



Trabalho de Conclusão de Curso

**Teoria da Informação e Estatística Computacional
no Processamento e Análise de Sinais – Uma
ferramenta para Análise de Séries Temporais**

Eduarda Tatiane Caetano Chagas
eduardachagas48@laccan.ufal.br

Orientador:
Prof. Dr. Alejandro Cesar Frery Orgambide

Maceió, Agosto de 2018

Eduarda Tatiane Caetano Chagas

**Teoria da Informação e Estatística Computacional
no Processamento e Análise de Sinais – Uma
ferramenta para Análise de Séries Temporais**

Monografia apresentada como requisito parcial
para obtenção do grau de Bacharel em Ciência
da Computação do Instituto de Computação da
Universidade Federal de Alagoas.

Orientador:

Prof. Dr. Alejandro Cesar Frery Orgambide

Maceió, Agosto de 2018

Monografia apresentada como requisito parcial para obtenção do grau de Bacharel em Ciência da Computação do Instituto de Computação da Universidade Federal de Alagoas, aprovada pela comissão examinadora que abaixo assina.

Prof. Dr. Alejandro Cesar Frery Orgambide - Orientador
Instituto de Computação
Universidade Federal de Alagoas

Prof. Dr. Osvaldo Anibal Rosso - Examinador
Instituto de Física
Universidade Federal de Alagoas

Prof. Dr. Heitor Soares Ramos Filho - Examinador
Instituto de Computação
Universidade Federal de Alagoas

Agradecimentos

Gostaria de agradecer primeiramente a minha mãe e meus avós, por todo o sacrifício realizado, mesmo diante de dificuldades, para sempre garantir uma educação de qualidade para mim e minha irmã.

Aos meus amigos, Glauber Braga, Pedro Almeida, Bruno Rafael e David Alexandre, que me apoiaram e suportaram todos as minhas crises e desabafos.

Ao Prof. Dr. Alejandro Frery, que se tornou não somente meu orientador, como também a minha maior inspiração de vida, me ensinando conhecimentos que levarei para toda o sempre.

À todas as meninas do Instituto de Computação, em especial as minhas amigas Vanessa Vieira, Ana Geórgia e Júlia Albuquerque, por me ensinarem que mulheres podem ser comuns neste meio da tecnologia e ainda assim permanecerem únicas.

Também agradeço aos membros do Laboratório de Computação Científica e Análise Numérica (LaCCAN) da UFAL por todo conhecimento adquirido ao longo desses anos.

Por fim, e não menos importante, agradeço à banca examinadora, pela leitura atenta, questionamentos e sugestões.

“... é fácil entender que sempre existe no mundo uma pessoa que espera a outra, seja no meio de um deserto, seja no meio das grandes cidades. E quando essas pessoas se cruzam e seus olhos se encontram, todo o passado e todo o futuro perdem qualquer importância; só existe aquele momento e aquela certeza incrível de que todas as coisas debaixo do sol foram escritas pela mesma Mão.”

– Paulo Coelho, *O Alquimista*

Eduarda Chagas

“Sem dor, como poderíamos reconhecer o prazer?”

– John Green, *A culpa é das estrelas*

Resumo

A análise de séries temporais é classicamente feita ou no domínio do tempo ou em algum domínio transformado (Fourier, Wavelet etc.). Mais recentemente, apareceram técnicas não-paramétricas e, dentre elas, a análise de descritores causais. Essas técnicas tem como grande vantagem a relativa pouca sensibilidade a perturbações dos dados, e a capacidade de revelar propriedades importantes da dinâmica subjacente ao processo. A análise dos descritores causais de uma série temporal possui uma ampla aplicabilidade em nossa rotina, por exemplo na análise de ações bancárias, no registro do comportamento da maré, nos índices da taxa de desemprego, nas temperaturas máximas e mínimas diárias de uma cidade, dentre outras incontáveis finalidades. Desse modo, relatamos aqui o processo de desenvolvimento de uma plataforma de análise dos descritores causais de uma série temporal oriundos da Teoria da Informação. A plataforma visa facilitar a análise dessas séries nos mais variados ramos da ciência. O sistema foi implementado na linguagem de programação R que, além de fornecer ferramentas gráficas, também possui uma grande precisão numérica, ambas características de extrema importância ao longo deste trabalho.

Palavras-chave: Séries Temporais; Teoria da Informação; Linguagem R.

Abstract

Time series analysis is classically performed either in the time domain or in a transformed domain (Fourier, Wavelet, etc.) More recently, nonparametric techniques have been proposed and, among them, the use of time causal descriptors. This class of techniques has the ability to reveal important properties of the underlying process and, at the same time, to be relatively insensitive to data contamination. The analysis of causal descriptors of a time series has a wide applicability, as in the analysis stock market, records of the behavior of the tides, index of the unemployment rates, maximum and minimum daily temperatures of a city, among others. We report here the process of developing a platform for analyzing causal descriptors of a time series using Information Theory. The platform aims to facilitate the analysis of such series in as many branches of science as possible. The system was implemented in the \mathbb{R} programming language, which besides providing graphical tools, also has a great numerical precision, both features of extreme importance throughout this work.

Keywords: Time Series; Information Theory; Language \mathbb{R} ;

Conteúdo

Lista de Figuras	vii
Lista de Tabelas	viii
1 Introdução	1
1.1 Motivação	1
1.2 Objetivo	3
1.3 Solução proposta	3
1.4 Contribuições	3
1.5 Estrutura do texto	4
2 Fundamentação Teórica	5
2.1 Representação do espaço de probabilidade	5
2.2 Método de simbolização de Bandt e Pompe	6
2.3 Distribuição de probabilidade de Bandt e Pompe	7
2.4 Entropia de permutação	8
2.5 Distância Estocástica	9
2.6 Complexidade Estatística	10
2.7 Plano Complexidade-Entropia	11
3 Metodologia	13
3.1 Estudo das funções a serem implementadas	13
3.2 Implementação e validação numérica	13
3.3 Análise de alternativas para o desenvolvimento da interface	14
3.4 Desenvolvimento de protótipos	14
3.5 Versão de produção da interface	15
3.6 Validação, verificação e preparação de manuais e tutoriais de uso	15
4 Resultados e Discussões	16
5 Demonstração de uso do Software	19
5.1 Upload de dados	19
5.2 Visualização da série temporal	20
5.3 Histograma da distribuição de Bandt-Pompe	20
5.4 Cálculo da Entropia de Shannon	21
5.5 Cálculo da Complexidade Estatística	21
5.6 Plano Complexidade-Entropia	21
6 Conclusões	24
6.1 Considerações Finais	24
6.2 Trabalhos futuros	24

A Manual de utilização das funções desenvolvidas	26
A.1 Pacotes necessários	26
A.2 Principais funções desenvolvidas	26
Referências bibliográficas	44

Lista de Figuras

2.1	Representação gráfica dos padrões com dimensão $D = 3$	6
2.2	Mapeamento por Permutação de Classificação (Traversaro et al., 2018)	7
2.3	Mapeamento por Permutação do Índice Cronológico (Traversaro et al., 2018)	7
2.4	Representação da Distância Euclidiana	10
2.5	Gráficos Complexidade-Entropia em relação à entropia de Shannon e as distâncias Euclidiana e de Wootters.	12
4.2	Representação gráfica da análise de uma série temporal de produção anual de cevada por acre.	17
4.1	Estrutura de organização dos componentes no RGtk2	18
4.3	Imagem atual do software.	18
5.1	Upload do arquivo	19
5.2	Gráfico do comportamento da Série Temporal	20
5.3	Histograma da distribuição da probabilidade de Bandt-Pompe	21
5.4	Resultado obtido da Entropia de Shannon	22
5.5	Resultado obtido da Complexidade Estatística	22
5.6	Caracterização do ruído $f^{-3/2}$ no Plano Complexidade-Entropia	23

Lista de Tabelas

2.1 Distâncias Estocásticas	10
---------------------------------------	----

1

Introdução

1.1 Motivação

Séries temporais estão presentes em todo o nosso cotidiano. São definidos como um conjunto de dados obtidos a partir de um processo observacional ao longo de um determinado período de tempo, não necessariamente dividido em espaços iguais, caracterizados pela dependência serial existente entre as observações.

A hipótese subjacente a toda essa análise é que os dados observados são o resultado da operação de um sistema causal sujeito a ruído observacional. Logo, esse sistema, ou dinâmica, é responsável pela criação de padrões que através de observações podemos inferir a respeito da dinâmica. Portanto, o estudo de tais dados auxilia na análise de diversas propriedades de sistemas.

Como comentado anteriormente, a aplicação deste conhecimento pode ser encontrada em múltiplas áreas do conhecimento científico como, por exemplo, na discriminação entre fenômenos estocásticos e caóticos (Ravetti et al., 2014), na identificação de padrões de comportamento em redes veiculares (Aquino et al., 2015), na classificação e verificação de assinaturas *online* (Rosso et al., 2016), na análise da eficiência informacional do mercado de petróleo (A. F. Bariviera and Rosso, 2016), na caracterização das séries temporais produzidas por eletroencefalogramas (Blanco et al., 1997), na análise da robustez de redes (Schieber et al., 2016), e na classificação de padrões de consumo de energia elétrica (Aquino et al., 2017).

Tradicionalmente o estudo de séries temporais costuma ser dividido em duas linhas de estudo, nos domínios do tempo e da frequência (Brockwell and Davis, 1991). No entanto, ambas abordagens utilizam diretamente os dados resultantes do processo observacional, que são sensíveis a efeitos provocados por diversos tipos de contaminação. Logo, surge assim a abordagem do uso de métodos não-paramétricos, como uma forma de evitar que tais efeitos invalidem as análises destes dados.

A Teoria da Informação surgiu como um ramo interdisciplinar, produzindo inúmeros resultados, tanto no ponto de vista teórico quanto nas aplicações e criação de novos métodos de extração de informações em sinais, abrangendo em suas soluções conceitos presentes na Probabilidade, Estatística e Telecomunicações.

O uso de suas ferramentas tem levado a resultados significativamente melhores do que aqueles obtidos com técnicas clássicas em diversas áreas do conhecimento. No trabalho de [Torres et al. \(2014\)](#), podemos ver uma grande contribuição no campo de processamento de imagens, onde este propõe uma técnica de filtragem que se adapta a cada ponto da imagem, observa uma janela de tamanho considerável e só emprega aquelas observações que não são muito discrepantes do valor central. Em [Bhattacharya et al. \(2015\)](#), vemos uma aplicação de distâncias estocásticas para obter uma decomposição polarimétrica otimizada. Já [Gambini et al. \(2015\)](#) propõe uma técnica de estimação de parâmetros minimizando distâncias estocásticas entre modelos e evidência empírica.

Entretanto, diversos desafios surgem na hora de tratar um problema com estes tipos de técnicas, pois ainda existem vários problemas analíticos e de ordem computacional em aberto, formando assim uma linha de pesquisa avançada, uma vez que requerem por parte dos envolvidos um bom domínio das teorias que dão sustentação às técnicas.

