

UNIVERSIDADE ESTADUAL DE MARINGÁ

PÓS-GRADUAÇÃO EM FÍSICA

Sergio de Picoli Junior

Distribuição q-exponencial de Tsallis e distribuição de Weibull: uma análise empírica

Orientador: Prof. Dr. Renio dos Santos Mendes

Dissertação de mestrado apresentada à
Universidade Estadual de Maringá para a
obtenção do título de mestre em Física.

Maringá, março de 2001.

UNIVERSIDADE ESTADUAL DE MARINGÁ
DEPARTAMENTO DE FÍSICA

Distribuição q -exponencial de Tsallis e distribuição de Weibull: uma análise empírica

Sergio de Picoli Junior

Dissertação de Mestrado
submetida ao Departamento de Física da
Universidade Estadual de Maringá

ORIENTADOR: Prof. Dr. Renio dos Santos Mendes

BANCA EXAMINADORA

Prof. Dr. Horácio O. Girotti (UFRGS)
Prof. Dr. Luiz Agostinho Ferreira (IFT)
Prof. Dr. Marcelo O. C. Gomes (IFUSP)
Prof. Dr. Paulo Teotônio Sobrinho (IFUSP)
Prof. Dr. Walter F. Wreszinski (IFUSP)

Maringá
2001

ABSTRACT

In this work we analyzed some complex systems data. Such data, related with the events frequency to each system, were adjusted to theoretical curves. We began investigating the presence of power law's in communication systems. In fact, considering the rarest events distribution (tail of the distribution) of the circulation of magazines and newspapers, of the audience of radio stations and of the cable TV subscribers, we observed a behavior that suggests a description with power law's. On the other hand, after a brief comment about Tsallis q -exponential distribution and Weibull distribution, we propose a distribution that interpolates them: the q -stretched exponential. Soon after, we analyzed data (in general, the whole interval of the distribution) coming of another systems using q -exponential, Weibull and q -stretched distributions. More specifically, we analyzed basketball baskets, highways length, cyclone victims, airports passengers, volcanos height and income due to the medicine sale. In theses fittings we noticed that not always it is clear to discern among q -exponential and Weibull distribution. With the objective of clearing up such situation, we made use of set analysis based on mistakes distributions between the data and the theoretical curves.

RESUMO

Neste trabalho analisamos dados provenientes de alguns sistemas complexos. Tais dados, que tem a ver com a frequência de eventos relacionados a cada sistema, foram ajustados à curvas teóricas. Iniciamos investigando a presença de leis de potência em sistemas de comunicação. Mais precisamente, considerando a distribuição dos eventos mais raros (cauda da distribuição) da circulação de revistas e jornais, da audiência de emissoras de rádio e dos assinantes de TV a cabo, observamos um comportamento que sugere uma descrição com leis de potência. A seguir, depois de uma breve explanação sobre a distribuição q -exponencial de Tsallis e sobre a distribuição de Weibull, propomos uma distribuição que as interpola: a q -stretched exponencial. Logo após, analisamos dados (em geral, todo o intervalo da distribuição) provenientes de outros sistemas usando as distribuições q -exponencial, de Weibull e q -stretched. Mais especificamente, analisamos as distribuições de cestas de basquete, de extensão de rodovias, de vítimas de ciclones, de passageiros em aeroportos, de altura de vulcões e de arrecadação devido à venda de remédios. Em nossos ajustes percebemos que nem sempre é claro discernir entre a q -exponencial e a distribuição de Weibull. Com o objetivo de clarear tal situação, fizemos uso de análises baseadas na distribuição de erros entre os dados e as curvas teóricas.

Dedicatória,
Aos Meus Pais: Sergio e Jurema

Agradecimentos

- Ao orientador, Dr. Renio do Santos Mendes pela realização desta dissertação. Boas idéias, entusiasmo, interesse e paciência de sua parte foram vitais para o bom andamento do trabalho. Apreciei a parceria com alguém que é destro na sua obra". (Provérbios 22:29);

- Ao professor Luis Malacarne por seu interesse e disponibilidade. Conversas sobre o tema e seu apoio técnico também contribuíram para este empreendimento;

- Ao professor Paulo Pedreira, à Akiko e demais funcionários do Departamento de Física que de alguma forma participaram, mesmo que indiretamente, neste trabalho;

- À minha família, especificamente aos meus pais, Sergio e Jurema, pelo apoio emocional e dedicação;

- Aos amigos Daniel, Frank e Márcio e aos colegas de curso, pela companhia;

- Ao Departamento de Física da Universidade Estadual de Maringá;

- Ao governo brasileiro que, por meio do CNPq, forneceu o apoio financeiro;

- Finalmente, ao nosso criador. "Digno és, Jeová, sim, nosso Deus, de receber a glória, e a honra, e o poder, porque criaste todas as coisas e porque elas existiram e foram criadas por tua vontade." (Revelação 4:11)

Conteúdo

1	Introdução	9
2	Ajuste de curvas	13
2.1	O método dos mínimos quadrados	13
2.2	O método da maximização do coeficiente de correlação linear	15
2.3	O método da máxima verossimilhança	16
3	Leis de potência	18
3.1	Introdução	18
3.2	Leis de potência e sistemas de comunicação	20
3.2.1	Revistas, jornais, rádio e TV a cabo	22
3.3	Limitações das leis de potência	24
4	Distribuição q-exponencial	28
4.1	Introdução	28
4.2	Análise dos parâmetros	30
4.3	Identificação e ajuste de uma q -exponencial	30
4.3.1	Máximo R^2 e q -exponencial	32
4.3.2	Mínimos quadrados, máxima verossimilhança e q -exponencial	33
4.4	Exemplos	35
4.4.1	Basquete	35
4.4.2	Rodovias	36
4.4.3	Ciclones	36
4.4.4	Aeroportos	36
4.4.5	Vulcões	37

4.4.6	Remédios	37
4.4.7	Significado do valor de corte (x_c)	40
4.5	q -exponencial e modelamento	43
5	Distribuição de Weibull	45
5.1	Introdução	45
5.2	Análise dos parâmetros	46
5.3	Identificação e ajuste de uma distribuição de Weibull	46
5.3.1	Distribuição de Weibull e máximo R^2	47
5.3.2	Mínimos quadrados, máxima verossimilhança e <i>stretched</i> exponenciais	47
5.4	Exemplos	50
6	Distribuição q-<i>stretched</i> exponencial	56
6.1	Introdução	56
6.2	Identificação e ajuste de uma q - <i>stretched</i> exponencial	58
6.2.1	q - <i>stretched</i> e máximo R^2	58
6.2.2	Mínimos quadrados, máxima verossimilhança e q - <i>stretched</i>	59
6.3	Exemplos	62
6.4	Distribuição q - <i>stretched</i> e modelamento	67
7	Estudo comparativo entre distribuições	69
7.1	Comparações visual e quantitativa	69
7.2	Análise de erros	78
7.3	Dados e aleatoriedade	99
8	Conclusão	102
A	Ajuste de mínimos quadrados para uma reta	106
B	Linearização e maximização de R^2	109
C	Mecânica estatística de Tsallis: uma breve introdução	114
C.1	A distribuição de Boltzmann-Gibbs	114
C.2	A distribuição de Tsallis	116

D	Generalizações da função exponencial	119
D.1	A q -exponencial	119
D.2	Outras generalizações	121
	Bibliografia	124

Capítulo 1

Introdução

Recentemente, o estudo de sistemas complexos tem despertado grande interesse na comunidade científica[1]. Em geral, tais sistemas não são de fácil descrição teórica, ou seja, difíceis de serem modelados. Por outro lado, é comum observar nestes sistemas regularidades que se manifestam na sua descrição estatística. De fato, certos dados provenientes de sistemas complexos muitas vezes podem ser ajustados por curvas simples.

Ao analisarmos o comportamento estatístico de sistemas complexos, não é difícil nos depararmos com curvas na forma de uma lei de potência, muitas vezes conhecida como lei de Pareto[2] ou lei de Zipf[3, 4]. Exemplos típicos de leis de potência podem ser encontrados no estudo de distribuições de renda[2], frequência de palavras num longo texto[3, 4], população de grandes cidades[5, 6, 7], fogo em florestas[8], fluxo de tráfego[9], citações científicas[10], guerras[11], epidemias em populações isoladas[12], frequência de sobrenomes[14], terremotos[15], entre outros.

É interessante ressaltar que leis de potência puras só seriam observadas em situações muito particulares, tais como em sistemas infinitamente grandes. Devido a isso, em geral, elas são encontradas apenas em regiões restritas dos dados. Por exemplo, a frequência de citações de artigos científicos segue aproximadamente uma lei de potência apenas num intervalo limitado da distribuição[10], ou seja, somente para os eventos mais raros (os poucos artigos mais citados). Em vista disso, torna-se interessante analisar outras distribuições. Um bom começo seria considerar distribuições que contenham as leis de potência como limite assintótico.

Nesta direção, Mandelbrot[16, 17], no contexto da frequência de palavras num longo

texto, propôs uma correção simples à lei de potência. Sua proposta, conhecida por lei de Zipf-Mandelbrot, tem como distribuição assintótica uma lei de potência. É interessante notar que nos últimos anos, devido aos trabalhos de Tsallis[18] e seus colaboradores, a lei de Zipf-Mandelbrot ressurgiu com uma nova forma. Tsallis propôs uma mecânica estatística generalizada [18, 19, 20, 21] que tem sido aplicada no estudo de sistemas não extensivos. Neste contexto, surge naturalmente uma distribuição chamada de q -exponencial (para $q=1$ temos a distribuição exponencial). A q -exponencial contém a lei de Zipf-Mandelbrot como um caso particular quando q é maior que um e torna-se uma possível extensão de tal lei quando q é menor que um.

A q -exponencial têm sido usada diretamente em ajustes de dados tais como na distribuição de citações de artigos científicos[22] e na distribuição de gols em campeonatos de futebol[23]. Além disso, outros sistemas têm sido estudados sob o enfoque da mecânica estatística de Tsallis, como por exemplo, turbulência bi-dimensional em plasma eletrônico[24], sistemas auto-gravitantes[25], sistemas granulares[26], neutrinos solares[27], difusão anômala correlacionada[28] e de Lèvy[29], espalhamento quântico[30], mapas uni-dimensionais[31], vários modelos de sistemas críticos auto-organizados[32], entre outros[20].

Outras distribuições também têm sido usadas para descrever dados provenientes de sistemas complexos. Em particular, podemos citar a distribuição de Weibull[33, 34]. Recentemente, Laherrere e Sornette[35] mostraram que a distribuição de Weibull, chamada por eles de *stretched*¹ exponencial, pode ser aplicada no ajuste de dados em diversas situações tais como na distribuição de emissões de ondas de rádio e luz das galáxias, tamanho de reservas de óleo, aglomerações urbanas, população dos países, variações nos mercados financeiros, eventos de extinção biológica, terremotos, variações de temperatura, entre outros.

Por outro lado, observamos que, em certas análises, alguns conjuntos de dados, apesar de apresentarem curvaturas bem definidas nos gráficos log-log, não se ajustam satisfatoriamente às distribuições q -exponencial e de Weibull. Analisamos, então, a possibilidade de tais dados serem ajustados por uma distribuição intermediária entre elas. Com este objetivo, propomos uma nova distribuição, por nós denominada q -*stretched* exponencial.

A análise estatística de dados provenientes de sistemas complexos envolve a escolha

¹O termo *stretched*, oriundo do idioma inglês, significa alongado. Assim sendo, o termo *stretched* exponencial, comumente empregado na literatura, denota uma função com uma cauda mais alongada que a exponencial. Usaremos neste trabalho o termo *stretched*-exponencial para designar a distribuição acumulada de Weibull.

da distribuição de frequência que melhor descreve certo sistema. Um outro passo seria encontrar valores precisos para os parâmetros da distribuição escolhida. Neste trabalho, focalizaremos nossa atenção principalmente na primeira questão, ou seja, na escolha da melhor distribuição de ajuste.

No caso das leis de potência, uma característica interessante é que elas apresentam um comportamento linear em gráficos log-log. Tal tipo de curva facilita o processo de ajuste, que pode se resumir a um ajuste linear dos dados. O método dos mínimos quadrados tem sido largamente usado, com eficiência, para o ajuste de um conjunto de dados a uma reta. Por este motivo, empregamos apenas o método dos mínimos quadrados para o nosso estudo envolvendo leis de potência. Enfatizamos também que o coeficiente de correlação linear, desde que usado corretamente, é um bom indicativo da qualidade do ajuste linear.

Em geral, ajustes de dados, cuja distribuição de frequência é não linear, não devem ser encarados de forma tão simples. Muitas vezes não é claro qual a distribuição que melhor se ajusta a tais conjuntos de dados. Com o objetivo de esclarecer melhor a questão da escolha da melhor distribuição, analisamos dados relacionados ao basquete, às rodovias, aos ciclones, aos aeroportos, aos vulcões e aos remédios. A seguir, ajustamos tais dados às distribuições q -exponencial, de Weibull e q -stretched. Neste estudo, visando uma análise comparativa mais detalhada dos ajustes, empregamos três métodos de ajuste: mínimos quadrados, máximo coeficiente de correlação linear e máxima verossimilhança. Neste contexto, empregamos também uma análise da distribuição de erros oriundos da comparação entre as curvas de ajuste e os dados. Em suma, intencionamos poder afirmar, com uma certa segurança, qual distribuição melhor se ajusta aos sistemas aqui analisados. Em particular, discernir se um conjunto de dados é melhor ajustado por uma distribuição q -exponencial ou de Weibull. Estas questões estão discutidas ao longo deste trabalho como segue.

No capítulo 2, consideramos três métodos que podem ser usados para ajustar uma curva teórica a um certo conjunto de dados. Descrevemos brevemente o método dos mínimos quadrados[36, 37], o método da máxima verossimilhança[37, 38] e um método que chamamos de máximo coeficiente de correlação linear.

Dedicamos o capítulo 3, que trata de leis de potência, à investigação do comportamento estatístico de sistemas de comunicação. Mais precisamente, analisamos a circulação de revistas e jornais, bem como a audiência de emissoras de rádio e assinantes de TV a cabo. Como leis de potência geralmente ocorrem apenas em regiões restritas, analisamos apenas

uma centena dos eventos mais raros relacionados a cada sistema. Utilizamos o método dos mínimos quadrados na análise de tais sistemas.

Por sua vez, no capítulo seguinte, considerando distribuições q -exponenciais, utilizamos os três métodos de ajuste, descritos no capítulo 2, na análise das distribuições de cestas de basquete, de extensão de rodovias, de vítimas de ciclones, de passageiros em aeroportos, de altura de vulcões e de arrecadação por marcas de remédios.

A seguir, no capítulo 5, visando uma comparação com as q -exponenciais, empregamos a distribuição de Weibull para a análise dos sistemas estudados no capítulo 4, utilizando os mesmos procedimentos.

No capítulo seguinte, introduzimos uma distribuição que interpola a q -exponencial e a distribuição de Weibull. Também, com o objetivo de fazer comparações entre ajustes, aplicamos tal distribuição no estudo do comportamento estatístico dos sistemas já analisados com a ajuda da q -exponencial e da distribuição de Weibull.

No capítulo 7, tendo em mãos uma análise estatística destes seis sistemas complexos, analisamos a qualidade dos ajustes feitos nos capítulos anteriores. Apresentamos maneiras de decidir qual distribuição se ajusta melhor a cada sistema estudado. Nesta direção, analisamos a distribuição de erros entre as curvas teóricas e os dados. Além disso, com respeito ao ajuste de curvas, discutimos brevemente a relevância do grau de aleatoriedade presente nos mecanismos de geração de dados.

No último capítulo, expomos nossas principais conclusões.

Temos também, neste trabalho, quatro apêndices. No primeiro deles, exemplificamos o uso do método dos mínimos quadrados para o caso particular da curva de ajuste ser uma reta. No apêndice B, descrevemos alguns exemplos referentes à aplicação do método da maximização do coeficiente de correlação linear. O apêndice seguinte contém uma breve revisão do formalismo da mecânica estatística de Tsallis. No apêndice D, discutimos as generalizações da função exponencial usadas neste texto.

Capítulo 2

Ajuste de curvas

De forma abreviada, expomos aqui três métodos de ajuste de dados empregados neste trabalho. Começamos descrevendo brevemente o bem conhecido método dos mínimos quadrados. A seguir, chamamos atenção para o coeficiente de correlação linear como sendo um indicativo da qualidade do ajuste de retas, sendo amplamente usado com este fim. O segundo método está baseado na linearização de curvas de forma a maximizar o quadrado de seu coeficiente de correlação linear. O método da máxima verossimilhança é o terceiro método de ajuste aqui apresentado.

2.1 O método dos mínimos quadrados

Consideremos um conjunto de dados tal que para cada x_i corresponda um y_i , com $i = 1, 2, \dots, N$. A partir destes dados, podemos fazer um diagrama de dispersão, ou seja, localizar os pontos (x_1, y_1) , (x_2, y_2) , ..., (x_N, y_N) em um sistema de coordenadas cartesianas. Frequentemente, é possível visualizar, neste diagrama, uma curva regular que se aproxime dos dados, chamada de curva de ajuste.

Um método largamente empregado como critério para indicar a melhor curva de ajuste é o método dos mínimos quadrados[36, 37], que passaremos a descrever brevemente. Os dados que queremos ajustar estão exemplificados na figura (2.1) pelos pontos (x_1, y_1) , (x_2, y_2) , ..., (x_N, y_N) , enquanto que C representa uma curva de ajuste, que por simplicidade tomamos como uma reta. Para um dado x_i , haverá uma diferença entre y_i e o

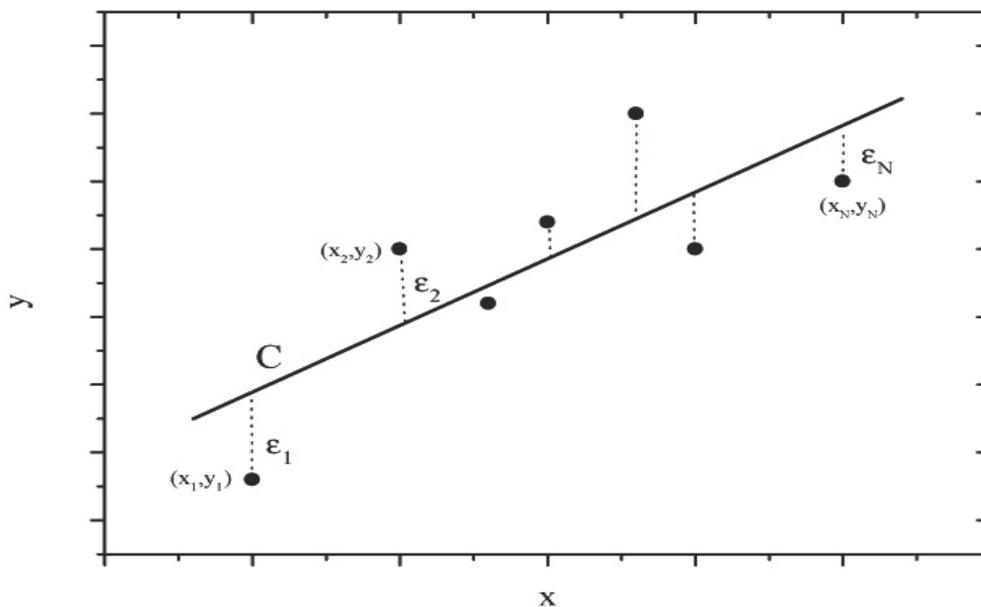


Figura 2.1: Os dados a serem ajustados estão representados pelos pontos (x_1, y_1) , (x_2, y_2) , ..., (x_N, y_N) . Por simplicidade, a reta C representa a curva de ajuste. Os valores ε_i representam a distância vertical (erros) entre os pontos e a curva de ajuste.

valor correspondente determinado na reta C . Conforme notamos na figura, tal diferença é representada por D_i .

A partir destas distâncias define-se

$$S = \sum_{i=1}^N D_i^2. \quad (2.1)$$

O método dos mínimos quadrados, basicamente, consiste em obter-se a curva, dentro de uma família de curvas pré-estabelecida, que minimiza S . Neste trabalho, ao fazermos uso deste método, utilizamos procedimentos numéricos para minimizar S . No entanto, se fizermos tal minimização analiticamente obteremos alguns resultados interessantes. No apêndice (A) minimizamos analiticamente a função S para o caso particular do ajuste por uma reta. Obtemos ainda neste apêndice uma expressão para o coeficiente de correlação linear e uma possível interpretação de seu significado.

2.2 O método da maximização do coeficiente de correlação linear

As propriedades apresentadas no apêndice (A) nos motivam a usar o coeficiente de correlação linear \mathcal{R} como sendo uma medida da qualidade do ajuste de um conjunto de dados a uma reta. De fato, o coeficiente \mathcal{R} é amplamente usado para este fim. Além disso, a definição para o coeficiente de correlação \mathcal{R} pode assumir uma forma mais geral, podendo ser empregada no estudo de relações não lineares. No entanto, mediante uma transformação adequada de variáveis, muitas vezes é de interesse linearizar tais relações. Tal procedimento é freqüentemente empregado, por exemplo, na linearização de gráficos de exponenciais e de potências usando-se escalas mono-log (semi-log) ou log-log, respectivamente.

Muitos conjuntos de dados, provenientes de sistemas complexos, podem apresentar curvaturas em gráficos log-log. O método da maximização de R^2 , que propomos aqui, está baseado na linearização destas curvas. Tal linearização é feita de modo que a reta de ajuste, obtida posteriormente por meio do método dos mínimos quadrados, tenha o máximo valor de R^2 . O apêndice (B) contém exemplos da aplicação deste método às distribuições usadas neste trabalho.

Diferentemente dos exemplos considerados no apêndice (B), em geral, dados observacionais apresentam flutuações quando comparados com sua curva de ajuste. Tal fato certamente influirá no valor do parâmetro R^2 , fazendo com que $R_{max}^2 < 1$. Sendo assim, as linearizações perfeitas exemplificadas no apêndice (B) dificilmente serão encontradas, seja qual for a natureza do sistema analisado. Assim, numa análise de ajuste de dados observacionais, obteríamos uma curva análoga àquelas apresentadas nas figuras (B.1 c), (B.2 c) e (B.3 c), porém deslocadas e com $R_{max}^2 < 1$.

Notemos também que no apêndice (B) citamos exemplos cujas relações entre x e y eram conhecidas. No entanto, nossos objetivos estão voltados para a análise de dados observacionais que, embora possamos suspeitar da forma de sua distribuição, possuem parâmetros com valores, a princípio, desconhecidos. Desta forma, o procedimento de ajuste a ser empregado tem como essência a obtenção dos parâmetros pertinentes que conduzem ao R_{max}^2 .

2.3 O método da máxima verossimilhança

Outro método, também muito conhecido, que pode ser usado para ajustar curvas a dados observacionais, é o método da máxima verossimilhança[37, 38]. Vamos supor que desejamos ajustar um conjunto de dados x_1, x_2, \dots, x_N a uma certa distribuição de frequência $P(x)$ que depende dos parâmetros $\alpha_1, \alpha_2, \dots, \alpha_k$. Por simplicidade, consideremos que haja apenas um parâmetro a ser estimado, α , por exemplo. Assim, uma vez que a distribuição $P(x)$ obedeça à condição de normalização

$$\int_{x_{min}}^{x_{max}} P(x) dx = 1, \quad (2.2)$$

podemos aplicar o método da máxima verossimilhança para estimar o valor do parâmetro α . Basicamente, tal método consiste em maximizar a probabilidade (em relação ao parâmetro envolvido) de que ocorra a seqüência x_1, x_2, \dots, x_N . Em outras palavras, o valor estimado $\tilde{\alpha}$ será aquele que maximiza a função

$$L(\alpha) = P(x_1) P(x_2) \dots P(x_N). \quad (2.3)$$

A eficiência deste método aumenta com o tamanho do sistema analisado, ou seja, quanto maior o valor de N mais confiável torna-se o método.

Para ilustrar o funcionamento deste método, poderíamos citar muitos exemplos[37]. No entanto, consideremos o ilustrativo problema de determinar a vida média τ de uma partícula instável. Muitas partículas semelhantes foram observadas experimentalmente. Observou-se que a primeira decaiu passado um tempo t_1 , a segunda num tempo t_2 e assim por diante. Queremos estimar o valor de τ para este tipo de partícula. Neste caso x , em (2.2), representa a variável t , sendo $x_{min} = 0$ e $x_{max} = \infty$.

A probabilidade de uma partícula de vida média τ decair num intervalo de tempo t a $t + dt$ é

$$P(t) dt = \frac{1}{\tau} \exp(-t/\tau) dt, \quad (2.4)$$

fornecendo então para a função L a forma

$$L(\tau) = \frac{1}{\tau} \exp(-t_1/\tau) \frac{1}{\tau} \exp(-t_2/\tau) \dots \frac{1}{\tau} \exp(-t_N/\tau) = \exp\left(-\frac{1}{\tau} \sum_{i=1}^N t_i - N \ln \tau\right). \quad (2.5)$$

Diferenciando L em relação a τ , ou equivalentemente $\ln L$, e igualando a zero, temos

$$0 = \frac{d}{d\tau} \ln L(\tau) = \frac{1}{\tau^2} \sum_{i=1}^N t_i - \frac{N}{\tau}. \quad (2.6)$$

O valor estimado

$$\tilde{\tau} = \frac{1}{N} \sum_{i=1}^N t_i, \quad (2.7)$$

é dado pela resolução de (2.6). Nota-se que neste exemplo $\tilde{\tau}$ é simplesmente a média aritmética dos tempos t_i observados. Tal resultado já era esperado, visto que a distribuição $P(x)$, neste caso, tem a forma de uma exponencial.

Neste exemplo, a equação a ser resolvida para maximizar L foi simples. No entanto, há muitos casos que, devido à complexidade da equação, torna-se necessário usar métodos numéricos, técnicas de aproximação ou mesmo soluções gráficas. Neste trabalho, temos usado um método numérico para tal tarefa.

Capítulo 3

Leis de potência

Neste capítulo, enfocamos uma das mais simples distribuições: as leis de potência. Em nossas considerações fazemos uso de distribuições acumuladas, que tem a vantagem de diminuir as flutuações oriundas dos dados. Investigamos a ocorrência de leis de potência em quatro sistemas relacionados à comunicação: revistas, jornais, rádio e TV a cabo. Para isso, utilizamos o ajuste via método dos mínimos quadrados, considerado no capítulo anterior. Expomos também algumas limitações das leis de potência, que sugerem o uso de outras formas de distribuição.

3.1 Introdução

Há cerca de um século, Pareto[2] propôs que a distribuição de renda de uma população segue uma lei simples e universal para os maiores rendimentos. Cerca de meio século mais tarde, Zipf[3, 4] observou que a distribuição de palavras de alguma linguagem natural, num longo texto, também segue uma lei empírica muito simples. Que tipo de lei foi identificada? Tratava-se de uma lei de potência,

$$P(x) = P_0 x^{-(1+\mu)} \quad (\mu > 0), \quad (3.1)$$

muitas vezes chamada de lei de Pareto ou lei de Zipf. Em (3.1), $P(x)$ representa a frequência de um evento com intensidade x . Por exemplo, no caso estudado por Pareto,

$P(x)$ é o número de pessoas com renda x , enquanto que no estudo de Zipf $P(x)$ é o número de palavras que ocorrem x vezes no texto.

Com o passar dos anos, distribuições de frequência que decaem como leis de potência vem sendo observadas em diferentes sistemas e em diversas situações, adquirindo assim um destaque especial. Por exemplo, elas têm sido usadas para descrever muitos fenômenos naturais, econômicos e sociais, tais como, população de grandes cidades[5, 6, 7], fogo em florestas[8], fluxo de tráfego[9], citações científicas[10], guerras[11], frequência de sobrenomes[14], epidemias em populações isoladas[12] e terremotos[15].

A seguir, apresentamos um procedimento que permite identificar e ajustar uma lei de potência através de uma análise estatística de um conjunto de dados. Tal análise estatística envolve o estudo da frequência com que cada elemento ocorre. Consideremos um conjunto de dados x_i , com $i = 1, 2, \dots, N$, com P_j sendo a frequência do elemento x_j (o número de vezes que x_j aparece na lista de dados).

Em geral, o gráfico de P_j por x_j apresenta flutuações em torno de uma curva lisa. Visando minimizar o efeito de tais flutuações usa-se a frequência acumulada

$$R_i = \sum_{j=1}^i P_j, \quad (3.2)$$

pois o efeito da flutuações são, em média, compensados.

Por outro lado, em muitas situações, temos $N \gg 1$ com $|x_i - x_{i-1}| \ll x_i$. Assim, com boa precisão, podemos substituir a variável discreta x_j , pela contínua x . Analogamente, empregamos $P(x)$ ao invés de P_j e

$$R(x) = \int_x^\infty P(z) dz \quad (3.3)$$

em substituição a R_j . Nas análises efetuadas neste trabalho, empregamos justamente $R(x)$ para ajustar R_j . Em particular, para o caso das leis de potência, temos

$$R(x) = \int_x^\infty P_0 z^{-(1+\mu)} dz = \frac{P_0}{\mu} x^{-\mu}. \quad (3.4)$$

Tomando-se o logaritmo em ambos os lados de (3.4) obtém-se

$$\ln R = \ln \frac{P_0}{\mu} - \mu \ln x. \quad (3.5)$$

De acordo com a equação (3.5), se um gráfico log-log de $R(x)$ por x apresentar-se como uma reta, isso indicaria uma relação com leis de potência. Esse comportamento pode ser facilmente visualizado nas figuras (3.1 a) e (3.1 b). Nestas figuras temos exemplos de leis de potência na sua forma acumulada para valores particulares de μ ($\mu = 1, 0, \mu = 1, 5$ e $\mu = 2, 0$). Logo após, vemos o mesmo gráfico numa escala log-log onde visualizamos claramente o comportamento linear.

Usamos o método dos mínimos quadrados, considerado no capítulo anterior, para o ajuste de dados às leis de potência. Como as leis de potência se apresentam como retas em gráficos log-log de $R(x)$ por x , tal ajuste se torna semelhante ao exemplo resolvido analiticamente no apêndice (A). A única diferença reside no fato de que no caso das leis de potência usamos gráficos log-log. Temos então que $D_i = (A_0 + A_1 \ln x_i) - \ln R_i$ é a diferença entre o logaritmo da curva teórica e o dado observacional $\ln R_i$. Portanto, os valores de A_0 e A_1 que minimizam a função

$$S = \sum_{i=1}^N D_i^2 = \sum_{i=1}^N [(A_0 + A_1 \ln x_i) - \ln R_i]^2, \quad (3.6)$$

são os valores estimados ou de ajuste fornecido pelo método dos mínimos quadrados. Os parâmetros P_0 e μ da distribuição são obtidos, de acordo com a equação (3.5), pelas equações

$$\mu = -A_1 \quad (3.7)$$

e

$$P_0 = -A_1 \exp(A_0). \quad (3.8)$$

3.2 Leis de potência e sistemas de comunicação

Além dos sistemas já citados na seção anterior, é possível estabelecer uma relação entre sistemas de comunicação e leis de potência.

No fim do século *XIX*, a mecanização das impressoras e o desenvolvimento do rádio passaram a ser utilizados amplamente para disseminar notícias. Logo a televisão também surgiu como um poderoso veículo de transmissão de notícias. Sistemas de TV a cabo e

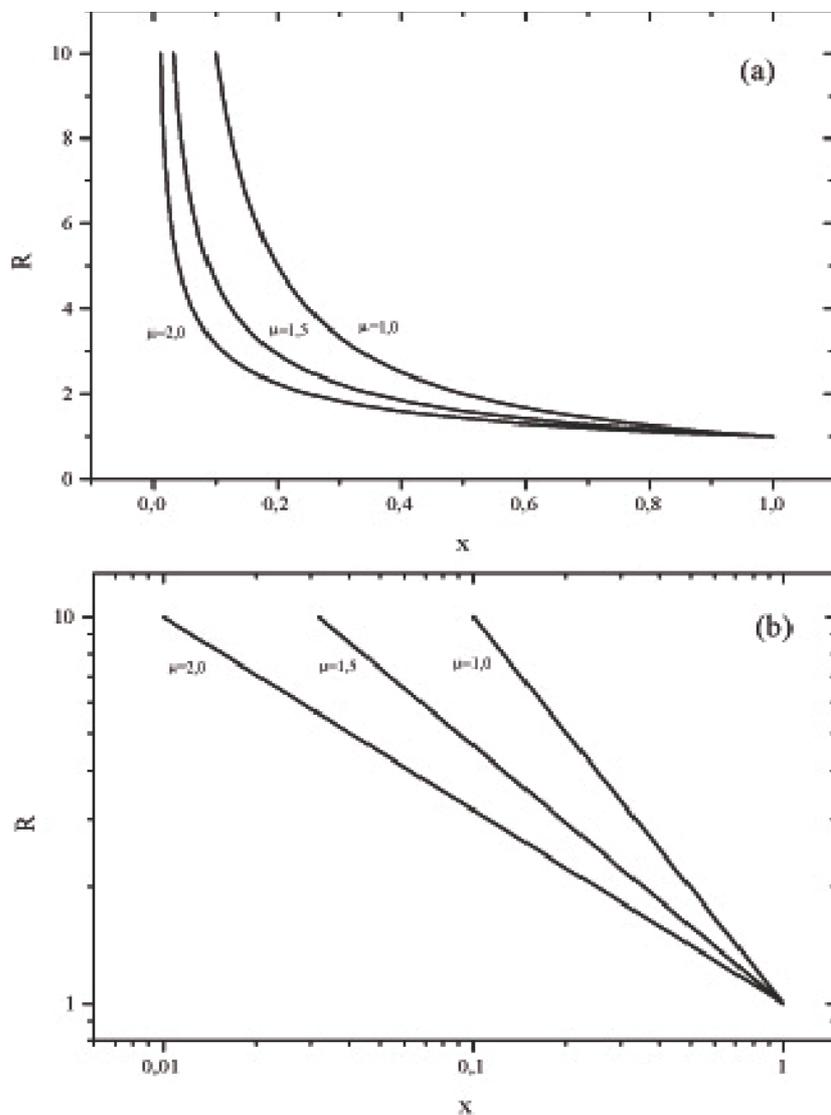


Figura 3.1: (a) - Gráfico de $R(x)$ por x de uma lei de potência $R(x) = (P_0/\mu)x^{-\mu}$ ($P_0/\mu = 1$) para alguns valores de μ ($\mu = 1, 0, 1, 5$ e $2, 0$). (b) - Gráfico anterior na escala log-log.

satélite oferecem agora canais de televisão que transmitem sua programação a um crescente número de assinantes. De uma maneira geral, não resta dúvida de que tais veículos noticiosos exercem cada dia mais influência na vida das pessoas. Neste sentido, questões relacionadas a sistemas de comunicação ocupam uma posição de destaque no contexto atual de nossa sociedade.

3.2.1 Revistas, jornais, rádio e TV a cabo

A indústria dos meios de comunicação (e propaganda) tem se tornado um campo fértil para pesquisas[39]. Diversos estudos já foram feitos utilizando-se dados provenientes de sistemas de comunicação relacionados com revistas[40], jornais[41], rádio[42] e TV a cabo[43]. Em nossa análise, entretanto, focalizamos a atenção na distribuição das revistas de maior circulação, dos jornais mais lidos, das emissoras de rádio mais ouvidas e dos sistemas de TV a cabo mais assinados.

Analizamos, de certa forma, os dados mais relevantes, possivelmente os responsáveis pela maior parte da influência sobre o público. Tais dados contêm o comportamento de cauda da distribuição de frequência relacionada ao sistema em estudo. Nos gráficos usados para essa análise, as variáveis de interesse, tais como a circulação e a audiência x , estão em função de sua distribuição acumulada $R(x)$.

Vemos na figura (3.2), a distribuição acumulada das cem revistas de maior circulação nos Estados Unidos[44] no ano de 1999. Tais dados se referem a uma circulação média de seis meses, referente ao último semestre de 1999.

Na figura (3.3), podemos visualizar a distribuição acumulada dos jornais diários com maior circulação em caráter mundial[45] no ano de 1999. Tais dados também estão relacionados a uma circulação média de seis meses terminando em setembro de 1999.

A distribuição acumulada das emissoras de rádio mais ouvidas pelos americanos[46], no período de abril a outubro de 2000, pode ser vista na figura (3.4). Os dados se referem a uma taxa relacionada ao número máximo de ouvintes durante o período da pesquisa. Visto não ser o número real de ouvintes, mas apenas uma taxa, multiplicamos esses dados por um fator de 1000 para melhor comparação entre os sistemas.

A distribuição acumulada dos cem maiores sistemas de TV a cabo dos Estados Unidos[47], em número de assinantes, podem ser vistos na figura (3.5). Os dados se referem ao mês

de abril de 2000.

Listamos, na tabela (3.1), além dos valores dos parâmetros obtidos pelo ajuste dos mínimos quadrados, valores de R^2 (coeficiente de correlação linear) e S/M (média da soma dos desvios ao quadrado), onde M é o número de pontos num gráfico $R(x)$ por x .

Sistema	μ	P_0	S/M	R^2
Revistas	1,25	$1,57 \times 10^4$	$1,630 \times 10^{-3}$	0,9907
Jornais	1,50	$5,93 \times 10^4$	$1,632 \times 10^{-3}$	0,9912
Rádio	1,51	$7,28 \times 10^3$	$3,033 \times 10^{-3}$	0,9862
TV a Cabo	2,02	$3,90 \times 10^5$	$2,430 \times 10^{-3}$	0,9850

Tabela (3.1) - Resultados obtidos do ajuste dos dados às leis de potência, sendo que R^2 e S/M fornecem indicações sobre a qualidade do ajuste.

Pode-se ver na figura (3.6 a) a distribuição acumulada dos quatro sistemas analisados neste capítulo. Além disso, vemos na figura (3.6 b) uma normalização desses dados. Dividindo x por seu valor médio m , é possível uma melhor comparação entre os sistemas. Estes gráficos sugerem, com certa aproximação, um comportamento universal para os diferentes sistemas de comunicação estudados. Este comportamento universal é mais pronunciado nos casos da circulação de revistas ($\mu \simeq 1,3$), circulação de jornais ($\mu \simeq 1,5$) e audiência de rádio ($\mu \simeq 1,5$). Nota-se que nestes casos os valores de μ assumem aproximadamente os mesmos valores.

No entanto, este comportamento universal não é tão bem compartilhado pela distribuição dos assinantes de TV a cabo ($\mu \simeq 2,0$). Uma possível fonte de discrepância entre o expoente $\mu \simeq 2,0$ e os demais têm a ver com o relativamente alto grau de seletividade econômica das assinaturas de TV a cabo. Sem nos atermos a uma discussão mais detalhada sobre o grau de universalidade dos sistemas de comunicação aqui estudados, passaremos a analisar algumas limitações gerais das leis de potência.

3.3 Limitações das leis de potência

Quanto à origem, muitos mecanismos podem levar a distribuições de lei de potência [48, 49, 50]. No entanto, tais leis são usualmente observadas de forma assintótica na natureza. Leis de potência, em seu pleno sentido, só seriam observadas em condições ideais tais como em situações com infinito controle de temperatura, simulações computacionais infinitamente grandes, ao esperar um tempo infinito para se obter o equilíbrio e ao desprezar os efeitos de gravidade. Estritamente falando, tais condições ideais não existem na natureza, portanto, num sistema finito uma descrição com leis de potência geralmente deve ser considerada como uma aproximação. Assim, curvaturas num gráfico log-log podem naturalmente ocorrer. Uma questão de interesse é identificar o motivo dessas curvaturas, se são apenas efeitos de tamanho finito ou se estão relacionadas com toda a distribuição, tornando necessário considerar outra distribuição de probabilidades. Portanto, torna-se conveniente analisarmos correções a esta lei.

Além disso, um aspecto importante que pode inviabilizar a aplicação de uma distribuição do tipo lei de potência é considerarmos que $x \in [0, \infty)$. Isto ocorre porque a integral $\int_0^\infty P(x)dx$ diverge, manifestando uma proliferação incontrolável de dados quando x tende a zero. Tal fato reforça a idéia de considerarmos distribuições alternativas às leis de potência.

