LUCAS MASSAROPPE

CARACTERIZAÇÃO DA CONECTIVIDADE ENTRE REGIÕES CEREBRAIS VIA ENTROPIA APROXIMADA E CAUSALIDADE DE GRANGER

Dissertação apresentada à Escola Politécnica da Universidade de São Paulo para obtenção do Título de Mestre em Engenharia Elétrica.

São Paulo 2011

LUCAS MASSAROPPE

CARACTERIZAÇÃO DA CONECTIVIDADE ENTRE REGIÕES CEREBRAIS VIA ENTROPIA APROXIMADA E CAUSALIDADE DE GRANGER

Dissertação apresentada à Escola Politécnica da Universidade de São Paulo para obtenção do Título de Mestre em Engenharia Elétrica.

Área de Concentração: Sistemas Eletrônicos

Orientador: Prof. Dr. Luiz Antonio Baccalá

São Paulo 2011

Este exemplar foi revisado e alterado em relação à versão original, sob responsabilidade única do autor e com a anuência de seu orientador.
São Paulo, 22 de Agosto de 2011.
Assinatura do autor
Assinatura do orientador

FICHA CATALOGRÁFICA

Massaroppe, Lucas

Caracterização da conectividade entre regiões cerebrais via entropia aproximada e causalidade de Granger/L. Massaroppe. — ed. rev. — São Paulo, 2011.

143 p.

Dissertação (Mestrado) — Escola Politécnica da Universidade de São Paulo. Departamento de Engenharia de Telecomunicações e Controle.

1. Causalidade 2. Entropia (Matemática aplicada) 3. Modelos não lineares 4. Análise de séries temporais 5. Inferência estatística 6. Teoria da informação I. Universidade de São Paulo. Escola Politécnica. Departamento de Engenharia de Telecomunicações e Controle. II. t.

Dedico este trabalho os meus pais, Maria Amélia Coló Massaroppe e José Antonio Massaroppe, pelo amor, apoio (de todas as formas), carinho que eles têm dedicado a mim, ao longo da minha formação e vida e por me ensinarem que o estudo é o bem mais valioso que se pode ter e dar.

AGRADECIMENTOS

Esta dissertação é o resultado de muitas leituras, dedicação, gosto por estudo, busca por conhecimento e pela pesquisa científica. Concluir essa etapa da minha carreira acadêmica, é uma grande vitória. Algumas pessoas são muito especiais nessa longa jornada:

Professor Luiz Antonio Baccalá, orientador deste trabalho, a quem devo respeito e gratidão. Nossas discussões, por vezes calorosas, foram muito úteis para meu crescimento técnico-científico e, também, humano. Devo muito a ele.

Laura, que desde o primeiro ano da graduação (aqui na EPUSP), acredita, incentiva, apóia e dedica enorme atenção a meu crescimento. Ela foi a primeira pessoa que me cativou com suas pesquisas e exemplo de seriedade, ética e amor à ciência, que todos os verdadeiros pesquisadores devem sempre ter. Ao passar dos anos, ensina-me a manter esse espírito tão digno e singular. Nossas discussões filosóficas são muito úteis para meu aperfeiçoamento científico e, principalmente, pessoal.

Silvia, a qual conheci no segundo ano de graduação e também sempre me estimulou a seguir em frente com minhas convicções.

Professora Lúcia Pereira Barroso, do Departamento de Estatística do Instituto de Matemática e Estatística da USP (MAE-IME-USP), especialista na área de *Estatística Multivariada*. Pelos ensinamentos em estatística, disponibilidade para me receber em sua sala e por responder aos meus *e-mails*.

Gilson Vieira (com formação pelo IME-USP) com quem tive longas e diversas discussões filosóficas, matemáticas e computacionais, nas quais ele sempre tinha razão, o que muito me ajudou a superar dificuldades no que se refere à realização deste trabalho.

Professores Birajara Soares Machado e João Ricardo Sato, pelas valiosas críticas e sugestões dadas ainda no período de qualificação e que, certamente, contribuíram para a elevação do nível da dissertação.

Professor Koichi Sameshima (Engenheiro Eletricista pela EPUSP e Médico pela FMUSP), que espero ter contribuído para seu trabalho de longa data, em assunto similar ao de minha dissertação, com quem meu orientador colabora diretamente há anos. A todos os colegas no Laboratório de Comunicações e Sinais (LCS) da Escola Politécnica da USP (EPUSP) que me apoiaram e ajudaram, de forma direta ou indireta, na realização deste trabalho.

Ligia Maria Prezia Lemos, revisora deste trabalho, pelo capricho, prontidão e pelas discussões sobre o como fazer um texto científico, que foram essenciais para mim.

Minha família, pelo apoio, imensa paciência e compreensão.

Por último, mas não menos importante, agradeço à **FAPESP** (Fundação de Amparo à Pesquisa do Estado de São Paulo), pelo apoio financeiro para a realização deste trabalho (*Processo FAPESP:* 2009/04397 - 7).

"…

Experimente coisas novas. Troque novamente. Mude, de novo. Experimente outra vez.

Você certamente conhecerá coisas melhores e coisas piores do que as já conhecidas, mas não é isso o que importa.

O mais importante é a mudança, o movimento, o dinamismo, a energia.

Só o que está morto não muda ! ..."

(Mudança — Clarice Lispector)

RESUMO

Essa dissertação apresenta o desenvolvimento de métodos para caracterização da conectividade entre séries temporais neurofisiológicas. Utilizam-se metodologias provenientes da Teoria da Informação — Entropias Aproximada e Amostral — para representar a complexidade da série no tempo, o que permite inferir como sua variabilidade se transfere a outras sequências, através do uso da coerência parcial direcionada. Para cada sistema analisado: (1) Faz-se uma transformação em outro, relacionando-o às medidas de entropia, (2) Estima-se a conectividade pela coerência parcial direcionada e (3) Avalia-se a robustez do procedimento via simulações de Monte Carlo e análise de sensibilidade. Para os exemplos simulados, a técnica proposta é capaz de oferecer resultados plausíveis, por meio da correta inferência da direção de conectividade em casos de acoplamento não-linear (quadrático), com número reduzido de amostras temporais dos sinais, em que outras abordagens falham. Embora de simples implementação, conclui-se que o processo mostra-se como uma extensão da causalidade de Granger para o caso não-linear.

Palavras-chave: Coerência parcial direcionada. Entropia aproximada. Entropia amostral. Modelos não lineares. Análise de séries temporais. Inferência estatística. Teoria da informação.

ABSTRACT

The purpose of this work is to present the development of methods for characterizing the connectivity between neurophysiological time series. Methodologies from Information Theory — Approximate and Sample Entropies — are used to represent the complexity of the series in a period of time, which allows inferring on how its variability is transferred to other sequences, using partial directed coherence. Methods: For each system under consideration, (1) It is done a transformation in another, relating it to measures of entropy, (2) The connectivity is estimated by the use of partial directed coherence and (3) The robustness of the procedure is analyzed via Monte Carlo simulations and sensitivity analysis. Results: For the simulated examples, the proposed technique is able to offer plausible results, through the correct inference of the connectivity direction, in cases of nonlinear coupling (quadratic), with a reduced number of signals samples, where other approaches fail. Conclusion: The process proves to be an extension of the Granger causality to the nonlinear case.

Keywords: Partial directed coherence. Approximate entropy. Sample entropy. Nonlinear models. Time series analysis. Statistical inference. Information theory.

LISTA DE SÍMBOLOS

Símbolos para matrizes e vetores

$\mathbf{A}(\cdot), \mathbf{B}(\cdot), \cdots$	Matriz
$a_{ij}(\cdot), b_{ij}(\cdot), \cdots$	Elemento de uma matriz na coluna i e linha j
\mathbf{I}_n	Matriz identidade de ordem n
$\mathbf{O}_{n imes m}$	Matriz nula de ordem $n \times m$
$\mathbf{x}(\cdot), \mathbf{y}(\cdot), \cdots$	Vetor
$\mathbf{\Sigma}_{[\cdot]}$	Matriz de variância-covariância do resíduo de predição

Símbolos para operadores

$[\cdot]^{1/2}$	Operador que denota a matriz triangular superior da Fa-
	toração de Cholesky
$[\cdot]^{\top}$	Operador matriz transposta
$[\cdot]^{\mathrm{H}}$	Operador matriz Hermitiana transposta
$[\cdot \circ \cdot]$	Operador produto de Hadamard
$\mathrm{COV}(\cdot, \cdot)$	Operador covariância
$\mathbb{E}[\cdot]$	Operador esperança
$\lim_{\{\cdot\}}\{\cdot\}$	Operador limite
$\min_{\{\cdot\}}\{\cdot\}$	Operador mínimo
$\mathbf{MSE}[\cdot]$	Operador erro quadrático médio (mean square error)
$VAR[\cdot]$	Operador variância
$\mathcal{O}(\cdot)$	Operador que denota a complexidade de um algoritmo

Símbolos para formas de convergência

- $\stackrel{d}{\rightarrow}$ Convergência em distribuição
- \xrightarrow{p} Convergência em probabilidade

Símbolos para funções

·	Função módulo
$\ \cdot\ _{\ell}$	Norma vetorial L_{ℓ}
$A_i^m(r) \in B_i^m(r)$	Contadores da "dimensão de correlação" do algoritmo da
	SampEn
$C_i^m(r)$	Dimensão de correlação (Correlation dimension)
$d[\cdot, \cdot]$	Função distância entre dois vetores
$H(\cdot)$	Entropia
$I(\cdot, \cdot)$	Informação mútua
$\det(\cdot)$	Determinante de uma matriz
g(x), g'(x)	A função $g(x)$ e sua primeira derivada $g'(x)$
$\ln(\cdot)$	Logaritmo na base e , ou logaritmo natural
$\log(\cdot)$	Logaritmo na base a
$\mathbb{P}(X = x_i)$	Probabilidade de um evento
$\pi(x)$	Função densidade de probabilidade marginal da variável
	aleatória x
$\bar{\pi}_{ij}(\cdot)$	(i, j)-ésimo coeficiente da coerência parcial direcionada
	(original)
$\iota \pi_{ij}(\cdot)$	(i, j)-ésimo coeficiente da coerência parcial direcionada in-
	formacional
$g\pi_{ij}(\cdot)$	(i, j)-ésimo coeficiente da coerência parcial direcionada ge-
	neralizada
$\mu(x,y)$	Função densidade de probabilidade conjunta das variáveis
	aleatórias $x \in y$
$\Theta(\cdot)$	Função degrau de Heaviside

Símbolos para distribuições

Variável aleatória, com distribuição Bernoulli e com pro-
babilidade de ocorrência p
Vetor aleatório Gaussiano, com vetor média $\mu_{1 \times N}$ e matriz
de covariância $\Sigma_{N \times N}$
Variável aleatória, com distribuição Gaussiana (Normal),
de média μ e variância σ^2
Variável aleatória, com distribuição Uniforme, no intervalo
(-a,a)

Símbolos para constantes

Probabilidade condicional (Conditional Probability)
Número de Euler
Frequência de amostragem
Comprimento do passo de deslocamento da janela desli-
zante
Unidade imaginária dos números complexos
Número de canais do processo autorregressivo
Dimensão do espaço de estados
Comprimento de uma série temporal
Ordem do processo autorregressivo ou significância, con-
forme o caso
Raio de tolerância
Erro padrão (Standard error)
Comprimento de uma janela deslizante
Quantil da Gaussiana padrão
Média do sinal $x(n)$
Desvio Padrão do sinal $x(n)$
Variância do sinal $x(n)$
Estimador de uma estatística
Erro tipo I
Erro tipo II

LISTA DE ABREVIAÇÕES E ACRÔNIMOS

Acrônimos para critérios de escolha de ordem do processo vetorial autorregressivo

AIC	Critério de informação de Akaike (Akaike Information Cri-
	terion)
BIC	Critério de informação Bayesiano de Schwartz (Schwartz's
	Bayesian Information Criterion)
FPE	Critério do erro de predição final (Final prediction error

- criterion)
- HQ Critério de Hannan-Quinn (Hannan-Quinn Criterion)

Acrônimos para termos estatísticos

ApEn	Entropia aproximada (Approximate Entropy)
DC	Coerência direcionada (Directed Coherence)
DTF	Directed Transfer Function
\mathcal{H}_0	Hipótese nula de teste estatístico
HQ	Critério de Hannan-Quinn (Hannan-Quinn Criterion)
IC	Intervalo de confiança
i.i.d.	Independente identicamente distribuído (Independent
	Identically Distributed)
PDC	Coerência parcial direcionada (<i>Partial Directed Coherence</i>)
pdf	Função densidade de probabilidade (Probability Density
	Function)
SampEn	Entropia amostral (Sample Entropy)
v.a.	Variável aleatória
FN	Falso negativo
FP	Falso positivo
VN	Verdadeiro negativo
VP	Verdadeiro positivo

Acrônimos para processos estocásticos

AR	Processo autorregressivo (Autoregressive Process)
GMP(p)	Processo Gaussiano de Markov (Gaussian Markov Pro-
	cess), com probabilidade de ocorrência p
MIX(p)	Processo de misturas (MIX), com probabilidade de
	ocorrencia p
$\operatorname{VAR}(p)$	Processo autorregressivo vetorial de ordem p (Vectorial Au-
	toregressive Process)
VARMA	Processo autorregressivo e de média móvel vetorial (Vecto-
	rial Autoregressive Moving Average Process)
VMA	Processo de média móvel vetorial (Vectorial Moving Ave- rage Process)
VECM	Modelo de correção de erros vetorial (Vector Error Correc-
	tion Model)
${x(n)}_{n\in\mathbb{Z}}$ ou ${x_n}_{n\in\mathbb{Z}}$	Processo estocástico

Acrônimos para termos matemáticos

- \mathcal{HI} Hipótese de indução finita
- \mathcal{PIF} Princípio de indução finita
- \mathcal{TI} Tese de indução finita

LISTA DE FIGURAS

1	Convergência assintótica de $\widehat{\operatorname{ApEn}}(m,r,N)$ para $\operatorname{ApEn}(m,r),$ no	
	caso de uma sequência <i>i.i.d.</i> uniforme entre $-\sqrt{3} e \sqrt{3}$	41
2	Evidência de normalidade assintótica da estatística $\widehat{\operatorname{ApEn}}(m,r,N),$	
	mostrada em (a) e (b), utilizando-se o processo $MIX(0,5), m = 1,$	
	r=0,18,N=1000.Foram utilizadas 1500 replicações de Monte	
	Carlo.	44
3	Evidência de normalidade assintótica da estatística $\widehat{\operatorname{ApEn}}(m, r, N)$,	
	mostrada em (a) e (b), utilizando-se o processo $GMP(0,2), m = 1,$	
	r = 0,18, N = 1000. For am utilizadas 200 replicações de Monte	
	Carlo.	45
4	Convergência assintótica da estatística Samp En $(m,0,18,N),$ para	
	seu valor teórico, no caso de um processo $i.i.d.$ uniforme no inter-	
	valo $(-\sqrt{3},\sqrt{3})$. Vê-se claramente que o viés é menor, do que no	
	caso da $\widehat{\operatorname{ApEn}}(m, 0, 18, N)$	47
5	Evidência de normalidade assintótica da estatística $\widehat{\text{SampEn}(m,r,N)}$,	
	mostradas em (a) e (b), utilizando-se o processo $MIX(0,5), m = 1,$	
	$r=0,\!18,N=1000$ e realizando 1500 replicações de Monte Carlo.	49
6	Evidência de normalidade assintótica da estatística $\widehat{\text{SampEn}(m,r,N)}$,	
	mostradas em (a) e (b), utilizando-se o processo $\text{GMP}(0,2), m =$	
	1, $r = 0,18$, $N = 1000$ e realizando 200 replicações de Monte Carlo.	50

7 Ilustração do método de geração das "séries informacionais". 53

8	Séries do Processo Autorregressivo Vetorial de ordem um (VAR(1)), da Equação 4.3.1.	57
9	PDC, típica, na forma padrão (ver [1]), entre séries do VAR(1). $% \left(\left(1, \left(1, \left($	58
10	Séries de entropia aproximada, calculadas com os dados da Fi- gura 8, usando $W = 150, h = 1, m = 1 e r = 0,15. \dots$	59
11	PDC, típica, na forma padrão (ver [1]), entre séries de ApEn, da Figura 10. Para se escolher a ordem, utilizou-se o AIC, que gerou um VAR(2)	60
12	Séries do Mapa Logístico Bivariado, das Equações 4.3.2	62
13	PDC, típica, na forma padrão (ver [1]), entre as séries da Figura 12. Para se escolher a ordem, utilizou-se o AIC, que gerou um VAR(6).	63
14	Séries de entropia aproximada, calculadas com os dados da Fi- gura 12, usando $W = 500, h = 1, m = 1 e r = 0,18. \dots$	64
15	PDC, típica, na forma padrão (ver [1]), entre séries de ApEn, da Figura 14. Para se escolher a ordem, utilizou-se o AIC, que gerou um VAR(4)	65
16	Séries de entropia amostral, calculadas com os dados da Figura 12, usando $W = 500, h = 1, m = 1 e r = 0,18.$	66
17	PDC, típica, na forma padrão (ver [1]), entre séries de SampEn, da Figura 16. Para se escolher a ordem, utilizou-se o AIC, que gerou um VAR(3)	67
18	Séries do Mapa de Acoplamento Quadrático Bivariado, calculadas usando as Equações 4.3.3 e os parâmetros $\beta=0,05$ e $R=0,99.$	70
19	PDC, típica, na forma padrão (ver [1]), entre as séries da Figura 18. Para se escolher a ordem, utilizou-se o AIC, que gerou um VAR(6).	71

20	Séries de entropia aproximada, calculadas com os dados da Fi- gura 18, usando $W = 150, h = 1, m = 1$ e $r = 0,18.$	72
21	PDC, típica, na forma padrão (ver [1]), entre séries de ApEn, da Figura 20. Para se escolher a ordem, utilizou-se o AIC, que gerou	
	um VAR(8)	73
22	Séries de entropia amostral, calculadas com os dados da Figura 18, usando $W = 150, h = 1, m = 1 e r = 0.18.$	74
23	PDC, típica, na forma padrão (ver [1]), entre séries de Samp En, da	
	Figura 22. Para se escolher a ordem, utilizou-se o AIC, que gerou um VAR(6)	75
24	Séries do Mapa de Biquadrático (Sistema 3D), calculadas usando as Equações 4.3.3 e o parâmetro $R = 0.99$	78
25	PDC, típica, na forma padrão (ver [1]), entre as séries da Figura 24. Para se escolher a ordem, utilizou-se o AIC, que gerou um VAR(16).	79
26	Séries da entropia aproximada, calculadas com os dados da Fi- gura 24, usando $W = 200, h = 1, m = 1 \text{ e } r = 0,15. \dots$	80
27	PDC, típica, na forma padrão (ver [1]), entre séries de ApEn, da Figura 26. Para se escolher a ordem, utilizou-se o AIC, que gerou	
	um VAR(6)	81
28	Séries de entropia amostral, calculadas com os dados da Figura 24, usando $W = 200, h = 1, m = 1 e r = 0.15.$	82
29	PDC, típica, na forma padrão (ver [1]), entre séries de SampEn, da Figura 28. Para se escolher a ordem, utilizou-se o AIC, que gerou	
	um VAR(13). \ldots	83

LISTA DE TABELAS

1	Casos simulados do sistema 2D	86
2	Exemplo de tabela padrão para contagem de VP, FN, FP e VN. $% \left({{{\rm{N}}} \right)$.	87
3	Porcentagens de VP e FP, referentes ao algoritmo da ApEn, para os casos da Tabela 1, do sistema 2D.	87
4	Porcentagens de VP e FP, referentes ao algoritmo da SampEn, para os casos da Tabela 1, do sistema 2D.	87
5	Tabela de contingência para $R = 0.99$, $h = 1$ e 10000 realizações, comparando-se os algoritmos da ApEn e SampEn, simulado com $W = 150, m = 2, r = 0.10. \dots \dots \dots \dots \dots \dots \dots \dots \dots \dots$	89
6	Tabela de contingência para $R = 0.99$, $h = 1$ e 10000 realizações, comparando-se os algoritmos da ApEn e SampEn, simulado com $W = 150, m = 1, r = 0.15. \dots$	89
7	Tabela de contingência para $R = 0.99$, $h = 1$ e 10000 realizações, comparando-se os algoritmos da ApEn e SampEn, simulado com $W = 200, m = 2, r = 0.10. \dots \dots \dots \dots \dots \dots \dots \dots \dots \dots$	89
8	Tabela de contingência para $R = 0,99, h = 1$ e 10000 realizações, comparando-se os algoritmos da ApEn e SampEn, simulado com W = 200, m = 1, r = 0,10.	89
9	Casos simulados do mapa de acoplamento biquadrático (Sistema 3D).	90

10	Porcentagens de VP e FP, referentes ao algoritmo da ApEn, para os	
	casos da Tabela 9 do mapa de acoplamento biquadrático (Sistema	
	3D)	91
11	Porcentagens de VP e FP, referentes ao algoritmo da SampEn,	
	para os casos da Tabela 9 do mapa de acoplamento biquadrático	
	(Sistema 3D)	92
12	Tabelas de contingência para $R = 0,99, h = 1$ e 10000 realizações,	
	comparando-se os algoritmos da ApEn e SampEn, simulado com	
	W = 200, m = 1, r = 0.15.	92
13	Tabelas de contingência para $R = 0,99, h = 1$ e 10000 realizações,	
	comparando-se os algoritmos da ApEn e SampEn, simulado com	
	W = 200, m = 2, r = 0.15.	93
14	Equações que fazem parte do algoritmo de Levinson multivariado	
	e condições iniciais do mesmo	119

SUMÁRIO

1	Intr	Introdução		
	1.1	Organ	ização da dissertação	22
	1.2	Contri	buições	23
2	Rev	v <mark>isão</mark> bi	ibliográfica	25
	2.1	Inferêr	ncia de conectividade por meio de causalidade	25
	2.2	Inform	nação e sua inferência	27
		2.2.1	Entropia aproximada	32
		2.2.2	Entropia amostral	34
	2.3	Sumár	io	36
3	Car	acteriz	zação dos estimadores de entropia	38
3	Car 3.1	eacteriz	z <mark>ação dos estimadores de entropia</mark> pia aproximada — ApEn	38 38
3	Car 3.1	eacteriz Entrop 3.1.1	cação dos estimadores de entropia pia aproximada — ApEn	38 38 38
3	Car 3.1	Entrop 3.1.1 3.1.2	cação dos estimadores de entropia pia aproximada — ApEn	 38 38 38 40
3	Car 3.1	Entrop 3.1.1 3.1.2 3.1.3	zação dos estimadores de entropia pia aproximada — ApEn	 38 38 38 40 41
3	Car 3.1	Entrop 3.1.1 3.1.2 3.1.3 3.1.4	cação dos estimadores de entropia pia aproximada — ApEn	 38 38 38 40 41 42
3	Car 3.1	Entrop 3.1.1 3.1.2 3.1.3 3.1.4 3.1.5	$xação$ dos estimadores de entropia $Dia a proximada - ApEn \dots \dots$	 38 38 40 41 42 43
3	Car 3.1 3.2	Entrop 3.1.1 3.1.2 3.1.3 3.1.4 3.1.5 Entrop	zação dos estimadores de entropia pia aproximada — ApEn	 38 38 38 40 41 42 43 46

		3.2.2	Viés da SampEn	46
		3.2.3	Comportamento assintótico	48
		3.2.4	Escolha dos parâmetros $m \in r$	48
	3.3	Sumái	rio	51
4	Des	critore	es informacionais de inferência causal	52
	4.1	Métod	lo	52
		4.1.1	Mapeamento	52
		4.1.2	Análise linear das séries mapeadas	54
	4.2	Conve	enções para se apresentar a PDC	54
	4.3	Exem	plos ilustrativos do uso da ApEn e da SampEn na deter-	
		minaç	ão de causalidade	55
		4.3.1	Modelo autorregressivo vetorial (VAR)	56
		4.3.2	Mapa logístico bivariado	61
		4.3.3	Mapa de acoplamento quadrático bivariado	68
		4.3.4	Mapa de acoplamento biquadrático (3D)	76
	4.4	Sumái	rio	84
5	Val	idação	estatística do método proposto	85
	5.1	Sistema 2D — Mapa de acoplamento quadrático (sistema de Equaçõe		1
		4.3.3)		85
		5.1.1	Sobre as PDCs entre as séries de ApEn e entre as séries	
			de SampEn, do mapa de acoplamento quadrático bivariado	
			das Equações 4.3.3	88

		5.1.2	Tabelas de contingência	88	
	5.2	Sistema 3D — Mapa de acoplamento biquadrático (sistema de			
		Equações 4.3.4–4.3.5)			
		5.2.1	Sobre as PDCs entre as séries de ApEn e entre as séries de		
		:	SampEn, do mapa de acoplamento biquadrático (Sistema		
		:	3D), das Equações 4.3.4–4.3.5	92	
		5.2.2	Tabelas de contingência	92	
	5.3	Discuss	ão dos resultados	93	
	5.4	Sumário	0	94	
6	Cor	nclusões	e trabalhos futuros	96	
	6.1	Trabalh	nos futuros	97	
R	e fer ê	ncias		98	
Ro	eferê nexo	ncias A – Lis	stagens dos códigos implementados em MATLAB	98 105	
	eferê nexo A.1	ncias A – Lis Entropi	stagens dos códigos implementados em MATLAB a aproximada: apen.m	98 105 105	
	eferê nexo A.1 A.2	ncias A – Lis Entropi Entropi	stagens dos códigos implementados em MATLAB a aproximada: apen.m	 98 105 105 106 	
Ro	eferê nexo A.1 A.2 A.3	ncias A – Lis Entropi Entropi Process	stagens dos códigos implementados em MATLAB a aproximada: apen.m a amostral: sampen.m o MIX: mix.m	 98 105 105 106 107 	
R	eferê nexo A.1 A.2 A.3 A.4	ncias A – Lis Entropi Entropi Process Process	stagens dos códigos implementados em MATLAB ia aproximada: apen.m ia amostral: sampen.m ia amostral: sampen.m io MIX: mix.m io Gaussian Markov Process (GMP): gmp.m	 98 105 105 106 107 108 	
R	eferê nexo A.1 A.2 A.3 A.4 A.5	ncias A – Lis Entropi Entropi Process Process Janelan	stagens dos códigos implementados em MATLAB ia aproximada: apen.m ia amostral: sampen.m io MIX: mix.m io Gaussian Markov Process (GMP): gmp.m inento de séries temporais: tswin.m	 98 105 105 106 107 108 109 	
R. As	eferê nexo A.1 A.2 A.3 A.4 A.5 nexo	ncias A – Lis Entropi Entropi Process Janelan B – O a	stagens dos códigos implementados em MATLAB ia aproximada: apen.m ia amostral: sampen.m ia amostral: sampen.m io MIX: mix.m io Gaussian Markov Process (GMP): gmp.m inento de séries temporais: tswin.m ialgoritmo de Levinson multivariado	 98 105 105 106 107 108 109 111 	
R. As	eferê nexo A.1 A.2 A.3 A.4 A.5 nexo B.1	ncias A – Lis Entropi Entropi Process Janelan B – O a Process	stagens dos códigos implementados em MATLAB ia aproximada: apen.m ia amostral: sampen.m ia amostral: sampen.m io MIX: mix.m io Gaussian Markov Process (GMP): gmp.m in ento de séries temporais: tswin.m ialgoritmo de Levinson multivariado ios autorregressivos vetoriais — VAR	 98 105 105 106 107 108 109 111 111 	

	B.2.1	Predição progressiva	112
	B.2.2	Predição regressiva	114
	B.2.3	Equações de Yule-Walker dos filtros de predição progressiva	
		e regressiva	115
	B.2.4	Algoritmo (generalizado) de Levinson multivariado	116
		B.2.4.1 Apresentação das equações	116
	B.2.5	O algoritmo	119
	B.2.6	A estrutura em treliça	119
B.3	Critéri	ios de seleção da ordem do VAR	120
Anexo	C – C	oerência parcial direcionada	122
C.1	Medin	do a PDC	122
	C.1.1	A causalidade de Granger	122
	C.1.2	A coerência parcial direcionada — PDC	124
		C.1.2.1 Sobre a causalidade de Granger instantânea	126
C.2	As var	iantes da PDC	126
	C.2.1	PDC (original) — $\bar{\pi}_{ij}(f)$	126
	C.2.2	PDC generalizada — $g\pi_{ij}(f)$	127
	C.2.3	PDC informational $- \iota \pi_{ij}(f)$	127
Anexo	D – D	emonstrações dos Teoremas do Capítulo 3	128
D.1	Demoi	nstração do Teorema 3.1.1	128
D.2	Demoi	nstração do Teorema 3.1.2	129

Anexo I	E – Sobre o intervalo de confiança da SampEn	131
E.1	Intervalo de confiança — IC \ldots	131
E.2	Estimativa da variância da SampEn	133
Anexo I	F – Publicações	135

1 INTRODUÇÃO

A maioria das pesquisas para compreender o funcionamento do cérebro são voltadas para localizar quais regiões estão ativas sob comportamentos, isto é, durante à realização de processos cognitivos [2, 3]. Esse contexto, comum em protocolos de neuroimagem funcional [4], por ser insuficiente para compreender o funcionamento cerebral, vem sendo substituido pelo estudo da *Conectividade* cerebral. Assim trata-se de ir além de apenas listar as regiões ativas e se estudar como elas interagem e de que forma suas relações se modificam dinamicamente.