Atualmente há diversas ferramentas que auxiliam na análise clássica de séries temporais; para a plataforma R, existem diversas bibliotecas para essa finalidade (ver <https://cran.rproject.org/web/views/TimeSeries.html>). Além destas opções, o usuário também pode contar com os softwares de visualização de séries temporais. No entanto, todas as alternativas são limitadas às opções de bibliotecas e softwares que trabalham, em sua grande parte, com técnicas paramétricas e exigem familiaridade do usuário com o ambiente utilizado.

Desse modo, existem dois principais pontos nessas linhas de pesquisa que podem originar ótimos trabalhos inovadores:

- a necessidade de tornar as técnicas acessíveis a usuários não especializados, e
- a necessidade de otimizar o desenvolvimento de novas técnicas.

O primeiro ponto pode ser solucionado por meio do desenvolvimento de sistemas com interface gráfica que encapsulem os algoritmos presentes na literatura. Já o segundo, consiste em utilizar técnicas de desenvolvimento de software científico.

Logo, é na esfera do domínio dos problemas computacionais que surgem na aplicação de ferramentas oriundas da Teoria da Informação a séries temporais, que este trabalho se insere.

Apresentamos, assim, o desenvolvimento de uma ferramenta portátil, rápida e de boa qualidade numérica que possibilita análises interativas e exploratórias dos dados de uma série temporal através de técnicas provenientes da Teoria da Informação. Com ela, o usuário

dispõe de um conjunto de técnicas de análise presentes na literatura para processar e examinar seus dados de modo eficiente e com um mínimo período de aprendizado. A ferramenta é extensível.

1.2 Objetivo

O objetivo geral deste trabalho é propor e desenvolver uma ferramenta inovadora, resultante de propostas recentes de pesquisas relacionadas a Teoria da Informação, para facilitar o uso de técnicas avançadas de processamento e análise de sinais.

1.3 Solução proposta

Realizamos o uso de técnicas modernas de análise de séries temporais. Uma série temporal é transformada em uma sequência de símbolos, através da técnica de simbolização de τ . Essa técnica consiste em transformar vetores de tamanho D em padrões ordinais de forma não-paramétrica e formar um histograma de ocorrência dos $D!$ padrões possíveis. Esse histograma é tratado como uma função de probabilidade, do qual são extraídos descritores oriundos da Teoria da Informação. Esses descritores são, depois, mapeados em um plano adequado, e a sua localização serve para identificar o tipo de dinâmica subjacente à série temporal. Há uma grande diversidade de descritores como, por exemplo, distâncias (Kullback-Leibler, Bhattacharya, Hellinger, Rényi, Triangular, Harmônica, dentre outras), e entropias (Jensen-Shannon, Rényi, Tsallis, dentre outras). O ambiente gráfico oferecerá essas opções, e permitirá experimentar com a sua expressividade.

1.4 Contribuições

As contribuições deste trabalho são:

- A compreensão e implementação de técnicas de análise não-paramétrica de séries temporais utilizando descritores causais oriundos da Teoria da Informação;
- A implementação de uma interface gráfica amigável para a aplicação de tais descritores, mantendo a portabilidade do software para os diversos sistemas operacionais e arquiteturas de hardware.

Note que essas contribuições podem facilitar este processo de análise e construção do conhecimento por parte do usuário, tornando tal experiência mais simples e completa, fornecendo para este novas funcionalidades e uma maior interação do gráfico da série com os seus padrões.

1.5 Estrutura do texto

Este trabalho foi dividido em 5 capítulos e um anexo. No capítulo 2 introduzimos algumas principais técnicas e ferramentas disponíveis na literatura para a análise não-paramétrica de séries temporais utilizando descritores da Teoria da Informação, focando nos conceitos e metodologias aplicados com sucesso em diversos ramos de pesquisa científica. No capítulo 3 apresentamos a metodologia do trabalho desenvolvido. No capítulo 4 mostramos os resultados obtidos. As funções implementadas ao longo do desenvolvimento do projeto se encontram presente no Anexo A. E, finalmente, no Capítulo 6 apresentamos as considerações finais, concluindo este trabalho.

2

Fundamentação Teórica

Para que se obtenha um melhor entendimento acerca do tema proposto, neste capítulo serão apresentadas as fundamentações teóricas, obtidas por meio da realização da revisão bibliográfica dos conceitos e técnicas presentes no estado da arte.

2.1 Representação do espaço de probabilidade

A transformação de uma série temporal em uma distribuição de probabilidade (PDF) permite avaliar o conteúdo informacional acerca da dinâmica do sistema e dos processos subjacentes, descrevendo-os de forma mensurável e observável (Gray, 1990). Através desta conversão é possível utilizar métricas do espaço PDF, permitindo comparar diferentes conjuntos e classificá-los de acordo com as propriedades dos processos subjacentes. Podemos assim, por exemplo, classificar uma série entre estocástica ou determinística.

A ideia das técnicas não-paramétricas consiste em construir o histograma de algum atributo da série temporal, e extrair dele métricas de Teoria da Informação. Os atributos são os mais variados (Kowalski et al., 2011), dentre eles:

- (a) Padrões ordinais (Rosso et al., 2009),
- (b) Histogramas (Cha, 2007; Micco et al., 2008),
- (c) Dinâmica simbólica binária (Mischaikow et al., 1999),
- (d) Análise de Fourier (Powell and Percival, 2001), e
- (e) Transformada wavelet (Rosso et al., 2001).

Todas estas metodologias são capazes de capturar aspectos globais de dinâmicas complexas. No entanto, não é trivial encontrar uma representação simbólica significativa da série original. Assim, por considerar a causalidade temporal dos dados, a abordagem de ? revela detalhes importantes da estrutura ordinal da série temporal.

2.2 Método de simbolização de Bandt e Pompe

De acordo com a abordagem de Bandt e Pompe, substituímos a série por sequências de postos, obtidos pela análise desta ao longo do tempo.

Dada uma série temporal a tempo discreto $X = x_t : 1 \leq t \leq T$, uma dimensão D e um tempo de atraso (delay) τ , o particionamento é efetuado por meio da reorganização do sistema em conjuntos seguindo os seguintes passos:

Composição dos grupos: A série inicialmente será particionada em conjuntos de tamanho D e delay τ , possuindo a seguinte estrutura:

$$(s) \mapsto (x_{(t-1)+\tau}, \dots, x_{(t-1)+\tau+D-1}).$$

Formação dos padrões: Cada grupo formado anteriormente é então relacionado a um padrão ordinal π de ordem D , como se observa abaixo:

$$\{0, 1, \dots, D-1\}.$$

Elaboração dos símbolos: É realizada então a permutação dos elementos dos grupos, de tal forma que estes estejam ordenados de forma crescente.

$$x_{(t-1)+\tau} \leq x_{(t-1)+\tau+1} \leq \dots \leq x_{(t-1)+\tau+D-1}.$$

De mesmo modo é impreterível que a permutação ocorra com os elementos dos padrões relacionados a cada grupo, pois estes corresponderam aos símbolos da série. No esquemático da Figura 2.1 podemos visualizar a relação de cada padrão no comportamento dos elementos da série, sendo perceptível o quanto de informação sobre a estrutura da dinâmica temporal do sistema podemos extrair com essa técnica de simbolização.

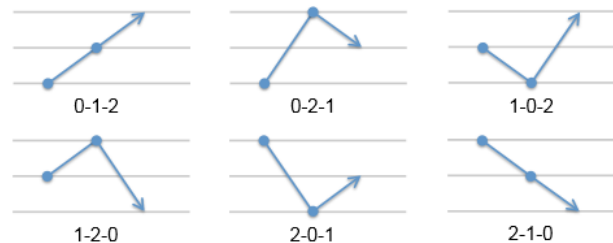


Figura 2.1: Representação gráfica dos padrões com dimensão $D = 3$.

A literatura apresenta duas maneiras de definir o mapeamento de padrões (Traversaro et al., 2018):

- (a) Ordenando as posições dos grupos em ordem cronológica (Permutação de Classificação), e
- (b) Ordenando os índices de tempo dos elementos dos subconjuntos (Permutação do Índice Cronológico).

Logo abaixo, observamos como se comporta a representação gráfica dos padrões ordinais quando aplicado cada um desses mapeamentos.

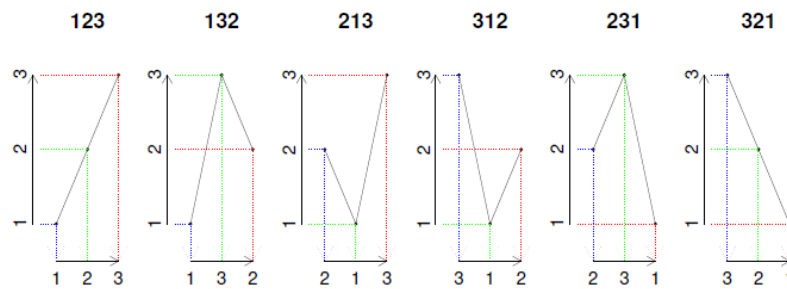


Figura 2.2: Mapeamento por Permutação de Classificação (Traversaro et al., 2018)

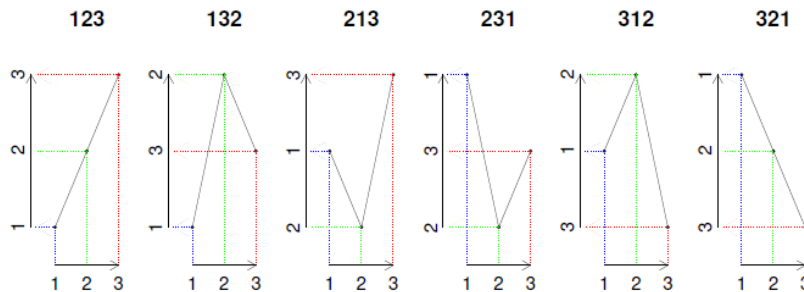


Figura 2.3: Mapeamento por Permutação do Índice Cronológico (Traversaro et al., 2018)

2.3 Distribuição de probabilidade de Bandt e Pompe

Em estatística, uma distribuição discreta de probabilidade refere-se à distribuição de frequências relativas para os resultados de um espaço amostral, apontando a quantidade de vezes em que um determinado elemento do conjunto assume cada um dos seus possíveis valores.

Logo:

$$\sum_{i=1}^n P_i = 1.$$

Considerando isto, a distribuição de probabilidade de Bandt & Pompe consiste no cálculo da distribuição dos símbolos da série diante das $D!$ possíveis permutações dos padrões ordinais π de comprimento D :

$$p(\pi) = \frac{\#\{t | t \leq T - (D-1)\tau, (x_{t+1}, \dots, x_{t+D}) \text{ do tipo } \pi\}}{T - (D-1)\tau}$$

Uma grande vantagem de sua utilização refere-se ao fato da distribuição de probabilidade tornar-se invariante com respeito às transformações monotônicas, propriedade extremamente desejada na análise das séries.

