

Laboratório de Modelagem, Análise e Controle de Sistemas Não-Lineares

Departamento de Engenharia Eletrônica

Universidade Federal de Minas Gerais

Av. Antônio Carlos 6627, 31270-901 Belo Horizonte, MG Brasil

Fone: +55 3499-4866 - Fax: +55 3499-4850

emmendes@cpdeee.ufmg.br



Métodos de Dados Sub-rogados Aplicados a Séries Temporais

Cássio Pascoal Costa

Dissertação submetida à banca examinadora designada pelo Colegiado do Programa de Pós-Graduação em Engenharia Elétrica da Universidade Federal de Minas Gerais, como parte dos requisitos necessários à obtenção do grau de Mestre em Engenharia Elétrica.

Orientador: Prof. Dr. Eduardo Mazoni Andrade Marçal Mendes

Belo Horizonte, março de 2008

Dedicatória

*A Deus, fonte e provedor do meu conhecimento
Aos meus pais, pelo amor
Aos meus irmãos, pela fraterna amizade
A minha amada, por compartilhar esse meu sonho
Aos meus companheiros de república, por me suportarem
por estes 2 anos.*

Agradecimentos

Agradeço a Deus!

Agradeço a você que está lendo essa dissertação!

Agradeço aos colegas do CPDEE e CPH.

Agradeço a CAPES por financiar esse trabalho.

Agradeço ao meu orientador: Eduardo Mendes

Agradeço àqueles que estudaram comigo: Davidson, Erlon e Leonardo Prudêncio.

Agradeço àqueles com quem convivi: Hellom, Ana Paula, Samuel e Jr.

Agradeço àqueles que me orientaram espiritualmente: Meus pais, irmãos e amigos

Agradeço àqueles que esqueci, e a eles peço o perdão.

Agradeço àquela que acreditou: minha mãe.

Agradeço ao meu pai: pelos apoio incondicional e zelo.

Agradeço a Kênia meu amor.

Agradeço por agradecer...

Sumário

Resumo	xi
Abstract	xiii
Lista de Figuras	xviii
Lista de Tabelas	xx
Lista de Símbolos	xxii
Lista de Abreviações	xxiii
1 Introdução	1
1.1 Objetivos	4
1.2 Motivação	5
1.3 Problemas	5
1.4 Relevância	7
1.5 Organização do Texto	7
2 Testes de Hipóteses Estatísticas	9
2.1 Introdução	9
2.2 Hipótese Estatística	10

2.3	Estatística de Teste	19
2.3.1	Métodos Paramétricos	20
2.3.2	Métodos Não-Paramétricos	24
2.4	Quantidade de séries de dados sub-rogados a serem geradas	28
3	Métodos de Dados Sub-rogados	29
3.1	Introdução	29
3.2	Embaralhamento Temporal (Shuffle): IID, GAUSS e BOOT - Algoritmo 0	31
3.3	Small-Shuffle - Algoritmo SS	32
3.4	FT - Algoritmo 1	34
3.5	AAFT - Algoritmo 2	36
3.6	WFT - Algoritmo 1 e AAWFT - Algoritmo 2	38
3.7	IAAFT - Algoritmo 2	40
3.8	CAAFT - Algoritmo 2	42
3.9	STAP - Algoritmo 2	44
3.10	Embaralhamento de ciclos - algoritmo CS	48
3.11	Sub-rogado pseudo-periódico - algoritmo PPS	50
3.12	Comentários Finais	53
4	Análise dos Métodos de Dados Sub-rogados: Proposição de um Pro- cedimento Geral	55
4.1	Introdução	55
4.2	Série simulada	57
4.2.1	Estatística de teste paramétrica	58
4.2.2	Estatística de teste não-paramétrica	62
4.3	Séries temporais financeiras	67
4.4	Processo linear gaussiano - AR, MA ou ARMA	71

4.5	Processo linear não-gaussiano - AR, MA ou ARMA	88
4.6	Transformação não-linear aplicada a processos lineares realimen- tados com inovações gaussianas e não-gaussianas	105
4.7	Mapas	112
4.8	Série de Mackey-Glass com atraso	120
4.9	Ressonância magnética nuclear - NMR	125
4.10	Séries temporais caóticas	128
4.11	Um estudo de caso	135
4.12	Estudo de caso “ <i>blind</i> ”	144
4.13	Comentários Finais	149
5	Conclusões	151
	Bibliografia	161
A	Determinismo vs. Estocasticidade	163
A.1	Introdução	163
A.2	Sistemas Dinâmicos	164
A.3	Processos Estocásticos	168
A.3.1	Conceitos Básicos	168
A.3.2	Definição de Processo Estocástico	177
B	Tabelas	179

Resumo

Nos últimos anos o método de dados sub-rogados tem sido amplamente aplicado por vários autores na análise de determinismo em séries temporais, principalmente aquelas geradas por não-linearidades (Theiler et al., 1992). Porém, as técnicas disponíveis na literatura nem sempre são adequadas quando se deseja determinar e analisar se a dinâmica que deu origem a uma série temporal é, por exemplo, determinística ou não, logo há de se lançar mão de outros métodos. Todavia, pouca atenção tem sido dada ao desenvolvimento de ferramentas específicas para lidar com esses tipos de questionamento.

Este trabalho propõe a investigação do comportamento dinâmico das séries temporais, contaminadas com ruídos gaussianos ou não-gaussianos, usando o método de teste de hipótese baseado em dados sub-rogados. As primeiras tentativas para encontrar tais respostas inerentes ao mecanismo gerador das séries temporais são identificadas por meio de testes estatísticos também chamados de discriminantes estatísticos que podem ser paramétricos e não-paramétricos.

Neste trabalho, procurou-se avaliar as técnicas disponíveis na literatura para a geração das séries temporais sub-rogadas. Da análise dos resultados, por meio de séries temporais simuladas com diversos tipos de comportamentos e séries temporais experimentais, são verificados quais algoritmos para dados sub-rogados (algoritmo 0, algoritmo 1, algoritmo 2, algoritmo SS, algoritmo CS e algoritmo PPS) são mais adequados para o teste de determinismo. Baseado na detecção de comportamento determinístico ou não das séries temporais sob análise foram obtidos resultados que implicam que nem sempre tais técnicas para geração dos dados sub-rogadas são adequadas e que existem situações particulares que conduzem a resultados errôneos para certas estruturas (modelos) geradoras de tais séries. Além disso, propõe-se, como resultado final desse trabalho, um procedimento para análise geral utilizando os métodos de dados sub-rogados, consistentes com as respectivas hipóteses nulas, que possibilita determinar aleatoriedade (ruído) ou determinismo (linear, não-linear, etc).

Abstract

In recent years the method of surrogate data has been widely applied to the analysis of determinism in time series, mainly those time series generated by a nonlinear mechanism (Theiler et al., 1992). Since the techniques available in the literature are not always prepared to deal with the problem of determining and analyzing whether underlying dynamics of the generating mechanism of the time series is deterministic or not, the study of new methods is necessary. However, little attention has been given to the development of specific tools to deal with this problem.

This work proposes the investigation of the dynamic behavior from a time series contaminated with Gaussian or Non-Gaussian noises by using the method of hypothesis testing based on the surrogate data methodology. The first attempts to find the generating mechanism of such time series are performed by using statistical tests, also called statistical discriminant analysis, that can be parametric and non-parametric.

In this work, several techniques available in the literature and used for generating surrogate time series are tested. From analyzing the results, using simulated time series with different dynamic behavior and also real time series, it is determined which surrogate data algorithm - algorithm 0, algorithm 1, algorithm 2, algorithm SS, algorithm CS or algorithm PPS - is more appropriate for the test of determinism. Based upon the detection or not of a deterministic behavior in the time series under analysis, the results obtained here reveal that such techniques for generation of the surrogate data are not always well tuned and that there are particular situations that lead to misleading results for certain model structures of the mechanism generator of such time series. As the major result, a procedure for general analysis is given using the surrogate method and appropriate null hypotheses that can determine randomness (pure noise) and determinism (linear, nonlinear and etc.) for each case illustrated in this work.

Lista de Figuras

2.1	Hierarquia das Hipóteses lineares	16
2.2	Região crítica para H_0	21
2.3	Intervalo de confiança para teste de S com $\alpha = 0,05$	22
3.1	Relação da amplitude A de um número aleatório gaussiano	33
3.2	FT aplicado ao sistema de Rössler	35
3.3	Tipos de janelamento de dados	40
4.1	Fluxograma geral para teste de hipótese nula (H_0) utilizando os métodos de dados sub-rogados.	56
4.2	Variáveis aleatórias gaussianas (1001-10000) com $N(0, 1)$	58
4.3	Gráfico da distribuição das estatísticas Q_n do algoritmo IID	60
4.4	Gráfico da distribuição das estatísticas Q_n do algoritmo GAUSS	60
4.5	Gráfico da distribuição das estatísticas Q_n do algoritmo BOOT	61
4.6	Gráfico da distribuição das estatísticas Q_n do algoritmo SS	61
4.7	Gráfico da FAC	62
4.8	Gráfico da FAC	63
4.9	Gráfico quantil-quantil:	63
4.10	Gráfico da FAC para a distribuição uniforme	64
4.11	Gráfico da FAC para a distribuição Beta	65
4.12	Gráfico da FAC para a distribuição Laplace	65
4.13	Gráfico da FAC para a distribuição Cauchy	65
4.14	Fluxograma geral para teste de hipótese nula (NH_0) utilizando os métodos de dados sub-rogados IID, GAUSS, BOOT e SS.	66
4.15	Série temporal financeira	68
4.16	Série financeira e dados sub-rogados da taxa diária de câmbio Real/USD	69
4.17	Gráfico QQ, da FAC e da IMM da taxa diária de câmbio Real/USD	70
4.18	Densidade espectral para o modelo MA	73

4.19	Gráfico quantil-quantil MA	73
4.20	Gráfico da FAC e da IMM para o MA	75
4.21	Séries temporais AR(1)	76
4.22	Gráfico quantil-quantil e densidade espectral do AR(1)	78
4.23	Gráfico da FAC e da IMM para AR(1)	78
4.24	Séries temporais geradas a partir do ARMA(6,1)	80
4.25	Gráfico da FAC e da IMM	81
4.26	Série temporal ARMA(6,1)	83
4.27	Gráfico da FAC e da IMM para a série temporal ARMA(6,1)	84
4.28	Gráfico da FAC e da IMM da série temporal financeira Real/USD)	85
4.29	Gráfico da FAC e da IMM da série temporal financeira Real/USD)	85
4.30	Gráfico da FAC e da IMM da série temporal financeira Real/USD)	86
4.31	Gráfico da FAC e da IMM da série temporal financeira Real/USD)	86
4.32	Gráfico da FAC e da IMM da série temporal financeira Real/USD)	87
4.33	Fluxograma geral para teste de hipótese nula (H_0) utilizando os métodos de dados sub-rogados algoritmo 0 (IID e SS), algoritmo 1 (FT) e algoritmo 2 (AFT, IAAFT-1, IAAFT-2, CAAFT e STAP) aplicados somente a séries temporais com estruturas do tipo AR, MA ou ARMA.	88
4.34	Série temporal MA(1) com inovações uniforme	89
4.35	Gráfico da FAC e da IMM para o MA(1) com inovações uniforme	90
4.36	Série temporal ARMA(1,1) com inovações uniforme	91
4.37	Gráfico da FAC e da IMM para o ARMA(1,1) com inovações uniforme	92
4.38	Série temporal ARMA(7,1) com ruído Beta	94
4.39	Ruído Beta	94
4.40	Gráfico da FAC e da IMM para o ARMA(7,1) com inovações beta	96
4.41	Série temporal ARMA(1,1) com inovações Laplace	97
4.42	Ruído de Laplace	98
4.43	Gráfico da FAC e da IMM para o ARMA(1,1) com inovações de Laplace	99
4.44	Série temporal ARMA(1,1) com inovações de Cauchy	100
4.45	Ruído de Cauchy	101
4.46	Gráfico da FAC e da IMM para o ARMA(1,1) com inovações de Cauchy	102
4.47	Fluxograma geral para teste de hipótese nula (H_0) utilizando os métodos de dados sub-rogados algoritmo 0 (IID, BOOT, GAUSS e SS), algoritmo 1 (FT) e algoritmo 2 (AFT, IAAFT-1, IAAFT-2, CAAFT e STAP) aplicados somente a séries temporais com estruturas do tipo AR, MA ou ARMA com inovações não-gaussianas.	104
4.48	Séries temporais NAWGN, AWNGN e NAWNGN	108

4.49	Gráficos quantil-quantil	108
4.50	Gráficos da FACs para NAWGN, AWNGN e NAWNGN	109
4.51	Gráficos IMM da NAWGN, AWNGN e NAWNGN	110
4.52	Histograma das séries NAWGN, AWNGN e NAWNGN	111
4.53	Mapas de recorrência:	114
4.54	Gráfico da FAC e da IMM da componente x_k do mapa de Ikeda com ruído	115
4.55	Gráfico da FAC e da IMM do mapa Logístico	117
4.56	Gráfico da FAC e da IMM da componente y do mapa de Hénon	118
4.57	Fluxograma geral para teste de hipótese nula (H_0) utilizando os métodos de dados sub-rogados algoritmo 0 (IID e SS), algoritmo 1 (FT) e algoritmo 2 (AFT, IAAFT-1, IAAFT-2, CAAFT e STAP) aplicados somente a séries temporais não-lineares estática ou dinâmica.	120
4.58	Série de Mackey-Glass com atraso $\tau = 17$	121
4.59	Série de Mackey-Glass	122
4.60	Gráficos quantil-quantil	123
4.61	FAC e IMM da série de Mackey-Glass	124
4.62	Comparação da dimensão de correlação	125
4.63	Séries do NMR	126
4.64	FAC e IMM da série NMR	127
4.65	Janela do sistema de Rössler caótico com ruído dinâmico	129
4.66	Sistema de Rössler	129
4.67	FAC e IMM do sistema de Rössler com ruído dinâmico	130
4.68	Dados sub-rogados PPS gerados da componente x da série temporal caótica do sistema de Rössler com ruído dinâmico	132
4.69	Dados sub-rogados PPS gerados da componente x da série temporal de Rössler de período 6 com ruído dinâmico	133
4.70	Comparação da dimensão de correlação	134
4.71	Série temporal caótica do sistema de Lorenz com inovações de laplace com desvio padrão de 0,05	135
4.72	Gráfico da FAC e da IMM para série caótica de Lorenz com inovações de Laplace	136
4.73	Gráfico da FAC e da IMM para série caótica de Lorenz com inovações de Laplace	137
4.74	Gráfico da FAC e da IMM para série caótica de Lorenz com inovações de Laplace	138
4.75	Gráfico da FAC e da IMM para série caótica de Lorenz com inovações de Laplace	139
4.76	Gráfico da FAC e da IMM para série caótica de Lorenz com inovações de Laplace	140
4.77	Gráfico da FAC e da IMM para série caótica de Lorenz com inovações de Laplace	141
4.78	Comparação da dimensão de correlação	142

4.79	Fluxograma com o procedimento geral para teste de hipótese nula (H_0) utilizando os métodos de dados sub-rogados algoritmo 0 (IID e SS), algoritmo 1 (FT) e algoritmo 2 (AFT, IAAFT-1, IAAFT-2, CAAFT e STAP) aplicados a qualquer série temporal.	143
4.80	Série temporal para o caso “ <i>blind</i> ”	144
4.81	Gráfico da FAC e da IMM para o caso “ <i>blind</i> ” com o IID e SS	145
4.82	Gráfico da FAC e da IMM para série do caso “ <i>blind</i> ” com FT	146
4.83	Gráfico da FAC e da IMM para a série do caso “ <i>blind</i> ” com AAFT	146
4.84	Gráfico da FAC e da IMM para a série do caso “ <i>blind</i> ” com IAAFT-1 e 2	147
4.85	Gráfico da FAC e da IMM para a série do caso “ <i>blind</i> ” com o CAAFT e STAP .	148
A.1	Órbita do sistema de Lorenz	168
A.2	Realizações típicas de dois processos estocásticos	178

Lista de Tabelas

2.1	Tipos de erro associados à realização dos testes estatísticos e suas respectivas probabilidades	14
2.2	Intervalo de confiança para teste	22
3.1	Distribuição dos algoritmos com suas respectivas técnicas para geração de dados sub-rogados	30
3.2	Hipóteses Nulas NH_0 para o algoritmo 0.	31
3.3	Simplex exemplo do método sub-rogado SS (Small e Nakamura (2005))	32
4.1	H_0 associada à realização dos testes estatísticos e sua respectiva probabilidade com 99 dados sub-rogados gerados.	59
4.2	H_0 associada à realização dos testes estatísticos e sua respectiva probabilidade com 39 dados sub-rogados gerados.	74
4.3	H_0 associada à realização dos testes estatísticos e sua respectiva probabilidade com 39 dados sub-rogados gerados.	77
4.4	Resultados da H_0 para o NAWGN, AWNGN e NAWNGN.	111
5.1	Quadro geral dos resultados com respeito a cada hipótese nula proposta para todos os modelos empregados. A rejeição ou não rejeição em negrito identifica uma decisão errônea.	152
A.1	Distribuições de probabilidade unidimensionais	172
B.1	H_0 associada ao tipo de teste estatístico e sua probabilidade com 99 dados sub-rogados gerados a partir da variável aleatória gaussiana.	179
B.2	H_0 associada ao tipo de teste estatístico e sua probabilidade com 99 dados sub-rogados gerados a partir da série temporal da distribuição Uniforme [0,1].	180

B.3	H_0 associada ao tipo de teste estatístico e sua probabilidade com 99 dados sub-rogados gerados a partir série temporal da distribuição Beta(2,5).	181
B.4	H_0 associada ao tipo de teste estatístico e sua probabilidade com 99 dados sub-rogados gerados a partir série temporal da distribuição Laplace(0,1).	182
B.5	H_0 associada ao tipo de teste estatístico e sua probabilidade com 99 dados sub-rogados gerados a partir série temporal da distribuição Cauchy(0,0.001).	183
B.6	H_0 associada ao tipo de teste estatístico e sua probabilidade com 39 dados sub-rogados gerados a partir da série da taxa diária do câmbio Real/USD, seção 4.3.	184

Lista de Símbolos

τ	Atraso de tempo;
$F_x(x)$	c.d.f.;
$cov(x, y)$	Covariância;
ρ_{xy}	Coefficiente de correlação;
$c_m = \langle\langle x^m \rangle\rangle$	Cumulantes;
σ	Desvio Padrão;
σ_Q	Desvio padrão do conjunto estatístico Q_1, \dots, Q_N ;
d_c	Dimensão de correlação;
d_e	Dimensão de imersão;
$N(0, 1)$	Distribuição normal com média zero e variância unitária;
$P(w)$	Espectro de Potência;
Q	Estatística de teste;
Q_0	Estatística de teste para o dado original;
$f_x(x)$	f.d.p. ou p.d.f.;
Φ_x	Função característica;
C_N	Função de correlação ou integral de correlação;
$r_{xx}(\tau)$	Função de Autocorrelação linear;
H_1	Hipótese alternativa;
H_0	Hipótese nula;
NH_0	Hipótese nula para o algoritmo 0;
NH_1	Hipótese nula para o algoritmo 1;
NH_2	Hipótese nula para o algoritmo 2;
NH_{CS}	Hipótese nula para o algoritmo cycle-shuffle;
NH_{SS}	Hipótese nula para o algoritmo Small-Shuffle;
NH_{PPS}	Hipótese nula para o algoritmo PPS;

μ	Média aritmética;
$\langle Q \rangle$	Média do conjunto estatístico Q_1, \dots, Q_N ;
$E[x^n] = \langle x^n \rangle$	Momento de ordem n ;
$m_n = \langle [x - \mu]^n \rangle$	Momento centrado de ordem n ;
α	Nível de significância ou probabilidade de cometer erro do tipo I;
N_s	Número de dados sub-rogados;
$f_{x,y}(x, y)$	p.d.f. conjunta;
β	Probabilidade de falhar em rejeitar a H_0 quando ela for falsa (erro tipo II);
S	Valor quantitativo calculado no método paramétrico;
σ^2	Variância;

Lista de Abreviações

AAFT	Ajuste de amplitude por transformada de Fourier (<i>amplitude adjusted Fourier transform</i>);
AR	Auto-regressivo;
ARMA	Auto-regressivo de média móvel (<i>autoregressive moving average</i>);
CAAFT	Corrected AAFT;
c.d.f.	Função de distribuição cumulativa ou função de distribuição de probabilidade;
CS	Embaralhamento dos ciclos (<i>cycle-shuffle</i>);
e.d.o.	Equações diferenciais ordinárias;
FAC	Função de autocorrelação;
FC	Função característica;
f.d.p. ou p.d.f	Função densidade de probabilidade;
FT	Transformada de Fourier;
IAAFT	Iterativo AAFT (<i>iterated AAFT</i>);
i.d.	Independetemente distribuído;
i.i.d.	Independentes e identicamente distribuído;
PPS	Sub-rogado pseudo-periódico (<i>pseudoperiodic surrogate</i>);
QQ	Quantile-quantile;
S	Valor quantitativo calculado no método paramétrico;
SS	Small-shuffle;
STAP	Processo autoregressivo estatisticamnete transformado (<i>statically transformed autoregressive process</i>);
TDF	Transformada discreta de Fourier;
TLC	Teorema do limite central;
v.a	Variável aleatória;

Introdução

A classe de fenômenos cujo processo observacional e conseqüente quantificação numérica gera uma seqüência de dados distribuídos no tempo é denominada série temporal (Morettin e Toloí, 2006).

A natureza de uma série temporal e a estrutura de seu mecanismo gerador está relacionada com o intervalo de ocorrência das observações no tempo. Caso o levantamento das observações da série possa ser feito a qualquer momento do tempo, a série temporal é dita contínua, sendo denotada por $x(t)$ (Granger e Newbold, 1977). Entretanto, de acordo com Granger e Newbold (1977) e Nelson (1973), na maioria das séries, as observações são tomadas em intervalos de tempo discretos e eqüidistantes.

Uma série temporal discreta pode ser representada por $X^T = \{x_1, x_2, \dots, x_T\}$, sendo que cada observação discreta está associada a um instante de tempo distinto, existindo uma relação de dependência serial entre essas observações (Souza, 1989).

Ao analisar uma ou mais séries temporais a representação gráfica dos dados seqüencialmente ao longo do tempo é fundamental e pode revelar padrões de comportamento importantes. Logo, a previsão desse comportamento é fundamental para análise. As previsões de séries temporais são bastante comuns nas ciências naturais, exatas, engenharias, áreas social e médica (Coelho et al., 2006). Nas últimas décadas tem-se verificado um grande crescimento no estudo em previsão de séries temporais introduzindo novas abordagens. Os estudos de séries temporais são umas das atividades mais antigas e relevantes da ciência, como bem expressa Ljung (1987) em seu livro *System Identification - Theory for the user*:

“Inferir modelos a partir de observações e estudar suas propriedades é o que realmente faz a ciência. Os modelos (“hipóteses”, “leis da natureza”, “paradigmas”, etc.) podem apresentar um caráter mais

ou menos formal, mas todos eles possuem uma mesma característica básica, que é a tentativa de encontrar algum padrão em observações.”¹

As técnicas disponíveis na literatura nem sempre são adequadas quando se deseja determinar se a dinâmica que deu origem à uma série temporal é, por exemplo, determinística (ou não). Logo há de se lançar mão de outros métodos para identificação de determinismo, como o teste de hipótese nula utilizando o método de dados sub-rogados. O procedimento utilizando teste de hipótese nula baseado em séries sub-rogadas (do inglês *surrogates*) descrito inicialmente por Theiler et al. (1992) tornou-se uma ferramenta indispensável para análise de séries temporais de sistemas com comportamento dinâmico em muitas áreas do conhecimento, tais como a fisiologia, geofísica, economia e áreas afins da engenharia (Kugiumtzis, 2002a).

O método de dados sub-rogados foi concebido inicialmente como uma ferramenta para verificar a presença de não-linearidades (Theiler et al., 1992; Schreiber e Schmitz, 1996; Schreiber, 1998; Kugiumtzis, 1999c) como, por exemplo, a identificação de caos (Stam et al., 1998; Schreiber et al., 1999; Small et al., 2001b; Small e Tse, 2002; Small et al., 2005a) ou qualquer outra dinâmica de interesse através da escolha apropriada de um teste estatístico para a hipótese nula (Theiler et al., 1992; Theiler e Prichard, 1995). A idéia principal de Theiler é gerar um conjunto de dados, denominados sub-rogados, a partir da série temporal original, que preservam determinadas características a qual se deseja analisar estatisticamente, tais como a distribuição de amplitude e o espectro de potência.

Atualmente, há uma gama bastante ampla de aplicações baseadas no método de hipótese nula aplicando dados sub-rogados para problemas de diversas natureza, abrangendo vários estudos sobre o comportamento dos sinais. Como citados a seguir: comportamento dos sinais biológicos ou biomédicos tais como o eletro-encefalograma (EEG) (Theiler et al., 1992; Theiler e Prichard, 1994; Theiler, 1995; Small e Judd, 1998a; Kugiumtzis, 1999a, 2000a; Small et al., 2001b; Kugiumtzis, 2002b, 2006; Furtado et al., 2006), o eletro-cardiograma (ECG) (Schreiber e Schmitz, 1998; Govindan et al., 1998; Small et al., 2001b; Small e Tse, 2002), os movimentos ou contrações abdominais (*inductance plethymography*) (Small e Judd, 1998b; Zhang et al., 1998; Small e Nakamura, 2006a); na identificação de sistemas com comportamento caótico (Stam et al., 1998; Schreiber et al., 1999; Small et al., 2001b,a; Small e Tse, 2002, 2003; Small et al., 2005a; Small

¹(Ljung, 1987, p. 1)

e Nakamura, 2005; Small et al., 2006a); na análise de dados físicos experimentais como o NMR laser (*Ressonância Magnética Nuclear*) (Schreiber e Schmitz, 1999; Schreiber et al., 1999; Nagarajan, 2005; Small e Nakamura, 2005; Small et al., 2006a), as manchas solares (*sunspot*) (Small e Tse, 2002; Small e Nakamura, 2005; Kugiumtzis, 2006; Small et al., 2006a; Small e Nakamura, 2006a), a taxa do isótopo de Oxigênio (Rial e Anaclerio, 2000; Schreiber e Schmitz, 2001); na análise de séries de resíduos (Coelho et al., 2006); na análise em dados climáticos como o comportamento das temperaturas nos grandes centros urbanos (Small et al., 2005b; Small e Nakamura, 2006a; Small et al., 2006a); na análise de modelos ARMA ou NARMA contaminados com ruído não-gaussiano (Nagarajan, 2005; Small et al., 2006b) e mapas (não-lineares) (Mendes et al., 2002; Nagarajan, 2005; Small et al., 2006b); para uma vasta aplicabilidade em dados financeiros (Schreiber e Schmitz, 1998; Kugiumtzis, 2000b; Small e Tse, 2002; Small et al., 2005b; Small e Nakamura, 2006b) e outras séries temporais desenvolvidas em diversos processos.

Nesse contexto amplo de aplicabilidade, o uso de ferramentas para auxiliar a análise e a escolha da modelagem a ser empregada é de grande interesse e, evidentemente, uma resposta bem definida a esse problema será de imensa importância em praticamente todas as aplicações científicas.

O método para gerar dados sub-rogados para teste de não-linearidade ficou popular nos últimos anos, especialmente com respeito à hipótese nula de que a série temporal examinada é gerada por um processo gaussiano linear que sofre uma transformação estática não-linear. Dados sub-rogados corretamente gerados para esta hipótese nula deveriam possuir a mesma autocorrelação (espectro de potência) e distribuição de amplitude da série original. Porém, nem sempre os algoritmos empregados provêm dados sub-rogados que preservam as correlações lineares originais ou a distribuição da amplitude, e isto pode conduzir a falsas rejeições (Kugiumtzis, 2002a). Isso é devido à estatística de teste aplicada não distinguir se o sinal era gaussiano, por isso poderia conduzir a resultados errôneos, sendo que na verdade rejeitava-se o fato delas terem distribuições diferentes. Logo são propostos outros testes estatísticos (não-paramétrico) para análise da não-linearidade como, por exemplo, a informação mútua média (IMM) e a dimensão de correlação (d_c).

Algumas das principais técnicas para gerar séries temporais sub-rogadas para análise estatística de processos são listadas a seguir: embaralhamento temporal dos dados (Theiler et al., 1992; Kugiumtzis, 2002c), aplicação da Transformada de Fourier (FT) (Theiler e Prichard, 1994; Schreiber, 1997) e Transformada

de Fourier com Ajuste de Amplitude(AAFT) (Theiler et al., 1992; Kugiumtzis, 1999c, 2000b).

A partir destas três foram propostas outras técnicas, tais como Interativo Transformada de Fourier com Ajuste de Amplitude (IAAFT) (Schreiber e Schmitz, 2000, 1996), o Janelamento da Transformada de Fourier (WFT) (Suzuki et al., 2005), o *Correct* Transformada de Fourier com ajuste de amplitude (CAAFT) (Kugiumtzis, 2000a), o embaralhamento curto (*Small-Shuffle- SS*) (Small e Nakamura, 2005), o embaralhamento de ciclos (*cycle-shuffle*) (Theiler, 1995), *statically transformed autoregressive process* (STAP) (Kugiumtzis, 2002b) e o sub-rogado pseudo-periódico (PPS) ou *pseudoperiodic surrogate* (Small et al., 2005a) para séries pseudo-periódicas.

A questão básica do trabalho é entender justamente o mecanismo gerador da série temporal bem como verificar se há algum outro tipo de comportamento dinâmico ou aleatório nesta série temporal através do teste de hipótese utilizando dados sub-rogados. Baseado na detecção de comportamento determinístico ou não das séries temporais sob análise é identificado que nem sempre tais técnicas sub-rogadas e as estatísticas de testes são adequadas e que existe situações que conduzem a decisões errôneas para certas estruturas de modelos geradores, relatadas nesse trabalho.

Em resumo, este trabalho propõe uma análise, por meio de um procedimento, para o quadro geral de possíveis comportamentos para a série temporal, ou seja, um procedimento utilizando de maneira consistente os métodos de dados sub-rogados o qual possibilita detectar aleatoriedade (ruído) ou determinismo (linear, não-linear, etc). Logo, visa-se estabelecer um procedimento o qual auxilia os usuários a melhor escolha dos algoritmos e índices de acordo com problema.

1.1 Objetivos

Em muitas situações práticas na engenharia é comum coletar um conjunto de dados composto de diversos sistemas subjacentes ou corrompidos por ruídos. Nessas situações, entender o tipo de comportamento dinâmico ou estocástico da série temporal facilita muito a identificação do possível modelo matemático a ser empregado. Diante desse fato e da crescente aplicação do teste de hipótese utilizando o método de dados sub-rogados em séries temporais, esse trabalho visa contribuir:

1. No estudo e formulação do conjunto de hipóteses e testes estatísticos em-

- pregados;
2. Estudo dos métodos para gerar dados sub-rogados;
 3. Implementação dos algoritmos geradores de dados sub-rogados para validação do método;
 4. Levantamento das condições e dos possíveis problemas dos métodos aplicados;
 5. Formulação de um algoritmo com um procedimento de análise para o quadro geral de possíveis comportamentos da série temporal.

1.2 Motivação

Nos últimos anos o método de dados sub-rogados tem recebido muita atenção e vem sendo amplamente aplicado por vários autores para análise de séries temporais, principalmente não-lineares (Kugiumtzis, 2002a). A motivação inicial com este trabalho é com a investigação do comportamento dinâmico da série temporal, ou mecanismo gerador da série, que tem fundamental importância e produz resultados imprescindíveis para modelos de previsões.

Muitos autores concordam que estes métodos para geração de séries sub-rogadas contêm algumas falhas, mas que nem sempre são fáceis de evitar (Theiler et al., 1992). Desta forma, a motivação secundária desse trabalho é em avaliar as técnicas disponíveis na literatura para geração das séries temporais sub-rogadas identificando as possíveis falhas para cada caso.

Todavia a principal motivação é a formulação de um procedimento de análise para o quadro geral de possíveis comportamentos da série temporal. Proporcionando ao usuário “dicas” de qual o melhor algoritmo a ser usado em uma dada situação.

1.3 Problemas

O teste de dados sub-rogados é freqüentemente um método aplicado para avaliar os resultados das análises de séries temporais, principalmente não-linear. Existem diversos problemas encontrados na literatura para essas análises levantadas. Uns dos problemas levantados é que um resultado negativo contra a hipótese nula de um processo estocástico linear gaussiano estacionário não só pode ser o

resultado de um sistema subjacente não-linear ou até mesmo um sistema caótico, mas também uma não-estacionariedade linear. Timmer (1998) investiga o poder do teste contra não-estacionariedades.

Schmitz e Schreiber (2004) ressalta o cuidado quanto a quantidade de observações e a estatística de teste que deve ser escolhida, desde que nenhuma teoria geral garante o teste corretamente.

Um outro problema encontrado é com o uso do sub-rogado FT devido a periodicidade artificial introduzida onde apenas a autocorrelação finita, cíclica é preservada, não a autocorrelação de uma duração teoricamente infinita (Schreiber e Schmitz, 2000). A autocorrelação cíclica pode diferir sistematicamente da original. Descontinuidade também pode ser introduzida, de forma que as características de energia de altas frequências dos sub-rogados são diferentes da original (Small et al., 2001b).

Coelho et al. (2006) sugere que o método sub-rogado FT (algoritmo 1) não é adequado para aplicação em séries temporais com distribuição não-gaussiana, uma vez que as séries temporais artificiais assim geradas terão distribuições diferentes da série original, supostamente gaussiana.

Nagaranjan (2005) revela que sub-rogados que testam procedimentos que re-têm a função de auto-correlação podem não ser apropriados na presença de ruídos não-gaussianos.

Segundo Kugiumtzis (2001) cada parâmetro livre propostos nos algoritmos geradores dos dados sub-rogados CAAFT e STAP, usado para o calculo da estatística não-linear, aumenta a probabilidade que os resultados do teste dependerão da escolha destes parâmetros livres.

Little et al. (2006) também sugere que os sub-rogados gerados nem sempre são realmente construídos conforme a hipótese nula proposta. Little et al. (2006) demonstra também evidências experimentais, comportamento vocal, que as suposição da hipótese de uma série temporal linear gaussiana não pode explicar a todo comportamento dinâmico real.

Outras análises críticas dos algoritmos sub-rogados podem ser achadas no artigo de Schreiber e Schmitz (2000), que revela a suscetibilidade das técnicas sub-rogadas para processos não-estacionarios, baseado em certas suposições, podem conduzir a resultados errôneos.

1.4 Relevância

Este trabalho procura fornecer subsídios para levantamento e descrição das hipóteses nulas, dos métodos para gerar dados sub-rogados e das estatísticas de testes aplicadas na literatura, bem como as análises para a tomada de decisão a favor ou contra a hipótese nula. Levantado as condições e os possíveis problemas dos métodos de dados sub-rogados aplicados .

Outro ponto relevante é o desenvolvimento de uma formulação de um procedimento geral para análise do comportamento de séries temporais utilizando os testes de hipóteses estatísticas com base no método de dados sub-rogados. Sendo assim, possível avaliar, por meio do conhecimento *à priori* do modelo, as técnicas disponíveis na literatura para a geração das séries sub-rogadas de acordo com a hipótese nula empregada.

A formulação de um procedimento de análise para o quadro geral de possíveis comportamentos da série temporal proporciona ao usuário “dicas” de qual o melhor algoritmo a ser usado em uma dada situação.

1.5 Organização do Texto

No Capítulo 2 são discutidos os conceitos essenciais e as propriedades para testes de hipóteses estatísticas. Na seção 2.2 descreve a formulação da hipótese nula para teste de acordo com o mecanismo gerador de dados sub-rogados. Na seção 2.3 as estatísticas de testes paramétricas e não-paramétricas são propostas para o teste de hipótese nula. Na seção 2.4 é feita uma análise para exigência mínima de significância para afirmação ou rejeição da hipótese nula com relação ao número de dados sub-rogados gerados para teste.

No Capítulo 3, uma revisão bibliográfica a respeito dos mecanismos geradores de dados sub-rogados. São apresentadas as principais técnicas (algoritmos) que aparecem na literatura. Os procedimentos dos mecanismos geradores são relatados com o objetivo de contextualizar as técnicas desenvolvidas para cada tipo de hipótese nula.

O Capítulo 4 é de cunho muito prático. São tratados os algoritmos geradores e as estatísticas de testes aplicadas a modelos conhecidos a priori (simulações) e dados experimentais (reais). A avaliação da decisão com respeito à hipótese nula empregada é aplicada e discutida com o intuito de identificar qual a característica (estocástico, determinismo, caos, pseudo-periódicos, linear, não-linear, etc) do mecanismo gerador da série temporal original, tratando de aspectos diversos com

respeito ao mecanismo gerador e sua hipótese. Neste capítulo são tratados os problemas e restrições utilizando as técnicas descritas no texto.

Finalmente, uma breve exposição das conclusões obtidas neste projeto de pesquisa são sintetizadas no Capítulo 5.

Testes de Hipóteses Estatísticas

2.1 Introdução

A aplicação formal do método de dados sub-rogados é expressa na linguagem estatística como teste de hipótese. Isto envolve dois ingredientes: uma hipótese nula contra a qual as observações são testadas, uma estatística de teste (fator discriminante estatístico) que é um número que quantifica ou análise gráfica de algum aspecto da série temporal em questão. Neste capítulo, os conceitos envolvidos nos testes de hipóteses estatísticas aplicados em séries temporais serão discutidos.

Definição 2.1.1 *“O teste de hipótese estatística é a afirmação ou rejeição (probabilística) concernente ao tipo de comportamento da série temporal em estudo.”*

A idéia principal de teste de hipótese nula aplicando o método de dados sub-rogados é gerar um conjunto de dados, denominados sub-rogados que será visto no Capítulo 3, a partir da série temporal original, que compartilham determinadas características. Essas características são analisadas usando estatísticas de testes que rejeitam ou não a hipótese nula de interesse. É importante enfatizar que a construção do dado sub-rogado é realizada de acordo com a hipótese nula de interesse (Theiler et al., 1992).

Neste capítulo, são discutidos os conceitos essenciais e as propriedades para testes de hipóteses estatísticas. A seção 2.2 a formulação da hipótese nula é descrito que pode ser simples ou composta, para teste de acordo com o mecanismo gerador de dados sub-rogados. Na seção 2.3, as estatísticas de testes paramétricas e não-paramétricas são propostas para o teste de hipótese nula. Na seção 2.4 é feita uma análise para exigência mínima de significância (95% ou 99%) para afirmação ou rejeição da hipótese nula com relação ao número de dados sub-rogados gerados para teste.

Neste Capítulo o conceito de dados sub-rogados será avaliado no contexto de testes de hipóteses. Não obstante, a definição e os vários tipos de dados sub-rogados serão tratados no Capítulo 3.

2.2 Hipótese Estatística

O campo da matemática estatística provê uma linguagem e uma ferramenta para lidar com as questões que surgem no estudo do comportamento dinâmico, principalmente não-linear, sendo uma destas, por exemplo, na dinâmica caótica. Para entender o mecanismo responsável para gerar as diferentes séries temporais é preciso responder algumas perguntas elementares primeiro (Small et al., 2006a): É caos ou meramente ruído, linear ou não-linear, estocástico ou determinístico, pseudo-periódico ou caos? Nesse tipo de pergunta que é feita freqüentemente na análise de séries temporais não-lineares, o teste de hipótese surge como uma ferramenta importante. Neste tipo de teste, inicia-se formulando perguntas sobre o conjunto de dados de interesse. Por exemplo (Theiler e Prichard, 1995),

- É não-gaussiana?
- Sua média é significativamente diferente de zero?
- Se é uma série temporal, há alguma correlação temporal?
- Há alguma estrutura não-linear nas correlações temporais?
- É caos?

A partir destas perguntas formula-se as hipóteses que se deseja testar. A hipótese que se deseja testar é chamada de Hipótese nula (ou H_0).

Muitos problemas na engenharia requerem decisão entre aceitar ou rejeitar uma afirmação acerca de algum parâmetro ou processo. A afirmação é chamada de **hipótese** e o procedimento de tomada de decisão sobre a hipótese é chamado de **teste de hipóteses**. Uma definição formal de uma hipótese estatística é dada a seguir (Callegari-Jacques, 2006):

Definição 2.2.1 “Uma *hipótese estatística* é uma afirmação ou suposição feita sobre algum valor (dos parâmetros) de uma ou mais populações¹.”

Usualmente a princípio, utiliza-se distribuição de probabilidades para representar populações ou a distribuição sobre estudo. Uma hipótese estatística também pode ser pensada como uma afirmação acerca da distribuição de probabilidade de uma variável. A hipótese envolverá um ou mais parâmetros dessa distribuição. É interessante imaginar o teste estatístico de hipótese como o estágio de

¹A população diz respeito a um conjunto de todos os elementos onde, cada um deles, apresenta uma ou mais características em comum. Quando se extrai um conjunto de observações da população ou seja, toma-se parte desta para a realização do estudo, tem-se a amostra.

análise dos dados de um experimento comparativo, tal como, comparar a média de uma população a um certo valor de interesse. Esses experimentos comparativos simples são frequentemente encontrados na prática e fornecem uma boa base para problemas mais complexos.

É importante, neste momento, distinguir dois tipos de hipótese nula: **simples** e **composta**. Uma hipótese nula simples afirma que um determinado conjunto de dados é uma realização aleatória de um único processo específico. A hipótese nula composta especifica uma família de processos, e afirma que o processo que de fato gerou os dados é um membro dessa família. Por exemplo, uma distribuição gaussiana com média zero e variância unitária é uma hipótese simples. A hipótese mais ampla é que os dados são membros de uma distribuição gaussiana, de média e variância não especificadas corresponderá a hipótese composta (Theiler e Prichard, 1995).

Definição 2.2.2 *Seja \mathbb{F} o espaço do processo em consideração, e seja ϕ uma hipótese específica e $F_\phi \subset \mathbb{F}$ o conjunto de processos (ou sistemas) que são consistentes com a hipótese nula. Considere que a hipótese nula diz que um processo particular \hat{F} que gerou os dados é um elemento do conjunto F_ϕ . Se este conjunto consistir em um simples membro, então a hipótese nula é simples. Caso contrário, a hipótese nula será composta, e se diz que os dados foram gerados por algum processo $\hat{F} \in F_\phi$, mas não especifica qual \hat{F} .*

Para melhor compreensão de testes de hipótese estatística propõe-se o seguinte exemplo idealizado retirado do site² do grupo *TOPUS* da Universidade de São Paulo (USP). Suponha que se tenha interesse na taxa de queima de um propelente sólido³, usado para fornecer energia aos sistemas de escapamento de aeronaves. A taxa de queima é uma variável aleatória que pode ser descrita por uma distribuição de probabilidades. Suponha que se esteja interessado na taxa média de queima do propelente sólido à 68,02 atm. Especificamente, está interessado em decidir se a taxa média de queima é ou não 1.529 cm/s. Pode-se expressar isso formalmente como

$$H_0 : \mu = 1,529 \text{ cm/s}$$

$$H_1 : \mu \neq 1,529 \text{ cm/s}$$

²<http://www.eesc.usp.br/smm/topus>

³Atualmente a USP trabalha apenas com propelentes sólidos compostos, ou seja, aqueles no qual o combustível e o oxidante são pulverizados e muito bem misturados em proporções adequadas para garantir a completa combustão.

A afirmação $H_0 : \mu = 1,529 \text{ cm/s}$ é chamada de **hipótese nula** e a afirmação $H_1 : \mu \neq 1,529 \text{ cm/s}$ é chamada de **hipótese alternativa**. Uma vez que a hipótese alternativa especifica valores de μ que poderiam ser maiores ou menores do que 1,529 cm/s, ela é chamada de **hipótese alternativa bilateral**. Em algumas situações, pode-se desejar formular uma **hipótese alternativa unilateral**, como em

$$H_0 : \mu = 1,529 \text{ cm/s}$$

$$H_1 : \mu > 1,529 \text{ cm/s}$$

ou

$$H_0 : \mu = 1,529 \text{ cm/s}$$

$$H_1 : \mu < 1,529 \text{ cm/s}$$

Em todas as análises feitas com teste de hipótese estatística com os métodos de dados sub-rogados observa-se que a *hipótese alternativa não será especificada*, isto é devido ao número indefinido de possíveis famílias não consistentes com a H_0 conforme a definição 2.2.2.

O objetivo do teste de hipóteses é geralmente determinar se o valor do parâmetro observado variou. É importante lembrar que hipóteses são sempre afirmações sobre a população ou distribuição em estudo. O valor do parâmetro especificado da população na hipótese nula (1,529 cm/s do exemplo dado) é geralmente determinado de três maneiras. Primeiro, ele pode resultar de experiência passada ou de conhecimento do processo ou mesmo de testes prévios. O segundo é se esse valor pode ser determinado a partir de alguma teoria ou do modelo relativo ao processo sob estudo. Uma terceira situação aparece quando o valor do parâmetro da população resulta de considerações extremas, tais como projeto ou especificações de engenharia.

Um procedimento levantado para uma tomada de decisão acerca de uma hipótese particular usam as informações de um conjunto de amostras provenientes da população de interesse. Se essa informação for consistente com a hipótese, então conclui-se que a hipótese é verdadeira; no entanto, se essa informação for inconsistente com a hipótese, conclui-se que a hipótese é falsa. Contudo é importante enfatizar que a verdade ou falsidade de uma hipótese particular pode nunca ser conhecida com certeza, a menos que se possa examinar a **população inteira**. Isso geralmente é impossível em muitas situações práticas. Desse modo, um procedimento de teste de hipótese deveria ser desenvolvido, tendo-se em mente a probabilidade de alcançar uma conclusão errada. Existem alguns procedimentos

de decisões que podem conduzir a conclusões erradas, tais como: o Erro Tipo I e Tipo II.

Definição 2.2.3 *A rejeição da hipótese nula H_0 , quando ela for verdadeira, é definida como um **Erro Tipo I**.*

Normalmente, um teste é projetado com um tamanho pré-especificado do teste “ α ” também chamado de **nível de significância**, para o qual corresponde uma esperada taxa de erro tipo I. Este nível de significância pode ser maior ou menor dependendo do nível desejado de confiança para evitar falsas afirmações. Convencionalmente, $\alpha = 0,05$ (ou 5%) é um valor considerado significativo. Um teste é preciso se o valor nominal de α corresponder à real probabilidade de cometer um erro tipo I. Alguns autores citam que este valor para teste de hipótese é o valor limiar de α ao qual a hipótese nula ainda seria rejeitada.

$$\alpha = \text{Prob}(\text{erro tipo I}) = \text{Prob}(\text{rejeitar } H_0 \text{ quando } H_0 \text{ for verdadeira}) \quad (2.1)$$

Geralmente, os testes de hipóteses possuem uma **região de rejeição** (também chamada região crítica) para teste onde se rejeita H_0 e uma outra região para a qual não se rejeita H_0 que por convenção chama-se geralmente de **região de aceitação**. Os limites entre as regiões críticas e de aceitação são chamados de **valores críticos** definindo assim o intervalo de confiança da estatística de teste, escolhidos através do nível de significância α .

Definição 2.2.4 *A não rejeição (aceitar) a hipótese nula H_0 , quando ela é falsa, é definida como um **Erro Tipo II**.*

A probabilidade disto acontecer é normalmente denotada β , e $1 - \beta$ é chamado de **potência** de um teste.

$$\beta = \text{Prob}(\text{erro tipo II}) = \text{Prob}(\text{falha rejeitar } H_0 \text{ quando } H_0 \text{ for falsa}) \quad (2.2)$$

A potência de um teste estatístico é a probabilidade de rejeitar a hipótese nula H_0 , quando a hipótese nula for falsa. A potência é uma medida concisa da sensibilidade de um teste estatístico, em que por sensibilidade entende-se a habilidade do teste de detectar diferenças.

A Tabela 2.1 relata resumidamente os tipos de erro e suas respectivas probabilidades.

Tabela 2.1: Tipos de erro associados à realização dos testes estatísticos e suas respectivas probabilidades

Decisão	H_0 É Verdadeira	H_0 É Falsa
Aceitar H_0	nenhum erro Probabilidade: $1 - \alpha$	erro tipo II Probabilidade: β
Rejeitar H_0	erro tipo I Probabilidade: α	nenhum erro Probabilidade: $1 - \beta$

Após essa definições pode-se apresentar pelo menos três pontos importantes para o teste de hipótese:

1. O tamanho da região crítica, e conseqüentemente a probabilidade de cometer o erro tipo I (α) pode sempre ser reduzido através da seleção apropriada dos valores críticos;
2. Os erros tipo I e tipo II estão relacionados. Uma diminuição na probabilidade de um tipo de erro sempre resulta em um aumento de probabilidade do outro, desde que o tamanho da amostra (N) não varie;
3. Um aumento no tamanho da amostra reduzirá, geralmente, α e β , desde que os valores críticos sejam mantidos constantes.

A idéia principal é gerar um conjunto de dados, denominados sub-rogados, a partir da série temporal original. Aplica-se uma estatística de teste ao dados sub-rogados gerados e a série temporal original. Posteriormente toma-se a decisão em rejeitar a H_0 com base nas análise paramétrica ou não paramétricas das estatísticas de testes aplicadas.

Segundo Small et al. (2006a) teste de hipótese nula utiliza medidas estatísticas do sistema e dos dados sub-rogados para determinar a probabilidade que uma hipótese proposta seja verdadeira (ou falsa). Os procedimentos comuns incluem:

1. Formular a hipótese nula de interesse e os riscos potenciais associados com uma decisão;
2. Escolher um estatística de teste q_i ;
3. Calcular a distribuição de frequência da estatística de teste sob a hipótese nula (estatística da série original q_0 e conjunto das estatísticas de testes dos dados sub-rogados $\{q_1, \dots, q_{N_s}\}$);

4. Com o padrão da distribuição de frequência estatísticas, escolha certo critério distintivo para determinar se rejeita (ou não) a hipótese nula. Método paramétrico (cálculo de S) ou método não-paramétrico (análise gráfica).

Tradicionalmente, para obter a distribuição de frequência da estatística de teste (para a série temporal original e dos dados sub-rogados), precisa escolher cuidadosamente a estatística de teste, tal que siga uma distribuição bem conhecida, como por exemplo, a distribuição normal $N(0, 1)$, comumente utilizada.

A idéia básica é produzir várias realizações diferentes sob a hipótese nula por meio da **simulação de Monte-Carlo**⁴ ou **teste de dados sub-rogados** (Small et al., 2006a). Na prática, estas realizações são normalmente geradas dos dados experimentais originais. Para o conjunto de dados sub-rogados gerados, pode-se calcular a distribuição empírica e o intervalo de confiança da estatística de teste, escolhidos através do nível de significância α . Nesse sentido, a distribuição de frequência dependerá essencialmente do algoritmo para geração de dados sub-rogados e da estatística escolhida. Logo, também poderia-se dizer que o algoritmo sub-rogado é um dos elementos que formam um teste de hipótese nula.

Três tipos de hipóteses nulas são comumente testadas para os chamados métodos geradores de dados **sub-rogados lineares**⁵, os quais são geralmente sinais estacionários e sem qualquer termo de tendência longo ou periodicidade, conhecidos como NH_0 , NH_1 e NH_2 que formam uma estrutura hierárquica (Small et al., 2006a):

- NH_0 : Para testar se o conjunto de dados da série temporal original são ruídos independentes e identicamente distribuídos (i.i.d.).
- NH_1 : Para testar se o conjunto de dados da série temporal original em teste é um processo estocástico linear gaussiano, onde os dados sub-rogados são produzidos por um processo estocástico gaussiano linearmente autocorrelacionado na forma de um modelo ARMA (*Autoregressive Moving Average*) com parâmetros desconhecidos.
- NH_2 : Para testar se o conjunto de dados da série temporal original em teste é um processo estocástico linear, onde os dados sub-rogados são obtidos

⁴O método de teste hipótes de Monte-Carlo é chamado freqüentemente de teste de dados sub-rogados, ou método de dados sub-rogados na literatura .

⁵O termo “sub-rogados lineares” foi muito usado por Schreiber (1997); Schreiber e Schmitz (2000). Também foi usado que a hipótese nula dos métodos proposto por Theiler et al. (1992) e Schreiber e Schmitz (2000) é uma hipótese nula linear de um processo estocástico linear.

aplicando um filtro estático, monotônico e não-linear (h) à série temporal originalmente gerada por um processo ARMA, ou seja, uma transformação estática, monotônica, não-linear de um processo linear (série original).

A Figura 2.1 ilustra a hierarquia das hipóteses nulas relatadas anteriormente.

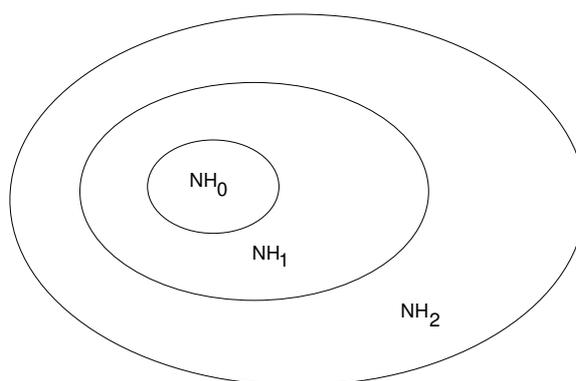


Figura 2.1: Hierarquia das Hipóteses lineares

É importante relatar que os dados (v.a.s i.i.d.) ou ruídos não precisam ser necessariamente gaussianos (Small, 2005). Em NH_1 os dados em teste são produzidos de um processo estocástico linear geralmente na forma um modelo auto-regressivo (AR) ou um modelo auto-regressivo de média móvel (ARMA) com parâmetros desconhecidos que são essencialmente um ruído i.i.d. gaussiano $N(0, \sigma^2)$ linear. Por outro lado, em NH_2 os dados em teste são obtidos aplicando uma transformação estática não-linear monotônica à série temporal original gerada por um processo ARMA.

De acordo com estas hipóteses, pode-se determinar se a série temporal observada é um ruído i.i.d. (NH_0) ou um processo estocástico linear (NH_1 e NH_2). Os três algoritmos para gerar estes dados sub-rogados são conhecidos como algoritmo 0, algoritmo 1 e algoritmo 2 correspondendo, respectivamente, a NH_0 , NH_1 e NH_2 . A natureza exata do algoritmo para gerar dados sub-rogados deve ser consistente com a hipótese escolhida.

Algoritmo 0: Os dados sub-rogados são gerados através do embaralhamento da ordem temporal (*Random Shuffle*) da série temporal original eliminando as-

sim qualquer correlação temporal. Em essência os sub-rogados são dados aleatórios (ruídos) i.i.d. consistentes com a mesma distribuição de probabilidade que o conjunto original de dados.

Algoritmo 1: Os dados sub-rogados são produzidos por um ruído gaussiano linear. Emprega-se a transformada discreta de Fourier (DFT) do conjunto original de dados e embaralha-se as fases dos pares conjugados complexos. O dado sub-rogado é a transformada discreta inversa de Fourier. Conseqüentemente, embora a correlação linear (o espectro de potência) seja preservada, qualquer estrutura determinística (não-linear) adicional é destruída. Em particular, a distribuição de probabilidade da série temporal original não é preservada nos dados sub-rogados gerados. Para evitar qualquer influência estatística, é preferível que os dados sub-rogados tenham a mesma distribuição de probabilidade dos dados originais.

Algoritmo 2: Dados sub-rogados gerados por este algoritmo são transformações estáticas não-lineares monotônicas de um processo gaussiano linear. Gerando dados sub-rogados que preservam a distribuição de probabilidade e o espectro de potência da série temporal original.

Em muitos algoritmos que geram dados sub-rogados não ocorrem a preservação do comportamento da estrutura global da série temporal original e geralmente as hipóteses empregadas nos métodos sub-rogados lineares não são claramente condizentes com as séries temporais existentes. Por causa destes problemas um nova técnica e hipótese nula surgiram (Small et al., 2001b, 2005a) para investigar dados com flutuações irregulares

- NH_{SS} : As flutuações irregulares são variáveis aleatórias independentemente distribuídas (i.d.).

A hipótese NH_{SS} , empregada para identificar séries com flutuações irregulares, geralmente dados financeiros, diz que as flutuações irregulares são variáveis aleatórias independentemente distribuídas (i.d.), em outras palavras, não há nenhuma dinâmica com termos curtos nas flutuações irregulares. Uma aplicação deste método de dados sub-rogados é detectar possivelmente correlações entre flutuações irregulares com termos de tendência longas (Small e Nakamura, 2005). A hipótese nula correspondente é que as flutuações irregulares estão distribuídas independentemente, que diferentemente da NH_0 , não requer a distribuição idêntica das flutuações. A premissa básica da técnica é que se as flutuações irregulares

não forem aleatórias há algum tipo de sistema dinâmico subjacente: modulado por qualquer tendência que contaminam os dados (ruídos).

Algoritmo SS: Em tal algoritmo, o índice (ordem) do próprio dado tem implicações importantes, independente da série temporal ser linear ou não-linear. Conseqüentemente, sempre que o índice dos dados mudam, o fluxo de informação também muda e a resultante da série temporal não reflete a dinâmica original. O algoritmo sub-rogado **Small-Shuffle** (SS) (Small et al., 2005b; Small e Nakamura, 2005) usa esta idéia. Isto é semelhante ao método sub-rogado *Random-Shuffle* (RS)⁶ proposto por Theiler et al. (1992), onde o método pode investigar se dados podem ser descritos completamente por variáveis aleatórias independentemente distribuídas (i.d.).

Freqüentemente, para sistemas experimentais observa-se dados com características periódicas fortes, e todas essas hipóteses lineares mostraram ser trivialmente falsas (Small et al., 2005b). Neste caso, é natural perguntar se há algum determinismo adicional ao sistema além da periodicidade. Com essa finalidade, Theiler (1995) propôs um outro tipo de hipótese: para termos determinísticos longos por embaralhamento dos ciclos (*Cycle-shuffle*) individuais dentro de uma série temporal. Mais recentemente, Small et al. (2001b) propôs um algoritmo melhorado que preserva ambas a estacionariedade e diferenciabilidade, não preservadas pelo algoritmo proposto por Theiler (1995), quando testado a uma hipótese semelhante.

Mais tarde, Small et al. (2001b) propôs o algoritmo **sub-rogado pseudo-periódico** (PPS - *pseudoperiodic surrogate*) com um outro ponto de vista.

A hipótese de Theiler (NH_{CS}) para sinais fortemente periódicos é bastante simples, mas poderosa. Theiler propõe que sub-rogados gerados pelo embaralhamento dos ciclos se direciona a hipótese que não há nenhuma correlação dinâmica entre os ciclos. Além disso, a técnica de embaralhamento dos ciclos é direcionada a uma hipótese, embora não-linear, ligeiramente diferente (semelhante a NH_0). Isto quer dizer que a dinâmica determinística existe dentro dos ciclos, mas não entre eles.

A hipótese NH_{PPS} é proposta por Small para sistemas dinâmicos contínuos que se identificam corretamente com a dependência estocástica linear entre os ciclos das órbitas pseudo-periódicas (quase-periódicas). A hipótese nula a ser testada é que a série temporal observada é consistente com um ruído (não correlacionado) governado por órbitas periódicas (ou pseudo-periódicas), ou seja, o

⁶Similar ao algoritmo IID.

conjunto de dados estacionário são pseudo-periódicos com componentes ruidosos que são (aproximadamente) identicamente distribuídos e não correlacionados para translações temporais suficientemente grandes.

Os algoritmos para gerarem estes dados sub-rogados são aqui denominados como **algoritmo CS** e **algoritmo PPS** correspondendo, respectivamente, a NH_{CS} e NH_{PPS} :

Algoritmo CS (*Cycle-shuffle*): A idéia é dividir todo o conjunto de dados em alguns segmentos (ciclos) e cada segmento contém exatamente um mesmo número inteiro de observações. Os dados sub-rogados são obtidos pelo embaralhamento aleatório desses segmentos que preservam a dinâmica intra-ciclo, mas destroem a dinâmica inter-ciclos pela embaralhamento aleatório da sequência temporal⁷ dos ciclos individualmente. A dificuldade em aplicar este algoritmo é que requer pré-conhecimento preciso na periodicidade (quebra dos ciclos), pois o embaralhamento de supostos ciclos poderiam conduzir a resultados espúrios, ou seja, indesejados (Small et al., 2001b).

Algoritmo PPS: Este método pode ser aplicado para testes contra a hipótese nula de uma órbita periódica com ruído não correlacionado para um número muito grande de sistemas experimentais que exibam comportamentos pseudo-periódicos. A principal idéia desses dados sub-rogados é gerar dados que exibam as mesmas estruturas periódicas que a série temporal original.

2.3 Estatística de Teste

Uma estatística de teste satisfatória deve ser selecionada para comparar a série temporal original com seus dados sub-rogados. Uma estatística de teste útil deve medir um parâmetro invariante não-trivial de um sistema dinâmico que seja dependente do modo como os sub-rogados são gerados.

