\mathcal{X}| = 3 > |\mathcal{Y}| = 2$ e $H(X) < H(Y)$.

Estes contraexemplos mostram que, na ausência de relação de refinamento, não há qualquer garantia sobre a relação entre as entropias, independentemente do número de resultados. $\square$

Definição 36. *Seja $\mathcal{H}$ o conjunto de todos os funcionais F definidos no espaço de variáveis aleatórias discretas que satisfazem:*

i) F é não-decrescente por refinamento;

ii) F é condicionalmente aditivo, isto é, $F(X, Y) = F(X) + F(Y | X)$ para quaisquer variáveis aleatórias X e Y .

Lema 1 (Continuidade de $F \in \mathcal{H}$). *Se $F \in \mathcal{H}$, então F é contínua em relação à topologia das distribuições de probabilidade.*

Demonstração. Seja X uma variável aleatória com distribuição $(p_1, \dots, p_n)$ e seja $X^{(m)}$ uma sequência de variáveis aleatórias com distribuições $(p_1^{(m)}, \dots, p_n^{(m)})$ convergindo para $(p_1, \dots, p_n)$.

Para cada m , construímos uma variável aleatória $Z^{(m)}$ que é um refinamento tanto de X quanto de $X^{(m)}$. Especificamente, particiona-se um conjunto uniforme suficientemente grande em subconjuntos que refinam ambas as partições correspondentes a X e $X^{(m)}$.

Pela propriedade de não-decrescimento por refinamento, tem-se:

$$F(X) \leq F(Z^{(m)}) \quad \text{e} \quad F(X^{(m)}) \leq F(Z^{(m)}).$$

Pela aditividade condicional, tem-se:

$$\begin{aligned} F(Z^{(m)}) &= F(X) + F(Z^{(m)} | X) \\ &= F(X^{(m)}) + F(Z^{(m)} | X^{(m)}). \end{aligned}$$

Quando $m \rightarrow \infty$, $F(Z^{(m)} | X) \rightarrow 0$ e $F(Z^{(m)} | X^{(m)}) \rightarrow 0$, portanto $F(X) - F(X^{(m)}) \rightarrow 0$.

Portanto, F é contínua. $\square$

Teorema 18 (Caracterização de $\mathcal{H}$). *Seja H a entropia de Shannon. Então*

$$\mathcal{H} = \{kH \mid k > 0\}.$$

Demonstração. Demonstra-se em duas etapas: (1) $\{kH \mid k > 0\} \subseteq \mathcal{H}$ e (2) $\mathcal{H} \subseteq \{kH \mid k > 0\}$.

Primeiramente, mostra-se que $\{kH \mid k > 0\} \subseteq \mathcal{H}$. Seja $F \in \{kH \mid k > 0\}$. Então existe $k > 0$ tal que $F = kH$. Note que pela Proposição 3, tem-se que

$$H(Y) \geq H(X),$$

quando Y é refinamento de X . Assim, $kH(Y) \geq kH(X)$ o que implicar que $F(Y) \geq F(X)$. Por outro lado, pela Proposição 2, tem-se que

$$H(X, Y) = H(X) + H(Y|X).$$

Logo, $kH(X, Y) = kH(X) + kH(Y|X)$ o que implicar em $F(X, Y) = F(X) + F(Y|X)$. Portanto, $F = kH \in \mathcal{H}$.

Agora prova-se que: $\mathcal{H} \subseteq \{kH \mid k > 0\}$. Seja $F \in \mathcal{H}$. Deseja-se mostrar que existe $k > 0$ tal que $F(X) = kH(X)$ para toda variável aleatória X .

Seja U_2 uma variável uniforme sobre dois elementos. Defina-se

$$k = \frac{F(U_2)}{H(U_2)} = \frac{F(U_2)}{\log 2}.$$

Como F é não-decrescente por refinamento, $F(U_2) > 0$, logo $k > 0$.