Uma vez calculado o histograma de padrões $\mathbf{p} = (p_1, \dots, p_{D!})$, isto é, a função de probabilidade, o próximo passo será obter descritores.

2.4 Entropia de permutação

A Entropia mede o desordem ou a imprevisibilidade de um sistema caracterizado por uma função de probabilidade \mathbf{p} . Neste trabalho, citaremos três modelos de entropia: Shannon, Tsallis e Rényi.

Proposta em 1948, a entropia de Shannon consiste de uma variação da Entropia de Boltzmann-Gibbs (Shannon, 2001). Seja, assim, $\mathbf{p} = (p_1, \dots, p_{D!})$ o histograma de proporções dos $D!$ padrões observados a partir da série temporal \mathbf{X} . Calculamos a entropia de Shannon:

$$S(\mathbf{p}) = - \sum_i^{D!} p_i \ln p_i. \quad (2.1)$$

Seu valor mínimo ocorre quando $S_{\min} = S(\mathbf{p}) = 0$, neste caso particular podemos assumir que temos conhecimento máximo sobre o sistema, uma vez que a probabilidade de um dado evento i ocorrer será unicamente determinada pela sua probabilidade p_i . No entanto, quando o comportamento do sistema é descrito por uma distribuição uniforme, ou seja, quando a sua probabilidade for determinada por $p_e = \{1/D! : i = 1, 2, \dots, D!\}$, teremos conhecimento mínimo dos dados analisados. Desse modo, $S_{\max} = S(\mathbf{p}) = \ln D!$.

Entretando, na literatura usualmente é utilizada a entropia normalizada de Shannon definida por Martin et al. (2006), dada por:

$$H(\mathbf{p}) = \frac{S(\mathbf{p})}{S_{\max}} \quad (2.2)$$

Uma vez que aplicada para estimar a desordem presente em uma distribuição de probabilidade de Bandt-Pompe, tal medida passa a ser chamada de Entropia de Permutação Normalizada (Bandt and Pompe, 2002), sendo definida por:

$$H(\mathbf{p}) = -\frac{1}{\ln D!} \sum_i^{D!} p_i \ln p_i \quad (2.3)$$

Tsallis propôs um novo modelo (Rényi, 1961), ampliando o conjunto de aplicações abordado por Boltzmann:

$$H_\alpha(\mathbf{p}) = (\alpha - 1)^{-1} (1 - \log \sum_{i=1}^{D!} p_i^\alpha), \quad \text{com } \alpha \neq 1. \quad (2.4)$$

A entropia de Rényi é uma generalização da entropia de Shannon, sendo aplicada em Teoria da Informação como um índice estatístico de diversidade ou aleatoriedade (Tsallis, 1988):

$$H_\alpha(\mathbf{p}) = (1 - \alpha)^{-1} \log \sum_{i=1}^{D!} p_i^\alpha. \quad (2.5)$$

2.5 Distância Estocástica

A capacidade da entropia de capturar propriedades do sistema é limitada, logo se faz necessário a utilização da mesma em conjunto de outros descritores, para assim realizar uma análise mais completa. Outras medidas interessantes são distâncias entre a função de probabilidade \mathbf{p} e uma medida de probabilidade que descreva um processo não informativo, tipicamente a distribuição uniforme.

Para mensurar a similaridade entre duas distribuições de séries temporais, todas as funções que calculam determinada característica devem respeitar algumas propriedades.

Sendo c_1, c_2 e c_3 objetos do universo de objetos, devem ser mantidas as seguintes particularidades:

- Simetria: $D(c_1, c_2) = D(c_2, c_1)$
- Similaridade: $D(c_1, c_1) = 0$
- Positividade: $D(c_1, c_2) = 0$ se, e somente se, $c_1 = c_2$
- Desigualdade triangular: $D(c_1, c_3) \leq D(c_1, c_2) + D(c_2, c_3)$

Também consideradas no estudo relatado, as chamadas divergências são aquelas na qual sequegem apenas duas das particularidades acima, positividade e similaridade.

A Tabela 2.1 mostra algumas possíveis medidas de distância $d(\mathbf{p}, \mathbf{q})$ entre duas funções de probabilidade $\mathbf{p} = (p_1, \dots)$ e $\mathbf{q} = (q_1, \dots)$, definidas sobre o mesmo suporte.

Outras distâncias e relações entre elas podem ser vistas no livro de Deza e Deza (Deza and Deza, 2009).

Euclidiana	$\sqrt{\sum_i (q_i - p_i)^2}$
Manhattan	$\sum_i q_i - p_i $
Chebyshev	$\max_i \{ q_i - p_i \}$
Kullback-Leibler	$\sum_i q_i \log \frac{q_i}{p_i}$
Jensen-Shannon	$\sum_i \left(p_i \log \frac{p_i}{q_i} + q_i \log \frac{q_i}{p_i} \right)$
Wotters	$\cos^{-1} \sum_i \sqrt{p_i q_i}$
Bhattacharya	$-\log \sum_i \sqrt{p_i q_i}$

Tabela 2.1: Distâncias Estocásticas



Figura 2.4: Representação da Distância Euclidiana

2.6 Complexidade Estatística

Por definição complexidade refere-se a um conjunto de coisas ligadas por um nexo comum. Inversamente à entropia, a complexidade estatística procura encontrar estruturas de interação e dependência entre os elementos de uma dada série, tratando-se de um fator extremamente importante no estudo de sistemas dinâmicos.

Essa propriedade é definida por meio da fórmula desenvolvida por Lopèz-Ruiz, Mancini e Calbet, onde uma Entropia e uma Distância, também chamada de desequilíbrio, podem ser combinadas no atributo Complexidade Estatística para aumentar o seu poder de descrição (Feldman et al., 2008; Feldman and Crutchfield, 1998; López-Ruiz et al., 1995):

$$C(\mathbf{h}, \mathbf{p}) = H(\mathbf{h})Q(\mathbf{h}, \mathbf{p}) \quad (2.6)$$

O desequilíbrio Q reflete como se comporta a arquitetura do sistema analisado. Quando tal sistema possui alguma estrutura privilegiada ou estados mais prováveis entre os acessíveis, esse valor será diferente de zero.

Uma escolha conveniente é a complexidade de Jensen-Shannon, dada por

$$C_{JS}(\mathbf{h}) = H_S(\mathbf{h}) \cdot Q_{JS}(\mathbf{h}, \mathbf{p}_e), \quad (2.7)$$

em que H_S é a entropia de Shannon normalizada, \mathbf{h} a função de probabilidade da série, \mathbf{p}_e a distribuição uniforme e Q_{JS} é a divergência de Jensen-Shannon, cuja importância da

discutida em [Lamberti et al. \(2004\)](#). Temos então:

$$Q(\mathbf{h}, \mathbf{p}_e) = Q_0 \cdot J(\mathbf{h}, \mathbf{p}_e), \quad (2.8)$$

Sendo,

$$J(\mathbf{h}, \mathbf{p}_e) = S\left(\frac{\mathbf{h} + \mathbf{p}_e}{2}\right) - \frac{S(\mathbf{h})}{2} - \frac{S(\mathbf{p}_e)}{2}, \quad (2.9)$$

e Q_0 uma constante de normalização, logo $0 \leq Q_0 \leq 1$, definida por:

$$Q_0 = -2 \left[\left(\frac{N+1}{N} \right) \ln(N+1) - 2 \ln 2N + \ln N \right]^{-1}. \quad (2.10)$$

2.7 Plano Complexidade-Entropia

O plano Complexidade-Entropia refere-se ao gráfico bidimensional entre a Entropia de Permutação Normalizada $H(p)$ (eixo horizontal) e a Complexidade Estatística $C(\mathbf{p}, \mathbf{p}_e)$ (eixo vertical).

Por intermédio de tal ferramenta é possível descobrir a natureza da série, determinando se esta corresponde a uma sequência caótica, estocástica ou determinística, analisando o seu comportamento, visto que estes possuem dinâmicas diferentes. De acordo com a segunda lei da termodinâmica:

A quantidade de entropia de qualquer sistema isolado termodinamicamente tende a incrementar-se com o tempo, até alcançar um valor máximo.

Como a entropia varia uniformemente com o tempo, podemos concluir que o plano Complexidade-Entropia além de analisar a interação entre estas duas características, também verifica a evolução temporal de $C(\mathbf{p}, \mathbf{p}_e)$.

O plano Entropia-Complexidade também é conhecido como “O plano de causalidade entre a entropia e a complexidade”, tendo em vista que no ramo da estatística causalidade refere-se a relação entre as causas dos fenômenos e seus respectivos efeitos e resultados. Assim, podemos inferir que como a própria nomenclatura sugere, o diagrama relaciona os dados resultantes do cálculo da entropia e da complexidade estatística e as suas características estimadas pela Teoria da Informação.

Cada série temporal X pode, então, ser mapeada no ponto $(H_S, C(\mathbf{p}, \mathbf{p}_e))$. O conjunto de todos os pontos possíveis forma o *mapa Entropia-Complexidade*, e a posição do ponto nesse plano é um descritor das propriedades da dinâmica subjacente à série ([Sinn and Keller, 2011](#)). A forma desse plano depende do comprimento D dos padrões ([Martin et al., 2006](#)).

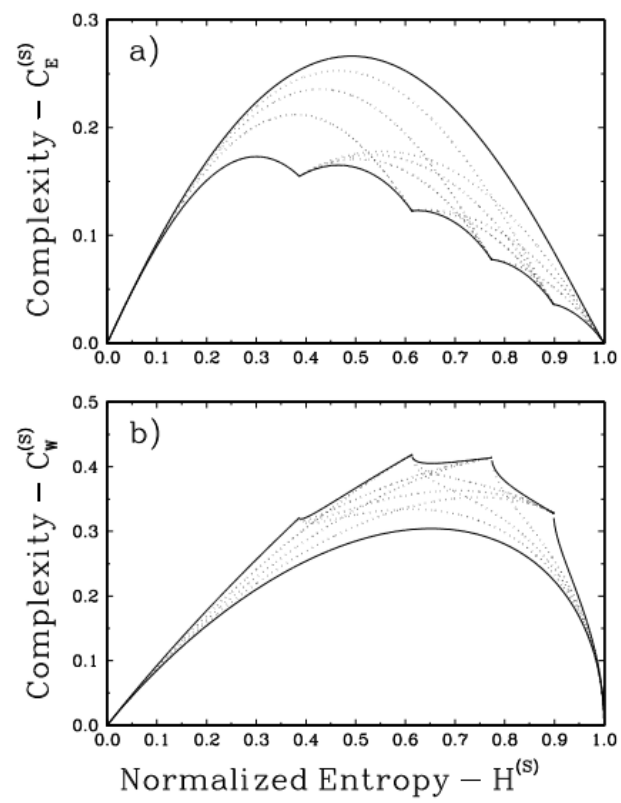


Figura 2.5: Gráficos Complexidade-Entropia em relação à entropia de Shannon e as distâncias Euclidiana e de Wootters.