Muitas abordagens têm sido propostas, todas baseadas na descrição dos sinais neurofisiológicos enquanto séries temporais, *i.e.*, medidas sequenciais no tempo ligadas a grandezas mensuráveis, empregadas em várias áreas do conhecimento [5–9]. São técnicas tanto para analisar séries isoladamente como para relacionar o comportamento temporal de grandezas distintas.

Grande parte dessas técnicas é voltada à descrição das relações entre séries, quando esse relacionamento ocorre de forma linear, como, por exemplo, a análise de correlação/coerência [10], equações estruturais [11] e coerência e coerência parcial direcionada [1], além de outras como, causalidade de Granger [12] e Informação mútua [13,14] que, em princípio, admitem ou requerem formulação mais geral e que possuem, todas, o uso penalizado por dificuldades de estimação quando os processos físico/biológicos responsáveis por sua geração são não-lineares [15, 16]. Encontra-se essa restrição, particularmente, quando se refere a métodos originários na idéia de *causalidade de Granger*, uma vez que o comportamento de preditores não-lineares, face a trechos curtos de dados, é frequentemente pouco confiável.

A idéia de *causalidade de Granger*, propriedade que uma série x_n possui sobre outra y_n , quando o passado da primeira pode ser útil à predição da segunda [1,12,17], pode servir de base para uma visão generalizada do problema conforme se argumenta em [18].

A questão que cabe neste contexto é: como utilizar estimadores robustos como os lineares e, ainda assim, capturar aspectos não-lineares do comportamento dinâmico?

A proposta central dessa investigação é usar descritores locais da informação associados às séries temporais e relacioná-los de modo linear.

Entre esses descritores encontram-se a chamada entropia aproximada sugerida por Pincus [19,20] e a amostral sugerida por Lake *et.al.* [21], que consistem em descrever o nível de complexidade/variabilidade de uma dada série no tempo. A questão é *se* e *como* essas formas de complexidade se propagam entre séries e se esta propagação pode ser capturada por meio dos estimadores disponíveis para causalidade.

Em grande medida, a escolha das entropias aproximada e amostral se prende ao fato de que essas têm sido sugeridas como meio de também detectar sincronia [22] e realimentação [23] e que, além do interesse fisiológico intrínseco, tais abordagens têm a mesma motivação que os estudos de causalidade mencionados [1, 12, 17, 18].

A partir do apresentado, o objetivo desse texto é estudar e desenvolver novas ferramentas estatísticas para inferir e classificar comportamentos dinâmicos de sistemas biológicos, por meio de Séries Temporais, com ênfase em modelos nãolineares, tendo como motivação que aquelas podem ser vistas como registros de medidas sequenciais no tempo ligadas a grandezas mensuráveis [5–9,24].

Há inúmeras técnicas para analisar séries isoladamente e relacionar o comportamento temporal de grandezas distintas. A maioria é voltada à descrição das relações entre séries quando essas se dão de forma linear. Contudo, encontramse dificuldades de estimativa quando processos físico/biológicos responsáveis por sua geração são não-lineares [15, 16]. Tais restrições são quando relacionadas a métodos ligados à ideia de causalidade de Granger [1, 12, 17].

Nesse trabalho apresenta-se a possibilidade de modelar séries segundo os estimadores de entropia: (1.) Aproximada, introduzida por Pincus, em [25] e (2.) Amostral, introduzida por Lake *et.al.*, em [21]. Tais métodos consistem na descrição do nível de complexidade e/ou variabilidade da série no tempo e têm a vantagem de serem não-paramétricos para, dessa forma, possibilitarem a inferência da direcionalidade do fluxo de informação entre séries temporais multivariadas, ao contrário do que é feito em [1], que solicita modelo paramétrico. Assim, pretendese verificar como tal complexidade se transfere para outras séries [15, 16, 26–31].

A título de ilustração, notam-se as técnicas anteriormente citadas que podem ser utilizadas em exames de eletroencefalograma (EEG), em que a resolução temporal é melhor do que a de imageamento por ressonância magnética funcional (fMRI). A partir dos sinais elétricos cerebrais, gerados pelos diferentes processos biológicos, físicos e químicos, pode-se fazer uma análise para inferir, mais detalhadamente, a direção do fluxo de informação entre as diferentes estruturas, com algoritmos que trazem uma medida direta de informação de tais processos.

A metodologia desse trabalho se dá, inicialmente, pelo estudo de técnicas para estimar dinamicamente a entropia de séries temporais e segue com a investigação do ajuste de modelos apropriados à explicitação da causalidade entre entropias no tempo. Para validar a metodologia, utilizam-se sistemas "toys", como mostrado no Capítulo 3, inferindo seu comportamento dinâmico, com o auxílio dos métodos para evidenciar a causalidade entre entropias.

A abordagem utilizada neste trabalho de mestrado é fundamentalmente numérica e experimental. As conclusões, em sua maioria, são baseadas na literatura e em simulações computacionais, com o auxílio do ambiente MATLAB.

1.1 Organização da dissertação

• Capítulo 2 — Revisão bibliográfica

Nesse capítulo inicia-se a descrição da pesquisa propriamente dita, com o aprofundamento da pesquisa bibliográfica realizada, focando trabalhos referentes às quantias conhecidas, respectivamente, como entropias aproximada e amostral.

• Capítulo 3 — Caracterização dos estimadores de entropia

No decorrer da pesquisa, houve a necessidade de se reproduzir certos resultados constantes na literatura. Essa descrição é realizada no Capítulo 3 e permitiu depurar métodos de cálculo das medidas referidas, cujas listagens em MATLAB estão contidas no Capítulo A.

• Capítulo 4 — Descritores informacionais de inferência causal

De posse dos resultados obtidos, passou-se, no Capítulo 4, à escolha de sistemas "toys" acoplados — modelos logístico e não-linear bivariados e um sistema 3D — que pudessem ter sua conectividade aferida via séries informacionais, com o auxílio do uso conjunto do cálculo de entropias associadas à coerências. A escolha desses modelos ocorreu devido à sua capacidade de emular um meio biológico e, também, por possuírem realimentação quadrática e a PDC entre as séries originais ser difícil de ser aferida.

• Capítulo 5 — Desempenho estatístico do método proposto

Neste capítulo é discutida a robustez do método proposto para avaliar a causalidade entre séries temporais não-lineares e, para tanto, são feitas simulações de Monte Carlo e uma pós análise de sensibilidade, expressa por meio de tabelas de contigência, a fim de se comparar o uso dos algoritmos da ApEn e SampEn.

• Capítulo 6 — Conclusões e trabalhos futuros

O texto finaliza por descrever as conclusões do trabalho e também oferece algumas perspectivas para sua continuação.

Seguem-se, então, os apêndices para referência.

1.2 Contribuições

As principais contribuições foram: (1.) o método de mapeamento por entropia deslizante e (2.) a extensão da causalidade de Granger para o caso não-linear.

Para capturar interações não-lineares, em que aproximações lineares falham, propõe-se abordagem híbrida (**mapeamento por** *entropia deslizante*), com o objetivo de calcular uma medida de complexidade dependente do tempo para série temporal. Produz-se assim outra sequência que retrata a evolução no tempo da complexidade da série original. A seguir, o método compara resultados desses mapeamentos entre séries por meio de técnicas lineares multivariadas.

No contexto de séries multivariadas, não-lineares, com um tom modulado (em uma frequência de ressonância bem definida) e realimentações quadráticas, o método proposto de mapeamento mostrou ser uma alternativa para realizar teste de causalidade de Granger, no dominío da frequência. Através de simulações de Monte Carlo em conjunto com análise de sensibilidade, foram geradas tabelas de contingência, para verificar a robustez do método, com resultados publicados em [32, 33].

2 REVISÃO BIBLIOGRÁFICA

A presente revisão consta de:

- Breve recapitulação de conceitos ligados à noção de conectividade inferida, a partir da noção de causalidade (Seção 2.1);
- 2. Discussão da noção de informação, seu contexto e inferência (Seção 2.2).

O capítulo termina com algumas considerações sobre as razões que levaram à escolha dos estimadores aqui inclusos.

2.1 Inferência de conectividade por meio de causalidade

A fim de caracterizar relações que permeiam a evolução temporal de grandezas econômicas, Granger [34] propôs examinar a melhoria sobre o conhecimento do passado que uma série temporal, x(n), pode ter sobre a previsibilidade de outra série, y(n). Tal proposta ficou conhecida como Causalidade de Granger para a distinguir de conceitos abstratos, usados em filosofia. Simbolicamente:

$$x(n) \xrightarrow[Causa]{Granger} y(n).$$

Um aspecto fundamental da causalidade de $Granger^1$ é a ausência de recipro-

¹Uma breve revisão sobre a *Causalidade de Granger* consta no Capítulo C.

cidade, *i.e.*:

$$x(n) \xrightarrow[Causa]{Granger} y(n),$$

não implica:

$$y(n) \xrightarrow[Causa]{Granger} x(n).$$

Essa ausência de reciprocidade distingue a causalidade de Granger de outras formas de descrição. Para esclarecer esse ponto, vale lembrar que um exemplo clássico de grandeza recíproca é o coeficiente de correlação, que é simétrico à permutação entre $x(n) \in y(n)$.

Na prática, há diversas formas de operacionalizar a inferência da causalidade, a maioria das quais envolve o ajuste de modelos de predição lineares [12], usando, por exemplo, modelos autorregressivos multivariados (ou vetoriais):

$$\mathbf{x}(n) = \sum_{r} \mathbf{A}(r)\mathbf{x}(n-r) + \mathbf{w}(n), \qquad (2.1.1)$$

em que, $\mathbf{x}(n)$ é um vetor K-dimensional de séries temporais e $\mathbf{w}(n)$ um processo de inovação, tal que:

$$\mathbf{A}(r) = \begin{bmatrix} a_{11}(r) & a_{12}(r) & \cdots & a_{1K}(r) \\ a_{21}(r) & a_{22}(r) & \cdots & a_{2K}(r) \\ \vdots & \vdots & \ddots & \vdots \\ a_{K1}(r) & a_{K2}(r) & \cdots & a_{KK}(r) \end{bmatrix},$$

tem os coeficientes $a_{ij}(r)$, responsáveis por medir o efeito da interação linear de $x_j(n-r)$ em $x_i(n)$, de sorte que a forma mais comum de aferir a existência de causalidade da série $x_j(n)$, para a série $x_i(n)$, consiste em examinar a nulidade desses coeficientes para todo r.

Testes estatísticos específicos podem ser apreciados em [12], para validar a Causalidade de Granger.

O exame recente e mais detalhado dessas questões, permitiu representar

relações de causalidade no domínio da frequência, dando lugar a grandezas, como as chamadas coerência direcionada [35], a DTF — directed transfer function [36] e coerência parcial direcionada [1,35] (PDC — partial directed coherence) que representam a Equação 2.1.1 no domínio da frequência.

Elas são complementares e se confundem quando K = 2. Vale, contudo, dizer que, quando K > 2, a PDC expõe as conexões diretas entre séries, enquanto a DTF exibe conexões entre as mesmas, ainda que indiretas [17]. Cabe, também, observar que os testes estatísticos descritos em [12] têm mapeamento direto sobre os relativos à PDC [37, 38].

O presente estudo envolve outros descritores de conectividade entre séries temporais e assume interesse adicional, em vista de ter demonstrado a relação que DTF e PDC possuem com a Teoria da Informação [18].

Os cálculos realizados no contexto do presente estudo têm por base o pacote desenvolvido pelo grupo de pesquisa "Processamento Digital de Sinais — LCS/LPS/EPUSP", liderado pelos professores Luiz Antonio Baccalá (LCS/E-PUSP) e Miguel Arjona Ramírez (LPS/EPUSP) e que está disponível em:

http://www.lcs.poli.usp.br/~baccala/pdc.

2.2 Informação e sua inferência

Conforme discutido na Seção 2.1, métodos que envolvem direta ou indiretamente somente descrição de segunda ordem de processos estocásticos são capazes de representar fielmente apenas aspectos relativos ao acoplamento linear entre séries temporais observadas.

Dessa limitação, surge a necessidade de se buscar outras abordagens. Como intuitivamente o acoplamento de grandezas dinâmicas têm como alicerce a ideia de que interações são essencialmente processos de troca de informação, segue naturalmente a necessidade de buscar elementos para generalizar tal descrição pelo uso da Teoria da Informação, cuja origem oficial é o trabalho de Shannon [39] e que tem profundos laços com Teoria de Probabilidades [13].

Ao invés de prover uma descrição rigorosa, como a de [14], parece mais interessante discorrer brevemente sobre alguns aspectos gerais, ainda que qualitativos.

A idéia central da teoria é associar a eventos aleatórios algum conceito de posse de "conhecimento" — atenção aqui dada às aspas. Embora a definição varie um pouco conforme o tipo do objeto aleatório envolvido O (evento, variável aleatória discreta ou continua ou processo)², ao valor E obtido num experimento envolvendo O, associa-se a informação $I(E) = -\log \mathbb{P}\{E\}$ no sentido de que quanto maior sua probabilidade, $\mathbb{P}\{E\}$, menor a surpresa na obtenção deste valor e, portanto, maior a informação disponível sobre sua obtenção.

Mais importante do que tipificar um dado valor E, é representar as características de O como um todo, algo que se faz com auxílio da noção de *entropia* que representa o valor médio da informação associada às probabilidades dos objetos de O. Nessas condições a entropia corresponde à esperança:

$$H(O) = \mathbb{E}_{\mathbb{P}}[O], \qquad (2.2.1)$$

definida sobre a medida de probabilidade \mathbb{P} associada aos elementos que compõem O.

Um resultado bem conhecido é de que a entropia de uma variável aleatória discreta e com distribuição uniforme é máxima, o que corresponde ao caso em que há uma forma de simetria relativa aos valores que podem ser obtidos num experimento aleatório, garantindo uma condição de máxima incerteza. Probabi-

²Aspectos como a eventual nulidade da probabilidade ou a natureza continua dos valores de O são os repensáveis por se ter que adotar definições ligeiramente diferentes para a informação em cada caso.

lidades que se repartem desigualmente implicam um vínculo de regularidade, no qual há menos incerteza. Esse cenário permite que se atribua à entropia o papel de medir o grau de incerteza associada a objetos aleatórios.

Como no problema de descrever o processo de comunicação estudado por Shannon, é natural considerar relações entre objetos aleatórios, como, por exemplo mensagem transmitida e recebida. Em teoria de probabilidades, a forma mais comum de enfrentar esse problema é por meio de probabilidades conjuntas, condicionais e marginais de que seguem imediatamente as definições respectivas, para entropias conjunta e condicional.

Neste caso, vale citar os seguintes resultados importantes:

1. Simetria da entropia conjunta $(O = O_1 \cup O_2)$:

$$H(O_1, O_2) = H(O_2, O_1) = H(O);$$

 Entropia condicional de O₂, dados os valores de O₁, corresponde à esperança da incerteza nos valores e O₂, conforme seja conhecido como os valores de O₁ se distribuem:

$$H_{O_1}(O_2) = \mathbb{E}_{\mathbb{P}_1} \left[H_{|O_1}(O_2) \right],$$

com

$$H_{|O_1}(O_2) = \mathbb{E}_{\mathbb{P}(O_2|O_1)} \left[I(O_2|O_1) \right]$$

em que $\mathbb{P}(O_2|O_1)$ representa a probabilidade de O_2 , dado O_1 . Equivalentemente:

$$H_{O_1}(O_2) = \mathbb{E}_{\mathbb{P}}\left[\log \mathbb{P}(O_2|O_1)\right];$$

 Decomposição da Entropia (consequência direta da noção de Probabilidade Condicional):

$$H(O_1, O_2) = H(O_1) + H_{O_1}(O_2);$$

- 4. Relações ligadas à independência entre O_1 e O_2 , em cujos casos:
 - (a) Conhecimento de um objeto não altera a incerteza sobre o outro:

$$H_{O_1}(O_2) = H(O_2);$$

(b) A incerteza conjunta é aditiva:

$$H(O_1, O_2) = H(O_1) + H(O_2).$$

Neste contexto, vale a pena também considerar a chamada *informação mútua*, dada pela diferença das incertezas associadas a O_2 isoladamente e a O_2 à luz do conhecimento de O_1 :

$$I(O_1, O_2) = H(O_2) - H_{O_1}(O_2), (2.2.2)$$

que, a exemplo da entropia conjunta, também é uma quantia simétrica, havendo independência sempre que for a mesma incerteza de uma variável, conhecida a outra. Pode-se mostrar que a Equação 2.2.2 é dada por:

$$I(O_1, O_2) = \mathbb{E}_{\mathbb{P}}\left[\log \frac{\mathbb{P}}{\mathbb{P}_1 \mathbb{P}_2}\right],$$
(2.2.3)

em que \mathbb{P} é probabilidade conjunta de O e \mathbb{P}_i , as marginais de O_i (i = 1, 2). A nulidade de $I(O_1, O_2)$, no caso de independência, segue diretamente da definição correspondente em teoria de probabilidades.

Nas considerações feitas até aqui, dois aspectos devem ser enfatizados: (a) as medidas são essencialmente simétricas, o que exclui a possibilidade imediata de se falar em causalidade entre $O_1 \in O_2 \in (b)$ a determinação empírica dos valores requer que se estimem densidades de probabilidades, de forma explicita, exigindo, em geral, uma quantidade grande de repetições dos experimentos aleatórios envolvidos.

Outro fator que deve ser lembrado é que quando os objetos O são processos
aleatórios, é necessário rever as definições de modo a permitir a inclusão de cada uma das variáveis aleatórias que compõem o processo, tomando-se o limite para todo o período infinito de tempo ocupado pelos processos. Nesse caso, as quantias referidas se tornam taxas de informação e de entropia já que se faz necessário dividir pelo número crescente de variáveis consideradas. Por exemplo, para se escrever a taxa de entropia de um processo $\{X_n\}_{n\in\mathbb{Z}}$, tem-se:

$$H(\{X_n\}) = \lim_{n \to \infty} \frac{1}{n} H(X_1, \dots, X_n),$$

em que $H(X_1, \ldots, X_n)$ é a entropia das variáveis aleatórias indexadas, X_i , que compõem o processo, $\{X_n\}_{n\in\mathbb{Z}}$. Note-se que valem definições análogas para as outras grandezas.

Um dos méritos principais da causalidade de Granger e da coerência parcial direcionada é a sua não reciprocidade inerente, sendo que isso permite associarlhe a noção de fluxo de informação. No contexto de processos/séries temporais, a simetria inerente às quantias ligadas à Teoria da Informação vistas aqui, até agora, são simétricas. Para quebrar essa simetria que as caracteriza, é preciso relacionar as variáveis que compõem os processos distintos ao considerá-las convenientemente deslocadas no tempo:

$$I({Y_1,...,Y_m}, {X_1,...,X_n}|X_1,...,X_n),$$

com m > n, conforme a interessante revisão de [40] e variantes desta idéia. Embora haja diversas formas de estimar as quantidades envolvidas [41, 42], não há, na prática, muita clareza quanto ao comportamento das estimativas. Isto é especialmente verdadeiro em termos do número de amostras temporais necessárias (duração do trecho observado da dinâmica). Note-se a quantidade de omissão quanto a aspectos práticos em [40].

De um modo geral, o problema comum a estas formas de estimação é a

necessidade de um grande número de observações para convergência, mesmo em circunstâncias favoráveis (e pouco práticas) na ausência de ruído aditivo.

Para superar essas dificuldades, Pincus propôs a chamada entropia aproximada [19, 20, 25, 43] (doravante, abreviada como ApEn — approximate entropy), voltada à medida da regularidade de séries temporais e com forte inspiração em abordagens voltadas à caracterização do comportamento de sistemas dinâmicos autônomos e seus atratores [44, 45], ao descrever-lhes as propriedades de recorrência nos respectivos espaços de estado. A proposta de Pincus tem sido largamente empregada com sucesso em problemas de reconhecimento de padrões (sadio versus patológico) [46,47] e se propõe a produzir uma quantia robusta (relativamente a ruído aditivo) baseada em séries empíricas de curta duração, algo que será melhor discutido no Capítulo 3, ao lado de uma outra proposta, a da entropia amostral (abreviada aqui, doravante como SampEn — sample entropy) introduzida no intuito de corrigir problemas de viés na estimativa da quantia introduzida por Pincus [48]. Uma breve porém interessante revisão comparativa destas formas de entropia, pode ser apreciada em [49]. A essa altura, vale a pena sublinhar a afirmação dos autores da SampEn, que esta seria aplicável não somente a trechos curtos de sinais, como a não estacionários, fatos que, em princípio, são auspiciosos para a aplicação pretendida aqui. A seguir, detalham-se as definições dessas entropias.

2.2.1 Entropia aproximada

Formalmente, define-se a ApEn como:

Definição 2.2.1. Dados uma série temporal $\{u(n)\}_{n=1}^{N}$ de comprimento N, um inteiro $m \leq N$ e uma tolerância r > 0, sejam os blocos $\mathbf{x}_m(i)$ e $\mathbf{x}_m(j)$, de comprimento m $(1 \leq i \leq N - m + 1)$, em que $\mathbf{x}_m(i) = (u(i), u(i+1), \cdots, u(i+1))$

m-1)), cuja distância é medida pela norma:

$$d[\mathbf{x}_m(i), \mathbf{x}_m(j)] = \max_{0 \le k \le m-1} (|u(i+k-1) - u(j+k-1)|) = \|\mathbf{x}_m(i) - \mathbf{x}_m(j)\|_{\infty}.$$

Então, conta-se o número de blocos no sinal $\mathbf{u}(\cdot)$ que dista de $\mathbf{x}(i)$ menos do que r, usando:

$$C_i^m(r) = \frac{1}{N-m+1} \sum_{j=0}^{N-m+1} \Theta(r - d[\mathbf{x}(i), \mathbf{x}(j)]),$$

em que $\Theta(\cdot)$ é função degrau de Heaviside. Agora, tomando média aritmética do logaritmo destas contagens:

$$\Phi^{m}(r) = \frac{1}{N-m+1} \sum_{i=1}^{N-m+1} \ln C_{i}^{m}(r)$$

obtém-se a estimativa da entropia aproximada como:

$$\widehat{\operatorname{ApEn}}(m,r,N)(\mathbf{u}) = \begin{cases} \Phi^m(r) - \Phi^{m+1}(r), & m \ge 1; \\ -\Phi^1(r), & m = 0; \end{cases}$$

que converge para a entropia aproximada:

ApEn
$$(m, r) = \lim_{N \to \infty} [\Phi^m(r) - \Phi^{m+1}(r)].$$
 (2.2.4)

Manipulações aritméticas [19], permitem concluir que:

$$-\widehat{\operatorname{ApEn}}(m, r, N)(\mathbf{u}) \xrightarrow{p} (\text{média aritmética sobre } i \text{ de:})$$

$$\ln[\text{frequência condicional de } (\|\mathbf{x}_m(i) - \mathbf{x}_m(j)\|_{\infty}),$$

$$\operatorname{dado} (\|\mathbf{x}_k(i) - \mathbf{x}_k(j)\|_{\infty})], \text{ para } 0 \le k \le m - 1, \forall i, j,$$

$$(2.2.5)$$

ou seja,

$$\widehat{\operatorname{ApEn}}(m,r,N)(\mathbf{u}) \xrightarrow{p} \mathbb{E} \left\{ \ln \left[\mathbb{P} \left(|u(j+m-1) - u(i+m-1)| \le r \right) \right] \right\}, \quad (2.2.6)$$
$$\operatorname{para} 0 \le k \le m-1, \forall i, j,$$

o que justifica a adoção do nome entropia já que se trata de uma esperança (ainda

que aproximada sobre probabilidades condicionais). Conceitualmente, está-se medindo a probabilidade de uma sequência de comprimento m possuir divergências na sua sequência temporal, conforme o comprimento m vai sendo aumentado. Quando, em média, esta frequência condicional de acertos se mantém, a medida que m cresce, o processo é bastante regular e previsível de modo que sua complexidade é baixa. Este é o modo pelo qual a ApEn mede a incerteza associada a uma dada sequência.

Adotando o relato de [48, 50], descritivamente, a ApEn quantifica a verossimilhança logarítmica de padrões similares (dentro de um raio r) em segmentos de m observações contíguas que se mantém próximas dentro do mesmo raio de tolerância r, de ainda se manterem juntos na seguinte comparação incremental, a medida que m aumenta.

A forma probabilística da ApEn garante robustez a "outliers" [48] e permite que seja interpretada como medida de correlação, persistência ou regularidade, ou seja, valores baixos significam maior persistência e correlação enquanto valores altos, mais independência e menos previsibilidade.

No contexto da Definição 2.2.1, cabem considerações relativas a escolha dos parâmetros $r \in m$ para uma adequada quantificação da ApEn. Comentários acerca desses aspectos são feitos no Capítulo 3, juntamente com a discussão da ferramenta de cálculo, implementada no contexto do presente estudo.

2.2.2 Entropia amostral

Uma das principais questões que induziram a definir a entropia amostral, enquanto alternativa à entropia aproximada, foi o problema do viés estatístico associado às estimativas dessa última e discutidos aqui na Seção 3.1.3, na qual se ilustra numericamente esse comportamento. Entropia amostral (SampEn), desenvolvida por Lake *et.al.* [21], também afere as similaridades de trechos de uma série temporal e, de uma mesma maneira, atribui valores mais altos à maior irregularidade nos dados [30]. De forma semelhante também requer a fixação de dois parâmetros o tamanho, m, das subsequências a comparar e um raio de tolerância, r, medido em termos de uma fração do desvio padrão, do sinal sob análise.

Formalmente,

Definição 2.2.2. Dada uma série temporal $u(n) := (u(1), \dots, u(N))$, sejam os blocos $\mathbf{x}_m(i) = (u(i), u(i+1), \dots, u(i+m-1))$, de comprimento m $(1 \le i \le N-m+1)$ e

$$B_i^m(r) = \frac{1}{N - m - 1} \sum_{\substack{j=1\\ j \neq i}}^{N - m} \Theta\left(r - \|\mathbf{x}_m(j) - \mathbf{x}_m(i)\|_{\infty}\right),$$

que conta a fração de vezes em que sequências de comprimento m se encontram em um raio de tolerância, r, de uma sequência de referência $\mathbf{x}_m(i)$, medianizadas com relação a todas sequências, i:

$$B^{m}(r) = \frac{1}{N-m} \sum_{i=1}^{N-m} B_{i}^{m}(r).$$

Relativamente ao aumento da dimensão em uma unidade, tem-se:

$$A_i^m(r) = \frac{1}{N-m-1} \sum_{\substack{j=1\\j\neq i}}^{N-m} \Theta\left(r - \|\mathbf{x}_{m+1}(j) - \mathbf{x}_{m+1}(i)\|_{\infty}\right),$$
$$A^m(r) = \frac{1}{N-m} \sum_{i=1}^{N-m} A_i^m(r),$$

cuja relação conduz à:

$$\widehat{\operatorname{SampEn}}(m, r, N) = -\ln\left[\frac{A(r)}{B(r)}\right],$$

que assume o valor assintótico do valor da entropia amostral, dada por:

SampEn
$$(m, r) = \lim_{N \to \infty} \left\{ -\ln \left[\frac{A(r)}{B(r)} \right] \right\}.$$

Pode-se entender a SampEn(m, r, N) como o negativo do logaritmo natural da probabilidade condicional $(\widehat{CP}(m, r))$ que um conjunto de dados de tamanho N, repete-se, através de um raio de tolerância r, por m pontos, sendo também similar por m + 1 pontos. Note-se que são explicitamente excluídas as autosimilaridades do ponto consigo mesmo ao contrário da ApEn [21], o que contribui para diminuir o viés na estimativa.

Entre as vantagens da SampEn, encontra-se um menor viés estatístico e a possibilidade explicita de se fornecer intervalos de confiança teóricos para o comportamento do seu estimador, assintoticamente Gaussiano (ver Seção 3.2.3).

2.3 Sumário

Nesse capítulo foram discutidos aspectos ligados à inferência da conectividade por meio de estimadores lineares e da Teoria da Informação.

Aos estimadores lineares deu-se um panorama da causalidade de Granger. Revisou-se, também, essa forma de causalidade no domínio da frequência, ditas DC, DTF e PDC, em que as duas últimas possuem relação com a Teoria da Informação.

Já para a Teoria da Informação, fez-se uma recapitulação sobre informação, entropia, a inferência dessas quantias para processos estocásticos e foi destacada a inviabilidade de boas estimativas e, também, de um bom controle sobre as mesmas.

Finalmente, foram apresentadas e definidas as **Entropias Aproximada** e **Amostral**, que são medidas de regularidade, inspiradas em caracterização de sistemas dinâmicos e utilizadas em reconhecimentos de padrões.

Essas últimas formas de entropia são muito mais convenientes para se estimar, pois não há necessidade de se ter o conhecimento da pdf do sinal, uma vez que são aplicadas a séries de curta duração e possuem imunidade a ruído e outliers.

No Capítulo 3 será tratada a validação dos algoritmos da ApEn e SampEn e alguns resultados estatísticos importantes, referentes aos mesmos.

