UNIVERSIDADE FEDERAL DE JUIZ DE FORA
INSTITUTO DE CIÊNCIAS EXATAS
PROGRAMA DE PÓS-GRADUAÇÃO EM MATEMÁTICA

Mateus Dutra Rhodes

Entropia Sequencial de Sistemas Dinâmicos

Juiz de Fora
2021

Mateus Dutra Rhodes

Entropia Sequencial de Sistemas Dinâmicos

Dissertação apresentada ao Programa de Pós-Graduação em Matemática da Universidade Federal de Juiz de Fora, na área de Matemática Pura, como requisito parcial para obtenção do título de Mestre em Matemática.

Orientador: Prof. Dr. José Barbosa Gomes

Juiz de Fora

2021

Ficha catalográfica elaborada através do Modelo Latex do CDC da UFJF
com os dados fornecidos pelo(a) autor(a)

RHODES, M. D..
Entropia Sequencial de Sistemas Dinâmicos / Mateus Dutra Rhodes.
– 2021.
78 f. : il.

Orientador: José Barbosa Gomes
Dissertação de Mestrado – Universidade Federal de Juiz de Fora, Instituto de Ciências Exatas. Programa de Pós-Graduação em Matemática, 2021.

I. GOMES, J. B., orient. II. Título.

Mateus Dutra Rhodes

Entropia Sequencial de Sistemas Dinâmicos

Dissertação apresentada ao Programa de Pós-Graduação em Matemática da Universidade Federal de Juiz de Fora, na área de Matemática Pura, como requisito parcial para obtenção do título de Mestre em Matemática.

Aprovada em 30 de abril de 2021

BANCA EXAMINADORA

José Barbosa
Prof. Dr. José Barbosa Gomes - Orientador
Universidade Federal de Juiz de Fora

Mário Jorge Dias Carneiro
Prof. Dr. Mário Jorge Dias Carneiro
Universidade Federal de Minas Gerais

Prof. Dr. Magno Branco Alves
Universidade Federal de Juiz de Fora

André Junqueira da S. Corrêa
Prof. Dr. André Junqueira da Silva Corrêa
Universidade Federal de Viçosa

Sara Cristina Campos Borges
Profa. Dra. Sara Cristina Campos Borges
Universidade Federal de Juiz de Fora

AGRADECIMENTOS

Agradeço a Deus, pelo dom da vida.

Aos meus pais, família, amigos, pelo apoio e amor incondicionais.

À minha namorada, pelo carinho e companhia.

Ao Professor José Barbosa, pela orientação e contributo, e Professor Magno, pelo préstimo desde que ingressei na UFJF,

À UFJF, pela oportunidade de aprender um pouco mais.

À CAPES, pela bolsa recebida durante este período.

A todos aqueles que contribuíram direta ou indiretamente para que este trabalho fosse feito.

RESUMO

Este trabalho tem como objetivo expor a entropia sequencial, um conceito introduzido por A. G. Kushnirenko e T. N. T. Goodman, generalizado as entropias métrica e topológica clássicas, respectivamente. A entropia sequencial é um número que permite extrair mais informações sobre o comportamento de um sistema dinâmico, tanto em cenários topológicos quanto quando munidos de medida. Abordamos paralelamente algumas das propriedades das entropias métrica e topológica usuais e suas correspondentes na entropia sequencial, mas nem todas são válidas na última. Nesse escopo, apresentaremos subsídios que embasam o Princípio Variacional para entropia sequencial, além de um exemplo onde falham suas hipóteses. O ponto cardeal deste texto é um teorema de Kushnirenko, que usa entropia sequencial para caracterizar espectro discreto de um sistema dinâmico.

Palavras-chave: Entropia sequencial. Princípio Variacional.

ABSTRACT

This work aims to expose sequence entropy, a concept introduced by A. G. Kushnirenko and T. N. T. Goodman, generalizing the classical metric and topological entropies, respectively. Sequence entropy is a number that allows extracting more information about the behavior of a dynamical system, both in topological scenarios and when provided with a measure. We approach in parallel some of the usual entropy properties and their correspondent ones in sequence entropy, but not all of them are valid in the second one. In this scope, we present subsidies that base the Variational Principle for sequence entropy are discussed, besides an example where its hypotheses fail. The cardinal point of the text is a theorem by Kushnirenko, which uses sequence entropy to characterize discrete spectrum of a dynamical system.

Keywords: Sequence entropy. Variational Principle.

SUMÁRIO

1 INTRODUÇÃO

A entropia topológica foi introduzida de duas maneiras distintas: uma por Adler *et al* (1), em 1965, e a outra de maneira independente por E. Dinaburg (2), em 1970, e R. Bowen (3), em 1971. A primeira definição utiliza o número mínimo de elementos que pode ter alguma subcobertura de uma dada cobertura aberta, e a segunda, o número crescente de órbitas distinguíveis por um medidor abstrato. As duas noções parecem muito diferentes à primeira vista; porém, num espaço métrico compacto, elas são iguais.

A entropia métrica foi introduzida em 1958, por A. Kolmogorov (4), e lapidada em 1959 por Y. Sinai (5). Esta definição se baseia no conceito de C. Shannon em Teoria da Informação (6), adaptada para sistemas dinâmicos em espaços de probabilidade, sendo um indicador da informação média assintótica obtida pelo comportamento das órbitas de cada ponto no espaço. Neste contexto, V. Rokhlin (7) fez significativas contribuições.

Estes conceitos de entropia estão interligados pelo Princípio Variacional, demonstrado pela primeira vez por T. N. T. Goodman em 1971 (8).

A entropia métrica sequencial foi idealizada em 1967 por A. G. Kushnirenko (9) que, no mesmo texto, apresentou um forte resultado, o qual caracteriza sistemas com espectro discreto, a ser explicitado no capítulo 4:

Teorema 1.0.1. *Se X é compacto, μ é invariante por f e f é invertível, então o sistema (f,μ) possui espectro discreto se, e somente se, a entropia sequencial $h_{\mu,\mathcal{S}}(f)$ é nula, para toda sequência $\mathcal{S}$ de inteiros não negativos.*

O mesmo T. N. T. Goodman trouxe a noção de entropia topológica sequencial em 1973 (10), juntamente com o Princípio Variacional para entropia sequencial.

Outros estudiosos de destaque em entropia sequencial são D. Newton (11), e J. Canovas (12, 13).

Abordaremos as ideias expostas acima ao longo do texto.

No Capítulo 2, apresentamos alguns resultados conhecidos em geral.

No Capítulo 3, expõe-se as definições de entropia topológica e entropia métrica, e o Princípio Variacional, que mostra a relação entre elas.

No Capítulo 4, são dadas as definições de entropia métrica e topológica sequenciais, que generalizam as do capítulo 3, e a relação entre elas, o Princípio Variacional Sequencial, que exige restrições adicionais.

No Capítulo 5, discorremos sobre um forte teorema, que caracteriza espectro discreto utilizando entropia sequencial, além de um exemplo onde falham as hipóteses do Princípio Variacional Sequencial.

O Capítulo 6 traz conclusões e disposições finais.

2 PRELIMINARES

Nesta seção, apresentaremos alguns resultados preliminares necessários ao transcorrer do texto.

2.1 RESULTADOS GERAIS

Recapitulemos que um espaço X e uma medida μ, com domínio nos elementos da σ-álgebra boreliana $\mathcal{B}$ de modo que $\mu(X) = 1$, compõem o *espaço de probabilidade* $(X, \mathcal{B}, \mu)$. Os elementos de $\mathcal{B}$ são chamados de *borelianos* ou, em contexto mais amplo, de *mensuráveis* (14).

Lema 2.1.1 (Continuidade da medida)**.** *Sejam* $(X, \mathcal{B}, \mu)$ *um espaço de probabilidade e* $\{A_i\}_{i\in\mathbb{N}}$ *uma sequência de mensuráveis de* X*.*

1. *Se* $A_1 \subset A_2 \subset \cdots \subset A_k \subset \ldots$*, então* $\mu\left(\bigcup_{i=1}^{\infty} A_i\right) = \lim_k \mu(A_k)$.

2. *Se* $A_1 \supset A_2 \supset \cdots \supset A_k \supset \ldots$*, então* $\mu\left(\bigcap_{i=1}^{\infty} A_i\right) = \lim_k \mu(A_k)$.

Dem:. 1. Defina a sequência $(B_i)_{i\in\mathbb{N}}$ fazendo $B_1 = A_1$ e $B_k = A_k \backslash A_{k-1}$, para cada $k \in \mathbb{N}$. Note que os B_k são disjuntos e sua união é igual à união dos A_k. Assim,

$$\mu\left(\bigcup_{i=1}^{\infty} B_i\right) = \sum_{i=1}^{\infty} \mu(B_i) = \lim_n \sum_{i=2}^{n} (\mu(A_k) - \mu(A_{k-1}))) + \mu(A_1) = \lim_n \mu(A_n).$$

Logo, $\mu\left(\bigcup_{i=1}^{\infty} A_i\right) = \lim_n \mu(A_n)$.

2. Temos $\mu(A_1) < \infty$, pois $\mu(X) = 1$, e

$$\left(A_1 \Big\backslash \bigcap_{i=1}^{\infty} A_i\right) = \mu(A_1) - \mu\left(\bigcap_{i=1}^{\infty} A_i\right). \tag{2.1}$$

Por outro lado,

$$A_1 \Big\backslash \bigcap_{i=1}^{\infty} A_i = A_1 \cap \left(\bigcap_{i=1}^{\infty} A_i\right)^{\mathsf{c}} = A_1 \cap \left(\bigcup_{i=1}^{\infty} A_i^{\mathsf{c}}\right) = \bigcup_{i=1}^{\infty} (A_1 \backslash A_i).$$

Defina a sequência $(B_i)_{i\in\mathbb{N}}$ pondo $B_i = A_1 \backslash A_i$, para cada i. Então $B_1 \subset B_2 \subset \ldots B_k \subset \ldots$. Usando o item anterior,

$$\mu\left(A_1 \Big\backslash \bigcap_{i=1}^{\infty} A_i\right) = \mu\left(\bigcup_{i=1}^{\infty} (A_1 \backslash A_i)\right) = \mu\left(\bigcup_{i=1}^{\infty} B_i\right) = \lim_n \mu(B_n) = \lim_n \mu(A_1) - \lim_n \mu(A_n)$$

Por 2.1, segue que $\mu\left(\bigcap_{i=1}^{\infty} A_i\right) = \lim_n \mu(A_n)$. ■

Uma medida de probabilidade μ num espaço topológico X é dita *regular* se, dados $\varepsilon > 0$ e B mensurável, existem F fechado e A aberto, tais que $F \subset B \subset A$, e $\mu(A/F) < \varepsilon$. De maneira equivalente, tem-se $\mu(B/F) < \varepsilon$ ou $\mu(A/B) < \varepsilon$.

Seja (X, d) um espaço métrico. A *δ-vizinhança* de um subconjunto B de X, B^δ, é definida pelos pontos que estão a uma distância menor que δ de B: $B^\delta = \{x \in X; d(x, B) < \delta\}$.

Lema 2.1.2. *Seja (X, d) um espaço métrico. Dados $F \subset X$ conjunto fechado e $\varepsilon > 0$, existe função Lipschitz $\ell_\varepsilon^F : X \to [0,1]$ tal que $\ell_\varepsilon^F(x) = 1$, para todo $x \in F$, e $\ell_\varepsilon^F(x) = 0$ para todo $x \notin F^\varepsilon$.*

Dem.: Considere a função $h : \mathbb{R} \to [0,1]$ dada por: $h(x) = 1$, se $x \leq 0; h(x) = 1 - x$, se $0 \leq x \leq 1; h(x) = 0$, se $x \geq 1$.

Defina $\ell_\varepsilon^F(x) := h\left(\dfrac{d(x,F)}{\varepsilon}\right)$. Tal função é Lipschitz, por ser composição de h e a função distância a um conjunto, duas funções Lipschitz.

Um ponto x está em F se, e somente se, $d(x, F) = 0$, o que implica $\ell_\varepsilon^F(x) = 1$. Se x não está em F^ε implica $d(x, F) \geq \varepsilon$, resultando em $\ell_\varepsilon^F(x) = 0$. ■

Proposição 2.1.3 (Regularidade da medida)**.** *Se $\mu : \mathcal{B} \to \mathbb{R}$ é medida de probabilidade no espaço métrico X, então μ é regular.*

Dem.: Seja $\mathcal{G}$ o conjunto dos subconjuntos mensuráveis B tais que, dado $\varepsilon > 0$ arbitrário, existem F fechado e A aberto de modo que $F \subset B \subset A$ e $\mu(A \backslash F) < \varepsilon$.

Assim $F^c \supset B^c \supset A^c$, com $\mu(F^c \backslash A^c) = \mu(F^c \cap A) = \mu(A \backslash F) < \varepsilon$. Logo, $B \in \mathcal{G} \Rightarrow B^c \in \mathcal{G}$.

Note que $\mathcal{G}$ contém os fechados de X. De fato, pela continuidade de μ (Lema 2.1.1), se B é fechado, $B = \bigcap_\delta B^\delta$ e, por isto, $\mu(B^\delta \backslash B) \to 0$ se $\delta \to 0$.

Logo, podemos tomar $F = B$ e $A = B^\delta$, para δ suficientemente pequeno. Como complementares de abertos são fechados, $\mathcal{G}$ contém também os abertos de X.

Considere uma família $(B_n)_{n\in\mathbb{N}}$ de elementos de $\mathcal{G}$ e seja $B = \bigcup_{i=1}^{\infty} B_i$.

Dado $\varepsilon > 0$, para todo $n \in \mathbb{N}$, existem F_n fechado e A_n aberto de forma que $F_n \subset B_n \subset A_n$ e $\mu(A_n \backslash F_n) < \varepsilon/2^{n+1}$. A união $A = \bigcup_{n=1}^{\infty} A_n$ é um aberto e, para cada $k \in \mathbb{N}$, a união finita $F = \bigcup_{n=1}^{k} F_n$ é um fechado.

Novamente, pela continuidade de μ, podemos escolher k_0 suficientemente grande tal que $\mu\left(\bigcup_{n=1}^{\infty} F_n \Big\backslash F\right) < \varepsilon/2$, onde $F = \bigcup_{n=1}^{k_0} F_n$ é um fechado. Então, $F \subset B \subset A$ e

$$\mu(A\backslash F) = \mu\left(\bigcup_{n=1}^{\infty}\big((A_n\backslash F_n)\cup(F_n\backslash F)\big)\right) \leq \sum_{n=1}^{\infty}\mu(A_n\backslash F_n) + \mu\left(\bigcup_{n=1}^{\infty}F_n\Big\backslash F\right) <$$
$$< \sum_{n=1}^{\infty}\frac{\varepsilon}{2^{n+1}} + \frac{\varepsilon}{2} = \varepsilon.$$

Deste modo, B está em $\mathcal{G}$, assinalando que $\mathcal{G}$ é uma σ-álgebra que contém os fechados de X e , por consequência, contém os mensuráveis de X. ■

Denotamos $\mathcal{M}(X)$ o espaço das medidas de probabilidade borelianas em X.

Uma medida μ é dita *invariante por f*, ou, no mesmo sentido, diz-se f *preserva* μ, se $\mu(A) = \mu(f^{-1}(A))$, para todo A mensurável. A medida *imagem*, ou *iterada* de μ *por* f, $f_*\mu$, é definida pondo $f_*\mu(A) = \mu(f^{-1}(A))$, para todo A mensurável.

Assim, μ é invariante por f se, e somente se, $f_*\mu = \mu$

Chamaremos de $\mathcal{M}(X, f) \subset \mathcal{M}(X)$ o subespaço das medidas invariantes por f. Na ocasião de X ser espaço métrico compacto, $\mathcal{M}(X, f)$ é não vazio e f é contínua (19, 17).

Definição 2.1.4. Diz-se que um ponto $x \in X$ é *recorrente* para $f : X \to X$ se existe uma sequência $\{n_k\}_{k\in\mathbb{N}}$ de sorte que $f^{n_k}(x) \to x$.

Definição 2.1.5. Um ponto $x \in X$ é dito *não errante* por f se, para toda vizinhança aberta V contendo x, existe k natural de modo que $f^k(V) \cap V \neq \varnothing$. Denotamos o conjunto dos pontos não errantes por f por Ω_f.

O conjunto Ω_f é fechado: se $x \notin \Omega_f$, então existe vizinhança V de x tal que $f^n(V) \cap V = \varnothing$, para todo $n \in \mathbb{N}$. Assim, todos os pontos em V não pertencem a Ω_f.

Todo ponto recorrente é não errante, mas a recíproca nem sempre é verdadeira (15), (16).

Definição 2.1.6. O *suporte* de uma medida μ, supp μ, é o conjunto dos pontos cujas as vizinhanças têm medida não nula: supp $\mu = \{x \in X; x \in G \subset X, G \text{ aberto} \Rightarrow \mu(G) > 0\}$.

Note que todo conjunto não contido em supp μ tem medida nula.

De fato, seja A um conjunto disjunto de supp μ.

Se A é aberto, então A possui medida nula, visto que supp μ é o complementar da união de todos os abertos com medida nula.

Se A é fechado, então podemos cobrir A com $\bigcup_{x\in A} V_x$, onde cada V_x tem medida nula. Como $A \subset X$, A é compacto. Então, podemos cobrir A com $\bigcup_{i=1}^{\infty} V_{x_i}$, donde segue que $\mu(A) = 0$.

Observe, também, que o suporte de uma medida é sempre um conjunto fechado: se $x \notin \text{supp}\ \mu$, então existe V vizinhança de x com $\mu(V) = 0$. Assim, V está contida no complementar de supp μ.

O conjunto supp μ pode ser interpretado como o lugar onde "mora"a medida μ.

Teorema 2.1.7 (Recorrência de Poincaré, versão mensurável)**.** *Seja $\mu \in \mathcal{M}(X, f)$ e $E \in X$ um conjunto mensurável com medida não nula. Então, para μ quase todo ponto $x \in E$, existem infinitos valores de k para os quais $f^n(x)$ está em E.*

Dem.: Seja $E_0 \coloneqq \{x \in E; f^n(x) \notin E, \forall n\}$ o conjunto dos pontos que nunca voltam a E.

Afirma-se que as pré-imagens $f^{-n}(E_0), n \in \mathbb{N}$ são disjuntas duas a duas.

De fato, suponha que existem $1 \leq n < m$ tais que $f^{-n}(E_0) \cap f^{-m}(E_0) \neq \varnothing$. Seja x um ponto nesta interseção, e $y = f^n(x)$. Então $f^{m-n}(y) = f^m(x) \in E_0$, ou seja, y volta a E_0, contradizendo sua definição.

Por μ ser invariante por f, $\mu(f^{-n}(E_0)) = \mu(E_0)$, para todo $n \in \mathbb{N}$.

Assim, $\infty > \mu\left(\bigcup_{i=1}^{\infty} f^{-i}(E_0)\right) = \sum_{i=1}^{\infty} \mu(f^{-i}(E_0)) = \sum_{i=1}^{\infty} \mu(E_0)$, pois μ é probabilidade.

Se $\mu(E_0)$ fosse positivo, a sentença acima não seria verdadeira. Logo, $\mu(E_0) = 0$.

Agora, chamemos de $E' \coloneqq \{x \in E; f^n(x) \in E$ para finitos índices $n\}$ o conjunto dos pontos que voltam para E finitas vezes.

Por definição, todo $x \in E'$ possui um iterado por f em E_0, isto é, $E' \subset \bigcup_{k=0}^{\infty} f^{-k}(E_0)$.

Como $\mu(E_0) = 0$, e μ é invariante, $\mu(E') \leq \mu\left(\bigcup_{i=0}^{\infty} f^{-i}(E_0)\right) = \sum_{i=0}^{\infty} \mu(E_0) = 0 \quad \blacksquare$.

Dizemos que um espaço X é *separável* quando possui um subconjunto $Y \subset X$ enumerável denso em X.

Teorema 2.1.8 (Recorrência de Poincaré, versão topológica)**.** *Sejam X um espaço separável e $\mu \in \mathcal{M}(X, f)$. Então, μ quase todo ponto $x \in X$ é recorrente para f.*

Dem.: Como X é espaço métrico separável, então existe uma família enumerável de abertos $\{V_k\}_{k\in\mathbb{N}}$ que é base para X, ou seja, todo aberto de X pode ser escrito como união de elementos V_i desta família.

Para cada $i \in \mathbb{N}$, chamemos $\tilde{V}_i \coloneqq \{x \in V_i; f^n(x) \notin V_i, \forall n\}$ o conjunto dos pontos de V_i que nunca voltam a V_i.

O Teorema 2.1.7 afirma que $\tilde{V}_i$ tem medida nula, para todo i. Consequentemente, a união $\tilde{V} = \bigcup_{i=1}^{\infty} \tilde{V}_i$ tem medida nula.

Vamos mostrar que todo ponto x fora de $\tilde{V}$ é recorrente, concluindo a demonstração.

Sejam $x \in X \backslash \tilde{V}$ e G uma vizinhança qualquer de x.

Por definição, existe algum V_{k_0} tal que $x \in V_{k_0} \subset G$. Como $x \notin \tilde{V}$, $x \notin \tilde{V}_{k_0}$, ou seja, existe $n \geq 1$ de forma que $f^n(x)$ está em V_{k_0}.

Como a vizinhança G é arbitrária, conclui-se que x é um ponto recorrente. ■

Teorema 2.1.9 (Desigualdade de Jensen)**.** *Seja $\varphi : I \subset \mathbb{R} \to \mathbb{R}$ uma função côncava. Se μ é medida de probabilidade em X e $f \in L^1(\mu)$, com $\int f d\mu \in I$, então: $\varphi\left(\int f d\mu\right) \geq \int \varphi \circ f d\mu$.*

Dem.: Sejam $a = \int f d\mu$ e $T(x) = Px + Q$, a reta tangente a φ em $x = a$, com $P, Q \in \mathbb{R}$.

Tem-se $\int T \circ f d\mu = \int (Pf + Q) d\mu = Pa + Q = T(a) = \varphi(a)$.

Como, por concavidade, $\varphi \leq T$, $\varphi(a) = \int T \circ f d\mu \geq \int \varphi \circ f d\mu$. ■

Como caso particular, sejam $\varphi : \mathbb{R} \to \mathbb{R}$ e $(\lambda_i)_{i \in \mathbb{N}}$ uma sequência de números reais satisfazendo $\sum_{i=1}^{\infty} \lambda_i \leq 1$, e $\{a_i\}_{i \in \mathbb{N}}$ uma sequência limitada de números reais.

Seja o intervalo $X = [0, 1]$ munido da medida de Lebesgue e $f : X \to \mathbb{R}$ uma função da forma $\sum_{i=1}^{\infty} a_i \mathcal{X}_{A_i}$, onde os A_i são conjuntos mensuráveis disjuntos tais que $\mu(A_i) = \lambda_i$. Aplicar a desigualdade à função f resulta em:

$$\varphi\left(\sum_{i=1}^{\infty} \lambda_i a_i\right) \geq \sum_{i=1}^{\infty} \lambda_i \varphi(a_i).$$

É este um resultado que será usado ao longo do texto.

Lema 2.1.10 (Fekete)**.** *Seja $\{a_n\}_n$ uma sequência não negativa de números reais com a propriedade subaditiva, isto é, $a_{m+n} \leq a_m + a_n$, $\forall\ m, n \geq 1$. Então existe, sendo realizado por um ínfimo, o limite*

$$\lim_n \frac{a_n}{n} = \inf_n \frac{a_n}{n}.$$

Dem.: Denote $L = \inf_n \frac{a_n}{n}$.

Por definição de ínfimo, para cada $\varepsilon > 0$, é possível obter n_0 de forma que $L > \frac{a_{n_0}}{n_0} - \varepsilon$, ou $a_{n_0} < n_0(L + \varepsilon)$.

Seja $K = \max_{1 \leq i \leq n_0} a_i$. Tome $m \geq n_0$. Então existem $0 \leq r \leq n_0$ e $b \in \mathbb{N}$ tais que $m = bn_0 + r$.

Por subaditividade, $a_m = a_{bn_0+r} \leq \underbrace{a_{n_0} + a_{n_0} + \cdots + a_{n_0}}_{b \text{ vezes}} + a_r \leq ba_{n_0} + K$.

Então,

$$\frac{a_m}{m} \leq \frac{ba_{n_0}}{m} + \frac{K}{m} < \frac{bn_0(L + \varepsilon)}{m} + \frac{K}{m} \leq L + \varepsilon + \frac{K}{m}$$

o que implica $\lim_{m\to\infty} \frac{a_m}{m} \leq L + \varepsilon$.

Logo, como ε é arbitrário, vale $\lim_{m\to\infty} \frac{a_m}{m} = L = \inf_m \frac{a_m}{m}$. ■

Dado $A \subset \mathbb{N}$, denote por $\gamma_A(n)$ a cardinalidade do conjunto $A \cap \{1, \dots, n\}$.

A *densidade assintótica*, ou simplesmente *densidade*, de um conjunto $A \subset \mathbb{N}$ é o limite, se existir, $\lim_n \frac{1}{n}\#(A \cap [1, n]) = \lim_n \frac{1}{n}\gamma_A(n)$.

Note que a união finita de conjuntos de densidade zero também possui densidade zero.

Lema 2.1.11 (Koopman-Von Neumann)**.** *(17) Seja $\{a_n\}_{n\in\mathbb{N}}$ uma sequência limitada de números reais. São equivalentes:*

1. $\lim_n \frac{1}{n}\sum_{i=1}^{n} |a_i| = 0$*;*
2. *Existe um subconjunto $J \subset \mathbb{N}$ com densidade zero, tal que $\lim_n a_n = 0$, $n \notin J$.*
3. $\lim_n \frac{1}{n}\sum_{i=1}^{n} |a_i|^2 = 0$.

Dem.: $(1 \Rightarrow 2)$: Seja $J_k := \{n \in \mathbb{N}; |a_n| \geq 1/k\}$. Tem-se $J_1 \subset J_2 \subset J_3 \cdots$.

Cada um dos J_k possui densidade zero, pois $\frac{1}{n}\sum_{i=1}^{n} |a_i| \geq \frac{1}{n}\frac{1}{k}\gamma_{J_k}(n)$.

Assim, existem naturais $\ell_0 < \ell_1 < \ell_2 < \cdots$ tais que que, se $n \geq \ell_k$, então $\frac{1}{n}\gamma_{J_{k+1}}(n) < \frac{1}{k+1}$.

Ponhamos $J = \bigcup_{i=0}^{\infty} \Big[[J_{k+1} \cap [\ell_k, \ell_{k+1})\Big]$. Vamos mostrar que J tem densidade zero.

Pelo fato de $J_1 \subset J_2 \subset \cdots$, se $\ell_k \leq n < \ell_{k+1}$, então

$$J \cap [1, n] = \Big[J \cap [1, \ell_k]\Big] \cup \Big[J \cap [\ell_k, n]\Big] \subset \Big[J_k \cap [1, \ell_k)\Big] \cup \Big[J_{k+1} \cap [1, n]\Big].$$

Portanto,

$$\frac{1}{n}\gamma_J(n) = \frac{1}{n}\#(J \cap [1, n]) \leq \frac{1}{n}(\gamma_{J_k}(\ell_k) + \gamma_{J_{k+1}}(n)) \leq \frac{1}{n}(\gamma_{J_k}(n) + \gamma_{J_{k+1}}(n)) < \frac{1}{k} + \frac{1}{k+1}.$$

Consequentemente, $\frac{1}{n}\gamma_J(n) \to 0$ quando $n \to \infty$, o que significa que J tem densidade zero.

Novamente, se $\ell_k \leq n < \ell_{k+1}$ mas $n \notin J$, então $n \notin J_{k+1}$, e, por consequência, $|a_n| < 1/(k+1)$. Isto, por sua vez, implica que $\lim_n |a_n| = 0, n \notin J$.

$(2 \Rightarrow 1)$: Como $\{a_n\}_n$ é limitada, existe $A \in \mathbb{R}$ tal que $|a_n| \leq A \in \mathbb{R}^+$. Dado $\varepsilon > 0$, existe n_0 de maneira que, $n > n_0, n \notin J$ implica, simultaneamente, em $|a_n| < \varepsilon$ e $\gamma_J(n)/n < \epsilon$. Logo, se $n > n_0$,

$$\frac{1}{n}\sum_{i=1}^{n}|a_i| = \frac{1}{n}\left(\sum_{i\in J\cap[1,n]}|a_i| + \sum_{i\notin J\cap[1,n]}^{n}|a_i|\right) < \frac{A}{n}\gamma_J(n) + \varepsilon < (A+1)\varepsilon.$$

Assim, $\frac{1}{n}\sum_{i=1}^{n}|a_i| = 0$.

($1 \iff 3$): Pelos itens acima, $\frac{1}{n}\sum_{i=1}^{n}|a_i| \to 0$ se, e somente se, $\lim_n |a_i| = 0, n \notin J$, tendo J densidade zero.

Isto, por sua vez, ocorre se, e somente se $\lim_n |a_i|^2 = 0, n \notin J$. Este último equivale a dizer que $\frac{1}{n}\sum_{i=1}^{n}|a_i|^2 \to 0$. ■

Convém observar que não é necessária a hipótese de que a sequência $\{a_n\}_n$ seja limitada para provar a passagem ($1 \Rightarrow 2$).

2.2 A TOPOLOGIA FRACA*

A topologia fraca* é uma topologia útil de $\mathcal{M}(X)$. Ela é definida por uma família de seminormas formuladas com funções contínuas de X em $\mathbb{R}$, cujo espaço denotamos $C^0(X)$. Este é dual a $\mathcal{M}(X)$, pelo Teorema 2.2.1 (18).

Teorema 2.2.1 (Riesz-Markov-Kakutani)**.** *Sejam X um espaço métrico compacto e $F : C^0(X) \to \mathbb{R}$ um funcional linear positivo contínuo. Então, existe uma única medida boreliana finita μ em X tal que*

$$F(\varphi) = \int \varphi d\mu$$

Além disso, μ é medida de probabilidade se, e somente se, $F(1) = 1$.

Uma demonstração deste teorema está apresentada em (14).

Os dois resultados a seguir são centrais em Teoria da Medida; são úteis para realizar cálculos com convergência de integrais conjugado com convergência de funções. Suas demonstrações podem ser encontradas em (14).

Lema 2.2.2 (Lema de Fatou)**.** *Seja $f_n : X \to \mathbb{R}$ uma sequência de funções mensuráveis não negativas. Então*

$$\int \liminf_n f_n d\mu \leq \liminf_n \int f_n d\mu.$$

Em adição, se existe $g : X \to \mathbb{R}$ com $g \geq f_n$, para todo n, então

$$\limsup_n \int f_n d\mu \leq \int \limsup_n f d\mu.$$

Teorema 2.2.3 (Convergência Dominada)**.** *Seja $f_n : X \to \mathbb{R}$ uma sequência de funções mensuráveis, e g integrável tal que $|f_n(x)| \leq g(x)$ para μ quase todo ponto $x \in X$. Se $\{f_n\}_n$ converge em μ quase todo ponto para $f : X \to \mathbb{R}$, então $f_n, n \in \mathbb{N}$ e f são integráveis e vale:*

$$\lim_n \int f_n d\mu = \int f d\mu$$

Proposição 2.2.4 ("Bolo em Camadas")**.** *Seja $f : X \to \mathbb{R}$ uma função mensurável, não negativa e limitada, e μ uma medida boreliana em X. Então:*

$$\int f d\mu = \int_0^\infty \mu(\{x \in X; f(x) > t\})dt.$$

Dem.: Seja $B_t = \{x \in X; f(x) > t\}$. Então,

$$\int_0^\infty \mu(B_t)dt = \int_0^\infty \int \mathcal{X}_{B_t} d\mu dt = \int \int_0^\infty \mathcal{X}_{B_t} dt d\mu = \int \int_0^{f(x)} dt d\mu = \int f d\mu \qquad \blacksquare$$

Definição 2.2.5. Dada uma medida $\mu \in \mathcal{M}(X)$ e um conjunto finito $\Phi = \{\phi_1, \dots, \phi_N\}$ de funções contínuas limitadas $\phi_i : X \to \mathbb{R}$, definimos uma *$\varepsilon$-vizinhança* de μ pelo conjunto:

$$V(\mu, \Phi, \varepsilon) = \left\{\nu \in \mathcal{M}(X); \left|\int \phi_i d\nu - \int \phi_i d\mu\right| < \varepsilon, \forall\ i = 1, \dots, N\right\}$$

A *topologia fraca** é a topologia definida por estas bases de vizinhanças de cada $\mu \in \mathcal{M}(X)$. Dito de outra forma, os abertos da topologia fraca* são os conjuntos $A \subset \mathcal{M}(X)$ tais que, para todo $\mu \in A$, existe alguma vizinhança $V(\mu, \Phi, \varepsilon)$ contida em A.

Lema 2.2.6. *Uma sequência $(\mu_n)_n \subset \mathcal{M}(X)$ converge para uma medida $\mu \in \mathcal{M}(X)$ em fraca* se, e somente se,*

$$\int \phi d\mu_n \to \int \phi d\mu, \text{ para toda } \phi \in C^0(X).$$

Dem.: ($\Rightarrow$): Considere $\phi \in C^0(X)$ arbitrária e tome o conjunto $\Phi = \{\phi\}$. Como $\{\mu_n\}_n \to \mu$, dado $\varepsilon > 0$, existe $n_0 = n_0(\varepsilon)$ de sorte que $\mu_n \in V(\mu, \Phi, \varepsilon)$ para todo $n > n_0$.

Isto significa que $\left|\int \phi d\mu_n - \int \phi d\mu\right| < \varepsilon,$ para todo $n > n_0$, o que equivale a dizer que $\int \phi d\mu_n \to \int \phi d\mu$.

