



Universidade Estadual de Maringá  
Centro de Ciências Exatas  
Departamento de Física

Trabalho de Conclusão de Curso

**Análise de séries temporais via curvas  
complexidade-entropia em múltiplas escalas**

Alvaro Franco Martins

Maringá, 16 de outubro de 2018.



Universidade Estadual de Maringá  
Centro de Ciências Exatas  
Departamento de Física

Trabalho de Conclusão de Curso

## **Análise de séries temporais via curvas complexidade-entropia em múltiplas escalas**

Trabalho de conclusão de curso apresentado ao Departamento de Física da Universidade Estadual de Maringá como parte dos requisitos para obtenção do título de bacharel em Física.

Orientador: Prof. Dr. Haroldo Valentin Ribeiro

Alvaro Franco Martins

Maringá, 16 de outubro de 2018.

# Sumário

<b>Resumo</b>	<b>1</b>
<b>Abstract</b>	<b>2</b>
<b>Introdução</b>	<b>3</b>
<b>1 Entropia de Shannon e suas generalizações</b>	<b>5</b>
1.1 Algumas formas de entropia . . . . .	5
1.2 Entropia de permutação: abordagem de Bandt e Pompe . . . . .	7
<b>2 Medidas de Complexidade</b>	<b>12</b>
2.1 A complexidade estatística . . . . .	12
2.2 Plano complexidade-entropia . . . . .	14
2.3 Generalização do plano complexidade-entropia para entropias monopara- métricas . . . . .	16
2.3.1 Curvas $q$ -complexidade-entropia de Tsallis . . . . .	17
2.3.2 Curvas $\alpha$ -complexidade-entropia de Rényi . . . . .	19
<b>3 Análise das curvas complexidade-entropia via <i>embedding delay</i></b>	<b>23</b>
3.1 Família de curvas complexidade-entropia . . . . .	23
3.2 Sinal periódico com ruído gaussiano . . . . .	24
3.3 Ruído harmônico . . . . .	27
3.4 Séries caóticas . . . . .	30
<b>Conclusões</b>	<b>34</b>
<b>Apêndice A - Séries Temporais</b>	<b>35</b>
A.1 Séries temporais estocásticas . . . . .	36
A.1.1 Sinal periódico com ruído . . . . .	36
A.1.2 Ruído harmônico . . . . .	36
A.2 Séries temporais caóticas . . . . .	36
A.2.1 Mapa logístico . . . . .	37
A.2.2 Mapa de Burgers . . . . .	37

A.2.3 Mapa cúbico . . . . .	37
A.2.4 Mapa Gingerbreadman . . . . .	37
A.2.5 Mapa de Hénon . . . . .	37
A.2.6 Mapa de Ricker . . . . .	38
A.2.7 Mapa seno . . . . .	38
A.2.8 Mapa Tinkerbell . . . . .	38
<b>Referências bibliográficas</b>	<b>38</b>

# Resumo

A ideia desse trabalho é empregar um novo parâmetro para análise de séries temporais regulares, caóticas e estocásticas dentro do contexto das chamadas curvas de complexidade-entropia. Esse novo parâmetro está associado às diferentes escalas temporais de uma série e produz uma família de curvas que pode ser útil no estudo e caracterização de séries temporais. Introduzimos o uso desse novo parâmetro para as curvas complexidade-entropia no contexto das entropias monoparamétricas de Tsallis e de Rényi, as quais fazem o papel de realçar as diferentes características entrópicas das séries temporais. De maneira mais específica, esse trabalho apresenta uma generalização da abordagem de curvas de complexidade-entropia, introduzindo um novo parâmetro, o chamado *embedding delay*, o qual permite investigar séries temporais em diferentes escalas de tempo. Ilustramos o uso dessa nova técnica em séries temporais do ruído harmônico, mapas caóticos e sinais periódicos ruidosos. Nessa última aplicação, observamos que quando o *embedding delay* coincide com o período do sinal, as curvas complexidade-entropia colapsam numa pequena área, indicando que nossa generalização pode ser útil para encontrar comportamentos periódicos em sinais ruidosos.

**Palavras-chave:** séries temporais, medidas de complexidade, entropia de permutação, curvas complexidade-entropia, sistemas complexos.

# Abstract

In this work, we employ a new parameter to analyze regular, chaotic and stochastic time series within the context of the so-called complexity-entropy curves. This new parameter is associated with different time scales and produces a family of curves that can be useful in the study and characterization of time series. We introduce the use of this new parameter for the complexity-entropy curves in the context of the single-parametric entropies of Tsallis and Rényi, which play the role of highlighting the different entropic characteristic of the time series. More specifically, this work generalizes the complexity-entropy curve approach, introducing a new parameter called embedding delay, which allows investigating time series at different time scales. We illustrate the use of this new technique in time series such as harmonic noise, chaotic maps and periodic noisy signals. In this latter application, we observe that when the embedding delay parameter coincides with the signal period, the complexity-entropy curves collapse into a small area, indicating that our generalization may be useful for finding periodic behaviors in noisy signals.

**Keywords:** time series, complexity measures, permutation entropy, complexity-entropy causality plane, complex systems.

# Introdução

Uma das tarefas mais comuns ao lidarmos com sistemas complexos é a busca por padrões ou regularidades em séries temporais. Embora comum, não se trata de uma tarefa trivial e, nesse contexto, a entropia [1] desempenha um papel importante para mensurar a quantidade de informações que podemos extrair dessas séries. Entretanto, a entropia por si só não é uma medida que quantifica o grau de estrutura ou os padrões presentes em um processo. Um exemplo cabal dessa situação ocorre ao analisar a entropia termodinâmica de um gás ideal e de um cristal. No primeiro caso a entropia é grande e no segundo é pequena; porém, do ponto de vista da ideia mais intuitiva de complexidade, os dois sistemas podem ser considerados pouco complexos, ou seja, ambos podem ser descritos por leis relativamente simples da Física. Por conta disso, medidas de complexidade estatística são necessárias para permitir uma quantificação do grau de estruturas presente no sistema. Essa propriedade não pode ser obtida apenas por medidas de aleatoriedade (como a entropia) e, por isso, é comum empregar os conceitos de complexidade e entropia simultaneamente [2,3].

Intuitivamente, a complexidade de um sistema pode ser pensada como uma medida de sua ordem interna. Essa medida deve estar intimamente ligada à noção de informação e de seu principal quantificador, a saber, a entropia de informação. Exemplos de sistemas como o gás ideal e um cristal demonstram que medidas de aleatoriedade não quantificam o grau de estrutura presente em um determinado processo ou sistema e a complexidade para esses dois sistemas deve ser pequena. Dessa maneira, medidas de complexidade, e mais especificamente, de complexidade estatística, devem ser capazes de quantificar a estrutura de sistemas de uma forma mais abrangente. Um quantificador de complexidade, portanto, deve se situar em algum ponto entre os dois extremos de alta e baixa complexidade, além de preservar a noção intuitiva de complexidade discutida anteriormente.

O conceito de complexidade, no entanto, não é único. A diversidade na origem das séries temporais exige uma gama de definições diferentes para a ideia de complexidade. Como consequência, algumas dessas definições não se aplicam a todo tipo de séries temporais. Além disso, grande parte das medidas de complexidade dependem de algoritmos específicos e parâmetros sensíveis a ajustes, dificultando a reprodutibilidade dos resultados. Nesse contexto, a abordagem de Bandt e Pompe (também chamada de entropia de permutação) [4] é uma medida de complexidade computacionalmente simples e robusta.

Esse método se baseia na comparação de valores vizinhos e pode ser facilmente aplicada a qualquer série temporal.

A fim de empregar os conceitos de complexidade e entropia simultaneamente, usaremos as chamadas curvas de complexidade-entropia. Os valores de complexidade e entropia são avaliados por meio da distribuição de probabilidade do sistema, ao passo que essas curvas são geradas em um plano no qual o eixo vertical corresponde aos valores de complexidade e o eixo horizontal corresponde aos valores de entropia. Mais especificamente, os valores da entropia no plano complexidade-entropia são calculados no contexto da abordagem de Bandt e Pompe, e o usual é utilizar essa abordagem avaliada pela entropia de Shannon tradicional. Porém, nesse trabalho utilizamos a substituição dessa entropia por suas generalizações: entropia  $q$  de Tsallis [5] e entropia  $\alpha$  de Rényi [6]. As curvas nos planos de complexidade-entropia para essas quantidades são geradas variando os parâmetros  $q$  e  $\alpha$ , respectivamente. Além disso, por dependerem de um parâmetro, essas medidas capturam melhor a forma entrópica da dinâmica do sistema. Isto é, esses parâmetros são capazes de acessar diferentes valores de entropia a fim de realçar diferentes escalas da entropia do sistema. Assim, com relação ao plano complexidade-entropia, essas generalizações mostraram-se muito úteis [7, 8] para classificar e caracterizar diferentes séries temporais, superando em performance o plano complexidade-entropia usual [9].

No contexto das curvas de complexidade-entropia, a principal proposta do presente trabalho é generalizar a análise dessas curvas introduzindo o parâmetro *embedding delay*, que é capaz de acessar diferentes escalas da dinâmica temporal de um sistema. Dessa maneira, além dos parâmetros entrópicos que nos permitem acessar simultaneamente diferentes escalas probabilísticas, as curvas de complexidade-entropia terão um parâmetro adicional que irá acessar diferentes escalas temporais.

No que se segue desse trabalho, utilizamos o primeiro capítulo para introduzir o conceito da entropia de informação por meio da entropia de Shannon, discutindo suas principais generalizações monoparamétricas e apresentando a abordagem de Bandt e Pompe para o cálculo da entropia de permutação. Já no segundo capítulo, nos dedicamos a mostrar o funcionamento das medidas de complexidade estatística nos planos de complexidade-entropia, bem como a generalização desse plano com respeito às entropias monoparamétricas. No terceiro e último capítulo, estudamos o efeito da variação do parâmetro *embedding delay* nas curvas  $q$ -complexidade-entropia e  $\alpha$ -complexidade-entropia. Por fim, apresentamos nossas conclusões e perspectivas.



# Capítulo 1

## Entropia de Shannon e suas generalizações

### 1.1 Algumas formas de entropia

Um poderoso conceito da teoria da informação é a chamada entropia de Shannon [1]. Essa medida é útil porque quantifica a informação de um sistema. De forma geral, pode-se extrair a quantidade de informação contida em um conjunto de dados por meio da sua distribuição de probabilidade. A distribuição, por sua vez, é moldada a partir dos dados do sistema e a entropia utiliza essa distribuição para quantificar a informação que o sistema possui. Nesse sentido, a entropia não é determinada somente pelo número de diferentes estados do sistema, mas também pela imprevisibilidade na ocorrência desses estados. A entropia de Shannon associada a uma distribuição de probabilidade  $P = \{p_1, p_2, \dots, p_n\}$  é definida como [1]

$$S(P) = \sum_{i=1}^n p_i \log_2 \frac{1}{p_i}, \quad (1.1)$$

na qual  $n$  é o número de estados acessíveis ao sistema. Essa medida pode ser pensada como o valor esperado para a informação do sistema. Caso a entropia  $S$  seja nula, temos completa certeza sobre futuros estados do sistema. Nesse caso, nosso conhecimento sobre o processo que descreve a distribuição de probabilidade é máximo. Por outro lado, caso a distribuição associada ao sistema seja uma distribuição uniforme  $U$ , nosso conhecimento é mínimo sobre a informação nele contida. Assim, se  $S(U)$  indica o valor máximo da entropia, a entropia de Shannon normalizada pode ser definida como

$$H(P) = \frac{S(P)}{S(U)}. \quad (1.2)$$

A título de exemplo, a entropia de Shannon no caso de um sistema de dois estados ( $A$  e  $B$ ) é

$$S = -(p \log_2 p + q \log_2 q), \quad (1.3)$$

sendo  $p$  a probabilidade de ocorrer o estado  $A$  e  $q$  a probabilidade de ocorrer  $B$ . Vale notar que  $p + q = 1$  por conta da normalização. A figura 1.1 mostra o comportamento da entropia em função de  $p$ . Observamos que a entropia é nula se  $p = 0$  ou  $p = 1$  ( $q = 0$ ), o que corresponde aos casos em que apenas um dos estados do sistema ocorre. Por outro lado, a entropia é máxima quando  $p = q = 1/2$ , o que corresponde ao caso em que  $A$  e  $B$  são equiprováveis.

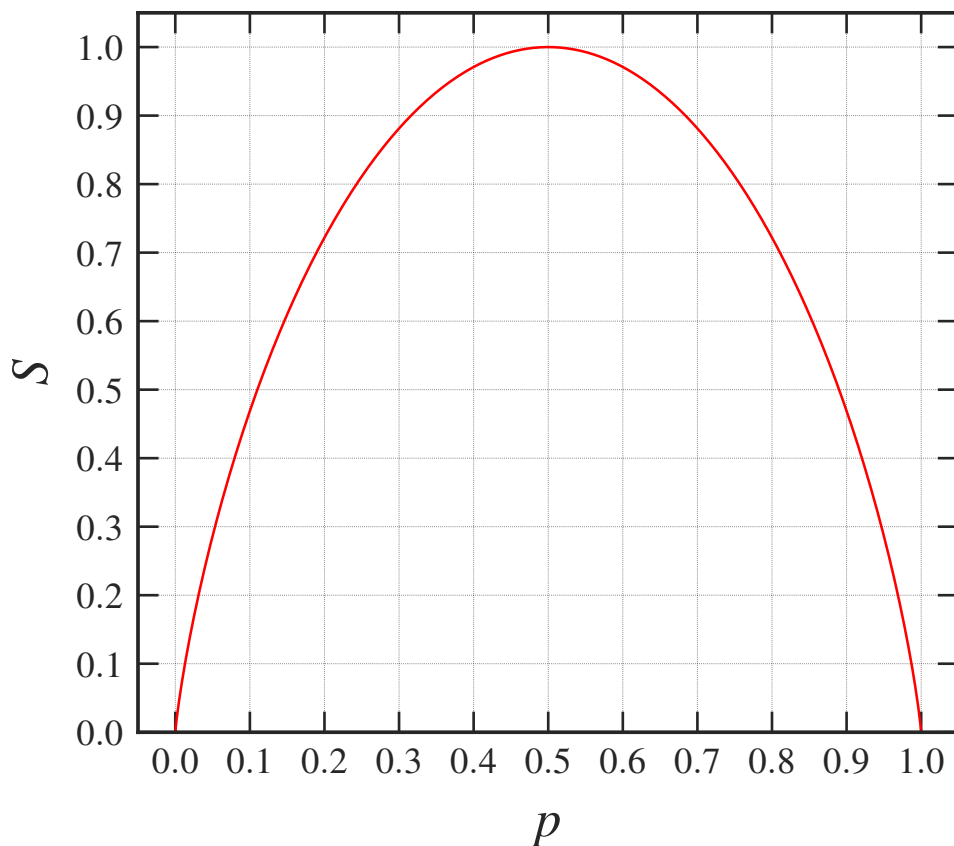


Figura 1.1: Entropia de Shannon para a distribuição  $P = \{p, 1 - p\}$  em função de  $p$ .