Considerem-se agora variáveis uniformes U_{2^m} sobre 2^m elementos. Tomem-se m cópias independentes de U_2 , denotadas por $U_2^{(1)}, U_2^{(2)}, \dots, U_2^{(m)}$. Pela aditividade condicional aplicada iterativamente

$$F(U_2^{(1)}, U_2^{(2)}, \dots, U_2^{(m)}) = F(U_2^{(1)}) + F(U_2^{(2)} \mid U_2^{(1)}) + \dots + F(U_2^{(m)} \mid U_2^{(1)}, \dots, U_2^{(m-1)}).$$

Pela independência, $F(U_2^{(i)} \mid U_2^{(1)}, \dots, U_2^{(i-1)}) = F(U_2^{(i)}) = F(U_2)$ para todo i , portanto

$$F(U_2^{(1)}, U_2^{(2)}, \dots, U_2^{(m)}) = mF(U_2).$$

Mas a sequência $(U_2^{(1)}, \dots, U_2^{(m)})$ tem distribuição uniforme sobre 2^m elementos, logo

$$\begin{aligned} F(U_{2^m}) &= mF(U_2) \\ &= mk \log 2 \\ &= k \log(2^m) \\ &= kH(U_{2^m}). \end{aligned}$$

Para uma variável aleatória arbitrária X com distribuição $(p_1, \dots, p_n)$, onde cada $p_i = \frac{m_i}{2^M}$ para inteiros m_i e algum M suficientemente grande, aproximação diádica (veja definição 31), constrói-se o seguinte refinamento.

Seja Y uma variável aleatória definida da seguinte forma: particiona-se um conjunto uniforme de 2^M elementos em n subconjuntos com tamanhos $m_1, m_2, \dots, m_n$. Define-se Y como o índice do subconjunto ao qual um elemento uniformemente escolhido pertence. Então Y tem

distribuição $(p_1, \dots, p_n)$ e é um refinamento de U_{2^M} .

Pela propriedade de não-decrescimento por refinamento, tem-se $F(Y) \leq F(U_{2^M}) = k \log(2^M) = kM \log 2$. Por outro lado, condicionando em Y , tem-se

$$F(U_{2^M}) = F(Y) + F(U_{2^M} | Y).$$

Mas $F(U_{2^M} | Y = i) = F(U_{m_i}) = k \log m_i$, portanto

$$\begin{aligned} F(U_{2^M} | Y) &= \sum_{i=1}^n p_i k \log m_i \\ &= k \sum_{i=1}^n p_i \log m_i. \end{aligned}$$

Assim,

$$\begin{aligned} F(Y) &= F(U_{2^M}) - F(U_{2^M} | Y) \\ &= kM \log 2 - k \sum_{i=1}^n p_i \log m_i \\ &= -k \sum_{i=1}^n p_i \log p_i \\ &= kH(X). \end{aligned}$$

Para probabilidades não diádicas, usa-se um argumento de limite baseado na densidade das distribuições diádicas (veja observação 10). Seja X uma variável aleatória com distribuição arbitrária $(p_1, \dots, p_n)$. Para cada inteiro $M > 0$, definem-se aproximações diádicas $p_i^{(M)} = m_i^{(M)} / 2^M$ onde $m_i^{(M)}$ são inteiros não-negativos tais que $\sum_{i=1}^n m_i^{(M)} = 2^M$ e $|p_i^{(M)} - p_i| \rightarrow 0$ quando $M \rightarrow \infty$.

Constroem-se variáveis aleatórias $X^{(M)}$ com distribuições $(p_1^{(M)}, \dots, p_n^{(M)})$ usando o mesmo esquema de refinamento anterior. Pela construção, cada $X^{(M)}$ é um refinamento de U_{2^M} e tem-se

$$F(X^{(M)}) = kH(X^{(M)}).$$

Quando $M \rightarrow \infty$, as distribuições $(p_1^{(M)}, \dots, p_n^{(M)})$ convergem para $(p_1, \dots, p_n)$. Pelo Lema de Continuidade, F é contínua, logo

$$\begin{aligned} F(X) &= \lim_{M \rightarrow \infty} F(X^{(M)}) \\ &= \lim_{M \rightarrow \infty} kH(X^{(M)}) \\ &= kH(X), \end{aligned}$$

onde a última igualdade segue da continuidade da função entropia.

Deste modo, $F(X) = kH(X)$ para toda variável aleatória X , o que conclui a demonstração

de que $\mathcal{H} \subseteq \{kH \mid k > 0\}$.

