

**Medidas de dependência entre séries temporais:
Estudo comparativo, análise estatística e aplicações em
neurociências**

Carlos Stein Naves de Brito

DISSERTAÇÃO APRESENTADA
AO
PROGRAMA INTERUNIDADES EM BIOINFORMÁTICA
DA
UNIVERSIDADE DE SÃO PAULO
PARA
OBTENÇÃO DO TÍTULO
DE
MESTRE EM CIÊNCIAS

Programa: Mestrado Interunidades em Bioinformática

Orientador: Prof. Dr. Koichi Sameshima

Durante o desenvolvimento deste trabalho o autor recebeu auxílio financeiro da CAPES

São Paulo, julho de 2010

Agradecimentos

Agradeço a todas as pessoas que me apoiaram neste projeto. Precisei de muitos conselhos pelo caminho até encontrar minha possível contribuição para essa promissora área de neurociências.

Agradeço ao Prof. Koichi Sameshima que me orientou e ajudou em todas as horas, fornecendo apoio científico e pessoal para cada etapa.

Agradeço aos colaboradores, Prof. Luiz Baccalá, Prof. João Sato, Prof. Edson Amaro, Daniel Takahashi, Luiz Lana, Eduardo Schenberg, Edgard Morya, Taisa Miranda, entre muitos outros que figuram nesta lista de pessoas fabulosas com as quais pude trabalhar e principalmente aprender.

Reconheço o quanto o apoio de uma família de veias científicas contribui no meu progresso, em especial meus queridos pais Sofia e Adriano, que sempre apoiaram todas as minhas escolhas, dando todo suporte e carinho imagináveis.

Agradeço à minha querida avó Suzana, revisora do texto, sábia e afetuosa conselheira e amiga.

Agradeço à minha querida irmã Elissa e todos meus amigos, presentes de uma forma ou de outra, em especial ao Glauber, com sua preciosa revisão gramatical, e à Gleyce, protagonistas nesta importante época.

E à Paula, minha companheira em absolutamente todos momentos deste projeto.

Resumo

Medidas de dependência entre séries temporais são estudadas com a perspectiva de evidenciar como diferentes regiões do cérebro interagem, por meio da aplicação a sinais eletrofisiológicos. Baseado na representação auto-regressiva e espectral de séries temporais, diferentes medidas são comparadas entre si, incluindo coerência espectral e a coerência parcial direcionada, e introduz-se uma nova medida, denominada transferência parcial direcionada. As medidas são analisadas pelas propriedades de parcialização, relações diretas ou indiretas e direcionalidade temporal, e são mostradas suas relações com a correlação quadrática. Conclui-se que, entre as medidas analisadas, a coerência parcial direcionada e a transferência parcial direcionada possuem o maior número de características desejáveis, fundamentadas no conceito de causalidade de Granger. A estatística assintótica é desenvolvida para todas as medidas, incluindo intervalo de confiança e teste de hipótese nula, assim como sua implementação computacional. A aplicação a séries simuladas e a análise de dados eletrofisiológicos reais ilustram o estudo comparativo e a aplicabilidade das novas estatísticas apresentadas.

Palavras-chave: Eletrofisiologia, Neurociência computacional, Conectividade funcional, Séries temporais, Inferência estatística, Medidas de dependência

Abstract

Measures of dependence between temporal series are studied in the context of revealing how different brain regions interact, through their application to electrophysiology. Based on the spectral and autoregressive model of time series, different measures are compared, including coherence and partial directed coherence, and a new measure is introduced, named partial directed transfer. The measures are analyzed through the properties of partialization, direct or indirect relations and temporal directionality, and their relation to quadratic correlation is shown. It results that among the presented measures, partial directed coherence and partial directed transfer reveal the highest number of desirable properties, being grounded on the concept of Granger causality. The asymptotic statistics for all measures are developed, including confidence intervals and null hypothesis testing, as well as their computational implementation. The application to simulated series and the analysis of electrophysiological data illustrate the comparative study and the applicability of the newly presented statistics. **Keywords:** Electrophysiology, Computational neuroscience, Functional connectivity, Time series, Statistical inference, Measure of dependence

Sumário

Lista de Abreviaturas	vii
Lista de Símbolos	viii
1 Introdução	1
2 Séries temporais	4
2.1 Séries temporais estacionárias	6
2.2 Processos lineares	7
2.3 Representação auto-regressiva	10
2.4 Representação por médias móveis	11
2.5 Representação espectral	12
2.6 Representação inversa	15
3 Medidas de dependência linear	17
3.1 Dependência e correlação	18
3.2 Coerência Espectral	19
3.3 Coerência Parcial	22
3.4 Conceitos de Causalidade	24
3.4.1 Causalidade de Granger	25
3.4.2 Causalidade de Granger condicionada	27
3.4.3 Causalidade indireta de Akaike	28
3.5 Coerência Direcionada	31
3.6 Coerência Parcial Direcionada	33
3.6.1 Transferência Parcial Direcionada	36
3.7 Comparação entre medidas de dependência	38
4 Estimação e estatística assintótica	41
4.1 Estimação	41
4.1.1 Estimação de AR(p)	42
4.1.2 Estimação da ordem p	43
4.2 Estatística assintótica	44
4.2.1 Método delta	45

4.2.2	Estatística para PDC	46
4.2.3	Estatística para DC, SDT e DTF	50
4.2.4	Estatística para Coerência Parcial	53
4.2.5	Estatística para Coerência Espectral	54
4.2.6	Estatística para PDT	56
5	Análise com exemplos	57
5.1	Implementação computacional em Python	57
5.2	Análise gráfica para dados simulados	59
5.2.1	Exemplo comparativo entre PDC e PDT	64
5.3	Estatística assintótica para dados finitos	65
5.4	Aplicação em Potencial de Campo Local em ratos	69
6	Conclusão	74
	Referências Bibliográficas	76

Lista de Abreviaturas

AR	Representação auto-regressiva (<i>Autoregressive model</i>).
AR(p)	Representação auto-regressiva de ordem finita p .
MA	Representação pro médias móveis (<i>Moving Average model</i>).
FT	Transformada de Fourier (<i>Fourier Transform</i>).
Cov	Covariância (<i>Covariance</i>).
Var	Variância (<i>Variance</i>).
GC	causalidade de Granger (<i>Granger causality</i>).
AC	causalidade de Akaike (<i>Akaike causality</i>).
Coh	Coerência Espectral (<i>Spectral Coherence</i>).
PC	Coerência Parcial (<i>Partial Coherence</i>).
DC	Coerência Direcionada (<i>Directed Coherence</i>).
SDT	Transferência Direcionada Espectral (<i>Spectral Directed Transfer</i>).
DTF	Função de Transferência Direcionada (<i>Directed Transfer Function</i>).
PDC	Coerência Parcial Direcionada (<i>Partial Directed Coherence</i>).
gPDC	Coerência Parcial Direcionada generalizada (<i>generalized PDC</i>).
iPDC	Coerência Parcial Direcionada informacional (<i>informational PDC</i>).
PDT	Transferência Parcial Direcionada (<i>Partial Directed Transfer</i>).
IPT	Transferência Parcial Instantânea (<i>Instantaneous Partial Transfer</i>).
LSE	Estimador de mínimos quadrados (<i>Least Square Estimator</i>).
AIC	Critério de Informação de Akaike (<i>Akaike Information Criterion</i>).
LFP	Potencial de Campo Local (<i>Local Field Potential</i>).
REM	Sono R.E.M. (Rapid Eye Movement).
SWS	Sono de ondas lentas (Slow-Wave Sleep).

Lista de Símbolos

B^T	Matriz transposta de B .
B^*	Conjugado complexo de B .
B^H	Matriz transposta hermitiana de B , satisfazendo $B^H = (B^T)^*$.
$Re(B)$	Parte real de B .
$Im(B)$	Parte imaginária de B .
$(B)_{ij}$	Elemento da i -ésima linha e j -ésima coluna de B .
b_{ij}	O mesmo que $(B)_{ij}$.
$det(B)$	Determinante de B .
$ B $	Valor absoluto de B .
I_k	Matriz identidade de dimensão k .
$x(t)$	Série temporal univariada.
$X(t)$	Série multivariada de ordem n : $X(t) = [x_1 x_2 \dots x_n]^T$.
$x_i(t)$	i -ésima série univariada que compõe $X(t)$.
n	Número de séries univariadas que compõe $X(t)$.
N	Tamanho dos dados amostrados.
$E[X(t)]$	Valor esperado (esperança) de $X(t)$.
X^i	Série $X(t)$ excluindo a série univariada x_i .
X^{t-}	Valores passados de $X(t)$: $(X(t-1), X(t-2), \dots)$.
$\Gamma(\tau)$	Matriz de autocovariância de $X(t)$, para atraso τ .
A_k	k -ésima matriz da representação AR, para $k \leq 1$.
H_k	k -ésima matriz da representação AR. Supõe-se $H_0 = I_n$.
p	Ordem do modelo AR.
D	Operador de atraso, satisfaz $D^k X(t) = X(t-k)$.
$\varepsilon(t)$	Série ortogonal das inovações nos modelos AR e MA.
Σ	$Var(\varepsilon(t))$.
σ_{ij}	Elemento da linha i e coluna j de Σ .

$dZ_X(\lambda)$	Representação espectral de $X(t)$.
$X(\lambda)$	Denota o mesmo que $dZ_X(\lambda)$.
$f_X(\lambda)$	Densidade espectral (ou espectro) de $X(t)$.
$f_{ii}(\lambda)$	Densidade espectral de $x_i(t)$.
$f_{ij}(\lambda)$	Densidade espectral cruzada de $x_i(t)$ e $x_j(t)$.
$A(\lambda)$	Transformada de Fourier das matrizes A_k , $k \geq 1$.
$\bar{A}(\lambda)$	Transformada de Fourier de $(A_0 = I_n, -A_1, -A_2, \dots)$. Satisfaz $\bar{A}(\lambda) = I_n - A(\lambda)$.
$H(\lambda)$	Transformada de Fourier das matrizes H_k . Satisfaz $H(\lambda) = \bar{A}(\lambda)^{-1}$.
h_j	j -ésima linha de $H(\lambda)$.
a_j	j -ésima coluna de $\bar{A}(\lambda)$.
${}^iX(t)$	Processo inverso de $X(t)$.
$g(\lambda)$	Densidade espectral do processo inverso ${}^iX(t)$.
$f_{ii x^i}(\lambda)$	Densidade espectral parcializada (ou espectro parcial) de x_i pelo espectro das outras séries, $x^i(t)$. Satisfaz $f_{ii x^i}(\lambda) = g_{ii}^{-1}(\lambda)$.
$f_{ij X^{i,j}}(\lambda)$	Espectro cruzado de x_i e x_j parcializado pelas outras séries, $X^{i,j}(t)$. Satisfaz $f_{ij X^{i,j}}(\lambda) = g_{ij}(\lambda)$.
$\eta_i(t)$	Inovação de x_i , $\varepsilon_i(t)$, parcializada pelas outras inovações, $\varepsilon^i(t)$. Satisfaz $Var(\eta(t)) = \Sigma^{-1}$.
$\rho(A, B)$	Correlação linear de Pearson entre A e B .
$\rho^2(A, B)$	Correlação quadrática, denota $ \rho(A, B) ^2$.
PDC_{ij}	Medida PDC direcionada de x_j para x_i . Mesma notação vale para todas medidas de dependência.
\hat{y}	Estimador de y .
A	$[A_1 \dots A_p]$.
α	$vec(A)$
ε	$vec(\Sigma)$
Ω_α	Variância assintótica de α .
Ω_ε	Variância assintótica de ε .
I_{ij}	$I_{ij}vec(A)$ seleciona elemento da i -ésima linha e j -ésima coluna de A .
I_j^l	$I_j^lvec(A)$ seleciona j -ésima linha de A .
I_j^c	$I_j^cvec(A)$ seleciona j -ésima coluna de A .
D_n	Matriz de duplicação, satisfaz $D_nvech(A) = vec(A)$, para $A_{n \times n}$.
A^+	Pseudo-inversa generalizada de Moore-Penrose de A .

Capítulo 1

Introdução

Nas neurociências, assim como em diversas outras áreas científicas incluindo a bioinformática, meteorologia e economia, observam-se rápido desenvolvimento de instrumentos e técnicas de aquisição de dados e, em consequência, uma crescente quantidade de informação a ser processada (Brown *et al.* (2004)).

Nas últimas duas décadas foram realizados enormes avanços na análise e interpretação destes dados, principalmente para relações entre uma tarefa cognitiva e quando, como, onde uma atividade neural acontece. Grande parte dessas análises se constituem de estatísticas descritivas de medidas univariadas, relacionando independentemente cada sinal, relativo a uma região do cérebro, a uma tarefa cognitiva (Cox e Savoy (2003)).

É sabido que o cérebro não é organizado em módulos independentes, mas sim composto de estruturas em rede, que interagem entre si. Nos últimos anos, tem havido um esforço da comunidade científica em ir além da análise univariada e lidar simultaneamente com várias áreas e inferir como elas interagem ao realizar uma tarefa cognitiva. Nessa perspectiva, avançam projetos de conectomas, interactomas e estudos de circuitos e redes, que tentam mapear e estudar as conexões entre regiões nas diferentes escalas espaciais e durante diferentes estados cognitivos (Greicius *et al.* (2003), Babiloni *et al.* (2005)).

Uma forma de analisar a interação de diversas áreas é empregar métodos multivariados. Esta dissertação contribui para o desenvolvimento e aplicação de modelos multivariados lineares que detectam dependência entre séries temporais. As medidas de dependência estudadas neste trabalho evidenciam, pela análise de dados coletados simultaneamente, se estes apre-

sentam interações lineares, se há interferência temporalmente direcionada ou realimentação, no domínio da frequência.

Pressupõe-se que a detecção de dependência entre sinais coletados em diferentes regiões seja evidência para a existência de conectividade funcional entre estas. A conectividade funcional tem definição ainda sob debate, referindo-se de forma abrangente a interações entre as atividades neuronais de cada região (Horwitz (2003)).

Causalidade será definida num sentido operacional e estará no cerne da discussão, pois estudaremos métodos que decompõem a simples correlação linear em uma correlação direcionada temporalmente.

Os métodos aqui estudados se baseiam na teoria de séries temporais estacionárias e suas representações na forma de regressão linear auto-regressiva (AR) e de médias móveis (MA). A partir da estimação destas representações multivariadas lineares, podemos criar medidas de dependência entre as séries, que têm relação direta com correlação linear entre variáveis aleatórias.

As medidas estudadas incluem a coerência (Coh), coerência parcial (PC), coerência parcial direcionada (PDC) e função de transferência direcionada (DTF), incluindo formulações alternativas.¹

Estas medidas estão sendo utilizadas de forma crescente em estudos de eletrofisiologia neural, caracterizando como diferentes regiões cerebrais se relacionam dependendo do contexto cognitivo e em que frequência do espectro as interações ocorrem. O resultado destas análises pode ser representado como uma rede de conexões direcionadas (conhecida como grafo direcionado).

Há certa confusão na literatura sobre a interpretação de cada medida de dependência, suas fundamentações teóricas e limitações práticas. Iremos analisar e comparar cada medida, e espera-se contribuir para a resolução deste debate, principalmente por meio da fundamentação em diferentes conceitos de causalidade e formulação matemática em forma de correlações quadráticas.

Ao aplicar estes métodos em dados experimentais, surge a necessidade de um instrumento

¹As abreviações por sigla deste trabalho estão de acordo com a nomenclatura inglesa encontrada na literatura de referência.

estatístico que forneça intervalos de confiança, p-valores e testes de hipóteses, validando os resultados encontrados. Obter estas estatísticas a partir da formulação dessas medidas não é uma tarefa trivial.

Uma contribuição central deste trabalho é a derivação teórica e implementação computacional das estatísticas assintóticas dos estimadores das quatro medidas de dependência citadas acima. O trabalho cobre tanto o caso geral, calculando intervalos de confiança, quanto o caso particular da hipótese nula, fornecendo testes de hipótese.

Por meio de simulação computacional, a estatística assintótica é comparada à distribuição amostral por simulação de Monte-Carlo. Exemplos para modelos com séries temporais finitas são apresentados, e aplicações em dados reais de eletrofisiologia ilustram a relevância do estudo.

Para a implementação computacional foi utilizada a plataforma código aberto Python. Com desenho amigável para a análise interativa de dados científicos, as rotinas foram projetadas para serem distribuídas para a comunidade científica e diretamente aplicadas a estudos metodológicos e análise de dados eletrofisiológicos.

Com uma implementação acessível possibilita-se um estudo computacional sistemático dos diferentes métodos, não acessível até então.

Espera-se que a disponibilidade das medidas, com intervalos de confiança e p-valores, implementadas em software de fácil uso, estimule o uso de medidas de conectividade em problemas relevantes em neurociências da comunidade científica nacional e internacional.

Capítulo 2

Séries temporais

As séries temporais são conjuntos de dados ubíquos encontrados em protocolos de pesquisas que coletam dados ao longo do tempo.

O desenvolvimento do estudos de séries temporais foi fortemente motivado pelo interesse em áreas aplicadas, como econometria e previsão de séries econômicas (Granger (1969)), engenharia e processamento de sinais em telecomunicações e biologia (Box e Jenkins (1994)), análises climáticas e geológicas (Mann e Lees (1996)).

Uma série temporal é um conjunto de variáveis aleatórias com indexação (normalmente indexada pelo tempo t), em que o índice assume valores inteiros (série discreta)

$$\mathbf{x} = (\dots, x(t-1), x(t), x(t+1), \dots). \quad (2.1)$$

Uma série temporal multivariada X é definida da mesma forma, sendo cada $X(t)$ um conjunto de n variáveis aleatórias

$$X(t) = \begin{bmatrix} x_1(t) \\ x_2(t) \\ \vdots \\ x_n(t) \end{bmatrix}. \quad (2.2)$$

Séries temporais contínuas no tempo podem ser igualmente definidas, tendo análise teórica semelhante ao caso discreto, mas não serão estudadas neste trabalho. Consideraremos

apenas séries temporais reais, com espaçamento temporal discreto e constante entre elementos.

Ao estudar séries temporais coletadas por experimentos, assumimos que esta série é uma realização de um processo gerador desconhecido. Supõe-se que exista um processo implícito, descrito probabilisticamente, que pode gerar diferentes séries temporais. Destes possíveis valores observamos uma amostra, finita, denominada realização.

Assumindo a existência desse processo subjacente, podemos definir e estimar suas propriedades, como estacionariedade, estabilidade, variância e média. Para que a estimação de propriedades como média e variância faça sentido, é necessário supor que os dados tenham propriedades que sejam atemporais, o que chamamos de estacionariedade.

Cada elemento da série é definido como uma variável aleatória, e o conjunto de elementos é definido por um conjunto de variáveis aleatórias com uma certa distribuição conjunta. A partir desta distribuição podemos inferir todas as propriedades desejadas.

Estamos interessados em dois problemas: definir propriedades relevantes de um processo estacionário e estimar tais propriedades a partir de realizações coletadas.

Neste trabalho estamos interessados nos seguintes conceitos: estacionariedade, estabilidade, média, covariância, representações lineares, representações espectrais e medidas de dependência.

Por meio de medidas de dependência desejamos inferir, no caso de séries multivariadas, como o comportamento de uma série depende das outras. Para isso, daremos destaque ao estudo de métodos lineares multivariados e conceitos de dependência.

A motivação desta pesquisa é o estudo de sinais neurais eletrofisiológicos, tendo como características recorrentes:

- grande número de séries temporais (sinais) registradas simultaneamente,
- sinais de duração variada,
- estacionariedade apenas local,
- ausência de um modelo realista do processo gerador dos dados,
- propriedades relevantes no domínio da frequência,

- interação desconhecida entre canais .

O estudo leva em consideração essas características.

O conceito de estacionariedade é importante para estudar sinais longos, com propriedades possivelmente apenas locais. Modelos lineares são gerais o suficiente para extrair características relevantes de um processo implícito desconhecido. Propriedades e dependência no domínio da frequência explicitam características relevantes dos sinais e de suas interações em diferentes faixas do espectro.

2.1 Séries temporais estacionárias

Séries temporais são interpretadas como um conjunto de variáveis, com distribuição conjunta definida. Esta definição é excessivamente abrangente, de forma que não é possível extrair propriedades gerais de interesse para toda série temporal.

Como o elemento especial neste conjunto de variáveis é o tempo, é razoável assumirmos que a distribuição conjunta tem relação com o tempo. Uma propriedade que nos interessa, sem a qual torna-se difícil seguir adiante, é a estacionariedade, na qual as propriedades estatísticas não mudam com o tempo.

Em uma série temporal estacionária, quaisquer subconjuntos ordenados da série, com mesmas relações temporais, terão mesma distribuição estatística. Numa série estacionária, as variáveis $x(1)$, $x(2)$ e $x(999)$ têm a mesma distribuição, e os seguintes conjuntos ordenados também:

$$\{x(-10), x(-9), x(-7)\}, \{x(1), x(2), x(4)\}, \{x(10), x(11), x(13)\}. \quad (2.3)$$

Uma série temporal é definida como estacionária de segunda ordem, ou estacionária em sentido amplo, se sua média e função de autocovariância existem e não dependem do tempo (dois primeiros momentos estacionários). Uma grande parte de processos estáveis na natureza satisfazem essa propriedade.

Ao nos referirmos a séries estacionárias neste trabalho, será entendido estacionariedade de segunda ordem.

A função de autocovariância, com atraso τ , de uma série temporal x é definida por

$$\Gamma_x(t, \tau) = Cov(x(t), x(t - \tau)), \forall t, \tau \in \mathbb{Z}. \quad (2.4)$$

Em particular $\Gamma_x(t, 0) = Var(x(t)) = \sigma_x^2$ e $\Gamma_x(t, \tau) = \Gamma_x(t, -\tau)$.

Para uma série estacionária temos $\Gamma_x(\tau) = \Gamma_x(0, \tau) = \Gamma_x(t, \tau), \forall t, \tau \in \mathbb{Z}$, dependendo apenas do atraso τ . Em todos resultados teóricos deste trabalho assumimos séries estacionárias de segunda ordem, o que nos permite representar as séries por propriedades atemporais.

No caso de séries temporais multivariadas, uma série n -dimensional X é conjuntamente estacionária se a matriz de autocovariância $Cov(X(t), X(t - \tau))$ não depende de t , isto é,

$$\begin{aligned} \Gamma_X(\tau) &= Cov(X(t), X(t - \tau)) & (2.5) \\ &= \begin{bmatrix} Cov(x_1(t), x_1(t - \tau)) & \cdots & Cov(x_1(t), x_n(t - \tau)) \\ \vdots & \ddots & \vdots \\ Cov(x_n(t), x_1(t - \tau)) & \cdots & Cov(x_n(t), x_n(t - \tau)) \end{bmatrix}, \forall t, \tau \in \mathbb{Z}. \end{aligned}$$

A função de autocovariância pode ser utilizada para definir um processo estacionário de segunda ordem, incluindo suas propriedades espectrais.

