

Universidade Federal de Campina Grande
Centro de Engenharia Elétrica e Informática
Programa de Pós-Graduação em Engenharia Elétrica

Abordagens da Teoria da Informação para análise de Causalidade e Conectividade em Processos Contínuos

Milena Marinho Arruda

Área de Concentração: Processamento da Informação

Orientador: Prof. Dr. Francisco Marcos de Assis

Orientadora: Profa. Dra. Luciana Ribeiro Veloso

Campina Grande, Paraíba, Brasil
©Milena Marinho Arruda, Julho de 2018

Abordagens da Teoria da Informação para análise de Causalidade e Conectividade em Processos Contínuos

Milena Marinho Arruda

Dissertação de Mestrado apresentada à Coordenação do Programa de Pós-Graduação em Engenharia Elétrica da Universidade Federal de Campina Grande, em cumprimento às exigências para obtenção do Grau de Mestre em Ciências no Domínio da Engenharia Elétrica.

Universidade Federal de Campina Grande
Centro de Engenharia Elétrica e Informática
Programa de Pós-Graduação em Engenharia Elétrica

Orientador: Prof. Dr. Francisco Marcos de Assis

Orientadora: Profa. Dra. Luciana Ribeiro Veloso

Campina Grande, Paraíba, Brasil
©Milena Marinho Arruda, Julho de 2018

A779a

Arruda, Milena Marinho.

Abordagens da Teoria da Informação para análise de causalidade e conectividade em processos contínuos / Milena Marinho Arruda. - Campina Grande, 2018.

67 f. : il. color.

Dissertação (Mestrado em Engenharia Elétrica) - Universidade Federal de Campina Grande, Centro de Engenharia Elétrica e Informática, 2018.

"Orientação: Prof. Dr. Francisco Marcos de Assis, Profª. Dra. Luciana Ribeiro Veloso".

Referências.

1. Máquina de Vetor de Suporte (SVM). 2. Entropia Diferencial. 3. Entropia de Transferência. 4. Lesão Cerebral. 5. Processos Contínuos. I. Assis, Francisco Marcos de. II. Veloso, Luciana Ribeiro. III. Título.

CDU 621.3(043)

**"ABORDAGENS DA TEORIA DA INFORMAÇÃO PARA ANÁLISE DE CAUSALIDADE E
CONECTIVIDADE EM PROCESSOS CONTÍNUOS"**

MILENA MARINHO ARRUDA

DISSERTAÇÃO APROVADA EM 31/07/2018



FRANCISCO MARCOS DE ASSIS, Dr., UFCG
Orientador(a)



LUCIANA RIBEIRO VELOSO, D.Sc., UFCG
Orientador(a)



BRUNO BARBOSA ALBERT, D.Sc., UFCG
Examinador(a)



BENEMAR ALENCAR DE SOUZA, D.Sc., UFCG
Examinador(a)



EDMAR CANDEIA GURJAO, D.Sc., UFCG
Examinador(a)

CAMPINA GRANDE - PB

À minha irmã Aline Marinho Arruda.

Agradecimentos

Agradeço, primeiramente, aos meus pais, Alex e Gildete, por acreditarem sempre em mim e nas minhas responsabilidades perante aos compromissos da vida, e, sobretudo, obrigada pela lição de amor que me ensinaram. À minha irmã, Aline, afinal, ter uma irmã é ter, pra sempre, uma vida lembrada com segurança em outro coração.

À minha família e amigos que vibraram minhas conquistas, carregando a certeza de que nunca estarei só. Vocês foram essenciais para que meus dias se tornassem mais leves e divertidos. A todas as pessoas que muitas vezes, anonimamente, fizeram uma diferença enorme na minha vida.

Aos voluntários e voluntárias do Ramo Estudantil IEEE UFCG, em especial do IEEE Women in Engineering UFCG, com os quais descobri que é possível desenvolver habilidades de liderança, comunicação e trabalho em grupo, mesmo em uma rotina diferenciada quanto esta que enfrentamos no mestrado acadêmico. Com eles também consegui formar uma rede de frIEEEnds, ao redor do Brasil e do mundo, com os quais aprendi a importância de devolver à sociedade todo o conhecimento que adquirimos ao longo de nossa formação.

À todos os Professores, mestres e doutores que foram responsáveis por minha formação, dedicando seu apoio, atenção, paciência, amizade e compreensão. Aos Professores do Departamento de Engenharia Elétrica que contribuíram com o meu crescimento, dando lições não só técnicas, mas também de vida. Em especial, agradeço aos meus orientadores, Professor Francisco Marcos e Professora Luciana Veloso, pela confiança, por estar sempre à disposição moldando o meu conhecimento e, além de tudo, pelos conselhos sobre a vida pessoal e profissional que eu levarei para sempre.

Ao Instituto de Estudos em Computação e Informação Quânticas (IQuanta), pela estrutura física necessária para a realização desse trabalho. Ao Programa de Pós-Graduação em Engenharia Elétrica (PPgEE - COPELE) da UFCG, pelo suporte administrativo durante o período do mestrado. À CAPES, pelo suporte financeiro para o desenvolvimento dessa dissertação.

*“Talvez um dia, para além dos dias,
Encontres o que queres porque o queres.”
Fernando Pessoa*

Resumo

A teoria da informação consolida abordagens matemáticas para medir a influência causal entre séries temporais multivariadas. Embora as formulações matemáticas das medidas de informação sejam objetivas, na prática, sua estimação pode ser um processo complexo, uma vez que os estimadores assumem a falta de conhecimento das propriedades estatísticas dos processos estocásticos analisados. Atualmente, algumas das áreas com aplicação de medidas de informação são: economia, neurociência, diagnóstico biomédico, detecção de conectividade em plantas de processo industrial, e assim por diante. Nesta dissertação, um método é proposto para caracterização de suporte efetivo para a aplicabilidade do método de aprendizado supervisionado, Máquina de Vetor de Suporte (SVM, do inglês: *Support Vector Machine*), para estimação de entropia de variáveis aleatórias contínuas. Esse método fornece resultados assintoticamente bons quando comparados com valores analíticos e valores estimados a partir de técnicas de histograma, suavização de kernel e distâncias de vizinhos. Além disso, discute-se o uso da entropia de transferência, estimada a partir do estimador baseado em distâncias entre vizinhos, como um auxílio no diagnóstico biomédico em casos de lesões cerebrais e na detecção da conectividade entre equipamentos industriais em um sistema com quatro tanques conectados.

Palavras-chave: SVM. Entropia diferencial. Entropia de transferência. Lesão Cerebral. Processos industriais.

Abstract

Information theory consolidates a mathematical approach to measure the causal influence among multivariate time series. Although the concepts in information theory are relatively simple and mathematical formulations are objective, in practice, their estimation can be a complex process, since the estimators have few or none knowledge about the statistical properties of the stochastic processes. Currently, some of the areas with application of information measures are: economics, neuroscience, biomedical diagnosis, connectivity detection in industrial process. In this dissertation, a method is proposed for characterization of effective support for the applicability of the supervised learning method, Support Vector Machine (SVM), in estimation of entropy of continuous random variables. This method provides asymptotically good results when compared to analytical values and values estimated from histogram techniques, kernel smoothing and distances of nearest neighbors. In addition, we discuss the use of transfer entropy, estimated from the estimator based on distances between nearest neighbors, as an assist in the biomedical diagnosis in cases of brain lesions and in the detection of connectivity in a system with four tanks.

Keywords: SVM. Differential entropy. Transfer entropy. Traumatic brain injury. Industrial processes.

Lista de Ilustrações

Figura 1.1 – Grafos de mapa causal direcional para relações entre os processos \mathbf{X} e \mathbf{Y} a fim de identificar se a relação é: a) ao longo de um caminho direto ou indireto através de \mathbf{Z} ou; b) a partir de uma fonte comum, \mathbf{Z} .	21
Figura 2.1 – Estimativa da função densidade de probabilidade por meio das técnicas baseadas em: a) histograma e; b) densidade kernel (curva azul).	29
Figura 2.2 – Representação gráfica da interpretação do estimador Kozachenko-Leonenko considerando $k = 3$ então têm-se pelo menos uma amostra com distância $r \in [\epsilon/2, \epsilon/2 + d\epsilon/2]$ até x_i (em azul), $k - 1$ amostras com distâncias menores que r (em vermelho) e, $N - k - 1$ amostras com distâncias maiores que r (branco).	30
Figura 2.3 – Classificação baseada no algoritmo SVM de uma classe: amostras de treinamento são separados no espaço característico por meio de vetores de suporte.	31
Figura 2.4 – Função densidade de probabilidade para uma variável aleatória X com distribuição exponencial, gamma, normal e uniforme. A área preenchida indica a região com densidade menor que 15% da densidade máxima.	33
Figura 2.5 – Estimação de entropia considerando 10^5 realizações de variáveis aleatórias exponencialmente distribuídas: $X_1 \sim \text{Exp}(2)$ com $h(X_1) = 1,69 \text{ nats}$, curvas mais acima e; $X_2 \sim \text{Exp}(1)$ com $h(X_2) = 1 \text{ nat}$, curvas mais abaixo.	35
Figura 2.6 – Estimação de entropia considerando 10^5 realizações de variáveis aleatórias gamma distribuídas: $X_1 \sim \Gamma(3; 2)$ com $h(X_1) = 2,09 \text{ nats}$, curvas mais acima e; $X_2 \sim \Gamma(0,5; 1)$ com $h(X_2) = 0,09 \text{ nats}$, curvas mais abaixo.	35
Figura 2.7 – Estimação de entropia considerando 10^5 realizações de variáveis aleatórias normalmente distribuídas: $X_1 \sim \mathcal{N}(2; 3)$ com $h(X_1) = 2,51 \text{ nats}$, curvas mais acima e; $X_2 \sim \mathcal{N}(0; 1)$ com $h(X_2) = 1,41 \text{ nats}$, curvas mais abaixo.	36

Figura 2.8 – Estimação de entropia considerando 10^5 realizações de variáveis aleatórias uniformemente distribuídas: $X_1 \sim \mathcal{U}(-1; 1)$ com $h(X_1) = 0,69 \text{ nats}$, curvas mais acima e; $X_2 \sim \mathcal{U}(0; 0,5)$ com $h(X_2) = -0,69 \text{ nats}$, curvas mais abaixo.	36
Figura 3.1 – Grafos de mapa causal direcional para os processos \mathbf{X} , \mathbf{Y} e \mathbf{Z} quando apenas $T_{\mathbf{X} \rightarrow \mathbf{Y}}$, $T_{\mathbf{X} \rightarrow \mathbf{Z}}$ e $T_{\mathbf{Z} \rightarrow \mathbf{Y}}$ são maiores que zero. Além disso, as possíveis situações de conectividade dos processos acontecem quando: (a) $D_{\mathbf{X} \rightarrow \mathbf{Y}} > 0$; (b) $D_{\mathbf{X} \rightarrow \mathbf{Y}} > 0$ e $D_{\mathbf{Z} \rightarrow \mathbf{Y}} = 0$; (c) $D_{\mathbf{X} \rightarrow \mathbf{Y}} = 0$	41
Figura 3.2 – Determinação dos parâmetros para os estimadores propostos por Krasnikov, Stögbauer e Grassberger. (a) Considerando o primeiro algoritmo com $k = 1$, $n_x = 5$ e $n_y = 3$; (b) Considerando o segundo algoritmo com $k = 2$, $n_x = 5$ e $n_y = 4$	43
Figura 3.3 – Grafo direcional do fluxo de informação para o modelo autorregressivo tridimensional (B.1) com $\alpha = 0.8$, $\beta = 0.6$, $\gamma = 0.4$, $\delta = 0.5$ e $\epsilon = 0$. Os dois primeiros grafos são resultantes do cálculo da entropia de transferência que representa a causalidade total do sistema. O último grafo é obtido após a análise com a entropia de transferência direta e corresponde à causalidade direta e verdadeira, interpretada como a conectividade do sistema.	45
Figura 3.4 – Convergência do estimador de entropia de transferência baseado nas distâncias entre vizinho para para as relações com valores analíticos maiores que zero: $T_{\mathbf{X} \rightarrow \mathbf{Y}} = 0,2365$, $T_{\mathbf{Y} \rightarrow \mathbf{Z}} = 0,2516$ and $T_{\mathbf{X} \rightarrow \mathbf{Z}} = 0,0666$	46
Figura 3.5 – Esquemático do sistema com quatro tanques (reservatórios) com água e uma bomba cuja taxa de fluxo da água varia com o valor médio $4 \text{ m}^3/\text{s}$ e variância $0.1 \text{ m}^3/\text{s}$. em um comportamento gaussiano.	48
Figura 3.6 – Grafo direcional do fluxo de informação para o sistema com quatro tanques com água resultante do cálculo da entropia de transferência que representa a causalidade total do sistema.	49
Figura 3.7 – Grafos direcionais do fluxo de informação para o sistema com quatro tanques com água. O primeiro grafo é resultante do cálculo da entropia de transferência que representa a causalidade total do sistema. Os grafos seguintes são obtidos sequencialmente a partir da análise de conectividade das variáveis que estão interligadas por meio de uma seta vermelha. O último grafo corresponde à causalidade direta e verdadeira, interpretada como a conectividade do sistema.	50
Figura 3.8 – Sinais de pressão sanguínea (ABP) e eletrocardiograma (ECG) para indivíduos saudáveis.	52
Figura 3.9 – Sinais de pressão sanguínea (ABP) e eletrocardiograma (ECG) para indivíduos diagnosticados com traumatismo craniano e lesões associadas.	52

Figura 3.10–Gráfico de barra de erro vertical para a média e o desvio padrão médio das medidas de entropia de transferência para indivíduos saudáveis e indivíduos diagnosticados com TBI e lesões associadas. 54

Lista de Tabelas

Tabela 2.1 – Entropia diferencial de Shannon para as distribuições exponencial, gamma, normal e uniforme.	33
Tabela 2.2 – Parâmetros ν para as distribuições exponencial, gamma, normal e uniforme.	33
Tabela 3.1 – Valores estimados de entropia de transferência usando o estimador (3.12) para o sistema autorregressivo tridimensional (B.1) com $\alpha = 0.8$, $\beta = 0.6$, $\gamma = 0.4$, $\delta = 0.5$ e $\epsilon = 0$	46
Tabela 3.2 – Parâmetros de operação dos quatro tanques que compõem o sistema da Figura 3.5 em que $i = 1, 2, 3, 4$	47
Tabela 3.3 – Valores estimados de entropia de transferência usando o estimador (3.12) para o sistema com quatro tanques ilustrado na Figura 3.5.	47
Tabela 3.4 – Entropia de transferência direta para relações direcionais que podem ter uma ou mais variáveis intermediárias.	49
Tabela 3.5 – Média e desvio padrão de 10 amostras, cada uma com 30 segundos, para entropia de transferência (em bits). São considerados os dois possíveis sentidos de fluxo de informação entre os sinais de ABP e ECG para cada voluntário saudável que compõe a base de dados PRCP.	53
Tabela 3.6 – Média e desvio padrão de 10 amostras, cada uma com 30 segundos, para entropia de transferência (em bits). São considerados os dois possíveis sentidos de fluxo de informação entre os sinais de ABP e ECG para cada voluntário diagnosticado com TBI ou lesões associadas que compõe a base de dados CHARIS DB.	53

Lista de Abreviaturas e Siglas

ABP	<i>Arterial Blood Pressure</i>
AEP	<i>Asymptotic Equipartition Property</i>
CTW	<i>Context-Tree Weighting</i>
CHARIS	<i>Cerebral Haemodynamic Autoregulatory Information System</i>
DATASUS	Departamento de Informática do Sistema Único de Saúde
ECG	Eletrocardiograma
MEG	Magnetoencefalografia
KDE	<i>Kernel Density Estimator</i>
KL	Kozachenko-Leonenko
MST	<i>Minimum Spanning Tree</i>
PDDL	<i>Public Domain Dedication an License</i>
PRCP	<i>Physiologic Response to Changes in Posture</i>
RBF	<i>Radial Basis Function</i>
SBRT	Simpósio Brasileiro de Telecomunicações e Processamento de Sinais
SVM	<i>Support Vector Machine</i>
TBI	<i>Traumatic Brain Injury</i>
WFDB	<i>WaveForm DataBase</i>

Lista de Símbolos

X, Y, Z	Variáveis aleatórias contínuas
$\mathbf{X}, \mathbf{Y}, \mathbf{Z}$	Processo estocástico
X_{n-k+1}^n	Sequência de comprimento k : X_{n-k+1}, \dots, X_n
X_n	n -ésima saída de um processo
$f(\cdot)$	Função densidade de probabilidade
$\hat{f}(\cdot)$	Estimativa da função densidade de probabilidade
$H(\cdot)$	Entropia
$\hat{H}(\cdot)$	Estimativa da entropia
$h(\cdot)$	Entropia diferencial
$\hat{h}(\cdot)$	Estimativa da entropia diferencial
$T_{\mathbf{X} \rightarrow \mathbf{Y}}$	Entropia de transferência de \mathbf{X} para \mathbf{Y}
$\hat{T}_{\mathbf{X} \rightarrow \mathbf{Y}}$	Estimativa da entropia de transferência de \mathbf{X} para \mathbf{Y}
$D_{\mathbf{X} \rightarrow \mathbf{Y}}$	Entropia de transferência direta de \mathbf{X} para \mathbf{Y}
$\hat{D}_{\mathbf{X} \rightarrow \mathbf{Y}}$	Estimativa da entropia de transferência direta de \mathbf{X} para \mathbf{Y}
$[\cdot]$	Suporte
$\psi(\cdot)$	Função <i>digamma</i>

Sumário

1	Introdução	18
1.1	Motivação	18
1.1.1	Aplicações em Biomédica	21
1.1.2	Aplicações em Processos Industriais	22
1.2	Notação e Terminologia	22
1.3	Contribuições	23
1.4	Organização do texto	23
2	Estimação de Entropia Diferencial	24
2.1	Introdução	24
2.2	Entropia Diferencial	25
2.3	Estimadores	27
2.3.1	Histograma	28
2.3.2	Estimador de Densidade Kernel	28
2.3.3	Estimador Kozachenko-Leonenko	30
2.3.4	SVM de uma classe	31
2.4	Resultados	32
2.5	Considerações	37
3	Abordagens para Detecção de Causalidade em Sistemas Contínuos	38
3.1	Introdução	38
3.2	Entropia de Transferência	39
3.3	Estimadores	42
3.4	Resultados	44
3.4.1	Modelo Autorregressivo Tridimensional	44
3.4.2	Causalidade e Conectividade em Sistema com Quatro Tanques	47
3.4.3	Identificação de Padrão para Diagnóstico de Traumatismo Craniano	51
3.5	Considerações	55
4	Conclusões	56
4.1	Trabalhos Futuros	57
4.2	Produção Científica	58
	Referências	60
	APÊNDICE A Estimadores de Medidas de Informação	64

A.1	Estimador Kozachenko-Leonenko	64
A.2	Estimador SVM de uma classe	65
APÊNDICE B Modelo Tridimensional Autorregressivo		68

Capítulo 1

Introdução

Neste capítulo serão apresentados a motivação, a notação e terminologia, as contribuições e a organização do texto desta dissertação. No cenário da motivação inclui-se a revisão bibliográfica, bem como o estado da arte, a fim de auxiliar nos fundamentos desenvolvidos ao longo desta pesquisa. A notação e terminologia tem o intuito de padronizar as discussões subsequentes. E, as contribuições visam identificar os pontos nos quais esta dissertação colabora com a comunidade científica.

1.1 Motivação

Apesar de ser uma ciência relativamente recente, a teoria da informação, originalmente proposta por Claude Shannon em 1948 com a publicação do artigo “*A Mathematical Theory of Communication*”, consolida uma abordagem matemática para o estudo de três principais áreas: armazenamento, comunicação e quantificação de informação [1]. Nesse contexto, o significado técnico da palavra “informação” não difere totalmente do significado comum. Se uma mensagem é conhecida, então a quantidade de informação que a mesma transmite é nula.

Claude Shannon introduz os conceitos de entropia, entropia diferencial e informação mútua a partir do estudo da compressão e transmissão de dados [1]. Além disso, medidas de informação como distância de Kullback-Leibler [2], informação direcional [3, 4] e entropia de transferência [5], desempenham um papel importante na teoria da informação. A escolha da base logarítmica para essas medidas determina a unidade da grandeza, sendo as mais comuns: bit (logaritmo na base 2) e nat (logaritmo na base neperiana).

Embora os conceitos da teoria da informação e as formulações matemáticas sejam objetivas, na prática, sua estimação pode ser um processo complexo. Portanto, esforços científicos são destinados ao estudo de técnicas que refinem estas estimativas. Em geral, recorre-se a estimadores que assumem o desconhecimento das propriedades estatísticas das variáveis aleatórias envolvidas.

Limitações como o número finito de amostras e propriedades intrínsecas dos dados da série temporal, podem resultar em estimativas divergentes do valor verdadeiro do parâmetro, neste caso, diz-se que o estimador é enviesado, tendencioso ou viciado. Para variáveis discretas, a abordagem para estimativa numérica pode ser simples e feita a partir da contagem das configurações para cada probabilidade a ser estimada, também chamada de *plug-in*, ou mais complexa a partir do algoritmo da árvore de contexto ponderada (CTW, do inglês: *Context-Tree Weighting*) [6, 7].

Contudo, o processo de estimação das medidas de informação para variáveis contínuas deve ser mais cuidadoso. Embora seja possível discretizar os dados e usar os estimadores para variáveis discretas, essa metodologia pode comprometer a precisão dos resultados. Portanto, ao longo das últimas décadas, foram propostos algoritmos com foco na estimação dessas medidas. Esses algoritmos podem depender da estimativa da função densidade de probabilidade, ou ser estimadores diretos da medida de informação.

A abordagem mais simples e comum é baseada em histograma, que representa graficamente através de barras (retângulos) um conjunto de dados dividido em classes. Sendo assim, a distribuição de frequência tende a uma distribuição de densidade de probabilidades quando o volume de dados aumenta indefinidamente e o intervalo de classes tende a zero. Essa estimação é enviesada, portanto, alguns métodos para correção foram desenvolvidos, por exemplo, o *Panzeri-Treves* [8] e extrapolação quadrática [9].

Outro estimador de densidade de probabilidade é o de densidade kernel (KDE, do inglês: *Kernel Density Estimator*) que depende essencialmente de dois parâmetros: função de suavização e largura de banda para controle da suavidade da curva de densidade resultante. Bowman e Azzalini (1997) [10] discutem técnicas para determinar de forma ótima esses parâmetros.