Embora amplamente difundida, a entropia de Shannon não é a única medida disponível para quantificar a informação contida na dinâmica dos sistemas. De fato, a definição de entropia de Shannon pode ser generalizada a fim de fornecer informações adicionais sobre a dinâmica dos sistemas levando em conta, por exemplo, diferentes escalas probabilísticas que ponderem de modo diferente os valores da distribuição de probabilidade do sistema.

Medidas de entropia que podem fornecer essa propriedade usualmente assumem formas que dependem de potências das probabilidades, isto é,  $p_i^\gamma$ . Notamos que essa forma é mais sensível a eventos mais frequentes se  $\gamma > 1$  e mais sensível a eventos raros quando  $\gamma < 1$ . Duas entropias monoparamétricas que apresentam essa característica são a entropia de Tsallis [5] e a de Rényi [6].

A entropia de Tsallis (ou entropia  $q$ ) é definida como [5]

$$S_q(P) = - \sum_{i=1}^n p_i \ln_q p_i, \quad (1.4)$$

sendo  $q$  um parâmetro real e

$$\ln_q(x) = \begin{cases} \ln(x) & \text{se } x > 0 \text{ e } q = 1 \\ \frac{x^{1-q} - 1}{1-q} \text{ ou } \int_1^x t^{-q} dt & \text{se } x > 0 \text{ e } q \neq 1, \\ \text{Indefinido} & \text{se } x \leq 0 \end{cases} \quad (1.5)$$

uma generalização da função logarítmica. Vale observar que  $S_1$  recupera a entropia de Shannon. Além disso, podemos escrever a entropia  $q$  normalizada como

$$H_q(P) = \frac{S_q(P)}{S_q(U)}, \quad (1.6)$$

na qual  $S_q(U) = \ln_q(n)$  é o valor máximo para a entropia de Tsallis, que ocorre quando  $p_i = 1/n$ .

Já a entropia de Rényi (ou entropia  $\alpha$ ) pode ser escrita como [6]

$$S_\alpha(P) = \frac{1}{1-\alpha} \ln \sum_{i=1}^n p_i^\alpha, \quad (1.7)$$

sendo  $\alpha > 0$  um parâmetro real. No limite em que  $\alpha \rightarrow 1$ , a entropia de Rényi também recupera a entropia de Shannon. Além disso, a entropia normalizada de Rényi pode ser escrita como

$$H_\alpha(P) = \frac{S_\alpha(P)}{S_\alpha(U)}, \quad (1.8)$$

na qual  $S_\alpha(U) = \ln(n)$  é o valor máximo para essa entropia, que também ocorre quando os estados do sistema são equiprováveis.

## 1.2 Entropia de permutação: abordagem de Bandt e Pompe

Baseado na entropia de Shannon, a entropia de permutação [4] é um método de caracterização de séries temporais. Trata-se de uma abordagem que pode ser aplicada à séries temporais arbitrárias e que é bastante robusta com relação à presença de ruídos, além de não depender do ajuste de parâmetros.

A entropia de permutação é, de fato, facilmente obtida para qualquer tipo de série tem-

poral, sendo uma medida computacionalmente rápida e de simples interpretação. Além disso, devido à sua capacidade de distinguir entre séries temporais regulares, caóticas e aleatórias, esse método é amplamente difundido entre a comunidade científica [9–11].

Para uma série temporal de  $n$  elementos

$$\{x_1, x_2, \dots, x_n\} = \{x_t\}_{t=1,2,\dots,n}, \quad (1.9)$$

o método de Bandt e Pompe ou entropia de permutação quantifica o grau de incerteza na ocorrência dos padrões ordinais de valores consecutivos. Em outras palavras, essa medida nos fornece a dinâmica de ordenamento de parcelas consecutivas de elementos da série. Esse conjunto de valores consecutivos podem ser pensados como partições de tamanho  $d > 1$ , a chamada *embedding dimension*. O número de partições que pode ser construído é  $n - (d - 1)\tau$  e cada partição pode ser representada por

$$(\vec{s}) = \{x_{s-(d-1)\tau}, x_{s-(d-2)\tau}, \dots, x_{s-\tau}, x_s\}, \quad (1.10)$$

com  $s = d, d + 1, \dots, n$ . Nessa relação,  $\tau$  é o chamado *embedding delay*, o qual permite selecionar elementos espaçados em  $\tau$  unidades de tempo. Inicialmente iremos usar  $\tau = 1$ ; porém, veremos que a variação desse parâmetro pode realçar a dinâmica do sistema em diferentes escalas temporais.

Vamos ilustrar essa técnica por meio de um exemplo concreto. Consideremos uma série com  $n = 8$  termos, isto é,

$$x = \{5, 8, 10, 11, 7, 12, 4, 3\}. \quad (1.11)$$

Por simplicidade, vamos escolher  $d = 2$ , de modo que podemos representar essas  $n - d + 1 = 7$  partições por meio dos seguintes vetores:

$$\begin{aligned} (\vec{2}) &= (5, 8), \\ (\vec{3}) &= (8, 10), \\ (\vec{4}) &= (10, 11), \\ (\vec{5}) &= (11, 7), \\ (\vec{6}) &= (7, 12), \\ (\vec{7}) &= (12, 4), \\ (\vec{8}) &= (4, 3). \end{aligned}$$

Uma vez que  $d = 2$ , existem  $d! = 2$  grupos de permutações ou possíveis maneiras para ordenar os elementos em cada partição  $(\vec{s})$ . A primeira permutação indica o caso

em que o primeiro elemento é menor do que o segundo e é representada por “01”. A segunda permutação ocorre quando o segundo elemento é menor do que o primeiro e é representada por “10”. O quadro a seguir mostra a classificação das 7 partições anteriores de acordo com o tipo de permutação.

permutação “01”	permutação “10”
$(\vec{2}) = (5, 8)$	$(\vec{5}) = (11, 7)$
$(\vec{3}) = (8, 10)$	$(\vec{7}) = (12, 4)$
$(\vec{4}) = (10, 11)$	$(\vec{8}) = (4, 3)$
$(\vec{6}) = (7, 12)$	

Nesse exemplo, existem quatro pares do tipo “01” e três pares do tipo “10”. Dessa forma, temos a probabilidade de encontrar vetores do tipo “01” dada por  $p(\text{“01”}) = 4/7$  e do tipo “10” dada por  $p(\text{“10”}) = 3/7$ . Assim, escrevemos a entropia de permutação como

$$S = -\frac{4}{7} \log_2 \left( \frac{4}{7} \right) - \frac{3}{7} \log_2 \left( \frac{3}{7} \right) \approx 0,985. \quad (1.12)$$

Nesse exemplo, a base do logaritmo é 2, de modo que a entropia é medida em *bits*. Da mesma maneira, quando temos partições de tamanho  $d = 3$ , os  $n - d + 1 = 6$  vetores são dados por:

$$\begin{aligned} (\vec{3}) &= (5, 8, 10), \\ (\vec{4}) &= (8, 10, 11), \\ (\vec{5}) &= (10, 11, 7), \\ (\vec{6}) &= (11, 7, 12), \\ (\vec{7}) &= (7, 12, 4), \\ (\vec{8}) &= (12, 4, 3). \end{aligned}$$

Em relação à ordem dos elementos, teremos  $d! = 6$  possibilidades. Os elementos dos vetores  $(\vec{3})$  e  $(\vec{4})$  estão na ordem crescente e, portanto, são representados por “012”. Respeitando a ordem de cada componente, os vetores  $(\vec{5})$  e  $(\vec{7})$  são representados por “201”, o vetor  $(\vec{8})$  por “210” e o vetor  $(\vec{6})$  por “102”. Note que nesse exemplo os padrões ordinais “021” e “120” não ocorrem. Desse modo, as probabilidades ficam

$$\begin{aligned} p(\text{“012”}) &= p(\text{“201”}) = 2/6, \\ p(\text{“102”}) &= p(\text{“210”}) = 1/6, \end{aligned}$$

conduzindo a entropia de permutação

$$S = -2 \left( \frac{2}{6} \right) \log_2 \left( \frac{2}{6} \right) - 2 \left( \frac{1}{6} \right) \log_2 \left( \frac{1}{6} \right) \approx 1,918. \quad (1.13)$$

De maneira geral, o procedimento anterior para encontrar as probabilidades de cada padrão ordinal pode ser expresso por

$$p(\pi_i) = \frac{\#(s|s \leq (n - (d - 1)\tau); (\vec{s}) \text{ do tipo } \pi_i)}{n - (d - 1)\tau}, \quad (1.14)$$

com  $\#$  representando o número de ocorrências da permutação  $\pi_i$ . Usando essa notação, a entropia de permutação pode ser escrita como

$$S[P] = - \sum_{i=1}^{d!} p(\pi_i) \log_2 p(\pi_i), \quad (1.15)$$

com  $P = \{p(\pi_i)\}_{i=1,2,\dots,d!}$  e  $d!$  sendo o número de permutações acessíveis para a *embedding dimension*  $d$ .

Para o caso em que todas as permutações são igualmente prováveis, ou seja,

$$p(\pi_i) = \frac{1}{d!}, \quad (1.16)$$

a entropia de permutação terá seu valor máximo, isto é,

$$S_{\max} = S[1/d!] = \log_2 d!.$$

Assim, podemos definir a entropia de permutação normalizada como

$$H[P] = \frac{S[P]}{\log_2 d!}, \quad (1.17)$$

a qual está limitada no intervalo  $0 \leq H[P] \leq 1$ , com  $H[P] = 0$  para uma série que apresenta apenas uma permutação e  $H[P] = 1$  para uma série na qual todas as permutações são equiprováveis.

Para ilustrar a aplicação, Bandt e Pompe [4] estudaram a série temporal das amplitudes do sinal sonoro obtido a partir da pronúncia da frase “entropia de permutação mede complexidade” (em inglês), a fim de encontrar os segmentos contendo voz ativa. Essa série é mostrada na figura 1.2(a). Eles também utilizaram uma medida chamada *zero-crossing rate* (ZCR) [12] para comparar com a entropia  $H$ . Esse método mede a taxa com que a amplitude do sinal muda de positivo para negativo dentro de um intervalo de tempo, isto é, a taxa local com que a amplitude do sinal passa pelo valor zero.

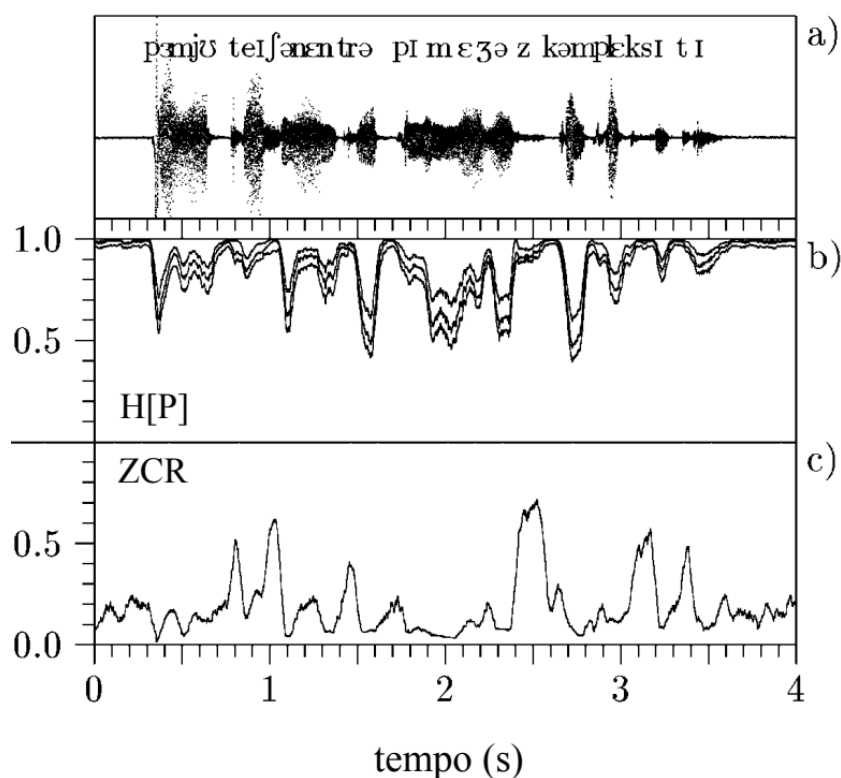


Figura 1.2: (a) Série temporal das amplitudes sonoras para a pronúncia da frase “entropia de permutação mede complexidade” (em inglês). A transcrição fonética da fala se encontra acima da série. A taxa de amostragem é de 11 kHz e a unidade da amplitude é arbitrária. (b) Entropia de permutação normalizada  $H[P]$  para  $d = 2, 3$  e  $4$  (de cima para baixo). (c) Valor da medida *zero-crossing rate* (ZCR). Ambas quantidades calculadas se situam dentro de uma janela de tempo móvel com tamanho igual a 46 milissegundos. Figura extraída da referência [13].

Muito usada para detectar segmentos sonoros de curta duração, a ZCR assume que a energia está concentrada em baixas frequências para a voz ativa, acarretando em poucas oscilações por unidade de tempo e, conseqüentemente, uma contagem baixa de passagens pelo zero [12]. Dessa forma, é esperado um valor baixo para a ZCR quando a voz ativa estiver presente e um valor alto caso contrário. Na figura 1.2(a), é possível observar que o sinal começa e termina com ruído e, portanto, os valores para a entropia  $H$  são bem próximos a um, como visto pela figura 1.2(b). O mesmo ocorre para os chamados *unvoiced sounds*, sons consonânticos (de baixa intensidade) e para as pausas durante a fala. Por outro lado, é possível notar um decréscimo no valor da entropia nos segmentos que contém voz ativa. A figura 1.2(b) indica que as duas técnicas se comportam similarmente, mas é possível notar que a entropia de permutação identifica melhor presença de voz ativa. Podemos observar, por exemplo, que mesmo onde há silêncio (entre 0 a 0,3 segundos ou a partir de 3,6 segundos) um valor baixo da medida ZCR corresponde à presença de voz. Já a entropia de permutação próxima de 1 indica silêncio, sendo claramente uma melhor medida nesse caso.