Daqui em diante consideraremos apenas o caso multivariado, com $X(t)$ n -dimensional.

Simplificaremos as análises considerando processos com média nula: $X'(t) = X(t) - E[X(t)]$. Para dados finitos, com N amostras no tempo, assumiremos a estimação usual para a média: $\hat{E}[X(t)] = \frac{1}{N} \sum_{t=1}^N X(t)$.

Observe que processos gaussianos, definidos por satisfazerem que qualquer subconjunto da série tem distribuição normal multivariada, são completamente definidos por sua média e covariância. Logo, séries estacionárias de segunda ordem gaussianas são estacionárias de modo completo (estacionariedade de estatística de qualquer ordem).

2.2 Processos lineares

Neste momento desejamos inferir propriedades mais interessantes que estacionariedade e primeiros momentos. Podemos lembrar que os dados que desejamos analisar estão na natu-

reza, logo possivelmente refletem relações causais entre elementos temporalmente próximos da série.

Um exemplo simples é o processo estocástico definido por

$$X(t) = \mathbf{A}X(t-1) + \varepsilon(t), \quad (2.6)$$

em que há uma relação determinística entre elementos sequenciais da série e um elemento estocástico $\varepsilon(t)$.

Podemos considerar $\varepsilon(t)$ como um ruído branco, que denota uma série estacionária de média zero e autocovariância zero para atraso não nulo: $\Gamma_\varepsilon(\tau) = 0, \forall \tau \neq 0$. O ruído branco não é correlacionado com seu passado (nem futuro), podendo ser interpretado nesse modelo como uma inovação, que não pode ser explicada pelo passado de $X(t)$.

\mathbf{A} é uma matriz $n \times n$, que indica a relação linear determinística de cada $X(t)$ com seu passado.

Supondo que $X(t)$ seja 3-dimensional, podemos ilustrar as relações entre as séries unidimensionais

$$\begin{bmatrix} x_1(t) \\ x_2(t) \\ x_3(t) \end{bmatrix} = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix} \begin{bmatrix} x_1(t-1) \\ x_2(t-1) \\ x_3(t-1) \end{bmatrix} + \begin{bmatrix} \varepsilon_1(t) \\ \varepsilon_2(t) \\ \varepsilon_3(t) \end{bmatrix}. \quad (2.7)$$

Temos, por exemplo, $x_1(t) = a_{11}x_1(t-1) + a_{12}x_2(t-1) + a_{13}x_3(t-1) + \varepsilon_1(t)$. O termo a_{12} define, na representação AR, como x_2 influencia o futuro de x_1 . Isso ilustra como a representação AR expõe relações lineares entre as séries que compõem a série multivariada.

Definiremos $\Sigma = Var(\varepsilon(t))$. Em condições gerais, $\varepsilon(t)$ pode não ser gaussiano, quando essas representações decompõem o processo em relações lineares, mas elementos não correlacionados não são necessariamente independentes. Neste caso, análises não lineares podem revelar propriedades adicionais.

Estudaremos apenas representações lineares, sendo que inferência de relações não-lineares não serão abordadas.

A motivação de (2.6) é que a série só dependa de seu passado imediato, somado a uma variável não correlacionada. Essa classe de processos é conhecida como auto-regressiva de

ordem 1, AR(1), ou processos de Markov lineares.

Observe que $X(t)$ não tem relação direta com $X(t-2)$, apenas através de $X(t-1)$, ou seja, $Cov(X(t), X(t-2)|X(t-1)) = 0$. Mas sem condicionar em $X(t-1)$, a correlação não é nula.

Para calcular os momentos de $X(t)$, devemos escrevê-lo em função das inovações, que são os elementos com distribuição conhecida. Temos que

$$X(t) = \varepsilon(t) + A\varepsilon(t-1) + A^2\varepsilon(t-2) + \dots = \sum_{k=0}^{+\infty} A^k \varepsilon(t-k). \quad (2.8)$$

Essa representação de $X(t)$ em função das inovações passadas é denominada representação por médias móveis (MA), nesse caso de ordem infinita.

Temos $E[X(t)] = 0$ e $E[X(t)X(t)^T] = \sum_{k=0}^{+\infty} A^k \Sigma (A^k)^T$.

Podemos reescrever (2.6) com o operador de atraso D ($D^k X(t) = X(t-k)$),

$$X(t) = ADX(t) + \varepsilon(t) \implies (I - AD)X(t) = \varepsilon(t), \quad (2.9)$$

em que I é a matriz identidade $n \times n$. A representação por médias móveis se torna

$$X(t) = \sum_{k=0}^{+\infty} A^k D^k \varepsilon(t) = (I - AD)^{-1} \varepsilon(t). \quad (2.10)$$

A condição de estabilidade é definida como

$$\det(\bar{A}(z)) = \det(I - Az) \neq 0, \forall |z| \leq 1, \quad (2.11)$$

ou para representação MA

$$\det(H(z)) = \det((I - Az)^{-1}) \neq 0, \forall |z| \geq 1. \quad (2.12)$$

2.3 Representação auto-regressiva

Vimos um exemplo na seção anterior de um processo linear simples, em que um elemento depende apenas do elemento anterior e de um termo de inovação, que não tem relação com elementos passados.

Podemos gerar processos de ordens maiores $AR(p)$, para qualquer p , em que um elemento depende de p elementos passados,

$$X(t) = \sum_{k=1}^p A_k X(t-k) + \varepsilon(t), \quad (2.13)$$

tendo $Cov(x_i(t), x_j(t-p-1) | X(t-1), X(t-2), \dots, X(t-p)) = 0, \forall i, j \in \{1, \dots, n\}$.

O interessante é que processos dessa forma já podem representar séries temporais estacionárias de forma satisfatória. Isso pode ser testado empiricamente, ajustando um modelo $AR(p)$ a dados reais e testando o ajuste. Adicionalmente podemos provar que, sob condições gerais, esse processo pode representar qualquer processo estacionário, com qualquer grau de aproximação.

Satisfeitas as condições de limitação, definidas à frente em (2.25), uma série temporal estacionária X pode ser representada de forma única por AR de ordem infinita (Priestley (1994)),

$$X(t) = \sum_{k=1}^{\infty} A_k X(t-k) + \varepsilon(t), \quad (2.14)$$

sendo que $\sum_{k=1}^{\infty} |A_k| < \infty$.

Inversamente, uma série AR infinita será estacionária se for estável,

$$\det(\bar{A}(D)) = \det\left(I - \sum_{k=1}^{\infty} A_k D^k\right) \neq 0, \forall |D| \leq 1. \quad (2.15)$$

Como uma série estacionária com representação AR satisfaz $\sum_{k=1}^{\infty} |A_k| < \infty$, as matrizes A_k convergem rapidamente para zero. Dessa forma podemos ignorar termos a partir de certo k e ainda assim obter uma boa aproximação para o modelo.

Dada uma série AR , e a representação $AR(p)$ dessa série com erro quadrático $\sigma^2(p)$

mínimo, temos a aproximação assintótica por erros quadráticos,

$$p \rightarrow \infty \implies |\sigma^2(p) - \sigma^2| \rightarrow 0. \quad (2.16)$$

Erro quadrático é a denominação para a variância da inovação no contexto de previsão ótima de $X(t)$ em função dos termos passados. Nesse caso, a inovação não é necessariamente branca, pois a representação $AR(p)$ é uma aproximação.

Este é um resultado importante, que torna o $AR(p)$ uma representação suficientemente geral, e útil por ter número finito de parâmetros.

Observe que ao representar um processo desconhecido em uma representação linear, por exemplo AR , não afirmamos que este seja seu processo gerador na natureza. Por exemplo, em grande parte dos fenômenos estacionários reais, o processo gerador é contínuo e não-linear.

Mas com essa representação (discreta e linear) podemos estimar relações lineares que podem ser úteis para uma análise exploratória do mecanismo e para calcular propriedades implícitas de interesse, na ausência de conhecimento sobre o processo gerador real. Diferentemente de modelos que necessitam especificar parâmetros a priori, o $AR(p)$ é uma útil ferramenta exploratória para evidenciar relações entre séries com o mínimo de pressupostos.

2.4 Representação por médias móveis

Na representação de médias móveis, $X(t)$ depende diretamente das inovações passadas, e não dos valores passados de $X(t)$. Igualmente ao $AR(p)$, podemos definir $MA(p)$, de ordem p ,

$$X(t) = \sum_{k=0}^p H_k \varepsilon(t - k). \quad (2.17)$$

Assim $X(t)$ pode ser interpretado como um filtro causal sobre um ruído branco.

Novamente temos o importante resultado de que, dadas as condições de limitação definidas em (2.25), toda série estacionária pode ser representada de forma única por médias móveis de ordem infinita (Priestley (1994)),

$$X(t) = \sum_{k=0}^{\infty} H_k \varepsilon(t - k), \quad (2.18)$$

em que $H_0 = I$.

Observe que existe uma relação entre as representações AR e MA, se ambas existem, dada por

$$(I - \sum_{k=1}^{\infty} A_k D^k) (\sum_{k=0}^{\infty} D^k H_k) = I. \quad (2.19)$$

Dada a existência da representação AR, sempre existirá a representação MA. Dada uma representação MM, existirá a representação AR se $H(D)$ for inversível.

De modo geral, um processo AR(p), de ordem finita, terá representação MA de ordem infinita. E, inversamente, um processo MM(p), de ordem finita, terá representação AR de ordem infinita.

Outra representação relevante é a mistura do AR e MA, chamado auto-regressivo de médias móveis (ARMA), mas que em geral não possui representação única

$$X(t) = \sum_{k=1}^{\infty} A_k X(t-k) + \sum_{k=0}^{\infty} H_k \varepsilon(t-k). \quad (2.20)$$

2.5 Representação espectral

A representação espectral é útil em várias situações por ser uma técnica conhecida para revelar propriedades implícitas de funções e variáveis, especialmente propriedades de periodicidade. Também podemos representar séries estacionárias por suas propriedades na frequência, sendo uma ferramenta para definir de forma única qualquer série estacionária e compará-las entre si.

Adicionalmente, no caso de estudo em Neurociências, propriedades espectrais são centrais na caracterização de estados cognitivos, codificação de informação e comunicação entre áreas.

Seja $X(t)$ um processo estacionário, existe um processo estocástico contínuo ortogonal $Z(\lambda)$, no intervalo $(-\pi, \pi)$, tal que (Priestley (1994))

$$X(t) = \int_{-\pi}^{\pi} e^{it\lambda} dZ(\lambda), \forall t \in \mathbb{Z}, \quad (2.21)$$

conhecida como transformada de Fourier-Stieltjes.

Neste caso, definimos a ortogonalidade por meio da covariância, sendo que para todos

conjuntos S e T , tal que $S \cap T = \emptyset$, $Cov(Z(S), Z(T)) = 0$. A igualdade é válida por média quadrática.

Da mesma forma, podemos representar a função de autocovariância de um processo estacionário espectralmente

$$\Gamma(\tau) = \int_{-\pi}^{\pi} e^{i\tau\lambda} dF(\lambda), \forall \tau \in \mathbb{Z}, \quad (2.22)$$

sendo $F(\lambda)$ uma função de distribuição de probabilidades (logo, não-decrescente, $F(-\pi) = 0, F(\pi) = 1$), satisfazendo

$$Var(dZ(\lambda)) = dF(\lambda). \quad (2.23)$$

Se $F(\lambda)$ é absolutamente contínua em relação à medida de Lebesgue, podemos definir a densidade espectral $f(\lambda) = \frac{\partial F(\lambda)}{\partial \lambda}$.

Para X multivariado, $f(\lambda)$ é uma matriz $n \times n$, incluindo a densidade espectral de cada série univariada e a denominada densidade espectral cruzada entre cada par. A densidade espectral cruzada é a representação espectral da covariância entre duas séries. Por exemplo, para $X(t)$ 3-dimensional, temos

$$f_X(\lambda) = \begin{bmatrix} f_{11}(\lambda) & f_{12}(\lambda) & f_{13}(\lambda) \\ f_{21}(\lambda) & f_{22}(\lambda) & f_{23}(\lambda) \\ f_{31}(\lambda) & f_{32}(\lambda) & f_{33}(\lambda) \end{bmatrix}. \quad (2.24)$$

A densidade espectral de séries com valores reais, como estudadas nesta dissertação, satisfazem as propriedades $f_{ij}(\lambda) = f_{ji}(\lambda)^*$, $f_{ii}(-\lambda) = f_{ii}(\lambda)^*$ e $f_{ij}(\lambda) = Cov(dZ_{x_i}(\lambda), dZ_{x_j}(\lambda))$.

Denotaremos livremente $dZ_X(\lambda)$ por $X(\lambda)$ e $Cov(dZ_X(\lambda), dZ_Y(\lambda))$ por $(X(\lambda)Y(\lambda))^*$.

Ao apresentar as representações AR e MA, nos referimos a condições a serem satisfeitas para sua existência. Agora podemos definir uma condição para uma série estacionária que garante ambas representações.

Uma série estacionária n-variada $X(t)$ satisfaz a condição de limitação se sua densidade espectral f_X satisfaz

$$c_1 I_n \leq f_X(\lambda) \leq c_2 I_n, \quad (2.25)$$

para $0 < c_1 < c_2 < \infty$ e quase todo $\lambda \in [-\pi, \pi)$. Note que $A \leq B$ se $A - B$ é negativa semi-definida.

Na representação MA, vimos que $X(t)$ pode ser representado como um filtro (convolução) sobre uma série de inovações (processo com elementos ortogonais). Em termos espectrais, isso pode ser reescrito por

$$X(t) = \int_{-\pi}^{\pi} e^{it\lambda} H(\lambda) dZ_{\varepsilon}(\lambda), \quad (2.26)$$

com $H(\lambda)$ sendo a transformada de Fourier de $H(t)$, série formada pelas matrizes H_k na representação MA, calculada por

$$H(\lambda) = \sum_{k=0}^{\infty} H_k e^{ik\lambda}. \quad (2.27)$$

Referiremos à forma espectral da MA pela igualdade

$$X(\lambda) = H(\lambda)\varepsilon(\lambda). \quad (2.28)$$

Essa relação justifica $H(\lambda)$ ser conhecida como função de transferência.

De forma geral, podemos representar convoluções pela covariância da multiplicação espectral, de onde tiramos a versão espectral da representação AR,

$$X(\lambda) = A(\lambda)X(\lambda) + \varepsilon(\lambda), \quad (2.29)$$

em que $A(\lambda)$ é a transformada de Fourier da série $A(t)$, formada pelas matrizes A_k na representação AR.

A relação entre representações MA e AR pode ser descrita pela igualdade:

$$H(\lambda) = \bar{A}(\lambda)^{-1}, \quad (2.30)$$

em que $\bar{A}(\lambda) = I - A(\lambda)$, denominação utilizada doravante.

A existência das representações espectrais de $A(t)$ e $H(t)$ é garantida para séries estacionárias satisfazendo a condição de limitação (2.25).

2.6 Representação inversa

Uma representação adicional será de interesse, onde tanto o conhecimento do passado quanto o futuro são descontados das séries. Denominaremos de representação inversa (também garantida pela condição de limitação (2.25)),

$$X(t) = \sum_{k=1}^{\infty} A_k X(t-k) + \sum_{k=1}^{\infty} B_k X(t+k) + W(t). \quad (2.31)$$

Esta representação é interpretada como representação de $X(t)$ parcializada pelo passado e futuro de todas as séries.

Observe que neste caso, o processo não explicado W não é branco, não sendo equivalente ao processo não explicado ε das representações AR e MA. O processo W , normalizado por sua variância, $Var(W(0))$, tem como matriz espectral a inversa da matriz espectral de X . Definimos então o espectro inverso de X por

$$g(\lambda) = f_W(\lambda) = f_X(\lambda)^{-1}, \lambda \in [-\pi, \pi] \quad (2.32)$$

e o processo inverso associado a X ,

$${}^i X(t) = \frac{W(t)}{Var(W(0))}. \quad (2.33)$$

O processo inverso está diretamente associado à parcialização de uma série por todas as outras, retirando suas influências, como veremos adiante. Temos que o espectro parcial é dado por

$$f_{ii|x^i}(\lambda) = g_{ii}^{-1}(\lambda), \quad (2.34)$$

em que x^i denota o conjunto de todas as séries, exceto x_i . O espectro parcial cruzado é dado por

$$f_{ij|x^{i,j}}(\lambda) = g_{ij}(\lambda). \quad (2.35)$$

A representação MA de ${}^i X(t)$ está diretamente relacionada à representação AR (2.14)

de $X(t)$:

$${}^iX(t) = \sum_{k=0}^{\infty} A_k^T \eta(t+k), \quad (2.36)$$

com $\eta(t) = \Sigma^{-1}\varepsilon(t)$, relação que pode ser ilustrada pelo seguinte fato no plano espectral

$$f_X(\lambda) = H(\lambda)\Sigma H(\lambda)^H \implies f_w(\lambda) = f_x(\lambda)^{-1} = \bar{A}^H(\lambda)\Sigma^{-1}\bar{A}(\lambda), \quad (2.37)$$

em que $(\cdot)^H$ denota a matriz hermitiana.

Capítulo 3

Medidas de dependência linear

Neste capítulo analisaremos diferentes formas de abordar a dependência entre séries temporais. Apresentaremos diferentes medidas de dependência, discutindo os conceitos pressupostos por cada medida, suas formulações para séries temporais e as propriedades e limitações de cada medida.

Em particular, séries temporais apresentam características temporais, o que justifica a presença de conceitos e medidas de dependência baseadas em definições de causalidade e precedência temporal.

No caso de séries multivariadas, com mais de duas séries, desejamos abordar a dependência bivariada na presença de influência de outras variáveis, o que introduz o conceito de parcialização.

Dados temporais provenientes de ativação cerebral são conhecidos por apresentarem propriedades espectrais. Essa característica motivou o desenvolvimento na comunidade de neurociências de medidas de dependência fundamentadas na representação espectral.

Faremos uso do seguinte modelo AR 5-dimensional para discutir as diferenças qualitativas entre as medidas de dependência daqui em diante. O modelo é uma variação do exemplo 3

de Baccalá e Sameshima (2001)

$$\begin{cases} x_1(t) = 0,95\sqrt{2}x_1(t-1) - 0,9025x_1(t-2) + e_1(t) \\ x_2(t) = 0,5x_2(t-1) + e_2(t) \\ x_3(t) = -0,4x_1(t-3) + e_3(t) \\ x_4(t) = -0,5x_1(t-2) + 0,8x_2(t-1) + 0,25\sqrt{2}x_4(t-1) + 0,25\sqrt{2}x_5(t-1) + e_4(t) \\ x_5(t) = -0,25\sqrt{2}x_4(t-1) + 0,25\sqrt{2}x_5(t-1) + e_5(t) \end{cases} \quad (3.1)$$

com $\Sigma = I_5$. Podemos representar as dependências diretas existentes neste modelo pelo grafo direcionado da Figura 3.1. Cada vértice é uma série temporal, cada aresta representa uma influência direcionada direta entre pares de séries.

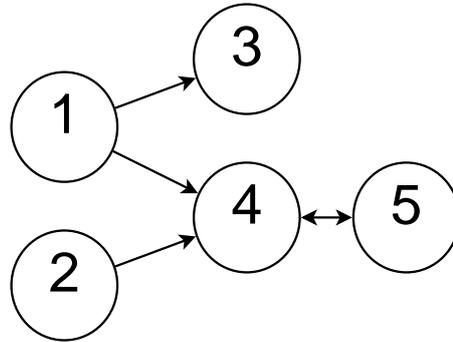


Figura 3.1: Grafo direcionado que representa as dependências diretas do modelo AR de (3.1).

3.1 Dependência e correlação

A ideia de dependência está fundamentada na distribuição de probabilidade conjunta. Duas variáveis aleatórias são (estatisticamente) dependentes se e somente se

$$P(X = x, Y = y) \neq P(X = x)P(Y = y) \quad (3.2)$$

Essa definição exige o conhecimento da distribuição conjunta das séries, que em geral não é conhecida ou não é satisfatoriamente estimável. Estatísticas mais simples, como momentos de primeira e segunda ordem (definindo a média e covariância), são utilizadas para uma definição de dependência mais fraca, porém amplamente aplicável.

Duas variáveis aleatórias A e B são correlacionadas se e somente se $Cov(A, B) \neq 0$. A

correlação é o fundamento das medidas de dependência estudadas neste trabalho.

Observe que, no caso de variáveis gaussianas, os processos são totalmente definidos por seus dois primeiros momentos. Neste caso as definições de dependência probabilística e correlação se equivalem.

Para quantificar a correlação entre duas variáveis, definimos o coeficiente de correlação de Pearson entre duas variáveis aleatórias,

$$\rho(A, B) = \frac{Cov(A, B)}{\sigma_A \sigma_B}, \quad (3.3)$$

em que $\sigma_X = \sqrt{Var(X)}$. Esta medida satisfaz $|\rho(A, B)| \leq 1$, sendo um índice para a força da relação linear entre as duas variáveis. Para variáveis reais, $\rho(A, B)$ assume tanto valores positivos quanto negativos, indicando correlação positiva ou negativa.

Definimos adicionalmente a correlação quadrática,

$$\rho^2(A, B) = |\rho(A, B)|^2, \quad (3.4)$$

que é real e positiva, satisfazendo $0 \leq \rho^2(A, B) \leq 1$.

Neste trabalho iremos analisar medidas quadráticas, baseadas em $\rho^2(A, B)$. Embora se perca informação sobre o sinal da relação, a correlação quadrática é útil por quantificar a proporção da variância de A explicada linearmente por B,

$$Var\left(A - \frac{Cov(A, B)}{Var(B)}B\right) = (1 - \rho^2(A, B))Var(A). \quad (3.5)$$

3.2 Coerência Espectral

Para séries estacionárias, podemos definir suas estatísticas independentemente do tempo, permitindo definições de dependência e correlação.

Duas séries estacionárias, X e Y , são não-correlacionadas se e somente se

$$Cov(X(t), Y(t - \tau)) = 0, \forall \tau \in \mathbb{Z}. \quad (3.6)$$

Na representação MA, a matriz de covariância pode ser calculada por

$$\Gamma(\tau) = Cov\left(\sum_{k=0}^{\infty} H_k \varepsilon(t-k), \sum_{k=0}^{\infty} H_k \varepsilon(t-\tau-k)\right) = \sum_{k=0}^{\infty} H_{k+\tau} \Sigma H_k^T, \quad (3.7)$$

em que $\Sigma = Cov(\varepsilon(t))$.

Como vimos em (2.5), a covariância Γ tem sua decomposição em frequência $f_X(\lambda)$, a partir da qual definimos uma medida de dependência na frequência,

$$Coh_{ij}(\lambda) = \frac{|f_{ij}(\lambda)|^2}{f_{ii}(\lambda)f_{jj}(\lambda)}, \quad (3.8)$$

denominada coerência espectral. A coerência espectral é simétrica e normalizada entre 0 e 1 e, com índices iguais, é trivialmente igual a 1.