O princípio do método de hipótese estatística com dados sub-rogados consiste em gerar um conjunto de dados sub-rogados que preservam algumas características estatísticas da série temporal original e destroem outras, tais como, a distribuição (média e variância). O espectro de potência (funções de autocorrelações) ou apenas as estruturas globais ou locais das séries temporais (Small et al., 2005b). Além disso, tais dados sub-rogados devem ser consistentes com a hipótese nula (H_0) de interesse. Também é importante que os dados sub-rogados

⁷Algoritmo 0.

sejam suficientemente “semelhante” para com a série temporal original. Finalmente aplica-se algum fator estatístico discriminante (estatística de teste) a série temporal original e também em cada uma das séries temporais sub-rogadas geradas através da série temporal original. Se o valor do fator discriminante da série temporal original for significativamente diferente do obtido para o conjunto de dados sub-rogados, então pode-se rejeitar H_0 .

Definição 2.3.1 (Estatística de Teste) *Valor calculado com base em dados observados e utilizado para testar uma hipótese nula. A estatística de teste é construída de forma a corresponder a uma distribuição conhecida se a hipótese nula é verdadeira; assim, a H_0 é rejeitada se não parecer plausível que o valor observado da estatística de teste prevenha daquela distribuição Callegari-Jacques (2006).*

Theiler (1995) sugere que haja dois tipos, fundamentalmente diferentes, de estatísticas de testes: **pivotal** e **não-pivotal**.

Definição 2.3.2 *Uma estatística de teste Q é pivotal se a densidade de probabilidade $P_{Q,F}$ é a mesma para todos os processos F consistentes com a hipótese nula; caso contrário é não-pivotal.*⁸

Uma vantagem notável da estatística de teste que é pivotal, como pode ser visto da definição, é que essa sempre obterá a mesma distribuição estatística $P_Q(t)$, que é independente do processo F escolhido do conjunto F_ϕ consistente com H_0 . Se a estatística de teste for não-pivotal, então não há nenhuma garantia que $P_{Q,F_i}(t) = P_{Q,F_j}(t)$ (com $i \neq j$) aconteça para arbitrários processos F_i e F_j em F .

Como o conhecimento exato da distribuição estatística não é frequentemente disponível, recorre-se a certo critério distintivo para ajudar na tomada da decisão e determinar o nível de confiança correspondente. Os critérios distintivos populares na literatura incluem duas classes: paramétricas e não-paramétricas.

2.3.1 Métodos Paramétricos

O critério paramétrico assume que a estatística de teste segue uma distribuição Gaussiana ou Normal, e os parâmetros da distribuição, isto é, a média e a variância, são estimados das amostras finitas. Pode-se rejeitar H_0 examinando se a

⁸Dado uma hipótese nula composta ϕ e um processo F consistente com ϕ , onde denota-se a estatística de teste escolhida por Q e a função de distribuição de probabilidade correspondente (p.d.f.) sob a hipótese nula de $P_{Q,F}(t) \equiv Prob(Q < q | F \in F_\phi)$. Se para qualquer processo F_i e F_j (com $i \neq j$) do conjunto F , tem-se que $P_{Q,F_i}(t) = P_{Q,F_j}(t)$, então a estatística Q é dito **pivotal**; caso contrário é **não-pivotal** (Theiler e Prichard (1995); Small et al. (2006a)).

estatística da série temporal original segue a distribuição estatística dos dados sub-rogados, enquanto o nível de confiança (α) correspondente para conclusão pode ser calculado da distribuição estatística estimada.

O **discriminante estatístico** “Q” (estatística de teste) é uma estimativa numérica de uma característica dos dados originais e sua variação é tal que nos permite decidir se a série temporal original é consistente com H_0 ou não. Quando a distribuição de Q sob H_0 for analiticamente conhecida, a região de rejeição estará nas extremidades (caudas) da distribuição de acordo com um determinado nível de significância como, por exemplo, mostrado na Figura 2.2.

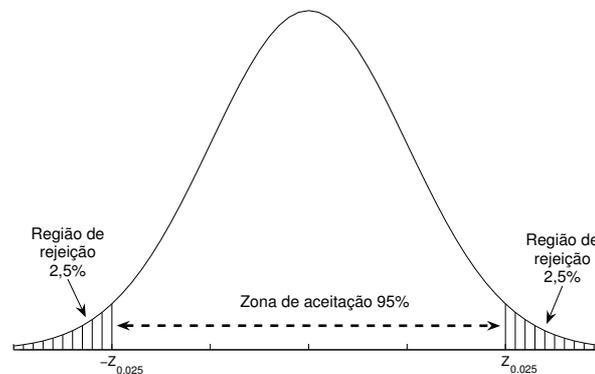


Figura 2.2: Região crítica para H_0

Caso contrário, a distribuição é formada numericamente em valores Q_1, \dots, Q_{N_s} da estatística computada em um conjunto de N_s dados sub-rogados consistente com H_0 .

O método de dados sub-rogados é determinado usando um valor quantitativo S, contanto que a distribuição Q_1, \dots, Q_{N_s} são distribuídos aproximadamente Normal, definido como

$$S = \frac{|Q_0 - \langle Q \rangle|}{\sigma_Q} \quad (2.3)$$

em que $\langle Q \rangle$ e σ_Q são, respectivamente, a média e o desvio padrão estimados do conjunto Q_1, \dots, Q_{N_s} . O nível de significância α defini os valores críticos em que o valor S rejeita ou não H_0 .

Este teste confere, quantitativamente, se Q_0 é incluído nas extremidades da distribuição empírica do conjunto Q_i ou não. Se Q_0 estiver entre a extremidade maior $(N_s + 1)\sigma/2$ e o menor $(N_s + 1)\sigma/2$ do conjunto de Q_i , a hipótese nula

é rejeitada. Uma boa aproximação prática (95%) para rejeição da H_0 utiliza um desvio padrão de 2σ

Seja como for, o valor numérico calculado pelo teste deve ser confrontado com valores críticos, que constam em tabelas apropriadas a cada teste, tabela da distribuição T. Essas tabelas da distribuição T geralmente associam dois parâmetros, que permitem localizar o valor crítico tabelado: nível de probabilidade ou nível de significância, usualmente 5% ($\alpha = 0,05$) ou 1% ($\alpha = 0,01$), e o número de graus de liberdade das amostras comparadas (geralmente igual ao tamanho da amostra N).

Assumindo S como um distribuição Normal $N(0, 1)$ ao nível de 5% os valores críticos são obtidos da tabela da distribuição t-student, como mostrada na Tabela 2.2.

Tabela 2.2: Intervalo de confiança para teste

Intervalo de confiança	Decisão
$-1,96 \leq S \leq 1,96$	falha em rejeitar H_0
$S \leq -1,96$ ou $S \geq 1,96$	rejeita-se H_0

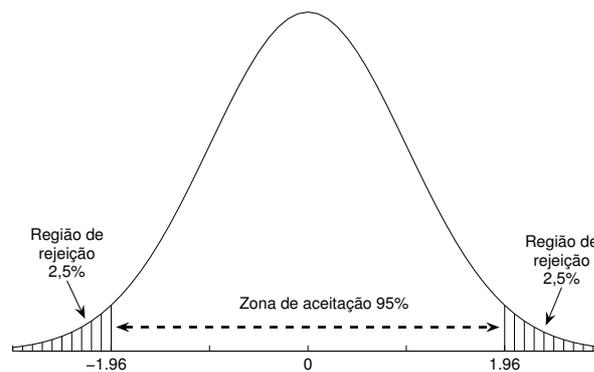


Figura 2.3: Intervalo de confiança para teste de S com $\alpha = 0,05$

A **região crítica**, Figura 2.3.1, para rejeitar H_0 encontra-se nas extremidades $S < -1,96$ e $S > 1,96$, com nível de significância $\alpha = 0,05$ (5%) e a **região de**

aceitação de H_0 está entre $1,96 \leq S \leq 1,96$, onde estaria configurado o que se chama de não-significância estatística, ou de aceitação da hipótese nula H_0 , ou de nulidade. Para $\alpha = 0,01$ (1%) a região de rejeição é $S < -2,58$ e $S > 2,58$.

Os testes estatísticos paramétricos propostos foram empregados a princípio para dados de processos com distribuições normais o que não é estritamente aplicado nos sistemas em geral, sendo este, um dos problemas para aplicação desses tipos de testes (paramétricos) para vários processos reais. Algumas estatísticas de testes paramétricas Q são comumente empregadas na literatura e serão discutidas a seguir.

Momentos Centrados dos dados

É a estatística linear padrão, particularmente no caso das hipóteses lineares padrão. Deixe m_n denotar o n -ésimo momento centrado do dado $X = \{x_i\}_{i=1}^N$.

$$m_n = \langle (X - \langle X \rangle)^n \rangle = \langle (X - \mu)^n \rangle, \quad (2.4)$$

em que μ denota a média aritmética populacional ou amostral. Em particular m_2 ou σ^2 é a variância, γ a assimetria e κ a kurtose que são definidas por:

$$\sigma^2 = \langle (x - \mu)^2 \rangle = m_2, \quad (2.5)$$

$$\gamma = \frac{\langle (x - \mu)^3 \rangle}{\sigma^3} = \frac{m_3}{m_2 \sqrt{m_2}}, \quad (2.6)$$

$$\kappa = \frac{\langle (x - \mu)^4 \rangle}{\sigma^4} - 3 = \frac{m_4}{m_2^2} - 3. \quad (2.7)$$

Cada uma dessas estatísticas medem quantidades relacionadas a distribuição dos dados; a assimetria é o grau de assimetria na distribuição; a kurtose mede quão “gorda” é a distribuição, ou convencionalmente, quão “pontiagudo” é a distribuição.

Outras estatísticas tais como reversão assimétrica temporal Q_{rev} (Timmer, 1998), o sexto cumulante padronizado C_6 (Mendes et al., 2002) e três pontos da autocorrelação Q_{TPA} (Kugiumtzis, 1999b) podem ser empregadas. A formulação dessas

estatísticas de testes é relatada a seguir:

$$Q_{rev} = \frac{\langle (x_i - x_{i-\tau})^3 \rangle}{\langle (x_i - x_{i-\tau})^2 \rangle} \quad (2.8)$$

$$C_6 = \frac{\langle (x - \mu)^6 \rangle}{\sigma^4} - 15\kappa - 10\gamma - 15 \quad (2.9)$$

$$Q_{TPA} = \frac{\langle (x_i - \mu)(x_{i-\tau} - \mu)(x_{i-2\tau} - \mu) \rangle}{\langle (x_i - \mu)^3 \rangle} \quad (2.10)$$

em que τ é o tempo de atraso desejado.

Todas estas estatísticas são funções do grau da distribuição dos dados.

2.3.2 Métodos Não-Paramétricos

Nos testes não-paramétricos, supõe-se que a distribuição de seus dados experimentais não seja normal, ou que não tenha elementos suficientes para poder afirmar que seja.

O critério estatístico não-paramétrico examina então a ordem (rank) da sequência dos valores estatísticos da série temporal original e seus dados sub-rogados gerados. Supõe-se que a estatística da série temporal original é Q_0 e que os valores de sub-rogados são $\{Q_i\}_{i=1}^{N_s}$ determinadas por N_s dados sub-rogados. Então, se o fator estatístico de teste da série temporal original e os sub-rogados seguirem a mesma distribuição, a probabilidade será de $1/(N_s + 1)$ para Q_0 seja menor ou maior dentre todos os valores $\{Q_1, \dots, Q_{N_s}\}$. Assim, quando encontramos Q_0 menor ou maior que a sequência de valores das estatísticas de testes $\{Q_i\}_{i=1}^{N_s}$, é muito provável que Q_0 siga uma distribuição diferente do conjunto dos dados sub-rogados, isso só é possível se N_s for grande.

Conseqüentemente o critério rejeita a hipótese nula sempre que a estatística original Q_0 for menor ou maior que $\{Q_1, \dots, Q_{N_s}\}$. A falsa taxa de rejeição ou a probabilidade em rejeitar a H_0 é considerada como $1/(N_s + 1)$ para testes unilaterais e $2/(N_s + 1)$ para testes bilaterais.

A seguir são descritas algumas testes utilizadas nos métodos não-paramétricos.

Gráfico Quantil-Quantil

Os gráficos Quantis-quantis são usados para determinar se duas amostras vêm da mesma família de distribuição. São gráficos espalhados de quantis computados de cada amostra, com uma linha desenhada entre o primeiro e terceiro quantil. Se

as quedas de dados próximos da linha forem pouco destorcidas (ou seja, retas), assumi-se que as duas amostras vêm da mesma distribuição. O método é robusto com respeito a mudanças locais e escalares de qualquer distribuição.

Essa relação linear aproximada sugere que as duas amostras possam vir da mesma família de distribuição. Portanto, um gráfico quantil-quantil linear é frequentemente suficiente para tal afirmação, com respeito apenas a distribuição dos dados.

Função de Autocorrelação (FAC)

A função de correlação entre dois sinais é uma medida da dependência temporal entre eles, isto é, uma estimativa da correlação linear dos dados. Se tais sinais forem independentes diz-se que os sinais são não correlacionados, ou em outras palavras, é uma medida de quão relacionados (no tempo) estão os sinais. A função de autocorrelação (FAC) é definida como,

$$r_{xx}(\tau, t) = E[x(t)x^*(t + \tau)] \quad (2.11)$$

se $x(t)$ for considerado real $x^*(t + \tau) = x(t + \tau)$, ergódico⁹.

No caso discreto, a definição da equação anterior torna-se

$$r_{xx}(k) = \lim_{N \rightarrow \infty} \frac{1}{2N + 1} \sum_{i=-N}^N x(i)x(i + k). \quad (2.12)$$

Como aplicar o FAC como teste estatístico de H_0 : a estatística de teste é feita construindo o gráfico da função de autocorrelação da série temporal original e dos N_s dados sub-rogados gerados. Se a função de auto-correlação da série temporal original cair dentro do conjunto da estatística de teste (FAC) dos dados sub-rogados gerados não se pode rejeitar H_0 . Caso algum ponto saia do conjunto da estatística de teste a hipótese nula pode ser rejeitada. Quando a estatística da série temporal original cair dentro da distribuição do conjunto de dados sub-rogados, considera-se que a série temporal original e os sub-rogados podem vir da mesma população.

⁹A esperança matemática pode ser substituída pela média temporal.

Informação Mútua Média (IMM)

Diferentemente da função de autocorrelação, a informação mútua considera também interdependências não-lineares. A informação mútua captura informações de momentos superiores, diferente da correlação linear que somente captura informações de segunda ordem. Portanto, a teoria da informação provê uma medida para a dependência não-linear dentro e entre séries temporais. Quando uma sucessão de medidas de uma variável é tomada durante um certo tempo, pode-se calcular a incerteza na predição da próxima medida dada as medidas precedentes.

Pode-se defini-la da seguinte forma: dada uma série temporal $\{x(t)\}_{t=1}^N$, a média mútua sobre uma medida $x_{t+\tau}$ dada uma medida x_t no tempo $t, \forall t$, é a informação mútua média $I(\tau)$,

$$I(\tau) = \frac{1}{N} \sum_{t=1}^N p(x_t, x_{t+\tau}) \log_2 \left[\frac{p(x_t, x_{t+\tau})}{p(x_t)p(x_{t+\tau})} \right] \quad (2.13)$$

em que τ é o atraso, $p(x_t)$ é a probabilidade de observar x_t na série temporal e $p(x_t, x_{t+\tau})$ é a probabilidade conjunta de observar x_t e $x_{t+\tau}$ na série temporal.

A função de informação mútua aplicada dentro de uma única série temporal é muito similar à função de auto-correlação que é uma medida do grau de dependência. A FAC também requer que as distribuições de x_t e $x_{t+\tau}$ sejam juntamente normais para que tenha um cálculo preciso de dependência, sendo que a função IMM não requer tal suposição. A vantagem da FAC é que pode ser calculada rapidamente e descreve muito bem um sistema linear. A desvantagem é que a FAC sempre assume que o processo subjacente é linear e calcula um valor para esta dependência; se o processo subjacente for não-linear, o valor poderia estar incorreto. A IMM é muito utilizada para análises não-lineares. Pode-se dizer que a IMM é uma versão em geral não-linear da FAC em uma série temporal Small e Nakamura (2006b).

O método testa e confere, conforme visto na FAC, se o calculo estatístico IMM da série temporal original cai dentro ou fora da distribuição estatística IMM dos dados sub-rogados. Quando a estatística da série temporal original cair dentro da distribuição do conjunto de dados sub-rogados, considera-se que a série temporal original e os sub-rogados podem vir da mesma população e então H_0 não pode ser rejeitada. Caso contrário, se cair fora da distribuição estatística dos dados sub-rogados a H_0 pode ser rejeitada.

Dimensão de correlação (d_c)

A dimensão de correlação (d_c) é uma medida estatística para avaliar a auto-similaridade da geometria de um conjunto de pontos no espaço fase Fielder-Ferrara e Prado (1994). Esta quantidade define o número de variáveis independentes que seriam necessárias para descobrir a dimensão de um sistema, isto é, os graus de liberdade do mesmo.

A dimensão de correlação d_c é um dos primeiros parâmetros utilizados para descrever atratores. Basicamente, fornece uma medida da complexidade do sistema em relação ao número de graus de liberdade do mesmo. Considerando o fato que d_c converge a um valor finito, no caso de sistemas determinísticos, e que não converge no caso de sistemas estocásticos, pode ser útil para avaliar a natureza determinística ou estocástica de um sistema. Grassberger e Procaccia (1983) propuseram um método para definir d_c de maneira experimental.

Usando uma série temporal Y_t são construídos N pontos num espaço fase com d_e dimensões (d_e é chamada de dimensão de imersão) de acordo com o teorema de Takens (Takens, 1981), onde $\{z_t\}_{t=1}^N$ é o conjunto de vetores d_e -dimensionais do espaço de fase \mathfrak{R}^{d_e} . A função de correlação $C_N(\epsilon)$ é definida por:

$$C_N(\epsilon) = \binom{N}{2}^{-1} \sum_{0 \leq i < j \leq N} \Theta(\|z_i - z_j\| < \epsilon) \quad (2.14)$$

onde $\Theta(X)$ é uma função de Heaviside cujo valor é 1 quando a condição X é satisfeita ($X > \epsilon$) e 0 caso contrário ($X < \epsilon$), e $\|\cdot\|$ é a função da distância em \mathfrak{R}^{d_e} . A soma $\sum_i \Theta(\|z_i - z_j\| < \epsilon)$ é o número de pontos dentro de uma distância ϵ de z_j . Espera-se que

$$C_N(\epsilon) \propto \epsilon^{d_c} \quad (2.15)$$

em que d_c é a dimensão do objeto. É natural definir a dimensão de correlação como

$$d_c = \lim_{\epsilon \rightarrow 0} \lim_{N \rightarrow \infty} \frac{\log(C_N(\epsilon))}{\log(\epsilon)} \quad (2.16)$$

A estimativa da dimensão de correlação utilizada é descrita por Judd (1992, 1994). Judd mostra que, para alguma constante ϵ_0 pequena, tem-se para todo $\epsilon < \epsilon_0$,

$$C_N(\epsilon) \propto \epsilon^{d_c} p(\epsilon) \quad (2.17)$$

onde $d_c(\epsilon_0)$ é a dimensão de correlação do objeto e $p(\epsilon)$ é um polinômio de ordem t , dimensão topológica. O algoritmo empregado estima a dimensão de correlação

d_c como uma função da observação escalar ϵ_0 , e conseqüentemente os resultados apresentados neste trabalho são curvas de $d_c(\epsilon)$. Ao aplicar a dimensão de correlação (d_c), precisa-se determinar uma dimensão de imersão d_e e o atraso de imersão τ . Devem ser selecionados valores ótimos de d_e e τ para o teste dos dados e então emprega-se também estes mesmos valores (d_e e τ) nos dados sub-rogados, assim é selecionada efetivamente uma estatística de teste. Selecionou-se τ como o primeiro zero do autocorrelação dos dados (o mesmo para os dados e sub-rogados).

2.4 Quantidade de séries de dados sub-rogados a serem geradas

Para um teste unilateral, normalmente usa-se o seguinte número de dados sub-rogados para o teste de hipótese:

$$N_s = \frac{1}{\alpha} - 1 \quad (2.18)$$

em que N_s é o número de séries sub-rogadas geradas e α o nível de significância, geralmente $\alpha = 0,05$. Para um teste bilateral,

$$N_s = \frac{2}{\alpha} - 1 \quad (2.19)$$

Portanto, para uma exigência mínima S de 95%, precisa-se pelo menos de 19 ou 39 séries sub-rogadas para um teste unilateral ou bilateral, respectivamente. Para uma exigência mínima de significância de 99%, precisa-se pelo menos de 99 séries temporais sub-rogadas para o teste.

O nível de significância não é fixo neste trabalho, ou seja, serem utilizados $\alpha = 0,05$ ou $\alpha = 0,01$ nas estatísticas de testes.

No capítulo seguinte será tratada a metodologia para gerar a série temporal sub-rogada com diferentes técnicas, conforme as hipóteses nulas tratadas no Capítulo 3.

Métodos de Dados Sub-rogados

3.1 Introdução

No Capítulo 2 foi visto como é formulada a hipótese nula¹ e algumas estatísticas de testes aplicadas para rejeitar ou não a H_0 de interesse. Neste Capítulo 3, as técnicas (algoritmos) para geração os dados sub-rogados serão tratados. Portanto, os conceitos dos métodos de dados sub-rogados serão sucintamente vistos, bem como, os métodos (algoritmos e técnicas) para gerar dados sub-rogados.

O método de **dados sub-rogados** (do inglês *surrogates data*) é utilizado para verificar determinadas características em dados gerados por processos estocásticos ou determinísticos. Isto é feito por intermédio de testes de hipótese, visto no Capítulo 2. A idéia principal é gerar um conjunto de dados, denominados sub-rogados “s(k)”, a partir da série temporal original, que compartilham determinadas características (distribuição de amplitude e o espectro de potência) a qual se deseja analisar. Tendo em mente qual característica deve ser investigada pode-se definir qual o método que deve ser usado para gerar os dados sub-rogados.

As principais técnicas, encontradas na literatura, para gerarem séries temporais sub-rogadas para análise estatística de processos são: algoritmo 0 - embaralhamento temporal dos dados (Theiler et al., 1992; Kugiumtzis, 2002c), algoritmo 1 - aplicação da Transformada de Fourier (FT) (Theiler e Prichard, 1994; Schreiber, 1997), algoritmo Transformada de Fourier com Ajuste de Amplitude (AAFT) (Theiler et al., 1992; Kugiumtzis, 1999c, 2000b).

A partir destes três foram propostos outras técnicas, tais como iterativo AAFT (IAAFT) (Schreiber e Schmitz, 2000, 1996) e outros tais como o WFT (Suzuki et al., 2005), o qual utiliza o mesmo procedimento adotado pelo FT, porém, faz-se o janelamento dos dados antes do cálculo da FT, para contornar o problema de altas frequências espúrias. Este janelamento também pode ser adotado na técnica AAFT originando o AAWFT. A técnica *correct* AAFT (CAAFT) (Kugiumtzis,

¹Hipóteses nulas: NH_0 , NH_1 , NH_2 , NH_{SS} , NH_{CS} e NH_{PPS} . Veja Capítulo 2.

2000a) adota o mesmo procedimento do AAFT, mas preserva a autocorrelação real dos dados originais. O *statically transformed autoregressive process* (STAP) (Kugiumtzis, 2002b) é baseado na idéia proposta pelo CAAFT.

Para gerar o dado sub-rogado por meio do embaralhamento temporal da série original, foram usadas três técnicas: o GAUSS, cuja H_0 estabelece que os dados são ruído i.i.d. com distribuição gaussiana; a técnica IID, que testa a H_0 que os dados são ruído i.i.d. e o BOOT em que embaralhamento é feito usando o método *bootstrap*, ou seja, o embaralhamento temporal da série original é feito com reposição.

Um outro algoritmo, recentemente proposto por Small e Nakamura (2005), é o *Small-Shuffle* (SS) que é similar ao algoritmo 0, mas cuja H_0 formulada é que os dados são variáveis aleatórias independentemente distribuídas (i.d.). Theiler (1995) também propôs um algoritmo de embaralhamento de ciclos (*cycle-shuffle*), para séries temporais pseudo-periódicas. Mais tarde, Small et al. (2005a) propôs o algoritmo sub-rogado pseudo-periódico (PPS ou *pseudoperiodic surrogate*) para séries pseudo-periódicas.

A Tabela 3.1 ilustra a distribuição dos algoritmos com suas respectivas técnicas para geração de dados sub-rogados, dando uma visão geral das técnicas a serem tratadas neste capítulo.

Tabela 3.1: Distribuição dos algoritmos com suas respectivas técnicas para geração de dados sub-rogados

Algoritmo	Técnica
Algoritmo 0	IID, GAUSS e BOOT
Algoritmo 1	FT e WFT
Algoritmo 2	AAFT, AAWFT, IAAFT, CAAFT e STAP
Algoritmo SS	Small Shuffle
Algoritmo CS	Embaralhamento dos ciclos
Algoritmo PPS	Sub-rogado pseudo-periódico

3.2 Embaralhamento Temporal (Shuffle): IID, GAUSS e BOOT - Algoritmo 0

Dada uma série temporal $x_k = \{x(k)\}_{k=1}^N$ os dados sub-rogados são gerados através do embaralhamento aleatório da ordem temporal da série temporal original. Este é o Algoritmo 0. Dessa forma, a correlação linear e não-linear da série temporal original é destruída e a distribuição de probabilidade ou amplitude é preservada. A hipótese nula, neste caso, é que a série original é um ruído aleatório não-correlacionado ou independente e identicamente distribuído (i.i.d.). Sendo possível testar, a partir dessas técnicas, se há indício de alguma dinâmica determinística nos dados originais. Baseado na proposta de Theiler et al. (1992) de embaralhamento temporal da série foram desenvolvidas três técnicas para gerar o algoritmo 0: IID, GAUSS e BOOT.

Na técnica IID proposta por Theiler et al. (1992), o embaralhamento ocorre sem reposição. Uma outra versão o GAUSS, a H_0 estabelece que os dados são ruído i.i.d. com distribuição gaussiana. Nessa técnica apenas os parâmetros média e variância da distribuição original são preservadas, ou seja, gera-se dados sub-rogados gaussianos com a mesma média e desvio padrão que a série temporal original. No BOOT o embaralhamento é feito usando o método *bootstrap*², ou seja, o embaralhamento temporal da série original é feito **com reposição**, diferentemente do proposto em Theiler et al. (1992). Os métodos IID e GAUSS são muito similares ao método de bootstrap (Efron, 1986), contudo eles são mais preferíveis, porque preservam a distribuição dos dados explicitamente. Assintoticamente, os métodos são equivalentes. A Tabela 3.2 mostra resumidamente as hipóteses (similares) para cada técnica com o algoritmo 0.

Tabela 3.2: Hipóteses Nulas NH_0 para o algoritmo 0.

Técnica	Hipótese Nula (NH_0)
IID	A série original é um ruído i.i.d.
GAUSS	A série original é um ruído i.i.d. com distribuição gaussiana.
BOOT	A série original é um ruído i.i.d.

²O método bootstrap (Efron, 1986; Efron e Tibshirani, 1993; Zoubir e Boashash, 1998) é um método de reamostragem para determinar regiões críticas ou intervalos de confiança que não podem ser aplicados para análise de séries temporais devido à dependência seqüencial no processo estocástico em análise.

3.3 Small-Shuffle - Algoritmo SS

Este método, recentemente desenvolvido por Small e Nakamura (2005) é descrito para testar flutuações irregulares. Seja o conjunto de dados da série temporal original $x(k)$, em que $i(k)$ corresponde ao índice de $x(k)$ ³, seja $g(k)$ um conjunto de números aleatório e $s(k)$ a série sub-rogada gerada. Os passos do algoritmo são descritos a seguir:

1. Obtenha $i'(k) = i(k) + A * g(k)$, em que A é o valor escalar da amplitude desejada, logo, têm-se números aleatórios gaussianos $g(k)$ acrescentados aos índices da série temporal original. Note que o índice $i(k)$ será uma seqüência de inteiros enquanto que a seqüência perturbada $i'(k)$ não.
2. Ordene de forma ascendente $i'(k)$ (função sort do Matlab) e faça o índice de $i'(k)$ ser $\hat{i}(k)$ (grau ordenado do índice perturbado, gerando assim um índice ligeiramente perturbado dos dados originais)
3. Obtenha o dado sub-rogado $s(k) = x(\hat{i}(k))$ (reordene a série temporal original com o índice perturbado $\hat{i}(k)$).

Tabela 3.3: Simples exemplo do método sub-rogado SS (Small e Nakamura (2005))

k	$i(k)$	$x(k)$	$i'(k)$	Ordenar $i'(k)$	$\hat{i}(k)$	$s(k) = (x(\hat{i}(k)))$
1	1	13	0.13	-1.33	2	12(= $x(2)$)
2	2	12	-1.33	0.13	1	13(= $x(1)$)
3	3	14	3.25	2.71	5	15(= $x(5)$)
4	4	11	4.58	3.25	3	14(= $x(3)$)
5	5	15	2.71	4.58	4	11(= $x(4)$)

Os índices (ordem) dos dados têm implicações importantes, independente se a série temporal é linear ou não-linear. Conseqüentemente, sempre que o índice do dado muda ou embaralhado, o fluxo de informação também muda e a resultante da série temporal não reflete a dinâmica original. O método *Small-Shuffle* (SS) usa esta idéia. Este método pode investigar se os dados amostrados, geralmente flutuações irregulares, podem ser descritos completamente por variáveis aleatórias independentemente distribuídas (NH_{SS}), isto é, em outras palavras, uma investigação de que se os dados são variáveis simplesmente aleatórias ou têm correlação

³Isto é, $i(k) = k$, sendo assim $x(i(k)) = x(k)$.

temporal. Os dados sub-rogados SS conservam a mesma distribuição de probabilidade que a série temporal original. O índice de um determinado dado é embaralhado em uma pequena escala para não destruir estruturas locais e preservar o comportamento estrutural global em relação a série temporal original.

Conseqüentemente chama-se o método de Small-Shuffle (SS) devido à escolha desta amplitude (A) pequena, ou seja, realiza-se um embaralhamento curto. Segundo Small et al. (2005b) e Small e Nakamura (2005) $A = 1$ é o valor mais apropriado na maioria dos casos. O valor de A depende provavelmente das características dos dados, e podem ser justificados valores menores ou maiores em algumas situações.

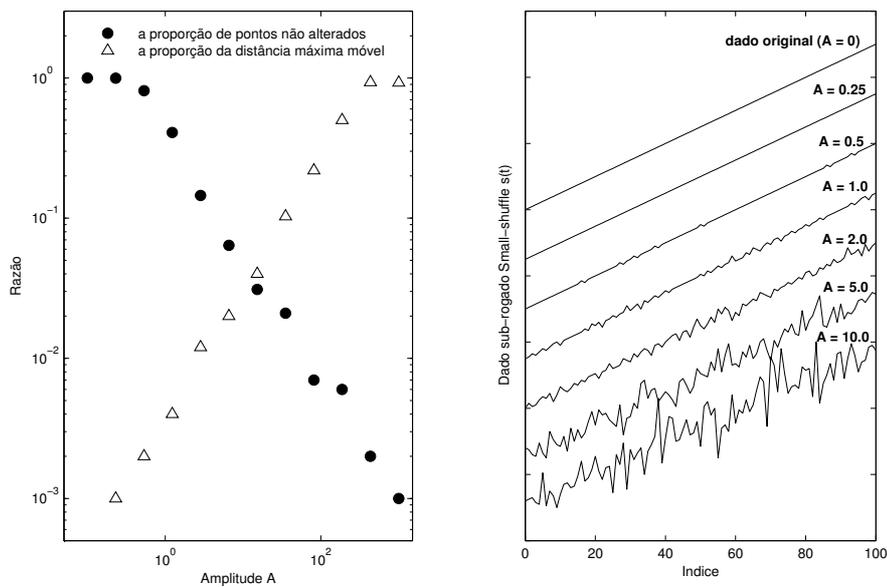


Figura 3.1: Relação da amplitude A de um número aleatório gaussiano e o índice. Painel a esquerda ilustra (como uma função da amplitude A): a proporção de pontos que não são alterados pelo algoritmo SS (\bullet); e, a distância máxima que um ponto qualquer da série temporal original é perturbado no sub-rogado (Δ , expressa como uma fração do comprimento dos dados). Painel a direita ilustra o efeito de valores diferentes de A . A série temporal original é gerada por meio $x(k) = k$, $1 \leq k \leq 100$. Se o sub-rogado SS e a série temporal original forem idênticos, então a curva deveria ser uma linha reta (como no caso de $A = 0$), Small e Nakamura (2005)

Na Figura 3.1 é mostrada a influência da amplitude A . O painel da esquerda mostra que quando A aumenta, o número de pontos de dados que não são alte-

rados diminui e a relação da distância máxima móvel⁴ aumenta. Para mostrar a influência da amplitude visualmente, compara-se os dados originais e os dados sub-rogados SS diretamente para diferentes amplitudes de A , onde os valores de A são 0.25, 0.5, 1.0, 2.0, 5.0 e 10.0. O painel da direita mostra que até A aproximadamente 2.0, o comportamento de $s(k)$ se mostra quase igual a série temporal original ($A = 0$), com o aumento de A , o comportamento de $s(k)$ se torna mais estocástico. Observa-se que embora espera-se $A = 1.0$, mais apropriado na maioria dos casos, o valor de A provavelmente depende das características dos dados, e podem ser justificados valores menores ou maiores em algumas situações.

O método sub-rogado SS pode indicar quais dados são simples v.a.s ou que têm correlações lineares. Contudo, o método sub-rogado SS não poderia discriminar entre sistemas dinâmicos lineares e não-lineares, logo este algoritmo é mais indicado para teste de determinismo.

3.4 FT - Algoritmo 1

Esta técnica FT (Algoritmo 1) proposta por Theiler et al. (1992) é baseado na hipótese nula de que o conjunto de dados da série temporal original vêm de um processo linear gaussiano. Na FT ou também chamado de fase aleatória (*phase randomized*), aplica-se a transformada discreta de Fourier (DFT) na série temporal original $X_k = \{x_k\}_{k=0}^{N-1}$, definida como

$$\tilde{x}_k = \frac{1}{N} \sum_{n=0}^{N-1} |X_n| e^{-j2\pi n \frac{k}{N}}, \quad (3.1)$$

em que $0 < k \leq N - 1$ e $\frac{k}{N} = f_k$ é a k -ésima frequência de Fourier.

Multiplicando as componentes \tilde{x}_k por fases aleatórias, obtém-se

$$\tilde{x}'_k = \tilde{x}_k e^{j\phi_k} \quad (3.2)$$

em que ϕ_k é uma variável aleatória uniformemente distribuída em $[0, 2\pi]$.

Para que a transformada discreta inversa de Fourier seja real (nenhum componente imaginário), deve-se ter necessariamente fases anti-simétricas tal que

$$\phi(f) = -\phi(-f) \quad (3.3)$$

⁴DMM é a distância máxima entre todas as distâncias da posição de um ponto na série temporal original para a posição deslocada no dado sub-rogado SS

As fases resultantes dos pares complexos conjugados são embaralhadas mantendo a amplitude espectral, de modo a garantir o mesmo espectro de potência da série original nas séries temporais sub-rogadas. Posteriormente, a seqüência gerada é apresentada no domínio do tempo aplicando a transformada discreta inversa de Fourier como mostrado a seguir

$$s_k = \frac{1}{N} \sum_{k=0}^{N-1} \tilde{x}'_k e^{j2\pi k \frac{n}{N}} \quad (3.4)$$

Esta técnica é denominada como fase aleatória (*phase randomization*) e diferentes realizações $\{\phi_k\}_{k=0}^{N-1}$ geram novos sub-rogados com as mesmas propriedades espectrais da série original $\{x_k\}_{k=0}^{N-1}$, veja Figura 3.2. Desta forma quaisquer outras estruturas não-lineares das séries são destruídas.

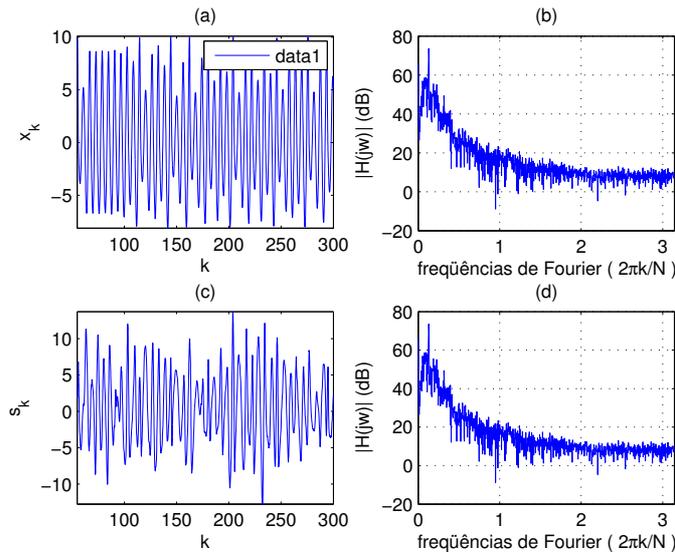


Figura 3.2: FT aplicado ao sistema de Rössler: (a) série temporal original, (b) espectro da série original (c) sub-rogado FT, e (d) espectro do sub-rogado FT.

A hipótese nula, neste caso, é que a série original é um ruído gaussiano linearmente autocorrelacionado (NH_1). Este resultado somente indicaria uma evidência estatística da linearidade em uma série temporal estacionária e gaussiana. Dessa forma, fica claro a dificuldade em se usar este algoritmo, uma vez que a rejeição da hipótese nula, pode estar associada a não admissão de que a série original seja linear ou que esta seja um processo gaussiano.

Este método para gerar dados sub-rogados não é adequado para séries com *distribuição não gaussiana*, uma vez que as séries artificiais assim geradas serão gaussianas e terão distribuições diferentes da série original.

3.5 AAFT - Algoritmo 2

O método FT, relatado na Seção 3.4, não é satisfatório quando a série temporal original não é normalmente distribuída, isto é, a FT do dado sub-rogado possui uma distribuição diferente da série temporal original. A técnica **Transformada de Fourier com Ajuste de Amplitude** (AAFT) ou *amplitude adjusted*, proposto por Theiler et al. (1992), foi desenvolvido para contornar as limitações do algoritmo FT.

Um modo para generalizar a hipótese nula (NH_1) anterior para casos onde os dados são não-gaussianos é supor que embora a dinâmica seja linear, a função de observação pode ser não-linear. Em particular, hipoteticamente já existiria uma série temporal subjacente " s_k ", consistente com a hipótese nula de um ruído linear gaussiano, cuja série temporal observada x_k é dada por

$$x_k = h(s_k) \quad (3.5)$$

em que s_k é conjunto da **série temporal normal** com $k = 1, \dots, N$.

A hipótese nula H_0 mais generalizada para o teste é que a série temporal original x_k é gerado por um processo gaussiano (normal) medido por uma transformação estática⁵ " h " e possivelmente não-linear. Para permitir a geração de dados sub-rogados, tem-se que assumir mais adiante (como parte da hipótese nula - NH_2) que a função de observação $h(\cdot)$ seja efetivamente inversível ($s_k = h^{-1}(x_k)$).

A técnica AAFT começa com a suposição importante que h é uma função monotônica, isto é, h^{-1} existe ($s_k = h^{-1}(x_k)$). A idéia é simular h^{-1} primeiro, reordenando os dados do **ruído branco** y_k gerados para terem a mesma estrutura como x_k .

A técnica para gerar este tipo de dados sub-rogados, considerando a série temporal original como sendo $X_k = \{x(k)\}_{k=1}^N$, é descrita da seguinte forma:

1. A série temporal original é re-escalado para uma distribuição normal (\tilde{Y}_k). Isto é feito gerando um conjunto de números aleatórios gaussianos $Y_k =$

⁵Desde que x_k , dependa somente do valor atual de s_k , e não em valores derivados ou passados, o filtro $h(\cdot)$ é dito ser "estático" ou "instantâneo".

$\{y(k)\}_{k=1}^N$. Cada elemento é gerado independentemente a partir de um gerador de números pseudo-aleatórios gaussianos. Posteriormente é ordenado \tilde{Y}_k para que tenha o mesmo grau de distribuição que X_k .

2. Calcula-se a FT desta nova sequência submetendo-a a um processo de embaralhamento de fases, como no algoritmo FT, gerando \tilde{Y}_k^{FT} .
3. A sequência final é gerada re-escalando à última sequência \tilde{Y}_k^{FT} , isto é, ordenando a série original X_k para criar um sub-rogado s_k que tenha o mesmo grau de distribuição que \tilde{Y}_k^{FT} .

A habilidade desta técnica para preservar corretamente o espectro de potência depende da existência de uma transformação estática da distribuição original. Se a distribuição de dados incluir singularidades ou pontuais transições, então esta técnica pode produzir resultados imprevisíveis ou não condizentes. Como a distribuição da série temporal original é, em geral, desconhecida, a transformação deve ser determinada empiricamente. Portanto, a re-escala no primeiro passo constitui uma transformação empírica da distribuição da série temporal original para uma realização finita específica de um processo de distribuição normal gaussiana $N(0, 1)$.

$$\tilde{y}(k) = \frac{y(k) - \mu_y}{\sigma_y} \sigma_x + \mu_x \quad (3.6)$$

em que $y(k)$ é o conjunto de números aleatórios gaussianos com média μ_y e desvio padrão σ_y ; e μ_x e σ_x são, respectivamente, a média e o desvio padrão da série temporal original.

Um conceito importante para essa aproximação é que para qualquer série temporal, pode-se construir uma série temporal normal que sob uma transformação monotônica possua a mesma auto-correlação e distribuição de amplitude da série temporal original. De acordo com a H_0 , um conjunto de dados sub-rogados ($Z = s_k^n$) tem que preservar a auto-correlação $r_x(\tau)$ da amostra da série temporal original, isto é, $r_x(\tau) = r_z(\tau)$ ($\tau = 1, \dots, \tau_{max}$), e a distribuição de amplitude da amostra, isto é, $F_x(x) = F_z(z)$, onde $F_x(x)$ é a amostra da função de densidade cumulativa (c.d.f.) de x_k .

A re-escala no terceiro passo resulta então numa realização independente, de uma diferente distribuição gaussiana no processo anterior, na distribuição da série temporal original. As transformações feitas antes e posteriormente não são exatamente inversas uma da outra, e isto resulta em mudanças do espectro de potência

do sub-rogado final. O resultado desta alteração é tentar “*embranquecer*”⁶ o espectro de potência da série temporal original. Claro que o “embranquecimento” que acontece depende do comprimento da série temporal (para séries temporais longas a distribuição será quase contínua, reduzindo o efeito) e do grau para o qual a distribuição original é não-gaussiana.

Portanto, os dados sub-rogados obtidos via AAFT conservam o espectro de potência e a distribuição da série temporal original. A H_0 testada, neste caso, é que os dados são gerados por uma aplicação de uma transformação estática monotônica não-linear de um processo gaussiano linear (NH_2).

Existem alguns problemas com esta técnica e alguns autores propõem soluções (Schreiber e Schmitz, 1996). Kugiumtzis (2000a) discute que o AAFT é adequado se a transformação estática em questão “não” seja não-linear, ou seja, a transformação “*h*” deve ser linear. O AAFT é indicado para identificação de processos lineares gaussianos.

3.6 WFT - Algoritmo 1 e AAWFT - Algoritmo 2

Um problema nos algoritmos é que para aplicar a transformada discreta de Fourier os dados tem que assumir periodicidade. Uma possível solução é limitar as fases aleatórias. Outra solução é a técnica de janelamento WFT e AAWFT (Theiler et al., 1992; Suzuki et al., 2005), onde é proposta uma mudança no algoritmo FT da seção 3.4. Durante o processo que gera dados sub-rogados calcula-se a transformada Fourier, este cálculo fica rápido e preciso quando se utiliza a transformada rápida de Fourier (FFT). Porém, existe uma possibilidade de piora da precisão dos dados gerados. Por exemplo, se a amostra da série temporal original for uma curva senoidal, este têm que possuir uma única componente de frequência cujo espectro de potência calculado possua também apenas um único pico. Todavia, se o tamanho das amostras da série temporal original não igualar (combinar), ao período completamente, várias frequências surgem sobre o a frequência verdadeira. Estas frequências que são chamadas de *frequências espúrias*⁷ afetam a estimação do espectro de potência. A razão é que no cálculo da FT, é assu-

⁶O termo “*embranquecer*” ou branco significa tornar-se igual ao espectro do ruído branco. Deve ser notado que esta nomenclatura diz respeito a propriedades estocásticas do sinal. Em outras palavras o fato do ruído ser branco indica que todas as frequências do sinal são igualmente importantes. Por outro lado, a qualificação branco nada diz com respeito às propriedades estatísticas do ruído, ou seja, a distribuição da amplitude que pode ter diversas distribuições como a gaussiana ou a uniforme.

⁷Frequências não desejadas.

mido que os dados originais são periódicos. Se os dados originais não possuem o mesmo período, como no instante de observação, frequências espúrias surgem e a precisão do cálculo do espectro de potência será conseqüentemente comprometida.

Na maioria dos casos de análise de séries temporais reais, raramente os dados originais possuem o mesmo período que no instante observado. Até mesmo quando se usa a reamostragem de vários períodos dos dados originais, para reduzir as frequências espúrias. Neste caso, pode ocorrer o comprometimento de informações essenciais que os dados originais poderiam conter.

Nesta seção será descrita então uma transformação dos dados originais através do janelamento destes (Theiler et al., 1992; Suzuki et al., 2005), reduzindo para zero ou quase zero as frequências espúrias. Logo, para o cálculo da transformada de Fourier com janelamento alguns tipos de janelamento de dados são propostos na literatura, (Theiler et al., 1992; Suzuki et al., 2005):

- O janelamento senoidal

$$w(k) = \text{sen}\left(\frac{\pi k}{N-1}\right) \quad (3.7)$$

- O janelamento Parzen

$$w_p(k) = 1 - \left| \frac{k - \frac{1}{2}(N+1)}{\frac{1}{2}(N-1)} \right| \quad (3.8)$$

- O janelamento Hanning

$$w_H(k) = \frac{1}{2} \left[1 - \cos\left(\frac{2\pi k}{N-1}\right) \right] \quad (3.9)$$

- O janelamento Welch

$$w_W(k) = 1 - \left[\frac{k - \frac{1}{2}(N+1)}{\frac{1}{2}(N-1)} \right]^2 \quad (3.10)$$

em que N é o tamanho da amostra de dados e $k = 0, \dots, N-1$.

A Figura 3.6 ilustra os tipos de janelamento propostos.

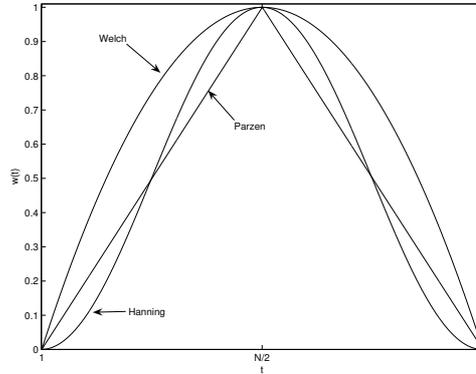


Figura 3.3: Tipos de janelamento de dados

Tomando a transformada discreta de Fourier $\tilde{X}(k)$ da série temporal original $x(k)$ com o janelamento $w(k)$. Tem-se

$$\begin{aligned}
 \tilde{X}(k+1) &= \sum_{n=0}^{N-1} x(k+1)w(k+1)e^{-j2\pi kk/N} \\
 \tilde{X}(k+1) &= \sum_{k=0}^{N-1} \left[\frac{1}{N} \sum_{k'=0}^{N-1} X(k'+1)e^{-j2\pi k'k/N} \right] w(k+1)e^{-j2\pi kk/N} \\
 \tilde{X}(k+1) &= \frac{1}{N} \sum_{k'=0}^{N-1} X(k'+1) \sum_{k=0}^{N-1} w(k+1)e^{-j2\pi(k-k')k/N} \\
 \tilde{X}(k+1) &= \frac{1}{N} \sum_{k'=0}^{N-1} X(k'+1)W(k-k'+1), \tag{3.11}
 \end{aligned}$$

em que $X(k+1)$ e $W(k)$ são a $(k+1)$ -ésima componente da transformada de Fourier de $x(k)$ e $w(k)$, respectivamente, em que $k = 0, 1, \dots, N-1$. A utilização do janelamento de dados implica em suavizar $X(k)$ com o peso $W(k)$. Na realidade estas amostras são multiplicadas por uma dessas funções de janelamento ou qualquer outra função de janelamento.

3.7 IAAFT - Algoritmo 2

Este método **Transformada de Fourier com iterativo ajuste de amplitude** (IAAFT) foi desenvolvido por Schreiber e Schmitz (1996, 2000) que se interessaram em preservar o espectro de potência, executando uma série de iterações

(repetições), com maior precisão que o AAFT, visto na Seção 3.5. Além disso, o IAAFT preservar, completamente, o histograma empírico da série temporal original. A técnica IAAFT é descrita a seguir:

1. Inicie com uma permutação aleatória da ordem temporal da série temporal original X_k , resultando r_n . Re-ordene o dado X_k encontrando X_{rank} , sinal de distribuição desejado. Calcule o espectro de potência (amplitudes de Fourier) da série original pela transformada de Fourier.
2. A transformada de Fourier é aplicada em r_n . Aqui o espectro de potência (amplitudes de Fourier) de r_n é substituído pelo da série temporal original calculada no primeiro passo; porém, a fase do espectro de r_n é preservada.
3. A transformada inversa de Fourier é aplicada ao dado obtido no passo 2, resultando em s_n . Embora a série temporal gerada exiba o mesmo espectro de potência que a série temporal original, seu histograma empírico (distribuição) é diferente da série temporal original. A série temporal chamada de dado sub-rogado IAAFT-1 tem o mesmo espectro de potência da série original.
4. Para preservar o histograma empírico da série temporal original, a sequência final s_n é re-escalada para obter o mesmo grau ordenado (mesma distribuição) da série temporal original (X_{rank}), resultando num novo r_n . A série temporal re-escalada é chamada de dado sub-rogado IAAFT-2 (mesma distribuição de probabilidade ou amplitude).
5. Se a discrepância entre o espectro de potência da série original e do dado sub-rogado não for suficientemente pequena, repete-se os passos anteriores, a partir do passo 2, substituindo o dado r_n no passo 2 pelo dado obtido no passo anterior. Repetindo várias vezes estes passos, o espectro de potência da série temporal original pode ser preservado com maior precisão que o método AAFT.

As suposições básicas deste método são que para cada iteração as amplitudes de Fourier são ajustadas para serem menores que na interação anterior, e que a alteração do espectro potência quando re-escalada será também menor que na interação anterior. Na realidade, Schreiber mostrou que, para um processo auto-regressivo não-linear (NARMA), o procedimento dessas interações convergir para o espectro de potência da série temporal original até que um ponto de saturação seja alcançado, onde o ajuste da amplitude de Fourier é tão pequena

que a re-escala posiciona os dados na ordem exata que tinham antes do ajuste da amplitude.

O sub-rogado IAAFT-2 tem exatamente a mesma distribuição de amplitude que o original, mas as discrepâncias no espectro de potência são prováveis em ocorrer. Por outro lado, procura-se melhorar o espectro de potência deixando as discrepâncias para a distribuição de amplitude, chamado de IAAFT-1.

A técnica IAAFT não faz nenhuma suposição para a forma da transformação h , isto é, pela construção, a técnica IAAFT pode representar a hipótese nula não obstante a formulação de h , quando a técnica AAFT não puder ser representada (quando h não é monotônico). Pode-se discutir que o desvio do espectro de potência devido a possível não-monotonicidade de h é pequena e não afeta os resultados do teste, mas isto está correto somente para os discriminantes estatísticos que não são sensíveis às correlações lineares dos dados, contudo a maioria dos métodos não-lineares, incluindo todos os modelos de predição não-linear, são sensíveis às correlações dos dados e, conseqüentemente, são supostos para ter o poder de distinguir correlações não-linear de linear.

Este é um teste usado na área de dinâmica não-linear para distinguir entre séries temporais que se originam de sistemas lineares e/ou de sistemas não-lineares. Como o interesse é na dinâmica não-linear e não na não-linearidade estática, a hipótese nula para o teste estatístico, neste caso, é que a série temporal se origina de um processo dinâmico linear (ARMA) que é modificado por um filtro estático não-linear.

3.8 CAAFT - Algoritmo 2

Uma série temporal sub-rogada corretamente projetada Z deve possuir a mesma autocorrelação que a série temporal original, $r_z(\tau) = r_x(\tau)$, para um atraso escalar τ , e de mesma distribuição de amplitude, $F_z(z_i) = F_x(x_i)$ ⁸, e sobretudo aleatória.

A técnica CAAFT ou **corrected AAFT** (Kugiumtzis, 2000a) corrige a autocorrelação⁹ do AAFT usando uma típica relação de aproximação. Uma série temporal é gerada por uma modelo AR (auto-regressivo) apropriado e é transformada por meio de uma “gaussianização inversa” para possuir uma c.d.f. exatamente como a série original e também suas correlações lineares. Esta aproximação envolve dois parâmetros livres, a ordem do modelo AR e o número de tentativas (k)

⁸ $F_x(x_i)$ é a função cumulativa da densidade (c.d.f.) de x_i

⁹A autocorrelação está intimamente relacionada com o espectro de potência, relatada no capítulo 2

necessárias para encontrar o “melhor” modelo (o parâmetro k é geralmente fixo e igual ao número de sub-rogados a serem gerados).

1. Faça os três passos do AAFT (visto na seção 3.5) gerando \tilde{Y}_k , \tilde{Y}_k^{FT} e X_{surr} , que são respectivamente, a série temporal original re-escalado para uma distribuição normal, o dado gerado a partir do algoritmo FT e sub-rogado gerado pelo AAFT.
2. Escolha τ_{max} e calcule as auto-correlações $r_{X_{surr}}(\tau)$, $r_x(\tau)$, $r_{\tilde{y}}(\tau)$, para $\tau = 1, \dots, \tau_{max}$.
3. Encontre uma interpolação linear simples $\tilde{\phi}$ de $r_{\tilde{y}}(\tau)$ como uma função de $r_{X_{surr}}(\tau)$.
4. Calcule $r_u(\tau)$ de $r_u(\tau) = \tilde{\phi}(r_x(\tau))$. A função $\tilde{\phi}$ é arbitrariamente definida como $r_u = r_x(\tau)[r_{\tilde{y}}(\tau^*)/r_{X_{surr}}(\tau^*)]$, onde τ^* é o menor atraso (lag) de $r_{X_{surr}}$.
5. Escolha uma ordem apropriada “ p ” para o modelo AR (por exemplo, $p = \tau_{max}$), e estime os coeficientes de AR(p) de r_u .
6. Gere u usando o modelo estabilizado de AR.
7. Transforme u em w ($w_i = F_x^{-1}(F_0(u_i)) = h(u_i)$, $i = 1, \dots, N$).
8. Repita os passos de 1 a 7 num total de k -vezes criando w^1, \dots, w^k .
9. Calcule $\{r_w^1, \dots, r_w^k\}$ e encontre um r_w^k mais próximo de r_x .
10. Utilize o modelo AR que corresponde ao k (melhor ou ótimo) selecionado no passo anterior e gere N_s dados sub-rogados repetindo os passos 6 e 7 (N_s vezes).

Conseqüentemente, o CAAFT é um modelo AR (reordenando no passo 7) cujo modelo foi combinado aproximadamente pela autocorrelação do dado observado. A razão deste método é encontrar um processo gaussiano satisfatório que sob uma transformação monotônica gere dados sub-rogados com aproximadamente a mesma autocorrelação que os dados originais.

Os dados sub-rogados gerados são como o AAFT, mas corrigido para terem a mesma autocorrelação que os dados originais. Para a correção, uma função de interpolação linear ($\tilde{\phi}$) é encontrada. Usando esta função de interpolação, para a autocorrelação da série temporal original, a autocorrelação da respectiva série

temporal gaussiana é estimada. Baseado nessa autocorrelação, os coeficientes do correspondente modelo AR de ordem p são estimados e uma série temporal AR é gerada, e transformada para combinar com a distribuição de amplitude da série temporal original. Este processo é feito k vezes para obter um modelo AR para candidato. O mais apropriado é selecionado, no sentido que a autocorrelação dos sub-rogados gerados seja a melhor aproximação da autocorrelação da série temporal original. Baseado neste modelo AR, as N_s realizações são geradas e transformadas para terem a mesma distribuição de amplitude que a série temporal original.

O sub-rogado CAAFT estima a correlação linear original sem tendência, mas normalmente com uma extensa variação do que os sub-rogados AAFT. Como resultado, a não-linearidade estática usando o CAAFT é frequentemente mais extensa (maior variação) que quando usado o AAFT (a menor é obtida com IAAFT). Isto faz com que o teste com o sub-rogado CAAFT geralmente seja mais conservador do que com os outros dois AAFT e IAAFT. É importante também observar que a custo computacional depende principalmente do tamanho de k . Portanto, o CAAFT possui maior tempo computacional do que o AAFT quando $k > N_s$.

3.9 STAP - Algoritmo 2

A característica chave para a execução bem sucedida do teste de dados sub-rogados em séries temporais é a geração de dados sub-rogados que representam exatamente a hipótese nula (processo estocástico normal transformado estaticamente - NH_2 do algoritmo 2). Uma aproximação conceitual é a técnica para a geração de dados sub-rogados proposta por Kugiumtzis (2002b), chamado **processo autoregressivo estaticamente transformado** ou *statically transformed autoregressive process* (STAP).

Os resultados errôneos são prováveis de ocorrer principalmente devido à ineficiência dos algoritmos para gerarem dados sub-rogados que preservem as correlações lineares originais. O algoritmo STAP proposto por Kugiumtzis (2002b) é desenvolvido em cima da idéia do algoritmo CAAFT. Formulando assim uma aproximação conceitual para a geração dos dados sub-rogados consistentes com NH_2 , que se resolve então analiticamente. O novo algoritmo, chamado STAP, gera dados sub-rogados com realizações de um apropriado processo auto-regressivo estaticamente transformado (STAP), ou seja, o processo sob a H_0 é projetado como uma transformação estática de um processo normal apropriado.

A idéia principal por trás do algoritmo STAP é que para qualquer processo estacionário $X = [X_1, X_2, \dots]$, com medidas escalares finitas \mathbf{x} , este terá um processo estocástico linear escalar Z com a mesma autocorrelação ρ_x e c.d.f. marginal Φ_x de acordo com o processo observado ($\rho_z(\tau) = \rho_x(\tau)$ e $\Phi_z(z_i) = \Phi_x(x_i)$), isto é, Z é uma “cópia linear” escalar do dado observado X . O objetivo é derivar Z por uma transformação monotônica estática g de um processo normal escalar U com uma autocorrelação ρ_u própria, ou seja, $Z_i = g(u_i)$. Assim g e ρ_u devem ser selecionados corretamente, de forma que Z tenha as mesmas propriedades desejadas da série temporal original. Na prática, o conjunto de dados sub-rogados é uma realização finita $z = \{z_i\}_{i=1}^n$ do processo Z , onde g e ρ_u são estimados baseados somente em \mathbf{x} . Note que g e U são, em geral, diferentes de h e s da H_0 vista no AAFT da seção 3.5 (estes são os mesmos se h for monotônico). Assim com esta aproximação a H_0 pode ser formulada mais generalizada, isto é, a série temporal é gerada por um processo estocástico linear (conforme NH_2).

Seja Φ_0 a c.d.f. marginal de um processo normal padrão U . Uma escolha satisfatória para g , de forma que $\Phi_z(z_i) = \Phi_x(x_i)$, está definida como

$$z_i = g(u_i) = \Phi_x^{-1}(\Phi_0(u_i)), \quad (3.12)$$

em que g é monotônico por construção. Assumindo que $\Phi_x(x_i)$ é contínuo e estritamente crescente e que $-1 < \rho_u < 1$, ambos são verdadeiros para todo propósito prático, há uma função ϕ que depende de g , tal que $\rho_z(\tau) = \phi(\rho_u(\tau))$ para qualquer atraso τ . Se g tiver uma forma analítica, então pode ser possível encontrar uma boa expressão analítica para ϕ . Neste caso, dado que ρ_x seja conhecido e fixando $\rho_z = \rho_x$, pode-se inverter ϕ para encontrar $\rho_u = \phi^{-1}(\rho_x)$, se ϕ^{-1} existir.