Estimar diretamente a medida de informação, isto é, sem necessariamente estimar a função densidade de probabilidade da variável, pode reduzir os problemas de enviesamento. Têm-se, então, os estimadores que se baseiam nas distâncias e contagem de vizinhos e estimam: entropia diferencial [11, 12], informação mútua [13, 14] e entropia de transferência [15]. Além disso, a estimação de suporte, a partir da técnica de aprendizado supervisionado de uma classe da máquina de vetor de suporte (SVM, do inglês: *Support Vector Machine*), também têm sido explorada [16, 17].

Nesse contexto, a aplicabilidade do SVM foi investigada para estimação da densidade de variáveis [18, 19] e estimação de medidas de informação considerando a formulação de Alfréd Rényi (1961) [20] para entropia diferencial de ordem zero [17]. A propriedade de equipartição assintótica (AEP, do inglês: *Asymptotic Equipartition Property*) garante que a diferença entre entropia diferencial de Rényi de ordem zero e entropia diferencial de Shannon está na caracterização do suporte [21]. Sendo assim, dependendo da caracterização desse suporte, a estimação pode convergir para quaisquer uma das duas entropias.

A possibilidade de utilizar o algoritmo SVM de uma classe para estimação da entropia diferencial de Shannon é apontada originalmente por Schölkopf et al. (2001) [16]. Contudo, na literatura ainda existe espaço para investigar metodologias de caracterização do conjunto de suporte efetivo a fim de expandir o uso do algoritmo SVM.

Atualmente, essas medidas e técnicas de estimação tem sido aplicadas em áreas como economia [7], neurociência [22, 23], diagnóstico biomédico [24, 25], detecção de conectividade em plantas de processos industriais [26, 27, 28] e etc. Nessa dissertação, investiga-se a convergência de estimadores para entropia diferencial e discute-se o uso da entropia de transferência para detecção de causalidade e conectividade em sistemas de processos contínuos.

Contudo, identificar as relações causais entre componentes de um sistema a partir de observações de séries temporais é desafiador quando o conhecimento sobre a dinâmica subjacente é parcial. Embora essas relações possam ser detectadas, também, por meio da modelagem matemática, para processos complexos em grande escala é difícil estabelecer modelos matemáticos práticos e precisos.

A teoria da informação fornece uma variedade de abordagens para medir a influência causal entre séries temporais multivariadas. A primeira definição de causalidade foi proposta por Wiener (1956) [29] na qual, um processo é dito causar outro se a capacidade de prever o segundo é melhorada pela incorporação de informações sobre o primeiro. Posteriormente, essa definição foi formalizada por Granger (1969) [30] que aplicou principalmente para análises na economia.

Supondo dois processos estocásticos \mathbf{X} e \mathbf{Y} , considera-se duas situações: 1) tentar prever X_{n+1} usando apenas termos passados de \mathbf{X} e; 2) tentar prever X_{n+1} usando termos passados de \mathbf{X} e \mathbf{Y} . Se a segunda situação atinge significativamente mais êxito, então o passado de \mathbf{Y} contém informação útil para prever X_{n+1} . Neste caso, \mathbf{Y} é dito causar \mathbf{X} no sentido da causalidade de Wiener e Granger [31].

Barnett, Barrett e Seth (2009) [32] demonstrou a equivalência entre a causalidade no sentido de Wiener e Granger e a entropia de transferência para variáveis com distribuições gaussianas. Ambas as medidas fornecem valiosas informações estatísticas sobre os sinais observados e sua conectividade.

Denomina-se de conectividade a identificação das vias de propagação direta de informação em um sistema. Ou seja, detectar a conectividade de um sistema significa detectar se a influência causal entre um par de variáveis é ao longo de uma via direta sem quaisquer variáveis intermediárias, ou ao longo de vias indiretas através de variáveis intermediárias, Figura 1.1. Em geral, a conectividade dos sistemas é representada por meio de grafos direcionais.



Figura 1.1 – Grafos de mapa causal direcional para relações entre os processos \mathbf{X} e \mathbf{Y} a fim de identificar se a relação é: a) ao longo de um caminho direto ou indireto através de \mathbf{Z} ou; b) a partir de uma fonte comum, \mathbf{Z} .

1.1.1 Aplicações em Biomédica

Recorrer aos métodos da teoria da informação pode ser valioso para a solução de problemas de processamento de dados biomédicos. Avanços nas pesquisas em neurociência e medicina, como por exemplo, detecção de interações cérebro-máquina [33], é um potencial no suporte do diagnóstico médico e no desenvolvimento de tecnologias que tragam segurança e comodidade à saúde das pessoas.

Segundo Blokh e Stambler (2014) [34], dados biomédicos são processos complexos com características não-lineares, estocásticas e não-analíticas, além disso, essas características são determinadas pela própria natureza do objeto em estudo: o organismo humano. Portanto, embora os processos não possam ser adequadamente descritos por modelos simples, a modelagem adequada é crucial para o processamento dos dados.

Na área de biomédica as aplicações são diversas e as contribuições visam oferecer mais uma ferramenta para o suporte no diagnóstico clínico. Vicente et al. (2011) [25] investiga a aplicabilidade da entropia de transferência como uma medida em um teste para conectividade efetiva de dados eletrofisiológicos baseados em simulações e gravações de magnetoencefalografia (MEG) em uma tarefa motora simples.

Buscando um método livre de modelo, Marzbanrad et al. (2015) [24] utiliza a entropia de transferência para quantificar os acoplamentos da frequência cardíaca materna e fetal em ambas as direções, revelando informações que podem fornecer novos marcadores clínicos de desenvolvimento fetal e bem-estar.

Gao et al. (2018) [35] analisa acoplamentos eletromiográficos no acidente vascular cerebral a partir da entropia de transferência simbólica. O autor conclui que a abordagem de análise de entropia de transferência simbólica de escala variável proposta pode ser usado para caracterizar quantitativamente as características de sincronização não linear e a interação de informações entre o córtex motor primário e os músculos.

Embora a entropia de transferência assuma um espaço importante para o desenvolvimento de pesquisas nessa área, medidas como informação mútua, informação direcional e causalidade de Granger, também são investigadas e aplicadas [36, 23, 7, 22, 37].

1.1.2 Aplicações em Processos Industriais

Com o atual estágio da indústria e com o desenvolvimento da automação e da mecânica vivencia-se a indústria 4.0 ou quarta revolução industrial. A indústria 4.0 acontece após três processos históricos transformadores [38], entre os quais: a primeira revolução industrial que marcou o ritmo da produção manual a mecanizada, entre 1760 e 1830; a segunda revolução industrial, por volta de 1850, que trouxe a eletricidade e permitiu a manufatura em massa, e; a revolução digital, em meados do século 20, com a chegada da eletrônica, da tecnologia da informação e da telecomunicação.

Mecanismos para automatizar a detecção e o diagnóstico de anormalidades e distúrbios em toda a planta são desafios importantes na indústria de processos, além de ser alvo de estudo da indústria 4.0 de modo a ter um processo flexível, no qual, a planta é dinamicamente alterada de acordo com a demanda.

Em um processo industrial complexo, os elementos não estão apenas conectados entre si, também são mutuamente dependentes [17]. Portanto, é possível utilizar o conceito de causalidade para descrever as relações causa-efeito entre variáveis ou eventos através de grafos direcionais, estabelecendo uma rede chamada de mapa causal [28].

Publicações recentes exploram a aplicabilidade de medidas de informação, principalmente da entropia de transferência, no setor industrial. Yao et al. (2017) [39] propõe o estudo da transferência de informação entre os dados de consumo de eletricidade entre indústrias com base na entropia de transferência e no modelo árvore de abrangência mínima (MST, do inglês: *Minimum Spanning Tree*).

Em dois artigos, Duan et al. (2013, 2015) [27, 17] propõe métodos para explorar aplicações em plantas de processos industriais utilizando entropia de transferência a partir de metodologias e algoritmos diferentes, entre os quais: estimador kernel e estimador SVM de uma classe. A aplicabilidade do método foi testada a partir de simulações que incluem desde modelos autoregressivos (AR) até o estudo de caso de processos industriais.

Além disso, em 2014, Yang et al. [26] publicou o livro intitulado "*Capturing Connectivity and Causality in Complex Industrial Processes*" que discute sobre as análises de causalidade e conectividade em processos industriais complexos.

1.2 Notação e Terminologia

Com o intuito de padronizar nossas discussões, nesta dissertação, denotamos variáveis aleatórias com letras maiúsculas, processos estocásticos com letras maiúsculas em negrito, seus alfabetos com letras caligráficas (por exemplo, \mathcal{X} denota o alfabeto da variável aleatória X) e sua função de densidade de probabilidade com $f(\cdot)$. Sobrescritos de uma variável aleatória indicam sequências de comprimento n e é representado por

$X^n = \{X_1, X_2, \dots, X_n\}$. A n -ésima saída de um processo é indicada por subscritos, X_n . A sequência de comprimento finito com a ordem (k) de um processo estocástico é denotado por sobrescritos, $X_{n-k+1}^n = \{X_{n-k+1}, \dots, X_n\}$. Ao longo deste trabalho, log refere-se à base neperiana, e, neste caso, a informação medida será em nats.

1.3 Contribuições

As principais contribuições nesta dissertação que a distingue de outros trabalhos estão listadas abaixo:

1. Proposta de caracterização do suporte de uma variável aleatória para utilização do algoritmo SVM de uma classe na estimação da entropia diferencial de Shannon.
2. Avaliação da convergência de quatro estimadores, para entropia diferencial de Shannon, baseados em técnicas de: histograma, suavização kernel, distâncias de vizinhos e aprendizado supervisionado.
3. Discussão sobre metodologias de suporte para decisão clínica em casos de lesões cerebrais.
4. Demonstração da aplicabilidade do estimador baseado nas distâncias entre vizinhos para entropia de transferência na detecção de causalidade e conectividade em sistemas industriais.

1.4 Organização do texto

No capítulo 2 será apresentado o uso do método de aprendizado supervisionado, Máquina de Vetor de Suporte (SVM, do inglês: *Support Vector Machine*), para estimação de entropia diferencial de Shannon em comparação com valores analíticos e outros valores estimados a partir de estimadores baseados em: histograma, suavização kernel e distâncias de vizinhos

No capítulo 3 será discutido a utilização da entropia de transferência como abordagem para detecção da causalidade e da conectividade em sistemas contínuos. Além disso, será demonstrado a aplicabilidade da medida estimada a partir do estimador baseado em distâncias dos vizinhos no auxílio de diagnóstico biomédico e na detecção de conectividade entre equipamentos industriais.

Por fim, no capítulo 4 serão apresentadas as conclusões gerais dessa dissertação e, serão apontados os trabalhos futuros em relação ao estudo sobre estimadores de medidas da Teoria da Informação para variáveis contínuas, e em relação à importância do estudo sobre detecção de causalidade e conectividade em sistemas contínuos.

Capítulo 2

Estimação de Entropia Diferencial

Neste capítulo será apresentado o uso do método de aprendizado supervisionado, máquina de vetor de suporte (SVM, do inglês: *Support Vector Machine*), para estimação de entropia de variáveis aleatórias contínuas. O método baseia-se na estimativa do suporte da função densidade de probabilidade e, dependendo da caracterização deste suporte, pode convergir tanto para entropia de Shannon quanto para entropia de Rényi de ordem zero. Resultados obtidos a partir de simulações indicam que o método proposto para caracterização do suporte fornece resultados assintoticamente bons para a estimação da entropia de Shannon. Esses resultados foram comparados a valores analíticos e valores estimados a partir das técnicas de: histograma, suavização kernel e distâncias de vizinhos.

2.1 Introdução

A entropia é o conceito mais comum da teoria da informação, ciência originalmente proposta em 1948 por Claude Shannon [1]. Em geral, a entropia refere-se ao grau de desordem, incerteza ou quantidade de informação. Algo é dito informativo quando transmite mensagens pouco esperadas ou desconhecidas. Quantitativamente, mais informação é transmitida a partir de mensagens menos prováveis, ou seja, quando a mensagem transmitida é conhecida, então a quantidade de informação recebida é nula.

Contudo, o interesse sobre medidas de informação é anterior ao trabalho de Claude Shannon que tratou as medidas de informação em um sentido estatístico. Em 1928, Ralph Hartley [40] sugeriu que a quantidade de informação de uma mensagem está associada ao logaritmo do número de mensagens possíveis. Anos depois, em 1961, Alfréd Rényi [20] propõe uma generalização destas medidas atribuindo-as uma ordem e servindo como base para o conceito de dimensões generalizadas.

Embora os conceitos da teoria da informação sejam relativamente simples e as formulações matemáticas sejam objetivas, na prática, sua estimação pode ser um processo complexo. Estimadores para variáveis contínuas baseados em histograma e em funções

kernel de suavização são amplamente usados para estimar a entropia diferencial. Entretanto, quando aplicados para estimação de outras medidas podem produzir viés [13, 37]. Tais problemas são minimizados utilizando estimadores baseados nas distâncias de vizinhos tanto para estimação de entropia quanto de outras mediadas [11, 13].

Além disso, o algoritmo de uma classe do SVM foi desenvolvido para estimação de suporte [16]. Contudo, embora alguns autores abordem a estimação da densidade de variáveis [18, 19], e diagnóstico de causalidade em plantas industriais com base na entropia de Rényi de ordem zero [17], o algoritmo ainda não foi amplamente aplicado para estimação de entropia diferencial de Shannon.

A propriedade de equipartição assintótica (AEP, do inglês: *Asymptotic Equipartition Property*), para variáveis contínuas, interpreta a entropia diferencial de Shannon como o logaritmo do menor volume (suporte) que contém amostras com maior densidade [21], e, assim, viabiliza o uso do algoritmo SVM de uma classe para estimação da entropia de Shannon. Isso porque, a diferença entre entropia diferencial de Rényi de ordem zero (logaritmo da medida de Lebesgue [41] do suporte) e entropia diferencial de Shannon (logaritmo do suporte efetivo) está na caracterização do suporte.

2.2 Entropia Diferencial

Embora uma generalização para entropia tenha sido proposta por Alfréd Rényi, o conceito mais comum na Teoria da Informação foi estabelecido por Claude Shannon. De acordo com Shannon, a entropia é uma medida do número de bits necessários, em média, para descrever variáveis aleatórias discretas e, a entropia diferencial está relacionada ao logaritmo do volume efetivo no qual variáveis aleatórias contínuas estão inseridas. Ao contrário da entropia, a entropia diferencial pode ser negativa e ocorre quando esse volume é menor que um.

Definição 2.1 *A entropia de Rényi de ordem α de uma variável aleatória contínua X com função densidade de probabilidade $f(x)$ é definida como*

$$h_\alpha(X) = \frac{1}{1-\alpha} \log \left[\int f^\alpha(x) dx \right], \quad (2.1)$$

em que $0 < \alpha < \infty$, $\alpha \neq 1$ e \log é o logaritmo natural e expressa a entropia em nats. Quando $\alpha \rightarrow 1$ têm-se a entropia de Shannon

$$h(X) = h_1(X) = - \int f(x) \log f(x) dx, \quad (2.2)$$

e quando $\alpha \rightarrow 0$, têm-se a entropia de Rényi de ordem zero

$$h_0(X) = \log \phi(\llbracket X \rrbracket), \quad (2.3)$$

em que $\phi(\cdot)$ é a medida de Lebesgue e $\llbracket \cdot \rrbracket$ é o conjunto no qual $f(x) > 0$ (suporte).

Exemplo 2.1 Considerando uma variável aleatória X , normalmente distribuída, com parâmetros μ e σ^2 que identificam a média e a variância da distribuição, respectivamente, cuja função densidade de probabilidade é:

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}, \quad (2.4)$$

em que $x \in (-\infty, \infty)$. A entropia diferencial para esta variável aleatória é dada por:

$$\begin{aligned} h_\alpha(X) &\stackrel{(a)}{=} \frac{1}{1-\alpha} \log \left[\int_{-\infty}^{\infty} \left(\frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \right)^\alpha dx \right] \\ &\stackrel{(b)}{=} \frac{1}{1-\alpha} \log \left[\frac{\sqrt{2\pi\sigma^2}^{1-\alpha}}{\alpha^{\frac{1}{2}}} \int_{-\infty}^{\infty} \frac{\alpha^{\frac{1}{2}}}{\sqrt{2\pi\sigma^2}} e^{-\frac{\alpha(x-\mu)^2}{2\sigma^2}} dx \right] \\ &\stackrel{(c)}{=} \frac{1}{1-\alpha} \log \left[\sqrt{2\pi\sigma^2}^{1-\alpha} \alpha^{-\frac{1}{2}} \right] \\ &\stackrel{(d)}{=} \frac{1}{2} \log \left[2\pi\sigma^2 \alpha^{\frac{1}{\alpha-1}} \right], \end{aligned} \quad (2.5)$$

em que a igualdade (a) segue da definição, equação (2.1), a igualdade (b) resulta da reorganização das variáveis de modo a obter a integral de uma $f(x)$ com variância $\frac{\sigma^2}{\alpha}$, em (c) a integral tem valor unitário e em (d) a equação é reescrita. Fazendo $\alpha \rightarrow 0$ e $\alpha \rightarrow 1$ na equação (2.5) obtêm-se $h_0(X) \rightarrow \infty$ e $h(X) = \frac{1}{2} \log(2\pi e\sigma^2)$, respectivamente.

A entropia diferencial está relacionada ao grau de dispersão de uma variável aleatória. Isto é, pequenos valores de entropia implica que a variável aleatória está confinada a um pequeno volume e conforme a entropia aumenta indica que o volume no qual a variável aleatória está inserida está aumentando, aumentando assim, sua dispersão.

Uma das funções importantes da entropia para variáveis aleatórias contínuas está na propriedade de equipartição assintótica. As definições e teoremas que formalizam a AEP permitem definir um conjunto típico e caracterizar o comportamento de sequências típicas. De forma geral, é possível dividir o conjunto de todas as sequências em dois conjuntos, o conjunto típico (a entropia de amostra está próxima da entropia verdadeira) e o conjunto não-típico (demais sequências). Qualquer propriedade comprovada para as sequências típicas é verdadeira, com alta probabilidade, e determina o comportamento médio de uma amostra grande [21].

Teorema 2.1 Seja X_1, X_2, \dots, X_n uma sequência de variáveis aleatórias independentes e identicamente distribuídas (i.i.d.) com função densidade de probabilidade $f(x)$. Então,

$$-\frac{1}{n} \log f(X_1, X_2, \dots, X_n) \rightarrow E[-\log f(X)] = h(X), \quad (2.6)$$

converge em probabilidade isto é, para todo $\epsilon > 0$,

$$Pr \left\{ \left| -\frac{1}{n} \log f(X_1, X_2, \dots, X_n) - h(X) \right| > \epsilon \right\} \rightarrow 0. \quad (2.7)$$

A prova do Teorema 2.1 é uma consequência direta da lei fraca dos grandes números, na qual, para variáveis aleatórias independentes e identicamente distribuídas (i.i.d.), $\frac{1}{n} \sum_{i=1}^n X_i$ se aproxima do valor esperado $E[X]$ ao considerar grandes valores de n .

Definição 2.2 Para $\epsilon > 0$ e qualquer n , um conjunto típico $A_\epsilon^{(n)}$ com relação à função densidade de probabilidade $f(x_1, x_2, \dots, x_n)$ é definido como

$$A_\epsilon^{(n)} = \left\{ (x_1, x_2, \dots, x_n) \in \llbracket X \rrbracket : \left| -\frac{1}{n} \log f(x_1, x_2, \dots, x_n) - h(X) \right| \leq \epsilon \right\} \quad (2.8)$$

em que (x_1, x_2, \dots, x_n) são realizações independentes e identicamente distribuídas da variável aleatória X e $f(x_1, x_2, \dots, x_n) = \prod_{i=1}^n f(x_i)$.

Teorema 2.2 O conjunto típico $A_\epsilon^{(n)}$ tem as seguintes propriedades:

1. $Pr\left(A_\epsilon^{(n)}\right) > 1 - \epsilon$ para n suficientemente grande.
2. $Vol\left(A_\epsilon^{(n)}\right) \leq 2^{n(h(X)+\epsilon)}$ para todo n .
3. $Vol\left(A_\epsilon^{(n)}\right) \geq (1 - \epsilon)2^{n(h(X)-\epsilon)}$ para n suficientemente grande.

Uma interpretação para entropia diferencial segundo Shannon de uma variável aleatória é resultante das definições e propriedades da AEP: logaritmo do menor hiper-volume que contém maior densidade de amostras. De acordo com o Teorema 2.2 esse hiper-volume é aproximadamente $2^{nh(X)}$.

Exemplo 2.2 Considerando um hipercubo de dimensão n , aresta a e volume a^n . Cada aresta tem medida $(2^{nh(X)})^{\frac{1}{n}} = 2^{h(X)}$.

Desta forma, para variáveis aleatórias contínuas, pode-se estabelecer uma relação entre entropia de Rényi de ordem zero e entropia diferencial de Shannon ao que remete à caracterização do suporte no qual cada medida se refere. Para entropia de Rényi de ordem zero têm-se o logaritmo da medida de Lebesgue do suporte e para entropia diferencial de Shannon têm-se o logaritmo do suporte efetivo [21].

2.3 Estimadores

O processo de estimação de entropia para variáveis contínuas requer atenção e soluções aparentemente simples podem comprometer esse processo, por exemplo, a discretização dos dados pode resultar em estimativas enviesadas. Nesta seção, quatro estimadores para entropia diferencial serão apresentados, dentre os quais dois dependem da estimativa da função de densidade de probabilidade e dois são estimadores diretos de entropia diferencial.

2.3.1 Histograma

A abordagem mais simples e comum para estimar entropia de uma variável aleatória contínua é a técnica baseada em histograma, que representa a distribuição de frequência (absoluta ou relativa) por meio de retângulos, cujas larguras representam intervalos e cujas áreas são proporcionais às frequências.

Na técnica de histograma por frequência relativa, o suporte no qual N realizações de uma variável aleatória estão inseridas, é quantizado em intervalos de comprimento Δ e, portanto, assume-se que a função densidade de probabilidade é constante em cada intervalo. O número de amostras presentes no mesmo intervalo de quantização é contado e a estimação da função densidade de probabilidade da variável aleatória, para qualquer valor real de x , é dada por:

$$\hat{f}(x) = \frac{\text{número de observações no mesmo intervalo}}{N\Delta}. \quad (2.9)$$

Após determinação da função densidade de probabilidade, estima-se a entropia diferencial utilizando a relação entre variáveis aleatórias discretas e contínuas [21],

$$\hat{H}(X) + \log \Delta \rightarrow h(X), \quad (2.10)$$

em que $\hat{H}(X)$ é a entropia da variável quantizada,

$$\hat{H}(X) = - \sum \Delta \hat{f}(x) \log \hat{f}(x) - \log \Delta, \quad (2.11)$$

sendo uma estimação verdadeira quando a função densidade de probabilidade é integrável segundo Riemann.

