

TESE DE DOUTORADO

Análise Multifractal e Seções de Lévy de
Flutuações Heterocedásticas

CÉSAR MOURA NASCIMENTO

CÉSAR MOURA NASCIMENTO

Análise Multifractal e Seções de Lévy de
Flutuações Heterocedásticas

Tese apresentada por César Moura
Nascimento ao Instituto de Física
da Universidade Federal de Alagoas
como pré-requisito para obtenção
do título de Doutor em Ciências.

Orientador: Prof. Madras Viswanathan Gandhi Mohan

Co-orientador: Prof. Iram Marcelo Gléria

Maceió, Alagoas – Brasil

Janeiro – 2008

Catálogo na fonte
Universidade Federal de Alagoas
Biblioteca Central
Divisão de Tratamento Técnico
Bibliotecária Responsável: Helena Cristina Pimentel do Vale

N244a Nascimento, César Moura.
Análise multifractal e seções de Lévy de flutuações heterocedásticas / César Moura Nascimento. – Maceió, 2008.
xii, 99 f. : graf.

Orientador: Madras Viswanathan Gandhi Mohan.

Co-Orientador: Iram Marcelo Gléria.

Tese (doutorado em Física da Matéria Condensada) – Universidade Federal de Alagoas. Instituto de Física. Maceió, 2008.

Bibliografia: f. 95-99.

1. Econofísica. 2. Análise de séries temporais. 3. Análise multifractal. I. Título.

CDU: 53:519.246.8

**PARECER DA BANCA EXAMINADORA DE DEFESA DE
TESE DE DOUTORADO**

"Análise multifractal de sessões de Levy de flutuações heterocedásticas"

por

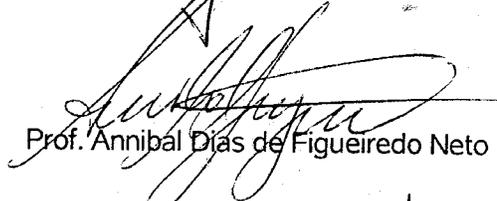
César Moura Nascimento

A Banca Examinadora composta pelos professores Madras Viswanathan Gandhi Mohan (Orientador), do Instituto de Física da Universidade Federal de Alagoas, Annibal Dias de Figueiredo Neto, do Departamento de Física da Universidade de Brasília, Paulo Sérgio da Silva Borges, do Departamento de Informática e Estatística da Universidade Federal de Santa Catarina, José Carlos Cressoni, do Instituto de Física da Universidade Federal de Alagoas e Marcelo Leite Lyra, do Instituto de Física da Universidade Federal de Alagoas, consideram o candidato aprovado com grau "A".

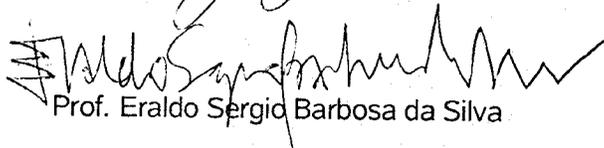
Maceió, 30 de janeiro de 2008



Prof. Madras Viswanathan Gandhi Mohan



Prof. Annibal Dias de Figueiredo Neto



Prof. Eraldo Sérgio Barbosa da Silva



Prof. José Carlos Cressoni



Prof. Marcelo Leite Lyra

Dedico este trabalho ao meu pai,
José Adilson do Nascimento (*in memoriam*),
cuja importância é inexprimível. Talvez
Honestidade e Responsabilidade
sejam os melhores valores que o definem.

AGRADECIMENTOS

Gostaria de expressar meus sinceros agradecimentos à toda minha família, principalmente à minha mãe: Ana Maria de M. Nascimento; minha companheira: Ana Carolina Matos Campos (e família); meus Irmãos: Leonardo, Bruno e Éder M. Nascimento; minha cunhada: Afra Moreira; minha sobrinha Clarinha; minhas primas: Cinara P.Sadala (e família) e Maria Helena; minhas tias Maria Augusta, Regina C. de Moura e Maria A. de Moura, meu tio: Geraldo de Moura (e família) .

Meus sinceros agradecimentos a todos meus amigos: Alexandre P. da Silva (e família), Glauco S. F. da Silva, Cleber D. Moreira (e família), Luiz F. de Ávila (e família), Marcelo P. Gomes (e família), Wallace L. da Silva, Charles P. Oliveira, Marciano A. Carneiro, Magnória V. dos Santos, Daniel M. Fernandes (e família), Arley A. Teymeny (e família), Luís Fernando, Warley Carvalho, Alessandra C. Senna (e família), Guilherme Alves (e família), Ricardo D. Dividório (e família), Rogério F. de Souza sua esposa Valdirene e seu filhinhos: Nathálya e Renato; Francisco de Assis R. Filho, Agnaldo J. dos Santos, Ildemir F. dos Santos, Caio L. Faustino, Ítalo N. de Oliveira, Elder S. Claudino, Maria Socorro S. Pereira, Wagner F. Silva, Wandearley S. Dias, Andre L. Moura, Manoel P. S. Júnior, Nuccia Carla A. Sousa, Arlan S. Ferreira, Marcelo F. Lima, José Ricardo R. Duarte, Vinícius M. Vieira, Cícero R. Silva, Elton M. Nascimento (e esposa), Itamar V. S. Lima, Pablo Young (e a todos da Cia. do Chapéu), Edmilson J. G. G. Vidal, Gentil L. S. II , Fabiano B. Santos, Juarez G. Silva,

Agradeço a todos meus amigos e colegas da Graduação do Instituto de Física da UFAL. Agradeço também ao pessoal dos bastidores, indispensáveis para todos: Ester M. Farias, Lays Rosa C. B. Leite, Arlete, Marcos Antônio S. Pereira e Maria

Edileuza.

Meus sinceros agradecimentos ao meu orientador Gandhi M. Viswanathan e ao meu co-orientador Iram M. Gléria, pelo apoio, discussões e principalmente pelo companheirismo, peças importantíssimas na realização deste trabalho.

Agradeço a M. Serva e a H. D. Jennings pelo fornecimentos dos dados aqui analisados.

Gostaria de expressar meus sinceros agradecimentos a todos os professores do IF da UFAL, em partircular a Jandir Miguel Hickmann — principal responsável pela minha vinda para o Nordeste, lugar de grandes realizações em ambos aspectos: profissional e pessoal; Marcelo Leite Lyra — chefe do nosso grupo e Francisco Anacleto Barros Fidelis de Moura — atual coordenador da Pós-Graduação. Estes dois últimos me auxiliaram, com muito boa vontade, em questões burocráticas que de certa forma garantiram o meu direito legal à apresentação deste trabalho.

Meus sinceros agradecimentos à Capes e ao CNPq pelo apoio financeiro.

Por fim agradeço a Deus por todas estas oportunidades em minha vida.

Resumo

Um importante problema em Física está relacionado ao estudo de processos estocásticos e flutuações de variáveis dinâmicas. Em uma variedade de sistemas, algumas das variáveis observadas têm uma qualidade macroscópica, no sentido de que elas representam a média ou a soma sobre o espaço ou tempo de quantidades microscópicas. Quando efeitos de memória de longo alcance ou correlação não desempenharem um papel significativo, então as condições necessárias e suficientes para a validade do Teorema do Limite Central podem ser satisfeitas. Frequentemente o segundo momento da variável em questão não diverge. Conseqüentemente em muitos exemplos importantes, as flutuações de muitos sistemas seguem uma estatística Gaussiana. Em contraste, sistemas complexos geram flutuações que muitas vezes os desviam da estatística Gaussiana. Aqui, nós focamos em duas propriedades relacionadas à flutuações Gaussianas: (i) monofractalidade e (ii) homocedasticidade. Especificamente, discutimos primeiro a questão geral sobre a natureza da relação entre multifractalidade e heterocedasticidade. Aplicamos a “multifractal detrended fluctuations analysis” a uma série temporal financeira não estacionária e de alta frequência referente à taxa cambial. Como um segundo teste, aplicamos a mesma técnica de análise para a série de áudio da quinta sinfonia de Beethoven. Obtivemos resultados que indicam que a heterocedasticidade pode causar ou aumentar a multifractalidade. Também investigamos em detalhes a convergência para o regime homocedástico e monofratal Gaussiano usando o método matemático de seções de Lévy, como previamente aplicado a séries temporais. Apresentamos conclusões relacionadas a estes questionamentos e discutimos a generalidade destes resultados no contexto da Física de sistemas complexos.

Abstract

An important problem in Physics concerns the study of stochastic processes and fluctuations away from the mean of dynamical variables. In a wide range of systems, some of the observed variables have a macroscopic quality, in the sense that they represent averages or sums over time or space of "microscopic" quantities. When long-range memory or correlation effects do not play a significant role, then the necessary and sufficient conditions for the Central Limit Theorem to hold can become satisfied. Quite often, the second moments of the studied dynamical variable do not diverge, hence in many important instances, the fluctuations of many systems follow Gaussian statistics. On the other hand, complex systems generate some variabilities that often deviate them from Gaussian statistics. Here, we focus on two properties related to Gaussian fluctuations: (i) monofractality and (ii) homoskedasticity. Specifically, we first address the general question about the nature of the relationship between multifractality and heteroscedasticity. We applied multifractal detrended fluctuation analysis to a nonstationary high frequency financial time series obtained from currency markets. As a second test, we applied the technique to the audio time series of Beethoven's fifth symphony. We obtained results suggesting that heteroscedasticity can cause or increase multifractality. We also investigate in greater detail the convergence to the homoskedastic and monofractal Gaussian regime, using the mathematical formalism of Lévy sections, as previously applied to time series. We report several conclusions related to these questions and discuss the generality of these results in the context of the physics of complex systems.

Sumário

Folha de rosto	i
Dedicatória	iii
Agradecimentos	iv
Resumo	vi
Abstract	vii
1 Introdução Geral	1
1.1 Física Estatística e Complexidade	1
1.2 Escala e Universalidade	2
1.3 Econofísica	4
1.4 Objetivos e Organização da Tese	8
2 Propriedades Estatísticas em Séries Temporais	10
2.1 Variáveis Aleatórias	10
2.2 Distribuições de Probabilidade	12
2.3 Exemplos Ilustrativos de Distribuições	19
2.4 Teorema do Limite Central	26
2.5 TLC Generalizado e α -estabilidade de Lévy	30
2.6 Estacionaridade e Tempo de Correlação	32

3	Multifractalidade versus monofractalidade	37
3.1	Auto-similaridade e dimensão fractal	38
3.2	Fractalidade em Series Temporais	43
3.3	Monofractais versus Multifractais	48
4	Seções de Lévy	56
4.1	Seções de Lévy e Generalização do TLC	56
4.2	Seções de Lévy em Séries Temporais	60
4.3	Significado Físico dos Momentos de Altas Ordens	63
5	Metodologia	70
6	Resultados e Discussões	73
6.1	Heterocedasticidade e Multifractalidade	73
6.2	Seções de Lévy e Convergência à Gaussiana	83
6.3	Estudo da Convergência das Seções de Lévy	86
7	Conclusões	90
7.1	Flutuações Heterocedásticas Multifractais	90
7.2	Relevância dos Resultados	91
7.3	Perspectivas Futuras	92

Lista de Figuras

2.1	(a) PDF referente ao lançamento de um dado de faces desiguais. . . .	15
2.2	Formas analíticas para as distribuições de Lévy.	24
3.1	Ilustração do método de contagem de caixas para o caso 2D. O objeto é seqüencialmente dividido.	39
3.2	Curva triádica Koch. Cada parte da figura representa uma geração diferente.	41
3.3	Ilustração do processo de divisão de um objeto empregado pelo mé- todo de contagens de caixas para a curva triádica de Koch.	41
3.4	Auto-similaridade para série retornos financeiros referente à taxa de câmbio da <i>Libra Esterlina</i> (GBP) frente ao <i>Dólar</i>	47
3.5	(a) Série contendo $1. 10^3$ de um total de $5. 10^4$ pontos, gerados arti- ficialmente. (b) Expoente de Hurst	54
4.1	Distribuição de freqüência das realizações de variáveis aleatórias: (a) distribuição com assimetria nula, (b) assimetria positiva e (c) assime- tria negativa.	65

4.2	Distribuição de freqüência das realizações de variáveis aleatórias: (a) curtose nula (mesocúrtica), (b) curtose positiva (leptocúrtica) e (c) curtose negativa (platicúrtica)	69
6.1	(a) Amostra de $6 \cdot 10^4$ de um total de 1 472 240 pontos contidos na série cambial do marco frente ao dólar.	74
6.2	Análise Multifractal feita a partir do método MF-DFA. (a) Expoente de Hurst generalizado. (b) Expoente τ da função partição	76
6.3	Ajuste quadrático dos espectros multifractais	78
6.4	Análise Multifractal feita a partir do método MF-DFA para a volatilidade e para a volatilidade embaralhada	79
6.5	Ajuste quadrático dos espectros multifractais	80
6.6	Aproximadamente 20% do primeiro movimento da quinta sinfonia de Beethoven	82
6.7	Comportamento da curtose para as séries das seções S_t com parâmetro $q = 2$ (linha tracejada) e para a série normal S_n (linha contínua) .	84
6.8	Comportamento da curtose para as séries das seções S_t para diferentes valores do parâmetro q	84
6.9	Comportamento da assimetria para as séries das seções S_t com parâmetro $q = 2$ (linha tracejada) e para a série normal S_n (linha contínua)	85
6.10	Comportamento da assimetria para as séries das seções S_t para diferentes valores do parâmetro q	85
6.11	Séries das seções para diferentes valores de t	87
6.12	(a) Comportamento de $\log F_2(s)$ versus $\log(s)$ das séries das seções para valores crescentes de t	88

Lista de Tabelas

3.1	Relação para 3 processos sucessivos de divisão de um objeto, empregado no método de contagens de caixas	39
5.1	Tabela ilustrativa das técnicas de embaralhamento de dados.	71

Capítulo 1

Introdução Geral

1.1 Física Estatística e Complexidade

Sistemas complexos tais como ecossistemas e mercados financeiros, possuem algo em comum com aqueles analisados por Física Estatística. A questão central de Física Estatística é tratar de fenômenos macroscópicos que resultam de interações microscópicas entre vários “indivíduos”. Considere o movimento de objetos inanimados, tais como pó de serra lançados num fluido turbulento. O comportamento de uma partícula sobre o “oceano” é influenciado por cada colisão que ela sofre a cada picosegundo ($10^{-12}s$) com aproximadamente 10^{10} moléculas microscópicas do fluido. Contudo, os avanços na área de turbulência não surgem da resolução das 10^{10} colisões num computador. O entendimento das observações é feito em termos de fenômenos coletivos que obedecem não as Leis de Newton, mas sim leis de escala em turbulência.

Como exemplo ilustrativo, considere um sistema complexo de interesse em

economia, tal como o mercado acionário. Neste caso, o valor de uma determinada ação é influenciada por todos os compradores sob variados graus. Uma vez que cada grupo de compradores decide se comportar de uma certa maneira, o valor da ação irá mudar. Contudo, cada comprador está seguindo regras bem definidas e do mesmo modo que as partículas de pó de serra seguem certas regras. Não sabemos muito sobre a completa conexão entre a quantidade microscópica e macroscópica observada, o valor da ação e as regras individuais seguidas pelos compradores. Similarmente, não sabemos muito a respeito da conexão entre a velocidade das partículas de pó de serra e as leis microscópicas seguidas pelas 10^{10} moléculas de água que colidem com as partículas a cada picosegundo.

Uma vez que a Física Estatística busca explicar o comportamento macroscópico que resulta de interações de muitos componentes microscópicos, logo ela é bastante útil para o problema geral de tentar entender sistemas complexos.

1.2 Escala e Universalidade

Fenômenos complexos têm pelo menos dois aspectos quantificáveis em comum:

- (*i*) Ausência de uma escala bem definida, possuindo uma simetria de invariância de escala.
- (*ii*) Ausência de segundo momento estacionário nos sinais gerados, ou seja, sua variância não é constante no tempo. Dizemos que os sinais possuem uma

dinâmica heterocedástica ¹.

Existe uma relação direta entre estes dois aspectos da dinâmica de sistemas complexos? A resposta é afirmativa e, uma das chaves para o entendimento desta relação vem de observações obtidas de estudos anteriores, onde fenômenos heterocedásticos seguem uma dinâmica multifractal [1] e não monofractal [2]. Monofractais referem-se aos casos onde um único expoente de escala ou *dimensão fractal* caracteriza inteiramente o sistema. Em contraste, sistemas cujas dinâmicas são multifractais necessitam de mais de um expoente para serem caracterizados completamente. Esta questão será abordada em maior detalhe no Cap. (3).

Simetria de invariância de escala é freqüentemente observada em sistemas que se encontram longe do equilíbrio, nas proximidades de *pontos críticos* (veja ref. em [3]), as vezes conhecida como *Fenômeno Crítico*. De forma contra-intuitiva, em vez de a correlação $C(r)$ entre sub-unidades de um sistema complexo separadas por uma distância r decair de forma exponencial com r : $C(r) \sim e^{-r/\xi}$, onde ξ é o comprimento de correlação (comprimento de escala característico acima do qual a função de correlação se torna desprezível), os resultados de experiências, bem como de diversos outros resultados teóricos, mostram que o decaimento exponencial da função de correlação (espacial ou temporal) é válido somente para sistemas distantes de seus pontos críticos. Nas proximidades de tais pontos a função de correlação se comporta sob a forma de lei de potência: $C(r) \sim r^{\eta-1}$, onde η é um exemplo de *expoente crítico*.

Expoentes críticos, tais como η , possuem um caráter *universal*, sendo in-

¹ O termo Heterocedástico tem sua origem no Grego “hetero” (diferente) e “skedastios” (dispersão) — dispersão diferente. O termo heteroscedástico também é correto e freqüentemente utilizado.

dependente da natureza do sistema. A descoberta desta universalidade implica a existência de mecanismos mais profundos, geralmente simples, responsáveis pelo comportamento dos sistemas críticos [4]. Tais idéias têm guiado os físicos em problemas de investigação interdisciplinares e puseram em evidência semelhanças entre problemas e disciplinas aparentemente muito diferentes. Retomaremos a tais questões posteriormente.

Sem perda de generalidade e admitindo a importância de fenômenos econômicos, optamos por analisar as propriedades estatísticas e de escala de séries temporais financeiras. A riqueza e complexidade dinâmica do mercado financeiro deu origem a uma nova área denominada *Econofísica* [1]. O interesse pela Econofísica aumentou bastante com a publicação dos trabalhos de Rosario Mantegna e de H. E. Stanley, que em 1995 analisaram as variações no índice Standard & Poor's 500 da bolsa de Nova York [5].

1.3 Econofísica

A Econofísica é o estudo de mecanismos econômicos a partir de uma óptica empírica e lógica visando uma correlação com a Física. As origens da Econofísica podem ser remetidas ao ano de 1960, com os trabalhos do matemático Benoit Mandelbrot [1] ou mesmo ao ano de 1900, com a tese de doutorado de Louis Bachelier, sobre a especulação financeira, intitulada “Théorie de la spéculation” [6] (veja também ref. em [7]).

Mesmo antes da descoberta do *movimento browniano* (MB) em 1827 pelo

botânico Robert Brown [8], ao observar o movimento desordenado de grãos de pólen “dissolvidos” em água, cuja explicação se deu somente com os trabalhos de Einstein em 1905 ², Bachelier propôs um modelo tipo MB para a dinâmica de preços de ativos financeiros. Do ponto de vista econômico, Bachelier investigou variações no preço, enquanto economistas tratavam principalmente de variações no logaritmo dos preços; fato este que não diminui o valor do trabalho pioneiro de Bachelier [5].

Retorno e Volatilidade Financeira

Mais importante que saber o valor de ativo financeiro em um dado instante de tempo k , é saber o quanto de lucro é obtido numa transação envolvendo a compra e a venda deste ativo, separadas por um dado intervalo de tempo. Tome o preço do ativo em instantes consecutivos de tempo. A diferença dos preços entre estes dois instantes é dada por $\Delta S(k) = S(k) - S(k - 1)$. Define-se a variação relativa de preços ou retorno líquido simples, entre os mesmos instantes de tempo, por:

$$R(k) = \frac{S(k) - S(k - 1)}{S(k - 1)}. \quad (1.1)$$

Note que, ao reescrever a eq. 1.1 na forma $1 + R(k) = \frac{S(k)}{S(k-1)}$, aplicando-se o logaritmo na expressão anterior obtém-se:

$$\ln \left[\frac{S(k)}{S(k - 1)} \right] = \ln[1 + R(k)] \quad (1.2)$$

e, como para pequenos valores de $R(k)$ (como geralmente é o caso), pode-se escrever

² O MB foi o tema da tese de doutoramento de Einstein e, apesar de ele ter ganhado o prêmio Nobel pelo *Efeito Fotoelétrico*, sua teoria para o MB é considerada um dos trabalhos mais notáveis da sua carreira, sendo responsável por um grande avanço nos campos da química e da física.

$\ln[1 + R(k)] \approx R(k)$, “definimos” o *log-retorno* como:

$$r(k) = \ln \left[\frac{S(k)}{S(k-1)} \right]. \quad (1.3)$$

Desta forma, analisar a série de log-retorno dos preços é equivalente a analisar aquela contendo o retorno líquido simples. Esta possui a informação relevante que é o lucro (ganho) e/ou perda, obtido em uma dada transação.

Uma segunda quantidade de grande importância em finanças é a medida do risco de um investimento. Esta medida está relacionada a outra quantidade, denominada de volatilidade. A volatilidade é definida como o desvio padrão (estas definições estatísticas são abordadas no Cap. 2) do retorno para um dado intervalo de tempo τ , matematicamente:

$$\sigma(r_\tau) \approx |r_\tau|. \quad (1.4)$$

Logo, neste trabalho a volatilidade é definida como o valor absoluto do retorno financeiro.

Enquanto a série temporal de preços de um ativo financeiro possui ambos — média (primeiro momento estatístico) e variância (segundo momento estatístico) — dependentes do tempo, a série de retornos, obtida da série original de preços através da transformação dada pela eq. 1.3, possui o primeiro momento aproximadamente constante, mantendo apenas o caráter heterocedástico da variância.

Hipótese de Mercado Eficiente

A Hipótese do Mercado Eficiente foi formulada durante década de 70 por Eugene Fama [9], que em sua tese de doutorado demonstrou que em mercados onde

há um grande número de investidores bem informados, o preço atual de um ativo financeiro reflete todas as informações e expectativas dos participantes do mercado. Consequentemente, nenhum lucro pode ser obtido a partir de negociações baseadas em informações do mercado, pois tais informações já foram incorporadas no preço do ativo [5], em outras palavras, a hipótese de mercado eficiente afirma que a cotação de um ativo é completamente imprevisível. Um mercado como este possui um comportamento similar àquele das partículas de pó de serra num fluido, ou seja, é um tipo de random walk unidimensional, onde a série dos preços é descorrelacionada no tempo.