3 CARACTERIZAÇÃO DOS ESTIMADORES DE ENTROPIA

O presente capítulo discute a implementação e os resultados de validação dos algoritmos, para o caso das entropias aproximada e amostral. Ele se organiza apresentando os desenvolvimentos associados a cada um dos casos e revisando alguns resultados teóricos conhecidos e que permitam aferir se os algoritmos possuem as propriedades esperadas teoricamente.

Cada um dos casos inicia-se pela descrição dos algoritmos seguida de suas propriedades, que são ilustradas a seguir, a partir de seu uso.

O capítulo termina com a inclusão de uma ilustração preliminar que demonstra a forma que se utilizam os dados para aferir a causalidade entre séries.

3.1 Entropia aproximada — ApEn

3.1.1 Descrição do algoritmo da ApEn

A seguir, descreve-se passo-a-passo o algoritmo implementado [46, 51, 52], apresentado-o na forma de pseudo-linguagem (Algoritmo 1), enquanto sua implementação em MATLAB pode ser apreciada no Capítulo A.

Dada uma série temporal $\mathbf{u} = [u(1), \cdots, u(N)]^{\top}$, selecione-se uma janela de tamanho *m* amostras (ou *tamanho padrão* na literatura) da qual se monta uma

sequência de vetores alocados na matriz de dados:

$$\mathbf{X} = [\mathbf{x}(1), \cdots, \mathbf{x}(N-m+1)], \qquad (3.1.1)$$

cujas colunas são dadas por:

$$\mathbf{x}(i) = [u(i), \cdots, u(i+m-1)]^{\top}.$$
 (3.1.2)

Selecione-se, também, o limiar de comparação, r^1 , para caracterizar a distância entre vetores $\mathbf{x}(i) \in \mathbf{x}(j)$, que são contados como similares, quando:

$$|x(i+k) - x(j+k)| < r, \forall k, 0 \le k \le m - 1.$$
(3.1.3)

Essas considerações estão resumidas no Algoritmo 1, cuja complexidade computacional é $\mathcal{O}(N^2)$. O código, em MATLAB, é fornecido no Capítulo A.

Algoritmo 1 Cálculo de $\widehat{\operatorname{ApEn}}(m, r, N)(\mathbf{x})$ (*MATLAB–Like*)

1: function Entropia Aproximada = $\mathbf{ApEn}(\mathbf{x}, m, r)$ 2: $N = \text{length}(\mathbf{x})$ 3: $\mathbf{x} = \frac{\mathbf{x} - \mu_{\mathbf{x}}}{\sqrt{\sigma_{\mathbf{x}}^2}}$ 4: $\Phi^m = 0$ 5: for all i tal que $1 \le i \le N - m + 1$ do $c_m = 0$ 6: for all j tal que $1 \le j \le N - m + 1$ do 7: 8: if $\|\mathbf{x}_m(i) - \mathbf{x}_m(j)\|_{\infty} < r$ then $c_m = c_m + 1$ 9: end if 10:end for $\Phi^m = \Phi^m + \ln\left(\frac{c_m}{N - m + 1}\right)$ 11: 12:13: **end for** 14: $\Phi^m = \frac{1}{N - m + 1} \Phi^m$ 15: Aumentar a dimensão de m em uma unidade: $m \to m+1$ 16: Com a nova dimensão, repetir os passos (5-15)17: Entropia Aproximada = $\Phi^{m+1} - \Phi^m$

 $^{^{1}}$ Uma discussão mais detalhada sobre como obter os valores de m e r estão contidas nas Seções 3.1.5 e 3.2.4.

3.1.2 Expressões analíticas para ApEn

Em muitos casos, podem-se prover expressões analíticas, com probabilidade um, para ApEn(m, r). Duas delas associam-se aos Teoremas 1 e 2 de [25], repetidas aqui, para referência²:

Teorema 3.1.1. Para um processo estacionário u(i) com representação em um espaço de estados contínuo munido da medida de probabilidades contínua $\mu(x, y)$ conjunta em \mathbb{R}^2 sendo $\pi(x)$ a probabilidade de equilíbrio de x, então

$$\widehat{\operatorname{ApEn}}(1,r,N) \xrightarrow{p} - \int_{-\infty}^{\infty} \mu(x,y) \ln \left(\frac{\int_{z=y-r}^{z=y+r} \int_{w=x-r}^{w=x+r} \mu(w,z) dw dz}{\int_{w=x-r}^{w=x+r} \pi(w) dw} \right) dx dy.$$
(3.1.4)

De modo análogo,

Teorema 3.1.2. Para um processo i.i.d. com função densidade de probabilidade $\pi(x)$

$$\widehat{\operatorname{ApEn}}(m,r,N) \xrightarrow{p} - \int_{-\infty}^{\infty} \pi(y) \ln\left(\int_{z=y-r}^{z=y+r} \pi(z) dz\right) dy.$$
(3.1.5)

O Teorema 3.1.1 pode ser estendido a fim de se derivar uma expressão para ApEn(m, r), em termos da distribuição de probabilidades conjunta. Assim, podese facilmente calcular a ApEn(m, r), para processos Gaussianos, a partir de suas distribuições conjuntas expressas pela suas matrizes de covariância.

O Teorema 3.1.2 pode ser usado para testar hipóteses de independência em processos identicamente distribuídos, já que, neste caso, o valor da ApEn independe de m, bastando, para isso, verificar a igualdade ApEn(1,0,18) =ApEn(2,0,18) = ApEn $(m,0,18), \forall m$, dentro dos limites de erro estatístico.

Para um processo *i.i.d.* com distribuição uniforme no intervalo $(-\sqrt{3},\sqrt{3})$,

 $^{^2\}mathrm{A}$ demonstração desses teoremas consta no Capítulo D

usando o Teorema 3.1.2, tem-se, [48]:

$$-\operatorname{ApEn}(m,r) = \left(1 - \frac{r}{\sqrt{3}}\right) \ln\left(\frac{r}{\sqrt{3}}\right) + \frac{1}{\sqrt{3}} \int_0^r \ln\left(\frac{r+y}{2\sqrt{3}}\right) dy$$
$$= \left(1 - \frac{r}{\sqrt{3}}\right) \ln\left(\frac{r}{\sqrt{3}}\right) + \frac{r}{\sqrt{3}} \left[\ln\left(\frac{2}{\sqrt{3}}\right) + \ln(r) - 1\right],$$
(3.1.6)

que assume o valor ApEn $(m, 0, 18) \approx 2,295$, quando r = 0,18.

O comportamento do estimador $\widehat{\operatorname{ApEn}}(m, r, N)$, calculado pelo Algoritmo 1, com respeito à sua dependência de N, está ilustrado na Figura 1, em que fica claro o erro sistemático, discutido na Seção 3.1.3 (a seguir).



Figura 1: Convergência assintótica de $\widehat{\operatorname{ApEn}}(m, r, N)$ para $\operatorname{ApEn}(m, r)$, no caso de uma sequência *i.i.d.* uniforme entre $-\sqrt{3} \in \sqrt{3}$.

3.1.3 Viés da ApEn

Como observa Pincus [48], o estimador da ApEn é viesado, fato que, deu o impulso original para a introdução da SampEn [21]. Isso significa que, quando calculada a partir de sequências empíricas, a ApEn possui sempre um erro sistemático, no seu valor numérico, ao ser comparada com o valor teórico. Esse viés, segundo [48], advém de duas fontes:

- 1. No cálculo de $C_i^m(r)$ é levada em conta a semelhança do vetor de referência consigo mesmo. Isso é feito para assegurar que os logaritmos possuam valor finito, já que, no caso contrário, $C_i^m(r)$ assumiria valor igual a zero e a ApEn não seria definida. Porém, ao se fazer isso, as estimativas das probabilidades condicionais, são consistentementes subestimadas. Se há poucos de $\mathbf{x}(j)$ próximos $\mathbf{x}(i)$ dentro do raio r, o procedimento discutido pode acarretar um viés de 20% a 30%;
- O logaritmo é uma função côncava; o valor de C^m_i(r) é estimado pelo valor amostral, mas se toma o logaritmo da média e não a média do logaritmo o que fica regido conforme a desigualdade de Jensen para o caso logarítmico (*i.e.*, E[log(X)] < log(E[X])).

A SampEn modifica a ApEn e não leva em conta o próprio vetor de referência na contagem [21,30,53,54]. Já o viés devido à concavidade logarítmica é comum às estatísticas de regularidade [48].

3.1.4 Normalidade assintótica

Uma propriedade importante de um estimador é sua distribuição assintótica. No caso da $\widehat{\operatorname{ApEn}}(m, r, N)$, há evidências de que sua distribuição assintótica seja razoavelmente bem aproximada por uma distribuição normal (Gaussiana), no contexto de diversos processos de interesse [48]. Para confirmar esse comportamento, são apresentados os resultados relativos a dois processos: (a) um de misturas $MIX(p)^3$ e (b) um processo de Markov GMP(p) ('Gaussian Markov Process')⁴ ambos retirados de [48], cujas distribuições e normalidade aproximada podem ser apreciados respectivamente nas Figuras 2 (MIX(0,5)) e 3 (GMP(0,2)), calculados usando m = 2, r = 0.18 e N = 1000.

Observação: A listagem em MATLAB da implementação dos processos MIX(p)e GMP(p) estão contidas no Capítulo A.

3.1.5 Escolha dos parâmetros $m \in r$

Com base em cálculos que incluem análises teóricas de processos determinísticos e estocásticos [25,48] e aplicações clínicas [43,47,55,56], conclui-se que para m = 2, $N = 1000 \text{ e } 0,1\sigma_{u(i)} \leq r \leq 0,25\sigma_{u(i)}$, tem-se uma boa aproximação da ApEn. A escolha dos valores de m e r, para a estatística $\widehat{\text{ApEn}}(m,r,N)$, depende do número de dados disponíveis. Geralmente, deve-se ter r tão pequeno e m tão grande quanto possível. Em muitas aplicações, o número de pontos de um sinal tem variado entre 100 e 5000.

³Seja $0 \le p \le 1$, com $p \in \mathbb{R}$. MIX(p) é dado por

 $\mathrm{MIX}(p)_j = (1 - Z_j)X_j + Z_jY_j,$

em que $X_j = \sqrt{2} \operatorname{sen}\left(\frac{2\pi}{12}j\right), \forall j, \{Y_j\}_{j\in\mathbb{Z}}$ é uma v.a. $i.i.d. \sim \mathsf{U}(-\sqrt{3},\sqrt{3})$ e $\{Z_j\}_{j\in\mathbb{Z}}$ é uma v.a. $i.i.d. \sim \mathsf{Ber}(p)$.

Essa família de processos estocásticos amostra a senóide para p = 0, é *i.i.d.* uniforme para p = 1 e, intuitivamente, se torna mais aleatória a medida que o valor de p aumenta (MIX possui média zero e desvio padrão igual a um, para todo p e, portanto, esses momentos não discriminam os membros da família MIX).

⁴Seja $0 \le p \le 1$, com $p \in \mathbb{R}$. GMP(p) é dado por

$$GMP(p) = X_i = pX_{i-1} + (1 - p^2)Z_i, X_1 = Z_1,$$

em que $\{Z_i\}_{i \in \mathbb{Z}}$ é uma v.a. *i.i.d.*~ N(0, 1). Note que essa outra família de processos estocásticos é um processo autorregressivo de ordem um (AR(1)) para 0 e é*i.i.d.*normal padrão para <math>p = 0.



Figura 2: Evidência de normalidade assintótica da estatística $\widehat{\text{ApEn}}(m, r, N)$, mostrada em (a) e (b), utilizando-se o processo MIX(0,5), m = 1, r = 0,18, N = 1000. Foram utilizadas 1500 replicações de Monte Carlo.



Figura 3: Evidência de normalidade assintótica da estatística $\widehat{\operatorname{ApEn}}(m, r, N)$, mostrada em (a) e (b), utilizando-se o processo $\operatorname{GMP}(0,2)$, m = 1, r = 0,18, N = 1000. Foram utilizadas 200 replicações de Monte Carlo.

3.2 Entropia amostral — SampEn

3.2.1 Descrição do algoritmo da SampEn

Usando notação semelhante às das Seções 3.1.1 e 2.2.2, resulta uma possível implementação em pseudo-código para calcular a entropia amostral, contida no Algoritmo 2, que possui complexidade $\mathcal{O}(N^2)$. Seu código, em MATLAB, é fornecido no Capítulo A.

Algoritmo 2 Cálculo de Samp $En(m, r, N)(\mathbf{x})$ (*MATLAB–Like*)

```
1: function Entropia Amostral = \mathbf{SampEn}(\mathbf{x}, m, r)
 2: N = \text{length}(\mathbf{x})
 3: \mathbf{x} = \frac{\mathbf{x} - \mu_{\mathbf{x}}}{\sqrt{\sigma_{\mathbf{x}}^2}}
4: B_m = 0
 5: A_m = 0
 6: for all i tal que 1 \le i \le N - m do
          \mathbf{x}_{m}^{(i)}(1:m) = \mathbf{x}(i:(i+m-1))
 7:
          \mathbf{x}_{m+1}^{(i)}(1:(m+1)) = \mathbf{x}(i:(i+m))
 8:
          for all j tal que j \neq i do
 9:
               \mathbf{x}_{m}^{(j)}(1:m) = \mathbf{x}(j:(j+m-1))
10:
               \mathbf{x}_{m+1}^{(j)}(1:(m+1)) = \mathbf{x}(j:(j+m))
11:
               d_{\max m} = \|\mathbf{x}_{m}^{(i)} - \mathbf{x}_{m+1}^{(j)}\|_{\infty}
d_{\max m+1} = \|\mathbf{x}_{m+1}^{(i)} - \mathbf{x}_{m+1}^{(j)}\|_{\infty}
B_{m} = B_{m} + (r > d_{\max m})
12:
13:
14:
               A_m = A_m + (r > d_{\max_{m+1}})
15:
          end for
16:
17: end for
18: Entropia Amostral = -\ln\left(\frac{A_m}{B_m}\right)
```

3.2.2 Viés da SampEn

Como já foi dito, a idéia de se introduzir a SampEn foi a de reduzir o viés da ApEn, como é discutido em [21, 30, 53, 54]. Outras considerações sobre o viés da SampEn são revistas na Seção 3.1.3.



Figura 4: Convergência assintótica da estatística $\widehat{\text{SampEn}}(m, 0, 18, N)$, para seu valor teórico, no caso de um processo *i.i.d.* uniforme no intervalo $(-\sqrt{3}, \sqrt{3})$. Vê-se claramente que o viés é menor, do que no caso da $\widehat{\text{ApEn}}(m, 0, 18, N)$.

3.2.3 Comportamento assintótico

Repetindo o processo de investigação empírica usada para o caso da ApEn, obtém-se as Figuras 5 e 6 para os processos MIX(0,5) e GMP(0,2), respectivamente. Dessa forma, conclui-se que o estimador da SampEn converge em distribuição para uma Gaussiana, ou seja:

$$\widetilde{\operatorname{SampEn}}(m, r, N) \xrightarrow{d} \mathsf{N}(\mu, \sigma^2).$$

É interessante observar que é relativamente fácil obter expressões aproximadas e fechadas para a variância da $\widehat{\text{SampEn}}(m, r, N)$ e, daí, construir intervalos de confiança, que podem ser usados em testes estatísticos [21, 53]. Uma discussão mais detalhada está contida no Capítulo E.

3.2.4 Escolha dos parâmetros $m \in r$

Apesar da característica não-paramétrica do estimador (SampEn), em [21] os autores sugerem a realização de um modelo autorregressivo (AR) dos dados, para que se faça a escolha de m (o valor de m é o valor da ordem do AR), determinando uma possível faixa de valores. O mínimo dessa, será o m escolhido. Note-se que [53] discute sistemáticas baseadas em otimização dos intervalos de confiança da SampEn para definir r.

Aqui, por conveniência, aproveitando os estudos sistemáticos [21,53,54], bem como os de Pincus, adotam-se na prática os valores:

$$\begin{cases} m \in \{1, 2\}; \\ 0, 1\sigma_{u(i)} \le r \le 0, 25\sigma_{u(i)}. \end{cases}$$



Figura 5: Evidência de normalidade assintótica da estatística $\widehat{\text{SampEn}(m, r, N)}$, mostradas em (a) e (b), utilizando-se o processo MIX(0,5), m = 1, r = 0,18, N = 1000 e realizando 1500 replicações de Monte Carlo.



Figura 6: Evidência de normalidade assintótica da estatística $\widehat{\text{SampEn}(m, r, N)}$, mostradas em (a) e (b), utilizando-se o processo GMP(0,2), m = 1, r = 0,18, N = 1000 e realizando 200 replicações de Monte Carlo.

3.3 Sumário

Nesse capítulo foram apresentadas as implementações, em pseudo-linguagem, dos algoritmos da ApEn e da SampEn, assim como suas propriedades.

Mostrou-se que é possível a obtenção de expressões analíticas para o cálculo do valor assintótico da ApEn. Para a SampEn, graças ao estimador possuir normalidade assintótica, no Capítulo E mostra-se que é possível obter, com razoável facilidade, seu desvio padrão, para determinar o intervalo de confiança.

Por fim, discutiram-se os métodos utilizados na literatura para escolha dos parâmetros $m \in r$, que são os mesmos usados nesta dissertação.

No Capítulo 4, será discutida uma aplicação desses algoritmos com a finalidade de reconstrução de séries temporais, mapeando-as em outras, em termos de suas medidas de entropia, variando no tempo.

4 DESCRITORES INFORMACIONAIS DE INFERÊNCIA CAUSAL

O objetivo deste capítulo é discutir, mais profundamente, a possibilidade de usar descritores informacionais de séries temporais, como os discutidos no Capítulo 3 e, além disso, inferir suas relações causais.

Para tanto, analisam-se, além do método para gerar as séries de ApEn e SampEn (Seção 4.1), exemplos ilustrativos (Seção 4.3) do uso desses descritores informacionais na determinação da causalidade. Apresenta-se, em primeiro lugar, um panorama da potencialidade do uso desses descritores para inferir a causalidade.

O capítulo termina por verificar que, com a utilização dos descritores informacionais há a possibilidade de uma melhoria da correta inferência da direção de conectividade.

4.1 Método

As duas fases do método são (a) o mapeamento e, (b) a análise linear das séries mapeadas.

4.1.1 Mapeamento

Considere uma série temporal $x_i(n)$, com N pontos, em que n e i são, respectivamente, os indexadores do tempo discreto e da série. Associada a $x_i(n)$, seja a sequência $\xi_i(n)$ gerada, a partir de uma janela deslizante de tamanho W, com passo de deslocamento de comprimento h:

$$\begin{bmatrix} x_i(n-W+h) & \cdots & x_i(n) \end{bmatrix},$$

que reflete uma medida de complexidade da série original na qual os pontos da janela são usados para calcular estatíticas como a entropia aproximada de Pincus [48] e a entropia amostral de Lake *et.al.* [21] (essa última corrige o viés estatístico da primeira). Portanto, fica justificada a terminologia doravante adotada como *entropia deslizante*.

Além do comprimento da janela, W e do passo, h (que indica como a janela se movimenta sobre a série), é necessário considerar os parâmetros de cálculo das entropias propriamente ditas, *i.e.*, a dimensão do espaço de estados, m e o raio de tolerância, r, para que se possa contar quantos padrões similares ocorrem em blocos com comprimento m, em uma distância r dentro da janela, segundo as Definições 2.2.1 e 2.2.2.



Figura 7: Ilustração do método de geração das "séries informacionais".

Dessa construção resulta a possibilidade de testar as relações entre as "séries informacionais" a partir de modelamento linear autorregressivo, como discutido na Revisão Bibliográfica (Seção 2.1). Embora a proposta do mapeamento de $x_i(n)$ seja geral, as escolhas destas medidas de entropia se justificam por sua rápida convergência, em termos do número de pontos em que são calculadas ($W \ll N$) e sua razoável imunidade a ruído conforme relatado em [57]. Outra vantagem é a Gaussianidade assintótica dos respectivos estimadores, que se faz consistente com os testes estatísticos adotados.

4.1.2 Análise linear das séries mapeadas

A próxima etapa consiste em aplicar a análise de causalidade ao mapeamento, relacionando as séries temporais $\xi_i(n)$, entre si, ao invés de se fazer diretamente com $x_i(n)$.

Embora existam muitas alternativas, aqui se utiliza a coerência parcial direcionada (*partial directed coherence* -PDC). Uma revisão da mesma, encontra-se no Capítulo C.

4.2 Convenções para se apresentar a PDC

Note-se que, neste texto, a convenção matricial usual [1] de se apresentar a PDC é adotada. Os gráficos dispostos na diagonal principal representam as densidades espectrais das séries (em escala logarítmica, mas com unidades arbitrárias) e aqueles que se apresentam em contradiagonais retratam as PDC, *i.e.*, a representação da conectividade espectral $i \leftarrow j$.

Nos gráficos, há diferentes cores, que indicam:

- traçado contínuo em azul: densidade espectral das séries;
- traçado contínuo em vermelho: valores estatisticamente significantes da PDC;

- traçado descontínuo em preto: limiar da aproximação de Patnaik [38], com nível de significância igual a α;
- traçado contínuo em verde: valores estatisticamente não-significantes da PDC;

Convenções similares são aplicadas a todos os gráficos de PDC.

4.3 Exemplos ilustrativos do uso da ApEn e da SampEn na determinação de causalidade

O objetivo desta seção é confrontar o comportamento da PDC entre as séries dos próprios modelos e suas respectivas "informacionais," com a motivação de que a PDC é bem observada entre séries lineares, mas não o é entre séries não-lineares. Para tanto, selecionou-se um modelo autorregressivo vetorial (Equação 4.3.1) e outros três, não-lineares: o logístico (Equações 4.3.2) e o de acoplamento quadrático, ambos bivariados (Equações 4.3.3) e o de acoplamento biquadrático, sistema 3D (Equações 4.3.4–4.3.5).

Para investigar a viabilidade da determinação de causalidade, cogitou-se a utilização de estimadores lineares robustos. Porém, sabe-se que esses não são capazes de capturar aspectos não-lineares do comportamento dinâmico. Assim, optou-se pelos modelos logístico e de acoplamento quadrático, bivariados e um mapa biquadrático (3D).

O uso desses modelos não-lineares ocorreu por sua maleabilidade de atuação em meio biológico. A escolha de modelos de acoplamento quadráticos de potências quadráticas e pares se deve ao fato de que é impossível a PDC ser detectada por essa forma de acoplamento e que, radiobiologicamente, é possível explicar fenômenos de radiação por modelos quadráticos [58]. Como contraponto, utilizou-se o modelo autorregressivo multivariado — linear — que dispensa o uso dos métodos informacionais.

4.3.1 Modelo autorregressivo vetorial (VAR)

Nessa seção, considera-se um modelo autorregressivo vetorial de ordem um (VAR(1)), dado a seguir:

$$\mathbf{x}_n = \mathbf{A}_1 \mathbf{x}_{n-1} + \mathbf{w}_n, \tag{4.3.1}$$

em que $\{\mathbf{w}_n\}_{n\in\mathbb{Z}} \sim \mathsf{N}(\mathbf{O}_{1\times 2}, \mathbf{I}_2)$ e

$$\mathbf{A}_1 = \begin{bmatrix} -0.5 & 0\\ 0.2 & 0.1 \end{bmatrix}.$$

Realizações típicas dos processos $\{x_{1n}\}_{n\in\mathbb{Z}} \in \{x_{2n}\}_{n\in\mathbb{Z}} (\mathbf{x}_n = [x_{1n} \ x_{2n}]^\top)$, podem ser apreciadas na Figura 8.

Dos resultados de [1, 38], a conectividade se dá no seguinte sentido:

$$x_{1n} \xrightarrow{Granger} x_{2n},$$

(Figura 9). Porém, ao se aplicar o mapeamento pelo algoritmo da ApEn (realizações típicas das séries de ApEn estão presentes na Figura 10), há um mascaramento da detectabilidade da direção da Causalidade de Granger, como consta na Figura 11.

A fim de confrontar esse resultado, nas Seções 4.3.2 e 4.3.3 foram usados modelos com realimentações não-lineares, para que, com a transformação em "séries informacionais," a PDC entre estas fosse observada.



Figura 8: Séries do Processo Autor
regressivo Vetorial de ordem um (VAR(1)), da Equação 4.3.1.



Figura 9: PDC, típica, na forma padrão (ver [1]), entre séries do $\mathrm{VAR}(1).$



Figura 10: Séries de entropia aproximada, calculadas com os dados da Figura 8, usando W = 150, h = 1, m = 1 e r = 0,15.



Figura 11: PDC, típica, na forma padrão (ver [1]), entre séries de ApEn, da Figura 10. Para se escolher a ordem, utilizou-se o AIC, que gerou um VAR(2).

4.3.2 Mapa logístico bivariado

Para fins de exemplo do uso das técnicas acima, simulou-se o seguinte mapa logístico bivariado, retirado de [59]:

$$\begin{cases} x_{n+1} = 1 - ax_n^2 + s\tau_n, & \{\tau_n\}_{n \in \mathbb{Z}} \sim \mathsf{N}(0, 1) \\ y_{n+1} = (1-b)(1 - ay_n^2) + b(1 - ax_n^2) + s\gamma_n, & \{\gamma_n\}_{n \in \mathbb{Z}} \sim \mathsf{N}(0, 1), & (4.3.2) \\ a = 1, 8, b = 0, 37, s = 0, 01 \end{cases}$$

explicitado na Figura 12 com N = 1500, usando m = 1, r = 0.18, W = 500 e h = 1 (Figuras 14 e 16). Tomou-se o cuidado de adotar os N pontos, descartando os 1.000 primeiros da simulação, para evitar problemas com transitórios. Note, na Figura 13, a dificuldade de detecção da Causalidade de Granger para tal modelo, sem antes completar o mapeamento, referido na Seção 4.1.

Após utilizar os algoritmos informacionais, têm-se as Figuras 14 e 16, que são realizações típicas de processos de ApEn e SampEn. Nas Figuras 15 e 17 são mostradas PDCs típicas, na forma padrão (ver [1]), calculadas entre, respectivamente, as séries das Figuras 14 e 16.

Dessa forma, utilizando a transformação, a determinação da PDC e usando as séries de entropia, obtiveram-se os resultados resumidos nas Figuras 15 e 17, que revelam serem significantes as relações:

$$\begin{cases} \xi_1^a(n) \xrightarrow{Granger}_{Causa} \xi_2^a(n) \\ \xi_1^s(n) \xrightarrow{Granger}_{Causa} \xi_2^s(n) \end{cases}$$

mas não suas inversas, que estão sob a linha tracejada. Note que foram usados os resultados estatísticos de [38].



(b) Série y_n do Mapa Logístico Bivariado.

Figura 12: Séries do Mapa Logístico Bivariado, das Equações 4.3.2.



Figura 13: PDC, típica, na forma padrão (ver [1]), entre as séries da Figura 12. Para se escolher a ordem, utilizou-se o AIC, que gerou um VAR(6).



Figura 14: Séries de entropia aproximada, calculadas com os dados da Figura 12, usando W = 500, h = 1, m = 1 e r = 0,18.



Figura 15: PDC, típica, na forma padrão (ver [1]), entre séries de ApEn, da Figura 14. Para se escolher a ordem, utilizou-se o AIC, que gerou um VAR(4).



Figura 16: Séries de entropia amostral, calculadas com os dados da Figura 12, usando W = 500, h = 1, m = 1 e r = 0,18.


Figura 17: PDC, típica, na forma padrão (ver [1]), entre séries de SampEn, da Figura 16. Para se escolher a ordem, utilizou-se o AIC, que gerou um VAR(3).

4.3.3 Mapa de acoplamento quadrático bivariado

Nessa seção, considerou-se o seguinte mapa de acoplamento quadrático bivariado para as simulações:

$$\begin{cases} x_n = 2R\cos(2\pi f)x_{n-1} - R^2 x_{n-2} + w_{1,n}, & \{w_{1,n}\}_{n \in \mathbb{Z}} \sim \mathsf{N}(0,1) \\ y_n = \alpha y_{n-1} + \beta x_{n-1}^2 + w_{2,n} & \{w_{2,n}\}_{n \in \mathbb{Z}} \sim \mathsf{N}(0,1) \cdot & (4.3.3) \\ \alpha = -0.9, f = 0.1 \end{cases}$$

Observe que o processo $\{x_n\}_{n\in\mathbb{Z}}$ é um autorregressivo de ordem 2 (AR(2)), com um tom modulado na frequência (normalizada — a frequência de amostragem, f_a , utilizada foi igual a um) f = 0,1, que está acoplado quadraticamente ao processo $\{y_n\}_{n\in\mathbb{Z}}$, que é um filtro passa-baixas, cuja conectividade pode ser aferida pelo parâmetro β . Assim, o esperado é que se tenha ao menos uma ressonância na segunda harmônica de f e a PDC, entre as séries x_n e y_n , apesar de muito difícil de ser capturada (como mostra a Figura 19). A Figura 18 exibe um par representativo de x_n e y_n , respectivamente, do conjunto de Equações 4.3.3, explicitado, com N = 1500 e gerados com R = 0,99 e $\beta = 0,1$. Aqui, novamente, adotou-se os N pontos, descartando os 1000 primeiros da simulação, para evitar os transitórios, assim como realizado na Seção 4.3.2.