($\Leftarrow$) : Se ocorre $\left(\int \phi d\mu_n\right)_n \to \int \phi d\mu$, para toda $\phi \in C^0(X)$, então, dados $\Phi = \{\phi_1, \dots, \phi_N\} \subset C^0(X)$ e $\varepsilon > 0$, existem $n_1, n_2 \dots, n_N \in \mathbb{N}$ de modo que

$$\left|\int \phi_i d\mu_n - \int \phi_i d\mu\right| < \varepsilon, \text{ para todo } n > n_i.$$

Tomando $n_0 = \max\{n_1, \dots, n_N\}$, obtém-se $\mu_n \in V(\mu, \Phi, \varepsilon)$, para todo $n > n_0$. $\blacksquare$.

O conjunto B é dito *conjunto de continuidade* para a medida μ se seu bordo tem medida nula: $\mu(\partial B) = 0$.

Teorema 2.2.7. *(19, 20) Sejam $\mu \in \mathcal{M}(X)$ e $\{\mu_n\}_{n\in\mathbb{N}} \subset \mathcal{M}(X)$. São equivalentes:*

1. *$\mu_n \to \mu$ na topologia fraca*.*
2. *$\displaystyle\int \psi d\mu_n \to \int \psi d\mu$, para toda $\psi : X \to \mathbb{R}$ Lipschitz limitada.*
3. *$\limsup_n \mu_n(F) \leq \mu(F)$, para todo $F \subset X$ fechado.*
4. *$\liminf_n \mu_n(A) \geq \mu(A)$, para todo $A \subset X$ aberto.*

Dem.: $(1 \Rightarrow 2)$: Suponha que exista $\psi : X \to \mathbb{R}$ Lipschitz limitada tal que $\displaystyle\int \psi d\mu_n \not\to \int \psi d\mu$. Porém, ψ é contínua e limitada. Logo, haveria de ocorrer $\displaystyle\int \psi d\mu_n \to \int \psi d\mu$, o que é uma contradição.

$(2 \Rightarrow 3)$: Para cada fechado F e $\varepsilon > 0$, pelo Lema 2.1.2, existe função $\psi_F = \ell_\varepsilon^F$ Lipschitz de forma que $\mathcal{X}_F \leq \ell_\varepsilon^F \leq \mathcal{X}_{F^\varepsilon}$.

Então, tem-se $\mu(F) \leq \displaystyle\int \psi_F d\mu \leq \mu(F^\varepsilon)$, e $\mu_n(F) \leq \displaystyle\int \psi_F d\mu_n \leq \mu_n(F^\varepsilon)$.

Pela continuidade de μ (Lema 2.1.1), $\lim_{\varepsilon\to 0} \mu(F^\varepsilon) = \mu(F)$.

Por hipótese, $\displaystyle\lim_n \int \psi_F d\mu_n = \int \psi_F d\mu$. Assim,

$$\limsup_n \mu_n(F) = \limsup_n \int \mathcal{X}_F d\mu_n \leq \lim_n \int \psi_F d\mu_n = \int \psi_F d\mu \leq \mu(F^\varepsilon).$$

Passar ao limite com $\varepsilon > 0$ resulta em $\limsup_n \mu_n(F) \leq \mu(F)$.

$(3 \Leftrightarrow 4)$: As afirmações 3 e 4 são equivalentes por passagem ao complementar.

$(3, 4 \Rightarrow 1)$: Tome $\phi : X \to \mathbb{R}$ contínua limitada. Sem perda de generalidade, assuma $0 \leq \phi \leq 1$. Seja $B_t = \{x \in X; \phi(x) > t\}$.

Pela Proposição 2.2.4, $\displaystyle\int \phi d\mu = \int_0^1 \mu(B_t) dt$, para toda $\mu \in \mathcal{M}(X)$.

Como ϕ é contínua, B_t é aberto e, pelo item 3, $\liminf_n \mu_n(B_t) \geq \mu(B_t)$.

Pelo Lema de Fatou aplicado em $\mu_n(B_t)$:

$$\liminf_n \int \phi d\mu_n = \liminf_n \int_0^1 \mu_n(B_t) dt \geq \int_0^1 \liminf_n \mu_n(B_t) dt \geq \int_0^1 \mu(B_t) dt = \int \phi d\mu.$$

Por outro lado, de forma análoga, usando o mesmo lema para $-\phi$, tem-se

$$\liminf_n \int -\phi d\mu_n = -\limsup_n \int_0^1 \mu_n(B_t) dt \geq -\int_0^1 \limsup_n \mu_n(B_t) dt =$$
$$= \liminf_n \int_0^1 -\mu(B_t) dt = \int -\phi d\mu.$$

Logo, resulta $\displaystyle\lim_n \int \phi d\mu_n = \int \phi d\mu$. ■

Teorema 2.2.8. *(19, 20) Se uma sequência $\{\mu_n\}_{n\in\mathbb{N}} \subset \mathcal{M}(X)$ converge para uma medida $\mu \in \mathcal{M}(X)$ na topologia fraca*, então $\mu_n(B) \to \mu(B)$ para todo conjunto de continuidade mensurável B.*

Dem.: Seja B um conjunto de continuidade para μ, $\overline{B}$ seu fecho (fechado) e B^o seu interior (aberto). Então $\mu(B) = \mu(B^o) = \mu(\overline{B})$. Daí, pelos itens 3 e 4 do Teorema 2.2.7:

$$\mu(B) = \mu(B^o) \leq \liminf_n \mu_n(B^o) \leq \liminf_n \mu_n(B) \leq \limsup_n \mu_n(\overline{B}) \leq \mu(\overline{B}) = \mu(B),$$

donde se conclui que $\mu(B) = \lim_n \mu_n(B)$. ■

Proposição 2.2.9. *Sejam X um espaço métrico, $f : X \to X$ uma função mensurável, $\mu, \nu \in \mathcal{M}(X)$ com $\mu = f_*\nu$. Dada $\phi : X \to \mathbb{R}$ integrável, tem-se $\int \phi d\mu = \int \phi \circ f d\nu$. Em particular, se $\nu \in \mathcal{M}(X, f)$, então $\int \phi d\mu = \int \phi \circ f d\mu$.*

Dem.: Seja A um conjunto mensurável. Tem-se

$$\mu(A) = \int \mathcal{X}_A d\mu = f_*\nu(A) = \nu(f^{-1}(A)) = \int \mathcal{X}_{f^{-1}(A)} d\nu = \int \mathcal{X}_A \circ f d\nu.$$

Isto é, o enunciado vale para funções características de conjuntos mensuráveis.

Por linearidade, vale para funções simples $s = \sum_{i=1}^{N} a_i \mathcal{X}_{A_i}$: $\int s d\mu = \int s \circ f d\nu$.

Se $\phi : M \to \mathbb{R}$ mensurável, então ϕ é limite de uma sequência de funções simples $s_1 \leq s_2 \leq \cdots \leq \cdots \leq \phi$: $\phi(x) = \lim_n s_n(x)$, para cada $x \in X$.

Assim, aplicando o Teorema 2.2.3 duas vezes:

$$\int \phi d\mu = \int \lim_n s_n d\mu = \lim_n \int s_n d\mu = \lim_n \int s_n \circ f d\nu = \int \lim s_n \circ f d\nu = \int \phi \circ f d\nu.$$

Logo, o resultado vale para qualquer $\phi : X \to \mathbb{R}$ mensurável. ■

Proposição 2.2.10. *(19, 17) Sejam X um espaço métrico compacto , $f : X \to X$ e $\{\nu_n\}_n$ sequência em $\mathcal{M}(X)$. Então, todo ponto de acumulação da sequência $\{\mu_n\}_n$, definida por $\mu_n = \frac{1}{n}\sum_{i=0}^{n-1} f_*^i \nu_n$. é uma medida invariante.*

Dem.: Sejam $\mu = \lim_k \mu_{n_k}$ ponto de acumulação de $\{\mu_n\}_n$ e $\phi : X \to \mathbb{R}$ mensurável. Então

$$\left| \int \phi \circ f d\mu - \int \phi d\mu \right| = \lim_k \left| \int \phi \circ f d\mu_{n_k} - \int \phi d\mu_{n_k} \right|. \tag{2.2}$$

Devido ao Teorema 2.2.9, $\int \phi d\mu_{n_k} = \frac{1}{n_k} \int \sum_{i=0}^{n_k - 1} \phi \circ f^i d\nu_{n_k}$.

Logo, a diferença 2.2 se torna uma soma telescópica:

$$\lim_k \left| \frac{1}{n_k} \int \sum_{i=0}^{n_k-1} (\phi \circ f^{i+1} - \phi \circ f^i) d\nu_{n_k} \right| = \lim_k \left| \frac{1}{n_k} \int (\phi \circ f^{n_k} - \phi) d\nu_{n_k} \right|$$

O último termo é limitado por $\frac{2\|f\|_1}{n_k}$, o que implica que o limite é zero. Isto é, μ é invariante por f. $\blacksquare$.

Os pontos de acumulação mencionados na Proposição 2.2.10 existem pela compacidade de $\mathcal{M}(X)$, que decorre da compacidade de X (17), (19).

Um subconjunto $B \subset X$ é dito um *átomo* da medida μ se $\mu(B) > 0$ e $\mu(A) = 0$ para todo subconjunto A estritamente contido em B. Se o espaço X em questão for separável, então todo átomo de μ consiste num único ponto. (21)

O operador Δ denota a *diferença simétrica* entre conjuntos: $A\Delta B = (A \backslash B) \cup (B \backslash A) = (A \cup B) \backslash (A \cap B)$.

Dizemos que um conjunto mensurável $B \subset X$ é *invariante* por f se sua pré-imagem por f não difere de B, a menos de medida nula: $\mu(B \Delta f^{-1}(B)) = 0$

Uma medida invariante μ é dita *ergódica* se todo conjunto invariante por f tem medida 0 ou medida 1.

Proposição 2.2.11. *Se μ é uma medida ergódica que possui átomo no espaço de probabilidade separável X, então μ está suportada na órbita de algum átomo.*

Dem.: Seja x um átomo de μ. Como μ é invariante por f, a medida da órbita de x, $\mathcal{O}(x)$, é maior que zero.

A órbita de x não pode conter infinitos pontos, pois todos possuem medida igual e positiva. Portanto, a órbita é periódica.

Por ergodicidade, $\mu(\mathcal{O}(x)) = 1$, ou seja, o suporte de μ é a órbita de x. $\blacksquare$

3 ENTROPIA

O termo *entropia* foi cunhado em 1865 pelo físico alemão Rudolf Clausius, um dos fundadores da Termodinâmica (19, 22), como uma amálgama dos termos gregos *érgon* (trabalho, energia) e *tropé* (transformação). Clausius definiu a mudança na entropia de um corpo pela troca de calor dividida pela temperatura absoluta deste. Maxwell e Boltzmann trouxeram uma teoria atomística da troca de calor baseada em probabilidades - a mecânica estatística - trazendo significado intuitivo ao conceito de entropia (22). Claude Shannon idealizou uma definição alternativa, mas afim, de entropia, associada à informação em uma sequência de caracteres (6).

Neste capítulo, tratar-se-á acerca destes conceitos adaptados para Sistemas Dinâmicos, a saber: as entropias métrica e topológica, exibindo propriedades, exemplos, além da relação entre as duas, concretizada no Princípio Variacional.

3.1 ENTROPIA TOPOLÓGICA

Aqui, trataremos sobre duas definições de entropia topológica: uma baseada números de elementos de subcoberturas, e outra baseada na taxa de crescimento de órbitas distintas. Dentro de certas condições, as duas são equivalentes.

No decorrer desta seção, subentende-se $f : X \to X$ uma aplicação contínua, e X um espaço topológico compacto.

3.1.1 Entropia Topológica segundo Adler, Konheim e McAndrew

Este número foi introduzido por Adler *et al.* em 1965 (1).

Uma família $\alpha = (A_i)_{i\in I}$ de conjuntos é uma *cobertura* de X se $X = \bigcup\limits_{i\in I} A_i$. Uma cobertura é dita *aberta* se todos os A_i, $i \in I$, são abertos em X. Se $\alpha_1 = (A_i)_{i\in J}$, com $J \subset I$, é uma cobertura de X, diz-se que α_1 é uma *subcobertura* de α, denotando-se $\alpha_1 \subset \alpha$.

A partir daqui, pressupomos X como o conjunto coberto, e todas as respectivas coberturas sendo abertas.

Uma subcobertura é dita *minimal* se tem cardinalidade mínima.

Uma cobertura β é dita um *refinamento* de uma cobertura α se, para todo $B \in \beta$, existir $A \in \alpha$ de modo que $B \subseteq A$. Denota-se $\alpha \prec \beta$. Tal relação é uma ordem parcial ampla no conjunto das coberturas de X.

Dadas α e β duas coberturas de X, define-se a *junção* de α com β, $\alpha \vee \beta$, como a cobertura cujos elementos são interseções de elementos de α com elementos de β: $\alpha \vee \beta \coloneqq \{A \cap B; A \in \alpha, B \in \beta\}$.

É claro que $\alpha \vee \beta$ refina α e refina β, e toda cobertura refina a si mesma.

A operação $\vee$ é comutativa e associativa.

Se α, α_1, β, β_1 são coberturas e $\alpha \prec \alpha_1$, $\beta \prec \beta_1$, então $\alpha \vee \beta \prec \alpha_1 \vee \beta_1$.

Dadas α e β coberturas, a cobertura $\alpha \vee \beta$ é a cobertura menos fina que refina α e β. Isto é, se γ é cobertura que refina α e β, então $\alpha \vee \beta \prec \gamma$.

Se α é uma cobertura de X, o conjunto $f^{-1}(\alpha) := \{f^{-1}(A); A \in \alpha\}$ é, também, cobertura de X, pois f é contínua. Chamamos este conjunto de *cobertura pré imagem por* f de α. Indutivamente, definimos a *cobertura pré imagem n-ésima por* f de α como $f^{-n}(\alpha)$.

Definição 3.1.1. Seja α uma cobertura de X. Definimos a cobertura *n-ésima f-junção de* α, α_f^n, por

$$\alpha_f^n = \alpha \vee f^{-1}(\alpha) \vee \cdots \vee f^{-n+1}(\alpha) = \bigvee_{i=0}^{n-1} f^{-i}(\alpha).$$

Definição 3.1.2. A *entropia de uma cobertura* α, $H(\alpha)$, é o logaritmo da cardinalidade de uma subcobertura minimal de α, $N(\alpha)$. Isto é,

$$H(\alpha) = \log N(\alpha) = \log \inf_{\beta \subset \alpha} \{\#\beta; \beta \text{ cobre } X\}.$$

Ordinariamente, a base do logaritmo é tomada como o número 2 ou o número e. Adotaremos a segunda opção em todos os casos deste texto.

É claro que a entropia é um valor não negativo.

Proposição 3.1.3. *Dadas α e β coberturas, então:*

1. *Se $\alpha \prec \beta$, ocorre:*

   a) $f^{-1}(\alpha) \prec f^{-1}(\beta)$;

   b) $N(\alpha) \leq N(\beta)$, *o que implica* $H(\alpha) \leq H(\beta)$;

   c) $N(\alpha \vee \beta) = N(\beta)$, *o que implica* $H(\alpha \vee \beta) = H(\beta)$.

2. $f^{-1}(\alpha \vee \beta) = f^{-1}(\alpha) \vee f^{-1}(\beta)$.

3. $N(\alpha \vee \beta) \leq N(\alpha)N(\beta)$, *o que implica* $H(\alpha \vee \beta) \leq H(\alpha) + H(\beta)$.

4. $N(\alpha) \geq N(f^{-1}(\alpha))$, *o que implica* $H(\alpha) \geq H(f^{-1}(\alpha))$. *Se f é sobrejetiva, ocorrem as igualdades.*

Dem.: 1. a) Para cada $B \in \beta$, encontra-se $A \in \alpha$ tal que $B \subseteq A$. Logo, $f^{-1}(B) \subseteq f^{-1}(A) \Rightarrow f^{-1}(\alpha) \prec f^{-1}(\beta)$.

b) Seja $\{B_1, B_2, \ldots, B_{N(\beta)}\}$ uma subcobertura minimal de β. Para cada $B_i \in \beta$, existe $A_i \in \alpha$ tal que $B_i \subseteq A_i, \forall\, i = 1, \ldots, N(\beta)$.. Logo, $\{A_1, A_2, \ldots, A_{N(\beta)}\}$ é uma subcobertura de α, não necessariamente minimal. Portanto, $N(\alpha) \leq N(\beta)$. e, assim, $H(\alpha) \leq H(\beta)$.

c) É claro que $N(\beta) \leq N(\alpha \vee \beta)$. Por outro lado, seja $\{B_1, \ldots, B_r\}$ uma subcobertura minimal de β. Como $\alpha \prec \beta$, para todo B_i, $i = 1, \ldots, r$, existe $A_i \in \alpha$ tal que $B_i \subset A_i$, o que implica $B_i \cap A_i \subset A_i$. Assim, $\{B_1 \cap A_1, \ldots, B_r \cap A_r\}$ é subcobertura, não necessariamente minimal, de $\alpha \vee \beta$. Logo, $N(\alpha \vee \beta) \leq N(\beta)$.

2. Temos $f^{-1}(\alpha \vee \beta) = \{f^{-1}(A \cap B); A \in \alpha, B \in \beta\} = \{f^{-1}(A) \cap f^{-1}(B); A \in \alpha, B \in \beta\} = f^{-1}(\alpha) \vee f^{-1}(\beta)$.

3. Sejam $\{A_1, A_2, \ldots, A_{N(\alpha)}\}$ e $\{B_1, B_2, \ldots, B_{N(\beta)}\}$ subcoberturas minimais de α e β, respectivamente. Então, a cobertura $\{A_i \cap B_j; i = 1, \ldots, N(\alpha) \text{ e } j = 1, \ldots, N(\beta)\}$ é subcobertura de $\alpha \vee \beta$, não necessariamente minimal, com $N(\alpha)N(\beta)$ elementos.

   Logo, $N(\alpha \vee \beta) \leq N(\alpha)N(\beta)$ e $H(\alpha \vee \beta) = \log(N(\alpha \vee \beta)) \leq \log(N(\alpha)N(\beta)) = \log(N(\alpha)) + \log(N(\beta)) = H(\alpha) + H(\beta)$.

4. Seja $\{A_1, A_2, \ldots, A_{N(\alpha)}\}$ uma subcobertura minimal de α. Tem-se que $\{f^{-1}(A_1), f^{-1}(A_2), \ldots, f^{-1}(A_{N(\alpha)})\}$ é subcobertura, não necessariamente minimal, de $f^{-1}(\alpha)$. Logo, $N(\alpha) \geq N(f^{-1}(\alpha))$ e $H(\alpha) \geq H(f^{-1}(\alpha))$.

   Se f é sobrejetiva e $\{f^{-1}(A_1), f^{-1}(A_2), \ldots, f^{-1}(A_{N(\alpha)})\}$ não é minimal, então pode-se obter $\{f^{-1}(B_1), f^{-1}(B_2), \ldots, f^{-1}(B_m)\}$ subcobertura com $m < N(\alpha)$ elementos. Então:

$$X = f(X) = f\left(\bigcup_{i=1}^{m} f^{-1}(B_i)\right) = f\left(f^{-1}\left(\bigcup_{i=1}^{m} B_i\right)\right) = \bigcup_{i=1}^{m} B_i$$

   Logo, $\displaystyle\bigcup_{i=1}^{m} B_i$ é subcobertura de α com $m < N(\alpha)$ elementos, uma contradição. ■

Proposição 3.1.4. *Se α é uma cobertura de X, então existe o limite*

$$\lim_n \frac{1}{n} H(\alpha \vee f^{-1}(\alpha) \vee \cdots \vee f^{-n+1}(\alpha)) = \lim_n \frac{1}{n} H(\alpha_f^n).$$

Dem.: Ponha, para cada $n \geq 1$, $H_n = H(\alpha_f^n)$.

Então:

$$\begin{aligned}
H_{m+n} &= H(\alpha_f^{-m-n+1}) = H(\alpha \vee \cdots \vee f^{-m+1}(\alpha) \vee f^{-m}(\alpha) \vee \cdots \vee f^{-m-n+1}(\alpha)) \leq \\
&\leq H(\alpha \vee \cdots \vee f^{-m+1}(\alpha)) + H(f^{-m}(\alpha) \vee \cdots \vee f^{-m-n+1}(\alpha)) = \\
&= H(\alpha_f^m) + H(f^{-m}(\alpha \vee \cdots \vee f^{-n+1}(\alpha))) \leq \\
&\leq H(\alpha_f^m) + H(\alpha \vee \cdots \vee f^{-n+1}(\alpha)) = H(\alpha_f^m) + H(\alpha_f^n) = H_m + H_n.
\end{aligned}$$

Portanto, pelo Lema 2.1.10, o limite $\lim_n \frac{1}{n} H_n$ existe. ■

Definição 3.1.5. A *entropia topológica* de uma aplicação f com respeito a uma cobertura α, $h(f,\alpha)$, é definida pelo limite

$$h(f,\alpha) = \lim_n \frac{1}{n} H(\alpha_f^n).$$

Proposição 3.1.6. *Dadas α e β coberturas de X com entropia finita, então:*

1. $h(f,\alpha) \leq H(\alpha)$.
2. *Se* $\alpha \prec \beta$, *então* $h(f,\alpha) \leq h(f,\beta)$.
3. $h(f,\alpha) = h(f,\alpha_f^k)$ *para todo* $k \in \mathbb{N}$.

Dem.: 1. $h(f,\alpha) = \lim_n \frac{1}{n} H(\alpha_f^n) \leq \lim_n (H(\alpha) + H(f^{-1}(\alpha)) + \cdots + H(f^{-n+1}(\alpha)) \leq$
$\leq \lim_n \frac{1}{n} \underbrace{(H(\alpha) + \cdots + H(\alpha))}_{n \text{vezes}} = \lim_n H(\alpha) = H(\alpha)$.

2. Como $\alpha \prec \beta$, então $f^{-n}(\alpha) \prec f^{-n}(\beta)$, para todo $n \in \mathbb{N}$. Disto, segue $\alpha_f^n \prec \beta_f^n$. Logo, $h(f,\alpha) = \lim_n \frac{1}{n} H(\alpha_f^n) \leq \lim_n \frac{1}{n} H(\beta_f^n) = h(f,\beta)$.

3. Observe que

$$(\alpha_f^k)_f^n = (\alpha_f^k \vee \cdots \vee f^{-n+1}(\alpha_f^k)) = (\alpha \vee \cdots \vee f^{-k+1}(\alpha) \bigvee f^{-1}(\alpha) \vee \cdots \vee f^{-k}(\alpha) \bigvee \ldots$$
$$\cdots \bigvee f^{-n+1}(\alpha) \vee \cdots \vee f^{-n-k+2}(\alpha)) = \alpha_f^{n+k-1}.$$

Portanto, $h(f,\alpha_f^k) = \lim_n \frac{1}{n} H(\alpha_f^{n+k-1}) = \lim_n \frac{1}{n} H(\alpha_f^n) = h(f,\alpha)$. ■

Definição 3.1.7. A *entropia topológica segundo Adler, Konheim e McAndrew de uma aplicação* f, $h(f)$, é definida como o supremo do conjunto $\{h(f,\alpha); \alpha$ é cobertura com entropia finita$\}$:

$$h(f) = \sup_\alpha h(f,\alpha).$$

Definição 3.1.8. Dizemos que duas aplicações $f : A \to A$ e $g : B \to B$ são *topologicamente conjugadas* se existe $\psi : A \to B$ homeomorfismo tal que $g = \psi^{-1} \circ f \circ \psi$. ψ é chamada de *conjugação topológica* de f e g.

Neste contexto, omitiremos o símbolo $\circ$ entre funções compostas: $g = \psi^{-1} f \psi$.

Teorema 3.1.9. *A entropia topológica de uma aplicação é um invariante topológico, isto é, duas aplicações topologicamente conjugadas têm a mesma entropia topológica.*

Dem.: Sejam $f, g : X \to X$ duas funções topologicamente conjugadas, α uma cobertura de X e ψ uma conjugação de f e g.

Note que $g^{-k}(\psi^{-1}(\alpha)) = (\psi^{-1} f \psi)^{-k}(\psi^{-1}(\alpha)) = \psi^{-1} f^{-k} \psi \psi^{-1}(\alpha)) = \psi^{-1} f^{-k}(\alpha)$, para todo $k \in \mathbb{N}$.

Então:

$$\begin{aligned}
&h(g,\psi^{-1}(\alpha)) = \\
&= \lim_n \tfrac{1}{n} H(\psi^{-1}(\alpha) \vee g^{-1}(\psi^{-1}(\alpha)) \vee \cdots \vee g^{-n+1}(\psi^{-1}(\alpha)) = \\
&= \lim_n \tfrac{1}{n} H(\psi^{-1}(\alpha) \vee \psi^{-1} f^{-1}(\alpha) \vee \cdots \vee \psi^{-1} f^{-n+1}(\alpha)) = \\
&= \lim_n \tfrac{1}{n} H(\psi^{-1}(\alpha \vee f^{-1}(\alpha) \vee \cdots \vee f^{-n+1}(\alpha)) = \\
&= \lim_n \tfrac{1}{n} H(\alpha \vee \cdots \vee f^{-n+1}(\alpha)) = h(f,\alpha)
\end{aligned}$$

Quando α varia sobre as coberturas de X, $\psi^{-1}(\alpha)$ também varia sobre as coberturas de X, pois ψ é homeomorfismo.

Então, $h(g) = \sup_\alpha \{h(g,\psi^{-1}(\alpha)\} = \sup_\alpha \{h(f,\alpha)\} = h(f)$. ■

Proposição 3.1.10. *$h(f^k) = kh(f)$, para todo $k \in \mathbb{N}$.*

Dem.: Dada α cobertura, e $k \in \mathbb{N}$:

$$\begin{aligned}
&h(f^k) \geq h(f^k, \alpha_f^k) = \lim_n \frac{1}{n} H(\alpha_f^k \vee f^{-k}(\alpha_f^k) \vee \cdots \vee f^{(-n+1)k}(\alpha_f^k)) = \\
&= k \lim_n \frac{1}{nk} H\Big((\alpha \vee \cdots \vee f^{-k+1}(\alpha) \bigvee f^{-k}(\alpha) \vee \cdots \vee f^{-2k+1}(\alpha)) \bigvee \ldots \\
&\cdots \bigvee f^{(-n+1)k}(\alpha) \vee \cdots \vee f^{-nk+1}(\alpha)\Big) = kh(f,\alpha).
\end{aligned}$$

Como $\alpha \prec \alpha_f^k$, temos $h(f^k) \geq kh(f,\alpha) = h(f^k, \alpha_f^k) \geq h(f^k, \alpha)$.

Tomando o supremo sobre as coberturas α, resulta em $h(f^k) \geq kh(f) \geq h(f^k)$, o que implica $kh(f) = h(f^k)$. ■

O *diâmetro de um conjunto* A é definido por diam $A := \sup_{x,y \in A} \{d(x,y)\}$.

O *diâmetro de uma cobertura* α é definido por diam $\alpha := \sup_{A \in \alpha} \{\text{diam } A\}$.

Definição 3.1.11. Seja α uma cobertura. O número $\varepsilon > 0$ dito *número de Lebesgue* de α se for o maior número tal que toda bola de diâmetro menor ou igual a ε está contida em algum elemento de α.

Proposição 3.1.12. *Seja α uma cobertura aberta de (X,d) espaço métrico compacto. Então α admite um número de Lebesgue.*

Dem.: Suponha que $\alpha = (A_i)_{i \in I}$ não admita número de Lebesgue.

Assim, pode-se obter, para cada $n \in \mathbb{N}$, uma bola aberta $B_n = B(x, 1/n)$, tal que nenhum elemento de α contenha B_n.

Para cada n, escolhamos um ponto $y_n \in B_n$.

Passando a alguma subsequência, se necessário, temos que $\lim_n y_n = y \in X$, pois X é compacto.

Tem-se, ainda, $y \in A_{i_0}$ para algum $i_0 \in I$. Como α é aberta, existe $\varepsilon > 0$ tal que $B(y, \varepsilon) \subset A_{i_0}$.

Escolha $n_0 \in \mathbb{N}$ tal que $\dfrac{1}{n_0} < \dfrac{\varepsilon}{2}$ e $d(y, y_{n_0}) < \dfrac{\varepsilon}{2}$.

Então, se $z \in B_{n_0} \Rightarrow d(y, z) \leq d(y, y_{n_0}) + d(y_{n_0}, z) < \dfrac{1}{n_0} + \dfrac{\varepsilon}{2} < \varepsilon$.

Ou seja, $B_{n_0} \subset B(y, \varepsilon) \subset A_{i_0}$, uma contradição. Logo, α admite número de Lebesgue. ■

Proposição 3.1.13. *Seja* $\{\alpha_k\}_{k\in\mathbb{N}}$ *uma sequência de coberturas de* X *tais que* diam α_k converge para zero. Então $h(f) = \sup_k h(f, \alpha_k) = \lim_k h(f, \alpha_k)$.

Dem.: Seja β uma cobertura qualquer de X com número de Lebesgue $\varepsilon > 0$.

Escolha n_0 de modo que diam $\alpha_k < \varepsilon$, para todo $k \geq n_0$.

Disto segue que todo elemento de α_k está contido em algum elemento de β, ou seja, $\beta \prec \alpha_k$.

Portanto, $h(f, \beta) \leq h(f, \alpha_k)$, o que implica $h(f) \leq \liminf_k h(f, \alpha_k)$.

Por outro lado, $h(f) \geq \sup_k h(f, \alpha_k) \geq \limsup_k h(f, \alpha_k)$. Combinando estas afirmações, obtemos $h(f) = \lim_k h(f, \alpha_k)$. ■

Exemplo 3.1.14. (23) Seja $\mathcal{A} = \{1, \ldots, k\}$ um conjunto, e $\mathcal{A}^{\mathbb{N}}$ o espaço formado por sequências unilaterais infinitas, constituídas por elementos de $\mathcal{A}$: $\mathcal{A}^{\mathbb{N}} = \{x_0x_1x_2\ldots; x_i \in \mathcal{A}, \forall i = 0, 1, 2, \ldots\}$.

Seja, também, $X = (\mathcal{A}^{\mathbb{N}}, d)$ um espaço métrico, munido com a métrica $d(x, y) = \sup_n\{2^{-n}; x_n \neq y_n\}$.

Este espaço é compacto, pois a topologia gerada por d coincide com a topologia produto; e nesta, há compacidade, pelo teorema de Tychonoff (25).

Denotamos o cilindro $[m, x_1, \ldots, x_n]$ como conjunto dos elementos de $\mathcal{A}^{\mathbb{N}}$ que têm $m - n$ elementos fixados, iguais a $x_1, \ldots, x_n$, nesta ordem, a partir da posição m.

Considere a aplicação *shift* $\sigma : \mathcal{A}^{\mathbb{N}} \to \mathcal{A}^{\mathbb{N}}$, pondo $\sigma((x_n)_n) = (x_{n+1})_n$, ou seja, $\sigma(x_0x_1x_2\ldots) = x_1x_2x_3\ldots$.

Seja $\alpha = \{A_1, ..., A_k\}$ uma cobertura de X, $A_i = \{x \in \mathcal{A}; x_0 = i\}, i = 1\ldots, k$. Observe que α_σ^n é a cobertura por cilindros disjuntos $[0; x_1, ..., x_k]$ com n elementos fixados, e $N(\alpha_\sigma^n) = k^n$, para todo n, pois os elementos de α_σ^n são disjuntos entre si.

Seja a sequência de coberturas $\{\alpha_i\}_i = \{\alpha_\sigma^i\}_i$. Ocorre $h(\sigma, \alpha_i) = h(\sigma, \alpha_\sigma^i) = h(\sigma, \alpha)$, para todo $i \in \mathbb{N}$, pelo item 3.1.6-3. Note que diam $\alpha_i \to 0$.

Ainda, $h(\sigma, \alpha_i) = \lim_n \frac{1}{n} H(\alpha_\sigma^{n+i+1}) = \lim_n \frac{1}{n}(n+i-1)\log k = \log k$, para todo $i \in \mathbb{N}$. Logo, pelo item 3.1.13, $h(\sigma) = \lim_n h(\sigma, \alpha_n) = \log k$.

3.1.2 Entropia Topológica segundo Bowen e Dinaburg

Esta definição de entropia é, a princípio, muito distinta da noção apresentada por Adler *et al*: está relacionada com o crescimento do número de segmentos de órbitas distintos, à medida que o número de iterados de tais segmentos cresce, e a resolução de um aparelho medidor que as distingue tende a zero (2, 3).

Nesta subseção, assumimos que X está munido com uma métrica d.

Definição 3.1.15. Sejam $x \in X$, $n \in \mathbb{N}$ e $\varepsilon > 0$. Definimos a *bola dinâmica* de centro x, raio ε e comprimento n, $B(x, n, \varepsilon)$, por

$$\begin{aligned} B(x,n,\varepsilon) &\coloneqq \{y \in X; d(f^m(x), f^m(y)) < \varepsilon \ \forall\ 0 \leq m \leq n-1\} = \\ &= \{y \in X; f^m(y) \in B(f^m(x), \varepsilon) \ \forall\ 0 \leq m \leq n-1\} = \\ &= \bigcap_{i=0}^{n-1} f^{-i}(B(f^i(x), \varepsilon)). \end{aligned}$$

Definição 3.1.16. Um conjunto $S \in X$ é dito (n, ε)-*separado* se, para todos $x, y \in S$ distintos, existe $0 \leq k \leq n-1$ inteiro, de modo que $d(f^k(x), f^k(y)) > \varepsilon$.

Intuitivamente, um conjunto (n, ε)-separado S é aquele tal que um aparelho de observação com resolução ε consegue distinguir quaisquer dois segmentos de f-órbitas com comprimento n de dois pontos nele distintos (24).

Definição 3.1.17. Um conjunto $R \subset X$ é dito (n, ε)-*gerador* se, para todo $x \in X$, existe $y \in R$ tal que, para todo $0 \leq k \leq n-1$ inteiro, se tenha $d(f^k(x), f^k(y)) \leq \varepsilon$. Isto equivale a dizer que o conjunto $\bigcup_{y \in R} B(y, n, \varepsilon)$ cobre X.

Intuitivamente, um conjunto (n, ε)-gerador R é aquele tal que um aparelho de observação com resolução ε não consegue distinguir o segmento da f-órbita com comprimento n de qualquer ponto do espaço do segmento da f-órbita com comprimento n de algum ponto do conjunto (24).