Portanto $\mathcal{H} = \{kH \mid k > 0\}$. $\square$

Definição 37 (Subconjunto de aditividade independente). *Seja $\mathcal{H}_b$ o subconjunto de $\mathcal{H}$ consistindo dos funcionais F que satisfazem a aditividade para variáveis independentes:*

$$\mathcal{H}_b = \{F \in \mathcal{H} : F(X, Y) = F(X) + F(Y) \text{ para quaisquer } X, Y \text{ independentes}\}.$$

Teorema 19 (Correspondência entre $\mathcal{B}$ e $\mathcal{H}_b$). *Existe uma correspondência biunívoca $\varphi : \mathcal{B} \rightarrow \mathcal{H}_b$ dada por $\varphi(k \ln) = kH$ que preserva a estrutura aditiva multiplicativa.*

Demonstração. Defina-se a aplicação $\varphi : \mathcal{B} \rightarrow \mathcal{H}_b$ por

$$\varphi(k \ln) = kH,$$

onde $k \ln$ denota a função $x \mapsto k \ln x$ em $\mathcal{B}$, e kH denota o funcional $X \mapsto kH(X)$ em $\mathcal{H}_b$.

Primeiro, verifica-se que φ está bem definida. Pelo Corolário 4, cada elemento de $\mathcal{B}$ é da forma $k \ln$ com $k > 0$. Pelo Teorema 18, cada elemento de $\mathcal{H}_b$ é da forma kH com $k > 0$. Portanto, a associação $k \ln \mapsto kH$ estabelece uma relação bem definida entre os conjuntos.

Agora, mostra-se que φ é uma correspondência biunívoca que preserva a estrutura:

(i) φ preserva a estrutura aditiva multiplicativa: para $f = k \ln \in \mathcal{B}$ e $x, y > 0$,

$$\begin{aligned} f(xy) &= k \ln(xy) \\ &= k \ln x + k \ln y \\ &= f(x) + f(y), \end{aligned}$$

e para $\varphi(f) = kH \in \mathcal{H}_b$ e variáveis independentes X, Y ,

$$\begin{aligned} \varphi(f)(X, Y) &= kH(X, Y) \\ &= kH(X) + kH(Y) \\ &= \varphi(f)(X) + \varphi(f)(Y). \end{aligned}$$

(ii) φ é injetora: se $\varphi(k_1 \ln) = \varphi(k_2 \ln)$, então $k_1 H = k_2 H$, o que implica que $k_1 = k_2$, logo $k_1 \ln = k_2 \ln$.

(iii) φ é sobrejetora: dado qualquer $F \in \mathcal{H}_b$, pelo Teorema 18, existe $k > 0$ tal que $F = kH = \varphi(k \ln)$.

(iv) φ preserva combinações lineares: para quaisquer $k_1, k_2 > 0$ e $c \in \mathbb{R}$,

$$\begin{aligned} \varphi(c(k_1 \ln) + k_2 \ln) &= \varphi((ck_1 + k_2) \ln) \\ &= (ck_1 + k_2)H \\ &= c(k_1 H) + k_2 H \\ &= c\varphi(k_1 \ln) + \varphi(k_2 \ln). \end{aligned}$$

Portanto, φ é uma correspondência biunívoca entre $\mathcal{B}$ e $\mathcal{H}_b$ que preserva a estrutura aditiva multiplicativa. $\square$

Observação 14. *Este resultado revela uma conexão entre a entropia de Shannon e a entropia de Boltzmann-Gibbs. A definição moderna de entropia na mecânica estatística, devida a Gibbs, é dada por*

$$S_{BG} = -k \sum_i p_i \ln p_i,$$

que é exatamente a entropia de Shannon a menos da constante k e da base do logaritmo.

Enquanto Boltzmann originalmente considerou o caso especial de probabilidades iguais, $S_B = k \ln W$, Gibbs generalizou este conceito para distribuições arbitrárias. O teorema acima mostra que quando as variáveis são independentes, a entropia de Shannon herda as propriedades fundamentais das funções entrópicas do conjunto $\mathcal{B}$, justificando matematicamente por que a forma $-k \sum p_i \ln p_i$ é a generalização natural da entropia de Boltzmann para sistemas com distribuições de probabilidade não uniformes.

Esta conexão unifica a teoria da informação de Shannon com os fundamentos da mecânica estatística, mostrando que a mesma estrutura matemática subjaz a ambos os conceitos de entropia.