2.3.2 Estimador de Densidade Kernel

O estimador de densidade Kernel (KDE, do inglês: *Kernel Density Estimator*) é uma estimativa não-paramétrica da função de densidade de probabilidade de uma variável aleatória contínua [42, 43]. Essa densidade é estimada, essencialmente, a partir de dois parâmetros: 1) função de suavização (estabelece a forma da curva usada para estimação) e; 2) valor de largura de banda (controla a suavidade da curva de densidade resultante).

Algumas das funções kernel de suavização comumente utilizadas são: uniforme, triangular, epanechnikov e gaussiana. Uma função kernel deve ser não negativa, de valor real e satisfazer dois requisitos:

1. Normalização: $\int_{-\infty}^{\infty} \Theta(u) du = 1$; e
2. Simetria: $\Theta(u) = \Theta(-u)$.

Nessa técnica, a estimativa é um problema fundamental de suavização de dados. Considerando x_1, x_2, \dots, x_N realizações de uma variável aleatória, em cada realização x_i é alocada uma função kernel e o somatório destas funções resulta na estimativa de densidade. Desta forma, para qualquer valor real x , o estimador de densidade Kernel é,

$$\hat{f}(x) = \frac{1}{Nr} \sum_{i=1}^N \Theta\left(\frac{x - x_i}{r}\right), \quad (2.12)$$

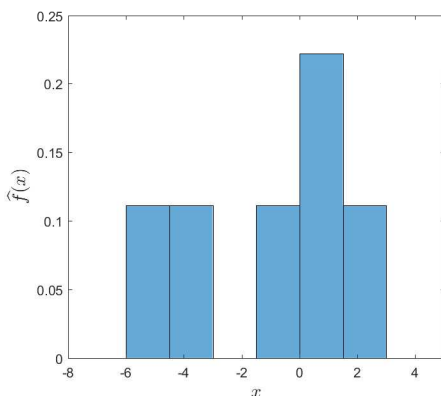
em que x_1, x_2, \dots, x_N são realizações da variável aleatória, N é o total de amostras analisadas, r é a largura de banda e, ao longo deste capítulo, $\Theta(\cdot)$ será a função kernel Gaussiana

$$\Theta(u) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}u^2}. \quad (2.13)$$

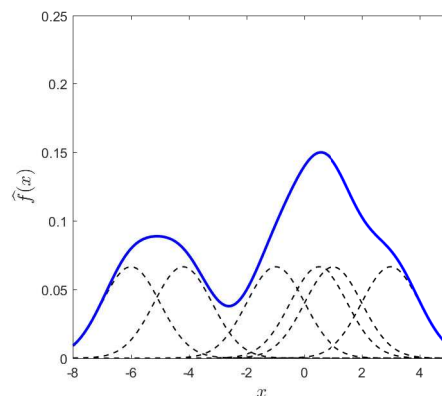
Após determinação da função densidade de probabilidade, a estimação da entropia diferencial ocorre através da relação (2.10).

Exemplo 2.3 Considerando um vetor de realizações de uma variável aleatória desconhecida, $\mathbf{x} = [-6 \quad -4,2 \quad -1 \quad 0,5 \quad 1 \quad 3]$, a função densidade probabilidade pode ser estimada a partir das técnicas:

1. *Histograma:* Divide-se o suporte de x em intervalos igualmente espaçados, nesse caso, $\Delta = 1,5$ e então, estima-se a $f(x)$ por meio da equação 2.9. Esse método é ilustrado na Figura 2.1a.
2. *Densidade kernel:* Define-se a largura de banda, nesse caso $r = 1$ e seis funções kernel Gaussiana (equação (2.13)) com variância unitária e ponderadas por $\frac{1}{N=6}$ estão centradas em cada uma das amostras do vetor \mathbf{v} (curvas pretas tracejadas da Figura 2.1b). A soma dessas seis curvas resulta na estimativa da $f(x)$ (curva azul da Figura 2.1b).



(a) Histograma.



(b) Estimativa densidade kernel.

Figura 2.1 – Estimativa da função densidade de probabilidade por meio das técnicas baseadas em: a) histograma e b) densidade kernel (curva azul).

2.3.3 Estimador Kozachenko-Leonenko

Ao contrário dos últimos dois métodos de estimação apresentados, o estimador proposto por Kozachenko e Leonenko (1987) [11] é um método direto, isto é, estima entropia sem necessariamente estimar a função densidade de probabilidade das amostras.

O estimador Kozachenko-Leonenko (KL) considera que a distribuição de probabilidade da distância entre x_i e seu k -ésimo vizinho é igual a chance de ter:

1. 1 amostra com distância $r \in [\epsilon/2, \epsilon/2 + d\epsilon/2]$ até x_i ;
2. $k - 1$ outras amostras com distâncias menores que $\epsilon/2$, e;
3. $N - k - 1$ amostras com distâncias maiores que $\epsilon/2$.

Desta forma, modela-se uma distribuição trinomial e após manipulações algébricas, que podem ser acompanhadas no Apêndice A, obtém-se o seguinte estimador para entropia diferencial:

$$\hat{h}(X) = -\psi(k) + \psi(N) + \log c_d + \frac{d}{N} \sum_{i=1}^N \log \epsilon(i), \quad (2.14)$$

em que $\psi(\cdot)$ é a função *digamma*¹, N é o número total de amostras analisadas, c_d é o volume da bola unitária d -dimensional (para norma Euclideana $c_d = \pi^{d/2}/\Gamma(1 + d/2)$) e $\epsilon(i)$ é duas vezes a distância de x_i ao seu k -ésimo vizinho.

Exemplo 2.4 Considerando o vetor de realizações de uma variável aleatória desconhecida do Exemplo 2.3, $\mathbf{x} = [-6 \quad -4,2 \quad -1 \quad 0,5 \quad 1 \quad 3]$, a representação gráfica para interpretação do estimador Kozachenko-Leonenko é expressa na Figura 2.2. Nesse caso, $N = 6$, $d = 1$ e $c_d = 2$. Deve-se fixar o k -ésimo vizinho analisado assim,

- se $k = 2$, então para $x_i = -1$ têm-se $\epsilon(i) = 4$ e para $x_i = 0,5$ têm-se $\epsilon(i) = 3$;
- se $k = 3$, então para $x_i = -1$ têm-se $\epsilon(i) = 6,4$ e para $x_i = 0,5$ têm-se $\epsilon(i) = 5$.

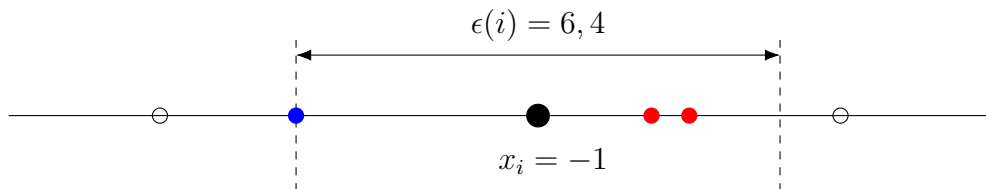


Figura 2.2 – Representação gráfica da interpretação do estimador Kozachenko-Leonenko considerando $k = 3$ então têm-se pelo menos uma amostra com distância $r \in [\epsilon/2, \epsilon/2 + d\epsilon/2]$ até x_i (em azul), $k - 1$ amostras com distâncias menores que r (em vermelho) e, $N - k - 1$ amostras com distâncias maiores que r (branco).

¹ A função *digamma* é definida como a derivada do logaritmo da função *gamma*: $\psi(x) = \frac{d}{dx} \log(\Gamma(x))$.

2.3.4 SVM de uma classe

O algoritmo SVM de uma classe foi proposto por Schölkopf et al. (2001) [16] para estimar o suporte de uma distribuição de grande dimensão. Além disso, foi originalmente usado para a estimar entropia de Rényi de ordem zero por Duan et al. (2015) [17].

Por ser um método de aprendizado supervisionado, o algoritmo SVM de uma classe recebe como entrada um conjunto de dados, chamado de dados de treinamento. Esses dados, pertencentes a uma única classe, são projetados em um espaço característico a partir de uma função kernel por intermédio da operação produto interno.

Após essas etapas, um modelo para o conjunto de amostras é definido, isto é, é determinado a linha de separação, mais comumente chamada de hiperplano, que isola os dados da classe avaliada dos chamados *outliers* (esse método é ilustrado na Figura 2.3.). Para tanto, Schölkopf et al. (2001) [16] propõe o seguinte problema quadrático,

$$\begin{aligned} \min_{\alpha} \quad & \frac{1}{2} \sum_{ij} \alpha_i \alpha_j k(\mathbf{v}_i, \mathbf{v}_j) \\ \text{sujeito a} \quad & 0 \leq \alpha_i \leq \frac{1}{\nu N}, \quad \sum_i \alpha_i = 1, \end{aligned} \quad (2.15)$$

em que α é um vetor com os coeficientes dos vetores de suporte, $\nu \in (0, 1)$ indica um limite superior na fração de pontos de treinamento fora da região (*outliers*) e $k(\mathbf{v}_i, \mathbf{v}_j) = e^{-\gamma \|\mathbf{v}_i - \mathbf{v}_j\|^2}$ mapeia as amostras por meio da operação produto interno.

Desta forma, amostras, até então desconhecidas, podem ser classificadas a partir do modelo delimitado por vetores de suporte. Para isso, uma região de interesse na qual, o suporte da variável aleatória está inserida, é delimitada e quantizada em n_{bins} intervalos de volume Δ . Então, a estimativa de entropia diferencial para uma variável aleatória é,

$$\hat{h}(X) = \log(n \times \Delta), \quad (2.16)$$

em que n é o número de pontos classificados como pertencentes à única classe treinada.

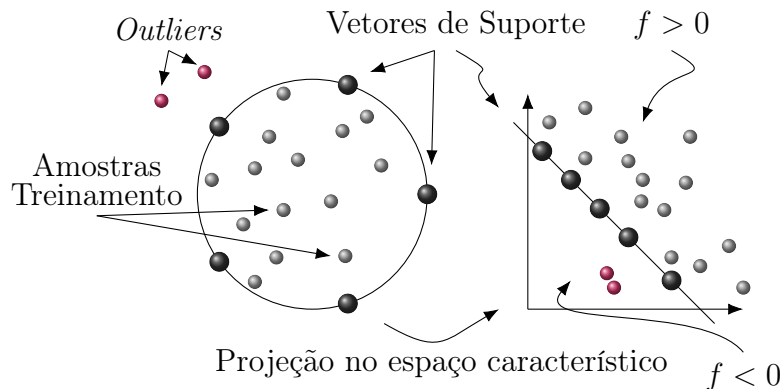


Figura 2.3 – Classificação baseada no algoritmo SVM de uma classe: amostras de treinamento são separados no espaço característico por meio de vetores de suporte.

A formulação matemática e o pseudocódigo do estimador de entropia a partir do SVM de uma classe estão no Apêndice A. O método atende à estimação da entropia de Rényi de ordem zero ao considerar que a fração de *outliers* é muito pequena [17] contudo, para atender à entropia de Shannon, é necessário caracterizar o suporte efetivo das amostras.

Caracterização do suporte efetivo

As aplicações que recorrem ao algoritmo SVM de uma classe para estimar a entropia foram usados para atender a entropia de Rényi de ordem zero [17]. Contudo, já foi discutido que a partir da AEP, ambas as estimativas, entropia de Rényi de ordem zero e entropia de Shannon, diferem apenas na caracterização do conjunto de suporte.

Visto que as estimativas de entropia de Shannon são feitas a partir do logaritmo do tamanho do conjunto de suporte efetivo, é necessário caracterizar esse suporte. O suporte efetivo é definido como aquele que contém a maior parte da amostra de dados. Nesta seção, propomos uma maneira de caracterizar o suporte efetivo, a fim de utilizar o algoritmo SVM de uma classe na estimação da entropia de Shannon, por meio do ajuste da fração de *outliers* no algoritmo.

Considerando a função densidade de probabilidade de alguma distribuição bem definida, propomos que a fração de *outliers* equivale a fração de amostras que ocorre com densidade menor que 15% do valor de densidade máxima (respectivo à moda da distribuição). Se essa área é nula ou infinita, $\nu = 0,01$.

Exemplo 2.5 Considerando uma variável aleatória X , gamma-distribuída, com parâmetros k e θ que identificam a forma e a escala da distribuição, respectivamente, sua função densidade de probabilidade é:

$$f(x) = \frac{1}{\Gamma(k)\theta^k} x^{k-1} e^{-x/\theta}, \quad (2.17)$$

em que $\Gamma(\cdot)$ é a função gamma e $x \in (0, \infty)$. Supondo $k = 3$ e $\theta = 2$, o valor máximo de $f(x)$ é 0,1353. A área com probabilidade menor que $0,15 \times 0,1353 = 0,0203$ representa aproximadamente 6% de toda a área de densidade das amostras, então, $\nu = 0,06$.

2.4 Resultados

Com o propósito de avaliar a técnica de caracterização de suporte efetivo proposta para o estimador SVM de uma classe, e compará-la com os demais métodos de estimação apresentados, foram consideradas 10^5 amostras independentes e identicamente distribuídas de variáveis aleatórias contínuas com densidade de probabilidade bem definida, dentre as quais: exponencial, gamma, normal e uniforme. Além disso, os valores teóricos para a entropia diferencial de Shannon dessas distribuições, Tabela 2.1, serão nossos referenciais durante as simulações.

Para o algoritmo de estimação SVM de uma classe, a caracterização do suporte, para cada uma das quatro distribuições supracitadas, está ilustrado na Figura 2.4. Na Tabela 2.2 estão dispostos os valores do parâmetro ν (indicador da fração de *outliers*) utilizando a metodologia já apresentada e considerando dois conjuntos de parâmetros em cada caso.

Tabela 2.1 – Entropia diferencial de Shannon para as distribuições exponencial, gamma, normal e uniforme.

Distribuição	Entropia
$\text{Exp}(\lambda)$	$1 - \log(\lambda)$
$\Gamma(k, \theta)$	$\log(\theta) + \log(\Gamma(k)) + (1 - k)\psi(k) + k$
$\mathcal{N}(\mu, \sigma^2)$	$\frac{1}{2} \log(2\pi e\sigma^2)$
$\mathcal{U}(a, b)$	$\log(b - a)$

Tabela 2.2 – Parâmetros ν para as distribuições exponencial, gamma, normal e uniforme.

Distribuição	Exp(1)	$\Gamma(3, 2)$	$\mathcal{N}(2, 9)$	$\mathcal{U}(-1, 1)$
ν	0,15	0,0606	0,0514	0,01
Distribuição	Exp(2)	$\Gamma(2, 0.5)$	$\mathcal{N}(0, 1)$	$\mathcal{U}(0, 0.5)$
ν	0,15	0,0694	0,0478	0,01

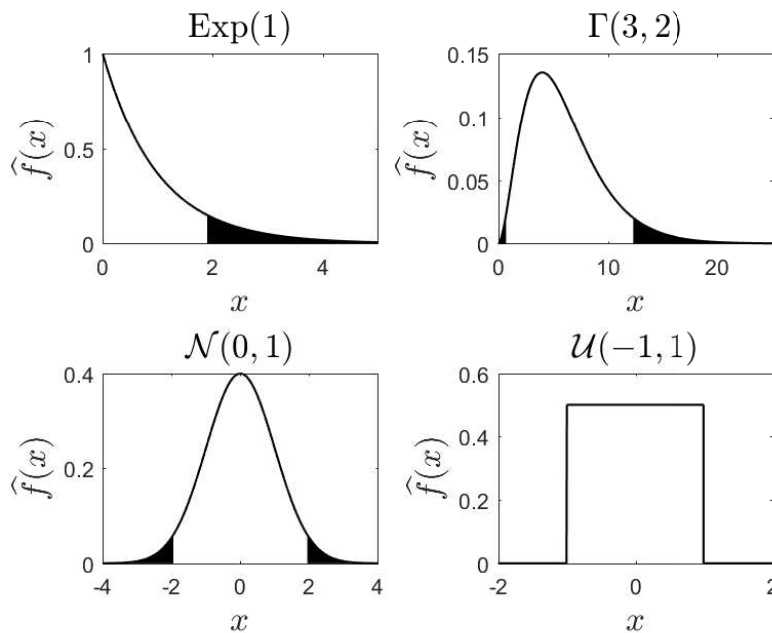


Figura 2.4 – Função densidade de probabilidade para uma variável aleatória X com distribuição exponencial, gamma, normal e uniforme. A área preenchida indica a região com densidade menor que 15% da densidade máxima.

Em cada método de estimação foram utilizados os seguintes parâmetros de simulação:

1. Histograma: 100 intervalos;
2. Estimador de densidade Kernel: $r = 0,01$;
3. Kozachenko-Leonenko: $k = 1$;
4. SVM de uma classe: Tabela 2.2.

São apresentados nas Figuras 2.5, 2.6, 2.7 e 2.8 os resultados para estimativa de entropia em função do crescimento do número de amostras, considerando dois conjuntos de parâmetros diferentes para as distribuições exponencial, gamma, normal e uniforme, respectivamente. Em todas as figuras, as linhas preta indicam o valor de entropia analítica e as linhas magenta, verde, vermelha e azul indicam a utilização das técnicas de estimação por histograma, densidade kernel, distâncias de vizinho e SVM de uma classe, respectivamente.

Nas simulações considerando as amostras com distribuição exponencial, gamma e normal, o estimador de densidade kernel foi o que precisou de um maior número de amostras para garantir uma boa convergência. Contudo, a largura de banda é um importante parâmetro para essa estimação e, nas simulações realizadas esse parâmetro foi considerado o mesmo em todas as situações. Métodos de otimização desse parâmetro podem melhorar esses resultados.

Os algoritmos que não dependem da estimação da função densidade de probabilidade (Kozachenko-Leonenko e SVM de uma classe) apresentaram convergência mais rápida em função do número de amostras quando comparados aos algoritmos baseados em histograma e kernel. O termo de correção introduzido por meio da função *digamma* no estimador KL é importante para redução do viés do estimador, principalmente quando considera-se um grande número de amostras [14].

Um diferencial do método SVM de uma classe é que agregando a proposta de caracterização de suporte efetivo, é possível optar por estimar tanto a entropia de Rényi de ordem zero quanto a entropia de Shannon. O algoritmo proposto forneceu resultados que convergem para o valor analítico de entropia, além disso, em algumas situações, o algoritmo foi o primeiro a convergir.

No entanto, se a função densidade de probabilidade tende a assumir valor infinito em alguma realização da variável aleatória, a caracterização de suporte, conforme proposto, não será eficiente para estimação da entropia de Shannon. Nas simulações realizadas, isso acontece quando $\mathcal{X} \sim \Gamma(0,5;1)$, Figura 2.6, e, nesse caso, a estimativa converge para a entropia de Rényi de ordem zero. Além disso, a técnica de histograma e o KDE também apresentaram erros mais significativos para essa estimação.

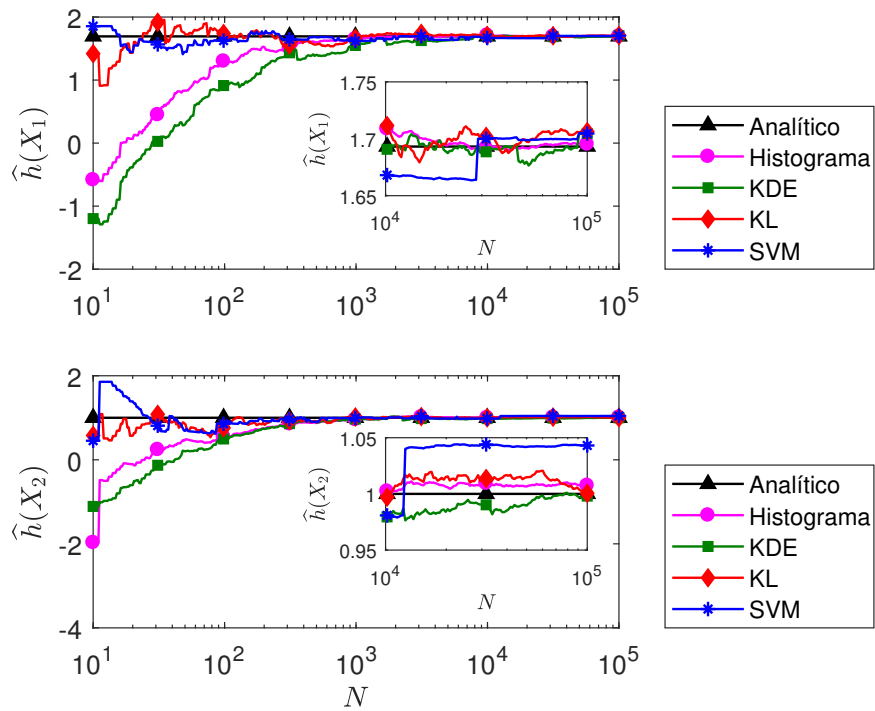


Figura 2.5 – Estimação de entropia considerando 10^5 realizações de variáveis aleatórias exponencialmente distribuídas: $X_1 \sim \text{Exp}(2)$ com $h(X_1) = 1,69 \text{ nats}$, curvas mais acima e; $X_2 \sim \text{Exp}(1)$ com $h(X_2) = 1 \text{ nat}$, curvas mais abaixo.

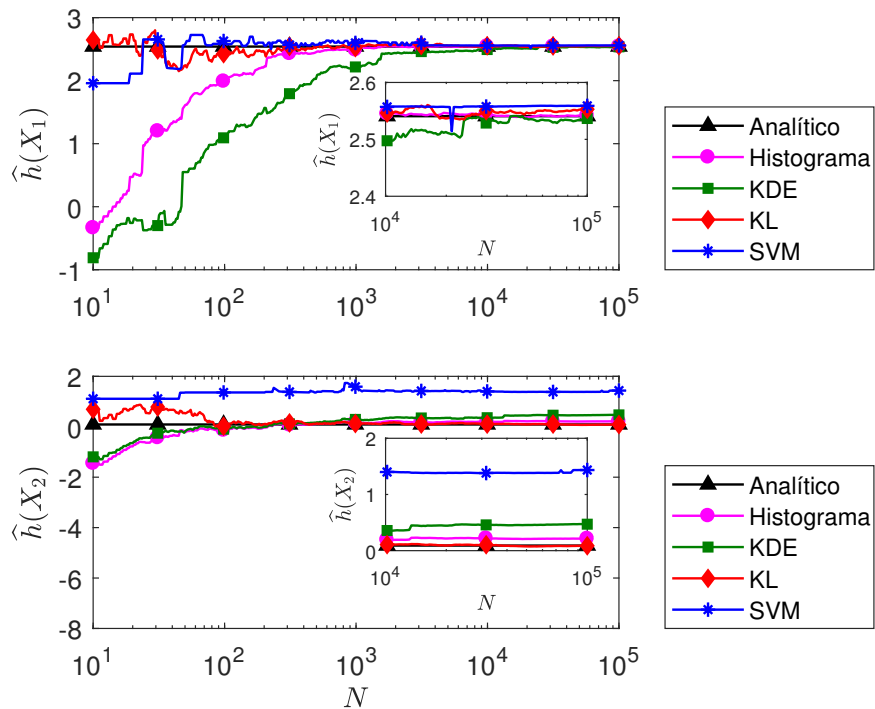


Figura 2.6 – Estimação de entropia considerando 10^5 realizações de variáveis aleatórias gamma distribuídas: $X_1 \sim \Gamma(3;2)$ com $h(X_1) = 2,09 \text{ nats}$, curvas mais acima e; $X_2 \sim \Gamma(0,5;1)$ com $h(X_2) = 0,09 \text{ nats}$, curvas mais abaixo.