Chamamos de $s_n(f, \varepsilon)$ o máximo das cardinalidades de todos os conjuntos (n, ε)-separados, e de $r_n(f, \varepsilon)$ o mínimo das cardinalidades de todos os conjuntos (n, ε)-geradores.

Defina $r(f, \varepsilon) \coloneqq \limsup_n \frac{1}{n} \log(r_n(f, \varepsilon))$, e $s(f, \varepsilon) \coloneqq \limsup_n \frac{1}{n} \log(s_n(f, \varepsilon))$.

Defina, também, $r(f) \coloneqq \lim_{\varepsilon \to 0} r(f, \varepsilon)$ e $s(f) \coloneqq \lim_{\varepsilon \to 0} s(f, \varepsilon)$.

Observação 3.1.18. Se dado conjunto $S \subset X$ é (n, ε)-separado maximal, então S é (n, ε)-gerador. De fato, dado $y \in X \backslash S$, $S \cup \{y\}$ não é (n, ε)-separado, pois S tem cardinalidade máxima. Portanto, existem $x \in S$ e $0 \leq m \leq n-1$ tais que $d(f^m(x), f^m(y)) \leq \varepsilon$, isto é, S é um conjunto (n, ε)-gerador.

Observação 3.1.19. O espaço X pode ser coberto por um número finito, digamos, $k \in \mathbb{N}$ de bolas dinâmicas de raio ε e comprimento n, por ser compacto: $X = \bigcup_{i=1}^{k} B(x_k, n, \varepsilon)$. Assim, o conjunto $\{x_1, \ldots, x_k\}$ é (n, ε)-gerador, e $r_n(f, \varepsilon) \leq k < \infty$.

As grandezas $s_n(f, \varepsilon)$ e $r_n(f, \varepsilon)$ são monótonas com relação a variável ε, isto é, se $\varepsilon > \varepsilon_1 > 0$, então $s_n(f, \varepsilon_1) \geq s_n(f, \varepsilon)$, e $r_n(f, \varepsilon_1) \geq r_n(f, \varepsilon)$. Consequentemente, $s(f, \varepsilon) \geq s(f, \varepsilon_1)$, e $r(f, \varepsilon) \geq r(f, \varepsilon_1)$.

Lema 3.1.20. *Para todos $n \geq 1$ e $\varepsilon > 0$, tem-se $r_n(f, \varepsilon) \leq s_n(f, \varepsilon) \leq r_n(f, \varepsilon/2)$.*

Dem.: Seja S um conjunto (n, ε)-separado maximal.

Pela observação 3.1.18, S é (n, ε)-gerador.

Logo, $r_n(f, \varepsilon) \leq \#S = s_n(f, \varepsilon)$.

Para a outra desigualdade, seja S um conjunto (n, ε)-separado e R um conjunto $(n, \varepsilon/2)$-gerador. Dado $x \in S$ arbitrário, existe $y \in R$ tal que $d(f^m(x), f^m(y)) \leq \varepsilon/2$, para todo $0 \leq m \leq n-1$.

Defina $\varphi : S \to R$ pondo $\varphi(x) = y$, como definido no parágrafo anterior. Tal aplicação é injetiva. De fato, sejam $x, z \in S$ tais que $\varphi(x) = \varphi(z) = y$. Então:

$$d(f^m(x), f^m(z)) \leq d(f^m(x), f^m(y)) + d(f^m(y), f^m(z)) \leq \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon,$$

para todo $0 \leq m \leq n-1$. Pelo fato de S ser (n, ε)-separado, $x = z$.

Disto segue que $\#S \leq \#R$, o que implica $s_n(f, \varepsilon) \leq r_n(f, \varepsilon/2)$. ■

Pelo Lema 3.1.20 e pelas observações 3.1.18 e 3.1.19, segue que $s_n(f, \varepsilon)$ é finito.

Dado $\varepsilon > 0$, o Lema 3.1.20 implica $r(f, \varepsilon) \leq s(f, \varepsilon) \leq r(f, \varepsilon/2)$.

Tomando o limite com $\varepsilon \to 0$, obtemos $r(f) = s(f)$.

Definição 3.1.21. A *entropia topológica segundo Bowen e Dinaburg de uma aplicação f* é definida por $r(f) = s(f)$.

Observação 3.1.22. A noção de entropia segundo a Definição 3.1.21 para uma métrica d_1 é a mesma que para outra métrica d_2, desde que estas sejam topologicamente equivalentes.

De fato, dado $\varepsilon > 0$, pode se obter $\delta = \delta(\varepsilon) > 0$ de sorte que $B_{d_2}(x, \delta) \subset B_{d_1}(x, \varepsilon)$, para todo $x \in X$.

Isto significa que todo conjunto (n, ε)-separado na métrica d_1 é (n, δ)-separado na métrica d_2, resultando em $s_n(f, \varepsilon, d_1) \leq s_n(f, \delta, d_2)$.

Daí, obtemos $s(f,d_1) \leq s(f,d_2)$.

De modo análogo, trocando-se os papéis de d_1 e d_2, depreende-se que $s(f,d_2) \leq s(f,d_1)$.

Proposição 3.1.23. *A entropia segundo Bowen e Dinaburg é a mesma entropia segundo Adler, Konheim e McAndrew, isto é,* $r(f) = h(f) = s(f)$.

Dem.: Pelo Lema 3.1.20, basta mostrar que $s(f) \leq h(f) \leq r(f)$.

Fixemos $\varepsilon > 0$ e $n \in \mathbb{N}$. Seja $S \subset X$ um conjunto (n,ε)-separado maximal e α cobertura aberta, com diâmetro menor que ε. Se x e y estão no mesmo elemento de α_f^n, então

$$d(f^m(x), f^m(y)) \leq \text{diam } \alpha < \varepsilon, \ \forall\ 0 \leq m \leq n-1.$$

Em particular, cada elemento de α_f^n não contém mais que um elemento de S, pois $x, y \in S$ resulta em $d(f^m(x), f^m(y)) > \varepsilon$ para algum $0 \leq m \leq n-1$.

Portanto, $s_n(f,\varepsilon) \leq N(\alpha_f^n)$, para todo $n \in \mathbb{N}$. Então

$$s(f,\varepsilon) = \limsup_n \frac{1}{n} \log s_n(f,\varepsilon) \leq \lim_n \frac{1}{n} \log N(\alpha_f^n) = h(f,\alpha) \leq h(f).$$

Fazendo $\varepsilon \to 0$, obtemos $s(f) \leq h(f)$.

Por outro lado, sejam α uma cobertura de X e $\varepsilon > 0$ o número de Lebesgue de α. Seja $R \subset X$ um conjunto $(n,\varepsilon/2)$-gerador com cardinalidade mínima. Para cada $x \in R$ e $0 \leq m \leq n-1$, existe $A_m^x \in \alpha$ tal que $B(f^m(x), \varepsilon/2) \subset A_m^x$.

Assim, $B(x,n,\varepsilon/2) \subset \bigcap\limits_{i=0}^{n-1} f^{-i}(A_i^x)$.

Pelo fato de R ser $(n,\varepsilon/2)$-gerador, o conjunto $\{B(x,n,\varepsilon/2); x \in R\}$ é uma cobertura de X. Também o é, então, o conjunto $\beta = \left\{\bigcap\limits_{i=0}^{n-1} f^{-i}(A_i^x); x \in R\right\}$, que é uma subcobertura de α_f^n com, no máximo, $\#R$ elementos.

Segue disto que $N(\alpha_f^n) \leq \#R = r_n(f,\varepsilon/2)$, para todo $n \in \mathbb{N}$.

Portanto, $h(f,\alpha) = \lim\limits_n \frac{1}{n} H(\alpha_f^n) \leq \limsup\limits_n \frac{1}{n} \log r_n(f,\varepsilon/2) = r(f,\varepsilon/2)$.

Fazendo $\varepsilon \to 0$, obtemos $h(f,\alpha) \leq r(f) \Rightarrow h(f) \leq r(f)$. ■

Exemplo 3.1.24. Seja f é uma isometria, isto é, $d(f(x), f(y)) = d(x,y)$, para todo $x, y \in X$. Então, a cardinalidade dos conjuntos (n,ε)-geradores é constante em relação a n, pois $d(f^k(x), f^k(y)) = d(x,y)$, para todo x, y. Logo, a entropia de f é zero:

$$s(f,\varepsilon) = \limsup_n \frac{1}{n} \log(r_n(f,\varepsilon)) = 0, \text{ o que implica } h(f) = \lim_{\varepsilon \to 0} s(f,\varepsilon) = 0$$

Exemplo 3.1.25. Seja $f : S^1 \to S^1$ um homeomorfismo e α uma cobertura de S^1 formada por finitos intervalos.

Denote $\partial\alpha$ o conjunto formado pelos extremos dos intervalos de α.

Para cada n, a cobertura α_f^n é formada por intervalos com extremos em $\partial\alpha_f^n = \partial\alpha \cup f^{-1}(\partial\alpha) \cup \cdots \cup f^{-n+1}(\partial\alpha)$.

Note que não há mais elementos em α_f^n do que em $\partial\alpha_f^n$. Além disso, o número de elementos em $\partial\alpha_f^n$ não excede $\#\partial\alpha$, pois, a cada iteração pela pré imagem de f em $\partial\alpha$, é acrescentado, no máximo, o número de elementos de $\partial\alpha$. Portanto, $\#\alpha_f^n \leq \#\partial\alpha_f^n \leq n\#\partial\alpha$.

Daí, $h(f,\alpha) = \lim_n \frac{1}{n} H(\alpha_f^n) \leq \liminf_n \frac{1}{n}\log(\#\alpha_f^n) \leq \liminf_n \frac{1}{n}\log(n\#\partial\alpha) = 0$

Considerando $\{\alpha_k\}_{k\in\mathbb{N}}$ sequência de coberturas finitas com diam $\alpha_k < 1/k$, temos $h(f,\alpha_k) = 0$, para todo k. Pela Proposição 3.1.13, $h(f) = \sup_k h(f,\alpha_k) = 0$.

3.2 ENTROPIA MÉTRICA, OU ENTROPIA DE KOLMOGOROV-SINAI

Esta definição de entropia foi confeccionada por Andrey Kolmogorov e Yakov Sinai (19, 22), adaptada do conceito de Shannon em Teoria da Informação (6).

Consideremos a seguinte situação: um canal de comunicação que transmite símbolos de um alfabeto $\mathcal{A}$ em sucessão. Cada símbolo $a \in \mathcal{A}$ possui uma probabilidade p_a de ser utilizado, independente dos outros símbolos.

Denotamos por $\mathcal{A}^n$ o conjunto das palavras de comprimento n.

No contexto de Shannon, a informação fornecida pelo caracter a é dada por $I(a) = -\log p_a$, e a informação fornecida pela palavra $a_1, \ldots, a_k$ é dada por $I(a_1a_2\ldots a_k) = -\log(a_1\ldots a_k) = -\log(p_{a_1}\ldots p_{a_k})$. Esta formulação perfaz duas propriedades intuitivas:

- Quanto maior a probabilidade de um caracter ocorrer, menor a informação fornecida por ele: se for solicitado a alguém que adivinhe uma palavra que começa com a letra A, a probabilidade de acerto é menor do que se a palavra começar com a letra Z.
- A informação fornecida por uma palavra é a soma das informações individuais dos caracteres que compõem a palavra.

No mesmo contexto, a entropia, ou informação média, do alfabeto $\mathcal{A}$ é dada por $I(\mathcal{A}) = \sum_{a\in\mathcal{A}} -p_a \log p_a$.

Por fim, a entropia, ou informação, do canal de comunicação é definida pelo informação média fornecida pelo conjunto das palavras de comprimento igual a n, quando n tende ao infinito:

$$I = \lim_n \frac{1}{n} I(\mathcal{A}^n)$$

Kolmogorov e Sinai aplicaram conceitos parecidos em Sistemas Dinâmicos, empregando partições do espaço de estados em questão.

Aqui, subentende-se $(X, \mathcal{B}, \mu)$ é um espaço de probabilidade, com X um espaço métrico compacto, $\mathcal{B}$ a σ-álgebra de Borel, e μ uma medida de probabilidade boreliana, além de $f : X \to X$ mensurável.

Uma família $\mathcal{P} = \{P_i\}_{i\in\mathbb{N}}$ de subconjuntos mensuráveis de X disjuntos dois a dois é dita uma *partição* de X se $\mu\left(\bigcup_{i=1}^{\infty} P_i\right) = 1$. Dado $x \in X$, denotamos $\mathcal{P}(x)$ o elemento de $\mathcal{P}$ que contém x.

A partir daqui, pressupomos X como o conjunto particionado.

Uma partição $\mathcal{Q}$ é dita um *refinamento* de uma partição $\mathcal{P}$ se, para todo $Q \in \mathcal{Q}$, existir $P \in \mathcal{P}$ tal que $Q \subseteq P$, a menos de medida nula. Denota-se $\mathcal{P} \prec \mathcal{Q}$. Tal relação é uma ordem parcial ampla no conjunto das partições de X.

Dadas $\mathcal{P}$ e $\mathcal{Q}$ duas partições, define-se a *junção* de $\mathcal{P}$ com $\mathcal{Q}$, $\mathcal{P} \vee \mathcal{Q}$, como a partição cujos elementos são interseções de elementos de $\mathcal{P}$ com elementos de $\mathcal{Q}$: $\mathcal{P} \vee \mathcal{Q} := \{P \cap Q; P \in \mathcal{P}, Q \in \mathcal{Q}\}$.

É claro que $\mathcal{P} \vee \mathcal{Q}$ refina $\mathcal{P}$ e refina $\mathcal{Q}$, e toda partição refina a si mesma.

A operação $\vee$ é comutativa e associativa.

Se $\mathcal{P}$, $\mathcal{P}_1$, $\mathcal{Q}$, $\mathcal{Q}_1$ são partições, e $\mathcal{P} \prec \mathcal{P}_1$, $\mathcal{Q} \prec \mathcal{Q}_1$, então $\mathcal{P} \vee \mathcal{Q} \prec \mathcal{P}_1 \vee \mathcal{Q}_1$.

Dadas $\mathcal{P}$ e $\mathcal{Q}$ partições, a partição $\mathcal{P} \vee \mathcal{Q}$ é a partição menos fina que refina $\mathcal{P}$ e $\mathcal{Q}$. Isto é, se $\mathcal{R}$ é partição que refina $\mathcal{P}$ e $\mathcal{Q}$, então $\mathcal{P} \vee \mathcal{Q} \prec \mathcal{R}$.

Se $\mathcal{P}$ é uma partição, o conjunto $f^{-1}(\mathcal{P}) := \{f^{-1}(P); P \in \mathcal{P}\}$ é uma partição, pois f é mensurável. Chamamos este conjunto de *partição pré imagem por f* de $\mathcal{P}$. Indutivamente, definimos a *partição pré imagem n-ésima por f* de $\mathcal{P}$ como $f^{-n}(\mathcal{P})$.

Definição 3.2.1. Definimos a partição *n-ésima f-junção de $\mathcal{P}$* pondo

$$\mathcal{P}_f^n = \mathcal{P} \vee f^{-1}(\mathcal{P}) \vee \cdots \vee f^{-n+1}(\mathcal{P}) = \bigvee_{i=0}^{n-1} f^{-i}(\mathcal{P}).$$

Dizemos que duas partições $\mathcal{P}$ e $\mathcal{Q}$ são *independentes* se $\mu(P \cap Q) = \mu(P)\mu(Q)$, para todo $P \in \mathcal{P}$ e todo $Q \in \mathcal{Q}$.

Definição 3.2.2. A cada partição $\mathcal{P}$ e medida μ, associamos a uma *função de informação* $I_{\mathcal{P}}^{\mu} : X \to \mathbb{R}$, dada por $I_{\mathcal{P}}^{\mu}(x) = -\log \mu(\mathcal{P}(x))$.

Tal função é mensurável, pois é composição de três funções mensuráveis:

- $\pi : X \to \mathcal{P}$, dada por $\pi(x) = \mathcal{P}(x)$;
- $g : \mathcal{P} \to [0, 1]$, dada por $g(P) = \mu(P)$;
- $f : [0, 1] \to \mathbb{R}$, dada por $f(x) = -\log(x)$.

Assim, pode-se falar na integral desta.

Definição 3.2.3. A *entropia*, ou *informação média*, de uma partição $\mathcal{P}$ *com respeito a uma medida* μ, $H_\mu(\mathcal{P})$, é o valor médio da função de informação, isto é,

$$H_\mu(\mathcal{P}) = \int I^\mu_\mathcal{P} d\mu = \sum_{P\in\mathcal{P}} -\mu(P)\log\mu(P).$$

Nesta seção, denotaremos por ϕ a função $x \to -x\log x$, com domínio no intervalo $[0,1]$. Faremos a convenção $0\log 0 = \lim\limits_{x\to 0} x\log x = 0$. Note que ϕ é côncava, pois sua segunda derivada é sempre menor que zero.

Definição 3.2.4. A *entropia condicional* de uma partição $\mathcal{P}$ com relação a outra partição $\mathcal{Q}$, $H_\mu(\mathcal{P}|\mathcal{Q})$ é definida por

$$H_\mu(\mathcal{P}|\mathcal{Q}) = \sum_{P\in\mathcal{P}}\sum_{Q\in\mathcal{Q}} -\mu(P\cap Q)\log\frac{\mu(P\cap Q)}{\mu(Q)}.$$

Esta pode ser escrita de outra forma:

$$\begin{aligned} H_\mu(\mathcal{P}|\mathcal{Q}) &= \sum_{P\in\mathcal{P}}\sum_{Q\in\mathcal{Q}} -\mu(P\cap Q)\log\mu(P\cap Q) + \sum_{P\in\mathcal{P}}\sum_{Q\in\mathcal{Q}} \mu(P\cap Q)\log\mu(Q) = \\ &= H_\mu(\mathcal{P}\vee\mathcal{Q}) - \sum_{Q\in\mathcal{Q}}\log\mu(Q)\sum_{P\in\mathcal{P}} -\mu(P\cap Q) = H_\mu(\mathcal{P}\vee\mathcal{Q}) - H_\mu(\mathcal{Q}). \end{aligned}$$

O número $H_\mu(\mathcal{P}|\mathcal{Q})$ pode ser visto como a informação adicional oferecida por $\mathcal{P}$ quando já havia a informação dada por $\mathcal{Q}$.

Lema 3.2.5. *Se $\mathcal{P}$ é uma partição finita, então $H_\mu(\mathcal{P}) \leq \log\#\mathcal{P}$, isto é, toda partição finita tem entropia finita.*

Dem.: Seja $\mathcal{P} = \{P_1,\ldots,P_k\}$ uma partição. Como a função ϕ é côncava, utilizamos a desigualdade 2.1.9:

$$\frac{1}{k}H_\mu(\mathcal{P}) = \sum_{i=1}^{k}\frac{1}{k}\phi(\mu(P_i)) \leq \phi\left(\sum_{i=0}^{k}\frac{1}{k}\mu(P_i)\right) = \phi\left(\frac{1}{k}\right) = \frac{\log k}{k},$$

o que implica em $H_\mu(\mathcal{P}) \leq \log k$. ■

Proposição 3.2.6. *Se μ e ν são probabilidades, então, para $t\in[0,1]$ e $\mathcal{P}$ é uma partição qualquer,*

$$H_{t\mu+(1-t)\nu}(\mathcal{P}) \geq tH_\mu(\mathcal{P}) + (1-t)H_\nu(\mathcal{P}).$$

Mais geralmente, se $a_1,\ldots,a_n$ são números não negativos tais que $a_1+\cdots+a_n = 1$ e $\mu_1,\ldots,\mu_n$ são probabilidades, então

$$H_{\sum\limits_i a_i\mu_i}(\mathcal{P}) \geq \sum_i a_i H_{\mu_i}(\mathcal{P}).$$

Dem.: Por definição, $H_{t\mu+(1-t)\nu}(\mathcal{P}) = \sum_P \phi(t\mu(P) + (1-t)\nu(P))$.

Como a função ϕ é côncava,

$$H_{t\mu+(1-t)\nu}(\mathcal{P}) \geq \sum_P t\phi(\mu(P)) + (1-t)\phi(\nu(P)) = tH_\mu(\mathcal{P}) + (1-t)H_\nu(\mathcal{P}).$$

Isto significa que a função $\mu \mapsto H_\mu$ é côncava. Aplicar a desigualdade 2.1.9 a esta função resulta na segunda afirmação. ■

Se μ é invariante por f e $\mathcal{P}$ é uma partição, então é claro que $H_\mu(f^{-1}(\mathcal{P})) = H_\mu(\mathcal{P})$. Mais geralmente, $H_\mu(f^{-k}(\mathcal{P})) = H_\mu(\mathcal{P})$, para todo $k \in \mathbb{N}$.

Lema 3.2.7. *Sejam $\mathcal{P}, \mathcal{Q}$ e $\mathcal{R}$ partições com entropia finita. Então:*

1. $H_\mu(\mathcal{P}|\mathcal{Q}) = 0$ *se, e somente se,* $\mathcal{P} \prec \mathcal{Q}$.

2. *Se* $\mathcal{P} \prec \mathcal{Q}$, *então*

   a) $f^{-1}(\mathcal{P}) \prec f^{-1}(\mathcal{Q})$.

   b) $H_\mu(\mathcal{P}|\mathcal{R}) \leq H_\mu(\mathcal{Q}|\mathcal{R})$. *Em particular, tomando* $\mathcal{R} = \{X\}$, *tem-se* $H_\mu(\mathcal{P}) \leq H_\mu(\mathcal{Q})$.

   c) $H_\mu(\mathcal{R}|\mathcal{P}) \geq H_\mu(\mathcal{R}|\mathcal{Q})$.

3. $H_\mu(\mathcal{P} \vee \mathcal{Q}|\mathcal{R}) = H_\mu(\mathcal{P}|\mathcal{R}) + H_\mu(\mathcal{Q}|\mathcal{P} \vee \mathcal{R})$. *Pelo item anterior, tem-se também* $H_\mu(\mathcal{P} \vee \mathcal{Q}|\mathcal{R}) \leq H_\mu(\mathcal{P}|\mathcal{R}) + H_\mu(\mathcal{Q}|\mathcal{R})$. *Em particular, tomando* $\mathcal{R} = \{X\}$, $H_\mu(\mathcal{P} \vee \mathcal{Q}) \leq H_\mu(\mathcal{P}) + H_\mu(\mathcal{Q}|\mathcal{P}) \leq H_\mu(\mathcal{P}) + H_\mu(\mathcal{Q})$.

4. *Se* f *preserva* μ, *então*

   a) $H_\mu(f^{-1}(\mathcal{P})|f^{-1}(\mathcal{Q})) = H_\mu(\mathcal{P}|\mathcal{Q})$. *Mais geralmente,* $H_\mu(f^{-k}(\mathcal{P})|f^{-k}(\mathcal{Q})) = $ $= H_\mu(\mathcal{P}|\mathcal{Q})$, *para todo* $k \in \mathbb{N}$.

   b) $H_\mu(\mathcal{P}_f^n|\mathcal{Q}_f^n) \leq nH_\mu(\mathcal{P}|\mathcal{Q})$.

O item 1 diz que $\mathcal{Q}$ refina $\mathcal{P}$ é equivalente a afirmar que toda informação cedida por $\mathcal{Q}$ contém a informação dada por $\mathcal{P}$ e, assim, não há acréscimo de informação.

O item 3 descreve a cumulatividade da informação: inserir $\mathcal{P} \vee \mathcal{Q}$ havendo previamente $\mathcal{R}$ equivale a inserir $\mathcal{P}$ após sucedido $\mathcal{R}$, depois $\mathcal{Q}$, ocorrido $\mathcal{P}$ juntamente com $\mathcal{R}$

Dem.: 1. $H_\mu(\mathcal{P}|\mathcal{Q}) = 0$ se, e somente se, para todo $P \in \mathcal{P}$ e todo $Q \in \mathcal{Q}$, ocorrer $\mu(P \cap Q) = 0$ ou $\mu(P \cap Q) = \mu(Q)$. Isso equivale a dizer que Q é disjunto de P, a menos de medida nula, ou Q está contido em P, a menos de medida nula. Assim, $H_\mu(\mathcal{P}|\mathcal{Q}) = 0$ é o mesmo que $\mathcal{P} \prec \mathcal{Q}$.

2. Suponha $\mathcal{P} \prec \mathcal{Q}$. Então:

   a) É idêntico ao item 3.1.6-1.

   b) $H_\mu(\mathcal{P}|\mathcal{R}) = \sum_{P,R} -\mu(P \cap R) \log \dfrac{\mu(P \cap R)}{\mu(R)} = \sum_{P} \sum_{R} \sum_{Q \subset P} -\mu(Q \cap R) \log \dfrac{\mu(P \cap R)}{\mu(R)} \leq$
   $\leq \sum_{P} \sum_{R} \sum_{Q \subset P} -\mu(Q \cap R) \log \dfrac{\mu(Q \cap R)}{\mu(R)} = \sum_{R} \sum_{P} \sum_{Q \subset P} -\mu(Q \cap R) \log \dfrac{\mu(Q \cap R)}{\mu(R)} =$
   $= \sum_{Q,R} -\mu(Q \cap R) \log \dfrac{\mu(Q \cap R)}{\mu(R)} = H_\mu(\mathcal{Q}|\mathcal{R})$.

   c) Utilizando a :a função ϕ:

$$H_\mu(\mathcal{R}|\mathcal{P}) = \sum_{R,P} -\mu(R \cap P) \log \frac{\mu(R \cap P)}{\mu(P)} = \sum_{R,P} \mu(P)\phi\left(\frac{\mu(R \cap P)}{\mu(P)}\right) \geq$$
$$= \sum_{R,P} \mu(P)\phi\left(\sum_{Q \subset P} \frac{\mu(Q)}{\mu(P)} \frac{\mu(R \cap Q)}{\mu(Q)}\right) \geq \sum_{R,P} \mu(P) \sum_{Q \subset P} \frac{\mu(Q)}{\mu(P)} \phi\left(\frac{\mu(R \cap Q)}{\mu(Q)}\right) =$$
$$= \sum_{R,Q} \mu(Q)\phi\left(\frac{\mu(R \cap Q)}{\mu(Q)}\right) = \sum_{R,Q} -\mu(R \cap Q) \frac{\mu(R \cap Q)}{\mu(Q)} = H_\mu(\mathcal{R}|\mathcal{Q}).$$

3. $H_\mu(\mathcal{P} \vee \mathcal{Q}|\mathcal{R}) = \sum_{P,Q,R} -\mu(P \cap Q \cap R) \log \dfrac{\mu(P \cap Q \cap R)}{\mu(R)} =$
$= \sum_{P,Q,R} -\mu(P \cap Q \cap R) \log \dfrac{\mu(P \cap Q \cap R)}{\mu(P \cap R)} + \sum_{P,Q,R} -\mu(P \cap Q \cap R) \log \dfrac{\mu(P \cap R)}{\mu(R)} =$
$= H_\mu(\mathcal{Q}|\mathcal{P} \vee \mathcal{R}) + \sum_{P,R} \log \dfrac{\mu(P \cap R)}{\mu(R)} \sum_{Q} -\mu(P \cap Q \cap R) =$
$= H_\mu(\mathcal{Q}|\mathcal{P} \vee \mathcal{R}) + \sum_{P,R} -\mu(P \cap R) \log \dfrac{\mu(P \cap R)}{\mu(R)} = H_\mu(\mathcal{Q}|\mathcal{P} \vee \mathcal{R}) + H_\mu(\mathcal{P}|\mathcal{R})$.

4. Dado que f preserva μ:

   a) $H_\mu(f^{-1}(\mathcal{P})|f^{-1}(\mathcal{Q})) = \sum_{P,Q} -\mu(f^{-1}(P) \cap f^{-1}(Q)) \log \dfrac{\mu(f^{-1}(P) \cap f^{-1}(Q))}{\mu(f^{-1}(Q))} =$
   $= \sum_{P,Q} -\mu(P \cap Q) \log \dfrac{\mu(P \cap Q)}{\mu(Q)} = H_\mu(\mathcal{P}|\mathcal{Q})$.

   b) Usando o caso particular do item 3, tem-se $H_\mu(\mathcal{P}_f^n|\mathcal{Q}_f^n) \leq H_\mu(\mathcal{P}|\mathcal{Q}_f^n) +$
   $+ H_\mu(f^{-1}(\mathcal{P})|\mathcal{Q}_f^n) + \cdots + +H_\mu(f^{-n+1}(\mathcal{P})|\mathcal{Q}_f^n)$.
   Como $f^{-i}(\mathcal{Q}) \prec \mathcal{Q}_f^n$, para $i = 0, \ldots, n-1$, pelo item 2c, $H_\mu(\mathcal{P}_f^n|\mathcal{Q}_f^n) \leq$
   $\leq H_\mu(\mathcal{P}|\mathcal{Q}) + H_\mu(f^{-1}(\mathcal{P})|f^{-1}(\mathcal{Q})) + \cdots + H_\mu(f^{-n+1}(\mathcal{P})|f^{-n+1}(\mathcal{Q})) =$
   $= nH_\mu(\mathcal{P}|\mathcal{Q})$. pelo item 4a. ■

No caso de duas partições $\mathcal{P}$ e $\mathcal{Q}$ serem independentes, temos $I^\mu_{\mathcal{P} \vee \mathcal{Q}} = I^\mu_{\mathcal{P}} + I^\mu_{\mathcal{Q}}$ e, portanto, $H_\mu(\mathcal{P} \vee \mathcal{Q}) = H_\mu(\mathcal{P}) + H_\mu(\mathcal{Q})$. Assim, $H_\mu(\mathcal{P}|\mathcal{Q}) = H_\mu(\mathcal{P})$, ou seja, a informação cedida por $\mathcal{P}$ é independente da informação dada por $\mathcal{Q}$.

Definição 3.2.8. Chamamos de $\mathcal{Z}_\mu$ o espaço das classes de equivalência das partições de X com entropia finita, com a relação $\mathcal{P} \sim \mathcal{Q}$ se, e somente se, os elementos de $\mathcal{P}$ são iguais aos de $\mathcal{Q}$ a menos de medida μ nula.

Será útil o emprego da *métrica de Rokhlin* $\rho_\mu : \mathcal{Z}_\mu \times \mathcal{Z}_\mu \to \mathbb{R}$, definida por

$$\rho_\mu(\mathcal{P}, \mathcal{Q}) = H_\mu(\mathcal{P}|\mathcal{Q}) + H_\mu(\mathcal{Q}|\mathcal{P}) = 2H_\mu(\mathcal{P} \vee \mathcal{Q}) - H_\mu(\mathcal{P}) - H_\mu(\mathcal{Q}).$$

A função ρ_μ é claramente não negativa. Além disso,

- $\rho_\mu(\mathcal{P}, \mathcal{Q}) = 0$ se, e somente se, $H_\mu(\mathcal{P}|\mathcal{Q}) = H_\mu(\mathcal{Q}|\mathcal{P}) = 0$, o que implica $\mathcal{P} \prec \mathcal{Q}$ e $\mathcal{Q} \prec \mathcal{P}$ a menos de medida μ nula, ou $\mathcal{P} = \mathcal{Q}$ em $\mathcal{Z}_\mu$.

- Dadas $\mathcal{P}, \mathcal{Q}, \mathcal{R} \in \mathcal{Z}_\mu$, e utilizando o Lema 3.2.7-3:

$$\begin{aligned}\rho_\mu(\mathcal{P}, \mathcal{Q}) &= H_\mu(\mathcal{P}|\mathcal{Q}) + H_\mu(\mathcal{Q}|\mathcal{P}) \leq H_\mu(\mathcal{R} \vee \mathcal{P}|\mathcal{Q}) + H_\mu(\mathcal{R} \vee \mathcal{Q}|\mathcal{P}) = \\ &= H_\mu(\mathcal{R}|\mathcal{Q}) + H_\mu(\mathcal{P}|\mathcal{R} \vee \mathcal{Q}) + H_\mu(\mathcal{R}|\mathcal{P}) + H_\mu(\mathcal{Q}|\mathcal{R} \vee \mathcal{P}) \leq \\ &\leq H_\mu(\mathcal{R}|\mathcal{Q}) + H_\mu(\mathcal{P}|\mathcal{R}) + H_\mu(\mathcal{R}|\mathcal{P}) + H_\mu(\mathcal{Q}|\mathcal{R}) = \\ &= \rho_\mu(\mathcal{P}, \mathcal{R}) + \rho_\mu(\mathcal{R}, \mathcal{Q}).\end{aligned}$$

Isto confirma que ρ_μ é uma métrica.

Lembremos do operador de *diferença simétrica* entre conjuntos: $A\Delta B = (A \backslash B) \cup (B \backslash A) = (A \cup B) \backslash (A \cap B)$.

Proposição 3.2.9. *(17) Para todo $\varepsilon > 0$, existe $\delta > 0$ de forma que, se $\mathcal{P} = \{P_1, \ldots, P_n\}$ e $\mathcal{Q} = \{Q_1, \ldots, Q_n\}$ são partições tais que $\mu(P_i \Delta Q_i) < \delta$, então $\rho_\mu(\mathcal{P}, \mathcal{Q}) < \varepsilon$. Em particular, $H_\mu(\mathcal{P}|\mathcal{Q}), H_\mu(\mathcal{Q}|\mathcal{P}) < \varepsilon$.*

Esta proposição diz que partições finitas cujos elementos são similares fornecem aproximadamente a mesma informação.

Dem.: Seja $\varepsilon > 0$. Escolha $\delta > 0$ de modo que $\delta < 1/4$ e $-n(n-1)\delta \log \delta - (1-\delta)\log(1-\delta) < \varepsilon/2$.

Seja, também, ξ a partição formada por elementos que são interseções $P_i \cap Q_j$, com $i, j = 1, \ldots, n$ e $i \neq j$, e o conjunto $\bigcup_{i=1}^{n} (P_i \cap Q_i)$.

Note que $\mathcal{P} \vee \mathcal{Q} = \mathcal{Q} \vee \xi$ e que $(P_i \cap Q_j) \subset \bigcup_{k=1}^{n} (P_k \Delta Q_k)$.