Em geral, a função g , como definido na equação (3.12), não tem uma forma analítica, porque Φ_x não é analiticamente conhecido, mas pode ser aproximado por uma função analítica, por exemplo, um polinômio p_m de grau m ,

$$z_i = g(u_i) \approx p_m(u_i) = a_0 + \sum_{j=1}^m a_j u_i^j. \quad (3.13)$$

Foram usados polinômios de baixo grau para aproximar tal transformação. Logo usando a definição para a autocorrelação, a expressão aproximada para ϕ é dada como,

$$\rho_x = \phi(r_u) = \frac{\sum_{s=1}^m \sum_{t=1}^m a_s a_t (\mu_{s,t} - \mu_s \mu_t)}{\sum_{s=1}^m \sum_{t=1}^m a_s a_t (\mu_{s+t} - \mu_s \mu_t)} \quad (3.14)$$

em que um arbitrário comprimento τ é incluído como argumento para as autocor-

relações, μ_s é o s -ésimo momento central de U_i sendo $\mu_{2k+1} = 0$, $\mu_{2k} = (2k - 1)!$, $k \geq 0$, e $\mu_{s,t}$ é o (s,t) -ésimo momento central conjunto da distribuição normal padrão bivalente (duas variáveis) de $(U_i, U_{i-\tau})$, definido como se segue:

$$\mu_{2k,2l} = \frac{(2k)!(2l)!}{2^{k+l}} \sum_{j=0}^{\nu} \frac{(2\rho_u)^{2j}}{(k-j)!(l-j)!(2j)!} \quad (3.15)$$

$$\mu_{2k+1,2l+1} = \frac{(2k+1)!(2l+1)!}{2^{k+l+1}} \sum_{j=0}^{\nu} \frac{(2\rho_u)^{2j+1}}{(k-j)!(l-j)!(2j+1)!} \quad (3.16)$$

em que $\mu_{i,j} = 0$ se $k+l$ for ímpar e $\nu = \min(k, l)$. Substituindo a expressão para os momentos na equação (3.14), a expressão para ϕ pode ser transcrita para uma forma polinomial de mesma ordem m ,

$$\rho_x = \phi(r_u) = \sum_{j=1}^m c_j \rho_u^j \quad (3.17)$$

em que o vetor de coeficientes $\mathbf{c} = \{c_k\}_{k=1}^m$ só é expresso em termos de $\mathbf{a} = \{a_k\}_{k=1}^m$.

Expressões simples podem ser derivadas usando os polinômios de Tchebycheff-Hermite (Bhatt e Dave, 1964). Dessa forma, uma expressão analítica para ρ_u é possível se a equação (3.17) puder ser resolvida com respeito a ρ_u . Para simulações, conjectura-se que se g for monotônico, então ϕ é também monotônico dentro do intervalo $[-1, 1]$. Então ϕ^{-1} existe e uma única solução para ρ_u pode ser encontrada na equação (3.17). O próprio processo normal padrão U é completamente definido por meio ρ_u , e aplicando a transformação g da equação (3.12) para os componentes de U , a “cópia linear” Z de um determinado processo X é construída.

Note que a solução para ρ_u é analiticamente determinada da aproximação polinomial de g e requer somente o conhecimento dos coeficientes “ \mathbf{a} ” do polinômio e a autocorrelação ρ_x .

Em prática, opera-se com uma única série temporal \mathbf{x} em lugar de um processo X e com as amostras estimadas F_x e r_x para Φ_x e ρ_x , respectivamente. Os passos do algoritmo são como segue (Kugiumtzis, 2002b) :

1. Estime o vetor de coeficientes $\mathbf{a} = [a_1, \dots, a_m]'$ do polinômio p_m do gráfico $x_i = F_x^{-1}(F_0(w_i))$ (gráfico de x versus w) onde \mathbf{w} é um ruído branco normal padrão ($N(0, 1)$).
2. Calcule $\mathbf{c} = [c_1, \dots, c_m]'$ para o determinado “ \mathbf{a} ” nas equações (3.14) e

(3.17).

3. Encontre r_u na equação (3.17) para o determinado \mathbf{c} e r_x , usando as amostras estimadas r em vez de ρ . A prática comum é que exista solução e que seja única. Se não for o caso, repita os passos 1 e 2 para um novo \mathbf{w} até que uma única solução seja obtida.
4. Gere uma realização \mathbf{u} de um processo normal padrão com autocorrelação r_u . Escolheu-se fazer isso simplesmente por meio de um modelo autoregressivo de alguma ordem p , $\text{AR}(p)$. Os parâmetros $\mathbf{b} = [b_0, b_1, \dots, b_p]$ do $\text{AR}(p)$ são encontrados por meio de r_u usando as equações normais resolvidas efetivamente pelo algoritmo Levinson (Brockwell e Davis, 1991). O modelo $\text{AR}(p)$ é realizado para gerar u ,

$$u_{i+1} = b_0 + \sum_{j=1}^p b_j u_{i-(j-1)} + e_i, \quad e_i \rightarrow N(0, 1) \quad (3.18)$$

5. Transforme \mathbf{u} em \mathbf{z} reordenando \mathbf{x} para a mesma amplitude de distribuição de \mathbf{u} , ou seja, $z_i = F_x^{-1}(F_0(u_i))$.

Note que \mathbf{u} possui a mesma c.d.f. marginal normal F_0 e r_u apropriado, de forma que \mathbf{z} possua $F_z = F_x$, $r_z = r_x$, e é sobre tudo aleatório, como desejado. No entanto, na prática, a igualdade $r_z = r_x$ não é exata e r_z pode variar substancialmente em torno de r_x . Duas possíveis razões para isto são: a aproximação insuficiente de g no passo 1 e a variação inevitável da autocorrelação gerando \mathbf{u} no passo 4, que diminui com o aumento do tamanho dos dados. A primeira razão é devida a limitada potência dos polinômios (p deve ser menor que 15) aproximados por funções monotônicas, e esta deficiência causa repetições ocasionais dos primeiros passos do algoritmo como exposto no passo 3. A segunda razão constitui uma propriedade inerente da denominada “realizações típicas” aproximadas (ou seja, um modelo é usado para gerar os dados sub-rogados (Theiler e Prichard, 1995)) e que não podem ser controladas. Porém, menor variação na autocorrelação é alcançada quando o modelo $\text{AR}(p)$ é aperfeiçoado fazendo os seguintes passos, da mesma maneira como o CAAFT.

1. Aplique o algoritmo apresentado k vezes, obtendo z^1, \dots, z^k séries temporais sub-rogadas.
2. Calcule r_z^1, \dots, r_z^k e encontre um r_z^j mais próximo de r_x .

3. Use os parâmetros \mathbf{b} da repetição j para gerar os N_s dados sub-rogados (passos 4 e 5 do algoritmo anterior).

As k repetições anteriores bem como também as repetições do passos 1-3 da primeira parte do algoritmo pode reduzir a velocidade do algoritmo se a série temporal for longa, mas não têm nenhum impacto na função principal do algoritmo. Simplesmente, algumas realizações do ruído branco w são descartadas na procura dos parâmetros \mathbf{b} do modelo AR mais satisfatório que geram os dados sub-rogados (através da transformação g).

Os parâmetros livres do algoritmo STAP são: o grau m da aproximação polinomial de g , a ordem p do modelo AR, o número k de repetições para o otimização de $AR(p)$, e o máximo atraso τ_{max} , usado para comparar r_z^1, \dots, r_z^k com r_x . Normalmente, um m pequeno ($m \leq 10$) é suficiente (Kugiumtzis, 2002b). Para p , não existe nenhum gama ótima de valores, mas pode variar conforme a forma de r_x , por exemplo, um decaimento lento de r_x pode ser melhor modelado por um p grande. Kugiumtzis (2002b) propôs fixar $k = N_s = 39$ que corresponde ao número de sub-rogados gerados e $\tau_{max} = p$.

3.10 Embaralhamento de ciclos - algoritmo CS

Alguns dados são incompatíveis com ruídos lineares (ou uma transformação estática monotônica). Portanto, muitas séries temporais experimentais reais são claramente distintas de tal processo: os dados podem exibir assimetria¹⁰, determinismo com termos curtos ou uma tendência pseudo-periódica forte.

Definição 3.10.1 *Por pseudo-periódico significa uma série temporal que exiba uma tendência periódica forte que manifesta com um pico claro no sinal do espectro de potência¹¹. O sistema pode ser periódico com elementos aditivos ou ruídos dinâmicos, ou pode ser um entre muitos sistemas caóticos oscilatórios (por exemplo, o sistema de Lorenz e o de Rössler).*

A hipótese formulada por Theiler (1995) para sinais fortemente periódicos é bastante simples, mas muito poderosa. Theiler propõe que o sub-rogado gerado por embaralhamento dos ciclos se dirija a hipótese que não há nenhuma correlação

¹⁰Tal assimetria é uma indicação de não-linearidade no sistema.

¹¹Para ser mais preciso deve-se quantificar o que quer dizer por "picos claros" e "tendência periódica forte". Porém, escolhe não fazer isto com a imposição de um limite arbitrário. Esta definição é deliberadamente vaga.

dinâmica entre os ciclos. A estrutura do algoritmo CS utiliza os seguintes passos (Theiler, 1995):

1. Divide o sinal $X_k = \{x(k)\}_{k=1}^N$ em seus ciclos individuais (identifique o local dos picos, ou algum outro ponto conveniente dentro de cada ciclo).
2. Aleatoriamente, reordene os ciclos e forme uma nova série temporal s_k concatenando os ciclos individuais. Lembrando que, se a série temporal original X_k for ligeiramente (ou pouco) não-estacionária, os ciclos individuais certamente terão que ser deslocados verticalmente para preservar a “continuidade” da série temporal original X_t no sub-rogado.

Portanto, Theiler (1995) sugeriu uma aproximação alternativa análoga ao algoritmo 0. Em vez do embaralhamento dos dados individualmente em uma série temporal, embaralha-se os ciclos individuais. Esse embaralhamento de ciclos deve destruir qualquer estrutura com um período mais longo que o comprimento do ciclo. Logo, deve ser considerado que os dados sejam periódicos com período cíclico fixo, ou seja, este método não serve para séries quase-periódicas. Para os dados considerados por Theiler (1995), eletro-encefalograma epiléptico, isto não era um problema - os dados eram significativamente estacionários e exibiam fortes picos súbitos. Porém, se os dados forem suaves, será necessário forçar a identificação dos pontos de ruptura satisfatórios. Portanto, os sub-rogados necessariamente possuirão qualquer descontinuidade espúria (nos pontos onde os ciclos foram reajuntados) ou experimentam não-estacionariedades (devido a dependência intra-ciclo, mas não necessariamente de períodos mais longos que o pseudo-período). Igualmente com convenientes pontos de ruptura entre ciclos, Theiler e Rapp (1996a) observaram termos mais longos espúrios na autocorrelação relatadas para os sub-rogados CS (*cycle-shuffle*).

O grau em que este método torna aleatório o dado sub-rogado dependerá do número de ciclos presentes na série temporal em teste. Por exemplo, se fosse empregada uma medida dinâmica, tal como dimensão de correlação, como uma estatística de teste e a janela de imersão como d_w , então os pontos de imersão só difeririam dos verdadeiros pontos se a janela de imersão cruzasse o ponto de ruptura do ciclo. Se uma série temporal com N pontos tiver p ciclos, logo cada ciclo tem aproximadamente o comprimento de N/p , e $N/p - d_w + 1$ pontos imersos idênticos durante cada ciclo da série original e dos dados sub-rogados gerados (se $N/p > d_w$). Não obstante, tem-se que assegurar então que a proporção de cada sub-rogado imerso, idênticos a série original, $1 - (p/N)(d_w - 1)$, seja pequena.

3.11 Sub-rogado pseudo-periódico - algoritmo PPS

Nessa seção, o algoritmo sub-rogado descrito é chamado de pseudo-periódico (PPS) ou *pseudoperiodic surrogate* (Small et al., 2001b; Small e Tse, 2002; Small et al., 2005a). Este método é baseado no conhecimento do método de modelagem linear local descritos por Mees e Judd (1995) e Sugihara e May (1990). Previamente, Small e Judd (1998b) defenderam e implementaram a base radial não-linear como uma forma de testar a hipótese nula. O método PPS pode ser aplicado para testes contra a hipótese nula de uma órbita periódica com ruído não correlacionado para um número muito grande de sistemas experimentais com comportamento pseudo-periódico.

Métodos sub-rogados com padrões lineares só são úteis para séries temporais que não exibam nenhuma estrutura pseudo-periódica. Este algoritmo pode distinguir entre uma órbita periódica com ruído (dinâmico ou aditivo) ou ainda órbita periódica com ruído correlacionado. Possíveis origens de determinismo aperiódico inter-ciclos dinâmicos inclui: não-periódicos lineares ou dinâmicas não-lineares, ou caos. Este algoritmo é fundamentado em imitar a ampla dinâmica com um modelo local, mas suprimir as grandezas características como dinâmicas ruidosas.

O algoritmo PPS emprega um tempo de atraso de imersão (τ) no dado para extrair as características topológicas da dinâmica subjacente. Deseja-se gerar dados sub-rogados que preservem o comportamento em larga escala dos dados (a estrutura periódica), mas destrua qualquer estrutura intra-ciclo (por exemplo, estrutura linear ou não-linear determinística, caótica). Várias implementações de técnicas de modelagens são propostas, tais como: modelos locais lineares, simples modelos locais constantes e triangulações. Small e Tse (2002) estudado a aplicação destes vários métodos para geração de sub-rogados. Porém, foi encontrado que a aproximação mais simples na verdade é a com melhor execução. Portanto, a implementação empregada aqui será um modelo local constante em espaço de estados. Este método evita as complicações adicionais dessas estruturas alternativas.

A idéia desse algoritmo é que as séries temporais sub-rogadas são geradas obtendo a dinâmica subjacente de tal modelo local e contaminando uma trajetória do atrator com ruído dinâmico. Com uma escolha apropriada de nível de ruído dinâmicas intra-ciclos (dentro dos ciclos, pequena escala) são preservadas, mas dinâmicas inter-ciclos (ciclo propriamente dito, grande escala) não são. A hipótese nula desse sub-rogado é uma órbita periódica com ruído não-correlacionado.

Esta metodologia requer um algoritmo apropriado para gerar conjuntos de da-

dos sub-rogados consistentes com uma hipótese de interesse. Os mais comuns métodos de geração de sub-rogados permitem somente testar simples dinâmicas lineares estocásticas. Para dados que exibem componentes periódicos fortes, tais hipóteses são trivialmente falsas. O teste de dados sub-rogados descrito aqui é capaz de testar séries temporais pseudo-periódica para adicional (aperiódico) determinismo

O dado sub-rogado pseudo-periódico (PPS) tem que satisfazer dois critérios:

- (i) o dado PPS têm que exibir a mesma estrutura periódica que a série temporal original $x(k)$;
- (ii) o dado PPS não exibi nenhuma outra estrutura determinística.

Para alcançar estas metas tem-se que caracterizar a dinâmica responsável por $x(k)$ e discriminar entre comportamentos periódicos e aperiódicos. Para caracterizar a dinâmica subjacente emprega-se o **tempo de atraso de imersão** (*time delay embedding*) e técnicas de modelagem locais (efetivamente, emprega-se um modelo local constante). Tornar aleatório e discriminação entre estrutura periódica e aperiódica são alcançadas por meio das repetidas trajetórias aleatórias entre vizinhos (espaciais).

Seja $\{x_k\}_{k=1}^N$ uma série temporal escalar com N observações (tamanho da amostra). A **dimensão de imersão** (d_e) e o **tempo de atraso** (τ) reconstroem a dinâmica subjacente de acordo com a teoria da imersão de Takens (Takens, 1981):

$$z_k = (x_k, x_{k-\tau}, x_{k-2\tau}, \dots, x_{k-d_w}) \quad (3.19)$$

para $k = \{d_w + 1, \dots, N\}$ onde $d_w = (d_e - 1)\tau$ é a janela de imersão. Por conveniência, deixe re-indexar a série temporal imersa para ser $\{z_k\}_{k=1}^{N-d_w}$. Esse sub-rogado PPS pode ser demonstrado como se segue (Small e Tse, 2002):

1. Defina o vetor de atraso de imersão $\{z_k\}_{k=1}^{N-d_w}$ da série temporal escalar $\{x_k\}_{k=1}^N$ como $z_k = (x_k, x_{k-\tau}, x_{k-2\tau}, \dots, x_{k-d_w})$. Para simplicidade da anotação defina-se que a **janela de imersão** $d_w = (d_e - 1)\tau$, onde d_e e τ são a dimensão de imersão e o tempo de atraso de imersão, respectivamente.
2. Escolha aleatoriamente uma condição inicial $s_1 \in \{z_k : k = 1, \dots, N - d_w\}$.
3. Deixe $i = 1$.

4. Selecione aleatoriamente um dos vizinhos de s_i em $\{z_k : k = 1, \dots, N - d_w\}$, dito s_j , com probabilidade

$$Prob(s_j = z_k) \propto \exp \frac{-\|z_k - s_i\|}{\rho}, \quad (3.20)$$

o parâmetro ρ é um ruído radial.

5. Fixe $s_{i+1} = s_{j+1}$ e incremente i .
6. Repita este procedimento do passo 4 até $i = N$.
7. Tome como a série temporal sub-rogada $\{(s_k)_1 : t = 1, \dots, N\}$ em que $(\cdot)_1$ denota a primeira coordenada escalar do vetor.

O algoritmo requer seleção de três parâmetros: **dimensão de imersão** d_e , **tempo de atraso** τ , e o **ruído radial** ρ . O ruído radial deve ser selecionado cuidadosamente. Se ρ for muito grande então o algoritmo PPS introduzirá alta aleatoridades e os sub-rogados já não se assemelham mais com a série temporal original. Reciprocamente se ρ for muito pequeno, então o algoritmo PPS introduzirá aleatoridades insuficientes, e os sub-rogados conteriam determinismo aperiódico. Um valor intermediário de ρ pode ser selecionado computando o número máximo esperado de segmentos de pontos do sub-rogado que são idênticos a série temporal original (Small et al., 2001b).

Portanto, de acordo com Small et al. (2001b), este algoritmo pode ser aplicado para testar dinâmica determinística aperiódica em séries temporais pseudo-periódicas. Reciprocamente, sistemas dinâmicos consistentes com a hipótese nula deste sub-rogado exibem uma órbita periódica com ruído não correlacionado. Para séries temporais que exibem dinâmica pseudo-periódica a rejeição insinua a existência de dinâmica determinística não-periódica. Porém, para séries temporais que não exibem dinâmica pseudo-periódica (Small e Tse, 2002) o teste mostrará que os sub-rogados gerados por este método são realmente consistente com dinâmicas determinísticas com termos curtos. Rejeição deste teste é então evidência para dinâmicas determinísticas com termos longo. A distinção entre “termo longo” e “termo curto” está talvez mal definida, e certamente dependente da seleção de ρ . Porém, para o corrente estudo uma definição mais precisa está além de nossas exigências. É suficiente notar que o sub-rogado gerado por esta última hipótese se pareça muito mais semelhante aos dados sub-rogados gerados pelos Algoritmos 0, 1, e 2. Portanto, rejeitar a H_0 do deste PPS indicaria a presença de dinâmica determinística não-trivial nos dados.

3.12 Comentários Finais

Neste Capítulo 3 foi feita uma revisão bibliográfica a respeito dos mecanismos geradores de dados sub-rogados, onde foram apresentadas as principais técnicas (algoritmos) que aparecem na literatura contextualizando as técnicas sub-rogadas desenvolvidas para cada tipo de hipótese nula.

O próximo capítulo (Capítulo 4) é de cunho mais prático. Nele será tratada a aplicabilidade de cada um dos algoritmos geradores de sub-rogados com a análise da estatística de teste para diversos modelos conhecidos *a priori* (simulações) e dados experimentais (reais).

A avaliação da decisão (aceitação ou rejeição) da hipótese nula de interesse será discutida, tendo em mente, identificar as dificuldades relevantes que são levantadas de diversos aspectos concernentes à construção do dado sub-rogado. Logo, serão tratadas as restrições e os problemas levantados na aplicação do teste de hipótese nula utilizando o método de dados sub-rogados e será proposto um procedimento de análise para o quadro geral.

Análise dos Métodos de Dados Sub-rogados: Proposição de um Procedimento Geral

4.1 Introdução

No capítulo anterior foi visto a construção da série temporal sub-rogada de acordo com a hipótese nula de interesse que é de fundamental importância para análises posteriores. Neste Capítulo 4 será ilustrado e investigado a utilidade do teste de hipótese nula utilizando o método de dados sub-rogados aplicado em diversas séries temporais reais e simuladas.

Primeiramente, serão apresentadas evidências de que o método aplicado, a princípio, identifica algum tipo de comportamento para os problemas propostos. Porém, medir a confiabilidade da decisão tomada sobre a hipótese nula e a aproximação dos dados sub-rogados com os dados originais corretamente requer vários testes tanto com dados numéricos simulados quanto com dados experimentais. Os algoritmos geradores de dados sub-rogados e suas estatísticas de testes (paramétrica e não-paramétrica) aplicadas foram implementados utilizando o software MATLAB7.

O fluxograma na Figura 4.1 proporciona uma visão geral das possíveis decisões tomadas concernente a cada hipótese nula testada utilizando os algoritmos geradores de dados sub-rogados. Em face à decisão tomada de que tipo de modelo a ser usado, será proposto uma gama de exemplos abrangendo algumas das possíveis estruturas de modelos, segundo a seqüência proposta pelo fluxograma.

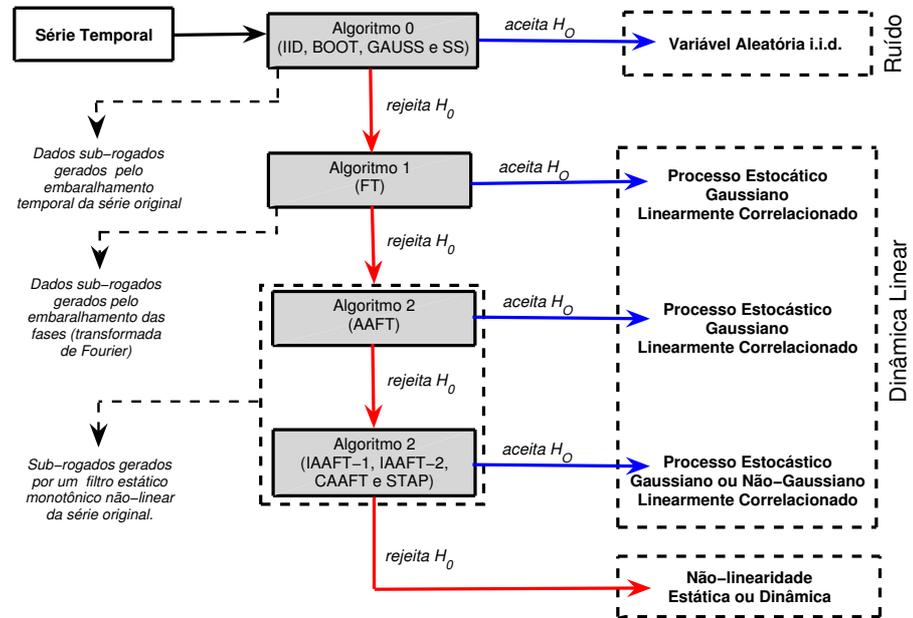


Figura 4.1: Fluxograma geral para teste de hipótese nula (H_0) utilizando os métodos de dados sub-rogados.

Os testes estatísticos paramétricos propostos neste trabalho foram empregados, a princípio, em dados de processos com distribuições normais (ou gaussianos) o que não é estritamente aplicado aos sistemas reais, sendo assim, um dos problemas levantados para aplicação desse tipo de teste (paramétrico) em vários processos reais.

Nos testes não-paramétricos, supõe-se que a distribuição de seus dados experimentais não seja normal (ou gaussianos), ou que não tenha elementos suficientes para poder afirmar que seja. O critério estatístico não-paramétrico examina então a ordem (rank) da sequência dos valores estatísticos da série temporal original e seus dados sub-rogados gerados. Os testes estatísticos não-paramétricos mais utilizados são a função de autocorrelação (FAC), a informação mútua média (IMM) e a dimensão de correlação (d_c). Tais estatísticas não-paramétricas são mais criteriosas para tomadas de decisões e serão preferencialmente aplicadas neste capítulo

para análise estatística.

O conjunto de séries temporais propostas para investigação, neste capítulo, são respectivamente: dados gerados por meio de variáveis aleatórias (ruídos) gaussianas e não-gaussianas, séries temporais financeiras, processos lineares gaussianos (AR, MA ou ARMA), processos lineares não-gaussianos (AR, MA ou ARMA), transformação não-linear aplicada à processos lineares realimentados com inovações não-gaussianas, mapas, a série de Mackey-Glass, algumas séries temporais caóticas, um dado experimental da série temporal do NMR (Ressonância Magnética Nuclear), um estudo de caso e um caso “blind” para proposição de um procedimento geral de análise.

Em todas as séries temporais propostas são investigadas as restrições ou problemas existentes nas análises, revelando que nem sempre tais técnicas são adequadas e que existe situações que conduzem a resultados errôneos para certas estruturas de modelos.

No final deste capítulo é proposto um procedimento geral para análise do quadro geral dos possíveis comportamentos da série temporal, ou seja, um fluxograma de procedimentos para análise do dado em estudo. Sendo assim, é feito um estudo de caso conhecido *a priori* e um estudo “blind” cujo comportamento do dado não é conhecido *a priori*, para avaliação dos resultados.

4.2 Série simulada

A primeira aplicação do teste de hipótese utilizando o método de dados sub-rogados é para com o caso mais simples, trata-se de um processo estocástico aleatório (ruído não-correlacionado). Tal série temporal é constituída de variáveis aleatórias independentes e identicamente distribuídas (i.i.d.). Essa proposição atribui a inexistência de algum tipo de dinâmica determinística nos dados em questão. Primeiramente, tal série temporal é gerada usando variáveis aleatórias com distribuição gaussiana de desvio padrão unitário, ou seja, uma distribuição normal $N(0, 1)$.

Por meio da série temporal original proposta são gerados 99 séries temporais sub-rogadas ($\alpha = 0,01$) para o teste de hipótese, ambas com $N = 9000$ pontos de um conjunto de 10000 pontos gerados, selecionados conforme mostrado na Figura 4.2 (entre as interações 1001 à 10000). São descartados os primeiros 1000 pontos da série temporal original.

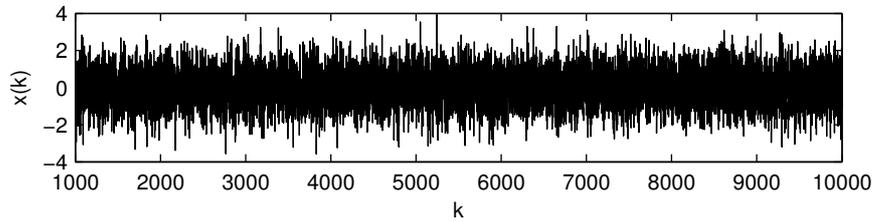


Figura 4.2: Variáveis aleatórias gaussianas (1001-10000) com $N(0, 1)$.

4.2.1 Estatística de teste paramétrica

Os resultados obtidos dos testes de hipóteses para os mecanismos geradores de dados sub-rogados utilizando o algoritmo 0 (técnicas IID, BOOT e GAUSS) e o algoritmo SS (com amplitude $A = 1$) são apresentados na Tabela 4.1, bem como as suas respectivas probabilidades¹ obtidas dos testes com um nível de significância α de 1%. A H_0 a ser testada é que a série temporal original é um ruído i.i.d.

As três hipóteses nulas a serem testada são: a série temporal original é um ruído i.i.d. (algoritmo IID e BOOT), a série temporal original é um ruído i.i.d. gaussiano (algoritmo GAUSS) ou a série temporal original é um ruído i.d. (algoritmo SS)

A Tabela 4.1 apresenta os resultados obtidos quanto a decisão tomada para a H_0 , os valores S calculados com as respectivas estatísticas de testes (série original e 99 séries sub-rogadas) escolhidas e suas probabilidades em rejeitar ou não a H_0 .

É importante lembrar que o limiar crítico (para $\alpha = 0,01$) utilizado nas estatísticas empregadas, determinado a partir da distribuição normal $N(0, 1)$, é de $-2,58 \leq S \leq 2,58$ para *não rejeitar* a H_0 e de $S < -2,58$ ou $S > 2,58$ para *rejeitar* a H_0 .

Os dois métodos paramétricos, reversão assimétrica temporal e três pontos da autocorrelação (TPA), utilizados foram muito úteis para o algoritmo IID² não rejeitando a H_0 , considerando, estatisticamente, que não existe nenhuma dinâmica determinística na série temporal original, lembrando que foi escolhido apenas estatísticas de alta ordem para coerência na análise neste tipo de algoritmo.

¹Probabilidade de não rejeitar H_0 é de $p > 0,01$ e a probabilidade em rejeitar H_0 (erro tipo I) é de $p < 0,01$.

²Lembrando que a média e variância dos sub-rogados gerados pelo IID e SS são iguais ou aproximadamente iguais com os dados originais, portanto não são um bom discriminante estatístico os demais testes paramétricos propostos.

Tabela 4.1: H_0 associada à realização dos testes estatísticos e sua respectiva probabilidade com 99 dados sub-rogados gerados.

Técnica	Estatística de teste	S	Probabilidade	Decisão H_0
IID	Reversão temporal	-0,20094	0,84074	não rejeita
	TPA	-0,43528	0,66336	não rejeita
GAUSS	Variância	-0,17095	0,86427	não rejeita
	Assimetria	0,68533	0,49313	não rejeita
	Kurtoses	-0,81966	0,41241	não rejeita
	C_6	-0,81096	0,41739	não rejeita
	Reversão temporal	-0,20206	0,83987	não rejeita
	TPA	-0,33416	0,73826	não rejeita
BOOT	Variância	0,096035	0,92349	não rejeita
	Assimetria	-0,037323	0,97023	não rejeita
	Kurtoses	0,066262	0,94717	não rejeita
	C_6	0,19999	0,84149	não rejeita
	Reversão temporal	-0,37171	0,71011	não rejeita
	TPA	-0,078853	0,93715	não rejeita
SS	Reversão temporal	-0,28681	0,77426	não rejeita
	TPA	-0,11977	0,90467	não rejeita

O resultado com o BOOT mostrou-se estar estatisticamente de acordo com a H_0 , não rejeitando H_0 , resultando que os dados originais são ruídos i.i.d. em todos os testes paramétricos analisados.

Na aplicação com o algoritmo GAUSS não se pode rejeitar a H_0 , portanto, estatisticamente, os dados originais são ruídos i.i.d. gaussianos.

Outra hipótese que não se pode rejeitar é a NH_{SS} , algoritmo SS, que afirma, estatisticamente, nesse exemplo que a série temporal original consiste de variáveis aleatórias i.d.

Uma outra análise pode ser construída por meio de um gráfico da distribuição dos 99 valores (quantitativos) encontrados nas estatísticas de testes paramétricas (Q_n) para cada dado sub-rogado gerado, sendo também identificado no mesmo gráfico a estatística de teste paramétrica da série temporal original (Q_0). Tal localização de Q_0 pode confirmar estatisticamente a rejeição ou não da hipótese nula em estudo, respeitando os limiares (traços nas extremidades) da região de aceitação³. A Figura 4.3, 4.4, 4.5 e 4.6 ilustram esses gráficos.

³Dentro das extremidades (zona de aceitação) estatisticamente 99% da não rejeição da H_0 ou aproximadamente $\mu \pm 3\sigma$.

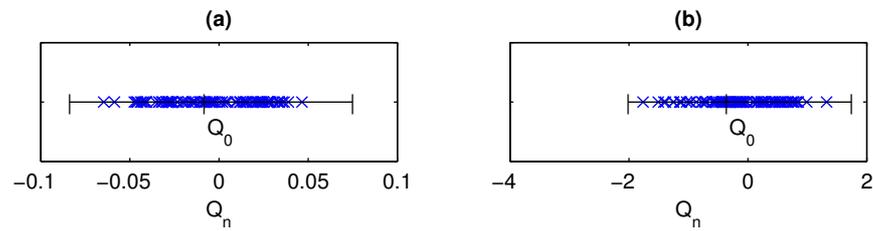


Figura 4.3: Distribuição das estatísticas Q_n IID: em (a) teste com a reversão temporal e em (b) teste com TPA. Em (x) os 99 valores da estatística de teste para as séries sub-rogadas e o traço vertical (Q_0) da série temporal original. Os traços nas extremidades identificam o limiar da região de aceitação do teste com 99%.

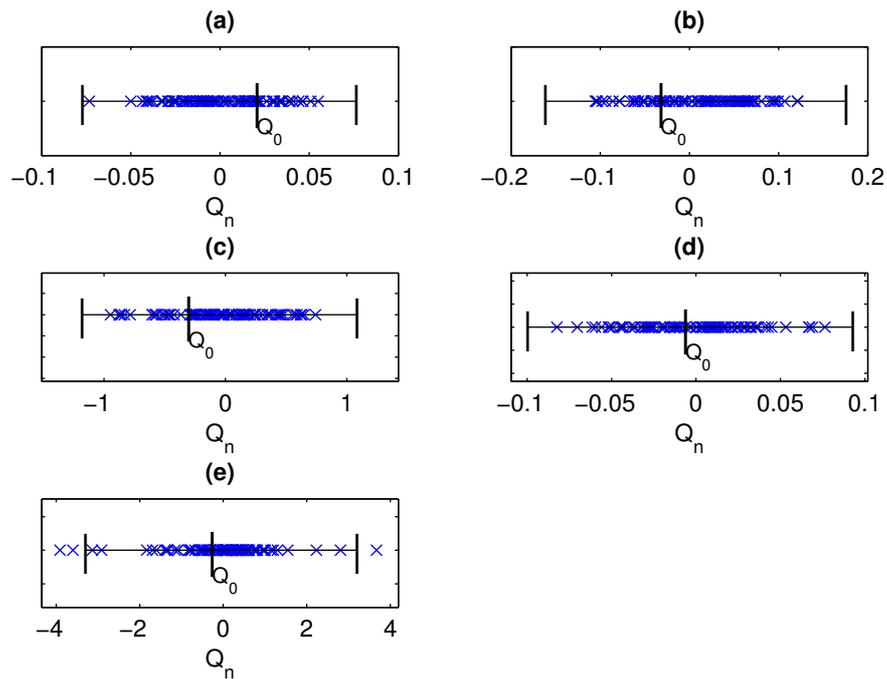


Figura 4.4: Distribuição das estatísticas Q_n GAUSS: em (a) teste com variância, em (b) teste com Assimetria, em (c) teste com Kurtoses, em (d) teste com C_6 , em (e) teste com a reversão temporal e em (f) teste com TPA . Em (x) os 99 valores da estatística de teste para as séries sub-rogadas e o traço vertical (Q_0) da série temporal original. Os traços nas extremidades identificam o limiar da região de aceitação do teste com 99%.

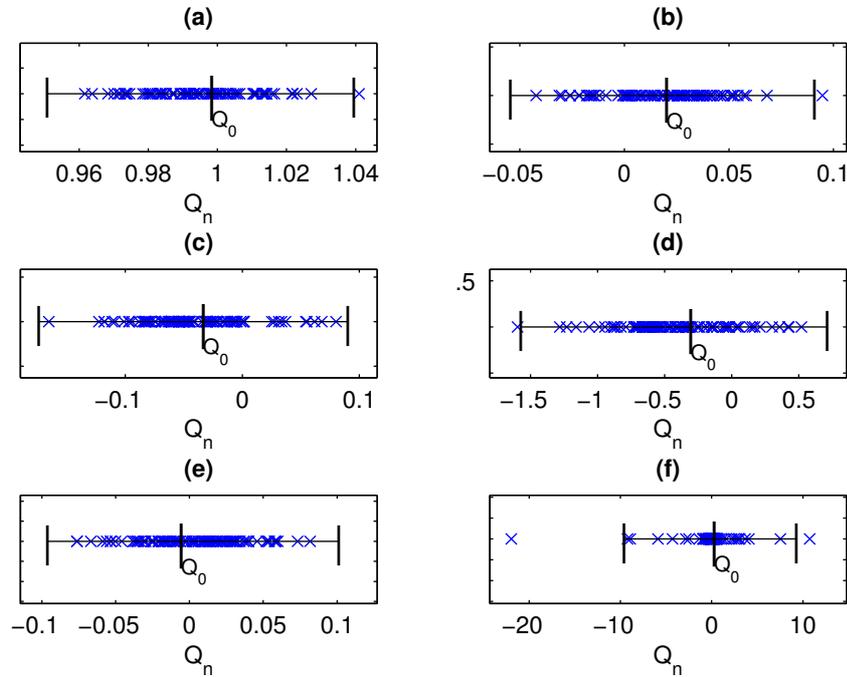


Figura 4.5: Distribuição das estatísticas Q_n BOOT: em (a) teste com variância, em (b) teste com Assimetria, em (c) teste com Kurtoses, em (d) teste com C_6 , em (e) teste com a reversão temporal e em (f) teste com TPA . Em (x) os 99 valores da estatística de teste para as séries sub-rogadas e o traço (Q_0) da série temporal original. Os traços nas extremidades identificam o limiar da região de aceitação do teste com 99%.

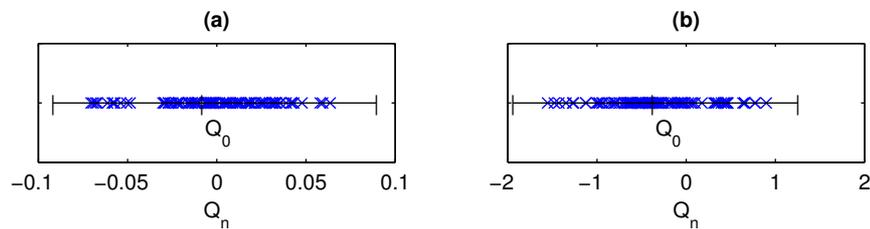


Figura 4.6: Distribuição das estatísticas Q_n SS: em (a) teste com a reversão temporal e em (b) teste com TPA. Em (x) os 99 valores da estatística de teste para as séries sub-rogadas e o traço vertical (Q_0) da série temporal original. Os traços nas extremidades identificam o limiar da região de aceitação do teste com 99%.

4.2.2 Estatística de teste não-paramétrica

A função de autocorrelação (FAC) é a medida da dependência temporal entre as observações, isto é, uma estimativa da correlação linear dos dados, como os dados são não-correlacionados não é necessário aplicar esta estatística de teste não-paramétrica, a princípio, para o algoritmo 0, ilustrado na Figura 4.7. Contudo, Small et al. (2005b) aplica a FAC e a informação mútua média (IMM) para o algoritmo SS, pelo fato que os dados sub-rogados gerados devem ser suficientemente semelhantes a série temporal original (Figura 4.8).

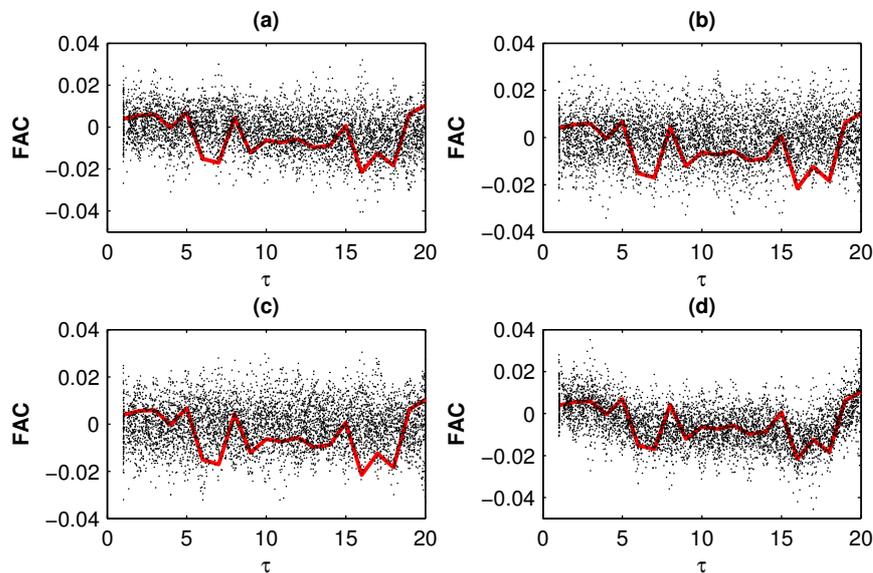


Figura 4.7: Gráfico da FAC: (a) IID em (b) GAUSS, (c) BOOT e (d) SS. A linha sólida é a FAC da série original e as linhas pontilhadas são as 99 FAC das séries sub-rogadas geradas.

A Figura 4.8 ilustra a FAC e IMM da série original e dos 99 séries sub-rogados geradas pelo algoritmo SS. A FAC e IMM não mostram nenhuma diferença significativa, ou seja, todos os dados originais caem dentro das distribuições dos dados sub-rogados gerados. Logo, não se pode rejeitar a hipótese nula NH_{SS} que a série temporal original é um conjunto de variáveis aleatórias i.d. Confirmando assim que não existe nenhuma dinâmica determinística.

Para rejeitar H_0 a FAC ou a IMM da série original devem cair fora da distribuição dos dados sub-rogados SS. Segundo Small e Nakamura (2006a) só uma estatística (FAC ou IMM) não é bastante para alguns casos, e esta é a razão principal para adotar-se neste trabalho duas estatísticas distintas, a FAC e a IMM

que mostram que quando os dados forem essencialmente aleatórios, a FAC e a IMM da série original tem que cair teoricamente dentro da distribuição dos dados sub-rogados, $r_x(\tau) \leq r_z(\tau)$ e $I_x(\tau) \leq I_z(\tau)$ ⁴.

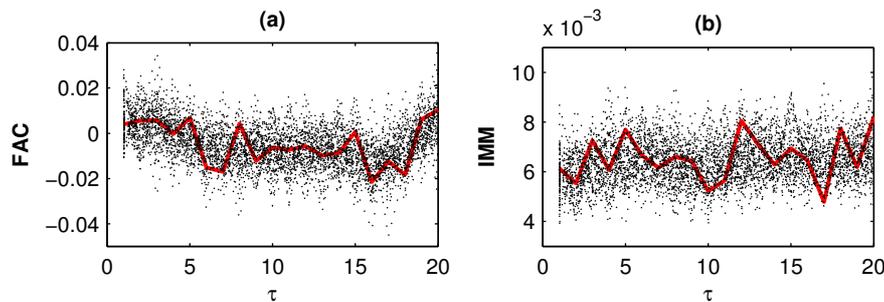


Figura 4.8: Gráfico da FAC em (a) e da IMM em (b). A linha sólida é a FAC da série original e as linhas pontilhadas são as 99 FAC das séries sub-rogadas SS geradas.

A Figura 4.9 ilustra os gráficos quantil-quantil utilizado para determinar se duas amostras, série original e dos dados sub-rogados⁵ para cada algoritmo, vêm da mesma família de distribuição.

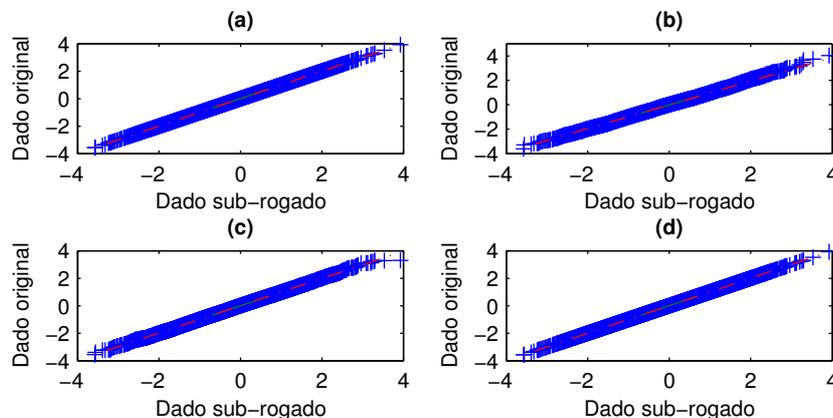


Figura 4.9: Gráfico quantil-quantil: (a) série original vs. 1 série sub-rogada IID (b) série original vs. 1 série sub-rogada GAUSS, (c) série original vs. 1 série sub-rogada BOOT e (d) série original vs. 1 série sub-rogada SS

Para algumas aplicações com o algoritmo 0, empregando as técnicas GAUSS e BOOT, podem revelar por meio do gráfico Quantil-quantil (QQ) uma ligeira

⁴ z é o conjunto de sub-rogados gerados, r é a FAC e I é a IMM.

⁵Apenas uma série entre as 99 geradas.

ou grande diferença nas distribuições. Tais discordâncias nos gráficos QQ são devido a construção dos dados sub-rogados propostos para tais técnicas. Essas construções, respectivamente, sugerem mudanças na escolha da sequência temporal dos dados originais (BOOT) e na distribuição de amplitude⁶ (GAUSS).

É importante lembrar que a H_0 é marginalmente rejeitada a um nível de confiança de 99% com um fator estatístico discriminante linear (FAC) e também com um fator estatístico discriminante não-linear (IMM), ou seja, existe a possibilidade de 1% em não se rejeitar a H_0 .

A necessidade de se identificar qualquer variável aleatória i.i.d. (ou ruído aleatório i.i.d.) e não apenas v.a.s gaussianas i.i.d. é verificada com a aplicação do método proposto de ruídos com distribuições não-gaussianas, tais como, a distribuição uniforme [0,1], distribuição beta (5,2), a distribuição de Laplace (0,1) e a distribuição de Cauchy (0,0.001), respectivamente. Todas essas distribuições são relacionadas na Tabela A.1 no Anexo A. Tais distribuições serão aplicadas e tratadas de forma mais sucinta nas seções seguintes.

As Figuras 4.10, 4.11, 4.12 e 4.13 mostram os gráficos das FAC obtidas no teste das 99 séries sub-rogados gerados, a partir de uma série temporal com v.a.s não-gaussianas ($N = 9000$ observações) propostas para o teste de hipótese nula aplicando apenas o algoritmo 0 (IID e SS).

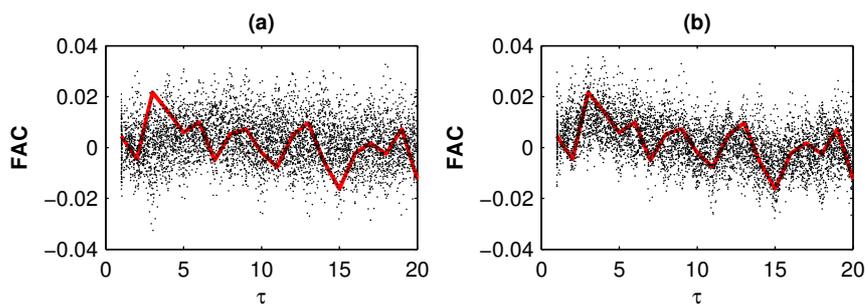


Figura 4.10: Gráfico da FAC da v.a com distribuição uniforme [0,1]: (a) sub-rogado IID e (b) sub-rogado SS. A linha sólida é a FAC da série original e as linhas pontilhadas são as 99 FAC das séries sub-rogadas geradas.

⁶Preserva apenas a média e a variância.

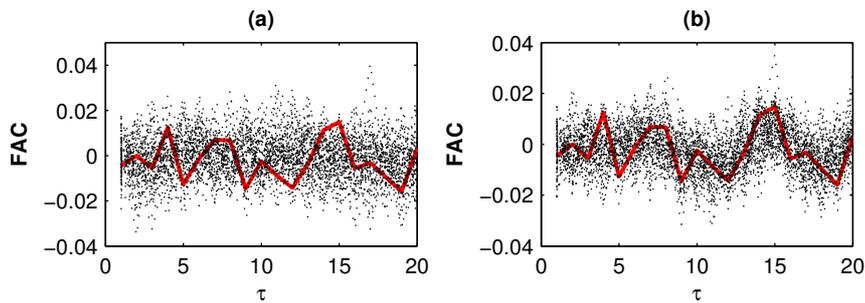


Figura 4.11: Gráfico da FAC da v.a com distribuição Beta (5,2): (a) sub-rogado IID e (b) sub-rogado SS. A linha sólida é a FAC da série original e as linhas pontilhadas são as 99 FAC das séries sub-rogadas geradas.

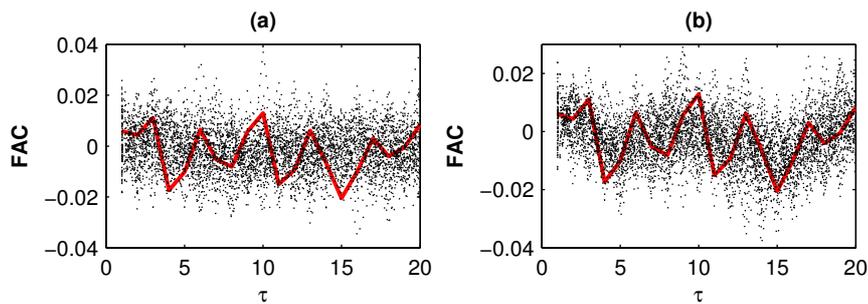


Figura 4.12: Gráfico da FAC da v.a com distribuição Laplace (0,1): (a) sub-rogado IID e (b) sub-rogado SS. A linha sólida é a FAC da série original e as linhas pontilhadas são as 99 FAC das séries sub-rogadas geradas.

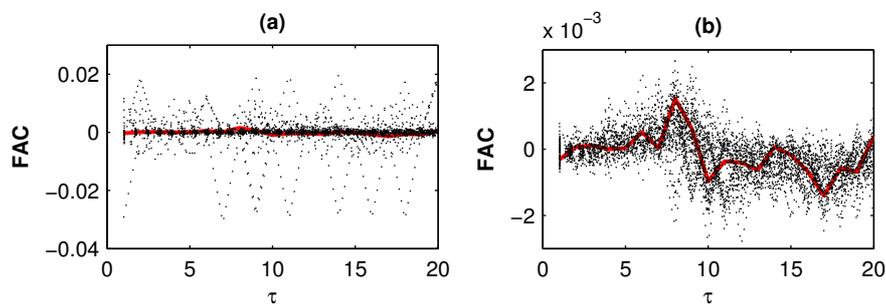


Figura 4.13: Gráfico da FAC da v.a com distribuição Cauchy (0,0.001): (a) sub-rogado IID e (b) sub-rogado SS. A linha sólida é a FAC da série original e as linhas pontilhadas são as 99 FAC das séries sub-rogadas geradas.

Todos as variáveis aleatórias i.i.d com suas respectivas distribuição propostas mostraram se condizentes com o algoritmo 0, não sendo possível rejeitar H_0 por meio dos métodos paramétricos (seção Tabelas no anexo B) e não-paramétricos com nível de confiança de 99% dos casos, ou seja, as séries temporais são possivelmente variáveis aleatórias i.i.d. ou i.d. Portanto, não existe estatisticamente nenhuma dinâmica determinística nas séries temporais em análise, o que conduz com o conhecimento “*a priori*” da série original.

Um fluxograma é apresentado na Figura 4.14 seqüenciando o procedimento geral inicial para o algoritmo 0.

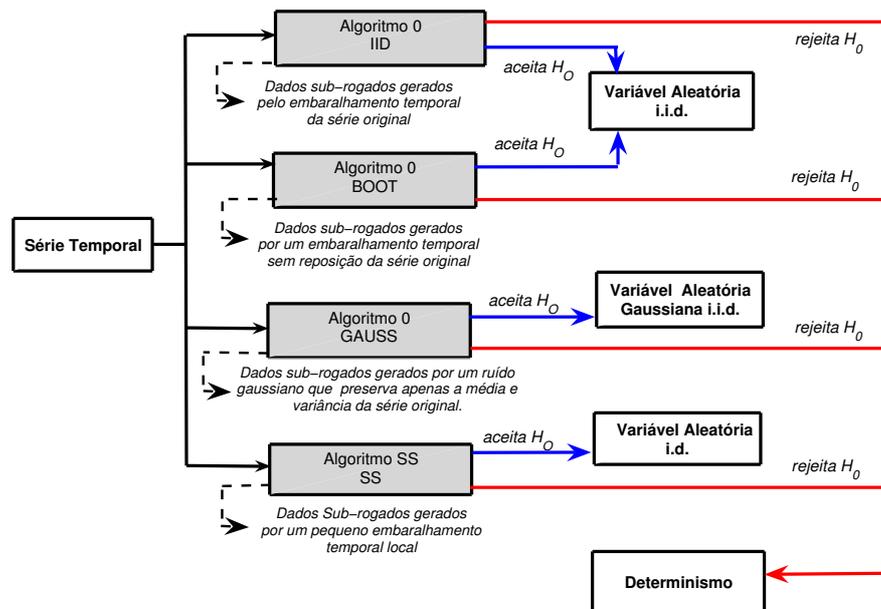


Figura 4.14: Fluxograma geral para teste de hipótese nula (NH_0) utilizando os métodos de dados sub-rogados IID, GAUSS, BOOT e SS.

A conclusão dessa primeira aplicação utilizando os métodos sub-rogados do algoritmo 0 (IID, GAUSS e BOOT) e algoritmo SS é que tais técnicas são de grande utilidade para identificação de ruídos, sejam eles gerados a partir de uma

distribuição gaussiana ou não-gaussiana. O método IID é considerado mais útil para identificação de determinismo em diversas aplicações, pois preserva a distribuição da amplitude da série temporal original e destroi qualquer tipo de correlação ou tendência existente entre os dados.

Na próxima seção é realizado uma aplicação em um caso prático para identificação de determinismo em uma série financeira. Para o procedimento geral é sugerido aplicar apenas os algoritmo IID e SS, respectivamente, devido a preservação da distribuição da amplitude que ambas as técnicas proporcionam.

4.3 Séries temporais financeiras

Questionou-se por muito tempo se as flutuações irregulares de dados financeiro são simplesmente variáveis aleatórias ou possuem algum tipo de dinâmica. Se fosse possível confirmar que estas flutuações irregulares não são variáveis aleatórias, isso implicaria a existência de algum tipo de estrutura dinâmica. Portanto, seria possível construir modelos para tais séries temporais. Claramente, tais modelos são de imenso valor para a compreensão e para a predição.

Segundo Nakamura e Small (2005) em mercados financeiros, onde é possível ter procedimentos financeiros no mundo inteiro, o fluxo de dinheiro é vasto, e há um número enorme de negociantes. A lógica para negociação e ação é bastante simples. Como o propósito é maximizar seus ganhos, compra-se e vende-se quando é esperado que o preço suba ou decline. Porém, as situações e condições para cada negociante não são as mesmas. Provavelmente, os negociantes comprarão independentemente e venderão dependendo das situações e condições. Conseqüentemente, embora o mercado financeiro seja um mundo artificial, é impossível controlar todos as especulações dos negociantes, sendo assim impossível controlar o mercado. Conseqüentemente, pode existir dois aspectos contrários para mudança de preço que normalmente se vê ou se obtém: o preço é um sistema dinâmico (determinístico), e o outro é um sistema aleatório (probabilístico). Como a ação de cada negociante como comprar e vender é determinística, cada mudança de preço deve ter algum papel que reflète na intenção dos negociantes.

Por outro lado, o número de negociantes é enorme, e as ações serão feitas quase independentemente. Conseqüentemente, a dinâmica pode ser tratada como um sistema de alta dimensão. Às vezes é mais apropriado tratar dados de sistemas de alta dimensão como aleatórios. Além disso, os dados financeiros que normalmente se vê são um conjunto de muitas ação de negócios. O conjunto de operações

pode cancelar individualidades e pode ocasionar um comportamento estocástico coletivo. Disso, espera-se também que estas mudanças de preço podem ser variáveis aleatórias. Isto quer dizer, mudanças de preço em mercados financeiros são fenômenos que a dinâmica de evolução microscópica é determinística o que reflete a intenção de cada negociante, porém, o comportamento coletivo aparenta ser estocástico.

Para investigar algumas características dos dados é proposto o método de dados sub-rogados, e a idéia tornou-se uma ferramenta central para validar os resultados da análise dinâmica, esta idéia é amplamente aplicada em forma de teste de hipótese (Nakamura e Small, 2005; Small e Nakamura, 2006b).

Uma série financeira é investigada: a taxa diária do câmbio da moeda brasileira Real (R\$) em U\$ dólar americano (Real/USD). A taxa diária do câmbio comercial entre a moeda Real e o Dólar (Real/USD) para teste com o método de hipótese utilizando dados sub-rogados são amostradas (Gazeta Mercantil - GM366_ERC366) entre as datas de criação da moeda Real (01/07/1994)⁷ até datas mais recentes (20/06/2007) contendo apenas os dias comerciais (2504 dados). É importante notar que todas as cotações são feitas com o preço de fechamento (final) do dia e não o valor médio diário, ou seja, são desconsiderados os finais de semanas e os feriados. A Figura 4.15 ilustra a série temporal financeira.

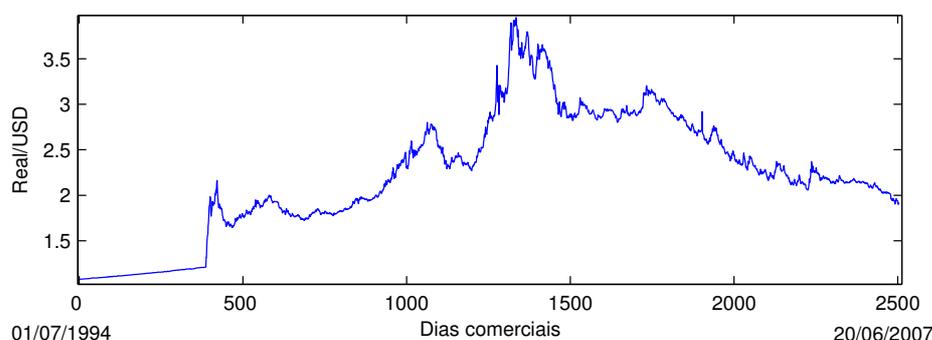


Figura 4.15: Série temporal financeira da taxa diária de câmbio Real/USD entre 01/07/1994 à 20/06/2007.

⁷As primeiras medidas de estabilização da economia que levaram ao Plano Real foram tomadas em 1993, com a criação da URV - Unidade Real de Valor, um indexador de preços e salários que deveria acompanhar a evolução do mercado. Diferentemente dos planos anteriores, as novas medidas buscavam obter a estabilização sem usar recursos tradicionais como o congelamento de preços e salários. Durante um ano, a URV preparou a estabilização e a implantação da nova moeda. Em 01/07/1994, o ministro da Fazenda, Fernando Henrique, lançou o Plano Real, convertendo a URV na nova moeda, o Real.

O algoritmo 0 (IID, GAUSS e BOOT) e o algoritmo SS são aplicados para o teste de determinismo na série financeira propostas gerando 39 séries temporais sub-rogadas para cada algoritmo, respectivamente. A Figura 4.16 ilustra a série temporal financeira da taxa diária câmbial do Real/USD bem como as séries sub-rogadas geradas pelos algoritmo 0.

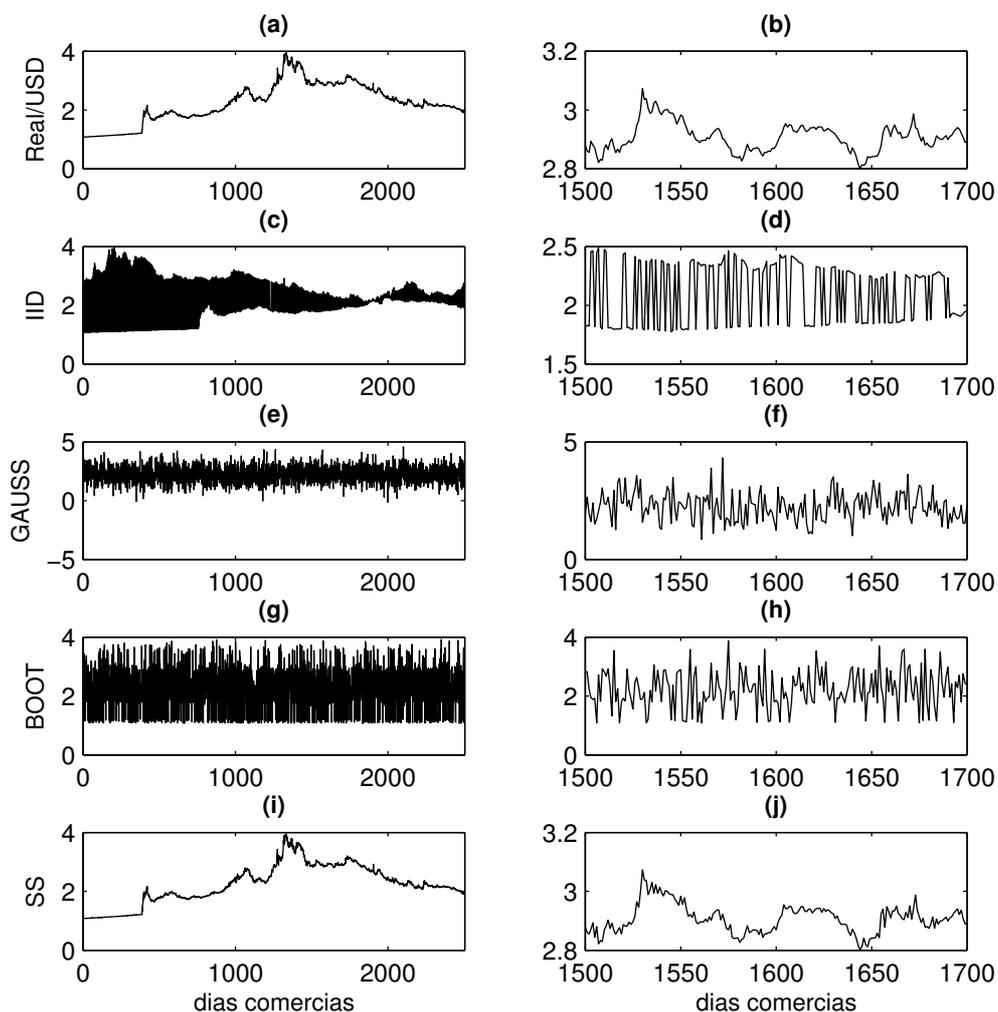


Figura 4.16: Série financeira da taxa diária de câmbio Real/USD entre 01/07/1994 à 20/06/2007: (a) série original e (b) janela da série original; (c) dado sub-rogado IID e (d) janela do sub-rogado IID; (e) dado sub-rogado GAUSS e (f) janela do sub-rogado GAUSS; (g) dado sub-rogado BOOT e (h) janela do sub-rogado BOOT; (i) dado sub-rogado SS e (j) janela do sub-rogado SS.