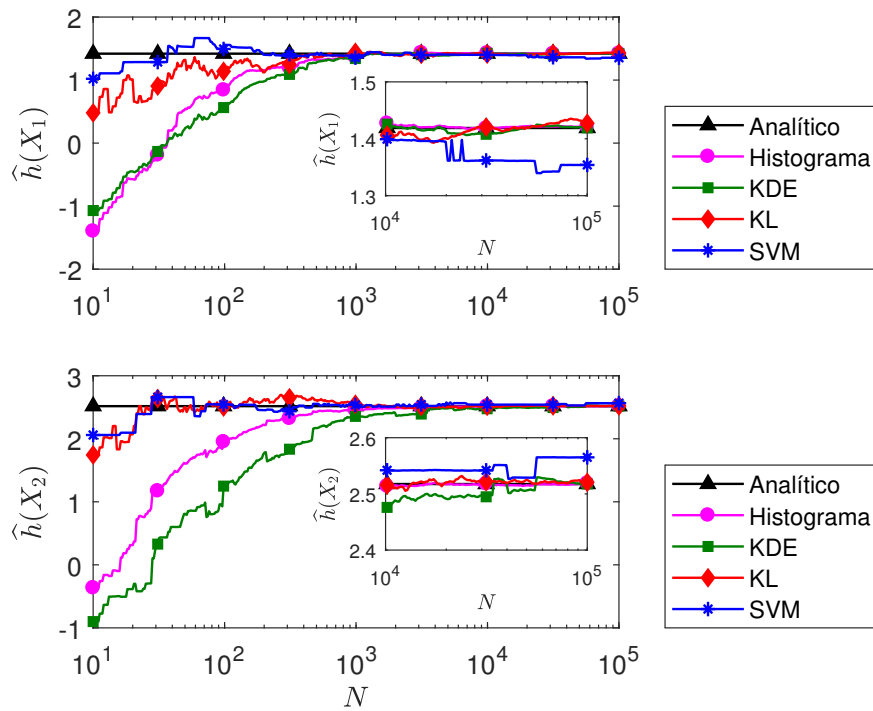


Figura 2.7 – Estimação de entropia considerando 10^5 realizações de variáveis aleatórias normalmente distribuídas: $X_1 \sim \mathcal{N}(2; 3)$ com $h(X_1) = 2,51 \text{ nats}$, curvas mais acima e; $X_2 \sim \mathcal{N}(0; 1)$ com $h(X_2) = 1,41 \text{ nats}$, curvas mais abaixo.

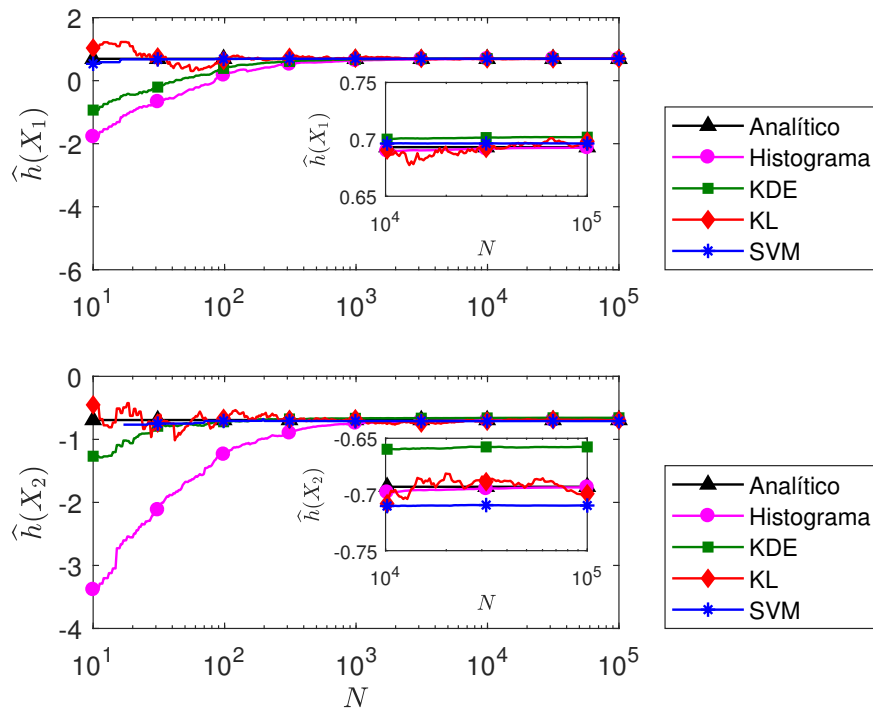


Figura 2.8 – Estimação de entropia considerando 10^5 realizações de variáveis aleatórias uniformemente distribuídas: $X_1 \sim \mathcal{U}(-1; 1)$ com $h(X_1) = 0,69 \text{ nats}$, curvas mais acima e; $X_2 \sim \mathcal{U}(0; 0,5)$ com $h(X_2) = -0,69 \text{ nats}$, curvas mais abaixo.

Quanto ao tempo de estimação, por ser um algoritmo de aprendizado de máquina, o estimador SVM de uma classe é o que mais demanda tempo de processamento. Esse tempo está diretamente relacionado ao parâmetro ν . Quando ν assume o valor mínimo, ou seja, $\nu = 0,01$, cada simulação, com 2×10^4 realizações, demanda aproximadamente $0,15$ s, enquanto que os métodos histograma, KDE e KL demandam $0,05$ s, $0,006$ s e $0,015$ s, respectivamente. Nas simulações realizadas, o valor máximo de ν foi $0,15$ e cada tentativa demandou aproximadamente $2,7$ s. As simulações foram realizadas em um computador com um processador de $2,70$ GHz.

2.5 Considerações

Embora o algoritmo do SVM de uma classe tenha sido inicialmente aplicado para estimar entropia de Rényi de ordem zero, a AEP para variáveis contínuas inclui uma interpretação para entropia de Shannon que relaciona as duas definições de entropia supracitadas, diferindo-as apenas ao que concerne a caracterização do suporte das amostras.

Foram realizadas oito simulações a fim de avaliar e comparar os quatro estimadores de entropia de Shannon já discutidos. Em cada simulação, foram analisadas 10^5 realizações de uma variável aleatória bem definida. Os resultados indicaram que não há diferença significativa entre os métodos de estimação. Contudo, em algumas situações, o algoritmo proposto para a caracterização do suporte efetivo aplicado à técnica do SVM de uma classe converge mais rápido em função do número de dados amostrais analisados.

Os algoritmos de estimação direta de entropia: estimador KL e estimador SVM de uma classe, estimam entropia sem priorizar a estimativa da função densidade de probabilidade da variável aleatória, desta forma, apresentaram melhor convergência no processo de estimação em função do número de realizações da variável. Isso é fundamentalmente importante para reduzir os erros de enviesamento do estimador. Contudo, o algoritmo SVM de uma classe ainda apresenta restrições de aplicação.

Com base na relação entre a entropia de Rényi de ordem zero e a entropia de Shannon, propusemos a caracterização do suporte efetivo que torna aplicável o método do SVM de uma classe nos dois casos. O método proposto fornece resultados assintoticamente bons para a estimativa da entropia de Shannon em sete dos oito casos analisados. O caso analisado que não converge, é um caso especial e, acontece quando a função densidade de probabilidade tende a assumir valor infinito. Nesse caso, o algoritmo estima todo o suporte da distribuição e a estimativa converge para a entropia de Rényi de ordem zero.

Apesar de o SVM de uma classe ser o algoritmo que requer mais tempo de processamento, ele pode ser útil em diversas aplicações, dentre as quais, na análise de causalidade considerando sinais não estacionários e aplicado à estimativa de outras medidas como entropia de transferência de ordem zero [17].

Capítulo 3

Abordagens para Detecção de Causalidade em Sistemas Contínuos

Neste capítulo será discutido a utilização da entropia de transferência como abordagem para detecção de causalidade e conectividade em sistemas contínuos. A entropia de transferência é uma medida assimétrica que identifica e quantifica relações direcionais lineares ou não lineares entre duas variáveis. Para estimação da medida, foram utilizados estimadores baseados em distâncias entre vizinhos. Os resultados obtidos a partir de simulações demonstram a aplicabilidade da medida na detecção de conectividade entre equipamentos industriais e no auxílio de diagnóstico biomédico.

3.1 Introdução

No contexto da Teoria da Informação, os conceitos de transferência de informação como: informação direcional [3, 4] e entropia de transferência [5], são usados como indicadores de causalidade [25, 23, 7, 24, 22]. Essas medidas quantificam a influência causal e a direcionalidade dessa influência entre dois processos.

Um processo \mathbf{X} é dito causar o processo \mathbf{Y} no sentido de Wiener (1956) [29] e Granger (1969) [30], se os valores passados de \mathbf{X} ajudam a prever o valor presente de \mathbf{Y} . O objetivo de usar medidas de informação é identificar e distinguir se a relação entre os processos \mathbf{X} e \mathbf{Y} é mútua (ambos influenciam aos outros) ou unidirecional (\mathbf{X} influencia \mathbf{Y} , mas \mathbf{Y} não influencia \mathbf{X}).

A entropia de transferência será explorada neste capítulo em aplicações com processos contínuos que inclui a detecção da conectividade em redes de processos (reproduzindo a verdadeira conectividade em plantas industriais) e a identificação de padrões (em sinais biomédicos). A captura da conectividade dos processos dar-se-á a partir da extensão da entropia de transferência proposta por Duan et al. (2013) [27]: a entropia de transferência direta.

Identificar a conectividade de um sistema é importante para distinguir se a influência causal ocorre ao longo de uma via direta sem quaisquer variáveis intermediárias, ou por meio de vias indiretas através de variáveis intermediárias. Os processos também podem estar relacionados a partir de uma fonte comum (causalidade espúria). Nesse caso, se o processo \mathbf{X} causa o processo \mathbf{Y} e o processo \mathbf{X} causa o processo \mathbf{Z} , então, os processos \mathbf{Y} e \mathbf{Z} podem está relacionados devido à influência do processo \mathbf{X} .

Durante o estudo de causalidade de processos estocásticos, as medidas que quantificam essas relações são resultantes de um processo de estimação isto porque análises analíticas não são triviais e portanto, recorre-se ao uso de estimadores de medidas da Teoria da Informação.

Neste capítulo as medidas de informação utilizadas podem ser interpretadas como somas de entropias diferenciais de Shannon sendo assim, diante dos resultados obtidos no capítulo 2, optou-se por utilizar estimadores diretos para cada medida de informação. Esses estimadores baseiam-se nas distâncias entre vizinhos (estimador KL) [13, 15] e, dessa forma, pode-se evitar o acúmulo de erros resultantes da soma de estimativas de entropias.

3.2 Entropia de Transferência

A entropia de transferência é uma medida não linear introduzida por Schreiber (2000) [5] com o objetivo de quantificar a troca de informação entre dois processos estocásticos, dependendo do acoplamento direcional existente, quando eles não são independentes e podem ser aproximados por um processo estacionário de Markov.

Inicialmente, a entropia de transferência foi definida para processos discretos [5] e estendida para processos contínuos [44]. Antes de introduzir sua interpretação matemática, é importante notar que os conceitos que envolvem entropia de transferência estão relacionados à teoria desenvolvida por Shannon, portanto, entropia refere-se à entropia de Shannon (seção 2.2).

Definição 3.1 *Dados dois processos estocásticos contínuos \mathbf{X} e \mathbf{Y} , a entropia de transferência de \mathbf{X} para \mathbf{Y} é definida como*

$$T_{\mathbf{X} \rightarrow \mathbf{Y}} = \int f(Y_{n+1}, Y_{n-l+1}^n, X_{n-k+1}^n) \log \frac{f(Y_{n+1} | Y_{n-l+1}^n, X_{n-k+1}^n)}{f(Y_{n+1} | Y_{n-l+1}^n)} d\mathbf{V}, \quad (3.1)$$

em que \mathbf{V} é o vetor aleatório $[Y_{n+1}, Y_{n-l+1}^n, X_{n-k+1}^n]$, $X_{n-k+1}^n = \{X_{n-k+1}, \dots, X_n\}$ e $Y_{n-l+1}^n = \{Y_{n-l+1}, \dots, Y_n\}$, isto é, k e l são os comprimentos das sequências de variáveis aleatórias de X e Y respectivamente.

A entropia de transferência mensura quanto de informação do processo \mathbf{X} está presente no processo \mathbf{Y} . Sua formulação matemática pode ser reescrita através da soma de entropias diferenciais ou por meio da informação mútua condicionada,

$$\begin{aligned} T_{\mathbf{X} \rightarrow \mathbf{Y}} &= h(X_{n-k+1}^n, Y_{n-l+1}^n) - h(Y_{n+1}, X_{n-k+1}^n, Y_{n-l+1}^n) \\ &\quad + h(Y_{n+1}, Y_{n-l+1}^n) - h(Y_{n-l+1}^n) \end{aligned} \quad (3.2)$$

$$= I(Y_{n+1}; X_{n-k+1}^n | Y_{n-l+1}^n). \quad (3.3)$$

Exemplo 3.1 Considerando um modelo autorregressivo bidimensional orientado por ruídos Gaussianos,

$$\begin{cases} X_{n+1} &= \eta_n^X \\ Y_{n+1} &= \beta X_n + \eta_n^Y \end{cases},$$

em que $\eta_n^X, \eta_n^Y \sim \mathcal{N}(0; 1)$ e $0 < \beta < 1$ (sistema é estável quando os pólos estão definidos dentro do círculo de raio unitário). Fazendo $k = l = 1$ o vetor aleatório para as duas possíveis análises da entropia de transferência é $[X_n, Y_n, X_{n+1}, Y_{n+1}]$ cuja matriz de covariância é dada por,

$$\text{cov}([X_n, Y_n, X_{n+1}, Y_{n+1}]) = \begin{bmatrix} 1 & 0 & 0 & \beta \\ 0 & \beta^2 + 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ \beta & 0 & 0 & \beta^2 + 1 \end{bmatrix}.$$

É importante lembrar que a entropia diferencial de um conjunto de variáveis aleatórias normalmente distribuídas com média μ e matriz de covariância K é,

$$h(X_1, X_2, \dots, X_n) = \frac{1}{2} \log [(2\pi e)^n |K|],$$

em que $|K|$ corresponde ao determinante de K . Sendo assim, a entropia de transferência é determinada analiticamente utilizando a equação 3.2,

$$\begin{aligned} T_{\mathbf{X} \rightarrow \mathbf{Y}} &= \frac{1}{2} \log \frac{|\text{cov}([X_n, Y_n])| |\text{cov}([Y_{n+1}, Y_n])|}{|\text{cov}([Y_{n+1}, X_n, Y_n])| |\text{cov}([Y_n])|} \\ &= \frac{1}{2} \log \frac{(\beta^2 + 1)(\beta^2 + 1)^2}{(\beta^2 + 1)(\beta^2 + 1)} \\ &= \frac{1}{2} \log \beta^2 + 1, \end{aligned}$$

$$\begin{aligned} T_{\mathbf{Y} \rightarrow \mathbf{X}} &= \frac{1}{2} \log \frac{|\text{cov}([X_n, Y_n])| |\text{cov}([X_{n+1}, X_n])|}{|\text{cov}([X_{n+1}, X_n, Y_n])| |\text{cov}([X_n])|} \\ &= \frac{1}{2} \log \frac{(\beta^2 + 1)1}{(\beta^2 + 1)1} \\ &= \frac{1}{2} \log 1 = 0. \end{aligned}$$

Contudo, a entropia de transferência não é suficiente para detectar a conectividade de uma rede de processos. Por exemplo, supondo três processos contínuos em que $T_{\mathbf{X} \rightarrow \mathbf{Y}}$, $T_{\mathbf{X} \rightarrow \mathbf{Z}}$ e $T_{\mathbf{Z} \rightarrow \mathbf{Y}}$ são todos maiores que zero, diz-se que \mathbf{X} causa \mathbf{Y} , \mathbf{X} causa \mathbf{Z} e \mathbf{Z} causa \mathbf{Y} . No entanto, este resultado não indica se o fluxo de informações de \mathbf{X} para \mathbf{Y} é ao longo de um caminho direto ou indireto, em outras palavras, se \mathbf{Z} é um processo intermediário. Essa verificação é dada a partir da entropia de transferência direta [27].

Definição 3.2 *Dados os processos estocásticos contínuos \mathbf{X} , \mathbf{Y} , $\mathbf{Z}_1, \dots, \mathbf{Z}_q$, a entropia de transferência direta de \mathbf{X} para \mathbf{Y} , considerando $\mathbf{Z}_1, \dots, \mathbf{Z}_q$ potenciais processos intermediários, é definida como*

$$D_{\mathbf{X} \rightarrow \mathbf{Y}} = \int f \left(Y_{n+1}, Y_{n-l+1}^n, Z_{1,n-j_1+1}^n, \dots, Z_{q,n-j_q+1}^n, X_{n-k+1}^n \right) \log \frac{f \left(Y_{n+1} | Y_{n-l+1}^n, Z_{1,n-j_1+1}^n, \dots, Z_{q,n-j_q+1}^n, X_{n-k+1}^n \right)}{f \left(Y_{n+1} | Y_{n-l+1}^n, Z_{1,n-j_1+1}^n, \dots, Z_{q,n-j_q+1}^n \right)} d\mathbf{W}, \quad (3.4)$$

em que \mathbf{W} é o vetor aleatório $\left[Y_{n+1}, Y_{n-l+1}^n, Z_{1,n-j_1+1}^n, \dots, Z_{q,n-j_q+1}^n, X_{n-k+1}^n \right]$, q é o número de processos intermediários e k, l e j_1, \dots, j_q são os comprimentos das sequências de variáveis aleatórias de \mathbf{X} , \mathbf{Y} e $\mathbf{Z}_1, \dots, \mathbf{Z}_q$ respectivamente.

De forma análoga à formulação matemática de entropia de transferência, considerando apenas uma variável aleatória intermediária, têm-se:

$$D_{\mathbf{X} \rightarrow \mathbf{Y}} = h \left(X_{n-k+1}^n, Y_{n-l+1}^n, Z_{n-j+1}^n \right) - h \left(Y_{n+1}, X_{n-k+1}^n, Y_{n-l+1}^n, Z_{n-j+1}^n \right) + h \left(Y_{n+1}, Y_{n-l+1}^n, Z_{n-j+1}^n \right) - h \left(Y_{n-l+1}^n, Z_{n-j+1}^n \right) \quad (3.5)$$

$$= I(Y_{n+1}; X_{n-k+1}^n | Y_{n-l+1}^n, Z_{n-j+1}^n). \quad (3.6)$$

Então, se $D_{\mathbf{X} \rightarrow \mathbf{Y}}$ é maior que zero, existe um caminho direto entre os processos, caso contrário, a influência causal é através da variável intermediária. Além disso, se houver uma causalidade direta entre \mathbf{X} e \mathbf{Y} , é possível que \mathbf{Z} não cause \mathbf{Y} e esta causalidade espúria está sendo gerada por uma fonte comum: \mathbf{X} . Portanto, $D_{\mathbf{Z} \rightarrow \mathbf{Y}}$ precisa ser calculado. Após esses procedimentos, a conectividade do sistema pode ser capturada e reproduzida em um mapa causal direcional, como ilustrado na Figura 3.1.

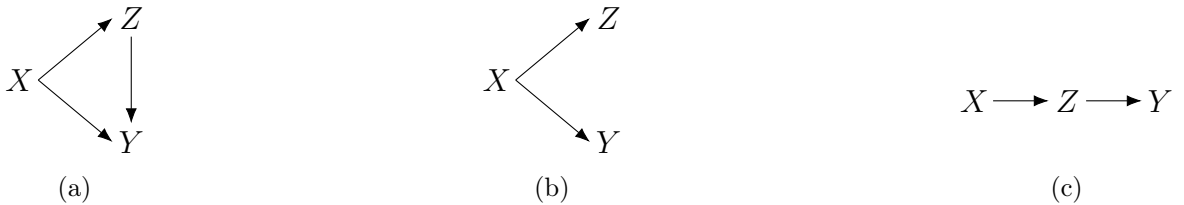


Figura 3.1 – Grafos de mapa causal direcional para os processos \mathbf{X} , \mathbf{Y} e \mathbf{Z} quando apenas $T_{\mathbf{X} \rightarrow \mathbf{Y}}$, $T_{\mathbf{X} \rightarrow \mathbf{Z}}$ e $T_{\mathbf{Z} \rightarrow \mathbf{Y}}$ são maiores que zero. Além disso, as possíveis situações de conectividade dos processos acontecem quando: (a) $D_{\mathbf{X} \rightarrow \mathbf{Y}} > 0$; (b) $D_{\mathbf{X} \rightarrow \mathbf{Y}} > 0$ e $D_{\mathbf{Z} \rightarrow \mathbf{Y}} = 0$; (c) $D_{\mathbf{X} \rightarrow \mathbf{Y}} = 0$.

3.3 Estimadores

Conforme exposto no capítulo 2, o processo de estimação de medidas da Teoria da Informação para variáveis contínuas requer atenção, uma vez que soluções aparentemente simples, como discretização dos dados, podem comprometer as estimativas. Além disso, entre os quatro estimadores discutidos, o estimador para entropia diferencial proposto por Kozachenko e Leonenko (1987) [11] (baseado nas distâncias dos vizinhos) foi o que demonstrou melhor desempenho, isto porque é um estimador direto e sem restrições de aplicabilidade. Nessa abordagem,

$$\widehat{h}(X) = -\psi(k) + \psi(N) + \log c_d + \frac{d}{N} \sum_{i=1}^N \log \epsilon(i), \quad (3.7)$$

em que $\psi(\cdot)$ é a função *digamma*, N é o número total de amostras analisadas, c_d é o volume da bola unitária d -dimensional (para norma Euclidiana $c_d = \pi^{d/2}/\Gamma(1 + d/2)$) e $\epsilon(i)$ é duas vezes a distância de x_i ao seu k -ésimo vizinho.