Da hipótese resultam $\mu(P_i \cap Q_j) < \delta$, $i \neq j$, e $\mu\left(\bigcup_{k=1}^{n} (P_k \cap Q_k)\right) > 1 - \delta$.

Como δ é pequeno, usando a função ϕ:

- $\phi(\mu(P_i \cap Q_j)) < \phi(\delta) = -\delta \log \delta$, se $i \neq j$;
- $\phi\left(\mu\left(\bigcup_{k=1}^{n} \mu(P_k \cap Q_k)\right)\right) < \phi(1-\delta) = -(1-\delta)\log(1-\delta)$.

Como há, no máximo, $n(n-1)$ conjuntos do tipo $P_i \cap Q_j, i \neq j$, então:

$$H_\mu(\xi) = \sum_{i \neq j} \phi(\mu(P_i \cap Q_j)) + \sum_{i=1}^{n} -\phi\left(\mu\left(\bigcup_{i=1}^{n} (P_i \cap Q_i)\right)\right) \leq$$
$$\leq -n(n-1)\delta \log \delta - (1-\delta)\log(1-\delta) < \varepsilon/2.$$

Assim,

$$H_\mu(\mathcal{P}|\mathcal{Q}) + H_\mu(\mathcal{Q}) = H_\mu(\mathcal{P} \vee \mathcal{Q}) = H_\mu(\mathcal{P} \vee \xi) \leq H_\mu(\mathcal{Q}) + H_\mu(\xi) < H_\mu(\mathcal{Q}) + \varepsilon/2.$$

Isto implica em $H_\mu(\mathcal{P}|\mathcal{Q}) < \varepsilon/2$.

Analogamente, como $\mathcal{P} \vee \mathcal{Q} = \mathcal{P} \vee \xi$, ocorre $H_\mu(\mathcal{Q}|\mathcal{P}) < \varepsilon/2$, donde concluímos a proposição. ■

Proposição 3.2.10. *Se $\mathcal{P}$ é uma partição e f preserva μ então existe o limite*

$$\lim_n \frac{1}{n} H_\mu(\mathcal{P} \vee f^{-1}(\mathcal{P}) \vee \cdots \vee f^{-n+1}(\mathcal{P})) = \lim_n \frac{1}{n} H_\mu(\mathcal{P}_f^n).$$

Dem.: Usando os Lemas 3.2.7 e 2.1.10, é análoga à demonstração da Proposição 3.1.4, substituindo α cobertura por $\mathcal{P}$ partição. ■

Definição 3.2.11. A *entropia* de uma aplicação f *com respeito a uma partição* $\mathcal{P}$ e uma medida invariante μ, $h_\mu(f, \mathcal{P})$, é definida pelo limite

$$h_\mu(f, \mathcal{P}) = \lim_n \frac{1}{n} H_\mu(\mathcal{P}_f^n).$$

A entropia, como definida acima, pode ser vista como a informação média, ao longo do tempo, que uma função - interpretada como um experimento - e uma partição fornecem. Conhecendo um elemento de $\mathcal{P}_f^n$, sabemos onde pontos do espaço de estados X estão e onde seus respectivos iterados por f estão.

Definição 3.2.12. A *entropia métrica de uma aplicação f com respeito a uma medida invariante* μ, $h_\mu(f)$, é definida como o supremo do conjunto $\{h_\mu(f, \mathcal{P}); \mathcal{P}$ é partição com entropia finita$\}$:

$$h_\mu(f) = \sup_{\mathcal{P}} h_\mu(f, \mathcal{P}).$$

Proposição 3.2.13. *Para toda partição $\mathcal{P}$ e $\mu \in \mathcal{M}(X, f)$, tem-se*

$$h_\mu(f, \mathcal{P}) = \lim_n H_\mu\left(\mathcal{P} \,\middle|\, \bigvee_{i=1}^{n} f^{-i}(\mathcal{P})\right)$$

Dem.: Por definição de entropia condicional,

$$H_\mu(\mathcal{P}_f^n) = H_\mu\left(\bigvee_{i=1}^{n-1} f^{-i}(\mathcal{P})\right) + H_\mu\left(\mathcal{P}\middle|\bigvee_{i=1}^{n-1} f^{-i}(\mathcal{P})\right) =$$
$$= H_\mu\left(\bigvee_{i=1}^{n-2} f^{-i}(\mathcal{P})\right) + H_\mu\left(\mathcal{P}\middle|\bigvee_{i=1}^{n-2} f^{-i}(\mathcal{P})\right) + H_\mu\left(\mathcal{P}\middle|\bigvee_{i=1}^{n-1} f^{-i}(\mathcal{P})\right).$$

Indutivamente, conclui-se que $H_\mu(\mathcal{P}_f^n) = H_\mu(\mathcal{P}) + \sum_{k=1}^{n-1} H_\mu\left(\mathcal{P}\middle|\bigvee_{i=1}^{k} f^{-i}(\mathcal{P})\right)$, usando que μ é invariante.

Então, $h_\mu(f,\mathcal{P})$ é dada por um limite Cesàro: $h_\mu(f,\mathcal{P}) = \lim_n \dfrac{1}{n}\sum_{k=1}^{n-1} H_\mu\left(\mathcal{P}\middle|\bigvee_{i=1}^{k} f^{-i}(\mathcal{P})\right)$.

Pelo Lema 3.2.7-2c, a sequência $\left\{H_\mu\left(\mathcal{P}\middle|\bigvee_{i=1}^{k} f^{-i}(\mathcal{P})\right)\right\}_{k\in\mathbb{N}}$ é decrescente.

Logo, o limite desta sequência existe e coincide com seu limite Cesàro. ■

Proposição 3.2.14. *Dadas $\mathcal{P}$ e $\mathcal{Q}$ duas partições com entropia finita, e $\mu \in \mathcal{M}(X,f)$:*

1. $h_\mu(f,\mathcal{P}) \le H_\mu(\mathcal{P})$.
2. *Se* $\mathcal{P} \prec \mathcal{Q}$, *então* $h_\mu(f,\mathcal{P}) \le h_\mu(f,\mathcal{Q})$.
3. $h_\mu(f,\mathcal{P}) \le h_\mu(f,\mathcal{Q}) + H_\mu(\mathcal{P}|\mathcal{Q})$.
4. $h_\mu(f,\mathcal{P}) = h_\mu(f,\mathcal{P}_f^k)$ *para todo* $k \in \mathbb{N}$.

Dem.: 1. $h(f,\mathcal{P}) = \frac{1}{n}H_\mu(\mathcal{P}_f^n) \le \lim_n (H(\mathcal{P}) + H(f^{-1}(\mathcal{P})) + \cdots + H(f^{n-1}(\mathcal{P})) \le$

$$\le \frac{1}{n}(\underbrace{H_\mu(\mathcal{P}) + \cdots + H_\mu(\mathcal{P})}_{n\text{vezes}}) = \lim_n H_\mu(\mathcal{P}) = H_\mu(\mathcal{P}).$$

2. Se $\mathcal{P} \prec \mathcal{Q}$, então $\mathcal{P}_f^n \prec \mathcal{Q}_f^n$, para todo n. Daí, $H_\mu(\mathcal{P}_f^n) \le H_\mu(\mathcal{Q}_f^n)$ e, então, $h_\mu(f,\mathcal{P}) \le h_\mu(f,\mathcal{Q})$.

3. Fazendo uso do item 3.2.7-4b,

$$\frac{1}{n}H_\mu(\mathcal{P}_f^n) \le \frac{1}{n}H_\mu((\mathcal{P}\vee\mathcal{Q})_f^n) = \frac{1}{n}H_\mu(\mathcal{P}_f^n \vee \mathcal{Q}_f^n) = \frac{1}{n}(H_\mu(\mathcal{P}_f^n|\mathcal{Q}_f^n) + H_\mu(\mathcal{Q}_f^n)) \le$$
$$\le \frac{1}{n}(nH_\mu(\mathcal{P}|\mathcal{Q}) + H_\mu(\mathcal{Q}_f^n) = H_\mu(\mathcal{P}|\mathcal{Q}) + \frac{1}{n}H_\mu(\mathcal{Q}_f^n).$$

Assim, $h_\mu(f,\mathcal{P}) \le h_\mu(f,\mathcal{Q}) + H_\mu(\mathcal{P}|\mathcal{Q})$.

4. Observe que

$$(\mathcal{P}_f^k)_f^n = (\mathcal{P}_f^k \vee \cdots \vee f^{-n+1}(\mathcal{P}_f^k)) = (\mathcal{P} \vee \cdots \vee f^{-k+1}(\mathcal{P}) \bigvee f^{-1}(\mathcal{P}) \vee \cdots \vee f^{-k}(\mathcal{P}) \bigvee \cdots$$
$$\cdots \bigvee f^{-n+1}(\mathcal{P}) \vee \cdots \vee f^{-n-k+2}(\mathcal{P})) = \mathcal{P}_f^{n+k-1}.$$

Portanto, $h(f,\mathcal{P}_f^k) = \lim_n \frac{1}{n}H_\mu(\mathcal{P}_f^{n+k-1}) = \lim_n \frac{1}{n}H_\mu(\mathcal{P}_f^n) = h(f,\mathcal{P})$. ■

Proposição 3.2.15. *A Definição 3.2.12 não é alterada se tomarmos o supremo sobre partições finitas.*

Dem.: Seja $\mathcal{P} = \{P_k\}_{k\in\mathbb{N}}$ uma partição com entropia finita e considere a sequência de partições finitas $\mathcal{P}_k = \{P_1, \ldots, P_k, Q_k\}$, onde $Q_k = \bigcup\limits_{i>k} P_i$.

Então, $h_\mu(f, \mathcal{P}_k) \leq h_\mu(f, \mathcal{P}) \leq h_\mu(f, \mathcal{P}_k) + H_\mu(\mathcal{P}|\mathcal{P}_k)$, para todo $k \in \mathbb{N}$.

Ainda, usando que $\mathcal{P}_k \prec \mathcal{P}$ e a função ϕ:

$$H_\mu(\mathcal{P}|\mathcal{P}_k) = H_\mu(\mathcal{P}) - H_\mu(\mathcal{P}_k) = \sum_{i=1}^{\infty}\phi(P_i) - \sum_{i=1}^{k}\phi(P_i) + \mu(Q_k)\log\mu(Q_k) =$$
$$= \sum_{i=k+1}^{\infty}\phi(P_i) - \sum_{i=k+1}^{\infty} -\mu(P_i)\log\mu(Q_k) = \sum_{i=k+1}^{\infty}\mu(P_i)(\log\mu(P_i) - \log\mu(Q_k))$$

Dado que a série $\sum\limits_{i=1}^{\infty}\phi(P_i)$ converge, $H_\mu(\mathcal{P}|\mathcal{P}_k) \to 0$ quando $k \to \infty$.

Logo, $h_\mu(f, \mathcal{P}) = \sup\limits_k h_\mu(f, \mathcal{P}_k)$. ■

Observação 3.2.16. Segue da Proposição 3.2.15 que o conjunto das partições finitas é denso em $(\mathcal{Z}_\mu, \rho_\mu)$.

De fato, nos moldes da Proposição 3.2.15, como $H_\mu(\mathcal{P}_k|\mathcal{P}) = 0$, para todo k, e $H_\mu(\mathcal{P}|\mathcal{P}_k) \to 0$ quando $k \to \infty$, segue que, dado $\varepsilon > 0$, existe k_0 tal que $\rho_\mu(\mathcal{P}, \mathcal{P}_{k_0}) = H_\mu(\mathcal{P}|\mathcal{P}_{k_0}) < \varepsilon$.

Lema 3.2.17. *(11) Sejam A um subconjunto de inteiros não negativos, $\mathcal{P}$ uma partição e $\mu \in \mathcal{M}(X, f)$. Então,*

$$H_\mu\left(\bigvee_{i\in A} f^{-i}(\mathcal{P})\right) \geq (\#A)h_\mu(f, \mathcal{P}).$$

Dem.: Vamos usar indução sobre $\#A$.

Para $\#A = 1$, o resultado é verdadeiro, pela Proposição 3.2.14-1.

Assuma validade da afirmação para $\#A = k > 1$.

Seja B um subconjunto dos inteiros não negativos com $\#B = k+1$, $b_0 = \max\{b; b \in B\}$, $b_1 = \min\{b; b \in B\}$ e $B_1 = B\backslash\{b_0\}$. Então, por definição de entropia condicional,

$$H_\mu\left(\bigvee_{i\in B} f^{-i}(\mathcal{P})\right) = H_\mu\left(f^{-b_1}(\mathcal{P}) \middle| \bigvee_{i\in B_1} f^{-i}(\mathcal{P})\right) + H_\mu\left(\bigvee_{i\in B_1} f^{-i}(\mathcal{P})\right)$$

Pela hipótese de indução e usando os itens 3.2.7-4a e 3.2.7-2c:

$$H_\mu\left(\bigvee_{i\in B} f^{-i}(\mathcal{P})\right) \geq H_\mu\left(\mathcal{P} \middle| \bigvee_{i\in B_1} f^{-(i-b_1)}(\mathcal{P})\right) + (\#B_1)h_\mu(f, \mathcal{P}) \geq$$
$$\geq H_\mu\left(\mathcal{P} \middle| \bigvee_{i=1}^{b_0} f^{-i}(\mathcal{P})\right) + (\#B_1)h_\mu(f, \mathcal{P}).$$

Usando a Proposição 3.2.13, obtemos

$$H_\mu\left(\bigvee_{i\in B} f^{-i}(\mathcal{P})\right) \geq h_\mu(f,\mathcal{P}) + (\#B_1)h_\mu(f,\mathcal{P}) = (\#B)h_\mu(f,\mathcal{P}). \qquad \blacksquare$$

Proposição 3.2.18. *$h_\mu(f^k) = kh_\mu(f)$, para todo $k \in \mathbb{N}$.*

Dem.: Análoga à demonstração da Proposição 3.1.6-3, substituindo α cobertura por $\mathcal{P}$ partição. $\blacksquare$

3.3 O PRINCÍPIO VARIACIONAL

Neste contexto, lembramos que $\mathcal{M}(X,f) \subset \mathcal{M}(X)$ é o espaço das medidas invariantes por f, munido da topologia fraca*. Este é um subconjunto não vazio do espaço de probabilidades em X, $\mathcal{M}(X)$ (19).

O conteúdo desta seção foi adaptado de (19), (15), (23) e (26).

O *diâmetro* de uma partição $\mathcal{P}$ é definido por diam $\mathcal{P} := \sup_{P\in\mathcal{P}}\{\text{diam } P\}$.

Proposição 3.3.1. *Se $\mu \in \mathcal{M}(X)$, então não existe família não enumerável de subconjuntos de X disjuntos dois a dois, cada um com medida positiva.*

Dem.: Seja $\mathcal{G}$ uma família qualquer de subconjuntos de X com medida positiva, e $\mathcal{G}_k = \{G \in \mathcal{G}; \mu(G) \geq 1/k\}$.

Todo $G \in \mathcal{G}$ está em algum $\mathcal{G}_k$, logo $\bigcup_{i=1}^{\infty} \mathcal{G}_k = \mathcal{G}$.

Se $\mathcal{G}$ é não enumerável, então, ao menos um dos $\mathcal{G}_k$ é, ao menos, infinito. Então $\mu\left(\bigcup_{G\in\mathcal{G}_k} G\right) \geq \sum_{i=1}^{\infty} 1/k$, que é não finito, contradizendo a finitude de μ. $\blacksquare$

Proposição 3.3.2. *Se X é um espaço métrico compacto e $\mu \in \mathcal{M}(X)$, então, dado $\varepsilon > 0$, existe uma partição $\mathcal{P} = \{P_1, \ldots, P_k\}$ de sorte que diam $P_i < \varepsilon$ e $\mu(\partial P_i) = 0, 1 \leq i \leq k$.*

Dem.: Dado $x \in X$ existe $\varepsilon_1 < \varepsilon$ de forma que $\mu(\partial B(x,\varepsilon_1)) = 0$. Isto porque, caso não exista, teríamos uma coleção não enumerável de conjuntos disjuntos dois a dois, com medida positiva, o que não pode acontecer, pela Proposição 3.3.1.

Logo, por compacidade, existe cobertura $\alpha = \{A_1, \ldots, A_r\}$ de bolas abertas com raio $\varepsilon_1 < \varepsilon$.

Sejam $B_1 = \overline{A_1}$ e $B_k = \overline{A_k} \setminus \bigcup_{i=1}^{k-1} \overline{A_i}$ para $1 < k \leq r$.

Logo, $\mathcal{P} = \{A_1, \ldots, A_r\}$ é uma partição, diam $A_n < \varepsilon$, e $\mu(\partial B_n) \leq \mu\left(\bigcup_{i=1}^{n} \partial A_i\right) = 0$. $\blacksquare$

Teorema 3.3.3 (Princípio Variacional)**.** *Se $f : X \to X$ é uma função contínua e X é uma espaço métrico compacto, então a entropia topológica $h(f)$ coincide com o supremo das entropias métricas $h_\mu(f)$ relativas às medidas de probabilidade invariantes por f.*

Vamos dividir a demonstração deste fato em dois lemas, cada um contendo uma das duas desigualdades.

Lema 3.3.4. *Nas hipóteses do Teorema 3.3.3,* $\sup\limits_{\mu \in \mathcal{M}(X,f)} h_\mu(f) \leq h(f)$.

Dem.: Seja $\mathcal{P} = \{P_1, \ldots, P_r\}$ uma partição finita de X.

Pela Proposição 2.1.3, μ é regular.

Como X é compacto, dado $\varepsilon > 0$, pode-se obter $K_1, \ldots, K_r$ compactos tais que $K_i \subset P_i$ e $\mu(P_i \backslash K_i) < \varepsilon$. Seja $\mathcal{Q} = \{K_0, K_1, \ldots, K_r\}$ uma partição de X, em que $K_0 = X \backslash \bigcup\limits_{i=1}^{r} K_i = \bigcap\limits_{i=1}^{r} K_i^c$.

Então $\mu(K_0) < r\varepsilon$. De fato:

$$\mu(K_0) = \mu\left(X \Big\backslash \bigcup_{i=1}^{r} K_i\right) = \mu\left(\bigcup_{i=1}^{r} P_i \Big\backslash \bigcup_{i=1}^{r} K_i\right) = \mu\left(\bigcup_{i=1}^{r} (P_i \backslash K_i)\right) < \sum_{i=1}^{r} \mu(P_i \backslash K_i) = r\varepsilon.$$

Logo, usando a função $\phi(x) = -x \log x$:

$$H_\mu(\mathcal{P}|\mathcal{Q}) = \sum_{i=1}^{r} \sum_{j=0}^{r} -\mu(P_i \cap K_j) \log \frac{\mu(P_i \cap K_j)}{\mu(K_j)} = \sum_{i=1}^{r} \sum_{j=0}^{r} \mu(K_j) \phi\left(\frac{\mu(P_i \cap K_j)}{\mu(K_j)}\right).$$

Para $j \neq 0$, ou $\mu(P_i \cap K_j) = \mu(K_j)$ ou $\mu(P_i \cap K_j) = 0$. Daí, como $\phi(1) = 0$,

$$H_\mu(\mathcal{P}|\mathcal{Q}) = \sum_{i=1}^{r} \mu(K_0) \phi\left(\frac{\mu(P_i \cap K_0)}{\mu(K_0)}\right) + \sum_{i=1}^{r} \sum_{j=1}^{r} \mu(K_j) \phi(1) = \sum_{i=1}^{r} \mu(K_0) \phi\left(\frac{\mu(P_i \cap K_0)}{\mu(K_0)}\right)$$

Como ϕ é côncava, lança-se mão da desigualdade de Jensen (2.1.9):

$$H_\mu(\mathcal{P}|\mathcal{Q}) = r\mu(K_0) \sum_{i=1}^{r} \frac{1}{r} \phi\left(\frac{\mu(P_i \cap K_0)}{\mu(K_0)}\right) \leq r\mu(K_0) \phi\left(\frac{1}{r} \sum_{i=1}^{r} \frac{\mu(P_i \cap K_0)}{\mu(K_0)}\right) =$$
$$= r\mu(K_0) \phi\left(\frac{1}{r}\right) < \varepsilon r \log r.$$

Escolhendo $\varepsilon < 1/(r \log r)$, tem-se $H_\mu(\mathcal{P}|\mathcal{Q}) < 1$. Para cada i, a união $K_0 \cup K_i$ é igual a X, removendo-se todos os K_j, com $j \neq i$: $K_0 \cup K_i = X \backslash \bigcup_{j \neq i} K_j$. Por isso, cada $K_0 \cup K_i$ é aberto.

Logo, $\alpha = \{K_0 \cup K_1, K_0 \cup K_2, \ldots, K_0 \cup K_r\}$ é uma cobertura aberta.

Todo elemento da cobertura α_f^n é da forma $(K_0 \cup K_{t_1}) \cap f^{-1}(K_0 \cup K_{t_2}) \cap \cdots \cap f^{-n+1}(K_0 \cup K_{t_n})$, com $t_i \in \{1, \ldots, r\}$, $i = 1, \ldots, n$.

Portanto, cada um destes contém algum elemento de $\mathcal{Q}_f^n$ da forma $K_{s_1} \cap f^{-1}(K_{s_2}) \cap \cdots \cap f^{-n+1}(K_{s_n})$, com $s_i \in \{1, \ldots, r\}$, $i = 1, \ldots, n$. Além disso, ocorre $t_i = s_i$ ou $s_i = 0$, para todo $i = 1, \ldots, n$, pois cada elemento de α contém no máximo dois elementos de $\mathcal{Q}$.

Assim, existem no máximo 2^n elementos de $\mathcal{Q}_f^n$ em cada elemento de α_f^n.

Tome uma subcobertura minimal de α_f^n, com N elementos.

Então, existem no máximo $N2^n$ elementos na partição $\mathcal{Q}_f^n$.

Pelo Lema 3.2.5, $H_\mu(\mathcal{Q}_f^n) \leq \log N2^n = n \log 2 + \log N$.

Dividindo por n e passando ao limite com $n \to \infty$, obtemos

$$h_\mu(f, \mathcal{Q}) \leq h(f, \alpha) + \log 2 \leq h(f) + \log 2.$$

Pela Proposição 3.2.14-3:

$$h_\mu(f, \mathcal{P}) \leq h_\mu(f, \mathcal{Q}) + H_\mu(\mathcal{P}|\mathcal{Q}) < h(f) + \log 2 + 1 \Rightarrow h_\mu(f) < h(f) + \log 2 + 1.$$

Esta desigualdade também vale para f^n, dado que esta também preserva μ. Pelas proposições 3.2.18 e 3.1.10, $h_\mu(f^n) = nh_\mu(f)$ e $h(f^n) = nh(f)$, logo

$$h_\mu(f) < h(f) + \frac{\log 2}{n} + \frac{1}{n}.$$

Passando ao limite com $n \to \infty$, infere-se o resultado: $\sup\limits_{\mu \in \mathcal{M}(X,f)} h_\mu(f) \leq h(f)$. ■

Lema 3.3.5. *Nas hipóteses do Teorema 3.3.3,* $\sup\limits_{\mu \in \mathcal{M}(X,f)} h_\mu(f) \geq h(f)$.

Dem.: É suficiente provar que, para todo $\varepsilon > 0$, existe $\mu \in \mathcal{M}(X, f)$ tal que $h_\mu(f) \geq s(f, \varepsilon)$, pois, como a entropia topológica é o limite crescente $h(f) = \lim\limits_{\varepsilon \to 0} s(f, \varepsilon)$, isto implicaria em $\sup\limits_{\mu \in \mathcal{M}(X,f)} h_\mu(f) \geq h(f)$.

Seja S_n um conjunto (n, ε)-separado de cardinalidade máxima, $s_n(f, \varepsilon)$.

Defina as medidas de probabilidade $\nu_n := \dfrac{1}{s_n(f, \varepsilon)} \sum\limits_{x \in S_n} \delta_x$ e $\mu_n := \dfrac{1}{n} \sum\limits_{i=0}^{n-1} f_*^i \nu_n$.

Fazendo uso da Proposição 2.2.10, considere uma sequência $\{n_k\}_{k \in \mathbb{N}}$ de modo que $\lim\limits_{n_k} \mu_{n_k} = \mu$, na topologia fraca*. Tem-se que μ invariante por f.

Tome uma partição $\mathcal{P} = \{P_1, \ldots, P_k\}$ com diâmetro menor que ε, tal que todos os seus elementos tenham bordo com medida nula (Proposição 3.3.2).

Observe que, se x, y estão no mesmo elemento de $\mathcal{P}_f^n$, então $d(f^i(x), f^i(y)) \leq \varepsilon, \forall$ $i = 0, \ldots, n-1$. Isso significa que cada elemento de $\mathcal{P}_f^n$ não contém mais que um elemento de S_n.

Caso $P \in \mathcal{P}_f^n$ não contenha elemento de S_n, então $\nu_n(P) = 0$.

Por isso, $H_{\nu_n}(\mathcal{P}_f^n) = \sum_{P\in\mathcal{P}_f^n} -\nu_n(P)\log(\nu_n(P)) = \sum_{x\in S_n} \frac{\log(s_n(f,\varepsilon))}{s_n(f,\varepsilon)} = \log(s_n(f,\varepsilon))$.

Vamos mostrar a desigualdade fundamental: para $q < n-1$, vale

$$\frac{1}{n}H_{\nu_n}(\mathcal{P}_f^n) \leq \frac{1}{q}H_{\mu_n}(\mathcal{P}_f^q) + \frac{2q\log k}{n}. \tag{3.1}$$

A ideia é a seguinte: a junção $\mathcal{P}_f^n$ possui n termos, com pré imagens indexadas desde 0 até $n-1$. Sendo $n-1 = aq+r$, com $0 \leq r \leq q < n-1$, subdividimos o conjunto $\{0,\ldots,n-1\}$ em a em conjuntos com q elementos cada, juntamente com um conjunto com r elementos.

Para cada $0 \leq j \leq q-1$, computamos os termos de $\mathcal{P}_f^n$ tomando blocos de $\mathcal{P}_f^q$, transladados $zq+j$, z indo de 1 até $b(j)-1$, $b(j) = \left\lfloor \frac{n-1-j}{q} \right\rfloor$, acrescentados dos elementos que restaram: de 0 até j, e de $b(j)-1$ até n. Por definição, $(b(j)-1)q+j$ está entre $n-1$ e $n-1-q$.

Chamemos de L_j o conjunto que contém os elementos restantes. Tem-se $\#L_j \leq 2q$, para todo $0 \leq j \leq q-1$. Isso é formalizado na igualdade

$$\mathcal{P}_f^n = \left(\bigvee_{z=0}^{b(j)-1} f^{-(zq+j)}(\mathcal{P}_f^q) \right) \vee \bigvee_{l\in L_j} f^{-l}(\mathcal{P}).$$

Então, para cada $0 \leq j \leq q-1$, acionando a convexidade de H_{ν_n}:

$$\begin{aligned} H_{\nu_n}(\mathcal{P}_f^n) &\leq \sum_{z=0}^{b(j)-1} H_{\nu_n}((f^{-(zq+j)}(\mathcal{P}_f^q)) + \sum_{l\in L_j} H_{\nu_n}(f^{-l}(\mathcal{P})) \leq \\ &\leq \sum_{z=0}^{b(j)-1} H_{f_*^{(zq+j)}\nu_n}((\mathcal{P}_f^q)) + \sum_{l\in L_j} H_{\nu_n}(f^{-l}(\mathcal{P})). \end{aligned} \tag{3.2}$$

O segundo termo do lado direito de 3.2 é menor ou igual a $2q\log k$, por 3.2.5.

Somando todas as q possibilidades de valores para j e dividindo por n, obtemos:

$$\frac{q}{n}\log s_n(f,\varepsilon) = \frac{q}{n}H_{\nu_n}(\mathcal{P}_f^n) = \frac{1}{n}\sum_{z=0}^{n-1} H_{f_*^{(z)}\nu_n}(\mathcal{P}_f^q) + \frac{2q^2\log k}{n} \tag{3.3}$$

Pelo fato de μ_n ser combinação convexa de $\nu_n, f_*(\nu_n), \ldots, f_*^{n-1}(\nu_n)$, o primeiro termo do lado direito de 3.3 é menor ou igual a $H_{\mu_n}(\mathcal{P}_f^q)$, pela Proposição 3.2.6:

$$\frac{q}{n}\log s_n(f,\varepsilon) = \frac{q}{n}H_{\nu_n}(\mathcal{P}_f^n) \leq H_{\mu_n}(\mathcal{P}_f^q) + \frac{2q^2\log k}{n}.$$

Assim, provamos a desigualdade 3.1.

Os membros de $\mathcal{P}_f^q$ têm bordo com medida nula. Portanto, $\mu_{n_k} \to \mu$ implica em $\mu_{n_k}(P) \to \mu(P)$, para todo $P \in \mathcal{P}_f^q$. Assim, substituindo n por n_k e passando ao limite superior com $k \to \infty$, obtemos:

$$qs(f,\varepsilon) \leq H_\mu(\mathcal{P}_f^q) \Rightarrow s(f,\varepsilon) \leq \frac{1}{q} H_\mu(\mathcal{P}_f^q) \Rightarrow s(f,\varepsilon) \leq h_\mu(f).$$

Dessa maneira, dado $\varepsilon > 0$, existe $\mu \in \mathcal{M}(X,f)$ de sorte que $s(f,\varepsilon) \leq h_\mu(f)$.

Logo, $\sup\limits_{\mu \in \mathcal{M}(X,f)} h_\mu(f) \geq h(f)$. ■

A seguir, uma consequência do Princípio Variacional, a qual diz que a entropia topológica pode ser computada tomando a restrição de f a seus pontos não errantes. Começamos com a

Proposição 3.3.6. *Se* $\mu \in \mathcal{M}(X,f)$, *então* supp $\mu \subset \Omega_f$.

Dem.: Por ser X um espaço métrico compacto, X é separável.

Pelo Teorema 2.1.8, μ quase todo ponto $x \in X$ é não errante. Seja M um conjunto com medida 1, no qual todo ponto em M é não errante. Note que M é fechado.

Se x está em supp μ, mas não está em M, então existe uma vizinhança V de x disjunta de M, de sorte que $\mu(V) > 0$, o que contradiz $\mu(M) = 1$. Logo, supp $\mu \subset M$.

Daí, segue que $h_\mu(f) = h_\mu\left(f|_{\Omega_f}\right)$. ■

Teorema 3.3.7. *(27) Se* $\mu \in \mathcal{M}(X,f)$, *as entropias métrica e topológica de* f *não mudam se o domínio for tomado como* Ω_f.

Dem.: Como X é um espaço métrico compacto, X é separável.

Devido à Proposição 3.3.6, $h_\mu(f) = h_\mu\left(f|_{\Omega_f}\right)$

Pelo Princípio Variacional (3.3.3),

$$h(f) = \sup_{\mu \in \mathcal{M}(X,f)} h_\mu(f) = \sup_{\mu \in \mathcal{M}(X,f)} h_\mu\left(f|_{\Omega_f}\right) = h\left(f|_{\Omega_f}\right) \qquad ■$$

4 ENTROPIA SEQUENCIAL

Neste Capitulo, trataremos sobre as extensões das definições de entropia topológica e métrica, idealizadas, respectivamente, por T. N. T. Goodman (10) e A. G. Kushnirenko (9).

Esta parte visa apresentar definições alternativas de entropia, que podem ofertar mais informações a respeito de um sistema dinâmico do que a definição usual (13).

Os operadores H e H_μ são os mesmos utilizados nas seções anteriores.

4.1 ENTROPIA TOPOLÓGICA SEQUENCIAL

Assim como na seção que trata de entropia topológica clássica, subentende-se $f : X \to X$ uma aplicação contínua e X um espaço métrico compacto.

Definição 4.1.1. Dada $\mathcal{S} = \{t_n\}_n$ uma sequência de inteiros não negativos e α uma cobertura de X, definimos a cobertura *n-ésima* $f, \mathcal{S}$*-junção de* α pondo

$$\alpha^n_{f,\mathcal{S}} = f^{-t_1}(\alpha) \vee f^{-t_2}(\alpha) \vee \cdots \vee f^{-t_n}(\alpha) = \bigvee_{i=1}^{n} f^{-t_i}(\alpha).$$

Definição 4.1.2. A *entropia topológica sequencial de uma aplicação f com respeito a uma cobertura α e uma sequência de inteiros não negativos* $\mathcal{S} = \{t_n\}_n$, $h_{\mathcal{S}}(f, \alpha)$, é definida pelo limite superior

$$h_{\mathcal{S}}(f, \alpha) = \limsup_n \frac{1}{n} H(\alpha^n_{f,\mathcal{S}}).$$

A definição usa limite superior, pois não há garantia da existência do limite.

A entropia topológica usual ocorre quando $\mathcal{S} = \{n - 1\}_{n \in \mathbb{N}}$.

A seguir, o conceito de entropia sequencial afim ao de Adler, Konheim e McAndrew:

Definição 4.1.3. A *entropia topológica sequencial de uma aplicação f com respeito a uma sequência de inteiros não negativos* $\mathcal{S} = \{t_n\}_n$, $h_{\mathcal{S}}(f)$, é definida como o supremo do conjunto $\{h_{\mathcal{S}}(f, \alpha); \alpha$ é cobertura com entropia finita$\}$:

$$h_{\mathcal{S}}(f) = \sup_\alpha h_{\mathcal{S}}(f, \alpha).$$

A entropia descrita acima também pode ser chamada de $\mathcal{S}$-entropia.

Teorema 4.1.4. *A entropia topológica sequencial de uma aplicação é um invariante topológico, isto é, duas aplicações topologicamente conjugadas têm a mesma entropia.*

Dem.: Sejam $f, g : X \to X$ duas funções topologicamente conjugadas, α uma cobertura, ψ uma conjugação de f e g e $\mathcal{S} = \{t_n\}_n$ uma sequência de inteiros não negativos.