A Figura 4.17 ilustra o gráfico quantil-quantil, a FAC e a IMM da série temporal financeira (Real/USD) com os dados sub-rogados gerados. Os testes estatísticos paramétricos encontram-se no anexo B - Tabelas.

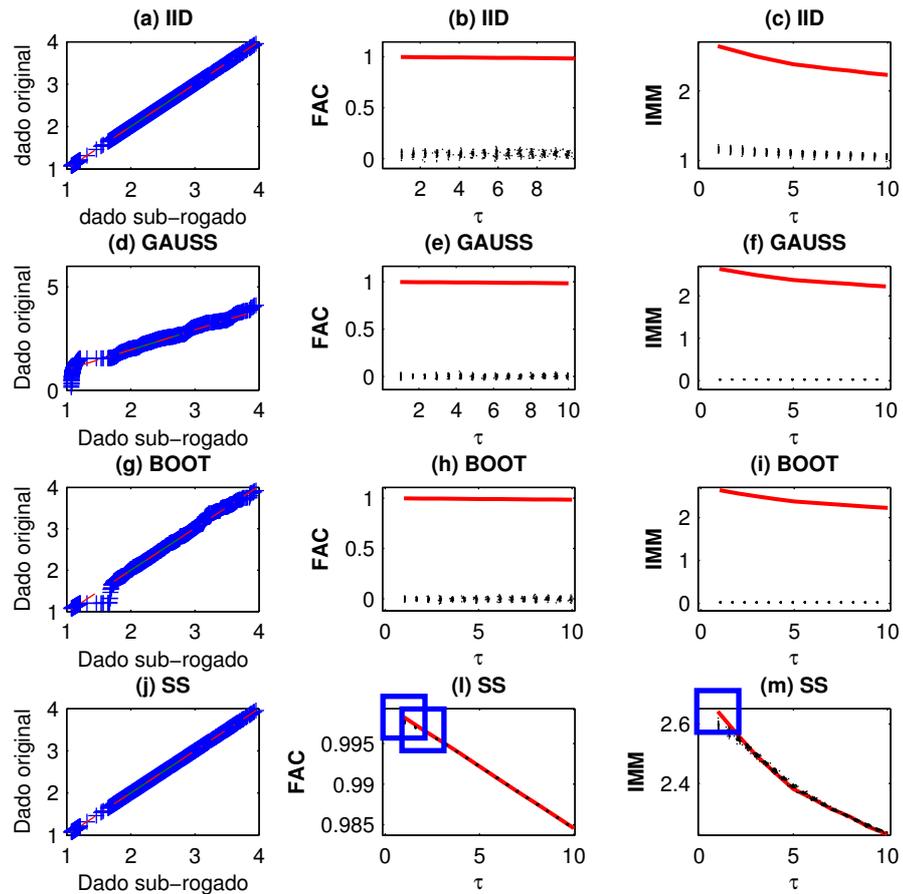


Figura 4.17: Gráfico QQ, da FAC e da IMM da taxa diária de câmbio Real/USD: (a), (b) e (c) para IID, (d), (e) e (f) para GAUSS, (g), (h) e (i) para BOOT, (j), (l) e (m) para SS. A linha sólida corresponde a série original e as linhas pontilhadas são as séries sub-rogadas geradas (39). As observações feitas em retângulos nas figuras são pontos ligeiramente fora da distribuição dos dados sub-rogados.

Note que as hipóteses nulas NH_0 e NH_{SS} são rejeitadas, concluindo assim que existe algum tipo de dinâmica determinística na série temporal da taxa diária cambial do Real/USD.

É importante lembrar que segundo Small et al. (2005) na aplicação da FAC e da IMM para o algoritmo SS, onde os dados sub-rogados gerados devem ser suficientemente semelhantes a série temporal original, a rejeição ou não da H_0 deve ocorrer nas duas estatísticas de testes não-paramétricas.

Pela análise dos testes estatísticos feitos é encontrado que as flutuações irregulares das séries temporais das variações de preços aqui utilizado para teste possuem algum tipo de comportamento dinâmico determinístico subjacente (rejeitam a hipótese nula - NH_0 e NH_{SS}).

Conseqüentemente, são considerados algumas implicações práticas desses resultados. Segundo (Nakamura e Small, 2005) os dados financeiros diários são considerados apenas pelos preços de fechamento. Quando as negociações do mercado diariamente está a ponto de terminar, os negociantes têm em mãos a maioria das informações do fluxo de mercado ao dia e também possuem algumas expectativas para os próximos negócios do mercado (dias seguintes ou semanas). Os negociantes finalizarão as compras ou vendas, ou não farão nada usando tais informações. Conseqüentemente, pode ser considerado que os preços finais refletem na maioria do fluxo de mercado daquele dia. Como o histórico dos dados de preço são coletâneas de tais dados, não é nenhuma surpresa esperar prever os preços futuros usando os históricos dos dados do fluxo do mercado. Conseqüentemente, é considerado que os resultados para o preço da barra de ouro diária em reais, da taxa diária de câmbio JPY/USD e da taxa diária de câmbio Real/USD reflitam isso.

4.4 Processo linear gaussiano - AR, MA ou ARMA

O caso mais simples citado anteriormente é da hipótese nula que os dados são de uma distribuição de probabilidade independente e identicamente (NH_0). Portanto, séries temporais sub-rogadas podem ser obtidas simplesmente embaralhando os dados colhidos aleatoriamente (algoritmo 0). Se forem encontradas correlações temporais significativamente diferentes na série e nos dados sub-rogados, **pode-se rejeitar a hipótese de que a série original seja i.i.d.**

O próximo passo seria explicar as estruturas encontradas pela autocorrelação linear entre os pontos dados. Uma hipótese nula correspondente é que os dados foram gerados por algum processo estocástico linear com incrementos gaussianos (ruídos). O mais generalizado modelo linear é o processo autoregressivo de média

móvel (ARMA), dado por,

$$x_k = a_0 + \sum_{i=1}^p a_i x_{k-i} + \sum_{i=1}^q b_i e_{k-i} \quad (4.1)$$

em que a_i e b_i são parâmetros fixos e e são incrementos aleatórios gaussianos não-correlacionados, ARMA(p,q).

A hipótese nula, neste caso, deve ser diferente, ou seja, a H_0 é que a série temporal é um processo gaussiano linearmente autocorrelacionado (NH_1). Conferi-se isto gerando uma série temporal para dois tipos de processos lineares simples:

- uma média móvel (MA),

$$x(k) = e(k) + ae(k - 1) \quad (4.2)$$

- e um auto-regressível AR(1),

$$x(k) = ax(k - 1) + e(k) \quad (4.3)$$

em que $e(k)$ é um ruído gaussiano com desvio padrão unitário.

O primeiro exemplo, de acordo com a equação 4.2, é dado pelo processo

$$x(k) = e(k) + 0,4e(k - 1) \quad (4.4)$$

onde $e(k)$ é um ruído gaussiano $N(0, 1)$.

Este teste procura desmentir, a princípio, o teste de hipótese NH_0 , ou seja, os dados observados não são ruídos independentes e identicamente distribuídos ou apenas independentemente distribuídos.

Aplica-se o teste de hipótese com o método de dados sub-rogados empregando o algoritmo 0, primeiramente, com as técnicas IID, GAUSS, BOOT e SS, respectivamente, e por último gera-se dados sub-rogados com o algoritmo 1. São produzidos 39 dados sub-rogados para o teste. A série possui 10000 pontos e os primeiros 1000 dados são descartados para evitar estados transientes.

A Figura 4.18 é mostrada a densidade espectral da série temporal original e dos dados sub-rogados (apenas um série entre as 39 geradas é ilustrada) gerados pelo algoritmo 1. Observe que na Figura 4.18 o algoritmo 1 preserva o espectro de potência da série original e também a distribuição de probabilidade, Figura 4.19. Todavia a distribuição de probabilidade não é preservada para séries com

distribuição não gaussianas, que será vista nas seções seguintes, uma vez que as séries sub-rogadas geradas terão distribuições diferentes da série original, isso será visto nos próximos exemplos.

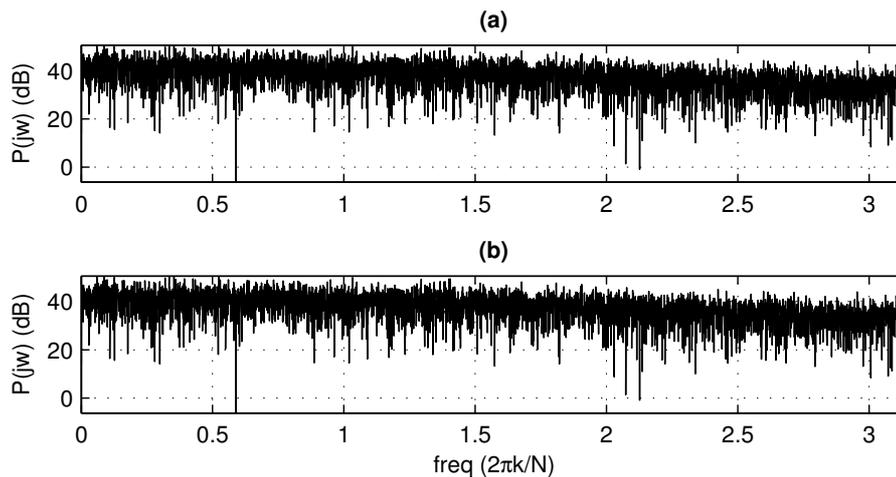


Figura 4.18: Densidade espectral para o modelo MA: (a) série original, (b) série sub-rogada 1 (FT).

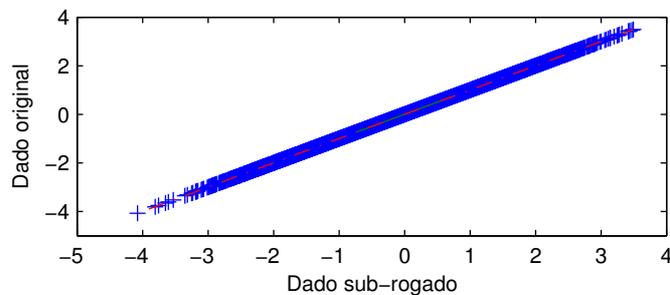


Figura 4.19: Gráfico quantil-quantil MA: série temporal original vs. 1 das séries sub-rogadas geradas pelo algoritmo 1.

Os resultados obtidos nos testes paramétricos para o algoritmo 0 (IID, GAUSS e BOOT), algoritmo SS e algoritmo 1 são apresentados na Tabela 4.2, com um nível de significância α de 5%. Nem todos os resultados obtidos nos testes paramétricos mostraram-se condizentes para com o modelo com correlação linear, sendo seus resultados contestados e não confiáveis, a princípio. Devido, aos resultados variáveis para os testes, não sendo possível chegara a alguma tomada de decisão.

Os testes paramétricos com o algoritmo FT identificaram estatisticamente algum tipo de dinâmica, não rejeitando NH_1 . Nesse caso a série temporal original é possivelmente (95%) um ruído gaussiano linearmente correlacionado. A não rejeição da H_0 pelo algoritmo 0 (IID, GAUSS e BOOT) e algoritmo SS implica erroneamente de que a série temporal original possui 95% de ser um ruído i.i.d. ou i.d. Portanto, pode-se ressaltar que o algoritmo 0 (IID, GAUSS e BOOT) e algoritmo SS nem sempre são indicados para processos lineares gaussianos com análise feitas apenas nas estatísticas de testes paramétricas, devem ser aplicadas outras estatísticas de teste, tais como a FAC e IMM (não-paramétricas).

Tabela 4.2: H_0 associada à realização dos testes estatísticos e sua respectiva probabilidade com 39 dados sub-rogados gerados.

Técnica	Estatística de teste	S	Probabilidade	Decisão H_0
IID	Reversão temporal	0,33979	0,73402	não rejeita
	TPA	-0,33252	0,73949	não rejeita
GAUSS	Variância	-0,25235	0,80077	não rejeita
	Assimetria	0,086022	0,93145	não rejeita
	Kurtoses	0,065681	0,94763	não rejeita
	C_6	-0,21455	0,83012	não rejeita
	Reversão temporal	0,13939	0,88914	não rejeita
	TPA	-0,59551	0,5515	não rejeita
BOOT	Variância	-0,19724	0,84364	não rejeita
	Assimetria	0,052561	0,95808	não rejeita
	Kurtoses	0,11352	0,90962	não rejeita
	C_6	-0,065706	0,94761	não rejeita
	Reversão temporal	0,48139	0,63024	não rejeita
	TPA	-0,31478	0,75293	não rejeita
SS	Reversão temporal	-0,16502	0,86893	não rejeita
	TPA	0,02395	0,98089	não rejeita
FT	Variância	-0,45513	0,64901	não rejeita
	Assimetria	0,39278	0,69448	não rejeita
	Kurtoses	0,37676	0,70635	não rejeita
	C_6	-0,89499	0,37079	não rejeita
	Reversão temporal	0,45914	0,64613	não rejeita
	TPA	-1,003	0,31586	não rejeita

A Figura 4.20 ilustra a FAC e a IMM da série temporal original MA e dos 39 dados sub-rogados gerados pelo algoritmo 0 (IID, GAUSS, BOOT), algoritmo

SS e algoritmo 1 (FT), respectivamente. Observe que na Figura 4.20(i) e 4.20(j) a H_0 não pode ser rejeitada, pois a FAC e a IMM da série original caem dentro da distribuição dos dados sub-rogados gerados indicando corretamente que a série temporal seja estatisticamente (95%) um ruído gaussiano linearmente correlacionado.

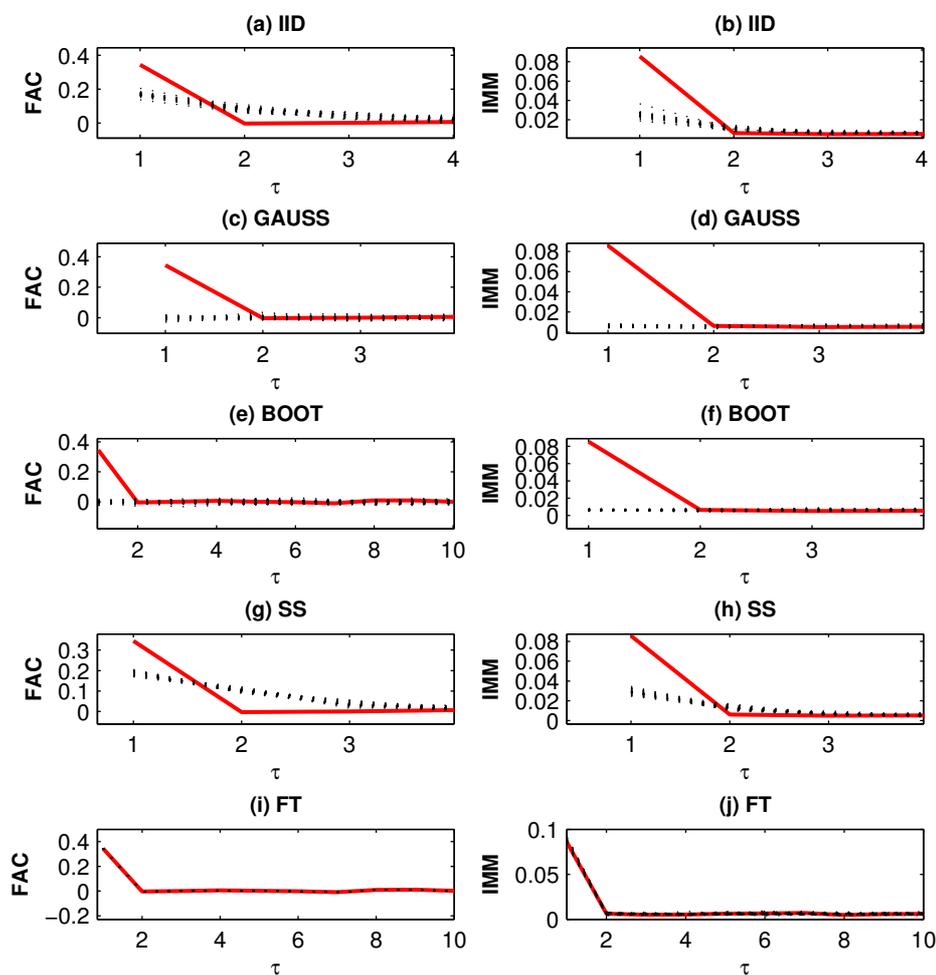


Figura 4.20: Gráfico da FAC e da IMM para o MA: (a) e (b) IID, (c) e (d) GAUSS, (e) e (f) BOOT, (g) e (h) SS, (i) e (j) FT. A linha sólida é a FAC da série original e as linhas pontilhadas são as 39 FAC das séries sub-rogadas geradas.

É importante deixar bem claro, a dificuldade em se usar, a princípio, o algoritmo FT proposto por Theiler e Prichard (1994), pois a rejeição da hipótese nula pode estar associada a não admissão de que a série temporal original seja linear ou que ainda esta seja um ruído i.i.d, ruído i.i.d. gaussiano ou ruído i.d.

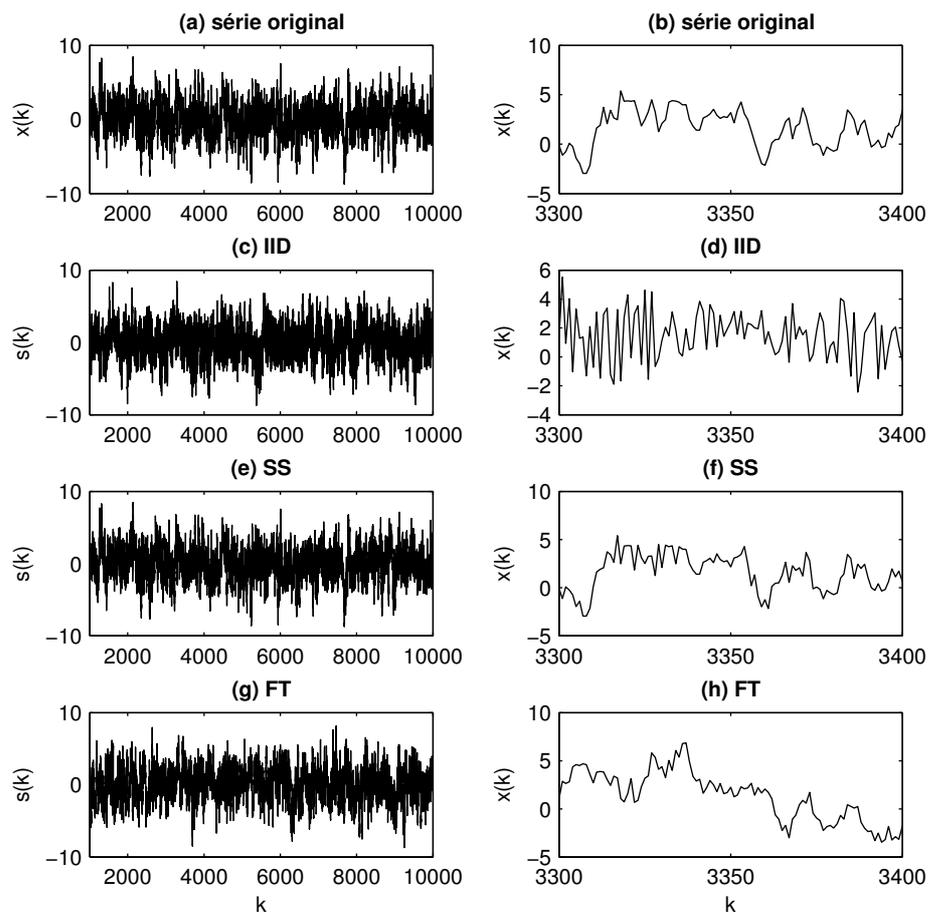


Figura 4.21: Séries temporais: (a) AR(1) e (b) janela do AR(1), (c) sub-rogado IID e (d) janela do sub-rogado IID, (e) sub-rogado SS e (f) janela do sub-rogado SS, (g) sub-rogado FT e (h) janela do sub-rogado FT.

O segundo processo proposto, conforme Equação 4.5, é um processo AR(1)

$$x(k) = 0.9x(k - 1) + e(k) \quad (4.5)$$

em que $e(k)$ é um ruído gaussiano $N(0,1)$.

Para a série temporal do modelo AR(1) são realizadas 10000 iterações, as primeiras 1000 são descartadas para evitar estados transientes e apenas os 9000 pontos restantes são tomados para teste, Figura 4.21. O teste de hipótese é aplicado gerando 39 ($\alpha = 0,05$) dados sub-rogados com o algoritmo 0 (IID), o algoritmo SS e o algoritmo 1, respectivamente.

Na Tabela 4.3 é apresentado os resultados da estatística de teste paramétrica para o modelo proposto. Novamente, nem todos os resultados obtidos nos testes estatísticos paramétricos para H_0 mostraram-se condizentes com o modelo com correlação linear AR(1), mais especificamente o algoritmo 0 (IID) e algoritmo SS cujas falhas, não rejeição da H_0 , foram relatadas no modelo anterior MA.

Tabela 4.3: H_0 associada à realização dos testes estatísticos e sua respectiva probabilidade com 39 dados sub-rogados gerados.

Técnica	Estatística de teste	S	Probabilidade	Decisão H_0
IID	Reversão temporal	-0.66595	0.50544	não rejeita
	TPA	1.4169	0.15652	não rejeita
SS	Reversão temporal	0.72948	0.46571	não rejeita
	TPA	0.49827	0.61829	não rejeita
FT	Variância	0.21138	0.83259	não rejeita
	Assimetria	-1.5361	0.12452	não rejeita
	Kurtoses	1.8145	0.069604	não rejeita
	C_6	1.7073	0.087758	não rejeita
	Reversão temporal	-1.2896	0.19718	não rejeita
	TPA	0,013402	0.98931	não rejeita

Por outro lado, os testes estatísticos paramétricos na Tabela 4.3 com os dados sub-rogados FT não podem ser rejeitados, comprovando que a série temporal é estatisticamente (95%) um processo linear gaussiano (NH_1).

O algoritmo 1 basicamente conserva o espectro de potência, mas a distribuição da série original não é preservada⁸ como mostra a Figura 4.22.

⁸Apenas um dado sub-rogado entre 39 é utilizado para a Figura 4.22(a). Lembrando que o gráfico gerado não é uma linha reta, ou seja, a distribuição do dado sub-rogado FT é ligeiramente diferente da original.

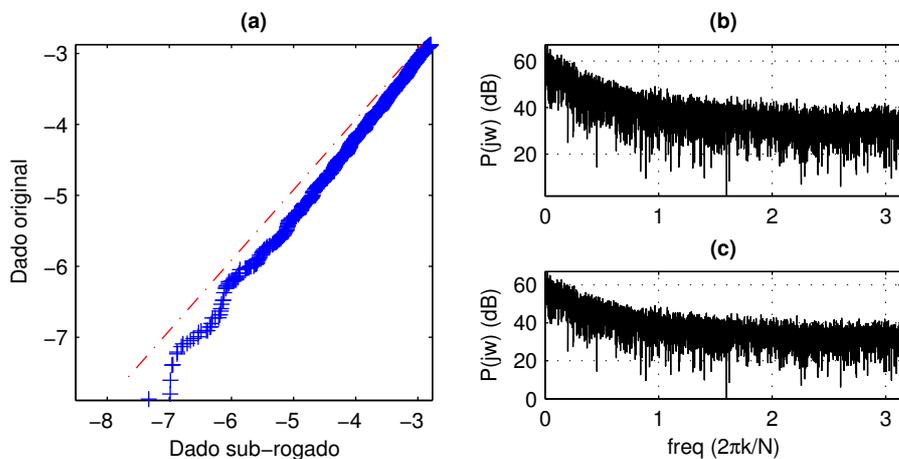


Figura 4.22: Gráfico quantil-quantil e densidade espectral do AR(1): em (a) série temporal original vs. 1 das séries sub-rogadas FT, em (b) densidade espectral original AR(1) e (c) densidade espectral da série sub-rogada FT.

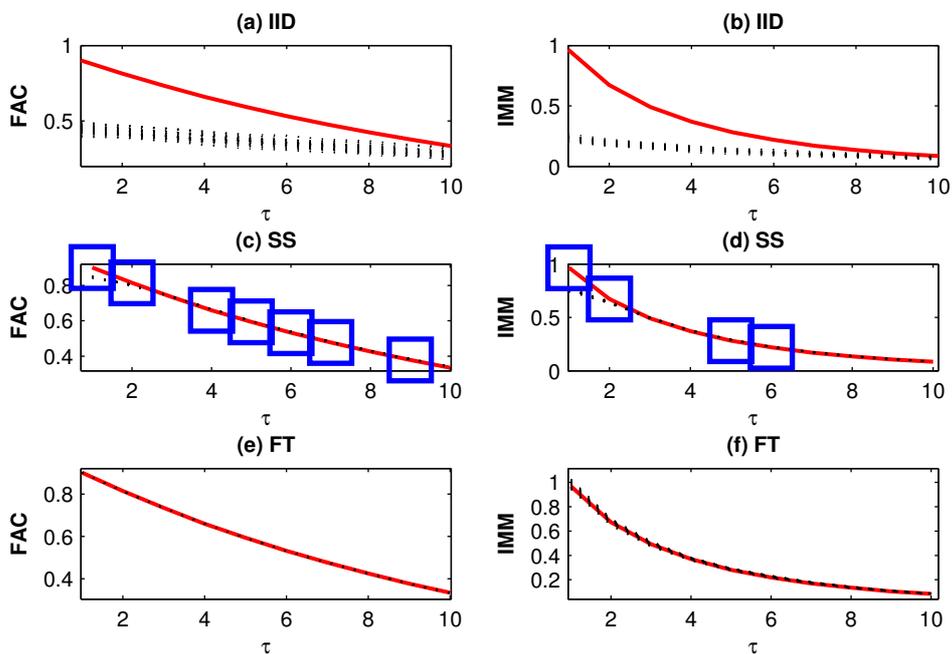


Figura 4.23: Gráfico da FAC e da IMM para AR(1): (a) e (b) para IID, (c) e (d) para SS, (e) e (f) para FT. A linha sólida é a série original e as linhas pontilhadas são as 39 séries sub-rogadas geradas. As observações feitas com retângulos nas figuras são pontos ligeiramente fora da distribuição dos dados sub-rogados.

A Figura 4.23 apresenta os resultados não-paramétricos (FAC e IMM) para o conjunto de dados da série temporal AR(1). A Figura 4.23(a) e 4.23(b) mostram que a FAC e IMM são distintas e caem fora das distribuições dos dados sub-rogados, conseqüentemente, pode-se rejeitar a hipótese NH_0 , consistindo assim que a série temporal original não seja um ruído i.i.d. Tal rejeição, conclui que a série temporal possui possivelmente (95%) algum tipo de dinâmica determinística subjacente.

Na Figura 4.23(c) e 4.23(d), referente aos dados sub-rogados SS, a FAC e IMM são distintas e caem fora das distribuições dos dados sub-rogados. Conseqüentemente, a hipótese NH_{SS} é rejeitada, considerando que a série temporal original **não seja** possivelmente uma variável aleatória i.d., enfatizando a existência de algum tipo de dinâmica determinística.

A FAC e IMM da Figura 4.23(e) e 4.23(f) para o FT não podem ser rejeitadas, pois a série original encontra-se dentro da distribuição dos dados sub-rogados. Sendo que tal hipótese nula (NH_1), não rejeitada (95%), é condizente com uma série temporal com correlações lineares (AR ou ARMA), ou seja, os dados sub-rogados são gerados por um processo gaussiano linear.

Um outro modelo linear ARMA(6,1) é proposto para análise. Tal modelo é dado por

$$x(k) = a_1x(k-1) + a_6x(k-6) + b_1e(k) + b_2e(k-1), \quad (4.6)$$

em que $a_1 = 0.3$, $a_6 = 0.2$, $b_1 = -1$, $b_2 = 0.5$ e $e(k)$ é um ruído gaussiano com desvio padrão unitário (Small e Judd, 1999).

O modelo ARMA(6,1) possui 10000 interações e as primeiras 1000 são descartadas para evitar estados transitórios. Apenas os 9000 pontos restantes são tomados para teste. São gerados 39 dados sub-rogados para o teste de hipótese nula usando o algoritmo 0 (IID), algoritmo 1 (FT) e algoritmo 2 (AAFT, IAAFT-1 e IAAFT-2), respectivamente, ilustrados na Figura 4.24.

O algoritmo 2 é aplicado a partir desse exemplo para os demais problemas que são propostos neste trabalho.

A técnica IAAFT do algoritmo 2 não faz nenhuma suposição para a forma da transformação h (principal idéia do algoritmo 2), isto é, pela construção, a técnica IAAFT pode representar a hipótese nula não obstante a formulação de h , quando a técnica AAFT não puder ser representada, ou seja, quando h não for monotônico.

A série temporal sub-rogada IAAFT-2 tem exatamente a mesma distribuição de amplitude que a série temporal original, mas as discrepâncias no espectro de potência são prováveis em ocorrer. Por outro lado, procura-se melhorar o espectro

de potência deixando as discrepâncias para a distribuição de amplitude, chamado de IAAFT-1.

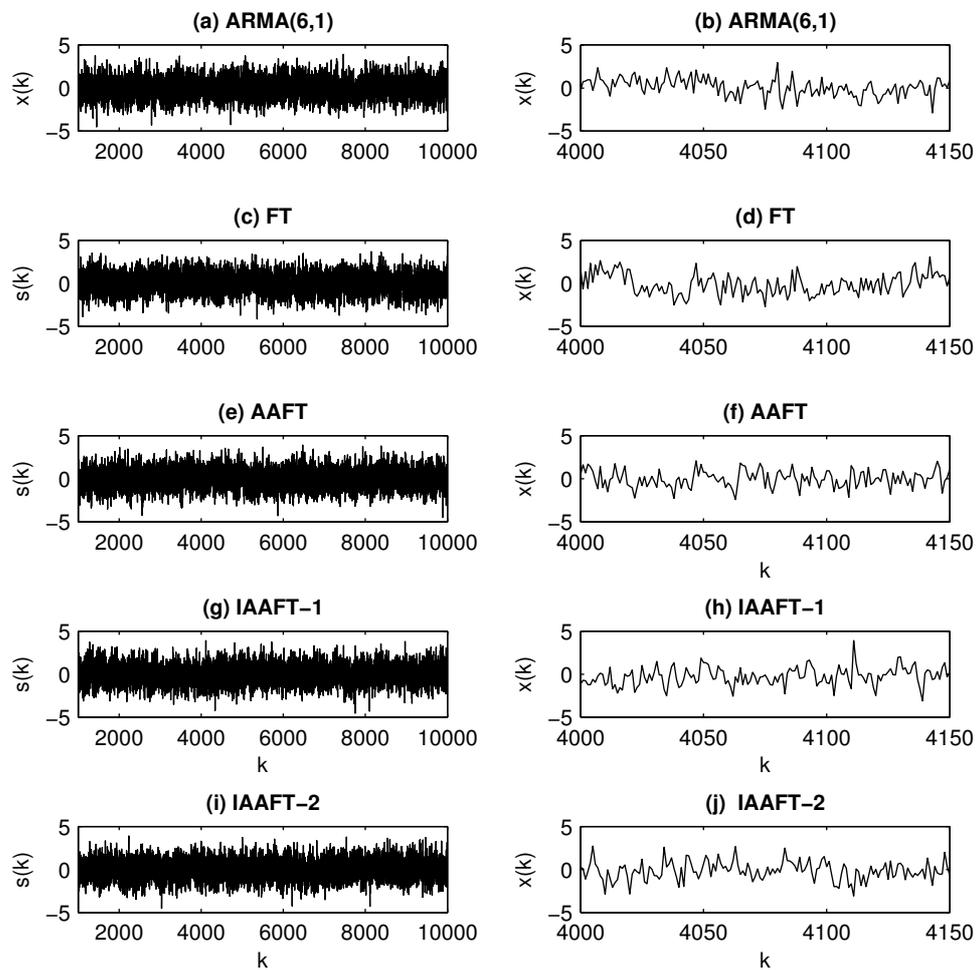


Figura 4.24: Séries temporais geradas a partir do ARMA(6,1): em (a) ARMA(6,1) e (b) janela do ARMA(6,1), (c) sub-rogado FT e (d) janela do sub-rogado FT, (e) sub-rogado AAFt e (f) janela do sub-rogado AAFt, (g) sub-rogado IAAFT-1 e (h) janela do sub-rogado IAAFT-1, (i) sub-rogado IAAFT-2 e (j) janela do sub-rogado IAAFT-2.

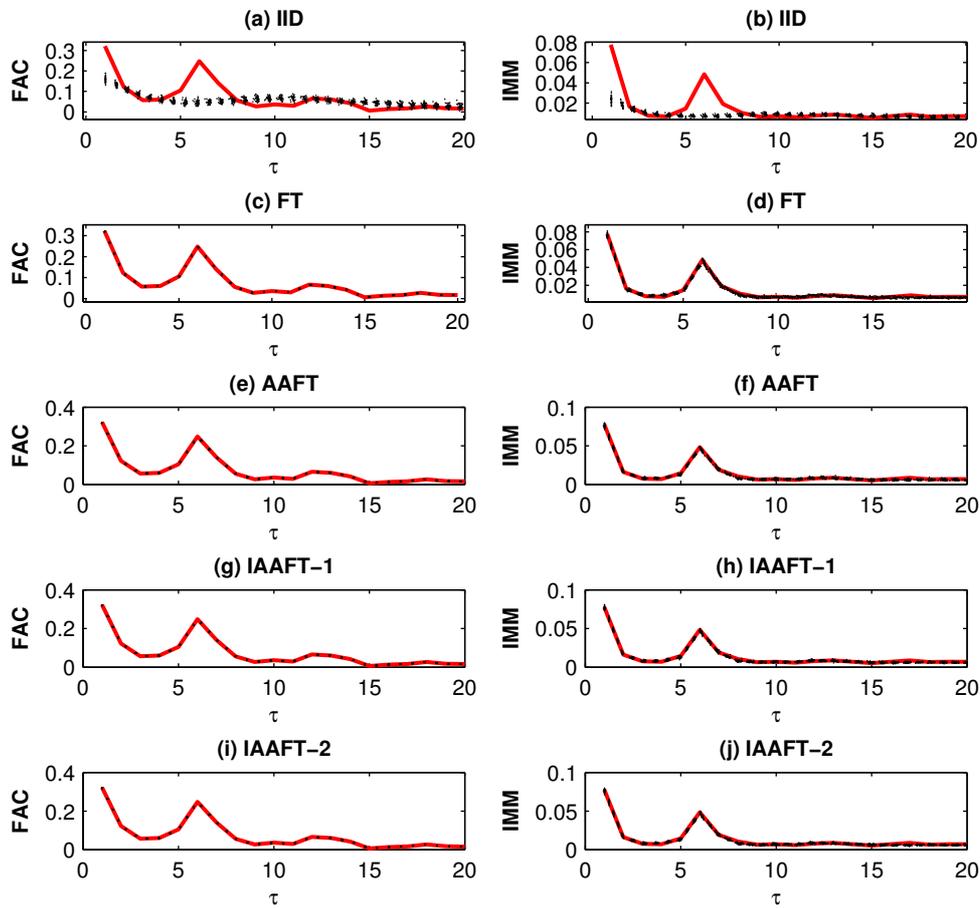


Figura 4.25: Gráfico da FAC e da IMM: (a) e (b) para IID, (c) e (d) para FT, (e) e (f) para AAFT, (g) e (h) para IAAFT-1, (i) e (j) para IAAFT-2. A linha sólida corresponde a série original e as linhas pontilhadas são as 39 séries sub-rogadas geradas.

A Figura 4.25 apresenta os resultados para o ARMA(6,1), cujas FACs e IMM não revelam nenhuma diferença significativa para os algoritmos 1 e algoritmo 2 (AAFT, IAAFT-1 e IAAFT-2), devido a FAC e IMM original cair dentro das distribuições dos dados sub-rogados gerados, menos para o algoritmo IID na Figura 4.25(a) e 4.25(b), uma vez que a rejeição assume que a série original não seja um possível (95%) ruído i.i.d.

Conseqüentemente, conclui-se que existe possivelmente (95%) algum tipo de

dinâmica proveniente de um processo estocástico linear gaussiano correlacionado (não rejeição NH_1) ou que os dados sub-rogados são gerados por uma transformação estática não-linear monotônica de um processo gaussiano linear da série temporal original, devido a não rejeição da NH_2 .

A monotonicidade de h é considerada como uma condição prévia para a implementação do teste e do algoritmo proeminente dos dados sub-rogados AAFT. Porém, não se pode afirmar a monotonicidade h ao lidar com dados reais. Essa limitação ou dificuldade do AAFT sob a não-monotonicidade é superado pelos algoritmos modificados apropriadamente para não-monotônica transformação, tais como, IAAFT-1 e 2, o CAAFT e o STAP.

Kugiumtzis discute esta limitação e propõe dois algoritmos: CAAFT - *corrected* AAFT (Kugiumtzis, 2000a) e STAP - *statically transformed autoregressive process* (Kugiumtzis, 2002b).

O algoritmo CAAFT é usado em vez do AAFT para qualquer transformação h . Assim, pelo uso de CAAFT estabelece a aplicabilidade do teste em aplicações reais.

Um conceito importante em nossa aproximação é que para qualquer série temporal, pode-se construir uma série temporal normal que sob uma transformação monotônica possui uma c.d.f. exatamente como a série original e também suas correlações lineares originais em comum.

O algoritmo STAP foi desenvolvido em cima do raciocínio de algum modo mais profundo do algoritmo CAAFT. Formulando assim uma aproximação conceitual para a geração dos dados sub-rogados consistentes com NH_2 , que se resolve então analiticamente. O novo algoritmo, chamado STAP, gera dados sub-rogados com realizações de um apropriado processo auto-regressivo estaticamente transformado (STAP), ou seja, o processo sob a H_0 é projetado como uma transformação estática de um processo normal apropriado.

A Figura 4.26 apresenta a série temporal do modelo ARMA(6,1) e dos dados sub-rogados (apenas um entre os 39 gerados) CAAFT e STAP gerados, respectivamente.

Os dados sub-rogados CAAFT foram gerados com os parâmetros: $\tau_{max} = p = 20$ (em que τ_{max} é o máximo atraso para análise da autocorrelação e p é a ordem do modelo AR) e $k = 39$ (k é o número de repetições ou o número de candidatos AR para serem estimados).

Os dados sub-rogados STAP possuem os seguintes parâmetros: $m = 10$ (ordem do polinômio), $\tau_{max} = p = 20$ (em que τ_{max} é o máximo atraso e p é a ordem do modelo AR), $k = 39$ (k é o número de repetições ou o número de candidatos

AR para serem estimados).

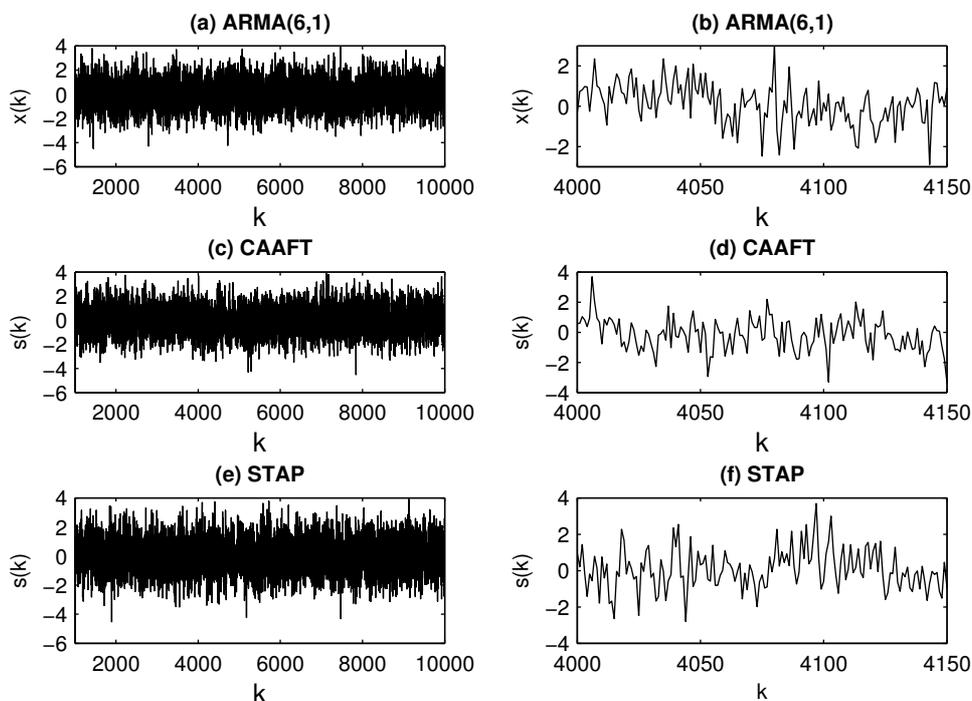


Figura 4.26: Série temporal ARMA(6,1): em (a) ARMA(6,1) e (b) uma janela do ARMA(6,1), (c) sub-rogado CAAFT e (d) uma janela do sub-rogado CAAFT, (e) sub-rogado STAP e (f) uma janela do sub-rogado STAP.

A Figura 4.27 apresenta a FAC e IMM da série temporal original ARMA(6,1) e das 39 séries temporais sub-rogadas geradas pelo CAAFT e STAP. Note que, conforme as demais técnicas (AAFT, IAAFT-1 e IAAFT-2) empregadas no algoritmo 2, não se pode também rejeitar (95%) a hipótese nula NH_2 , ou seja, os dados sub-rogados foram gerados por uma transformação estática não-linear monotônica de um processo gaussiano linear (série temporal original).

É importante notar que na Figura 4.27 fica mais claro e amplamente visível⁹ de que a série temporal ARMA(6,1) encontra-se dentro da distribuição dos dados sub-rogados CAAFT e STAP gerados. Portanto, as estatísticas lineares e a não-lineares usando a técnica CAAFT e STAP do algoritmo 2 é frequentemente mais extensa (ocorre maior variação) que quando usado o AAFT, a menor variação é obtida com o IAAFT.

⁹maior variação da FAC e IMM

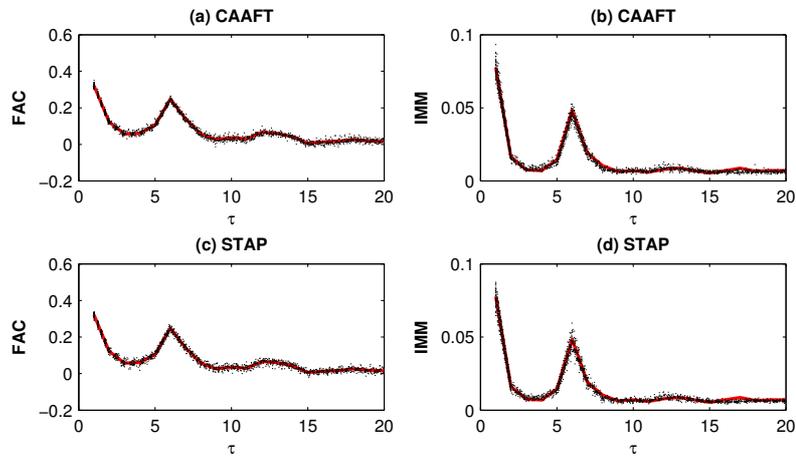


Figura 4.27: Gráfico da FAC e da IMM para a série temporal ARMA(6,1): em (a) e (b) para CAAFT e em (c) e (d) para STAP. A linha sólida corresponde a série original ARMA(6,1) e as linhas pontilhadas são as 39 séries sub-rogadas geradas.

Pode-se concluir que para as séries temporais propostas MA, AR(1) e ARMA(6,1) o uso das técnicas (IID, SS, FT, AAFT, IAAFT-1, IAAFT-2, CAAFT e STAP) mostraram-se favoráveis, de acordo com a hipótese nula empregada e a estatística de teste (FAC e IMM) não-paramétrica, de que existe estatisticamente algum tipo de dinâmica linear nas séries temporais em estudo.

A afirmação feita anterior para esse exemplo confirma a princípio a identificação de estruturas que são simplesmente processos gaussianos lineares, mas não é verificado a robustez dessa análise para casos não-gaussianos ou não-linearidades presentes nas séries. Na Seção 4.5 será tratado essas questões levantadas.

Antes de finalizar essa seção 4.4 é proposta a aplicação desse conjunto de dados sub-rogados para análise da série temporal financeira da taxa diária do câmbio Real/USD, Seção 4.3.

As Figuras 4.28, 4.29, 4.30, Figura 4.31 e 4.32 ilustram as estatísticas de teste FAC e IMM. Pela análise não-paramétrica (FAC e IMM) pode-se concluir que as FAC ou a IMM da série original caem fora da distribuição dos dados sub-rogados gerados, sendo assim, estatisticamente (95%) essa série temporal financeira possui algum tipo de dinâmica determinística, como demonstrado na seção anterior, mas possivelmente não-linear, pois rejeitam as hipóteses nulas de linearidade (NH_1 e NH_2).

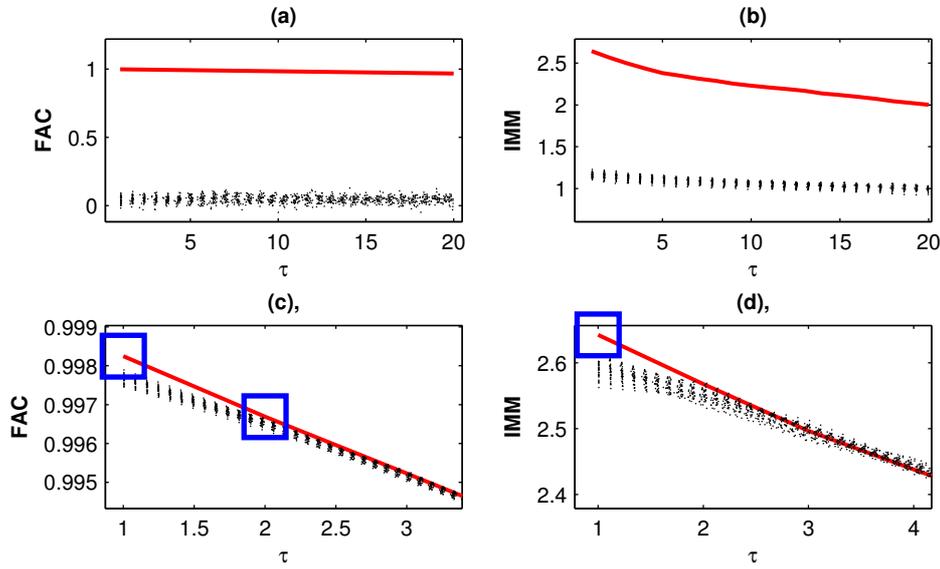


Figura 4.28: Gráfico da FAC e da IMM da série temporal financeira Real/USD: (a) e (b) para IID, (c) e (d) para SS. A linha sólida corresponde a série original e as linhas pontilhadas são as 39 séries sub-rogadas geradas. As observações feitas com retângulos nas figuras são pontos ligeiramente fora da distribuição dos dados sub-rogados.

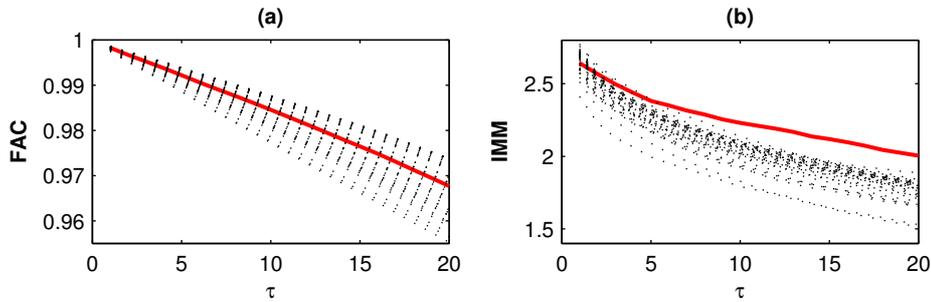


Figura 4.29: Gráfico da FAC e da IMM da série temporal financeira Real/USD: (a) e (b) para FT. A linha sólida corresponde a série original e as linhas pontilhadas são as 39 séries sub-rogadas geradas.

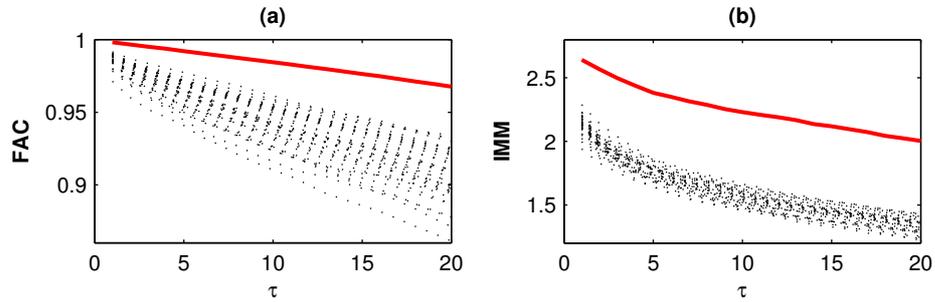


Figura 4.30: Gráfico da FAC e da IMM da série temporal financeira Real/USD: (a) e (b) para AAFT. A linha sólida corresponde a série original e as linhas pontilhadas são as 39 séries sub-rogadas geradas.

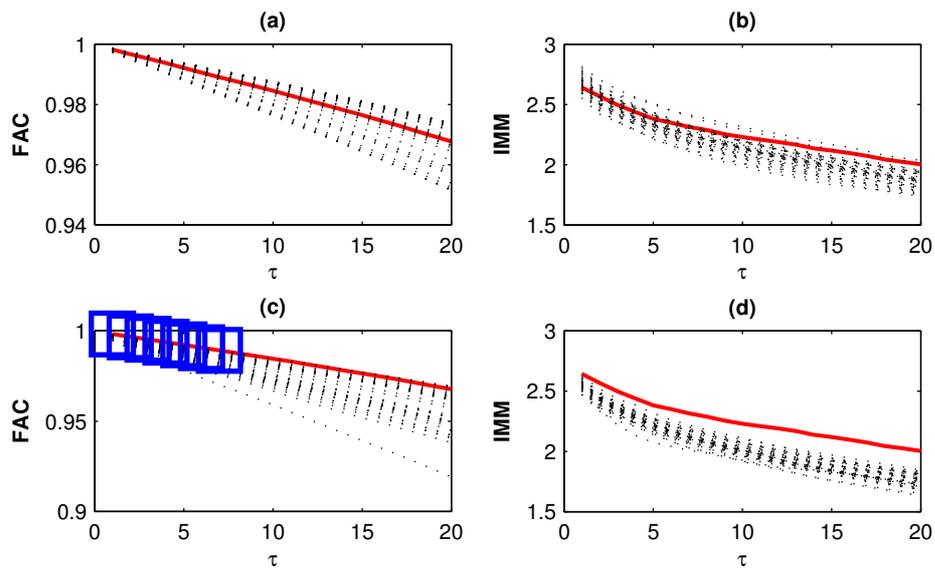


Figura 4.31: Gráfico da FAC e da IMM da série temporal financeira Real/USD: (a) e (b) para IAAFT-1, (c) e (d) para IAAFT-2. A linha sólida corresponde a série original e as linhas pontilhadas são as 39 séries sub-rogadas geradas. As observações feitas com retângulos nas figuras são pontos ligeiramente fora da distribuição dos dados sub-rogados.

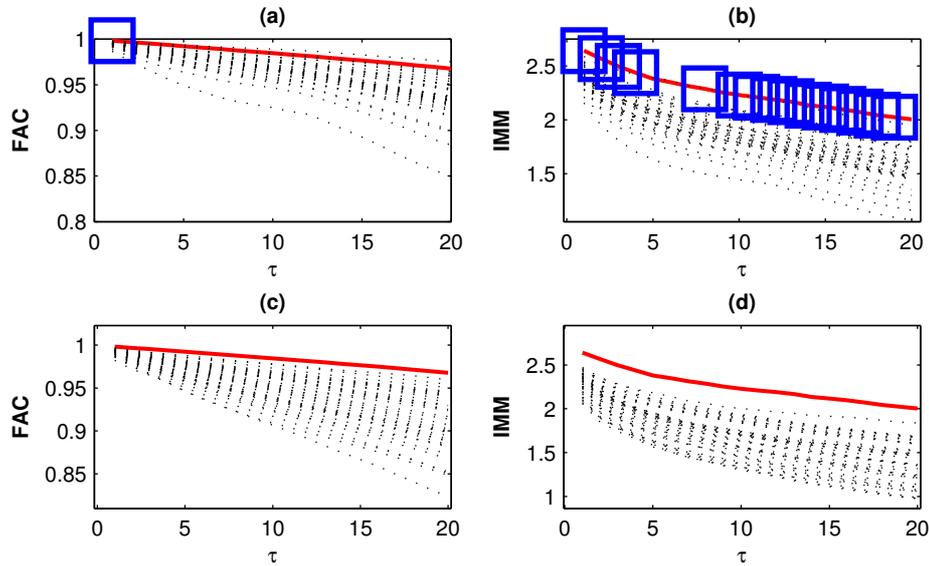


Figura 4.32: Gráfico da FAC e da IMM da série temporal financeira Real/USD: (a) e (b) CAAFT, (c) e (d) para STAP. A linha sólida corresponde a série original e as linhas pontilhadas são as 39 séries sub-rogadas geradas. As observações feitas com retângulos nas figuras são pontos ligeiramente fora da distribuição dos dados sub-rogados.

Um fluxograma é apresentado na Figura 4.33 seqüenciando o procedimento geral tomado nessa seção aplicado apenas para as estruturas MA, AR e ARMA com inovações gaussianas. Esse fluxograma identifica os possíveis resultados para esses casos.

Na seção seguinte será investigado os métodos de dados sub-rogados para processos lineares não-gaussianos.

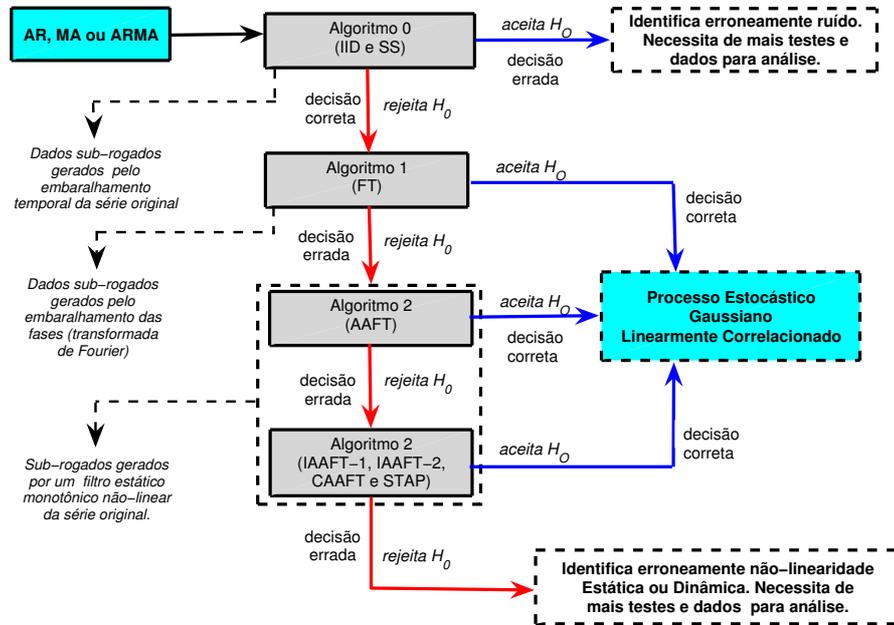


Figura 4.33: Fluxograma geral para teste de hipótese nula (H_0) utilizando os métodos de dados sub-rogados algoritmo 0 (IID e SS), algoritmo 1 (FT) e algoritmo 2 (AFT, IAAFT-1, IAAFT-2, CAAFT e STAP) aplicados somente a séries temporais com estruturas do tipo AR, MA ou ARMA.

4.5 Processo linear não-gaussiano - AR, MA ou ARMA

Uma pergunta adicional é esta: o que acontecerá se a série original em teste for linear, mas com condições de inovações (ruídos) assimétricas, ou seja não-gaussianas? Para responder esta pergunta é construído um exemplo de um modelo MA(1) com inovações que seguem uma distribuição não-gaussiana,

$$x_k = e_k - 0.4e_{k-1} \tag{4.7}$$

com e_k termos de inovações que segue a distribuição uniforme entre $[0,1]$.

São gerados 10000 pontos por meio da simulação do modelo. Os 1000 prime-

iros pontos são descartados, para evitar estados transientes. Um aspecto importante a ser notado é que o ruído no presente exemplo é não-gaussiano, conforme mostrado na Figura 4.34 que ilustra o histograma da série temporal MA(1) com inovações uniforme e_k no painel 4.34(a) e o histograma da distribuição uniforme e_k painel 4.36(b), que tem uma distribuição diferente da gaussiana.

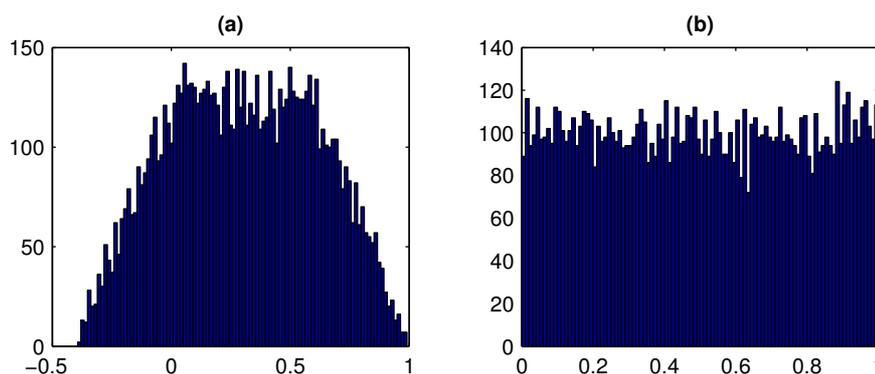


Figura 4.34: Série temporal MA(1) com inovações uniforme $[0,1]$: em (a) histograma da série original MA(1) e (b) histograma das inovações e_k com distribuição uniforme entre $[0,1]$.

Nesse exemplo deve-se investigar se uma determinada realização de um processo não-gaussiano pode falhar no teste de hipótese. A fim de investigação tais possíveis falhas, são gerados 99 dados sub-rogados pelo algoritmo 0 (IID), algoritmo 1, algoritmo 2 (AAFT, CAAFT e STAP).

A primeira conclusão é que se trata de um sistema determinístico, pois a hipótese nula NH_0 é rejeita, descartando estatisticamente (probabilidade de 99%) que os dados provem de um processo aleatório, veja Figuras 4.35(a) e (b). É importante observar as FAC das séries sub-rogadas e original na Figura 4.35, que demonstra que o algoritmo 1 (FT) e o algoritmo 2 (CAAFT e STAP) preservam o espectro de potência da série original. Porém, a técnica AAFT não preserva a FAC da série original, logo o algoritmo 2 (AAFT) não é indicado para aplicações em séries temporais contaminadas com ruídos não-gaussianos. Devido, por construção da técnica, ver Seção 3.5, gerar séries sub-rogados com comportamento linear gaussiano.

É importante observar que a FAC da série temporal original cai totalmente dentro da distribuição dos 99 dados sub-rogados gerados, o que não ocorre com as IMMs. Logo, encontra-se a primeira falha na análises onde não se pode rejeitar a hipótese nula NH_1 e NH_2 , observando somente as Figuras 4.35(a), (c), (g) e (i).