Existem duas maneiras distintas de abordagem em finanças [10]: *Abordagem Fundamentalista* e *Abordagem Técnica ou Grafista*. A primeira abordagem se baseia na idéia da existência de uma correlação lógica entre o valor intrínseco de uma ação e seu preço de mercado. Este valor levaria em conta todo o patrimônio da empresa cuja ação está associada, bem como fatores externos: desempenho e posição no respectivo setor de atuação, grau de atualização tecnológica dos empreendimentos da empresa, até mesmo fatores políticos relacionados à área de atuação da empresa. Já a segunda abordagem se utiliza de dados do passado (séries temporais) como fonte de previsibilidade do comportamento dos preços num instante de tempo futuro, tendo como objetivo identificar oportunidades de compra e venda de ações, determinar limites de oscilação de preços, tornando, desta forma, possível o estabelecimento de estratégias de risco.

Num mercado que satisfaça completamente a Hipótese de Mercado Eficiente, não existiria forma dos fundamentalistas obterem lucros acima da média e nem dos grafistas teriam como antecipar novas tendências ou movimento dos preços. Obter

lucro acima da média num mercado como este seria apenas uma questão de sorte.

De fato, a Hipótese de Mercado Eficiente é apenas uma idealização, ela serve como referência para criação de modelos matemáticos.

Entre os problemas abertos Econofísica estão os seguintes: *(i)* a origem das caudas grossas nas distribuições de retornos financeiros, *(ii)* a caracterização das propriedades multifractais destas séries, *(iii)* a importância do papel de correlações de longo alcance multifractais nos retornos absolutos e *(iv)* a convergência das séries para o regime Gaussiano. Nesta última linha, o uso de um teorema denominado Seções de Lévy tem se mostrado útil, voltaremos nesta questão no Cap. 4

1.4 Objetivos e Organização da Tese

Nesta tese investigamos o grau com que a multifractalidade e a heterocedasticidade estão relacionados. Um causa o outro? Ou são completamente decorrelacionados? Também investigamos o processo de convergência de séries temporais ao regime Gaussiano monofractal e homocedástico monofractal via Seções de Lévy. Esse regime representa um importante aspecto em Física Estatística de sistemas em equilíbrio. Analisamos também a “natureza” deste processo, ou seja, o modo que as Seções de Lévy promovem a convergência. No Cap. 6 desta tese estes questionamentos são respondidos.

Esta tese está dividida da seguinte forma: O Cap. 2 é dedicado ao estudo de conceitos estatísticos, no Cap. 3 mostramos os conceitos de fractalidade aplicados a séries temporais, no Cap. 4 estudamos o processo de convergência ao regime Gaus-

siano via Seções de Lévy, os resultados e suas discussões são apresentados no Cap. 6 e no Cap. 7 apresentaremos as conclusões.

Capítulo 2

Propriedades Estatísticas em Séries Temporais

Como caracterizar quantitativamente as propriedades estatísticas de séries temporais? Este capítulo é dedicado a esta questão. Apresentamos aqui conceitos estatísticos básicos tais como: o de variáveis aleatórias, funções densidade de probabilidade, momentos estatísticos, teorema do limite central (TLC), TLC generalizado, funções de correlação e espectro de potência.

2.1 Variáveis Aleatórias

Na Mecânica Clássica, uma vez conhecida a posição e velocidade de uma partícula para um instante de tempo t qualquer, estas quantidades se tornam bem definidas para qualquer instante de tempo posterior ao aplicar as *Leis de Newton* da Mecânica. Imagine agora um sistema composto por um número N de partículas. Na sua descrição por intermédio das leis da *Leis de Newton* seria necessário a resolução

de um conjunto de $3N$ equações diferenciais (uma para cada componente de posição) no caso de o sistema não possuir vínculo, além de ser necessário conhecer os valores da posição e velocidade de cada uma das partículas para um dado instante de tempo. Esse procedimento se torna inviável em ocasiões onde as condições iniciais não são bem definidas e/ou o número N de partículas for muito grande. Para estes casos um tratamento estatístico do sistema se torna essencial.

No primeiro caso onde existe uma relação matemática bem definida para cada componente de posição e momento da partícula em função do tempo, dizemos que estas quantidades são determinísticas. Já no segundo caso, onde não existe tal relação definida para as quantidades físicas do sistema, dizemos que estas quantidades são aleatórias ou estocásticas, e é comum a utilização de letras maiúsculas para representá-las. Digamos que alguém queira realizar um experimento, por exemplo, aquele consistindo no lançamento de um total de 10 moedas. Defina uma variável X como a diferença entre o número de moedas que caem com a face cara voltada para cima e aquele número em que a face coroa é a superior. O conjunto de realizações possíveis para variável X , comumente representados por letras minúsculas, é chamado de domínio da variável ou espaço amostral (é comum representar o espaço amostral relacionado a uma variável aleatória X por S), que para este exemplo em particular é composto por

$$x = 0, 1, \dots, 10.$$

Antes de as moedas serem lançadas só faz sentido falar na probabilidade desta ou daquela realização x para a variável X ocorrer. Uma vez realizado o experimento, falar de probabilidade perde o sentido pois sabemos o valor de x com certeza. Esta é a principal característica de uma dada variável aleatória: seu valor só se mostra

através de uma medida.

As variáveis aleatórias podem ser discretas ou contínuas. Ela será discreta quando a menor diferença entre dois possíveis valores de suas realizações for finita, como é o caso do exemplo anterior. No caso de uma variável contínua se, ao passar de um valor real a para um outro valor b , ela assumir todos os valores intermediários entre a e b . Como exemplo de uma variável aleatória contínua temos a medida da temperatura T em um dado experimento. Em geral, medições dão origem a dados contínuos e contagens ou enumerações a dados discretos [11].

2.2 Distribuições de Probabilidade

A probabilidade é a quantização das expectativas em relação ao resultado de um experimento. Seja a uma possível realização de um experimento e $p(a)$ sua respectiva probabilidade (a pode representar as flutuações na taxa cambial entre duas moedas quaisquer; poderia ser o número referente à face de um dado lançado, ou qualquer outra realização de uma variável aleatória); quando N experimentos idênticos forem realizados, espera-se que a ocorra $Np(a)$ vezes. No limite do número de experimentos tendendo a infinito ($N \rightarrow \infty$), espera-se que a fração do número de experimentos que resultam em a seja $p(a)$. Um caso especial e muito importante é aquele no qual um experimento possa resultar em n realizações igualmente prováveis. Se um número m destes resultados corresponde à realização de a , então, $p(a)$ será dado por:

$$p(a) = \frac{m}{n}.$$

Dada uma possível realização de uma variável discreta X e sua respectiva probabilidade de ocorrência $p(x)$, é possível “mapear” o espaço amostral S , definido pelo conjunto $\{x_i\}$ das realizações de X , no espaço de probabilidade definido pelo conjunto $\{p_i\}$ sobre S . Aqui o índice i é para diferenciar as possíveis realizações de X de suas respectivas probabilidades de ocorrência. O conjunto de probabilidades $\{p_i\}$ definido sobre o espaço S deverá satisfazer as seguintes condições:

1. Condição de positividade:

$$p_i \geq 0 \quad (2.1)$$

e

2. Condição de normalização:

$$\sum_{i=1}^n p_i = 1 . \quad (2.2)$$

A probabilidade em um experimento de a variável aleatória se encontrar entre os valores a e b será dada por:

$$P(a \leq x \leq b) = \sum_{a \leq x \leq b} p(x_i) , \quad (2.3)$$

onde a soma percorre todos os valores das possíveis realizações entre a e b .

A *Função Densidade de Probabilidade* (PDF em inglês), $P_X(x)$, é definida como:

$$P_X(x) = \sum_{i=1}^n p_i \delta(x - x_i) , \quad (2.4)$$

onde $\delta(x - x_i)$ é a função Delta de Dirac. Já a *Função Distribuição de Probabilidade* (DF em inglês), $F_X(x)$, é definida como:

$$F_X(x) = \int_{-\infty}^x dy P_X(y) = \sum_{i=1}^n p_i \theta(x - x_i) , \quad (2.5)$$

onde $\theta(x - a)$ é a função de passo Heaviside:

$$\theta(x - a) = \begin{cases} 0 & \text{se } x < a \\ \frac{1}{2} & \text{se } x = a \\ 1 & \text{se } x > a . \end{cases} \quad (2.6)$$

A DF representa a probabilidade da variável X ter sua realização pertencente ao intervalo $] - \infty, x]$. O fato da PDF ser positiva definida implica que a DF seja sempre uma função monotônica crescente de x , tendo como valores limites $F_X(-\infty) = 0$ e $F_X(\infty) = 1$. Note que a PDF pode ser obtida a partir da derivada da DF ($P_X(x) = dF_X(x)/dx$).

A Fig. 2.1 mostra PDF e DF de uma variável aleatória discreta X referente ao lançamento de um dado com “faces desiguais”. O domínio de X consiste dos valores x_i ($i = 1, \dots, 6$), referente à realização da i -ésima face num lançamento.

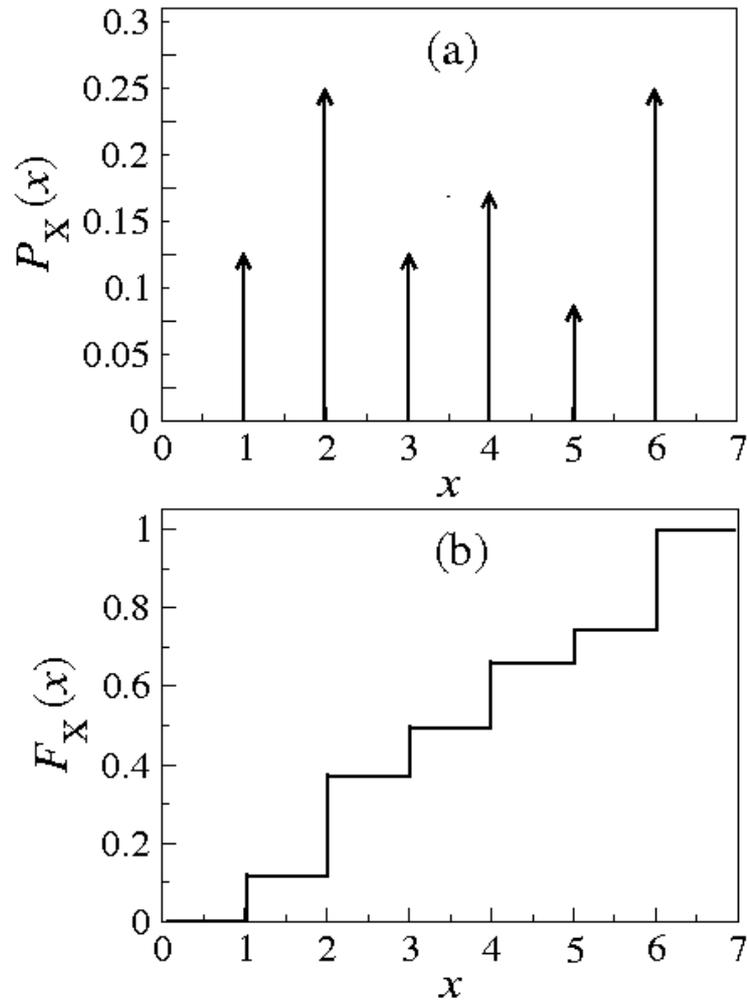


Figura 2.1: (a) PDF referente ao lançamento de um dado de faces desiguais. Os valores x_i com $i = 1, 2, \dots$ e 6 são as possíveis realizações de X , já o tamanho das setas indicam suas respectivas probabilidades de ocorrência: $p_1 = \frac{1}{8}$, $p_2 = \frac{1}{4}$, $p_3 = \frac{1}{8}$, $p_4 = \frac{1}{6}$, $p_5 = \frac{1}{12}$, $p_6 = \frac{3}{12}$. (b) Função Distribuição de Probabilidade associada a $P_X(x)$. Note o comportamento monotônico crescente da função DF.

No caso de X ser uma variável aleatória contínua, a eq. 2.3 deve ser repassada por

$$P(a \leq x \leq b) = \int_a^b dx P_X(x), \quad (2.7)$$

onde $P_X(x)$ é a PDF contínua (assumida existir) associada à variável aleatória X e, $P_X(x)dx$ é a probabilidade da variável X ser obtida para os valores compreendidos entre x e $x + dx$. A PDF deverá satisfazer as condições $P_X(x) \geq 0$ e $\int_{-\infty}^{+\infty} dy P_X(y) = 1$. Já a DF para uma variável aleatória contínua X é dada por

$$F_X(x) = \int_{-\infty}^x dy P_X(y). \quad (2.8)$$

Momentos Estatísticos

Todas as informações referentes a uma dada variável aleatória X estão contidas em sua PDF e na auto-correlação presente em suas várias realizações. Entretanto, dificilmente em situações reais a determinação da $P_X(x)$ é possível. Nestes casos, se faz necessária a obtenção de informações a partir dos momentos relacionados a X (além da função de auto-correlação descrita posteriormente).

O primeiro momento, ou simplesmente valor médio³, referente a um conjunto de realizações da variável aleatória X , é escrito como:

$$\langle X \rangle = \frac{1}{N} \sum_{i=1}^N x_i, \quad (2.9)$$

onde x_i é referente a i -ésima realização da variável X e N é o número de suas

³ O valor médio da variável X é freqüentemente confundido com outras duas quantidades: o valor mais provável x_p e a mediana x_m . O valor mais provável x_p é tal que $P_X(x_p)$ é máximo e a mediana é tal que $\int_{-\infty}^{x_m} y P_X(y) = \int_{x_m}^{\infty} y P_X(y) = \frac{1}{2}$.

realizações. O valor médio $\langle X \rangle$ é muitas vezes denotado por $E[X]$ ou simplesmente μ .

O momento de p -ésima ordem em torno da origem é escrito na forma:

$$\mu_p = \frac{1}{N} \sum_{i=1}^N x_i^p . \quad (2.10)$$

No caso de uma variável aleatória contínua X , o p -ésimo momento é definido como:

$$\langle X \rangle \equiv \int_{-\infty}^{\infty} dx x^p P_X(x) . \quad (2.11)$$

Um caso muito importante refere-se aos momentos de alta ordem em torno da média, os quais são obtidos repassando x_i por $(x_i - \mu)$ na eq. 2.10 e x por $(x - \mu)$ na eq. 2.11 no caso de variáveis contínuas:

$$\mu_p \equiv \int_{-\infty}^{\infty} dx (x - \mu)^p P_X(x) . \quad (2.12)$$

O segundo momento é conhecido como a variância, a qual, para conjunto de realizações da variável aleatória X é escrita na forma:

$$\sigma_X^2 = \frac{1}{N} \sum_{i=1}^N [x_i - \langle X_i \rangle]^2 . \quad (2.13)$$

Já o desvio padrão é obtido tomando a raiz quadrada da variância.

Enquanto o primeiro momento é referente a uma medida de tendência central (tende a se localizar em um ponto central), a variância é uma medida de dispersão das realizações de X em torno da média. Atualmente tem se tornado comum o uso de momentos de alta ordem normalizados no estudo das propriedades estatísticas de sistemas. Abordaremos esta questão na seção (4.3).

Funções geradora de momento e característica

A *função geradora de momento*, $M_X(k)$, referente à variável aleatória X cuja PDF é $P_X(x)$, é definida como

$$M_X(k) \equiv \langle e^{kX} \rangle = \int_{-\infty}^{\infty} dx e^{kx} P_X(x). \quad (2.14)$$

Expandindo em uma série de Taylor:

$$\begin{aligned} M_X(k) &= \int_{-\infty}^{\infty} dx e^{kx} P_X(x) \\ &= \int_{-\infty}^{\infty} dx \sum_{n=0}^{\infty} \left(\frac{kx}{n!} \right)^n P_X(x) \\ &= \int_{-\infty}^{\infty} dx \left(1 + \frac{kx}{1!} + \frac{(kx)^2}{2!} + O[k^2] \right) P_X(x) \\ &= \int_{-\infty}^{\infty} dx P_X(x) + \frac{k}{1!} \int_{-\infty}^{\infty} dx x P_X(x) + \frac{k^2}{2!} \int_{-\infty}^{\infty} dx x^2 P_X(x) + O[k^2] \\ &= 1 + \frac{k}{1!} \langle X \rangle + \frac{k^2}{2!} \langle X^2 \rangle + O[k^2] \\ &= \sum_{n=0}^{\infty} \frac{k^n \langle X^n \rangle}{n!}. \end{aligned} \quad (2.15)$$

A expansão na eq. 2.15 só faz sentido quando os momentos de alta ordem em torno da origem forem pequenos, garantindo a convergência da série [12, 13]. Quando este for o caso, vemos que o momento de p -ésima ordem em torno da origem pode ser obtido a partir da mesma ordem de derivada da *função geradora de momento*:

$$\mu_p = \langle X^p \rangle = \lim_{k \rightarrow 0} \frac{d^p}{dk^p} M_X(k). \quad (2.16)$$

A *função característica* (FC) pode ser obtida a partir da eq. 2.14 pela transformação $k \rightarrow ik$:

$$f_X(k) \equiv \langle e^{ikX} \rangle = \int_{-\infty}^{\infty} dx e^{ikx} P_X(x). \quad (2.17)$$

Note que a FC é basicamente a transformada de Fourier da PDF, e é uma função contínua de k com seguintes propriedades: $f_X(k=0) = 1$, $|f_X(k)| \leq 1$ e $f_X(-k) = f_X^*(k)$. Logo, a PDF é dada pela transformada de Fourier inversa da FC

$$P_X(x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} dk e^{-ikx} f_X(k). \quad (2.18)$$

Já a eq. 2.16 pode ser reescrita em termos da FC:

$$\mu_p = \langle X^p \rangle = \lim_{k \rightarrow 0} (-i)^p \frac{d^p}{dk^p} M_X(k). \quad (2.19)$$

2.3 Exemplos Ilustrativos de Distribuições

Distribuição Binomial

Eventos caracterizados por sucessivas tentativas, independentes entre si, cada qual com apenas duas possibilidades, são descritos pela *Distribuição Binomial*. Um exemplo clássico em física onde a Distribuição Binomial é aplicada é a caminhada aleatória.

Imagine uma partícula confinada a se mover em uma dimensão. Suponha que a partícula possa efetuar deslocamentos de tamanho fixo ℓ , com probabilidade p do deslocamento no sentido positivo (direita) e probabilidade $q = 1 - p$ no sentido negativo (esquerda), de forma que a posição da partícula em relação à origem pode ser escrita da forma $x = m\ell$, com $m \in \mathbb{Z}$ correspondente à diferença entre o número de passos para a direita (n_r) e aquele para a esquerda (n_l). É importante notar que,

uma vez fixada as probabilidades p e q , os passos tornam-se estatisticamente independentes. Desta forma, a posição da partícula no n -ésimo deslocamento dependerá de sua posição anterior. Já o fato de ela ter se deslocado para a esquerda ou direita não influencia no sentido do passo seguinte.

Utilizando o fato dos passos serem estatisticamente independentes, a probabilidade da partícula ter realizado n_r deslocamentos para a direita e n_l para a esquerda após n deslocamentos ($n = n_r + n_l$) pode ser escrita na forma:

$$\begin{aligned} W(n_r) &= \frac{n!}{n_r!n_l!} p^{n_r} q^{n_l} \\ &= \frac{n!}{n_r!(n - n_r)!} p^{n_r} (1 - p)^{n - n_r}, \end{aligned} \quad (2.20)$$

onde o fator $\frac{n!}{n_r!n_l!}$ é referente ao número de combinações possíveis de deslocamentos resultando na mesma situação final [14].

Utilizando as relações $m = n_r - n_l$ e $n = n_r + n_l$ na eq. 2.20 obtém-se expressão para a probabilidade da partícula ser encontrada numa posição $x = m\ell$ “qualquer” dados n passos:

$$P_n(m) = \frac{n!}{[\frac{1}{2}(n+m)]! [\frac{1}{2}(n-m)]!} p^{\frac{1}{2}(n+m)} (1-p)^{\frac{1}{2}(n-m)}. \quad (2.21)$$

Um caso mais simples é aquele em que $p = q = \frac{1}{2}$. Note que as condições de normalização (eq. 2.2) e positividade (eq. 2.1) são satisfeitas.

A difusão de um gás e aquela das partículas de açúcar em água são exemplos de caminhadas aleatórias tridimensionais com probabilidades iguais de deslocamento das moléculas em todas as direções.

Distribuição Normal

A *Distribuição Gaussiana ou Normal* (Fig. 2.2) de uma variável contínua X é definida pela PDF

$$P_X(x) = \frac{1}{\sqrt{2\pi} \sigma_X} \exp \left[-\frac{(x - \mu)^2}{2\sigma_X^2} \right], \quad (2.22)$$

onde μ e σ_X são a média e variância referentes à variável aleatória X .

A Distribuição normal também é conhecida como curva em forma de sino devido ao seu formato. A importância da distribuição Gaussiana se mostra na sua enorme ocorrência na natureza, fato decorrente do *Teorema do Limite Central* (TLC) que será apresentado na seção (2.4).

Distribuição de Lévy

O matemático francês Paul Pierre Lévy [15] desenvolveu uma classe “geral” de PDF cuja FC é da forma

$$f_X(k; \alpha, \beta, \gamma, \mu) = \exp[-ik\mu - \gamma|k|^\alpha (1 - i\beta \operatorname{sgn}(k) \tan(\pi\alpha/2))], \quad (2.23)$$

onde $\mu \in \mathbb{R}$ é referente à média da variável aleatória X ; γ é um fator de escala e assume valores positivos; $\alpha \in]0, 2]$ é um parâmetro referente ao achatamento da distribuição (curtose); já o parâmetro $\beta \in [-1, 1]$ refere-se a assimetria (skewness) da distribuição. A função $\operatorname{sgn} k$ é definida na forma

$$\operatorname{sgn} k = \begin{cases} -1 & \text{se } k < 0 \\ 0 & \text{se } k = 0 \\ 1 & \text{se } k > 0. \end{cases} \quad (2.24)$$

As distribuições de Lévy provenientes da FC descrita na eq. 2.23 se enquadram numa classe de distribuições do tipo cauda grossa (“fat tailed”). As distribuições de Lévy (para o caso de $\alpha < 2$) têm a propriedade de decair de forma lenta em função da variável aleatória. Este comportamento das caudas permite que eventos extremos sejam mais fáceis de serem observados do que no caso de uma distribuição com decaimento rápido, por exemplo a Gaussiana. Esta propriedade é decorrente do fato de que para valores grandes do módulo das realizações da variável aleatória X , as distribuições de Lévy se comportam em forma de lei de potência com expoente entre 1 e 3:

$$P_X(|x|) \sim \frac{1}{|x|^{1+\alpha}} . \quad (2.25)$$

Devido a esse comportamento da PDF, os momentos de p -ésima ordem ($\langle X^p \rangle$) divergem quando $p > \alpha$ e $\alpha < 2$. Veremos no cap. 3 que sistemas regidos por lei de potência possuem propriedades fractais.

As distribuições de Lévy são amplamente aplicadas em várias áreas, como exemplo podemos citar a economia (veja ref. em [16]) onde distribuições de Lévy são utilizadas para se ajustar a parte central das distribuições de retorno financeiro. Uma outra aplicação muito interessante das distribuições de Lévy é referente a estratégias de busca de alimentos por parte de animais. Acredita-se que, satisfeitos alguns requisitos como alto poder de detecção da presa e abundância de alimento, alguns animais realizam os chamados vôos de Lévy (“Lévy Flight”) na busca de alimento. Ao contrário do movimento aleatório comum (MB) onde os passos são frequentemente curtos — regido pela distribuição Gaussiana, em um vôo de Lévy é possível que passos arbitrariamente grandes seja efetuados. Desta forma, como uma maneira de aumentar as chances de sobrevivência, alguns animais poderiam adotar

os vôos de Lévy como uma estratégia de otimização de suas buscas.