Note que $g^{-k}(\psi^{-1}(\alpha)) = (\psi^{-1}f\psi)^{-k}(\psi^{-1}(\alpha)) = \psi^{-1}f^{-k}\psi\psi^{-1}(\alpha)) = \psi^{-1}f^{-k}(\alpha)$, para todo $k \in \mathbb{N}$.

Então:

$$\begin{aligned}
&h_{\mathcal{S}}(g, \psi^{-1}(\alpha)) = h_{\mathcal{S}}(\psi^{-1}f\psi, \psi^{-1}(\alpha)) = \\
&= \lim_n \tfrac{1}{n} H(g^{-t_1}(\psi^{-1}(\alpha)) \vee g^{-t_2}(\psi^{-1}(\alpha)) \vee \cdots \vee (g^{-t_n}(\psi^{-1}(\alpha)) = \\
&= \lim_n \tfrac{1}{n} H(\psi^{-1}f^{-t_1}(\alpha) \vee \psi^{-1}f^{-t_2}(\alpha) \vee \cdots \vee \psi^{-1}f^{-t_n}(\alpha)) = \\
&= \lim_n \tfrac{1}{n} H(\psi^{-1}(f^{-t_1}(\alpha) \vee f^{-t_2}(\alpha) \vee \cdots \vee f^{-t_n}(\alpha)) = \\
&= \lim_n \tfrac{1}{n} H(f^{-t_1}\alpha \vee \cdots \vee f^{-t_n}(\alpha)) = h_{\mathcal{S}}(f, \alpha)
\end{aligned}$$

Quando α varia sobre as coberturas de X, $\psi^{-1}(\alpha)$ também varia sobre as coberturas de X, pois ψ é homeomorfismo.

Então, $h_{\mathcal{S}}(g) = \sup_\alpha \{h_{\mathcal{S}}(g, \psi^{-1}(\alpha)\} = \sup_\alpha \{h_{\mathcal{S}}(f, \alpha)\} = h_{\mathcal{S}}(f)$. ■

Definição 4.1.5. Sejam $\mathcal{S} = \{t_n\}_n$ uma sequência de inteiros não negativos, $x \in X$, $n \in \mathbb{N}$ e $\varepsilon > 0$. Definimos a $\mathcal{S}$-*bola dinâmica* de centro x, raio ε e comprimento n, $B(\mathcal{S}, x, n, \varepsilon)$, por

$$\begin{aligned}
B(\mathcal{S}, x, n, \varepsilon) &:= \{y \in X; d(f^{t_m}(x), f^{t_m}(y)) < \varepsilon \ \forall\ 1 \leq m \leq n\} = \\
&= \{y \in X; f^{t_m}(y) \in B(f^{t_m}(x), \varepsilon) \ \forall\ 1 \leq m \leq n\} = \\
&= \bigcap_{i=1}^{n} f^{-t_i}(B(f^{t_i}(x), \varepsilon)).
\end{aligned}$$

Definição 4.1.6. Um conjunto $S \in X$ é dito $(\mathcal{S}, n, \varepsilon)$-*separado* se, para todos $x, y \in S$ distintos, existe $1 \leq k \leq n$ inteiro, de modo que $d(f^{t_k}(x), f^{t_k}(y)) > \varepsilon$.

Intuitivamente, um conjunto $(\mathcal{S}, n, \varepsilon)$-separado S é aquele tal que um aparelho de observação com resolução ε consegue distinguir quaisquer dois segmentos de $(\mathcal{S}, f)$-órbitas de dois pontos nele distintos.

Definição 4.1.7. Um conjunto $R \subset X$ é dito $(\mathcal{S}, n, \varepsilon)$-*gerador* se, para todo $x \in X$, existe $y \in R$ tal que, para todo $1 \leq k \leq n$ inteiro, se tenha $d(f^{t_k}(x), f^{t_k}(y)) \leq \varepsilon$. Isto equivale a dizer que o conjunto $\bigcup_{y \in R} B(\mathcal{S}, y, n, \varepsilon)$ cobre X.

Intuitivamente, um conjunto $(\mathcal{S}, n, \varepsilon)$-gerador R é aquele tal que um aparelho de observação com resolução ε não consegue distinguir o segmento da $(\mathcal{S}, f)$-órbita com comprimento n de qualquer ponto do espaço do segmento da $(\mathcal{S}, f)$-órbita com comprimento n de algum ponto do conjunto.

Chamamos de $s_n(\mathcal{S}, f, \varepsilon)$ o máximo das cardinalidades de todos os conjuntos $(\mathcal{S}, n, \varepsilon)$-separados, e de $r_n(\mathcal{S}, f, \varepsilon)$ o mínimo das cardinalidades de todos os conjuntos $(\mathcal{S}, n, \varepsilon)$-geradores.

Defina $r(\mathcal{S}, f, \varepsilon) \coloneqq \limsup_n \dfrac{1}{n} \log(r_n(\mathcal{S}, f, \varepsilon))$ e $s(\mathcal{S}, f, \varepsilon) \coloneqq \limsup_n \dfrac{1}{n} \log(s_n(\mathcal{S}, f, \varepsilon))$.

Defina, também, $r(\mathcal{S}, f) \coloneqq \lim_{\varepsilon\to 0} r(\mathcal{S}, f, \varepsilon)$, e $s(\mathcal{S}, f) \coloneqq \lim_{\varepsilon\to 0} s(\mathcal{S}, f, \varepsilon)$.

Observação 4.1.8. Se dado conjunto $S \subset X$ é $(\mathcal{S}, n, \varepsilon)$-separado maximal, então S é $(\mathcal{S}, n, \varepsilon)$-gerador. De fato, dado $y \in X \backslash S$, $S \cup \{y\}$ não é $(\mathcal{S}, n, \varepsilon)$-separado, pois S tem cardinalidade máxima. Portanto, existem $x \in S$ e $1 \leq m \leq n$ tais que $d(f^{t_m}(x), f^{t_m}(y)) \leq \varepsilon$, isto é, S é um conjunto $(\mathcal{S}, n, \varepsilon)$-gerador.

Observação 4.1.9. O espaço X pode ser coberto por um número finito, digamos, $k \in \mathbb{N}$ de $\mathcal{S}$-bolas dinâmicas de raio ε e comprimento n, por ser compacto: $X = \bigcup_{i=1}^{k} B(\mathcal{S}, x_k, n, \varepsilon)$. Assim, o conjunto $\{x_1, \ldots, x_k\}$ é $(\mathcal{S}, n, \varepsilon)$-gerador, e $r_n(\mathcal{S}, f, \varepsilon) \leq k < \infty$.

As grandezas $s_n(\mathcal{S}, f, \varepsilon)$ e $r_n(\mathcal{S}, f, \varepsilon)$ são monótonas com relação a variável ε, isto é, se $\varepsilon > \varepsilon_1 > 0$, então $s_n(\mathcal{S}, f, \varepsilon_1) \geq s_n(\mathcal{S}, f, \varepsilon)$, e $r_n(\mathcal{S}, f, \varepsilon_1) \geq r_n(\mathcal{S}, f, \varepsilon)$. Consequentemente, $s(\mathcal{S}, f, \varepsilon) \geq s(\mathcal{S}, f, \varepsilon_1)$, e $r(\mathcal{S}, f, \varepsilon) \geq r(\mathcal{S}, f, \varepsilon_1)$.

Lema 4.1.10. *Para todos $n \geq 1$ e $\varepsilon > 0$, tem-se $r_n(\mathcal{S}, f, \varepsilon) \leq s_n(\mathcal{S}, f, \varepsilon) \leq r_n(\mathcal{S}, f, \varepsilon/2)$*

Dem.: Análoga ao do Lema 3.1.20, trocando $0, \ldots, m, \ldots, n-1$ por $t_1, \ldots, t_m, \ldots, t_n$. ■

Pelo Lema 4.1.10 e pelas observações 4.1.8 e 4.1.9, segue que $s_n(\mathcal{S}, n, \varepsilon)$ é finito.

Dado $\varepsilon > 0$, o Lema 4.1.10 implica $r(\mathcal{S}, f, \varepsilon) \leq s(\mathcal{S}, f, \varepsilon) \leq r(\mathcal{S}, f, \varepsilon/2)$.

Tomando o limite com $\varepsilon \to 0$, obtemos $r(\mathcal{S}, f) = s(\mathcal{S}, f)$.

Observação 4.1.11. A noção de entropia segundo a Definição 4.1.12 para uma métrica d_1 é a mesma que para outra métrica d_2, desde que estas sejam topologicamente equivalentes.

O raciocínio é análogo ao da observação 3.1.22.

A seguir, o conceito de entropia sequencial afim ao de Bowen e Dinaburg:

Definição 4.1.12. A *entropia topológica sequencial de uma aplicação f* é definida como $r(\mathcal{S}, f) = s(\mathcal{S}, f)$.

Proposição 4.1.13 (Equivalência entre as definições)**.** *A entropia sequencial análoga ao conceito de Bowen e Dinaburg é a mesma entropia sequencial análoga ao conceito de Adler, Konheim e McAndrew. Isto é, $r(\mathcal{S}, f) = h_{\mathcal{S}}(f) = s(\mathcal{S}, f)$.*

Dem.: Pelo Lema 4.1.10, é suficiente mostrar que $s(\mathcal{S}, f) \leq h_{\mathcal{S}}(f) \leq r(\mathcal{S}, f)$.

Fixemos $\varepsilon > 0$ e $n \in \mathbb{N}$. Seja $V \subset X$ um conjunto $(\mathcal{S}, n, \varepsilon)$-separado maximal e α cobertura aberta, com diâmetro menor que ε. Se x e y estão no mesmo elemento de $\alpha^n_{f,\mathcal{S}}$, então

$$d(f^{t_k}(x), f^{t_k}(y)) \leq \text{diam } \alpha < \varepsilon, \ \forall\ 1 \leq k \leq n.$$

Em particular, cada elemento de $\alpha^n_{f,\mathcal{S}}$ não contém mais que um elemento de V, pois $x, y \in V$ resulta em $d(f^{t_k}(x), f^{t_k}(y)) > \varepsilon$ para algum $1 \leq k \leq n$.

Portanto, $\#V \leq N(\alpha^n_{f,\mathcal{S}}) \Rightarrow s_n(\mathcal{S}, f, \varepsilon) \leq N(\alpha^n_{f,\mathcal{S}})$, para todo $n \in \mathbb{N}$, e:

$$s(\mathcal{S}, f, \varepsilon) \leq \limsup_n \frac{1}{n} \log s_n(\mathcal{S}, f, \varepsilon) \leq \lim_n \frac{1}{n} \log N(\alpha^n_{f,\mathcal{S}}) = h_{\mathcal{S}}(f, \alpha) \leq h_{\mathcal{S}}(f).$$

Fazendo $\varepsilon \to 0$, obtemos $s(\mathcal{S}, f) \leq h_{\mathcal{S}}(f)$.

Por outro lado, sejam α uma cobertura e $\varepsilon > 0$ o número de Lebesgue de α. Seja $R \subset X$ um conjunto $(\mathcal{S}, n, \varepsilon/2)$-gerador com cardinalidade mínima. Para cada $x \in R$ e $1 \leq k \leq n$, existe $A^x_k \in \alpha$ tal que $B(f^{t_k}(x), \varepsilon/2) \subset A^x_k$.

Então: $B(\mathcal{S}, x, n, \varepsilon/2) \subset \bigcap_{i=1}^{n} f^{-t_i}(A^x_i)$.

Pelo fato de R ser $(\mathcal{S}, n, \varepsilon/2)$-gerador, o conjunto $\{B(\mathcal{S}, x, n, \varepsilon/2); x \in R\}$ é uma cobertura de X. Também o é, então, o conjunto $\beta = \left\{\bigcap_{i=1}^{n} f^{-t_i}(A^x_i); x \in R\right\}$, que é uma subcobertura de $\alpha^n_{f,\mathcal{S}}$ com, no máximo, $\#R$ elementos.

Segue disto que $N(\alpha^n_{f,\mathcal{S}}) \leq \#R = r_n(\mathcal{S}, f, \varepsilon/2)$, para todo $n \in \mathbb{N}$.

Portanto, $h_{\mathcal{S}}(f, \alpha) = \lim_n \frac{1}{n} H(\alpha^n_{f,\mathcal{S}}) \leq \limsup_n \frac{1}{n} \log r_n(\mathcal{S}, f, \varepsilon/2) = r(\mathcal{S}, f, \varepsilon/2)$.

Fazendo $\varepsilon \to 0$, obtemos $h_{\mathcal{S}}(f, \alpha) \leq r(\mathcal{S}, f) \Rightarrow h_{\mathcal{S}}(f) \leq r(\mathcal{S}, f)$. ■

Proposição 4.1.14. *(1, 10) Sejam X, Y compactos, $\mathcal{S} = \{t_n\}_n$ uma sequência de inteiros não negativos e $f : X \to X$, $g : Y \to Y$ aplicações contínuas. Então $h_{\mathcal{S}}(f \times g) \leq h_{\mathcal{S}}(f) + h_{\mathcal{S}}(g)$, onde $f \times g : X \times Y \to X \times Y$ é definida por $(f \times g)(x, y) = (f(x), g(y))$. Caso $\mathcal{S} = \{n-1\}_n$ ou $f = g$ e $X = Y$, vale a igualdade.*

Dem.: Sendo α e β coberturas, definimos a cobertura $\alpha \times \beta = \{A \times B; A \in \alpha, B \in \beta\}$.

A continuidade de f e g garante a continuidade de $f \times g$. Logo, $(f \times g)^{-1}(\alpha \times \beta)$ é uma cobertura de $X \times Y$.

Note que as coberturas de $X \times Y$ da forma $\alpha \times \beta = \{A \times B; A \in \alpha, B \in \beta\}$ possuem as seguintes propriedades: se α e α_1 são coberturas de X e β e β_1 são coberturas de Y, então

- $N(\alpha \times \beta) = N(\alpha)N(\beta)$;
- $(\alpha \times \beta) \vee (\alpha_1 \times \beta_1) = (\alpha \vee \alpha_1) \times (\beta \vee \beta_1)$;
- $(f \times g)^{-1}(\alpha \times \beta) = f^{-1}(\alpha) \times g^{-1}(\beta)$.

Daí,

$$(\alpha\times\beta)^n_{f\times g,\mathcal{S}} = f^{-t_1}(\alpha\times\beta)\vee(f\times g)^{-t_2}(\alpha\times\beta)\vee\cdots\vee(f\times g)^{-t_n}(\alpha\times\beta) =$$
$$= (f^{-t_1}(\alpha)\vee f^{-t_2}(\alpha)\vee\cdots\vee f^{-t_n}(\alpha))\times(f^{-t_1}(\beta)\vee f^{-t_2}(\beta)\vee\cdots\vee f^{-t_n}(\beta)) = \alpha^n_{f,\mathcal{S}}\times\beta^n_{g,\mathcal{S}}.$$

Portanto,

$$N((\alpha\times\beta)^n_{f\times g,\mathcal{S}}) = N(\alpha^n_{f,\mathcal{S}})N(\beta^n_{g,\mathcal{S}}) \Rightarrow H((\alpha\times\beta)^n_{f\times g,\mathcal{S}}) = H(\alpha^n_{f,\mathcal{S}}) + H(\beta^n_{g,\mathcal{S}}).$$

Segue disto que $h_{\mathcal{S}}(f\times g,\alpha\times\beta)\leq h_{\mathcal{S}}(f,\alpha)+h_{\mathcal{S}}(g,\beta) \Rightarrow h_{\mathcal{S}}(f\times g)\leq h_{\mathcal{S}}(f)+h_{\mathcal{S}}(g)$.

Caso $\mathcal{S}=\{n-1\}_n$, há um limite no lugar do limite superior, e ocorre a igualdade: $h(f\times g,\alpha\times\beta)=h(f,\alpha)+h(g,\beta)$, o que implica $h(f\times g)\geq h(f)+h(g)$. De fato, se isto não ocorresse, existiriam α', β' coberturas tais que $h(f\times g,\alpha'\times\beta')\leq h(f\times g) < h(f,\alpha')+h(g,\beta') < h(f)+h(g)$, gerando uma contradição.

Para a outra desigualdade, mostraremos que, se γ é cobertura de $X\times Y$, existe uma outra cobertura do tipo $\alpha\times\beta$ que refina γ, onde α cobre X e β cobre Y. Assim, $h(f\times g,\gamma)\leq h(f\times g,\alpha\times\beta)=h(f,\alpha)+h(g,\beta)$, o que implica em $h(f\times g)\leq h(f,\alpha)+h(g,\beta)$.

Como todo aberto de $X\times Y$ é união de conjuntos da forma $A\times B$, onde A é aberto de X e B é aberto de Y, podemos obter um refinamento de γ, e deste, uma subcobertura minimal, γ_1, escrita como $\{A_1\times B_1,\ldots,A_{N(\gamma_1)}\times B_{N(\gamma_1)}\}$.

Sejam $\alpha_1=\{A_1,\ldots,A_{N(\gamma_1)}\}$ uma cobertura de X, $\beta_1=\{B_1,\ldots,B_{N(\gamma_1)}\}$ uma cobertura de Y.

Dado $x\in X$, denote V_x a interseção de todos os elementos de α_1 que contêm x. Dado $y\in Y$, denote W_y a interseção de todos os elementos de β_1 que contêm y. Tais conjuntos, assim definidos, são abertos.

Pelo fato de X e Y serem compactos, existem finitos $x_1,\ldots,x_n\in X$ e finitos $y_1,\ldots,y_m$ tais que $\nu=\{V_{x_1},\ldots,V_{x_n}\}$ cobre X, e $\omega=\{W_{y_1},\ldots,W_{y_m}\}$ cobre Y.

Considere um conjunto $V_{x_i}\times W_{y_j}\in\nu\times\omega$, $i=1,\ldots,n; j=1,\ldots,m$.

Como γ_1 cobre $X\times Y$, $(x_i,y_j)\in A_k\times B_k$, para algum $1\leq k\leq N(\gamma_1)$.

O fato de $x_i\in A_k$ e $y_j\in B_k$ implica em $V_{x_i}\subset A_k$ e $W_{y_j}\subset B_k$, isto é, $V_{x_i}\times W_{y_j}\subset A_k\times B_k$.

Isto significa que $\gamma\prec\gamma_1\prec\nu\times\omega$, o que é exatamente o que queríamos provar.

Agora, se $f=g$ e $X=Y$, seja S um conjunto $(\mathcal{S},n,\varepsilon)$-separado maximal em X.

Então, $S\times S$ é um conjunto $(\mathcal{S},n,\varepsilon)$-separado, não necessariamente maximal, em $X\times X$.

Daí, $s_n(\mathcal{S},f\times f,\varepsilon)\geq\#(S\times S)=(\#S)^2=s_n(\mathcal{S},f,\varepsilon)^2$.

Por consequência,

$$h_{\mathcal{S}}(f \times f) \geq \limsup_{n} \frac{1}{n} \log s_n(\mathcal{S}, f \times f, \varepsilon)^2 = 2 \limsup_{n} \frac{1}{n} s_n(\mathcal{S}, f, \varepsilon),$$

donde se conclui que $h_{\mathcal{S}}(f \times f) \geq 2h_{\mathcal{S}}(f)$. ■

Observação 4.1.15. O texto de Lemańczyk (28) apresenta um exemplo que não satisfaz a igualdade na Proposição 4.1.14 quando f é diferente de g.

Proposição 4.1.16. *Seja $\{\alpha_k\}_{k\in\mathbb{N}}$ uma sequência de coberturas de X tal que* diam α_k converge para zero. Então $h_{\mathcal{S}}(f) = \sup_k h_{\mathcal{S}}(f, \alpha_k) = \lim_k h_{\mathcal{S}}(f, \alpha_k)$.

Dem.: Análoga à da Proposição 3.1.13. ■.

4.2 ENTROPIA MÉTRICA SEQUENCIAL

Como na seção que trata sobre entropia métrica clássica, subentende-se $(X, \mathcal{B}, \mu)$ um espaço de probabilidade, com X um espaço métrico compacto, $\mathcal{B}$ a σ-álgebra de Borel, e μ uma medida de probabilidade boreliana, além de $f : X \to X$ mensurável, com μ invariante por f.

Definição 4.2.1. Dada $\mathcal{S} = \{t_n\}_n$ uma sequência de inteiros não negativos, definimos a partição *n-ésima $f, \mathcal{S}$-junção de $\mathcal{P}$* pondo

$$\mathcal{P}^n_{f,\mathcal{S}} = f^{-t_1}(\mathcal{P}) \vee f^{-t_2}(\mathcal{P}) \vee \cdots \vee f^{-t_n}(\mathcal{P}) = \bigvee_{i=1}^{n} f^{-t_i}(\mathcal{P}).$$

Definição 4.2.2. A *entropia métrica sequencial de uma aplicação f com respeito a uma partição $\mathcal{P}$ e uma sequência de inteiros não negativos $\mathcal{S}$, $h_{\mu,\mathcal{S}}(f, \mathcal{P})$*, é definida pelo limite superior

$$h_{\mu,\mathcal{S}}(f, \mathcal{P}) = \limsup_{n} \frac{1}{n} H_\mu(\mathcal{P}^n_{f,\mathcal{S}}).$$

Definição 4.2.3. A *entropia métrica sequencial de uma aplicação f com respeito a uma sequência de inteiros não negativos $\mathcal{S}$, $h_{\mu,\mathcal{S}}(f)$*, é definida como o supremo do conjunto $\{h_{\mu,\mathcal{S}}(f, \mathcal{P}); \mathcal{P}$ é partição com entropia finita$\}$:

$$h_{\mu,\mathcal{S}}(f) = \sup_{\mathcal{P}} h_{\mu,\mathcal{S}}(f, \mathcal{P}).$$

A entropia descrita acima também pode ser chamada de $\mathcal{S}$-entropia.

Proposição 4.2.4. *Sejam $\mu \in \mathcal{M}(X, f)$, $\mathcal{S} = \{t_n\}_n$ uma sequência de inteiros não negativos e $\mathcal{P}, \mathcal{Q}$ duas partições. Então:*

1. $H_\mu(\mathcal{P}^n_{f,\mathcal{S}} | \mathcal{Q}^n_{f,\mathcal{S}}) \leq n H_\mu(\mathcal{P}|\mathcal{Q})$.

2. $h_{\mu,\mathcal{S}}(f, \mathcal{P}) \leq h_{\mu,\mathcal{S}}(f, \mathcal{Q}) + H_\mu(\mathcal{P}|\mathcal{Q})$.

Dem.: Análoga aos itens 3.2.7-4b e 3.2.14-3. ■

Proposição 4.2.5. *A Definição 4.2.3 não é alterada se tomarmos o supremo sobre partições finitas.*

Dem.: Análoga à da Proposição 3.2.15, fazendo uso da Proposição 4.2.4. ■.

Chamamos de $\epsilon = \{x\}_{x \in X}$ a partição de X em pontos.

Dizemos que uma partição $\mathcal{P}$ é f*-geradora* se $\bigvee_{i=0}^{\infty} f^{-i}(\mathcal{P}) = \epsilon$, a menos de medida nula.

Lema 4.2.6. *(7) Seja $\{\mathcal{P}_k\}_{k \in \mathbb{N}}$ uma sequência de partições tais que $\mathcal{P}_1 \prec \mathcal{P}_2 \prec \cdots \prec \mathcal{P}_n \prec \ldots$ e $\bigvee_{i=1}^{\infty} \mathcal{P}_i = \epsilon$. Então, o conjunto $\mathcal{J} = \{\mathcal{Q} \in \mathcal{Z}_\mu; \mathcal{Q} \prec \mathcal{P}_k$ para algum $k\}$ é denso no conjunto das partições finitas de $\mathcal{Z}_\mu$, na métrica de Rokhlin.*

Demonstração. Seja $\mathcal{R} = \{R_1, \ldots, R_m\}$ uma partição e $\varepsilon > 0$.

Como $\mathcal{P}_i \prec \mathcal{P}_{i+1}$ e $\bigvee_{i=1}^{\infty} \mathcal{P}_i = \epsilon$, dado $\delta > 0$, é possível obter $n_0 \in \mathbb{N}$ e uma partição $\beta = \{B_1, \ldots, B_m\}$ formada por uniões de elementos de $\mathcal{P}_{n_0}$, de sorte que $\mu(R_i \Delta B_i) < \delta$, $i = 1, \ldots, m$.

Note que $\beta \prec \mathcal{P}_{n_0}$.

Pela Proposição 3.2.9, escolhendo δ apropriadamente, ocorre $\rho_\mu(\mathcal{R}, \beta) < \varepsilon$. ■

Lema 4.2.7. *Sejam $\mathcal{P}, \mathcal{Q}$ duas partições, $\mathcal{S} = \{t_n\}_n$ uma sequência de inteiros não negativos e $\mu \in \mathcal{M}(X, f)$. Então $|h_{\mu,\mathcal{S}}(f, \mathcal{P}) - h_{\mu,\mathcal{S}}(f, \mathcal{Q})| \leq \rho_\mu(\mathcal{P}, \mathcal{Q})$, isto é, a aplicação $\mathcal{P} \mapsto h_{\mu,\mathcal{S}}(f, \mathcal{P})$ é Lipschitz em relação à métrica de Rokhlin.*

Dem.: Por definição, temos que

$$H_\mu(\mathcal{P}^n_{f,\mathcal{S}} \vee \mathcal{Q}^n_{f,\mathcal{S}}) = H_\mu(\mathcal{P}^n_{f,\mathcal{S}} | \mathcal{Q}^n_{f,\mathcal{S}}) + H_\mu(\mathcal{Q}^n_{f,\mathcal{S}}) = H_\mu(\mathcal{Q}^n_{f,\mathcal{S}} | \mathcal{P}^n_{f,\mathcal{S}}) + H_\mu(\mathcal{P}^n_{f,\mathcal{S}}). \tag{4.1}$$

Pela Proposição 4.2.4,

$$H_\mu(\mathcal{Q}^n_{f,\mathcal{S}} | \mathcal{P}^n_{f,\mathcal{S}}) \leq n H_\mu(\mathcal{Q}|\mathcal{P}), \text{ e } H_\mu(\mathcal{P}^n_{f,\mathcal{S}} | \mathcal{Q}^n_{f,\mathcal{S}}) \leq n H_\mu(\mathcal{P}|\mathcal{Q}). \tag{4.2}$$

Subtraindo as igualdades de 4.1 e usando as desigualdades de 4.2, obtemos

$$\left| \frac{1}{n} H_\mu(\mathcal{P}^n_{f,\mathcal{S}}) - \frac{1}{n} H_\mu(\mathcal{Q}^n_{f,\mathcal{S}}) \right| \leq H_\mu(\mathcal{P}|\mathcal{Q}) + H_\mu(\mathcal{Q}|\mathcal{P}) = \rho_\mu(\mathcal{P}, \mathcal{Q}).$$

Passando ao limite com $n \to \infty$, infere-se o resultado. ■

Teorema 4.2.8 (Kolmogorov-Sinai)**.** *(9) Sejam $\{\mathcal{P}_k\}_{k\in\mathbb{N}}$ uma sequência de partições finitas tais que $\mathcal{P}_1 \prec \mathcal{P}_2 \prec \cdots \prec \mathcal{P}_n \prec \ldots$ e $\bigvee_{i=1}^{\infty} \mathcal{P}_i = \epsilon$, e $\mu \in \mathcal{M}(X,f)$. Então,*

$$h_{\mu,\mathcal{S}}(f) = \lim_n h_{\mu,\mathcal{S}}(f, \mathcal{P}_n).$$

Dem.: Dado $\delta > 0$, seja $\mathcal{R}$ uma partição finita satisfazendo $h_{\mu,\mathcal{S}}(f) < h_{\mu,\mathcal{S}}(f,\mathcal{R}) + \delta/2$. Isto é possível por causa da Proposição 4.2.5.

Pelo Lema 4.2.6, existem k inteiro e $\mathcal{D} \in \mathcal{J} = \{\mathcal{Q} \in \mathcal{Z}_\mu; \mathcal{Q} \prec \mathcal{P}_i$ para algum $i\}$ tais que $\rho_\mu(\mathcal{R},\mathcal{D}) < \delta/2$, e $\mathcal{D} \prec \mathcal{P}_k$; este último implica em $h_{\mu,\mathcal{S}}(f) - h_{\mu,\mathcal{S}}(f,\mathcal{D}) \geq h_{\mu,\mathcal{S}}(f) - h_{\mu,\mathcal{S}}(f,\mathcal{P}_k)$.

Ainda, usando o Lema 4.2.7, $h_{\mu,\mathcal{S}}(f,\mathcal{R}) - h_{\mu,\mathcal{S}}(f,\mathcal{D}) \leq \rho_\mu(\mathcal{P},\mathcal{D}) < \delta/2$.

Logo, $h_{\mu,\mathcal{S}}(f) - h_{\mu,\mathcal{S}}(f,\mathcal{D}) < \delta$, o que implica em $h_{\mu,\mathcal{S}}(f) - h_{\mu,\mathcal{S}}(f,\mathcal{P}_k) < \delta$.

Portanto, $\lim_k h_{\mu,\mathcal{S}}(f,\mathcal{P}_k) = h_{\mu,\mathcal{S}}(f)$. ■

Corolário 4.2.9. *Se $\mathcal{P}$ é f-geradora e μ está em $\mathcal{M}(X,f)$, então $\mathcal{P}$ realiza o supremo $h_\mu(f) = h_\mu(f,\mathcal{P})$.*

Dem.: Forme a sequência $\{\mathcal{P}_k\}_{k\in\mathbb{N}}$ pondo $\mathcal{P}_k = \mathcal{P}_f^k$, para cada k.

Por $\mathcal{P}$ ser f-geradora, o Teorema 4.2.8 permite escrever $h_\mu(f) = \lim_n h_\mu(f,\mathcal{P}_f^n)$.

Usando o item 3.2.14-4, $h_\mu(f,\mathcal{P}_f^n) = h_\mu(f,\mathcal{P})$, para todo n.

Logo, $h_\mu(f) = h_\mu(f,\mathcal{P})$. ■

Note que o item 3.2.14-4 não é valido para uma sequência $\mathcal{S}$ qualquer. Por exemplo, se $\mathcal{S} = \{2^n\}_{n\in\mathbb{N}}$, então:

$$\begin{aligned}(\mathcal{P}_{f,\mathcal{S}}^3)_{f,\mathcal{S}}^2 &= f^{-1}(f^{-1}(\mathcal{P}) \vee f^{-2}(\mathcal{P}) \vee f^{-4}(\mathcal{P})) \bigvee f^{-2}(f^{-1}(\mathcal{P}) \vee f^{-2}(\mathcal{P}) \vee f^{-4}(\mathcal{P})) = \\ &= f^{-2}(\mathcal{P}) \vee f^{-3}(\mathcal{P}) \vee f^{-4}(\mathcal{P}) \vee f^{-5}(\mathcal{P}) \vee f^{-6}(\mathcal{P}).\end{aligned}$$

Mas $\mathcal{P}_{f,\mathcal{S}}^{3+2-1} = \mathcal{P}_{f,\mathcal{S}}^4 = f^{-1}(\mathcal{P}) \vee f^{-2}(\mathcal{P}) \vee f^{-4}(\mathcal{P}) \vee f^{-8}(\mathcal{P}) \neq (\mathcal{P}_{f,\mathcal{S}}^3)_{f,\mathcal{S}}^2$.

Exemplo 4.2.10. (19) Seja $\{\mathcal{A},\mathcal{B},\nu\}$ um espaço de probabilidade boreliano qualquer.

Denotamos por $\mathcal{A}^\mathbb{N}$ o espaço das sequências $(x_1 x_2 \ldots x_n \ldots) = (x_n)_n$, onde cada x_i, $i \in \mathbb{N}$ está em $\mathcal{A}$. Munimos tal espaço com σ-álgebra Σ gerada pela álgebra dos cilindros, que são subconjuntos da forma

$$[m; A_m, \ldots, A_n] \coloneqq \{x \in \mathcal{A}^\mathbb{N}; x_i \in A_i \subset \mathcal{A}, m \leq i \leq n\},$$

juntamente com a medida μ, dada por $\mu([m; A_m, \ldots, A_n]) = \nu(A_m) \cdots \nu(A_n)$.

Considere a aplicação *shift* $\sigma : \mathcal{A}^\mathbb{N} \to \mathcal{A}^\mathbb{N}$, pondo $\sigma((x_n)_n) = (x_{n+1})_n$.

Tomemos o caso particular $\mathcal{A} = \{1, \ldots, k\}$, com $\nu(\{i\}) = p_i = 1/k, 1 \leq i \leq k$. Este sistema, (σ, μ) é conhecido como um *shift de Bernoulli*, com pesos $(1/k, \ldots, 1/k)$.

Considere $\mathcal{P} = \{P_1, P_2, \ldots, P_k\}$ uma partição de $\mathcal{A}$, com $P_i = \{x \in \mathcal{A}; x_1 = i\}, i = 1 \ldots, k$. Observe que $\mathcal{P}_\sigma^n$ é a partição em cilindros $[0; x_1, \ldots, x_n]$ com n elementos fixados.

Daí,

$$\begin{aligned} H_\mu(\mathcal{P}_\sigma^n) &= \sum_{x_1,\ldots,x_n} -p_{x_1} \cdots p_{x_n} \log(p_{x_1} \cdots p_{x_n}) = \sum_{x_1,\ldots,x_n} -p_{x_1} \cdots p_{x_n} \left(\sum_r \log p_{x_r} \right) = \\ &= \sum_r \sum_{x_r} -p_{x_r} \log(p_{x_r}) \sum_{x_i, i \neq r} p_{x_1} \cdots p_{x_{r-1}} p_{x_{r+1}} \cdots p_{x_n} = \sum_r \sum_{x_r} -p_{x_r} \log p_{x_r} = \\ &= -n \sum_{i=1}^k p_i \log p_i = n \log k. \end{aligned}$$

Assim, $h_\mu(\sigma, \mathcal{P}) = \lim_n \frac{1}{n} H_\mu(\mathcal{P}_\sigma^n) = \lim_n \log k = \log k$.

Note, também, que μ é invariante por σ, pois o é em cada elemento da álgebra gerada pelos cilindros (19).

$\mathcal{P}$ é σ-geradora, pois a interseção de infinitos elementos das pré imagens de $\mathcal{P}$ contém um único elemento de $\mathcal{A}^\mathbb{N}$.