Como $f_{ij} = Cov(dZ_{x_i}, dZ_{x_j})$, observamos a importante relação

$$Coh_{ij}(\lambda) = \frac{|Cov(dZ_{x_i}, dZ_{x_j})|^2}{Var(dZ_{x_i}, dZ_{x_i})Var(dZ_{x_j}, dZ_{x_j})} = \rho^2(dZ_{x_i}, dZ_{x_j}), \quad (3.9)$$

implicando que a coerência é a correlação quadrática entre as componentes espectrais de cada sinal, justificando ser interpretada como acoplamento linear entre as suas séries em cada frequência.

Na representação MA, f_{ij} é calculada por

$$f_{ij} = x_i(\lambda)x_j(\lambda)^* = h_i \Sigma h_j^H, \quad (3.10)$$

em que h_i denota a i -ésima linha de $H(\lambda)$. Calculamos então a coerência espectral,

$$Coh_{ij}(\lambda) = \frac{|h_i \Sigma h_j^H|^2}{(h_i \Sigma h_i^H)(h_j \Sigma h_j^H)}. \quad (3.11)$$

A coerência espectral e a correlação são medidas amplamente aplicadas, principalmente por sua interpretação intuitiva como acoplamento linear entre variáveis.

Uma limitação óbvia destas medidas é supor relações lineares, de forma que relações não-lineares não serão apropriadamente quantificadas. Como neste trabalho não iremos analisar

medidas não-lineares, essa limitação estará presente em todas as medidas apresentadas.

A Figura 3.2 mostra o grafo correspondente às coerências espectrais não-nulas entre as séries para o modelo (3.1). Observe como todo par de séries que tem alguma influência comum, mesmo que indireta, possui coerência espectral não nula.

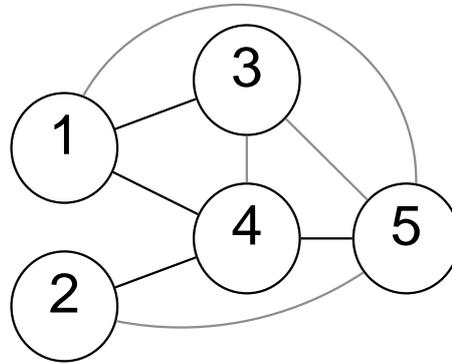


Figura 3.2: Grafo representa as coerências espectrais não nulas para o modelo (3.1). Dependências não existentes diretamente no modelo gerador estão em cinza.

A coerência espectral apresenta uma limitação adicional por não considerar a influência de outras variáveis. Tampouco permitem extrair conclusões sobre a direcionalidade da influência. O grafo direcionado mostrado na Figura 3.3 ilustra como uma variável oculta pode criar correlação entre duas variáveis sem que exista influência direta. Os vértices representam variáveis (possivelmente séries temporais) e as arestas, a dependência bivariada.

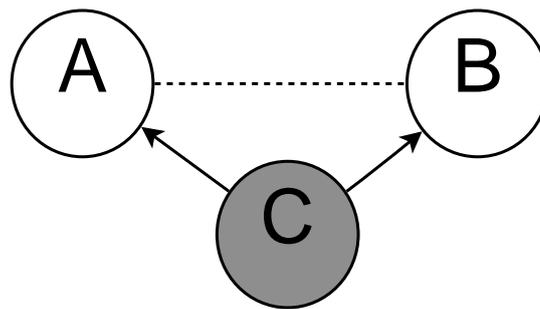


Figura 3.3: A variável oculta C influencia ambas as variáveis A e B . Ao considerar apenas a dependência entre A e B , desconsiderando a variável oculta C , haverá uma correlação não correspondente a uma interação direta no processo real.

Esse problema, conhecido como variável oculta, é recorrente no estudo de causalidade e em geral não pode ser evitado, a não ser por uma melhor inspeção do modelo suposto.

No caso em que se têm as medidas de três variáveis, podemos evidenciar que a correlação é espúria com a probabilidade condicional, como discutido a seguir.

3.3 Coerência Parcial

A coerência espectral e a correlação se baseiam na dependência linear entre duas variáveis. Na presença de variáveis adicionais, é possível que conclusões incorretas sejam obtidas, pois a dependência pode ser resultado de uma terceira variável influenciando a ambas.

A probabilidade condicional pode ser utilizada para eliminar a influência da outra variável. Definimos que X e Y são independentes, condicionadas a Z , se $P(X = x, Y = y|Z = z) = P(X = x|Z = z)P(Y = y|Z = z)$.

No contexto de relações lineares, podemos imaginar um exemplo em que X e Y dependem linearmente de Z :

$$\begin{cases} X = aZ + t \\ Y = bZ + u \end{cases} \quad (3.12)$$

em que $Cov(t, u) = Cov(Z, t) = Cov(Z, u) = 0$. Temos que $Cov(X, Y) = abVar(Z)$, mas $Cov(X - aZ, Y - bZ) = 0$.

Definimos a covariância linearmente parcial,

$$Cov(X, Y|Z) = Cov\left(X - \frac{Cov(X, Z)}{Var(Z)}Z, Y - \frac{Cov(Y, Z)}{Var(Z)}Z\right), \quad (3.13)$$

e o correspondente coeficiente de correlação parcial,

$$\rho(A, B|C) = \frac{Cov(A, B|C)}{\sqrt{Var(A|C)Var(B|C)}}. \quad (3.14)$$

Para séries estacionárias, torna-se relevante a representação inversa, definida em (2.31), que retira a dependência linear de $X(t)$ do passado e futuro das séries. A partir da definição (2.32) de $g(\lambda)$, espectro do processo inverso iX , definimos $f_{ii|X^i}$ e $f_{ij|X^i, j}$, os espectros e espectros cruzados das séries parcializadas, satisfazendo as seguintes relações

$$f_{ii|X^i}(\lambda) = g_{ii}^{-1}(\lambda) = f_{ii}(\lambda) - f_{x_i X^i}(\lambda) f_{X^i}(\lambda)^{-1} f_{X^i x_i}(\lambda) \quad (3.15)$$

e

$$f_{ij|X^i, j}(\lambda) = g_{ij}(\lambda) = f_{ij}(\lambda) - f_{x_i X^i, j}(\lambda) f_{X^i, j}(\lambda)^{-1} f_{X^i, j x_i}(\lambda). \quad (3.16)$$

Denominaremos a coerência espectral baseada no espectro inverso ${}^i X$ de coerência parcial (PC),

$$PC_{ij}(\lambda) = \frac{|g_{ij}(\lambda)|^2}{g_{ii}(\lambda)g_{jj}(\lambda)}, \quad (3.17)$$

que na representação AR pode ser definida por

$$PC_{ij}(\lambda) = \frac{|a_i^H \Sigma^{-1} a_j|^2}{(a_i^H \Sigma^{-1} a_i)(a_j^H \Sigma^{-1} a_j)}, \quad (3.18)$$

em que a_i denota a i -ésima coluna de $\bar{A}(\lambda)$. Assim como a coerência espectral, a coerência parcial é simétrica. A coerência parcial com índices iguais, PC_{ii} , é trivialmente igual a 1.

Pela sua definição, podemos interpretar a coerência parcial como uma medida de correlação para o processo inverso ${}^i X$,

$$PC_{ij}(\lambda) = \frac{|Cov(dZ_{i_{x_i}}, dZ_{i_{x_j}})|^2}{Var(dZ_{i_{x_i}}, dZ_{i_{x_i}})Var(dZ_{i_{x_j}}, dZ_{i_{x_j}})} = \rho^2(dZ_{i_{x_i}}, dZ_{i_{x_j}}). \quad (3.19)$$

A coerência parcial tem a importante propriedade de retirar influências do passado de outras séries. O grafo de dependências no modelo (3.1) para a coerência parcial é mostrado na Figura 3.4. Observe que a coerência parcial entre x_3 e x_4 é nula, pois a coerência espectral entre estes, devido à influência comum x_1 , é subtraída. Este tipo de situação justifica argumentos em favor de sua utilização em vez da coerência espectral em dados multivariados.

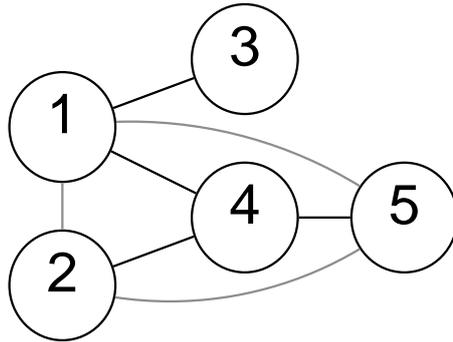


Figura 3.4: Grafo representa as coerências parciais não nulas para o modelo (3.1). Note a dependência entre 1 e 2, não existente no grafo de coerências.

Entretanto esse argumento não é inteiramente convincente. Ao nos referirmos à coerência parcial, dissemos que esta retira a influência das outras séries (parcialização). O detalhe importante é que ela retira influência do passado, mas também do futuro. Assim mesmo

que duas séries não tenham correlação nenhuma, a coerência parcial pode ser positiva, pois ambas influenciam uma mesma terceira série no futuro.

Na Figura 3.4, essa correlação espúria é ilustrada pela existência da coerência parcial entre x_1 e x_2 . Ambas as séries são não correlacionadas e sem nenhuma relação causal, mas apresentam coerência parcial por ambas influenciarem a série x_4 . O mesmo ocorre com os pares (x_1, x_5) e (x_2, x_5) , pois estes influenciam x_4 .

Esses exemplos ilustram como o uso da coerência parcial, aplicada sem conhecimento do processo gerador, fará inserir dependências não existentes na coerência espectral. Esse é o efeito oposto ao esperado de retirar dependências não relativas a relações causais.

Esta propriedade limita sua utilidade para situações em que não se tem informação a priori sobre a direcionalidade das dependências.

Para tentar contornar esta limitação, necessitamos de uma medida que tenha parcialização apenas em relação ao passado das séries.

3.4 Conceitos de Causalidade

Introduziremos medidas que considerem a ordem temporal, motivados pela suposição de que os dados foram gerados por modelos causais, em que o estado presente depende do passado. Para isso, torna-se necessário ter definições de causalidade que fundamentem as interpretações.

Adotam-se diversas definições de causalidade nas ciências, sem um consenso universal, gerando debates em áreas como filosofia, física, direito, saúde, ciências sociais e metodologia científica. Pearl (2000) traz uma extensa análise de modelos e inferências de causalidade.

Muitos argumentos teóricos envolvem condições de necessidade e suficiência (B só acontece se A ocorrer), que são muito restritivos, de forma que temos interesse na interpretação que utiliza condicionalidade probabilística.

Entre propriedades esperadas para determinação de causalidade de A em B, podemos citar algumas, estreitamente relacionadas entre si:

1. Ação direta de A em B dado o modelo assumido: bola de bilhar encosta em outra bola de bilhar, causando seu movimento, segundo modelo de física de corpos rígidos.

2. Condicionalidade do estado de B em relação ao estado de A: a probabilidade de ter câncer de pulmão é maior se fumar.
3. Mudanças em A acompanham mudanças em B: emissão de gases poluentes na atmosfera acompanha aumento da temperatura global.
4. Mudanças em A precedem temporalmente (prevêm) mudanças em B: diminuição do desemprego aumenta índice de bem estar.
5. Possibilidade de manipulação do estado de A e simultânea medição do estado de B, a fim de comprovar a dependência.

Vemos como o conceito de causalidade pode ser aplicado a diversas situações, muito distintas entre si, o que dificulta definições universais. Observe que o item 3 tem relação com o conceito de correlação, de forma que sozinho, não é possível inferir se A influencia B ou B que influencia A, a menos que um modelo plausível seja assumido. É muito comum uso inadequado dos termos, referindo-se à causalidade no lugar da correlação, sem explicitar um modelo plausível para justificação.

3.4.1 Causalidade de Granger

Para séries temporais, a representação AR sugere relações causais, pois eventos do presente dependem deterministicamente de eventos do passado. Por exemplo, considere um processo bivariado AR(1):

$$\begin{cases} x_1(t) = 0,5x_1(t-1) + 0,3x_2(t-1) + \varepsilon_1(t) \\ x_2(t) = 0,4x_2(t-1) + \varepsilon_2(t), \end{cases} \quad (3.20)$$

com $\Sigma = I_2$.

Vemos que nesta representação x_1 depende de valores do passado de x_2 , mas x_2 depende apenas do seu próprio passado e não depende de valores do passado de x_1 . Essas propriedades assimétricas inspiraram a definição de uma causalidade operacional, introduzida por [Granger \(1969\)](#), por isso conhecida por causalidade de Granger (*GC*).

A causalidade de Granger se inspira na previsão e ordem temporal. Para duas séries temporais $x(t)$ e $y(t)$, $x(t)$ *Granger-causa* $y(t)$ se o conhecimento dos valores passados de $x(t)$ melhoram a previsão do valor presente de $y(t)$, isto é,

$$Var(y(t)|x^{t-}, y^{t-}) < Var(y(t)|y^{t-}) \Rightarrow (x(t) \xrightarrow{GC} y(t)). \quad (3.21)$$

Nesta definição, a variância da previsão de $y(t)$ é menor quando, além dos valores passados de $y(t)$ (descrito por y^{t-}), usa-se informação do passado de $x(t)$ (x^{t-}).

Note que a causalidade de Granger não é recíproca, o que fica claramente ilustrado no exemplo (3.20), evidenciando sua interpretação como uma relação direcionada.

Esta definição desconsidera um modelo implícito às séries. Exige apenas uma ordem temporal e condicionalidade, de forma que os valores passados de um influenciem na previsão do valor presente de outro.

Em relação às características de um modelo causal citadas anteriormente, estaria inspirada nos itens 2) e 4). Não assume itens 1) e 5) (ação direta e possibilidade de intervenção), sendo útil para casos em que não se assume um conhecimento do processo gerador, e lida-se com dados já coletados fixos, não passíveis de manipulação.

O conceito de causalidade de Granger pode ser aplicado imediatamente ao modelo AR, pois as relações lineares utilizam dados passados para reduzir a variância do componente estocástico do presente. Isso equivale a minimizar o erro quadrático (variância) do preditor linear do valor presente.

No exemplo (3.20), temos que $a_{12} \neq 0$. No caso de não termos conhecimento sobre o passado de x_2 , é equivalente a modelarmos x_1 de forma univariada, ou seja, com $a_{12} = 0$,

$$x_1(t) = b_1 x_1(t-1) + b_2 x_1(t-2) + \dots + \nu_1(t). \quad (3.22)$$

É sabido que o modelo original é equivalente ao melhor preditor de x_1 , tendo σ_{11} mínimo. Teremos que $Var(\nu_1) > Var(\varepsilon_1)$, o que satisfaz a condição para causalidade de Granger, concluindo que x_2 *Granger-causa* x_1 .

No entanto x_2 depende apenas de seu próprio passado (x_2^{t-}), de forma que, condicionado a x_2^{t-} , não é correlacionado ao passado de x_1 . Teremos que $Var(y(t)|x^{t-}, y^{t-}) = Var(y(t)|y^{t-})$,

não havendo causalidade de Granger, $x_1 \not\rightarrow^{GC} x_2$.

3.4.2 Causalidade de Granger condicionada

Para um modelo com mais de duas séries necessitamos de maior cuidado. Como discutido anteriormente, é possível que a suposta causalidade seja devida a uma terceira variável oculta.

No caso da GC, devemos incluir o conhecimento de todas séries do sistema, o que em nossa interpretação probabilística implica calcular relações condicionadas ao passado de todas as séries.

Dadas n séries temporais, $x_1(t), x_2(t), \dots, x_n(t)$, expandimos o conceito de GC para

$$Var(x_i(t)|X^{t-}) < Var(x_i(t)|X^{t-} \setminus x_j^{t-}) \Rightarrow (x_j(t) \xrightarrow{GC} x_i(t)), \quad (3.23)$$

em que X^{t-} é o passado de todas as n séries. Assim, o conhecimento de x_j^{t-} , quando acrescentado ao conhecimento dos valores passados de todas as outras séries, ajuda na previsão de $x_i(t)$.

Esta definição é um caso particular da definição geral de GC,

$$Var(x_i(t)|\Omega^{t-}) < Var(x_i(t)|\Omega^{t-} \setminus x_j^{t-}) \Rightarrow (x_j(t) \xrightarrow{GC} x_i(t)), \quad (3.24)$$

em que Ω^{t-} representa toda informação disponível do universo sobre o passado. Como no nosso caso temos apenas as informações de nossas séries, contentamo-nos com a definição (3.23).

Para nossa representação AR multivariada, novamente temos uma interpretação direta de causalidade de Granger. Como a representação AR já considera todas as séries, se $(A_k)_{ij} = 0$ para todo k , quer dizer que $x_i(t)$ não é correlacionado a $x_j(t-k)$, para todo k , condicionado ao passado das outras séries.

É importante notar como o condicionamento é em relação apenas ao passado das séries, diferentemente da coerência parcial.

Inversamente, se $(A_k)_{ij} \neq 0$, para algum k , se representarmos as séries por um modelo reduzido, sem a série x_j , teremos uma variância maior no erro de predição de x_i .

Definimos então a causalidade de Granger para representações AR como

$$\exists k \in \mathbb{Z}^+, (A_k)_{ij} \neq 0 \Leftrightarrow (x_j(t) \xrightarrow{GC} x_i(t)). \quad (3.25)$$

O conceito de causalidade de Granger para AR satisfaz o primeiro quesito de causalidade listado anteriormente, pois de acordo com o modelo assumido (AR), há relação direta entre os processos. Mas, mais rigorosamente, sabemos que o modelo assumido pode não ter relação com o processo real, de forma que interpretações devem ser feitas com cautela.

A causalidade de Granger se mostra então uma definição útil, pois respeita condicionalidade, temporalidade e ação direta na representação AR.

No caso do estudo de fisiologia do cérebro, é extremamente difícil definir um conceito de causalidade entre os sinais neurais. Além de ser uma rede complexa de numerosos neurônios interconectados, têm-se em geral apenas informações de parte deles e muitas vezes de forma indireta. Resultados devem ser interpretados com precaução e pensados como uma evidência exploratória.

De todo modo, qualquer outro método de análise irá sofrer dos mesmos problemas, não sendo uma característica própria deste conceito ou das representações lineares. Assim sendo, considera-se que o conceito de causalidade é útil, e até necessário, para substituir análises fundamentadas em correlações, que não consideram relações temporais.

3.4.3 Causalidade indireta de Akaike

Iniciamos aqui uma discussão sobre definições alternativas de causalidade.

A questão central é diferença entre ação direta e indireta. Suponha um modelo de três séries (Figura 3.5), em que x_3 influencia x_2 , que por sua vez influencia x_1 . Podemos afirmar que x_3 causa x_1 ?

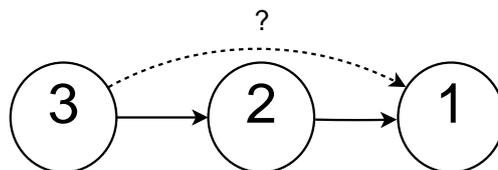


Figura 3.5: Causalidade indireta e direta. Definição de causalidade deve definir se há causalidade de 3 para 1.

Essa é uma pergunta conceitual, e dependerá apenas da definição de causalidade escolhida. A definição de causalidade de Granger estabelece que devemos condicionar a relação de x_3 e x_1 a x_2 . Dessa forma não haverá GC de x_3 para x_1 . Isso está relacionado à ação direta, assim mesmo que a influência de x_3 altere o futuro de x_1 , será de forma indireta, através de x_2 .

Uma definição alternativa, que define causalidade indireta, surge ao analisarmos a representação MA. Suponha o modelo simples com representação AR:

$$\begin{cases} x_1(t) = x_2(t-1) + \varepsilon_1(t) \\ x_2(t) = x_3(t-1) + \varepsilon_2(t) \\ x_3(t) = \varepsilon_3(t). \end{cases} \quad (3.26)$$

Claramente ele satisfaz o grafo de causalidade da Figura 3.5. Teremos a seguinte representação MA correspondente:

$$\begin{cases} x_1(t) = \varepsilon_1(t) + \varepsilon_2(t-1) + \varepsilon_3(t-2) \\ x_2(t) = \varepsilon_2(t) + \varepsilon_3(t-1) \\ x_3(t) = \varepsilon_3(t). \end{cases} \quad (3.27)$$

Na representação MA, temos a dependência sobre inovações passadas de série. Nesta perspectiva, sabemos a origem de cada alteração nas variáveis. No sentido de causalidade que aceita ações indiretas, de forma a saber a origem das alterações, deveríamos afirmar que x_3 causa x_1 , pois uma alteração em x_3 irá influenciar x_1 .

Esta definição está relacionada à resposta ao impulso da função de transferência de x_i para x_j , e iremos denominá-la causalidade de Akaike (AC), por ter relação com a medida definida em Akaike (1968) como contribuição relativa em potência. Hosoya (2001) sugeriu esta alternativa como o conceito recomendado para definir dependências. Um argumento é exposto a seguir.

Considere o seguinte modelo, sugerido por Hosoya (2001) e comentado por Takahashi

(2008), com a representação AR dada por

$$\begin{cases} X(t) = -0,25Y(t-2) + 0,5Z(t-1) + \varepsilon(t) \\ Y(t) = \xi(t) \\ Z(t) = 0,5Y(t-1) + \nu(t) \end{cases} \quad (3.28)$$

e representação MA por

$$\begin{cases} X(t) = \varepsilon(t) + 0,5\nu(t) \\ Y(t) = \xi(t) \\ Z(t) = \nu(t) + 0,5\xi(t). \end{cases} \quad (3.29)$$

Observamos a peculiar situação em que, embora Y tenha ação direta em X (causalidade de Granger), não há causalidade de Akaike. A ação das inovações de Y sobre X são canceladas com a ação indireta através de Z, assim X e Y são não correlacionados. Logo um impulso em Y (ε_Y) não tem influência sobre X.

Hosoya argumenta que esse caso torna a causalidade de Granger falha, pois identifica uma causalidade que não ocorre. Como foi notado em [Takahashi \(2008\)](#), a diferença não está na representação AR ou MA, mas na escolha do condicionamento. Assim a causalidade de Granger estuda a relação de X(t) e o passado de Y(t), condicionado ao passado de Z(t), enquanto a causalidade de Akaike estuda a relação de X(t) e a inovação $\xi(t)$, sem condicionamento a outras séries.

Como argumento adicional, este exemplo de Hosoya é um caso extremamente particular, e em geral a causalidade de Granger implica a causalidade de Akaike. Nesse modelo particular, há relação indireta de Y para X, mas não há AC, assim formalmente a definição de AC é da correlação entre uma série e a inovação de outra. Em geral, relações indiretas (e diretas) implicam a presença de AC.

Podemos definir a causalidade de Akaike, para uma representação MA, como

$$\exists k \in \mathbb{Z}^k, (H_k)_{ij} \neq 0 \Leftrightarrow (x_j(t) \xrightarrow{AC} x_i(t)). \quad (3.30)$$

Baseadas na definição de GC e AC, diversas medidas de influência causal foram definidas, com muitas controvérsias sobre qual seria a medida que estaria mais de acordo com a defini-

ção de causalidade; e circularmente qual seria a melhor definição de causalidade. [Takahashi \(2008\)](#) estuda o fundamento dessas medidas e a surpreendente relação com medidas de correlação e informação mútua.