Em geral não existem formas analíticas para as distribuições de Lévy e, somente para alguns casos especiais, expressões analíticas podem ser obtidas. Segue listado abaixo tais casos:

- Distribuição Normal, $\alpha = 2$:

$$P_X(x) = \frac{1}{\sqrt{4\gamma\pi}} \exp \left[-\frac{(x-\mu)^2}{4\gamma} \right]; \quad (2.26)$$

- Distribuição de Cauchy ou Lorentziana, $\alpha = 1$, $\beta = 0$:

$$P_X(x) = \frac{\gamma}{\pi} \frac{1}{(x-\mu)^2 + \gamma^2}; \quad (2.27)$$

- Distribuição de Lévy-Smirnov, $\alpha = \frac{1}{2}$, $\beta = 1$:

$$P_X(x) = \sqrt{\frac{\gamma}{2\pi}} \frac{\exp \left[-\frac{\gamma}{2(x-\mu)} \right]}{(x-\mu)^{\frac{3}{2}}}. \quad (2.28)$$

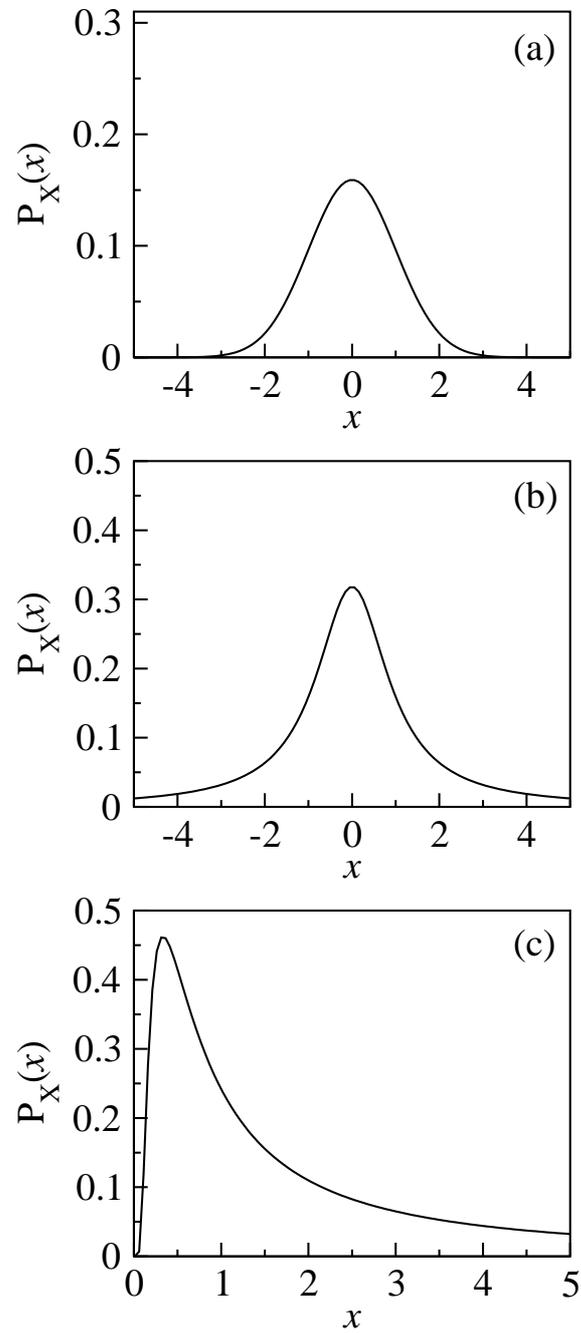


Figura 2.2: Formas analíticas para as distribuições de Lévy. (a) Distribuição Gaussiana para o caso de média nula e variância unitária. (b) Distribuição de Cauchy ou Lorentziana com média nula e variância unitária. (c) Distribuição de Lévy-Smirnov.

Estabilidade Gaussiana

Um conceito muito importante em estatística é aquele de estabilidade de uma distribuição. Seja S_2 a soma de 2 variáveis aleatórias: $S_2 = X_1 + X_2$, a distribuição de S_2 é dita estável se ela for idêntica àquela referente a X_1 e X_2 , ou seja, se X_1 , X_2 e S_2 forem governadas pela mesma PDF [17]. A PDF de S_2 para as variáveis independentes (elas não necessitam possuir idênticas PDF) é dada pela convolução entre as PDF's de cada uma das variáveis:

$$P_{S_2}(s_2) = \int_{-\infty}^{\infty} dx_1 P_{X_1}(x_1)P_{S_2}(s_2 - x_1) = P_{X_1}(x_1) \otimes P_{X_2}(x_2). \quad (2.29)$$

O teorema da convolução [13] diz que a transformada de Fourier da convolução de duas funções é igual ao produto da transformada de Fourier de cada função. Simbolizando a transformada de Fourier por \mathcal{F} tem-se que

$$\begin{aligned} \mathcal{F}[P_{S_2}(s_2)] &= \mathcal{F}[P_{X_1}(x_1) \otimes P_{X_2}(x_2)] \\ &= \mathcal{F}[P_{X_1}(x_1)]\mathcal{F}[P_{X_2}(x_2)]. \end{aligned} \quad (2.30)$$

Para o caso da soma de n variáveis aleatórias independentes tem-se que

$$\mathcal{F}[P_{S_n}(s_n)] = \prod_{i=1}^n \mathcal{F}[P_{X_i}(x_i)]. \quad (2.31)$$

Logo, a partir da eq. 2.17 percebe-se que a eq. 2.31 é a própria FC referente à S_n . Supondo que as variáveis além de independentes possuem idênticas PDF (variáveis *i.i.d.*)

$$f_{S_n}(k) = [f_X(x)]^n. \quad (2.32)$$

Para o caso de uma distribuição Gaussiana tem-se que

$$f_X(k) = \exp\left[-\frac{\sigma_X^2}{2}k^2\right] \quad (2.33)$$

logo

$$f_{S_n}(s_n) = [f_X(k)]^n = \exp \left[-\frac{\tilde{\sigma}_X^2}{2} k^2 \right] \quad (2.34)$$

onde $\tilde{\sigma}_X = \sigma_X \sqrt{n}$, conseqüentemente, a PDF referente à S_n também manterá sua forma funcional, fazendo da Gaussiana uma distribuição estável. A convolução de n Gaussianas é também Gaussiana. Isto é um exemplo de estabilidade.

2.4 Teorema do Limite Central

Dado um conjunto $\{x_i\}$ ($i = 1, 2, \dots, n$) de realizações independentes de uma variável aleatória X (com média e variância finitas), seja y_n o desvio da média das realizações de X

$$\begin{aligned} y_n &= \frac{1}{n} (x_1 + x_2 + \dots + x_n) - \langle X \rangle \\ &= z_1 + z_2 + \dots + z_n, \end{aligned} \quad (2.35)$$

com $z_i = \frac{1}{n} (x_i - \langle X \rangle)$, o TLC afirma que a PDF $P_{Y_n}(y_n)$ se aproxima de uma Gaussiana à medida que n tende a infinito; matematicamente:

$$\lim_{n \rightarrow \infty} P_{Y_n}(y_n) = \sqrt{\frac{n}{2\pi\sigma_X^2}} \exp \left[-\frac{ny_n^2}{2\sigma_X^2} \right]. \quad (2.36)$$

A demonstração deste teorema pode ser feito via função característica. Faça $f_Z(k, n)$ a FC relacionada a $z = \frac{1}{n} (x - \langle X \rangle)$:

$$\begin{aligned}
f_Z(k, n) &= \int_{-\infty}^{\infty} dx \exp \left[i \frac{k}{n} (x - \langle X \rangle) \right] P_X(x) \\
&= \sum_{p=0}^{\infty} \frac{1}{p!} \left(\frac{ik}{n} \right)^p \int_{-\infty}^{\infty} dx (x - \langle X \rangle)^p P_X(x) \\
&= 1 - \frac{1}{2} \left(\frac{k}{n} \right)^2 \sigma_X^2 + \sum_{p=3}^{\infty} \frac{1}{p!} \left(\frac{ik}{n} \right)^p \mu_p.
\end{aligned}$$

Neste ponto podemos desprezar os momentos de alta ordem, o que é justificável tanto pelo fato de supor a convergência da série de momentos quanto pelo fato de considerar valores grandes de n , obtendo:

$$f_Z(k, n) \equiv f_Z \left(\frac{k}{\sqrt{n}} \right) \approx 1 - \frac{1}{2n} \left(\frac{k}{\sqrt{n}} \right)^2 \sigma_X^2, \quad (2.37)$$

onde σ_X^2 é variância de X (segundo momento em torno da média) definida na seção (2.2). A FC, $f_{Y_n}(y_n)$, pode ser escrita na forma

$$\begin{aligned}
f_{Y_n}(k) &= \left[f_Z \left(\frac{k}{\sqrt{n}} \right) \right]^n \\
&\approx \left(1 - \frac{1}{2n} \left(\frac{k}{\sqrt{n}} \right)^2 \sigma_X^2 \right)^n \\
&\approx \exp \left[-\frac{1}{2} \frac{k^2 \sigma_X^2}{n} \right].
\end{aligned} \quad (2.38)$$

Utilizando a eq. 2.18 temos que

$$\begin{aligned}
P_{Y_n}(y_n) &= \frac{1}{2\pi} \int_{-\infty}^{\infty} dk \exp[-iky_n] \exp \left[-\frac{1}{2} \frac{k^2 \sigma_X^2}{n} \right] \\
&= \frac{1}{2\pi} \exp \left[-\frac{ny_n^2}{2\sigma_X^2} \right] \int_{-\infty}^{\infty} dk \exp \left[-\frac{1}{2n} \left(k\sigma_X - \frac{iny_n}{\sigma_X} \right)^2 \right].
\end{aligned}$$

Fazendo a mudança de variável: $\zeta = k\sigma_X - \frac{iny_n}{\sigma_X}$, logo $d\zeta = \sigma_X dk$

$$P_{Y_n}(y_n) = \frac{1}{2\pi\sigma_X} \exp \left[-\frac{ny_n^2}{2\sigma_X^2} \right] \int_{-\infty - \frac{iny_n}{\sigma_X}}^{\infty - \frac{iny_n}{\sigma_X}} d\zeta \exp \left[-\frac{1}{2n} \zeta^2 \right], \quad (2.39)$$

$$P_{Y_n}(y_n) = \frac{1}{2\pi\sigma_X} \exp\left[-\frac{ny_n^2}{2\sigma_X^2}\right] \sqrt{2n\pi},$$

o que nos leva à demonstração do TLC:

$$P_{Y_n}(y_n) = \sqrt{\frac{n}{2\pi\sigma_X^2}} \exp\left[-\frac{ny_n^2}{2\sigma_X^2}\right]. \quad (2.40)$$

Devemos notar que na demonstração do TLC foi utilizado a definição da PDF a partir da FC (eq. 2.18) e, as únicas suposições feitas foram sobre o conjunto de realizações da variável aleatória X . Isto implica que qualquer distribuição que satisfaça as condições do TLC tem a propriedade de convergir para a distribuição Gaussiana quando o número realizações tendem a infinito. Por isso é muito comum se dizer que a PDF Gaussiana representa um atrator no espaço das PDF's. A caráter ilustrativo mostraremos este resultado para a Distribuição Binomial apresentada anteriormente.

Para grandes valores de n e n_r , a distribuição binomial $W(n_r)$ (eq. 2.20) exibirá um máximo pronunciado em algum valor $n_r = \tilde{n}_r$ e sofrerá um decaimento rápido para valores próximos de \tilde{n}_r [14]. Próximo do máximo de $W(n_r)$ sua variação fracional quando n_r varia por uma unidade é muito pequena:

$$\frac{|W(n_r + 1) - W(n_r)|}{W(n_r)} \ll 1.$$

Desta forma, embora n_r só possa assumir valores inteiros, com boa aproximação pode-se tratar $W(n_r)$ como uma função contínua de n_r . A localização de $n_r = \tilde{n}_r$ do máximo de $W(n_r)$ é dada pela condição

$$\left. \frac{d[W(n_r)]}{dn_r} \right|_{n_r=\tilde{n}_r} = 0$$

ou equivalentemente

$$\left. \frac{d[\ln W(n_r)]}{dn_r} \right|_{n_r=\tilde{n}_r} = 0.$$

Por razões de conveniência podemos estudar o comportamento de $W(n_r)$ em torno de seu valor máximo fazendo a substituição: $n_r = \tilde{n}_r + \eta$ e expandindo a função $\ln[W(\tilde{n}_r + \eta)]$ numa série de Taylor:

$$\ln[W(n_r)] = \ln[W(\tilde{n}_r)] + B_1\eta + \frac{1}{2!}B_2\eta^2 + \frac{1}{3!}B_3\eta^3 + \dots, \quad (2.41)$$

onde

$$B_k = \left. \frac{d^k \left(\ln[W(\tilde{n}_r + \eta)] \right)}{d\tilde{n}_r^k} \right|_{\eta=0}. \quad (2.42)$$

A partir da eq. 2.20 pode-se analisar a expansão 2.41:

$$\ln [W(n_r)] = \ln n! - \ln n_r! - \ln (n - n_r)! + n_r \ln p + (n - n_r) \ln q.$$

Podemos usar os mesmos argumentos para $\ln n!$ para $n \gg 1$ como fizemos para a distribuição $W(n_r)$. Desta forma, temos que

$$\frac{d \ln n!}{dn} = \frac{\ln(n+1)! - \ln n!}{(n+1) - n} = \ln(n+1) \approx \ln n. \quad (2.43)$$

Consequentemente:

$$\frac{d [\ln W(n_r)]}{dn_r} = -\ln n_r + \ln (n - n_r) + \ln p - \ln q. \quad (2.44)$$

Como a expansão é em torno de \tilde{n}_r o termo B_1 é nulo:

$$\ln \left[\frac{n - \tilde{n}_r p}{\tilde{n}_r q} \right] = 0. \quad (2.45)$$

Utilizando o fato de que $p + q = 1$, obtem-se que $\tilde{n}_r = np$. Utilizando este resultado, o termo B_2 pode ser escrito como

$$B_2 = -\frac{1}{npq}. \quad (2.46)$$

Utilizando o mesmo procedimento é possível mostrar que $B_k < \eta^k / (npq)^{k-1}$.

Para valores grandes de $|B_2|$ pode-se ignorar termos de ordem superior a η^2 na eq 2.41 [14], o que nos gera:

$$W(n_r) = \frac{1}{(2\pi npq)^{1/2}} \exp \left[-\frac{(n_r - np)^2}{2npq} \right], \quad (2.47)$$

onde foi utilizada a condição de normalização: $W(\tilde{n}_r) \int_{-\infty}^{\infty} d\eta \exp \left[-\frac{1}{2}|B_2|\eta^2 \right] = 1$.

A PDF (discreta) para a caminhada aleatória é obtida repassando n_r por x na eq. 2.47

$$P_X(x) = \frac{1}{(2\pi npq)^{1/2}} \exp \left[-\frac{(x - np)^2}{2npq} \right]. \quad (2.48)$$

Deve-se ter em mente que a eq. 2.48 só é aplicável nos casos de x e n grandes.

Isto mostra mais uma vez o motivo das distribuições Gaussianas serem muito frequentes na natureza, uma vez que para grandes números, distribuições que satisfazem o TLC têm a propriedade de convergir para uma PDF Gaussiana.

2.5 TLC Generalizado e α -estabilidade de Lévy

Vimos na seção anterior que o Teorema do limite central afirma basicamente que a PDF referente à soma de n variáveis aleatórias *i.i.d.*, com média e desvio padrão finitos, converge para uma Gaussiana no limite que n tende para infinito, tendo como uma consequência a estabilidade da distribuição Gaussiana. Lévy e outros generalizaram este teorema. O TLC Generalizado afirma que a soma de n variáveis *i.i.d.* cujas distribuições possuem variância infinita (distribuições do tipo cauda grossa) tenderá para uma distribuição de Lévy à medida que n tende a infinito.

Não faremos aqui a demonstração rigorosa deste teorema, no entanto ele pode ser entendido a partir da conexão para o caso da estabilidade Gaussiana mostrada abaixo:

$$\begin{array}{ccc} \text{Gaussiana} & & \text{Lévy} \\ Y_n = \frac{1}{n} \sum_{i=1}^n (x_i - \langle X \rangle) & \rightarrow & Y_n = \frac{1}{a_n} \sum_{i=1}^n (x_i - \langle X \rangle) \end{array} \quad (2.49)$$

$$f_Z(k, n) \equiv f_Z\left(\frac{k}{\sqrt{n}}\right) \rightarrow f_Z(k, n) \equiv f_Z\left(\frac{k}{a_n}\right) \quad (2.50)$$

$$f_{Y_n}\left(\frac{k}{\sqrt{n}}\right) \equiv \left[f_Z\left(\frac{k}{\sqrt{n}}\right)\right]^n \rightarrow f_{Y_n}\left(\frac{k}{a_n}\right) \equiv \left[f_Z\left(\frac{k}{a_n}\right)\right]^n \quad (2.51)$$

$$a_n = n^{1/2} \rightarrow a_n \sim n^{1/\alpha} \quad (2.52)$$

$$f_{Y_n}\left(\frac{k}{a_n}\right) = \left[f_Z\left(\frac{k}{n^{1/\alpha}}\right)\right]^n \quad (2.53)$$

A questão que devemos responder é: “uma FC com a dependência funcional compatível com a eq. 2.53 permite estabilidade no limite que n tende a infinito?”

Vamos definir uma nova função $\psi(k)$ da forma:

$$\psi(k) \equiv \log f_{Y_n}\left(\frac{k}{a_n}\right), \quad (2.54)$$

logo

$$\psi(k) = n \psi(k/n^{1/\alpha}). \quad (2.55)$$

Por substituição direta, tem-se que uma possível solução para $\psi(k)$ é da forma:

$$\psi(k) \sim |k|^\alpha \quad (2.56)$$

o que gera:

$$f_Z(k) = \exp[-\gamma |k|^\alpha]. \quad (2.57)$$

Como resposta ao nosso questionamento, a eq. 2.57 representa um FC estável para a soma de n variáveis. Uma generalização da α -estabilidade é obtida reescrevendo a eq. 2.57 na forma (eq. 2.23):

$$f_X(k; \alpha, \beta, \gamma, \mu) = \exp[-ik\mu - \gamma|k|^\alpha(1 - i\beta\text{sgn}(k)\tan(\pi\alpha/2))],$$

e a PDF é obtida por

$$P_X(x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} dk \exp[-ikx] f_X(k; \alpha, \beta, \gamma, \mu) \quad (2.58)$$

onde μ representa a média, β é responsável pela assimetria, γ pela escala de $P_X(x)$ e $\alpha \in]0, 2]$.

A convolução de duas PDF's α -estáveis de Lévy é outra PDF α -estável de Lévy. Este é o caso mais geral de estabilidade para variáveis *i.i.d.* .

2.6 Estacionaridade e Tempo de Correlação

Dado um conjunto $\{x(t)\}$ de realizações de uma certa variável aleatória X , define-se a função de correlação para o conjunto $\{x(t)\}$ como:

$$\langle x(t_1)x(t_2) \rangle = C(t_1, t_2) . \quad (2.59)$$

Num sentido não muito restritivo, o conjunto $\{x(t)\}$ pode ser dito estacionário [5] se ele satisfaz as seguintes condições:

- (1) A média $\langle x(t) \rangle \equiv \mu$ é independente do tempo .
- (2) A função de correlação é da forma: $C(t_1, t_2) \equiv C(\tau)$ com $\tau = t_2 - t_1$.

(3) $\langle x^2(t) \rangle \equiv C(0)$. Logo, a variância do processo $C(0) - \mu^2$ é independente do tempo.

Por conveniência, mas sem perda de generalidade, consideremos um processo estacionário com média nula ($\mu = 0$) e variância unitária ($\sigma^2 = 1$). Podemos agora retomar algumas questões levantadas no Cap. 1, com respeito à escala típica de um processo (tempo de memória).

Para um processo estacionário, a informação a respeito do tempo de memória pode ser obtida por meio da integral da função de correlação $C(\tau)$. A área abaixo de $C(\tau)$ pode apresentar três valores:

$$\int_0^{\infty} d\tau C(\tau) = \begin{cases} \text{finito} \\ \text{infinito} \\ \text{indeterminado.} \end{cases} \quad (2.60)$$

A $\int_0^{\infty} d\tau C(\tau)$ finita implica a existência de um tempo de memória τ_0 chamado tempo de correlação do processo.

Como mencionado no Cap. 1, o decaimento exponencial da função de correlação refere-se a um exemplo típico de um tempo de escala bem definido:

$$\int_0^{\infty} d\tau e^{-\tau/\tau_0} = \tau_0. \quad (2.61)$$

Já uma dependência do tipo lei de potência implica a ausência de uma escala bem definida:

$$\int_0^{\infty} d\tau \tau^{\eta-1} = \infty, \quad (2.62)$$

se $0 < \eta \leq 1$.

Mediante a discussão anterior é conveniente analisar o comportamento da variância relacionada à variável S_n definida pela soma de n variáveis aleatórias X_i , num processo estacionário. Similar à eq. 2.13, $\sigma_{S_n}^2$ é escrito como:

$$\sigma_{S_n}^2 = \frac{1}{N} \sum_{i=1}^N \left[s_n^i - \langle S_n \rangle \right]^2, \quad (2.63)$$

logo:

$$\begin{aligned} \sigma_{S_n}^2 &= \frac{1}{N} \sum_{i=1}^N \left[\sum_{j=1}^n x_j^i - \left\langle \sum_{j'=1}^n X_{j'}^i \right\rangle \right]^2, \\ &= \frac{1}{N} \sum_{i=1}^N \left[\sum_{j=1}^n x_j^i - \left\langle \sum_{j'=1}^n X_{j'}^i \right\rangle \right] \left[\sum_{k=1}^n x_k^i - \left\langle \sum_{k'=1}^n X_{k'}^i \right\rangle \right], \\ &= \frac{1}{N} \sum_{i=1}^N \left[\sum_{j,k=1}^n x_j^i x_k^i - \sum_{j,k'=1}^n x_j^i \langle X_{k'}^i \rangle - \sum_{k,j'=1}^n x_k^i \langle X_{j'}^i \rangle + \sum_{k',j'=1}^n \langle X_{k'}^i \rangle \langle X_{j'}^i \rangle \right] \\ &= \frac{1}{N} \left[\sum_{i,j,k=1}^{N,n} x_j^i x_k^i - \sum_{i,j,k'=1}^{N,n} x_j^i \langle X_{k'}^i \rangle - \sum_{i,k,j'=1}^{N,n} x_k^i \langle X_{j'}^i \rangle + N \sum_{k',j'=1}^n \langle X_{k'}^i \rangle \langle X_{j'}^i \rangle \right] \\ &= \left[\sum_{j,k=1}^n \langle X_j X_k \rangle - \sum_{j,k'=1}^n \langle X_j \rangle \langle X_{k'} \rangle - \sum_{k,j'=1}^n \langle X_k \rangle \langle X_{j'} \rangle + \sum_{k',j'=1}^n \langle X_{k'} \rangle \langle X_{j'} \rangle \right] \\ &= \left[\sum_{j,k=1}^n \langle X_j X_k \rangle - \sum_{j,k'=1}^n \langle X_j \rangle \langle X_{k'} \rangle - \sum_{k,j'=1}^n \langle X_k \rangle \langle X_{j'} \rangle + \sum_{k',j'=1}^n \langle X_{k'} \rangle \langle X_{j'} \rangle \right] \\ &= \sum_{i=1}^n \left(\langle X_i^2 \rangle - \langle X_i \rangle^2 \right) + \sum_{\substack{i=1 \\ j \neq i}}^n \left(\langle X_i X_j \rangle - \langle X_i \rangle \langle X_j \rangle \right) \\ &= \sum_{i=1}^n \sigma_{X_i}^2 + \sum_{\substack{i=1 \\ j \neq i}}^n \left(\langle X_i X_j \rangle - \langle X_i \rangle \langle X_j \rangle \right). \end{aligned} \quad (2.64)$$

Como o processo é estacionário, a eq. 2.64 pode ser reescrita como:

$$\sigma_{S_n}^2 = n\sigma_{X_i}^2 + 2 \sum_{\substack{i=1 \\ j < i}}^n Cov(X_i, X_j), \quad (2.65)$$

onde $Cov(X_i, X_j)$ é a covariância, a qual possui um comportamento idêntico àquele da correlação para processos com média nula.