3

Metodologia

A metodologia da pesquisa desenvolvida consistiu em dois grandes momentos, a etapa teórica e a implementação das funcionalidades.

Para o desenvolvimento do projeto descrito neste trabalho, foram planejadas as seguintes etapas de execução.

3.1 Estudo das funções a serem implementadas

O estudo das funções a serem implementadas foi realizado a partir da análise de um conjunto de referências bibliográficas de qualidade, visando ampliar os conhecimentos a cerca do tema proposto.

Foram estudados ao longo deste momento, temas como séries temporais, suas propriedades e aplicações, Teoria da Informação, entropias (Salicrú et al., 1993), distâncias estocásticas (Pardo, 2006), complexidades estatísticas, plano Complexidade-Entropia e a linguagem de programação R.

3.2 Implementação e validação numérica

Após o término da revisão bibliográfica da literatura existente, foi dado então início à implementação do trabalho, desenvolvido em R e sempre fazendo uso de boas práticas de desenvolvimento de software científico.

Para que tal ferramenta seja aplicada na análise de dados é de suma importância realizar a verificação de suas propriedades numéricas. Portanto, a avaliação da qualidade numérica das funcionalidades desenvolvidas foi feita utilizando uma metodologia própria baseada em sistemas dinâmicos com saídas conhecidas.

3.3 Análise de alternativas para o desenvolvimento da interface

Um dos grandes objetivos da pesquisa consistia em ampliar a aplicabilidade das técnicas de extração de informações de séries temporais, por meio de uma ferramenta portátil e interativa de análise. Assim, foram avaliadas algumas opções de ferramentas de GUI que fossem capaz de suportar as funcionalidades desenvolvidas em R na primeira etapa.

Foi então feita uma pesquisa sobre as alternativas existentes sendo considerado os seguintes fatores:

- Portabilidade do software para os diversos sistemas operacionais e arquiteturas de hardware;
- Facilidade de instalação, pois como queremos por meio do desenvolvimento do projeto facilitar de um modo geral a análise de séries temporais na experiência do usuário, esta não deverá apresentar problemas no processo de instalação;
- Integração com a linguagem de programação R.

Desse modo, RGtk2 e Java Swing foram as alternativas iniciais para o desenvolvimento da interface gráfica. No entanto, após estudos sobre o funcionamento destas GUIs (*Graphical User Interface*), verificamos que a implementação da interface utilizando Java Swing apresentava certos empecilhos em relação a portabilidade do software em diferentes sistemas operacionais, não satisfazendo ao item 1 de nossas exigências, seria necessário a implementação individual do software para cada sistema operacional, já que o programa deveria ser capaz de reconhecer o sistema utilizado pelo cliente e assim executar seguindo as regras e padrões deste. Outro fator decisivo foram as dificuldades de comunicação entre o código Java e o script em R.

Portanto, optamos pelo RGtk2, por ser uma biblioteca própria do ambiente de desenvolvimento R e pela sua maior facilidade em manter a portabilidade do sistema.

3.4 Desenvolvimento de protótipos

Foram desenvolvidos alguns protótipos de modelos de interface com as alternativas de bibliotecas gráficas citadas anteriormente, sempre com foco na experiência do usuário.

No entanto, por possuímos como objetivo o desenvolvimento de uma ferramenta *Desktop* algumas alterações foram realizadas para se adequar as funções oferecidas pela biblioteca escolhida.

3.5 Versão de produção da interface

Após a finalização do processo de escolha da biblioteca `RGtk2`, foi então dado início a implementação da interface. Esta etapa consistiu basicamente da realizada da integração entre o ambiente gráfico do sistema e as funções de análise de séries temporais implementadas em fases anteriores.

3.6 Validação, verificação e preparação de manuais e tutoriais de uso

Como já citado, é de fundamental importância para tal projeto a verificação da qualidade numérica do software desenvolvido, portanto um dos seus objetivos consistiu em validar a interface e as funções com usuários finais.

Foram também desenvolvidos manuais de uso das funções implementadas, informando as suas funcionalidades, parâmetros de entrada e o resultado final computado. Todas essas descrições se encontram apresentados no apêndice A deste trabalho.



Resultados e Discussões

Apresentamos o desenvolvimento de uma ferramenta portátil, rápida e de boa qualidade numérica que possibilita gerar novos métodos de interação do usuário com o sistema de análise, permitindo que este seja capaz de analisar os diferentes descritores oriundos da Teoria da Informação e permitir a análise gráfica dos resultados.

Seguindo o modelo de engenharia de software em espiral, o sistema foi projetado e desenvolvido de forma modular, composto pelas seguintes unidades:

- Módulo de simbolização;
- Módulo de análise;
- Módulo de visualização e interação (Em fase de desenvolvimento);

Esses módulos foram e estão sendo desenvolvidos seguindo um cronograma. Depois passaram pelas seguintes etapas:

- Integração dos módulos em um sistema;
- Teste e validação do sistema;
- Geração da interface gráfica.

Permite-se a leitura de dados em vários formatos (TXT, CSV ou XLSX), e o usuário a seguir poderá escolher:

- Gerar o gráfico da série (ver Figura 1);
- Calcular seus diversos valores de Entropia;
- Calcular seus diversos valores de Distâncias Estocásticas;

- Calcular complexidades estatísticas;
- Identificar padrões no gráfico da série temporal;
- Gerar planos de Entropias;
- Gerar planos de Distâncias Estocásticas;
- Gerar o histograma de padrões (ver Figura 1);
- Identificar o ponto característico da série no plano Entropia-Complexidade (ver Figura 1).

Um elemento original do sistema é a vinculação entre o histograma de padrões, formado através do processo de simbolização de Bandt-Pompe (Bandt and Pompe, 2002), e a série temporal. Escolhendo um ou mais elementos do histograma, os valores correspondentes na série temporal aparecem realçados. Esta funcionalidade permite a análise visual da distribuição temporal dos padrões, possibilitando futuramente a realização de outros testes.

O teste e a validação do sistema foram tarefas contínuas ao longo do desenvolvimento do projeto, bem como o incremento do desenvolvimento de novas funcionalidades.

Com a troca da ferramenta de interface, foi necessário primeiramente um estudo de documentações referentes ao pacote gráfico (Lawrence and Temple Lang, 2010). Uma vez que ocorreu uma mudança de paradigmas, pois a biblioteca escolhida funciona por meio de blocos verticais e horizontais, onde os horizontais se são distribuídos diante dos verticais, foram encontrados os seguintes problemas durante a implementação:

- A reprodução do modelo do protótipo;
- A implementação da função referente a `file.choose` em R, pois o escopo das variáveis declaradas dentro das funções de tratamento de interrupções é local;
- A implementação das funções de tratamento de interrupção;
- O desenvolvimento da parte estética do software.

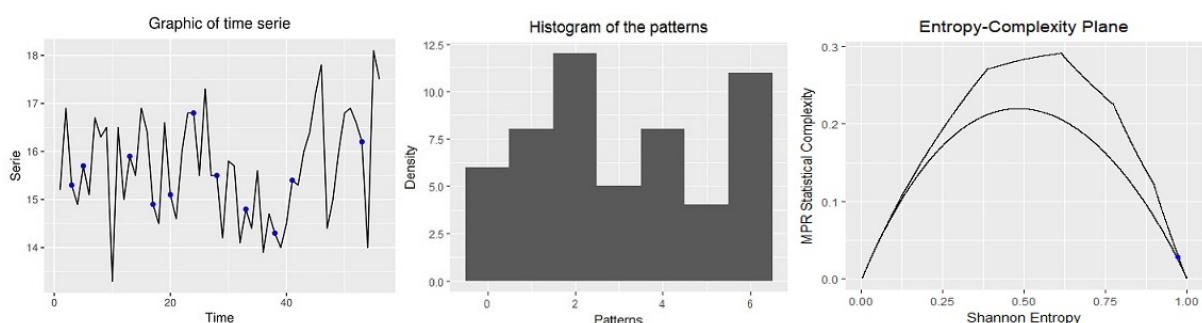


Figura 4.2: Representação gráfica da análise de uma série temporal de produção anual de cevada por acre.

Figura 4.1: Estrutura de organização dos componentes no RGtk2

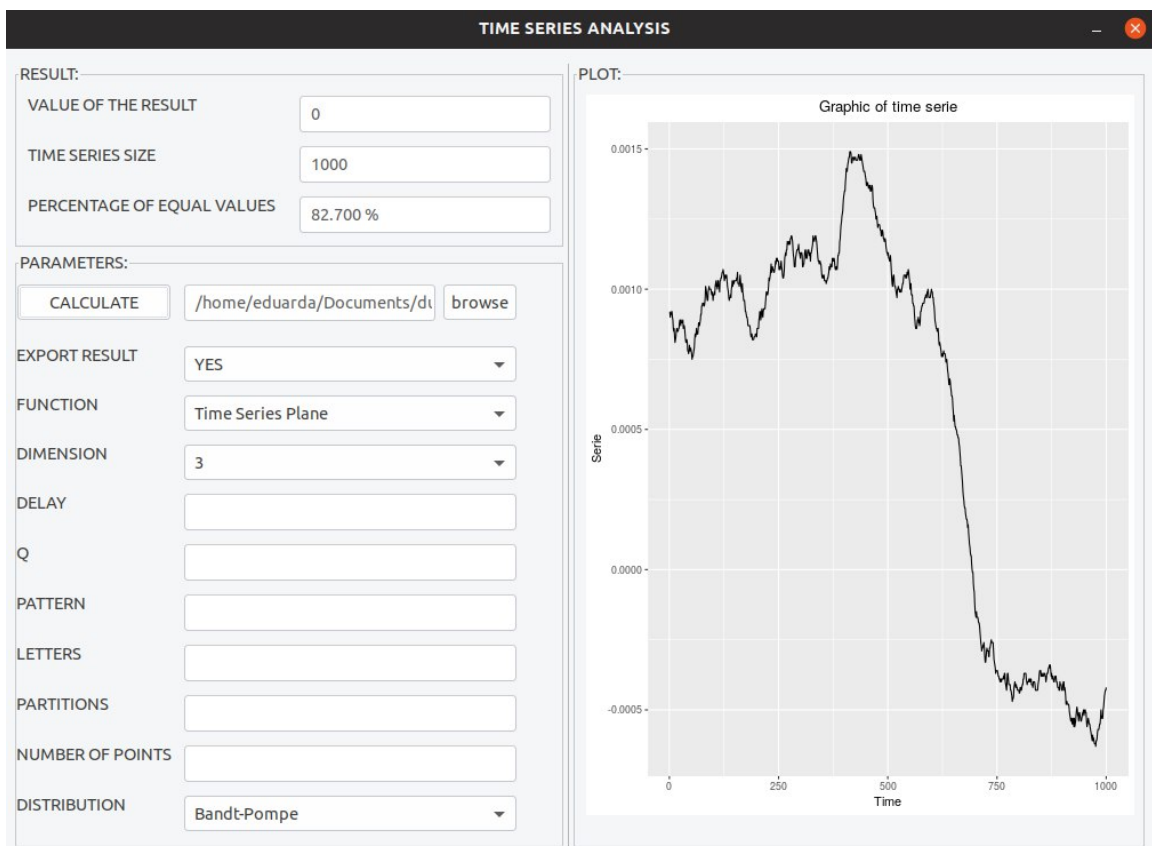
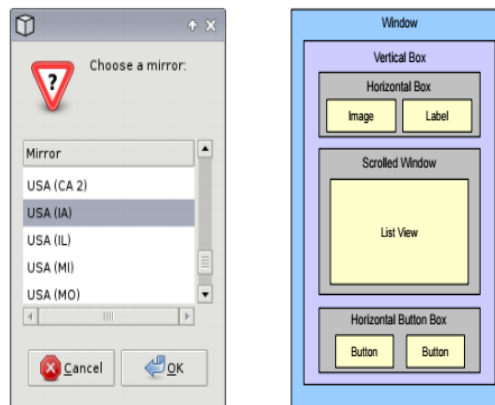


Figura 4.3: Imagem atual do software.