3.5 Coerência Direcionada

A causalidade de Akaike motiva uma medida de dependência direcionada, que relacione uma série à inovação de outra. Podemos definir uma medida como a correlação entre essas duas quantidades, ou seja, $\rho^2(dZ_{x_i}(\lambda), dZ_{\varepsilon_j}(\lambda))$.

Apesar de ser uma medida válida, observamos que, se existe correlação entre as inovações (Σ não diagonal), por exemplo entre ε_j e ε_k ($\sigma_{jk} \neq 0$), essa medida estará considerando uma correlação que está presente também em $\rho^2(dZ_{x_i}(\lambda), dZ_{\varepsilon_k}(\lambda))$. Isso motiva uma parcialização da inovação ε_j pelas outras inovações, que denotaremos η_j . Calculamos que

$$Var(\eta_j(t)) = Var(\varepsilon_j(t)|\varepsilon^j(t)) = \sigma_{jj} - Cov(\varepsilon_j \varepsilon^j) Var(\varepsilon^j)^{-1} Cov(\varepsilon^j \varepsilon_j). \quad (3.31)$$

Observe que $\eta_j(t) = \varepsilon_j(t)$ para Σ diagonal.

Definimos assim a coerência direcionada (DC) como

$$DC_{ij}(\lambda) = \rho^2(dZ_{x_i}(\lambda), dZ_{\eta_j}(\lambda)). \quad (3.32)$$

Notamos na representação MA que o espectro de x_i é uma função linear das inovações,

$$x_i(\lambda) = h_{i1}(\lambda)\varepsilon_1(\lambda) + h_{i2}(\lambda)\varepsilon_2(\lambda) + \dots + h_{in}(\lambda)\varepsilon_n(\lambda), \quad (3.33)$$

o que implica

$$DC_{ij}(\lambda) = \frac{|h_{ij}|^2 Var(\eta_j)}{f_{x_i}} = \frac{|h_{ij}|^2 Var(\eta_j)}{h_i \Sigma h_i^H}. \quad (3.34)$$

Na decomposição acima, fica clara a interpretação de DC como proporção do espectro de x_i explicada exclusivamente por ε_j , sendo h_{ij} a função de transferência entre as duas variáveis. Essa decomposição motiva definições de medidas alternativas.

Podemos decompor $f_{ii}(\lambda) = XX^*$ em

$$f_{ii}(\lambda) = |h_{i1}|^2\sigma_{11} + |h_{i2}|^2\sigma_{22} + \dots + |h_{in}|^2\sigma_{nn} + \sum_{j \neq k} h_{ij}h_{ik}^*\sigma_{jk}. \quad (3.35)$$

Vemos nessa decomposição como o termo $|h_{ij}|^2\sigma_{jj}$ é a contribuição de ε_j , desconsiderando sua interação com a contribuição de outras inovações. Podemos assim definir a transferência direcionada espectral (SDT), sendo a contribuição em potência espectral, normalizada por todas contribuições,

$$SDT_{ij}(\lambda) = \frac{|h_{ij}|^2\sigma_{jj}}{\sum_k |h_{ik}|^2\sigma_{kk}}. \quad (3.36)$$

Outra medida espectral para séries multivariadas, relacionada à causalidade de Akaike, foi apresentada por Kaminski e Blinowska (1991) como função de transferência direcionada quadrática (DTF),

$$DTF_{ij}(\lambda) = \frac{h_{ij}}{h_i h_i^H}. \quad (3.37)$$

A DTF equivale à DC quando $\Sigma = kI_n, k \in \mathbb{R}$. Apesar de evitar consideração sobre a correlação entre inovações, é sensível à diferença de escala entre as séries por não considerar Σ .

A DTF e SDT têm a propriedade de serem normalizadas pelo espectro do influenciado x_i , satisfazendo $\sum_j DTF_{ij} = 1$ e $\sum_j SDT_{ij} = 1$.

Podemos ver a DC como uma decomposição da coerência espectral em duas direções, evidenciada pela representação MA. Considerando Σ diagonal, e a coerência espectral e DC não-quadráticas, temos

$$Coh_{ij} = \frac{f_{ij}}{\sqrt{f_{ii}f_{jj}}} = \frac{h_i\sqrt{\Sigma}}{\sqrt{f_{ii}}} \frac{\sqrt{\Sigma}h_j^H}{\sqrt{f_{jj}}} = DC_i DC_j^H, \quad (3.38)$$

em que $DC_i = [DC_{i1} DC_{i2} \dots DC_{in}]$.

Todas essas medidas, DC, SDT e DTF, têm correspondência com a causalidade de Akaike, tendo a existência de dependência condicionada ao termo h_{ij} . Essas medidas detectam dependência direcionadas indiretas, de forma que o grafo para a DTF para o modelo de referência (3.1) é mostrado na Figura 3.6. Observe como a influência indireta entre de x_1 para x_4 e

x_2 para x_4 são detectadas. Em comparação com a coerência espectral, correlações devido à influência comum, como entre x_3 e x_4 , são nulas.

A medida para índices iguais (p.ex. DC_{ii}) é a proporção explicada pela própria inovação.

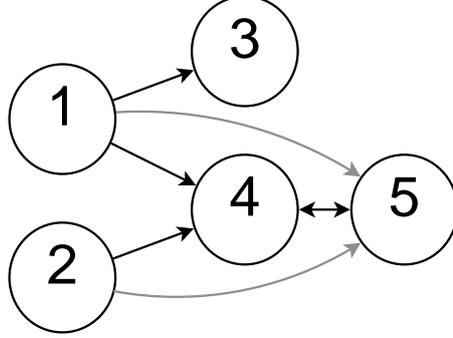


Figura 3.6: Grafo representa as coerências direcionadas não nulas para o modelo (3.1).

3.6 Coerência Parcial Direcionada

Podemos definir uma medida semelhante à coerência direcionada para o conceito de causalidade de Granger. De modo análogo, iremos definir a dependência entre as séries parcializadas, derivadas do processo inverso iX , e sua inovação $\nu(t) = \Sigma^{-1}\varepsilon(t)$, parcializada pelas outras inovações. Definimos assim a coerência parcial direcionada informacional (Takahashi *et al.* (2010)),

$${}^iPDC_{ij}(\lambda) = \rho^2(dZ_{i x_i}, dZ_{i \eta_j}), \quad (3.39)$$

em que ${}^i\eta_j$ é a inovação parcializada do processo inverso. Temos que $Var({}^i\eta_j) = \sigma_{jj}^{-1}$ (Takahashi (2008)) e que o espectro inverso é o inverso do espectro parcializado, $g_i(\lambda) = f_{ii|x^i}^{-1}(\lambda)$.

Dada a relação entre a representação AR de X e a MA de iX (2.36), concluímos que

$${}^iPDC_{ij}(\lambda) = \frac{|a_{ij}|^2 \sigma_{jj}^{-1}}{g_{ii}(\lambda)} = \frac{|a_{ij}|^2 \sigma_{ii}^{-1}}{a_j^H \Sigma^{-1} a_j}, \quad (3.40)$$

em que a_j é a j -ésima coluna de $\bar{A}(\lambda)$. A denominação *informacional* é usada para diferenciar da PDC originalmente proposta, e se justifica pelo fato de, sendo uma relação de correlação ρ , tem uma justificativa na Teoria da Informação quando as séries são gaussianas (Takahashi (2008)).

Novamente, temos definições alternativas na literatura, incluindo a PDC originalmente proposto (Baccalá e Sameshima (2001)),

$$PDC_{ij} = \frac{|a_{ij}|^2}{a_j^H a_j}, \quad (3.41)$$

e a PDC generalizada (gPDC), que utiliza apenas a diagonal de Σ (Baccala *et al.* (2007)),

$$gPDC_{ij} = \frac{|a_{ij}|^2 \sigma_{ii}^{-1}}{a_j^H \Sigma_d^{-1} a_j}, \quad (3.42)$$

em que $\Sigma_d^{-1} = (\Sigma \odot I_n)^{-1}$, sendo a inversa da diagonal da matriz de covariâncias. \odot representa o produto de Hadamard.

Tanto PDC quanto gPDC satisfazem $\sum_i PDC_{ij} = 1$. Observe que a normalização do PDC é pela série *fonte* x_j , diferentemente das medidas DC e DTF. iPDC pode ser interpretada como o acoplamento linear direto entre as séries (termo a_{ij}), normalizado por todos acoplamentos que têm x_j como fonte.

Assim como a DTF, a PDC original é sensível a mudanças de escala entre as séries, o que motivou a definição da gPDC em Baccala *et al.* (2007).

A medida para índices iguais (p.ex. PDC_{ii}) é a proporção do espectro inverso não explicada pelas outras séries.

A PDC, em todas suas formulações, detecta dependência baseada em a_{ij} , de forma que tem correspondência direta com a causalidade de Granger. Essas medidas detectam dependência direcionadas *diretas*, de forma que o grafo para a PDC para o modelo de referência (3.1) é mostrado na Figura 3.7. As dependências detectadas pelo PDC são as mesmas que as representadas na descrição original do modelo (Figura 3.1). Isso é imediato, pois assumimos o modelo AR como modelo gerador.

A PDC pode ser interpretada como uma decomposição em duas direções da coerência parcial, em suas representações não quadráticas. Para a PDC informacional, com Σ diagonal, temos

$$PC_{ij} = \frac{g_{ij}}{\sqrt{g_{ii}g_{jj}}} = \frac{a_j^H \sqrt{\Sigma^{-1}} \sqrt{\Sigma^{-1}} a_i}{\sqrt{g_{jj}} \sqrt{g_{ii}}} = iPDC_j iPDC_i^H, \quad (3.43)$$

em que $iPDC_j = [iPDC_{1j} iPDC_{2j} \dots iPDC_{nj}]$.

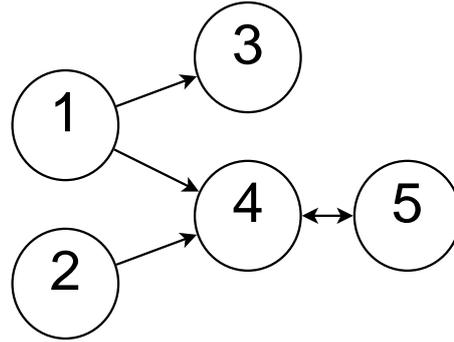


Figura 3.7: Grafo representa as coerências parciais direcionadas não nulas para o modelo (3.1). As relações de dependências são as mesmas que as dependências diretas no processo gerador.

Na literatura há a discussão sobre a relação entre causalidades de Granger e de Akaike, e em particular sobre qual das medidas, PDC ou DTF, seria a mais apropriada para aplicações em neurociências.

Para sistemas de apenas duas variáveis, em que não há parcialização por outras séries, as duas definições são equivalentes, valendo

$$n = 2 \implies {}_iPDC_{ij} = DC_{ij}. \quad (3.44)$$

Por apresentarem fundamento em definições de causalidade diferentes, nenhuma é incorreta, sendo alternativas dependendo do que se deseja inferir. A normalização diferenciada (por fonte ou alvo) torna seus valores bastante distintos.

Desconsiderando a diferença de normalização, a PDC se mostra uma medida mais útil em comparação ao DTF por detectar apenas as dependências diretas, juntando as características de direcionalidade e parcialização. Mesmo contendo a mesma informação, a PDC apresenta as dependências de forma mais explícita, para imediata interpretação.

No estudo funcional do cérebro, por exemplo, é possível imaginar que todas áreas apresentem conexões indiretas com todas outras, tornando a existência de causalidade de Akaike trivial, logo irrelevante. A PDC pode inferir evidências de regiões que têm conexão direta, podendo ser a evidência funcional correspondente a conexões estruturais.

Para a PDC detectar dependências diretas, assumimos que o modelo AR de fato é um bom modelo para o processo gerador das séries. Mas a mesma condição vale para o DTF e todas outras medidas lineares deste trabalho, não sendo um argumento diferencial nesta

comparação.

3.6.1 Transferência Parcial Direcionada

A PDC apresenta propriedades de parcialidade e direcionalidade, tendo condições de existência equivalente à existência da causalidade de Granger. Essas propriedades sustentam a PDC como a medida linear, entre as apresentadas, com maior informação sobre a estrutura geradora do processo.

A PDC pode ser vista como decomposição da coerência parcial, que retira a influência do passado e futuro das outras séries na equação (3.43). Essa decomposição fundamenta a normalização pela fonte x_j . Essa normalização gera casos em que a medida pode contrariar a expectativa e causar problemas de interpretação sobre o valor quantitativo da PDC.

Considere o seguinte processo AR:

$$\begin{cases} x_1(t) = 0,5x_1(t-1) + 1,0x_2(t-1) + \varepsilon_1(t) \\ x_2(t) = \varepsilon_2(t) \\ x_3(t) = \beta x_2(t-1) + \varepsilon_3(t) \end{cases} \quad (3.45)$$

com $\Sigma = I_3$.

Observamos como x_2 tem grande influência sobre x_1 . Podemos quantificar esse fato pela coerência direcionada, que mede a proporção do espectro de x_1 explicada pela inovação ε_2 , calculando $DC_{12}(\lambda = 0) = 0,5$. Para $\beta = 0$, a PDC será equivalente, $PDC_{12}(\lambda = 0) = 0,5$.

No entanto, ao introduzirmos uma interação de x_2 para x_3 , poderíamos esperar que a dependência entre x_1 e x_2 não fosse alterada. Por exemplo, x_3 pode ser uma segunda medição, com amplificação e atraso, do sinal x_2 . Mas para $\beta = 10$, temos $PDC_{12}(0) < 0,01$, podendo ser interpretada como conexão espúria.

Notamos que, com o auxílio das estatísticas calculadas adiante, pode-se inferir que $PDC_{12}(0)$ é estatisticamente significativa em qualquer dos casos. Mas seu valor baixo pode dificultar a interpretação acerca da força de dependência.

Com essa motivação, introduzimos aqui uma nova medida de dependência. Desejamos preservar a parcialização e direcionalidade, condicionando a dependência nos termos a_{ij} .

Por outro lado, desejamos ser possível uma normalização pelo alvo da dependência, com a interpretação como proporção do espectro explicado pela dependência.

Temos a decomposição espectral de $x_1(t)$ na representação AR

$$x_i(\lambda) = a_{i1}x_1(\lambda) + a_{i2}x_2(\lambda) + \dots + a_{in}x_n(\lambda) + \varepsilon_i(\lambda), \quad (3.46)$$

da qual podemos calcular sua densidade espectral

$$f_{ii}(\lambda) = x_i x_i^* = |a_{i1}|^2 f_{11} + |a_{i2}|^2 f_{22} + \dots + |a_{in}|^2 f_{nn} + \sigma_{ii} + \sum_{j \neq k} a_{ij} a_{ik}^* f_{jk} + \sum_j (\varepsilon_i a_{ij}^* x_j^* + a_{ij} x_j \varepsilon_i^*). \quad (3.47)$$

As duas somas ao final consideram as interações entre as séries, e entre séries e a inovação, que estão influenciando x_i . O mesmo efeito ocorre na decomposição do espectro em termo de inovações na representação MA na expressão (3.35), quando há correlação entre as inovações.

Considerando apenas o que é transmitido em potência espectral por x_j para x_i , antes de sua interação com as outras séries, definimos a transferência parcial direcionada (PDT),

$$PDT_{ij} = \frac{|a_{ij}|^2 f_{jj}}{\sum_k |a_{ik}|^2 f_{kk} + \sigma_{ii}}, \quad (3.48)$$

observando que a_{ij} é a representação espectral da sequência $(A_k)_{ij}$, para $k > 0$, desconsiderando a matriz $A_0 = I$, presente na definição da PC e PDC.

Por completeza, definiremos a contribuição instantânea da inovação como transferência parcial instantânea: $IPT = \frac{\sigma_{ii}}{\sum_k |a_{ik}|^2 f_{kk} + \sigma_{ii}}$. A PDT satisfaz $\sum_j PDT_{ij} + IPT_i = 1$.

A PDT pode ser interpretada como potência espectral medida no canal de transmissão de x_j para x_i , normalizada por todas transmissões. A PDT é normalizada pelo alvo e não tem o mesmo efeito que a PDC no modelo acima, em que $PDT_{12}(0) = 0.25$ para qualquer β .

Para índices iguais (PDT_{ii}), a PDT releva a potência espectral do sinal de realimentação de x_i .

Tendo interpretação como potência espectral transmitida, sua versão sem normalização pode ser útil quando se deseja saber a transmissão total, sem depender das outras transmis-

sões. Definimos a transferência parcial direcionada total (TPDT),

$$TPDT_{ij} = |a_{ij}|^2 f_{jj}. \quad (3.49)$$

A PDT não pode ser representada trivialmente como uma correlação de variáveis. O motivo é que sua definição não se baseia na parcialização das séries. A PDT usa a decomposição de uma série como combinação das outras (representação AR). Apesar de relevante, a decomposição em séries não ortogonais dificulta a interpretação por correlação, assim como a decomposição de vetores em bases não ortogonais dificulta sua representação por produto interno.

3.7 Comparação entre medidas de dependência

Sumarizamos na Tabela 3.1 as propriedades de cada uma das medidas apresentadas, para facilitar o entendimento e a escolha da medida mais adequada para cada necessidade.

A coerência espectral, medida mais conhecida e amplamente usada, baseia-se simplesmente no espectro cruzado bivariado, não utilizando informações de outras séries. Essa restrição implica não apresentar nenhuma das propriedades citadas. Podemos apresentá-la pela correlação $\rho^2(x_i(\lambda), x_j(\lambda))$.

A coerência parcial retira a influência comum das outras séries, sendo formulada a partir do espectro inverso das séries, a qual é diretamente relacionada ao espectro parcializado. Esta é útil quando se tem informação a priori de quais séries exercem tal influência comum. Caso contrário, por retirar também informações futuras, irá, em geral, inserir dependências não existentes. Essa característica faz da coerência parcial uma escolha duvidosa para análises exploratórias. Podemos apresentá-la pela correlação $\rho^2({}^i x_i(\lambda), {}^i x_j(\lambda))$.

A coerência parcial direcionada (e suas variantes, gPDC e iPDC) tem condição equivalente à existência de causalidade de Granger. Este conceito de causalidade considera apenas relações com direcionalidade temporal e parcializadas por todas outras séries, resultando em dependências direcionadas e diretas, desconsiderando relações indiretas. A PDC pode ser interpretada como uma decomposição da coerência parcial nas duas direções de influência,

podendo ser apresentada pela correlação entre inovação e série parciais, $\rho^2({}^i\eta_i(\lambda), {}^i x_j(\lambda))$.

Foi introduzida uma nova medida de dependência, a transferência parcial direcionada (PDT), que também se fundamenta sobre a representação AR e causalidade de Granger. A PDT se diferencia da PDC por não utilizar o espectro inverso, mas sim uma decomposição do espectro direto. A PDT pode ser interpretada como a potência espectral medida na transmissão do canal x_j para o canal x_i , ignorando como esta interfere com as outras transmissões. Desta forma adiciona a importante propriedade de ser normalizada pelo alvo, enquanto a PDC é normalizada pela fonte, gerando resultados possivelmente não desejados.

Notamos como entre os diferentes conceitos de dependência, a causalidade de Granger apresenta o maior conjunto de propriedades desejáveis discutidas. Entre suas formulações como medida, destacamos a *iPDC* por ser fundamentada na correlação quadrática entre variáveis, sendo invariante à escala, e a *gPDC* por apresentar a interessante propriedade $\sum_i gPDC_{ij} = 1$.

Também destacamos como a nova medida proposta, PDT, apresenta a propriedade adicional importante de ser normalizada por o que recebe, sendo uma medida para a proporção do espectro total explicado por cada dependência.

Tabela 3.1: Sumário das propriedades e formulações de cada conceito e medida de dependência.

Conceito	Correlação	Correlação Parcial	causalidade de Akaike	causalidade de Granger					
Condição de existência	$f_{ij} \neq 0$	$g_{ij} \neq 0$	$h_{ij} \neq 0$	$a_{ij} \neq 0$					
Sem influência comum	Não	Sim ^a	Sim	Sim					
Direcionalidade	Não	Não	Sim	Sim					
Relações diretas	Não	Sim ^b	Não	Sim					
Grafo para modelo 3.1									
Medida espectral	Coerência	Coerência Parcial	DTF	SDT	DC	PDC	gPDC	iPDC	PDT
Formulação AR/MA	$\frac{ h_i \Sigma h_j^H ^2}{(h_i \Sigma h_i^H)(h_j \Sigma h_j^H)}$	$\frac{ a_i^H \Sigma^{-1} a_i }{(a_i^H \Sigma^{-1} a_i)(a_j^H \Sigma^{-1} a_j)}$	$\frac{h_{ij}}{\sum_k h_{ik} ^2}$	$\frac{ h_{ij} ^2 \sigma_{jj}}{\sum_k h_{ik} ^2 \sigma_{kk}}$	$\frac{ h_{ij} ^2 Var(\eta_j)}{h_i \Sigma h_i^H}$	$\frac{ a_{ij} ^2}{a_j^H a_j}$	$\frac{ a_{ij} ^2 \sigma_{ii}^{-1}}{a_j^H \Sigma^{-1} a_j}$	$\frac{ a_{ij} ^2 \sigma_{ii}^{-1}}{a_j^H \Sigma^{-1} a_j}$	$\frac{ a_{ij} ^2 f_{jj}}{\sum_k a_{ik} ^2 f_{kk}}$
Formulação outra	$\frac{ f_{ij} ^2}{f_{ii} f_{jj}}$	$\frac{ g_{ij} ^2}{g_{ii} g_{jj}}$	-	-	$\frac{f_{ii}}{ h_{ij} ^2 Var(\eta_j)}$	-	-	$\frac{g_i}{ a_{ij} ^2 \sigma_{jj}^{-1}}$	-
Invariância à escala	Sim	Sim	Não	Sim	Sim	Não	Sim	Sim	Sim
Normalização	-	-	Alvo	Alvo	Alvo	Fonte	Fonte	Fonte	Alvo
Soma 1 ^c	Não	Não	sum_j	sum_j	Não	sum_i	sum_i	Não	sum_j^d

^aRetira influência comum passada, mas insere influência comum futura.

^bRelações diretas, mas adiciona influência direta comum futura.

^cSomatório em um índice resulta em unidade, p.ex. $\sum_j DTF_{ij} = 1$.

^dSoma 1 quando adicionado termo relativo à inovação instantânea (IPT): $\sum_j PDT_{ij} + IPT_i = 1$.

Capítulo 4

Estimação e estatística assintótica

Para a aplicação das medidas de dependência propostas, enunciaremos fórmulas para sua estimação. É fundamental também um estudo para o comportamento desses estimadores. Com a estatística assintótica para os estimadores, desenvolvida neste capítulo, podemos testar hipóteses e ter intervalos de confiança com dados finitos suficientemente grandes.