# Capítulo 2

## Medidas de Complexidade

### 2.1 A complexidade estatística

Quantificadores de complexidade podem fornecer importante intuição sobre o caráter estrutural de um dado sistema. Em seu artigo [4], Bandt e Pompe mencionam estudos em que medidas de complexidade aplicadas a dados obtidos do coração e do cérebro são capazes de distinguir entre indivíduos saudáveis e doentes e até prever ataques cardíacos ou ataques epiléticos.

Para criar uma medida de complexidade que seja condizente com o que foi discutido na introdução, isto é, quantificar o grau de estrutura do sistema e ao mesmo tempo considerar ambos os extremos de entropia no caso do gás ideal e um cristal, López-Ruiz *et al.* [2] definem uma medida chamada “desequilíbrio”,  $\mathcal{D}$ , que pode ser entendida como uma medida da distância entre a distribuição equiprovável e a distribuição de probabilidade dos estados acessíveis ao sistema. A ideia é que essa medida de “desequilíbrio” leve em conta a existência de uma hierarquia das probabilidades existentes no sistema. Caso haja estados privilegiados ou mais prováveis dentre os acessíveis,  $\mathcal{D}$  será diferente de zero.

No exemplo de um cristal perfeito, os átomos estão completamente ordenados e, devido à sua simetria, a distribuição de probabilidade dos estados acessíveis está centrada em torno de um único estado. Para essa hierarquia, o “desequilíbrio” tem um valor máximo. Por outro lado, para um gás ideal, o sistema tem a mesma probabilidade de ser encontrado em qualquer um dos estados acessíveis e, portanto, obtemos um desequilíbrio mínimo, pois há a ausência de uma hierarquia.

A figura 2.1 ilustra o fato de que as medidas de entropia e desequilíbrio não são capazes, por si só, de quantificar a complexidade do sistema. A solução para esse problema, proposta por López-Ruiz *et al.*, foi considerar o produto entre as medidas de entropia  $H$  e o desequilíbrio  $\mathcal{D}$ . Dessa forma, essa nova quantidade apresenta as características requeridas anteriormente, ou seja, tender a zero tanto para o cristal quanto para o gás ideal e ser diferente de zero para outros sistemas mais complexos.



Aliada às ideias de Bandt e Pompe, a proposta de López-Ruiz *et al.* permite obter uma medida de complexidade estatística que realça informações de séries temporais que outrora não eram quantificadas pelas medidas de entropia de permutação.

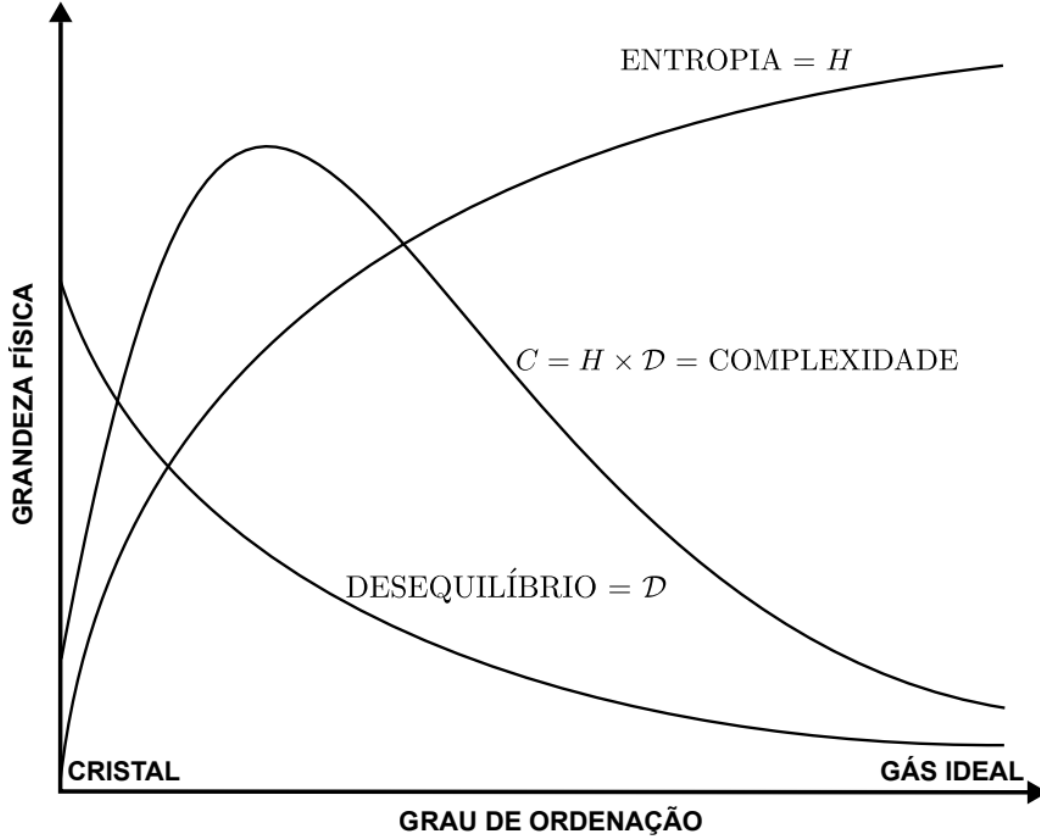


Figura 2.1: Imagem que descreve o comportamento esperado para a entropia  $H$ , desequilíbrio  $\mathcal{D}$  e complexidade  $C = H\mathcal{D}$  de um sistema. Cada uma dessas medidas são funções do grau de ordenação do sistema. Essa definição ilustra ambos os extremos: o cristal perfeito e o gás ideal. Imagem extraída da referência [13].

No contexto da abordagem de Bandt e Pompe, a complexidade estatística de López-Ruiz pode ser expressa como

$$C[P] = H[P]\mathcal{D}[P] = -C_0 \left[ \sum_{i=1}^{d!} p(\pi_i) \log p(\pi_i) \right] \left[ \sum_{i=1}^{d!} \left( p(\pi_i) - \frac{1}{d!} \right)^2 \right], \quad (2.1)$$

na qual  $C_0$  é uma constante de normalização,  $H[P]$  é a entropia normalizada e  $\mathcal{D}[P]$  é a distância euclidiana entre a distribuição de probabilidade do sistema  $P = \{p(\pi_i)\}_{i=1,2,\dots,d!}$  e a distribuição uniforme  $U = \{1/d!\}_{i=1,2,\dots,d!}$ .

Essa medida de complexidade satisfaz os dois casos extremos, isto é, para o cristal perfeito o desequilíbrio  $\mathcal{D}$  será grande por conta da sua simetria (presença de um estado privilegiado), mas sua entropia baixa acarreta em  $C \rightarrow 0$ . Para o gás ideal, por sua vez, teríamos  $H$  grande e o desequilíbrio  $\mathcal{D} \rightarrow 0$ , acarretando também em  $C \rightarrow 0$ . Mais

adiante iremos considerar outra definição para o desequilíbrio  $\mathcal{D}$ .

Antes de prosseguirmos, é importante realçar o fato de que a medida de complexidade proposta por López-Ruiz *et al.* não é uma função unívoca da entropia. Isso acontece porque diferentes distribuições de probabilidade podem levar ao mesmo valor de entropia e, ao mesmo tempo, diferentes valores de desequilíbrio  $\mathcal{D}$ , ao passo que se obtém um intervalo de valores para a complexidade  $C$  ao fixar o valor de  $H$ .

A título de exemplo, iremos considerar a distribuição de probabilidade  $P = \{p, q, 1 - (p + q)\}$ , com  $p > 0$ ,  $q > 0$  e  $p + q \leq 1$ , de modo que a entropia é

$$S = -p \log_2 p - q \log_2 q - [1 - (p + q)] \log_2 [1 - (p + q)]. \quad (2.2)$$

O desequilíbrio para essa probabilidade fica

$$\mathcal{D} = \left(p - \frac{1}{3}\right)^2 + \left(q - \frac{1}{3}\right)^2 + \left\{[1 - (p + q)] - \frac{1}{3}\right\}^2. \quad (2.3)$$

Por fim, a complexidade para esse caso é dada por

$$C = (-p \log_2 p - q \log_2 q - [1 - (p + q)] \log_2 [1 - (p + q)]) \left[ \left(p - \frac{1}{3}\right)^2 + \left(q - \frac{1}{3}\right)^2 + \left([1 - (p + q)] - \frac{1}{3}\right)^2 \right]. \quad (2.4)$$

Variando os valores de  $p$  e  $q$ , e fazendo um gráfico com os valores de complexidade em função da entropia, podemos verificar que para um dado valor de entropia podem ocorrer mais de um valor para complexidade. Esse gráfico é mostrado na figura 2.2, na qual podemos concluir que a complexidade não é uma função unívoca da entropia.

## 2.2 Plano complexidade-entropia

Embora a entropia de permutação seja uma medida simples e poderosa para caracterizar séries temporais, Rosso *et al.* [9] perceberam que essa medida não é capaz de distinguir processos estocásticos de caóticos. Foi observado, por exemplo, que séries temporais do movimento browniano fracionário podem apresentar a mesma entropia de séries do mapa logístico. O diagrama da figura 2.2, o qual é composto pela entropia no eixo horizontal e pela complexidade no eixo vertical, é chamado de plano complexidade-entropia. O fato descrito nessa figura serviu como motivação para Rosso *et al.* usarem os valores de entropia e complexidade para distinguir sinais caóticos e estocásticos.

De fato, para distinguir esses tipos de séries, eles propuseram o uso de uma representação baseada no trabalho de López-Ruiz *et al.* [2], na qual duas medidas são avaliadas: a entropia de permutação normalizada  $H$  e a complexidade estatística  $C$  associada à divergência de Jensen-Shannon [9].

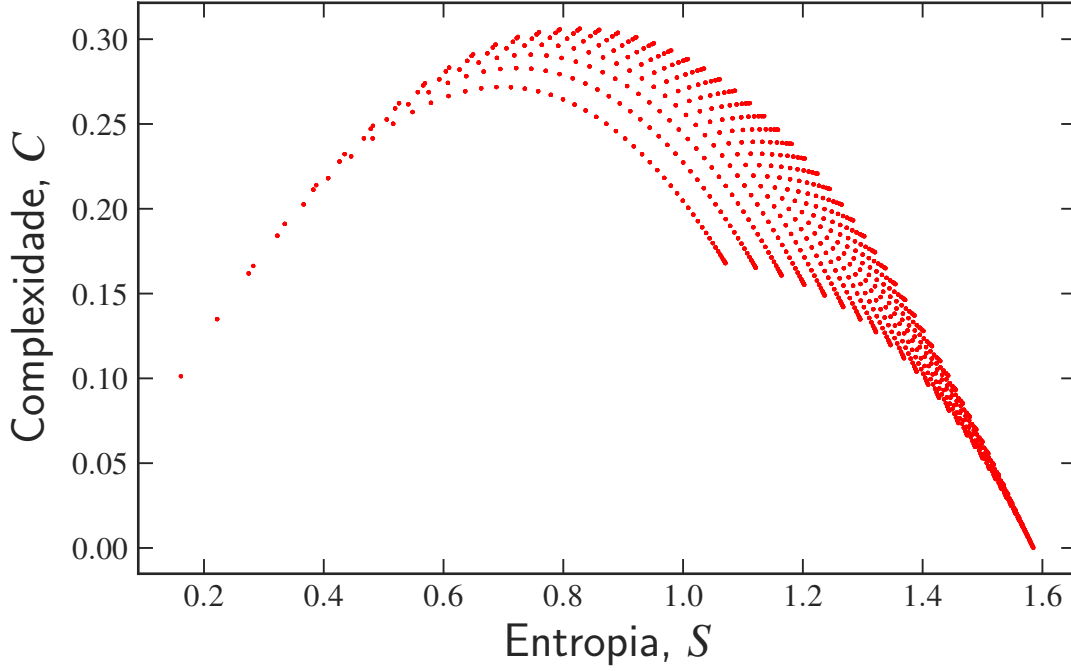


Figura 2.2: Gráfico dos valores complexidade versus entropia para as expressões (2.2) e (2.4) ao variarmos os valores de  $p$  e  $q$  entre 0 e 1 com passos de tamanho 0,01 respeitando a condição de que  $p + q \leq 1$ .

De maneira mais específica, Rosso *et al.* utilizaram a abordagem de Bandt e Pompe para extrair a distribuição dos padrões ordinais  $P = \{p(\pi_i)\}_{i=1,2,\dots,d!}$  de várias séries temporais simuladas cuja dinâmica era conhecida ser caótica ou estocástica. Usando essas distribuições, eles calcularam a entropia de permutação normalizada

$$H[P] = -\frac{1}{\ln d!} \sum_i^{d!} p(\pi_i) \ln p(\pi_i), \quad (2.5)$$

e a complexidade estatística

$$C[P] = \frac{H[P]\mathcal{D}[P]}{\mathcal{D}_0}, \quad (2.6)$$

sendo  $\mathcal{D}[P]$  o desequilíbrio medido pela divergência de Jensen-Shannon [9], ou seja,

$$\mathcal{D}[P] = \left[ S\left(\frac{P+U}{2}\right) - \frac{S[P]}{2} - \frac{S[U]}{2} \right], \quad (2.7)$$

com

$$\frac{P+U}{2} = \left\{ \frac{p(\pi_i) + 1/d!}{2} \right\}_{i=1,2,\dots,d!}, \quad (2.8)$$

e  $\mathcal{D}_0$  uma constante de normalização de valor

$$\mathcal{D}_0 = -\frac{1}{2} \left[ \frac{d! + 1}{d!} \ln(d! + 1) - \ln d! - 2 \ln 2 \right]. \quad (2.9)$$

O plano complexidade-entropia foi estimado para diversos tipos de séries temporais, como mostrado na figura 2.3. Levando em conta a posição que ocupam nesse gráfico, Rosso *et al.* conseguiram diferenciar séries temporais. Eles ressaltam, por exemplo, que sistemas caóticos estão localizados próximos à região de máxima complexidade. Esse fato está relacionado à ausência de certos padrões ordinais, o que acarreta em um grau maior de desequilíbrio e, portanto, uma maior complexidade. Dessa forma, a complexidade estatística quantifica, além da aleatoriedade, esse grau de estruturas existentes na série temporal.