Nas figuras 20 e 22 estão, respectivamente, realizações típicas dos processos $\{\xi_1^a(n)\}_{n\in\mathbb{Z}}, \{\xi_2^a(n)\}_{n\in\mathbb{Z}}, \{\xi_1^s(n)\}_{n\in\mathbb{Z}} \in \{\xi_2^s(n)\}_{n\in\mathbb{Z}}, \text{ gerados da mesma maneira}$ como explicado na Seção 4.1, mas com os seguintes parâmetros para as entropias aproximada e amostral: W = 150, h = 1, m = 1 e r = 0, 15.

No intuito de se determinar as PDC entre as séries de ApEn e SampEn, as séries do conjunto de Equações 4.3.3 foram simuladas empregando-se R = 0,99e com diferentes valores de β , no seguinte intervalo: {0,05,0,1}. Além disto, os seguintes parâmetros foram empregados para se obter as séries de ApEn e SampEn:

- h = 1;
- $m \in \{1, 2\};$
- $r \in \{0, 10, 0, 15, 0, 18\};$
- $W \in \{150, 200\}.$

Na Figura 21, pode ser vista a PDC entre as séries de ApEn (dadas na Figura 20), assim como na Figura 23, entre as séries de SampEn (mostradas na Figura 22), quando foram empregados os seguintes parâmetros: R = 0.99, $\beta = 0.1$, W = 150, h = 1, m = 1 e r = 0.18. Repare que, ao usar os resultados estatísticos de [38], é possível inferir que as relações:

$$\begin{cases} \xi_1^a(n) \xrightarrow{Granger} \xi_2^a(n) \\ \xi_1^s(n) \xrightarrow{Granger} Causa \\ \xi_2^s(n) \end{cases} \xrightarrow{Granger} \xi_2^s(n)$$

são significantes, mas não suas inversas.



(b) Série y_n do Mapa de Acoplamento Quadrático Bivariado.

Figura 18: Séries do Mapa de Acoplamento Quadrático Bivariado, calculadas usando as Equações 4.3.3 e os parâmetros $\beta = 0.05$ e R = 0.99.



Figura 19: PDC, típica, na forma padrão (ver [1]), entre as séries da Figura 18. Para se escolher a ordem, utilizou-se o AIC, que gerou um VAR(6).



Figura 20: Séries de entropia aproximada, calculadas com os dados da Figura 18, usando W = 150, h = 1, m = 1 e r = 0,18.



Figura 21: PDC, típica, na forma padrão (ver [1]), entre séries de ApEn, da Figura 20. Para se escolher a ordem, utilizou-se o AIC, que gerou um VAR(8).



Figura 22: Séries de entropia amostral, calculadas com os dados da Figura 18, usando W = 150, h = 1, m = 1 e r = 0,18.



Figura 23: PDC, típica, na forma padrão (ver [1]), entre séries de SampEn, da Figura 22. Para se escolher a ordem, utilizou-se o AIC, que gerou um VAR(6).

4.3.4 Mapa de acoplamento biquadrático (3D)

Nessa seção, considerou-se o seguinte mapa de acoplamento biquadrático (3D) para as simulações:

$$\begin{cases} x_{1,n} = 2R\cos(2\pi f)x_{1,n-1} - R^2 x_{1,n-2} + w_{1,n}, & \{w_{1,n}\}_{n \in \mathbb{Z}} \sim \mathsf{N}(0,1) \\ x_{2,n} = \alpha_1 x_{2,n-1} + 0.05 x_{1,n-1}^2 + w_{2,n} & \{w_{2,n}\}_{n \in \mathbb{Z}} \sim \mathsf{N}(0,1) \\ x_{3,n} = \alpha_2 x_{3,n-1} + x_{1,n-1}^4 + w_{3,n} & \{w_{3,n}\}_{n \in \mathbb{Z}} \sim \mathsf{N}(0,1) \\ \alpha_1 = -0.9, \alpha_2 = 0.5, f = 0.1, \end{cases}$$

$$(4.3.4)$$

$$\begin{cases} y_{1,n} = \frac{x_{1,n} - \mu_{x_{1,n}}}{\sigma_{x_{1,n}}} \\ y_{2,n} = \frac{x_{2,n} - \mu_{x_{2,n}}}{\sigma_{x_{2,n}}} \\ y_{3,n} = \frac{x_{3,n} - \mu_{x_{3,n}}}{\sigma_{x_{3,n}}}. \end{cases}$$
(4.3.5)

Repare que as Equações de $x_{1,n}$ e $x_{2,n}$ são as mesmas, respectivamente, de x_n e y_n , do mapa de acoplamento quadrático. Nesse modelo incluiu-se a equação de $x_{3,n}$, que é um filtro passa-altas, cuja conexão com $x_{1,n}$ é biquadrática. Dessa forma, o esperado é que se tenha ressonância de f pelo menos em 2f e em 4f, com as seguintes relações causais:

$$\begin{cases} x_{1,n} \xrightarrow{Granger} x_{2,n} \\ x_{1,n} \xrightarrow{Granger} x_{3,n}. \end{cases}$$

As transformações descritas nas Equações 4.3.5, do processo vetorial $\mathbf{x}_{\mathbf{n}} = [x_{1,n} \ x_{2,n} \ x_{3,n}]^{\top}$ para a Gaussiana padrão, ou seja, $\mathbf{y}_{\mathbf{n}} = [y_{1,n} \ y_{2,n} \ y_{3,n}]^{\top} \sim \mathsf{N}(\mathbf{O}_{3\times 1}, \mathbf{I}_3)$, foi realizada para que se pudesse limitar a amplitude do processo $\mathbf{x}_{\mathbf{n}}$.

A Figura 24 exibe uma tripla representativa de $y_{1,n}$, $y_{2,n}$ e $y_{3,n}$, respectivamente, do conjunto de Equações 4.3.4–4.3.5, explicitado, com N = 1500 e geradas com R = 0.99. Aqui, novamente, adotou-se os N pontos, descartando os 1000 primeiros da simulação, para evitar os transitórios, assim como foi feito nas Seções 4.3.2 e 4.3.3.

Nas figuras 26 e 28 estão, respectivamente, realizações típicas dos processos vetoriais $\boldsymbol{\xi}^{a}(n) = [\xi_{1}^{a}(n) \ \xi_{2}^{a}(n) \ \xi_{3}^{a}(n)]^{\top}$ e $\boldsymbol{\xi}^{s}(n) = [\xi_{1}^{s}(n) \ \xi_{2}^{s}(n) \ \xi_{3}^{s}(n)]^{\top}$, gerados da mesma maneira como explicado na Seção 4.1, mas com os seguintes parâmetros para as entropias aproximada e amostral: W = 150, h = 1, m = 1 e r = 0.15.

De maneira análoga à Seção 4.3.3, para se determinar as PDC entre as séries de ApEn e SampEn, as séries do conjunto de Equações 4.3.4–4.3.5 foram simuladas empregando-se R = 0.99. Além disso, os seguintes parâmetros, do espaço dos métodos de entropia, (h, W, m, r), foram empregados:

- h = 1;
- $m \in \{1, 2\};$
- r = 0.15;
- W = 200;

Na Figura 27, pode ser vista a PDC entre as séries de ApEn (dadas na Figura 26), assim como na Figura 29, entre as séries de SampEn (mostradas na Figura 28), quando foram empregados os seguintes parâmetros: R = 0.99, W = 150, h = 1, m = 1 e r = 0.15. Repare que, ao usar os resultados estatísticos de [38], é possível inferir que as relações:

$$\begin{cases} \xi_1^a(n) \xrightarrow{Granger}_{Causa} \xi_2^a(n) \\ \xi_1^a(n) \xrightarrow{Granger}_{Causa} \xi_3^a(n) \end{cases}$$
$$\begin{cases} \xi_1^s(n) \xrightarrow{Granger}_{Causa} \xi_2^s(n) \\ \xi_1^s(n) \xrightarrow{Granger}_{Causa} \xi_3^s(n) \end{cases}$$



são significantes, mas não suas inversas.

Figura 24: Séries do Mapa de Biquadrático (Sistema 3D), calculadas usando as Equações 4.3.3 e o parâmetro R = 0.99.



Figura 25: PDC, típica, na forma padrão (ver [1]), entre as séries da Figura 24. Para se escolher a ordem, utilizou-se o AIC, que gerou um VAR(16).



Figura 26: Séries da entropia aproximada, calculadas com os dados da Figura 24, usando W = 200, h = 1, m = 1 e r = 0,15.



Figura 27: PDC, típica, na forma padrão (ver [1]), entre séries de ApEn, da Figura 26. Para se escolher a ordem, utilizou-se o AIC, que gerou um VAR(6).



Figura 28: Séries de entropia amostral, calculadas com os dados da Figura 24, usando W = 200, h = 1, m = 1 e r = 0.15.



Figura 29: PDC, típica, na forma padrão (ver [1]), entre séries de SampEn, da Figura 28. Para se escolher a ordem, utilizou-se o AIC, que gerou um VAR(13).

4.4 Sumário

Nesse capítulo foi apresentada e proposta uma metodologia de mapeamento para séries temporais, utilizando-se métodos de entropia. As novas sequências obtidas podem, assim, ter suas relações causais testadas por um modelamento autorregressivo, utilizando-se da PDC.

Pelas figuras obtidas, observa-se que a detectabilidade da causalidade de Granger no dominío da frequência possui uma melhora através do mapeamento por "entropia deslizante," para os modelos de séries não-lineares, o que não ocorreu com o VAR. Para modelos de séries lineares, a utilização da metodologia acarreta no mascaramento do conteúdo espectral da causalidade de Granger. Assim, não se deve utilizar a técnica para modelos lineares.

No Capítulo 5 serão feitas simulações de Monte Carlo, com o mapa de acoplamento quadrático (Sistema 2D) e, também, com o mapa de biquadrático (Sistema 3D), para se verificar a robustez do método proposto.

5 VALIDAÇÃO ESTATÍSTICA DO MÉTODO PROPOSTO

Para se verificar robustez do método, nesse capítulo são feitas simulações de Monte Carlo, através de N_R realizações dos processos das séries de entropia, escolhendo-se o modelo de acoplamento quadrático (Sitema 2D) e, também, modelo de acoplamento quadrático (Sistema 3D), ambos apresentados no Capítulo 4, por serem não-lineares. Dessa forma, pode-se ter uma ideia da robustez do método proposto no capítulo anterior e de sua dependência no espaço dos parâmetros de cálculo das entropias, (h, W, m, r).

Com essas N_R replicações de Monte Carlo, uma Análise de Sensibilidade é feita quanto à detecção da Causalidade de Granger, no domínio da frequência. São propostas tabelas de contingência para se comparar o desempenho entre os algoritmos (ApEn *versus* SampEn).

Ao final deste capítulo, conclui-se que o algoritmo da SampEn é o mais adequado para o uso neste contexto, por possuir uma taxa menor de erros do que o algoritmo da ApEn.

5.1 Sistema 2D — Mapa de acoplamento quadrático (sistema de Equações 4.3.3)

A fim de se validar o método utilizado na Seção 4.3.3, foram feitas simulações de Monte Carlo, com o mapa de acoplamento quadrático bivariado das Equações 4.3.3, dos casos da Tabela 1, para se comparar os algoritmos da ApEn e SampEn, como mostram, mais adiante, as Tabelas 3 versus 4.

	W - 150	m = 2	r = 0,10
$R = 0.99, \ h = 1, \ \beta \in \{0.05, 0.10\}$	W = 100	m = 1	r = 0.15
	W - 200	m = 2	r = 0.10
	<i>vv</i> = 200	m = 1	/ = 0,10

Tabela 1: Casos simulados do sistema 2D.

Em cada cenário citado na Tabela 1, utilizando-se $N_R = 10000$ replicações Monte Carlo, foi feita uma Análise de Sensibilidade, quanto à detecção da Causalidade de Granger, no domínio da frequência.

Com a finalidade de que tal análise pudesse ser realizada, estabeleceu-se um critério de contagem de verdadeiros-positivos e falsos-negativos, assim como os de falsos-positivos e verdadeiros-negativos (aqui abreviados por VP, FN, FP e VN, respectivamente), utilizando-se a seguinte hipótese nula (\mathcal{H}_0) [60]:

$$\mathcal{H}_0: |\pi_{ij}(f)|^2 = 0,$$

em que $\pi_{ij}(f)$ é o (i, j)-ésimo coeficiente da coerência parcial direcionada (PDC), na frequência normalizada f.

Assim, através de um nível de significância α , calcula-se um valor de limiar, em uma certa frequência, que é nomeado de $\gamma_{ij}(f)$, para cada uma das N_R realizações. Como as Equações 4.3.3 geram outro par de séries de ApEn, ou de SampEn, com frequência característica f^* ($f^* = 0,1$), deve-se ter:

- se $\pi_{12}(2f^*)$ é estatisticamente diferente de zero, ou seja, $\pi_{12}(2f^*) > \gamma_{12}(2f^*)$, então há um verdadeiro-positivo, caso contrário, ocorre um falso-negativo;
- se $\pi_{21}(2f^*)$ é estatisticamente diferente de zero, ou seja, $\pi_{21}(2f^*) > \gamma_{21}(2f^*)$, então há um falso-positivo, caso contrário, ocorre um verdadeiro-negativo.

	\mathcal{H}_0	\mathcal{H}_0
	Verdadeira	Falsa
\mathcal{H}_0	\mathbf{VP}	\mathbf{FP}
rejeitada	$(1 - \beta)$	(\boldsymbol{lpha})
\mathcal{H}_0	\mathbf{FN}	VN
aceita	$(oldsymbol{eta})$	$(1 - \alpha)$
Totais	100%	100%

A partir das regras de decisão acima, foi possível gerar tabelas, como, por exemplo, a Tabela 2.

Tabela 2: Exemplo de tabela padrão para contagem de VP, FN, FP e VN.

Escolhendo $\alpha = 5\%$, a seguir, mostram-se as Tabelas 3 e 4, indicando as porcentagens de VP, FP (FN e VN são os respectivos complementares, já que se utilizou 10000 replicações de Monte Carlo, sendo, portanto, omitidos), para cada caso de simulação citados no início da Seção 5.1 (Tabela 1).

			VP	FP	VP	\mathbf{FP}
R = 0,99, h = 1		$\beta = 0.05$		$\beta = 0,10$		
W = 150	m=2	r = 0,10	57,69	46,69	92,48	74,89
	m = 1	r = 0.15	88,21	15,97	99,92	24,54
W = 200	m=2	r = 0,10	79,44	56,47	98,23	82,78
VV = 200	m = 1		91,89	18,13	99,91	27,03

Tabela 3: Porcentagens de VP e FP, referentes ao algoritmo da ApEn, para os casos da Tabela 1, do sistema 2D.

			VP	FP	VP	FP
R = 0,99, h = 1		$\beta = 0.05$		$\beta = 0,10$		
W = 150	m=2	r = 0,10	94,24	9,87	99,93	17,72
	m = 1	r = 0,15	100,00	22,69	100,00	41,42
W = 200	m = 2	m = 0.10	$95,\!85$	12,52	99,94	20,22
W = 200	m = 1	1 = 0,10	100,00	23,21	100,00	37,62

Tabela 4: Porcentagens de VP e FP, referentes ao algoritmo da SampEn, para os casos da Tabela 1, do sistema 2D.

Em seguida, comenta-se a respeito das PDCs entre as séries de ApEn e entre

as séries de SampEn.

5.1.1 Sobre as PDCs entre as séries de ApEn e entre as séries de SampEn, do mapa de acoplamento quadrático bivariado das Equações 4.3.3

Observando-se as Tabelas 3 e 4, algumas características das PDCs, entre as séries de ApEn e entre as de SampEn, podem ser observadas pela variação dos parâmetros, $R, m, r \in W$, destacando:

- a direção preferencial de causalidade quando a transformação pelos algoritmos informacionais é corretamente obtida;
- a variação de *m* tem pouco efeito sobre as estimativas;
- a detectabilidade da direção da conectividade aumenta com β , $r \in W$ crescentes.

Observação: aqui são utilizados resultados estatísticos de [38].

5.1.2 Tabelas de contingência

Com a finalidade de comparar o desempenho dos dois algoritmos, foram feitas as Tabelas de Contigência 5, 6, 7 e 8, que demonstram o quanto os dois algoritmos acertam conjuntamente, erram conjuntamente e se apenas um deles acerta.

	W = 150, m = 2, r = 0.10					
	$\beta = 0$	0,05	$\beta = 0$	0,10		
ApEn SampEn	Acerto	Erro	Acerto	Erro		
Acerto	55,88	1,81	92,45	0,03		
Erro	38,36	96,05	7,48	99,96		

Tabela 5: Tabela de contingência para R = 0,99, h = 1 e 10000 realizações, comparando-se os algoritmos da ApEn e SampEn, simulado com W = 150, m = 2, r = 0,10.

	W = 150, m = 1, r = 0.15					
	$\beta =$	0,05	$\beta =$	$0,\!10$		
ApEn SampEn	Acerto	Erro	Acerto	Erro		
Acerto	88,21	0,00	99,92	$0,\!00$		
Erro	11,79	100,00	0,08	100,00		

Tabela 6: Tabela de contingência para R = 0.99, h = 1 e 10000 realizações, comparando-se os algoritmos da ApEn e SampEn, simulado com W = 150, m = 1, r = 0.15.

	W = 200, m = 2, r = 0.10					
	$\beta = 0$	0,05	$\beta = 0$),10		
ApEn SampEn	Acerto	Erro	Acerto	Erro		
Acerto	77,43	2,01	98,21	0,02		
Erro	18,42	97,86	1,73	99,96		

Tabela 7: Tabela de contingência para R = 0.99, h = 1 e 10000 realizações, comparando-se os algoritmos da ApEn e SampEn, simulado com W = 200, m = 2, r = 0.10.

	W = 200, m = 1, r = 0.10					
	$\beta =$	0,05	$\beta =$	$0,\!10$		
ApEn SampEn	Acerto	Erro	Acerto	Erro		
Acerto	91,89	0,00	99,91	0,00		
Erro	8,11	100,00	0,09	100,00		

Tabela 8: Tabela de contingência para R = 0,99, h = 1 e 10000 realizações, comparando-se os algoritmos da ApEn e SampEn, simulado com W = 200, m = 1, r = 0,10.

Observando-se as tabelas, percebe-se que o algoritmo da SampEn possui um desempenho superior ao da ApEn, pois, em todos os casos, o primeiro acerta, sozinho, mais vezes, do que o segundo.

Dessa forma, conclui-se que, neste caso, o conjunto de parâmetros mais favorável para uso do método é: W = 150, h = 1, m = 1, r = 0.15.

5.2 Sistema 3D — Mapa de acoplamento biquadrático (sistema de Equações 4.3.4–4.3.5)

Nessa sessão, utilizando-se o mapa de acoplamento biquadrático (sistema de Equações 4.3.4–4.3.5), novamente foram utilizadas $N_R = 10000$ replicações Monte Carlo, com os parâmetros da Tabela 9, a fim de realizar uma Análise de Sensibilidade, quanto à detecção da Causalidade de Granger, no domínio da frequência.

|--|

Tabela 9: Casos simulados do mapa de acoplamento biquadrático (Sistema 3D).

Uma vez que se sabe do comportamento da Gaussianidade assintótica dos estimadores da ApEn e da SampEn, a seguinte hipótese nula (\mathcal{H}_0) [60] ainda é usada:

$$\mathcal{H}_0: \left|\pi_{ij}(f)\right|^2 = 0,$$

em que $\pi_{ij}(f)$ é o (i, j)-ésimo coeficiente da coerência parcial direcionada (PDC), na frequência normalizada f.

Entretanto, sendo f^* a frequência caracerística ($f^* = 0,1$), o critério de contagem de verdadeiros-positivos e falsos-positivos, para esse caso é:

1. $\xi_1 \rightleftharpoons \xi_2$

- se $\pi_{21}(2f^*)$ é estatisticamente diferente de zero, ou seja, $\pi_{21}(2f^*)>$

 $\gamma_{21}(2f^*)$, então há um verdadeiro-positivo, caso contrário, ocorre um falso-negativo;

- se $\pi_{12}(2f^*)$ é estatisticamente diferente de zero, ou seja, $\pi_{12}(2f^*) > \gamma_{12}(2f^*)$, então há um falso-positivo, caso contrário, ocorre um verdadeironegativo;
- 2. $\xi_1 \rightleftharpoons \xi_3$
 - − se $\pi_{31}(2f^*)$ e $\pi_{31}(4f^*)$ são estatisticamente diferentes de zero, ou seja, [$\pi_{31}(2f^*) > \gamma_{31}(2f^*)$] ∧ [$\pi_{31}(4f^*) > \gamma_{31}(4f^*)$], então há um verdadeiropositivo, caso contrário, ocorre um falso-negativo;
 - se $\pi_{21}(2f^*)$ e $\pi_{31}(4f^*)$ são estatisticamente diferentes de zero, ou seja, $[\pi_{31}(2f^*) > \gamma_{31}(2f^*)] \wedge [\pi_{31}(4f^*) > \gamma_{31}(4f^*)]$, então há um falsopositivo, caso contrário, ocorre um verdadeiro-negativo;

3. $\xi_2 \rightleftharpoons \xi_3$

- se $\pi_{32}(2f^*)$ é estatisticamente diferente de zero, ou seja, $\pi_{32}(2f^*) > \gamma_{32}(2f^*)$, então há um falso-positivo, caso contrário, ocorre um verdadeironegativo;
- se $\pi_{23}(2f^*)$ e $\pi_{23}(4f^*)$ são estatisticamente diferentes de zero, ou seja, $[\pi_{23}(2f^*) > \gamma_{23}(2f^*)] \wedge [\pi_{23}(4f^*) > \gamma_{23}(4f^*)]$, então há um falsopositivo, caso contrário, ocorre um verdadeiro-negativo;

Dessa forma, os resultados obtidos podem apreciados nas Tabelas 10 e 11.

			$\xi_1 \rightleftharpoons \xi_2$		$\xi_1 \rightleftharpoons \xi_3$		$\xi_2 \rightleftharpoons \xi_3$	
R =	0,99, h	= 1	VP	\mathbf{FP}	VP	\mathbf{FP}	$\mathrm{FP}_{2 \rightarrow 3}$	$\mathrm{FP}_{3 ightarrow 2}$
W = 200	m = 1	r = 0.15	56,11	19,20	19,68	21,94	28,75	17,08
W = 200	m = 2	T = 0,15	60,44	38,32	7,91	1,61	20,47	7,47

Tabela 10: Porcentagens de VP e FP, referentes ao algoritmo da ApEn, para os casos da Tabela 9 do mapa de acoplamento biquadrático (Sistema 3D).

			$\xi_1 \rightleftharpoons \xi_2$		$\xi_1 \rightleftharpoons \xi_3$		$\xi_2 \rightleftharpoons \xi_3$	
R =	0,99, h	= 1	VP	\mathbf{FP}	VP	FP	$\mathrm{FP}_{2 \rightarrow 3}$	$\mathrm{FP}_{3 \rightarrow 2}$
W = 200	m = 1	r = 0.15	90,32	30,63	75,32	8,74	69,74	49,81
<i>W</i> - 200	m=2	T = 0,15	72,03	21,23	24,66	$19,\!39$	34,37	17,12

Tabela 11: Porcentagens de VP e FP, referentes ao algoritmo da SampEn, para os casos da Tabela 9 do mapa de acoplamento biquadrático (Sistema 3D).

5.2.1 Sobre as PDCs entre as séries de ApEn e entre as séries de SampEn, do mapa de acoplamento biquadrático (Sistema 3D), das Equações 4.3.4–4.3.5

Pelas Tabelas 10 e 11, percebe-se um comportamento semelhente das PDCs (do mapa de acoplamento quadrático), entre as séries de ApEn e entre as de SampEn do Sistema 3D, pois com a variação de m há pouco efeito sobre as estimativas, mas é possível observar uma direção preferencial de causalidade quando a transformação pelos algoritmos informacionais é obtida apenas com o algoritmo da SampEn.

Observação: aqui são utilizados resultados estatísticos de [38].

5.2.2 Tabelas de contingência

Nessa seção seguiram-se as mesmas direções da Seção 5.1.2, para que se pudesse comparar o desempenho dos dois algoritmos (ApEn *versus* SampEn).

	$\xi_1 \overleftarrow{\leftarrow}$	ξ_2	$\xi_1 \rightleftharpoons \xi_3$		
ApEn SampEn	Acerto	Erro	Acerto	Erro	
Acerto	55,28	0,83	14,77	4,91	
Erro	42,04	98,15	60,55	80,23	

Tabela 12: Tabelas de contingência para R = 0,99, h = 1 e 10000 realizações, comparando-se os algoritmos da ApEn e SampEn, simulado com W = 200, m = 1, r = 0,15.

Novamente, como se observa pelas Tabelas 12 e 13, o algoritmo da SampEn

	$\xi_1 \rightleftharpoons \xi_2$		$\xi_1 \rightleftharpoons \xi_3$	
ApEn SampEn	Acerto	Erro	Acerto	Erro
Acerto	14,87	45,57	2,03	$5,\!88$
Erro	9,90	70,34	22,74	$30,\!65$

Tabela 13: Tabelas de contingência para R = 0,99, h = 1 e 10000 realizações, comparando-se os algoritmos da ApEn e SampEn, simulado com W = 200, m = 2, r = 0,15.

possui desempenho superior, pelos mesmos motivos apresentados na Seção 5.1.2, ou seja, nos dois casos, a SampEn acerta, sozinha, mais vezes, do que a ApEn e se conclui que, neste caso, o conjunto de parâmetros mais favorável para uso do método é: W = 200, h = 1, m = 1, r = 0.15.

5.3 Discussão dos resultados

A idéia básica, apresentada aqui, é usar as flutuações das medidas de entropia para medir como a complexidade migra de uma série para outra. As simulações de Monte Carlo evidenciam como é crítica a reconstrução do espaço de fases para que o processo tenha sucesso, dando origem ao problema de escolha ótima dos parâmetros (W, m, r), que fica mais evidente ao se trabalhar com sistema de dimensões maiores do que dois, como apresentado.

Como esperarado, o algoritmo da SampEn possui um desempenho melhor, principalmente devido ao menor viés estatístico, em relação ao algoritmo da ApEn.

Ainda, é interessante notar que a maioria das séries tranformadas pelo mapeamento por *entropia deslizante* resultaram na significante de cointegração entre as sequências de entropia¹ e, após duas diferenciações das séries (o suficente para

¹Aqui, utilizando-se tanto o procedimento de Engle-Granger [61], como o de Johansen [62–64] para fins de testes estatísticos (que é uma generalização multivariada do Teste Aumentado de Dickey-Fuller). O pacote de MATLAB utilizado para empreender tais testes foi o *Econometrics Toolbox*, desenvolvido por James P. LeSage [65], disponível em:

as deixar estacionárias²), utilizou-se o procedimento da PDC sobre essas últimas. Com isso, foram observadas taxas de causalidade de Granger compatíveis com as apresentadas aqui.

5.4 Sumário

Pelos resultados obtidos neste capítulo, observa-se que, através do processo de mapeamento em séries de entropia, é possível identificar a correta direção de causalidade, para os modelos "toys" usados.

Também é possível notar que algoritmo da SampEn possui um comportamento mais uniforme, quando comparado ao da ApEn (devido ao maior viés estatístico que este último possui) e, portanto, recomenda-se o uso do primeiro por ser mais adequado. No que se refere à conectividade, que se dá no sentido $1 \leftarrow 2$, estabelecido por hipótese pelas equações do modelo 2D, observa-se que há uma realimentação com termo quadrático. No modelo tridimensional, além da conectividade que ocorre no mesmo sentido citado anteriormente, há também no sentido $1 \leftarrow 3$, devido à realimentação biquadrática. Pela razão das realimentações serem de forma quadrática, a PDC, entre as séries originais, é impossível de se detectar (como explicado na Seção 4.3), sendo, então, necessário aplicar o mapeamento por "entropia deslizante".

Dessa forma, torna-se necessário fazer a nova transformação proposta, sohttp://www.spatial-econometrics.com/.

Os testes utilizados, também seguem de perto o Modelo de Correção de Erro Vetorial (*Vector Error Correction Model* — *VECM*), como pode ser visto com mais detalhes em [12]. Neste, é possível realizar uma forma de teste para a Causalidade de Granger, em que se verificou que os VECM para as séries de entropia possuem valores significativos de causalidade de Granger, sendo, portanto, compatíveis com os resultados apresentados.

²Para se verificar o número de vezes necessárias para se diferenciar as séries, a fim de as deixar estacionárias, utilizou-se um processo iterativo através do teste estatístico de KPSS (Kwiatkowski-Phillips-Schmidt-Shin), em que a hipótese nula é que uma série temporal é estacionária em torno de uma curva de tendência. Ou seja, tal processo iterativo se inicia calculando a significância (p) do processo ser estacionário e, até que não se rejeite a hipótese nula $(p > \alpha)$, a iteração é repetida, em que α é o erro tipo I escolhido.

6 CONCLUSÕES E TRABALHOS FUTUROS

Este trabalho elaborou uma forma alternativa de estudar o efeito de não-linearidades em séries temporais multivariadas. Para tanto, utilizaram-se conceitos como a Entropia Aproximada e a Amostral, fundamentados na Teoria da Informação.