No próximo capítulo ilustramos com exemplos os cálculos desenvolvidos e a comparação com estatísticas amostrais.

Espera-se que com a disponibilidade da estatística assintótica, aumente a utilização e a confiabilidade de resultados experimentais baseados nestas medidas de dependência.

4.1 Estimação

Por motivos de estimação, desejamos um número finito de parâmetros no modelo. Utilizaremos assim modelos de ordem finita a partir deste momento, mais especificamente modelos $AR(p)$.

Consideramos dados com N pontos de uma realização, a partir da qual desejamos inferir as medidas de dependência propostas. Esta seção utiliza resultados diversos de [Lutkepohl \(2005\)](#).

4.1.1 Estimação de AR(p)

Reescrevemos os parâmetros em uma forma vetorial:

$$A = [A_1 \dots A_p], \quad (4.1)$$

$$\alpha = \text{vec}(A), \quad (4.2)$$

$$\varepsilon = \text{vec}(\Sigma). \quad (4.3)$$

O estimador de mínimos quadrados (LSE), conhecido como estimador de Yule-Walker, é dado por

$$\hat{A} = [\hat{\Gamma}(1) \dots \hat{\Gamma}(p)] \hat{\Gamma}_X(0)^{-1}, \quad (4.4)$$

$$\hat{\Sigma} = \hat{\Gamma}(0) - [\hat{\Gamma}(1) \dots \hat{\Gamma}(p)] \hat{\Gamma}_X(0)^{-1} [\hat{\Gamma}(1) \dots \hat{\Gamma}(p)]^T, \quad (4.5)$$

onde

$$\hat{\Gamma}_X(0) = \begin{bmatrix} \hat{\Gamma}(0) & \hat{\Gamma}(1) & \dots & \hat{\Gamma}(p-1) \\ \hat{\Gamma}(1) & \hat{\Gamma}(0) & \dots & \hat{\Gamma}(p-2) \\ \vdots & \vdots & \ddots & \vdots \\ \hat{\Gamma}(p-1) & \hat{\Gamma}(p-2) & \dots & \hat{\Gamma}(0) \end{bmatrix}, \quad (4.6)$$

e cada $\hat{\Gamma}(\tau)$ é a função de covariância amostral, estimada por

$$\hat{\Gamma}(\tau) = \frac{1}{N} \sum_{t=0}^{N-\tau} X(t)X(t-\tau)^T. \quad (4.7)$$

Ao inferirmos os parâmetros do modelo AR(p), teremos uma distribuição assintótica para o estimador, que será a base para o cálculo da distribuição assintótica de cada medida de dependência.

Ambos estimadores de α e ε são não viesados e assintoticamente normais. O estimador LSE de α satisfaz a seguinte convergência em distribuição,

$$\sqrt{N}(\hat{\alpha} - \alpha) \xrightarrow{d} N(0, \Omega_\alpha), \quad (4.8)$$

sendo $\Omega_\alpha = \Gamma_x(0)^{-1} \otimes \Sigma$, em que \otimes representa o produto de Kronecker.

Se Σ é desconhecido, teremos que estimá-lo também. Tem-se a variância assintótica deste estimador dada por:

$$\sqrt{N}(\hat{\varepsilon} - \varepsilon) \xrightarrow{d} N(0, \Omega_\varepsilon) \quad (4.9)$$

$$\Omega_\varepsilon = 2D_n D_n^+ (\Sigma \otimes \Sigma) D_n^{+T} D_n^T \quad (4.10)$$

sendo D_n é a matriz de duplicação tal que $D_n \text{vech}(A) = \text{vec}(A)$, e D^+ é a pseudo-inversa generalizada de Moore-Penrose de D .

Os estimadores $\hat{\alpha}$ e $\hat{\varepsilon}$ são assintoticamente independentes.

Na implementação computacional dos resultados, um algoritmo de estimação alternativo foi utilizado, chamado de Nuttall-Strand ou Burg multivariado. Em [Schlögl \(2006\)](#) é feito um estudo comparativo entre os estimadores e este apresenta os melhores resultados. É importante notar que o estimador de Nuttall-Strand tem mesma convergência assintótica do estimador de mínimos quadrados ([Hainz \(1994\)](#)).

4.1.2 Estimação da ordem p

Supondo que o modelo é AR(p) finito, mas de ordem p desconhecida, devemos estimar p . Para isso definiremos um dos muitos possíveis estimadores, o critério de informação de Akaike (AIC), baseado em teoria da informação.

O AIC estima a ordem pesando entre a variância e o viés do estimador. Quanto maior p , menor a variância Σ , mas maior a variância na estimação. O AIC tenta evitar sobreajuste, penalizando o número de parâmetros enquanto minimiza Σ .

O valor estimado de p será o valor que minimize a função

$$AIC(p) = N \log(\det(\Sigma)) + 2n^2 p, \forall p \in \mathbb{Z}^+. \quad (4.11)$$

4.2 Estatística assintótica

Com base nos estimadores definidos em (4.1), temos os parâmetros do modelo e, inserindo estes parâmetros nas definições das medidas, teremos as medidas estimadas.

Um resultado inédito deste trabalho é a derivação teórica, e respectiva implementação computacional, das distribuições assintóticas dos estimadores da PDC (PDC original, gPDC, iPDC), PDT (apenas sob hipótese nula), DTF, coerência parcial, coerência espectral.

Estatísticas assintóticas para estas medidas restaram ausentes por anos e valores subjetivos para limiares de significância eram adotados. Entre as estatísticas já publicadas, [Schelter *et al.* \(2006\)](#) mostrou a estatística para a PDC original sob a hipótese nula, e [Takahashi *et al.* \(2007\)](#) apresenta uma derivação alternativa, onde introduziu a notação e método utilizado neste trabalho, incluindo o caso da hipótese alternativa, calculando intervalos de confiança.

[Hrafnkelsson e Newton \(2000\)](#) calcularam a estatística para a coerência espectral a partir da representação auto-regressiva, mas com o descuido de não considerar o caso particular com coerência espectral nula. Em [Eichler \(2006\)](#), a estatística apenas para o caso nulo da DTF é mostrada.

Uma pequena parte dos resultados a seguir, incluindo a derivação da estatística para a gPDC, foi publicada em forma de artigo em [Brito *et al.* \(2010\)](#).

Para facilitar a compreensão do método utilizado, demonstraremos primeiramente as estatísticas para a PDC em maiores detalhes. Em seguida, demonstraremos as demais estatísticas mais sucintamente.

O método adotado reescreve cada medida em forma matricial, facilitando a derivação de cada medida por um mesmo método. Adicionalmente, simplifica a implementação em linguagens de programação com sintaxe vetorial.

Os cálculos propostos não são computacionalmente ótimos se implementados diretamente. Por exemplo, certas matrizes nos cálculos são esparsas, e sua manipulação termo a termo é recomendável quando for necessária eficiência.

Nos referiremos à hipótese nula por $H_0 : M_{ij}(\lambda) = 0$, sendo M a medida em estudo. A hipótese alternativa será denotada por $H_1 : M_{ij}(\lambda) \neq 0$. A diferenciação nestes dois casos

será necessária, pois sob H_0 teremos uma distribuição assintótica diferente do que sob H_1 .

4.2.1 Método delta

O procedimento utilizado faz uso extensivo do método delta para estatísticas assintóticas, descritos em [van der Vaart \(2000\)](#).

Dados um estimador \hat{y} para y , com variância assintótica Ω/N , e uma função $g = g(\hat{y})$. Se $G = \left. \frac{\partial g(z)}{\partial z} \right|_{z=y}$ existe e é contínua, temos

$$\sqrt{N}(g(\hat{y}) - g(y)) \xrightarrow{d} N(0, G\Omega G^T). \quad (4.12)$$

O método delta está relacionado à aproximação de primeira ordem da série de Taylor de $g(y)$.

Em nosso contexto, a variável y representa nossos parâmetros do AR(p) (\bar{A} e Σ), e a função $g(y)$, uma medida de dependência.

Como já possuímos a variância assintótica dos estimadores dos parâmetros do AR(p), resta-nos calcular $G = \frac{\partial g(y)}{\partial y}$, a derivada da medida por estes parâmetros. A partir da variância assintótica podemos calcular intervalos de confiança e p-valores.

É importante notar que, se a primeira derivada é nula ($\frac{\partial g(y)}{\partial y} \neq 0$), precisamos utilizar derivadas de ordens maiores. Este caso ocorre sob H_0 para todas medidas. Enunciamos o método delta de segunda ordem.

Dados um estimador \hat{y} para y , com variância assintótica Ω , e uma função $g = g(\hat{y})$ que satisfaz $\left. \frac{\partial g(z)}{\partial z} \right|_{z=y} = 0$. Se $G_2 = \left. \frac{\partial^2 g(z)}{\partial z \partial z^T} \right|_{z=y}$ existe e é contínua, temos

$$N(g(\hat{y}) - g(y)) \xrightarrow{d} \frac{1}{2} x^T G_2 x, \quad (4.13)$$

em que $x \xrightarrow{d} N(0, \Omega)$.

Podemos fazer uma transformação de variáveis que transforma este limite, que é uma forma quadrática de variáveis gaussianas, em uma soma de chi-quadrados. Seja $\Omega = LL^T$ a decomposição de Choleski de Ω , com L matriz triangular inferior. Defina $x = Lz$, logo $z = (L^T L)^{-1} L^T x$. Temos que $E[zz^T] = I$.

Se $D = L^T G_2 L$ e $D = U \Lambda U^T$ sua diagonalização, temos

$$x^T G_2 x = z^T L^T G_2 L z = z^T D z = z^T U \Lambda U^T z = \sum_k l_k z^T u_k u_k^T z = \sum_k l_k \nu_k^2, \quad (4.14)$$

pois $u_k u_k^T = I$, sendo l_k os autovalores de D e ν_k^2 chi-quadrados com 1 grau de liberdade, independentes entre si.

$$\text{Logo } N(g(\hat{y}) - g(y)) \xrightarrow{d} \frac{1}{2} \sum_k l_k \nu_k^2.$$

4.2.2 Estatística para PDC

Iniciaremos com a obtenção do teste contra a hipótese nula e dos intervalos de confiança para a PDC nas suas três variantes: iPDC, gPDC e PDC original.

A PDC é definida por

$$PDC_{ij}(\lambda) = \frac{S_{n_{ii}}^{-1} a_{ij} a_{ij}^*}{(a_j^H S_d^{-1} a_j)}, \quad (4.15)$$

em que S_n e S_d são as devidas normalizações para cada variante:

- iPDC: $S_n = \Sigma_d$, $S_d = \Sigma$;
- gPDC: $S_n = \Sigma_d$, $S_d = \Sigma_d$;
- PDC: $S_n = I_n$, $S_d = I_n$;

sendo que $\Sigma_d = \Sigma \odot I_n$ considera apenas a diagonal de Σ , anulando os outros termos. \odot denota o produto de Hadamard e I_n a matriz identidade de ordem n .

Inicialmente reescreveremos o AR(p) e a PDC em uma forma matricial que facilite as derivações seguintes. Em seguida derivaremos a estatística assintótica.

Formulação matricial

Na fórmula do gPDC dependência sobre $a(\lambda)$, no domínio da frequência. Elas foram definidas (na forma matricial) por

$$a^c(\lambda) = \text{vec}(I_n - \sum_{k=1}^p A_k \cdot e^{-j2\pi\lambda k}). \quad (4.16)$$

Como $a^c(\lambda)$ é um vetor complexo, iremos decompô-lo em sua parte real e imaginária, mantendo todas derivações restantes no domínio dos reais,

$$a(\lambda) = \begin{bmatrix} \text{Re}(a^c(\lambda)) \\ \text{Im}(a^c(\lambda)) \end{bmatrix}. \quad (4.17)$$

Assim, reescreveremos $a(\lambda)$ como uma derivação matricial de $\alpha = [A_0 \dots A_p]$,

$$a(\lambda) = \text{vec}([I_n O_n]) - \mathcal{C} \cdot \alpha, \quad (4.18)$$

em que \mathcal{C} é a matriz relativa à transformada de Fourier, separando a parte real e imaginária (\cos e \sin) e pode ser construída por

$$C = [\cos(-2\pi\lambda \cdot 1) \cos(-2\pi\lambda \cdot 2) \dots \cos(-2\pi\lambda \cdot p)] \quad (4.19)$$

$$S = [\sin(-2\pi\lambda \cdot 1) \sin(-2\pi\lambda \cdot 2) \dots \sin(-2\pi\lambda \cdot p)] \quad (4.20)$$

$$\mathcal{C}^* = \begin{bmatrix} C \\ S \end{bmatrix} \quad (4.21)$$

$$\mathcal{C} = \mathcal{C}^* \otimes I_{n^2}. \quad (4.22)$$

Como $\frac{\partial a}{\partial \alpha} = -\mathcal{C}$, utilizando o método delta (4.12), temos a seguinte distribuição assintótica para a :

$$\sqrt{N}(\hat{a} - a) \xrightarrow{d} N(0, \mathcal{C}\Omega_a\mathcal{C}^T) \quad (4.23)$$

de forma que definimos $\Omega_a = \mathcal{C}\Omega_\alpha\mathcal{C}^T$.

Reescrevemos a PDC em sua forma matricial,

$$PDC_{ij} = \frac{\text{num}_{PDC}}{\text{den}_{PDC}} = \frac{a^T I_{ij} \bar{S}_d^{-1} I_{ij} a}{a^T I_j^c \bar{S}_n^{-1} I_j^c a}. \quad (4.24)$$

Nesta segunda formulação, I_{ij} serve para selecionar os termos $a(\lambda)_{ij}$ (tanto a parte real quanto imaginária), pois $a_{ij}a_{ij}^* = \Re a_{ij}^2 + \Im a_{ij}^2$, de forma que $I_{ij}a$ mantém apenas estes elementos e zera os outros. I_j^c serve de forma semelhante para selecionar toda a coluna j de

a , zerando os outros termos. Eles podem ser definidos por

$$I_{ij} = \text{diag}([0 \dots 010 \dots 0]), \quad (4.25)$$

com apenas termo $i + n \cdot j$ não nulo;

$$I_{ij} = I_2 \otimes I_{ij} \quad (4.26)$$

$$I_j^* = \text{diag}([0 \dots 010 \dots 0]) \quad (4.27)$$

com apenas termo j não nulo;

$$I_j = I_j^* \otimes I_n \quad (4.28)$$

$$I_j^c = I_2 \otimes I_j. \quad (4.29)$$

As matrizes $\bar{S}_{(\cdot)}$ são formuladas convenientemente para manter a normalização desejada,

$$\bar{S}_{(\cdot)} = I_{2n} \otimes S_{(\cdot)}. \quad (4.30)$$

O interessante é podermos representar a PDC por uma razão de formas quadráticas, o que facilitará as derivações seguintes.

Derivada de PDC pelos parâmetros

Calculando a derivada do gPDC por a , obtemos

$$G_a = \frac{\partial PDC}{\partial a} = 2PDC \left(\frac{a^T I_{ij} \bar{S}_n^{-1} I_{ij}}{\text{num}_{PDC}} - \frac{a^T I_j^c \bar{S}_d^{-1} I_j^c}{\text{den}_{PDC}} \right). \quad (4.31)$$

Assim, se Σ for conhecido, a variância assintótica de $P\hat{D}C$ é dada por

$$\sqrt{N}(P\hat{D}C - PDC) \xrightarrow{d} N(0, G_a \Omega_a G_a^T). \quad (4.32)$$

Se Σ não for conhecido, teremos que estimá-lo. Como ele aparece na formulação da PDC, sua estimação influirá na variância assintótica do estimador. Assim desejamos encontrar

$$G_e = \frac{\partial PDC}{\partial \varepsilon}.$$

Novamente, após uma sequência de regras da cadeia, temos

$$G_e = PDC \left(\frac{(\mathbf{I}_{ij}\mathbf{a})^T \otimes (\mathbf{a}^T \mathbf{I}_{ij})}{num_{PDC}} \boldsymbol{\xi}_n - \frac{(\mathbf{I}_j^c \mathbf{a})^T \otimes (\mathbf{a}^T \mathbf{I}_j^c)}{den_{PDC}} \boldsymbol{\xi}_d \right) \boldsymbol{\Theta}_n, \quad (4.33)$$

em que $\boldsymbol{\xi}_{(\cdot)} = \frac{\partial S_{(\cdot)}^{-1}}{\partial \Sigma}$, sendo

$$iPDC : \boldsymbol{\xi}_n = -diag(vec(\bar{\Sigma}_d^{-2})), \boldsymbol{\xi}_d = -\bar{\Sigma}^{-1T} \otimes \bar{\Sigma}^{-1}; \quad (4.34)$$

$$gPDC : \boldsymbol{\xi}_n = -diag(vec(\bar{\Sigma}_d^{-2})), \boldsymbol{\xi}_d = -diag(vec(\bar{\Sigma}_d^{-2}));$$

$$PDC : \boldsymbol{\xi}_n = 0_n, \boldsymbol{\xi}_d = 0_n.$$

e

$$\boldsymbol{\Theta}_n = (\mathbf{T}_{2n,n} \otimes \mathbf{I}_{2n^2})(\mathbf{I}_n \otimes vec(\mathbf{I}_{2n} \otimes \mathbf{I}_n)), \quad (4.35)$$

sendo $\mathbf{T}_{L,M}$ a matriz de comutação (Lutkepohl (2005)).

A variância assintótica da PDC é a soma da variância em relação a estimação do α e do ε ,

$$\sqrt{N}(P\hat{D}C - PDC) \xrightarrow{d} N(0, G_a \Omega_a G_a^T + G_e \Omega_\varepsilon G_e^T). \quad (4.36)$$

Para os intervalos de confiança basta pegar os respectivos quantis da distribuição normal assintótica.

Observe que para o PDC original $G_e = 0$.

Teste contra a hipótese nula

Para a distribuição sob $H_0 : PDC_{ij}(\lambda) = 0$, temos que a primeira derivada será nula. Logo utilizaremos o método delta de segunda ordem. O mesmo será feito para todas medidas subsequentes.

Para ter a distribuição assintótica de segunda ordem necessitamos a segunda derivada G_{a2} , dada por

$$G_{a2} = \frac{\partial PDC}{\partial a \partial a^T} = \frac{2I_{ij} \bar{S}_n^{-1} I_{ij}}{num_{PDC}}. \quad (4.37)$$

As segundas derivadas da PDC em relação Σ são nulas, pois os termos I_{ija} não desapa-

recem.

Com G_{a2} , basta decompor Ω_a , calcular D e achar seus autovalores. Para o teste de hipótese, calculamos o quantil superior da soma de chi-quadrados $\sum_k l_k \nu_k^2$,

$$N(P\hat{D}C - PDC) \xrightarrow{d} \frac{1}{2} x^T G_{a2} x = \sum_k l_k \nu_k^2, \quad (4.38)$$

em que $x \xrightarrow{d} N(0, \Omega_a)$, l_k são autovalores de $L^T G_{a2} L$, L fator de Choleski de Ω_a e $\nu_k^2 \sim \chi_1^2$.

Temos $\text{rank}(D) \leq \text{rank}(G_2) \leq \text{rank}(I_{ij}) = 2$, logo $\text{rank}(D) = 2$, exceto quando $\lambda \in \{-0.5, 0, 0.5\}$ ou $p = 1$, onde temos $\text{rank}(D) = 1$.

Dessa forma, convergirá para uma soma ponderada de uma ou duas variáveis chi-quadrados. Para calcular computacionalmente os quantis de uma soma ponderada chi-quadrados, foi utilizada a aproximação de Patnaik, que aproxima a soma por uma chi-quadrado (Takahashi (2008)),

$$\sum_k l_k \nu_k^2 \sim c \chi_d^2 \quad (4.39)$$

$$c = \frac{\sum_k l_k^2}{\sum_k l_k}, d = \frac{(\sum_k l_k)^2}{\sum_k l_k^2}. \quad (4.40)$$

4.2.3 Estatística para DC, SDT e DTF

Similarmente, escreveremos as três medidas alternativas de forma unificada, a qual denotaremos por DC,

$$DC_{ij}(\lambda) = \frac{S_{n_{jj}} h_{ij} h_{ij}^*}{(h_i S_d h_i^H)}, \quad (4.41)$$

em que S_n e S_d são as devidas normalizações para cada variante,

- DC: $S_n = ((\Sigma^{-1})_d)^{-1}$, $S_d = \Sigma$;
- SDT: $S_n = \Sigma_d$, $S_d = \Sigma_d$;
- DTF: $S_n = I_n$, $S_d = I_n$;

em que $S_n = ((\Sigma^{-1})_d)^{-1}$ calcula a matriz diagonal com η 's. Definimos então a forma matricial,

$$DC_{ij} = \frac{num_{DC}}{den_{DC}} = \frac{h^T I_{ij} \bar{S}_d^{-1} I_{ij} h}{h^T I_i^l \bar{S}_n^{-1} I_i^l h}, \quad (4.42)$$

em que $h = vec\left[\begin{array}{cc} Re(H(\lambda)) & Im(H(\lambda)) \end{array} \right]$. A matriz I_i^l seleciona a linha i (diferentemente de I_j^c que seleciona a coluna j)

$$I_i^* = diag([0 \dots 0 1 0 \dots 0]), \quad (4.43)$$

com apenas termo i não nulo;

$$I_i = I_n \otimes I_i^* \quad (4.44)$$

$$I_i^l = I_2 \otimes I_i. \quad (4.45)$$

As matrizes $\bar{S}_{(\cdot)}$ são formuladas convenientemente para manter a normalização desejada (observe que há diferença em relação à normalização da PDC),

$$\bar{S}_{(\cdot)} = S_{(\cdot)} \otimes I_{2n}. \quad (4.46)$$

Para o cálculo de $h_a = \frac{\partial h}{\partial a}$ temos a relação $H = A^{-1}$, assim a derivada complexa será

$$B = \frac{\partial vec(H)}{\partial vec(A)} = -H^T \otimes H. \quad (4.47)$$

De forma que a derivada decompondo as partes reais e imaginárias é dada por

$$h_a = \begin{bmatrix} Re(B) & -Im(B) \\ Im(B) & Re(B) \end{bmatrix}, \quad (4.48)$$

de onde chegamos à sua convergência assintótica,

$$\sqrt{N}(\hat{h} - h) \xrightarrow{d} N(0, \Omega_h), \quad (4.49)$$

em que $\Omega_h = h_a \mathcal{C} \Omega_\alpha \mathcal{C}^T h_a^T$.

Vemos que a formulação da DC é análoga à PDC, substituindo h ao a e I_i^l no lugar de I_j^c . Desta forma recalcularemos as derivadas de acordo.