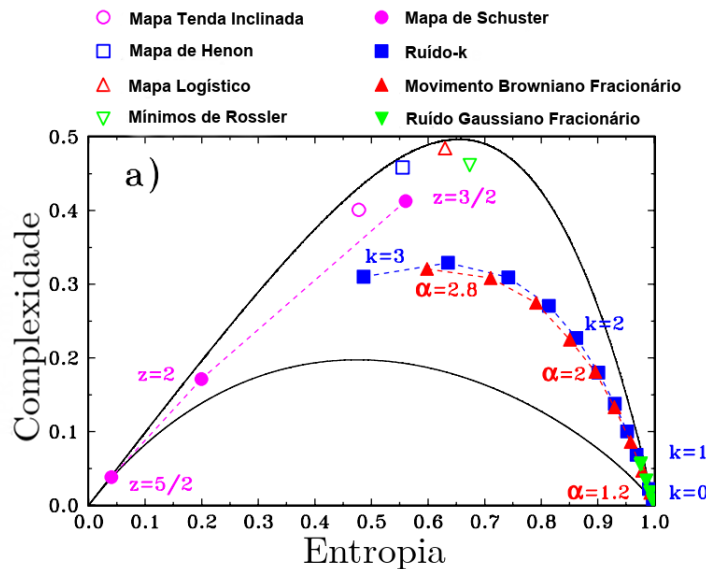


Figura 2.3: As linhas contínuas representam as complexidades mínimas  $C_{\min}$  e máximas  $C_{\max}$ , e a região entre esses dois valores delimita o plano complexidade-entropia. Os símbolos nesse gráfico indicam a localização de diferentes sinais caóticos (os cinco primeiros da legenda) e estocásticos (os três últimos) nesse plano. Podemos notar, por exemplo, que no caso de séries estocásticas do ruído- $k$ , embora tenham a mesma entropia de séries caóticas do mapa logístico, os valores das complexidades são completamente diferentes. Assim, mesmo possuindo a mesma entropia, esses sinais são facilmente distinguíveis pelo plano. Figura extraída e adaptada da referência [9].

## 2.3 Generalização do plano complexidade-entropia para entropias monoparamétricas

Apesar de ser aplicado com sucesso para o estudo de diversos sistemas, o plano complexidade-entropia de Rosso *et al.* não se mostra suficiente para distinguir certos

tipos de séries temporais [7, 8]. Motivados por este fato, alguns autores decidiram estender o plano complexidade-entropia considerando as entropias monoparamétricas de Tsallis e de Rényi. Nessa seção iremos apresentar tais generalizações.

### 2.3.1 Curvas $q$ -complexidade-entropia de Tsallis

Ribeiro *et al.* [7] propõem substituir a equação da entropia de Shannon normalizada (1.2) e da complexidade estatística (2.6) por generalizações monoparamétricas baseadas na entropia  $q$  de Tsallis. A entropia, portanto, será dada pela equação (1.6) normalizada, isto é

$$H_q(P) = \frac{S_q(P)}{S_q(U)},$$

enquanto a complexidade será

$$C_q(P) = \frac{D_q(P, U)H_q(P)}{D_q^*}, \quad (2.10)$$

com o desequilíbrio generalizado sendo

$$D_q(P, U) = \frac{1}{2}K_q\left(P\left|\frac{P+U}{2}\right.\right) + \frac{1}{2}K_q\left(U\left|\frac{P+U}{2}\right.\right), \quad (2.11)$$

na qual  $K_p(P, R) = -\sum p_i \ln_q r_i/p_i$  é a generalização da divergência de Kullback-Leibler no contexto da entropia de Tsallis [5]. Explicitamente, podemos escrever

$$D_q(P, U) = -\frac{1}{2} \sum_{i=1}^{d!} p_i \ln_q \frac{p_i + 1/d!}{2p_i} - \frac{1}{2} \sum_{i=1}^{d!} \frac{1}{d!} \ln_q \frac{p_i + 1/d!}{2/d!}, \quad (2.12)$$

com

$$D_q^* = \max D_q(P, U),$$

$$D_q^* = \frac{2^{2-q}d! - (1+d!)^{1-q} - d!(1+1/d!)^{1-q} - d! + 1}{(1-q)2^{2-q}d!},$$

sendo a constante de normalização.

As medidas  $H_q$  e  $C_q$  são generalizações normalizadas da entropia e da complexidade utilizadas por Rosso *et al.* Como os valores de complexidade e entropia dependem do parâmetro  $q$ , os valores  $\{H_q, C_q\}$  representam uma curva parametrizada pelos valores de  $q$ . Além disso, no limite de  $q \rightarrow 1$ , recuperamos o plano complexidade-entropia usual.

Para uma dada distribuição de probabilidade  $P = \{p_j(\pi_j)\}_{j=1, \dots, d!}$ , seja  $r$  o número de componentes diferentes de zero de  $P$  (ou seja, o número de permutações  $\pi_j$  que realmente ocorrem na série temporal) e  $\gamma = \frac{r-1}{d!-1}$  (isto é, a fração de permutações dentre todas  $d!$  possíveis). Assim, Ribeiro *et al.* enumeram e demonstram as seguintes propriedades

dessas curvas de complexidade-entropia:

- (i) Se  $r = 1$ , então  $H_q = 0$  e  $C_q(P) = 0$  para qualquer  $q > 0$ .
- (ii)  $H_q(P) \rightarrow \gamma$  e  $C_q(P) \rightarrow \gamma(1 - \gamma)$  quando  $q \rightarrow 0^+$ .
- (iii)  $r > 1$ , então  $H_q(P) \rightarrow 1$  e  $C_q(P) \rightarrow 1 - \gamma$  quando  $q \rightarrow \infty$ .

Essas propriedades gerais de  $H_q(P)$  e  $C_q(P)$  têm as seguintes consequências:

1. A curva  $q$  de complexidade-entropia de uma série temporal que só tem uma permutação  $\pi_j$  (isto é,  $r = 1$ ) se colapsa no ponto  $(0, 0)$ .
2. Para uma série temporal que possui todas as possíveis permutações  $\pi_j$  (isto é,  $r = d!$  e  $\gamma = 1$ ), a curva  $q$  de complexidade-entropia é uma curva fechada que começa no ponto  $(1, 0)$  para  $q = 0^+$  e termina no mesmo ponto para  $q \rightarrow \infty$ .
3. Para uma série temporal que não possui todas as permutações  $\pi_j$ , a curva  $q$  de complexidade-entropia começa no ponto  $(\gamma, \gamma(1 - \gamma))$  para  $q = 0^+$  e termina no ponto  $(1, 1 - \gamma)$  para  $q \rightarrow \infty$ . O número de permutações  $r$  que realmente acontecem pode ser obtido por  $\gamma$  via  $r = (d! - 1)\gamma + 1$ , no qual  $0 < \gamma < 1$ .

Ribeiro *et al.* comentam que séries temporais estocásticas são geralmente caracterizadas por curvas fechadas no plano complexidade-entropia, enquanto que séries caóticas têm curvas abertas (em especial para grandes valores da *embedding dimension*). Essa última característica está relacionada à existência de padrões ordinais proibidos na dinâmica caótica [7]. Além disso, os mesmos autores evidenciam que quanto menor a autocorrelação da série, menor será a área formada por sua curva no plano complexidade-entropia. Em outros termos, quanto mais aleatória for a série, menor será essa área. Isso faz sentido porque ao passo que a aleatoriedade da série aumenta (o caso máximo é quando a distribuição apresenta uma probabilidade uniforme), o desequilíbrio diminui e conseqüentemente a complexidade também.

A figura 2.4 mostra a dependência da entropia e complexidade para com o parâmetro  $q$ , começando de  $(H_q, C_q)$  para  $q = 0^+$  e fazendo  $q \rightarrow \infty$  para várias séries temporais caóticas (as primeiras duas linhas) e estocásticas (as duas últimas linhas).

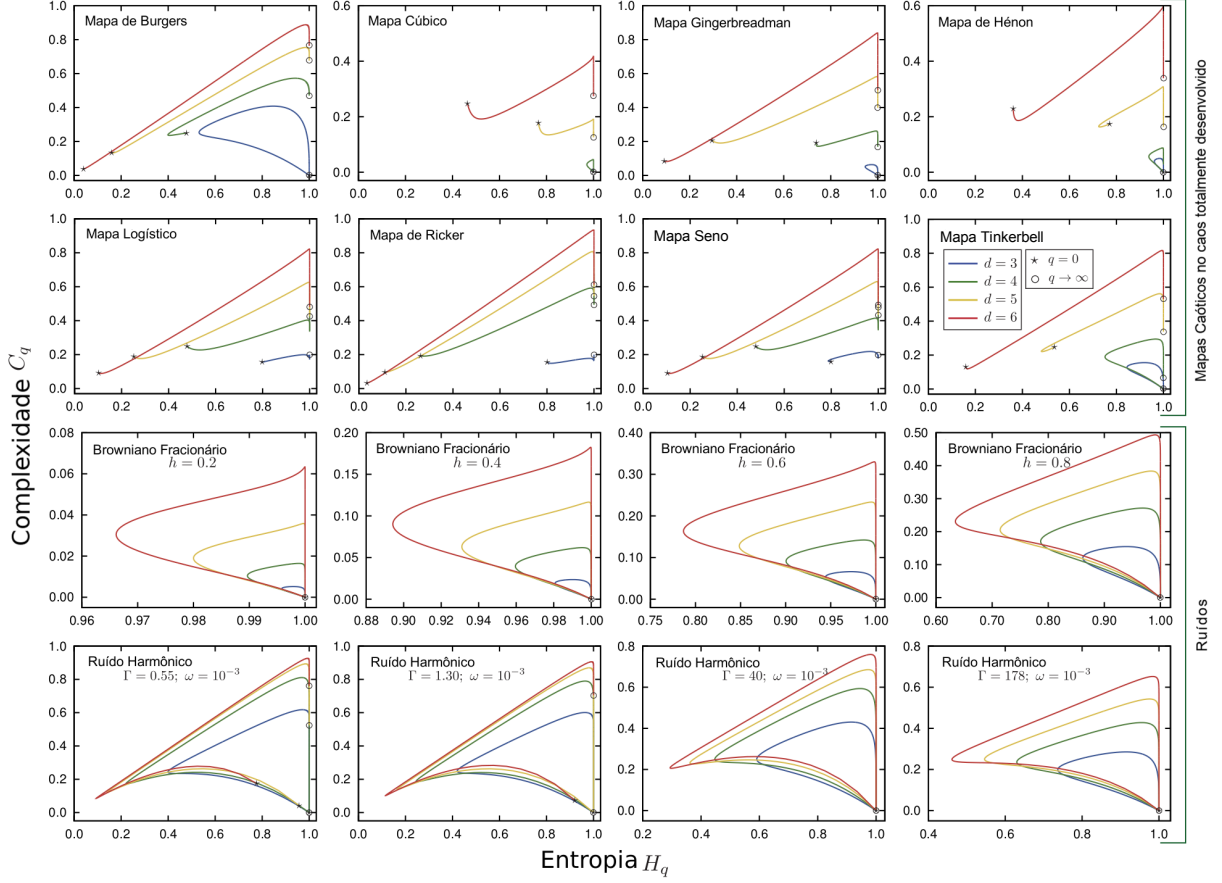


Figura 2.4: Cada painel mostra curvas  $q$ -complexidade-entropia para mapas caóticos ou processos estocásticos, com o parâmetro *embedding dimension* tomando valores  $d = 3, 4, 5, 6$ . Os marcadores estrela e círculo aberto representam pontos nos quais  $q = 0^+$  e  $q \rightarrow \infty$ , respectivamente. Figura extraída e adaptada de [7].

### 2.3.2 Curvas $\alpha$ -complexidade-entropia de Rényi

Além da generalização anterior, Jauregui *et al.* [8] propõem substituir a entropia usual de Shannon normalizada pela entropia  $\alpha$  de Rényi. Essa medida é dada pela equação (1.8), ou seja,

$$H_\alpha(P) = \frac{S_\alpha(P)}{S_\alpha(U)}, \quad (2.13)$$

enquanto a complexidade é

$$C_\alpha(P) = \frac{D_\alpha(p)H_\alpha(p)}{D_\alpha^*}, \quad (2.14)$$

sendo

$$D_\alpha(P, U) = \frac{1}{2(\alpha - 1)} \left[ \ln \sum_{i=1}^{d!} p_i^\alpha \left( \frac{p_i + 1/d!}{2} \right)^{1-\alpha} + \ln \sum_{i=1}^{d!} \frac{1}{(d!)^\alpha} \left( \frac{p_i + 1/d!}{2} \right)^{1-\alpha} \right], \quad (2.15)$$

e

$$D_\alpha^* = \frac{1}{2(\alpha - 1)} \ln \left[ \frac{(d! + 1)^{1-\alpha} + d! - 1}{d!} \left( \frac{d! + 1}{4d!} \right)^{1-\alpha} \right]. \quad (2.16)$$

A quantidade  $D_\alpha$  é a divergência de Jensen-Shannon generalizada. No contexto da entropia de Rényi, essa medida pode ser entendida como uma distância entre a distribuição de probabilidade  $P$  e a distribuição uniforme  $U$ .

Em contraste com as curvas  $q$ -complexidade-entropia, as curvas de Rényi nunca são fechadas. Jauregui *et al.* enunciam e demonstram as seguintes propriedades dessas curvas:

- (i) Se apenas uma permutação ocorrer, a curva  $\alpha$ -complexidade-entropia se reduz a um único ponto  $(0, 0)$ . Um exemplo é se a série temporal for estritamente monótona.
- (ii) Se todas as permutações permitidas ocorrerem, então a curva  $\alpha$ -complexidade-entropia começa no ponto  $(1, 0)$ , correspondente a  $\alpha \rightarrow 0^+$ , e termina no ponto  $(h_f, c_f)$  correspondente a  $\alpha \rightarrow \infty$ . O ponto final  $(h_f, c_f)$  nos dá o valor da componente mínima  $p_m$  da distribuição de probabilidade por meio da relação

$$p_m = \frac{4p_M}{d!p_M + 1} \left( \frac{d+1}{4d!} \right)^{c_f/h_f} - \frac{1}{d!}.$$

- (iii) Caso ocorram  $r$  permutações, com  $1 < r < d!$ , então a curva Rényi de complexidade-entropia começa no ponto  $(h_i, c_i)$  e termina no ponto  $(h_f, c_f)$ . O valor de  $r$  é obtido pela coordenada inicial da entropia,  $h_i$ , por meio da relação  $r = (d!)^{h_i}$ . Do ponto final, obtemos o valor máximo da componente  $p_M$  da distribuição de probabilidade usando  $p_M = (d!)^{-h_f}$ . Nesse caso, o valor da componente mínima de  $p$  é, portanto, zero.

Como consequência da propriedade (iii), quando as  $r$  permutações tiverem a mesma probabilidade, isto é,  $1/r$ , teremos  $p_M = 1/r$  e, de  $r = (d!)^{h_i}$  ficamos com  $p_M = (d!)^{-h_i}$ , obtendo  $h_f = h_i = \ln r / \ln d!$ . Dessa forma, a curva  $\alpha$ -complexidade-entropia de Rényi é uma linha vertical. A complexidade, nesse caso, ainda depende de  $\alpha$ . Essa situação surge para o caso no qual a série temporal exibe um padrão periódico.