Revisou-se o tema quanto aos algoritmos utilizados e, para validação dos mesmos, foi realizada uma análise teórica utilizando simulações de Monte Carlo. Isto permitiu concluir que tais descritores são assintoticamente normais, de acordo, também, com [31].

Em seguida, descreveu-se o uso dos algoritmos de ApEn e de SampEn, para obtenção de "séries informacionais", procedimento a que se deu o nome de **En**tropia **Deslizante** aplicado na detecção da Causalidade de Granger no domínio da frequência. Em uma primeira análise, esse novo método se mostrou atraente, pois para um sistema dinâmico caótico (mapa logístico acoplado) e outros dois sistemas não-lineares (mapas com tom modulado em frequência definida, de duas e três dimensões, respectivamente), mostrou-se apto identificar corretamente a direção do acoplamento causal. Dessa forma, para se verificar a robustez do método proposto, através de simulações de Monte Carlo, foi possível inferir a correta direção de causalidade dos sistemas não-lineares bi e tridimensionais, através de análise de sensibilidade e tabelas de contingência. Assim, conclui-se que o método de *Entropia Deslizante*, por meio dos estudos de casos apresentados em que o acoplamento causal é impossível de detectar por modelamento linear vetorial autorregressivo, aponta para a melhora da detecção da causalidade de Granger no domínio da frequência e, ainda, que não se justifica seu uso em sistemas lineares.

6.1 Trabalhos futuros

Este trabalho abre possibilidades de estudo da metodologia em casos de dimensões maiores. Sua continuidade poderia gerar o estudo do problema de obtenção ótima do espaço de fases (W, m, r), através do método de *entropia deslizante* e, também, a investigação da cointegração, ambas observadas entre as séries de entropia. Ainda no último caso, seria preciso considerar o modelo VECM (ver [12]), para teste da causalidade de Granger. Aqui, estão contemplados apenas dois casos não-lineares, em duas e três dimensões, com um tom modulado (em uma frequencia bem definida) e com realimentações quadráticas, já que esses facilitaram as análises. Porém, para aplicações em neurociências e em outras áreas biomédicas seria preciso considerar o efeito de sistemas dinâmicos mais gerais e a possibilidade da análise de um número maior de sinais simultâneos ($K \gg 3$).

Para o autor desta dissertação, como objetivo de trabalho futuro apresentamse como propostas: (1) aplicar o *Mapeamento por Entropia Deslizante* e (2) implementar métodos alternativos de mapeamento, como o uso da *complexidade de Lempel-Ziv* [66–68] (eliminando a dependência das variáveis m e r) e também da *correntropia*, que é uma função de correlação generalizada, recentemente introduzida [69,70], mostrando-se uma boa alternativa para o estudo da causalidade de Granger (no domínio do tempo) para modelos não-lineares [71], pretendendo-se extendender essa idéia ao domínio da frequência.

REFERÊNCIAS

- BACCALA, L. A.; SAMESHIMA, K. Partial directed coherence: a new concept in neural structure determination. *Biological Cybernetics*, v. 84, n. 6, p. 463–474, 2001.
- [2] PURVES, D.; FITZPATRICK, D.; AUGUSTINE, G. J. (Ed.). Neuroscience. 4th. ed. Sunderland, Massachusetts: Sinauer Associates, Incorporated, 2008.
- [3] BEAR, M. F.; CONNORS, B. W.; PARADISO, M. A. Neuroscience: exploring the brain. 3rd. ed. Baltimore, Philadelphia: Lippincott Williams & Wilkins, 2006. 857 p.
- [4] CABEZA, R.; KINGSTONE, A. Handbook of functional neuroimaging of cognition. Cambridge, Mass.; London: MIT Press, 2001.
- [5] BROCKWELL, P. J.; DAVIS, R. A. Time series: theory and methods. 2nd. ed. New York: Springer-Verlag, 1991. (Springer series in statistics).
- [6] HAMILTON, J. D. *Time series analysis*. Princeton, N.J.: Princeton University Press, 1994.
- [7] PRIESTLEY, M. B. Spectral analysis and time series. London: Academic Press, 1981.
- [8] SHUMWAY, R. H.; STOFFER, D. S. Time series analysis and its applications. New York: Springer, 2000. (Springer texts in statistics).
- [9] TSAY, R. S. Analysis of financial time series. 2nd. ed. Hoboken, N.J.: Wiley, 2005.
- [10] BENDAT, J. S.; PIERSOL, A. G. Engineering applications of correlation and spectral analysis. New York; Chichester: Wiley, 1980.
- [11] SCHUMACKER, R. E.; LOMAX, R. G. A beginner's guide to structural equation modeling. 2nd. ed. Mahwah, N.J.: Lawrence Erlbaum Associates, 2004.
- [12] LUTKEPOHL, H. New introduction to multiple time series analysis. Berlin: Springer, 2005.
- [13] APPLEBAUM, D. Probability and information: an integrated approach. 2nd. ed. Cambridge; New York: Cambridge University Press, 2008. 273 p.
- [14] COVER, T. M.; THOMAS, J. A. Elements of information theory. New York: Wiley, 2006. 774 p. (Wiley series in telecommunications).

- [15] FRANSES, P. H.; DIJK, D. v. Nonlinear time series models in empirical finance. New York: Cambridge University Press, 2000.
- [16] GENÇAY, R.; SELÇUK, F.; WHITCHER, B. An introduction to wavelets and other filtering methods in finance and economics. San Diego, California: Academic Press, 2002.
- [17] BACCALA, L. A.; SAMESHIMA, K. Overcoming the limitations of correlation analysis for many simultaneously processed neural structures. *Progress* in Brain Research, v. 130, n. Advances in Neural Population Coding, p. 33–47, 2001.
- [18] TAKAHASHI, D. Y. Medidas de fluxo de informação com aplicação em neurociência. Tese (Tese de Doutorado) — Programa Interunidades de Pósgraduação em Bioinformática da Universidade de São Paulo, São Paulo, 2009.
- [19] PINCUS, S. M.; GOLDBERGER, A. L. Physiological time-series analysis: what does regularity quantify? *American Journal of Physiology — Heart and Circulatory Physiology*, v. 266, n. 4 35-4, p. H1643–H1656, 1994.
- [20] PINCUS, S. M. Approximate entropy: a complexity measure for biological time series data. In: *Bioengineering*, *Proceedings of the Northeast Conference*. [S.l.: s.n.], 1991. p. 35–36.
- [21] LAKE, D. E. et al. Sample entropy analysis of neonatal heart rate variability. American Journal of Physiology — Regulatory, Integrative and Comparative Physiology, v. 283, p. R789–R797, September 2002.
- [22] LIU, P. Y. et al. Joint synchrony of reciprocal hormonal signaling in human paradigms of both ACTH excess and cortisol depletion. *American Journal of Physiology — Endocrinology and Metabolism*, v. 289, n. 1 52-1, p. E160–E165, 2005.
- [23] LIU, P. Y. et al. A noninvasive measure of negative-feedback strength, approximate entropy, unmasks strong diurnal variations in the regularity of LH secretion. American Journal of Physiology Endocrinology and Metabolism, v. 293, n. 5, p. E1409–E1415, 2007.
- [24] BROCKWELL, P. J.; DAVIS, R. A. Introduction to time series and forecasting. 2nd. ed. New York: Springer, 2002. (Springer texts in statistics).
- [25] PINCUS, S. M. Approximate entropy as a measure of system complexity. Proceedings of the National Academy of Sciences of the United States of America, v. 88, n. 6, p. 2297–2301, 1991.
- [26] HU, X. et al. Adaptive computation of approximate entropy and its application in integrative analysis of irregularity of heart rate variability and intracranial pressure signals. *Medical Engineering & Physics*, v. 30, n. 5, p. 631–639, 2008. ISSN 1350-4533.

- [27] LEWIS, M. J.; SHORT, A. L. Sample entropy of electrocardiographic RR and QT time-series data during rest and exercise. *Physiological Measurement*, v. 28, n. 6, p. 731–744, 2007.
- [28] LI, S. C. et al. Measurement of climate complexity using sample entropy. International Journal of Climatology, v. 26, n. 15, p. 2131–2139, 2006.
- [29] RICHMAN, J. S. Sample entropy statistics and testing for order in complex physiological signals. *Communications in Statistics — Theory and Methods*, v. 36, n. 5, p. 1005–1019, 2007.
- [30] RICHMAN, J. S.; MOORMAN, J. R. Physiological time-series analysis using approximate entropy and sample entropy. *American Journal of Physiology — Heart and Circulatory Physiology*, v. 278, n. 6, p. H2039–H2049, 2000.
- [31] RUKHIN, A. L. Approximate entropy for testing randomness. Journal of Applied Probability & Statistics, v. 37, n. 1, p. 88–100, 2000.
- [32] MASSAROPPE, L.; BACCALÁ, L. A.; SAMESHIMA, K. Semiparametric detection of nonlinear causal coupling using partial directed coherence. In: 2011 Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC). Boston: [s.n.], 2011. (Aceito para publicação).
- [33] MASSAROPPE, L.; BACCALÁ, L. A. Método semi-paramétrico para inferência de conectividade não-linear entre séries temporais. In: Congresso de Matemática Aplicada e Computacional — Região Sudeste (CMAC). Uberlândia: [s.n.], 2011. (Aceito para publicação).
- [34] GRANGER, C. W. J. Investigating causal relations by econometric models and cross-spectral methods. *Econometrica*, The Econometric Society, v. 37, n. 3, p. 424–438, 1969. ISSN 00129682.
- [35] BACCALÁ, L. A. et al. Studying the interaction between brain structures via directed coherence and Granger causality. *Applied Signal Processing*, v. 5, p. 40–48, 1998.
- [36] KAMIŃSKI, M. J.; CIESLAK-BLINOWSKA, K. J. A new method of the description of the information flow in the brain structures. *Biological Cyberne*tics, Springer Berlin / Heidelberg, v. 65, p. 203–210, 1991. ISSN 0340-1200.
- [37] TAKAHASHI, D. Y.; BACCALA, L. A.; SAMESHIMA, K. Partial directed coherence asymptotics for VAR processes of infinite order. *International Journal of Bioelectromagnetism*, v. 10, p. 31–36, 2008.
- [38] TAKAHASHI, D. Y.; BACCALA, L. A.; SAMESHIMA, K. Connectivity inference between neural structures via partial directed coherence: asymptotic results. *Journal of Applied Statistics*, v. 34, n. 10, p. 1259–1273, 2007.
- [39] SHANNON, C. E. A mathematical theory of communication. *Bell System Technical Journal*, v. 27, p. 379–423 and 623–656, July and October 1948.

- [40] HLAVÁČKOVÁ-SCHINDLER, K. et al. Causality detection based on information-theoretic approaches in time series analysis. *Physics Reports*, v. 441, n. 1, p. 1–46, March 2007. ISSN 03701573.
- [41] GRASSBERGER, P.; PROCACCIA, I. Estimation of the Kolmogorov entropy from a chaotic signals. *Physical Review A*, American Physical Society, v. 28, n. 4, p. 2591–2593, October 1983.
- [42] GREENSIDE, H. S. et al. Impracticality of a box-counting algorithm for calculating the dimensionality of strange attractors. *Physical Review A*, American Physical Society, v. 25, n. 6, p. 3453–3456, June 1982.
- [43] PINCUS, S. M.; GLADSTONE, I. M.; EHRENKRANZ, R. A. A regularity statistic for medical data analysis. *Journal of Clinical Monitoring and Computing*, v. 7, n. 4, p. 335–345, October 1991.
- [44] ECKMANN, J.-P.; RUELLE, D. Ergodic theory of chaos and strange attractors. *Reviews of Modern Physics*, American Physical Society, v. 57, n. 3, p. 617–656, July 1985.
- [45] FRASER, A. M. Information and entropy in strange attractors. *IEEE Tran*sactions on Information Theory, v. 35, n. 2, p. 245–262, March 1989. ISSN 0018-9448.
- [46] ABASOLO, D. et al. Electroencephalogram analysis with approximate entropy to help in the diagnosis of Alzheimer's disease. In: 4th International IEEE EMBS Special Topic Conference on Information Technology Applications in Biomedicine. [S.l.: s.n.], 2003. p. 222–225.
- [47] CALDIROLA, D. et al. Approximate entropy of respiratory patterns in panic disorder. The American Journal of Psychiatry, v. 161, p. 79–87, 2004.
- [48] PINCUS, S. M.; HUANG, W.-M. Approximate entropy: statistical properties and applications. *Communications in Statistics — Theory and Methods*, v. 21, n. 11, p. 3061–3077, January 1992.
- [49] RICHMAN, J. S.; LAKE, D. E.; MOORMAN, J. R. Sample entropy. In: JOHNSON, M. L.; BRAND, L. (Ed.). Methods in Enzymology — Numerical Computer Methods. [S.l.: s.n.], 2004. v. 384, p. 173–184.
- [50] PINCUS, S. M. Approximate entropy as an irregularity measure for financial data. *Econometric Reviews*, v. 27, n. 4–6, p. 329–362, 2008.
- [51] HORNERO, R. et al. Interpretation of approximate entropy: analysis of intracranial pressure approximate entropy during acute intracranial hypertension. *IEEE Transactions on Biomedical Engineering*, v. 52, n. 10, p. 1671–1680, October 2005.
- [52] MANIS, G. Fast computation of approximate entropy. Computer Methods and Programs in Biomedicine, v. 91, n. 1, p. 48–54, July 2008.

- [53] RAMDANI, S. et al. On the use of sample entropy to analyze human postural sway data. *Medical Engineering & Physics*, v. 31, n. 8, p. 1023–1031, October 2009.
- [54] ALCARAZ, R.; RIETA, J. J. A review on sample entropy applications for the non-invasive analysis of atrial fibrillation electrocardiograms. *Biomedical Signal Processing and Control*, v. 5, n. 1, p. 1–14, 2010.
- [55] PINCUS, S. M. Approximate entropy as a measure of irregularity for psychiatric serial metrics. *Bipolar Disorders*, v. 8, n. 5p1, p. 430–440, October 2006.
- [56] KAPLAN, D. T. et al. Aging and the complexity of cardiovascular dynamic. *Biophysical Journal*, v. 59, p. 945–949, April 1991.
- [57] PINCUS, S. M. Approximating markov chains. In: Proceedings of the National Academy of Science. [S.l.: s.n.], 1992. v. 89, p. 4432–4436.
- [58] DALE, R. Use of the linear-quadratic radiobiological model for quantifying kidney response in targeted radiotherapy. *Cancer Biotherapy and Radiopharmaceuticals*, v. 19, n. 3, p. 363–370, 2004.
- [59] MARINAZZO, D.; PELLICORO, M.; STRAMAGLIA, S. Kernel method for nonlinear Granger causality. *Physical Review Letters*, American Physical Society, v. 100, n. 14, p. 144103, April 2008.
- [60] BRITO, C. S. N. de et al. Asymptotic behavior of generalized partial directed coherence. In: 2010 Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC). Buenos Aires: [s.n.], 2010. p. 1718– 1721. ISSN 1557-170X.
- [61] ENGLE, R. F.; GRANGER, C. W. J. Co-Integration and error correction: representation, estimation, and testing. *Econometrica*, v. 55, n. 2, p. 251–276, March 1987.
- [62] JOHANSEN, S. Statistical analysis of cointegration vectors. Journal of Economic Dynamics and Control, v. 12, n. 2–3, p. 231–254, 1988.
- [63] JOHANSEN, S. The role of constant and linear terms in cointegration analysis of nonstationary variables. *Econometric Reviews*, v. 13, n. 2, p. 205–229, 1994.
- [64] JOHANSEN, S. Likelihood based inference in cointegrated vector autoregressive model. 2nd. ed. Oxford: Oxford University Press, 1996.
- [65] LESAGE, J. P.; PACE, R. K. Introduction to spatial econometrics. Boca Raton, London, New York: CRC Press/Taylor & Francis Group, 2009. 374 p. (Statistics: A Series of Textbooks and Monographs).
- [66] HU, J.; GAO, J.; PRÍNCIPE, J. C. Analysis of biomedical signals by the Lempel-Ziv Complexity: the effect of finite data size. *IEEE Transactions on Biomedical Engineering*, v. 53, n. 12, p. 2606–2609, December 2006. ISSN 0018-9294.
- [67] KASPAR, F.; SCHUSTER, H. G. Easily calculable measure for the complexity of spatiotemporal patterns. *Physical Review A*, American Physical Society, v. 36, n. 2, p. 842–848, July 1987.
- [68] LEMPEL, A.; ZIV, J. On the complexity of finite sequences. *IEEE Tran*sactions on Information Theory, v. 22, n. 1, p. 75–81, January 1976. ISSN 0018-9448.
- [69] SANTAMARÍA, I.; POKHAREL, P. P.; PRÍNCIPE, J. C. Generalized correlation function: definition, properties, and application to blind equalization. *IEEE Transactions on Signal Processing*, v. 54, n. 6, p. 2187–2197, June 2006. ISSN 1053-587X.
- [70] PRÍNCIPE, J. C. Information theoretic learning: Renyi's entropy and kernel perspectives. 1st edition. ed. New York: Springer Publishing Company, Incorporated, 2010. 448 p. (Information Science and Statistics, XIV).
- [71] PARK, I.; PRINCIPE, J. C. Correntropy based Granger causality. In: *IEEE International Conference on Acoustics, Speech and Signal Processing, 2008. ICASSP 2008.* [S.l.: s.n.], 2008. p. 3605–3608. ISSN 1520-6149.
- [72] TRINDADE, A. A. Modified burg algorithms for multivariate subset autoregression. Tese (Doutorado), Fort Collins, CO, USA, 2000. AAI3002104.
- [73] JR., S. L. M. Digital spectral analysis: with aplications. New Jersey: Prentice Hall, 1987. 492 p. (Prentice Hall Signal Processing Series).
- [74] ANDERSON, C. W.; STOLZ, E. A.; SHAMSUNDER, S. Multivariate autoregressive models for classification of spontaneous electroencephalographic signals during mental tasks. *IEEE Transactions on Biomedical Engineering*, v. 45, n. 3, p. 277–286, March 1998. ISSN 0018-9294.
- [75] PETERSEN, K. B.; PEDERSEN, M. S. The matrix cookbook. Technical University of Denmark, 2008. Version 2008/11/10. Disponível em: http://www2.imm.dtu.dk/pubdb/p.php?3274>.
- [76] LAUB, A. J. Matrix analysis for scientists and engineers. Philadelphia, PA, USA: Society for Industrial and Applied Mathematics, 2004. ISBN 0898715768.
- [77] HORN, R. A. Topics in matrix analysis. New York: Cambridge University Press, 1986. ISBN 0-521-30587-X.
- [78] HORN, R. A.; JOHNSON, C. R. Matrix analysis. New York: Cambridge University Press, 1990. ISBN 0521386322.

- [79] PERCIVAL, D. B.; WALDEN, A. T. Spectral analysis for physical applications: multitaper and conventional univariate techniques. New York: Cambridge University Press, 1993. 611 p.
- [80] HAYKIN, S. Adaptive filter theory. 4th. ed. New Jersey: Prentice Hall, 2002. 920 p. (Prentice Hall Information and Systems Sciences Series).
- [81] BACCALÁ, L. A.; SAMESHIMA, K. Partial directed coherence and neuronal connectivity inference. In: Proceedings of the 25th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, 2003. [S.l.: s.n.], 2003. v. 3, p. 2151. ISSN 1094-687X.
- [82] SCHELTER, B. et al. Testing for directed influences among neural signals using partial directed coherence. *Journal of Neuroscience Methods*, v. 152, n. 1-2, p. 210–219, 2006. ISSN 0165-0270.
- [83] BACCALÁ, L. A.; SAMESHIMA, K.; TAKAHASHI, D. Y. Generalized partial directed coherence. In: Proceedings of the 2007 15th International Conference on Digital Signal Processing (DSP 2007). [S.l.: s.n.], 2007. p. 163–166.
- [84] TAKAHASHI, D. Y.; BACCALÁ, L. A.; SAMESHIMA, K. Information theoretic interpretation of frequency domain connectivity measures. *Biological Cybernetics*, Springer Berlin / Heidelberg, v. 103, p. 463–469, 2010. ISSN 0340-1200.
- [85] TAKAHASHI, D. Y.; BACCALÁ, L. A.; SAMESHIMA, K. Frequency domain connectivity: an information theoretic perspective. In: 2010 Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC). Buenos Aires: [s.n.], 2010. p. 1726–1729. ISSN 1557-170X.

ANEXO A – LISTAGENS DOS CÓDIGOS IMPLEMENTADOS EM MATLAB

A.1 Entropia aproximada: apen.m

```
1 function [e] = apen(x, m, r, sflag)
2 % Input Variables
3 %
4 <sup>%</sup> X

    input data

5 % M
              - maximum template length
              - matching tolerance (d[x_{m}(i), x_{m}(j)] \leq r)
6 % r
             - flag to standardize signals (default yes / sflag = 1)
7 % sflag
  00
8
9 % Output Variables
10 %
11 % e

    approximate entropy estimate

12 \stackrel{\circ}{\sim}
13 % Reference: Steven M. Pincus and Wei-Min Huang. "Approximate entropy:
14 %
                Statistical properties and applications". Communications in
15 \frac{9}{8}
                Statistics ---- Theory and Methods, 21(11): 3061--3077, January,
16 \frac{9}{6}
                1992.
17 %
18 % Author: L.M. ---- JUL/2010
19
20 if ¬exist('m', 'var') || isempty(m), m = 5; end
21 if ¬exist('r','var') || isempty(r), r = .2*std(x); end
22 if ¬exist('sflag','var') || isempty(sflag), sflag = 1; end
23
24 x = x(:); % Same as: vec(x)
25 N = length(x); % Number of elements of the time series x
26 Nr = length(r); % Number of elements of the vector r
27
28 % Standardization of the signal
29 if sflag > 0
30
     x = (x - mean(x)) / std(x);
```

```
31 end
32
33 % Iteration for window of length m %
34 X = zeros(N - m + 1, m); \% Initialization of the buffer for length m
35 \text{ for } i = 1 : (N - m + 1)
       X(i, :) = x(i : (i + m - 1));
36
37 end
38 phim = zeros(1, Nr); % Initialization of the counter phim
  for i = 1 : (N - m + 1)
39
       cm = zeros(1, Nr); % Initialization of the counter cm
40
       for j = 1 : (N - m + 1)
41
           dmax = max(abs(X(j, :) - X(i, :)));
42
43
           cm = cm + (dmax < r);
       end
44
       phim = phim + log(cm / (N - m + 1));
45
46 end
47 phim = phim / (N - m + 1);
48
49 % Iteration for window of length m+1 %
50 X = zeros(N - m, m + 1); % Initialization of the buffer for length m+1
51 for i = 1 : (N - m)
52
       X(i, :) = x(i : (i + m));
53 end
54 phimplus1 = zeros(1, Nr); % Initialization of the counter phimplus1
  for i = 1 : (N - m)
55
       cmplus1 = zeros(1, Nr); % Initialization of the counter cmplus1
56
       for j = 1 : (N - m)
57
           dmaxplus1 = max(abs(X(j, :) - X(i, :)));
58
           cmplus1 = cmplus1 + (dmaxplus1 < r);
59
       end
60
       phimplus1 = phimplus1 + log(cmplus1 / (N - m));
61
62 end
63 phimplus1 = phimplus1 / (N - m);
64
65 e = phim - phimplus1; % Approximate entropy, log = natural lorgarithm.
66
67 % EOF %
```

A.2 Entropia amostral: sampen.m

```
1 function [e] = sampen(x, m, r, sflag)
2 %
3 % Input Variables
4 %
           - input data
5 % X
           - maximum template length
6
  % m
           - matching tolerance (d[x_{m} + 1](i), y_{m} + 1] \le r)
7 % r
8 % sflag - flag to standardize signals (default yes / sflag = 1)
9 %
10 % Output Variables
11 💡
12 % e

    – sample entropy estimate

13 %
```

```
14 % Reference: Douglas E. Lake; Joshua S. Richman; M. Pamela Griffin and
                J. Randall Moorman. "Sample entropy analysis of neonatal
15 %
                heart rate variability". American Journal of Physiology
16 %
                ---- Regulatory, Integrative and Comparative Physiology,
17 %
                283: R789-R797, September, 2002.
18 💡
19 💡
20 % Author: L.M. ---- JUL/2010
21
22 if ¬exist('m', 'var') || isempty(m), m = 5; end
23 if \negexist('r', 'var') || isempty(r), r = .2*std(x); end
24 if ¬exist('sflag','var') || isempty(sflag), sflag = 1; end
25
26 x = x(:); % Same as: vec(x)
27 N = length(x); % Number of elements of the time series x
28 Nr = length(r); % Number of elements of the vector r
29
30 % Standardization of the signal
_{31} if sflag > 0
      x = (x - mean(x)) / std(x);
32
33 end
34
35 Bm = zeros(1, Nr); % Initialization of the counters Bm and Am
36 Am = zeros(1, Nr);
37 \text{ for i} = 1 : (N - m)
      xmi(1 : m) = x(i : (i + m - 1));
38
      xmplus1i(1 : (m + 1)) = x(i : (i + m));
39
      for j = [1 : (i - 1) (i + 1) : (N - m)] % X(j=i) is not taken in count
40
           xmj(1 : m) = x(j:(j + m - 1));
41
           xmplus1j(1 : (m + 1)) = x(j : (j + m));
42
           dmaxm = max(abs(xmi - xmj));
43
           dmaxmplus1 = max(abs(xmplus1i - xmplus1j));
44
          Bm = Bm + (r > dmaxm);
45
           Am = Am + (r > dmaxmplus1);
46
       end
47
48 end
49 e = - log(Am ./ Bm); % Sample entropy, log = natural lorgarithm.
50
51 % EOF %
```

A.3 Processo MIX: mix.m

```
1 function W = mix(p, N)
2 % W = MIX(p, N)
3 %
4 % Mixture Process
5 %
6 % Input
7 % p - probability of a Bernoulli trial - real number in the range 0≤p≤1
8 % N - lenght of the process
9 %
10 % Output
11 % W - mixture process:
12 % W = (1-Z) *X+Z*Y, where Z¬Ber(p), Y¬U(-sqrt(3), sqrt(3))
```

```
13 %
                                   X=((.5)^(-.5))*sin(2*pi*j/12), for all j,
14 %
                                     sunch that j is in the range [1 N]
15 %
16 % Reference: Steven M. Pincus and Wei-Min Huang. "Approximate entropy:
                 Statistical properties and applications". Communications in
17 🖇
18 %
                 Statistics ---- Theory and Methods, 21(11): 3061--3077, January,
19
  00
                 1992.
20 %
  % Author: L.M. ---- JUL/2010
21
22
  rand('seed', sum(1e4*clock));
23
24
  alpha = sum(sin(2 * pi * (1 : 12)/12).^2) / 12;
25
26
27 i = 1 : N;
28 X = alpha<sup>(-.5)</sup> * sin((2 * pi * j) / 12);
29
30
  Y = sqrt(3) + (-sqrt(3) - sqrt(3)) * rand(1, N);
31
_{32} t = rand(1, N);
33
_{34} um = t < p;
35
_{36} zero = t > p;
37
_{38} Z (um) = 1;
39
40 \ Z(zero) = 0;
41
42 W = (1 - Z) \cdot X + Z \cdot Y;
```

A.4 Processo Gaussian Markov Process (GMP): gmp.m

```
1 function X = gmp(p, N)
2 \% X = GMP(p, N)
3 %
4
  % Gaussian Markov Process
  8
5
6 % Input
7 % p - real number in the range 0 \le p \le 1
8 % N − lenght of the process
9 %
10 % Output
  % W - Gaussian Markov process:
11
        X(i) = p * X(i-1) + (1-p^2) * Z(i), where Z \neg N(0, 1)
12
  00
13
  8
14 % Reference: Steven M. Pincus and Wei-Min Huang. "Approximate entropy:
15 \frac{9}{8}
                 Statistical properties and applications". Communications in
                 Statistics ---- Theory and Methods, 21(11): 3061--3077, January,
16 %
1992.
18 %
```

```
19 % Author: L.M. --- JUL/2010
20
21 randn('seed', sum(le4*clock));
22
23 Z = randn(1, N);
24
25 X = zeros(1, N);
26
27 X(1) = Z(1);
28
29 for i = 2 : N
30 X(i) = p * X(i - 1) + (1 - p^2) * Z(i);
31 end
```

A.5 Janelamento de séries temporais: tswin.m

```
1 function xw = tswin(x, W, h, nPoints, nDiscard)
2 % xw = TSWIN(x, W, h, nPoints, nDiscard)
3 %
4 % Time Series Windowing
5 %
6 % Input
           x - Multivariate time series
7 %
           W - Window length
8 %
      h — Step size
9 %
10 % nPoints - Number of analyzed samples points
11 % nDiscard - Number of points discarded at beginning of simulated series
12 %
13 % Output
     xw — Windowed multivariate time series
14 🖇
15 \frac{8}{6}
16 % Example of usage: [xw] = window(x,200,1,1501,1000);
                       => if size(x) = [10000 2], then size(xw) = [200 1300 2],
17 💡
                        where 2 = number of chanels, 200 = W and
18 %
                        1301 = nPoints - W + h.
19 %
20 %
21 % Author: L.M. ---- FEB/2011
22
23
24 [nrows ncols] = size(x);
25
26 if nrows < ncols
     y = x'; % Data must be organized column-wise
27
      nChannels = nrows;
28
29 else
      nChannels = ncols;
30
      y = x;
31
32 end
33
34 u = y((nDiscard + 1) : (nDiscard + nPoints), :);
35
36 x_w = zeros(floor(W/h), nPoints - W + h, nChannels);
37
```

ANEXO B – O ALGORITMO DE LEVINSON MULTIVARIADO

O foco deste apêndice está na revisão da maneira de se estimar os coeficientes de processos vetoriais (ou multivariados) autorregressivos (VAR), baseado nas equações de Yule-Walker multivariadas e na respectiva generalização do algoritmo de Levinson, que é utilizado para resolução recursiva, na ordem e no tempo de tais equações, em uma estrutura robusta: a treliça.