Fazendo a derivada do DC por h , obtemos

$$dcG_h = \frac{\partial DC}{\partial h} = 2DC \left(\frac{h^T I_{ij} \bar{S}_n I_{ij}}{num_{DC}} - \frac{h^T I_i^l \bar{S}_d I_i^l}{den_{DC}} \right) \quad (4.50)$$

e adicionalmente calculamos $dcG_\varepsilon = \frac{\partial DC}{\partial \varepsilon}$,

$$dcG_\varepsilon = DC \left(\frac{(\mathbf{I}_{ij} \mathbf{h})^T \otimes (\mathbf{h}^T \mathbf{I}_{ij})}{num_{DC}} \boldsymbol{\xi}_n - \frac{(\mathbf{I}_i^l \mathbf{h})^T \otimes (\mathbf{h}^T \mathbf{I}_i^l)}{den_{DC}} \boldsymbol{\xi}_d \right) \boldsymbol{\Theta}_n, \quad (4.51)$$

em que $\boldsymbol{\xi}_{(\cdot)} = \frac{\partial S_{(\cdot)}}{\partial \Sigma}$, respeitando cada variante,

$$DC : \boldsymbol{\xi}_n = \text{diag}(\text{vec}(\bar{\Sigma}_d^{-2})) (\bar{\Sigma}^{-1T} \otimes \bar{\Sigma}^{-1}), \boldsymbol{\xi}_d = I_{2n^2}; \quad (4.52)$$

$$SDT : \boldsymbol{\xi}_n = \text{diag}(\text{vec}(I_{2n^2})), \boldsymbol{\xi}_d = \text{diag}(\text{vec}(I_{2n^2}));$$

$$DTF : \boldsymbol{\xi}_n = 0_n, \boldsymbol{\xi}_d = 0_n.$$

Assim a variância assintótica da PDC é a soma da variância em relação a estimação do α e do ε ,

$$\sqrt{N}(\hat{DC} - DC) \xrightarrow{d} N(0, dcG_h \Omega_h dcG_h^T + dcG_\varepsilon \Omega_\varepsilon dcG_\varepsilon^T). \quad (4.53)$$

Observe que para a DTF original $dcG_\varepsilon = 0$.

Assim para ter a distribuição assintótica de segunda ordem necessitamos a segunda derivada G_{a2} , dada por

$$dcG_{h2} = \frac{\partial DC}{\partial h \partial h^T} = \frac{2I_{ij} \bar{S}_n I_{ij}}{num_{DC}}. \quad (4.54)$$

As segundas derivadas da DC em relação Σ são nulas, pois os termos $I_{ij} a$ não desaparecem, logo temos sob H_0

$$N(\hat{DC} - DC) \xrightarrow{d} \frac{1}{2} x^T dcG_{h2} x = \sum_k l_k \nu_k^2, \quad (4.55)$$

em que $x \xrightarrow{d} N(0, \Omega_h)$, l_k são autovalores de $L^T dcG_{h2} L$, L fator de Choleski de Ω_h e $\nu_k^2 \sim \chi_1^2$.

Temos $\text{rank}(D) = 2$, exceto quando $\lambda \in \{-0.5, 0, 0.5\}$ ou $p = 1$, onde temos $\text{rank}(D) = 1$.

4.2.4 Estatística para Coerência Parcial

A coerência parcial pode ser reescrita matricialmente como

$$PC_{ij} = \frac{\text{Re}(g_{ij})^2 + \text{Im}(g_{ij})^2}{g_{ii}g_{jj}}, \quad (4.56)$$

em que g_{ij} é a densidade espectral cruzada parcial, e

$$\text{Re}(g_{ij}) = (a^T I_i^c \bar{\Sigma}^{-1} I_j^c a) \quad (4.57)$$

$$\text{Im}(g_{ij}) = (a^T I_i^c \bar{\Sigma}^{-1} I_j^{c*} a) \quad (4.58)$$

em que $\bar{\Sigma} = I_{2n} \otimes \Sigma$ e

$$I_j^{c*} = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix} \otimes I_j. \quad (4.59)$$

Calculamos as derivadas em relação a a e e ,

$$RG_a = \frac{\partial \text{Re}(g_{ij})}{\partial a} = a^T (I_i^c \bar{\Sigma}^{-1} I_j^c + I_j^c \bar{\Sigma}^{-1} I_i^c) \quad (4.60)$$

$$RG_e = \frac{\partial \text{Re}(g_{ij})}{\partial e} = ((I_j^c a)^T \otimes a^T I_i^c) \boldsymbol{\xi} \boldsymbol{\Theta}_K \quad (4.61)$$

$$IG_a = \frac{\partial \text{Im}(g_{ij})}{\partial a} = a^T (I_i^c \bar{\Sigma}^{-1} I_j^{c*} + I_j^{c*} \bar{\Sigma}^{-1} I_i^c) \quad (4.62)$$

$$IG_e = \frac{\partial \text{Im}(g_{ij})}{\partial e} = ((I_j^{c*} a)^T \otimes a^T I_i^c) \boldsymbol{\xi} \boldsymbol{\Theta}_K, \quad (4.63)$$

em que $\boldsymbol{\xi} = -\bar{\Sigma}^{-1T} \otimes \bar{\Sigma}^{-1}$ e $\boldsymbol{\Theta}_n$ como definido em (4.35), e os limites assintóticos

$$\begin{aligned} \sqrt{N}(\text{Re}(\hat{g}_{ij}) - \text{Re}(g_{ij})) &\xrightarrow{d} N(0, RG_a \Omega_a RG_a^T + RG_e \Omega_e RG_e^T) \\ \sqrt{N}(\text{Im}(\hat{g}_{ij}) - \text{Im}(g_{ij})) &\xrightarrow{d} N(0, IG_a \Omega_a IG_a^T + IG_e \Omega_e IG_e^T). \end{aligned} \quad (4.64)$$

Assim temos as derivadas da PC em relação a $\text{Re}(g_{ij})$, $\text{Im}(g_{ij})$, $g_{ii} = \text{Re}(g_{ii})$ e $g_{jj} =$

$Re(g_{jj})$,

$$pcG = \begin{bmatrix} \frac{2Re(g_{ij})}{g_{ii}g_{jj}} & \frac{2Im(g_{ij})}{g_{ii}g_{jj}} & \frac{-PC}{g_{ii}} & \frac{-PC}{g_{jj}} \end{bmatrix} \quad (4.65)$$

$$\sqrt{N}(\hat{PC} - PC) \xrightarrow{d} N(0, pcG\Omega pcG^T), \quad (4.66)$$

em que Ω é a matriz de covariâncias assintóticas de $Re(g_{ij})$, $Im(g_{ij})$, $Re(g_{ii})$ e $Re(g_{jj})$.

Observando que sob H_0 temos $Re(g_{ij}) = Im(g_{ij}) = 0$, as únicas derivadas de segunda ordem não nulas são em relação a $Re(g_{ij})$ e $Im(g_{ij})$,

$$pcG_2 = \begin{bmatrix} \frac{2}{g_{ii}g_{jj}} & 0 \\ 0 & \frac{2}{g_{ii}g_{jj}} \end{bmatrix} \quad (4.67)$$

$$N(\hat{PC} - PC) \xrightarrow{d} \frac{1}{2}x^T pcG_2 x = \sum_k l_k \nu_k^2, \quad (4.68)$$

em que $E[xx^T] = \Omega$ é a matriz de covariâncias assintóticas de $Re(g_{ij})$ e $Im(g_{ij})$, l_k são autovalores de $L^T pcG_2 L$, L fator de Choleski de Ω_a e $\nu_k^2 \sim \chi_1^2$.

Temos $rank(D) = rank(pcG_2) = 2$, exceto quando $\lambda \in \{-0.5, 0, 0.5\}$ ou $p = 1$, onde temos $rank(D) = 1$.

4.2.5 Estatística para Coerência Espectral

A derivação para a coerência espectral é análoga à PC, com as devidas adaptações. A coerência espectral pode ser definida por

$$Coh_{ij} = \frac{Re(f_{ij})^2 + Im(f_{ij})^2}{f_{ii}f_{jj}}, \quad (4.69)$$

em que f_{ij} é a densidade espectral cruzada,

$$Re(f_{ij}) = (h^T I_i^l \bar{\Sigma} I_j^l h) \quad (4.70)$$

$$Im(f_{ij}) = (h^T I_i^l \bar{\Sigma} I_j^{*l} h), \quad (4.71)$$

em que $\bar{\Sigma} = \Sigma \otimes I_{2n}$ e

$$I_j^{l*} = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix} \otimes I_i. \quad (4.72)$$

Temos as derivadas em relação a h e e ,

$$RG_h = \frac{\partial Re(f_{ij})}{\partial h} = h^T (I_i^l \bar{\Sigma} I_j^l + I_j^l \bar{\Sigma} I_i^l) \quad (4.73)$$

$$RG_e = \frac{\partial Re(f_{ij})}{\partial e} = ((I_j^l h)^T \otimes h^T I_i^l) \Theta_n \quad (4.74)$$

$$IG_h = \frac{\partial Im(f_{ij})}{\partial h} = h^T (I_i^l \bar{\Sigma} I_j^{l*} + I_j^{l*T} \bar{\Sigma} I_i^l) \quad (4.75)$$

$$IG_e = \frac{\partial Im(f_{ij})}{\partial e} = ((I_j^{l*} h)^T \otimes h^T I_i^l) \Theta_n, \quad (4.76)$$

de onde calculamos os limites assintóticos,

$$\sqrt{N}(Re(\hat{f}_{ij}) - Re(f_{ij})) \xrightarrow{d} N(0, RG_h \Omega_h RG_h^T + RG_e \Omega_e RG_e^T) \quad (4.77)$$

$$\sqrt{N}(Im(\hat{f}_{ij}) - Im(f_{ij})) \xrightarrow{d} N(0, IG_h \Omega_h IG_h^T + IG_e \Omega_e IG_e^T). \quad (4.78)$$

Finalmente, temos as derivadas da coerência espectral em relação a $Re(f_{ij})$, $Im(f_{ij})$, $f_{ii} = Re(f_{ii})$ e $f_{jj} = Re(f_{jj})$ dadas por

$$cohG = \begin{bmatrix} \frac{2Re(f_{ij})}{f_{ii}f_{jj}} & \frac{2Im(f_{ij})}{f_{ii}f_{jj}} & \frac{-Coh}{f_{ii}} & \frac{-Coh}{f_{jj}} \end{bmatrix} \quad (4.79)$$

$$\sqrt{N}(\hat{Coh} - Coh) \xrightarrow{d} N(0, cohG \Omega cohG^T), \quad (4.80)$$

em que Ω é a matriz de covariâncias assintóticas de $Re(f_{ij})$, $Im(f_{ij})$, $Re(f_{ii})$ e $Re(f_{jj})$.

Observando que sob H_0 temos $Re(f_{ij}) = Im(f_{ij}) = 0$, as únicas derivadas de segunda ordem não nulas são em relação a $Re(f_{ij})$ e $Im(f_{ij})$,

$$cohG_2 = \begin{bmatrix} \frac{2}{f_{ii}f_{jj}} & 0 \\ 0 & \frac{2}{f_{ii}f_{jj}} \end{bmatrix} \quad (4.81)$$

$$N(\hat{Coh} - Coh) \xrightarrow{d} \frac{1}{2}x^T cohG_2x = \sum_k l_k \nu_k^2, \quad (4.82)$$

em que $E[xx^T] = \Omega$ é a matriz de covariâncias assintóticas de $Re(f_{ij})$ e $Im(f_{ij})$, l_k são autovalores de $L^T cohG_2L$, L fator de Choleski de Ω_a e $\nu_k^2 \sim \chi_1^2$.

Temos $rank(D) = rank(cohG_2) = 2$, exceto quando $\lambda \in \{-0.5, 0, 0.5\}$ ou $p = 1$, onde temos $rank(D) = 1$.

4.2.6 Estatística para PDT

Para a PDT derivaremos apenas a estatística sob H_0 , quando $a_{ij} = 0$.

Temos a formulação matricial da PDT,

$$PDT_{ij} = \frac{a^T I_{ij} a f_{jj}}{f_{jj}} \quad (4.83)$$

e a derivada segunda dada por

$$pdtG_{a2} = \frac{\partial PDT}{\partial a \partial a^T} = \frac{2I_{ij} f_{jj}}{f_{jj}}. \quad (4.84)$$

A partir das derivadas obtemos a estatística assintótica sob a hipótese nula,

$$N(P\hat{D}T - PDT) \xrightarrow{d} \frac{1}{2}x^T pdtG_{a2}x = \sum_k l_k \nu_k^2, \quad (4.85)$$

em que $x \xrightarrow{d} N(0, \Omega_a)$, l_k são autovalores de $L^T G_{a2}L$, L fator de Choleski de Ω_a e $\nu_k^2 \sim \chi_1^2$.

Temos $rank(D) = 2$, exceto quando $\lambda \in \{-0.5, 0, 0.5\}$ ou $p = 1$, onde temos $rank(D) = 1$.

Capítulo 5

Análise com exemplos

Para ilustrar as medidas de dependência, e respectivas estatísticas, estas foram implementadas computacionalmente. Neste capítulo apresentamos exemplos que ilustram as estatísticas assintóticas e analisam seu comportamento para amostras finitas, de forma a comparar com as distribuições amostrais por Monte-Carlo. Com isso teremos um estudo da aplicabilidade dessas estatísticas limites sobre dados finitos.

Em seguida, mostramos o resultado de cada medida para um modelo simulado e a dados reais de Potencial de Campo Local (LFP) em ratos.

5.1 Implementação computacional em Python

Um resultado substancial deste trabalho foi a implementação das medidas de dependência apresentadas, incluindo as estatísticas assintóticas desenvolvidas.

As formulações foram implementadas como rotinas na linguagem Python, com auxílio do pacote de distribuição Python(x,y)¹. Python(x,y) inclui um conjunto completo de pacotes científicos, fundamentados na biblioteca vetorial Numpy (Oliphant (2007)). Python cresce nos anos recentes com uma excelente plataforma para computação científica, sendo gratuita, código aberto e com crescente número de pacotes distribuídos pela comunidade científica.

O software desenvolvido inclui diversas rotinas, muitas não discutidas neste texto, entre as quais:

¹<http://www.pythonxy.com/>

- Medidas de dependência
 - Coerência, Coerência Parcial, PDC (PDC, gPDC, iPDC), PDT, DTF
 - Índice de Causalidade de Granger, Causalidade de Granger Instantânea
- Estatística assintótica
 - Coerência, Coerência Parcial, PDC (PDC, gPDC, iPDC), DTF
 - Estatística amostral por bootstrap
- Apresentação gráfica de resultados
 - Gráficos de dependência para cada par e frequência (ver Figura 5.1)
 - Gráficos de tempo-frequência (ver Figura 5.7)
- Simulação Monte-Carlo de modelos AR(p)
- Estimação de modelo AR(p)
 - Algoritmos Nuttall-Strand e Yule-Walker
- Dependência para potencial relativo a eventos (ERP)
 - Estimação AR(p) para pequenos trechos repetidos
 - Janelamento de dados por eventos
 - Janelamento de dados e estatística por estados
- Exemplos ilustrativos retirados da literatura

O software foi desenvolvido de forma modular para chamada por comandos simples. Uma série de parâmetros são acessíveis para escolha entre as rotinas e parâmetros (p.ex. resolução na frequência, método estatístico, opções gráficas, ordem do AR(p)). O código seguinte ilustra algumas chamadas para análise de dependência em uma sessão interativa de Python:

```

1 # Importa biblioteca PDC-Python desenvolvida.
2 from pdc import *
3 # Importa dados do usuário.
4 dados = loadtxt('meus_dados.txt')
5 # Rotina completa: estima AR(p), calcula PDC, calcula estatística
   assintótica e apresenta resultados gráficos.
6 pdc_full(dados)
7 # Retorna dados simulados para um modelo AR(p) fornecido, com 10000
   amostras.
8 dados_sim = ar_data(model = 1, nd = 10000)
9 # Rotina completa para a coerência, utilizando estatística por bootstrap
   e ordem AR(p) máxima 3.
10 coh_full(dados_sim, stat = 'boot', maxp = 3)

```

Espera-se que este pacote de ferramentas incentive o estudo e aplicação dos métodos apresentados.

5.2 Análise gráfica para dados simulados

Nas Figuras 5.1 a 5.3 mostramos graficamente o valor de cada medida de dependência para o modelo (3.1), para simulações com 500 pontos e limiar de significância de 1%. Cada subgráfico representa o valor da dependência para um par de séries, para cada frequência (eixo x). Os subgráficos da diagonal apresentam a densidade espectral logarítmica de cada série.

Nota-se como todas conexões previstas foram detectadas, estando em conformidade com os respectivos grafos de dependência.

Fica ilustrada como a coerência e coerência parcial são simétricas e apresentam dependência entre muitos pares, dificultando a interpretação sobre as interações. As medidas baseadas na decomposição destas, DTF, gPDC e PDT, mostram-se mais informativas, revelando as interações direcionadas entre os pares.

A DTF apresenta dependência significativa nos pares $1 \rightarrow 5$ e $2 \rightarrow 5$, não detectadas pela gPDC. Apesar de conter a mesma informação, fica evidente a dificuldade de inferir

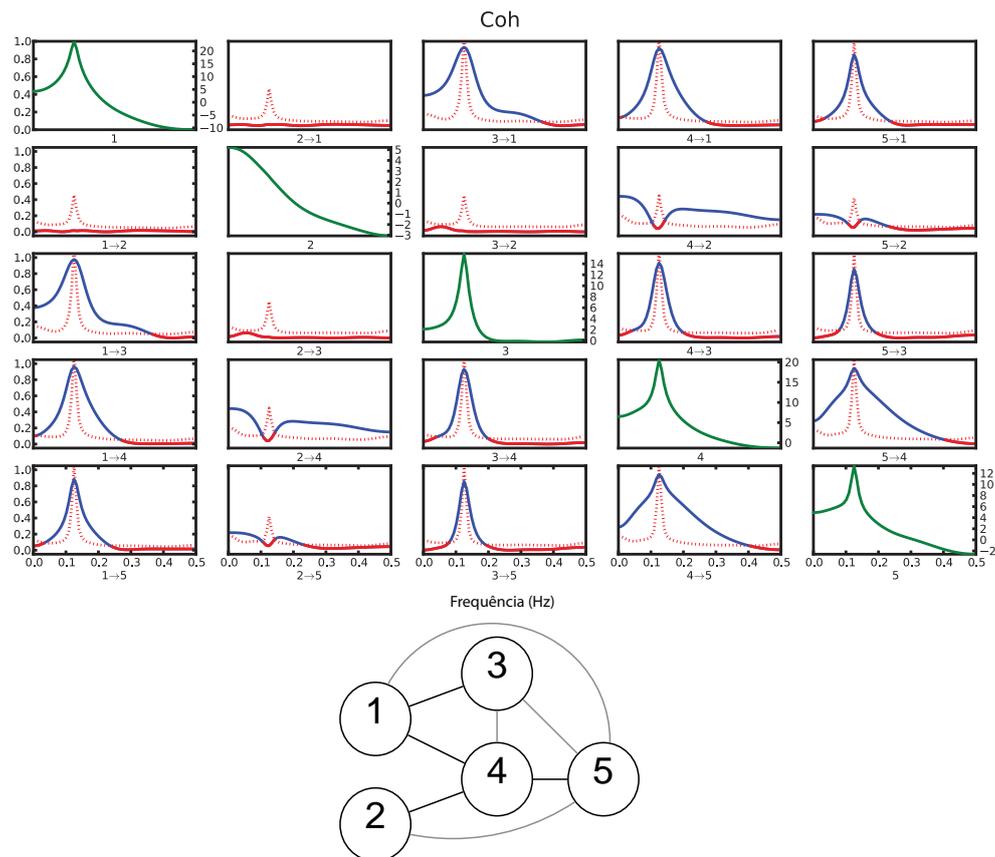


Figura 5.1: Resultado gráfico da coerência espectral para simulação com 500 pontos do modelo (3.1). Cada subgráfico representa a força da dependência para um par, para cada frequência. A diagonal apresenta a potência espectral de cada série na escala logarítmica. O limiar para 1%, calculado por meio da estatística assintótica, é mostrado em vermelho pontilhado. Valores da coerência abaixo do limiar são mostrados em vermelho; valores acima, em azul. O grafo é mostrado para referência das dependências esperadas para esta medida.

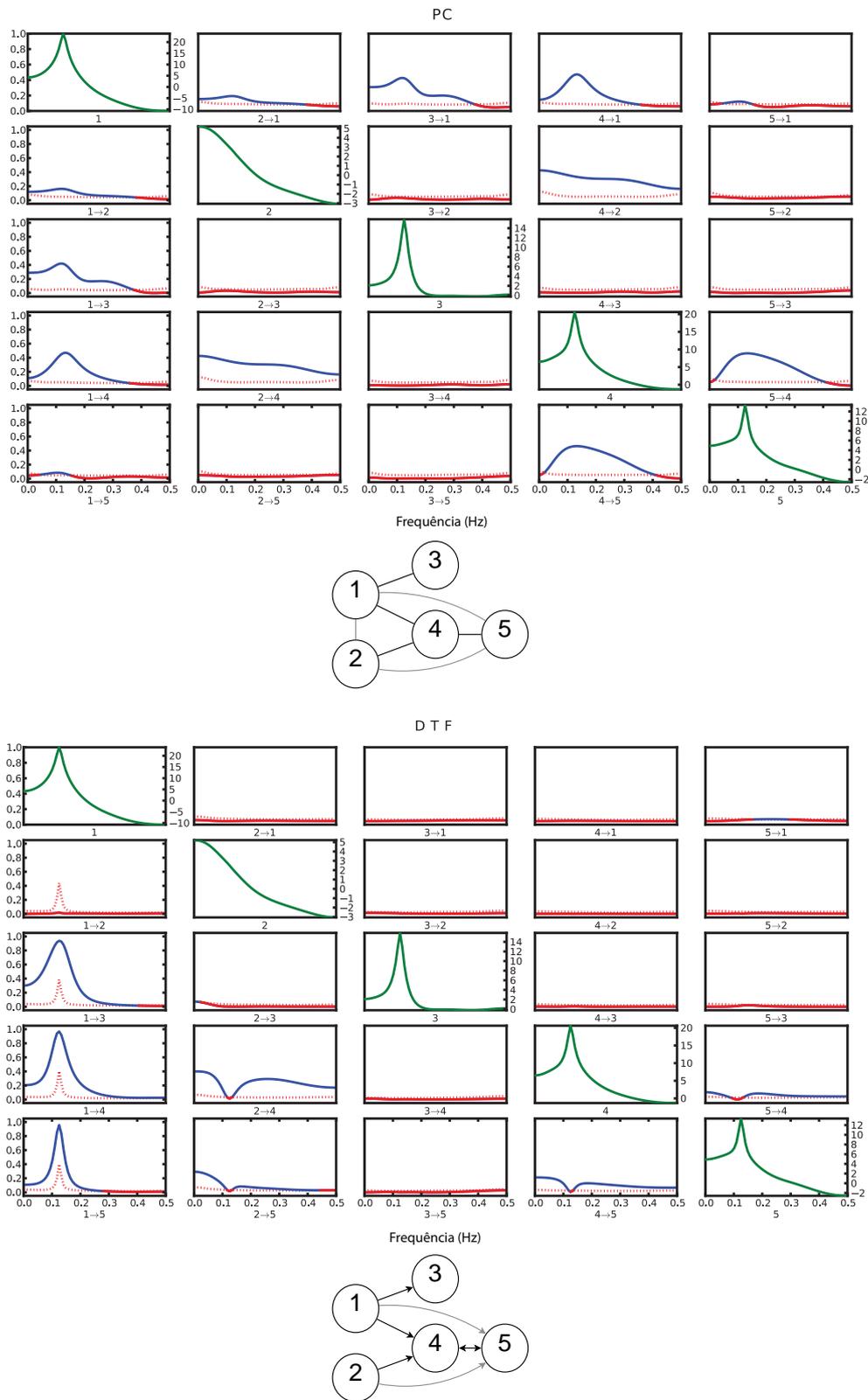


Figura 5.2: Resultado gráfico da coerência parcial e DTF para o modelo (3.1). Ver legenda da Figura 5.1 para maiores explicações.