Se a soma na eq. 2.65 permanece finita para grandes valores de n , então as variáveis X_i ($i = 1, 2, \dots, n$) são ditas fracamente correlacionadas. Entretanto, se a soma diverge, então as variáveis possuem correlações de longo alcance. Esse comportamento é observado em processos estocásticos caracterizados por uma função de correlação do tipo lei de potência ($C(\tau) \sim \tau^{\eta-1}$). Desta forma, estudando o comportamento da variância de S_n , obtemos informações sobre o tipo de correlação num processo.

Se as variáveis são *i.i.d.*, os termos no somatório da eq. 2.65 se cancelam ($\langle X_i X_j \rangle = \langle X_i \rangle \langle X_j \rangle$), restando:

$$\sigma_{S_n} = n^H \sigma_X, \quad (2.66)$$

onde H é o expoente de Hurst, o qual, para o caso de variáveis do tipo *i.i.d.* possui o valor $1/2$. O expoente H será abordado com maiores detalhes no Cap. 3.

Estas mesmas propriedades podem ser vistas no domínio da frequência. Para isso basta expressar a função de correlação da variável aleatória X como uma integral de Fourier:

$$C(\tau) = \int_{-\infty}^{\infty} d\omega J(\omega) \exp[i\omega\tau], \quad (2.67)$$

onde o coeficiente $J(\omega)$ é chamado de *espectro de potência ou densidade espectral* de X . Da eq. 2.67 segue que $J(\omega)$ pode ser expressado em termos da função de correlação $C(\tau)$:

$$J(\omega) = \frac{1}{2\pi} \int_{-\infty}^{\infty} d\tau C(\tau) \exp[-i\omega\tau]. \quad (2.68)$$

No caso da função de correlação não possuir uma escala bem definida:

$$C(\tau) \sim \tau^{\eta-1} \quad (2.69)$$

com $0 < \eta \leq 1$, o espectro de potência terá a dependência da forma:

$$J(f) = \frac{c}{|f|^\eta}. \quad (2.70)$$

Pode ser mostrado também que $J(\omega) \propto |\tilde{x}(\omega)|^2$.

A relação entre o expoente η e o expoente de Hurst é da forma: $\eta = 2H + 1$. Alguns valores de η recebem nomes especiais: $\eta = 0$ é referente a um ruído branco e representa um sinal totalmente descorrelacionado; $\eta = 2$ é o caso do MB (também recebe o nome de processo Wiener). Já o caso intermediário $\eta = 1$ recebe o nome de ruído $1/f$ e é observado em uma ampla variedade de fenômenos (veja ref. em [5]).

As eqs. 2.67 e 2.68 são conhecidas como relações de *Wiener-Khintchine* e, como vimos, são de grande importância no estudo de correlações em séries temporais. Não usaremos aqui o espectro de potência e outros métodos espectrográficos porque eles funcionam corretamente apenas para séries estacionárias.

Capítulo 3

Multifractalidade versus monofractalidade

Qual é o comprimento do litoral brasileiro? A resposta depende da escala de resolução utilizada na medida e está relacionada com o conceito de fractalidade que será abordado neste capítulo.

Muitos fenômenos e estruturas encontradas na natureza necessitam de um formalismo matemático especial para quantificá-los. Alguns padrões tais como o crescimento e a disposição de galhos e folhas em uma árvore, a formação de nuvens, podem ser recriados a partir de regras simples de construção geométrica, que quando executadas são capazes de criar estruturas de alta complexidade. Algumas dessas estruturas têm propriedades conhecidas como *fractais* [1, 18] e para defini-los são necessários dois conceitos: (i) *dimensão fractal* — característica marcante de uma estrutura do tipo fractal devido ao fato de frequentemente possuir valores fracionários, o que foge dos conceitos da dimensão topológica associada à geometria euclidiana e, (ii) *auto-similaridade* — invariância por mudança de escala, ela é fisi-

camente caracterizada pelo fato de poder observar a mesma estrutura independente da escala de observação.

Embora a idéia de objetos com dimensão fracionária date do início do século XIX, foi o francês Benoit Mandelbrot, considerado o pai da geometria fractal, quem a introduziu em 1975 [1]. A palavra fractal tem sua origem do latim “*fractus*”: fração, fragmentado, etc. Atualmente, os conceitos de fractalidade são utilizados amplamente em muitas áreas que se estendem desde a Economia, Física, Medicina, até à própria Arte.

3.1 Auto-similaridade e dimensão fractal

Para os objetos geométricos tradicionais existe uma relação simples entre a sua dimensão d , o número de “caixas” N (não sobrepostas) necessárias para recobrir todo o objeto e o comprimento das caixas. Esta é a idéia do método conhecido como método “Box-Counting” (contagem de caixas) [19]. Essa relação é da forma

$$N_n = \left[\frac{1}{L_n} \right]^d . \quad (3.1)$$

onde L_n é o comprimento da aresta na nova “escala”.

A critério de visualização, apliquemos o método de contagem de caixas ao caso bidimensional. A Fig. 3.1(a) mostra um quadrado de aresta unitária preenchendo completamente a parte central de um quadrado de aresta 3. As Figs. 3.1(b) e 3.1(c) são obtidas a partir da divisão sucessiva da primeira, ilustrando o processo empregado no método de contagem de caixas. Já a tabela 3.1 relaciona o número de células nas Figs. 3.1(a), 3.1(b) e 3.1(c) que contém parte do quadrado com a aresta de cada célula.

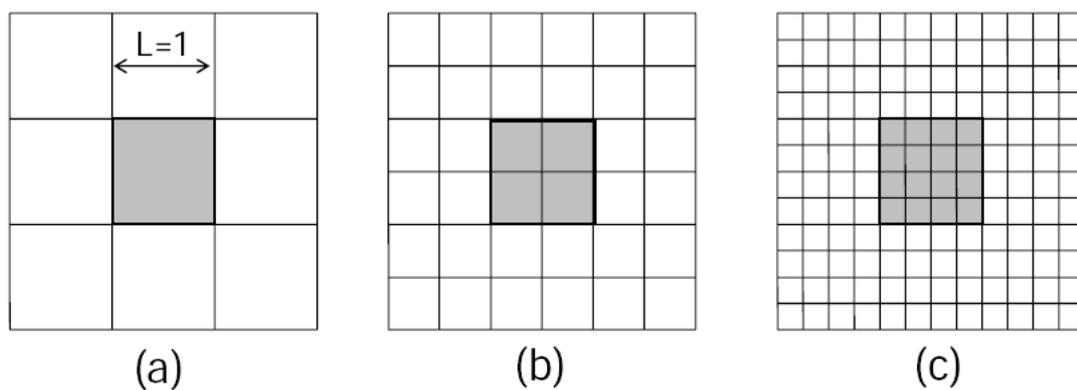


Figura 3.1: Ilustração do método de contagem de caixas para o caso 2D. O objeto é seqüencialmente dividido. A dimensão associada ao objeto pode ser obtida a partir da eq. 3.1, que relaciona d com a quantidade de caixas não sobrepostas necessárias para recobrir todo o objeto e o tamanho da aresta destas caixas.

Figura	Número de células	Comprimento das células
Fig. 3.1(a)	$N = 1$	$L = 1$
Fig. 3.1(b)	$N = 4$	$L = \frac{1}{2}$
Fig. 3.1(c)	$N = 16$	$L = \frac{1}{4}$

Tabela 3.1: Relação para 3 processos sucessivos de divisão de um objeto, empregado no método de contagens de caixas, entre o número de caixas N (contendo parte do objeto) e o comprimento L da aresta das caixas.

A partir da eq. 3.1 associamos o valor 2 para a dimensão do objeto na Fig. (3.1). Os valores de 0, 1 ou 3 seriam obtidos se no lugar do quadrado houvesse um ponto, uma reta ou um cubo, respectivamente.

Consideremos um segundo objeto mostrado na Fig. 3.2. A curva triádica de Koch, como esse objeto é nomeado, que pode ser obtida de um segmento de linha com tamanho unitário a partir do seguinte procedimento:

- 1 - Divida o segmento unitário (Fig. 3.2(a)) em três novos segmentos cujos comprimentos são $\delta = \frac{1}{3}$, retire a parte central e acrescente um quarto segmento formando a parte superior de um triângulo equilátero (Fig. 3.2(b)) — o resultado é a primeira geração: a curva de Koch.
- 2 - Repita o procedimento anterior para cada segmento da primeira geração. Como resultado, serão obtidos 16 novos segmentos de tamanho: $\delta = \left(\frac{1}{3}\right)^2$ (Fig. 3.2(c)) — o resultado é a segunda geração.

A cada repetição desse processo, uma nova geração da curva de Koch é obtida. É fácil perceber que na n -ésima geração (Fig. 3.2(e)) a curva possuirá 4^n segmentos cujo tamanho será dado por $\delta = \left(\frac{1}{3}\right)^n$.

Qual é a dimensão da curva triádica de Koch? A Fig. 3.3 [19] mostra o processo de divisão da curva triádica utilizado no método de contagens de caixas.

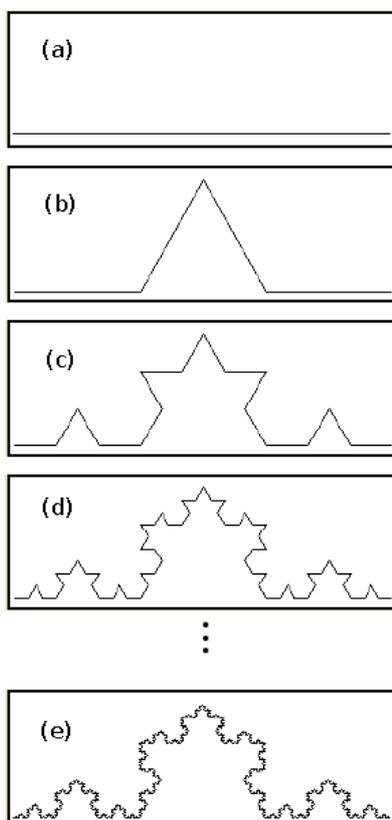


Figura 3.2: Curva triádica Koch. Cada parte da figura representa uma geração diferente. (a) Geração de ordem zero, (b) primeira geração, (c) segunda geração, (d) quarta geração e (e) geração de n -ésima ordem.

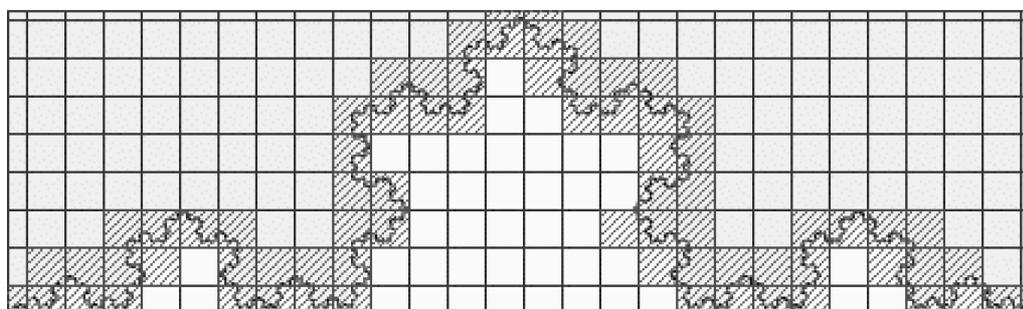


Figura 3.3: Ilustração do processo de divisão de um objeto empregado pelo método de contagens de caixas para a curva triádica de Koch.

Pode-se calcular a dimensão da curva de Koch tomando o comprimento de cada caixa como o tamanho mínimo de cada parte da mesma: $L_n = (\frac{1}{3})^n$, o que corresponde a $N_n = 4^n$ caixas. Utilizando a eq. 3.1 obtém-se $d = \frac{\log 4}{\log 3} \approx 1.26$.

Uma interpretação para a dimensão fracionária da curva triádica é que sua “rugosidade” faz com que ela ocupe mais espaço que uma simples reta (cuja dimensão é 1) e menos espaço que uma superfície (que possui dimensão 2). Desta forma, a dimensão fractal, definida pelo grau de “rugosidade” (irregularidade) de uma forma fractal, torna-se uma medida de comparação entre tais estruturas [1, 18].

A eq. 3.1 mostra que a relação entre o número de caixas utilizado no método de contagem e o comprimento de cada caixa obedece uma lei de potência. Isto permite a obtenção da dimensão d com o uso de uma escala logarítmica ($\log N_n = -d \log L_n$). Logo, a fractalidade pode ser entendida a partir dos conceitos de simetria de escala e dimensão fracionária.

Uma descoberta importante e surpreendente na última década foi que muitos sistemas físicos [20, 21], biológicos e até mesmo econômicos não possuem escalas características, isto é, eles têm propriedades fractais e mostram invariância por mudança de escala [3]. Enquanto certas classes de distribuições de probabilidade para uma dada variável aleatória X , por exemplo

$$P_X(x) \sim e^{-x/x_0}, \quad (3.2)$$

possui uma escala característica (neste caso, associada ao parâmetro x_0), outra classe de distribuições expressas por leis de potência, tais como

$$P_X(x) \sim \frac{1}{x^\mu}, \quad (3.3)$$

mostram invariância por mudança de escala (auto-similaridade):

$$P_X(\lambda x) \sim \lambda^{-\mu} P_X(x) . \quad (3.4)$$

A identificação de sistemas onde é observado um comportamento do tipo lei de potência tem crescido de forma surpreendente nos últimos anos. Como exemplos, citamos o comportamento de retornos no mercado financeiro [22, 23], os vôos de Lévy descritos na seção 2.3, o comportamento dos intervalos entre as batidas do coração [24] e até mesmo a estrutura do DNA que armazena toda a informação genética [25].

3.2 Fractalidade em Series Temporais

Na seção anterior definimos um objeto fractal como aquele que satisfaz dois critérios: auto-similaridade e dimensão fracionária. A auto-similaridade implica que um objeto é composto de sub-conjuntos (sub-unidades) em vários níveis de escala, os quais assemelham-se ao objeto como um todo. Esta ausência de escala apesar de ilimitada em termos matemáticos apresenta limites nos casos reais, tanto a menor quanto a maior escala em que a auto-similaridade pode ser observada. Como exemplo pode-se pensar em uma couve-flor, onde a menor escala de observação é definida pelo tamanho de sua menor “célula” macroscópica e a maior escala é aquela da própria verdura. O segundo critério para um objeto fractal é o mesmo possuir uma dimensão fractal fracionária. Para possuir propriedades fractais, não basta apenas satisfazer ao critério de auto-similaridade, o conceito de dimensão fracionária também deverá ser satisfeito.

A ausência de um comprimento de escala característico não se limita somente a objetos geométricos [1] (veja também ref. [3]), podendo ser aplicado a séries temporais complexas. O problema de se “detectar” e quantificar ausência de escala em séries temporais é que, embora uma série temporal seja representada em um espaço bidimensional, os eixos da abscissa e ordenada possuem unidades físicas diferentes. Como um exemplo pode-se pensar numa série temporal econômica para o preço ou retorno (definido na seção 1.3) relacionados a um dado ativo financeiro. Enquanto o eixo das ordenadas está em unidades monetárias (no caso de representar os preços) ou adimensionais (no caso dos retornos financeiros), o eixo das abscissas possui unidades de tempo (intervalo entre as transações financeiras: segundos, minutos, dias, etc.). Este não é o caso para objetos geométricos onde ambos os eixos possuem dimensão de distância. Enquanto que o estudo de auto-similaridade para o caso de objetos geométricos pode ser feito apenas por reescalar uma sub-unidade do objeto para o tamanho do objeto original (por multiplicar ambos eixos por um mesmo fator) e, depois disso comparar suas propriedades com aquelas do objeto original, para o caso de uma série temporal o estudo não é tão simples pois são necessários dois fatores de magnificação, uma vez que cada eixo representa uma unidade física diferente.

Em termos matemáticos tem-se que uma dada série temporal é auto-similar se:

$$y(t) \stackrel{d}{\equiv} a^\alpha y\left(\frac{t}{\alpha}\right) \quad (3.5)$$

onde $\stackrel{d}{\equiv}$ implica que as propriedades estatísticas de ambos os lados da equação são similares ou idênticas. Portanto, uma série temporal será auto-similar se, quando o eixo das ordenadas for reescalado na forma: $y \rightarrow a^\alpha y$, e o eixo temporal for

reescalado por um fator a ($t \rightarrow \frac{t}{a}$), a série resultante possui as mesmas propriedades estatísticas que a original. O expoente α é chamado de parâmetro de auto-similaridade.

Vimos no Cap. 2 que todas as propriedades estatísticas de um dado conjunto de dados pode ser dado por correlações e sua PDF, sendo, esta última, descrita por todos momentos estatísticos, o que torna impraticável o uso rigoroso da eq. 3.5. O que se é feito na maioria dos casos é considerar uma auto-similaridade em um nível mais fraco. Isso é feito por considerar apenas os dois primeiros momentos da distribuição estatística da série e de seus sub-conjuntos.

Portanto, o fator de magnificação M_x do eixo temporal pode escrito na forma:

$$M_x = \frac{\Delta t}{\tilde{\Delta t}} \quad (3.6)$$

onde Δt é a escala de tempo da série original e $\tilde{\Delta t}$ é a escala temporal do sub-conjunto. Já o fator de magnificação M_y é definido como:

$$M_y = \frac{S}{\tilde{S}} \quad (3.7)$$

onde S é o desvio padrão relacionado à série original e, \tilde{S} é aquele relacionado ao dado sub-conjunto da série.

O parâmetro de auto-similaridade α pode ser escrito como:

$$\alpha = \frac{M_y}{M_x} . \quad (3.8)$$

O parâmetro α é obtido por considerar sub-conjuntos de diferentes tamanhos (para maiores detalhes veja ref. [26]).

A Fig. 3.4(a) mostra a auto-similaridade na série de retornos financeiros, obtida gratuitamente na internet, referente à taxa de câmbio da *Libra Esterlina*

(GBP) frente ao *Dólar Americano* (USD) para um período de aproximadamente 8 mil dias de negociação. Pode-se perceber claramente a ausência de escala, onde os histogramas referentes aos retornos agregados para os períodos de $\tau = 2, 3, 5, 10, 20, 40, 60, 120$ e 240 dias praticamente se sobrepõem (após a reescala dos eixos) àquele de retornos diários ($\tau = 1 d$). A Fig. 3.4(b) mostra o gráfico $P(\tau, 0)$ versus τ , onde $P(\tau, 0)$ (símbolos) é referente a quantidade de pontos na distribuição de frequência (histograma) dos retornos referentes ao retorno financeiro nulo e, τ é o período de tempo (em dias) analisado.

O parâmetro de auto-similaridade foi obtido a partir da regressão por lei de potência [5, 27] da forma $P(\tau, 0) = P(1, 0)\tau^B$ (curva contínua) para os dados da Fig. 3.4(b), onde $\alpha = -B^{-1}$. Se os retornos na série temporal analisada se comportassem como um random walk, α teria o valor de 1/2.

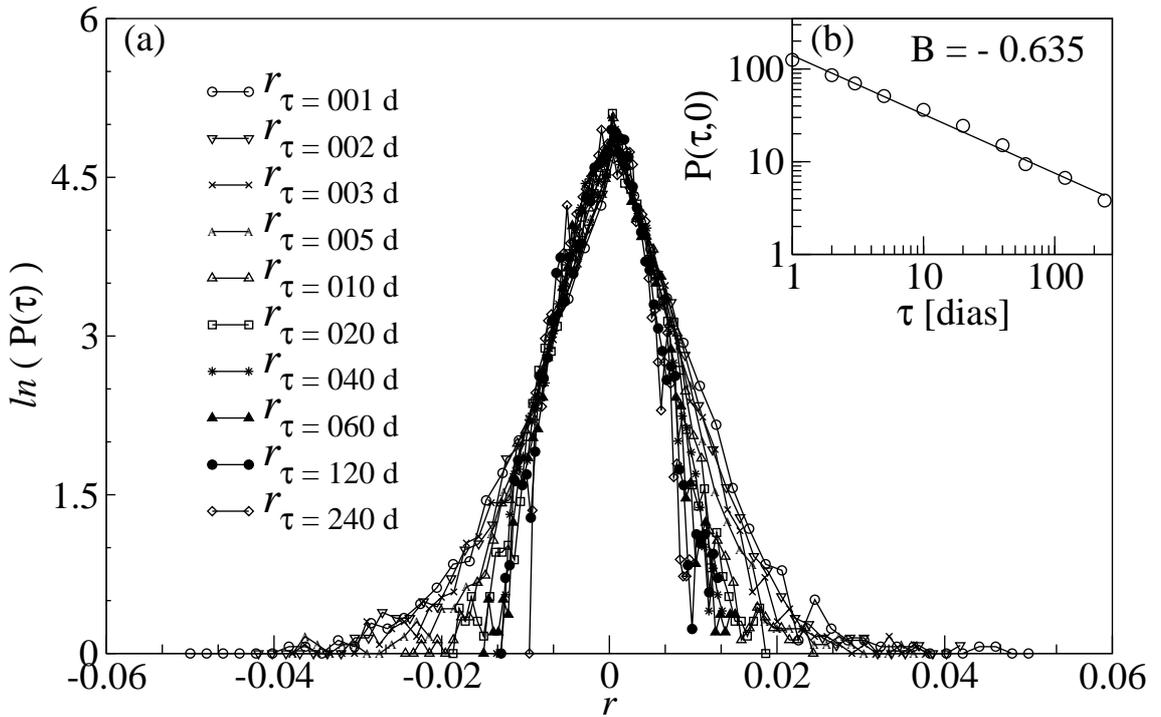


Figura 3.4: Auto-similaridade para série retornos financeiros referente à taxa de câmbio da *Libra Esterlina* (GBP) frente ao *Dólar Americano* (USD) para um período de aproximadamente 8 mil dias de negociação. (a) Sobreposição dos histogramas para os períodos de $\tau = 1, 2, 3, 5, 10, 20, 40, 60, 120$ e 240 dias. (b) A curva contínua é referente ao ajuste do tipo lei de potência da parte central do histograma e, os círculos são os valores experimentais de $P(\tau,0)$ para cada valor do período de tempo τ analisado. O coeficiente de auto-similaridade é dado pelo negativo do inverso do expoente da lei de potência. Encontramos aqui um comportamento que foge do previsto pelo TLC para variáveis *i.i.d.* .