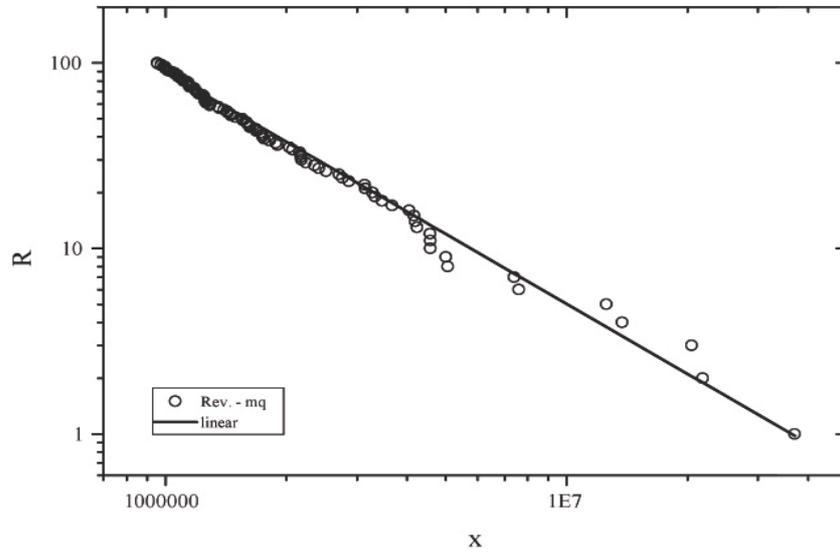


Figura 3.2: Gráfico log-log de $R(x)$ pelo número x de revistas que circularam nos EUA ano de 1999. O ajuste de mínimos quadrados fornece $A_0 = 9,439$ e $A_1 = -1,25$ que nos permite calcular os valores $\mu = 1,25$ e $P_0 = 15.710$.

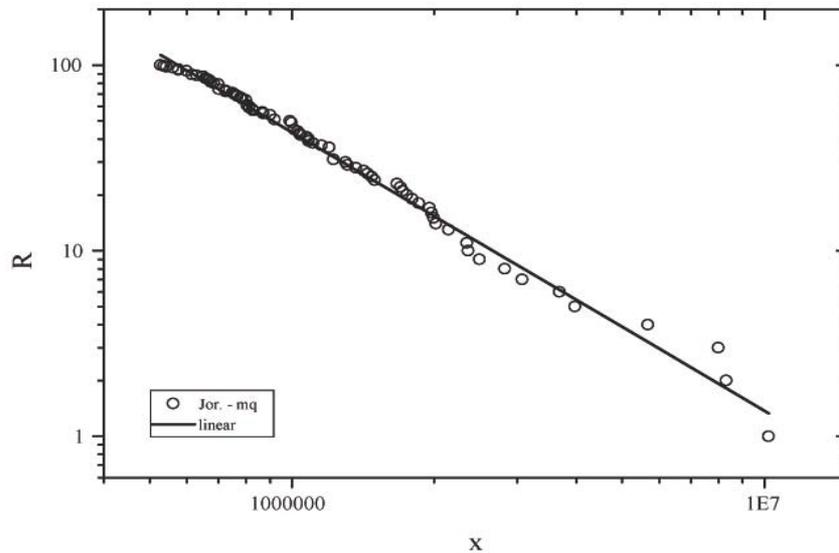


Figura 3.3: Gráfico log-log de $R(x)$ pelo número x de jornais que circularam em caráter mundial em 1999. Neste caso, ajuste de mínimos quadrados nos dá $A_0 = 10,585$ e $A_1 = -1,50$ e conseqüentemente $\mu = 1,50$ e $P_0 = 59.310$.

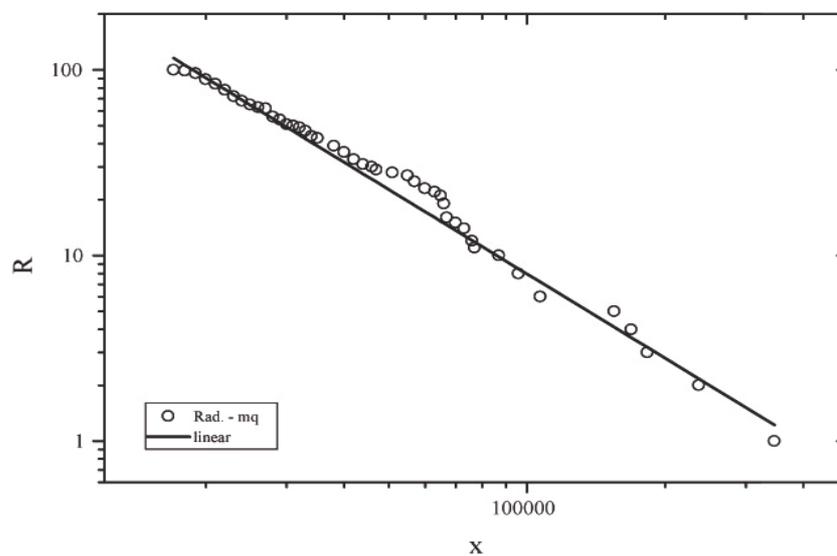


Figura 3.4: Neste caso, temos o gráfico log-log de $R(x)$ por x , onde x é uma taxa relacionada ao número máximo de ouvintes de emissoras de rádio nos EUA. O ajuste nos dá $A_0 = 8,481$ e $A_1 = -1,51$, fornecendo $\mu = 1,51$ e $P_0 = 7.280$.

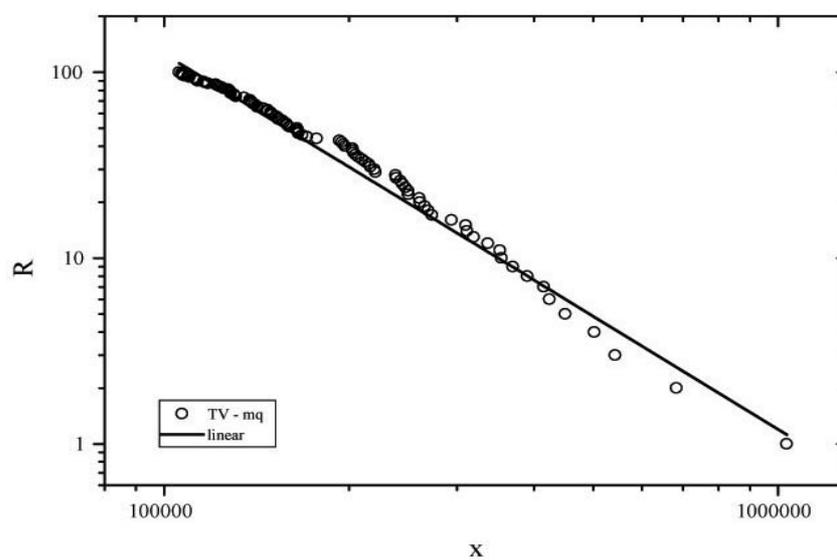


Figura 3.5: Aqui temos o gráfico log-log de $R(x)$ por x , onde x é o número de assinantes de sistemas de TV a cabo nos EUA. Neste caso, $A_0 = 12,170$ e $A_1 = -2,02$, permitindo calcular $\mu = 2,02$ e $P_0 = 389.690$.

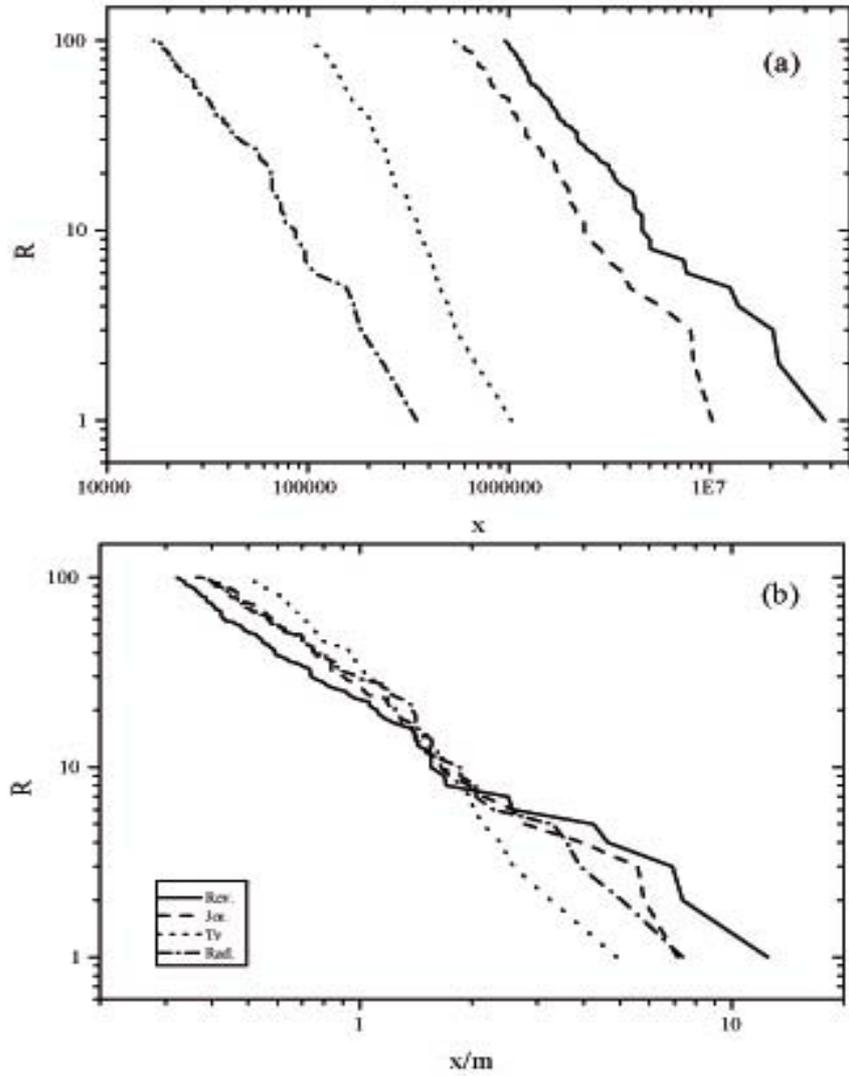


Figura 3.6: (a) - O gráfico log-log de $R(x)$ por x dos quatro sistemas analisados. Os valores de μ estão no intervalo $1,3 \leq \mu \leq 2,0$. (b) - O mesmo gráfico anterior, porém com os dados normalizados. Substituiu-se x por x/m , onde $m = 1/N \sum_{i=1}^N x_i$ é a média aritmética de x .

Capítulo 4

Distribuição q -exponencial

Visando investigar correções às leis de potência, consideramos aqui a distribuição q -exponencial, que surge no contexto da mecânica estatística não extensiva de Tsallis. A q -exponencial é equivalente à lei de Zipf-Mandelbrot, no caso q maior que um, sendo uma possível extensão de tal lei, no caso q menor que um. Discutimos o significado dos seus parâmetros e investigamos possíveis relações entre q -exponenciais e distribuições de cestas de basquete, extensão de estradas, vítimas de ciclones, movimentação de aeroportos, altitude de vulcões e arrecadação por marcas de medicamentos. Finalmente, com respeito ao modelamento de sistemas, verificamos que a equação diferencial cuja solução é uma q -exponencial assume uma forma simples.

4.1 Introdução

Como já dissemos, Zipf[3, 4] fez observações interessantes sobre linguística há algumas décadas. Ele mostrou que se ordenarmos as palavras de um longo texto de acordo com a frequência com que aparecem, obteremos, com certa precisão, uma lei de potência. Mais tarde, Mandelbrot argumentou que tal comportamento poderia ser um reflexo de algum tipo de fractalidade intrínseca do sistema e sugeriu uma correção à lei de potência[16, 17]. Essa correção ou generalização,

$$p(x) = \frac{c}{(1 + dx)^b}, \quad (4.1)$$

é usualmente conhecida como lei de Zipf-Mandelbrot, onde b , c e d são constantes reais positivas.

Como passaremos a discutir, a lei de Zipf-Mandelbrot surge também em outro contexto. Nos últimos anos, tem-se aumentado o interesse no estudo de sistemas não extensivos. Nesta direção, uma contribuição pioneira foi dada por Tsallis[19] ao introduzir uma mecânica estatística generalizada não extensiva que contém a mecânica estatística usual como um caso particular. Uma breve apresentação de seu formalismo pode ser vista no apêndice (C). No contexto da mecânica estatística não extensiva de Tsallis, surge naturalmente uma generalização da distribuição canônica. Baseando-se nesta generalização, pode-se definir uma função exponencial generalizada (q -exponencial):

$$e_q^x = \exp_q(x) \equiv [1 + (1 - q)x]^{\frac{1}{1-q}}, \quad (4.2)$$

sendo q um parâmetro real. Algumas de suas propriedades são discutidas no apêndice (D). Nota-se que uma simples redefinição das constantes em (4.1) leva à forma (4.2), quando $q > 1$. De fato, a distribuição de Zipf-Mandelbrot pode ser escrita como

$$p(x) = c \exp_q(-ax) = c[1 - (1 - q)ax]^{\frac{1}{1-q}} \quad (4.3)$$

e, portanto, por simples comparação, as constantes da equação (4.1) podem ser relacionadas com esta outra forma do seguinte modo:

$$b = \frac{1}{q-1} \quad \text{e} \quad d = -\frac{a}{q-1}. \quad (4.4)$$

Deve-se lembrar que, por definição, as constantes b e d da lei de Zipf-Mandelbrot são positivas implicando em $q > 1$. Assim, a forma (4.3) vai mais além que a equação (4.1), pois admite também o caso $q < 1$.

Observemos também que a distribuição q -exponencial tem sido empregada no estudo de uma variedade de sistemas físicos, tais como turbulência bi-dimensional em plasma eletrônico[24], sistemas auto-gravitantes[25], sistemas granulares[26], neutrinos solares[27], difusão anômala correlacionada[28] e de Lévy[29], espalhamento quântico[30], mapas uni-dimensionais[31], vários modelos de sistemas críticos auto-organizados[32] entre outros[20]. Em particular, a distribuição q -exponencial tem sido usada diretamente para o ajuste de conjuntos de dados, fornecendo bons resultados. Dois exemplos recentes

são a distribuição de gols em campeonatos de futebol[23] e a distribuição de citações de artigos científicos[22].

4.2 Análise dos parâmetros

Conforme vemos na equação (4.3), dois parâmetros de especial importância na distribuição q -exponencial são a e q , sendo c apenas uma constante de normalização. No apêndice (D), vimos que para $q \rightarrow 1$ a q -exponencial torna-se uma exponencial usual. Neste caso especial ($q = 1$), o parâmetro a dita quão rapidamente a q -exponencial (exponencial) decai, característica esta que permanecerá inalterada para $q \neq 1$. Por sua vez, o parâmetro q dita o quanto a distribuição se desvia de uma exponencial. Por exemplo, o caso $q > 1$ funciona como um anti-amortecimento, no sentido de fazer com que a distribuição decaia mais vagarosamente que a exponencial. Neste caso, para $x \gg [a(1 - q)]^{-1}$ o comportamento da distribuição aproxima-se muito de uma lei de potência, $P(x) \simeq c[a(q - 1)]^{1/(1-q)} x^{1/(1-q)}$. Por outro lado, o parâmetro $q < 1$ faz com que a distribuição decaia mais rapidamente que a exponencial, sendo necessário um corte em $x = [a(1 - q)]^{-1}$. Esses aspectos da q -exponencial estão devidamente ilustrados no apêndice (D). A partir dessas observações, deve estar claro que a forma de decaimento da distribuição q -exponencial (4.3) tem duas contribuições: a que vem de a e a relacionada com q .

4.3 Identificação e ajuste de uma q -exponencial

Visando analisar conjuntos de dados, empregaremos a forma acumulada da distribuição q -exponencial, isto é,

$$\begin{aligned} R(x) &= \int_x^\infty c \exp_q(-az) dz = \int_x^\infty c [1 - (1 - q)az]^{-\frac{1}{1-q}} dz \\ &= \frac{c}{a(2 - q)} [1 - (1 - q)ax]^{-\frac{2-q}{1-q}} = c' \exp_{q'}(-ax), \end{aligned} \quad (4.5)$$

onde

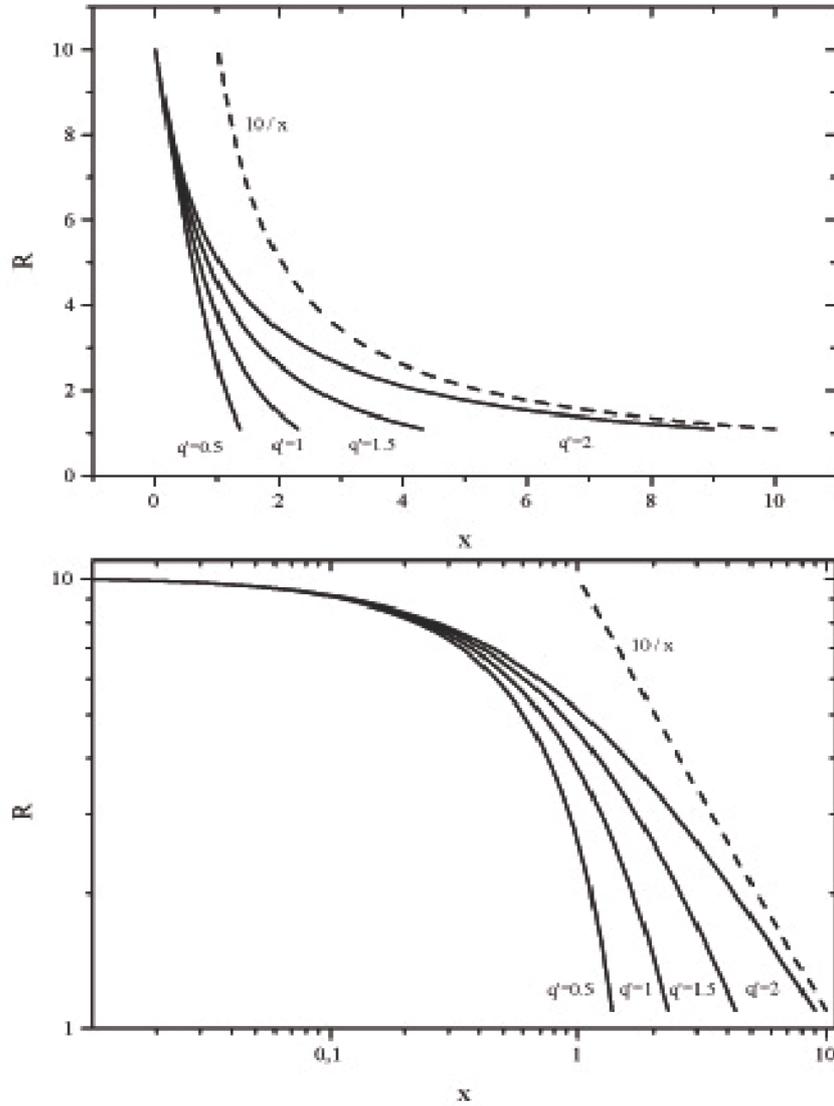


Figura 4.1: (a) - Gráfico $R(x)$ por x de uma distribuição acumulada q -exponencial: $R(x) = 10 \exp_{q'}(-x)$, com $q' = 0,5, 1, 1,5$ e 2 . A lei de potência $R(x) = 10/x$ (linha tracejada) representa a função $10 \exp_2(-x)$, quando $x \gg -1$. (b) - Gráfico anterior na escala log-log.

O comportamento desta forma acumulada da q -exponencial, para valores particulares de q' , pode ser visto na figura (4.1). Como podemos notar, q -exponenciais não tem o mesmo comportamento que as leis de potência, dificultando sua fácil identificação num gráfico log-log. A seguir, consideramos aplicações da distribuição acumulada q -exponencial empregando os métodos de ajuste descritos no capítulo 2.

4.3.1 Máximo R^2 e q -exponencial

Primeiramente, voltemos nossa atenção ao método da maximização de R^2 . Tomando-se o q -logaritmo (apêndice D) da equação (4.5), isto é,

$$\ln_{q'} R = \ln_{q'} [c' \exp_{q'} (-a'x)], \quad (4.6)$$

obtém-se, usando a relação (D.4),

$$\begin{aligned} \ln_{q'} R &= \ln_{q'} c' + \ln_{q'} [\exp_{q'} (-a'x)] + (1 - q) \ln_{q'} c' \ln_{q'} [\exp_{q'} (-a'x)] \\ &= \ln_{q'} c' - a' [1 + (1 - q') \ln_{q'} c'] x. \end{aligned} \quad (4.7)$$

Uma vez encontrado o parâmetro q' que maximiza o coeficiente R^2 da reta $\ln_{q'} R = A_0 + A_1 x$, usamos o método dos mínimos quadrados para obter os valores de A_0 e A_1 , que por sua vez, fornecem os valores

$$a' = \frac{-A_1}{1 + (1 - q')A_0} \quad (4.8)$$

e

$$c' = \exp_{q'}(A_0). \quad (4.9)$$

Portanto, se um determinado conjunto de dados apresentar um comportamento linear neste tipo de gráfico, isso indicaria uma relação com q -exponenciais caracterizadas pelo índice q' . Isso é ilustrado na figura (4.2), onde foram usados conjuntos de dados cuja distribuição acumulada tem exatamente a forma de uma q -exponencial com os mesmos valores particulares de q' empregados na figura (4.1). Além disso, usando-se as equações (4.3), podemos recuperar os valores de c , a e q da distribuição (4.3).

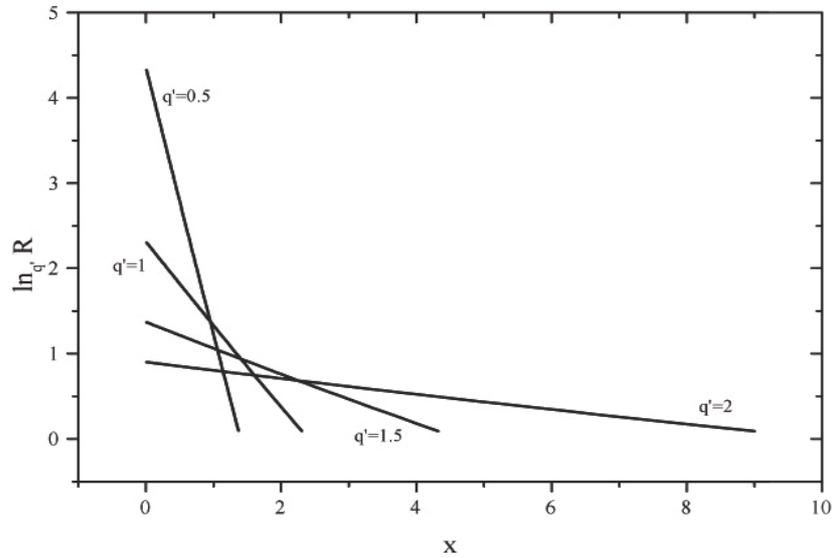


Figura 4.2: O mesmo gráfico da figura (4.1 a), porém na escala $\ln_{q'} R$ por x .

Deve estar claro, a partir do que foi exposto no parágrafo anterior, que obteremos uma reta somente se tomarmos $\ln_{q''} R$ por x , com $q'' = q'$, conforme podemos visualizar a partir da figura (4.3).

4.3.2 Mínimos quadrados, máxima verossimilhança e q -exponencial

No caso do método dos mínimos quadrados, a função S assume a forma

$$S = \sum_{i=1}^N D_i^2 = \sum_{i=1}^N [c' \exp_{q'}(-a' x_i) - R_i]^2. \quad (4.10)$$

Os valores de c' , a' e q' que minimizam esta função caracterizam a curva q -exponencial que melhor se ajusta (por este método) ao conjunto de dados analisados.

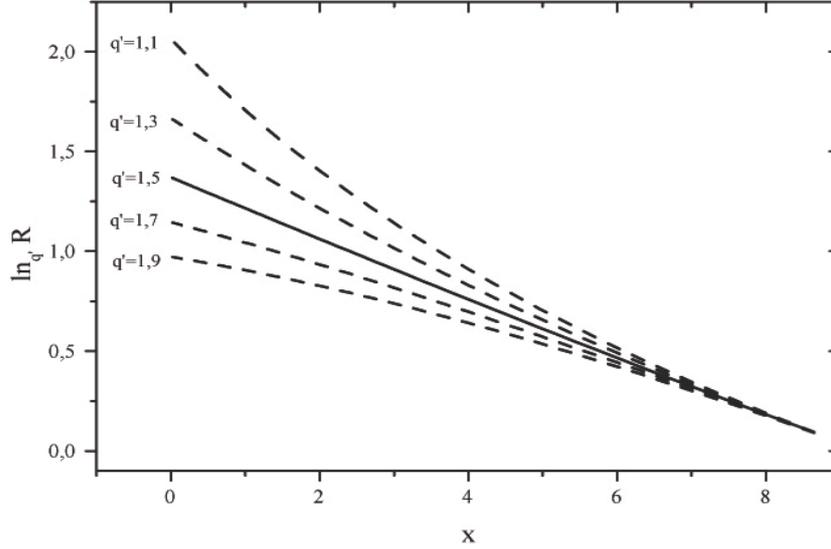


Figura 4.3: Gráfico de $\ln_{q'} R$ por x de uma distribuição acumulada q -exponencial: $R(x) = 10 \exp_{1,5}(-x)$. Dentre os diferentes valores de q'' (1, 1, 1, 3, 1, 5, 1, 7 e 1, 9), apenas $q'' = q' = 1,5$ fornece um comportamento linear.

Por fim, no caso da máxima verossimilhança, a normalização

$$\int_{x_{min}}^{\infty} n c \exp_q(-az_i) dz = 1, \quad (4.11)$$

que leva em conta a possibilidade de haver dados truncados (partindo-se de um valor de x diferente de zero) nos dá

$$n = a(2 - q) [\exp_q(-a x_{min})]^{q-2}. \quad (4.12)$$

A função L tem a forma

$$L(a, q) = n \exp_q(-ax_1) n \exp_q(-ax_2) \dots n \exp_q(-ax_N) = n^N \prod_{i=1}^N [1 - (1 - q)ax_i]^{-\frac{1}{1-q}}. \quad (4.13)$$

Tomando-se o logaritmo desta função temos

$$\ln L = N \ln n + \frac{1}{1 - q} \sum_{i=1}^N \ln[1 - (1 - q)ax_i]. \quad (4.14)$$

Assim, segundo a método da máxima verossimilhança, os valores de c , a e q que maximizam (4.14) caracterizam a distribuição q -exponencial de ajuste do sistema analisado.

4.4 Exemplos

Passamos agora a analisar a distribuição de frequência acumulada de alguns dados provenientes de sistemas complexos. Usamos os três métodos, que acabamos de discutir, para ajustar q -exponenciais a tais dados. Os parâmetros de ajuste obtidos para cada sistema estão listados na tabela (4.1). No capítulo 7 discutiremos a qualidade dos ajustes. Isto nos fornecerá uma base para decidirmos quando usar uma q -exponencial no ajuste de um conjunto de dados.

4.4.1 Basquete

Entre os sistemas citados no último parágrafo da seção (4.1), que têm sido estudados com o auxílio da distribuição q -exponencial, notamos que um deles está relacionado à distribuição de gols entre jogadores de futebol. Outros estudos, no contexto dos esportes[51, 52, 53], relacionam-se com futebol americano, golf, hóquei e basquete.

Nesta direção, nós analisamos um aspecto da famosa liga de basquete dos Estados Unidos, a NBA. Ela é composta de cerca de 400 jogadores, divididos em seus respectivos times. Durante o ano, mantém-se um registro da eficiência dos jogadores, especialmente dos pontos (cestas) que fazem[54]. Nós analisamos a distribuição do total de cestas de dois pontos durante o ano de 1999 entre todos os jogadores. Ajustamos tais dados a uma q -exponencial conforme vemos na figura (4.4). Nesta figura e nas subsequentes, mR^2 , m_q e m_v representam, respectivamente, máximo R^2 , mínimos quadrados e máxima verossimilhança.

4.4.2 Rodovias

O Sistema rodoviário de alguns países é consideravelmente significativo para sua economia. As rodovias geralmente ligam estrategicamente várias partes de um país, podendo chegar a muitos milhares de quilômetros. Acidentes em rodovias e sua relação com a extensão dos trechos onde ocorreram é apenas um exemplo dos estudos relacionados a tais sistemas[55]. Nós analisamos aqui a distribuição de extensão das 198 rodovias norte americanas[56]. O ajuste destes dados por uma q -exponencial pode ser visto na figura (4.5).

4.4.3 Ciclones

Enchentes repentinas, tais como as ocasionadas por ciclones, geralmente são consideradas como uma das catástrofes naturais mais significativas[57]. Estudos relacionados a tais fenômenos incluem a distribuição espacial e temporal de ciclones numa determinada região[58]. Além disso, na tentativa de identificar os processos climatológicos que levam à formação dos ciclones, têm-se usado dados diários fornecidos por satélites meteorológicos[59]. No entanto, torna-se difícil estudar ciclones que ocorreram há muitas décadas, quando tais tecnologias não estavam disponíveis. Dessa forma, o número de vítimas fatais poderia ser usado para obter alguma informação sobre tais desastres, mesmo podendo haver outros fatores que influenciem estes dados.

Sendo assim, nós analisamos a distribuição do número de vítimas fatais de ciclones tropicais atlânticos durante o período de 1492 a 1999[60]. Ao todo, neste período, estão catalogados cerca de 263 ciclones tropicais atlânticos com vítimas fatais. Vemos na figura (4.6) o gráfico log-log de $R(x)$ por x bem como as curvas de ajuste q -exponenciais.

4.4.4 Aeroportos

Os aeroportos podem ser considerados como vínculos essenciais da rede mundial de transporte. Eles têm assumido uma importância cada vez maior na evolução da economia global, contribuindo significativamente para o progresso da sociedade moderna. Na

tentativa de modelar tais sistemas, muitos de seus aspectos, incluindo a capacidade de grandes aeroportos, tem sido estudados[61, 62]. Em nossa investigação, analisamos os 598 aeroportos, em escala mundial, com maior capacidade[63], ou seja, aqueles pelos quais passam o maior número de passageiros. O número total de passageiros, por aeroporto, no ano de 1999, é um bom indicativo da relevância de cada um desses aeroportos. A figura (4.7) expõe os ajustes desses dados a uma q -exponencial.

4.4.5 Vulcões

O conjunto de vulcões espalhados ao redor da Terra formam um sistema complexo natural interessante. Estudos sobre emissões acústicas relacionadas a rochas vulcânicas[64], topologia, terremotos induzidos, fluxo de lava[65, 66] e avalanches em vulcões[67] detectaram a presença de leis de potência. Nossas análises envolvem outro aspecto dos vulcões: sua distribuição de altitude. Analisamos a distribuição de altitude dos 1397 vulcões[68] considerados ativos (que tiveram pelo menos uma erupção nos últimos 10.000 anos). Sua altura é dada em metros e medida a partir do nível do mar. Nesta análise, excluimos os vulcões submarinos pois estão sujeitos a mudanças mais rápidas e drásticas e muitos deles tem altura, até agora, desconhecida. Usando-se a distribuição q -exponencial obtivemos resultados que podem ser visualizados na figura (4.8).

4.4.6 Remédios

Devido ao amplo uso de medicamentos pela população em geral, a indústria farmacêutica ocupa um lugar de destaque em âmbito mundial. Em particular, ela é um grande negócio, movimentando grandes somas de dinheiro por ano. Nesta direção, estudos tem relacionado gastos farmacêuticos (na compra de remédios) com a renda dos consumidores[69]. Motivados por estes fatos, investigamos a distribuição das 200 mais rentáveis marcas de medicamentos nos Estados Unidos[70] no ano de 1999. Os ajustes destes dados com a distribuição q -exponencial podem ser vistos na figura (4.9).

Na tabela (4.1), apresentamos um resumo dos nossos ajustes empregando q -exponenciais.

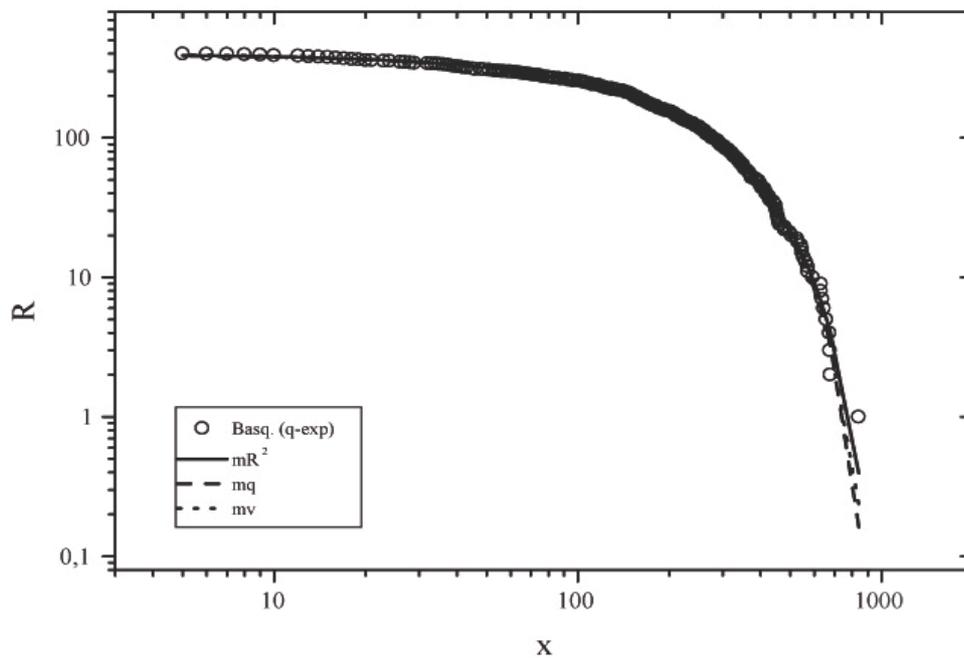


Figura 4.4: Ajuste, por uma q -exponencial, da distribuição acumulada de cestas de dois pontos em campeonatos de basquete, em 1999, na NBA.

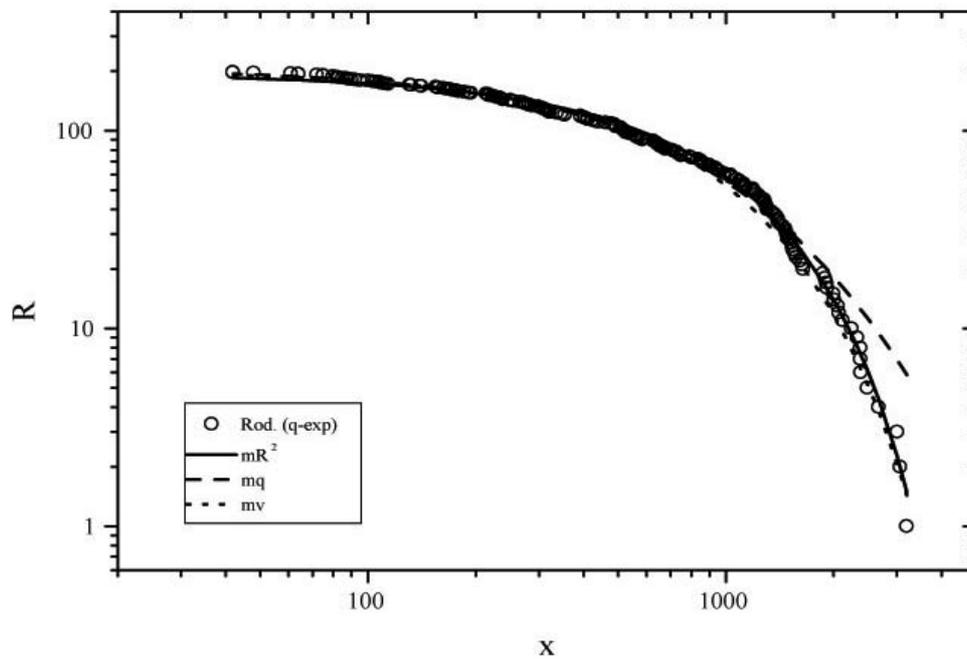


Figura 4.5: Ajuste, por uma q -exponencial, da distribuição acumulada da extensão (em milhas) das rodovias dos Estados Unidos.

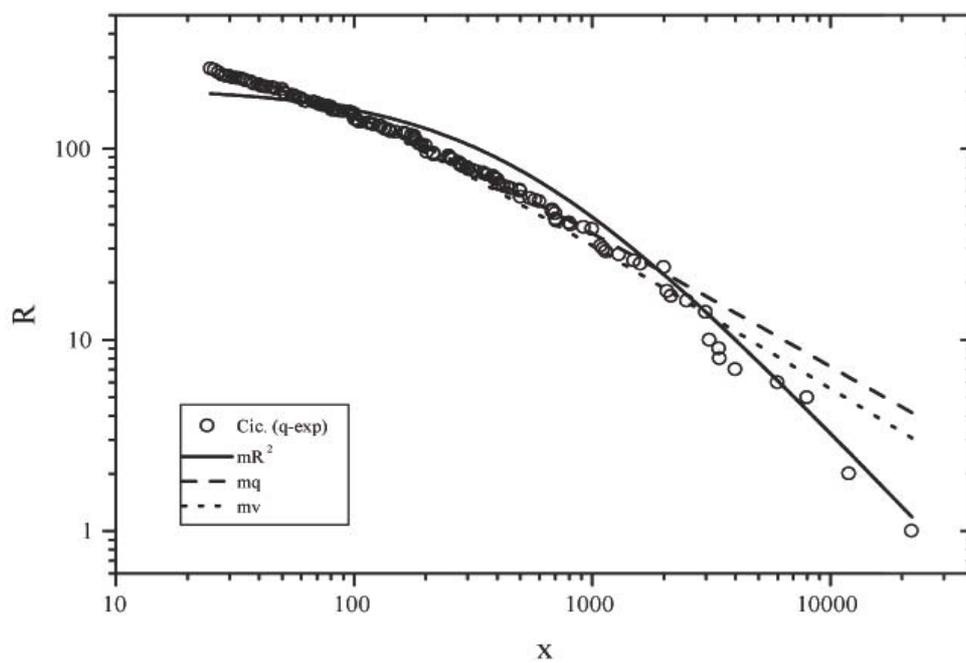


Figura 4.6: Ajuste, por uma q -exponencial, da distribuição acumulada de vítimas de ciclones.

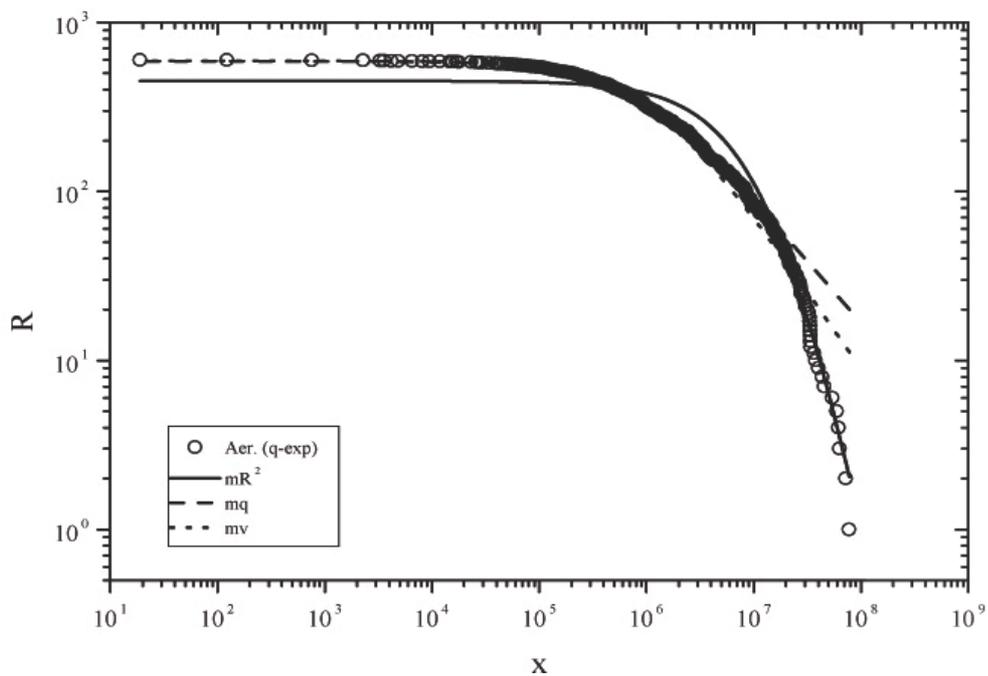


Figura 4.7: Ajuste, por uma q -exponencial, da distribuição acumulada de passageiros nos aeroportos.