A figura 2.5 mostra a forma das curvas  $\alpha$ -complexidade-entropia para os mapas caóticos em completo caos, variando-se o parâmetro  $\alpha$ . Para cada mapa os autores escolheram o parâmetro *embedding dimension*  $d \in \{3, 4, 5, 6\}$ .



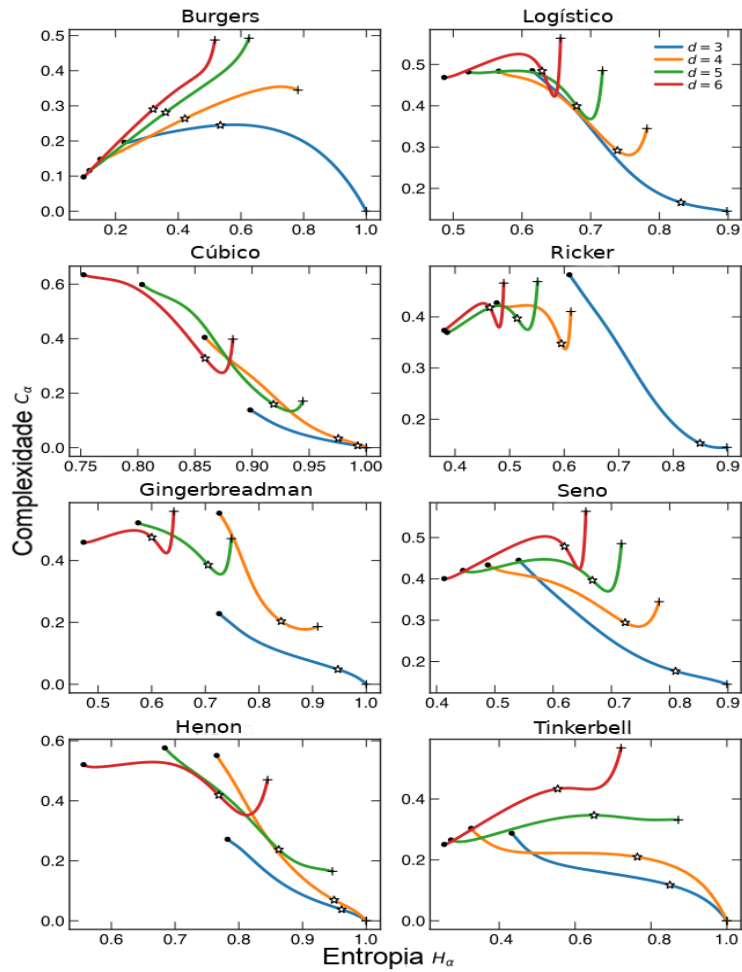


Figura 2.5: Curvas  $\alpha$ -complexidade-entropia para diferentes mapas caóticos. Os símbolos de sinais positivos e círculos abertos representam início e fim de cada curva, respectivamente. Os símbolos em forma de estrela indicam os pares associados com a entropia e complexidade usuais quando  $\alpha$  tende a 1. Figura extraída e adaptada de [8].

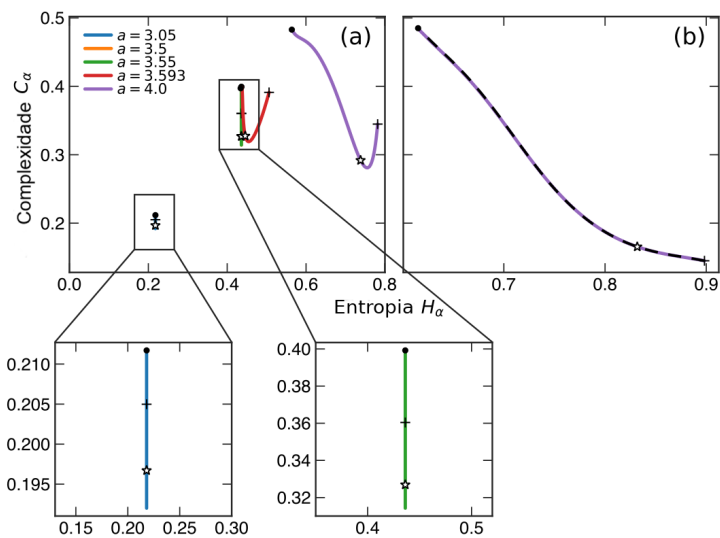


Figura 2.6: Curvas  $\alpha$ -complexidade-entropia para séries temporais do mapa logístico com *embedding dimension*  $d = 4$ . Figura extraída e adaptada de [8].

Baseando-se no fato de apresentarem curvas  $q$ -complexidade-entropia abertas ou fechadas, Ribeiro *et al.* verificaram que essa característica pode ser útil na distinção entre séries temporais caóticas e estocásticas. No caso de séries temporais estocásticas, devido ao fato de não existir padrões ordinais proibidos em sua dinâmica, as curvas são fechadas e o valor da área limitada pela curva está diretamente relacionada à aleatoriedade do processo gerador.

Sendo assim, com relação ao plano complexidade-entropia de Tsallis, para processos estocásticos a área da curva no plano complexidade-entropia é um potencial quantificador da aleatoriedade do sistema. No que tange aos processos caóticos nos quais as curvas formadas não são fechadas, iremos quantificar o papel do *embedding delay* através do comprimento da curva. Da mesma forma, no caso das curvas  $\alpha$ -complexidade-entropia de Rényi, nas quais as curvas nunca são fechadas, será avaliado o comprimento da curva. A ideia é que mesmo nos casos em que o processo gerador da série seja pouco aleatório ou exista certo padrão em seus valores consecutivos, ao variarmos o *embedding delay*, seremos capazes de capturar pontos que tem pouca ou quase nenhuma relação, aumentando ainda mais a aleatoriedade da série, e portanto diminuindo a área ou o comprimento das curvas no plano complexidade-entropia.

Jauregui *et al.* também utilizaram curvas de complexidade-entropia para distinguir séries temporais, mas com a entropia paramétrica  $\alpha$  de Rényi. Essas curvas, como já foi citado, nunca são fechadas. Eles mostram que essas curvas permitem a diferenciação entre séries temporais caóticas, estocásticas e de natureza periódica. Em particular, séries temporais de natureza estocástica são associadas à curvas que exibem uma curvatura positiva na vizinhança de seus pontos iniciais, ao passo que curvas relacionadas a fenômenos caóticos têm uma curvatura negativa. Eles estudam isso a partir do sinal das derivadas  $dC_\alpha/dH_\alpha$  para os gráficos da figura 2.5. Por fim, eles mostram que séries temporais periódicas são representadas por linhas retas verticais, como visto pelo gráfico da figura 2.6.

Assim para cada curva realizada pela variação dos parâmetros  $q$  (para a entropia de Tsallis) e  $\alpha$  (para a entropia de Rényi), ao invés de variarmos o *embedding dimension* como ambos os trabalhos fizeram, iremos variar o *embedding delay* a fim de capturarmos o papel desse parâmetro nas curvas complexidade-entropia de algumas séries temporais.

# Capítulo 3

## Análise das curvas complexidade-entropia via *embedding delay*

### 3.1 Família de curvas complexidade-entropia

Como já foi visto, no contexto da entropia de permutação, uma série de  $n$  elementos

$$\{x_1, x_2, \dots, x_n\} = \{x_t\}_{t=1,2,\dots,n},$$

produz  $n - (d - 1)\tau$  partições representadas por

$$(\vec{s}) = \{x_{s-(d-1)\tau}, x_{s-(d-2)\tau}, \dots, x_{s-\tau}, x_s\}.$$

O parâmetro *embedding delay* acessa diferentes escalas temporais das séries temporais porque ele nos permite selecionar elementos não consecutivos da série, isto é, para cada valor de  $\tau$  podemos imaginar que uma nova série é construída a partir dos elementos espaçados por  $\tau$  unidades de tempo da série original. De forma mais específica, isso significa que para uma mesma série, a variação desse parâmetro produz uma família de curvas no plano complexidade-entropia. No caso do plano  $q$ -complexidade-entropia, por exemplo, cada valor de  $\tau$  produz uma curva  $(H_q, C_q)$  diferente.

No contexto das curvas complexidade-entropia, para estudar quantitativamente a influência desse parâmetro no comportamento de séries temporais, iremos obter medidas de comprimento (no caso de curvas abertas) e área (no caso de curvas fechadas).

## 3.2 Sinal periódico com ruído gaussiano

Acessar a dinâmica temporal das séries pode ser útil, por exemplo, para encontrar um possível período de séries estocásticas. A ideia é que ao variar o *embedding delay*, se o valor desse parâmetro coincidir com o período da série, a dinâmica da série temporal será essencialmente a de um ruído branco e, assim, os valores das entropias aumentam e os da complexidade diminuem. Desse modo, as curvas  $q$ -complexidade-entropia devem apresentar uma área mínima, enquanto as curvas  $\alpha$ -complexidade-entropia devem apresentar um comprimento mínimo. Essa ideia também pode ser válida para séries temporais mais complexas que apresentem um caráter oscilatório. Nesses casos, analisando as curvas complexidade-entropia com um conjunto de valores para o *embedding delay*, podemos induzir que a série tem um comportamento periódico quando a área dessas curvas (no caso de Tsallis) ou o comprimento (no caso de Rényi) apresentem um comportamento oscilatório em função do *embedding delay*.

Para investigar essa possibilidade de maneira empírica, utilizamos uma série temporal composta pelos valores da função seno somada a um número aleatório gaussiano  $\xi$  de média nula e desvio padrão  $\sigma$ , conforme ilustra a figura 3.1 para  $\sigma = 0$  (à esquerda) e  $\sigma = 0,75$  (à direita).

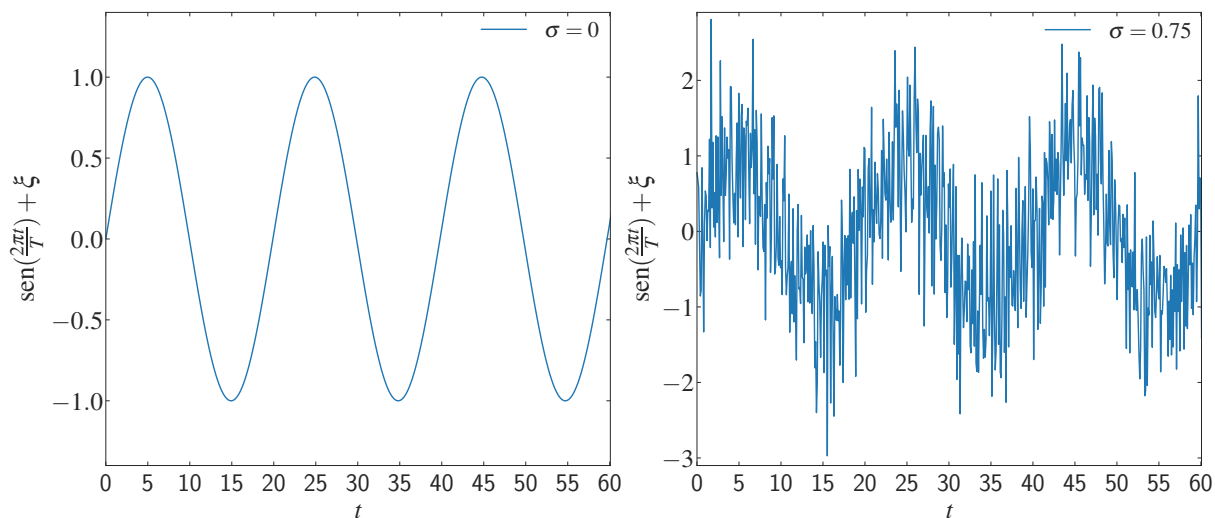


Figura 3.1: Sinal periódico na forma  $\text{sen}(\frac{2\pi t}{T})$  com período  $T = 20$  adicionado um ruído gaussiano de média nula e desvio padrão  $\sigma$  (como indicado no gráfico).

A figura 3.2 mostra exemplos de curvas  $q$ -complexidade-entropia para esse tipo de sinal periódico com  $T = 20$ ,  $\sigma = 0,75$ ,  $d = 3$  e diferentes valores para o *embedding delay*  $\tau$ . Notamos que a área com  $\tau = 20$  é muito menor do que as áreas para outros valores de  $\tau$ , confirmando nossa hipótese inicial. Para estudar o comportamento da área  $A$  em função de  $\tau$  de modo mais sistemático, construímos um conjunto de curvas  $q$ -complexidade-entropia para  $\tau \in \{1, 2, \dots, 20\}$ , estimando a área para cada curva. A figura 3.3(a) mostra o comportamento da área em função do *embedding delay*  $\tau$ .

Observamos que existem muitos valores de  $\tau$  para os quais a área é aproximadamente nula, não apenas quando o valor de  $\tau$  é um múltiplo inteiro do período da série. Entretanto, a variação dos valores de área é muito menor ao redor dos valores de  $\tau$  que são múltiplos inteiros do período da série. A figura 3.3(b) mostra a derivada do inverso da área com relação ao *embedding delay*, na qual observamos picos bem definidos quando  $\tau$  é um múltiplo do período  $T$  da série temporal. Sendo assim, uma análise como essa seria capaz de encontrar o período de um sinal ruidoso periódico.

Realizamos uma análise similar para as curvas  $\alpha$ -complexidade-entropia. Em particular, a figura 3.4 mostra essas curvas para diferentes valores de  $\tau$ . Observamos que para  $\tau = 20$  o comprimento da curva é muito menor do que o observado para os demais casos. Para estudar o comportamento dessas curvas em função de  $\tau$ , calculamos os seus comprimentos para um conjunto de valores de  $\tau$ , conforme mostra a figura 3.5(a). Calculamos também a derivada do inverso do comprimento com relação a  $\tau$ , como mostra a figura 3.5(b). Nessa figura, observamos que essa derivada apresenta picos bem definidos para valores de  $\tau$  que são múltiplos do período  $T$  da série temporal. Desse modo, analisando essa característica das curvas  $\alpha$ -complexidade-entropia também podemos inferir o período da série temporal.