Para processos multivariados autorregressivos e de médias móveis (VARMA) e apenas de médias móveis (VMA), pouca atenção tem-se dado, pois há pouca necessidade de se estimar os parâmetros de tais processos e os procedimentos se mostram difíceis.

B.1 Processos autorregressivos vetoriais — VAR

Os processos autorregressivos vetoriais, de ordem p (VAR(p)) são definidos de acordo com a seguinte recursão:

$$\mathbf{x}(n) = \sum_{r=1}^{p} \mathbf{A}(r)\mathbf{x}(n-r) + \mathbf{w}(n), \qquad (B.1.1)$$

em que, $\mathbf{x}(n)$ é um vetor K-dimensional de séries temporais e $\mathbf{w}(n)$ um processo de inovação Gaussiano, de média zero e matriz de variância-covariância positiva definida, $\Sigma_{\mathbf{w}} = \mathbb{E} \left[\mathbf{w}(n) \mathbf{w}^{\mathrm{H}}(n) \right]$. Já $\mathbf{A}(r) \in \mathbb{C}^{K \times K}$ são as matrizes dos coeficientes autor regressivos a serem estimadas, a partir do processo $\{\mathbf{x}(n)\}_{n=1}^{N}$ e p é a ordem do VAR.

B.2 Equações de Yule-Walker multivariadas

B.2.1 Predição progressiva

Considere o processo vetorial $\{\mathbf{x}(n)\}_{n\in\mathbb{Z}}$ de média zero, fracamente estacionário, com função de correlação, $\Gamma(h) = \mathbb{E} \left[\mathbf{x}(n+h)\mathbf{x}^{\mathrm{H}}(n)\right], h \in \mathbb{Z}$. Nesse processo o melhor preditor linear **progressivo** de $\mathbf{x}(n)$ é [72]:

$$\hat{\mathbf{x}}_p^f(n) = \sum_{r=1}^p \mathbf{A}_p(r) \mathbf{x}(n-r).$$
(B.2.1)

Assim, o resíduo de predição progressivo será [73,74]:

$$\mathbf{e}_p^f(n) = \mathbf{x}(n) - \hat{\mathbf{x}}_p^f(n) = \mathbf{a}_p \cdot \mathbf{x}_p(n), \qquad (B.2.2)$$

em que:

$$\mathbf{a}_{p} = \begin{bmatrix} \mathbf{I}_{K} & \mathbf{A}_{p}(1) & \cdots & \mathbf{A}_{p}(p) \end{bmatrix}$$
(B.2.3)
$$\mathbf{x}_{p}(n) = \begin{bmatrix} \mathbf{x}(n) \\ \vdots \\ \mathbf{x}(n-p) \end{bmatrix},$$

Para estabelecer uma relação entre as matrizes dos coeficientes do VAR(p) e a função de correlação, basta multiplicar ambos os lados da Equação B.2.2 por $\mathbf{x}_p^{\mathrm{H}}(n)$, à direita e aplicar o operador esperança, obtendo:

$$\mathbb{E}\left[\mathbf{e}_{p}^{f}(n)\mathbf{x}_{p}^{\mathrm{H}}(n)\right] = \mathbb{E}\left[\mathbf{a}_{p}\cdot\mathbf{x}_{p}(n)\mathbf{x}_{p}^{\mathrm{H}}(n)\right] = \mathbf{a}_{p}\mathbb{E}\left[\mathbf{x}_{p}(n)\mathbf{x}_{p}^{\mathrm{H}}(n)\right] = \mathbf{a}_{p}\boldsymbol{\Sigma}_{p}.$$
 (B.2.4)

Seja, agora, a matriz $\Sigma_p \in \mathbb{C}^{(p+1) \times (p+1)}$, em que os blocos são dados por $\Gamma(n) \in$

 $\mathbb{C}^{K \times K}$:

$$\begin{split} \boldsymbol{\Sigma}_{p} &= \mathbb{E} \left[\mathbf{x}_{p}(n) \mathbf{x}_{p}^{\mathrm{H}}(n) \right] \\ &= \begin{bmatrix} \mathbb{E} \left[\mathbf{x}(n) \mathbf{x}^{\mathrm{H}}(n) \right] & \mathbb{E} \left[\mathbf{x}(n) \mathbf{x}^{\mathrm{H}}(n-1) \right] & \cdots & \mathbb{E} \left[\mathbf{x}(n) \mathbf{x}^{\mathrm{H}}(n-p) \right] \\ \mathbb{E} \left[\mathbf{x}(n-1) \mathbf{x}^{\mathrm{H}}(n) \right] & \mathbb{E} \left[\mathbf{x}(n-1) \mathbf{x}^{\mathrm{H}}(n-1) \right] & \cdots & \mathbb{E} \left[\mathbf{x}(n) \mathbf{x}^{\mathrm{H}}(n-1) \right] \\ \vdots & \vdots & \ddots & \vdots \\ \mathbb{E} \left[\mathbf{x}(n-p) \mathbf{x}^{\mathrm{H}}(n) \right] & \mathbb{E} \left[\mathbf{x}(n-p) \mathbf{x}^{\mathrm{H}}(n-1) \right] & \cdots & \mathbb{E} \left[\mathbf{x}(n-p) \mathbf{x}^{\mathrm{H}}(n-p) \right] \end{bmatrix} \\ &= \begin{bmatrix} \mathbf{\Gamma}(0) & \mathbf{\Gamma}(1) & \cdots & \mathbf{\Gamma}(p) \\ \mathbf{\Gamma}(-1) & \mathbf{\Gamma}(0) & \cdots & \mathbf{\Gamma}(p-1) \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{\Gamma}(-p) & \mathbf{\Gamma}(-p+1) & \cdots & \mathbf{\Gamma}(0) \end{bmatrix} . \end{split}$$
(B.2.5)

Repare que a matriz Σ_p possui ambas estruturas *Hermitiana* e *Bloco-Toeplitz*. Ótimas referências para essas formas de matrizes podem ser encontras em [75–78]. Os blocos de Σ_p não possuem, em geral, simetria Hermitiana ($\Gamma(\varepsilon) \neq \Gamma(-\varepsilon)$), entretanto, tem-se que $\Gamma(-\varepsilon) = \Gamma^{\rm H}(\varepsilon)$.

Ainda, da Equação B.2.4, o vetor em bloco $\mathbb{E}[\mathbf{e}_p^f(n)\mathbf{x}_p^{\mathrm{H}}(n)]$, pode ser simplificado para:

$$\mathbb{E}\left[\mathbf{e}_{p}^{f}(n)\mathbf{x}_{p}^{\mathrm{H}}(n)\right] = \mathbb{E}\left[\mathbf{e}_{p}^{f}(n)\left(\mathbf{e}_{p}^{f}^{\mathrm{H}}(n) - \sum_{r=1}^{p}\mathbf{x}^{\mathrm{H}}(n-r)\mathbf{A}_{p}^{\mathrm{H}}(r)\right)\right]$$
$$= \mathbb{E}\left[\mathbf{e}_{p}^{f}(n)\mathbf{e}_{p}^{f}^{\mathrm{H}}(n) - \sum_{r=1}^{p}\underbrace{\mathbf{e}_{p}^{f}(n)\mathbf{x}^{\mathrm{H}}(n-r)}_{=0,n< r}\mathbf{A}_{p}^{\mathrm{H}}(r)\right], \quad (B.2.6)$$
$$=\left[\boldsymbol{\Sigma}_{p}^{f} \quad \mathbf{O}_{K\times K} \quad \cdots \quad \mathbf{O}_{K\times K}\right]$$

em que:

$$\boldsymbol{\Sigma}_{p}^{f} = \mathbb{E}\left[\mathbf{e}_{p}^{f}(n)\mathbf{e}_{p}^{f \mathrm{H}}(n)\right] = \boldsymbol{\Sigma}_{\mathbf{w}},\tag{B.2.7}$$

ou seja, a matriz de covariância do resíduo de predição progressivo é a mesma do processo de inovação de entrada. Vale lembrar que a última igualdade da Equação B.2.6 decorre da propriedade de que os resíduos são descorrelacionados com valores passados do processo VAR, para n < r. Dessa forma, a Equação B.2.4 pode ser rescrita como:

$$\mathbf{a}_p \mathbf{\Sigma}_p = \begin{bmatrix} \mathbf{\Sigma}_p^f & \mathbf{O}_{K \times K} & \cdots & \mathbf{O}_{K \times K} \end{bmatrix}, \qquad (B.2.8)$$

que é conhecida como a versão multivariada das equações normais de Yule-Walker, para o processo VAR(p), obtido a partir do filtro de predição progressiva multivariada de ordem p.

B.2.2 Predição regressiva

Analogamente à Seção B.2.1, o resíduo do processo VAR(p), advindo a partir do filtro de predição regressivo, no índice temporal n - 1, é:

$$\mathbf{e}_p^b(n-1) = \mathbf{x}(n-p-1) - \hat{\mathbf{x}}_p^b(n-1)$$
$$= \mathbf{x}(n-p-1) - \left(\sum_{r=1}^p \mathbf{B}_p(r)\mathbf{x}(n-p-1+r)\right), \quad (B.2.9)$$
$$= \mathbf{b}_p \cdot \mathbf{x}_p(n-1)$$

em que \mathbf{b}_p é o vetor linha (em bloco) dos parâmetros do filtro de predição regressiva multivariado:

$$\mathbf{b}_p = \begin{bmatrix} \mathbf{B}_p(p) & \cdots & \mathbf{B}_p(1) & \mathbf{I}_K \end{bmatrix}$$
(B.2.10)

Portanto, a partir da Equação B.2.9 é possível deduzir as equações normais multivariadas de Yule-Walker para o filtro de predição regressivo multivariado:

$$\mathbb{E}\left[\mathbf{e}_{p}^{b}(n-1)\hat{\mathbf{x}}_{p}^{b}^{\mathrm{H}}(n-1)\right] = \mathbb{E}\left[\mathbf{b}_{p}\cdot\hat{\mathbf{x}}_{p}^{b}^{\mathrm{H}}(n-1)\mathbf{x}_{p}(n-1)\right], \qquad (B.2.11)$$

chegando a:

$$\mathbf{b}_p \boldsymbol{\Sigma}_p = \begin{bmatrix} \mathbf{O}_{K \times K} & \mathbf{O}_{K \times K} & \cdots & \boldsymbol{\Sigma}_p^b \end{bmatrix}, \qquad (B.2.12)$$

em que:

$$\Sigma_p^b = \mathbb{E}\left[\mathbf{e}_p^b(n-1)\mathbf{e}_p^{b \ \mathrm{H}}(n-1)\right],\tag{B.2.13}$$

ou seja, a matriz de covariância do resíduo de predição regressivo é a mesma do processo de inovação de entrada. Note, no entanto, que, em geral, as matrizes $\Sigma_p^f \in \Sigma_p^b$ são diferentes, para processos vetoriais.

B.2.3 Equações de Yule-Walker dos filtros de predição progressiva e regressiva

Resumindo, as equações necessárias para se encontrar os parâmetros de um processo VAR(p), a partir dos respectivos filtros de predição progressiva e regressiva são:

$$\mathbf{a}_p \boldsymbol{\Sigma}_p = \begin{bmatrix} \boldsymbol{\Sigma}_p^f & \mathbf{O}_{K \times K} & \cdots & \mathbf{O}_{K \times K} \end{bmatrix}$$
(B.2.14a)

$$\mathbf{b}_p \boldsymbol{\Sigma}_p = \begin{bmatrix} \mathbf{O}_{K \times K} & \cdots & \mathbf{O}_{K \times K} & \boldsymbol{\Sigma}_p^f \end{bmatrix}, \quad (B.2.14b)$$

ou, em forma matricial,

$$\boldsymbol{\Sigma}_{p}^{\mathrm{H}} \begin{bmatrix} \mathbf{a}_{p}^{\mathrm{H}} & \mathbf{b}_{p}^{\mathrm{H}} \end{bmatrix} = \begin{bmatrix} \boldsymbol{\Sigma}_{p}^{f \ \mathrm{H}} & \mathbf{O}_{K \times K}^{\top} \\ \mathbf{O}_{K \times K}^{\top} & \vdots \\ \vdots & \mathbf{O}_{K \times K}^{\top} \\ \mathbf{O}_{K \times K}^{\top} & \boldsymbol{\Sigma}_{p}^{b \ \mathrm{H}} \end{bmatrix}, \qquad (B.2.15)$$

em que \mathbf{a}_p , \mathbf{b}_p , $\boldsymbol{\Sigma}_p$, $\boldsymbol{\Sigma}_p^f$ e $\boldsymbol{\Sigma}_p^b$ são dados pelas Equações B.2.3, B.2.10, B.2.5, B.2.7, B.2.13, respectivamente.

A seguir, na Seção B.2.4, descreve-se como resolver, recursivamente, na ordem e no tempo, o sistema de Equações B.2.14.

B.2.4 Algoritmo (generalizado) de Levinson multivariado

Nessa seção, considera-se uma das generalizações do algoritmo de Levinson para o caso multivariado. Serão tratadas, aqui, as recursões desenvolvidas por Wiggins e Robinson (1965), que são baseadas nas mesmas idéias do caso univariado [5, 12, 73, 79, 80].

B.2.4.1 Apresentação das equações

A matriz de correlação, bloco-Toeplitz, pode ser particionada de duas formas:

$$\boldsymbol{\Sigma}_{p+1} = \begin{bmatrix} \boldsymbol{\Sigma}_{p} & \boldsymbol{\psi}_{p+1}^{\mathrm{H}} \\ \boldsymbol{\psi}_{p+1} & \boldsymbol{\Gamma}(0) \end{bmatrix} = \begin{bmatrix} \boldsymbol{\Gamma}(0) & \boldsymbol{\sigma}_{p+1} \\ \boldsymbol{\sigma}_{p+1}^{\mathrm{H}} & \boldsymbol{\Sigma}_{p} \end{bmatrix}, \quad (B.2.16)$$

em que:

$$\boldsymbol{\sigma}_{p+1} = \mathbb{E}\left[\mathbf{x}_p(n-1)\mathbf{x}^{\mathrm{H}}(n)\right] = \begin{bmatrix} \boldsymbol{\Gamma}(1) & \cdots & \boldsymbol{\Gamma}(p+1) \end{bmatrix}$$
 (B.2.17a)

$$\boldsymbol{\psi}_{p+1} = \mathbb{E}\left[\mathbf{x}_p(n)\mathbf{x}^{\mathrm{H}}(n-1)\right] = \left[\boldsymbol{\Gamma}[-(p+1)] \cdots \boldsymbol{\Gamma}(-1)\right].$$
 (B.2.17b)

Com as Equações B.2.16 e B.2.17, é possível reescrever o sistema de Equações B.2.14, da seguinte maneira:

$$\begin{bmatrix} \mathbf{a}_p & \mathbf{O}_{K \times K} \end{bmatrix} \boldsymbol{\Sigma}_{p+1} = \begin{bmatrix} \boldsymbol{\Sigma}_p^f & \mathbf{O}_{K \times K} & \cdots & \mathbf{O}_{K \times K} & \boldsymbol{\Delta}_{p+1} \end{bmatrix}$$
(B.2.18a)

$$\begin{bmatrix} \mathbf{O}_{K \times K} & \mathbf{b}_p \end{bmatrix} \boldsymbol{\Sigma}_{p+1} = \begin{bmatrix} \boldsymbol{\nabla}_{p+1} & \mathbf{O}_{K \times K} & \cdots & \mathbf{O}_{K \times K} & \boldsymbol{\Sigma}_p^b \end{bmatrix}, \quad (B.2.18b)$$

ou, em forma matricial,

$$\boldsymbol{\Sigma}_{p+1}^{\mathrm{H}} \begin{bmatrix} \begin{bmatrix} \mathbf{a}_{p} & \mathbf{O}_{K \times K} \end{bmatrix}^{\mathrm{H}} & \begin{bmatrix} \mathbf{O}_{K \times K} & \mathbf{b}_{p} \end{bmatrix}^{\mathrm{H}} \end{bmatrix} = \begin{bmatrix} \boldsymbol{\Sigma}_{p}^{f \ \mathrm{H}} & \boldsymbol{\nabla}_{p+1}^{\mathrm{H}} \\ \mathbf{O}_{K \times K}^{\top} & \mathbf{O}_{K \times K}^{\top} \\ \vdots & \vdots \\ \mathbf{O}_{K \times K}^{\top} & \mathbf{O}_{K \times K}^{\top} \\ \boldsymbol{\Delta}_{p+1}^{\mathrm{H}} & \boldsymbol{\Sigma}_{p}^{b \ \mathrm{H}} \end{bmatrix}, \quad (B.2.19)$$

em que as matrizes Δ_{p+1} e ∇_{p+1} (ambas no espaço $\mathbb{C}^{K \times K}$) são definidas como:

$$\boldsymbol{\Delta}_{p+1} = \boldsymbol{\Gamma}(p+1) + \sum_{r=1}^{p} \mathbf{A}_{p}(n) \boldsymbol{\Gamma}(p+1-n) = \mathbf{a}_{p} \boldsymbol{\psi}_{p+1}^{\mathrm{H}}$$
(B.2.20a)

$$\boldsymbol{\nabla}_{p+1} = \boldsymbol{\Gamma}(-p-1) + \sum_{r=1}^{p} \mathbf{B}_{p}(n) \boldsymbol{\Gamma}(-p-1+n) = \mathbf{b}_{p} \boldsymbol{\sigma}_{p+1}^{\mathrm{H}}.$$
 (B.2.20b)

As equações de recursão, na ordem, das matrizes dos coeficientes autorregressivos, de predição progressiva e regressiva, são:

$$\mathbf{a}_{p+1} = \begin{bmatrix} \mathbf{a}_p & \mathbf{O}_{K \times K} \end{bmatrix} + \mathbf{A}_{p+1}(p+1) \begin{bmatrix} \mathbf{O}_{K \times K} & \mathbf{b}_p \end{bmatrix}$$
(B.2.21a)

$$\mathbf{b}_{p+1} = \begin{bmatrix} \mathbf{O}_{K \times K} & \mathbf{b}_p \end{bmatrix} + \mathbf{B}_{p+1}(p+1) \begin{bmatrix} \mathbf{a}_p & \mathbf{O}_{K \times K} \end{bmatrix}.$$
(B.2.21b)

Portanto, para uma matriz particionada em blocos, o filtro de predição linear, com estrutura matricial, deve obedecer às seguintes recursões (na ordem e no tempo):

$$\mathbf{A}_{p+1}(n) = \mathbf{A}_p(n) + \mathbf{A}_{p+1}(p+1)\mathbf{B}_{p+1}(p+1-n)$$
(B.2.22a)

$$\mathbf{B}_{p+1}(n) = \mathbf{B}_p(n) + \mathbf{B}_{p+1}(p+1)\mathbf{A}_{p+1}(p+1-n),$$
(B.2.22b)

para $n = 1, \dots, p$. Dessa forma, há duas matrizes de coeficientes de reflexão, dadas por:

$$\mathbf{A}_{p+1}(p+1) = -\mathbf{\Delta}_{p+1} \left(\mathbf{\Gamma}_p^b\right)^{-1}$$
(B.2.23a)

$$\mathbf{B}_{p+1}(p+1) = -\nabla_{p+1} \left(\Gamma_p^f \right)^{-1}.$$
 (B.2.23b)

A fim de verificar as igualdades das Equações B.2.21a e B.2.21b, basta multiplicar, à esquerda, ambos os lados das equações, por Γ_{p+1} e, após essa operação, reconhecer e substituir as Equações B.2.18a, B.2.18b, B.2.23a e B.2.23b.

Na etapa final dessa verificação, extrai-se a importante recursão da atualização da matriz de covariância dos resíduos de predição progressivo e regressivo, respectivamente:

$$\Gamma_{p+1}^{f} = \Gamma_{p}^{f} + \mathbf{A}_{p+1}(p+1)\nabla_{p+1} = [\mathbf{I}_{K} - \mathbf{A}_{p+1}(p+1)\mathbf{B}_{p+1}(p+1)]\Gamma_{p}^{f} \quad (B.2.24a)$$

$$\Gamma_{p+1}^{b} = \Gamma_{p}^{b} + \mathbf{B}_{p+1}(p+1)\mathbf{\Delta}_{p+1} = \left[\mathbf{I}_{K} - \mathbf{B}_{p+1}(p+1)\mathbf{A}_{p+1}(p+1)\right]\Gamma_{p}^{b}.$$
 (B.2.24b)

Utilizando, agora, a identidade matricial [73,75]:

$$\begin{bmatrix} \mathbf{a}_p & \mathbf{O}_{K \times K} \end{bmatrix} \boldsymbol{\Sigma}_{p+1} \begin{bmatrix} \mathbf{O}_{K \times K} \\ \mathbf{b}_p^{\mathrm{H}} \end{bmatrix} = \begin{bmatrix} \mathbf{O}_{K \times K} & \mathbf{b}_p \end{bmatrix} \boldsymbol{\Sigma}_{p+1} \begin{bmatrix} \mathbf{a}_p^{\mathrm{H}} \\ \mathbf{O}_{K \times K} \end{bmatrix}, \quad (B.2.25)$$

é possível verificar que:

$$\boldsymbol{\nabla}_{p+1} = \boldsymbol{\Delta}_{p+1}^{\mathrm{H}}.\tag{B.2.26}$$

A matriz Δ_{p+1} é a correlação cruzada entre os resíduos de predição progressiva e regressiva, com uma unidade de atraso:

$$\boldsymbol{\Delta}_{p+1} = \mathbb{E}\left[\mathbf{e}_p^f(n)\mathbf{e}_p^b(n-1)\right] = \boldsymbol{\Sigma}_p^{fb}$$
(B.2.27)

Para verificar essa relação, basta tomar as Equações B.2.17b, B.2.20a e notar que $\mathbf{x}(n-1) = \mathbf{b}_p \cdot \mathbf{x}_p(n) + \mathbf{w}(n).$

Cabe, agora, definir a matriz de correlações parciais normalizadas, $\Lambda_{p+1} \in \mathbb{C}^{K \times K}$:

Observações [77, 78]:

- O operador (·)^{−H} denota a matriz conjugada transposta da inversa e, a operação, ocorre na seguinte ordem: a primeira é a inversa da matriz e a segunda é a respectiva transposição da conjugada;
- 2. O operador $(\cdot)^{1/2}$ denota a matriz triangular superior obtida pela Fatoração

de Cholesky de uma matriz. Assim, se **M** é uma matriz Hermitiana, então:

$$\mathbf{M} = \left(\mathbf{M}^{1/2}\right) \left(\mathbf{M}^{1/2}\right)^{\mathrm{H}}.$$

Portanto, com essas considerações, é possível reescrever as Equações B.2.23a e B.2.23b, da seguinte maneira:

$$\mathbf{A}_{p+1}(p+1) = -\left(\boldsymbol{\Sigma}_{p}^{f\ 1/2}\right) \boldsymbol{\Lambda}_{p+1}\left(\boldsymbol{\Sigma}_{p}^{b\ 1/2}\right)^{-1}$$
(B.2.29a)

$$\mathbf{B}_{p+1}(p+1) = -\left(\Sigma_{p}^{b\ 1/2}\right) \Lambda_{p+1}^{\mathrm{H}} \left(\Sigma_{p}^{f\ 1/2}\right)^{-1}.$$
 (B.2.29b)

B.2.5 O algoritmo

Apresentadas as equações anteriores, o algoritmo de Levinson multivariado constitui em dois possíveis conjuntos de Equações, com condições iniciais:

Equações	Condições Iniciais
B.2.14a	
B.2.21a	
B.2.21b	
B.2.22a B.2.28	$\Sigma_0^f = \Sigma_0^b = \Gamma(0)$
B.2.22b B.2.29b	
B.2.23a	
B.2.23b	

Tabela 14: Equações que fazem parte do algoritmo de Levinson multivariado e condições iniciais do mesmo.

A natureza recursiva na ordem do algoritmo em questão se restringe na questão de que a matriz Σ_p deve ser positiva definida e não-singular, para todas as ordens, ou seja, para $r = 1, \dots, p$. Essa afirmação é verdadeira enquanto det $|\Sigma_r^f| > 0$ e det $|\Sigma_r^b| > 0$, para $r = 1, \dots, p$.

B.2.6 A estrutura em treliça

Como se sabe, o algoritmo de Levinson possui uma estrutura em treliça, tornando-o muito robusto para estimação dos coeficientes do processo AR(p). Para o caso vetorial não é diferente. A fim de verificar essa importante propriedade, basta multiplicar à esquerda, ambos os lados das igualdades das Equações B.2.21a e B.2.21b por $\mathbf{x}_p(n)$, reconhecendo que $\mathbf{e}_p^f(n) = \mathbf{a}_p \cdot \mathbf{x}_p(n)$ e $\mathbf{e}_p^b(n-1) = \mathbf{b}_p \cdot \mathbf{x}_p(n-1)$, chegando a:

$$\mathbf{e}_{p+1}^{f}(n) = \mathbf{e}_{p}^{f}(n) + \mathbf{A}_{p+1}(p+1)\mathbf{e}_{p}^{b}(n-1)$$
 (B.2.30a)

$$\mathbf{e}_{p+1}^{b}(n) = \mathbf{e}_{p}^{b}(n-1) + \mathbf{B}_{p+1}(p+1)\mathbf{e}_{p}^{f}(n),$$
 (B.2.30b)

ou, ainda,

$$\begin{bmatrix} \mathbf{e}_{p+1}^{f}(n) \\ \mathbf{e}_{p+1}^{b}(n) \end{bmatrix} = \begin{bmatrix} \mathbf{A}_{p+1}(p+1) & \mathbf{I}_{K} \\ \mathbf{I}_{K} & \mathbf{B}_{p+1}(p+1) \end{bmatrix} \begin{bmatrix} \mathbf{e}_{p}^{b}(n-1) \\ \mathbf{e}_{p}^{f}(n) \end{bmatrix}.$$
 (B.2.31)

B.3 Critérios de seleção da ordem do VAR

Embora a aplicação de testes estatísticos para avaliar a melhor ordem do VAR seja muito utilizada, aqueles não são totalmente satisfatórios para propósitos específicos, se um modelo específico é necessário. Ou seja, esta última abordagem pode ser muito útil para escolha de modelos, a fim de se obter uma boa previsão do evento em estudo. Dessa forma, torna-se útil quando é incluída na análise a seleção da ordem do VAR.

Os critérios listados nos Itens 1, 2, 3 e 4 tentam evitar sobreajuste, minimizando a variância da série de resíduos, com a penalidade do número de parâmetros usados. Em [12] é possível encontrar diversas propriedades dos estimadores, como, por exemplo, consistência, comparação entre eles e outras.

A fim de apresentar as equações dos critérios, considere as variáveis $m, p, \hat{\Sigma}_{\mathbf{w}}, T, e K$, como sendo, respectivamente: ordem estimada, intervalo de valores possíveis para a estimativa da ordem, estimativa da matriz de covariância da série de resíduos, comprimento da série temporal e número de canais. Dessa maneira,

os critérios mais utilizados são:

1. Critério de Informação de Akaike (Akaike Information Criterion -AIC):

$$\min_{m \in p} \left\{ \operatorname{AIC}(m) \right\} = \ln \left\{ \det \left[\hat{\boldsymbol{\Sigma}}_{\mathbf{w}}(m) \right] \right\} + \frac{2}{T} m K^2;$$

2. Critério de informação Bayesiano de Schwartz (Schwartz's Bayesian Information Criterion):

$$\min_{m \in p} \left\{ \text{BIC}(m) \right\} = \ln \left\{ \det \left[\hat{\boldsymbol{\Sigma}}_{\mathbf{w}}(m) \right] \right\} + \frac{\ln(T)}{T} m K^2;$$

3. Critério do Erro de Predição Final (*Final Prediction Error Criterion* — *FPE*):

$$\min_{m \in p} \left\{ \text{FPE}(m) \right\} = \left[\frac{T + Km + 1}{T - Km - 1} \right]^K \det \left[\hat{\Sigma}_{\mathbf{w}}(m) \right];$$

4. Critério de Hannan-Quinn (HQ):

$$\min_{m \in p} \left\{ \mathrm{HQ}(m) \right\} = \ln \left\{ \det \left[\hat{\boldsymbol{\Sigma}}_{\mathbf{w}}(m) \right] \right\} + \frac{2 \ln[\ln(T)]}{T} m K^2.$$

ANEXO C – COERÊNCIA PARCIAL DIRECIONADA

Este apêndice trata de abordagem muito utilizada no domínio da frequência para descrever causalidade entre séries temporais multivariadas.