Sistema	Parâmetro	mR ²	mq	mv
Basquete	q'	0,79	0,76	0,79
	x' ₀	230	238	223
	c'	399	395	- - -
Rodovias	q'	0,86	1,14	0,92
	x' ₀	908	693	784
	c'	192	206	- - -
Ciclones	q'	1,76	2,41	2,32
	x' ₀	335	60,2	47,9
	c'	182	349	- - -
Aeroportos	q'	1,30	2,35	2,08
	x' ₀	5,79 × 10 ⁶	1,11 × 10 ⁶	1,18 × 10 ⁶
	c'	405	589	- - -
Vulcões	q'	0,80	0,68	0,67
	x' ₀	2,10 × 10 ³	2,24 × 10 ³	2,54 × 10 ³
	c'	1,55 × 10 ³	1,55 × 10 ³	- - -
Remédios	q'	1,41	1,64	1,54
	x' ₀	1,72 × 10 ⁵	8,58 × 10 ⁴	9,68 × 10 ⁴
	c'	277	437	- - -

Tabela (4.1) - Valores obtidos pelos ajustes dos dados à curva q -exponencial, para os métodos: máximo R² (mR²), mínimos quadrados (mq) e máxima verossimilhança (mv).

4.4.7 Significado do valor de corte (x_c)

Entre os sistemas já citados, por exemplo, o futebol e a citação de artigos científicos, obteve-se para o parâmetro q um valor maior que um. Por outro lado, é digno de nota

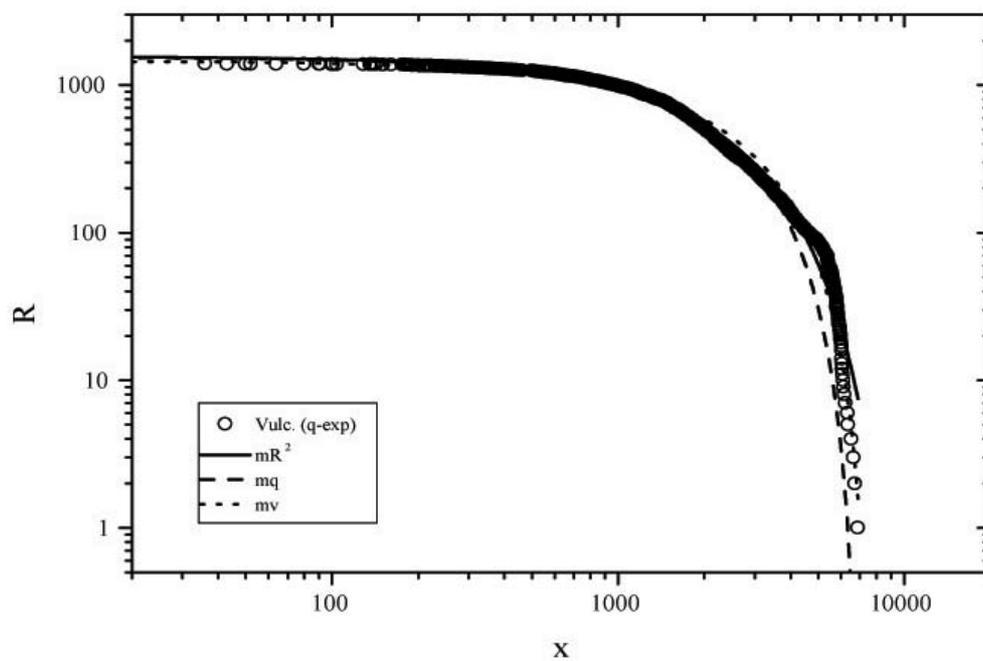


Figura 4.8: Ajuste, por uma q -exponencial, da distribuição acumulada de altura de vulcões.

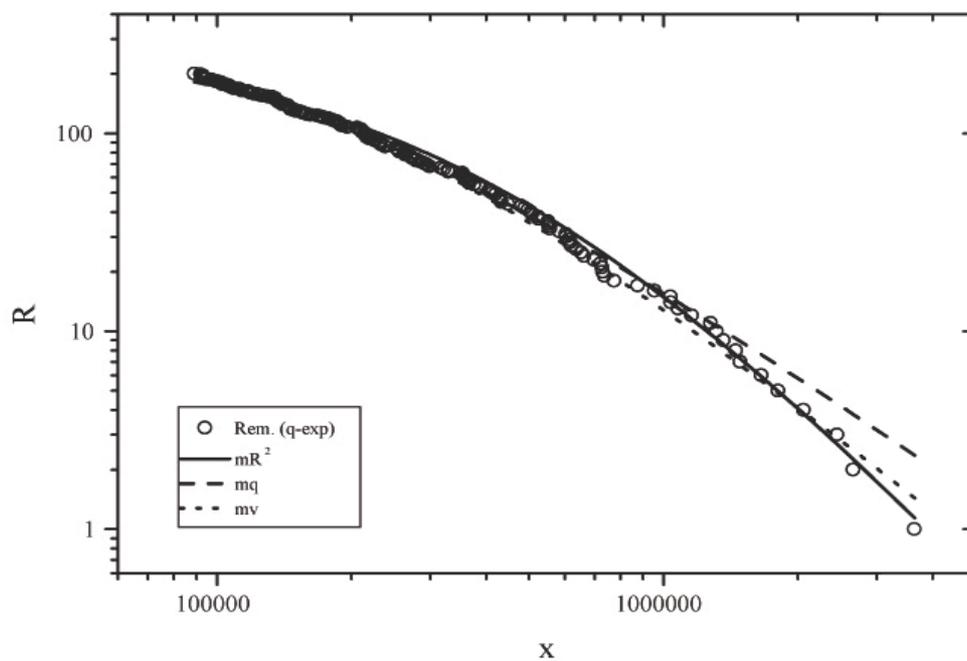


Figura 4.9: Ajuste, q -exponencial, da distribuição acumulada da arrecadação por marcas de remédios.

que os valores do parâmetro q encontrados para alguns dos sistemas aqui analisados são menores que um.

De modo geral, a distribuição q -exponencial, com $q < 1$, tem uma particularidade interessante, conforme ressaltado no apêndice (D). Tal distribuição torna-se nula a partir de um valor crítico de x . Esse corte evita que surjam probabilidades imaginárias. O valor de x crítico (x_c) é obtido a partir da equação $(1-q)ax_c = 1$. Uma vez que um determinado sistema seja realmente descrito por uma q -exponencial, x_c dá um limite máximo para os valores assumidos por x .

Conforme notamos na tabela (4.1), o ajuste da distribuição de cestas, usando-se q -exponenciais, forneceu $q < 1$. Assim, o método da maximização de R^2 , dos mínimos quadrados e da máxima verossimilhança forneceram respectivamente os valores $x_c \simeq 1.077$ ($q = 0,73$), $x_c \simeq 1084$ ($q = 0,74$) e $x_c \simeq 981$ ($q = 0,68$). Na medida em que tal afirmação for verdade, o valor de x_c deve nos dar alguma informação sobre o sistema. Estes valores nos indicariam que um jogador de basquete da NBA, por melhor ou mais eficiente que seja, dificilmente ultrapassaria durante o ano tal número de cestas de dois pontos. De alguma forma, a dinâmica deste sistema complexo, poderia limitar tal evento, tornando nula sua probabilidade de acontecer. De fato, o maior número de cestas de dois pontos alcançado, no respectivo ano, por um jogador na NBA foi de 839.

Logo após, na subseção (4.4.2), analisamos as extensões das rodovias dos Estados Unidos. Dos três métodos utilizados para o ajuste, dois deles forneceram valores de q menores que um: $q = 0,84$ do máximo R^2 ($x_c \simeq 6.600$) e $q = 0,91$ da máxima verossimilhança ($x_c \simeq 9.500$). O método dos mínimos quadrados nos deu $q = 1,12$ indicando ausência de corte. Tais valores poderiam indicar um limite máximo de extensão de uma rodovia nos Estados Unidos. A maior estrada dos Estados Unidos, de acordo com nossos dados, tem 3.205 milhas de extensão.

Já na seção (1.4.5), ao ajustarmos a distribuição de altitudes dos vulcões, obtivemos também valores de q menores que um: $q = 0,75$, $q = 0,53$ e $q = 0,52$. Tais valores nos dão $x_c \simeq 10.496$, $x_c \simeq 7.013$ e $x_c \simeq 7.833$ metros, sendo que o maior vulcão considerado em nossas análises tem 6.064 metros. Tal valor para x_c poderia indicar um limite máximo para a altura de um vulcão.

4.5 q -exponencial e modelamento

Além do ajuste de dados, modelamentos de sistemas são largamente estudados, por exemplo, no caso das populações isoladas (crescimento ou decaimento). Tem-se notado que em tais sistemas, no caso particular de não haver restrições (alimentos abundantes e ausência de predadores), a taxa de nascimento é diretamente proporcional ao tamanho da população. Isto equivale a dizer que o crescimento desta população obedece à equação

$$\frac{dy}{dx} = \tilde{a}y, \quad (4.15)$$

onde y é o tamanho da população, x o tempo decorrido e \tilde{a} é a constante de proporcionalidade. Verifica-se facilmente que a distribuição que satisfaz tal equação é uma exponencial na forma

$$p(x) = p_0 \exp(\tilde{a}x), \quad (4.16)$$

p_0 correspondendo à população em $x = 0$.

Por outro lado, se considerarmos algumas restrições para o sistema, a equação (4.15) começa a apresentar problemas, quando empregada no modelamento do crescimento da população. Devido a isso, têm-se sugerido mudanças nos modelos para acompanhar estes casos mais gerais. Por exemplo, têm-se usado equações do tipo $dy/dx = f(y)$.

No contexto deste trabalho, torna-se interessante considerarmos o caso $f(y) = \tilde{a}y^q$, isto é,

$$\frac{dy}{dx} = \tilde{a}y^q, \quad (4.17)$$

onde q é um parâmetro real. Neste caso, a taxa de crescimento é proporcional a uma potência do tamanho da população. A distribuição que satisfaz (4.17) é uma q -exponencial na forma

$$y(x) = y_0 [1 + (1 - q)ax]^{\frac{1}{1-q}} = y_0 \exp_q(ax), \quad (4.18)$$

onde y_0 e $a = \tilde{a}y_0^{q-1}$ são constantes. A definição de \tilde{a} indica que a equação (4.17) pode ser usada para descrever um crescimento ou decaimento com algum tipo de memória, onde q , de alguma forma, estaria conectado com a memória do sistema.

Um exemplo deste tipo de comportamento pode ser encontrado na cinética química[?]. A concentração de uma determinada substância A obedece a uma equação empírica para a

sua taxa de variação na forma $dC_A/dt = \tilde{k}C_A^\alpha C_B^\beta C_C^\gamma \dots$, com k representando a constante da reação e C_B e C_C sendo as concentrações das outras substâncias envolvidas. Pode-se, sob determinadas condições experimentais, considerar as concentrações C_B e C_C constantes. Assim,

$$\frac{dC_A}{dt} = \tilde{k}C_A^\alpha. \quad (4.19)$$

Comparando-se (4.17) com (4.19) vemos que $q = \alpha$, sendo α a ordem da reação relacionada à substância A . Assim, a solução de (4.19) é justamente uma q -exponencial

$$C_A(x) = C_{0A}[1 + (1 - q)kx]^{1/(1-q)} = C_{0A} \exp_q(kx), \quad (4.20)$$

com $q = \alpha$ e $k = \tilde{k}C_{0A}^{\alpha-1}$.

Portanto, quando certos sistemas apresentam alguma relação com q -exponenciais é natural considerarmos um modelamento na forma da equação (4.17).

Capítulo 5

Distribuição de Weibull

A distribuição proposta por Weibull e recentemente revisitada por Laherrere e Sornette é o objeto de enfoque deste capítulo. Investigamos o comportamento dos sistemas descritos no capítulo anterior sob este novo ponto de vista, usando os mesmos métodos de ajuste.

5.1 Introdução

Além da distribuição q -exponencial, considerada no capítulo anterior, curvaturas em gráficos de $R(x)$ por x em escala logarítmica podem assumir ainda outras formas. De fato, outros tipos de distribuição podem ser empregadas em várias análises de dados.

Recentemente, Laherrere e Sornette[35] exploraram e testaram a hipótese de que muitas curvaturas observadas nos gráficos de $R(x)$ por x , em escala logarítmica, seriam o resultado de um profundo desvio do paradigma das leis de potência, sendo preciso uma descrição alternativa para todo o domínio da distribuição. Dessa forma, foi proposto um modelo alternativo para as funções de distribuição de frequência, em termos de exponenciais alongadas,

$$P(x) = c \frac{r x^{r-1}}{x_0^r} \exp \left[- \left(\frac{x}{x_0} \right)^r \right], \quad (5.1)$$

que torna-se normalizável somente se $r > 0$, sendo que c e x_0 são parâmetros reais positivos. A distribuição (5.1) é comumente conhecida como distribuição de Weibull, devido a seus trabalhos[33, 34], de 1939.

Entre os sistemas naturais e econômicos que as distribuições de Weibull têm sido usadas para descrever, podemos citar[35] a distribuição de emissões de rádio e luz das galáxias, tamanho de reservas de óleo, aglomerações urbanas na França, nos Estados Unidos e mundiais, populações dos países, variações de mercados financeiros, eventos de extinção biológica, terremotos, variações de temperatura, entre outros. Neste capítulo, mais precisamente na seção (5.4), estudamos os sistemas analisados no capítulo 4 usando a distribuição de Weibull.

5.2 Análise dos parâmetros

Como vemos da equação (5.1), os parâmetros relevantes na distribuição são x_0 e r , sendo c apenas uma constante de normalização. Nota-se que (5.1) contém a exponencial usual como caso particular quando $r = 1$. Neste caso especial, o parâmetro x_0 dita quão rapidamente a exponencial decai. No caso $r \neq 1$ o parâmetro x_0 praticamente continuará desempenhando este mesmo papel. Além disso, é interessante observar que a distribuição (5.1) pode, em certo sentido, ser encarada como uma distribuição intermediária entre a exponencial e a lei de potência. Isto porque, se considerarmos o caso $0 < r < 1$, ela decai mais lentamente que uma exponencial e mais rapidamente que uma lei de potência, no que se refere ao seu comportamento de cauda. Na maioria das aplicações da distribuição de Weibull, tem-se usado normalmente o parâmetro r neste intervalo.

5.3 Identificação e ajuste de uma distribuição de Weibull

Visando a aplicação dos métodos de ajuste, usamos a forma acumulada da distribuição de Weibull (*stretched* exponencial), isto é,

$$R(x) = \int_x^\infty c \frac{rz^{r-1}}{x_0^r} \exp \left[- \left(\frac{z}{x_0} \right)^r \right] dz = c \exp \left[- \left(\frac{x}{x_0} \right)^r \right]. \quad (5.2)$$

O comportamento da distribuição (5.2) pode ser visualizado na figura (5.1) para valores particulares de r .

5.3.1 Distribuição de Weibull e máximo R^2

Tomando-se o logaritmo de ambos os lados de (5.2), isto é,

$$\ln R = \ln \left[c \exp \left[- \left(\frac{x}{x_0} \right)^r \right] \right], \quad (5.3)$$

obtem-se a relação

$$\ln R = \ln c - \frac{1}{x_0^r} x^r. \quad (5.4)$$

Portanto, se um determinado conjunto de dados apresentar-se como uma reta em gráficos do tipo $\ln R$ por x^r , isso indicará uma estreita relação com *stretched* exponenciais caracterizadas pelo índice r . Neste caso, $-x_0^{-r}$ é o coeficiente angular da reta e o termo independente de x fornece $R(0) = c$. Dito de outra forma, se escrevermos (5.4) na forma simplificada $\ln R = A + Bx^r$, obteremos os parâmetros x_0 e c usando-se as relações

$$c = \exp(A) \quad (5.5)$$

e

$$x_0 = \left(-\frac{1}{B} \right)^{\frac{1}{r}}. \quad (5.6)$$

Esse comportamento é ilustrado na figura (5.2), onde usamos conjuntos de dados exatamente descritos por *stretched* exponenciais. Deve estar claro, a partir destes comentários, que obteremos uma reta somente se tomarmos $x^{r'}$ com $r' = r$, fato que é ilustrado na figura (5.3). O procedimento a ser empregado no ajuste de r é o mesmo daquele usado na seção (4.3) para a q -exponencial. Mais detalhes sobre tal procedimento (a maximização de R^2) são encontrados no apêndice (B).

5.3.2 Mínimos quadrados, máxima verossimilhança e *stretched* exponenciais

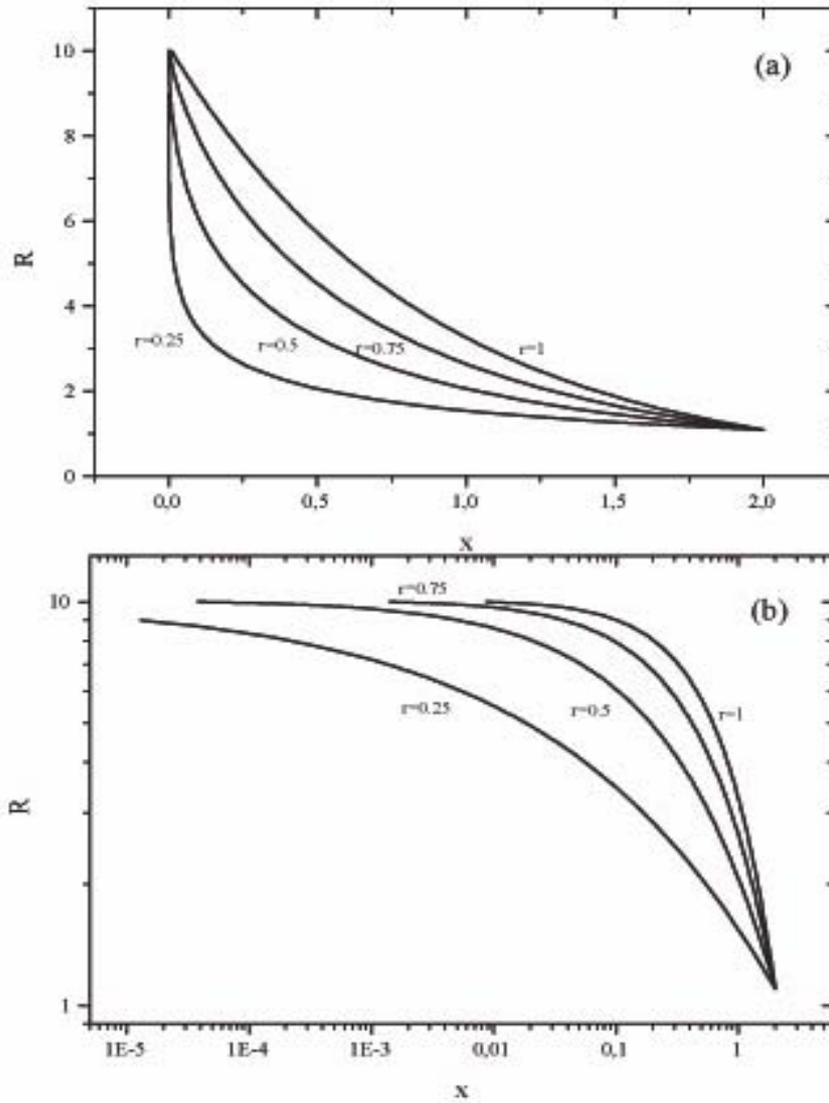


Figura 5.1: (a) - Gráfico de $R(x)$ por x de uma *stretched* exponencial: $R(x) = 10 \exp[-(x/2)^r]$, com $r = 0,25, 0,5, 0,75$ e 1 . (b) - Gráfico anterior na escala log-log.

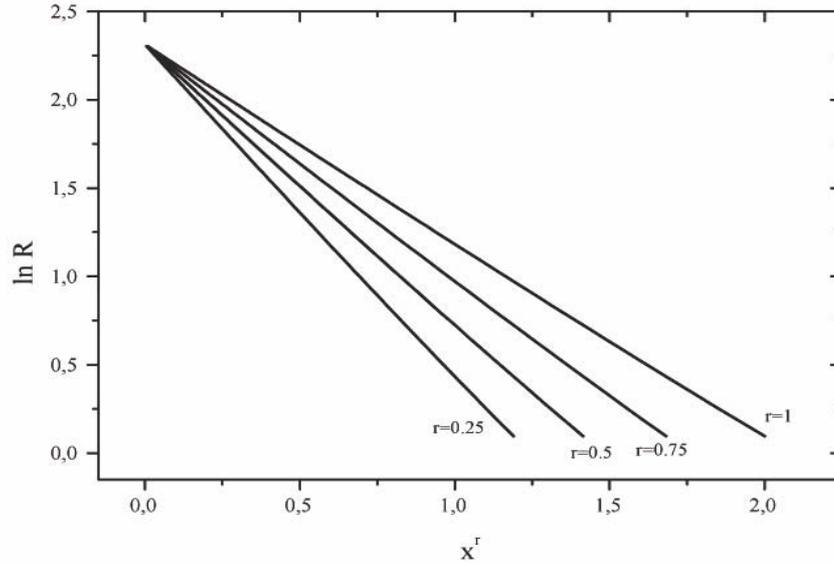


Figura 5.2: O mesmos gráficos da figura (5.1 a), porém na escala $\ln R(x)$ por x^r .

A função S , base do método dos mínimos quadrados, assume a forma

$$S = \sum_{i=1}^N D_i^2 = \sum_{i=1}^N \left(c \exp \left[- \left(\frac{x_i}{x_0} \right)^r \right] - R_i \right)^2, \quad (5.7)$$

onde R_i é a frequência acumulada observada do elemento x_i . Os valores de c , x_0 e r que minimizam esta função caracterizam a curva *stretched* exponencial que melhor (segundo este método) se ajusta ao conjunto de dados.

Por sua vez, no caso da máxima verossimilhança, a normalização da distribuição acumulada (5.2)

$$\int_{x_{min}}^{\infty} n \frac{r z^{r-1}}{x_0^r} \exp \left[- \left(\frac{z}{x_0} \right)^r \right] dz = 1, \quad (5.8)$$

que leva em conta a possibilidade de haver dados truncados (partindo-se de um valor de x diferente de zero) nos dá

$$n = \frac{1}{\exp \left[- \left(\frac{x_{min}}{x_0} \right)^r \right]}. \quad (5.9)$$

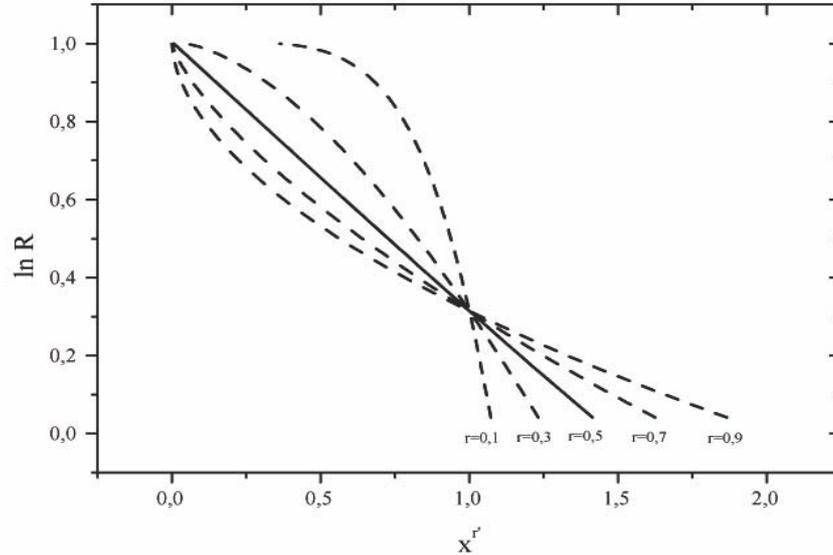


Figura 5.3: Gráfico de $\ln R(x)$ por x^r , com $R(x) = 10 \exp[-(x/2)^{0,5}]$. Nota-se que, dentre os diferentes valores de r' (0, 1, 0, 3, 0, 5, 0, 7 e 0, 9), apenas quando $r' = r = 0, 5$ obtém-se um comportamento linear.

Assim, a função L , assume a forma

$$\begin{aligned}
 L(x_0, r) &= n \frac{r x_1^{r-1}}{x_0^r} \exp \left[- \left(\frac{x_1}{x_0} \right)^r \right] n \frac{r x_2^{r-1}}{x_0^r} \exp \left[- \left(\frac{x_2}{x_0} \right)^r \right] \dots n \frac{r x_N^{r-1}}{x_0^r} \exp \left[- \left(\frac{x_N}{x_0} \right)^r \right] \\
 &= \left(\frac{n r}{x_0^r} \right)^N \exp \left[- \sum_{i=1}^N \left(\frac{x_i}{x_0} \right)^r \right] \prod_{i=1}^N x_i^{r-1}, .
 \end{aligned} \tag{5.10}$$

Os valores de c , x_0 e r que maximizam $L(x_0, r)$ caracterizam a melhor curva de ajuste (*stretched* exponencial) aos dados. Entretanto, por razões computacionais, é melhor obter c , x_0 e r tomando-se o máximo do logaritmo de L ,

$$\ln L = N \ln \left(\frac{n r}{x_0^r} \right) - \frac{1}{x_0^r} \sum_{i=1}^N x_i^r + (r - 1) \sum_{i=1}^N \ln x_i. \tag{5.11}$$

5.4 Exemplos

Visto que os sistemas que estamos analisando foram devidamente descritos no capítulo

anterior, nesta seção nos restringimos a apresentar apenas as figuras com os dados e ajustes. São elas: figura (5.4), (5.5), (5.6), (5.7), (5.8) e (5.9). Tais figuras, bem como a tabela (5.1), se referem ao ajuste dos sistemas analisados no capítulo anterior usando-se *stretched* exponenciais.

Sistema	Parâmetros	mR^2	mq	mv
Basquete	r	1,40	1,15	1,08
	x_0	236	212	190
	c	352	387	- - -
Rodovias	r	1,23	0,87	0,95
	x_0	918	712	689
	c	178	215	- - -
Ciclones	r	0,23	0,23	0,23
	x_0	4,15	4,41	3,20
	c	$1,18 \times 10^3$	$1,14 \times 10^3$	- - -
Aeroportos	r	0,49	0,51	0,57
	x_0	$2,34 \times 10^6$	$2,17 \times 10^6$	$2,98 \times 10^6$
	c	643	653	- - -
Vulcões	r	1,31	1,38	1,40
	x_0	$2,06 \times 10^3$	$2,01 \times 10^3$	$2,12 \times 10^3$
	c	$1,42 \times 10^3$	$1,43 \times 10^3$	- - -
Remédios	r	0,35	0,35	0,38
	x_0	$1,28 \times 10^4$	$1,30 \times 10^4$	$1,89 \times 10^4$
	c	$1,43 \times 10^3$	$1,41 \times 10^3$	- - -

Tabela (5.1) - Valores obtidos pelos ajustes dos dados à curva *stretched* exponencial, para os métodos: máximo R^2 (mR^2), mínimos quadrados (mq) e máxima verossimilhança (mv).

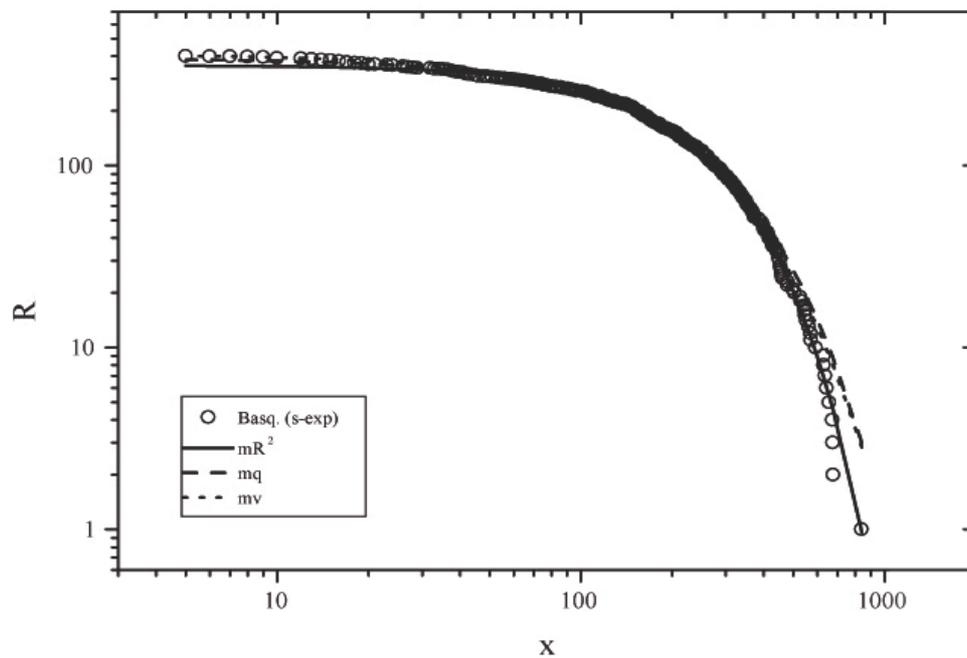


Figura 5.4: Ajuste, por uma *stretched* exponencial, da distribuição acumulada de cestas de dois pontos no basquete.

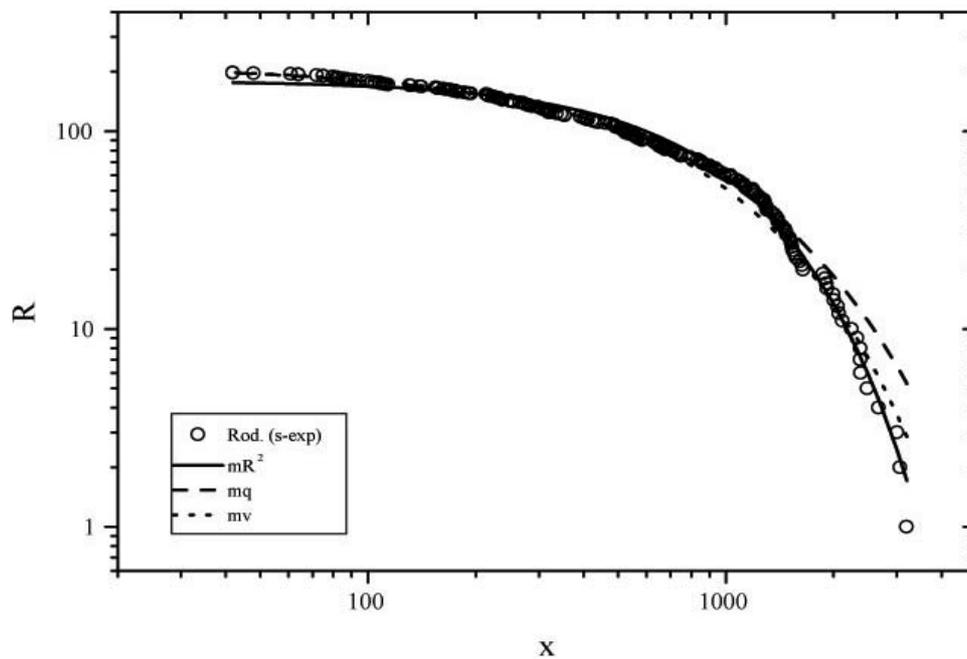


Figura 5.5: Ajuste, por uma *stretched* exponencial, da distribuição acumulada da extensão (em milhas) das rodovias dos Estados Unidos.

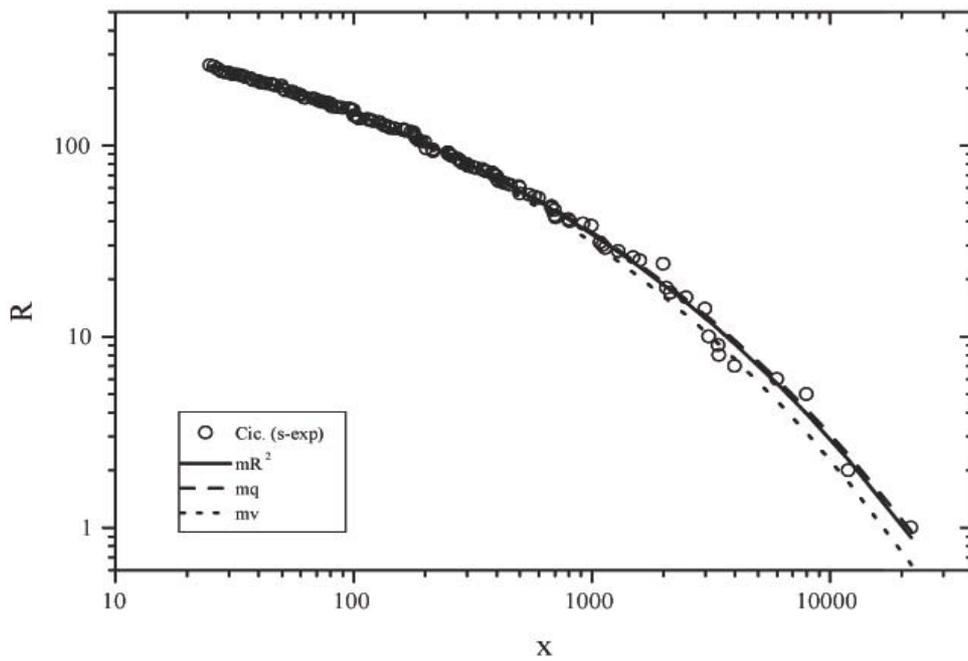


Figura 5.6: Ajuste, por uma *stretched* exponencial, da distribuição acumulada de vítimas de ciclones.

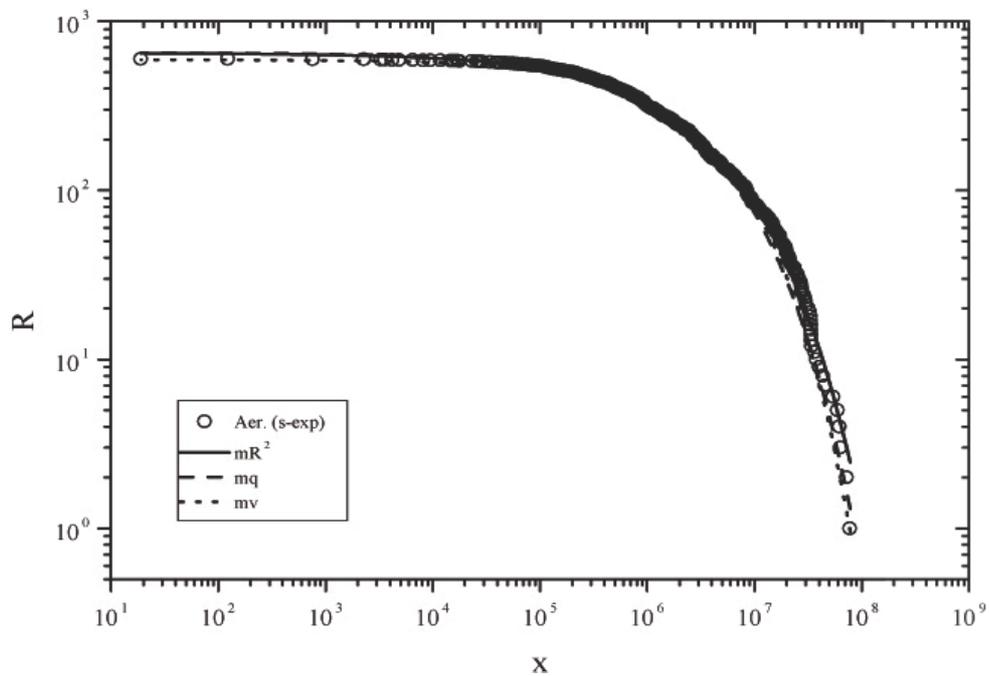


Figura 5.7: Ajuste, por uma *stretched* exponencial, da distribuição acumulada de passageiros nos aeroportos.

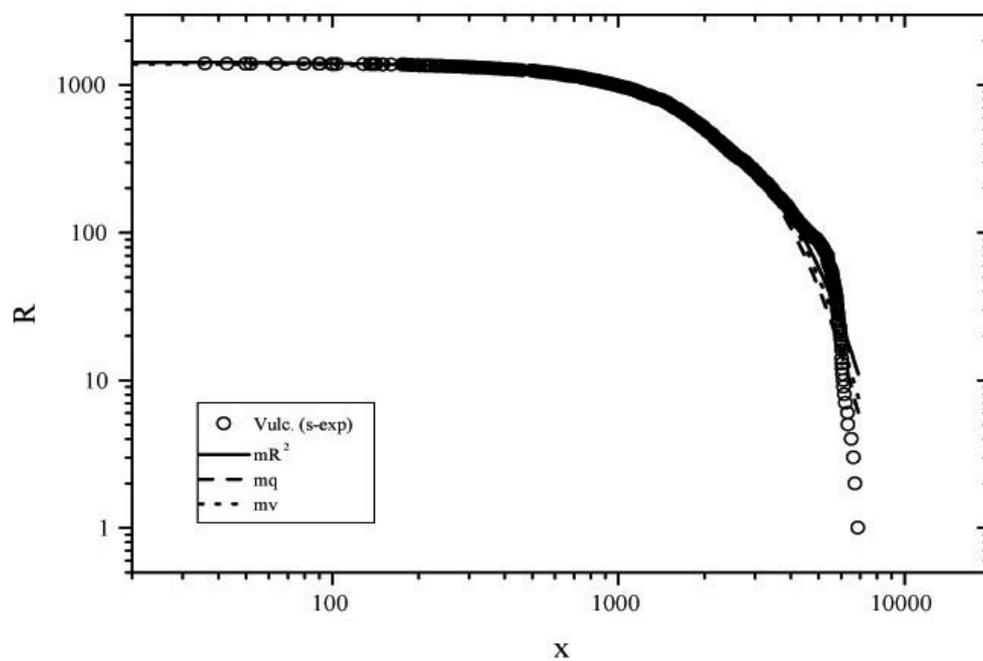


Figura 5.8: Ajuste, por uma *stretched* exponencial, da distribuição acumulada de altura de vulcões.

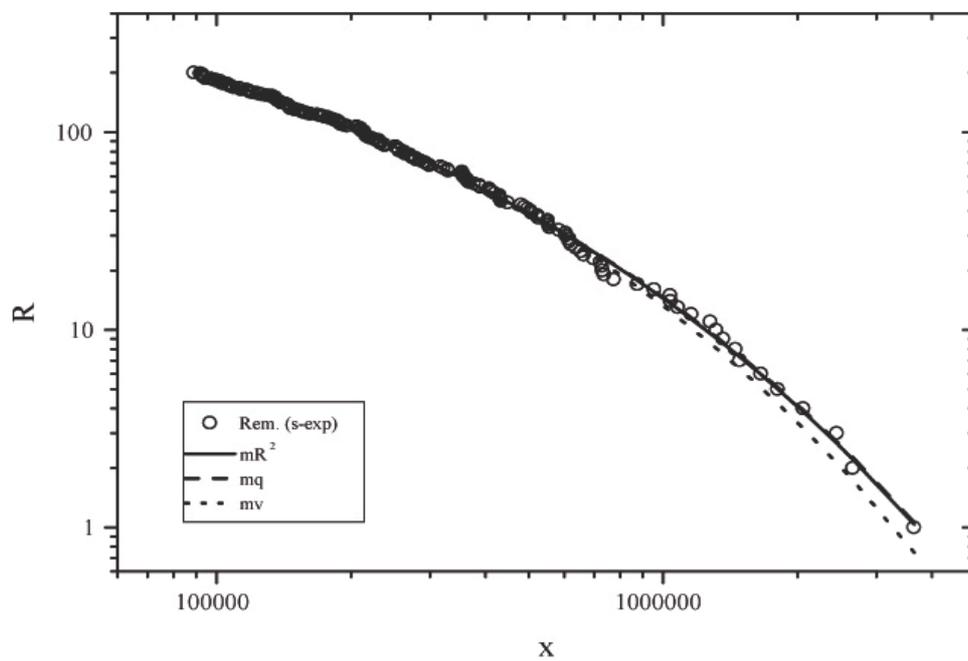


Figura 5.9: Ajuste, por uma *stretched* exponencial, da distribuição acumulada da arrecadação por marcas de remédios.