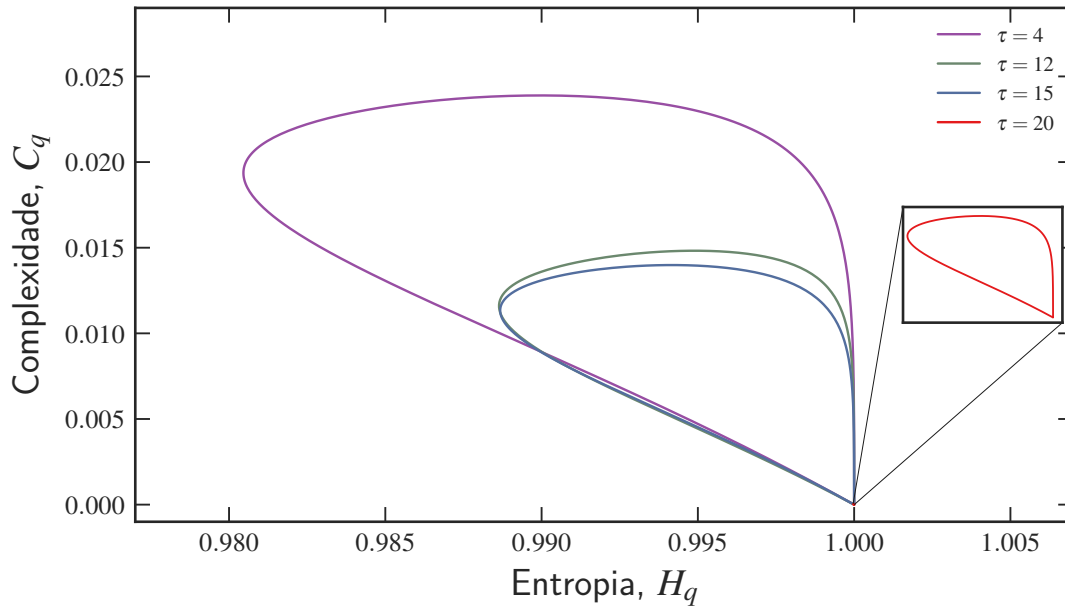


Figura 3.2: Curvas  $q$ -complexidade-entropia de um sinal periódico ruidoso ( $\sigma = 0,75$  e  $T = 20$ ) para diferentes valores do *embedding delay*  $\tau$ , conforme indicado no gráfico.

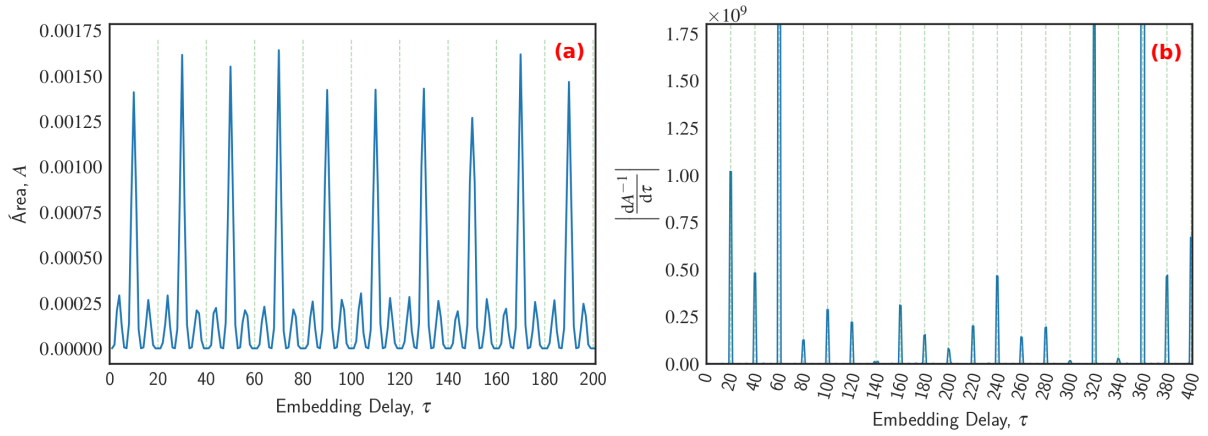


Figura 3.3: (a) Valor da área  $A$  das curvas  $q$ -complexidade-entropia em função do *embedding delay*  $\tau$  para um sinal periódico com ruído gaussiano ( $\sigma = 0,75$  e  $T = 20$ ). (b) Módulo da derivada do inverso da área em função de  $\tau$ . Notamos que existem picos consecutivos quando os valores de  $\tau$  coincidem com múltiplos do período  $T$  da série temporal (indicados por linhas verticais tracejadas).

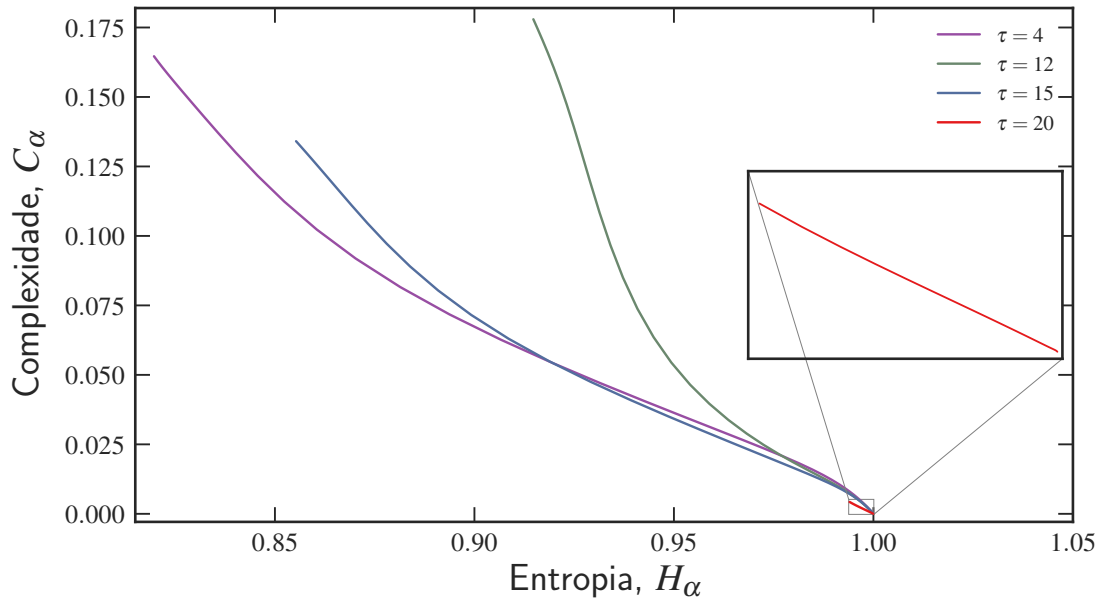


Figura 3.4: Curvas  $\alpha$ -complexidade-entropia de um sinal periódico ruidoso ( $\sigma = 0,75$  e  $T = 20$ ) para diferentes valores do *embedding delay*  $\tau$ , conforme indicado no gráfico.

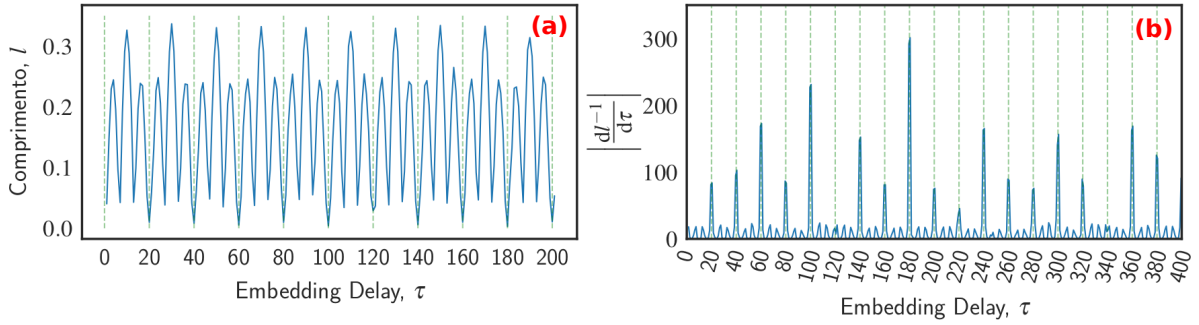


Figura 3.5: (a) Valor do comprimento  $l$  das curvas  $\alpha$ -complexidade-entropia em função do *embedding delay*  $\tau$  para um sinal periódico com ruído gaussiano. (b) Módulo da derivada do inverso do comprimento em função de  $\tau$ . Notamos que existem picos consecutivos quando os valores de  $\tau$  coincidem com múltiplos do período  $T$  da série temporal (indicado por linhas verticais tracejadas).

### 3.3 Ruído harmônico

Para uma série temporal periódica simples como a estudada na seção anterior, observamos que as derivadas do inverso da área (no caso de Tsallis) e do inverso do comprimento (no caso de Rényi) em função de  $\tau$  podem identificar o período dessas séries. Seria interessante, entretanto, se essa análise também revelasse o período de séries temporais mais complexas de natureza periódica. É nesse sentido que estudamos o ruído harmônico (Apêndice A.1.2). Um exemplo dessa série temporal é mostrado na figura 3.6. Essa série tem um período igual a  $\pi/5$ , embora sua identificação visual seja mais complicada analisando apenas a figura 3.6.

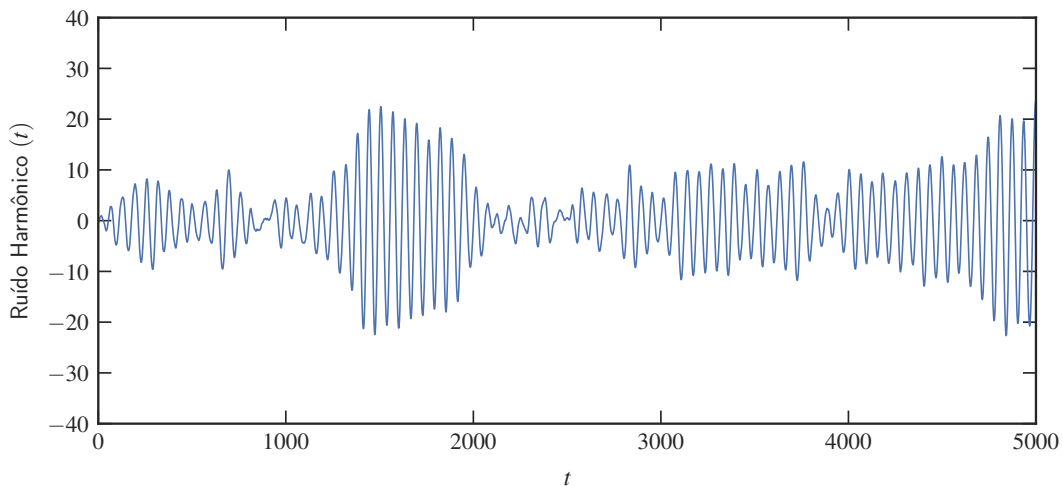


Figura 3.6: Ruído harmônico (Apêndice A.1.2) com período  $\pi/5$ .

Com relação às curvas complexidade-entropia de Tsallis para o ruído harmônico, pode-se notar pela figura 3.7, que para  $\tau = 1000\pi/5$  e  $\tau = 2000\pi/5$ , múltiplos do período da

série temporal, as áreas formadas nesse plano também diminuem.

Os valores das áreas dos gráficos em função do *embedding delay* se encontram na figura 3.8(a). Novamente, à primeira vista, existem diversos valores do parâmetro  $\tau$  que resultam na área ser aproximadamente nula, não somente quando o parâmetro  $\tau$  coincide com um múltiplo do período. Porém, ao calcularmos a derivada da área a fim de ressaltarmos bruscas variações, notamos picos de variações bruscas nos valores em que o parâmetro coincide com um múltiplo do período da série, como mostra a figura 3.8(b). Diferente dos mesmos gráficos para o sinal periódico com ruído, o caso do ruído harmônico apresenta um decaimento característico ao processo que gera essa série temporal. Esse gráfico, portanto, também pode nos fornecer o valor do período da função a ser analisada, basta estimar os valores consecutivos nos quais a derivada da área em relação ao *embedding delay* possui as variações mais bruscas.

A mesma análise foi realizada para as curvas complexidade-entropia de Rényi, mostrado na figura 3.9. Observamos, novamente, que quando o *embedding delay* assume valores próximos a múltiplos inteiros do período da série temporal, os comprimentos das curvas diminuem.

A figura 3.10(a) mostra o comprimento dessas curvas em função do *embedding delay*. Nesse caso podemos notar que quando o valor de  $\tau$  é um múltiplo do período da série, o valor do comprimento da curva sofre uma brusca variação. Para realçar esse comportamento, a figura 3.10(b) mostra o módulo da derivada do comprimento em relação ao *embedding delay*, na qual observamos picos no valor da derivada quando  $\tau$  é um múltiplo inteiro do período da série. Sendo assim, também é possível encontrar o período da série a partir de uma análise como essa.



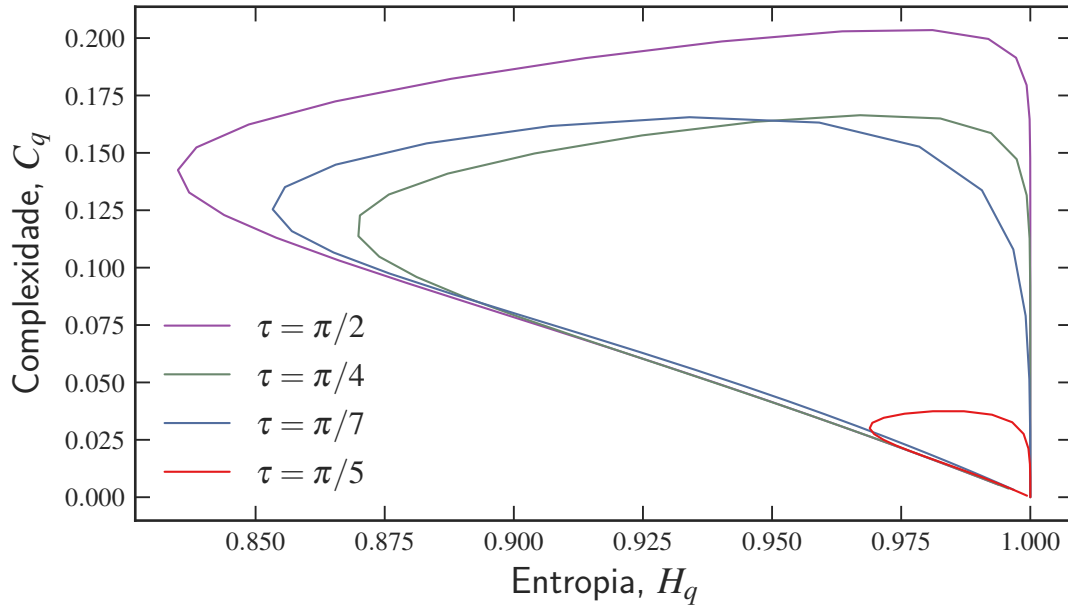


Figura 3.7: Curvas complexidade-entropia de Tsallis para o ruído harmônico com período  $\pi/5$  para diferentes valores do *embedding delay*  $\tau$ , como indicado no gráfico. Notamos que a área tende a diminuir quando  $\tau$  se aproxima de múltiplos do período da série.