Essa formulação foi expandida em dois algoritmos para estimação de informação mútua por Kraskov, Stögbauer e Grassberger (2004) [13]. Ambos os algoritmos, para estimar informação mútua entre duas variáveis aleatórias X e Y , usam a norma máxima para o espaço conjunto $Z = (X, Y)$,

$$\|z - z'\| = \max\{\|x - x'\|, \|y - y'\|\},$$

contudo, qualquer norma pode ser usada para $\|x - x'\|$ e $\|y - y'\|$ visto que os espaços são diferentes. Nesse contexto, utilizar-se-á a norma Euclidiana dada a sua ampla aplicabilidade na literatura. Seja $\epsilon(i)/2$ a distância entre z_i e seu k -ésimo vizinho, e $\epsilon_x(i)/2$ e $\epsilon_y(i)/2$ as distâncias entre os mesmos pontos projetados nos subespaços X e Y , então, $\epsilon(i) = \max\{\epsilon_x(i), \epsilon_y(i)\}$.

O estimador para informação mútua, $I(X, Y) = h(X) + h(Y) - h(X, Y)$, é obtido a partir da soma de estimativas conforme (3.7) para entropia diferencial. No espaço conjunto, a entropia diferencial é obtida substituindo d por $d_x + d_y$ e c_d por $c_{d_x} c_{d_y}$,

$$\widehat{h}(X, Y) = -\psi(k) + \psi(N) + \log c_{d_x} c_{d_y} + \frac{d_x + d_y}{N} \sum_{i=1}^N \log \epsilon(i). \quad (3.8)$$

Contudo, para qualquer k fixo, a distância entre uma amostra e o seu k -ésimo vizinho no espaço conjunto é maior do que no espaço marginal. Desta forma, mantendo o k fixo, o viés em $\widehat{h}(X)$, $\widehat{h}(Y)$ e $\widehat{h}(X, Y)$ não é cancelado.

Portanto, a fim de minimizar esse viés, se houver $n_x(i)$ pontos dentro da região delimitada pelas linhas verticais $x = x_i \pm \epsilon(i)/2$ então, $\epsilon(i)/2$ é a distância entre x_i e o seu $(n_x(i) + 1)$ -ésimo vizinho, assim,

$$\widehat{h}(X) = -\frac{1}{N} \sum_{i=1}^N \psi(n_x(i) + 1) + \psi(N) + \log c_{d_x} + \frac{d_x}{N} \sum_{i=1}^N \log \epsilon(i). \quad (3.9)$$

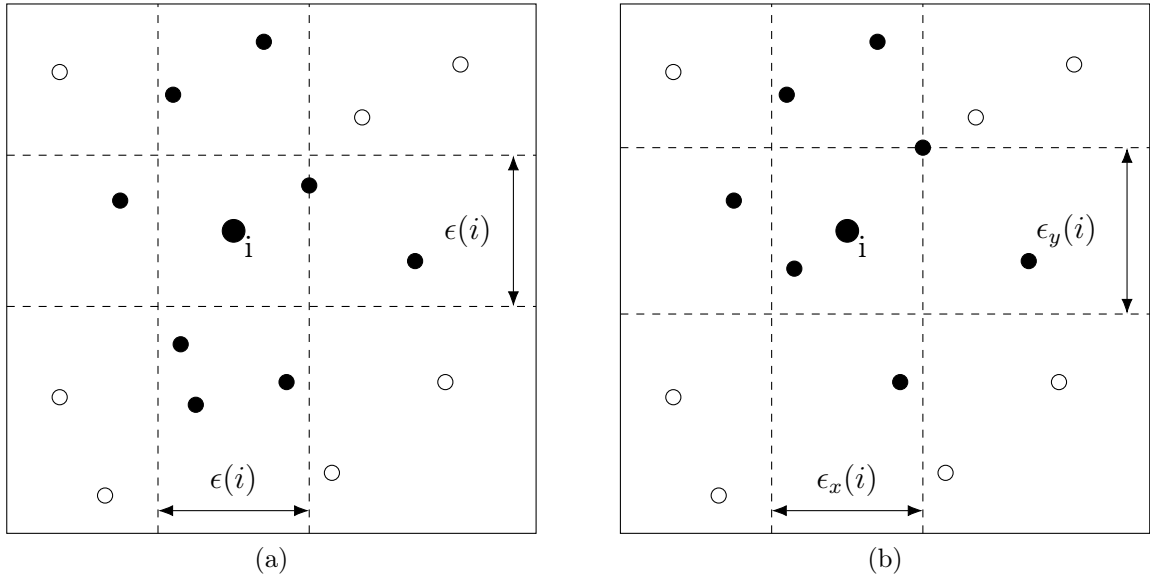


Figura 3.2 – Determinação dos parâmetros para os estimadores propostos por Kraskov, Stögbauer e Grassberger. (a) Considerando o primeiro algoritmo com $k = 1$, $n_x = 5$ e $n_y = 3$; (b) Considerando o segundo algoritmo com $k = 2$, $n_x = 5$ e $n_y = 4$.

A diferença entre os dois algoritmos está nas considerações de delimitação da região de busca de vizinhos. No primeiro algoritmo, $n_x(i)$ e $n_y(i)$ correspondem ao número de pontos com $\|x_i - x_j\|$ e $\|y_i - y_j\|$ menores que $\epsilon(i)/2$, Figura 3.2a, e a estimativa é,

$$I^{(1)}(X; Y) = \psi(k) - \frac{1}{N} \sum_{i=1}^N [\psi(n_x(i) + 1) + \psi(n_y(i) + 1)] + \psi(N), \quad (3.10)$$

em que N é o total de amostras analisadas. No segundo algoritmo, $n_x(i)$ e $n_y(i)$ correspondem ao número de pontos com $\|x_i - x_j\| \leq \epsilon_x(i)/2$ e $\|y_i - y_j\| \leq \epsilon_y(i)/2$. Além disso, ϵ_x e ϵ_y são determinado com base no mesmo ponto, Figuras 3.2b, e a estimativa é dada por,

$$I^{(2)}(X; Y) = \psi(k) - \frac{1}{k} - \frac{1}{N} \sum_{i=1}^N [\psi(n_x(i)) + \psi(n_y(i))] + \psi(N). \quad (3.11)$$

Em geral, as duas formulações apresentam resultados semelhantes, entretanto, em dimensões muito altas, $\epsilon(i)$ tende a ser muito diferente das distâncias marginais, $\epsilon_x(i)$ e $\epsilon_y(i)$, então, nesse caso, é preferível usar o segundo algoritmo [13]. Porém, nas aplicações que serão exploradas ao longo deste capítulo, os processos estocásticos são de dimensão unitária, então, optou-se por explorar apenas as extensões do primeiro algoritmo, para estimação de entropia de transferência, visto que sua formulação computacional é mais simples e objetiva.

Analogamente, Frenzel e Pompe (2007) [15] generalizam o primeiro algoritmo supracitado para formulação do estimador para informação mútua condicionada e, como con-

sequência da formulação matemática (3.3), este estimador é também um estimador para entropia de transferência,

$$\begin{aligned}\widehat{T}_{\mathbf{X} \rightarrow \mathbf{Y}} &= \widehat{I}(X; Y|Z) \\ &= \psi(k) + \frac{1}{N} \sum_{i=1}^N [\psi(n_z(i) + 1) - \psi(n_{xz}(i) + 1) - \psi(n_{yz}(i) + 1)],\end{aligned}\quad (3.12)$$

em que $n_z(i)$, $n_{xz}(i)$ e $n_{yz}(i)$ são o número de amostras inseridas nas regiões projetadas no subespaço Z , XZ e YZ , respectivamente, e delimitadas pelo k -ésimo vizinho.

Seguindo a mesma formulação, propõe-se o seguinte estimador para entropia de transferência direta,

$$\begin{aligned}\widehat{D}_{\mathbf{X} \rightarrow \mathbf{Y}} &= \widehat{I}(X; Y|Z, W) \\ &= \psi(k) + \frac{1}{N} \sum_{i=1}^N [\psi(n_{zw}(i) + 1) - \psi(n_{xzw}(i) + 1) - \psi(n_{yzw}(i) + 1)],\end{aligned}\quad (3.13)$$

em que $n_{zw}(i)$, $n_{xzw}(i)$ e $n_{yzw}(i)$ são o número de amostras inseridas nas regiões projetadas no subespaço ZW , XZW e YZW , respectivamente, e delimitadas pelo k -ésimo vizinho.

3.4 Resultados

Nas aplicações subseqüentes recorre-se estimadores (3.12) e (3.13), e em todas as situações considerou-se $k = 4$. A escolha baseia-se no trabalho de Kraskov, Stögbauer e Grassberger (2004) [13] que sugere que em aplicações práticas, deve-se escolher $k > 1$ para reduzir erros estatísticos, evitar valores altos de k e usar tipicamente $k = 2, 3$, ou 4 . Por razões computacionais, os comprimentos das sequências de variáveis aleatórias analisadas foram considerados iguais a 1.

3.4.1 Modelo Autorregressivo Tridimensional

Esta aplicação é fundamentalmente importante pois, entre processos estocásticos Gaussianos, podem ser obtidos os resultados analíticos das medidas de entropia de transferência e entropia de transferência direta. Desta forma, é possível analisar a confiabilidade dos estimadores (3.12) e (3.13) para a análise de causalidade e conectividade.

Considerando o modelo autorregressivo tridimensional,

$$\begin{cases} X_{n+1} = \alpha X_n + \eta_n^X \\ Y_{n+1} = \beta X_n + \gamma Y_n + \eta_n^Y \\ Z_{n+1} = \delta Y_n + \epsilon Z_n + \eta_n^Z \end{cases}, \quad (3.14)$$

em que $\eta_n^X, \eta_n^Y, \eta_n^Z \sim \mathcal{N}(0; 1)$, quando $0 < \alpha, \beta, \gamma, \delta, \epsilon < 1$ (pólos definidos dentro do círculo de raio unitário), existe fluxo de informação direto de \mathbf{X} para \mathbf{Y} e de \mathbf{Y} para \mathbf{Z} , e o fluxo de informação de \mathbf{X} para \mathbf{Z} é indireto através da variável intermediária \mathbf{Y} .

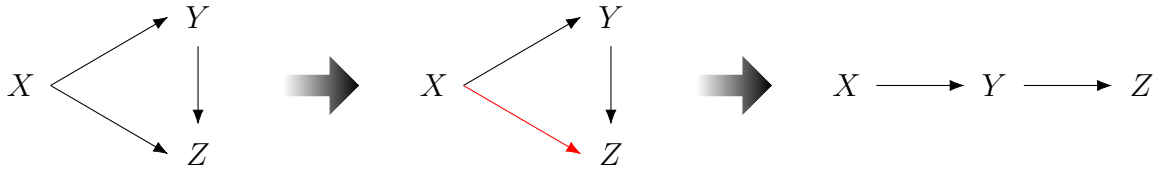


Figura 3.3 – Grafo direcional do fluxo de informação para o modelo autorregressivo tridimensional (B.1) com $\alpha = 0.8$, $\beta = 0.6$, $\gamma = 0.4$, $\delta = 0.5$ e $\epsilon = 0$. Os dois primeiros grafos são resultantes do cálculo da entropia de transferência que representa a causalidade total do sistema. O último grafo é obtido após a análise com a entropia de transferência direta e corresponde à causalidade direta e verdadeira, interpretada como a conectividade do sistema.

A análise teórica de causalidade e conectividade desse sistema inicia com a definição arbitrária do conjunto de parâmetros que o descrevem, portanto, $\alpha = 0,8$, $\beta = 0,6$, $\gamma = 0,4$, $\delta = 0,5$ e $\epsilon = 0$. Analiticamente, a entropia de transferência, considerando todos os possíveis pares de processos presentes nesse sistema, é dada por: $T_{\mathbf{X} \rightarrow \mathbf{Y}} = 0,2365 \text{ nats}$, $T_{\mathbf{Y} \rightarrow \mathbf{Z}} = 0,2516 \text{ nats}$, $T_{\mathbf{X} \rightarrow \mathbf{Z}} = 0,0666 \text{ nats}$ e $T_{\mathbf{Y} \rightarrow \mathbf{X}} = T_{\mathbf{Z} \rightarrow \mathbf{X}} = T_{\mathbf{Z} \rightarrow \mathbf{Y}} = 0$. Assim, as relações de causalidade são: \mathbf{X} causa \mathbf{Y} , \mathbf{Y} causa \mathbf{Z} e \mathbf{X} causa \mathbf{Z} .

Embora sejam assumidos três caminhos para o fluxo de informação, eles não representam a conectividade verdadeira do sistema. Nesse caso, é necessário verificar se existe troca de informação direta do processo \mathbf{X} para \mathbf{Z} , ou se o processo \mathbf{Y} é uma variável intermediária que está influenciando esse fluxo de informação.

Para tal análise, recorre-se à entropia de transferência direta. Visto que $D_{\mathbf{X} \rightarrow \mathbf{Z}} = 0$, então, não existe causalidade direta de \mathbf{X} para \mathbf{Z} . Portanto, conclui-se a análise teórica obtendo como resultado o grafo de representação da conectividade do sistema, ilustrado na Figura 3.3. Um detalhamento analítico da entropia de transferência e entropia de transferência direta é apresentado no Apêndice B.

Seguindo para a análise de convergência dos estimadores, são simuladas 8000 amostras de cada processo, dentre as quais foram descartadas as 3000 primeiras de modo a garantir a estabilidade dos processos autorregressivos [27]. À medida que o número de amostras que compõe o processo de estimação cresce, espera-se que as estimativas convirjam para os valores analíticos.

As relações de causalidade são identificadas por meio do estimador (3.12) e aquelas estimativas negativas ou positivas e da ordem de 10^{-3} , considera-se que não há relação de causalidade entre os processos. Neste exemplo, para as estimativas positivas, essa convergência é evidenciada no gráfico da Figura 3.4. Além disso, as estimativas para todas as possíveis relações de causalidade estão dispostas na Tabela 3.1.

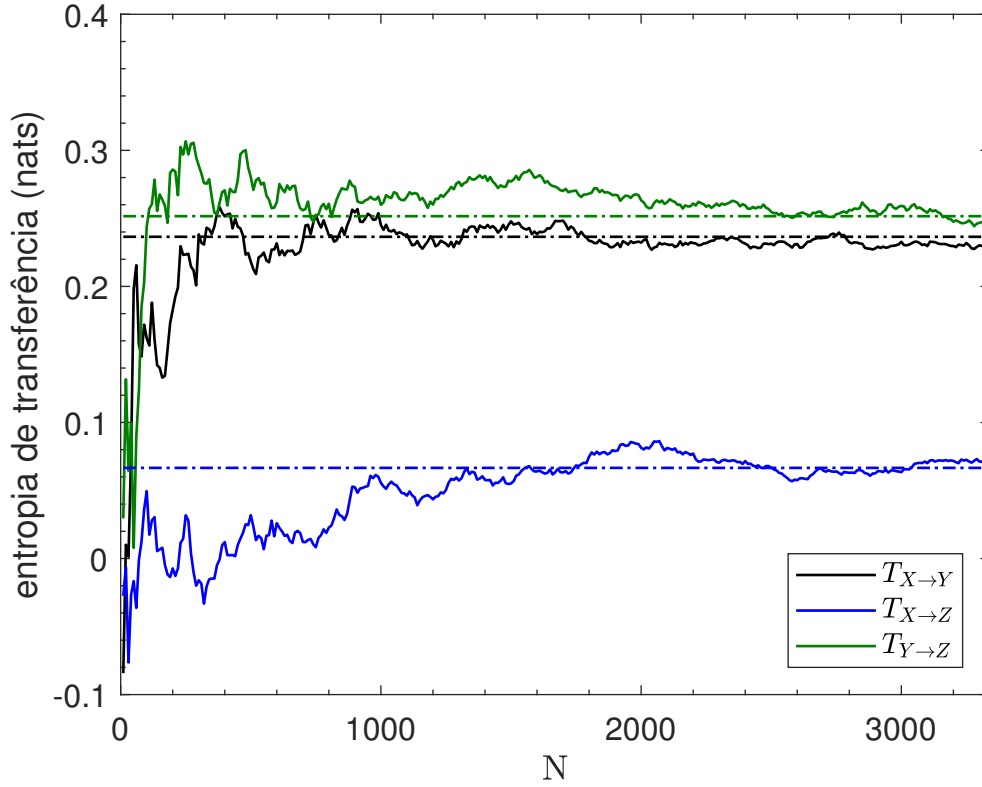


Figura 3.4 – Convergência do estimador de entropia de transferência baseado nas distâncias entre vizinho para para as relações com valores analíticos maiores que zero: $T_{\mathbf{X} \rightarrow \mathbf{Y}} = 0,2365$, $T_{\mathbf{Y} \rightarrow \mathbf{Z}} = 0,2516$ and $T_{\mathbf{X} \rightarrow \mathbf{Z}} = 0,0666$.

Tabela 3.1 – Valores estimados de entropia de transferência usando o estimador (3.12) para o sistema autorregressivo tridimensional (B.1) com $\alpha = 0.8$, $\beta = 0.6$, $\gamma = 0.4$, $\delta = 0.5$ e $\epsilon = 0$.

	$\hat{T}_{lin \rightarrow col}$		
	X	Y	Z
X	NA	0.2357	0.0606
Y	-0.0127	NA	0.2439
Z	-0.0036	-0.0361	NA

Assim como na análise teórica, para determinar a conectividade verdadeira do sistema é necessário determinar se existe causalidade direta entre **X** e **Z**. Essa relação é identificada por meio do estimador (3.13), e, novamente, em caso de estimativa negativa ou positivas e da ordem de 10^{-3} , esta será considerada nula. Visto que, $\hat{D}_{\mathbf{X} \rightarrow \mathbf{Z}} = -0,0066$, então, não existe causalidade direta de **X** para **Z**. O grafo direcional de conectividade desse sistema, obtido por meio de simulação e desconsiderando qualquer conhecimento estatístico dos processos, é exatamente igual ao obtido analiticamente e apresentado na Figura 3.3.

3.4.2 Causalidade e Conectividade em Sistema com Quatro Tanques

O sistema é composto por quatro reservatórios com água conectados conforme ilustrado na Figura 3.5. A bomba bombeia a água para os tanques 4 e 3, além disso, a vazão do tanque 4 e tanque 3 é conduzida para as entradas do tanque 1 e tanque 2, respectivamente.

Cada tanque que compõe o sistema foi simulado de acordo com o modelo base disponibilizado pela MathWorks no qual, a água entra pelo topo a uma taxa proporcional à tensão aplicada à bomba e sai através de uma abertura na base do tanque a uma taxa proporcional à raiz quadrada da altura da água no tanque. A equação diferencial que descreve cada tanque é

$$\frac{d}{dt}Vol = A\frac{dh}{dt} = bV - a\sqrt{h} \quad (3.15)$$

em que Vol é o volume de água, h é a altura de água no tanque, V é a tensão aplicada à bomba, A é a área da seção transversal do tanque, b é uma constante relacionada à vazão de entrada no tanque e a é uma constante relacionada à vazão de saída do tanque.

A Tabela 3.2 relaciona os parâmetros de operação de cada tanque para simulação do sistema. A bomba apresenta comportamento gaussiano, na qual, a taxa de fluxo da água para fora da bomba varia com o valor médio $4 \text{ m}^3/\text{s}$ e variância $0.1 \text{ m}^3/\text{s}$.

As estimativas de entropia de transferência entre cada par dos processos \mathbf{x} , \mathbf{h}_1 , \mathbf{h}_2 , \mathbf{h}_3 e \mathbf{h}_4 estão dispostas na Tabela 3.3. Para aquelas estimativas negativas ou positivas e da ordem de 10^{-3} , considera-se que não há relação de causalidade entre os processos.

Tabela 3.2 – Parâmetros de operação dos quatro tanques que compõem o sistema da Figura 3.5 em que $i = 1, 2, 3, 4$.

Parâmetro	Valor
A_i	1,44; 1,44; 2,25; 2,25
a_i	1; 1; 2; 2
b_i	0,85; 0,6; 0,4; 0,9

Tabela 3.3 – Valores estimados de entropia de transferência usando o estimador (3.12) para o sistema com quatro tanques ilustrado na Figura 3.5.