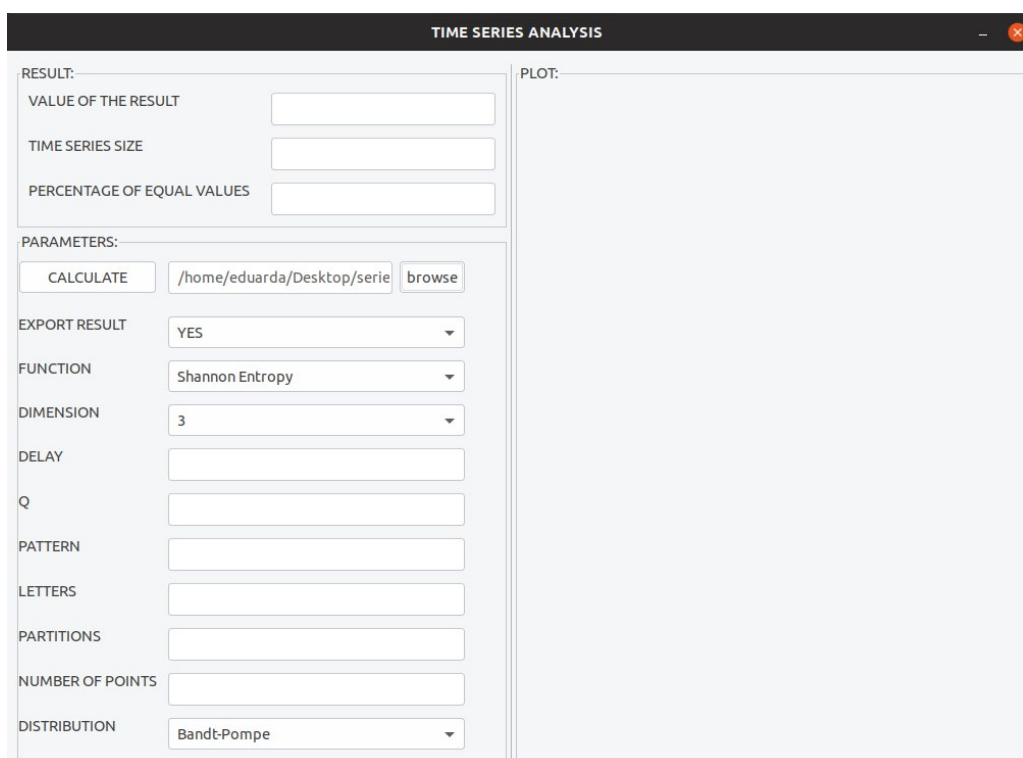
5

Demonstração de uso do Software

Nesta sessão, demonstraremos como utilizar a interface do Software desenvolvido para realizar a análise da caracterização do ruído colorido ¹ de espectro de potência $f^{-3/2}$.

5.1 Upload de dados

Primeiramente, iremos fazer upload do arquivo .csv que contém os dados que serão utilizados. Para isso iremos clicar no botão **BROWSE** e selecionar o arquivo desejado (Figura 5.1).



The screenshot shows the 'TIME SERIES ANALYSIS' software interface. The window title is 'TIME SERIES ANALYSIS'. The interface is divided into two main sections: 'RESULT' and 'PARAMETERS'. The 'RESULT' section contains three input fields: 'VALUE OF THE RESULT', 'TIME SERIES SIZE', and 'PERCENTAGE OF EQUAL VALUES'. The 'PARAMETERS' section contains a 'CALCULATE' button, a file path input field with a 'browse' button, and several dropdown and input fields: 'EXPORT RESULT' (set to YES), 'FUNCTION' (Shannon Entropy), 'DIMENSION' (3), 'DELAY', 'Q', 'PATTERN', 'LETTERS', 'PARTITIONS', 'NUMBER OF POINTS', and 'DISTRIBUTION' (Bandt-Pompe). A large empty 'PLOT' area is on the right side of the interface.

Figura 5.1: Upload do arquivo

¹<https://www.mathworks.com/matlabcentral/fileexchange/35381-noisefk-m>

5.2 Visualização da série temporal

O próximo passo será visualizar como se comporta a série temporal ao longo do tempo. Para isso, iremos selecionar dentro das possibilidades da variável `FUNCTION` a funcionalidade `Time Series Plane`.

Como podemos verificar, algumas informações básicas sobre os dados também são fornecidas, como o tamanho da série e o percentual de valores repetidos (Figura 5.2).

O software também disponibiliza a opção de exportar os resultados obtidos em cada iteração com o usuário, para isso é necessário apenas habilitar a opção na variável `EXPORT RESULT`. Todos os devidos arquivos resultantes serão armazenados no mesmo diretório que o sistema se encontra.

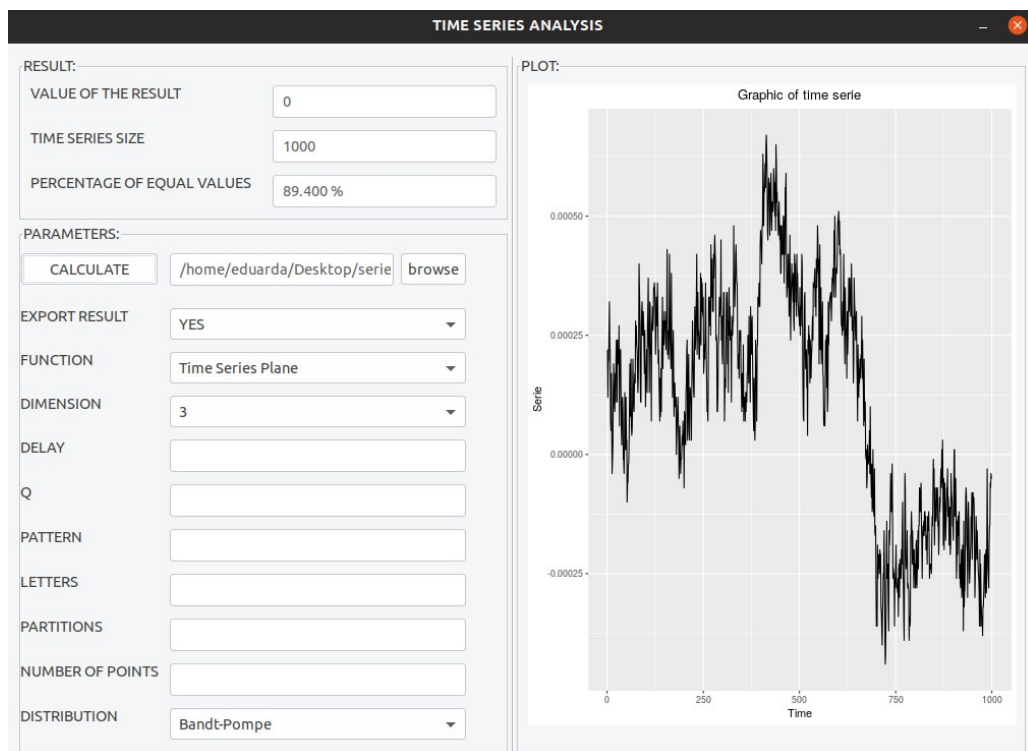


Figura 5.2: Gráfico do comportamento da Série Temporal

5.3 Histograma da distribuição de Bandt-Pompe

Assim como propõe a metodologia da simbolização, iremos agora visualizar como se comporta a distribuição dos padrões de Bandt-Pompe. Neste exemplo, aplicaremos valores de dimensão $D = 3$ e delay $\tau = 1$. Para isso, selecionaremos a funcionalidade `Histogram` e configuraremos a variável `DELAY` para o valor desejado (Figura 5.3).

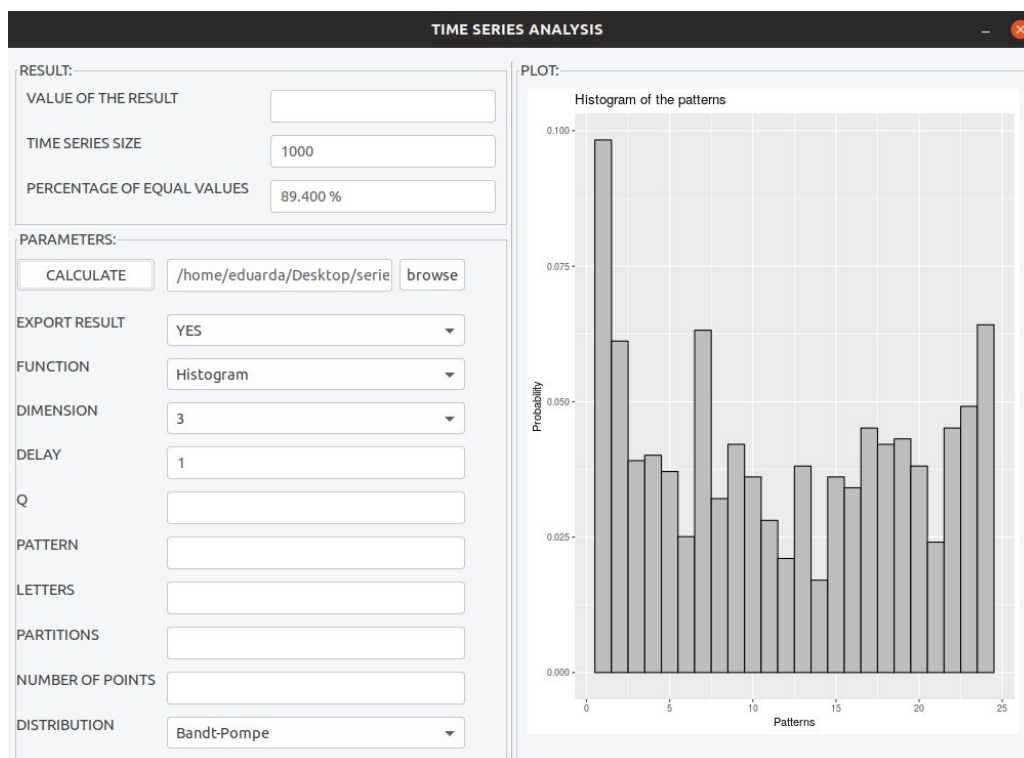


Figura 5.3: Histograma da distribuição da probabilidade de Bandt-Pompe

5.4 Cálculo da Entropia de Shannon

Para adquirir isoladamente o valor da Entropia de Permutação Normalizada de Shannon, devemos agora apenas selecionar a opção **Shannon Entropy** e pressionar o botão **CALCULATE** (Figura 5.4).