3.3 Monofractais versus Multifractais

Em muitos casos é necessário mais de um expoente para se caracterizar completamente um objeto como aquele na eq. 3.1. Em termos geométricos pode-se pensar que dimensões diferentes são atribuídas a diferentes subconjuntos fractais de um objeto. Já no caso de uma série temporal, momentos de ordem superiores se tornam importantes para a descrição de sua PDF. Se este for o caso, o objeto (ou a série) é dito multifractal sendo simplesmente fractal ou monofractal em caso contrário [1, 18]. Considere o momento de q -ésima ordem x^q , relacionado a um conjunto de realizações $\{x(t)\}$ de uma dada variável aleatória X , da seguinte forma:

$$\langle x^q \rangle \sim t^{qH(q)} . \quad (3.9)$$

No caso de H não depender de q (ser uma função constante) a série é dita monofractal e momentos de ordem superiores podem ser escritos em termos do momento de segunda ordem; em caso contrário, onde $H(q)$ possui uma dependência não linear em q , a série é dita multifractal. Neste caso, os dois primeiros momentos se tornam insuficientes para uma descrição satisfatória da PDF.

Métodos Fractais Aplicados a Séries Financeiras

É sabido da literatura que correlações temporais influenciam as propriedades fractais de séries, o que faz com que tais propriedades se tornem importantes no estudo e caracterização de tais séries. Esta parte da tese é dedicada à descrição de alguns formalismos utilizados na caracterização fractal de séries temporais.

Expoente de Hurst e análise R/S

A análise da estatística R/S tem sua origem no trabalho do Engenheiro Hidrologista britânico Harold Edwin Hurst (1880-1978) em 1951 [28] ao investigar por vários anos o nível do rio Nilo. Logo, a primeira aplicação da análise da estatística R/S foi relacionada a problemas de dimensionamento de represas. De posse das vazões de água para vários períodos de tempo, a idéia consistia em determinar quais eram os volumes máximo e mínimo no reservatório, pois se pretendia evitar tanto o transbordamento quanto a secagem do mesmo.

Hurst percebeu que a razão entre a diferença do valor máximo e mínimo da série integrada (R) dos dados e o desvio padrão (S) obedecia uma lei de potência em função do período de observação com o expoente H , atualmente conhecido com expoente de Hurst. Posteriormente Hurst verificou esta relação para vários outros fenômenos naturais. Este novo método estatístico, conhecido como análise R/S foi descrito detalhadamente em [5, 18, 28, 29].

Segue abaixo o procedimento empregado pela estatística R/S a uma série temporal $\{r_i\}$ contendo N termos.

Divida a série em n blocos, sendo $n\tau = N$, para cada bloco segue o seguinte procedimento [29]:

1. Calcule o retorno médio para o j -ésimo bloco:

$$\langle r \rangle_j = \frac{1}{n} \sum_{i=1}^{\tau} r_{j,i} \quad (3.10)$$

2. Calcule o desvio padrão dentro de cada bloco:

$$S_j = \left[\frac{1}{n} \sum_{i=1}^{\tau} (r_{j,i} - \langle r \rangle_j)^2 \right]^{\frac{1}{2}} \quad (3.11)$$

3. Calcule para o j -ésimo bloco o valor da quantidade $(R/S)_j$:

$$(R/S)_j = \frac{1}{S_j} \left[\max_{1 \leq t \leq \tau} \sum_{i=1}^t (r_{j,i} - \langle r \rangle_j) - \min_{1 \leq t \leq \tau} \sum_{i=1}^t (r_{j,i} - \langle r \rangle_j) \right] \quad (3.12)$$

4. Calcule a média para a quantidade $(R/S)_j$ para os n blocos (cada qual contendo τ termos):

$$\langle R/S \rangle = \frac{1}{n} \sum_{j=1}^n (R/S)_j \quad (3.13)$$

5. Repita o procedimento variando a quantidade τ de termos em cada bloco (variando a quantidade de blocos). A quantidade R/S poderá ser escrita como uma lei de potência da forma:

$$\langle R/S \rangle \sim \tau^H \quad (3.14)$$

Mandelbrot e Wallis verificaram que a descoberta empírica de Hurst podia ser aplicada ao estudo de correlações em séries temporais (veja ref. em [1]), onde pode-se interpretar o expoente H da seguinte forma: $0 < H < \frac{1}{2}$ representa uma antipersistência na série, $H = \frac{1}{2}$ implica numa série descorrelacionada do tipo “random walk” e $\frac{1}{2} < H < 1$ representa uma persistência na série.

O expoente de Hurst e a dimensão fractal da série estão relacionados [18] por

$$d = 2 - H . \quad (3.15)$$

Expoente de Hölder e o método *MF-DFA*

Com a finalidade de quantificar as propriedades multifractais de séries temporais, estudamos seu espectro multifractal. Ele contém informações sobre a correlação de n pontos e, desta forma, fornece mais informações que a função de correlação usual [30]. Segue abaixo o procedimento empregado na análise “multifractal detrended fluctuations analysis” (MF-DFA) [31, 32] — utilizado na obtenção do espectro multifractal para uma série temporal $\{r_t\}$ contendo N pontos:

1. Obtenha a série integrada (perfil):

$$y(t) \equiv \sum_{k=1}^t [r_k - \langle r \rangle]; \quad (3.16)$$

$$t = 1, \dots, N.$$

2. Divida o perfil $y(t)$ em segmentos não sobrepostos de tamanho s , tal que $N \equiv s N_s$;
3. Por meio de uma regressão do ν -ésimo segmento ($\nu = 1, \dots, N_s$), obtenha o melhor ajuste para a tendência local da série $y_\nu(i)$, $i = 1, \dots, s$.
4. Determine o segundo momento para o ν -ésimo segmento através de ⁴:

$$F_2(\nu, s) \equiv \frac{1}{s} \sum_{i=1}^s |y((\nu - 1)s + i) - y_\nu(i)|^2. \quad (3.17)$$

5. Calcule a função de flutuação de q -ésima ordem:

$$F_q(s) \equiv \frac{1}{N_s} \sum_{\nu=1}^{N_s} F_2(\nu, s)^{q/2}. \quad (3.18)$$

⁴ A análise DFA (abreviação de “Detrended Fluctuation Analysis” em inglês) [33] utilizada para estimar o expoente H segue o procedimento descrito nos itens 1 ao 4.

A eq. 3.18 seguirá a lei de escala

$$F_q(s) \sim s^{qh(q)} \quad , \quad (3.19)$$

onde $h(q)$ representa um expoente de Hurst generalizado. Enquanto valores positivos de q fazem com que grandes flutuações se sobressaiam, momentos negativos contribuem para pequenas flutuações. Uma série monofractal possui um “único” expoente de Hurst $h(q) = H$, já para aquelas multifractais o valor de $h(q)$ possui uma dependência não linear com q .

A relação entre $h(q)$ definido na eq. 3.19 no método MF-DFA com o expoente $\tau(q)$ usado convencionalmente para quantificar a função partição em livros textos [18, 23, 32, 34] é da forma: $\tau(q) = qh(q) - 1$.

O espectro multifractal $f(\alpha)$ pode ser obtido a partir de uma transformação de Legendre do expoente $\tau(q)$:

$$\alpha = d\tau(q)/dq \quad (3.20)$$

$$f(\alpha) = q\alpha - \tau(q) \quad . \quad (3.21)$$

Aqui, f é a dimensão fractal do subconjunto da série caracterizada pelo comprimento de singularidade α . Na literatura α é referido como expoente de Hölder ou expoente Lipschitz-Hölder. Dada uma função $S(x)$, define-se $\alpha(x_0)$, seu expoente de Hölder no ponto x_0 , como o maior expoente de forma a existir um polinômio $P_n(x)$ de grau n , tal que a relação $|S(x) - P_n(x)| \leq C|x - x_0|^{\alpha(x_0)}$ seja satisfeita na vizinhança de x_0 (C é uma constante). Logo, a máxima derivada que a função S irá possuir é de ordem n . O polinômio $P_n(x)$ corresponde à série de Taylor de $S(x)$ no ponto x_0 expandida até a ordem n . Assim, $\alpha(x_0)$ mede a singularidade no ponto x_0 . Quanto maior $\alpha(x_0)$, mais regular (mais suave) é a função nesse ponto.

Chamamos a atenção para o fato que na análise MF-DFA o expoente $h(q)$ (e conseqüentemente o expoente α e o espectro $f(\alpha)$) é definido a partir da lei de escala dada pela eq. 3.19, que por sua vez é construída a partir das flutuações de segunda ordem na série temporal (veja eqs. 3.17 e 3.18). Contudo, muitas vezes a lei de escala é obtida a partir de quantidades que têm uma medida de probabilidade e, quando for este o caso, o espectro multifractal ($f(\alpha)$ vs α) possuirá a propriedade de ser tangenciado pela reta com coeficiente angular 1 [35]. Logo, esta propriedade não será satisfeita pelos espectros obtidos a partir da análise MF-DFA.

A Fig. 3.5(a) mostra uma série temporal contendo 1.10^3 de um total de 5.10^4 pontos, gerados artificialmente. O experimento consiste na simulação do lançamento de 10 moedas; o eixo y representa a variável aleatória $\{y_i\}$ correspondente à diferença entre o número de caras e coroas resultante de cada lançamento, já o eixo dos x representa cada evento (lançamento). A Fig. 3.5(b) mostra o expoente H estimado usando a análise DFA. Já na Fig. 3.5(c) temos o gráfico do expoente de Hurst generalizado $H(q)$ versus q e o espectro multifractal na Fig. 3.5(d).

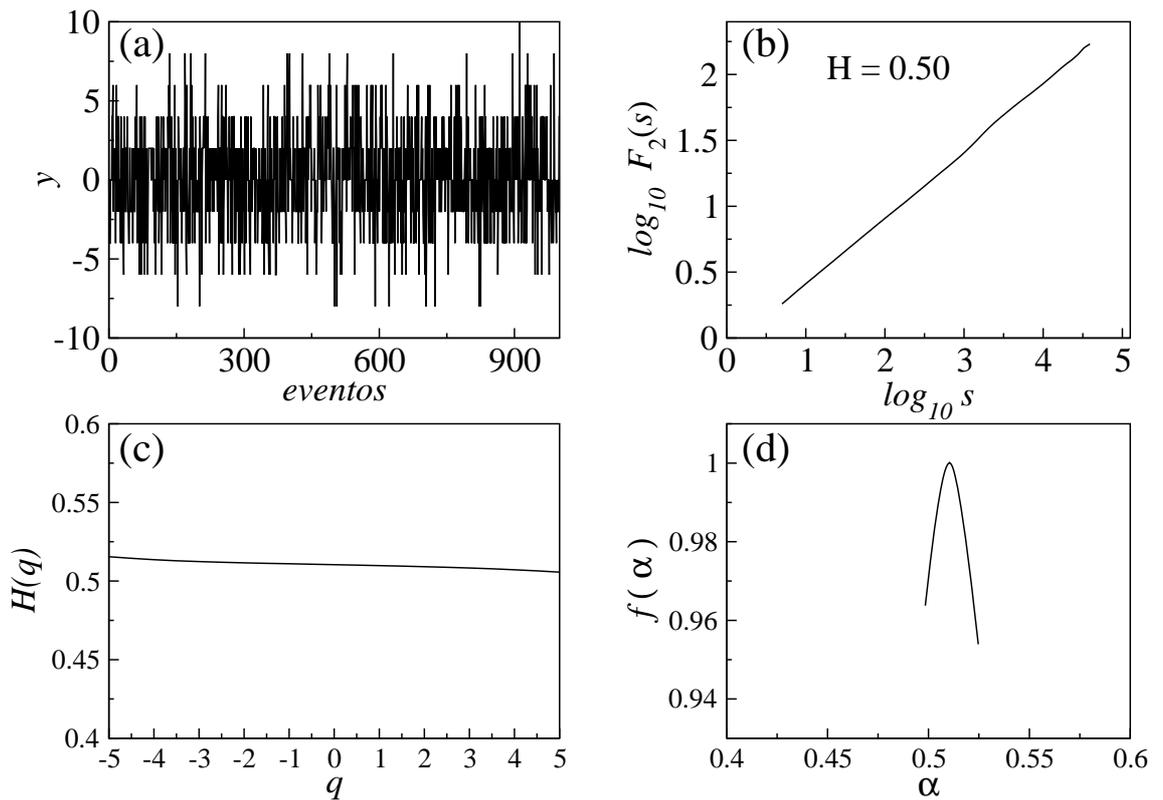


Figura 3.5: (a) Série contendo $1. \cdot 10^3$ de um total de $5. \cdot 10^4$ pontos, gerados artificialmente. (b) Expoente de Hurst estimado usando a análise DFA ($q = 2$). (c) Expoente de Hurst generalizado. (d) Espectro multifractal.

O valor do expoente $H \simeq 1/2$ (Fig. 3.5(b)), o fato de $H(q)$ não depender de q (Fig. 3.5(c)) e, tanto o fato do espectro multifractal $f(\alpha)$ estar centrado em $\alpha \simeq 1/2$ (Fig. 3.5(d)) quanto o fato dele ser bastante estreito, mostram a ausência de correlação entre os pontos e o caráter monofractal da série da Fig. 3.5(a).

Nesta tese, tanto a largura do espectro multifractal quanto o valor do expoente de Hölder em torno do qual o espectro está centralizado serão utilizados como meio de quantificar o grau de fractalidade das séries analisadas. Quanto mais largo o espectro mais o caráter multifractal da série é pronunciado. A presença de anti-correlação será indicada quando o máximo de $f_{\text{máx}}(\alpha)$ ocorrer para um valor de $\alpha < 1/2$. A ausência de correlação — série do tipo “random walk” — se mostra quando o máximo ocorrer em torno de $\alpha = 1/2$. Já a presença de correlações positivas se mostra quando o máximo ocorrer para $\alpha > 1/2$.

É importante ressaltar o fato que $f_{\text{máx}}(\alpha) = 1$ uma vez que nesta tese só estudamos propriedades de séries temporais sem suporte fractal [18] em uma dimensão.

Capítulo 4

Seções de Lévy

Um problema importante em física diz respeito a origem de raros eventos na dinâmica de sistemas complexos. Eventos como terremotos, epidemias, extinção de espécies e quebra de bolsas de valores podem ter sua origem devido a caudas grossas nas distribuições da específica variável aleatória ou até mesmo num comportamento não Gaussiano destas distribuições. Muitos trabalhos são voltados para o estudo da dinâmica das caudas grossas da variável aleatória [22, 36], outros estão voltados simplesmente para o estudo da convergência da variável aleatória para o regime Gaussiano [37, 38]. Neste capítulo estudaremos uma “generalização” para o teorema do limite central clássico para o caso de variáveis aleatórias contínuas e sua modificação para o caso de séries temporais [37, 38, 39].

4.1 Seções de Lévy e Generalização do TLC

Considere uma conjunto de variáveis estocásticas X_n com $n \in \mathbb{N}$. A probabilidade condicional referente à realização x_{n+1} , dado o conjunto de realizações

x_1, \dots, x_n , pode ser escrita como $P(x_{n+1}|x_1, \dots, x_n)$. Desta forma, a média e a variância condicional relacionadas a X_{n+1} podem ser escritas como:

$$\mu_n = \langle X_{n+1} \rangle_{x_1, \dots, x_n} \quad (4.1)$$

$$\mu_n = \int x_{n+1} P(x_{n+1}|x_1, \dots, x_n) dx_{n+1} \quad , \quad (4.2)$$

e

$$m_n^2 = \langle X_{n+1}^2 \rangle_{x_1, \dots, x_n} - \langle X_{n+1} \rangle_{x_1, \dots, x_n}^2 \quad (4.3)$$

$$m_n^2 = \int x_{n+1}^2 P(x_{n+1}|x_1, \dots, x_n) dx_{n+1} - \mu_n^2 \quad . \quad (4.4)$$

Ambos μ_n e m_n dependem tanto da seqüência das realizações x_1, \dots, x_n quanto de seus valores específicos. Vamos definir a quantidade λ_n

$$\lambda_n = \sum_{i=1}^n m_i^2 \quad (4.5)$$

onde m_i é a variância condicional definida na eq. (4.4). Considere um número real positivo t tal que a condição

$$\lambda_{n-1} \leq t < \lambda_n \quad (4.6)$$

seja satisfeita. Dizemos que o conjunto das realizações x_1, \dots, x_n pertence à seção t e a condição definida pela eq. 4.6 é dita *condição da seção t* . Podemos definir uma série referente a realizações de uma variável S_t cuja i -ésima “realização”, s_t^i , é dada pela soma das realizações de X_n pertencentes à seção t :

$$s_t^i = \sum_{j=1}^n x_j^i \quad (4.7)$$

O teorema da *seção de Lévy* [39] afirma o seguinte:

Teorema 1 *Para a média condicional $\mu_n = 0$ ($\forall n \in \mathbb{N}$) e variáveis estocásticas X_n ($\forall n \in \mathbb{N}$) que satisfaçam a condição de Lindeberg, a distribuição de probabilidade de S_t/\sqrt{t} é tal que*

$$\lim_{t \rightarrow \infty} P\left(\frac{S_t}{\sqrt{t}} < \eta\right) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\eta} dx e^{-\frac{x^2}{2}}.$$

Significado Físico das Seções de Lévy

Desta forma, o teorema 1 pode ser visto como uma extensão do *teorema do limite central* clássico. Aqui, a condição de que as variáveis sejam *i.i.d.* não é exigida. A distribuição de probabilidade da variável S_t/\sqrt{t} converge para uma Gaussiana de média nula e desvio padrão unitário à medida que t tende para infinito. Dada a variância de S_t , digamos $M_t = \sqrt{\langle S_t^2 \rangle - \langle S_t \rangle^2}$, a distribuição de probabilidade para a variável normalizada S_t/M_t também convergirá para uma Gaussiana de média nula e desvio padrão unitário. Enquanto que o desvio padrão da variável S_t/M_t permanece constante e igual a unidade durante o processo de convergência, aquele referente à variável S_t/\sqrt{t} convergirá gradualmente para a unidade.

Tempo de Volatilidade e Seções de Lévy

Dada a probabilidade condicional referente à variável aleatória X_n , sua distribuição de probabilidade é dada pela probabilidade marginal definida como

$$P_n(x_n) = \sum_{x_1 \dots x_n} P(x_n | x_1, \dots, x_{n-1}) \quad (4.8)$$

onde a soma considera todas as possíveis realizações das variáveis X_i ($i = 1, 2, \dots, n$).

Já a variância marginal pode ser calculada através de

$$\nu_n^2 = \langle X_n^2 \rangle - \langle X_n \rangle^2 = \int dx_n x_n^2 P_n(x_n) - \left(\int dx_n x_n P_n(x_n) \right)^2 . \quad (4.9)$$

Convém neste ponto definir a quantidade

$$\sigma_n^2 = \frac{1}{n} \sum_{i=1}^n \nu_i^2 \quad (4.10)$$

que chamaremos de variância média acumulada de X_n . Lembrando que a variância relacionada à variável aleatória S_n definida como a soma usual de n variáveis aleatórias X_i é da forma

$$M_{S_n}^2 = \sum_{i=1}^n \nu_i^2 + \sum_{\substack{i=1 \\ j \neq i}}^n \nu_i \nu_j (Cov(X_i, X_j)) \quad (4.11)$$

onde $Cov(X_i, X_j)$ é a covariância entre as variáveis X_i e X_j , podemos definir, a partir das eqs. 4.10 e 4.11, a seguinte unidade de tempo — que denominaremos *tempo de volatilidade*:

$$\tau_n = \frac{M_{S_n}^2}{\sigma_n^2} . \quad (4.12)$$

Para um melhor entendimento do *tempo de volatilidade* definido na eq. 4.12 considere o caso em que as variáveis X_i sejam descorrelacionadas, ou seja, $Cov(X_i, X_j) = 0$ para todo $i \neq j$. Neste caso o *tempo de volatilidade* torna-se o tempo “usual”, dado por: $\tau_n = n$. Outro exemplo é aquele em que a variância marginal é estacionária ($\nu_i = \nu_j = \nu$), para esta situação a eq. 4.12 torna-se

$$\tau_n = n + Cov(X_i, X_j) , \quad (4.13)$$

onde fica claro que a presença da correlação linear promove um atraso (no caso de correlações positivas) ou um adiantamento (no caso de correlações negativas) na medida do *tempo de volatilidade* comparada com a medida de tempo usual.

Enquanto que para alguns casos a variância relacionada a S_n segue uma lei de potência [1] da forma: $M_{S_n} \sim n^H$, onde H é o expoente de Hurst, o *tempo de volatilidade* escala como $\tau_n \sim n^{2H}$. O adiantamento ocorrerá no caso em que $H > \frac{1}{2}$ e o atraso no caso em que $H < \frac{1}{2}$.

Podemos agora estender o conceito do *tempo de volatilidade* para a série das realizações de S_t relacionada à seção t que, para o caso em que a volatilidade referente as realizações das variáveis X_n é estacionária é:

$$\tau_t = \frac{M_{S_t}^2}{\nu^2}. \quad (4.14)$$

4.2 Seções de Lévy em Séries Temporais

Dada uma série temporal $\{x_i\}$ onde o índice $i = 1, \dots, N$ é referente à realização da variável aleatória X_i , para se calcular a seção t primeiramente temos que considerar o cálculo da variância condicional (eq. 4.4), o que se torna um problema devido ao fato de não conhecermos a probabilidade condicional $P(x_{n+1}|x_1, \dots, x_n)$. Este problema pode ser contornado pelo cálculo de uma volatilidade local [37, 38] construída da seguinte maneira: considere um número qualquer $q \in \mathbb{Z}$, defina uma nova série $\{y_k\}$ a partir da série $\{x_i\}$ onde $y_k = x_{k+q}$ ($k = 1, \dots, N - 2q$), de modo que essa primeira série seja idêntica a segunda com a exceção do fato que ela possuirá $2q$ termos a menos (os q primeiros termos e os q últimos termos serão removidos). A variância condicional m_n^2 , que de agora em diante chamaremos de *variância local*, será redefinida para uma série temporal como

$$m_n^2 = \frac{1}{2q+1} \sum_{i=1}^{2q+1} x_i^{n^2} - \left(\frac{1}{2q+1} \sum_{i=1}^{2q+1} x_i^n \right)^2, \quad (4.15)$$

onde a soma na eq. 4.15 pode assumir os valores $n = 1, \dots, N - 2q$. Desta forma, a condição da seção t permanece a mesma com a variância local (eq. 4.15) no lugar da variância condicional (eq. 4.4).