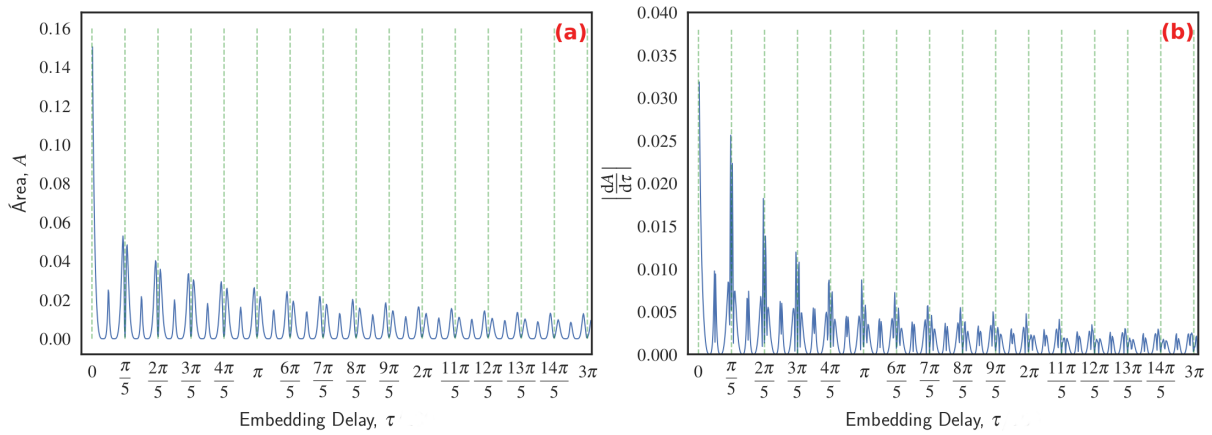


Figura 3.8: (a) Valor da área da curva  $q$ -complexidade-entropia em função do *embedding delay* para o ruído harmônico com período  $\pi/5$ . (b) Módulo da derivada da área em função de  $\tau$ . Podemos notar que há variações bruscas consecutivas quando  $\tau$  coincide com múltiplos do período da série. Além disso, observamos um decaimento nos valores da área de sua derivada com o aumento de  $\tau$ .

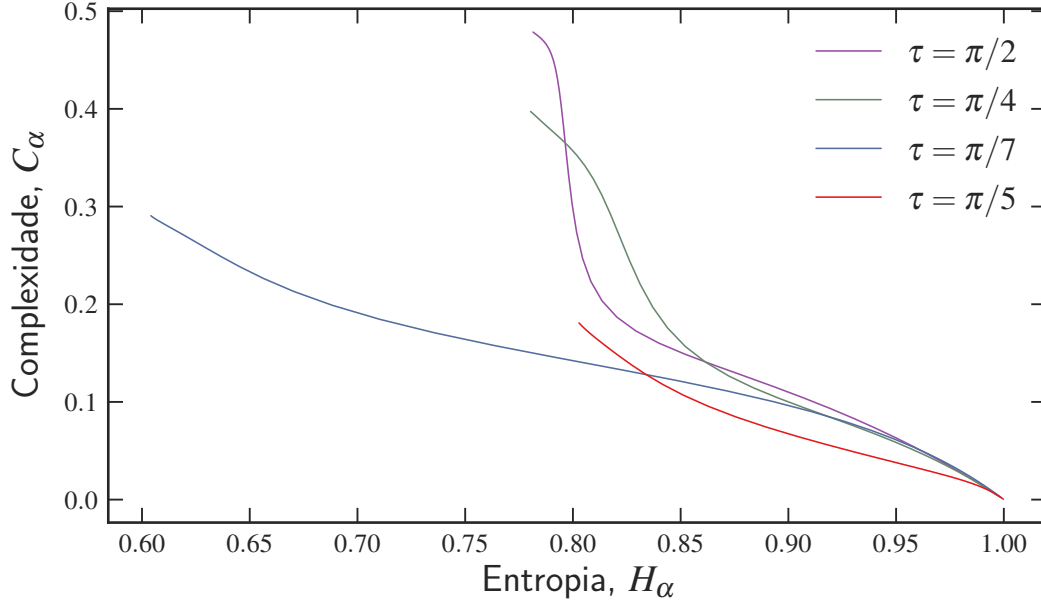


Figura 3.9: Curvas  $\alpha$ -complexidade-entropia para o ruído harmônico com período  $\pi/5$  para diferentes valores do *embedding delay* (indicado no gráfico).

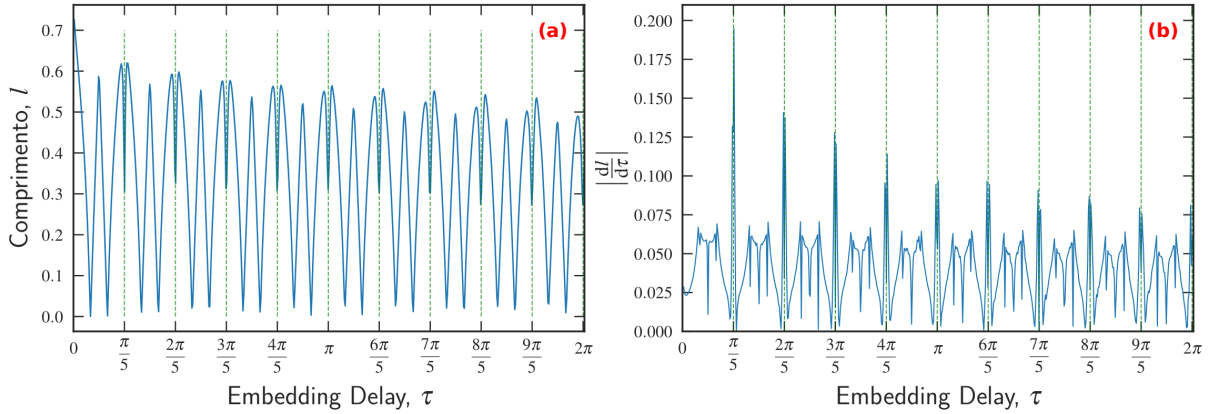


Figura 3.10: (a) Valor do comprimento da curva  $\alpha$ -complexidade-entropia em função do *embedding delay* para o ruído harmônico com período  $\pi/5$ . (b) Módulo da derivada do comprimento em função de  $\tau$ . Podemos notar que há variações bruscas consecutivas quando  $\tau$  coincide com múltiplos do período da série. Além disso, observamos um decaimento nos valores do comprimento de sua derivada com o aumento de  $\tau$ .

### 3.4 Séries caóticas

Nessa seção, apresentamos alguns resultados das curvas complexidade-entropia para séries temporais de natureza caótica. Para isso, geramos séries temporais de 8 mapas caóticos com parâmetros escolhidos de modo que todos eles estivessem operando no regime de caos completo. Esses mapas são o logístico, o de Burgers, o cúbico, o de Gingerbreadman,

o de Hénon, o de Ricker, o do seno e o de Tinkerbell. As definições de cada um, bem como os parâmetros que utilizamos para gerar as séries temporais são apresentadas no Apêndice A.2.

Iniciamos pelo cálculo das curvas  $q$ -complexidade-entropia para cada um dos 8 mapas para alguns valores do *embedding delay*  $\tau$ , conforme mostra a figura 3.11. Observamos que a maioria dos mapas apresenta uma curva fechada para essa *embedding dimension*  $d = 3$ . Apenas o mapa logístico e o mapa de Ricker são descritos por curvas abertas. Com relação ao aumento do *embedding delay*  $\tau$ , notamos que a maioria das curvas diminuem de comprimento ou até colapsam praticamente em um ponto (como no mapa logístico). Na tentativa de compreender melhor esse comportamento, estudamos como o comprimento das curvas  $q$ -complexidade-entropia varia ao aumentarmos o valor de  $\tau$ , conforme mostra a figura 3.12. Nessa figura observamos uma tendência geral de diminuição do comprimento com  $\tau$ , ainda que algumas flutuações consideráveis surjam para alguns mapas (por exemplo, no caso do mapa Gingerbreadman). Nenhum comportamento periódico é observado nessas curvas, uma vez que as séries temporais são de origem completamente caóticas.

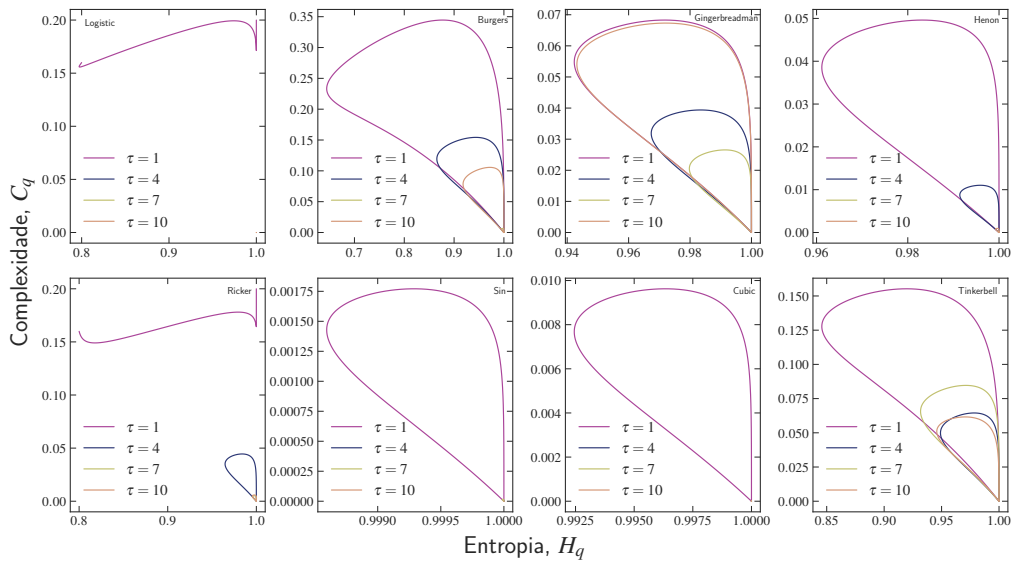


Figura 3.11: Curvas  $q$ -complexidade-entropia para séries temporais caóticas. Cada painel mostra as curvas correspondentes aos valores de  $\tau \in \{1, 2, \dots, 10\}$ , com *embedding dimension*  $d = 3$ .

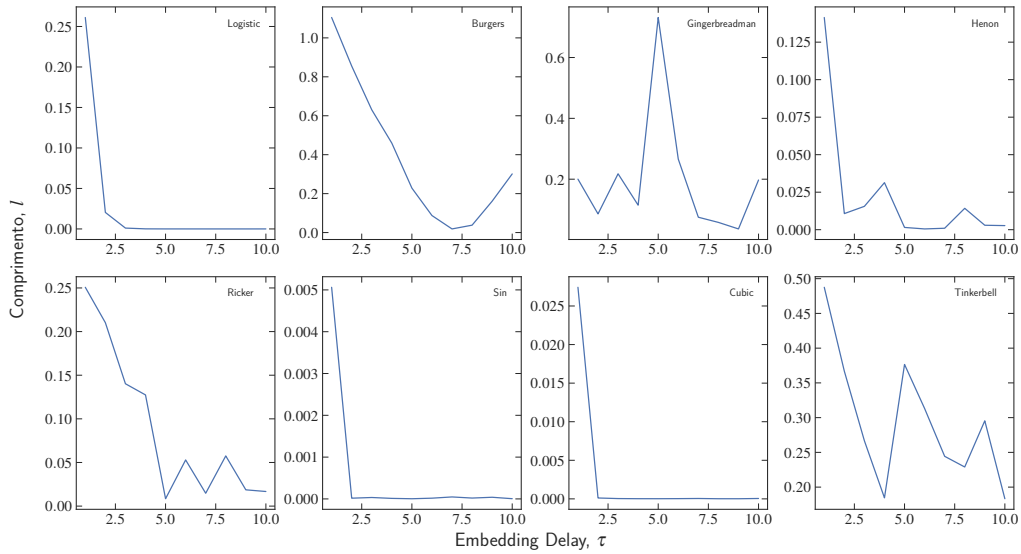


Figura 3.12: Valor do comprimento das curvas  $q$ -complexidade-entropia para séries temporais caóticas. Cada painel mostra o comprimento das curvas resultantes em função de  $\tau \in \{1, 2, \dots, 10\}$  para  $d = 3$ .

Já no cálculo das curvas  $\alpha$ -complexidade-entropia, utilizamos alguns valores do *embedding delay*  $\tau$  para cada um dos 8 mapas, conforme mostra a figura 3.13. Nesse plano todos dos mapas apresentam curvas abertas. Da mesma forma que anteriormente, com relação ao aumento do *embedding delay*, notamos que a maioria das curvas diminuem de comprimento ou até colapsam praticamente em um ponto (como no mapa logístico). Para compreender esse processo mais quantitativamente, calculamos como o comprimento das curvas  $\alpha$ -complexidade-entropia varia ao aumentarmos o valor de  $\tau$ , conforme mostra a figura 3.14. Nessa figura observamos uma tendência geral de diminuição do comprimento com  $\tau$ , ainda que algumas flutuações (parecidas com o caso do plano  $q$ -complexidade-entropia) consideráveis surjam para alguns mapas (por exemplo, no caso do mapa Gingerbreadman). Novamente, nenhum comportamento periódico é observado nessas curvas, uma vez que as séries temporais são de origem completamente caóticas.

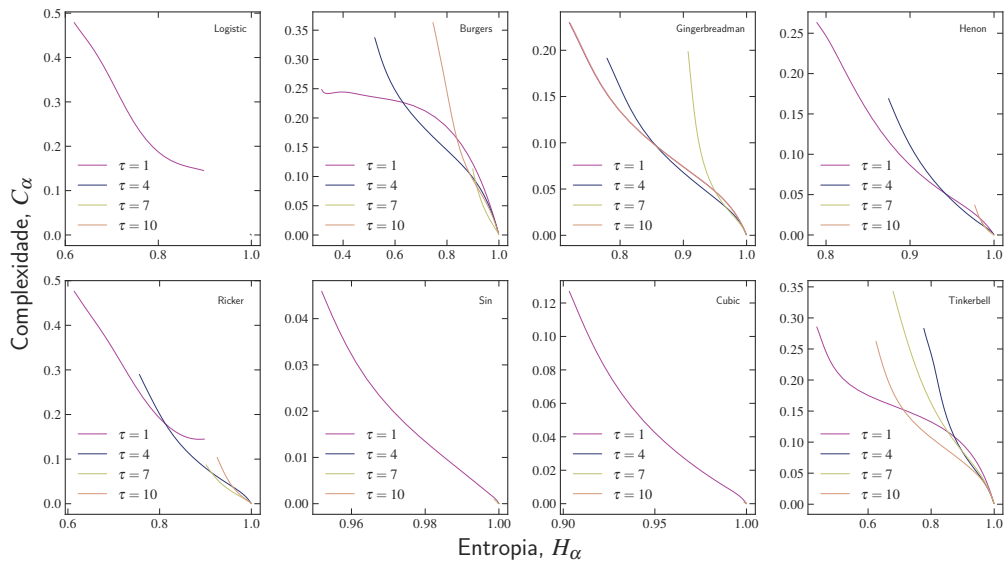


Figura 3.13: Curvas  $\alpha$ -complexidade-entropia para séries temporais caóticas. Cada painel mostra as curvas correspondentes aos valores de  $\tau \in \{1, 2, \dots, 10\}$ , com *embedding dimension*  $d = 3$ .