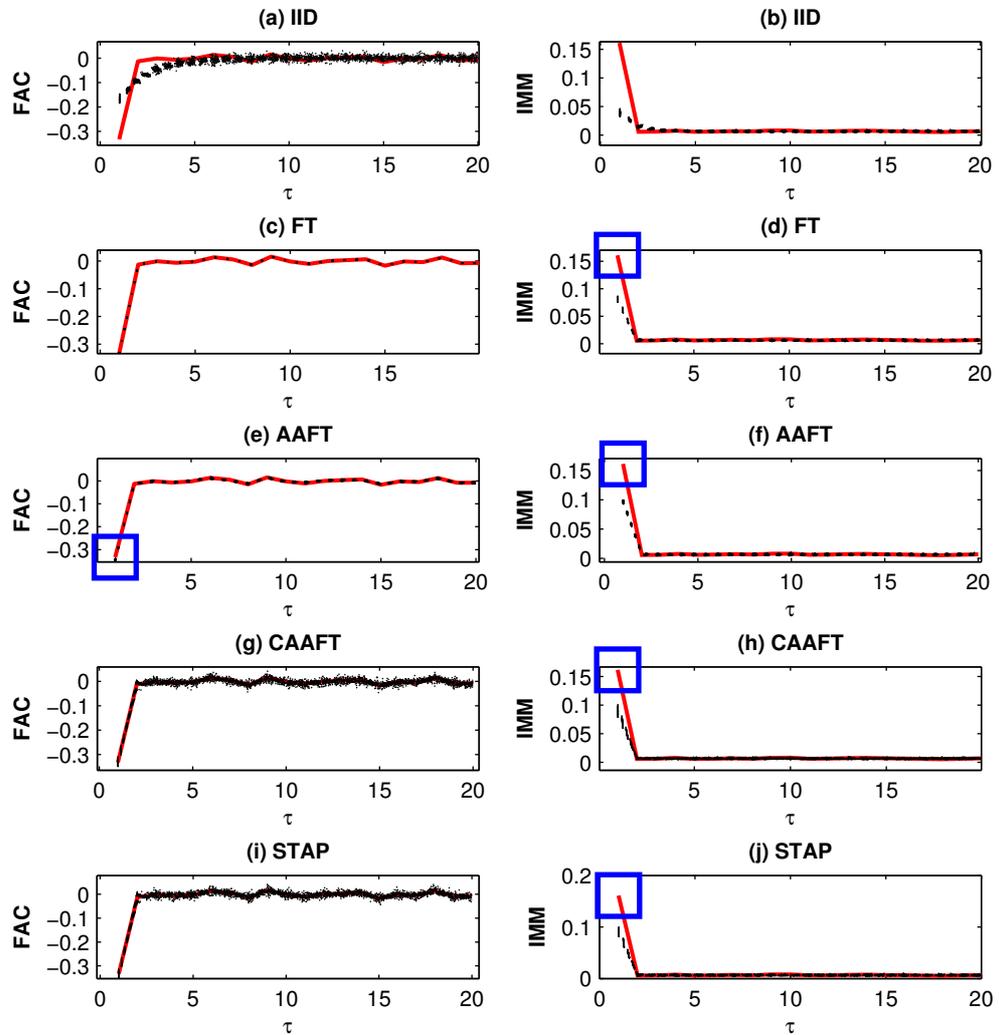


Figura 4.35: Gráfico da FAC e da IMM para o MA(1) com inovações uniforme [0,1]: (a) e (b) para IID, (c) e (d) para FT, (e) e (f) para AAFT, (g) e (h) para CAAFT, (i) e (j) para STAP. A linha sólida corresponde a série original e as linhas pontilhadas são as 99 séries sub-rogadas geradas. As observações feitas com retângulos nas figuras são pontos ligeiramente fora da distribuição dos dados sub-rogados.

Desta forma, pode-se concluir que existe possivelmente (99%) algum tipo de dinâmica subjacente, mas não se pode decidir com certeza por meio dessas análises se este comportamento é linear ou não-linear. Pode-se chegar a um resultado errôneo analisando de forma complementar a IMM, o que resultaria, segundo alguns autores (Small, 2005; Nagaranjan, 2005), indícios de alguma dinâmica não-linear, o que não é verdade de fato, pois se trata de um processo linear não-gaussiano, veja Equação (4.7).

É construído agora um outro modelo ARMA(1,1) com inovações que seguem uma distribuição uniforme,

$$x_k = 0.9x_{k-1} + e_k - 0.4e_{k-1} \quad (4.8)$$

com e_k termos de inovações que segue a distribuição uniforme entre $[0,1]$.

Para este modelo ARMA(1,1) com inovações uniforme são gerados 10000 pontos e os 1000 primeiros pontos são descartados, para evitar estados transientes.

Um aspecto importante a ser notado é que o ruído no presente exemplo é não-gaussiano, conforme mostrado na Figura 4.36(b). A Figura 4.36 ilustra o histograma da série temporal ARMA(1,1) com inovações uniforme e_k no painel 4.36(a) e o histograma da distribuição uniforme e_k no painel 4.36(b), que tem uma distribuição diferente da gaussiana. Observe que o histograma da Figura 4.36(a) vale o que diz o teorema do limite central (seção A.3.1 no anexo-A), alguns modelos (AR ARMA) convergem para uma distribuição normal.

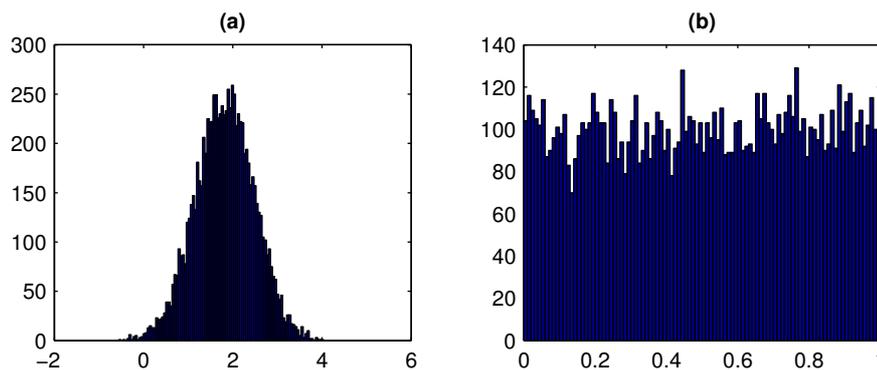


Figura 4.36: Série temporal ARMA(1,1) com inovações uniforme $[0,1]$: em (a) histograma da série original e (b) histograma das inovações e_k com distribuição uniforme entre $[0,1]$.

Com este exemplo deve-se investigar se uma determinada realização de um processo não-gaussiano pode falhar no teste de hipótese.

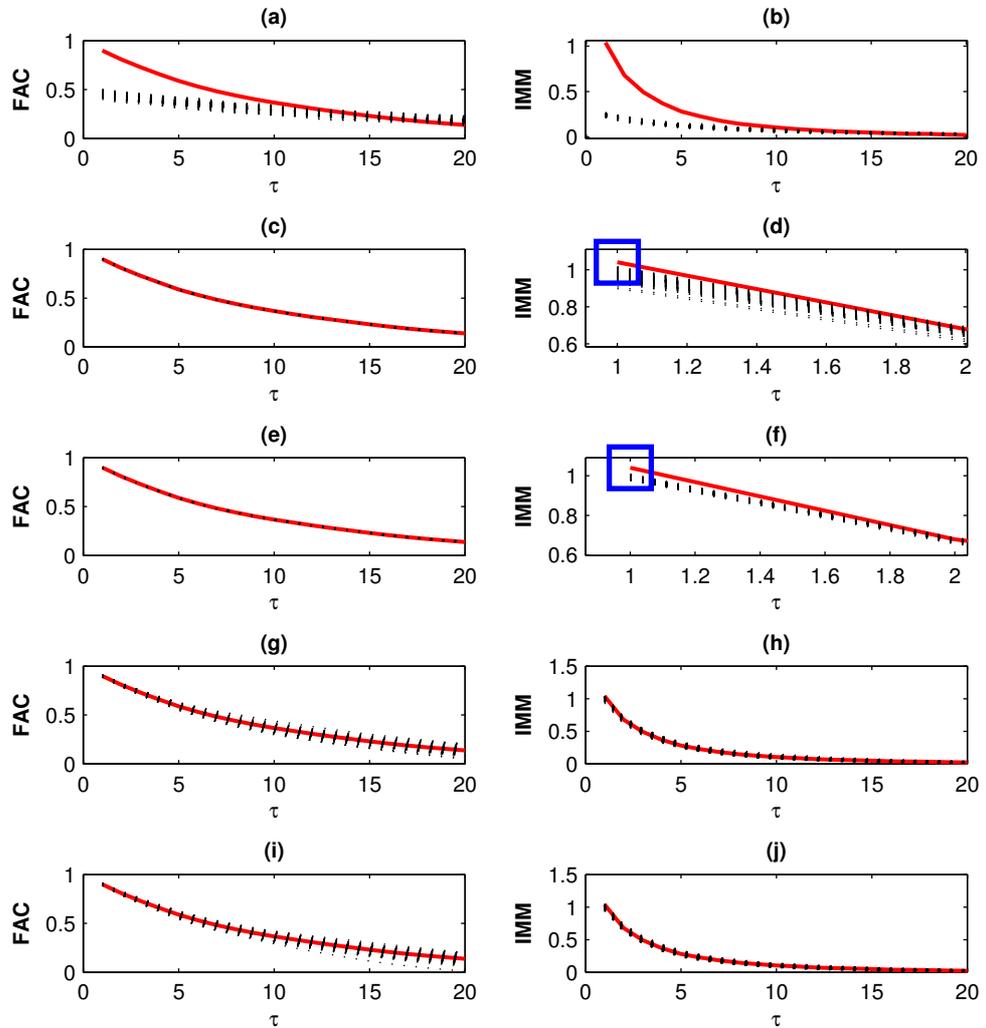


Figura 4.37: Gráfico da FAC e da IMM para o ARMA(1,1) com inovações uniforme [0,1]: (a) e (b) para IID, (c) e (d) para FT, (e) e (f) para AAFT, (g) e (h) para CAAFT, (i) e (j) para STAP. A linha sólida corresponde a série original e as linhas pontilhadas são as 99 séries sub-rogadas geradas. As observações feitas com retângulos são pontos ligeiramente fora da distribuição dos dados sub-rogados.

A fim de investigar tais possíveis falhas, são gerados 99 séries sub-rogadas pelo algoritmo 0 (IID), algoritmo 1, algoritmo 2 (AAFT, CAAFT¹⁰ e STAP¹¹).

É importante notar na Figura 4.37 a rejeição, painéis (a) e (b), do teste com o algoritmo 0 (IID) indica algum tipo de dinâmica determinística presente na série temporal. Também é observado a rejeição, painel (d), do teste com o algoritmo 1 (FT), pois o presente exemplo proposto não se trata de um processo linear gaussiano. E por fim, a não rejeição, painéis (g), (h), (i) e (j), do teste com o algoritmo 2, apenas com as técnicas CAAFT e STAP, nesse processo com inovações não-gaussianas. Observe que a série temporal original para o algoritmo 2 (CAAFT e STAP) cai totalmente dentro da distribuição dos 99 dados sub-rogados gerados, tanto para a FAC e a IMM, o que não ocorre com a técnica (AAFT). Lembrando que o AAFT é por construção¹² do algoritmo indício de um processo linear gaussiano.

Desta forma, pode-se concluir, por meio do CAAFT e STAP, que existe possivelmente (99%) alguma tipo de dinâmica linear subjacente, como a princípio exposta pela hipótese NH_2 , não necessariamente gaussiana¹³.

Um outro exemplo é um modelo ARMA(7,1) com inovações beta similar ao proposto por Small et al. (2006b),

$$x_k = a_6x_{k-1} + a_5x_{k-2} + a_4x_{k-3} + a_3x_{k-4} + a_2x_{k-5} + a_1x_{k-6} + a_0 + e_k - 0.3e_{k-1} \quad (4.9)$$

com e_k termos de inovações que segue a distribuição beta(2,5) e parâmetros $a_0 = 0.6$, $a_1 = 0$, $a_2 = 0.5$, $a_3 = 0$, $a_4 = -0.6$, $a_5 = 0.3$ e $a_6 = -0.1$. Com

$$beta(\alpha, \beta) \equiv \frac{1}{B(\alpha, \beta)} s^{\alpha-1} (1-s)^{\beta-1} \quad (4.10)$$

em que $B(\alpha, \beta)$ denota a função beta com $\alpha, \beta > 0$ (Tabela A.1 no Capítulo 2).

Neste exemplo são gerados 10000 dados, os primeiros 1000 pontos são descartados. A série temporal do modelo ARMA(7,1) com inovações beta (ruído beta) para teste é ilustrada na Figura 4.38. O ruído beta e_k proposto na Equação (4.9) é também não-gaussiano, conforme mostrado na Figura 4.39.

¹⁰Parâmetros do CAAFT: $\tau_{max} = p = 20$ e $k = 99$

¹¹Parâmetros do STAP: $m = 10$, $\tau_{max} = p = 20$ e $k = 39$.

¹²A transformação estática não-linear do AAFT é um re-escala ou uma normalização dos dados por uma gaussiana.

¹³Um resultado errôneo seria analisar séries temporais não-gaussianas por meio da IMM aplicando somente o algoritmo FT ou AAFT, pois resultaria, segundo Small (2005); Nagaranjan (2005), uma indícios de alguma dinâmica não-linear, o que não é verdade.

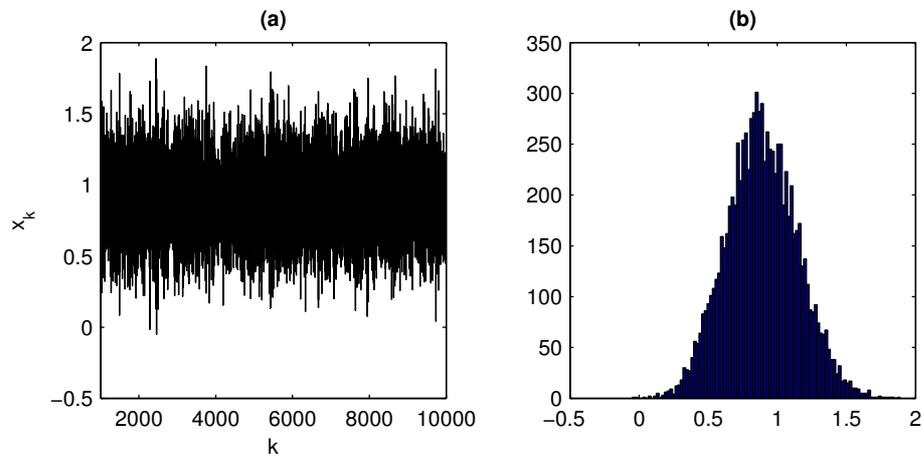


Figura 4.38: Série temporal em (a) do ARMA(7,1) com inovações que segue a distribuição beta(2,5) e (b) seu histograma da série temporal proposta.

É importante observar que a idéia *a priori* para os processos AR ou ARMA eram que estes tenderiam para uma distribuição gaussiana para um número muito grande de amostras, conforme exposto pelo TLC (seção A.3.1 no anexo-A). Contudo é observado que para certos processos com inovações não-gaussianas o TLC é verdadeiro apenas para casos de ordem elevada do modelo ARMA(p,q), pois o ruído (beta) é mascarado e a série temporal gerada semelha-se a um processo gaussiano.

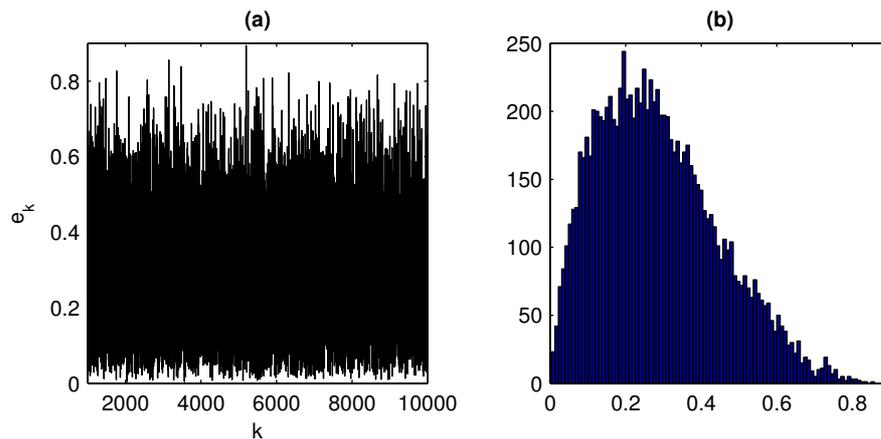


Figura 4.39: Ruído Beta: (a) série e_k das inovações com a distribuição beta(2,5) e em (b) o histograma da distribuição das inovações e_k beta.

Para investigar possíveis falhas na identificação de modelos ARMA(7,1) com

inovações não-gaussianas são gerados 99 séries sub-rogadas ($\alpha = 0,01$) utilizando o algoritmo 0 (IID), algoritmo 1, algoritmo 2 (AAFT, CAAFT e STAP). A Figura 4.40 mostra a análise das estatísticas de teste não-paramétrica da FAC e da IMM.

Deve ser notado que na Figura 4.40(a) e Figura 4.40(b) mostra a rejeição do teste de hipótese nula com o algoritmo 0 (IID), identificando estatisticamente algum tipo de dinâmica determinística.

Por outro lado, o teste de hipótese nula com o algoritmo 1 e algoritmo 2 (AAFT, CAAFT e STAP) não rejeitam as respectivas H_0 . Observe que a série temporal original cai dentro da distribuição dos dados sub-rogados gerados. Contudo, não se pode dizer o mesmo dos resultados obtidos, teoricamente errado, do teste com o algoritmo 1, Figura 4.40(c) e 4.40(d), e com o algoritmo 2 (AAFT), Figura 4.40(e) e 4.40(f), pois apesar da série temporal ser com inovações não-gaussianas têm uma semelhança com processos lineares gaussianos, conforme mostrado na Figura 4.38(b), quando o número de amostras (N) é grande e cujo modelo ARMA(p,q) é de alta ordem.

As técnicas CAAFT e STAP são para análise das séries temporais lineares quaisquer, ou seja, que podem ser geradas por um processo linear com ruído gaussiano ou não-gaussiano. Essa consideração também é válida para o algoritmo IAAFT-1 e IAAFT-2 que não são aplicados nesse exemplo, mas em outras seções adiante.

Sabendo do fato de que a alta ordem mascara o ruído não-gaussiano, pode-se concluir que existe alguma dinâmica linear subjacente e que os dados sub-rogados são gerados por uma transformação estática não-linear monotônica de um processo linear da série temporal original, aparentemente gaussiana.

Uma observação importante a princípio é que o TLC está correto para esse exemplo, devido a ordem elevado do modelo ARMA(p,q), pois sugere que a série temporal geradora seja um processo linear gaussiano, mas esse fato não ocorre em todas as séries temporais reais.

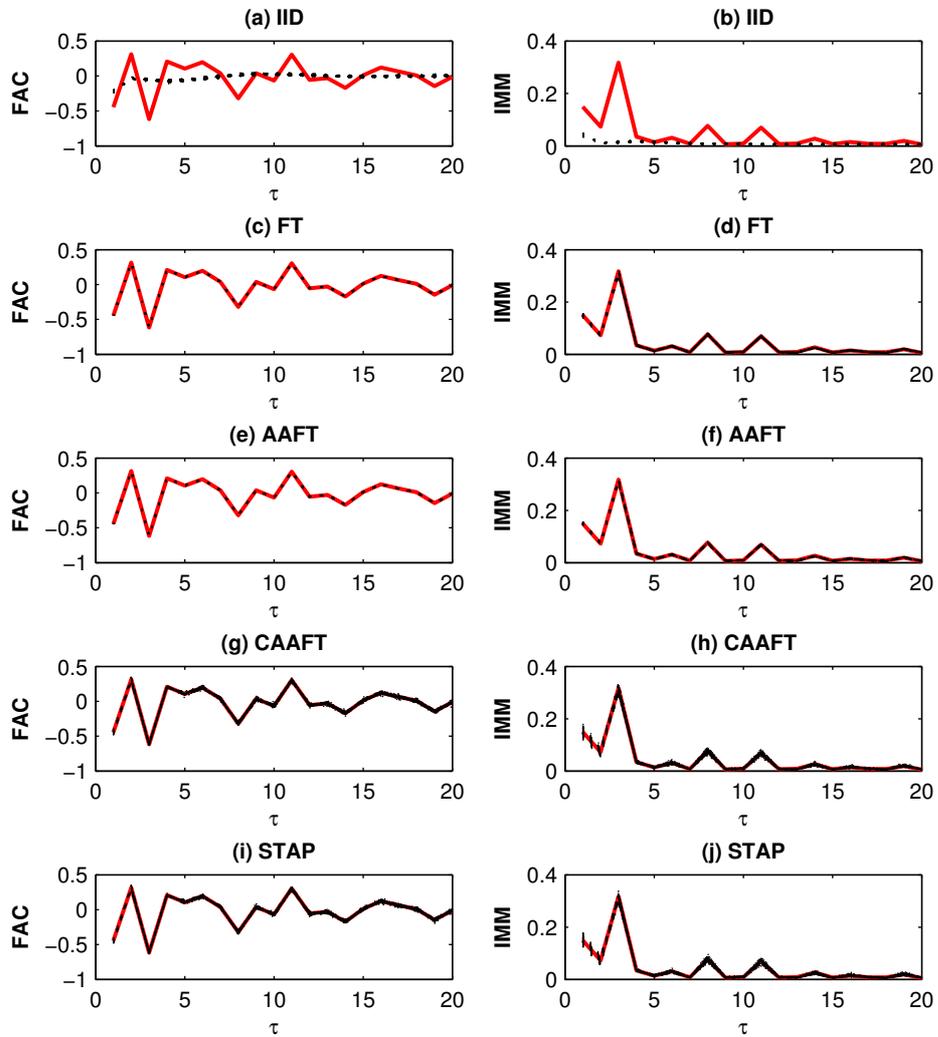


Figura 4.40: Gráfico da FAC e da IMM para o ARMA(7,1) com inovações beta: (a) e (b) para IID, (c) e (d) para FT, (e) e (f) para AAFT, (g) e (h) para CAAFT, (i) e (j) para STAP. A linha sólida corresponde a série original e as linhas pontilhadas são as 99 séries sub-rogadas geradas.

Um outro modelo com inovações ligeiramente não-gaussianas pode ser construído para teste. Seja o modelo ARMA(1,1) com inovações não-gaussianas dado por

$$x_k = 0.9x_{k-1} + e_k - 0.4e_{k-1}; \quad (4.11)$$

com e_k termos de inovações que segue a distribuição de Laplace (Tabela A.1) com parâmetros $\theta = 0$ e $\lambda = 1$ para uma distribuição de Laplace generalizada conforme,

$$f(x) = (y \leq \theta) \exp\left(\frac{y-\theta}{\lambda}\right) + (y > \theta) \left(1 - 0.5 \exp\left(\frac{y-\theta}{\lambda}\right)\right) \quad (4.12)$$

$$F(x) = \frac{1}{2\lambda} \exp\left(-\left|\frac{y-\theta}{\lambda}\right|\right) \quad (4.13)$$

Para o presente exemplo ARMA(1,1) com inovações de Laplace são realizadas 10000 interações e os primeiros 1000 pontos são descartados, para evitar estados transientes. A série temporal do modelo ARMA(1,1) com inovações de Laplace para teste é ilustrada na Figura 4.41. A Figura 4.42 mostra a série temporal das inovações e_k de Laplace geradas e um histograma de sua distribuição, muito próxima da distribuição gaussiana.

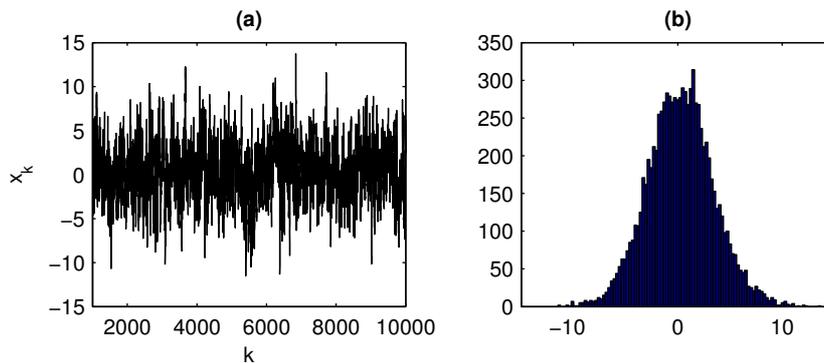


Figura 4.41: Série temporal em (a) do ARMA(1,1) com inovações que segue a distribuição de Laplace ($\theta = 0$ e $\lambda = 1$) e em (b) um histograma da série temporal obtida com as inovações propostas.

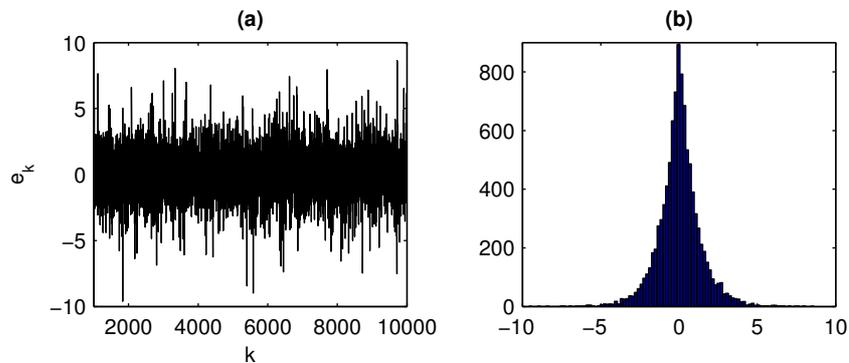


Figura 4.42: Ruído de Laplace: em (a) série temporal das inovações e_k de Laplace e em (b) o histograma da distribuição e_k de Laplace gerada.

São geradas 99 séries temporais sub-rogadas ($\alpha = 0,01$) pelo algoritmo 0 (IID) e algoritmo 2 (IAAFT-1, IAAFT-2, CAAFT e STAP), respectivamente. O sub-rogado CAAFT possui os seguintes parâmetros: $\tau_{max} = p = 20$ e $k = 99$. Os dados sub-rogados gerados pelo STAP possuem parâmetros: $m = 10$, $\tau_{max} = p = 20$ e $k = 99$. A Figura 4.43 ilustra o teste estatístico não-paramétrico da FAC e da IMM.

É importante notar que na Figura 4.43(a) e Figura 4.43(b) a hipótese nula NH_0 é rejeitada, comprovando a existência de dinâmica determinística na série temporal sobre análise.

Por outro lado, o algoritmo 2 (STAP e CAAFT) empregado não rejeita a hipótese nula NH_2 , mostrando que a série temporal original cai dentro das distribuições dos dados sub-rogados gerados, ilustrados na Figura 4.43(g), 4.43(h), 4.43(i) e 4.43(j). Portanto não se pode rejeitar a hipótese nula NH_2 .

Observe que na IMM para o algoritmo 2 (IAAFT-1 e IAAFT-2) rejeitam a hipótese nula NH_2 , Figuras 4.43(d) e (f). Essa rejeição causa uma dúvida na análise do resultado para tomada de decisão a favor ou não da hipótese nula NH_2 , não sendo possível afirmar se a dinâmica é linear ou não-linear. É proposto aqui para investigação se as técnicas CAAFT e o STAP serão mais acertivas ou não nas análises de resultados posteriores.

Por isso a princípio, pode-se concluir que existe alguma dinâmica determinística subjacente e que os dados sub-rogados construídos pelo algoritmo 2 (CAAFT e STAP) por meio de uma transformação estática não-linear monotônica de um processo linear da *série temporal original*, mas tal série temporal original não é necessariamente um processo linear com inovações gaussianas, ou seja, pode ser com inovações não-gaussianas.

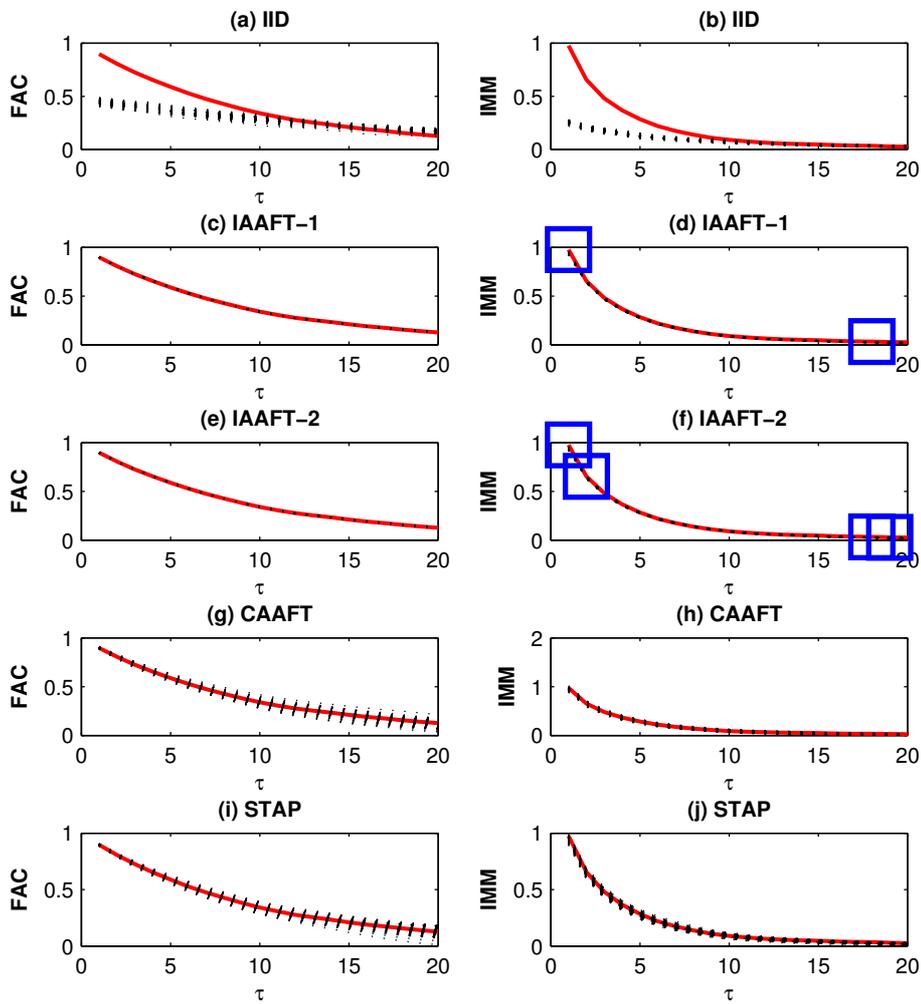


Figura 4.43: Gráfico da FAC e da IMM para o ARMA(1,1) com inovações de Laplace ($\theta = 0$ e $\lambda = 1$): (a) e (b) para IID, (c) e (d) para IAAFT-1, (e) e (f) para IAAFT-2, (g) e (h) para CAAFT, (i) e (j) para STAP. A linha sólida corresponde a série original e as linhas pontilhadas são as 99 séries sub-rogadas geradas.

Um outro modelo com inovações não-gaussianas pode ser construído para teste. Seja o modelo ARMA(1,1) com inovações não-gaussianas dado por

$$x_k = 0.9x_{k-1} + e_k - 0.4e_{k-1} \quad (4.14)$$

com e_k termos de inovações que segue a distribuição de Cauchy (Tabela A.1 no Anexo A) com parâmetros $t = 5$ e $s = 0,001$ para uma distribuição de Cauchy generalizada:

$$f(x) = 0.5 + \frac{\arctan\left(\frac{(y-t)}{s}\right)}{\pi}, \quad (4.15)$$

$$F(x) = \frac{s}{\pi(s^2 + (y-t)^2)} \quad (4.16)$$

Para o presente exemplo do ARMA(1,1) com inovações de Cauchy são realizadas 10000 interações e os primeiros 1000 pontos são descartados, para evitar estados transientes. A série temporal do modelo ARMA(1,1) com inovações de Cauchy para teste é ilustrada na Figura 4.44.

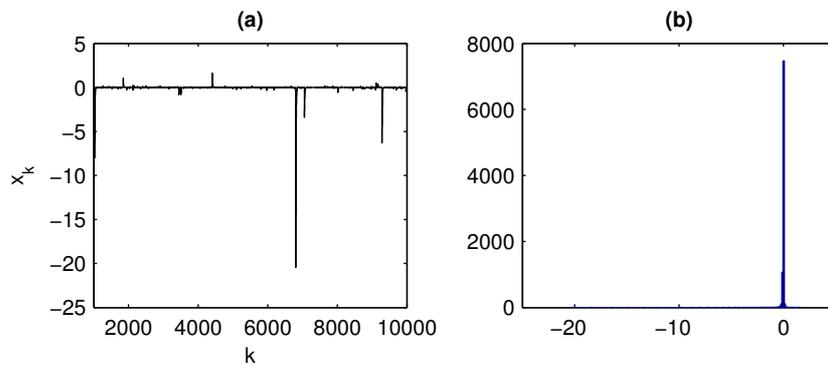


Figura 4.44: Série temporal em (a) ARMA(1,1) com inovações que segue a distribuição de Cauchy ($s = 0,001$ e $t = 0$) e (b) um histograma da série temporal obtida com as inovações propostas.

A Figura 4.45 mostra a série temporal das inovações de Cauchy e_k geradas e um histograma de sua distribuição, muito diferente da distribuição gaussiana.

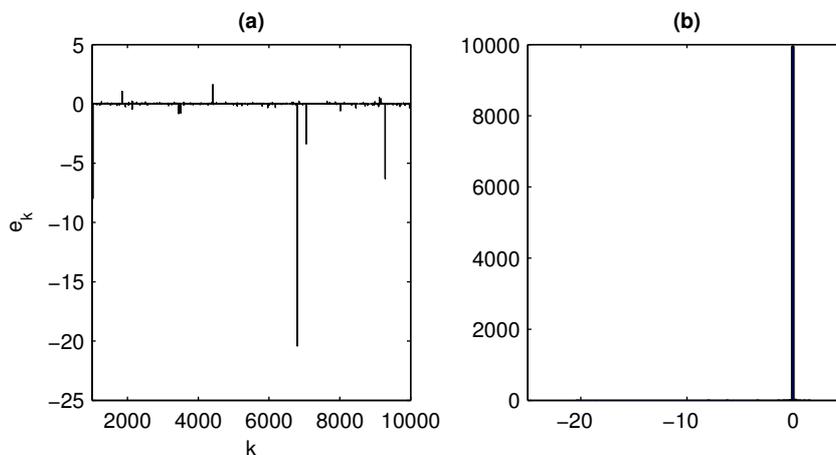


Figura 4.45: Ruído de Cauchy e_k : em (a) série temporal das inovações de Cauchy com distribuição $s = 1$ e $t = 0$, e em (b) o histograma da distribuição da série de Cauchy e_k gerada.

Estas inovações de Cauchy propostas são interessantes, pois podem servir de base para simulações, por exemplo, com possíveis erros de medições de sistema. Tais simulações identificariam a estrutura dinâmica ou não do modelo mesmo com ocorrência de leituras não procedentes do processo, seja pela interação de uma forma mais agressiva do sistema num determinado momento como, por exemplo, um outro processo causando distúrbios (picos elevados) nas medições, ou por quebra ou falha de algum dos instrumentos em um determinado momento de medição.

São geradas 99 séries temporais sub-rogadas ($\alpha = 0,01$) utilizando o algoritmo 0 (IID) e o algoritmo 2 (AAFT, IAAFT-1, CAAFT e STAP), respectivamente, para o teste de hipótese nula. O sub-rogado CAAFT possui os seguintes parâmetros: $\tau_{max} = p = 20$ e $k = 99$. Os dados sub-rogados gerados pelo STAP possuem parâmetros: $m = 10$, $\tau_{max} = p = 20$, $k = 99$.

A Figura 4.46 ilustra a aplicação da estatística de teste não-paramétrica da FAC e IMM, para os dados sub-rogados IID, AAFT, IAAFT-1, CAAFT e STAP, respectivamente.

É importante notar que na Figura 4.46(a) e 4.46(b) a hipótese NH_0 é rejeitada (1% para cometer erro do tipo I), identificando estatisticamente alguma dinâmica determinística na série original.

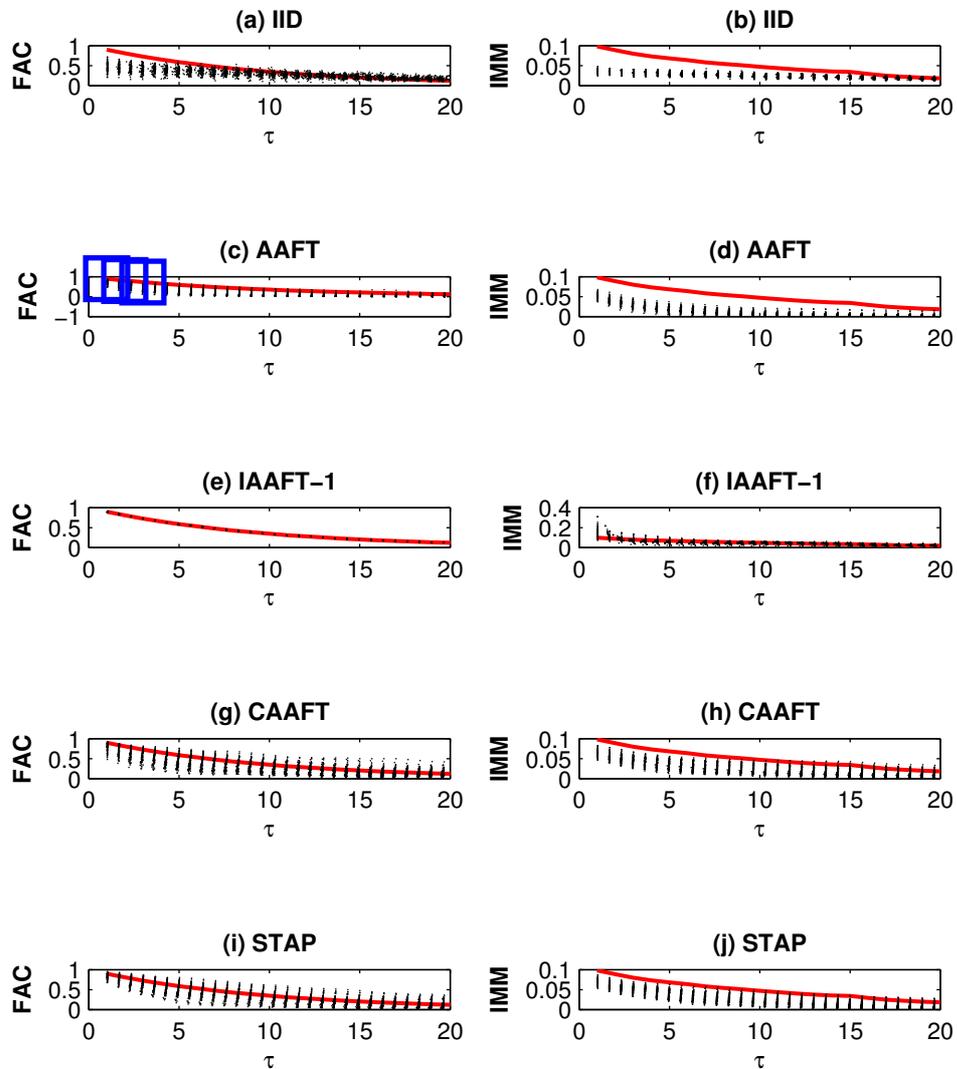


Figura 4.46: Gráfico da FAC e da IMM para o ARMA(1,1) com inovações de Cauchy ($s = 5$ e $t = 0.001$): (a) e (b) para IID, (c) e (d) para AAFT, (e) e (f) para IAAFT-1, (g) e (h) para CAAFT, (i) e (j) para STAP. A linha sólida corresponde a série original e as linhas pontilhadas são as 99 séries sub-rogadas geradas. As observações feitas com retângulos nas figuras são pontos ligeiramente fora da distribuição dos dados sub-rogados.

Por outro lado, o algoritmo 2 (técnicas AAFT, CAAFT e STAP) empregadas para teste rejeitam a hipótese nula NH_2 , Figura 4.46(c), 4.46(d), 4.46(g), 4.46(h) e 4.46(j).

O algoritmo 2 (IAAFT-1) obteve o melhor resultado, mais condizente com

a hipótese nula de interesse (NH_2), pois não se pode rejeitar a H_0 de interesse, como mostram as Figuras 4.46(e) e 4.46(f). É importante lembrar que o IAAFT não depende da transformação “ h ”. Portanto, o resultado do teste para a série original, por meio do IAAFT-1, é que existe alguma dinâmica determinística linear subjacente e que os dados sub-rogados do IAAFT-1 são gerados por uma transformação “ h ” (monotônica estática e não-linear) de um processo possivelmente (99%) linear, ou seja, para gerar o sub-rogado basta aplicar a transformação “ h ” na série temporal original (um processo linear).

Lembrando que a proposta para investigação das técnicas CAAFT e STAP, feita no exemplo ARMA(1,1) com inovações de Laplace, não é verdadeira, ou seja, nem sempre o CAAFT e STAP proporcionaram análises consistentes com a série temporal original. Portanto, retorna-se a dúvida em qual algoritmo ou técnica confiar. Essa conclusão é realizada, após análise de forma mais abrangente das possíveis estruturas das séries temporais aplicáveis, no Capítulo 5.

Uma observação importante que foi realizada, mas não ilustrada nos exemplos dessa seção, é que os testes estatísticos aplicados com apenas 39 séries temporais sub-rogadas (95%) não obtiveram bons resultados para modelos ARMA com inovações não-gaussianas, sendo que algumas hipóteses nulas foram rejeitadas indevidamente, logo se optou por um número maior de séries sub-rogadas (99) diminuindo a probabilidade para 1% de cometer um erro tipo I, ou seja, a probabilidade de 1% em rejeitar a H_0 quando ela for verdadeira. Tal observação é válida, pois para algumas análises obtidas nos testes o grau de confiabilidade dos resultados das estatísticas de testes podem ser elevados apenas aumentando o número de dados sub-rogados para o teste de hipótese. Porém, essa proposta tem um alto custo computacional, dificultando o tempo de análise, por exemplo, para análise (não existe nenhum exemplo na literatura) de alguma aplicação em séries temporais extraídas de algum processo real para identificação de estruturas dinâmicas não-lineares.

A Figura 4.47 resume, por meio de um fluxograma, um quadro geral para análise das séries dinâmicas lineares com inovações não-gaussianas (MA, AR e ARMA). Portanto, chega-se a uma conclusão nessa seção que dados contaminados com ruídos não-gaussianos são de difíceis investigações ou não triviais para o teste de hipótese nula utilizando os métodos de dados sub-rogados. Porém é observado que apenas o teste utilizando o algoritmo 0 (IID, BOOT, GAUSS e SS) é indicado para identificação de determinismo nas séries em estudo, não sendo possível em algumas vezes, afirmar ou não sobre nenhuma das hipóteses (NH_1 ou NH_2).

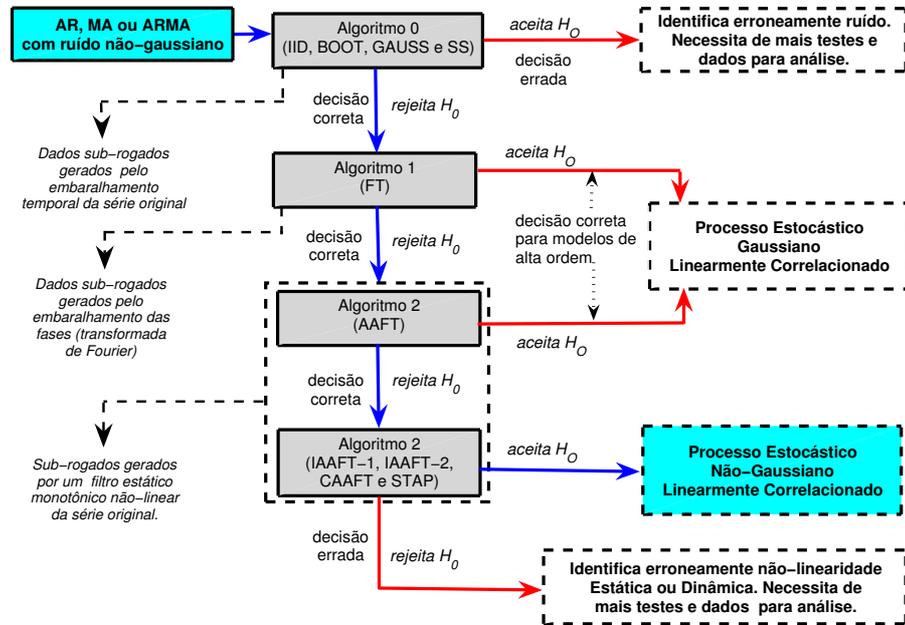


Figura 4.47: Fluxograma geral para teste de hipótese nula (H_0) utilizando os métodos de dados sub-rogados algoritmo 0 (IID, BOOT, GAUSS e SS), algoritmo 1 (FT) e algoritmo 2 (AFT, IAAFT-1, IAAFT-2, CAAFT e STAP) aplicados somente a séries temporais com estruturas do tipo AR, MA ou ARMA com inovações não-gaussianas.

4.6 Transformação não-linear aplicada a processos lineares realimentados com inovações gaussianas e não-gaussianas

Nesta seção, a utilidade dos sub-rogados FT, AAFT, IAAFT-1, IAAFT-2, CAAFT e STAP para modelos linearmente correlacionados com inovações não-gaussiana e com transformações não-lineares, estáticas e inversíveis empíricas são discutidas mais amplamente. Na seção 4.5 foi relatada apenas para modelos ARMA com inovações não-gaussianas, aqui serão empregados também modelos não-lineares (NARMA) com inovações não-gaussianas.

Para colocações experimentais, tem-se freqüentemente só aplicado uma única realização do processo dinâmico, conforme visto nas seções anteriores. Assumise que esta única realização é um representação da dinâmica subjacente. Tal suposição é válida especialmente para processos estacionários cujas propriedades estatísticas são invariantes no tempo. A conservação do espectro de potência dos dados nos sub-rogados FT insinua a conservação da função de autocorrelação. Rejeitar H_0 usando uma estatística discriminante sensível a não-linearidades (estática ou dinâmica) freqüentemente conduz à conclusão que a determinada amostra empírica exhibe correlações não-lineares. Subseqüentemente, algoritmos tais como AAFT, IAAFT (Nagaranjan, 2005) e CAAFT e STAP (Kugiumtzis, 1999c, 2002b) são usados para inferir a natureza não-linear. A hipótese nula dirigida aos sub-rogados AAFT, IAAFT, CAAFT e STAP são que a determinada amostra empírica é gerada por uma transformação não-linear, estática e inversível de um ruído linearmente correlacionado com inovações gaussianas.

Segundo (Nagaranjan, 2005), rejeitar a H_0 usando uma estatística discriminante sensível à não-linearidades dinâmicas, por exemplo, IMM, é atribuído à existência de não-linearidades dinâmicas na amostra empírica. O IAAFT é um das formas precursoras para inferir caos determinístico¹⁴ (Theiler et al., 1992).

Os resultados apresentados nesta seção proporcionam uma boa distinção entre não-linearidade e não-gaussianidade. A investigação da não-gaussianidade e não-linearidade é realizada nessa seção por meio de: (i) uma não-linear transformação estática e inversível de um processo linear de primeira-ordem realimentado com inovações gaussianas, (ii) um processo linear de primeira-ordem realimentado com inovações não-gaussianas e (iii) uma não-linear transformação estática

¹⁴Deve-se notar que caos determinístico é um exemplo de não-linearidade dinâmica e não envolve a classe inteira de dinâmica não-lineares.

e inversível de um processo linear de primeira-ordem realimentado com inovações não-gaussianas. Para isto são mostrados que os sub-rogados FT, AAFT e IAAFT podem não ser adequados em identificar amostras empíricas geradas por processos lineares realimentados¹⁵ com inovações não-gaussianas e suas transformações não-lineares.

O processo linear realimentado de primeira-ordem é representado na Equação 4.17, com inovações ϵ_k amostradas de um processo independente e identicamente distribuído (i.i.d) com média zero e variância unitária.

$$x_k = \theta x_{k-1} + \epsilon_k \quad (4.17)$$

onde $k = 1, \dots, N$.

Enquanto a definição clássica de ruído linearmente correlacionado assume implicitamente a gaussianidade de ϵ_k , tal necessidade de suposição não necessariamente é em geral verdade (Bickel e Buhlmann, 1996). É importante lembrar que o decaimento da correlação é governada pelo sinal e a magnitude do parâmetro θ do processo. Desta forma, tem-se que o processo da Equação 4.17 é estacionário quando $|\theta| < 1$.

Esta seção é focada nos processos não-gaussianos, para isso são propostos três exemplos simples:

1. Transformação não-linear, estática e inversível de um ruído linearmente correlacionado

$$x_k = \theta x_{k-1} + \epsilon_k \quad (4.18)$$

$$y_k = x_k \sqrt{|x_k|} \quad (4.19)$$

com parâmetro $\theta = 0.95$ e inovações ϵ_k amostradas de um processo i.i.d. normalmente distribuído com média zero e variância unitária. Nas discussões subsequentes este processo é referido como **transformação não-linear do ruído branco gaussiano** (nonlinear transformed additive white Gaussian noise) - NAWGN.

2. Ruído linearmente correlacionado

$$y_k = \theta y_{k-1} + \epsilon_k \quad (4.20)$$

¹⁵um processo realimentado linearmente correlacionado é também conhecido como ruído linearmente correlacionado.

com parâmetro $\theta = 0.95$ e inovações ϵ_k não-gaussianas amostradas de um processo de distribuição exponencial¹⁶ i.i.d. com média zero e variância unitária. Denominado como um processo linear do **ruído branco não-gaussiano** (additive white non-gaussian noise) - AWNGN.

3. Transformação não-linear, estática e inversível de um ruído linearmente relacionado

$$x_k = \theta x_{k-1} + \epsilon_k \quad (4.21)$$

$$y_k = x_k \sqrt{|x_k|} \quad (4.22)$$

com parâmetro $\theta = 0.95$ e inovações ϵ_k não-gaussianas amostradas de um processo de distribuição exponencial i.i.d. com média zero e variância unitária. Este processo é chamado como **transformação não-linear do ruído branco não-gaussiano** (nonlinear transformed additive white non-gaussian noise) - NAWNGN.

A escolha do parâmetro de processo $\theta = 0.95$ e da não-linearidade estática-inversível $y_k = x_k \sqrt{|x_k|}$ é encorajada para atestar a aplicabilidade dos sub-rogados. São gerados os processos NAWGN, AWNGN e NAWNGN com 10000 interações e descartados os primeiros 1000 pontos, a Figura 4.48.

Nesta seção são produzidos 99 dados sub-rogados¹⁷ (FT, AAFT, IAAFT-1, IAAFT-2, CAAFT e STAP) para cada processo, respectivamente, NAWGN, AWNGN e NAWNGN. O sub-rogado CAAFT possui os seguintes parâmetros: $\tau_{max} = p = 20$ e $k = 99$. Os dados sub-rogados gerados pelo STAP possuem parâmetros: $m = 10$, $\tau_{max} = p = 20$, $k = 99$.

¹⁶Pode ser mostrado analiticamente que os cumulantes de segunda-ordem a saber: média e variância, são suficientes para descrever inovações gaussianas de ϵ_k . Porém, para ϵ_k amostrados de uma distribuição exponencial, cumulantes de altas ordens são necessários para descrever o processo completamente. Considere $f(x) = \lambda e^{-\lambda x}$, $x \geq 0$, o k-ésimo cumulante c_k é dado pela expressão $c_k = (k-1)\lambda^{-k}$ que exibe um decaimento exponencial com k.

¹⁷Que proporciona um $\alpha = 0,01$ para o teste estatístico.

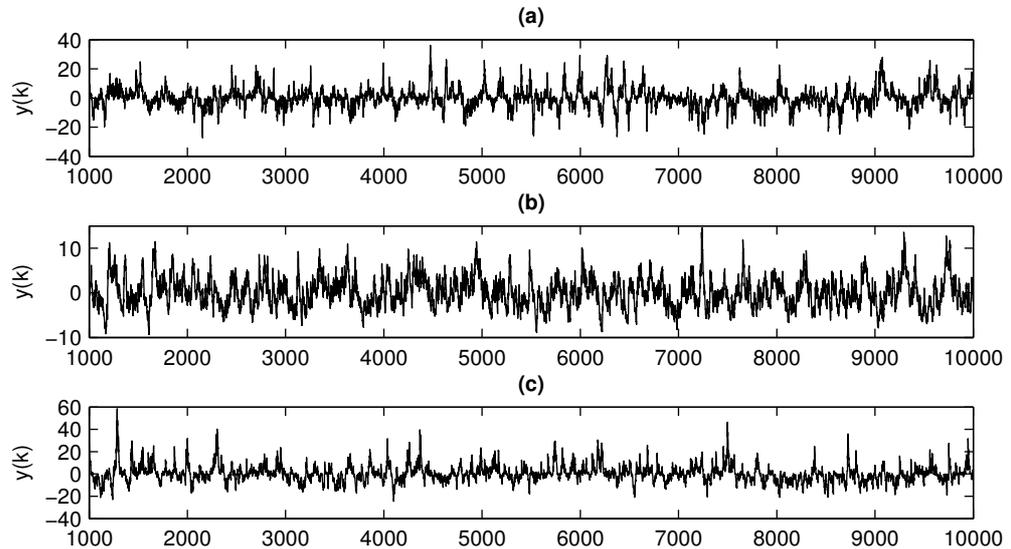


Figura 4.48: Séries temporais em (a) de uma transformação não-linear do ruído branco gaussiano - NAWGN, em (b) ruído branco não-gaussiano - AWNGN e em (c) de uma transformação não-linear do ruído branco não-gaussiano - NAWNGN.

Primeiramente, para o algoritmo FT que gera sub-rogados gaussianos, observa-se que este método não é adequado para séries com inovações não-gaussianas, uma vez que as séries artificiais assim geradas terão distribuições diferentes da série original, Figura 4.49.

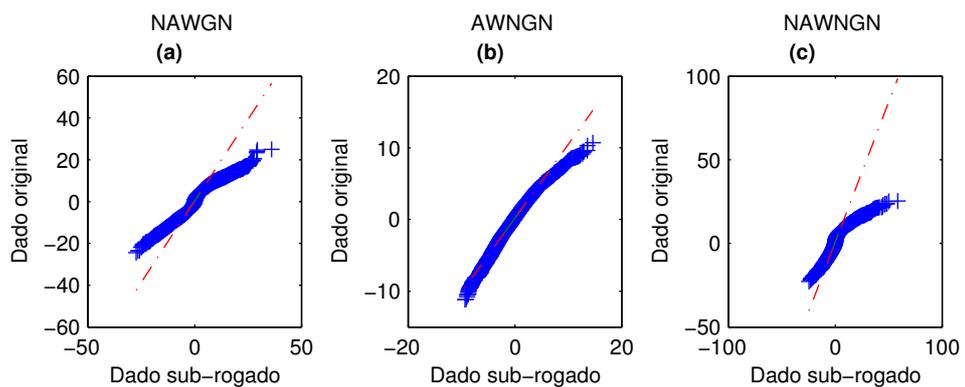


Figura 4.49: Gráficos quantil-quantil: em (a), (b) e (c) correspondendo respectivamente NAWGN, AWNGN e NAWNGN para o sub-rogado FT.

A Figura 4.50 e Figura 4.51 mostram as estatísticas de teste dos sub-rogados.

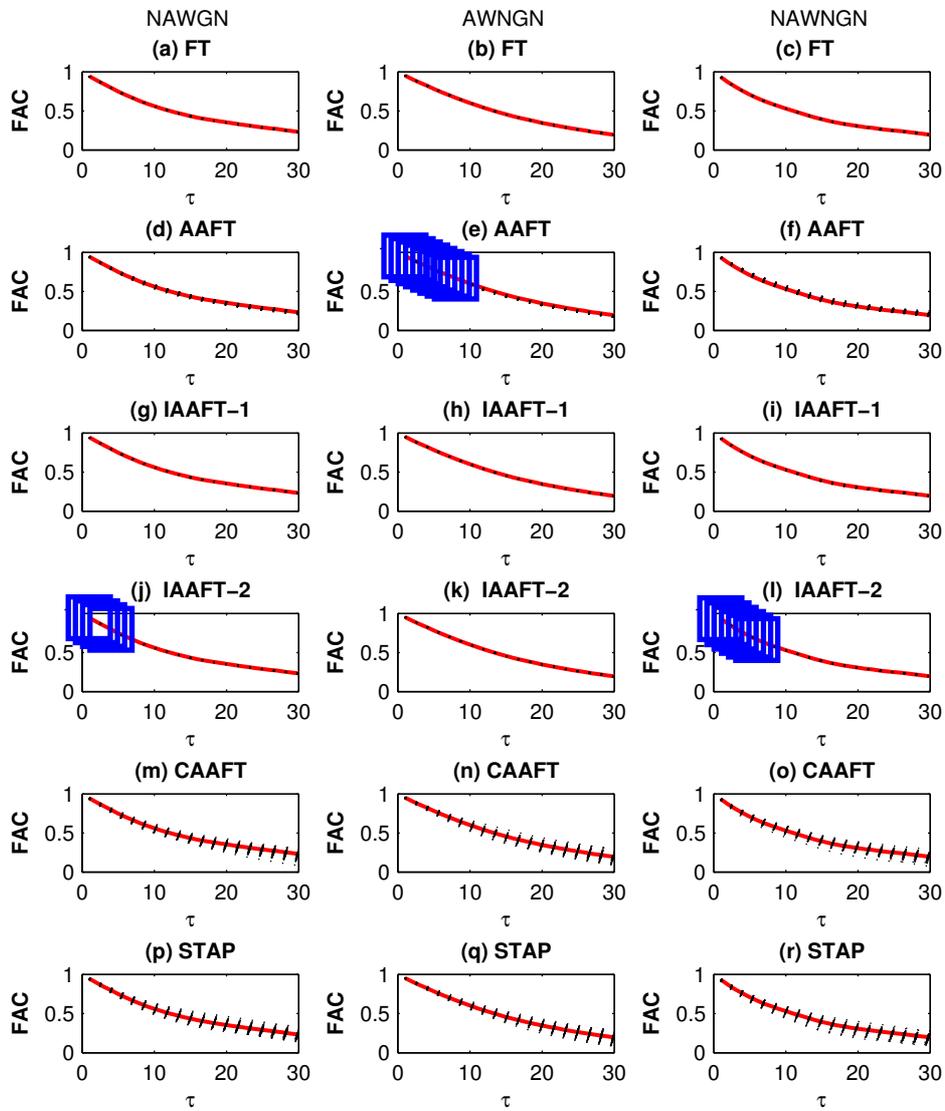


Figura 4.50: Gráficos da FACs. Colunas da esquerda para a direita, correspondem respectivamente: NAWGN, AWNGN e NAWNGN. As FACs são: em (a), (b) e (c) do FT; em (d), (e) e (f) do AAF; em (g), (h) e (i) do IAAFT-1; em (j), (l) e (k) do IAAFT-2; em (m), (n) e (o) para CAAFT e em (p), (q) e (r) do STAP. A linha sólida corresponde a série original e as linhas pontilhadas são as 99 séries sub-rogadas geradas. As observações feitas com retângulos nas figuras são pontos fora da distribuição dos dados sub-rogados gerados.

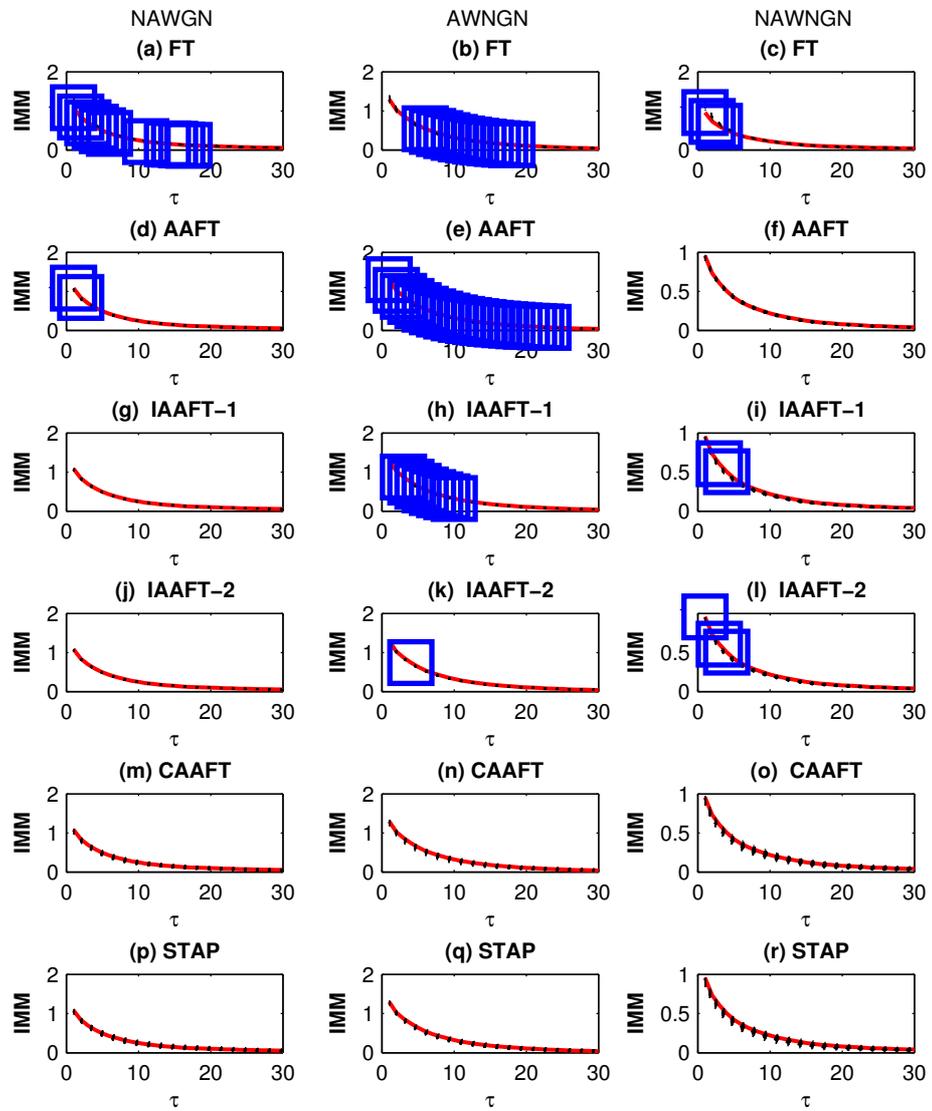


Figura 4.51: Gráficos da IMM. Colunas da esquerda para a direita, correspondem respectivamente: NAWGN, AWNGN e NAWNGN. As IMM são: em (a), (b) e (c) do FT; em (d), (e) e (f) do AAFT; em (g), (h) e (i) do IAAFT-1; em (j), (l) e (k) do IAAFT-2; em (m), (n) e (o) para CAAFT e em (p), (q) e (r) do STAP. A linha sólida corresponde a série original e as linhas pontilhadas são as 99 séries sub-rogadas geradas. As observações feitas com retângulos nas figuras são pontos fora da distribuição dos dados sub-rogados gerados.