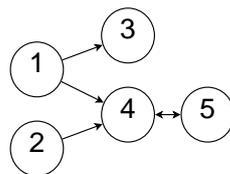
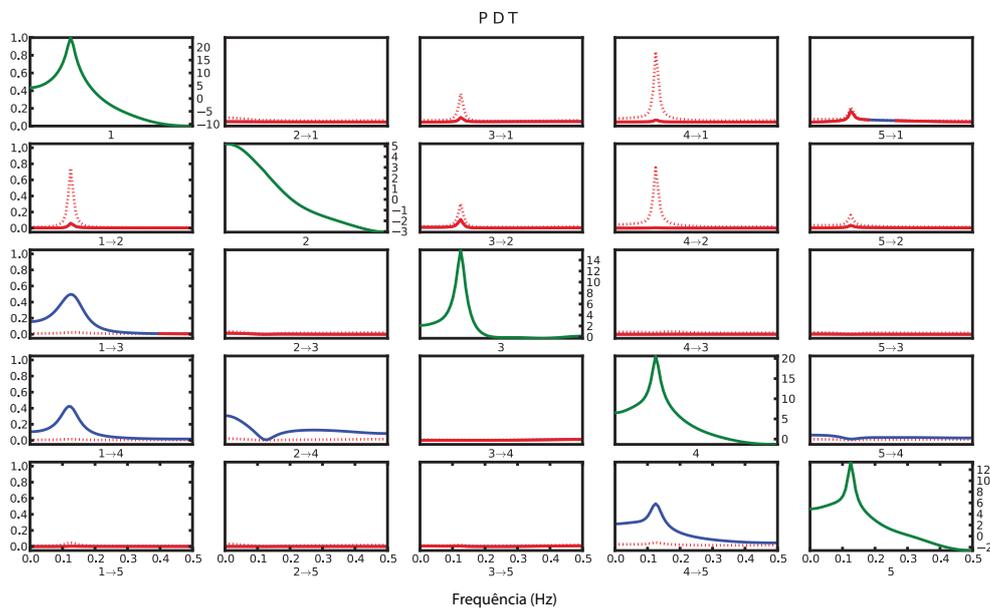
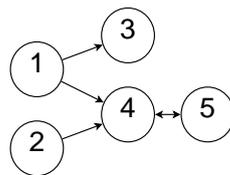
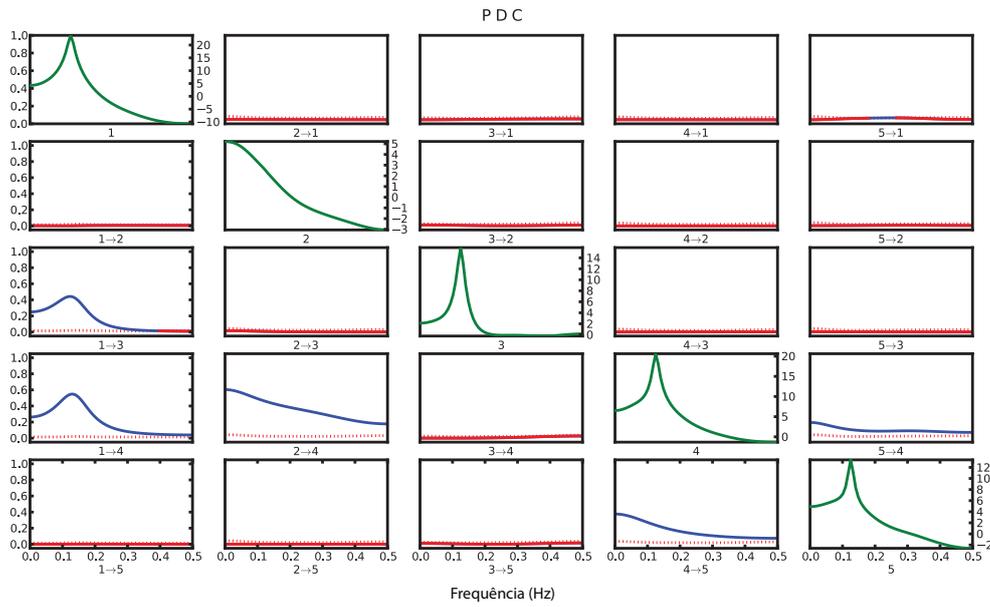


Figura 5.3: Resultado gráfico da PDC e PDT para o modelo (3.1). Ver legenda da Figura 5.1 para maiores explicações.

as interações diretas do processo AR gerador a partir da DTF. Por outro lado, pode-se observar a origem das influências. Vê-se como x_1 não recebe influência de nenhuma série e tem um pico em seu espectro em torno de $0,13Hz$. Essa potência é transferida para as séries influenciadas, x_3 , x_4 e x_5 , direta ou indiretamente. Deste modo a alta potência para estas séries nesta frequência tem origem em x_1 , o que é ilustrado pelo alto DTF para estas dependências.

A gPDC mostra como significativas as dependências que são relações diretas no modelo AR. Pode-se inferir com facilidade o grafo direcionado equivalente ao gerador dos dados, incluindo a realimentação entre x_4 e x_5 , sendo possivelmente mais informativa que as medidas anteriores. A gPDC apresenta os menores limiares entre as medidas (menor viés do estimador), devido à sua formulação direta sobre os parâmetros do AR.

A PDT apresenta as mesmas dependências da gPDC. A diferença fundamental é ser normalizada pelo alvo, sendo formulada sobre o espectro não parcializado. Isso fica claro na dependência $2 \rightarrow 4$, na qual há um vale em $0,13Hz$. Podemos assim interpretar que nesta frequência a potência de x_4 é devida à influência direta de x_1 , sendo a influência de x_2 desprezível. A exposição da origem da potência em cada frequência não é mostrada pela gPDC e suas variantes. A PDT tem estimador com maior viés (maior limiar) que a gPDC, pois necessita a inversão da matriz $A(\lambda)$.

Os pequenos trechos com falsos positivos na DTF, gPDC e PDT de $5 \rightarrow 1$ estão dentro do previsto devido aos múltiplos testes feitos. A correção de Bonferroni, ajustando o limiar de acordo com o número de testes feitos, pode ser utilizada, mas é dificultada pelo fato dos testes não serem independentes.

Os testes mostrados são pontuais, sendo que testes para faixas de frequência, incluindo a faixa com todo o espectro, não têm formulação trivial.

Fica claro como, em faixas do espectro em que uma medida é numericamente maior, há maior poder do teste. Já dependências mais fracas necessitam maior número de amostras para adquirir significância, como se vê na coerência parcial do par $1 \rightarrow 5$, que quase não é detectada.

Em alguns casos nota-se um maior viés e limiar do estimador na faixa do espectro com maior potência, como na DTF no par $1 \rightarrow 2$. Esta relação parece ser recorrente e é um

possível tópico de investigação teórica.

As medidas baseadas no espectro direto, coerência espectral, DTF e PDT, apresentam picos no limiar em algumas dependências por volta de $0,13HZ$, que é a frequência de maior potência da série x_1 . O alto limiar deve-se a variância do estimador, elevada por este basear-se na matriz $H(\lambda)$, que é calculada por uma inversão da matriz $A(\lambda)$. A dependência sobre a matriz $A(\lambda)$ torna-se uma vantagem para a coerência parcial e gPDC, tendo maior robustez no cálculo de seu valor e estatística.

Devemos notar que no cálculo dos limiares, utilizamos valor estimado para os parâmetros do modelo AR. Uma possível contestação é de que, por exemplo, no caso do gPDC, o teste é equivalente ao teste de nulidade de a_{ij} , sendo que o valor de a_{ij} utilizado no teste não é nulo. Isto pode gerar distorções no limiar estimado.

Uma alternativa seria estimar o modelo AR restito, com $a_{ij} = 0$ fixo. O problema é que a restrição pode ocasionar modelos estimados muito diferentes do original. Adicionalmente, outras medidas como a coerência espectral e DTF não são testes diretos de $a_{ij} = 0$, não havendo uma forma de restringir o modelo AR trivialmente.

5.2.1 Exemplo comparativo entre PDC e PDT

Para ilustrar a motivação da nova medida PDT, analisaremos uma alteração do modelo (3.1), no qual alteramos a conexão $1 \rightarrow 3$:

$$\left\{ \begin{array}{l} x_1(t) = 0,95\sqrt{2}x_1(t-1) - 0,9025x_1(t-2) + e_1(t) \\ x_2(t) = 0,5x_2(t-1) + e_2(t) \\ x_3(t) = \alpha x_1(t-3) + e_3(t) \\ x_4(t) = -0,5x_1(t-2) + 0,8x_2(t-1) + 0,25\sqrt{2}x_4(t-1) + 0,25\sqrt{2}x_5(t-1) + e_4(t) \\ x_5(t) = -0,25\sqrt{2}x_4(t-1) + 0,25\sqrt{2}x_5(t-1) + e_5(t) \end{array} \right. \quad (5.1)$$

sendo $\Sigma = I_5$.

Na Figura 5.4 vemos o resultado da gPDC e PDT para $\alpha = 10$, para uma simulação de 10.000 pontos e limiar de 1%.

Como a gPDC é normalizada pela fonte, ao aumentar a influência de $1 \rightarrow 3$, irá ocorrer

uma diminuição na gPDC $1 \rightarrow 4$. De fato, vemos como a gPDC $1 \rightarrow 4$ resulta perto de zero, sendo apenas significativa devido à estatística calculada, sendo que dificilmente seria interpretada como uma influência relevante. Para o modelo original tínhamos $PDC_{41}(0, 13) = 0,60$, sendo que com $\alpha = 10$ temos $PDC_{41}(0, 13) = 0,002$. Isso pode ser indesejável, pois a única alteração no modelo foi na relação entre $1 \rightarrow 3$, que não deveria influenciar na medida da relação $1 \rightarrow 4$.

Por outro lado, a PDT é normalizada pela fonte, e como o espectro de x_4 continua sendo explicado em grande parte por x_1 , não há nenhuma alteração na PDT de $1 \rightarrow 4$, sendo $PDT_{41}(0, 13) = 0,41$ para qualquer α . Nota-se apenas o aumento na influência $1 \rightarrow 3$, como esperado. Um contraponto é o maior viés no estimador da PDT, ilustrado pelos picos em $0,13Hz$ em dependências não existentes. Para dados com menos amostras o viés será maior e esta limitação pode ser um problema para a aplicação do método.

Um efeito similar ao ilustrado neste exemplo seria encontrado ao reduzir a variância de x_3 , aumentando assim a influência de x_1 sobre este.

5.3 Estatística assintótica para dados finitos

Por meio de simulações de Monte-Carlo, analisamos o comportamento das estatísticas assintóticas para diferentes tamanhos amostrais. Mostraremos o resultado para as estatísticas da Coerência, Coerência Parcial, DTF e gPDC. As outras variantes resultam em comportamentos similares.

Para comparar com a estatística amostral, simulamos 10.000 realizações do modelo (3.1) para 10.000 pontos, calculamos cada medida e a distribuição amostral do conjunto de realizações.

Na Figura 5.5, mostramos o gráfico de quantis para cada medida em relação à respectiva estatística assintótica. A linha de simetria é mostrada em azul nos gráficos em que a medida exata é não nula (quantis da normal assintótica) e em vermelho onde a medida é nula (quantis das chi-quadrados ponderadas). É mostrada a distribuição para frequência $0,2Hz$ (considerando uma resolução temporal fictícia de 1s) das relações $1 \rightarrow 2$, $3 \rightarrow 4$, $1 \rightarrow 4$ e $1 \rightarrow 5$.

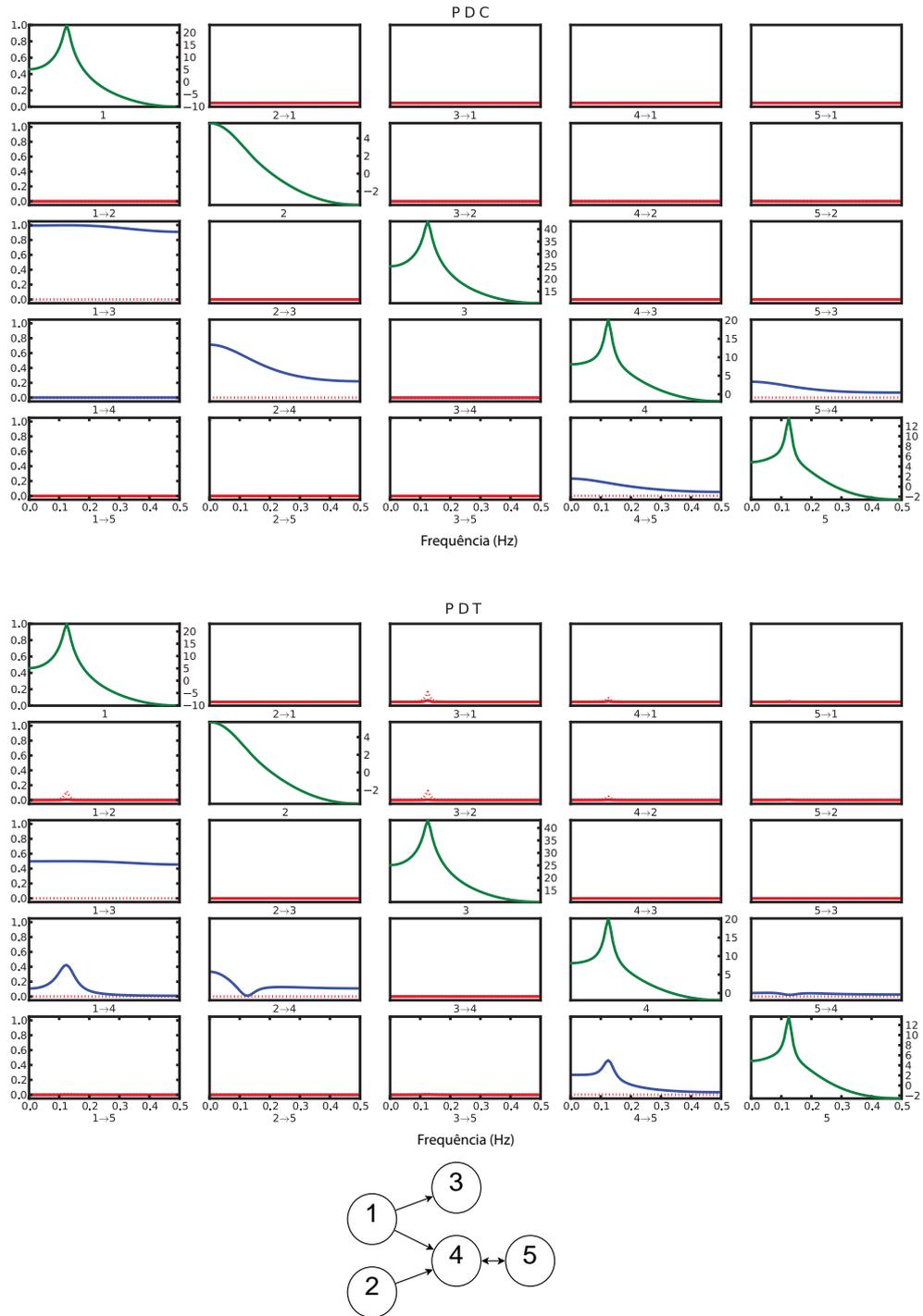


Figura 5.4: Resultado gráfico para a $gPDC$ e PDT para o modelo 5.1, com $\alpha = 10$. Por ser normalizado por quem recebe, na $gPDC$ a forte conexão $1 \rightarrow 3$ reduz muito o valor da influência $1 \rightarrow 4$, mesmo sendo significativa. O mesmo não acontece com a PDT , normalizada pelo alvo.

Há uma boa aproximação para ambos casos, sob H_0 e H_1 , ilustrando como a estatística está de acordo com a distribuição amostral dos respectivos casos, para todas as medidas.

Para a relação $1 \rightarrow 5$ temos medidas não nulas, porém, próximas de zero, para a coerência espectral, DTF e gPDC. Isso justifica a leve distorção nesse caso em relação à reta de simetria, pois a aproximação de primeira ordem tem valor pequeno, tornando significativos os termos de segunda ordem.

A distorção na extremidade das curvas relativas à hipótese nula é relativa aos quantis limítrofes, não interferindo em estatísticas a partir de 1% de significância. A distorção pode ser justificada pela substituição da soma de chi-quadrados pela aproximação de Patnaik.

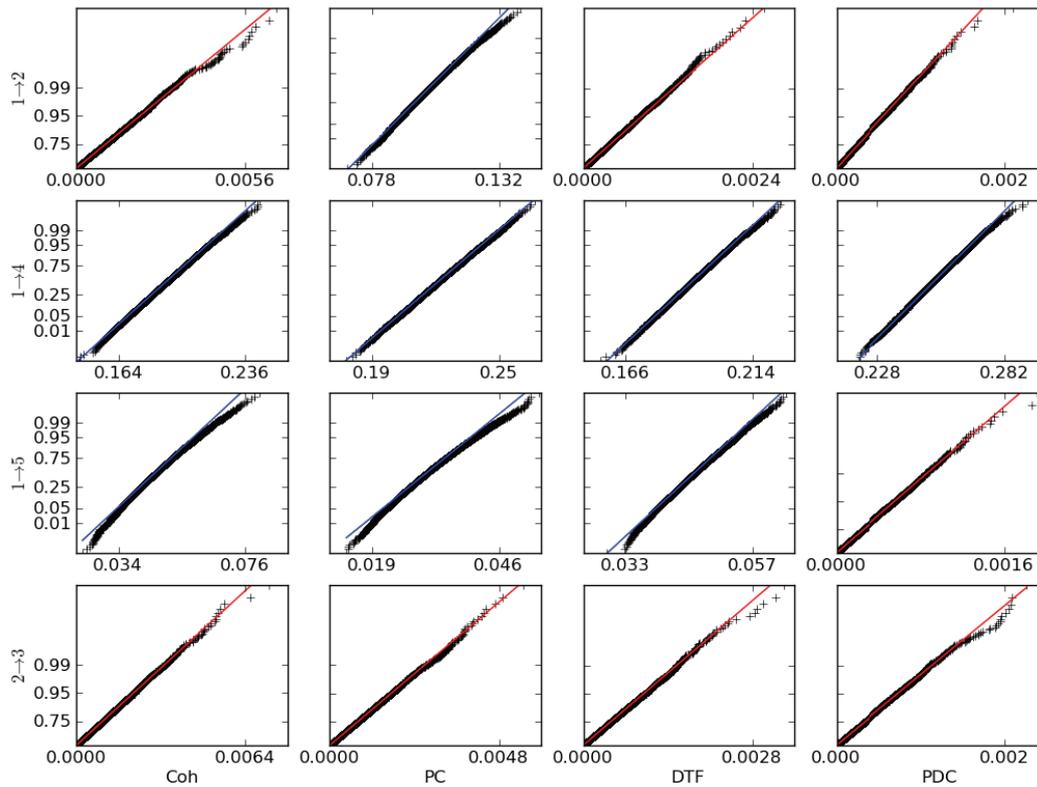


Figura 5.5: Gráfico de quantis para cada medida (por coluna) para as relações $1 \rightarrow 2$, $1 \rightarrow 4$, $1 \rightarrow 5$ e $2 \rightarrow 3$ (por linha), para 10.000 simulações. Os gráficos comparam quantis da distribuição assintótica teórica (eixo y) com a distribuição amostral (eixo x). Os gráficos com linha de simetria em vermelho correspondem à distribuição sob H_0 (soma ponderada de χ^2) e com linhas de simetria em azul correspondem à distribuição sob H_1 (normal).

Para ilustrar a densidade amostral e assintótica, mostramos na Figura 5.6 o histograma da coerência espectral para 0,2 Hz para as relações $1 \rightarrow 2$ e $1 \rightarrow 4$, sobrepostos pela

distribuição teórica assintótica, mostrando a adequação da distribuição normal e da soma de chi-quadrados.

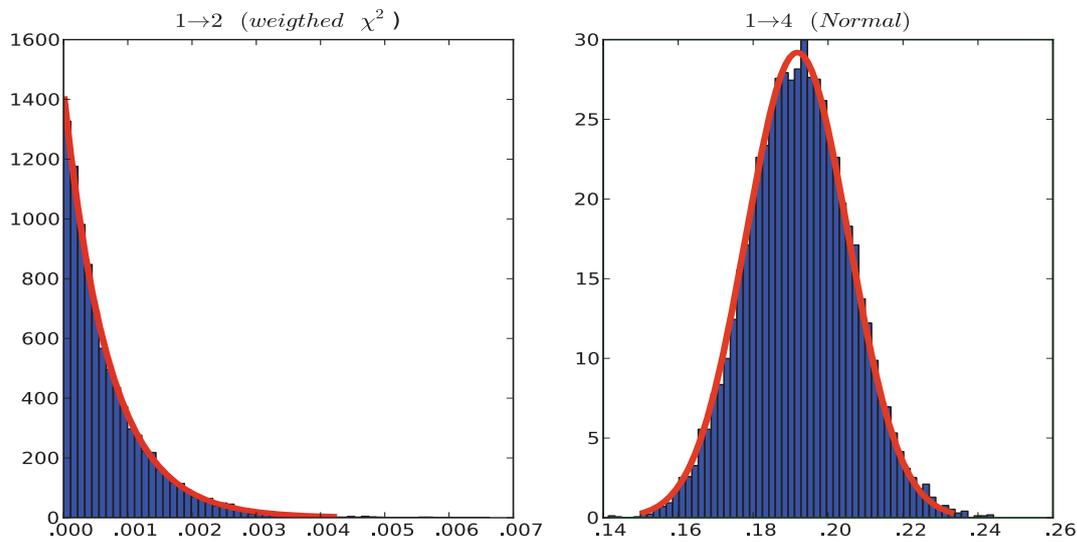


Figura 5.6: Histograma da amostragem para coerência espectral entre $1 \rightarrow 2$ e $1 \rightarrow 4$, sobreposto pela distribuição teórica (em vermelho), para $\lambda = 0.2Hz$.