Para concluirmos, teceremos alguns comentários quanto à origem da distribuição de Weibull, indicando possíveis mecanismos para o surgimento de tais distribuições. Assim como as leis de potência, elas podem ser obtidas de um mecanismo genérico em termos de processos multiplicativos. Recentemente, Frisch e Sornette[71], desenvolveram uma teoria de desvios extremos generalizando o Teorema do Limite Central que, quando aplicada a multiplicação de variáveis aleatórias, prediz a presença genérica da distribuição de Weibull. Mais tarde, Sornette[50] verificou teoricamente que a introdução de um corte em certos processos multiplicativos realmente gera distribuições de Weibull e listou algumas de suas aplicações.

Capítulo 6

Distribuição *q-stretched* exponencial

Neste capítulo, depois de uma breve comparação entre as distribuições *q*-exponencial e de Weibull, consideramos a possibilidade de haver casos intermediários entre elas. Propomos então a distribuição *q-stretched*. Sua flexibilidade extra engloba casos intermediários e unifica as distribuições *q*-exponencial e de Weibull. De maneira análoga aos capítulos anteriores, ajustamos os dados comentados no capítulo 4 à esta nova distribuição. Apresentamos também uma breve discussão sobre *q-stretched* e modelamento.

6.1 Introdução

Conforme vimos no capítulo 3 deste trabalho, as leis de potência têm sido úteis em descrever o comportamento de muitos sistemas. No entanto, elas apresentam certas limitações sendo necessário, em geral, considerar outros tipos de distribuição.

Em particular, no capítulo 4 deste trabalho, consideramos a distribuição *q*-exponencial

$$P(x) = c \exp_q \left[- \left(\frac{x}{x_0} \right) \right], \quad (6.1)$$

onde escrevemos $a = 1/x_0$. Tal distribuição aproxima-se, no caso $q > 1$, das leis de potência num limite apropriado ($x \gg x_0/(1 - q)$), contendo a distribuição exponencial como um caso particular quando $q = 1$. No entanto, apesar de assumir um comportamento mais geral, ainda restam situações que nos induzem à análise de outros tipos de

distribuição. Nesta direção, consideramos no capítulo 5 a distribuição de Weibull

$$P(x) = c \frac{rx^{r-1}}{x_0^r} \exp \left[- \left(\frac{x}{x_0} \right)^r \right]. \quad (6.2)$$

Entre suas particularidades, notamos que ela também contém a distribuição exponencial como um caso particular quando $r = 1$.

É digno de nota que, mesmo fazendo-se uso das distribuições (6.1) e (6.2), existem sistemas que fogem a tais descrições, dando a impressão de estarem numa situação intermediária. Diante desta situação, surgem as seguintes questões: haveria alguma distribuição que pudesse descrever satisfatoriamente tais supostas situações intermediárias? Em caso afirmativo, qual seria sua forma? O objetivo básico deste capítulo é justamente introduzir uma distribuição com tal propriedade.

Neste trabalho nós propomos a distribuição

$$P(x) = c \frac{rx^{r-1}}{x_0^r} \exp_q \left[- \left(\frac{x}{x_0} \right)^r \right], \quad (6.3)$$

visto que ela interpola as distribuições q -exponencial e de Weibull. Chamaremos (6.3) de distribuição q -stretched exponencial ou simplesmente q -stretched.

Como a distribuição q -stretched se origina de duas distribuições distintas, seus parâmetros relevantes (q , r e x_0) permanecem basicamente com o significado inalterado. A combinação entre tais parâmetros faz com que a distribuição q -stretched tenha uma grande flexibilidade de ajuste. Por outro lado, tal fato pode dificultar a obtenção de seus valores precisos.

Notemos que a q -stretched conserva a propriedade da q -exponencial de ter leis de potência como um caso limite. Tal distribuição aproxima-se da lei potência $P(x) \simeq c(1-q)^{1/(1-q)}(x/x_0)^{r/(1-q)}$ se $x \gg x_0(1-q)^{-1/r}$ e $q > 1$. Além disso, a q -stretched deve se anular a partir do valor crítico $x_c = x_0(1-q)^{-1/r}$ para evitar valores imaginários para a distribuição se $q < 1$. Além disso, os parâmetros q e r da distribuição q -stretched podem se combinar de forma a maximizar ou minimizar certo tipo de comportamento.

6.2 Identificação e ajuste de uma q -stretched exponencial

Seguindo o procedimento empregado nos capítulos anteriores, consideremos inicialmente a forma acumulada da distribuição q -stretched

$$\begin{aligned}
 R(x) &= \int_x^\infty c \frac{r z^{r-1}}{x_0^r} \exp_q \left[- \left(\frac{z}{x_0} \right)^r \right] = \frac{c}{2-q} \left[1 - (1-q) \left(\frac{x}{x_0} \right)^r \right]^{\frac{2-q}{1-q}} \\
 &= cq' \left[1 - \frac{(1-q')}{q'} \left(\frac{x}{x_0} \right)^r \right]^{\frac{1}{1-q'}} = c' \left[1 - (1-q') \left(\frac{x}{x'_0} \right)^r \right]^{\frac{1}{1-q'}} \\
 &= c' \exp_{q'} \left[- \left(\frac{x}{x'_0} \right)^r \right], \tag{6.4}
 \end{aligned}$$

onde os novos parâmetros q' , x'_0 e c' podem ser expressos em função dos parâmetros q , r , x_0 e c na forma

$$q' = \frac{1}{2-q}, \quad x'_0 = \frac{x_0}{(2-q)^{\frac{1}{r}}} \quad \text{e} \quad c' = \frac{c}{2-q}. \tag{6.5}$$

O comportamento de tal distribuição acumulada, para valores particulares de q' e de r , pode ser visualizado nas figuras (6.1) e (6.2).

6.2.1 q -stretched e máximo R^2

Ao molde dos capítulos anteriores, passemos a obter um gráfico linear que relacione $R(x)$ com x . Para tal, tomemos o q -logaritmo em ambos os lados de (6.4),

$$\ln_{q'} R = \ln_{q'} \left[c' \exp_{q'} \left[- \left(\frac{x}{x'_0} \right)^r \right] \right], \tag{6.6}$$

e usando-se a relação (D.4), temos que

$$\begin{aligned}
 \ln_{q'} R &= \ln_{q'} c' + \ln_{q'} \left[\exp_{q'} \left[- \left(\frac{x}{x'_0} \right)^r \right] \right] + (1-q) \ln_{q'} c' \ln_{q'} \left[\exp_{q'} \left[- \left(\frac{x}{x'_0} \right)^r \right] \right] \\
 &= - \frac{[1 + (1-q') \ln_{q'} c']}{(x'_0)^r} x^r + \ln_{q'} c'. \tag{6.7}
 \end{aligned}$$

Portanto, obtém-se um comportamento linear ao usarmos a escala $\ln_{q'} R$ por x^r , conforme podemos visualizar na figura (6.3).

Os parâmetros q' e r a serem escolhidos são aqueles que maximizam o coeficiente R^2 da reta $\ln_{q'} R = A_0 + A_1 x^r$. Assim,

$$x'_0 = \left[\frac{1 + (1 - q')A_0}{-A_1} \right]^{\frac{1}{r}} \quad (6.8)$$

e

$$c' = \exp_{q'}(A_0). \quad (6.9)$$

Uma vez encontrados tais parâmetros as equações (6.5) possibilitam recuperar os valores de q , x_0 e c da distribuição (6.3).

6.2.2 Mínimos quadrados, máxima verossimilhança e q -stretched

Com relação ao método dos mínimos quadrados, os melhores valores de c' , x'_0 , r , e q' são aqueles que minimizam a função

$$S = \sum_{i=1}^N D_i^2 = \sum_{i=1}^N \left(c' \exp_{q'} \left[- \left(\frac{x_i}{x'_0} \right)^r \right] - R_i \right)^2. \quad (6.10)$$

No que se refere ao método da máxima verossimilhança, a normalização

$$\int_{x_{min}}^{\infty} n \frac{r z^{r-1}}{(x_0)^r} \exp_q \left[- \left(\frac{z}{x_0} \right)^r \right] dz = 1 \quad (6.11)$$

conduz a

$$n = (2 - q) \left(\frac{x_0^r}{q - 1} \right)^{2 \frac{2-q}{q-1}} \frac{1}{\left[\exp_q \left[- \left(\frac{x_{min}}{x_0} \right)^r \right] \right]^{2-q}}. \quad (6.12)$$

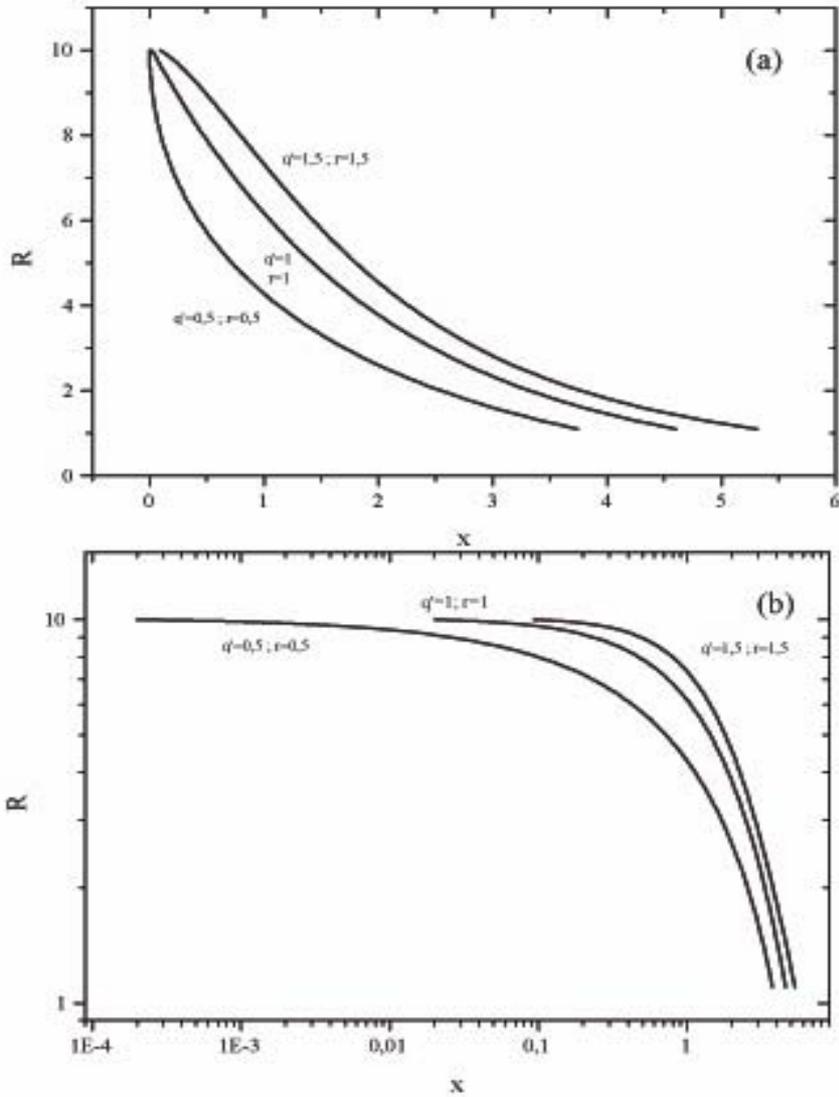


Figura 6.1: (a) - Gráfico de $R(x)$ por x de uma distribuição acumulada q -stretched: $R(x) = 10 \exp_{q'}[-(x/2)^r]$, com $q' = 0,5$ e $r = 0,5$, $q' = 1$ e $r = 1$, $q' = 1,5$ e $r = 1,5$. (b) - Gráfico anterior na escala log-log.

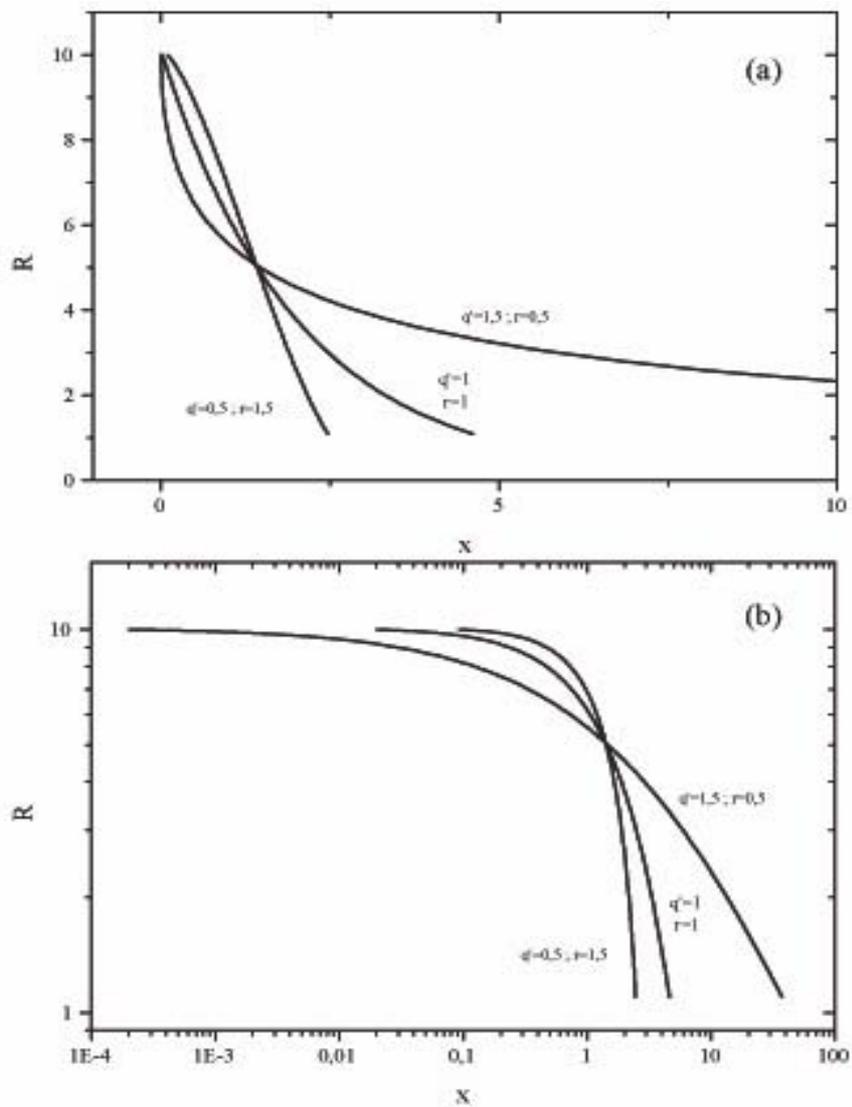


Figura 6.2: (a) - Gráfico de $R(x)$ por x de uma distribuição acumulada q -stretched: $R(x) = 10 \exp_{q'}[-(x/2)^r]$, com $q' = 0,5$ e $r = 1,5$, $q' = 1$ e $r = 1$, $q' = 1,5$ e $r = 0,5$. (b) - Gráfico anterior na escala log-log.

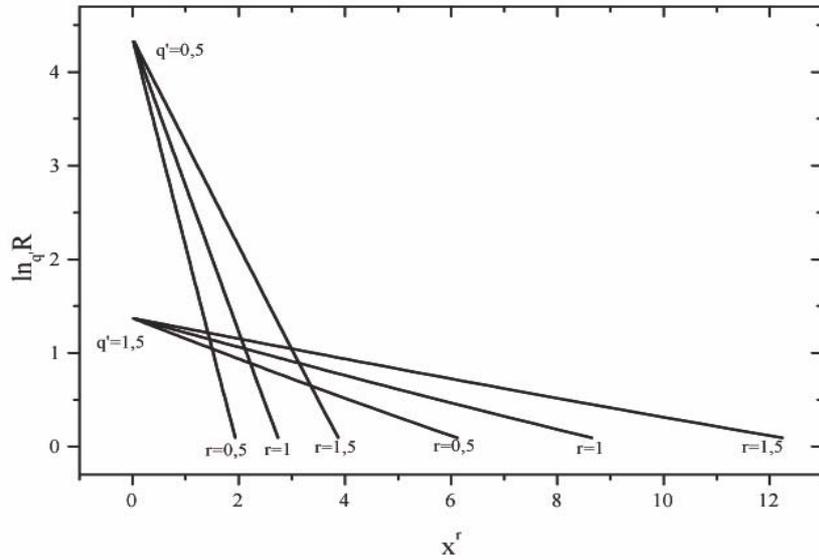


Figura 6.3: Gráfico de $R(x) = 10 \exp_{q'}[-(x/2)^r]$ por x para valores particulares de q' e r na escala $\ln_{q'} R$ por x^r .

Assim, a função L , ou melhor, $\ln L$ a ser maximizada é

$$\ln L = N \ln \left(\frac{n r}{x_0^r} \right) + \sum_{i=1}^N \ln \left[1 - (1 - q) \left(\frac{x_i}{x_0} \right)^r \right] + (r - 1) \sum_{i=1}^N \ln x_i. \quad (6.13)$$

6.3 Exemplos

Nesta seção apresentamos os ajustes dos sistemas estudados nos dois últimos capítulos à distribuição q -stretched. Tais ajustes podem ser visualizados nas figuras (6.4), (6.5), (6.6), (6.7), (6.8) e (6.9). Os parâmetros obtidos estão listados na tabela (6.1).

Sistema	Parâmetros	mR ²	mq	mv
Basquete	q'	0,79	0,67	0,70
	r	1,01	0,91	0,87
	x' ₀	232	249	231
	c'	396	403	- - -
Rodovias	q'	0,57	0,52	0,61
	r	0,60	0,57	0,68
	x' ₀	899	942	913
	c'	239	242	- - -
Ciclones	q'	0,75	0,95	0,74
	r	0,06	0,22	0,06
	x' ₀	$7,98 \times 10^{-6}$	5,06	$1,09 \times 10^{-5}$
	c'	$1,18 \times 10^4$	$1,12 \times 10^3$	- - -
Aeroportos	q'	0,86	1,82	1,39
	r	0,38	0,80	0,71
	x' ₀	$1,97 \times 10^6$	$1,33 \times 10^6$	$1,92 \times 10^6$
	c'	737	652	- - -
Vulcões	q'	0,73	1,48	1,13
	r	0,86	1,80	1,51
	x' ₀	$2,01 \times 10^3$	$1,77 \times 10^3$	$1,97 \times 10^3$
	c'	$1,68 \times 10^3$	$1,38 \times 10^3$	- - -
Remédios	q'	0,94	0,89	0,89
	r	0,27	0,25	0,26
	x' ₀	$4,23 \times 10^3$	$4,41 \times 10^3$	$5,33 \times 10^3$
	c'	$2,31 \times 10^3$	$2,22 \times 10^3$	- - -

Tabela (6.1) - Valores obtidos pelos ajustes dos dados à curva *q-stretched*, usando-se os métodos: máximo R² (mR²), mínimos quadrados (mq) e máxima verossimilhança (mv).

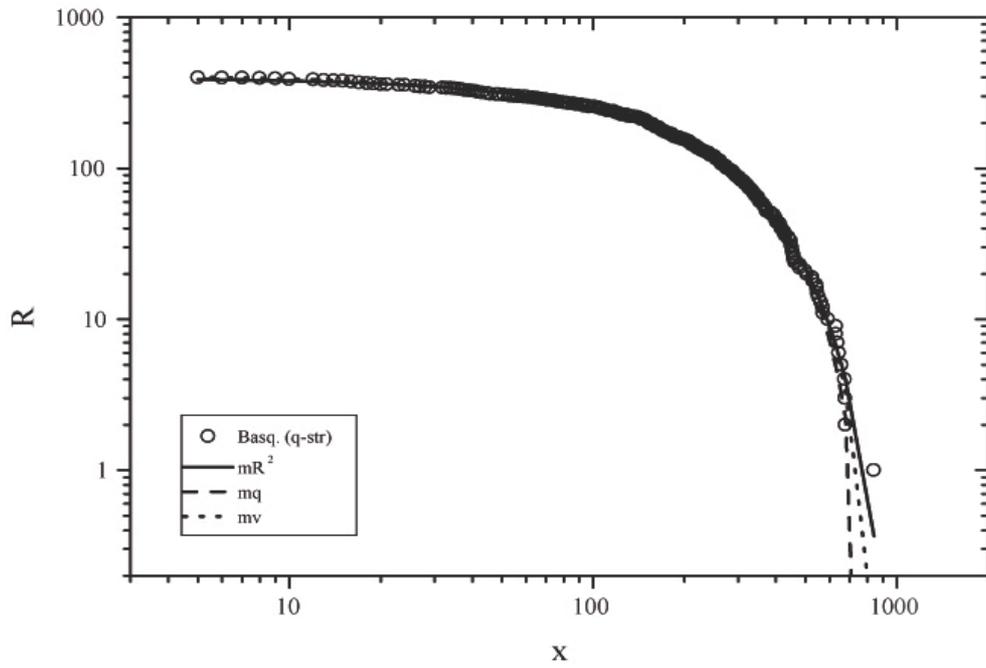


Figura 6.4: Ajuste, por uma q -stretched, da distribuição acumulada de cestas de dois pontos no basquete.

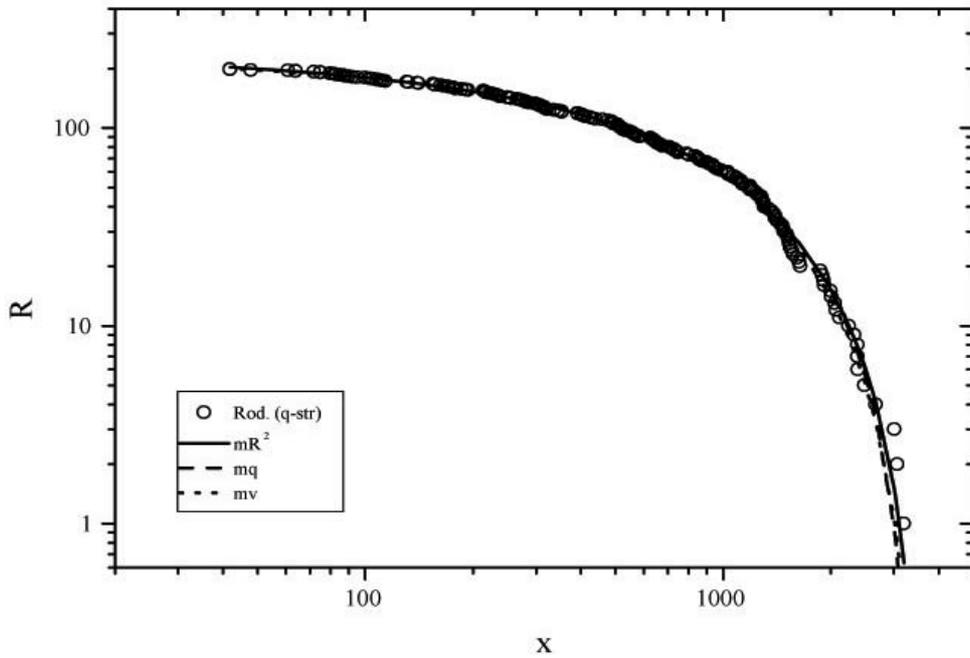


Figura 6.5: Ajuste, por uma q -stretched, da distribuição acumulada da extensão (em milhas) das rodovias dos Estados Unidos.

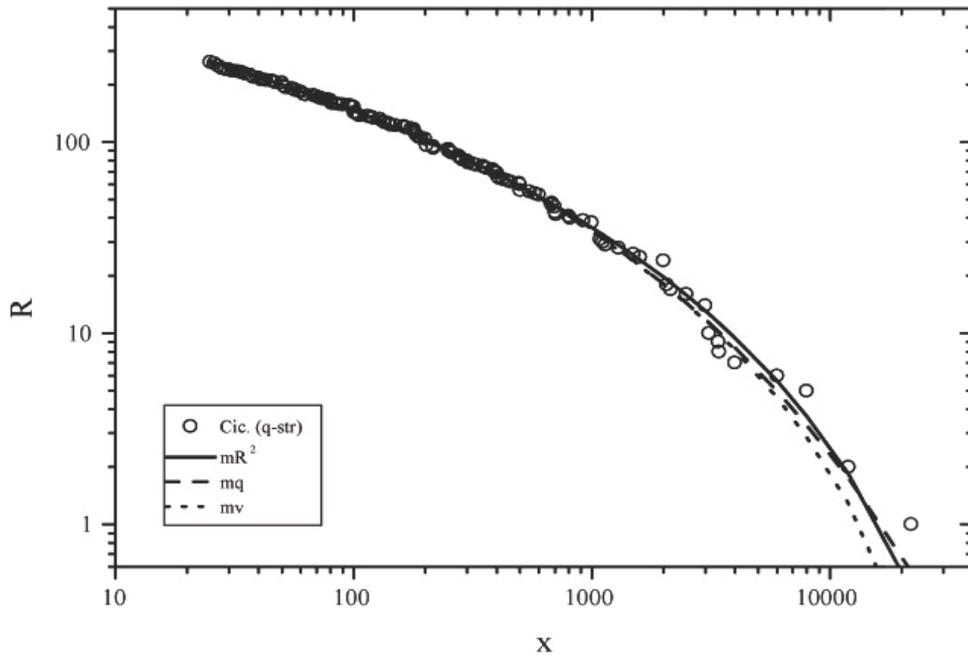


Figura 6.6: Ajuste, por uma q -stretched, da distribuição acumulada de vítimas de ciclones.

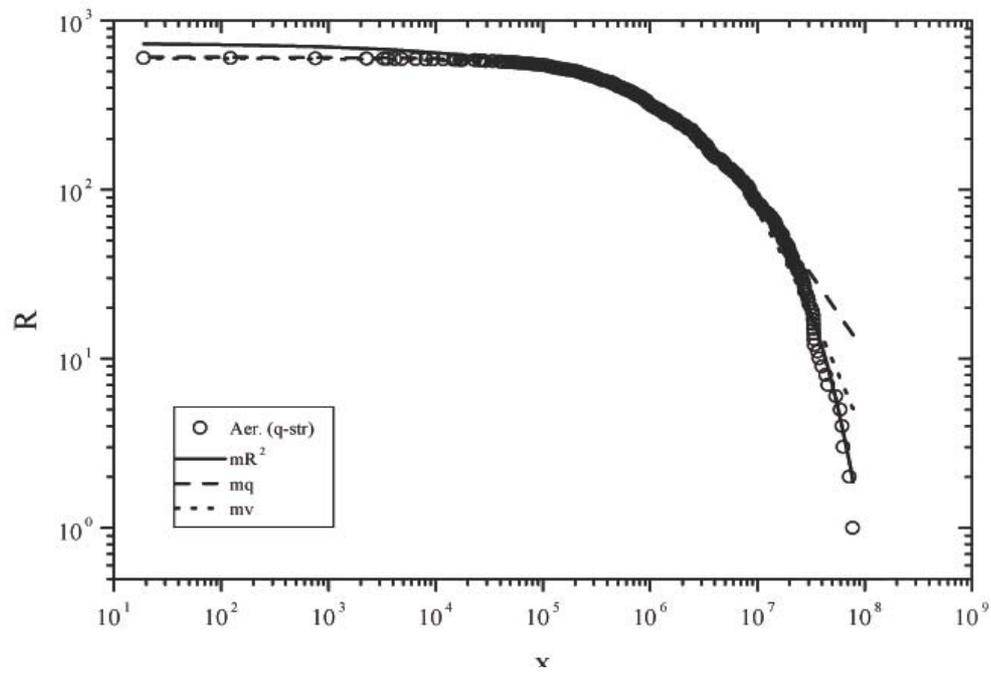


Figura 6.7: Ajuste, por uma q -stretched, da distribuição acumulada de passageiros nos aeroportos.

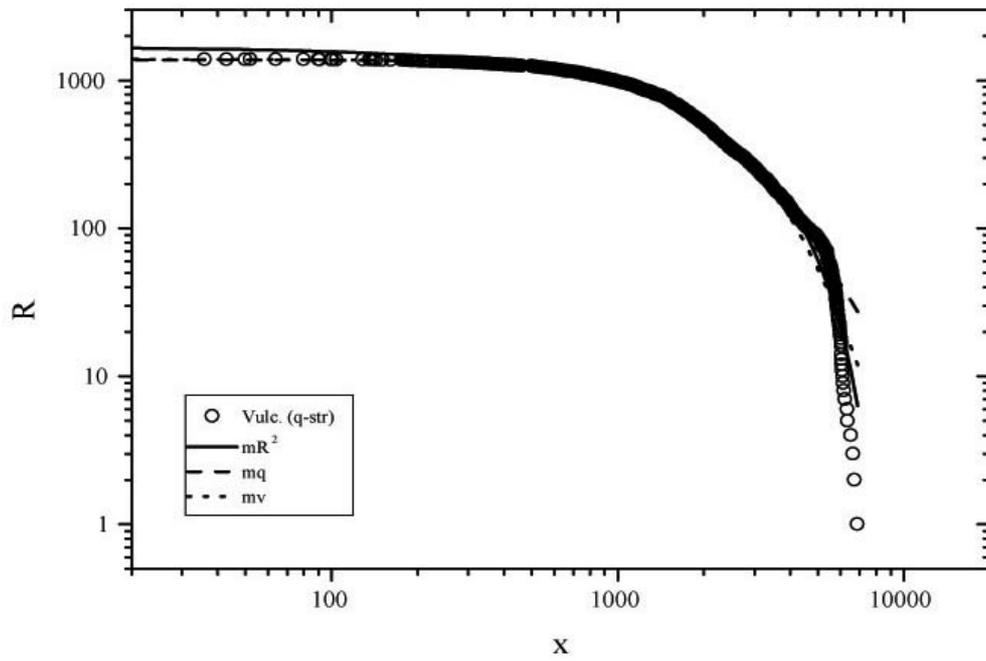


Figura 6.8: Ajuste, por uma *q-stretched*, da distribuição acumulada de altura de vulcões.

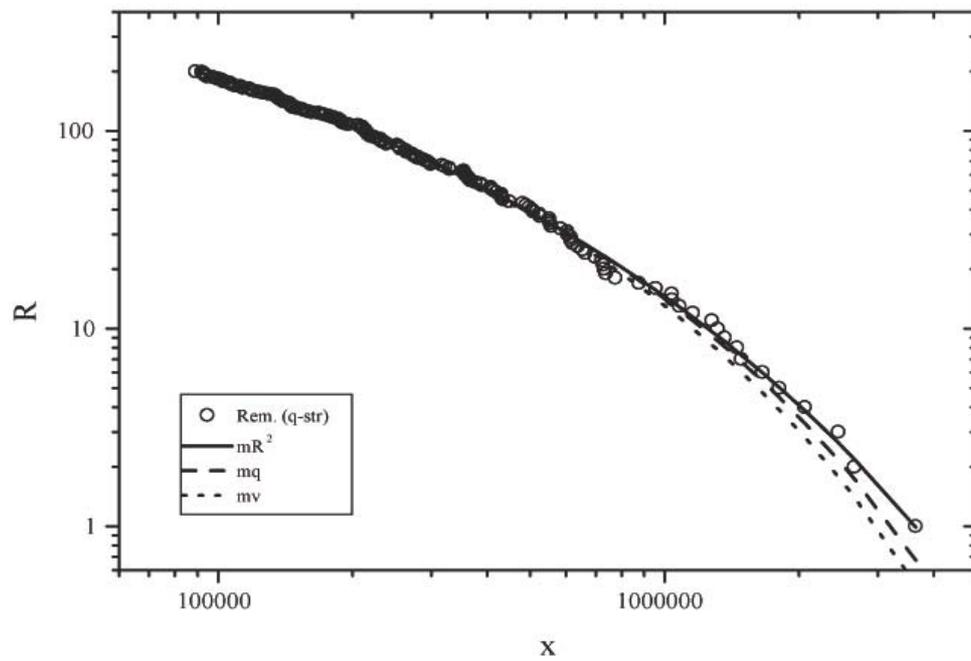


Figura 6.9: Ajuste, por uma *q-stretched*, da distribuição acumulada da arrecadação por marcas de remédios.

6.4 Distribuição q -stretched e modelamento

Antes de concluirmos este capítulo, discutiremos brevemente sobre um possível modelo (equação) que tem como solução a q -stretched. Para isso, voltemos nossa atenção, por um instante, para a equação que a distribuição de Weibull

$$y(x) = c \frac{r x^{r-1}}{x_0^r} \exp \left[- \left(\frac{x}{x_0} \right)^r \right] \quad (6.14)$$

satisfaz. Procedendo analogamente ao que fizemos para o caso q -exponencial, obtemos

$$\frac{dy}{dx} = a(x)y, \quad (6.15)$$

onde

$$a(x) = \frac{r}{x} \left(\frac{r-1}{r} - \frac{x^r}{x_0^r} \right). \quad (6.16)$$

Nota-se que a equação (6.15) não assume uma forma tão simples quanto a correspondente q -exponencial (ver equação (4.17)). Pior ainda seria se considerássemos a q -stretched

$$y(x) = c \frac{r x^{r-1}}{x_0^r} \exp_q \left[- \left(\frac{x}{x_0} \right)^r \right], \quad (6.17)$$

pois teríamos em (6.15)

$$a(x) = \frac{r}{x} \left(\frac{r-1}{r} - \frac{x^r}{x_0^r - (1-q)x^r} \right). \quad (6.18)$$

Uma equação diferencial bem mais elegante e que relembra a estrutura da (4.17) pode ser obtida para a distribuição acumulada (6.4),

$$R(x) = c' \exp_{q'} \left[- \left(\frac{x}{x_0'} \right)^r \right]. \quad (6.19)$$

De fato, constatamos que $R(x)$ satisfaz a equação

$$\frac{dR}{dx} = b x^{r-1} R^{q'}, \quad (6.20)$$

com $b = r(x'_0)^{-r}(c')^{1-q}$. Em particular, no caso da *stretched* exponencial temos $dR/dx = bx^{r-1}R$ e $dR/dx = bR^q$ para a q -exponencial.

Capítulo 7

Estudo comparativo entre distribuições

Neste capítulo, expomos, inicialmente, uma comparação entre as distribuições q -exponencial, de Weibull e q -stretched, no que se refere aos ajustes dos sistemas analisados nos últimos três capítulos. Visando obter informações sobre a melhor curva de ajuste, apresentamos uma comparação visual entre as curvas assim como os respectivos coeficientes R^2 , S/M e $(\ln L)/N$. Apresentamos também dois procedimentos, baseados na distribuição de erros, que podem ser decisivos na escolha da melhor curva de ajuste. Além disso, discutimos brevemente a presença de possíveis aleatoriedades nos sistemas complexos. Ilustramos como tal fato poderia dificultar ainda mais a obtenção de valores precisos para os parâmetros, bem como a escolha da melhor distribuição de ajuste.

7.1 Comparações visual e quantitativa

Dados relacionados ao basquete, às rodovias, aos ciclones, aos aeroportos, aos vulcões e aos remédios foram ajustados pela distribuição q -exponencial no capítulo 4, pela distribuição de Weibull no capítulo 5 e por uma distribuição que as interpola, a q -stretched, no capítulo 6. Porém, até agora, pouco falamos sobre a possibilidade de escolhermos, para cada um destes sistemas, a melhor distribuição de ajuste. Visando cobrir esta lacuna, iniciaremos nossa comparação entre as distribuições visualizando, figuras (7.1) a (7.6), as

curvas obtidas para cada sistema estudado.

Para uma comparação quantitativa entre tais curvas de ajuste, usamos valores numéricos dos coeficientes característicos de cada método de ajuste (R^2 , S e $(\ln L)$). Por exemplo, usando-se o método dos mínimos quadrados para ajustar dados de certo sistema a diferentes curvas de ajuste, a distribuição que fornecer o mínimo valor de S torna-se uma boa candidata a ser a melhor curva de ajuste para aquele sistema. Desta maneira, usamos os coeficientes R^2 , S/M e $(\ln L)/N$ para tal tipo de análise. Nestes coeficientes, M e N são o número de pontos num gráfico de R por x e de P por x , respectivamente. Uma listagem de tais coeficientes, para cada sistema em estudo, esta apresentada na tabela (7.1).

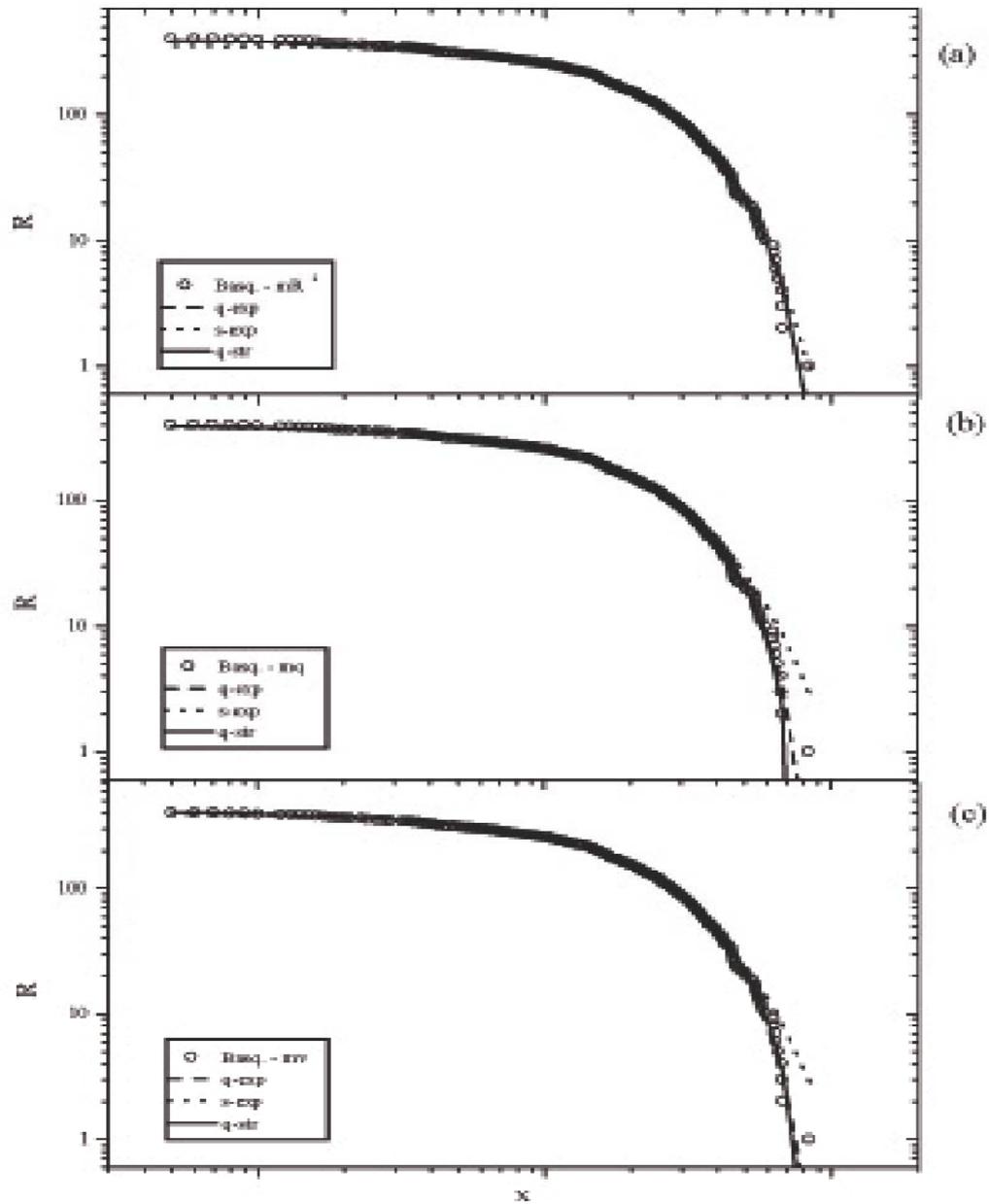


Figura 7.1: Ajustes da distribuição de cestas de basquete às curvas q -exponencial (q -exp), $stretched$ -exponencial (s -exp) e q - $stretched$ (q -str) usando-se os métodos (a) - máximo R^2 (mR^2), (b) - mínimos quadrados (mq) e (c) - máxima verossimilhança (mv).