Dado um valor de t , o i -ésimo termo da seção será da forma ⁵

$$s_t^i = y_1^i + y_2^i + \dots + y_{k-1}^i + y_k^i \quad (4.16)$$

onde $i + k \leq N - 2q$ e a condição

$$m_i^2 + m_{i+1}^2 + \dots + m_{i+k-1}^2 < t < m_i^2 + m_{i+1}^2 + \dots + m_{i+k-1}^2 + m_{i+k}^2 \quad (4.17)$$

é satisfeita. Logo, a variância local, juntamente com a eqs. 4.5 e a condição de seção t limitam a quantidade de termos da série $\{y_k\}$ presentes em cada realização de S_t . Uma vez que o número de termos presentes na série $\{y_k\}$ impõe uma limitação ao número de termos presentes na série da seção t , esta última possuirá no máximo $N - 2q$ termos, que será o caso de se considerar um valor para t inferior a qualquer das variâncias locais.

Vale a pena fazer uma comparação entre a série S_n e aquela relacionada a seção S_t . No primeiro caso, dado um valor de n , tanto a quantidade de termos quanto cada termo da série são bem definidos:

$$S_n = (s_n^1, s_n^2, \dots, s_n^{N-n}, s_n^{N-n+1}) \quad (4.18)$$

$$S_n = \left(\sum_{i=1}^n x_i^1, \sum_{i=1}^n x_i^2, \dots, \sum_{i=1}^n x_i^{N-n}, \sum_{i=1}^n x_i^{N-n+1} \right). \quad (4.19)$$

O mesmo não será válido para o segundo caso. Uma vez que a i -ésima realização da variável estocástica S_t é dada pelo somatório das realizações das variáveis Y_n pertencentes à seção t , cada realização de S_t será proveniente da soma de quantidades

⁵ O termo y_j^i refere-se aquele y_{j-1+i} da série $\{y_k\}$.

de termos diferentes. Já a eq. 4.15 referente à variância local garante a existência de um número inteiro $j \in [0, N - 2q]$ tal que a condição da seção t não será satisfeita, o que limita o número de termos presentes na série referente às realizações de S_t

$$S_t = (s_t^1, s_t^2, \dots, s_t^{j-1}) \quad (4.20)$$

$$S_t = \left(\sum_{i=1}^{n_1} y_i^1, \sum_{i=1}^{n_2} y_i^2, \dots, \sum_{i=1}^{n_{j-1}} y_i^{j-1} \right). \quad (4.21)$$

A forma em que a série referente à seção t é construída supõe a estacionariedade da série de $\{y_i\}$. Isto faz com que a extensão da seção de Lévy para o caso de séries temporais possa vir a falhar no caso de séries não estacionárias. Este problema poderia ser contornado por se tirar a tendência da série $\{y_k\}$, entretanto o fato de cada realização de S_t possuir diferentes números de termos agregados (da série $\{y_k\}$) inviabiliza este procedimento, sendo o caso mais simples a subtração da média local em cada termo da agregação.

Uma questão intrigante surge aqui. O teorema 1 implica a convergência da distribuição de probabilidade referente à série da seção t para a Gaussiana à medida que t tende para infinito. Isto nos leva intuitivamente a pensar num expoente de Hurst tendendo para o valor $1/2$ no processo de convergência. Por outro lado, o fato de cada realização de S_t ser proveniente da soma de quantidades de termos diferentes implica num aumento do grau de heterocedasticidade da série $\{s_t^i\}$ com o aumento no valor de t .

4.3 Significado Físico dos Momentos de Altas Ordens

Historicamente é comum o uso da estatística Gaussiana na descrição e modelagem de sistemas, sendo quantidades como a média e o desvio padrão da média muito importantes. Recentemente isto tem mudado e cada vez mais o uso de momentos de altas ordens (terceira e quarta ordens) têm se tornado comum no estudo das propriedades estatísticas de séries com significativo número de pontos.

Enquanto que os dois primeiros momentos possuem a mesma dimensão que a da variável aleatória analisada, a assimetria (momento de terceira ordem) e a curtose (momento de quarta ordem) são definidas de forma adimensional. Aqui, a assimetria e a curtose serão utilizadas como meio de verificação da convergência das séries das seções para o regime Gaussiano.

Assimetria

A assimetria de uma variável aleatória X é definida por:

$$S_{K_X} = \frac{1}{N} \sum_{i=1}^N \left[\frac{x_i - \langle X \rangle}{\sigma_X} \right]^3. \quad (4.22)$$

Seja agora um conjunto de N variáveis aleatórias $\{X_i\}$ ($i = 1, \dots, N$) e a variável soma $S_n = X_1 + \dots + X_n$.

A sua assimetria será:

$$S_{k_{S_n}} = \frac{1}{N} \sum_{i=1}^N \left[\frac{s_n^i - \langle S_n \rangle}{\sigma_{S_n}} \right]^3 \quad (4.23)$$

onde $s_n^i = x_1^i + x_2^i + \dots + x_n^i$ indica a i -ésima realização da variável aleatória S_n e σ_{S_n} o desvio padrão referente às realizações de S_n .

Um valor positivo da assimetria significa que a distribuição de freqüência das realizações de S_n possui uma calda alongada na direção de valores positivos. Já um valor de assimetria negativo implica que a calda é alongada na direção negativa destas realizações e um valor nulo de assimetria é referente a um caso de distribuição aproximadamente simétrica, conforme mostra a Fig. (4.1).

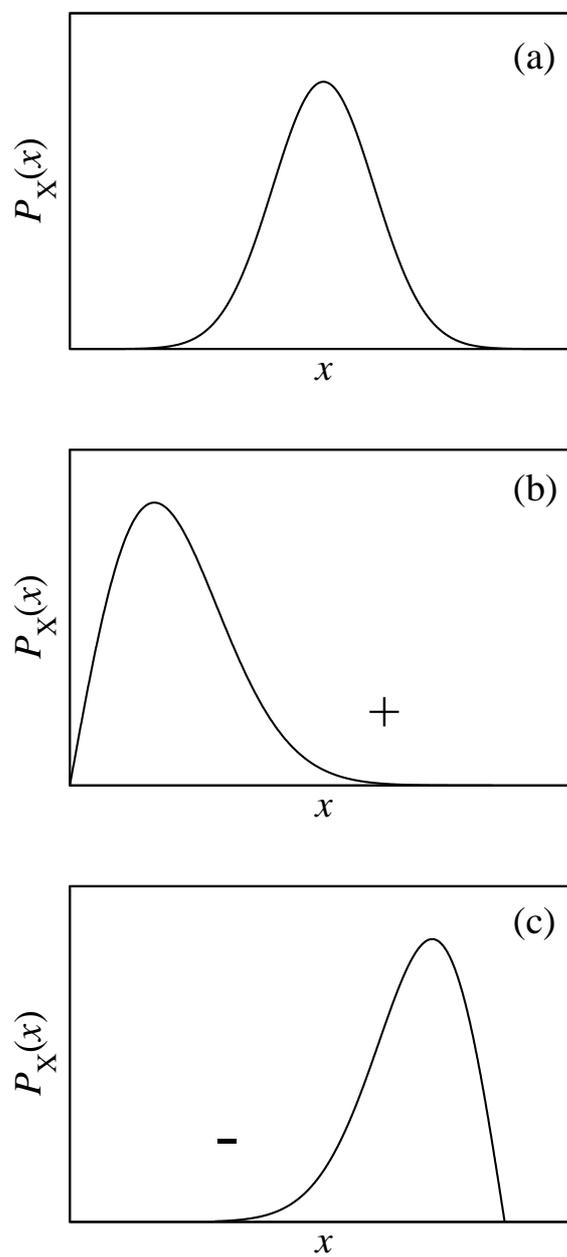


Figura 4.1: Distribuição de frequência das realizações de variáveis aleatórias: (a) distribuição com assimetria nula, (b) assimetria positiva e (c) assimetria negativa.

Existe uma relação importante entre as eqs. 4.22 e 4.23 quando as variáveis aleatórias X_i forem *i.i.d.*. Escrevendo $\langle S_n \rangle$ de forma explícita:

$$\langle S_n \rangle = \frac{1}{N} \sum_{i=1}^N s_n^i \quad (4.24)$$

onde s_n^i é dado por

$$s_n^i = \sum_{k=1}^n x_k^i \quad (4.25)$$

logo

$$\langle S_n \rangle = \sum_{k=1}^n \frac{1}{N} \sum_{i=1}^N x_k^i \quad (4.26)$$

$$\langle S_n \rangle = \sum_{k=1}^n \langle X_k \rangle. \quad (4.27)$$

Podemos reescrever a eq. 4.23 como:

$$\sigma_{S_n}^3 S_{k_{S_n}} = \frac{1}{N} \sum_{i=1}^N \left(s_n^{i3} - 3 s_n^{i2} \langle S_n \rangle + 3 s_n^i \langle S_n \rangle^2 - \langle S_n \rangle^3 \right) \quad (4.28)$$

utilizando as eqs. 4.25 e 4.27 temos:

$$\begin{aligned} \sigma_{S_n}^3 S_{k_{S_n}} &= \frac{1}{N} \sum_{i=1}^N \left[\sum_{j,k,l=1}^n x_j^i x_k^i x_l^i - 3 \sum_{j,k,l=1}^n x_j^i x_k^i \langle X_l \rangle + \right. \\ &\quad \left. + 3 \sum_{j,k,l=1}^n x_j^i \langle X_k \rangle \langle X_l \rangle - \sum_{j,k,l=1}^n \langle X_j \rangle \langle X_k \rangle \langle X_l \rangle \right] \\ &= \left[\sum_{j,k,l=1}^n \frac{1}{N} \sum_{i=1}^N x_j^i x_k^i x_l^i - 3 \sum_{j,k,l=1}^n \frac{1}{N} \sum_{i=1}^N x_j^i x_k^i \langle X_l \rangle + \right. \\ &\quad \left. + 3 \sum_{j,k,l=1}^n \frac{1}{N} \sum_{i=1}^N x_j^i \langle X_k \rangle \langle X_l \rangle - \frac{1}{N} \sum_{i=1}^N \sum_{j,k,l=1}^n \langle X_j \rangle \langle X_k \rangle \langle X_l \rangle \right] \\ &= \left[\sum_{j,k,l=1}^n \langle X_j X_k X_l \rangle - 3 \sum_{j,k,l=1}^n \langle X_j X_k \rangle \langle X_l \rangle + \right. \\ &\quad \left. + 3 \sum_{j,k,l=1}^n \langle X_j \rangle \langle X_k \rangle \langle X_l \rangle - \sum_{j,k,l=1}^n \langle X_j \rangle \langle X_k \rangle \langle X_l \rangle \right] \end{aligned}$$

$$\begin{aligned} \sigma_{S_n}^3 S_{K_{S_n}} = & \sum_{j=1}^n \left[\langle X_j^3 \rangle - 3 \langle X_j^2 \rangle \langle X_j \rangle + 2 \langle X_j \rangle^3 \right] + \sum_{\substack{j,k,l=1 \\ j \neq k}}^n \left[\langle X_j X_k X_l \rangle + \right. \\ & \left. - 3 \langle X_j X_k \rangle \langle X_l \rangle + 3 \langle X_j \rangle \langle X_k \rangle \langle X_l \rangle - \langle X_j \rangle \langle X_k \rangle \langle X_l \rangle \right]. \end{aligned} \quad (4.29)$$

No caso de variáveis *i.i.d.*, as seguintes igualdades: $\langle X_j X_k \rangle \langle X_l \rangle = \langle X_j \rangle \langle X_k \rangle \langle X_l \rangle$ e $\langle X_j X_k X_l \rangle = \langle X_j \rangle \langle X_k \rangle \langle X_l \rangle$ são válidas e o segundo somatório se torna nulo. Utilizando a relação $\sigma_{S_n}^3 = \sigma_X^3 n^{\frac{3}{2}}$ obtém-se:

$$S_{K_{S_n}} = \frac{S_{K_X}}{\sqrt{n}}. \quad (4.30)$$

Curtose

A curtose é uma medida de dispersão que caracteriza o “achatamento” de uma determinada distribuição de frequência das realizações de um conjunto de variáveis aleatórias $\{X_i\}$ ($i = 1, \dots, N$) quando comparada com a distribuição de frequência Gaussiana. Ela é definida como:

$$K_X = \frac{1}{N} \sum_{i=1}^N \left[\frac{x_i - \langle X_i \rangle}{\sigma_X} \right]^4 - 3, \quad (4.31)$$

onde σ_X é o desvio padrão referente às variáveis aleatórias X_i . A subtração do valor 3 na eq. 4.31 é devido ao fato de a curtose referente a uma distribuição Gaussiana ter o valor 3; logo, a eq. 4.31 é sempre referente a uma comparação com uma distribuição de frequência Gaussiana.

Se o valor é positivo então a distribuição em questão é mais alta (afunilada) e concentrada que a distribuição normal. Diz-se desta função probabilidade que

é leptocúrtica, ou que a distribuição tem caudas pesadas (o significado é que é relativamente fácil obter valores que se afastam da média a vários múltiplos do desvio padrão). No caso de um valor negativo, a distribuição é denominada platicúrtica — mais achatada que aquela Gaussiana e a denominação de mesocúrtica é dada para aquelas distribuições com curtose nula, conforme mostra a Fig. (4.2).

Similar à assimetria existe uma relação simples entre a eq. 4.31 e a curtose referente a S_n para o caso das variáveis aleatórias X_i *i.i.d.*. Esta relação é da forma:

$$K_{S_n} = \frac{K_X}{n}. \quad (4.32)$$

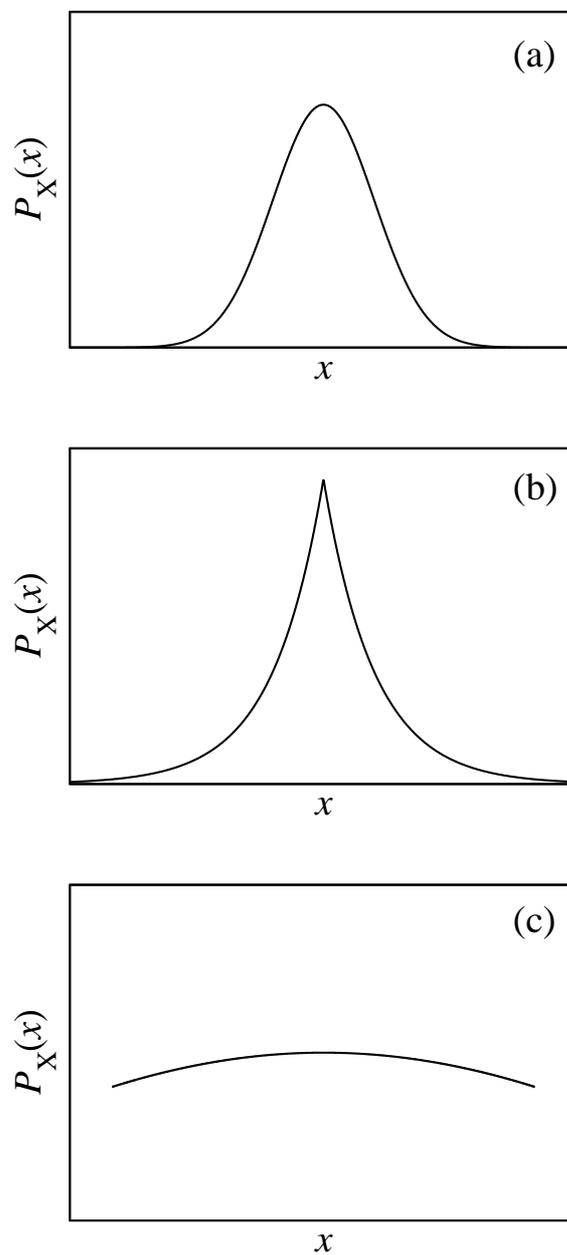


Figura 4.2: Distribuição de freqüência das realizações de variáveis aleatórias: (a) curtose nula (mesocúrtica), (b) curtose positiva (leptocúrtica) e (c) curtose negativa (platicúrtica).

Capítulo 5

Metodologia

O estudo das propriedades fractais das séries temporais foi feito a partir dos métodos DFA e MF-DFA, descritos no Cap. 3. Já a investigação da influência da organização temporal e da correlação dos sinais sobre as propriedades multifractais das séries foi feita mediante o uso de técnicas de embaralhamento de dados. No primeiro caso, a comparação das propriedades multifractais da série antes e depois de ter seus dados embaralhados mostra o efeito da organização temporal sobre tais propriedades. Já no segundo caso, dada uma série temporal $\{x_i\}$ com $i = 1, \dots, N$, tomemos cada termo da série como: $x_i = s_i|x_i|$, onde $s_i = \pm 1$. A comparação entre as propriedades multifractais da série original e aquelas referentes à série com os sinais s_i embaralhados revela a influência da correlação dos sinais. A Tab. 5.1 ilustra as técnicas de embaralhamento de dados utilizadas no nosso trabalho.

A primeira coluna da Tab. 5.1 traz um conjunto de valores originais. A segunda coluna mostra o mesmo conjunto de valores após ter sua ordem embaralhada. A série de dados embaralhados fica livre de toda correlação temporal e, uma vez integrada (seção 3.3), ela se comporta como um “random walk” cujo expoente H é

série original	série embaralhada	sinal embaralhado
-1.0	-1.0	1.0
1.5	-2.0	-1.5
-0.5	3.0	0.5
3.0	-0.5	-3.0
-2.0	1.5	-2.0

Tabela 5.1: Tabela ilustrativa das técnicas de embaralhamento de dados.

1/2. A terceira coluna da Tab. 5.1 mostra o mesmo conjunto de dados da primeira, porém, com a ordem dos sinais embaralhados. É importante ressaltar que apesar da série com sinais embaralhados ser isenta de toda correlação temporal, ela mantém as suas correlações na volatilidade (em se tratando de séries financeiras) uma vez que a modulação inicial é mantida.

O estudo da convergência das seções de Lévy ao regime Gaussiano foi feito mediante a comparação entre os momentos normalizados da série S_n e aqueles da série das seções S_t , todos discutidos previamente. Neste ponto, chamamos a atenção para o fato de que, tanto S_n , definida pela soma de n realizações da série original, quanto a série das seções S_t , são formadas por termos sobrepostos (veja eqs. 4.19 e 4.21). Pode-se compreender uma vantagem da definição das séries com termos sobrepostos sobre aquela com termos não sobrepostos, a partir da análise da própria série S_n . Dada uma série temporal $\{x_i\}$ ($i = 1, 2, \dots, N$), a i -ésima “realização” da série S_n com termos não sobrepostos é dada por:

$$s_n^i = \sum_{j=i-n+1}^n x_j . \quad (5.1)$$

Isto faz com que o número de termos de S_n seja da ordem de $\frac{N}{n}$. Já aquela série

definida na eq. 4.19 ($s_n^i = \sum_{j=1}^n x_j^i \equiv \sum_{j=1}^n x_{i+j-1}$) possuirá $N-n$ termos, garantindo melhores análises estatísticas.

Capítulo 6

Resultados e Discussões

Primeiramente apresentamos os resultados da análise multifractal para a série cambial do Marco Alemão — German Deutsche Mark (DEM) frente ao Dólar Americano (USD), em seguida apresentamos os resultados da aplicação das seções de Lévy à mesma série. Por último mostramos alguns resultados preliminares referentes ao estudo das propriedades fractais das seções de Lévy.

A série cambial DEM versus USD foi obtida de Reuters EFX (fornecida por Olsen & Associates). Ela é referente a um período de 1 ano (1 de Outubro de 1992 a 30 de Setembro de 1993) e corresponde a um total de 1 472 240 pontos, o que equivale aproximadamente a um ponto a cada 20 s (Fig. 6.1).

6.1 Heterocedasticidade e Multifractalidade

A Fig. 6.1(a) é referente à série cambial de DEM versus USD. Já as Fig. 6.1(b) e (c) referem-se às séries de retorno e volatilidade obtidas a partir da série cambial mediante a aplicação das eqs. 1.3 e 1.4, respectivamente.

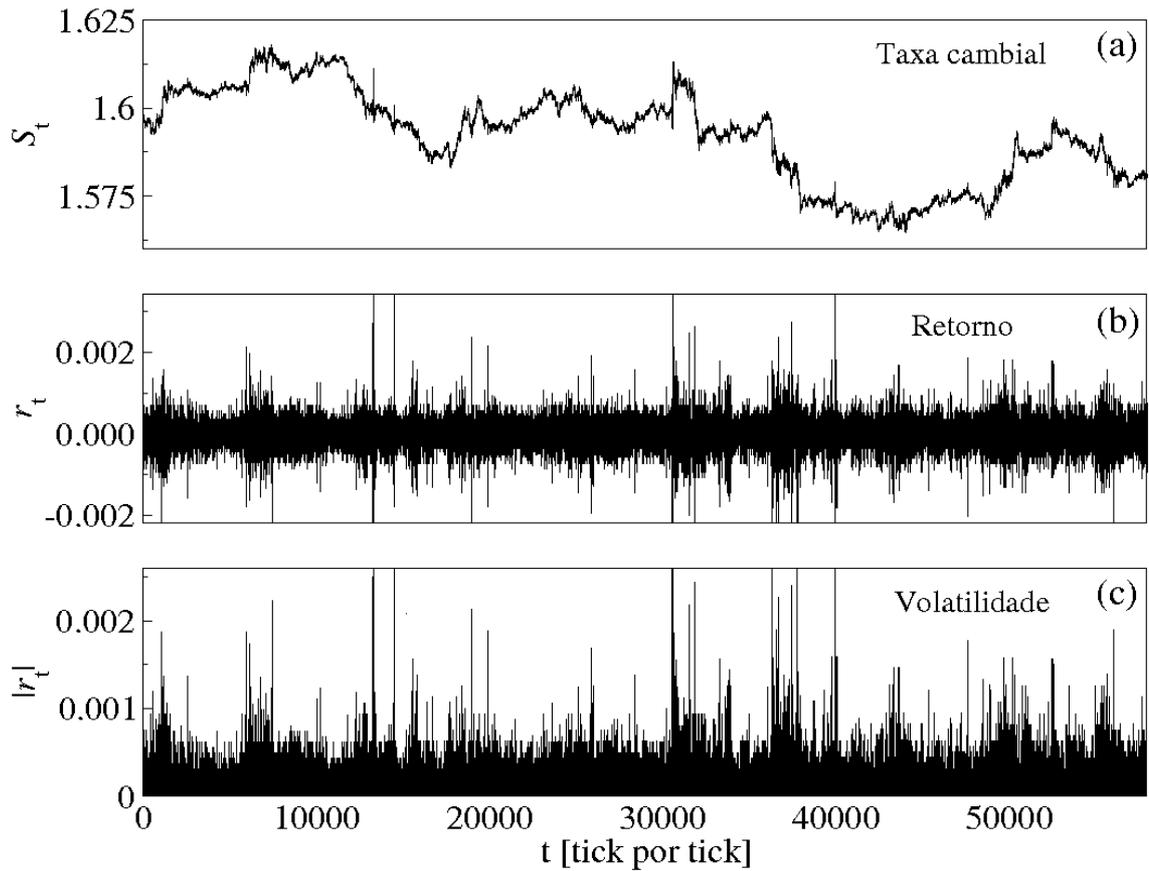


Figura 6.1: (a) Amostra de $6.0 \cdot 10^4$ de um total de 1 472 240 pontos contidos na série cambial do marco frente ao dólar. Os dados são referentes ao período de 1 de Outubro 1992 a 30 de Setembro de 1993. (b) Série de retornos financeiros e (c) Volatilidade. O termo tick (em inglês) é frequentemente utilizado para se referir a transações eletrônicas. Note o caráter heterocedástico em ambas Fig. 6.1(b) e (c).