	$\widehat{T}_{lin \rightarrow col}$				
	\mathbf{h}_1	\mathbf{h}_2	\mathbf{h}_3	\mathbf{h}_4	\mathbf{x}
\mathbf{h}_1	NA	0,1017	0,0171	0,0086	-0,0024
\mathbf{h}_2	0,3618	NA	0,0013	-0,0051	-0,0003
\mathbf{h}_3	0,3136	0,4269	NA	-0,0037	0,0004
\mathbf{h}_4	0,2001	0,2858	-0,0045	NA	-0,0018
\mathbf{x}	-0,0487	-0,0462	0,9229	1,0876	NA

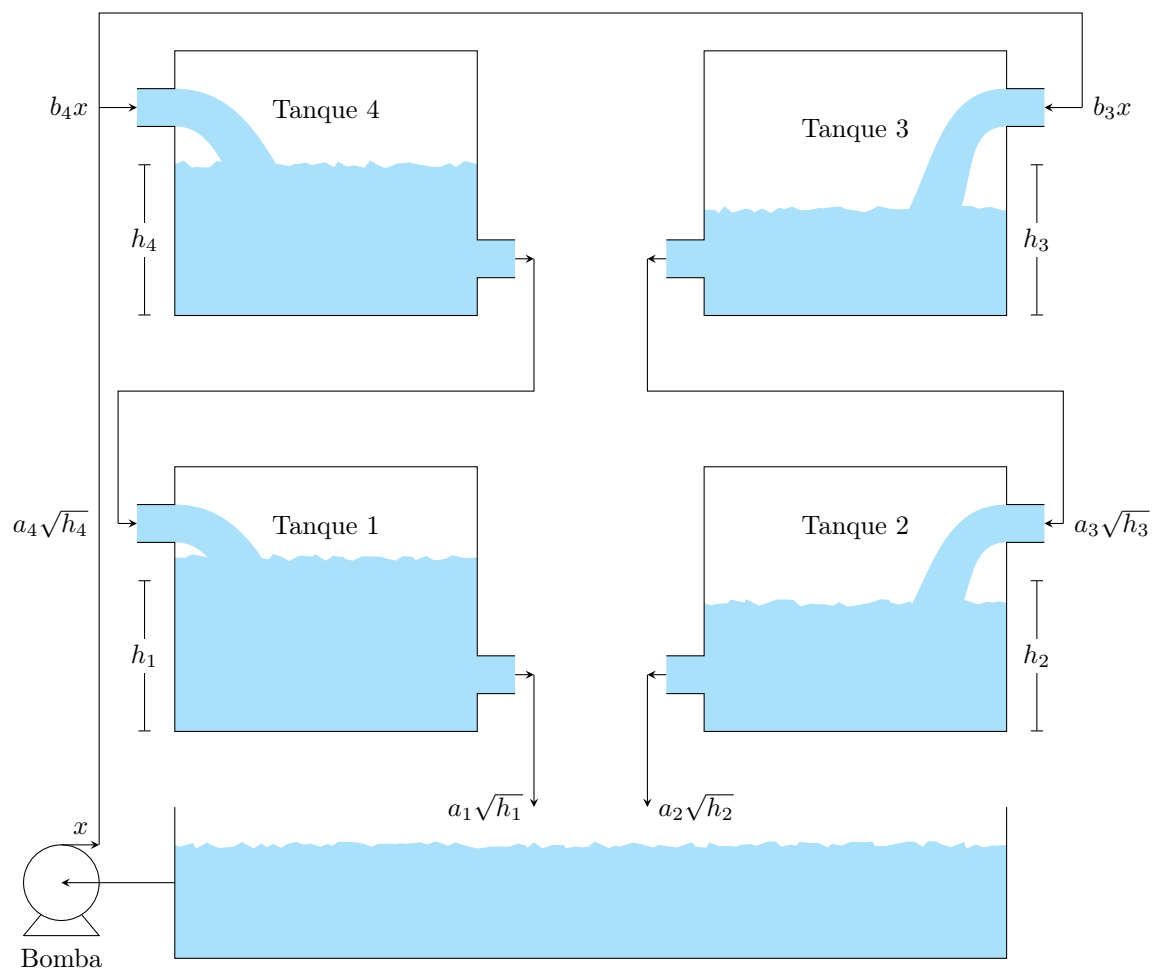


Figura 3.5 – Esquemático do sistema com quatro tanques (reservatórios) com água e uma bomba cuja taxa de fluxo da água varia com o valor médio $4 \text{ m}^3/\text{s}$ e variância $0.1 \text{ m}^3/\text{s}$. em um comportamento gaussiano.

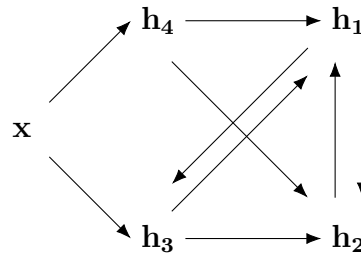


Figura 3.6 – Grafo direcional do fluxo de informação para o sistema com quatro tanques com água resultante do cálculo da entropia de transferência que representa a causalidade total do sistema.

Tabela 3.4 – Entropia de transferência direta para relações direcionais que podem ter uma ou mais variáveis intermediárias.

	Processos intermediários	$\widehat{D}_{\mathbf{x} \rightarrow \mathbf{y}}$
$\mathbf{h}_1 \rightarrow \mathbf{h}_2$	$\mathbf{h}_3, \mathbf{h}_4, \mathbf{x}$	0,00005
$\mathbf{h}_2 \rightarrow \mathbf{h}_1$	\mathbf{h}_4, \mathbf{x}	-0,0002
$\mathbf{h}_4 \rightarrow \mathbf{h}_2$	$\mathbf{h}_1, \mathbf{h}_3, \mathbf{x}$	0,0002
$\mathbf{h}_3 \rightarrow \mathbf{h}_1$	\mathbf{h}_2, \mathbf{x}	0,0071
$\mathbf{h}_1 \rightarrow \mathbf{h}_3$	\mathbf{x}	0,0001
$\mathbf{h}_4 \rightarrow \mathbf{h}_1$	\mathbf{h}_2	0,0432
$\mathbf{h}_3 \rightarrow \mathbf{h}_2$	\mathbf{h}_1	0,0689

Sendo assim, com base na Tabela 3.3, o fluxo de informação nesse sistema é ilustrado no grafo da Figura 3.6. Entretanto, o objetivo é identificar a verdadeira conectividade do sistema, isto é, determinar se essas relações são diretas ou indiretas. Para tal análise, recorre-se à entropia de transferência direta.

Na Tabela 3.4 estão relacionadas as estimativas da entropia de transferência direta bem como os possíveis processos intermediários entre um par de sinais. Da mesma forma, aquelas estimativas negativas ou positivas e da ordem de 10^{-3} , considera-se que não há relação de conectividade entre os processos. As etapas de cálculo e identificação das vias de fluxo de informação são apresentadas na Figura 3.7.

Primeiro verificou-se se a causalidade entre \mathbf{h}_1 e \mathbf{h}_2 é verdadeira e direta considerando \mathbf{h}_3 como as possível variável intermediária e \mathbf{h}_4 e \mathbf{x} como possíveis variáveis fonte comum. Visto que a entropia de transferência direta ente \mathbf{h}_1 e \mathbf{h}_2 é muito pequena, conclui-se que não há causalidade direta entre eles. Em seguida, investigou-se outras quatro relações de causalidade direta de $\mathbf{h}_2 \rightarrow \mathbf{h}_1$, $\mathbf{h}_4 \rightarrow \mathbf{h}_2$, $\mathbf{h}_3 \rightarrow \mathbf{h}_1$ e $\mathbf{h}_1 \rightarrow \mathbf{h}_3$, contudo, todas essas estimativas ou são negativas ou são pequenas, descartando-se assim, a conectividade dessas variáveis.

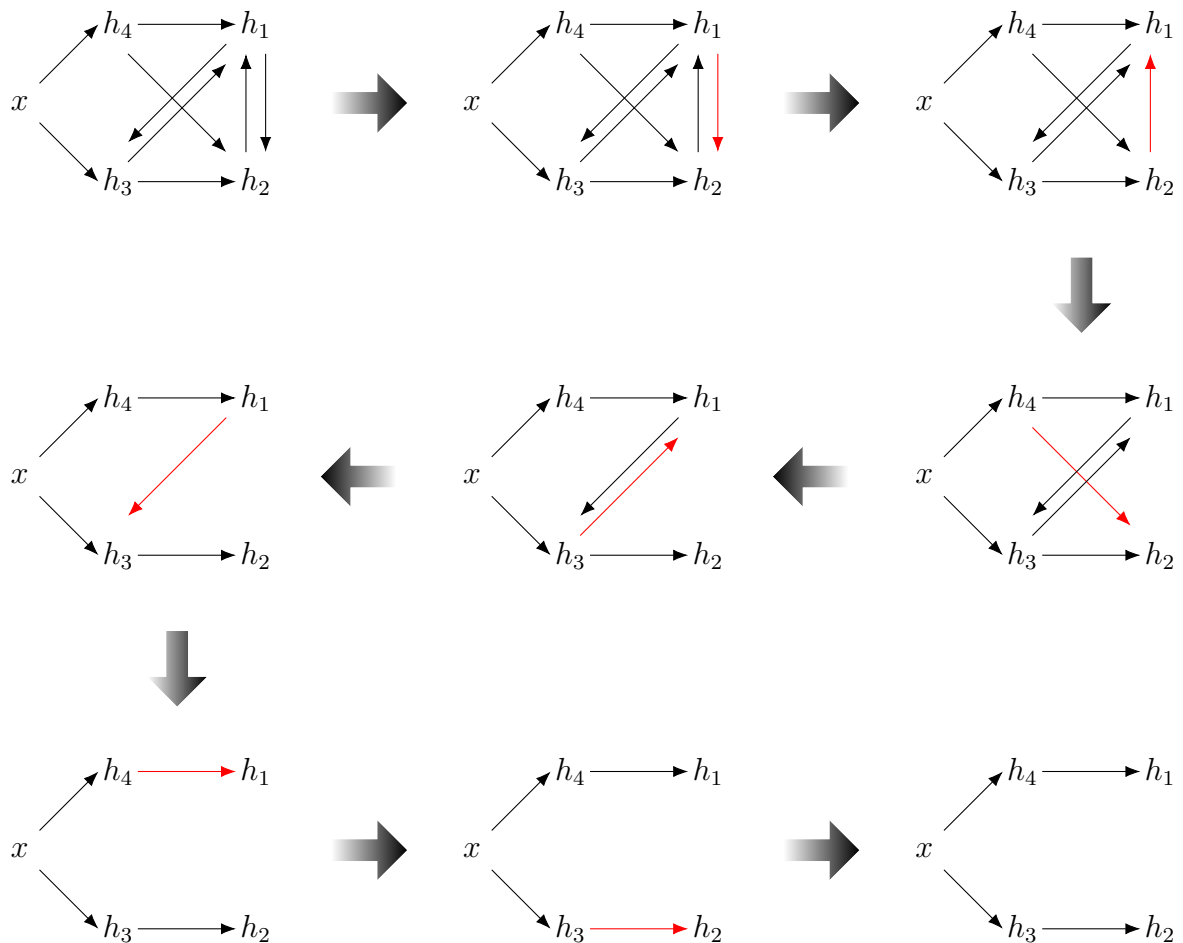


Figura 3.7 – Grafos direcionais do fluxo de informação para o sistema com quatro tanques com água. O primeiro grafo é resultante do cálculo da entropia de transferência que representa a causalidade total do sistema. Os grafos seguintes são obtidos sequencialmente a partir da análise de conectividade das variáveis que estão interligadas por meio de uma seta vermelha. O último grafo corresponde à causalidade direta e verdadeira, interpretada como a conectividade do sistema.

Além disso, não é necessário detectar se existe causalidade em $x \rightarrow h_3$ e $x \rightarrow h_4$ uma vez que não existe uma variável intermediária ou uma variável que é fonte comum em suas vias. Por fim, com base em todas as estimativas conclui-se que, exceto pela causalidade de $h_4 \rightarrow h_1$ e $h_3 \rightarrow h_2$, as demais causalidades detectadas são indiretas ou espúrias. O último grafo da Figura 3.7 corresponde à verdadeira conectividade do sistema.

3.4.3 Identificação de Padrão para Diagnóstico de Traumatismo Craniano

A lesão cerebral traumática (TBI, do inglês: *Traumatic Brain Injury*) é uma agressão não degenerativa e não-congênita ao cérebro causada por uma força mecânica externa, levando a danos permanentes ou temporários das funções cognitivas, físicas e psicossociais, com um estado associado de redução ou alteração da consciência [45].

Informações sobre os casos de TBI no Brasil estão disponíveis no banco de dados do Departamento de Informática do Sistema Único de Saúde (DATASUS), mantido pelo Ministério da Saúde e disponível online em <http://datasus.saude.gov.br/>. Segundo os registros, entre janeiro de 2008 e dezembro de 2016, ocorreram mais de 130 mil internações hospitalares com diagnóstico de TBI, correspondendo a uma incidência de 65,2 internações por 100 mil habitantes por ano. Além disso, o custo médio das despesas hospitalares por internação foi de quase US\$ 500 e a taxa de mortalidade foi de 7,6%.

Nesse cenário, com o objetivo de auxiliar no diagnóstico médico, foram analisados para indivíduos saudáveis e diagnosticados com TBI e lesões associadas, sinais de pressão sanguínea (ABP, do inglês: *Arterial Blood Pressure*) e eletrocardiograma (ECG). Estes sinais integram dois bancos de dados de acesso livre da PhysioNet [46] e estão disponibilizados sob a licença e dedicação de domínio público v1.0 (PDDL, do inglês: *Public Domain Dedication and License*). Além disso, é oferecida uma biblioteca de funções (WFDB, do inglês: *WaveForm DataBase*), para leitura, escrita e processamento dos sinais fisiológicos que compõem a PhysioNet [47].

No caso dos indivíduos saudáveis e que não apresentam sinais de doença cardiovascular, os sinais estão disponíveis na base de resposta fisiológica à mudanças na postura (PRCP, do inglês: *Physiologic Response to Changes in Posture*) [48]. Enquanto que para os indivíduos com diagnóstico de lesão craniana, os sinais estão disponíveis na base de dados de sistema de informação autorregulatório hemodinâmico cerebral (CHARIS, do inglês: *Cerebral Haemodynamic Autoregulatory Information System*) [49].

O conjunto de amostras dessas bases compreende vinte e um sujeitos (9 homens, 8 mulheres e 4 não identificados) com idade média de $38,9 \pm 18,6$ anos, entre os quais, dez eram voluntários saudáveis e onze com diagnóstico de lesão cerebral traumática. Séries temporais de sinais ABP e ECG para quatro voluntários em cada banco de dados são apresentadas nas Figuras 3.8 e 3.9.

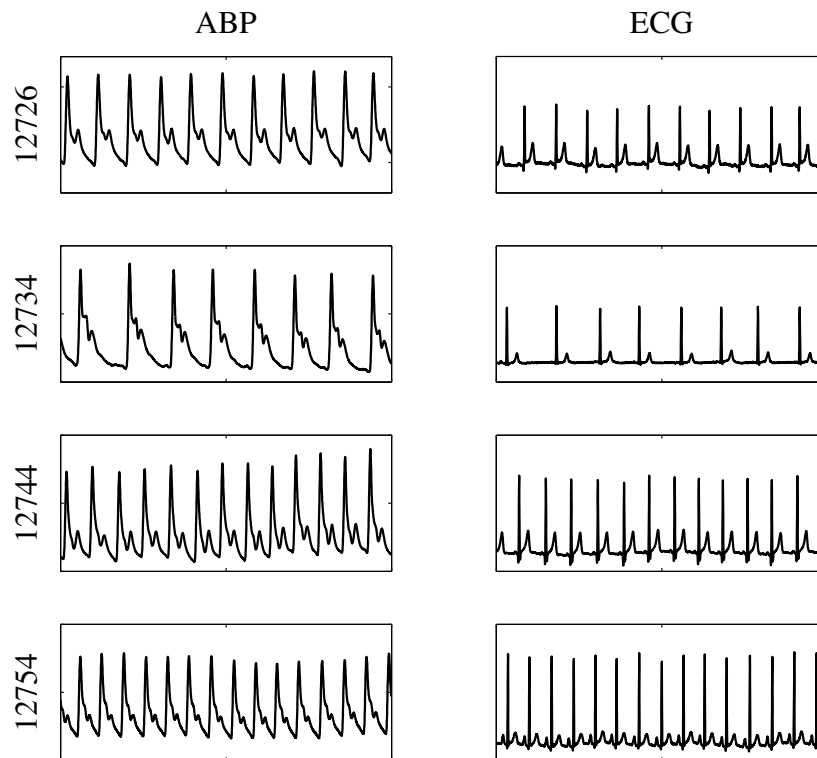


Figura 3.8 – Sinais de pressão sanguínea (ABP) e eletrocardiograma (ECG) para indivíduos saudáveis.

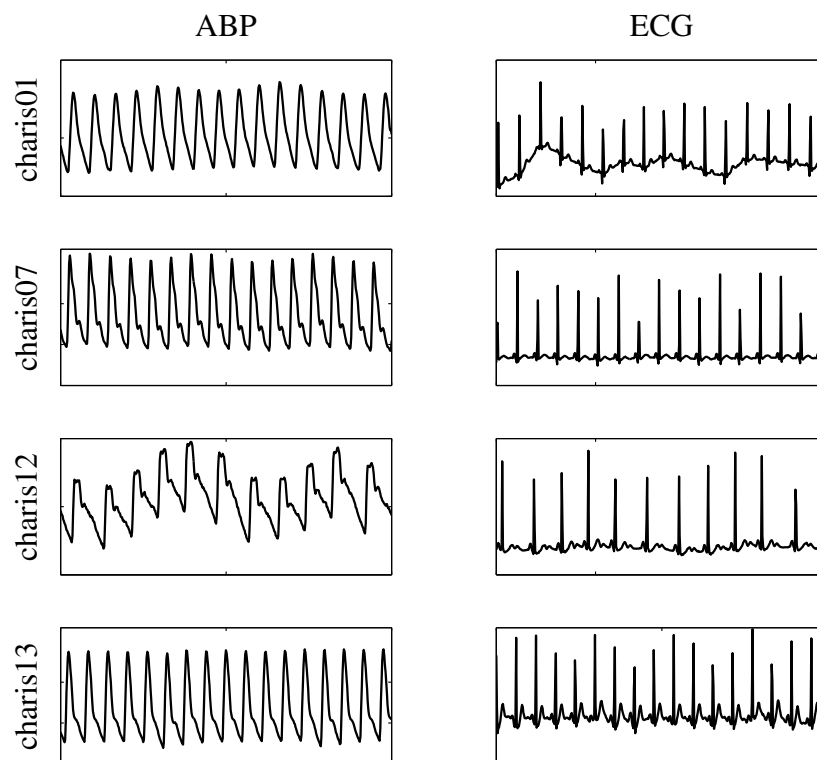


Figura 3.9 – Sinais de pressão sanguínea (ABP) e eletrocardiograma (ECG) para indivíduos diagnosticados com traumatismo craniano e lesões associadas.

Tabela 3.5 – Média e desvio padrão de 10 amostras, cada uma com 30 segundos, para entropia de transferência (em bits). São considerados os dois possíveis sentidos de fluxo de informação entre os sinais de ABP e ECG para cada voluntário saudável que compõe a base de dados PRCP.

Voluntário	ABP \rightarrow ECG	ECG \rightarrow ABP
12726	0,2063 \pm 0,0303	0,3435 \pm 0,0139
12734	0,1244 \pm 0,0085	0,1340 \pm 0,0165
12744	0,2849 \pm 0,0086	0,3718 \pm 0,0093
12754	0,1890 \pm 0,0246	0,2765 \pm 0,0311
12755	0,2686 \pm 0,0284	0,4015 \pm 0,0335
12814	0,1488 \pm 0,0075	0,2414 \pm 0,0158
12815	0,1645 \pm 0,0069	0,3121 \pm 0,0362
12819	0,3095 \pm 0,0173	0,3707 \pm 0,0286
12821	0,2748 \pm 0,0052	0,4431 \pm 0,0073
13960	0,1312 \pm 0,0242	0,3051 \pm 0,0404
$\mu \pm \sigma$	0,2102 \pm 0,0161	0,3200 \pm 0,0233

Tabela 3.6 – Média e desvio padrão de 10 amostras, cada uma com 30 segundos, para entropia de transferência (em bits). São considerados os dois possíveis sentidos de fluxo de informação entre os sinais de ABP e ECG para cada voluntário diagnosticado com TBI ou lesões associadas que compõe a base de dados CHARIS DB.

Voluntário	ABP \rightarrow ECG	ECG \rightarrow ABP
charis01	0,4945 \pm 0,1319	0,2652 \pm 0,1083
charis03	0,5935 \pm 0,0585	0,3752 \pm 0,0211
charis04	0,3548 \pm 0,0990	0,0305 \pm 0,0260
charis06	0,6110 \pm 0,0514	0,4666 \pm 0,0729
charis07	0,7500 \pm 0,0514	0,5089 \pm 0,0384
charis08	0,7672 \pm 0,0666	0,2403 \pm 0,0345
charis09	0,3529 \pm 0,0731	0,1000 \pm 0,0670
charis10	0,5762 \pm 0,0698	0,3665 \pm 0,0323
charis11	0,7900 \pm 0,1111	0,4138 \pm 0,0302
charis12	0,1611 \pm 0,0837	0,1244 \pm 0,0781
charis13	0,8637 \pm 0,0601	0,5200 \pm 0,0220
$\mu \pm \sigma$	0,5741 \pm 0,0779	0,3101 \pm 0,0483

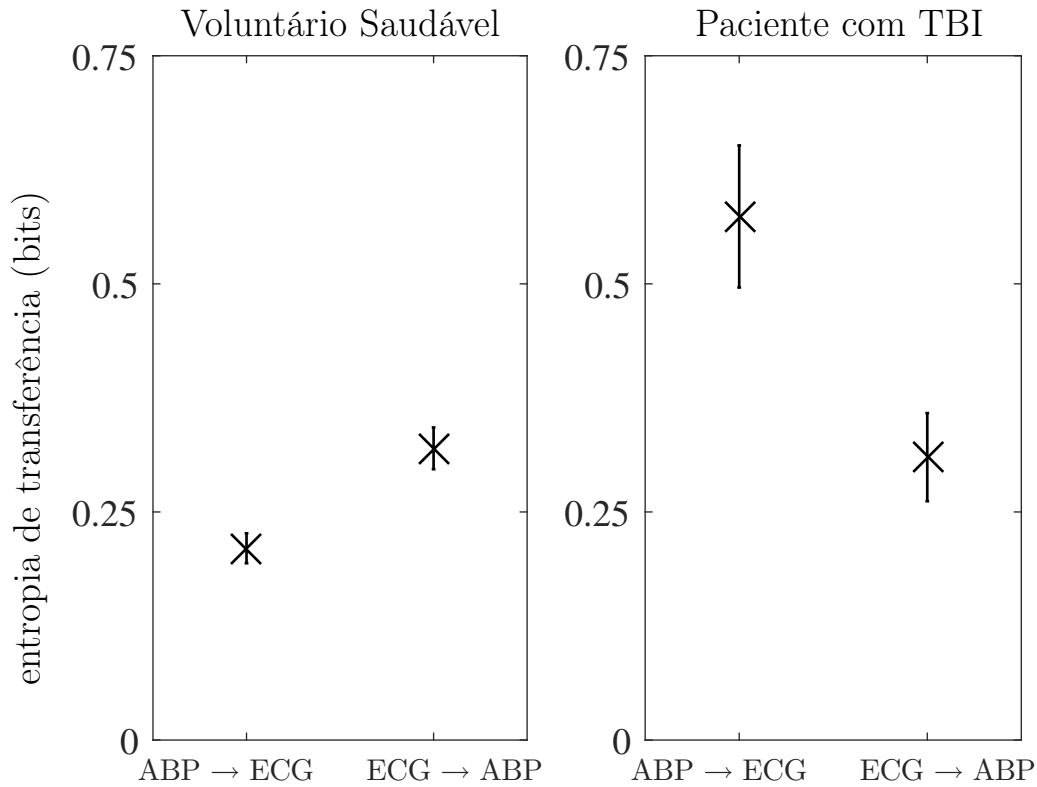


Figura 3.10 – Gráfico de barra de erro vertical para a média e o desvio padrão médio das medidas de entropia de transferência para indivíduos saudáveis e indivíduos diagnosticados com TBI e lesões associadas.