Corolário 5. *A entropia de Shannon coincide com a entropia de Boltzmann a menos de uma constante multiplicativa quando as probabilidades são uniformes.*

Demonstração. Pela Observação anterior, a entropia de Boltzmann-Gibbs é dada por $S_{BG} = -k \sum_i p_i \ln p_i$, que equivale à entropia de Shannon com constante k e logaritmo natural. No caso particular de probabilidades uniformes $p_i = \frac{1}{W}$, tem-se

$$S_{BG} = -k \sum_{i=1}^W \frac{1}{W} \ln \frac{1}{W} = k \ln W,$$

que é exatamente a entropia de Boltzmann S_B . Portanto, a entropia de Shannon difere da entropia de Boltzmann apenas por uma constante multiplicativa $\frac{k}{\log e}$ quando as probabilidades são uniformes. $\square$

5. CONCLUSÕES

Este trabalho dedicou-se ao estudo sistemático do conjunto $\mathcal{B}$ das funções de Boltzmann, definido como o conjunto de todas as funções $f : (0, \infty) \rightarrow \mathbb{R}$ que são estritamente crescentes e satisfazem a propriedade de aditividade multiplicativa $f(xy) = f(x) + f(y)$ para todos $x, y > 0$. Partindo desta caracterização inicial, investigou-se a estrutura matemática deste conjunto, respondendo às questões levantadas na introdução.

Do ponto de vista das propriedades de regularidade, demonstrou-se que toda função $f \in \mathcal{B}$ é

- **Contínua** em $(0, \infty)$;
- **Diferenciável** em $(0, \infty)$ com $f'(x) = \frac{k}{x}$ para algum $k \in \mathbb{R}$;
- **Analítica** em $(0, \infty)$;
- **Infinitamente diferenciável** (C^∞) em $(0, \infty)$.

Estes resultados mostram que as propriedades funcionais impostas pela definição de $\mathcal{B}$ são extremamente restritivas, forçando um alto grau de regularidade em suas funções.

No que diz respeito às propriedades topológicas, considerando $\mathcal{B}$ como subconjunto do espaço $C((0, \infty), \mathbb{R})$ com a topologia da convergência uniforme em compactos, estabeleceu-se que

- $\mathcal{B}$ **não é aberto**;
- $\mathcal{B}$ **não é fechado**;
- $\mathcal{B}$ **é convexo**;
- $\mathcal{B}$ **não é compacto**;
- $\mathcal{B}$ **é conexo por caminhos**.

Estas propriedades topológicas revelam a estrutura global do conjunto $\mathcal{B}$, mostrando que ele forma um conjunto convexo e conexo, mas que não é topologicamente bem-comportado no sentido de ser aberto ou fechado no espaço ambiente.

A caracterização completa do conjunto $\mathcal{B}$ foi obtida com a demonstração de que

$$\mathcal{B} = \{k \ln x \mid k > 0\},$$

mais precisamente, o conjunto das funções de Boltzmann coincide exatamente com o conjunto dos múltiplos positivos do logaritmo natural.

Este resultado final possui profundo significado físico. Ele mostra que a exigência de que a entropia seja extensiva (aditiva para sistemas independentes) e crescente, com o número de microestados, determina unicamente a forma funcional da entropia, a menos de uma constante multiplicativa. A constante k , que emerge naturalmente da demonstração, pode ser identificada com a constante de Boltzmann na formulação estatística da termodinâmica.

A partir desta caracterização fundamental do conjunto $\mathcal{B}$, estendeu-se a investigação para o domínio da teoria da informação, onde se estudou a entropia de Shannon. A análise revelou uma hierarquia natural de generalizações no conceito de entropia.

1. **Boltzmann (1877):** $S_B = k \ln W \rightarrow$ **Aditividade e probabilidades uniformes e independentes** (sistemas microcanônicos).
2. **Boltzmann-Gibbs (1902):** $S_{BG} = -k \sum_i p_i \ln p_i \rightarrow$ **Aditividade mantida, probabilidades não-uniformes, mas ainda independentes** (fatoráveis).
3. **Shannon (1948):** $H(X) = -\sum p_i \log p_i \rightarrow$ **Aditividade condicional, sem restrição de independência.**

Através do Teorema 19, correspondência entre $\mathcal{B}$ e $\mathcal{H}_b$, demonstrou-se que a entropia de Shannon, quando restrita a sistemas estatisticamente independentes, herda exatamente as propriedades matemáticas das funções de Boltzmann. Esta conexão não é apenas formal. Quando aplicada a sistemas com distribuições de probabilidade uniformes, a entropia de Shannon reduz-se à entropia de Boltzmann a menos de uma constante multiplicativa.