5.5 Cálculo da Complexidade Estatística

De modo semelhante a Entropia, para possui o valor da Complexidade Estatística, devemos selecionar a opção **Statistical Complexity** e pressionar o botão **CALCULATE** (Figura 5.5).

5.6 Plano Complexidade-Entropia

Por fim, uma vez que os valores referentes a dimensão D e o delay τ já se encontram configurados, para gerar o Plano Complexidade-Entropia devemos apenas selecionar a opção **HC Plane** e informar em quantas partições queremos analisar a série, caso o valor informado seja superior a 1, a série irá ser dividida em subconjuntos e exibido os pontos correspondentes a cada um destes (Figura 5.6).

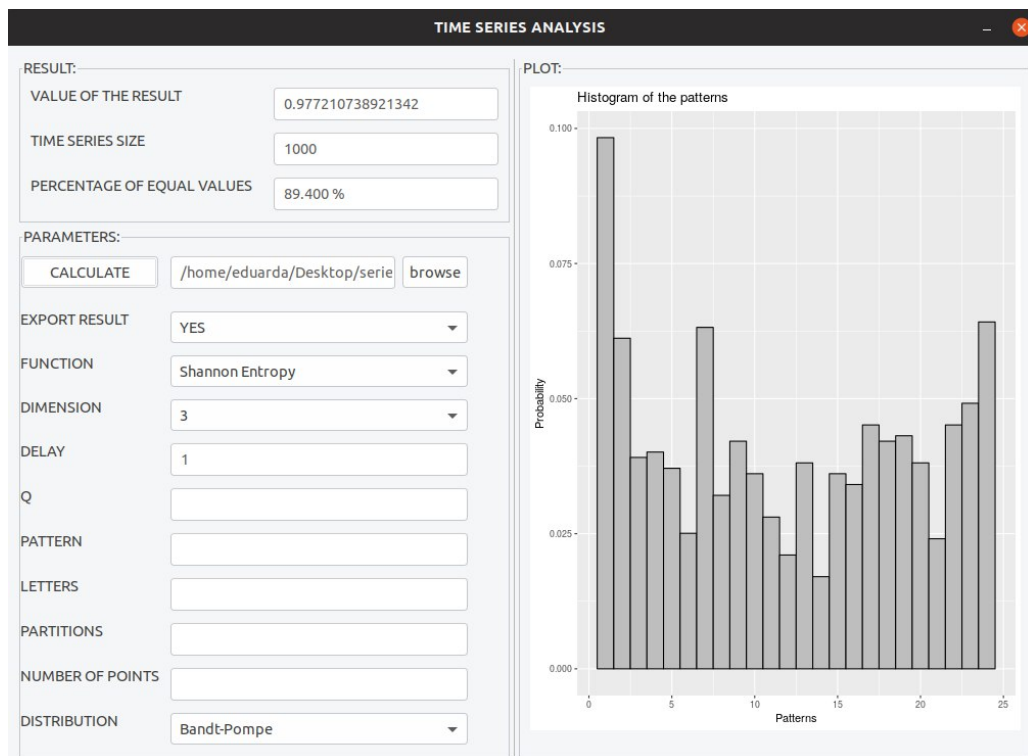


Figura 5.4: Resultado obtido da Entropia de Shannon

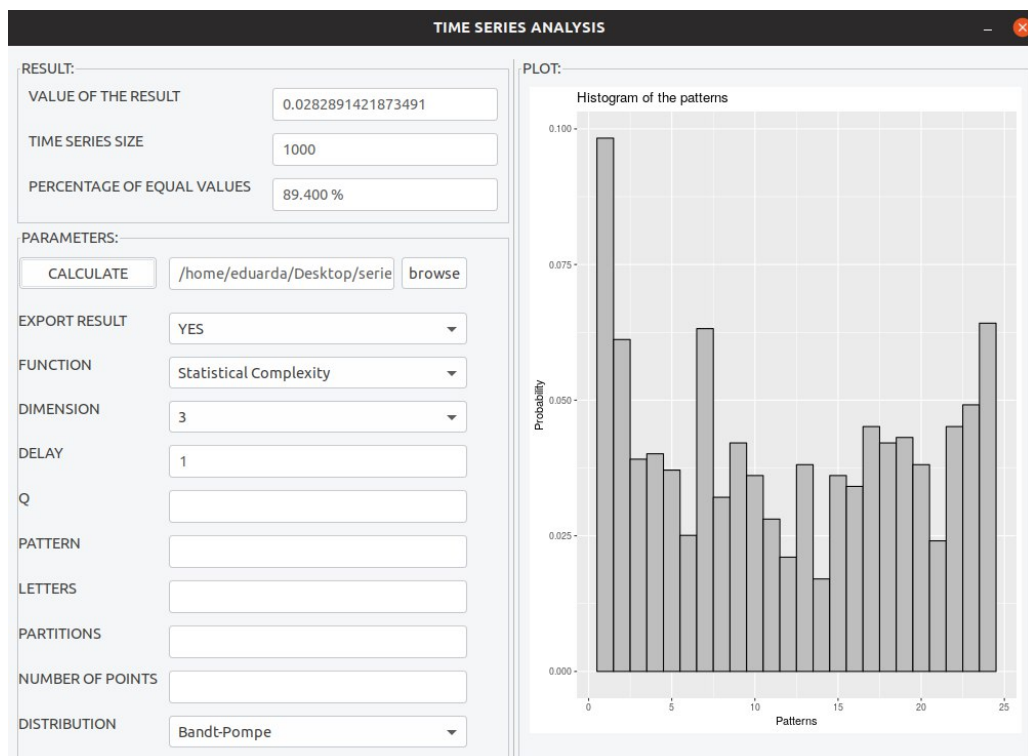


Figura 5.5: Resultado obtido da Complexidade Estatística

Como podemos observar, o comportamento descrito no plano corresponde ao valor já esperado na literatura (Zemp, 2012), o ruído $f^{-3/2}$ possui um alto valor de Entropia, ou seja

alta desordem na estrutura da dinâmica dos seus dados e um baixo valor de Complexidade.

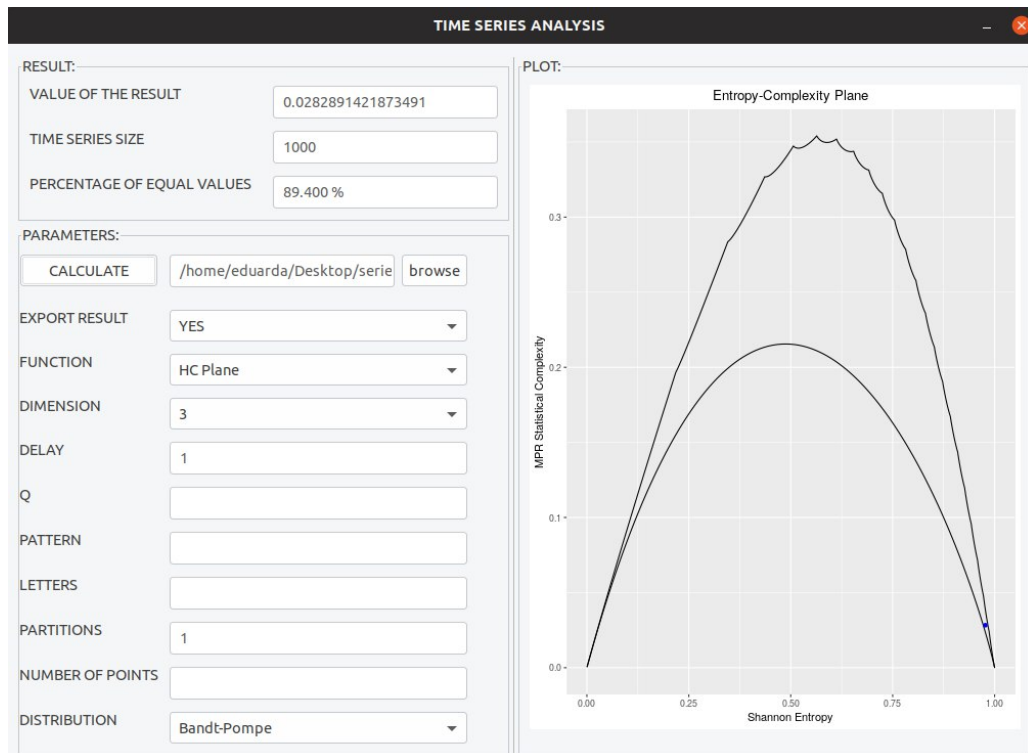


Figura 5.6: Caracterização do ruído $f^{-3/2}$ no Plano Complexidade-Entropia

6

Conclusões

Neste capítulo serão abordados os avanços no meio científico e a importância proporcionada através do desenvolvimento deste trabalho. Além disso, também apresentaremos sugestões para futuros trabalhos.

6.1 Considerações Finais

Este trabalho propôs o desenvolvimento de uma ferramenta portátil, rápida e de boa qualidade numérica que possibilita análises de uma série temporal através de descritores provenientes da Teoria da Informação. Para atribuir uma função de distribuição de probabilidade utilizamos o método de simbolização de Bandt-Pompe. A caracterização dos dados é dada por meio dos seus descritores, sendo então disponibilizadas diversas entropias, distâncias estocásticas e complexidade estatística.

Um elemento original do sistema é a vinculação entre o histograma de padrões e a série temporal. Escolhendo um ou mais elementos do histograma, os valores correspondentes na série temporal aparecem realçados. Esta funcionalidade permite a análise visual da distribuição temporal dos padrões, possibilitando futuramente a realização de outros testes.

O projeto também oferece aos pesquisadores a facilidade de utilização de técnicas sofisticadas da computação científica por meio de uma interface simples e intuitiva, sendo possível realizar em poucos passos atividades antes realizadas apenas por meio de scripts, exigindo assim mínimo conhecimento com programação por parte do usuário.