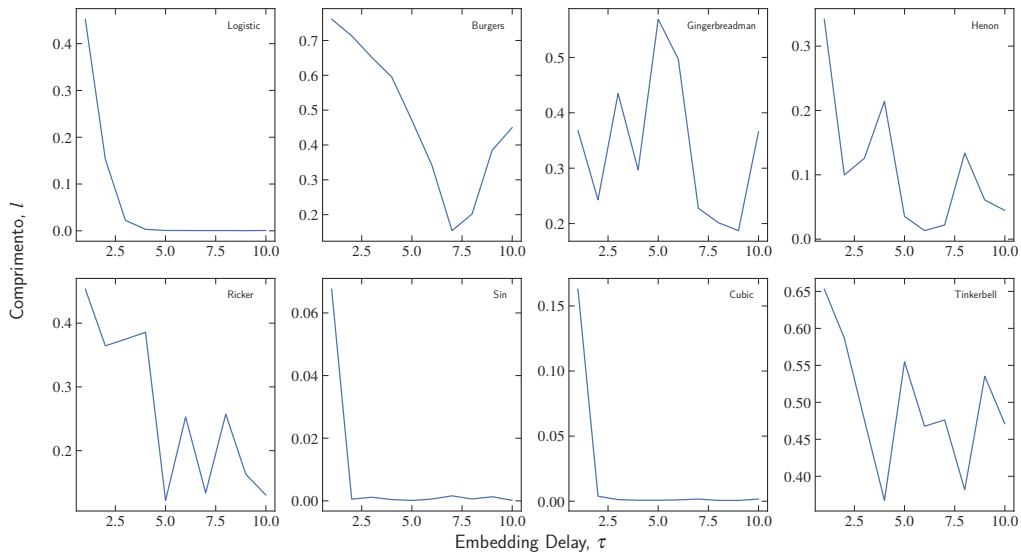


Figura 3.14: Valor do comprimento das curvas  $\alpha$ -complexidade-entropia para séries temporais caóticas. Cada painel mostra o comprimento das curvas resultantes em função de  $\tau$  para  $d = 3$ .

# Conclusões

Esse trabalho apresentou uma generalização da abordagem das curvas de complexidade-entropia por meio do parâmetro *embedding delay*. O estudo foi realizado com base na entropia de Shannon, medida essa que nos permite calcular a quantidade de informação contida em um sistema. No entanto, esse conceito por si só não se mostra suficiente para descrever o grau de estrutura presente nos sistemas. Vimos que aliada ao conceito de entropia, a ideia de complexidade definida para quantificar a estrutura do sistema nos permite uma melhor descrição do mesmo. Com o intuito de facilitar o cálculo da complexidade foi utilizado o método de Bandt Pompe, uma abordagem simples e computacionalmente rápida para qualquer série temporal. Além disso, vimos que usando essas medidas podemos criar curvas complexidade-entropia que podem ser útil no estudo de séries temporais. Entretanto, para realçar melhor as características entrópicas dos sistemas, o estudo foi realizado a partir de duas definições, que são generalizações da entropia de Shannon: as entropias monoparamétricas  $q$  de Tsallis e  $\alpha$  de Rényi. Com essas generalizações criamos curvas  $q$ -complexidade-entropia e  $\alpha$ -complexidade-entropia que são úteis no estudo e na caracterização de séries temporais. Nesse contexto, usamos o parâmetro *embedding delay* que é capaz de acessar a dinâmica temporal dessas séries e notamos que variação desse parâmetro cria uma família de curvas complexidade-entropia. A análise dessa família de curvas, por sua vez, evidencia o papel desse parâmetro nas séries temporais. No caso da série temporal periódica com ruído gaussiano, por exemplo, observamos que quando esse parâmetro coincide com o período da série, as curvas fechadas formadas no plano  $q$ -complexidade-entropia se colapsam a uma pequena área. Observamos que isso poderia ser útil para encontrar o período de outras séries temporais estocásticas que tenham natureza periódica. Estudamos então a variação desse parâmetro no caso do ruído harmônico e notamos que a área da curva diminui quando o *embedding delay* coincide com o período da série. Já para o caso de séries temporais caóticas, mostramos que há uma tendência de diminuição da área e comprimento das curvas ao aumentarmos o valor do *embedding delay* para ambas as famílias de curvas. Isso condiz com o fato de que elementos mais espaçados dessas séries temporais possuem menos correlação, causando um aumento na entropia e uma diminuição na complexidade, diminuindo assim o comprimento e área das curvas. Pretendemos explorar essa família de curvas complexidade-entropia para estudar outras séries temporais sintéticas e também séries empíricas. Acreditamos que essa família de

curvas pode ser útil para o processo de classificação de séries temporais, podendo inclusive melhorar a performance dessas classificações, uma vez que os valores  $\tau$  produzem mais informações sobre a dinâmica das séries temporais.

# Apêndice A - Séries Temporais

## A.1 Séries temporais estocásticas

### A.1.1 Sinal periódico com ruído

Definimos um sinal periódico com ruído usando

$$f(t) = \sin\left(\frac{2\pi t}{T}\right) + \xi, \quad (\text{A.1})$$

na qual  $\xi$  é um número aleatório gaussiano de média nula, desvio padrão  $\sigma$  e  $T$  é o período da série.

### A.1.2 Ruído harmônico

O ruído harmônico [14] é um processo estocástico definido pelo seguinte sistema de equações de Langevin:

$$\begin{aligned} \dot{y} &= s, \\ \dot{s} &= -\Gamma s - \Omega^2 y + \sqrt{2\varepsilon}\Omega^2 \xi(t), \end{aligned} \quad (\text{A.2})$$

na qual  $\Omega$  é a frequência de oscilação,  $\Gamma$  e  $\varepsilon$  são constantes positivas e  $\xi(t)$  é o ruído Gaussiano com média zero (isto é,  $\langle \xi(t) \rangle = 0$ ) sem correlação  $\langle \xi(t)\xi(t') \rangle = \delta(t - t')$ . Trata-se da equação de movimento de um oscilador harmônico amortecido e forçado por uma força aleatória, sendo

$$T = \frac{2\pi}{\sqrt{\Omega^2 - (\Gamma/2)^2}} \quad (\text{A.3})$$

o período de oscilação. Para produzir séries temporais a partir do ruído harmônico integra-se o sistema de equações (A.2) usando o método de Euler. O tamanho dos passos usado foi  $dt = 0,01$ , com tempo máximo de integração de 300 unidades e valores de  $\Gamma = 0,01$ ,  $\Omega = 10$  e  $\varepsilon = 0,1$ , os quais produzem um período  $T = \pi/5$ .



## A.2 Séries Temporais Caóticas

### A.2.1 Mapa logístico

O mapa de logístico é definido como [15], [7]

$$x_{k+1} = ax_k(1 - x_k), \quad (\text{A.4})$$

em que  $a$  é um parâmetro do mapa e os valores de  $x_k$  são obtidos iterando essa equação a partir de uma condição inicial  $x_0$ . Em nossas análises usamos  $x_0 = 0,1$  e  $a = 4$ , sendo essa última condição usada para gerar séries temporais no regime completamente caótico.

### A.2.2 Mapa de Burgers

O mapa de Burgers é definido como [15], [7]

$$\begin{aligned} x_{k+1} &= ax_k - y_k^2, \\ y_{k+1} &= by_k + x_k y_k. \end{aligned} \quad (\text{A.5})$$

O mapa é bidimensional e, por isso, a série temporal analisada é  $(x_k + y_k)^2$ . Os valores de  $x_k$  e  $y_k$  são obtidos pelas condições iniciais  $x_0 = -0,1$  e  $y_0 = 0,1$ , respectivamente. As constantes  $a$  e  $b$  são parâmetros do mapa e em nossas análises usamos  $a = 0,75$  e  $b = 1,75$ .

### A.2.3 Mapa cúbico

O mapa cúbico é definido como [15], [7]

$$x_{k+1} = ax_k(1 - x_k^2), \quad (\text{A.6})$$

em que  $a$  é um parâmetro do mapa e os valores de  $x_k$  são obtidos iterando essa equação a partir de uma condição inicial  $x_0$ . Em nossas análises usamos  $x_0 = 0,1$  e  $a = 3$ .

### A.2.4 Mapa Gingerbreadman

O mapa Gingerbreadman é definido como [15], [7]

$$\begin{aligned} x_{k+1} &= 1 - y_k + |x_k|, \\ y_{k+1} &= x_k. \end{aligned} \quad (\text{A.7})$$

O mapa é bidimensional e, por isso, a série temporal analisada é  $(x_k + y_k)^2$ . Os valores de  $x_k$  e  $y_k$  são obtidos pelas condições iniciais  $x_0 = 0,5$  e  $y_0 = 3,7$ , respectivamente.

### A.2.5 Mapa de Hénon

O mapa de Hénon é definido como [15], [7]

$$\begin{aligned}x_{k+1} &= 1 - ax_k^2 + y_k, \\y_{k+1} &= bx_k.\end{aligned}\tag{A.8}$$

O mapa é bidimensional e, por isso, a série temporal analisada é  $(x_k + y_k)^2$ . Os valores de  $x_k$  e  $y_k$  são obtidos pelas condições iniciais  $x_0 = 0$  e  $y_0 = 0,9$ , respectivamente. As constantes  $a$  e  $b$  são parâmetros do mapa e em nossas análises usamos  $a = 1,4$  e  $b = 0,3$ .

### A.2.6 Mapa de Ricker

O mapa de Ricker é definido como [15], [7]

$$x_{k+1} = ax_k e^{-x_k},\tag{A.9}$$

em que  $a$  é um parâmetro do mapa e os valores de  $x_k$  são obtidos iterando essa equação a partir de uma condição inicial  $x_0$ . Em nossas análises usamos  $x_0 = 0,1$  e  $a = 20$ .

### A.2.7 Mapa seno

O mapa Seno é definido como [15], [7]

$$x_{k+1} = a \sin(\pi x_k),\tag{A.10}$$

em que  $a$  é um parâmetro do mapa e os valores de  $x_k$  são obtidos iterando essa equação a partir de uma condição inicial  $x_0$ . Em nossas análises usamos  $x_0 = 0,1$  e  $a = 0,1$ .

### A.2.8 Mapa Tinkerbell

O mapa de Tinkerbell é definido como [15], [7]

$$\begin{aligned}x_{k+1} &= x_k^2 - y_k^2 + ax_k + by_k, \\y_{k+1} &= 2x_k y_k + cx_k + dy_k.\end{aligned}\tag{A.11}$$

O mapa é bidimensional e, por isso, a série temporal analisada é  $(x_k + y_k)^2$ . Os valores de  $x_k$  e  $y_k$  são obtidos pelas condições iniciais  $x_0 = -0,1$  e  $y_0 = 0,1$ , respectivamente. As constantes  $a$ ,  $b$ ,  $c$  e  $d$  são parâmetros do mapa e em nossas análises usamos  $a = 0,1$ ,  $b = -0,6$ ,  $c = 2,0$  e  $d = 0,5$ .

# Referências Bibliográficas

- [1] C. E. Shannon, “A mathematical theory of communication,” *The Bell System Technical Journal*, vol. 27, no. 3, pp. 379–423, 1948.
- [2] R. López-Ruiz, H. L. Mancini, and X. Calbet, “A statistical measure of complexity,” *Physics Letters A*, vol. 209, no. 5, pp. 321–326, 1995.
- [3] L. Zunino, M. C. Soriano, I. Fischer, O. A. Rosso, and C. R. Mirasso, “Permutation-information-theory approach to unveil delay dynamics from time-series analysis,” *Physical Review E*, vol. 82, no. 4, p. 046212, 2010.
- [4] C. Bandt and B. Pompe, “Permutation entropy: a natural complexity measure for time series,” *Physical Review Letters*, vol. 88, no. 17, p. 174102, 2002.
- [5] C. Tsallis, “Possible generalization of Boltzmann-Gibbs statistics,” *Journal of Statistical Physics*, vol. 52, no. 1-2, pp. 479–487, 1988.
- [6] A. Rényi, *Probability Theory*. North-Holland Publishing Company, Amsterdam, Springer, New York, 2009.
- [7] H. V. Ribeiro, M. Jauregui, L. Zunino, and E. K. Lenzi, “Characterizing time series via complexity-entropy curves,” *Physical Review E*, vol. 95, no. 6, p. 062106, 2017.
- [8] M. Jauregui, L. Zunino, E. K. Lenzi, R. S. Mendes, and H. V. Ribeiro, “Characterization of time series via Rényi complexity-entropy curves,” *Physica A: Statistical Mechanics and its Applications*, vol. 498, pp. 74–85, 2018.
- [9] O. A. Rosso, H. A. Larrondo, M. T. Martin, A. Plastino, and M. A. Fuentes, “Distinguishing noise from chaos,” *Physical Review Letters*, vol. 99, no. 15, p. 154102, 2007.
- [10] Y. Cao, W.-w. Tung, J. B. Gao, V. A. Protopopescu, and L. M. Hively, “Detecting dynamical changes in time series using the permutation entropy,” *Physical Review E*, vol. 70, no. 4, p. 046217, 2004.

- [11] L. Zunino, M. Zanin, B. M. Tabak, D. G. Pérez, and O. A. Rosso, “Forbidden patterns, permutation entropy and stock market inefficiency,” *Physica A: Statistical Mechanics and its Applications*, vol. 388, no. 14, pp. 2854–2864, 2009.
- [12] L. R. Rabiner and R. W. Schafer, *Digital processing of speech signals*. Englewood Cliffs, N.J.: Prentice-Hall, 1978.
- [13] Higor Sigaki, *Física Estatística Aplicada ao Estudo de Obras de Arte*. Dissertação de Mestrado, Universidade Estadual de Maringá, Maringá, 2017.
- [14] L. Schimansky-Geier and C. Zülicke, “Harmonic noise: Effect on bistable systems,” *Zeitschrift für Physik B Condensed Matter*, vol. 79, no. 3, pp. 451–460, 1990.
- [15] J. Amigó, *Permutation Complexity in Dynamical Systems: Ordinal Patterns, Permutation Entropy and All That*. Springer Series in Synergetics, Berlin Heidelberg: Springer-Verlag, 2010.