A Figura 4.52 apresenta um histograma das séries temporais NAWGN, AWNGN e NAWNGN, respectivamente.

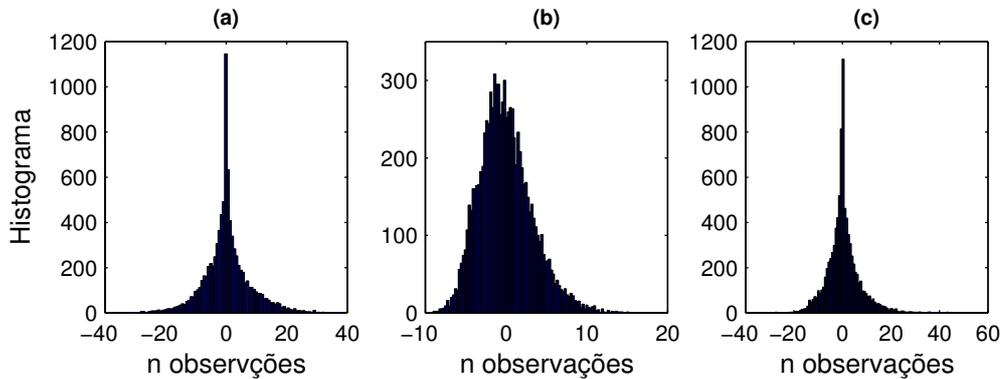


Figura 4.52: Histograma das séries: (a) NAWGN, (b) AWNGN e (c) NAWNGN.

A Tabela 4.4 resume o resultado encontrado das estatísticas de teste da FAC e da IMM em conjunto, como proposta por Small et al. (2006a). Small relata que se ocorrer a rejeição da FAC ou da IMM rejeita-se a hipótese nula aplicada. Isso torna o teste mais criterioso aceitação da hipótese nula proposta. Portanto, a Tabela 4.4 segue esse idéia do Small.

Tabela 4.4: Resultados da H_0 para o NAWGN, AWNGN e NAWNGN.

Técnica	NAWGN	AWNGN	NAWNGN
FT	rejeita	rejeita	rejeita
AAFT	rejeita	rejeita	não rejeita
IAAFT-1	não rejeita	rejeita	rejeita
IAAFT-2	rejeita	rejeita	rejeita
CAAFT	não rejeita	não rejeita	não rejeita
STAP	não rejeita	não rejeita	não rejeita

A presente Tabela 4.4 descreve as decisões para cada H_0 mais claramente. A hipótese NH_1 , algoritmo 1 (FT), é rejeitada corretamente para o NAWGN, o AWNGN e o NAWNGN, pois não são processos lineares gaussianos. Lembrando que o AWNGN é um processo linear com ruído não-gaussiano.

O algoritmo 2 (AAFT) rejeita corretamente a H_0 para o NAWGN e AWNGN, pois a aceitação da H_0 do AAFT é para processos lineares gaussianos. Todavia, para NAWNGN não rejeita a H_0 , o que é incorreto, pois NAWNGN é não-linear.

O IAAFT-1 acerta apenas ao rejeitar a H_0 para a série temporal NAWNGN. O IAAFT-2 rejeita corretamente a H_0 para as séries temporais NAWGN (não-linear e gaussiana) e NAWNGN (não-linear e não-gaussiana).

Analisando a FAC (Figuras 4.50(m), 4.50(n), 4.50(o), 4.50(p), 4.50(q) e 4.50(r)) e a IMM (Figura 4.51(n), 4.51(o), 4.51(p), 4.51(q) e 4.51(r)), observa-se que o CAAFT e STAP não rejeita a hipótese NH_2 , apesar da não-gaussianidade e das transformações não-lineares impostas. É importante observar que a aceitação da H_0 para as séries NAWGN e NAWNGN, afirmam que tais séries são processos lineares gaussianos ou não-gaussianos, uma vez que tais séries são não-lineares (uma transformação estática não-linear).

Até o momento, parece não haver uma comparação crítica abrangente sobre diversas representações matemáticas para sistemas não-lineares. Uma conclusão importante vista é que os resultados das estatísticas de testes não-paramétricas (FAC) apresentadas indicam uma conservação do espectro de potência, confirmando, o que segundo Nagaranjan (2005), pode não ser suficiente em discernir não-linearidades em sinais não-gaussianos. Este fato não é uma limitação dos algoritmos sub-rogados existentes, mas apenas uma limitação inerente.

O presente estudo também encoraja a interpretação mais judiciosa de sub-rogados testados em conjuntos de dados experimentais, geralmente não-gaussianos e as transformações estáticas desses dados. Os algoritmos CAAFT e STAP mostraram ser não aplicáveis para transformações estáticas não-lineares de processos lineares realimentados com ruído gaussiano ou não-gaussiano.

Nessa seção foram tratados casos com não-linearidades estáticas, na seção seguinte será investigado não-linearidade dinâmica com inovações gaussianas ou não-gaussianas.

4.7 Mapas

Existem várias razões para estudar mapas do ponto de vista do seu comportamento dinâmico. Com efeito, há inúmeros modelos cujas soluções aparecem sob forma de relações de recorrência e, portanto por uma dinâmica descrita por mapas. Três deles são tratados aqui para análise do método de hipótese com dados sub-rogados ambos com 10000 pontos simulados, onde são descartados os 1000 primeiros dados, para evitar estados transitórios.

- Mapa de Ikeda dado por (Ikeda, 1973)

$$\begin{pmatrix} x_{k+1} \\ y_{k+1} \end{pmatrix} = \begin{pmatrix} 1 + \mu(x_k \cos \theta - y_k \sin \theta) \\ \mu(x_k \sin \theta + y_k \cos \theta) \end{pmatrix}, \quad (4.23)$$

em que $\theta = a - \frac{b}{(1 + x_k^2 + y_k^2)}$ com $\mu = 0.83$, $a = 0.4$ e $b = 6.0$.

- Mapa Logístico dado por (May, 1976)

$$x_k = ax_{k-1}(1.0 - x_{k-1}) \quad (4.24)$$

onde usa-se $a = 4.0$.

- Mapa de Hénon dado por (Fielder-Ferrara e Prado, 1994)

$$\begin{aligned} x_{k+1} &= a_0 + a_1 x_k^2 + y_k \\ y_{k+1} &= b x_k \end{aligned} \quad (4.25)$$

onde $a_0 = 1.0$, $a_1 = -1.4$ e $b = 0.3$.

O mapa logístico é um modelo usado para dinâmica de populações biológicas e é um mapa unimodal, ou seja, são mapas definidos num intervalo finito e com um único máximo¹⁸. Estes mapas não são inversíveis, podendo em um ponto apresentar duas pré-imagens. Os mapas de Ikeda e de Hénon são bidimensionais. A Figura 4.53 ilustra o mapa de Ikeda, o mapa Logístico e o mapa de Hénon, respectivamente, e um trecho (janela) do comportamento dinâmico das componentes escolhidas para o teste de hipótese.

Com o intuito de uma investigação mais ampla e robusta dos modelos de dinâmicas não-lineares é proposto a introdução de um ruído aos mapas, lembrando que esse ruído não é dinâmico, ou seja, o ruído não faz parte da dinâmica geradora do processo, logo têm-se um processo com comportamento do tipo $y_k = x_k + e_k$. Os ruídos introduzidos são respectivamente: um ruído gaussiano para as componentes do mapa de Ikeda, ruído uniforme $[0,1]$ para o mapa Logístico e para as componentes do mapa de Henon um ruído beta (2,5). Todos estes ruídos já foram descritos e aplicados nas seções anteriores.

¹⁸mais exatamente, mapas unimodais são mapas continuamente diferenciáveis que mapeam o intervalo unitário $[0,1]$ nele mesmo, que apresentam um máximo (único) em $x = 0.5$ e são monotônicos para $0 \leq x \leq 0.5$ e $0.5 < x \leq 1$

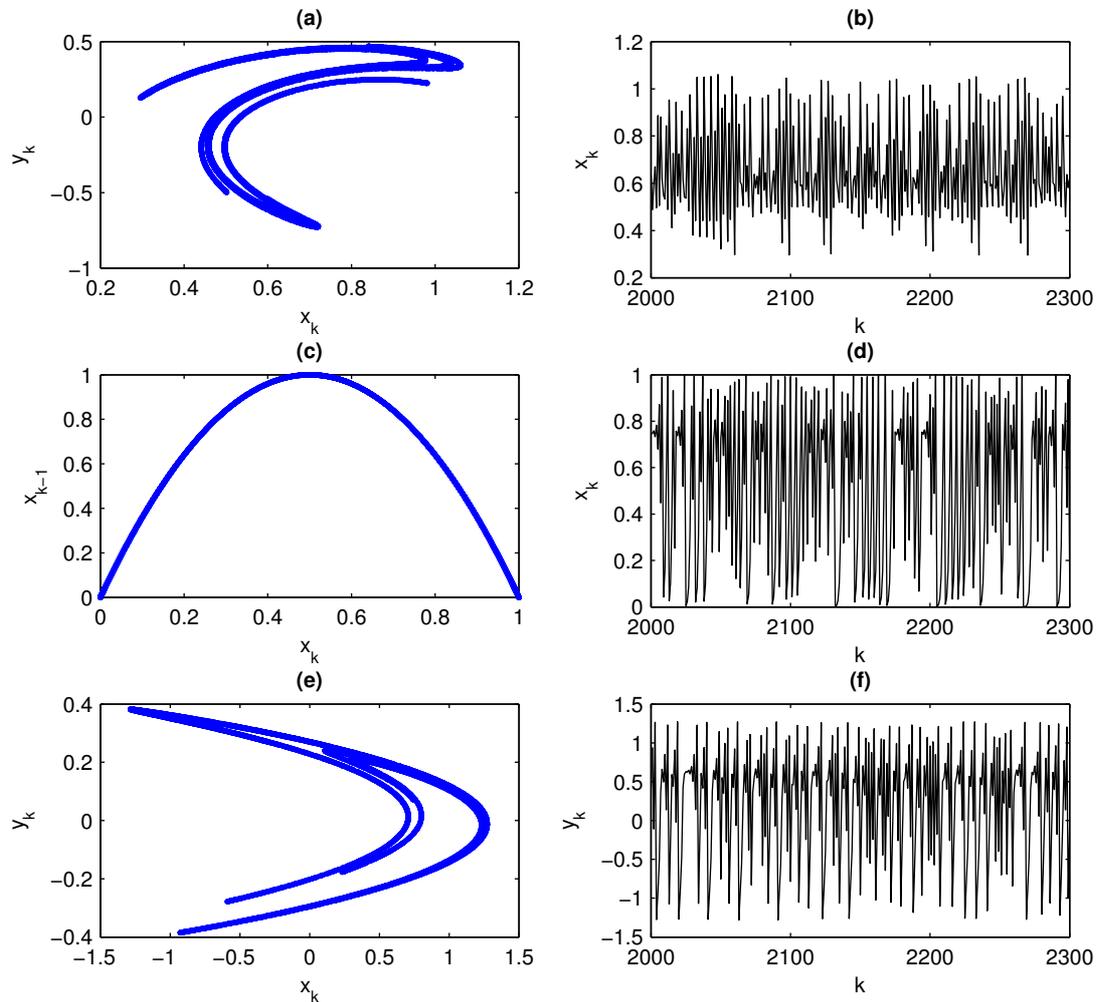


Figura 4.53: Mapas de recorrência: (a) Mapa de Ikeda e em (b) janela da série temporal da componente x_k do Mapa de Ikeda; (c) Mapa Logístico e em (d) janela da série temporal da componente x_k do Mapa Logístico; (e) Mapa de Hénon e em (f) janela da série temporal da componente y_k do Mapa de Hénon.

Primeiramente, são geradas 99 séries sub-rogadas ($\alpha = 0,01$) para teste da série temporal da componente x_k do mapa de Ikeda com ruído gaussiano utilizando o algoritmo 0 (IID), algoritmo 1 e algoritmo 2 (AAFT, CAAFT e STAP). Os dados sub-rogados CAAFT são gerados utilizando os seguintes parâmetros: $\tau_{max} = p = 20$ e $k = 99$. Finalmente, os dados sub-rogados STAP, cujos parâmetros são: $m = 10$, $\tau_{max} = p = 20$ e $k = 99$.

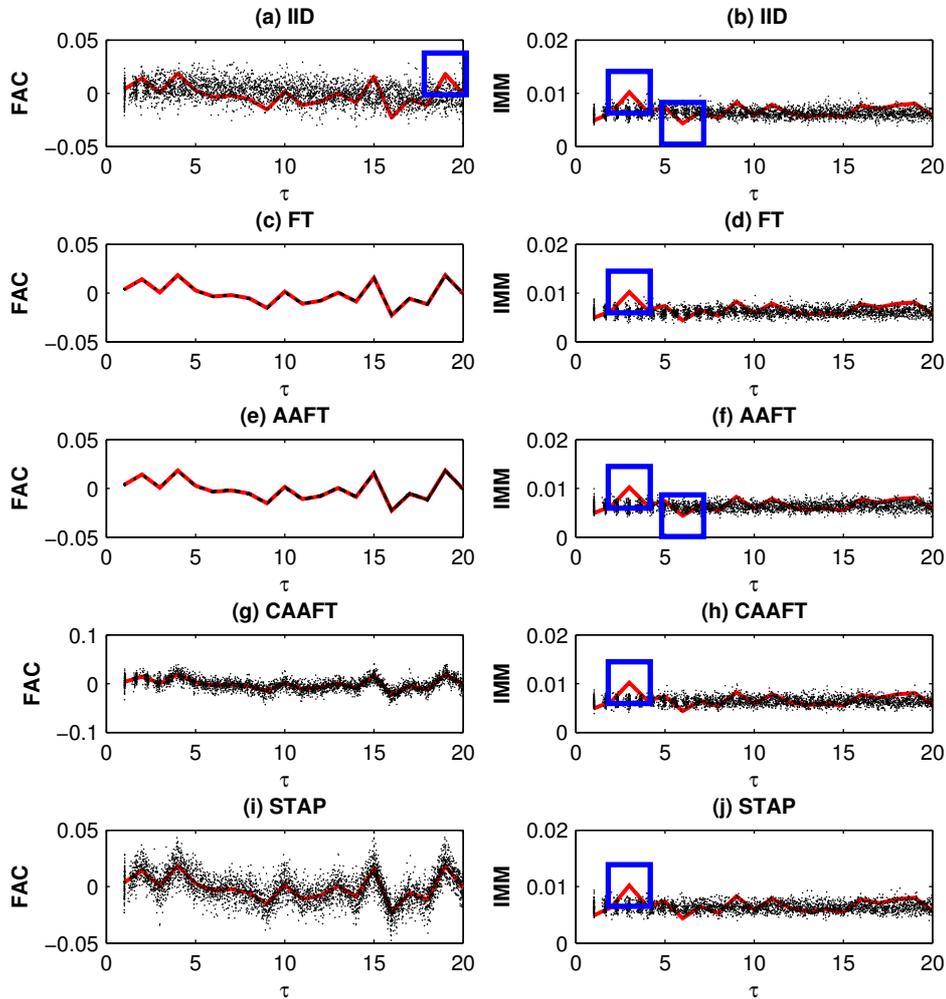


Figura 4.54: Gráfico da FAC e da IMM da componente x_k do mapa de Ikeda com ruído gaussiano: (a) e (b) para IID, (c) e (d) para FT, (e) e (f) para AAFT, (g) e (h) para CAAFT, (i) e (j) para STAP. A linha sólida corresponde a série original e as linhas pontilhadas são as 99 séries sub-rogadas geradas.

A Figura 4.54 apresenta a análise da estatística de teste não-paramétrica para a série temporal da componente x do mapa de Ikeda com ruído gaussiano. É importante notar que na Figura 4.54(a) da FAC e Figura 4.54(b) da IMM rejeitam a hipótese nula NH_0 , indicando que a série temporal da componente x do mapa de Ikeda com ruído gaussiano possui algum tipo de dinâmica subjacente. Observe

que a Figura 4.54(c), 4.54(e), 4.54(g) e 4.54(i) da FAC são favoráveis a hipótese nula (não se pode rejeitar H_0), contudo as Figuras 4.54(d), 4.54(f), 4.54(h), e 4.54(j) da IMM posicionam contra a hipótese nula (rejeita-se H_0).

Segundo Kugiumtzis (2002b) se a série temporal for rejeitada pela estatística de teste não-paramétrica da FAC com os algoritmos CAAFT ou STAP a evidência para não-linearidade é mais segura que ao utilizar os sub-rogados AAFT ou IAAFT-1 e IAAFT-2. Contudo é importante enfatizar que a evidência mais forte de dinâmica não-linear deve ocorrer por meio da rejeição da estatística de teste não-paramétrica com a IMM.

Note que nesse caso houve a rejeição das três hipóteses lineares: NH_0 , NH_1 e NH_2 , característica principal a favor da não-linearidade. Sendo assim, existe possivelmente (99%) algum tipo de dinâmica não-linear.

Para o outro exemplo proposto são gerados 99 dados sub-rogados para teste da série temporal do mapa Logístico com ruído uniforme $[0,1]$, mas utilizando o algoritmo 0 (IID) e algoritmo 2 (AAFT, IAAFT-1, CAAFT e STAP). Parâmetros do CAAFT: $\tau_{max} = p = 20$ e $k = 99$ e parâmetros do STAP: $m = 10$, $\tau_{max} = p = 20$ e $k = 99$.

A mesma análise pode ser feita para o Mapa Logístico. A FAC da Figura 4.55(a) e a IMM da Figura 4.55(b) rejeitam a hipótese NH_0 , indicando que a série temporal da componente x do mapa Logístico com ruído uniforme tem algum tipo de dinâmica subjacente. É importante notar também que a FAC das Figuras 4.55(e), 4.55(g) e 4.55(i) não rejeitam a H_0 , contudo a IMM das Figuras 4.55(d), 4.55(f), 4.55(h) e 4.55(j) - a IMM para o AAFT, IAAFT, CAAFT e STAP, respectivamente - rejeitam a hipótese NH_2 . Isso também é lógico devido o mapa de Logístico possuir comportamento dinâmico não-linear.

Por meio do algoritmo 0 (IID) e do algoritmo 2 (IAAFT-1, CAAFT e STAP), conclui-se que ocorre a rejeição das hipóteses nulas NH_0 e NH_2 (também ocorre a rejeição da hipótese NH_1 , não demonstrada aqui), ou seja, existe algum comportamento dinâmico não-linear na série da componente x do mapa Logístico.

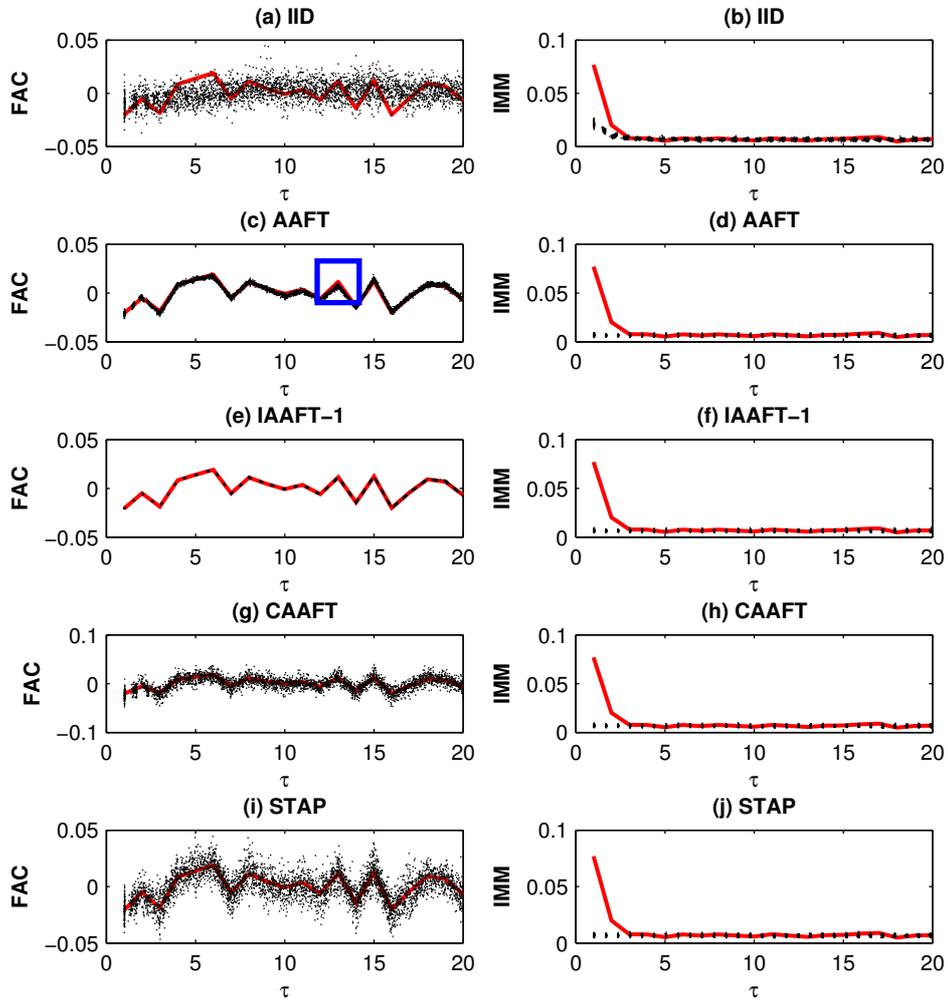


Figura 4.55: Gráfico da FAC e da IMM do mapa Logístico com ruído uniforme [0,1]: (a) e (b) para IID, (c) e (d) para AAFT, (e) e (f) para IAAFT-1, (g) e (h) para CAAFT, (i) e (j) para STAP. A linha sólida corresponde a série original e as linhas pontilhadas são as 99 séries sub-rogadas geradas. O destaque em quadrado na FAC mostram pontos que caem fora da distribuição dos dados sub-rogados, rejeita-se a H_0 .

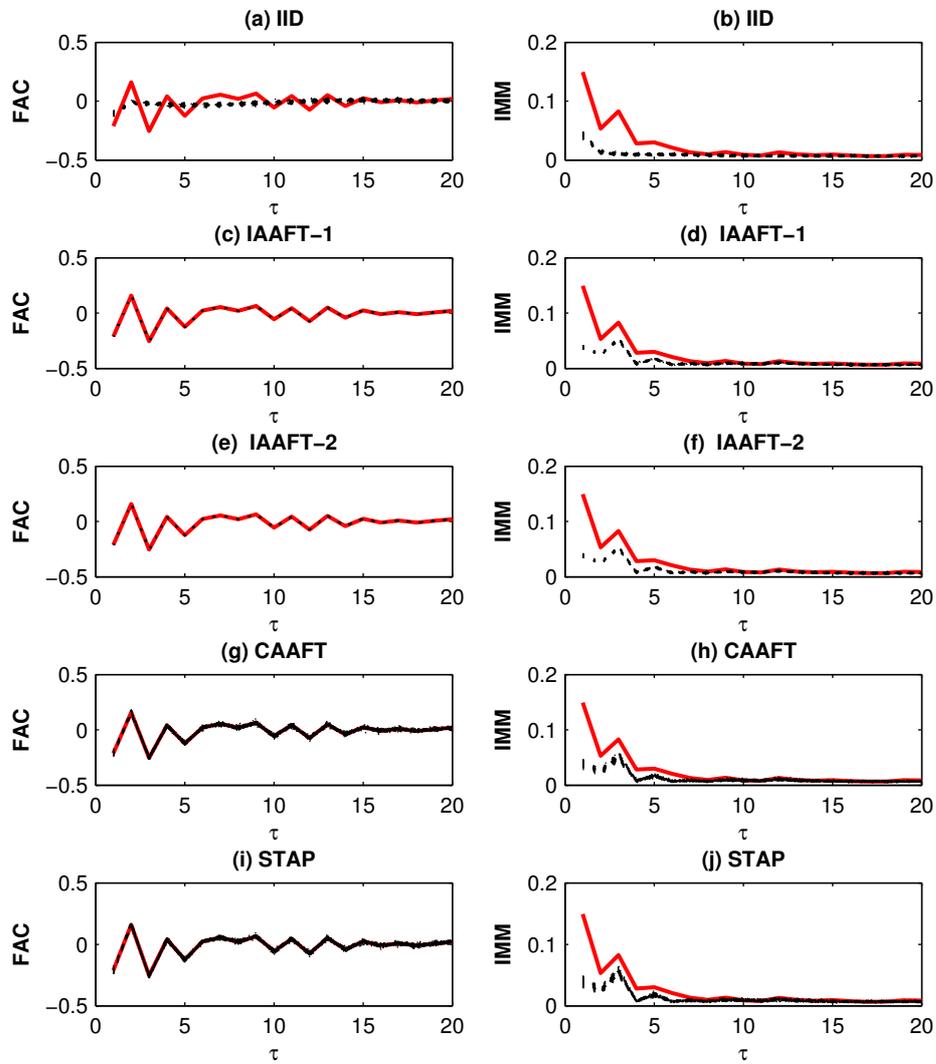


Figura 4.56: Gráfico da FAC e da IMM da componente y do mapa de Hénon: (a) e (b) para IID, (c) e (d) para IAAFT-1, (e) e (f) para IAAFT-2, (g) e (h) para CAAFT, (i) e (j) para STAP. A linha sólida corresponde a série original e as linhas pontilhadas são as 99 séries sub-rogadas geradas.

Finalmente a análise da estatística de teste para o Mapa de Hénon. São geradas 99 séries sub-rogadas para teste da série temporal da componente y do mapa Hénon com ruído beta (2,5), mas utilizando o algoritmo 0 (IID) e algoritmo 2 (IAAFT-1, IAAFT-2, CAAFT e STAP). Os dados sub-rogados CAAFT são gerados utilizando os seguintes parâmetros: $\tau_{max} = p = 20$ e $k = 99$. Os dados sub-rogados STAP possuem parâmetros: $m = 10$, $\tau_{max} = p = 20$ e $k = 99$.

Na Figura 4.56 a FAC e IMM, Figura 4.56(a) e 4.56(b), rejeitam a hipótese NH_0 , indicando que a série temporal da componente y do mapa de Hénon possui algum tipo de dinâmica. Note também que todas as FAC, menos Figura 4.56(a), são favoráveis novamente a hipótese nula (não se pode rejeitar H_0), contudo todas as IMM posicionam contra a hipótese (rejeita-se H_0). Mais uma vez, esse resultado é lógico devido o mapa Hénon ter comportamento dinâmico não-linear. Todavia, para o algoritmo 2 (AAFT), não ilustrado, rejeita-se a hipótese NH_2 . Portanto, o teste de hipótese resulta que na série temporal da componente y do mapa de Hénon existe algum comportamento dinâmico não-linear.

Conclui-se que a rejeição das hipóteses NH_0 , NH_1 e NH_2 são indícios fortes de não-linearidades, geralmente dinâmica, sendo mais fáceis de serem identificadas com os algoritmos sub-rogados propostos que para os casos aplicados na seção 4.6 das não-linearidades estáticas.

O fluxograma da Figura 4.57 sintetiza o procedimento geral para não-linearidade estática e dinâmica.

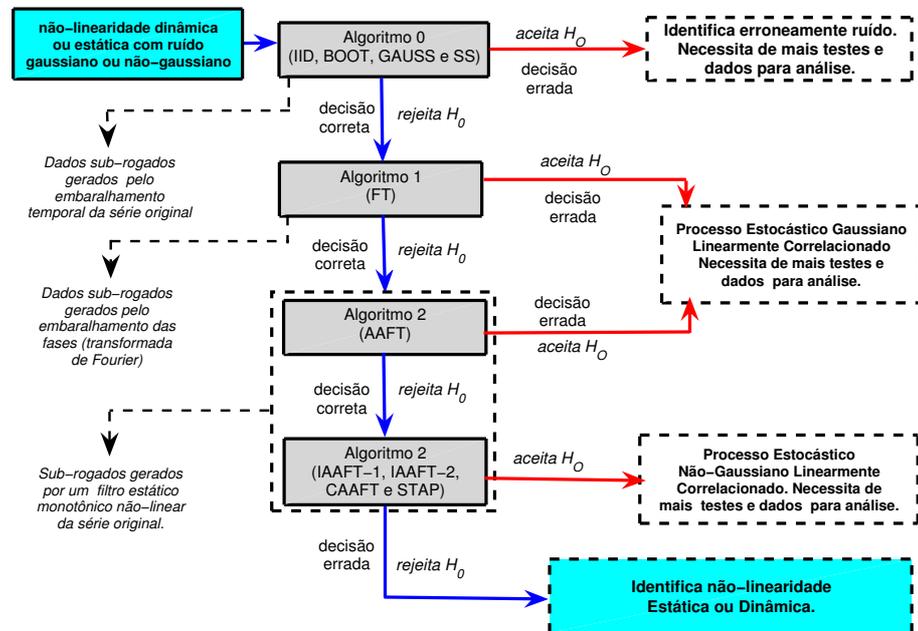


Figura 4.57: Fluxograma geral para teste de hipótese nula (H_0) utilizando os métodos de dados sub-rogados algoritmo 0 (IID e SS), algoritmo 1 (FT) e algoritmo 2 (AFT, IAAFT-1, IAAFT-2, CAAFT e STAP) aplicados somente a séries temporais não-lineares estática ou dinâmica.

4.8 Série de Mackey-Glass com atraso

Para dados de séries temporais que exibem comportamento pseudo-periódico forte, as hipóteses nulas lineares são obviamente incorretas para serem aplicadas. Então, estes algoritmos 0, 1 e 2 existentes são de uso limitado para tais dados.

Para uma série temporal pseudo-periódica, Theiler e Rapp (1996b) sugeriram um algoritmo alternativo: o sub-rogado com embaralhamento de ciclos (*cycle shuffle*) ou algoritmo CS. Análogo ao sub-rogado do algoritmo 0, os sub-rogados CS são produzidos embaralhando os ciclos individuais dentro de uma série temporal. Conseqüentemente, as dinâmicas intra-ciclos são preservadas, mas dinâmicas inter-ciclos não. Contudo, até mesmo com esta aproximação de Theiler notou-

se alguns termos longos indesejados correlacionados nos gráficos da função de autocorrelação para os sub-rogados CS. Além disso, se os picos ou vales não acontecerem precisamente nos respectivos valores ¹⁹ os sub-rogados gerados por este método não podem preservar a estacionariedade e a continuidade dos dados. A troca vertical dos ciclos individuais, podem ser realizada para ciclos emparelhados, preservando assim a sua continuidade. Porém, tal transformação introduz não-estacionariedades nos sub-rogados que não existem ou estão ausentes na série original.

Cada uma das técnicas lineares são aplicadas para confirmar comumente (ou rejeitar) observações de não-linearidades em séries temporais experimentais. Porém, para dados cíclicos, isto é, pseudo-periódicos, os métodos atuais são executados equivocadamente ou simplesmente não são aplicáveis. Um exemplo é a equação diferencial de Mackey-Glass de atraso τ (Mackey e Glass, 1977) dada por

$$\frac{dx(k)}{dk} = -bx(k) + \frac{ax(k - \tau)}{1 + [x(k - \tau)]^{10}} \quad (4.26)$$

com $a = 0.2$, $b = 0.1$ e $\tau = 17$ (condição inicial $x(0) = 1.2$).

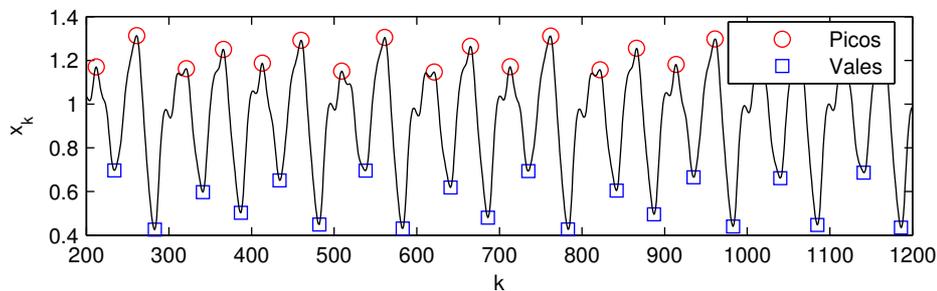


Figura 4.58: Série de Mackey-Glass com atraso $\tau = 17$

A Figura 4.58 ilustra uma série pseudo-periódica, tal série temporal é chamada de série de Mackey-Glass. São gerados na simulação 1200 pontos, cujos 200 primeiros pontos são descartados. Este método para gerar dados sub-rogados por meio do embaralhamento dos ciclos da série (algoritmo CS), direciona-se a hipótese que não há nenhuma correlação dinâmica entre os ciclos (semelhante a NH0), isto quer dizer que há dinâmica dentro dos ciclos, mas não entre eles. São gerados 39 séries sub-rogadas ($\alpha = 0,05$) para o teste de hipótese nula. Na Figura 4.58,

¹⁹sem deslocamentos, ou seja, não ocorre deslocamento nas ligações dos pontos de rupturas entre os ciclos.

é importante notar as possíveis formas para rupturas dos ciclos, que podem ser pontos de rupturas por meio dos picos ou dos vales.

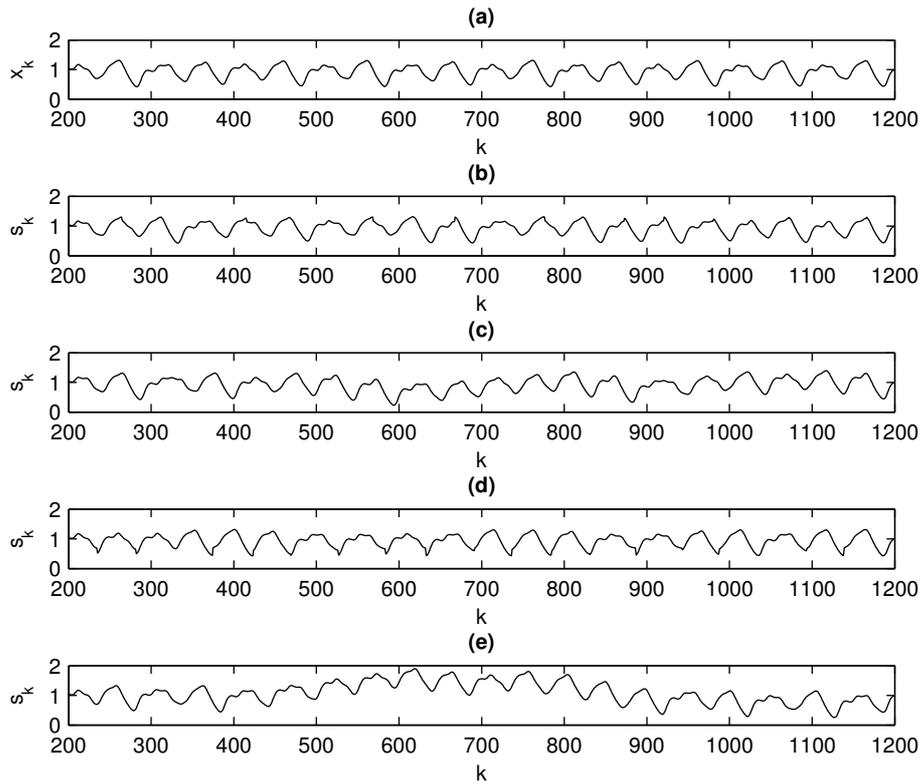


Figura 4.59: Série de Mackey-Glass: (a) série original, (b) embaralhamento dos ciclos (picos) sem deslocamento vertical, (c) embaralhamento dos ciclos (picos) com deslocamento vertical, (d) embaralhamento dos ciclos (vales) sem deslocamento vertical, (e) embaralhamento dos ciclos (vales) com deslocamento vertical.

A escolha do tipo de ponto de ruptura (picos ou vales) implicam em duas ocorrências, vistas na Figura 4.59. As junções dos pontos de rupturas dos ciclos podem ser feitas de duas maneiras: com deslocamento vertical ou sem deslocamento vertical.

O deslocamento vertical é usado para evitar um possível problema com a introdução de aparentes descontinuidades (perde-se na estacionariedade, distribuição de probabilidade dos dados aparentemente diferente) que são deslocados verticalmente para ajuste dos pontos de ligação da série sub-rogado gerada. Se nenhum

deslocamento vertical é executado, evitaria assim emissões inadequadas na estacionariedade dos dados, mas pode introduzir aparentes descontinuidades. As descontinuidade introduzidas sem o deslocamentos verticais dos ciclos são visíveis na Figura 4.60(b) e 4.60(d), que mostram as perdas na distribuição de probabilidade.

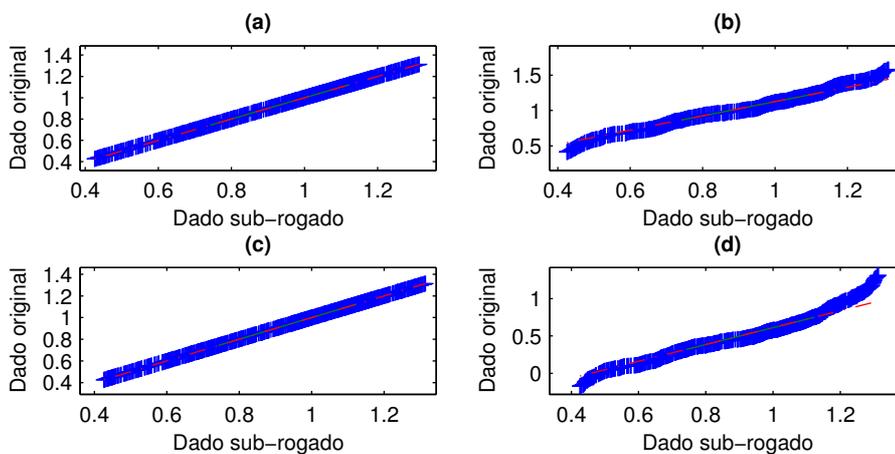


Figura 4.60: Gráficos quantil-quantil: em (a) série original e CS (picos) sem deslocamento vertical, em (b) série original e CS (picos) com deslocamento vertical, em (c) série original e CS (vales) sem deslocamento vertical e em (d) série original e CS (vales) com deslocamento vertical.

Conforme dito no início dessa seção, são encontrados termos longos correlacionados nos gráficos da FAC para os sub-rogados CS com e sem deslocamentos verticais (com rupturas dos ciclos por picos ou vales), Figura 4.61. Theiler (1995) propõe que os sub-rogados gerados pelo CS se direciona a hipótese que não há nenhuma correlação dinâmica entre os ciclos. Além disso, a técnica de embaralhamento dos ciclos é direcionada a uma hipótese, embora não-linear, ligeiramente diferente (semelhante a NH_0). É importante notar que se rejeita a H_0 para a série de Mackey-Glass. Portanto, pode-se concluir que existe possivelmente (95%) um comportamento dinâmico (não-linear) determinístico entre os ciclos.

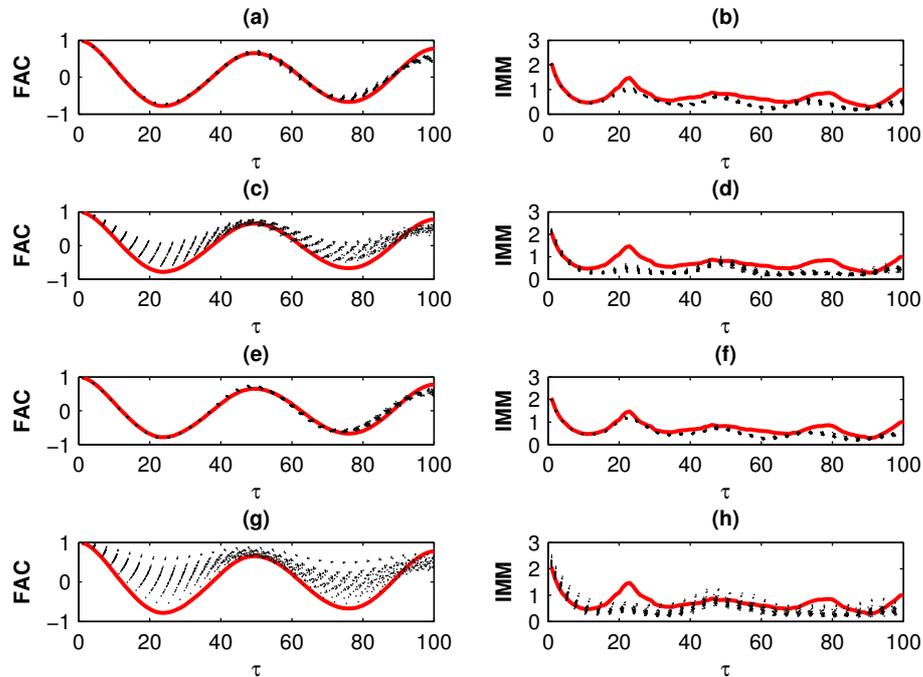


Figura 4.61: FAC e IMM da série de Mackey-Glass para (a) e (b) sub-rogados CS (picos) sem deslocamento vertical, em (c) e (d) sub-rogados CS (picos) com deslocamento vertical, em (e) e (f) sub-rogados CS (vales) sem deslocamento vertical e em (g) e (h) sub-rogados CS (vales) com deslocamento vertical. A linha sólida corresponde a série original e as linhas pontilhadas são as 39 séries sub-rogadas geradas.

A análise da dimensão de correlação na Figura 4.62 mostra distinções clara entre a série temporal original e a distribuição dos dados sub-rogados CS gerados. A dimensão de imersão utilizada é $d_e = 3$ e o algoritmo CS é aplicado com rupturas dos ciclos entre os picos e sem deslocamento vertical. A hipótese nula utilizando a dimensão de correlação como estatística de teste também é rejeitada. Logo, o teste de hipótese nula com o algoritmo CS aplicado à série de Mackey-Glass identifica comportamento dinâmico determinístico entre os ciclos. A dimensão de correlação encontrada é aproximadamente 1,607 para uma dimensão de imersão $d_e = 3$ e $\tau = 17$, utilizando o algoritmo de Gauss-Kernel (GKA).

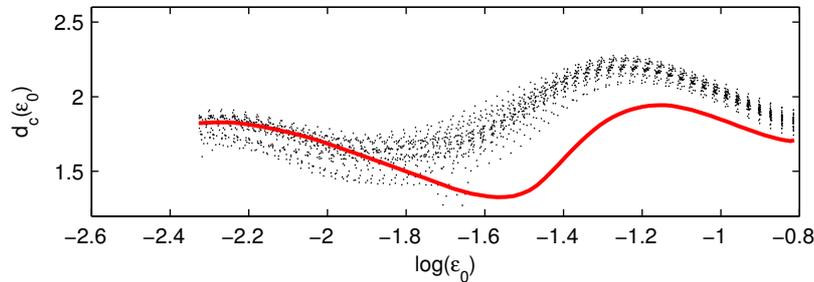


Figura 4.62: Comparação da dimensão de correlação da série original (linha sólida) e do dado CS (picos) sem deslocamento vertical (linhas pontilhadas) para a série de Mackey-Glass com atraso $\tau = 17$. A dimensão de correlação estimada $d_c(\epsilon_0)$ é plotada contra $\log(\epsilon_0)$.

4.9 Ressonância magnética nuclear - NMR

A ressonância magnética nuclear (NMR ou do inglês *Nuclear Magnetic Resonance*) é um fenômeno físico baseado nas propriedades quânticas mecânica-magnéticas do núcleo de um átomo. O NMR se refere também comumente a uma família de métodos científicos que exploram a ressonância magnética nuclear para estudos das moléculas.

Todos os núcleos que contêm números ímpares de prótons ou nêutrons têm um momento magnético intrínseco e impulso angular. As mais comuns medidas nucleares são hidrogênio-1 (o mais receptivo isótopo em abundância natural) e carbono-13, embora também podem ser observados núcleos de isótopos de muitos outros elementos.

O NMR estuda núcleos magnéticos alinhando-os com um campo magnético externo muito poderoso e perturbando este alinhamento usando um campo eletromagnético. A resposta resultante para a perturbação externa do campo eletromagnético é o fenômeno que é explorado no estroboscópico NMR e na imagem da ressonância magnética.

O estroboscópio NMR é uma das principais técnicas usadas para obter a física, química, elétrica e informações estruturais sobre as moléculas. É uma técnica poderosa que pode prover informação detalhada sobre a topologia, a dinâmica e a estrutura tridimensional das moléculas em solução e em estado sólido. Este conjunto de séries temporais contém dados de um corte transversal da saída de potência do estroboscópio de um laser NMR. Como a série exibe um comportamento pseudo-periódico, é empregado para o teste da hipótese nula o algoritmo

CS (embaralhamento de ciclos) gerando 39 séries sub-rogadas para o presente teste.

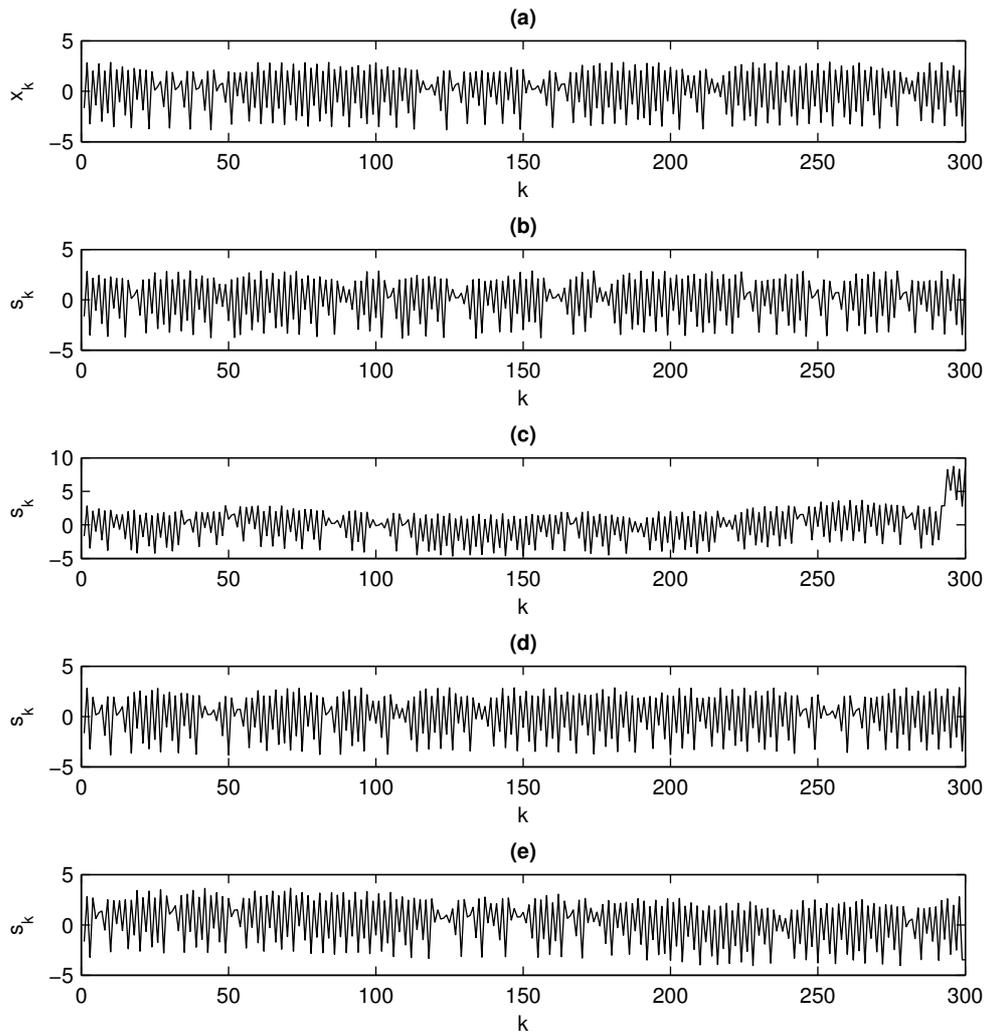


Figura 4.63: Séries do NMR: (a) dado NMR, (b) embaralhamento dos ciclos (picos) sem deslocamento vertical, (c) embaralhamento dos ciclos (picos) com deslocamento vertical, (d) embaralhamento dos ciclos (vales) sem deslocamento vertical, (e) embaralhamento dos ciclos (vales) com deslocamento vertical.

A Figura 4.63 mostram as formas propostas para a construção dos dados sub-

rogados CS. A Figura 4.63(c) mostra os problemas ocorridos com o deslocamento vertical onde é visível a introdução de uma descontinuidade forte nos dados sub-rogados gerados.

Os gráficos da FAC e IMM, Figura 4.64, rejeitam a H_0 da correlação entre os ciclos (inter-ciclos), podendo afirmar que a série original possui dinâmica determinística não-linear, como de fato da série NMR são conhecidos por serem não-lineares (Kantz e Schreiber, 1997).

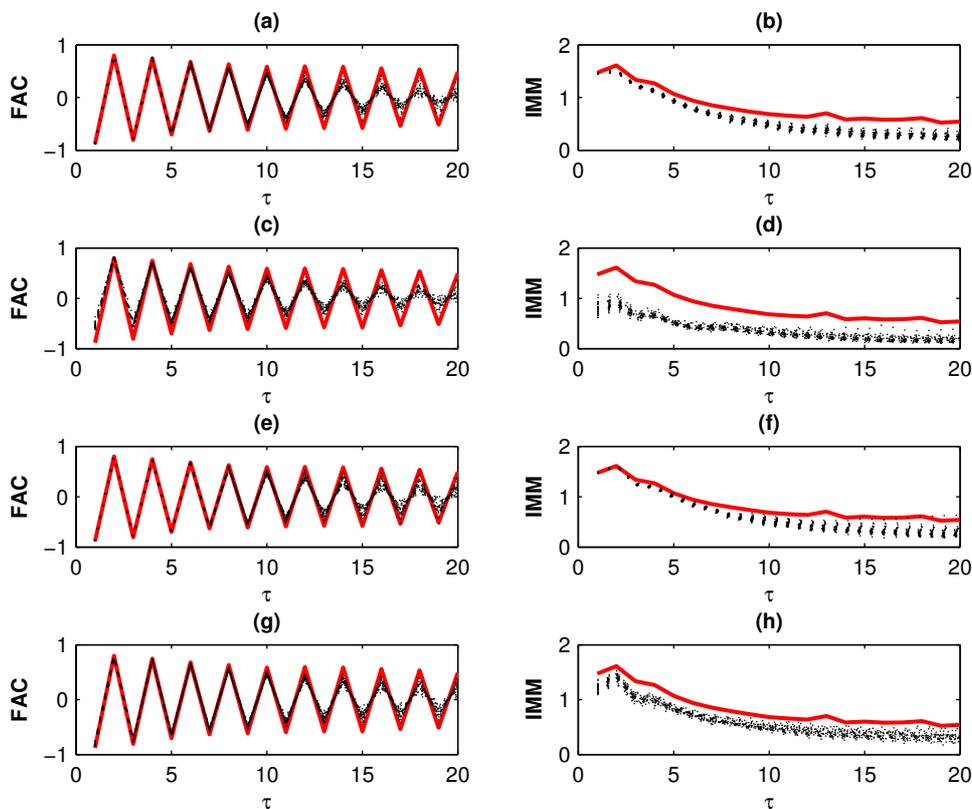


Figura 4.64: FAC e IMM da série NMR para (a) e (b) sub-rogados CS (picos) sem deslocamento vertical, em (c) e (d) sub-rogados CS (picos) com deslocamento vertical, em (e) e (f) sub-rogados CS (vales) sem deslocamento vertical e em (g) e (h) sub-rogados CS (vales) com deslocamento vertical.

4.10 Séries temporais caóticas

Padrão linear para métodos sub-rogados só são úteis para séries temporais que não exibam nenhuma estrutura pseudo-periódica. O algoritmo PPS (Small et al., 2001b)(dado sub-rogado pseudo-periódico) pode distinguir entre uma órbita periódica com ruído e determinismo aperiódico inter-ciclos dinâmicos. Possíveis origens de determinismo aperiódico inter-ciclos dinâmicos inclui: aperiódicidades lineares ou dinâmicas não-lineares, ou caos. Alguns modelos caóticos são encontrados na literatura. O sistema de Rössler é um exemplo de comportamento dinâmico caótico, cujas equações dinâmicas não-lineares são dadas por,

$$\begin{aligned}\dot{x} &= -y - z \\ \dot{y} &= x + ay \\ \dot{z} &= b + z(x - c)\end{aligned}$$

em que $a = 2$, $b = 4$ e $c = 0,398$ com condições iniciais $x(0) = y(0) = z(0) = 0,1$ exibe caos e para $c = 0,39095$ o sistema de Rössler exibe comportamento com ciclo limite de período 6.

Para diferentes valores dos parâmetros este sistema exibe caos. O regime de integração deste sistema com 10000 dados é de 0.2 unidades temporais e os primeiros 1000 pontos são descartados para evitar incluir estados transientes. São investigadas duas séries: a série caótica e a série de período 6 de Rössler, em ambas as séries são adicionadas ruídos gaussianos dinâmicos de desvio padrão 0,05 a cada passo, para cada uma das componentes.

É empregado primeiramente o algoritmo sub-rogado sugerido por Theiler e Rapp (1996) com embaralhamento de ciclos (*cycle shuffle* - CS), gerando 39 sub-rogados. São produzidos embaralhamentos dos ciclos individuais dentro da série caótica de Rossler com ruído dinâmico de desvio padrão 0,05. A Figura 4.65 mostra uma janela da série caótica de Rössler com o ruído dinâmico. Na Figura 4.66 são demonstrados os gráficos em (a) da série original, em (b) os sub-rogados gerados com rupturas nos ciclos por picos e sem deslocamento vertical, em (c) com rupturas nos ciclos por picos e com deslocamento vertical, em (d) com rupturas nos ciclos por vales e sem deslocamento vertical e em (e) com rupturas nos ciclos por vales e com deslocamento vertical.

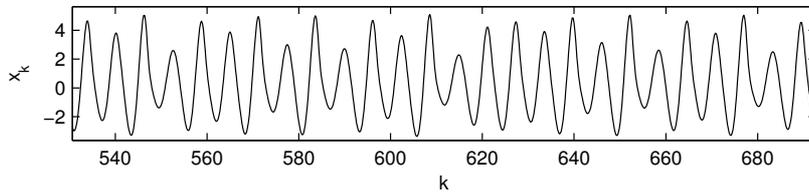


Figura 4.65: Janela do sistema de Rossler caótico com ruído dinâmico.

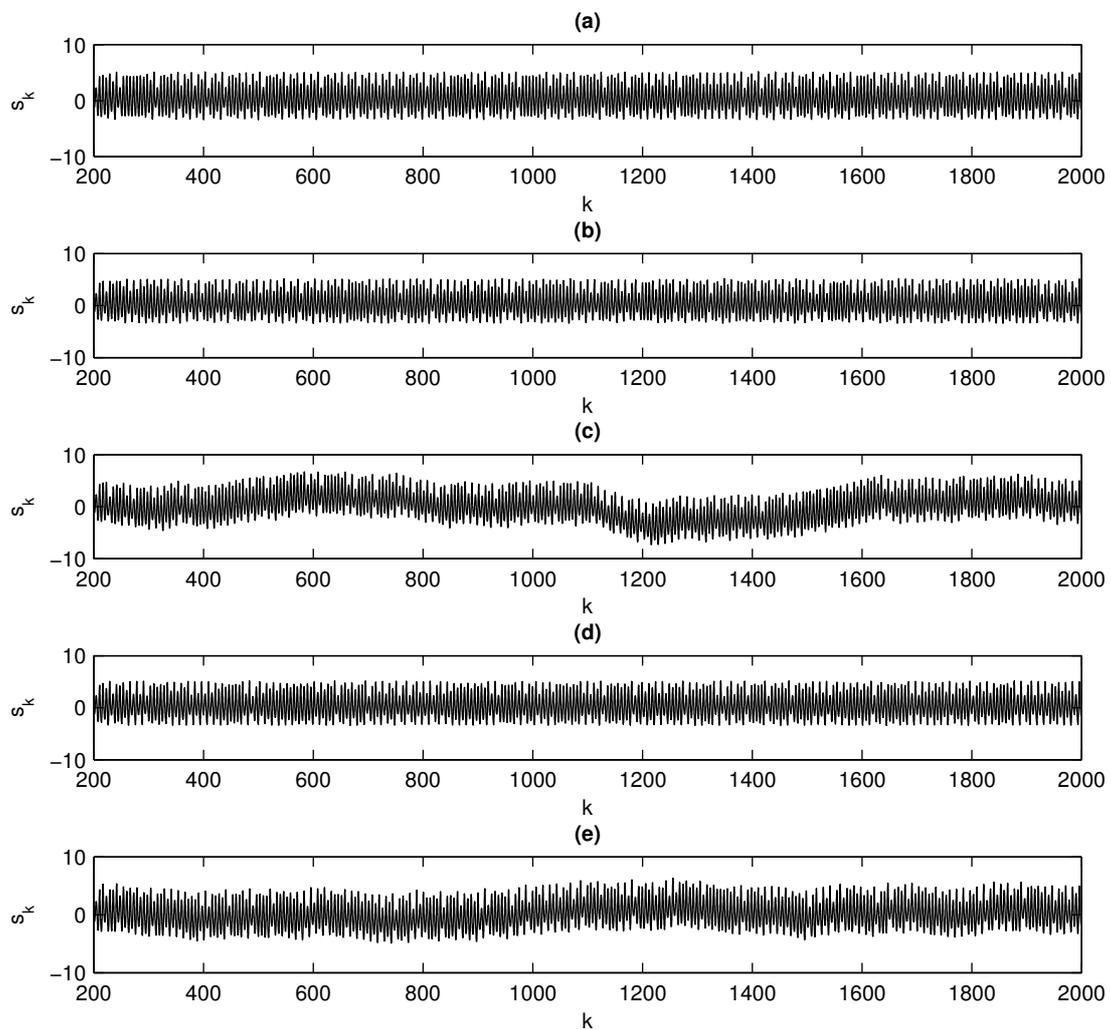


Figura 4.66: Sistema de Rössler: (a) série original, (b) CS (picos) sem deslocamento vertical, (c) CS (picos) com deslocamento vertical, (d) CS (vales) sem deslocamento vertical, (e) CS (vales) com deslocamento vertical.

A FAC e a IMM são ilustradas na Figura 4.67 para cada sub-rogado gerado, conforme respectivamente citado e demonstrado na Figura 4.66.

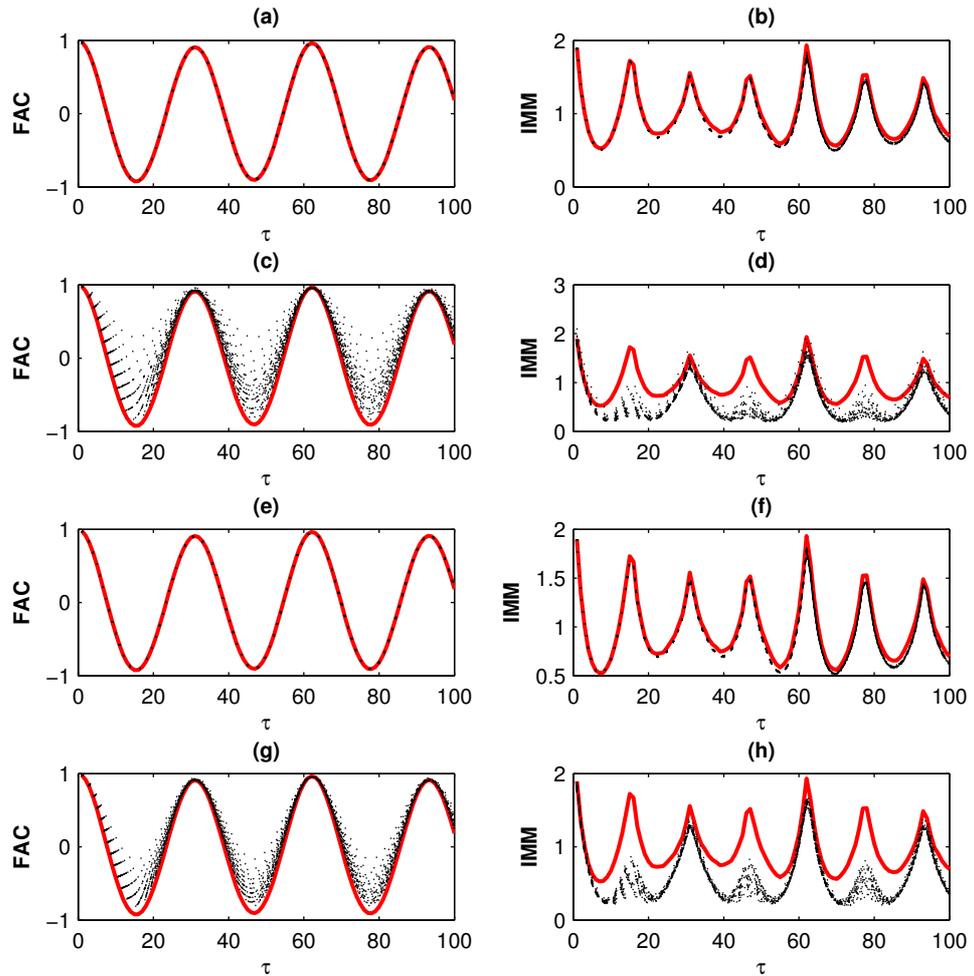


Figura 4.67: FAC e IMM do sistema de Rössler com ruído dinâmico (desvio padrão de 0,05) para (a) e (b) sub-rogados CS (picos) sem deslocamento vertical, em (c) e (d) sub-rogados CS (picos) com deslocamento vertical, em (e) e (f) sub-rogados CS (vales) sem deslocamento vertical e em (g) e (h) sub-rogados CS (vales) com deslocamento vertical.