A Fig. 6.2 é referente a análise multifractal feita a partir do método MF-DFA para os retornos heterocedásticos, retornos embaralhados não heterocedásticos e para os retornos com apenas os sinais embaralhados. É importante frisar que apesar deste último ter suas auto-correlações de retorno destruídas, ele mantém tanto seu caráter heterocedástico quanto suas correlações de volatilidade uma vez que a modulação de sua volatilidade foi mantida. Os expoentes $\tau(q)$ e $h(q)$ são relacionados pela eq. $\tau(q) = qh(q) - 1$ (veja na seção 3.3).

Podemos observar tanto o efeito da organização temporal quanto da correlação dos sinais sobre os retornos (losangos sem preenchimento), ao comparar o comportamento de seu expoente h em função de q , com aqueles dos retornos embaralhados (losangos preenchidos) e os retornos com sinais embaralhados (símbolos de adição). Enquanto h é aproximadamente constante para os retornos embaralhados, mostrando uma forte dependência com a organização temporal, seu comportamento não sofre uma mudança muito drástica no caso dos retornos com sinais embaralhados.

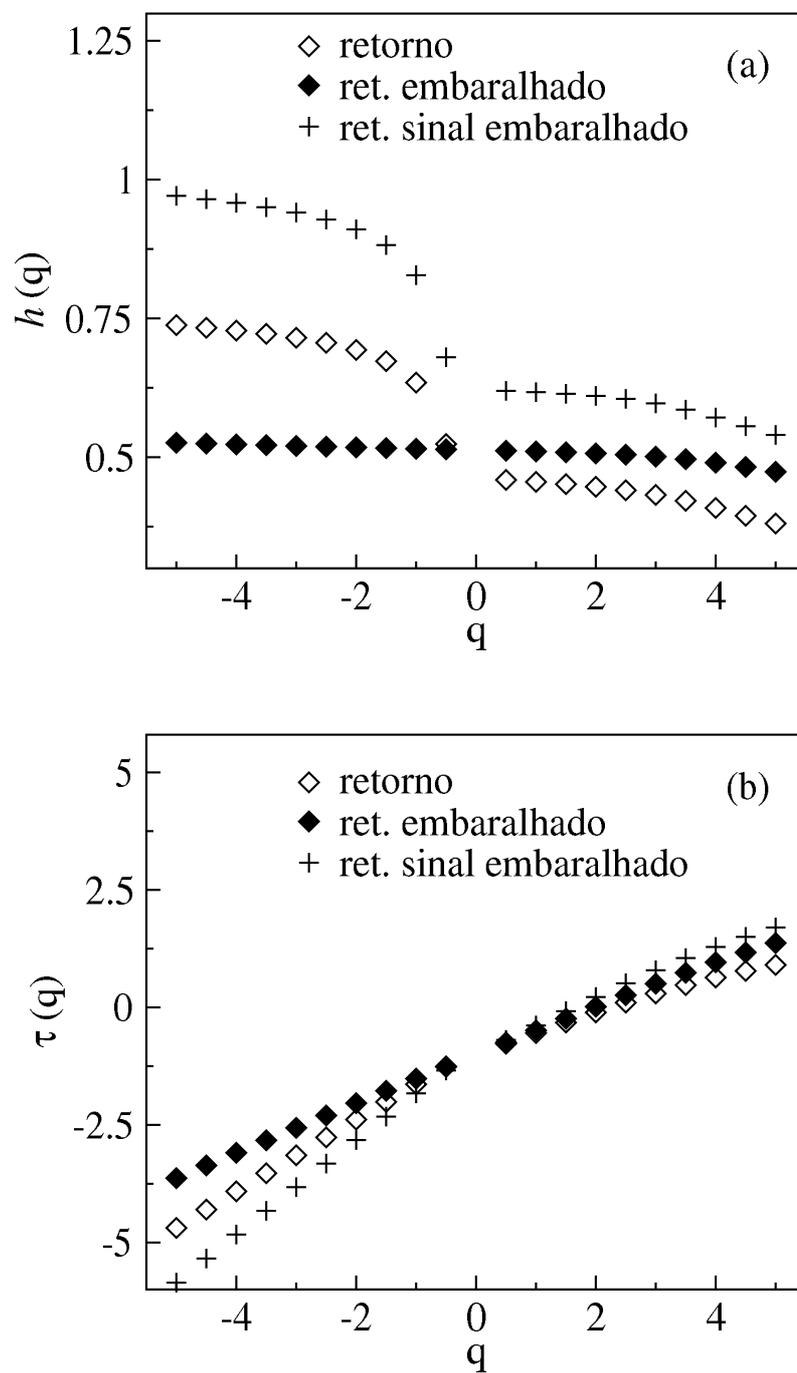


Figura 6.2: Análise Multifractal feita a partir do método MF-DFA. (a) Expoente de Hurst generalizado. (b) Expoente τ da função partição.

A Fig. 6.3 é referente ao ajuste quadrático dos espectros multifractais para os retornos heterocedásticos, retornos embaralhados não heterocedásticos e para os retornos com apenas os sinais embaralhados. O espectro multifractal é obtido a partir da transformação de Legendre do expoente τ (eq. 3.21 da seção 3.3).

Aqui, o efeito da organização temporal se mostra no fato do espectro multifractal referente aos retornos embaralhados ser muito mais estreito que o espectro referente aos retornos. O mesmo estreitamento do espectro para a série de retornos com sinais embaralhados não é observada.

Observamos um deslocamento no espectro dos retornos com sinais embaralhados, onde o coeficiente de Hölder dominante (para $f = 1$) se desloca para valores distantes de $\alpha = 1/2$. Provavelmente, isto ocorra devido a correlação negativa entre volatilidade-retorno conhecido na literatura como efeito de alavanca (“leverage effect” em inglês). Se o retorno diminui, a volatilidade tende a aumentar e se o retorno aumenta, a volatilidade tende a diminuir; o mesmo raciocínio não é válido a partir da análise na volatilidade. Uma consequência do efeito de alavanca é uma assimetria negativa na PDF dos retornos [40], que por sua vez é eliminada mediante o embaralhamento dos sinais.

As Figs. 6.4 e 6.5 são referentes a análise multifractal feita a partir do método MF-DFA e ao ajuste quadrático dos espectros multifractais para a volatilidade e volatilidade embaralhada, respectivamente.

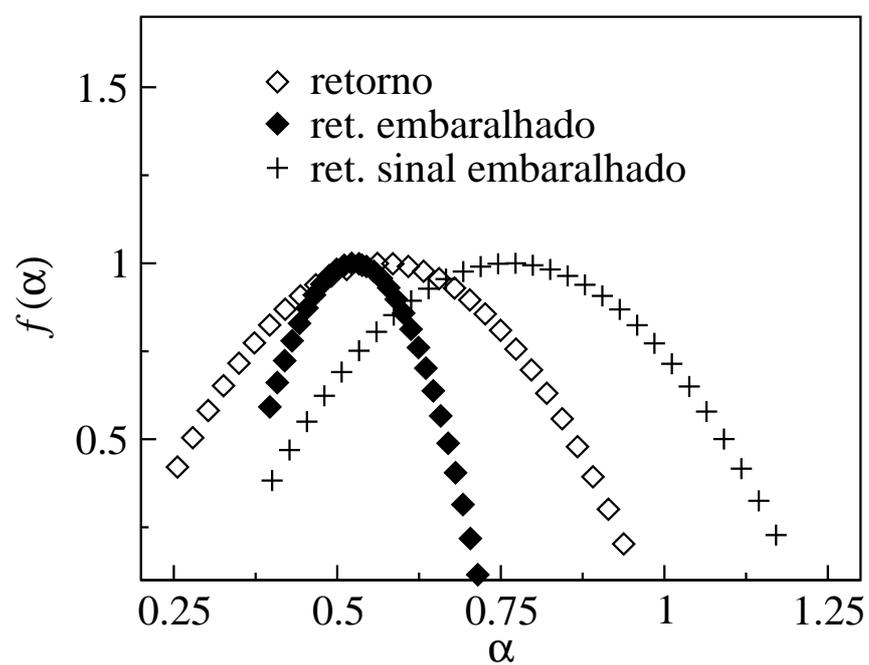


Figura 6.3: Ajuste quadrático dos espectros multifractais. Note que a série não heterocedástica (retorno embaralhado) possui o menor grau de multifractalidade.

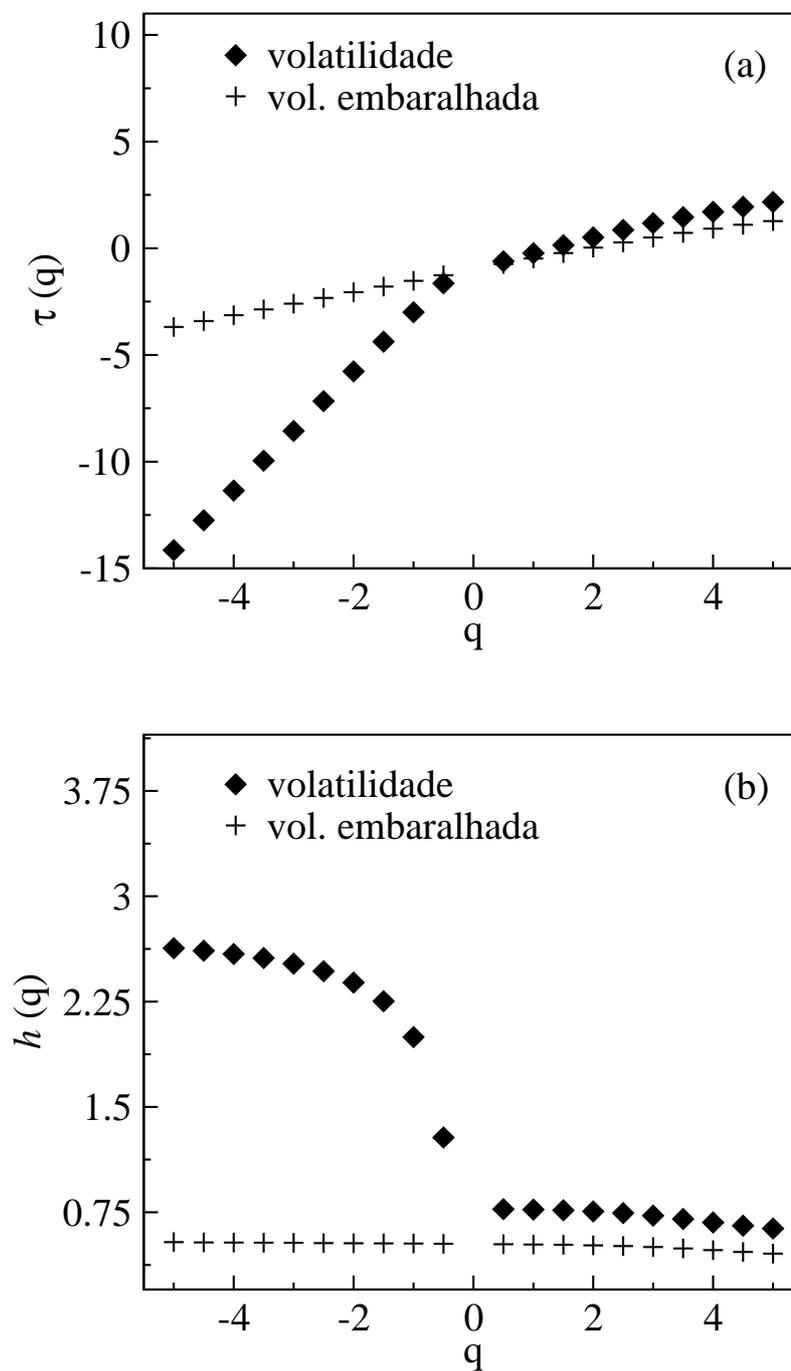


Figura 6.4: Análise Multifractal feita a partir do método MF-DFA para a volatilidade e para a volatilidade embaralhada. (a) Expoente de Hurst generalizado. (b) Expoente τ da função partição.

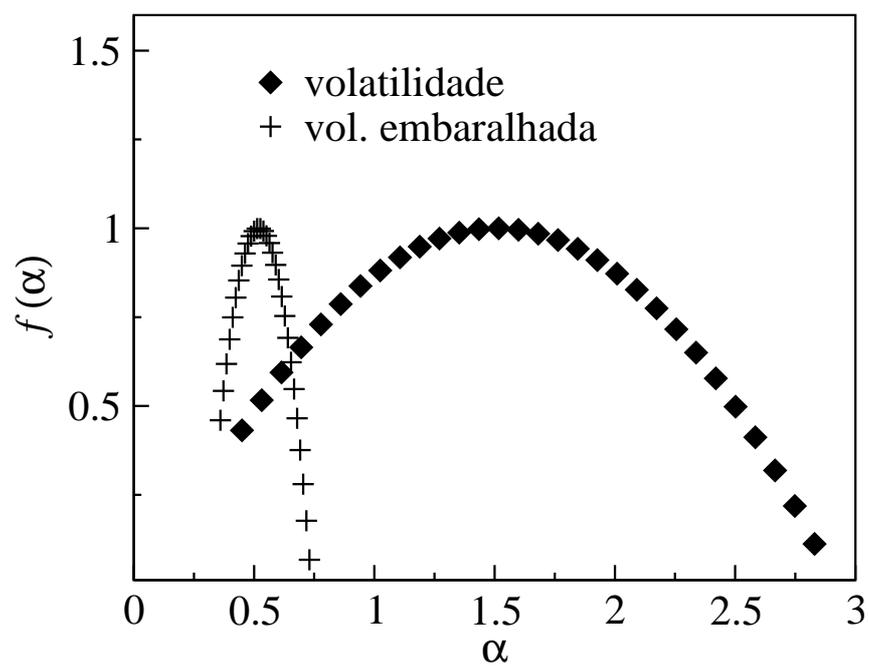


Figura 6.5: Ajuste quadrático dos espectros multifractais. Note que a série não heterocedástica, neste caso a volatilidade, tem seu caráter multifractal drasticamente reduzido mediante a eliminação de sua heterocedasticidade.

Similares observações podem ser feitas para a volatilidade. A Fig. 6.5 mostra que eliminar a heterocedasticidade dos retornos absolutos conduz a uma significativa redução nas propriedades fractais. Parte da importância das Fig. 6.4 e 6.5 situa-se no fato que, enquanto o espectro multifractal está bem próximo de $\alpha = 1/2$ para os retornos, como esperado da *Hipótese de Mercado Eficiente*, o espectro multifractal da volatilidade é centrado em um valor maior de α , consistente com a existência de correlações de longo alcance. Acreditamos que a heterocedasticidade possivelmente causa multifractalidade introduzindo correlação de longo alcance no valor absoluto da variável aleatória.

Na tentativa de testar nossa hipótese de que multifractalidade e heterocedasticidade estão relacionadas, aplicamos o método MF-DFA a uma diferente base de dados. Escolhemos estudar um sistema completamente diferente: a série temporal de áudio [41] do primeiro movimento da quinta sinfonia em *C* menor, Op. 67 (Fig. 6.6(a)) de Ludwig van Beethoven .

Note o comportamento fortemente heterocedástico da série. As Figs. 6.6(b) e 6.6(c) mostram a série embaralhada e o espectros multifractais para a série original e a embaralhada. Como esperado, a multifractalidade reduz-se drasticamente com a eliminação da heterocedasticidade.

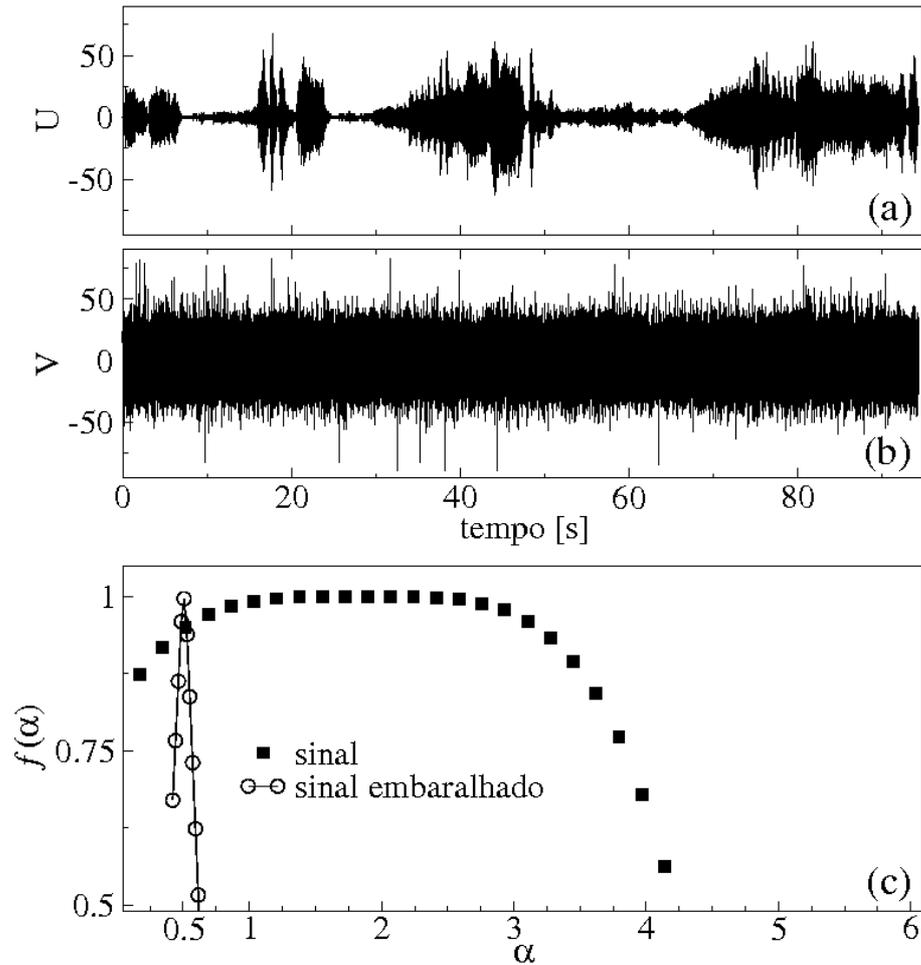


Figura 6.6: Aproximadamente 20% do primeiro movimento da quinta sinfonia de Beethoven. (a) O sinal original U_i , (b) O sinal embaralhado V_i . (c) Ajuste de 4ª ordem (quadrados) e interpolação (círculos) dos espectros multifractais. A multifractalidade desaparece quando a heterocedasticidade do sinal é eliminada.

6.2 Seções de Lévy e Convergência à Gaussiana

O estudo da convergência à Gaussiana das seções de Lévy foi feito a partir da análise dos momentos normalizados de terceira e quarta ordem da série DEM-USD. O valor inicial no parâmetro t da seção foi da ordem de 10^{-15} , de forma a garantir a coincidência entre a série da seção inicial e aquela original. Já os incrementos utilizados no parâmetro t da seção foram da ordem de 10^{-6} garantindo uma suavidade tanto nas curvas de curtose quanto naquelas referentes à assimetria.

A Fig. 6.7 é referente ao comportamento da curtose para a série das seções t e a série normal S_n . Já a Fig. 6.8 mostra o comportamento da série das seções em função do número de termos utilizados na estimativa da variância local (eq. 4.15 da seção 4.2).

As Figs. 6.9 e 6.10 são similares às Figs. 6.7 e 6.8 e mostram o comportamento da assimetria para a série das seções S_t e a série S_n , bem como o comportamento da assimetria em função do parâmetro q .

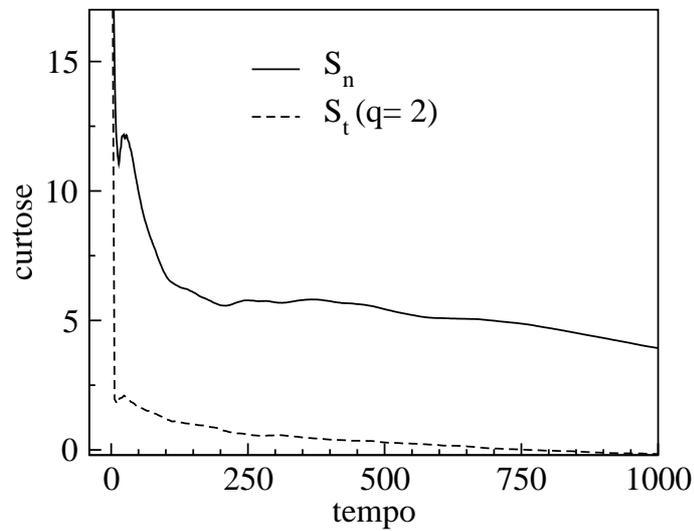


Figura 6.7: Comportamento da curtose para as séries das seções S_t com parâmetro $q = 2$ (linha tracejada) e para a série normal S_n (linha contínua), ambas provenientes da taxa cambial DEM versus USD. Note a rápida convergência da série das seções.

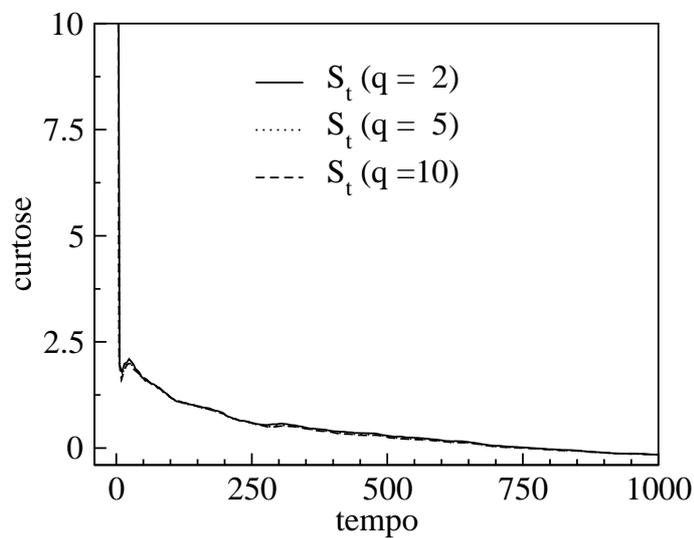


Figura 6.8: Comportamento da curtose para as séries das seções S_t para diferentes valores do parâmetro q . Note a pequena sensibilidade da curtose frente ao parâmetro q .

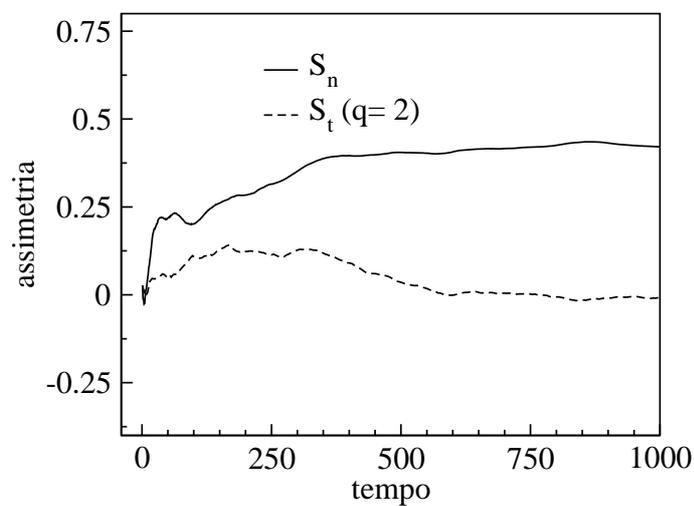


Figura 6.9: Comportamento da assimetria para as séries das seções S_t com parâmetro $q = 2$ (linha tracejada) e para a série normal S_n (linha contínua), ambas provenientes da taxa cambial DEM versus USD. Note a rápida convergência da série das seções.