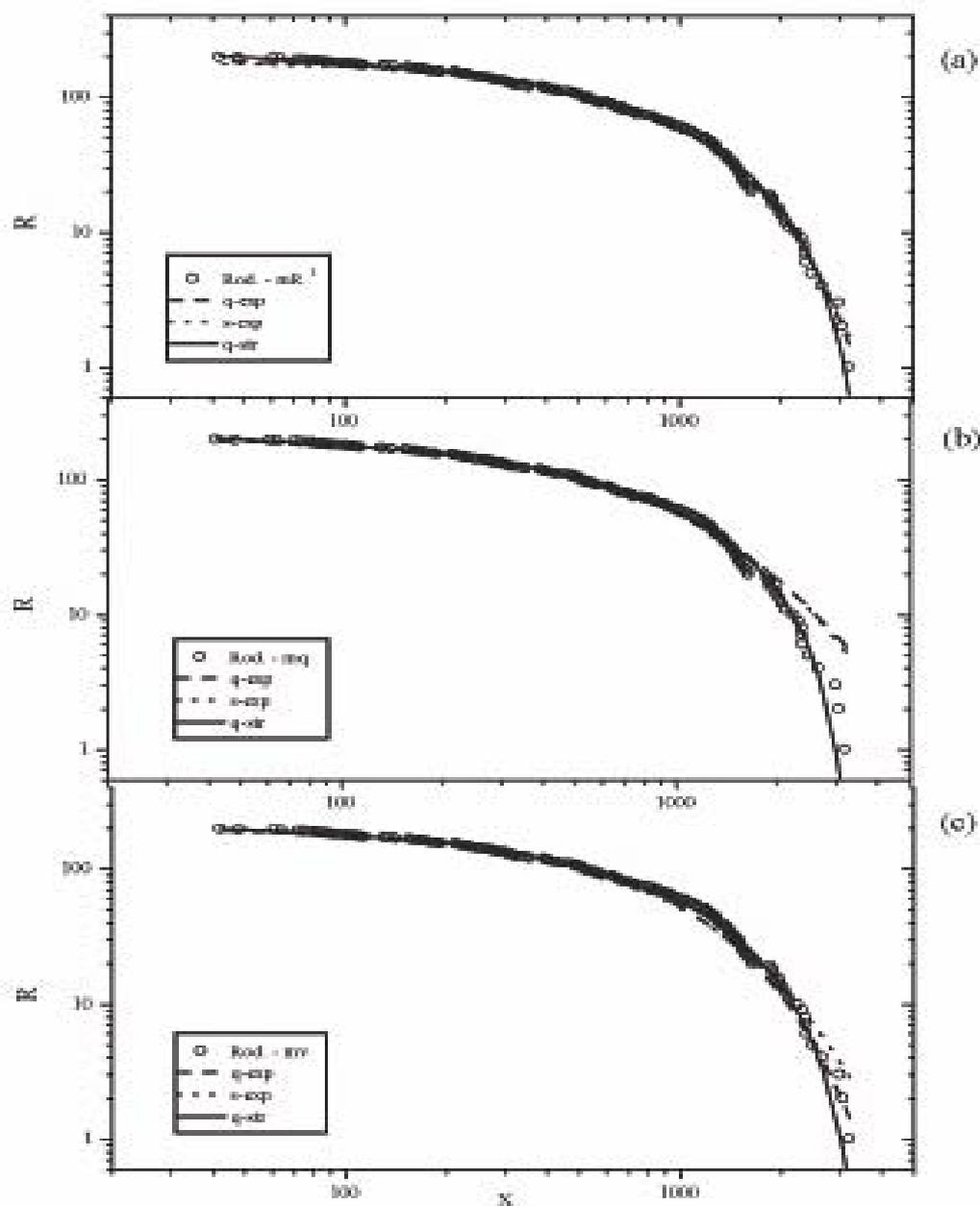


Figura 7.2: Ajustes da distribuição de extensão de rodovias às curvas q -exponencial (q -exp), $stretched$ -exponencial (s -exp) e q - $stretched$ (q -str) usando-se os métodos (a) - máximo R^2 (mR^2), (b) - mínimos quadrados (mq) e (c) - máxima verossimilhança (mv).

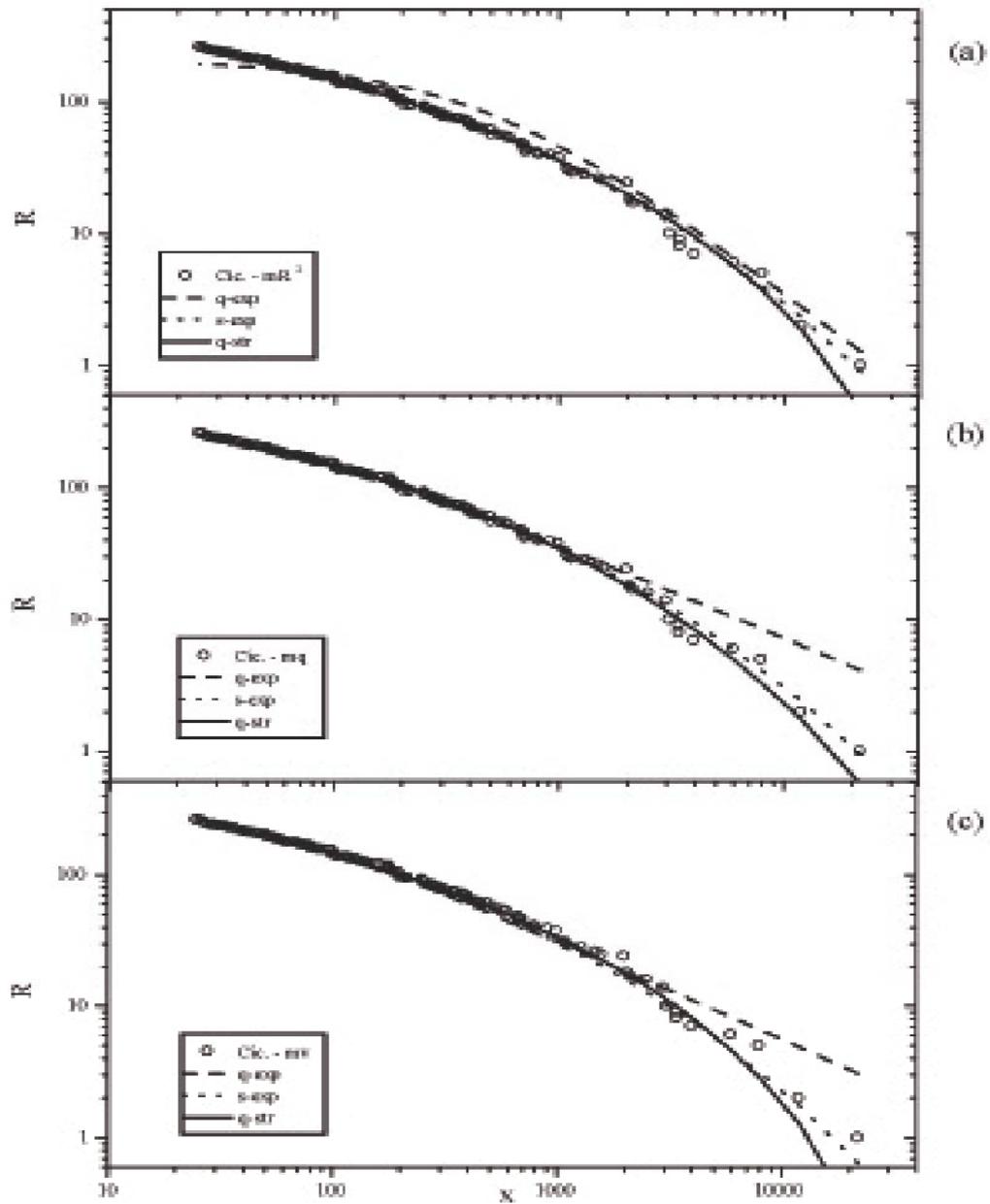


Figura 7.3: Ajustes da distribuição de vítimas de ciclones às curvas q -exponencial (q -exp), $stretched$ -exponencial (s -exp) e q -stretched (q -str) usando-se os métodos (a) - máximo R^2 (mR^2), (b) - mínimos quadrados (mq) e (c) - máxima verossimilhança (mv).

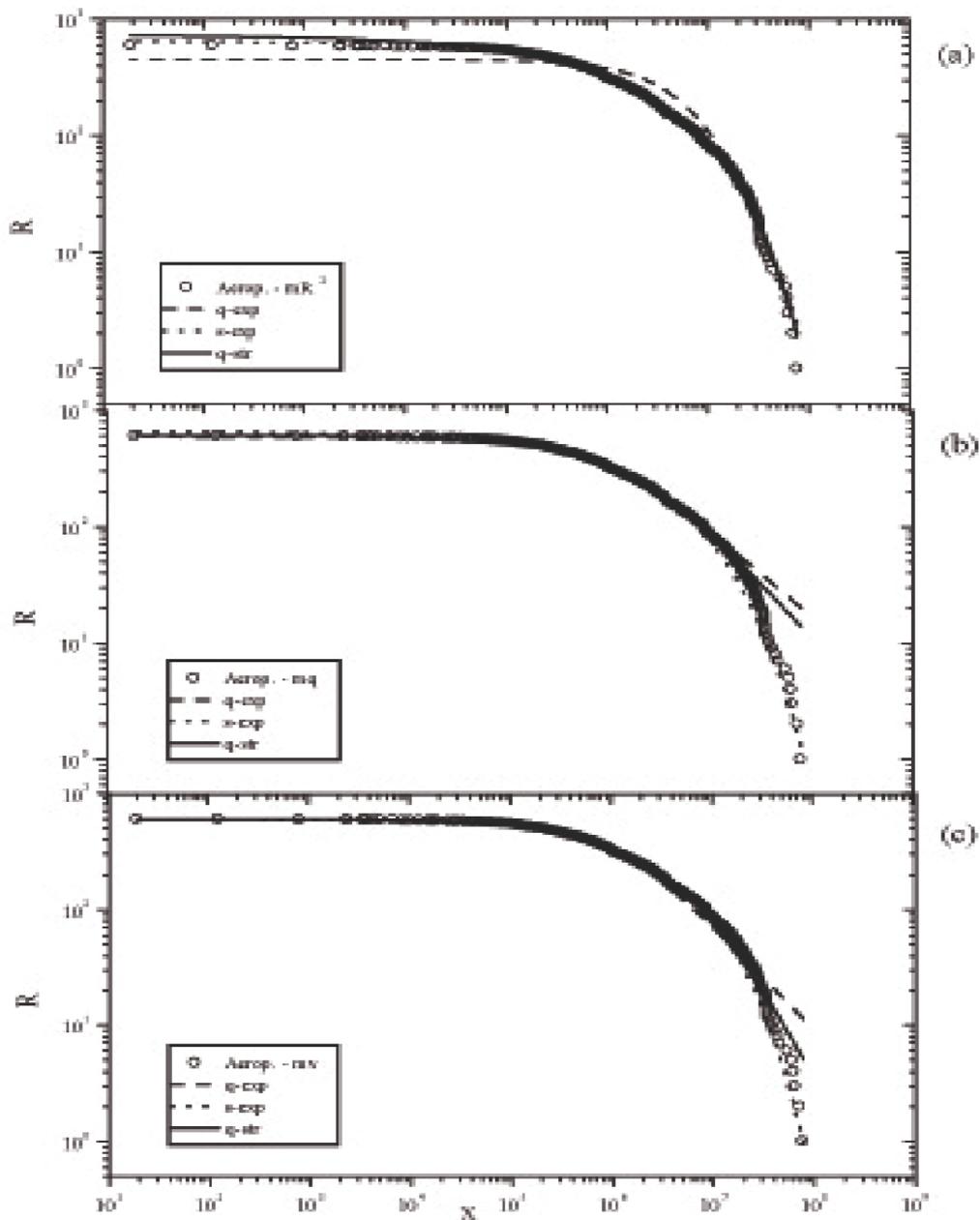


Figura 7.4: Ajustes da distribuição de movimentação de aeroportos às curvas q -exponencial (q -exp), $stretched$ -exponencial (s -exp) e q - $stretched$ (q -str) usando-se os métodos (a) - máximo R^2 (mR^2), (b) - mínimos quadrados (mq) e (c) - máxima verossimilhança (mv).

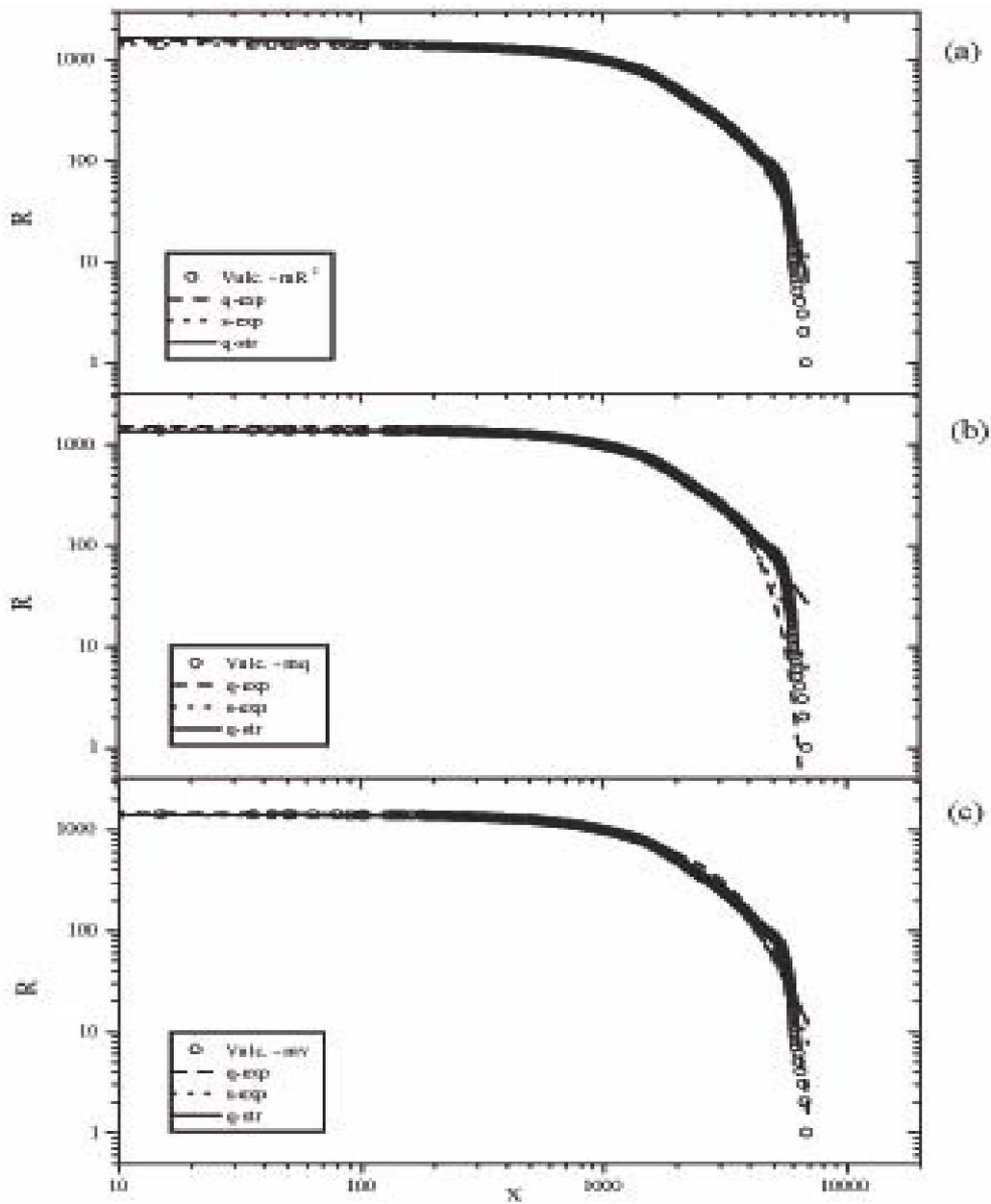


Figura 7.5: Ajustes da distribuição de altura de vulcões às curvas q -exponencial (q -exp), $stretched$ -exponencial (s -exp) e q - $stretched$ (q -str) usando-se os métodos (a) - máximo R^2 (mR^2), (b) - mínimos quadrados (mq) e (c) - máxima verossimilhança (mv).

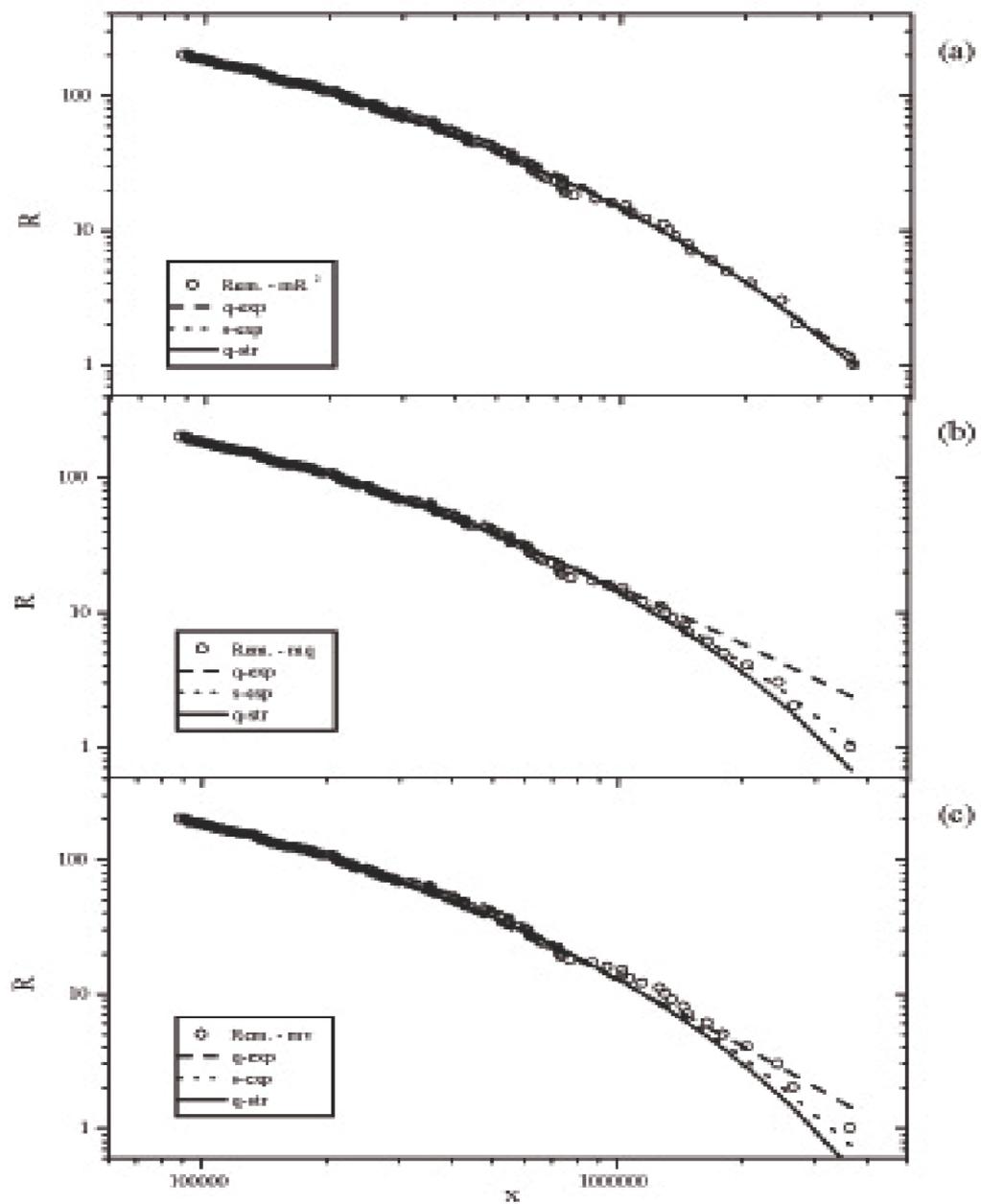


Figura 7.6: Ajustes da distribuição de arrecadação devido à venda de remédios (por marca) às curvas q -exponencial (q -exp), $stretched$ -exponencial (s -exp) e q - $stretched$ (q -str) usando-se os métodos (a) - máximo R^2 (mR^2), (b) - mínimos quadrados (mq) e (c) - máxima verossimilhança (mv).

Sistema	Distribuição	R^2	S/M	$(-\ln L)/N$
Basquete	q-exp	0,9978	20,06	6,195
	s-exp	0,9950	34,27	6,360
	q-str	0,9978	16,81	6,162
Rodovias	q-exp	0,9951	11,79	7,576
	s-exp	0,9922	09,20	7,576
	q-str	0,9972	05,02	7,560
Ciclones	q-exp	0,9908	11,63	6,722
	s-exp	0,9962	06,63	6,703
	q-str	0,9966	06,74	6,700
Aeroportos	q-exp	0,9701	45,07	16,072
	s-exp	0,9932	124,2	16,063
	q-str	0,9960	28,37	16,052
Vulcões	q-exp	0,9883	1071	8,507
	s-exp	0,9713	344,2	8,478
	q-str	0,9897	110,1	8,477
Remédios	q-exp	0,9951	04,72	13,425
	s-exp	0,9983	04,05	13,415
	q-str	0,9983	04,01	13,415

Tabela (7.1) - Valores dos coeficientes R^2 , S/M e $(\ln L)/N$ para cada sistema em estudo.

Observemos que na tabela (7.1) parece não haver grandes discrepâncias entre os coeficientes R^2 , S/M e $(\ln L)/N$. Independentemente do método usado, os coeficientes parecem apontar para uma mesma distribuição de ajuste. Por exemplo, no sistema relacionado com o basquete, os três coeficientes apontam para a distribuição q -exponencial, enquanto que no caso dos remédios, eles apontam para a distribuição de Weibull. Notamos também que a distribuição q -stretched, na maioria dos casos forneceu melhores resultados em vista de sua maior flexibilidade, tendo um parâmetro de ajuste a mais que as distribuições q -exponencial e de Weibull.

7.2 Análise de erros

Como já comentamos, o método dos mínimos quadrados consiste em minimizar a soma quadrática de erros ou diferenças (verticais ou horizontais) entre os dados e sua curva teórica de ajuste. No entanto, apenas uma soma dos erros talvez não seja suficiente para podermos decidir, com confiança, qual a melhor curva de ajuste. Neste ponto torna-se interessante analisar como estes erros estariam distribuídos ao longo da curva. Existiria alguma relação entre a qualidade do ajuste e a possível forma da distribuição de erros?

No século XVIII pesquisadores observaram que discrepâncias entre medidas repetidas de uma mesma quantidade física apresentavam um surpreendente grau de regularidade. Tais regularidades podiam muito bem ser aproximadas por uma distribuição contínua, então chamada de curva normal de erros. As propriedades matemáticas e a base teórica dessa distribuição foram investigadas por Laplace, Moivre e Gauss. Tal distribuição, muitas vezes é conhecida como distribuição normal[?]. Neste contexto, consideremos, como exemplo ilustrativo, o caso simples de medir o comprimento de uma barra de metal. Depois de várias medidas nota-se que, na ausência de erros grosseiros, as diferenças entre o comprimento real da barra e as medidas observadas geralmente seguem uma distribuição normal. Uma outra situação interessante seria considerar a produção de um certo número de barras de forma que deveriam ter o mesmo comprimento. Depois de medir cuidadosamente (apenas uma vez) cada uma das barras, obteríamos para a distribuição de erros um comportamento similar ao caso anterior, ou seja, uma distribuição normal. Neste último exemplo, os erros são as diferenças entre o comprimento ideal da barra e o comprimento real. Flutuações no processo de produção são possíveis fontes de tais erros. Por analogia, num ajuste entre uma curva teórica e uma distribuição de frequência acumulada de certos dados, os erros são as diferenças entre o valor observado R_i e o valor teórico $R(x_i)$. Desta analogia podemos esperar que um bom ajuste entre um conjunto de dados R_i e uma distribuição $R(x_i)$ deveria apresentar uma *distribuição de erros normal* $P(\varepsilon)$ em todo seu intervalo.

Além da forma da distribuição de erros, um outro aspecto que pode fornecer informações sobre a qualidade de um ajuste tem a ver com a aleatoriedade dos erros. Como exemplo, voltemos ao caso das barras de metal. Vamos considerar novamente a fabricação de um certo número delas. Vamos supor que $g(p)$ representam barras acima(abaixo) do

seu comprimento ideal. As barras produzidas podem então ser representadas por uma sequência do tipo

$$ppppgggppppppppppggppp\dots \quad (7.1)$$

O número de grupos com letras repetidas nesta sequência poderia ser um bom indicativo do grau de aleatoriedade do sistema (produção de barras). A análise desses grupos de repetição pode ajudar a detectar alguma tendência ou um possível comportamento cíclico, fato que poderia indicar alguma deficiência no processo de produção das barras. Quanto mais aleatória for a sequência, menor será a possibilidade de haver algum comportamento cíclico ou alguma tendência. Novamente, como analogia, uma *distribuição aleatória* dos erros ao longo da curva de ajuste é um indicativo de um bom ajuste.

Podemos visualizar os gráficos $P(\varepsilon_i)$ por ε_i e ε_i por i para o basquete, nas figuras (7.7) a (7.9); para as rodovias, nas figuras (7.10) a (7.12); para os ciclones, nas figuras (7.13) a (7.15); para os aeroportos, nas figuras (7.16) a (7.18); para os vulcões, nas figuras (7.19) a (7.21) e para os remédios, nas figuras (7.22) a (7.24). Nestas figuras, $\varepsilon_i = R(x_i) - R_i$ e $P(\varepsilon_i)$ é a frequência de ε_i .

Consideremos, como exemplo, o sistema relacionado com a extensão de rodovias. A distribuição de erros entre as curvas de ajuste (q -exponencial, *stretched*-exponencial e q -*stretched*) e a distribuição de extensão das rodovias, pode ser visualizada nas figuras (7.7) (máximo R^2), (7.8) (mínimos quadrados) e (7.9) (máxima verossimilhança). Notamos que a q -*stretched*, independente do método de ajuste usado, apresenta a forma mais próxima de uma distribuição normal de erros. Quanto à aleatoriedade dos erros, vemos que, novamente, a q -*stretched* não apresenta um comportamento muito definido, sendo aparentemente mais aleatório que o comportamento dos erros nas outras duas distribuições. Portanto, no que se refere às rodovias, todas as nossas análises, inclusive a comparação visual, indicam que a melhor distribuição de ajuste é a q -*stretched*.

Aplicando tal tipo de análise aos outros sistemas, podemos ter uma base mais sólida para decidir entre as distribuições q -exponencial e de Weibull. Chegamos à conclusão que, de uma maneira geral, a q -exponencial se ajusta relativamente bem ao basquete, que a distribuição de Weibull descreve bem a arrecadação por marcas de remédio e as vítimas de ciclones e que por fim, a q -*stretched* se ajusta melhor às rodovias, aeroportos e vulcões.

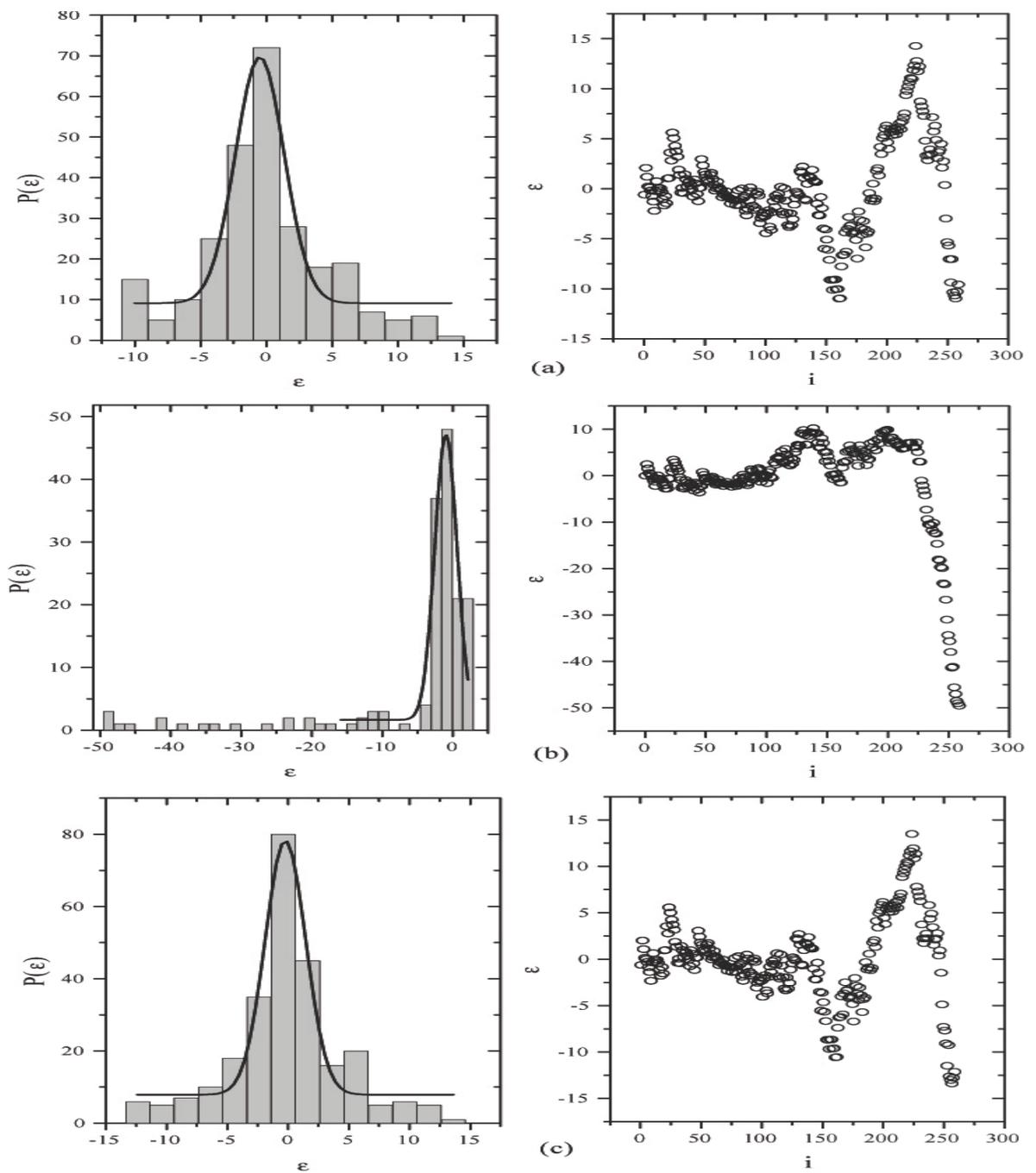


Figura 7.7: Sistema: Basquete. Método: máximo R^2 . Gráficos $P(\epsilon)$ por ϵ (frequência de erros) e ϵ por i (aleatoriedade) para as curvas de ajuste a) *q*-exponencial, b) *stretched*, c) *q*-*stretched*.

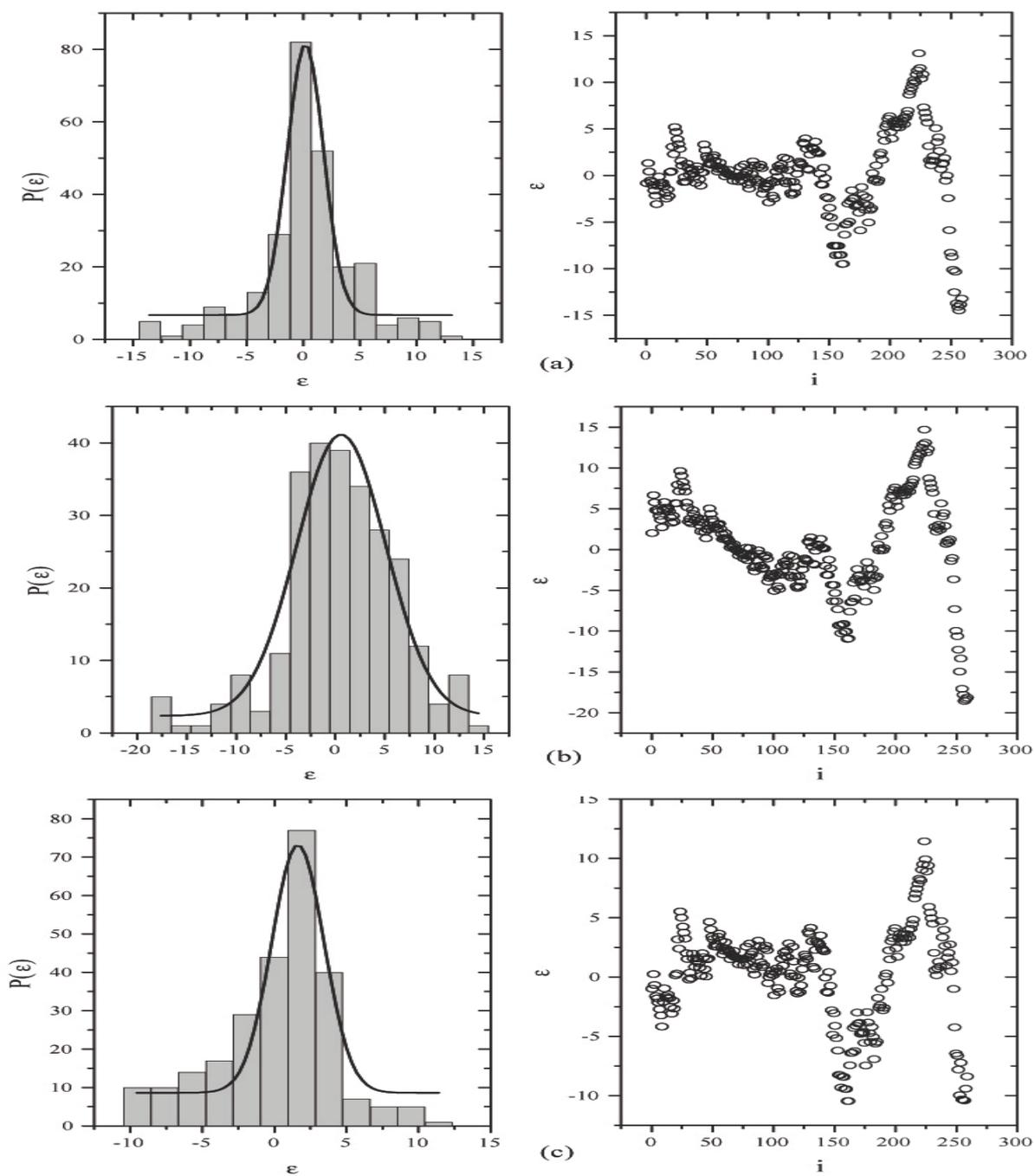


Figura 7.8: Sistema: Basquete. Método: mínimos quadrados. Gráficos $P(\varepsilon)$ por ε (frequência de erros) e ε por i (aleatoriedade) para as curvas de ajuste a) q -exponencial, b) stretched, c) q -stretched.

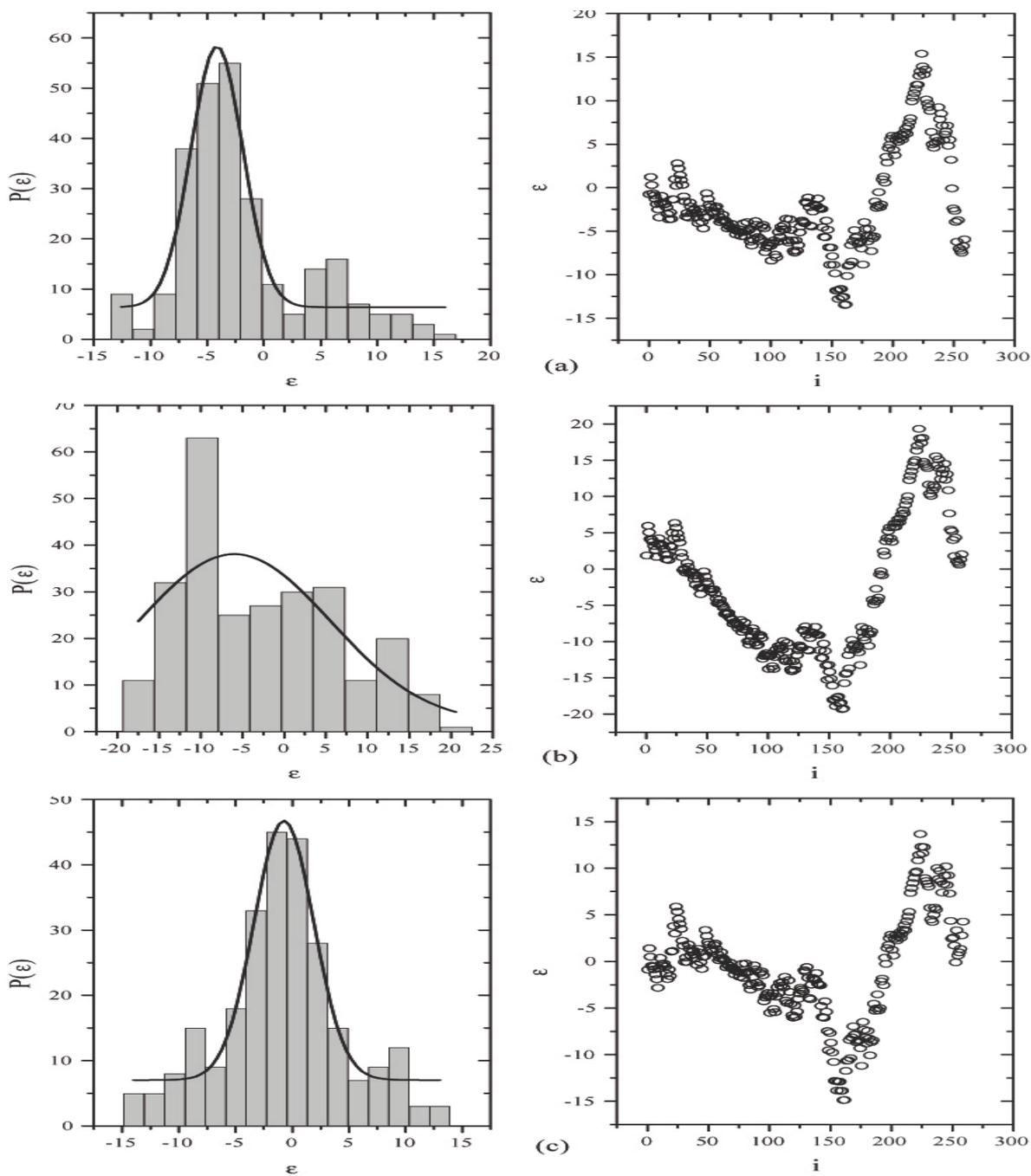


Figura 7.9: Sistema: Basquete. Método: máxima verossimilhança. Gráficos $P(\epsilon)$ por ϵ (frequência de erros) e ϵ por i (aleatoriedade) para as curvas de ajuste a) q -exponencial, b) *stretched*, c) q -*stretched*.