6.2 Trabalhos futuros

Pretendemos expandir as funcionalidades do sistema, dando agora ênfase ao problema da imputação de padrões ausentes. Para tanto, pretendemos atingir os seguintes objetivos:

- Estudar e implementar técnicas para imputação de padrões ausentes ocasionados por dados repetidos;
- Analisar a capacidade de reconstrução de informações dessas técnicas quando a série temporal é armazenada com menos precisão do que a ideal;
- Analisar a distribuição temporal dos padrões originais e imputados.

Apêndice A

Manual de utilização das funções desenvolvidas

A.1 Pacotes necessários

Para que seja possível utilizar plenamente as funções desenvolvidas ao longo deste projeto será necessário que os seguintes pacotes estejam instalados no ambiente *RStudio*:

- `combinat`
- `ggplot2`
- `dygraphs`
- `ggthemes`

Após a instalação, o usuário pode realizar normalmente as chamadas das funções implementadas.

A.2 Principais funções desenvolvidas

`equalitiesValues`

Percentual de dados repetidos em uma dada série temporal

Uso

1 `equalitiesValues (serie)`

Argumentos

<i>serie</i>	Um vetor numérico onde estará instanciada a série temporal que deve ser avaliada pela função.
--------------	---

<code>removeDuplicate</code>	<i>Remove os dados duplicados de uma série temporal</i>
------------------------------	---

Uso

```
1 removeDuplicate(serie)
```

Argumentos

<i>serie</i>	Um vetor numérico onde estará instanciada a série temporal que deve ser avaliada pela função.
--------------	---

<code>distribution</code>	<i>Distribuição de probabilidade de Bandt-Pompe</i>
---------------------------	---

Uso

```
1 distribution(serie, dimension, delay)
```

Argumentos

<i>serie</i>	Um vetor numérico onde estará instanciada a série temporal que deve ser avaliada pela função.
--------------	---

<i>dimension</i>	Dimensão dos padrões ordinais.
------------------	--------------------------------

<i>delay</i>	Delay utilizado na formação dos padrões.
--------------	--

WPE	Distribuição de probabilidade de Bandt and Pompe Weigth.
-----	--

Uso

1 `WPE (serie, dimension, delay)`

Argumentos

<i>serie</i>	Um vetor numérico onde estará instanciada a série temporal que deve ser avaliada pela função.
<i>dimension</i>	Dimensão dos padrões ordinais.
<i>delay</i>	Delay utilizado na formação dos padrões.

shannonEntropy	<i>Entropia de Shannon de uma dada distribuição de probabilidade</i>
----------------	--

Uso

1 `shannonEntropy (p)`

Argumentos

<i>p</i>	Uma distribuição de padrão ordinal.
----------	-------------------------------------

shannonNormalized

Entropia normalizada de Shannon de uma distribuição de probabilidade

Uso

1 `shannonNormalized(p)`

Argumentos

p Uma distribuição de padrão ordinal.

tsallisEntropy

Entropia de Tsallis de uma dada distribuição de probabilidade

Uso

1 `tsallisEntropy(p, q)`

Argumentos

p Uma distribuição de padrão ordinal.

q A ordem da entropia. Permite apenas números positivos.

tsallisNormalized

Entropia normalizada de Tsallis de uma distribuição de probabilidade

Uso

1 `tsallisNormalized(p, q)`

Argumentos

p

Uma distribuição de padrão ordinal.

q

A ordem da entropia. Permite apenas números positivos.

renyiEntropy

Entropia de Renyi de uma distribuição de probabilidade

Uso

1 `renyiEntropy(p, q)`

Argumentos

p

Uma distribuição de padrão ordinal.

q

A ordem da entropia. Permite apenas números positivos.

renyiNormalized

Entropia normalizada de Renyi de uma distribuição de probabilidade

Uso

1 `renyiNormalized(p, q)`

Argumentos

p

Uma distribuição de padrão ordinal.

q

A ordem da entropia. Permite apenas números positivos.

PME

Entropia entropia de mínima permutação de uma distribuição de probabilidade

Uso

1 `PME(p)`

Argumentos

p

Uma distribuição de padrão ordinal.

euclidianDistance

Distância euclidiana de uma dada distribuição de probabilidade e a distribuição de probabilidade uniforme

Uso

1 `euclidianDistance (p)`

Argumentos

p

Uma distribuição de padrão ordinal.

squaredDistance

Distância euclidiana quadrada de uma dada distribuição de probabilidade e a distribuição de probabilidade uniforme

Uso

1 `squaredDistance (p)`

Argumentos

p

Uma distribuição de padrão ordinal.

manhattanDistance

Distância de Manhattan de uma dada distribuição de probabilidade e a distribuição de probabilidade uniforme

Uso

1 `manhattanDistance(p)`

Argumentos

p

Uma distribuição de padrão ordinal.

chebyshevDistance

Distância de Chebyshev de uma dada distribuição de probabilidade e a distribuição de probabilidade uniforme

Uso

1 `chebyshevDistance(p)`

Argumentos

p

Uma distribuição de padrão ordinal.

hellingerDistance

Distância de Hellinger de uma dada distribuição de probabilidade e a distribuição de probabilidade uniforme

Uso

1 hellingerDistance (p)

Argumentos

p

Uma distribuição de padrão ordinal.

jensenDivergence

Medida de desequilíbrio generalizado para distribuições de probabilidade com base na divergência de Jensen-Shannon

Uso

1 jensenDivergence (p)

Argumentos

p

Uma distribuição de padrão ordinal.

woottersDistance

Distância de Wootters de uma dada distribuição de probabilidade e a distribuição de probabilidade uniforme

Uso

1

```
woottersDistance (p)
```

Argumentos p

Uma distribuição de padrão ordinal.

kullbackDivergence

Medida de desequilíbrio generalizado para distribuições de probabilidade com base na divergência de kullback-Leibler

Uso

1

```
kullbackDivergence (p)
```

Argumentos p

Uma distribuição de padrão ordinal.

bhattacharyyaDistance

Distância de Bhattacharyya de uma dada distribuição de probabilidade e a distribuição de probabilidade uniforme

Uso

1 `bhattacharyyaDistance(p)`

Argumentos

p

Uma distribuição de padrão ordinal.

Ccomplexity

Complexidade estatística de uma distribuição de probabilidade

Uso

1 `Ccomplexity(p)`

Argumentos

p

Uma distribuição de padrão ordinal.

timeSeries

Gráfico da série temporal

Uso

```
1 timeSeries(serie)
```

Argumentos

serie

Um vetor numérico onde estará instanciada a série temporal que deve ser avaliada pela função.

histogram

Histograma dos padrões de Bandt-Pompe de uma série temporal

Uso

```
1 histogram(serie, dimension, delay)
```

Argumentos

serie

Um vetor numérico onde estará instanciada a série temporal que deve ser avaliada pela função.

dimension

Dimensão dos padrões ordinais.

delay

Delay utilizado na formação dos padrões.

`patternsOnGraph`

Localiza os pontos de uma série temporal pertencentes a um certo padrão ordinal

Uso

```
1 patternsOnGraph (serie, dimension, delay, pattern)
```

Argumentos

serie

Um vetor numérico onde estará instanciada a série temporal que deve ser avaliada pela função.

dimension

Dimensão dos padrões ordinais.

delay

Delay utilizado na formação dos padrões.

pattern

Padrão que deverá ser analisado. Valor baseado nos padrões demonstrados no histograma.

entropyPlane

Plota o gráfico da entropia de uma certa série temporal ou seus particionamentos

Uso

```
1 entropyPlane (serie , partitions , dimension , delay ,
2               distribution , option , q)
```

Argumentos

serie

Um vetor numérico onde estará instanciada a série temporal que deve ser avaliada pela função.

partitions

Quantidade de partições que a série deve ser dividida para a análise.

dimension

Dimensão dos padrões ordinais.

delay

Delay utilizado na formação dos padrões.

distribution

Distribuição que deve ser utilizada. O parâmetro deverá ser 1 para a distribuição de Bandt and Pompe, caso contrário a distribuição que será aplicada será a de Bandt and Pompe weighth.

option

Entropia que deve ser analisada. O parâmetro deve ser 1 para a entropia de shannon, 2 para a entropia de Tsallis ou 3 para a entropia de Renyi. Caso contrário, deverá ser aplicada a min entropy.

q

(Pode não ser necessário dependendo da entropia selecionada) Ordem da entropia.

distancePlane

Plota o gráfico da distância estocástica de uma certa série temporal ou seus particionamentos

Uso

```
1 distancePlane<-function(serie,partition,dimension,delay,
2 optionD=1,optionP=1,q=1){
```

Argumentos

serie

Um vetor numérico onde estará instanciada a série temporal que deve ser avaliada pela função.

partition

Quantidade de partições que a série deve ser dividida para a análise.

dimension

Dimensão dos padrões ordinais.

delay

Delay utilizado na formação dos padrões.

optionD

Distância estocástica que deve ser analisada. Devem ser usados os seguintes valores para acessar as opções disponíveis: Euclidiana (1), euclidiana quadrática (2), manhattan (3), chebyshev (4), divergência de Kullback-Leibler (5), helinger (6), divergência de Jensen Shannon (7), wootters (8) e bhattacharyya (9)

optionP

Distribuição que deve ser utilizada. O parâmetro deverá ser 1 para a distribuição de Bandt and Pompe, caso contrário a distribuição que será aplicada será a de Bandt and Pompe weighth.

q (Pode não ser necessário dependendo da entropia selecionada) Ordem da entropia.

partitionMPR *Gráfico entropia complexidade de uma série temporal ou de seus particionamentos*

Uso

```
1 partitionMPR(serie, dimension, delay, partitions)
```

Argumentos

serie Um vetor numérico onde estará instanciada a série temporal que deve ser avaliada pela função.

dimension Dimensão dos padrões ordinais.

delay Delay utilizado na formação dos padrões.

partitions Quantidade de partições que a série deve ser dividida para a análise.

saxPlot *Representação gráfica da classificação da série de acordo com o método Symbolic Aggregate Approximation*

Uso

```
1 saxPlot(serie, letters, partitions)
```

Argumentos

<i>serie</i>	Um vetor numérico onde estará instanciada a série temporal que deve ser avaliada pela função.
<i>letters</i>	O número de letras.
<i>partitions</i>	Quantidade de partições que a série deve ser dividida para a análise.
PIP	<i>Pontos encontrados na série fornecida pela técnica Perceptually Important Points</i>

Uso

1 `PIP (serie, numberOfPoints)`

Argumentos

<i>serie</i>	Um vetor numérico onde estará instanciada a série temporal que deve ser avaliada pela função.
<i>numberOfPoints</i>	Números de pontos que devem ser encontrados pela função.

plotPAA

Exibe o gráfico dos valores adquiridos após o cálculo da Piecewise aggregate approximation

Uso

```
1 plotPAA(serie, partitions)
```

Argumentos

serie

Um vetor numérico onde estará instanciada a série temporal que deve ser avaliada pela função.

partitions

Quantidade de partições que a série deve ser dividida para a análise.