Nota-se claramente termos longos correlacionados nos gráficos da FAC da Figura 4.67(a) e 4.67(e), contudo os gráficos da IMM para os sub-rogados CS com e sem deslocamento vertical e com rupturas dos ciclos por picos ou vales rejeitam a H_0 da correlação entre os ciclos (inter-ciclos), podendo afirmar que a série original possui dinâmica determinística possivelmente (95%) não-linear, como de fato ocorre em sistemas caóticos (não-lineares).

Um outro método de dados sub-rogados pseudo-periódicos (PPS) também pode ser aplicado para testes contra a hipótese nula de uma órbita periódica com ruído não correlacionado para um número muito grande de sistemas experimentais que exibem comportamento pseudo-periódico, tais como estes apresentados pelo sistema de Rössler. Este algoritmo pode distinguir entre uma órbita periódica com ruído e determinismo aperiódico inter-ciclos dinâmicos (caos). Duas séries são investigadas: a série caótica e a série de período 6 de Rössler, em ambas as séries são adicionadas ruídos gaussianos dinâmicos de desvio padrão 0,05.

De acordo com o teorema de Takens (1981), a trajetória completa de um sistema em um espaço de fases pode ser reconstruída a partir da medida de uma única variável independente. Ou seja, é possível definir um espaço de fase de baixa dimensão que capture a dinâmica em uma estrutura geométrica imersa nesse espaço.

O conjunto geométrico imerso é chamado de atrator reconstruído e ele é topologicamente equivalente ao atrator que seria produzido pela solução numérica do sistema dinâmico de equações, caso elas fossem conhecidas. As duas séries temporais, junto com os dados sub-rogados PPS (39 gerados para cada série) e os atratores reconstruídos (por meio da imersão) da componente x e para um dos dados sub-rogados PPS gerado das duas séries temporais são ilustradas nas Figuras 4.68 e 4.69.

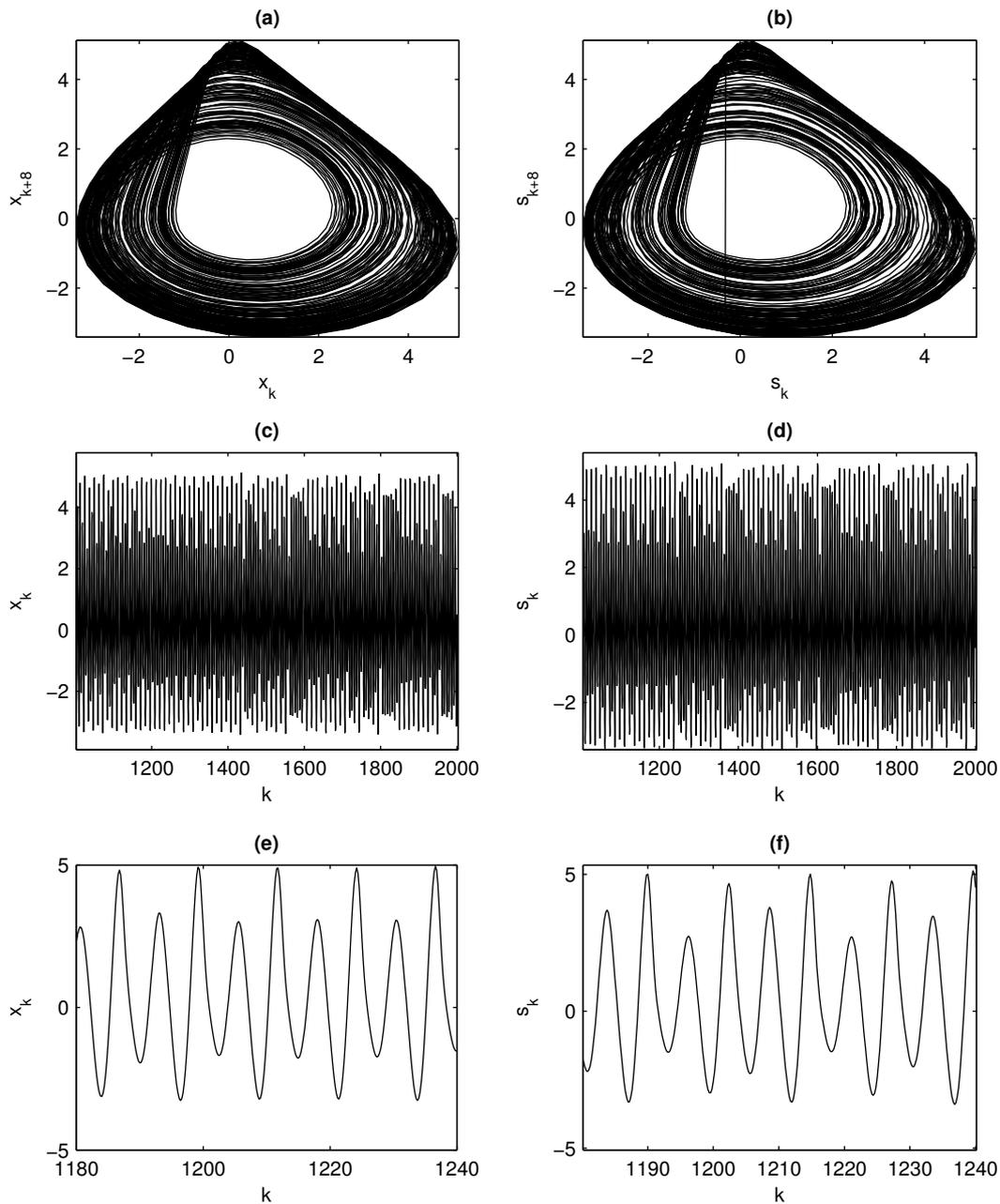


Figura 4.68: Dados sub-rogados PPS gerados da componente x da série temporal caótica do sistema de Rössler com ruído dinâmico de desvio padrão 0,05. No topo do painel mostram os atratores reconstruídos ($d_e = 3$, $\tau = 8$) por meio da componente x : (a) da série original e (b) um dos dados sub-rogados PPS ($\rho = 0.005$). Também é ilustrado: (c) a série temporal original, em (d) um dos dados sub-rogados PPS e uma pequena janela em (e) da série original e em (f) do dado sub-rogado PPS. Note que as séries temporais parecem visualmente indistinguível e os atratores reconstruídos são qualitativamente semelhantes (embora não idênticos).

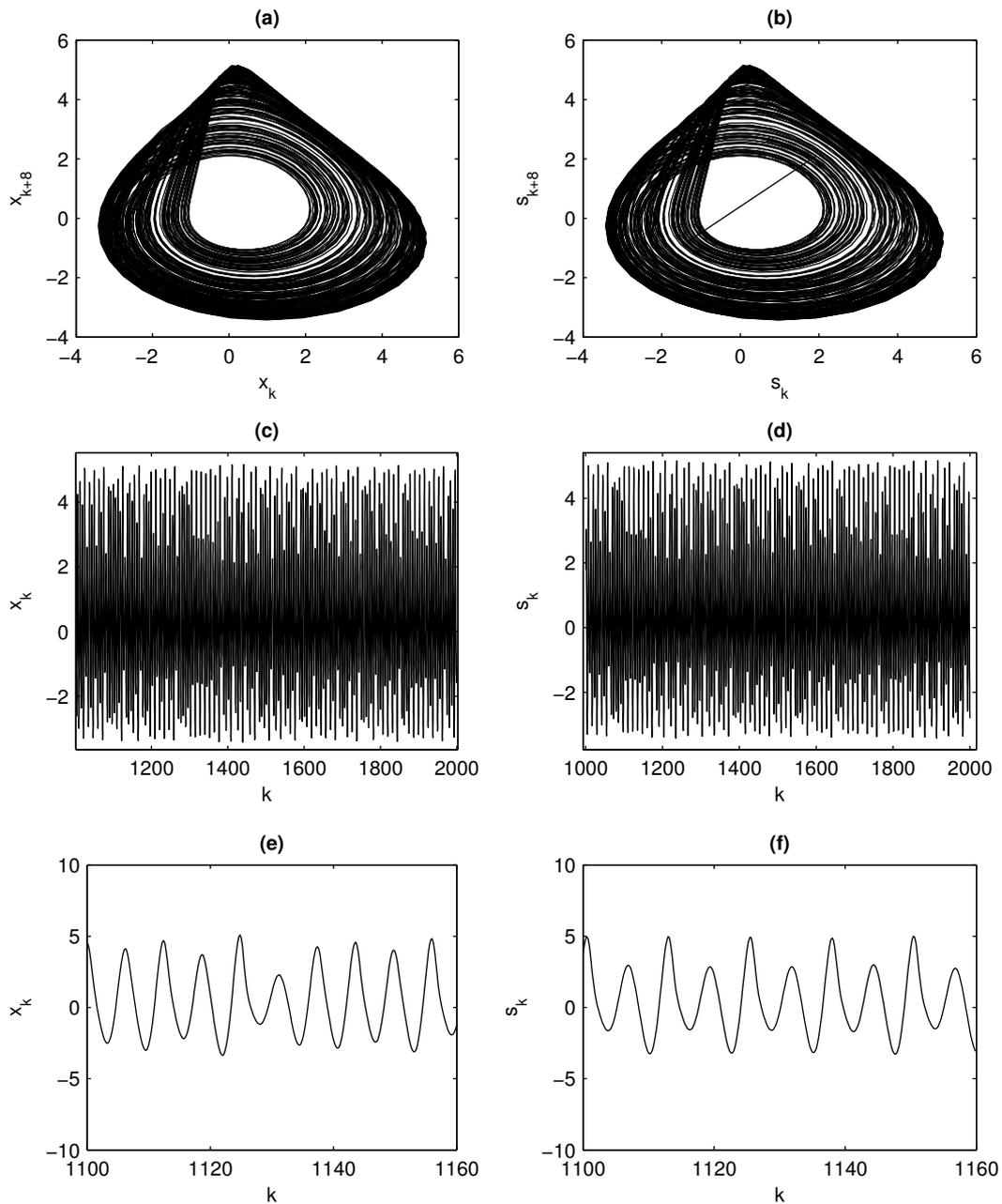


Figura 4.69: Dados sub-rogados PPS gerados da componente x da série temporal de Rössler de período 6 com ruído dinâmico gaussiano de desvio padrão 0,05. No topo do painel mostram os atratores reconstruídos ($d_e = 3$, $\tau = 8$) por meio da componente x : (a) da série original e (b) um dos dados sub-rogados PPS ($\rho = 0.001$). Também é ilustrado: (c) a série temporal original, em (d) um dos dados sub-rogados PPS e uma pequena janela em (e) da série original e em (f) do dado sub-rogado PPS. Note que a série temporal se parece visualmente indistinguível e os atratores reconstruídos são qualitativamente semelhantes (embora não idênticos).

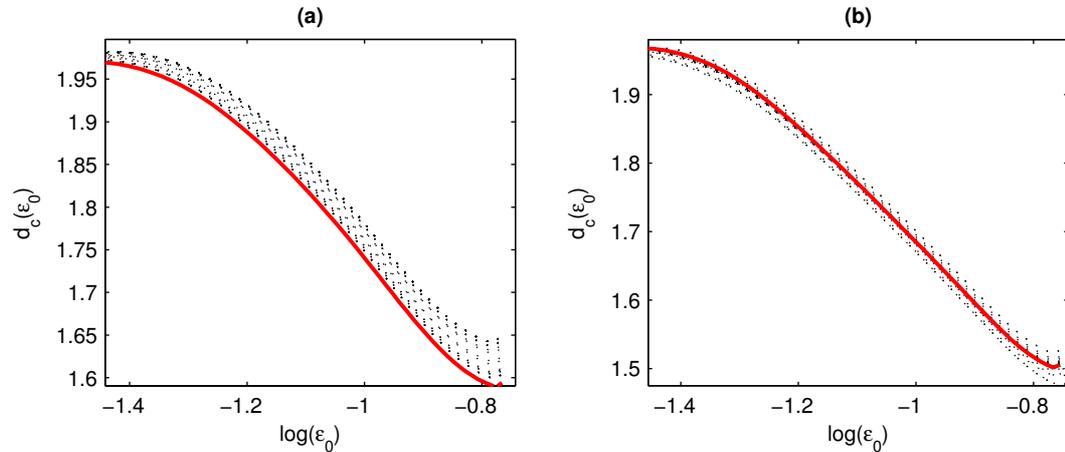


Figura 4.70: Comparação da dimensão de correlação: em (a) componente x do sistema de Rössler caótico ($d_e = 3$ e $\tau = 8$) contaminado com ruído dinâmico gaussiano de desvio padrão 0,05 e os dados sub-rogados PPS gerados com $\rho = 0.005$ e em (b) componente x do sistema de Rössler de período 6 contaminado com ruído dinâmico gaussiano de desvio padrão 0,05 ($d_e = 3$ e $\tau = 8$) e os dados sub-rogados PPS gerados com $\rho = 0.005$. A dimensão de correlação estimada $d_c(\epsilon_0)$ é plotada contra $\log(\epsilon_0)$. O valor de $d_c(\epsilon_0)$ é demonstrado para a série original como uma linha sólida e para cada sub-rogado PPS gerado como uma linha pontilhada. Para a série temporal caótica com ruído dinâmico, painel (a), a série original e os sub-rogados são claramente distinguíveis, porém para o sistema periódico (Rössler de período 6) contaminado com ruído dinâmico, painel (b), não há nenhuma distinção.

Todos os métodos de dados de sub-rogados requerem uma estatística de teste para comparar os dados originais com os sub-rogados gerados. É escolhido aplicar o algoritmo de estimação da dimensão de correlação descrito por Judd (1992, 1994). Este algoritmo calcula a dimensão de correlação d_c em função de ϵ_0 , ou seja, $d_c(\epsilon_0)$. Segundo Judd o algoritmo é robusto quando empregado como um teste estatístico com dados sub-rogados. Os resultados da análise da dimensão de correlação são mostradas na Figura 4.70. Na Figura 4.70(a) mostram a série original e os sub-rogados PPS gerados são ligeiramente distintos: o algoritmo PPS identifica dinâmica determinística não-periódica (característico de sistemas caóticos). Na análise da série de Rössler de período 6 contaminada com ruído dinâmico na Figura 4.70(b) a dimensão de correlação calculada mostra uma situação oposta. Neste caso o algoritmo PPS não pode distinguir entre a série original e os sub-rogados PPS gerados, e então corretamente conclui que o dado sob consideração é gerado por uma órbita periódica contaminada com ruído. Significativamente, a série original (Rössler de período 6 contaminado com ruído dinâmico) e os dados sub-rogados PPS mostrados Figura 4.70(b) são qualitativamente semelhante um

para com o outro.

4.11 Um estudo de caso

Para um entendimento mais amplo e esclarecedor da aplicabilidade do teste de hipótese nula utilizando o método de dados sub-rogados é proposto um estudo de caso. Tal estudo é importante para identificar passo à passo as etapas na investigação da estrutura geradora da série temporal original. O sistema caótico de Lorenz é proposto para o estudo de caso passo à passo,

$$\begin{aligned}\dot{x} &= -\sigma(x - y), \\ \dot{y} &= rx - y - xz, \\ \dot{z} &= xy - bz,\end{aligned}\tag{4.27}$$

em que $(x, y, z) \in \mathcal{R}^3$, $\sigma = 10$, $b = \frac{8}{3}$ e $r = 28$ com condições iniciais $x(0) = y(0) = z(0) = 0.1$ que exibe comportamento caótico, vista na Figura A.1 da seção A.2 em anexos, para possuir comportamento com órbita periódica qualquer outro valor real de r deve ser proposto.

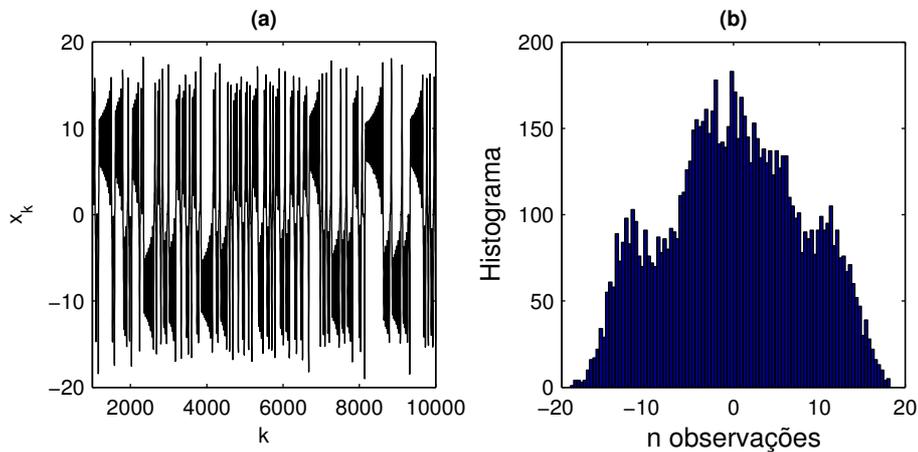


Figura 4.71: Componente x da série temporal caótica do sistema de Lorenz com inovações de laplace ($\theta = 0$ e $\lambda = 1$) com desvio padrão de 0,05 em (a) e em (b) um histograma da série temporal proposta.

Além da estrutura dinâmica é adicionado inovações (ruído) de acordo com a distribuição de laplace ($\theta = 0$ e $\lambda = 1$) com desvio padrão de 0,05 em cada

componente do sistema de Lorenz. O regime de integração para este sistema é de 0,02 unidades temporais, apenas a componente x do sistema é tomada para teste por meio de 10000 pontos gerados e cujos primeiros 1000 dados são descartados. A série temporal da componente x com inovações Laplacianas é ilustrada na Figura 4.71.

São gerados em todas as etapas a serem analisadas para o teste de hipótese estatística 99 dados sub-rogados ($\alpha = 0,01$). A primeira etapa de análise é verificar se a série temporal é um conjunto de variáveis aleatórias independentes e identicamente distribuídas ou independentemente distribuídas. A análise para o teste estatístico pode ser feita utilizando o hipótese nula NH_0 do algoritmo 0 abordando apenas a técnica IID e a hipótese nula NH_{SS} do algoritmo SS. A Figura 4.72 ilustra a análise estatística por meio da FAC e IMM aplicando o algoritmo 0 (IID e SS). É importante observar que nessa primeira etapa ocorre a rejeição da hipótese nula NH_0 e NH_{SS} , sendo então, provavelmente (99%), que a série temporal proposta possui algum tipo de dinâmica determinística.

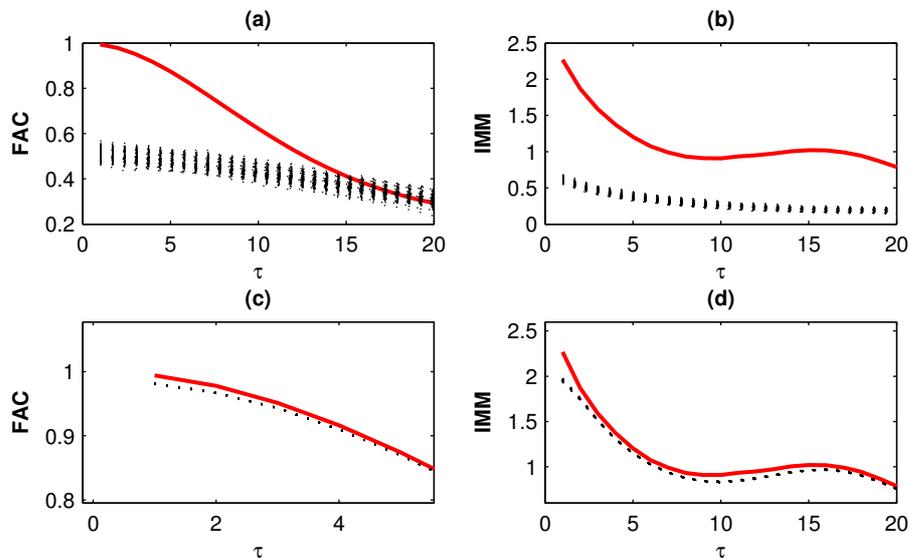


Figura 4.72: Gráfico da FAC e da IMM para a série caótica de Lorenz com inovações de laplace ($\theta = 0$ e $\lambda = 1$): (a) e (b) para os sub-rogados IID; (c) e (d) para os sub-rogados SS. A linha sólida corresponde a série original e as linhas pontilhadas são as 99 séries sub-rogadas geradas.

Na segunda etapa é verificada a possibilidade da série temporal em análise ser possivelmente (99%) um processo gaussiano linearmente autocorrelacionado. Para isso, é empregada a hipótese nula NH_1 por meio da aplicação do algoritmo 1

(FT) na série temporal original. Lembrando que tal teste estatístico verifica a possibilidade da série temporal ser um processo estocástico linear (AR) ou processo estocástico linear com ruído gaussiano (ARMA), sendo limitado apenas para ruídos gaussianos. É notável na Figura 4.73 que o teste estatístico não-paramétrico falha (99%), ou seja, rejeita a hipótese NH_1 .

Essa rejeição nos relata três possíveis razões: a série temporal possui alguma dinâmica linear, mas os dados sub-rogados gerados não são apropriados para análise (espectros de potências iguais, mas as distribuições das amplitudes diferentes); a série temporal possui alguma dinâmica linear, mas não necessariamente gaussiana; ou a que a série temporal possui possivelmente algum tipo de não-linearidade (estática ou dinâmica).

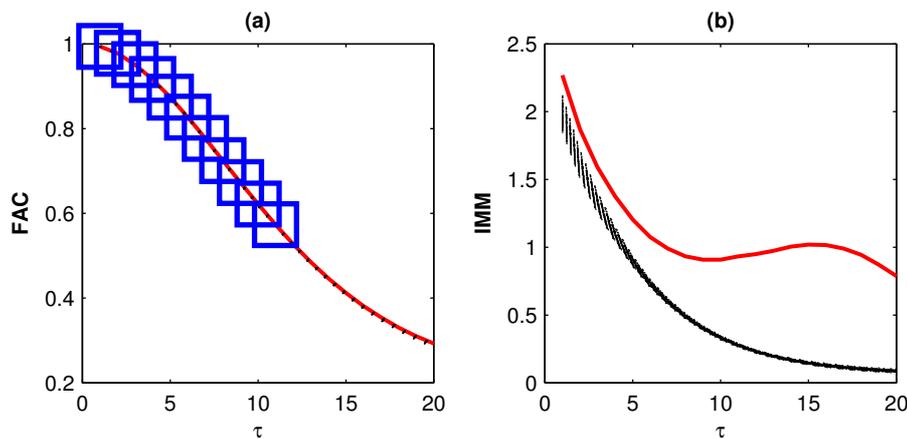


Figura 4.73: Gráfico da FAC e da IMM para a série temporal caótica de Lorenz com inovações de laplace ($\theta = 0$ e $\lambda = 1$) para os sub-rogado FT. A linha sólida corresponde a série original e as linhas pontilhadas são as 99 séries sub-rogadas geradas. As observações feitas com retângulos nas figuras são pontos ligeiramente fora da distribuição dos dados sub-rogados.

A terceira etapa é por meio da análise da estatística de teste da hipótese nula NH_2 . Tal hipótese propõe uma transformação (“ h ”) para gerar dados sub-rogados condizentes com a série temporal original (mesmo espectro de potência e mesma distribuição de amplitude). O algoritmo 2 é dividido aqui em três grupos de técnicas geradoras de dados sub-rogados: o AAFT; o IAAFT-1 e 2; o CAAFT e STAP. A princípio, esses grupos foram divididos dessa forma devido as semelhanças na construção proposta para cada um dos dados sub-rogados baseada na série temporal original.

Nessa terceira etapa aplica-se apenas o AAFT cujo teste de hipótese nula relata que os dados sub-rogados são gerados por meio de uma transformação estática

não-linear e monotônica de um processo linear gaussiano (série temporal original), ou seja, uma forma mais refinada de assegurar a linearidade, mas lembrando ainda que são para processos gaussianos. A Figura 4.74 ilustra a aplicação dos testes estatísticos (FAC e IMM), sendo possível ainda observar a rejeição (99%) da hipótese nula proposta.

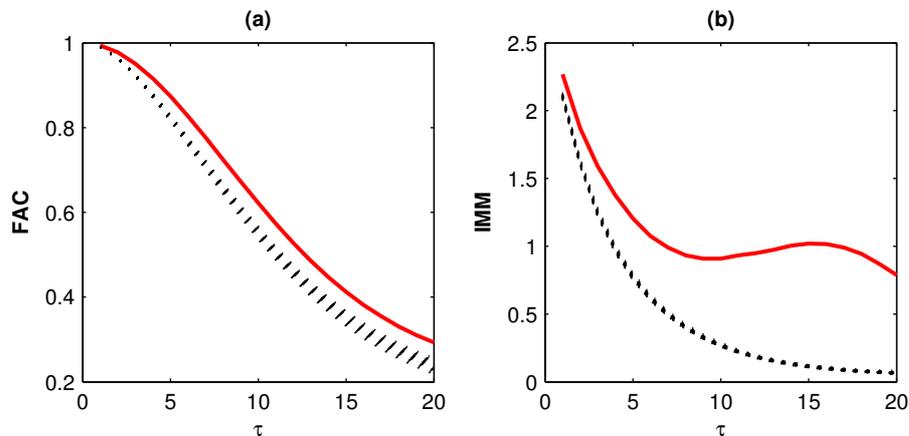


Figura 4.74: Gráfico da FAC e da IMM para a série temporal caótica de Lorenz com inovações de laplace ($\theta = 0$ e $\lambda = 1$) para os sub-rogado AAFT. A linha sólida corresponde a série original e as linhas pontilhadas são as 99 séries sub-rogadas geradas.

A quarta etapa é aplicada ainda a hipótese nula NH_2 , mas o algoritmo 2 com as técnicas IAAFT-1 (mesmo espectro de potência) e IAAFT-2 (mesma distribuição da amplitude). Lembrando que agora a transformação estática não-linear e monotônica (“ h ”) do processo linear é realizada por meio de várias interações até aproximar o espectro de potência (IAAFT-1) ou a distribuição de amplitude (IAAFT-2) da série temporal original. Essa é a primeira tentativa para encontrar evidências de não-linearidades nas séries temporais.

A Figura 4.75 ilustra a análise da estatística de teste não-paramétrica com a FAC e a IMM para os dados sub-rogados do IAAFT-1 e do IAAFT-2. É importante notar que ocorre a rejeição (99%) da hipótese nula, sendo possivelmente que tal série temporal proposta tenha alguma tipo de não-linearidade (estática ou dinâmica).

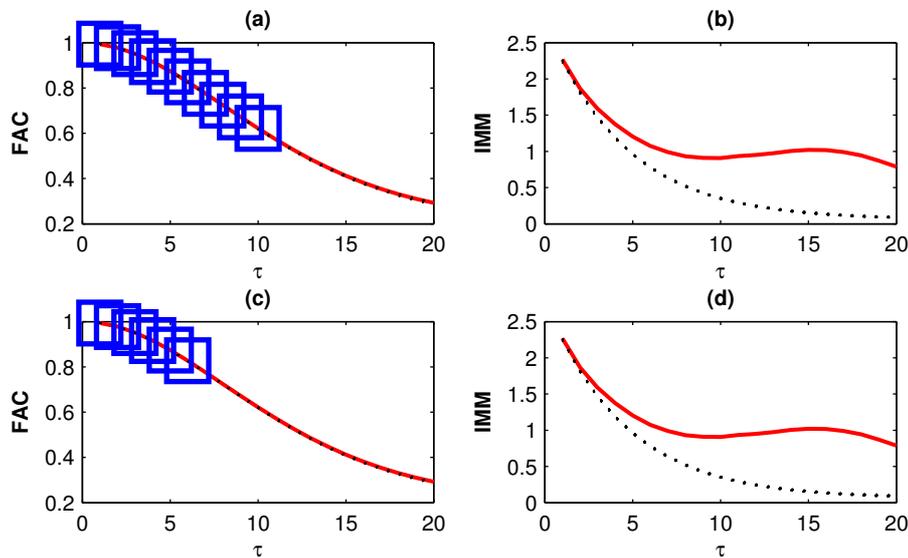


Figura 4.75: Gráfico da FAC e da IMM para a série temporal caótica de Lorenz com inovações de laplace ($\theta = 0$ e $\lambda = 1$): (a) e (b) para os sub-rogados IAAFT-1; (c) e (d) para os sub-rogados IAAFT-2. A linha sólida corresponde a série original e as linhas pontilhadas são as 99 séries sub-rogadas geradas. As observações feitas com retângulos nas figuras são pontos ligeiramente fora da distribuição dos dados sub-rogados.

A quinta etapa e última para a possível confirmação estatística ou não da dinâmica linear da série temporal original é por meio das técnicas CAAFT e STAP do algoritmo 2. Lembrando que a construção dos dados sub-rogados utilizando tais técnicas não restringem a hipótese nula de que a série temporal original seja um processo linear gaussiano apenas, possibilitando uma gama de análise de séries temporais não-gaussiana o que de fato ocorre na prática em quase todos os processos reais.

A Figura 4.76 ilustra a análise da estatística de teste não-paramétrica com a FAC e a IMM para os dados sub-rogados gerados pelo CAAFT e STAP, respectivamente. Nessa ilustração também mostra variação das FAC e das IMM do conjunto de dados sub-rogados gerados por estes algoritmos. Consequentemente, a análise destas estatísticas de testes confirmam possivelmente (99%), por meio da rejeição da hipótese nula, que a série temporal original proposta tenha algum tipo de não-linearidade (estática ou dinâmica).

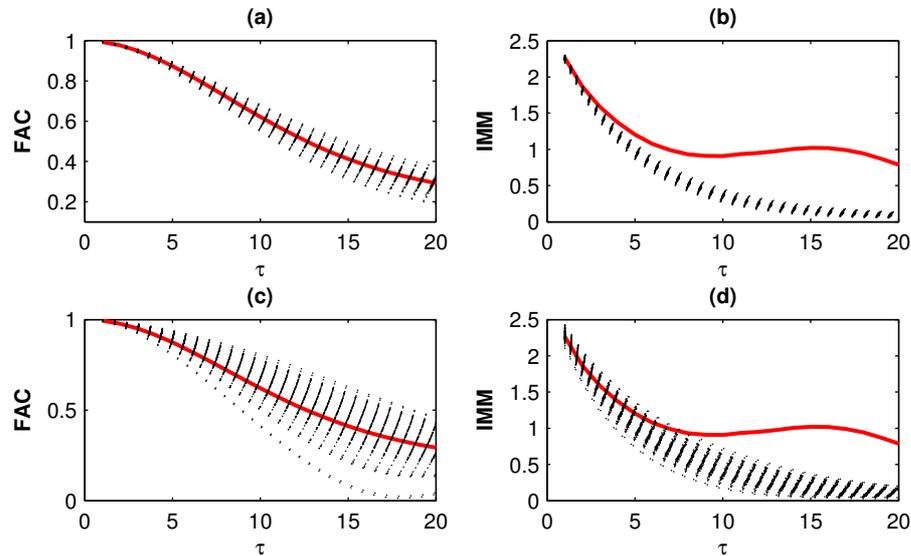


Figura 4.76: Gráfico da FAC e da IMM para a série temporal caótica de Lorenz com inovações de laplace ($\theta = 0$ e $\lambda = 1$): (a) e (b) para os sub-rogados CAAFT; (c) e (d) para os sub-rogados STAP. A linha sólida corresponde a série original e as linhas pontilhadas são as 99 séries sub-rogadas geradas.

É importante enfatizar que estas 5 etapas de análise propostas são necessárias para o teste de hipóteses nulas lineares e que tais rejeições das três hipóteses nulas (NH_0 ou NH_{SS} , NH_1 e NH_2) são necessárias para confirmar estatisticamente possíveis (99% nesse caso) não-linearidades, exceto para alguns casos de não-linearidades estática, como visto anteriormente na seção 4.6.

Alguns dados são incompatíveis com ruídos lineares ou uma transformação estática monotônica do mesmo. Após a confirmação estatística (99%) da possível não-linearidade presente será investigada nas etapas seguintes se a série temporal possui alguma dinâmica pseudo-periódica entre os ciclos, ou seja, a série temporal é cíclica.

Logo, a sexta etapa proposta é então a aplicação do algoritmo CS (*cycle-shuffle*), cuja hipótese nula verifica a correlação entre ciclos da série temporal original por meio da construção de dados sub-rogados embaralhando os ciclos encontrados pelas rupturas por meio de picos ou vales e com ou sem deslocamento vertical dos pontos de rupturas encontrados, ou seja, dirija-se a hipótese que não há nenhuma correlação dinâmica entre ciclos (NH_{CS}).

A Figura 4.77 ilustra a análise da estatística de teste não-paramétrica da FAC e da IMM. Em todos os gráficos mostrados ocorrem a rejeição da hipótese nula

NH_{CS} , portanto estatisticamente (99%) existe correlação dinâmica entre os ciclos.

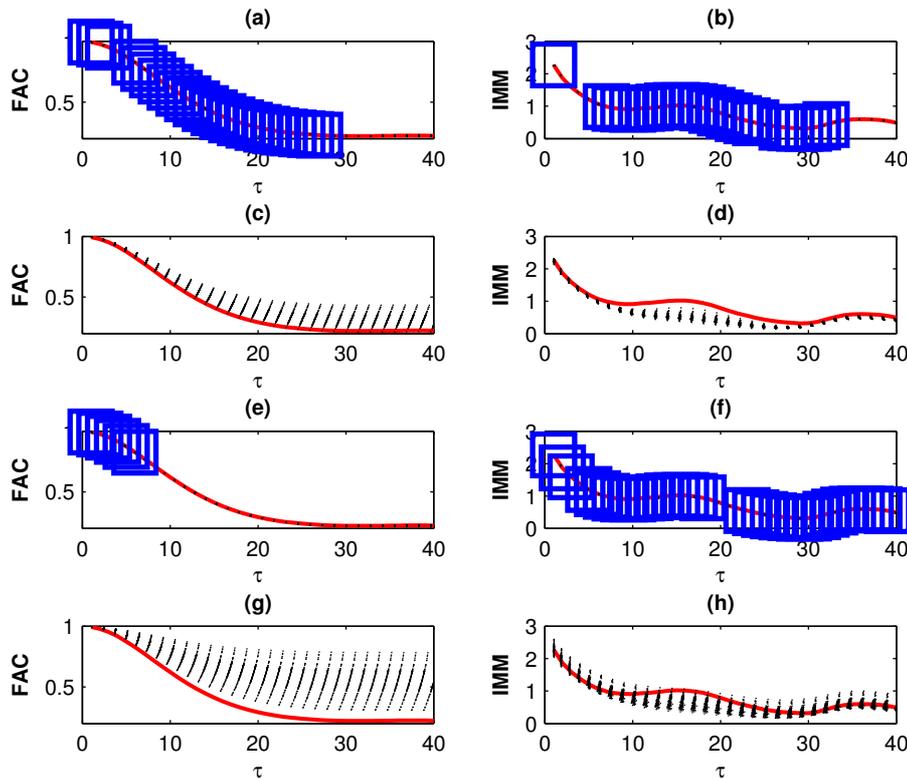


Figura 4.77: Gráfico da FAC e da IMM para a série temporal caótica de Lorenz com inovações de laplace ($\theta = 0$ e $\lambda = 1$) gerando dados sub-rogados CS: (a) e (b) com ruptura nos picos e sem deslocamento vertical; (c) e (d) com ruptura nos picos e com deslocamento vertical; (e) e (f) com ruptura nos vales e sem deslocamento vertical; (g) e (h) com ruptura nos vales e com deslocamento vertical. A linha sólida corresponde a série original e as linhas pontilhadas são as 99 séries sub-rogadas geradas. As observações feitas com retângulos nas figuras são pontos ligeiramente fora da distribuição dos dados sub-rogados.

A sétima e última etapa é uma investigação se a dinâmica não-linear da série temporal encontrada possa ser caótica. O algoritmo empregado é o PPS (sub-rogado pseudo-periódico). Este método pode ser aplicado para testes contra a hipótese nula NH_{PPS} de uma órbita periódica com ruído não correlacionado para um número muito grande de sistemas experimentais com comportamento pseudo-periódico.

O algoritmo PPS pode distinguir entre uma órbita periódica com ruído (dinâmico ou não) e determinismo aperiódico inter-ciclos dinâmicos, caso ocorra a rejeição da hipótese NH_{PPS} .

Possíveis origens de determinismo aperiódico inter-ciclos dinâmicos inclui: não-periódicos lineares ou dinâmicas não-lineares, ou caos. Lembrando que a etapa anterior (sexta etapa) estatisticamente confirmou correlação entre os supostos ciclos na série temporal sob estudo, ou seja, existe alguma dinâmica entre os ciclos.

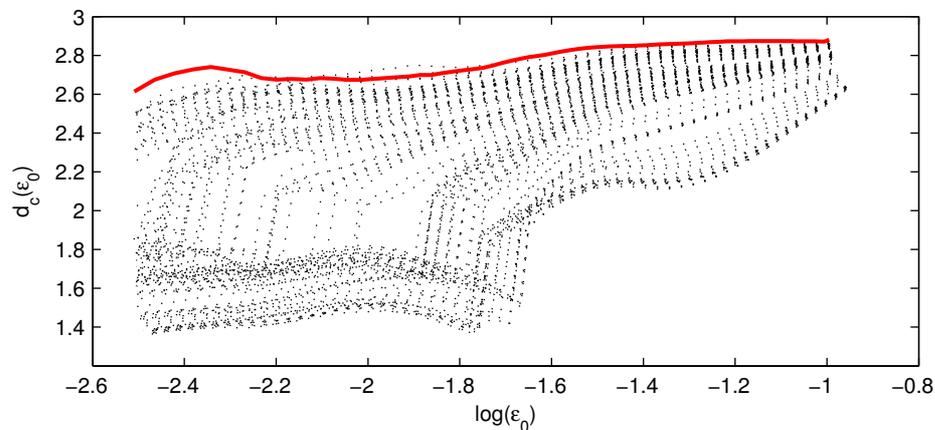


Figura 4.78: Comparação da dimensão de correlação para a série temporal da componente x do sistema de Lorenz caótico ($d_e = 3$ e $\tau = 251$) contaminado com inovações laplacianas ($\theta = 0$ e $\lambda = 1$) com desvio padrão 0,05 e os dados sub-rogados PPS gerados com $\rho = 0,05$. A dimensão de correlação estimada $d_c(\epsilon_0)$ é plotada contra $\log(\epsilon_0)$. O valor de $d_c(\epsilon_0)$ é demonstrado para a série temporal original como uma linha sólida e para cada sub-rogado PPS gerado como uma linha pontilhada. A série temporal original e os sub-rogados PPS gerados são claramente distinguíveis.

A Figura 4.78 demonstra graficamente que a série temporal original e as 99 séries sub-rogadas PPS geradas são claramente distinguíveis, ou seja, ocorre a rejeição da hipótese nula NH_{PPS} . Portanto, estatisticamente (99%) conclui-se que a série temporal sob análise possui dinâmica não-linear aperiódica (sistema caótico).

Todas as etapas propostas não precisam ser necessariamente realizadas, mas a escala ascendente de análise das etapas deve ser respeitada, ou seja, caso não ocorra em alguma das etapas a rejeição das respectivas hipóteses nulas propostas pode-se afirmar (após a análise da etapa seguinte), com um certa probabilidade estatística, a estrutura ou mecanismo gerador da série temporal como sendo a proposta pela hipótese nula empregada na respectiva etapa. É importante lembrar

que tais afirmações ou negações são idealizadas estatisticamente e não com 100% de certeza, existindo uma possibilidade de erro (erro tipo I).

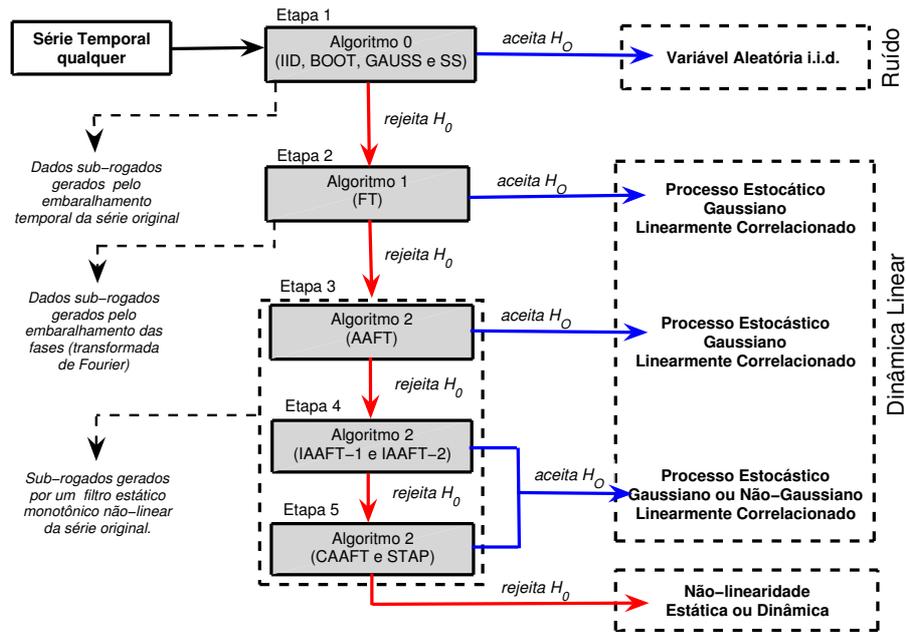


Figura 4.79: Fluxograma com o procedimento geral para teste de hipótese nula (H_0) utilizando os métodos de dados sub-rogados algoritmo 0 (IID e SS), algoritmo 1 (FT) e algoritmo 2 (AFT, IAAFT-1, IAAFT-2, CAAFT e STAP) aplicados a qualquer série temporal.

4.12 Estudo de caso “blind”

A idéia principal deste estudo de caso “blind” (as cegas) é verificar a dinâmica do sistema sem conhecimento do processo, ou seja, sem nenhuma informação sobre o comportamento da série temporal sob análise, a princípio. São colhidos para teste 31025 dados, sendo que os primeiros 1025 dados são descartados. A série temporal é ilustrada na Figura 4.80, juntamente com seu histograma.

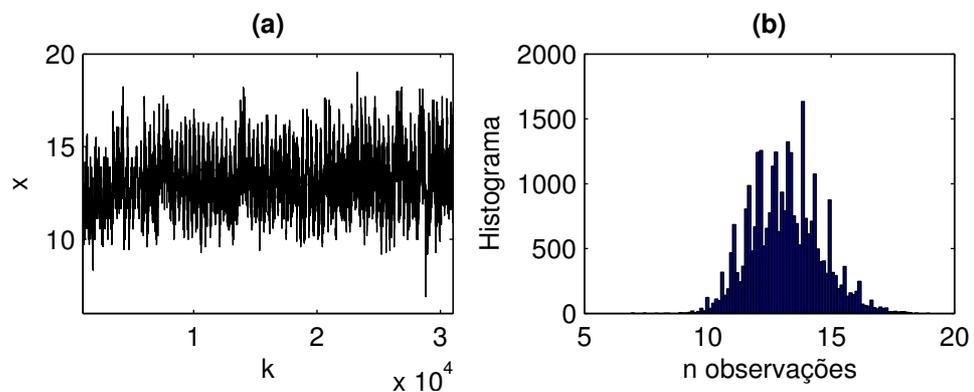


Figura 4.80: Série temporal para o caso “blind”: em (a) a série temporal e em (b) um histograma.

São gerados em todas as etapas a serem analisadas para o teste de hipótese estatística 99 dados sub-rogados ($\alpha = 0,01$).

A primeira etapa de análise é verificar se a série temporal é um conjunto de variáveis aleatórias independentes e identicamente distribuídas ou independentemente distribuídas. A análise para o teste estatístico pode ser feita utilizando o hipótese nula NH_0 do algoritmo 0 abordando apenas a técnica IID e a hipótese nula NH_{SS} do algoritmo SS.

A Figura 4.81 ilustra a análise estatística por meio da FAC e IMM aplicando o algoritmo 0 (IID e SS). É importante observar que nessa primeira etapa ocorre a rejeição da hipótese nula NH_0 e NH_{SS} , sendo então, provavelmente (99%), que a série temporal proposta possui algum tipo de dinâmica determinística.

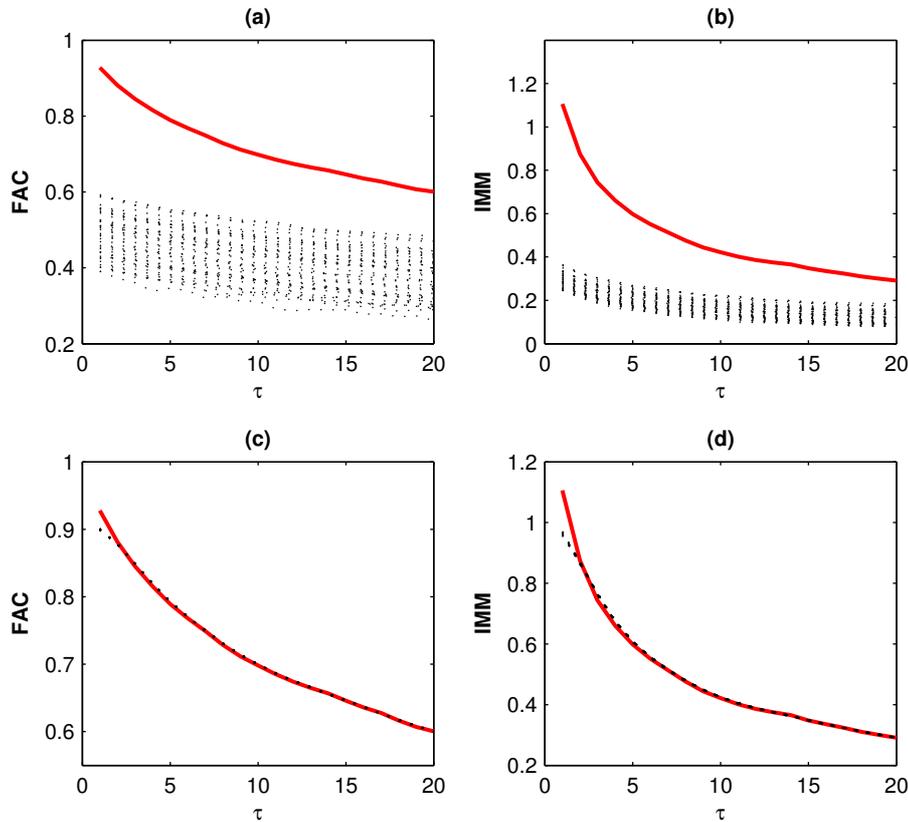


Figura 4.81: Gráfico da FAC e da IMM para a série do caso *blind*: (a) e (b) para os sub-rogados IID; (c) e (d) para os sub-rogados SS. A linha sólida corresponde a série original e as linhas pontilhadas são as 99 séries sub-rogadas geradas.

Na segunda etapa é verificada a possibilidade da série temporal sob análise ser possivelmente (99%) um processo gaussiano linearmente autocorrelacionado (NH_1). É notável na Figura 4.82 que o teste estatístico não-paramétrico não rejeita (99%), ou seja, aceita a hipótese NH_1 . Essa aceitação estatisticamente relata que a série temporal possui alguma dinâmica determinística linear com ruído gaussiano. Porém, não se pode tirar nenhuma conclusão sem analisar as outras técnicas sub-rogadas para o teste de hipótese nula.

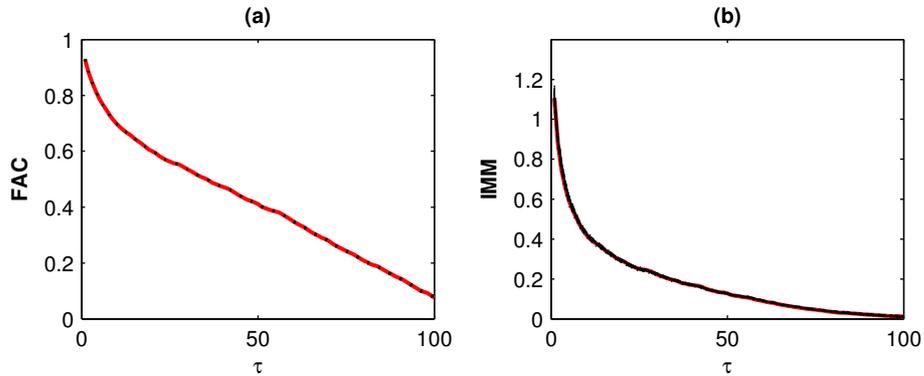


Figura 4.82: Gráfico da FAC e da IMM para a série do caso *blind* e para os sub-rogado FT. A linha sólida corresponde a série original e as linhas pontilhadas são as 99 séries sub-rogadas geradas.

Nessa terceira etapa aplica-se apenas o AAFT cujo teste de hipótese nula relata que os dados sub-rogados são gerados por meio de uma estática transformação não-linear e monotônica de um processo linear gaussiano (série temporal original), ou seja, uma forma mais refinada de assegurar a lineariedade, mas lembrando ainda que são para processos gaussianos.

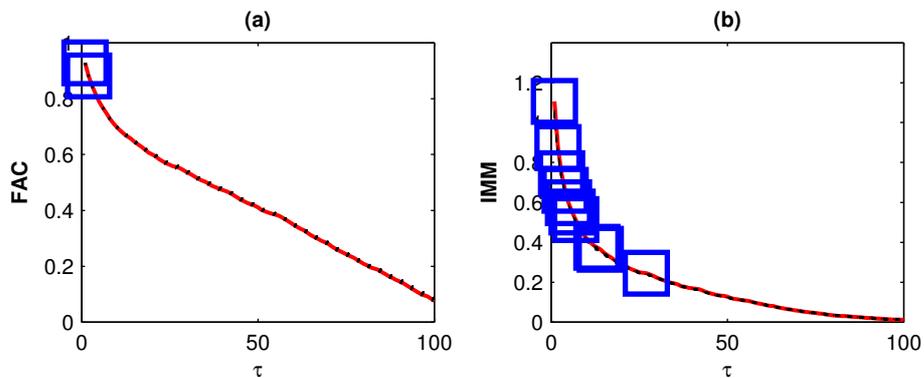


Figura 4.83: Gráfico da FAC e da IMM para a série do caso *blind* para os sub-rogado AAFT. A linha sólida corresponde a série original e as linhas pontilhadas são as 99 séries sub-rogadas geradas. As observações feitas com retângulos nas figuras são pontos fora da distribuição dos dados sub-rogados gerados.

A Figura 4.83 ilustra a aplicação dos testes estatísticos (FAC e IMM), sendo possível ainda observar a rejeição (99%) da hipótese nula proposta. Observe que essa decisão desmente a decisão anterior, ou seja, o algoritmo FT pode conduzir

a decisões errôneas, mas somente será confirmado tal erro aplicando as demais técnicas sub-rogadas do algoritmo 2 da hipótese nula NH_2 .

A quarta etapa é aplicada ainda a hipótese nula NH_2 , mas o algoritmo 2 com as técnicas IAAFT-1 (mesmo espectro de potência) e IAAFT-2 (mesma distribuição da amplitude). Lembrando que agora a estática transformação não-linear e monotônica (“ h ”) do processo linear é realizada por meio de várias interações até aproximar o espectro de potência (IAAFT-1) ou a distribuição de amplitude (IAAFT-2) da série temporal original.

A Figura 4.84 ilustra a análise da estatística de teste não-paramétrica com a FAC e a IMM para os dados sub-rogados do IAAFT-1 e do IAAFT-2. É importante notar que ocorre a rejeição (99%) da hipótese nula, sendo possivelmente que tal série temporal proposta tenha alguma tipo de não-linearidade (estática ou dinâmica).

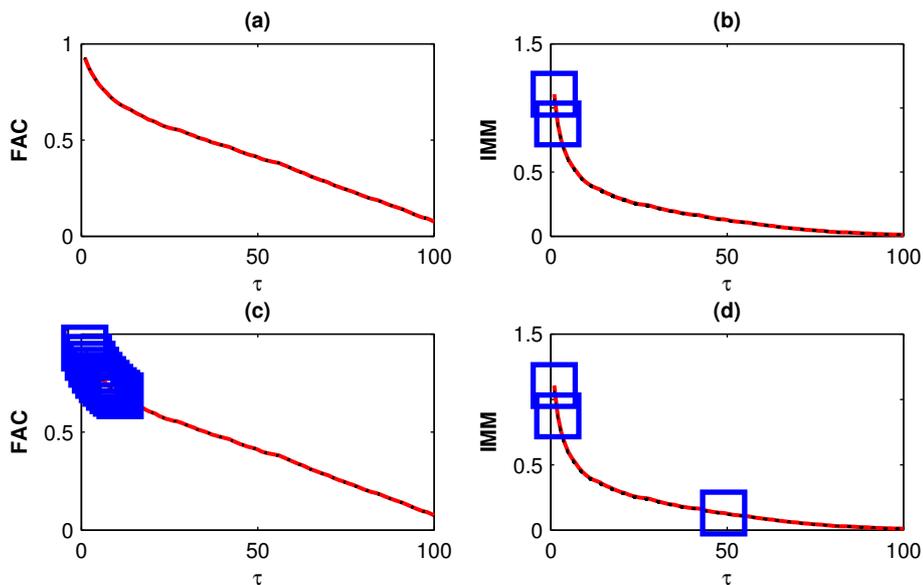


Figura 4.84: Gráfico da FAC e da IMM para a série do caso “blind” com IAAFT-1 e 2: (a) e (b) para os sub-rogados IAAFT-1; (c) e (d) para os sub-rogados IAAFT-2. A linha sólida corresponde a série original e as linhas pontilhadas são as 99 séries sub-rogadas geradas. As observações feitas com retângulos nas figuras são pontos ligeiramente fora da distribuição dos dados sub-rogados.

A quinta etapa e última para a possível confirmação estatística ou não da dinâmica linear da série temporal original é por meio das técnicas CAAFT e STAP do algoritmo 2, Figura 4.85. Lembrando que a construção dos dados sub-rogados utilizando tais técnicas não restringem a hipótese nula de que a série temporal

original seja um processo linear gaussiano apenas, possibilitando uma gama de análise de séries temporais não-gaussianas o que de fato ocorre na prática em quase todos os processos existentes.

A Figura 4.85 ilustra a análise da estatística de teste não-paramétrica com a FAC e a IMM para os dados sub-rogados gerados pelo CAAFT e STAP, respectivamente. Nessa ilustração também mostra que as estatísticas de testes não-paramétricas confirmam estatisticamente (99%), por meio da rejeição da hipótese nula NH_2 , que a série temporal original proposta tenha algum tipo de não-linearidade (estática ou dinâmica).

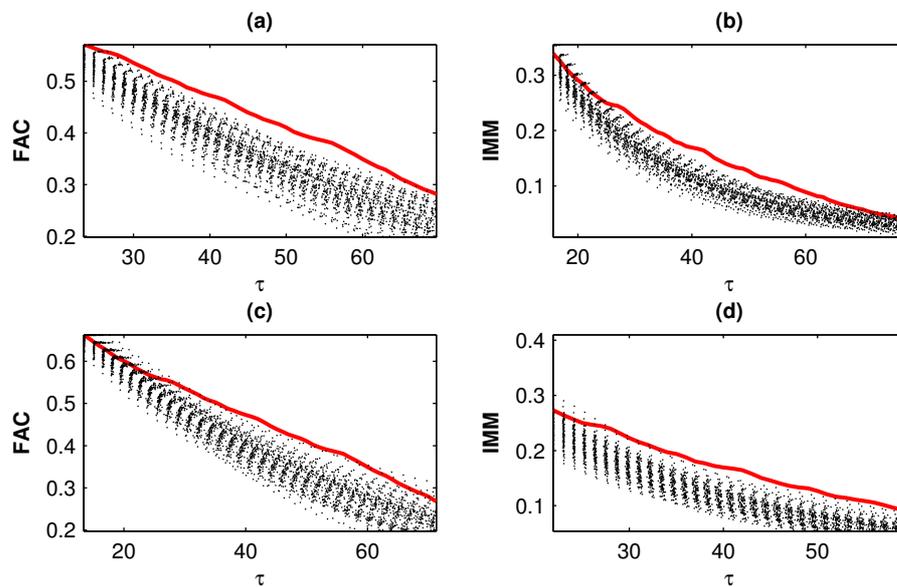


Figura 4.85: Gráfico da FAC e da IMM para a série do caso *blind*: (a) e (b) para os sub-rogados CAAFT; (c) e (d) para os sub-rogados STAP. A linha sólida corresponde a série original e as linhas pontilhadas são as 99 séries sub-rogadas geradas.

É importante salientar que a escolha do τ para análise da FAC e IMM deve ser feita de forma cuidadosa. Para pequenos valores de τ (até 20) algumas estatísticas de teste podem indicar a aceitação da hipótese nula e não a sua rejeição. Portanto, é importante ampliar o valor de τ para encontrar casos em que ocorra a rejeição da H_0 , principalmente para séries temporais com grande volume de dados amostrados, cujas correlações não-lineares podem ser mascaradas devido ao princípio do teorema do limite central (TLC), no anexo A, séries temporais muito grandes tendem a uma distribuição normal ou gaussiana.

Pode-se concluir que este estudo *blind* com dados não conhecidos revela, com base nas análises de dados sub-rogados construídos e testes estatísticos realizados, que o conjunto de dados da série temporal sob estudo é possivelmente (99%) um modelo determinístico não-linear (estático ou dinâmico), devido a rejeição das hipóteses nulas lineares analisadas em todas as etapas, menos a da etapa 2 do algoritmo FT. Essa etapa enganaria a decisão da H_0 proposta. Por isso, deve-se analisar pelo menos uma das etapas seguintes para a aceitação ou não da decisão relatada na etapa anterior.

É revelado após efetuada a análise dos resultados a estrutura da série temporal *blind*. Os aspectos de não-lineares observados são da temperatura da superfície do mar em Monterey Bay (Breaker, 2006), baseado em um registro de 85 anos de observações diárias no Pacífico Grove, Califórnia. Em Breaker (2006) são descritos processos oceânicos que afetam as águas de Monterey Bay, processos que poderiam contribuir num registro de não-linearidade. A análise de dados explanatória revela que o registro no Pacífico Grove é também não-gaussiano e, provável, não-estacionário.

4.13 Comentários Finais

A proposta desse capítulo foi de demonstrar de forma sucinta a aplicação do teste de hipótese nula utilizando o método de dados sub-rogados. Tal método mostrou-se ser adequado para a maioria dos modelos com comportamento dinâmico e estocástico propostos, mas existem alguns casos de restrição como, por exemplo, o processo MA com ruído não-gaussiano e em transformações estáticas não-lineares de processos lineares gaussianos e não-gaussianos.

Ressalta-se que a composição do procedimento final proposto resulta da formulação de novos conceitos sobre dados sub-rogados e da investigação das aplicações desenvolvidas ao longo desse trabalho. Portanto, a decisão tomada sob a hipótese nula de interesse deve ser cuidadosamente analisada conforme o procedimento proposto para análise do quadro geral da série temporal em estudo.

O objetivo deste procedimento proposto é garantir que o modelo possua características semelhantes ao sistema original, idéia principal do método de dados sub-rogados, e maior confiabilidade nas decisões tomadas sobre a hipótese nula de interesse com base nas estatísticas de testes (FAC e IMM) não-paramétricas.

Finalmente, uma breve exposição das conclusões obtidas nesta pesquisa são sintetizadas no Capítulo 5 de forma mais sucinta, dando uma visão geral dos re-

sultados encontrados.

Conclusões

O estudo conduzido nesse trabalho, sobre a aplicabilidade de teste de hipótese nula utilizando o método de dados sub-rogados, mostrou-se realizável e apresentou resultados relevantes para os casos estudados, revelando que uma investigação mais aprofundada dos sistemas dinâmicos lineares e não-lineares podem trazer vários benefícios para pesquisas aplicadas em engenharia.

Neste sentido, o trabalho desenvolvido contribui com uma extensa revisão bibliográfica relativa aos possíveis casos, analisados por meio das estatísticas de testes paramétricas ou não-paramétricas, de sistemas lineares, não-lineares, ruídos, caos ou sinais pseudo-periódicos, utilizando dados sub-rogados gerados conforme as características do mecanismo gerador original da série temporal idealizada. Tais análises foram de fundamental importância para identificação dos casos aplicáveis e condições prováveis de erro utilizando os métodos propostos de dados sub-rogados, tendo em mente que tais métodos são aplicados amplamente na literatura por diversos autores.

Uma das principais contribuições desta dissertação foi tratada no Capítulo 4, com os resultados construídos de diversos modelos ou casos a serem identificados. É identificado que a aplicação do método de dados sub-rogados não proporciona uma confiável decisão de acordo com a estrutura proposta para os seguintes casos: para modelos MA não-gaussianos (H_0 rejeitada em todas as aplicações), para alguns modelos ARMA não-gaussianos (ARMA(1,1) com distribuição uniforme e ARMA(1,1) com distribuição de Cauchy) e para as transformações estáticas não-lineares de processos lineares gaussianos e não-gaussianos (NAWGN, AWNGN e NAWNGN).