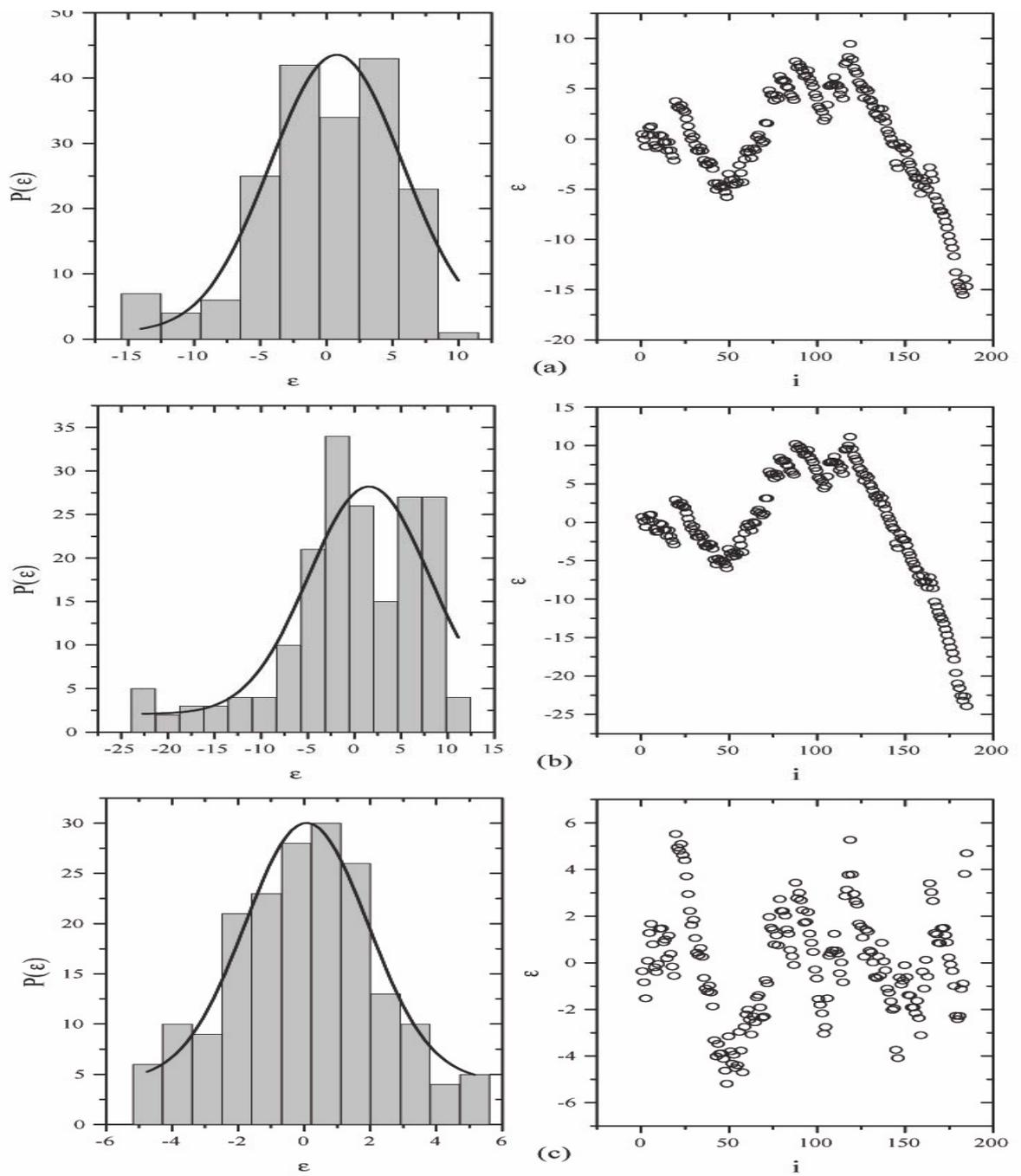


Figura 7.10: Sistema: Rodovias. Método: máximo R^2 . Gráficos $P(\epsilon)$ por ϵ (frequência de erros) e ϵ por i (aleatoriedade) para as curvas de ajuste a) q -exponencial, b) $stretched$, c) q - $stretched$.

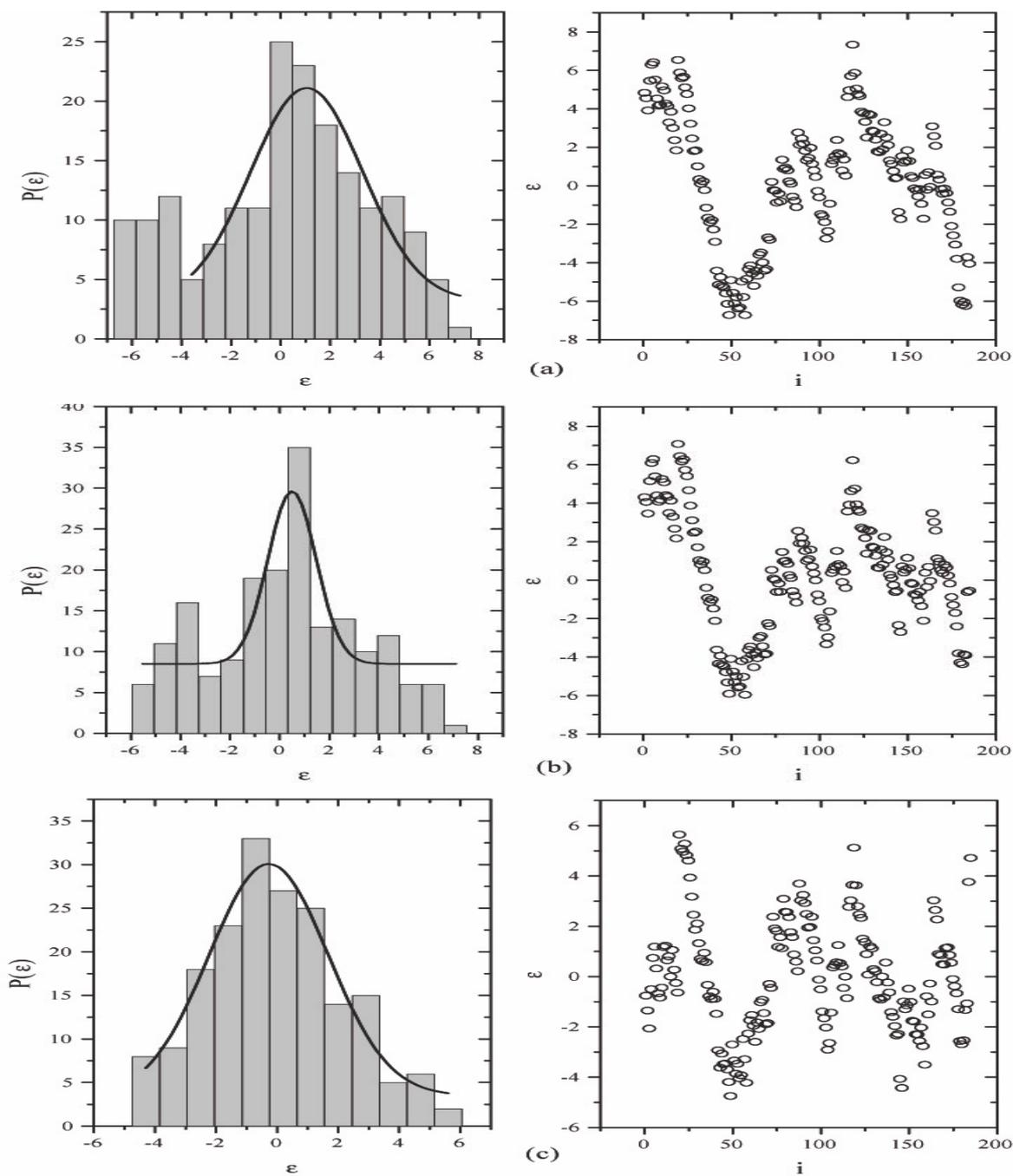


Figura 7.11: Sistema: Rodovias. Método: mínimos quadrados. Gráficos $P(\epsilon)$ por ϵ (frequência de erros) e ϵ por i (aleatoriedade) para as curvas de ajuste a) q -exponencial, b) $stretched$, c) q - $stretched$.

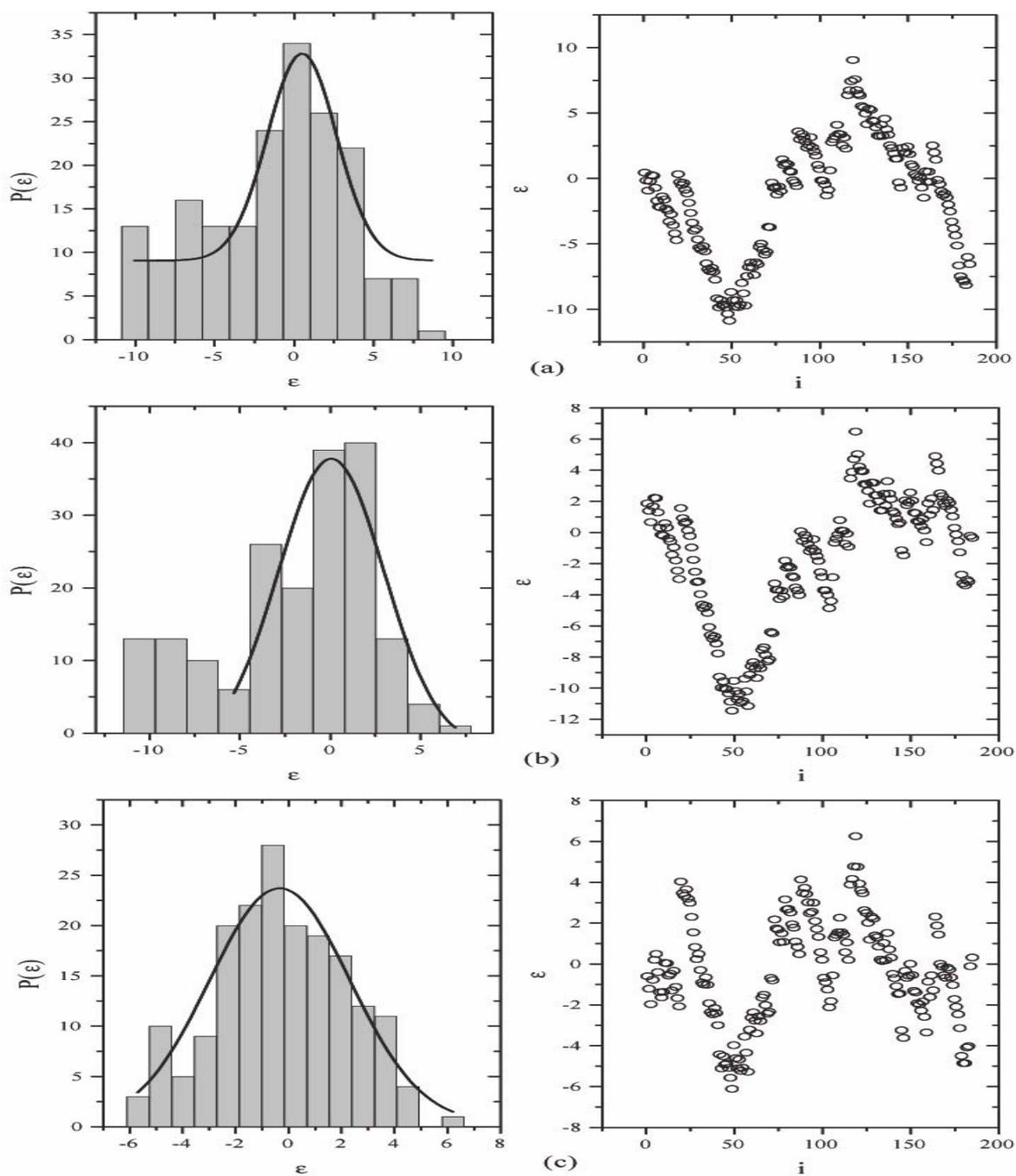


Figura 7.12: Sistema: Rodovias. Método: máxima verossimilhança. Gráficos $P(\varepsilon)$ por ε (frequência de erros) e ε por i (aleatoriedade) para as curvas de ajuste a) q -exponencial, b) $stretched$, c) q - $stretched$.

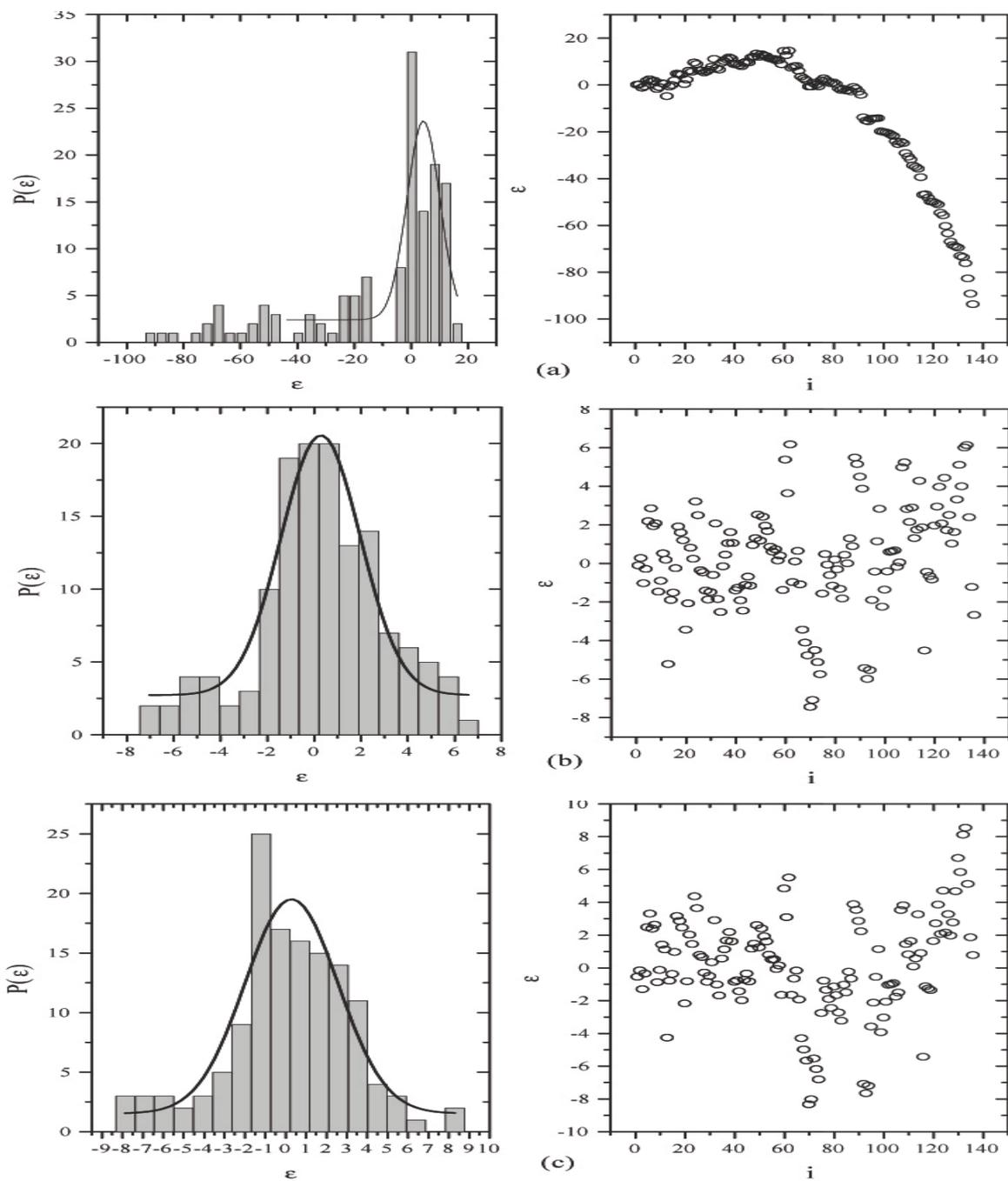


Figura 7.13: Sistema: Ciclones. Método: máximo R^2 . Gráficos $P(\epsilon)$ por ϵ (frequência de erros) e ϵ por i (aleatoriedade) para as curvas de ajuste a) q -exponencial, b) $stretched$, c) q - $stretched$.

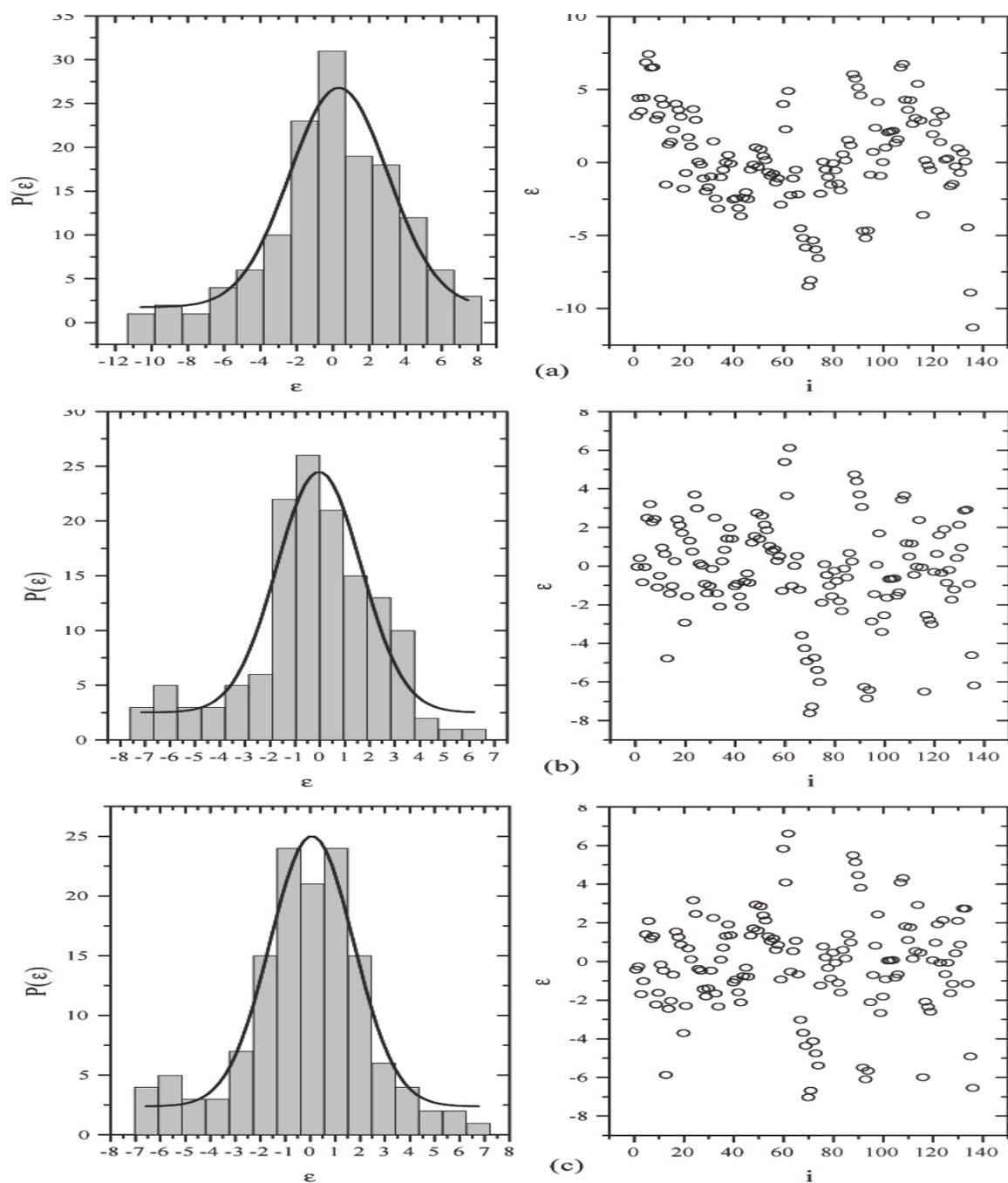


Figura 7.14: Sistema: Ciclones. Método: mínimos quadrados. Gráficos $P(\varepsilon)$ por ε (frequência de erros) e ε por i (aleatoriedade) para as curvas de ajuste a) q -exponencial, b) stretched, c) q -stretched.

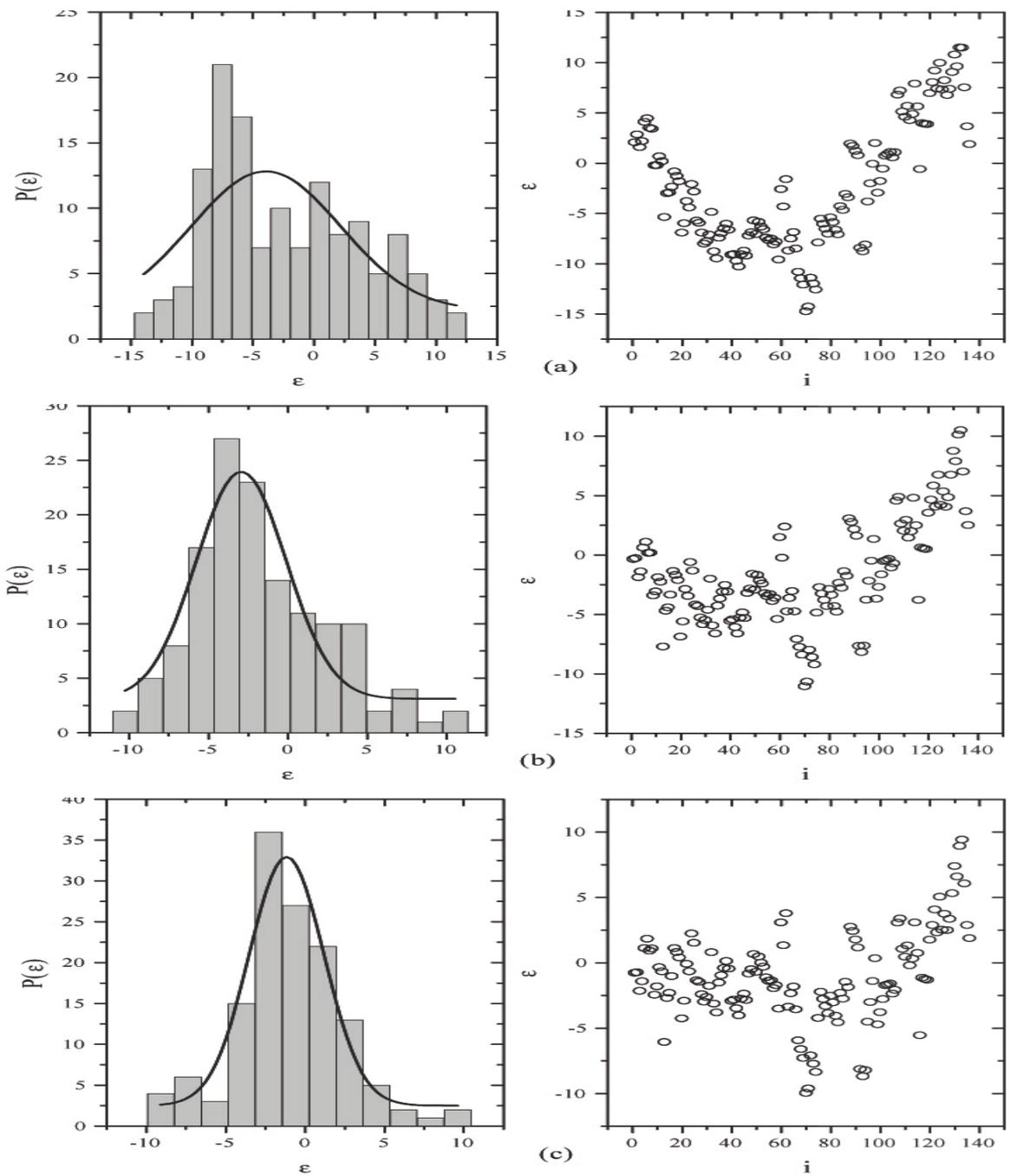


Figura 7.15: Sistema: Ciclones. Método: máxima verossimilhança. Gráficos $P(\varepsilon)$ por ε (frequência de erros) e ε por i (aleatoriedade) para as curvas de ajuste a) q -exponencial, b) $stretched$, c) q - $stretched$.

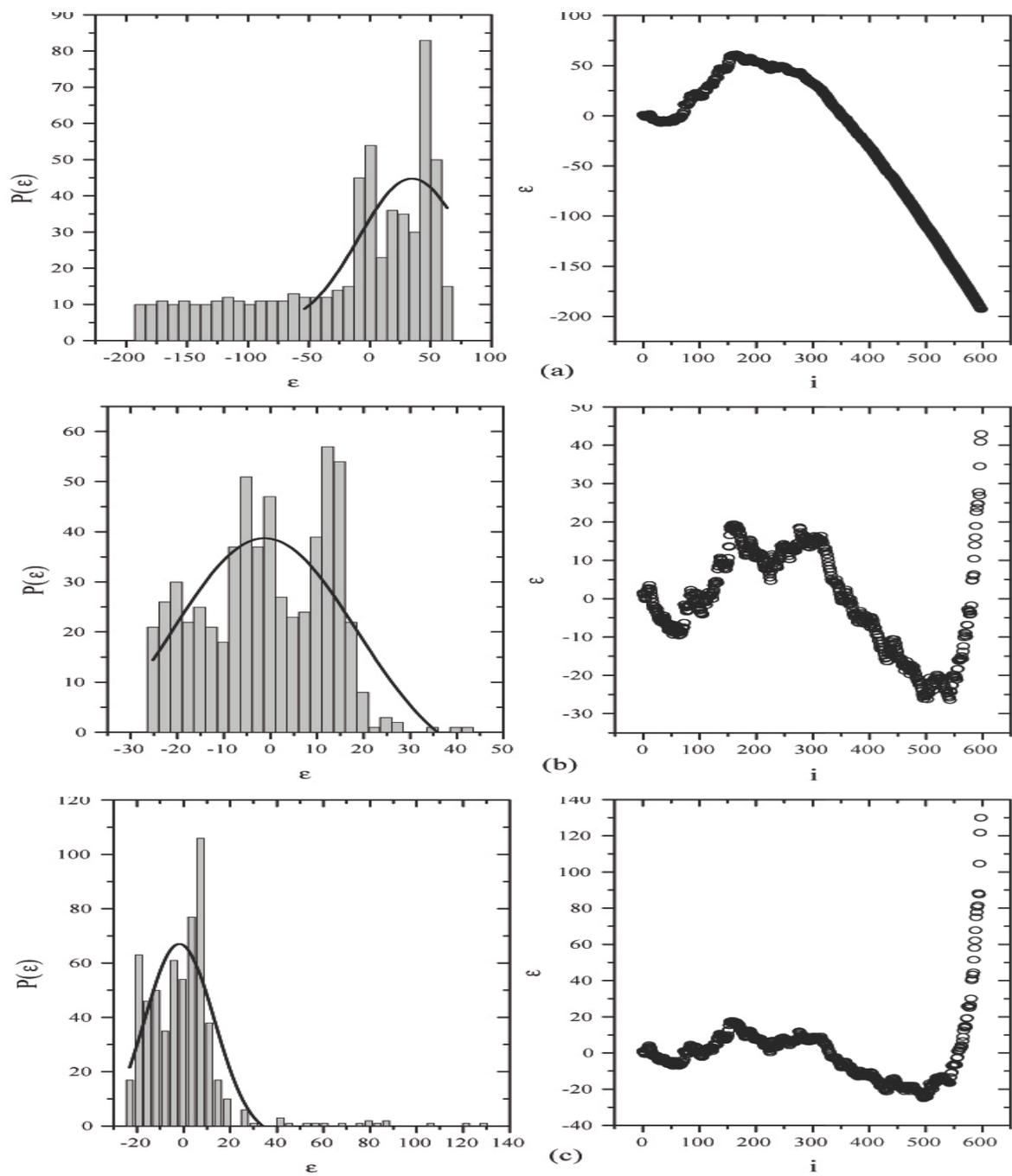


Figura 7.16: Sistema: Aeroportos. Método: máximo R^2 . Gráficos $P(\varepsilon)$ por ε (frequência de erros) e ε por i (aleatoriedade) para as curvas de ajuste a) q -exponencial, b) $stretched$, c) q - $stretched$.

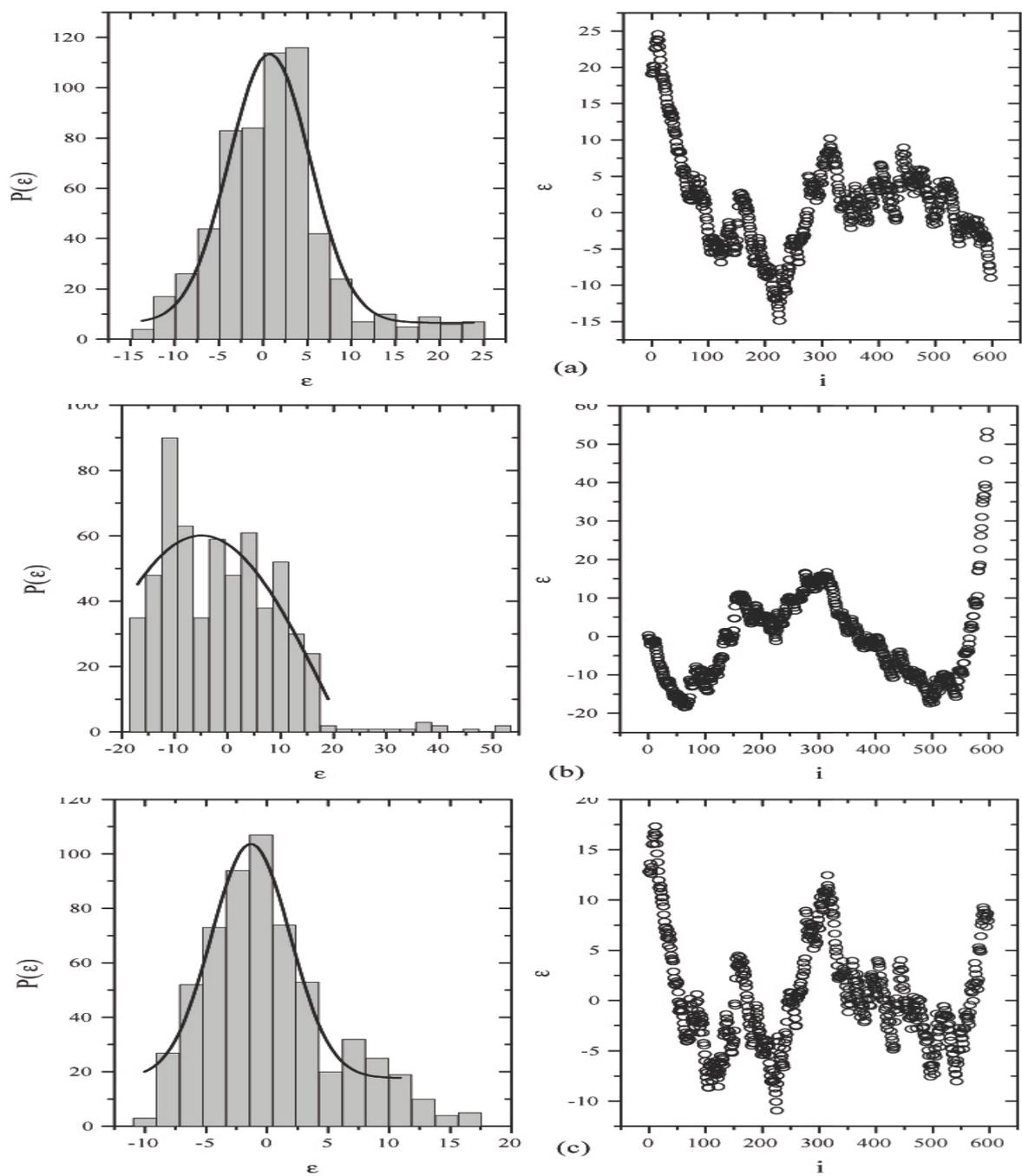


Figura 7.17: Sistema: Aeroportos. Método: mínimos quadrados. Gráficos $P(\epsilon)$ por ϵ (frequência de erros) e ϵ por i (aleatoriedade) para as curvas de ajuste a) q -exponencial, b) $stretched$, c) q - $stretched$.

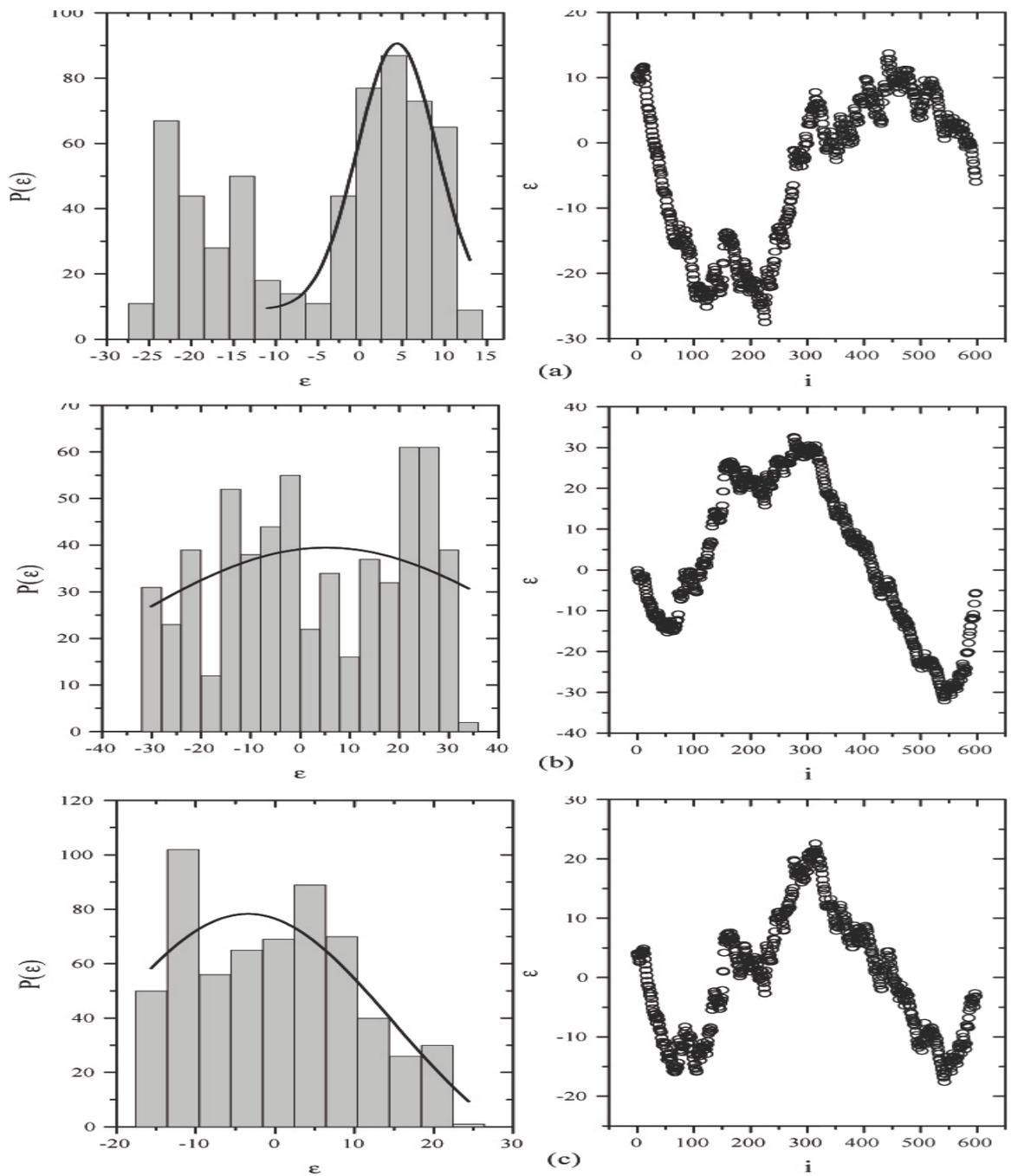


Figura 7.18: Sistema: Aeroportos. Método: máxima verossimilhança. Gráficos $P(\epsilon)$ por ϵ (frequência de erros) e ϵ por i (aleatoriedade) para as curvas de ajuste a) q -exponencial, b) *stretched*, c) q -*stretched*.

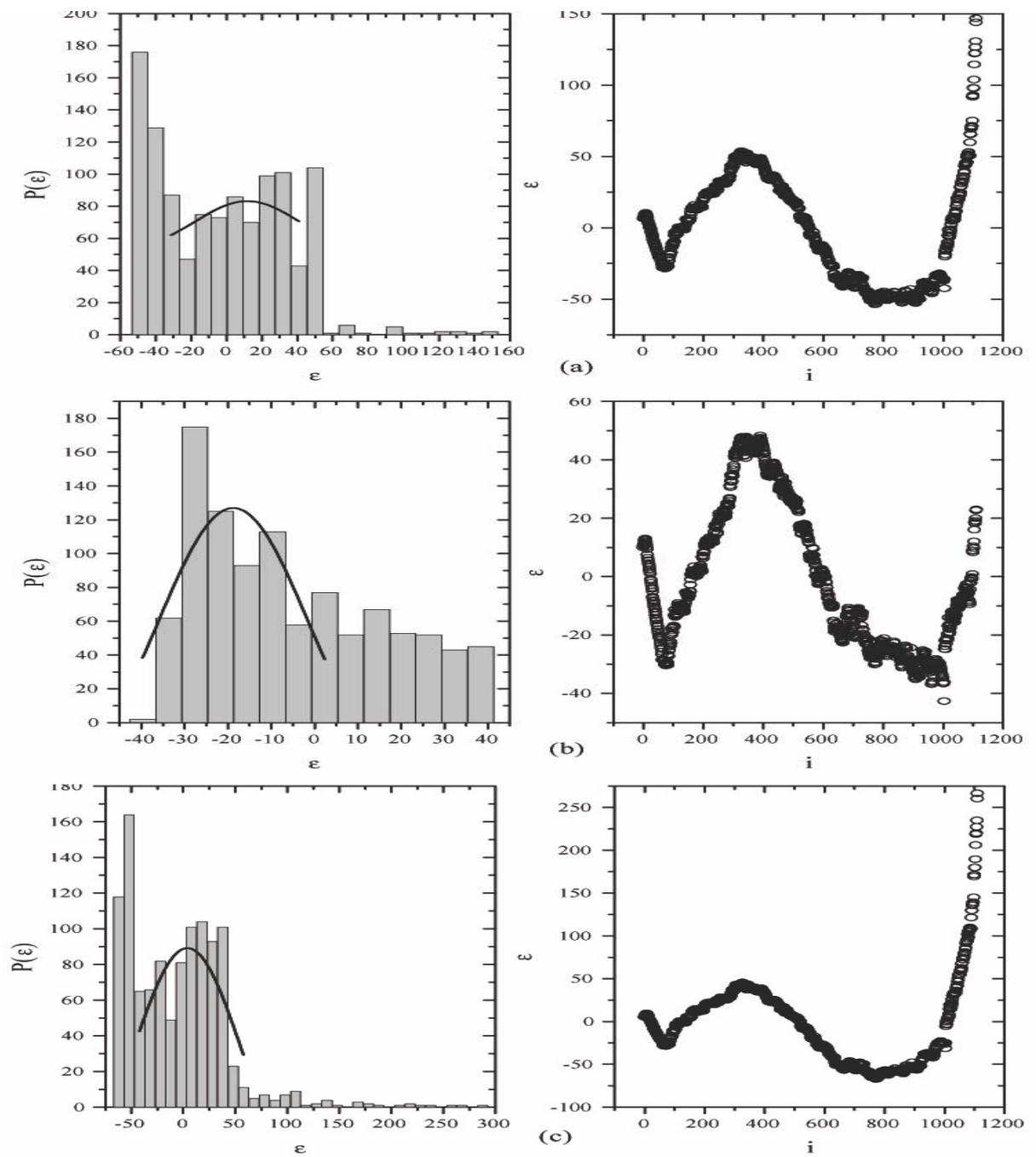


Figura 7.19: Sistema: Vulcões. Método: máximo R^2 . Gráficos $P(\epsilon)$ por ϵ (frequência de erros) e ϵ por i (aleatoriedade) para as curvas de ajuste a) q -exponencial, b) *stretched*, c) q -*stretched*.