Referências bibliográficas

- L. Zunino A. F. Bariviera and O. A. Rosso. Crude oil market and geopolitical events: An analysis based on information-theory-based quantifiers. *Fuzzy Economic Review*, 21(1): 41–51, 2016. URL <https://EconPapers.repec.org/RePEc:fzy:fuzeco:v:21:y:2016:i:1:p:41-51>.
- A. L. L. Aquino, T. S. G. Cavalcante, E. S. Almeida, A. C. Frery, and O. A. Rosso. Characterization of vehicle behavior with information theory. *The European Physical Journal B: Condensed Matter and Complex Systems*, 88(10):257–269, Oct 2015. ISSN 1434-6036. DOI [10.1140/epjb/e2015-60384-x](https://doi.org/10.1140/epjb/e2015-60384-x). URL <http://dx.doi.org/10.1140/epjb/e2015-60384-x>.
- A. L. L. Aquino, H. S. Ramos, A. C. Frery, L. P. Viana, T. S. G. Cavalcante, and O. A. Rosso. Characterization of electric load with information theory quantifiers. *Physica A*, 465: 277–284, 2017. DOI [10.1016/j.physa.2016.08.017](https://doi.org/10.1016/j.physa.2016.08.017).
- C. Bandt and B. Pompe. Permutation entropy: A natural complexity measure for time series. *Physical review letters*, 88:174102, 05 2002.
- A. Bhattacharya, A. Muhuri, S. De, S. Manickam, and A. C. Frery. Modifying the yamaguchi four-component decomposition scattering powers using a stochastic distance. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 8(7): 3497–3506, July 2015. ISSN 1939-1404. DOI [10.1109/JSTARS.2015.2420683](https://doi.org/10.1109/JSTARS.2015.2420683).
- S. Blanco, S. Kochen, R. Q. Quiroga, L. Riquelme, O. A. Rosso, and P. Salgado. *Characterization of Epileptic EEG Time Series (I): Gabor Transform and Nonlinear Dynamics Methods*. Birkhäuser Boston, Boston, MA, 1997. ISBN 978-1-4612-2010-7. DOI [10.1007/978-1-4612-2010-7-9](https://doi.org/10.1007/978-1-4612-2010-7-9). URL <https://doi.org/10.1007/978-1-4612-2010-7-9>.
- P. J. Brockwell and R. A. Davis. *Time Series: Theory and Methods*. Springer-Verlag, Berlin, 2 edition, 1991.

- S.-H. Cha. Comprehensive survey on distance/similarity measures between probability density functions. *Int. J. Math. Model. Meth. Appl. Sci.*, 1, 01 2007.
- M. M. Deza and E. Deza. *Encyclopedia of Distances*. Springer Berlin Heidelberg, 2009. ISBN 978-3-642-00234-2. DOI [10.1007/978-3-642-00234-2](https://doi.org/10.1007/978-3-642-00234-2).
- D. P. Feldman and J. P. Crutchfield. Measures of statistical complexity: Why? *Physics Letters A*, 238(4):244–252, 1998. ISSN 0375-9601. DOI [https://doi.org/10.1016/S0375-9601\(97\)00855-4](https://doi.org/10.1016/S0375-9601(97)00855-4). URL <http://www.sciencedirect.com/science/article/pii/S0375960197008554>.
- D. P. Feldman, C. S. McTague, and J. P. Crutchfield. The organization of intrinsic computation: Complexity-entropy diagrams and the diversity of natural information processing. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, 18(4):043106, 2008. DOI [10.1063/1.2991106](https://doi.org/10.1063/1.2991106). URL <https://doi.org/10.1063/1.2991106>.
- J. Gambini, J. Casseti, M. M. Lucini, and A. C. Frery. Parameter estimation in sar imagery using stochastic distances and asymmetric kernels. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 8(1):365–375, Jan 2015. ISSN 1939-1404. DOI [10.1109/JSTARS.2014.2346017](https://doi.org/10.1109/JSTARS.2014.2346017).
- R. M. Gray. *Entropy and Information Theory*. Springer-Verlag, Berlin, Heidelberg, 1990. ISBN 0-387-97371-0.
- A. M. Kowalski, M. T. Martín, A. Plastino, O. A. Rosso, and M. Casas. Distances in probability space and the statistical complexity setup. *Entropy*, 13:1055–1075, 2011.
- PW Lamberti, M.T Martin, A Plastino, and O.A Rosso. Intensive entropic non-triviality measure. *Physica A: Statistical Mechanics and its Applications*, 334(1):119 – 131, 2004. ISSN 0378-4371. DOI <https://doi.org/10.1016/j.physa.2003.11.005>. URL <http://www.sciencedirect.com/science/article/pii/S0378437103010963>.
- Michael Lawrence and Duncan Temple Lang. Rgtk2: A graphical user interface toolkit for r. *Journal of Statistical Software*, 037(i08), 2010. URL <http://EconPapers.repec.org/RePEc:jss:jstsof:v:037:i08>.
- R. López-Ruiz, H. L. Mancini, and X. Calbet. A statistical measure of complexity. *Physics Letters A*, 209(5):321 – 326, 1995. ISSN 0375-9601. DOI [https://doi.org/10.1016/0375-9601\(95\)00867-5](https://doi.org/10.1016/0375-9601(95)00867-5). URL <http://www.sciencedirect.com/science/article/pii/0375960195008675>.
- M. T. Martin, A. Plastino, and O. A. Rosso. Generalized statistical complexity measures: Geometrical and analytical properties. *Physica A: Statistical Mechanics and its*

- Applications*, 369(2):439–462, 2006. ISSN 0378-4371.
DOI <https://doi.org/10.1016/j.physa.2005.11.053>. URL
<http://www.sciencedirect.com/science/article/pii/S0378437106001324>.
- L. De Micco, C. M. González, H. A. Larrondo, M. T. Martin, A. Plastino, and O. A. Rosso. Randomizing nonlinear maps via symbolic dynamics. *Physica A: Statistical Mechanics and its Applications*, 387(14):3373–3383, 2008. ISSN 0378-4371.
DOI <https://doi.org/10.1016/j.physa.2008.02.037>. URL
<http://www.sciencedirect.com/science/article/pii/S0378437108002318>.
- K. Mischaikow, M. Mrozek, J. Reiss, and A. Szymczak. Construction of symbolic dynamics from experimental time series. *Phys. Rev. Lett.*, 82:1144–1147, Feb 1999.
DOI [10.1103/PhysRevLett.82.1144](https://doi.org/10.1103/PhysRevLett.82.1144). URL
<https://link.aps.org/doi/10.1103/PhysRevLett.82.1144>.
- L. Pardo. *Statistical Inference Based on Divergence Measures*. Number 185 in Statistics, textbooks and monographs. Chapman & Hall/CRC, Boca Raton, 2006.
- G. E. Powell and I. Percival. A spectral entropy method for distinguishing regular and irregular motion of Hamiltonian systems. *Journal of Physics A: Mathematical and General*, 12:2053, 01 2001.
- M. G. Ravetti, L. C. Carpi, B. A. Gonçalves, A. C. Frery, and O. A. Rosso. Distinguishing noise from chaos: objective versus subjective criteria using Horizontal Visibility Graph. *PLOS ONE*, 9(9):1–15, 2014. **DOI** [10.1371/journal.pone.0108004](https://doi.org/10.1371/journal.pone.0108004).
- A. Rényi. On measures of entropy and information. In *4th Berkeley Symposium on Mathematical Statistics and Probability*, volume 1, pages 547–561, 1961.
- O. A. Rosso, S. Blanco, J. Yordanova, V. Kolev, A. Figliola, M. Schürmann, and E. Başar. Wavelet entropy: a new tool for analysis of short duration brain electrical signals. *Journal of Neuroscience Methods*, 105(1):65 – 75, 2001. ISSN 0165-0270.
DOI [https://doi.org/10.1016/S0165-0270\(00\)00356-3](https://doi.org/10.1016/S0165-0270(00)00356-3). URL
<http://www.sciencedirect.com/science/article/pii/S0165027000003563>.
- O. A. Rosso, H. Craig, and P. Moscato. Shakespeare and other english renaissance authors as characterized by information theory complexity quantifiers. *Physica A: Statistical Mechanics and its Applications*, 388(6):916 – 926, 2009. ISSN 0378-4371.
DOI <https://doi.org/10.1016/j.physa.2008.11.018>. URL
<http://www.sciencedirect.com/science/article/pii/S0378437108009461>.
- O. A. Rosso, R. Ospina, and A. C. Frery. Classification and verification of handwritten signatures with time causal information theory quantifiers. *PLOS ONE*, 11(12):e0166868, 2016. **DOI** [10.1371/journal.pone.0166868](https://doi.org/10.1371/journal.pone.0166868).

- M. Salicrú, M. L. Mendéndez, and L. Pardo. Asymptotic distribution of (h, ϕ) -entropy. *Communications in Statistics – Theory Methods*, 22(7):2015–2031, 1993.
- T. A. Schieber, L. Carpi, A. C. Frery, O. A. Rosso, P. M. Pardalos, and M. G. Ravetti. Information theory perspective on network robustness. *Physics Letters A*, 380:359–364, 2016. DOI [10.1016/j.physleta.2015.10.055](https://doi.org/10.1016/j.physleta.2015.10.055).
- C. E. Shannon. A mathematical theory of communication. *SIGMOBILE Mob. Comput. Commun. Rev.*, 5(1):3–55, January 2001. ISSN 1559-1662. DOI [10.1145/584091.584093](https://doi.org/10.1145/584091.584093). URL <http://doi.acm.org/10.1145/584091.584093>.
- M. Sinn and K. Keller. Estimation of ordinal pattern probabilities in Gaussian processes with stationary increments. *Computational Statistics & Data Analysis*, 55(4):1781–1790, 2011. DOI [10.1016/j.csda.2010.11.009](https://doi.org/10.1016/j.csda.2010.11.009).
- Leonardo Torres, Sidnei J. S. Sant’Anna, Corina da Costa Freitas, and Alejandro César Frery. Speckle reduction in polarimetric sar imagery with stochastic distances and nonlocal means. *Pattern Recognition*, 47:141–157, 2014.
- F. Traversaro, F. Redelico, M. Risk, A. C. Frery, and O. Rosso. Bandt-pompe symbolization dynamics for time series with tied values: a data-driven approach. *Chaos: an Interdisciplinary Journal of Nonlinear Science*, 2018. ISSN 1054–1500.
- C. Tsallis. Possible generalization of boltzmann-gibbs statistics. *Journal of Statistical Physics*, 52(1):479–487, Jul 1988. ISSN 1572-9613. DOI [10.1007/BF01016429](https://doi.org/10.1007/BF01016429). URL <https://doi.org/10.1007/BF01016429>.
- Delphine Zemp. *The Complexity of the Fraction of Absorbed Photosynthetically Active Radiation on a global scale*. PhD thesis, 06 2012.