O conceito, ao qual é dado o nome de Coerência Parcial Direcionada (*Partial Directed Coherence — PDC*), mostra ser uma interpretação da Causalidade de Granger no domínio da frequência e, também, possui a capacidade de transmitir a idéia de como duas ou mais estruturas estão conectadas funcionalmente e se isto ocorre [1].

C.1 Medindo a PDC

C.1.1 A causalidade de Granger

Formalmente, define-se a Causalidade de Granger como:

Definição C.1.1. Diz-se que uma série temporal $x_1(n)$ causa $x_2(n)$ no sentido de Granger [34] se:

$$\mathbf{MSE} [x_1(n) \mid \{x_1(n_-), x_2(n_-)\}] < \mathbf{MSE} [x_1(n) \mid \{x_1(n_-)\}],$$

em que **MSE** $[x(n) | \{\cdot\}]$ é o erro quadrático médio (*Mean Square Error* — *MSE*) de predição, dado o conjunto de informação que compreende exclusivamente as observações passadas $\{x_1(n_-)\}$ de $x_1(n)$, ou conjuntamente de $x_1(n)$ e $x_2(n)$, via $\{x_1(n_-), x_2(n_-)\}$. Na prática, há diversas formas de operacionalizar a inferência da causalidade, a maioria das quais envolve o ajuste de modelos de predição lineares [12], usando, por exemplo, modelos autorregressivos multivariados (ou vetoriais):

$$\mathbf{x}(n) = \sum_{r} \mathbf{A}(r)\mathbf{x}(n-r) + \mathbf{w}(n), \qquad (C.1.1)$$

em que, $\mathbf{x}(n)$ é um vetor K-dimensional de séries temporais e $\mathbf{w}(n)$ um processo de inovação, tal que:

$$\mathbf{A}(r) = \begin{bmatrix} a_{11}(r) & a_{12}(r) & \cdots & a_{1K}(r) \\ a_{21}(r) & a_{22}(r) & \cdots & a_{2K}(r) \\ \vdots & \vdots & \ddots & \vdots \\ a_{K1}(r) & a_{K2}(r) & \cdots & a_{KK}(r) \end{bmatrix}$$

tem os coeficientes $a_{ij}(r)$, responsáveis por medir o efeito da interação linear de $x_j(n-r)$ em $x_i(n)$, de sorte que a forma mais comum de aferir a existência de causalidade da série $x_j(n)$, para a série $x_i(n)$, consiste em examinar a nulidade desses coeficientes para todo r.

Testes estatísticos específicos podem ser apreciados em [12], para validar a causalidade de Granger.

Conceitualmente, a causalidade de Granger é advinda do contexto da econometria, em que é possível apreciá-la através de diferentes testes estatísticos [12]. Sua principal idéia é considerar **exclusivamente** amostras passadas na melhoria da predição. Quando esta se dá por **amostra presente**, do sinal em estudo, é usual dizer que há *causalidade de Granger instantânea*. Assim, é importante ter em mente essas formas distintas de causalidade, no que se refere à previsibilidade de séries temporais:

 a primeira se dá pelo exclusivo uso do passado de séries temporais, para ajudar a prever outra; a segunda já se utiliza do efeito do presente de uma série temporal sobre outra.

Dessa forma, a PDC foi introduzida para se inferir a causalidade de Granger no domínio da frequência [1,81,82] e na Seção C.1.2 será apresentada uma maneira de como fazer sua medida.

C.1.2 A coerência parcial direcionada — PDC

Considere a transformada \mathcal{Z} da Equação B.1.1, que relaciona a entrada e saída do sistema, dada por:

$$\mathbf{X}(z) = \mathbf{H}(z)\mathbf{W}(z), \qquad (C.1.2)$$

em que:

$$\mathbf{H}(z) = \bar{\mathbf{A}}^{-1}(z) = \left[\mathbf{I}_K - \sum_{r=1}^p \mathbf{A}(r) z^{-r}\right]^{-1},$$
 (C.1.3)

 com

$$\mathbf{A}(r) = \begin{bmatrix} a_{11}(r) & a_{12}(r) & \cdots & a_{1K}(r) \\ a_{21}(r) & a_{22}(r) & \cdots & a_{2K}(r) \\ \vdots & \vdots & \ddots & \vdots \\ a_{K1}(r) & a_{K2}(r) & \cdots & a_{KK}(r) \end{bmatrix}.$$

Assim, a transformada de Fourier de tempo discreto (TFTD), de $\mathbf{H}(z)$ é:

$$\mathbf{H}(f) = \bar{\mathbf{A}}^{-1}(f) = \left[\mathbf{I}_K - \sum_{r=1}^p \mathbf{A}(r) z^{-r} \Big|_{z=e^{-i2\pi f}} \right]^{-1}, \quad (C.1.4)$$

em que $\mathbf{i}^2 = -1$ ($\mathbf{i} \in \mathbb{C}$ é a unidade imaginária).

Tomando-se a seguinte identidade, a partir da Equação C.1.4:

$$\bar{\mathbf{A}}(f) = \mathbf{I}_{K} - \sum_{r=1}^{p} \mathbf{A}(r) e^{-\mathbf{i}2\pi f r}$$

$$= \begin{bmatrix} \bar{\mathbf{a}}_{1}(f) & \bar{\mathbf{a}}_{2}(f) & \cdots & \bar{\mathbf{a}}_{K}(f) \end{bmatrix}$$
(C.1.5)

e, portanto, o conceito do *fator da coerência parcial direcional* (PDCF), que foi introduzida em [1], é o seguinte:

$$\pi_{ij}(f) = \frac{|\bar{A}_{ij}(f)|}{\sqrt{\bar{\mathbf{a}}_j^{\mathrm{H}}(f)\boldsymbol{\Sigma}_{\mathbf{w}}^{-1}\bar{\mathbf{a}}_j(f)}},$$
(C.1.6)

em que, $\bar{A}_{ij}(f)$ é o i, j-ésimo elemento da matriz $\mathbf{A}(f)$ e $\Sigma_{\mathbf{w}}$ é a matriz de variância-covariância do processo de inovação, de média zero e estacionário.

Uma vez que:

$$\bar{A}_{ij} = \begin{cases} 1 - \sum_{r=1}^{p} a_{ij}(r) e^{-\mathbf{i}2\pi f r}, & \text{se} \quad i = j \\ -\sum_{r=1}^{p} a_{ij}(r) e^{-\mathbf{i}2\pi f r}, & \text{caso contrário,} \end{cases}$$
(C.1.7)

a PDCF depende, principalmente, dos coeficientes $a_{ij}(r)$, que descrevem as relações que surgem entre o presente da série temporal $x_i(n)$ e o passado de $x_j(n)$, quando comparado o efeito do passado de $x_j(n)$, a outras séries.

Como $\Sigma_{\mathbf{w}}$ está presente no denominador da Equação C.1.6, a PDCF mede ambas, ou seja, causalidade de Granger e causalidade de Granger instantânea. Para eliminar essa ambiguidade e dar preferência apenas pela causalidade de Granger, define-se a PDC da seguinte maneira: [1]:

$$\bar{\pi}_{ij}(f) = \frac{|\bar{A}_{ij}(f)|}{\sqrt{\bar{\mathbf{a}}_j^{\mathrm{H}}(f)\bar{\mathbf{a}}_j(f)}},$$
(C.1.8)

com as seguintes propriedades:

- P1. $0 \leq |\bar{\pi}_{ij}(f)|^2 \leq 1 \Rightarrow$ Designaldade de Cauchy-Schwartz;
- P2. $\sum_{r=1}^{N} |\bar{\pi}_{ij}(f)|^2 = 1, \forall j, 1 \le j \le N.$
- P3. Devido à normalizada da Equação C.1.8, $\bar{\pi}_{ij}(f)$ representa o acoplamento relativo, de uma dada estrutura, na qual a fonte provém de j com relação à estrutura comparada, i;

P4. Quando i = j, tem-se que $\bar{\pi}_{jj}(f)$ reflete a densidade espectral do processo, com fonte em j.

Na Seção C.1.2.1, há um comentário sobre porquê a matriz $\Sigma_{\mathbf{w}}$ deve ser diagonal para que haja a causalidade de Granger instantânea e, portanto, com a Equação C.1.8 não há ambiguidade entre ambas formas de causalidade.

C.1.2.1 Sobre a causalidade de Granger instantânea

Para verificar a causalidade de Granger instantânea, pode-se mostrar que $\Sigma_{\mathbf{w}}$ é diagonal, se o estado presente da série adicionar pouco à predição da outra série [12]. Relações entre amostras presentes de séries temporais $x_i(n)$ estão presentes exclusivamente nas correlações dos processo de inovações $w_i(n)$. Diferentes testes para testar se a matriz $\Sigma_{\mathbf{w}}$ é diagonal podem ser apreciados em [12].

C.2 As variantes da PDC

Como foi visto na Seção C.1, a PDC é uma relação de conectividade efetiva, baseada em como uma série depende de outra, de modo direto, ou seja, na representação VAR. Dessa forma, a única dependência são dos coeficientes das matrizes $\mathbf{A}(r)$, da Equação B.1.1.

Nessa seção, são apresentadas as três variações possíveis para a PDC: a original, a generalizada e a informacional.

C.2.1 PDC (original) — $\bar{\pi}_{ij}(f)$

Conceito introduzido em [1]:

$$\bar{\pi}_{ij}(f) = \frac{|\bar{A}_{ij}(f)|}{\sqrt{\bar{\mathbf{a}}_j^{\mathrm{H}}(f)\bar{\mathbf{a}}_j(f)}}$$
(C.2.1)

C.2.2 PDC generalizada — $g\pi_{ij}(f)$

Conceito introduzido em [83]:

$$g\pi_{ij}(f) = \frac{|\bar{A}_{ij}(f)|\sigma_{ii}^{-1/2}}{\sqrt{\bar{\mathbf{a}}_j^{\mathrm{H}}(f)\boldsymbol{\Sigma}_{\mathbf{d}}^{-1}\bar{\mathbf{a}}_j(f)}},$$
(C.2.2)

em que $\sigma_{jj} = \mathbb{E}[w_j(n)^2], \Sigma_d = \Sigma_w \circ \mathbf{I}_K \in (\circ)$ é o produto de Hadamard.

C.2.3 PDC informational — $\iota \pi_{ij}(f)$

Conceito introduzido em [84,85]:

$$\iota \pi_{ij}(f) = \frac{|\bar{A}_{ij}(f)| \sigma_{ii}^{-1/2}}{\sqrt{\bar{\mathbf{a}}_j^{\mathrm{H}}(f) \boldsymbol{\Sigma}_{\mathbf{w}}^{-1} \bar{\mathbf{a}}_j(f)}},$$
(C.2.3)

em que $\sigma_{jj} = \mathbb{E} \left[w_j(n)^2 \right]$.

ANEXO D – DEMONSTRAÇÕES DOS TEOREMAS DO CAPÍTULO 3

Este apêndice é dedicado às demonstrações dos Teoremas 3.1.1 e 3.1.2, apresentados no Capítulo 3.

D.1 Demonstração do Teorema 3.1.1

Demonstração. Por estacionariedade, é suficiente mostrar que o negativo do lado direito da igualdade da Equação 3.1.4 é igual a

$$\mathbb{E}\left\{-\ln\left[\frac{C_1^{(2)}(r)}{C_1^{(1)}(r)}\right]\right\} = \mathbb{E}\left\{-\ln\left[\mathbb{P}\left(|x_j - x_i| \le r \mid |x_{j-1} - x_{i-1}| \le r\right)\right]\right\},\$$

que por sua vez é a própria interpretação probabilística da ApEn, dada pelas Equações 2.2.5 e 2.2.6, quando m = 1. Como se sabe, a probabilidade condicional é:

$$\mathbb{P}\left(|x_j - x_i| \le r \mid |x_{j-1} - x_{i-1}| \le r\right) = \frac{\mathbb{P}\left(|x_j - x_i| \le r \& |x_{j-1} - x_{i-1}| \le r\right)}{\mathbb{P}(|x_{i-1} - x_{j-1}| \le r)}.$$

Por definição [14], entropia é dada por

$$\mathbb{E}[\ln(\mathbf{X})] = \int_{-\infty}^{\infty} \ln(\mathbf{X}) \pi(\mathbf{x}) d\mathbf{x}.$$

Logo, aplicando o operador esperança a $-\ln \left[\mathbb{P}\left(|x_j - x_i| \leq r \mid |x_{j-1} - x_{i-1}| \leq r\right)\right]$, chega-se à Equação 3.1.4, em que $\mu(x, y)$ é a pdf conjunta (ou probabilidade estacionária) dos processos $x \in y \in \pi(x)$ a pdf marginal (ou probabilidade de equilíbrio) de x.

D.2 Demonstração do Teorema 3.1.2

Demonstração. Como, por hipótese, o processo $\{\mathbf{x_n}\}_{n\in\mathbb{Z}} \in i.i.d.$, então $\mu(x_1, \dots, x_n) = \prod_{i=1}^N \pi(x_i)$. Ainda, do Teorema 3.1.1, sabe-se que o estimador $\widehat{\operatorname{ApEn}}(m = 1, r, N)$ converge em probabilidade para $\mathbb{P}(|x_j - x_i| \leq r \mid |x_{j-1} - x_{i-1}| \leq r)$. Dessa forma, tem-se:

$$\mathbb{P}(|x_j - x_i| \le r \mid |x_{j-1} - x_{i-1}| \le r) = \frac{\mathbb{P}(|x_j - x_i| \le r \& |x_{j-1} - x_{i-1}| \le r)}{\mathbb{P}(|x_{i-1} - x_{j-1}| \le r)} \\
= \frac{\mathbb{P}(|x_j - x_i| \le r) \mathbb{P}(|x_{j-1} - x_{i-1}| \le r)}{\mathbb{P}(|x_{i-1} - x_{j-1}| \le r)} \\
= \mathbb{P}(|x_j - x_i| \le r)$$

e, portanto, para processos *i.i.d.*, tem-se: $\widehat{\operatorname{ApEn}}(m = 1, r, N) \xrightarrow{p} \mathbb{P}(|x_j - x_i| \le r).$

Como já foi verificado que para o caso mais simples (m = 1), é possível obter uma expressão para $\widehat{\operatorname{ApEn}}(m = 1, r, N)$, com boa aproximação. Dessa forma, a demonstração segue pelo *Princípio de Indução Finita* — \mathcal{PIF} :

Hipótese de Indução (\mathcal{HI}): $\exists k, 0 \leq k \leq m-1, m \in \mathbb{Z}_{+}^{\star}, \forall i, j, tal que:$

$$\mathbb{P}\left(|x_{j+m-1} - x_{i+m-1}| \le r \mid |x_{j+k-1} - x_{i+k-1}| \le r\right) = \mathbb{P}\left(|x_{j+m-1} - x_{i+m-1}| \le r\right)$$

Tese de Indução (\mathcal{TI}): $\exists k + 1, 1 \leq k + 1 \leq m, m \in \mathbb{Z}_+^{\star}, \forall i, j, tal que:$

$$\mathbb{P}\left(|x_{j+m-1} - x_{i+m-1}| \le r \mid |x_{j+k} - x_{i+k}| \le r\right) = \mathbb{P}\left(|x_{j+m-1} - x_{i+m-1}| \le r\right)$$

Portanto, como o processo $\{\mathbf{x}_n\}_{n\in\mathbb{Z}}$ é *i.i.d.*, por hipótese, desenvolvendo-se o

primeiro membro da \mathcal{TI} , deve-se ter:

$$\mathbb{P}\left(|x_{j+m-1} - x_{i+m-1}| \leq r \mid |x_{j+k} - x_{i+k}| \leq r\right) \iff$$

$$\iff \frac{\mathbb{P}\left(|x_{j+m-1} - x_{i+m-1}| \leq r \& |x_{j+k} - x_{i+k}| \leq r\right)}{\mathbb{P}(|x_{j+k} - x_{i+k}| \leq r)} \iff$$

$$\iff \frac{\mathbb{P}\left(|x_{j+m-1} - x_{i+m-1}| \leq r\right) \mathbb{P}\left(|x_{j+k} - x_{i+k}| \leq r\right)}{\mathbb{P}(|x_{j+k} - x_{i+k}| \leq r)} \iff$$

$$\iff \mathbb{P}\left(|x_{j+m-1} - x_{i+m-1}| \leq r\right),$$

que é igual ao segundo membro da \mathcal{TI} .

Portanto, para processos *i.i.d.*, tem-se:

$$\widehat{\operatorname{ApEn}}(m,r,N) \xrightarrow{p} \mathbb{P}\left(|x_{j+m-1} - x_{i+m-1}| \le r \right), \forall m \in \mathbb{Z}_{+}^{\star}$$

e, assim, chega-se à Equação 3.1.5.

ANEXO E – SOBRE O INTERVALO DE CONFIANÇA DA SampEn

Este apêndice discute como obter o intervalo de confiança (IC) da estatística $\widehat{\text{SampEn}}(m, r, N)$, para o caso em que a distribuição deste estimador é assintoticamente Gaussiana (Normal), ou seja:

$$\widehat{\operatorname{SampEn}}(m,r,N) \xrightarrow{d} \mathsf{N}(\mu,\sigma^2)$$

em que μ e σ^2 são, respectivamente, a média e a variância, da respectiva distribuição.

Os resultados principais são extraídos de [21].

E.1 Intervalo de confiança — IC

Como foi discutido na Definição 2.2.2, a estatística $\widehat{\operatorname{CP}}(m,r) = \frac{A(r)}{B(r)}$ estima a probabilidade condicional de que duas séries temporais tendem a ficar similares em m + 1 pontos, dado que são parecidas em m pontos. A precisão da estimativa da $\operatorname{CP}(m,r)$ pode ser regida pelo tamanho do IC, que é proporcional ao erro padrão $\left(\widehat{\operatorname{SE}}(m,r) = \frac{\widehat{\sigma}_{\widehat{\operatorname{CP}}(m,r)}}{\widehat{\operatorname{CP}}(m,r)}\right)$ da SampEn.

Dessa forma, se B(r) fosse fixo e todas as séries similares de m pontos independentes, umas das outras, então a variável aleatória (v.a.) A(r) possuiria distribuição binomial e a variância da $\widehat{CP}(m, r)$ seria, simplesmente:

$$\frac{\widehat{\operatorname{CP}}(m,r)\left(1-\widehat{\operatorname{CP}}(m,r)\right)}{B(r)}$$

Entretanto, a situação é mais complicada, pois B(r) também é uma v.a. e é correlacionado com A(r). Essa última afirmação é um tanto óbvia, pois não se pode deixar de considerar os casos em que há uma sobreposição entre as séries similares por m + 1 e m pontos.

Na Seção E.2 mostra-se como calcular uma estimativa da variância da SampEn, que é dada por:

$$\hat{\sigma}_{\widehat{\operatorname{CP}}(m,r)}^2 = \frac{\widehat{\operatorname{CP}}(m,r)\left(1-\widehat{\operatorname{CP}}(m,r)\right)}{B(r)} + \frac{1}{B^2(r)}\left[\hat{K}_A(m,r) - \hat{K}_B(m,r)\widehat{\operatorname{CP}}^2(m,r)\right],$$
(E.1.1)

em que $\hat{K}_A(m,r)$ e $\hat{K}_B(m,r)$ são as estimativas dos números de pares em que há sobreposições entre séries similares m + 1 e m pontos, respectivamente.

Usando a seguinte aproximação (chamada de *padrão*, na literatura):

$$\sigma_{g[\widehat{CP}(m,r)]} \approx \left| g' \left[\widehat{CP}(m,r) \right] \right| \hat{\sigma}_{\widehat{CP}(m,r)}, \tag{E.1.2}$$

com $g(x) \in g'(x)$ dados, respectivamente, por:

$$\begin{cases} g[\widehat{CP}(m,r)] = -\ln\left[\widehat{CP}(m,r)\right] \\ g'[\widehat{CP}(m,r)] = -\frac{1}{\widehat{CP}(m,r)}, \end{cases}$$
(E.1.3)

o $\widehat{\operatorname{SE}}(m,r)$ da Samp En pode ser estimado por:

$$\widehat{SE}(m,r) = \frac{\hat{\sigma}_{\widehat{CP}(m,r)}}{\widehat{CP}(m,r)}$$
(E.1.4)

Assim, para valores de m pequeno o suficiente e r grande o suficiente, porém obedecendo as faixas de valores discutidas no Capítulo 3, pode-se assumir que a SampEn possui distribuição Gaussiana (Normal — $N(\mu, \sigma^2)$) e o IC é calculado

da seguinte maneira:

$$IC = \left[-\ln\left[\widehat{CP}(m,r)\right] - \left|z_{\alpha/2}\right| \frac{\hat{\sigma}_{\widehat{CP}(m,r)}}{\widehat{CP}(m,r)}; -\ln\left[\widehat{CP}(m,r)\right] + \left|z_{\alpha/2}\right| \frac{\hat{\sigma}_{\widehat{CP}(m,r)}}{\widehat{CP}(m,r)} \right],$$
(E.1.5)

em que $|z_{\alpha/2}|$ é o $\frac{\alpha}{2}$ quantil da Gaussiana padrão e α é o *erro tipo I*.

E.2 Estimativa da variância da SampEn

Dada a série temporal $u(n) := (u(1), \dots, u(N))$, de comprimento N e assumida ser *i.i.d*, sejam os blocos $\mathbf{x}_m(i) := (u(i), u(i+1), \dots, u(i+m-1))$, de comprimento m ($1 \le i \le N - m + 1$). Relativamente à Definição 2.2.2, as expressões de B(r) e A(r) podem ser rescritas da seguinte forma:

$$B(r) = \sum_{i,j} U_{ij}^B(m,r)$$
 (E.2.1a)

$$A(r) = \sum_{i,j} U_{ij}^{A}(m,r),$$
 (E.2.1b)

em que:

$$U_{ij}^{B}(m,r) = \begin{cases} 1, & \text{se } \mathbf{x}_{m}(i) \text{ coincide com } \mathbf{x}_{m}(j) \\ 0, & \text{caso contrário.} \end{cases}$$
(E.2.2)

$$U_{ij}^{A}(m,r) = \begin{cases} 1, & \text{se } \mathbf{x}_{m+1}(i) \text{ coincide com } \mathbf{x}_{m+1}(j) \\ 0, & \text{caso contrário.} \end{cases}$$
(E.2.3)

Portanto, a variância da probabilidade condicional, $\widehat{CP}(m, r) = \frac{A(r)}{B(r)}$, fica:

$$\hat{\sigma}_{\widehat{CP}(m,r)}^{2} = \frac{1}{B(r)} \operatorname{VAR}\left[A(r)\right] = \frac{1}{B^{2}(r)} \sum_{i,j} \sum_{k,l} \operatorname{COV}\left(U_{ij}^{A}(m,r), U_{kl}^{A}(m,r)\right).$$
(E.2.4)

No somatório da Equação E.2.4, para os pares em que há coincidência de pares

de tamanho m, quando i = k e j = l, deve-se ter:

$$\operatorname{COV}\left(U_{ij}^{A}(m,r), U_{kl}^{A}(m,r)\right) = \operatorname{VAR}\left[A(r)\right] = \widehat{\operatorname{CP}}(m,r)\left(1 - \widehat{\operatorname{CP}}(m,r)\right), \quad \operatorname{se} i = k \operatorname{e} j = l$$
(E.2.5)

Se, nos pares contados, não há pontos em comum, então eles são independentes e, portanto, descorrelacionados, acarretando a:

$$COV(U_{ij}^{A}(m,r), U_{kl}^{A}(m,r)) = 0.$$
 (E.2.6)

Porém, se os blocos se sobrepõem, a covariância pode ser estimada por:

$$\widehat{\text{COV}}\left(U_{ij}^{A}(m,r), U_{kl}^{A}(m,r)\right) = U_{ij}^{A}(m,r)U_{kl}^{A}(m,r) - \widehat{\text{CP}}^{2}(m,r)$$
$$= \begin{cases} 1 - \widehat{\text{CP}}^{2}(m,r) & \text{se dois blocos de } m+1 \text{ pontos coincidem} \\ -\widehat{\text{CP}}^{2}(m,r) & \text{caso contrário.} \end{cases}$$
(E.2.7)

Dessa forma, a variância pode ser estimada pela Equação E.1.1.

Note que o cálculo das quantias $\hat{K}_A(m,r)$ e $\hat{K}_B(m,r)$ não é trivial. São necessárias as contagens de todos os pares referentes a $B^2(r)$ e, como do ponto de vista de algoritmo, a contagem de B(r) é $\mathcal{O}(N^2)$, $\hat{\sigma}^2_{\widehat{CP}(m,r)}$ será $\mathcal{O}(N^4)$. Dessa forma, para valores pequenos m e grandes de r, a complexidade pode se tornar alta.

Vale, ainda, ressaltar que a condição de que um bloco (i, j) sobreponha um outro (l, j) é equivalente a:

$$\min(i - k, i - l, j - k, j - l) \le m,$$
(E.2.8)

e, de acordo com a teoria de contagem, é necessário cuidado para que os pares que se sobrepõem não sejam contados duas vezes.

ANEXO F – PUBLICAÇÕES

Neste apêndice são apresentadas as publicações referentes a este trabalho:

- MASSAROPPE, L.; BACCALÁ, L. A.; SAMESHIMA, K. Semiparametric Detection of Nonlinear Causal Coupling Using Partial Directed Coherence. In: 2011 Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC). Boston: [s.n.], 2011. (Aceito para publicação).
- 2 MASSAROPPE, L.; BACCALÁ, L. A. Método semi-paramétrico para inferência de conectividade não-linear entre séries temorais. In: Congresso de Matemática Aplicada e Computacional — Região Sudeste. Uberlândia: [s.n.], 2011. (Aceito para publicação).

Semiparametric Detection of Nonlinear Causal Coupling Using Partial Directed Coherence

Lucas Massaroppe

Luiz Antonio Baccalá

Koichi Sameshima

Abstract—Infering causal relationships from observed time series has attracted much recent attention. In cases of nonlinear coupling, adequate inference is often hindered by the need to specify coupling details that call for many parameters and global minimization of nonconvex functions. In this paper we use an example to investigate a new concept, termed here *running entropy mapping*, whereby time series are mapped onto other entropy related time sequences whose analysis via a linear parametric time series methods, such as partial directed coherence, is able to expose the presence of formerly linearly undetectable causal relationships.

Keywords:Approximate Entropy, Sample Entropy, Granger Causality, Partial Directed Coherence

I. INTRODUCTION

Specially in neuroscience [1], [2], [3], [4], [5], [6], [7], [8], but also in other biomedical applications [9], [10], [11], much recent attention has been paid to methods for infering the relationship between observations that evolve in time. This endeavour has become known as the study of 'connectivity' and is now seen as a pre-requisite for elucidating the brain's inner workings. The reason behind this interest is furthered by the fact that these techniques enable constructing plausible causal explanations for the time evolution of observations in connection to brain states while avoiding often invasive and possibly harmful direct intervention procedures.

Many currently promising techniques somehow ultimately rely explicitly or not on the idea of Granger causality [12]. The reason for this interest, in addition, is the possibility of providing precise measures of information flow [13], [14] which includes the interaction direction as opposed to mere correlation based methods [15].

To date, possibly thanks to their well understood convergence properties, the most successfull techniques employ adequately fitted linear multivariate models to simultaneously acquired time series data [16], [17]. It is noteworthy that such techniques have even proved successful in detecting some instances of nonlinear interactions given sufficiently high model order and lengthy observations [18].

However, in some cases such as for quadratic coupling (see Sec. III below), linear model approximations fail. Whereas alternatives obviously exist, both nonparametric [19], [20], [21], [22] and parametric [10], [23], they often require many

Lucas Massaroppe and Luiz Antonio Baccalá are with Escola Politécnica, Department of Telecommunications and Control Engineering, University of São Paulo, Av. Prof. Luciano Gualberto, travessa 3, São Paulo, Brazil, 05508 – 900. baccala@lcs.poli.usp.br, lucasmassaroppe@gmail.com Koichi Sameshima is with Faculdade de Medicina, Department of Ra-

Koichi Sameshima is with Faculdade de Medicina, Department of Radiology, University of São Paulo, Av. Dr. Enéas de Carvalho Aguiar, 255, São Paulo, Brazil, 01246 – 903. ksameshi@usp.br observations for converging. In the parametric case, additional difficulties arise from reliance on 'ad hoc' structural assumptions and from the usually non convex nature of the functionals employed to obtain parameter estimates.

In this paper, to capture the presence of nonlinear interactions whose presence is linearly undetectable, we investigate a new hybrid approach we term *running entropy mapping*. The main idea is to compute a time dependent measure of a time series's complexity over a suitably long running window. This produces an allied time series that portrays how its complexity evolves in time. The next step consists of comparing the resulting mapped time series among themselves via linear multivariate methods.

The rest of this paper is organized as follows: Sec. II describes two entropy running measures and their computation and briefly recaps *partial directed coherence* (PDC) [3] whose use is made in Sec. III to illustrate the effectiveness of the proposal for a simple model. This is followed by a brief discussion and conclusions in Sec. IV.

II. THE METHOD

There are two steps to the method: (a) entropy mapping and (b) linear analysis of the resulting mapped series. Whereas many alternatives exist for the second step, here for definiteness we employ *partial directed coherence* (PDC).

A. Running Entropy Mapping

Consider a time series $x_i(n)$ comprising N sequential observations. Associate it to another time series $\xi_i(n)$ sequence generated from a sliding window $x(n-W+1), \ldots, x(n)$ and constructed so as to reflect some measure of the original time series complexity.