Por 4.2.8, a entropia métrica de σ com relação a μ é é $h_\mu(\sigma) = h_\mu(\sigma, \mathcal{P}) = \log k$.

Proposição 4.2.11. *(9) Sejam $\mathcal{S} = \{t_n\}_n$ uma sequência de inteiros não negativos, $(X, , \mathcal{B}, \mu)$, $(Y, \mathcal{C}, \nu)$ espaços de medida com respectivas σ-álgebras de Borel, e $f : X \to X$, $g : Y \to Y$ aplicações mensuráveis, tais que $\mu \times \nu$ é invariante por $f \times g$. Então $h_{\mu\times\nu,\mathcal{S}}(f \times g) \leq h_{\mu,\mathcal{S}}(f) + h_{\nu,\mathcal{S}}(g)$, onde $f \times g : X \times Y \to X \times Y$ é definida por $(f \times g)(x, y) = (f(x), g(y))$. Caso $\mathcal{S} = \{n-1\}_n$, ou $f = g$, $X = Y$ e $\mu = \nu$, ocorre a igualdade.*

Dem.: Sendo $\mathcal{P}$ uma partição de X e $\mathcal{Q}$ uma partição de Y, definimos a partição $\mathcal{P} \times \mathcal{Q} = \{P \times Q; P \in \mathcal{P}, Q \in \mathcal{Q}\}$.

A mensurabilidade de f e g garante que $f \times g$ é mensurável. Logo, $(f \times g)^{-1}(\mathcal{P} \times \mathcal{Q})$ é uma partição de $X \times Y$.

Sejam ϵ_X e ϵ_Y as partições de X e Y em pontos, respectivamente.

Note que as partições de $X \times Y$ da forma $\mathcal{P} \times \mathcal{Q} = \{P \times Q; P \in \mathcal{P}, Q \in \mathcal{Q}\}$ possuem as seguintes propriedades: se $\mathcal{P}$ e $\mathcal{P}_1$ são partições de X e $\mathcal{Q}$ e $\mathcal{Q}_1$ são partições de Y, então

- $(\mathcal{P} \times \mathcal{Q}) \vee (\mathcal{P}_1 \times \mathcal{Q}_1) = (\mathcal{P} \vee \mathcal{P}_1) \times (\mathcal{Q} \vee \mathcal{Q}_1)$;
- $(f \times g)^{-1}(\mathcal{P} \times \mathcal{Q}) = f^{-1}(\mathcal{P}) \times g^{-1}(\mathcal{Q})$.

De modo análogo ao desenvolvido na Proposição 4.1.14, $(\mathcal{P} \times \mathcal{Q})_{f\times g,\mathcal{S}}^n = \mathcal{P}_{f,\mathcal{S}}^n \times \mathcal{Q}_{g,\mathcal{S}}^n$.

Então,

$$H_{\mu\times\nu}\Big((\mathcal{P}\times\mathcal{Q})^n_{f\times g,\mathcal{S}}\Big) = \sum_{P\in\mathcal{P}^n_{f,\mathcal{S}}}\sum_{Q\in\mathcal{Q}^n_{g,\mathcal{S}}}(\mu\times\nu)(P\times Q)\log((\mu\times\nu)(P\times Q) =$$
$$= \sum_{P,Q}\mu(P)\nu(Q)\log(\mu(P)\nu(Q)) =$$
$$= \sum_{P,Q}\mu(P)\nu(Q)\log\mu(P) + \sum_{P,Q}\mu(P)\nu(Q)\log\nu(Q) = H_\mu(\mathcal{P}^n_{f,\mathcal{S}}) + H_\nu(\mathcal{Q}^n_{g,\mathcal{S}}). \tag{4.3}$$

Lembrando que o limite superior da soma é menor ou igual à soma dos limites superiores, $h_{\mu\times\nu,\mathcal{S}}(f\times g,\mathcal{P}\times\mathcal{Q}) \leq h_{\mu,\mathcal{S}}(f,\mathcal{P}) + h_{\nu,\mathcal{S}}(g,\mathcal{Q})$.

Tomemos sequências de partições $\{\mathcal{P}_k\}_{k\in\mathbb{N}}$ e $\{\mathcal{Q}_k\}_{k\in\mathbb{N}}$ tais que $\mathcal{P}_1 \prec \cdots \prec \mathcal{P}_n \prec \ldots$, $\mathcal{Q}_1 \prec \cdots \prec \mathcal{Q}_n \prec \ldots$, $\bigvee_{i=1}^{\infty}\mathcal{P}_i = \epsilon_X$ e $\bigvee_{i=1}^{\infty}\mathcal{Q}_i = \epsilon_Y$.

Então $\mathcal{P}_1\times\mathcal{Q}_1 \prec \cdots \prec \mathcal{P}_n\times\mathcal{Q}_n \prec \ldots$ e $\bigvee_{i=1}^{\infty}(\mathcal{P}_i\times\mathcal{Q}_i) = \epsilon_X\times\epsilon_Y$.

Pelo Teorema 4.2.8, $h_{\mu\times\nu,\mathcal{S}}(f\times g) = \lim_k h_{\mu\times\nu,\mathcal{S}}(f\times g,\mathcal{P}_k\times\mathcal{Q}_k)$.

Mas, para cada k, $h_{\mu\times\nu,\mathcal{S}}(f\times g,\mathcal{P}_k\times\mathcal{Q}_k) \leq h_{\mu,\mathcal{S}}(f,\mathcal{P}_k) + h_{\nu,\mathcal{S}}(g,\mathcal{Q}_k)$.

Logo, passando ao limite em k, obtemos $h_{\mu\times\nu,\mathcal{S}}(f\times g) \leq h_{\mu,\mathcal{S}}(f) + h_{\nu,\mathcal{S}}(g)$.

No caso em que $\mathcal{S} = \{n-1\}_{n\in\mathbb{N}}$, vale a igualdade, pois toma-se o limite usual, em vez do limite superior, na expressão $\frac{1}{n}H_\mu(\mathcal{P}^n_f)$, dado que tal existe.

Sob as cincurstâncias $f = g$, $X = Y$ e $\mu = \nu$, pelo uso de 4.3, obtemos

$$H_{\mu\times\mu}\Big((\mathcal{P}\times\mathcal{P})^n_{f\times f,\mathcal{S}}\Big) = 2H_\mu(\mathcal{P}^n_{f,\mathcal{S}}),\ \text{o que implica}\ h_{\mu\times\mu,\mathcal{S}}(f\times f,\mathcal{P}\times\mathcal{P}) = 2h_{\mu,\mathcal{S}}(f,\mathcal{P}).$$

Daí, tomamos o supremo sobre as partições $\mathcal{P}$: $\sup_{\mathcal{P}} h_{\mu\times\mu,\mathcal{S}}(f\times f,\mathcal{P}\times\mathcal{P}) = 2h_{\mu,\mathcal{S}}(f)$.

O número $\sup_{\mathcal{R}} h_{\mu\times\mu,\mathcal{S}}(f\times f,\mathcal{R})$, em que $\mathcal{R}$ varre todas as partições possíveis de $X\times X$, é maior ou igual que o supremo que considera somente partições do tipo $\mathcal{P}\times\mathcal{P}$, onde $\mathcal{P}$ particiona X.

Logo, $h_{\mu\times\mu,\mathcal{S}}(f\times f) \geq \sup_{\mathcal{P}} h_{\mu\times\mu,\mathcal{S}}(f\times f,\mathcal{P}\times\mathcal{P}) = 2h_{\mu,\mathcal{S}}(f)$ e, então, vale a igualdade. ■

Observação 4.2.12. O texto de Lemańczyk (28) apresenta um exemplo que não satisfaz a igualdade na Proposição 4.2.11 quando f é diferente de g.

A identidade $h(f^k) = kh(f), k \in \mathbb{N}$, não é, em geral, verdadeira para entropia sequencial. A seguir, vamos expor um contraexemplo (12).

Novamente, consideremos a aplicação *shift* $\sigma : \mathbb{N}^{\mathbb{N}} \to \mathbb{N}^{\mathbb{N}}$, pondo $\sigma(x_1x_2\ldots) = (x_2x_3\ldots)$. Podemos compô-la indutivamente: $\sigma^k(x_1x_2\ldots) = (x_{k+1}x_{k+2}\ldots)$.

Se $\mathcal{S} = \{t_i\}$ é uma sequência de inteiros não negativos, a sequência *shift* de $\mathcal{S}$ deslocada k vezes é definida pondo $\sigma^k(\mathcal{S}) = \{t_i\}_{i>k}$.

Lema 4.2.13. *Seja $\mathcal{S} = \{t_n\}_n$ uma sequência de inteiros não negativos e $f : X \to X$.*

1. *Se f é mensurável, invariante por μ e $\mathcal{P}$ é partição finita, então $h_{\mu,\mathcal{S}}(f,\mathcal{P}) = h_{\mu,\sigma^k(\mathcal{S})}(f,\mathcal{P})$.*

2. *Se f é contínua e α é cobertura finita, então $h_{\mathcal{S}}(f,\alpha) = h_{\sigma^k(\mathcal{S})}(f,\alpha)$.*

Em particular, ocorrem $h_{\mu,\mathcal{S}}(f) = h_{\mu,\sigma^k(\mathcal{S})}(f)$e $h_{\mathcal{S}}(f,\alpha) = h_{\sigma^k(\mathcal{S})}(f,\alpha)$.

Dem.: 1. Se $n \geq k$, então $H_\mu(\mathcal{P}^n_{f,\mathcal{S}}) \leq H_\mu(\mathcal{P}^k_{f,\mathcal{S}}) + H_\mu\left(\bigvee_{i=k+1}^{n} f^{-t_i}(\mathcal{P})\right)$.

Passando ao limite com $n \to \infty$, obtemos $h_{\mu,\mathcal{S}}(f,\mathcal{P}) \leq h_{\mu,\sigma^k(\mathcal{S})}(f,\mathcal{P})$

Por outro lado, $H_\mu(\mathcal{P}^n_{f,\mathcal{S}}) \geq H_\mu\left(\bigvee_{i=k+1}^{n} f^{-t_i}(\mathcal{P})\right)$,

Passando ao limite novamente, logra-se $h_{\mu,\mathcal{S}}(f,\mathcal{P}) \geq h_{\mu,\sigma^k(\mathcal{S})}(f,\mathcal{P})$.

2. Análoga ao do item (1), substituindo $\mathcal{P}$ por α e H_μ por H. ■

Proposição 4.2.14. *Sejam $k \in \mathbb{N}$ e $\mathcal{S} = \{k^n\}_{n\in\mathbb{N}}$.*

1. *Se f é invariante por μ, então, para todo $m \in \mathbb{N}$ $h_{\mu,\mathcal{S}}(f) = h_{\mu,\mathcal{S}}(f^{k^m})$.*

2. *Se f é contínua, então, para todo $m \in \mathbb{N}$, $h_{\mathcal{S}}(f) = h_{\mathcal{S}}(f^{k^m})$.*

Dem.: 1. Sejam dados $\mathcal{P}$ uma partição e $m \in \mathbb{N}$. Note que, se $\mathcal{S} = \{t_n\}_n$ é sequência de inteiros não negativos, então $h_{\mu,\mathcal{S}}(f^m,\mathcal{P}) = h_{\mu,m\mathcal{S}}(f,\mathcal{P})$, onde $m\mathcal{S} = \{mt_n\}_n$.

Daí, dado $k \in \mathbb{N}$, se $t_n = k^n$ para todo n, então $h_{\mu,\mathcal{S}}(f^{k^m},\mathcal{P}) = h_{\mu,k^m\mathcal{S}}(f,\mathcal{P})$

Mas $k^m\mathcal{S} = \{t_n\}_{n>m-1} = \sigma^{m-1}(\mathcal{S})$. Pelo Lema 4.2.13, $h_{\mu,\mathcal{S}}(f^{k^m},\mathcal{P}) = h_{\mu,\sigma^{m-1}(\mathcal{S})}(f,\mathcal{P}) = h_{\mu,\mathcal{S}}(f,\mathcal{P})$.

Logo, $h_{\mu,\mathcal{S}}(f) = h_{\mu,\mathcal{S}}(f^{k^m})$.

2. Análogo ao item anterior, substituindo $\mathcal{P}$ por α e H_μ por H. ■

Exemplo 4.2.15. (19) Se a medida μ está suportada num número finito de átomos, então sua entropia métrica sequencial é nula.

De fato, μ só pode tomar um conjunto finito de valores, nesse caso. Por consequência, a entropia $H_\mu(\mathcal{P})$ também só toma um número finito de valores, quando consideramos as partições $\mathcal{P}$ finitas.

Por isto, para qualquer sequência $\mathcal{S}$ de inteiros não negativos, $\lim_n \frac{1}{n} H_\mu(\mathcal{P}^n_f) = 0$, para toda partição finita $\mathcal{P}$.

Logo, $h_{\mu,\mathcal{S}}(f) = 0$.

No caso em que μ é ergódica e possui algum átomo, seu suporte será dado por uma órbita periódica, pela Proposição 2.2.11. Assim, ocorre $h_{\mu,\mathcal{S}}(f) = 0$.

4.3 O PRINCÍPIO VARIACIONAL SEQUENCIAL

A seguir, serão apresentados três resultados preliminares de Goodman (10), que servirão de alicerce para suster o Princípio Variacional para entropia sequencial, ou simplesmente Princípio Variacional Sequencial.

Definição 4.3.1. A *ordem* de uma coleção de conjuntos $\alpha = \{A_i\}_{i\in I}$, Ord α, é o número máximo de conjuntos de α que intersectam entre si.

Definição 4.3.2. A *dimensão de recobrimento* de um espaço topológico X, $\dim X$, é o menor valor de n de modo que toda cobertura aberta finita de X possui refinamento de ordem no máximo $n + 1$.

Observação 4.3.3. A dimensão de recobrimento de $\mathbb{R}^n$ coincide com a dimensão como espaço vetorial, isto é, n (29).

Exemplo 4.3.4. (25) Qualquer subespaço compacto de $\mathbb{R}$ possui dimensão de recobrimento no máximo igual a 1.

Para provar isto, em primeiro lugar, definimos as coleções de intervalos abertos

$$\alpha_1 = \{(n, n + 1); n \in \mathbb{Z}\} \text{ e } \alpha_2 = \{(n - 1/2, n + 1/2); n \in \mathbb{Z}\}.$$

Assim, $\alpha = \alpha_1 \cup \alpha_2$ é uma cobertura aberta de $\mathbb{R}$ com ordem 2, pois cada número real está em, no máximo, 2 elementos de α.

Agora, seja X um subconjunto compacto de $\mathbb{R}$. Dada uma cobertura aberta β de X, esta possui número de Lebesgue (Proposição 3.1.12), digamos, $\delta > 0$.

Isto significa que toda cobertura aberta de X com diâmetro menor ou igual a δ é um refinamento de β.

Considere o homeomorfismo $f : \mathbb{R} \to \mathbb{R}$ dado por $f(x) = \delta x/2$.

Note que $f(\alpha) = \{f(A); A \in \alpha\}$ é uma cobertura aberta de $\mathbb{R}$, com diâmetro $\delta/2$.

As interseções dos elementos de $f(\alpha)$ com X formam uma cobertura aberta de X que refina β. Assim, obtemos um refinamento de β com ordem, no máximo 2, o que implica $\dim X \leq 1$.

Afirma-se, também, que um intervalo fechado qualquer na reta possui dimensão de recobrimento igual a 1.

De fato, seja $Y = [a, b]$, com $a < b$ números reais. Temos $\dim Y \leq 1$, pois Y é compacto.

Seja $\gamma = \{[a,b),(a,b]\}$ uma cobertura de Y. Seus elementos são abertos em Y.

Se γ_1 é uma cobertura aberta de Y que refina γ, então γ_1 possui, ao menos, 2 elementos.

Sejam U um dos elementos de γ_1 e V a união dos outros elementos.

Se γ_1 tivesse ordem 1, então U e V seriam disjuntos, o que iria contradizer a conexidade do intervalo Y.

Logo, γ_1 possui, ao menos, ordem 2, o que implica $\dim Y \geq 1$.

De maneira análoga, pode-se mostrar que o círculo unitário S^1 também possui dimensão de recobrimento 1.

Teorema 4.3.5. *Suponha que X possua dimensão de recobrimento finita. Então, para quaisquer sequência de inteiros não negativos $\mathcal{S}$ e medida μ em $\mathcal{M}(X,f)$, tem-se $h_{\mu,\mathcal{S}}(f) \leq h_{\mathcal{S}}(f)$.*

Dem.: Ponhamos $\dim X = k$ e seja $\mathcal{S} = \{t_n\}_n$ uma sequência de inteiros não negativos.

Dada uma cobertura $\alpha_1 = \{A_\lambda\}_{\lambda \in L}$, esta possui subcobertura finita. Por $\dim X$ ser finita, tal subcobertura finita possui refinamento com ordem finita, digamos, $k+1 \in \mathbb{N}$ ou menor. Chamemos $\alpha = \{A_1, \ldots, A_\ell\}$ tal refinamento.

Construamos uma partição de X, $\mathcal{P} = \{P_1, \ldots, P_l\}$, fazendo $P_1 = A_1$ e $P_k = A_k \Big\backslash \bigcup\limits_{i=1}^{k-1} A_i$ para $1 < k \leq \ell$.

Daí, para cada $x \in X$, podemos escolher uma vizinhança $V_x \ni x$, intersectando no máximo $k+1$ elementos de $\mathcal{P}$.

De fato, cada $x \in X$ não está em mais que $k+1$ elementos de α. Tome o conjunto formado pela interseção destes elementos de α que contêm x. Este conjunto é aberto, então contém alguma vizinhança de x. Chame tal vizinhança de V_x. V_x não intersecta mais do que $k+1$ elementos de $\mathcal{P}$ pois, se o fizesse, intersectaria mais que $k+1$ elementos de α.

Chamemos de ν uma subcobertura finita de $\{V_x; x \in X\}$. Seja $\varepsilon > 0$ o número de Lebesgue de ν. Dado $y \in X$, chamemos também $V(y)$ um elemento de ν que contém y.

Sejam E_n um conjunto $(\mathcal{S}, n, \varepsilon)$-gerador de cardinalidade mínima.

Sejam também $Y_n \coloneqq \{(i_1, ...i_n);$ existe $x \in X$ com $f^{t_k}(x) \in P_{i_k}, \forall\ 1 \leq k \leq n\}$ e $Z_n \coloneqq \{(i_1, ...i_n);$ existe $y \in E_n$ com $V(f^{t_k}(y)) \cap P_{i_k} \neq \varnothing, \forall\ 1 \leq k \leq n\}$.

Tome $(i_1, \ldots, i_n) \in Y_n$. Então, existe $x \in X$ com $f^{t_k}(x) \in P_{i_k}$, para todo $1 \leq k \leq n$.

Além disso, existe $y \in E_n$ tal que, para todo $1 \leq k \leq n$, tem -se $d(f^{t_k}(x), f^{t_k}(y)) \leq \varepsilon$. Isto significa que $f^{t_k}(x) \in V(f^{t_k}(y))$, para todo $1 \leq k \leq n$.

Logo, $V(f^{t_k}(y)) \cap P_{i_k} \neq \varnothing$, para todo $1 \leq k \leq n$, o que significa que $Y_n \subset Z_n$, implicando em $\#Y_n \leq \#Z_n$.

Para cada $y \in E_n$, fixado $1 \leq k \leq n$, $V(f^{t_k}(y))$ intersecta, no máximo, $k+1$ elementos da partição $\mathcal{P}$. Como isto ocorre para todos os k entre 1 e n, temos que, para cada $y \in E_n$, existem no máximo $(k+1)^n$ elementos de Z_n associados.

Portanto, $\#Z_n \leq (k+1)^n \#E_n = (k+1)^n r_n(\mathcal{S}, f, \varepsilon)$.

Dada $\mu \in \mathcal{M}(X, f)$, $H_\mu(\mathcal{P}^n_{f,\mathcal{S}}) \leq \log \#\mathcal{P}^n_{f,\mathcal{S}}$.

Para cada $x \in P \in \mathcal{P}^n_{f,\mathcal{S}}$, tem-se $f^{t_k}(x) \in P_{i_k}$, para todo $1 \leq k \leq n$. Então $\#\mathcal{P}^n_{f,\mathcal{S}} \leq \#Y_n$.

Assim,

$$H_\mu(\mathcal{P}^n_{f,\mathcal{S}}) \leq \log \#Y_n \leq \log \#Z_n \leq \log((k+1)^n \#E_n) = \log r_n(\mathcal{S}, n, \varepsilon) + n \log(k+1). \quad (4.4)$$

Por consequência, $h_{\mu,\mathcal{S}}(f, \mathcal{P}) \leq r(\mathcal{S}, f, \varepsilon) + \log(k+1) \leq h_{\mathcal{S}}(f) + \log(k+1)$.

Tome $\varepsilon_1 > 0$, e uma partição finita $\mathcal{Q} = \{Q_1, \ldots, Q_\ell\}$ de modo que $h_{\mu,\mathcal{S}}(f, \mathcal{Q}) > h_{\mu,\mathcal{S}}(f) - \varepsilon_1/2$.

Escolhamos $\delta > 0$ tal que, se houver partição $\mathcal{Q}' = \{Q'_1, \ldots, Q'_\ell\}$ com $\mu(Q_i \Delta Q'_i) < \delta$ (Proposição 3.2.9), então $H_\mu(\mathcal{Q}|\mathcal{Q}') < \varepsilon_1/2$.

Como μ é regular (Proposição 2.1.3), podemos obter compactos $K_i \subset Q_i$, $1 \leq i \leq \ell$, com $\mu(Q_i \backslash K_i) < \delta/\ell$.

Defina a cobertura $\beta_1 = \{B_1, \ldots, B_\ell\}$ pondo $B_i = \bigcap\limits_{j \neq i} K_j^c, 1 \leq i \leq \ell$. Seja β um refinamento de β_1 com ordem no máximo $k+1$.

Analogamente ao que foi feito até a equação 4.4, podemos tomar uma partição $\mathcal{R}$ com $\beta \prec \mathcal{R}$ e $h_{\mu,\mathcal{S}}(f, \mathcal{R}) \leq h_{\mathcal{S}}(f) + \log(k+1)$.

Em seguida, defina uma partição $\Gamma = \{M_1, \ldots, M_\ell\}$ pondo

$$\begin{aligned} M_i &= \bigcup \{R \in \mathcal{R}; R \cap K_i \neq \varnothing\}, 1 \leq i \leq \ell - 1; \\ M_\ell &= \bigcap M_i^c. \end{aligned}$$

O fato de $\mathcal{R}$ refinar β garante que nenhum $R \in \mathcal{R}$ intersecta mais do que um $K_i, 1 \leq i \leq \ell$. Isto certifica que Γ seja partição, e que $\Gamma \prec \mathcal{R}$.

Como $K_i \subset M_i$, para todo $1 \leq i \leq \ell$, $\mu(Q_i \backslash M_i) \leq \mu(Q_i \backslash K_i) \leq \delta/\ell$, para todo $1 \leq i \leq \ell$.

Se um dado ponto está em algum M_i, e não está em Q_i, para algum $1 \leq i \leq \ell$, então ele está em algum Q_j e não está em M_j, para algum $1 \leq j \leq \ell, j \neq i$. Assim, $M_i \backslash Q_i \subset \bigcup\limits_{j \neq i} (Q_j \backslash M_j)$.

Por consequência,

$$\mu(M_i \backslash Q_i) \leq \mu\left(\bigcup_{j\neq i}(Q_j\backslash M_j)\right) \leq \sum_{j=1}^{\ell} \mu(Q_j\backslash M_j) - \mu(Q_i\backslash M_i) = \left(1-\frac{1}{\ell}\right)\delta$$

Daí, $\mu(Q_i \Delta M_i) < \delta$, o que implica em $H_\mu(\mathcal{Q}|\Gamma) < \dfrac{\varepsilon_1}{2}$.

Usando o item 4.2.4-2: $h_{\mu,\mathcal{S}}(f,\mathcal{Q}) \leq h_{\mu,\mathcal{S}}(f,\Gamma) + H_\mu(\mathcal{Q}|\Gamma) < h_{\mu,\mathcal{S}}(f,\Gamma) + \varepsilon_1/2$.

Isto implica em

$$\begin{aligned} h_{\mu,\mathcal{S}}(f) &< h_{\mu,\mathcal{S}}(f,\mathcal{Q}) + \frac{\varepsilon_1}{2} \leq h_{\mu,\mathcal{S}}(f,\Gamma) + \varepsilon_1 \leq h_{\mu,\mathcal{S}}(f,\mathcal{R}) + \varepsilon_1 \leq \\ &\leq h_{\mathcal{S}}(f) + \log(k+1) + \varepsilon_1. \end{aligned}$$

Como ε_1 é arbitrário, $h_{\mu,\mathcal{S}}(f) \leq h_{\mathcal{S}}(f) + \log(k+1)$.

Agora, dado $n \in \mathbb{Z}$, sejam $Y = \prod_{i=1}^{n} X_i$ e $g = \prod_{i=1}^{n} f_i$, onde $(X_i, f_i) = (X, f), i = 1, \ldots, n$.

Por um resultado de (29), página 20, tem-se $\dim Y \leq nk$.

Aplicando o obtido acima a (Y, g), obtemos $h_{\mu,\mathcal{S}}(g) \leq h_{\mathcal{S}}(g) + \log(nk+1)$.

Usar as proposições 4.1.14 e 4.2.11 resulta em $nh_{\mu,\mathcal{S}}(f) \leq nh_{\mathcal{S}}(f) + \log(nk+1)$

Dividindo por n e passando ao limite com $n \to \infty$, decorre o resultado. ∎

A seguir, trataremos sobre o Princípio Variacional para entropia sequencial, ou Princípio Variacional Sequencial.

A prova deste resultado se dá pela amálgama de três resultados.

Começamos por introduzir o número auxiliar $K(\mathcal{S})$.

Dada uma sequência de inteiros não negativos $\mathcal{S} = \{t_n\}_n$, vamos definir o número auxiliar $K(\mathcal{S})$ da seguinte maneira:

Para $k, n \in \mathbb{N}$, sejam $U_{\mathcal{S}}(n,k) \coloneqq \bigcup_{i=1}^{n}\{t_i, t_i+1, \ldots, t_i+k-1\}$ e $\Theta_{\mathcal{S}}(n,k) = \#U_{\mathcal{S}}(n,k)$.

Note que, por $\Theta_{\mathcal{S}}(n,k)$ ser uma função crescente positiva em k, $\limsup_n \frac{1}{n}\Theta_{\mathcal{S}}(n,k)$ é uma função crescente não negativa em k.

Definimos $K(\mathcal{S}) \coloneqq \lim_k \limsup_n \dfrac{1}{n}\Theta_{\mathcal{S}}(n,k)$.

Note que, se $\mathcal{S} = \{n-1\}_{n\in\mathbb{N}}$, então, para k fixo e $n > k$, que $\Theta_{\mathcal{S}}(n,k) = n+k-1$. Logo, $\limsup_n \frac{1}{n}\Theta_{\mathcal{S}}(n,k) = 1$ e, assim, $K(\mathcal{S}) = 1$.

Lema 4.3.6. *Seja $\mathcal{S}$ uma sequência de inteiros não negativos. Então,*

$$h_{\mathcal{S}}(f) \leq \begin{cases} 0, & se\ K(\mathcal{S}) = 0; \\ K(\mathcal{S})h(f), & se\ 0 < K(\mathcal{S}) < \infty. \end{cases}$$

Dem.: Seja α uma cobertura e $a_n = \dfrac{1}{n}H(\alpha_f^n)$. Chame $a = \lim\limits_{n\to\infty} a_n = h(f,\alpha)$.

Denote $\varepsilon_n = \sup\limits_{r\geq n}(a_r - a)$. Note que $\{\varepsilon_n\}_{n\in\mathbb{N}}$ é uma sequência decrescente, convergindo para zero.

Sejam $k, n \in \mathbb{N}$, $U_{\mathcal{S}}(n,k)$ e $\Theta_{\mathcal{S}}(n,k)$ definidos como imediatamente antes deste lema.

Defina a relação de equivalência em $U_{\mathcal{S}}(n,k)$ da seguinte maneira: para i, j em $U_{\mathcal{S}}(n,k)$, decretamos $i \sim j$ se, e somente se, ℓ entre i e j ($i \leq \ell \leq j$ ou $j \leq \ell \leq i$) implica $\ell \in U_{\mathcal{S}}(n,k)$.

Denote $U_1, \ldots, U_r$ as classes de equivalência: $U_{\mathcal{S}} = U_1 \dot\cup \ldots \dot\cup\, U_r$.

Seja $b_i = \#U_i, i = 1, \ldots, r$.

Note que $k \leq b_i$, para todo $1 \leq i \leq r$, pois cada classe contém ao menos k elementos consecutivos de $U_{\mathcal{S}}(n,k)$.

Seja $\beta(n,k) = f^{-t_1}(\alpha_f^k) \vee \cdots \vee f^{-t_n}(\alpha_f^k) = \bigvee\limits_{i\in U_{\mathcal{S}}(n,k)} f^{-i}(\alpha)$ uma cobertura.

Então:

$$
\begin{aligned}
H(\beta(n,k)) &= H\left(\bigvee_{i\in U_{\mathcal{S}}(n,k)} f^{-i}(\alpha)\right) \leq \\
&\leq H\left(\bigvee_{i\in U_1} f^{-i}(\alpha)\right) + H\left(\bigvee_{i\in U_2} f^{-i}(\alpha)\right) + \cdots + H\left(\bigvee_{i\in U_r} f^{-i}(\alpha)\right).
\end{aligned}
$$

Os membros de uma classe são inteiros consecutivos entre si. Portanto

$$
H(\beta(n,k)) \leq H(\alpha_f^{b_1}) + H(\alpha_f^{b_2}) + \cdots + H(\alpha_f^{b_r}) = b_1 a_{b_1} + b_2 a_{b_2} + \cdots + b_r a_{b_r}.
$$

Mas, $a_{b_i} \leq a + \varepsilon_k, i = 1, \ldots, r$.

Então, $H(\beta(n,k)) \leq (a+\varepsilon_k)(b_1 + \cdots + b_r) = (a+\varepsilon_k)\Theta_{\mathcal{S}}(n,k)$.

Por consequência,

$$
h_{\mathcal{S}}(f, \alpha_f^k) = \limsup_n \frac{1}{n} H(\beta(n,k)) \leq \limsup_n \frac{1}{n}\Theta_{\mathcal{S}}(n,k)(a+\varepsilon_k) \leq K(\mathcal{S})(a+\varepsilon_k).
$$

Então, para cada $k \in \mathbb{N}$, $h_{\mathcal{S}}(f,\alpha) \leq h_{\mathcal{S}}(f,\alpha_f^k) \leq K(\mathcal{S})(a+\varepsilon_k)$.

Logo, $K(\mathcal{S}) = 0$ implica $h_{\mathcal{S}}(f,\alpha) = 0$. Além disso, $\varepsilon_k \to 0$ quando $k \to \infty$ e, assim, $0 < K(\mathcal{S}) < \infty$ implica $h_{\mathcal{S}}(f,\alpha) \leq K(\mathcal{S})h(f,\alpha) \leq K(\mathcal{S})h(f)$.

Como α é arbitrária, pode-se concluir o resultado tomando o supremo sobre as coberturas. ∎

Proposição 4.3.7. *Seja $\mathcal{S} = \{t_n\}_n$ uma sequência de inteiros não negativos, e $\mu \in \mathcal{M}(X, f)$. Então:*

$$h_{\mu,\mathcal{S}}(f) \geq \begin{cases} K(\mathcal{S})h_\mu(f), & se\ 0 < h_\mu(f) < \infty; \\ \infty, & se\ K(\mathcal{S}) > 0, h_\mu(f) = \infty. \end{cases}$$

Dem.: Seja $\mathcal{P}$ uma partição e $\mathcal{S} = \{t_n\}_n$ uma sequência de inteiros não negativos.

Usando o Lema 3.2.17, dados n, k inteiros positivos, tem-se:

$$H_\mu\left(\bigvee_{i=1}^{n} f^{-t_i}(\mathcal{P}_f^k)\right) = H_\mu\left(\bigvee_{i \in U_\mathcal{S}(n,k)} f^{-i}(\mathcal{P})\right) \geq (\#U_\mathcal{S}(n,k))h_\mu(f,\mathcal{P}) = \Theta(n,k)h_\mu(f,\mathcal{P}).$$

Donde segue que

$$h_{\mu,\mathcal{S}}(f,\mathcal{P}_f^k) = \limsup_n \frac{1}{n} H_\mu\left(\bigvee_{i \in U_\mathcal{S}(n,k)} f^{-i}(\mathcal{P})\right) \geq \limsup_n \frac{1}{n}\Theta(n,k)h_\mu(f,\mathcal{P})$$

Então,

$$h_{\mu,\mathcal{S}}(f) \geq \lim_k h_{\mu,\mathcal{S}}(f,\mathcal{P}_f^k) \geq \lim_k \limsup_n \frac{1}{n}\Theta(n,k)h_\mu(f,\mathcal{P}) = K(\mathcal{S})h_\mu(f,\mathcal{P}).$$

Assim, $h_{\mu,\mathcal{S}}(f) \geq K(\mathcal{S})h_\mu(f)$. ■

Teorema 4.3.8 (Princípio Variacional Sequencial)**.** *Suponha que X possua dimensão de recobrimento finita. Então*

$$h_\mathcal{S}(f) = \begin{cases} 0, & se\ K(\mathcal{S}) = 0; \\ 0, & se\ K(\mathcal{S}) < \infty, h(f) = 0; \\ K(\mathcal{S})h(f), & se\ 0 < h(f) < \infty; \\ \infty, & se\ K(\mathcal{S}) > 0, h(f) = \infty. \end{cases}$$

De modo equivalente, se $h(f) > 0$ ou $K(\mathcal{S}) < \infty$:

$$h_\mathcal{S}(f) = \sup_{\mu \in \mathcal{M}(X,f)} h_{\mu,\mathcal{S}}(f).$$

Dem.: Pelo Teorema 4.3.5, Lema 4.3.6 e Proposição 4.3.7:

$$K(\mathcal{S})h(f) \geq h_\mathcal{S}(f) \geq h_{\mu,\mathcal{S}}(f) \geq K(\mathcal{S})h_\mu(f).$$

Usando o Princípio Variacional ordinário (Teorema 3.3.3) e tomando o supremo sobre as medidas invariantes por f:

$$K(\mathcal{S})h(f) \geq h_\mathcal{S}(f) \geq \sup_{\mu \in \mathcal{M}(X,f)} h_{\mu,\mathcal{S}}(f) \geq \sup_{\mu \in \mathcal{M}(X,f)} K(\mathcal{S})h_\mu(f) = K(\mathcal{S})h(f).$$

O resultado segue de imediato. ■

Teorema 4.3.9. *(27) Se $\mu \in \mathcal{M}(X, f)$, dada $\mathcal{S}$ uma sequência de inteiros não negativos, as entropias métrica e topológica sequenciais de f não mudam se o domínio for tomado como Ω_f, desde que $K(\mathcal{S}) < \infty$ ou $h(f) > 0$.*

Dem.: Devido à Proposição 3.3.6, $h_{\mu,\mathcal{S}}(f) = h_{\mu,\mathcal{S}}\left(f|_{\Omega_f}\right)$.