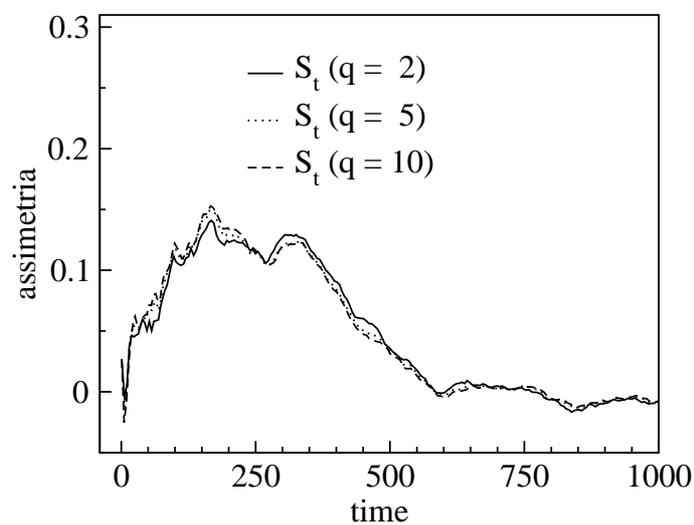


Figura 6.10: Comportamento da assimetria para as séries das seções S_t para diferentes valores do parâmetro q . Note a pequena sensibilidade da assimetria frente ao parâmetro q .

6.3 Estudo da Convergência das Seções de Lévy

Na busca do entendimento do processo de convergência das Seções de Lévy, estimamos o expoente de Hurst para várias séries, cada uma com um valor específico de seção t . A estimativa do expoente H foi feita a partir do método DFA.

A Fig. 6.11 mostra as séries das seções em ordem crescente do valor da seção. Já a Fig. 6.12(a) mostra o comportamento de $\log F_2(s)$ versus $\log(s)$ das séries das seções para valores crescentes de t . Já a Fig. 6.12(b) mostra o comportamento de $\log F_2(s)$ versus $\log(s)$ de algumas das séries das seções mostradas na Fig. 6.12(a) bem como um ajuste linear para cada uma delas. O coeficiente angular obtido na regressão linear é referente ao expoente de Hurst da série.

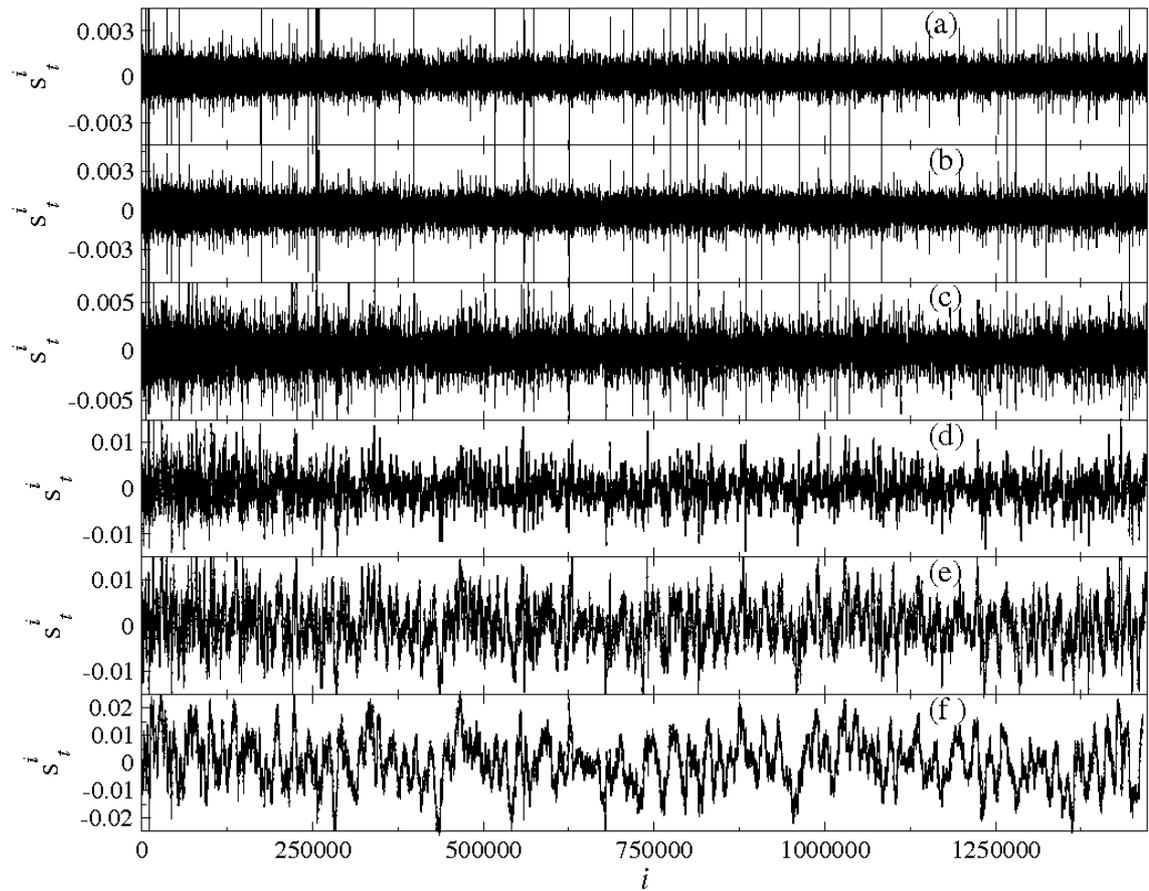


Figura 6.11: Séries das seções para diferentes valores de t . (a) $t = 2.5 \cdot 10^{-7}$, (b) $t = 1.0 \cdot 10^{-6}$, (c) $t = 1.0 \cdot 10^{-5}$, (d) $t = 8.5 \cdot 10^{-5}$, (e) $t = 1.8 \cdot 10^{-4}$ e (f) $t = 5.2 \cdot 10^{-4}$. Note a mudança no perfil da série com o aumento do valor da seção, quanto maior o valor da seção, mais suave é a série.

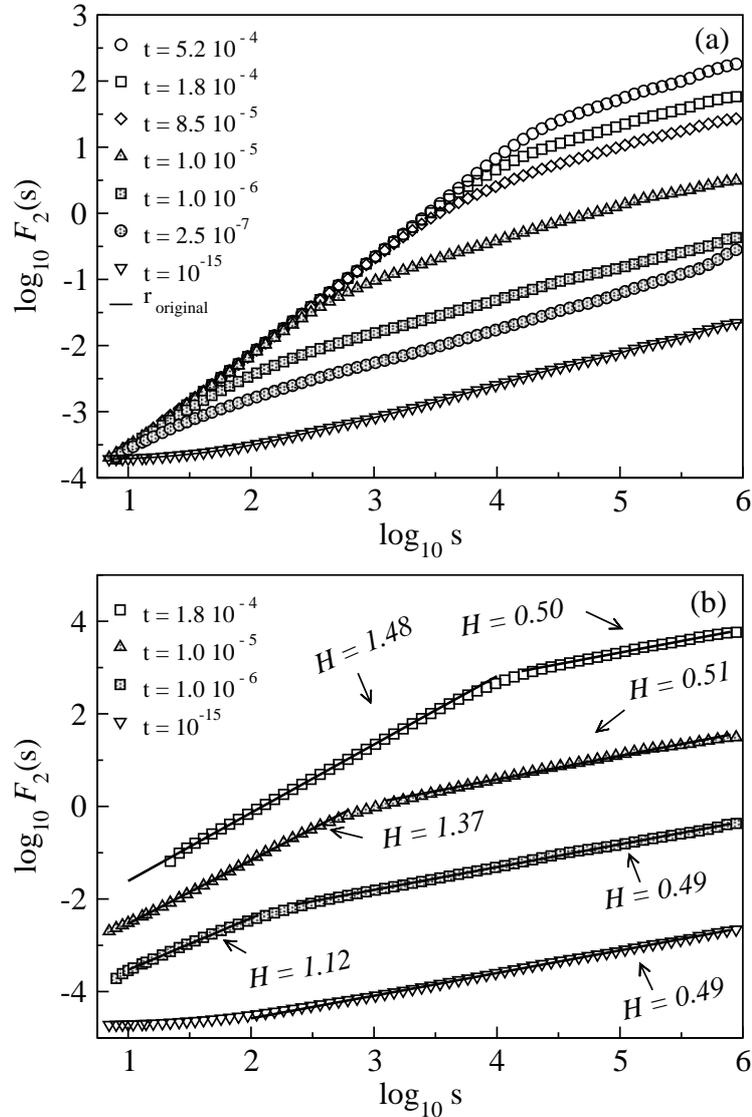


Figura 6.12: (a) Comportamento de $\log F_2(s)$ versus $\log(s)$ das séries das seções para valores crescentes de t . (b) Comportamento de $\log F_2(s)$ versus $\log(s)$ das séries das seções para alguns valores de t (símbolos) mostrados em Fig. 6.12(a) juntamente com as regressões lineares (linhas contínuas). Note que podemos associar dois valores do expoente H a cada curva, isto indica um *joelho* ou “crossover” na escala de $F_2(q)$. As séries na Fig. 6.12(b) foram deslocadas para efeitos de visualização.

Podemos perceber a existência de correlações de curto alcance na série de retornos DEM versus USD, onde o valor H associado a pequenos valores de s é diferente de $1/2$, porém, para grandes valores de s a série se comporta como um movimento aleatório (ruído branco) com expoente H muito próximo de $1/2$. Fica claro na Fig. 6.12(b) que a medida que aumentamos o valor da seção, a curva $\log F_2(s)$ versus $\log(s)$ passa a ter um *joelho* ou “crossover” conectando dois valores de expoente H . Enquanto o expoente H para pequenos valores de s tende à $3/2$ (característico de um ruído browniano — integração de um ruído branco [5]) à medida que o valor da seção aumenta, ele H permanece em torno de $1/2$ para grandes valores de s . Notamos também que à medida que aumentamos o valor da seção, a faixa de s com predominância do tipo ruído branco diminui, enquanto aquela referente ao ruído browniano aumenta. Portanto, a forma com que a seção agrega os termos da série se comporta como uma segunda integração à medida que o valor da seção t aumenta.

Capítulo 7

Conclusões

7.1 Flutuações Heterocedásticas Multifractais

Mostramos neste trabalho que no caso de efeitos de correlação em sinais e caudas de lei de potência não desempenharem um papel significativo, então as condições necessárias e suficientes para a validade do Teorema do Limite Central podem ser satisfeitas. Isto mostra que o regime Gaussiano é um caso especial, sendo o regime multifractal-heterocedástico o caso geral. Nesta tese investigamos a relação entre natureza heterocedástica de sinais e suas propriedades multifractais e mostramos que a existência da primeira é condição suficiente mas não necessária para causar a segunda. Investigamos também o processo de convergência de séries de retornos financeiros heterocedásticas multifractais para o regime homocedástico monofractal usando o método matemático das seções de Lévy.

7.2 Relevância dos Resultados

Sabemos que conceitos importantes como mono e multifractalidade vão além dos casos geométricos, e surgem também em uma diversa gama de sistemas complexos. Mostramos a conexão de tais conceitos com séries temporais. Enquanto que no caso geométrico a fractalidade é mostrada por meio da auto-similaridade e dimensão fracionária, em séries temporais a auto-similaridade se apresenta como uma simetria de invariância de escala da função densidade de probabilidade dos sinais. Já a medida da dimensão fractal é feita de forma indireta, a partir do seu expoente de Hurst H que corresponde à rugosidade da série. Portanto, o expoente H representa uma medida global, sendo sua relação com a dimensão fractal da forma: $d = 2 - H$. Por outro lado, o expoente de Hölder reflete medidas locais da série, sendo muito importante no estudo de sistemas onde são necessários mais de um único expoente para serem caracterizados.

Mostramos uma relação direta entre heterocedasticidade e multifractalidade, onde vimos que o primeiro é causa suficiente mas não necessária para o segundo [42].

Vimos que a função de correlação de um sistema está intimamente relacionada com a variância de seus sinais, o que faz da variância uma medida fundamental do sistema. Mostramos também que apesar de a série de retornos financeiros possuir correlação de curto alcance, a série de volatilidade possui correlações de longo alcance, mas nos dois casos as correlações têm sua origem diretamente relacionada com a organização temporal dos sinais. Isto nos leva a acreditar que a heterocedasticidade introduz correlação na auto-correlação do módulo dos sinais. Analisamos também o processo de convergência da PDF dos retornos financeiros ao regime Gaus-

siano via Seções de Lévy [38]. Percebemos uma rápida convergência das séries das seções para o regime Gaussiano. Mostramos que a rápida convergência das séries das seções pode ser entendida como uma segunda integração da mesma, sendo a primeira realizada pelo próprio método DFA, o qual foi utilizado na análise das séries das seções e a segunda integração como intrínseco do processo de convergência realizado pelo método das Seções de Lévy. Esta segunda integração se mostra a partir do surgimento de um joelho na escala da função $\log F_2(s)$ versus $\log s$. Enquanto que para grandes valores de s (à direita do joelho), o valor do expoente de escala da série da seção permanece o mesmo da série original, para pequenos valores de s (à esquerda do joelho), observamos uma tendência do expoente de escala para $3/2$ (típico de um ruído browniano) à medida que aumentamos o valor da seção t , promovendo desta forma uma maior suavidade na série das seções. Acreditamos que no limite assintótico de $t \rightarrow \infty$ o joelho deixa de ser observado, restando somente a característica de ruído browniano. É provável que este comportamento seja independente do expoente de escala da série original. Se for este o caso, sinais provenientes das seções de Lévy podem ser modelados por equações “simples”.

7.3 Perspectivas Futuras

Estas mesmas análises multifractais podem ser realizadas para séries tais como: sísmicas, eletrocardiogramas (ECG) e eletroencefalogramas (EEG). Sinais sísmicos apresentam períodos de flutuações uniformes e também períodos de grandes variações nos sinais provocados por terremotos: este comportamento reflete o caráter

heterocedástico de tais sinais. O mesmo ocorre em sinais eletrocardiográficos ou eletroencefalográficos, onde a ocorrência de ataques epiléticos podem causar uma dinâmica heterocedástica.

Acredito que estas análises podem ser aplicadas até mesmo em problemas de cunho industrial, com repercursões muito abrangentes. Suponha um processo de produção em série de algum produto que possua um alto nível de acompanhamento pós-venda. Como exemplo pensemos na indústria de automóveis, onde um veículo passa por revisões periódicas. A análise de séries temporais contendo a informação de defeito ou perfeito funcionamento de uma parte de um veículo (tais séries poderiam ser fornecidas por oficinas autorizadas), poderia fornecer informações sobre a eficiência dos processos de amostragem realizados pelas indústrias. Um valor de expoente $H = 1/2$, indicaria aleatoriedade nos eventos (defeitos). Ainda neste exemplo, suponha que uma vez que uma peça apresente um defeito, os dados obtidos daquele veículo a partir de então passassem a ser admitidos em um segundo grupo. Os coeficientes H obtidos para este segundo grupo poderiam inferir sobre a qualidade da assistência técnica oferecida na oficina autorizada cujos dados são oriundos. A partir de tais informações, um consumidor optaria por esta ou aquela marca de veículo (não poderíamos comparar dados de veículos luxuosos com os populares) apenas pelo conforto ou valor comercial, no caso de as marcas comparadas apresentarem o mesmo coeficiente referente ao processo de montagem. Isso promoveria uma maior competitividade de mercado (os preços dos veículos iriam baixar). Além disso, de posse de informações do valor de H para o grupo de veículo que apresentou defeito, o consumidor teria um melhor critério na escolha de uma oficina autorizada. Estes resultados poderiam ser utilizados por empresas (até mesmo por

prefeituras de municípios, governos de estados, etc.) como critério adicional em processos de licitações. Marcas emergentes, em busca de “bons” valores dos expoentes, ofereceriam melhores condições aos consumidores. Uma vez obtido boas análises de seus produtos, tais empresas gastariam muito menos com propaganda.

Nossa análise pode também ser aplicada a problemas ecológicos como extinção de espécies. Tais processos são não estacionários e seus sinais são heterocedásticos e a origem da heterocedasticidade pode estar associada a situações distintas. Por exemplo, escassez de alimentos, competição entre duas espécies ou variações climáticas bruscas. Não é óbvio que em todos estes casos a heterocedasticidade desempenhe o mesmo papel.

Enfim, as perspectivas para o futuro nesta linha de pesquisa são diversas e amplas.

FIM

Referências Bibliográficas

- [1] B. B. Mandelbrot, *The fractal geometry of nature*, Freeman, New York (1982).
- [2] M. Serva, U. L. Fulco, I. M. Gléria, M. L. Lyra, F. Petroni, and G. M. Viswanathan, *Physica A*, **363**, 393 (2006).
- [3] I. Gléria, R. Matsushita, S. Silva, *Revista Brasileira de Ensino de Física*, **26**, 99 (2004).
- [4] M. Buchanan, *Ubiquity: The Science of History ... or Why the World Is Simpler Than We Think*. Weidenfeld & Nicolson, Londres (2000).
- [5] R. N. Mantegna, H. E. Stanley, *An Introduction to Econophysics, Correlations and Complexity in Finance*, Cambridge: Cambridge University Press (2000).
- [6] L. Bachelier, *Ann. Sci. Ec. Norm. Super.*, **17**, 21 (1900).
- [7] *Louis Bachelier* — Disponível em:
http://en.wikipedia.org/wiki/Louis_Bachelier#Citations. Acesso em:
28 out. 2007.
- [8] R. Brown, *Phil. Mag.*, **4(1)**, 161 (1828).
- [9] E. F. Fama, *J. Finance*, **25**, 383 (1970).

- [10] *Curso de Análise Fundamentalista e Análise de Balanços* — Disponível em:
http://www.ricardoborges.com/analise_fundamentalista.htm. Acesso em: 28 out. 2007.
- [11] Murray Ralph Spiegel, *Estatística*, MacGraw-Hill do Brasil, São Paulo (1971).
- [12] L.E. Reichl, *A Modern Course in Statistical Physics*, John Wiley & Sons, USA (1998).
- [13] George B. Arfken and Hans J. Weber, *Mathematical Methods for Physicists*, Academic Press, USA (1995).
- [14] Frederick Reif, *Fundamentals of Statistical and Thermal Physics*, International Edition. MacGraw-Hill (1965).
- [15] P. Lévy, *Calcul des probabilités*, Gauthier-Villars, Paris (1925).
- [16] C. L. Faustino, L. R. da Silva, M. G. E. da Luz, E. P. Raposo and G. M. Viswanathan, *Europhysics Letters*, **77**, 30002-p1 (2007).
- [17] Sergio da Silva, Raul Matsushita, Iram Gleria, Annibal Figueiredo and Pushpa Rathie, *Physica A*, **10(4)**, 365 (2005).
- [18] J. Feder, *Fractals*, Plenum Press, New York (1988).
- [19] *Dimension by Box-Counting*. Disponível em:
<http://classes.yale.edu/fractals/Labs/CoastlineLab/CoastlineLab.html>. Acesso em: 2 nov. 2007.
- [20] A.-L. Barabási and H. E. Stanley, *Fractal Concepts in Surface Growth*, Cambridge University Press, New York, USA (1995).

-
- [21] H.E. Stanley, P. Meakin, *Multifractal phenomena in physics and chemistry*, **335**, 405 (1988).
- [22] G. M. Viswanathan, U. L. Fulco, M. Lyra and M. Serva, *Physica A*, **329**, 273 (2003).
- [23] M. Serva, U. L. Fulco, M. Lyra and G. M. Viswanathan, *Physica A*, **63 (2)**, 393 (2006).
- [24] G. M. Viswanathan, C. -K. Peng, H. E. Stanley, A. L. Goldberger, *Phys. Rev. E*, **55**, 845 (1997).
- [25] G. M. Viswanathan, S. V. Buldyrev, S. Havlin, H. E. Stanley, *Biophysics Journal*, **72**, 866 (1997).
- [26] *Fractal Objects and Self-Similar Processes*. Disponível em:
<http://www.physionet.org/tutorials/fmnc/node3.html#SECTION00021000000000000000>. Acesso em: 2 nov. 2007.
- [27] Iram Gleria, Raul Matsushita and Sergio Da Silva, *Economics Bulletin*, **7(3)**, 1 (2002).
- [28] H.E. Hurst, Long Term Storage Capacity of Reservoirs, *Transactions of the American Society of Civil Engineers*, **116(1)**, 770 (1951).
- [29] S.R.S. Souza, B.M. Tabak, D.O. Cajueiro, *Revista Brasileira de Economia*, **60(2)**, 193 (2006).
- [30] A.L. Barabasi and T. Vicsek, *Phys. Rev. A*, **44**, 2730 (1991).

-
- [31] K. Matia, Y. Ashkenazy and H.E. Stanley, *Europhysics Letters*, **61(3)**, 422 (2003).
- [32] J.W. Kantelhardt, S.A. Zschiegner, E. Koscielny-Bunde, S. Havlin, A. Bunde and H.E. Stanley, *Physica A*, **316**, 87 (2002).
- [33] K. Hu, P.C. Ivanov, Z. Chen, P. Carpena, H.E. Stanley, *Phys. Rev. E*, **64**, 011114.1 (2001).
- [34] H.-O. Peitgen, H. Jürgens and D. Saupe, *Chaos and Fractals*, Springer-Verlag, New York (1992).
- [35] T. C. Halsey, M. H. Jensen, L. P. Kadanoff, I. Procaccia and I. Shraiman, *Physical Review A*, **33(2)**, 1141 (1986).
- [36] H.B. do Nascimento Jr, U.L. Fulco, M.L. Lyra, M. Serva e G.M. Viswanathan, *Revista Brasileira de Ensino de Física*, **29(3)**, 341 (2007).
- [37] A. Figueiredo, Iram Gleria, R. Matsushita, S. da Silva, *Physica A*, **40**, 5783 (2007).
- [38] A. Figueiredo, R. Matsushita, S. da Silva, M. Serva, G.M. Viswanathan, C. Nascimento, Iram Gleria, *Physica A*, **386**, 756 (2007).
- [39] P. Lévy, *Théorie de l'addition de variables aléatoires*, Gauthiers-Villars, Paris (1927).
- [40] J. P. Bouchaud, A. Maticz and M. Potters, *Physical Review Letters*, **87(22)**, 228701-1 (2001).

-
- [41] H.D. Jennings, P.Ch. Ivanov, A.M. Martins, P.C. Silva, G.M. Viswanathan, *Physica A*, **336**, 585 (2004).
- [42] C. M. Nascimento, H. B. N. Júnior, H. D. Jennings, M. Serva, Iram Gleria and G. M. Viswanathan, *Europhysics Letters*, **81**, 18002 (2008).