In this paper we examine two such measures: (a) Pincus's *approximate entropy* [24], [25], [26] and (b) Lake's et al. [27] bias corrected *sample entropy* which justifies the *running entropy mapping* terminology adopted here.

In addition to the window length, W, the latter entropies require defining an embedding dimension m and a radius rand consist of counting the odds of sample m length packets in the series that are close to each given such packet to within a distance r.

Whereas the proposal of the latter filtering of $x_i(n)$ is a quite general one, the choice of the latter entropy measures is justified by their fast convergence in terms of number of observed points as W << N and their reasonable reported immunity to the presence of additive noise [28]. An added advantage is their asymptotic gaussian behaviour consistent with the statistical tests [29] adopted herein.

The next step consists of applying causality analysis relating mapped $\xi_i(n)$ time series among themselves rather than the original $x_i(n)$ series.

B. PDC Analysis

Partial directed coherence was introduced [3] as a means of exposing linear Granger causal relationships in the frequency domain. When relating K simultaneously observed time series PDC is given by:

$$\pi_{ij}(f) = \frac{\bar{A}_{ij}(f)}{\sqrt{\sum_{l=1}^{K} |\bar{A}_{lj}(f)|^2}}.$$
(1)

where

,

$$\bar{A}_{ij}(f) = \begin{cases} 1 - \sum_{l=1}^{p} a_{ij}(l) e^{-j2\pi f l}, \text{ if } i = j \\ - \sum_{l=1}^{p} a_{ij}(l) e^{-j2\pi f l}, \text{ otherwise} \end{cases}$$
(2)

for $\mathbf{j} = \sqrt{-1}$ and where $a_{ij}(l)$ are the coefficients of an adequately fit multivaritate autoregressive model which in the present proposal relates the associated entropy time series rather than the original observations.

III. SIMULATION RESULTS

To examine the proposed approach, consider the following model describing a linear stochastically fed oscillator

$$\begin{cases} x_1(n) = 2R\cos(.2\pi)x_1(n-1) - R^2x_1(n-2) + w_1(n), \\ x_2(n) = -.9x_2(n-1) + \beta x_1^2(n-1) + w_2(n), \end{cases}$$

that is quadratically connected to a low pass system filter whose connectivity strength is gauged through β . Both $w_i(n)$ were taken as gaussian zero mean mutually uncorrelated white driving processes. The simulations used a sharp resonance, i.e. R = .99.



Fig. 1. A realization of (a) $x_1(n)$ and (b) $x_2(n)$ from the model in Eq. 3 comprising 1,500 time samples (R = .99 and $\beta = .05$).



Fig. 2. The computed PDC, in standard form (see text), for the data in Fig. 1 showing that the existing nonlinear influence from $x_1(n)$ to $x_2(n)$ is not captured.

A sample run (1,500 data points) of such vector process is shown on Fig. 1 whose model led to the PDC portrayed in Fig. 2 where no causality can be detected at 5% as the estimates are below the dashed line threshold [29]. In Fig. 2, the usual matrix convention [3] of portraying PDC is adopted. The graphs along the main diagonal represent the series power spectral (arbitrary unit log scale) whereas the counter diagonal portrays PDCs, i.e. the spectral connectivity representations where the bottom left graph corresponds to the $x_1(n) \rightarrow x_2(n)$ connection and the upper right graph to $x_2(n) \rightarrow x_1(n)$. Similar conventions apply to all PDC graphs used herein.

A. Approximate Entropy

(3)



Fig. 3. Approximate running entropy series computed for the data in Fig. 1 using W = 150, m = 1 and r = .15.





Fig. 5. Running sample entropy reconstruction from the data in Fig. 1.

Fig. 4. PDC results for the traces in Fig. 3 in standard form as described in the text showing correctly inferred directionality, i.e. $PDC_{1\rightarrow2}$ (left bottom graph) is above the dashed threshold, whereas $PDC_{2\rightarrow1}$ is mostly below threshold.

Reconstruction of the allied approximate entropy time series $[\xi_1^a(n) \xi_2^a(n)]^T$ (using W = 150, r = 0.15, h = 1 and m = 1) is shown on Fig. 3 and its associated computed PDC on Fig. 4 where significant $\xi_1^a(n) \rightarrow \xi_2^a(n)$ is present above threshold and correctly infers interaction direction whereas no significant interaction happens in the reverse direction.

	ТР	FP	TP	FP		
R = 0,99, h = 1	$\beta = 0.05$		$\beta = 0,10$			
m = 2, W = 150, r = 0.10	82.69	24.32	98.64	82.45		
m = 1, W = 150, r = 0.15	93.92	2.38	99.98	5.51		
m = 2, W = 200, r = 0.10	95.88	45.05	99.87	87.59		
m = 1, W = 200, r = 0.10	95.60	3.21	100.00	5.84		
TABLE I						

Approximate Entropy results for 10,000 trials as a function of coupling strength (β) and space reconstruction parameters. The TP label refers to the percentage of times

 $\xi_1^a(n)\to\xi_2^a(n)$ is correctly detected while the FP label refers to the rate of reverse incorrectly detected connections

 $(\xi_2^a(n)\to\xi_1^a(n)).$

To assess method robustness and its dependence on entropy space parameters , 10,000 realizations of the process in Eq. (3) were used in obtaining the values of Table I where the crucial nature of reconstruction parameter choice becomes apparent, the best results happen for m = 1, W = 150, h = 1 and r = 0.15 even for small β and are to within the expected 5% test significance.

B. Sample Entropy

Similar results are obtained using the sample entropy (reconstructed $[\xi_1^s(n) \xi_2^s(n)]^T$ series on Fig. 5 for W = 150, h = 1 and m = 1) and its allied PDC (Fig. 6) where again connectivity is correctly inferred at 5% (see also Table II).



Fig. 6. Standard form PDC (see text) between the traces in Fig. 5 show correct directional connectivity inference.

	TP	FP	TP	FP		
R = 0,99, h = 1	$\beta = 0.05$		$\beta = 0.10$			
m = 2, W = 150, r = 0,10	94,67	5,05	99,99	3,37		
m = 1, W = 150, r = 0.15	93,96	2,36	100,00	11,68		
m = 2, W = 200, r = 0,10	98,68	9,39	99,97	6,84		
m = 1, W = 200, r = 0,10	93,90	3,40	100,00	13,25		
TABLE II						

Sample Entropy results for 10,000 trials as a function of coupling strength (β) and space reconstruction parameters. The TP label refers to the percentage of times $\xi_1^{s}(n) \rightarrow \xi_2^{s}(n)$ is correctly detected while refers to the rate of reverse incorrectly detected connections ($\xi_2^{s}(n) \rightarrow \xi_1^{s}(n)$).
IV. DISCUSSION AND FUTURE WORK

Though in many ways still preliminary, the present results point to the potential of using suitable transformations of time series to still infer causality by fitting linear models between the resulting transformed series in those cases where the causal coupling between the original time series is in principle not even approximately detectable via linear vector autoregressions.

The basic idea presented herein is that of using the fluctuations in entropy measures to gauge how complexity flows from one time series to another. The extensive simulations portray how critical phase space reconstruction is for the process to work properly thus giving rise to the new problem of optimal (W, m, r) parameter choice in the present context.

As perhaps expected, sample entropy proves slightly superior by generating a lower false positive rate.

It is interesting to note that testing many of the mapped running entropy series resulted in the presence of significant cointegration between traces which passed specific Granger causality tests at rates comparable to the ones presented here.

The present study case points to the interest in studying the present methodology further specially in cases of models comprizing larger dimensions. Exploratory investigation of further examples is under way. The study of alternative running maps is also in progress.

V. ACKNOWLEDGMENTS

The authors gratefully acknowledge support from the FAPESP Grant 2005/56464-9 (CInAPCe). L.M to FAPESP Grant 2009/04397-7. L.A.B. to CNPq Grant 304404/2009-8.

REFERENCES

- [1] M. Kaminski and K. Blinowska, "A new method of the description of the information flow in the brain structures," *Biological Cybernetics*, vol. 65, pp. 203–210, 1991.
- A. Baccalá, K. Sameshima, G. Ballester, A. C. do Valle, and [2] C. Timo-Iara, "Studying the interaction between brain structures via directed coherence and granger causality," Applied Signal Processing, vol. 5, pp. 40-48, 1998.
- [3] L. A. Baccalá and K. Sameshima, "Partial directed coherence: a new concept in neural structure determination," Biological Cybernetics,
- concept in neural structure determination, *Biological Cybernetics*, vol. 84, no. 6, pp. 463–474, 2001.
 [4] M. Ding, S. L. Bressler, W. Yang, and H. Liang, "Short-window spectral analysis of cortical event-related potentials by adaptive multi-variate autoregressive modeling: data preprocessing, model validation, and usriability accounterment," *Biological Cybernetics*, vol. 82, pp. 26. and variability assessment," Biological Cybernetics, vol. 83, pp. 35-45 2000
- L. Astolfi, F. Cincotti, D. Mattia, M. Marciani, L. Baccala, F. Fallani, [5] S. Salinari, M. Ursino, M. Zavaglia, L. Ding, J. Edgar, G. Miller, B. He, and F. Babiloni, "Comparison of different cortical connectivity estimators for high-resolution eeg recordings," Human Brain Mapping, vol. 28, no. 2, pp. 143–157, 2007.
 L. Astolfi, F. Cincotti, D. Mattia, M. Marciani, L. Baccala,
- [6] L. [6] L. Astolli, F. Cincolti, D. Mattia, M. Marciani, L. Baccala, F. De Vico Fallani, S. Salinari, M. Ursino, M. Zavaglia, and F. Ba-biloni, "Assessing cortical functional connectivity by partial directed coherence: Simulations and application to real data," *IEEE Transac-tions on Biomedical Engineering*, vol. 53, no. 9, pp. 1802–1812, 2006.
 [7] K. Blinowska, R. Kus, M. Kaminski, and J. Janiszewska, "Transmis-
- sion of brain activity during cognitive task," *Brain Topography*, vol. 23, no. 2, pp. 205–213, 2010.
- no. 2, pp. 203–213, 2010.
 J. R. Sato, D. Y. Takahashi, S. M. Arcuri, K. Sameshima, P. A. Morettin, and L. A. Baccalá, "Frequency domain connectivity identification: an application of partial directed coherence in fMRI." *Hum. Brain. Mapp.*, vol. 30, no. 2, pp. 452–461, 2009.

- [9] L. Faes, A. Porta, R. Cucino, S. Cerutti, R. Antolini, and G. Nollo, 'Causal transfer function analysis to describe closed loop interaction
- Causa transfer function analysis to describe closed roop interactions between cardiovascular and cardiorespiratory variability signals," *Biol. Cybern.*, vol. 90, pp. 390–399, 2004.
 L. Faes, G. Nollo, and K. H. Chon, "Linear and nonlinear parametric model identification to assess granger causality in short-term cardiorespinatory of the proceeding of the proceedi vascular interactions," Computers in Cardiology, vol. 35, pp. 793-796, 2008
- [11] A. Fujita, P. Severino, J. R. Sato, and S. Miyano, "Granger causality in systems biology: modeling gene networks in time series microarray data using vector autoregressive models," in *Proceedings of the Ad*vances in bioinformatics and computational biology, and 5th Brazilian conference on Bioinformatics and computational notacitation conference on Bioinformatics, ser. BSB'10. Berlin, Heidelberg: Springer-Verlag, 2010, pp. 13–24.
 C. W. J. Granger, "Investigating causal relations by econometric models and cross-spectral methods." *Econometrica*, vol. 37, no. 3,
- pp. 424–438, 1969.
 [13] D. Y. Takahashi, L. A. Baccalá, and K. Sameshima, "Frequency
- domain connectivity: an information theoretic perspective," in Pro-ceedings of the 32nd Annual International Conference of the IEEE Engineering in Medicine and Biology Society, Buenos Aires, IEEE. UEE Derg. 2010. pr. 1726–1730. IEEE Press, 2010, pp. 1726–1729.
 —, "Information theoretic interpretation of frequency domain con-
- [14] nectivity measures," Biological Cybernetics, vol. 103, pp. 463-469, 2010.
- [15] L. A. Baccalá and K. Sameshima, "Overcoming the limitations of correlation analysis for many simultaneously processed neural struc tures," Progress in Brain Research, vol. 130, no. Advances in Neural Population Coding, pp. 33–47, 2001. H. Lütkepohl, New introduction to multiple time series analysis.
- [16] Berlin: Springer, 2005. A. Schlögl, "A comparison of multivariate autoregressive estimators
- Signal Processing, vol. 86, no. 9, pp. 2426 2429, 2006, special
 Section: Signal Processing in UWB Communications.
 B. Schelter, M. Winterhalder, M. Eichler, M. Peifer, B. Hellwig,
- [18] B. Guschlbauer, H. Lücking, R. Dahlhaus, and J. Timmer, "Testing for directed influences among neural signals using partial directed coherence," Journal of Neuroscience Methods, vol. 152, pp. 210-219,
- [19] K. Hlaváčková-Schindler, M. Paluš, M. Veimelka, and J. Bhattacharva
- [19] K. Havackova-Schnider, M. Falus, M. Vejnerka, and J. Bhattacharya, "Causality detection based on information-theoretic approaches in time series analysis," *Physics Reports*, vol. 441, no. 1, pp. 1–46, 2007.
 [20] D. Marinazzo, M. Pellicoro, and S. Stramaglia, "Kernel method for nonlinear granger causality," *Physical Review Letters*, vol. 100, no. 14, 10102 2000. 144103 2008
- M. Paluš, V. Komárek, Z. Hrncír, and K. Sterbová, "Synchronization as adjustment of information rates: Detection from bivariate time series. *Phys. Rev. E*, vol. 63, no. 4, p. 046211, 2001.
 [22] M. Paluš and A. Stefanovska, "Direction of coupling from phases of
- interacting oscillators: An information-theoretic approach," Physical Review E, vol. 67, no. 5, pp. 1–4, 2003.
- E. Pereda, R. Q. Quiroga, and J. Bhattacharya, "Nonlinear multivariate analysis of neurophysiological signals," *Progress in Neurobiology*, vol. 77, pp. 1–37, 2005.
 S. M. Pincus, "Approximate entropy as an irregularity measure for financial data," *Econometric Reviews*, vol. 27, no. 4–6, pp. 329–362, 2005.
- 2008.
- [25] "Approximate entropy as a measure of irregularity for psychi atric serial metrics," Bipolar Disorders, vol. 8, no. 5p1, pp. 430-440, 2006.
- [26] ---, "Approximate entropy (apen) as a complexity measure," *Chaos*, vol. 5, no. 1, pp. 110–117, 1995.
 [27] D. E. Lake, J. S. Richman, M. P. Griffin, and J. R. Moorman, "Sample
- or Lake, J. Stakenhalt, M. Formin, and J. K. Mooman, Sampe entropy analysis of neonatal heart rate variability," *American Journal* of *Physiology Regulatory, Integrative and Comparative Physiology*, vol. 283, pp. R789–R797, 2002.
 S. M. Pincus and W.-M. Huang, "Approximate entropy: Statistical Methods of the physical structure of the physical structure
- properties and applications," Communications in Statistics Theory
- [29] D. Y. Takahashi, L. A. Baccalá, and K. Sameshima, "Connectivity inference via partial directed coherence: asymptotic results," *Journal of Applied Statistics*, vol. 34, pp. 1259–1273, 2007.

Método Semi-Paramétrico para Inferência de Conectividade Não-Linear entre Séries Temporais

Lucas Massaroppe, Luiz Antonio Baccalá,

Depto de Engenharia de Telecomunicações e Controle, EP, USP, 05508-900, São Paulo, SP E-mail: lucasmassaroppe@gmail.com, baccala@lcs.poli.usp.br.

Palavras-chave: Entropia aproximada, entropia amostral, causalidade de Granger, coerência parcial direcionada.

Resumo: Inferência de relações causais entre séries temporais tem atraído muita atenção recente. Quando o acoplamento é não-linear, essa tarefa torna-se particularmente difícil pois é preciso usar modelos não-lineares específicos. Neste trabalho, usa-se um exemplo para investigar uma alternativa, denominada de **mapeamento por "entropia deslizante**", na qual séries são mapeadas em outras sequências temporais relacionadas à sua entropia e que podem ser analisadas usando métodos lineares, como a coerência parcial direcionada, para expor a presença de relações causais, dispensando o uso de modelos paramétricos não-lineares em sua detecção.

1 Introdução

Até hoje, possivelmente graças ao bom entendimento das propriedades de convergência, as práticas mais bem sucedidas para a inferência da Causalidade de Granger [2] empregam técnicas de regressões a modelos de dados multivariados de séries temporais observadas [4]. Para alguns casos de acoplamento não-linear, técnicas meramente lineares se mostram capazes de detectar interações desde que se usem observações e modelos suficientemente longos. Entretanto, há casos nos quais essas aproximações lineares falham, como o de acoplamento quadrático (veja Seção 3). Aqui, para se capturar a presença de interações não-lineares, propõe-se uma nova abordagem híbrida, dita **mapeamento por** *entropia deslizante*, cuja ideia principal é calcular uma medida de complexidade, dependente do tempo, para uma série temporal, usando uma janela deslizante, isso é, um intervalo de tamanho fixo ao do tempo. Isso produz outra sequência que retrata como a complexidade da série original evolui. O passo seguinte consiste em compara resultados destes mapeamentos entre séries de interesse usando métodos lineares multivariados.

2 Método

As duas fases do método são: mapeamento e análise linear das séries mapeadas.

2.1 Mapeamento

Considere uma série temporal $x_i(n)$, com N pontos, em que $n \in i$ são, respectivamente, os indexadores do tempo discreto e da série. Associada a $x_i(n)$, seja a sequência $\xi_i(n)$ gerada, a partir de uma janela deslizante de tamanho W, com passo de deslocamento de comprimento h:

$$\begin{bmatrix} x_i(n-W+h) & \cdots & x_i(n) \end{bmatrix},$$

que reflete uma medida de complexidade da série original na qual os pontos da janela são usados para calcular estatíticas como a entropia aproximada de Pincus [5] e a entropia amostral de Lake *et.al.* [3] (essa última corrige o viés estatístico da primeira).

Além do comprimento da janela, W e do passo, h (que indica como a janela se movimenta sobre a série), é necessário considerar os parâmetros de cáclulo das entropias propriamente ditas, *i.e.*, a dimensão do espaço de estados, m e o raio de tolerância, r, para que se possa contar quantos padrões similares ocorrem em blocos com comprimento m, em uma distância r dentro da janela, segundo as definições dadas por Pincus [5] e Lake *et.al.* [3].

Embora a proposta do mapeamento de $x_i(n)$ seja geral, as escolhas destas medidas de entropia se justificam por sua rápida convergência, em termos do número de pontos em que são calculadas (W é muito menor do que N) e sua razoável imunidade a ruído conforme relatado em [5]. Outra vantagem é a gaussianidade assintótica dos respectivos estimadores, que se faz consistente com os testes estatísticos adotados.

2.2 Análise linear das séries mapeadas

A etapa seguinte consiste em aplicar a análise linear usual de causalidade às séries mapeadas, relacionando as sequências $\xi_i(n)$ entre si, em lugar de usar as séries $x_i(n)$ originais.

Embora haja diversas alternativas, aqui emprega-se a coerência parcial direcionada (*partial directed coherence* — *PDC*), introduzida em [1], como uma forma de se interpretar a *causalidade de Granger* no domínio da frequência. Quando K séries temporais são observadas simultaneamente, a PDC é dada por:

$$\bar{\pi}_{ij}(f) = \frac{|A_{ij}(f)|}{\sqrt{\sum_{l=1}^{K} |\bar{A}_{lj}(f)|}},\tag{1}$$

sendo:

$$\bar{A}_{ij}(f) = \begin{cases} 1 - \sum_{\substack{r=1 \ p}}^{p} a_{ij}(r) e^{-i2\pi fr}, & \text{se} \quad i = j \\ -\sum_{\substack{r=1 \ r=1}}^{p} a_{ij}(r) e^{-i2\pi fr}, & \text{caso contrário,} \end{cases}$$
(2)

em que $\mathbf{i}^2 = -1$, a_{ij} são coeficientes de um ajuste adequado por um modelo autorregressivo vetorial (vectorial autoregressive — VAR), que diz respeito às series "informacionais" (geradas com os métodos de entropia), em lugar das das séries originais, $|\cdot|$ denota módulo, f é a frequência normalizada em que a PDC é calculada e p é a ordem do modelo VAR, escolhido por um critério, e.g., Akaike, bayesiano de Schwartz, erro de predição final ou Hannan-Quinn (ver [4]).

3 Resultados de simulações

Com a finalidade de se validar o método proposto, considere o seguinte modelo:

$$\mathbf{x}_{n} = \begin{bmatrix} 2R\cos(0,2\pi) & 0\\ 0 & -0,9 \end{bmatrix} \mathbf{x}_{n-1} + \begin{bmatrix} -R^{2} & 0\\ 0 & 0 \end{bmatrix} \mathbf{x}_{n-2} + \begin{bmatrix} 0\\ \beta x_{1,n-1}^{2} \end{bmatrix} + \mathbf{w}_{n},$$
(3)

 $\mathbf{x}_n = (x_{1,n} \ x_{2,n})^\top \in \{\mathbf{w}_n\}_{n \in \mathbb{Z}} \sim \mathsf{N}(\mathbf{O}_{2 \times 1}, \mathbf{I}_2).$

Observe que a equação de $x_{1,n}$ descreve um oscilador com frequência de ressonância bem definida (f = 0, 1 - ver Figura 1(a)), que está acoplada, de maneira quadrática à segunda $(x_{2,n})$, que é um filtro passa-baixas¹, cuja conectividade pode ser aferida pelo parâmetro β

 $^{^{1}}$ Filtro passa-baixas é um sistema que permite a passagem de baixas frequências sem dificuldades e atenua (ou reduz) a amplitude das frequências maiores que a de corte.



Figura 1: Forma padrão das PDCs (conforme texto) calculadas entre as séries indicadas nos itens (a), (b) e (c). Repare que nos itens (b) e (c) há uma evidente melhora da correta inferência da direcionalidade, *i.e.*, $PDC_{1\rightarrow 2}$ (gráfico abaixo e à esquerda) está acima do limiar em tracejado, enquanto que para $PDC_{2\rightarrow 1}$ está sempre abaixo do limiar. Já ao se observar o painel (a), é impossível inferir a correta direção de causalidade.

(fator de acoplamento entre as séries). As simulações foram feitas tomando-se R = 0,99, para que o espalhamento espectral do pólo fosse pequeno e h = 1.

Uma PDC típica entre os processos da Equação (3) (com 1.500 pontos), na Figura 1(a), mostra que é impossível detectar a correta direção da causalidade, considerando um nível de significancia de 5%, já que as estimativas estão abaixo das curvas de limiar, tracejadas [6]. Nessa mesma figura, a maneira usual [1] de se apresentar a PDC é adotada. Os gráficos na diagonal principal representam a densidade espectral (em unidades arbitrárias logarítimicas) e na antidiagonal, estão as PDCs, *i.e.*, as representações espectrais da conectividade, em que o gráfico abaixo e à esquerda representa a conexão $x_{1,n} \rightarrow x_{2,n}$ e o gráfico acima e à direita, $x_{2,n} \rightarrow x_{1,n}$. Convenções similares são usadas em todos os gráficos das PDCs, no texto.

3.1 Robustez

As Figuras 1(b) e 1(c) mostram as PDCs entre, respectivamente, as sequências de séries de entropia aproximada $(\xi_1^a(n) \ \xi_2^a(n))^\top$ e as séries de entropia amostral $(\xi_1^s(n) \ \xi_2^s(n))^\top$. Nos dois casos, foram usados os seguintes parâmetros para reconstrução: W = 150, r = 0.15, h = 1 e m = 1. Repare que a causalidade é sempre identificada de maneira correta: $\xi_1^a(n) \rightarrow \xi_2^a(n)$ e $\xi_1^s(n) \rightarrow \xi_2^s(n)$, pois estão acima do limiar, mas suas inversas não são significantes.

Para se verificar a robustez do método, realizou-se simulação de Monte Carlo, com 10.000 replicações, sempre com R = 0.99 e h = 1. Os parâmetros W, $m \in r$ são mostrados na Tabela 1.

3.1.1 Tabelas de contingência

As tabelas de contingência mostram o quanto os dois algoritmos acertam conjuntamente, erram conjuntamente e se apenas um deles acerta.

Observando-se as Tabelas 1(a), 1(b), 1(c) e 1(d), percebe-se que o algoritmo da SampEn possui um desempenho superior ao da ApEn, pois, em todos os casos, o primeiro acerta, sozinho, mais vezes, do que o segundo.

Dessa forma, conclui-se que o conjunto de parâmetros mais favorável para uso do método é: W = 150, h = 1, m = 1, r = 0,15.

4 Discussão e trabalhos futuros

Embora com resultados preliminares, há potencialidade de se usar transformações satisfatórias de séries temporais, para se inferir causalidade por ajuste de modelos lineares, entre as

(a) Caso 1.						(b) Caso 2.			
	W = 150, m = 2, r = 0,10]	W = 150, m = 1, r = 0.15			
	$\beta = 0.05$		$\beta = 0,10$]	$\beta = 0.05$		$\beta = 0,10$	
ApEn SampEn	Acerto	Erro	Acerto	Erro		Acerto	Erro	Acerto	Erro
Acerto	55,88	1,81	92,45	0,03		88,21	0,00	99,92	0,00
Erro	38,36	96,05	7,48	99,96]	11,79	100,00	0,08	100,00
(c) Caso 3.						(d) Caso 4.			
	(c) Caso	3.					(d) C	aso 4.	
[(c) Caso W =	3. 200, m	=2, r=0	,10		W	(d) C = $200, m$	aso 4. = 1, $r = 0$,10
	(c) Caso $W = \beta = 0$	3. $200, m$	$= 2, r = 0$ $\beta = 0$,10),10		W $\beta =$	(d) C = $200, m$ 0,05	aso 4. = 1, $r = 0$ $\beta =$,10 0,10
ApEn SampEn	(c) Caso $\frac{W}{\beta} = 0$ Acerto	3. 200, m 0,05 Erro	$= 2, r = 0$ $\beta = 0$ Acerto	,10),10 Erro		W $\beta =$ Acerto	(d) C = 200, $m = 0,05$ Erro	aso 4. = 1, $r = 0$ $\beta =$ Acerto	,10 0,10 Erro
ApEn SampEn Acerto	(c) Caso $W = \frac{\beta}{\beta} = 0$ Acerto 77,43	3. 200, m 0,05 Erro 2,01	$= 2, r = 0$ $\beta = 0$ Acerto $98,21$,10),10 Erro 0,02		$\frac{W}{\beta} =$ Acerto 91,89	(d) C = $200, m = 0,05$ Erro 0,00	$\begin{array}{c} \text{aso } 4.\\ = 1, \ r = 0\\ \hline \beta =\\ \hline \text{Acerto}\\ 99,91 \end{array}$,10 0,10 Erro 0,00

Tabela 1: Tabelas de contingência, comparando os resultados dos algoritmos da ApEn e SampEn, para 10.000 realizações, como função do fator de acoplamento (β). Os parâmetros dos casos simulados são mostrados nos itens (a), (b), (c) e (d).

sequências transformadas, especialmente nos casos em que a conectividade é indetectável.

A ideia básica apresentada aqui, é usar as flutuações das medidas de entropia para medir como a complexidade migra de uma série para outra. As simulações de Monte Carlo deixam evidente como é crítico reconstruir o espaço de fases para que se tenha sucesso, dando origem ao problema de escolha ótima dos parâmetros (W, m, r).

Como era de se esperar, o algoritmo da SampEn possui um desempenho melhor do que o da ApEn.

O presente estudo indica o interesse em investigar a metodologia em modelos de maiores dimensões. Análise exploratória de tais exemplos também está em curso e métodos alternativos de mapeamento estão sendo investigados.

5 Agradecimentos

Os autores gentilmente agradecem à FAPESP, pelo apoio financeiro, através do processo 2005/56464 - 7 (Projeto CInAPCe). L.M., à FAPESP, pelo processo 2009/04397 - 7. L.A.B., ao CNPq, pelo processo 304404/2009 - 8.

Referências

- L. A. Baccalá e K. Sameshima, Partial directed coherence: a new concept in neural structure determination, *Biol. Cybern.*, 84(6) (2001) 463–474.
- [2] C. W. J. Granger, Investigating causal relations by econometric models and cross-spectral methods. *Econometrica*, 37(3) (1969) 424–438.
- [3] D. E. Lake, J. S. Richman, M. P. Griffin e J. R. Moorman, Sample entropy analysis of neonatal heart rate variability, Am. J. Physiol. Regul. Integr. Comp. Physiol., 283 (2002) R789–R797.
- [4] H. Lütkepohl, "New introduction to multiple time series analysis", Berlin: Springer, 2005.
- [5] S. M. Pincus e W.-M. Huang, Approximate entropy: statistical properties and applications. *Commun. Stat. — Theory M.* 21(11) (1992) 3061–3077.
- [6] D. Y. Takahashi, L. A. Baccalá e K. Sameshima, Connectivity inference between neural structures via partial directed coherence: asymptotic results. J. Appl. Statist. 34(10) (2007) 1259–1273.