Pelo Princípio Variacional Sequencial (4.3.8),

$$h_{\mathcal{S}}(f) = \sup_{\mu \in \mathcal{M}(X,f)} h_{\mu,\mathcal{S}}(f) = \sup_{\mu \in \mathcal{M}(X,f)} h_{\mu,\mathcal{S}}\left(f|_{\Omega_f}\right) = h_{\mathcal{S}}\left(f|_{\Omega_f}\right). \qquad \blacksquare$$

4.4 O EXEMPLO DE SZLENK

Nesta seção, mostraremos um sistema dinâmico simbólico que foge às hipóteses do Princípio Variacional Sequencial, isto é, possui o supremo das entropias da medida sequenciais diferente da entropia topológica sequencial. Este exemplo foi elaborado por Szlenk (30). Ver também (31).

O espaço X será um subespaço fechado das sequências formadas por zeros ou uns, $\{0,1\}^{\mathbb{N}}$, munido com a dinâmica *shift* σ, conforme disposto nos exemplos 4.2.10 e 3.1.14.

Seja $n \geq 0$ um inteiro positivo, e r_n, ℓ_n dois inteiros não negativos tais que $r_n = 2^n \ell_n$.

Considere a família de todos os segmentos de sequência $v_n = (v_1, v_2, \ldots, v_{r_n})$, onde $v_k = 0$ se $k \neq i\ell_n$, e v_k é arbitrário, dito *livre* (0 ou 1), para $k = i\ell_n$, $i = 1, \ldots, 2^n$

Por definição, há 2^n dígitos livres em cada v_n. Logo, há 2^{2^n} formas diferentes de se compor v_n segundo esta regra.

Denotemos todos os v_n por $v_n^{(1)}, v_n^{(2)}, \ldots, v_n^{(2^{2^n})}$.

Seja V_n a sequência formada pela concatenação dos v_n: $V_n = v_n^{(1)} v_n^{(2)} \ldots, v_n^{(2^{2^n})}$.

Concatenando, por sua vez, os V_i, definimos $x_0 = V_0 V_1 \ldots V_n \ldots \in \{0,1\}^{\mathbb{N}}$.

Coloquemos $\ell_0 = 1 = r_0$, isto é, $v_0^{(1)} = (0), v_0^{(2)} = (1)$.

Requer-se que ℓ_n seja maior que o comprimento do segmento de sequência $V_0 V_1, \ldots, V_{n-1}$. A ideia é que o intervalo entre dois dígitos livres consecutivos tenda a infinito muito rapidamente.

Como o comprimento de V_i é igual a $2^{2^i} r_i$, ℓ_n deve ser maior que $\sum_{i=0}^{n-1} 2^{2^i} r_i$.

Então, fazemos $\ell_n = n 2^{2^n} r_{n-1} = n 2^{2^n + n - 1} \ell_{n-1}$.

Para fins de ilustração, vejamos os três primeiros V_i: $i = 0, 1, 2$.

V_0, por definição, é dado por 01.

Tem-se $\ell_1 = 1 \cdot 2^{2^1 + 1 - 1} \cdot 1 = 4$, e $r_1 = 2^1 \ell_1 = 8$. Ou seja, cada segmento $v_1^{(i)}$ possui 8 símbolos, sendo dois deles livres, o quarto e o oitavo. O segmento V_1, assim, é composto pela concatenação dos $2^{2^1} = 4$ segmentos $v_1^{(i)}$.

Agora, $\ell_2 = 2 \cdot 2^{2^2+2-1} \cdot 4 = 256$, e $r_2 = 2^2 \ell_2 = 1024$. Isto é, cada segmento $v_2^{(i)}$ possui 1024 símbolos, sendo quatro deles livres: o 256º, o 512º o 768º e o 1024º. O segmento V_2, assim, é composto pela concatenação dos $2^{2^2} = 16$ segmentos $v_2^{(i)}$.

Denote as coordenadas de x_0 por $a_1, a_2, \dots$ e, para todo $n \in \mathbb{N}$, seja p_n inteiro positivo tal que $(a_0, a_1, \dots, a_{p_n}) = V_0 V_1 \dots V_n$.

Isto é, $p_n - p_{n-1}$ é o comprimento de V_n, $n > 0$. Como $p_0 = 2$, temos $p_1 = 34$, $p_2 = 16418$, e assim por diante.

Denote $x_n = \sigma^n(x_0)$, e defina X como o fecho de $\{x_n\}_n$.

Proposição 4.4.1. *O espaço X contém somente pontos da forma:*

1. x_n, $n \in \mathbb{N}$;
2. $e_n = (0, 0, \dots, 0, 1, 0, \dots)$ - *o número 1 aparece na n-ésima posição, somente;*
3. $(0, 0, 0, \dots) = \overline{0}$.

Dem.: É claro que $(0, 0, 0, \dots)$ é ponto de acumulação de $\{\sigma^n(x_0)\}_n$, pois, dado N natural qualquer, existem infinitos termos de $\{\sigma^n(x_0)\}_n$ tal que todos os dígitos até a posição N são 0. Isto se deve ao fato de que $\ell_n \to \infty$ sempre que $n \to \infty$,

Suponha, então, que $a = (a_1, a_2, a_3, \dots)$ é ponto de acumulação de $\{\sigma^n(x_0)\}_n$, com $a_m = 1$, para dado m natural, sem perda de generalidade.

É fato que existem infinitos termos de $\{\sigma^n(x_0)\}_n$ tal que seus m-ésimos dígitos são iguais a 1, por definição de x_0.

Chamemos tais termos de $\{\sigma^{n_k}(x_0)\}_k$.

Existe, ainda, um número natural M de sorte que, se $k > M$, então todos os dígitos, do primeiro ao $m-1$, de $\sigma^{n_k}(x_0)$ são iguais a zero, bem como os dígitos de $m+1$ até $2m$. Isto também se deve ao fato de $\ell_n \to \infty$ quando $n \to \infty$.

Isto indica que a é exatamente $(0, 0, \dots, 0, 1, 0, \dots)$, onde o dígito 1 aparece na m-ésima posição.

Além disso, $e_k \to \overline{0}$ quando $k \to \infty$.

Assim, exaurimos as possibilidades para os elementos de X. ■

Proposição 4.4.2. *A entropia sequencial métrica do sistema (X, σ), $h_{\mu,\mathcal{S}}(\sigma)$, é nula, para qualquer sequência de inteiros não negativos $\mathcal{S}$.*

Dem.: Sejam $\mu \in \mathcal{M}(X, \sigma)$ e $Y = \bigcap_{i=0}^{\infty} \sigma^i(X)$.

Conforme a Proposição 4.4.1, sejam X_1 o conjunto contendo todos os x_n e X_2 o conjunto contendo todos os e_n, $n \in \mathbb{N}$.

Tem-se $Y = X_2 \cup \{\overline{0}\}$. Note que Y é fechado.

Além disso, Y não intersecta X_1, pois, para todo k, $x_k \notin \sigma^{k+1}(X)$.

Como, para todo i, $\sigma^i(X) \subset \sigma^{-1}(\sigma^{i+1}(X))$ e σ é invariante por μ, então

$$\mu(\sigma^{i+1}(X)) = \mu(\sigma^{-1}(\sigma^{i+1}(X)) \geq \mu(\sigma^i(X)).$$

Daí, como $\mu(X) = 1$, $\mu(\sigma^i(X)) = 1$, para todo $i \in \mathbb{N}$.

Pela continuidade da medida 2.1.1, $\mu(Y) = 1$.

Pelo fato de $\sigma^{-i}(e_0) = e_i$, para todo i, segue que $\mu(e_0) = \mu(e_i)$, para todo i.

Cada e_i tem medida nula, pois, caso contrário, X_2 teria medida infinita. Por $\mu(Y) = 1$ resulta $\mu(\overline{0}) = 1$.

Logo, todo subconjunto de X tem medida 0 ou 1.

Assim, dada $\mathcal{S}$ sequência qualquer de inteiros não negativos, tem-se $h_{\mu,\mathcal{S}}(\sigma) = 0$. ■

Proposição 4.4.3. *Existe uma sequência tal que a entropia topológica sequencial do sistema (X, σ), $h_{\mathcal{S}}(\sigma)$, é positiva.*

Dem.: Vamos particionar X de maneira similar do feito em 3.1.14.

Recapitulemos que p_i é a quantidade de dígitos até o fim de V_i, para todo $i = 0, 1, 2, \ldots$.

Seja $A_n = \bigcup_{i=1}^{n} \{p_{i-1} + s\ell_i\}_{s=1}^{2^i}$, para todo $n = 1, 2, \ldots$, e $\mathcal{S} = \{t_i\}_i$ a sequência contendo todos os termos de $\bigcup_{i=1}^{\infty} A_i$, em ordem crescente. Esta, escrita por extenso, seria:

$$\begin{aligned}
&\{p_0 + \ell_1, p_0 + 2\ell_1\} \cup \{p_1 + \ell_2, p_1 + 2\ell_2, p_1 + 3\ell_2, p_1 + 4\ell_2\} \cup \\
&\cup \{p_2 + \ell_3, p_2 + 2\ell_3, \ldots, p_2 + 8\ell_3\} \cup \ldots \\
&\cup \{p_{n-1} + \ell_n, p_{n-1} + 2\ell_n, \ldots, p_{n-1} + 2^n\ell_n\}.
\end{aligned}$$

O conjunto A_{n+1} contém 2^{n+1} termos a mais que o conjunto A_n.

Note que os termos de A_n indicam, dentro de x_0, as posições dos 2^i dígitos livres do primeiro v_i de cada V_i, a partir de V_1 até V_n,

Ainda, A_n possui os $\sum_{i=1}^{n} 2^i = 2^{n+1} - 2$ primeiros termos da sequência $\mathcal{S}$.

Seja $\alpha = \{A_0, A_1\}$ uma cobertura, em que $A_i = \{x \in X; x_0 = i\}, i = 0, 1$.

Em vista da definição de v_n, V_n e ℓ_n, a cobertura $\alpha_n := \bigvee_{i \in A_n} \sigma^{-i}(\alpha)$ contém, ao menos, 2^{2^n} elementos. De fato, a cobertura α_n contém, ao menos, todos os elementos de X que possuam 0 ou 1 nas posições dos índices de A_n. Considerando somente os índices relativos aos dígitos livres de V_n em x_0, α_n contém ao menos, todas as combinações possíveis para v_n, isto é, 2^{2^n} elementos.

Observe, também, que a cobertura α_n é minimal, pois seus elementos são disjuntos dois a dois.

A entropia sequencial de σ com relação a α é $h_{\mathcal{S}}(\sigma,\alpha) = \limsup_n \frac{1}{n} \log N \left(\bigvee_{i=1}^{n} \sigma^{-t_i}(\alpha) \right)$

Chame $y_n = \frac{1}{n} \log N \left(\bigvee_{i=1}^{n} \sigma^{-t_i}(\alpha) \right)$.

Note que $\frac{1}{\#A_n} \log N(\alpha_n) = y_{2^{n+1}-2}$ e que $h_{\mathcal{S}}(f,\alpha) \geq \limsup_n y_{2^{n+1}-2}$.

Por isto,

$$\frac{1}{\#A_n} \log N(\alpha_n) \geq \frac{1}{2^{n+1}-2} \log 2^{2^n} = \frac{2^n}{2^{n+1}-2} \log 2.$$

Consequentemente,

$$h_{\mathcal{S}}(f,\alpha) \geq \limsup_n \frac{1}{\#A_n} \log N(\alpha_n) \geq \frac{1}{2} \log 2.$$

Logo, a entropia topológica sequencial de σ é positiva. $\blacksquare$

Corolário 4.4.4. *A entropia topológica sequencial do sistema (X,σ) em relação à sequência $\mathcal{S}$ não é atingida no conjunto de pontos não errantes de σ.*

De fato, a entropia topológica sequencial do sistema (X,σ) é positiva, segundo a Proposição 4.4.3, mas, quando restrita a Ω_σ, é zero. Para esta última afirmação, note, por Ω_σ ser invariante por σ, que $\Omega_\sigma \subset Y$, em que Y é dado pela Proposição 4.4.2. Além disso, $\overline{0} \in \Omega_\sigma$ e $e_m \to \overline{0}$ quando $m \to +\infty$. Isto quer dizer que Ω_σ contém, além de $\overline{0}$, possivelmente, todos os e_i, com i maior do que algum $j \in \mathbb{N} \cup 0$.

Caso Ω_σ contenha somente $\overline{0}$, segue de imediato que $h_{\mathcal{S}}\left(\sigma|_{\Omega_\sigma}\right) = 0$.

Caso contenha também os e_i, com i maior do que algum $j \in \mathbb{N} \cup 0$, considere uma bola de raio $\varepsilon > 0$, centrada no ponto $\overline{0}$. Fora desta bola, existem finitos pontos do tipo e_i, digamos, N pontos.

Dada uma sequência de inteiros não negativos $\mathcal{S} = \{t_i\}_i$, um conjunto $(\mathcal{S},1,\varepsilon)$-gerador seria formado por $\overline{0}$, os N pontos fora da ε-bola centrada em $\overline{0}$, e mais N pontos dentro da bola, cujas respectivas imagens por f^{t_1} seriam estes mesmos N pontos, totalizando $2N+1$ elementos. Isto é $r_1(\mathcal{S},f,\varepsilon) \leq 2N+1$.

Um conjunto $(\mathcal{S},2,\varepsilon)$-gerador teria que incluir, além dos pontos do conjunto $(\mathcal{S},1,\varepsilon)$-gerador, no máximo, mais N pontos dentro da ε-bola centrada em $\overline{0}$, de sorte que suas imagens por f^{t_2} sejam os N pontos fora desta bola. Assim $r_2(\mathcal{S},f,\varepsilon) \leq 3N+1$.

Continuando o raciocínio, um conjunto $(\mathcal{S},n,\varepsilon)$-gerador minimal de Ω_σ tem cardinalidade, no máximo, igual a $(n+1)N+1$.

Assim, $\limsup_n \frac{1}{n} r(\mathcal{S}, f, \varepsilon) \leq \limsup_n \frac{1}{n} \log((n+1)N+1) = 0$.

Isto é, pela definição de entropia por conjuntos $(\mathcal{S}, n, \varepsilon)$-geradores, temos que $h_{\mathcal{S}}\left(\sigma|_{\Omega_\sigma}\right) = 0$, não importando a sequência $\mathcal{S}$ de inteiros não negativos.

Com isto, pelo Princípio Variacional clássico (Teorema 3.3.3), concluímos que $h(\sigma) = h\left(\sigma|_{\Omega_\sigma}\right) = 0$.

Mas, $h_{\mathcal{S}}(\sigma) > 0$, para $\mathcal{S}$ dada conforme a Proposição 4.4.2.

Este exemplo materializa o fato da entropia topológica sequencial poder distinguir um sistema que possui entropia topológica nula.

5 O TEOREMA DE KUSHNIRENKO

Neste capítulo, apresentaremos alguns resultados propedêuticos, a fim de provar um importante teorema de Kushnirenko (9), o qual caracteriza espectro discreto de um sistema dinâmico com entropia sequencial positiva. Primeiramente, algumas definições:

Definição 5.1. O *operador de Koopman* de uma sistema (f, μ), $U_f : L^2(X, \mu) \to L^2(X, \mu)$, é definido pondo $U_f(\varphi) = \varphi \circ f$.

Note que, se μ está em $\mathcal{M}(X, f)$, o operador de Koopman é uma isometria: dada $\varphi \in L^2(X, \mu)$,

$$\|U_f(\varphi)\|_2 = \sqrt{\left\|\int |U_f(\varphi)|^2 d\mu\right\|} = \sqrt{\left\|\int |\varphi|^2 \circ f d\mu\right\|} = \sqrt{\left\|\int |\varphi|^2 d\mu\right\|} = \|\varphi\|_2.$$

Devido a isto, os autovalores de U_f são sempre unitários.

Dizemos que um coleção de elementos ortonormal de um espaço de Hilbert X, $\{\mathbf{e}_i\}_{i\in\mathbb{N}}$ é *base de Hilbert* se, para todo $x \in X$, ocorra $x = \sum_{i=1}^{\infty} a_i \mathbf{e}_i$.

Definição 5.2. Dizemos que um sistema (f, μ) possui *espectro discreto* se os autovetores do respectivo operador de Koopman U_f geram, no sentido de base de Hilbert, o espaço $L^2(X, \mu)$. Isto é o mesmo que dizer que U_f possui espectro discreto.

Assim, se (f, μ) possui espectro discreto, qualquer função em $\varphi \in L^2(X, \mu)$ pode ser escrita como uma série $\varphi = \sum_{i=1}^{\infty} a_i \mathbf{e}_i$, onde $U_f(\mathbf{e}_i) = \lambda_i \mathbf{e}_i$, $|\lambda_i| = 1$.

Agora, o enunciado do teorema:

Teorema 5.3 (Kushnirenko)**.** *Se $\mu \in \mathcal{M}(X, f)$ e f é invertível, então o sistema (f, μ) possui espectro discreto se, e somente se, $h_{\mu,\mathcal{S}}(f) = 0$ para toda sequência $\mathcal{S}$ de inteiros não negativos.*

Neste contexto, além de X ser espaço compacto, assumimos que μ é não atômica. Começamos pela

Proposição 5.4. *Seja (K, d) um espaço métrico compacto. Dada uma sequência $\{x_n\}_{n\in\mathbb{N}}$ em K e $\varepsilon > 0$, existe $N \in \mathbb{N}$ de sorte que, para cada $x_n, n > N$, pode-se obter $x_i, i \leq N$, tal que $d(x_n, x_i) < \varepsilon$.*

Dem.: Suponha que a afirmação seja falsa.

Então, existe $\varepsilon > 0$ de sorte que, para todo $N \in \mathbb{N}$, existe um $x_n, n > N$, de maneira que $d(x_n, x_i) \geq \varepsilon$, para todo $i \leq N$.

Para $N = 1$, existe $n_1 > N$ tal que $d(x_{n_1}, x_1) \geq \varepsilon$.

Para $N = n_1$, existe $n_2 > N$ tal que $d(x_{n_2}, x_{n_1}), d(x_{n_2}, x_1) \geq \varepsilon$.

Para $N = n_2$, existe $n_3 > N$ tal que $d(x_{n_3}, x_{n_2}), d(x_{n_3}, x_{n_1}), d(x_{n_3}, x_1) \geq \varepsilon$.

Procedendo desta maneira, construímos uma sequência $\{x_1, x_{n_1}, x_{n_2}, \dots\}$ cujos elementos distam ao menos ε de todos os outros. Tal sequência não possui, assim, subsequência convergente, contradizendo a compacidade sequencial de K. ■

Definição 5.5. Chamamos de $\mathcal{W}_\mu$ o conjunto das funções características de subconjuntos de X com medida 1/2. Para cada $T \in \mathcal{W}_\mu$, denotamos por $\xi_T = \{T^{-1}(1), T^{-1}(0)\}$ sua partição correspondente. Chamamos também de $\psi\mathcal{W}_\mu \subset \mathcal{Z}_\mu$ o conjunto das partições correspondentes às funções de $\mathcal{W}_\mu$.

Trataremos $\mathcal{W}_\mu$ com a métrica ζ_μ, induzida pela norma de $L^2(X, \mu)$: dados $T, R \in \mathcal{W}_\mu$, $\zeta_\mu(T, R) \coloneqq \|T - R\|_2$. Note que $|T - R|(x)$ é não nulo se, e somente se, x está ou em $T^{-1}(1)$ ou em $R^{-1}(1)$. Portanto, $\zeta_\mu(T, R) = \sqrt{\mu(T^{-1}(1)\Delta R^{-1}(1))}$.

Proposição 5.6. *A função $\psi : \mathcal{W}_\mu \to \psi\mathcal{W}_\mu$, que leva T em ξ_T, desta forma definida, é contínua.*

Dem.: Sejam T,R funções em $\mathcal{W}_\mu$, e $\xi_T = \{T^{-1}(1), T^{-1}(0)\}, \xi_R = \{R^{-1}(1), R^{-1}(0)\} \in \psi W$ suas partições correspondentes. Chamemos $A_1 = T^{-1}(1), A_2 = T^{-1}(0), A_3 = R^{-1}(1), A_4 = R^{-1}(0)$. A menos de medida nula, $A_1 = A_2^c$ e $A_3 = A_4^c$. Chamemos também $\alpha = \mu(A_1 \cap A_3)$, $\beta = \mu(A_1 \cap A_4)$, $\gamma = \mu(A_2 \cap A_3)$, $\delta = \mu(A_2 \cap A_4)$.

Vamos usar a função ϕ, além de $\mu(A_i) = 1/2$, obtendo

$$H_\mu(\xi_T|\xi_R) = \phi(\alpha) + \phi(\beta) + \phi(\gamma) + \phi(\delta) - \log 2(\alpha + \beta + \gamma + \delta).$$

Ainda, o fato de $H_\mu(\xi_T|\xi_R) = H_\mu(\xi_R|\xi_T)$ resulta em

$$\rho_\mu(\xi_T, \xi_R) = 2(\phi(\alpha) + \phi(\beta) + \phi(\gamma) + \phi(\delta)) - 2\log 2(\alpha + \beta + \gamma + \delta).$$

Ademais, por definição dos A_i:

$$\begin{aligned} \beta + \gamma &= \mu(A_1 \backslash A_3) + \mu(A_3 \backslash A_1) = \mu(A_1 \Delta A_3) = \zeta_\mu^2(T, R) \\ \alpha + \delta &= \mu(A_1 \cap A_3) + \mu(A_1^c \cap A_3^c) = 1 - \mu(A_1 \Delta A_3) = 1 - \zeta_\mu^2(T, R). \end{aligned} \tag{5.1}$$

Isto implica em $\rho_\mu(\xi_T, \xi_R) = 2(\phi(\alpha) + \phi(\beta) + \phi(\gamma) + \phi(\delta)) - 2\log 2$.

Em adição:

$$\begin{aligned} \delta + \gamma = 1/2.\ \beta + \delta = 1/2; \\ \alpha + \gamma = 1/2;\ \alpha + \beta = 1/2; \end{aligned} \tag{5.2}$$

De (4.1) e (4.2), resulta $\alpha = \delta = (1 - \zeta_\mu^2(T, R))/2$ e $\beta = \gamma = \zeta_\mu^2(T, R)/2$.

Por conseguinte, $\rho_\mu(\xi_T, \xi_R)$ se torna igual a

$$\begin{aligned} &2(\alpha \log \alpha + \beta \log \beta + \gamma \log \gamma + \delta \log \delta) - 2 \log 2 = \\ &= -2((1 - \zeta_\mu^2(T,R)) \log(1 - \zeta_\mu^2(T,R)) + \zeta_\mu^2(T,R) \log \zeta_\mu^2(T,R)). \end{aligned} \tag{5.3}$$

Pelo fato de $\lim\limits_{x\to 0}(1 - x^2)\log(1 - x^2) = \lim\limits_{x \to 0} x^2 \log x^2 = 0$, para todo $\varepsilon > 0$, pode-se obter $\delta > 0$ de sorte que, se $\zeta_\mu(T,R) < \delta$, então $\rho_\mu(\xi_T, \xi_R) < \varepsilon$. ■

Observação 5.7. Note que ψ leva exatamente dois elementos do domínio em um elemento da imagem: dado $T \in \mathcal{W}_\mu$, $\psi(T) = \psi(1 - T)$. Por este fato, e por ψ ser contínua, se K é fechado e não compacto, então $\psi(K)$ é não compacto.

De fato, ψ é dois para um, pois cada partição $\xi_T = \{T^{-1}(1), T^{-1}(0)\} \in \psi\mathcal{W}_\mu$ possui duas funções em $\mathcal{W}_\mu$ que são levadas a ela por ψ, a saber, T e $1 - T$.

Ainda, $\mathcal{W}_\mu$ não é discreto, pois, dado $f \in \mathcal{W}_\mu$ e $\varepsilon > 0$, é possível obter $g \in \mathcal{W}_\mu$ com $\zeta_\mu(f, g) = \sqrt{\mu(f^{-1}(1)\Delta g^{-1}(1))} < \varepsilon$, pela continuidade da medida μ.

Se K é fechado, mas não compacto, existe sequência $\{f_n\}_{n\in\mathbb{N}}$ sem subsequência convergente, isto é, existe $\varepsilon > 0$ tal que $\zeta_\mu(f_i, f_j) \geq \varepsilon$, para todos $i \neq j$.

Como ψ é dois para um, a sequência $\{\psi(f_n)\}_{n\in\mathbb{N}}$ possui infinitos termos distintos. Suponha que tal sequência possua subsequência convergente, digamos, para $\xi_0 \in \psi(K)$.

Sejam $g_0, 1 - g_0$ as funções levadas a ξ_0: $\psi(g_0) = \psi(1 - g_0) = \xi_0$.

Usando a continuidade de ψ, escolha $\delta > 0$ tal que a pré imagem de $B_{\rho_\mu}(\xi_0, \delta)$ contenha $B_{\zeta_\mu}(g_0, \varepsilon)$ e $B_{\zeta_\mu}(1 - g_0, \varepsilon)$.

O conjunto $B_{\rho_\mu}(\xi_0, \delta)$ contém infinitos elementos de $\{\psi(f_n)\}_{n\in\mathbb{N}}$. Então, algum dos dois conjuntos $B_{\zeta_\mu}(g_0, \varepsilon)$ ou $B_{\zeta_\mu}(1 - g_0, \varepsilon)$ contém infinitos termos de $\{f_n\}_{n\in\mathbb{N}}$, o que contradiz $\zeta_\mu(f_i, f_j) \geq \varepsilon$, para $i \neq j$.

Lema 5.8. *Seja $\delta > 0$ e sejam $\{a_1, \ldots, a_k\}$, $\{x_1, \ldots, x_k\}$ dois conjuntos de números reais não negativos tais que $\sum_{i=1}^{k} a_i = 1$ e $\sum_{i=1}^{k} a_i x_i < \delta$.*

Então, existe R real tal que, se $M \in \mathbb{N}, M > R$, existem subconjuntos

$$\{a_{\ell_1}, \ldots, a_{\ell_n}\}, \{x_{\ell_1}, \ldots, x_{\ell_n}\}, \text{ com } n \leq k,$$

tais que $\sum_{i=1}^{n} a_{\ell_i} \geq 1 - 1/M$ e $x_{\ell_i} \leq M\delta$, para todo $i = 1, \ldots, n$.

Dem.: Suponha que a assertiva seja falsa.

Note que não é possível que que todos os $a_i, i = 1, \ldots, k$, sejam maiores do que $M\delta$, pois isso implicaria em $\sum_{i=1}^{k} a_i x_i > \delta$.

Assim, existe $\ell < k$ tal que, reordenando os elementos, se necessário, $a_i \leq M\delta$, para $i = 1 \ldots, \ell$, e $a_i > M\delta$, para $i = \ell + 1, \ldots, k$.

Se ocorresse $\sum_{i=1}^{\ell} a_i < 1 - 1/M$, então teríamos $\sum_{i=\ell+1}^{k} a_i > 1/M$. Daí, $\sum_{i=\ell+1}^{k} a_i x_i > M\delta(1/M) = \delta$, o que contradiz a hipótese. ■

Definição 5.9. Dados P, Q dois conjuntos mensuráveis, chamamos de $\mu(P|Q) := \dfrac{\mu(P \cap Q)}{\mu(Q)}$ a *medida condicional de P dado Q*.

Definição 5.10. Definimos a entropia condicional de uma partição $\mathcal{P}$ em relação a um conjunto mensurável Q por $H_\mu(\mathcal{P}|Q) = \sum_{P \in \mathcal{P}} -\mu(P|Q) \log(\mu(P|Q))$.

Proposição 5.11. *Para todo $\varepsilon > 0$ existe $\delta > 0$ de modo que, para qualquer partição $\mathcal{P}$ finita, se ocorrer $H_\mu(\mathcal{Q}|\mathcal{P}) < \delta$ para algum $\mathcal{Q} \in \psi\mathcal{W}_\mu$, então existe $\mathcal{P}' \prec \mathcal{P}$ que $\rho_\mu(\mathcal{Q}, \mathcal{P}') < \varepsilon/2$.*

Dem.: Denotemos por Q_1 e Q_2 os elementos de $\mathcal{Q}$, e por $P_1, \ldots, P_k$ os elementos de $\mathcal{P}$.

Chamemos, usando a função ϕ: $H_\mu(\mathcal{Q}|P_j) := \phi(\mu(Q_1|P_j)) + \phi(\mu(Q_2|P_j))$.

Note que $\mu(Q_1|P_j) + \mu(Q_2|P_j) = 1$. Pondo $\mu(Q_1|P_j) = \gamma$, então $H_\mu(\mathcal{Q}|P_j) = \phi(\gamma) + \phi(1 - \gamma)$ que, como função de γ, não ultrapassa 1. Isto se deve à partição $\mathcal{Q}$ ter dois elementos e à base do logaritmo ser e. Portanto, $H_\mu(\mathcal{Q}|P_j) \leq 1$.

Assim, $H_\mu(\mathcal{Q}|\mathcal{P}) = \sum_{i=1}^{k} \mu(P_j) H_\mu(\mathcal{Q}|P_j) < \delta$, por hipótese - uma média ponderada.

Como $\sum_{i=1}^{k} \mu(P_j) = 1$, faremos uso do lema 5.8: para todo M suficientemente grande, existe um subconjunto $\mathcal{P}_1$ de elementos de $\mathcal{P}$ tal que $\sum_{P_j \in \mathcal{P}_1} \mu(P_j) \geq 1 - 1/M$, e tal que $H_\mu(\mathcal{Q}|P_j) \leq M\delta$ se $P_j \in \mathcal{P}_1$. Basta tomar $a_i = \mu(P_i)$ e $x_i = H_\mu(\mathcal{Q}|P_i)$, $i = 1, \ldots, k$, nos moldes do lema. Atribua a tais conjuntos de $\mathcal{P}_1$ os rótulos $P_1, \ldots, P_m$.

Afirmação: Para θ suficientemente pequeno, a desigualdade $H_\mu(\mathcal{Q}|P_j) \leq M\delta < -\theta \log \theta - (1 - \theta) \log(1 - \theta)$ somente pode implicar em ou $\mu(Q_1|P_j) < \theta$ ou $\mu(Q_2|P_j) < \theta$.

Sem perda de generalidade, assuma que

- para $1 \leq j \leq \ell$ ocorre $\mu(Q_1|P_j) < \theta$ e $\mu(Q_2|P_j) > 1 - \theta$;
- para $\ell < j \leq m$ ocorre $\mu(Q_2|P_j) < \theta$ e $\mu(Q_1|P_j) > 1 - \theta$.

Seja $\mathcal{P}' = \{P_1', P_2', P_3'\}$ uma partição, onde $P_1' = \bigcup_{i=1}^{\ell} P_i$, $P_2' = \bigcup_{i=\ell+1}^{m} P_i$, e $P_3' = \bigcup_{i=m+1}^{k} P_i$.

Então, $\mu(P_3') \leq 1/M$ e, por isso, $\mu(Q_1 \cap P_3') \leq 1/M$, $\mu(Q_2 \cap P_3') \leq 1/M$. Outrossim,

- $\mu(Q_1 \cap P_1') = \sum_{i=1}^{\ell} \mu(Q_1 \cap P_i) < \theta \sum_{i=1}^{\ell} \mu(P_i) = \theta\mu(P_1') \leq \theta$;
- $\mu(Q_2 \cap P_2') = \sum_{i=\ell+1}^{m} \mu(Q_2 \cap P_i) < \theta \sum_{i=\ell+1}^{m} \mu(P_i) = \theta\mu(P_2') \leq \theta$;
- $\mu(Q_2 \cap P_1') = \sum_{i=1}^{\ell} \mu(Q_2 \cap P_i) > (1-\theta) \sum_{i=1}^{\ell} \mu(P_i) = (1-\theta)\mu(P_1')$;
- $\mu(Q_1 \cap P_2') = \sum_{i=\ell+1}^{m} \mu(Q_2 \cap P_i) > (1-\theta) \sum_{i=\ell+1}^{m} \mu(P_i) = (1-\theta)\mu(P_2')$.

Tais fatos implicam, usando o crescimento e decrescimento de ϕ e o fato de θ ser pequeno, em:

- $\phi(\mu(Q_1|P_1')) < \phi(\theta)$;
- $\phi(\mu(Q_2|P_2')) < \phi(\theta)$;
- $\phi(\mu(Q_2|P_1')) < \phi(1-\theta)$;
- $\phi(\mu(Q_1|P_2')) < \phi(1-\theta)$.