A verdadeira generalização representada pela entropia de Shannon vai além da mera extensão a probabilidades não-uniformes. Enquanto a formulação de Boltzmann-Gibbs assume implicitamente a independência estatística entre subsistemas ($P(X, Y) = P(X)P(Y)$), a entropia de Shannon relaxa esta restrição através da propriedade de **aditividade condicional**:

$$H(X, Y) = H(X) + H(Y|X).$$

Esta estrutura matemática permite descrever sistemas com correlações arbitrárias, algo essencial para sistemas físicos reais. Para ilustrar esta necessidade, considere um gás contendo partículas carregadas (íons positivos e negativos) em um recipiente isolado. Devido à interação eletrostática de longo alcance, as posições das partículas carregadas não são independentes, a probabilidade de encontrar um íon negativo em uma determinada região depende fortemente da posição dos íons positivos próximos, ou seja, $P(\text{íon negativo} | \text{íon positivo}) \neq P(\text{íon negativo})$. Neste sistema, a distribuição de probabilidade conjunta não pode ser fatorada como produto de distribuições marginais, violando a suposição de independência da entropia de Boltzmann-Gibbs. A descrição termodinâmica adequada requer a estrutura da entropia de Shannon

$$S(\text{sistema}) = S(\text{íons positivos}) + S(\text{íons negativos} | \text{íons positivos}),$$

onde o segundo termo captura exatamente as correlações entre as partículas carregadas.

O exemplo do gás carregado demonstra que sistemas físicos reais frequentemente envolvem correlações que não podem ser ignoradas. Enquanto a entropia de Boltzmann-Gibbs é adequada para sistemas ideais com interações de curto alcance ou fracamente correlacionados, a entropia de Shannon fornece o arcabouço matemático necessário para sistemas com interações de longo alcance e correlações espaciais significativas.

Matematicamente, a análise mostra que a transição de $\mathcal{H}_b$ (aditividade para variáveis independentes) para $\mathcal{H}$ (aditividade condicional para variáveis arbitrárias) corresponde à substituição da aditividade simples pela aditividade condicional. Esta mudança não é apenas uma generalização técnica, mas reflete uma compreensão mais profunda da natureza da informação e da desordem em sistemas correlacionados.

Assim como Gibbs generalizou Boltzmann ao permitir probabilidades não uniformes, Shannon generalizou Gibbs ao remover a restrição de independência estatística. Esta hierarquia de generalizações não é acidental, mas reflete o progresso natural na compreensão dos fenômenos estatísticos, desde sistemas físicos ideais até sistemas complexos com correlações intrínsecas.

A equivalência matemática entre a entropia de Shannon e a forma de Boltzmann-Gibbs, no caso independente, juntamente com a capacidade da primeira de lidar com correlações arbitrárias, estabelece a entropia de Shannon como o conceito mais geral e fundamental. Ela unifica a termodinâmica estatística com a teoria da informação, proporcionando uma linguagem comum para descrever a desordem, a informação e a complexidade em sistemas naturais.

Em síntese, este trabalho não apenas respondeu completamente às questões matemáticas levantadas sobre a estrutura do conjunto $\mathcal{B}$, mas também forneceu uma fundamentação para a forma funcional da entropia na física estatística, conectando diferentes formulações da entropia e estabelecendo uma ponte conceitual entre a termodinâmica estatística e a teoria da informação. Através do estudo da relação entre $\mathcal{B}$, $\mathcal{H}_b$ e $\mathcal{H}$, revelou-se como a estrutura matemática da entropia evolui para acomodar sistemas cada vez mais complexos, culminando na formulação de Shannon que representa o ponto mais geral desta evolução conceitual.

REFERÊNCIAS BIBLIOGRÁFICAS

- [1] Callen, H. B.: *Thermodynamics and an Introduction to Thermostatistics*. John Wiley & Sons, 2^a ed., 1985.
- [2] Clausius, R.: *The Mechanical Theory of Heat – With its Applications to the Steam Engine and to Physical Properties of Bodies*. John van Voorst, London, 1865.
- [3] Eddington, A.: *The Nature of the Physical World*. Cambridge University Press, Cambridge, 1928.
- [4] Kurashima, K. H. e Oler, J. G.: *Topologia das funções entrópicas*. C.Q.D. - Revista Eletrônica Paulista de Matemática, 24:e24007, 2024. <https://sistemas.fc.unesp.br/ojs/index.php/revistacqd/index>.
- [5] Lima, E. L.: *Análise Real*, vol. 1 de *Coleção Matemática Universitária*. IMPA, 2016, ISBN 978-8524401183. 8^a edição.
- [6] Nussenzveig, H. M.: *Curso de Física Básica: Flúidos, Oscilações e Ondas, Calor*, vol. 2. Editora Blucher, 5^a ed., 2014.
- [7] Shannon, C. E.: *A mathematical theory of communication*. The Bell System Technical Journal, 27(3):379–423, 1948.