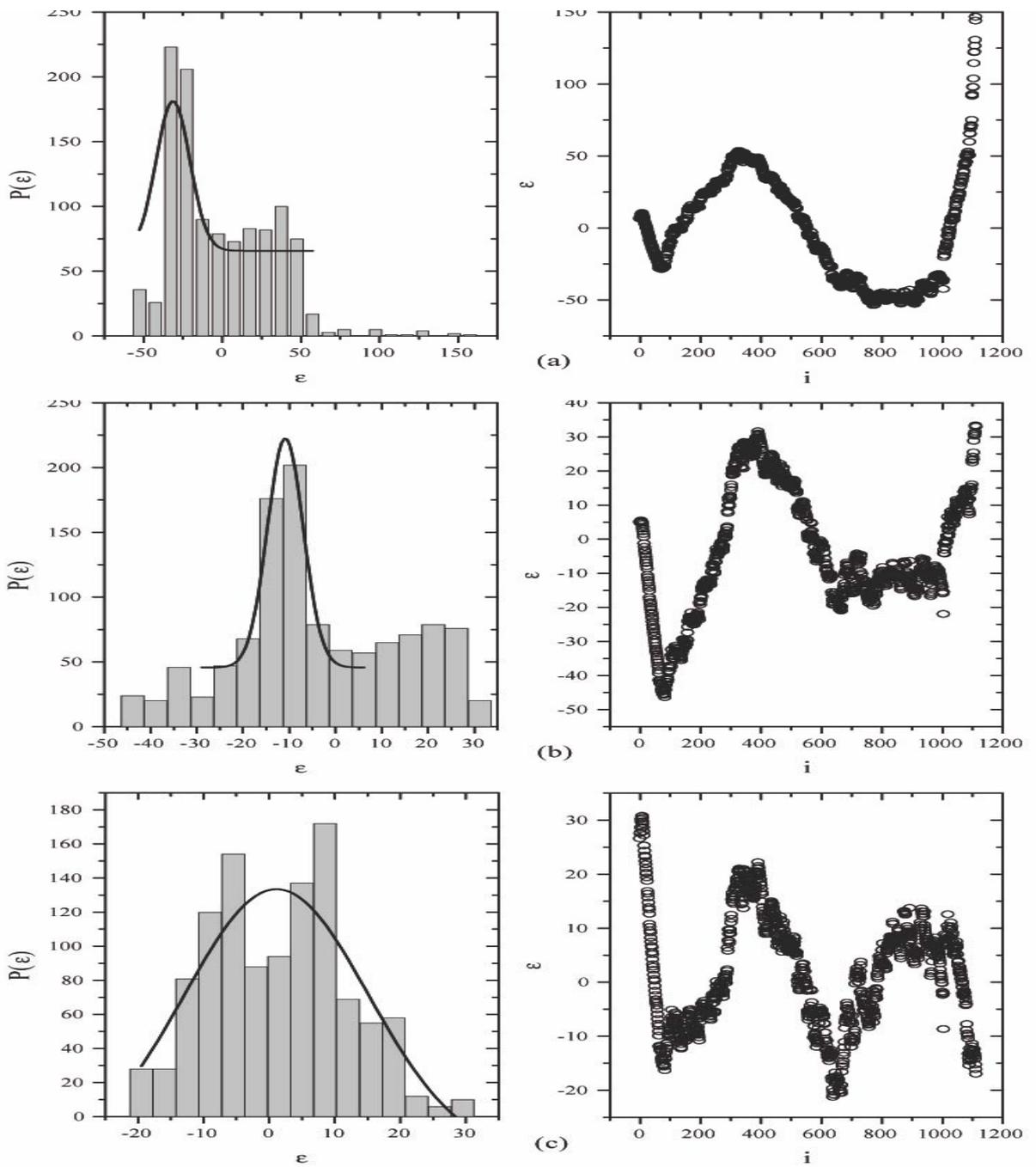


Figura 7.20: Sistema: Vulcões. Método: mínimos quadrados. Gráficos $P(\epsilon)$ por ϵ (frequência de erros) e ϵ por i (aleatoriedade) para as curvas de ajuste a) q -exponencial, b) $stretched$, c) q - $stretched$.

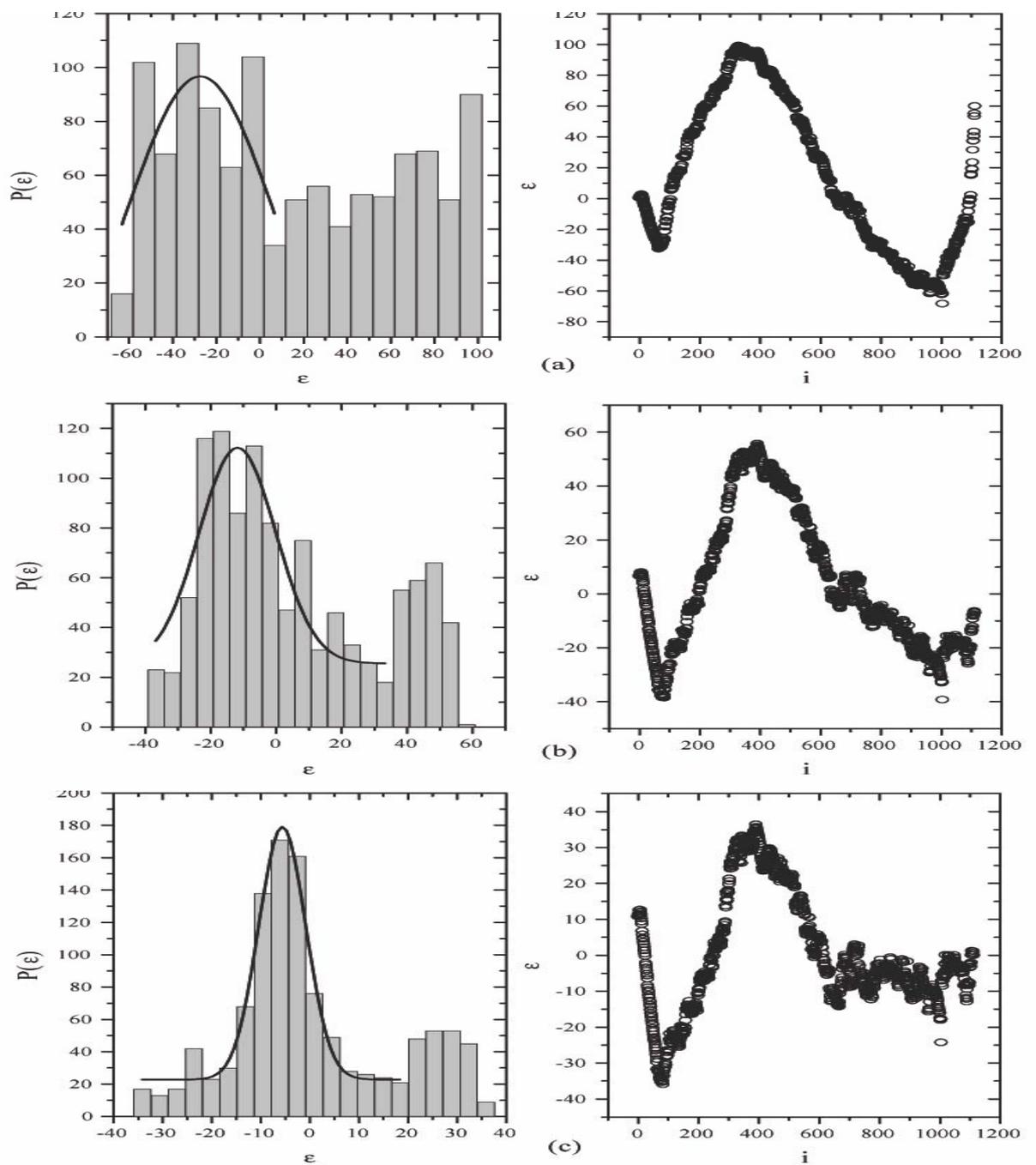


Figura 7.21: Sistema: Vulcões. Método: máxima verossimilhança. Gráficos $P(\varepsilon)$ por ε (frequência de erros) e ε por i (aleatoriedade) para as curvas de ajuste a) q -exponencial, b) $stretched$, c) q - $stretched$.

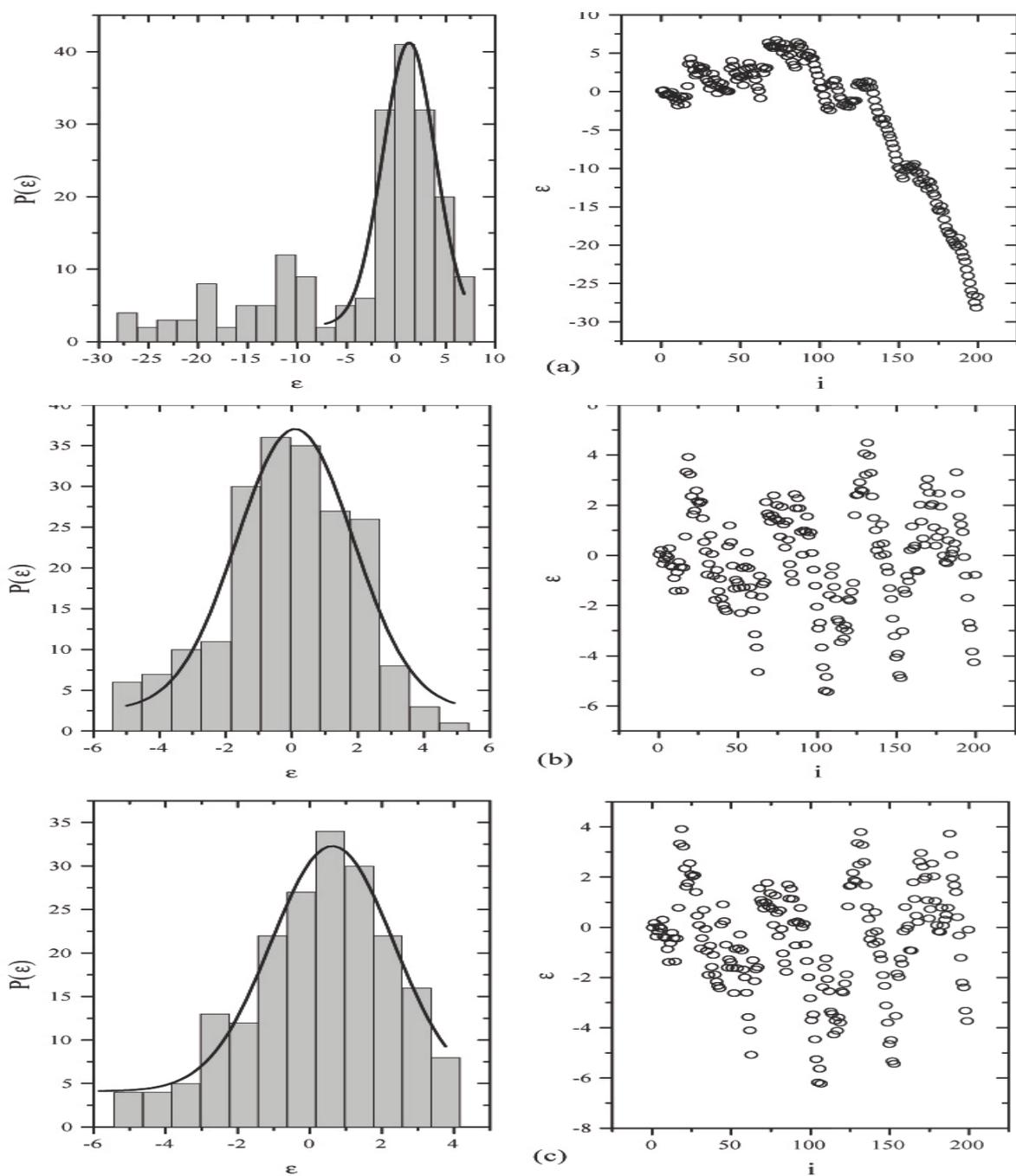


Figura 7.22: Sistema: Remédios. Método: máximo R^2 . Gráficos $P(\epsilon)$ por ϵ (frequência de erros) e ϵ por i (aleatoriedade) para as curvas de ajuste a) *q*-exponencial, b) *stretched*, c) *q*-*stretched*.

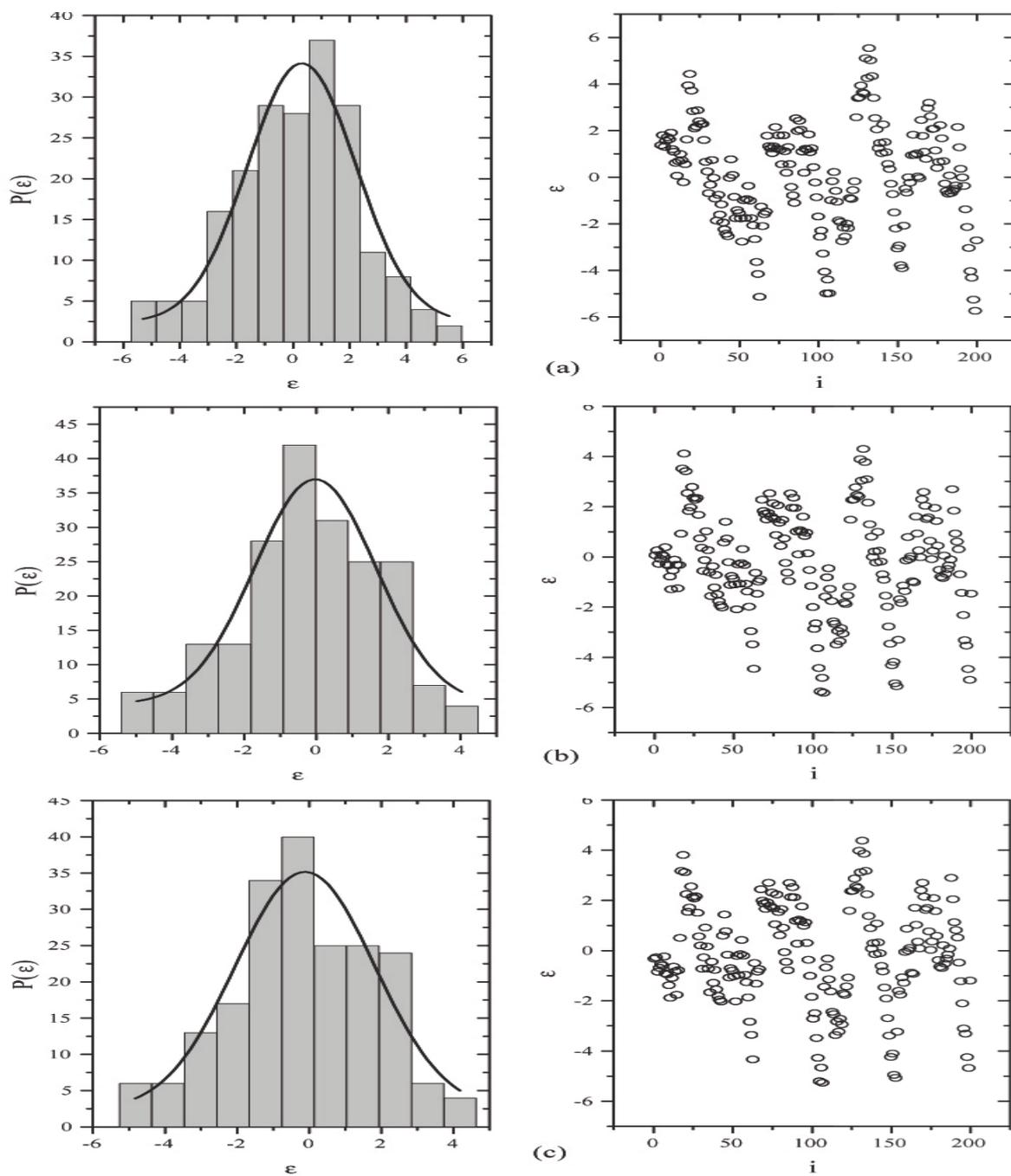


Figura 7.23: Sistema: Remédios. Método: mínimos quadrados. Gráficos $P(\varepsilon)$ por ε (frequência de erros) e ε por i (aleatoriedade) para as curvas de ajuste a) q -exponencial, b) $stretched$, c) q - $stretched$.

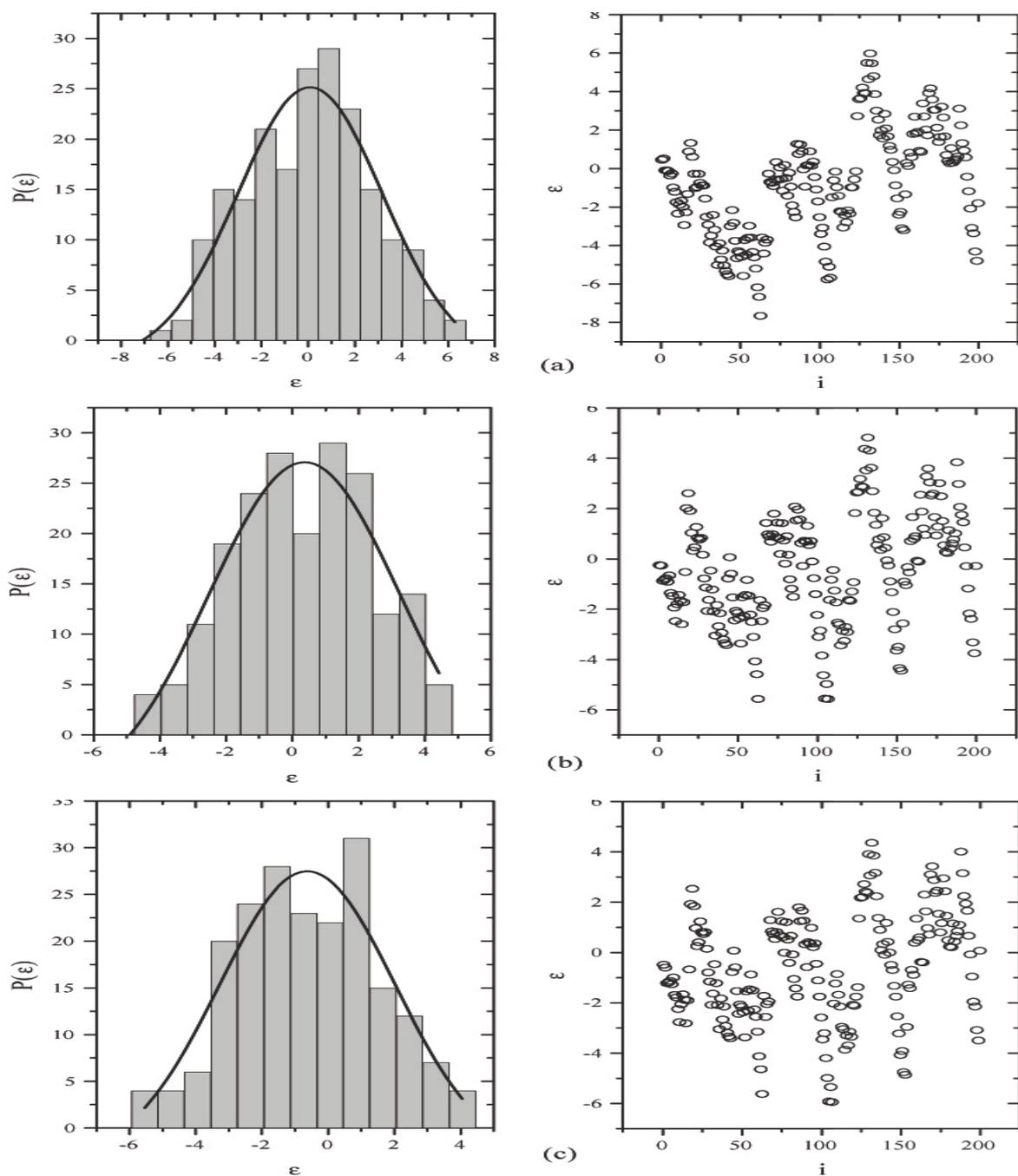


Figura 7.24: Sistema: Remédios. Método: máxima verossimilhança. Gráficos $P(\varepsilon)$ por ε (frequência de erros) e ε por i (aleatoriedade) para as curvas de ajuste a) q -exponencial, b) $stretched$, c) q - $stretched$.

7.3 Dados e aleatoriedade

Para concluirmos este capítulo, gostaríamos de apresentar uma pequena discussão sobre um conjunto genérico de dados e como eles podem ser gerados, relacionando isso ao ajuste por uma curva lisa. Ao mesmo tempo, desejamos que este estudo tenha uma ligação com as formas das distribuições de ajuste empregadas neste trabalho e apresente a maior simplicidade possível. Sob este enfoque, a distribuição interessante a ser investigada é a exponencial, visto que as demais empregadas neste trabalho, em algum sentido, a generalizam.

Em geral, espera-se que um sistema complexo possa ser modelado, pelo menos parcialmente, por algum sistema de equações. A depender do modelo, as equações podem ser determinísticas ou conter variáveis aleatórias (como é o caso da equação de Langevin). Assim, quando supomos que os dados relativos a algum sistema podem ser ajustados por uma curva suave, as flutuações em torno desta curva podem ser basicamente oriundas de duas fontes: erros na aquisição dos dados ou algum tipo de aleatoriedade presente no sistema. Focalizaremos nossa atenção na segunda possibilidade. Para tal consideraremos o caso extremo no qual os dados de um hipotético sistema obedecem a uma lei totalmente probabilística, através de uma distribuição exponencial.

Concretamente, geramos números reais positivos, x , com a probabilidade

$$p(x) = a \exp(-a x) , \quad (7.2)$$

sendo $a = 2$. É interessante observar que a partir de uma realização de tal distribuição de probabilidade não obtemos, em geral, $a = 2$. Sendo que os desvios em torno de $a = 2$ são mais pronunciados nos conjuntos de dados menores. Isto está claramente ilustrado na figura (7.10), onde consideramos um conjunto de 10 realizações para conjuntos de 100 e 1000 números. Tais realizações foram obtidas usando o gerador de números aleatórios do programa *Mathematica*[72]. Para ajustar os dados, empregamos o método da máxima verossimilhança. Poderíamos ter empregado qualquer outro método, mas isto não interfere nas nossas conclusões.

A partir da discussão prévia, vemos que desvios dos parâmetros de ajuste, que advém da própria natureza do sistema, podem ocorrer. Isto porque fontes de aleatoriedade podem estar presentes em maior ou menor grau na descrição do sistema. Assim, este aspecto

aumenta a dificuldade na escolha da melhor curva de ajuste, bem como a obtenção de valores precisos para os seus parâmetros. Em particular, isto torna-se mais pronunciado quando temos apenas uma realização do conjunto de dados. Devemos considerar a possibilidade de os conjuntos de dados analisados neste trabalho apresentarem, em certo grau, um comportamento aleatório. Isso, aliado ao fato de que a maioria dos sistemas analisados neste trabalho não estão replicados, ou seja, só apresentam uma realização, gera uma incerteza extra no ajuste de seus parâmetros, assim como na escolha da melhor curva de ajuste.

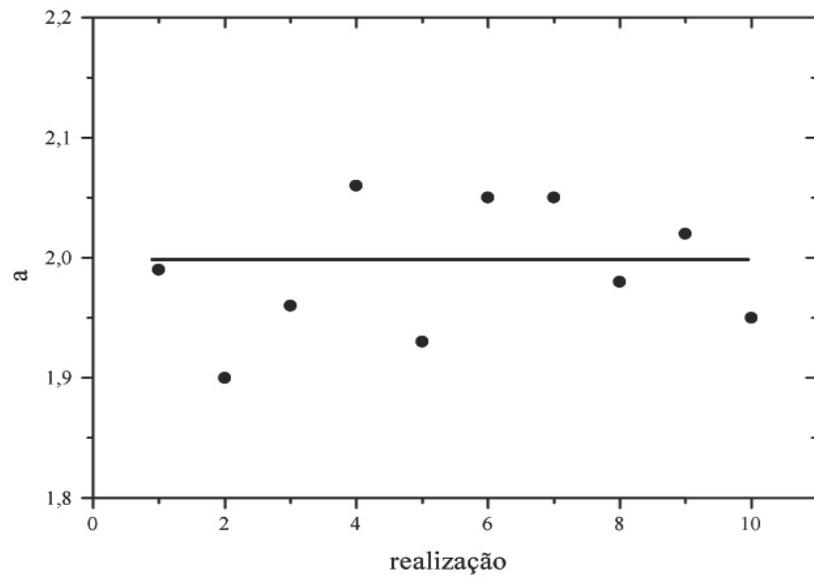
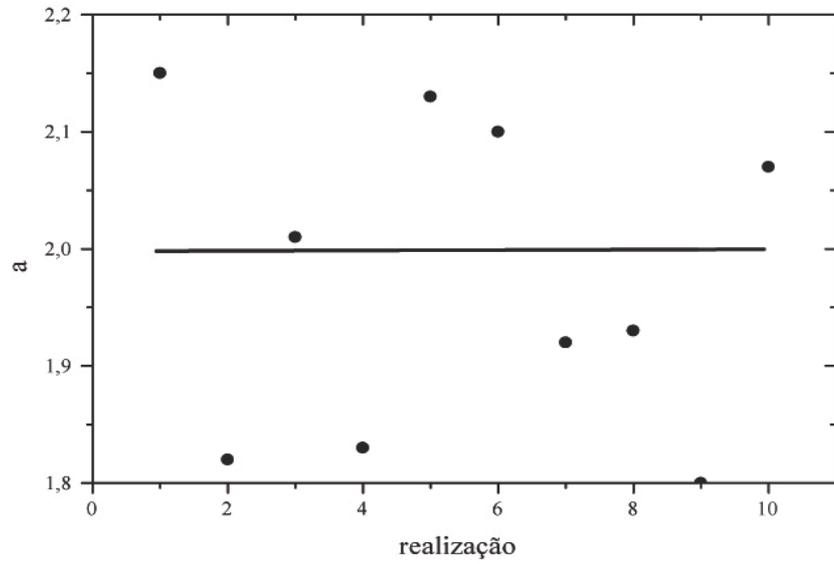


Figura 7.25: Gráficos de a (estimado) por n , onde n representa a n -ésima realização, para conjuntos de (a) 100 números e (b) 1000 números.

Capítulo 8

Conclusão

Em nossa análise estatística de dados, provenientes de sistemas complexos, identificamos regularidades que se manifestam na forma relativamente suave das distribuições de frequência acumulada.

Por exemplo, conforme vimos no capítulo 2, distribuições do tipo leis de potência descrevem relativamente bem os dados dos sistemas de comunicação analisados neste trabalho. O fato de leis de potência apresentarem-se como retas em gráficos log-log da distribuição de frequência facilita seu ajuste. Devido a isso, para definir a qualidade das retas de ajuste, nos baseamos apenas nos valores do coeficiente de correlação linear R^2 e nos valores de S/M . Os valores de R^2 próximos à unidade e os relativamente baixos valores de S/M indicam bons ajustes por leis de potência.

Além disso, os valores do parâmetro μ , que caracteriza a potência (x^μ), sugerem um certo comportamento universal para estes sistemas de comunicação analisados. Os sistemas relacionados a revistas ($\mu \simeq 1,3$), jornais ($\mu \simeq 1,5$) e emissoras de rádio ($\mu \simeq 1,5$) parecem demonstrar mais fortemente tal tipo de comportamento, pois nestes casos o parâmetro μ assume valores próximos entre si. No entanto, este comportamento universal não é tão bem pronunciado no caso do sistema relacionado com assinantes de TV a cabo ($\mu \simeq 2,0$). Um dos motivos desta discrepância com os demais sistemas de comunicação talvez seja algum mecanismo mais seletivo presente no caso dos assinantes de TV a cabo. Realmente, o acesso a assinaturas de TV a cabo, geralmente, são financeiramente mais dispendiosas do que o acesso a revistas, jornais ou a emissoras de rádio.

Além destes sistemas de comunicação, analisamos também distribuições de frequência

provenientes de outros sistemas complexos. Da grande variedade de tais sistemas na natureza, escolhemos sistemas relacionados ao basquete, às rodovias, aos ciclones, aos aeroportos, aos vulcões e aos remédios. Nesta análise, empregamos a forma acumulada das distribuições q -exponencial ($\exp_{q'}(x/x_0)$), de Weibull ($\exp[(x/x_0)^r]$) e q -stretched ($\exp_{q'}[(x/x_0)^r]$). Usamos também três métodos de ajuste de curvas: mínimos quadrados, máximo R^2 e máxima verossimilhança.

Com relação ao basquete, obtivemos, do ajuste pela distribuição q -stretched, valores de r próximos da unidade. Isto poderia indicar uma certa tendência do sistema ser descrito por uma q -exponencial. Usando-se a q -exponencial, obtivemos valores de q' ($q < 1$) significativamente próximos entre si, o que sugere um bom ajuste. Outro fato a favor da q -exponencial tem a ver com os valores dos coeficientes R^2 , S/M e $(\ln L)/N$. Todos eles são melhores que os obtidos com a *stretched*-exponencial. Além disso, constatamos que os erros entre os dados e a q -exponencial se distribuem de uma forma muito próxima à distribuição normal e que apresentam uma boa aleatoriedade ao longo dos dados. Tudo isso nos leva a afirmar que, dentre as distribuições q -exponencial e de Weibull, a melhor distribuição de ajuste para as cestas de basquete é uma q -exponencial, com q menor que um.

No caso das marcas de remédios, os ajustes com a distribuição q -stretched forneceram valores de q' próximos à unidade, indicando assim uma tendência para a distribuição de Weibull. Obtivemos dos ajustes com a *stretched* exponencial valores de r muito próximos entre si. Os coeficientes R^2 , S/M e $(\ln L)/N$ superaram em qualidade os obtidos com a q -exponencial. Na análise de erros, constatamos que a *stretched* exponencial fornece a distribuição que melhor se ajusta à normal e apresenta a maior aleatoriedade. Portanto, a distribuição de arrecadação por marca de remédios é, seguramente, melhor descrita por uma distribuição de Weibull.

De modo análogo aos anteriores, analisamos a distribuição da extensão de rodovias. Ao usarmos a q -stretched para ajustar tais dados, percebemos que, diferentemente dos casos anteriores, os valores dos parâmetros q' e r não estavam próximos da unidade ($q' \sim 0,55$ e $r \sim 0,6$) e além disso apresentaram apenas uma pequena variação ao mudarmos o método de ajuste. Observamos também que a q -stretched forneceu, visualmente, o melhor ajuste. Os coeficientes R^2 , S/M e $(\ln L)/N$ mostraram-se significativamente melhores que das outras distribuições. Além disso, com respeito à distribuição de erros, nota-se claramente um certo comportamento assimétrico fornecido pelo ajuste às distribuições

q -exponencial e de Weibull. Em contraste, a distribuição de erros devido ao ajuste pela q -*stretched* pode ser bem identificada com uma distribuição normal, apresentando também uma boa aleatoriedade de erros. Tais constatações nos levam a afirmar que a distribuição de extensão de rodovias nos Estados Unidos apresenta-se como um caso intermediário entre a q -exponencial e a distribuição de Weibull, sendo melhor ajustada por uma q -*stretched*.

No caso da distribuição de vítimas de ciclones, percebemos que os valores de q' obtidos do ajuste pela q -*stretched* são próximos da unidade. Este comportamento é mais evidente no ajuste pelo método dos mínimos quadrados. Nos outros ajustes, por algum motivo, o parâmetro r assumiu valores muito baixos ($r < 0,1$), favorecendo a diminuição do valor de q' . Mesmo assim, tais valores de q' próximos da unidade sugerem uma tendência para a distribuição de Weibull. É digno de nota que no ajuste com a *stretched* exponencial o valor do parâmetro r foi o mesmo ao empregarmos os três métodos de ajuste. Sua distribuição de erros não apresenta graves assimetrias, sendo bem identificada por uma distribuição normal. Observamos, neste caso, uma notável aleatoriedade de erros, sem qualquer comportamento definido. Portanto, concluímos que a distribuição de vítimas em ciclones tropicais é melhor ajustada por uma distribuição de Weibull do que por uma q -exponencial.

No que se refere aos aeroportos, a análise com a distribuição q -*stretched* revelou uma variação significativa nos valores dos parâmetros q' e r ao usarmos os métodos de ajuste. O mesmo aconteceu com as outras duas distribuições. Esta instabilidade dos parâmetros, talvez devido à existência de muitas flutuações no sistema, causa uma certa indefinição na escolha da melhor curva de ajuste. Além disso, observamos, na análise dos coeficientes R^2 e S/M , uma certa indefinição entre a q -exponencial (melhor S/M) e a *stretched* exponencial (melhor R^2). Por outro lado, obtivemos valores mais coerentes usando a q -*stretched*. A análise dos erros revelou algumas assimetrias significativas na distribuição de erros da q -exponencial e da *stretched* exponencial. A q -*stretched* forneceu uma distribuição de erros mais simétrica e mais próxima da normal. Em vista destes resultados, podemos dizer que, dentre as distribuições analisadas, a q -*stretched* é a curva que melhor se ajusta à distribuição de passageiros em aeroportos, porém com um grau de discernimento menor do que aqueles dos exemplos anteriores.

A análise da distribuição de altura de vulcões com a q -*stretched* revelou, a exemplo dos aeroportos, uma variação nos parâmetros q' e r para diferentes métodos de ajuste.

Os parâmetros das demais distribuições apresentaram também uma certa variação. Os coeficientes obtidos também apresentaram uma certa incoerência, fornecendo um melhor valor de R^2 para a q -exponencial e um melhor valor de S/M para a *stretched* exponencial. Por outro lado, novamente os valores obtidos com a q -*stretched* foram melhores e mais coerentes (melhor valor para os três coeficientes). Neste caso também encontramos assimetrias na distribuição de erros, sendo menos pronunciadas na curva q -*stretched*. Sendo assim, dentre as distribuições aqui analisadas, a q -*stretched* pode ser considerada como melhor distribuição de ajuste para a altura dos vulcões.

A análise dos casos acima sugere que tais sistemas são bem ajustados por distribuições da família da exponencial. Em particular, não é claro, a priori, se um sistema é descrito por uma distribuição q -exponencial ou de Weibull. Isto, por sua vez, indica que, em determinadas situações, pode ser difícil discernirmos algo sobre os mecanismos básicos que norteiam os sistemas em estudo. Isso porque poderíamos, num primeiro momento, relacionar a q -exponencial à estatística de Tsallis e a distribuição de Weibull aos processos multiplicativos. Neste sentido, uma possível continuação deste trabalho seria uma investigação baseada no modelamento dos sistemas aqui analisados.

Apêndice A

Ajuste de mínimos quadrados para uma reta

Neste apêndice, exemplificamos o método dos mínimos quadrados no caso de um ajuste linear[36, ?]. Introduzimos também o conceito de coeficiente de correlação linear.

Consideremos, como exemplo do método dos mínimos quadrados, o tipo mais simples de curva de ajuste: a reta. Sua equação pode ser escrita como

$$y = A_0 + A_1x \tag{A.1}$$

ou alternativamente

$$x = B_0 + B_1y. \tag{A.2}$$

No caso da família de curvas ser representada por retas na forma (A.1) teremos

$$D_i = A_0 + A_1x_i - y_i \tag{A.3}$$

e o valor de $S = \sum_{i=1}^N D_i^2$ será mínimo quando

$$0 = \frac{\partial S}{\partial A_0} = \frac{\partial}{\partial A_0} \sum_{i=1}^N (A_0 + A_1x_i - y_i)^2 \tag{A.4}$$

e

$$0 = \frac{\partial S}{\partial A_1} = \frac{\partial}{\partial A_1} \sum_{i=1}^N (A_0 + A_1 x_i - y_i)^2. \quad (\text{A.5})$$

Essas expressões fornecem as equações

$$\sum_{i=1}^N y_i = A_0 N + A_1 \sum_{i=1}^N x_i \quad (\text{A.6})$$

e

$$\sum_{i=1}^N x_i y_i = a_0 \sum_{i=1}^N x_i + a_1 \sum_{i=1}^N x_i^2, \quad (\text{A.7})$$

cuja resolução simultânea nos dá

$$A_0 = \frac{(\sum_{i=1}^N y_i)(\sum_{i=1}^N x_i^2) - (\sum_{i=1}^N x_i)(\sum_{i=1}^N x_i y_i)}{N \sum_{i=1}^N x_i^2 - (\sum_{i=1}^N x_i)^2} \quad (\text{A.8})$$

e

$$A_1 = \frac{N \sum_{i=1}^N x_i y_i - (\sum_{i=1}^N x_i)(\sum_{i=1}^N y_i)}{N \sum_{i=1}^N x_i^2 - (\sum_{i=1}^N x_i)^2}. \quad (\text{A.9})$$

Visando simplificar a notação, costuma-se fazer uma transformação de coordenadas: $X = x - \bar{x}$ e $Y = y - \bar{y}$. Fazendo-se uso destas novas coordenadas e dos valores de A_0 e A_1 , a equação (A.1) pode ser escrita na forma

$$Y = A_1 X = \frac{\sum_{i=1}^N X_i Y_i}{\sum_{i=1}^N X_i^2} X. \quad (\text{A.10})$$

Analogamente, se as distâncias D_i fossem medidas horizontalmente e usássemos a família de equações (A.2), obteríamos

$$X = B_1 Y = \frac{\sum_{i=1}^N X_i Y_i}{\sum_{i=1}^N Y_i^2} Y, \quad (\text{A.11})$$

com

$$B_0 = \frac{(\sum_{i=1}^N x_i)(\sum_{i=1}^N y_i^2) - (\sum_{i=1}^N y_i)(\sum_{i=1}^N x_i y_i)}{N \sum_{i=1}^N y_i^2 - (\sum_{i=1}^N y_i)^2} \quad (\text{A.12})$$

e

$$B_1 = \frac{N \sum_{i=1}^N x_i y_i - (\sum_{i=1}^N x_i)(\sum_{i=1}^N y_i)}{N \sum_{i=1}^N y_i^2 - (\sum_{i=1}^N y_i)^2}. \quad (\text{A.13})$$

Percebe-se que as retas de ajuste obtidas pelo método dos mínimos quadrados, (A.10)

e (A.11), passam pelo ponto (\bar{x}, \bar{y}) .

É digno de nota que as equações de ajuste (A.10) e (A.11) serão idênticas se e somente se todos os pontos do diagrama de dispersão caírem sobre uma reta, ou seja, quando $A_1 B_1 = 1$. Entretanto, podemos escrever

$$A_1 B_1 = \frac{\sum_{i=1}^N X_i Y_i}{\sum_{i=1}^N X^2} \times \frac{\sum_{i=1}^N X_i Y_i}{\sum_{i=1}^N Y^2} = R^2, \quad (\text{A.14})$$

sendo

$$R = \frac{\sum_{i=1}^N X_i Y_i}{[(\sum_{i=1}^N X^2)(\sum_{i=1}^N Y^2)]^{\frac{1}{2}}}. \quad (\text{A.15})$$

Portanto, $R = \pm 1$ indica que os dados são exatamente descritos por uma reta. Neste caso, $R = 1$ (-1) corresponde a $A_1 > 0$ ($A_1 < 0$). No entanto, quando os dados não se ajustam perfeitamente a uma reta, as inclinações A_1 e B_1 conduzem a $|R| < 1$. Em particular, se $R = 0$ as retas (A.10) e (A.11) são perpendiculares ($A_1 B_1 = 0$). Assim, o parâmetro R fornece uma medida do afastamento angular entre as duas retas de ajuste. É interessante notar que a quantidade R conserva-se a mesma, quer se considere x ou y como variável independente.

Uma forma compacta para R é obtida quando escrevemos

$$s_{xy} = \frac{\sum_{i=1}^N X_i Y_i}{N}, \quad s_x = \left(\frac{\sum_{i=1}^N X_i^2}{N} \right)^{\frac{1}{2}} \quad \text{e} \quad s_y = \left(\frac{\sum_{i=1}^N Y_i^2}{N} \right)^{\frac{1}{2}}, \quad (\text{A.16})$$

onde s_{xy} é conhecido como covariância de x e y e s_x e s_y são os desvios padrões das variáveis x e y . Adotando estes símbolos, a equação (A.15) pode ser escrita, de forma simplificada, como

$$R = \frac{s_{xy}}{s_x s_y}. \quad (\text{A.17})$$

Apêndice B

Linearização e maximização de R^2

Neste apêndice, exemplificamos como ajustar um conjunto de dados maximizando o coeficiente de correlação linear R^2 .

Consideremos inicialmente a q -exponencial

$$y = 10 \exp_{1,5}(-x), \quad (\text{B.1})$$

que está devidamente definida no apêndice (D). Aplicando-se o q -logaritmo (também definido no apêndice (D)), $\ln_q(x)$, com $q = 1,5$, em ambos os lados de (B.1), obtemos

$$\ln_{1,5} y = \ln_{1,5} 10 - 0,5 \ln_{1,5} 10 x. \quad (\text{B.2})$$

Notamos que x e $\ln_{1,5} y$ estão linearmente relacionados, o que não ocorre com x e y . Vemos isso mais claramente na figura (B.1 a). Atentemos para o fato que uma escolha apropriada do valor de q possibilita, neste caso, uma linearização perfeita ($R^2 = 1$ quando $q = 1,5$), indicando que a melhor reta de ajuste é obtida com $q = 1,5$. No entanto, se q tivesse um valor diferente de $1,5$ obteríamos, simplesmente, $R^2 < 1$, conforme ilustrado nas figuras (B.1 b) e (B.1 c).