Portanto, com o auxílio da entropia de transferência, foram identificados padrões ao que se refere à direcionalidade do fluxo de informação entre os sinais de ABP e ECG, permitindo assim, a caracterização de indivíduos saudáveis e com patologia.

Considerando os sinais fisiológicos de cada sujeito estimou-se a entropia de transferência em ambas as direcionalidades para 10 amostras diferentes de 30 segundos cada e, foram anotados a média (μ) e o desvio padrão (σ) das medidas nas Tabelas 3.5 e 3.6.

Para indivíduos saudáveis, a entropia de transferência média no sentido ABP para ECG é menor que no sentido ECG para ABP. Contudo, para pacientes diagnosticados com TBI e lesões associadas ocorre o inverso, a medida no sentido ABP para ECG é maior que no sentido ECG para ABP. Além disso, o desvio padrão médio das estimativas de causalidade dos sinais de pessoas saudáveis é menor quando comparadas às estimativas dos sinais de indivíduos com patologia.

Na Figura 3.10 têm-se um gráfico de barra de erro vertical no qual, o ponto central representa a média e os valores máximos e mínimos indicam o desvio padrão médio das medidas de entropia de transferência.

3.5 Considerações

A identificação das relações causais e da conectividade entre sinais contínuos de um sistema complexo é desafiador, visto que, em geral, o conhecimento sobre a dinâmica subjacente é parcial. Entretanto, a teoria da informação fornece uma variedade de abordagens para medir essas relações entre séries temporais multivariadas, por exemplo, a entropia de transferência.

De acordo com a abordagem apresentada, valores não nulos da entropia de transferência indicam uma relação de causalidade entre os sinais analisados e, valores não nulos da entropia de transferência direta indicam conectividade entre os sinais. O estudo analítico de causalidade e conectividade do sistema tridimensional autorregressivo, combinado com simulações que desconsideram qualquer conhecimento estatístico dos processos, é essencial para demonstrar, nesses casos, a confiabilidade dos estimadores baseados nas distâncias de vizinhos, (3.12) e (3.13).

A detecção de causalidade e conectividade no sistema com quatro reservatórios com água é a aplicação base para introduzir a discussão da aplicabilidade do método em processos industriais. Assim como para a análise do sistema autorregressivo, nesse caso também foi possível recuperar a conectividade dos sistemas assumindo o desconhecimento estatístico dos processos.

Além disso, a partir dos resultados com sinais fisiológicos, pode-se introduzir uma nova discussão sobre metodologias de suporte para decisão clínica em casos de lesões cerebrais. Embora, para as duas classes analisadas: saudáveis e com diagnóstico de TBI, o fluxo médio de informação no sentido ECG para ABP sejam aproximadamente iguais, o suporte clínico está em identificar, para os sinais de ABP e ECG, em qual direcionalidade o fluxo de informação é maior.

Capítulo 4

Conclusões

Um dos desafios da teoria da informação é identificar causalidade entre sinais a partir da estimação de medidas de informação. Esta dissertação apresenta resultados ao que concerne uma nova metodologia para estimação de entropia diferencial de Shannon bem como metodologia para análise de causalidade em sinais biomédicos para diagnósticos de lesões cerebrais e, análise de causalidade e conectividade em sinais provenientes de um sistema com quatro reservatórios com água e uma bomba.

O capítulo 2 foi dedicado ao estudo e análise de quatro estimadores para entropia diferencial de Shannon fundamentados em métodos como: histograma, suavização kernel, distâncias dos vizinhos mais próximos e aprendizado supervisionado de máquina. Além disso, foi proposta uma metodologia para caracterização do suporte de uma variável aleatória a fim de expandir a aplicabilidade de algoritmos de aprendizado supervisionado, como máquina de vetor de suporte (SVM, do inglês: *Support Vector Machine*), também para estimação da entropia diferencial de Shannon, que foi justificada a partir da propriedade de equipartição assintótica (AEP, do inglês: *Asymptotic Equipartition Property*).

Com suporte na AEP, para variáveis aleatórias contínuas a relação entre entropia de Rényi de ordem zero (logaritmo da medida de Lebesgue do suporte) e entropia diferencial de Shannon (logaritmo do suporte efetivo) remete à caracterização do suporte no qual cada medida se refere. A aplicabilidade dos estimadores e do método proposto foi evidenciado a partir da simulação de 10^5 amostras independentes e identicamente distribuídas de variáveis aleatórias contínuas com densidade de probabilidade bem definida, dentre as quais: exponencial, gamma, normal e uniforme.

Quando comparados os resultados obtidos para estimação da entropia diferencial de Shannon pelos quatro estimadores supracitados com o valor analítico, observou-se que o algoritmo proposto forneceu resultados assintoticamente bons para as estimativas em sete dos oito casos analisados. O único caso que não convergiu, foi considerado um caso especial e, aconteceu quando a função densidade de probabilidade tender a assumir valor infinito para alguma realização. Nesse caso, o algoritmo estimou todo o suporte da distribuição e

a estimativa convergiu para a entropia de Rényi de ordem zero. Além disso, o algoritmo proposto demandou mais tempo de processamento contudo, em algumas simulações o método convergiu primeiramente em função do número de amostras analisadas.

Ao longo do capítulo 3 foram utilizados os conceitos de entropia de transferência e entropia de transferência direta para identificação das relações de causalidade e conectividade, respectivamente, entre variáveis contínuas. O primeiro sistema analisado foi um modelo autorregressivo tridimensional, fundamentalmente importante pois, entre processos estocásticos gaussianos, é possível obter os resultados analíticos das medidas de informação supracitadas. Desta forma, estas simulações, que desconsideravam qualquer conhecimento estatístico dos processos, foram essenciais para demonstrar a confiabilidade dos estimadores baseados nas distâncias de vizinhos mais próximo na análise de causalidade e conectividade.

Também foi explorada a aplicabilidade da entropia de transferência como medida de causalidade em sinais fisiológicos para auxílio no diagnóstico clínico em casos de lesões cerebral. A partir dos sinais de pressão sanguínea (ABP, do inglês: *Arterial Blood Pressure*) e eletrocardiograma (ECG) de voluntário saudáveis e diagnosticados com algum tipo de lesão cerebral traumática (TBI, do inglês: *Traumatic Brain Injury*), observou-se um comportamento padrão no fluxo de informação entre os dois sinais analisados. Isto é, para indivíduos saudáveis, a entropia de transferência média no sentido ABP para ECG é menor que no sentido ECG para ABP e para pacientes diagnosticados com TBI e lesões associadas ocorre o inverso, a medida no sentido ABP para ECG é maior que no sentido ECG para ABP.

Identificar a conectividade verdadeira de um sistema é importante para distinguir se a influência causal ocorre ao longo de uma via direta sem quaisquer variáveis intermediárias, ou por vias indiretas através de variáveis intermediárias e/ou variáveis que são fonte comum. Nesse sentido, diante da aplicação analisada, um sistema com quatro tanques com água e uma bomba, foi possível, utilizando abordagens da teoria da informação, recuperar fielmente a conectividade do sistema.

Desta forma, conclui-se que o trabalho alcançou os objetivos almejados e trás contribuições no cenário ao qual está incluído.

4.1 Trabalhos Futuros

A pesquisa sobre análise de causalidade e conectividade em sistemas com variáveis contínuas é uma área ativa e ainda existem algumas questões que não foram totalmente exploradas portanto, merecem atenção e continuidade da pesquisa, entre as quais:

- **Estimação de medidas de informação**

Quatro algoritmos para estimação de entropia diferencial de Shannon foram abordados, entre os quais, três são amplamente utilizados na literatura e um foi proposto nessa dissertação. Além disso, o algoritmo baseado nas distâncias entre vizinhos foi explorado para estimação da entropia de transferência na análise de causalidade e conectividade em sistemas contínuos. Contudo, embora as estimativas converjam ao valor analítico para ambas as medidas de informação supracitadas, cada algoritmo apresenta limitações e portanto, esforços ainda podem ser destinados para o desenvolvimento de estimadores mais precisos e eficientes.

- **Estimador SVM de uma classe para entropia diferencial**

Foi proposto um método de caracterização para o algoritmo SVM de uma classe a fim de aplicá-lo na estimação da entropia diferencial de Shannon. Embora em alguns testes o algoritmo seja o primeiro a convergir em função do número de amostras analisadas, quando a variável aleatória apresenta valor infinito na função densidade de probabilidade para alguma realização, o algoritmo converge para entropia diferencial de Rényi. Portanto, nesse cenário, pode-se aprimorar o método de caracterização do suporte a partir de parametrizações mais eficazes e, explorar a análise matemática de variância e convergência do estimador.

- **Análise de causalidade em sinais biomédicos**

A aplicação explorada nessa dissertação visou introduzir um auxílio no diagnóstico de traumatismo craniano a partir da análise de causalidade dos sinais fisiológicos: eletrocardiograma e pressão sanguínea. Nesse cenário, é necessário validar a metodologia proposta e expandir os resultados obtidos. Isso pode ser feito a partir da análise de outras bases de dados e até mesmo a partir de ensaios reais.

- **Análise de causalidade e conectividade em sinais de plantas industriais**

A aplicação explorada nessa dissertação foi introdutória e analisou um sistema não-linear composto por quatro reservatórios com água. Apesar de sistemas industriais também terem sido explorados por Duan et al. (2013) [27], exploramos o uso do estimador baseado nas distâncias dos vizinhos nessa aplicação. Nesse cenário, é interessante investigar a detecção de faltas a partir dos grafos de conectividade reproduzidos. Além disso, é pertinente aumentar o grau de complexidade do sistema analisado.

4.2 Produção Científica

Como resultado deste trabalho, dois artigos foram publicados ou aceitos em congresso nacional da área, sendo eles:

- “Transfer Entropy Characterization of TBI cases”
Autores: Milena M. Arruda, Luciana R. Veloso e Francisco M. de Assis
Publicado em: XXXV Simpósio Brasileiro de Telecomunicações e Processamento de Sinais (SBrT), 2017.
- “Differential Entropy Estimation via One-Class SVM”
Autores: Milena M. Arruda, Luciana R. Veloso e Francisco M. de Assis
Aceito em: XXXVI Simpósio Brasileiro de Telecomunicações e Processamento de Sinais (SBrT), 2018.

Referências

- 1 SHANNON, C. E. A mathematical theory of communication. *The Bell System Technical Journal*, v. 27, p. 379–423, july, october 1948. Citado nas páginas 18 e 24.
- 2 KULLBACK, S.; LEIBLER, R. A. On information and sufficiency. *Ann. Math. Statist.*, v. 22, n. 1, p. 79–86, 1951. Citado na página 18.
- 3 MARKO, H. The bidirectional communication theory - a generalization of information theory. *IEEE Transactions on Communications*, v. 21, n. 12, p. 1345–1351, December 1973. ISSN 0090-6778. Citado nas páginas 18 e 38.
- 4 MASSEY, J. L. *Causality, Feedback And Directed Information*. 1990. Citado nas páginas 18 e 38.
- 5 SCHREIBER, T. Measuring information transfer. *Phys. Rev. Lett.*, American Physical Society, v. 85, p. 461–464, Jul 2000. Citado nas páginas 18, 38 e 39.
- 6 WILLEMS, F. M. J.; SHTARKOV, Y. M.; TJALKENS, T. J. The context-tree weighting method: basic properties. *IEEE Transactions on Information Theory*, v. 41, n. 3, p. 653–664, May 1995. ISSN 0018-9448. Citado na página 19.
- 7 JIAO, J.; PERMUTER, H. H.; ZHAO, L.; KIM, Y. H.; WEISSMAN, T. Universal estimation of directed information. *IEEE Transactions on Information Theory*, v. 59, n. 10, p. 6220–6242, Oct 2013. ISSN 0018-9448. Citado nas páginas 19, 20, 21 e 38.
- 8 TREVES, A.; PANZERI, S. The upward bias in measures of information derived from limited data samples. *Neural Computation*, v. 7, n. 2, p. 399–407, March 1995. ISSN 0899-7667. Citado na página 19.
- 9 STRONG, S. P.; KOBERLE, R.; STEVENINCK, R. R. d. R. van; BIALEK, W. Entropy and information in neural spike trains. *Physical review letters*, APS, v. 80, n. 1, p. 197, 1998. Citado na página 19.
- 10 BOWMAN, A. W.; AZZALINI, A. *Applied smoothing techniques for data analysis: the kernel approach with S-Plus illustrations*. [S.l.]: Oxford: Clarendon Press; New York: Oxford University Press, 1997. ISBN 0198523963. Citado na página 19.
- 11 KOZACHENKO, L.; LEONENKO, N. N. Sample estimate of the entropy of a random vector. *Problemy Peredachi Informatsii*, Russian Academy of Sciences, Branch of Informatics, Computer Equipment and Automatization, v. 23, n. 2, p. 9–16, 1987. Citado nas páginas 19, 25, 30, 42 e 64.
- 12 CHARZYŃSKA, A.; GAMBIN, A. Improvement of the k-nn entropy estimator with applications in systems biology. *Entropy*, Multidisciplinary Digital Publishing Institute, v. 18, n. 1, p. 13, 2015. Citado na página 19.

- 13 KRASKOV, A.; STÖGBAUER, H.; GRASSBERGER, P. Estimating mutual information. *Phys. Rev. E*, American Physical Society, v. 69, p. 066138, Jun 2004. Citado nas páginas 19, 25, 39, 42, 43 e 44.
- 14 GAO, W.; OH, S.; VISWANATH, P. Demystifying fixed k-nearest neighbor information estimators. In: *2017 IEEE International Symposium on Information Theory (ISIT)*. [S.l.]: IEEE, 2017. Citado nas páginas 19 e 34.
- 15 FRENZEL, S.; POMPE, B. Partial mutual information for coupling analysis of multivariate time series. *Phys. Rev. Lett.*, American Physical Society, v. 99, p. 204101, Nov 2007. Citado nas páginas 19, 39 e 43.
- 16 SCHÖLKOPF, B.; PLATT, J. C.; SHAWE-TAYLOR, J.; SMOLA, A. J.; WILLIAMSON, R. C. Estimating the support of a high-dimensional distribution. *Neural Computation*, MIT Press - Journals, v. 13, n. 7, p. 1443–1471, jul 2001. Citado nas páginas 19, 20, 25, 31, 65 e 66.
- 17 DUAN, P.; YANG, F.; SHAH, S. L.; CHEN, T. Transfer zero-entropy and its application for capturing cause and effect relationship between variables. *IEEE Transactions on Control Systems Technology*, Institute of Electrical and Electronics Engineers (IEEE), v. 23, n. 3, p. 855–867, may 2015. Citado nas páginas 19, 22, 25, 31, 32, 37 e 66.
- 18 MUKHERJEE, S.; VAPNIK, V. Support vector method for multivariate density estimation. *Center for Biological and Computational Learning. Department of Brain and Cognitive Sciences, MIT. CBCL*, v. 170, 1999. Citado nas páginas 19 e 25.
- 19 MUNOZ, A.; MOGUERZA, J. M. One-class support vector machines and density estimation: The precise relation. In: SPRINGER. *Iberoamerican Congress on Pattern Recognition*. [S.l.], 2004. p. 216–223. Citado nas páginas 19 e 25.
- 20 RÉNYI, A. On measures of entropy and information. In: *Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability, Volume 1: Contributions to the Theory of Statistics*. Berkeley, Calif.: University of California Press, 1961. p. 547–561. Citado nas páginas 19 e 24.
- 21 COVER, T. M.; THOMAS, J. A. *Elements of Information Theory 2nd Edition (Wiley Series in Telecommunications and Signal Processing)*. [S.l.]: Wiley-Interscience, 2006. ISBN 0471241954. Citado nas páginas 19, 25, 26, 27 e 28.
- 22 ASSIS, J. M. de; ASSIS, F. M. de. An application of directed information to infer synaptic connectivity. In: *Anais do XXXIV Simpósio Brasileiro de Telecomunicações e Processamento de Sinais, Santarém, Brazil*. [S.l.: s.n.], 2016. p. 528–532. Citado nas páginas 20, 21 e 38.
- 23 QUINN, C. J.; COLEMAN, T. P.; KIYAVASH, N.; HATSOPOULOS, N. G. Estimating the directed information to infer causal relationships in ensemble neural spike train recordings. *Journal of computational neuroscience*, Springer, v. 30, n. 1, p. 17–44, 2011. Citado nas páginas 20, 21 e 38.
- 24 MARZBANRAD, F.; KIMURA, Y.; ENDO, M.; PALANISWAMI, M.; KHANDOKER, A. H. Transfer entropy analysis of maternal and fetal heart rate coupling. In: *2015 37th Annual International Conference of the IEEE Engineering in*

Medicine and Biology Society (EMBC). [S.l.: s.n.], 2015. p. 7865–7868. ISSN 1094-687X. Citado nas páginas 20, 21 e 38.

25 VICENTE, R.; WIBRAL, M.; LINDNER, M.; PIPA, G. Transfer entropy—a model-free measure of effective connectivity for the neurosciences. *Journal of Computational Neuroscience*, v. 30, n. 1, p. 45–67, Feb 2011. ISSN 1573-6873. Citado nas páginas 20, 21 e 38.

26 YANG, F.; DUAN, P.; SHAH, S. L.; CHEN, T. *Capturing Connectivity and Causality in Complex Industrial Processes*. 1. ed. [S.l.]: Springer International Publishing, 2014. (SpringerBriefs in Applied Sciences and Technology). ISBN 978-3-319-05379-0,978-3-319-05380-6. Citado nas páginas 20 e 22.

27 DUAN, P.; YANG, F.; CHEN, T.; SHAH, S. L. Direct causality detection via the transfer entropy approach. *IEEE Transactions on Control Systems Technology*, Institute of Electrical and Electronics Engineers (IEEE), v. 21, n. 6, p. 2052–2066, nov 2013. Citado nas páginas 20, 22, 38, 41, 45 e 58.

28 BAUER, M.; COX, J. W.; CAVENESS, M. H.; DOWNS, J. J.; THORNHILL, N. F. Finding the direction of disturbance propagation in a chemical process using transfer entropy. *IEEE Transactions on Control Systems Technology*, v. 15, n. 1, p. 12–21, Jan 2007. ISSN 1063-6536. Citado nas páginas 20 e 22.

29 WIENER, N. The theory of prediction. In: *Beckebach EF, editor. Modern mathematics for engineers*. McGraw-Hill, New York, 1956. Citado nas páginas 20 e 38.

30 GRANGER, C. W. J. Investigating causal relations by econometric models and cross-spectral methods. *Econometrica*, [Wiley, Econometric Society], v. 37, n. 3, p. 424–438, 1969. ISSN 00129682, 14680262. Citado nas páginas 20 e 38.

31 BRESSLER, S. L.; SETH, A. K. Wiener–granger causality: A well established methodology. *NeuroImage*, v. 58, n. 2, p. 323 – 329, 2011. ISSN 1053-8119. Citado na página 20.

32 BARNETT, L.; BARRETT, A. B.; SETH, A. K. Granger causality and transfer entropy are equivalent for gaussian variables. *Physical Review Letters*, v. 103, 2009. Citado na página 20.

33 OMAR, C.; AKCE, A.; JOHNSON, M.; BRETL, T.; MA, R.; MACLIN, E.; MCCORMICK, M.; COLEMAN, T. P. A feedback information-theoretic approach to the design of brain–computer interfaces. *Intl. Journal of Human–Computer Interaction*, Taylor & Francis, v. 27, n. 1, p. 5–23, 2010. Citado na página 21.

34 BLOKH, D.; STAMBLER, I. Applying information theory analysis for the solution of biomedical data processing problems. *American Journal of Bioinformatics*, Science Publications, v. 3, n. 1, p. 17, 2014. Citado na página 21.

35 GAO, Y.; REN, L.; LI, R.; ZHANG, Y. Electroencephalogram–electromyography coupling analysis in stroke based on symbolic transfer entropy. *Frontiers in Neurology*, v. 8, p. 716, 2018. ISSN 1664-2295. Citado na página 21.

36 SCHAMBERG, G.; COLEMAN, T. P. A sample path measure of causal influence. *arXiv preprint arXiv:1805.03333*, 2018. Citado na página 21.

- 37 ASSIS, J. M. de; SANTOS, M. O.; ASSIS, F. M. de. Auditory stimuli coding by postsynaptic potential and local field potential features. *PLOS ONE*, Public Library of Science (PLoS), v. 11, n. 8, p. e0160089, aug 2016. Citado nas páginas 21 e 25.
- 38 ZHOU, K.; LIU, T.; ZHOU, L. Industry 4.0: Towards future industrial opportunities and challenges. In: *2015 12th International Conference on Fuzzy Systems and Knowledge Discovery (FSKD)*. [S.l.: s.n.], 2015. p. 2147–2152. Citado na página 22.
- 39 YAO, C.-Z.; KUANG, P.-C.; LIN, Q.-W.; SUN, B.-Y. A study of the transfer entropy networks on industrial electricity consumption. *Entropy*, v. 19, n. 4, p. 159, 2017. Citado na página 22.
- 40 HARTLEY, R. V. Transmission of information. *Bell Labs Technical Journal*, Wiley Online Library, v. 7, n. 3, p. 535–563, 1928. Citado na página 24.
- 41 RUDIN, W. et al. *Principles of mathematical analysis*. [S.l.]: McGraw-hill New York, 1964. v. 3. Citado na página 25.
- 42 ROSENBLATT, M. Remarks on some nonparametric estimates of a density function. *The Annals of Mathematical Statistics*, JSTOR, p. 832–837, 1956. Citado na página 28.
- 43 PARZEN, E. On estimation of a probability density function and mode. *The annals of mathematical statistics*, JSTOR, v. 33, n. 3, p. 1065–1076, 1962. Citado na página 28.
- 44 KAISER, A.; SCHREIBER, T. Information transfer in continuous processes. *Physica D: Nonlinear Phenomena*, Elsevier, v. 166, n. 1-2, p. 43–62, 2002. Citado na página 39.
- 45 SYED, A.; LONE, N.; WANI, M. A.; BHAT, A. Clinical management of patients with minor head injuries. *International journal of health sciences*, Qassim University, v. 1, n. 1, p. 131, 2007. Citado na página 51.
- 46 GOLDBERGER, A. L.; AMARAL, L. A.; GLASS, L.; HAUSDORFF, J. M.; IVANOV, P. C.; MARK, R. G.; MIETUS, J. E.; MOODY, G. B.; PENG, C.-K.; STANLEY, H. E. Physiobank, physiotoolkit, and physionet. *Circulation*, Am Heart Assoc, v. 101, n. 23, p. e215–e220, 2000. Citado na página 51.
- 47 SILVA, I.; MOODY, G. B. An open-source toolbox for analysing and processing physionet databases in matlab and octave. *Journal of open research software*, NIH Public Access, v. 2, n. 1, 2014. Citado na página 51.
- 48 HELDT, T.; OEFINGER, M.; HOSHIYAMA, M.; MARK, R. Circulatory response to passive and active changes in posture. In: IEEE. *Computers in Cardiology, 2003*. [S.l.], 2003. p. 263–266. Citado na página 51.
- 49 KIM, N.; KRASNER, A.; KOSINSKI, C.; WININGER, M.; QADRI, M.; KAPPUS, Z.; DANISH, S.; CRAELIUS, W. Trending autoregulatory indices during treatment for traumatic brain injury. *Journal of clinical monitoring and computing*, Springer, v. 30, n. 6, p. 821–831, 2016. Citado na página 51.