Assim, fazendo uso de $H_\mu(\mathcal{Q}|P_j) \leq 1$:

$$
\begin{aligned}
&H_\mu(\mathcal{Q}|\mathcal{P}') = \sum_{i=1}^{3} \mu(P_i) H_\mu(\mathcal{Q}|P_i) < H_\mu(\mathcal{Q}|P_1) + H_\mu(\mathcal{Q}|P_2) + 1/M = \\
&= \phi(Q_1|P_1') + \phi(Q_1|P_2') + \phi(Q_2|P_1') + \phi(Q_2|P_2') + 1/M < \\
&< 2\phi(\theta) + 2\phi(1-\theta) + 1/M = -2(\theta\log\theta + (1-\theta)\log(1-\theta)) + 1/M.
\end{aligned}
$$

Como $\mu(Q_1) = \mu(Q_2) = 1/2$ e $2\mu(Q_1 \cap P_2')$, $2\mu(Q_2 \cap P_1') > 1 - 2\theta$,

$$
\begin{aligned}
&H_\mu(\mathcal{P}'|\mathcal{Q}) = \sum_{i=1}^{2} \mu(Q_i) H_\mu(\mathcal{P}'|Q_i) \leq \sum_{i=1}^{2} H_\mu(\mathcal{P}'|Q_i) = \\
&= \sum_{i=1}^{3} \phi(2\mu(P_i' \cap Q_1)) + \sum_{i=1}^{3} \phi(2\mu(P_i' \cap Q_2)) \leq 2\left(\phi(2\theta) + \phi(1-2\theta) + \phi\left(\frac{2}{M}\right)\right) = \\
&= -2\left(2\theta\log 2\theta + (1-2\theta)\log(1-2\theta) + \frac{2}{M}\log\frac{2}{M}\right).
\end{aligned}
$$

Juntando as partes,

$$
\rho_\mu(\mathcal{Q}, \mathcal{P}') < 2\phi(\theta) + 2\phi(1-\theta) + \frac{1}{M} + 2\left(\phi(2\theta) + \phi(1-2\theta) + \phi\left(\frac{2}{M}\right)\right).
$$

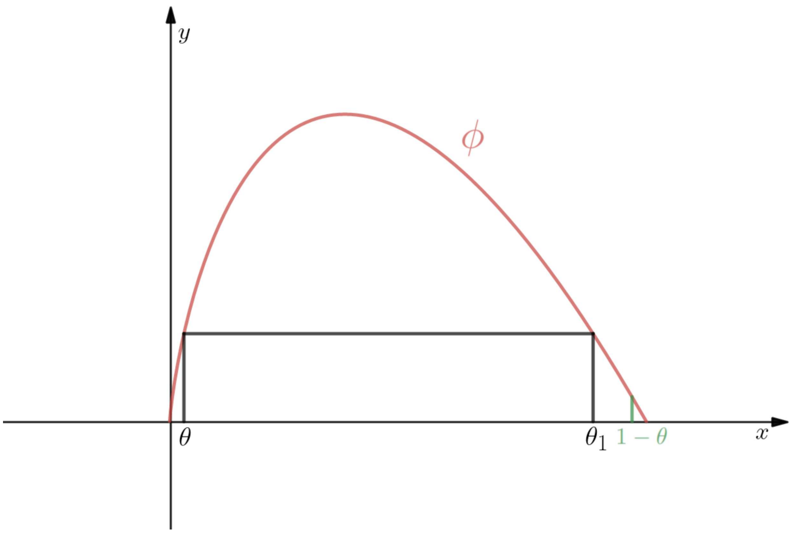
Como $\lim_{\theta\to 0} \phi(2\theta) = \lim_{\theta\to 0} \phi(1-2\theta) = 0$, podemos escolher θ tão pequeno quanto se queira e M tão grande quanto se queira, de modo que $\rho_\mu(\mathcal{Q}, \mathcal{P}') < \varepsilon/2$.

Daí, escolhemos δ a partir de $M\delta < -\theta\log\theta - (1-\theta)\log(1-\theta)$. Como desejado, δ depende somente de ε. $\blacksquare$

Prova da afirmação: Seja θ um número real suficientemente pequeno ($< 1/e$) e seja θ_1 a outra pré imagem da imagem de θ por ϕ. Note que $\theta_1 < 1-\theta$, pois $|\phi'|$ é maior perto de 0 do que perto de 1.

Seja $q = \mu(Q_1|P_j)$. Há quatro opções possíveis. Isto está ilustrado na figura 1:

Figura 1 – Ilustração das quatro opções possíveis para q.



Fonte: Elaborada pelo autor.

1. $0 \leq q < \theta$. Neste caso, $\phi(q) < \phi(\theta)$ e $1-q > 1-\theta$, o que implica $\phi(1-q) < \phi(1-\theta)$. Então, $\phi(q) + \phi(1-q) < \phi(\theta) + \phi(1-\theta)$, o que é admissível;

2. $\theta \leq q \leq \theta_1$. Neste caso, $\phi(q) \geq \phi(\theta)$ e $1-\theta \geq 1-q \geq 1-\theta_1$, o que implica $\phi(1-\theta) \leq \phi(1-q)$. Então, $\phi(q)+\phi(1-q) \geq \phi(\theta)+\phi(1-\theta)$, o que não é admissível;

3. $\theta_1 < q \leq 1-\theta$. Neste caso, $\phi(q) \geq \phi(1-\theta)$ e $1-\theta_1 > 1-q \geq \theta$, o que implica $\phi(1-q) \geq \phi(\theta)$. Então, $\phi(q) + \phi(1-q) \geq \phi(\theta) + \phi(1-\theta)$, o que não é admissível;

4. $q > 1-\theta$. Neste caso, $\phi(q) < \phi(1-\theta)$ e $1-q < \theta$, o que implica $\phi(1-q) < \phi(\theta)$. Então, $\phi(q) + \phi(1-q) < \phi(\theta) + \phi(1-\theta)$, o que é admissível.

Logo, somente é possível ocorrer os casos 1 e 4, ou seja, $q < \theta$ ou $1-q < \theta$. ■

Observação 5.12. Sob as hipóteses da Proposição 5.11, o conjunto dos elementos $\{\mathcal{Q}_i\}_{i\in I} \subset \psi\mathcal{W}_\mu$ tais que ocorre simultaneamente $\rho_\mu(\mathcal{Q}_i, \mathcal{Q}_j) \geq \varepsilon$, se $i \neq j$ e $H_\mu(\mathcal{Q}_i|\mathcal{P}) < \delta$, é finito.

De fato, suponha que não o seja. Como $\mathcal{P}$ é um conjunto finito, o número de partições tais que $\mathcal{P}' \prec \mathcal{P}$ nos moldes da demonstração de 5.11 é finito. Seja $\{\mathcal{P}'_1, \ldots, \mathcal{P}'_k\}$ o conjunto destas partições.

Seja a função $g : \{Q_i\}_{i\in I} \to \{\mathcal{P}'_1, \ldots, \mathcal{P}'_k\}$ que leva cada $\mathcal{Q}_i$ em algum $\mathcal{P}'_j$ de forma que aconteça $\rho_\mu(\mathcal{Q}_i, \mathcal{P}'_j) < \varepsilon/2$.

Note que g é injetiva, pois, se não o for, não poderá acontecer $\rho_\mu(\mathcal{Q}_i, \mathcal{Q}_j) \geq \varepsilon, i \neq j$.

Logo, pelo fato da imagem de g ser finita, segue que $\mathcal{L}$ também o é.

Lema 5.13. *Um conjunto fechado $K \in \psi\mathcal{W}_\mu$ é compacto se, e somente se, para toda sequência $\{\mathcal{P}_n\}_{n\in\mathbb{N}}$ em K, tem-se $\lim_n \frac{1}{n} H_\mu\left(\bigvee_{i=1}^{n} \mathcal{P}_i\right) = 0$.*

Dem.: ($\Rightarrow$) Seja K um compacto. Segundo a Proposição 5.4, para toda sequência $\{\mathcal{P}_n\}_{n\in\mathbb{N}}$ em K e todo $\varepsilon > 0$, existe N natural, de sorte que, para todo $\mathcal{P}_n, n > N$, pode-se obter $\mathcal{P}_i$, $i \leq N$, tal que $\rho_\mu(\mathcal{P}_n, \mathcal{P}_i) < \varepsilon$.

Então, para todo $n > N$, utilizando o Lema $3.2.7 - 2c$:

$$H_\mu\left(\bigvee_{i=1}^{n} \mathcal{P}_i\right) - H_\mu\left(\bigvee_{i=1}^{n-1} \mathcal{P}_i\right) = H_\mu\left(\mathcal{P}_n \middle| \bigvee_{i=1}^{n-1} \mathcal{P}_i\right) \leq H_\mu(\mathcal{P}_n|\mathcal{P}_i) \leq \rho_\mu(\mathcal{P}_n, \mathcal{P}_i) < \varepsilon.$$

Então, se $m, n > N$, temos $H_\mu\left(\bigvee_{i=1}^{m+n} \mathcal{P}_i\right) - H_\mu\left(\bigvee_{i=1}^{n} \mathcal{P}_i\right) < m\varepsilon$, o que implica $\limsup_m \frac{1}{m} H_\mu\left(\bigvee_{i=1}^{m} \mathcal{P}_i\right) \leq \varepsilon$. Como ε é arbitrário, tem-se $\limsup_m \frac{1}{m} H_\mu\left(\bigvee_{i=1}^{m} \mathcal{P}_i\right) = 0$.

($\Leftarrow$) Se K é fechado, mas não compacto, podemos obter $\varepsilon > 0$ e uma sequência $\{\mathcal{P}_i\}_{i\in\mathbb{N}}$ de pontos de K, de sorte que $\rho_\mu(\mathcal{P}_i, \mathcal{P}_j) \geq \varepsilon$, se $i \neq j$.

Dado $\delta > 0$ e $\mathcal{P}$ uma partição finita, o número de termos de $\{\mathcal{P}_i\}_{i\in\mathbb{N}}$ que satisfazem $H_\mu(\mathcal{P}_i, \mathcal{P}) < \delta$ é finito, por causa da observação 5.12.

Logo, para infinitos índices i, dada $\mathcal{P}$ partição finita, temos $H_\mu(\mathcal{P}_i|\mathcal{P}) \geq \delta$.

Agora, formemos uma subsequência $\mathcal{P}_{n_1}, \mathcal{P}_{n_2}, \ldots, \mathcal{P}_{n_k}, \ldots$ indutivamente.

Escolha $\mathcal{P}_{n_1}$ e assuma que os $k-1$ primeiros elementos da subsequência foram determinados, satisfazendo $H_\mu\left(\mathcal{P}_{n_j} \middle| \bigvee_{i=1}^{j-1} \mathcal{P}_{n_i}\right) \geq \delta$, para todo $1 \leq j \leq k-1$.

Fazendo $\mathcal{P} = \bigvee_{i=1}^{k-1} \mathcal{P}_{n_i}$, existe um dentre os infinitos termos da sequência $\{\mathcal{P}_i\}_{i\in\mathbb{N}}$ satisfazendo $H_\mu(\mathcal{P}_i|\mathcal{P}) \geq \delta$. Chame-o de $\mathcal{P}_{n_k}$.

Geramos, assim, sequência $\{\mathcal{P}_{n_k}\}_{k\in\mathbb{N}}$, de forma que $H_\mu\left(\mathcal{P}_{n_k} \middle| \bigvee_{i=1}^{k-1} \mathcal{P}_{n_i}\right) \geq \delta$.

Daí, de modo análogo ao feito na parte 'somente se', $H_\mu\left(\bigvee_{i=1}^{m+n} \mathcal{P}_{n_i}\right) - H_\mu\left(\bigvee_{i=1}^{n} \mathcal{P}_{n_i}\right) \geq m\delta$, o que implica $\limsup_k \frac{1}{k} H_\mu\left(\bigvee_{i=1}^{k} \mathcal{P}_{n_k}\right) \geq \delta > 0$. ■

O seguinte lema é parte de um teorema de Paul Halmos. Sua demonstração utiliza o Teorema Espectral, e não será exibida neste texto. O leitor interessado poderá consultar (19), p. 222-224, ou (32), p. 39-41.

Lema 5.14. *Se a função $\varphi \in L^2(X,\mu)$ é ortogonal a todos os autovetores do operador unitário U_f, então $\frac{1}{n}\sum_{i=0}^{n-1}|\langle U_f^i(\varphi),\varphi\rangle|^2 \to 0$ se $n \to \infty$.*

Lema 5.15. *O operador unitário $U_f : L^2(X,\mu) \to L^2(X,\mu)$ possui espectro discreto se, e somente se, o fecho do conjunto $\{U_f^n(\varphi)\}_n$ é compacto, para toda $\varphi \in L^2(X,\mu)$.*

Dem.: ($\Rightarrow$) Se (f,μ) possui espectro discreto, então existe base ortonormal $\{\mathbf{e}_k\}_{k\in\mathbb{N}}$ de $L^2(X,\mu)$, tal que $U_f(\mathbf{e}_k) = \lambda_k\mathbf{e}_k, |\lambda_k| = 1$, $k = 1,2,\ldots$.

Toda $\varphi \in L^2(X,\mu)$ pode ser escrita como $\varphi = \sum_{i=1}^{\infty} a_i\mathbf{e}_i$, com $\sum_{i=1}^{\infty}|a_i|^2 < \infty$.

Para todo $n \in \mathbb{N}$, temos que $U_f^n(\varphi) = \sum_{i=1}^{\infty} b_i\mathbf{e}_i$, onde $b_i = \lambda_i^n a_i$. É claro que $|b_i| \le |a_i|$, para todo i.

Assim, o fecho de $\{U_f^n(\varphi)\}_{n\in\mathbb{N}}$ é compacto, pois está contido no conjunto

$$\left\{g = \sum_{i=1}^{\infty} c_i\mathbf{e}_i; |c_i| \le |a_i|\right\}.$$

Este conjunto é compacto, por ser sequencialmente compacto. A prova deste fato segue as ideias do mesmo argumento para um cubo de Hilbert (33).

($\Leftarrow$): Suponha que (f,μ) não possua espectro discreto. Então, existe um vetor no complemento ortogonal do subespaço gerado pelas autofunções de U_f, isto é, existe $\varphi \in L^2(X,\mu)$ ortogonal a todas às autofunções de U_f.

Pelo Lema 5.14, $\frac{1}{n}\lim_n \sum_{i=0}^{n-1}|\langle U_f^i(\varphi),\varphi\rangle|^2 = 0$.

Isto equivale a dizer que $\lim_n |\langle U_f^n(\varphi),\varphi\rangle| = 0$, se n não toma valores em um conjunto J de densidade zero, pelo Lema 2.1.11.

Por f ser invertível, o operador U_f é unitário. Então, $\langle U_f(\varphi), U_f(\psi)\rangle = \langle\varphi,\psi\rangle$, para todas $\varphi,\psi \in L^2(X,\mu)$. Disto segue que, para quaisquer $a, b \in \mathbb{N}$ com $a \ge b$, tem-se $\langle U_f^a(\varphi), U_f^b(\varphi)\rangle = \langle U_f^{a-b}(\varphi),\varphi\rangle$.

Seja $\varepsilon > 0$ dado. Pode-se obter $n_1 \notin J_1$ com densidade zero, de sorte que $|\langle U_f^{n_1}(\varphi),\varphi\rangle| < \varepsilon$.

De forma semelhante, pode-se obter $n_2 > n_1$ de sorte que $n_2 \notin J_2$ com densidade zero, e $|\langle U_f^{n_2-n_1}(\varphi),\varphi\rangle| = |\langle U_f^{n_2}(\varphi), U_f^{n_1}(\varphi)\rangle| < \varepsilon$ e $|\langle U_f^{n_2}(\varphi),\varphi\rangle| < \varepsilon$.

Ainda, pode-se obter $n_3 > n_2$ de forma que $n_3 \notin J_3$ com densidade zero, $|\langle U_f^{n_3-n_2}(\varphi),\varphi\rangle| = |\langle U_f^{n_3}(\varphi), U_f^{n_2}(\varphi)\rangle| < \varepsilon$, e $|\langle U_f^{n_3-n_2}(\varphi),\varphi\rangle| = |\langle U_f^{n_3}(\varphi), U_f^{n_2}(\varphi)\rangle| < \varepsilon$.

Continuando o raciocínio indutivo, obtemos n_k de forma que $n_k > n_{k-1}$, $n_k - n_i \notin J_k$ de densidade zero, $i = 1,\ldots,k-1$, ocorrendo $|\langle U_f^{n_k}(\varphi), U_f^{n_j}(\varphi)\rangle| < \varepsilon, i = 1,\ldots,k-1$ e $|\langle U_f^{n_k}(\varphi)\cdot\varphi\rangle| < \varepsilon$.

O fato da união finita de conjuntos de densidade zero ser, também, um conjunto de densidade zero, garante que todos os índices $n_1, \ldots, n_k$ e $n_i - n_j, 1 \leq i < j \leq k$, estejam fora de um conjunto de densidade zero, a saber, $J_1 \cup \cdots \cup J_k$.

Assim, para $i \neq j$, tem-se

$$\zeta_\mu(U_f^{n_i}(\varphi), U_f^{n_j}(\varphi)) = \|U_f^{n_i}(\varphi) - U_f^{n_j}(\varphi)\|_2 =$$
$$= \sqrt{\|U_f^{n_i}(\varphi)\|_2^2 + \|U_f^{n_j}(\varphi)\|_2^2 - \langle U_f^{n_i}(\varphi), U_f^{n_j}(\varphi)\rangle - \langle U_f^{n_j}(\varphi), U_f^{n_i}(\varphi)\rangle} > \sqrt{2\|\varphi\| - 2\varepsilon} > 0,$$

ou seja, a sequência $\{U_f^{n_i}(\varphi)\}_{i\in\mathbb{N}}$ não possui subsequência convergente, implicando que o conjunto $\{U_f^n(\varphi)\}_{n\in\mathbb{N}}$ não é compacto. ■

No próximo lema, faremos uso da hipótese que μ é não atômica, juntamente com o resultado de Sierpinski (21):

Teorema 5.16. *Se μ é medida não atômica em X, então, para todo conjunto A mensurável e $a \in [0, \mu(A)]$, existe $B \subset A$ tal que $\mu(B) = a$.*

Lema 5.17. *Seja $\mathcal{P} \in \mathcal{Z}_\mu$ uma partição finita. Dado $\varepsilon > 0$, existe $k > 0$ e $\xi_1, \ldots, \xi_k \in \psi\mathcal{W}_\mu$ de sorte que $\rho_\mu(\mathcal{P}, \xi_1, \ldots, \xi_k) < \varepsilon$.*

Dem.: Para provar a afirmação, basta provar que existe um conjunto infinito $\{\mathcal{P}_1, \ldots, \mathcal{P}_n, \ldots\}$, onde $\mathcal{P}_i = \bigvee_{j=1}^{i} \xi_j$, $\xi_j \in \psi\mathcal{W}_\mu$, de sorte que $\bigvee_{i=1}^{\infty} \mathcal{P}_i = \bigvee_{i=1}^{\infty} \xi_i = \epsilon$, usando a Proposição 4.2.6.

Suponha que isto não aconteça: para cada coleção de partições $\{\xi_1, \ldots, \xi_n, \ldots\}$, existe um conjunto K de medida positiva, contido em algum elemento de $\bigvee_{i=1}^{\infty} \xi_i$.

Tome duas coleções de partições $\{\eta_1, \ldots, \eta_n, \ldots\}$ e $\{\xi_1, \ldots, \xi_n, \ldots\}$, sendo K_1 o conjunto com medida positiva contido em algum elemento de $\bigvee_{i=1}^{\infty} \eta_i$ e K_2 o conjunto com medida positiva contido em algum elemento de $\bigvee_{i=1}^{\infty} \xi_i$.

Se $K_1 \cap K_2 = \varnothing$, então a coleção $\{\eta_1, \ldots, \eta_n, \ldots, \xi_1, \ldots, \xi_n, \ldots\}$ é tal que $\bigvee_{i=1}^{\infty} \xi_i \vee \bigvee_{i=1}^{\infty} \eta_i = \epsilon$.

Então, sem perda de generalidade, podemos supor que existe conjunto K com medida positiva (digamos, $\mu(K) = a < 1$) tal que, para toda coleção $\{\xi_1, \ldots, \xi_n, \ldots\}$, K está contido em algum elemento de $\bigvee_{i=1}^{\infty} \xi_i$.

Seja $\alpha = \{A, B\}$ uma partição formada da seguinte maneira: tome P subconjunto de K com $0 < \mu(P) = b < a$. Isto é possível, pelo Teorema 5.16. Tome, também, Q conjunto disjunto de K, com $\mu(Q) = 1/2 - b$. Faça $A = P \cup Q$ e $B = X \backslash A$.

Observe $\alpha \vee \bigvee_{i=1}^{\infty} \xi_i$ contradiz a hipótese, e concluímos a prova. ■

Proposição 5.18. *Dada uma sequência de inteiros não negativos* $\mathcal{S} = \{t_n\}_n$*, ocorre* $h_{\mu,\mathcal{S}}(f) = 0$ *se, e somente se,* $h_{\mu,\mathcal{S}}(f, \mathcal{P}) = 0$ *para toda* $\mathcal{P} \in \psi\mathcal{W}_\mu$*.*

Dem.: ($\Leftarrow$): Suponha $h_{\mathcal{S}}(f, \xi) = 0$ para todo $\xi \in \psi\mathcal{W}_\mu$.

Seja $\mathcal{P}$ uma partição. Pelo Lema 5.17, para todo $\varepsilon > 0$, existem $k \in \mathbb{N}$ e partições $\xi_1, \ldots, \xi_k$ em $\psi\mathcal{W}_\mu$ tais que $\rho_\mu(\mathcal{P}, \xi_1 \vee \cdots \vee \xi_k) < \varepsilon$.

Note que $h_{\mu,\mathcal{S}}(f, \xi_1 \vee \cdots \vee \xi_k) \leq h_{\mu,\mathcal{S}}(f, \xi_1) + \cdots + h_{\mu,\mathcal{S}}(f, \xi_k) = 0$.

Então, pelo Lema 4.2.7, $h_{\mu,\mathcal{S}}(f, \mathcal{P}) < \varepsilon$.

Logo, $h_{\mu,\mathcal{S}}(f, \mathcal{P}) = 0$, para toda partição $\mathcal{P}$, o que implica $h_{\mu,\mathcal{S}}(f) = 0$

($\Rightarrow$): Como $h_{\mu,\mathcal{S}} = 0$, então $h_{\mu,\mathcal{S}}(f, \xi) = 0$ para toda partição ξ, em particular, para $\xi \in \psi\mathcal{W}_\mu$. ■

Vamos, agora, à prova do teorema:

($\Rightarrow$) Suponha que (f, μ) possua espectro discreto.

Pelo Lema 5.15, para toda $\varphi \in \mathcal{W}_\mu$, o fecho do conjunto $\{U_f^n(\varphi)\}_{n\in\mathbb{N}} = \{\varphi \circ f^n\}_{n\in\mathbb{N}}$ é compacto.

Como a função ψ é contínua (Observação 5.7), o fecho de $\{\xi_{\varphi\circ f^n}\}_{n\in\mathbb{N}} = \{f^{-n}(\xi_\varphi)\}_{n\in\mathbb{N}}$ é compacto.

Repare que $h_{\mu,\mathcal{S}}(f, \xi_\varphi) = \limsup\limits_n \dfrac{1}{n} H_\mu\left(\bigvee\limits_{i=1}^n f^{-t_n}(\xi_\varphi)\right)$.

Pelo Lema 5.13, $h_{\mu,\mathcal{S}}(f, \xi_\varphi) = 0$, para toda $\varphi \in \mathcal{W}_\mu$.

A Proposição 5.18 garante que $h_{\mu,\mathcal{S}}(f) = 0$.

($\Leftarrow$) Suponha que (f, μ) não possua espectro discreto.

Novamente, pelo Lema 5.13, para toda $\varphi \in \mathcal{W}_\mu$, o fecho de $\{\varphi \circ f^n\}_{n\in\mathbb{N}}$ não é compacto em $\psi\mathcal{W}_\mu$.

Como ψ é contínua (Observação 5.7), o fecho de $\{f^{-n}(\xi_\varphi)\}_{n\in\mathbb{N}}$ não é compacto em $\psi\mathcal{W}_\mu$.

Logo, o Lema 5.13 permite afirmar que existe subsequência $\{f^{-n_k}(\xi_\varphi)\}_{k\in\mathbb{N}}$ tal que $\limsup\limits_k \dfrac{1}{k} H_\mu\left(\bigvee\limits_{i=1}^k f^{-n_k}(\xi_\varphi)\right) > 0$, isto é, $h_{\mu,\mathcal{S}}(f, \xi_\varphi) > 0$, para $\mathcal{S} = \{n_k\}_{k\in\mathbb{N}}$.

Assim, a Proposição 5.18 certifica $h_{\mu,\mathcal{S}}(f) > 0$, e concluímos a prova. ■

Exemplo 5.19. (19) Considere o toro $\mathbb{T}^n = \mathbb{R}^n/\mathbb{Z}^n$.

Podemos munir este espaço com uma medida de probabilidade, da seguinte maneira: seja $\pi : \mathbb{R}^n \to \mathbb{T}^n$ a projeção que associa cada ponto $x \in \mathbb{R}^n$ à sua classe de equivalência $[x]$ em $\mathbb{T}^n$.

Diremos que um conjunto $A \in \mathbb{T}^n$ é mensurável se $\pi^{-1}(A)$ é um conjunto mensurável em $\mathbb{R}^n$. Definimos a medida de Lebesgue ν no toro fazendo $\nu(A) = \mu(\pi^{-1}(A) \cap [0,1)^n)$, em que μ é a medida de Lebesgue em $\mathbb{R}^n$.

O espaço $L^2(\nu)$ das funções quadrado integráveis de $\mathbb{T}^n$ em $\mathbb{R}$ tem como base de Hilbert a família de funções de Fourier $\mathcal{F} = \{\phi_k(x) = e^{2\pi i k\cdot x}; k \in \mathbb{Z}^n\}$,

Seja $R_\theta : \mathbb{T}^n \to \mathbb{T}^n$ a rotação por um vetor $\theta = (\theta_1, \ldots, \theta_n)$: dado $x = (x_1, \ldots, x_n)$, tem-se $R_\theta(x) = (x_1 + \theta_1, \ldots, x_n + \theta_n) \bmod 1$.

Então,

$$U_{R_\theta}(\phi_k)(x) = \phi_k \circ R_\theta(x) = \phi_k(x+\theta) = e^{2\pi i k\cdot\theta}\phi_k(x),$$

o que significa que os elementos de $\mathcal{F}$ são autovetores de U_{R_θ} e, portanto, o sistema (R_θ, ν) possui espectro discreto.

Pelo Teorema 5.3, dada $\mathcal{S}$ uma sequência de inteiros não negativos, $h_{\mu,\mathcal{S}}(R_\theta) = 0$.

6 CONCLUSÕES

Este trabalho realizou uma exposição de uma das várias maneiras que é possível estender o conceito de entropia, a saber, a entropia sequencial, de modo a melhor caracterizar um sistema dinâmico.

A entropia sequencial oportuniza fortes possibilidades. O teorema de Kushnirenko diz respeito a um sistema com entropia sequencial nula; o teorema abaixo, por J. Smítal e N. Franzová (34), traz o caso positivo para aplicações em intervalos da reta: uma transformação $f : I \subset \mathbb{R} \to I$ é caótica no sentido Li-Yorke se, e somente se, existe sequência de inteiros não negativos $\mathcal{S}$ de sorte que $h_{\mathcal{S}}(f) > 0$.

Além da entropia sequencial, há outras generalizações, como a entropia lenta (*slow entropy*), entropia média de Fried (*Fried average entropy*) (27) e entropia polinomial (*polynomial entropy*) (35).

REFERÊNCIAS

1 ADLER, R. L; KONHEIM, A. G.; McANDREW, M. H.. ***Topological Entropy***. *Transactions of the American Mathematical Society*, v. 114, no. 2, p. 309-319, 1965.

2 DINABURG, E.. ***A Correlation Bewteen Topological Entropy and Metric Entropy***. *Proceedings of the Russian Academy of Sciences*, p. 19-22, 1970.

3 BOWEN, R.. ***Entropy for Group Endomorphisms and Homogeneous Spaces***. *Transactions of the American Mathematical Society*, v. 153, p. 401-414, 1971.

4 KOLMOGOROV. A.. ***A New Metric Invariant of Transient Dynamical Systems and Automorphisms in Lebesgue Spaces***. *Proceedings of the Russian Academy of Sciences*, v. 119, p. 861-864, 1958.

5 SINAI, Y.. ***On The Notion of Entropy of a Dynamical System***. *Proceedings of the Russian Academy of Sciences*, v. 124, p. 768-771, 1959.

6 SHANNON, C. E.. ***A Mathematical Theory of Communication***. *The Bell System Communication Journal*, vol 27. p. 379-423, 1948.

7 ROKHLIN, V.. ***Lectures on The Entropy Theory of Measure Preserving Transformations***. *Russian Mathematical Surveys*, v. 22, p. 1-52, 1967.

8 GOODMAN, T. N. T.. ***Relating Topological Entropy and Measure Theory***. *Bulletin of The London Mathematical Society*, v. 3, p. 176-180, 1971.

9 KUSHNIRENKO, A. G.. ***On Metric Invariants Of Entropy Type***. *Russian Mathematical Surveys*, v. 22, p. 53-61, 1967.

10 GOODMAN, T. N. T.. ***Topological Sequence Entropy***. *Proceedings of the London Mathematical Society*, v. 29, p. 331-350, 1974.

11 NEWTON, D.; KRUG, E.. ***On Sequence Entropy of Automorphisms of a Lebesgue Space***. *Probability Theory and Related Fields*, v. 24, p. 211-214, 1974.

12 CANOVAS, J. S.; BALIBREA, F; LÓPEZ, V. J.. ***Some Results on Entropy and Sequence Entropy***. *International Journal of Bifurcation and Chaos*, v. 9, no. 9, p. 1731–1742, 1999.

13 CANOVAS, J. S.. ***A Guide to Topological Sequence Entropy***. Janeiro de 2008. Disponível em: `https://www.researchgate.net/publication/286636259_A_guide_to_topological_sequence_entropy`. Acesso em 25/02/2021.

14 HALMOS. P. R.. ***Measure Theory***. 1ª Edição. Nova Iorque, *Springer*, 1974.

15 KATOK, A.; HASSELBLAT, B.. ***Introduction to The Modern Theory of Dyamical Systems***. 1ª Edição. Cambridge: *Cambridge University Press*, 1995.

16 MERRY, W. J.. ***Non-Wandering Sets and Its Friends***. Notas de Aula. Disponível em `https://www.merry.io/dynamical-systems/3-the-non-wandering-set-and-its-friends/`. Acesso em 27/02/2021.

17 WALTERS, P.. ***Introduction to Ergodic Theory***. 1ª Edição. Nova Iorque: *Springer*, 1982.

18 POMBO, D. P.. **Introdução à Análise Funcional**. 2ª Edição. Rio de Janeiro: UFF, 2010.

19 OLIVEIRA, K.; VIANA, M.. **Fundamentos da Teoria Ergódica**. 2ª Edição. Rio de Janeiro: SBM, 2019.

20 ZITKOVIC, G.. ***Theory of Probability I*** Notas de Aula. Disponível em: `https://web.ma.utexas.edu/users/gordanz/notes/theory_of_probability_I.pdf`. Acesso em 19/12/2020.

21 DUDLEY, R. M.; NORVAIŠA, R. ***Concrete Functional Calculus***. 1ª Edição. Nova Iorque: *Springer Monographs on Mathematics*, 2011.

22 KATOK, A.. ***Fifty Years of Entropy in Dynamics: 1958-2007***. *Journal of Modern Dynamics*, v. 1, no. 4, p. 545-596, 2007.

23 COLLI, E.. **Teoria Ergódica**. Notas de Aula. Disponível em: `https://www.ime.usp.br/~colli/Ergodica/Ergodica.pdf`. Acesso em 25/02/2021.

24 IMPA. **Programa de Doutorado: Teoria Ergódica Diferenciável**. Disponível em: `https://www.youtube.com/watch?v=UahWFdLS1SE&list=PLo4jXE-LdDTSBGqH-EQ0oqzffTq8uE6SN`. Acesso em 04/12/2020.

25 MUNKRES, J. R.. ***Topology: Pearson New International Edition***. 2ª Edição. Harlow: *Pearson Associated Limited*, 2014.

26 VARÃO, R.. **Dinâmica Hiperbólica e Teoria Ergódica**. Notas de Aula. Disponível em: `http://www.ime.unicamp.br/~regisvarao/out/din.hip.pdf`. Acesso em 25/02/2021.

27 KATOK, A.; KANIGOWSKI, A.; WEI, D.. ***Survey on Entropy-type Invariants of Sub-exponential Growth in Dynamical Systems***. Abril de 2020. Disponível em: `https://arxiv.org/abs/2004.04655v1`.

28 LEMAŃCZYK, M.. ***The Sequence Entropy for Morse Shifts and Other Counterexamples***. *Studia Mathematica*, v. LXXXIL, p. 221-241, 1985.

29 NAGATA, J.. ***Modern Dimension Theory***. 1ª Edição. *Amsterdam: North-Holland Publishing Company*, 1965.

30 SZLENK, W.. ***On Weakly* Conditionally Compact Dynamical Systems***. *Studia Mathematica*, v. LXVI, p. 25-32, 1979.

31 HYB, W. ***Remark on the supremum theorem for the sequence entropy***. *Annales Societatis Mathematicae Polonae, Series I: Commentationes Mathematicae*, v. XXIII, p. 215-218, 1983.

32 HALMOS. P. R.. ***Lectures on Ergodic Theory***. 1ª Edição. Nova Iorque, *Chelsea Publishing Company*, 1956.

33 STEEN, A. L.; SEEBACH, J. A.. ***Counterexamples in Topology***. *Dover Publications*, 1995.

34 SMÍTAL, J; FRANZOVÁ, N. ***Positive Sequence Topological Entropy Characterizes Chaotic Maps***. *Proceedings of the American Mathematical Society*, v. 112, No 4, p. 1083-1086, 1991.

35 MARCO, J.. ***Polynomial Entropies and Integrable Hamiltonian Systems***. *Regular and Chaotic Dynamics*, v. 18, p. 623-655, 2013.