Consideremos agora uma *stretched* exponencial (veja novamente o apêndice (D)) com $r = 0,5$

$$y = 10 \exp \left[- \left(\frac{x}{2} \right)^{0,5} \right]. \quad (\text{B.3})$$

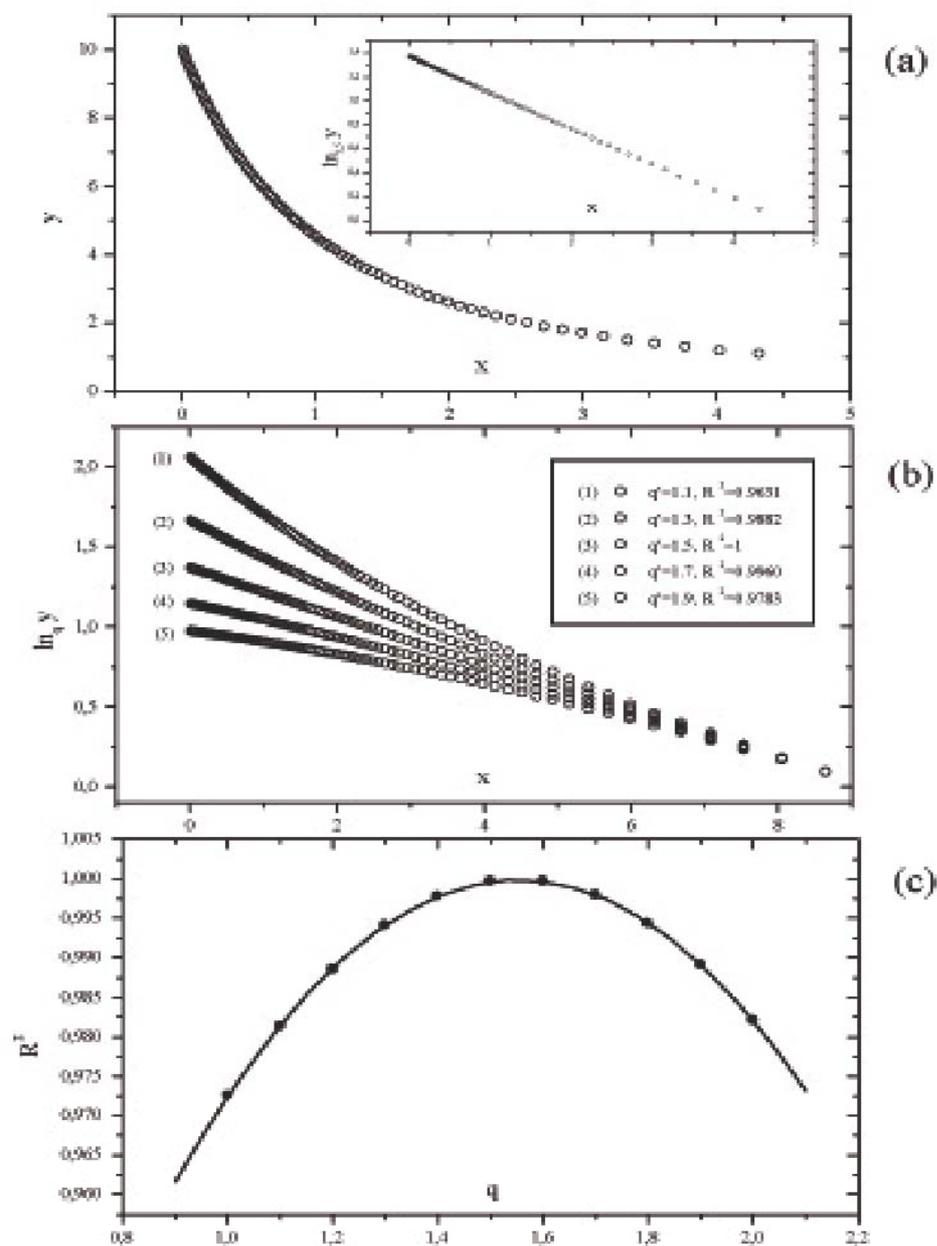


Figura B.1: (a) - Gráfico de y por x , com $y = 10 \exp_{1.5}(-x)$. No detalhe: Gráfico de $\ln_{1.5} y$ por x . (b) - Gráfico de $\ln_q y$ por x , com $q = 1, 1, 1, 3, 1, 5, 1, 7$ e $1, 9$, e seus respectivos valores de R^2 . (c) - O comportamento de R^2 , para o exemplo da figura (b), ao tomarmos q no intervalo $1 < q < 2$.

Tomando-se o logaritmo em ambos os lados de (B.3), obtemos

$$\ln y = \ln 10 - 2^{-0,5} x^{0,5}. \quad (\text{B.4})$$

Tal mudança de variáveis possibilita a linearização perfeita ($R^2 = 1$) da relação (B.3), fato ilustrado na figura (B.2 a). Se tomarmos x^r com $r \neq 0,5$ também observaremos decréscimos no valor de R^2 similares aos encontrados no exemplo anterior, conforme vemos nas figuras (B.2 b) e (B.2 c).

Como último exemplo, consideremos o caso

$$y = 10 \exp_{1,5} \left[- \left(\frac{x}{2} \right)^{0,5} \right], \quad (\text{B.5})$$

um tipo de mistura entre os dois exemplos anteriores. Assim como no primeiro exemplo, ao tomarmos o q -logaritmo em ambos os lados da equação (B.5), obtemos a relação

$$\ln_{1,5} y = - \frac{0,5 \ln_{1,5} 10}{2^{0,5}} x^{0,5} + \ln_{1,5} 10. \quad (\text{B.6})$$

Percebemos que o gráfico assume um comportamento linear perfeito ($R^2 = 1$), conforme visualizamos na figura (B.3 a). No entanto, se tomarmos $\ln_q(y)$ com $q \neq 1,5$ e/ou x^r com $r \neq 0,5$ obteremos novamente $R^2 < 1$ perdendo assim o comportamento linear, conforme enfatizado nas figuras (B.3 b) e (B.3 c).

Uma vez encontrados os parâmetros que maximizam o coeficiente R^2 , o problema se reduz a um ajuste linear de dados. Usamos então outro método, os mínimos quadrados, por exemplo, para obter a reta de ajuste. Esta, por sua vez, é usada na obtenção dos outros parâmetros de ajuste da distribuição.

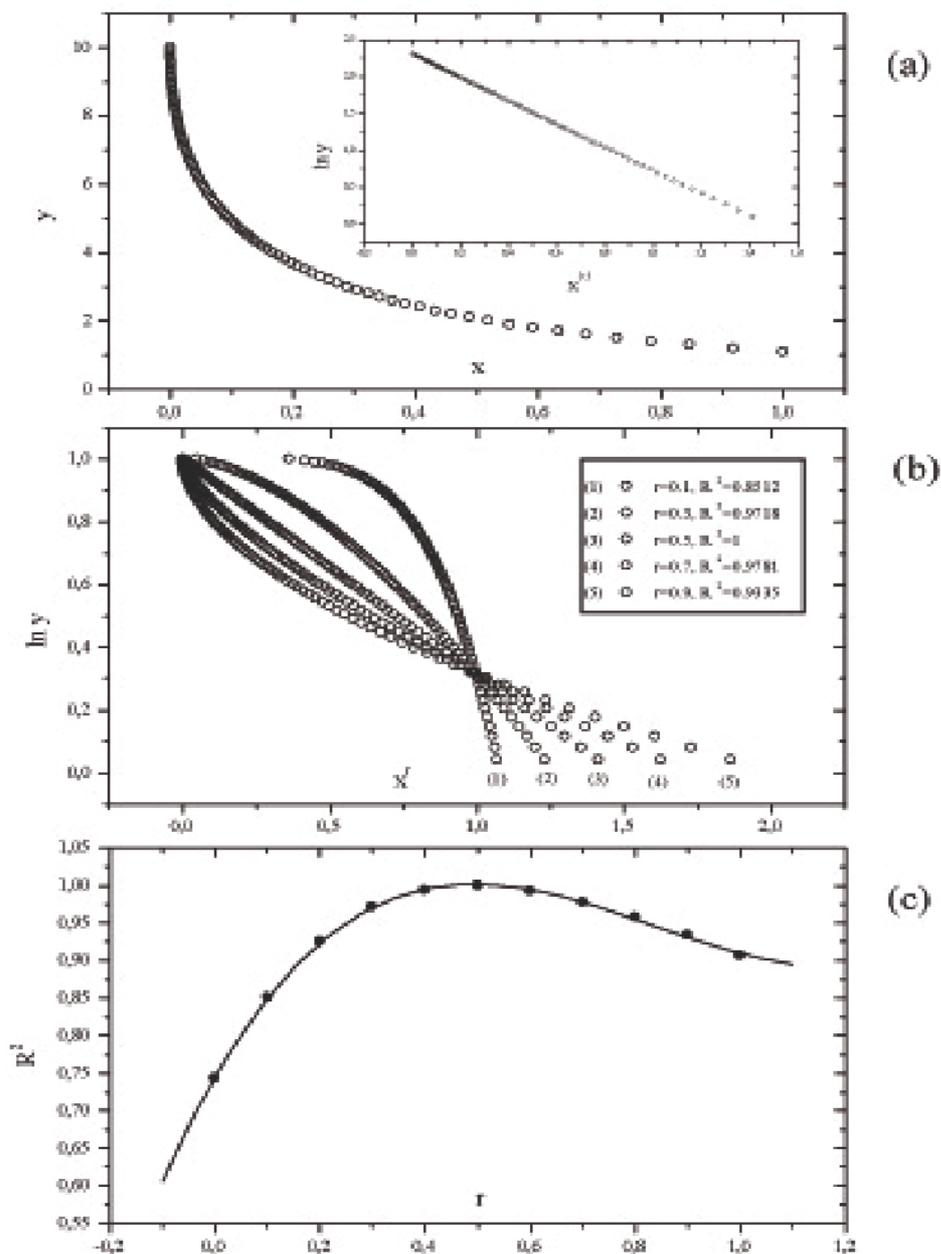


Figura B.2: (a) - Gráfico de y por x , com $y = 10 \exp[-(x/2)^{0.5}]$. No detalhe: Gráfico de $\ln(y)$ por $x^{0.5}$. (b) - Gráfico de $\ln y$ por x^r , com $r = 0, 1, 0, 3, 0, 5, 0, 7$ e $0, 9$, e seus respectivos valores de R^2 . (c) - O comportamento de R^2 , para o exemplo da figura (b), ao tomarmos r no intervalo $0 < r < 1$.

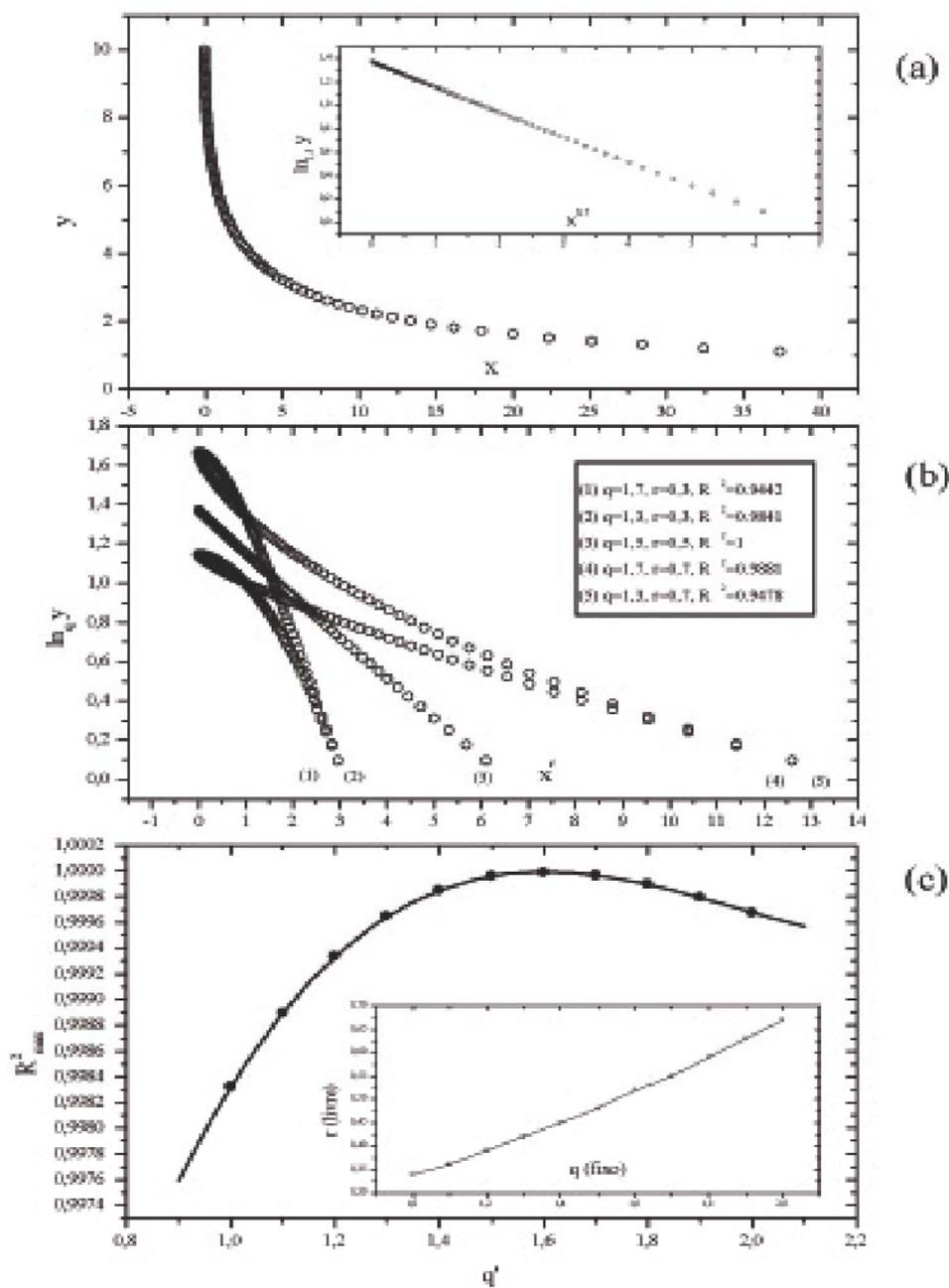


Figura B.3: (a) - Gráfico de y por x , com $y = 10 \exp_{1,5}[-(x/2)^{0,5}]$. No detalhe: Gráfico de $\ln_{1,5} y$ por $x^{0,5}$. (b) - Gráfico de $\ln_q y$ por x^r , com $q = 1,3$ e $r = 0,3$, $q = 1,3$ e $r = 0,7$, $q = 1,5$ e $r = 0,5$, $q = 1,7$ e $r = 0,7$ e $q = 1,7$ e $r = 0,3$, bem como seus respectivos valores de R^2 . (c) - O comportamento de R^2 , para o exemplo da figura (b), ao tomarmos valores fixos de q no intervalo $1 < q < 2$, com o respectivo r que conduz a R^2 máximo.

Apêndice C

Mecânica estatística de Tsallis: uma breve introdução

Apresentamos aqui breves comentários sobre a mecânica estatística de Boltzmann-Gibbs, desde a definição de entropia, a entropia de Boltzmann-Gibbs-Shannon, até a distribuição de probabilidades canônica. A seguir, fazemos uma breve revisão sobre o formalismo da mecânica estatística não extensiva de Tsallis. A partir da definição de uma nova forma para a entropia, a entropia de Tsallis, a distribuição q -exponencial surge como uma generalização da distribuição canônica exponencial.

C.1 A distribuição de Boltzmann-Gibbs

A escolha de uma forma entrópica S e de um conjunto de vínculos fisicamente apropriados pode ser usada como um método para se estudar a mecânica estatística de equilíbrio bem como para introduzir uma nova mecânica estatística. Tal fato, vinculado ao postulado de que em qualquer estado final de equilíbrio termodinâmico a entropia deve ser máxima, conduz a um princípio variacional do qual podemos obter as probabilidades relativas aos diversos estados do sistema.

Neste contexto, a entropia de Boltzmann-Gibbs-Shannon (B-G-S)

$$S = -k \sum_{i=1}^W p_i \ln p_i, \quad (\text{C.1})$$

é de vital importância na mecânica estatística usual, onde p_i é o conjunto de probabilidades associadas aos W estados acessíveis ao sistema em estudo e k é a constante de Boltzmann. Os valores assumidos por p_i , em princípio, variam de acordo com o estado i . O que define o modo como estes valores se distribuem são as condições macroscópicas as quais o sistema está submetido.

Quando o volume e o número de partículas do sistema permanecem constantes, mas deixamos livre a energia, porém fixando seu valor médio, temos o *ensemble* canônico. Neste caso, para maximizarmos a entropia acima em relação ao conjunto p_i devemos usar o vínculo (normalização das probabilidades)

$$1 = \sum_{i=1}^W p_i \quad (\text{C.2})$$

e o vínculo

$$u = \sum_{i=1}^W \epsilon_i p_i, \quad (\text{C.3})$$

que está relacionado ao valor médio da energia, onde ϵ_i é o valor da energia no i -ésimo estado. Utilizando o método dos multiplicadores de Lagrange, podemos obter o máximo da entropia (C.1) sujeita aos vínculos desejados (C.2) e (C.3). Para isso, construímos a função auxiliar

$$R_1(p_i) = - \sum_{i=1}^W p_i \ln p_i + \beta_1 \left(1 - \sum_{i=1}^W p_i \right) + \beta_2 \left(u - \sum_{i=1}^W \epsilon_i p_i \right), \quad (\text{C.4})$$

cuja extremização em relação a p_i nos permite escrever

$$\frac{dR_1(p_i)}{dp_j} = - \ln p_j - 1 - \beta_1 - \beta_2 \epsilon_j = 0, \quad (\text{C.5})$$

que implica

$$p_i = C \exp(-\beta \epsilon_i), \quad (\text{C.6})$$

com

$$C = \frac{1}{Z_1} = \frac{1}{\sum_{i=1}^W \exp(-\beta \epsilon_i)}. \quad (\text{C.7})$$

Associando β com a quantidade $\beta = 1/kT$, sendo T a temperatura do sistema, identificamos a equação (C.6) com a distribuição de Boltzmann, sendo Z_1 a função de partição do sistema.

É digno de nota que a expressão (C.3) representa a forma padrão para o cálculo do valor médio de qualquer observável arbitrário O na estatística usual (Boltzmann-Gibbs), isto é,

$$\langle O \rangle = \sum_{i=1}^W O_i p_i. \quad (\text{C.8})$$

C.2 A distribuição de Tsallis

Neste ponto estamos em condições de introduzir a mecânica estatística generalizada de Tsallis. Ele postula que a entropia generalizada[19] S_q está relacionada com as probabilidades p_i dos microestados por

$$S_q = k \frac{1 - \sum_{i=1}^W (p_i)^q}{q - 1}. \quad (\text{C.9})$$

A entropia S_q apresenta várias propriedades importantes tais como a positividade e a concavidade definida. Além disso, no limite $q \rightarrow 1$, verifica-se facilmente que S_q reduz-se à entropia de B-G-S, ou seja

$$\lim_{q \rightarrow 1} S_q = \lim_{q \rightarrow 1} k \frac{1 - \sum_{i=1}^W p_i \exp[(q - 1) \ln p_i]}{q - 1} = - \sum_{i=1}^W p_i \ln p_i = S_1, \quad (\text{C.10})$$

onde expandimos a exponencial em série de potências em $(q - 1)$.

No entanto S_q não é extensiva, mas de um modo geral, o caso $q \rightarrow 1$ recupera todo o formalismo usual da mecânica estatística, inclusive a extensividade. Tal fato é bem ilustrado ao considerarmos uma distribuição de probabilidades obtida a partir de dois

subsistemas A e B estatisticamente independentes

$$P_{ij}^{(AUB)} = P_i^{(A)} P_j^{(B)} \quad (\text{C.11})$$

com $i = 1, \dots, W_A$, $j = 1, \dots, W_B$ e $W = W_A W_B$. Usando a distribuição (C.11) e a entropia (C.9), verificamos que

$$S_q^{(AUB)} = S_q^{(A)} + S_q^{(B)} + (1 - q) S_q^{(A)} S_q^{(B)}. \quad (\text{C.12})$$

Desta forma, o parâmetro q pode ser considerado como uma medida da não extensividade da entropia S_q . Nota-se também facilmente que no caso de $q \rightarrow 1$, S_q recupera sua extensividade.

Neste ponto, podemos maximizar a entropia (C.9) mantendo o vínculo de normalização (C.2) inalterado, porém mudando o vínculo (C.3) para [19, ?]

$$u_q = \sum_{i=1}^W \epsilon_i p_i^q. \quad (\text{C.13})$$

Seguindo o mesmo procedimento empregado para a entropia de B-G-S, temos

$$R_q(p_i) = \frac{1 - \sum_{i=1}^W p_i^q}{q - 1} + \beta_1 \left(1 - \sum_{i=1}^W p_i \right) + \beta_2 \left(u_q - \sum_{i=1}^W \epsilon_i p_i^q \right) \quad (\text{C.14})$$

e

$$\frac{dR_q(p_i)}{dp_j} = \frac{q p_j^{q-1}}{q - 1} + \beta_1 + \beta_2 \epsilon_j q p_j^{q-1} = 0, \quad (\text{C.15})$$

conduzindo a

$$p_i = C_q [1 - (1 - q) \beta \epsilon_i]^{-\frac{1}{1-q}} \quad (\text{C.16})$$

com $C_q = [\beta_1(1 - q)/q]^{1/(1-q)}$ e $\beta = \beta_2$. Levando-se em conta a normalização, obtemos

$$C_q = \frac{1}{Z_q} = \frac{1}{\sum_{i=1}^W [1 - (1 - q) \beta \epsilon_i]^{-\frac{1}{1-q}}}. \quad (\text{C.17})$$

Em analogia à mecânica estatística usual, identificamos β com $\beta = 1/kT$ e notamos que u_q representa a energia interna generalizada do sistema. Assim, identificamos a equação (C.16) com a distribuição generalizada de Tsallis, sendo Z_q a função de partição

generalizada que garante a normalização das probabilidades.

Mais recentemente, tem sido considerado outro tipo de vínculo na forma[?]

$$u'_q = \frac{u_q}{\sum_{i=1}^W p_i^q}. \quad (\text{C.18})$$

Entretanto, para as discussões apresentadas neste trabalho, os pormenores que distinguem as mecânicas estatísticas generalizadas relacionadas aos vínculos u_q e u'_q não são relevantes.

Apêndice D

Generalizações da função exponencial

Consideramos neste apêndice algumas propriedades da função q -exponencial, que surge naturalmente na mecânica estatística não extensiva de Tsallis. Relacionado a esta função, introduzimos o q -logaritmo, uma generalização do logaritmo usual. Logo após citamos brevemente outra generalização da função exponencial, a *stretched* exponencial. Uma composição destas duas funções permite construir uma função mais geral, que preserva algumas propriedades de suas funções de origem.

D.1 A q -exponencial

A função conhecida como q -exponencial,

$$e_q^x = \exp_q(x) \equiv [1 + (1 - q)x]^{\frac{1}{1-q}}, \quad (\text{D.1})$$

é largamente usada neste trabalho. Ela emerge naturalmente no contexto da mecânica estatística não extensiva de Tsallis, generalizando o peso estatístico de Boltzmann-Gibbs $\exp(-\beta\epsilon)$ (ver equação (C.16)).

Em particular, a função q -exponencial reduz-se à exponencial (usual) no limite $q \rightarrow 1$.

Isso pode ser facilmente verificado escrevendo-se a função exponencial na forma

$$e^x = \exp(x) \equiv \lim_{\alpha \rightarrow \infty} \left(1 + \frac{1}{\alpha}x\right)^\alpha. \quad (\text{D.2})$$

Se considerarmos $\alpha = 1/(1 - q)$, o limite $\alpha \rightarrow \infty$ é equivalente ao limite $q \rightarrow 1$. Assim, fica claro que $e_1^x = e^x$.

Quando a função q -exponencial é usada na forma de distribuição de probabilidades, torna-se necessário fazer $e_q^x = 0$, se $1 + (1 - q)x < 0$. Este corte deve existir, pois, de outra forma, obteríamos resultados incompatíveis com a teoria das probabilidades ($p_i \propto \exp_q(-\beta\epsilon_i)$ deixaria de ser real).

Outra propriedade é que, para $q > 1$ e $x \gg 1$, a exponencial generalizada $\exp_q(-x)$ obedece a uma lei de potência simples, tornando-se proporcional a $x^{\frac{1}{1-q}}$. É justamente esta propriedade que permite conectar uma q -exponencial a uma lei de potência. Algumas dessas propriedades podem ser observadas diretamente da figura (D.1).

Assim como a q -exponencial,

$$\ln_q x \equiv \frac{x^{(1-q)} - 1}{1 - q} \quad (\text{D.3})$$

pode ser considerada como uma generalização para a função logarítmica, que de modo análogo, reduz-se ao logaritmo usual no limite $q \rightarrow 1$.

Esta função obedece a identidade

$$\ln_q AB = \ln_q A + \ln_q B + (1 - q) \ln_q A \ln_q B, \quad (\text{D.4})$$

que pode ser considerada como sendo uma generalização de $\ln AB = \ln A + \ln B$ (note que a identidade (D.4) está intimamente relacionada à identidade (C.12)). Além disso, observemos que

$$e_q^{(\ln_q x)} = \left[1 - (1 - q) \left(\frac{x^{1-q} - 1}{1 - q}\right)\right]^{\frac{1}{1-q}} = (x^{1-q})^{\frac{1}{1-q}} = x \quad (\text{D.5})$$

e

$$\ln_q e_q^x = \frac{\left([1 - (1 - q)x]^{\frac{1}{1-q}}\right)^{1-q} - 1}{1 - q} = \frac{(1 - q)x}{1 - q} = x. \quad (\text{D.6})$$

Assim, permanece válida a identidade

$$e_q^{(\ln_q x)} = \ln_q e_q^x, \quad (\text{D.7})$$

que pode ser encarada como uma generalização de $e^{(\ln x)} = \ln e^x$.

D.2 Outras generalizações

Além da q -exponencial, duas outras generalizações da exponencial serão usadas neste trabalho. Uma delas é a função

$$\exp(x^r), \quad (\text{D.8})$$

também conhecida como *stretched* exponencial (ou exponencial alongada). O comportamento da $\exp(-x^r)$ é ilustrado na figura (D.2).

Por fim, consideremos um tipo de composição entre a q -exponencial e a *stretched* exponencial, isto é,

$$e_q^{x^r} = \exp_q(x^r) \equiv [1 + (1 - q)x^r]^{\frac{1}{1-q}}. \quad (\text{D.9})$$

Tal função, embora tenha características mais gerais, apresenta vestígios do comportamento tanto da q -exponencial quanto da *stretched* exponencial. Algumas de tais características podem ser observadas na figura (D.3).

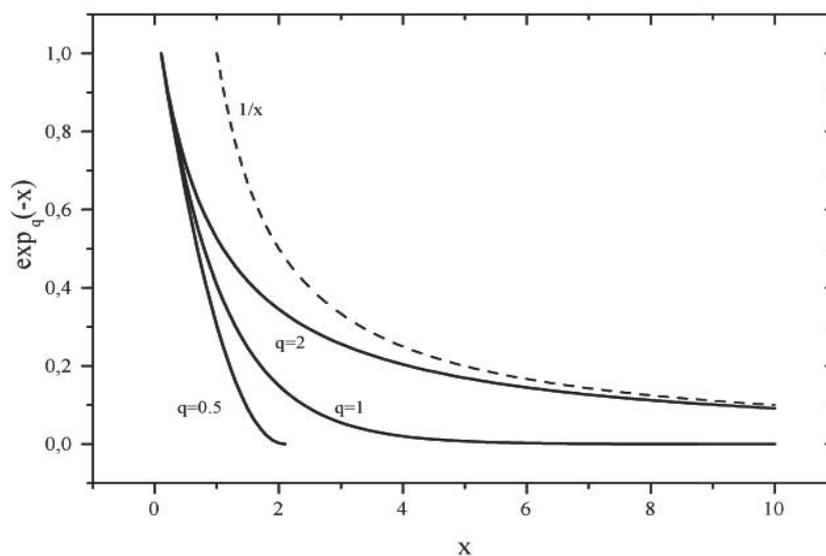


Figura D.1: A função $y = \exp_q(-x)$ para valores particulares de q . Em particular temos um corte em $x_c = 2$ para $q = 0.5$. Notemos que $\exp_2(-x) \sim x^{-1}$ para $x \gg 1$.

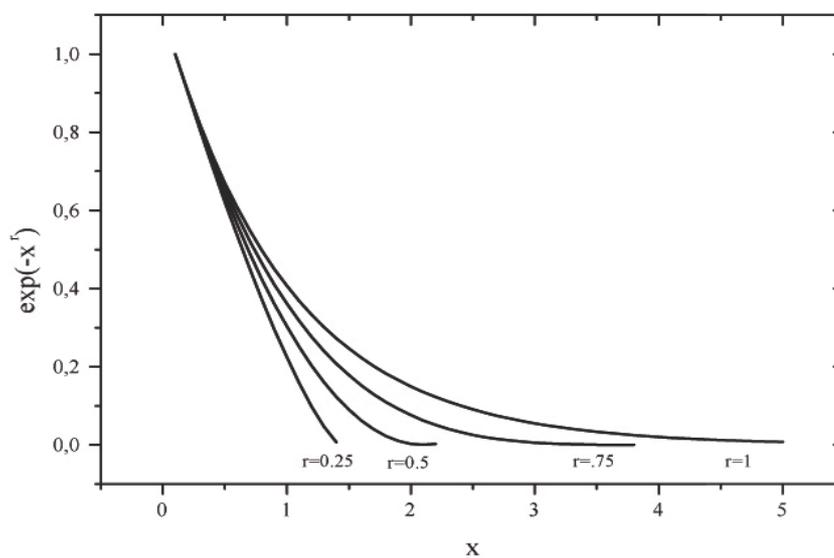


Figura D.2: A função $y = \exp(-x^r)$ para valores particulares de r .

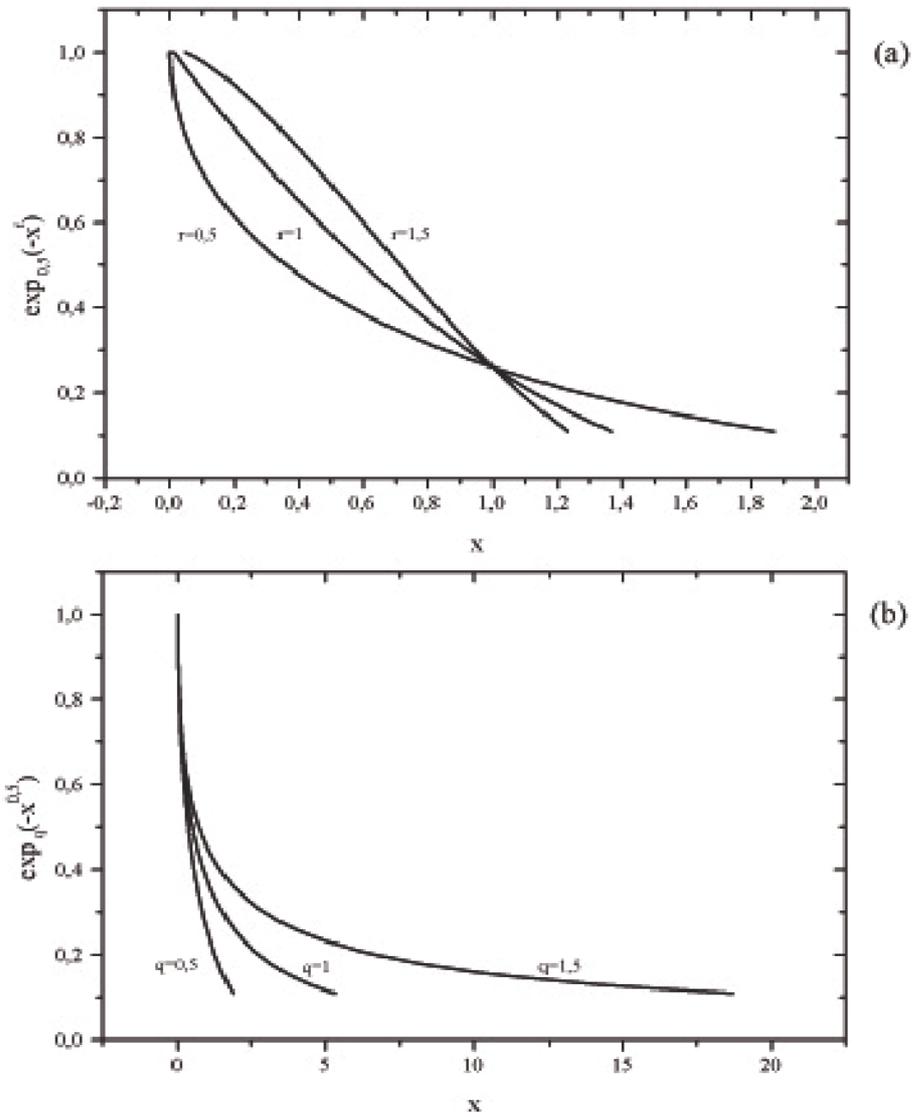


Figura D.3: (a) - A função $y = e_q^{-x^r}$, com $q = 0,5$ e valores particulares de r . (b) - A função $y = e_q^{-x^r}$ com $r = 0,5$ e valores particulares de q .

Bibliografia

- [1] P. Bak e M. Paczuski, *Phys World* 6: (12) 39-43 (1993)
- [2] V. Pareto, *Cours d'Economie Politique*, Lausanne and Paris: Rouge and Cie, (1897).
- [3] G. K. Zipf, *Human behavior and the principle of least effort* (Addison-Wesley, Cambridge-MA, 1949)
- [4] G. K. Zipf, *The psycho-biology of language*, M.I.T Press, Cambridge, 1965
- [5] A. H. Makse, S. Havlin e H. E. Stanley, *Nature* **377**, 608 (1995)
- [6] D. H. Zanette e S. C. Manrubia, *Phys. Rev. Lett.* **79**, 523 (1997)
- [7] M. Marsilli e Y. C. Zhang, *Phys. Rev. Lett.* **80**, 2741 (1998)
- [8] B. D. Malamud, G. Morein e D. L. Turcotte, *Science* **281**, 1840 (1998)
- [9] D. L. Gerlough e M. J. Huber, *Traffic flow teory* National Research Council, Washington, D.C., 1975.
- [10] S Redner, *Eur Phys J B* 4:(2) 131-134 (1998)
- [11] D. C. Roberts e D. L. Turcotte, *Fractals* 6 (1998) 351.
- [12] C. J. Rhodes e M. Anderson, *Nature* **381**, 600 (1996)
- [13] M. Levy e S. Solomon *Physica A* 242: (1-2) 90-94 (1997)
- [14] S. Miyazima, Y. Lee, T. Nagamine, et al. *Physica A* 278: (1-2) 282-288 (2000)
- [15] P. Bak, C. Tang, *J. Geophys*, 94 (1989). K. Ito, *Phys. Rev. E*, 52 (1995).

- [16] B. Mandelbrot */Betterworths* (1953)
- [17] B. Mandelbrot Freeman (1982)
- [18] *Nonextensive Statistical Mechanics and Its Applications* 2000
- [19] C. Tsallis *J Stat Phys* 52: (1-2) 479-487 (1988)
- [20] C. Tsallis *Braz J Phys* 29: (1) 1-35 (1999)
- [21] C. Tsallis, G. Bemsiki e R. S. Mendes, *Phys Lett A* 257: (1-2) 93-98 (1999)
- [22] C. Tsallis e M. P. de Albuquerque *Eur Phys J B* 13: (4) 777-780 (2000)
- [23] (Regularities in football goal distributions)
- [24] B. M. Boghosian, *Phys. Rev. E* **53** 4754 (1996)
- [25] A. R. Plastino e A. Plastino *Phys. Lett. A* **174** 384 (1993) e 193 251 (1994).
- [26] Y. H. Taguchi e H. Takayasu, *Europhys. Lett.* **30** 499 (1995)
- [27] C. Kaniadakis, A. Lavagno e P. Quarati, *Phys. Lett. B* **369**, 308 (1996)
- [28] A. R. Plastino e A. Plastino, *Physica A* **222**, 347 (1995); C. Tsallis e D. J. Bukman, *Phys. Rev. E* **54**, R2197 (1996); A. M. C. Souza e C. Tsallis, *Physica A* **236**, 52 (1997).
- [29] D. H. Zanette e P. A. Alemany, *Phys. Rev. Lett.* **75**, 366 (1995); C. Tsallis, S. V. F. Levy, A. M. C. de Souza e R. Maynard, *Phys. Rev. Lett.* **77**, 5442 (1996).
- [30] D. B. Ion e M. L. D. Ion, *Phys. Rev. Lett.* **81**, 5714 (1998).
- [31] C. Tsallis, A. R. Plastino e W. M. Zheng, *Chaos, Solitons and Fractals* **8**, 885 (1997); M. L. Lyra e C. Tsallis, *Phys. Rev. Lett.* **80**, 53 (1998).
- [32] F. A. Tamarit, S. A. Cannas e C. Tsallis, *Eur. Phys. J. B.* **1**, 545 (1998); A. R. R. Papa e C. Tsallis, *Phys. Rev. E* **57**, 3923 (1998).
- [33] W. Weibull *Ingeniors Vetenskaps Akademiens Handligar*, Stockholm 151 (1939)
- [34] W. Weibull *Ingeniors Vetenskaps Akademiens Handligar*, Stockholm 153 (1939)

- [35] J. Laherrere e D. Sornette *Eur Phys J B* 2: (4) 525-539 (1998)
- [36] Murray R. Spiegel, *Estatística*, McGraw-Hill, (1961)
- [37] Leo Breiman *Statistics: With a View Toward Applications*, Houghton Mifflin Company, (1973)
- [38] Jon Mathews e Robert L. Walker *Mathematical Methods of Physics*, (Addison-Wesley, California Institute of Tecnology, 1964)
- [39] P. J. Danaher *American Statistician* 46: (4) 254-260 (1992)
- [40] P. J Danaher *Journal of Business and Economic Statistics* 10: (4) 401-407 (1992)
- [41] M. G. Van Buer, D. L. Woodruff, R. T. Olson *European Journal of Operational Research* 115: (2) 237-253 (1999)
- [42] C. Ilie *Journal of Pragmatics* 33: (2) 209-254 (2001)
- [43] C. Vassilakis, M. Paterakis, P. Triantafillou *Multimedia Systems* 8: (2) 92-104 (2000)
- [44] www.adage.com/dataplace/archives/dp458.html
- [45] www.mediainfo.com/ephone/research/researchhtm/world100.htm
- [46] www.americantop40.com/rtsl-share.html
- [47] www.cablevisionmagazine.com/database/db/topsy/asp
- [48] M. Levy e S. Solomon, *Int J Mod Phys C* 7: (4) 595-601 (1996)
- [49] D. Sornette e R. Cont, *J. Phys. I France* **7**, 431 (1997).
- [50] D. Sornette *Phys Rev E* **57** 4811-4813 (1998)
- [51] T. R. Vollmer e J. Bourret *Journal of Applied Behavior Analysis* 33: (2) 137-150 (2000)
- [52] P. S. Gill *American Statistician* 54: (2) 94-99 (2000)
- [53] F. Mosteller *American Statistician* 51: (4) 305-310 (1997)

- [54] cbs.sportslines.com/u/basketball/stats/sfgpanba.htm
- [55] I. Thomas *Accident Analysis and Prevention* 28: (2) 251-264 (1996)
- [56] www.members.nbci.com/us98/ushwy/.htm
- [57] C. B. Chang *Meteorology And Atmospheric Physics* 66: (3-4) 215-227 (1998)
- [58] P. Maheras, H. A. Flocas, I. Patrikas, C. Anagnostopoulou *International Journal of Climatology* 21: (1) 109-130 (2001)
- [59] R. S. Cerveny e L. E. Newman *Monthly Weather Review* 128: (9) 3329-3336 (2000)
- [60] www.nhc.noaa.gov/pastdeadly1.html
- [61] J. I. Daniel *Econometrica* 63: (2) 327-370 (1995)
- [62] M. Hansen *Journal of Transportation Engineering-ASCE* 121: (6) 453-460 (1995)
- [63] www.airports.org/traffic/index.html
- [64] P. Diodati, F. Marchesoni e S. Piazza, *Phys. Rev. Lett.* **67**, 2239 (1991).
- [65] F. Mazzarini, P. Armienti *Bulletin of Volcanology* 62: (6-7) 420-430 (2001)
- [66] J. R. Grasso, P. Bachelery *Geophysical Research Letters* 22: (21) 2897-2900 (1995)
- [67] D. A. Noever, *Phys. Rev. E* **47**, 724 (1993).
- [68] www.volcano.si.edu/gvp/volcano/vbd_alph.htm
- [69] C. Huttin *Applied Economics* 29: (4) 497-503 (1997)
- [70] dt.pdr.net/dt/public.htm
- [71] U. Frisch e D. Sornette *J Phys I* 7: (9) 1155-1171 (1997)
- [72] S. Wolfram *The Mathematica* fourth edition