APÊNDICE A

Estimadores de Medidas de Informação

A.1 Estimador Kozachenko-Leonenko

Considerando uma variável aleatória X com função densidade de probabilidade $f(x)$ e cujas distâncias entre quaisquer duas realizações de X podem ser calculadas por meio de alguma métrica, por exemplo, norma máxima ou norma Euclidiana, a entropia diferencial segundo Shannon é definida como:

$$h(X) = - \int f(x) \log f(x) dx \quad (\text{A.1})$$

em que \log é o logaritmo natural e expressa a entropia em nats. O estimador proposto por Kozachenko e Leonenko (1987) [11] para a equação A.1 objetiva determinar, para N amostras de uma variável aleatória, uma média de $\log f(x)$, isto é:

$$\widehat{h}(X) = -N^{-1} \sum_{i=1}^N \widehat{\log f(x_i)} \quad (\text{A.2})$$

Para estimar $\widehat{\log f(x_i)}$, os autores consideram que a distribuição de probabilidade da distância entre x_i e seu k -ésimo vizinho, $P_k(\epsilon)$ é igual a chance de ter:

1. 1 amostra com distância $r \in [\epsilon/2, \epsilon/2 + d\epsilon/2]$ até x_i ;
2. $k - 1$ outras amostras com distâncias menores que r e;
3. $N - k - 1$ amostras com distâncias maiores que r .

Portanto, seja p_i a massa da bola centrada em x_i , $p_i(\epsilon) = \int_{\|\xi - x_i\| < \epsilon/2} f(\xi) d\xi$, modela-se a seguinte distribuição trinomial:

$$P_k(\epsilon) d\epsilon = \frac{(N - 1)!}{1!(k - 1)!(N - k - 1)!} \times \frac{dp_i(\epsilon)}{d\epsilon} d\epsilon \times p_i(\epsilon)^{k-1} \times (1 - p_i(\epsilon))^{N-k-1}. \quad (\text{A.3})$$

O estimador do logaritmo da função densidade de probabilidade, $\widehat{\log f(x)}$, é obtido assumindo que $f(x)$ é constante no entorno da amostra x_i , assim:

$$p_i(\epsilon) \approx c_d \epsilon^d f(x_i), \quad (\text{A.4})$$

em que d é a dimensão da variável aleatória X e c_d é o volume da bola unitária d -dimensional (para norma Euclideana $c_d = \pi^{d/2}/\Gamma(1 + d/2)$).

Calculando o valor esperado do logaritmo em ambos os lados da equação A.4, têm-se:

$$E \log f(x_i) \approx \log p_i(\epsilon) - dE \log \epsilon - \log c_d, \quad (\text{A.5})$$

mas,

$$\begin{aligned} E(\log p_i(\epsilon)) &= \int_0^1 P_k(\epsilon) \log p_i(\epsilon) d\epsilon \\ &= k \binom{N-1}{k} \int_0^1 p_i(\epsilon)^{k-1} (1-p_i(\epsilon))^{N-k-1} \log p_i(\epsilon) dp_i(\epsilon) \\ &= \psi(k) - \psi(N), \end{aligned} \quad (\text{A.6})$$

em que p_i pode ser interpretada como uma variável aleatória beta-distribuída na qual, a função densidade de probabilidade é:

$$f(p_i; k, N-k) = \frac{p_i^{k-1} (1-p_i)^{N-k-1}}{\mathbf{B}(k, N-k)}. \quad (\text{A.7})$$

Portanto,

$$\widehat{h}(X) = -\psi(k) + \psi(N) + \log c_d + \frac{d}{N} \sum_{i=1}^N \log \epsilon(i) \quad (\text{A.8})$$

em que $\epsilon(i)$ é duas vezes a distância entre x_i e seu k -ésimo vizinho.

A.2 Estimador SVM de uma classe

O algoritmo SVM de uma classe foi proposto por Schölkopf et al. (2001) [16] para estimar o suporte de uma distribuição de grande dimensão.

Seja \mathbf{v} um vetor d -dimensional de realizações de variáveis aleatórias definido por uma matriz $1 \times d$, denominado vetor de treinamento,

$$\mathbf{v} = \begin{bmatrix} x_1 & x_2 & \cdots & x_d \end{bmatrix}, \quad (\text{A.9})$$

e, $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_N \in \mathcal{V}$ em que N é o número de observações de \mathbf{v} , \mathcal{V} é um subconjunto de \mathbb{R}^d e d é a dimensão do vetor de treinamento. Uma função de mapeamento denominada $\Psi(\cdot)$, mapeia esses vetores em um espaço característico de produto interno $\mathcal{V} \rightarrow F$ tal que a imagem é calculada por alguma função kernel [16],

$$k(\mathbf{v}_i, \mathbf{v}_j) = \langle \Psi(\mathbf{v}_i), \Psi(\mathbf{v}_j) \rangle, \quad (\text{A.10})$$

em que $\langle \cdot \rangle$ denomina o produto interno. Nesta dissertação foi utilizada a função kernel gaussiana (RBF, do inglês: *Radial Basis Function*) dada por:

$$k(\mathbf{v}_i, \mathbf{v}_j) = e^{-\gamma \|\mathbf{v}_i - \mathbf{v}_j\|^2}. \quad (\text{A.11})$$

Duan et al. (2015) [17] recorre a técnica de validação cruzada para determinar γ na equação (A.11). Entretanto, visto que trata-se de um algoritmo de aprendizado supervisionado, para aplicar a validação cruzada no SVM de uma classe é necessário pré-especificar uma taxa de acerto desejada e, em seguida, procurar o parâmetro cuja precisão de validação cruzada é semelhante.

Contudo, especificar essa taxa é equivalente à razão entre a quantidade de pontos que compõem o suporte das amostras e a quantidade total de pontos de uma região delimitada na qual o suporte está inserido, mas, essa configuração não é conhecida. Sendo assim, a partir de simulações optou-se por utilizar $\gamma = 0, 1$.

A ideia base do SVM de uma classe consiste em mapear o vetor de treinamento em um espaço característico com o objetivo de separar os vetores da origem com margem máxima. Para tanto, Schölkopf et al. (2001) [16] propõe o seguinte problema quadrático,

$$\begin{aligned} \min_{\alpha} \quad & \frac{1}{2} \sum_{ij} \alpha_i \alpha_j k(\mathbf{v}_i, \mathbf{v}_j) \\ \text{sujeito a} \quad & 0 \leq \alpha_i \leq \frac{1}{\nu N}, \quad \sum_i \alpha_i = 1, \end{aligned} \quad (\text{A.12})$$

em que α é um vetor com os coeficientes dos vetores de suporte e $\nu \in (0, 1)$ indica um limite superior na fração de pontos de treinamento fora da região (*outliers*).

Os parâmetros e soluções determinados a partir das equações (A.11) e (A.12) são denominados modelo do sistema. Após determinação do modelo, utiliza-se a função de decisão, $f(\mathbf{v})$, para verificar quantos pontos em uma região de interesse e quantizada fazem parte do conjunto de suporte do modelo,

$$f(\mathbf{v}) = \text{sgn} \left(\sum_i \alpha_i (\mathbf{v}_i, \mathbf{v}) - \rho \right), \quad (\text{A.13})$$

em que $\rho = \sum_j \alpha_j k(\mathbf{v}_i, \mathbf{v}_j)$ é a distância entre o hiperplano modelado até a origem e,

$$\text{sgn}(u) = \begin{cases} 1, & u \geq 0, \\ -1, & \text{caso contrário.} \end{cases} \quad (\text{A.14})$$

A região de interesse supracitada identifica a região máxima na qual o conjunto de suporte está definido. Esta região é delimitada por um limitante inferior (\mathbf{x}_{lb}) e o limitante superior (\mathbf{x}_{ub}) que são definidos com valores mínimo e máximo das amostras contidas no

Algoritmo 1: SVM de uma classe para estimação de entropia

Entrada: $X, nbins$ // X é uma matriz $N \times d$
Saída: \hat{h}
 1 $\gamma = 0,1$ // parâmetro do kernel
 2 $\nu = 0,01$ // fração de *outliers*
 3 $vbins = 1$
 4 **para** $i = 1$ até d **faça**
 5 Determine os vetores de suporte e o δ para $X(:, i)$
 // Equações (A.12) e (A.15) ou (A.16)
 // $X(:, i)$ é um vetor coluna
 6 $lb(i) = \min_{1:n} X(:, i) - \delta$
 7 $ub(i) = \max_{1:n} X(:, i) + \delta$
 8 $\Delta(i) = \frac{ub(i) - lb(i)}{nbins}$
 9 $vbins = vbins \times \Delta(i)$
 10 $\widehat{\text{supp}}(X) = [lb(1) : \Delta(1) : ub(1)] \times \cdots \times [lb(d) : \Delta(d) : ub(d)]$
 // delimita a região de busca de suporte
 11 Determine os vetores de suporte para X // Equação (A.12)
 12 $n = |f(\widehat{\text{supp}}(X)) == 1|$
 13 $\hat{h} = \log(n \times vbins)$
 14 **retorna** \hat{h}

vetor de treinamento compensados de um fator δ de acordo com as seguintes desigualdades

$$\sum_{i=1}^N \alpha_i e^{-\gamma \|\mathbf{v}_{lb} - \mathbf{v}_i\|^2} < \sum_{i=1}^N \alpha_i e^{-\gamma \|\mathbf{v}_{lb} - \mathbf{v}_{min}\|^2} = e^{-\gamma \delta^2} \quad (\text{A.15})$$

$$\sum_{i=1}^N \alpha_i e^{-\gamma \|\mathbf{v}_{ub} - \mathbf{v}_i\|^2} < \sum_{i=1}^N \alpha_i e^{-\gamma \|\mathbf{v}_{ub} - \mathbf{v}_{max}\|^2} = e^{-\gamma \delta^2}, \quad (\text{A.16})$$

em que $\mathbf{v}_{lb} = \mathbf{v}_{min} - \delta$, $\mathbf{v}_{ub} = \mathbf{v}_{max} + \delta$ e $e^{-\gamma \delta^2} = \rho$. Para delimitação dessa região no caso de variáveis multivariadas deve-se calcular ρ e δ considerando variáveis unidimensionais e então, determina-se o suporte máximo através do produto cartesiano das regiões delimitadas em cada dimensão.

Em seguida, o suporte máximo é uniformemente quantizado em $nbins$ intervalos de volume Δ . Então, a estimativa de entropia diferencial para uma variável aleatória é,

$$\hat{h}(X) = \log(n_x \times \Delta), \quad (\text{A.17})$$

em que n é o número de pontos classificados como pertencentes à única classe treinada. O Algoritmo 1 relaciona o pseudocódigo do estimador de entropia a partir do SVM de uma classe para variáveis aleatórias multivariadas.

APÊNDICE B

Modelo Tridimensional Autorregressivo

Considerando o modelo autorregressivo tridimensional do estudo realizado no capítulo 3 seção 3.4.1,

$$\begin{cases} X_{n+1} = \alpha X_n + \eta_n^X \\ Y_{n+1} = \beta X_n + \gamma Y_n + \eta_n^Y \\ Z_{n+1} = \delta Y_n + \epsilon Z_n + \eta_n^Z \end{cases}, \quad (\text{B.1})$$

em que $\eta_n^X, \eta_n^Y, \eta_n^Z \sim \mathcal{N}(0; 1)$, quando $0 < \alpha, \beta, \gamma, \delta, \epsilon < 1$ (sistema é estável quando os pólos estão definidos dentro do círculo de raio unitário). Fazendo $k = l = 1$ o vetor aleatório para as análises de entropia de transferência e entropia de transferência direta é $[X_n, Y_n, Z_n, X_{n+1}, Y_{n+1}, Z_{n+1}]$, cujas covariâncias são:

$$\begin{aligned} \sigma_x^2 &= \text{cov}(X_n, X_n) = \frac{1}{1-\alpha^2} & \sigma_{xy} &= \sigma_{yx} = \text{cov}(X_n, Y_n) = \frac{\alpha\beta\sigma_x^2}{1-\alpha\gamma} \\ \sigma_y^2 &= \text{cov}(Y_n, Y_n) = \frac{\beta^2\sigma_x^2 + 2\beta\gamma\sigma_{xy} + 1}{1-\gamma^2} & \sigma_{xz} &= \sigma_{zx} = \text{cov}(X_n, Z_n) = \frac{\alpha\delta\sigma_{xy}}{1-\alpha\epsilon} \\ \sigma_z^2 &= \text{cov}(Z_n, Z_n) = \frac{\delta^2\sigma_y^2 + 2\delta\epsilon\sigma_{yz} + 1}{1-\epsilon^2} & \sigma_{yz} &= \sigma_{zy} = \text{cov}(Y_n, Z_n) = \frac{\beta\delta\sigma_{xy} + \beta\epsilon\sigma_{xz} + \gamma\delta\sigma_y^2}{1-\gamma\epsilon} \\ \sigma_{xx^+} &= \text{cov}(X_n, X_{n+1}) = \alpha\sigma_x^2 & \sigma_{xz^+} &= \text{cov}(X_n, Z_{n+1}) = \delta\sigma_{xy} + \epsilon\sigma_{xz} \\ \sigma_{yy^+} &= \text{cov}(Y_n, Y_{n+1}) = \beta\sigma_{xy} + \gamma\sigma_y^2 & \sigma_{zx^+} &= \text{cov}(Z_n, X_{n+1}) = \alpha\sigma_{xz} \\ \sigma_{zz^+} &= \text{cov}(Z_n, Z_{n+1}) = \delta\sigma_{yz} + \epsilon\sigma_z^2 & \sigma_{yz^+} &= \text{cov}(Y_n, Z_{n+1}) = \delta\sigma_y^2 + \epsilon\sigma_{yz} \\ \sigma_{xy^+} &= \text{cov}(X_n, Y_{n+1}) = \beta\sigma_x^2 + \gamma\sigma_{xy} & \sigma_{zy^+} &= \text{cov}(Z_n, Y_{n+1}) = \beta\sigma_{xz} + \gamma\sigma_{yz} \\ \sigma_{yx^+} &= \text{cov}(Y_n, X_{n+1}) = \alpha\sigma_{xy} \end{aligned}$$

Na análise já realizada foi visto que nesse caso existe fluxo de informação direto de \mathbf{X} para \mathbf{Y} e de \mathbf{Y} para \mathbf{Z} , e o fluxo de informação de \mathbf{X} para \mathbf{Z} é indireto através da variável intermediária \mathbf{Y} . Portanto, têm-se as seguintes expressões analíticas para todas as relações de entropia de transferência e para a relação que é maior que zero de entropia de transferência direta:

$$\begin{aligned}
 T_{\mathbf{X} \rightarrow \mathbf{Y}} &= \frac{1}{2} \log \frac{|\text{cov}([X_n, Y_n])| |\text{cov}([Y_{n+1}, Y_n])|}{|\text{cov}([Y_{n+1}, X_n, Y_n])| |\text{cov}([Y_n])|} \\
 &= \frac{1}{2} \log \left[-\frac{-\alpha^4 \gamma^2 + 2\alpha^3 \gamma + \alpha^2 \gamma^2 - \alpha^2 - 2\alpha \beta^2 \gamma - 2\alpha \gamma + \beta^4 + 2\beta^2 + 1}{(\alpha \gamma - 1)(\gamma \alpha^3 - \alpha^2 + \gamma \alpha \beta^2 - \gamma \alpha + \beta^2 + 1)} \right], \\
 T_{\mathbf{Y} \rightarrow \mathbf{X}} &= \frac{1}{2} \log \frac{|\text{cov}([X_n, Y_n])| |\text{cov}([X_{n+1}, X_n])|}{|\text{cov}([X_{n+1}, X_n, Y_n])| |\text{cov}([X_n])|} \\
 &= \frac{1}{2} \log 1 = 0, \\
 T_{\mathbf{X} \rightarrow \mathbf{Z}} &= \frac{1}{2} \log \frac{|\text{cov}([X_n, Z_n])| |\text{cov}([Z_{n+1}, Z_n])|}{|\text{cov}([Z_{n+1}, X_n, Z_n])| |\text{cov}([Z_n])|} \\
 &= \frac{1}{2} \log \left[\frac{(-\alpha^2 \beta^2 \gamma^2 \delta^2 + \alpha^2 \beta^2 \delta^2 - \alpha^2 \gamma^4 + \alpha^2 \gamma^2 \delta^2 + \alpha^2 \gamma^2 + 2\alpha \gamma^3 - 2\alpha \gamma \delta^2 - 2\alpha \gamma + \beta^2 \delta^2 - \gamma^2 + \delta^2 + 1)(-\alpha^4 \gamma^4 + \alpha^4 \gamma^2 \delta^4 + 2\alpha^4 \gamma^2 \delta^2 + \alpha^4 \gamma^2 + 2\alpha^3 \gamma^3 - 2\alpha^3 \gamma \delta^4 + \alpha \beta^2 \gamma \delta^2 + \alpha \gamma^3 - \alpha \gamma \delta^2 - \alpha \gamma + \beta^2 \delta^2 - \gamma^2 + \delta^2 + 1)(-\alpha^2 \beta^2 \gamma^2 \delta^2 - 4\alpha^3 \gamma \delta^2 - 2\alpha^3 \gamma + 2\alpha^2 \beta^2 \gamma^2 \delta^2 + \alpha^2 \gamma^4 - \alpha^2 \gamma^2 \delta^4 - 2\alpha^2 \gamma^2 \delta^2 - 2\alpha^2 \gamma^2 + \alpha^2 \delta^4 + \alpha^2 \beta^2 \delta^4 + \alpha^2 \beta^2 \delta^2 - \alpha^2 \gamma^4 + \alpha^2 \gamma^2 \delta^4 + 2\alpha^2 \gamma^2 \delta^2 + \alpha^2 \gamma^2 - 2\alpha \beta^2 \gamma \delta^4 + 2\alpha^2 \delta^2 + \alpha^2 + 2\alpha \beta^2 \gamma \delta^4 - 2\alpha \gamma^3 + 2\alpha \gamma \delta^4 + 4\alpha \gamma \delta^2 + 2\alpha \gamma - \beta^4 \delta^4 - 2\beta^2 \delta^4 + 2\alpha \gamma^3 - 2\alpha \gamma \delta^4 - 4\alpha \gamma \delta^2 - 2\alpha \gamma + \beta^4 \delta^4 + 2\beta^2 \delta^4 + 2\beta^2 \delta^2 - 2\beta^2 \delta^2 + \gamma^2 - \delta^4 - 2\delta^2 - 1)}{(\alpha \gamma - 1)(-\alpha^3 \gamma^3 + \alpha^3 \gamma \delta^2 + \alpha^3 \gamma + \alpha^2 \gamma^2 - \alpha^2 \delta^2 - \alpha^2 - \gamma^2 + \delta^4 + 2\delta^2 + 1)} \right], \\
 T_{\mathbf{Z} \rightarrow \mathbf{X}} &= \frac{1}{2} \log \frac{|\text{cov}([X_n, Z_n])| |\text{cov}([X_{n+1}, X_n])|}{|\text{cov}([X_{n+1}, X_n, Z_n])| |\text{cov}([X_n])|} \\
 &= \frac{1}{2} \log 1 = 0, \\
 T_{\mathbf{Y} \rightarrow \mathbf{Z}} &= \frac{1}{2} \log \frac{|\text{cov}([Y_n, Z_n])| |\text{cov}([Z_{n+1}, Z_n])|}{|\text{cov}([Z_{n+1}, Y_n, Z_n])| |\text{cov}([Z_n])|} \\
 &= \frac{1}{2} \log \left[\frac{-\alpha^4 \gamma^4 + \alpha^4 \gamma^2 \delta^4 + 2\alpha^4 \gamma^2 \delta^2 + \alpha^4 \gamma^2 + 2\alpha^3 \gamma^3 - 2\alpha^3 \gamma \delta^4 - 4\alpha^3 \gamma \delta^2 - 2\alpha^3 \gamma + 2\alpha^2 \beta^2 \gamma^2 \delta^2 + \alpha^2 \gamma^4 - \alpha^2 \gamma^2 \delta^4 - 2\alpha^2 \gamma^2 \delta^2 - 2\alpha^2 \gamma^2 + \alpha^2 \delta^4 + 2\alpha^2 \delta^2 + \alpha^2 + 2\alpha \beta^2 \gamma \delta^4 - \alpha^2 \delta^2 - \alpha^2 + \alpha \beta^2 \gamma \delta^2 + \alpha \gamma^3 - \alpha \gamma \delta^2 - 2\alpha \gamma^3 + 2\alpha \gamma \delta^4 + 4\alpha \gamma \delta^2 + 2\alpha \gamma - \beta^4 \delta^4 - 2\beta^2 \delta^2 - 2\beta^2 \delta^2 + \gamma^2 - \delta^4 - 2\delta^2 - 1}{(\alpha \gamma - 1)(-\alpha^3 \gamma^3 + \alpha^3 \gamma \delta^2 + \alpha^3 \gamma + \alpha^2 \gamma^2 - \alpha^2 \delta^2 - \alpha^2 - \alpha \gamma + \beta^2 \delta^2 - \gamma^2 + \delta^2 + 1)} \right], \\
 T_{\mathbf{Z} \rightarrow \mathbf{Y}} &= \frac{1}{2} \log \frac{|\text{cov}([Y_n, Z_n])| |\text{cov}([Y_{n+1}, Y_n])|}{|\text{cov}([Y_{n+1}, Y_n, Z_n])| |\text{cov}([Y_n])|} \\
 &= \frac{1}{2} \log 1 = 0, \\
 D_{\mathbf{X} \rightarrow \mathbf{Z}} &= \frac{1}{2} \log \frac{|\text{cov}([X_n, Y_n, Z_n])| |\text{cov}([Z_{n+1}, Y_n, Z_n])|}{|\text{cov}([Z_{n+1}, X_n, Y_n, Z_n])| |\text{cov}([Y_n, Z_n])|} \\
 &= \frac{1}{2} \log 1 = 0.
 \end{aligned}$$