A Tabela 5.1 ilustra um visão geral desenvolvida com intuito de mapear as possíveis falhas do método proposto. A possibilidade vista neste quadro dessas análises, em síntese, é o principal identificador para a aplicabilidade do método para o conjunto de casos (modelos) levantados e propostos.

Além disso, é chegada a conclusão da utilização em conjunto das estatísticas de testes (FAC e IMM) não-paramétricas, pois demonstram ser mais consistentes

Tabela 5.1: Quadro geral dos resultados com respeito a cada hipótese nula proposta para todos os modelos empregados. A rejeição ou não rejeição em negrito identifica uma decisão errônea.

Modelo	IID	SS	FT	AAFT	IAAFT-1	IAAFT-2	CAAFT	STAP
Gaussiano	não rejeita	não rejeita	-	-	-	-	-	-
Uniforme [0,1]	não rejeita	não rejeita	-	-	-	-	-	-
Beta(2,5)	não rejeita	não rejeita	-	-	-	-	-	-
Laplace(0,1)	não rejeita	não rejeita	-	-	-	-	-	-
Cauchy(0,0.001)	não rejeita	não rejeita	-	-	-	-	-	-
Real/Dolar	rejeita	rejeita	rejeita	rejeita	não rejeita	rejeita	rejeita	rejeita
MA	rejeita	rejeita	não rejeita	-	-	-	-	-
AR(1)	rejeita	rejeita	não rejeita	-	-	-	-	-
ARMA(6,1)	rejeita	-	não rejeita	não rejeita	não rejeita	não rejeita	não rejeita	não rejeita
MA(1) Uniforme	rejeita	-	rejeita	rejeita	-	-	rejeita	rejeita
ARMA(1,1) Uniforme	rejeita	-	rejeita	rejeita	-	-	não rejeita	não rejeita
ARMA(7,1) Beta similar gaussiano	rejeita	-	não rejeita	não rejeita	-	-	não rejeita	não rejeita
ARMA(1,1) Laplace	rejeita	-	-	-	rejeita	rejeita	não rejeita	não rejeita
ARMA(1,1) Cauchy	rejeita	-	-	rejeita	não rejeita	-	rejeita	rejeita
NAWGN	-	-	rejeita	rejeita	não rejeita	rejeita	não rejeita	não rejeita
AWNGN	-	-	rejeita	rejeita	rejeita	rejeita	não rejeita	não rejeita
NAWNGN	-	-	rejeita	não rejeita	rejeita	rejeita	não rejeita	não rejeita
Mapa Ikeda gaussiano	rejeita	-	rejeita	rejeita	-	-	rejeita	rejeita
Mapa Logístico uniforme	rejeita	-	-	rejeita	rejeita	-	rejeita	rejeita
Mapa Henon beta	rejeita	-	-	-	rejeita	rejeita	rejeita	rejeita
Lorenz Laplace(0,1)	rejeita	rejeita	rejeita	rejeita	rejeita	rejeita	rejeita	rejeita
Caso Blind	rejeita	rejeita	não rejeita	rejeita	rejeita	rejeita	rejeita	rejeita

na tomada de decisão final com respeito a descoberta da natureza do mecanismo gerador da série temporal original, conservação da distribuição da amplitude ou do espectro de potência, inferindo assim aspectos lineares (FAC) e não-lineares (IMM) importantes para investigação na tomada da decisão. Foi colocado também que a ocorrência da rejeição em qualquer uma dessas estatísticas de testes (FAC ou IMM) a hipótese nula deve ser rejeitada, apresentado assim um teste mais restrito e criterioso para tomada de decisão final.

Os métodos sub-rogados utilizando o embaralhamento de ciclos (algoritmo CS) e o pseudo-periódico (algoritmo PPS) são expostos apenas para visualizar os métodos estudados e aplicados atualmente. Contudo, não foi dada ênfase em tais métodos por se tratarem de métodos de difícil implementação e aplicação das estatísticas de testes, mais especificamente dimensão de correlação. Sendo gasto um alto tempo computacional para tais investigações de dinâmica não-linear aperiódica ou dinâmica pseudo-periódica contaminada com ruído, e resultados, às vezes foram inconsistentes. Deixando aberto para estudos futuros investigações mais aprofundadas na aplicação desses métodos sub-rogados. Uma importante observação para o algoritmo CS é que este não serve para séries quase-periódicas.

Uma das principais contribuições desta dissertação é o procedimento geral proposto desenvolvido no final do Capítulo 4. Ressalta-se que a composição de um procedimento geral proposto resulta da aplicação de alguns conceitos e formulações sobre dados sub-rogados desenvolvidas ao longo desse trabalho. Portanto, a decisão tomada sob a hipótese nula de interesse deve ser cuidadosamente analisada com as etapas ou procedimentos propostos para análise do quadro geral da série temporal sob estudo, o que atualmente, não acontece na literatura.

Os testes estatísticos paramétricos são empregados, a princípio, para dados de processos com distribuições normais o que não é estritamente aplicado nos sistemas reais, sendo assim, este um dos problemas para aplicação desses tipos de testes estatísticos paramétricos para vários processos reais. Devido a esse fato, tais testes não foram tratados neste trabalho.

O resultado final do trabalho é um procedimento geral proposto utilizando os métodos de dados sub-rogados consistentes que possibilita determinar aleatoriedade (ruído) ou determinismo (linear ou não-linear). Lembrando que o procedimento proposto leva em conta uma combinação das decisões encontradas para cada método sub-rogado aplicado, para chegar a uma conclusão final, que nem sempre é trivial. Ressalta-se que devem ser feitos estudos mais abrangentes para ponderação de cada decisão encontrada individualmente.

Espera-se que esta dissertação sirva como ponto de partida para estudos poste-

riores que busquem explorar cada vez mais a investigação das estruturas (modelos) geradoras das séries temporais existentes.

A proposta deste trabalho é apresentar uma metodologia para aplicação do teste de hipótese nula utilizando vários métodos de dados sub-rogados aplicados a um exemplo (caso *Blind*) sem conhecimento dos dados *a priori*. Tais métodos mostraram ser adequados, mas com algumas restrições. Teste de hipótese nula utilizando o FT e o AAFT não são indicados para processos lineares não-gaussianos.

Ressalta-se que a composição de um procedimento geral proposto resulta da aplicação de alguns conceitos e formulações sobre dados sub-rogados desenvolvidas ao longo desse trabalho. Portanto, a decisão tomada sob a hipótese nula de interesse deve ser cuidadosamente analisada com as etapas ou procedimentos propostos para análise do quadro geral da série temporal sob estudo, o que atualmente não acontece na literatura.

O resultado final do trabalho é um procedimento geral proposto utilizando os métodos de dados sub-rogados consistentes que possibilita determinar aleatoriedade (ruído) ou determinismo (linear ou não-linear). Lembrando que o procedimento proposto leva em conta uma combinação das decisões encontradas para cada método sub-rogada aplicado, para chegar a uma conclusão final, que nem sempre é trivial. Os fluxogramas apresentados proporcionam um diagnóstico das causas das possíveis falhas, como por exemplo, a necessidade de mais dados para análise. Resalta que devem ser feitos estudos mais abrangentes para ponderação de cada decisão encontrada endividualmente.

Como trabalho futuro é proposto:

- Um estudo mais abrangente dos algoritmos CS e PPS entre outros;
- A aplicação de outras estatísticas de teste;
- Desenvolvimento de uma rede neural para tomada da decisão, levando em consideração um fator de contribuição da decisão para cada algoritmo ou técnica utilizada;
- Desenvolvimento de um algoritmo para identificação de dinâmica não-linear, baseado na análise da informação mútua média (interdependências não-lineares, informações de ordem superiores).

Referências Bibliográficas

- Aguirre, L. (2004). *Introdução à Identificação de Sistemas - Técnicas Lineares e Não-lineares Aplicadas a Sistemas Reais*. Editora UFMG.
- Anteneodo, C. (2004). Processos estocásticos. V Escola do Centro Brasileiro de Pesquisas Físicas.
- Bhatt, N. M. e Dave, P. H. (1964). A note on the correlation between polynomial transformations of normal variates. *Journal of the Indian Statistical Association*, 2:177–181.
- Bickel, P. e Buhlmann, P. (1996). What is a linear process? *Proc. Natl. Acad. Sci. USA*, 93:12128–12131.
- Breaker, L. C. (2006). Nonlinear aspects of sea surface temperature in monterey bay. *Elsevier*, páginas 61–89.
- Brockwell, P. e Davis, R. (1991). *Time Series: Theory and Methods*,. Springer Series in Statistics.
- Callegari-Jacques, S. M. (2006). *Bioestatística: Princípios e aplicações*.
- Coelho, M. C. S., A., A. L., e Mendes, E. M. A. M. (2006). Análise de determinismo em series temporais baseada em dados sub-rogados. *I SPGEE*, páginas 1–6.
- Doebelin, E. (1980). *System modeling and response - Theoretical and experimental approaches*. John Wiley & Sons, Inc.
- Efron, B. (1986). Bootstrap methods for standard errors, confidence intervals and other measures of statistical accuracy. *Statistic Science*, páginas 54–75.

- Efron, B. e Tibshirani, R. (1993). An introduction to the bootstrap. *Chapman and Hall*.
- Fielder-Ferrara, N. e Prado, C. P. C. (1994). *Caos: uma introdução*. Editora Edgar Blücher LTDA.
- Furtado, E. C., Mendes, E. M. A., e de Sá, A. M. M. L. F. (2006). Análise bispectral de sinais eeg utilizando o método surrogate. Seminários Internos do CPH II.
- Govindan, R. B., Narayanan, K., e Gopinathan, M. S. (1998). On the evidence of deterministic chaos in eeg: Surrogate and predictability analysis. *Chaos*, 8:495–502.
- Granger, C. W. J. e Newbold, P. (1977). *Forecasting Economic Time Series*. New York : Academic Press.
- Grassberger, P. e Procaccia, I. (1983.). Measuring the strangeness of strange attractors. *Physica D*, 8.
- Ikeda, K. (1973). Multiple-valued stationary state and its instability of the transmitted light by a ring cavity system. *Opt. Commun*, 30:257–261.
- Judd, K. (1992). An improved estimator of dimension and some comments on providing confidence intervals. *Physica D*, 56:216–228.
- Judd, K. (1994). Estimating dimension from small samples. *Physica D*, 71:421–429.
- Kantz, H. e Schreiber, T. (1997). *Nonlinear time series analysis*. Cambridge, UK: Cambridge University Press.
- Kugiumtzis, D. (1999a). Complications in applying the method of surrogate data to eeg. *World Scientific*, páginas 1–4. Max-Planck Institute for Physics of Complex Systems.
- Kugiumtzis, D. (1999b). Complications in applying the method of surrogate data to eeg. *World Scientific*, páginas 1–4.
- Kugiumtzis, D. (1999c). Test your surrogate data before you test for nonlinearity. *Physical Review E*, 60(3):2808–2816.

- Kugiumtzis, D. (2000a). Surrogate data test for nonlinearity including monotonic transformations. *Physical Review E*, 62(1):2808–2816.
- Kugiumtzis, D. (2000b). Surrogate data test on time series.
- Kugiumtzis, D. (2002a). *Modelling and Forecasting Financial Data, Techniques of Nonlinear Dynamics*. Kluwer Academic.
- Kugiumtzis, D. (2002b). Statically transformed autoregressive process and surrogate data test for nonlinearity. *Physical Review E*, 66:025201–1–4.
- Kugiumtzis, D. (2002c). *Surrogate data test on time series, In Modelling and Financial Data*. Kluwer Academic.
- Kugiumtzis, D. (2006). Test your surrogate data before you test for nonlinearity. 2006.
- Kugiumtzis, D. (2001). On the reliability of the surrogate data test for nonlinearity in the analysis of noisy time series. *Int. J. Bifurcation Chaos Appl. Sci. Eng.*, 11:1881–1896.
- Little, M. A., McSharry, P. E., Moroz, I. M., e Roberts, S. J. (2006). Testing the assumptions of linear prediction analysis in normal vowels. *Acoustical Society of America*, páginas 549–558.
- Ljung, L. (1987). *System Identification: Theory for the User*. Prentice-Hall, London.
- Mackey, M. C. e Glass, L. (1977). Oscillation and chaos in physiological control systems. *Science*, 197:287–289.
- May, R. M. (1976). Simple mathematical models with very complicated dynamics. *Nature*, 261:459–467.
- Mees, A. e Judd, K. (1995). On selecting models for nonlinear time series. *Physica D*, 82:426–444.
- Mendes, E. M. A. M., Hinich, M. J., e Stone, L. (2002). A comparison between standart bootstrap and theiler’s surrogate methods.
- Morettin, P. A. e Toloí, C. M. C. (2006). *Análise de Séries Temporais*.

- Nagarajan, R. (2005). Local analysis of dissipative dynamical systems. *Nonlinear Sciences - Chaotic Dynamics*, páginas 1–85.
- Nagaranjan, R. (2005). Surrogate testing of linear feedback processes with non-gaussian innovations. 2005.
- Nakamura, T. e Small, M. (2005). Testing for dynamics in the irregular fluctuations of financial data. *Elsevier Science*.
- Nelson, C. R. (1973). *Applied Time Series Analysis*. San Francisco : Holden-Day.
- Rial, J. e Anaclerio, C. (2000). Understanding nonlinear responses of the climate system to orbital forcing. *Quaternary Science Reviews*, páginas 1709–1722.
- Schmitz, A. e Schreiber, T. (2004). Surrogate data for non-stationary signals. *World Scientific*.
- Schreiber, T. (1997). Detecting and analysing nonstationarity in a time series using nonlinear cross prediction. *Physical Review Letters*, (78):843–846.
- Schreiber, T. (1998). Interdisciplinary application of nonlinear time series methods. *Chaotic Dynamics*, páginas 1–86.
- Schreiber, T., Hegger, R., e Kantza, H. (1999). Practical implementation of nonlinear time series methods: The tisean package. *Chaos*, 9:413–435.
- Schreiber, T. e Schmitz, A. (1996). Improved surrogate data for nonlinearity tests. *Physical Review Letters*, 77(4):635–638.
- Schreiber, T. e Schmitz, A. (1998). Constrained randomisation of time for hypothesis testing. *Chaos*.
- Schreiber, T. e Schmitz, A. (1999). On the discrimination power of measures for nonlinearity in a time series. *Chaos*, 1:1–6.
- Schreiber, T. e Schmitz, A. (2000). Surrogate time series. *Physica D*, (142):346–382.
- Schreiber, T. e Schmitz, A. (2001). Surrogate time series. *Chaos*, páginas 1–28.
- Small, M. (2005). *Applied Nonlinear Time Series Analysis: Applications in Physics, Physiology and Finance*. World Scientific.

- Small, M., Harrison, R., e Tse, C. (2001a). A surrogate test for inter-cycle determinism in oscillatory time series data. *International Symposium on Nonlinear Theory and its Applications*.
- Small, M. e Judd, K. (1998a). Correlation dimension: A pivotal statistic for non-constrained realizations of composite hypotheses in surrogate data analysis. *Physica D*, 120:386–400.
- Small, M. e Judd, K. (1998b). Detecting nonlinearity in experimental data. *International Journal of Bifurcation and Chaos*, 8.
- Small, M. e Judd, K. (1999). Detecting periodicity in experimental data using linear modeling techniques. *Physical Review E*, 59:1379–1385.
- Small, M., Luo, X., e Nakamura, T. (2006a). Surrogate data method applied to nonlinear time series. *Elsevier Science*.
- Small, M., Luo, X., Zhang, J., e Moroz, I. (2006b). Exact nonparametric inference for detection of nonlinearity. *Elsevier*, páginas 1–22.
- Small, M. e Nakamura, T. (2005). Small-shuffle surrogate data testing for dynamics in fluctuating data with trends. *Physical Review E*.
- Small, M. e Nakamura, T. (2006a). Applying the method of small-shuffle surrogate data: Testing for dynamics in fluctuating data with trends. *International Journal of Bifurcation and Chaos*, 16:3581–3603.
- Small, M. e Nakamura, T. (2006b). Testing for dynamics in the irregular fluctuations of financial data. *Elsevier Science*, páginas 1–16.
- Small, M., Nakamura, T., e Luo, X. (2005a). Surrogate test to distinguish between chaotic and pseudoperiodic time series. *Physical Review E*, 71:026230–1–8.
- Small, M. e Tse, C. K. (2002). Applying the method of surrogate data to cyclic time series. *Physica D*, 164:187–201.
- Small, M. e Tse, C. K. (2003). Detecting determinism in time series: The method of surrogate data. *IEEE*, 50:663–672.
- Small, M., Yu, D., e Harrison, R. (2001b). Surrogate test for pseudoperiodic time series data. *Physical Review Letters*, 87:188101/1–6.

- Small, T., Nakamura, T., Hiratab, Y., e Aiharab, K. (2005b). Small-shuffle surrogate method and the application. *International Symposium on Nonlinear Theory and its Applications*.
- Souza, R. C. (1989). Modelos estruturais para previsão de séries temporais : Abordagens clássica e bayesiana.
- Stam, C. J., Pijn, J. P. M., e Pritchard, W. S. (1998). Reliable detection of nonlinearity in experimental time series with strong periodic components. *Physica D*, 112:361–380.
- Sugihara, G. e May, R. (1990). Nonlinear forecasting as a way of distinguishing chaos from measurement error in time series. *Nature*, (344):737–740.
- Suzuki, T., Ikeguchi, T., e Suzuki, M. (2005). Effects of data windows on the methods of surrogate data. *Physical Review E*, 71:056708/1–5.
- Takens, F. (1981). Detecting strange attractors in turbulence. *Lect. Note Math.*, 898:366–381.
- Theiler, J. (1995). On the evidence for low-dimensional chaos in an epileptic electroencephalogram. *Physical Letter A*.
- Theiler, J., Eubank, S., Longtin, A., Galdrikian, B., e farmer, J. D. (1992). Testing for nonlinearity in time series: the method of surrogate data. *Physica D*, 58:77–94.
- Theiler, J. e Prichard, D. (1994). Generating surrogate data for time series with several simultaneously measured variables. *Physical Review Letters*, 73:951–955.
- Theiler, J. e Prichard, D. (1995). Constrained realization monte carlo method for hypothesis testing.
- Theiler, J. e Rapp, P. (1996a). Re-examination of the evidence for low-dimensional, nonlinear structure in the human electroencephalogram. *Electroencephalography and clinical neurophysiology*, 98:213–222.
- Theiler, J. e Rapp, P. (1996b). Re-examination of the evidence for low-dimensional, nonlinear structure in the human electroencephalogram. *Electroencephalogr Clin Neurophysiol*, 98:213–222.

- Timmer, J. (1998). Power of surrogate data testing with respect to nonstationarity. *Physical Review E*, 58:5153–5156.
- Zhang, J., Luo, X., e Small, M. (1998). Chaotic correlation among cycles in human electrocardiograms.
- Zoubir, A. M. e Boashash, B. (1998). The bootstrap signal processing applications. *IEEE Transactions on Signal Processing*, 15(1):55–76.

Determinismo vs. Estocasticidade

A.1 Introdução

Representar, por meio de modelos matemáticos, sistemas e fenômenos observados sempre foi um desafio. Em muitos problemas práticos ou teóricos é necessário o conhecimento prévio de modelos matemáticos que representem sistemas de qualquer natureza. Logo, o conhecimento prévio¹ e a obtenção de modelos que descreva adequadamente o sistema em questão é em si um problema digno de extrema importância.

Há diversas formas de se obter um modelo para um sistema em questão. Uma forma direta é recorrer à ciência e, a partir do conhecimento sobre o comportamento de partes do sistema e suas inter-relações tentar construir o modelo. Essa abordagem leva a modelos cuja precisão está relacionada com a profundidade sobre o sistema em estudo e a complexidade do mesmo. Nem sempre essa abordagem é possível, seja por levar a modelos proibitivamente complexos, seja por falta de conhecimento sobre o sistema ou poucos termos para elaborá-los.

Existem profundas diferenças entre métodos para previsão de sistemas dinâmicos e de processos estocásticos. Tal classificação é um passo decisivo para estabelecer estratégias de modelagem do sistema. Logo, o **método de dados sub-rogados** (do inglês *Surrogates*) é utilizado para verificar ou identificar determinadas características em dados gerados por processos estocásticos e determinísticos. Isto é feito por intermédio de teste de hipótese, tratado no Capítulo 2. Evidentemente uma resposta bem definida a esse problema será de imensa importância em praticamente todas as aplicações científicas nesse contexto.

No presente capítulo, inicialmente descreve-se de forma introdutória a teoria dos sistemas dinâmicos e, na seqüência, dos processos estocásticos.

¹Entenda aqui como conhecimento prévio a investigação do tipo de comportamento que o sistema ou fenômeno observado realiza, para tal estudo propõe-se o método de hipótese nula utilizando dados sub-rogados.

A.2 Sistemas Dinâmicos

Representar um sistema dinâmico², ou simplesmente um fenômeno físico, por uma equação matemática é algo que há muito tempo desperta o interesse de pesquisadores. O assunto remonta aos primórdios da ciência. Vários cientistas procuraram descrever fenômenos físicos através de leis matemáticas, como Isaac Newton com as leis do movimento e gravitação universal, Charles Coulomb com as leis da repulsão e atração de cargas elétricas, Maxwell e Boltzmann com a teoria cinética dos gases, dentre outros.

As equações diferenciais formam uma das primeiras soluções aplicadas para reproduzir comportamentos dinâmicos de sistemas físicos. Este método, em geral, é fundamentado na modelagem física do sistema dinâmico (Doebelin, 1980). A busca da representação de comportamentos, principalmente não-lineares, juntamente com o advento dos computadores, produziu ao longo dos anos várias propostas de estruturas de equações matemáticas para cumprir tal objetivo.

Existem dois tipos de equações que geram um sistema dinâmico. O primeiro tipo é definido como equações diferenciais ordinárias (e.d.o)

$$\dot{x} = f(t, x) \quad (\text{A.1})$$

onde $t \in \mathfrak{R}$ e $x \in \mathfrak{R}^n$. O índice n representa o número de equações ordinárias, esse índice está diretamente relacionado com a dimensão do vetor x .

Este sistema é dito *não-autônomo* se a função f dependa explicitamente do tempo. Caso contrário (se $f=f(x)$) o sistema é dito *autônomo*. Na prática chama-se de sistema autônomo o que não possui entradas, seja por elas não serem conhecidas ou porque o número de entradas seja tão grande que se torna impossível levá-las em conta. Sistemas não-autônomos possuem entradas e saídas bem definidas. Portanto, para descrever tais sistemas normalmente são usados modelos de entrada e saída (Aguirre, 2004), como descrito no seguinte exemplo

$$\tau \frac{dy}{dt} + y(t) = ku(t), \quad (\text{A.2})$$

sendo que $y(t)$ e $u(t)$ representam a saída e a entrada de um sistema contínuo hipotético, τ e k são constantes do sistema.

O segundo tipo de equação descreve sistemas discretos e pode ser definido por

²*Sistema Dinâmico*: um sistema que evolui, que se transforma como o tempo. Os sistemas dinâmicos tem memória e são descrito por equações diferenciais no caso de sistemas contínuos, e por equações de diferença no caso de sistemas discretos

equações de diferença

$$x_{k+1} = g(x_k), \quad k \in \mathbb{Z}^+ \quad (\text{A.3})$$

pode-se dizer que o sistema dinâmico contínuo sofreu um processo de discretização.

Um sistema dinâmico diferenciável pode ser representado tanto em tempo discreto quanto em tempo contínuo. Fenômenos representados por séries temporais geralmente acontecem de forma contínua e deveriam ter uma representação contínua. Todavia, as respectivas séries temporais são obtidas com uma determinada frequência (processo de amostragem), atribuído ao sistema uma representação discreta.

Um outro conceito importante é a invariância temporal³, isto é, o comportamento do sistema em questão não varia com o tempo. Isto não significa que as variáveis do sistema possuem valores constantes. Pelo contrário, normalmente os valores das variáveis que caracterizam um sistema flutuam com o tempo, sendo que tal evolução temporal é determinada por um comportamento dinâmico. Portanto, ser invariante no tempo não quer dizer que o sistema esteja “estático”, mas certamente implica que a dinâmica que está regulando a evolução temporal é a mesma. Essa é uma das premissas mais importantes em modelagem matemática, a saber, que a dinâmica não se altera significativamente no período de tempo em que se considera o sistema.

Infelizmente, a maioria dos sistemas reais o comportamento varia ao longo do tempo, seja em função de flutuações de variáveis que afetam a sua operação (por exemplo, flutuações de temperatura ao longo do dia ou do ano), seja como consequência de envelhecimento (por exemplo, perda de elasticidade de uma mola) ou simplesmente devido a uma forma diferente de operação (por exemplo, a dinâmica associada aos batimentos cardíacos é diferente se o paciente está dormindo ou se está acordado).

Um conceito relacionado ao de invariância é o de *estacionariedade*⁴. Este, entretanto, é mais usado no contexto de sinais e processos estocásticos.

³Formalmente, se diz que um sistema é invariante se um deslocamento no tempo na entrada causa um deslocamento no tempo na saída. Suponha-se o caso discreto em que um sistema com entrada $u(k)$ e saída $y(k)$. Se esse sistema for invariante no tempo, $u(k - k_0)$ produzirá $y(k - k_0)$. Informalmente, diz-se que um sistema é invariante no tempo se sua dinâmica (incluindo também seu ganho estático) não varia com o tempo.

⁴Um processo é dito estacionário se as leis de probabilidade que regem não variam com o tempo ou cujas propriedades estatísticas (momentos) não se alteram com o tempo. Pode-se dizer que um processo estacionário está em equilíbrio estático. O conceito de estacionariedade está intimamente ligado ao de invariância.

Os sistemas dinâmicos podem ser divididos em duas classes: *sistemas lineares* e *não-lineares*, que divergem entre si na sua relação de causa e efeito. Na primeira a resposta do sistema a um distúrbio é diretamente proporcional à intensidade deste, por exemplo, equação A.2. Já na segunda a resposta do sistema não é necessariamente proporcional à intensidade do distúrbio. Sistemas lineares têm a vantagem da simplicidade e do elaborado ferramental de análise disponível, porém não são capazes de exibir uma série de comportamentos como, por exemplo, bifurcações e caos. Representações não-lineares, apesar de mais complexas, são capazes de captar esses regimes dinâmicos.

Formalmente se diz que um sistema é linear se ele satisfaz o *princípio da superposição*⁵. Uma consideração freqüentemente feita é a de supor que o sistema em questão comporta-se de forma aproximadamente linear. Tal suposição é normalmente verificada observando-se o comportamento de um sistema numa faixa relativamente estreita de operação. Informalmente, pode-se dizer que o sistema linear tem o mesmo tipo de comportamento independentemente do ponto de operação. De fato, para um sistema que obedeça ao princípio da superposição, é possível descrever e estudar o comportamento linear a partir de um conjunto reduzido de casos particulares.

Um sistema é dito não-linear se ele não satisfaz o *princípio da superposição*. Um exemplo de equação diferencial não-linear é dado:

$$u(t)\frac{dy}{dt} + y(t) - ku(t) = 0, \quad (\text{A.4})$$

sendo que $y(t)$ e $u(t)$ representam a saída e a entrada de um sistema hipotético e k é uma constantes do sistema. A equação A.4 é não-linear uma vez que a derivada primeira do sinal de saída é multiplicada pela entrada e não por uma constante apenas como na equação A.2.

Todo sistema real é em princípio não-linear. A dinâmica de sistemas não-lineares normalmente depende da amplitude do sinal de entrada, bem como do ponto de operação do sistema. Em torno do ponto de operação, alguns sistemas não-lineares podem ser aproximados por modelos lineares.

Nas últimas décadas tem-se verificado um grande desenvolvimento no estudo dos fenômenos não-lineares com introdução de novas abordagens e conceitos no

⁵A resposta produzida pela aplicação simultânea de diversas excitações diferentes é igual a soma das respostas individuais a cada excitação. Considere, por exemplo, um sistema que ao ser excitado pela entrada $u_1(t)$ produz saída $y_1(t)$ e quando excitado por $u_2(t)$ produz $y_2(t)$. Se tal sistema satisfaz o princípio da superposição então, quando excitado por $au_1(t) + bu_2(t)$, sua saída será $ay_1(t) + by_2(t)$, sendo a e b constantes possivelmente complexas (Aguirre, 2004).

tratamento de sistemas dinâmicos conservativos e dissipativos. Um dos aspectos centrais dos novos desenvolvimentos reside no comportamento caótico determinístico que pode ocorrer já em sistemas com pelo menos três graus de liberdade (Fielder-Ferrara e Prado, 1994).

A teoria do caos⁶ estuda o comportamento aleatório e imprevisível dos sistemas, mostrando uma faceta onde podem ocorrer irregularidades na uniformidade da natureza como um todo. Isto ocorre a partir de pequenas alterações que aparentemente nada têm a ver com o evento futuro, alterando toda uma previsão física dita precisa. A idéia é então que uma pequena variação nas condições iniciais (dependência sensitiva) em determinado ponto de um sistema dinâmico pode ter conseqüências de proporções inimagináveis.

Uma das idéias centrais desta teoria, é que os comportamentos casuais (aleatórios) também são governados por leis e que estas podem predizer dois resultados para uma entrada de dados. O primeiro é uma resposta ordenada e superfície lisa e cujo futuro dos eventos ocorre dentro de margens estatísticas de erros previsíveis. O segundo é uma resposta também ordenada, onde porém a resultante futura dos eventos é corrugada, onde a superfície é áspera, caótica, ou seja, ocorre uma contradição neste ponto onde é previsível que os resultados de um determinado sistema será caótico.

O sistema de Lorenz é um exemplo de comportamento dinâmico caótico.

$$\begin{aligned}\dot{x} &= -\sigma(x - y), \\ \dot{y} &= rx - y - xz, \\ \dot{z} &= xy - bz,\end{aligned}\tag{A.5}$$

onde $(x, y, z) \in \mathfrak{R}^3$, $(\sigma, r, b > 0)$.

Com o interesse na previsão do tempo, o modelo foi criado para estudar o movimento convectivo descrito por uma massa de gás atmosférico que, quando estivesse próximo a superfície, deveria esquentar e subir e, quando estivesse na parte superior da atmosfera, deveria esfriar e descer. O modelo de Lorenz não representa o movimento real descrito pelos gases atmosféricos, mas sua notável contribuição na verificação do fenômeno de alta sensibilidade às condições iniciais não pode ser ignorada. As funções $x(t)$, $y(t)$ e $z(t)$ não representam coordenadas espaciais, elas possuem significados físicos precisos no modelo: $x(t)$ é proporcional à intensidade de convecção; $y(t)$ é proporcional à diferença de temperatura entre as correntes ascendentes e descendentes; e $z(t)$ é proporcional à

⁶fenômenos não-lineares

distorção do perfil de temperatura vertical em relação a um perfil linear.

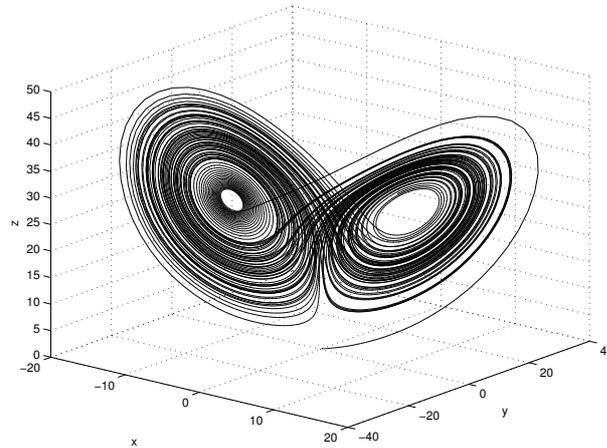


Figura A.1: Órbita do sistema de Lorenz para $\sigma = 10$, $b = \frac{8}{3}$ e $r = 28$ com condições iniciais $(x(0) = y(0) = z(0) = 0.1)$

A.3 Processos Estocásticos

Praticamente qualquer sistema (físico ou não) está sujeito a complicadas influências que não podem ser inteiramente conhecidas. Estas influências, tipicamente associadas a um grande número de graus de liberdade envolvidos, impedem prever com precisão arbitrária o estado do sistema em cada instante. A formulação de leis matemáticas com os conceitos de probabilidade é de antemão fundamental para caracteriza o fenômeno como sendo um processo estocástico.

A.3.1 Conceitos Básicos

Um experimento aleatório (ou não-determinista) é tal que não seja possível afirmar a priori o resultado que ocorrerá, podendo o resultado ser diferente mesmo ao se repetir o ensaio em condições praticamente inalteradas. Os resultados podem parecer errôneos nas primeiras tentativas, entretanto, após um grande número de repetições, aparecem regularidades.

Quando se lida com fenômenos aleatórios, pode-se conhecer, em geral, o conjunto dos possíveis resultados a serem observados ao realizar uma determinada experiência. A partir de um modelo, pode-se também atribuir aos resultados ou conjuntos de resultados possíveis, números que representem as suas chances de

ocorrência. Estes números, positivos e menores que a unidade, que se associam a um evento aleatório, e que se medem pela frequência relativa da sua ocorrência numa longa sucessão de eventos, são denominadas **probabilidades**.

Variáveis Aleatórias (v.a.)

Dado um experimento aleatório ξ , o espaço amostral Ω é o conjunto dos resultados possíveis (numéricos ou não). Para cada evento A (qualquer subconjunto de Ω) pode ser associado um número real não-negativo $P(A)$ denominado probabilidade, tal que $P(A \cup B) = P(A) + P(B)$, para eventos A e B mutuamente excludentes, e $P(\Omega) = 1$.

Uma **variável aleatória** (v.a.) unidimensional X é uma função $\Omega \mapsto \Omega_X \subset \mathfrak{R}$ que associa a cada elemento $\omega \in \Omega$ um (único) número real $X(\omega)$. A variável X pode ser discreta (Ω_X finito ou infinito numerável) ou contínua (Ω_X infinito não-numerável).

Distribuições de Probabilidade

Uma v.a. é frequentemente descrita por uma função de distribuição. A distribuição de probabilidade de uma v.a. X cujo contradomínio ⁷ é $\Omega_X = \{x_1, x_2, \dots\}$, caso discreto, é dada pelo conjunto de pares (x_i, p_i) , com $i = 1, 2, \dots$, onde $p_i \triangleq P(X = x_i)$, probabilidade de x_i com $p_i \geq 0$ e $\sum_{i \geq 1} p_i = 1$.

No caso contínuo, a distribuição de probabilidade de uma variável aleatória X é dada pela função f_X , chamada **função densidade de probabilidade** ou p.d.f., tal que $f_X(x)dx$ representa a probabilidade $P(x \leq X \leq x + dx)$. Assim sendo, a p.d.f. permite calcular a probabilidade de que X se encontre dentro de um intervalo real $[a, b]$:

$$P(a \leq X \leq b) = \int_a^b f_X(x)dx \quad (\text{A.6})$$

onde $f_X(x) \geq 0, \forall x$, e $\int_{-\infty}^{+\infty} f_X(x)dx = 1$.

A **função distribuição cumulativa** (c.d.f) ou **função distribuição de probabilidade** é dada por

$$F_X(x) = \int_{-\infty}^x f_X(\omega)d\omega = P(X \leq x) \quad (\text{A.7})$$

⁷ou espaço amostral com relação a X

onde $F_X(x)$ é monotônico crescente⁸.

Inversamente, a densidade não-cumulativa ou p.d.f. é obtida por diferenciação:

$$f_X(x) = \frac{\partial F_X}{\partial x} \quad (\text{A.8})$$

Por exemplo, seja X uma v.a. de densidade gaussiana, escreve-se, respectivamente, a p.d.f e a c.d.f. como

$$f_X(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \quad (\text{A.9})$$

$$F_X(x) = \int_{-\infty}^x \frac{1}{\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx = P(X \leq x), \quad (\text{A.10})$$

onde μ e σ^2 são respectivamente a média e a variância.

A p.d.f. conjunta de duas (ou mais) variáveis aleatórias X e Y é $f_{X,Y}(x, y)$, tal que $f_{X,Y}(x, y)dx dy = P(x \leq X \leq x + dx, y \leq Y \leq y + dy)$. O par (X, Y) representa uma v.a. bidimensional. A c.d.f. conjunta é

$$F_{X,Y}(x, y) = P(X \leq x, Y \leq y) \quad (\text{A.11})$$

A p.d.f. marginal, por exemplo, da variável X é dada por

$$f_X(x) = \int_{-\infty}^{+\infty} f_{X,Y}(x, y) dy, \quad (\text{A.12})$$

enquanto a c.d.f. marginal é $F_X(x) = F_{X,Y}(+\infty, x)$ e $F_Y(y) = F_{X,Y}(y, +\infty)$.

A p.d.f. condicional⁹ de X , dado um certo valor de $Y = y$, é

$$f_{X|Y}(x, y) = \frac{f_{X,Y}(x, y)}{f_Y(y)}, \quad \text{com } f_Y > 0 \text{ (p.d.f. marginal de } Y) \quad (\text{A.13})$$

⁸Se $x_1 < x_2$, isto significa que $F_X(x_1) \leq F_X(x_2)$.

⁹Regras de Bayes

Momentos

O **momento**¹⁰ de ordem n da variável X é dado por:

$$E[X^n] = \langle X^n \rangle = \int_{-\infty}^{+\infty} x^n f_X(x) dx. \quad (\text{A.14})$$

onde o momento de ordem 1 é chamado de média aritmética.

O **momento centrado** de ordem n da variável X é dado por

$$m_k = \langle [X - \mu_X]^n \rangle = \int_{-\infty}^{+\infty} [x - \mu_X]^n f_X(x) dx, \quad (\text{A.15})$$

onde $\mu_X \equiv \langle X \rangle$ é a média aritmética. A variância ou desvio quadrático é o momento centrado de ordem 2,

$$\sigma_X^2 = \langle [X - \mu_X]^2 \rangle, \quad (\text{A.16})$$

e o desvio padrão a raiz quadrada da variância.

A média conjunta de duas v.a. (X, Y) é dada por

$$\mu_{XY} = \langle X, Y \rangle = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} x \cdot y f_{X,Y}(x, y) dx dy. \quad (\text{A.17})$$

Se X e Y são v.a. independentes, então $E[X \cdot Y] = E[X] \cdot E[Y]$.

Função característica

A **função característica** (FC) da variável X é a transformada de Fourier da sua p.d.f.

$$\Phi_X(k) = \langle e^{ikx} \rangle = \int_{-\infty}^{+\infty} e^{ikx} f_X(x) dx. \quad (\text{A.18})$$

onde $i = \sqrt{-1}$ e k um parâmetro real.

A função característica Φ_X pode ser considerada uma função geratriz de momentos, já que os coeficientes da expansão de Taylor estão relacionados com os momentos da p.d.f associada

$$\Phi_X(k) = \sum_{n \geq 1} \frac{(ik)^n}{n!} \langle X^n \rangle \quad (\text{A.19})$$

¹⁰Também chamado de esperança matemática $E[\cdot]$ ou valor esperado

Tabela A.1: Distribuições de probabilidade unidimensionais

Distribuição	p.d.f.	média	variância
Uniforme	$\frac{1}{b-a}$	$\frac{a+b}{2}$	$\frac{(b-a)^2}{12}$
Exponencial	$\alpha e^{-\alpha x}, \quad x \geq 0 (\alpha > 0)$	$\frac{1}{\alpha}$	$\frac{1}{\alpha^2}$
Gaussiana $N(\mu, \sigma^2)$	$\frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$	μ	σ^2
Beta	$\frac{x^{\alpha-1}(1-x)^{\beta-1}}{\text{beta}(\alpha, \beta)}$ com $\alpha, \beta > 0$	$\frac{\alpha}{\alpha + \beta}$	$\frac{\alpha\beta}{(\alpha + \beta)^2(\alpha + \beta + 1)}$
Cauchy	$\frac{s}{\pi(s^2 + (x-t)^2)}$ onde $s > 0$ e t a localização	-	-
Laplace	$\frac{1}{2\lambda} e^{-\left \frac{y-\theta}{\lambda}\right }$ onde $\lambda > 0$ e θ a localização	θ	$2\lambda^2$
t-student	$\frac{\Gamma(\frac{n+1}{2})}{\sqrt{n\pi}\Gamma(\frac{n}{2})} \frac{1}{\left(1 + \frac{x^2}{n}\right)^{\frac{n+1}{2}}}$	$0 \quad (n > 1)$ $* \quad (n = 1)$	$\infty \quad (n \neq 2)$ $\frac{n}{n-2}$
χ_n^2	$\frac{x^{\left(\frac{n}{2}-1\right)} e^{-\frac{x}{2}}}{2^{\frac{n}{2}} \Gamma\left(\frac{n}{2}\right)}$ $x \geq 0 \quad (n = 1, 2, \dots)$	n	$2n$

que podem ser obtidos como sendo

$$\langle X^n \rangle = (-i)^n \frac{\partial^n \Phi_X}{\partial k^n} \Big|_{k=0} \quad (\text{A.20})$$

A FC também permite obter os **cumulantes** $\langle\langle X^n \rangle\rangle$ (combinações dos momentos), definidos como

$$C_n = \langle\langle X^n \rangle\rangle = (-i)^n \frac{\partial^n \Phi_X}{\partial k^n} \Big|_{k=0} \quad (\text{A.21})$$

Covariância e autocorrelação

A **covariância** de duas variáveis aleatórias X e Y é dada por

$$\begin{aligned} \text{cov}(X, Y) &= \langle [X - \mu_x] \rangle \langle [Y - \mu_y] \rangle \\ &= \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} [x - \mu_x][y - \mu_y] f_{X,Y}(x, y) dx dy \end{aligned} \quad (\text{A.22})$$

e o coeficiente de correlação é dado por

$$\rho_{XY} = \frac{\text{cov}(X, Y)}{\sigma_X \sigma_Y} \quad (\text{A.23})$$

Duas variáveis X e Y são ditas descorrelacionadas se a sua covariância for nula.

Por outro lado, dado um processo estocástico $X(t)$, se o processo for estacionário, a sua **função de autocorrelação** depende somente da diferença de tempos:

$$E[X_t X_\tau^*] = r_{xx}(t, \tau) \equiv r(t, \tau) \quad (\text{A.24})$$

onde o asterisco denota o complexo conjugado. Se o processo é estacionário e ergódico, $r_{xx}(\tau)$ independe do tempo e o valor esperado pode ser substituída por uma média temporal sobre um intervalo de tempo suficientemente longo $2T$ de uma dada realização $x(t)$ do processo, assim

$$r_{xx}(\tau) = \lim_{T \rightarrow \infty} \frac{1}{2T} \int_{-T}^T x(t)x(t + \tau) dt \quad (\text{A.25})$$

Na forma discreta tem-se

$$r_{xx}(\tau) = \lim_{N \rightarrow \infty} \frac{1}{N} \sum_{k=1}^N x(k)x(k + \tau) \quad (\text{A.26})$$

onde τ é o atraso.

Densidade espectral de potência ou espectro de potência

A identificação de processos regulares pode ser feita através do uso de métodos clássicos. Nesses se estuda o tipo de irregularidade (ou regularidade) existente na série temporal $X(t)$. São métodos clássicos a análise espectral (espectro de potência) e a função de autocorrelação.

Dada uma série $X(t)$, é útil conhecer sua decomposição em funções senoidais do tempo através da análise de Fourier. Qualquer função $f(t)$ pode ser representada pela superposição de um número (eventualmente infinito) de componentes periódicas. A determinação do peso relativo a cada uma dessas componentes é chamada *análise espectral*. Se $f(t)$ é periódica, seu espectro pode ser representado como a combinação linear de oscilações cujas frequências são múltiplos inteiros de uma frequência básica w . Essa combinação linear é chamada série de Fourier. Quando $f(t)$ é não-periódica, o que é mais freqüente, o espectro de frequência varia continuamente e usa-se a chamada transformada de Fourier para representar $f(t)$ em termos dessas frequências. Escreve-se a transformada de Fourier de $f(t)$ como

$$f(w) = \int_{-\infty}^{\infty} e^{-iwt} f(t) dt \quad (\text{A.27})$$

onde $f(w)$ indica o peso relativo com que a frequência w comparece na composição de $f(t)$. O espectro de potência é definido como o módulo quadrado de $f(w)$, ou seja,

$$P(w) = |f(w)|^2 \quad (\text{A.28})$$

Na situação de interesse prático dispõe-se de uma série temporal finita discreta X_k e cuja transformada de Fourier é definida como,

$$\tilde{x}(n) = \frac{1}{\sqrt{N}} \sum_{k=1}^N x(k) e^{-j2\pi nk/N} \quad (\text{A.29})$$

onde $n = 1, \dots, N$ e $w = 2\pi \frac{n}{N}$.

O espectro de potência $P(w)$ para uma série temporal discreta é definido por

$$P(w) = |\tilde{x}_k|^2 \quad (\text{A.30})$$

O espectro de potência é proporcional à transformada de Fourier da função de

autocorrelação

$$P(w) = \int_{-\infty}^{\infty} e^{-iw\tau} r_{xx}(\tau) d\tau \quad (\text{A.31})$$

ou para o caso discreto

$$P(w) = |x_k|^2 = \sum_{n=1}^N r_{xx}(k) e^{-i2\pi n \frac{k}{N}} \quad (\text{A.32})$$

Este é um resultado importante, conhecido como relação de Wiener-Khinchin, que relaciona a espectro de potência $P(w)$ com a função de autocorrelação r_{xx} .

Independência estatística

Duas v.a.s de uma distribuição conjunta X_i e X_j , $i \neq j$, são ditas independentes se os eventos $\{X_i \leq x_i\}$ e $\{X_j \leq x_j\}$ forem independentes, ou seja, cada evento realizado independe do outro. Isto leva a

$$F_{X_i, X_j}(x_i, x_j) = F_{X_i}(x_i) F_{X_j}(x_j), \quad (\text{A.33})$$

e em consequência

$$f_{X_i, X_j}(x_i, x_j) = \frac{\partial^2 F_{X_i, X_j}(x_i, x_j)}{\partial x_i \partial x_j} = f_{X_i}(x_i) f_{X_j}(x_j), \quad (\text{A.34})$$

As v.a.s X_1, \dots, X_n são mutuamente independentes se

$$F_{X_1, \dots, X_n}(x_1, \dots, x_n) = \prod_{i=1}^n F_{X_i}(x_i), \quad (\text{A.35})$$

em consequência

$$f_{X_1, \dots, X_n}(x_1, \dots, x_n) = \prod_{i=1}^n f_{X_i}(x_i), \quad (\text{A.36})$$

Teorema do limite central

Sejam X_i ($i = 1, \dots, N$) variáveis aleatórias independentes e identicamente distribuídas (i.i.d.) com média μ finita e variância $0 < \sigma^2 < \infty$, segundo o *teorema*

do limite central (TLC) a distribuição da variável

$$\frac{\sum_{i=1}^N X_i - n\mu}{\sigma \sqrt{n}} \xrightarrow{D} N(0, 1) \quad (\text{A.37})$$

tende para uma distribuição normal $N(0, 1)$ quando $N \rightarrow \infty$.

É importante notar que para processos autoregressíveis (AR) com ruído (análise feita para séries infinitas) o TLC prova que, os modelos AR convergem para uma distribuição normal. Porém, para a maioria dos modelos autoregressivos de média móvel - ARMA o TLC não é aplicável.

Estacionariedade

Os processos estocásticos podem ser *estacionários* ou *não-estacionários* consoante as propriedades estatísticas sejam “semelhantes” entre si ou não e de “idêntico aspecto” ao longo do tempo.

- Processo estacionário no sentido estrito (ESE): Todas as propriedades estatísticas são invariantes para uma translação do tempo.

$$f_X(x_t) = f_X(x_{t+\tau}) = f_X(x) \quad (\text{A.38})$$

isto significa que,

$$\begin{aligned} \int_{-\infty}^{+\infty} x_t f_X(x_t) dt &= \int_{-\infty}^{+\infty} x_{t+\tau} f_X(x_{t+\tau}) d(t+\tau) \\ \mu_X &= \mu_X(t+\tau) = \text{constante} \end{aligned} \quad (\text{A.39})$$

- Processo estacionário no sentido amplo (ESA)¹¹: satisfaz simultaneamente a ESE e $r_{xx}(t, t+\tau)$ depende apenas de τ

$$\begin{aligned} \mu_X &= \text{constante} \\ r_{xx}(t, t+\tau) &= r_{xx}(\tau) \end{aligned} \quad (\text{A.40})$$

O processo estocástico é dito estacionário em sentido amplo, se possui segundos momentos finitos e se a $cov(X(t_i), X(t_i + \tau))$ depende somente de $\tau, \forall \tau \in T$.

¹¹fracamente estacionário

A.3.2 Definição de Processo Estocástico

Os processos estocásticos $X(t)$ são famílias arbitrárias de variáveis aleatórias indexadas por t onde $t \in \mathbb{T}$ (sendo \mathbb{T} um conjunto qualquer, geralmente \mathbb{N}^+) e $X(t)$ representa o estado do processo no tempo t . O conjunto índice do processo t pode ser visto como um indexador de tempo ou espaço.

- Se \mathbb{T} é um conjunto enumerável, então $X(t)$, $t \in \mathbb{T}$, é um processo estocástico discreto no tempo.
- Se \mathbb{T} é um conjunto não enumerável ou \mathbb{T} é um intervalo aberto ou fechado da reta, então $X(t)$, $t \in \mathbb{T}$, é um processo estocástico contínuo no tempo.

Os resultados assumidos por $X(t)$ (ou X_t) são denotados como conjunto dos estados do processo. Os processos estocásticos são definidos de maneira estritamente formal através do Teorema Fundamental dos Processos Estocásticos (Anteneodo, 2004):

Definição A.3.1 *Se a função distribuição de probabilidade conjunta das variáveis aleatórias $X_{t_1}, X_{t_2}, \dots, X_{t_n}$ é conhecida para todo n enumerável positivo, e para todo conjunto valores t_1, t_2, \dots, t_n onde t_k qualquer, pertence a um conjunto \mathbb{T} , podemos denotar conjunto destas variáveis, X_{t_i} , por Processo Estocástico.*

Intuitivamente, se uma variável aleatória uni-dimensional é um número real que varia aleatoriamente, um processo estocástico é uma função que varia aleatoriamente.

Dessas formulações de conceitos, vem que um processo estocástico X_t é uma v.a. para cada instante t , ou seja, trata-se de uma função de dois argumentos, $X(t)$, para $t \in \mathbb{T}$. Essa dependência indica que em t , obtém-se uma v.a. descrita por uma função de distribuição $f_{X_t}(x)$. Se X_t e X_τ forem v.a.s independentes para $t \neq \tau$ e, além disso, apresentam a mesma distribuição, então define-se o processo como **independente e identicamente distribuído** (i.i.d.).

Existem vários tipos de processos estocásticos. Como exemplos, temos os Processos Gaussianos, Markovianos, Poisson, Lévy, Martingalas e outros.

Processos puramente aleatórios

A p.d.f condicional $f_{1|n-1}$, com $n > 1$, independe dos valores em instantes anteriores, ou seja

$$f_{1|n-1}(x_n, t_n | x_1, t_1; \dots; x_{n-1}, t_{n-1}) = f_1(x_n, t_n) \quad (\text{A.41})$$

Um processo assim produz resultados diferentes e imprevisíveis a cada vez em que é executado, puramente aleatório. Um exemplo de processo puramente aleatório é a seqüência de resultados ao se lançar muitas vezes um dado. Neste caso particular trata-se de um processo estocástico de tempo discreto em que x_t são variáveis aleatórias independentes e identicamente distribuídas com distribuição uniforme.

Um processo estocástico associa-se em geral a presença de ruído. No caso extremo de um sinal completamente aleatório tem-se o chamado ruído branco, semelhante a um processo gaussiano com $N(0, \sigma^2)$. A Figura A.2 ilustra dois processos estocásticos, ruído branco e ruído browniano.

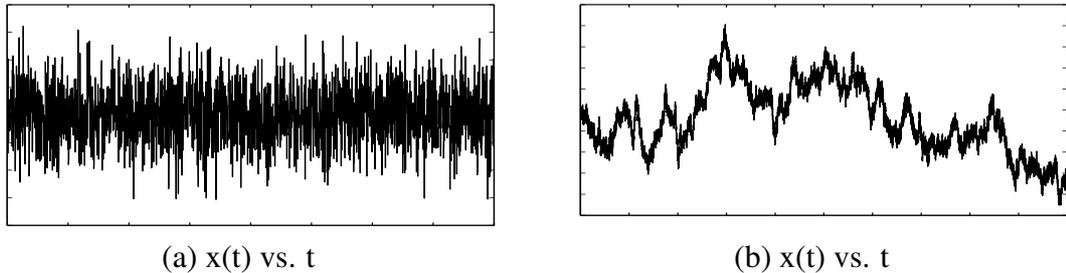


Figura A.2: Realizações típicas de dois processos estocásticos: (a) ruído branco e (b) ruído browniano.

Tabelas

Tabela B.1: H_0 associada ao tipo de teste estatístico e sua probabilidade com 99 dados sub-rogados gerados a partir da variável aleatória gaussiana.

Técnica	Estatística de teste	Significância	Probabilidade	Decisão H_0
IID	Reversão temporal	-0,11789	0,9062	não rejeita
	TPA	-0,043877	0,9650	não rejeita
GAUSS	Variância	0	1	não rejeita
	Assimetria	-0,43111	0,6664	não rejeita
	Kurtoses	-1,1129	0,2658	não rejeita
	C_6	0,81292	0,4163	não rejeita
	Reversão temporal	-0,14649	0,8835	não rejeita
	TPA	0,25113	0,8017	não rejeita
BOOT	Variância	-0,10066	0,9198	não rejeita
	Assimetria	-0,18759	0,8512	não rejeita
	Kurtoses	0,017617	0,9859	não rejeita
	C_6	0,068766	0,9452	não rejeita
	Reversão temporal	-0,2075	0,8356	não rejeita
	TPA	-0,10492	0,9164	não rejeita
SS	Reversão temporal	-0,15304	0,8784	não rejeita
	TPA	-0,0031656	0,9975	não rejeita

Tabela B.2: H_0 associada ao tipo de teste estatístico e sua probabilidade com 99 dados sub-rogados gerados a partir da série temporal da distribuição Uniforme [0,1].

Técnica	Estatística de teste	Significância	Probabilidade	Decisão H_0
IID	Reversão temporal	0,58202	0,5606	não rejeita
	TPA	1,2692	0,2044	não rejeita
GAUSS	Variância	-0,11936	0,9050	não rejeita
	Assimetria	77,0761	0	rejeita*
	Kurtoses	155,0518	0	rejeita*
	C_6	578,3909	0	rejeita*
	Reversão temporal	1,686	0,0918	não rejeita
	TPA	0,022494	0,9821	não rejeita
BOOT	Variância	0,085036	0,9322	não rejeita
	Assimetria	0,24684	0,8050	não rejeita
	Kurtoses	0,25308	0,8002	não rejeita
	C_6	0,23311	0,8157	não rejeita
	Reversão temporal	0,96638	0,3339	não rejeita
	TPA	1,676	0,0937	não rejeita
SS	Reversão temporal	0,61196	0,5406	não rejeita
	TPA	0,94541	0,3444	não rejeita

Tabela B.3: H_0 associada ao tipo de teste estatístico e sua probabilidade com 99 dados sub-rogados gerados a partir série temporal da distribuição Beta(2,5).

Técnica	Estatística de teste	Significância	Probabilidade	Decisão H_0
IID	Reversão temporal	0,34973	0,7265	não rejeita
	TPA	-0,64305	0,5202	não rejeita
GAUSS	Variância	0,19956	0,8418	não rejeita
	Assimetria	-23,4838	0	rejeita*
	Kurtoses	-1,0615	0,2885	não rejeita
	C_6	9,0613	0	rejeita*
	Reversão temporal	0,37654	0,7065	não rejeita
	TPA	0,15807	0,8744	não rejeita
BOOT	Variância	-0,051486	0,9589	não rejeita
	Assimetria	-0,047292	0,9623	não rejeita
	Kurtoses	0,0040555	0,9968	não rejeita
	C_6	0,013249	0,9894	não rejeita
	Reversão temporal	0,35236	0,7246	não rejeita
	TPA	-1,0189	0,3083	não rejeita
SS	Reversão temporal	0,54095	0,5885	não rejeita
	TPA	-1,0399	0,2984	não rejeita

Tabela B.4: H_0 associada ao tipo de teste estatístico e sua probabilidade com 99 dados sub-rogados gerados a partir série temporal da distribuição Laplace(0,1).

Técnica	Estatística de teste	Significância	Probabilidade	Decisão H_0
IID	Reversão temporal	0,50335	0,6147	não rejeita
	TPA	-1,1497	0,2503	não rejeita
GAUSS	Variância	0,23533	0,8140	não rejeita
	Assimetria	0,63184	0,5275	não rejeita
	Kurtoses	53,0197	0	rejeita*
	C_6	76,6841	0	rejeita*
	Reversão temporal	1,0908	0,2754	não rejeita
	TPA	-0,077621	0,9381	não rejeita
BOOT	Variância	0,095274	0,9241	não rejeita
	Assimetria	0,1086	0,9135	não rejeita
	Kurtoses	-0,02403	0,9808	não rejeita
	C_6	0,015248	0,9878	não rejeita
	Reversão temporal	0,90567	0,3651	não rejeita
	TPA	-0,47671	0,6336	não rejeita
SS	Reversão temporal	0,53039	0,5958	não rejeita
	TPA	-0,81129	0,4172	não rejeita

Tabela B.5: H_0 associada ao tipo de teste estatístico e sua probabilidade com 99 dados sub-rogados gerados a partir série temporal da distribuição Cauchy(0,0.001).

Técnica	Estatística de teste	Significância	Probabilidade	Decisão H_0
IID	Reversão temporal	-0,00017505	0,9999	não rejeita
	TPA	0,057345	0,9543	não rejeita
GAUSS	Variância	0,22433	0,8225	não rejeita
	Assimetria	1799,5614	0	rejeita*
	Kurtoses	66560,5708	0	rejeita*
	C_6	236940,3286	0	rejeita*
	Reversão temporal	-0,46128	0,6446	não rejeita
	TPA	-0,050189	0,9600	não rejeita
BOOT	Variância	-0,072689	0,9421	não rejeita
	Assimetria	0,64694	0,5177	não rejeita
	Kurtoses	0,81968	0,4124	não rejeita
	C_6	0,84531	0,3979	não rejeita
	Reversão temporal	-0,0019104	0,9985	não rejeita
	TPA	-0,10716	0,9147	não rejeita
SS	Reversão temporal	-1,1555	0,2479	não rejeita
	TPA	-0,56045	0,5752	não rejeita

Tabela B.6: H_0 associada ao tipo de teste estatístico e sua probabilidade com 39 dados sub-rogados gerados a partir da série da taxa diária do câmbio Real/USD, seção 4.3.

Técnica	Estatística de teste	Significância	Probabilidade	Decisão H_0
IID	Reversão temporal	-3,2423	0,0011858	rejeita*
	TPA	0,58942	0,55558	não rejeita
SS	Reversão temporal	-1,2599	0,20772	não rejeita
	TPA	3,627	0,00028673	rejeita*
FT	Variância	0,33288	0,73923	não rejeita
	Assimetria	0,34262	0,73189	não rejeita
	Kurtoses	1,6045	0,10861	não rejeita
	C_6	-1,6947	0,090127	não rejeita
	Reversão temporal	-9,5164	0	rejeita*
	TPA	-16,0765	0	rejeita*
AAFT	Variância	0,66504	0,50602	não rejeita
	Assimetria	0,6411	0,52146	não rejeita
	Kurtoses	-1,687	0,091595	não rejeita
	C_6	0,73645	0,46145	não rejeita
	Reversão temporal	-1,3356	0,18168	não rejeita
	TPA	-1,2251	0,22054	não rejeita
IAAFT-1	Variância	0,27087	0,78649	não rejeita
	Assimetria	0,057668	0,95401	não rejeita
	Kurtoses	1,4094	0,15871	não rejeita
	C_6	-1,3185	0,18733	não rejeita
	Reversão temporal	-1,3367	0,18132	não rejeita
	TPA	-2,028	0,042557	não rejeita
IAAFT-2	Variância	0,73454	0,46262	não rejeita
	Assimetria	0,27394	0,78413	não rejeita
	Kurtoses	-1,4477	0,14771	não rejeita
	C_6	0,91455	0,36043	não rejeita
	Reversão temporal	-0,79427	0,42704	não rejeita
	TPA	-3,597	0,00032186	rejeita*
CAAFT	Variância	1,2201	0,22243	não rejeita
	Assimetria	0,64567	0,51849	não rejeita
	Kurtoses	-1,6488	0,099191	não rejeita
	C_6	0,72777	0,46675	não rejeita
	Reversão temporal	-0,34565	0,7296	não rejeita
	TPA	-0,85591	0,39205	não rejeita
STAP	Variância	1,2249	0,2206	não rejeita
	Assimetria	0,40596	0,68477	não rejeita
	Kurtoses	-1,6878	0,091444	não rejeita
	C_6	0,94801	0,34312	não rejeita
	Reversão temporal	-1,3426	0,17939	não rejeita
	TPA	-0,96024	0,33693	não rejeita