Para analisar o comportamento das estatísticas para dados finitos, estimamos os tamanhos dos intervalos de confiança e limiares para testes de hipótese em relação, comparando com os valores nominais esperados. Na Tabela 5.1 mostramos, para cada medida, os falsos positivos para a relação $2 \rightarrow 3$ para um erro do tipo I de 1% e 5%, e o tamanho amostral dos intervalos de confiança para a relação $1 \rightarrow 4$, para a frequência $0,2Hz$. O modelo foi simulado 10.000 vezes, para tamanhos amostrais de 100, 500, 1.000 e 10.000 pontos.

Observa-se como para todas as medidas a estatística converge para o valor nominal com o aumento do tamanho da amostra, como esperado. Para amostras curtas, de 100 ou 500 pontos, a estatística tende a ser conservadora, com taxa de erro menor que o valor nominal. O desvio para essas amostras curtas é esperado, pois a distribuição assintótica não é uma boa aproximação para estes casos. Entretanto, não se observa grandes desvios em relação ao valor nominal, de forma que a estatística pode ainda sim ser aplicada, com a devida precaução na escolha dos valores nominais.

Tabela 5.1: Erro tipo I (em %) para cada medida, sob um valor nominal de 1% e 5%, para 0, 2Hz. O modelo (3.1) foi simulado 10.000 vezes, para amostras de 100, 500, 1.000 e 10.000 pontos. A dependência $1 \rightarrow 4$ é não nula para todas medidas, logo estima erro do intervalo de confiança. A dependência $2 \rightarrow 3$ é nula, estimando erro do limiar. Vemos uma boa aproximação da estatística em relação ao valor nominal com 10.000 pontos e tendência a ser conservadora para amostras menores.

Coh		N			
		100	500	1.000	10.000
$1 \rightarrow 4$	1%	0.5	0.72	0.83	1.04
	5%	4.66	4.49	4.62	5.19
$2 \rightarrow 3$	1%	1.44	1.15	1.04	1.14
	5%	5.94	4.91	4.69	5.19

PC					
$1 \rightarrow 4$	1%	0.55	0.74	0.92	1.07
	5%	4.6	4.41	4.85	5.15
$2 \rightarrow 3$	1%	0.58	1.01	1.01	1.07
	5%	3.49	5.05	4.88	5.31

DTF					
$1 \rightarrow 4$	1%	0.8	0.7	0.72	1.03
	5%	5.11	3.72	4.3	5.37
$2 \rightarrow 3$	1%	1.63	0.95	1.16	1.16
	5%	5.08	4.95	5.33	5.13

gPDC					
$1 \rightarrow 4$	1%	0.55	0.57	0.81	0.88
	5%	3.4	3.22	4.42	4.92
$2 \rightarrow 3$	1%	0.6	1.05	0.98	1.21
	5%	3.42	5.09	4.85	5.34

5.4 Aplicação em Potencial de Campo Local em ratos

O registro de atividade neural por implante de microeletrodos tornou-se uma técnica central na área de eletrofisiologia neural (Buzsaki *et al.* (1992)). Eletrodos isolados eletricamente são inseridos no cérebro de forma a ter sua ponta exposta na região de interesse, medindo o potencial elétrico local.

O sinal medido detecta simultaneamente dois padrões elétricos de interesse. A componente de baixa frequência do sinal está relacionada à atividade elétrica média da região. Este sinal, obtido por filtros passa-baixa, é denominado potencial de campo local (LFP), e acredita-se medir a resultante das atividades elétricas devidas a potenciais de ação que chegam à região medida. Desta forma, estaria relacionada ao potencial elétrico sub-limiar dos neurônios de um raio de aproximadamente 3 milímetros do local de medição (Juergens *et al.*

(1999)).

Observa-se também uma atividade de alta frequência, resultante de potenciais de ação em neurônios próximos do eletrodo (até $\sim 300\mu m$), medindo a atividade de saída de neurônios individuais (Gray *et al.* (1995)).

Em colaboração com o trabalho de doutorado de Eduardo Schenberg (Schenberg *et al.* (2010)), analisamos atividades de LFP de ratos Wistar machos com as medidas de dependência estudadas.

Dois matrizes com 32 eletrodos, de tungstênio revestidos com teflon, foram implantadas na região do hipocampo, nos hemisférios direito e esquerdo. A atividade elétrica foi registrada durante ciclos de sono-vigília e atividade LFP foi filtrada por filtro passa-baixa. Por meio de análise simultânea do LFP e suas propriedades espectrais, e vídeo comportamental, diferentes estados de sono e vigília foram determinadas.

O hipocampo é implicado no processo de memorização, tanto em ratos quanto em humanos, sendo intensamente investigado. Adicionalmente, acredita-se que ciclos do sono estejam diretamente relacionados a processos mnemônicos. Ilustramos as medidas em dados respectivos às fases de ondas lentas (SWS) e REM do sono, a fim de diferenciar padrões de atividade no hipocampo em cada estado cognitivo.

Os estágios do sono são caracterizados por significantes diferenças no potencial espectral de seus sinais LFP. No sono REM apresenta-se uma alta atividade em toda faixa de baixa frequência, entre 0 e 20 Hz. Por outro lado, no sono de ondas lentas observa-se uma atividade significativa apenas na faixa entre 6 e 12 Hz, denominado ritmo teta (Gervasoni *et al.* (2004)).

Utilizamos nesta análise quatro eletrodos, das regiões CA1 e CA3 do hipocampo de cada hemisfério, selecionados por qualidade do sinal, exame de histologia e coordenadas estereotáxicas do implante.

O sinal LFP foi coletado a $1KHz$ e decimado para $500Hz$. Por meio do critério de Akaike 4.11 aplicado à análise de alguns trechos, fixamos a ordem do modelo VAR em 25 pontos de atraso. Os dados foram divididos em janelas de 10 segundos para cálculo das medidas.

Na Figura 5.7 observa-se o resultado gráfico das medidas gPDC e coerência espectral ao longo do tempo, para um intervalo de 600 segundos. As cores correspondem ao valor da medida, observando que as escalas da gPDC e coerência são distintas, para facilitar a análise

da gPDC que apresenta em geral valores reduzidos em relação à coerência.

As cores da barra horizontal superior indicam a fase do sono do animal, REM (verde) ou SWS (vermelho). A cor negra representa que estágio do sono não foi apropriadamente caracterizado ou representa transição entre estágios.

Os subgráficos na diagonal apresentam a potência espectral ao longo do tempo (espectrograma). Observamos como as características respectivas a cada estágio do sono (ritmo teta ou baixas frequência) ficam evidentes.

Os resultados da coerência mostram um forte acoplamento entre as estruturas CA1 de cada hemisfério ("CA1d" e "CA1e") e entre as estruturas CA3 de cada hemisfério ("CA3d" e "CA3e"), nas faixa espectrais de maior potência de cada estágio do sono. Com base em conhecimento anatômicos, esperava-se uma dependência maior entre as duas regiões CA3, e entre regiões CA3 e CA1 de um mesmo hemisférios. Os resultados indicam que mesmo sem uma forte conexão anatômica, as regiões CA1 de cada hemisfério estão funcionalmente acopladas.

A gPDC mostra dependência fundamentada em deslocamento temporal. Seu baixo valor em muitos pares de conexões indica que grande parte da coerência é devida à atividade acoplada simultânea. A atividade simultânea pode significar que há uma fonte comum atuando instantaneamente sobre ambos sinais.

No SWS, a gPDC apresenta alta dependência na direção $CA3e \rightarrow CA3d$ entre 0 e 20 Hz, com uma dependência mais fraca no sentido inverso. Há uma dependência em $CA1e \rightarrow CA1d$ entre 0 e 10Hz, com uma dependência no sentido oposto entre 8 e 20 Hz.

Na fase REM, a dependência $CA3e \rightarrow CA3d$ decai drasticamente, havendo maior dependência no sentido oposto $CA3d \rightarrow CA3e$ entre 0 e 20 Hz. Já entre as estruturas Ca1 ocorre o efeito contrário, havendo forte dependência nas duas direções, sendo $CA1e \rightarrow CA1d$ mais evidente entre 0 e 10 Hz, e sentido $CA1d \rightarrow CA1e$ mais evidente entre 8 e 20 Hz.

Estes resultados mostram como a análise espectral releva propriedades bem definidas e diferenciadas dos estágios do sono. O resultado evidencia acoplamentos funcionais que podem ser distintos das conexões anatômicas.

A gPDC retira dependências instantâneas, identificando se há uma relação de atraso temporal no acoplamento. Mostra-se como em alguns casos há dependência em diferentes

faixas do espectro para cada direção da dependência, como em REM entre as regiões CA1.

Estudos adicionais são necessários para analisar a reprodutibilidade destas medidas, possivelmente dependentes do local exato do implante na região.

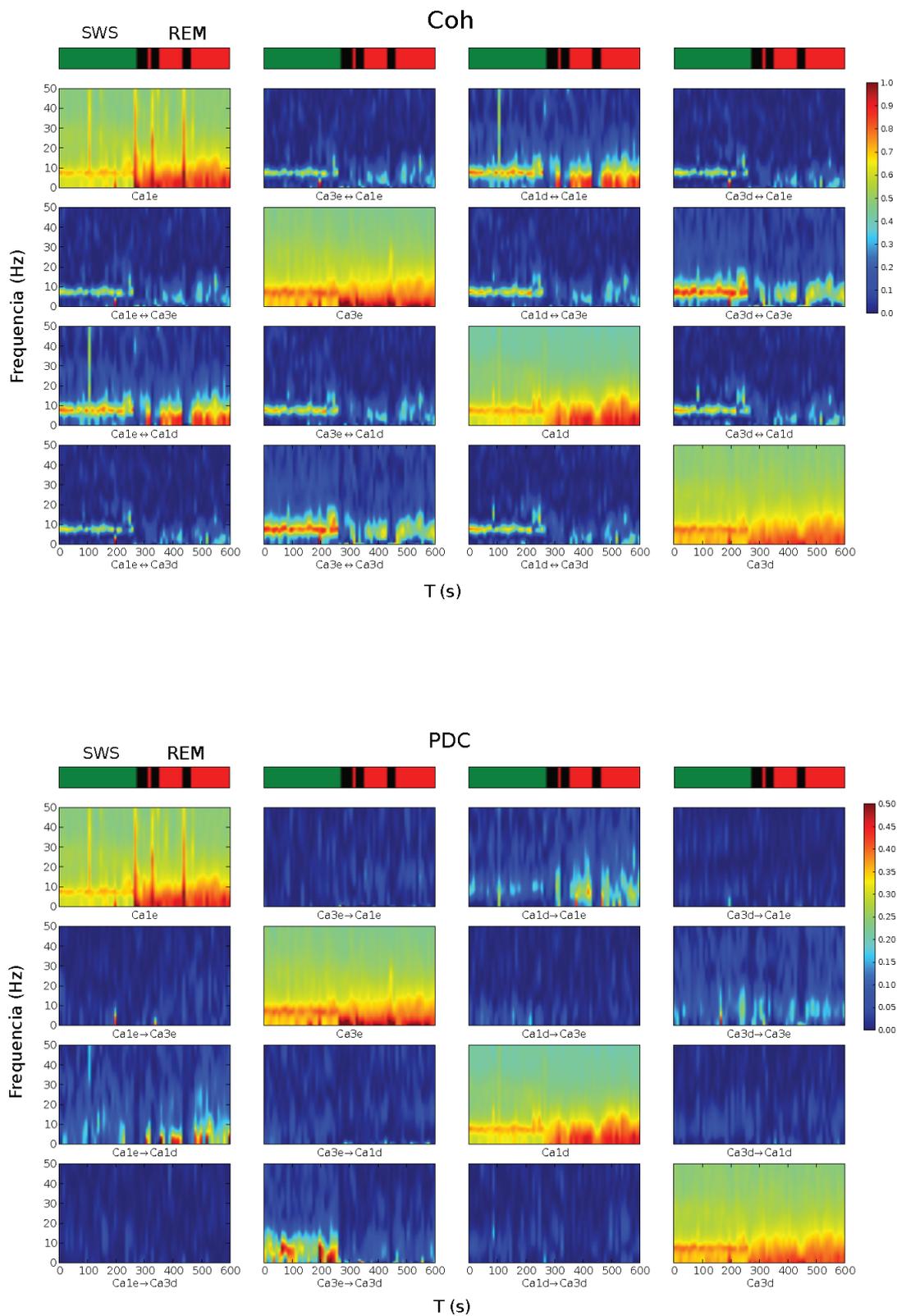


Figura 5.7: Progressão temporal da coerência espectral e *gPDC*. Mostra-se o valor da medida para cada frequência ao longo do tempo. Trecho de 600s foi dividido em janelas de 10s para estimação do modelo AR. A cor representa a intensidade da medida, observando a diferença entre as escalas. Os subgráficos na diagonal mostram o potencial espectral de cada sinal. A barra superior representa o estágio do sono (REM ou SWS) em cada momento do trecho.

Capítulo 6

Conclusão

Medidas de dependência entre séries temporais com o intuito de evidenciarem conexões reais entre os processos que as geraram foram estudadas e comparadas. Com as propriedades das medidas e respectivos resultados, espera-se obter informações sobre se, e como, as interações acontecem.

O estudo de dados multivariados se fundamentou na teoria de séries temporais. Foram apresentadas as suposições necessárias sobre séries, incluindo estacionariedade, estabilidade e invertibilidade. Representações lineares auto-regressiva e médias móveis, e respectivas representações espectrais, foram definidas como base para interpretação de relações lineares e medidas de dependência.

Estudamos as propriedades de algumas medidas lineares no domínio da frequência, que medem o acoplamento entre séries, tendo sua potência decomposta no domínio espectral. Entre as propriedades estudadas se destacam a capacidade de retirar a influência comum de uma terceira série (parcialização), a capacidade de detectar apenas relações diretas, e a fundamentação em conceitos de causalidade, sendo capaz de definir direcionalidade temporal na dependência entre as séries.

O método desenvolvido para análise das diferentes medidas de dependência possibilita uma comparação sistemática entre estas, incluindo propriedades possuídas, formulações alternativas, definições como correlação quadrática e conceitos de dependência que as fundamentam. A Tabela 3.1 sumariza o resultado deste estudo.

A coerência espectral, amplamente utilizada, é a medida menos informativa, não apre-

sentando nenhuma destas propriedades. Ao contrário, as medidas baseadas na causalidade de Granger (PDC e PDT) satisfazem todas estas condições, sendo recomendável como substituta para a Coerência em análises multivariadas.

Foi introduzida uma nova medida, a transferência parcial direcionada, também fundamentada na causalidade de Granger, mas baseada no processo direto (e não inverso ou parcial). A PDT pode ser interpretada como decomposição do espectro direto pelas dependências diretas direcionadas, tendo a importante propriedade de normalização pelo alvo.

Como contribuição para a aplicabilidade destas medidas, a estatística assintótica para estas foi desenvolvida e implementada computacionalmente. O cálculo destas estatísticas seguiu uma metodologia sistemática, possibilitando o desenvolvimento de todas as estatísticas por uma mesma técnica.

Tem-se com isso intervalos de confiança e p-valores, sendo possível validar estatisticamente resultados obtidos com essas medidas. Possíveis trabalhos futuros podem dar continuidade a este estudo estatístico abordando estatísticas de grupos e múltiplas realizações, além de estatísticas para faixas do espectro.

As estatísticas demonstradas foram ilustradas com simulações para grandes amostras, por meio das quais ilustrou-se a convergência esperada. Também mostramos como estas se comportam com amostras menores. As estatísticas mostraram-se robustas para tamanhos amostrais suficientemente limitados, logo podem e devem ser utilizadas para estudo de significância de resultados.

O exemplo de aplicação a dados de potencial de campo local mostrou como as medidas apresentadas podem revelar propriedades não inferidas pela coerência espectral. A PDC e PDT mostram como diferentes regiões interagem direcionalmente em cada frequência, para diferentes estados cognitivos.

Este método pode ser aplicado em dados de quaisquer regiões para diferentes atividades cognitivas, sendo uma ferramenta exploratória importante do mapeamento de interações funcionais no cérebro.

Referências Bibliográficas

- Akaike(1968)** H. Akaike. On the use of a linear model for the identification of feedback systems. *Annals of the Institute of statistical mathematics*, 20(1):425–439. Citado na pág. [29](#)
- Babiloni et al.(2005)** F. Babiloni, F. Cincotti, C. Babiloni, F. Carducci, D. Mattia, L. Astolfi, A. Basilisco, P. M. Rossini, et al. Estimation of the cortical functional connectivity with the multimodal integration of high-resolution eeg and fmri data by directed transfer function. *NeuroImage*, 24(1):118–131. Citado na pág. [1](#)
- Baccala et al.(2007)** L. A Baccala, D. Y. Takahashi, e K. Sameshima. Generalized partial directed coherence. Em *Proceedings of the 2007, 15th International Conference on Digital Signal Processing (DSP2007)*. *IEEE*, volume 1, páginas 162–166. Citado na pág. [34](#)
- Baccalá e Sameshima(2001)** L. A Baccalá e K. Sameshima. Partial directed coherence: a new concept in neural structure determination. *Biological Cybernetics*, 84(6):463–474. Citado na pág. [18](#), [34](#)
- Box e Jenkins(1994)** George Edward Pelham Box e Gwilym M. Jenkins. *Time Series Analysis: Forecasting and Control*. Prentice Hall PTR. ISBN 0130607746. Citado na pág. [4](#)
- Brito et al.(2010)** C. S. N. Brito, Luiz Antonio Baccala, Daniel Yasumasa Takahashi, e Koichi Sameshima. Asymptotic behavior of generalized partial directed coherence. *IEEE EMBS Conference*. Citado na pág. [44](#)
- Brown et al.(2004)** Emery N Brown, Robert E Kass, e Partha P Mitra. Multiple neural spike train data analysis: state-of-the-art and future challenges. *Nat Neurosci*, 7(5):456–461. ISSN 1097-6256. Citado na pág. [1](#)

- Buzsaki et al.(1992)** G. Buzsaki, Z. Horvath, R. Urioste, J. Hetke, e K. Wise. High-frequency network oscillation in the hippocampus. *Science*, 256(5059):1025. Citado na pág. 69
- Cox e Savoy(2003)** D. D Cox e R. L Savoy. Functional magnetic resonance imaging (fMRI). *Neuroimage*, 19(2):261?270. Citado na pág. 1
- Eichler(2006)** Michael Eichler. On the evaluation of information flow in multivariate systems by the directed transfer function. *Biological Cybernetics*, 94(6):469–482. Citado na pág. 44
- Gervasoni et al.(2004)** D. Gervasoni, S. C Lin, S. Ribeiro, E. S Soares, J. Pantoja, e M. A.L Nicolelis. Global forebrain dynamics predict rat behavioral states and their transitions. *Journal of Neuroscience*, 24(49):11137. Citado na pág. 70
- Granger(1969)** C. W. J. Granger. Investigating causal relations by econometric models and cross-spectral methods. *Econometrica*, 37(3). Citado na pág. 4, 25
- Gray et al.(1995)** Charles M. Gray, Pedro E. Maldonado, Mathew Wilson, e Bruce McNaughton. Tetrodes markedly improve the reliability and yield of multiple single-unit isolation from multi-unit recordings in cat striate cortex. *Journal of Neuroscience Methods*, 63(1-2):43–54. ISSN 0165-0270. Citado na pág. 70
- Greicius et al.(2003)** M. D Greicius, B. Krasnow, A. L Reiss, e V. Menon. Functional connectivity in the resting brain: a network analysis of the default mode hypothesis. *Proceedings of the National Academy of Sciences of the United States of America*, 100(1):253. Citado na pág. 1
- Hainz(1994)** G. Hainz. The asymptotic properties of burg estimators. *Beitrage zur Statistik*, no, 18. Citado na pág. 43
- Horwitz(2003)** B. Horwitz. The elusive concept of brain connectivity. *Neuroimage*, 19(2):466?470. Citado na pág. 2
- Hosoya(2001)** Y. Hosoya. Elimination of third-series effect and defining partial measures of causality. *Journal of time series analysis*, 22(5):537–554. Citado na pág. 29

- Hrafnkelsson e Newton(2000)** B. Hrafnkelsson e H. J. Newton. Asymptotic simultaneous confidence bands for vector autoregressive spectra. *Biometrika*, 87(1). Citado na pág. 44
- Juergens et al.(1999)** E. Juergens, A. Guettler, e R. Eckhorn. Visual stimulation elicits locked and induced gamma oscillations in monkey intracortical-and EEG-potentials, but not in human EEG. *Experimental Brain Research*, 129(2):247–259. Citado na pág. 69
- Kaminski e Blinowska(1991)** M. Kaminski e K. Blinowska. A new method of the description of the information flow in the brain structures. *Biological Cybernetics*, 65(3):203–210. Citado na pág. 32
- Lutkepohl(2005)** H. Lutkepohl. *New introduction to multiple time series analysis*. Springer. Citado na pág. 41, 49
- Mann e Lees(1996)** M. E Mann e J. M Lees. Robust estimation of background noise and signal detection in climatic time series. *Climatic Change*, 33(3):409–445. Citado na pág. 4
- Oliphant(2007)** T. E Oliphant. Python for scientific computing. *Computing in Science & Engineering*, páginas 10–20. Citado na pág. 57
- Pearl(2000)** J. Pearl. *Causality: models, reasoning, and inference*. Cambridge Univ Pr. Citado na pág. 24
- Priestley(1994)** M. B. Priestley. *Spectral analysis and time series*. Academic Press Limited, London. Citado na pág. 10, 11, 12
- Schelter et al.(2006)** B. Schelter, M. Winterhalder, M. Eichler, M. Peifer, B. Hellwig, B. Guschlbauer, C. H Lücking, R. Dahlhaus, e J. Timmer. Testing for directed influences among neural signals using partial directed coherence. *Journal of neuroscience methods*, 152(1-2):210–219. Citado na pág. 44
- Schenberg et al.(2010)** E. E. Schenberg, R. P. Proença, E. Morya, M. A. L. Nicolelis, e K. Sameshima. Electrophysiological dynamics of hippocampal subregions during the sleep-wake cycle. *FENS Forum Abstracts*, 5(130.72). Citado na pág. 70

- Schlögl(2006)** Alois Schlögl. A comparison of multivariate autoregressive estimators. *Signal Processing*, 86(9):2426–2429. Citado na pág. 43
- Takahashi(2008)** Daniel Yasumasa Takahashi. Medidas de fluxo de informacao com aplicacao em neurociencia. *Tese de Doutorado - Bioinformática - USP*. Citado na pág. 29, 30, 31, 33, 50
- Takahashi et al.(2010)** Daniel Yasumasa Takahashi, Luiz Antonio Baccalá, e Koichi Sameshima. Frequency domain connectivity: an information theoretic perspective. *IEEE EMBS Conference*. Citado na pág. 33
- Takahashi et al.(2007)** D.Y. Takahashi, L.A. Baccala, e K. Sameshima. Connectivity inference between neural structures via partial directed coherence. *Journal of Applied Statistics*, 34(10):1259–1273. Citado na pág. 44
- van der Vaart(2000)** A. W. van der Vaart. *Asymptotic statistics*. Cambridge University Press. Citado na pág. 45